Some computational approaches to modelling the behaviour of metabolic systems

David Gilbert
drg@dcs.gla.ac.uk

Bioinformatics Research Centre University of Glasgow

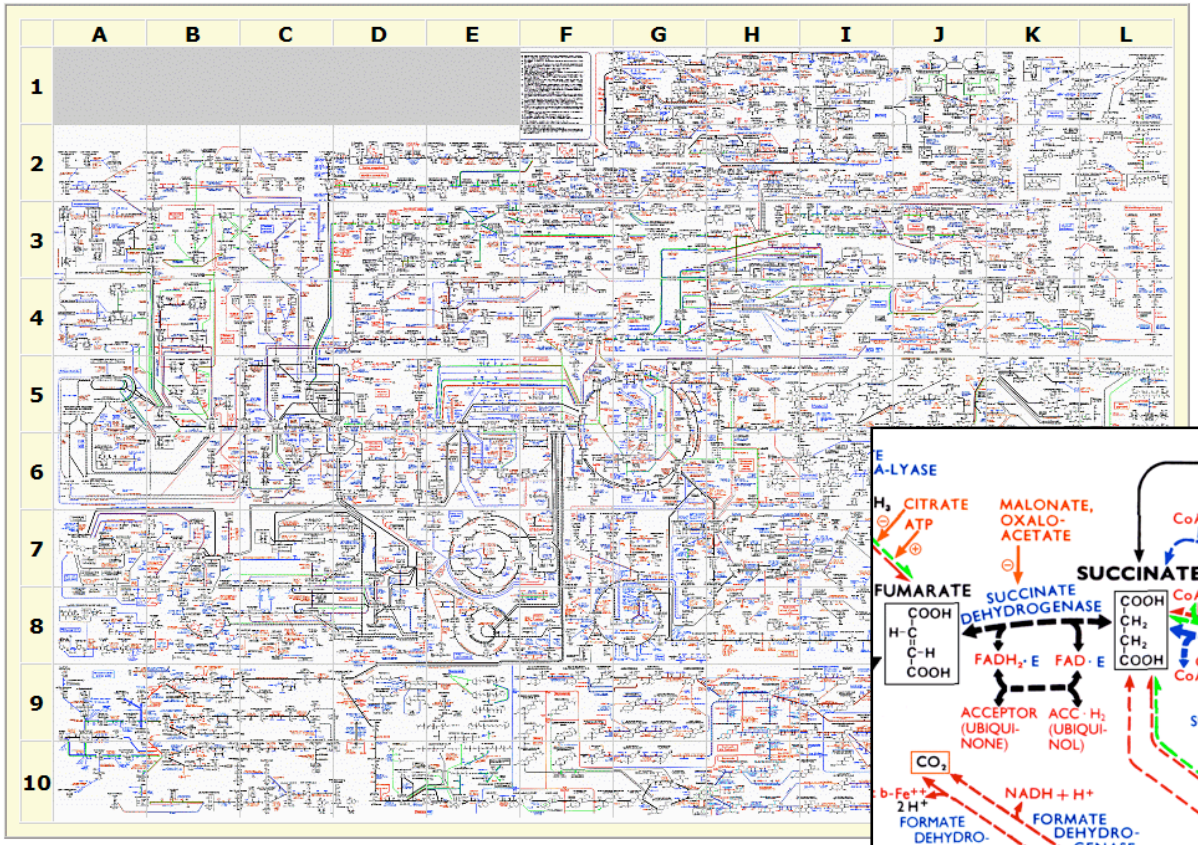
UNIVERSITY of GLASGOW

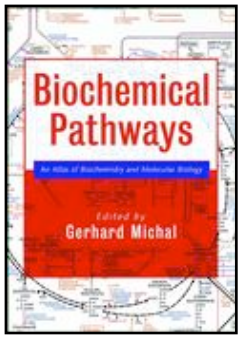
Outline

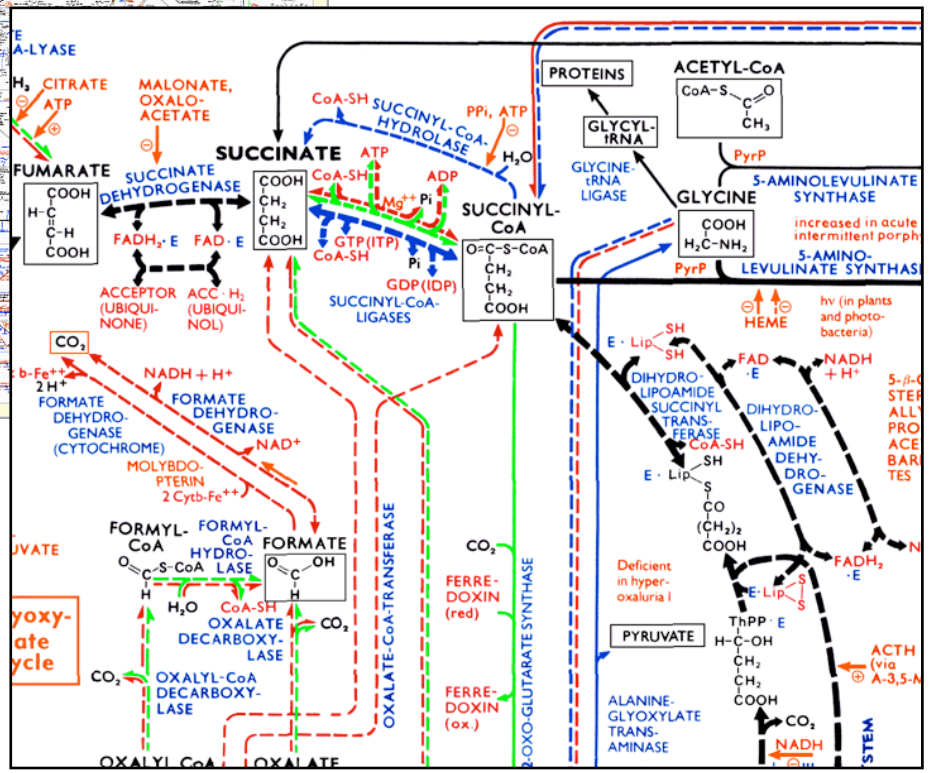
- Data models & databases
-
- Computations over static models
- Qualitative to quantitative
- Simulation
- Analysis
- Model checking

UNIVERSITY of GLASGOW

Metabolic Pathways







What can we do computationally?

- Generate / gather data
- Construct networks (various types)
  - Static
  - Dynamic
- Create databases of network data
- Display (visualise) network
- Analyse static network properties
  - Global, local, motifs, …
- Navigate the networks
  - Data queries e.g. pathfinding
- Simulate dynamic behaviour
- Compare networks (static, dynamic properties)
- Analyse dynamic properties
- Predict effects of interventions / re-engineering

UNIVERSITY of GLASGOW

Terminology: Pathways or Networks?

- Pathways implies ‘paths’ - sequences of objects
  - An ordered sequence of proteins and substrates
  - A series of biochemical reactions
  - An evolutionary product
  - A biological system (living cell)

- Networks - more complex connectivity

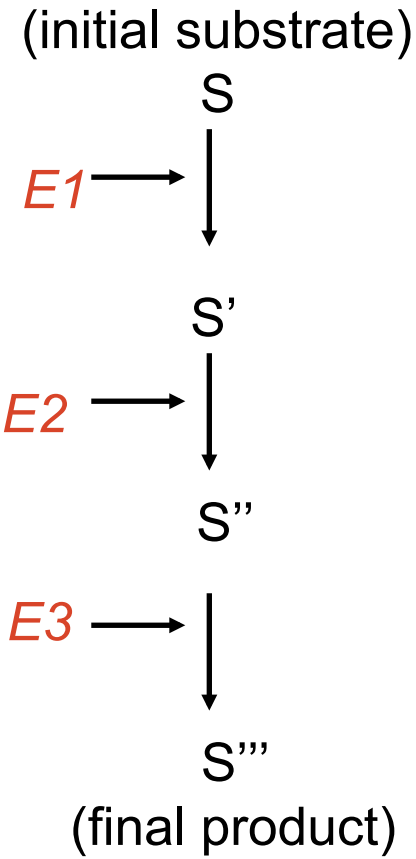
- Both are represented by ***graphs***

- Networks: generic; Pathways: specific (?)
  - ‘Metabolic networks’
  - ‘The glycolytic pathway’

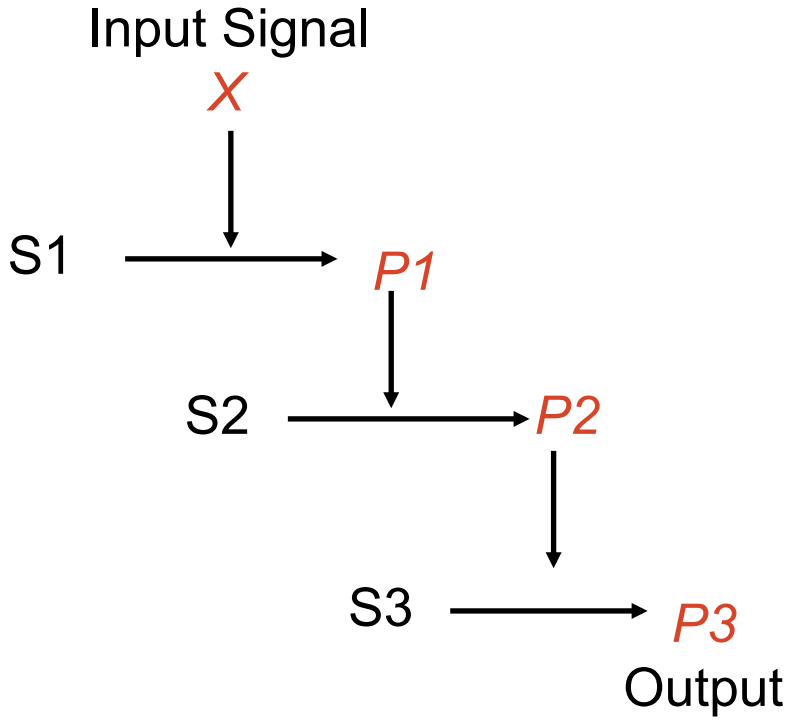
UNIVERSITY of GLASGOW

Metabolic pathways vs Signalling Pathways

Metabolic



Signalling cascade



Classical enzyme-product pathway

Product become enzyme at next stage

Enzymes are in RED

Database models

- Aim to represent data
  - to store them
  - to take advantage of the DBMS's data storage, management, and retrieval facilities
- Often unsuitable to analyse the structure of biochemical networks

Y. Deville, D. Gilbert, J. van Helden & S. Wodak. An Overview of Data Models for the Analysis of Biochemical Pathways, Briefings in Bioinformatics, 2003 4:3, 246-259

Graph-based data models for pathways

- Compound graph
- Reaction graph
- Bipartite graph
- Hypergraph
- Object-oriented models

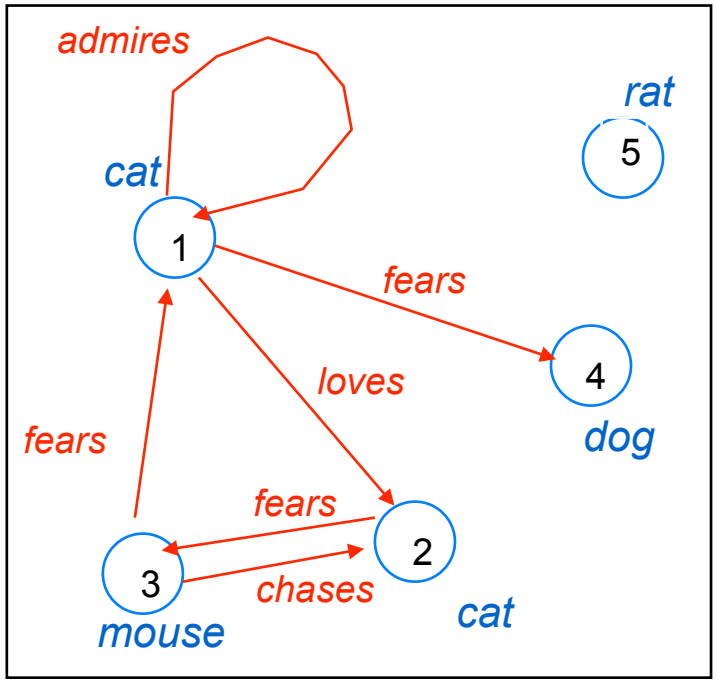
Y. Deville, D. Gilbert, J. van Helden & S. Wodak. An Overview of Data Models for the Analysis of Biochemical Pathways, Briefings in Bioinformatics, 2003 4:3, 246-259

Graphs

Graph = (V,A) V = set of vertices (nodes) A = set of arcs

A graph is either directed or not

If directed then A - arcs. If undirected then A - edges



G = (V,A)

V = { 1 , 2 , 3, 4 , 5 }

A = {1→2, 2→3, 3→2, 3→1, 1→4 , 1→1}

Paths (some)

P1 = (2→3, 3→1)

P1 = (2→3, 3→1, 1→4)

P3 = (2→3, 3→1, 1→1)

Circuits

C1 = (1→2, 2→3, 3→1) length = 3

C2 = (1→1) length = 1

Optionally label vertices & arcs

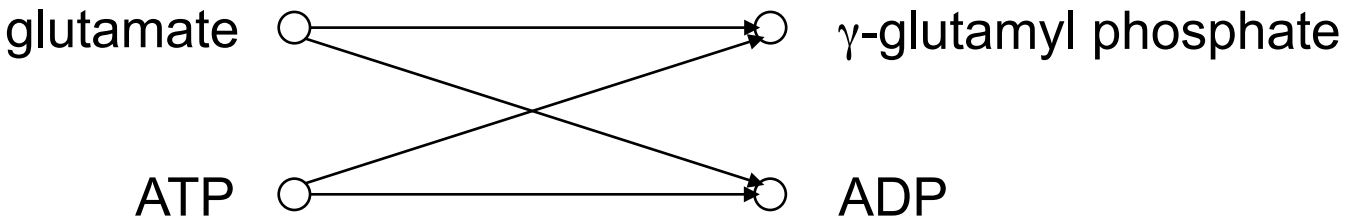
UNIVERSITY of GLASGOW

Compound graph

- To model (bio-)chemical reactions
- Nodes are (bio-)chemical compounds
- Directed edges connect compound A to compound B if A is a substrate and B is a product in the same reaction

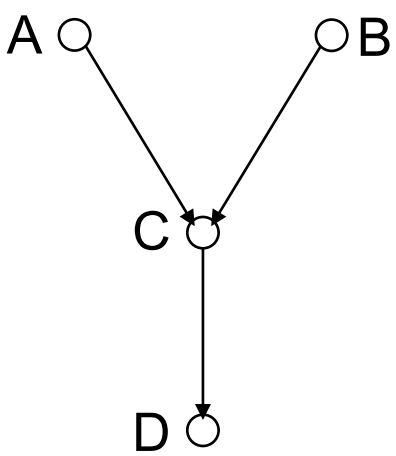
glutamate + ATP ⟶ γ-glutamyl phosphate + ADP

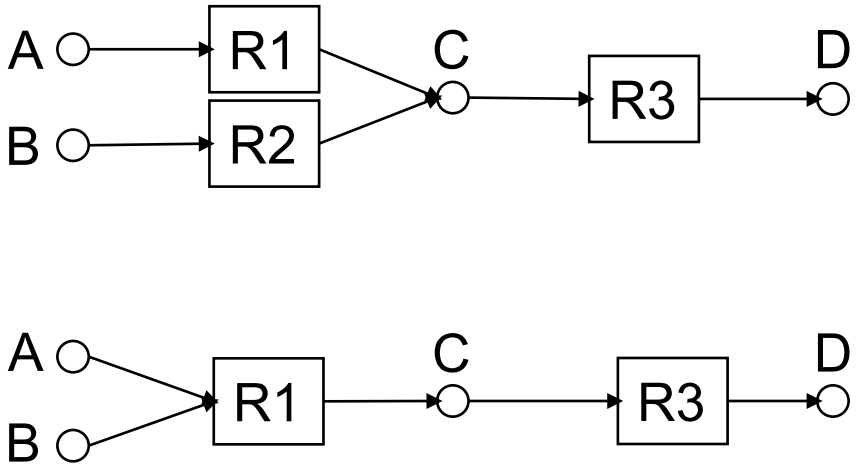
•Catalysed by γ-glutamyl kinase (EC 2.7.2.11)



Compound graph - problems

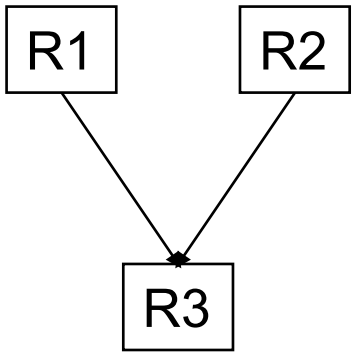
- Can be used to represent metabolic or regulatory pathways
- Can not be used to combine them
  - Would require different nodes for compound or genes
  - Different edges for chemical reactions or regulatory events
- Don’t contain information about the enzymes catalysing the reactions
- Ambiguous: different reactions can lead to the same graph

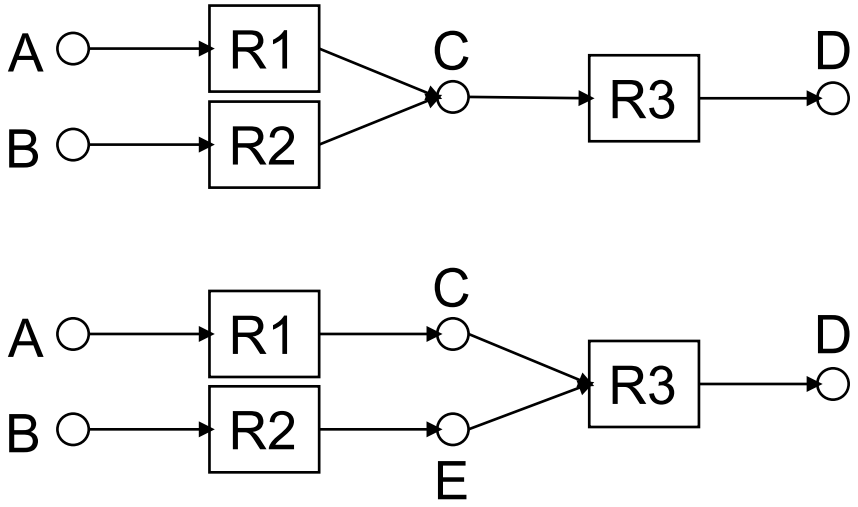




Reaction graph

- Nodes are (bio-)chemical reactions
- Edges are between nodes if there is a compound which is the product of one reaction and the substrate of a second
- Edges can be directed or undirected (if reactions are reversible)
- Similar limitations to compound graphs
- Ambiguous:



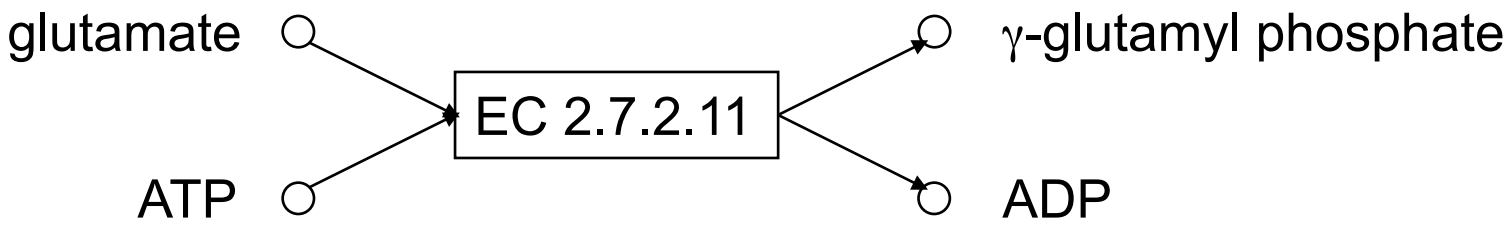


Why compound and reaction graphs?

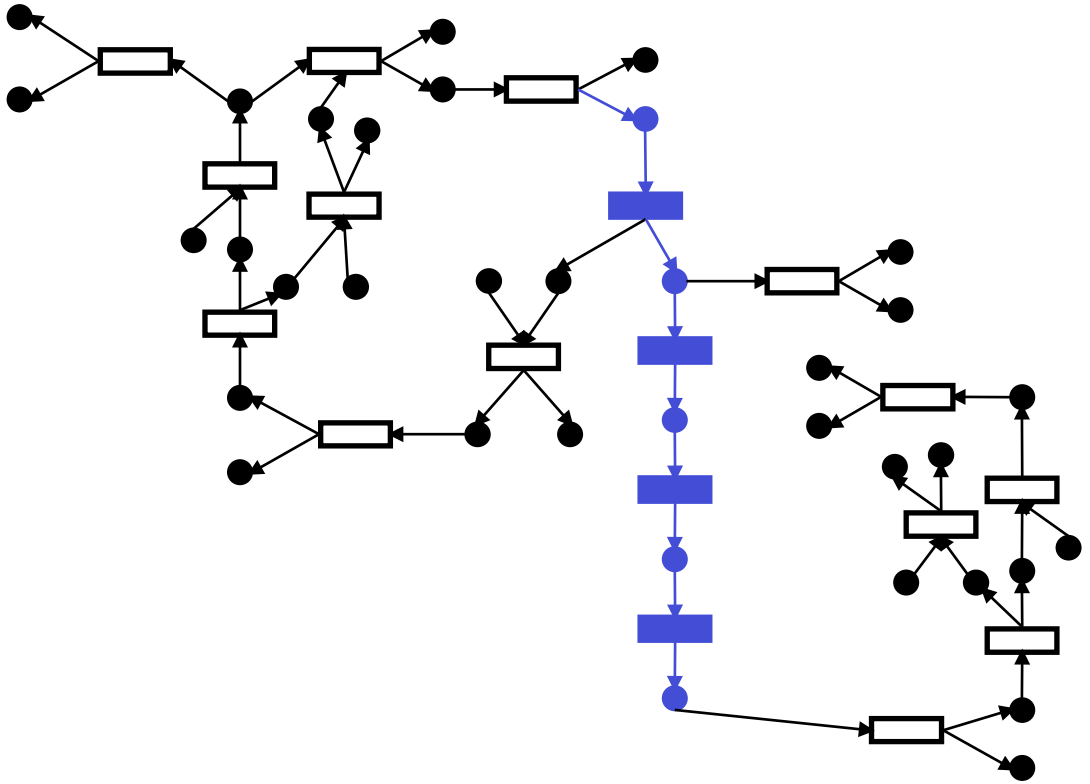
- Simple
- Sufficient for some analysis such as topological or statistical properties
- Discovery of basic patterns
- Useful in specific applications

Bipartite graphs

- Two classes of nodes, compounds and reactions
- Edges can not relate nodes from the same set
  - Edges occur between a compound and a reaction
- Edges can be directed or undirected
- Directed edge from compound to reaction denotes a substrate of the reaction and vice versa
- No ambiguity



Reactions and compounds as directed bipartate graph



compounds

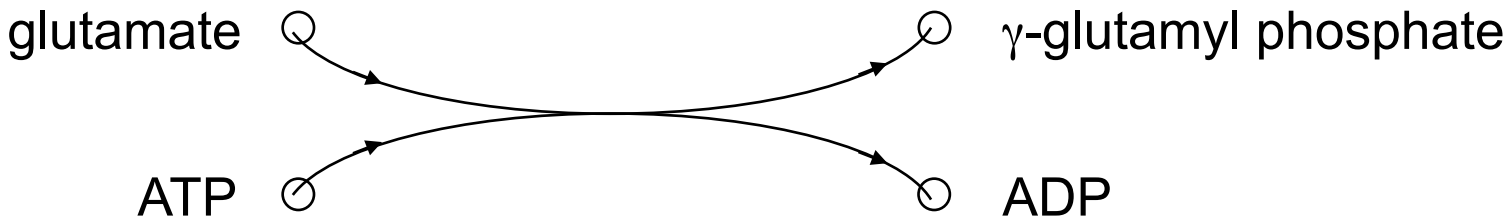
reactions

substrate → reaction

reaction → product

Hypergraphs

- Like bipartite graphs
- Hyperedge relates a set of substrates to a set of products
- Can be converted to bipartite graph or vice versa



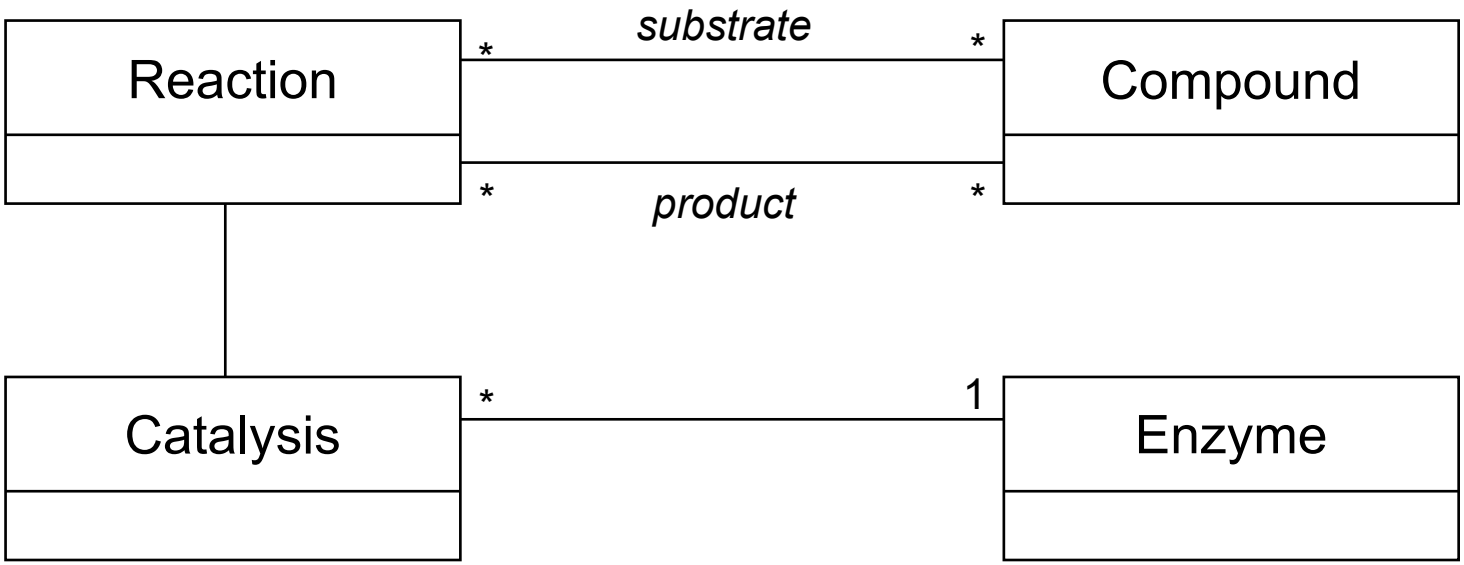
Bipartite graphs and hypergraphs - limitations

- Control mechanisms of reactions can not be explicitly represented
  - e.g. catalysis, inhibition, activation, etc.
- Limited to reactions and compounds

- However, this is sufficient for:
  - Analysis of topological properties
  - Path finding
  - Pathway reconstruction/synthesis
  - Pathway prediction

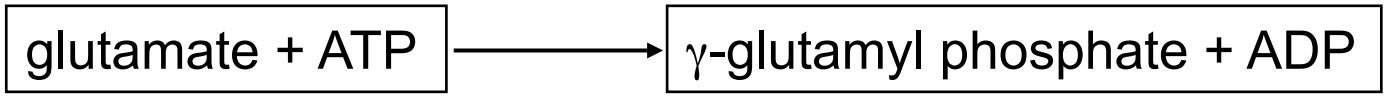
Object models

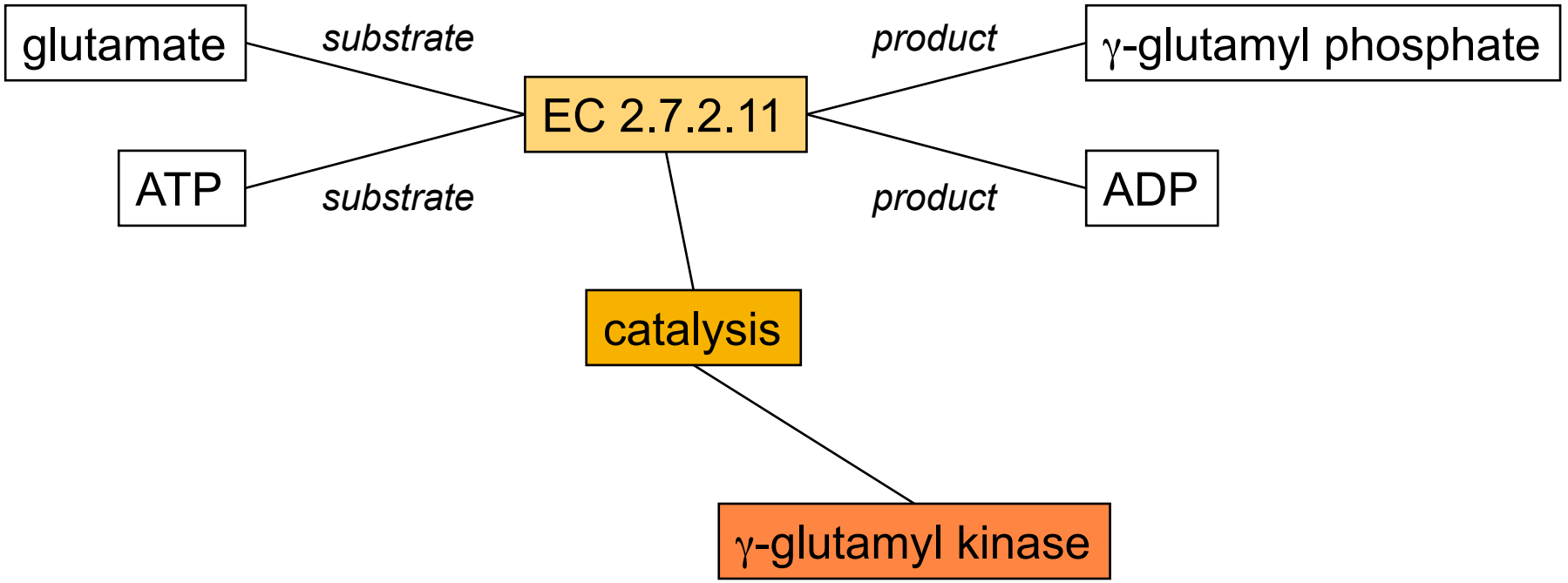
- Required if regulatory information is to be included
- Generalisation of bipartite graphs
- Nodes are typed, permit more detailed description
- Allow inheritance



Object models - example

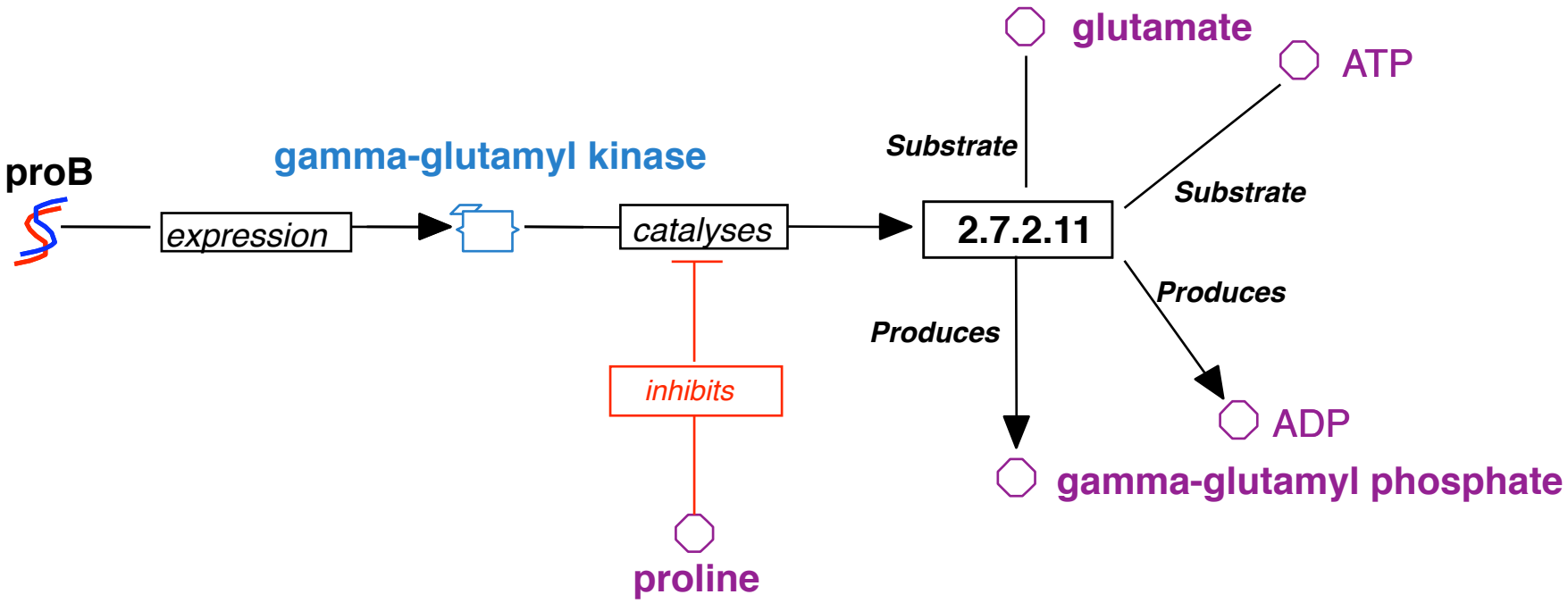
- The reaction catalysed by γ-glutamyl kinase

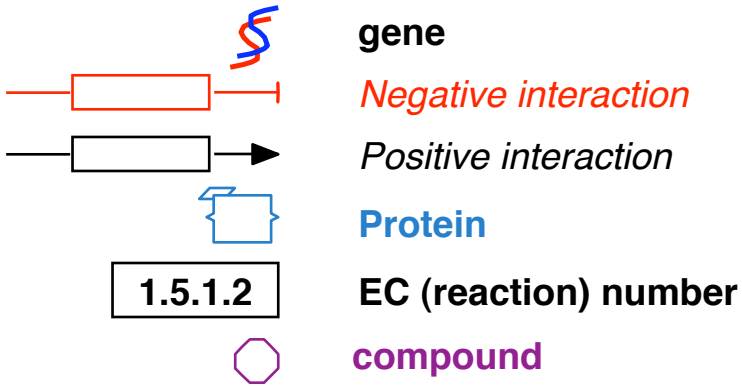




Metabolic Step

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



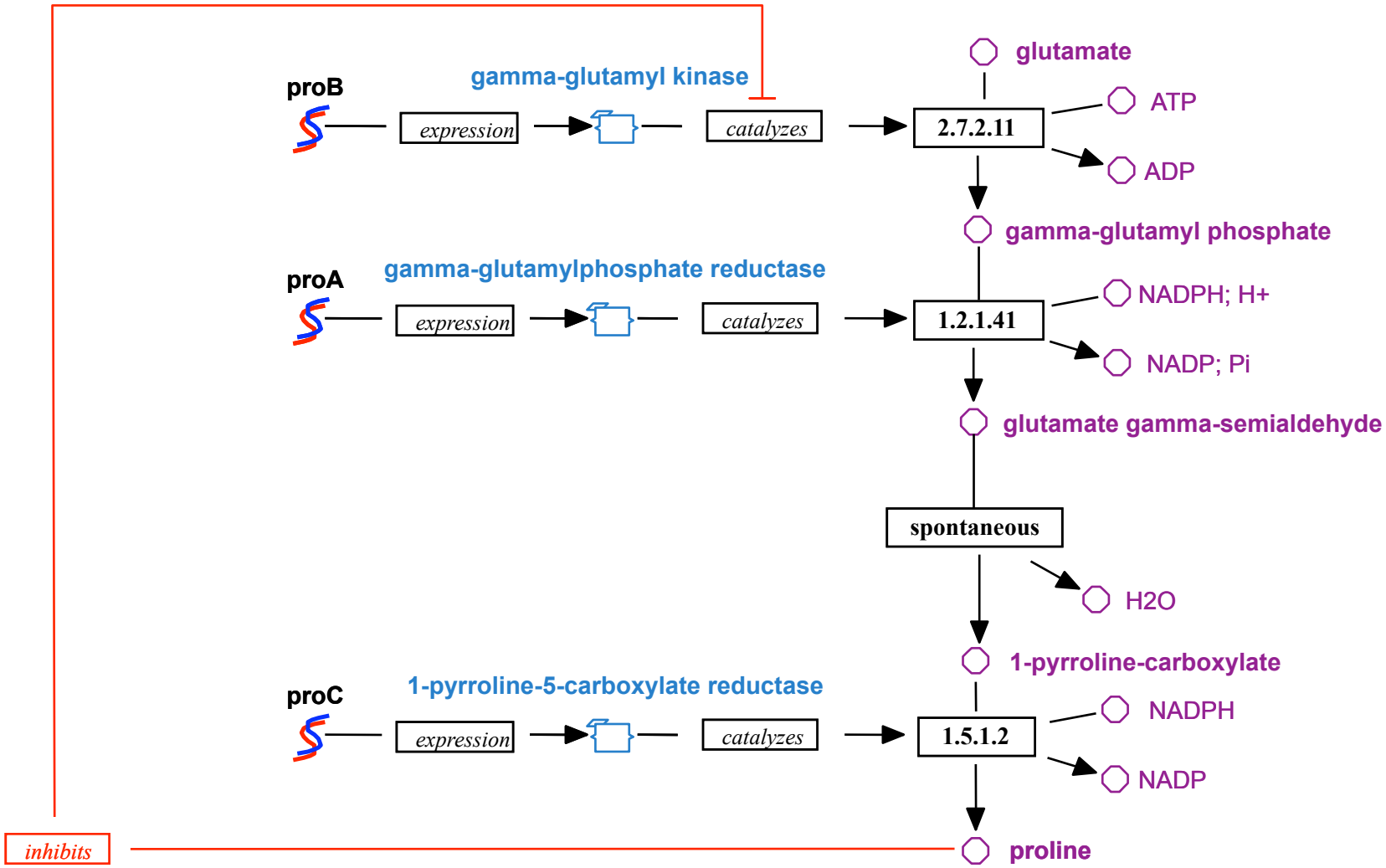


Biochemical Entity

-o [Inhibits] ->

Reaction Catalysis

Metabolic Pathway: Proline Biosynthesis



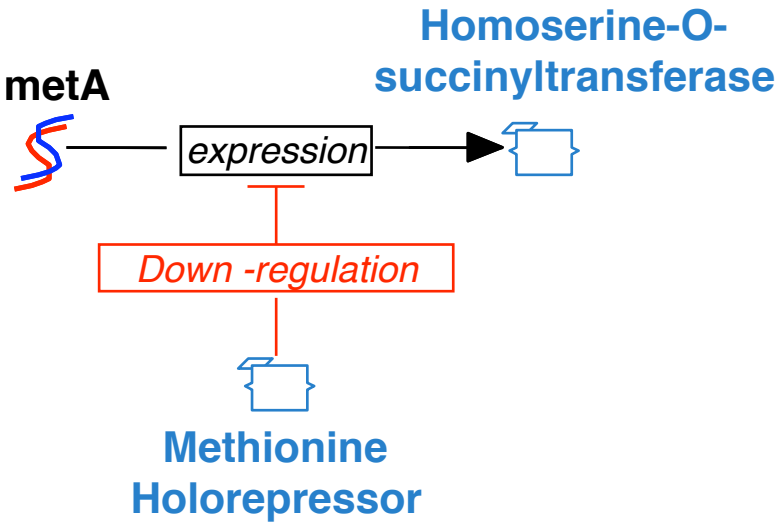
Transcriptional Regulation

Transcriptional repression (down-regulation)

Protein

-o [down-regulates] ->

expression

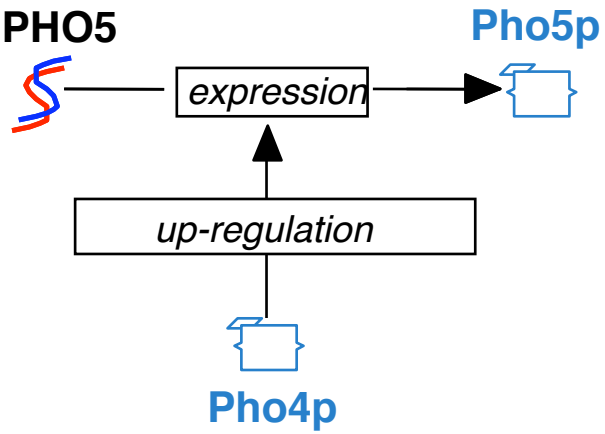


Transcriptional activation (up-regulation)

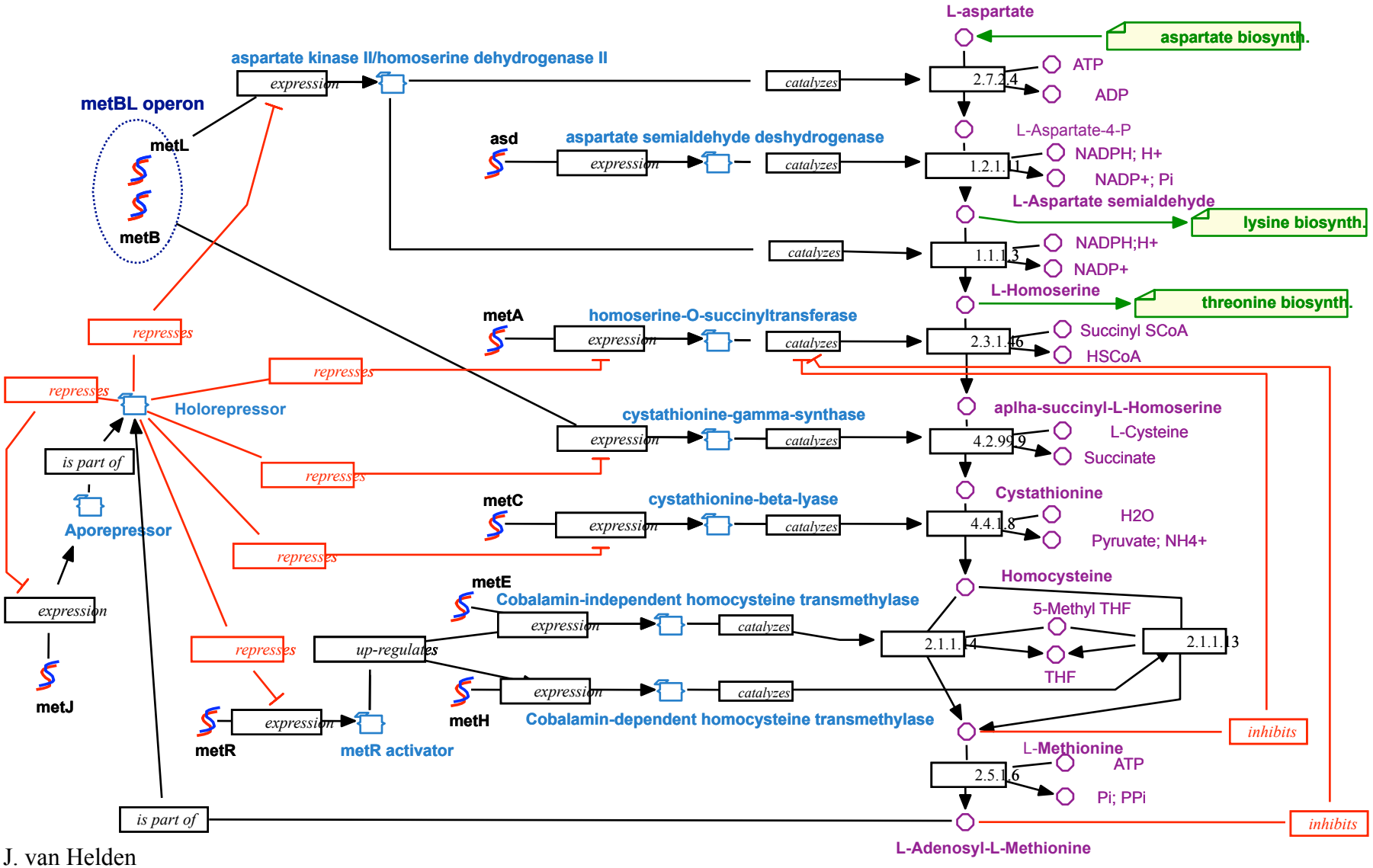
Protein

-o [up-regulates] ->

expression

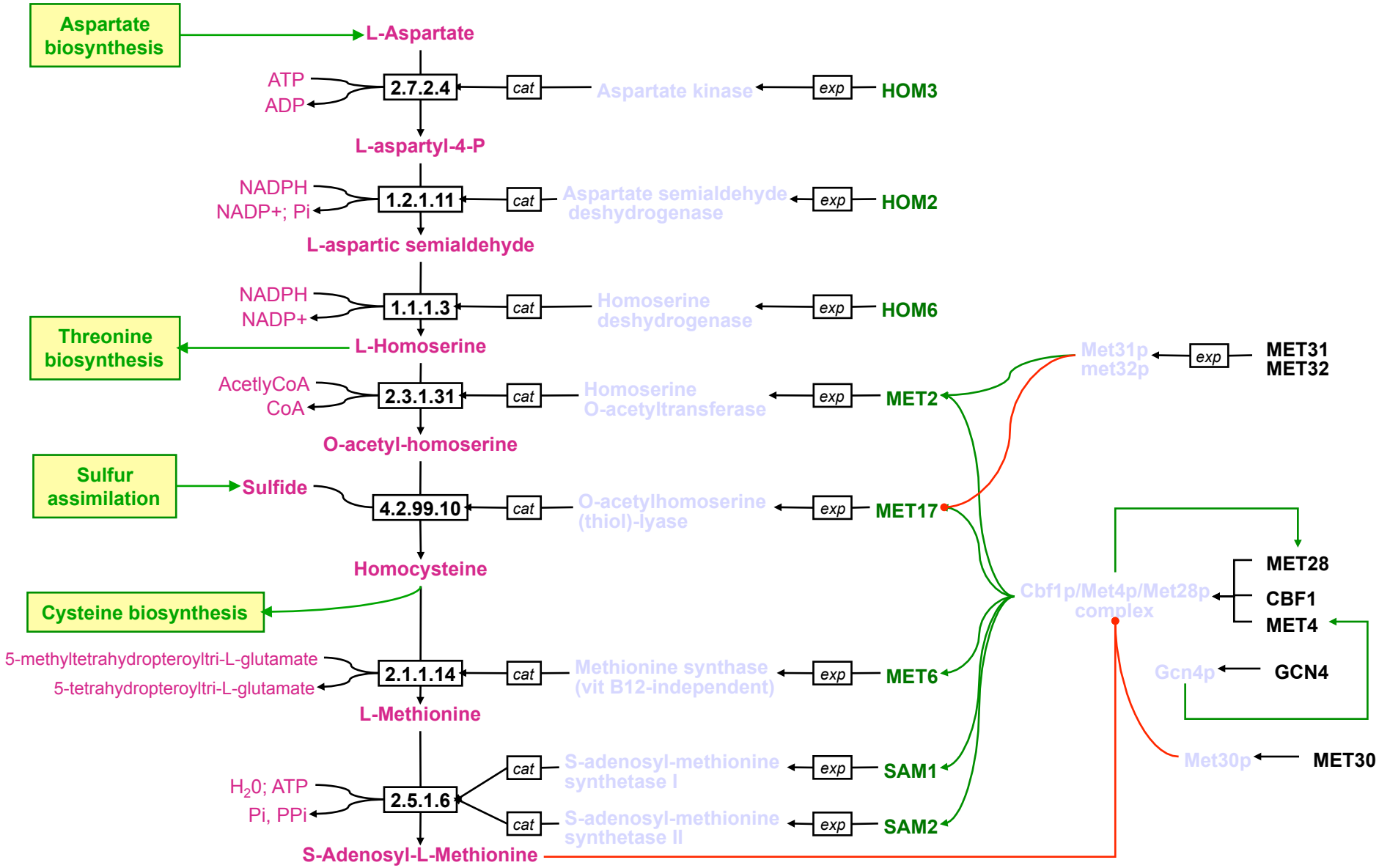


Methionine Biosynthesis in E.coli



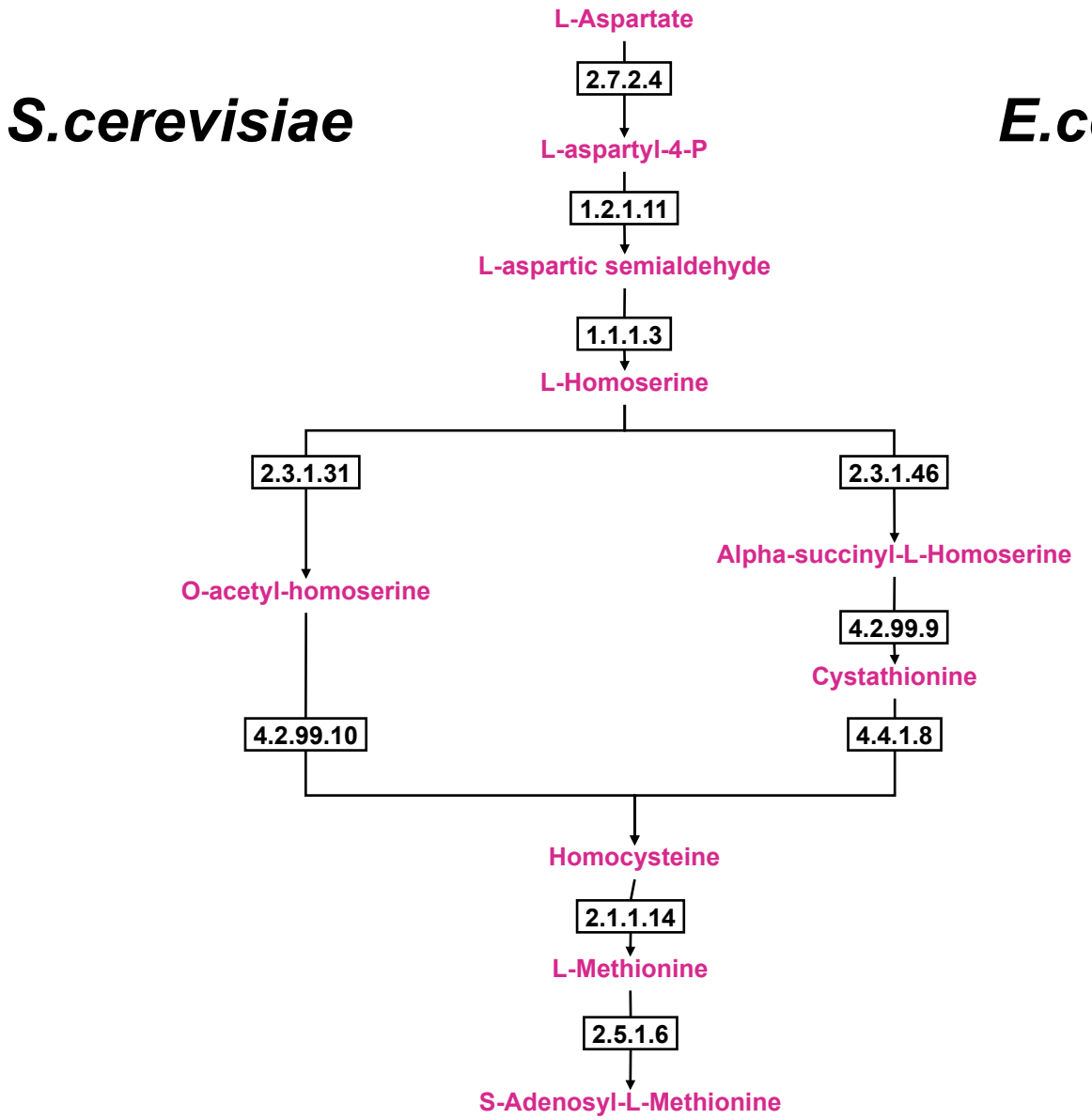
J. van Helden

Methionine Biosynthesis in S.cerevisiae

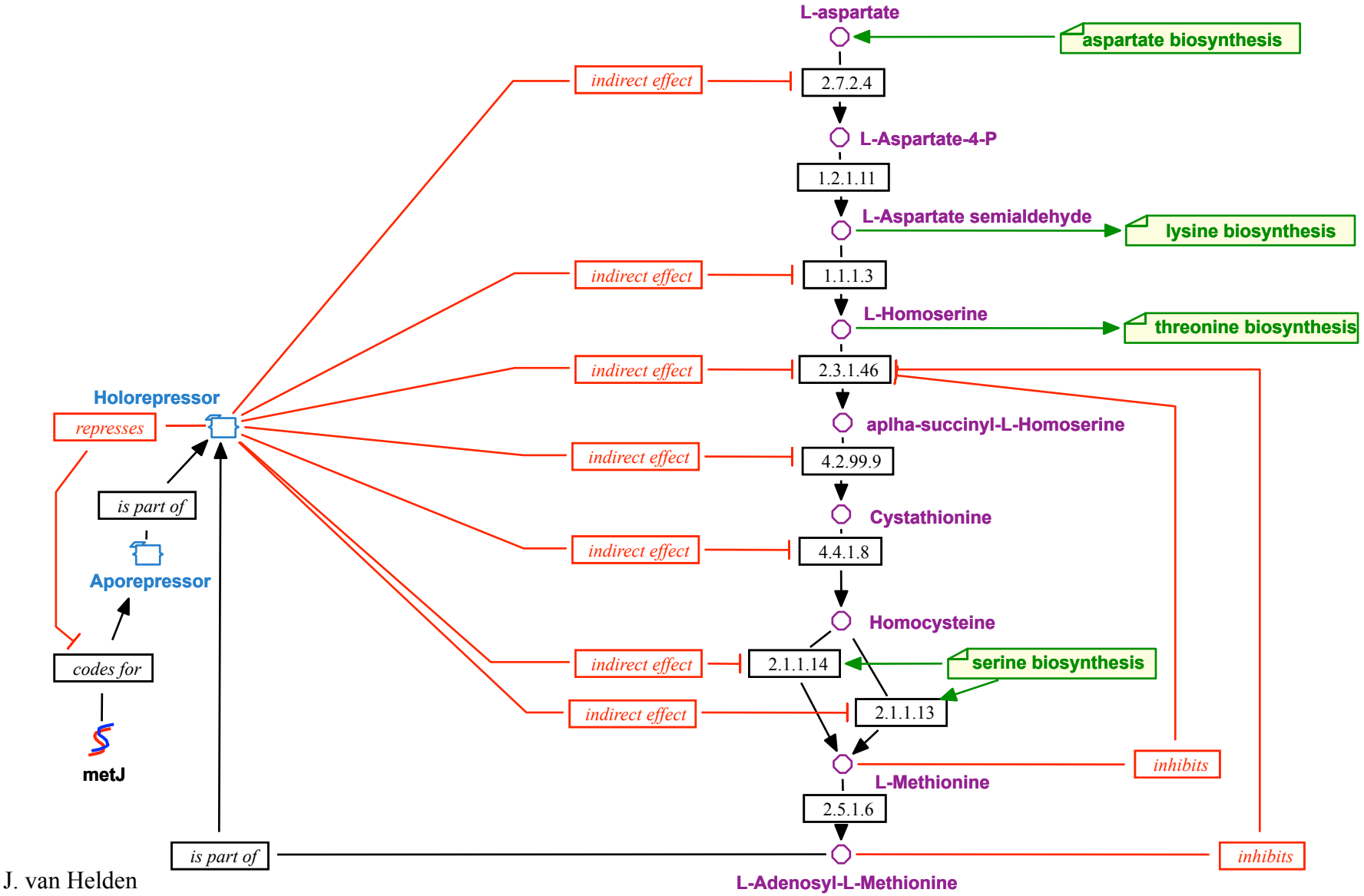


J. van Helden

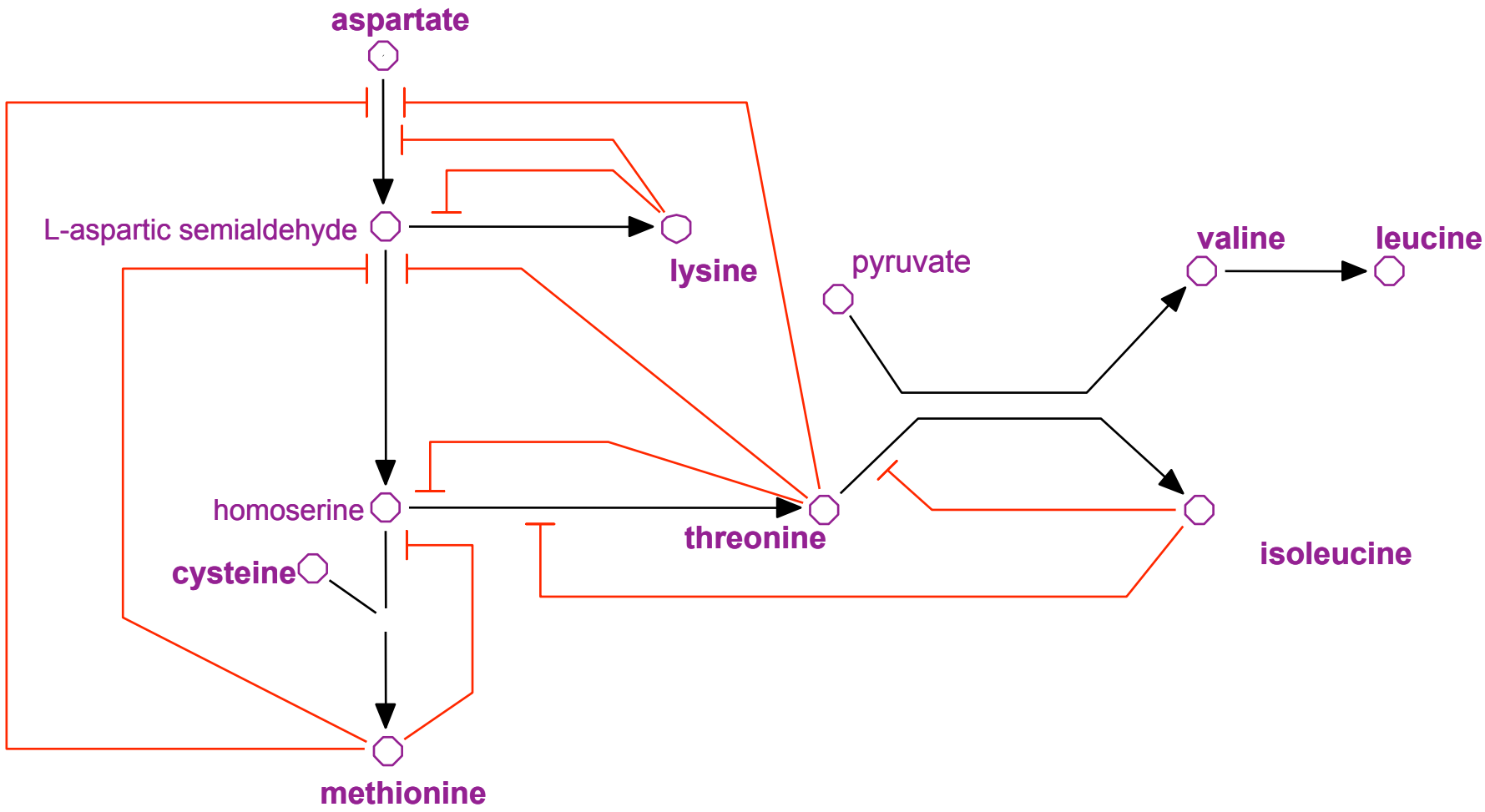
Alternative methionine pathways



Shortcut Representation



High-level Abstraction



Other Important Issues

- ***Partial information*** (indirect interactions), and subsequent filling of the missing steps.
- ***Negative results*** (elements that have been shown not to interact, enzymes missing in an organism).
- ***Putative interactions*** resulting from computational analyses

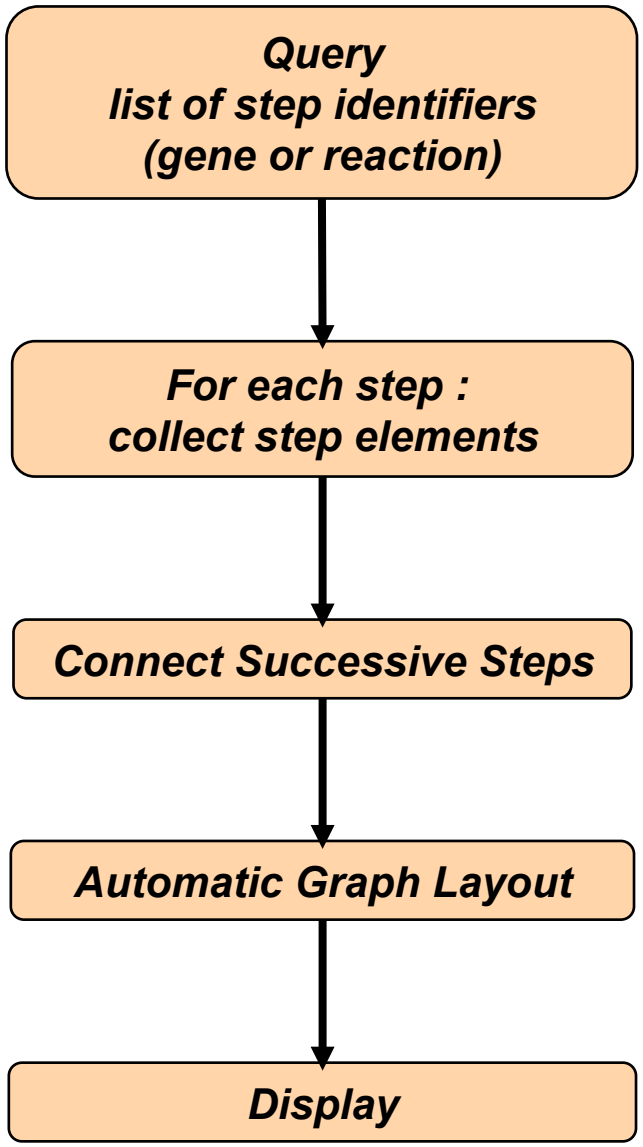
Requirements: Network navigation

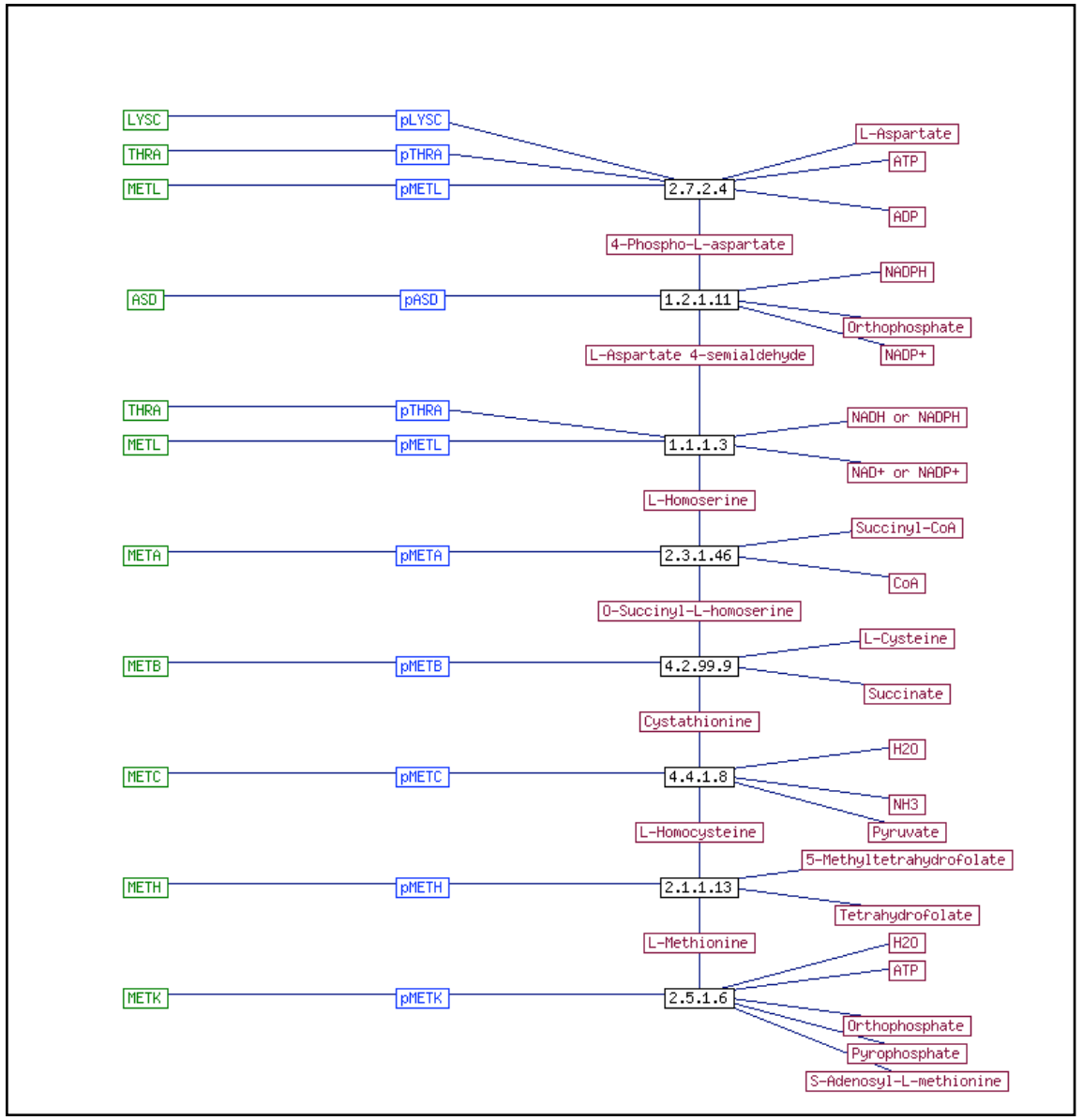
- ***How many pathways*** & ***how many steps*** within each pathway, from compound A to compound B
- Give all the ***pathways that contain or lack specified compounds*** or processes
- ***Highlight pathways/networks***: level of certainty of the information, eliminating trivial pathways (e.g. production consumption of water); rank according to fitness of match
- Which ***paths / pathways may be affected*** when gene/proteins turned off / missing.
- ***Compare biochemical pathways:*** from different organisms and tissues; highlight common features and differences; predict missing elements ('reconstruction')
- Represent pathways at ***different resolution levels***
- Compile repertoires of recurrent ***network motifs*** at different resolution levels
- Identify all ***positive/negative regulatory cycles*** in a pathway graph.

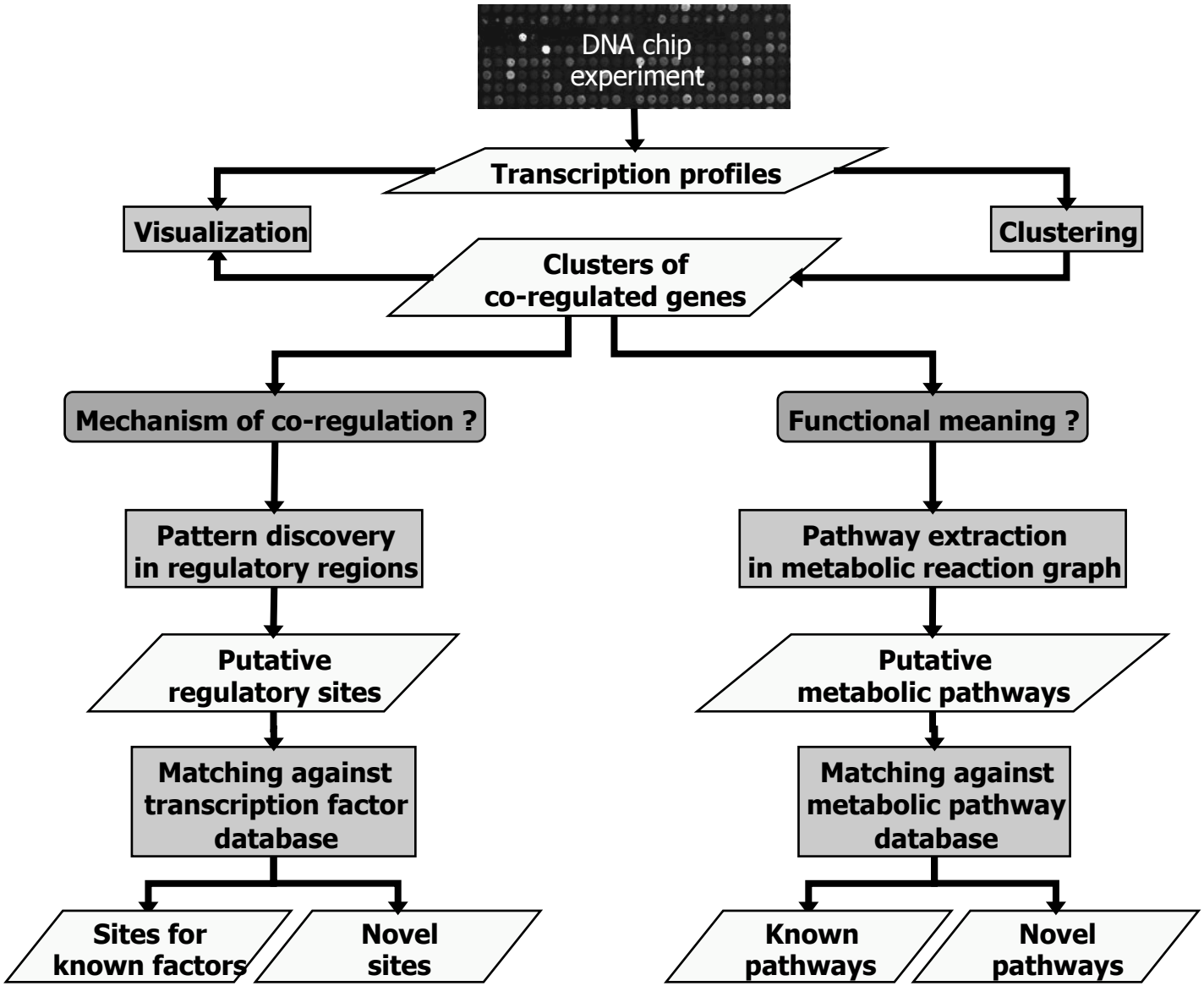
Jacques van Helden, Lorenz Wernisch, David Gilbert, and Shoshana Wodak. "Graph-based analysis of metabolic networks". in Ernst Schering Research Foundation Workshop Volume 38: Bioinformatics and Genome Analysis. Springer-Verlag, 2002

Metabolic Graph Layout

Metabolic pathway: Query on EC numbers:
E.coli, methionine biosynthesis



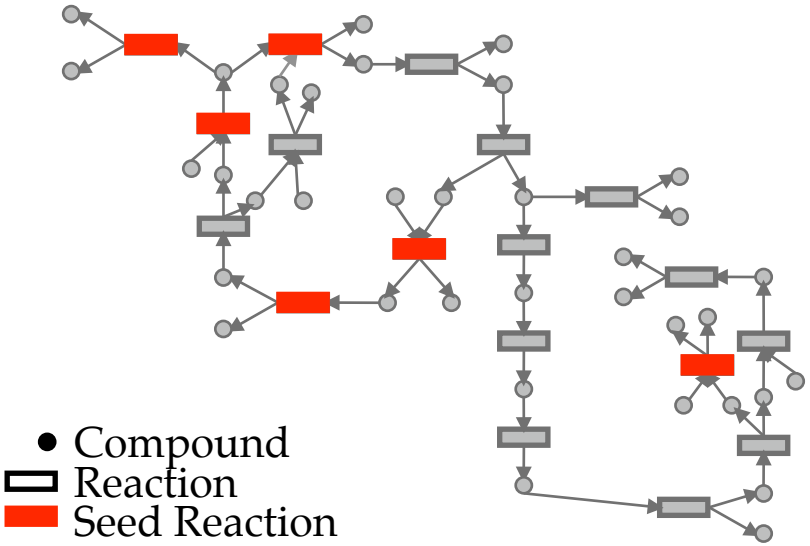




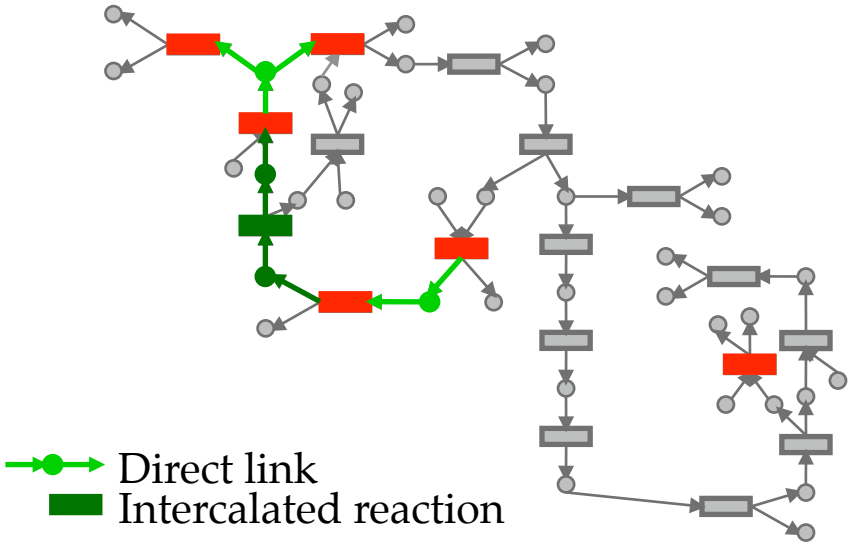
J van Helden, D Gilbert, L Wernisch, M Schroeder, and S Wodak, Application of Regulatory Sequence Analysis and Metabolic Network Analysis to the Interpretation of Gene Expression Data, in Computational Biology (Olivier Gascuel and Marie-France Sagot, Eds), LNCS 2006, 147-163, 2001

Queries - subgraph extraction

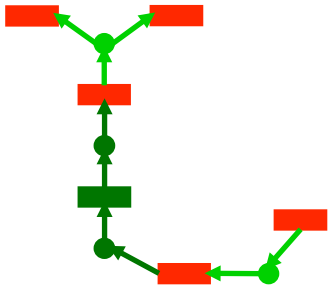
A. Seed reactions



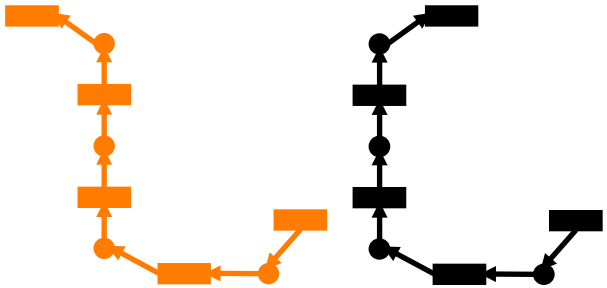
B. Reaction linking



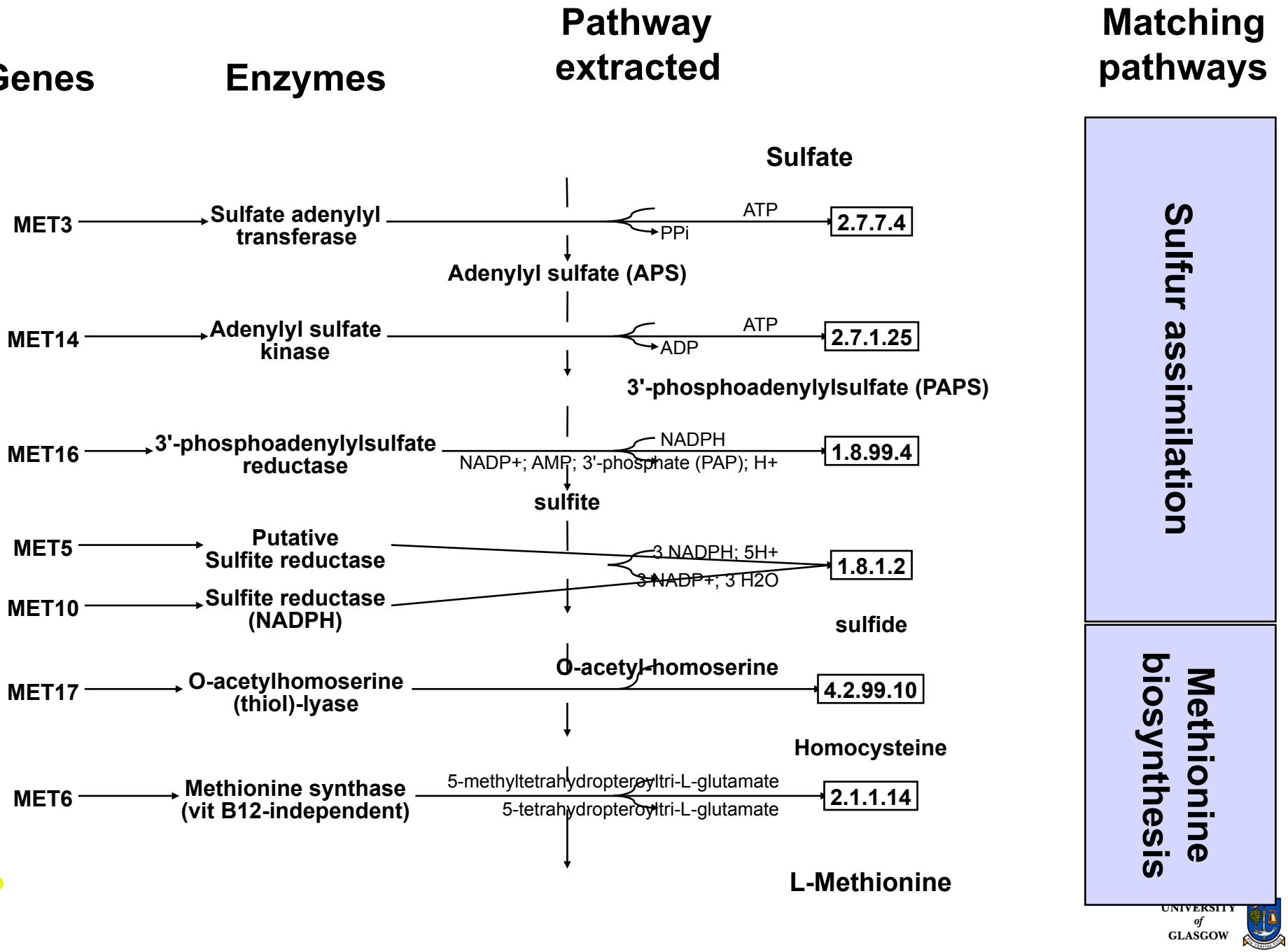
C. Subgraph extraction



D. Linear Path Enumeration



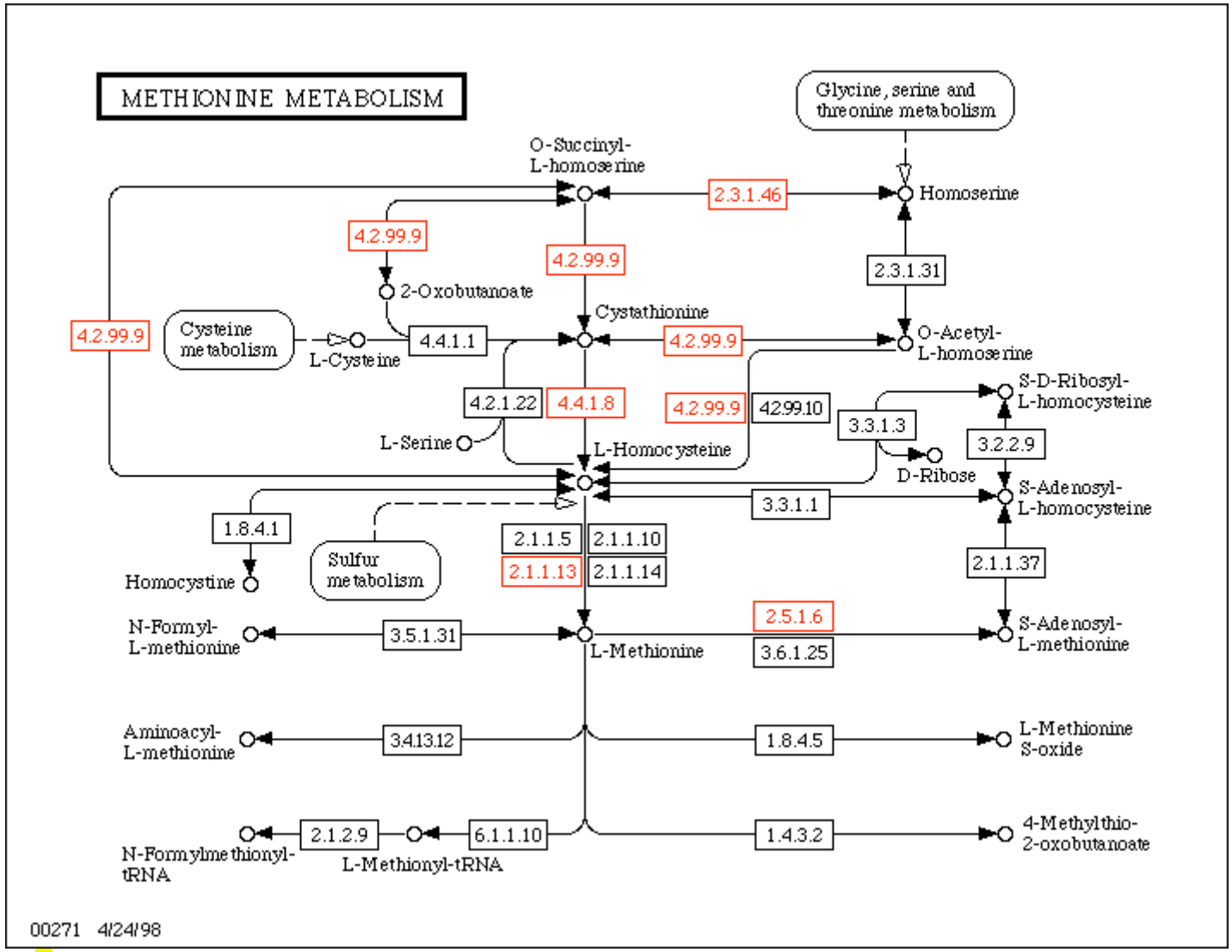
Maximal Pathway extracted from a cluster of cell-cycle regulated genes



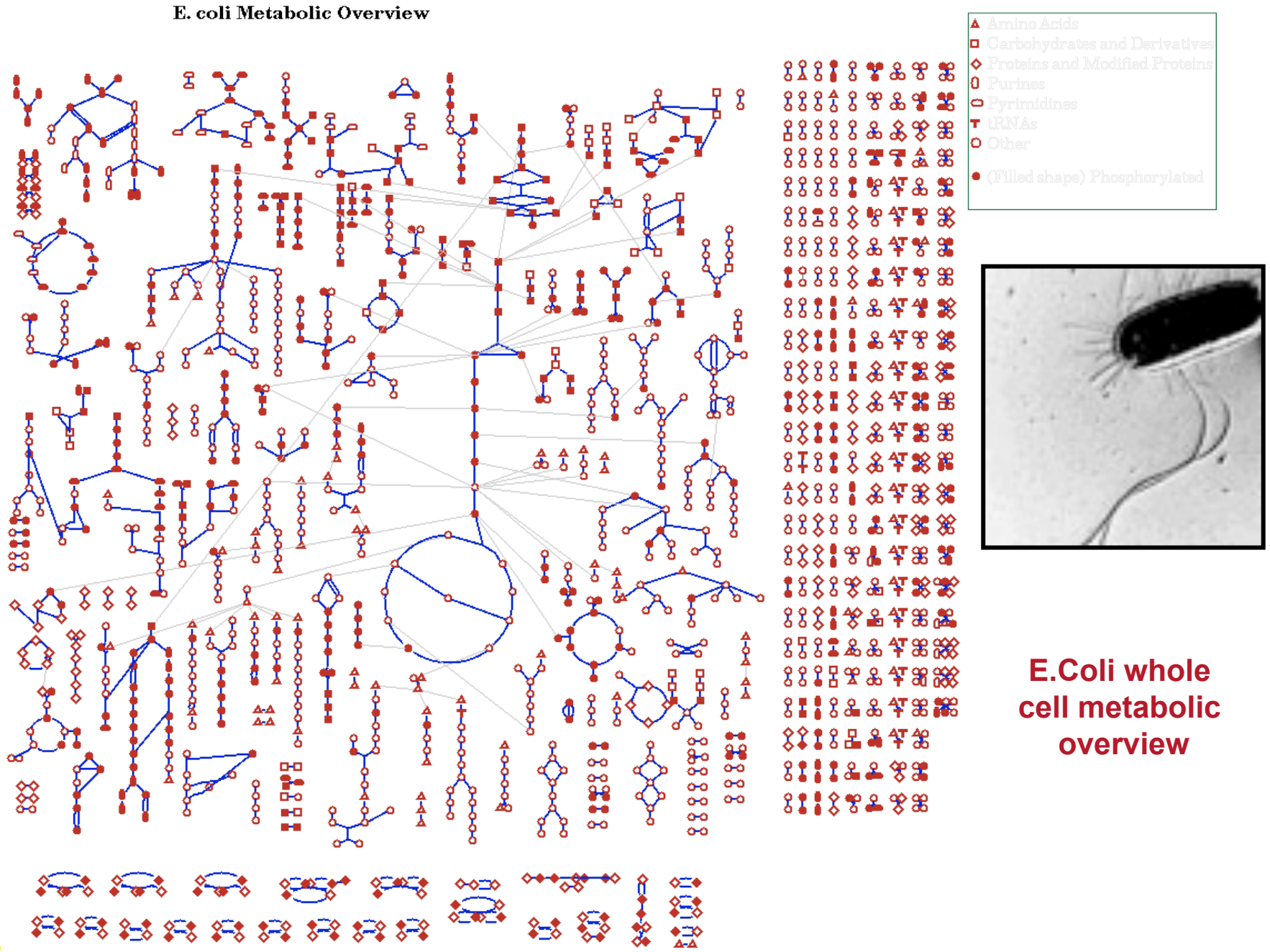
Databases & systems available

- Enzyme function and metabolic pathways :
  - **KEGG**
  - **BioCyc: EcoCyc** (E.Coli), **MetaCyc** (900 organisms); **+368 predicted** (PathoLogic program)
  - **AMAZE (**metabolic, regulatory and signal transduction pathways)
  - **BRENDA** - enzyme function only.
- Querying facilities - various levels of complexity. Simple browsing & basic queries (string search on the values of selected fields), to pathway analysis.
- Some path-finding tools, which find all paths between two specified elements, or from a specified element to any other.
- Results display: colouring paths found on pre-drawn static maps (KEGG), or on a dynamically generated diagram

KEGG Query & result

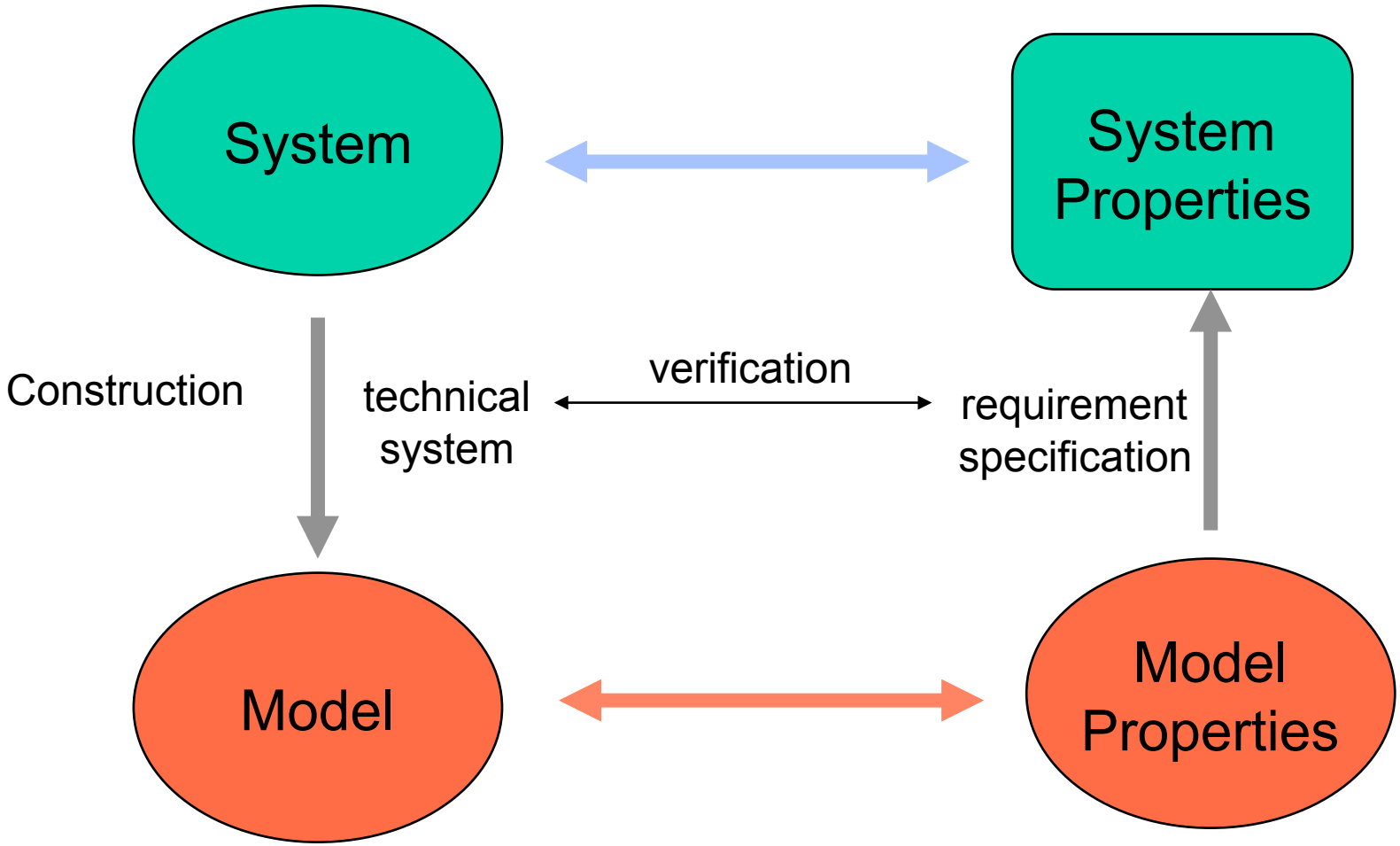


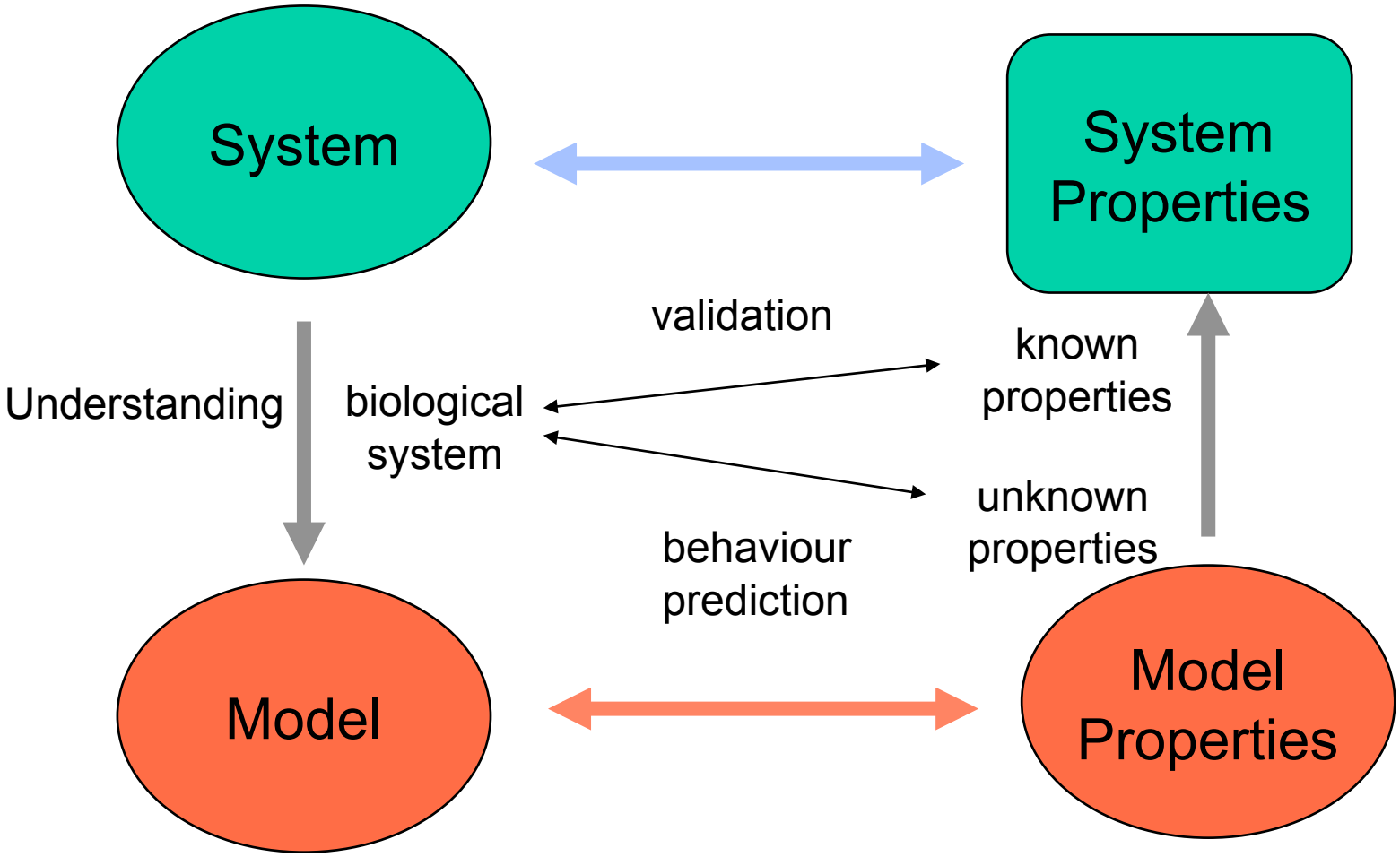
Query =
2.7.2.4
1.2.1.11
1.1.1.3
2.3.1.46
4.2.99.9
4.4.1.8
2.1.1.13
2.5.1.6
map00271
Methionine
metabolism



E.Coli whole cell metabolic overview

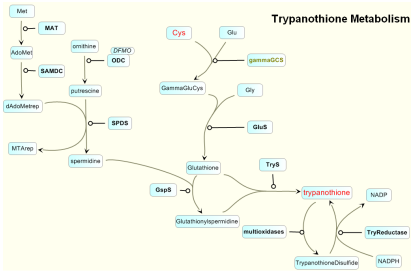
UNIVERSITY of GLASGOW


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

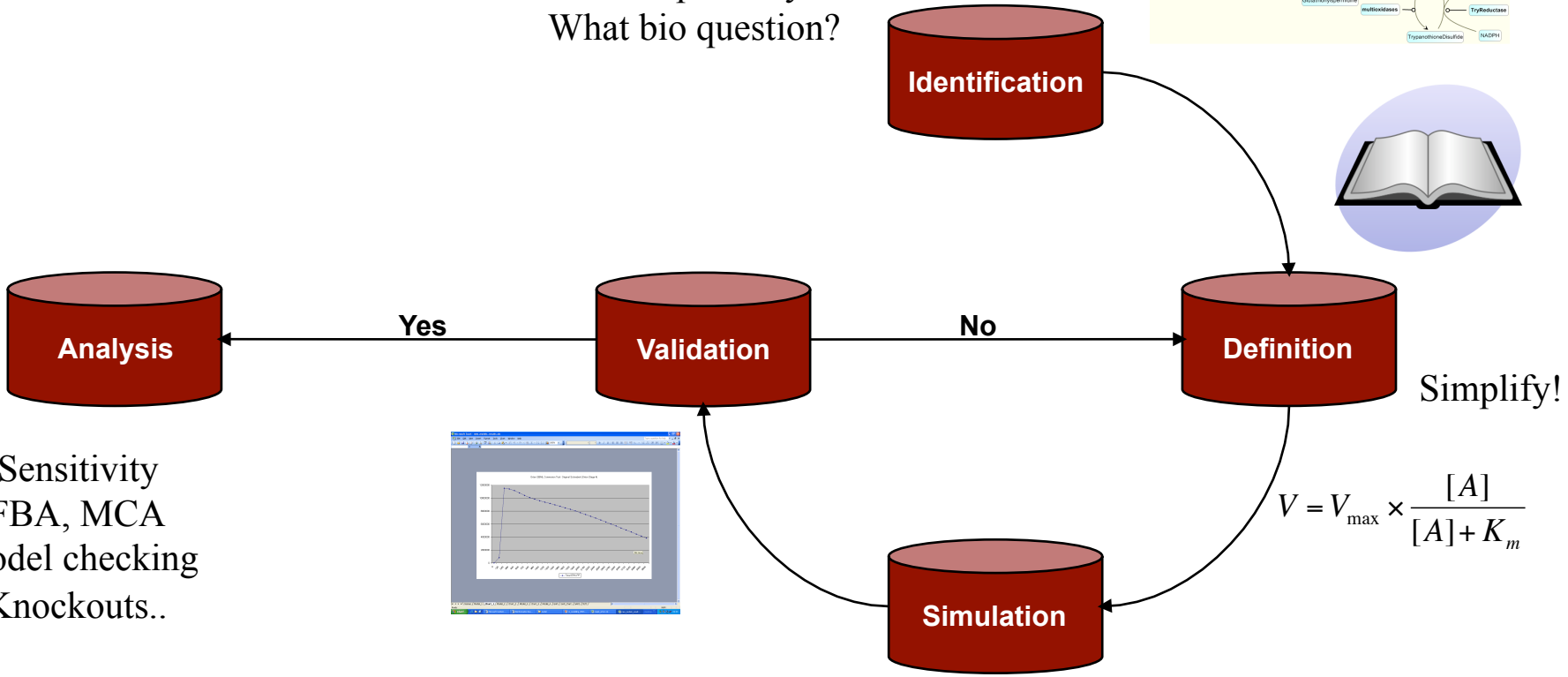

System
System Properties
validation
Understanding
biological system
known properties
unknown properties
behaviour prediction
Model
Model Properties

Slide from
Richard Orton

How to model

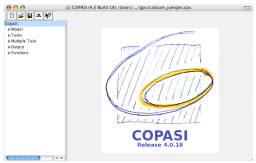


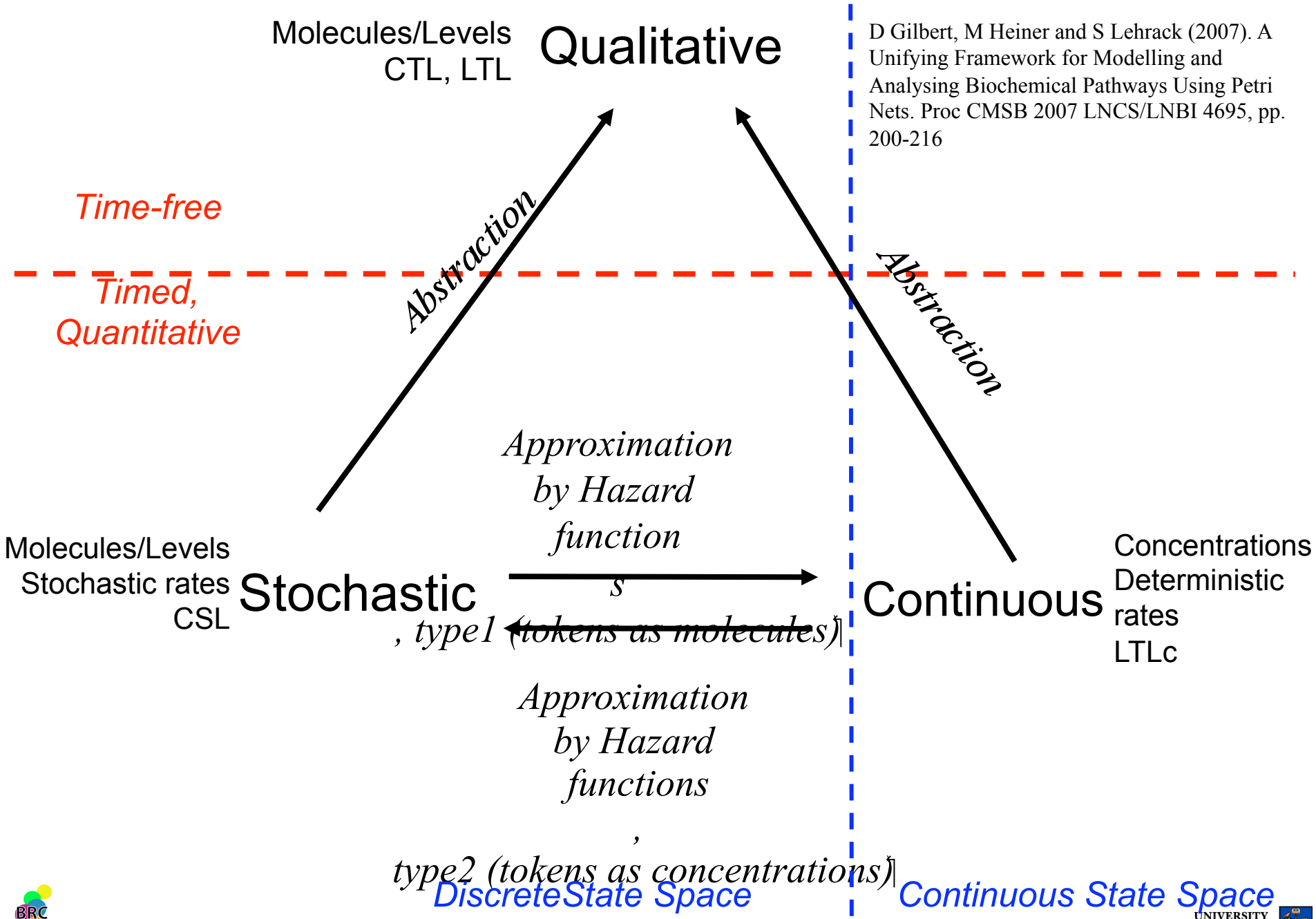
Which pathway?
What bio question?



Sensitivity
FBA, MCA
Model checking
Knockouts..

$$V = V_{\max} \times \frac{[A]}{[A] + K_m}$$




Molecules/Levels
CTL, LTL
Qualitative
D Gilbert, M Heiner and S Lehrack (2007). A Unifying Framework for Modelling and Analysing Biochemical Pathways Using Petri Nets. Proc CMSB 2007 LNCS/LNBI 4695, pp. 200-216
Time-free
Timed, Quantitative
Abstraction
Abstraction
Approximation by Hazard function
s
, type1 (tokens as molecules)
Approximation by Hazard functions
,
type2 (tokens as concentrations)
Molecules/Levels
Stochastic rates
CSL
Stochastic
Continuous
Concentrations
Deterministic rates
LTLc
DiscreteState Space
Continuous State Space

Hazard functions

- Hazard function type1 (tokens as molecules)

$$h_t := c_t \cdot \prod_{p \in \bullet t} \binom{m(p)}{f(p,t)}$$

- c_t transition specific stochastic rate constant
- m(p) current number of tokens on pre-place p of transition t
- binomial coefficient number of non-ordered combinations of the f(p,t) molecules, required for the reaction, out of the m(p) available ones.

- Hazard function type2 (tokens as concentrations)

$$h_t := k_t \cdot N \cdot \prod_{p \in \bullet t} \left(\frac{m(p)}{N} \right)$$

- k_t transition deterministic rate constant
- N number of levels
- Levels: Calder et al, Trans Comp Sys Bio VI, LNBI 4220, 2006

UNIVERSITY of GLASGOW

Dynamic behaviour - modelling

http://bib.oxfordjournals.org - Brief Bioinform -- Gilbert et al. 7 (4): 339 Table 2

Table 2: Comparison of methods for description, simulation and analysis of biochemical systems

Method	Depiction/model	Simulation	Analysis
Pathway chart	Biochemical reactions/no formal model	None	None
Ordinary differential equations (ODEs)	Mathematical equations	Deterministic numerical solution: time-discretisation	Symbolic and numerical analysis (e.g. bifurcation analysis)
Partial differential equations (PDEs)	Mathematical equations	Deterministic numerical solution: space-time-discretisation	Symbolic and numerical analysis
Stochastic differential equations	Mathematical equations with random terms	Stochastic numerical simulation: time-discretisation	Symbolic and numeric analysis
Discrete Petri nets	Graph, labelled transition system	Animation via tokens	Qualitative: structural analysis and temporal logic
Continuous Petri nets	Graph, labelled transition system, rate information	Via ODEs	See ODEs
SBML-based graphical formalisms	Graph, rate information	Various (e.g. ODEs, Gillespie)	Various, tool-dependent
Stochastic π-calculus	Algebraic formulae	Stochastic numerical simulation via Gillespie algorithm	None
Process algebra (PEPA)	Algebraic terms, stochastic temporal logic	Stochastic numerical simulation via Gillespie algorithm; simulation from logical analysis	Quantitative, via temporal logic over models
Cellular automata	Spatiotemporal explicit model based on state and simple rules	Step-wise application of rules to discrete space state	Analysis of emergent properties
Agents	Spatiotemporal explicit model based on autonomous intelligent object behaviour	Representation of object(s) behaviour determined by history of encounters with environment	Analysis of emergent properties

Done

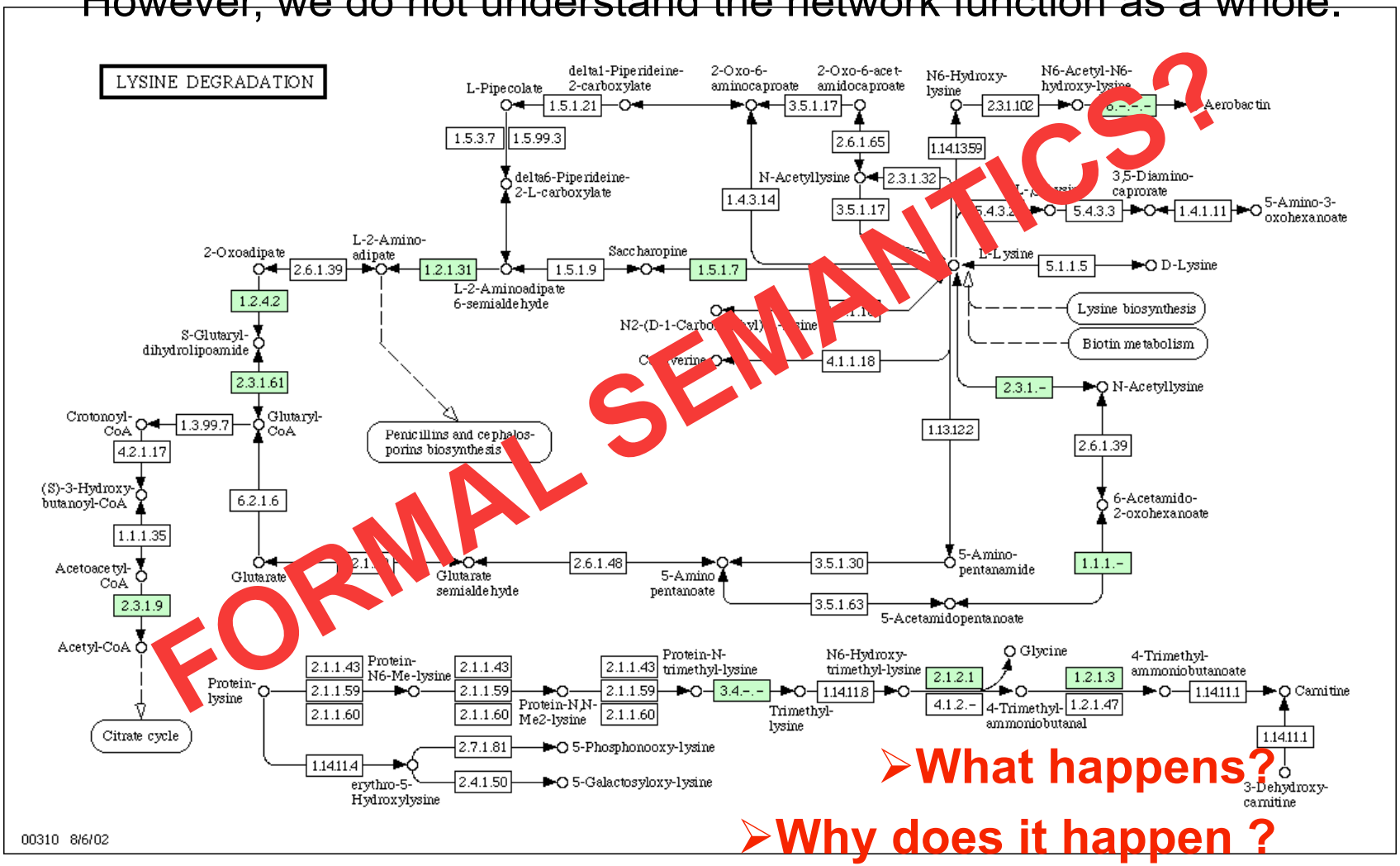
Gilbert, D. et al. Brief Bioinform 2006 7:339-353; doi:10.1093/bib/bbl043

Briefings in Bioinformatics

Biochemical networks

We can describe the general topology and single biochemical steps.

~~However, we do not understand the network function as a whole.~~



FORMAL SEMANTICS?

➢What happens?

➢Why does it happen ?

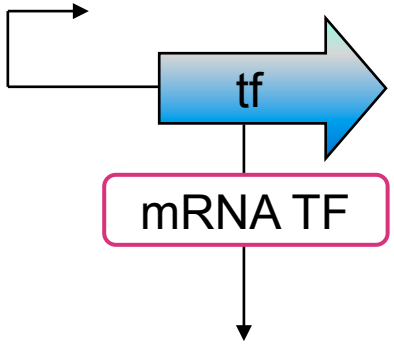
$$\frac{d\alpha}{dt} = -v_1$$
$$\frac{dSte2}{dt} = -v_2 + v_3 - v_5$$
$$\frac{dSte2_{active}}{dt} = v_2 - v_3 - v_4$$
$$\frac{dSst2_{active}}{dt} = v_{46} - v_{47}$$
$$\frac{dG\alpha\beta\gamma}{dt} = -v_6 + v_9$$
$$\frac{dG\alpha GTP}{dt} = v_6 - v_7 - v_8$$
$$\frac{dG\alpha GDP}{dt} = v_7 + v_8 - v_9$$
$$\frac{dG\beta\gamma}{dt} = v_6 - v_9 - v_{10} + v_{11} + v_{21} + v_{23} + v_{25} + v_{27} + v_{32} - v_{42} + v_{43}$$
$$\frac{dSte5}{dt} = -v_{12} + v_{13} + v_{17} + v_{21} + v_{23} + v_{25} + v_{27} + v_{32}$$
$$\frac{dSte11}{dt} = -v_{12} + v_{13} + v_{17} + v_{21} + v_{23} + v_{25} + v_{27} + v_{32}$$
$$\frac{dSte7}{dt} = -v_{14} + v_{15} + v_{17} + v_{21} + v_{23} + v_{25} + v_{27} + v_{32}$$
$$\frac{dFus3}{dt} = -v_{14} + v_{15} + v_{17} + v_{21} + v_{23} + v_{25} + v_{27} - v_{29} + v_{30} + v_{33}$$
$$\frac{dSte20}{dt} = -v_{18} + v_{19} + v_{21} + v_{23} + v_{25} + v_{27} + v_{32}$$

$$v_1 = \alpha[t] \cdot Bar1_{active}[t] \cdot k_1$$
$$v_2 = Ste2[t] \cdot \alpha[t] \cdot k_2$$
$$v_3 = Ste2_{active}[t] \cdot k_3$$
$$v_4 = Ste2_{active}[t] \cdot k_4$$
$$v_5 = Ste2[t] \cdot k_5$$
$$v_6 = Ste2_{active}[t] \cdot G\alpha\beta\gamma[t] \cdot k_6$$
$$v_7 = G\alpha GTP[t] \cdot k_7$$
$$v_8 = G\alpha GTP[t] \cdot Sst2_{active}[t] \cdot k_8$$
$$v_9 = G\alpha GDP[t] \cdot G\beta\gamma[t] \cdot k_9$$
$$v_{10} = G\beta\gamma[t] \cdot C[t] \cdot k_{10}$$
$$v_{11} = D[t] \cdot k_{11}$$
$$v_{12} = Ste5[t] \cdot Ste11[t] \cdot k_{12}$$
$$v_{13} = A[t] \cdot k_{13}$$
$$v_{14} = Ste7[t] \cdot Fus3[t] \cdot k_{14}$$
$$v_{15} = B[t] \cdot k_{15}$$
$$v_{16} = A[t] \cdot B[t] \cdot k_{16}$$
$$v_{17} = C[t] \cdot k_{17}$$
$$v_{18} = D[t] \cdot Ste20[t] \cdot k_{18}$$

READABILITY?

Glasgow iGEM team

Simplifying a Model



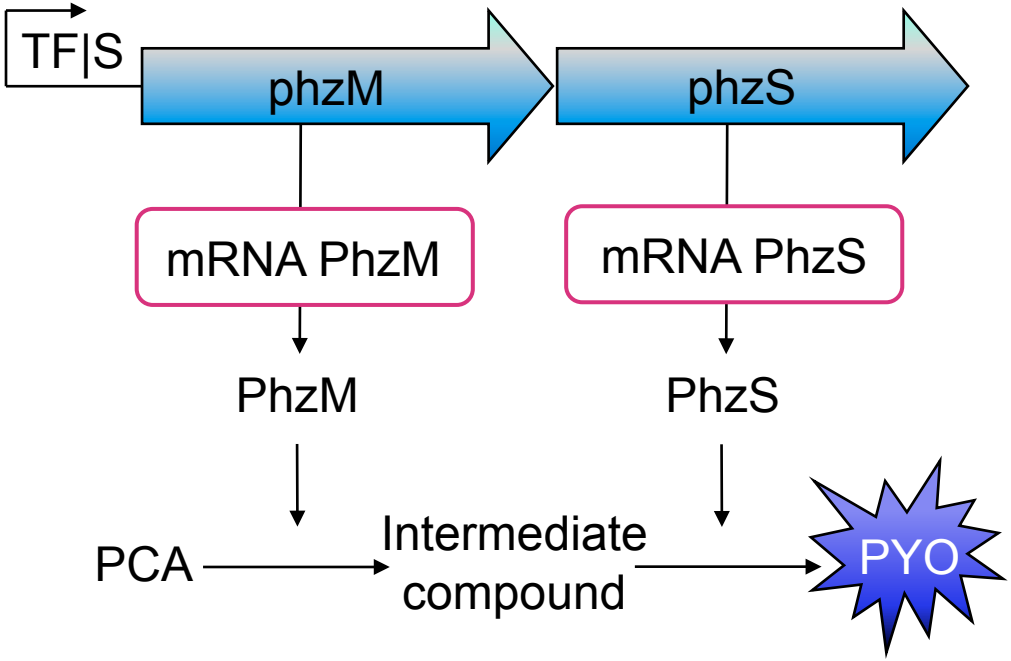
- Merge transcription and translation
- Merge phzM with phzS (Parsons 2007)

TF + S ⇌ TF|S

TF: Dntr or Xylr

S: signal

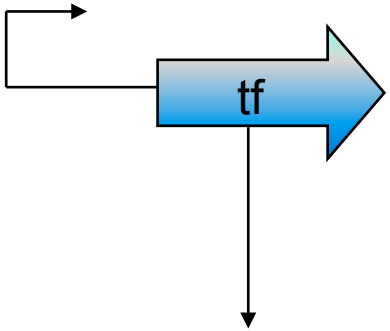
TF|S: complex



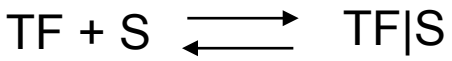
UNIVERSITY of GLASGOW

Glasgow iGEM team

Simplifying a Model



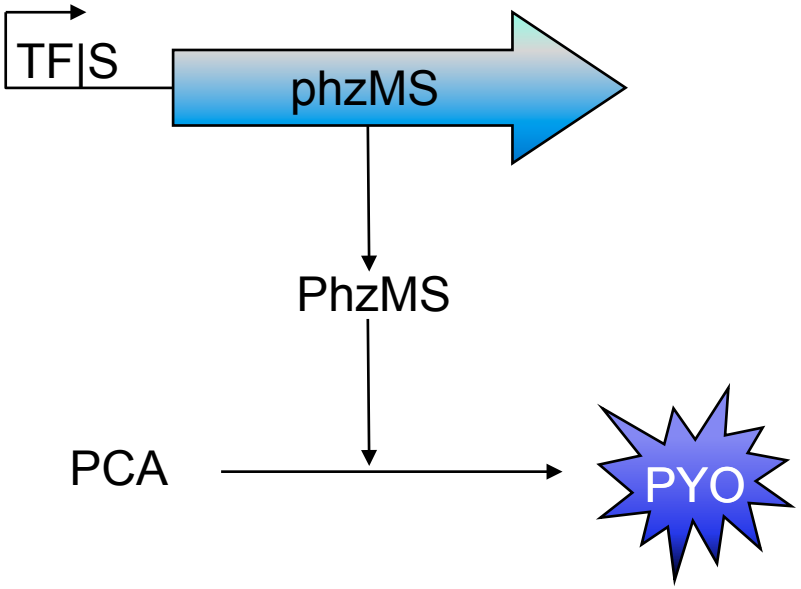
- Merge transcription and translation
- Merge phzM with phzS (Parsons 2007)

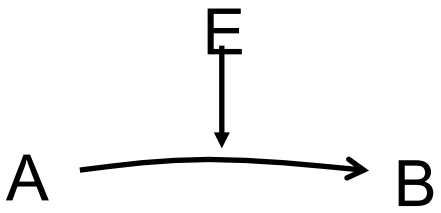


TF: Dntr or Xylr

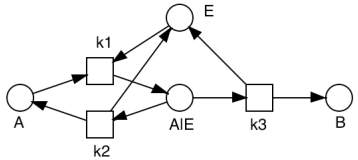
S: signal

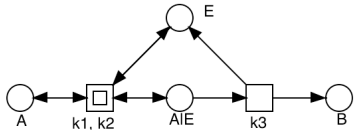
TF|S: complex

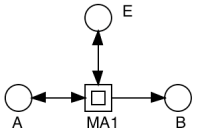




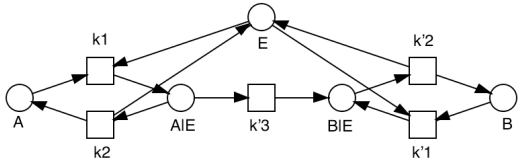
$$E + A \underset{k_2}{\overset{k_1}{\rightleftarrows}} E \mid A \xrightarrow{k_3} E + B$$

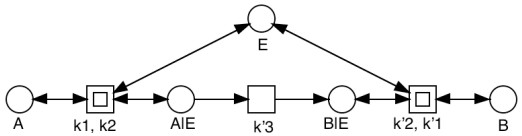


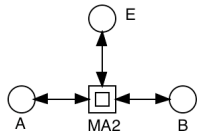




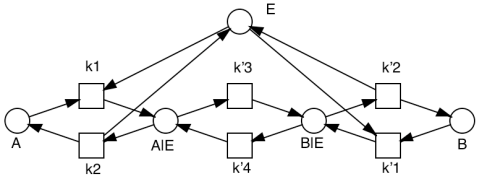
$$E + A \underset{k_2}{\overset{k_1}{\rightleftarrows}} E \mid A \xrightarrow{k_3} E \mid B \underset{k_5}{\overset{k_4}{\rightleftarrows}} E + B$$

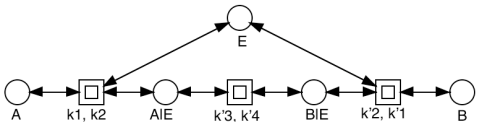


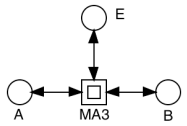




$$E + A \underset{k_2}{\overset{k_1}{\rightleftarrows}} E \mid A \underset{k_6}{\overset{k_3}{\rightleftarrows}} E \mid B \underset{k_5}{\overset{k_4}{\rightleftarrows}} E + B$$

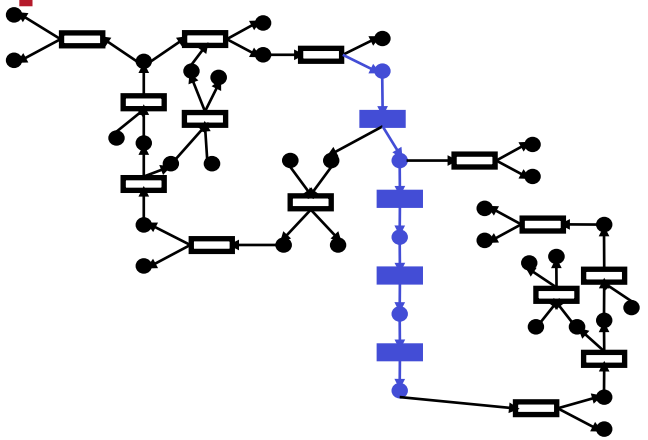




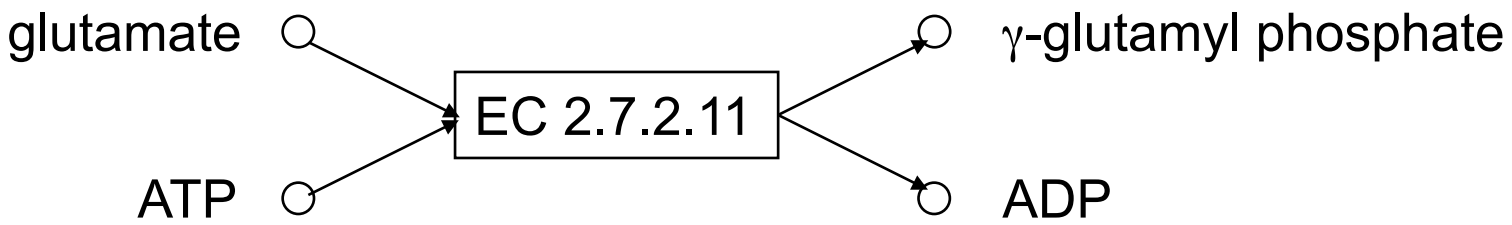


UNIVERSITY of GLASGOW

…..Bipartite graphs!



- Two classes of nodes, compounds and reactions
- Edges can not relate nodes from the same set
  - Edges occur between a compound and a reaction
- Edges can be directed or undirected
- Directed edge from compound to reaction denotes a substrate of the reaction and vice versa
- No ambiguity

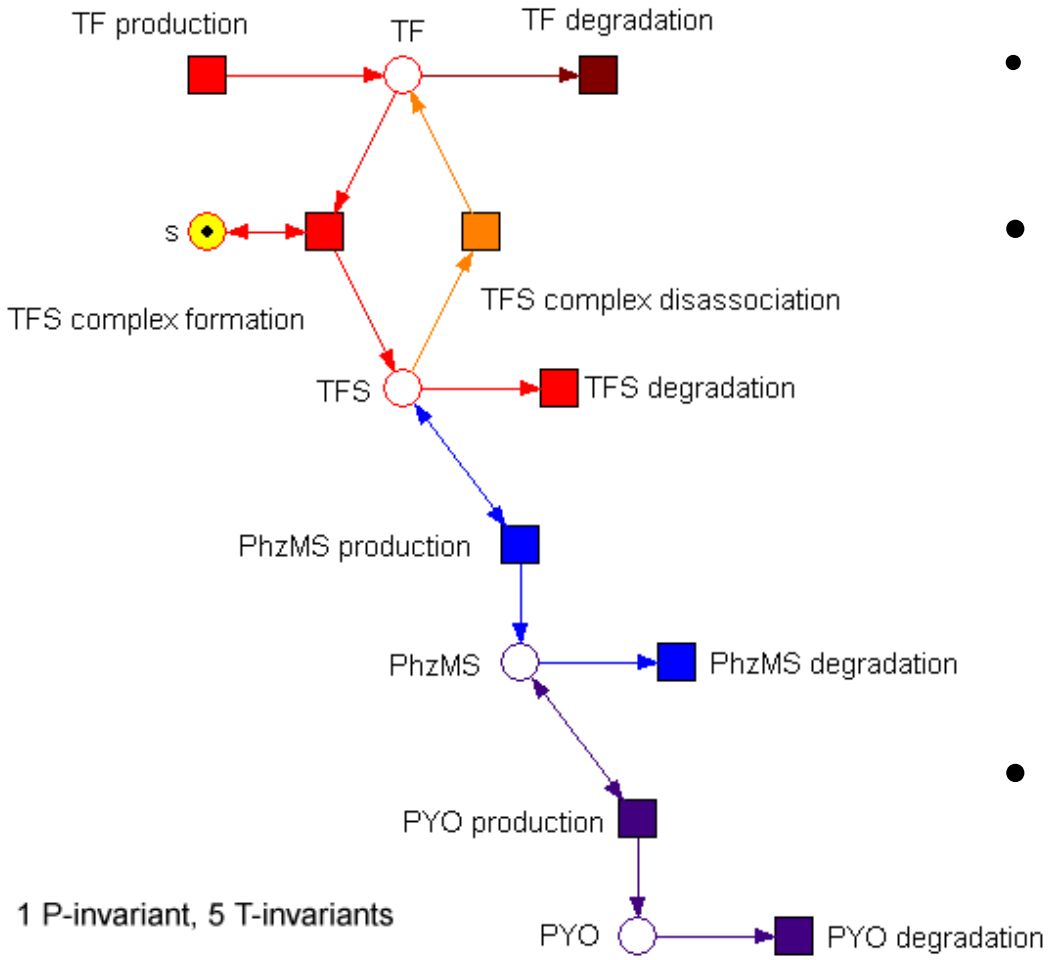


Petri-net analysis

- Place invariants (P-invariants) - sets of places where the sum of tokens remains constant over any firing.
- Transition invariants (T-invariants) - sets of transitions which have a zero effect on the marking of the system.
- If the T-Invariants cover the entire Petri net, it shows that the system can have cyclic behaviour, while incorporating all system parts, which suggests that the system might have been modelled correctly.

Glasgow iGEM team
& Monika Heiner

Qualitative Petri-Net
Modelling & Analysis



- Graphical representation-- Snoopy
- Qualitative analysis Charlie
  - T invariants (cyclic behavior in pink)
  - P invariants
  - (constant amount of output)
- Quantitative Analysis by continuous Petri Net
  - ODE Simulation

BRC

UNIVERSITY of GLASGOW

Petri net analysis

PUR - The Petri net is not pure, i.e. there are pairs of nodes, connected in both directions. This structure corresponds to read arcs. There are two (three) read arcs in the given net.

ORD - The Petri net is ordinary, i.e. all arc weights are equal to 1. This includes homogeneity (see the next bullet) and non-blocking multiplicity (see the next but one bullet).

HOM - The Petri net is homogeneous, i.e. all outgoing arcs of a given place have the same multiplicity.

NBM - The Petri net has the non-blocking multiplicity property, which is of importance in combination with the deadlock trap property (DTP)

CSV - The Petri net is not conservative, i.e. there are transitions which do not preserve the total token amount by their firing, i.e. they increase or decrease the total token amount when firing. Obviously, this applies to the input and output transitions.

SCF - The Petri net is not free of static conflicts, i.e. there are transi- tions sharing a pre-place. This structural property holds e.g. for the two transitions T F degradation and T F S complex production, sharing the pre-place T F , which means that a token on the place T F can either be broken down or follow the way of TFS complex production.

CON - The Petri net is connected, i.e. it holds for all pairs of nodes a and b that there is an undirected path from a to b.

SC - The Petri net is not strongly connected, i.e. it does not hold for all pairs of nodes a and b that there is a directed path from a to b. For example, there is no path from the transition called P Y O degradation back to the transition called T F generation.

FT0, TF0 - The Petri net has input transitions and ouput transitions, i.e. it is an open system. Input transitions are always enabled, therefore they are able to fire arbitrarily often, making the Petri net unbounded.

FP0, PF0 - There are neither input places nor output places.

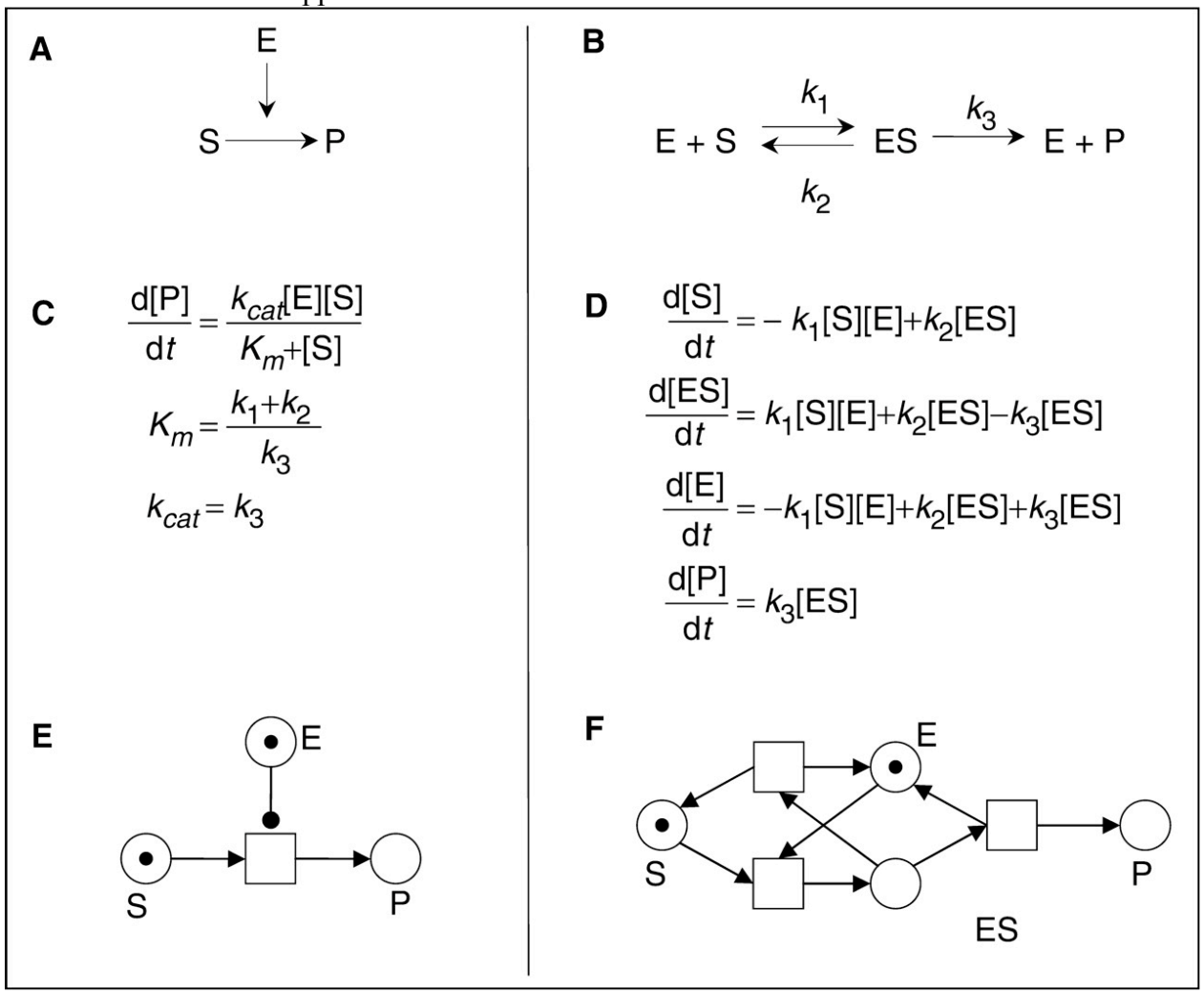
NC - The Petri net belongs to the structural net class Extended Simple

Petri net analysis

- DTP - The Petri net has the deadlock trap property, The DTP involves liveness for ordinary ES nets. Because the net is live, there are no dead transitions and no dead states.
- CPI - The Petri net is not covered by P-invariants. Actually, there is only one minimal P-invariant, which comprises merely the place s. This means that the token number on this place never changes under any firing. Therefore, this place requires at least one token in the initial marking to allow its post-transition to fire sometimes. Contrary, all other places are unbounded, i.e. the token amount may amount up to infinity.
- CTI - The Petri net is covered by T-invariants. There are the following minimal T-invariants for the Petri net without the positive feedback:
  - y1 = {T F generation, T F degradation},
  - y2 = {T F S complex production, T F S complex disassociation},
  - y3 = {T F generation, T F S complex production, T F S degradation},
  - y4 = {P hz M S production, P hz M S degradation},
  - y5 = {P Y O production, P Y O degradation} .
- The Petri net with the positive feedback has additionally the following two T-invariants:
  - y6 = {pf b, T F degradation},
  - y7 = {pf b, T F S complex production, T F S degradation, } .
- SCTI - The Petri net is not strongly covered by T-invariants, i.e. by non- trivial T-invariants only.

A single enzyme-catalysed reaction in various modelling representations

Michaelis–Menten approximations | Mass-action kinetics



(A, B) Conventional notation of the chemical reactions and kinetic constants.

(C, D) A possible ODE representation.
The differential equations mathematically describe the temporal change of each molecular species.

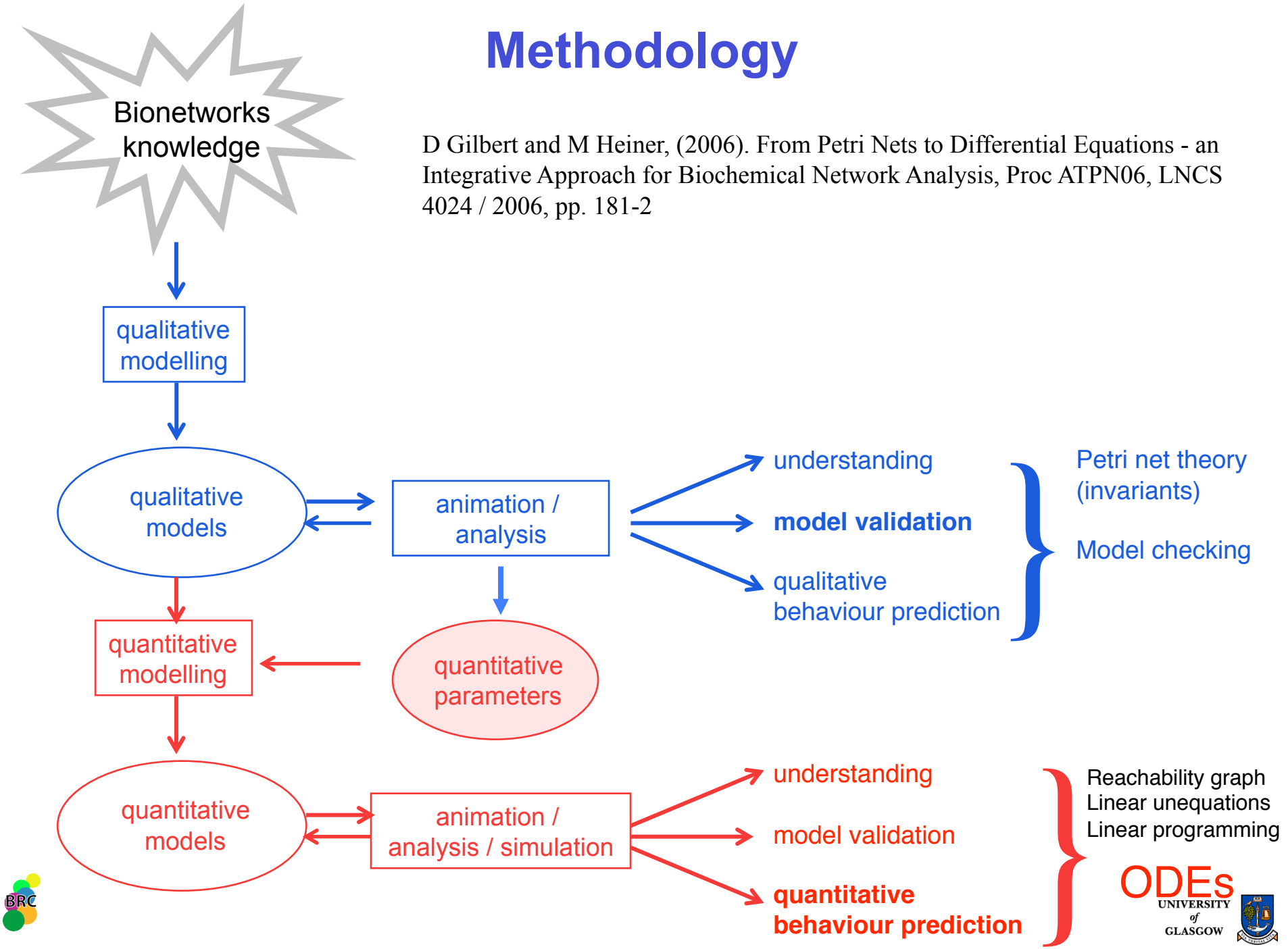
(E, F) Discrete Petri net description.
Circular nodes - biochemical entities and boxes represent reactions. Enzymatic catalysis in E is represented using a special read arc (circled end). The marking of circular nodes with tokens indicates whether the biochemical entity is present in the state of the model. Reactions may occur if their preceding biochemical entities are marked.

Gilbert, D. et al. Brief Bioinform 2006 7:339-353; doi:10.1093/bib/bbl043

Briefings in Bioinformatics

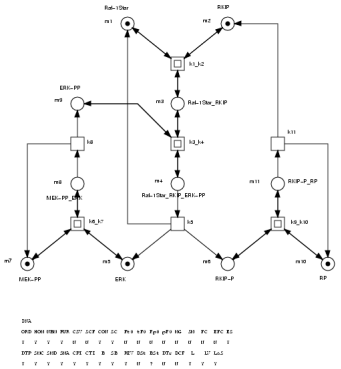
Methodology

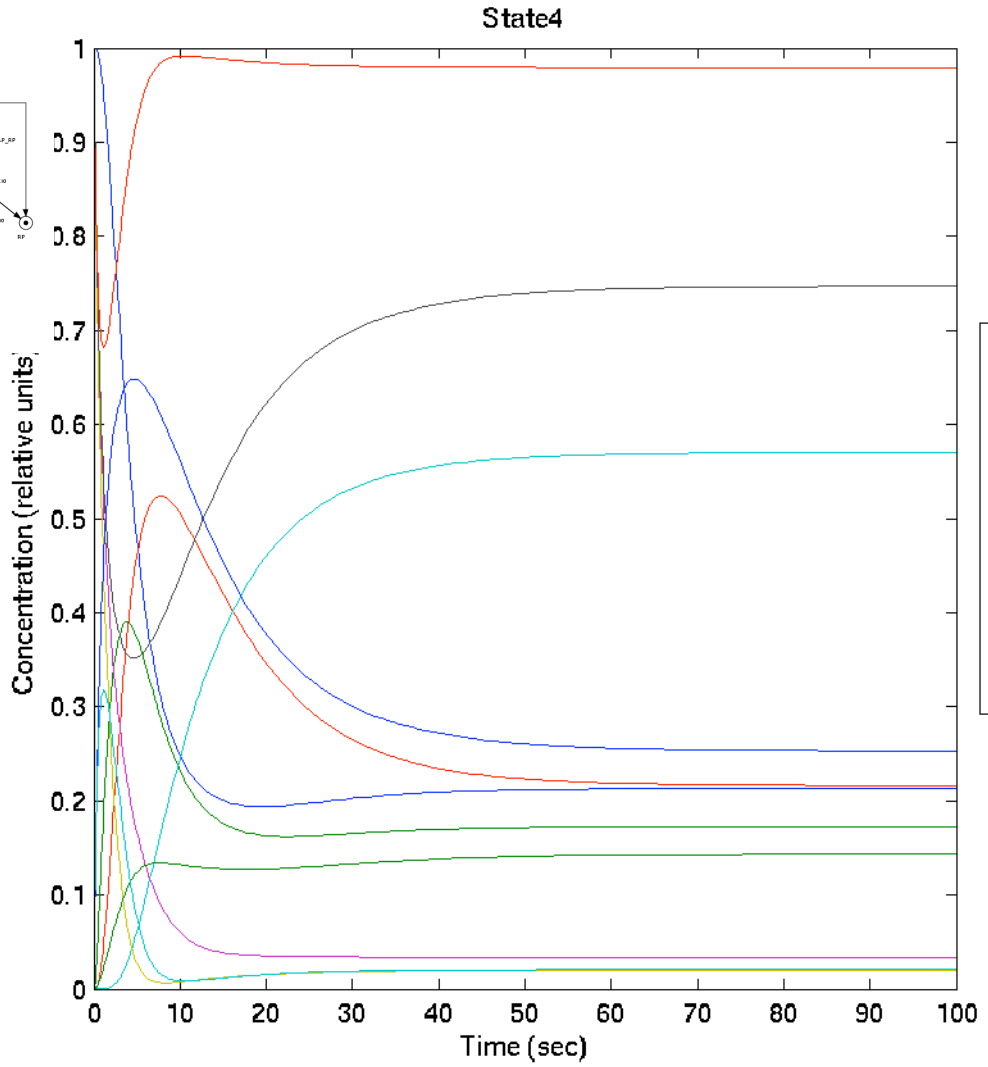
D Gilbert and M Heiner, (2006). From Petri Nets to Differential Equations - an Integrative Approach for Biochemical Network Analysis, Proc ATPN06, LNCS 4024 / 2006, pp. 181-2



From Petri Nets to Differential Equations - an Integrative Approach

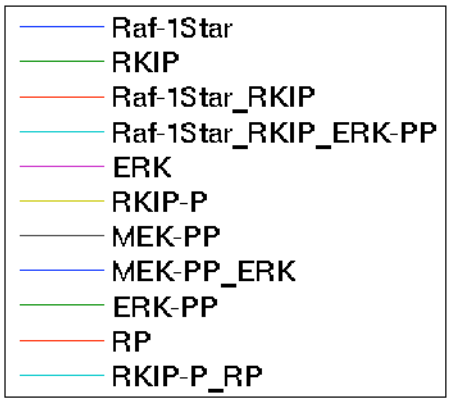
David Gilbert & Monika Heiner

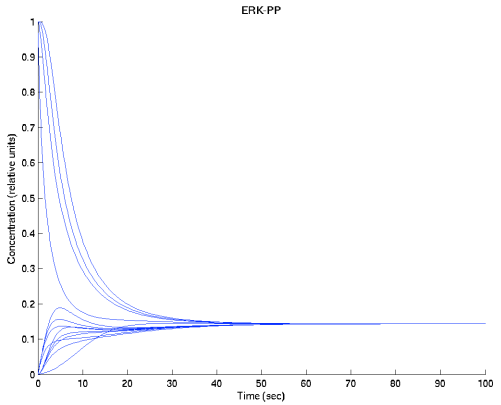




Species	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Raf-1*	1	0	0	1	1	1	1	1	0	0	1	1	1
RKIP	1	0	0	0	0	0	0	1	0	0	1	0	0
Raf-1*	0	1	0	0	0	0	0	0	1	1	0	0	0
Raf-1*/RKIP/ERK-PP	0	0	1	0	0	0	0	0	0	0	0	0	0
ERK	0	0	0	1	0	0	1	1	1	0	0	0	0
RKIP-P	0	0	0	1	1	0	0	0	0	0	0	0	1
MEK-PP	1	1	1	1	0	0	1	1	1	0	0	1	1
MEK-PP/ERK	0	0	0	0	1	1	0	0	0	1	1	0	0
ERK-PP	1	1	0	0	0	0	0	0	0	0	0	1	1
RP	1	1	1	1	1	0	0	1	1	1	1	0	1
RKIP-P/RP	0	0	0	0	0	1	1	0	0	0	0	1	0

Initial 13 'good' state configurations





UNIVERSITY of GLASGOW

Stochastic representations of the single enzyme-catalysed reaction

A. PEPA

$E_H = (r_1,k_1).E_L$

$E_L = (r_2,k_2).E_H + (r_3,1).E_H$

$S_H = (r_1,k_1).S_L$

$S_L = (r_2,k_2).S_H$

$SE_H = (r_2,k_2).SE_L + (r_3,k_3).SE_L$

$SE_L = (r_1,k_1).SE_H$

$P_L = (r_3,k_3).P_H$

$P_H = (\text{stop},1).P_H$

$$S_H \underset{\{r_1,r_2\}}{\bowtie} (SE_L \underset{\{r_1,r_2,r_3\}}{\bowtie} E_H) \underset{\{r_3\}}{\bowtie} P_L$$

B. Stochastic π-calculus

$E(k_1) \stackrel{\Delta}{=} (\nu k_2)(\nu k_3)(!k_1(k_2,k_3).(?k_2.E(k_1) + ?k_3.E(_1))$

$S(k_1) \stackrel{\Delta}{=} ?k_1(k_2,k_3).(!k_2.K(k_1) + !k_3.P())$

run 100 of $E(a) \mid S(a)$

(A): Stochastic process algebra description in PEPA

•The upper part defines the biochemical components, where the concentrations of each one can be either high or low (e.g. for the substrate either SH or SL). The reactions are referred to by the labels r1,r2,r3 and k1,k2,k3 represent the rates.

•The last line describes how the components are composed together to form the model.

•Simulations are via ODEs or the Gillespie algorithm and queries about the model can be made with the PRISM model checker.

(B): Stochastic π-calculus description

•The first two lines are rules describing the behaviours of the enzyme and substrate respectively.

•The product is also defined in the second rule.

•The third line is the instruction to simulate the model with 100 molecules each of the enzyme and substrate using the Gillespie algorithm.

Gilbert, D. et al. Brief Bioinform 2006 7:339-353; doi:10.1093/bib/bbl043

Briefings in Bioinformatics

Advantages and disadvantages of stochastic modelling

- Living systems are intrinsically stochastic due to low numbers of molecules that participate in reactions
- Gives a better prediction of the model on a cellular level
- Allows random variation in one or more inputs over time
- Slow simulation time

Glasgow iGEM team

Chemical Master Equations

A set of linear, autonomous ODE's, one ODE for each possible state of the system. The system may be written:

- **Φ → TF** - production of TF
- **TF → Φ** - degradation of TF
- **TF+S → TFS** - association of TFS
- **TFS → TF+S** - dissociation of TFS
- **TFS → Φ** - degradation of TFS
- **Φ → PhzMS** - production of PhzMS
- **PhzMS → Φ** - degradation of PhzMS
- **PhzMS → PYO** - production of pyocyanin
- **PYO → Φ** - degradation of pyocyanin

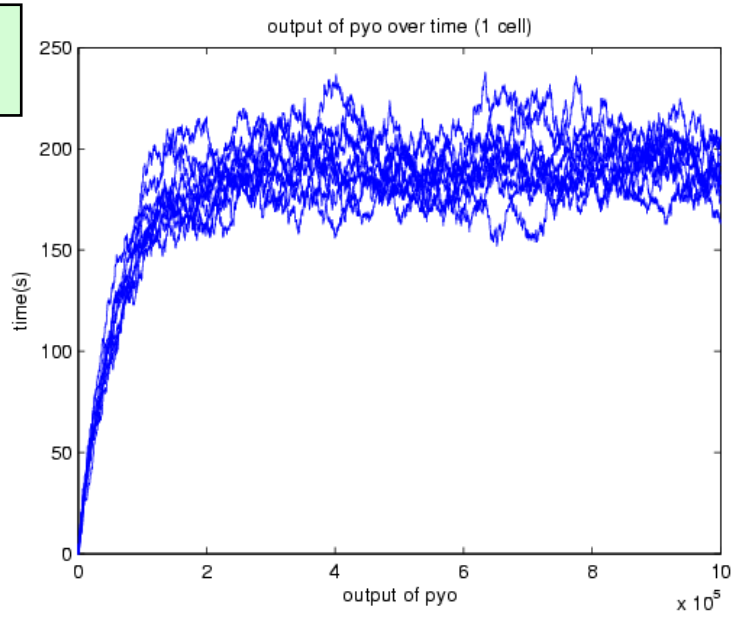
- Propensity functions:

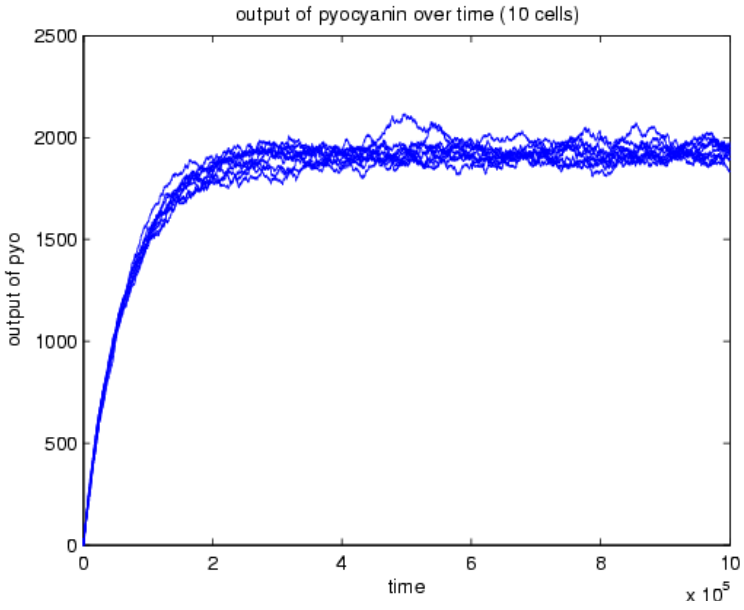
reaction	rate constant	propensity function
$\phi \rightarrow TF$	$\alpha = c(1)$	$a(1) = c(1)$
$TF \rightarrow \phi$	$\delta_{TF} = c(2)$	$a(2) = c(2) * X(1)$
$TF + S \rightarrow TFS$	$K1 * S = c(3)$	$a(3) = c(3) * X(1)$
$TFS \rightarrow TF + S$	$Km1 = c(4)$	$a(4) = c(4) * X(2)$
$TFS \rightarrow \phi$	$\delta_{TFS} = c(5)$	$a(5) = c(5) * X(2)$
$\phi \rightarrow P3$	$\frac{\beta * TFS}{\gamma + TFS} = c(6)$	$a(6) = c(6)$
$P3 \rightarrow \phi$	$\delta_{P3} = c(7)$	$a(7) = c(7) * X(3)$
$P3 \rightarrow P4$	$\alpha_2 = c(8)$	$a(8) = c(8) * X(3)$
$P4 \rightarrow \phi$	$\delta_{P4} = c(9)$	$a(9) = c(9) * X(4)$

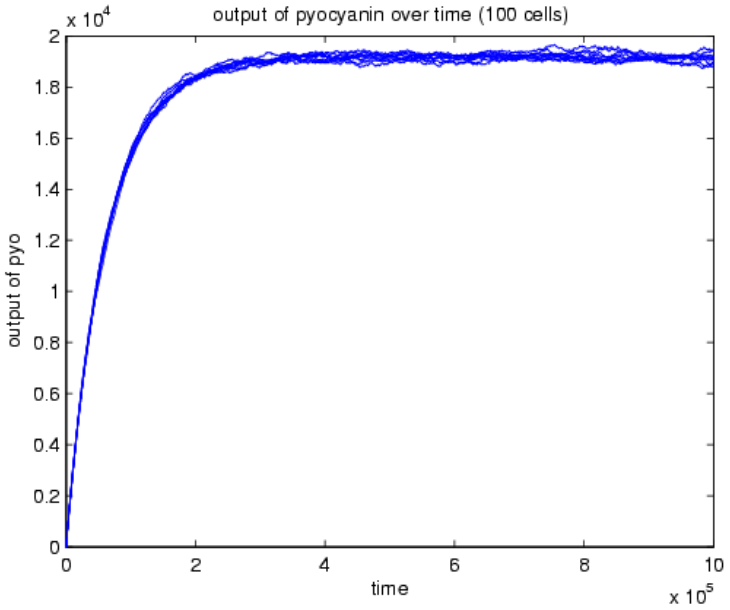
BRC

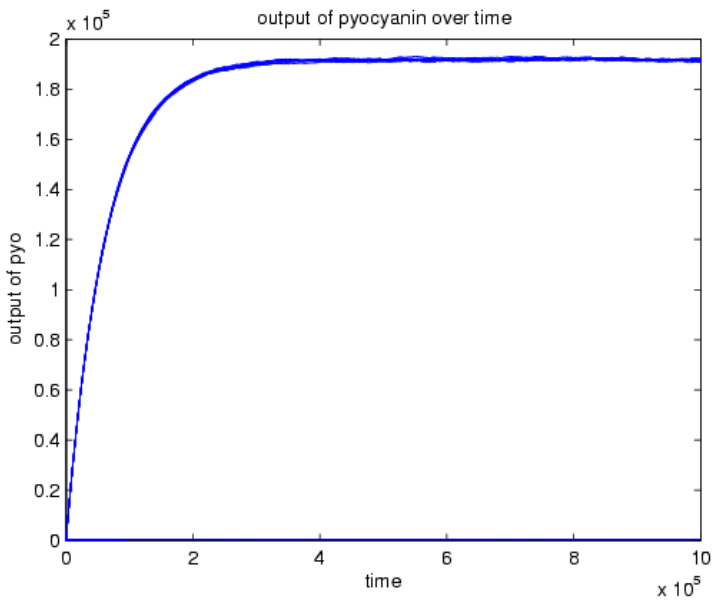
UNIVERSITY of GLASGOW

Glasgow iGEM team









UNIVERSITY of GLASGOW

Robin Donaldson

Probabilistic temporal logic

X	next
F	finally
G	globally
U	until
{…}	filter (check property from when property becomes true)
P	probability

$P_{=?}$ [([ProteinX] = L) U ([ProteinX] > L) {[ProteinX] = L}]

What is the probability of the concentration of ProteinX increasing, when starting in a state where the level is already at K?

Can also query about oscillations

F(d[ProtX]>0 ∧ F(d[ProtX]<0 ∧ F(d[ProtX]>0 ∧…)))

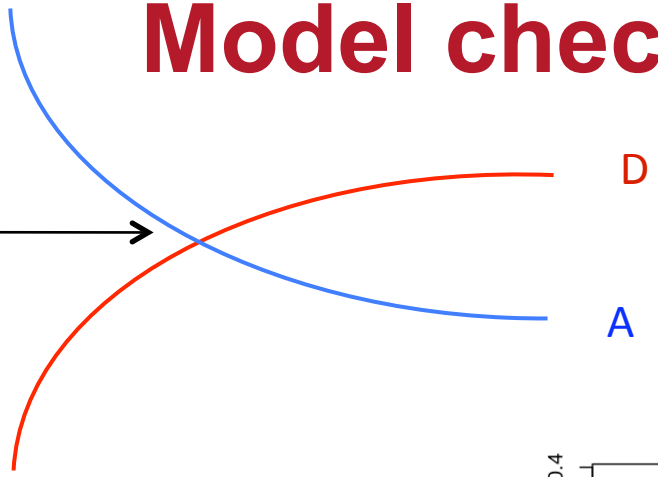
UNIVERSITY of GLASGOW

Robin Donaldson

Model checking

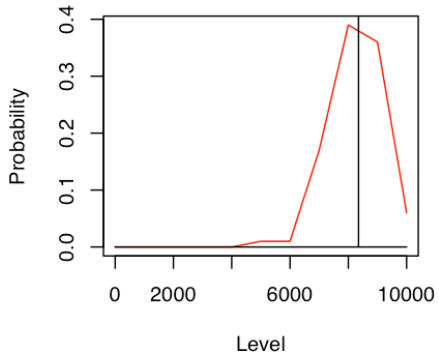
Two reactions:
(1) A->B
(2) C->D

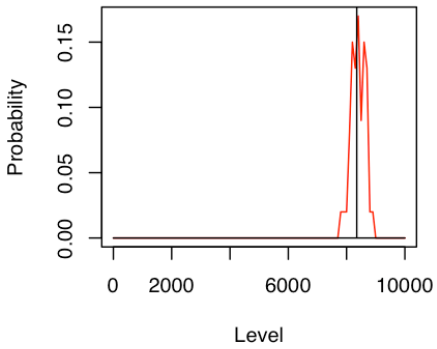
At what concentration?



Property: P=?[([A] = X) { [A]=[D] }]

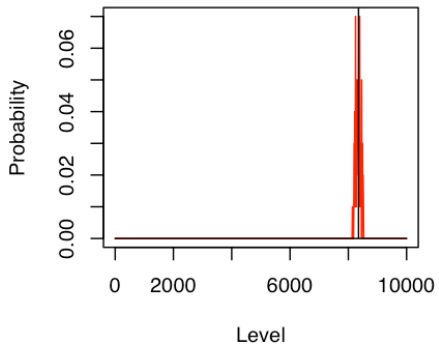
Assessing [X] at which reactant [A] equals product [D]

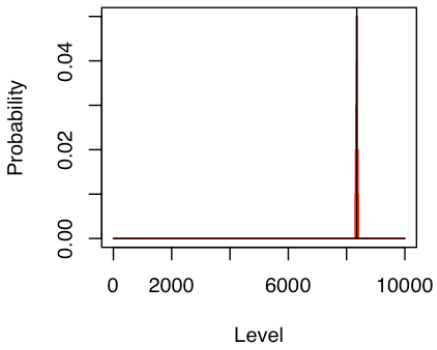




Results using 10, 100, 1000, 10000 levels.

1000 levels: peaks at 837, I.e. 8.37 the most probable concentration when [A] = [D]





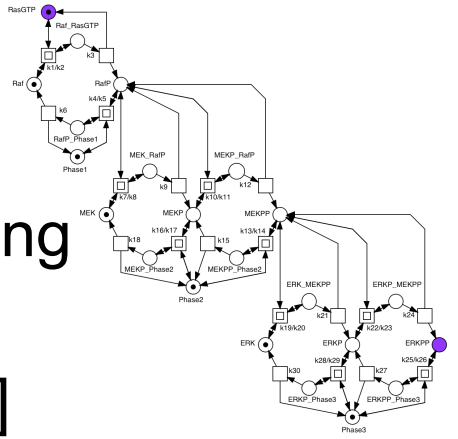
ODE simulation; [A] = [D] at concentration ~8.35

BRC

UNIVERSITY of GLASGOW

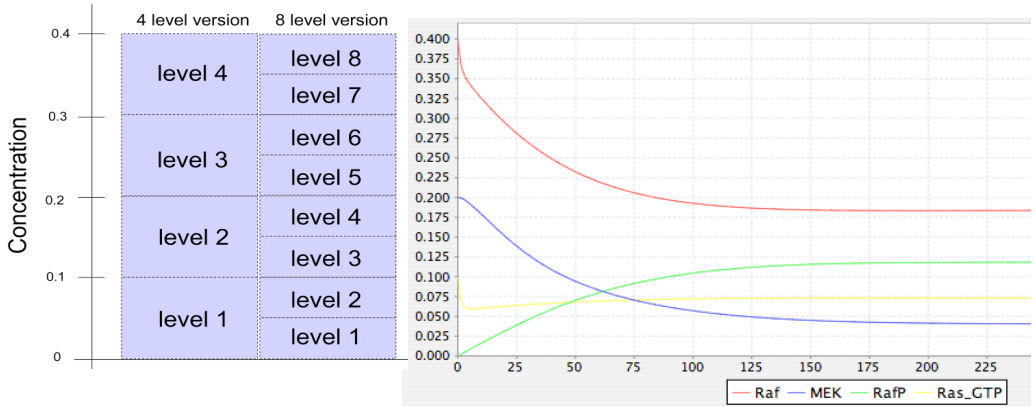
Robin Donaldson

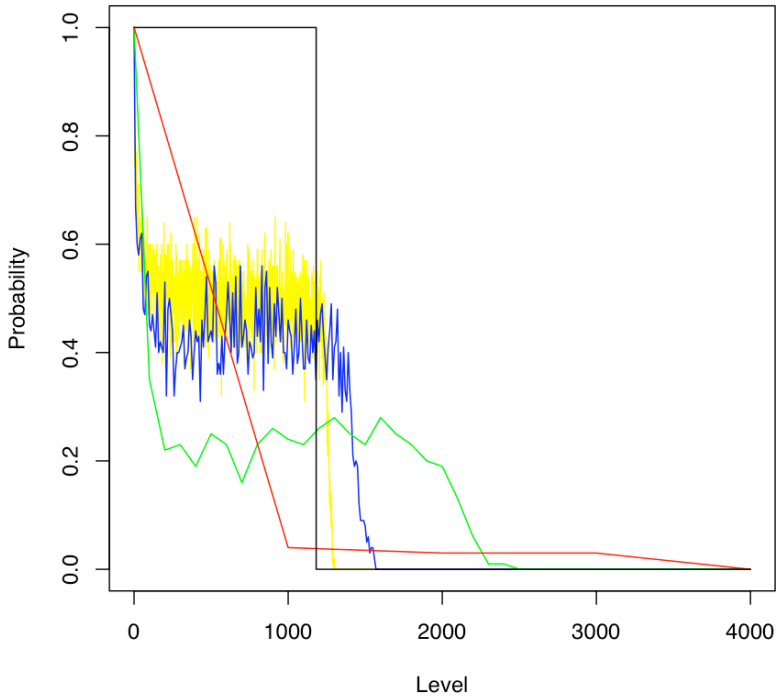
Probabilistic model checking



- Property S1: What is the probability of the concentration of RafP increasing, when starting in a state where the level is already at K?

$$P_{=?}\ [\ ([RafP] = L)\ U\ ([RafP] > L)\ \{[RafP] = L\}]$$





- *Stochastic: 4 (red), 40 (green) 400 (blue), 4000 (yellow) levels*
- *Extensible to thousands*
- *Approximates to deterministic behaviour (**black**) 0.1182...*

M. Calder, V. Vyshemirsky, D. Gilbert and R. Orton. (2006). Analysis of Signalling Pathways using Continuous Time Markov Chains, Trans.on Computat. Syst. Biol. VI, 4220 pp 44-67 Springer Verlag

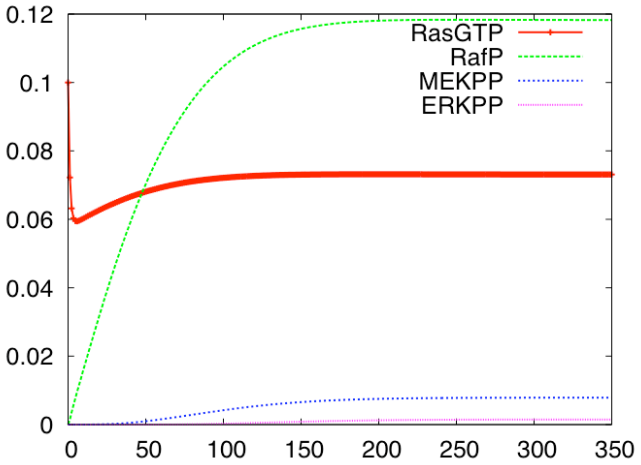
UNIVERSITY of GLASGOW

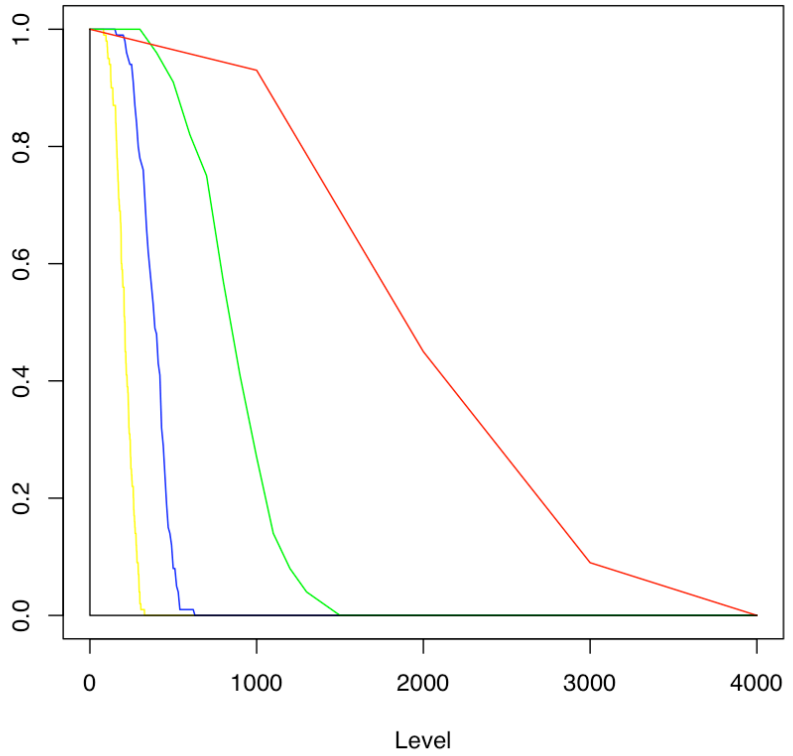
Robin Donaldson

Probabilistic model checking

- Property S2: What is the probability RafP being the first species to react?

$$P_{=?}\,[\,((\,[MEKPP] = 0) \wedge (\,[ERKPP] = 0)) \,U\, (\,[RafP] > L)$$
$$\{(\,[MEKPP] = 0) \wedge (\,[ERKPP] = 0) \wedge (\,[RafP] = 0)\}$$

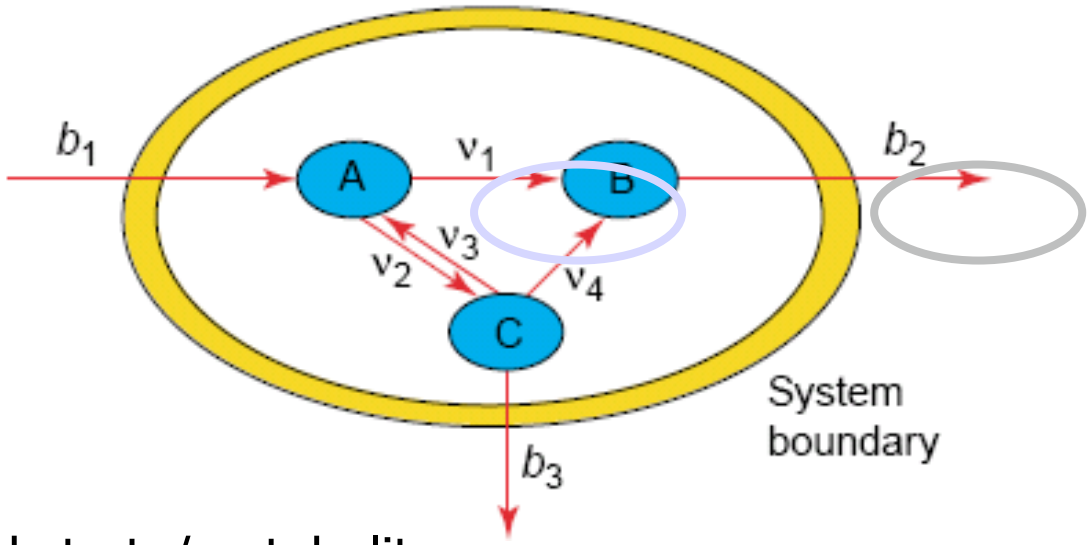




- *Stochastic: 4 (red), 40 (green)*
  400 (blue), 4000 (yellow) levels
- *Extensible to thousands*

Lecture notes by Eran Eden

Flux balance analysis



Vertex - substrate/metabolite concentration.

Edge - flux (conversion mediated by enzymes of one substrate into the other)

Internal flux edge

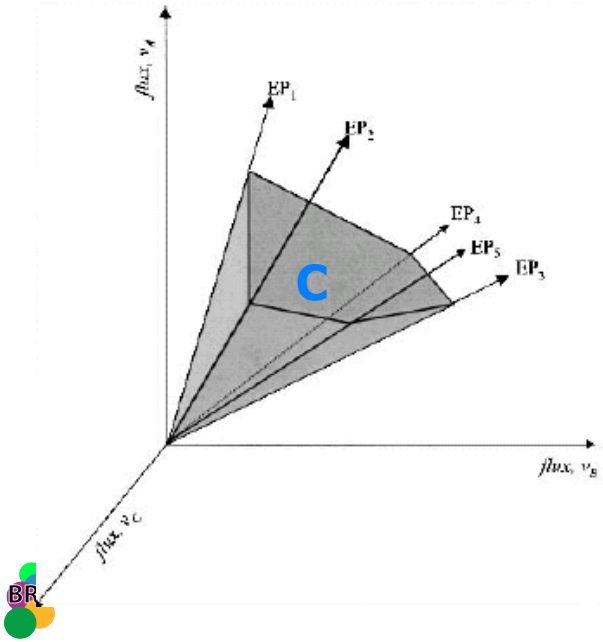
External flux edge

Flux cone and metabolic capabilities

$$0 = S \cdot v$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (columns) · v (rows)



The number of reactions considerably exceeds the number of metabolites

↓

The S matrix will have more columns than rows

↓

The null space of viable solutions to our linear set of equations contains an infinite number of solutions.

What about the constraints?

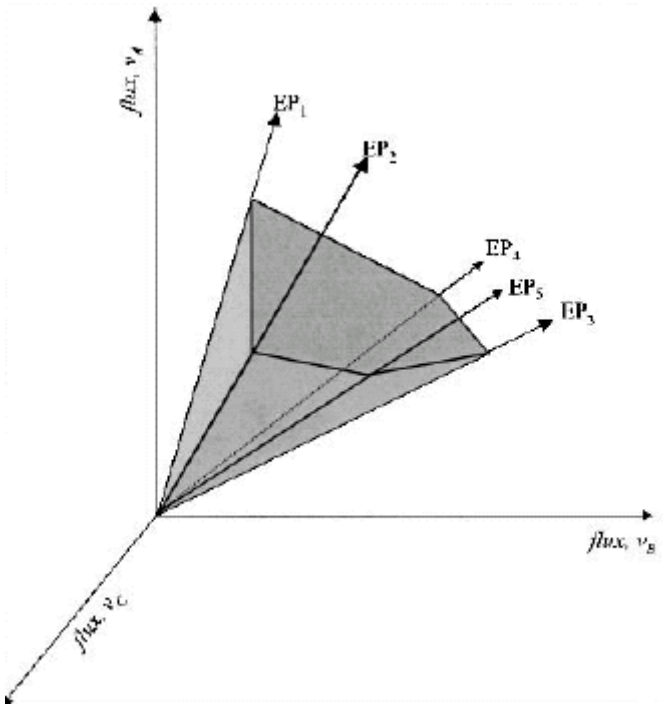
↓

"The solution space for any system of linear homogeneous equations and inequalities is a convex polyhedral cone." - Schilling 2000

Our flux cone contains all the points of the null space with non negative coordinates (besides exchange fluxes that are constrained to be negative or unconstrained)

UNIVERSITY of GLASGOW

Flux cone and metabolic capabilities



What is the significance of the flux cone?

- It defines what the network can do and cannot do!
- Each point in this cone represents a flux distribution in which the system can operate at steady state.
- The answers to the following questions (and many more) are found within this cone:
  - what are the building blocks that the network can manufacture?
  - how efficient is energy conversion?
  - Where is the critical links in the system?

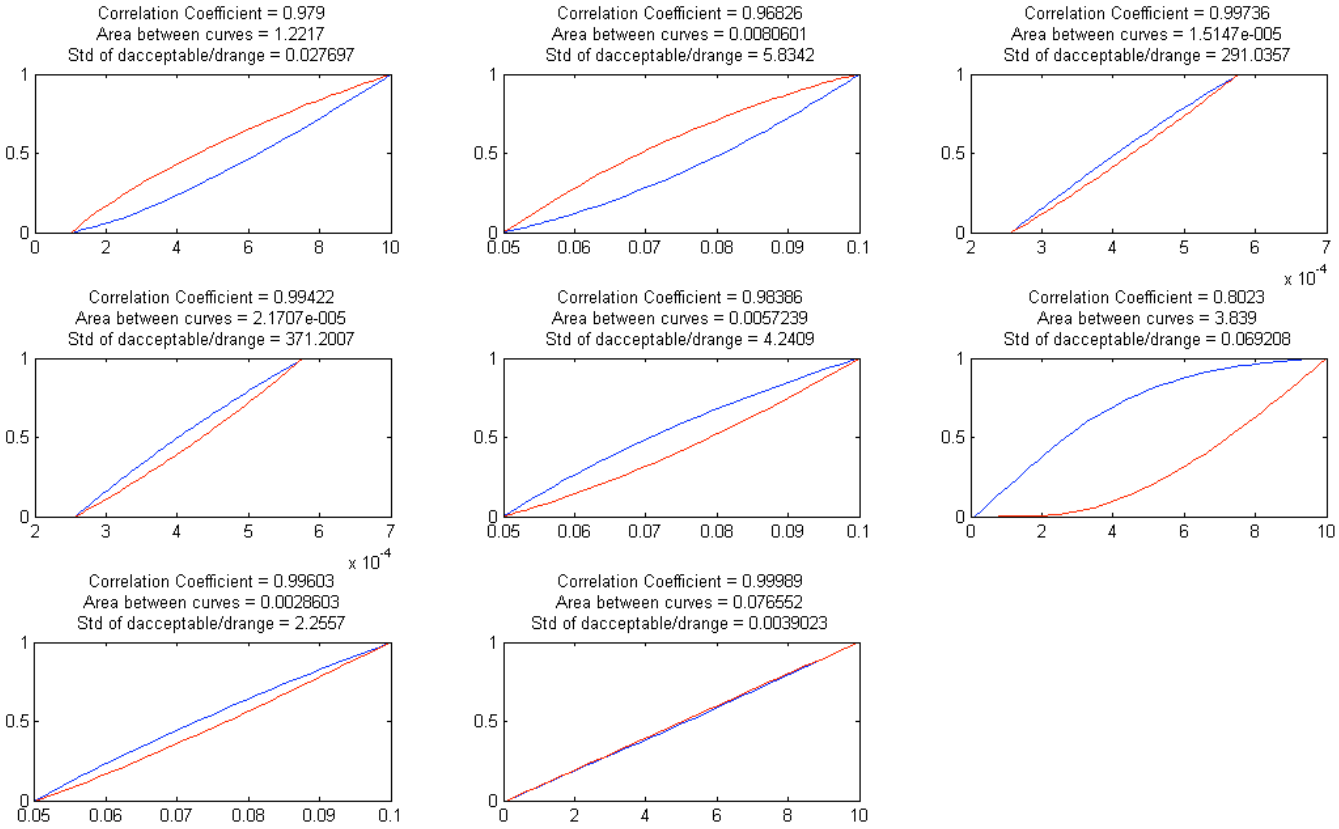
Metabolic control analysis

- Quantitative sensitivity analysis of fluxes and metabolite concentrations.
- In MCA one studies the relative control exerted by each step (enzyme) on the system's variables (fluxes and metabolite concentrations).
- This control is measured by applying a perturbation to the step being studied and measuring the effect on the variable of interest after the system has settled to a new steady state.

Glasgow iGEM team

Model Parameter Refinement

- Modified MPSA



UNIVERSITY of GLASGOW

Fitting and optimization

- Genetic Algorithms
- Simulated Annealing

Dynamic behaviour analysis

- Bifurcation analysis (to discover oscillations)

Databases & tools

Databases

Name	Content	Website
TRANSPATH [62]	Signalling pathways	http://www.biobase.de/pages/index.php?id=39
aMAZE [63]	Annotated protein interactions	http://www.amaze.ulb.ac.be/
KEGG [64]	Annotated metabolic and signalling pathways	www.genome.ad.jp/kegg
BRENDA [65]	Enzyme function and kinetic data	www.brenda.uni-koeln.de
KDBI [66]	Kinetic data	xin.cz3.nus.edu.sg/group/kdbi/kdbi.asp
BioModels [69]	Dynamic model repository	www.ebi.ac.uk/biomodels
DOQCS [70]	Dynamic model repository	doqcs.ncbs.res.in
CellML model repository [67]	Dynamic model repository	www.cellml.org/models

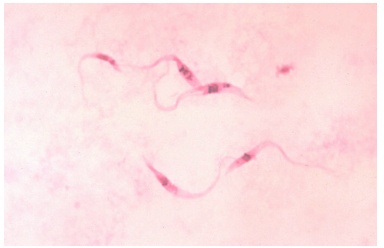
Tools

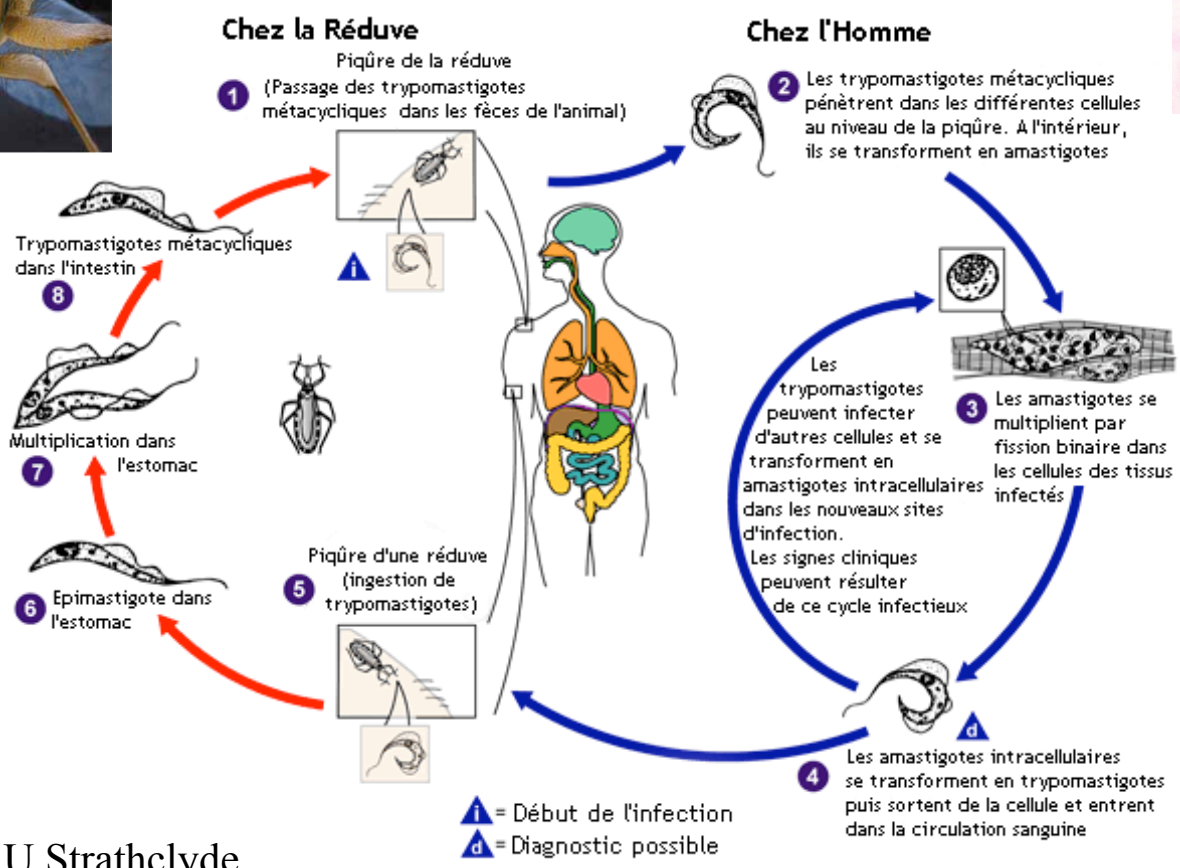
Name	Category	Model Representation	Function	URL
MATLAB, with SimBiology Toolbox [71]	Continuous and stochastic	Mathematical (e.g. ODE)	General-purpose mathematical environment, simulation and analysis	www.mathworks.com
XPPAut	Continuous and stochastic	ODE	General purpose; simulation, analysis	www.math.pitt.edu/~bard/xpp/xpp.html
Copasi [73]	Continuous and stochastic	ODE	Simulation and analysis	www.copasi.org
Virtual Cell [75]	Continuous and stochastic	ODE-based, PDE	Simulation and parameter sensitivity analysis	www.nrcam.uchc.edu
Systems Biology Workbench [76], including Jarnac and JDesigner	Discrete, continuous and stochastic	ODE/SBML	Data-exchange framework for modelling, simulation and analysis	sbw.kgi.edu
Narrator [15]	Continuous and stochastic	Graphical, ODE-based	Modelling and simulation	www.narrator-tool.org
STOCHSIM [78]	Stochastic	Probabilistic	General-purpose biochemical simulator	www.pdn.cam.ac.uk/groups/comp-cell/StochSim.html
E-CELL [77]	Continuous	Object-oriented	Modelling and simulation	www.e-cell.org
SPiM [83]	Stochastic	Π-calculus	Simulation	http://www.doc.ic.ac.uk/~anp/spim/
BioSigNet [85]	Discrete	Graphical	Reasoning, hypothesis testing	http://www.public.asu.edu/~cbaral/biosignet
BIOCHAM [84]	Discrete and continuous	Logical + kinetic models	Simulation and analysis	contraintes.inria.fr/BIOCHAM
PRISM [24]	Discrete	Stochastic process algebra	General purpose; Analysis (model checking)	http://www.cs.bham.ac.uk/~dxp/prism
PEPA Workbench [20]	Discrete	Stochastic process algebra	General purpose; Analysis	www.dcs.ed.ac.uk/pepa/tools

Done

UNIVERSITY of GLASGOW

Example - SyTryp project





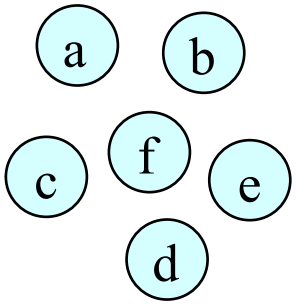
U.Glasgow, U.Strathclyde
INRA, INRIA, Bordeaux

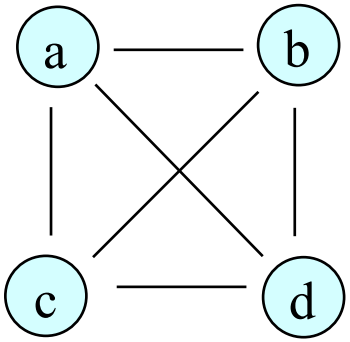
UNIVERSITY of GLASGOW

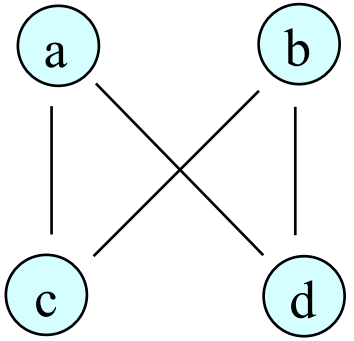
From connectivities to dynamic graphs (1)

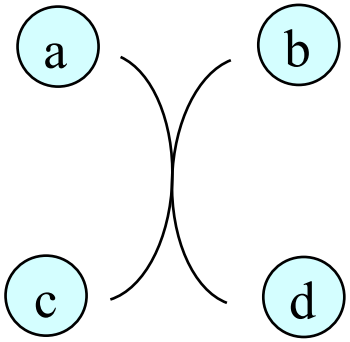
- Start with *metabolite relation networks* generated by inference techniques
- Focus on selected modules of interest which have been identified by clustering and confirmed, using visualization proposed to be of potential interest by the biologists.
- Transform these into *reaction networks*
  - manually mapping metabolite relationships onto known metabolic networks from databases for model organisms such as E.coli using data from MetaCyc and Kegg,
  - automated using Bayesian networks (see review Werhli et al, Bioinformatics, 2006).
- Reaction networks: *bipartite graphs* (Petri nets),
  - metabolites are represented by one type of vertex,
  - reactions & the enzymes catalyzing the reaction by another vertex type.
  - edges decorated with stoichiometric information derived from the databases.
  - Default: reactions as reversible, unless sufficient information.
- Validate qualitative reaction networks using Petri net analysis techniques & tools
  - consistency check of elementary graph properties
  - identification of mass-conserving and state-reproducing subnetworks
  - identification of smallest possible functional units
  - model checking of expected qualitative behavioural properties (e.g flux balance and elementary mode analysis).
- Derive meaningful initial markings (hence initial concentrations for the quantitative models)

From ab-initio & correlations to reactions…

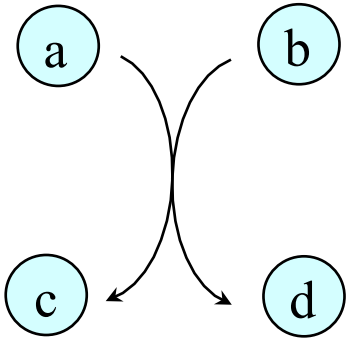




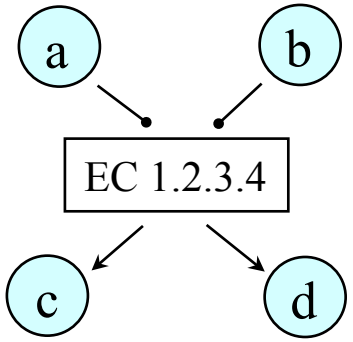




Prune edges using bayesian approach? (Muggleton)

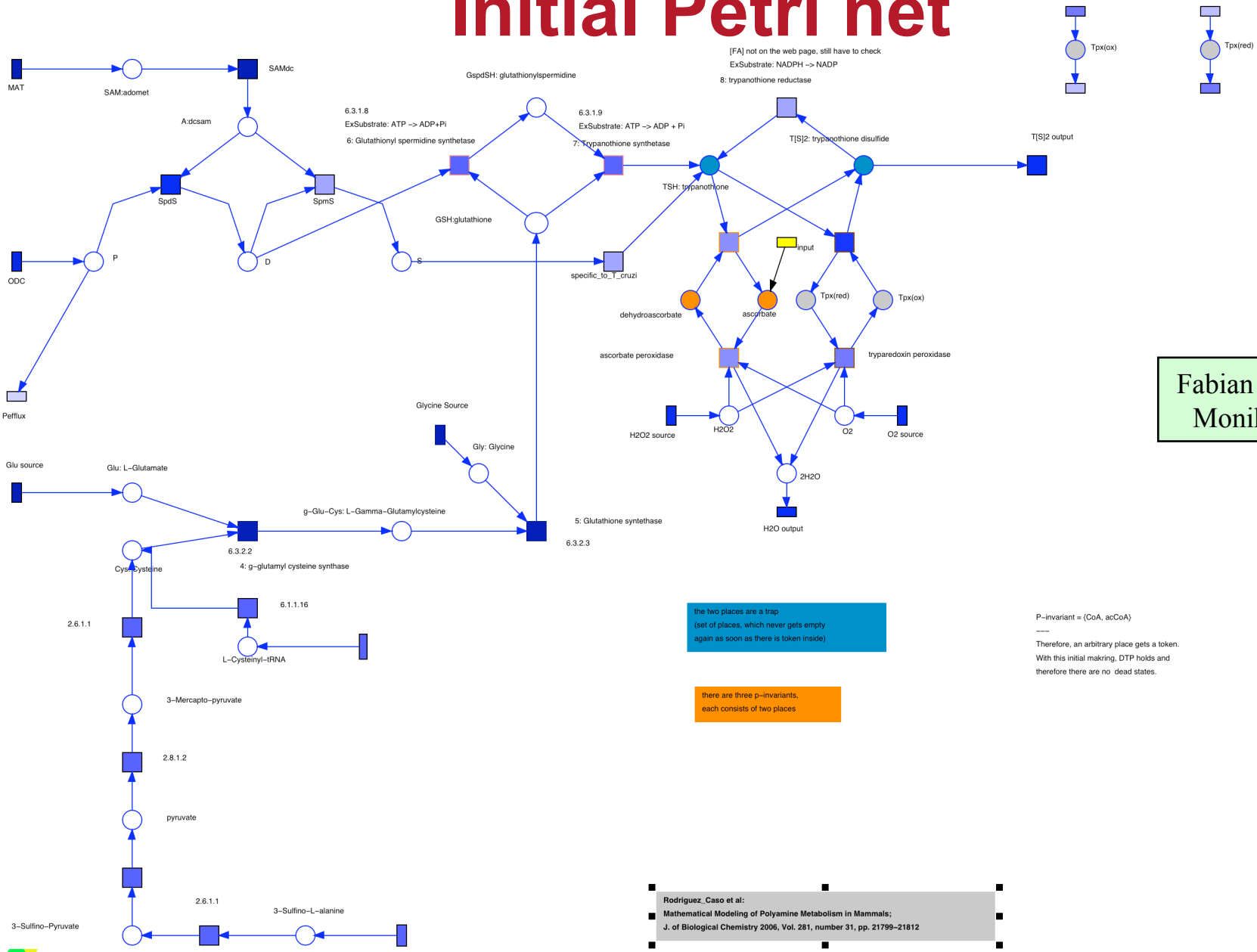


Directionality by mapping onto existing reaction databases?



Bipartite graph

Initial Petri net



Fabian Jourdan &
Monika Heiner

the two places are a trap
(set of places, which never gets empty
again as soon as there is token inside)

there are three p–invariants,
each consists of two places

P–invariant = {CoA, acCoA}

Therefore, an arbitrary place gets a token.
With this initial makring, DTP holds and
therefore there are no dead states.

Rodriguez_Caso et al:
Mathematical Modeling of Polyamine Metabolism in Mammals;
J. of Biological Chemistry 2006, Vol. 281, number 31, pp. 21799–21812

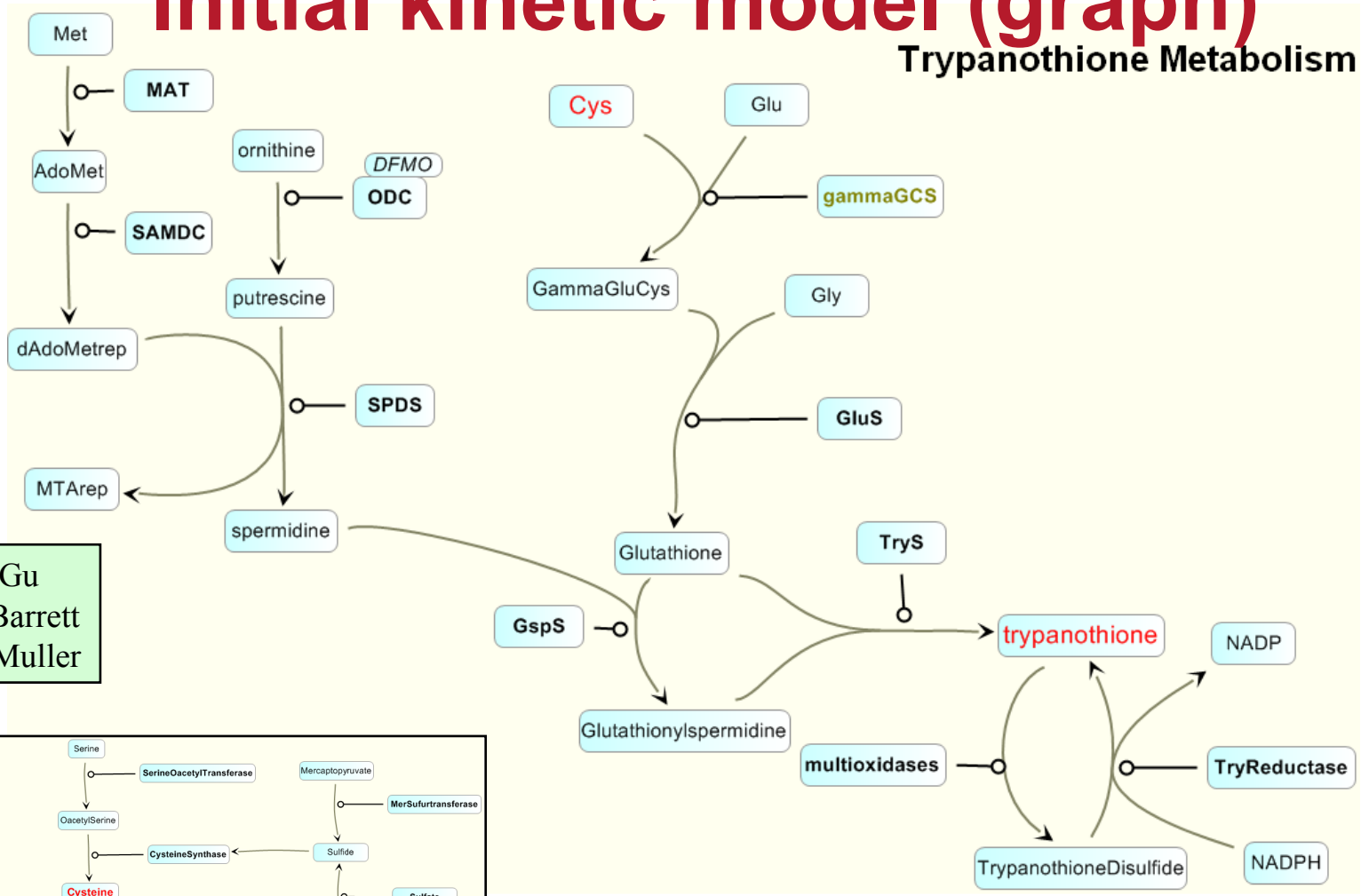
BRC

UNIVERSITY of GLASGOW

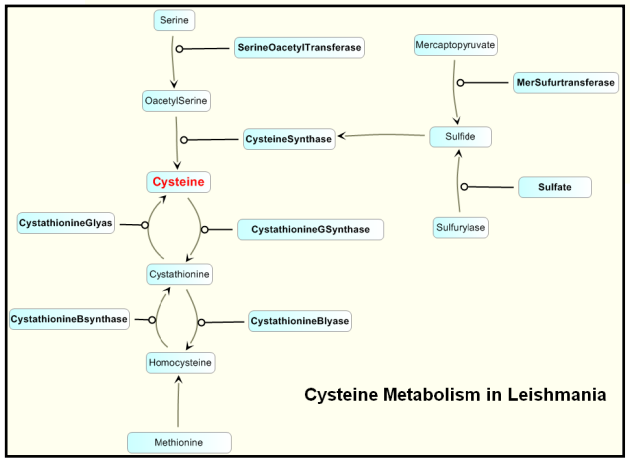
From connectivities to dynamic graphs (2)

- Transform validated qualitative model into quantitative models (stochastic & continuous Petri nets) by retaining the structure and adding quantitative parameters.
- Rate parameters from public domain databases (e.g. MetaCyc, Kegg, Brenda), or literature (via PubMed queries, or the text-mining).
- Estimate the remaining parameters for steady state behaviour, using previously generated knowledge of state-reproducing subnetworks.
- Validate stochastic/continuous Petri nets using
  - probabilistic/continuous model checking of the stochastic/continuous counterparts of the qualitative behavioural properties
  - simulation-based transient and steady-state analyses.
- Sensitivity analysis
- Refine the rate parameters by scanning or fitting.
  - Bayesian inference of model parameters from the experimental data, (Vyshemirsky & Girolami) - families of behaviours, generating distributions over rate parameters. Also identification of the most likely reaction network topologies from alternatives generated from the metabolite relation networks.
- **The model-based design of knockout experiments will additionally help in validating the developed quantitative models**

Initial kinetic model (graph)



Xu Gu
Mike Barrett
Sylke Muller



Trypanothionine Initial ODE model (1)

Equations and parameters	References
ODE	
$V_{ODC} = \frac{V_{max}^{ODC} * [\text{Orn}]}{K_M^{ODC}\left(1+\frac{[\text{P}]}{K_{ip}^{ODC}}\right)+[\text{Orn}]}$	(11)
SAMdc	
$V_{SAMdc} = \frac{V_{max}^{SAMdc}}{1+\frac{[\text{S}]}{K_{is}^{SAMdc}}} * \frac{[\text{SAM}]}{K_M^{SAMdc}\left(1+\frac{K_{aP}^{SAMdc}}{[\text{P}]}+\frac{[\text{dSAM}]}{K_{idSAM}^{SAMdc}}\right)+[\text{SAM}]}$	(11)
MAT	
$V_{MAT} = \frac{V_{max}^{MAT}}{1+\left(\frac{K_M^{MAT}}{[\text{Met}]}\right)*\left(1+\frac{[\text{SAM}]}{K_{iMet}^{MAT}}\right)}$	(11)
SpdS[a]	
$V_{SpdS} = \frac{V_{max}^{SpdS} * [\text{dSAM}] * [\text{P}]}{K_{dSAM}^{SpdS}*\left(1+\frac{[\text{MTA}]}{K_{iMTA}^{SpdS}}\right)*K_p^{SpdS}*\left(1+\frac{[\text{D}]}{K_{iD}^{SpdS}}\right)+K_P^{SpdS}*\left(1+\frac{[\text{D}]}{K_{iD}^{SpdS}}\right)*[\text{dSAM}]+K_{dSAM}^{SpdS}*\left(1+\frac{[\text{dSAM}]}{K_{iMTA}^{SpdS}}\right)*[\text{P}]+[\text{dSAM}]*[\text{P}]}$	(1)
SpmS[a]	
$V_{SpmS} = \frac{V_{max}^{SpmS} * [\text{dSAM}] * [\text{D}]}{K_{dSAM}^{SpmS}*\left(1+\frac{[\text{MTA}]}{K_{iMTA}^{SpmS}}\right)*K_D^{SpmS}*\left(1+\frac{[\text{S}]}{K_{iS}^{SpmS}}\right)+K_D^{SpmS}*\left(1+\frac{[\text{S}]}{K_{iS}^{SpmS}}\right)*[\text{dSAM}]+K_{dSAM}^{SpmS}*\left(1+\frac{[\text{dSAM}]}{K_{iMTA}^{SpmS}}\right)*[\text{D}]+[\text{dSAM}]*[\text{D}]}$	(1)
γGCS[b]	
$V_{\gamma GCS} = \frac{\frac{V_{max}^{\gamma GCS} * [\text{Glu}]*[\text{Aba}]*[\text{ATP}]}{\alpha*\beta*\gamma*K_{Glu}*K_{Aba}*K_{ATP}}}{1+\frac{[\text{Glu}]}{K_{Glu}}+\frac{[\text{Aba}]}{K_{Aba}}+\frac{[\text{ATP}]}{K_{ATP}}+\frac{[\text{Glu}]*[\text{Aba}]}{\gamma*K_{Glu}*K_{Aba}}+\frac{[\text{Glu}]*[\text{ATP}]}{\beta*K_{Glu}*K_{ATP}}+\frac{[\text{Aba}]*[\text{ATP}]}{\alpha*K_{Aba}*K_{ATP}}+\frac{[\text{Glu}]*[\text{Aba}]*[\text{ATP}]}{\alpha*\beta*\gamma*K_{Glu}*K_{Aba}*K_{ATP}}}$	(2)
γGCS[c]	
$V_{\gamma GCS} = \phi_0 + \frac{\phi_1}{[\text{Glu}]} + \frac{\phi_2}{[\text{ATP}]} + \frac{\phi_3}{[\text{Ala(CI)}]} + \frac{\phi_{12}}{[\text{ATP}]*[\text{Glu}]} + \frac{[\text{GSH}]}{[\text{Glu}]}\left(\frac{\phi_1}{K_{ig}}+\frac{\phi_{12}}{K_{ig}*[\text{ATP}]}\right) + \frac{[\text{GSH}]*\phi_2}{[\text{ATP}]*K_{ig'}} + \left(\frac{[\text{Glu-Ala(CI)}]}{[\text{Glu}]}\right)\left(\frac{\phi_1}{K_{id}}+\frac{\phi_{12}}{[\text{ATP}]*K_{id}}\right) + \frac{[\text{ADP}]*\phi_2}{[\text{ATP}]*K_{iADP'}} + \left(\frac{[\text{ADP}]*\phi_2}{[\text{Ala(CI)}]}\right)\left(\frac{1}{K_{iADP}}+\frac{1}{[\text{ATP}]*K_{iADP}*K_{aATP}}+\frac{1}{[\text{ATP}]*[\text{Glu}]*K_{iADP}*K_{aATP}*K_{KaGlu}}\right)$	(4)

[a] The equation takes MTA into account, which behaves behaves as competitive inhibitor onto dAdoMet (dSAM)

[b] Only one Cys residue (Cys-319 in T.brucei γGCS) is invariant. Mutation of Cys-319 to Ala in T. brucei γGCS renders the enzyme insensitive to cystamine inactivation without significantly affecting the enzyme's catalytic effciency, kinetic mechanisms or substrate affinities.

[c] the equation includes the inhibitory terms resulting from the presence of glutathione (GSH) and all the inhibitor terms containing phosphate concentration have been omitted due to the lack of phosphate binding to enzymes species.

Xu Gu

UNIVERSITY of GLASGOW

Trypanothionine Initial ODE model (2)

Equations and parameters	References
GS[a]	
$V_{GS} = V_{max}^{GS}\left(\frac{\alpha*K_m*[\gamma\text{GluCys}]+[\gamma\text{GluCys}]^2}{\alpha*(K_m)^2+2*\alpha*K_m*[\gamma\ \text{GluCys}]+[\gamma\text{GluCys}]^2}\right)$	(5; 7; 8)
GS[b]	
$V_{GS} = \frac{V_{max}^{GS}}{1+\frac{[\text{GSH}]}{K_{iGSH}^{GS}}} * \frac{[\gamma\text{GluCys}]*[\text{Gly}]}{K_{iGly}^{GS}*K_{\gamma GC}^{GS}+K_{\gamma GC}^{GS}*[\text{Gly}]+K_{Gly}^{GS}*[\gamma\text{GluCys}]+[\gamma\text{GluCys}]*[\text{Gly}]}$	(6; 3)
GspS[c]	
$V_{GspS} = \frac{V_{max}^{GspS}}{1+\frac{[\text{TSH}]}{K_{iTSH}^{GspS}}} * \frac{[\text{D}]*[\text{GSH}]}{K_D^{GspS}\left(1+\frac{[\text{GspdSH}]}{K_{hD}^{GspS}}\right)*K_{GSH}^{GspS}\left(1+\frac{[\text{GspdSH}]}{K_{hG}^{GspS}}\right)+K_D^{GspS}\left(1+\frac{[\text{GspdSH}]}{K_{hD}^{GspS}}\right)+K_G^{GspS}\left(1+\frac{[\text{GspdSH}]}{K_{hG}^{GspS}}\right)+[\text{GSH}]*[\text{D}]}$	
TryS[d]	
$V_{TryS} = \frac{V_{max}^{TryS}*[\text{GSH}]*[\text{GspdSH}]}{K_{iGSH}^{TryS}*K_{GspdSH}^{TryS}+K_{GspdSH}^{TryS}*[\text{GSH}]+K_{GSH}^{TryS}*[\text{GspdSH}]+[\text{GSH}]*[\text{GspdSH}]}$	
Ts2S[e]	
$V_{max}^{Ts2S} = \frac{V_{max}^{Ts2S}*[\text{TryS}]}{K_m^{Ts2S}+[\text{Ts2S}]}$	
TryR	
$\text{TSH} + \text{NADP} \underset{\rightarrow}{\leftarrow} T[S]_2 + \text{NADPH}$	

[a] As investigated in (5), GS catalyzes the formation of a γ-glutamylcysteine phosphate-enzyme intermediate from γ-glutamylcysteine and ATP, then this acyl-phosphate intermediate is attacked by glycine to form GSH. Also, there is not product inhibition in the presence of the enzyme deficiency.

[b] As discussed in (6), the negative co-operativity observed for γ-glutamylcysteine binding to the rate enzyme was not found for the parasite protein. This may be due to the alteration of several amino acids in the γ-glutamylcysteine-binding site. GSH was dispalyed as uncompetitive inhibitor (10).

[c] As reported in (12), TSH did competitively inhibit catalysis of Gsp synthetase with 10mM GSH and 10mM spermidine. Also, as proved in (9), GspS involves two catalytic activities, first is to catalyze the synthesis of GspdSH and the second is to hydrolyze the substance-enzyme compound. Hence, GspdSH can be considered as competitive inhibitor as presented in (10)

[d] Assume there is no product inhibition back on the enzyme.

[e] *in silico parameters?*

Description	Experimental measurements	References
s-adenosylmethionine decarboxylase	$K_m = 0.38 \pm 0.15$ $V_{max} = 3s^{-1}(4\mu mol/min/mg)$ $[\text{AdoMetDC}]_{initial} = 6nM$ $[\text{AdoMet}]_{initial} = 0.04mM$	(Willert et al., 2007)
Ornithine decarboxylase	$K_m = 280 \pm 30\mu M$ $K_{iDFMO} = 220 \pm 70\mu M$ $V_{max} = 2.7 \times 10^6 nmolCO_2/h/mg$ $[\text{Ornithine}]_{initial} = 50\mu M$	(Phillips et al., 1998)

Xu Gu

UNIVERSITY of GLASGOW

Xu Gu

Kinetic data

Description	Experimental measurements	References
s-adenosylmethionine decarboxylase	$K_m = 0.38 \pm 0.15$ $V_{max} = 3s^{-1}(4\mu mol/min/mg)$ $[\text{AdoMetDC}]_{initial} = 6nM$ $[\text{AdoMet}]_{initial} = 0.04mM$	(Willert et al., 2007)
Ornithine decarboxylase	$K_m = 280 \pm 30\mu M$ $K_{iDFMO} = 220 \pm 70\mu M$ $V_{max} = 2.7 \times 10^6 nmolCO_2/h/mg$ $[\text{Ornithine}]_{initial} = 50\mu M$	(Phillips et al., 1998)

A black art?

UNIVERSITY of GLASGOW

Parameter identification

- Given network topology, reaction equations + observations
  - Derive/refine kinetic parameters from observed data

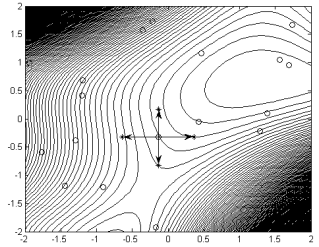
- Issue: Computational efficiency (time)

- Challenges - Data:
  - Partial (few time points, not all species)
  - Sparse (few repeated observations)
  - Noisy (experimental error, system variability)

- Can return multiple solutions

- Methods: multiple shooting, bayesian inference, …
- plus: sensitivity analysis, indentifiability of parameter dependence
- Optimisation problem (global, local)
- Model decomposition (helps with partial data)

PSwarm



- PSwarm - global optimization solver for bound constrained problems (author I.Vaz)
- Combines pattern search & particle swarm.
- PSO similar to evolutionary computation (e.g. GA)
  - system initialized with population of random solutions
  - searches for optima by updating generations.
  - no evolution operators (crossover and mutation…)
  - potential solutions (particles) fly through problem space by following the current optimum particles.
  - Faster convergence than GA
- Disadvantages: initial population & control parameters dependent; single solution.

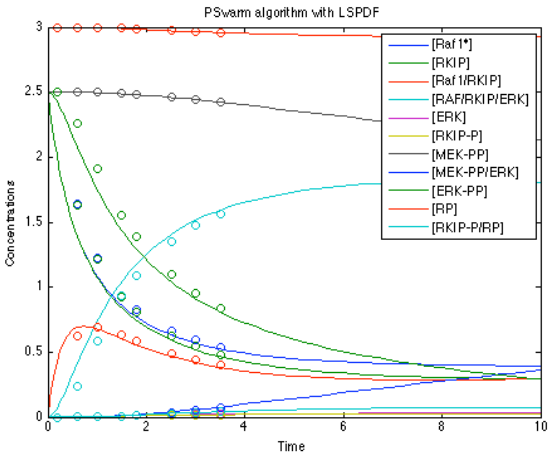
UNIVERSITY of GLASGOW

Xu Gu

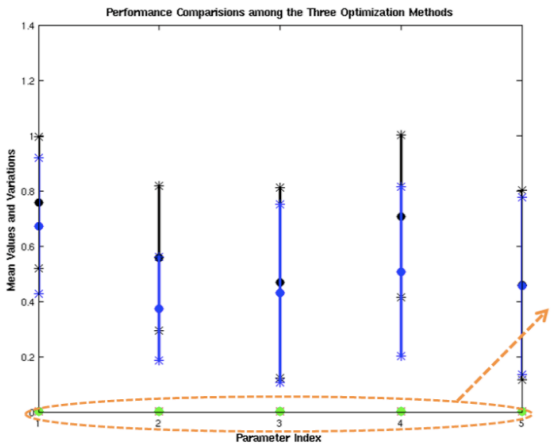
PSwarm+

- Apply root-finding method to constrain & fragment initial search space
- Multiple inititial states (fragments) → multiple solutions
- Computationally efficient

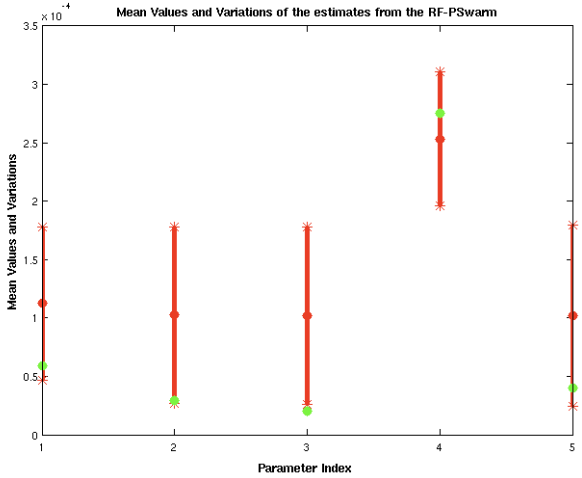
RKIP - signalling



Isomerisation of a-pinene metabolic pathway



GA, PSwarm



RF-PSwarm

UNIVERSITY of GLASGOW

Systems Biology Markup Language

- Machine-readable format for representing computational models in SB
  - Expressed in XML using an XML Schema
  - Intended for software tools—not for humans

- Tool-neutral exchange language for software applications in SB
  - Simply an enabling technology
- Used quite widely in biological modelling
- It is supported by over 40 software systems including Gepasi
- Good documentation, user community and publicly available tools
- www.sbml.org
- Also www.ebi.ac.uk/biomodels

SBML Example Reaction

```
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1">
  <model id="newModel">
    <listOfCompartments>
      <compartment id="compartment" size="1"/>
    </listOfCompartments>
    <listOfSpecies>
      <species id="A" compartment="compartment" initialConcentration="5"/>
      <species id="B" compartment="compartment" initialConcentration="1"/>
    </listOfSpecies>
    <listOfParameters>
      <parameter id="K1" value="1"/>
    </listOfParameters>
    <listOfReactions>
      <reaction id="Ak1B" reversible="false">
        <listOfReactants>
          <speciesReference species="A"/>
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="B"/>
        </listOfProducts>
        <kineticLaw>
          <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
              <times/>
              <ci> K1 </ci>
              <ci> A </ci>
            </apply>
          </math>
        </kineticLaw>
      </reaction>
    </listOfReactions>
  </model>
</sbml>
```

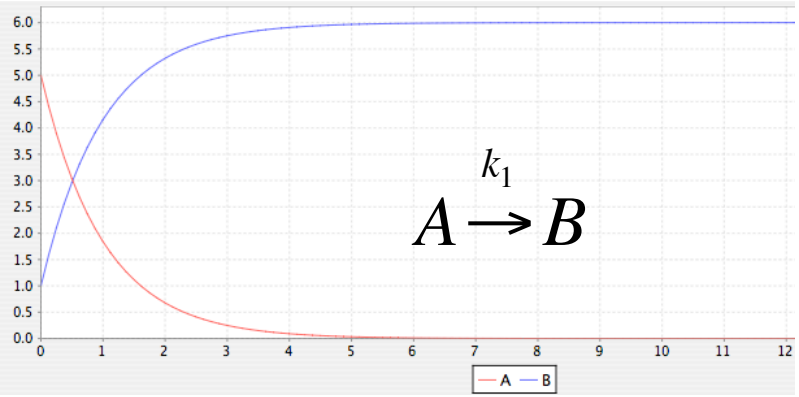
$$A \xrightarrow{k_1} B$$

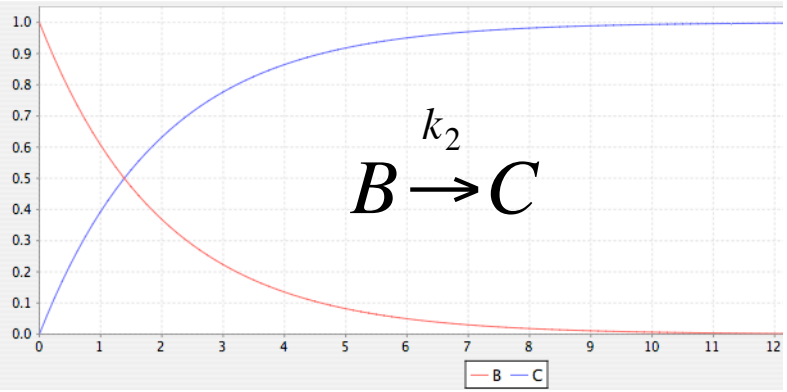
UNIVERSITY of GLASGOW

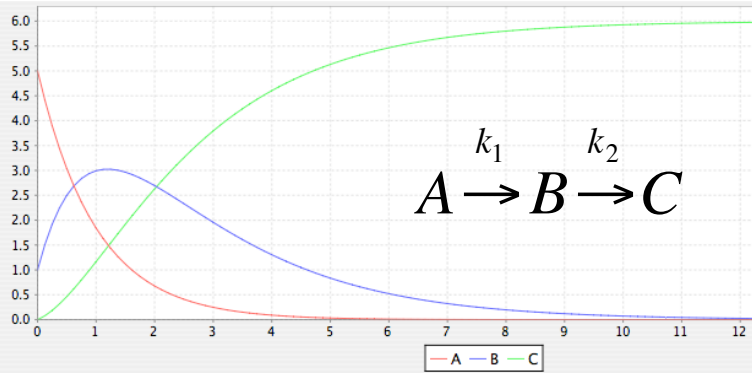
Martin Goodfellow

Composition of SBML models

- Fusion: Merge N models into 1 model (lose sub-model identities)
- Hierarchical composition (collection of sub-models - SBML3?)
  - Aggregation: defined interfaces to models
  - Computation via parallel execution?



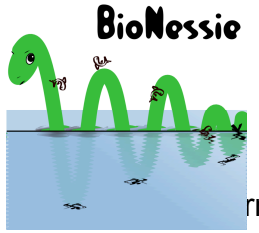




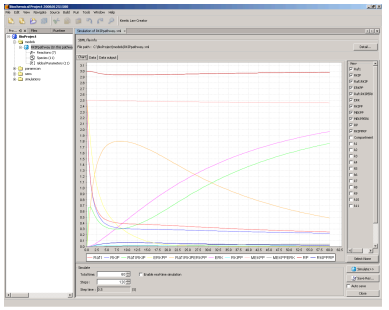
UNIVERSITY of GLASGOW

Related Efforts

- Some similarity to CellML (**`www.cellml.org`**)
  - SBML is somewhat closer to rep. used in simulators
  - CellML is somewhat more abstract and broader
  - Both SBML and CellML teams are working together
    - Committed to bringing them closer together
    - SBML Level 2 adopted features from CellML
- BioPAX (**`www.biopax.org`**)
  - A common exchange format for databases of pathways
  - SBML & BioPAX are complementary, not competing
  - SBML and BioPAX teams working together to define linkages between SBML and BioPAX representations

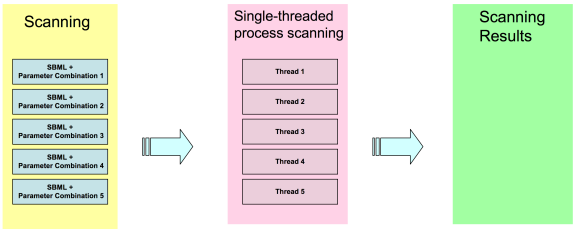


BioNessie ODE workbench



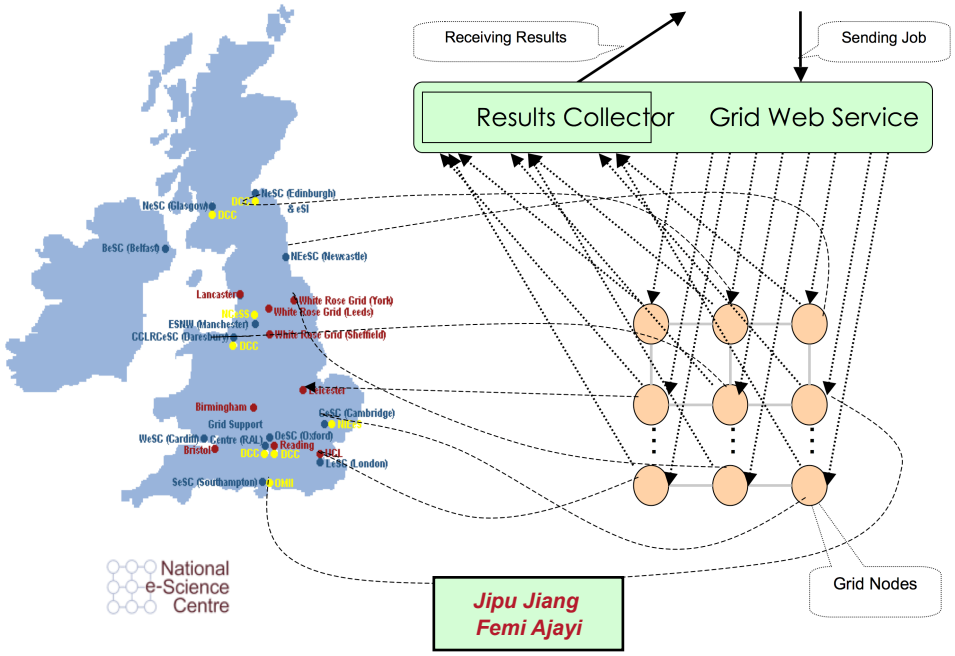
- rm independent
  - Windows, Linux (i386 or AMD64) and Mac Os with Intel i386.
  - Released on 5th October 2006 for internal use.
  - JAVA Web Start

Multithreaded Parameter Scan



- Simulation
  - Multithreaded: simulation of different models at the same time.
  - User-friendly data viewer and printable data output
- SBML model construction
  - Graphical tool supports creation & editing of SBML biochemical models
  - Kinetic Law creation and management

- Grid
- Multithreading
- Parameter Scanning
- Sensitivity Analysis
- Model Version Control System
- Model Development Management
- Optimisation
- Model checking



Xuan Liu

Jipu Jiang
Femi Ajayi

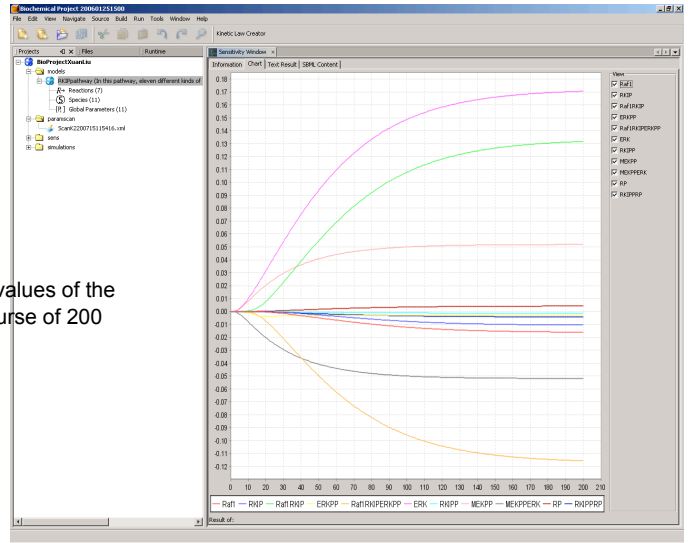
UNIVERSITY of GLASGOW

Sensitivity analysis

- Sensitivity analysis investigates the changes in the system outputs or behavior with respect to the parameter variations. It is a general technique for establishing the contribution of individual parameter values to the overall performance of a complex system.

- Sensitivity analysis is an important tool in the studies of the dependence of a system on external parameters, and sensitivity considerations often play an important role in the design of control systems.

- Parameter sensitivity analysis can also be utilised to validate a model's response and iteratively, to design experiments that support the estimation of parameters

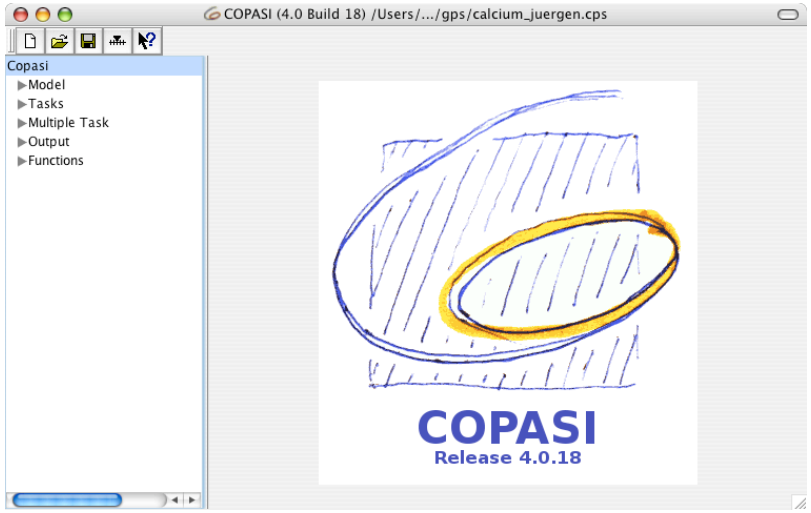


Sensitivity of species to the values of the parameter K6 for the timecourse of 200 timesteps of 200 time units.

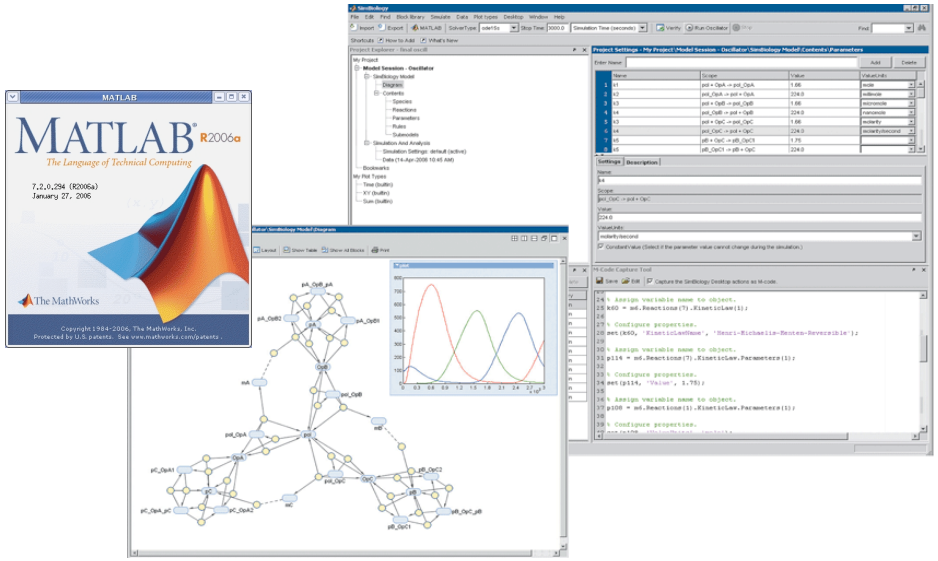
Slide from Xuan Liu

UNIVERSITY of GLASGOW

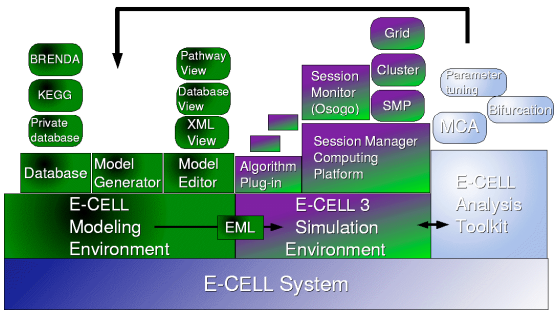
Other simulators include…



Systems Biology Workbench



E-CELL Development Overview



BRC

UNIVERSITY of GLASGOW

Acknowledgements

Glasgow:

- Robin Donaldson
- Xu Gu
- Xuan Liu
- Richard Orton
- Muffy Calder
- Vladislav Vyshemirsky
- Gary Gray
- Jipu Jiang
- Femi Ajayi
- iGEM Team

Cottbus:

- Monika Heiner
- Sebastian Lehrack

UL Brussels

- Jacques van Helden

MRC Cambridge

- Lorenz Wernisch

U Toronto

- Shoshana Wodak