

Simulation multi-agent de réseaux génétiques : les rythmes circadiens d'*Ostreococcus tauri*

Sébastien Picault¹
picault@lifl.fr

Florence Corellou²

Christian Schwartz²

François-Yves Bouget²

francois-yves.bouget@obs-banyuls.fr

¹Laboratoire d'Informatique Fondamentale de Lille
UMR CNRS 8022
Université des Sciences et Technologies de Lille
59655 Villeneuve d'Ascq Cedex – FRANCE

²Laboratoire Modèles en Biologie Cellulaire et Evolutive
UMR CNRS-Paris 6 7628
Observatoire Océanologique de Banyuls-sur-Mer
BP 44, 66651 Banyuls-sur-Mer Cedex – FRANCE

Résumé

Nous décrivons ici des travaux sur la modélisation des rythmes biologiques circadiens (i.e. ~ 24 h) observés chez une algue verte unicellulaire, *Ostreococcus tauri*. Ces rythmes sont produits par des processus de régulation de la transcription génétique. Nous présentons plusieurs modèles issus d'un groupe de travail pluridisciplinaire et qui ont donné lieu à des simulations par agents. Ils visent à identifier les acteurs moléculaires de l'horloge circadienne, en confrontant hypothèses biologiques, résultats de simulation et mesures expérimentales.

Mots-clés : Simulation multi-agent, Modèles de comportement, Biologie cellulaire

Abstract

We report here our research on the modelling of biological circadian rhythms (i.e. ~ 24 h) in a unicellular green alga, *Ostreococcus tauri*. Such rhythms are the result of the regulation of genetic transcription. We present several models designed by physicians and computer scientists, in order to identify molecular actors of the circadian clock. We therefore compare biological assumptions, simulation results and experimental data.

Keywords: Multi-agent simulation, Models of behavior, Cell biology

1 Introduction

Les travaux que nous présentons ici s'inscrivent dans un projet de modélisation des rythmes biologiques *circadiens* ; nous traiterons plus spécifiquement de l'utilisation de systèmes multi-agents pour simuler ces phénomènes, en montrant comment cette approche se positionne par rapport à d'autres méthodes de modélisation.

Les êtres vivants sont sujets à des variations rythmiques de leur activité. La plupart d'entre eux disposent notamment de rythmes appelés *circadiens* (i.e. dont la période avoisine 24 h). Ils sont *endogènes*, c'est-à-dire produits par des processus biologiques propres à l'organisme considéré, et peuvent se *synchroniser* régulièrement avec les variations environnementales (alternance jour/nuit principalement). Enfin, une horloge circadienne possède une « signature », sa *courbe de réponse de phase*, qui peut être obtenue expérimentalement en mesurant la réponse de l'horloge à un décalage horaire.

Les rythmes circadiens résultent généralement de mécanismes de régulation génétique, aussi bien chez les plantes que chez les animaux [Harmer et al., 2001], et interagissent souvent avec le cycle de division cellulaire (c'est-à-dire que les cellules ne peuvent se diviser qu'à certaines heures de la journée). La compréhension des liens entre horloge circadienne et division cellulaire comporte des enjeux médicaux notamment pour le traitement de tumeurs cancéreuses [Filipski et al., 2005].

Ces réseaux de régulation sont assez difficiles à étudier, en raison du grand nombre d'acteurs moléculaires susceptibles de jouer un rôle dans les mécanismes de l'horloge. Nous nous intéressons ici à une algue verte unicellulaire, *Ostreococcus tauri* [Courties et al., 1994], qui est le plus petit eucaryote (cellule à noyau) connu. Il présente donc des caractéristiques communes avec les êtres pluricellulaires (animaux et plantes supérieurs), tout en restant relativement simple avec un génome très compact [Derelle et al., 2006] (donc relativement moins d'acteurs moléculaires). Il est doté d'un rythme circadien dont la période naturelle est d'environ 25 h et son cycle de division est sous contrôle de

l'horloge circadienne. Il constitue donc un modèle biologique particulièrement prometteur.

L'objectif des travaux, dont nous présentons ici une phase préliminaire, consiste à proposer une modélisation et des outils de simulation afin :

- d'examiner, en termes de systèmes dynamiques, la *plausibilité des hypothèses biologiques* concernant la structure de l'horloge moléculaire ;
- de *comparer plusieurs modèles* et d'identifier des acteurs moléculaires susceptibles d'intervenir dans l'horloge circadienne ;
- d'étudier diverses hypothèses relatives aux voies d'entrée des signaux lumineux ;
- de suggérer aux biologistes, sur la base du comportement des systèmes simulés, des expériences complémentaires pour conforter ou infirmer leurs hypothèses ;
- dans une phase ultérieure, étendre l'étude au contrôle de la division cellulaire par l'horloge.

Ces recherches sont menées au sein d'un groupe de travail pluridisciplinaire sur la modélisation des horloges circadiennes, impliquant diverses équipes de Lille I : le Laboratoire de Physique des Lasers, Atomes, Molécules ; l'Unité de Glycobiologie Structurale et Fonctionnelle ; le Laboratoire d'Informatique Fondamentale de Lille. Les compétences représentées vont ainsi de la dynamique non-linéaire à la simulation multi-agent, en passant par le calcul formel. Ces diverses approches permettent de soumettre les modèles élaborés en collaboration avec l'Observatoire Océanologique de Banyuls à divers critères de validation.

Nous allons d'abord indiquer la méthode suivie pour la conception des modèles et leur validation ; nous détaillerons ensuite les deux modèles préliminaires d'horloge retenus et leur traduction multi-agent ; enfin nous présenterons et discuterons les résultats des simulations distribuées ainsi que les modèles en cours d'étude.

2 Contraintes et méthodologie

Les modèles multi-agents qui peuvent être élaborés pour rendre compte de l'horloge circadienne d'*O. tauri* sont soumis à des contraintes fortes qui déterminent des choix de conception spécifiques.

2.1 Contraintes sur les données

Les données expérimentales relatives aux rythmes circadiens d'*O. tauri* sont d'abord

longues et coûteuses à acquérir. Par ailleurs, toutes les données concernent des populations de cellules : il est impossible de savoir ce qui se passe au sein d'une cellule unique. En effet, les modèles biologiques font tous implicitement l'hypothèse que les millions ou milliards d'individus d'un milieu de culture sont interchangeables : ils sont issus d'une même souche donc génétiquement identiques, et soumis aux mêmes conditions environnementales. Pourtant, lorsqu'une oscillation disparaît ou change de période et d'amplitude, on ne peut pas déterminer absolument si c'est parce qu'il n'y a plus d'oscillations dans les cellules, ou si celles-ci ne font que se désynchroniser !

Une autre difficulté réside dans les fortes incertitudes de mesure (cf. fig. 3) et la variabilité quantitative d'une expérience à une autre ; toutes les mesures sont en outre normalisées en amplitude donc la comparaison entre l'expression de deux gènes est surtout qualitative, et il est difficile de détecter des oscillations de faible amplitude. Toutefois, les expériences donnent des résultats qualitatifs nettement reproductibles, en termes d'acuité des pics, de déphasage entre courbes, etc. : c'est donc ces caractéristiques que l'on peut s'attacher à reproduire par simulation.

Enfin, dans tous les cas les mesures sont limitées dans le temps : il faut environ deux à trois jours d'entraînement pour synchroniser la population en alternance jour/nuit, mais au total on ne peut dépasser une semaine de culture, car les cellules sont alors trop nombreuses dans le milieu.

2.2 Contraintes sur les modèles

Les modèles que nous cherchons à élaborer doivent en tout premier lieu présenter une plausibilité biologique suffisante, c'est-à-dire faire intervenir des acteurs supposés de l'horloge, et utiliser des paramètres (et rapport de paramètres) conformes aux données de la littérature lorsqu'elles sont connues, notamment pour les échelles de temps.

Nous avons veillé tout particulièrement à la parcimonie des hypothèses, en examinant d'abord les modèles les plus simples (le moins possible d'acteurs, d'interaction entre eux), en n'introduisant de nouvel acteur ou de nouvelle relation que pour obtenir des résultats plus proches, sur plusieurs critères, des données.

Par ailleurs, spécialement dans le cas de la modélisation par agents, nous tenons à ce que les modèles possèdent un pouvoir explicatif : on

doit être en mesure de comprendre la dynamique à l'œuvre dans le système, le rôle joué par chaque acteur, etc.

Quant au caractère prédictif de nos modèles, ils visent non pas tant les aspects quantitatifs du phénomène étudié qu'à une fidélité qualitative, comme la reproduction des profils des courbes (pics, déphasages), ainsi qu'à suggérer des expériences complémentaires.

2.3 Contraintes sur les résultats

En l'absence de données biologiques précises quant aux paramètres requis par le modèle pour *O. tauri* (taux de transcription, demi-vie des ARNm et protéines, etc.), deux voies sont envisageables. La première consiste à chercher à identifier ces paramètres par diverses méthodes d'apprentissage, en comparant les courbes simulées pour divers jeux de paramètres aux courbes expérimentales. Cela pourrait être envisagé dans les modèles équationnels, sans garantie toutefois d'obtenir des paramètres biologiquement plausibles, surtout à haut niveau de bruit.

Dans le cas des simulations distribuées que nous avons réalisées, identifier des paramètres n'a aucun sens puisque les échelles (d'espace, de nombre d'agents, de dimension, etc.) ne sont pas respectées. En revanche, puisque l'on s'attache aux propriétés dynamiques des circuits de régulation génétique, on doit veiller au respect de certains critères liés à l'évolution temporelle du système :

- paramètres restant dans l'ordre de grandeur des paramètres biologiques connus dans la littérature (p.ex. rapport entre taux de transcription libre et taux réprimé)
- le déphasage entre les courbes d'ARNm et de protéines ;
- le déphasage entre les courbes d'expression des acteurs du modèle à 2 gènes (déphasage entre l'ARNm de *A* et celui de *B*, entre la protéine *A* et la protéine *B*) ;
- la possibilité d'entraîner le système à une fréquence légèrement inférieure à sa fréquence naturelle ;
- la possibilité pour le système de continuer à osciller en lumière ou en obscurité continues ;
- l'aspect de la courbe de réponse de phase.

3 Recherche de modèles minimaux

L'étape suivante consiste à rechercher les modèles d'horloges les plus simples possibles : à

un seul gène ou deux. Les modèles présentés ici ont fait l'objet d'une étude approfondie au sein du groupe de modélisation pluridisciplinaire de Lille.

3.1 Étude préliminaire : modèle à 1 gène

Le modèle de circuit génétique le plus simple qui puisse être envisagé est celui d'un gène *auto-réprimé* (cf. fig. 1) : les protéines issues de la transcription puis de la traduction du gène *G* peuvent inhiber de façon plus ou moins forte la transcription de *G*. Il s'agit d'une boucle de rétroaction négative simple dont on peut s'attendre à ce qu'elle produise des oscillations : en effet, l'ARNm et les protéines subissent une dégradation au cours du temps, donc la probabilité que le gène repasse dans l'état libre augmente au cours du temps, et la transcription peut redémarrer.

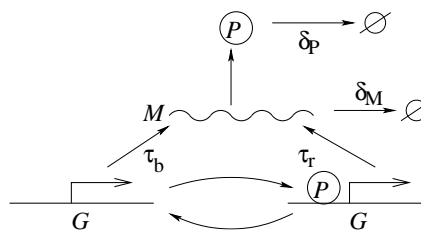


FIG. 1 – Modèle le plus simple : gène réprimé par sa propre protéine. Le gène *G* à l'état non régulé (à gauche) est transcrit sous forme d'ARNm (*M*) à un taux supposé constant au cours du temps (τ_b). *M* est à son tour traduit en protéine (*P*). *M* et *P* subissent une dégradation exponentielle au cours du temps (resp. δ_M et δ_P). La protéine *P* peut se fixer sur un site régulateur de *G* et réprimer sa transcription (à droite) : *G* est alors transcrit à un taux $\tau_r \ll \tau_b$.

On peut généraliser ce modèle en tenant compte de la coopérativité de la régulation, c'est-à-dire du fait que le répresseur de *G* peut être un multimère de la protéine. Il faut alors que n protéines *P* s'assemblent avant de pouvoir se fixer sur *G*. L'hypothèse adoptée dans la plupart des modèles de ce genre est que la dégradation de l'ARNm et des protéines suit une loi exponentielle (le taux de dégradation est constant au cours du temps : $\frac{dM}{dt} = -\delta_M M$). Or, l'étude menée par [Boulier et al., 2007], démontre analytiquement que le système ne présente pas de bifurcation de Hopf pour $n < 9$; les simulations numériques menées pour de nombreux jeux de paramètres confirment l'absence d'oscillations pour $n < 9$. Un degré de coopérativité aussi élevé constitue une contrainte biologique assez peu plausible et il faut donc envisager des mécanismes supplémentaires pour un système os-

cillant.

Or, ces mécanismes additionnels susceptibles de conduire à des oscillations ne manquent pas : délais dans les équations [Lewis, 2003], transport de protéines entre cytoplasme et noyau [Leloup et al., 1999], modifications fonctionnelles des protéines [Goldbeter, 1995] ou dégradation des protéines par une enzyme [Goldbeter, 1996]. En l'absence d'indications biologiques précises, il semble plus raisonnable de se focaliser sur la cinétique de dégradation des protéines, qui est un phénomène bien connu [Goldbeter, 1996]. Les simulations numériques de [Morant et al., 2007] montrent d'ailleurs qu'il suffit de faibles contraintes sur la dégradation de l'ARNm et la protéine pour obtenir des oscillations.

Ces contraintes sont réalisées notamment par un processus de dégradation enzymatique représenté par la réaction suivante :



dans lequel la protéine P forme avec une enzyme (protéase) E un complexe C plus ou moins stable, qui dégrade la protéine en libérant l'enzyme. En situation quasi-stationnaire, la dégradation des protéines suit alors la dynamique dite de Michaelis-Menten : $\frac{dP}{dt} = -\frac{V_{\max}P}{K+P}$ (où V_{\max} et K s'expriment en fonction des constantes cinétiques de la réaction). L'intérêt de ce mécanisme est de produire un effet de *saturation* lorsque la quantité de protéine à dégrader est très supérieure au seuil K , car on a alors $\dot{P} \sim -V_{\max}$ (dégradation linéaire, donc lente), alors que pour une quantité de protéines assez inférieures à K , on a au contraire $\dot{P} \sim -\frac{V_{\max}}{K}P$ (dégradation exponentielle, donc rapide) : autrement dit, ce processus tend à instaurer un niveau haut de protéine (avec dégradation relativement faible) et un niveau bas (avec dégradation relativement forte).

3.2 Modèle à 2 gènes

Si le circuit à 1 gène auto-réprimé offre l'avantage de la parcimonie, il est en revanche probablement moins robuste (aux mutations, au bruit, aux changements environnementaux...); de fait, plusieurs horloges circadiennes de plantes semblent plutôt mettre en œuvre deux gènes ou plus. En l'occurrence, on peut s'appuyer sur les homologies connues entre les gènes d'*O. tauri* et ceux d'autres espèces végétales pour rechercher des gènes candidats dans le fonctionnement de l'horloge circadienne.

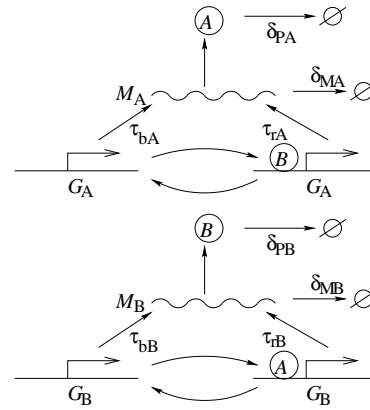


FIG. 2 – Modèle à deux gènes, avec régulation croisée : les protéines A et B , produites respectivement par transcription et traduction des gènes G_A et G_B , régulent la transcription respective de B et A . Il peut s'agir d'une activation ou d'une répression. Dans un modèle d'horloge inspiré d'*A. thaliana*, on a $\tau_{rA} \ll \tau_{bA}$ (répression de A par B) et $\tau_{rB} \gg \tau_{bB}$ (activation de B par A).

Dans le modèle d'étude des végétaux supérieurs (*Arabidopsis thaliana*), deux gènes centraux de l'horloge [Salome and McClung, 2004, Locke et al., 2005] ont des homologues chez *O. tauri* : ce sont TOC1 et CCA1, qui selon l'équipe de l'Observatoire océanologique de Banyuls sont de bons candidats. Le mécanisme présumé est le suivant : la protéine TOC1 active la transcription du gène CCA1, et la protéine CCA1 réprime la transcription de TOC1 (cf. fig. 2 en posant $A = \text{TOC1}$ et $B = \text{CCA1}$). Les résultats expérimentaux obtenus à l'Observatoire montrent des oscillations sur ces gènes au niveau des transcrits et des protéines, aussi bien en alternance jour/nuit qu'en lumière ou en nuit continues (cf. fig. 3).

3.3 Action de la lumière

Une fois la forme générale du modèle choisie (acteurs moléculaires et régulations), la première étape consiste lors des simulations (numériques ou distribuées) à s'assurer qu'il existe des jeux de paramètres biologiquement plausibles permettant d'obtenir des oscillations avec les propriétés adéquates.

Pour aller plus loin, il faut élaborer ensuite des hypothèses sur la façon dont la lumière agit sur le système : sur quel(s) acteur(s) moléculaire(s) a-t-elle un effet (ARNm, protéine, intermédiaires...) et lequel (dégradation, modification de la fonction biologique, changement de compartiment cellulaire...)?

Dans cet article, nous présentons les résultats

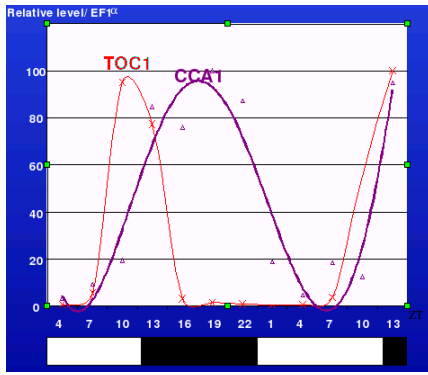


FIG. 3 – Données expérimentales du groupe Horloge circadienne et cycle cellulaire de l’Observatoire océanologique de Banyuls, donnant les niveaux d’expression (quantité normalisée d’ARNm) de deux gènes d’*O. tauri* (TOC1 et CCA1) en fonction du temps circadien (ZT), mesurés en alternance jour/nuit (bandes blanches et noires). L’instant initial ZT 0 correspond au début de la période d’éclairage. On observe un pic étroit de TOC1 juste avant la fin du jour (vers ZT 10), suivi d’un long passage à zéro entre ZT 17 et ZT 7 ; l’expression de CCA1 est plus étalée dans le temps, avec un maximum au cours de la nuit vers ZT 17 et un minimum vers ZT 7.

obtenus sous une hypothèse résolument simpliste : comme la protéine TOC1 n’apparaît qu’à la tombée de la nuit (quelques heures après le pic d’ARNm TOC1, cf. fig. 3) pour diminuer fortement avec le début de l’éclairage, nous supposons que *la lumière agit directement sur la protéine TOC1 en la dégradant*. Les limitations de cette hypothèse sont discutées au § 6. Dans toute la suite, on supposera donc que la lumière dégrade soit la protéine *P* (modèle à 1 gène), soit la protéine *A* (modèle à 2 gènes).

4 Agentification et expériences

À partir des modèles à 1 ou 2 gènes proposés pour rendre compte des oscillations d’*O. tauri*, nous avons élaboré et simulé des modèles multi-agents. Ceux-ci n’ont vocation ni à la prédiction quantitative, ni à l’identification de paramètres biologiques : il s’agit d’abord de tester la plausibilité de certaines hypothèses biologiques, relatives d’une part aux acteurs moléculaires impliqués dans l’oscillateur circadien et aux relations qu’ils entretiennent, d’autre part aux voies d’entrée de l’horloge (entraînement par la lumière). Nous avons donc d’emblée décidé de nous placer dans une situation résolument artificielle (2D, nombre d’agents impliqués, etc.) en faisant l’hypothèse (forte) qu’on peut passer par des changements d’échelle et de dimension de la situation *in vivo* à la simulation. Ce que nous attendons de la simulation concerne donc

la dynamique globale du système en fonction des acteurs et de leurs interactions ainsi que ses réponses aux contraintes environnementales et aux perturbations endogènes.

4.1 L’environnement

Nous avons choisi de ne simuler qu’une cellule : les modèles biologiques sont en effet toujours relatifs au comportement d’un individu, ses clones étant supposés réagir identiquement dans les mêmes conditions environnementales.

De plus, la plupart des expérimentations ont été réalisées en 2D, sur une surface torique (pour s’affranchir des problèmes de frontière) : nous avons simplement vérifié que les comportements en 3D étaient qualitativement identiques (moyennant bien sûr de changer des paramètres liés au passage surface/volume, tels que les coefficients de diffusion, et de conserver la même densité d’agents).

En l’état actuel des connaissances, nous n’avons pas jugé utile de différencier les comportements des agents selon qu’ils sont dans le noyau ou dans le cytoplasme. Toutefois, dans une véritable cellule, c’est surtout à la périphérie du noyau que se concentrent les ribosomes, organites chargés de la traduction des ARNm en protéines : nous avons donc respecté cette importante hétérogénéité spatiale en délimitant une zone centrale qui joue le rôle de noyau et détermine la position des ribosomes.

4.2 Les agents

Tous les agents mis en œuvre dans nos simulations sont réactifs. La simulation est à temps discret, chaque agent est susceptible d’effectuer une action élémentaire à chaque cycle.

Nous avons distingué les agents suivants (un aperçu de l’environnement et des agents d’une simulation à 1 gène est donné sur la fig. 4) :

- Les *gènes* codent des protéines. Un gène peut être dans l’état libre, activé ou réprimé, chacun étant caractérisé par un *taux de transcription* qui donne la probabilité pour ce gène, à chaque cycle, d’être transcrit en ARNm (p. ex. pour G_A : $\tau_{bA} = 0,9$ et $\tau_{rA} = 10^{-3}$).
- Les *ARNm* sont issus de la transcription d’un gène. Ils diffusent à l’intérieur de la cellule et subissent une décroissance exponentielle : autrement dit, à chaque cycle un agent ARNm a une probabilité constante $\delta_{M/MA/MB}$ de dis-

paraître. Enfin, ils peuvent être traduits en protéines par des ribosomes.

- Les *ribosomes* ont pour fonction d’effectuer la traduction d’un ARNm en la protéine correspondante. Ils sont placés à la périphérie du noyau.
- Les *protéines* sont issues de la traduction des ARNm par les ribosomes. Elles diffusent à l’intérieur de la cellule et peuvent soit subir une décroissance exponentielle (probabilité constante $\delta_{P/A/B}$ au cours du temps), soit être dégradées par des enzymes (autres protéines) spécifiques appelées *protéases*. Une protéine peut jouer un rôle régulateur et activer ou inhiber la transcription d’un gène cible.
- Les *protéases* sont des enzymes qui ne font que diffuser dans la cellule et dégrader les protéines qu’elles rencontrent (selon le processus de Michaelis-Menten mentionné ci-dessus). Elles ne sont donc présentes que dans certaines simulations.

Une fois définis ces agents et leurs primitives comportementales, la définition d’un modèle de simulation consiste à déterminer les interactions que ces agents peuvent effectuer les uns avec les autres, puis évaluer les résultats obtenus avec divers jeux de paramètres.

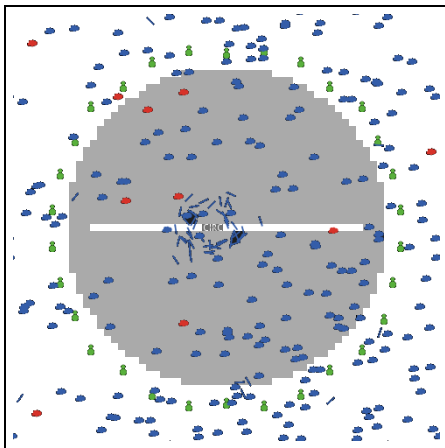


FIG. 4 – Vue du modèle de simulation distribuée, ici pour un modèle à 1 gène. **Environnement** : une seule cellule qu’on suppose fermée ; on est dans cet exemple en 2D sans bords (tore) ; le noyau, en grisé, définit la position des ribosomes. **Agents** : Au centre, un brin d’ADN contenant un agent *gène* (CIRC) : il produit des ARNm (bâtonnets) qui diffusent dans la cellule. Lorsqu’ils passent à proximité de *ribosomes* (en vert à la périphérie du noyau), ceux-ci les traduisent en *protéines* (en bleu). Ces protéines peuvent diffuser, se fixer près du gène pour le réprimer, ou être détruits par les *protéases* (en rouge).

4.3 Les comportements

Les comportements des agents sont exprimés sous forme d’une matrice d’interactions (comme préconisé dans la méthode IODA [Mathieu et al., 2007]) : schématiquement, une interaction est un comportement exprimé indépendamment d’une catégorie particulière d’agent, et qui peut être effectué par un agent, appelé source, sur un autre agent, appelé cible. Les comportements qui agissent sur l’environnement (déplacement) ou sur l’agent lui-même (disparition spontanée) sont représentés par des interactions sans cible définie.

sources \ cibles	\emptyset	G	ARNm	P
G	transcr.			
ARNm	die (1) diffuse (0)			
ribosome			transl.	
P	diffuse (0)	repress (1)		
Δ	diffuse (0)			delete (1)

TAB. 1 – Matrice d’interactions d’un modèle à 1 gène G dont la protéine P réprime la transcription. Dans la situation présentée ici, les ARNm subissent une décroissance exponentielle (interaction die), tandis que les protéines P sont dégradées par des protéases (agents Δ). Les entiers éventuels indiquent la priorité relative des interactions du point de vue de la source (une protéine P effectue repress plutôt que diffuse si les deux sont réalisables).

sources \ cibles	\emptyset	G_A	G_B	ARNm $_{A/B}$
$G_{A/B}$	transcr.			
ARNm $_{A/B}$	die (1) diffuse (0)			
ribosome				transl.
A	die (2) diffuse (0)		activate (1)	
B	die (2) diffuse (0)	repress (1)		

TAB. 2 – Matrice d’interactions d’un modèle à deux gènes G_A et G_B , la protéine A activant la transcription du gène G_B et la protéine B réprimant la transcription du gène G_A . Ici, les ARNm et les protéines subissent une décroissance exponentielle.

Les tableaux 1 et 2 donnent les matrices d’interactions respectives des modèles à 1 gène auto-réprimé avec dégradation enzymatique de P et un modèle à 2 gènes avec décroissance exponentielle de A et B . Ces matrices font appel aux interactions suivantes (réalisables lorsque les agents sont en contact) :

die décroissance exponentielle : l’agent peut disparaître spontanément avec une probabilité constante dans le temps

delete dégradation d'un agent par un autre
diffuse diffusion d'un agent dans la cellule (déplacement aléatoire)
transcr. transcription d'un gène : le gène qui effectue cette interaction peut produire un ARNm avec une probabilité qui dépend de son état (libre, activé ou réprimé)
transl. traduction par un ribosome d'un ARNm en la protéine correspondante
repress répression d'un gène par un facteur de transcription (en général une protéine) : l'agent répresseur se fixe sur le gène et diminue fortement sa probabilité d'être transcrit
active activation d'un gène par un facteur de transcription : l'agent activateur se fixe sur le gène et augmente fortement sa probabilité d'être transcrit
forward déplacement en ligne droite d'un bord à l'autre de l'environnement

Les simulations ont été ensuite codées en Netlogo 3.1, d'accès relativement facile à des non-informaticiens.

4.4 Rôle de la lumière

Le comportements décrits ci-dessus décrivent le fonctionnement d'une horloge génétique en régime libre (sans signal lumineux). Pour étudier l'hypothèse d'influence directe de la lumière sur la protéine P (modèle à 1 gène) ou sur la protéine A (modèle à 2 gènes), il suffit de définir des agents *photons* (notés φ) qui sont créés par l'environnement pendant toute la durée des périodes d'éclairement, et, pour représenter leur comportement, d'ajouter aux matrices d'interaction de ces modèles (cf. resp. tab. 1 et 2) les lignes suivantes respectivement (cf. tab. 3) :

Modèle à 1 gène	cibles	\emptyset	P
	sources	φ	forward delete
Modèle à 2 gènes	cibles	\emptyset	A
	sources	φ	forward delete

TAB. 3 – Modifications à apporter aux matrices d'interactions des modèles à 1 et 2 gènes respectivement (tab. 1 et 2) pour prendre en compte l'action de la lumière (dégradation directe de la protéine P/A)

5 Analyse et étape suivante

En régime libre, *O. tauri* présente des oscillations d'une période voisine de 25 h. La dé-

marche générale que nous avons appliquée pour ces simulations est donc la suivante :

1. déterminer des jeux de paramètres (biologiquement plausibles) donnant lieu à des oscillations ;
2. calculer la période T^* (en nombre de cycles de simulation) de ces oscillations en régime libre ;
3. introduire un signal lumineux de période T légèrement inférieure à T^* ; T représente par construction la durée d'une journée (24 h) et permet d'établir la durée que représente un cycle de simulation ;
4. vérifier que le système oscille en alternance jour/nuit, en jour continu et en nuit continue ;
5. vérifier le déphasage entre les courbes d'ARNm et de protéine d'un même produit (entre 2 et 5 h) ;
6. dans le modèle à 2 gènes, vérifier les déphasages entre A et B , ainsi que l'heure des pics ;
7. si possible établir une courbe de réponse de phase.

Les résultats que nous présentons ci-dessous sont encore à un stade préliminaire. En particulier, l'hypothèse faite sur le mode d'action de la lumière entraîne des résultats moins bon en nuit continue ; de plus nous avons laissé de côté la dernière étape pour la réserver à des modèles plus détaillés (cf. §6).

5.1 Le modèle à 1 gène

Le modèle à 1 gène avec dégradation enzymatique des protéines doit osciller pour certaines valeurs de paramètres (d'après l'étude mathématique de [Morant et al., 2007]). Toutefois, de très nombreux jeux de paramètres donnent des courbes du type de celle de la fig. 5 : on observe une succession de pics (plus nettement marqués pour l'ARNm) qui ne sont pas parfaitement réguliers. Ces « pseudo-oscillations » tiennent à l'aspect spatial du phénomène, absent des modèles équationnels classiques, qui s'apparente dans une certaine mesure aux équations à délais [Lewis, 2003] qui prévoient des oscillations. Le manque de régularité des pics, en revanche, est lié au caractère stochastique de la fixation d'une protéine sur l'ADN ou de sa libération, et donc à l'initiation ou à l'arrêt de la transcription du gène.

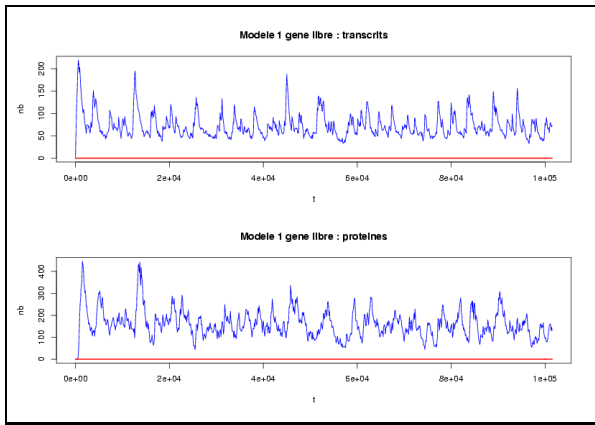


FIG. 5 – Résultats typiques de la simulation à 1 gène : en haut, courbe de l'ARNm ; en bas, celle des protéines. Axe des abscisses : nombre de cycles de simulation (1 cycle = 1 action élémentaire pour tous les agents).

Ces résultats prennent tout leur intérêt si l'on se penche sur ce qui se passe lorsque l'on entraîne de tels systèmes par un signal lumineux. Avec le même jeu de paramètres, et en donnant un signal lumineux carré dont la période est la moyenne de l'écart entre les pics de la fig. 5, on obtient des pics très réguliers. Mais surtout, lorsque l'on entraîne la cellule simulée en alternance jour/nuit suivie d'une phase de lumière continue (fig. 6) comme dans les conditions de laboratoire, les oscillations persistent bien plus nettement qu'en régime libre.

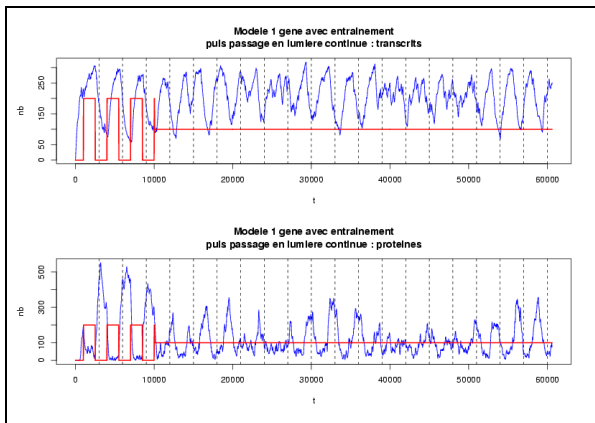


FIG. 6 – Résultats obtenus (ARNm/protéines) avec l'introduction d'un signal lumineux carré puis d'un signal lumineux continu.

5.2 Le modèle à 2 gènes

Le modèle suivant dans l'ordre de complexité est le modèle à 2 gènes, avec décroissance exponentielle des protéines. De nombreux jeux de paramètres, en particulier avec des taux de transcription et de dégradation élevés,

conduisent à des oscillations en régime libre. Ces oscillations peuvent être forcées en alternance jour/nuit, et l'on obtient également une nette persistance des oscillations lors du passage en lumière continue (cf. fig. 7).

Ces résultats sont limités une fois encore par l'hypothèse d'action directe de la lumière : en nuit continue, les oscillations persistent plus difficilement. De plus on peut constater que le pic d'ARNm de TOC1 est légèrement en retard par rapport à ce que l'on peut en attendre au vu des données expérimentales (cf. fig. 3). En revanche les déphasages entre les quatre courbes sont proches de ce que l'on peut observer *in vivo*.

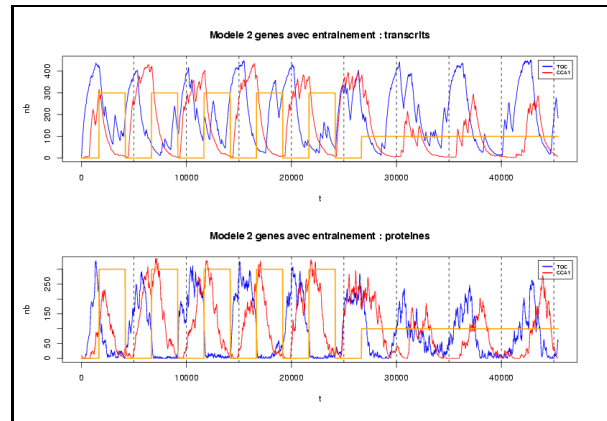


FIG. 7 – Résultats typiques de la simulation à 2 gènes : en haut, courbe des ARNm de TOC1 et CCA1 ; en bas, courbe des protéines. Axe des abscisses : nombre de cycles de simulation. La cellule est entraînée d'abord par un signal lumineux carré ($T = 5000$ cycles) puis on passe en lumière continue. Ici on a $\delta_{MA} = \delta_{MB} = 1,5 \cdot 10^{-3}$, ce qui équivaut à une demi-vie des ARNm d'un peu plus de 2 h (compatible avec les données). À noter : le déphasage progressif en jour continu, le système retournant à sa période libre légèrement supérieure à 24 h.

5.3 Comparaison avec les modèles équationnels

La modélisation de circuits génétiques passe majoritairement par des systèmes équationnels décrivant l'évolution des concentrations des divers acteurs moléculaires impliqués en fonction du temps [Goldbeter, 1996, Locke et al., 2005, Boulter et al., 2007]. Ces modèles présentent le double avantage de pouvoir être étudiés analytiquement selon des méthodes issues de la dynamique non-linéaire ou du calcul formel, ainsi que de donner des résultats d'ensemble via une intégration numérique. Ils permettent également une modélisation uniforme des circuits de régulation génétique [François, 2005]. Mais ils im-

pliquent aussi, par leur forme même, un ensemble d'hypothèses parfois fortes :

- la cellule est traitée comme un réacteur chimique homogène, sans tenir compte des hétérogénéités spatiales ;
- la notion de concentration est supposée toujours pertinente (alors que le nombre de molécules mis en jeu peut devenir trop faible, ou que la croissance de la taille de la cellule durant sa réplication provoque en fait une dilution, etc.)
- les équations sont déterministes : se posent donc la question de la robustesse de ces systèmes au bruit et aux événements stochastiques (fixation d'un régulateur transcriptionnel sur l'ADN par exemple)
- il est difficile d'introduire des notions spatiales, telles que le simple fait que transcription et traduction ont lieu à des endroits très différents de la cellule.

Pour compenser ces problèmes, de nombreux auteurs soumettent leurs modèles à des simulations stochastiques pour évaluer l'influence du bruit [Gonze et al., 2002] ou introduisent des délais dans les équations pour remplacer le transport de molécules au sein de la cellule [Lewis, 2003]. Mais la simulation multi-agent, par sa nature intrinsèquement stochastique et spatialisée, constitue une approche plus adéquate pour estimer la robustesse des modèles.

En l'occurrence, les résultats de ces simulations multi-agents corroborent les études menées sur les modèles équationnels à 1 et 2 gènes, et ce dans un contexte discret et stochastique (les concentrations sont remplacées par un nombre d'agents), où l'espace joue un rôle crucial. Ils permettent donc de vérifier notamment que les modèles d'oscillateurs minimaux sont robustes au bruit.

Les observations faites sur le « pseudo-oscillateur » du modèle à 1 gène donnent par ailleurs un éclairage intéressant sur les approches mathématiques, qui visent à trouver de « véritables » oscillateurs. En effet, il arrive fréquemment qu'un système dynamique présente des régimes transitoires qui ressemblent à des oscillations mais disparaissent après un temps plus ou moins long. Or, cette distinction entre régime oscillant et régime transitoire n'est peut-être pas si pertinente dans ce cadre, puisqu'on se trouve en fait dans les mêmes conditions d'étude (expérimentation et interprétation) que pour le système naturel : la question est bien de savoir si, durant les quelques « jours » d'observation, le système simulé présente les mêmes caractéristiques dynamiques que le système vivant, et

non de savoir si ce système serait capable en théorie d'osciller indéfiniment.

Enfin, les simulations par agents donnent une représentation explicite des modèles (acteurs moléculaires et interactions) à l'intention de nos interlocuteurs biologistes et se prêtent aisément à l'évaluation d'une hypothèse « à la volée ».

6 Conclusion, perspectives

Nous avons montré dans cet article comment la simulation multi-agent pouvait s'inscrire dans une démarche pluridisciplinaire de modélisation, venant compléter des études analytiques de modèles mathématisés.

Les résultats préliminaires obtenus sur ces modèles simples ont permis d'identifier les points qui requièrent un approfondissement : il s'agit en premier lieu d'obtenir des profils d'expression (heures des pics en particulier) plus proches des données. Cela semble possible en introduisant dans le modèle à 2 gènes une dégradation enzymatique explicite des protéines [Morant et al., 2007]. Nous envisageons également de représenter plus finement les mécanismes de transcription, notamment en reproduisant des dynamiques de « bursts » (c'est-à-dire l'existence d'un délai d'allumage d'un gène suivi de sa transcription intensive en ARNm pendant de courtes périodes, donc des processus poissoniens), et tenir compte de la « marche » des facteurs de transcription le long de l'ADN (par l'introduction d'un biais de déplacement des agents par exemple).

De plus, l'hypothèse d'action directe de la lumière sur la dégradation de la protéine TOC1, même si elle a permis d'effectuer une première étude fructueuse des mécanismes de l'oscillateur, s'est avérée évidemment simpliste. Elle fait donc actuellement l'objet d'une révision pour tenir compte d'homologies avec d'autres systèmes biologiques et d'une possible variation de la photosensibilité selon l'heure circadienne. Des courbes de réponse de phase pourront alors établir plus précisément si le modèle à 2 gènes révisé suffit à rendre compte de toutes les données. À plus long terme, nous chercherons des mécanismes robustes de transmission du signal d'horloge vers des voies de sortie (pour le contrôle du cycle de division cellulaire).

Nous avons en outre commencé à étudier les moyens de formaliser mathématiquement certains aspects de la simulation agent, moyennant

certaines hypothèses simplificatrices, pour établir un pont entre les paramètres des systèmes d'équations différentielles et les paramètres de simulation multi-agent relatifs à un même modèle biologique.

Quelques-uns des problèmes que nous n'avons que mentionnés ici mériteraient une étude séparée et de plus vaste ampleur, en particulier le lien entre hétérogénéités spatiales et échelles de temps, le passage de modèles « globaux » aux modèles spatiaux (notamment en comparant la simulation distribuée à des équations de diffusion), ainsi que le problème du changement d'échelle dans les systèmes dynamiques.

Remerciements

Nous souhaitons remercier tous les participants du groupe de travail sur la modélisation des rythmes circadiens de Lille la qualité des discussions, ainsi que les membres de l'équipe de F.-Y. Bouget, qui nous ont chaleureusement accueilli à Banyuls et ont supporté nos questions naïves avec une patience remarquable.

Références

- [Boulier et al., 2007] Boulier, F., Lefranc, M., Lemaire, F., Morant, P.-E., and Ürgüplü, A. (2007). On proving the absence of oscillations in models of genetic circuits. In *Proceedings of Algebraic Biology*.
- [Courties et al., 1994] Courties, C., Vaquer, A., Troussellier, M., Lautier, J., Chrétiennot-Dinet, M. J., Neveux, J., Machado, C., and Claustre, H. (1994). Smallest eukaryotic organism. *Nature*, 370(6487) :255.
- [Derelle et al., 2006] Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piégu, B., Ball, S. G., Ral, J.-P., Bouget, F.-Y., Pignoneau, G., De Baets, B., Picard, A., Delseny, M., Demaille, J., Van de Peer, Y., and Moreau, H. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 103 :11647–11652.
- [Filipski et al., 2005] Filipski, E., Innominat, P., Wu, M., Li, X., Iacobelli, S., Xian, L., and Lévi, F. (2005). Effects of light and food schedules on liver and tumor molecular clocks in mice. *Journal of the National Cancer Institute*, 97(7).
- [François, 2005] François, Paul et Hakim, V. (2005). Core genetic module : the mixed feedback loop. *Phys. Rev. E*, 72(031908).
- [Goldbeter, 1995] Goldbeter, A. (1995). A model for circadian oscillations in the *Drosophila* period protein (PER). *Proc. R. Soc. Lond. B*, 261 :319.
- [Goldbeter, 1996] Goldbeter, A. (1996). *Biochemical Oscillations and Cellular Rhythms : The molecular bases of periodic and chaotic behaviour*. Cambridge University Press, Cambridge.
- [Gonze et al., 2002] Gonze, D., Halloy, J., and Goldbeter, A. (2002). Deterministic versus stochastic models for circadian rhythms. *J. Biol. Phys.*, 28 :637–653.
- [Harmer et al., 2001] Harmer, S., Panda, S., and Kay, S. (2001). Molecular bases of circadian rhythms. *Annu. Rev. Cell. Dev. Biol.*, 17 :215–253.
- [Leloup et al., 1999] Leloup, J.-C., Gonze, D., and Goldbeter, A. (1999). Limit cycle models for circadian rhythms based on transcriptional regulation in *Neurospora* and *Drosophila*. *J. Biol. Rhythms*, 14 :433–448.
- [Lewis, 2003] Lewis, J. (2003). Autoinhibition with transcriptional delay : a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.*, 13 :1398.
- [Locke et al., 2005] Locke, J. C. W., Millar, A. J., and Turner, M. S. (2005). Modelling genetic networks with noisy and varied experimental data : the circadian clock in *Arabidopsis thaliana*. *J. Theor. Biol.*, 234 :383.
- [Mathieu et al., 2007] Mathieu, P., Picault, S., and Routier, J.-C. (2007). Donner corps aux interactions (l'interaction enfin concrétisée). In *Actes de la conférence Modèles Formels pour l'Interaction (MFI'07)*.
- [Morant et al., 2007] Morant, P.-E., Vandermoere, C., Parent, B., Lemaire, F., Corellou, F., Schwartz, C., Bouget, F.-Y., and Lefranc, M. (2007). Oscillateurs génétiques simples. applications à l'horloge circadienne d'une algue unicellulaire. In *Actes des Rencontres du non linéaire*, Paris.
- [Salome and McClung, 2004] Salome, P. and McClung, C. (2004). The *Arabidopsis thaliana* clock. *J. Biol. Rhythms*, 19 :425–435.