

Extensions AltaRica pour la modélisation en biologie intégrative

Application de méthodes formelles pour l'étude de processus
biologiques

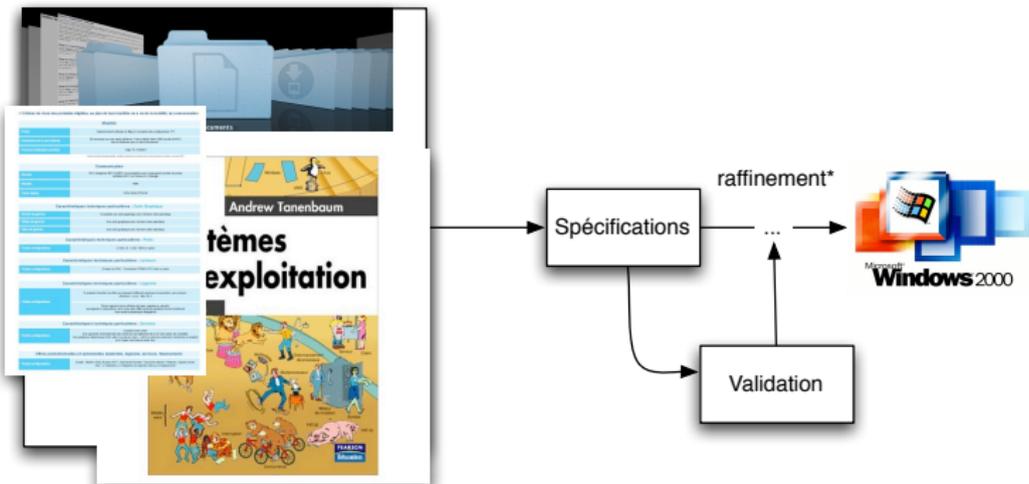
H. Soueidan M. Nikolski D.J. Sherman G. Sutre

LaBRI
Université de Bordeaux 1

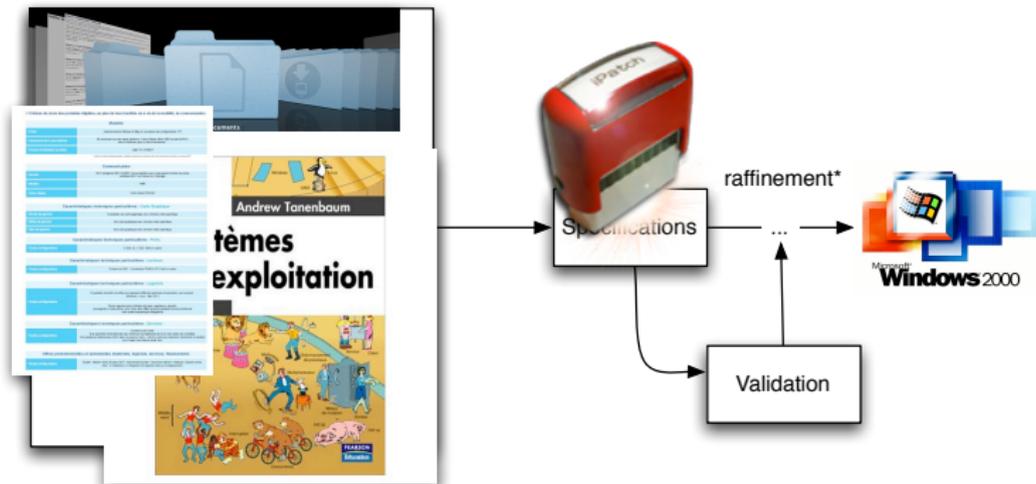
Novembre 2007



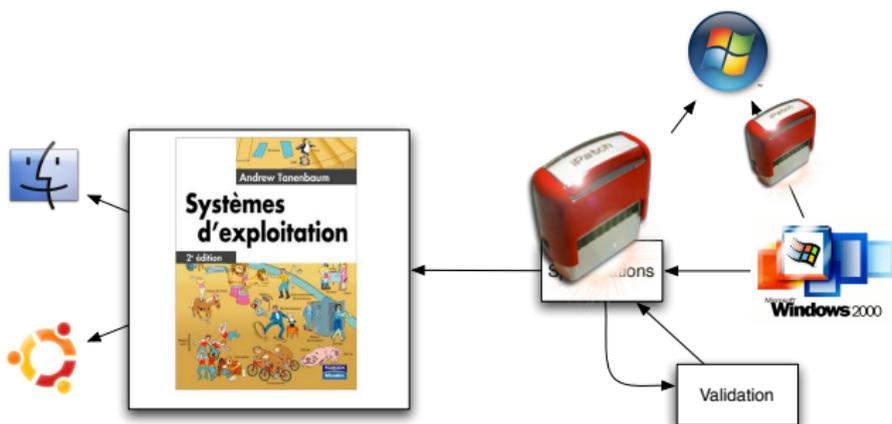
Approche ingénierie d'un système d'exploitation



Approche ingénierie d'un système d'exploitation



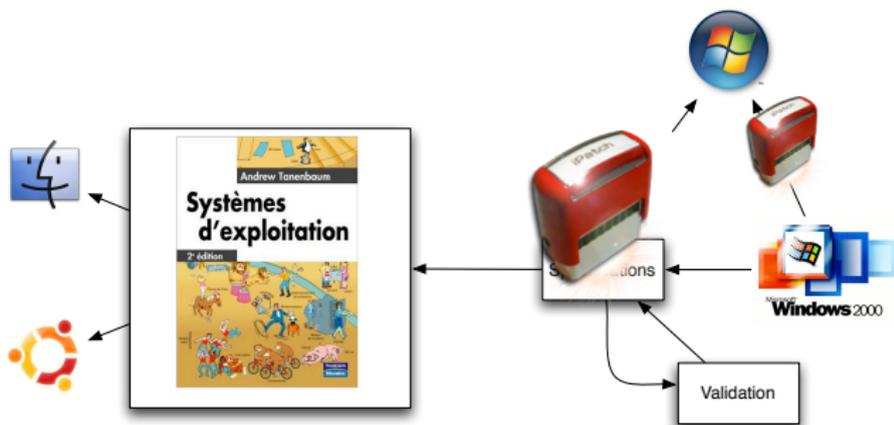
Rétro-ingénierie



- A partir du code binaire! Objectifs :
 - Apporter une modification mineure à un système d'exploitation, s'assurer que la modification est correcte
 - Expliquer le fonctionnement du système
 - Concevoir/Comprendre de nouveaux systèmes reprenant des fonctionnalités de systèmes connus



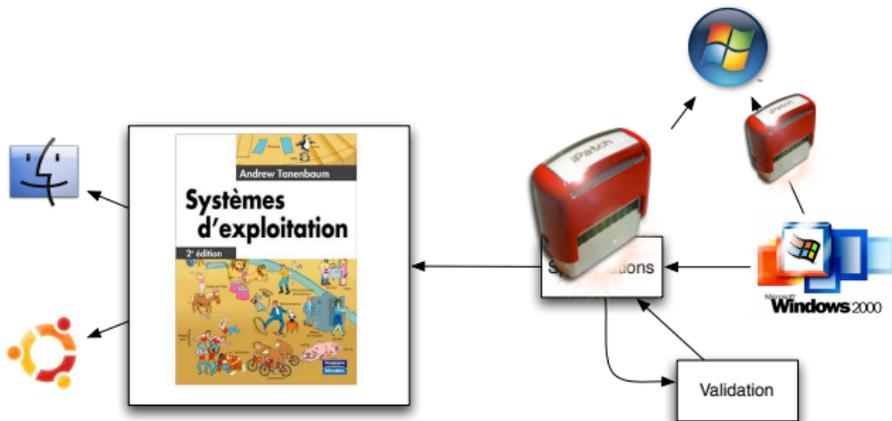
Rétro-ingénierie



- Désassemblage → liste de fonctions
- Recherche expérimentale des spécifications
- Validation de la modification vis-à-vis des spécifications
- Il suffit de modifier la bonne fonction !



Rétro-ingénierie

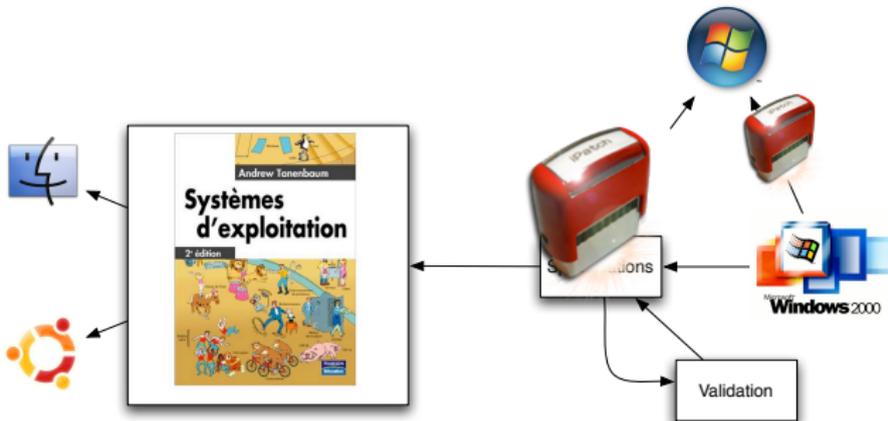


Mais :

- Massivement parallèle
- Auto correcteur
- Robuste aux modifications
- Modes dégradés
- Fonctions récursives, co-récursives
- S'adaptant à son environnement



Rétro-ingénierie

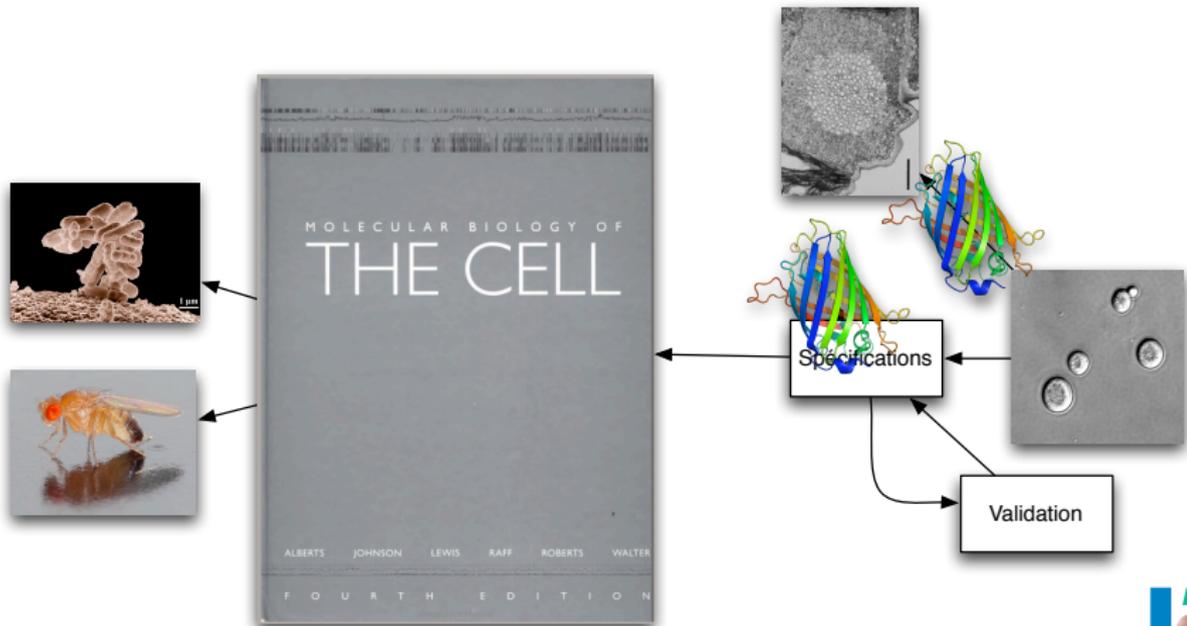


Mais :

- Obtenir une trace d'appel des fonctions peut coûter 10 k€
- Modifier un système existant peut coûter 100 k€ et prendre 6 mois



Rétro-ingénierie en Biologie



Objectifs de la biologie intégrative

- Pharmacologie, traitements
 - Minimiser les effets secondaires
 - Aide a la recherche de molécules cibles
- GénoLevures, YSNB
 - Recherche de nouvelles molécules, modifiant le comportement de levures
 - Levures pathogènes (*Candida Glabrata*, *Candida Albicans*)
 - Levure du boulanger, impact agro-alimentaire, optimisation de la fermentation
 - Ingénierie de nouveaux organismes pour : Production industrielle d'enzymes, de biopolymères, médicaments

(*Yeast systems biology to unravel the network of life*, R.Mustachi et al., 2006)

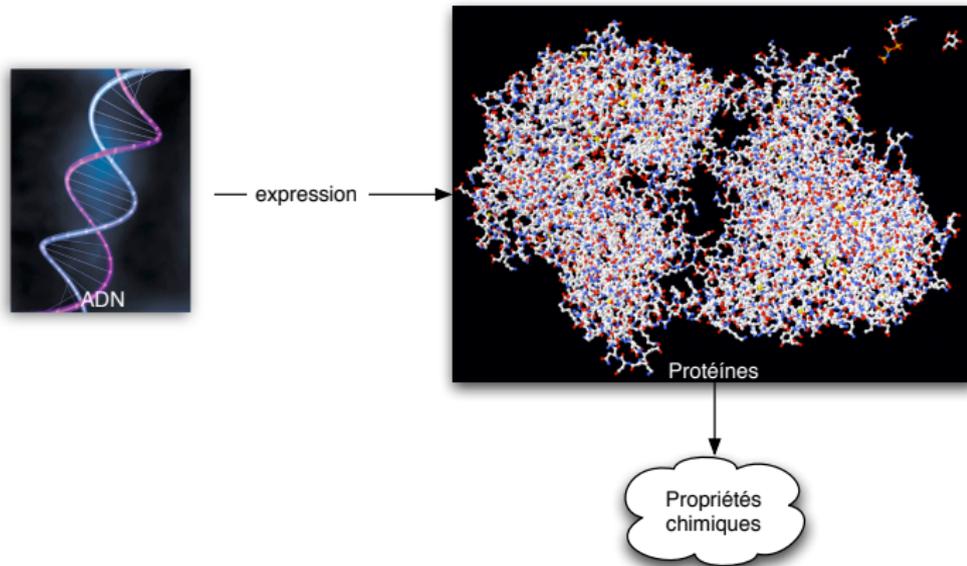


Plan

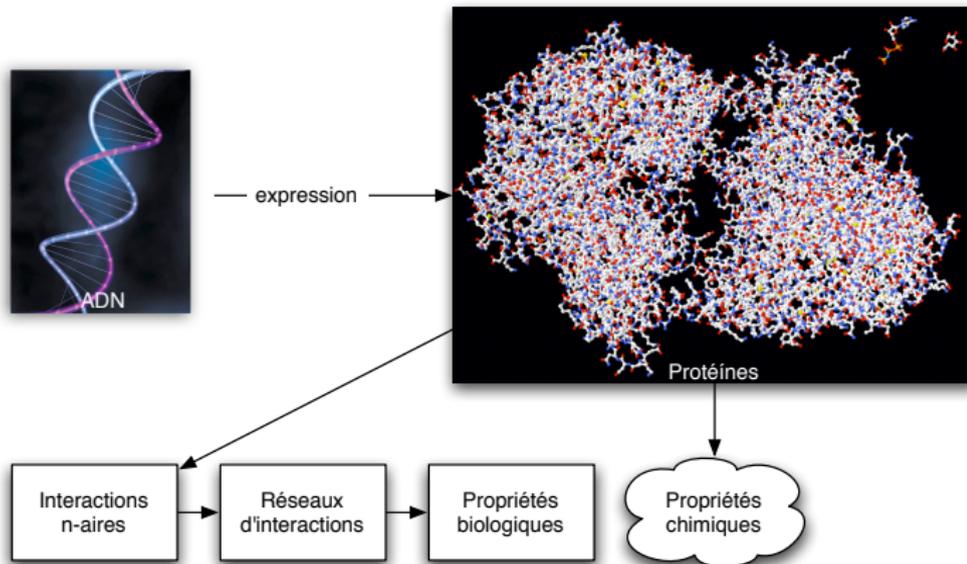
- 1 Biologie intégrative, Méthodes formelles et validation
- 2 Extension stochastique d'AltaRica dataflow
- 3 Spécification et vérification de modèles dynamiques
- 4 Application au vieillissement cellulaire chez la levure
- 5 Conclusion



Dogme central en biologie, approche "Réductionniste"

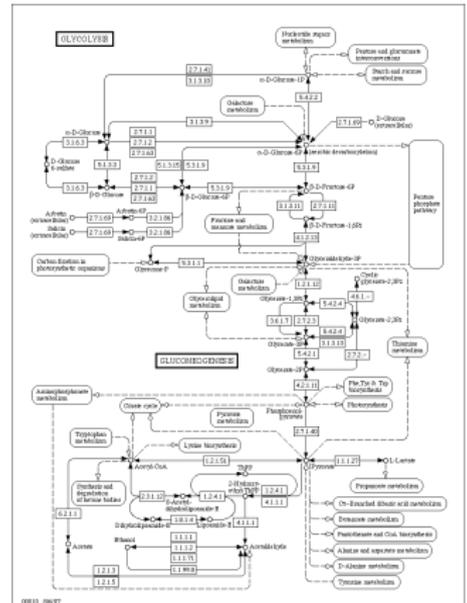


Dogme en biologie intégrative



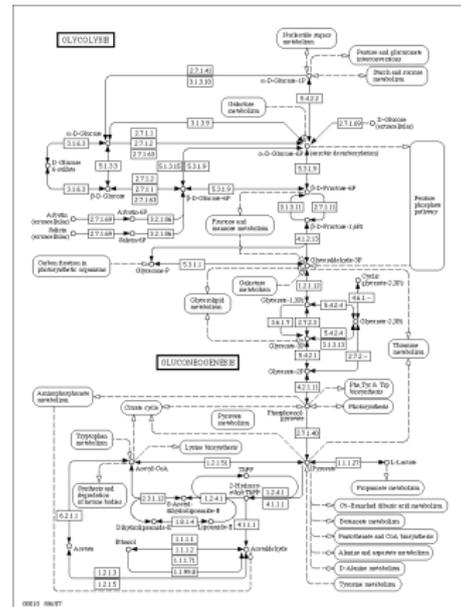
État actuel des connaissances biologiques

- Lois physico-chimique d'interactions bien connues
- Séquençage et annotation des génomes de plusieurs espèces
- Données d'expression accessibles
- Beaucoup d'interactions potentielles caractérisées (KEGG)



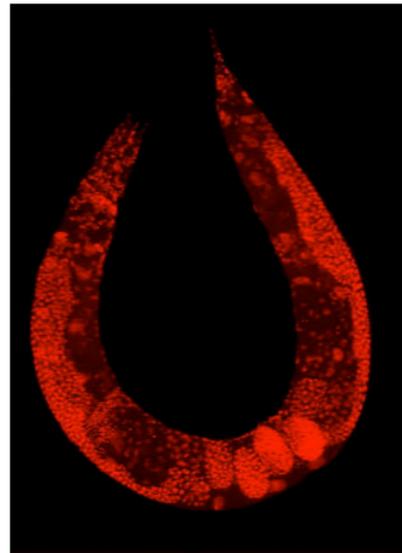
État actuel des connaissances biologiques

- Lois physico-chimique d'interactions bien connues
- Séquençage et annotation des génomes de plusieurs espèces
- Données d'expression accessibles **mais coûteuses et sensibles aux conditions expérimentales**
- Beaucoup d'interactions potentielles caractérisées (KEGG)
- **Peu de connaissances permettant de relier interactions possibles et observations**



Démarche expérimentale hypothético-déductive

- Caractérisation d'un phénotype "sauvage"
- Perturbation des conditions naturelles (mutation de gènes, environnement dégradé) et observation des phénotypes résultants, dits "mutants"
- Hypothèse explicative des phénotypes mutants (e.g. absence d'un gène, modification séquentielle de l'environnement)



Caenorhabditis elegans
(Wikimedia Commons)



Apport des méthodes formelles

Permettent :

- Formalisation des observations (Formules logiques, automates), constitution de bases de données structurées et non ambiguës des observations
- Détection d'incohérence
- Prédiction d'expérimentations non effectuées
- Explication de résultats "surprenants" (Recherche des pré-conditions / scénarios menant à telle observation in-vitro)

Formalismes : Réseaux de Petri, diagrammes de séquences, LTL, CTL, automates finis

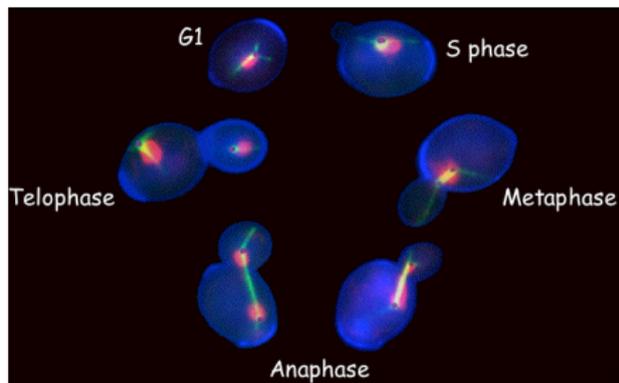
(Formal Modeling of C.elegans Development : A Scenario-Based Approach, Pnueli et al., 2003)



Exemple de modélisation biologique, cycle cellulaire

Propriétés émergentes observées in-vitro pour le phénotype sauvage

- Oscillations
- Irréversibilité des transitions
- Taille critique atteinte avant la division



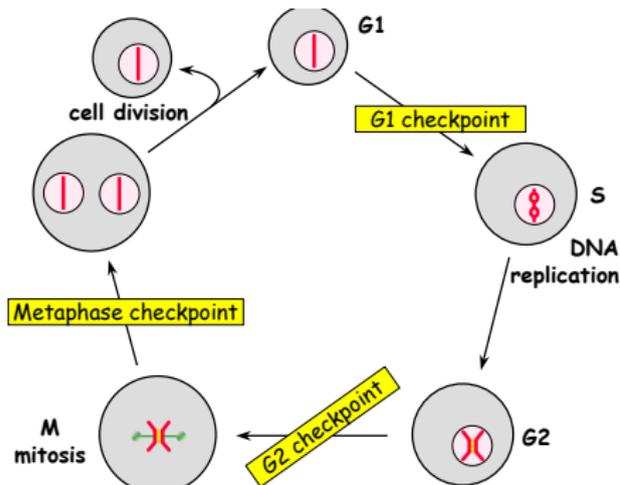
(*Networks Dynamics and Cell Physiology, Lessons from Modeling the Yeast Cell Cycle, J.J. Tyson, 2007*)



Exemple de modélisation biologique, cycle cellulaire

Propriétés émergentes observées in-vitro pour le phénotype sauvage

- Oscillations
- Irréversibilité des transitions
- Taille critique atteinte avant la division

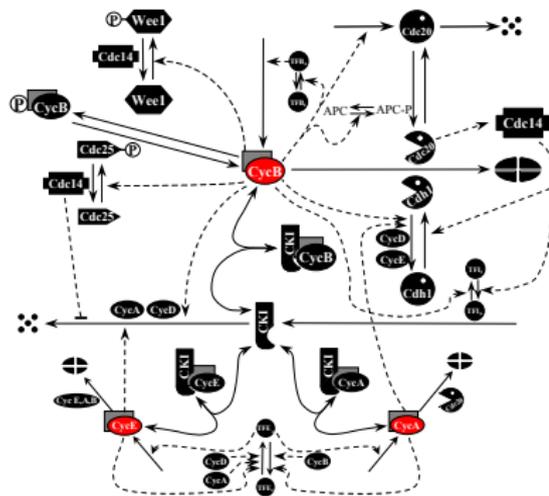


(*Networks Dynamics and Cell Physiology, Lessons from Modeling the Yeast Cell Cycle, J.J. Tyson, 2007*)

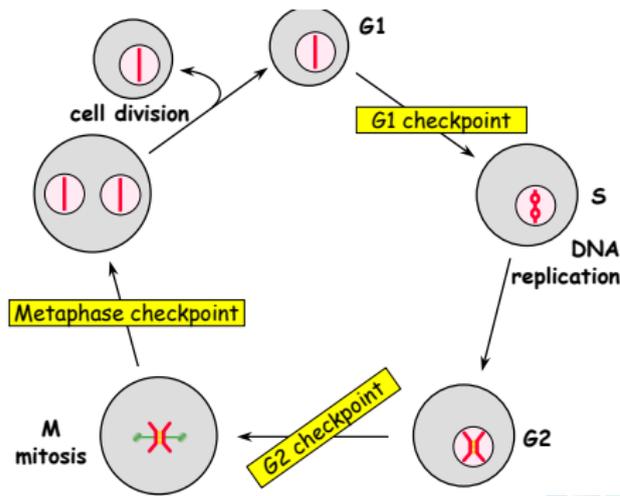


Exemple de modélisation biologique, cycle cellulaire

Modèle d'interaction "explicatif", interactions protéine-protéine



(Networks Dynamics and Cell Physiology, Lessons from Modeling the Yeast Cell Cycle, J.J. Tyson, 2007)



Exemple de modélisation biologique, cycle cellulaire

Validation du modèle explicatif :

- Méthode la plus répandue : Transformation du réseau d'interaction en modèle différentiel plat
- Analyse et simulation du modèle complet, phénotype sauvage retrouvée ?
- Prise en compte de phénotypes mutants :
 - Caractérisation des gènes mutés, dégradation du modèle d'interaction (e.g. Inhibition d'une protéine)
 - Analyse et simulation des modèles dégradés, phénotypes mutants retrouvés ?
- Pour chaque nouveau phénotype observé, modèle d'interaction raffiné, re-validation manuelle des observations précédentes
- Pour J.J. Tyson, plus de 130 phénotypes mutants observés et pris en compte par ce modèle.

(*Networks Dynamics and Cell Physiology, Lessons from Modeling the Yeast Cell Cycle*, J.J. Tyson) ▶ ◀ ☰ ☱ ☲ ☳ ☴ ☵ ☶ ☷



Apport des méthodes formelles

Permettent :

- Recherche exhaustive de combinaisons d'inhibitions renforçant un comportement donné (Klamt,2006)
- Description formelle des transitions discrètes (division cellulaire)
- Validation de propriétés temporelles
- Analyse automatisée des propriétés de systèmes différentiels (Analyse de bifurcations, calculs symboliques de prédécesseurs)

Formalismes : Description booléenne de réseaux d'interactions, Automates temporisés, Automates hybrides



Non déterminisme dans les modèles biologiques

Les modèles différentiels sont complètement déterministes, mais :

- Sémantique différentielle des réseaux d'interactions basée sur des cinétiques
 - Invalide pour des protéines présentes en quantité infinitésimale (Facteurs de transcription)
 - Suppose que certains réactifs sont en surnombre (cinétique enzymatique)
 - Forte dépendance à des paramètres cinétiques non mesurables in-vitro
- Équilibres dépendant des conditions initiales (phénomènes de bistabilité)
- Certains processus de bas-niveau sont plus fidèlement décrits par des automates discrets (Choix lyse-lysogénie chez le *phage lambda*)
- Processus d'expression de gènes intrinsèquement aléatoire

Stochastic gene expression in a single cell, Elowitz et al., 2002



Apport des méthodes formelles

Permettent :

- Analyse symbolique des comportements dans des intervalles de paramètres et de conditions initiales pour des systèmes linéaires
- Mécanismes aléatoires décrits par des processus stochastiques, comparaison des distributions à l'équilibre avec distributions in-vitro
- Application de modèles de concurrence pour la sémantique et simulation correcte de formalismes composés



Défis

Exprimés par la communauté
biologique :

- Hiérarchisation
- Découpage modulaire
- Interchangeabilité de modules
- Raisonnement compositionnel
- Modèle complet d'un organisme,
vérifiant toutes les propriétés
connues

*(Challenges for modeling and simulation methods in systems biology, D.Harel et al., 2007
Bounded Asynchrony, Fischer et al., 2006)*

En méthodes formelles :

- Modélisation de plus haut niveau,
plus proche du biologiste
- Logiques et langages de données
(Quantification sur les variables)
- Vérification de propriétés de
modèles basés ODEs
- Formalisme composant formalismes
existants
- Conception de la concurrence pour
systèmes multi-échelles



Plan

- 1 Biologie intégrative, Méthodes formelles et validation
- 2 Extension stochastique d'AltaRica dataflow**
- 3 Spécification et vérification de modèles dynamiques
- 4 Application au vieillissement cellulaire chez la levure
- 5 Conclusion



Objectifs

- Modèle formel pour les systèmes biologiques
 - Stochastique
 - Compositionnel
 - Hiérarchique
 - Hybride
- Admettant une simulation efficace et une compilation vers des formalismes décidables
- Permettant de composer large gamme de modèles existants, décrits par des formalismes distincts



AltaRica Dataflow Stochastique

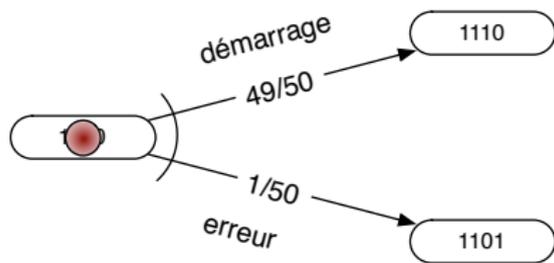
- Basé sur AltaRica Dataflow et les automates de modes (Rauzy, 2002)
- Compositions stochastiques
- Choix aléatoire entre événements concurrents via pondération
- Temporisation aléatoire des événements via distributions de temporisation paramétrées par les variables du système
- Fluctuation non déterministe des variables de flux entrantes



Aspects stochastiques

- Choix du successeur par tirage aléatoire

```
node a
  state
    i : [0, 2];
  event
    demarrage, erreur;
  trans
    i=0 ⊢ erreur → i:=1;
    i=0 ⊢ demarrage → i:=2;
  extern
    choice <demarrage> : 49;
    choice <erreur> : 1;
edon
```



Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés

[...]

extern

```
choice<demarrage>:49;
```

```
choice<erreur>:1;
```

```
Law<demarrage>:
```

```
  Exponential(I1);
```

```
Law<erreur>:
```

```
  Exponential(I2);
```

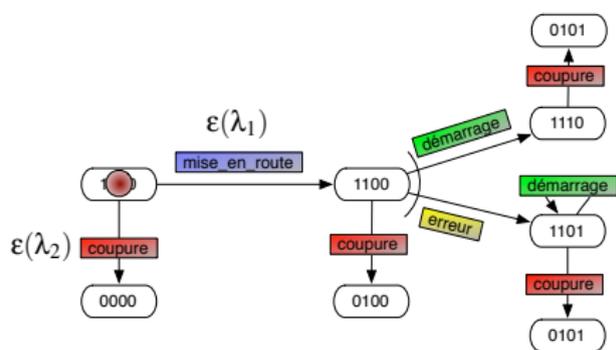
```
Law<mise_en_route>:
```

```
  Exponential(I1);
```

```
Law<coupure>:
```

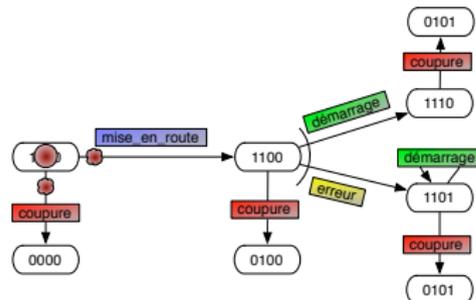
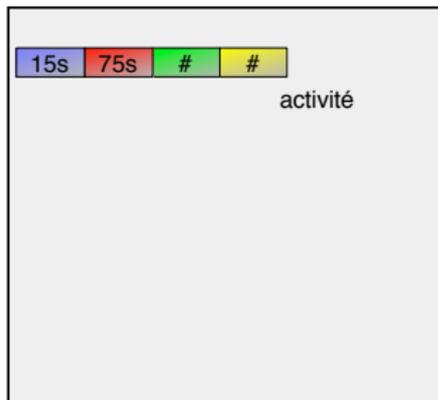
```
  Exponential(I2);
```

[...]



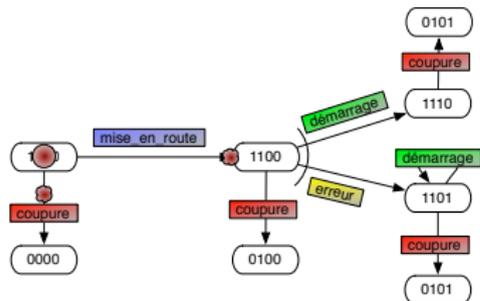
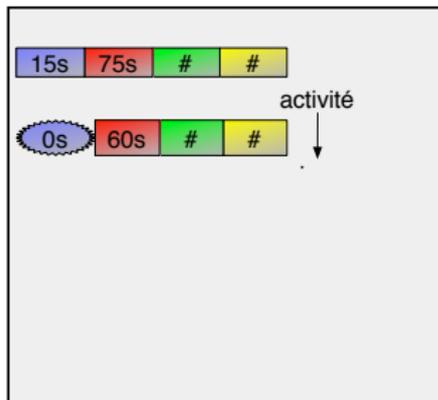
Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés



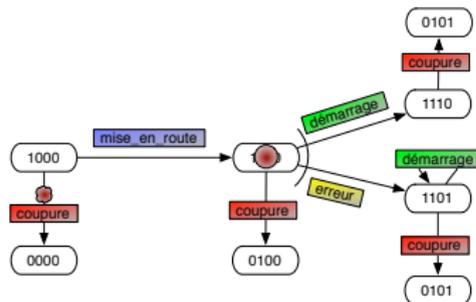
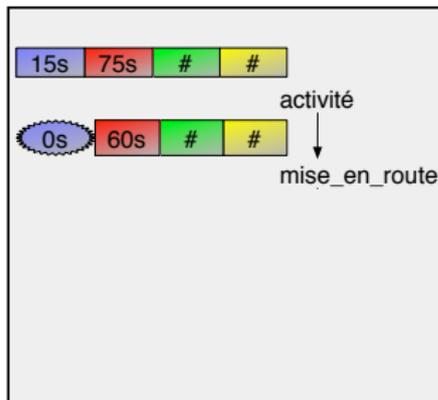
Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés



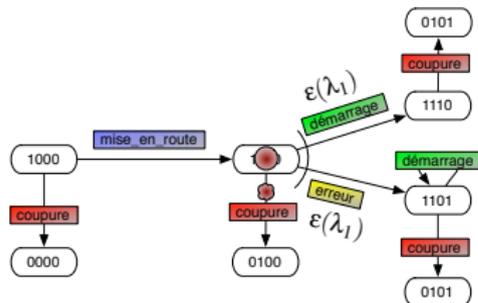
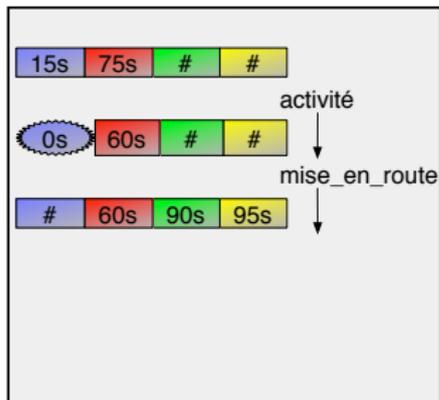
Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés



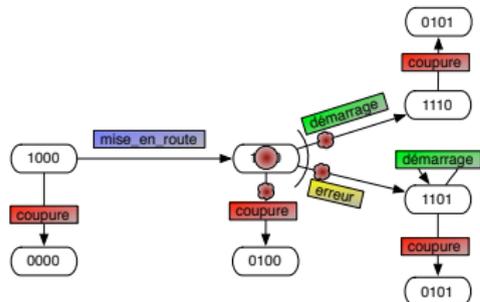
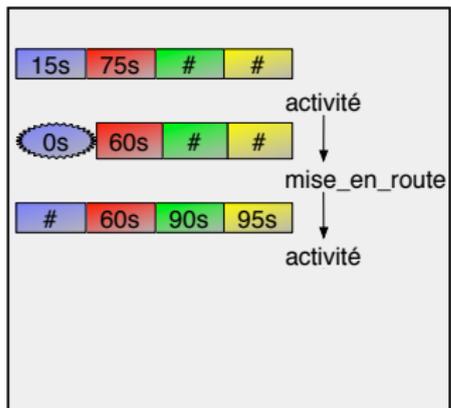
Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés



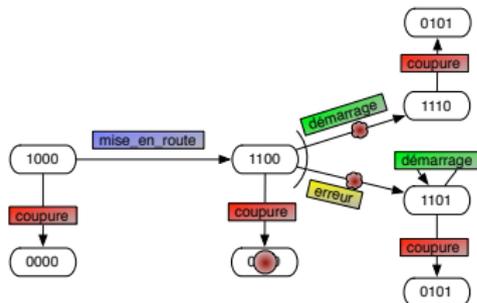
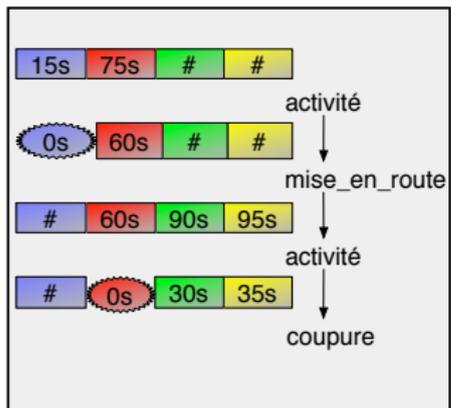
Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés



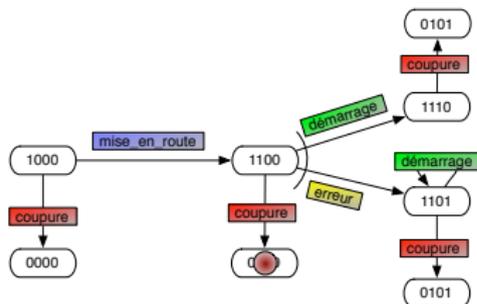
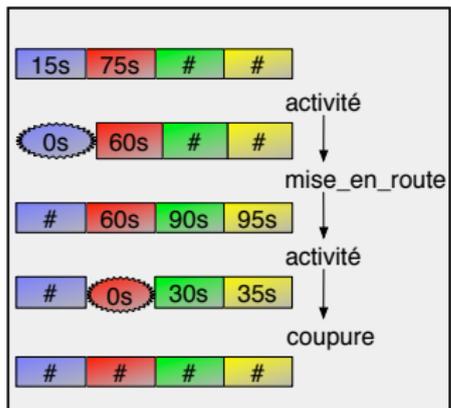
Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés



Aspects stochastiques

- Choix du successeur par tirage aléatoire
- Événements temporisés



Aspects stochastiques

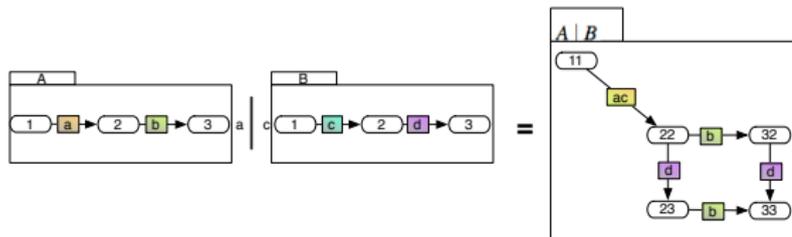
- Choix du successeur par tirage aléatoire
- Événements temporisés
- Sémantique paramétrée par politique de mise à jour
- Sémantique donnée sous forme de processus décisionnels stochastiques continus



Compositions

Définition formelle des extensions stochastiques des compositions

- Parallèle
- Synchronisée (Arnold et al, 1994)
- Par connexion de variable de flux dirigée (Rauzy, 2002)



Modélisation de systèmes différentiels

Intégrateur “numérique” capturé via les distributions de duration paramétrée,

- Discrétisation des variables du système différentiel et
 - Application directe de la loi d'action de masse pour déterminer temps d'incrémentatation / décrémentatation de chaque variable
 - Ou application de la méthode de Gillespie pour simulation atomiste stochastique (Gillespie, 1977)
 - Ou méthode hybride adaptative
- Ou utilisation d'une variable discrète comptant le nombre d'événements d'un système différentiel encapsulé par une distribution de duration
- Combinaison de distribution de probabilité avec intégrateur permettent des intégrations randomisées



Expressivité

Un noeud ARD peut décrire des processus discrets ou stochastiques ou non déterministes, et capture donc :

- Chaînes de markov, Chaînes de markov continues, processus Poissoniens, processus semi-markoviens généralisés (Glynn,1989)
- Processus décisionnel de markov, processus semi-markoviens décisionnels généralisés (Younes,2004)
- Automates finis
- Réseaux de Pétri bornés et leurs variantes stochastiques
- Équations différentielles, systèmes hybrides

BioRica : A multi-model description and simulation system, Soueidan et al, 2007



Outils et implémentations

- Génération de simulateurs via compilation vers du code C++
- Module OCaml des automates de modes stochastiques
- Importation / Exportation de modèles SBML v2 (A.Finney et al, 2004)
- Calculs automatiques de projections vers des formalismes décidables



Plan

- 1 Biologie intégrative, Méthodes formelles et validation
- 2 Extension stochastique d'AltaRica dataflow
- 3 Spécification et vérification de modèles dynamiques**
- 4 Application au vieillissement cellulaire chez la levure
- 5 Conclusion



Objectifs

Description de processus de population avec suivi individuel d'historique, par exemple

- Toutes les cellules finiront par être dans un état stable
- Une protéine p53 nouvellement synthétisée peut être liée à une protéine MDM2 et être déplacé dans le cytosol
- Une protéine p53 dont le site N est phosphorylé puis lié à une protéine Pin1 ne peut plus être lié à une protéine MDM2

Suivi individuel fréquent en biologie moléculaire grâce à l'utilisation de sondes fluorescentes (GFP)



© Toru Shimada, Univ. of Tokyo

Besoins

- Description formelle de processus créant et détruisant nombre non borné de composants
- Admettant un typage des composants via des prédicats unaires
- Admettant la description de liaison n-aires via des prédicats n-aires
- Description formelle des propriétés de ces processus
- Model checking de ces propriétés
- **Problématique proche de la vérification de programmes avec mémoire dynamique**



Description des processus : Automates d'ensembles

- Basés sur la notion générique *d'entité*, représentant quelque chose de créé mais n'ayant pas de dynamique intrinsèque
- Des variables d'états pointant vers ces entités, pouvant les manipuler
 - Allocation, Destruction
 - Affectation a d'autres variables
 - Affectation a un prédicat
- Admettent une sémantique en terme de traces
- (En cours) Sémantique compositionnelle, stochasticité



Description des processus : Automates d'ensembles, exemple

```
node A
  state
    a, b;
    P(1), Q(1), R(2);
  event
    add, remove, move, link;
  trans
     $\mathbb{T} \vdash \text{add} \rightarrow a = \text{alloc};$ 
     $\mathbb{T} \vdash \text{remove} \rightarrow a = \text{free};$ 
     $\exists(x). (P(x), Q(x)) \vdash \text{link} \rightarrow R += (x, x);$ 
     $\mathbb{T} \vdash \text{move} \rightarrow P += (a);$ 
edon
```



Description des propriétés : AITL (Distefano,2005)

Termes $t ::= a|x$

Formule $\phi ::= \mathbb{T} \mid \mathbb{F} \mid \text{alive}(t) \mid \text{new}(t) \mid t = t \mid$
 $(\neg\phi) \mid (\phi \wedge \phi) \mid (\exists x.\phi) \mid X\phi \mid \phi U\phi$

Sémantique Basée sur LTL, étendue avec quantification existentielle et alternance entre quantification et combinateurs temporels



Résultats théoriques

- Quantification existentielle et prédicats unaires permettent d'encoder machine à 2 compteurs avec test à zéro
- Existence d'une abstraction complète et correcte pour modèles sans la quantification existentielle
- Model checking de LTL sur les modèles sans quantification est décidable
- (En cours) Model checking de AITL sur les modèles sans quantification est décidable



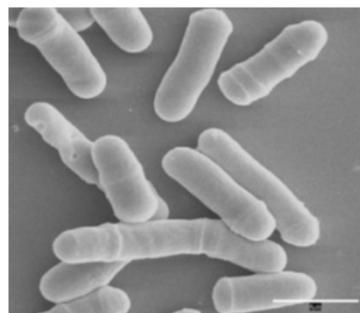
Plan

- 1 Biologie intégrative, Méthodes formelles et validation
- 2 Extension stochastique d'AltaRica dataflow
- 3 Spécification et vérification de modèles dynamiques
- 4 Application au vieillissement cellulaire chez la levure**
- 5 Conclusion



Objectifs

- En collaboration avec M. Cvijovic (Göteborg Univ.) et E.Klipp (Max Planck Inst.,Berlin)
- Construction d'un modèle du vieillissement chez les levures et simulation avec les outils BioRica
- Modèle générique, décrivant les stratégies de division de différentes souches :
 - **Symétrique** Fission binaire chez *Schizosaccharomyces Pombe*
 - **Asymétrique** Bourgeonnement chez *la levure du boulanger*
- Levures sont un organisme modèle pour l'étude de la longévité d'organismes multi-cellulaires



©Prof. Rosa Aligue Alemany,
Univ. de Barcelonne

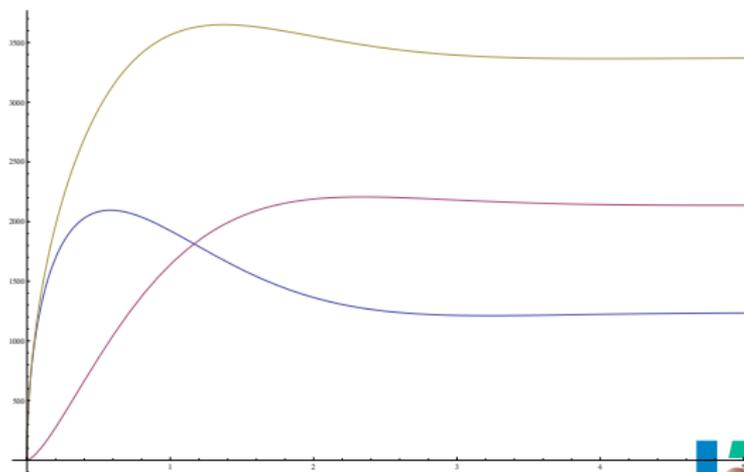


Principe du modèle

Modèle différentiel non linéaire décrivant l'évolution de protéines endommagées et intactes dans une cellule , augmenté de deux transitions discrètes :

- Masse critique atteinte entraîne la division cellulaire
- Seuil critique du taux de variation du nombre de protéines endommagées entraîne la mort cellulaire

Chaque division cellulaire entraîne l'instantiation d'une nouvelle cellule indépendante, générant un arbre de filiation.

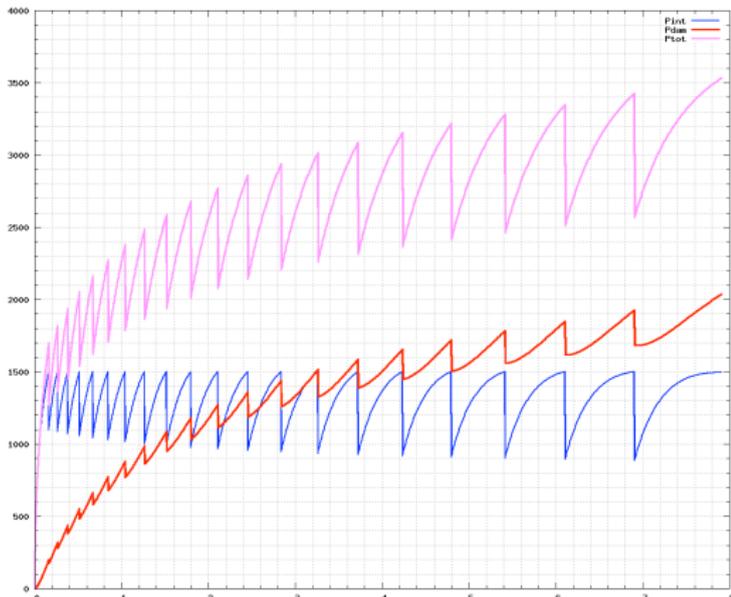


Principe du modèle

Modèle différentiel non linéaire décrivant l'évolution de protéines endommagées et intactes dans une cellule, augmenté de deux transitions discrètes :

- Masse critique atteinte entraîne la division cellulaire
- Seuil critique du taux de variation du nombre de protéines endommagées entraîne la mort cellulaire

Chaque division cellulaire entraîne l'instantiation d'une nouvelle cellule indépendante, générant un arbre de filiation.

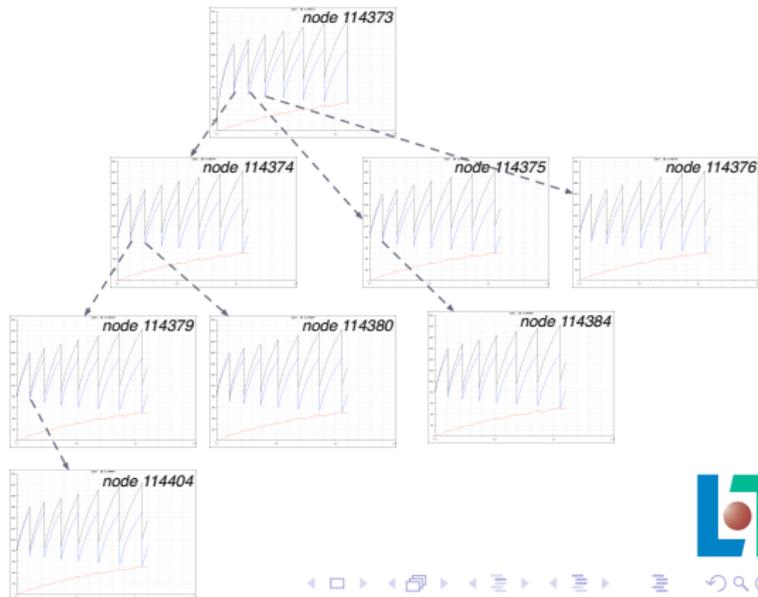


Principe du modèle

Modèle différentiel non linéaire décrivant l'évolution de protéines endommagées et intactes dans une cellule, augmenté de deux transitions discrètes :

- Masse critique atteinte entraîne la division cellulaire
- Seuil critique du taux de variation du nombre de protéines endommagées entraîne la mort cellulaire

Chaque division cellulaire entraîne l'instantiation d'une nouvelle cellule indépendante, générant un arbre de filiation.



Analyse du modèle

Vérification de quelques propriétés non triviales

- Un facteur de rétention ou une division asymétrique entraîne toujours la mort cellulaire
- Il n'existe pas de population ayant des cellules immortelles et mortelles
- Un facteur de rétention entraîne le rajeunissement des dernières cellules filles

à vérifier dans un espace de paramètres à 9 dimensions.

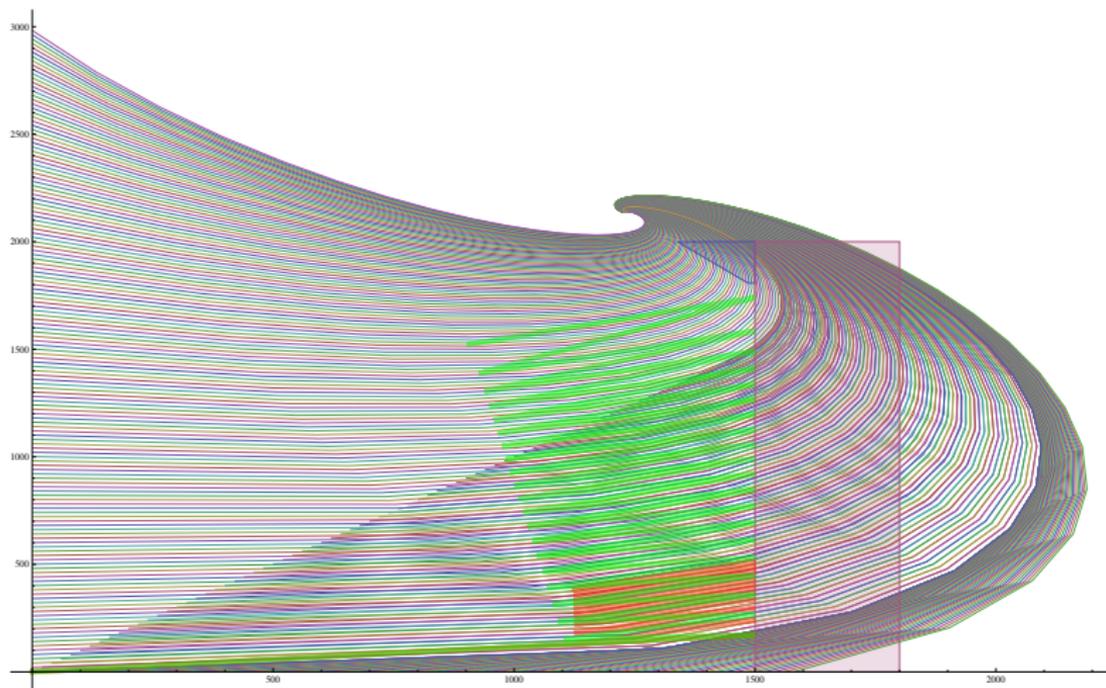


Analyse du modèle, principe de la validation

- Vérification hors de portée des logiciels disponibles de vérification de systèmes hybrides
- Analyse symbolique paramétrée des régions
 - Calcul des régions de mort cellulaire et de division
 - Calcul des ensembles de prédécesseur de ces régions
 - Recherche parmi ces prédécesseurs des trajectoires menant à la mort cellulaire après 1 division
- Pour ce modèle, le calcul de prédécesseurs a toujours convergé
- **Nécessite des représentations symboliques d'ensembles bornés par des polynômes**



Analyse du modèle, principe de la validation



Devenir de la méthode

- Preuve de convergence du calcul des régions
- Identification de classe de modèles hybrides admettant une recherche automatique de régions
- Recherche de représentations symboliques plus adaptées
- Simplification des régions via abstractions ?
- Application à d'autres études de populations de levures, considérée pour l'instant comme non dynamique (Henson et al, 2007)



Devenir du modèle

- Validation expérimentale de l'impact du facteur de rétention
- Evaluation de la robustesse au bruit de l'activation de la division
- Composition avec des modèles du cycle cellulaire (Tyson,2007) et des modèles du stress cellulaire
- Utilisation des bases de données de génomiques comparées (GénoLevures) pour expliciter voies métaboliques influant facteur de rétention



Plan

- 1 Biologie intégrative, Méthodes formelles et validation
- 2 Extension stochastique d'AltaRica dataflow
- 3 Spécification et vérification de modèles dynamiques
- 4 Application au vieillissement cellulaire chez la levure
- 5 Conclusion**



Conclusions et perspectives

- Méthodes formelles permettent trois types de description pertinente en biologie intégrative : normative, explicative et prédictive
- Approches formelles en biologie sont particulières, **mais se basent sur les mêmes outils théoriques que pour l'ingénierie**
- Expressivité forte d'une approche de modélisation de populations à partir de comportements individuels et d'interactions, **mais encore peu explorée**
- Collaborateurs très réceptifs aux principes hiérarchiques, de compositionnalité et aux approches multi-modèles



Merci !

