

Petite Classe n° 6
Vendredi 21 Octobre 2016

Réseaux génétiques, comportements moyens et fluctuations

Le processus élémentaire d'expression génétique et de synthèse de protéines est dessiné en Fig. 1a. Pour établir un modèle mathématique de ce processus, nous introduisons les taux de réaction schématisés en Fig. 1b.

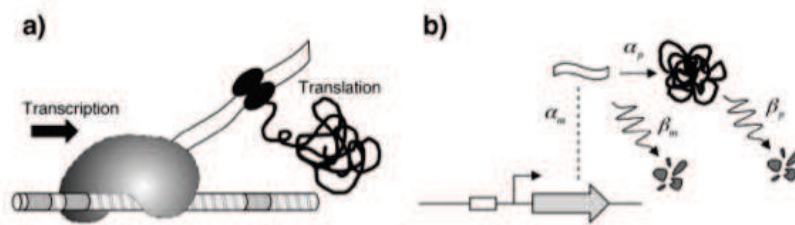


FIGURE 1 – Modèle simplifié de l'expression d'un gène. a) la polymérase transcrit l'ADN en ARN messenger, qui est ensuite traduit en protéine par le ribosome. b) Schéma représentant ce processus. On reconnaît le site promoteur et le gène sur l'ADN, l'ARN messenger et la protéine. Les taux α représentent la production de ces molécules, alors que les taux β représentent leur dégradation.

Les ordres de grandeurs des taux sont les suivants : synthèse de l'ARN messenger (α_m) = de 1 nM/min à 60 nM/min ; temps de vie moyen de l'ARN messenger ($1/\beta_m$) = de 1 à 10 min ; taux de traduction (α_p) = quelques protéines par ARN messenger/min ; temps de vie moyen d'une protéine ($1/\beta_p$) = temps de doublement de la bactérie (dû à la dilution, peut être accéléré en cas de protéolyse).

Dans un premier temps, nous allons modéliser le comportement moyen de ce processus et nous étudierons ensuite les fluctuations dues au bruit à l'échelle moléculaire. Nous verrons enfin comment ces modèles peuvent être étudiés numériquement avec l'algorithme de Gillespie.

1 Comportement moyen

1. Ecrire les équations différentielles auxquelles obéissent les concentrations d'ARN messenger $[m]$ et de protéines $[p]$. Calculer les valeurs des concentrations dans l'état stationnaire. A quel nombre de molécules ces concentrations correspondent-elles ?

Dans la suite, on va tenir compte du fait que le produit de l'expression (protéine) peut réguler cette expression. Pour cela, on considère que les protéines produites s'accrochent au promoteur avec une probabilité dépendant de leur concentration et modulent ainsi la production d'ARN messenger. On introduit la fonction d'activité du promoteur $g([p])$, dessinée en Fig. 2 pour les cas d'un répresseur (r) et d'un activateur (a).

2. On modélise g sous la forme suivante :

$$g_a([p]) = \frac{K_a^n + \omega[p]^n}{K_a^n + [p]^n} \quad g_r([p]) = \frac{K_r^n + (1/\omega)[p]^n}{K_r^n + [p]^n}$$

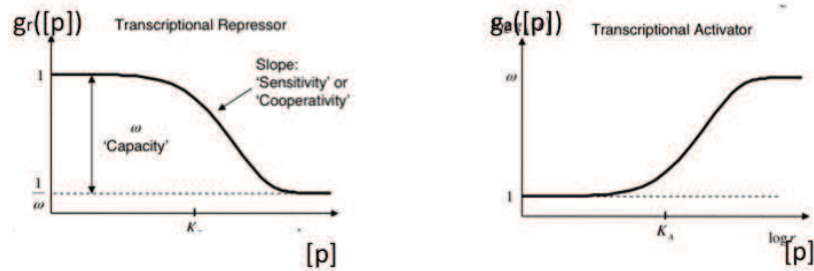


FIGURE 2 – Fonctions d'activité du promoteur.

A quels processus ces expressions correspondent-elles en pratique ? Quelles sont les interprétations de ω et n ? Montrer que, plus n est grand, plus la fonction d'activité varie abruptement avec la concentration.

- On considère pour commencer le cas très simple du réseau auto-activateur schématisé en Fig.3. Ecrire les équations auxquelles obéissent les concentrations d'ARN messenger $[m]$ et de protéines $[r]$. Quelle équation satisfait l'état stationnaire ? Déterminer son régime de stabilité.

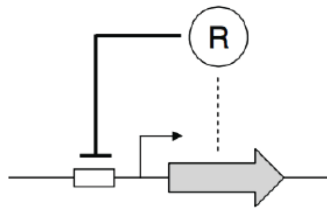


FIGURE 3 – Schéma d'un réseau auto-réresseur. La protéine produite se lie au promoteur et tend à empêcher l'expression du gène.

- On considère maintenant le réseau de la figure 4, qui couple deux gènes. Ecrire les équations différentielles auxquelles obéissent les concentrations d'ARN messagers $[m_1]$ et $[m_2]$ et des protéines correspondantes $[r_1]$ et $[r_2]$. Quels sont les états stationnaires possibles ? Pour quelles valeurs des paramètres l'appellation commune de ce réseau, interrupteur à bascule, est-elle justifiée ?

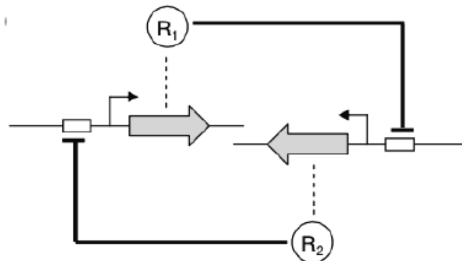


FIGURE 4 – Schéma d'un réseau dit d'interrupteur à bascule. Les protéines produites par chacun des deux gènes se lient au promoteur de l'autre gène et tendent à empêcher son expression.

2 Fluctuations

Le faible nombre de molécules en jeu fait que la description en terme de concentration moyenne n'est pas très bien justifiée. Il faut tenir compte des fluctuations autour de ces valeurs moyennes. Pour cela, nous allons écrire et résoudre des équations d'évolution de la probabilité du nombre de molécules à un temps donné, appelée équation maîtresse.

5. Considérons à nouveau le processus de la Fig. 1b. Soit $m = 0, 1, 2, \dots$ le nombre de molécules d'ARN messenger. Cette grandeur est stochastique, et nous appelons $P(m, t)$ la distribution de m au temps t . Montrer que P satisfait l'équation d'évolution :

$$\frac{dP(m, t)}{dt} = V\alpha_m P(m-1, t) + \beta_m(m+1)P(m+1, t) - (V\alpha_m + \beta_m m)P(m, t)$$

où V est le volume d'une bactérie.

6. Calculer la valeur moyenne $\langle m \rangle$ et la variance du nombre de molécules $\sigma_m = \langle m^2 \rangle - \langle m \rangle^2$ d'ARN messenger à l'état stationnaire. On pourra utilement introduire la fonction génératrice $F(t, z) = \sum_m P(m, t)z^m$. Que vaut le rapport $\sigma_m / \langle m \rangle$?
7. Etendre l'analyse de la question précédente à la production non régulée de protéines (Fig. 1b). On écrira en particulier l'équation satisfaite par la distribution $P(m, p, t)$ des nombres m d'ARN messagers et p de protéines au temps t . En déduire la valeur moyenne $\langle p \rangle$ et la variance $\sigma_p = \langle p^2 \rangle - \langle p \rangle^2$ du nombre de protéines produites en régime stationnaire. Comment se comparent les fluctuations du nombre de protéines par rapport à celles du nombre d'ARN messagers ?
La figure 5 montre des résultats expérimentaux obtenus par le groupe de S. Xie à Harvard. Des techniques de fluorescence sur molécule unique permettent de mesurer le nombre de protéines produites en fonction du temps.
8. Ces mesures expérimentales vous semblent-elles compatibles avec les résultats précédents ?

3 Simulations : l'algorithme de Gillespie

Lorsque l'expression est régulée, il peut devenir très difficile de résoudre l'équation satisfaite par la fonction génératrice F . On utilise souvent un algorithme de simulation pour comprendre le fonctionnement du système. La méthode la plus utilisée est celle de Gillespie. Reprenons le réseau de la figure 4. On simplifie son fonctionnement en supposant qu'il peut se décrire comme quatre processus stochastiques parallèles :

$$r_1 \rightarrow r_1 - 1 \quad \text{avec le taux : } \nu_1 = \beta_r r_1 \quad (1)$$

$$r_2 \rightarrow r_2 - 1 \quad \text{avec le taux : } \nu_2 = \beta_r r_2 \quad (2)$$

$$r_1 \rightarrow r_1 + b \quad \text{avec le taux : } \nu_3 = \alpha_m g_r(r_2) \quad (3)$$

$$r_2 \rightarrow r_2 + b \quad \text{avec le taux : } \nu_4 = \alpha_m g_r(r_1) \quad (4)$$

Le paramètre $b = \alpha_r / \beta_m$ est ici la taille du burst, que l'on suppose constante. Dans l'algorithme de Gillespie, on part de nombres r_1 et r_2 de protéines au temps $t = 0$. Puis on attend un temps τ , distribué aléatoirement suivant la loi $D(\tau)$ et on choisit une des quatre réactions ci-dessus selon les probabilités q_1, q_2, q_3 et q_4 . On passe alors du temps t au temps $t + \tau$, on met à jour les nombres r_1 et r_2 selon la réaction choisie, puis on itère la procédure.

9. Comment faut-il choisir $D(\tau)$ et les q_i pour que cet algorithme soit correct ?
10. La figure 6 montre le résultat de deux simulations de Gillespie, voir légende pour les valeurs des paramètres. Commenter les résultats de ces simulations.

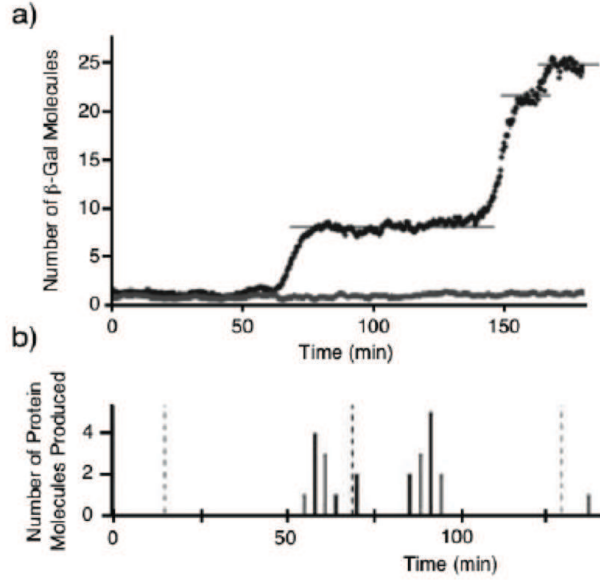


FIGURE 5 – Suivi de la production de protéine en temps réel par imagerie. Les protéines produites sont fusionnées avec des protéines fluorescentes et peuvent être comptées. a) Nombre de beta-galactosidases produites par une bactérie *E. Coli* en fonction du temps avec un lac opéron sous forte répression indépendante du temps. (Extrait de : *Nature* 440, p358 (2006)). b) Nombres de Tsr-Venus (Tsr est une protéine membranaire, Venus est une protéine fluorescente émettant dans le jaune, toutes deux sont fusionnées et insérées dans un lac opéron) produites au cours du temps. Les lignes pointillées verticales représentent les divisions cellulaires, et on suit une lignée de descendants d'une *E. Coli* initiale. Le nombre moyen de protéines produites dans un burst est proche de 5, la distribution est exponentielle. Le nombre moyen de bursts par cycle cellulaire est 1.2. Extrait de *Science* 311, 1600 (2006).

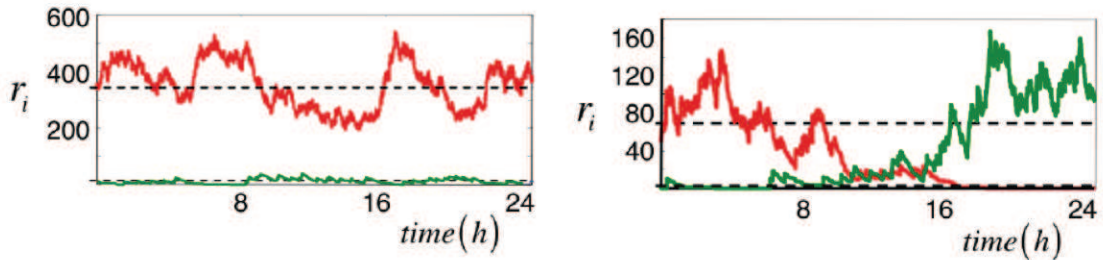


FIGURE 6 – Simulations de Gillespie du système (3), extrait de M. Scott, *Tutorial : genetic circuits and noise* (2006). Les valeurs des paramètres sont pour la simulation de gauche : $b = 10$, $\alpha_m = 1$ nM/min, $1/\beta_r = 50$ min, $K_R = 25$ nM, $\omega = 200$ avec $q_R = [r]/(2K_R)$. Dans la simulation de droite, les paramètres sont les mêmes à l'exception de α_m et K_r qui sont 5 fois plus petits. Le point de départ est l'équilibre stable $(r_1, r_2) = (11, 341)$ à gauche et $(r_1, r_2) = (11/5, 341/5)$ à droite.