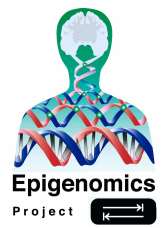
On the use of temporal formal logic to deduce the parameters of a gene regulatory network

Gilles Bernot, Jean-Paul Comet and Janine Guespin

University of Nice sophia antipolis, I3S laboratory, France

Observability Group of the Epigenomics Project



Menu

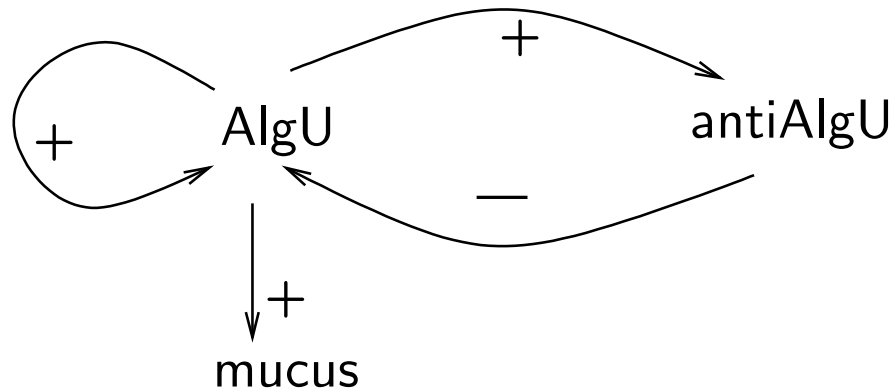
1. Modelling biological regulatory networks
2. Discrete framework for biological regulatory networks
3. Temporal logic and Model Checking for biology
4. Computer aided elaboration of formal models
5. Pedagogical example: *Pseudomonas aeruginosa*

Static Graph *v.s.* Dynamic Behaviour

Difficulty to predict the result of combined regulations

Difficulty to measure the strength of a given regulation

Example of "competitor" circuits



Positive *v.s.* Negative circuits

Even *v.s.* Odd number of "—" signs

Multistationarity *v.s.* Homeostasy

René Thomas, Snoussi, . . . , Soulé, Richard

Functional circuits "pilot" the behaviour

Mathematical Models and Simulation

1. Rigorously encode sensible knowledge, into ODEs for instance
2.
   - A few parameters are approximatively known
   - Some parameters are limited to some intervals
   - Many parameters are *a priori* unknown
3. Perform lot of simulations, compare results with known behaviours, and propose some credible values of the unknown parameters which produce robust acceptable behaviours
4. Perform additional simulations reflecting novel situations
5. If they predict interesting behaviours, propose new biological experiments
6. Simplify the model and try to go further

Mathematical Models and Validation

"Brute force" simulations are not the only way to use a computer.
We can offer computer aided environments which help:

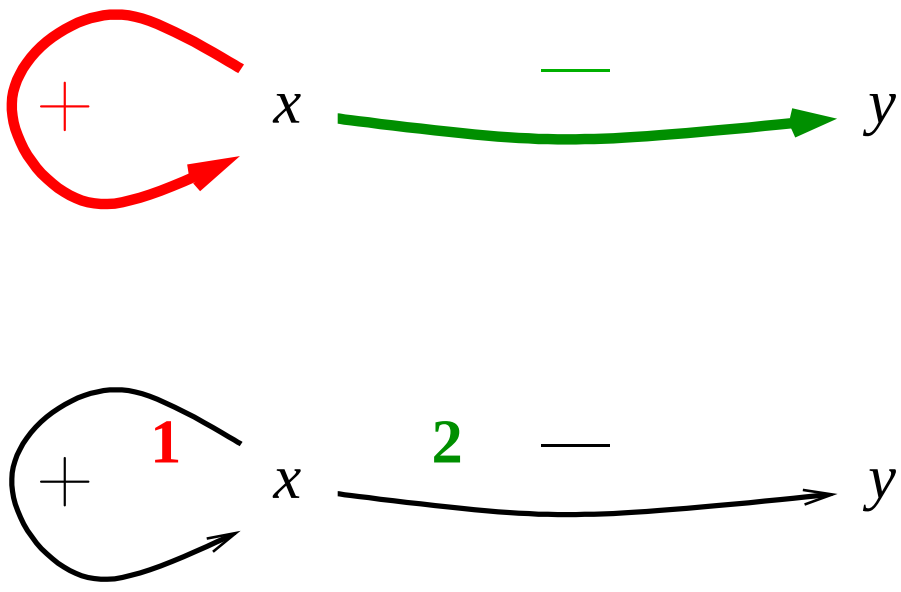
- to consider simplified models that can be anatically solved
- to avoid models that can be "tuned" *ad libitum*
- to validate models with a reasonable number of experiments
- to define only models that could be experimentally refuted
- to prove refutability w.r.t. experimental capabilities
- to establish a *methodology*: models $\leftrightarrow$ experiments

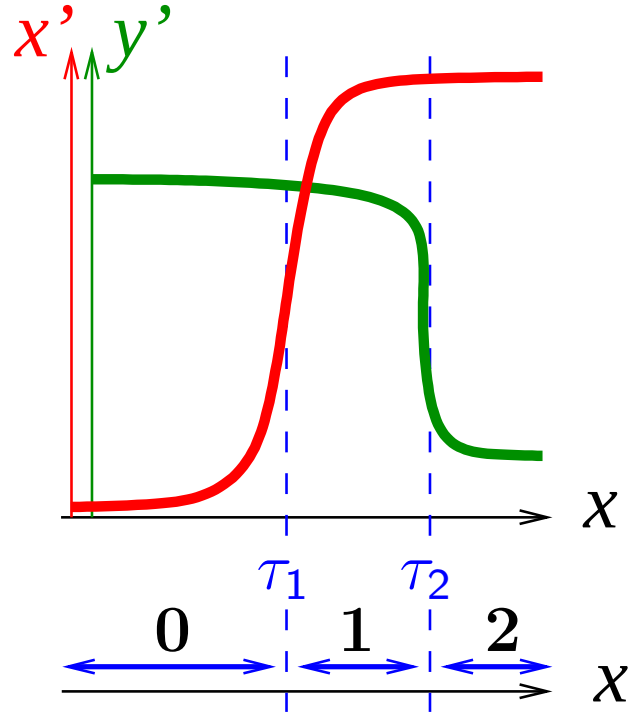
Observability issues:
Observability Group, Epigenomics Project.

Menu

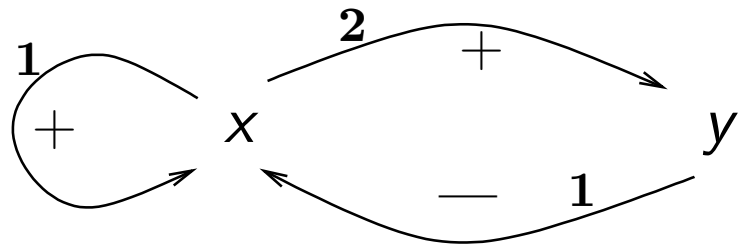
1. Modelling biological regulatory networks
2. Discrete framework for biological regulatory networks
3. Temporal logic and Model Checking for biology
4. Computer aided elaboration of formal models
5. Pedagogical example: *Pseudomonas aeruginosa*

Multivalued Regulatory Graphs





Regulatory Networks (R. Thomas)



Basal level : K_x K_y

x helps : $K_{x,x}$ $K_{y,x}$

Absent y helps : $K_{x,\overline{y}}$

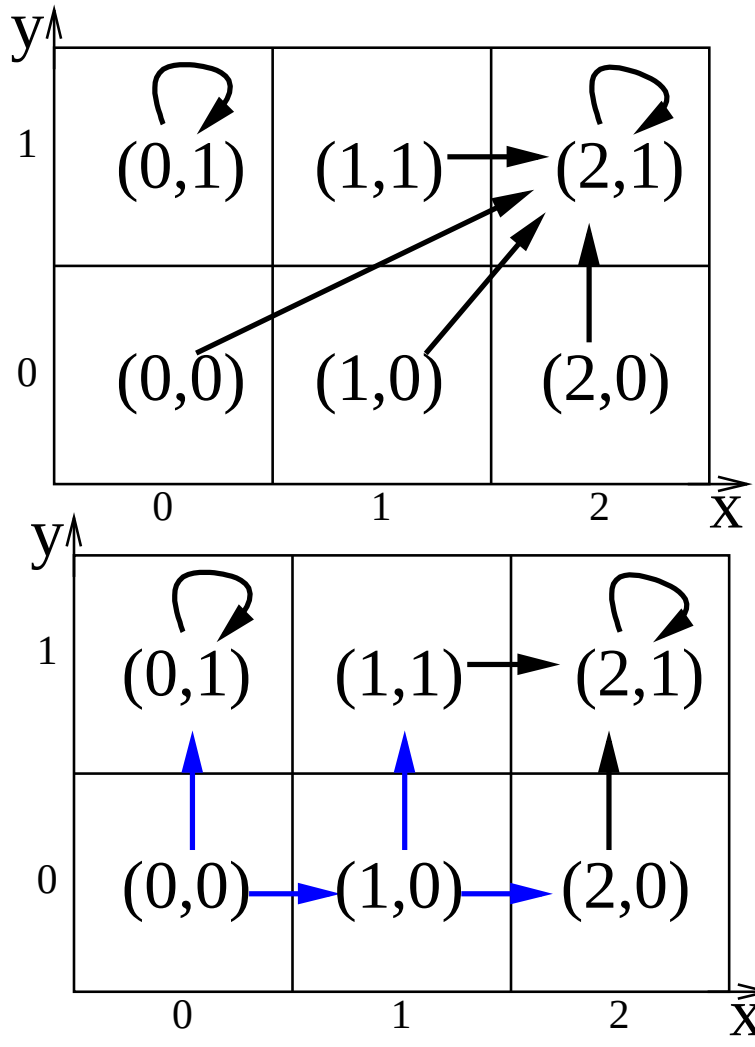
Both : $K_{x,x\overline{y}}$

(x,y)	*Focal Point*
(0,0)	$(K_{x,\overline{y}}, K_y)$
(0,1)	(K_x, K_y)
(1,0)	$(K_{x,x\overline{y}}, K_y)$
(1,1)	$(K_{x,x}, K_y)$
(2,0)	$(K_{x,x\overline{y}}, K_{y,x})$
(2,1)	$(K_{x,x}, K_{y,x})$

State Graphs

(x,y)	*Focal Point*
(0,0)	$(K_{x,\overline{y}}, K_y)=(2,1)$
(0,1)	$(K_x, K_y)=(0,1)$
(1,0)	$(K_{x,x\overline{y}}, K_y)=(2,1)$
(1,1)	$(K_{x,x}, K_y)=(2,1)$
(2,0)	$(K_{x,x\overline{y}}, K_{y,x})=(2,1)$
(2,1)	$(K_{x,x}, K_{y,x})=(2,1)$

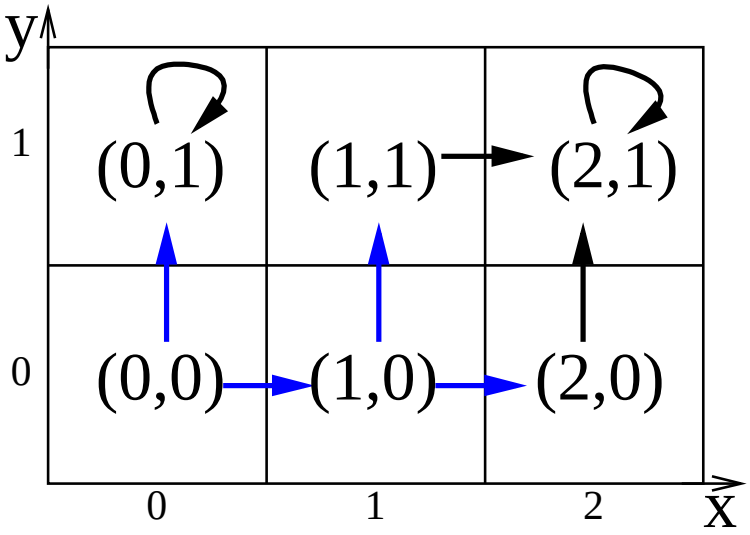
"desynchronization" $\longrightarrow$
by units of Manhattan distance



Menu

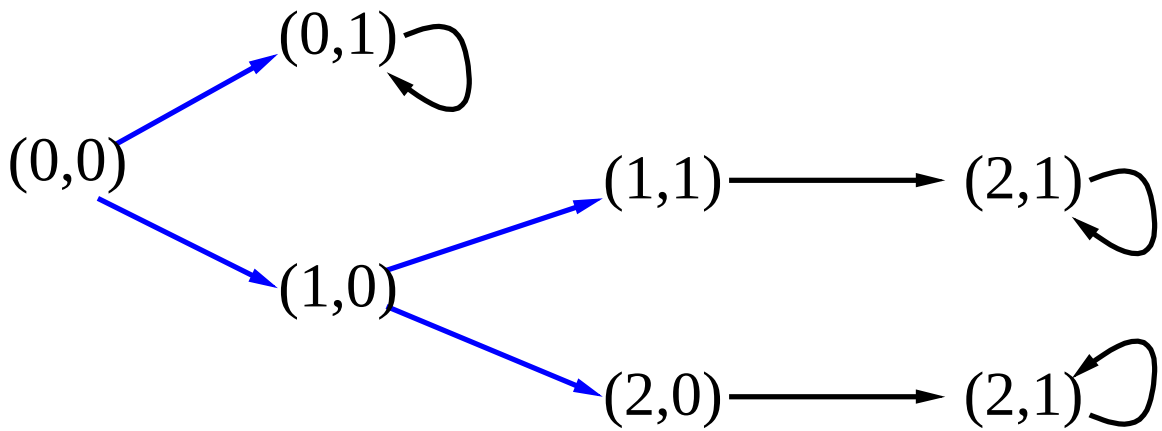
1. Modelling biological regulatory networks
2. Discrete framework for biological regulatory networks
3. Temporal logic and Model Checking for biology
4. Computer Aided elaboration Of Formal models
5. Pedagogical example: *Pseudomonas aeruginosa*

Time has a tree structure



As many possible state graphs as possible parameter sets. . . (huge number)

From an initial state:



CTL = Computation Tree Logic

Atoms = comparaisons : (x=2) (y>0) ...

Logical connectives: $(\varphi_1 \wedge \varphi_2)$ $(\varphi_1 \Longrightarrow \varphi_2)$ $\cdots$

Temporal connectives: made of 2 characters

first character	second character
A = for **A**ll path choices	X = ne**X**t state
	F = for some **F**uture state
E = there **E**xist a choice	G = for all future states (**G**lobally)
	U = **U**ntil

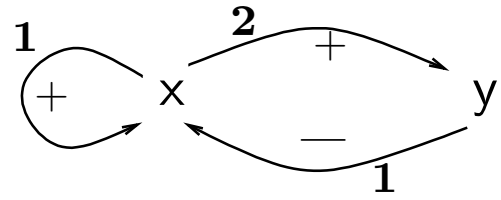
$AX(y = 1)$: the concentration level of y belongs to the interval 1 in all states directly following the considered initial state.

$EG(x = 0)$: there exists at least one path from the considered initial state where x always belongs to its lower interval.

CTL to encode Biological Properties

Common properties:

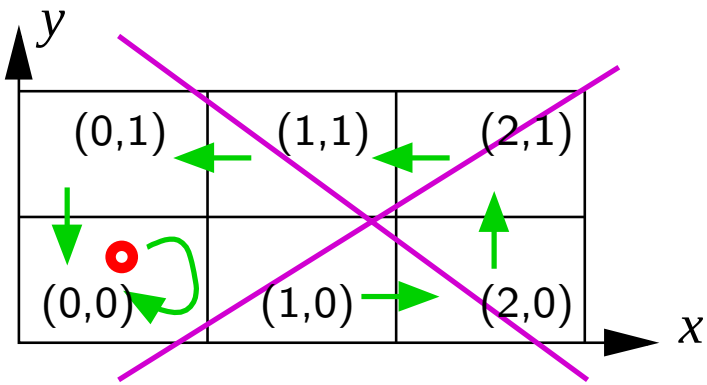
"functionality" of a sub-graph

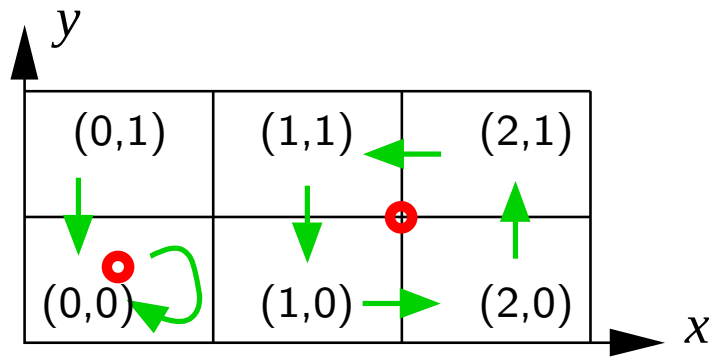


Special role of "feedback loops"

– positive: *multistationnarity* (even number of —)

– negative: *homeostasy* (odd number of —)





Characteristic properties: $\begin{cases} (x = 2) \Longrightarrow AG(\neg(x = 0)) \\ (x = 0) \Longrightarrow AG(\neg(x = 2)) \end{cases}$

They express *"the positive feedback loop is functional"*

(satisfaction of these formulae relies on the parameters $K_{...}$)

Model Checking

Efficiently computes all the states of a state graph which satisfy a given formula: $\{ \eta \mid M \models_\eta \varphi \}$.

Efficiently select the models which globally satisfy a given formula.

Theoretical Models $\leftrightarrow$ Experiments

CTL formulae are satisfied (or refuted) w.r.t. a set of paths from a given initial state

- They can be tested against the possible paths of the theoretical models ($M \models_{Model\ Checking} \varphi$)
- They can be tested against the biological experiments ($Biological\_Object \models_{Experiment} \varphi$)

CTL formulae link theoretical models and biological objects together

Menu

1. Modelling biological regulatory networks
2. Discrete framework for biological regulatory networks
3. Temporal logic and Model Checking for biology
4. Computer aided elaboration of formal models
5. Pedagogical example: *Pseudomonas aeruginosa*

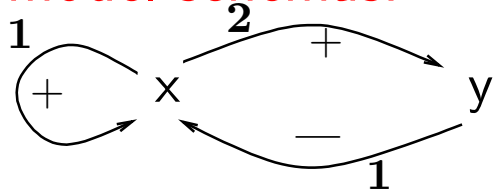
Computer Aided Elaboration of Models

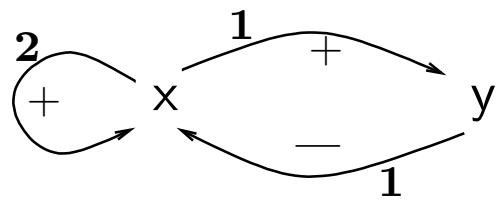
From biological knowledge and/or biological hypotheses, it comes:

- properties:

  "Without stimulus, if gene x has its basal expression level, then it remains at this level."

- model schemas:



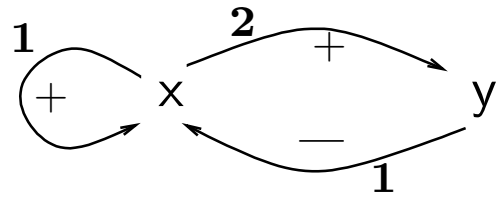


. . .

Formal logic and formal models allow us to:

- verify hypotheses and check consistency
- elaborate more precise models incrementally
- suggest new biological experiments to efficiently reduce the number of potential models

The Two Natural Questions

$\Phi = \{\varphi_1, \varphi_2, \cdots, \varphi_n\}$ and $\mathcal{M} =$ 

. . .

1. Is it possible that Φ and $\mathcal{M}$?
   Consistency of knowledge and hypotheses. Means to select models belonging to the schemas that satisfy Φ.
   $(\exists ?\ M \in \mathcal{M} \ \mid \ M \models \varphi)$
2. If so, is it true *in vivo* that Φ and $\mathcal{M}$?
   Compatibility of one of the selected models with the biological object. Require to propose experiments to **validate** or **refute** the selected model(s).

$\rightarrow$ Computer aided *proofs* and *validations*

Question 1 = Consistency

1. Draw all the sensible regulatory graphs with all the sensible threshold allocations. It defines $\mathcal{M}$.
2. Express in CTL the known behavioural properties as well as the considered biological hypotheses. It defines Φ.
3. Consider all the possible state graphs derived from $\mathcal{M}$ (i.e., all possible parameters $K_{...}$) and check each of them against Φ. Our software plateform SMBioNet handles this automatically.
4. If no model survive to the previous step, then reconsider the hypotheses and perhaps extend model schemas...
5. If at least one model survives, then the biological hypotheses are consistent. Possible parameters $K_{...}$ have been established.

Now Question 2 has to be addressed

Question 2 = Validation

1. Among all possible formulae, some are "observable" i.e., they express a possible result of a possible biological experiment. Let *Obs* be the set of all observable formulae.
2. Let $Th(\Phi, \mathcal{M})$ be the set of consequences of Φ and $\mathcal{M}$. $Th(\Phi, \mathcal{M}) \cap Obs$ is the set of experiments able to validate the survivors of Question 1. Unfortunately it is infinite in general.
3. Select a finite subset of $Th(\Phi, \mathcal{M}) \cap Obs$ that maximizes the chance to refute the survivors
4. Perform these experiments.

Sometimes a *complete* and small set of experiments exists.
It has been the case of the mucus production of *P.aeruginosa*.

Menu

1. Modelling biological regulatory networks
2. Discrete framework for biological regulatory networks
3. Temporal logic and Model Checking for biology
4. Computer aided elaboration of formal models
5. Pedagogical example: *Pseudomonas aeruginosa*

Mutation, Epigenesis, Adaptation

Terminology about phenotype modification:

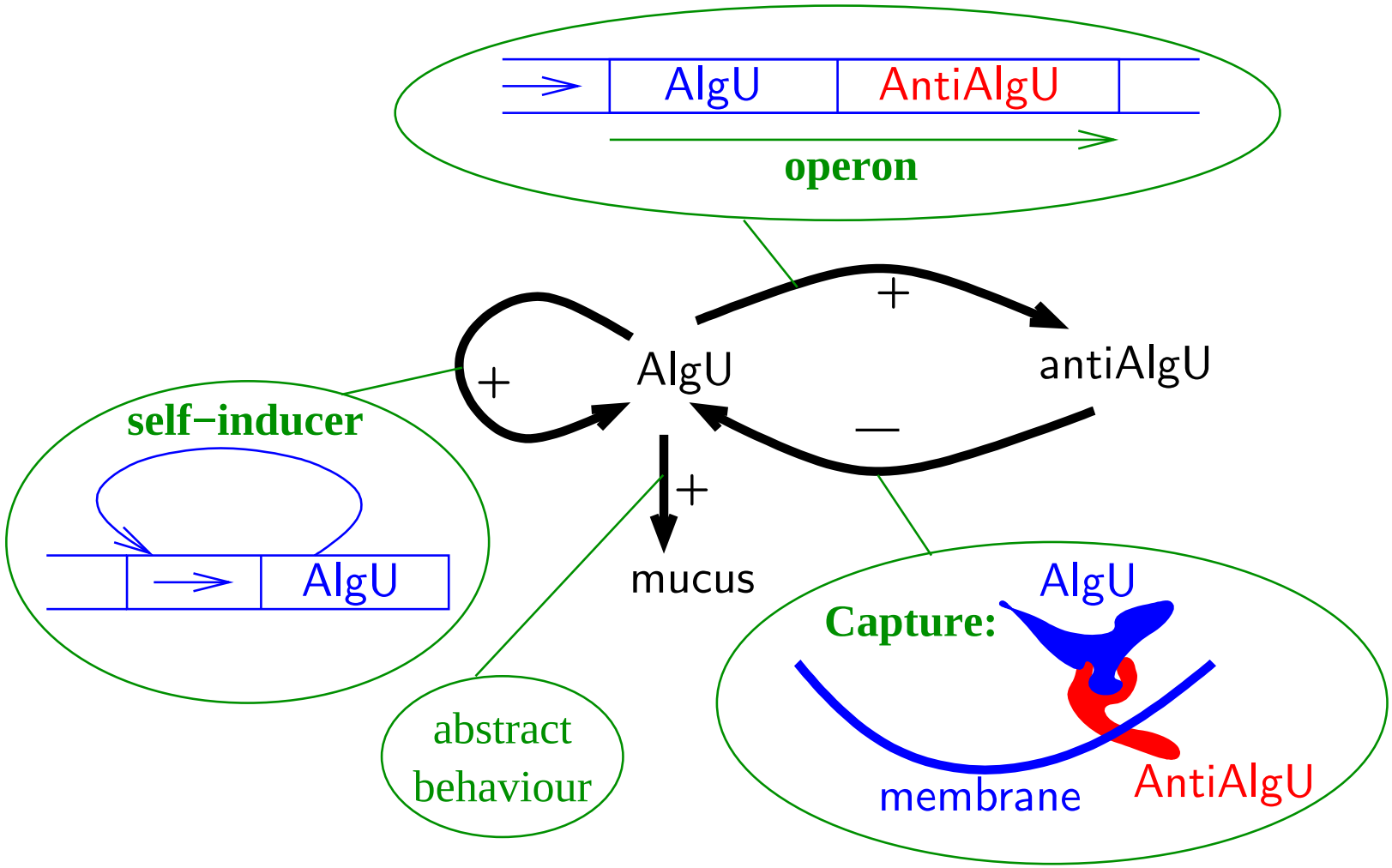
genetic modification: inheritable and not reversible (mutation)
epigenetic modification: inheritable and reversible
adaptation: not inheritable and reversible

Pseudomonas aeruginosa is an opportunistic bacteria that produces mucus in the lungs of patients (often lethal in cystic fibrosis)

The biological question (Janine Guespin):
could **mucus production** in *P. aeruginosa* be the result of an epigenetic switch ?

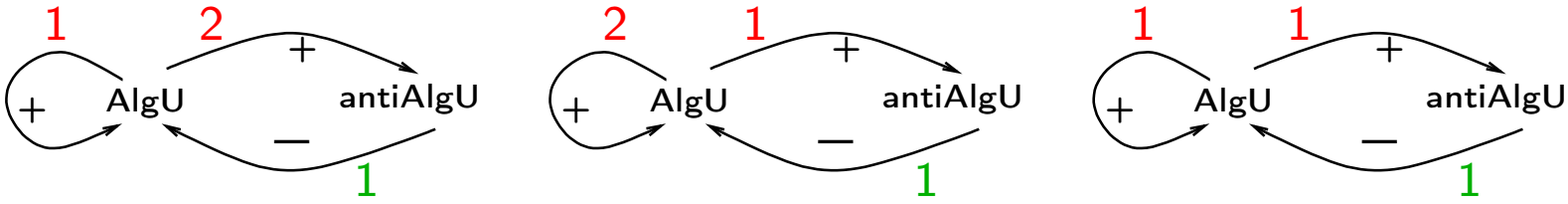
It would open the door to new possible therapies

Mucus Production in *P. aeruginosa*



Parameters & thresholds: unknown

Thresholds for AlgU in *P.aeruginosa* are unknown:



and parameters are unknown:

$$3^4 \times 2^2 \qquad 3^4 \times 2^2 \qquad 2^4 \times 2^2$$

712 possible models

One CTL formula for each stable state:

$$(\mathrm{AlgU} = 2) \Longrightarrow AXAF(\mathrm{AlgU} = 2)$$
$$(\mathrm{AlgU} = 0) \Longrightarrow AG(\neg(\mathrm{AlgU} = 2))$$

Question 1, consistency: proved by *Model Checking*

$\rightarrow$ 10 models among the 712 models are extracted by SMBioNet

Validation of the epigenetic hypothesis

Question 2 = to validate bistationnarity *in vivo*

Non mucoid state: $(\mathrm{AlgU} = 0) \Longrightarrow AG(\neg(\mathrm{AlgU} = 2))$

P. aeruginosa, with a basal level for AlgU *does not produce mucus spontaneously*: actually validated

Mucoid state: $(\mathrm{AlgU} = 2) \Longrightarrow AX(AF(\mathrm{AlgU} = 2))$

Experimental limitation (1999-2000):
— AlgU can be saturated but it cannot be measured.
— Mucus production can be observed.

Experiment:

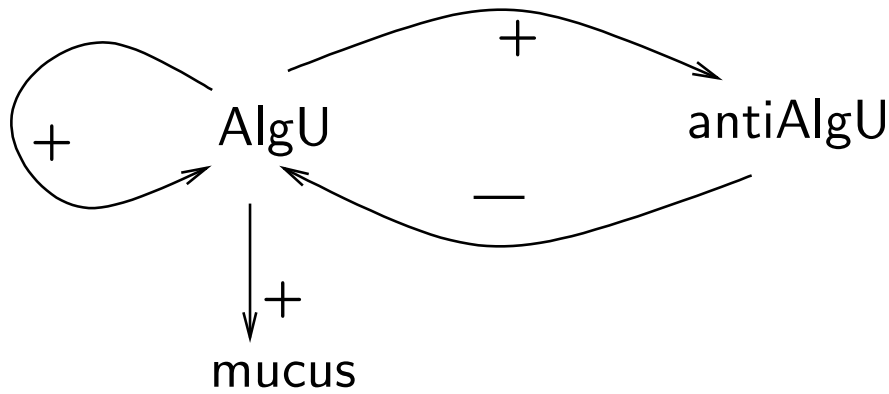
to pulse AlgU *and then to test if mucus production remains*

($\Longleftrightarrow$ to verify a hysteresis)

This experiment can be generated automatically

To test $(\mathrm{AlgU}{=}2) \Longrightarrow AXAF(\mathrm{AlgU}{=}2)$

$\mathrm{AlgU} = 2$ cannot be directly verified but $mucus = 1$ can be verified.



Lemma: $AXAF(\mathrm{AlgU} = 2) \Longleftrightarrow AXAF(mucus = 1)$
(. . . formal proof by computer . . .)

$\rightarrow$ To test: $(\mathrm{AlgU} = 2) \Longrightarrow AXAF(mucus = 1)$

$$(\mathrm{AlgU} = 2) \Longrightarrow AXAF(mucus = 1)$$

Karl Popper:

to validate = to try to refute

thus A=false is useless

experiments must begin with a pulse

$A \Longrightarrow B$	*true*	*false*
true	true	false
false	true	true

The pulse forces the bacteria to reach the initial state $\mathrm{AlgU} = 2$.
If the state were not directly controlable we had to prove lemmas:

$$(\textit{something reachable}) \Longrightarrow (\mathrm{AlgU} = 2)$$

General form of a test:

$$(\textit{something reachable}) \Longrightarrow (\textit{something observable})$$

Concluding Comments

Behavioural *properties* (Φ) are as much important as *models* ($\mathcal{M}$)

Modelling is significant only with respect to the considered experimental *reachability* and *observability* (*Obs*)

Formal proofs can suggest wet experiments

Based on the same ideas as SMBioNet, more elaborated approaches exist:

- Hybrid approaches with chronometric considerations
- BIOCHAM also considers metabolic networks
- Computer aided weakening of inconsistent hypotheses
- ...