



HAL
open science

Cells as Machines: Towards Deciphering Biochemical Programs in the Cell (keynote talk)

François Fages

► **To cite this version:**

François Fages. Cells as Machines: Towards Deciphering Biochemical Programs in the Cell (keynote talk). 21st European Conference on Artificial Intelligence, 2014, Prague, Czech Republic. hal-01103333

HAL Id: hal-01103333

<https://inria.hal.science/hal-01103333>

Submitted on 14 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cells as Machines: towards deciphering biochemical programs in the cell



François Fages
Inria Paris-Rocquencourt
<http://lifeware.inria.fr>



To tackle the complexity of biochemical reaction systems, investigate:

- Programming theory concepts
- Formal methods of circuit and program verification
- Artificial Intelligence techniques

Implemented in BIOCHAM v3.5 (Biochemical Abstract Machine)

Systems Biology Challenge

Gain system-level understanding of multi-scale biological processes in terms of their elementary interactions at the molecular level. [Kitano 1999]



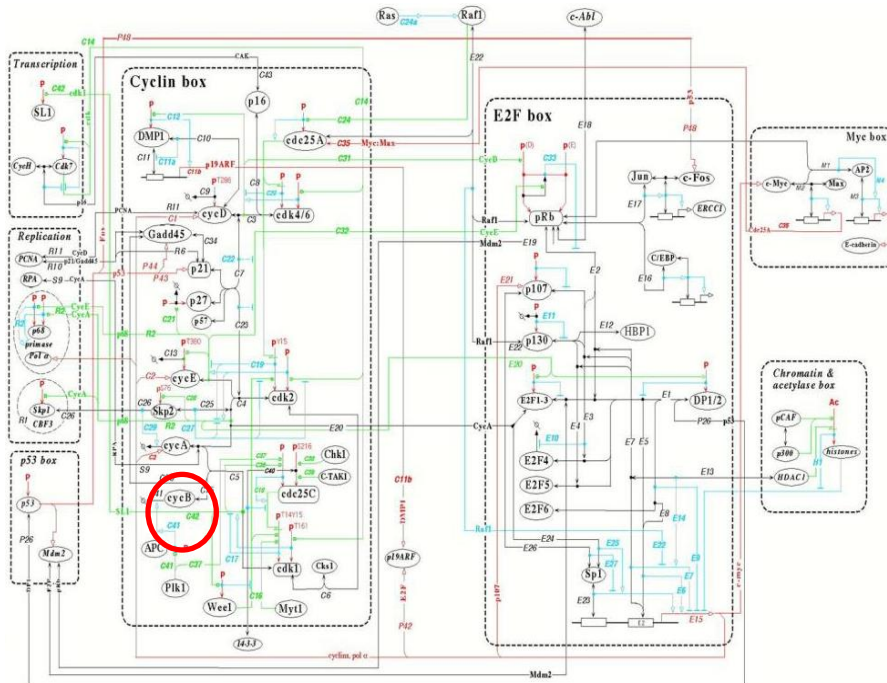
Cell mitosis videos [Lodish et al. 03]

Follow-up of Human Genome Project (90s)

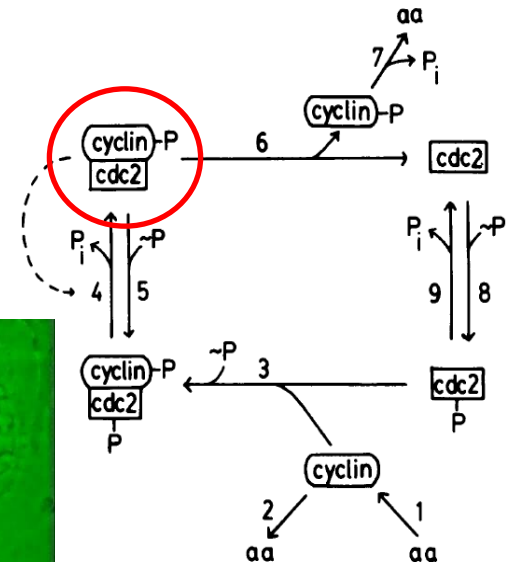
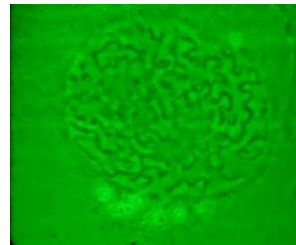
- Beyond genomic data: **protein-protein interaction** databases, RNAs, ...
- **Model repositories of cell processes** : e.g. `biomodels.net` 1000 models
- Systems Biology Markup Language (SBML): model exchange format
- Modeling environments (Cell designer, Cytoscape, Copasi, **BIOCHAM**,...)
- Simulation of a whole-cell *mycoplasma genitalium* [Karr Covert et al 12]

Biochemical reaction models

- 1) Models for representing knowledge : the more detailed the better
- 2) Models for answering questions : the more abstract the better



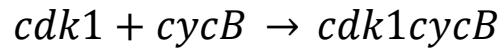
Mammalian cell cycle control interaction map [Kohn 1999]



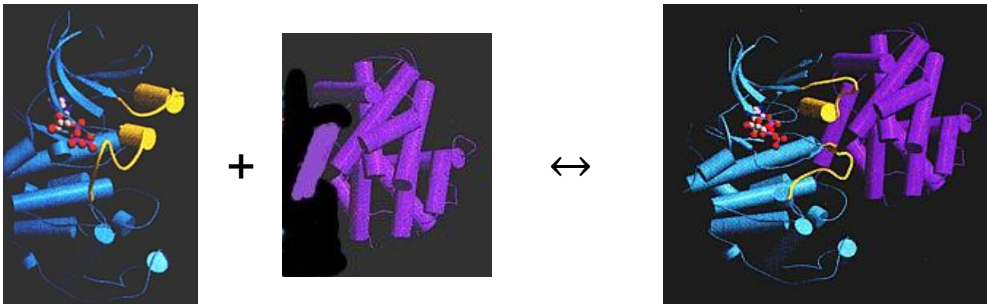
Simplified yeast cell cycle model [Tyson 1991]

Biochemical reactions

- Binding, complexation:

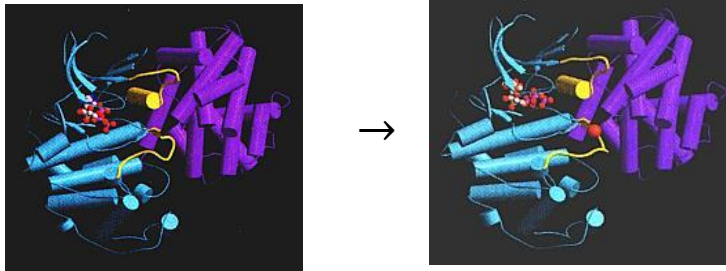


- Unbinding, decomplexation:



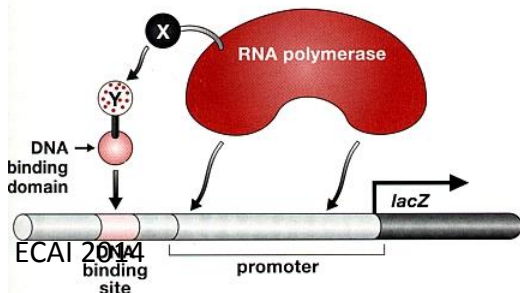
Biochemical reactions

- Binding, complexation: $A + B \rightarrow C$
 $cdk1 + cycB \rightarrow cdk1cycB$
- Unbinding, decomplexation: $A \rightarrow B + C$
- Transformation, phosphorylation, transport: $A \rightarrow B$ $(A + K \rightarrow C \rightarrow B + K)$
 $cdk1cycB \rightarrow cdk1cycBp$



Biochemical reactions

- Binding, complexation:** $A + B \rightarrow C$
 $cdk1 + cycB \rightarrow cdk1cycB$
- Unbinding, decomplexation:** $A \rightarrow B + C$
- Transformation, phosphorylation, transport:** $A \rightarrow B$ $(A + K \rightarrow C \rightarrow B + K)$
 $cdk1cycB \rightarrow cdk1cycBp$
- Gene expression, synthesis:** $A \rightarrow A + B$
 $E2Fa \rightarrow E2Fa + RNAcycA$



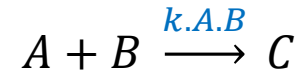
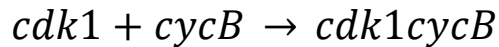
Biochemical reactions

- **Binding, complexation:** $A + B \rightarrow C$
 $cdk1 + cycB \rightarrow cdk1cycB$
- **Unbinding, decomplexation:** $A \rightarrow B + C$
- **Transformation, phosphorylation, transport:** $A \rightarrow B$ $(A + K \rightarrow C \rightarrow B + K)$
 $cdk1cycB \rightarrow cdk1cycBp$
- **Gene expression, synthesis:** $A \rightarrow A + B$
 $E2Fa \rightarrow E2Fa + RNAcycA$
- **Degradation:** $A \rightarrow _$

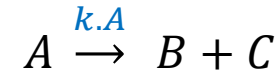
Biochemical reaction rates

Time matters

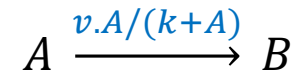
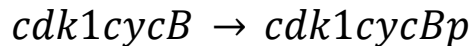
- Binding, complexation:



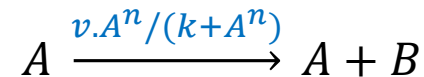
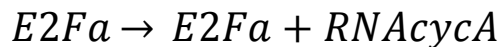
- Unbinding, decomplexation:



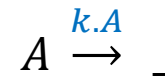
- Transformation, phosphorylation, transport:



- Gene expression, synthesis:



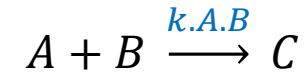
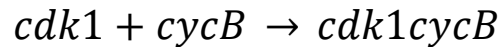
- Degradation:



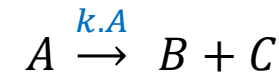
Biochemical reaction rates

Time matters

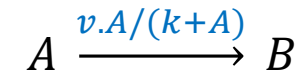
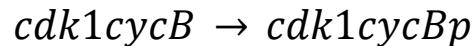
- Binding, complexation:



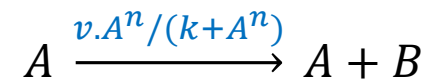
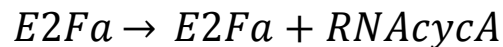
- Unbinding, decomplexation:



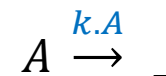
- Transformation, phosphorylation, transport:



- Gene expression, synthesis:



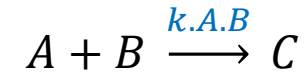
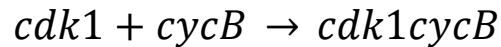
- Degradation:



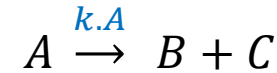
Biochemical reaction rates

Time matters

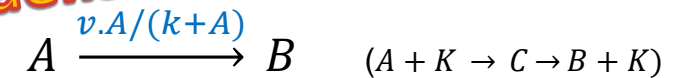
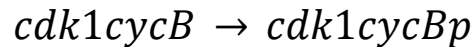
- Binding, complexation:



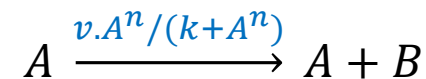
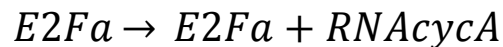
- Unbinding, decomplexation:



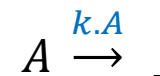
- Transformation, phosphorylation, transport:



- Gene expression, synthesis:



- Degradation:

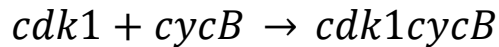


Mass action law kinetics

Michaelis Menten kinetics

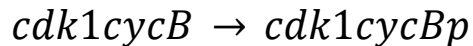
Biochemical reaction rates

- Binding, complexation:

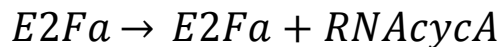


- Unbinding, decomplexation:

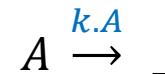
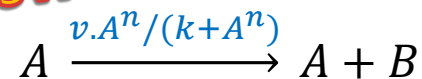
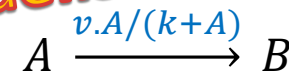
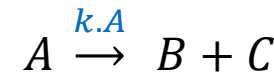
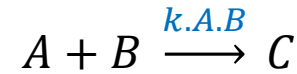
- Transformation, phosphorylation, transport:



- Gene expression, synthesis:



- Degradation:



Time matters

Mass action law kinetics

Michaelis Menten kinetics

Hill sigmoid: switch like behavior

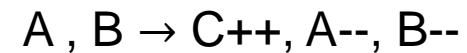
Semantics of Reactions $A+B \xrightarrow{f(A,B)} C$

Boolean Semantics: presence-absence of molecules $A \wedge B \rightarrow C \wedge A/\neg A \wedge B/\neg B$
Asynchronous Transition System

Semantics of Reactions $A+B \xrightarrow{f(A,B)} C$

Boolean Semantics: presence-absence of molecules $A \wedge B \rightarrow C \wedge A/\neg A \wedge B/\neg B$
Asynchronous Transition System

Petri Net Semantics: numbers of molecules
Multiset rewriting



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

Semantics of Reactions $A+B \xrightarrow{f(A,B)} C$

Boolean Semantics: presence-absence of molecules $A \wedge B \rightarrow C \wedge A/\neg A \wedge B/\neg B$
Asynchronous Transition System

Petri Net Semantics: numbers of molecules $A, B \rightarrow C++, A--, B--$
Multiset rewriting
CHAM [Berry Boudol 90] [Banatre Le Metayer 86]

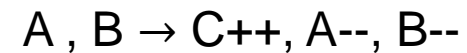
Stochastic Semantics: reaction probabilities, time of next reaction in a given state S_i
Continuous Time Markov Chain (CTMC) $A, B \xrightarrow{p(S_i), t(S_i)} C++, A--, B--$

Semantics of Reactions $A+B \xrightarrow{f(A,B)} C$

Boolean Semantics: presence-absence of molecules $A \wedge B \rightarrow C \wedge A/\neg A \wedge B/\neg B$
Asynchronous Transition System

Logical time

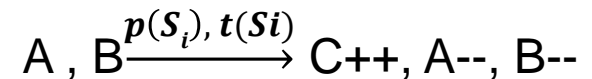
Petri Net Semantics: numbers of molecules
Multiset rewriting



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

Stochastic Semantics: reaction probabilities, time of next reaction in a given state S_i

Continuous Time Markov Chain (CTMC)



Real time

Semantics of Reactions $A+B \xrightarrow{f(A,B)} C$

Boolean Semantics: presence-absence of molecules $A \wedge B \rightarrow C \wedge A/\neg A \wedge B/\neg B$
Asynchronous Transition System

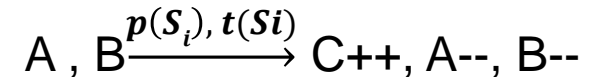
Logical time

Petri Net Semantics: numbers of molecules $A, B \rightarrow C++, A--, B--$
Multiset rewriting

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

Stochastic Semantics: reaction probabilities, time of next reaction in a given state S_i

Continuous Time Markov Chain (CTMC)



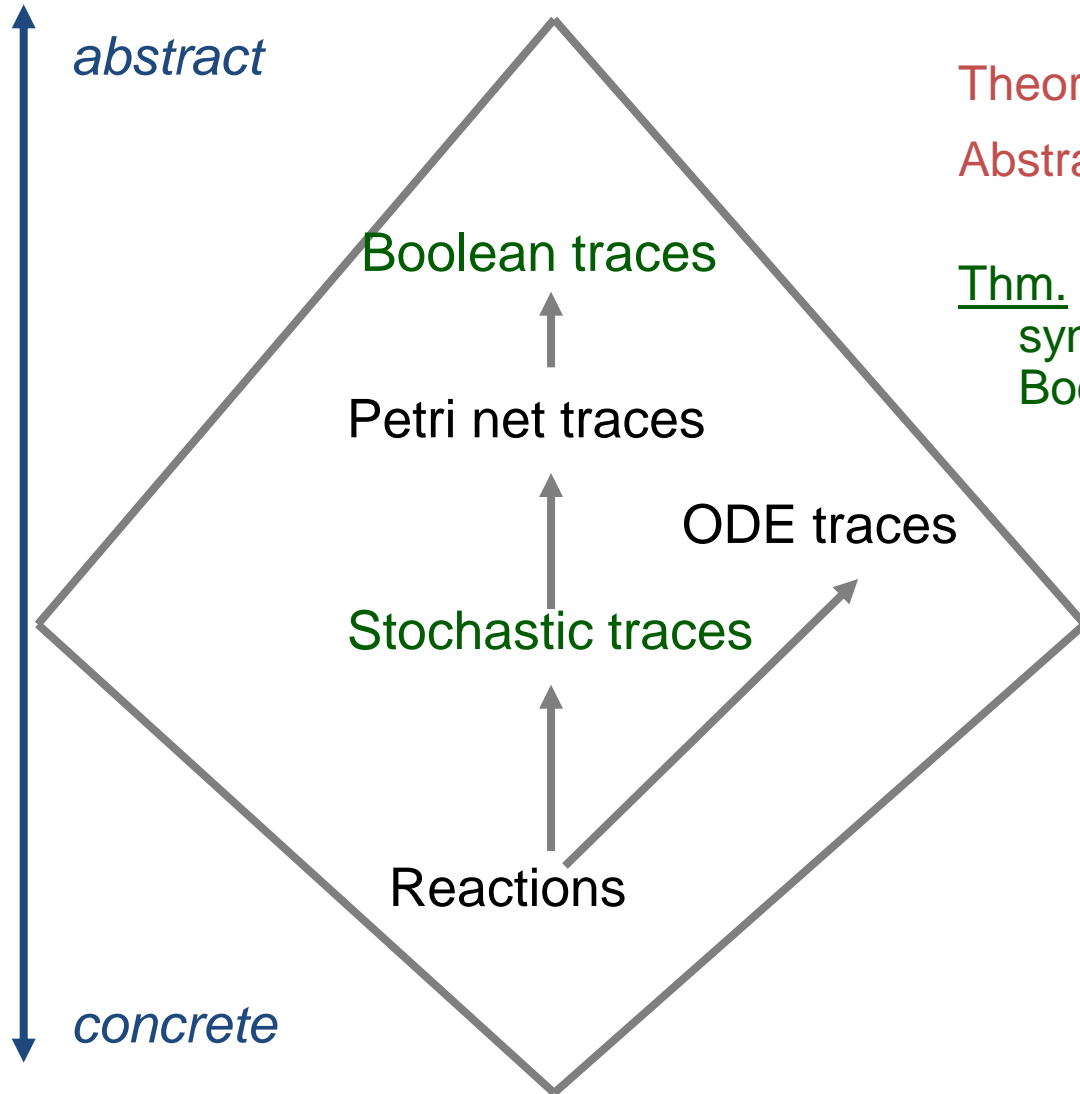
Real time

Continuous Semantics: concentrations, continuous evolution

Ordinary Differential Equations (ODE)

$$\dot{A}_i = \sum_{r=1}^n f_r \times \delta_r(A_i)$$

Abstraction Relationships



Theory of abstract Interpretation

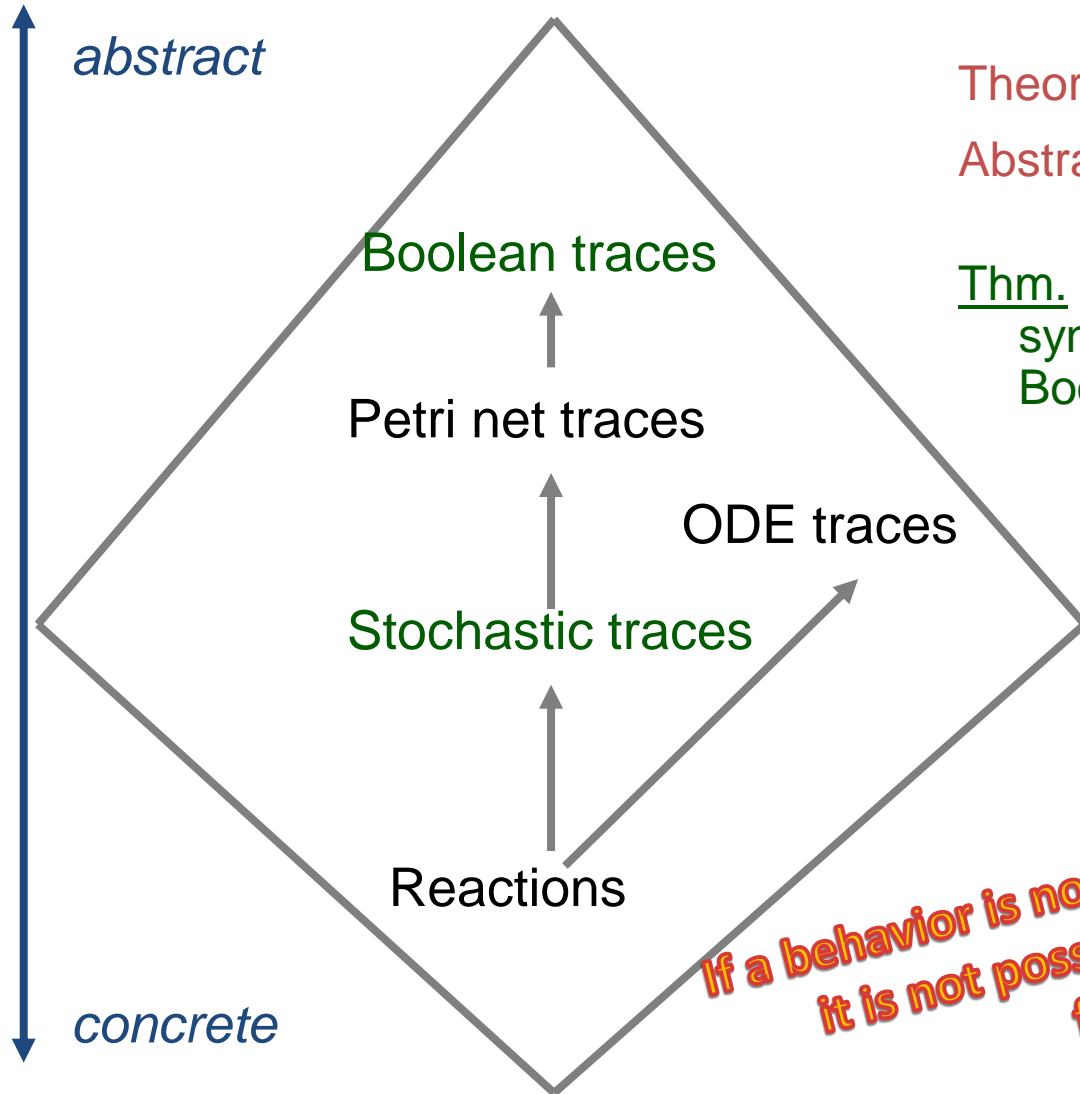
Abstractions as Galois connections

[Cousot Cousot POPL'77]

Thm. Galois connections between the
syntactical, stochastic, Petri Net and
Boolean semantics

[FF Soliman CMSB'06, TCS'08]

Abstraction Relationships



Theory of abstract Interpretation

Abstractions as Galois connections

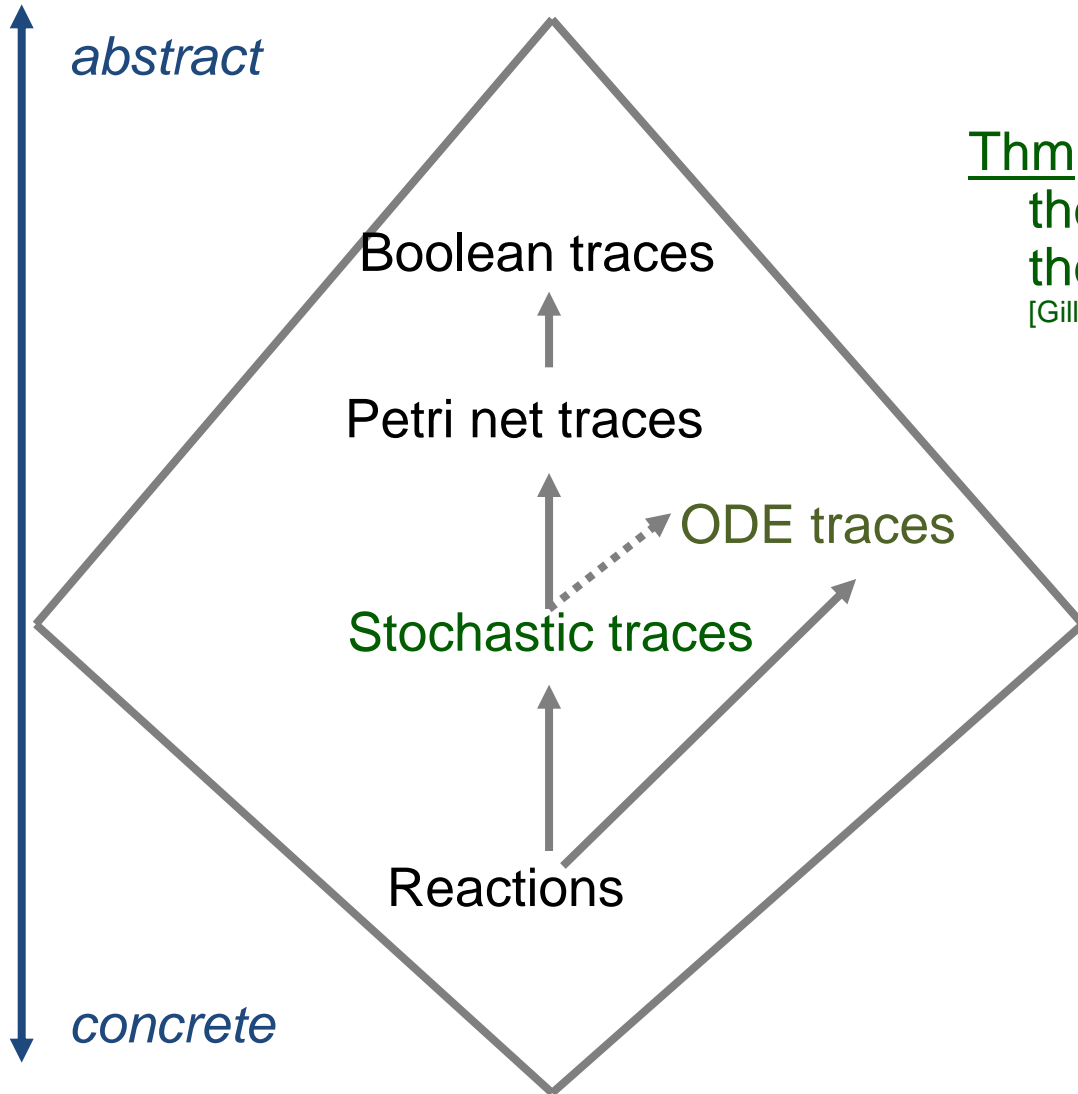
[Cousot Cousot POPL'77]

Thm. Galois connections between the syntactical, stochastic, Petri Net and Boolean semantics

[FF Soliman CMSB'06, TCS'08]

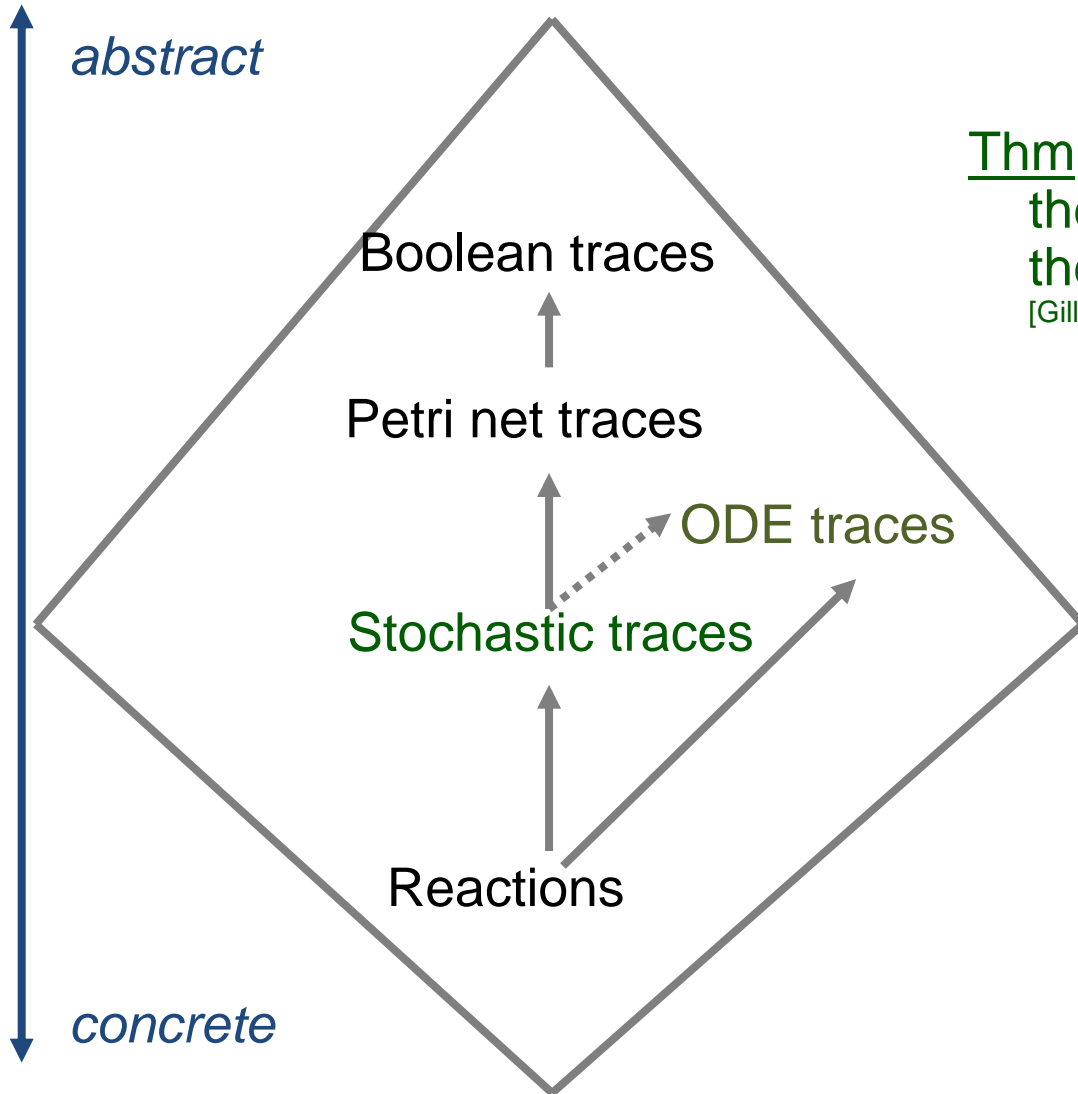
**If a behavior is not possible in the Boolean semantics
it is not possible in the stochastic semantics
for any reaction rates**

Abstraction Relationships



Thm. Under large number conditions
the ODE semantics approximates
the mean stochastic behavior
[Gillespie 71]

Abstraction Relationships

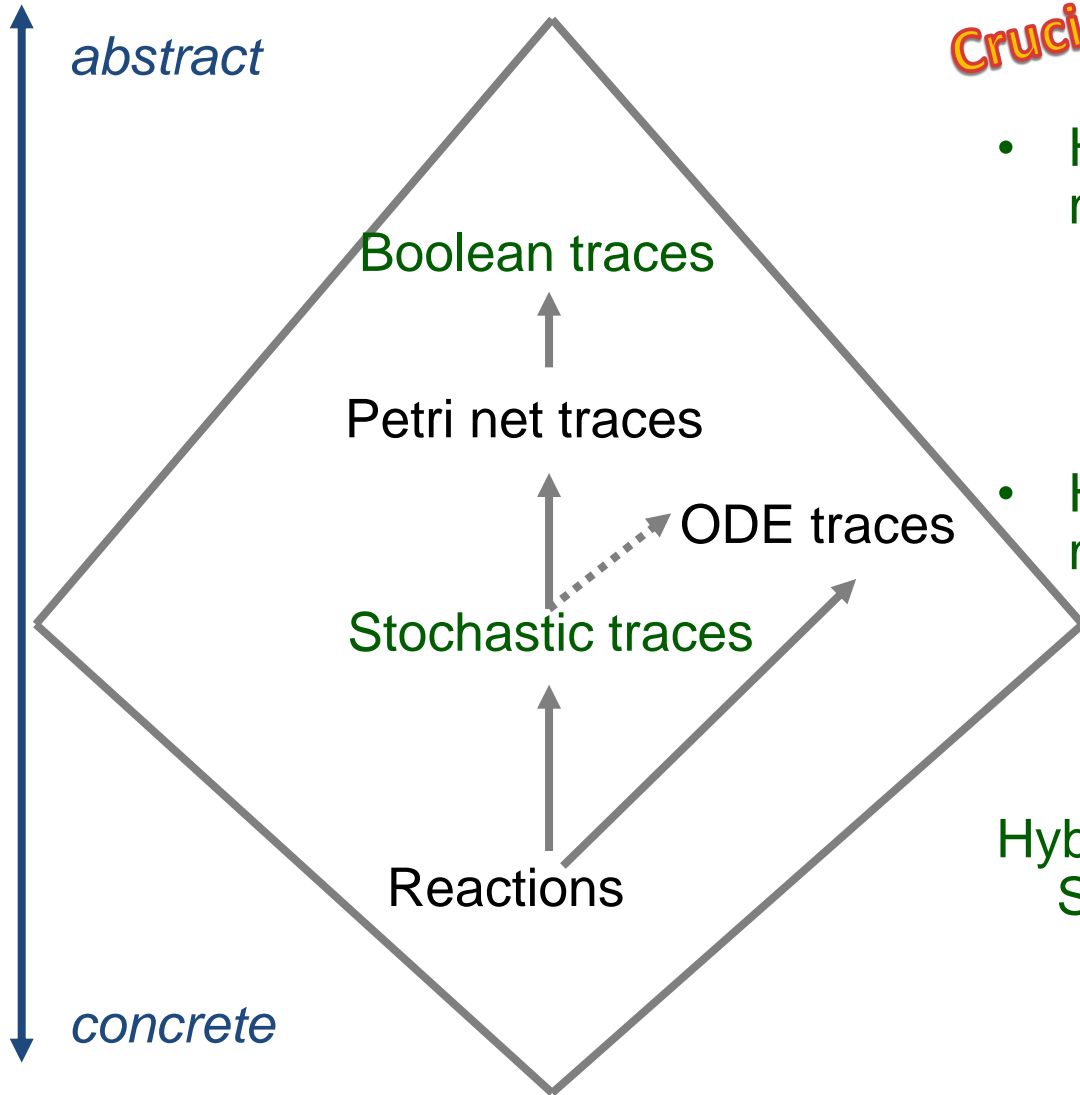


Thm. Under large number conditions
the ODE semantics approximates
the mean stochastic behavior
[Gillespie 71]

Hot topic:
higher order moments
ODE for mean, variance,...

Hybrid Models

Crucial for model reduction and multi-scale models



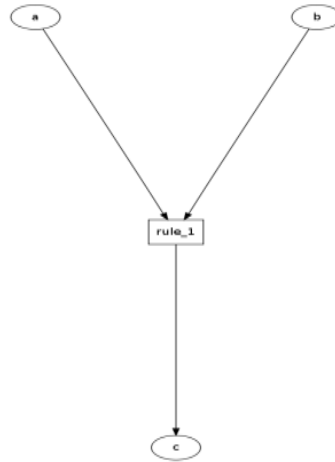
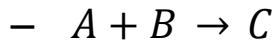
- Hybrid Boolean-continuous models (hybrid automata)
Boolean gene expression + continuous protein activation
- Hybrid stochastic-continuous models (CTMC+ODE)
Stochastic gene expression + continuous protein activation

Hybrid simulators for free with SBML's reactions+events

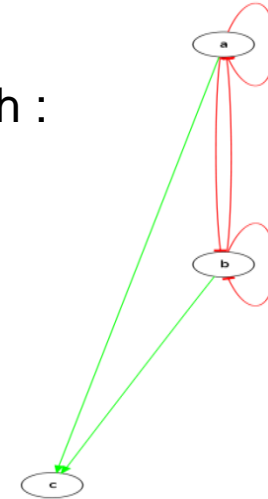
[Chiang FF Huang Soliman 13 cmsb]

Before going to simulations: structural analyses

- Reaction hypergraph :



- Influence graph :



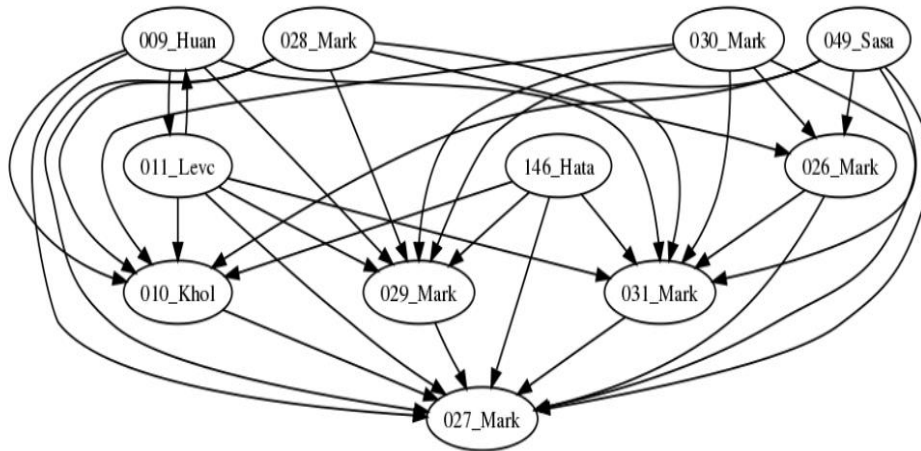
- Large independence of the influence graph from the kinetics [FF Soliman 08 fmsb]
- Existence of positive circuits in the influence graph
 - **necessary condition for multi-stationarity** (cell differentiation, homeostasis) in Boolean and ODE semantics [Soliman 13 bmb, Soule 03, Thomas 82]
- Existence of negative circuits in the influence graph
 - **necessary condition for oscillations** in ODE semantics [Snoussi 93, Thomas 82]

Structural analysis: reaction graph reduction

- Subgraph epimorphisms (SEPI) [Gay FF Martinez Soliman 13 dam]
 - Structural notion of model reduction by **deletion & merge of species & reactions** in reaction hypergraphs

Structural analysis: reaction graph reduction

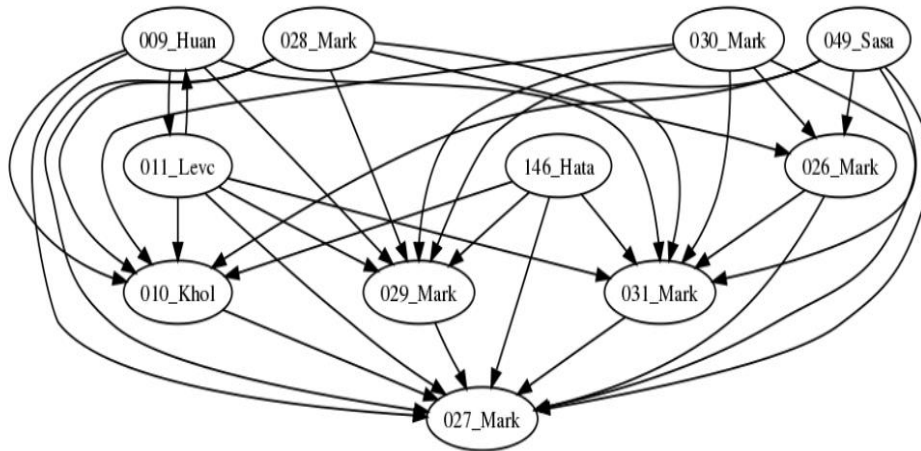
- Subgraph epimorphisms (SEPI) [Gay FF Martinez Soliman 13 dam]
 - Structural notion of model reduction by **deletion & merge of species & reactions** in reaction hypergraphs
- Sufficient to detect model reductions in biomodels.net
 - Reconstructed hierarchy of MAPK models in **biomodels.net** [Gay Soliman FF 10 bi]



Possibility to query model repositories by the structure of reaction networks

Structural analysis: reaction graph reduction

- Subgraph epimorphisms (SEPI) [Gay FF Martinez Soliman 13 dam]
 - Structural notion of model reduction by **deletion & merge of species & reactions** in reaction hypergraphs
- Sufficient to detect model reductions in biomodels.net
 - Reconstructed hierarchy of MAPK models in **biomodels.net** [Gay Soliman FF 10 bi]



Possibility to query model repositories by the structure of reaction networks

- The SEPI existence problem is NP-complete
Quite efficient computation of SEPI by **CLP(FD) /SAT solvers** [Gay FF Santini Soliman 13 wcb]

Structural analysis: Petri net invariants

- P-invariants: linear invariants on the number of tokens in all reachable markings
 - [linear conservation laws](#), elimination of variables in ODE

Structural analysis: Petri net invariants

- P-invariants: linear invariants on the number of tokens in all reachable markings
 - [linear conservation laws](#), elimination of variables in ODE
- T-invariants: cyclic behaviors
 - extreme fluxes in metabolic network (main tool for analysis and optimization)

Structural analysis: Petri net invariants

- P-invariants: linear invariants on the number of tokens in all reachable markings
 - [linear conservation laws](#), elimination of variables in ODE
- T-invariants: cyclic behaviors
 - extreme fluxes in metabolic network (main tool for analysis and optimization)
- Siphons: set of molecules that once empty remain empty
- Traps: set of molecules that once populated remain populated
 - Sufficient and necessary condition for liveness (absence of deadlock on transitions)

Structural analysis: Petri net invariants

- P-invariants: linear invariants on the number of tokens in all reachable markings
 - [linear conservation laws](#), elimination of variables in ODE
- T-invariants: cyclic behaviors
 - extreme fluxes in metabolic network (main tool for analysis and optimization)
- Siphons: set of molecules that once empty remains empty
- Traps: set of molecules that once populated remains populated
 - Sufficient and necessary condition for liveness (absence of deadlock on transitions)

The existence of a minimal siphon containing a given place is NP-complete

However, very efficient enumeration of all minimal siphons in all biomodels.net
by [CLP\(FD\)](#) and/or [SAT solvers](#) [Nabli Martinez FF Soliman 12 CP]

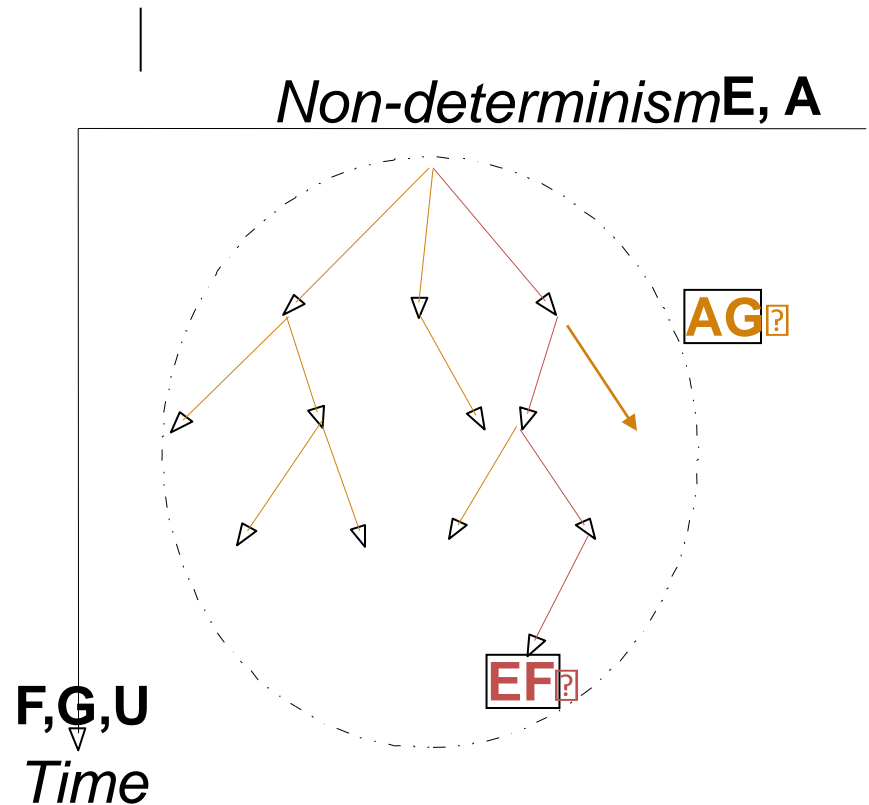
Structural analysis of dynamical behaviors

- Temporal logic FO patterns
 - Stability $\mathbf{G}\phi$
 - Reachability $\mathbf{F}\phi$, thresholds $\mathbf{F}([A]>0.1)$,
 - Peaks of concentration $\mathbf{F}([A]<V \wedge \mathbf{X}([A]=V \wedge \mathbf{X}([A]<V))$
 - Amplitude
 - Periods and phases as distance between peaks
 - ...
- More flexible than curve fitting, can deal with imprecise data
- Constraints for model validation/synthesis w.r.t. experimental observations
- Constraints for parameter search
 - Good fit \rightarrow model-based predictions
 - No good fit \rightarrow revisit structure \rightarrow model-based contribution to biology

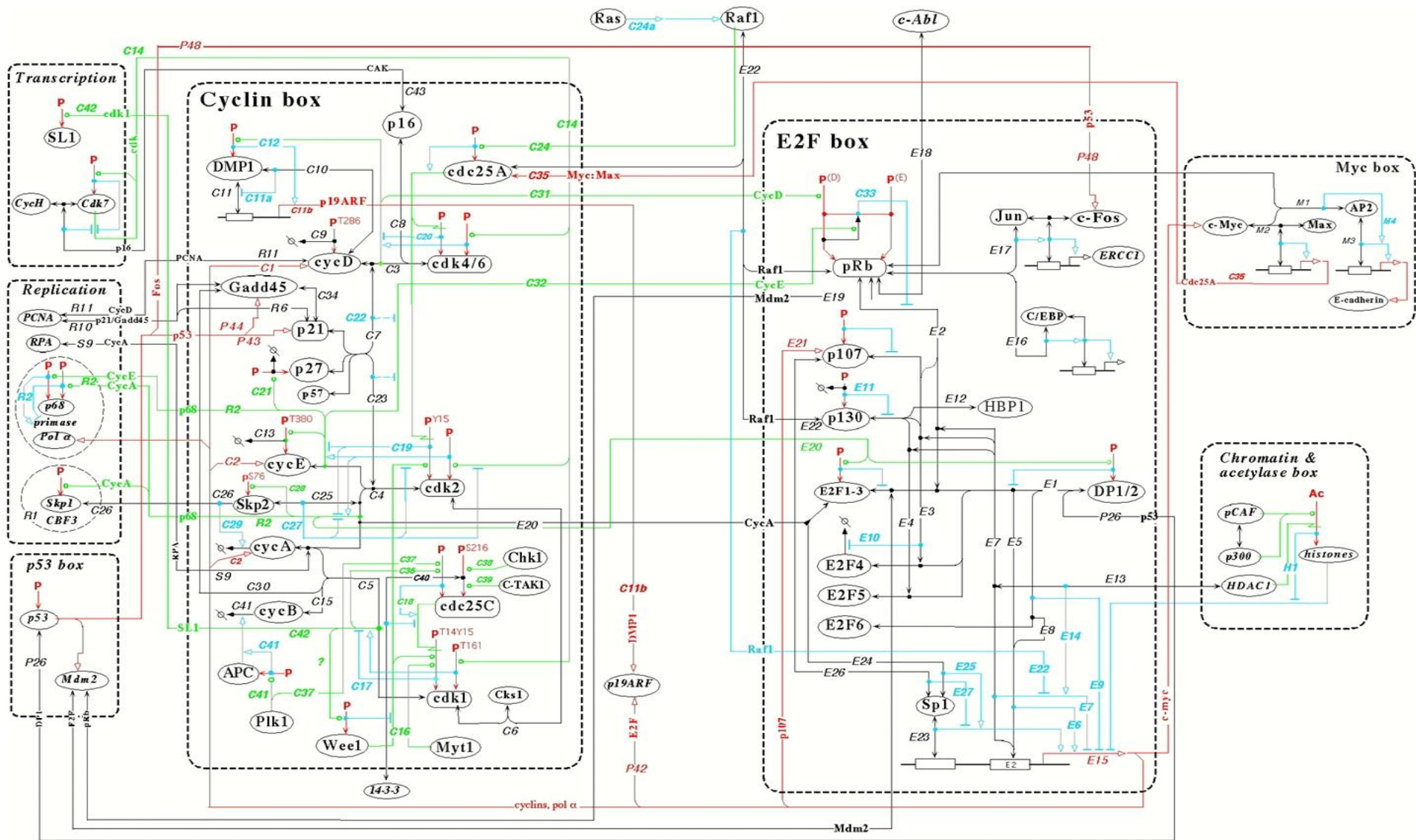
Boolean Semantics Queries

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 (φ_1 U φ_2)	A (φ_1 U φ_2)



Mammalian Cell Cycle Control Map [Kohn 99]



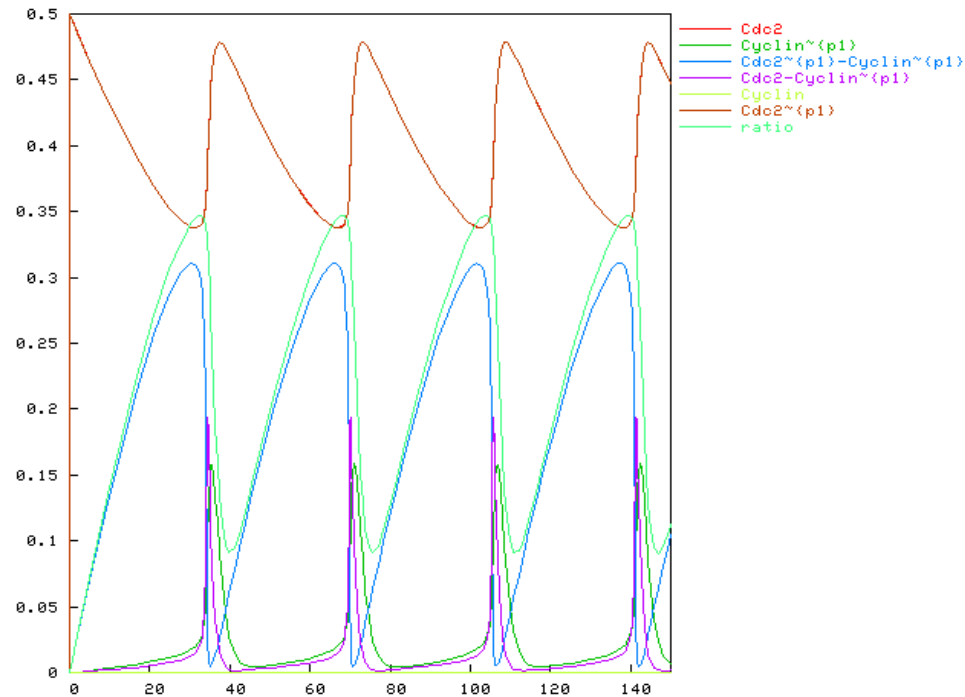
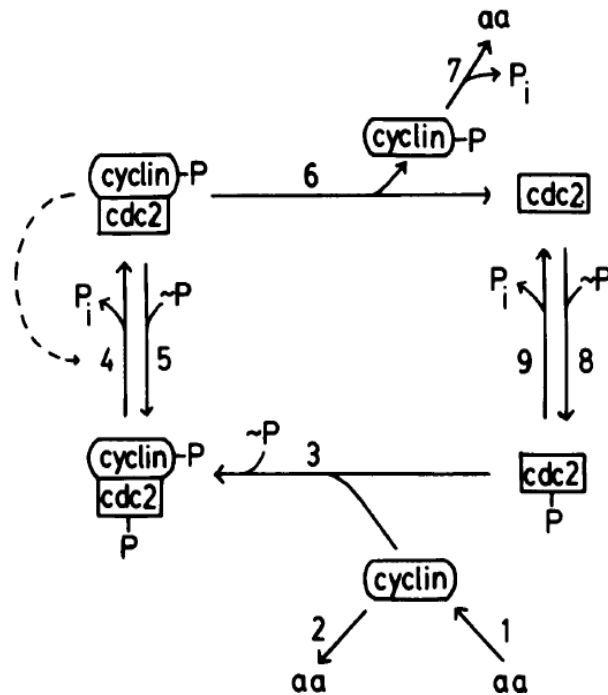
Kohn's Map Boolean Model-Checking

800 reactions, 165 proteins and genes, 500 variables, 2^{500} states.

Biocham NuSMV symbolic model-checker time in seconds [Chabrier et al. TCS 04]

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 EF (\neg Cdc25~{Nterm} U Cdk1~{Thr161}-CycB)	2.2
Oscillations CycA	EG ((EF \neg CycA) & (EF CycA))	31.8
Oscillations CycB	EG ((EF \neg CycB) & (EF CycB))	6

Dynamical Model of Cell Cycle Control [Tyson 91]



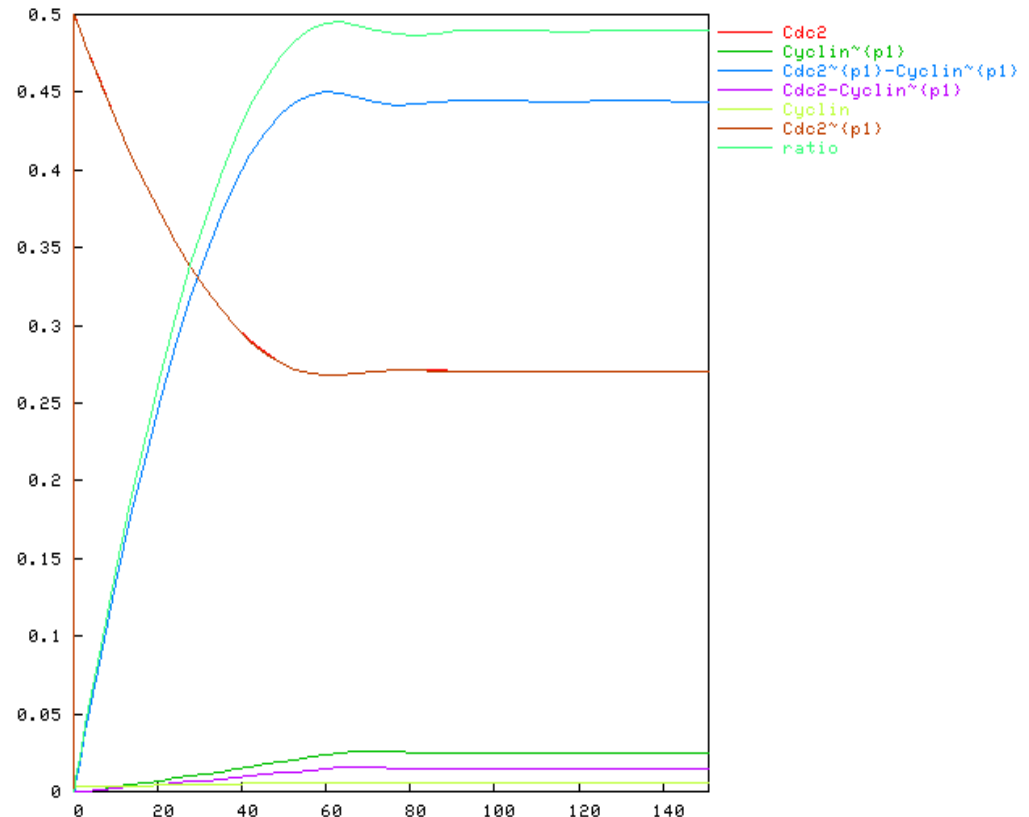
- 6 molecular species
- 8 reactions
- 8 kinetic parameters

ODE semantics

Parameter Search from LTL(R) Properties

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

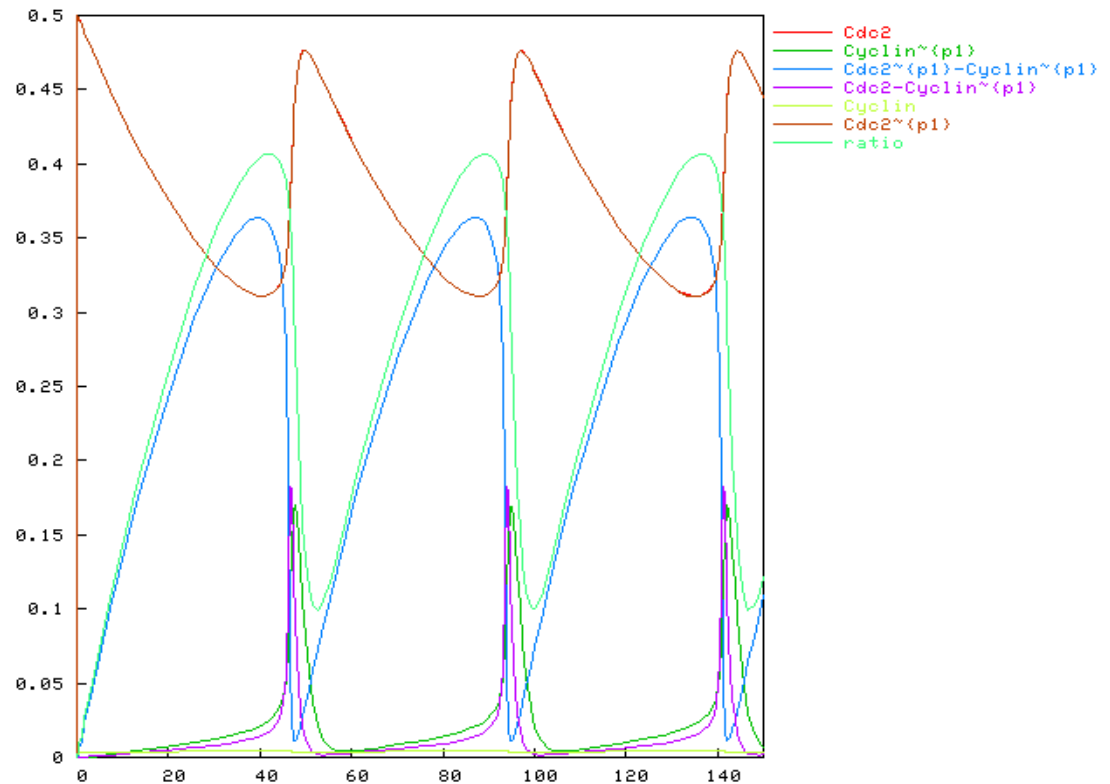
First values found :
parameter(k3,10).
parameter(k4,70).



Parameter Search from LTL(R) Properties

biocham: search_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).

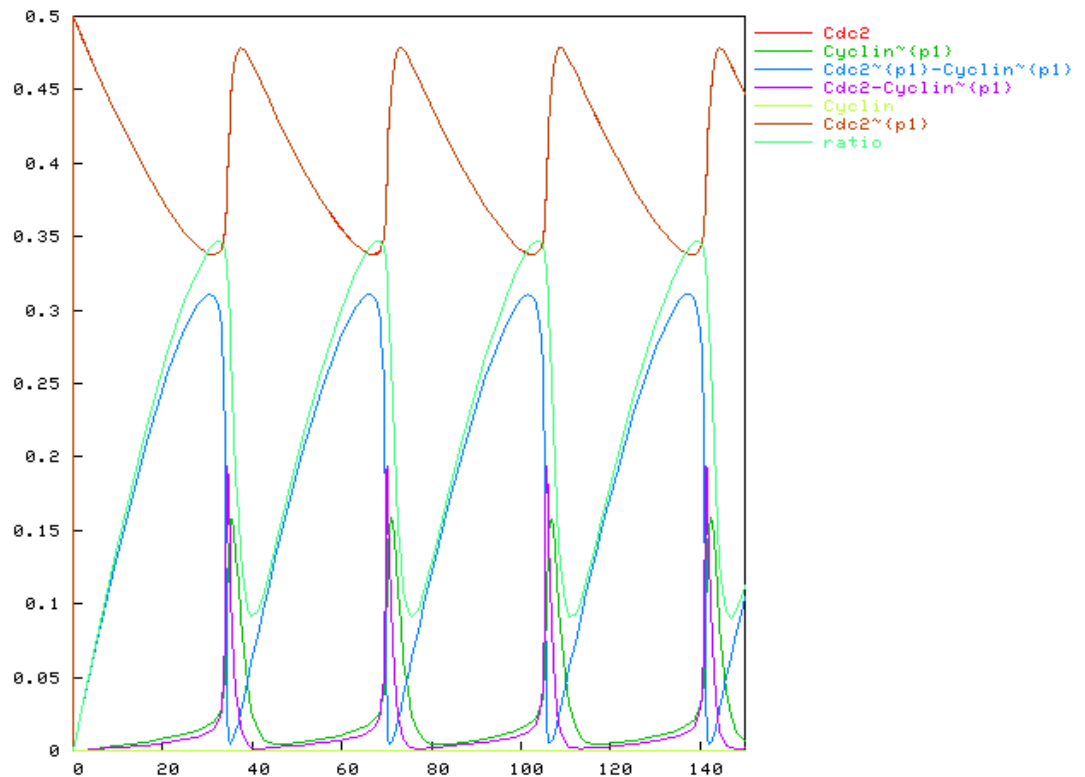


Parameter Search from LTL(R) Properties

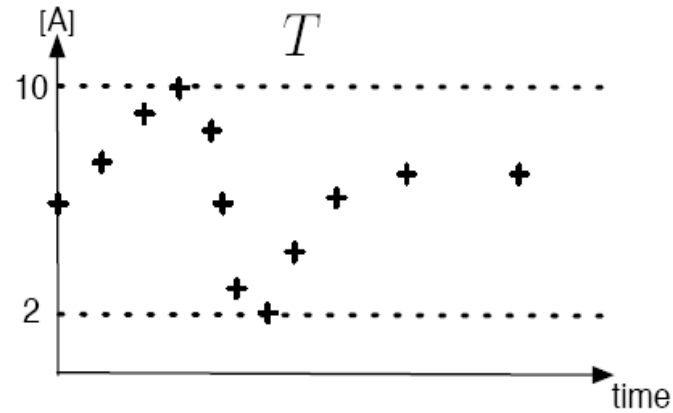
biocham: search_parameter([k3,k4],[(0,200),(0,200)],20,
period(Cdc2-Cyclin~{p1},35), 150).

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

**But parameter scanning
does not scale up
beyond 3 parameters...**



Model-Checking Generalized to Constraint Solving



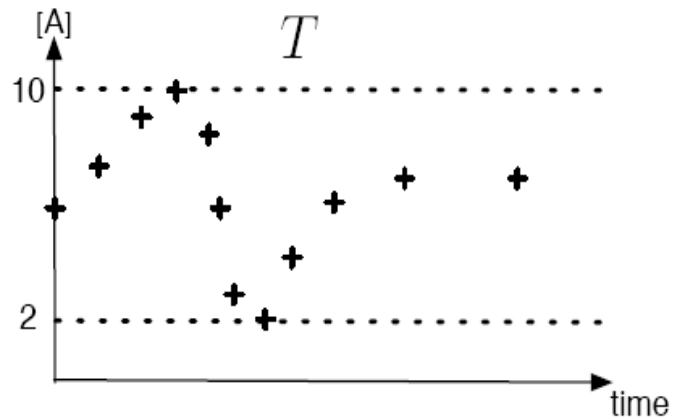
$LTL(\mathbb{R})$

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

Model-checking

the formula is false

Model-Checking Generalized to Constraint Solving



$LTL(\mathbb{R})$

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

Model-checking

the formula is false

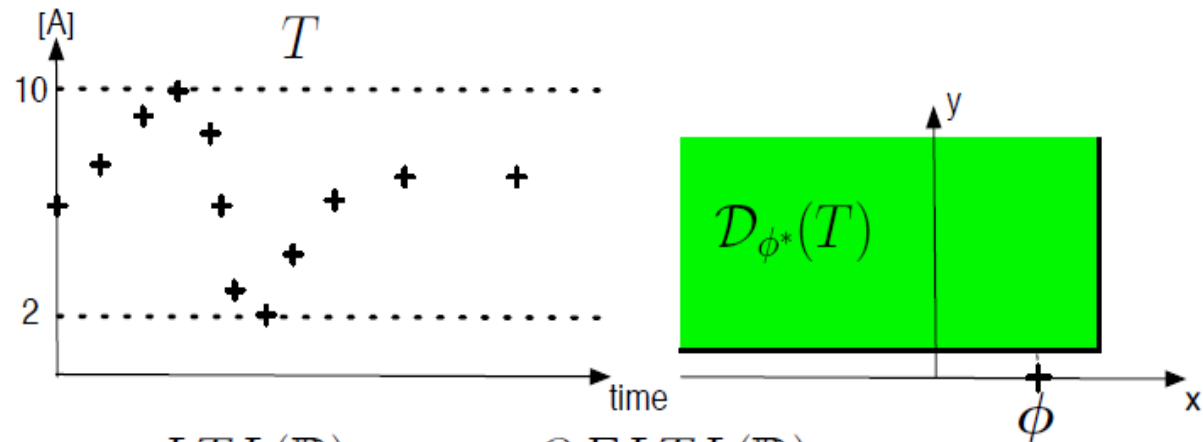
$QFLTL(\mathbb{R})$

$$\Phi^* = F([A] \geq x) \wedge F([A] \leq y)$$

Constraint solving

the formula is true for any
 $x \leq 10 \wedge y \geq 2$

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 $vd=2$ $sd=1/3$

$QFLTL(\mathbb{R})$

$$\Phi^* = F([A] \geq x) \wedge F([A] \leq y)$$

Constraint solving

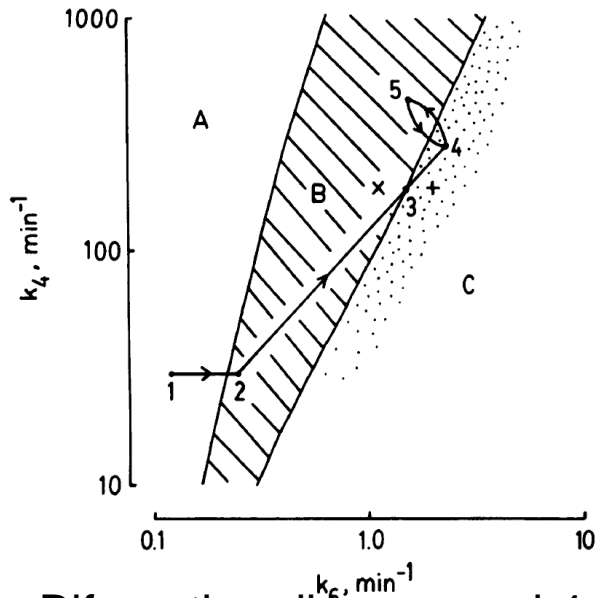
the formula is true for any $x \leq 10 \wedge y \geq 2$

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

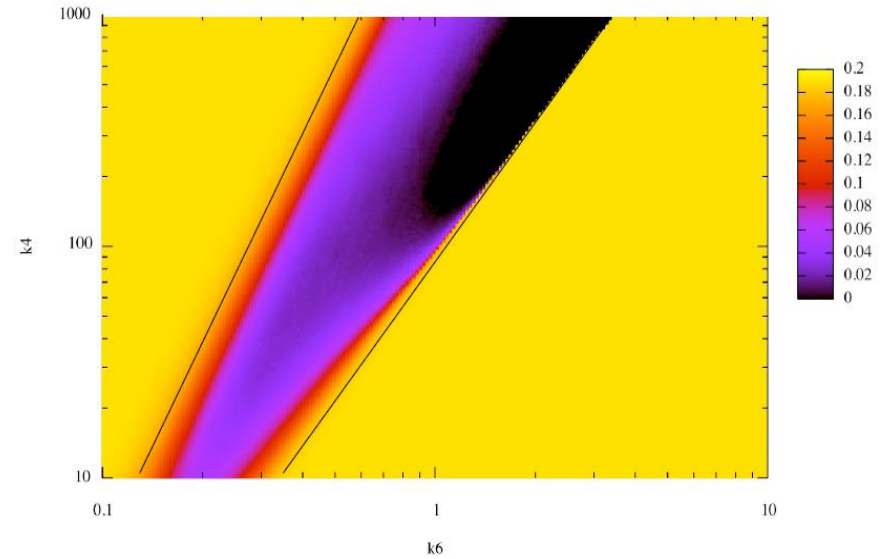
Violation degree $vd(T, \phi) = \text{distance}(\text{val}(\phi), \mathcal{D}_{\phi^*}(T))$

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

LTL(R) Continuous Satisfaction Degree in [0,1]



Bifurcation diagram on k_4 , k_6
[Tyson 91]

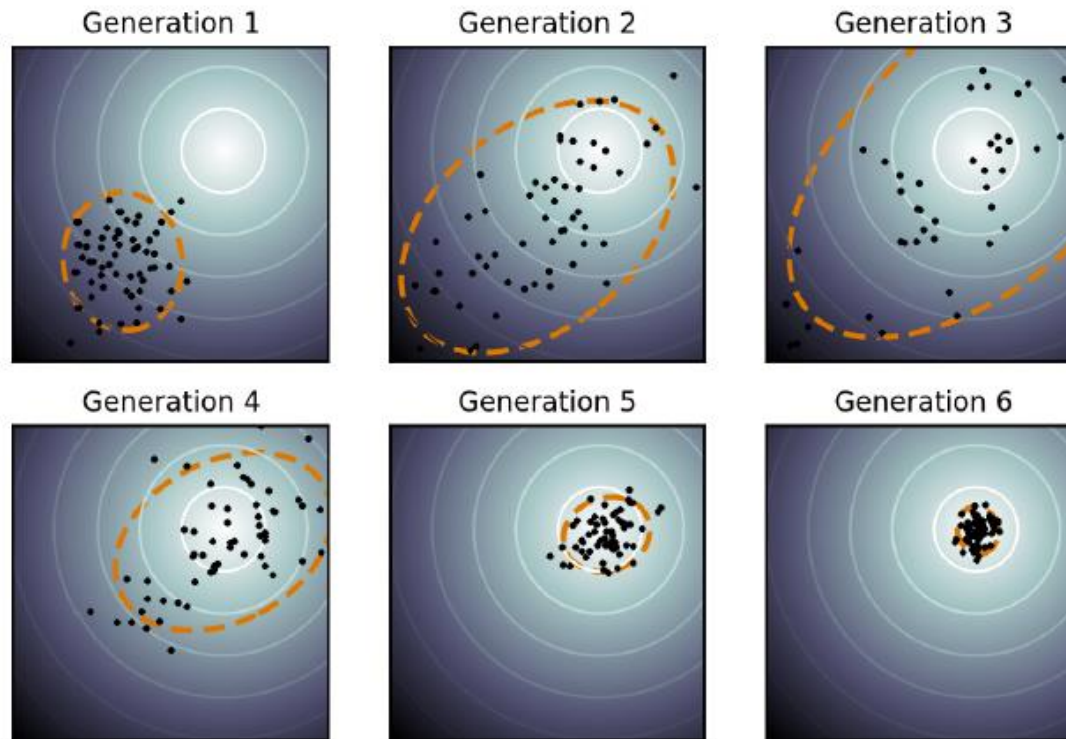


Continuous satisfaction degree in [0,1]
of the LTL(R) formula for oscillation
with amplitude constraint [Rizk Batt FF Soliman CMSB 08]

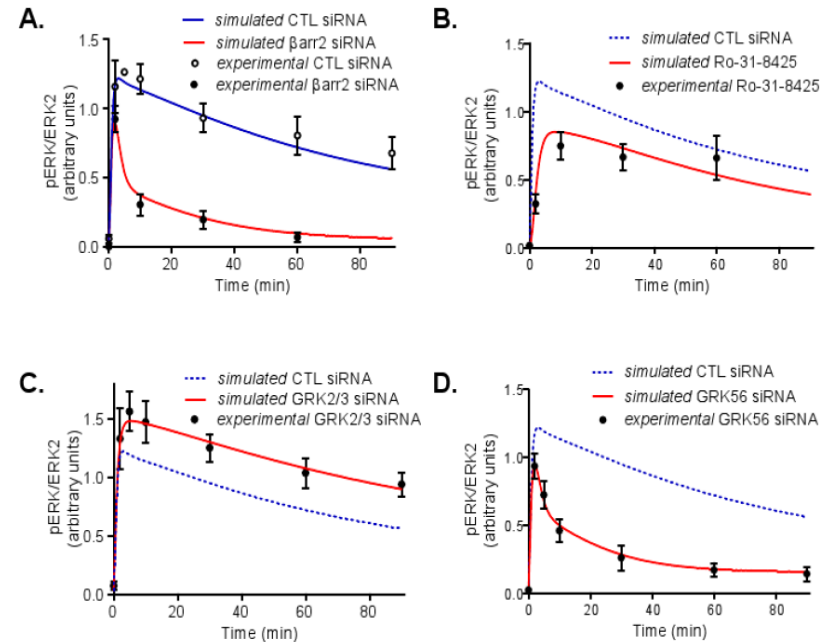
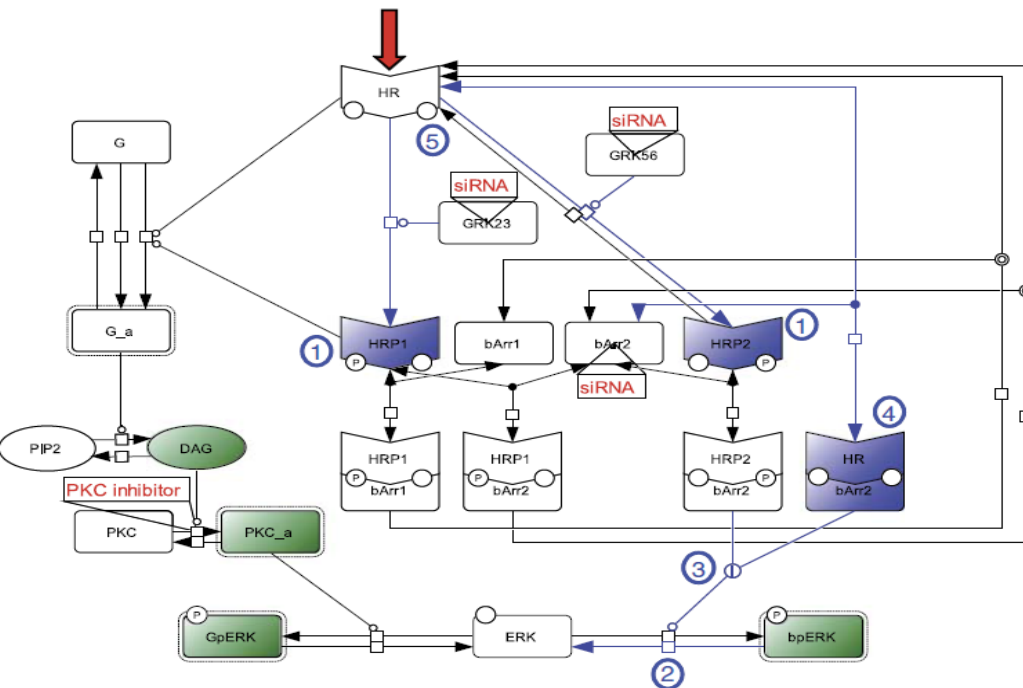
- **Parameter search** under LTL(R) constraints in high dimension (100 parameters) by continuous optimization (evolutionary algorithm CMA-ES)
- **Robustness and sensitivity** analyses w.r.t. LTL(R) specification

Covariance Matrix Adaptation Evolutionary Strategy

- CMA-ES maximizes a black box fitness function ($sd(\phi)$) in continuous domain (k'_i 's) [Hansen Osermeier 01, Hansen 08]
- CMA-ES uses a probabilistic neighborhood and updates information in covariance matrix at each move



Success Story in GPCR Signaling



Failure to find parameters fitting the data under perturbed conditions in BIOCHAM with CMAES-LTL(R) was the key to **revisit the structure** of the interactions

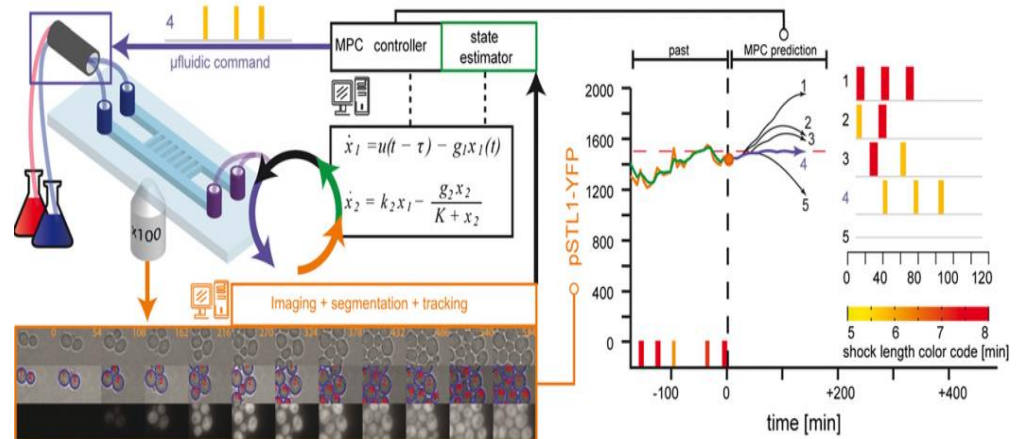
[D. Heitzler, ..., FF, R. Lefkowitz, E. Reiter 2012 *Molecular Systems Biology*]

Model-based control of gene expression in yeast

Perception - learning - action loop

1. Microscope, cell tracking
2. Model fitting (parameter opt.)
3. Osmotic pressure (parameter opt.)

[Uhlendorf ... Batt Hersen PNAS 109(35) 2012]



SM 5 : Real time yeast single cell control to a constant target profile

Long-term model predictive control of gene expression at the population and single-cell levels

J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt* & P. Hersen*

✉ pascal.hersen@univ-paris-diderot.fr 🔗 <http://inryurl.com/TeamHersen>
 ✉ gregory.batt@inria.fr 🔗 <http://www-roq.inria.fr/~batt/>



SM 7 : Real time yeast single cell control to a trapeze target profile

Long-term model predictive control of gene expression at the population and single-cell levels

J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt* & P. Hersen*

✉ pascal.hersen@univ-paris-diderot.fr 🔗 <http://inryurl.com/TeamHersen>
 ✉ gregory.batt@inria.fr 🔗 <http://www-roq.inria.fr/~batt/>



SM 6 : Real time yeast single cell control to a sine wave target profile

Long-term model predictive control of gene expression at the population and single-cell levels

J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt* & P. Hersen*

✉ pascal.hersen@univ-paris-diderot.fr 🔗 <http://inryurl.com/TeamHersen>
 ✉ gregory.batt@inria.fr 🔗 <http://www-roq.inria.fr/~batt/>



Wrap-up

- Chemical reaction network theory
 - Sufficient/necessary conditions for dynamical behaviors
 - Hybrid models and hybrid simulations

Crucial for reducing models

Wrap-up

- Chemical reaction network theory
 - Sufficient/necessary conditions for dynamical behaviors
 - Hybrid models and hybrid simulations
- Static analyzers of reaction programs
 - Structural invariants
 - Model reduction relationships
 - Efficient CLP(FD) and SAT solvers

Crucial for reducing models

What can be said from the structure of a reaction network ?

Wrap-up

- Chemical reaction network theory
 - Sufficient/necessary conditions for dynamical behaviors
 - Hybrid models and hybrid simulations
- Static analyzers of reaction programs
 - Structural invariants
 - Model reduction relationships
 - Efficient CLP(FD) and SAT solvers
- Dynamic analyzers of behaviors using temporal logic patterns
 - Parameter search for model building and control

Crucial for reducing models

What can be said from the structure of a reaction network ?

The nerve of war in biological modeling

Wrap-up

- **Chemical reaction network theory**
 - Sufficient/necessary conditions for dynamical behaviors
 - Hybrid models and hybrid simulations
- **Static analyzers** of reaction programs
 - Structural invariants
 - Model reduction relationships
 - Efficient CLP(FD) and SAT solvers
- **Dynamic analyzers** of behaviors using temporal logic patterns
 - **Parameter search** for model building and control
- **Biochemical programming**
 - Synthetic biology: correct or implement new functions in the cell or in non living vesicles
 - Systems biology: elucidate natural circuits as optimized biochemical code

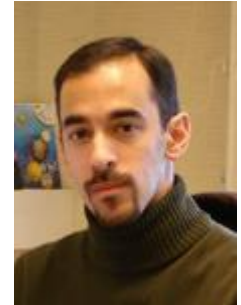
Crucial for reducing models

What can be said from the structure of a reaction network ?

The nerve of war in biological modeling

**“What I cannot create I do not understand”
R. Feynman**

Lifeware team at Inria



Research Scientists

Grégory Batt (Inria, Researcher, Synthetic Biology Principal Investigator)
Sylvain Soliman (Inria, Researcher)

Associates

Pascal Hersen (CNRS MSC lab, Researcher)
Denis Thieffry (ENS Paris, Professor)



PhD students

François Bertaux (AMX, Ecole Polytechnique, with EPI Bang)
Katherine Chiang (National Taiwan University)
Xavier Duportet (Inria, CORDI-S, MIT Weiss lab)
David Fournier (Cifre General Electric Transportation)
Steven Gay (Inria, Post doc Univ. Louvain la Neuve, Belgium)
Jean-Baptiste Lugagne (Inria, CORDI-C, CNRS MSC lab)
Artemis Llamosi (CNRS MSC lab, Inria)
Pauline Traynard (AMX, Ecole Polytechnique, with ENS)

Engineers

François-Marie Floch (Inria ADT, Engineer)
Thierry Martinez (Research Engineer, SED, Inria Rocquencourt)
Philippe Morignot (Inria, Engineer)