

Réseaux biologiques: des mesures et des math

La Reposée, novembre 2005

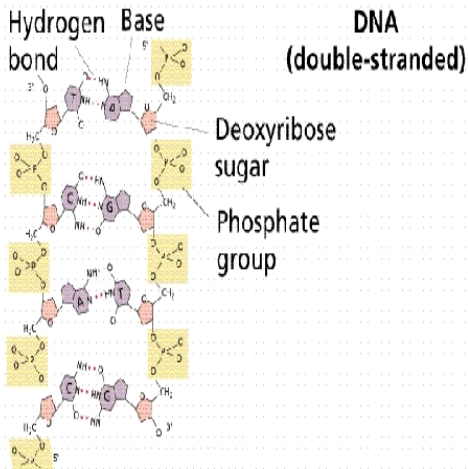
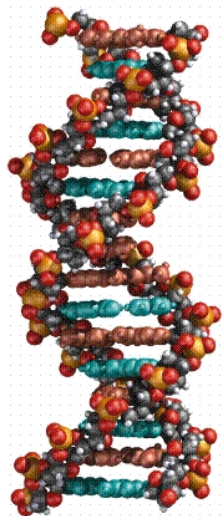
Comment construire des modèles?

- Mesures en cours
- Connaissances en stock (BD?)
- Statique comparative
- Et la dynamique?
- Back to sequences ...

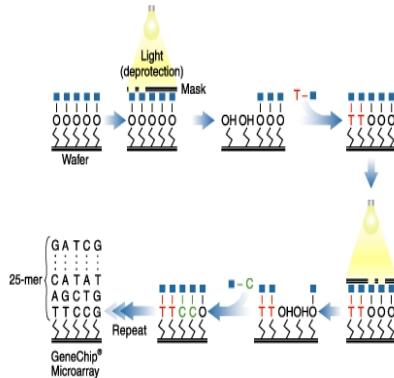
On prend de nouvelles mesures

- microarrays
- spectrométrie de masse
- double hybride

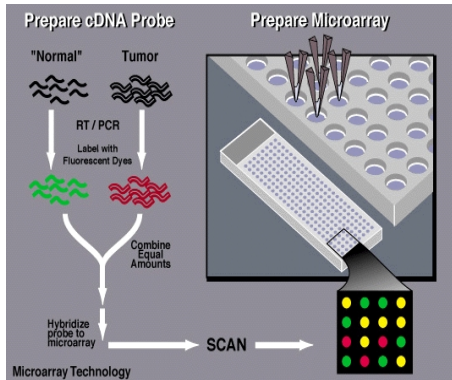
Microarray: l'ADN s'hybride



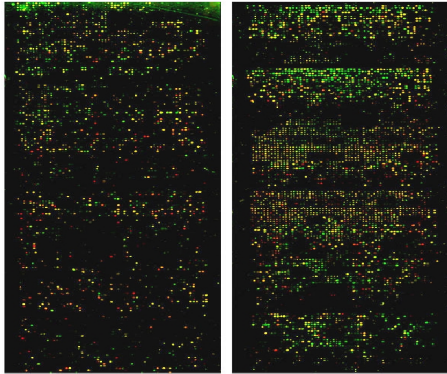
Technologie microarray(1)



Technologie microarray(2)



Technologie microarray (3)



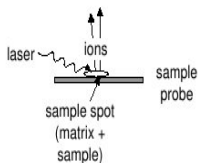
Problemes

- selection des échantillons de tissus
- synchronisation
- préparation des cDNA
- qualité d'hybridation
- traitement d'image
- traitement statistiques
- mesures différentielles

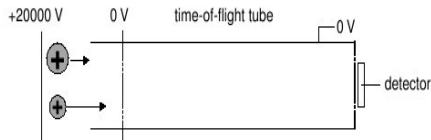
Spectrométrie de masse

AFDEK[AWSVAR[VVAGVANALAHK[EK[FVIEIR[....

(a) MALDI



(b) TOF



Spectrométrie de masse (2)

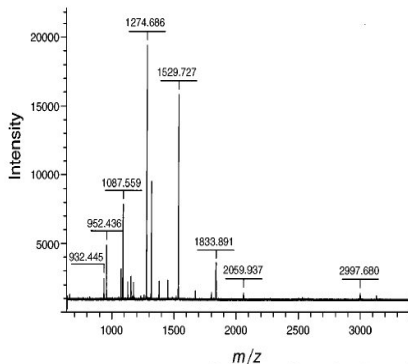


Figure 2. MALDI-TOF spectrum of a mixture of peptides obtained from tryptic digestion of hemoglobin (human).

Double hybride

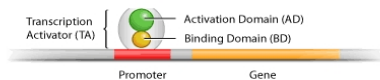


Figure 1. Normal Transcription. Normal transcription requires both the DNA-binding domain (BD) and the activation domain (AD) of a transcriptional activator (TA).

Double hybride

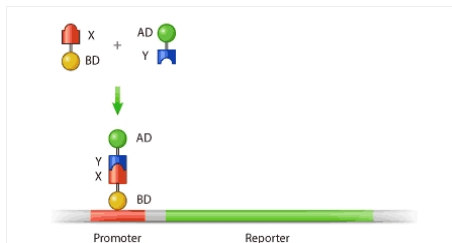


Figure 2. Yeast two-hybrid transcription. The yeast two-hybrid technique measures protein-protein interactions by measuring transcription of a reporter gene. If protein X and protein Y interact, then their DNA-binding domain and activation domain will combine to form a functional transcriptional activator (TA). The TA will then proceed to transcribe the reporter gene that is paired with its promoter.

Problèmes

- résultats qualitatifs
- complémentaire micro-array (spectro)
- ne donnent pas d'informations comportementales
- observation simultanée d'un grand nombre d'évènements

Le capital de connaissances

- peu ou pas formalisées
- dispersion dans la littérature
- extraction difficile
- comparaison difficile:
 - espèces différentes
 - conditions différentes (in vivo, in vitro, en culture)
 - grande variabilité

BD de connaissances

- BD références bibliographiques (Medline)
- BD de réseaux métaboliques/géniques (Kegg, Ecocyc ...)
- BD de connaissance (Gene ontology)
- Outils de fouille de données (Bibliosphère ...)
- Langages de description de systèmes biologique.

Quelle est l'attente?

- rien
- tout
- critères de qualité flous
- expérimentation à l'épuisette

Un modèle doit donner les moyens d'être contredit

Un exemple en statique comparative

- contexte: métabolisme des lipides et contrôle génétique
- données a-priori (BD)
- données de micro-arrays
- comparaison de deux situations: a jeun/nourri

Hypothèses fondamentales

- on observe deux points stationnaires différents
- les connaissances à priori représentent des constantes
- les lois de la biochimie s'appliquent
- les mesures d'ARNm sont traduit en niveaux des protéines

Du modèle différentiel au modèle discret

$$\frac{dX}{dt} = F(X) \quad (1)$$

$$F(X_{eq}) = 0 \quad (2)$$

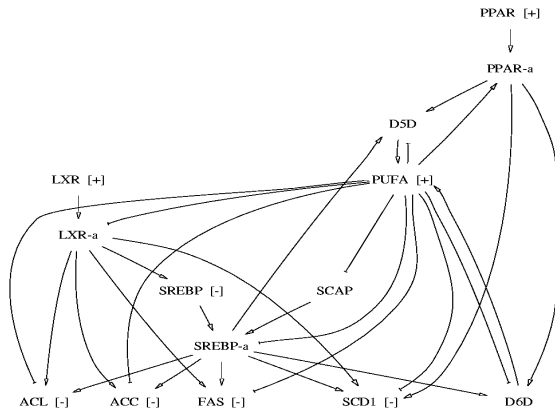
Variations = déplacement d'équilibre

Graphe d'interactions

- noeuds $V = \{1, \dots, n\}$ correspondent aux produits
- arcs $E = \{(j, i) \mid \frac{\partial F_i}{\partial X_j} \neq 0\}$. étiquetés par $s(j, i) = \text{sgn}(\frac{\partial F_i}{\partial X_j})$

En fait personne n'a vu F , par contre le graphe modélise la connaissance à priori.

Exemple



Fonction implicite et variations (1)

- $\frac{\partial F_i}{\partial X_i} < 0$
- $X_i \rightarrow F_i(X_i, \hat{X}_i)$ strictement décroissante
- $F_i(X_i, \hat{X}_i) = 0$ a au plus une solution en X_i
- $F_i(X_i, \hat{X}_i) = 0$ ssi $X_i = \Phi_i(\hat{X}_i)$

Fonction implicite et variations (2)

$$d\Phi_i = - \left(\frac{\partial F_i}{\partial X_i} \right)^{-1} \left(\sum_{k \in \text{pred}(i)} \frac{\partial F_i}{\partial X_k} dX_k \right) \quad (3)$$

$$X_{eq_i}^1 - X_{eq_i}^2 = \delta X_i = \int_C - \left(\frac{\partial F_i}{\partial X_i} \right)^{-1} \sum_{k \in \text{pred}(i)} \frac{\partial F_i}{\partial X_k} dX_k \quad (4)$$

Passage dans l'algèbre des signes (1)

$++- = ?$ $+++ = +$ $-+- = -$ $+ \times - = -$ $+ \times + = +$ $- \times - = +$
 $?+- = ?$ $?++ = ?$ $?+? = ?$ $? \times - = ?$ $? \times + = ?$ $? \times ? = ?$

\approx	$+$	$-$	$?$
$+$	T	F	T
$-$	F	T	T
$?$	T	T	T

Passage dans l'algèbre des signes

$$X_{eq_i}^1 - X_{eq_i}^2 = \delta X_i = \int_C - \left(\frac{\partial F_i}{\partial X_i} \right)^{-1} \sum_{k \in \text{pred}(i)} \frac{\partial F_i}{\partial X_k} dX_k \quad (5)$$

$$s(\Delta X_i) \approx \sum_{k \in \text{pred}(i)} s(k, i) s(\Delta X_k) \quad (6)$$

Exemple

- (1) $\text{PPAR-a} = \text{PPAR} + \text{PUFA}$
- (2) $\text{LXR-a} = -\text{PUFA} + \text{LXR}$
- (3) $\text{SREBP} = \text{LXR-a}$
- (4) $\text{SREBP-a} = \text{SREBP} + \text{SCAP} - \text{PUFA}$
- (5) $\text{ACL} = \text{LXR-a} + \text{SREBP-a} - \text{PUFA}$
- (6) $\text{ACC} = \text{LXR-a} + \text{SREBP-a} - \text{PUFA}$
- (7) $\text{FAS} = \text{LXR-a} + \text{SREBP-a} - \text{PUFA}$
- (8) $\text{SCD1} = \text{LXR-a} + \text{SREBP-a} - \text{PUFA} + \text{PPAR-a}$
- (9) $\text{SCAP} = -\text{PUFA}$
- (10) $\text{D5D} = \text{PPAR-a} + \text{SREBP-a} - \text{PUFA}$
- (11) $\text{D6D} = \text{PPAR-a} + \text{SREBP-a} - \text{PUFA}$

Résolution

- algèbre spéciale
- existant: résolution par règle de réécriture
- problème NP complet
- codage sur des corps finis simples ($\mathbb{Z}/p\mathbb{Z}$)
- système d'équations algébriques

Codage (1)

- Toute fonction $f : (\mathbb{Z}/p\mathbb{Z})^n \rightarrow \mathbb{Z}/p\mathbb{Z}$ avec n arguments $\mathbb{Z}/p\mathbb{Z}$ est une fonction polynomiale
- si \uplus représente l'opération $f \uplus g = f^{(p-1)} + g^{(p-1)}$, alors tout système d'équations $p_1(X) = 0, \dots, p_k(X) = 0$ a les mêmes solutions que l'équation unique $p_1 \uplus p_2 \uplus \dots \uplus p_k(X) = 0$.

Codage (2)

- $\mathbb{Z}/3\mathbb{Z}$
- $+$ → 1
- $-$ → -1
- $?$ → 0
- $X \oplus Y \rightarrow sq(X, Y) \stackrel{\text{def}}{=} -XY(X + Y)$
- $X \approx Y \rightarrow eq(X, Y) \stackrel{\text{def}}{=} XY(X - Y)$.

Représentation des fonctions polynomiales

Décomposition de Shannon

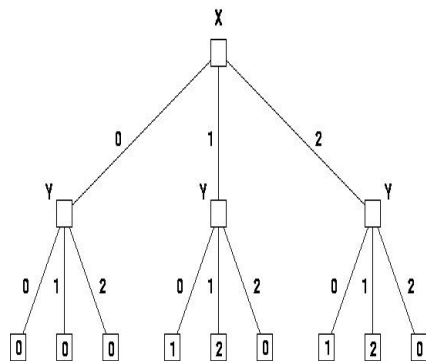
$$p(X_1, X) = (1 - X_1^2)p_{[X_1=0]}(X) + X_1(-X_1 - X_1^2)p_{[X_1=1]}(X) + X_1(X_1 - X_1^2)p_{[X_1=2]}(X)$$

Opérations binaires

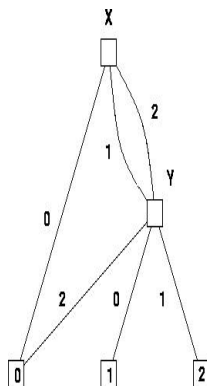
$$p^i(X_1, X) = (1 - X_1^2)p_0^i(X) + X_1(-X_1 - X_1^2)p_1^i(X) + X_1(X_1 - X_1^2)p_2^i(X)$$

$$p^1 \Delta p^2 = (1 - X_1^2)(p_0^1 \Delta p_0^2) + X_1(-X_1 - X_1^2)(p_1^1 \Delta p_1^2) + X_1(X_1 - X_1^2)(p_2^1 \Delta p_2^2)$$

Implémentation



Logiciel Sigali



Modèles et expérimentations (1)

Cohérence:

- codage du graphe d'interactions en système d'équations
- calcul de l'équation unique équivalente
- substitution des valeurs expérimentales
- existences de solutions pour le système résultant
- éventuellement résolution

Modèles et expérimentations (2)

Correction de valeurs erronées

- on corrige le minimum de valeurs
- distance de Hamming / espace des solutions
- projection de Hamming
- peut-être appliqué aux valeurs des arcs

Modèles et expérimentations (3)

Plan d'expérience

- experimentation - j variables observées
- calcul de la proportion de valeurs contraintes
- mesure de la qualité de l'expérience
- détermination des composantes dures

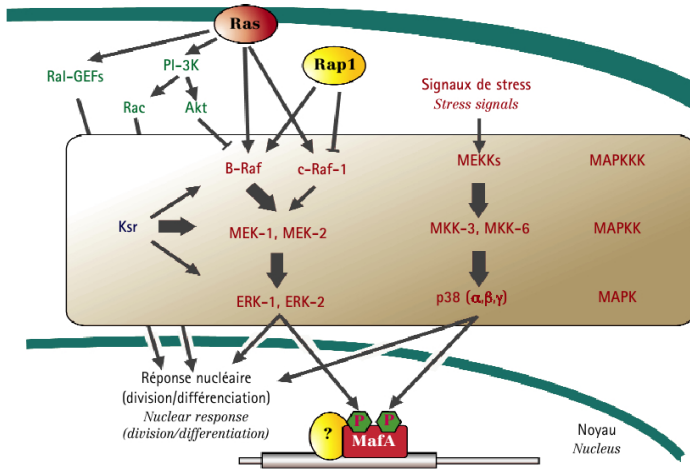
Et la dynamique?

- modèles à évènements discrets
- observations qualitatives
- ordonnancement d'évènements
- logique temporelle
- vérification: model checking

Systemes dynamiques: identification

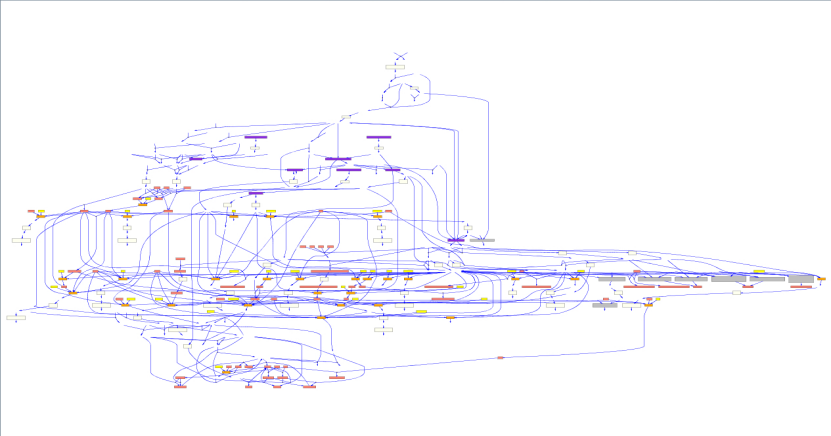
- modes de type Thomas
- paramétrisation des ordres des seuils
- paramétrisation des point focaux
- comportement spécifié par logique temporelle
- ensemble de paramètres compatibles
- problème "dual" du contrôle

Et les séquences dans tout cela?



Où sont branchés les spaghettis?

File Select Export Macro Analyses Paths



Commands
>>

Informations
Graphicvoid
Loading finished
Number of products: 337
Number of actions: 730

Navigation icons: back, forward, search, refresh, etc.

Comment obtenir le réseau de gènes?

- sélectionner les gènes d'intérêt (microarray)
- utiliser la partie régulatrice d'un gène
- trouver les sites de fixation des facteurs de transcription
- travail débutant ...

Conclusion

- pas de classe de modèles dominante
- modèles mathématiques classiques inadaptés
- modèles informatiques? (processus, agents, ...)
- ne pas dédaigner les modèles simples
- veille technologique

