Computer-Aided Systems Biology

David Šafránek

sybila
systems biology laboratory

Masaryk University
Czech Republic

Outline

Outline

Biology

- since ancient times
- empirically studies **life** and **living organisms**
- studied aspects: structure, **function**, growth, development and evolution
- used concepts:
  - the **cell** – the unit of life
  - the **gene** – the unit of inherited **information**
  - the **evolution** – the mechanism of species creation

Biophysics

- since the mid of 19th century
- living organism = open (thermodynamic) **system**
- the goal: **why** and **how** the living matter works?
- uses mathematical apparatus
- a fascinating phenomenon: homeostasis
  - maintain a **stable condition** in a **changing environment**
  - **robust** (up-to certain limits)

reductionism

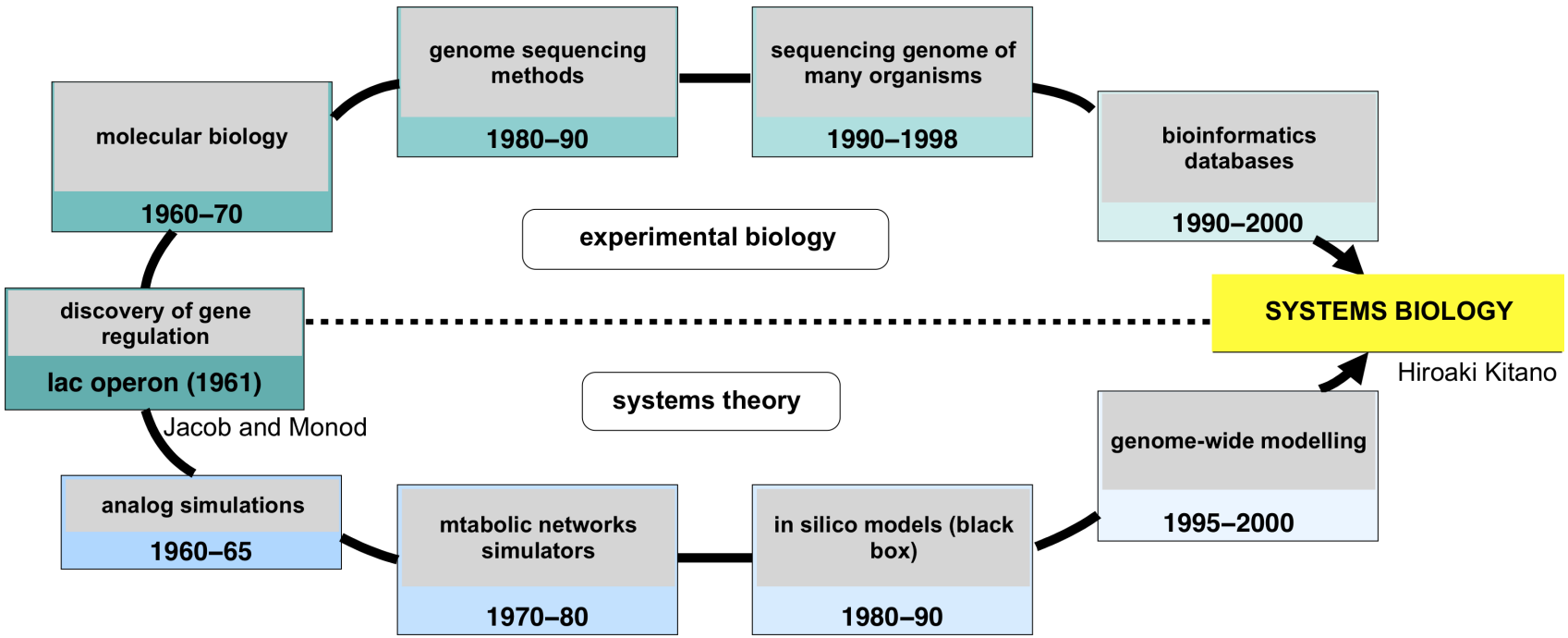
?

holism

"A Whole is Greater Than the Sum of Its Parts."

– Aristoteles

Development of Systems Biology



"Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?"

Philip Merilees

Motivation: Rigorously Answer Biological Questions

- biology is goal-oriented
- biological problems typically address complex processes

Examples of biological problems

How the bacteria cell **utilises** particular nutritions?

Which nutritions imply fastest **growth** under given conditions?

Motivation: Rigorously Answer Biological Questions

- biology is goal-oriented
- biological problems typically address complex processes

Examples of biological problems

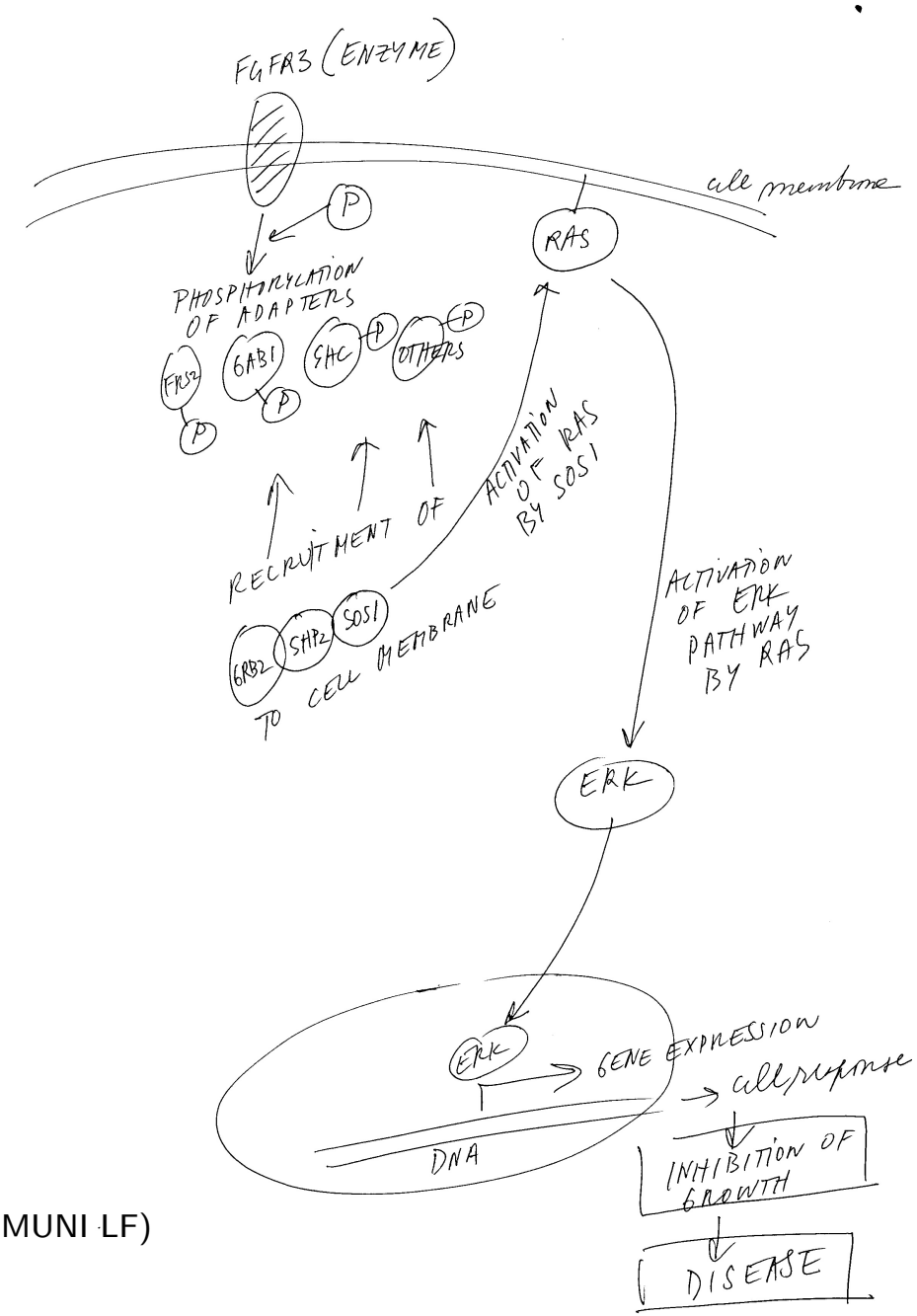
How the bacteria cell **utilises** particular nutritions?

Which nutritions imply fastest **growth** under given conditions?

The answer should fullfil specific requirements

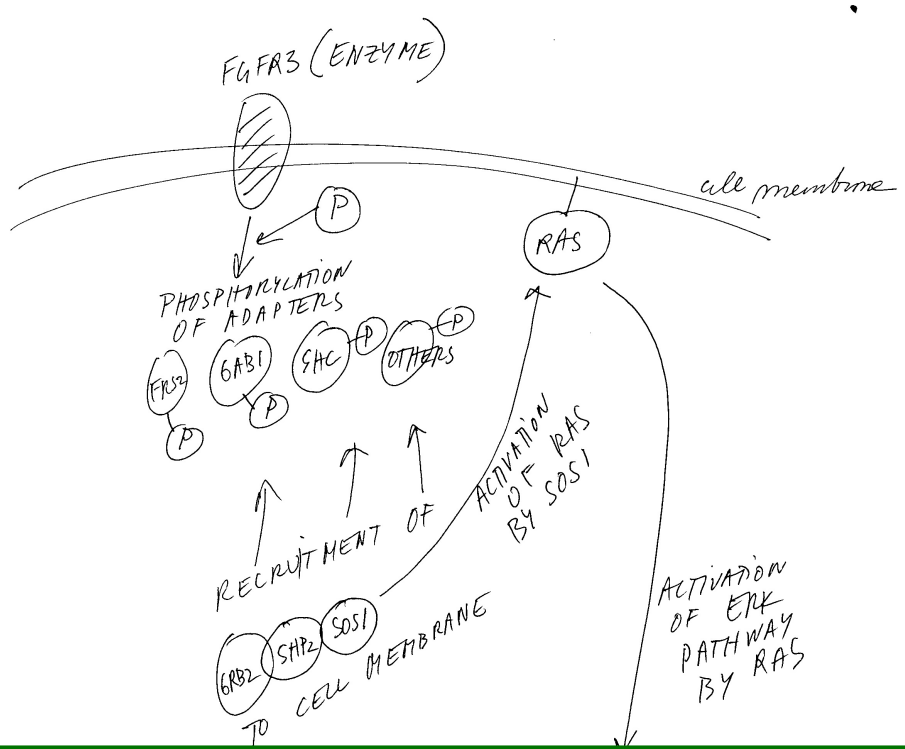
- to formulate and analyse a biological problem **holistically**
- to give mechanistic explanation based on known facts – mechanistic means in the context of laws of physics/chemistry
- to project the mechanistic details onto the genetic information

From Biologist's Table



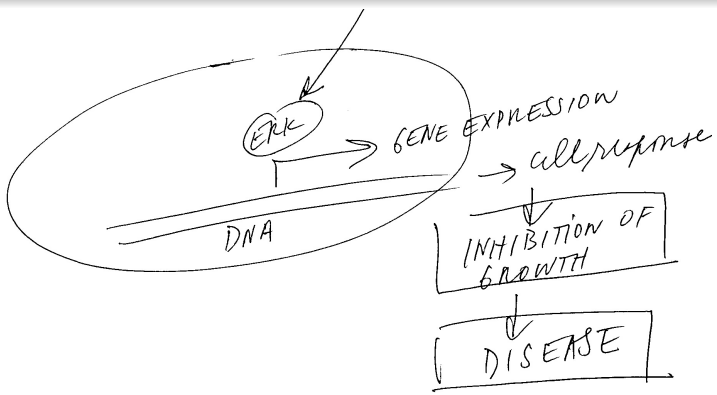
slide credits: Pavel Krejčí (MUNI LF)

From Biologist's Table



Biological Problem

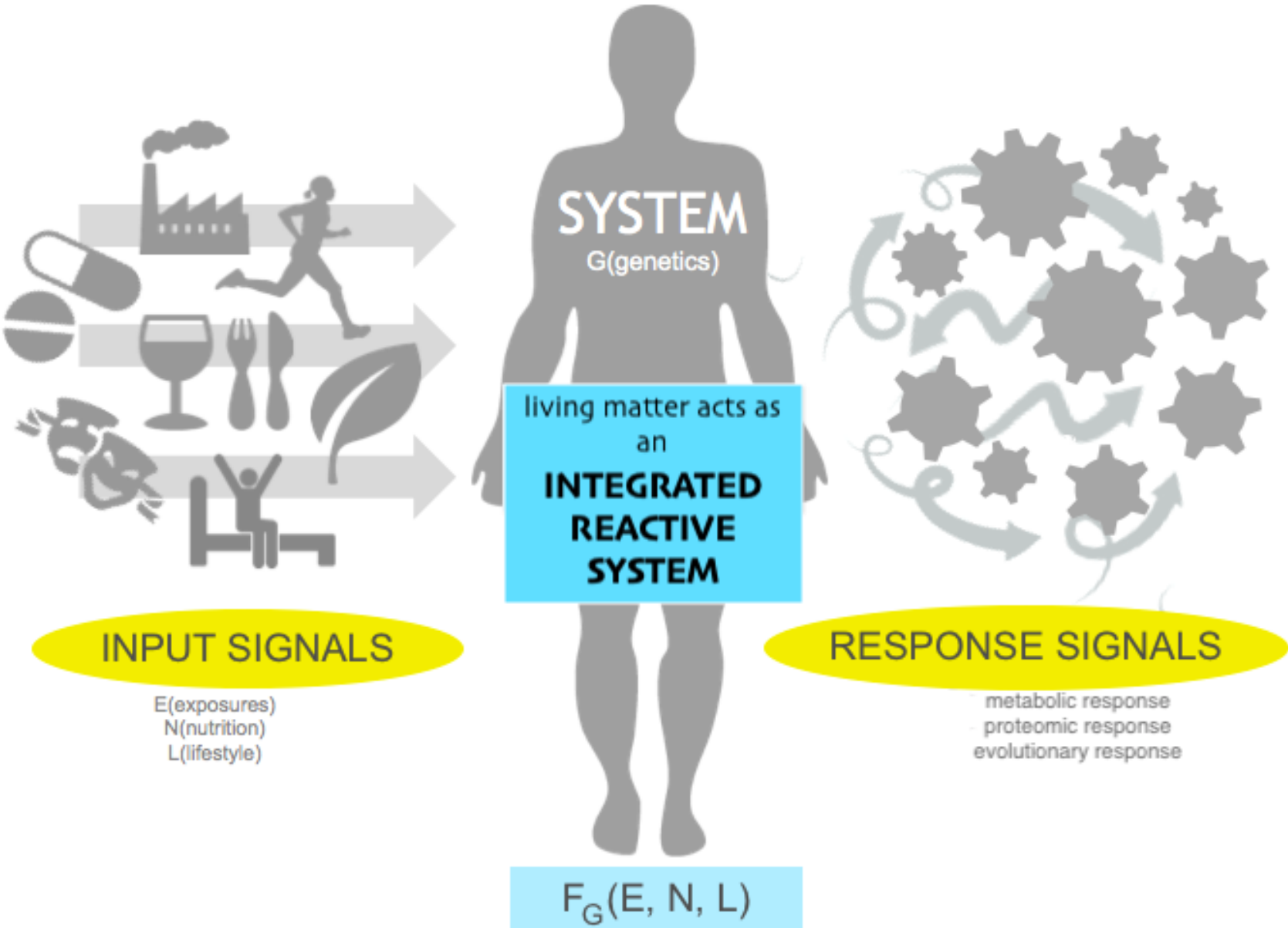
Why a human fibroblast cell misinterprets a certain growth factor?



slide credits: Pavel Krejčí (MUNI LF)

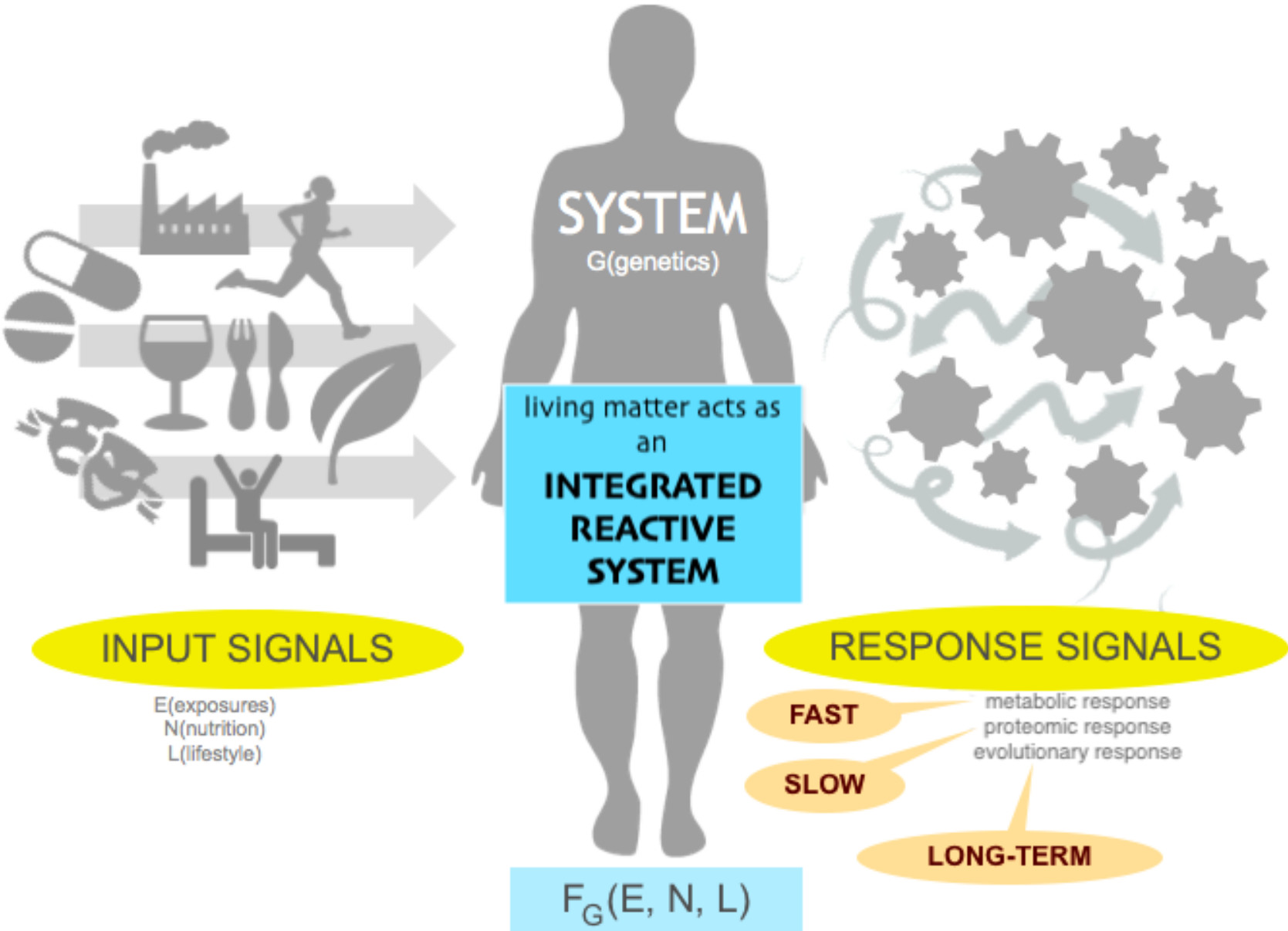
Systems Approach: The Grand Challenge

Complex Organism as a System



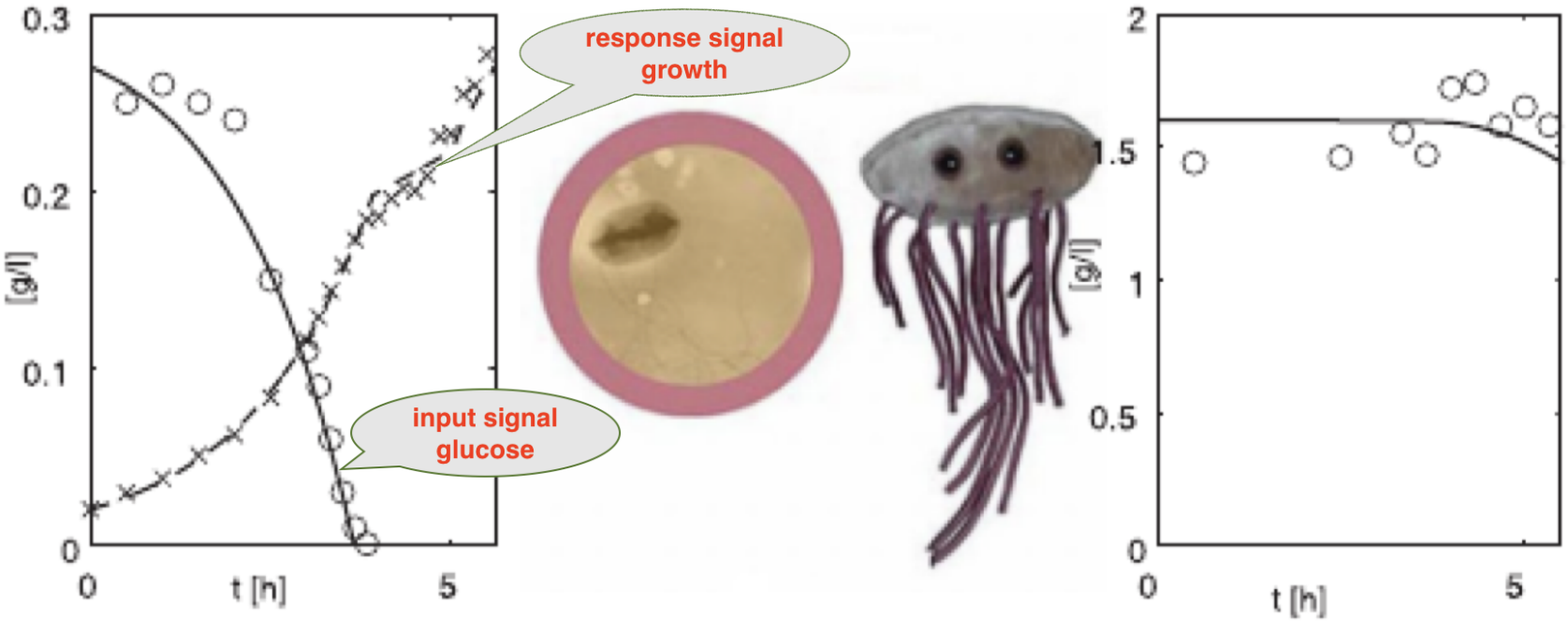
Systems Approach: The Grand Challenge

Complex Organism as a System



Systems Approach: A Moderate Challenge

Population of Bacteria as a System

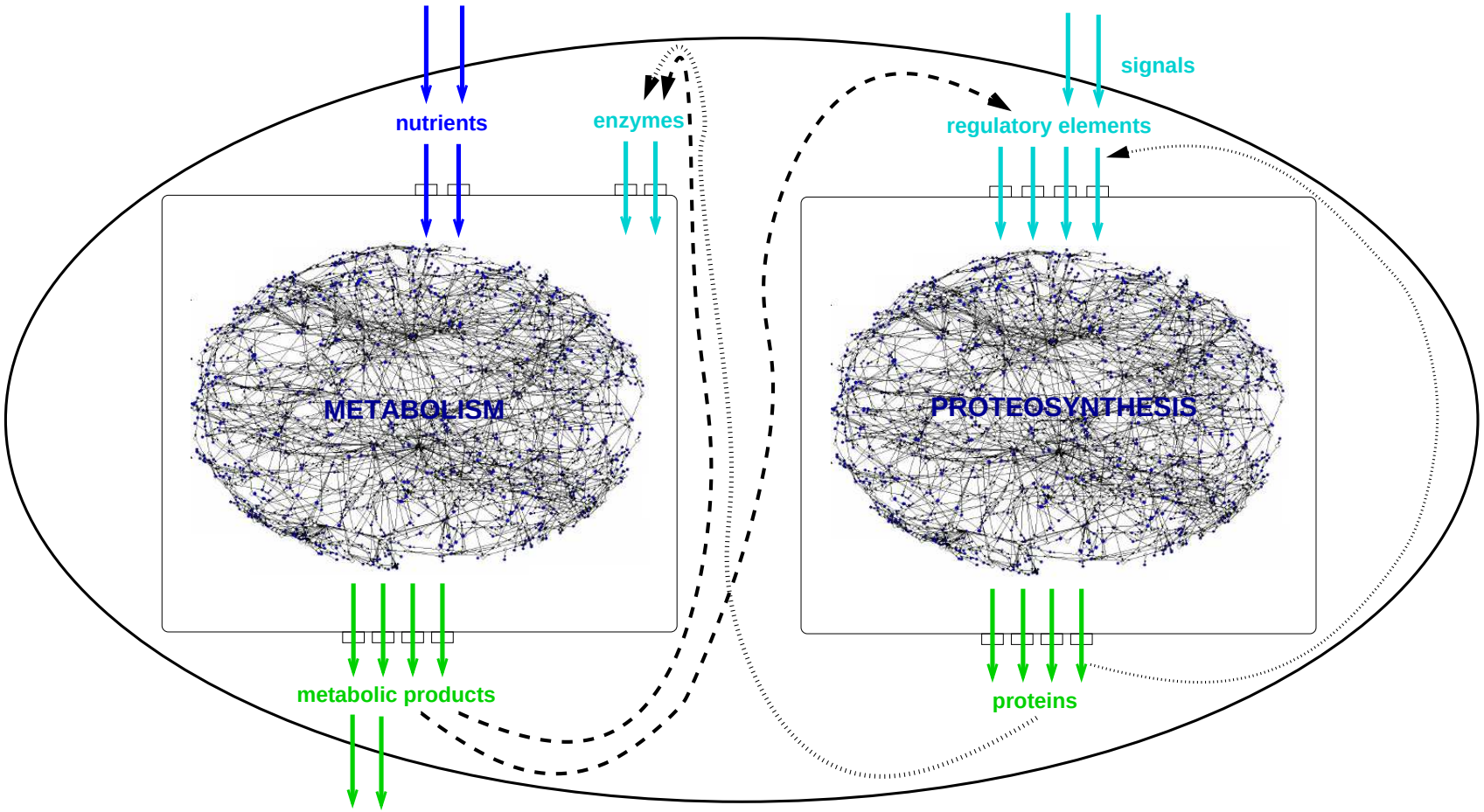


for a particular set of genes G

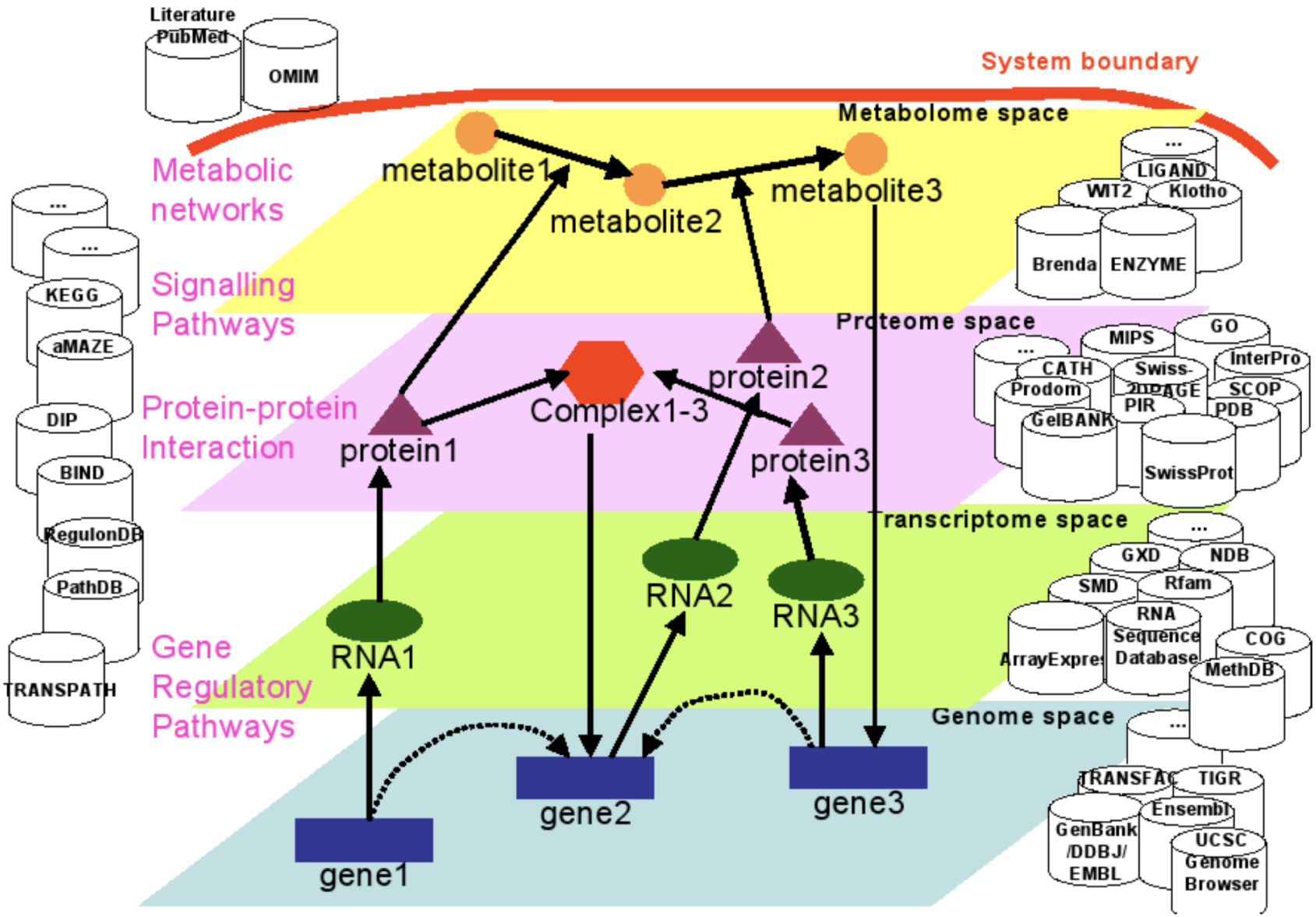
F_G : (environment exposure, nutrition) $\rightarrow$ growth profile

slide credits: Ralf Steuer (HU Berlin)

Systems View of Processes Driving the Cell



The Cell as a Complex Interaction Network



slide credits: David Gilbert (Brunel Univ.)

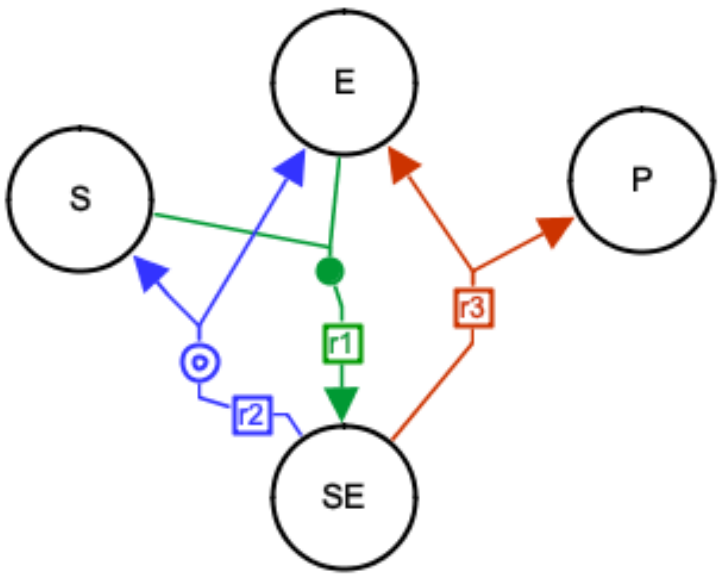
Outline

The Paradigm of Computational Systems Biology

Assumptions

- The biological reality (a biophysical process) is understood as a **biological system**.
- A biological system is given as a **network** $\mathcal{N}$ of biochemical **components** connected by chemical/physical **interactions**.
- The components include relevant genes and gene products.

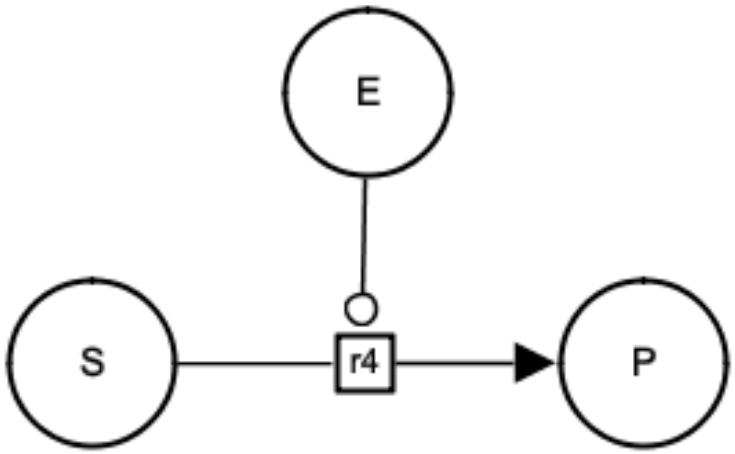
- basic form: **chemical reaction networks (CRNs)**
  - elementary chemical reactions
  - represent the flow of the mass
- example:



$$S + E \underset{r_2}{\overset{r_1}{\rightleftarrows}} SE \xrightarrow{r_3} P + E$$

SBGN standard notation, see https://www.sbgn.org

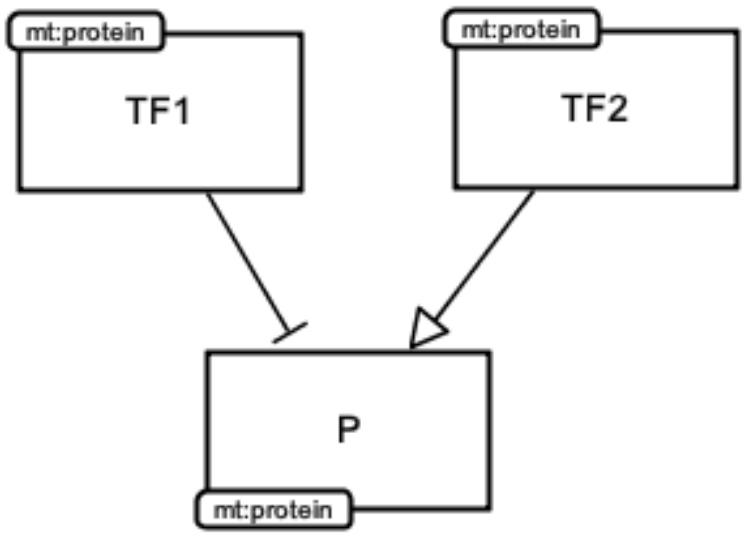
- simplified form: **reaction-influence networks (RINs)**
  - chemical reactions influenced by other molecules
  - represent the modulated flow of the mass
- example:



$$S \xrightarrow{r_4} P; E$$

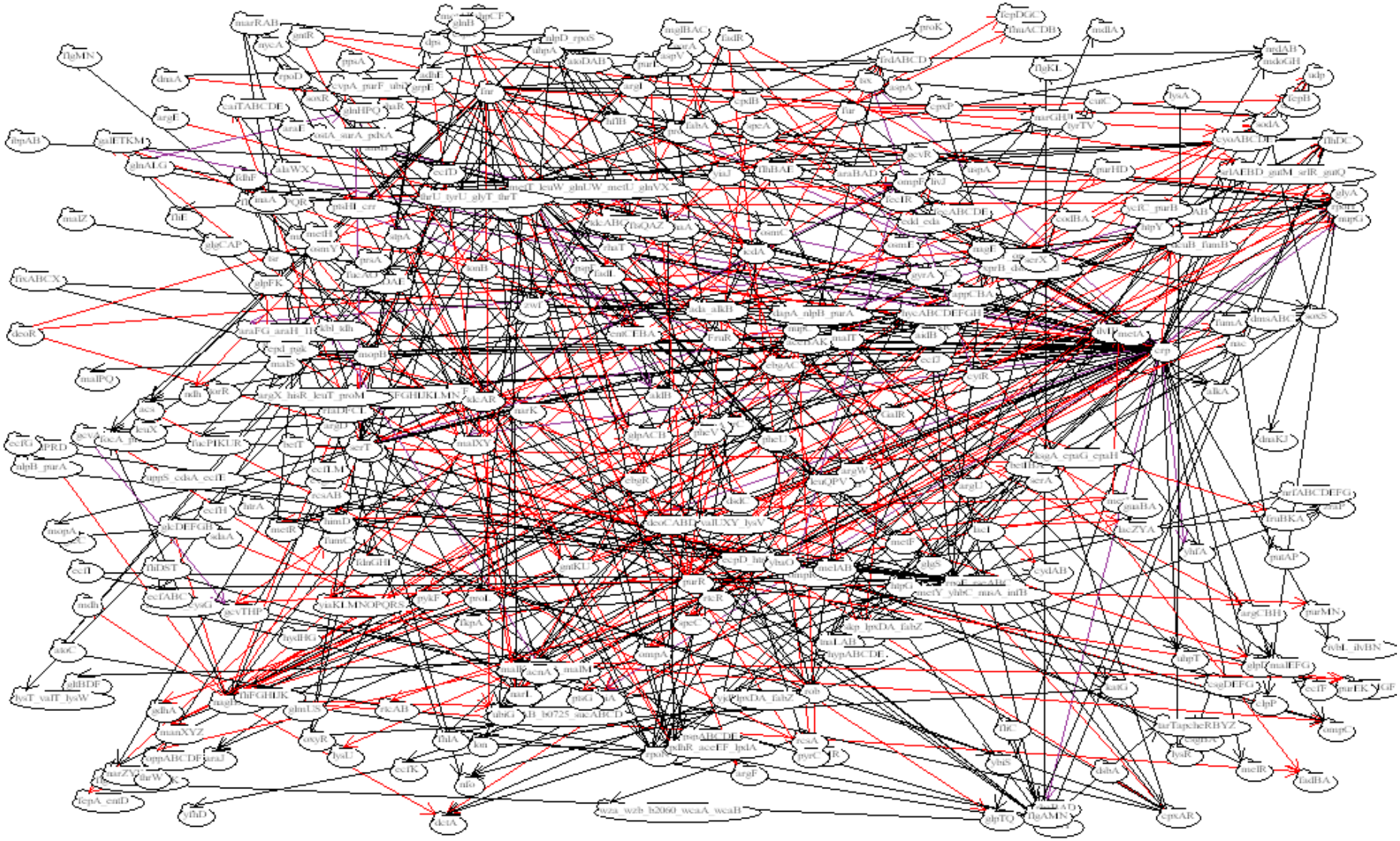
SBGN standard notation, see `https://www.sbgn.org`

- abstract form: **influence networks (INs)**
  - represent positive/negative influences among molecules
  - well fit incomplete knowledge
  - typically gene regulatory networks, signalling pathways
- example:



SBGN standard notation, see `https://www.sbgn.org`

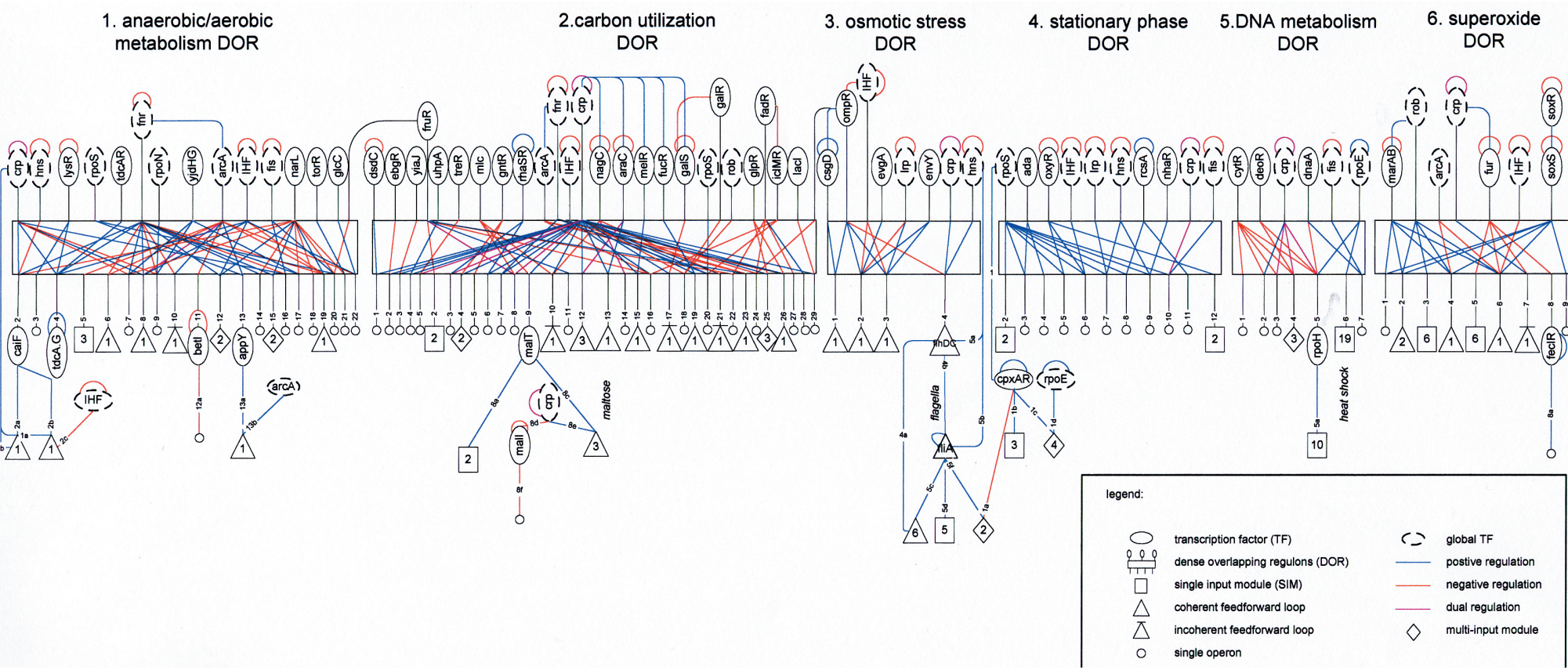
Biological Networks are Large



genetic regulatory network of *E. coli*

see `https://reactome.org` for more...

...But Organised



genetic regulatory network of *E. coli*

slide credits: Uri Alon

The Goal of Computational Systems Biology

The General Goal

For a biological system given by a network $\mathcal{N}$ **reconstruct** the system's **dynamics**:

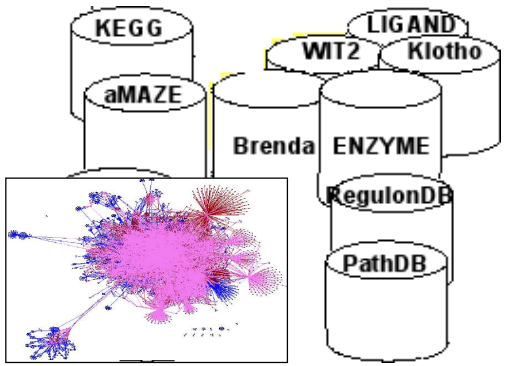
Define a function that encodes the information (signal) processing occuring in **all** *components of the system in time.*

$$F_{\mathcal{N}} : (\text{input stimuli}, \text{environment signals}) \rightarrow \text{response signals}$$

Model-Based Workflow

network reconstruction

biological knowledge databases

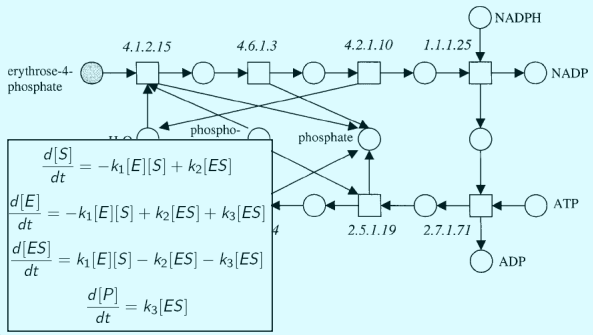


biological network

hypothesis

model specification

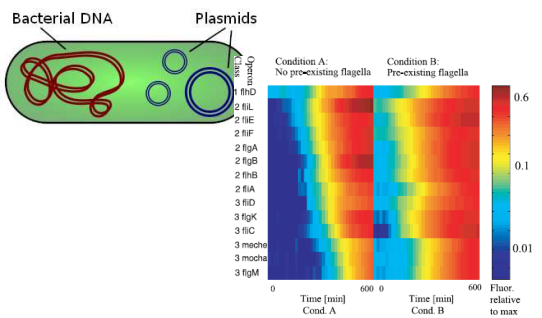
Petri Net, ODEs, rule-based, process calculus, Boolean network, …



computer lab

model validation

gene reporters, DNA microarray, mass spectrometry, ...

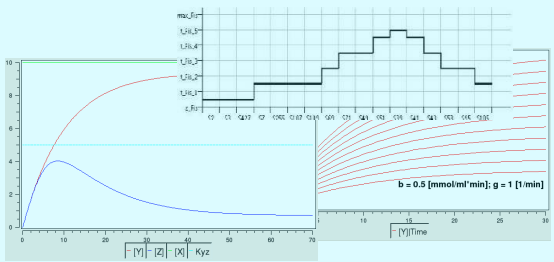


emergent properties

model questions

model analysis

static analysis, numerical simulation, analytical methods, model checking



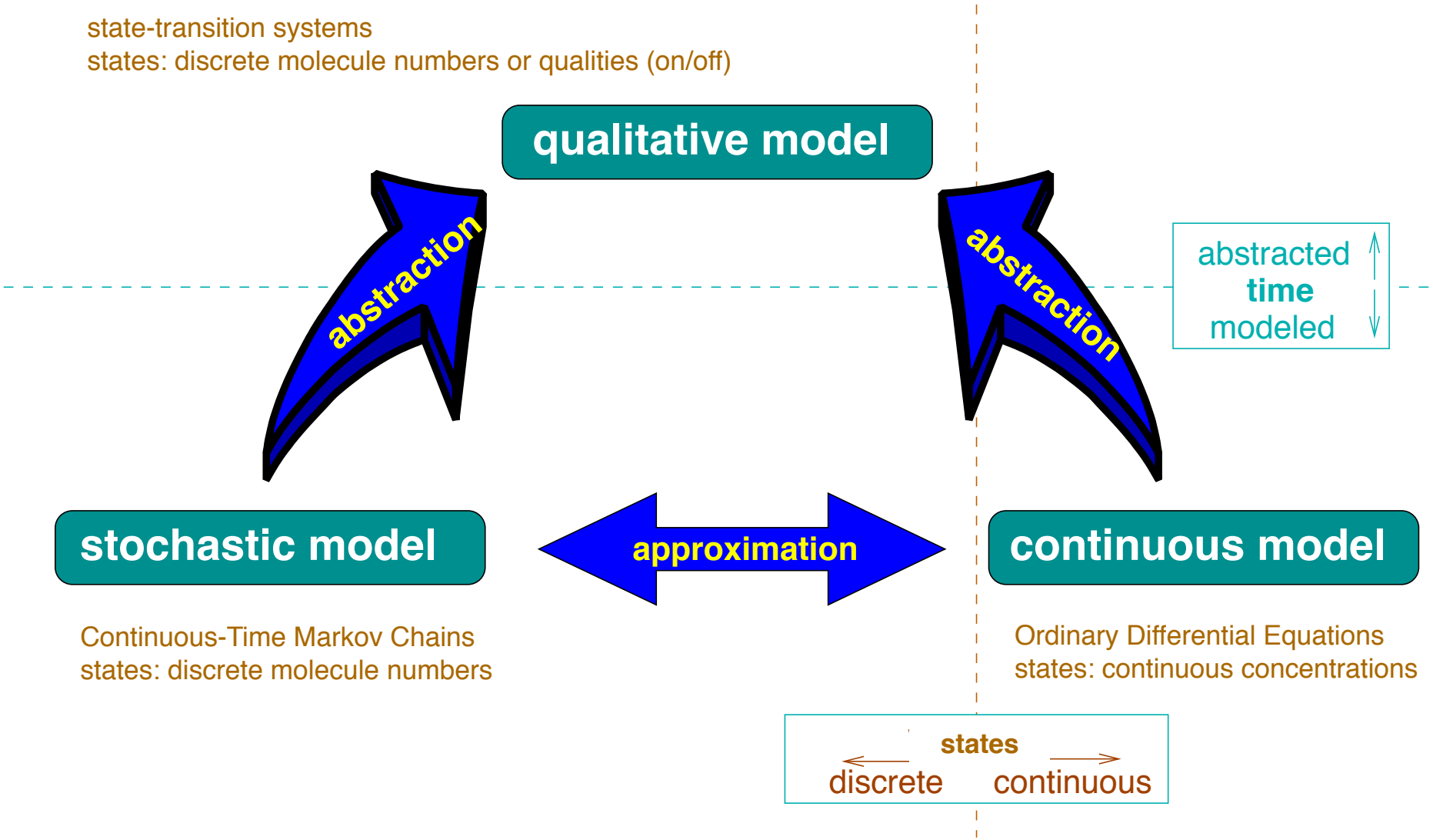
hypothesis testing, property detection, new hypothesis inference

From Networks to Models

How to reconstruct the dynamics $F_{\mathcal{N}}$ of a network $\mathcal{N}$?

1. choose a modelling framework
2. associate every interaction in $\mathcal{N}$ with a suitable **kinetic rule**
   - describes how a **state** of affected components changes in time
3. build a computational **model** by combining kinetic rules

Modelling Frameworks



deterministic continuous-time dynamics of molecule population
molecules dissolved in the cell volume (a well-stirred "pool")

molar concentration $[M]$=$[mol \cdot l^{-1}]$

$$A + B \longrightarrow AB$$

biophysical law of **mass action kinetics**:

$$v = k[A][B] \qquad [M \cdot s^{-1}]$$

deterministic continuous-time dynamics of molecule population
molecules dissolved in the cell volume (a well-stirred "pool")

molar concentration $[M]$=$[mol \cdot l^{-1}]$

$$A + B \longrightarrow AB$$

biophysical law of **mass action kinetics**:

$$v = k[A][B] \qquad [M \cdot s^{-1}]$$

$$\begin{aligned} \frac{d[A]}{dt} &= -v \\ \frac{d[B]}{dt} &= -v \\ \frac{d[AB]}{dt} &= v \end{aligned}$$

k $[s^{-1}]$ is a reaction-specific **parameter**

Waage, P.; Guldberg, C.M. (1864). Studier over Affiniteten. Forhandlinger i Videnskabs-selskabet i Christiania (Transactions of the Scientific Society in Christiania) (in Danish): 35–45.

stochastic continuous-time dynamics of molecule population
molecules distributed uniformly in the cell volume (well-stirred)

states describe molecules number (#)

$$A + B \rightarrow AB$$

biophysical law of **stochastic mass action** describes an expected rate of reaction event occurence:

$$v = k \cdot \#A \cdot \#B \qquad [s^{-1}]$$

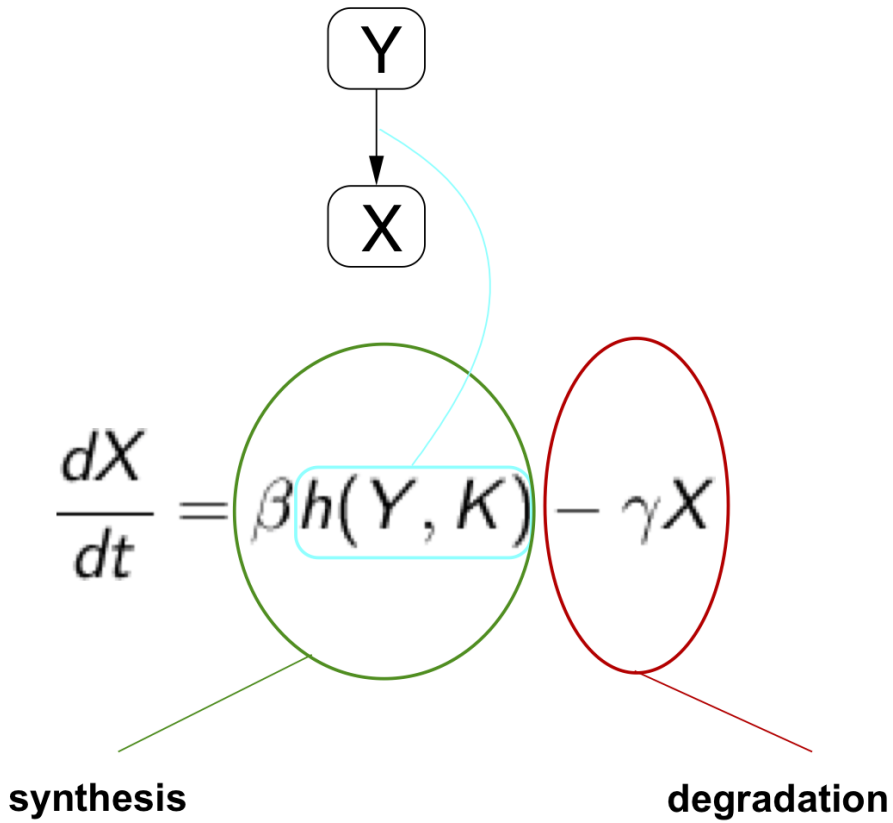
$$\begin{pmatrix} \#A = 5 \\ \#B = 2 \\ \#AB = 0 \end{pmatrix} \xrightarrow{v} \begin{pmatrix} \#A = 4 \\ \#B = 1 \\ \#AB = 1 \end{pmatrix}$$

k $[s^{-1}]$ is a reaction-specific **parameter**, depends on cell volume

Gillespie, Daniel T. (1977). "Exact Stochastic Simulation of Coupled Chemical Reactions". The Journal of Physical Chemistry. 81 (25): 2340–2361.

controlled synthesis of X and its pontaneous degradation

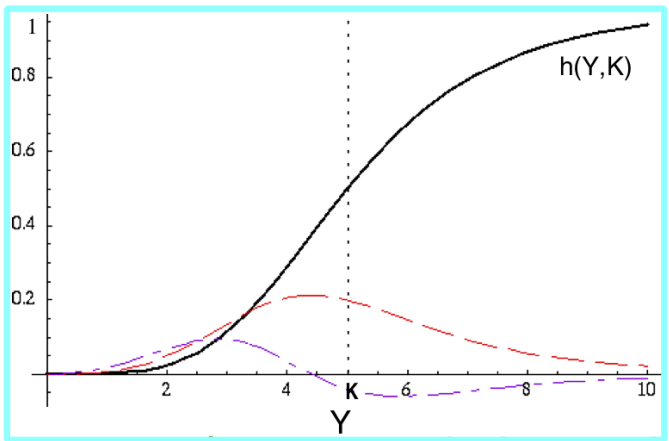
so-called (sigmoidal) Hill kinetics



β, γ ... production and degradation parameters;

controlled synthesis of X and its pontaneous degradation

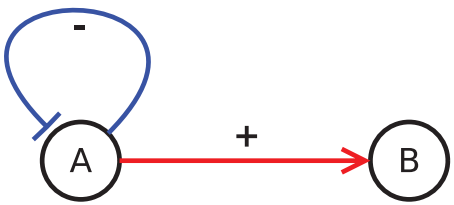
so-called (sigmoidal) Hill kinetics



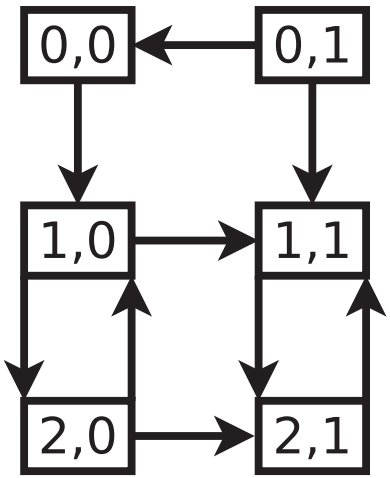
$$\frac{dX}{dt} = \beta h(Y, K) - \gamma X$$

synthesis degradation

β, γ ... production and degradation parameters;
K ... parameter of the influence



$A \in \{0, 1, 2\},\ B \in \{0, 1\}$
$t_{AA} = 2,\ t_{AB} = 1$
$K_{A,\emptyset} = 2$
$K_{A,\{A\}} = 0$
$K_{B,\emptyset} = 0$
$K_{B,\{A\}} = 1$



- introduced by René Thomas [1973]
- refined by Chaouiya et al. [2003]

- for a given network a model is built on first principles
- continuous models
  - **denotational** semantics defines the continuous flow

$$\dot{x} = f(x, p)$$

- stochastic and qualitative models
  - **operational** semantics defines events execution
  - quantitative or discrete parameters

- for a given network a model is built on first principles
- continuous models
  - **denotational** semantics defines the continuous flow

$$\dot{x} = f(x, p)$$

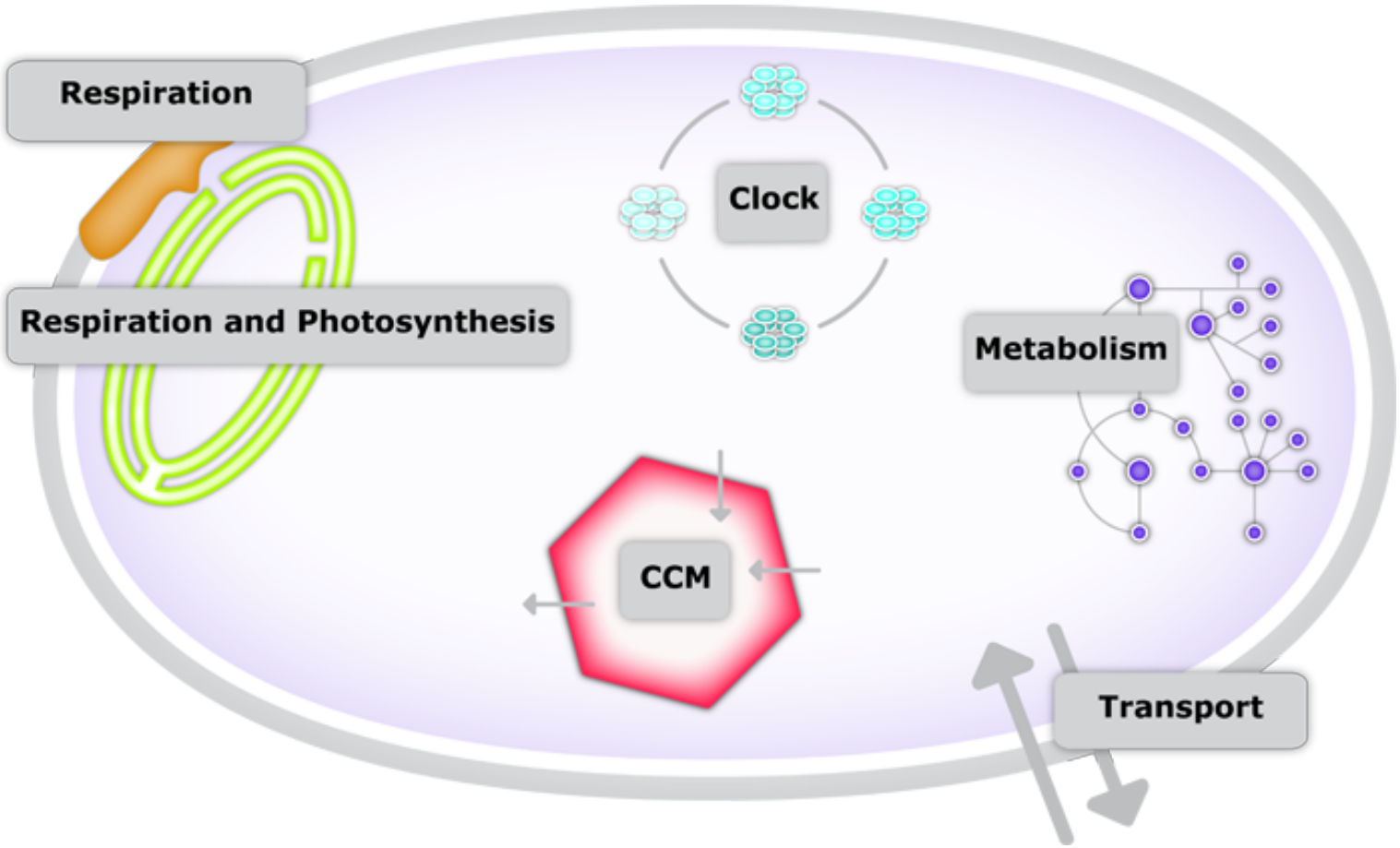
- stochastic and qualitative models
  - **operational** semantics defines events execution
  - quantitative or discrete parameters

Challenges (for Computer Science!)

How to deal with large, potentially infinite networks?

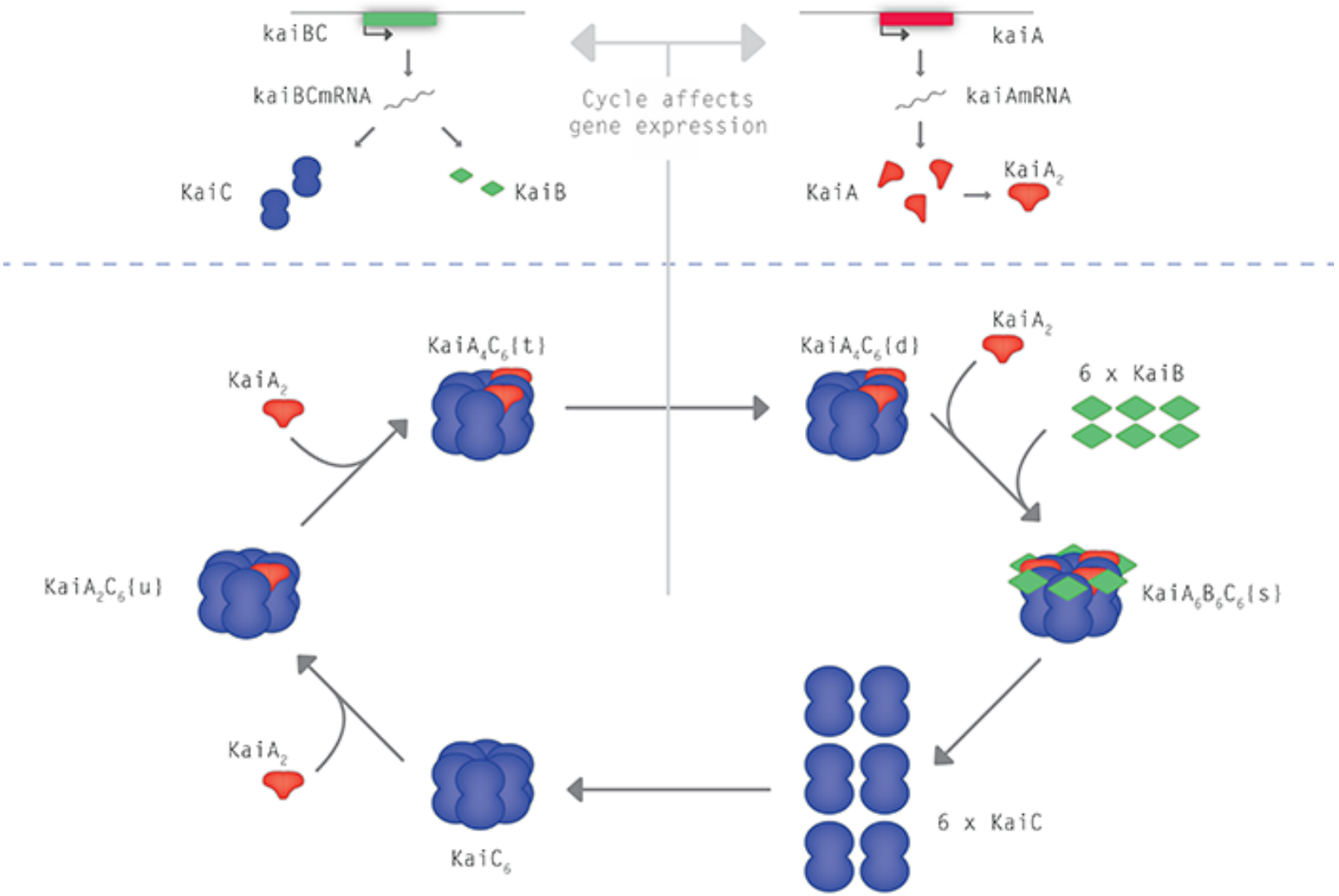
How to deal with unknown parameters?

Model Specification
Comprehensive Modelling Platform (CMP)



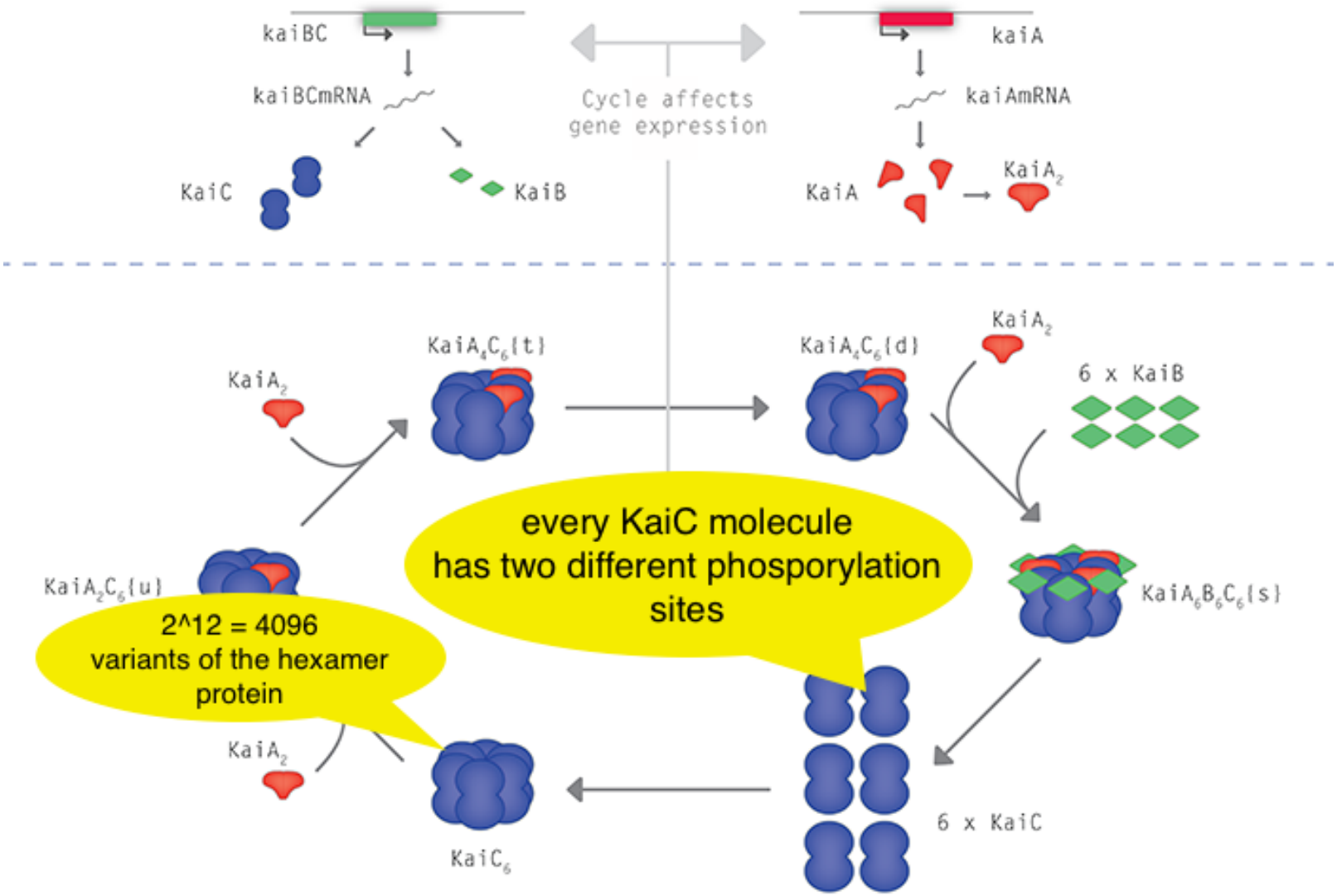
https://e-cyanobacterium.org

Model Specification
Example: Cyanobacteria Clock



https://e-cyanobacterium.org

Model Specification
Example: Cyanobacteria Clock



https://e-cyanobacterium.org

- rule-based: generalise reactions to rules
- components have states (e.g., phosphorylated sites)
- rules are executed in a solution (soup of entities) in a context-free manner (match $\rightarrow$ apply)
- all kind of models can be generated
- BCSL example:

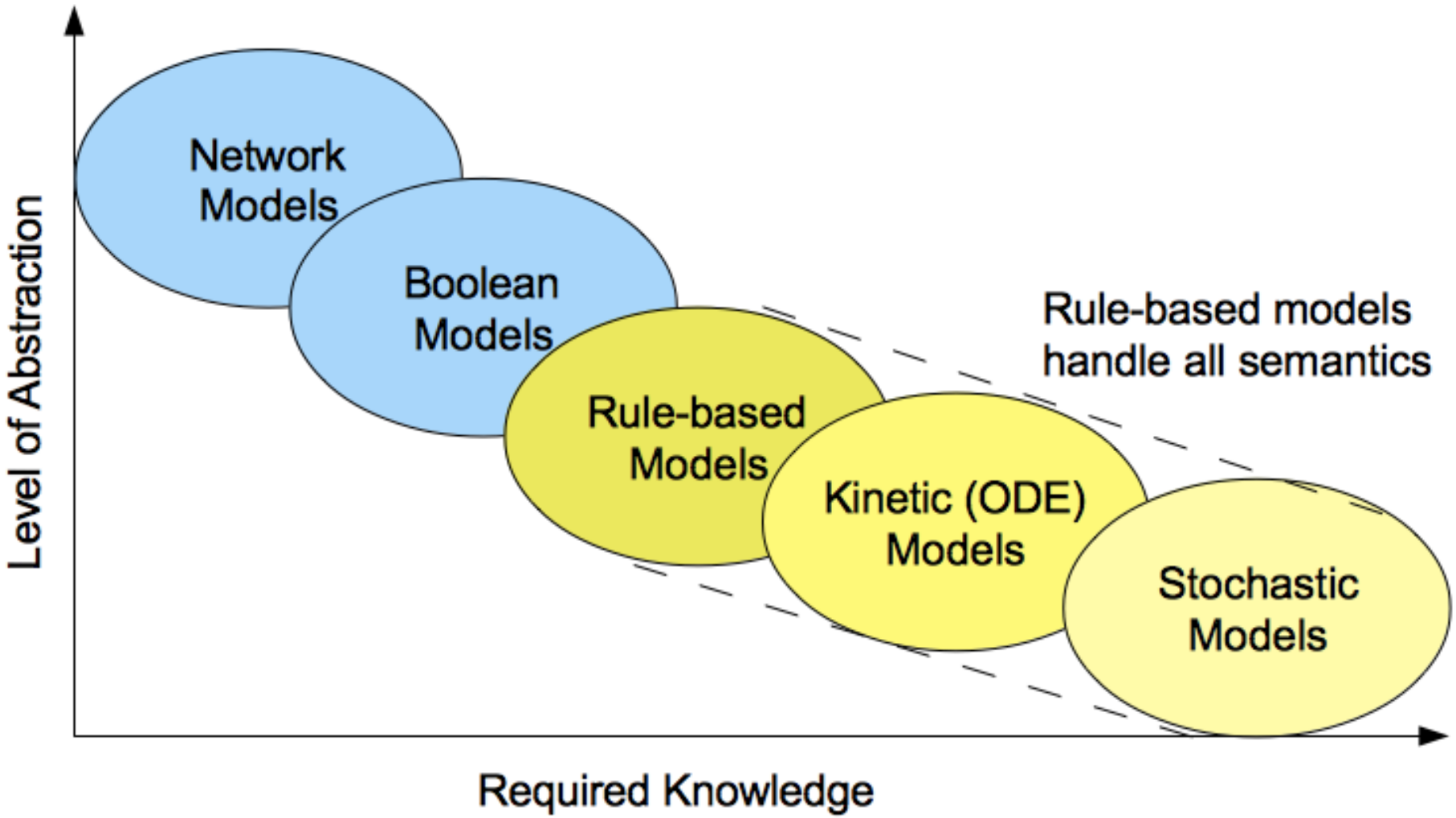
```
S{u}::KaiC::KaiC6::cyt <=> S{p}::KaiC::KaiC6::cyt
```

executes on any serine site of any KaiC in a hexamer

Other Languages

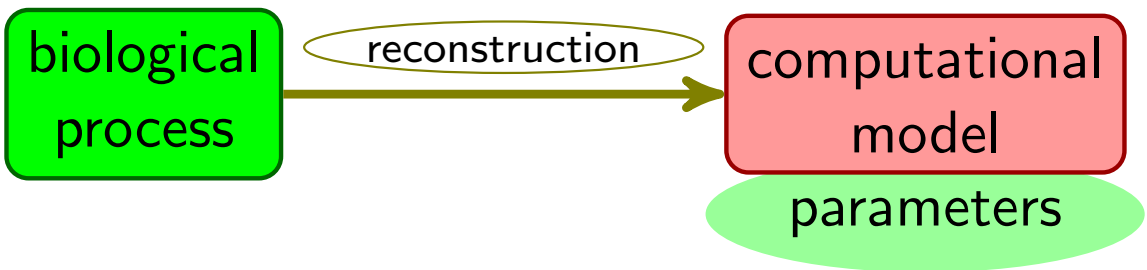
- Kappa, BNGL, Chromar, LBS, ... (rule-based)
- SPiM, BioSPI, ... (process-algebraic)

Required Knowledge



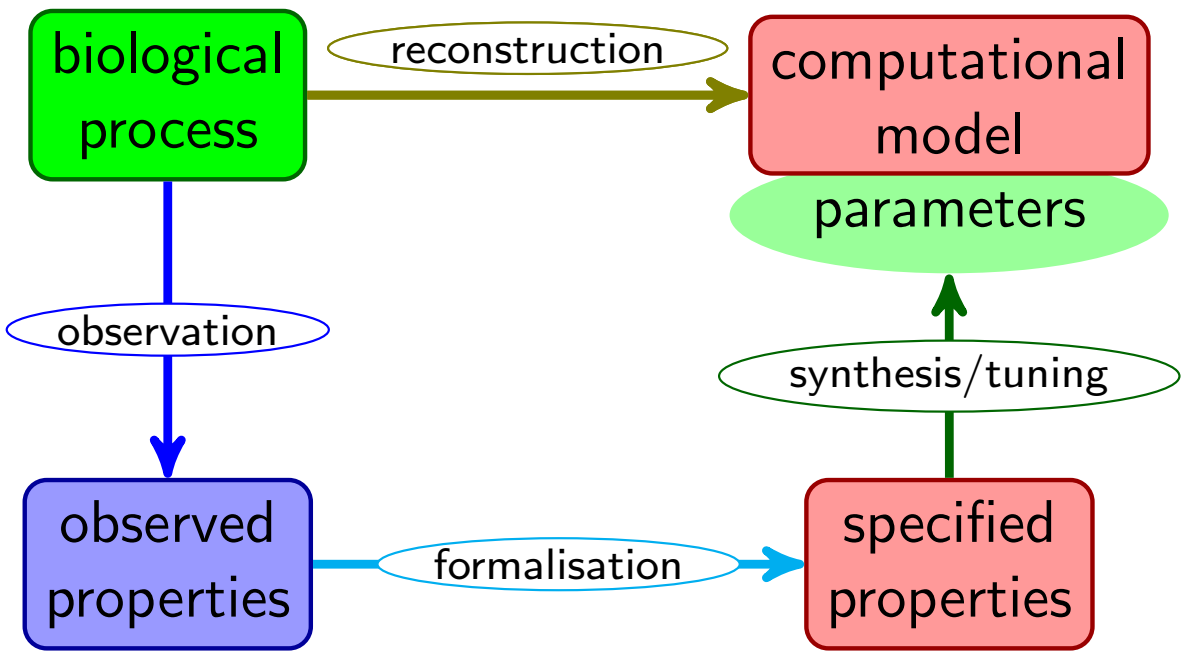
Model Construction
Dealing with Unknown Parameters

- traditional approaches – parameter estimation
  $\Rightarrow$ finding a single "optimal" value fitting experimental data



Model Construction
Dealing with Unknown Parameters

- traditional approaches – parameter estimation
  $\Rightarrow$ finding a single "optimal" value fitting experimental data
- computer science approach – **parameter synthesis**
  $\Rightarrow$ finding values satisfying given dynamical properties/hypotheses

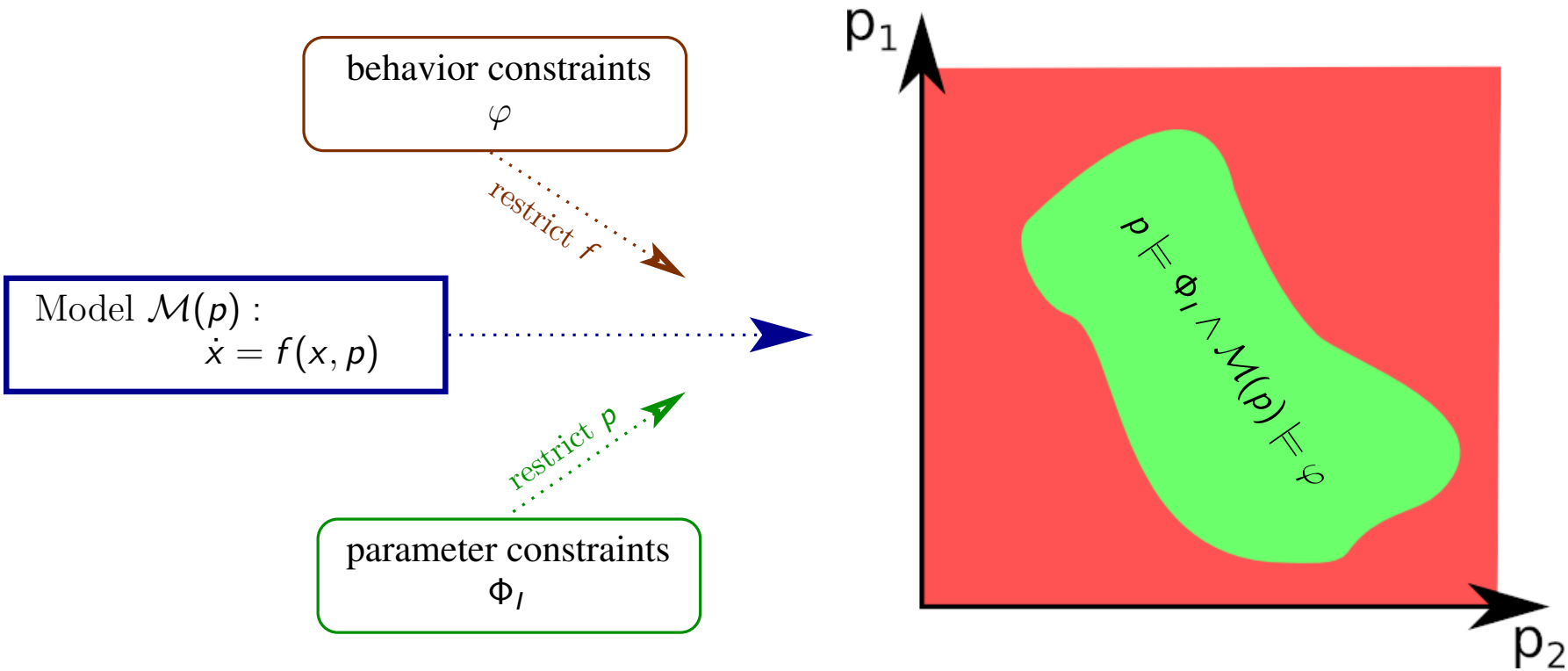


Temporal Logics for Biological Systems

- qualitative properties (LTL, CTL, HCTL)
  - modalities (possibilities/necessities in future behaviour)
  - reachability of particular (sets of) states
  - temporal ordering of events, monotonicity
  - temporal correlations of model variables
  - stability (attractors, basins of attraction)
- quantitative properties
  - deterministic (MTL, MITL, STL, STL*)
    - enhance modalities with (dense) time information
    - exact timing of events, time-bounds
    - value-freezing (HSB 2012)
  - stochastic (PLTL, PCTL, CSL)
    - probability of property satisfaction
    - stochasticity combined with time

Problem Formulation

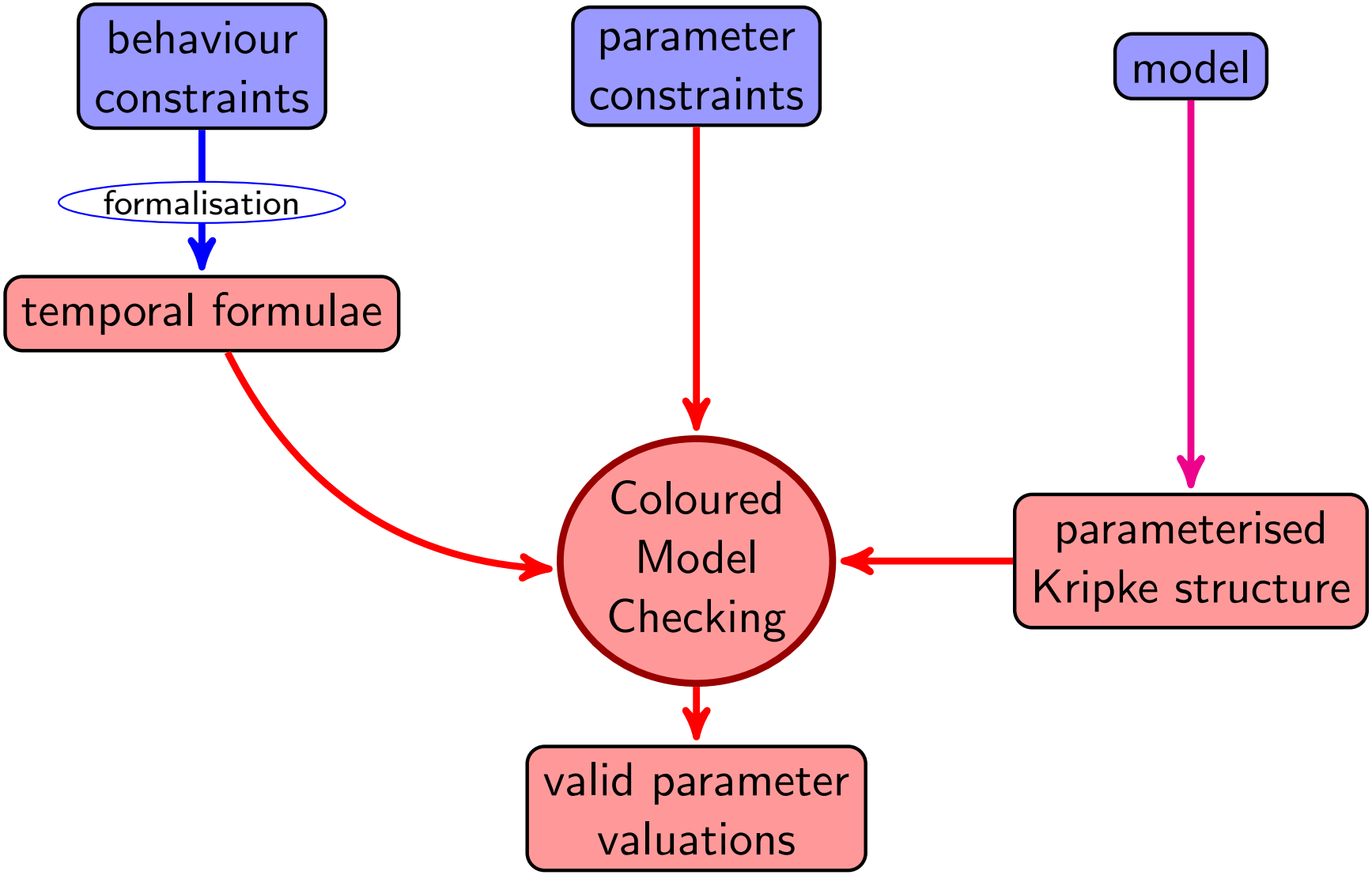
Parameter Synthesis for Dynamical Systems



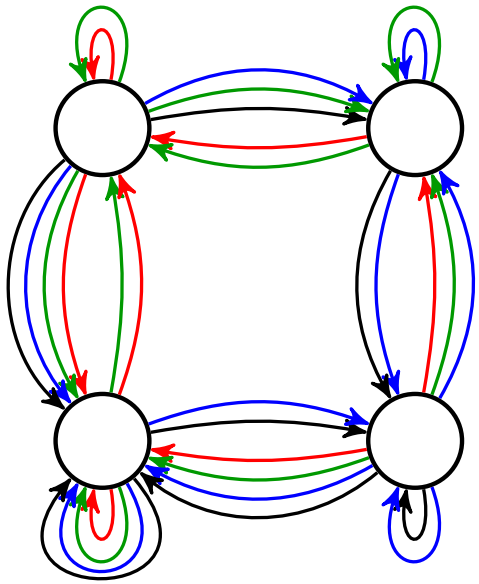
Parameter Synthesis Problem

Assume $\mathcal{P}$ is the admissible **parameter space**. Given a *behaviour constraint* φ, *parameter constraint* Φ_I, and a *parameterised model* $\mathcal{M}$, **find the maximal set** $P \subseteq \mathcal{P}$ **of parameterisations** such that $p \models \Phi_I$ and $\mathcal{M}(p) \models \varphi$ for all $p \in P$.

Parameter Synthesis Workflow

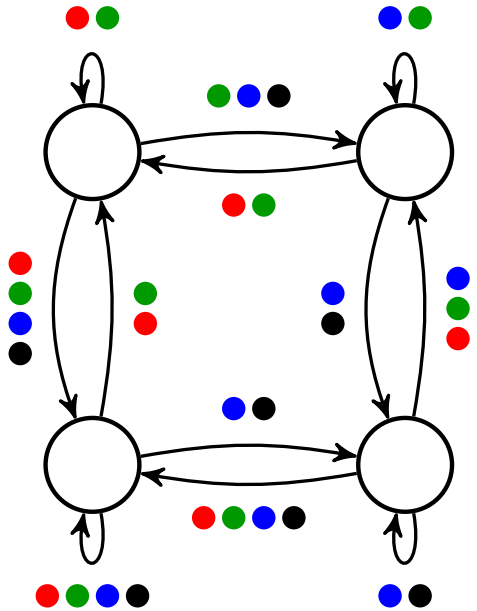


Transitions with Parameters (coloured transitions)



- each parameter valuation represents one Kripke structure
- shared state space, different transition space

Transitions with Parameters (coloured transitions)



- each parameter valuation represents one Kripke structure
- shared state space, different transition space
- symbolic representation of parameters
- symbolic PKS: every transition is associated with a formula

From ODE Models to Kripke Structures

Rectangular Abstraction of Reaction Kinetics

[Belta, Habets, Schuppen]

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} P + E$$

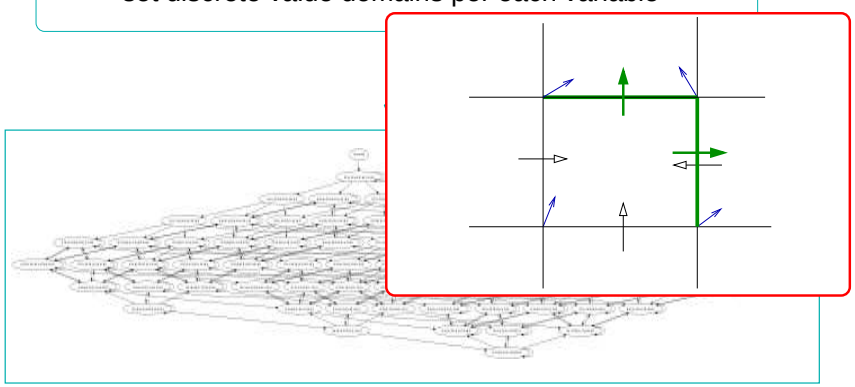
mass action kinetics

with limitation of 1 molecule per each reactant species

multi−affine ODEs

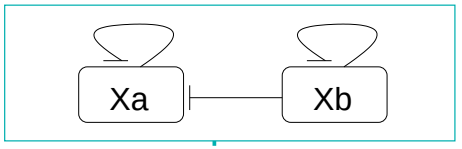
$$\frac{d[S]}{dt} = -k_1[E][S] + k_2[ES]$$
$$\frac{d[E]}{dt} = -k_1[E][S] + k_2[ES] + k_3[ES]$$
$$\frac{d[ES]}{dt} = k_1[E][S] - k_2[ES] - k_3[ES]$$
$$\frac{d[P]}{dt} = k_3[ES]$$

set discrete value domains per each variable



Rectangular Abstraction of Regulatory Kinetics

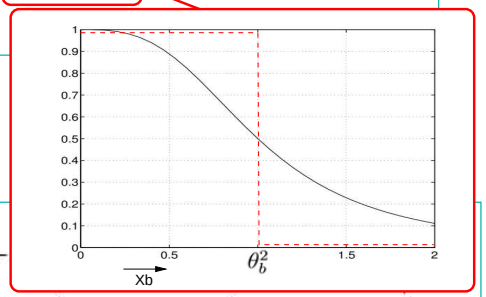
[de Jong, Batt]

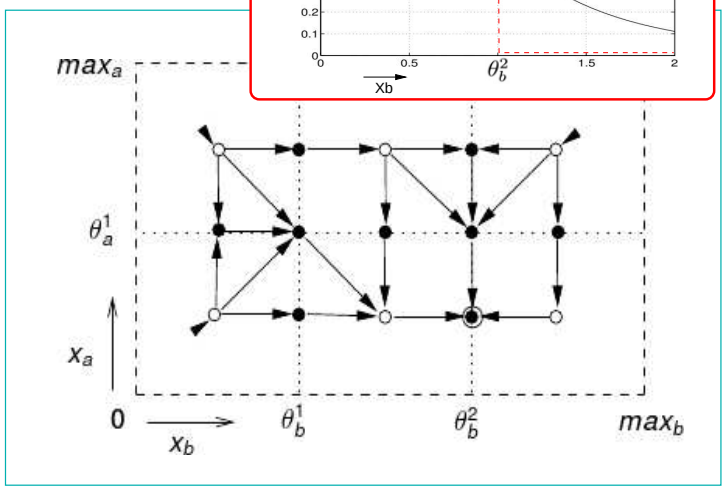


Hill kinetics

piece−wise affine ODEs

$$\frac{dx_a}{dt} = \kappa_a\, s^-(x_a, \theta_a^1)\, s^-(x_b, \theta_b^1) - \gamma_a\, x_a,$$
$$\frac{dx_b}{dt} = \kappa_b\, s^-(x_b, \theta_b^2) - \gamma_b\, x_b$$



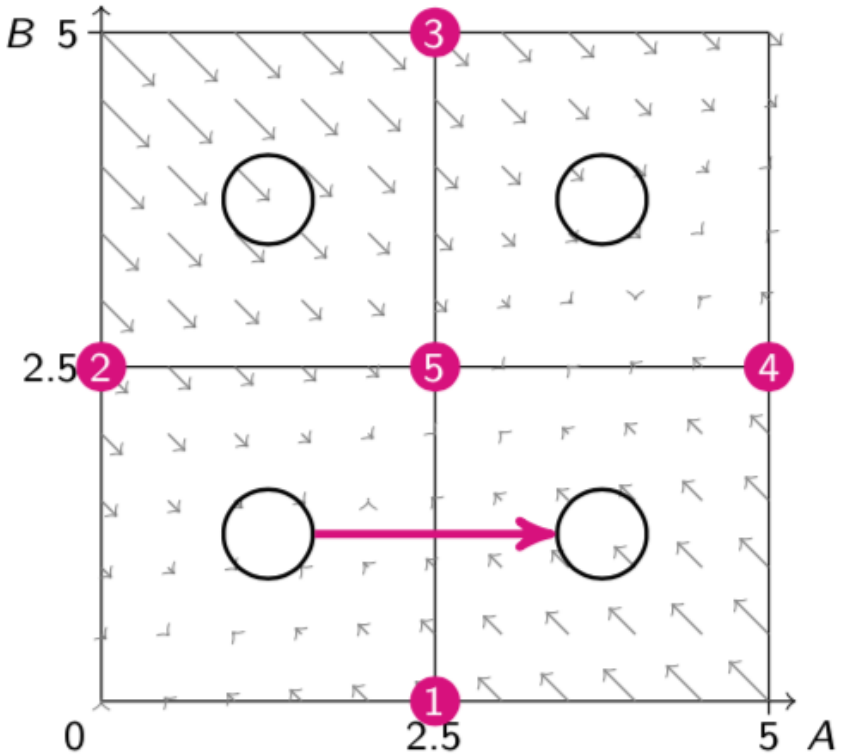


$$\frac{dA}{dt} = -k_1 \cdot A + k_2 \cdot B$$

$$\frac{dB}{dt} = k_1 \cdot A - k_2 \cdot B$$

$$k_2 = 0.8$$

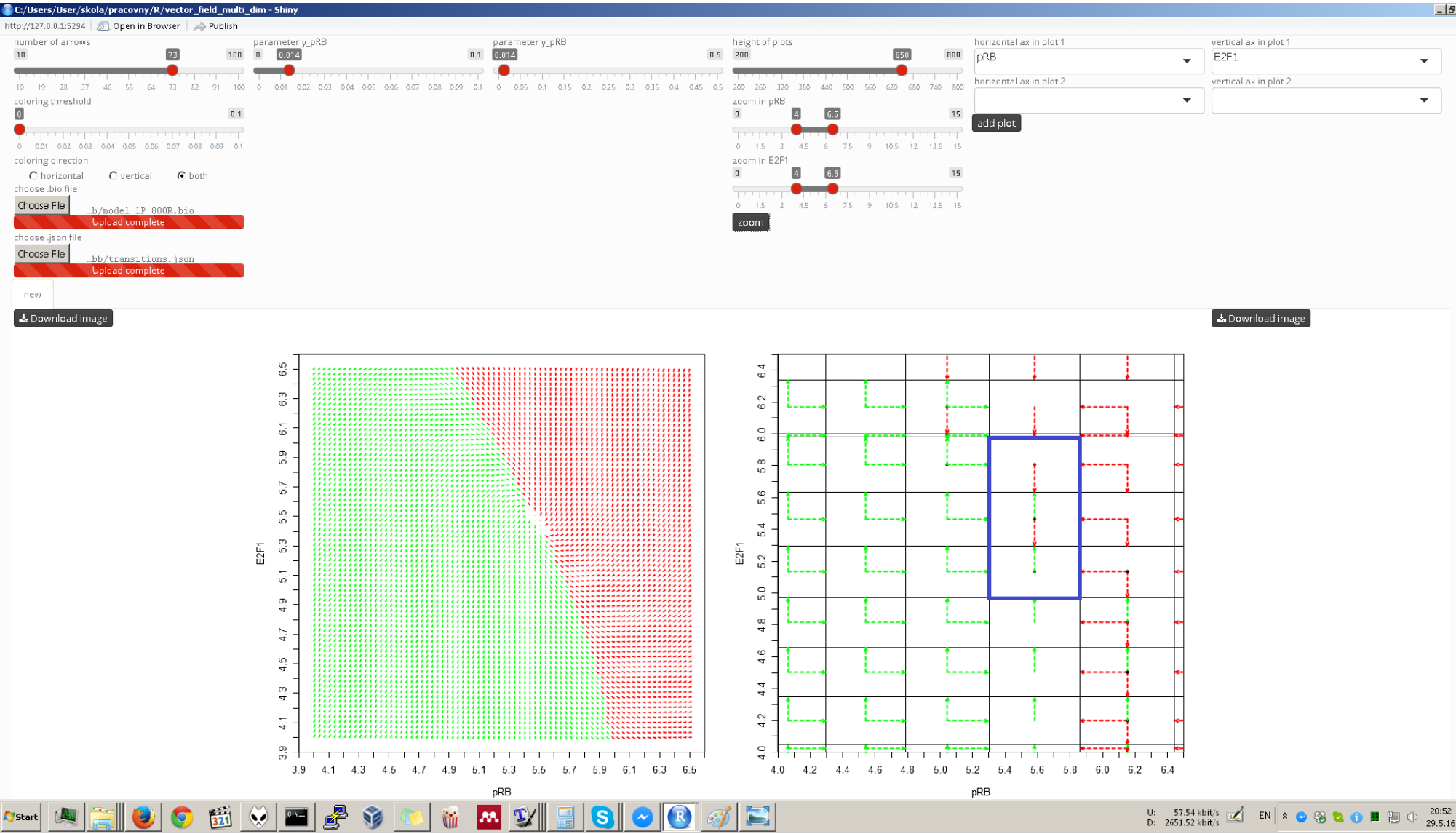
$$k_1 = ?$$



$$\Phi_{\text{state00}\to\text{state10}} := -2.5 \cdot k_1 > 0 \vee -2.5 \cdot k_1 + 2.5 \cdot k_2 > 0$$

The transition exists if and only if the formula is **satisfiable**.
Local parameter constraints are **predicates over reals**.

Pithya Tool



http://pithya.ics.muni.cz

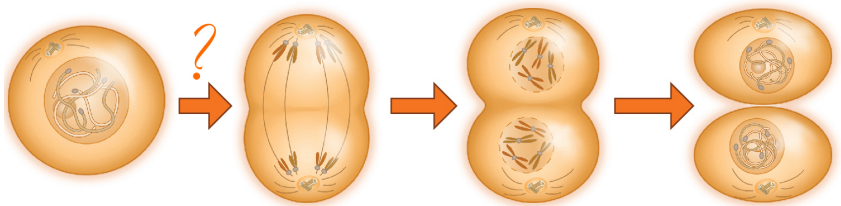
[CAV 2017]

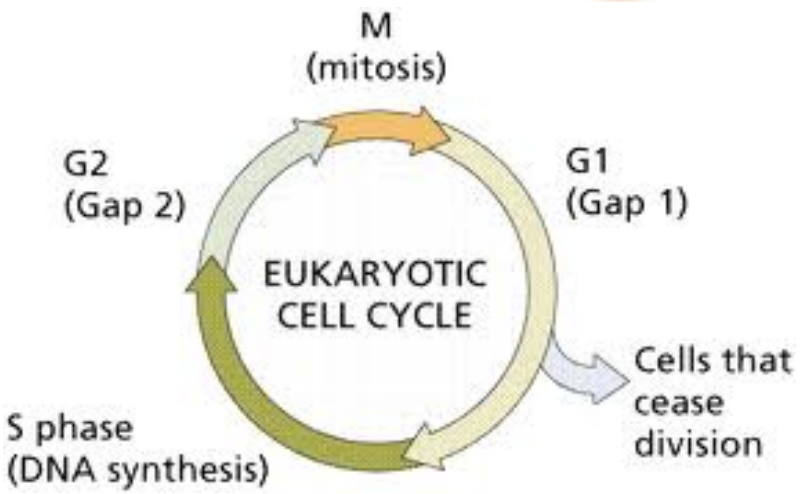
Outline

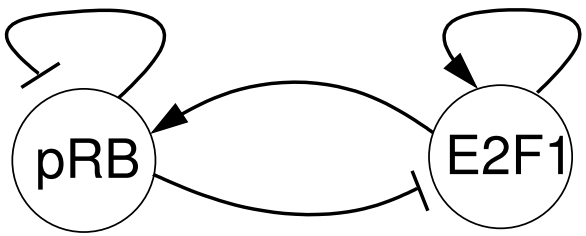
Case Study: Cell Cycle Control System

Example: decision making in living cells

— to divide or not to divide?







[Swat et al. 2004]

decisions implemented by circuits of positive and negative interactions
modelling of cell cycle since 1970 [Goldbetter et al.]

Case Study: Cell Cycle Control System

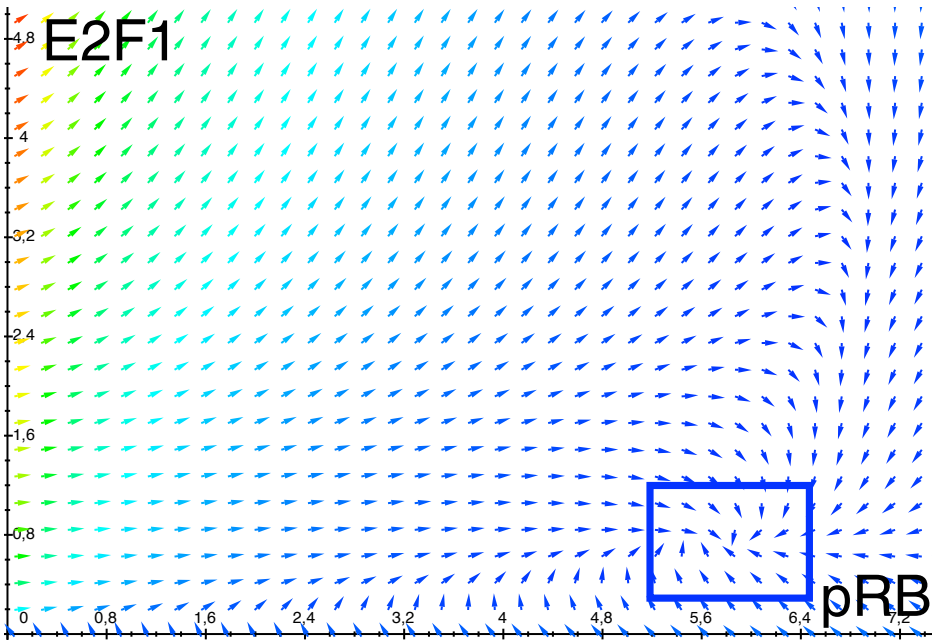
Parameterised Non-Linear Mathematical Model

$$\dot{x} = f(x(t), p)$$

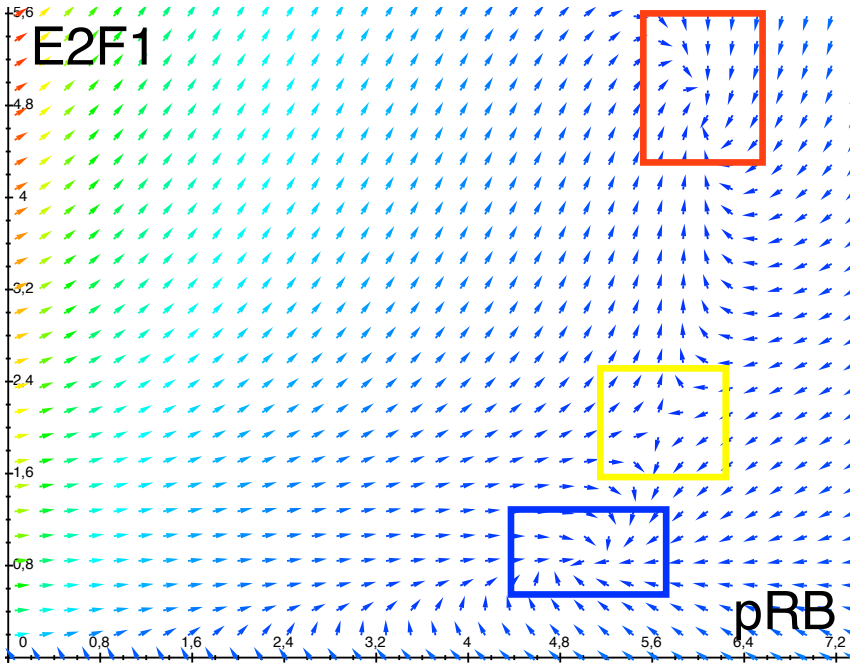
f ... phase space (vector field), $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$

x ... state vector ($\mathbb{R}^n$)

p ... parameter vector ($\mathbb{R}^m$)



$p = 0.006$



$p = 0.012$

Properties Specification: HCTL

- HCTL — hybrid CTL with past
- state formulae

$$\varphi ::= \mathit{true} \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbf{E}\,\psi \mid \mathbf{A}\,\psi \mid$$
$$\hat{\mathbf{E}}\,\psi \mid \hat{\mathbf{A}}\,\psi \mid x \mid \downarrow x.\varphi \mid @x.\varphi \mid \exists x.\varphi$$

- path formulae

$$\psi ::= \mathbf{X}\,\varphi \mid \varphi\,\mathbf{U}\,\varphi$$

Properties Specification: HCTL

Single-state patterns

- sink (stable steady state): $\downarrow s.\, \mathbf{AX}\, s$
- source (only self-loops, no other incoming): $\downarrow s.\, \hat{\mathbf{AX}}\, s$

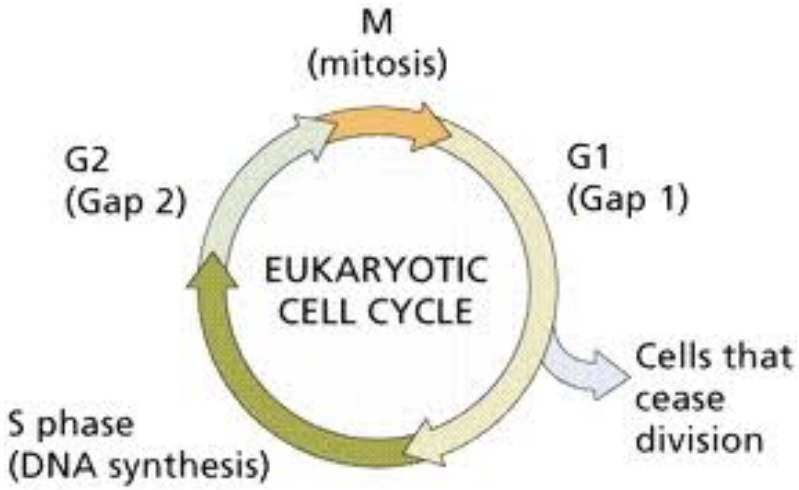
Multi-state patterns

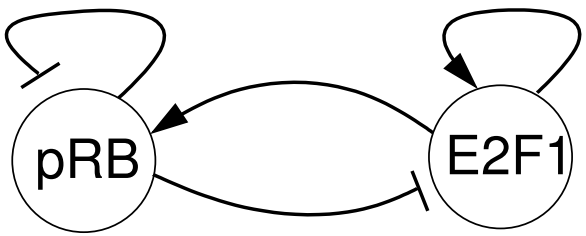
- state in a nontrivial SCC: $\downarrow s.\, \mathbf{EX}\, \mathbf{EF}\, s$
- state in a final SCC (generalised sink): $\downarrow s.\, \mathbf{AG}\, \mathbf{EF}\, s$

Relations among patterns

- at least two sinks in the whole system:
  $\exists s. \exists t. (@s. \neg t \wedge \mathbf{AX}\, s) \wedge (@t.\, \mathbf{AX}\, t)$

Case Study: Regulation of G_1/S Cell Cycle Transition





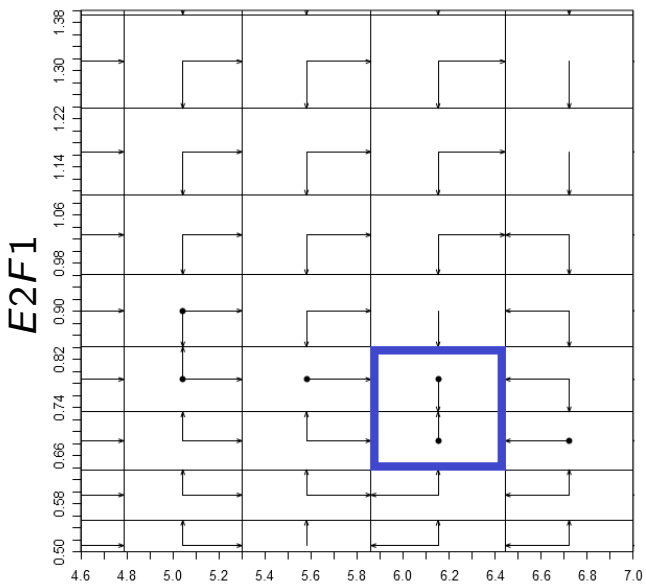
[Swat et al. 2004]

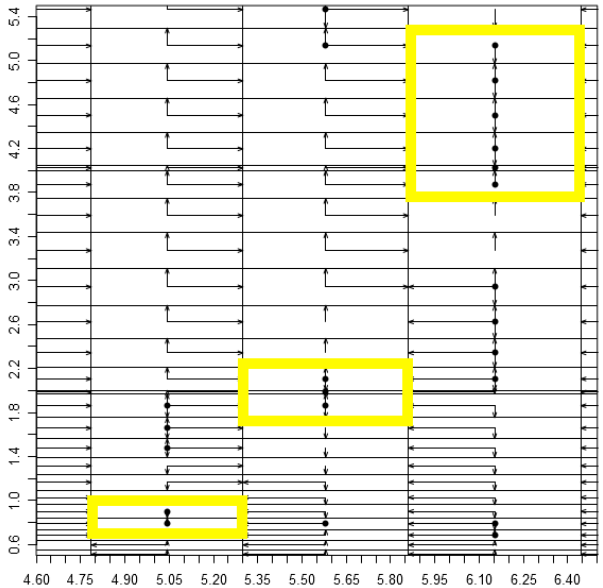
$$\frac{d[pRB]}{dt} = k_1 \frac{[E2F1]}{K_{m1}+[E2F1]} \frac{J_{11}}{J_{11}+[pRB]} - \phi_{pRB}[pRB]$$
$$\frac{d[E2F1]}{dt} = k_p + k_2 \frac{a^2+[E2F1]^2}{K_{m2}^2+[E2F1]^2} \frac{J_{12}}{J_{12}+[pRB]} - \phi_{E2F1}[E2F1]$$

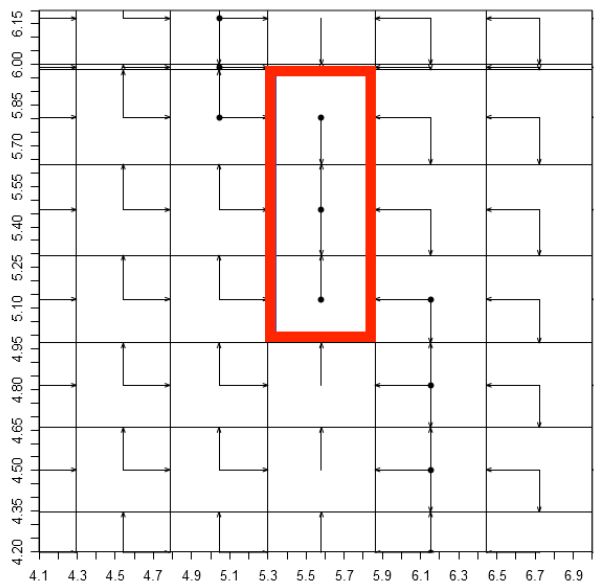
unknown parameter: ϕ_{pRB}

- $\varphi_1 := \exists s.\exists t.(@s.\,\mathbf{AG}\,\mathbf{EF}\,s) \wedge (@t.\neg\,\mathbf{EF}\,s \wedge \mathbf{AG}\,\mathbf{EF}\,t)$
- $\varphi_2 := \neg\varphi_1 \wedge \downarrow s.\,\mathbf{AG}\,\mathbf{EF}\,s \wedge E2F1 < 4$
- $\varphi_3 := \neg\varphi_1 \wedge \downarrow s.\,\mathbf{AG}\,\mathbf{EF}\,s \wedge E2F1 > 4$

Case Study: Results







φ_2

$\phi_{pRB} = 0.0075$

$[0.002, 0.011]$

φ_1

$\phi_{pRB} = 0.0115$

$[0.011, 0.0136]$

φ_3

$\phi_{pRB} = 0.014$

$[0.0136, 0.5]$

results agree with numerical methods up-to precision of approximation/discretisation

Parameter Synthesis Chronology

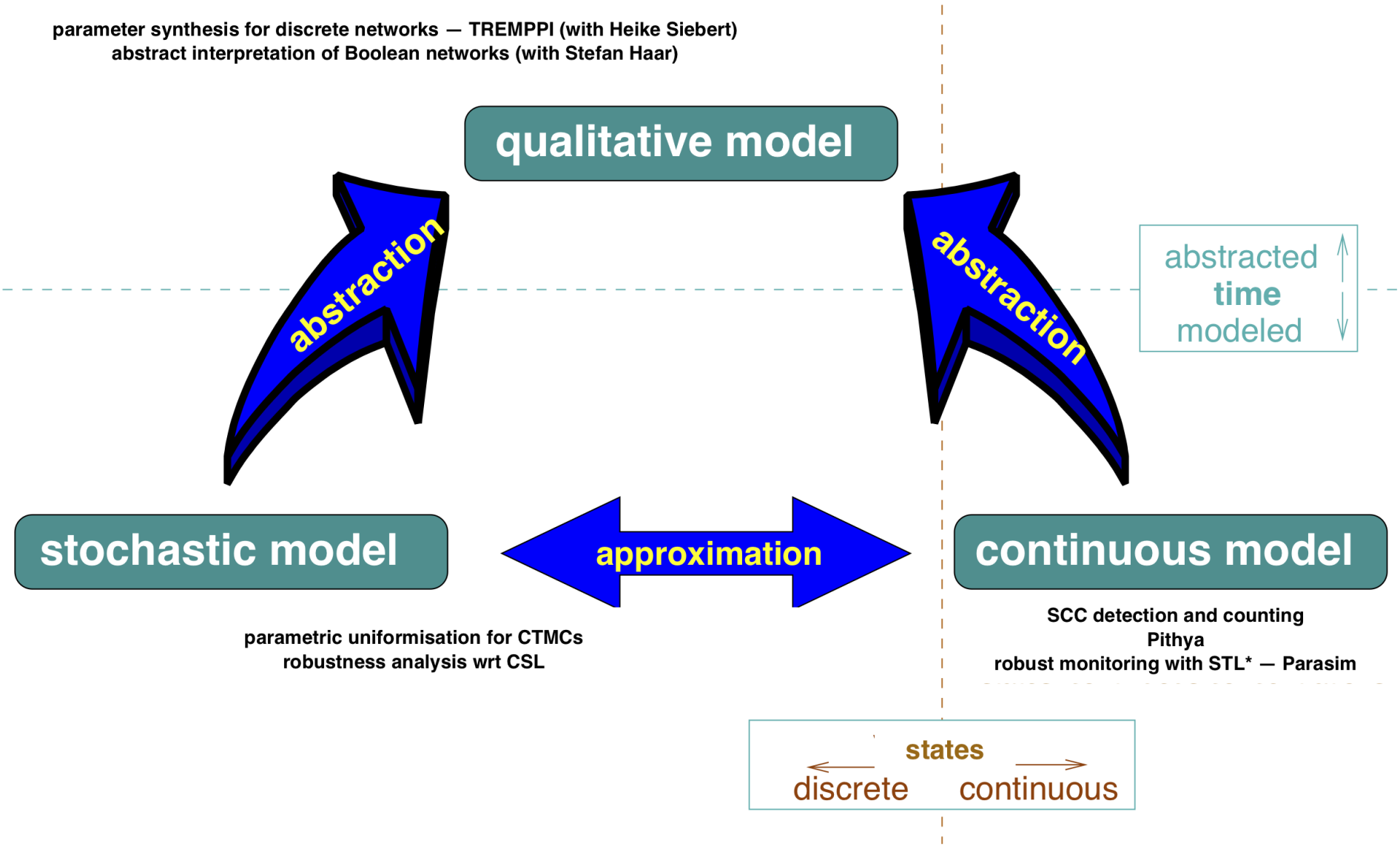
Related Work

- Batt et al. 2007: *RoverGene*, BDD/Polytopes-based approach
- Batt et al. 2010: *GNA*, symbolic approach, piecewise affine
- Grosu et al. 2011: *RoverGene* revisited, approximation improved
- Bogomolov et al. 2015, *SpaceEx*, multi-affine hybrid automata

Our Contribution

- HIBI 2010, TCCB 2012: coloured LTL model checking, piecewise multi-affine, parallel algorithm
- CMSB 2015: coloured CTL model checking, piecewise multi-affine, parallel algorithm
- ATVA 2016, CMSB 2016: parameters represented in first order logic, SMT solver employed, interdependent parameters
- HSB 2015, FM 2016: discrete bifurcation analysis
- CMSB 2017, ICCTCS 2018: analysis of terminal SCCs, application to cyanobacteria models
- TACAS 2019: application to bifurcation analysis of TCP

Contribution Overview



Conclusions

- using methods of computer science we can specify biological systems rigorously
- formal methods allow exhaustive exploration of models under parameter uncertainty
- use of formal methods is important for synthetic biology – we want to know what we construct!
- analysis becomes a push-button technology
- applications in cyber-physical systems
- problems:
  - the grand challenge not yet targeted
  - modellers trained in biophysics and computer science needed
  - scalability
  - we need methods giving results up to given precision instead of insisting on exact results
- Machine Learning to learn $F_{\mathcal{N}}$?

Credits

Computer Science

Luboš Brim, Marta Kwiatkowska, Jiří Barnat, Thomas Henzinger, Loïc Paulevé, Ezio Bartocci, Luca Bortolussi, Jérôme Feret, Andrzej Mizera, Alessandro Abate, Jan Van Schuppen, Milan Češka, Nikola Beneš, Stefan Haar, Heike Siebert, Hidde de Jong, Ivana Černá

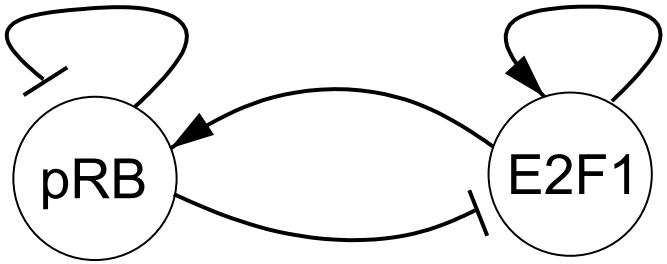
Biology

Ralf Steuer, Louis Mahadevan, Jan Červený, Dušan Lazár, Pavel Krejčí, Stefanie Hertel, Christoff Flamm

Students

- Samuel Pastva, Matej Troják, Sven Dražan, Jana Dražanová, Martin Demko, Matej Hajnal
- Tomáš Vejpustek, Juraj Kolčák, Jan Papoušek, Vojtěch Brůža, Juraj Nižnan, Lukrécia Mertová, Petr Dluhoš, Simon Van Goethem

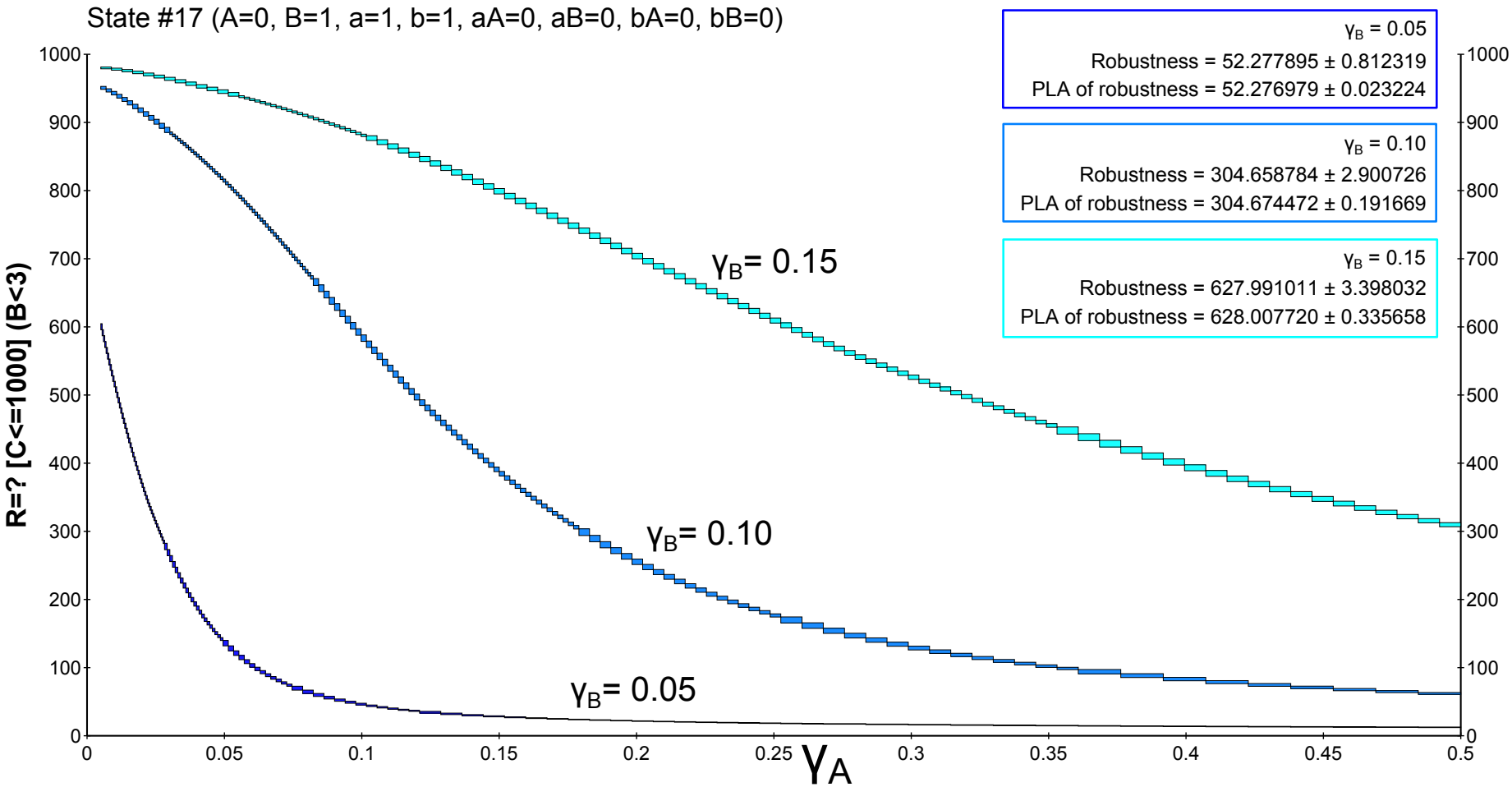
Parameter Exploration of Stochastic Models



Gene a interactions		Gene b interactions	
$a \rightarrow a + A$	1	$b \rightarrow b + B$	0.05
$aB \rightarrow aB + A$	1	$bB \rightarrow bB + B$	1
$A + a \leftrightarrow aA$	100; 10	$A + b \leftrightarrow bA$	100; 10
$B + a \leftrightarrow aB$	100; 10	$B + b \leftrightarrow bB$	100; 10
Proteins degradation			
$A \rightarrow$	γ_A	$B \rightarrow$	γ_B

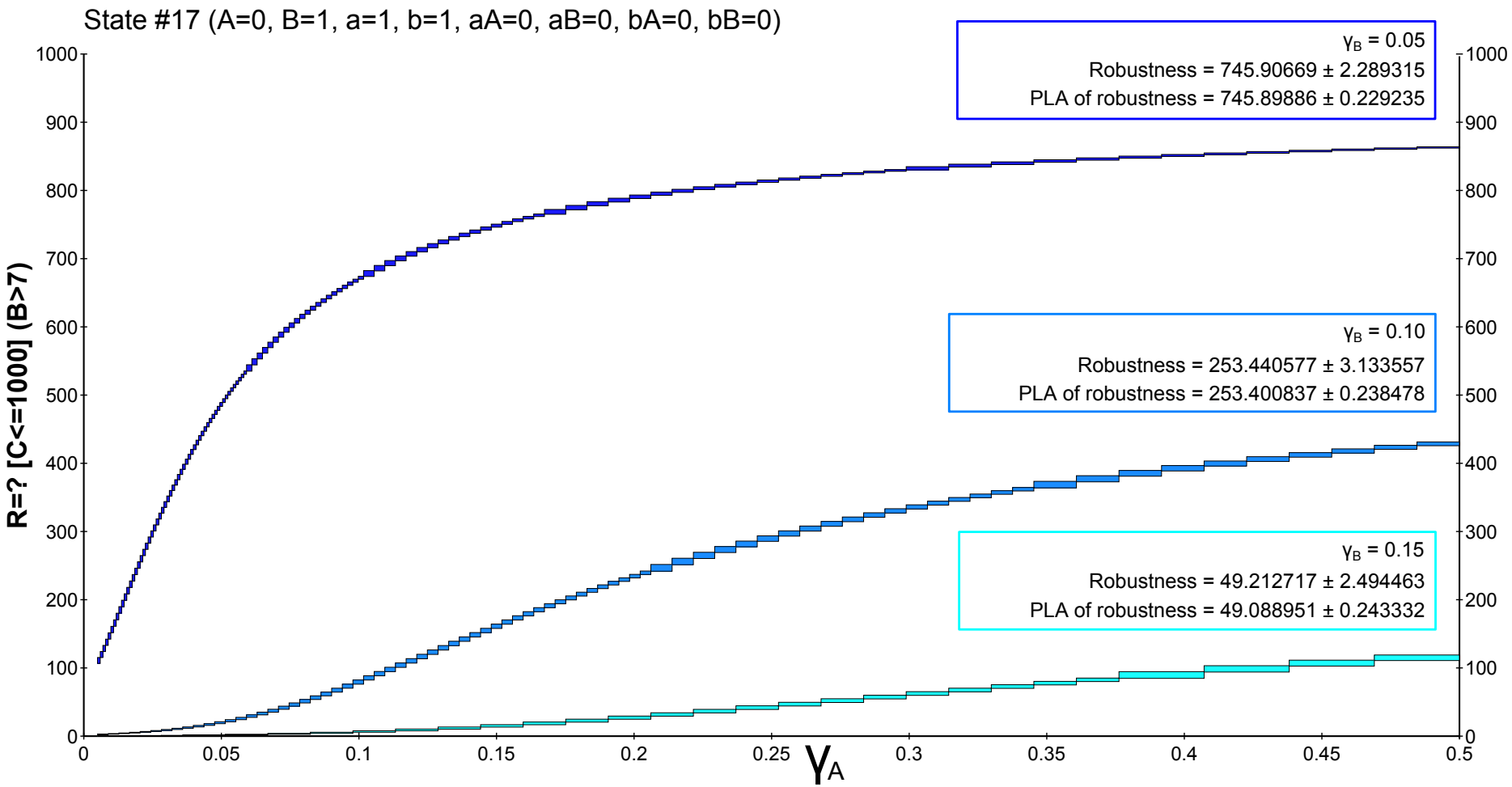
- CTMC with 1078 states and 5919 transitions
- hypothesis about stability of B in low/high population
  - expected time spent in states with low/high population of B
  - formalization in CSL using cumulative rewards
  - $R_{=?}[C^{\leq 1000}](B < 3)$, $R_{=?}[C^{\leq 1000}](B > 7)$

Robustness analysis - stability in low population of B



- in average $3.0 \cdot 10^6$ reaction steps, 100 subspaces, 7 hours

Robustness analysis - stability in high population of B



- in average $3.0 \cdot 10^6$ reaction steps, 100 subspaces, 7 hours