A Couple of Systems Biological Stories told by Formal Methods

David Šafránek

sybila systems biology laboratory

Masaryk University
Czech Republic

Outline

1. Introduction and Motivation
2. Methodology
   - Biological Networks
   - Modelling Problems
3. Story I: Signalling Pathways of Fibroblast Growth Factors
4. Story II: Synthetic Biology: Trichlorpropane Degradation

Outline

1. Introduction and Motivation

2. Methodology
   - Biological Networks
   - Modelling Problems

3. Story I: Signalling Pathways of Fibroblast Growth Factors

4. Story II: Synthetic Biology: Trichlorpropane Degradation

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 and Theoretical Biology

Domain Roots

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)

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

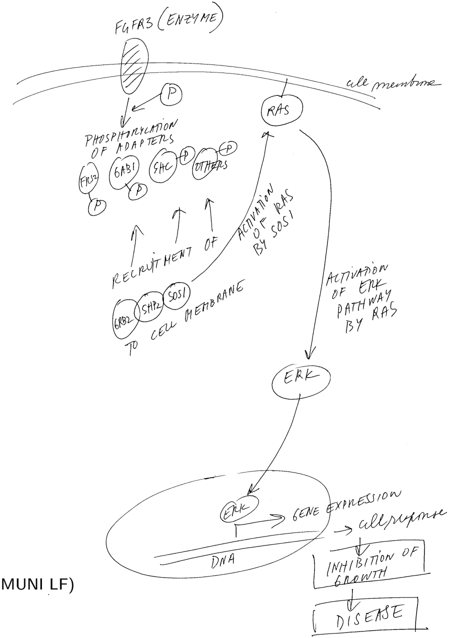
How the bacteria cell **utilises** particular nutritions?

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

The answer should fullfil specific requirements

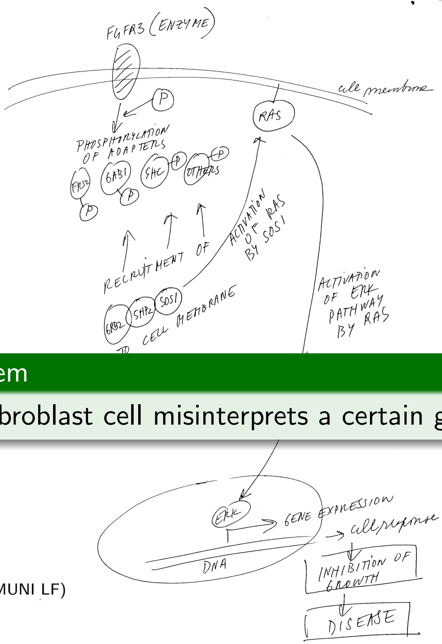
- to formulate and analyse a biological problem **holistically**
- to give mechanistic explanation based on known facts – mechanistic means in the context of laws of physics/chemistry
- to project the mechanistic details onto the genetic information

From Biologist's Table



slide credits: Pavel Krejčí (MUNI LF)

From Biologist's Table



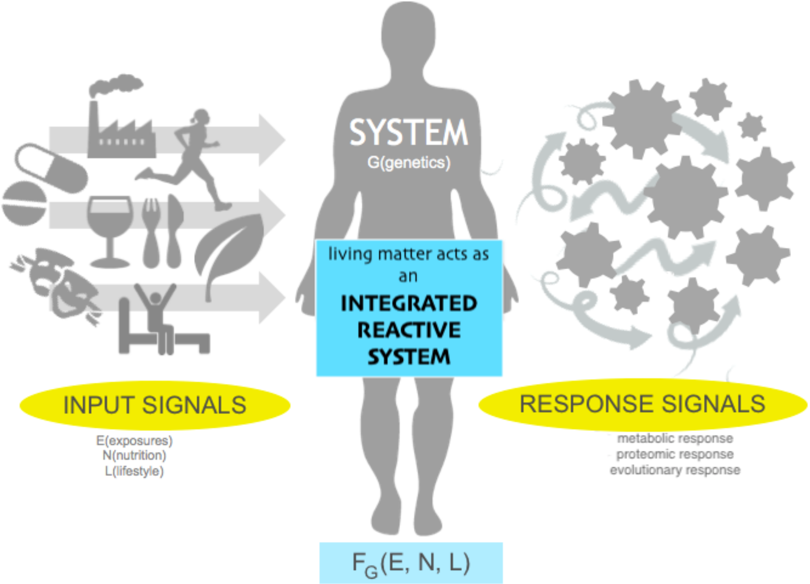
Biological Problem

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



SYSTEM

G(genetics)

living matter acts as an

INTEGRATED REACTIVE SYSTEM

INPUT SIGNALS

E(exposures)
N(nutrition)
L(lifestyle)

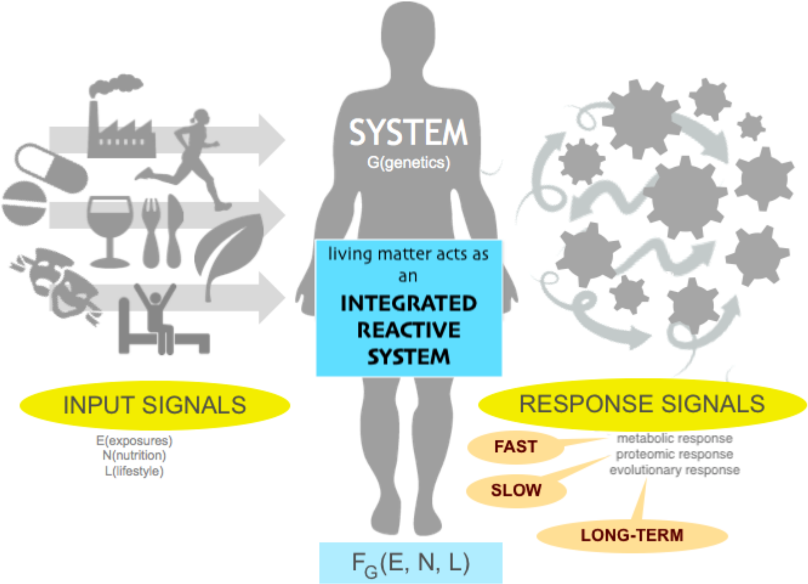
RESPONSE SIGNALS

metabolic response
proteomic response
evolutionary response

$F_G(E, N, L)$

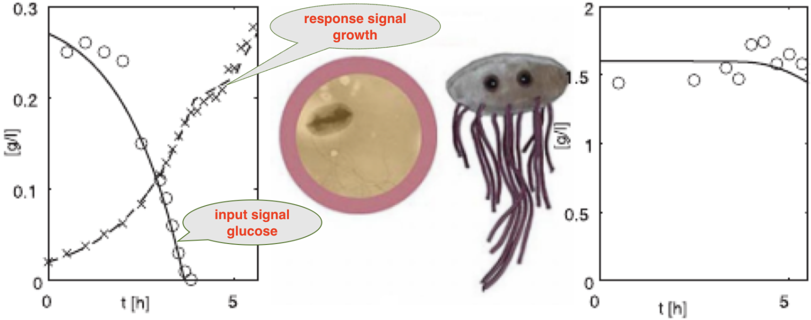
Systems Approach: The Grand Challenge

Complex Organism as a System



Systems Approach: A Moderate Challenge

Population of Bacteria as a System

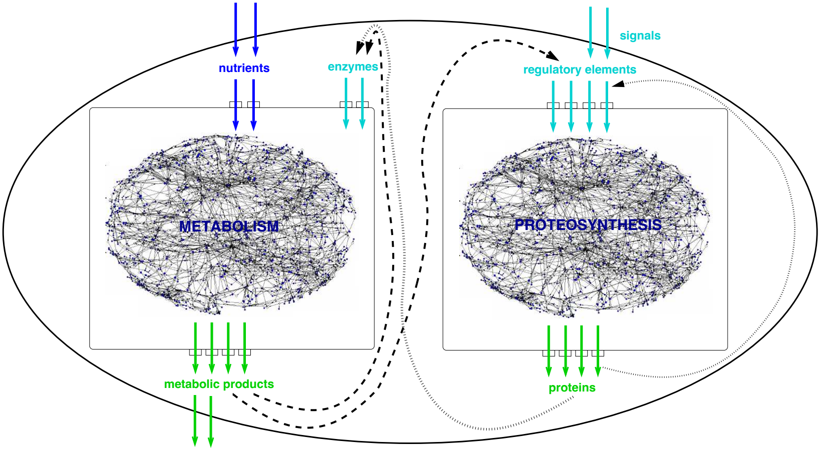


for a particular set of genes G

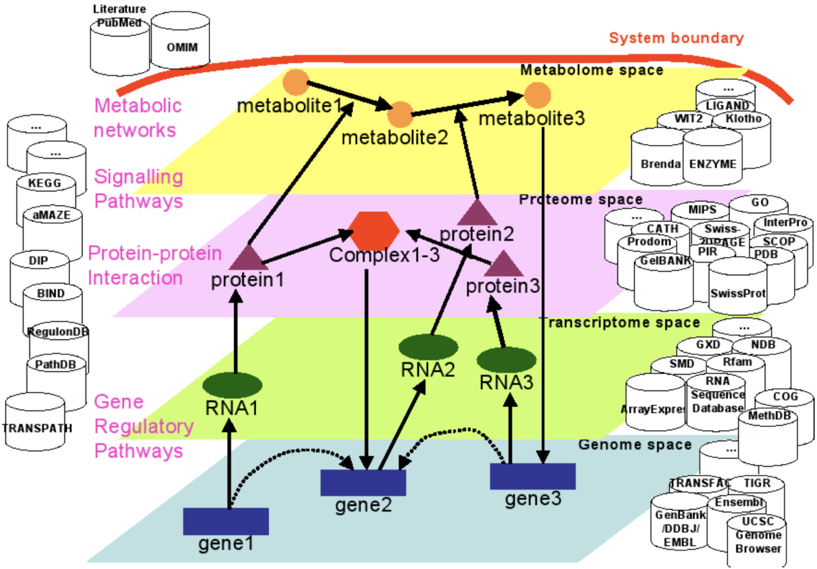
F_G : (environment exposure, nutrition) → growth profile

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

1. Introduction and Motivation

2. Methodology
   - Biological Networks
   - Modelling Problems

3. Story I: Signalling Pathways of Fibroblast Growth Factors

4. Story II: Synthetic Biology: Trichlorpropane Degradation

The Paradigm of Computational Systems Biology

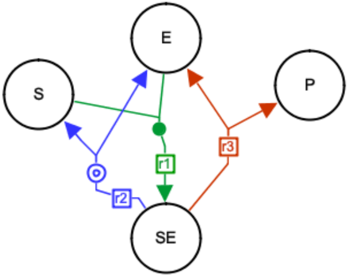
Assumptions

- 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.

Biological Networks

CRNs

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



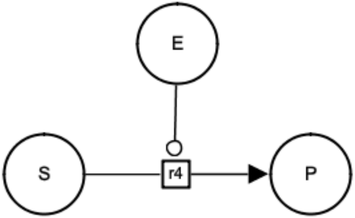
$$S + E \underset{r_2}{\overset{r_1}{\rightleftarrows}} SE \xrightarrow{r_3} P + E$$

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

Biological Networks

Reaction-Influence Networks

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



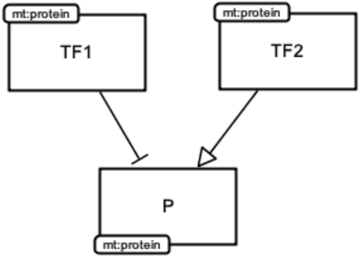
$$S \xrightarrow{r_4} P;\ E$$

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

Biological Networks

Influence Networks

- 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`

The Goal of Computational Systems Biology

The General Goal

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

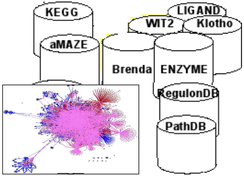
Define a function that encodes the information (signal) processing occuring in **all** *components of the system in time.*

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

Model-Based Workflow

network reconstruction

biological knowledge databases

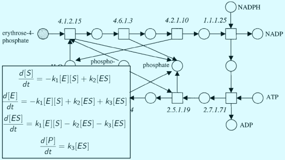


biological network

hypothesis

model specification

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



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

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

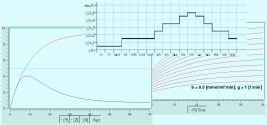
$$\frac{d[ES]}{dt} = k_1[E][S] - k_2[ES] - k_3[ES]$$

$$\frac{d[P]}{dt} = k_3[ES]$$

computer lab

model analysis

static analysis, numerical simulation, analytical methods, model checking



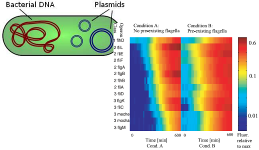
hypothesis testing, property detection, new hypothesis inference

emergent properties

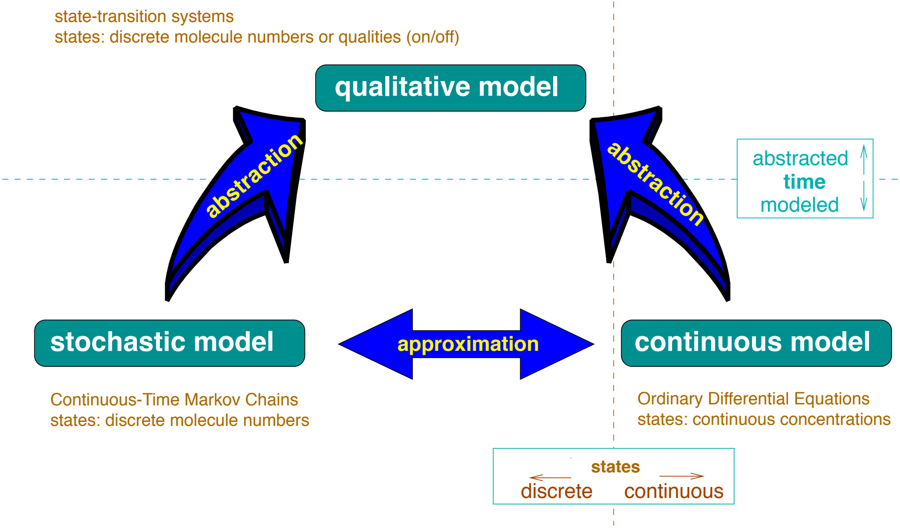
model questions

model validation

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



Modelling Frameworks



state-transition systems
states: discrete molecule numbers or qualities (on/off)

qualitative model

abstraction

abstraction

abstracted
time
modeled

stochastic model

approximation

continuous model

Continuous-Time Markov Chains
states: discrete molecule numbers

Ordinary Differential Equations
states: continuous concentrations

states
discrete continuous

Outline

1. Introduction and Motivation

2. Methodology
   - Biological Networks
   - Modelling Problems

3. Story I: Signalling Pathways of Fibroblast Growth Factors

4. Story II: Synthetic Biology: Trichlorpropane Degradation

Why to model?

Achondroplasia

Thanatophoric Dysplasia

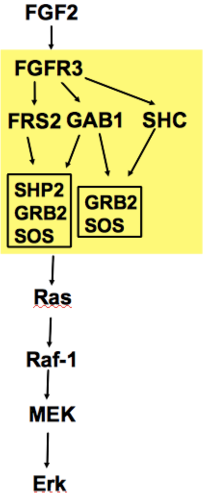
- short long bones
- brachydactyly
- macrocephaly
- low nasal bridge
- spinal stenosis
- temporal lobe malformations

Nat Genet 1995, 9:321-8.

e.g., FGFR3-related skeletal dysplasia

Why to model?

Need to explain...



FGF2

FGFR3

FRS2 GAB1 SHC

SHP2 GRB2 SOS

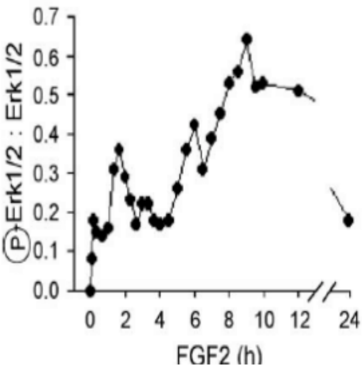
GRB2 SOS

Ras

Raf-1

MEK

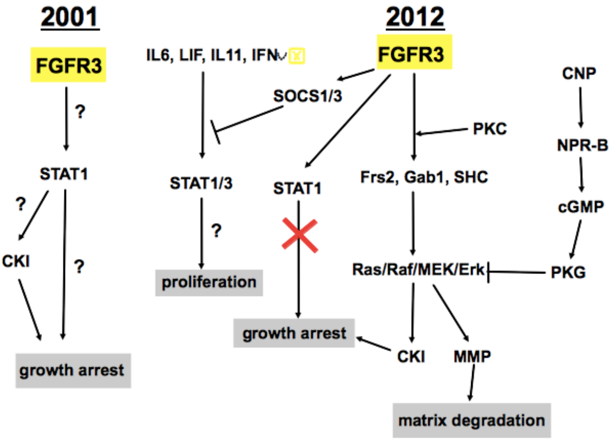
Erk



P. Krejčí, Masaryk University

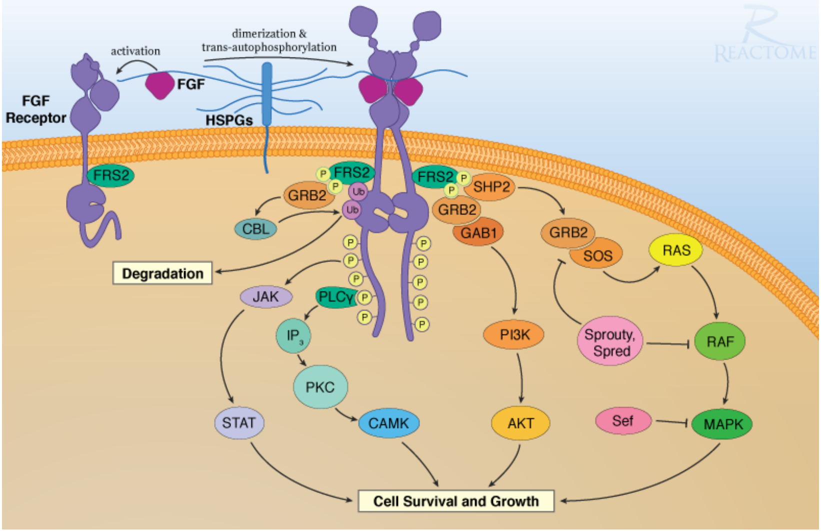
Why to model?

Try to fill in the unknown information...



Current State of Knowledge

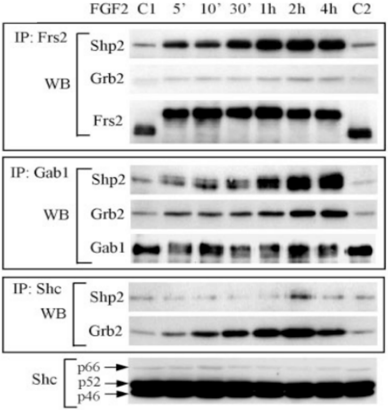
Reactome Database



also available in SBGN on Reactome.org

Wet-lab Measurements

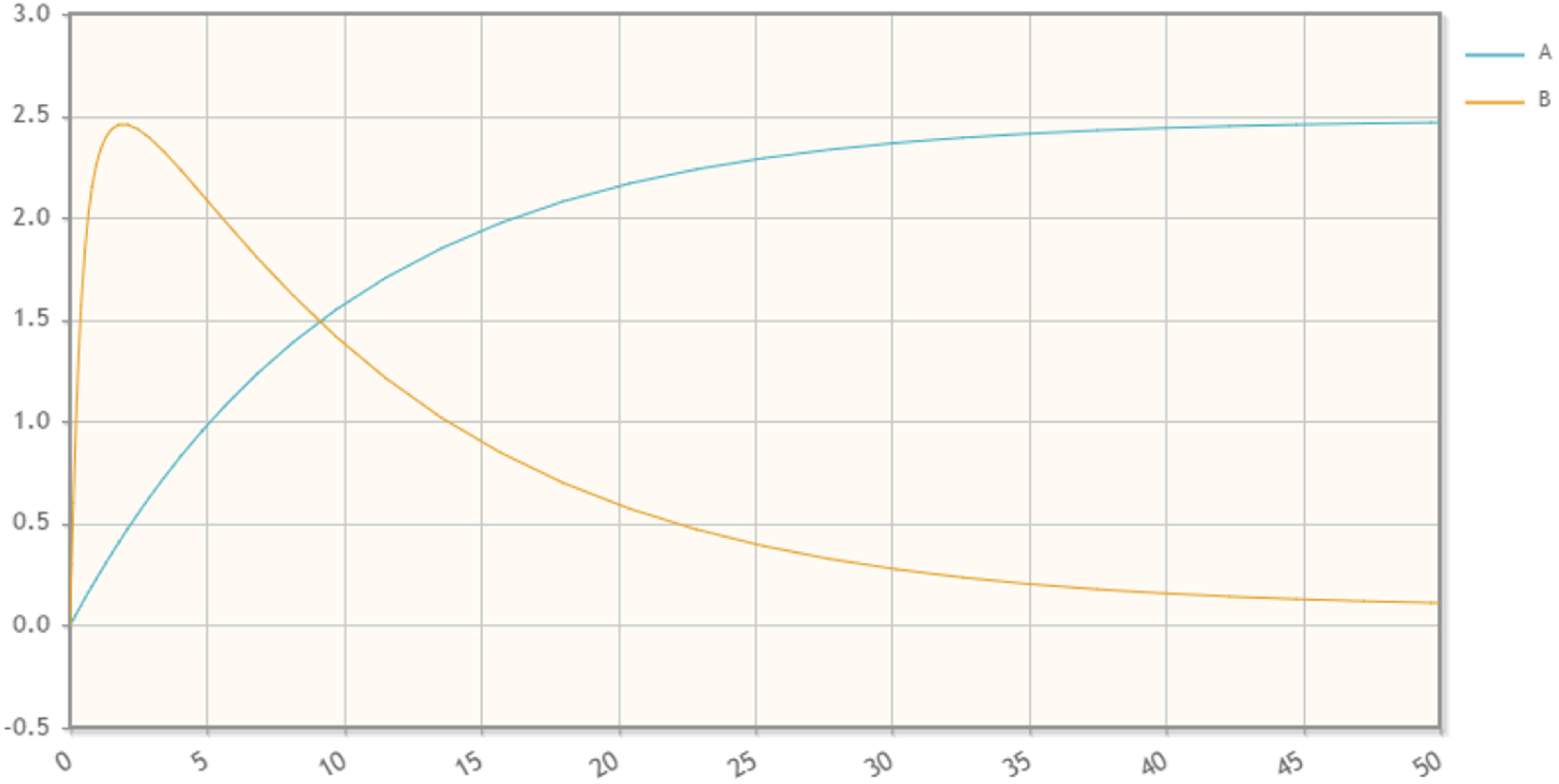
- western blots
- measurements of protein binding (presence of certain proteins)



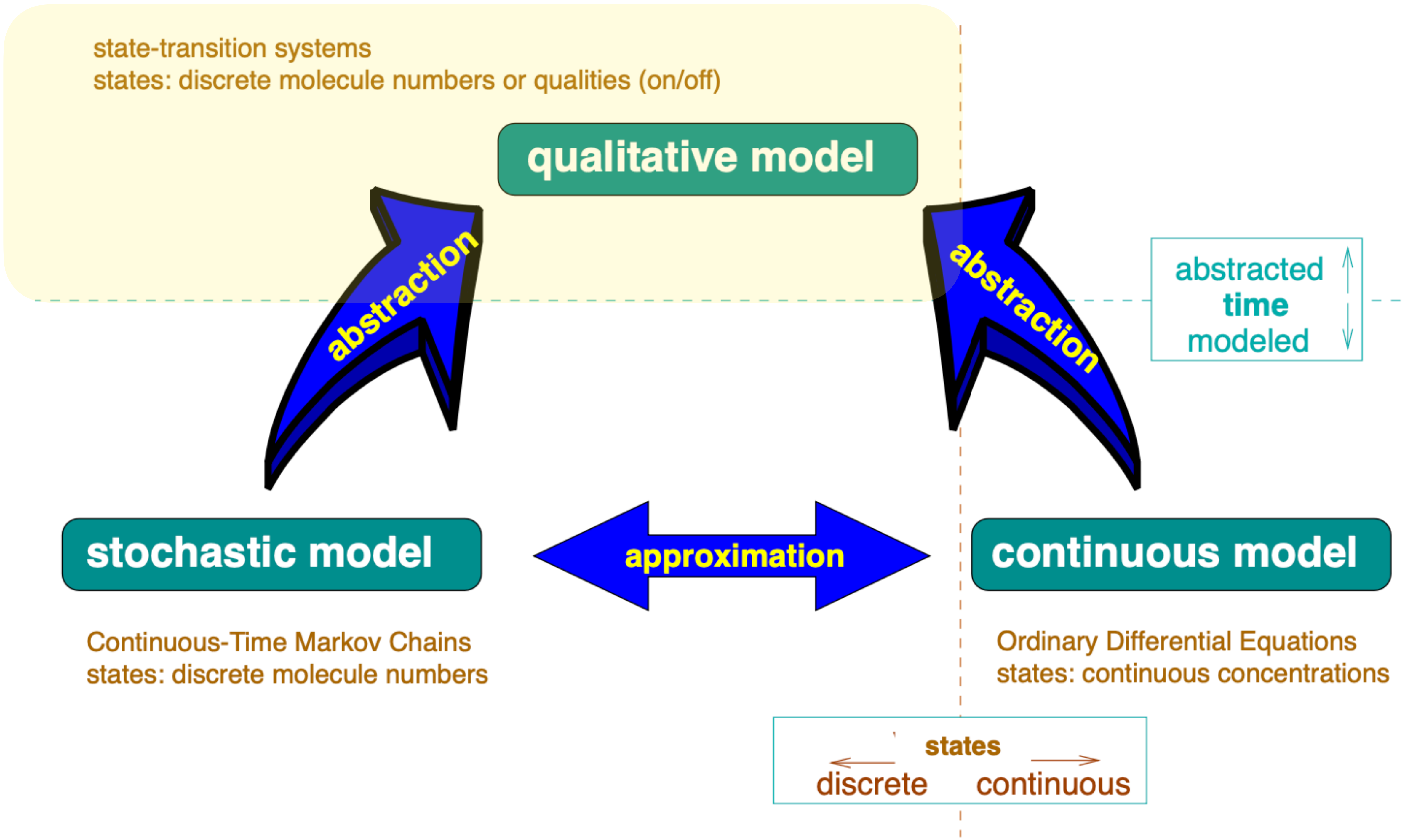
Signal Response Modes

Transient vs. Sustained Effector Activation

- In both cases we assume constantly active growth factor (FGF)
- Pathological cells display sustained response (A)
- Healthy cells display non-monotonous response (B)



Modelling Frameworks



state-transition systems
states: discrete molecule numbers or qualities (on/off)

qualitative model

stochastic model

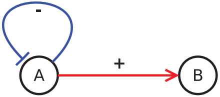
continuous model

Continuous-Time Markov Chains
states: discrete molecule numbers

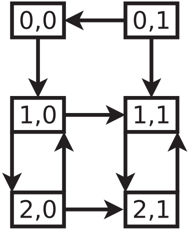
Ordinary Differential Equations
states: continuous concentrations

Model Construction

Qualitative View of Influence Nets – Discrete Regulatory Networks



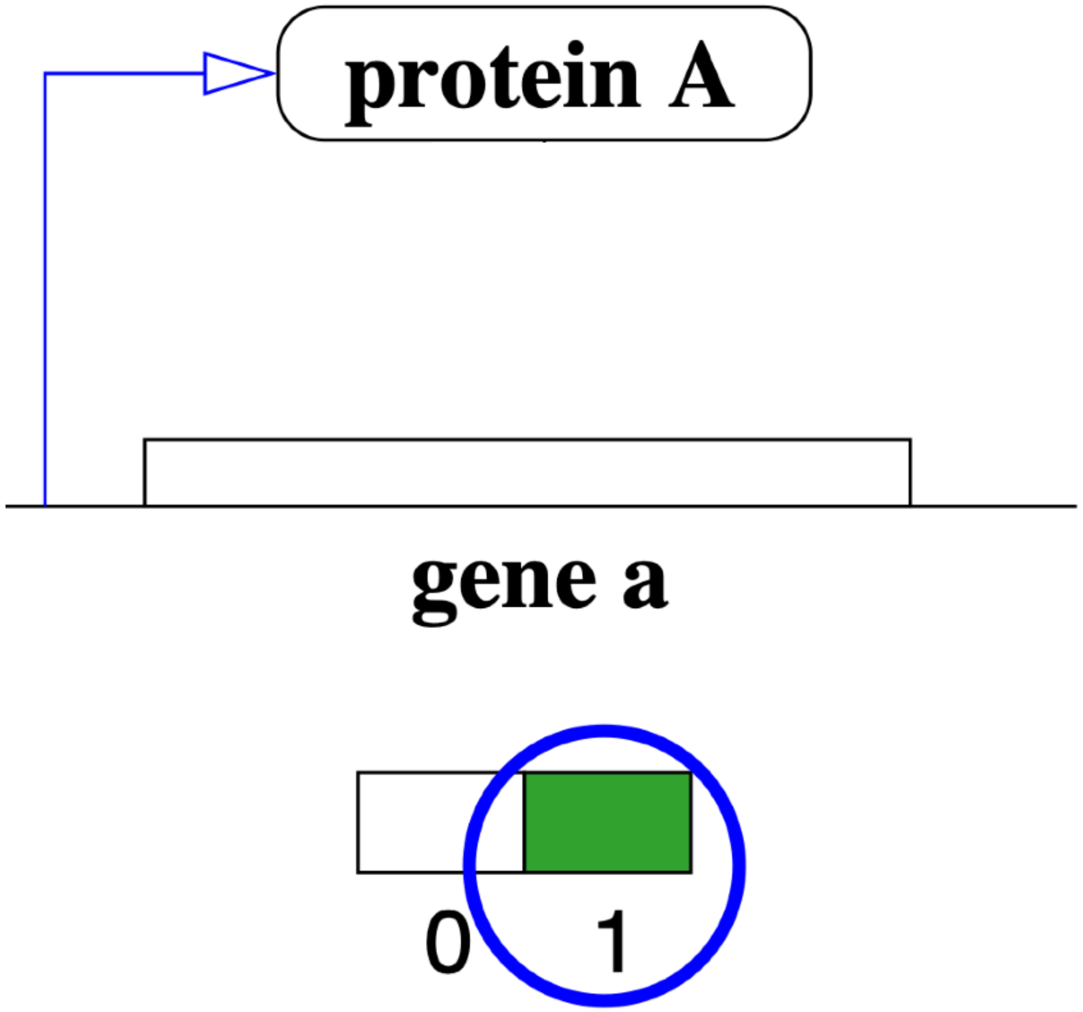
$$A \in \{0,1,2\},\ B \in \{0,1\}$$
$$t_{AA} = 2,\ t_{AB} = 1$$
$$K_{A,\emptyset} = 2$$
$$K_{A,\{A\}} = 0$$
$$K_{B,\emptyset} = 0$$
$$K_{B,\{A\}} = 1$$



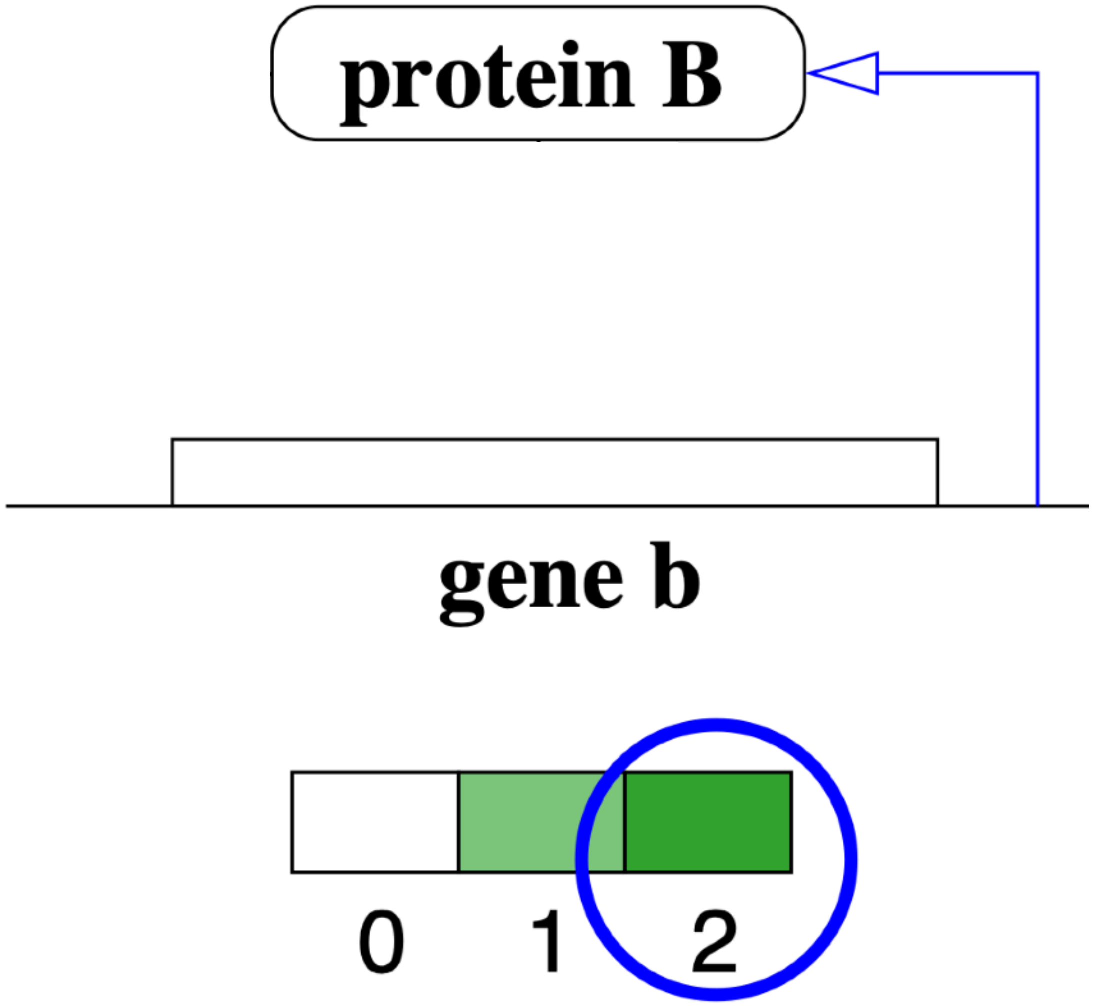
- introduced by René Thomas [1973]
- refined by Chaouiya et al. [2003]
- simplistic case: binary domain $\Longrightarrow$ Boolean Networks

Model Construction

Discrete Models of Influence (Regulatory) Networks



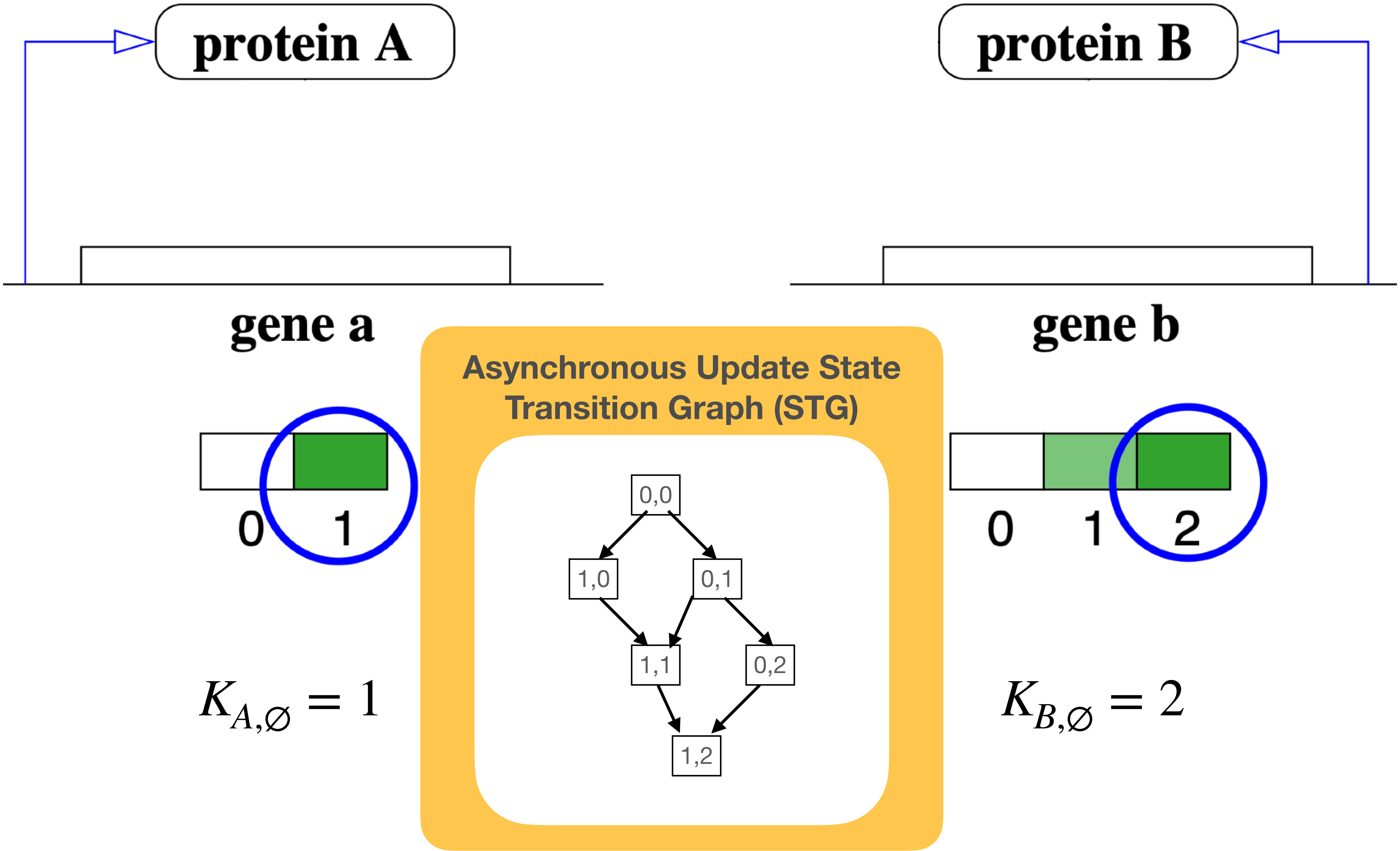
$$K_{A,\varnothing} = 1$$



$$K_{B,\varnothing} = 2$$

Model Construction

Discrete Models of Influence (Regulatory) Networks



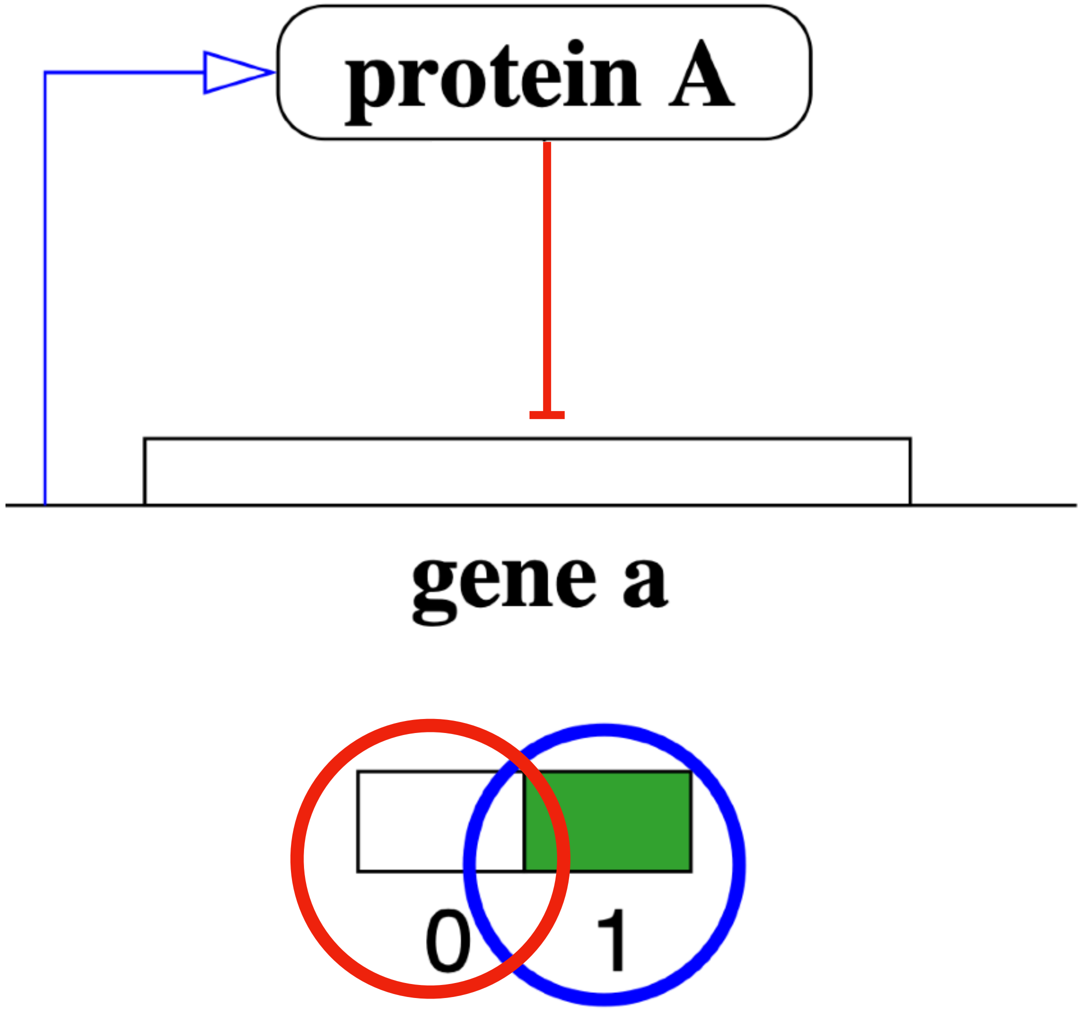
Asynchronous Update State Transition Graph (STG)

$K_{A,\varnothing} = 1$

$K_{B,\varnothing} = 2$

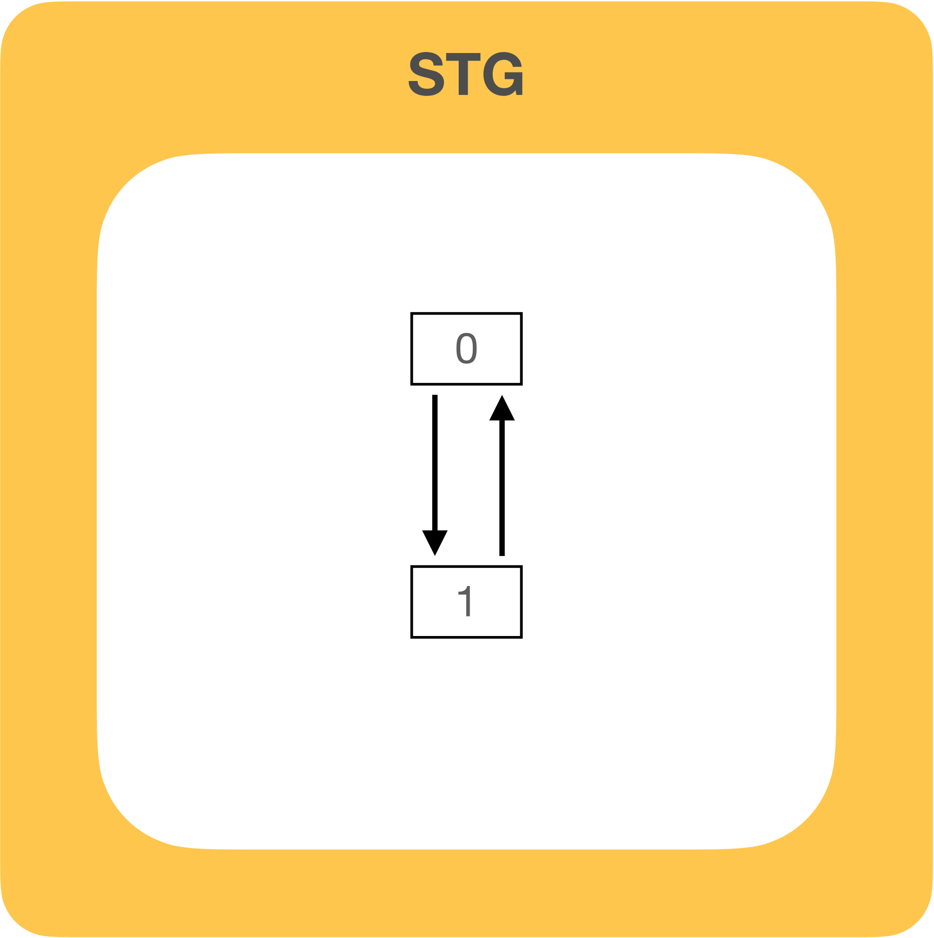
Model Construction

Negative Influence (Inhibition)



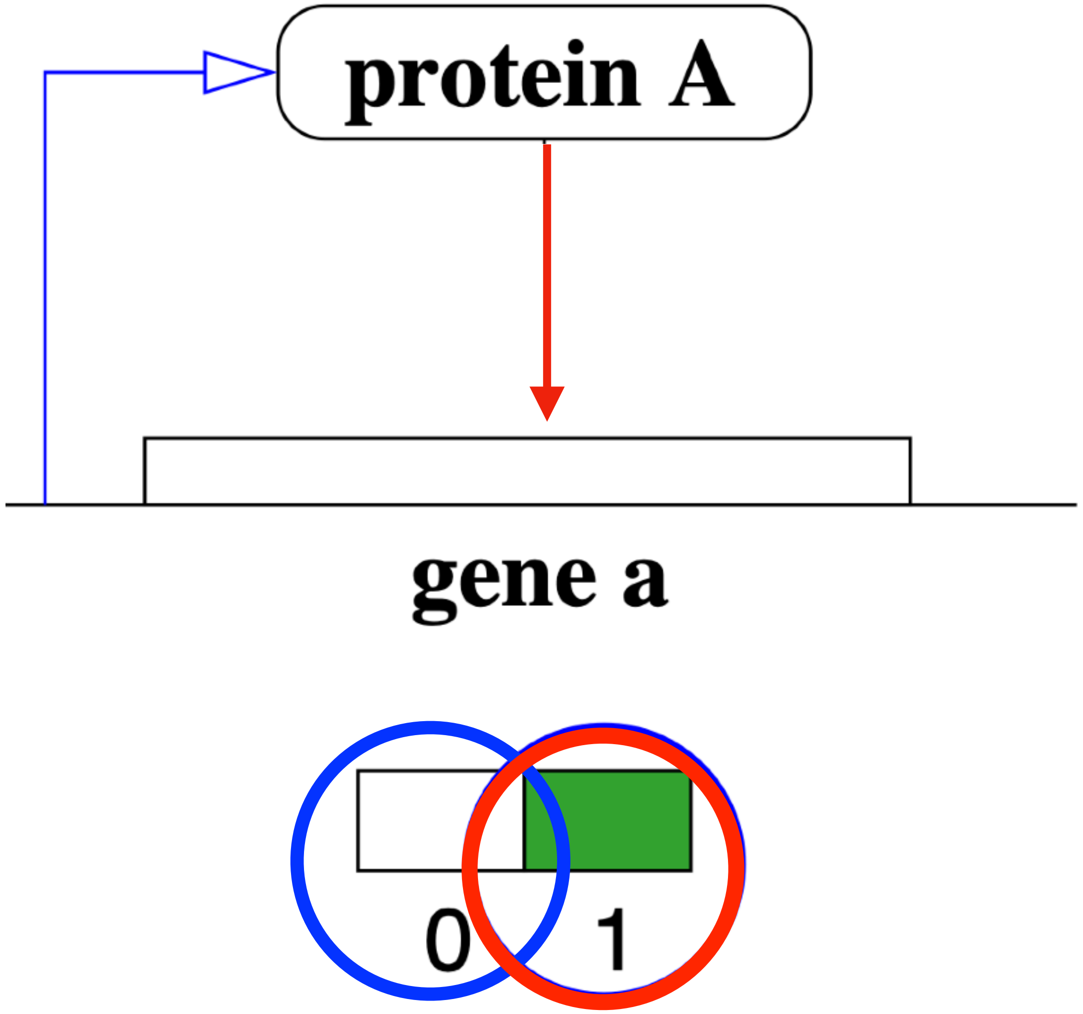
$$K_{A,\emptyset} = 1$$

$$K_{A,\{A\}} = 0$$



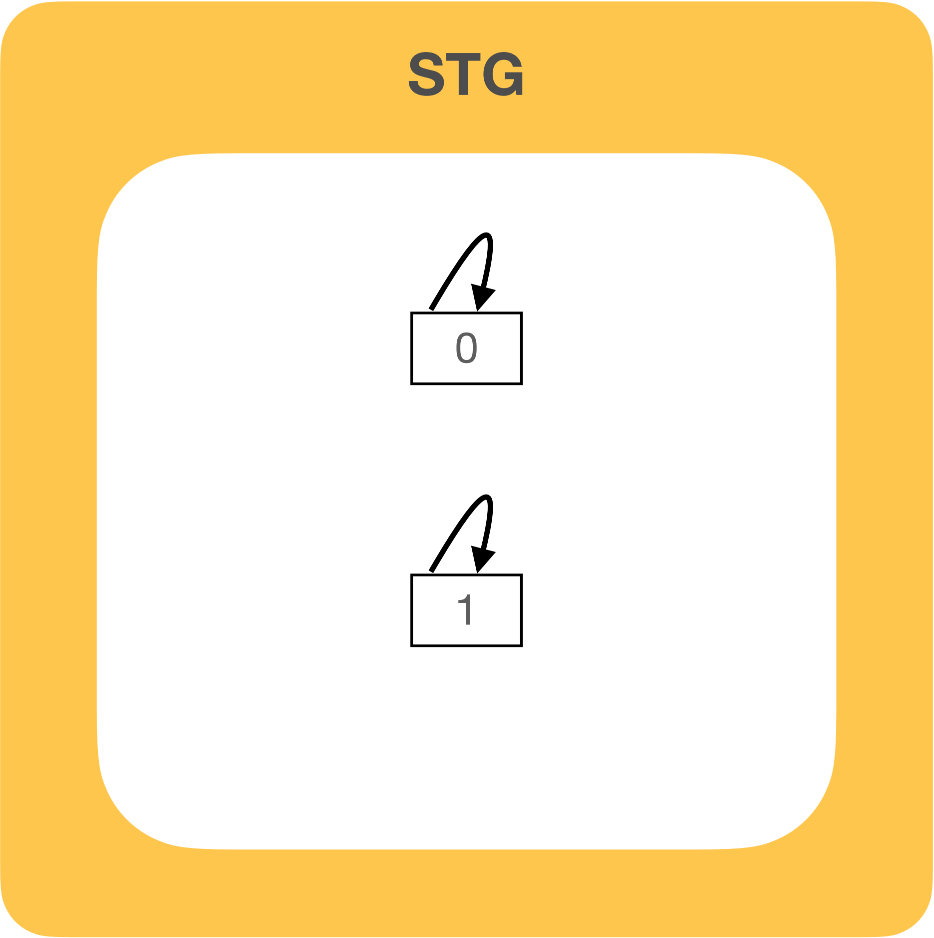
Model Construction

Negative Influence (Inhibition)



$$\color{red}{K_{A,\emptyset} = 0}$$

$$\color{blue}{K_{A,\{A\}} = 1}$$

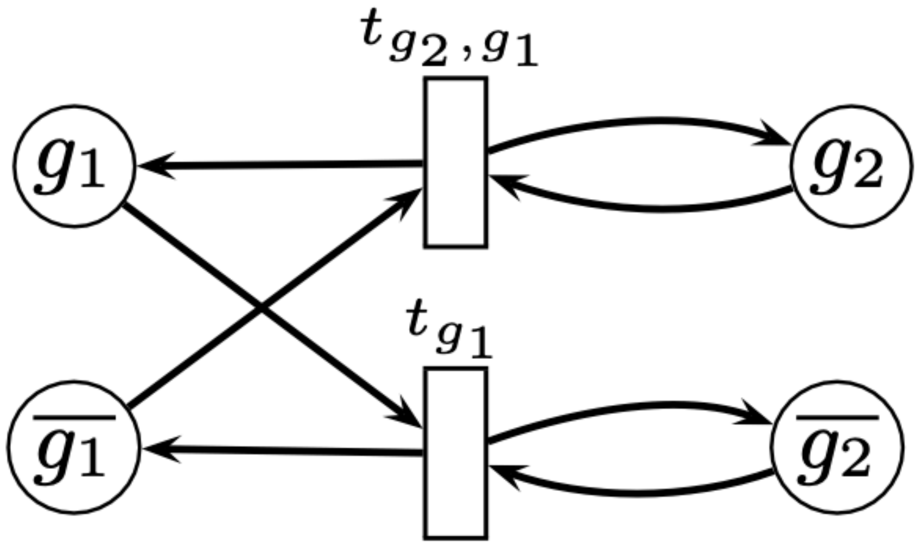


Model Construction — A Remark

Regulatory Networks as Petri Nets

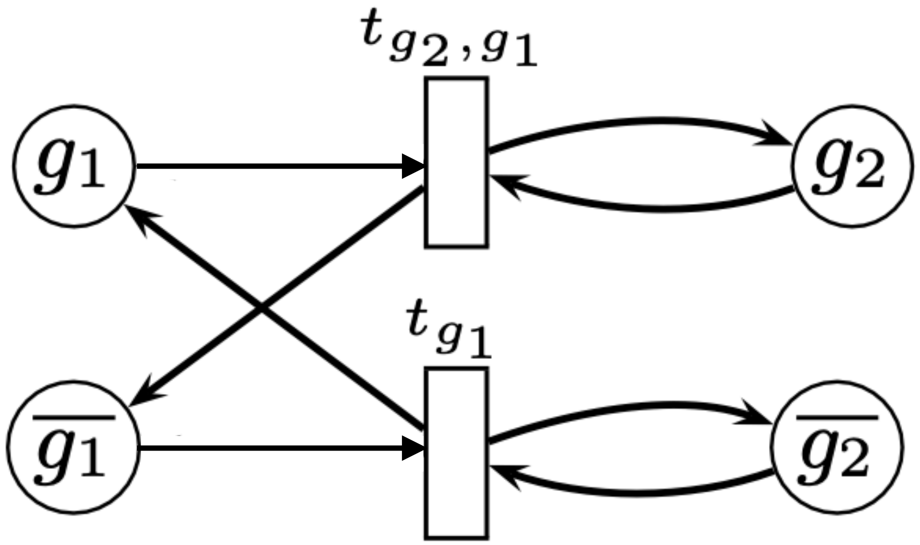
$$K_1(\emptyset) = 0, K_1(\{g_2\}) = 1$$

$$g_1 \longleftarrow g_2$$



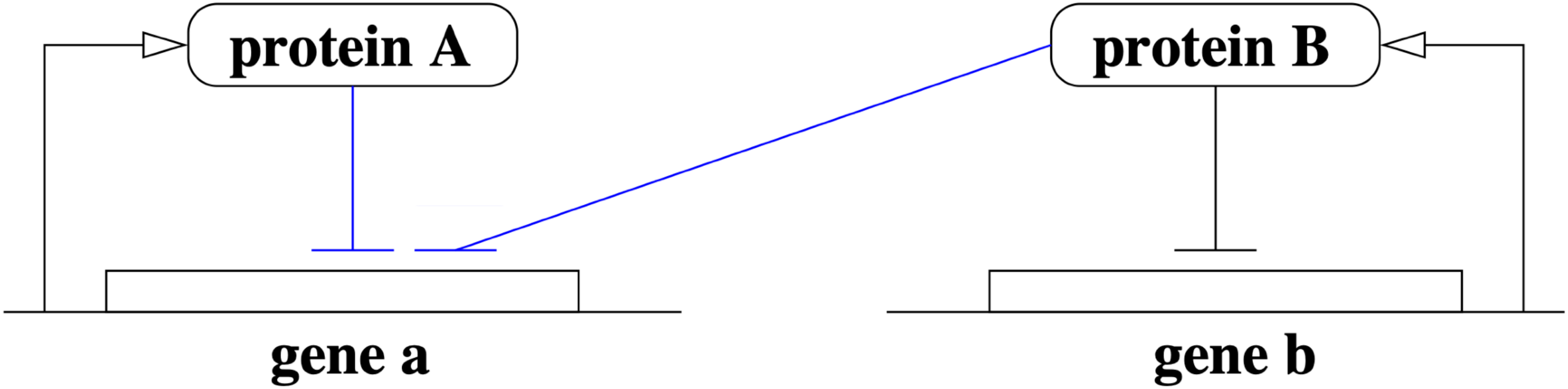
$$K_1(\emptyset) = 1, K_1(\{g_2\}) = 0$$

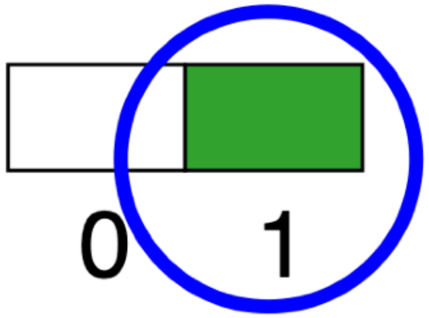
$$g_1 \dashv g_2$$



Model Construction

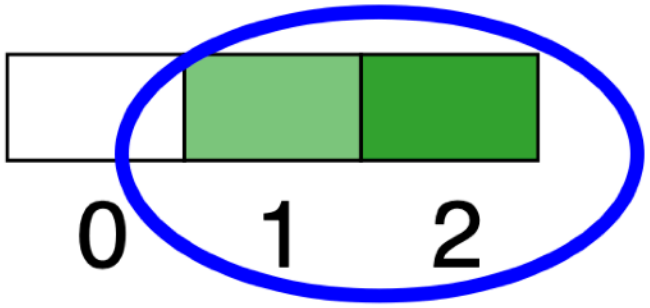
Discrete Models of Influence (Regulatory) Networks





$t_{AA} = 1$

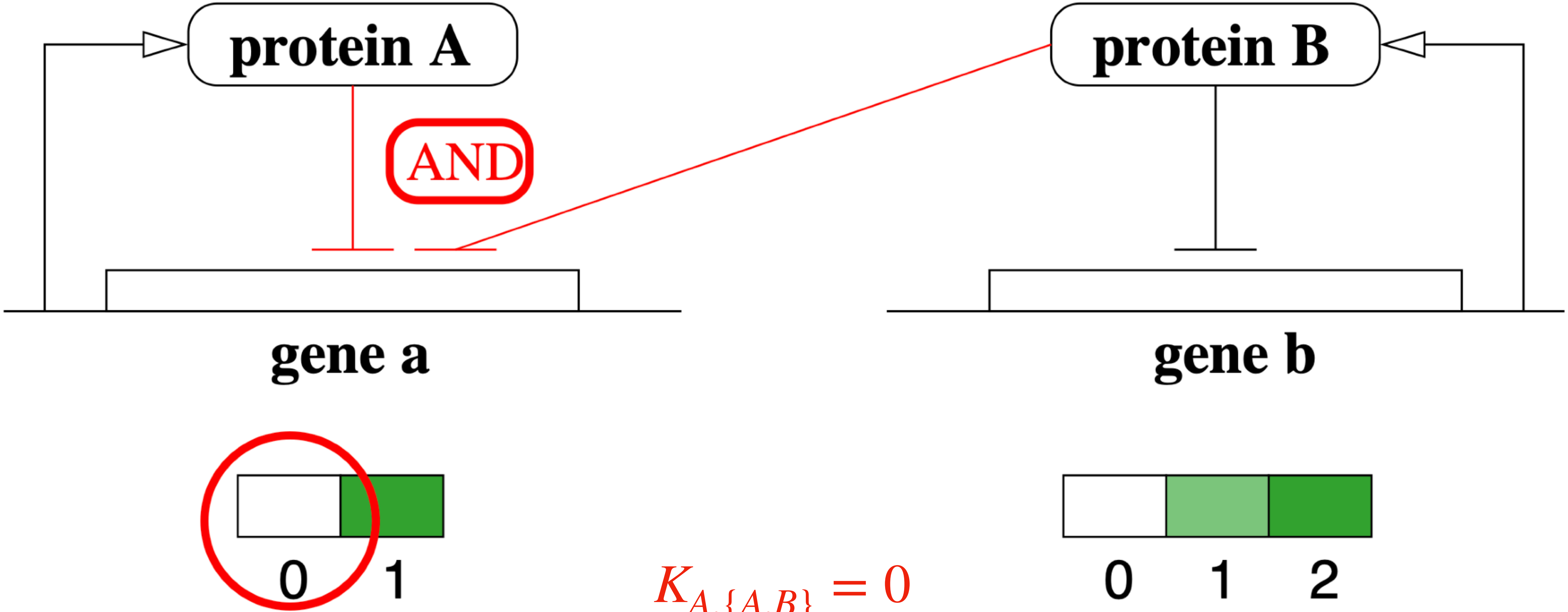
$$K_{A,\{A,B\}} = ?$$
$$K_{A,\{A\}} = ?$$
$$K_{A,\{B\}} = ?$$
$$K_{A,\varnothing} = 1$$



$t_{BA} = 1$

Model Construction

Discrete Models of Influence (Regulatory) Networks



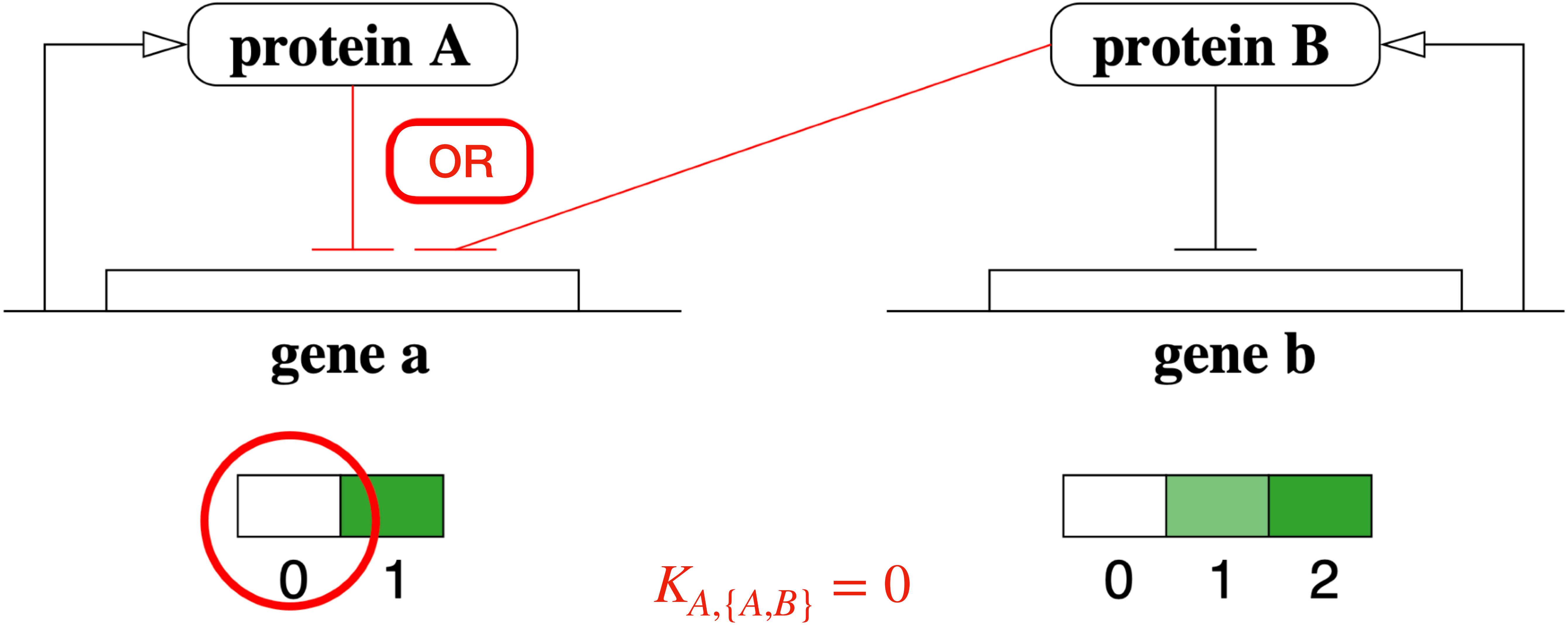
$$
\begin{aligned}
K_{A,\{A,B\}} &= 0\\
K_{A,\{A\}} &= 1\\
K_{A,\{B\}} &= 1\\
K_{A,\varnothing} &= 1
\end{aligned}
$$

$t_{AA} = 1$

$t_{BA} = 1$

Model Construction

Discrete Models of Influence (Regulatory) Networks

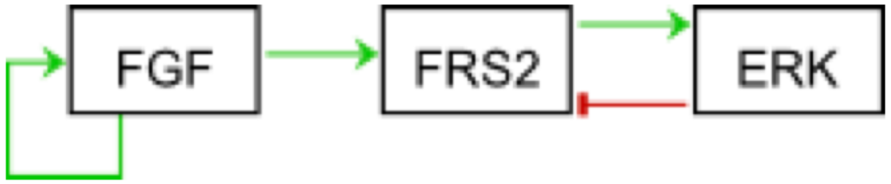


$$K_{A,\{A,B\}} = 0$$
$$K_{A,\{A\}} = 0$$
$$K_{A,\{B\}} = 0$$
$$K_{A,\varnothing} = 1$$

$t_{AA} = 1$

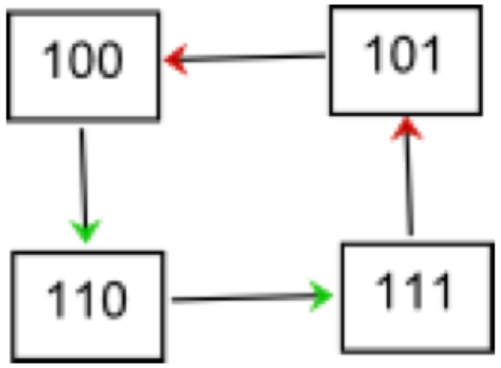
$t_{BA} = 1$

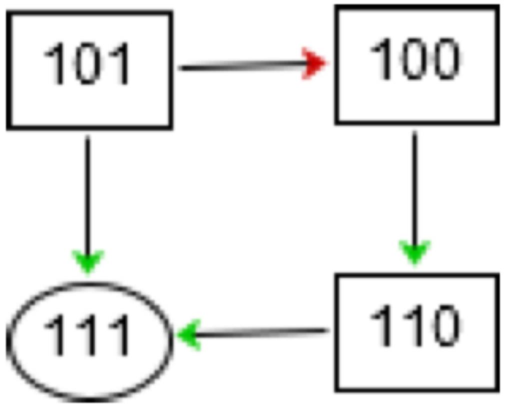
Revealing the Story Behind
Boolean Network Approach



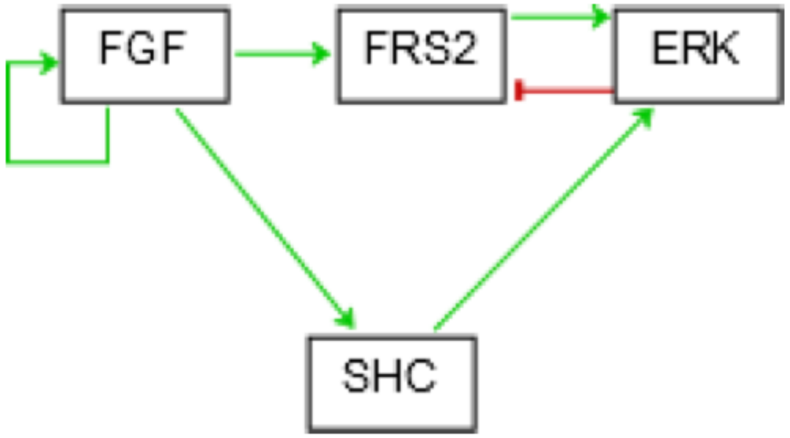
$$K_{FGF}, \emptyset = 0$$
$$K_{FGF}, \{FGF\} = 1$$
$$K_{FRS2}, \emptyset = 0$$
$$K_{FRS2}, \{FGF\} = 1$$
$$K_{FRS2}, \{ERK\} = 0$$
$$K_{FRS2}, \{FGF, ERK\} = 0$$
$$K_{ERK}, \emptyset = 0$$
$$K_{ERK}, \{FRS2\} = 1$$

$$K_{FGF}, \emptyset = 0$$
$$K_{FGF}, \{FGF\} = 1$$
$$K_{FRS2}, \emptyset = 0$$
$$K_{FRS2}, \{FGF\} = 1$$
$$K_{FRS2}, \{ERK\} = 0$$
$$K_{FRS2}, \{FGF, ERK\} = 1$$
$$K_{ERK}, \emptyset = 0$$
$$K_{ERK}, \{FRS2\} = 1$$

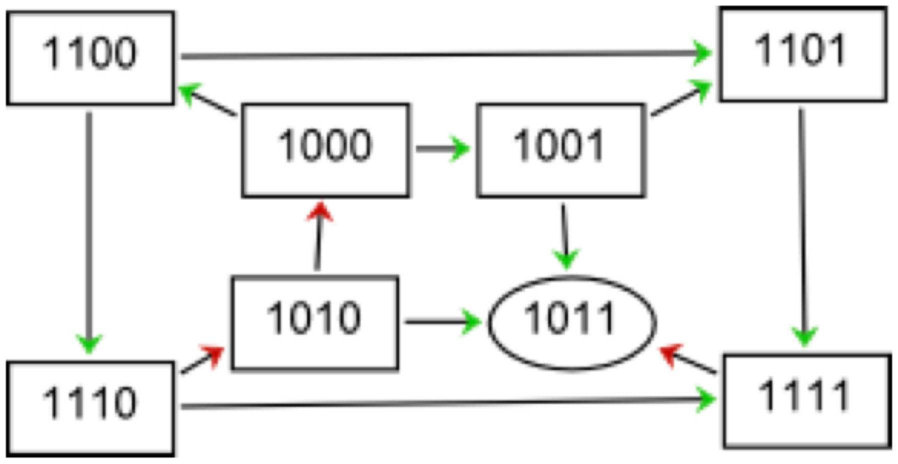




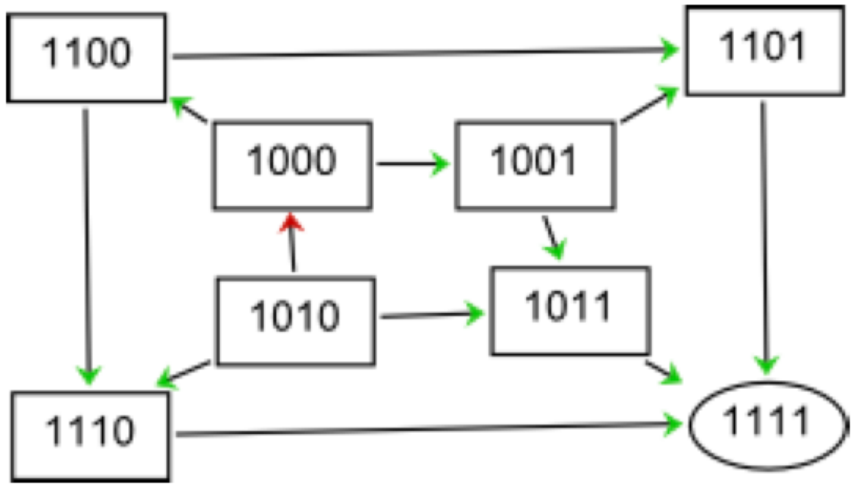
Revealing the Story Behind
Boolean Network Approach



$$
\begin{aligned}
&K_{FRS2}, \{FGF, ERK\} = 0\\
&K_{ERK}, \emptyset = 0\\
&K_{ERK}, \{FRS2\} = 1\\
&K_{ERK}, \{SHC\} = 1\\
&K_{ERK}, \{FRS2, SHC\} = 1\\
&K_{Shc}, \emptyset = 0\\
&K_{Shc}, \{FGF\} = 1
\end{aligned}
$$

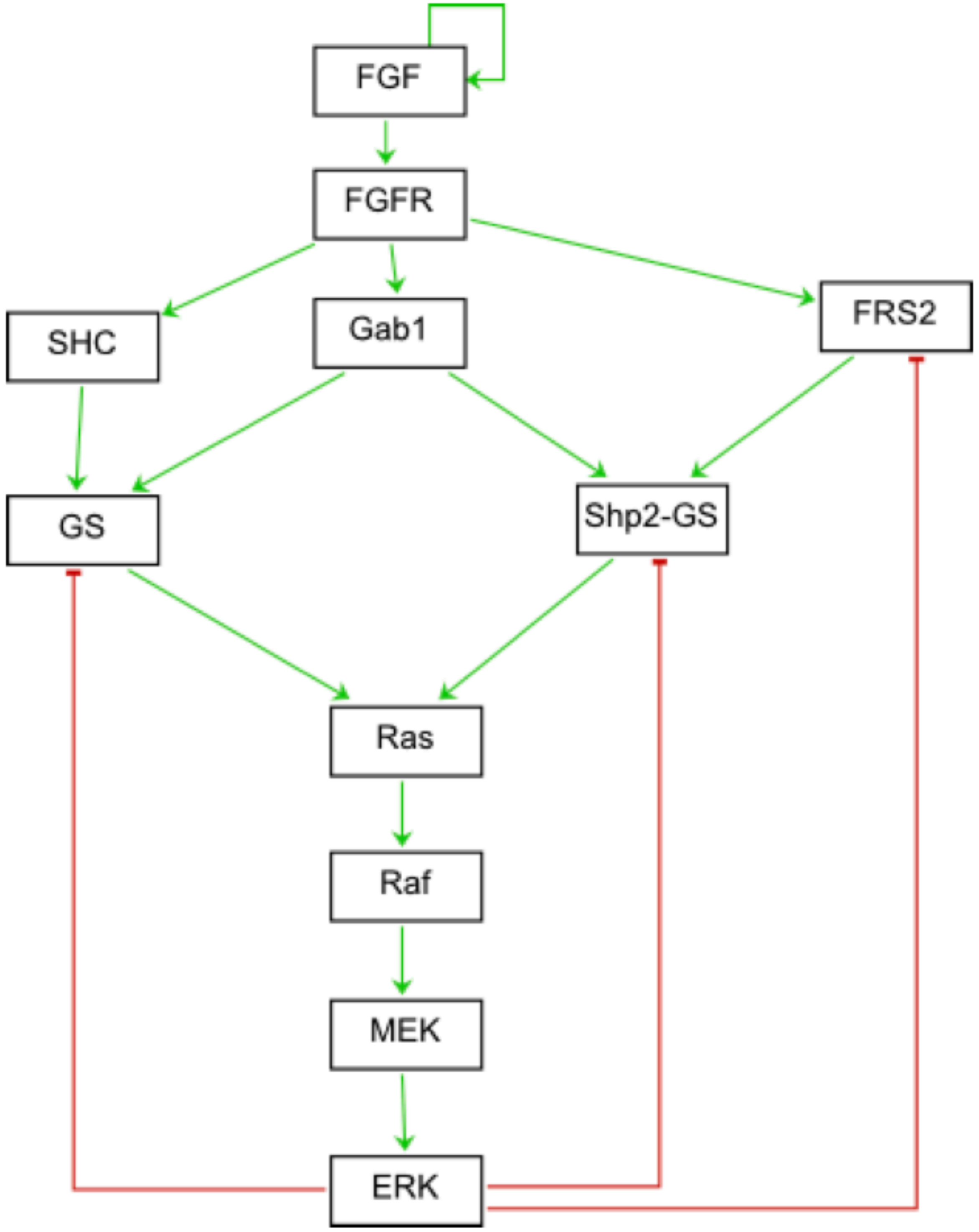


$$
\begin{aligned}
&K_{FRS2}, \{FGF, ERK\} = 1\\
&K_{ERK}, \emptyset = 0\\
&K_{ERK}, \{FRS2\} = 1\\
&K_{ERK}, \{SHC\} = 1\\
&K_{ERK}, \{FRS2, SHC\} = 1\\
&K_{Shc}, \emptyset = 0\\
&K_{Shc}, \{FGF\} = 1
\end{aligned}
$$



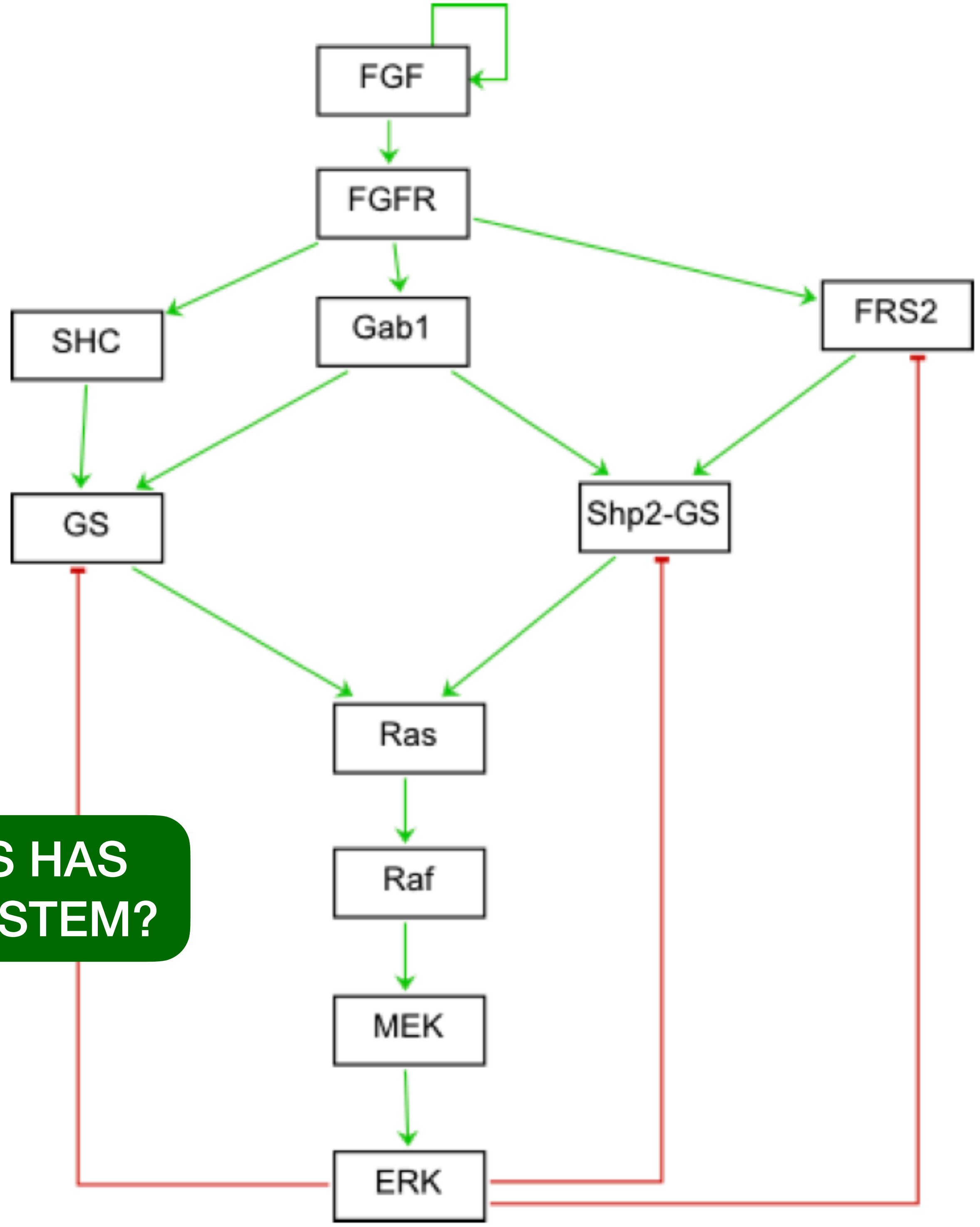
Revealing the Story Behind

Boolean Network Approach



Revealing the Story Behind

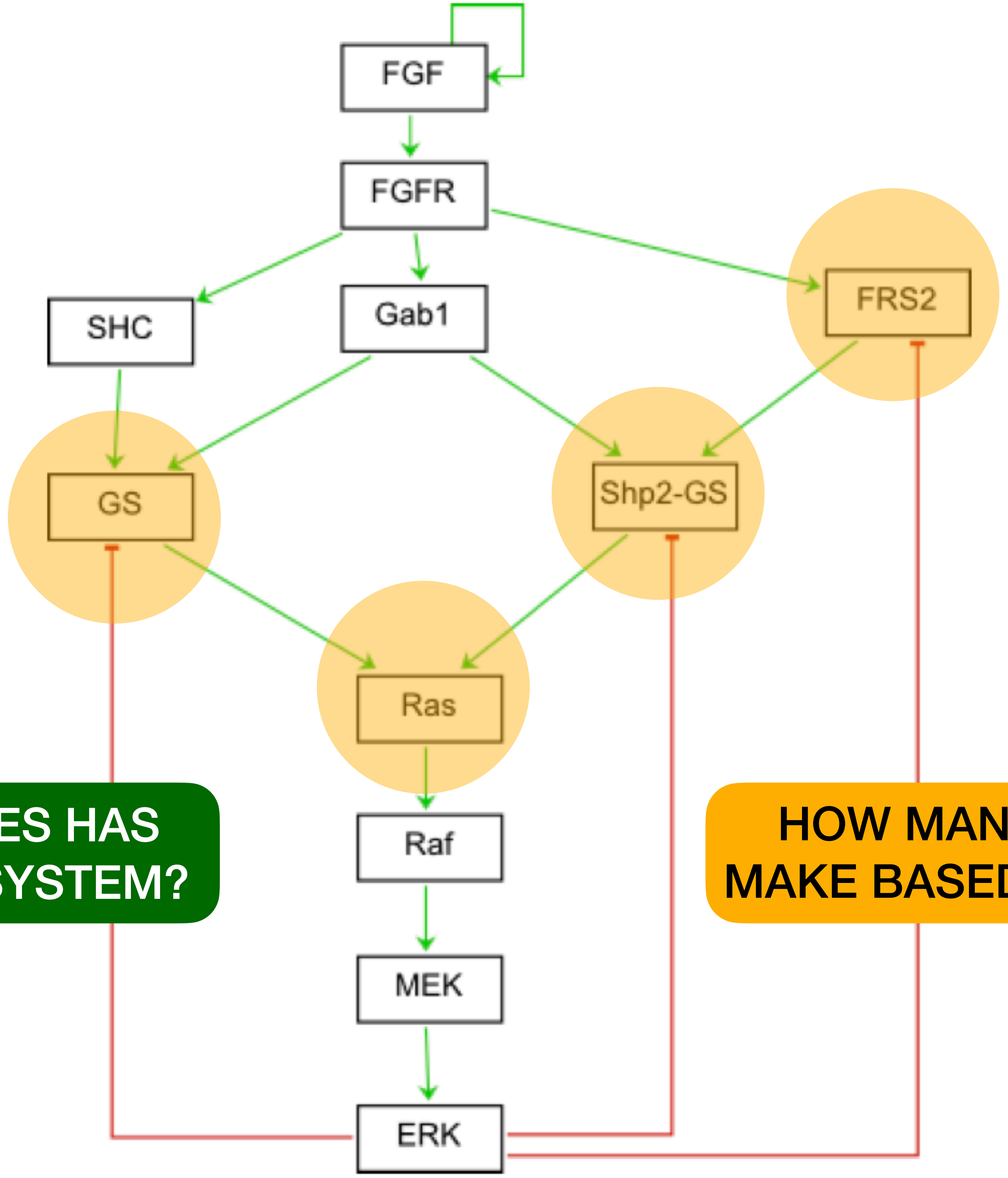
Boolean Network Approach



HOW MANY STATES HAS THE TRANSITION SYSTEM?

Revealing the Story Behind

Boolean Network Approach

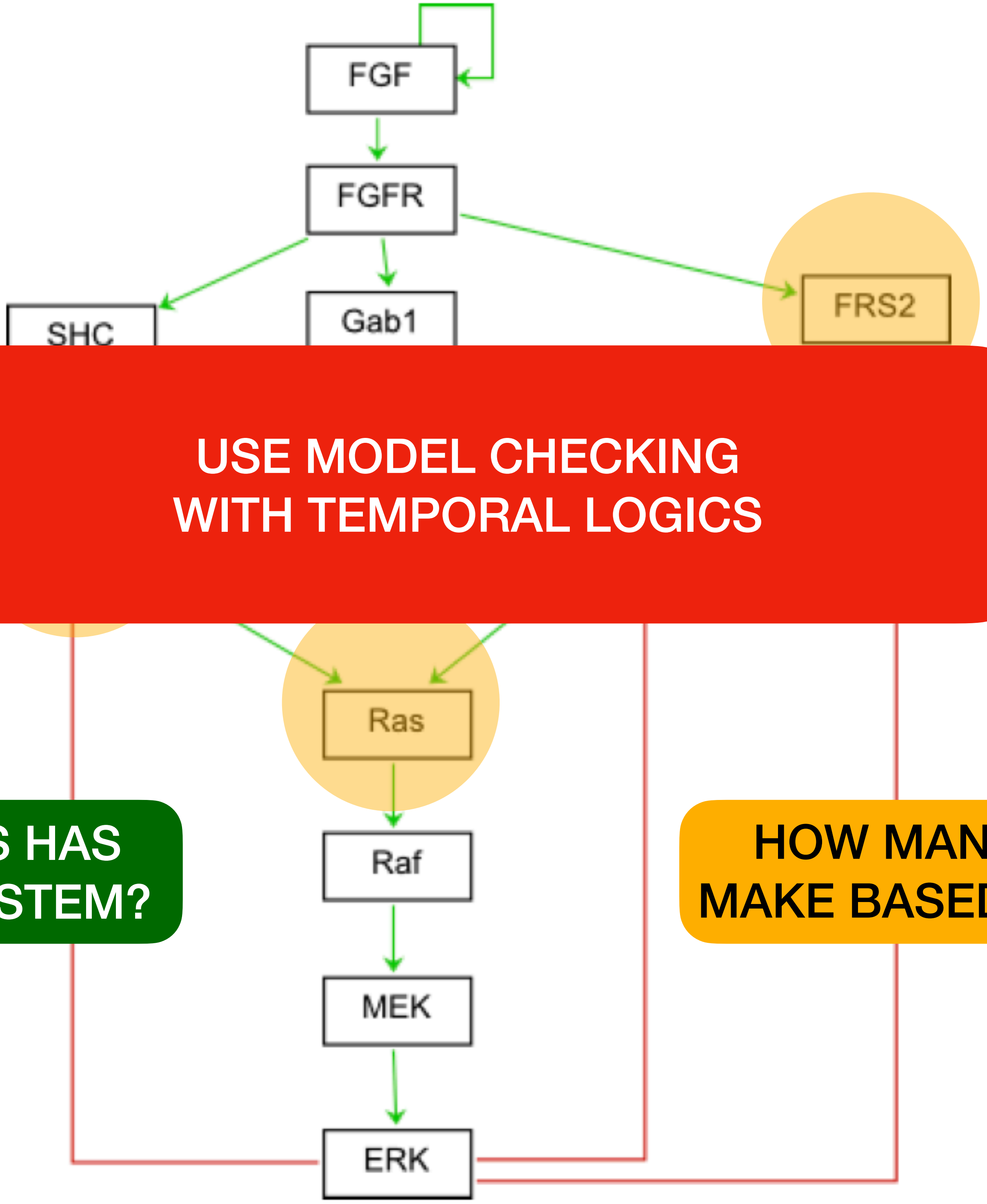


HOW MANY STATES HAS THE TRANSITION SYSTEM?

HOW MANY MODELS WE CAN MAKE BASED ON THIS NETWORK?

Revealing the Story Behind

Boolean Network Approach



USE MODEL CHECKING
WITH TEMPORAL LOGICS

HOW MANY STATES HAS
THE TRANSITION SYSTEM?

HOW MANY MODELS WE CAN
MAKE BASED ON THIS NETWORK?

Kripke Structure

Definition

Let AP be the set of *atomic propositions* (logical expressions over model variables, typical inequalities). *Kripke structure* is the quadruple $K = \langle S, S_0, T, L\rangle$ where:

- S is the finite set of states
- $S_0 \subseteq S$ is the set of inititial states
- $T \subseteq S \times S$ such that $\forall s \in S, \exists s' \in S : \langle s, s'\rangle \in T$
- L is the labeling $L : S \to 2^{AP}$

Kripke structure – properties

- for a state $s \in S$, $L(s)$ represents the set of all atomic propositions satisfied in s
- unfolding of the Kripke structure from any initial state is always an infinite-depth tree
  - maximal paths in the unfolding represent individual (infinite) executions of the Kripke structure

Linear-time Temporal Logic – syntax

Let AP be the set of atomic propositions. Formula φ is *linear temporal logic (LTL) formula* iff the following holds:

- $\varphi = p$ for any $p \in AP$
- If φ_1 and φ_2 LTL formulae then:
  - $\neg\varphi_1$, $\varphi_1 \wedge \varphi_2$ and $\varphi_1 \vee \varphi_2$ are LTL formulae
  - $\mathbf{X}\varphi_1$, $\mathbf{F}\varphi_1$ a $\mathbf{G}\varphi_1$ are LTL formulae
  - $\varphi_1\mathbf{U}\varphi_2$ a $\varphi_1\mathbf{R}\varphi_2$ are LTL formulae

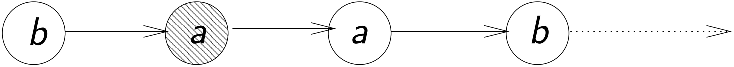
Linear Temporal Logic – semantics

Let $\pi = s_0, s_1, ..., s_i, ...$ be an infinite sequence of states (a path) in a Kripke structure K. For $j > 0$ we denote π^j the suffix $s_j, s_{j+1}, ..., s_i, ...$. Satisfiability relation $\models$ is defined by induction:

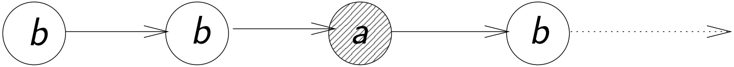
- $\pi \models p$ iff $p \in L(s_0)$
- $\pi \models \neg\varphi$ iff $\pi \not\models \varphi$
- $\pi \models \varphi_1 \wedge \varphi_2$ iff $\pi \models \varphi_1$ and $\pi \models \varphi_2$
- $\pi \models \varphi_1 \vee \varphi_2$ iff $\pi \models \varphi_1$ or $\pi \models \varphi_2$
- $\pi \models \mathbf{X}\varphi$ iff $\pi^1 \models \varphi$
- $\pi \models \mathbf{F}\varphi$ iff $\exists i \geq 0.\, \pi^i \models \varphi$
- $\pi \models \mathbf{G}\varphi$ iff $\forall i \geq 0.\, \pi^i \models \varphi$
- $\pi \models \varphi_1 \mathbf{U} \varphi_2$ iff $\exists j \geq 0.\, \pi^j \models \varphi_2$ and $\forall i < j.\, \pi^i \models \varphi_1$
- $\pi \models \varphi_1 \mathbf{R} \varphi_2$ iff $\forall j \geq 0, \forall 0 \leq i < j.\, \pi^i \not\models \varphi_1 \Rightarrow \pi^j \models \varphi_2$.

Linear Temporal Logic – semantics

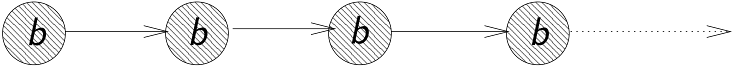
$\mathbf{X}a$



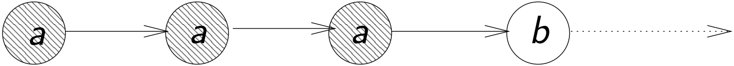
$\mathbf{F}a$



$\mathbf{G}b$



$a\mathbf{U}b$



Model checking

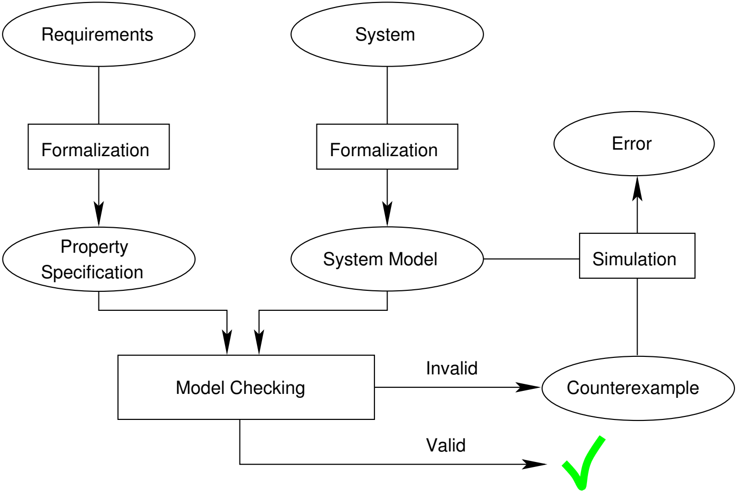
Kripke structure as a model for a formula

Let K be a Kripke structure. A formula φ is satisfied by K, $K \models \varphi$ iff for each execution $\pi = s_0, \ldots$ such that $s_0 \in S_0$ it holds $\pi \models \varphi$.

Model Checking Problem

Model checking problem is to deside for a given Kripke structure K and a temporal property Φ the problem $K \models \Phi$.
If the result is negative, a path π such that $\pi \not\models \Phi$ is returned (a so-called *counterexample*).

Model-Checking Overview



Temporal Properties

Transient vs. Sustained Dynamics

Sustained Dynamics

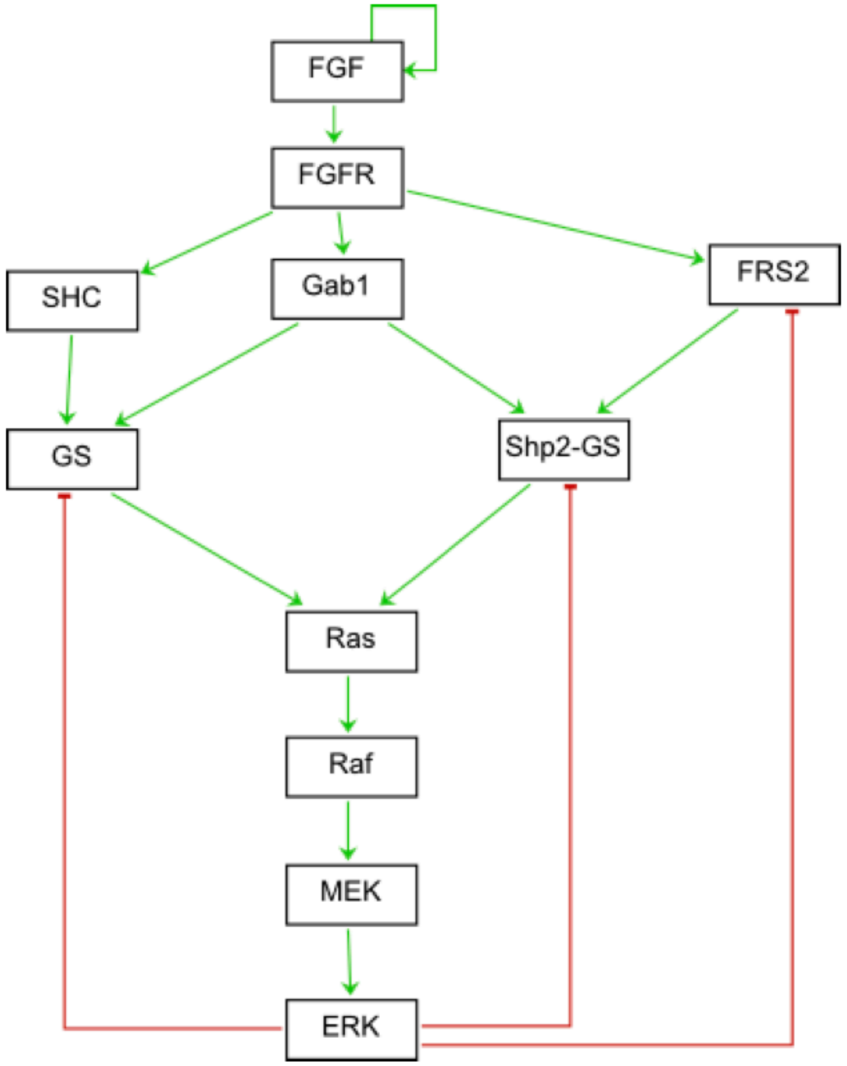
- $\mathbf{FG}(ERK > 0)$
- $\mathbf{FG}(FRS2 > 0)$
- $\mathbf{FG}(ERK > 0) \wedge \mathbf{FG}(FRS2 > 0)$

Transient Dynamics

- $\mathbf{F}(ERK > 0) \wedge \mathbf{GF}(ERK < 1)$
- $\mathbf{F}(ERK > 0) \wedge \mathbf{GF}(ERK < 1)$

Revealing the Story Behind Growth Factor Signalling

Boolean Network Approach



$K_{Shp2-GS}, \{FRS2\} = 1$

$K_{Shp2-GS}, \{Gab1\} = 1$

$K_{Shp2-GS}, \{FRS2, Gab1\} = 1$

$K_{Shp2-GS}, \{ERK, FRS2\} = 1$

$K_{Shp2-GS}, \{ERK, Gab1\} = 1$

$K_{Shp2-GS}, \{ERK, FRS2, Gab1\} = 1$

$K_{Shp2-GS}, \{ERK\} = 0$

$K_{g_i}, \emptyset = 0$

$K_{FGFR}, \{FGF\} = 1$

$K_{SHC}, \{FGFR\} = 1$

$K_{Gab1}, \{FGFR\} = 1$

$K_{FRS2}, \{FGFR\} = 1$

$K_{FRS2}, \{FGFR, ERK\} = 1$

$K_{FRS2}, \{ERK\} = 0$

$K_{Ras}, \{GS\} = 1$

$K_{Ras}, \{Shp2 - GS\} = 1$

$K_{Ras}, \{GS, Shp2 - GS\} = 1$

$K_{Raf}, \{Ras\} = 1$

$K_{MEK}, \{Raf\} = 1$

$K_{ERK}, \{MEK\} = 1$

$K_{GS}, \{SHC\} = 1$

$K_{GS}, \{Gab1\} = 1$

$K_{GS}, \{SHC, Gab1\} = 1$

$K_{GS}, \{ERK, Gab1\} = 0$

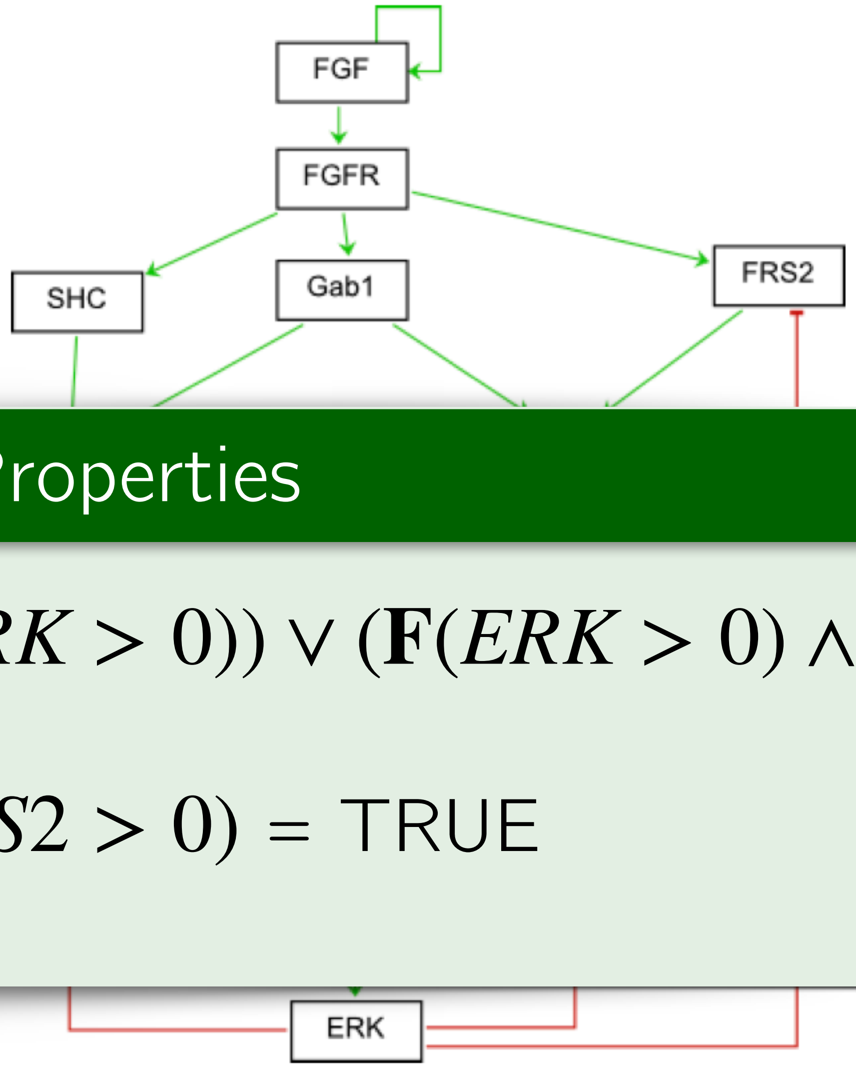
$K_{GS}, \{ERK, SHC\} = 0$

$K_{GS}, \{ERK, SHC, Gab1\} = 0$

$K_{GS}, \{ERK\} = 0$

Revealing the Story Behind Growth Factor Signalling

Boolean Network Approach



$K_{g_i}, \emptyset = 0$
$K_{FGFR}, \{FGF\} = 1$
$K_{SHC}, \{FGFR\} = 1$
$K_{Gab1}, \{FGFR\} = 1$
$K_{FRS2}, \{FGFR\} = 1$

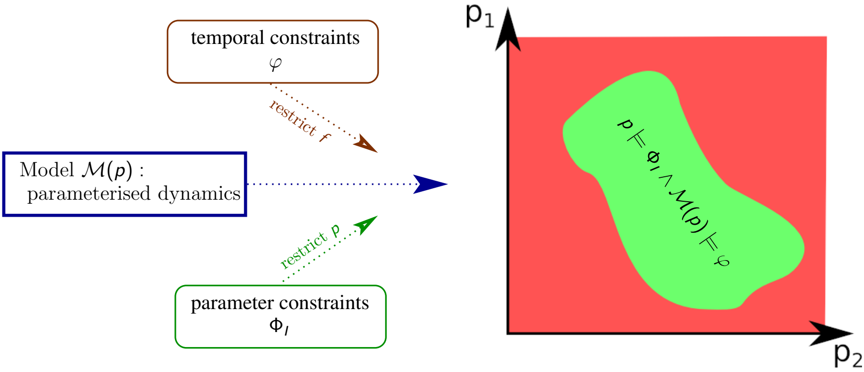
Satisfied Properties

- $(\mathbf{FG}(ERK > 0)) \vee (\mathbf{F}(ERK > 0) \wedge \mathbf{GF}(ERK < 1))$ = TRUE
- $\mathbf{FG}(FRS2 > 0)$ = TRUE

$K_{Shp2-GS}, \{FRS2\} = 1$
$K_{Shp2-GS}, \{Gab1\} = 1$
$K_{Shp2-GS}, \{FRS2, Gab1\} = 1$
$K_{Shp2-GS}, \{ERK, FRS2\} = 1$
$K_{Shp2-GS}, \{ERK, Gab1\} = 1$
$K_{Shp2-GS}, \{ERK, FRS2, Gab1\} = 1$
$K_{Shp2-GS}, \{ERK\} = 0$

$K_{GS}, \{SHC\} = 1$
$K_{GS}, \{Gab1\} = 1$
$K_{GS}, \{SHC, Gab1\} = 1$
$K_{GS}, \{ERK, Gab1\} = 0$
$K_{GS}, \{ERK, SHC\} = 0$
$K_{GS}, \{ERK, SHC, Gab1\} = 0$
$K_{GS}, \{ERK\} = 0$

From Model Checking to Parameter Synthesis



temporal constraints φ

Model $\mathcal{M}(p)$: parameterised dynamics

parameter constraints Φ_I

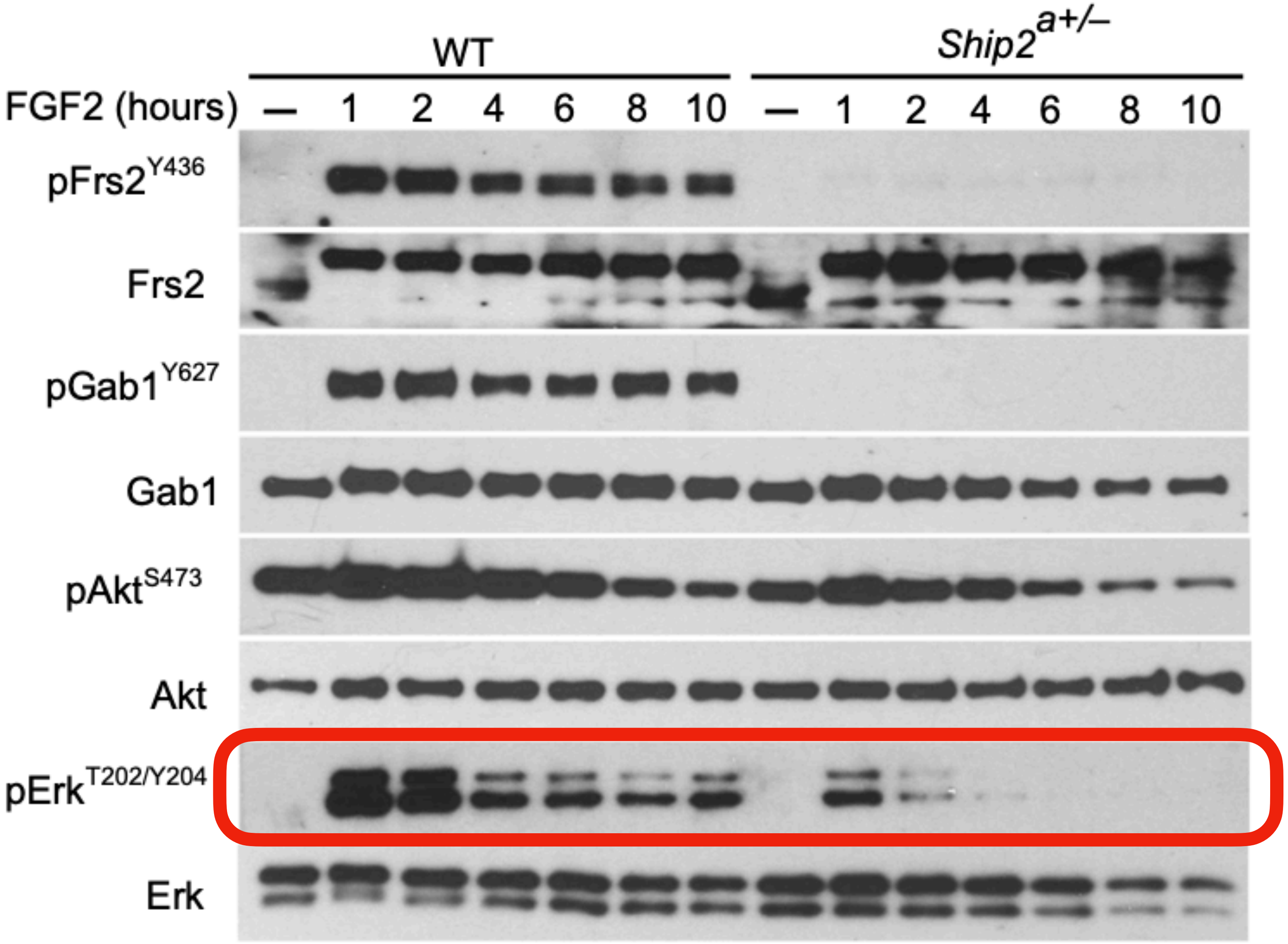
Parameter Synthesis Problem

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 $p \in P$.

Tools

- GINsim, `http://ginsim.org`
  non-parameterised, edit, generate STG, graph-based analysis
- CellCollective, `https://cellcollective.org`
  non-parameterised, simulation, annotation, repository
- TREMPPI, `http://tremppi.fi.muni.cz`
  parameterised, LTL parameter synthesis, model ranking
- AEON,
  `https://sybila.fi.muni.cz/tools_html/aeon.html`
  parameterised, attractor analysis, knowledge inference
- many others...

New Observations — The Role of SHIP2 Protein

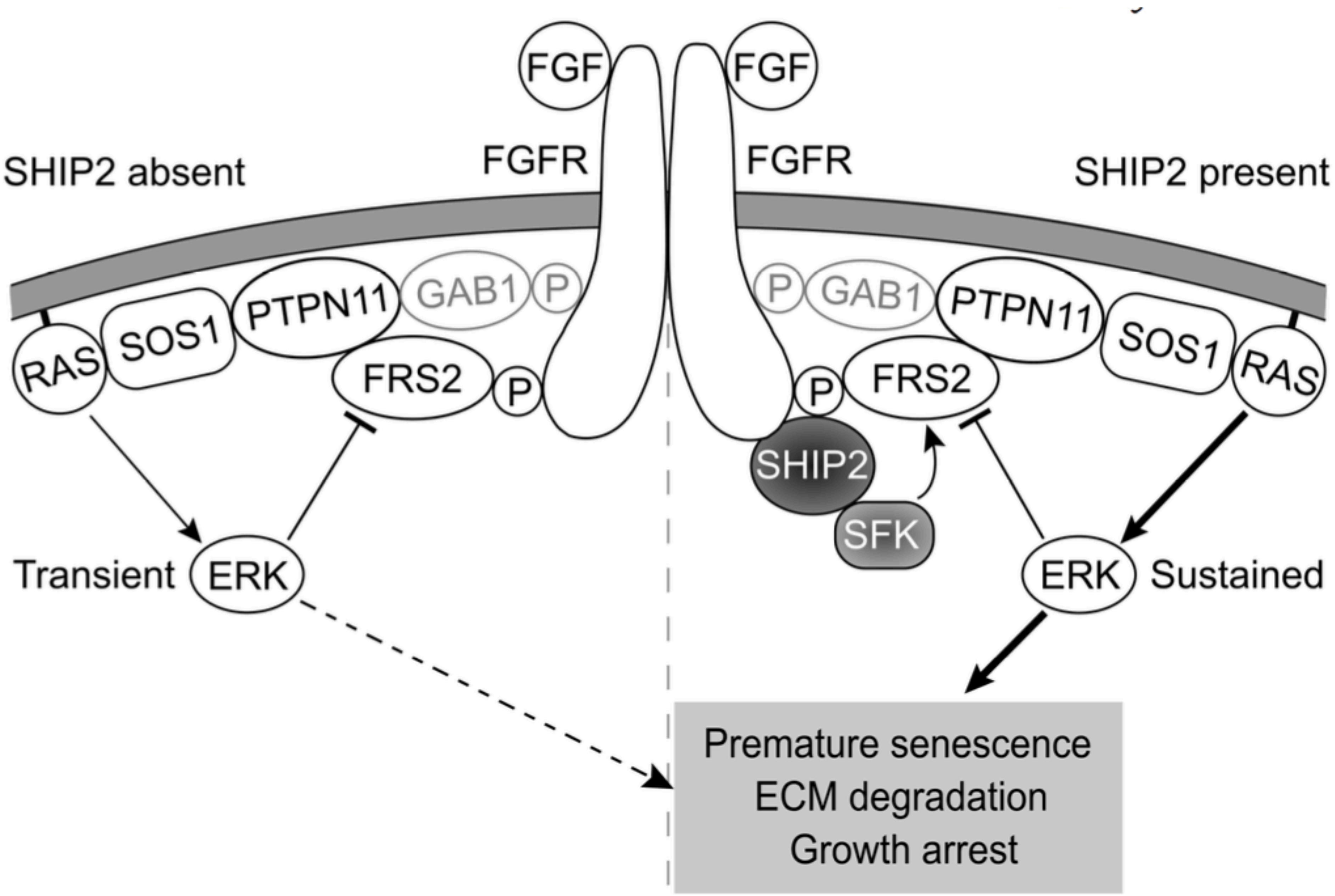


Signaling Dynamics with and without SHIP2

- wildtype includes the inositol phosphatase SHIP2
- when SHIP2 is removed, transient dynamics of pERK is obtained

Fafilek et al. *Science Signaling* (2018)

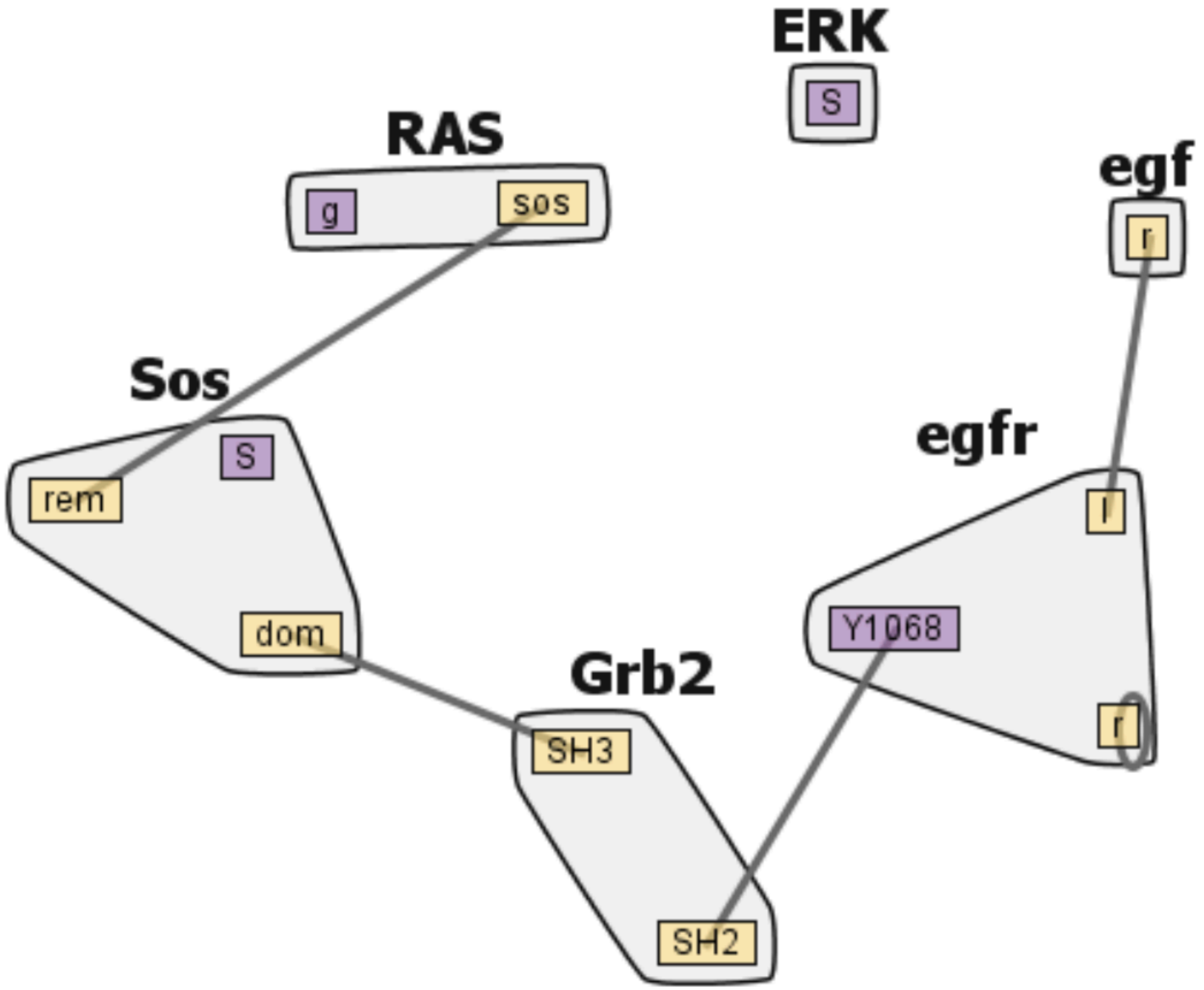
Role of SHIP2 Protein — The Hypothesis



Fafilek et al. *Science Signaling* (2018)

Revealing the Story Behind Growth Factor Signalling

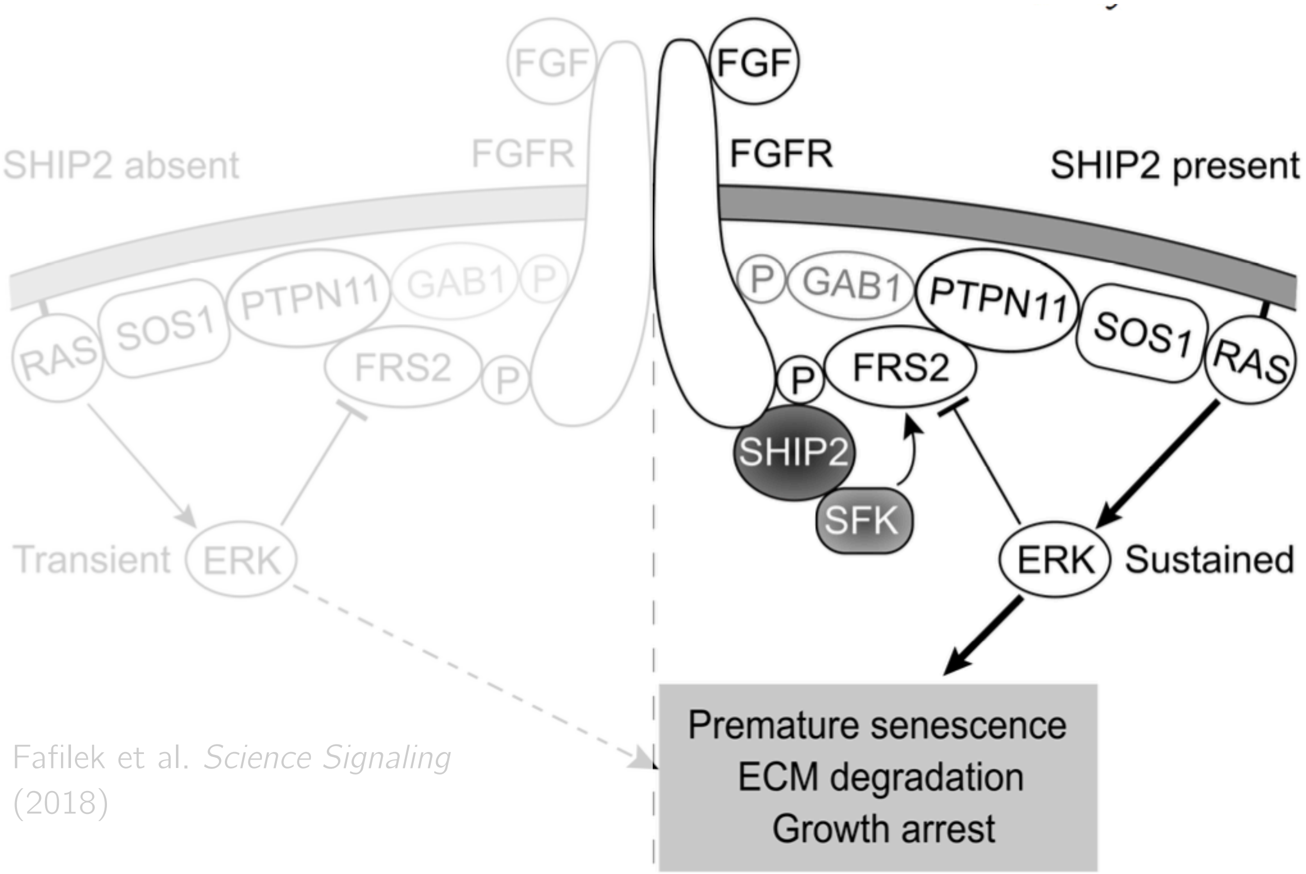
Rule-Based Approach



- modelling the exact binding and modifications of the proteins
- BNGL rule-based language (BioNetGen and RuleBender tools)

Revealing the Story Behind Growth Factor Signalling

Rule-Based Approach



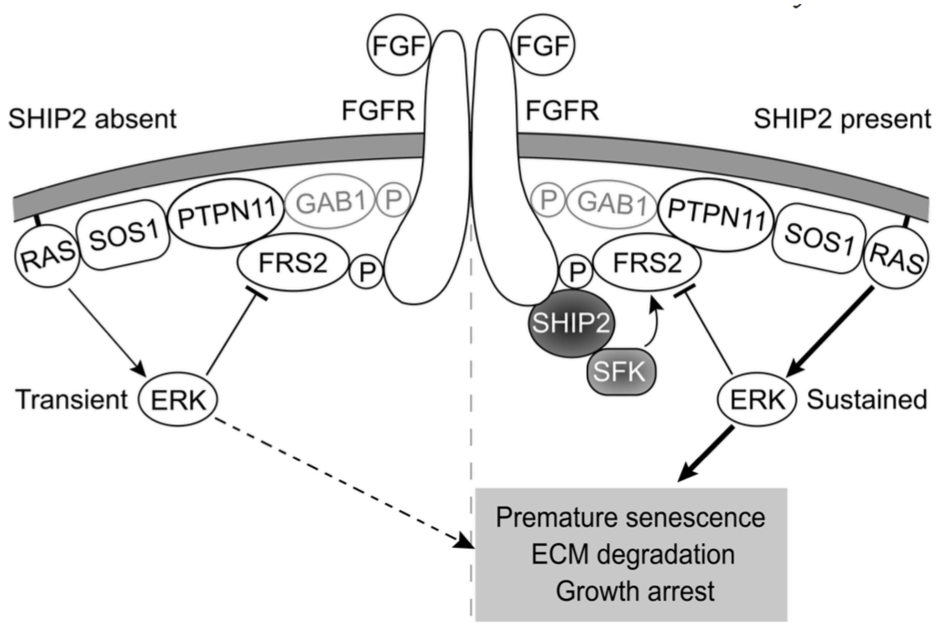
Fafilek et al. *Science Signaling* (2018)

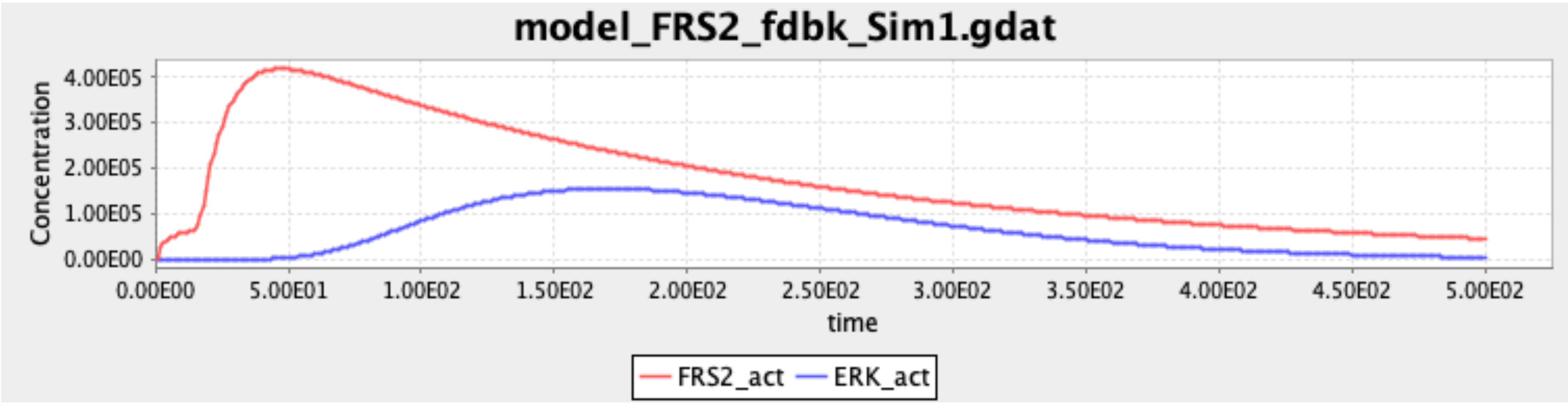
```
begin molecule types
  FGF(R)
  FGFR(L,R,Y1~u~p,Y2~u~p)
  FRS2(pR,Y1~u~p,Th1~u~p,S)
  GS(SH2,REM)

  SHIP(Y~u~p)
  SFK(Y~u~p)

  RAS(sos,g~GDP~GTP)
  RasGAP(ras)
  RAF(S~A~I~P)
  MEK(T292~U~P,S~U~P~PP)
  ERK(S~U~P~PP)
end molecule types
```

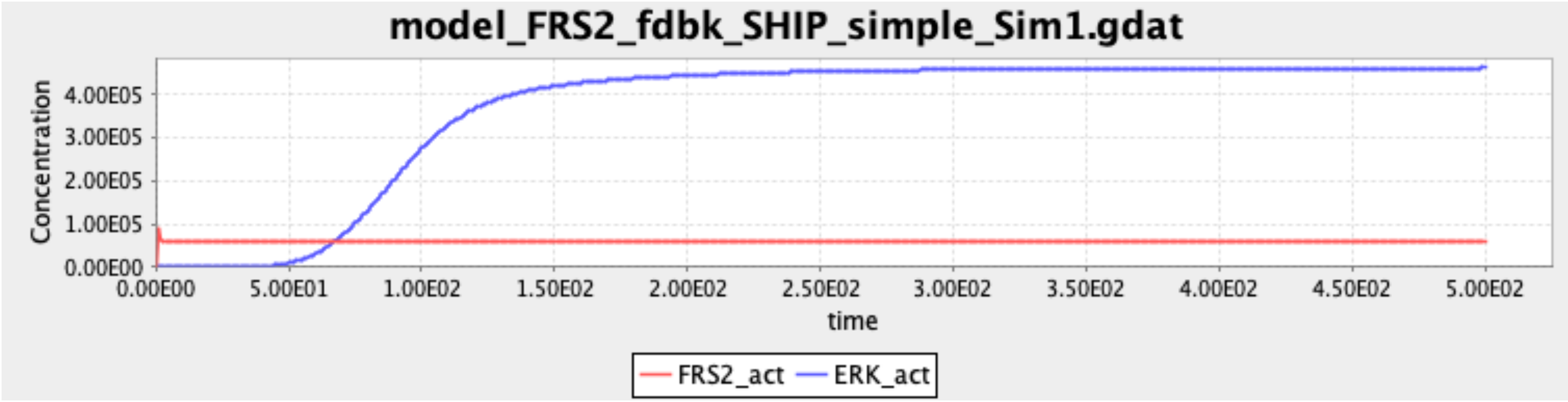
```
begin reaction rules
# Ligand-receptor binding (ligand-monomer)
  FGFR(L,R)   + FGF(R) <-> FGFR(L!1,R).FGF(R!1)   kp1, km1

# Receptor-dimerisation
  FGFR(L!+,R) + FGFR(L!+,R) <-> FGFR(L!+,R!3).FGFR(L!+,R!3)  kp2, km2

# FRS2 phosphorylation by FGFR
  FGFR(R!+,Y1~p!1).FRS2(pR!1,Y1~u,Th1~u) -> FGFR(R!+,Y1~p!1).FRS2(pR!1,Y1~p,Th1~u)  kp14
  FRS2(pR!+,Y1~p)  ->  FRS2(pR!+,Y1~u)  km14

#SFK-mediated FRS2 phosphorylation (hypothetised)
  FRS2(Y1~u) + SFK(Y~p) -> FRS2(Y1~p) + SFK(Y~p)  1000*kp14
...
```

Revealing the Story Behind Growth Factor Signalling

Rule-Based Approach





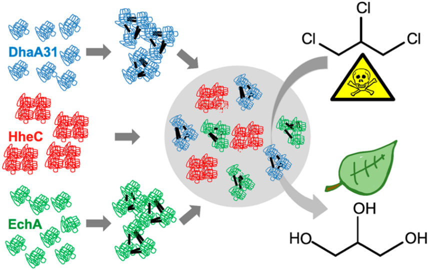


Outline

1. Introduction and Motivation
2. Methodology
   - Biological Networks
   - Modelling Problems
3. Story I: Signalling Pathways of Fibroblast Growth Factors
4. Story II: Synthetic Biology: Trichlorpropane Degradation

Biodegradation of Trichloropropane in *E. coli*

$$\textbf{TCP} \xrightarrow{\text{DhaA}} \textbf{DCP} \xrightarrow{\text{HheC}} \textbf{ECH} \xrightarrow{\text{EchA}} \textbf{CPD} \xrightarrow{\text{HheC}} \textbf{GDL} \xrightarrow{\text{EchA}} \textbf{GLY}$$



$$\frac{d[TCP]}{dt} = -\frac{k_1 \cdot DhaA \cdot [TCP]}{K_{m,1} + [TCP]}$$

$$\frac{d[DCP]}{dt} = \frac{k_1 \cdot DhaA \cdot \cdot [TCP]}{K_{m,1} + [TCP]} - \frac{k_2 \cdot HheC \cdot [DCP]}{K_{m,2} + [DCP]}$$

$$\frac{d[ECH]}{dt} = \frac{k_2 \cdot HheC \cdot [DCP]}{K_{m,2} + [DCP]} - \frac{k_3 \cdot EchA \cdot [ECH]}{K_{m,3} + [ECH]}$$

$$\frac{d[CPD]}{dt} = \frac{k_3 \cdot EchA \cdot [ECH]}{K_{m,3} + [ECH]} - \frac{k_4 \cdot HheC \cdot [CPD]}{K_{m,4} + [CPD]}$$

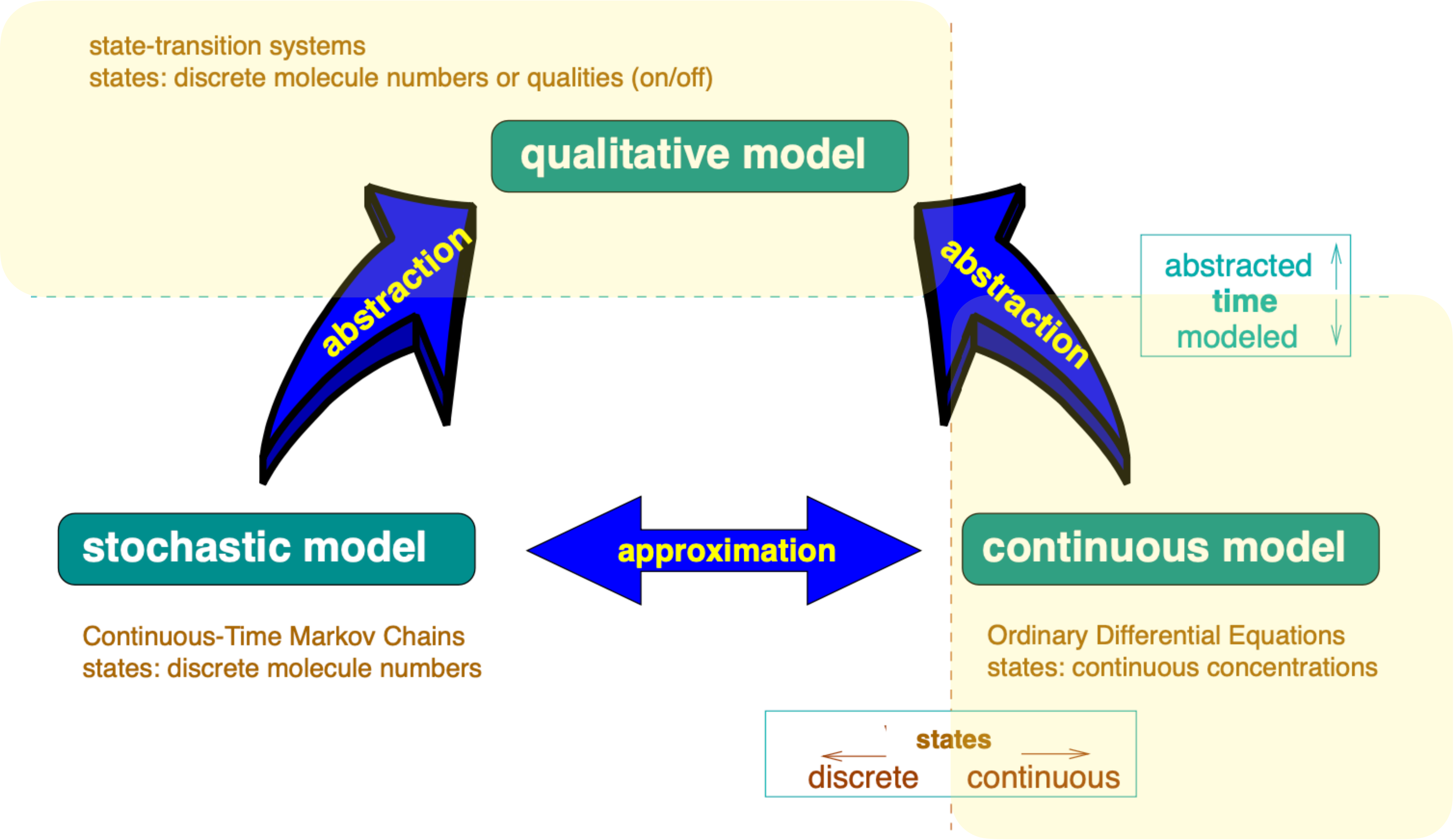
$$\frac{d[GDL]}{dt} = \frac{k_4 \cdot HheC \cdot [CPD]}{K_{m,4} + [CPD]} - \frac{k_5 \cdot HheC \cdot [GDL]}{K_{m,5} + [GDL]}$$

$$\frac{d[GLY]}{dt} = \frac{k_5 \cdot HheC \cdot [GDL]}{K_{m,5} + [GDL]}$$

- biodegradation of toxic substrate and intermediates
- synthetic pathway utilising enzymes from two other bacteria

  Rhodococcus rhodochrous NCIMB 13064; *Agrobacterium radiobacter* AD1
- find optimal enzymes concentration balancing *metabolic burden* and *toxicity*

Modelling Frameworks



state-transition systems
states: discrete molecule numbers or qualities (on/off)

qualitative model

abstracted
time
modeled

stochastic model

Continuous-Time Markov Chains
states: discrete molecule numbers

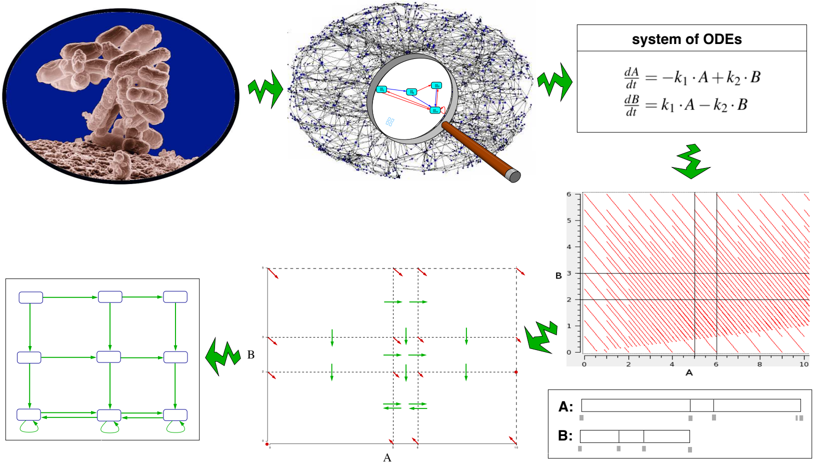
continuous model

Ordinary Differential Equations
states: continuous concentrations

states
discrete continuous

The Approach: Rectangular Abstraction of ODE

From a Continuous System to a Discrete Finite Quotient



system of ODEs

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

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

P. Collins, L. Habets, J.H. van Schuppen, I. Černá, J. Fabriková, and D. Šafránek. Abstraction of Biochemical Reaction Systems on Polytopes. In Proceedings of 18th IFAC World Congress, 2011.

Parameter Synthesis over Rectangular Abstraction

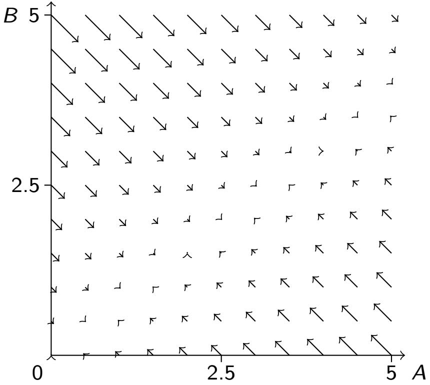
Phase Space Discretisation Leads to Parameter Space Discretisation

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

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

$$k_2 = 0.8$$

$$k_1 = 0.6$$



Parameter Synthesis over Rectangular Abstraction

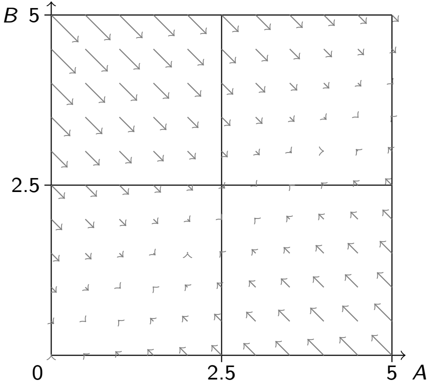
Phase Space Discretisation Leads to Parameter Space Discretisation

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

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

$$k_2 = 0.8$$

$$k_1 = 0.6$$



Parameter Synthesis over Rectangular Abstraction

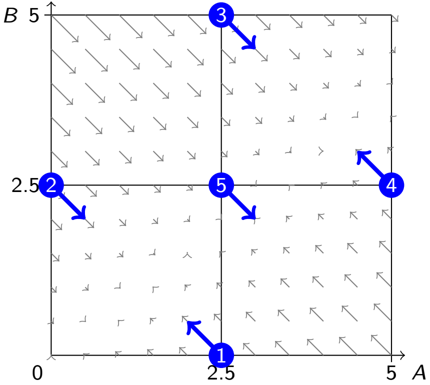
Phase Space Discretisation Leads to Parameter Space Discretisation

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

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

$$k_2 = 0.8$$

$$k_1 = 0.6$$



Parameter Synthesis over Rectangular Abstraction

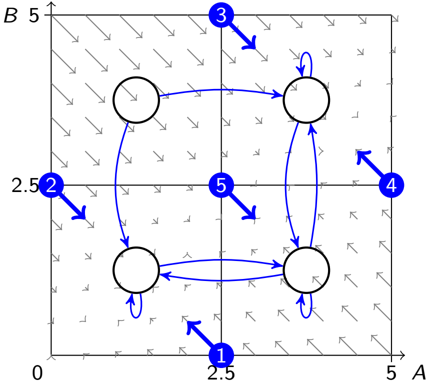
Phase Space Discretisation Leads to Parameter Space Discretisation

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

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

$$k_2 = 0.8$$

$$k_1 = 0.6$$



Parameter Synthesis over Rectangular Abstraction

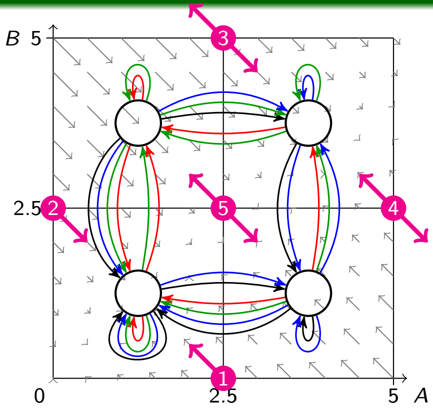
Phase Space Discretisation Leads to Parameter Space Discretisation

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

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

$$k_2 = 0.8$$

$$k_1 = ?$$



	(0,0.4)	(0.4,0.8)	(0.8,1.6)	(1.6,max)
①	↖	↖	↖	↖
②	↘	↘	↘	↘
③	↘	↘	↘	↖
④	↘	↖	↖	↖
⑤	↘	↘	↖	↖

Parameter Synthesis over Rectangular Abstraction

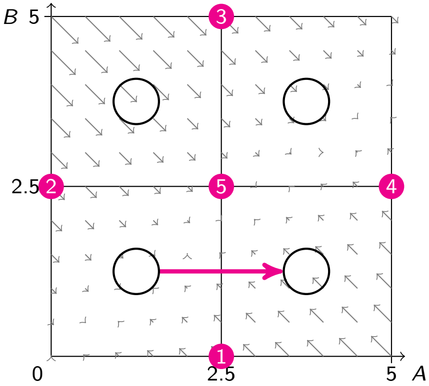
Phase Space Discretisation Leads to Parameter Space Discretisation

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

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

$$k_2 = 0.8$$

$$k_1 = ?$$



	(0,0.4)	(0.4,0.8)	(0.8,1.6)	(1.6,max)
①	↖	↖	↖	↖
②	↘	↘	↘	↘
③	↘	↘	↘	↖
④	↘	↖	↖	↖
⑤	↘	↘	↖	↖

$-2.5 \cdot k_1 > 0$

$-2.5 \cdot k_1 + 2.5 \cdot k_2 > 0$

Parameter Synthesis over Rectangular Abstraction

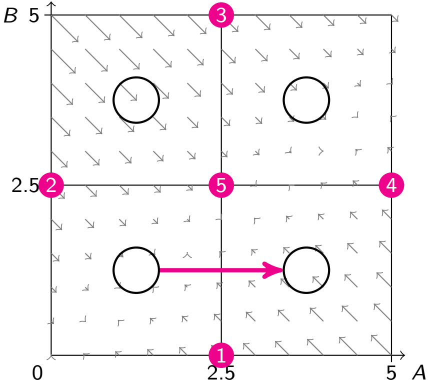
Phase Space Discretisation Leads to Parameter Space Discretisation

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

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

$$k_2 = 0.8$$

$$k_1 = ?$$

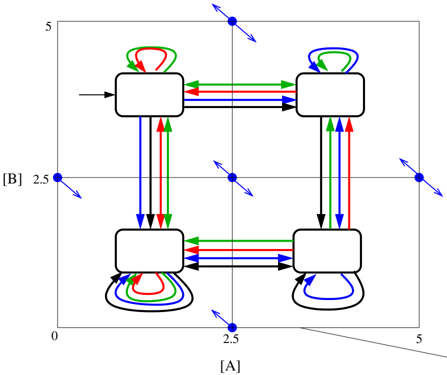


$$\Phi_{\text{state00}\rightarrow\text{state10}} := -2.5 \cdot k_1 > 0 \lor -2.5 \cdot k_1 + 2.5 \cdot k_2 > 0$$

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

Parameter Synthesis over Rectangular Abstraction

parameterized Kripke structure of the model



LTL or (A)CTL specification

CMC

identify states and colors for which the property does/doesn’t hold

YES

parameter intervals where the specification is guaranteed
(some might be missing)

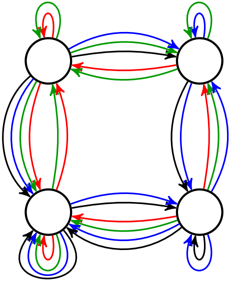
NO

parameter intervals where the specification might be violated

Parameterised Kripke Structures

State Transition Systems with Parameters

Transitions with Parameters (coloured transitions)

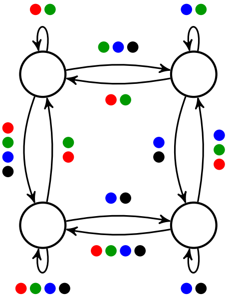


- each parameter valuation represents one Kripke structure
- shared state space, different transition space

Parameterised Kripke Structures

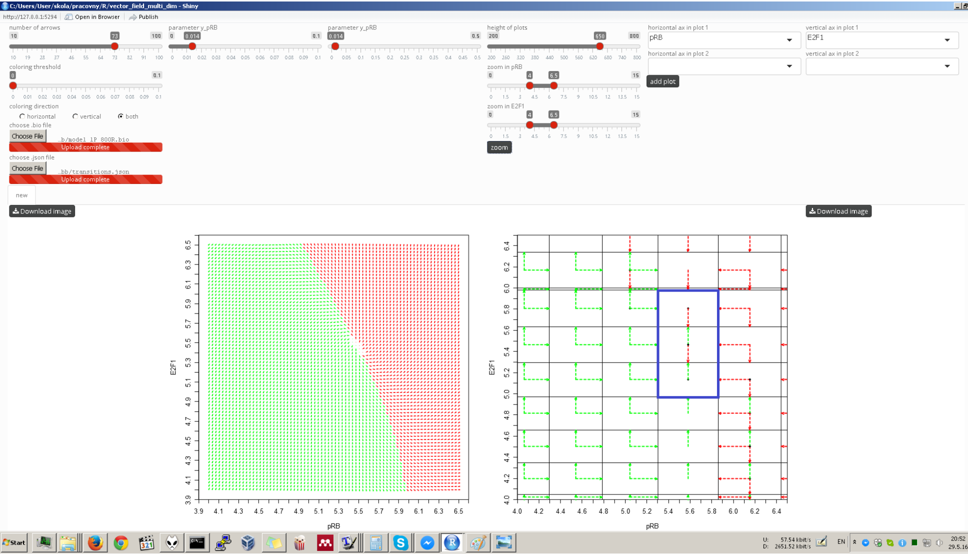
State Transition Systems with Parameters

Transitions with Parameters (coloured transitions)



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

Pithya Tool



C:/Users/User/skola/pracovny/R/vector_field_multi_dim - Shiny

http://127.0.0.1:5294

number of arrows: 73

parameter y_pRB: 0.014

parameter y_pRB: 0.014

height of plots: 650

coloring threshold: 0

coloring direction: ○ horizontal ○ vertical ● both

choose .bio file: ...b/model 1P 800R.bio — Upload complete

choose .json file: ...bb/transitions.json — Upload complete

zoom in pRB: 4 – 6.5

zoom in E2F1: 4 – 6.5

horizontal ax in plot 1: pRB

vertical ax in plot 1: E2F1

horizontal ax in plot 2:

vertical ax in plot 2:

`http://pithya.ics.muni.cz`

[CAV 2017]

Biodegradation of Trichloropropane in *E. coli*

Desired behaviour:

"TCP is finally completely degraded and the concentration of intermediates does not exceed given bounds"

Formally:

$$\varphi_1 = (([TCP] > x)\mathbf{U}(\mathbf{FG}\ [TCP] < y)),$$

$$\varphi_2 = (([GLY] < y)\mathbf{U}(\mathbf{FG}\ [GLY] > x)),$$

$$\varphi_3 = (\mathbf{G}\ [DCP] < v) \wedge (\mathbf{G}\ [GDL] < w),$$

$$\varphi = (\varphi_1 \wedge \varphi_2 \wedge \varphi_3),$$

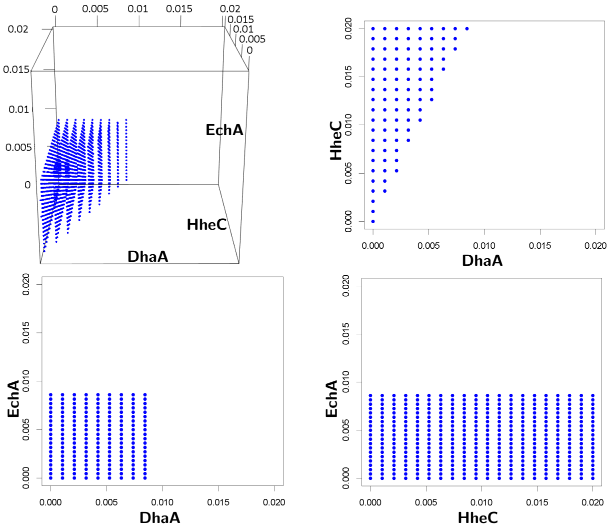
where x, y, v and w are estimated values making an instance of this property:

- $x = 1.9$ (according to authors[1] using the value 2 mM),
- $y = 0.01$ (obviously, cannot be zero),
- $v \in \{0.5, 0.3, 0.1\}$ (variations based on experimental data observation)
- $w \in \{0.5, 0.25, 0.1\}$ (variations based on experimental data observation)

---

[1] Kurumbang et al., ACS Synthetic Biology, 2013

Biodegradation of Trichloropropane in *E. coli*

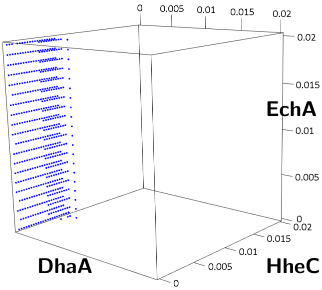


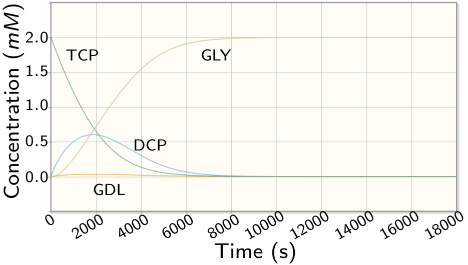
A sample of the resulting parameter space for a particular initial state:
TCP $\in [1.9, 1.9586]$, DCP $\in [0.448898, 0.5]$, GDL $\in [0.0, 0.0669138]$, GLY $\in [0.0, 0.01]$

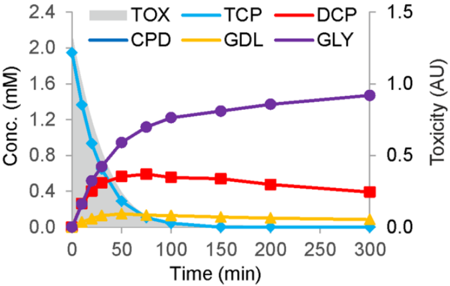
Dotted area corresponds to φ ($v = 0.5$, $w = 0.25$).

Biodegradation of Trichloropropane in *E. coli*

Preliminary Biological Validation

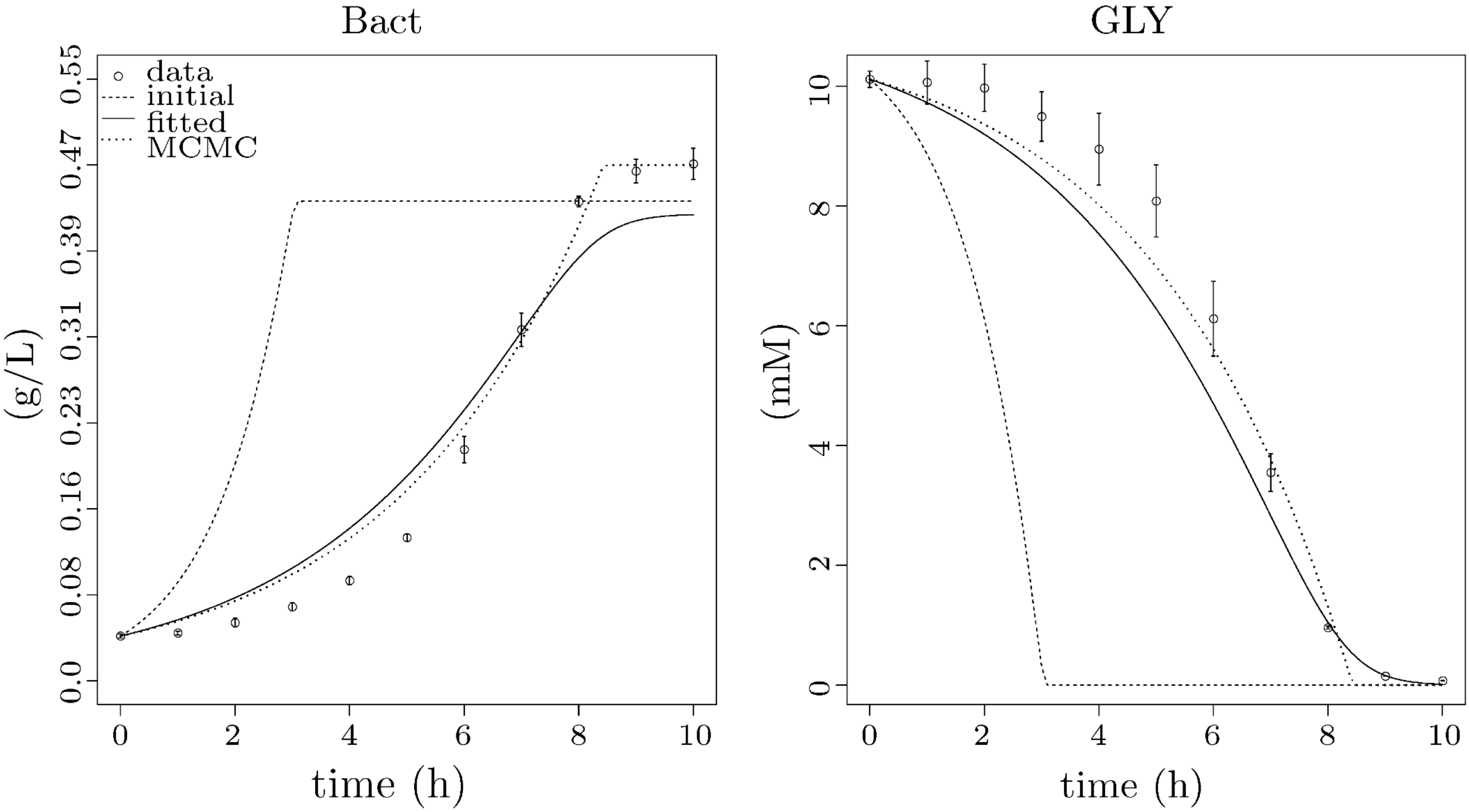






Biodegradation of Trichloropropane in *E. coli*

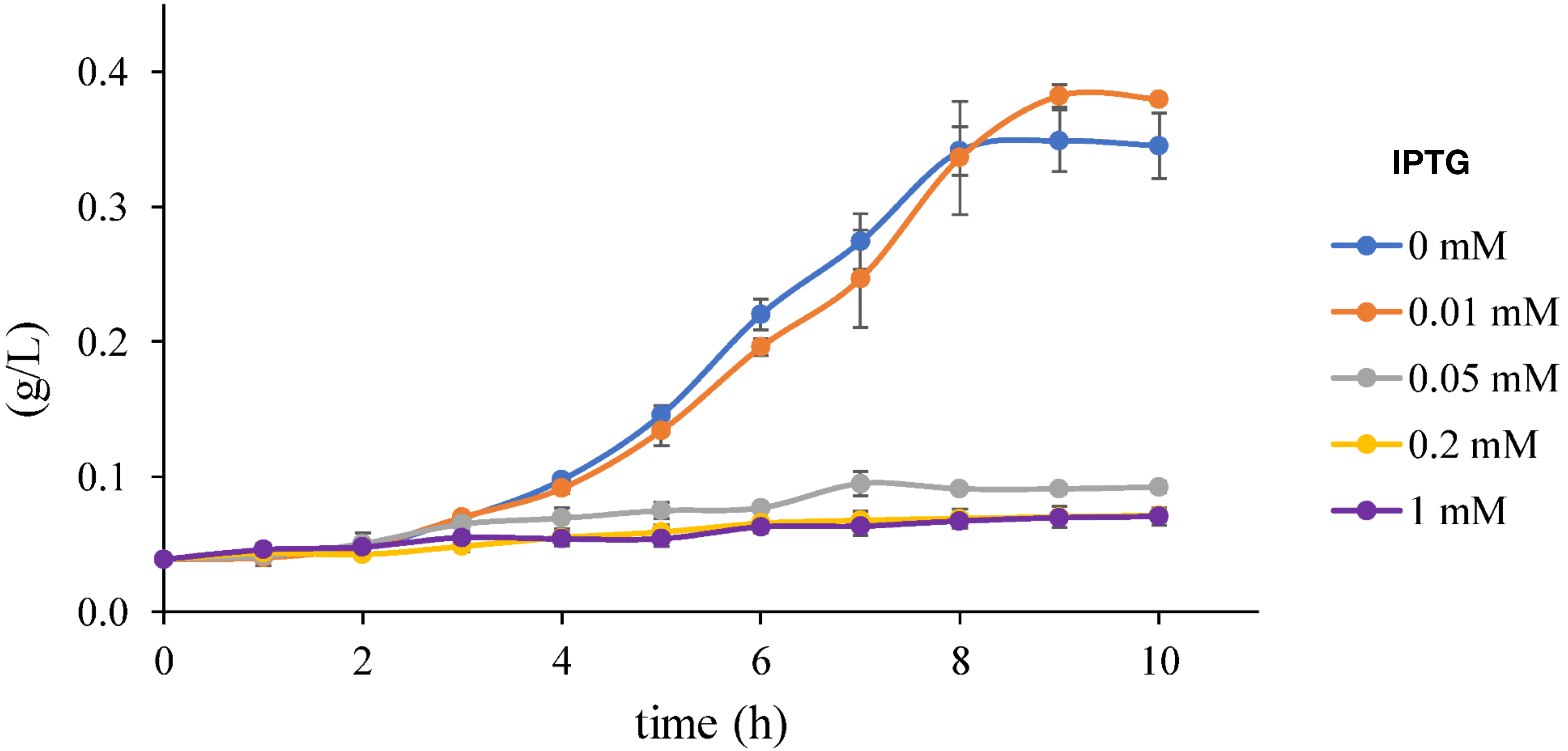
Extending the Model with Population Growth



Demko et al. *Microorganisms* (2019)

Biodegradation of Trichloropropane in *E. coli*

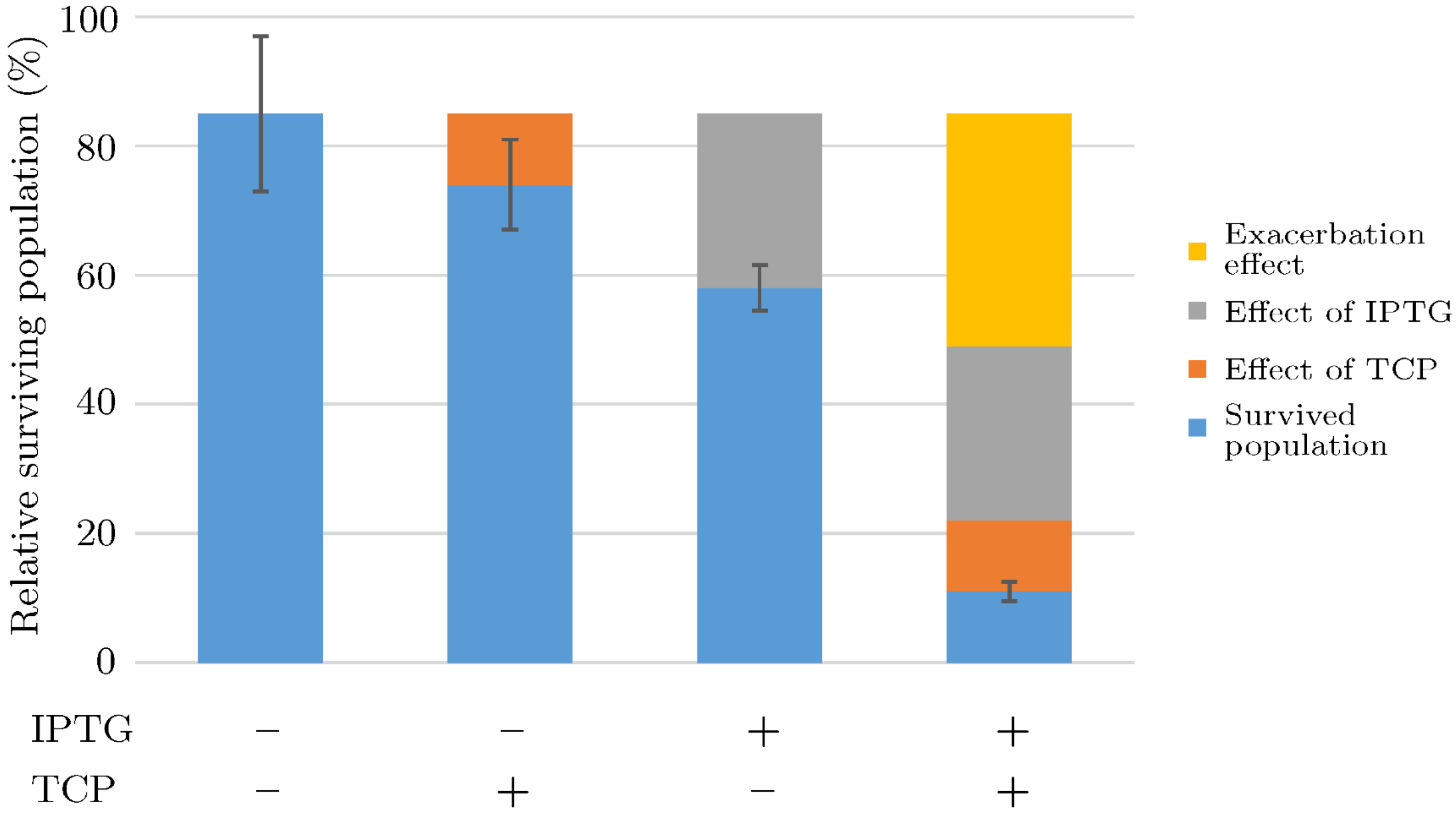
Extending the Model with Population Growth



Demko et al. *Microorganisms* (2019)

Biodegradation of Trichloropropane in *E. coli*

Extending the Model with Population Growth



Demko et al. *Microorganisms* (2019)

Biodegradation of Trichloropropane in *E. coli*

Extending the Model with Population Growth

$$\frac{d[\mathrm{Bact}]}{dt} = + \boxed{\text{growth on GLY}} - \boxed{\text{toxicity by TCP}} - \boxed{\text{toxicity on ECH}} - \boxed{\text{toxicity by GDL}} - \boxed{\text{toxicity by CPD}}$$

$$\frac{d[\mathrm{GLY}]}{dt} = + \boxed{\text{GDL into GLY}} - \boxed{\text{utilisation of GLY}}$$

$$\frac{d[\mathrm{TCP}]}{dt} = - \boxed{\text{TCP into } (R)\text{-DCP}} - \boxed{\text{TCP into } (S)\text{-DCP}}$$

$$\frac{d[(R)\text{-DCP}]}{dt} = + \boxed{\text{TCP into } (R)\text{-DCP}} - \boxed{(R)\text{-DCP into ECH}}$$

$$\frac{d[(S)\text{-DCP}]}{dt} = + \boxed{\text{TCP into } (S)\text{-DCP}} - \boxed{(S)\text{-DCP into ECH}}$$

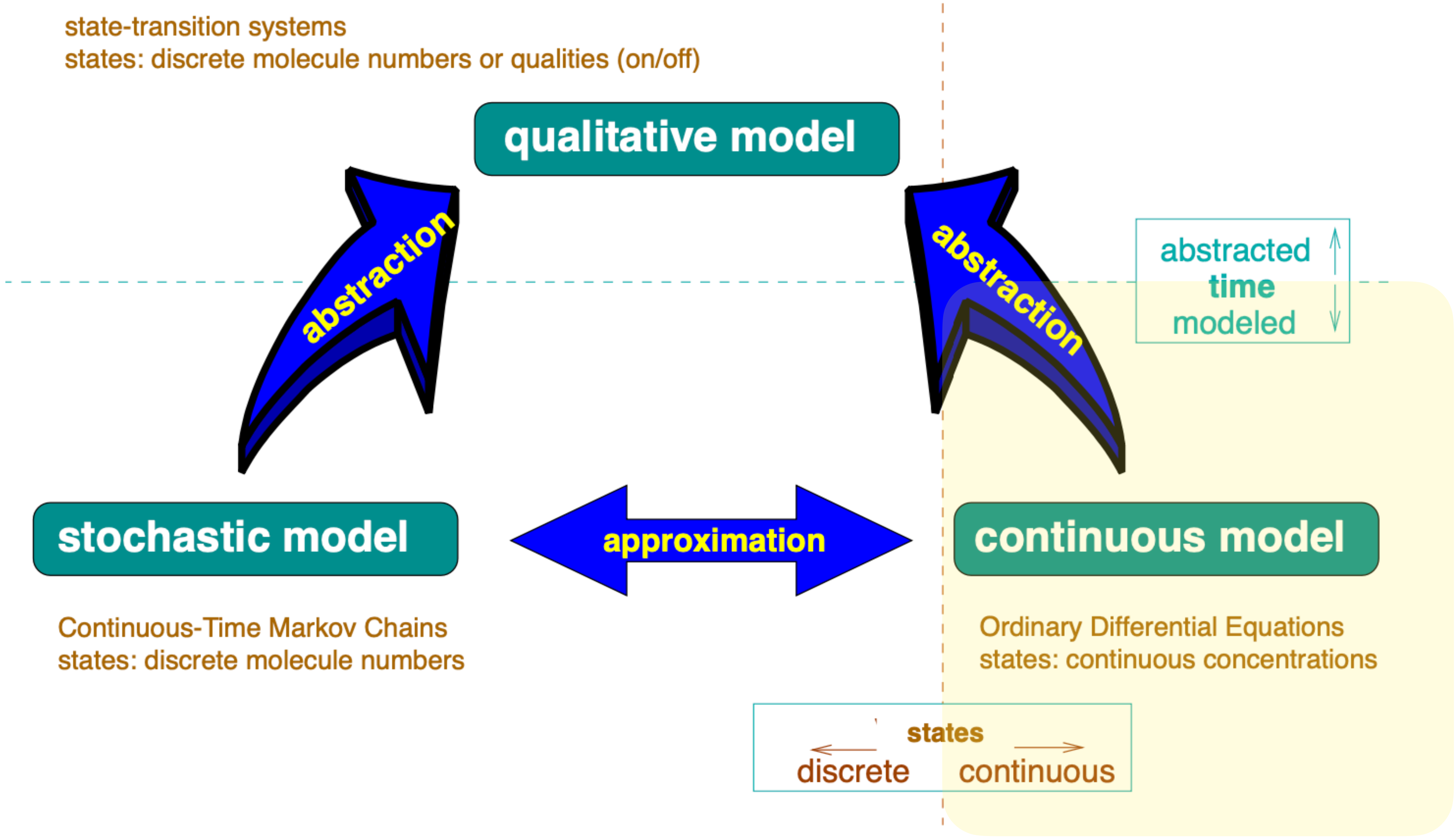
$$\frac{d[\mathrm{ECH}]}{dt} = + \boxed{(R)\text{-DCP into ECH}} + \boxed{(S)\text{-DCP into ECH}} - \boxed{\text{ECH into CPD}}$$

$$\frac{d[\mathrm{CPD}]}{dt} = + \boxed{\text{ECH into CPD}} - \boxed{\text{CPD into GDL}}$$

$$\frac{d[\mathrm{GDL}]}{dt} = + \boxed{\text{CPD into GDL}} - \boxed{\text{GDL into GLY}}$$

Demko et al. *Microorganisms* (2019)

Modelling Frameworks



state-transition systems
states: discrete molecule numbers or qualities (on/off)

qualitative model

abstracted
time
modeled

stochastic model

continuous model

Continuous-Time Markov Chains
states: discrete molecule numbers

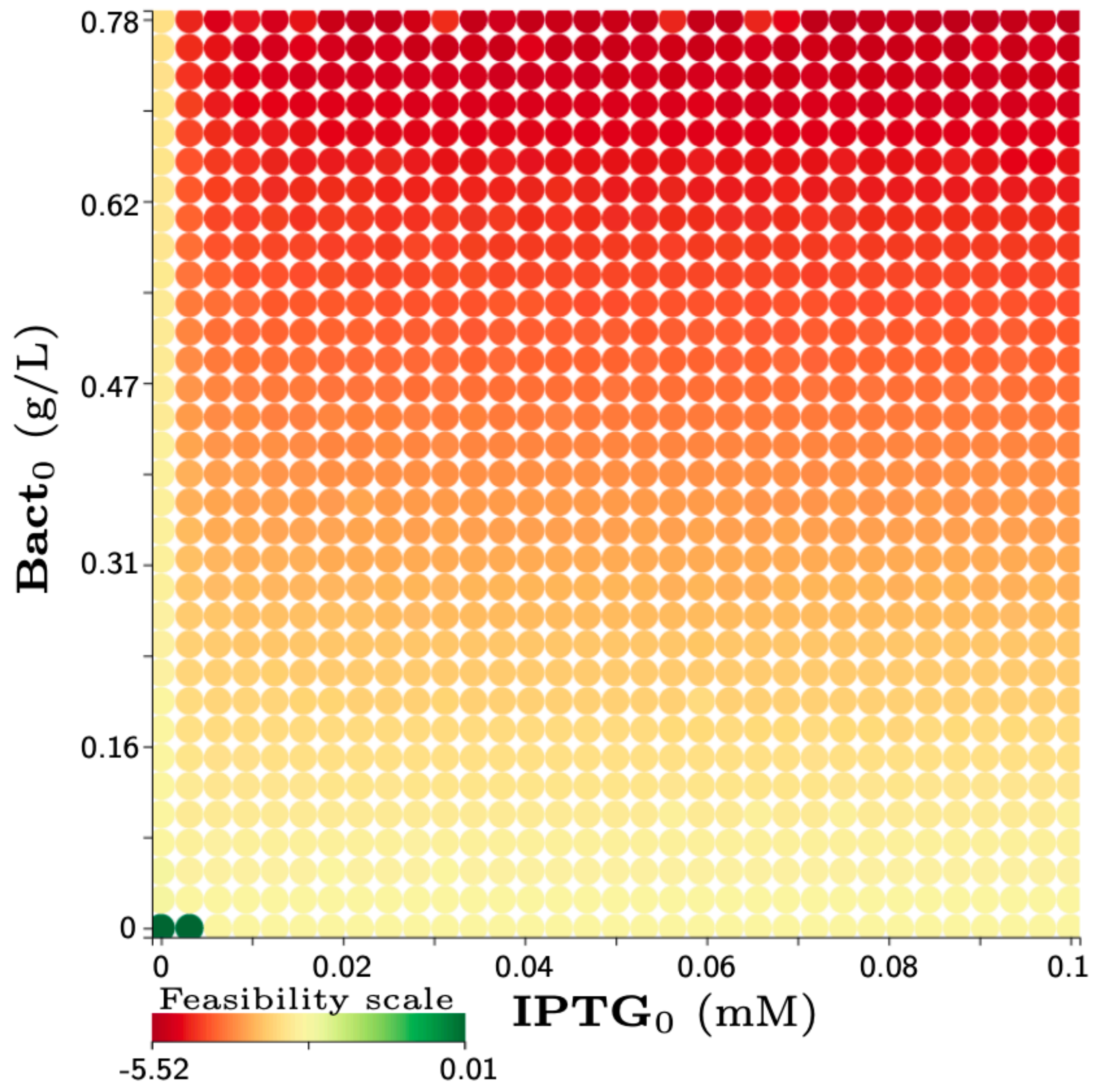
Ordinary Differential Equations
states: continuous concentrations

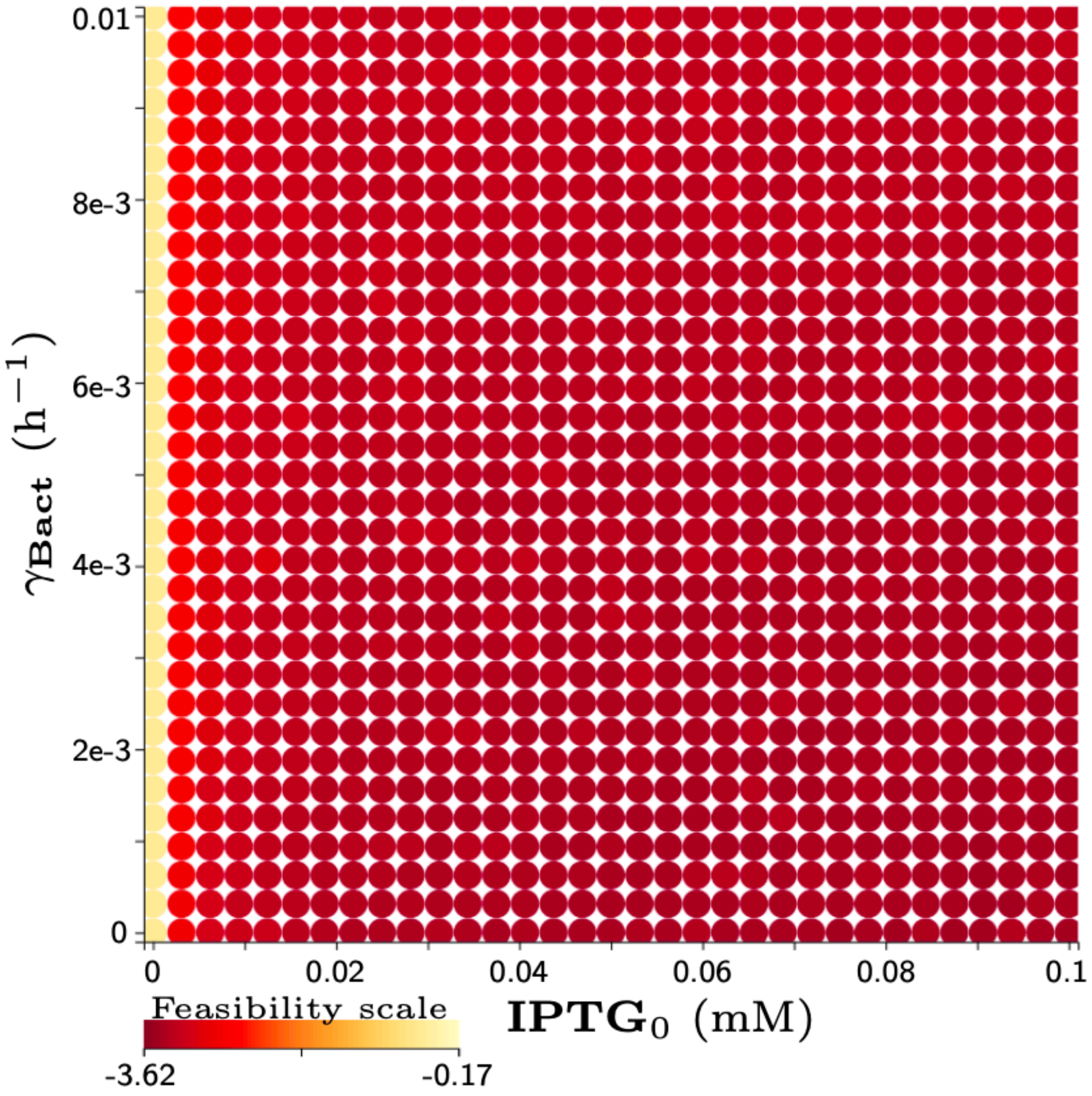
states
discrete continuous

Biodegradation of Trichloropropane in *E. coli*

Extending the Model with Population Growth - Results

The population dies eventually (drops below 0.01 g/L) while TCP does not degrade entirely (does not drop below 0.1 mM) in the 5 h horizon.



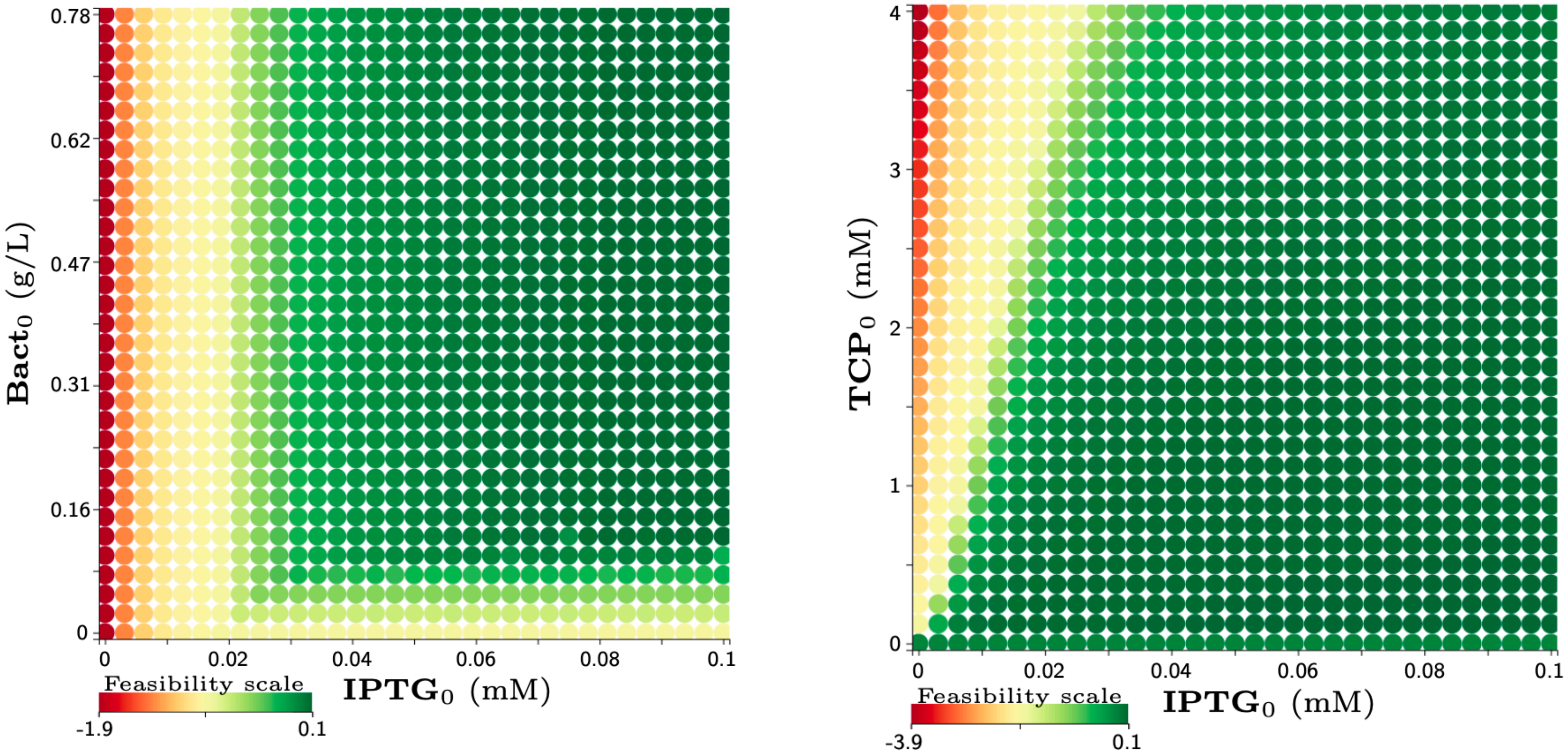


Demko et al. *Microorganisms* (2019)

Biodegradation of Trichloropropane in *E. coli*

Extending the Model with Population Growth - Results

The population will never drop below half of its initial value in the 5 h scope and TCP will degrade (drop below 0.01 mM) in the 2.5 h scope at the same time.



Demko et al. *Microorganisms* (2019)

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!
- applications in cyber-physical systems
- problems:
  - the grand challenge not yet targeted
  - experts trained in life sciences 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