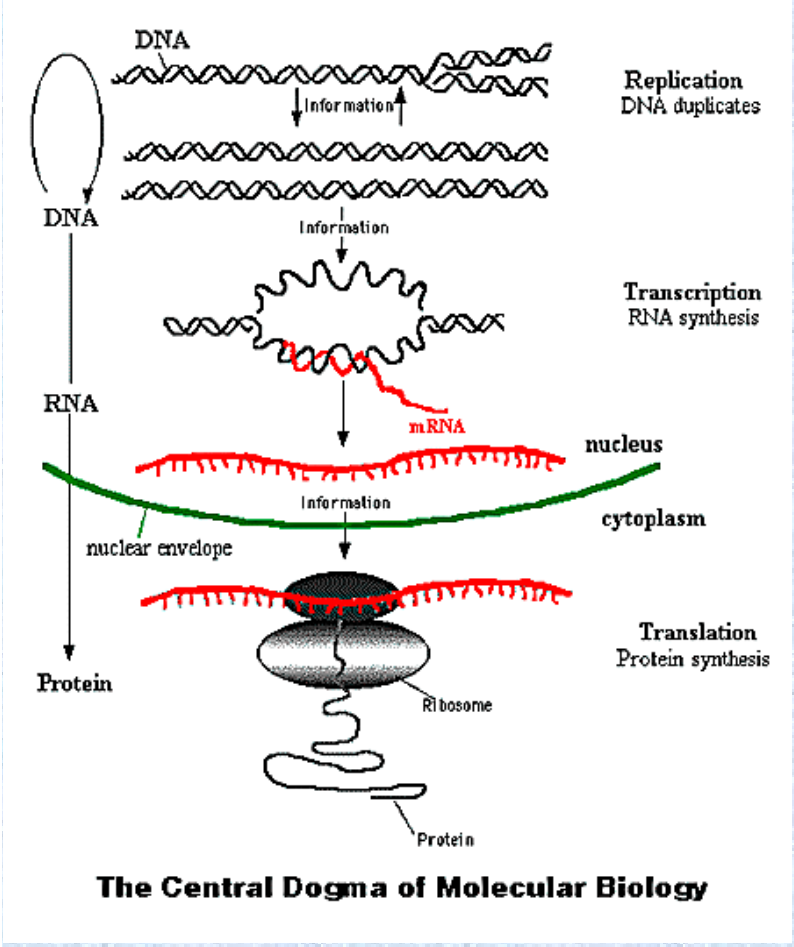
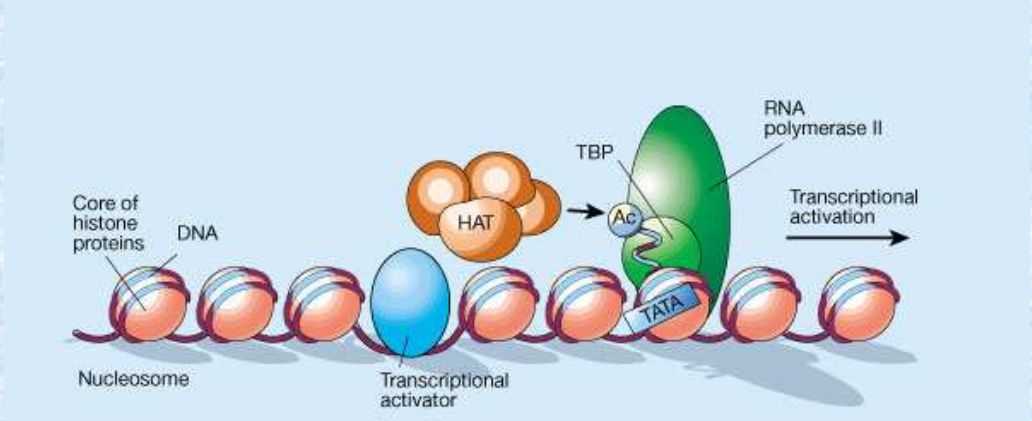
Dependence of dynamic properties of gene interaction networks on the network topology and parameter values

Juris Viksna, Estonian-Latvian Theory Days ,2010

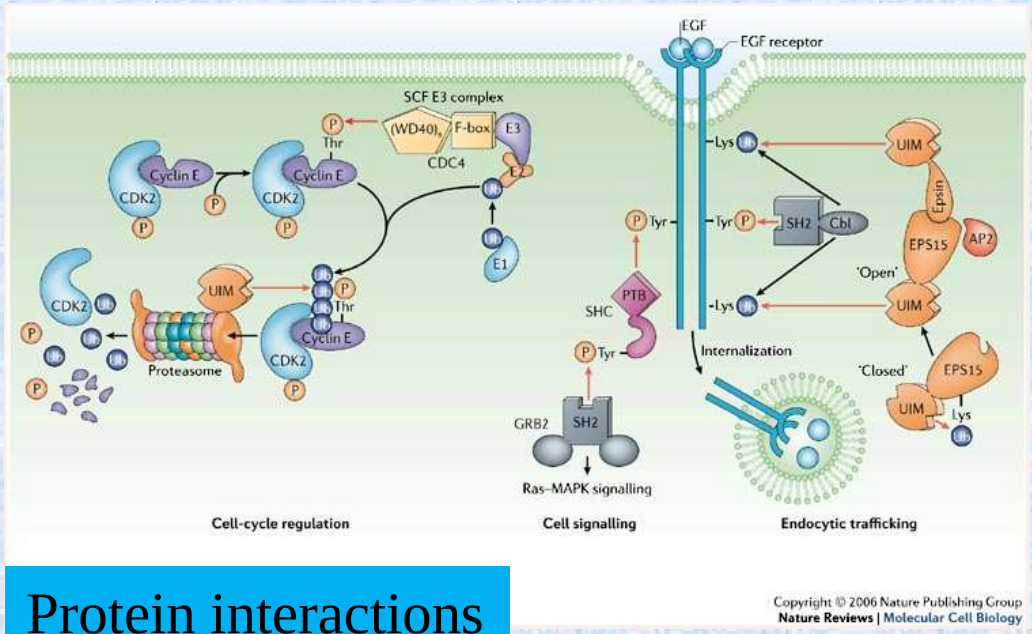
Modelling of biological processes



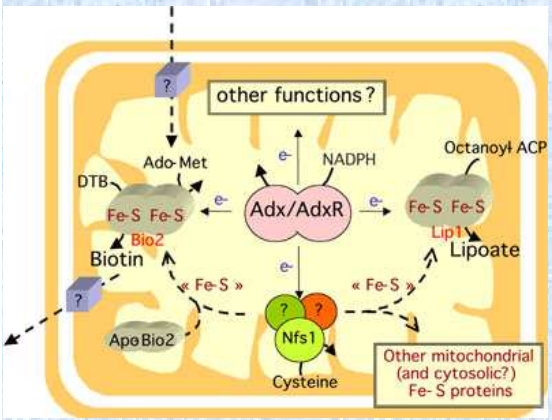
The Central Dogma of Molecular Biology



Regulation of transcription

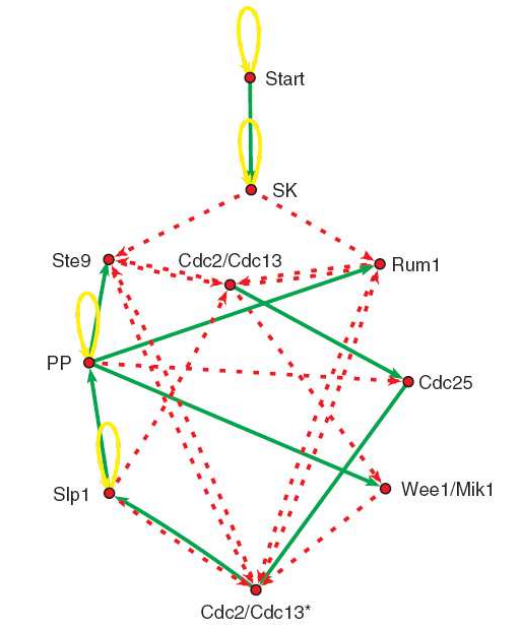


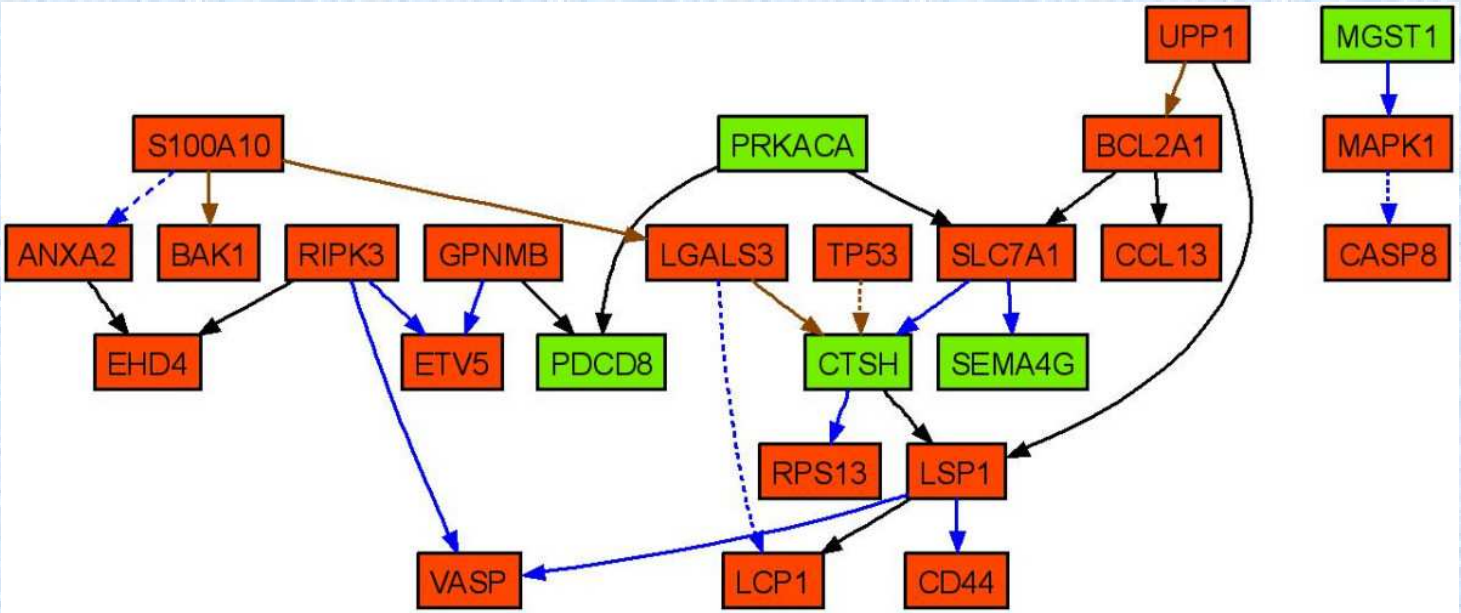
Protein interactions



Metabolic interactions

What we can gain from models of biological processes?

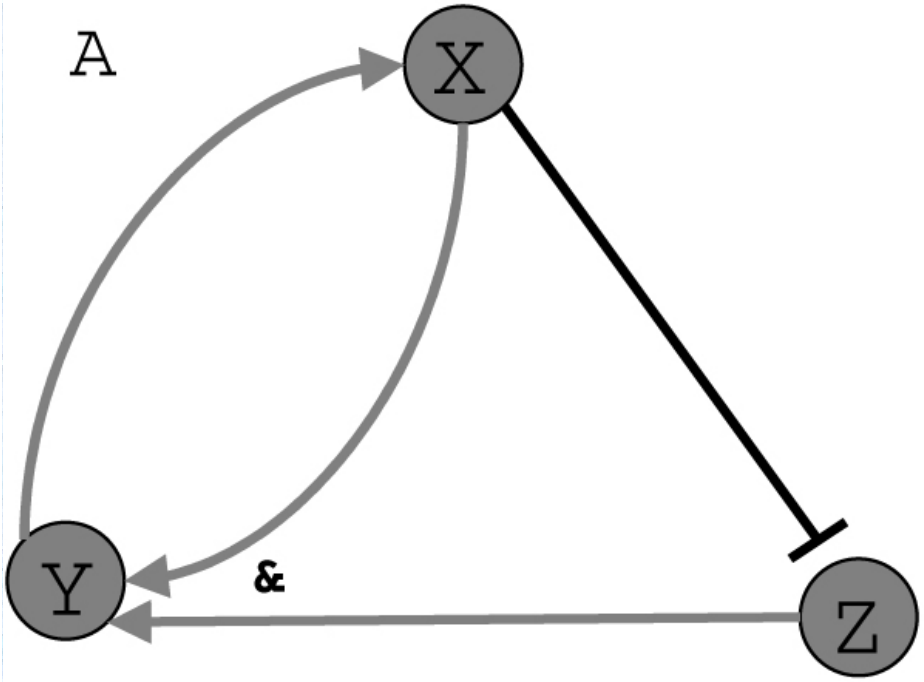




1. ***Simulation***. For given initial conditions compute how the system evolves with time.
2. ***System dynamics***. Find all the possible "stable bahaviours" that the system might exhibit.
3. ***Constraints***. For a given "stable behaviour" find the minimal requirements on system parameters that must be satisfied to ensure that this behaviour takes place.

Boolean networks

A

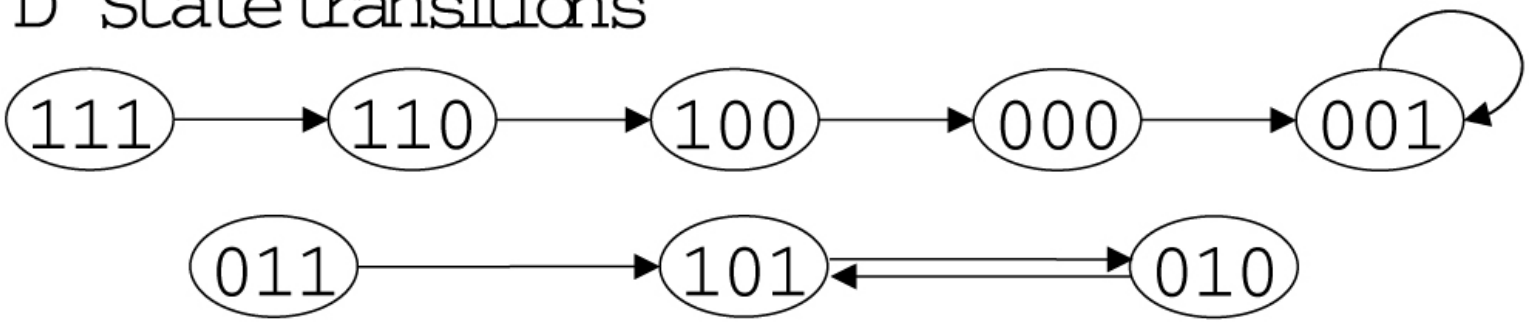


B Y=X&Z, X=Y, Z= ¬X

C

t			t+1		
X	Y	Z	X	Y	Z
0	0	0	0	0	1
0	0	1	0	0	1
0	1	0	1	0	1
0	1	1	1	0	1
1	0	0	0	0	0
1	0	1	0	1	0
1	1	0	1	0	0
1	1	1	1	1	0

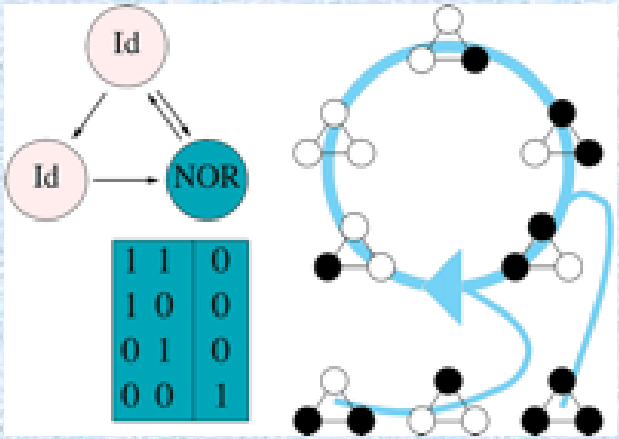
D State transitions

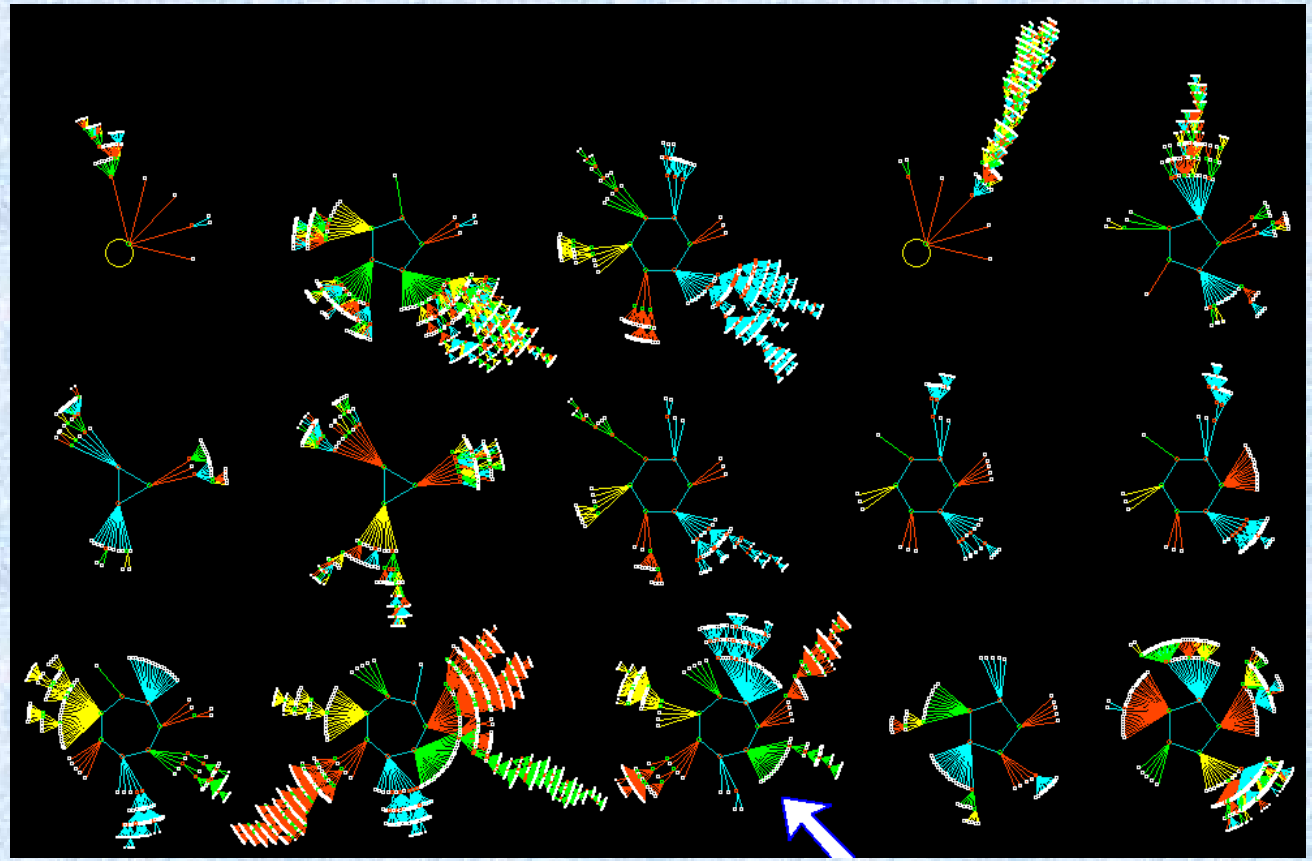


Introduced and studied by S.Kauffman already in 1969 (!)

Boolean networks

Attractors and ***attractor basins***

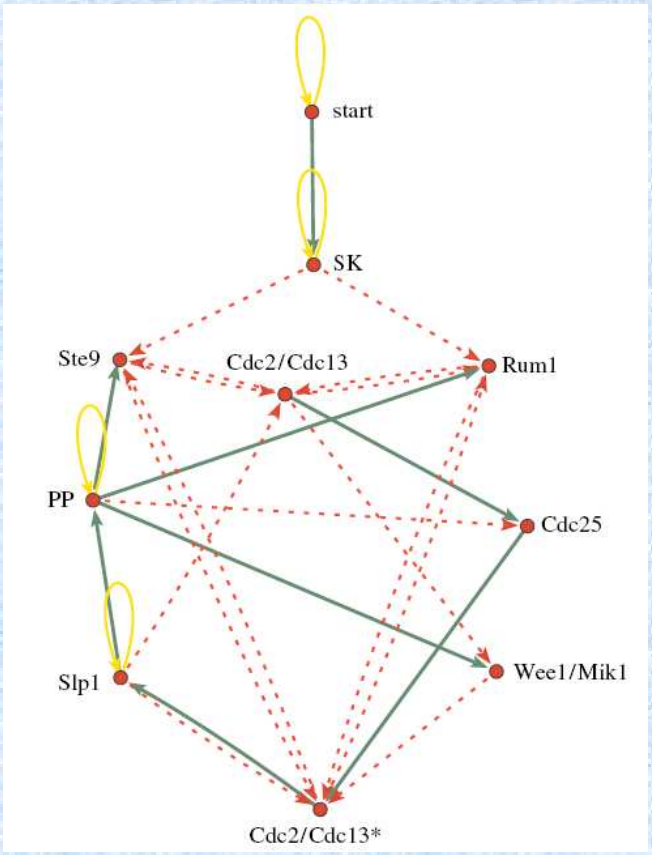


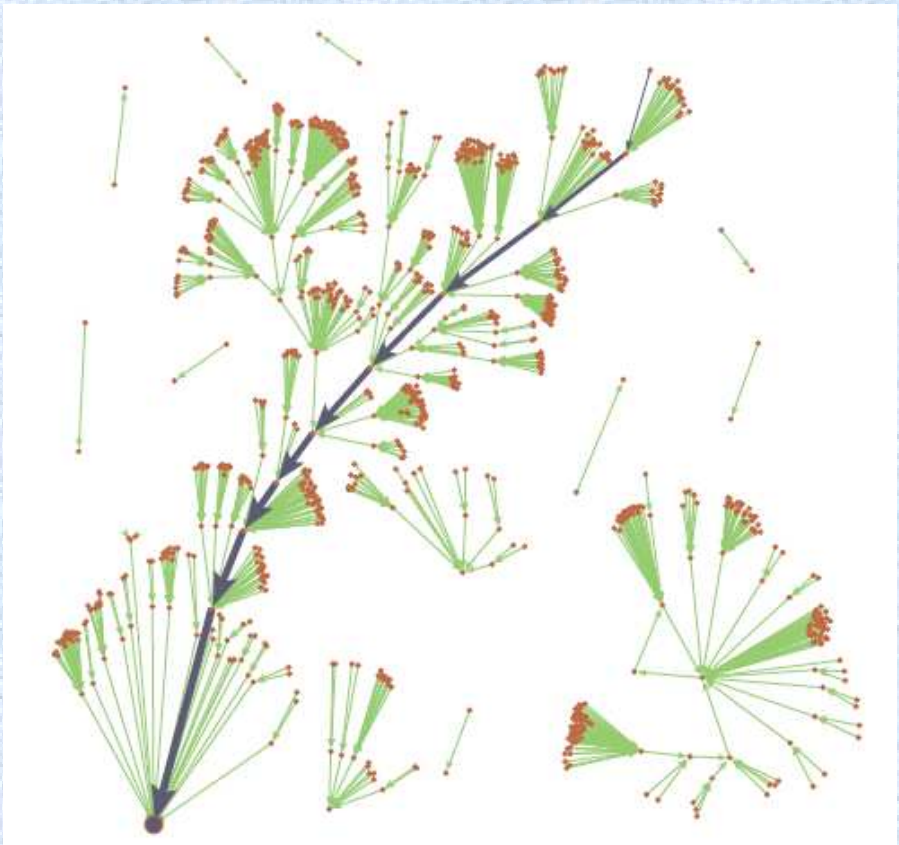


Each attractor basin corresponds to a particular "stable behaviour" of the system.

Apart from computational restrictions all 3 problems of network analysis can be solved.

Boolean networks

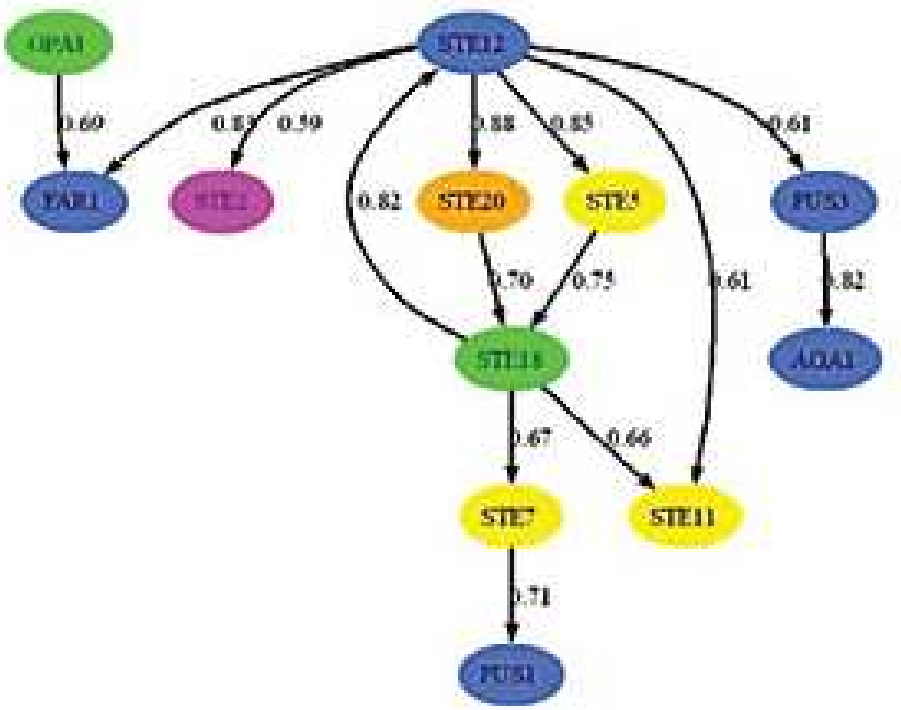


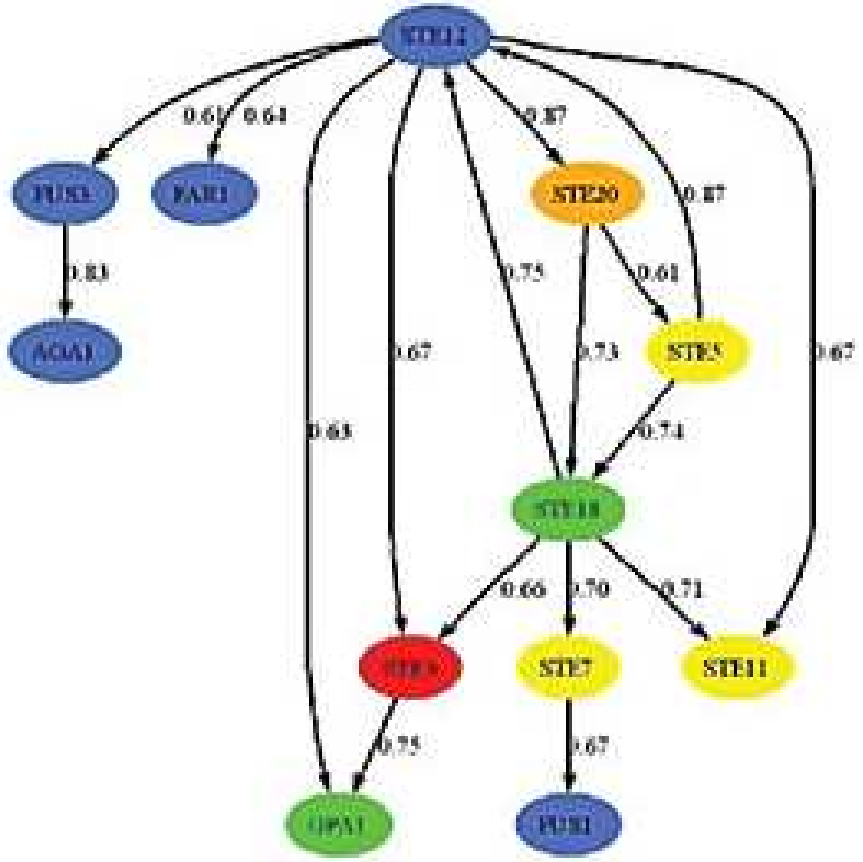


Davidich & Bornholdt (2008) Boolean model of yeast cell cycle network.

Unfortunately Boolean networks don't allow in any way to model time of interactions...

Bayesian models





For each pair of genes A and B assigns a probability p(A,B) that A influences B. The influence of several genes is cumulative.
Likely there doesn't exist a universally accepted definition of Bayesian network.
Probably we can't do much more than simulate the network behaviour...

Modelling with differential equations

Equation 3.

$$\frac{dx_i}{dt} = \frac{-e^{0.5h} + e^{-h(\omega_i - 0.5)}}{(1 - e^{0.5h})(1 + e^{-h(\omega_i - 0.5)})} - \gamma_i x_i$$

$$\omega_i = \begin{cases} \left(\frac{1+\sum\alpha_n}{\sum\alpha_n}\right)\left(\frac{\sum\alpha_n x_n^a}{1+\sum\alpha_n x_n^a}\right)\left(1-\left(\frac{1+\sum\beta_m}{\sum\beta_m}\right)\left(\frac{\sum\beta_m x_m^i}{1+\sum\beta_m x_m^i}\right)\right) & \S \\ \left(\frac{1+\sum\alpha_n}{\sum\alpha_n}\right)\left(\frac{\sum\alpha_n x_n^a}{1+\sum\alpha_n x_n^a}\right) & \S\S \\ \left(1-\left(\frac{1+\sum\beta_m}{\sum\beta_m}\right)\left(\frac{\sum\beta_m x_m^i}{1+\sum\beta_m x_m^i}\right)\right) & \S\S\S \end{cases}$$

$0 \le x_i \le 1$

$0 \le \omega_i \le 1$

$h, \alpha_n, \beta_m, \gamma_i > 0$

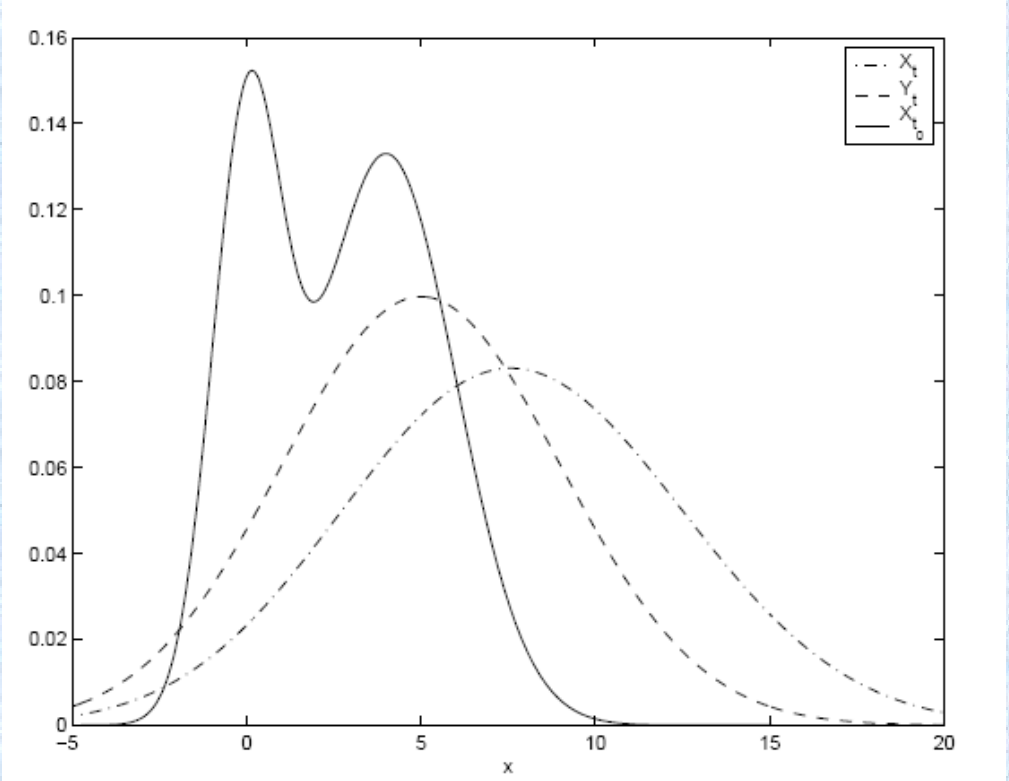
$\{x_n^a\}$ is the set of activators of x_i

$\{x_n^i\}$ is the set of inhibitors of x_i

$\S$ is used if x_i has activators and inhibitors

$\S\S$ is used if x_i has only activators

$\S\S\S$ is used if x_i has only inhibitors

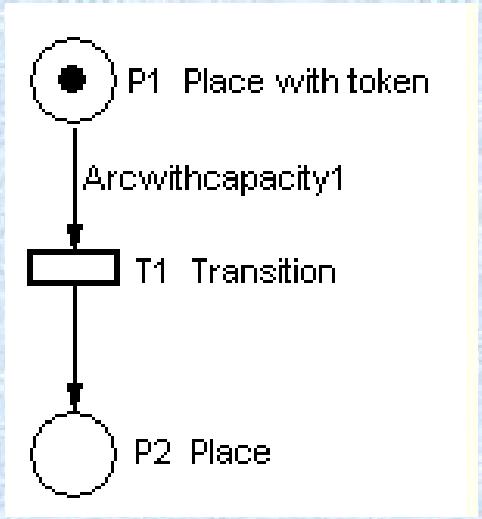


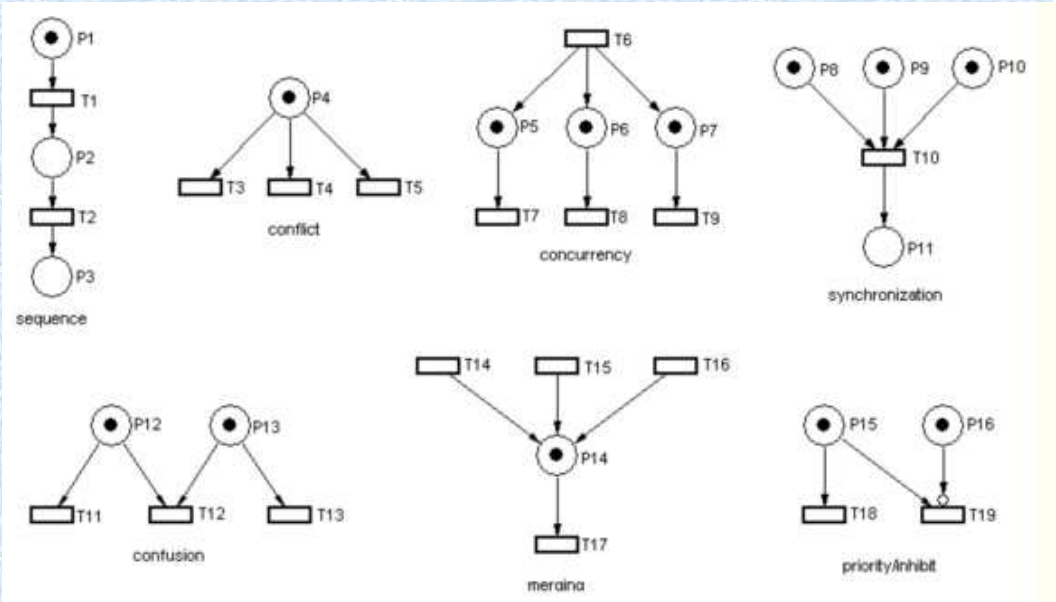
With well tuned parameters may work well for simulation of biological systems.

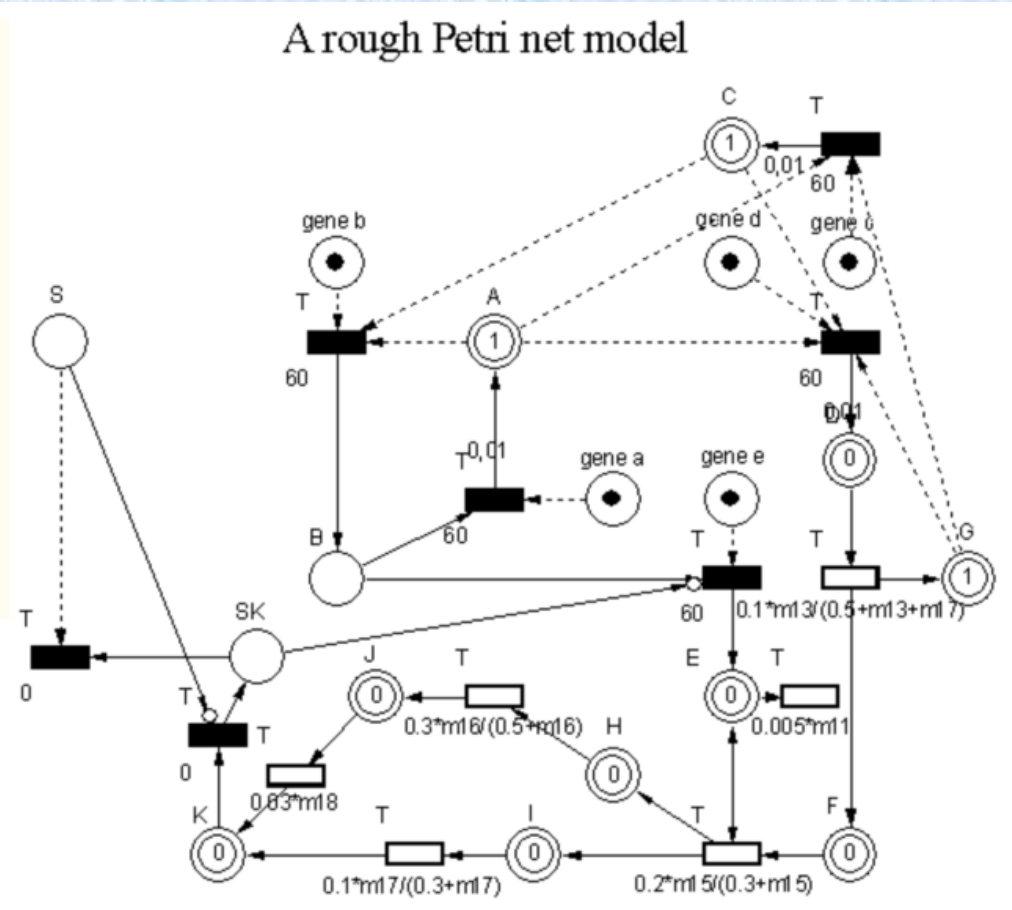
Finding the "right" parameters, however, already isn't a simple problem.

Petri net models

Petri nets were invented in August 1939 by Carl Adam Petri (at the age of 13 !!!) for the purpose of describing chemical processes.

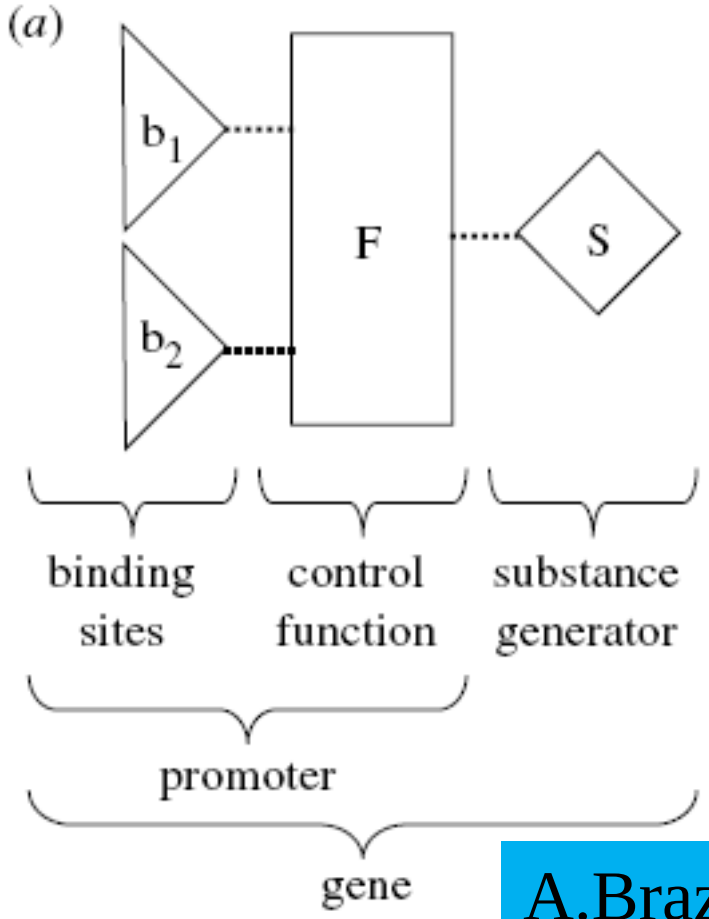


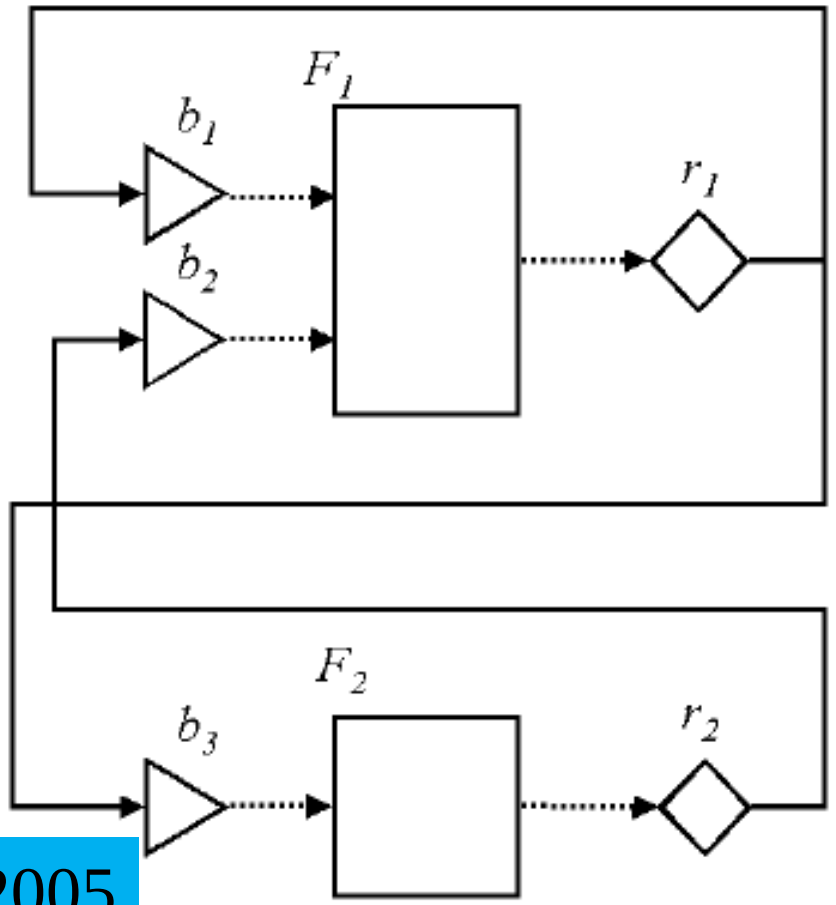




Simulation and "model checking".

Finite state linear models (FSLM)

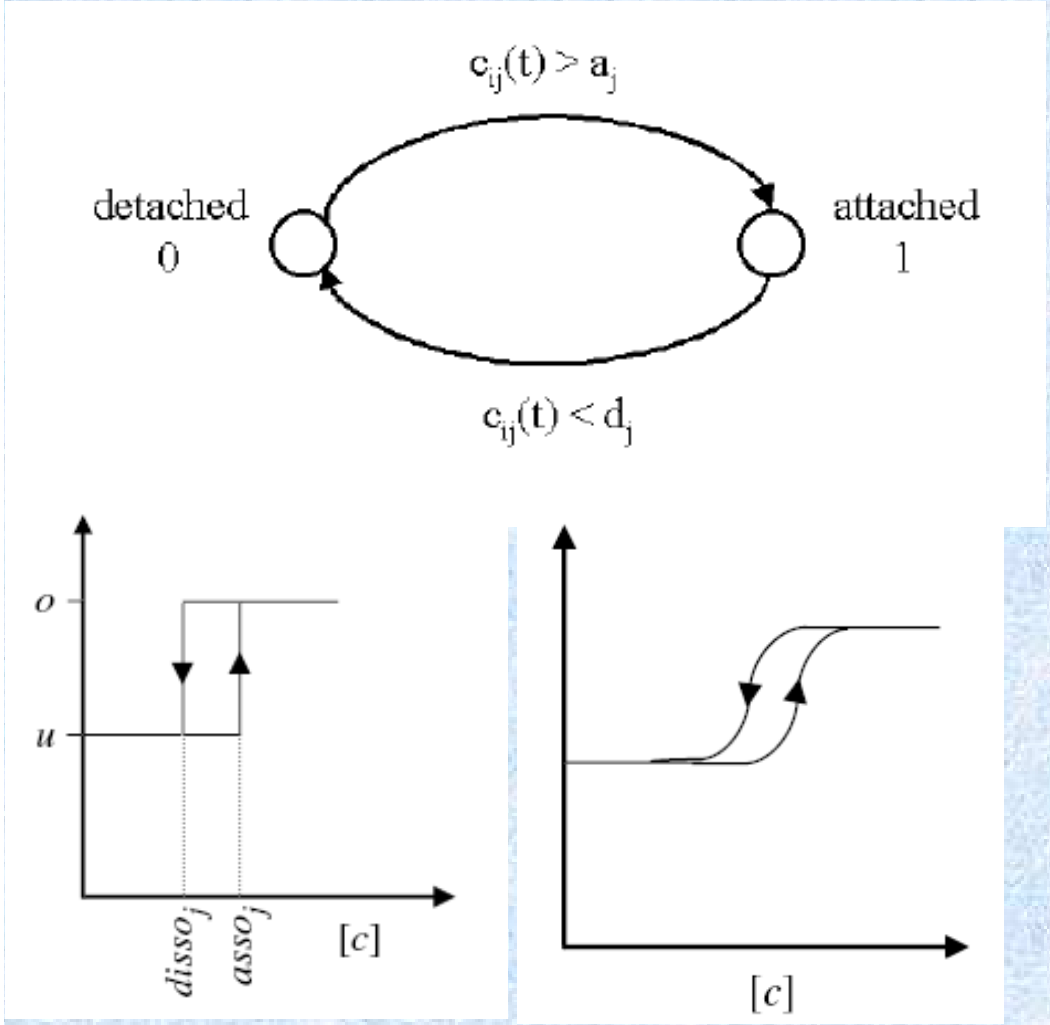


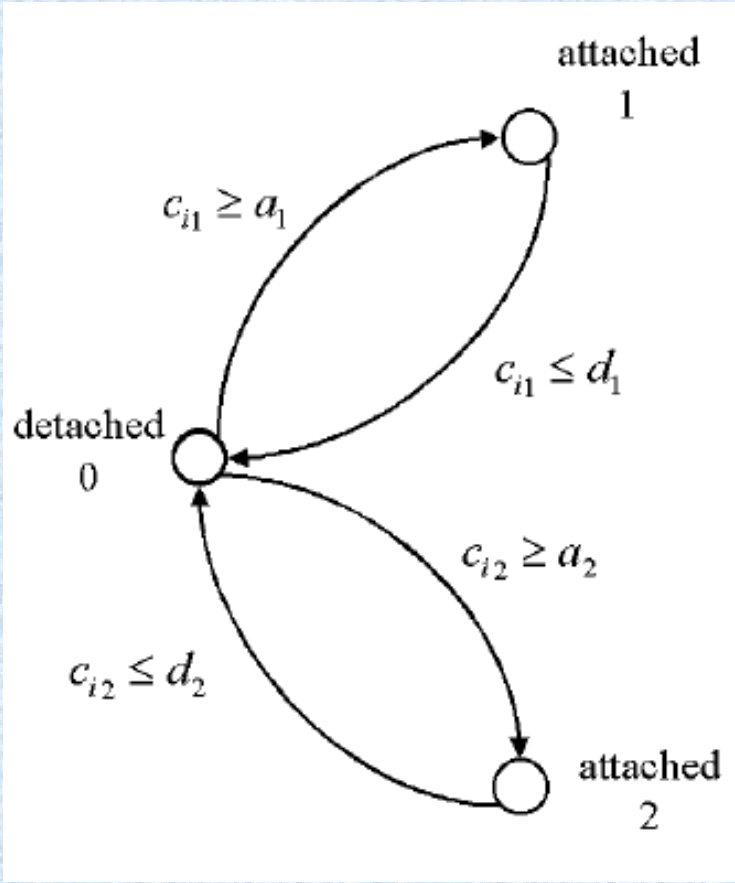


A.Brazma, T.Schlitt, 2005

- can be regarded as extensions of Boolean models which also accounts for the changes of protein concentrations over time
- gene state is a Boolean function from the sates of binding sites
- protein concentrations changes linearly

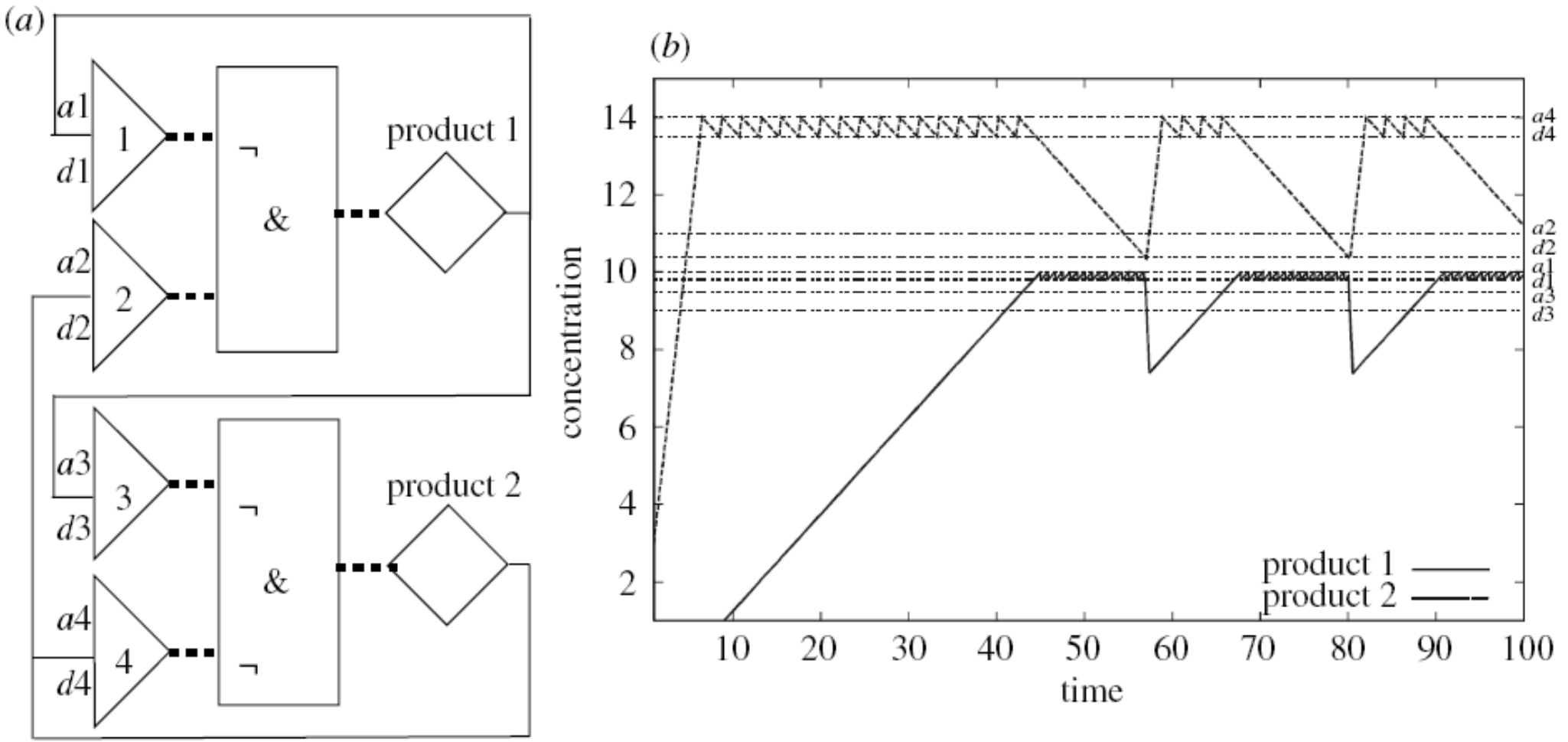
Modelling of binding sites



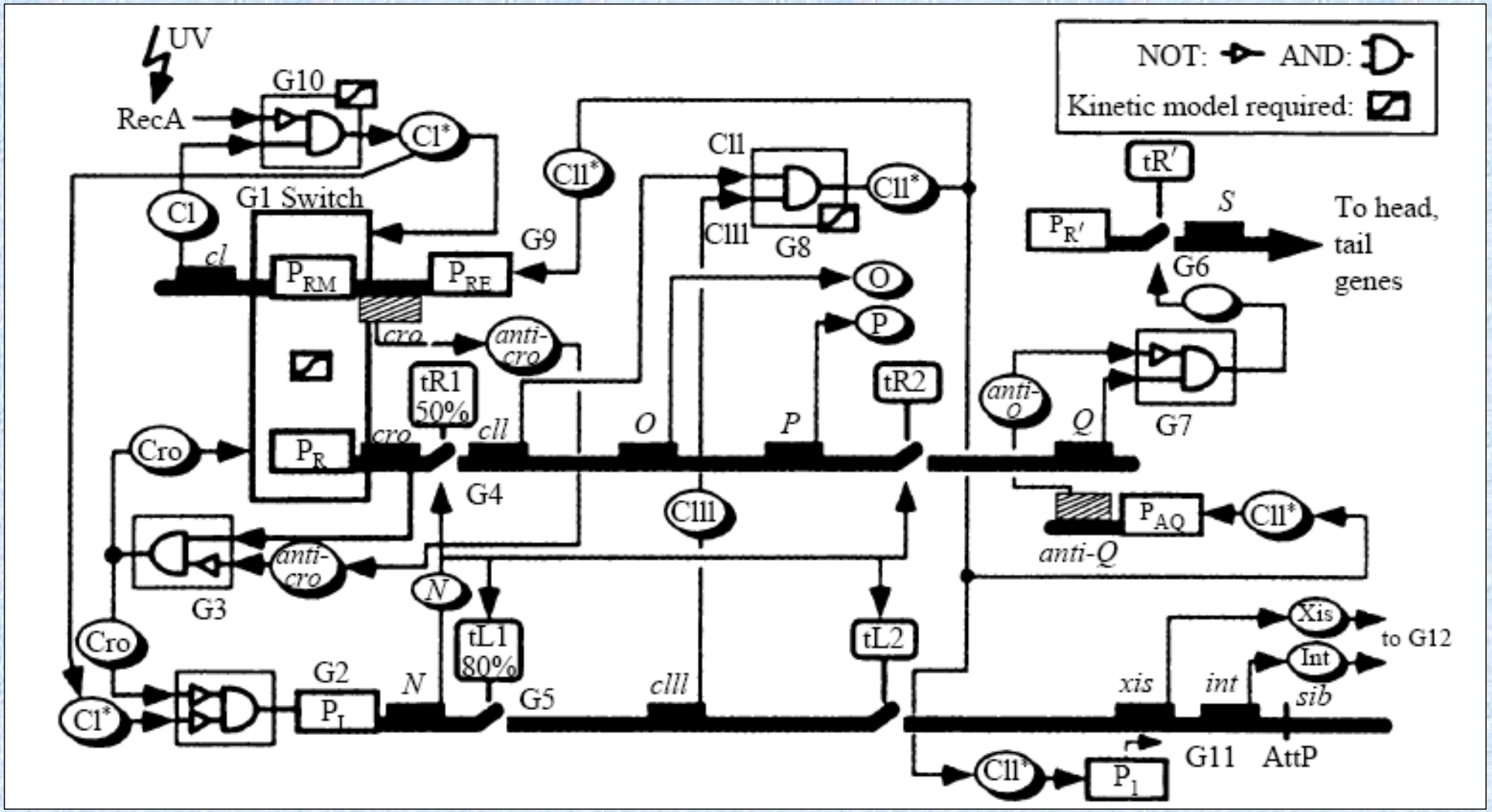


- in simplest case two constants $a>d$, sites becomes active when concentration c exceeds d and inactive when c falls below d
- several binding factors can compete for one binding site, this also could influence the protein growth rate

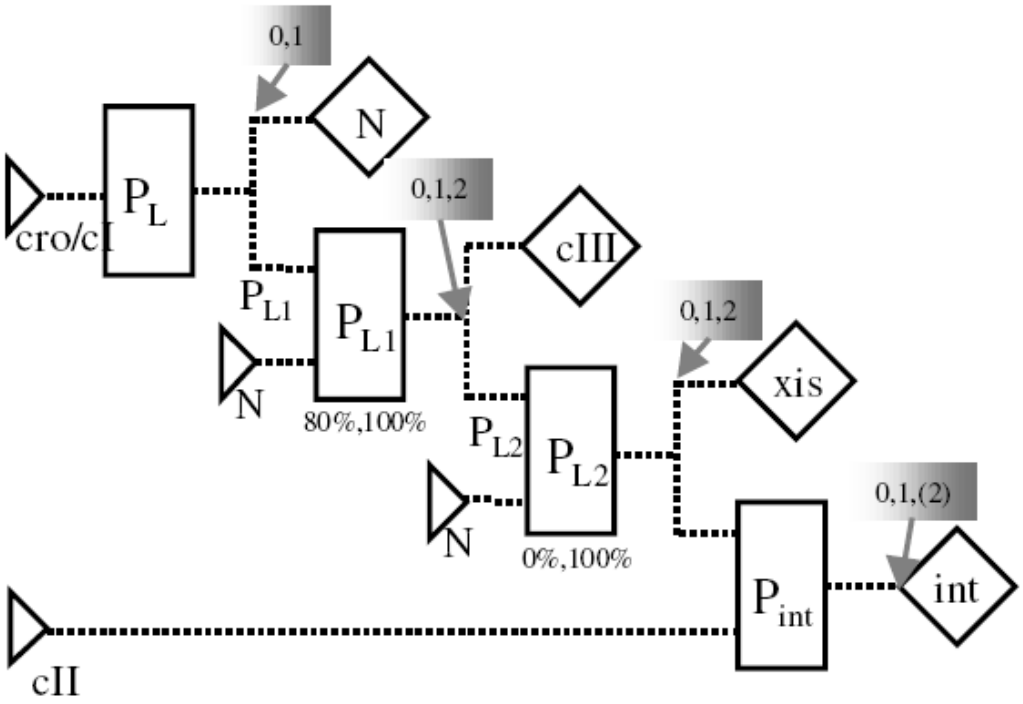
Example of FSLM

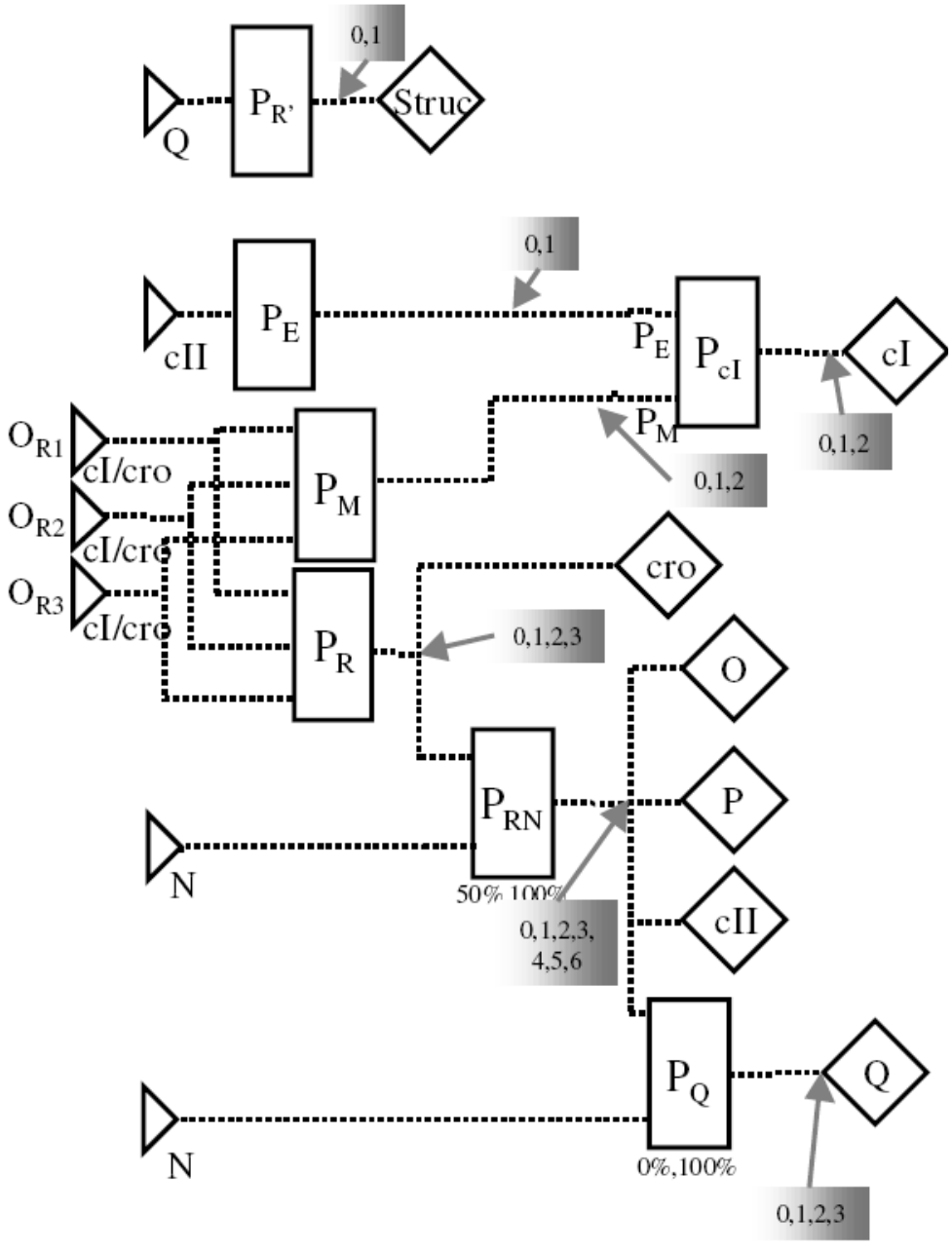


Shapiro model of lambda phage

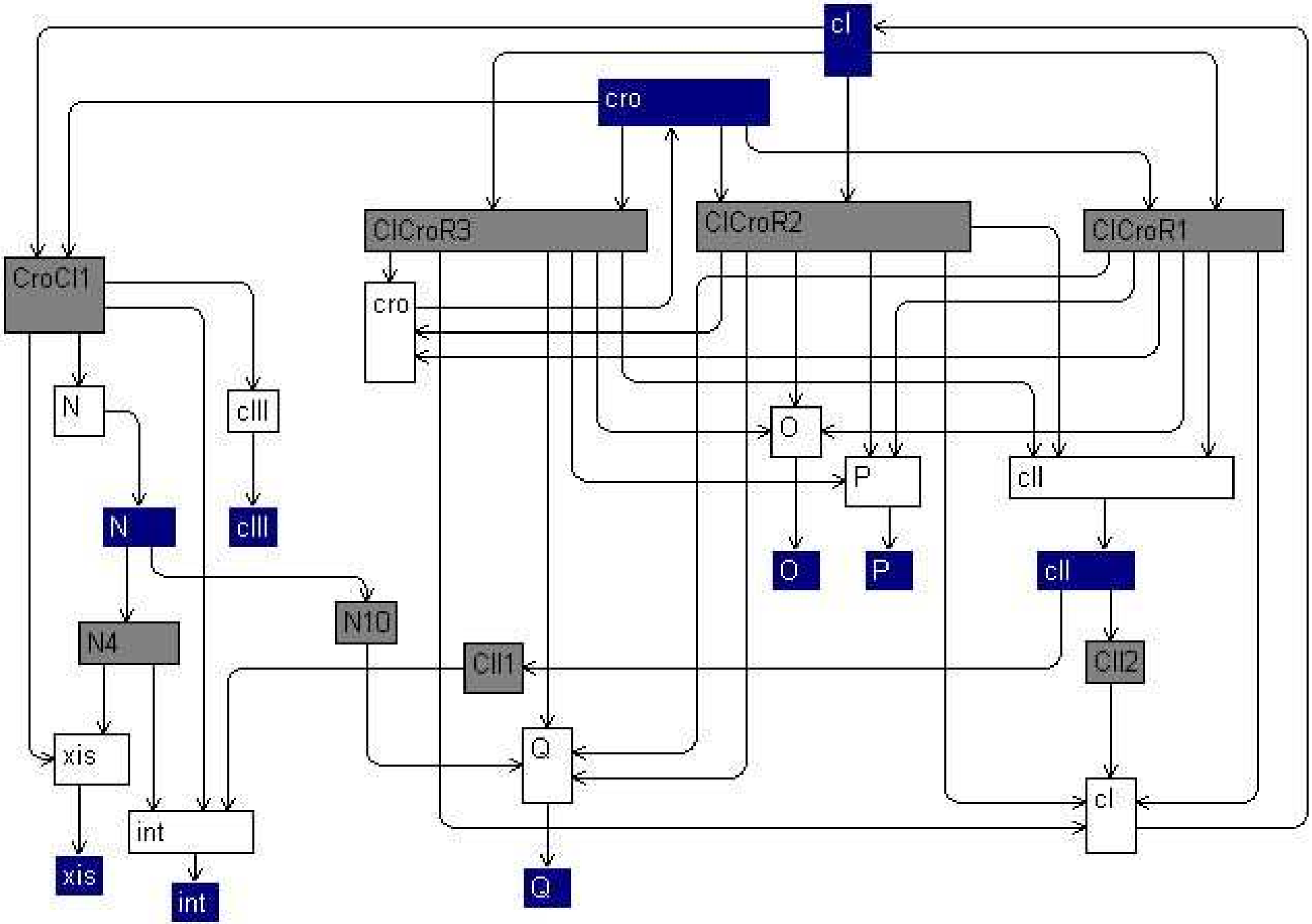


FSLM for lambda phage

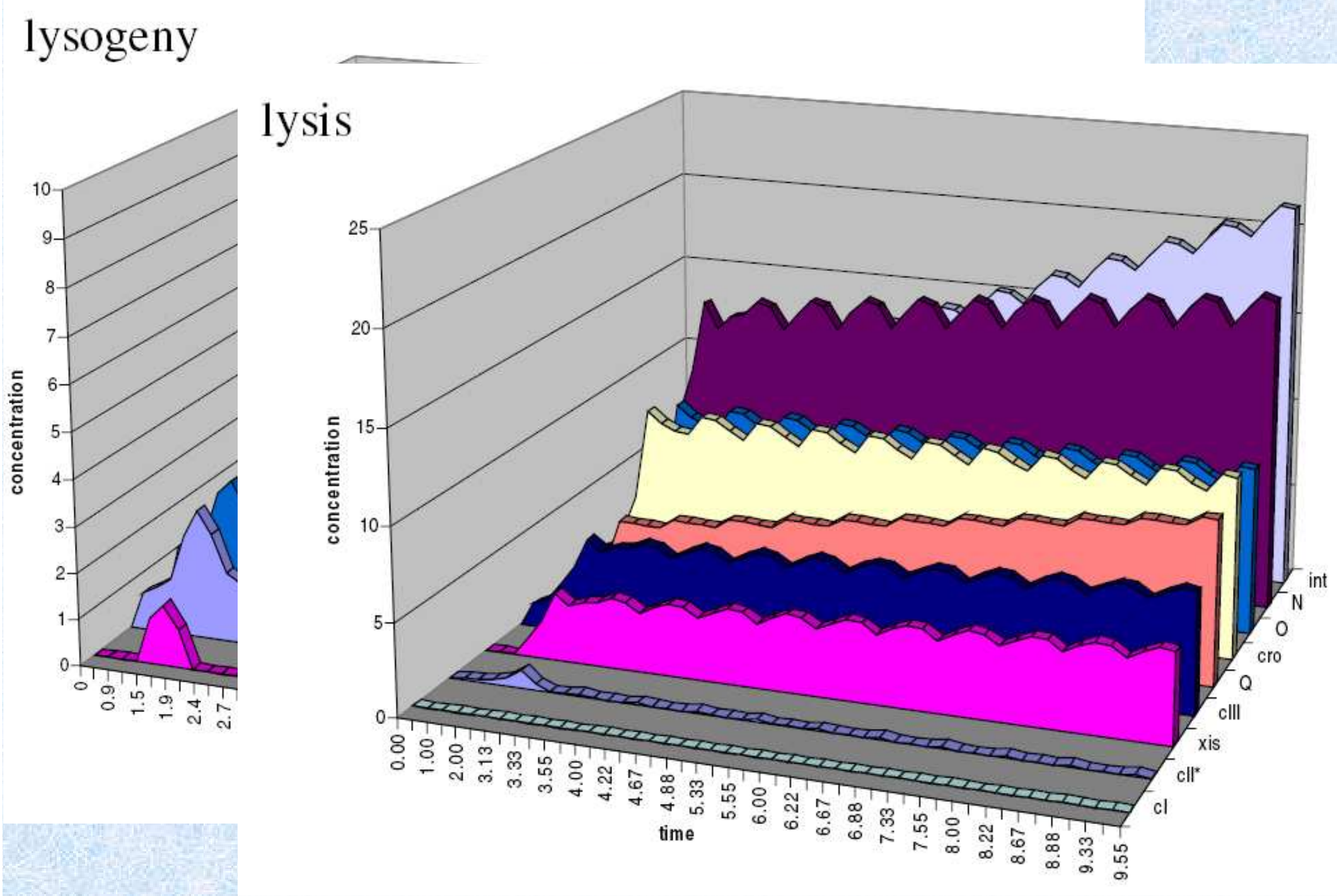




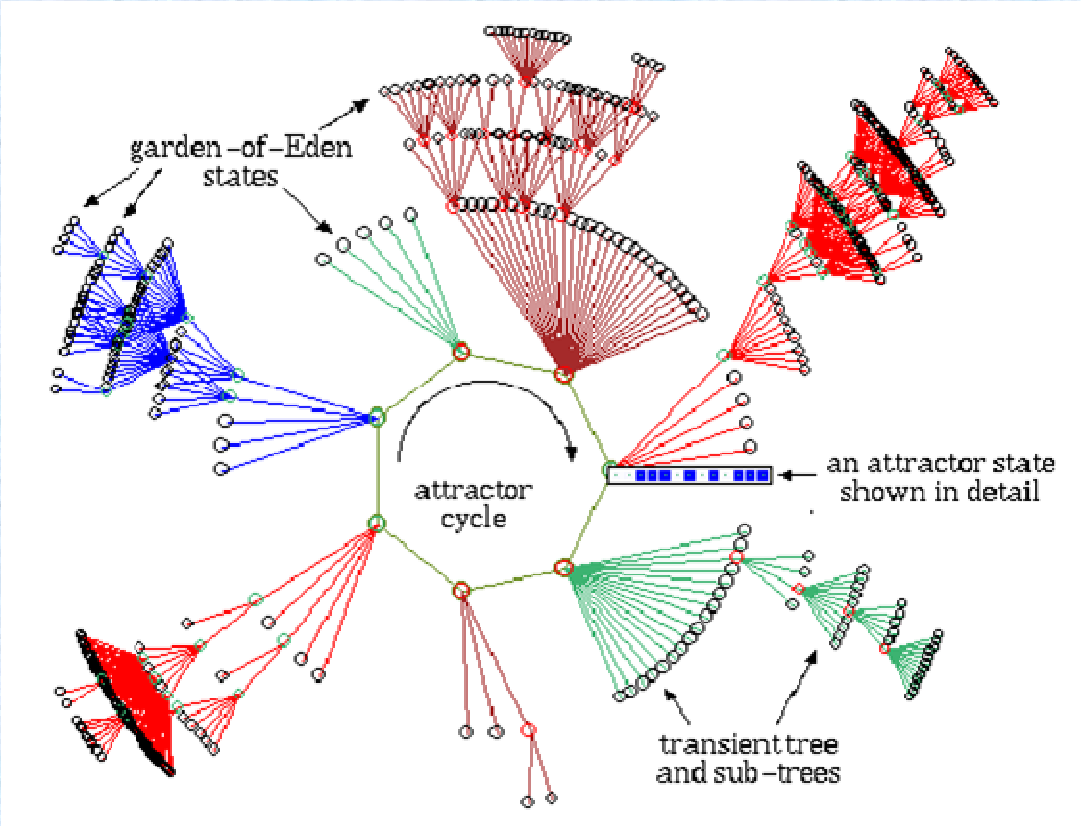
FSLM for lambda phage



Model simulation



Other notions of network "stability"



Attractors in Boolean networks

In general case we will be able to find just one attractor containing all FSLM states.

Can we be more lucky with real biological models?

States representing FSLM dynamics

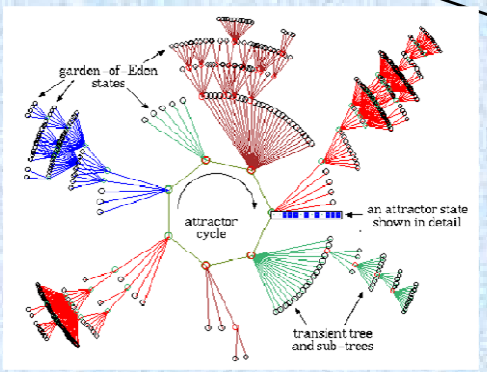
B_1 = bound by p_j
...
B_i = free
$t_{g-1} < conc(p_1) < t_g$ ↑
...
$t_{h-1} < conc(p_j) < t_h$ ↓

$conc(p_1) \geq t_{g+1}$

$conc(p_1) < t_g$

B_1 = bound by p_j
...
B_i = bound by p_1
$t_g \leq conc(p_1) \leq t_{g+1}$ ↓
...
$t_{h-1} < conc(p_j) < t_h$ ↓

$conc(p_j) \leq t_{h-1}$



B_1 = free
...
B_i = bound by p_1
$t_{g-1} < conc(p_1) < t_g$ ↑
...
$t_{h-2} < conc(p_j) \leq t_{h-1}$ ↓

We translate continuous FSLM state space to number of discrete states.
Transition between two discrete states is possible *only* if there is a transition between corresponding continuous states.
In principle this means computing something similar to state space of Boolean networks.

States representing FSLM dynamics

Each discrete state is characterized by:

- states of all binding sites (free, occupied)
- for each protein p_i the information:
-- whether concentration of p_i is increasing or decreasing
-- the position of p_i within an increasing list of thresholds $t_1(i)$, ..., $t_k(i)$

Computation of state space graph similar to depth-first search and takes $O(V + E \log V)$ time.

Analysis of state graphs

We start by looking for strongly connected components.

A strongly connected component is called *locally stable*, if it consists of a single loop or single state.

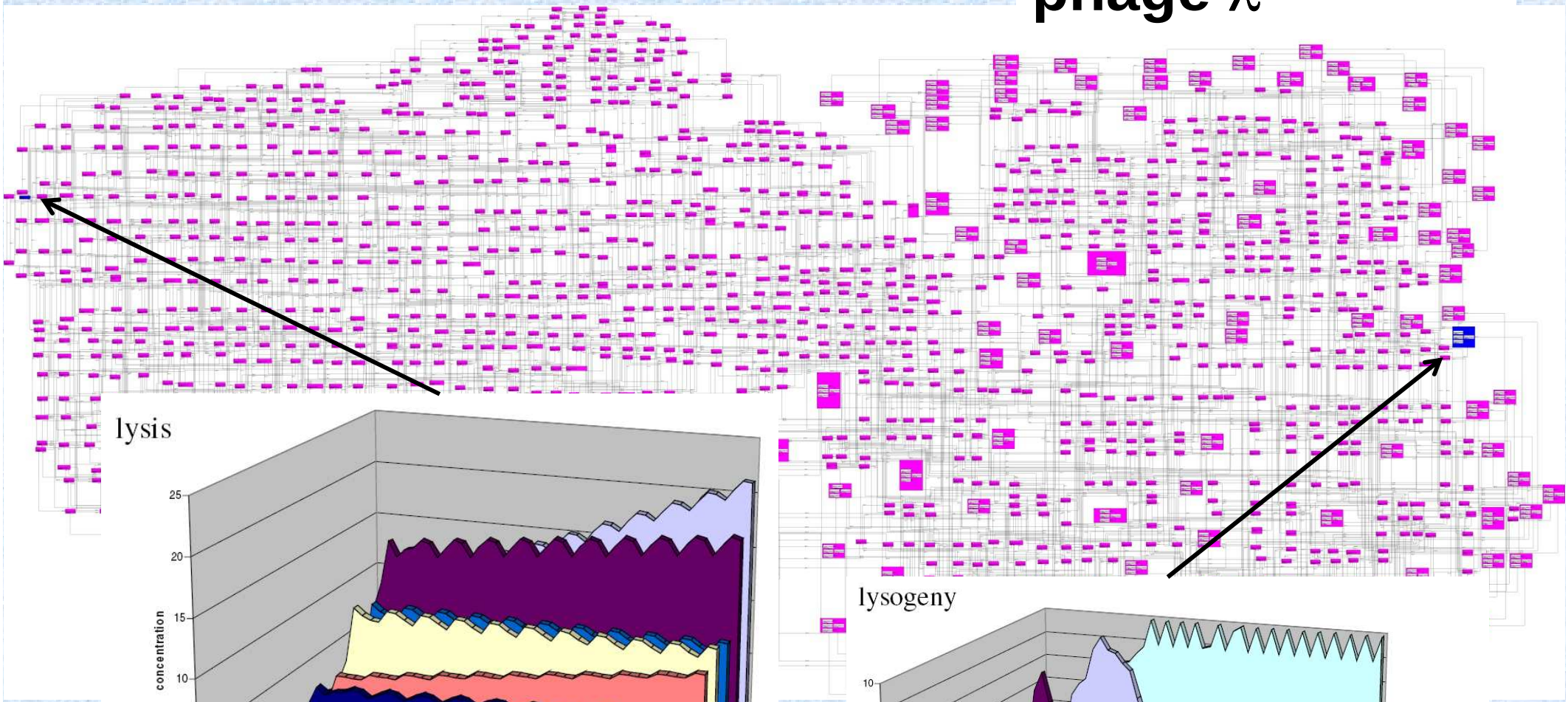
Network behaviour is *stable*, while its state is inside locally stable component.

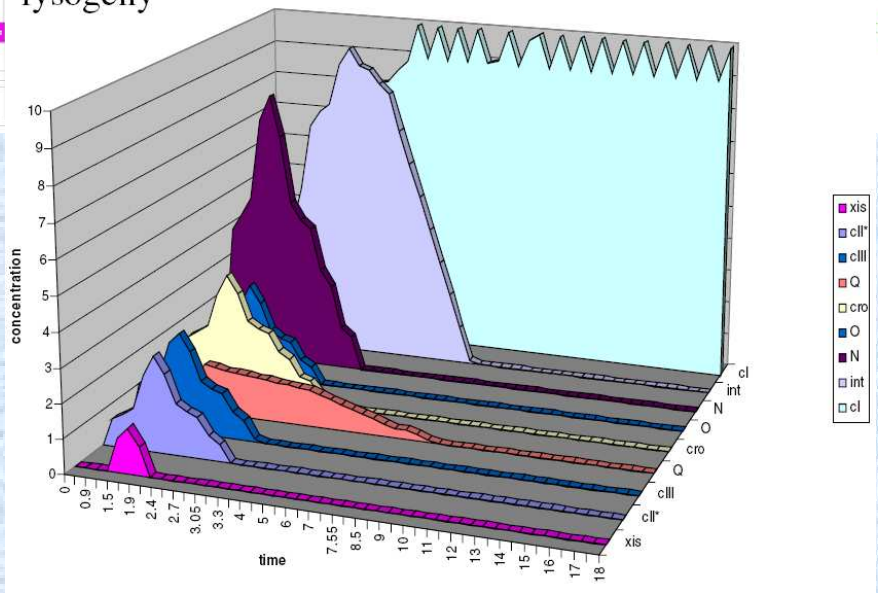
A strongly connected component is called *temporary* if it is not possible to stay inside it infinitely (as yet un-finalized set of rules – e.g. protein concentration must stay with given bounds, but it isn't regulated within the component). *Most single states tend to be temporary*.

Remaining non-temporary (and, ideally, locally stable :) components are regarded as potential analogs of attractors.

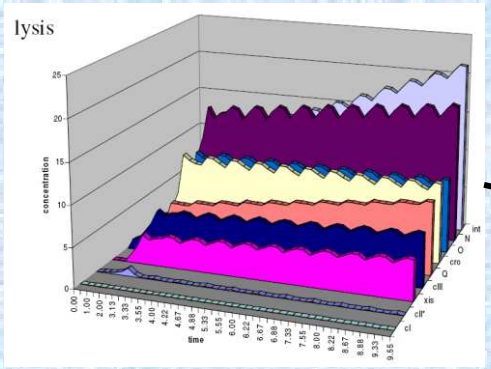
Analysis of state graphs

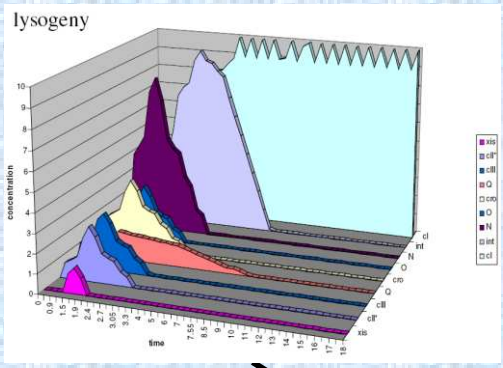
Attractors of phage λ

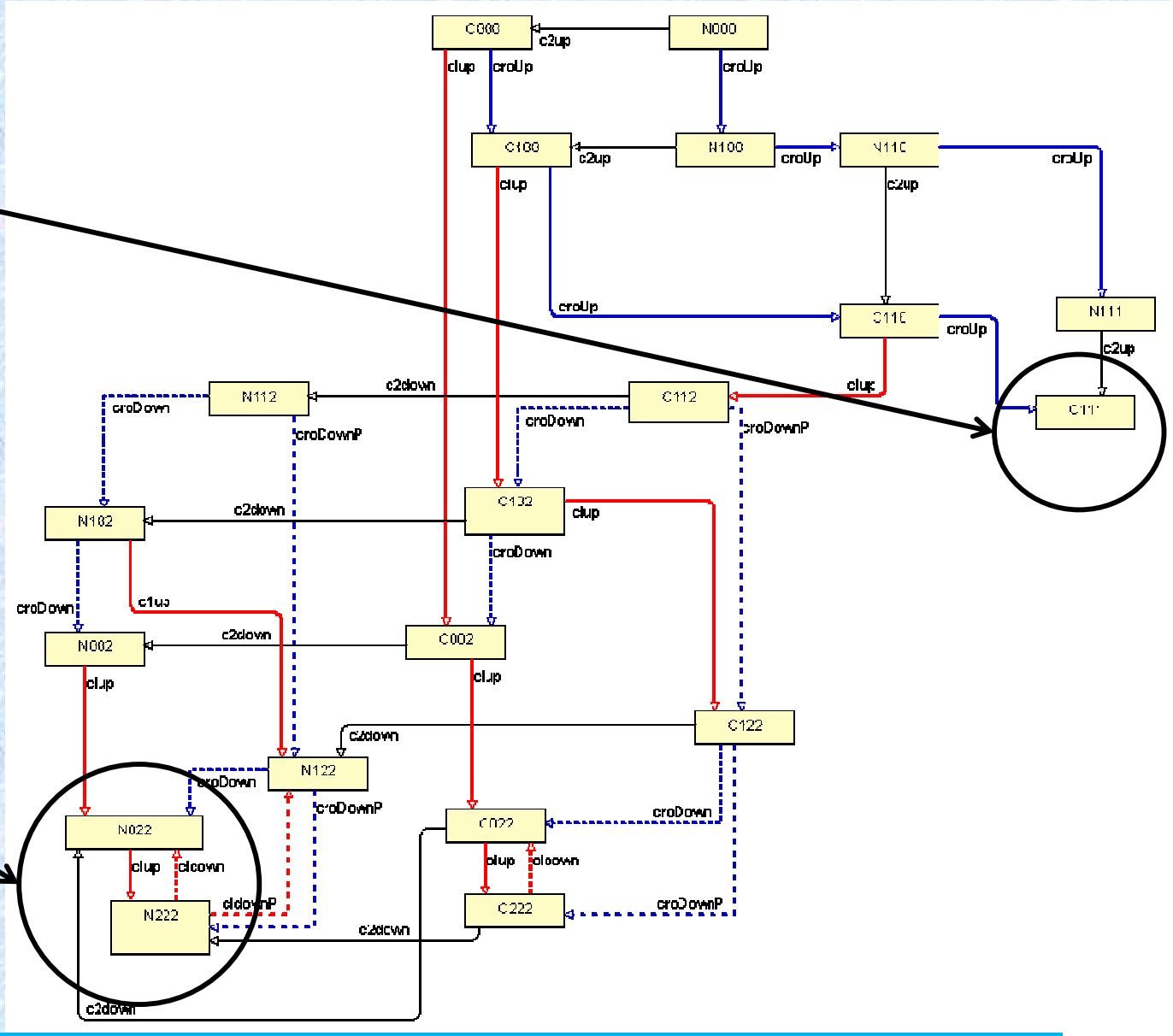




Analysis of state graphs



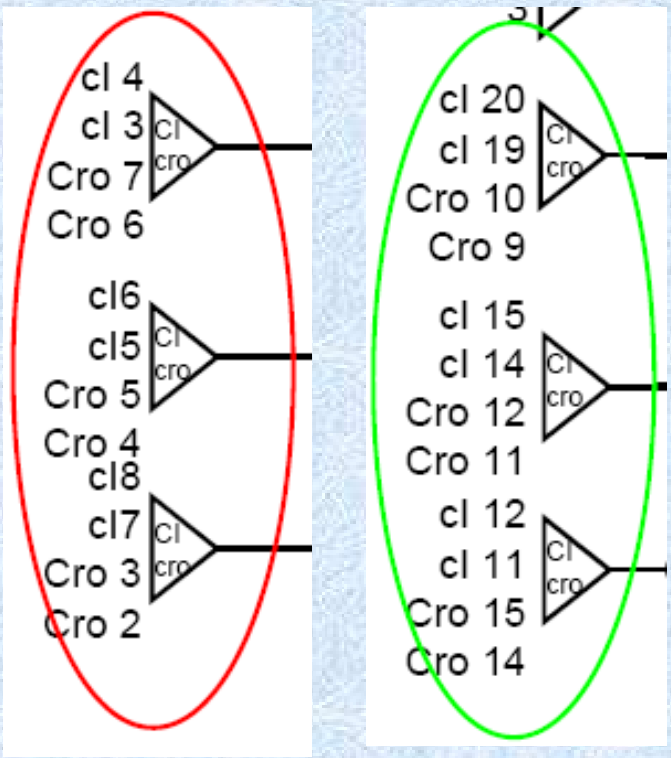




A fragment of state space with strongly connected components.

Model parameters

Topology of state graphs depends from the order of thresholds $t_1(i)$, ..., $t_k(i)$ for each protein.



For lambda phage model the attractors turned out to be influenced only by thresholds for binding sites with competing proteins.

By changing this order it is possible to obtain two other stable behaviours that doesn't seem to occur in nature.

FSLM - how good are they?

The reachability problem in state space graph in general is algorithmically unsolvable (D.Ruklisa, 2005).

At least in some "practical cases" it can be solved and even allows finding of "generalized attractors".

For most of the analysis we don’t need *linearity* of FSLM, what is important is that protein concentration either grows or decreases *monotonically*.

The constraints for particular "stable behaviours" could be quite simple - ordering of some activation/disactivation thresholds and relations between some growth and/or decrease rates.

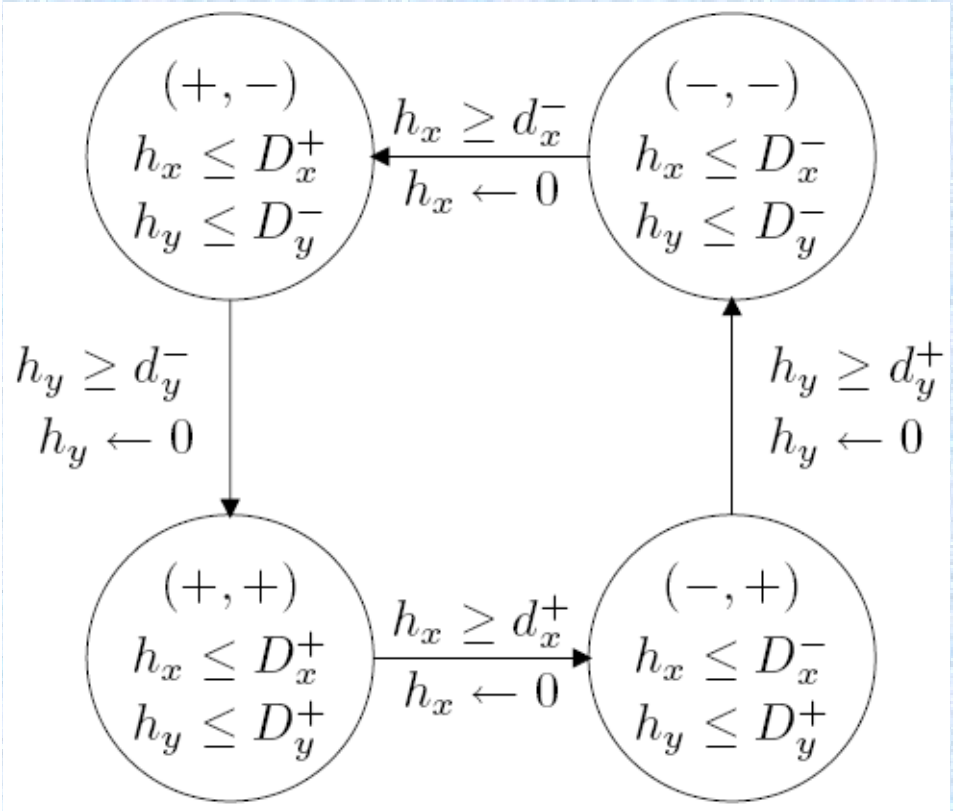
Not all important biological processes fit into this model (e.g. protein interactions).

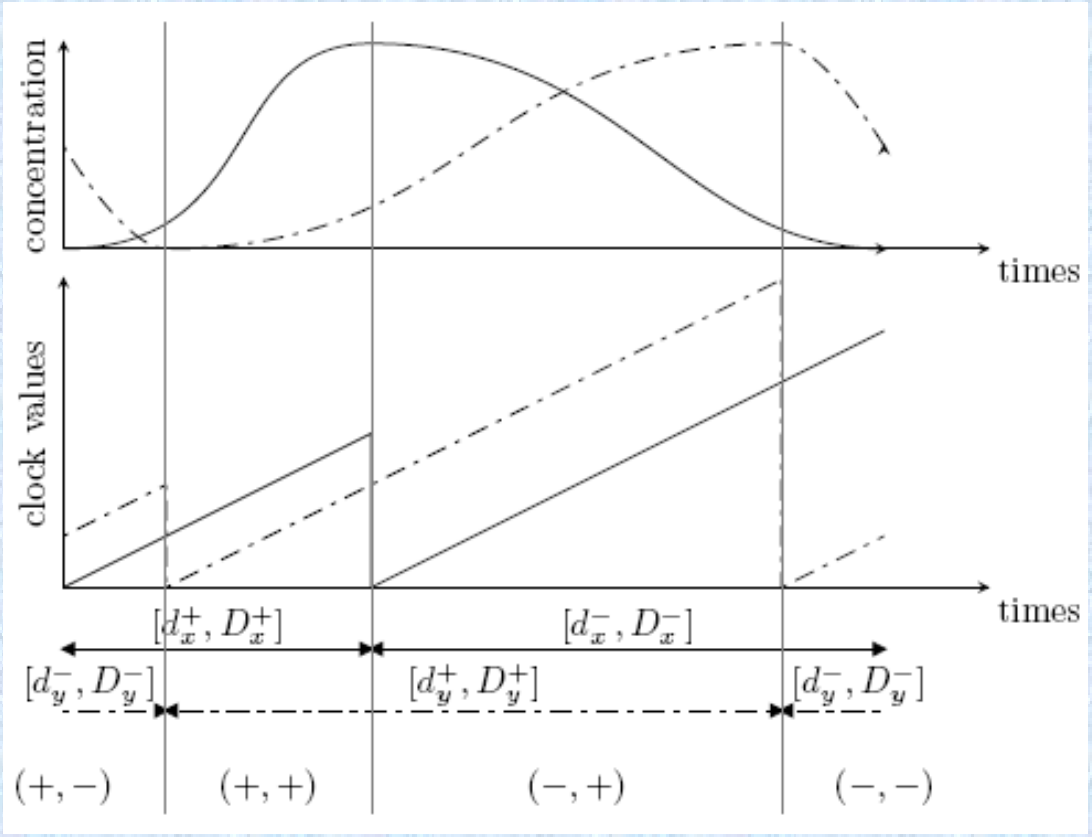
Can we somehow simplify/generalize it, whilst keeping "nice" properties?

On hybrid systems based models

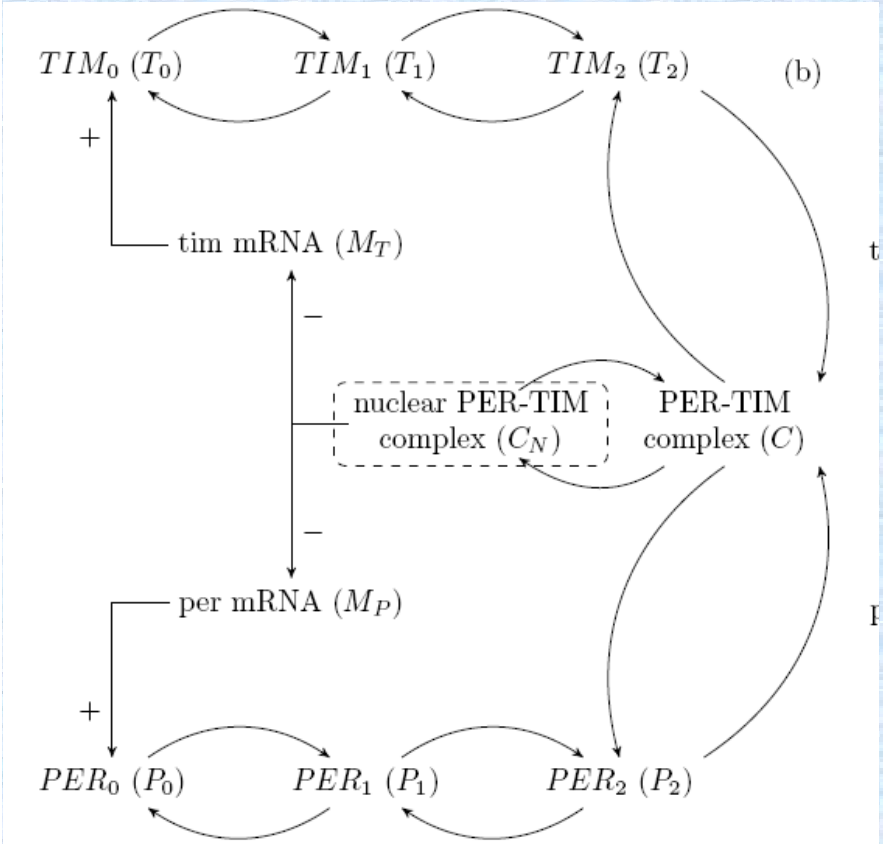
HS-based formalisms are becoming increasingly popular in latest few years.

Adaptation of Temporal Evolution Model for Drosophila circadian cycle modelling (Fromentin et al, 2010).





On hybrid systems based models



These model specifications imply three constraints that are necessary for the existence of such a cycle:

$$\begin{cases} D^-_{M_P} \geq 5 & \text{and} \quad D^-_{M_T} \geq 5 & \text{(c1)} \\ D^-_{C_N} \geq d_{M_P} - 5 & \text{and} \quad D^-_{C_N} \geq d^-_{M_T} - 5 & \text{(c2)} \\ D^+_{M_P} + D^-_{M_P} \geq d^-_{M_T} & \text{and} \quad D^+_{M_T} + D^-_{M_T} \geq d^-_{M_P} & \text{(c3)} \end{cases}$$

(Biological) model of circadian cycle and constraints that are needed for existence of such cycle.

This was obtained by using known HS formalism and existing hybrid model-checker (Phaver).

Specialized HS models for biological systems (HSM)

The general idea is to build the "weakest possible" HS model that is still sufficient to describe biological processes, thus trying to maximize our possibilities to analyze such systems.

More formally, we define a hybrid system (HS) as an 5-tuple $\mathcal{H} = \langle M, X, T, F, MF \rangle$, where:

1. $M = \{\mu_1, \ldots, \mu_k\}$ is a finite set of *modes*.
2. $X = \{x_1, \ldots, x_m\}$ is a finite set of *continuous variables* having values in $\mathbf{R}_+$. We denote the value of variable x at the time t by $x(t)$ ($x(t) \in \mathbf{R}_+$). Also we denote by $X(t) \in \mathbf{R}_+^m$ the vector of all values of variables $(x_1(t), \ldots, x_m(t))$.
3. T is a set of *mode transitions*, where each transition $\tau \in T$ has the form $\tau = (\mu_i, \mu_j, p_{i,j,l})$, where $\mu_i, \mu_j \in M$, $p_{i,j,l} : \mathbf{R}_+ \to \{true, false\}$ and predicate $p_{i,j,l}$ is defined either by $p_{i,j,l}(t) = true$ iff $x_l(t) \leq c$ or by $p_{i,j,l}(t) = true$ iff $x_l(t) \geq c$ for some constant $c \in \mathbf{R}_+$. Predicate $p_{i,j,l}$) is called *guard* and we require that for each $i, j \in \{1, \ldots, k\}$ there is at most one transition $\tau = (\mu_i, \mu_j, p_{i,j,l})$.
4. $F = \{f_1, \ldots, f_n\}$ is a set of *mode functions*, where each $f_i : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ is continuous and monotonous in both arguments. We will also call them growth/degradation functions.
5. $MF : M \times X \to F$ is a set of *mode-function assignments*. We denote $MF(\mu_i, x_l) = f_{i,l}$.

From the perspective of "descriptional power" HSM is more general than FSLM.

At the same time we still have the same, if not improved, possibilities to analyze such systems.

The team

Alvis Brāzma (EBI, Hinxton)
Dace Rukliša (EBI, Hinxton)
Karlis Čerāns (IMCS, Riga)
Juris Vīksna (IMCS, Riga)
Thomas Schlitt (King's College, London)