Biological Regulatory Networks: Logical Description and Model Checking

G. Bernot,
Programme ÉPIGÉNOMIQUE, Genopole®

Special thanks to J. Guespin, J-P. Comet & the Observability group
1. Modelling biological regulatory networks
2. Formal framework for biological regulatory networks
3. Temporal logic and Model Checking
4. Computer aided elaboration of formal models
5. Example: mucus production in *Pseudomonas aeruginosa*
Molecular Biology & Causality

Heaviness of “causality networks”

Causality loops

Counter-intuitive resulting behaviours

Predicting dynamics from models

Computer aided modelling methodologies

• quantitative approaches → e.g. differential equations
• qualitative approaches → logic & computer science can help
• mixed approaches → cf. Marcelline Kaufman

Biological questions are often of qualitative nature
Formal Logic: syntax/semantics/deduction

Syntax
Formulae

Semantics
Models

Deduction
Rules

Formulae $\Phi \vdash \varphi$

Models $M \models \varphi$

Rules proved=satisfied

Semantics correctness completeness

cyan=Computer

green=Mathematics

red=Computer Science
Regulatory Networks

To model direct or indirect regulations between biological objects (e.g. gene, macromolecule, signal, ...)

Direct: transcription factor, operon, repressor, ...

Indirect: cascade of events, capture of macromolecules, ...

\[ x \quad \text{induces} \quad y : \quad x \quad \rightarrow^{+} \quad y \]

\[ x \quad \text{inhibits} \quad y : \quad x \quad \rightarrow^{-} \quad y \]
Mucus Production in *P. aeruginosa*

**Operon**

- **AlgU**
- **AntiAlgU**

**Capture:**

- **AlgU**
- **AntiAlgU**

**Membrane**

**Abstract Behaviour**

- **Self-inducer**
- **Mucus**
Static Graph & 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

Functional circuits “pilot” the behaviour
1. Modelling biological regulatory networks
2. Formal framework for biological regulatory networks
3. Temporal logic and Model Checking
4. Computer aided elaboration of formal models
5. Example: mucus production in *Pseudomonas aeruginosa*
Multivalued Regulatory Graphs
Definition of Regulatory Graphs

A labelled directed graph \((\mathcal{V}, \mathcal{E})\)

- each node of \(\mathcal{V}\) is a variable \(x\) with a boundary \(b_x \in \mathbb{N}\), less or equal to the out-degree of \(x\).
- each edge \(x \rightarrow y\) of \(\mathcal{E}\) is labelled by \(\varepsilon \in \{+, -\}\) and by \(s \in [0 \cdots b_x]\).

**Variant:** bipartite graph

- complexation of two proteins
- inhibition of a regulation
- external conditions...
Regulatory Networks (R. Thomas)

Basal level: $K_x$

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

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

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

<table>
<thead>
<tr>
<th>$(x,y)$</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>$(K_x, \overline{y}, K_y)$</td>
</tr>
<tr>
<td>(0,1)</td>
<td>$(K_x, K_y)$</td>
</tr>
<tr>
<td>(1,0)</td>
<td>$(K_x, x\overline{y}, K_y)$</td>
</tr>
<tr>
<td>(1,1)</td>
<td>$(K_x, x, K_y)$</td>
</tr>
<tr>
<td>(2,0)</td>
<td>$(K_x, x, K_y, x)$</td>
</tr>
<tr>
<td>(2,1)</td>
<td>$(K_x, x, K_y, x)$</td>
</tr>
</tbody>
</table>
Resources in a Regulatory Network

States:

\[ \eta : \mathcal{V} \rightarrow \mathbb{N} \ (\approx \text{vector of integers}) \]

\[ \eta(x) = \text{abstract concentration level of } x \]

Variant: singular states (values can be the thresholds \( \tau_1, \tau_2, \ldots \))

Resources:

For each \( x \xrightarrow{+,s} y, \) \( x \) is a resource of \( y \) iff \( \eta(x) > s \)

For each \( x \xrightarrow{-,s} y, \) \( x \) is a resource of \( y \) iff \( \eta(x) \leq s \)

Parameters:

Partial function \( K : \mathcal{V} \times \mathcal{P}(\mathcal{V}) \rightarrow \mathbb{N} \)

Image:

Vector of the \( K(y, \omega) \) where \( \omega \) is the set of resources of \( y \)
State Graphs

<table>
<thead>
<tr>
<th>(x,y)</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>(K_{x,y}, K_y) = (2,1)</td>
</tr>
<tr>
<td>(0,1)</td>
<td>(K_x, K_y) = (0,1)</td>
</tr>
<tr>
<td>(1,0)</td>
<td>(K_{x,x}, K_y) = (2,1)</td>
</tr>
<tr>
<td>(1,1)</td>
<td>(K_{x,x} x, K_y) = (2,1)</td>
</tr>
<tr>
<td>(2,0)</td>
<td>(K_{x,x}, K_{y,x}) = (2,1)</td>
</tr>
<tr>
<td>(2,1)</td>
<td>(K_x, x, K_{y,x}) = (2,1)</td>
</tr>
</tbody>
</table>

“desynchronization” \rightleftharpoons by units of Manhattan distance
Time has a tree structure

From an initial state:
Parameters & thresholds: often unknown

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

\[ 3^4 \times 2^2 \quad \text{and} \quad 3^4 \times 2^2 \quad \text{and} \quad 2^4 \times 2^2 \]

712 possible models

Some criteria exist to reduce the number of models,

but formal logic is needed to go further automatically

*Note:* some models are observably equivalent
1. Modelling biological regulatory networks
2. Formal framework for biological regulatory networks
3. Temporal logic and Model Checking
4. Computer Aided elaboration Of Formal models
5. Example: mucus production in *Pseudomonas aeruginosa*
**CTL = Computation Tree Logic**

Atoms = comparisons: \((x=2) \quad (y>0) \quad \ldots\)

Logical connectives: \((\varphi_1 \land \varphi_2) \quad (\varphi_1 \implies \varphi_2) \quad \ldots\)

Temporal connectives: made of 2 characters

<table>
<thead>
<tr>
<th>first character</th>
<th>second character</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) = for (\text{All path choices})</td>
<td>(X) = ne(X)t state</td>
</tr>
<tr>
<td>(E) = there (\text{Exist a choice})</td>
<td>(F) = for some (\text{Future state})</td>
</tr>
<tr>
<td></td>
<td>(G) = for all future states ((G)lobally)</td>
</tr>
<tr>
<td></td>
<td>(U) = (\text{Until})</td>
</tr>
</tbody>
</table>

\(AX(y = 1)\): 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.
Temporal Connectives of CTL

next state:

$EX\varphi : \varphi$ can be satisfied in a next state
$AX\varphi : \varphi$ is always satisfied in the next states

eventually in the Future:

$EF\varphi : \varphi$ can be satisfied in the future
$AF\varphi : \varphi$ will be satisfied at some state in the future

Globally:

$EG\varphi : \varphi$ can be an invariant in the future
$AG\varphi : \varphi$ is necessarily an invariant in the future

Until:

$E[\psi U \varphi] :$ there exist a path where $\psi$ is satisfied until a state where $\varphi$ is satisfied
$A[\psi U \varphi] :$ $\psi$ is always satisfied until some state where $\varphi$ is satisfied
Semantics of Temporal Connectives

- $EX\varphi$
- $AX\varphi$
- $EF\varphi$
- $AF\varphi$
- $EG\varphi$
- $AG\varphi$
- $E[\psi U \varphi]$
- $A[\psi U \varphi]$
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:

$(x = 2) \implies AG(\neg(x = 0))$

and $(x = 0) \implies AG(\neg(x = 2))$

They express “the positive feedback loop is functional”

(satisfaction of these formulae relies on the parameters $K$...)

\[ (0,0) \quad (1,0) \quad (2,0) \]

\[ (0,1) \quad (1,1) \quad (2,1) \]

\[ (0,0) \quad (1,0) \quad (2,0) \]

\[ (0,1) \quad (1,1) \quad (2,1) \]
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 ($M \models \eta \varphi$)
- They can be tested against the biological experiments ($Biological\_Object \models_{\text{experiment}} \varphi$)

CTL formulae link theoretical models and biological objects together
Model Checking

Computes all the states of a theoretical model which satisfy a given formula: \( \{ \eta \mid M \models_{\eta} \varphi \} \).

Idea 1: work on the state graph instead of the path trees.

Idea 2: check first the atoms of \( \varphi \) and then check the connectives of \( \varphi \) with a bottom-up computation strategy.

Idea 3: (computational optimization) group some cases together using BDDs (Binary Decision Diagrams).

Example: \( (x = 0) \implies AG(\neg(x = 2)) \)

Obsession: travel the state graph as less as possible
\[(x = 0) \implies AG(\neg (x = 2))\]

\[x = 0 \quad x = 2\]

\[\neg (x = 2) \quad \text{and} \quad AG(\neg (x = 2))?\]

\ldots one should **travel all** the paths from any green box and check if successive boxes are green: *too many boxes to visit.*

**Trick:** \(AG(\neg (x = 2))\) is equivalent to \(\neg EF(x = 2)\)

start from the red boxes and follow the transitions backward.
1. Modelling biological regulatory networks
2. Formal framework for biological regulatory networks
3. Temporal logic and Model Checking
4. Computer aided elaboration of formal models
5. Example: mucus production in *Pseudomonas aeruginosa*
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:**

  $1 \xrightarrow{+} x \xrightarrow{-} y \xrightarrow{1} 1$

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, \ldots, \varphi_n\} \quad \text{and} \quad \mathcal{M} = \{\}

1. Is it possible that \( \Phi \) and \( \mathcal{M} \)?

**Consistency** of knowledge and hypotheses. Means to select models belonging to the schemas that satisfy \( \Phi \).

\( (\exists? \ M \in \mathcal{M} \mid M \models \varphi) \)

2. If so, is it true *in vivo* that \( \Phi \) 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 $\Phi$.

3. Automatically generate all the possible regulatory networks derived from $\mathcal{M}$ according to all possible parameters $K$... Our software platform SMBioNet handles this automatically.

4. Check each of these models against $\Phi$.
   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 $K$... have been indirectly established. Now Question 2 has to be addressed.
1. Among all possible formulae, some are “observable” i.e., they express a possible result of a possible biological experiment. Let $\text{Obs}$ be the set of all observable formulae.

2. Let $\Lambda$ be the set of theorems of $\Phi$ and $\mathcal{M}$. $\Lambda \cap \text{Obs}$ 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 mucus production of $P.\text{aeruginosa}$. 

**Question 2 = Validation**
1. Modelling biological regulatory networks
2. Formal framework for biological regulatory networks
3. Temporal logic and Model Checking
4. Computer aided elaboration of formal models
5. Example: mucus production in *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

The biological questions (Janine Guespin):
are **mucus production** and/or **cytotoxicity** in *Pseudomonas aeruginosa* due to an epigenetic switch?
[→ cystic fibrosis]
Mucus production in *P. aeruginosa*

*Pseudomonas aeruginosa*: (J.Guespin, M.Kaufman)

Epigenetic Hypothesis (without mutation) =
→ The positive feedback circuit is functional, with a mucoid stable state and another non mucoid stable state.
→ An external signal (in the cystic fibrosis’ lungs) could switch AlgU from its lower stable state to the higher one.
→ The mutation could be favored later because the inhibitor complex is toxic for the bacteria. ⇒ New possible therapy.
Cytotoxicity in *P. aeruginosa*

(Janine Guespin)

![Cytotoxicity Diagram]

**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:

\[(\text{ExsA} = 2) \implies AXAF(\text{ExsA} = 2)\]

\[(\text{ExsA} = 0) \implies AG(\neg(\text{ExsA} = 2))\]

Question 1, consistency: proved by Model Checking
→ 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: \((\text{ExsA} = 0) \implies AG(\neg(\text{ExsA} = 2))\)

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

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

Experimental limitation:

ExsA can be saturated but it cannot be measured.

Experiment:

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

\((\iff \text{to verify a hysteresis})\)

This experiment can be generated automatically
To test \((\text{ExsA}=2) \iff AXAF(\text{ExsA}=2)\)

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

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

\(\ldots\) formal proof by computer \(\ldots\)

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

Karl Popper:

to validate = to try to refute

thus A=false is useless

experiments must begin with a pulse

<table>
<thead>
<tr>
<th>A \implies B</th>
<th>true</th>
<th>false</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
</tbody>
</table>

The pulse forces the bacteria to reach the initial state ExsA = 2.
If the state were not directly controllable we had to prove lemmas:

(ExsA = 2) \iff (something reachable)

General form of a test:

(something reachable) \implies (something observable)
Concluding Slogans

- Behavioural properties ($\Phi$) 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 ($\text{Obs}$).
- 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.