Computational Methods in Systems and Synthetic Biology

François Fages
INRIA Paris-Rocquencourt

mailto:Francois.Fages@inria.fr
http://contraintes.inria.fr

Overview of the Lectures

1. Formal molecules and reaction models in BIOCHAM
2. Kinetics
3. Qualitative properties formalized in temporal logic CTL
4. Quantitative properties formalized in LTL(R) and pLTL(R)
5. Reaction hypergraphs and influence graphs
6. Hierarchy of semantics and typing for systems biology by abstract interpretation
7. Learning parameters from temporal logic properties
8. Robustness analysis
   - Definition of robustness
   - Robustness analysis of cell cycle model
   - Robustness analysis of synthetic transcriptional cascade in *E. Coli*

Defining Robustness

Robustness of

- a system s
- w.r.t. a set of possible perturbations $p \in P$
- and an evaluation D_a^s of functionality property a

Defining Robustness

Robustness of

- a system s
- w.r.t. a set of possible perturbations $p \in P$
- and an evaluation D_a^s of functionality property a

General definition of Kitano [Mol. Syst. Biol. 2007]

$$R_{a,P}^s = \int_{p \in P} prob(p)\ D_a^s\ dp \tag{1}$$

The *evaluation function* D_a^s of the system should determine if the system still maintains its function under a perturbation and to what degree. The evaluation function needs to be defined for each specific problem in an ad-hoc manner and re-implemented for the computation of the robustness.

Temporal Logic based definition of robustness

Definition using the satisfaction degree of a functionality specification in temporal logic QFLTL:

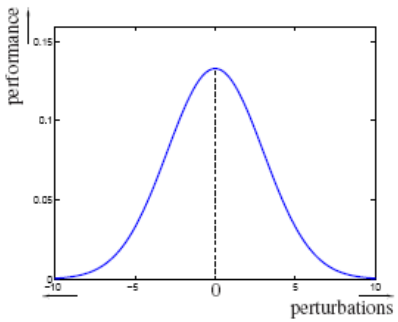
$$R^{s}_{\phi,P} = \int_{p \in P} prob(p)\ sd(T_p, \phi)\ dp, \tag{2}$$

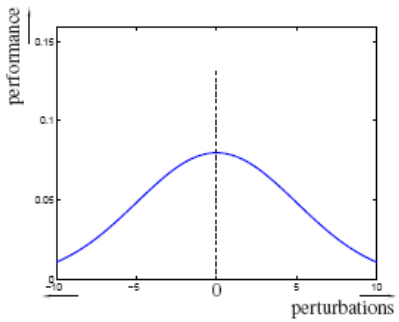
where ϕ is the specification of the functionality in temporal logic and T_p is the trace representing the behavior of the system under perturbation p.

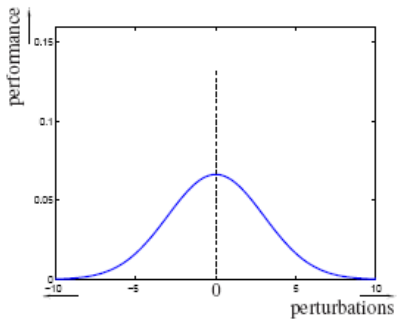
This notion of *robustness* corresponds to a mean functionality, that is, describes on average how the system behaves under perturbations.

Robustness as mean functionality over perturbations

Plots 1-3 of the performance D_a^s (satisfaction degree) of three systems in the face of perturbations







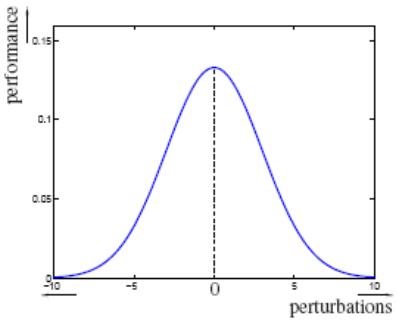
Because the average is the same in 1 and 2, the robustness of these two systems are equal for evenly-distributed perturbations.
For example in a bioengineering context, if the "property" reflects the quantity of some product exported by cells, these two systems will indeed produce on average the same quantity of the desired product.

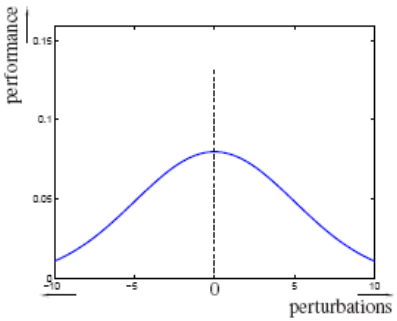
The performance of system 3 is half of the one of system 1.
the absolute robustness of 3 is less than 1, 2.

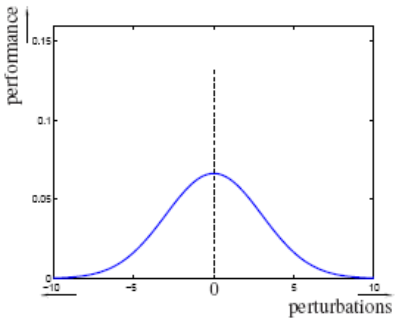
Relative robustness w.r.t. nominal behavior

$$R^{s,p*}_{\phi,P} = R^{s}_{\phi,P} / sd(T_{p*}, \phi), \tag{3}$$

where T_{p^*} denotes the unperturbed, nominal behavior of the system.







one can distinguish the relative robustness of systems 1 and 2 w.r.t. their nominal performance,
reflecting that the performance is more impacted by perturbations in system 1 than in system 2.

The performance of system 3 is half of the one of system 1.
the absolute robustness of 3 is less than 1, 2.
the relative robustness of 1 and 3 are the same
the relative robustness of 2 is less than 1, 3.

Computing Robustness

For the computation of $R^{s}_{\phi,P}$, and $R^{s,p*}_{\phi,P}$, one needs to distinguish

- whether the set of perturbations is finite (*e.g.* gene knockouts) in which case computation can be exact
- or infinite (*e.g.* normally-distributed parameter variations) in which case the value can be estimated by sampling the perturbation set for sufficiently many perturbations.

input: a (model of the) system f, (QF)LTL formulae ϕ and $\phi(\mathbf{y})$, a set of perturbations P and their probabilities, and nominal behavior p^*
output: robustness estimates $R_{\phi,P}$, $R^{p*}_{\phi,P}$, and $Rsd_{\phi,P}$

Given an ODE model f, a set P of perturbations of initial conditions or parameters, and (QF)LTL properties ϕ and $\phi(\mathbf{y})$, the BIOCHAM computes the absolute and relative robustness w.r.t. perturbations.

Implementation

The computation of the trace T_p is done by numerical integration.
The computation of the satisfaction domain $\mathcal{D}_{T_p,\phi(\mathbf{y})}$ is made by induction on the formula structure, using for each subformula a direct implementation of the definition.
Polytopes operations are implemented in BIOCHAM using a standard polyhedral library.
for every perturbation $p \in P \cup \{p^*\}$
$T_p :=$ Compute_tracef, p
$\mathcal{D}_{T_p,\phi(\mathbf{y})} :=$ Compute_sat_domain $T_p, \phi(\mathbf{y})$

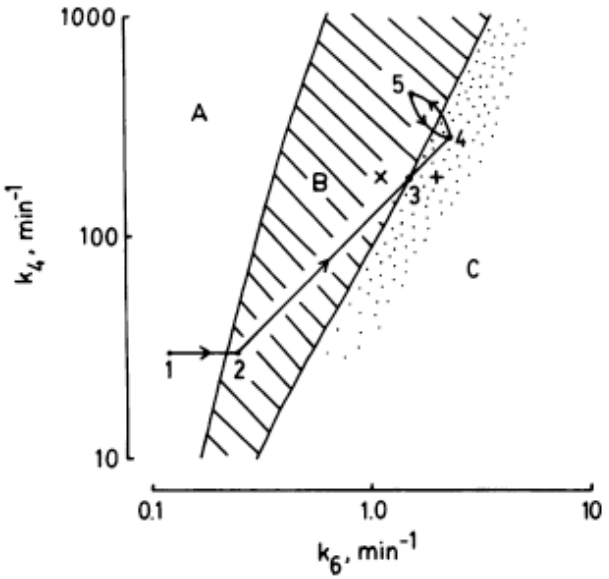
$R_{\phi,P} := \sum_{p \in P} prob(p)(1 + dist(\mathcal{D}_{T_p,\phi(\mathbf{y})}, \phi))^{-1}$

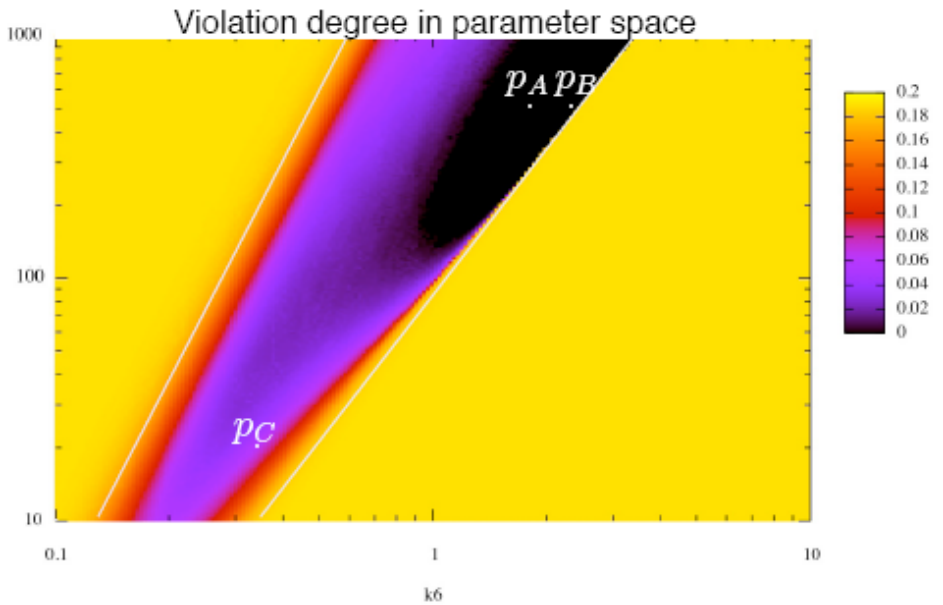
Implementation

The computation of the trace T_p is done by numerical integration.
The computation of the satisfaction domain $\mathcal{D}_{T_p,\phi(\mathbf{y})}$ is made by induction on the formula structure, using for each subformula a direct implementation of the definition.
Polytopes operations are implemented in BIOCHAM using a standard polyhedral library.
for every perturbation $p \in P \cup \{p^*\}$
$T_p :=$ Compute_tracef, p
$\mathcal{D}_{T_p,\phi(\mathbf{y})} :=$ Compute_sat_domain $T_p, \phi(\mathbf{y})$

$$R_{\phi,P} := \sum_{p \in P} prob(p)(1 + dist(\mathcal{D}_{T_p,\phi(\mathbf{y})}, \phi))^{-1}$$

$$R^{p*}_{\phi,P} := \; R_{\phi,P} * (1 + dist(\mathcal{D}_{T_{p^*},\phi(\mathbf{y})}, \phi))$$





The robustness degree of this property is compared for three different values of k4 and k6. These three points in the parameter space of k4 and k6 are indicated by the three points $\vec{k}_A$, $\vec{k}_B$ and $\vec{k}_C$.
The robustness is respectively 133, 12.9 and 13.5.

Improving global robustness of oscillations

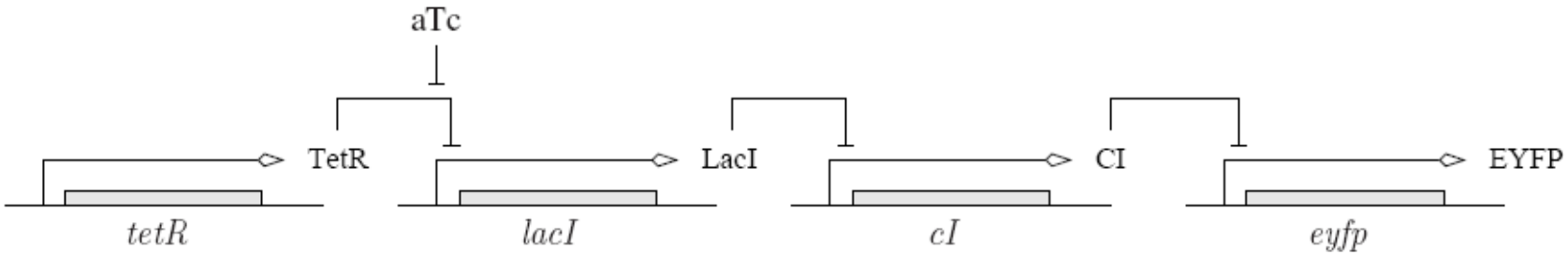
The robustness degree can be estimated for perturbations on any number of parameters.

For instance, by computing a robustness estimate for perturbations on all parameters,
with coefficient of variation 0.2 for specification ϕ_2^* and parameter values $\vec{k}_{Tyson}$ and $\vec{k}_3$,
the estimated robustness degrees for $\vec{k}_{Tyson}$ and $\vec{k}_3$ are 20.7 and 27.1 respectively.

This indicates that the oscillations are more robust to variations of the parameters values for $\vec{k}_3$ than for the parameters given in the original model of Tyson.

Synthetic biollogy in *E. Coli*

We consider a cascade of transcriptional inhibitions built in *E.coli* [Ron Weiss et al PNAS 05]



The fluorescence of the system by protein EYFP is the measured output.

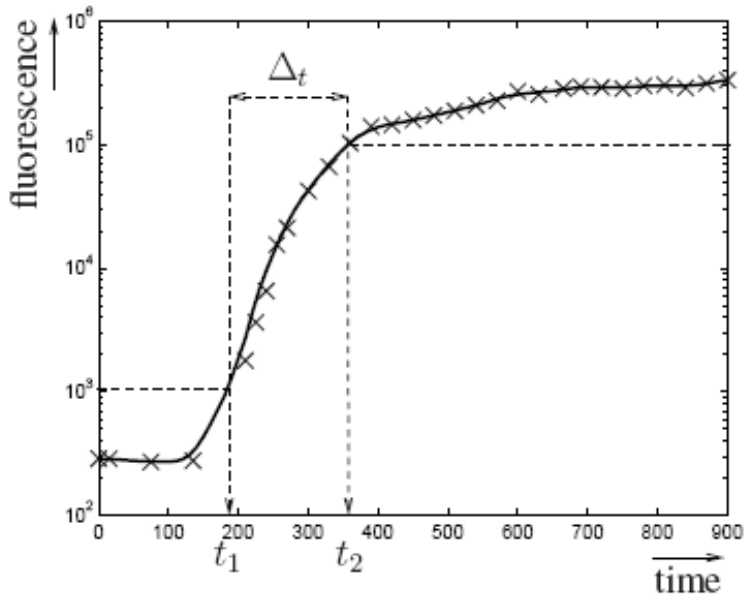
The system can be controlled by the addition or removal of a small diffusible molecule, aTc, in the growth media. More precisely, aTc binds to TetR and relieves the repression of *lacI*. The aTc concentration thus serves as a controllable input to the system.

Intuitively, the output (*i.e.* the fluorescence) of the system at steady state will be low for low inputs (*i.e.* aTc concentration), and high for high inputs.

Specifying the expected behavior in QFLTL

it has been shown that the time response of the system to an inducer addition is characterized by a rapid increase of the fluorescence, preceded by a significant lag-phase. Unfortunately, a h

Here we consider that the system is well-timed if the fluorescence remains below 10^3 for at least 150 minutes, then exceeds 10^5 after at most 450 minutes, and switches rapidly from low to high levels, that is, in less than 150 minutes.



Specifying the expected behavior in QFLTL

These specifications can be formalized in temporal logic as follows:

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

which is abstracted into

$$\begin{aligned}\phi(t_1, t_2, b_1, b_2, b_3) = \quad & \mathbf{G}(time < t_1 \rightarrow [\mathtt{EYFP}] < 10^3) \\ \wedge \quad & \mathbf{G}(time > t_2 \rightarrow [\mathtt{EYFP}] > 10^5) \\ \wedge \quad & t_1 > b1 \wedge t_2 < b_2 \wedge t_2 - t_1 < b_3\end{aligned}$$

for the computation of validity domains and satisfaction degree in a given trace.

High cell cell variability

High cell-to-cell variability has also been observed.

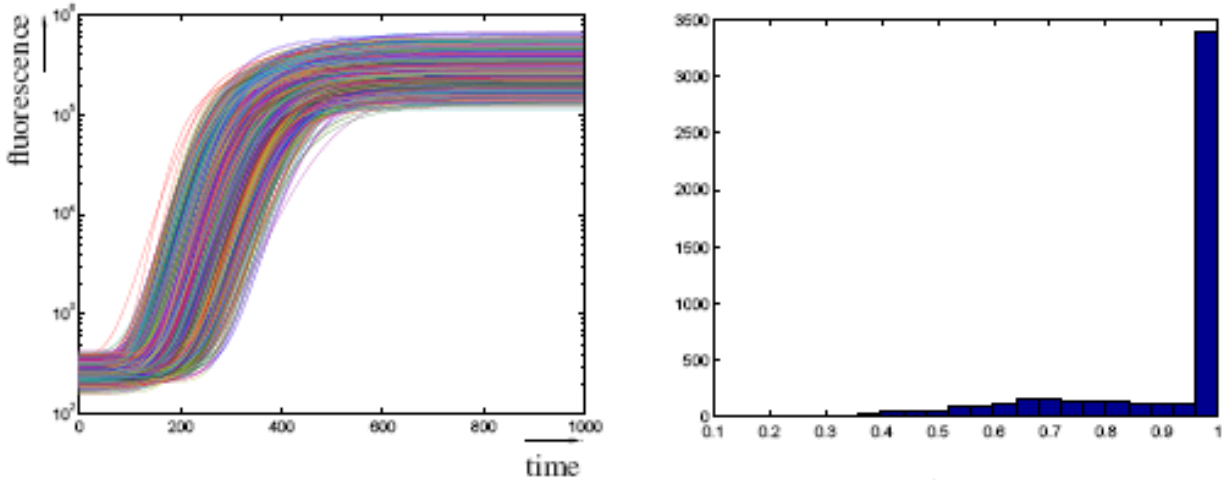
The heterogeneity of the cell responses makes it difficult to use this system as a biological timer, for example for developmental programs as suggested in [Weiss et al PNAS 05].

In this context, as for many synthetic biology applications, having even a low proportion of cells sending a signal too early or too long might compromise the correct functioning of the whole system.

Modeling the system's variability with log-normal distributions

ODE model with reference parameters $\mathbf{p}^*$
average of 5000 numerical simulations of with log-normal distributed parameters

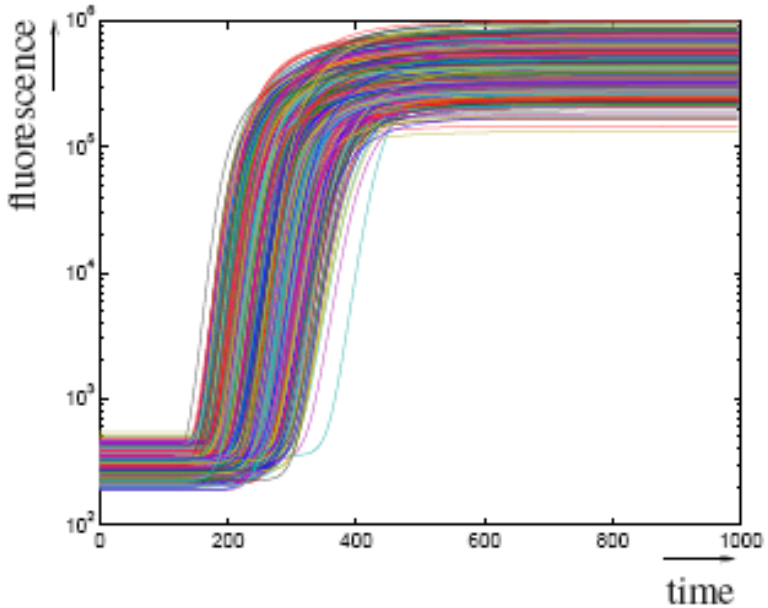
Distribution of the satisfaction degree following addition of aTc:

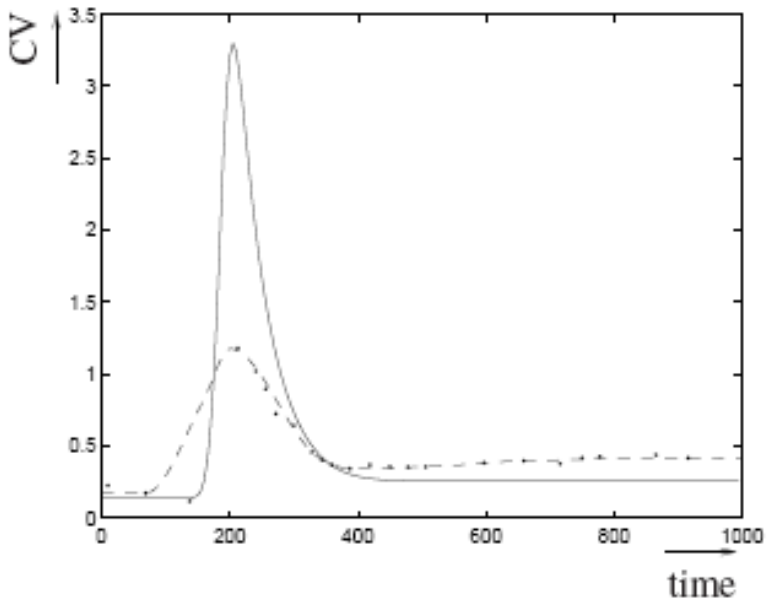


On 5000 log-normally distributed parameter values, the robustness of the system is $\hat{R}_{\phi,P} = 0.9$: the specification is not robustly satisfied.
The property is not satisfied by all cells

Improving robustness

Parameter optimization $\tilde{\mathbf{p}}$ w.r.t. robustness criterion using CMAES: EYFP production rate and the Hill coefficients η must be significantly increased.





Parameter influence on robust behavior

When a measure (in our case the robustness) is affected by variations of several parameters, one can statistically assess the importance of the variations of each parameter by computing its *sensitivity index*:

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

These sensitivity indices and higher order sensitivity indices quantify how the variance of a parameter P_i or a group of parameters contributes to the variance of R.

Sensitivity Indexes

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

γ variations have a very strong impact on the the cascade.

aTc variations seem to have a very low impact

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

Because κ_{cI} has strong effect, the basal production of EYFP is due to an incomplete repression of the promoter by CI (high effect of κ_{cI}) rather than a constitutive leakage of the promoter (low effect of κ^0_{eyfp}).

Conclusion

The continuous degree of satisfaction of LTL formulae can be used to compute robustness of systems w.r.t. high-level specification of its functionality in temporallogic:

- absolute robustness
- relative robustness
- sensibility analysis to parameter (and initial condition) variations

Conclusion

The continuous degree of satisfaction of LTL formulae can be used to compute robustness of systems w.r.t. high-level specification of its functionality in temporallogic:

- absolute robustness
- relative robustness
- sensibility analysis to parameter (and initial condition) variations

Use in synthetic biology:

- model of perturbation for transcriptional cascade among E. Coli cells
- estimation of robustness
- integration of robustness as objective criterion for parameter optimization