

Semantical and Algorithmic Aspects of the Living

François Fages



INRIA Paris-Rocquencourt



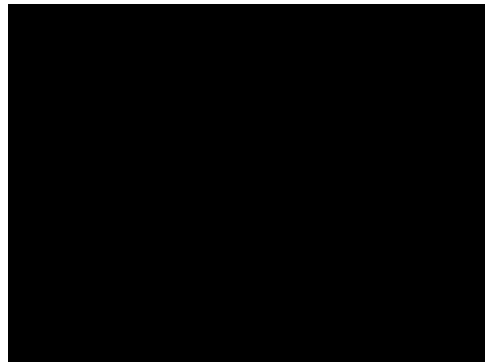
To tackle the complexity of biological systems, investigate:

- Programming Theory Concepts
- Formal Methods of Circuit and Program Verification
- Constraint Modeling and Automated Reasoning Tools

Implementation in the Biochemical Abstract Machine BIOCHAM v3.3

Systems Biology

Aims at system-level understanding of multi-scale biological processes in terms of their elementary interactions at the molecular level.

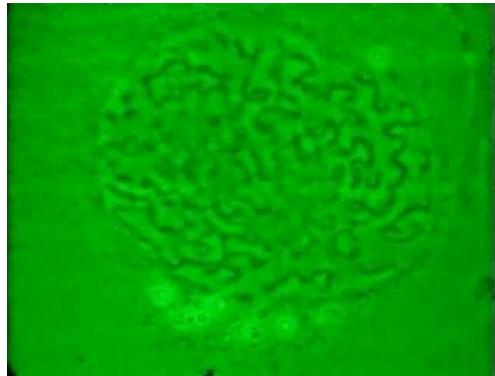


- Systems Biology Markup Language (SBML): model exchange format
- Model repositories: e.g. biomodels.net 1000 models of cell processes
- Modeling environments (Cell designer, Cytoscape, Copasi, Biocham,...)

The limits on what is observable and uncertainty of data make reasoning with **partial information structures** necessary → symbolic, logic, constraints

Semantics of Living Processes ?

Formally, the semantics of a system depend on *our choice* of observables.

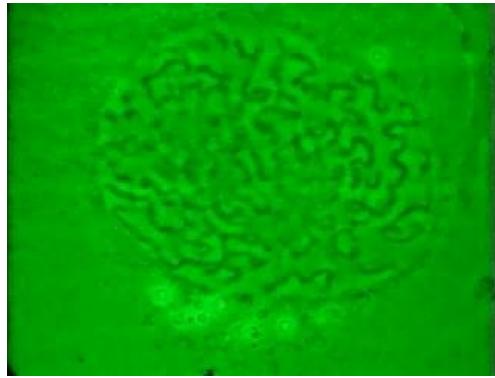


Mitosis movie [Lodish et al. 03]

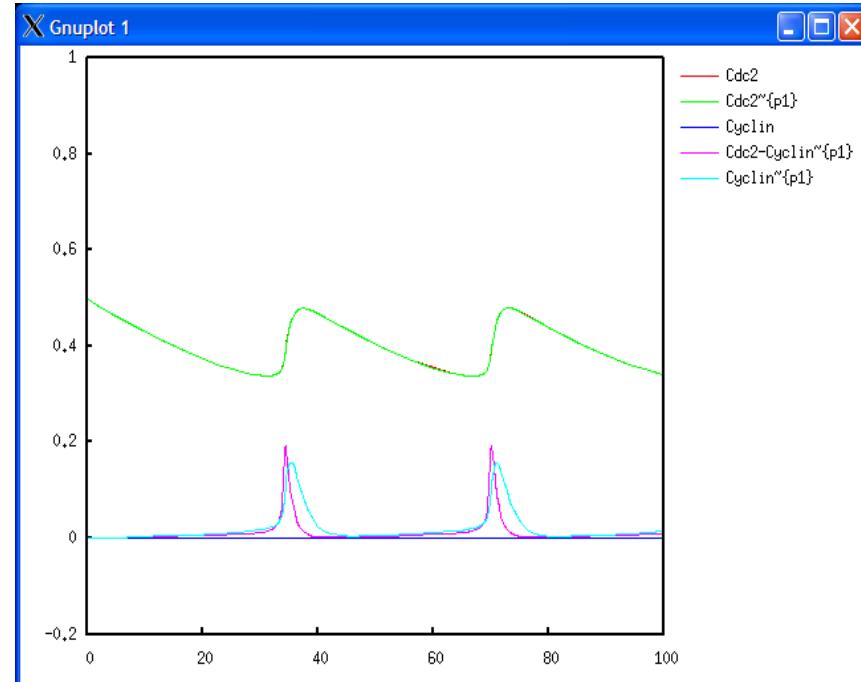
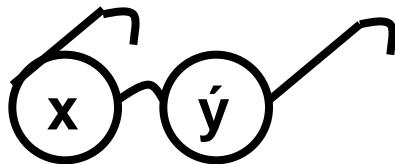
Continuous Differential Semantics

Formally, the semantics of a system depend on *our choice* of observables.

- Concentrations of molecules
- Reaction rates
- Ordinary Differential Equation (ODE) model



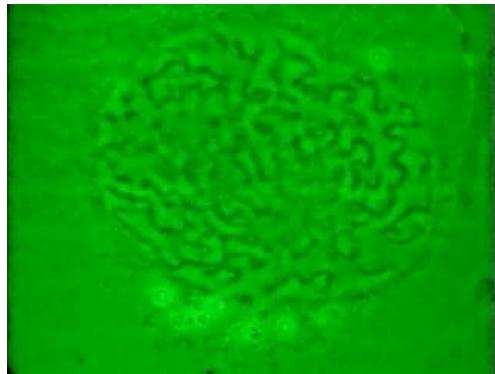
Mitosis movie [Lodish et al. 03]



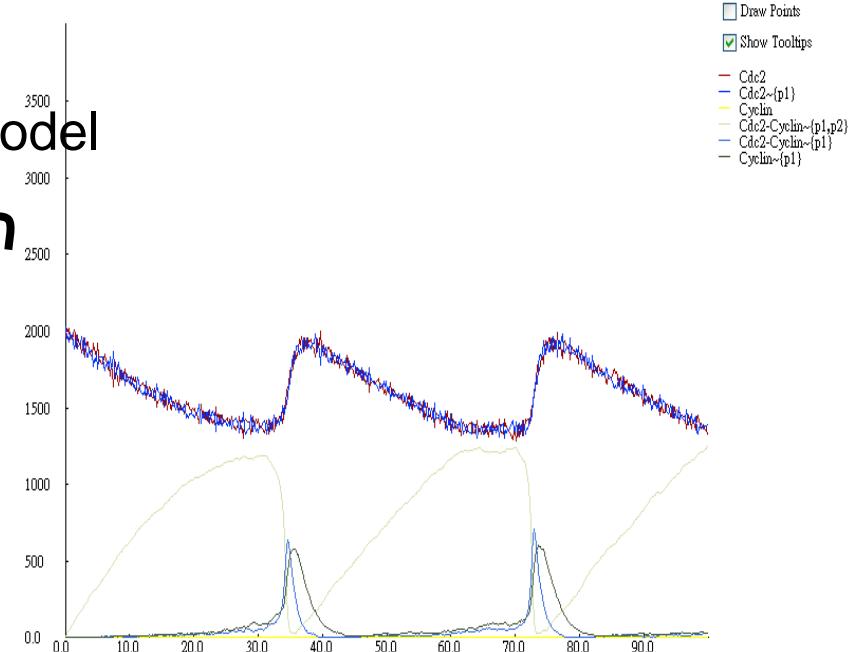
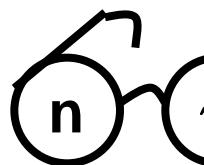
Stochastic Semantics

Formally, the semantics of a system depend on *our choice* of observables.

- (Small) numbers of molecules
- Probabilities of reaction
- Continuous Time Markov Chain (CTMC) model



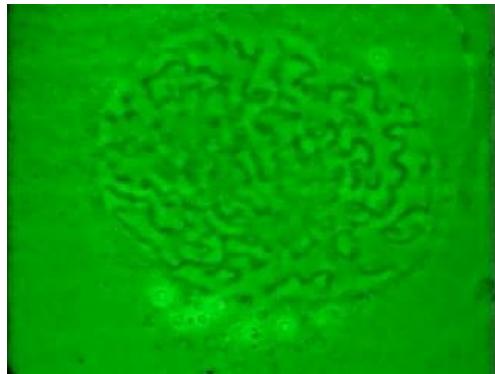
Mitosis movie [Lodish et al. 03]



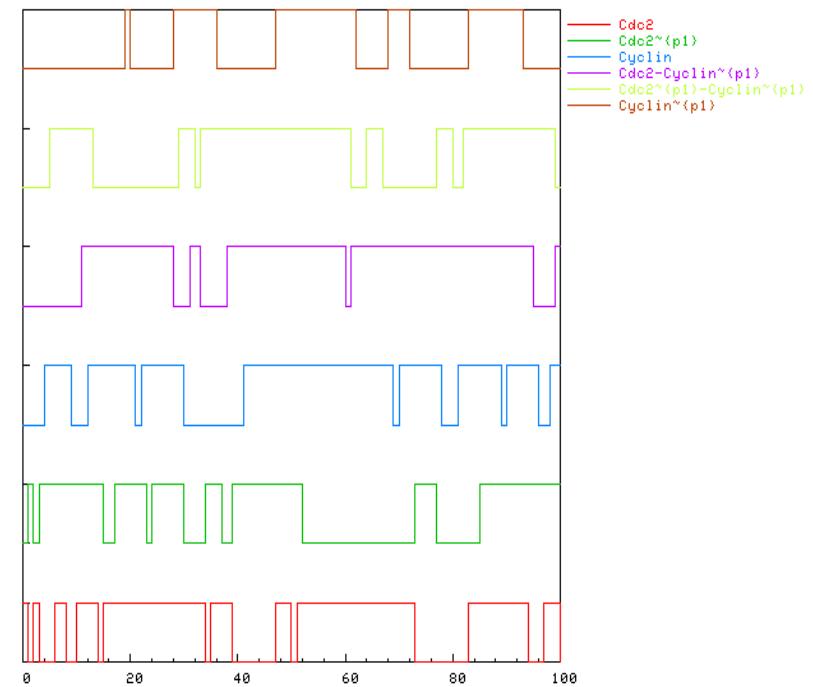
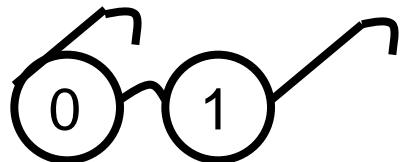
Boolean Semantics

Formally, the semantics of a system depend on *our choice* of observables.

- Presence/absence of molecules
- Boolean transition model



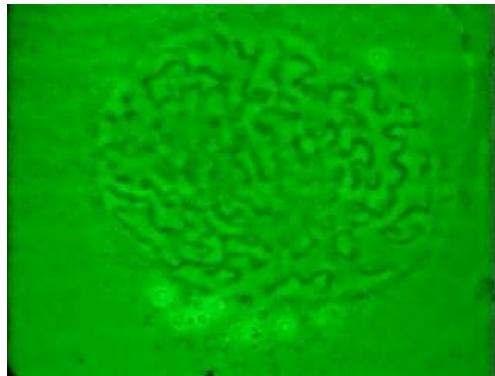
Mitosis movie [Lodish et al. 03]



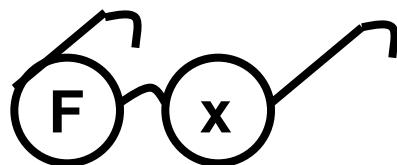
Propositional Temporal Logic

Formally, the semantics of a system depend on *our choice* of observables.

- Presence/absence of molecules
- Temporal logic formulas on Boolean traces



Mitosis movie [Lodish et al. 03]



$F\ x$

$F(x \wedge F(\neg x \wedge y))$

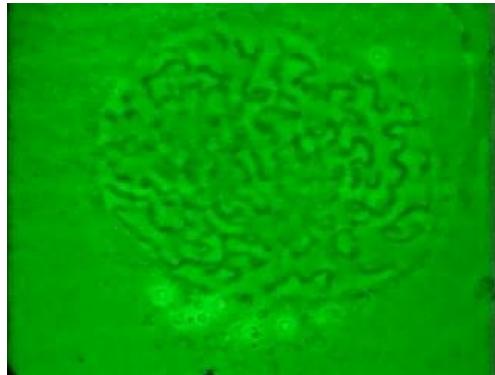
FG ($x \vee y$)

...

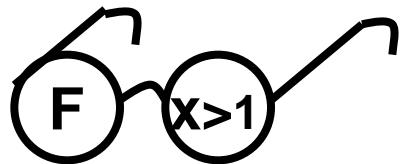
Constraint Temporal Logic LTL(R)

Formally, the semantics of a system depend on *our choice* of observables.

- Concentrations of molecules
- Temporal logic with constraints over R on quantitative traces



Mitosis movie [Lodish et al. 03]



F ($x > 0.2$)

F ($x > 0.2 \wedge F (x < 0.1 \wedge y > 0.2)$)

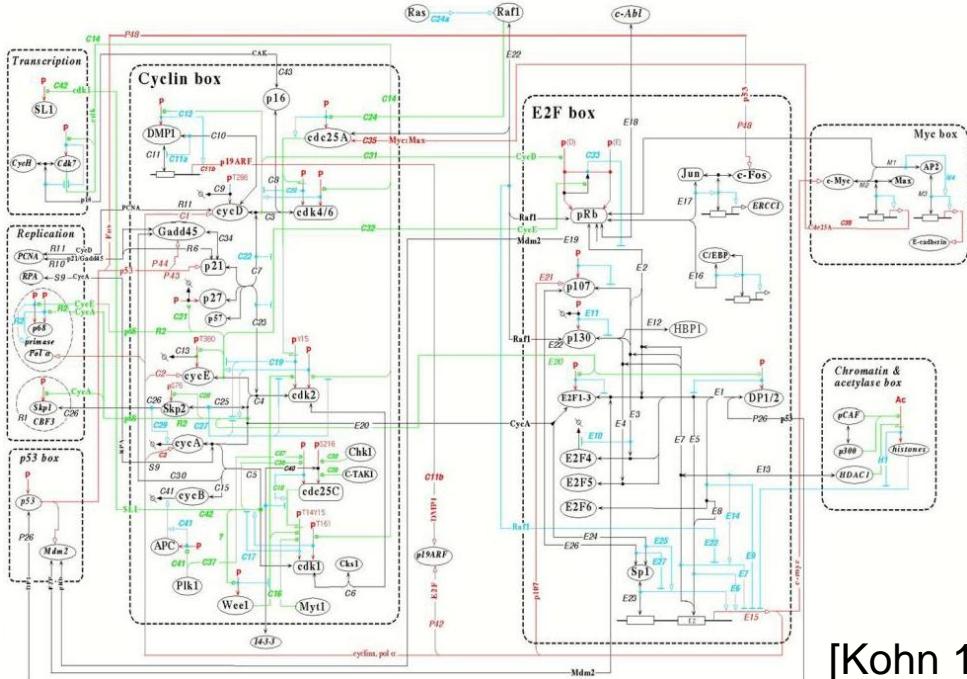
FG ($x > 0.2 \vee y > 0.2$)

...

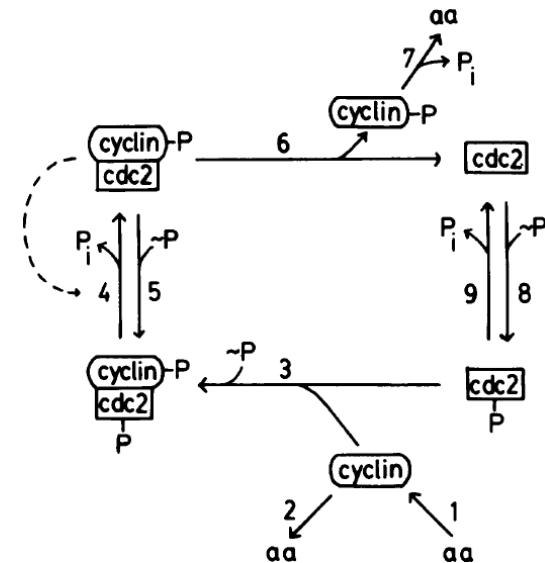
Models in Systems Biology

Models are built in Systems Biology with two contradictory perspectives :

- 1) Models for representing knowledge : the more detailed the better
- 2) Models for making predictions : the more abstract the better



[Kohn 1999]

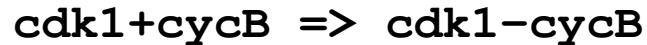


[Tyson 1991]

→ Organize **models** and **formalisms** in hierarchies of abstractions

Formal Reaction Rules

Complexation: $A + B \Rightarrow A-B.$



Decomplexation $A-B \Rightarrow A + B.$

Phosphorylation: $A = [K] \Rightarrow A\sim\{p\}.$

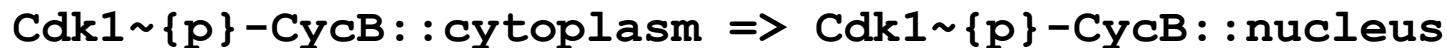


Synthesis: $_ = [G] \Rightarrow A.$



Degradation: $A = [C] \Rightarrow _.$

Transport: $A::L1 \Rightarrow A::L2.$



Semantics of Reaction Rules $\{e_i \text{ for } l_i \Rightarrow r_i\}_{i=1,\dots,n}$



Discrete Semantics: numbers of molecules

Multiset rewriting, Petri net



CHAM [Berry Boudol 90] [Banatre Le Metayer 86]

Boolean Semantics: presence-absence of molecules

Asynchronous Transition System



Differential Semantics: concentrations

Ordinary Differential Equations

$$\frac{dx_k}{dt} = \sum_{i=1}^n r_i(x_k) * e_i - \sum_{j=1}^n l_j(x_k) * e_j$$

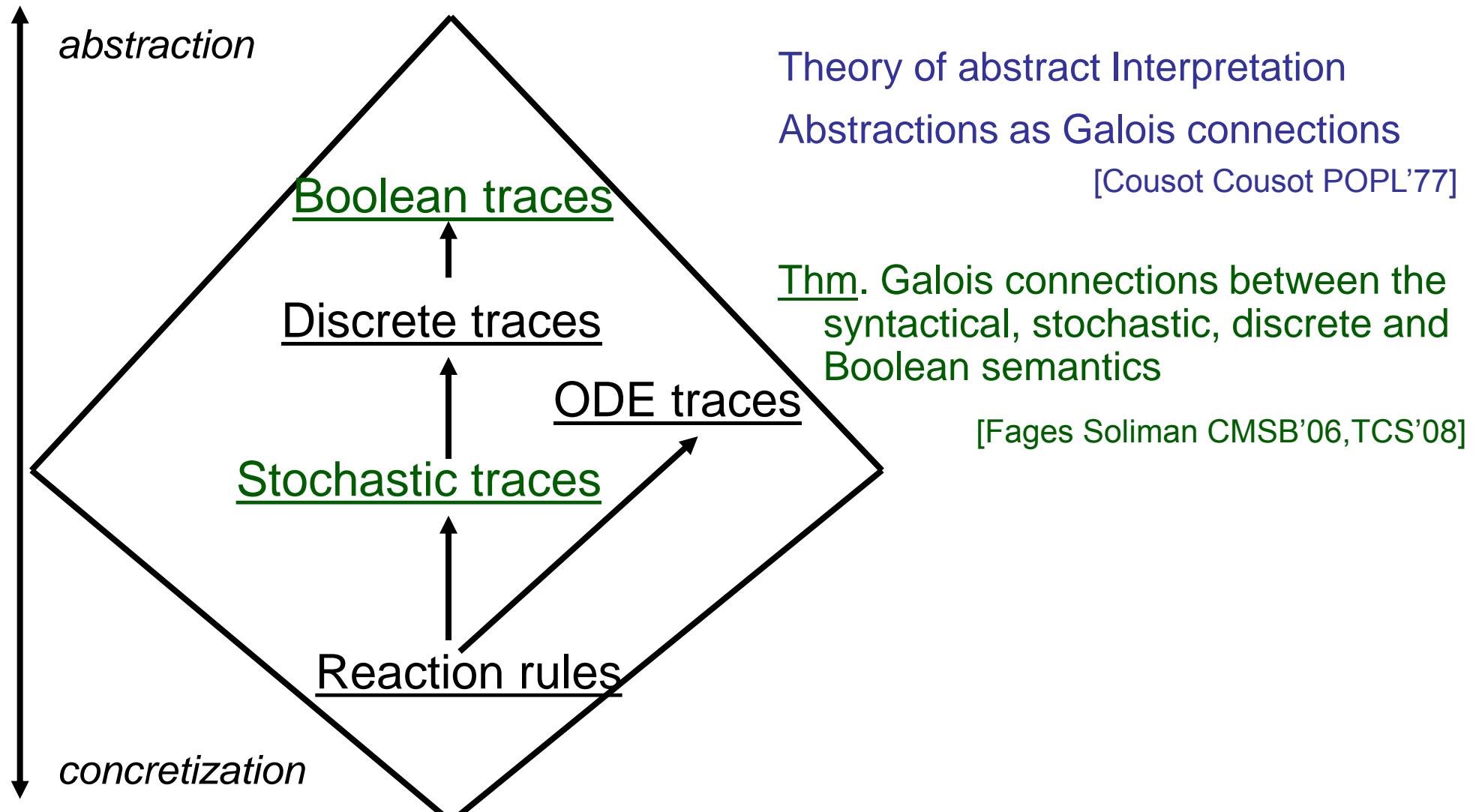
Hybrid Automata (ODE+events)

Stochastic Semantics: numbers of molecules

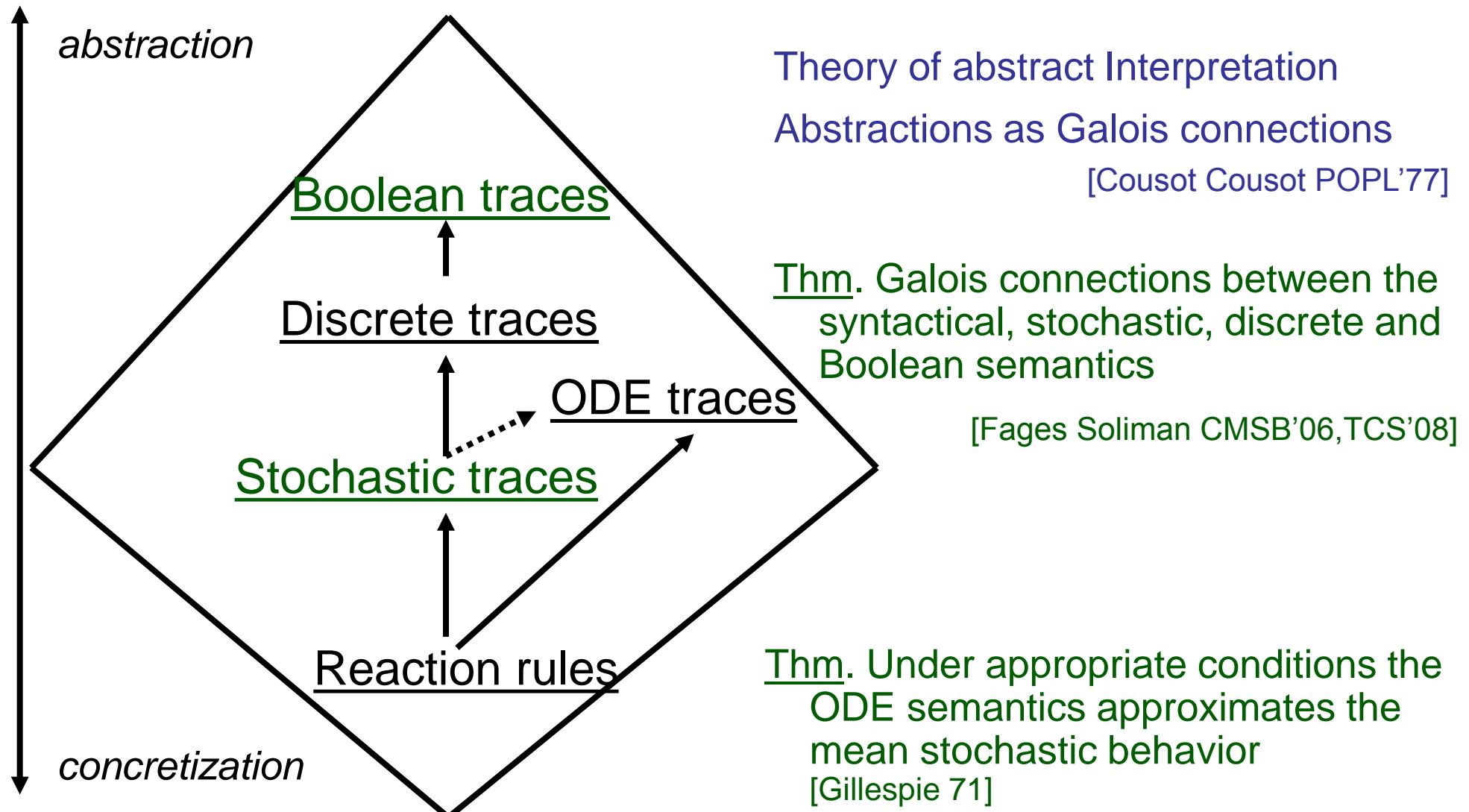
Continuous time Markov chain

$$p_{ij} = \frac{\tau_{ij}}{\sum_{(S_i, S_k, \tau_{ik}) \in s} \tau_{ik}}$$

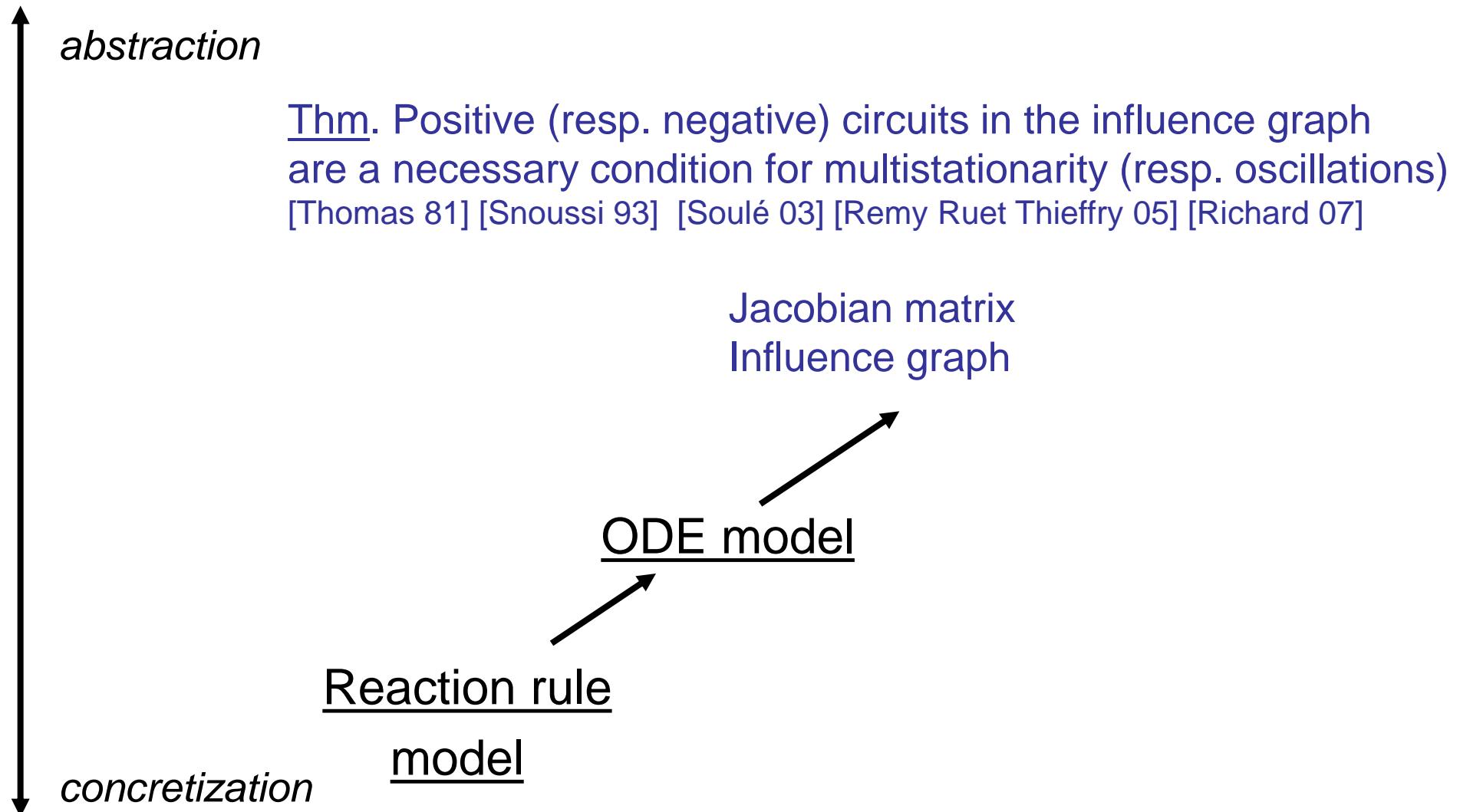
Hierarchy of Semantics



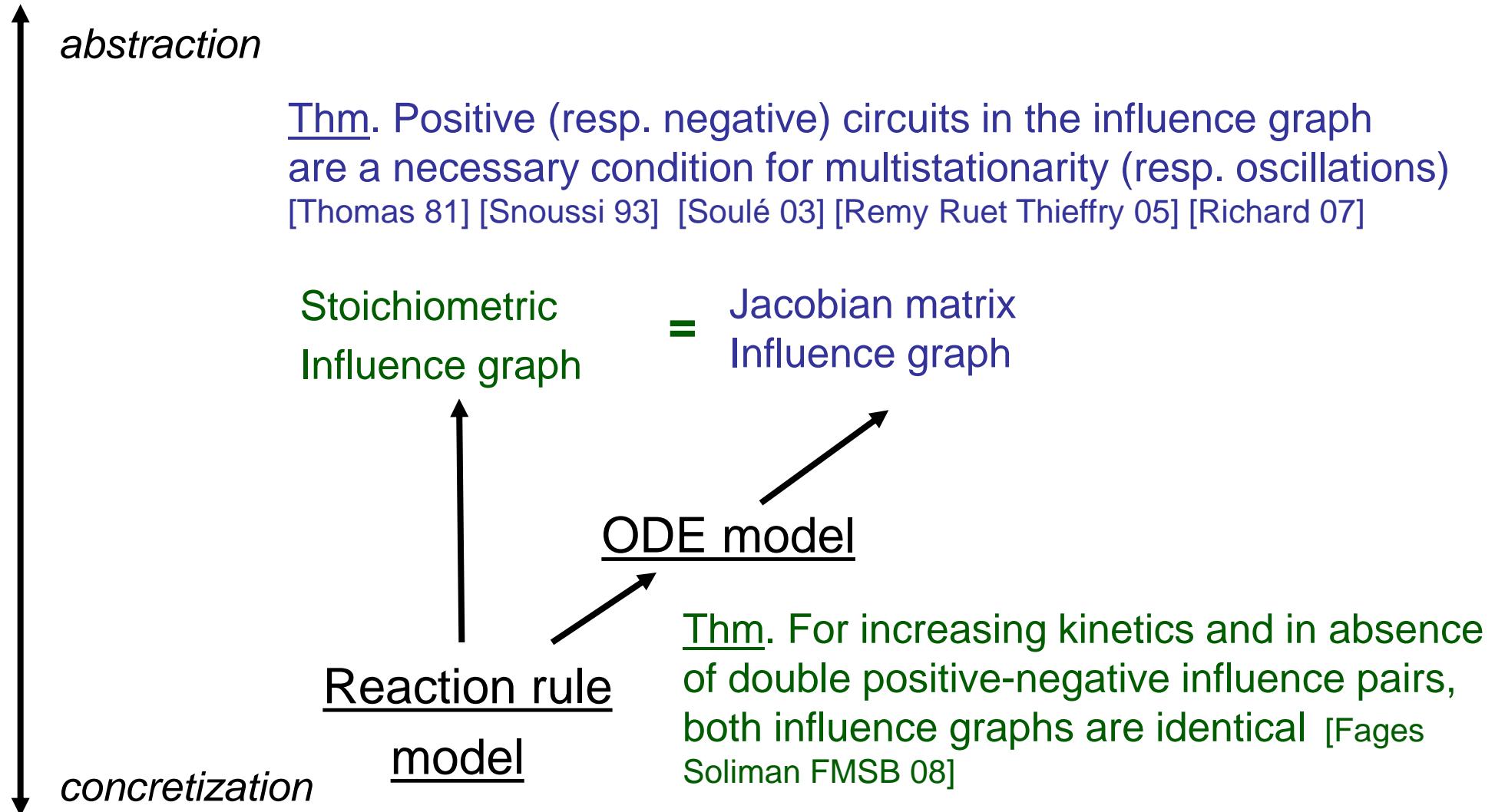
Hierarchy of Semantics



Influence Graph Abstraction

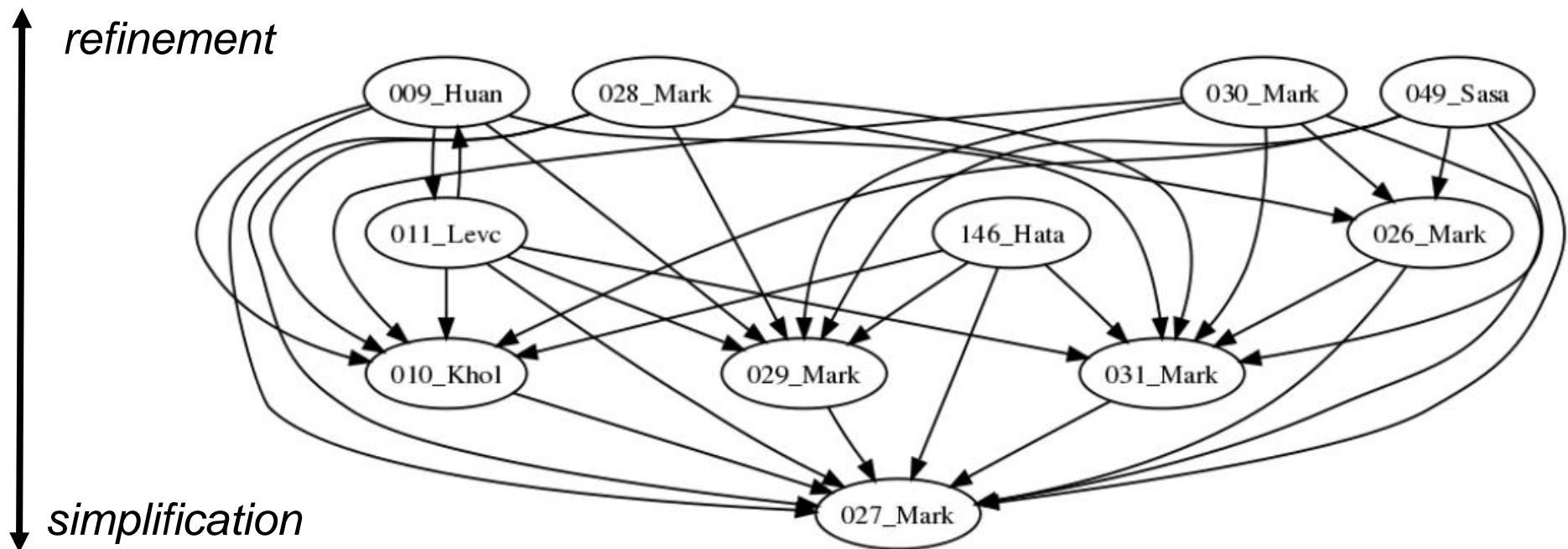


Influence Graph Abstraction



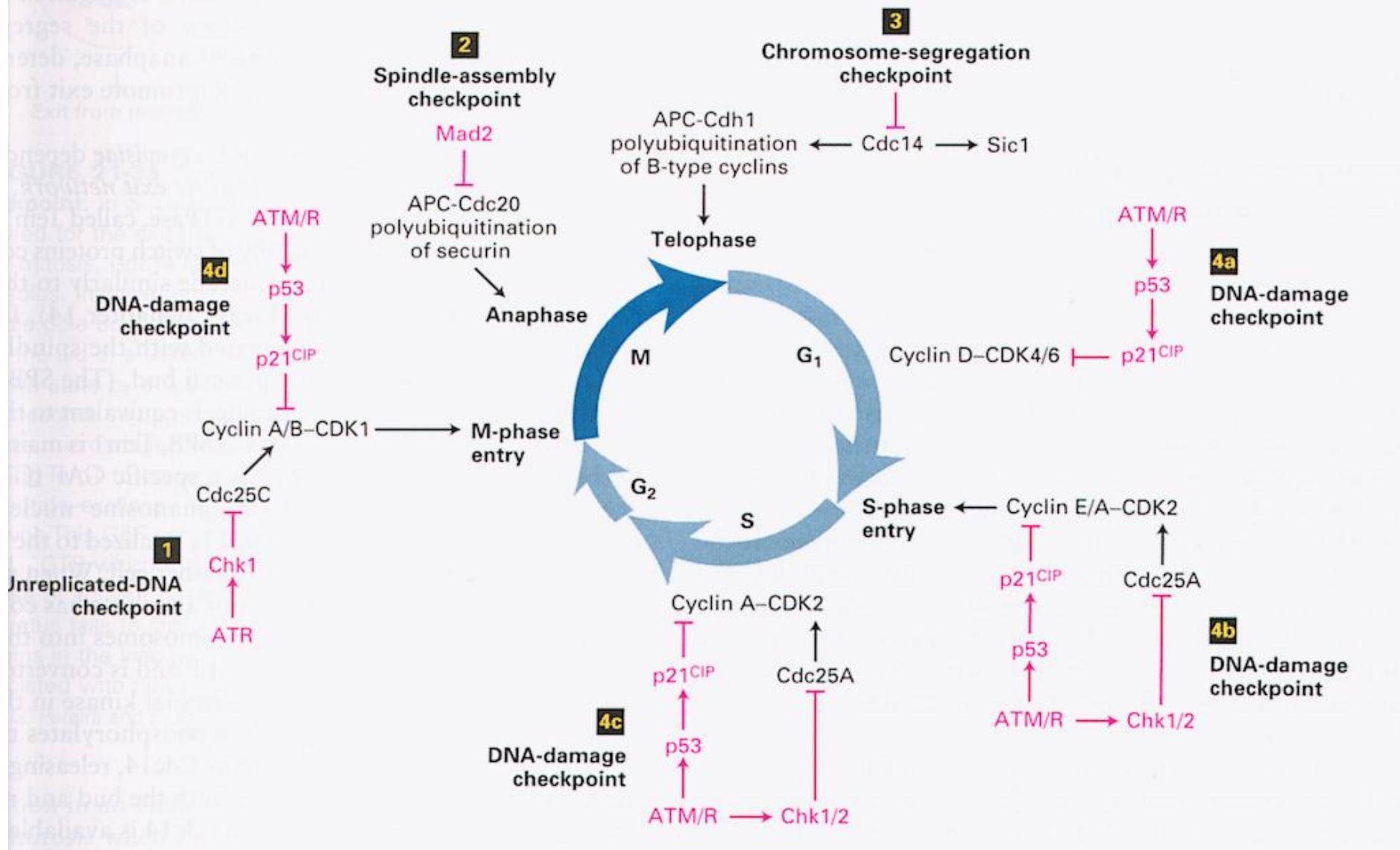
Model Reductions / Reaction Graph

Subgraph epimorphisms: reductions by deletion/merging of species/reactions
NP-complete existence problem, Constraint Logic Program to solve it

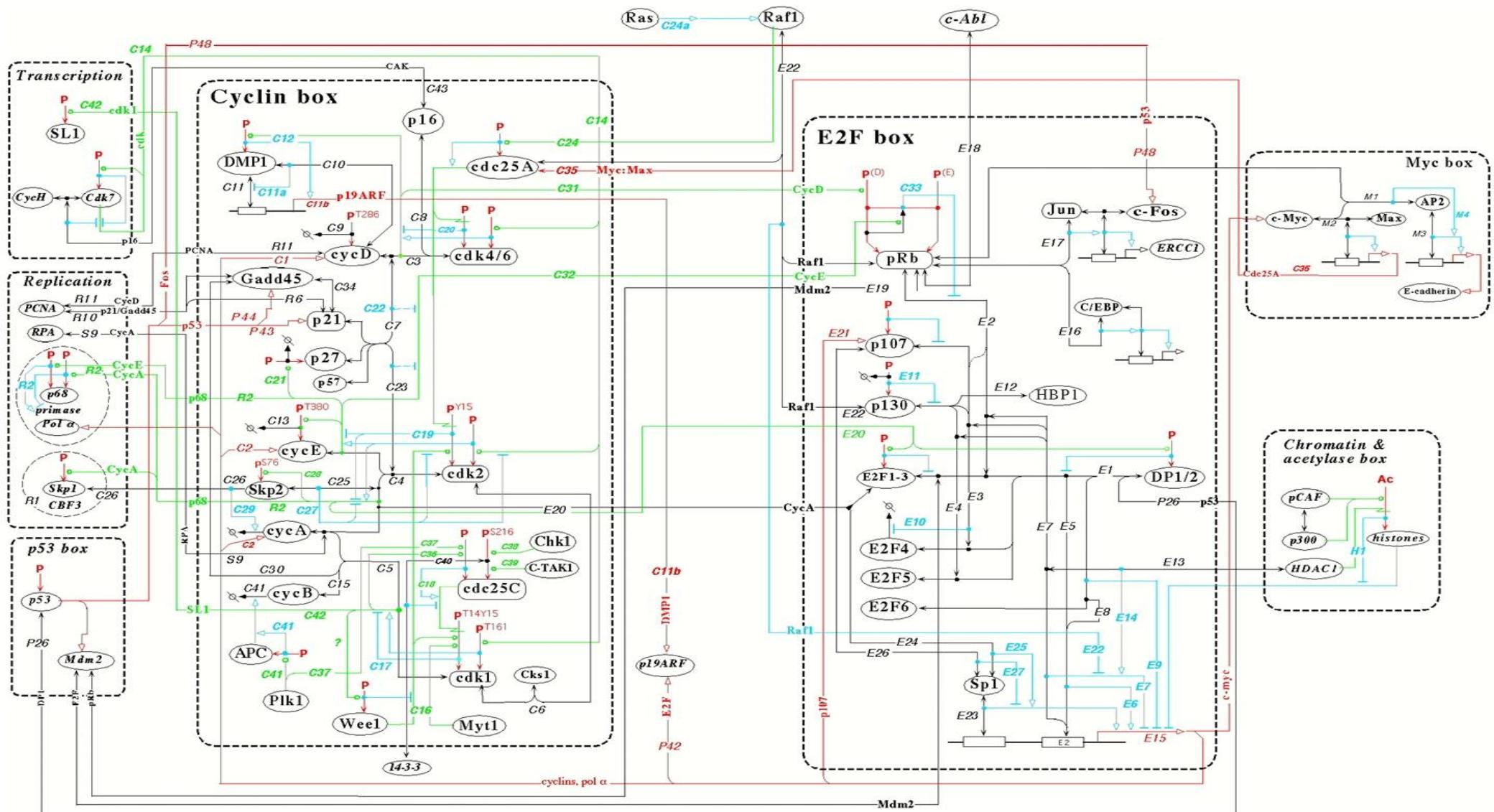


Computed hierarchy of MAPK signaling models in <http://www.biomodels.net>
[Gay Fages Soliman, A graphical method to reduce and relate models in systems biology 2010 bi]

Cell Cycle Control



Mammalian Cell Cycle Control Map [Kohn 99]



Kohn's map detail for Cdk2

Complexations with CycA and CycE

$\text{cdk2} \sim \$P + \text{cycA} - \$C \Rightarrow \text{cdk2} \sim \$P - \text{cycA} - \$C$

where \$C in {_, cks1} .

$\text{cdk2} \sim \$P + \text{cycE} \sim \$Q - \$C \Rightarrow \text{cdk2} \sim \$P - \text{cycE} \sim \$Q - \C

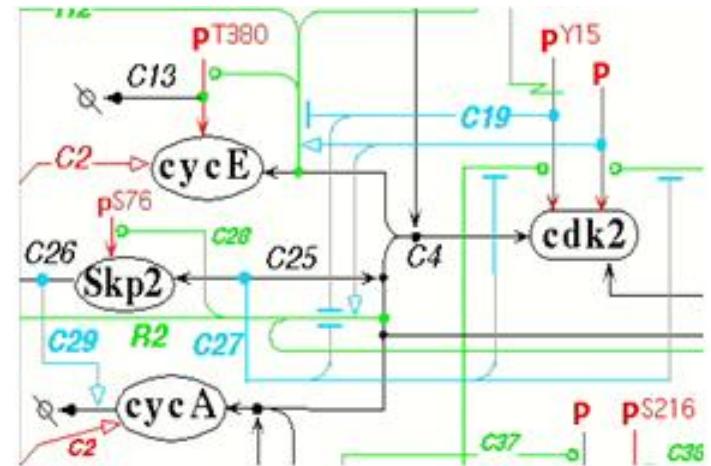
where \$C in {_, cks1} .

$p57 + \text{cdk2} \sim \$P - \text{cycA} - \$C \Rightarrow p57 - \text{cdk2} \sim \$P - \text{cycA} - \$C$

where \$C in {_, cks1} .

$\text{cycE} - \$C = [\text{cdk2} \sim \{p2\} - \text{cycE} - \$S] \Rightarrow \text{cycE} \sim \{T380\} - \C

where \$S in {_, cks1} and \$C in {_, cdk2 ~ ?, cdk2 ~ ? - cks1}



147 rule patterns \rightarrow 2733 expanded rules [Chabrier Chiaverini Danos Fages Schachter 04]

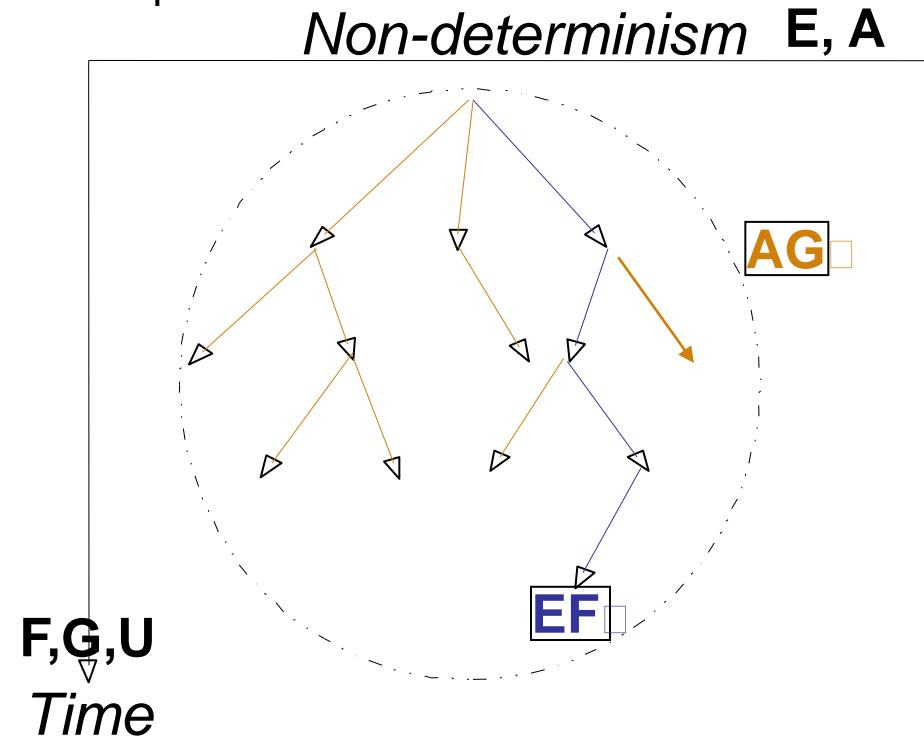
How to query the possible dynamical properties of such a system ?
reachability, checkpoints, steady states, oscillations ?

Temporal Logic Queries

Temporal logics introduced for program verification by [Pnueli 77]

Computation Tree Logic CTL [Emerson Clarke 80]

Non-det. Time	E exists	A always
X next time	EX (φ)	AX (φ)
F finally	EF (φ) $\neg AG(\neg \varphi)$	AF (φ) <i>liveness</i>
G globally	EG (φ) $\neg AF(\neg \varphi)$	AG (φ) <i>safety</i>
U until	E ($\varphi_1 \mathbf{U} \varphi_2$)	A ($\varphi_1 \mathbf{U} \varphi_2$)



Kohn's Map Model-Checking

147-2733 rules, 165 proteins and genes, 500 variables, 2^{500} states.

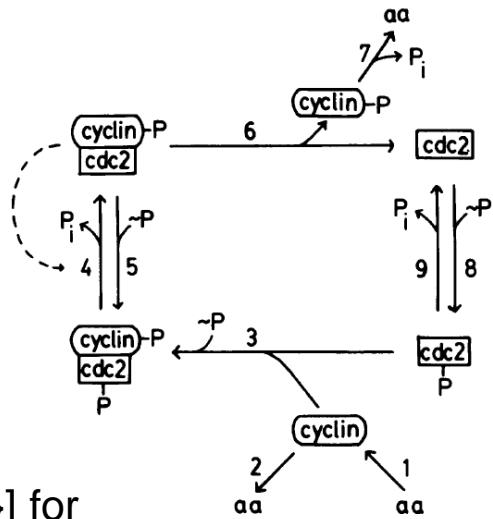
Biocham NuSMV model-checker time in seconds [Chabrier Fages CMSB 03]

Initial state G2	Query:	Time in sec.
	compiling	29
Reachability G1	EF CycE	2
Reachability G1	EF CycD	1.9
Reachability G1	EF PCNA-CycD	1.7
Checkpoint for mitosis complex	$\neg \mathbf{EF} (\neg \text{Cdc25}\sim\{\text{Nterm}\})$ $\mathbf{U} \text{Cdk1}\sim\{\text{Thr161}\}\text{-CycB}$	2.2
Oscillations CycA	EG ((EF \neg CycA) & (EF CycA))	31.8
Oscillations CycB	EG ((EF \neg CycB) & (EF CycB)) false !	6

Constraint Linear Time Logic LTL(R)

- $\mathbf{F}([A]>10)$: the concentration of A eventually gets above 10.
- $\mathbf{FG}([A]>10)$: the concentration of A eventually reaches and remains above value 10.
- $\mathbf{F}(\text{Time}=t_1 \wedge [A]=v_1 \wedge \mathbf{F}(\dots \wedge \mathbf{F}(\text{Time}=t_N \wedge [A]=v_N)\dots))$
Numerical data time series (e.g. experimental curves)
- $\mathbf{G}([A]+[B]<[C])$: the concentration of C is always greater than the sum of the concentrations of A and B.
- $\mathbf{F}([M] < V \wedge \mathbf{X}([M] = V) \wedge \mathbf{X}([M] < V)))$: local maximum V
- oscillations, period constraints, etc.

Quantitative Model of Cell Cycle Control [Tyson 91]



k1 for $_ \rightarrow \text{Cyclin}$.

$k3 * [\text{Cyclin}] * [\text{Cdc2} \sim \{\text{p1}\}]$ for
 $\text{Cyclin} + \text{Cdc2} \sim \{\text{p1}\} \Rightarrow \text{Cdc2} \sim \{\text{p1}\} - \text{Cyclin} \sim \{\text{p1}\}$.

$k6 * [\text{Cdc2-Cyclin} \sim \{\text{p1}\}]$ for
 $\text{Cdc2-Cyclin} \sim \{\text{p1}\} \Rightarrow \text{Cdc2} + \text{Cyclin} \sim \{\text{p1}\}$.

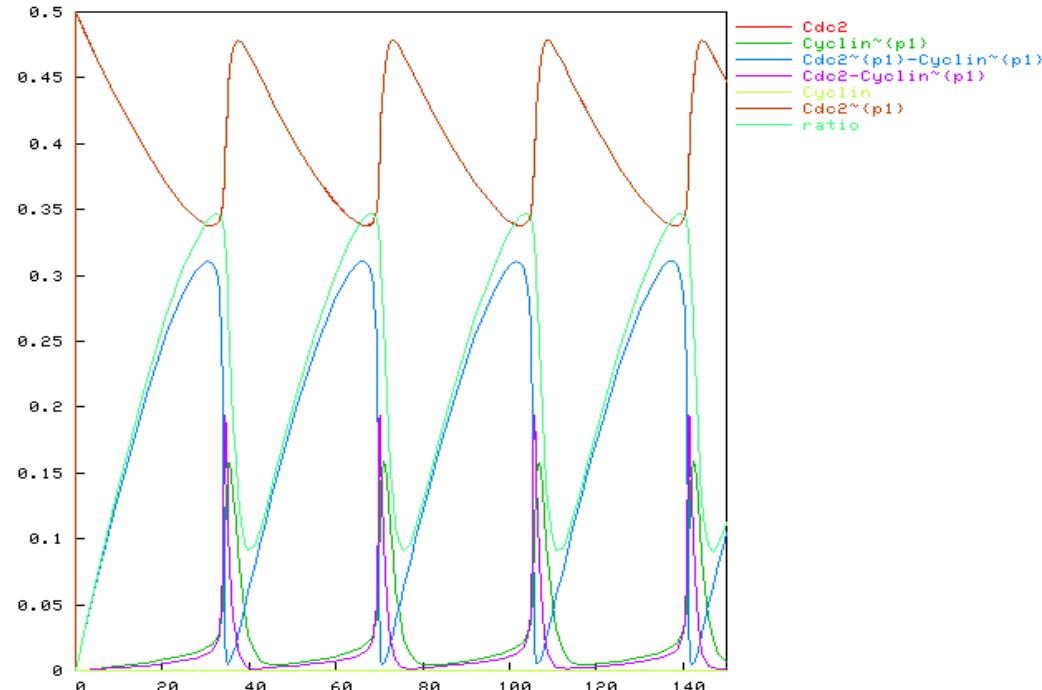
$k7 * [\text{Cyclin} \sim \{\text{p1}\}]$ for $\text{Cyclin} \sim \{\text{p1}\} \Rightarrow _$.

$k8 * [\text{Cdc2}]$ for $\text{Cdc2} \Rightarrow \text{Cdc2} \sim \{\text{p1}\}$.

$k9 * [\text{Cdc2} \sim \{\text{p1}\}]$ for $\text{Cdc2} \sim \{\text{p1}\} \Rightarrow \text{Cdc2}$.

$k4p * [\text{Cdc2} \sim \{\text{p1}\} - \text{Cyclin} \sim \{\text{p1}\}]$ for $\text{Cdc2} \sim \{\text{p1}\} - \text{Cyclin} \sim \{\text{p1}\} \Rightarrow \text{Cdc2-Cyclin} \sim \{\text{p1}\}$.

$k4 * [\text{Cdc2-Cyclin} \sim \{\text{p1}\}]^2 * [\text{Cdc2} \sim \{\text{p1}\} - \text{Cyclin} \sim \{\text{p1}\}]$ for
 $\text{Cdc2} \sim \{\text{p1}\} - \text{Cyclin} \sim \{\text{p1}\} = [\text{Cdc2-Cyclin} \sim \{\text{p1}\}] \Rightarrow \text{Cdc2-Cyclin} \sim \{\text{p1}\}$.



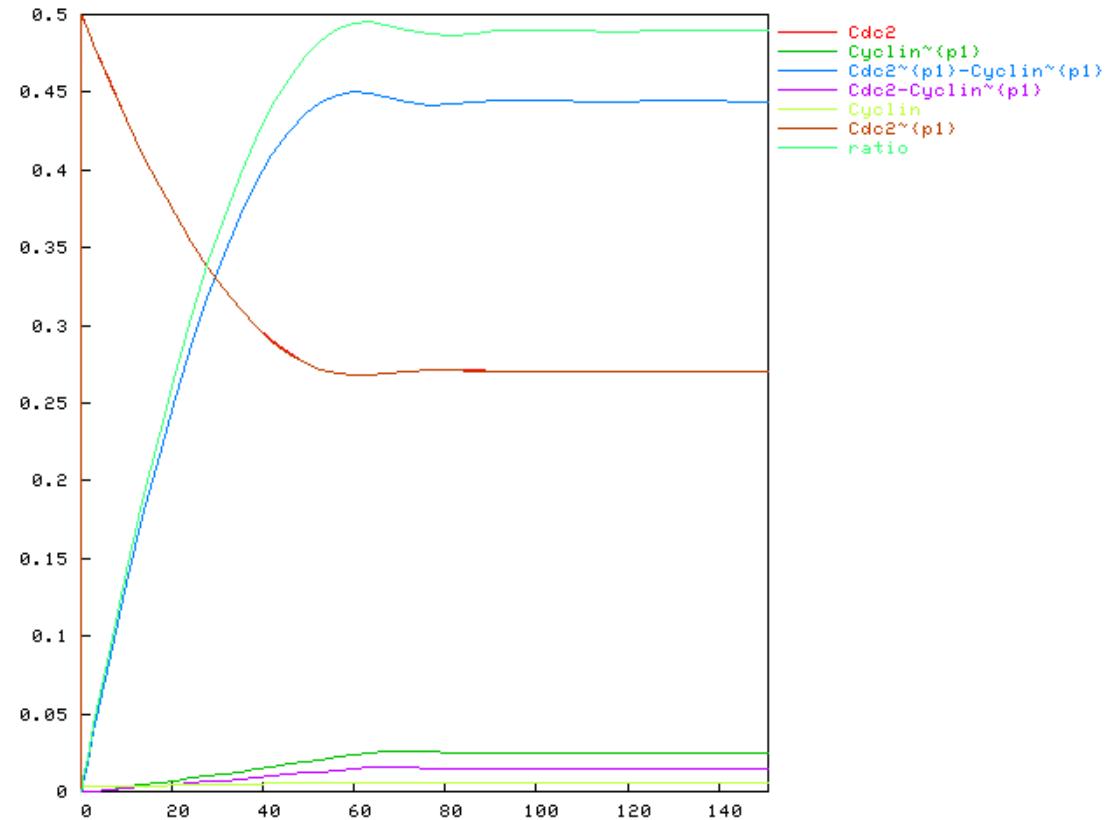
Parameter Search from LTL(R) Properties

```
biocham: learn_parameter([k3,k4],[(0,200),(0,200)],20,  
oscil(Cdc2-Cyclin~{p1},3),150).
```

Parameter Search from LTL(R) Properties

```
biocham: learn_parameter([k3,k4],[(0,200),(0,200)],20,  
oscil(Cdc2-Cyclin~{p1},3),150).
```

First values found :
parameter(k3,10).
parameter(k4,70).



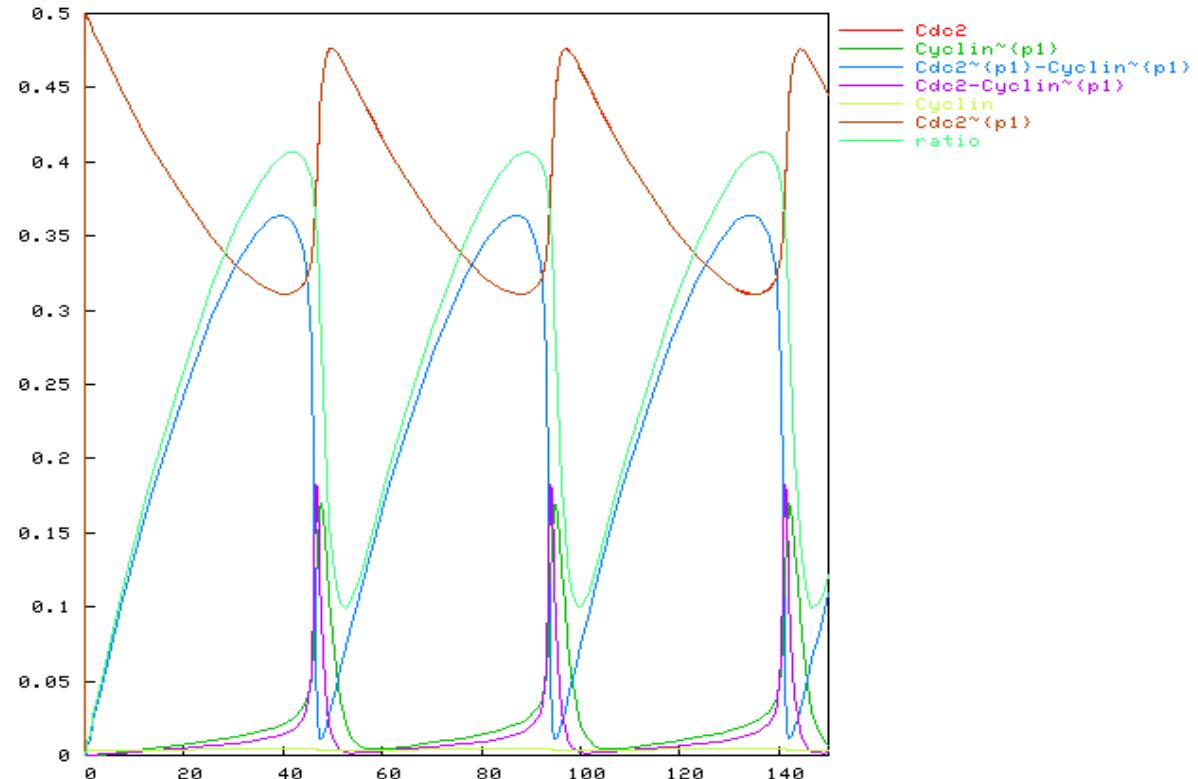
Parameter Search from LTL(R) Properties

```
biocham: learn_parameter([k3,k4],[(0,200),(0,200)],20,  
oscil(Cdc2-Cyclin~{p1},3) & F([Cdc2-Cyclin~{p1}]>0.15), 150).
```

First values found :

parameter(k3,10).

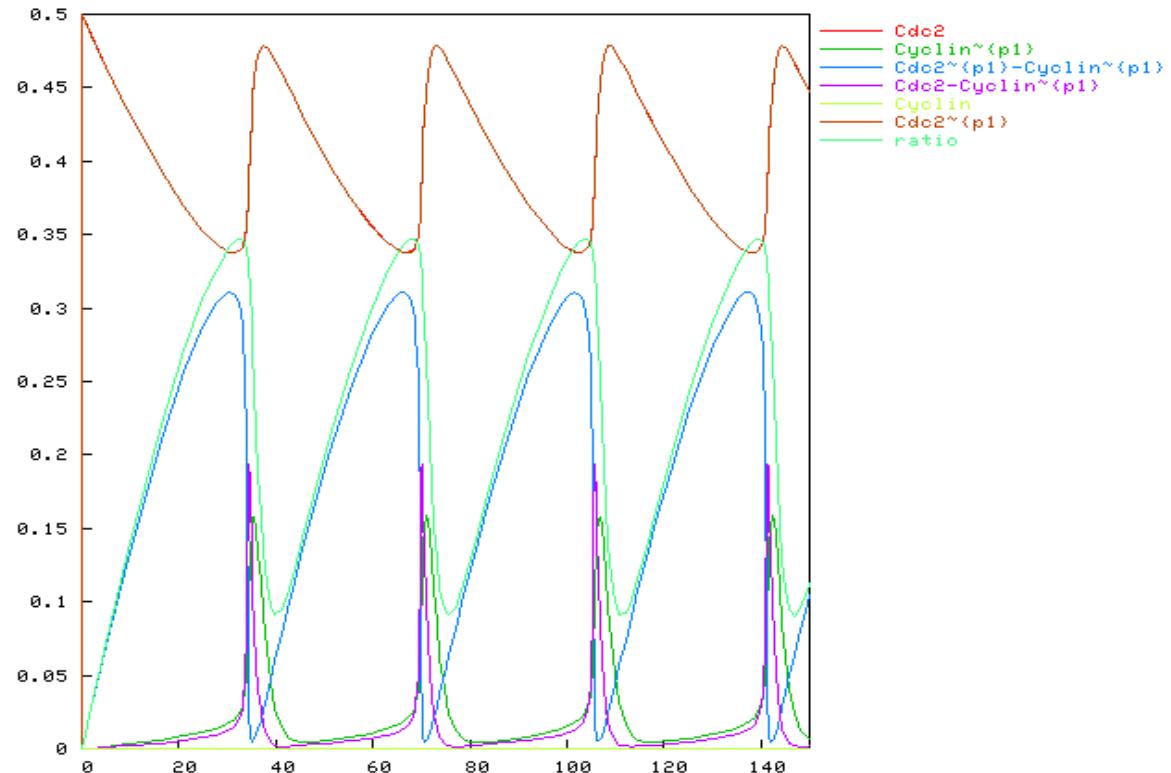
parameter(k4,120).



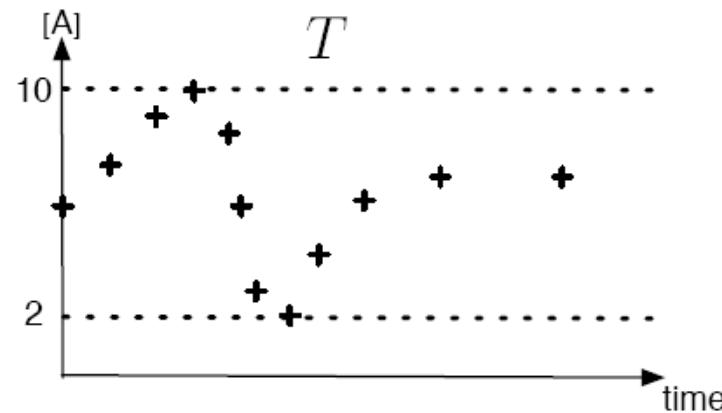
Parameter Search from LTL(R) Properties

```
biocham: learn_parameter([k3,k4],[(0,200),(0,200)],20,  
period(Cdc2-Cyclin~{p1},35), 150).
```

First values found:
parameter(k3,10).
parameter(k4,280).



Model-Checking Generalized to Constraint Solving



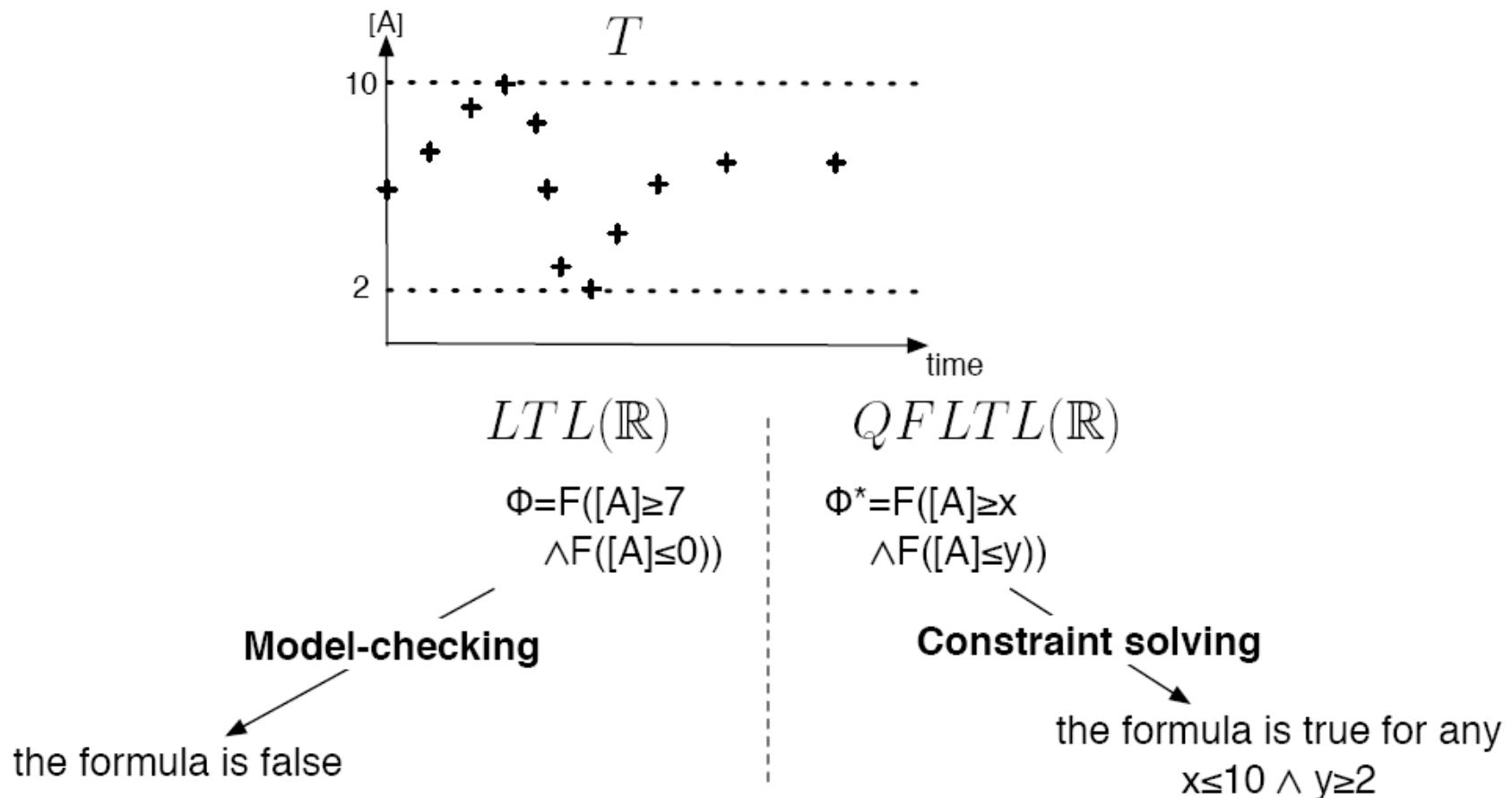
$LTL(\mathbb{R})$

$$\Phi = F([A] \geq 7 \wedge F([A] \leq 0))$$

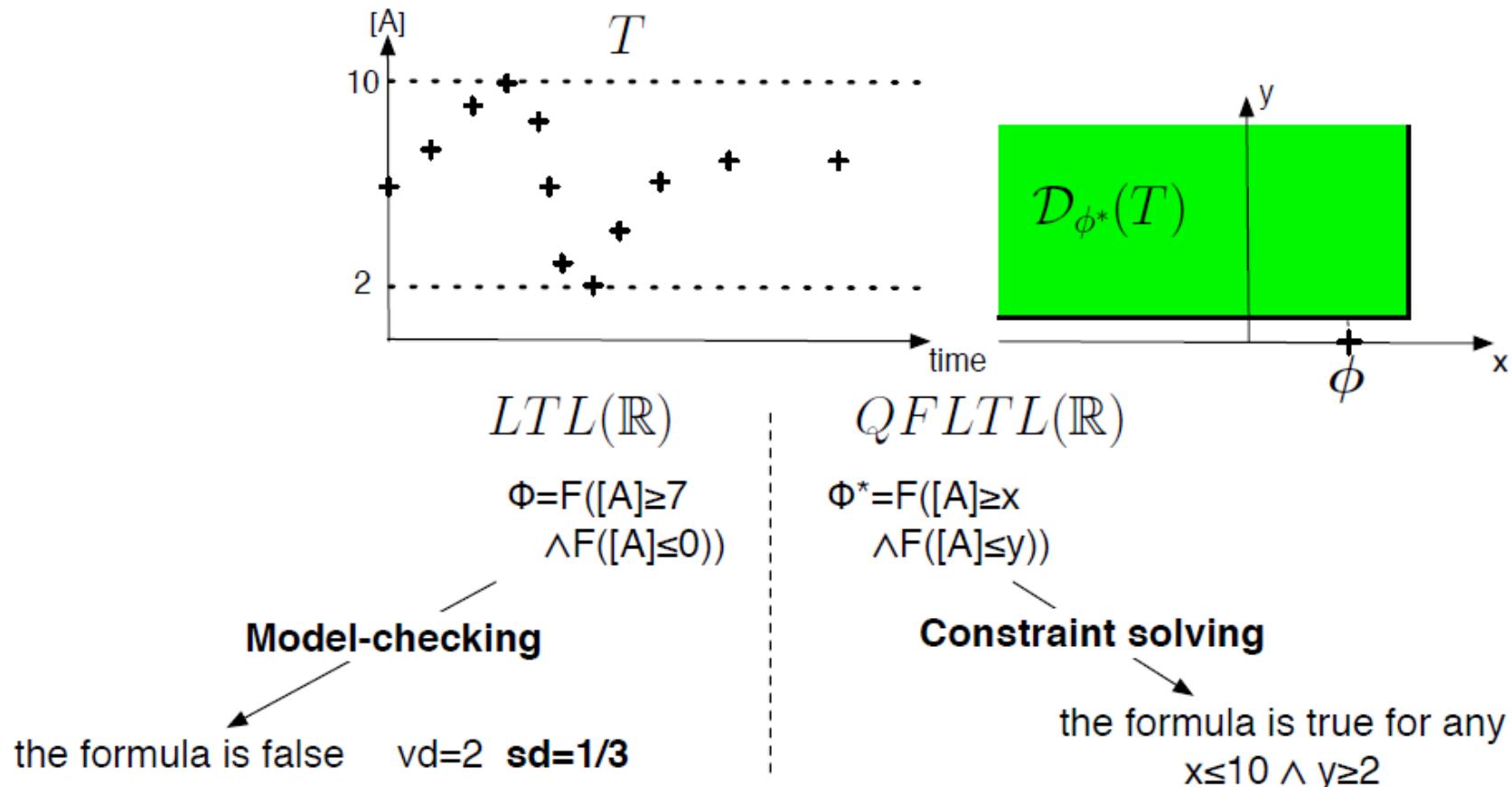
Model-checking

the formula is false

Model-Checking Generalized to Constraint Solving



Model-Checking Generalized to Constraint Solving

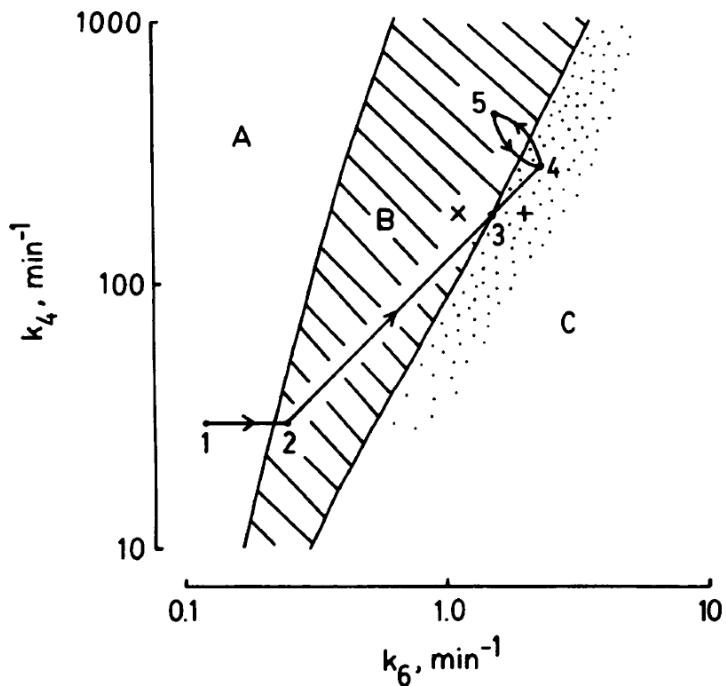


Validity domain $\mathcal{D}_{\phi^*}(T)$ for the **free variables** in ϕ^* [Fages Rizk CMSB'07]

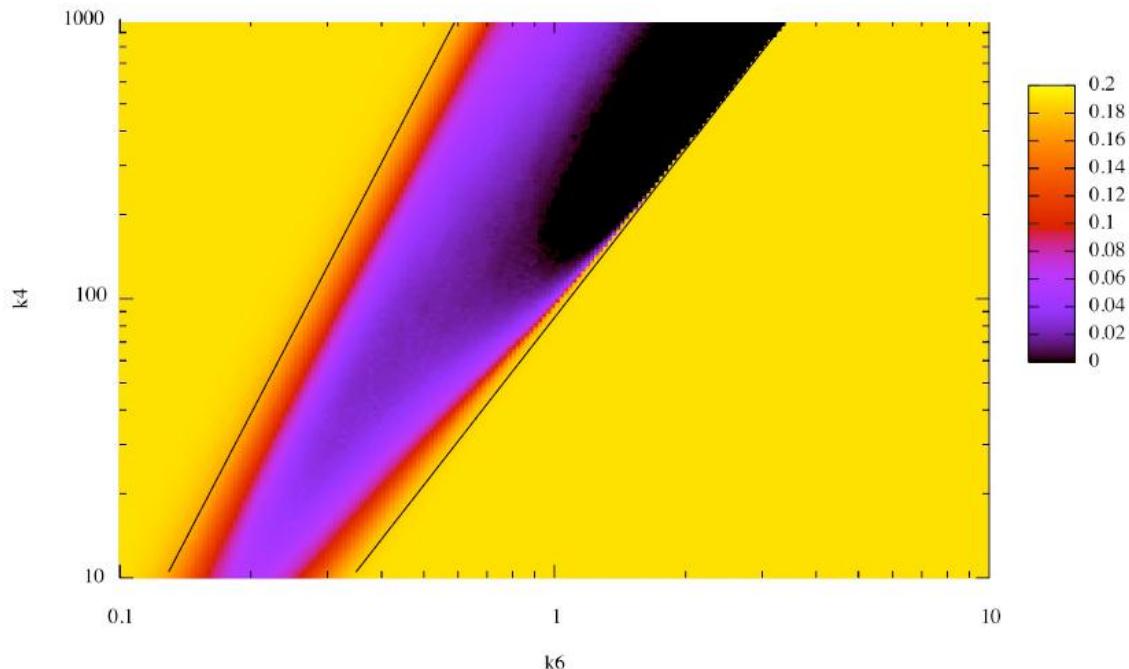
Violation degree $vd(T, \phi) = \text{distance}(\text{val}(\phi), \mathcal{D}_{\phi^*}(T))$

Satisfaction degree $sd(T, \phi) = \frac{1}{1+vd(T, \phi)} \in [0, 1]$

LTL(R) Satisfaction Degree



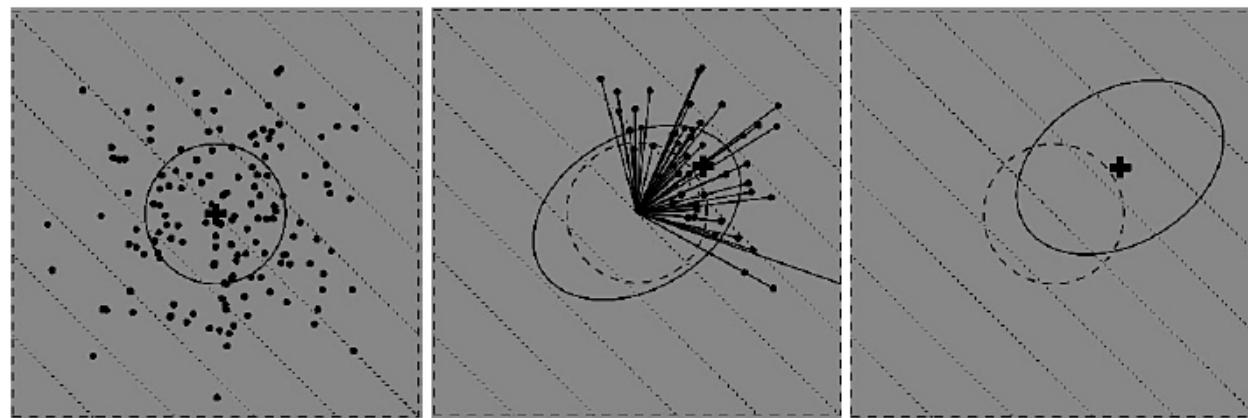
Bifurcation diagram on k_4, k_6
[Tyson 91]



Continuous satisfaction degree in [0,1]
of the LTL(R) formula for oscillation
with amplitude constraint
[Rizk Batt Fages Soliman CMSB 08]

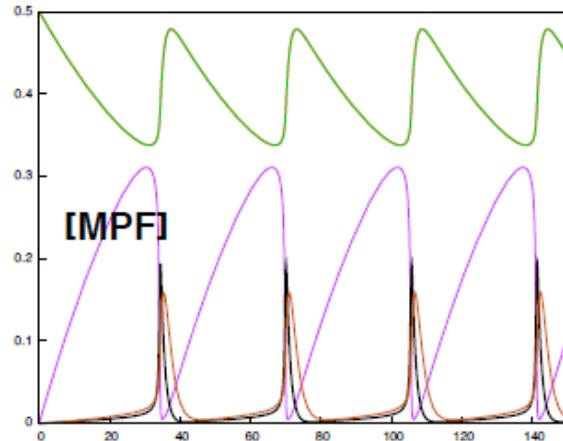
Black-box Randomized Non-linear Optimization Method

- Use existing non-linear optimization toolbox for kinetic parameter search using satisfaction degree as fitness function
- We use the state-of-the-art Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [Hansen Osermeier 01, Hansen 08]
- CMA-ES maximizes an objective function in continuous domain in a black box scenario
- CMA-ES uses a probabilistic neighborhood and updates information in covariance matrix at each move



Learning kinetic parameter values from LTL specifications

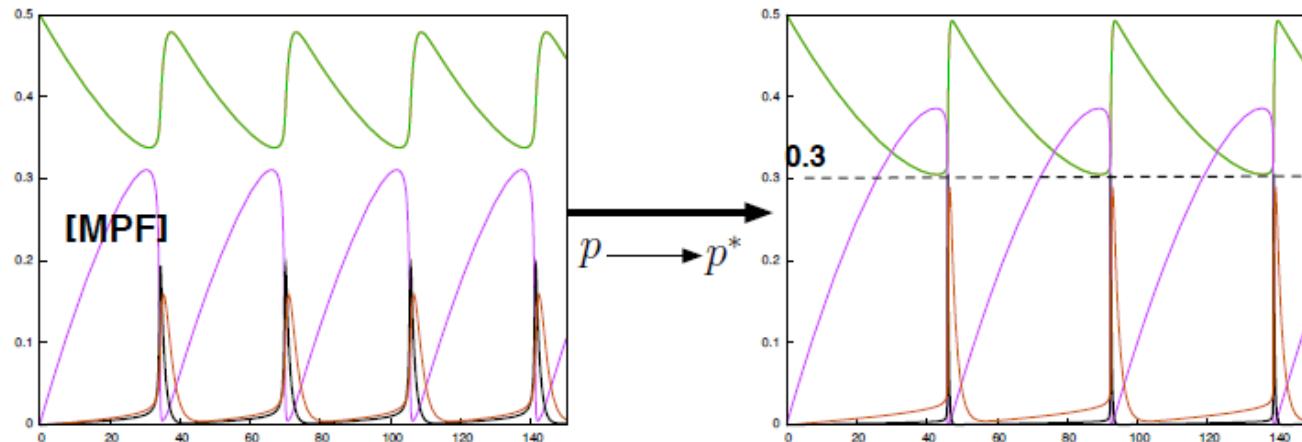
- simple model of the yeast cell cycle from [Tyson PNAS 91]
- models Cdc2 and Cyclin interactions (6 variables, 8 kinetic parameters)



- Pb : find values of 8 parameters such that amplitude is ≥ 0.3
 $\phi^*: \mathbf{F}([A]>x \wedge \mathbf{F}([A]<y))$
amplitude $z=x-y$
goal : $z = 0.3$

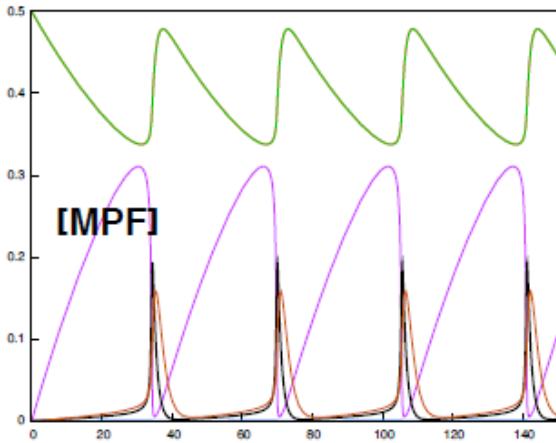
Learning kinetic parameter values from LTL specifications

- simple model of the yeast cell cycle from [Tyson PNAS 91]
- models Cdc2 and Cyclin interactions (6 variables, 8 kinetic parameters)



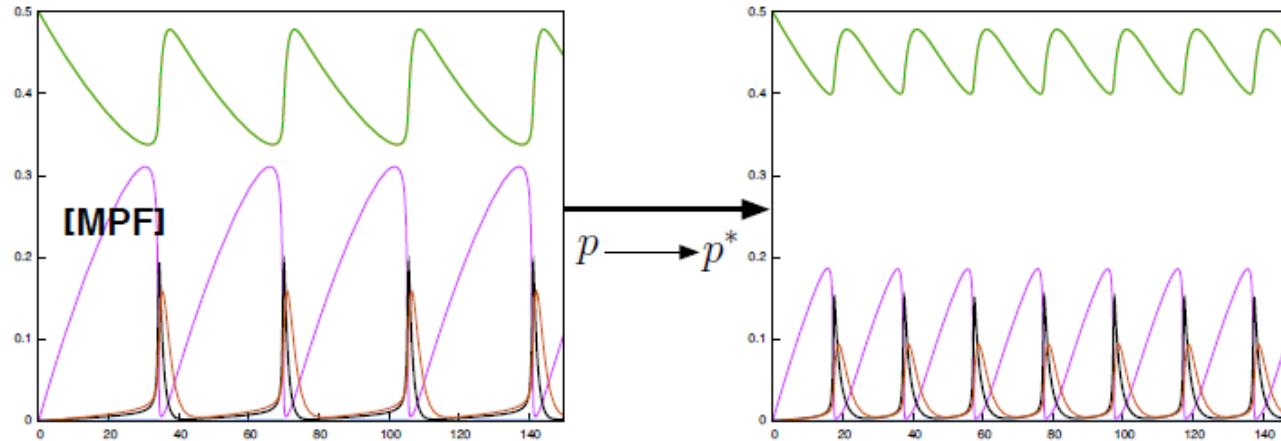
- Pb : find values of 8 parameters such that amplitude is ≥ 0.3
 $\phi^*: \mathbf{F}([A]>x \wedge \mathbf{F}([A]<y))$
amplitude $z=x-y$
goal : $z = 0.3$
- → solution found after 30s (100 calls to the fitness function)

Learning Parameter Values from Period Constraints in LTL



- Pb : find values of 8 parameters such that period is 20
 $\phi^* : \mathbf{F}(\text{MPF}_{localmaximum} \wedge \text{Time}=t1 \wedge \mathbf{F}(\text{MPF}_{localmaximum} \wedge \text{Time}=t2))$
(with $\text{MPF}_{localmaximum}$: $d([\text{MPF}])/dt > 0 \wedge \mathbf{X}(d([\text{MPF}])/dt < 0)$)
period $z=t2-t1$
goal $z=20$

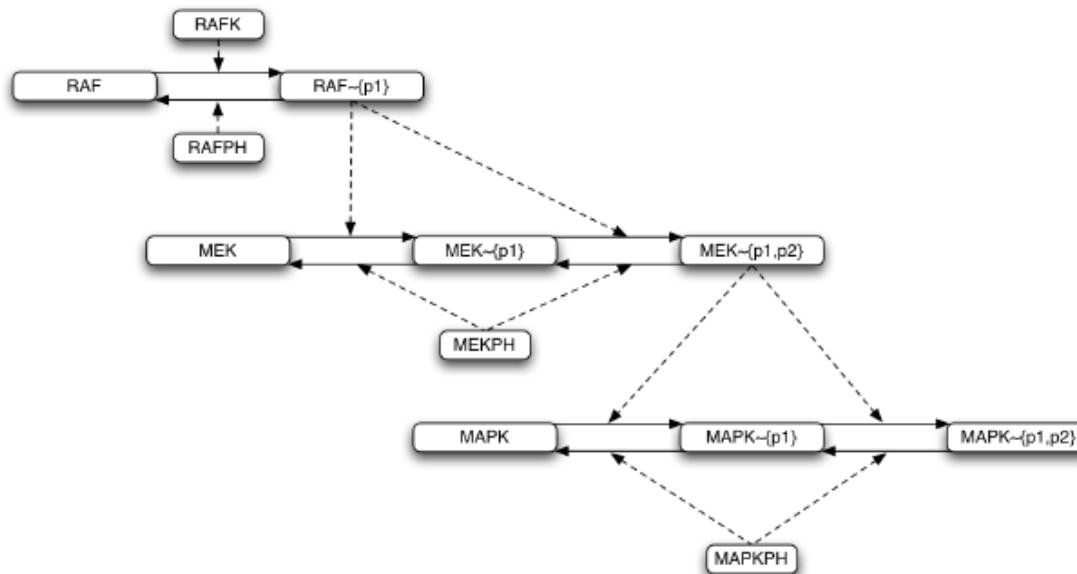
Learning Parameter Values from Period Constraints in LTL



- Pb : find values of 8 parameters such that period is 20
 $\phi^* : \mathbf{F}(\text{MPF}_{\text{localmaximum}} \wedge \text{Time}=t1 \wedge \mathbf{F}(\text{MPF}_{\text{localmaximum}} \wedge \text{Time}=t2))$
(with $\text{MPF}_{\text{localmaximum}}$: $d([\text{MPF}])/dt > 0 \wedge \mathbf{X}(d([\text{MPF}])/dt < 0)$)
period $z=t2-t1$
goal $z=20$
- → Solution found after 60s (200 calls to the fitness function)

Oscillations in MAPK signal transduction cascade

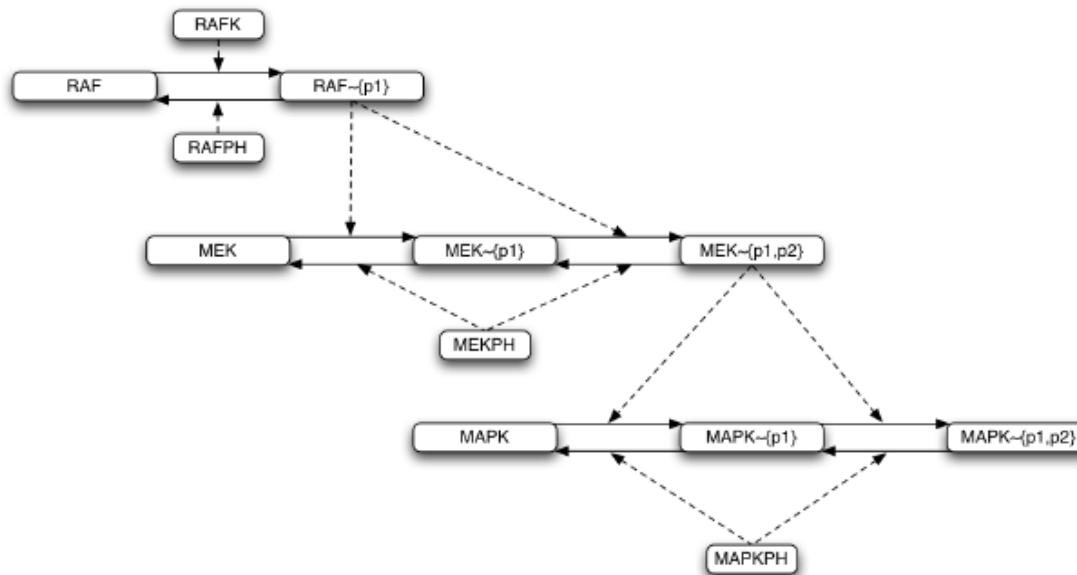
- **MAPK signaling model** [Huang Ferrel PNAS 93]



- **search for oscillations in 37 dimensions** (30 parameters and 7 initial conditions)
→ solution found after 3 min (200 calls to the fitness function)
Oscillations already observed by simulation [Qiao et al. 07]

Oscillations in MAPK signal transduction cascade

- **MAPK signaling model** [Huang Ferrel PNAS 93]



- **search for oscillations in 37 dimensions** (30 parameters and 7 initial conditions)
 - solution found after 3 min (200 calls to the fitness function)
 - Oscillations already observed by simulation [Qiao et al. 07]
- No negative feedback in the **reaction graph**, but negative circuits in the **influence graph** [Fages Soliman FMSB'08, CMSB'06]

Robustness Measure Definition

Robustness defined with respect to :

- a biological system
- a functionality property D_a
- a set P of perturbations
- General notion of robustness proposed in [Kitano MSB 07]:

$$\mathcal{R}_{a,P} = \int_{p \in P} D_a(p) \ prob(p) \ dp$$

Robustness Measure Definition

Robustness defined with respect to :

- a biological system
- a functionality property D_a
- a set P of perturbations
- General notion of robustness proposed in [Kitano MSB 07]:

$$\mathcal{R}_{a,P} = \int_{p \in P} D_a(p) \ prob(p) \ dp$$

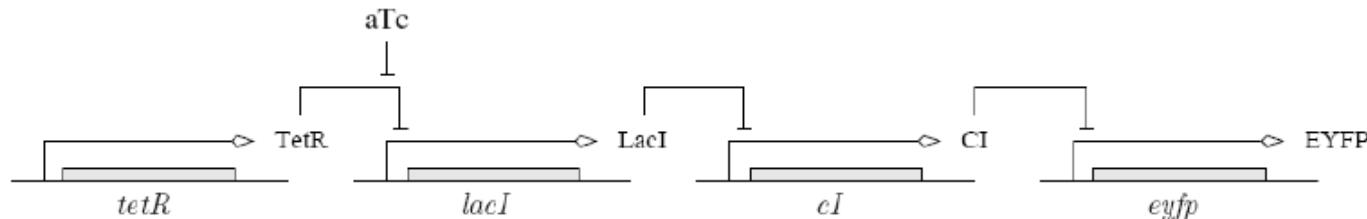
- Our computational measure of robustness w.r.t. LTL(\mathbb{R}) spec:
Given an ODE model with initial conditions, a TL formula ϕ and a set of perturbations P (on initial conditions or parameters),

$$\mathcal{R}_{\phi,P} = \int_{p \in P} sd(T(p), \phi) \ prob(p) \ dp$$

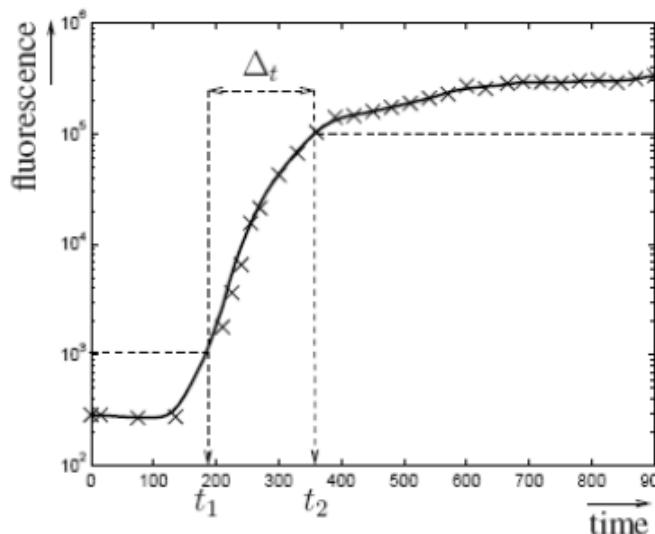
where $T(p)$ is the trace obtained by numerical integration of the ODE for perturbation p

Application to Synthetic Biology in *E. Coli*

Cascade of transcriptional inhibitions implemented in *E.coli* [Weiss et al PNAS 05]



The output protein EYFP is controlled by the small input molecule aTc



The system is well-timed if EYFP remains below 10^3 for at least 150 min., then exceeds 10^5 after at most 450 min., and switches from low to high levels in less than 150 min.

Specifying the expected behavior in LTL(\mathbb{R})

The timing specifications can be formalized in temporal logic as follows:

$$\begin{aligned}\phi(t_1, t_2) = & \quad \mathbf{G}(\text{time} < t_1 \rightarrow [\text{EYFP}] < 10^3) \\ & \wedge \quad \mathbf{G}(\text{time} > t_2 \rightarrow [\text{EYFP}] > 10^5) \\ & \wedge \quad t_1 > 150 \wedge t_2 < 450 \wedge t_2 - t_1 < 150\end{aligned}$$

which is abstracted into

$$\begin{aligned}\phi(t_1, t_2, b_1, b_2, b_3) = & \quad \mathbf{G}(\text{time} < t_1 \rightarrow [\text{EYFP}] < 10^3) \\ & \wedge \quad \mathbf{G}(\text{time} > t_2 \rightarrow [\text{EYFP}] > 10^5) \\ & \wedge \quad t_1 > b_1 \wedge t_2 < b_2 \wedge t_2 - t_1 < b_3\end{aligned}$$

for computing validity domains for b_1, b_2, b_3

with the objective $b_1 = 150, b_2 = 450, b_3 = 150$ for computing the satisfaction degree in a given trace.

Variance-based global sensitivity indices

$$S_i = \frac{\text{Var}(E(R|P_i))}{\text{Var}(R)} \in [0, 1]$$

S_γ	20.2 %	$S_{\kappa_{eyfp}, \gamma}$	8.7 %
$S_{\kappa_{eyfp}}$	7.4 %	$S_{\kappa_{cl}, \gamma}$	6.2 %
$S_{\kappa_{cl}}$	6.1 %	$S_{\kappa_{cl}^0, \gamma}$	5.0 %
$S_{\kappa_{lacI}^0}$	3.3 %	$S_{\kappa_{cl}^0, \kappa_{eyfp}}$	2.8 %
$S_{\kappa_{cl}^0}$	2.0 %	$S_{\kappa_{cl}, \kappa_{eyfp}}$	1.8 %
$S_{\kappa_{lacI}}$	1.5 %	$S_{\kappa_{eyfp}^0, \gamma}$	1.5 %
$S_{\kappa_{eyfp}^0}$	0.9 %	$S_{\kappa_{cl}^0, \kappa_{cl}}$	1.1 %
$S_{u_{aTc}}$	0.4 %	$S_{\kappa_{cl}^0, \kappa_{lacI}}$	0.5 %
total first order	40.7 %	total second order	31.2 %

degradation factor γ has the strongest impact on the cascade.

aTc variations have a very low impact

different importance of the basal κ_{eyfp}^0 and regulated κ_{eyfp} EYFP production rates

the basal production of EYFP is due to an incomplete repression of the promoter by CI (high effect of κ_{cl}) rather than a constitutive leakage of the promoter (low effect of κ_{eyfp}^0). [Rizk Batt Fages Soliman ISMB'09 Bioinformatics]

Conclusion

- New focus in Systems Biology: formal methods from Computer Science
 - Beyond diagrammatic notations: formal semantics, abstract interpretation
 - Beyond simulation: symbolic execution
 - Beyond curve fitting: high-level specifications in temporal logic
 - Automatic model-checking, parameter optimization, model reductions
- New focus in Programming: numerical methods
 - Beyond discrete machines: stochastic or continuous or hybrid dynamics
 - Quantitative transition systems, hybrid systems
 - Continuous satisfaction degree of Temporal Logic specifications, optimization
 - From model checking (back) to model synthesis [Pnueli 77]

Acknowledgments

- **Contraintes group** at INRIA Paris-Rocquencourt on this topic:
Grégory Batt, Sylvain Soliman, Szymon Stoma, Steven Gay, Thierry Martinez, Faten Nabli, Xavier Duportet, Janis Ulhendorf, Raphael Martin
- **EraNet SysBio C5Sys** (follow up of FP6 Tempo) on cancer chronotherapies, coord.
Francis Lévi, INSERM; Jean Clairambault INRIA; ...
Coupled models of cell and circadian cycles, p53/mdm2, cytotoxic drugs.
- **INRIA/INRA project Regate** coord. *F. Clément INRIA; E. Reiter, INRA Tour, D. Heitzler* Modeling of GPCR Angiotensine and FSH signaling networks
- **ANR project Calamar**, coord. *C. Chaouiya, D. Thieffry , ENS, L. Calzone, Curie...*
Modularity and Compositionality in regulatory networks.
- **OSEO Biointelligence**, coord. *Dassault-Systèmes*
- **ANR Iceberg**, coord *Grégory Batt, with P. Hersen, Physics, O Gandrillon CNRS, ...*
Model-based control of gene expression in microfluidic
- **ANR Syne2arti**, coord. *Grégory Batt, INRIA with D. Drasdo INRIA, O. Maler, CNRS Verimag, R. Weiss MIT* Model-based control of tissue growth in synthetic biology