

# BIOCHAM

Une approche langage de la biologie des systèmes

Laurence Calzone   Sylvain Soliman

Travail commun avec François Fages et Nathalie Chabrier-Rivier

INRIA – Projet CONTRAINTES

Conférence OUEST-genopole – 8 Septembre 2005, Rennes

# Plan

- 1 Introduction
  - Biologie des Systèmes
  - Modélisation
  - Approche langage
- 2 BIOCHAM
  - Description
  - Langage pour les règles d'interaction
  - Langage pour les propriétés biologiques
- 3 Exemples
  - Abstraction booléenne
  - Abstraction de concentrations
- 4 Conclusion

# Biologie des Systèmes

H. Kitano, ICSB 2000

“Systems Biology aims at systems-level understanding [which] requires a set of principles and methodologies that links the behaviors of molecules to systems characteristics and functions.”

- **Analyser** une grande quantité de données ;
- **Intégrer** des données hétérogènes de manière consensuelle : résumer en un diagramme ou un modèle des informations sur un problème spécifique ;
- **Comprendre** des diagrammes complexes ;
- **Prédire** des comportements ou des interactions dans des réseaux (de protéines ou de gènes).



## Types de modèles

De nombreux outils ont été développés pour répondre à des **problèmes spécifiques** et utilisent des approches différentes :

- Statistique ;
- Discrète (booléen ou avec seuils) ;
- Continue, équations différentielles (ODEs ou PDEs) ;
- Stochastique ;
- Hybride ;
- ...

**Formalisation**  $\Rightarrow$  **raisonner** sur le comportement **global** du système étudié, **échanger et réutiliser** des modèles existants (SBML)

# Approche langage

## Modèles qualitatifs : de la notation *diagrammatique* à

- Réseaux booléens [Thomas 73]
- $\pi$ -calcul [Regev et al. 99-01, Nagasali et al. 00]
- Systèmes de Transition Concurrents BIOCHAM-1 [Chabrier-Fages 03]
- Pathway logic [Eker et al. 02]
- Bio-ambients [Regev et al. 03]

## Modèles quantitatifs : des systèmes d'*équations différentielles* à

- Réseaux de Petri Hybrides [Hofstadt-Thelen 98, Matsuno et al. 00]
- Automates Hybrides [Alur et al. 01, Ghosh-Tomlin 01]
- Hybrid CC [Bockmayr-Courtois 01]
- BIOCHAM-2 [Chabrier-Fages-Soliman 04]

## “Que peut apporter un nouvel outil de modélisation ?”

### la formalisation de règles d'interaction biochimiques

- selon le niveau de connaissance disponible (qualitatif, quantitatif) ;
- avec un recours minimal à une connaissance mathématique ou informatique.

### la formalisation de propriétés expérimentales

- conditions d'oscillation ;
- conditions de stabilité ;
- checkpoints ;

⇒ Adjonction au modèle

# The BIOChemical Abstract Machine

## “Qu'est-ce que BIOCHAM ?”

- Un environnement de modélisation ;
- Un langage formel pour décrire les interactions (moléculaires) et les résultats expérimentaux connus sur le système biologique étudié ;
- Un outil d'analyse à 3 niveaux d'abstraction : présence/absence de molécules, concentrations, populations.

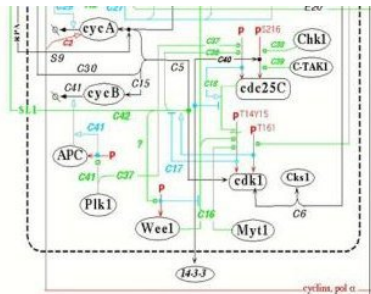
## “Que modéliser ?”

- pour l'instant, les réseaux protéiques et géniques et leur dynamique au niveau intracellulaire ;
- dans le futur, populations de cellules, tissus, ...

## Exemple : Le Cycle Cellulaire de Kohn

Extrait d'une transcription du diagramme du cycle cellulaire des mammifères de Kurt Kohn [MBC 1999].

Environ 800 règles pour 500 variables.



$cdk1\{p1,p2,p3\} + cycA \Rightarrow cdk1\{p1,p2,p3\}-cycA.$

$cdk1\{p1,p2,p3\} + cycB \Rightarrow cdk1\{p1,p2,p3\}-cycB.$

...

$cdk1\{p1,p3\}-cycA = [ Wee1 ] \Rightarrow cdk1\{p1,p2,p3\}-cycA.$

$cdk1\{p1,p3\}-cycB = [ Wee1 ] \Rightarrow cdk1\{p1,p2,p3\}-cycB.$

$cdk1\{p2,p3\}-cycA = [ Myt1 ] \Rightarrow cdk1\{p1,p2,p3\}-cycA.$

$cdk1\{p2,p3\}-cycB = [ Myt1 ] \Rightarrow cdk1\{p1,p2,p3\}-cycB.$

...

$cdk1\{p1,p2,p3\} = [ cdc25C\{p1,p2\} ] \Rightarrow cdk1\{p1,p3\}.$

$cdk1\{p1,p2,p3\}-cycA = [ cdc25C\{p1,p2\} ] \Rightarrow cdk1\{p1,p3\}-cycA.$

$cdk1\{p1,p2,p3\}-cycB = [ cdc25C\{p1,p2\} ] \Rightarrow cdk1\{p1,p3\}-cycB.$

...

$\_ = [ E2F13-DP12-gE2 ] \Rightarrow cycA.$

...

$cycB = [ APC\{p1\} ] \Rightarrow \_.$

...

# Expressions Cinétiques et Stochastiques

Quand l'information est disponible, on peut associer à chaque règle une expression, interprétée comme une vitesse ou une probabilité de réaction.

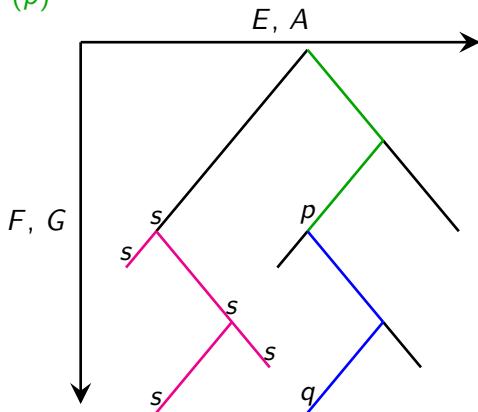
```
present(MPF~{p}, 1).
absent({MPF, cdc25C~{p1,p2}, Wee1, Wee1~{p}, APC}).
...
parameter(k2u, 3).
parameter(k1cc, 0.25).
...
k1cc                for _ => MPF~{p}.
k2u*[APC]*[MPF]     for MPF =[ APC ]=> _ .
k2u*[APC]*[MPF~{p}] for MPF~{p} =[ APC ]=> _ .

k3cc*[cdc25C~{p1,p2}]*[MPF~{p}]   for MPF~{p} =[ cdc25C~{p1,p2} ]=> MPF .
k4cc*[Wee1]*[MPF]                 for MPF =[ Wee1 ]=> MPF~{p} .
...
```

Au niveau booléen (présence/absence de molécules), ces expressions seront ignorées.

# Formalisation des propriétés biologiques du système

- d'activation, d'accessibilité  $EF(p)$
- sur les "chemins"
  - produit intermédiaire  
 $EF(p \wedge EF(q))$
  - checkpoint  
 $\neg E(\neg p U q)$
- de stabilité
  - état stationnaire  
 $s \Rightarrow EG(s)$
  - comportement cyclique  
 $EG((t \Rightarrow EF(\neg t)) \wedge (\neg t \Rightarrow EF(t)))$





# Model-Checking avec NuSMV

```
biocham: check_reachable(cdk46~{p1,p2}-cycD~{p1}).  
Ei(EF(cdk46~{p1,p2}-cycD~{p1})) is true  
  
biocham: check_checkpoint(cdc25C~{p1,p2}, cdk1~{p1,p3}-cycB).  
Ai(! (E(! (cdc25C~{p1,p2}) U cdk1~{p1,p3}-cycB))) is true  
  
biocham: nusmv(Ai(AG(! (cdk1~{p1,p2,p3}-cycB) -> checkpoint(Wee1, cdk1~{p1,p2,p3}-cycB))))).  
Ai(AG(! (cdk1~{p1,p2,p3}-cycB)->! (E(! (Wee1) U cdk1~{p1,p2,p3}-cycB)))) is false  
  
biocham: why.  
-- Loop starts here  
  cycB-cdk1~{p1,p2,p3} is present  
  cdk7 is present  
  cycH is present  
  cdk1 is present  
  Myt1 is present  
  cdc25C~{p1,p2} is present  
  
rule_114   cycB-cdk1~{p1,p2,p3}=[cdc25C~{p1,p2}]=>cycB-cdk1~{p2,p3}.  
  cycB-cdk1~{p2,p3} is present  
  cycB-cdk1~{p1,p2,p3} is absent  
  
rule_74   cycB-cdk1~{p2,p3}=[Myt1]>cycB-cdk1~{p1,p2,p3}.  
  cycB-cdk1~{p2,p3} is absent  
  cycB-cdk1~{p1,p2,p3} is present
```

Au plus quelques dizaines de secondes...

# LTL avec contraintes numériques et PCTL

LTL est une restriction de CTL aux opérateurs  $F$  et  $G$ .

$$F([MPF] > 0.2 \ \& \ F(d([APC])/dt < 0))$$
$$period(MPF, 22)$$

LTL et PCTL (Probabilistic CTL) permettent de formaliser des propriétés de simulations stochastiques.

$$P(EF(cdk46 \sim \{p1, p2\} - cycD \sim \{p1\})) > 0.5$$

Vérification (LTL et PCTL) par model-checking (traces ou ensemble de simulations).

## Processus de modélisation

La formalisation dans BIOCHAM permet :

- la vérification/compréhension de diagrammes d'interaction (ex. Kohn);
- l'aide à la construction de modèle.

Formalisation(propriétés+modèle)  $\Rightarrow$   
Automatisation(vérification+correction)

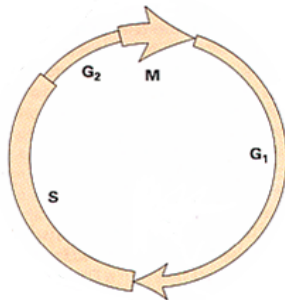
Questions :

- Le modèle rend-il compte de ce que l'on veut expliquer ?
- Le modèle est-il cohérent ?

# Le Cycle Cellulaire

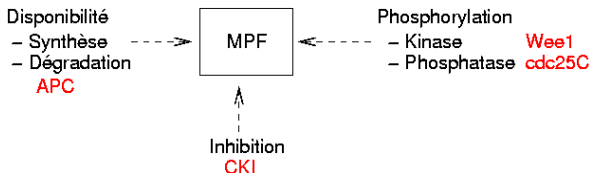
**Def. :** Le cycle cellulaire est une succession d'événements pendant lesquels la cellule grossit et se divise en deux cellules filles, chacune contenant l'information nécessaire pour répéter le processus.

G1	Gap entre M et S
S	Synthèse ADN
G2	Gap entre S et M
M	Mitose



# Un modèle générique du Cycle Cellulaire

- Un complexe au centre du cycle cellulaire : **MPF** (cycB-cdk1)
- Différentes manières de réguler l'activité de ce complexe



	G1	S	G2	M
CKI				MPF
Wee1				cdc25C
				APC

⇒ Deux phases considérées : G1-S-G2 et M.



# Propriétés biologiques du système

## Ajout d'une spécification CTL :

```
add_specs({
  Ei(reachable(MPF~{p})),
  Ei(reachable(MPF)),
  Ei(reachable(cdc25C)),
  Ei(reachable(cdc25C~{p1,p2})),
  Ei(reachable(Wee1)),
  Ei(reachable(APC)),
  ...
  Ai(loop(MPF~{p},MPF)),
  Ai(oscil(cdc25C)),
  Ai(oscil(cdc25C~{p1,p2})),
  Ai(loop(Wee1,Wee1~{p})),
  Ai(oscil(CKI)),
  Ai(oscil(APC)),
  ...
  Ai(EG((!(MPF))>checkpoint(cdc25C~{p1,p2},MPF)))
}).
```

# Recherche automatique de règles d'interaction

Supposons l'oubli de la règle d'activation de MPF par la forme active de cdc25C.

Vérification de la spécification :

```
biocham: check_all.
```

The specification is not satisfied.

This formula is the first not verified:  $E_i(EF(MPF))$

Suggestion de corrections :

```
biocham: learn_one_rule(elementary_interaction_rules).
```

Rules tested: 4218

Good rules to be added: 3

```
_=[cdc25C~{p1,p2}]=>MPF
```

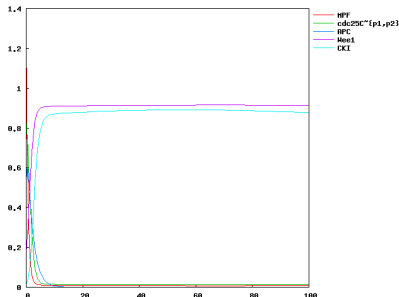
```
MPF~{p}=[cdc25C~{p1,p2}]=>MPF
```

```
CKI+MPF~{p}=[cdc25C~{p1,p2}]=>CKI-MPF
```

# Recherche de valeurs de paramètres

Donnons une cinétique à la règle trouvée :

```
k3cc*[MPF~{p}]*[cdc25C~{p1,p2}] for MPF~{p}=[cdc25C~{p1,p2}]=>MPF.  
parameter(k3cc,0.1).
```

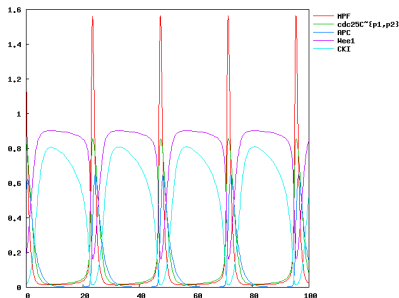


Le comportement du modèle ne correspond toujours pas à ce qui est observé : le “cycle cellulaire” ne montre pas d’oscillations. . .

# Recherche de valeurs de paramètres

## Recherche des valeurs satisfaisantes :

```
biocham: trace_get([k3cc], [(0,5)], 20, oscil(MPF,4) & F([MPF]<0.05), 100).
```



Found parameters that make `oscil(MPF,4) & F([MPF]<0.05)` true:  
`parameter(k3cc, 2.5)`.

La première valeur proposée pour `k3cc` est 2,5.

# Conclusion

- Un langage simple de **description des processus biologiques**
  - sémantique de population de molécules
  - sémantique de concentrations (ODE)
  - sémantique booléenne : présence/absence de molécules
- Un langage de **description des propriétés biologiques** du système
  - logique temporelle CTL (NuSMV)
  - logique temporelle LTL avec contraintes numériques (Prolog)
- Un système original d'aide à la modélisation
  - **recherche de règles d'interaction** (à partir d'une spec. CTL)
  - **estimation de paramètres** (à partir d'une spec. LTL)

# Perspectives

- Langage de modélisation du système
  - Lieux ;
  - Liaison avec Gene Ontology pour les noms (alias) ;
  - Notion de fonction.
- Langage de formalisation des propriétés
  - Théorie de la bifurcation ;
  - Aide à la vérification de modèles (SBML) avant publication.
- Modélisation
  - Composition/réutilisation de modèles ;
  - Modèles multi-cellulaires ;
  - Réseaux d'influence/d'interaction.

# BIOCHAM

STREP APRIL 2  
 NoE REVERSE

