Formal Methods from Computer Science to study Biological Networks

Gilles Bernot

Programme d'Épigénomique, Genopole®

Epigenomics Project

genopole®

Menu

1. Simulation *vs.* Validation
2. Formal Methods for the Modelling Activity
3. Example of Regulatory Networks & Temporal Logic
4. Example of the Example: *Pseudomonas aeruginosa*

Mathematical Models and Simulation

1. Rigorously encode sensible knowledge into mathematical formulae
2. - Some parameters are well defined, e.g. from biochemical knowledge
   - Some parameters are limited to some intervals
   - Some 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 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:

- Avoid models that can be “tuned” *ad libitum*
- Validate models with a reasonable number of experiments
- Only define models that could be experimentally refuted
- Prove refutability w.r.t. experimental capabilities

Observability issues:

Groupe Observabilité, Programme d’Épigénomique.

Modeling for Understanding

Computer aided modelling approaches

- Elementary modes of metabolic pathways
- Process Algebras
- Chemical Abstract Machine [BioCHAM]
- Discrete modeling of regulatory network (René Thomas) [SMBioNet, GNA]
- ...

Underlying theories:

- Operational Research
- Pi-Calculus
- Temporal logics
- ...

Different Mathematical Cultures

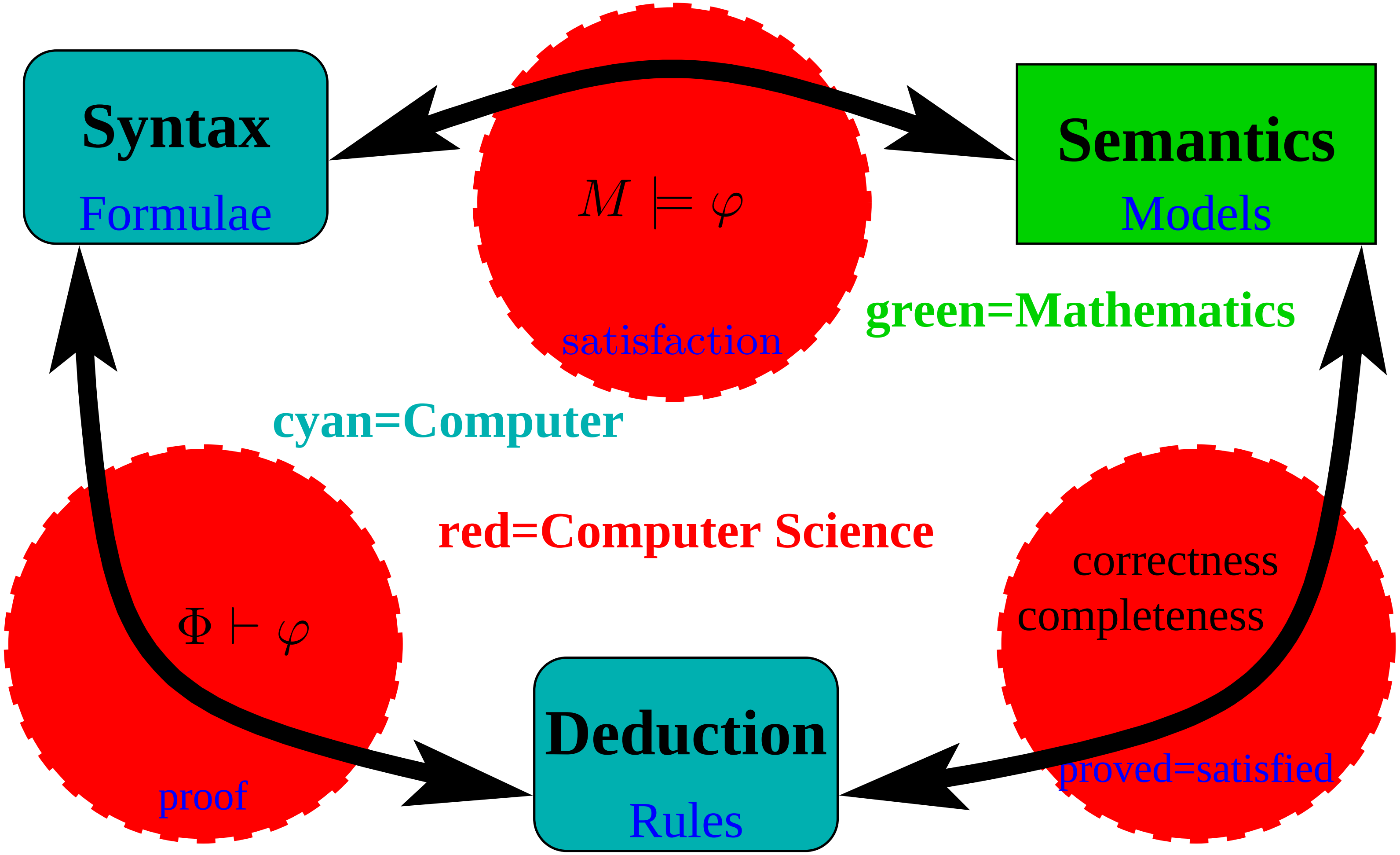
- Analytical *vs.* Algebraic Mathematics
- Continuous *vs.* Discrete computational approaches

Difficulty to manage hybrid approaches: ongoing researches.

Menu

1. Simulation *vs.* Validation
2. Formal Methods for the Modelling Activity
3. Example of Regulatory Networks & Temporal Logic
4. Example of the Example: *Pseudomonas aeruginosa*

Formal Logic: syntax/semantics/deduction



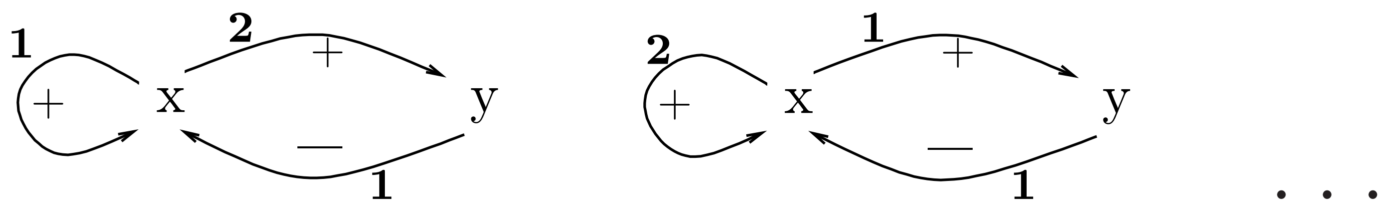
Computer Aided Elaboration of Models

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."

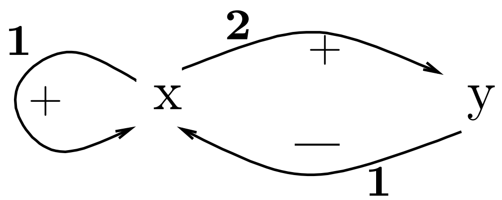
- model schemas:



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 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 $\mathcal{M}$?

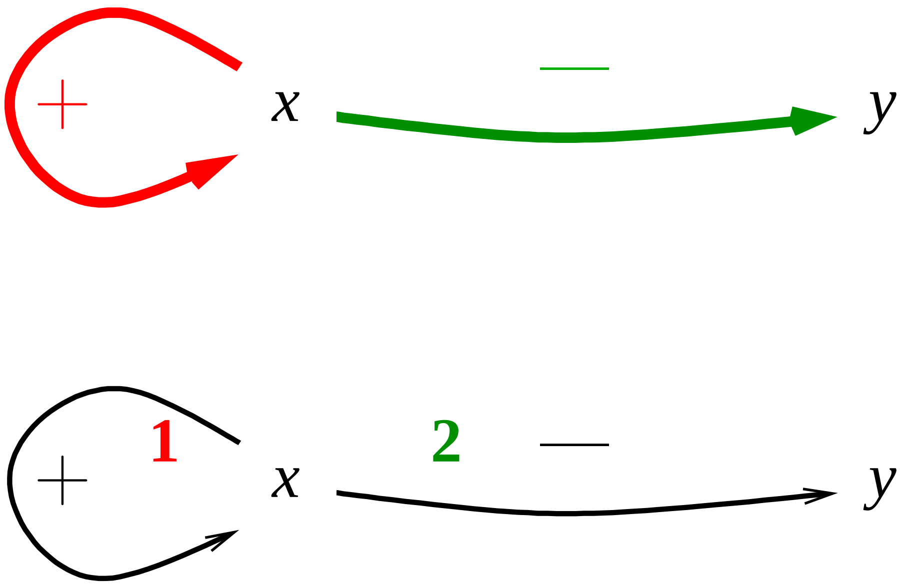
   Compatibility of one of the selected models with the biological object. Require to propose experiments to **validate** (or **refute**) the selected model(s).

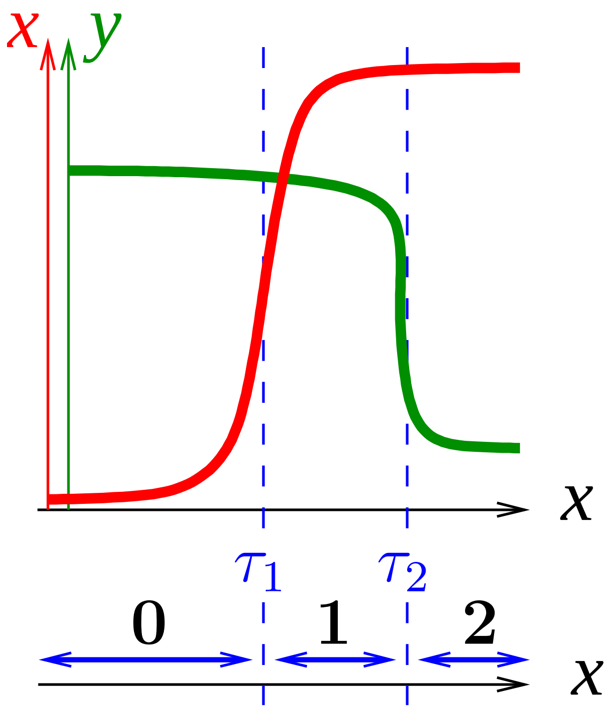
$\rightarrow$ Computer aided *proofs* and *validations*

Menu

1. Simulation *vs.* Validation
2. Formal Methods for the Modelling Activity
3. Example of Regulatory Networks & Temporal Logic
4. Example of the Example: *Pseudomonas aeruginosa*

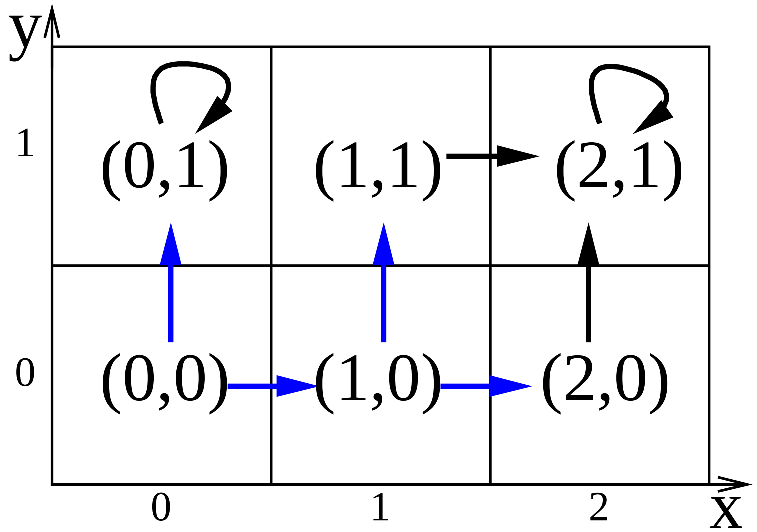
Multivalued Regulatory Graphs



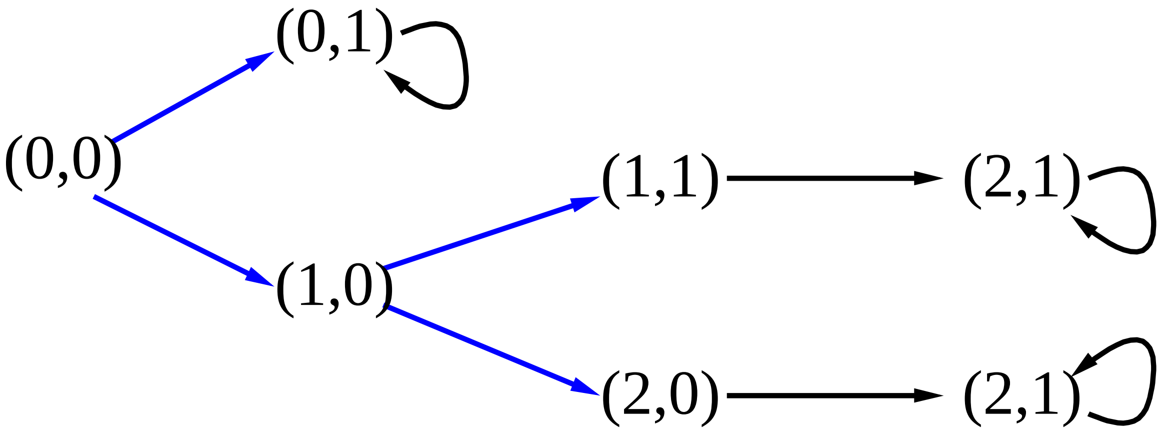


State Graphs

(x,y)	*Image*
(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)



Time has a tree structure:



CTL = Computation Tree Logic

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

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

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.

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

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. Automatically generate all the possible regulatory networks derived from $\mathcal{M}$ according to all possible parameters $K_{\ldots}$.
   Our software plateform SMBioNet handles this automatically.
4. Check each of these models against Φ.
   SMBioNet uses model checking to perform this step.
5. If no model survive to the previous step, then reconsider the hypotheses and perhaps extend model schemas. . .
6. If at least one model survives, then the biological hypotheses are consistent. Possible parameters $K_{\ldots}$ have been indirectly 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 Λ be the set of theorems of Φ and $\mathcal{M}$. $\Lambda \cap Obs$ is the set of experiments able to validate the survivors of Question 1. Unfortunately it is infinite in general.
3. Testing frameworks from computer science aim at selecting a finite subsets of these observable formulae, which maximize the chance to refute the survivors.
4. These subsets are often too big but in some cases, these testing frameworks can be applied to regulatory networks. It has been the case of the cytotoxicity of *P.aeruginosa*.

Menu

1. Simulation *vs.* Validation
2. Formal Methods for the Modelling Activity
3. Example of Regulatory Networks & Temporal Logic
4. Example of the Example: *Pseudomonas aeruginosa*

Example of *P.aeruginosa*

Terminology about phenotype modification:

Genetic modification: inheritable and not reversible (mutation)

Epigenetic switch: inheritable and reversible

Adaptation: not inheritable and reversible

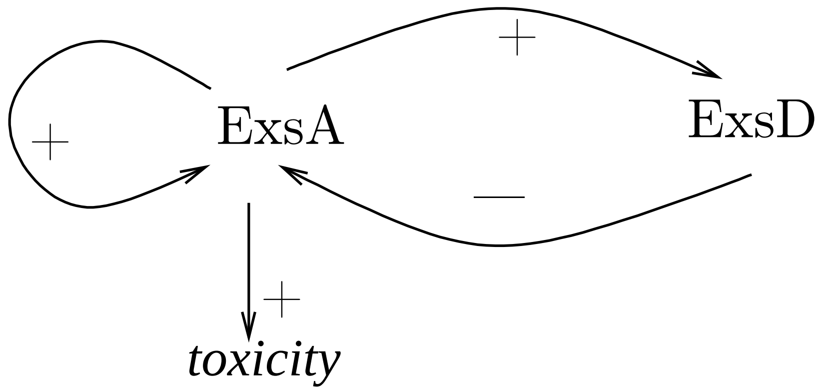
The biological questions (Janine Guespin):

is **cytotoxicity** in *Pseudomonas aeruginosa* due to an epigenetic switch ?

[$\rightarrow$ cystic fibrosis]

Cytotoxicity in *P. aeruginosa*

(Janine Guespin)

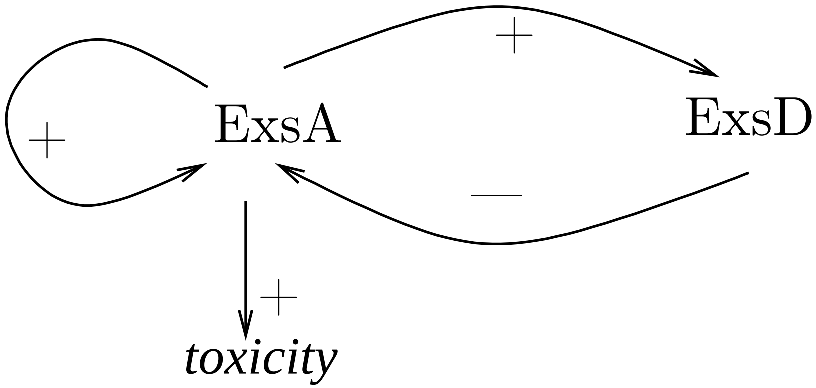


Epigenetic hypothesis =

→ The positive feedback circuit is functional, with a cytotoxic stable state and the other one is not cytotoxic.

→ An external signal (in the cystic fibrosis' lungs) could switch ExsA from its lower stable state to the higher one.

Consistency of the Hypothesis



One CTL formula for each stable state:

$$(\text{ExsA} = 2) \Longrightarrow AX\,AF(\text{ExsA} = 2)$$
$$(\text{ExsA} = 0) \Longrightarrow AG(\neg(\text{ExsA} = 2))$$

Question 1, consistency: proved by *Model Checking*

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

Question 2: and *in vivo* ? ...

Validation of the epigenetic hypothesis

Question 2 = to validate bistationnarity *in vivo*

Non cytotoxic state: $(\text{ExsA} = 0) \Longrightarrow AG(\neg(\text{ExsA} = 2))$

P. aeruginosa, with a basal level for ExsA *does not become spontaneously cytotoxic*: actually validated

Cytotoxic state: $(\text{ExsA} = 2) \Longrightarrow AXAF(\text{ExsA} = 2)$

Experimental limitation:

ExsA can be saturated but it cannot be measured.

Experiment:

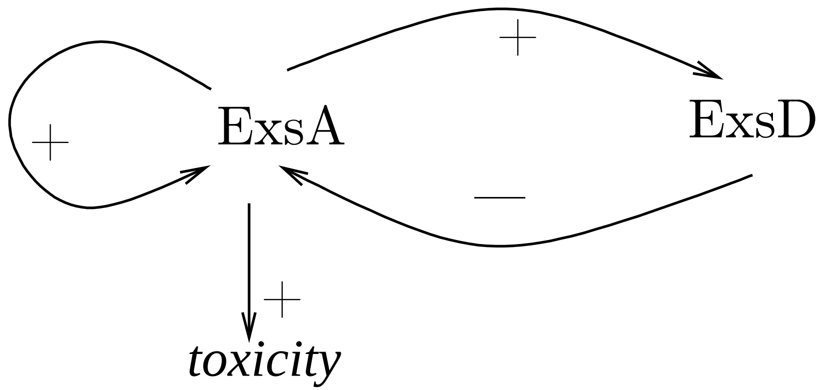
to pulse ExsA *and then to test if toxin production remain.*

($\Longleftrightarrow$ to verify a hysteresis)

This experiment can be generated automatically

To test $(\text{ExsA}=2) \Longrightarrow AXAF(\text{ExsA}=2)$

ExsA $= 2$ cannot be directly verified but $toxicity = 1$ can be verified.



Lemma: $AXAF(\text{ExsA} = 2) \Longleftrightarrow AXAF(toxicity = 1)$

(... formal proof by computer ...)

$\rightarrow$ $\boxed{\text{To test: } (\text{ExsA} = 2) \Longrightarrow AXAF(toxicity = 1)}$

$$(\text{ExsA} = 2) \Longrightarrow AXAF(toxicity = 1)$$

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 ExsA = 2.

If the state were not directly controlable we had to prove lemmas:

$$(\text{ExsA} = 2) \Longleftarrow (something\ reachable)$$

General form of a test:

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

Concluding Slogans

- Behavioural *properties* (Φ) are as much important as models ($\mathcal{M}$) for the modelling activity
- Modelling is significant only with respect to the considered experimental *reachability* and *observability* (Obs)
- The bigger is the risk of *refutation*, the better are the "surviving" models (Popper), thus models should be "simple" with few non observable parameters (Occam)

Formal methods (*syntax/semantics/proofs*) facilitate *abstraction* and consequently they simplify models

- They ensure *consistency* of the modelling activity
- They allow us to perform computer aided *validations* of models
- They take benefit of 30 years of researches in computer sciences