

Réponse d'un système biologique à une perturbation: théorie et application à l'analyse de données

Anne Siegel, O. Radulescu, M. Le Borgne, C. Guziolowski, P. Veber, S. Lagarrigue

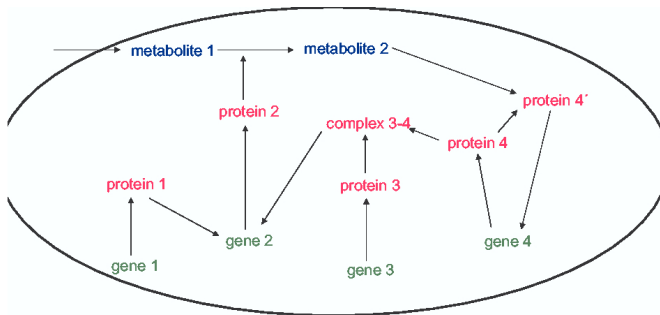
Décembre 2007

Contexte

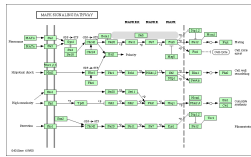
- On dispose d'un catalogue (plus ou moins exact) des **éléments fonctionnels** dans la cellule.
- Des nouvelles technologies fournissent des données en masse
 - **Mécanismes élémentaires**: interactions protéines-protéines, interactions protéines-ADN, réactions biochimiques...
 - **Etat d'une cellule**: expression d'ARN, expression de protéines, phénotype, concentrations de métabolites, flux métaboliques...
- **Le niveau gène/protéine atteint ses limites** pour expliquer le fonctionnement du vivant.

Réseau biomoléculaire

Réseau d'interactions entre des entités biologiques dans une cellule



Source: D. Ropers (Inria Grenoble)



Représentation par graphe...

Interactions

Sommets

- **Gène** (séquence d'ADN)
- **Protéine** (codées à partir d'un gène)
- Complexe protéique
- **Métabolites**
- Autres composés (ATP, ADP, ...)

Transformations

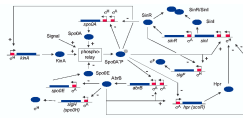
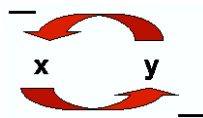
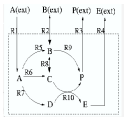
- Réaction chimique
- Expression d'un gène
- Formation d'un complexe

Contrôle d'une transformation

- Catalyse d'une réaction par une enzyme
- Régulation de l'expression d'un gène

Divers

- Co-occurrence dans la littérature



Sources: J. Gagneur (EMBL), D. Thieffry (IBDM, Marseille), H. De Jong (Inria Grenoble)

Différents types de réseaux biomoléculaires

● Réseau de régulation génétique

Interaction de régulation entre des gènes, des protéines et des petites molécules.

Fonction: production régulée de protéines

Sommets: gènes

Arêtes: régulations

Données: séquence, puces

● Réseau métabolique

Série de réactions chimiques contrôlées éventuellement par des enzymes.

Fonction: exploiter et transformer les ressources disponibles en énergie

Sommets: métabolites

Arêtes: consommation, production

Données: séquence, biochimie, spectrométrie

● Réseau de signalisation

Associations et dissociations de molécules (phosphorylation...).

Fonction : Transport rapide de l'information

Sommets: protéines et état d'activité

Arêtes: interactions modifiant les états

Données: mesure de modification post-traductionnelles.

● Réseau d'interaction protéine-protéine

Interaction physique entre molécules

Sommets: protéines

Arêtes: interactions physiques

Données: double-hybrides, spectrométrie

Question concrète

- **Nouvelles techniques** en biologie moléculaire:
 - **puces à ADN.** Variations de tous les produits géniques d'une cellule
Dans une cellule, on compare la quantité de milliers de protéines à un instant donné (sous la forme d'ARN messagers).
 - **CHip-on-Chip.** Cibles d'une protéine donnée
On choisit une protéine. On peut trouver tous les morceaux d'ADN sur lesquels la protéine peut se fixer pour déclencher la transcription d'une protéine.
- **Connaissances** existant dans la littérature sur le mode d'action d'un produit sur un autre. Effort naissant de standardisation.
- Applications importantes en diagnostic.

[Biologiste] J'ai des gigaoctets de données sur mon disque dur. Mais finalement je ne sais pas quoi en faire et comment les exploiter...

Ambitions du biologiste

Le développement et la fonction d'un organisme émergent des interactions dans et entre ces différents types de réseaux...

- Aider à structurer, **représenter** et interpréter des données expérimentales
 - Intégrer différents types de données.
 - Relier les mécanismes décrits aux états du système.
- Comprendre les processus cellulaires
 - **Prédire** l'état d'une cellule
 - Prédire l'effet d'une modification du système sur ses éléments.
- Aider les **expérimentations**, en produisant des hypothèses vérifiables sur un système.
- Comparer des réseaux pour différents organismes ou tissus, en rapport avec leur évolution.
- Mieux comprendre la **fonction** de gènes donnés.

Questions mathématiques et informatiques

- (Re)construire/identifier des réseaux à partir de données expérimentales ou théoriques
[Statistiques, programmation par contraintes, apprentissage]
- Définir des types de **modélisation**
[en fonction du système biologique, des prédictions recherchées, des données expérimentales disponibles]
- Analyser le **comportement** d'un modèle.
[Simulation, Stabilité, Prédiction de comportements. Effets du bruit.]
- Analyser la **structure** des réseaux
[propriétés structurelles (motifs). Modularité. Evolution différents organismes.]
- **Applications**
[Fonction d'une protéine. Contrôler le système ?]

Différents modèles

Graphes d'interactions

Modèles discrets/booléens

Les composants ont un petit nombre d'états qualitatifs.
Fonctions de transitions.

Linéaires par morceaux

formalisme différentiel
les composants ont un nombre fini d'états qualitatifs.
Fonctions seuils.

Modèles différentiels

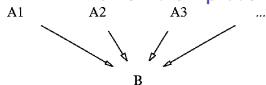
Grand nombre de molécules.
Coefficients stochiométriques et paramètres.

Modèles stochastiques

Petit nombre de molécules.
Lois de probabilités.

Modélisation différentielle

Le flux d'un produit est la somme des influences qui y arrivent.

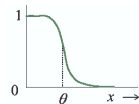


$$\frac{dB}{dt} = K_1 v(A_1) + K_2 v(A_2) + \dots$$

- Une équation pour chaque noeud ayant des arêtes entrantes
- les fonctions cinétiques $v(A)$ **dépendent du type d'interaction**
 - paliers si réaction enzymatique;
 - seuil/sigmoïde si interaction génétique;
 - linéaire si réaction chimique usuelle.



$$\begin{cases} \frac{dx}{dt} = k_x F_{s_x}^-(y) - k_{-x} x \\ \frac{dy}{dt} = k_y F_{s_y}^-(x) - k_{-y} y \end{cases}$$



$$F_s^-(x) = \frac{x^n}{x^n + s^n}.$$

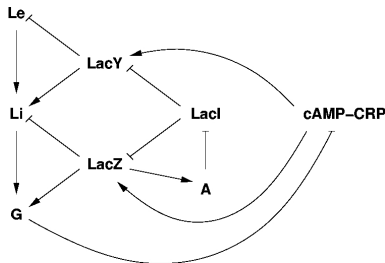
Modèle différentiel

- Vecteur des concentrations $X_{(i)}$ des molécules: \mathbf{X}
- Ensemble de paramètres de contrôle du système \mathbf{P}
- Système différentiel

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \mathbf{P})$$

- Graphe d'interaction: matrice jacobienne

$$i \rightarrow j \quad \text{ssi} \quad \frac{\partial F_{X_{(j)}}}{\partial X_{(i)}} \neq 0$$



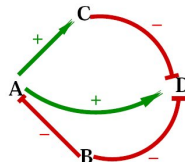
Connaissances sur le modèle

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \mathbf{P})$$

- Réseaux métaboliques et de signalisation: \mathbf{F} est souvent explicite(abl)e.
Simulation de dynamiques
- Réseaux de régulation génétique:
 - Existence d'arcs [données (promoteurs, chip-on-chip) et/ou littérature]
 - Signe des interactions [littérature]
 - Modélisation par systèmes linéaires par morceaux [littérature, model-checking]

On ne peut pas connaître explicitement \mathbf{F}

On peut (en travaillant) connaître le graphe d'interaction (signes de la matrice jacobienne)



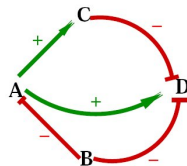
Données à interpréter (massivement): puces à ADN

Dans une cellule, on compare le nombre d'ARN messagers (précurseurs de protéines) de milliers de gènes à un instant donné.

- Pas de séries temporelles (quelques points espacés au mieux)
- Bruitées
- **Qualitatives**: on donne un ratio (plus ou moins valable) par rapport à une valeur de référence.



On peut savoir si la quantité de produits génétiques a augmenté ou diminué entre des intervalles de temps assez distants



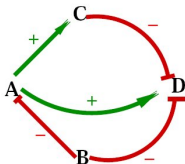
A augmente
C diminue

Résumé

- On suppose que l'action des produits sur les autres est régie par un **système différentiel** $\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \mathbf{P})$
- On peut éventuellement connaître les **signes des différentielles partielles** de \mathbf{F}
- On connaît le **signe de la variation de certaines variables** pendant une expérimentation

Déplacement entre deux états d'équilibre

Que peut-on apprendre sur le système ? Comment exploiter ces données et connaissances ?



La concentration de A augmente
la concentration de C diminue

Que peut-on dire ?

Une règle intuitive...

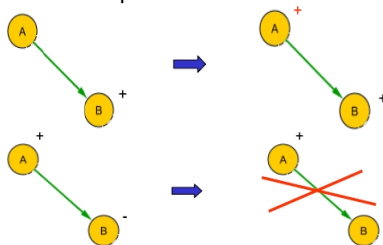
La variation d'un élément doit être expliquée par la variation d'un de ses prédecesseurs.



Cette règle est-elle seulement vraie ?

Une règle intuitive...

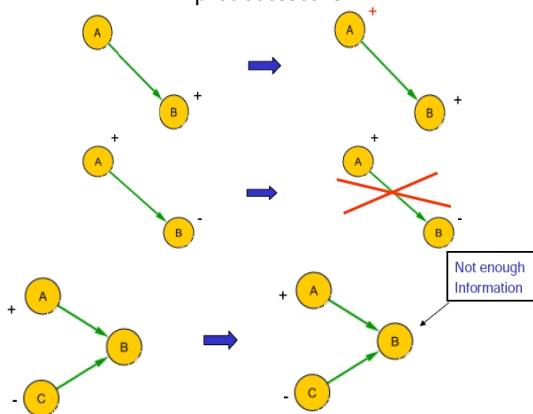
La variation d'un élément doit être expliquée par la variation d'un de ses prédecesseurs.



Cette règle est-elle seulement vraie ?

Une règle intuitive...

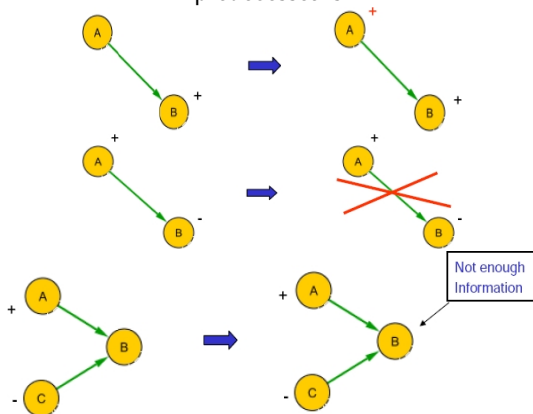
La variation d'un élément doit être expliquée par la variation d'un de ses prédécesseurs.



Cette règle est-elle seulement vraie ?

Une règle intuitive...

La variation d'un élément doit être expliquée par la variation d'un de ses prédecesseurs.



Cette règle est-elle seulement vraie ?

Calcul des petites variations

- Système stable $\mathbf{F}(\mathbf{X}, \mathbf{P}) = \mathbf{0}$
- Variation des paramètres puis stabilisation du système.
- Mesure des variations des produits.

Proposition

Si $X_{(i)}$ n'est pas influencé directement par les paramètres ($\frac{\partial \mathbf{F}_{X_{(i)}}}{\partial P_{(k)}} = 0$) et $\left(\frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(i)}}\right) \neq 0$, alors les variations de $X_{(i)}$ peuvent se déduire des variations de ses prédécesseurs dans le graphe:

$$\delta X_{(i)} = - \left(\frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(i)}} \right)^{-1} \sum_{k \neq i, k \rightarrow i} \frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(k)}} \delta X_{(k)}.$$

Peut-on intégrer cette relation ?

$$\delta X_{(i)} = - \left(\frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(i)}} \right)^{-1} \sum_{k \neq i, k \rightarrow i} \frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(k)}} \delta X_{(k)}.$$

On a $\mathbf{F}_{X_{(i)}} = \mathbf{F}_{X_{(i)}}(X_{(i)}, \hat{X}^{(i)})$. avec $\hat{X}^{(i)} = (X_{(k_1)}, \dots, X_{(k_p)})$, $k_j \neq i, k_j \rightarrow i$.

Il faut exprimer globalement $X_{(i)}$ comme fonction de $\hat{X}^{(i)}$

Si $\frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(i)}} < -C$, alors $\mathbf{F}_{X_{(i)}}(X_{(i)}, \hat{X}^{(i)})$ est monotone en $X_{(i)}$ et négative en l'infini.

Puisque $\mathbf{F}_{X_{(i)}}(\{X, X_{(i)} = 0\}) > 0$, la fonction s'annule exactement une fois:

$$X_{(i)} = \Phi_i(\hat{X}^{(i)}).$$

Fonctions implicites $d\Phi_i = - \left(\frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(i)}} \right)^{-1} \sum_{k \neq i, k \rightarrow i} \frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(k)}} dX_{(k)}.$

On prend un chemin différentiable entre deux états d'équilibre de \mathbf{F} qui vérifie $\mathbf{F}_i(\Phi_i(C^{1,2}), \hat{X}^{(i)}(C^{1,2})) = 0$ et évite les points singuliers de $\mathbf{F}_{X_{(i)}}$.

$$\Delta X_i = \int_{C^{1,2}} - \left(\frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(i)}} \right)^{-1} \sum_{k \neq i, k \rightarrow i} \frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(k)}} dX_{(k)}.$$

La trajectoire de Φ_i est bornée quand $\hat{X}^{(i)}$ parcourt $C^{1,2}$. Donc $C^{1,2}$ est proche des valeurs d'équilibre expérimentales. Les signes restent constant sur cette trajectoire.

Généralisation à de larges variations

Les variations d'un produit sont transmises par ses prédécesseurs dans le graphe.

Théorème

Si

- $\frac{\partial \mathbf{F}_{X(i)}}{\partial X(i)} < -C, \quad C > 0.$
- Il n'y a pas d'influence directe de \mathbf{P} sur $X(i)$
- $\mathbf{F}_{X(i)}(\{\mathbf{X}, X(i) = 0\}) > 0$
- Pour tout $k \rightarrow i$, le signe de $\frac{\partial \mathbf{F}_{X(i)}}{\partial X(k)}$ est constant sur une plage d'expérimentation

Alors les signes des variations vérifient la contrainte suivante.

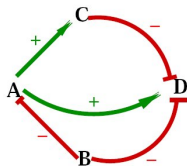
$$\text{signe}(\Delta X(i)) \simeq \sum_{k \neq i, k \rightarrow i} \text{signe} \left(\frac{\partial \mathbf{F}_{X(i)}}{\partial X(k)} \right) \times \text{signe}(\Delta X(k)).$$

Les relations d'addition, multiplication et d'égalité sont valables dans l'**algèbre des signes** $\{[+], [-], [?]\}$

$$[+] + [-] = [?], \quad [+] \simeq [?] \quad [-] \simeq [?]$$

Variation de concentrations entre deux états d'équilibre

$$\text{signe}(\Delta X_{(i)}) \simeq \sum_{k \neq i, k \rightarrow i} \text{signe} \left(\frac{\partial F_{X_{(i)}}}{\partial X_{(k)}} \right) \times \text{signe}(\Delta X_{(k)}).$$

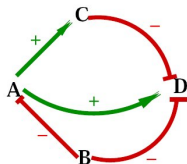


- $\text{signe}(\Delta C) \approx \text{signe}(\Delta A)$
- $\text{signe}(\Delta A) \approx -\text{signe}(\Delta B)$
- $\text{signe}(\Delta D) \approx \text{signe}(\Delta A) - \text{signe}(\Delta B) - \text{signe}(\Delta C)$

A ne peut pas diminuer si C augmente

Concrètement

- On construit un réseau biologique à partir des connaissances
 - Graphe orienté signé
 - Sommets: produits d'un réseaux biologique
 - Arcs: l'origine influence la production de la cible [Activation, Repression, Double action]
- on suppose que toutes les interactions vérifie les bonnes hypothèses
- S'il n'y a pas de boucle de rétro-régulation positive sur un sommet, on écrit l'équation de dépendance en fonction de ses prédécesseurs.

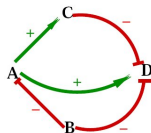


- $\text{signe}(\Delta C) \approx \text{signe}(\Delta A)$
- $\text{signe}(\Delta A) \approx -\text{signe}(\Delta B)$
- $\text{signe}(\Delta D) \approx \text{signe}(\Delta A) - \text{signe}(\Delta B) - \text{signe}(\Delta C)$

Etude des solutions du système

Solution du système:

- Chaque variable doit prendre la valeur **+** ou **-**
- Toutes les équations sont vérifiées simultanément



$$\begin{aligned}
 + &\approx + \\
 + &\approx -(-) \\
 + &\approx + - (-) - (+)
 \end{aligned}$$

4 jeux de variations
(parmi 16) autorisés

A	B	C	D
+	-	+	+
+	-	+	-
-	+	-	+
-	+	-	-

Etude des solutions

- **PYQUALI / BIOQUALI**: Librairie Python développée à l'IRISA
- Les équations sont codées comme des polynômes sur le corps $\mathbb{Z}/3\mathbb{Z}$. Recherche du pgcd de tous les polynômes.
- Le système est réduit en un système plus petit équivalent
- Les solutions sont représentées par des arbres de décisions ternaires Ternaires (TDD)
- L'existence de solution est étudiée sur le petit système

A quoi ca sert ?

- Tester la **compatibilité** entre un réseau et des données
Le système d'équations doit avoir une solution.
- Proposer des **corrections** en cas d'incompatibilité
Trouver un sous-système en dehors duquel une solution existe
- Prédire de nouvelles **variations**
Solutions fixées dans toutes les solutions du système.

Concrètement: un serveur web...

Plateforme Bio-informatique GENOUEST

Rechercher

Accueil FAQ Plan Contacts English Imprimer

La plate-forme Outils Banques Séminaires Formations Services Aide

Vous êtes ici : Accueil > Outils > Réseaux d'Interactions > Bioquali

Bioquali

BioQuali: Network Compatibility and products variation inference in a biological network

What's this framework for?
Data example

Step 1 :

- Interaction's data file - Use the previous graph ☐

Step 2 :

- Experimental data set

Upload a graph file

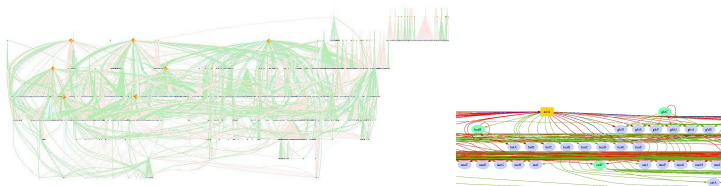
Data example

Regulatory network of an organism O

Experimental data associated to organism O

Analysis

Exemple Grande échelle: E. Coli



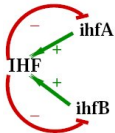
- Réseau large échelle
- **Structure hiérarchique** (87% des gènes sont régulés par 13%)
- 160 interactions avec un **double signe**
- 7 **facteurs principaux** : (> 80 successeurs): CRP, FNR, IHF, FIS, ArcA, NarL and Lrp
- **1100 équations qualitatives**, 1258 variables

Number of nodes	1258
Number of interactions	2526
Nodes without successor	1101
Nodes with more than 80 successors	7
protein complex	4

Le système d'équations qualitatives pour le réseau transcriptionnel d'E. Coli n'a aucune solutions.

Corrections

- Problème (étude automatique des équations):



$$IHF \approx ihfA + ihfB \quad (1)$$

$$ihfA \approx -IHF \quad (2)$$

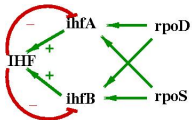
$$ihfB \approx -IHF \quad (3)$$

système inconsistant

(aucune solution)

ihfA	ihfB	IHF	Conflict
+	+	+	(2), (3)
+	+	-	(1)
+	-	+	(1)
+	-	-	(1)
-	+	+	(1)
-	+	-	(1)
-	-	+	(1)
-	-	-	(2), (3)

- Solution: Ajout de nouvelles interactions (facteurs sigma)



$$IHF \approx ihfA + ihfB$$

$$ihfA \approx -IHF + rpoD + rpoS$$

$$ihfB \approx -IHF + rpoD + rpoS$$

Système consistant (18

solutions parmi 32)

rpoD	rpoS	ihfA	ihfB	IHF
+	+	+	+	+
+	+	+	-	+
+	+	-	+	+
-	-	-	-	-
-	-	-	+	-
-	-	+	-	-
+/ -	-/+	+	+	+
+/ -	-/+	+	-	+
+/ -	-/+	+	-	-
+/ -	-/+	-	+	+
+/ -	-/+	-	-	-

Protein	Gene	Function
σ^{70}	rpoD	Transcribes most genes in growing cells
σ^{38}	rpoS	The starvation/stationary phase sigma-factor
σ^{28}	rpoF	The flagellar sigma-factor
σ^{32}	rpoH	The heat shock sigma-factor
σ^{24}	rpoE	The extracytoplasmic stress sigma factor

Cohérence interne pour E. Coli



Number of nodes	1529
Number of interactions	3883
Nodes without successor	1365
Nodes with more than 80 successors	10
sigma-factors	6
protein complex	4

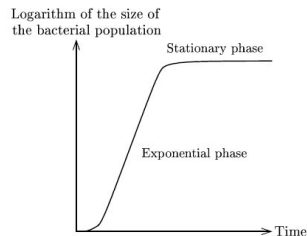
- Ajout des interactions pour les facteurs-sigma et leurs cibles
- Pas d'équations pour les facteurs sigmas, puisqu'ils reçoivent des influences de l'extérieur

Protein	Gene	Induced Genes
σ^{70}	rpoD	1047
σ^{38}	rpoS	114
σ^{54}	rpoN	100
σ^{24}	rpoE	48
σ^{32}	rpoH	29
σ^{19}	fecI	7

Le réseau de régulation transcriptionnel pour E. Coli et ses facteurs sigma est cohérent.

Et si on rajoute des données ?

- Croissance exponentielle de E.coli sous stress nutritionnel
- Variation de produits entre deux états d'équilibre (phases exponentielle et stationnaire)
- 40 données expérimentales (lues dans la littérature par des spécialistes)



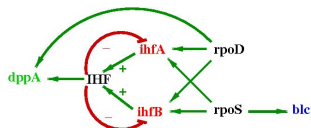
Ropers *et al.*, Biosystems, 2004.

gene	effect	gene	effect	gene	effect	gene	effect	gene	effect
acnA	+	csiE	+	gadC	+	osmB	+	recF	+
acrA	+	cspD	+	hmp	+	osmE	+	rob	+
adhE	+	dnaN	+	hns	+	osmY	+	sdaA	-
appB	+	dppA	+	hyaA	+	otsA	+	sohB	-
appC	+	fic	+	ihfA	-	otsB	+	treA	+
appY	+	gabP	+	ihfB	-	polA	+	yeiL	+
blc	+	gadA	+	lrp	+	proP	+	yfiD	+
bolA	+	gadB	+	mpl	+	proX	+	yihI	-

Incohérence: Il n'y a pas de solutions du système qui se superpose aux observations

Que faire

Sous graphe incohérent: (plus petit polynome premier avec le reste)

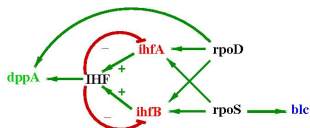


24 solutions cohérentes :

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	+	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	-	-	-	-	+	-

Que faire

Sous graphe incohérent: (plus petit polynome premier avec le reste)



4 solutions cohérentes :

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
-	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	+	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Observations utilisées

● *ihfA* = -

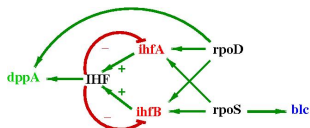
● *ihfB* = -

● *blc* = +

● *dppA* = +

Que faire

Sous graphe incohérent: (plus petit polynome premier avec le reste)



1 solution cohérente :

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	+	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Observations utilisées

● *ihfA* = -

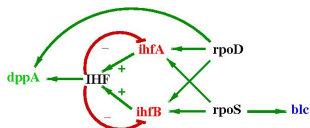
● *ihfB* = -

● *blc* = +

● *dppA* = +

Que faire

Sous graphe incohérent: (plus petit polynome premier avec le reste)



0 solution cohérente :

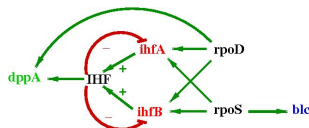
rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Observations utilisées

- *ihfA* = -
- *ihfB* = -
- *blc* = +
- *dppA* = +

Correction du modèle

Sous graphe incohérent: (plus petit polynome premier avec le reste)



2 solutions cohérentes:

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/ -	- / +	+	-	+	- / +	+/ -
-	+	+	-	+	+	+
+	+	+	-	-	-	+
+/ -	- / +	+	-	-	- / +	-
+/ -	- / +	-	+	+	- / +	+/ -
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/ -	- / +	-	+	-	- / +	+/ -
+	-	-	-	-	-	+
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Proposition de corrections

- *ihfA* = +
- *ihfB* = +
- *blc* = +
- *dppA* = +

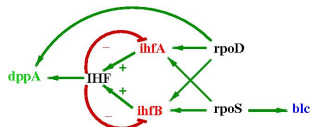
Lecture approfondie des articles de biologie

Les données fournies par la base de données sur la littérature étaient incorrectes

Le réseau transcriptionnel d'E. Coli et ses facteurs sigma avec les 40 observations corrigées est cohérent (enfin !)

Correction du modèle

Sous graphe incohérent: (plus petit polynome premier avec le reste)



2 solutions cohérentes:

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/ -	- / +	+	-	+	- / +	+/ -
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/ -	- / +	+	-	-	- / +	-
+/ -	- / +	+	-	-	- / +	+/ -
+	+	-	+	+	+	+
+	-	-	+	-	-	-
+/ -	- / +	-	+	-	- / +	+/ -
+	-	-	-	-	-	+
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Proposition de corrections

- *ihfA* = +
- *ihfB* = +
- *blc* = +
- *dppA* = +

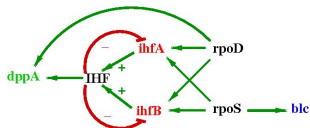
Lecture approfondie des articles de biologie

Les données fournies par la base de données sur la littérature étaient incorrectes

Le réseau transcriptionnel d'E. Coli et ses facteurs sigma avec les 40 observations corrigées est cohérent (enfin !)

Prédictions

- On part des observations compatibles avec le réseau



Observations:

ihfA = + *ihfB* = +

blc = + *dppA* = +

2 solutions cohérentes:

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
-	+	+	+	+	+	+

- Inférence :

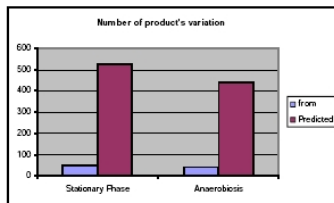
- Pas de valeur constante pour *rpoD*
- rpoS* = +
- IHF* = +

- Cas réel: 40 observations sur 1529 sommets. $5.2 \cdot 10^{16}$ solutions aux équations.

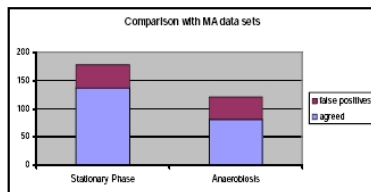
les 40 observations permettent de prédire la variation de 381 nouveaux produits.

Validation des prédictions

- Microarray data set:
 - from Faith *et al.* 2007 and Allen *et al.* 2003
 - Condition: Stationary Phase after 720min vs log phase
 - Threshold = 2-fold
 - 720 products



- 526 variations predicted
- Comparison between predictions and microarray
 - 177 common observed products
 - 77% of them agreed
- Similar results with other experimental conditions



Les prédictions sont validées à 80%.

Origine des erreurs: il manque des interactions (surtout post-traductionnelles) dans le réseau.

Résumé de la démarche

Bio On part de connaissances biologiques

Math On suppose qu'il existe un système dynamique différentiel qui modélise bien tout cela

Math Les connaissances renseignent certaines propriétés du système différentiel (signe du Jacobien)

Math Système de contraintes vérifiées par les variations des concentrations des produits

Info+Math Outils informatiques performants pour résoudre ces contraintes

le modèle proposé n'était pas bon: pas de cohérence interne

Info Localisation des problèmes

Bio proposition de corrections du modèle.

Info Le modèle n'est pas compatible avec les données.

Info Localisation des problèmes

Bio Proposition de corrections des données.

Prédictions de comportements

Ingénieurs Mise à disposition des méthodes.

Il faut faire travailler ensemble les biologistes, informaticiens et mathématiciens pour parvenir à un résultat...

Conclusion

Formalisme d'analyse, diagnostic et prédictions à partir d'un réseau biologique décrit par un graphe d'influence.

- Nouvelles méthodes pour identifier les erreurs et proposer des corrections?
- Peut-on trouver les parties du réseau qui le contraignent le plus ? (observations à privilégier)
- Quelles sont les combinaisons de produits “manipulables” qui permettent d'obtenir une variation de produit fixée ?
- Peut-on dire quelque chose sur la **dynamique du système** à partir des connaissances disponibles ?
- On a fait des approximations pour pouvoir exploiter les données biologiques.

Sont-elles graves ?

Dans tous les cas, les systèmes biologiques fournissent des classes de systèmes dynamiques très particulier et les biologistes sont très demandeurs d'outils de prédictions à partir de leurs données.

Les acteurs

:

Info. M. Le Borgne, P. Veber, C. Guziolowski

Math. O. Radulescu, M. Le Borgne

Bio. S. Lagarrigue, C. Guziolowski

Ingé. G. Ranchy, A. Bourdé