

# Logique temporelle et Model Checking pour les réseaux de régulation biologique

Gilles Bernot & Jean-Paul Comet

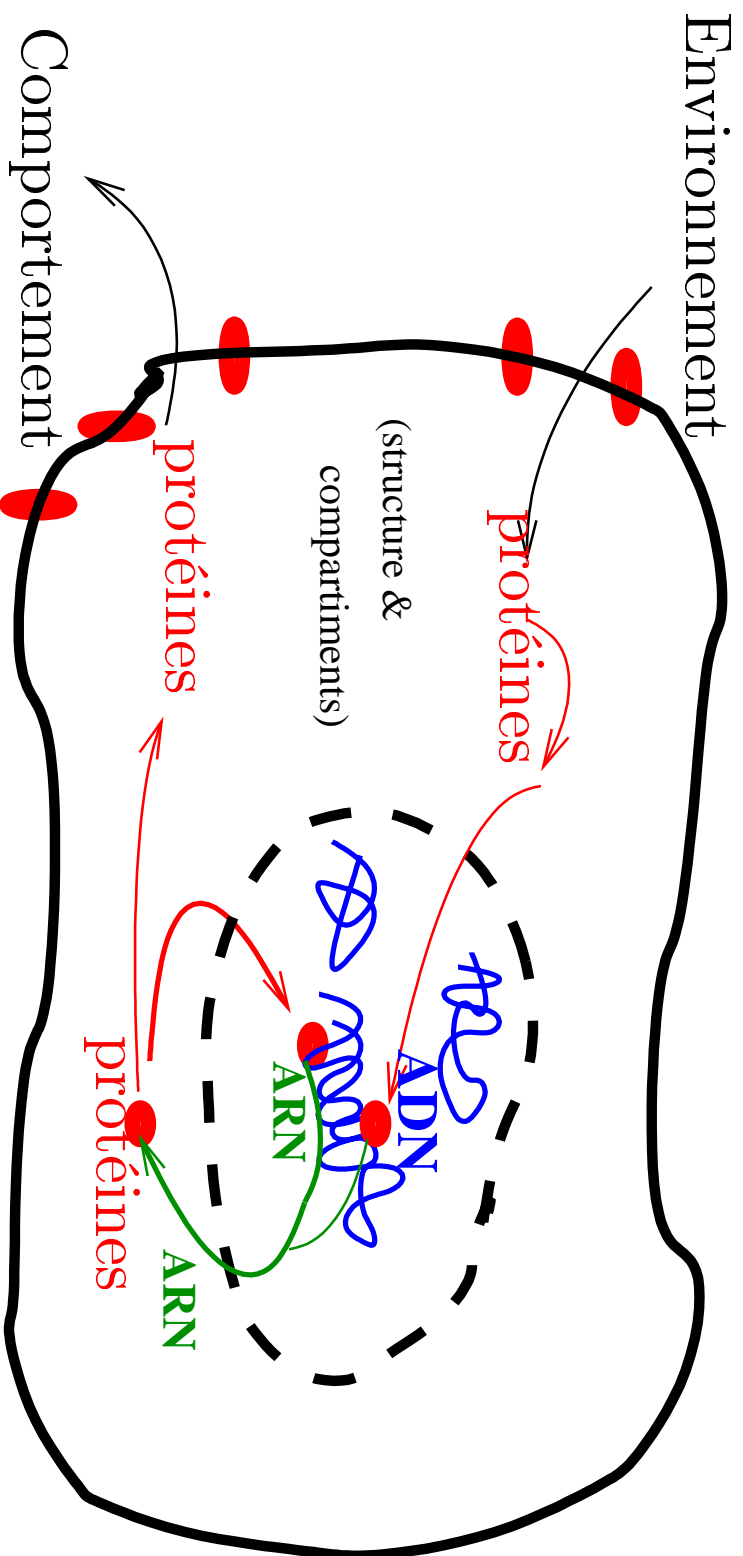
LAMI

(Laboratoire de Méthodes Informatiques)

C.N.R.S. / Université d'Évry UMR 8042 et **genopole**<sup>®</sup>



# Programme Épigénomique (Univ. Évry)



Modéliser, « reconstruire » le comportement *au dessus* du génome

# Postgénomique, Grande échelle, et autres slogans

Population de gènes & macromolécules

Fonctionnement & Combinatoire des interactions

Simulation & Comportements émergeants

*Modéliser pour comprendre*

Modèles & Affirmations comportementales

Connaissances parcellaires et expérimentations limitées

⇒ classes d'équivalence de modèles

*Predictivité*

Retour « à la pailleasse »

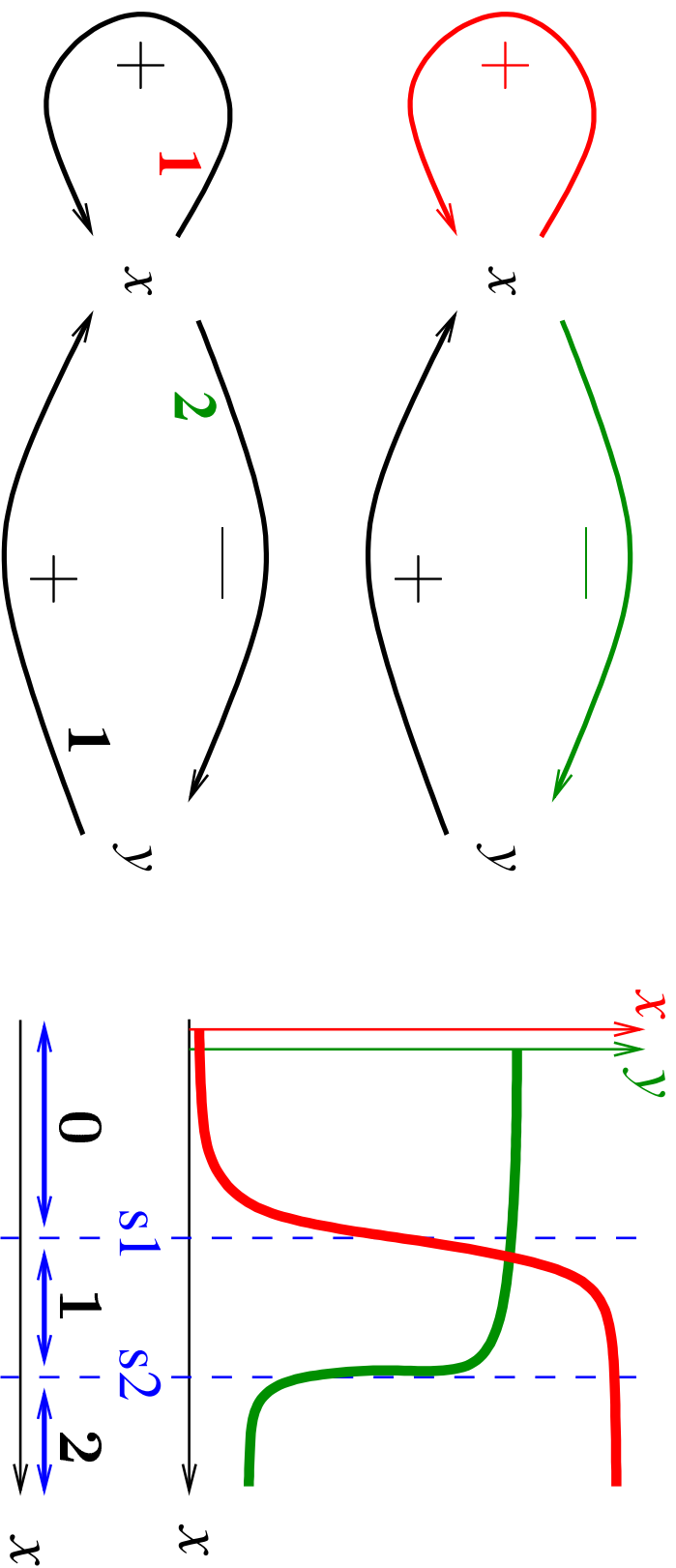
# Plan

*Ce travail apporte un cadre formel à la théorie des réseaux de régulation de René Thomas et intègre les outils informatiques correspondants.*

1. Réseaux de régulation biologique multivalués
2. Sémantique, automates & traces
3. Logique temporelle CTL
4. Formalisation de questions biologiques
5. La bactérie *Pseudomonas aeruginosa*
6. Mise au point de modèles par Model Checking et test

# Graphes de régulation *multivaluée*

Abstraire par une variable : un gène, l'ARN  
et la protéine pour laquelle il code



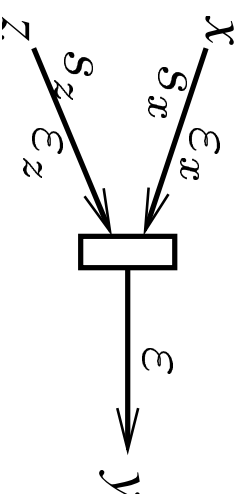
# Formalisation des graphes de régulation

Graphe à seuils : couple  $(\mathcal{N}, \mathcal{A})$

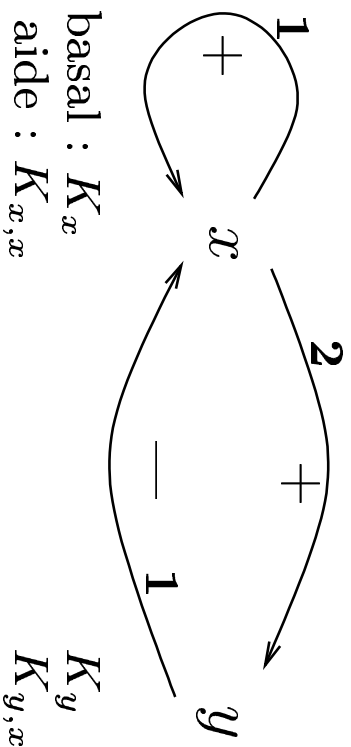
- Nœud = variable  $x$  munie d'une *borne*  $x \in \mathbb{N}$ , inférieure au degré sortant de  $x$ .
- Arc  $x \rightarrow y$  étiqueté par  $\varepsilon \in \{+, -\}$  et  $s \in [0 \dots \text{borne}_x]$ .

Variante : graphe bipartite

- formation de complexes protéiques
- inhibition d'une influence
- représentation d'un contexte



# Réseaux de régulation : paramètres



niveau basal :  $K_x$   
 $x$  aide :  $K_{x,x}$   
 $y$  absent aide :  $K_{x,\bar{y}}$   
 les deux :  $K_{x,x\bar{y}}$

$K_y$   
 $K_{y,x}$

$(x, y)$	<i>Attracteur</i>
(0,0)	$(K_{x,\bar{y}}, K_y)$
(0,1)	$(K_x, K_y)$
(1,0)	$(K_{x,x\bar{y}}, K_y)$
(1,1)	$(K_{x,x}, K_y)$
(2,0)	$(K_{x,x\bar{y}}, K_{y,x})$
(2,1)	$(K_{x,x}, K_{y,x})$

# Formalisation des réseaux de régulation

Marquage de niveaux de concentration :

$\eta : \mathcal{N} \rightarrow \mathbb{N}$  (= vecteur de valeurs entières, inférieures à la borne pour chaque nœud).  $\eta(x) = \text{niveau de concentration de } x$

**Variante** : état *singulier* (certaines valeurs sont les seuils  $s_1, s_2, \dots$ )

**Arc franchissable** :

niveau d'entrée supérieur (resp. inférieur) à  $s$  si  $\varepsilon = +$  (resp.  $\varepsilon = -$ )

**Paramètres de transitions** :

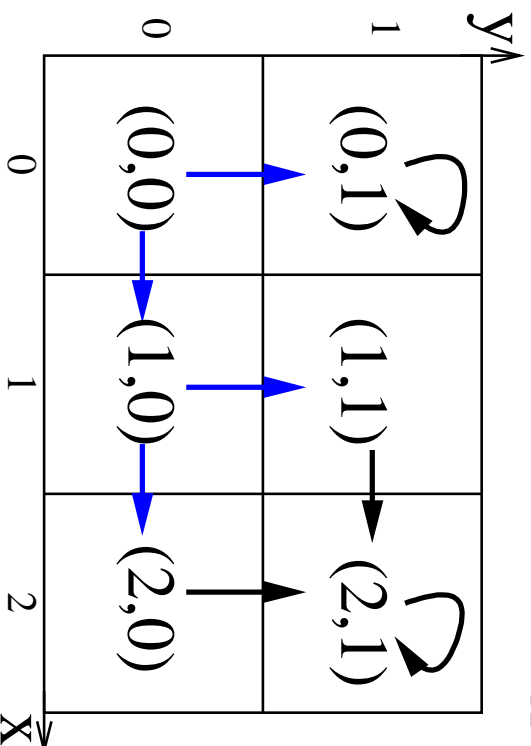
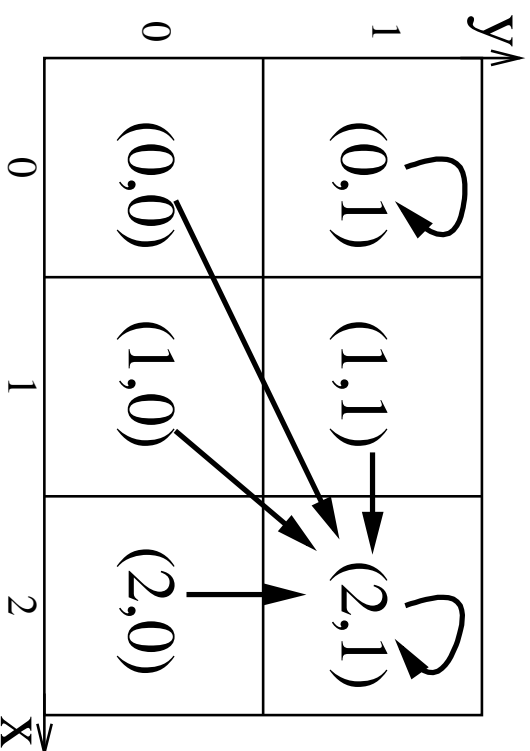
fonction partielle  $K : \mathcal{N} \times \mathcal{P}(\mathcal{N}) \rightarrow \mathbb{N}$

**Vecteur attracteur** :

état (régulier) « attracteur » donné par les  $K(x, \omega)$

# Réseau de régulation → Graphes d'états

$(x,y)$	<u>Attracteur</u>
(0,0)	$(K_{x,\bar{y}}, K_y) = (2,1)$
(0,1)	$(K_x, K_y) = (0,1)$
(1,0)	$(K_{x,x\bar{y}}, K_y) = (2,1)$
(1,1)	$(K_{x,x}, K_y) = (2,1)$
(2,0)	$(K_{x,x\bar{y}}, K_{y,x}) = (2,1)$
(2,1)	$(K_{x,x}, K_{y,x}) = (2,1)$



« désynchronisation » →  
par unités de distance Manhattan

Arbre des traces, etc.

# Computation Tree Logic (CTL)

**Atomes**=comparaisons ( $x=2$ ), ( $y>0$ )... + **Connecteurs** standard

**Prochain état** :

$EX \varphi$  :  $\varphi$  peut être vraie à l'étape d'après

$AX \varphi$  :  $\varphi$  toujours vraie à l'étape d'après

**Un jour dans le futur** :

$EF \varphi$  :  $\varphi$  peut être vraie plus tard

$AF \varphi$  :  $\varphi$  toujours vraie plus tard

**Modalités**

**Globalement** :

$EG \varphi$  :  $\varphi$  peut être un invariant à l'avenir

$AG \varphi$  :  $\varphi$  toujours un invariant à l'avenir

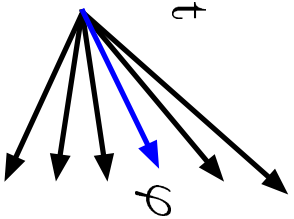
**Jusqu'à** :

$E[\psi U \varphi]$  il existe un chemin avec  $\psi$  vrai jusqu'à ce que  $\varphi$  le soit

$A[\psi U \varphi]$   $\psi$  reste toujours vrai jusqu'à ce que  $\varphi$  le soit

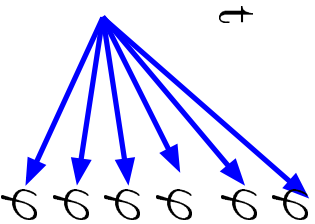
# Sémantique des modalités sur les traces

$t+1$



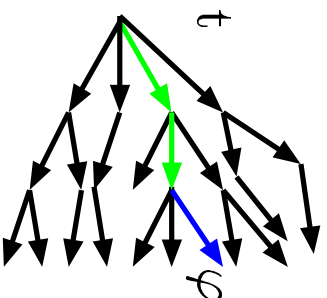
$EX\varphi$

$t+1$



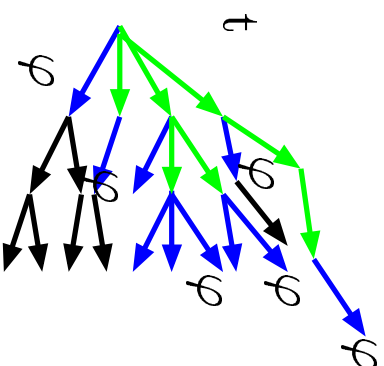
$AX\varphi$

$t+k$



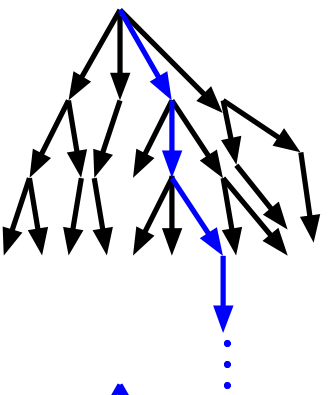
$EFF\varphi$

$t$



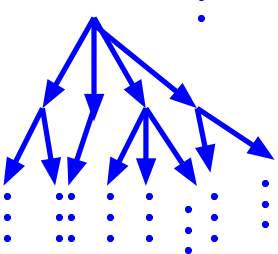
$AFF\varphi$

...



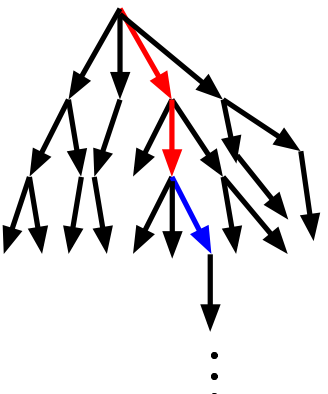
$EFG\varphi$

...



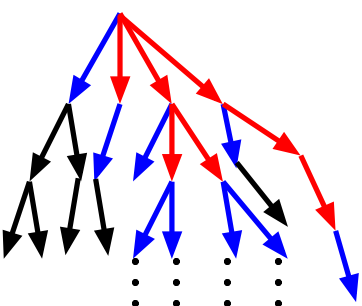
$AG\varphi$

...



$E[\psi U\varphi]$

...



$A[\psi U\varphi]$

## CTL et biologie

Les formules ainsi construites sont valides ou invalides par rapport à un ensemble de traces donné

Elles peuvent être confrontées à l'ensemble des traces possibles du modèle théorique

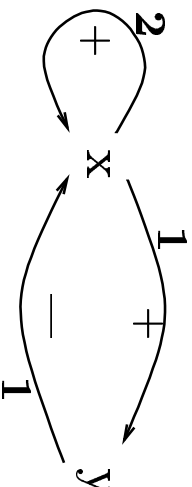
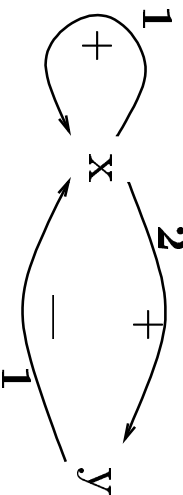
Elles peuvent être confrontées à l'ensembles des expériences connues

*Elles font donc le lien entre modèles et objets biologiques*

# Formalisation de questions biologiques

Des connaissances et/ou des hypothèses biologiques on extrait :

– des schémas de modèles :



•••

– des propriétés :

« Sans stimuli, si  $x$  est à son niveau basal, il  $y$  reste ».

soit :  $(x = 0) \implies AG(x = 0)$

Les deux questions à traiter :

1. **Cohérence** des connaissances et des hypothèses.  
→  $\exists$  ? modèles issus des schémas satisfaisant les propriétés.
2. **Compatibilité** d'un des modèles trouvés avec l'objet biologique.  
→ Proposer de nouveaux plans d'expériences pour le tester.

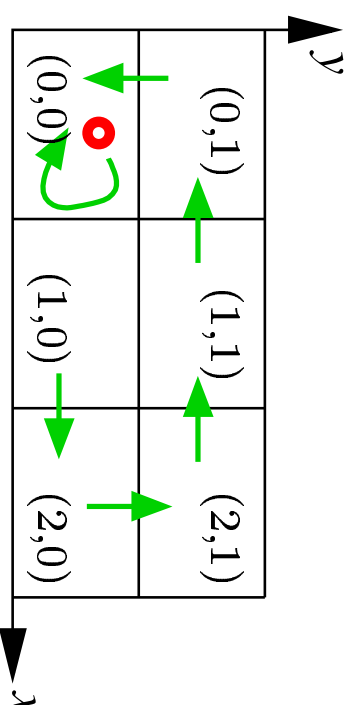
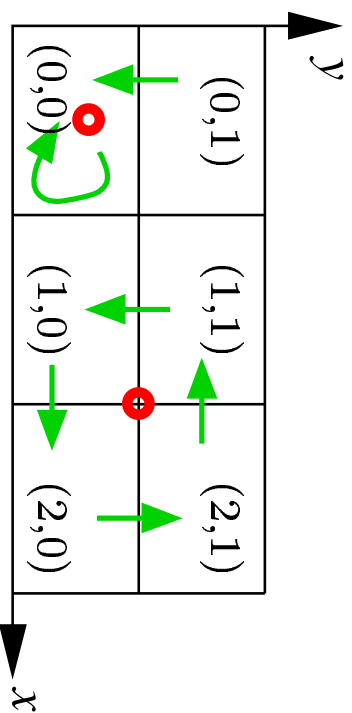
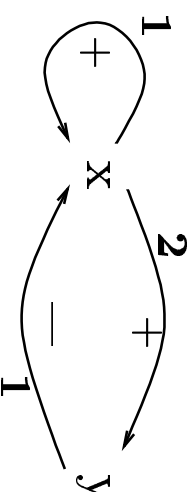
# Formalisation de propriétés biologiques

Propriétés souvent rencontrées :

« **fonctionnalité** » d'un sous-graphe

Rôle majeur des **boucles de rétroaction**

- positive : *bistationnarité* (nombre pair de —)
- négative : *homéostasie* (nombre impair de —)



Propriétés distinctives :

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

$$\text{et } (x = 0) \implies AG(\neg(x = 2))$$

traduisent *la boucle positive est fonctionnelle* (dépendent des  $K...$ )

## Un exemple biologique « simple »

Modification de phénotype, terminologie :

**modification génétique** : héritable et non réversible (mutation)

**modification épigénétique** : héritable mais réversible

**adaptation** : non héritable et réversible

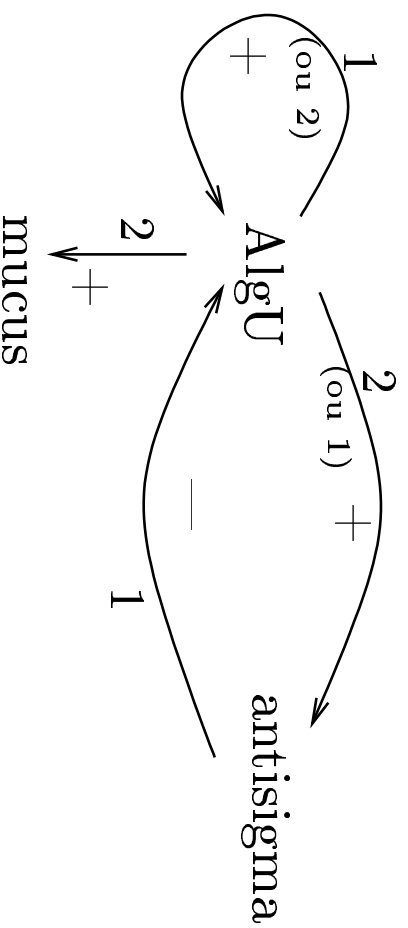
**La question biologique (Janine Guespin)** : la synthèse de mucus chez la bactérie *Pseudomonas aeruginosa* est-elle une modification génétique ou épigénétique ?

[→mucoviscidose]

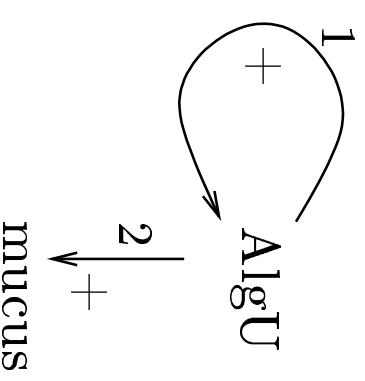
**Fait** : après stabilisation des populations productrices de mucus, beaucoup sont mutantes

# *Pseudomonas aeruginosa*

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



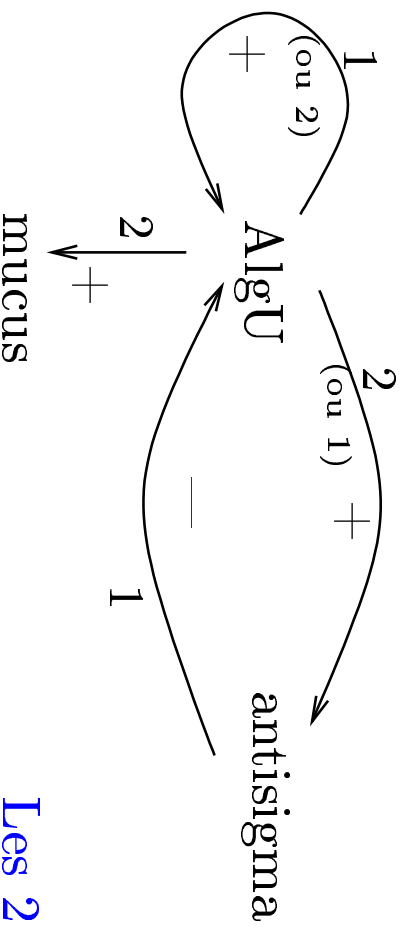
... et mutée :



Hypothèse épigénétique (i.e. sans mutation) =

- Le cycle positif est fonctionnel malgré le cycle négatif, avec un état non mucoïde et l'autre mucoïde.
- Un signal extérieur (produit par le poumon malade) pourrait éventuellement faire passer AlgU de l'état bas à l'état haut.
- La pression de sélection favorise *ultérieurement* les mutants en milieu muqueux.  $\implies$  Nouvelles perspectives de thérapie.

## Preuves sèches et preuves humides...



Les 2 états stables :

$$(AlgU = 2) \implies AX AF (AlgU = 2)$$

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

**Question 1, cohérence** : prouvée par Model Checking (8 modèles parmi 648, extraits par notre logiciel dédié SMBioNet)

**Question 2, compatibilité *in vivo*** : une seule expérience clef est nécessaire et suffisante (preuve similaire au test fonctionnel)

Elle est en préparation au CHU de Grenoble.

## Méthode par Model Checking et test

- Tracer tous les réseaux de régulation avec toutes les combinaisons de seuils crédibles
- Formaliser en CTL les propriétés comportementales connues et les hypothèses (boucles fonctionnelles, etc.)
- Engendrer automatiquement tous les graphes de régulation selon toutes les valeurs possibles des  $K$ ...
- Les **confronter** par Model-Checking, *retenir les modèles rescapés*
- S'il n'en reste aucun alors remettre en cause les hypothèses

S'il en reste passer à la question 2. : chercher les conséquences révélatrices qui peuvent donner lieu à des **expériences humides**

Actuellement la détermination des paramètres  $K$ ... est un goulot d'étranglement de la construction de modèles

## Conclusion

Les méthodes formelles informatiques pour la biologie, ça marche !

En cours :

- Environnements de manipulation de modèles assistée par ordinateur ([MeRoVinGien](#) et [SMBioNet](#))
- Extensions du pouvoir d'expression des réseaux de régulation
- Introduction de délais d'activation/inactivation d'un arc
- Nouveaux exemples biologiques
- Liens avec des notions d'observabilité / réfutabilité
- Préservation de propriétés par inclusions de réseaux
- etc.

À Évry : Programme Épigénomique, groupe G<sup>3</sup>, Ateliers genopole<sup>®</sup>