

N° d'ordre : 40723

**THÈSE DE DOCTORAT
DE L'UNIVERSITÉ DES SCIENCES ET TECHNOLOGIES
DE LILLE I**

Spécialité :

LASER, MOLÉCULES ET RAYONNEMENT ATMOSPHERIQUE

présentée par

Constant VANDERMOËRE

Pour obtenir le grade de

DOCTEUR de L'UNIVERSITE DE LILLE I

Sujet de la thèse

**Exemples d'oscillateurs génétiques :
le gène auto-réprimé à dynamique
transcriptionnelle lente et l'oscillateur central de
l'horloge circadienne de l'algue *Ostreococcus Tauri***

Soutenue le 15 Décembre 2011 devant le jury composé de :

Christophe LETELLIER	Professeur, Université de Rouen	Rapporteur
Aziz ALAOUÏ	Professeur, Université du Havre	Rapporteur
Marc LEFRANC	Professeur, Université Lille I	Directeur de thèse
Didier GONZE	Premier Assistant, Université Libre Bruxelles	Examineur
Ralf BLOSSEY	DR CNRS, Université Lille I	Examineur

Remerciements

Ayant effectué ma thèse au laboratoire PhLAM, les remerciements que je dois à son Directeur, M. WLODARCZAK, au Directeur du CERLA, M. DEROZIER, ainsi qu'aux Directeurs d'Etudes Doctorales successifs, M. ROHART et TAKI, ne sont pas uniquement dus à la convenue. Elles émanent d'un sentiment profond de reconnaissance pour ceux qui m'ont aidé lors de ma recherche d'un financement de thèse.

Bien naturellement, l'expression de ma gratitude la plus élevée revient à mon Directeur de Thèse, M. Marc LEFRANC, qui m'a fait partager ses qualités de physicien. La science exige de nombreuses qualités qui doivent être coordonnées, et j'ai pu apprécier chez lui la rigueur, la culture, la compréhension, l'inspiration. Les mots de Nicolas Boileau : "Ce que l'on conçoit bien s'énonce clairement, Et les mots pour le dire arrivent aisément" lui conviennent parfaitement.

Je remercie également M. Quentin THOMMEN qui a apporté un élan et une grande fraîcheur lors de son arrivée dans l'équipe dirigée par M. LEFRANC. Sa vision juste et ses grandes qualités de modélisateur ont été pour moi une source majeure d'inspiration. Je remercie aussi M. Pierre-Emmanuel MORANT pour l'aide qu'il m'a apportée à de nombreuses reprises.

La science résulte de la symbiose des efforts d'un grand nombre de personnes. Lorsqu'on se penche sur le développement de la science, on constate que la plupart des créations de l'esprit sont inspirées par celles qui les ont précédées. La formalisation de la "Nature" par un physicien théoricien n'est possible que grâce à la description qu'en fait l'expérience qui, en fin de compte, dirige l'évolution de la pensée. Mon travail n'aurait pu voir le jour sans les travaux expérimentaux de M. François-Yves BOUGET et Mme CORRELOU. Ces biologistes hors pair ont réalisé de formidables prouesses expérimentales, qui sont à la base de ma thèse. Je les remercie pour cela.

Je tiens à remercier également l'ensemble du jury, M. LETELLIER, M. ALAOUI, M. GONZE et M. BLOSSEY, qui ont accepté de juger et valider mon travail, et par là de le voir reconnu. C'est également un grand honneur pour moi que des scientifiques d'une telle envergure participent à mon jury de thèse.

Le travail de doctorant n'est pas limité à la recherche car il faut également apprendre à enseigner. Je remercie donc les personnes m'ayant permis de dispenser des enseignements. M. BRIQUEZ est une des rares personnes qui excelle autant d'un point de vue scientifique que sportif. Sa force de volonté lui a permis de faire ce que la plupart font en deux vies. Je remercie également Mme Carole GORS, actuelle Directrice de l'UFR de physique, pour m'avoir communiqué une certaine vision de l'enseignement et m'avoir apporté son expérience.

Enfin je souhaite remercier mes parents et ma femme Linda pour l'aide et le soutien moral qu'ils m'ont apporté durant cette thèse.

Résumé

Les gènes situés sur la molécule d'ADN au coeur de nos cellules non seulement portent une information héréditaire, mais de plus participent dynamiquement au fonctionnement de la cellule en synthétisant plus ou moins activement des protéines. Ces dernières sont susceptibles de nombreuses interactions non linéaires avec d'autres acteurs, menant à la formation de réseaux biochimiques qui peuvent présenter des comportements dynamiques divers et complexes. En particulier, la régulation de l'activité des gènes (leur taux de transcription) par des protéines spécifiques peut aboutir à la formation de boucle de rétroaction. Ces boucles tissent des réseaux génétiques où un ensemble de gènes régulent réciproquement leurs expressions.

Des travaux expérimentaux récents ont montré que la régulation de l'activité du gène n'est pas toujours instantanée. Nous avons alors étudié l'influence d'une dynamique transcriptionnelle intrinsèque dans un circuit simple constitué d'un gène réprimé par sa propre protéine. Nous obtenons un critère analytique pour l'apparition des oscillations, qui nous permet de montrer que ces dernières sont favorisées lorsque le temps de réponse du gène prend une certaine valeur. L'échelle de temps ainsi repérée est pertinente à la fois dans une description déterministe et dans une description stochastique.

Ce sont également des réseaux génétiques qui sont à l'origine de certains rythmes biologiques, induisant les oscillations de l'expression de protéines clés. Ces réseaux sont alors des horloges endogènes qui permettent à de nombreux êtres vivants d'anticiper les modifications cycliques de l'environnement. Parmi celles-ci, l'horloge circadienne permet à l'organisme de s'adapter au cycle diurnal en étant entraîné par l'alternance du jour et de la nuit.

À partir de données expérimentales, nous avons modélisé l'horloge circadienne de l'algue unicellulaire *Ostreococcus tauri*. Le modèle est basé sur une boucle de rétroaction transcriptionnelle négative impliquant deux gènes se régulant mutuellement. L'accord entre le modèle et les données expérimentales est excellent et met en évidence une absence de signature du couplage dans les données lorsque l'horloge est à l'heure, ce qui révèle une propriété de robustesse aux fluctuations d'éclairement.

mots clefs : dynamique transcriptionnelle, oscillateur génétique, boucle de rétroaction, horloge circadienne, rythme biologique, réseau génétique

Genes located on the DNA macromolecule inside our cells do not only carry hereditary information. They also contribute dynamically to biological functions by synthesizing proteins at a variable rate. Proteins are subject to many nonlinear interactions with other actors, forming vast biochemical networks which may display a number of complex behaviors. In particular, regulation of gene activity (i.e., of their transcription rate) by specific proteins creates feedback loops. These loops form genetic networks where a set of genes regulate their expressions reciprocally.

Recent experimental studies have shown that gene regulation is not always instantaneous. We have thus studied the influence of an intrinsic transcriptional dynamics in the simple circuit where a gene is repressed by its own protein. We have obtained an analytical criterion for the appearance of oscillations, which allows us to show that oscillations are favored when gene response time is close to a characteristic value. The time scale thus identified is relevant both in a deterministic and a stochastic description.

Gene regulatory networks are also at work in some biological rhythms, inducing oscillations in the concentrations of some key proteins. These networks then serve as endogenous clocks, which allow many living organisms to anticipate periodic changes in the environment. In particular, the circadian clock is used by organisms to adapt to the diurnal cycle by being entrained by the day/night cycle.

Using experimental data, we have constructed a mathematical model of the circadian clock of the unicellular alga *Ostreococcus tauri*. This model is based on a transcriptional negative feedback loop, which involves two genes regulating each other. Agreement between numerical simulations and experimental data is excellent and unveils the fact that there is no signature of coupling in data when the clock is on time. This reveals a strong robustness to daylight fluctuations.

keywords : transcriptional dynamics, genetic oscillator, feedback loop, circadian clock, biological rhythm, genetic network

Table des matières

Avant-propos	11
1 Réseaux de régulation génétiques	15
1.1 Expression des gènes par les protéines	15
1.1.1 Rôles fonctionnels des protéines	15
1.1.2 Support de l'information du vivant : les gènes	16
1.1.3 Variation de l'activité transcriptionnelle du gène	17
1.1.4 Rôle de l'ARN messenger	18
1.2 Description mathématique des processus moléculaires	19
1.2.1 Loi d'action de masse	19
1.2.2 Activité du gène	19
1.2.3 Traduction	20
1.2.4 Dégradations	21
1.2.5 Coefficient de Hill	22
1.2.6 Autres processus	22
1.2.7 Effets stochastiques	23
1.3 Dynamique induite par les réseaux de régulation génétique	24
1.3.1 Réseau génétique à boucle de rétroaction	24
1.3.1.1 Formation d'un réseau biochimique	24
1.3.1.2 Régulation simple	26
1.3.1.3 Auto-régulation	26
1.3.1.4 Boucle de rétroaction	27
1.3.2 Multiple dynamique d'un réseau génétique : le "Mixed Feedback Loop"	28
1.3.2.1 Présentation des équations du MFL	28
1.3.2.2 Diagramme de bifurcation	29
1.3.3 Réseaux artificiels oscillant : le Repressilator	31
1.4 Horloges circadiennes	34
1.4.1 Définition du rythme biologique	34
1.4.2 Définition du rythme circadien	35
1.4.3 Ubiquité et topologie de l'horloge circadienne	36
1.4.4 Analogie avec un oscillateur auto-entretenu	38
1.4.5 Structures d'horloge circadienne de quelques organismes	41
1.4.5.1 Horloge circadienne de <i>Drosophila melanogaster</i>	41
1.4.5.2 Modèle <i>Arabidopsis Thaliana</i>	43
1.5 Mesures quantitatives de l'expression génétique	45
1.5.1 Séquençage de l'ADN	45

1.5.2	Micropuces à ADN	46
1.5.3	Gène rapporteur de la luciférase	46
2	Réseau génétique oscillant à un seul gène	51
2.1	Mécanismes permettant les oscillations dans un réseau génétique	51
2.1.1	Modèle de rétroaction négative à délai	51
2.1.2	Conditions d'oscillations	52
2.2	Gène auto-régulé à dynamique transcriptionnelle rapide	54
2.2.1	Exemple d'oscillateur à un seul gène : l'oscillateur de Goodwin	54
2.2.2	Exemple de gène auto-régulé contenant un délai temporel : la segmentation des somites	56
2.2.3	Modèle de gène auto-régulé de la drosophile : délai par multiphosphorylation	58
2.2.3.1	Phosphorylation	58
2.2.3.2	Modèle de Goldbeter	58
2.2.4	Dynamique d'activation lente du gène	60
2.3	Dynamique d'un gène auto-régulé à temps de réponse variable du gène	61
2.3.1	Présentation du modèle	61
2.3.2	Critère analytique d'apparition des oscillations	62
2.3.2.1	Équations sans dimension et détermination des points fixes	62
2.3.2.2	Critère d'oscillation	63
2.3.2.3	Condition nécessaire d'obtention des oscillations	64
2.3.2.4	Relation entre deux paramètres normalisés	65
2.3.2.5	Forme simplifiée du critère	65
2.4	Analyse du critère de stabilité	66
2.4.1	Dynamique pour certains types de fonction de dégradation	66
2.4.2	Influence de la dynamique du gène sur l'apparition d'oscillations	67
2.4.2.1	Temps de réponse du gène et non-linéarité des équations	67
2.4.2.2	Diagramme de phase	68
2.4.2.3	Application avec une dégradation allostérique	69
2.4.2.4	Régions du diagramme de phase	69
2.4.3	Phénomène de résonance : facteur d'optimisation de l'apparition des oscillations	70
2.4.3.1	Temps de réponse optimale du gène	70
2.4.3.2	Intérêt de l'étude stochastique	71
2.5	Conclusion	72
3	Nouveau modèle circadien d' <i>Ostreococcus Tauri</i>	73
3.1	Stratégie d'étude du rythme circadien	73
3.1.1	Interface biologie et mathématique	73
3.1.2	Modèle biologique	73
3.1.3	Modèle mathématique	74
3.2	Rythme circadien d' <i>Ostreococcus Tauri</i>	75
3.2.1	Présentation de l'organisme	75
3.2.2	Identification de deux gènes de l'horloge : TOC et CCA	76
3.2.3	Données d'expression de l'ARN messager par les micropuces à ADN	77

3.2.4	Données de luminescences de l'activité génique et des concentrations de protéines	78
3.2.4.1	Lignées de fusion transcriptionnelle	78
3.2.4.2	Lignées de fusion traductionnelle	79
3.2.4.3	Signaux d'activité transcriptionnelle et traductionnelle en cycle jour nuit	80
3.2.4.4	Signaux d'activité transcriptionnelle et traductionnelle en jour continu	81
3.2.5	Effets de la surexpression et de la répression des gènes de l'horloge	82
3.2.6	Régulation directe de la protéine <i>CCAI</i> sur le promoteur du gène <i>tocI</i>	84
3.2.7	Réseau génétique à deux gènes au coeur de l'horloge circadienne	85
3.3	Modèle de l'horloge circadienne	86
3.3.1	Introduction	86
3.3.2	Équations de la boucle à deux gènes	86
3.3.2.1	Algorithme d'identification de la boucle de régulation	86
3.3.2.2	Voie d'entrée de la lumière dans le modèle	88
3.3.3	Signaux de luminescence de l'activité transcriptionnelle et de l'activité traductionnelle	89
3.3.3.1	Signaux de luminescence en cycle jour nuit	89
3.3.3.2	Signaux de luminescence en jour continu	91
3.3.4	Biais des signaux expérimentaux	94
3.3.4.1	Analyse du bruit de fond	94
3.3.4.2	Biais du plancher de luminescence	95
3.3.5	Traitement des signaux expérimentaux des puces à ADN et de luminescence	99
3.3.5.1	Signaux des micropuces à ADN	99
3.3.5.2	Détermination d'un profil cible de l'ARN à partir des micropuces à ADN	101
3.3.5.3	Détermination d'un profil cible à partir des lignées de fusion traductionnelle	103
3.3.5.4	Quantification de l'accord entre le profil expérimental et le profil numérique	105
3.3.6	Résultats de l'ajustement sans couplage à la lumière	107
3.3.6.1	Ajustement des profils biaisés par le plancher de luminescence	107
3.3.6.2	Ajustement des profils sans le plancher de luminescence	109
3.3.7	Ajustement avec couplage à la lumière	110
3.4	Conclusion	113

4 Modèles de l'horloge circadienne prenant en compte la dynamique de la luciférase 115

4.1	Introduction	115
4.2	Nouveaux modèles de l'horloge circadienne	116
4.2.1	Cinétique de la luciférase	116
4.2.2	Modèle des lignées de fusion donnant l'activité promoteur	116
4.2.3	Modèles décrivant des lignées traductionnelles	117
4.2.3.1	Compétition entre facteurs de transcription	117

4.2.3.2	Compétition entre protéines pour l'enzyme de dégradation . . .	118
4.2.3.3	Modèles de l'activité traductionnelle	119
4.3	Stratégie d'ajustement des données par le modèle	120
4.3.1	Traitement des données par Transformée de Fourier Rapide	120
4.3.2	Fonction de score	122
4.3.3	Méthodes d'ajustement des courbes	122
4.4	Influence des paramètres cinétiques sur l'interprétation des expériences	123
4.4.1	Variation de la durée d'émission de lumière par la luciférase	123
4.4.2	Temps d'inactivation de la luciférase	124
4.5	Résultat de l'ajustement par les modèles détaillés	127
4.5.1	Protocole des ajustements	127
4.5.2	Ajustement des lignées <i>pTOC :Luc</i> et <i>pCCA :LUC</i>	127
4.5.3	Comparaison entre les lignées <i>pTOC :LUC</i> et <i>pCCA :LUC</i>	128
4.5.4	Indice de saturation et de régulation	131
4.5.5	Ajustement des lignées traductionnelles <i>TOC :LUC</i> et <i>CCA :LUC</i>	131
4.6	Discussions des paramètres des ajustements des lignées traductionnelles	134
4.7	Conclusion	136
5	Conclusion	137
A	Annexe au chapitre 2 : Oscillations in the Expression of a Self-Repressed Gene Induced by a Slow Transcriptional Dynamics	139
B	Annexe au chapitre 3 : A robust two-gene oscillator at the core of <i>Ostreococcus tauri</i> circadian clock	145
C	Fonction de transcription et de dégradation	159
C.1	Fonction de transcription	159
C.2	Fonction de dégradation	160
D	Compétition entre facteurs de transcription	163
E	Tableaux des valeurs de paramètres du modèle tenant compte de la luciférase	165
E.1	Lignée <i>pTOC :LUC</i>	166
E.2	Lignée <i>pLHY :LUC</i>	167
E.3	Lignées <i>TOC :LUC</i>	168
E.4	Lignées <i>LHY :LUC</i>	169

Introduction

La vie est complexe. Elle existe grâce à la synchronisation harmonieuse de phénomènes ou de fonctions biologiques assurant l'évolution d'un organisme de sa naissance à sa mort [1]. La cellule se divise, se spécialise pour former des tissus et des organes [2], se renouvelle et meurt, parfois de manière programmée ou pour répondre à l'environnement [3]. Pour survivre et s'adapter, les organismes vivants ont développé des rythmes biologiques présents dans les cellules, le tissu et le cerveau afin de contrôler nos comportements, comme le sommeil et l'alimentation pour les mammifères [4], la floraison pour les fleurs [5, 6]... Certains de ces rythmes sont communs entre les organismes cellulaires les plus complexes et la plus simple des bactéries [7]. La dynamique de ces rythmes a pour origine l'Acide Désoxyribonucléique, repris par la suite sous la forme trigramme "ADN".

La recherche et les développements technologiques ont permis une révolution scientifique majeure grâce à l'essor de la génétique et de la biologie moléculaire. Au coeur de ces découvertes, il y a l'ADN [2]. L'ADN est une molécule complexe portant une information héréditaire transmise et portée par chaque cellule. L'ADN contient le génome, la cartographie de l'ensemble des gènes d'un organisme vivant [8]. La mission générale des gènes est connue : les gènes contiennent l'information pour synthétiser les protéines (voir figure 1)[2]. Les protéines sont les briques élémentaires de la vie, les différentes parties de la machinerie moléculaire d'une cellule en ayant différents rôles [2, 9, 10] : elles constituent la structure de la cellule, elles permettent la signalisation, la réponse immunitaire, les rythmes biologiques... Cependant, même le catalogue des gènes complètement connu, le recoupement de données expérimentales, les outils d'analyses et les comparaisons de séquences génétiques entre différents organismes [11, 9] ne donnent qu'une partie des fonctionnalités des protéines dans un organisme donné. Ainsi, la connaissance de la structure moléculaire d'une protéine ne permet pas toujours de savoir quelles sont les tâches qui lui incombent, comment et quand elle les effectue. Quelle est donc l'information manquante ?

Les gènes et les protéines ne sont pas des molécules dont l'activité est individuelle, statique et prédéfinie. Elles fonctionnent en faisant partie d'un groupe d'acteurs moléculaires dont les facultés d'actions sont collectives. Ce dynamisme naît grâce aux interactions moléculaires qui modifient les propriétés chimiques menant à la régulation, au transport des biomolécules ... La multiplicité de ces interactions crée un réseau d'interactions moléculaires. Certains gènes sont au coeur de tels réseaux et leur rôle dépend de l'ensemble des interactions auxquelles ils sont soumis [12, 13, 14] et ne découle pas directement du décryptage du génome. Ces réseaux permettent à la cellule de répondre à de nombreux signaux inter ou extra-cellulaire, à s'adapter à des modifications environnementales ou de se confronter aux changements physiologiques.

À la figure 2, nous présentons un réseau d'interactions complexe mais partiel du cycle cellulaire du mammifère [3]. Le schéma montre de nombreuses voies d'interactions chimiques contrôlant l'expression des gènes. Ces voies peuvent être reliées entre elles, induisant des ré-

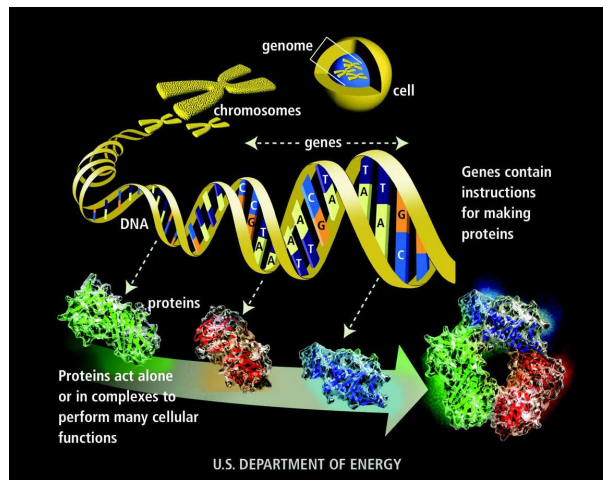


FIGURE 1 – Vision simplifiée de la synthèse des protéines à partir des gènes (Illustration de l’American Academy of Environmental Medicine). L’ensemble des gènes constituant le génome contient l’information pour fabriquer les protéines. Les protéines sont immédiatement actives ou sont modifiées pour réaliser de nombreuses fonctions nécessaires pour la cellule.

seaux imbriqués les uns aux autres. Ce type de réseau est susceptible d’avoir de nombreux comportements dynamiques des concentrations moléculaires. La détermination des comportements dynamiques d’un tel réseau est nécessaire à l’avancée de la compréhension de la biologie moléculaire, au-delà d’un niveau où l’intuition est dépassée par la complexité des réseaux. L’incorporation de nouvelles méthodes mathématiques et l’utilisation d’outils théoriques deviennent alors pertinentes [1] car elles permettront à la biologie ou la médecine de diagnostiquer, de prédire [15] ou de corriger des dérèglements génétiques [16].

La modélisation mathématique est multiple (déterministe, délai, stochastique ...) [17, 1, 18, 19, 20, 21]. Mais quelles que soient les méthodes utilisées, elles nécessitent toujours d’identifier les acteurs moléculaires, les interactions (la fonction mathématique représentant l’interaction) et donc finalement les valeurs des paramètres implicites aux équations mathématiques [14].

Les équations mathématiques des réseaux sont souvent (pour ainsi dire toujours) fortement non-linéaires et aboutissent à de nombreux comportements dynamiques [22, 23, 24]. Ces comportements dynamiques répondent à des besoins vitaux pour les organismes vivants. Ainsi, la bistabilité (ou la multistabilité) permet de faire mémoriser le chemin emprunté (certains processus biologique “se souviennent” de réactions passées) [25, 26] ou alors permet d’enclencher l’irréversibilité de certaines étapes du cycle cellulaire [27]. La multistabilité accorde à l’organisme le choix de l’expression génétique la plus adaptée à l’environnement. L’existence des oscillations est discutée plus amplement dans le manuscrit car elles sont fondamentales pour le vivant. Issues de la dynamique de nombreux processus bio-chimiques, elles sont, par exemple, à la base des rythmes biologiques.

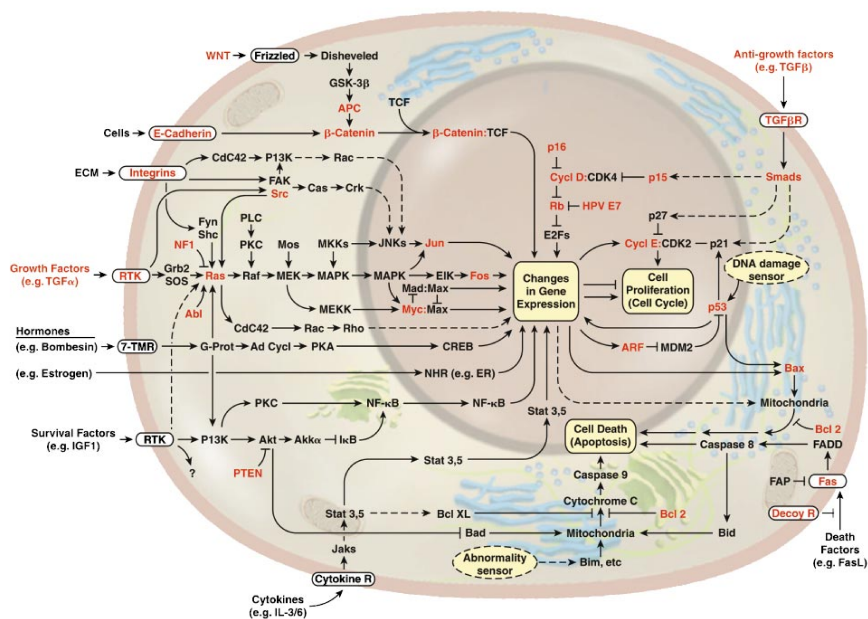


FIGURE 2 – (Tirée de [3]). Représentation schématique partielle du réseau d'interaction des principales protéines impliquées dans le contrôle du cycle et de la mort programmée de la cellule chez un mammifère. Dans les cases jaunes apparaissent des fonctionnalités primordiales du cycle cellulaire contrôlant l'expression du gène. Selon les stimulus extérieurs et les accidents du hasard (DNA damage), l'expression du gène s'altère. Le comportement ou l'état dynamique du réseau change au gré des modifications des paramètres du système, signe d'une grande adaptabilité et flexibilité des processus du vivant.

Cette thèse est composée de quatre parties. Dans le chapitre 1, nous présentons quelques aspects du fonctionnement du gène et en particulier les mécanismes biologiques impliqués au niveau de l'ADN qui permettent l'interaction avec des protéines. Nous définissons alors le terme de réseau génétique. Nous montrons quelques exemples de comportements dynamiques possibles dans ces réseaux génétiques. Ces derniers sont également impliqués dans l'existence de certains rythmes biologiques, comme le rythme circadien. Ce rythme est fondamental pour de nombreux organismes. Il est illustré à partir de modèles biologiques étudiés par des équations mathématiques. Pour clore ce chapitre, quelques techniques de mesures de grandeurs biologiques sont expliquées.

Dans le chapitre 2, nous détaillons des modèles mathématiques s'évertuant à distinguer parmi les nombreuses réactions bio-chimiques quels sont les processus clefs aboutissant aux oscillations. Nous soulignerons en particuliers l'importance de la non-linéarité, de la variété des échelles de temps et de l'omniprésence d'une rétroaction négative agissant sur le gène (ce dernier terme sera expliqué plus en détail). A partir de ces propriétés, nous établissons un modèle de réseau génétique simple constitué d'un seul gène. Notre modèle se différencie des autres grâce à une influence du temps de réponse du gène non négligeable et vitale dans l'apparition des oscillations. Cette particularité met en évidence la présence d'une résonance favorisant les oscillations mais également pertinente lorsque le système biologique est vu sous un angle stochastique.

Dans le chapitre 3, nous appliquons un modèle mathématique au réseau génétique sous-jacent l'horloge circadienne d'un nouvel organisme microscopique appelé *Ostreococcus Tauri*. Cette algue unicellulaire et microscopique est génétiquement simple mais elle est assez complexe pour contenir le rythme circadien et les caractéristiques de plantes avancées, comme la capacité à faire la photosynthèse. L'organisme peut donc devenir un modèle de référence d'étude biologique. La première partie consiste à présenter et à décrire les deux gènes centraux liés au rythme circadien. Certains points expérimentaux nécessitent d'être confirmés par un modèle mathématique. Pour cela, une phase de traitement des données expérimentales est faite et précède la phase de comparaison des profils numériques issus de l'intégration de nos équations. Un bon accord est trouvé entre la théorie et l'expérience. Nous pourrions alors valider la structure centrale du réseau génétique circadien.

Dans le quatrième chapitre, le modèle mathématique du chapitre 3 est transformé afin de vérifier certaines hypothèses liés aux données expérimentales. En effet, ces dernières sont supposées être directement assimilables à des concentrations de protéines. La détermination de certains paramètres cinétiques confirment cette dernière hypothèse. Malgré tout, la mise en place expérimentale aboutit à une légère transformation génétique de l'organisme vivant. Nous devons alors vérifier que cette perturbation ne compromet pas nos résultats.

Chapitre 1

Réseaux de régulation génétiques

La définition des réseaux génétiques se fait à travers la présentation de trois acteurs moléculaires majeurs : le gène, l'acide ribonucléique messager (dit ARNm) et la protéine. Après avoir expliqué leurs rôles, nous détaillons la nature des interactions et des processus biochimiques existante. Nous montrons comment modéliser ces réactions par des équations mathématiques. Nous illustrons la dynamique des réseaux génétiques grâce à un réseau appelé le "mixed feedback loop". Ce dernier modèle a quatre comportements dynamiques possibles, selon la valeur des constantes cinétiques. Ensuite, nous considérons un réseau génétique artificiel dont les quantités de protéines oscillent au cours du temps. Ce réseau est appelé le "repressilator". Les rythmes biologiques, dont l'existence dépend de l'oscillation des concentrations de protéines, sont définis par la suite. Nous insistons sur la définition du très répandu rythme circadien. La structure génétique d'organismes vivants et la stratégie mathématiques d'étude de ce rythme sont discutées. En conclusion, nous analysons certaines techniques de mesures quantitatives et qualitatives de données biologiques permettant de travailler sur les rythmes biologiques.

1.1 Expression des gènes par les protéines

1.1.1 Rôles fonctionnels des protéines

Les protéines sont des macromolécules se composant d'acides aminés dont le nombre varie de quelques uns à plusieurs centaines. Ces acides sont reliés par des liaisons peptidiques [2].

Les attributions des protéines sont très nombreuses puisqu'elles assurent la majorité des fonctions cellulaires. Les protéines de structure constituent les différents éléments de l'organisme. Par exemple, les protéines du cytosquelette consolident la cellule. Des protéines jouent un rôle de communication comme certaines hormones ou la protéine G (voir figure 1.1). D'autres protéines assurent le transport de l'hémoglobine qui fournissent les cellules en oxygène. Elles servent aussi de catalyseurs de réactions biochimiques, comme les protéases. Ces dernières engendrent la dégradation d'autres protéines en cassant les liaisons peptidiques. Elles ont d'autres rôles dont nous comprenons de plus en plus l'importance : certaines protéines contrôlent le rythme cyclique de processus physiologiques ou comportementaux. Ce contrôle s'opère à travers la variation périodique de la concentration de ces protéines.

Les différentes cellules ne produisent pas toutes les mêmes protéines, ce qui engendre la spécialisation cellulaires dans les organismes vivants. L'ensemble des protéines actives d'un individu définit alors des caractères morphologiques et physiologiques qui lui confèrent un aspect

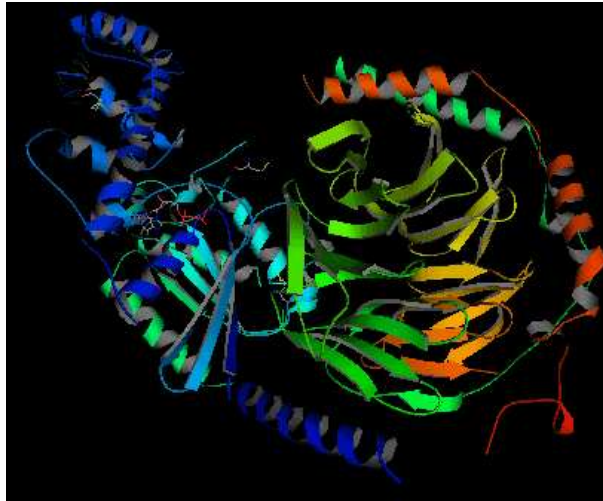


FIGURE 1.1 – Schéma de la structure moléculaire de la protéine G (tiré de [28]). La protéine G (guanine nucleotide binding proteins) permet la transduction du signal, c'est-à-dire le transfert d'information à l'intérieur de la cellule. Si un récepteur est activé à la surface de la cellule par un stimuli extérieur, la protéine G aura un effet inhibiteur ou excitateur via une cascade de signalisation [29].

identifiable appelé le phénotype. Nous nous intéresserons en particuliers aux processus permettant la variation de la concentrations de certaines protéines au cours du temps et participant au phénotype. La régulation de la création des protéines est un élément clef dans l'apparition de ces oscillations et agit au niveau des gènes.

1.1.2 Support de l'information du vivant : les gènes

L'ADN (Acide DésoxyriboNucléique) est une macromolécule formée d'une succession de molécules appelés nucléotides (voir figure 1.2). Il existe 4 types de nucléotides dont la combinaison permet à l'ADN de contenir des unités d'information appelées gènes. L'ADN circule de cellules en cellules grâce à la division cellulaire.

L'ensemble des gènes constitue le génome d'un organisme. Ce génome contient les différents plans architecturaux du vivant synthétisant les principaux constituants cellulaires : les protéines. La synthèse d'une protéine se réduit en deux étapes essentielles. La première étape est la phase de transcription. Au cours de celle-ci, le gène est transcrit en ARN messenger. Une fois l'ARN messenger produit, il vient la deuxième étape : la phase de traduction. L'ARN messenger est traduite par les ribosomes¹ en protéine. De nombreuses transformations post-transcriptionnelles modifient également les propriétés biochimiques de la protéine.

Le spectre des protéines synthétisées par l'intermédiaire des gènes est l'expression génétique ou le phénotype d'une cellule. Mais le nombre de protéines varie au cours du temps grâce à des mécanismes régulateurs afin que leur concentration satisfasse les besoins de la cellule. Une défaillance de la transcription créant une absence ou une surabondance de la quantité des protéines aboutit à de graves problèmes physiologiques. Ainsi, la compréhension des

1. Ribosome : complexe composé de protéines et d'ARN synthétisant les protéines par le décodage de l'information contenue dans l'ARNm.

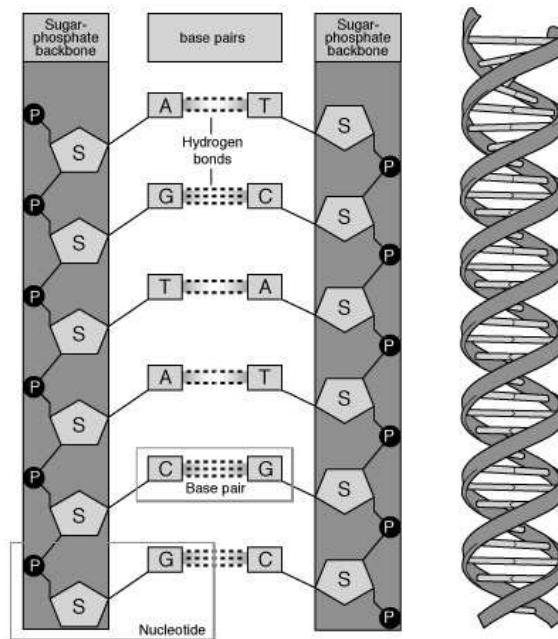


FIGURE 1.2 – Double brin constituant l’ADN. L’ADN est constitué de quatre nucléotides (Amine, Cytosine, Guanine et Thymines). Les différentes combinaisons correspondent à différentes informations portées par l’ADN [2]. *National Human Genome Research*.

mécanismes régulateurs des quantités de protéines permet d’avoir une vision plus précise du fonctionnement de la cellule.

La cellule réduit ou augmente les quantités de certaines protéines en réponse à des stimuli ou à tout autres facteurs. Elle utilise plusieurs moyens de contrôle comme la modification du taux de synthèse de l’ARN messenger au fil du temps, ou autrement dit, la modification du taux transcriptionnel du gène.

1.1.3 Variation de l’activité transcriptionnelle du gène

Le taux de transcription du gène varie au fil du temps. La variation de ce taux est fonction des réponses à des communications cellulaires, à des facteurs environnementaux ou à une “mécanique interne”, c’est-à-dire que le taux de transcription de certains gènes varie de manière autonome comme une horloge.

La cellule régule l’expression de différentes protéines vitales pour son fonctionnement. La variation directe de ce taux transcriptionnel est due à des protéines spécifiques appelées facteurs de transcription. Au cours de la transcription, un ensemble de réactions biochimiques assure la copie locale de la région codante du gène en ARN messenger grâce à une enzyme appelée l’ARN polymérase (voir figure 1.3)². L’ARN polymérase se fixe sur une zone de l’ADN qui est la région promotrice de transcription. Cette dernière se trouve en amont de la partie codante du gène. Ensuite, l’ARN polymérase migre le long de l’ADN pour transcrire le gène en ARN messenger. Le facteur de transcription se fixe sur le promoteur afin d’exercer une régulation de la transcription. Selon la nature et le type de gène, le facteur de transcription favorise (active)

2. Les mécanismes biologiques de la transcription peuvent différer suivant l’organisme considéré [2]).

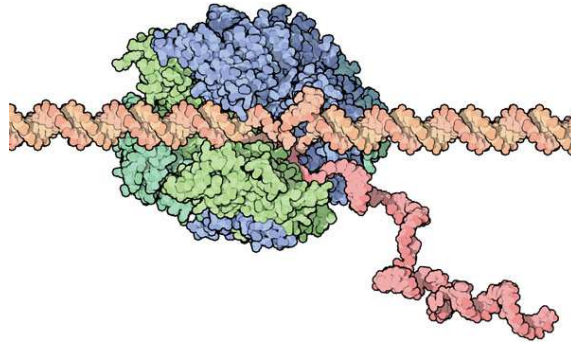


FIGURE 1.3 – L'ARN polymérase se fixe sur le double brin de l'ADN (Image de David S. Goodsell, The Scripps Research Institute). L'ADN (en rouge) se déplie ce qui permet la lecture des nucléotides par l'ARN polymérase pour synthétiser l'ARNm (en rose).

Legend: A transcription factor molecule binds to the DNA at its binding site, and thereby regulates the production of a protein from a gene.

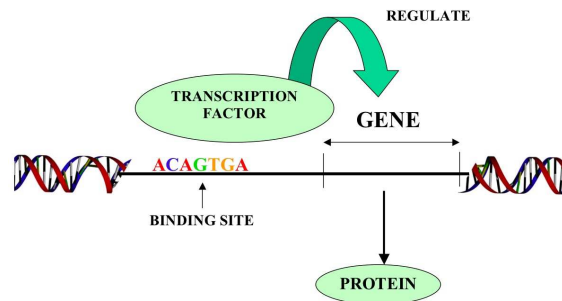


FIGURE 1.4 – Régulation du gène par le facteur de transcription (Image de Howard Hugues). Le facteur de transcription possède une séquence de reconnaissance du promoteur permettant la liaison. Une fois fixé, le facteur de transcription va recruter ou bloquer l'ARN polymérase.

ou bloque (réprime) le recrutement d'ARN polymérase (voire figure 1.4)³. Ainsi, le taux de transcription est une fonction dépendante de la concentration de certains types de facteurs de transcription avec lesquelles le promoteur du gène possède une affinité chimique.

1.1.4 Rôle de l'ARN messenger

L'ARN est l'Acide RiboNucléique. Elle est l'une des trois macromolécules majeures du vivant avec l'ADN et la protéine. Elle assure le transport de l'information génétique contenue dans l'ADN vers les ribosomes. Au niveau des ribosomes, l'ARN est traduit en protéines lors de la phase de traduction. À cause de son rôle de transport d'information, des modèles mathématiques furent établis sans tenir compte explicitement de l'ARNm. Cependant, il s'agit d'une vision trop simple de l'utilité de l'ARNm car ce dernier est soumis à des processus de régulations [31] et de dégradation non-linéaire. L'ARNm joue alors un rôle majeur pour l'apparition de comportements dynamiques, comme les oscillations [32].

3. En fait, les régions régulatrices sont parfois plus grandes et complexes que la région codante et sont appelées les régions cis-régulatrices. Elles comprennent plusieurs zones promotrices et les facteurs de transcription sont susceptibles de se fixer sur plusieurs sites [30].

1.2 Description mathématique des processus moléculaires

1.2.1 Loi d'action de masse

Les processus biologiques ayant lieu au coeur d'une cellule comportent un nombre important d'acteurs et de réactions biochimiques. La dynamique d'évolution de ces grandeurs biochimiques est obtenue par la loi cinétique d'action des masses. Cette loi est basée sur l'hypothèse que le taux de réaction est proportionnel aux concentrations des réactifs [18]. Considérons la réaction suivante :



L'évolution des variables (les concentrations) dans le temps est :

$$\frac{d[C]}{dt} = k_1[A][B] - k_{-1}[C] \quad (1.2)$$

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -k_1[A][B] + k_{-1}[C] \quad (1.3)$$

Dans ces équations, le terme $\frac{d}{dt}$ signifie la dérivée par rapport au temps t . Les paramètres k_1 et k_{-1} sont les constantes cinétiques des réactions.

1.2.2 Activité du gène

Comme les gènes dans une cellule se trouvent en faible nombre de copies, ces derniers sont considérés comme une variable statistique ayant deux états [33], éteint ou allumé (activé ou réprimé) lorsqu'il est régulé par un facteur de transcription. Dans l'état éteint, l'ARNm n'est pas synthétisé, tandis que dans l'état allumé, il y a synthèse d'ARNm. S'il y a un nombre important de transitions de l'état du gène, la dynamique du taux de transition du gène atteint une limite d'évolution quasi-continue. Dans ce dernier cas, nous modélisons la cinétique du taux de transition du gène par la loi d'action de masse en considérant le gène comme une variable continue, statistiquement équivalente à une activité moyenne dans le temps [32].

Les réactions liées à la régulation du gène par le facteur de transcription sont :



Le gène existe sous deux états : l'état libre G et l'état lié $G : F$. L'état lié résulte de la fixation du promoteur par le facteur de transcription F . Le taux de fixation dépend de la constante cinétique α . L'état libre du gène correspond au cas où le promoteur n'est pas lié au facteur de transcription F . Le taux de libération du promoteur dépend de la constante cinétique θ .

Comme le gène n'est ni créé, ni dégradé, nous avons la relation de conservation :

$$G + G : F = N \quad (1.7)$$

où N est un facteur tenant compte du nombre de copie du gène considéré (et généralement $N = 1$). ARN est la concentration de l'ARNm synthétisé par le gène G . Quand le promoteur

est occupé, la constante du taux de transcription est ρ_b et lorsque le promoteur est libre, l'ARNm est initié avec la constante ρ_f .

La nature de la régulation dépend du signe de la différence entre les valeurs ρ_f et ρ_b . Si le paramètre ρ_b est supérieur au paramètre ρ_f , alors le facteur de transcription active la transcription : le gène est actif. Sinon, le facteur réprime la transcription. Les équations dynamiques liées à la régulation du gène par le facteur de transcription s'écrivent :

$$\frac{dG}{dt} = -\frac{dG : F}{dt} = \theta G : F - \alpha G F \quad (1.8)$$

$$\frac{dARN}{dt} = \rho_f G + \rho_b G : F \quad (1.9)$$

La valeur de la variable G est compris dans l'intervalle $[0; 1]$. Si le gène est activé par le facteur de transcription, alors le gène est pleinement actif si $G : F = 1$. En fait, la modélisation par la loi d'action de masse du gène est sujet à discussions et dépend finalement du type d'étude du système biologique. Dans la partie 2.3, nous établissons des grandeurs liées à l'activité du gène. Elles donnent accès aux échelles de temps où la variable de l'activité peut être considérée comme déterministe ou stochastique.

La formation du complexe $G : F$ est parfois très rapide. Si l'échelle de temps de cette réaction est très petite par rapport aux autres processus ayant lieu, alors la réaction de formation est essentiellement à l'état stationnaire et nous pouvons poser $\frac{dG}{dt} \approx 0$. Nous en déduisons une fonction algébrique donnant l'activité du gène en fonction de la variable F et de $G : F$. L'activité du gène obéit à la fonction suivante, si nous éliminons $G : F$ grâce à la loi de conservation 1.7 :

$$G = \frac{1}{1 + \frac{F}{F_0}} \quad (1.10)$$

Le paramètre F_0 est égal au rapport $\frac{\theta}{\alpha}$. Ce paramètre correspond au seuil de régulation du gène. Lorsque la concentration du facteur de transcription est importante par rapport à ce paramètre, le gène est fortement régulé. La dynamique de création de l'ARNm s'écrit :

$$\frac{dARN}{dt} = \rho_b + \frac{\rho_f - \rho_b}{1 + F/F_0} \quad (1.11)$$

L'équation (1.11) est utilisée pour décrire la dynamique de la concentration de l'ARNm dans le chapitre 3.

1.2.3 Traduction

La traduction est l'étape de synthèse des protéines par l'ARNm. Nous supposons que le taux de traduction des protéines est simplement proportionnel au nombre d'ARN messenger, noté M , aboutissant à la réaction suivante :



La variable P est la protéine traduite selon la constante cinétique β . Dans cette réaction, l'ARNm est conservé lors de la traduction. La loi d'action de masse donne l'équation d'évolution de la concentration de protéines suivante :

$$\frac{dP}{dt} = \beta M \quad (1.13)$$

La traduction peut être régulée [34], mais nous nous limiterons à sa dynamique la plus simple (1.13) [18].

1.2.4 Dégradations

Les dégradations des protéines et des ARN sont respectivement effectuées par les protéases et les ribonucléases. Ils se passent des réactions biochimiques enzymatiques dont les équations dynamiques sont établies et vérifiées expérimentalement [35]. Le schéma réactionnel est le suivant :



Les variables P , E et C représentent respectivement la protéine (ou l'ARN) se dégradant, l'enzyme et le complexe enzyme-protéine. La quantité d'enzyme E est synthétisée et détruite de manière continue et équilibrée, respectant la loi de conservation enzymatique :

$$E + C = e_0 \quad (1.15)$$

où le paramètre e_0 est la quantité totale d'enzyme. Elle est supposée constante.

L'enzyme étant un catalyseur, le complexe C est donc instable et change rapidement d'état. Cette réaction est donc si rapide qu'elle est essentiellement à l'équilibre. Nous pouvons alors invoquer l'hypothèse d'état quasi-stationnaire [35] :

$$\frac{dC}{dt} = k_1 P E - (k_{-1} + k_2) C \approx 0 \quad (1.16)$$

L'élimination de la variable E grâce à la loi de conservation enzymatique (1.15) injectée dans l'équation (1.16) permet d'établir la fonction d'évolution de C en fonction de P :

$$C = \frac{e_0 P}{\kappa + P} \quad \text{avec} \quad \kappa = \frac{k_{-1} + k_2}{k_1} \quad (1.17)$$

Le paramètre κ est appelé la constante de Michaelis-Menten et correspond à une constante de saturation (voir figure 1.5). En substituant l'expression (1.17) de C dans l'équation dynamique de P , nous trouvons :

$$\frac{dP}{dt} = -\frac{K_2 P}{\kappa + P} \quad (1.18)$$

avec $K_2 = k_2 e_0$. L'équation (1.18) a deux régimes de fonctionnement bien distincts (voir figure 1.5) : un régime à dégradation linéaire ($[P] \ll \kappa$) et un régime à dégradation constante ($[P] \gg \kappa$). La saturation de la dégradation favorise l'apparition d'oscillations dans les réseaux biochimiques [36] comme nous le verrons plus loin (chapitre 2 page 51).

Il existe néanmoins d'autres formes de dégradation enzymatique, comme les dégradations allostériques, les processus à compétition [35]. Dans l'annexe C.2, nous étudierons la construction du modèle dynamique de dégradation enzymatique lorsqu'il y a une compétition entre plusieurs protéines pour la même enzyme.

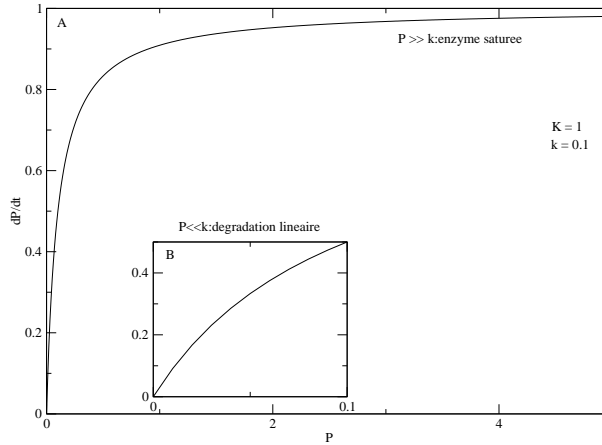


FIGURE 1.5 – Évolution de la pente $\frac{dP}{dt}$ en fonction de la concentration P à la figure A. On fixe le paramètre K à 1 h^{-1} (inverse du temps) et la constante de saturation κ à 0.1 nM (nanoMole). On distingue deux régimes : lorsque $[P] \gg \kappa$, la pente est égale à une constante, lorsque $[P] \ll \kappa$, la dégradation est linéaire (voir encadrée B).

1.2.5 Coefficient de Hill

L'efficacité de la liaison peut dépendre du nombre de protéines se fixant sur l'enzyme (ou au promoteur d'un gène) et induit ainsi une coopérativité entre les protéines. Cette coopérativité peut se faire grâce à la formation d'un polymère. Pour quantifier cet effet, le coefficient de Hill n apparaît. Ce coefficient correspond au nombre de protéines identiques se liant pour former un polymère. La fonction de régulation (1.11) tenant compte d'une telle coopérativité devient alors :

$$\frac{dARN}{dt} = \rho_b + F_0^n \frac{\rho_f - \rho_b}{F_0^n + F^n} \quad (1.19)$$

Cette dernière forme est utilisée dans l'élaboration du modèle biochimique du chapitre 3, à la page 73.

Le coefficient de Hill est un paramètre important dans la dynamique d'un réseau biochimique. L'augmentation de ce coefficient élargit le domaine paramétrique permettant les oscillations des concentrations des protéines [18, 37].

1.2.6 Autres processus

De nombreux autres processus biologiques peuvent être importants dans la modélisation de réseaux bio-chimiques, comme le transport nucléaire [38] ou les réactions de phosphorylation [39]. Cependant, ces réactions sont parfois spécifiques à certains modèles biologiques. Sans preuves expérimentales rendant ces processus nécessaires pour la modélisation, il est plus judicieux d'étudier un réseau en commençant par la structure la plus simple (transcription, traduction et dégradation). Si les résultats obtenus ne correspondent pas aux résultats biologiques, alors l'incorporation de processus supplémentaires est envisageable.

1.2.7 Effets stochastiques

Bien que cette thèse ne traite pas de modèles stochastiques, une part de cette approche est expliquée en raison de la nature des réactions chimiques ayant cours au sein de la cellule. Ainsi, cette méthode est certainement la plus réaliste pour de nombreuses descriptions et est largement utilisée [17].

Les réactions précédentes sont écrites en supposant que l'évolution du nombre de molécules suit une loi déterministe et continu et sont modélisées par des équations différentielles. Mais de manière rigoureuse, ces évolutions temporelles ne sont pas continues puisque le nombre de molécules change de manière discrète. De plus, si on suppose que les réactions ont lieu dans un milieu relativement homogène et en équilibre thermique, il n'y a pas de direction privilégiée du mouvement des molécules. La rencontre de ces dernières est donc due aux fruits du hasard et l'évolution des grandeurs n'est donc pas déterministe [40]. A priori, l'approximation par équations différentielles est acceptable s'il y a au moins l'hypothèse d'un grand nombre de molécules dans le milieu [41]. Mais cet hypothèse n'est pas toujours valable dans les cellules car le nombre mesuré de certaines bio-molécules, comme les facteurs de transcription ou les gènes, est en faible quantité [2]. La méthode la plus adaptée semble être la description par les équations maîtresses [41]. Ces équations donnent l'évolution temporelle de la probabilité du nombre de molécules à l'instant t . Cependant, la résolution numérique de ces équations étant extrêmement difficile, on lui préfère son équivalent en algorithme, l'algorithme de simulation stochastique de Gillespie [40]. Cet algorithme engendre des séquences de réactions chimiques se produisant au hasard suivant les lois de probabilité de la cinétique chimique.

Le choix de la méthode peut être déterminant pour l'interprétation correcte du modèle biologique. Les modèles stochastiques de systèmes non linéaires sont plus généraux [41] et montrent parfois des effets ou des propriétés absentes lors de l'analyse déterministe [42]. Cependant, la simulation stochastique n'est pas la méthode la plus adaptée pour nos travaux car nos résultats théoriques obtenus par des équations différentielles ordinaires restent pertinent pour l'approche stochastique (voir section 2.4.3.2), ou bien décrivent parfaitement l'expérience (voir section 3.3.6).

1.3 Dynamique induite par les réseaux de régulation génétique

Dans cette partie, nous analysons la dynamique non-linéaire des réseaux génétiques possédant des boucles de rétroactions.

1.3.1 Réseau génétique à boucle de rétroaction

1.3.1.1 Formation d'un réseau biochimique

Les techniques de séquençages de l'ADN largement automatisées mettent à jour le très grand nombre d'acteurs moléculaires (ARN, protéines venant de gène identifié) agissant au cœur de la cellule. Mais ce séquençage ne prédit pas avec exactitude l'expression génétique qui dépend fortement des interactions biomoléculaires. Ces dernières régulent les quantités et les fonctions moléculaires au cours du temps [30]. Différents techniques (méthode SAGE, puces à ADN, gel de protéines) permettent un accès quantitatif et qualitatif aux concentrations ou bien mettent à jour les interactions possibles (puces à ADN double-brin, technique du double hybride, voir figure 1.6 ...).

Une représentation schématique des réactions biochimiques est possible en considérant l'ensemble { gène, ARN, protéine } comme un nœud. La détection d'une interaction entre deux ensembles (ou même d'un ensemble avec lui-même) correspond à un lien, comme une interaction protéine-protéine ou protéine-gène. La liaison de ces ensembles donne naissance à un réseau d'interactions (voir figure 1.6).

La compréhension du rôle d'un réseau ne dépend pas que de la connaissance des acteurs et de l'interaction. En effet, la nature des interactions est multiple et variée (section 1.2) : les interactions peuvent être la dégradation, la régulation transcriptionnelle, la traversée nucléaire [38, 43], la protection d'une protéine contre la dégradation [44] . . . Il faut donc déduire la nature exacte de l'interaction pour analyser un réseau car les fonctions mathématiques les représentant sont différentes.

Comme les réseaux forment des systèmes topologiques complexes [45, 46] rendant impossible la compréhension par l'intuition, la méthode d'analyse la plus fiable est l'étude de la dynamique non-linéaire de ces réseaux [19] afin de combler le fossé entre les données génomiques et la caractérisation fonctionnelle des macromolécules. La nature de la régulation des facteurs de transcription (activation ou répression) joue un rôle prépondérant dans la dynamique de la concentration des protéines des réseaux génétiques [47, 37]. Nous décrivons quelques exemples dynamiques simples provoqués par différentes régulations.

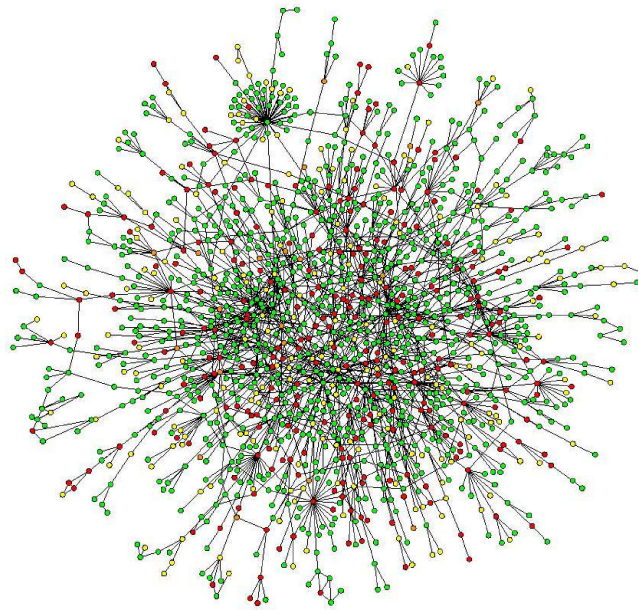


FIGURE 1.6 – (Illustration de *H. Jeong* [48]). De nouvelles techniques d'identification des interactions entre protéines ont permis d'obtenir beaucoup d'informations sur les réseaux de la levure [49, 50]. Cela nous permet de comprendre le rôle des réseaux dans les processus cellulaires. Nous montrons un exemple de réseau de la levure établi par Jeong et al. [48]. Sur la figure ci-dessus, chaque point représente un noeud {gène, ARN, protéine}. Un lien entre deux noeuds représente une interaction entre protéines.

1.3.1.2 Régulation simple

Une régulation simple est rencontrée lorsqu'un facteur de transcription Y active un gène X. La transcription est induite par une liaison de Y avec une molécule, par une transduction de signal ou autre. L'ARNm, noté M, du gène X est synthétisé à taux constant et augmente jusqu'à une valeur correspondante à l'état stationnaire, c'est-à-dire au moment où la dérivée du produit M est nulle :

$$\frac{dM}{dt} = \rho - \delta M(t) \approx 0 \quad (1.20)$$

Nous voyons alors que la concentration est égale au rapport entre le taux de production ρ et le taux de dégradation δ .

Lorsque l'activation s'arrête ($\rho = 0$), après avoir atteint l'état d'équilibre, la concentration $M(t)$ décroît suivant une loi exponentielle :

$$M(t) = \frac{\rho}{\delta} e^{-\delta t} \quad (1.21)$$

Il est en fait courant que certains gènes soient transcrits à taux constant, mais des situations plus complexes interviennent en présence de facteurs de transcription.

1.3.1.3 Auto-régulation

Deux types d'auto-régulation existent et sont décrites par la suite : l'auto-régulation négative et positive.

Dans l'auto-régulation négative, un facteur de transcription réprime l'expression de son propre gène X [51, 52, 53]. La fonction de transcription donne la synthèse de la quantité d'ARNm, noté M, du gène X :

$$\frac{dM(t)}{dt} = \mu + \frac{\lambda}{1 + (P/P_0)^n} \quad (1.22)$$

La variable P représente la quantité de protéine du gène X. Le paramètre P_0 est le seuil de répression. Lorsque le gène est complètement réprimé, le taux basal de transcription est μ et lorsque le gène n'est pas réprimé, le taux basal de transcription est $\mu + \lambda$. L'exposant n est le coefficient de Hill.

Commun dans la bactérie *E. Coli* [51, 54] et dans certains eucaryotes [55], cette régulation a deux effets vérifiés expérimentalement [51]. Elle provoque une vitesse de convergence vers l'état stationnaire de la protéine X plus rapide que pour une régulation simple [51]. Cela est avantageux pour rendre les dispositifs cellulaires de réparation plus rapide, comme le système de réparation de l'ADN de *E. Coli*, dans lequel le régulateur maître *LexA* régule son propre gène [56]. Elle réduit également la variation du nombre de certaines protéines entre différentes cellules d'une même population, diminuant ainsi le bruit d'amplitude [53, 56, 57].

Dans l'auto-régulation positive, le facteur de transcription active la transcription de son propre gène :

$$\frac{dM(t)}{dt} = \mu + \lambda \frac{(P/P_0)^n}{1 + (P/P_0)^n} \quad (1.23)$$

Le paramètre P_0 est le seuil d'activation. Lorsque le gène est complètement activé, le taux basal de transcription est $\mu + \lambda$. Si la quantité de protéine P est très faible devant le seuil d'activation, le taux basal de transcription est μ .

Il a été montré théoriquement [58] et expérimentalement [59] que la convergence vers l'état stationnaire est moins rapide que pour la régulation simple. Cette régulation mène à la bistabilité (ou à une distribution bimodale) des concentrations de la protéine régulatrice, c'est-à-dire qu'il y a une grande variabilité de la concentration de la protéine entre différentes cellules : tandis que certaines cellules ont une faible concentration, d'autres cellules ont une concentration élevée. Cette distribution bimodale permet à l'organisme de conserver plusieurs phénotypes au sein d'une même population, afin d'avoir une meilleure adaptation à l'environnement [60, 59, 61, 62].

1.3.1.4 Boucle de rétroaction

Un réseau de régulation génétique est formé d'un ensemble de gènes interagissant entre eux grâce aux régulations impliquant leur facteur de transcription. La régulation du gène par sa propre protéine constitue une boucle de rétroaction transcriptionnelle et suffit à créer un réseau de régulation génétique. En général, un réseau génétique contient de multiples régulations transcriptionnelles formant une boucle car un gène régule indirectement sa propre expression. Le gène produit un facteur de transcription régulant un autre gène, ce dernier produit lui-même un facteur de transcription régulant un tiers gène, etc. jusqu'à ce que le gène original soit régulé par un facteur de transcription d'un gène impliqué dans la boucle [63, 64]. Dans un tel réseau génétique, si une boucle de rétroaction est faite de n régulations, la boucle de rétroaction est dite négative si le nombre de régulations répressives (facteur de transcription réprimant l'expression du gène) est impair. Autrement, la boucle de rétroaction est positive [65].

Un réseau génétique transcriptionnel peut facilement contenir de nombreux gènes (allant de la dizaine à la centaine) dans divers organismes [66, 55, 67, 68, 69, 70]. La détermination de la dynamique d'un réseau n'est pas simple car, même si nous identifions clairement les régulations présentes, de simples interactions ou modulations de paramètres cinétiques modifient éventuellement la dynamique.

Nous discutons maintenant d'un modèle théorique de réseau génétique possédant une boucle de régulation transcriptionnelle. Sa topologie est fortement représentée dans la levure. La particularité de ce réseau est la possibilité d'avoir de nombreux comportements ou états dynamiques juste en changeant la valeur de deux paramètres liés aux taux de transcription d'un gène.

1.3.2 Multiple dynamique d'un réseau génétique : le "Mixed Feedback Loop"

Le "Mixed Feedback Loop" ou MFL est un réseau génétique simple étudié par François et al. [71]. Il a été théoriquement trouvé à partir d'un algorithme de développement et d'évolution de réseau génétique [72]. La topologie de ce réseau est présente en grand nombre dans la levure [70]. Une variation de certains paramètres de ce réseau fait passer d'un commutateur bistable à un oscillateur. Ce système est donc idéal comme exemple de modèle biologique et mathématique afin de montrer la richesse dynamique des réseaux.

1.3.2.1 Présentation des équations du MFL

Le MFL est un réseau à deux gènes g_a et g_b produisant respectivement la protéine A et la protéine B. La protéine A régule la transcription du gène g_b et interagit directement avec la protéine B pour former un dimère AB . François et al. considèrent un volume cellulaire unitaire et donc la concentration d'une protéine A, notée $[A]$, est le nombre de protéines dans la cellule. Un schéma du réseau est donné à la figure 1.7. La modélisation se fait par la loi cinétique d'action de masse [18]. La régulation de la transcription du gène g_b par la protéine A donne :

$$\frac{d[g_b]}{dt} = \theta[g_b : A] - \alpha[g_b][A] = \theta(1 - [g_b]) - \alpha[g_b][A] \quad (1.24)$$

$$\frac{d[r]}{dt} = \rho_f[g_b] + \rho_b(1 - [g_b]) - \delta_r[r] \quad (1.25)$$

Le gène est noté $[g_b]$. $[r]$ est la concentration de l'ARNm du gène g_b . Le gène est supposé exister sous deux formes : une forme avec A lié au promoteur $[g_b : A]$ et une forme libre $[g_b]$. Le gène se conserve : $[g_b] + [g_b : A] = 1 \Leftrightarrow [g_b : A] = 1 - [g_b]$. En injectant la dernière forme dans le second membre de l'équation (1.24), il vient la partie de droite de (1.24). Les protéines se lient au promoteur du gène $[g_b]$ selon la constante cinétique α et se libèrent avec le taux θ . Lorsque A est lié au promoteur de $[g_b]$, le taux de transcription est ρ_b et si le promoteur du gène est libre, le taux de transcription est ρ_f . Si $\rho_b > \rho_f$, alors le gène est activé (activation transcriptionnelle) et si $\rho_f > \rho_b$, alors le gène est réprimé (répression transcriptionnelle). L'ARNm r se dégrade avec une constante de dégradation δ_r . Les équations suivantes décrivent les synthèses, la dimérisation et la dégradation des protéines $[A]$ et $[B]$:

$$\frac{d[A]}{dt} = \rho_A - \gamma[A][B] + \frac{d[g_b]}{dt} - \delta_A[A] \quad (1.26)$$

$$\frac{d[B]}{dt} = \beta[r_b] - \gamma[A][B] - \delta_B[B] \quad (1.27)$$

$$\frac{d[AB]}{dt} = \gamma[A][B] - \delta_{AB}[AB] \quad (1.28)$$

Les variables $[A]$, $[B]$ et $[AB]$ sont respectivement les concentrations des protéines A, B et du dimère AB. Dans ce modèle, la protéine A est produite avec un taux de base ρ_A , sans régulation du gène g_A . La protéine A et la protéine B se dimérisent avec le taux γ . Si la protéine A est liée au gène g_B , B ne peut pas interagir avec A. La protéine A se dégrade avec le taux de dégradation δ_A et la protéine B se dégrade avec le taux δ_B . L'équation (1.27) décrit la traduction de l'ARN messager $[r]$ avec un taux β . La dernière équation (1.28) décrit la dynamique du complexe formé par A et B. Paul François et al. ont montré que la dynamique dépendait de la valeur de deux paramètres redimensionnés [71], les deux taux de transcription du gène g_b .

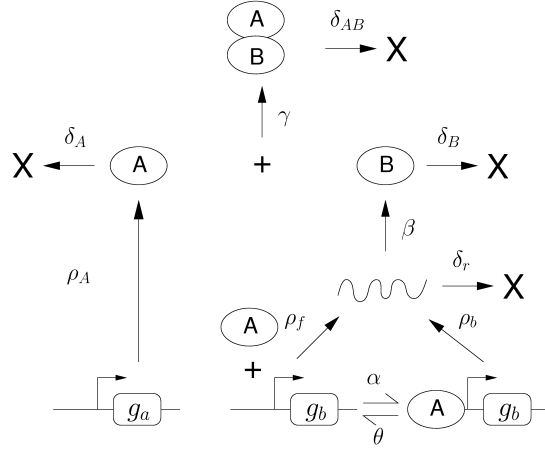


FIGURE 1.7 – Schéma du “Mixed Feedback Loop” (illustration de *François* [71]). Le gène g_a transcrit avec un taux constant ρ_a la protéine A. La protéine A se dégrade linéairement avec le taux δ_A . La protéine A régule le gène g_b en se liant avec le taux α . La protéine se libère avec le taux θ . Le taux de transcription est soit ρ_f si le gène est libre, ou soit ρ_b si le gène est lié. L’ARN (en trait ondulé) se dégrade avec le taux δ_r et est traduit en protéine B avec la constante β . La protéine B se dégrade selon la constante δ_B . La protéine A et B peuvent se dimériser avec le taux γ . Le dimère se dégrade avec le taux δ_{AB} .

1.3.2.2 Diagramme de bifurcation

Le système d’équations de la MFL est redimensionné pour faire apparaître les deux paramètres suivants :

$$\rho_0 = \frac{\beta\rho_f}{\delta_r} \times \frac{1}{\rho_A} = \frac{\beta\rho_f}{\rho_A\delta_r} \quad (1.29)$$

$$\rho_1 = \frac{\beta\rho_b}{\delta_r} \times \frac{1}{\rho_A} = \frac{\beta\rho_b}{\rho_A\delta_r} \quad (1.30)$$

Le comportement dynamique du réseau dépend de ces deux paramètres. Trois comportements sont observables. L’existence d’un seul état stable, un état de bistabilité et le comportement oscillant (figure 1.8).

L’état stationnaire monostable de la concentration de protéine B est bas. Sur la figure (1.8), cela correspond au domaine tel que $\rho_0 < 1, \rho_1 < 1$. Dans ce cas de figure, les deux taux de transcription renormalisés de l’ARN du gène g_b , qui sont $\beta\rho_f/\delta_r$ et $\beta\rho_b/\delta_r$, sont plus bas que ρ_A . La concentration de la protéine A est plus haute que celle de B, le promoteur g_b est occupé par la protéine A. La protéine B se dimérise avec la protéine A et se retrouve en faible nombre, quelque soit la nature de la régulation.

Dans le cas monostable avec B haut (1.8), le domaine est délimité par les conditions suivantes : $\rho_0 > 1$ et $\rho_1 > 1$. En suivant le même raisonnement que précédemment, les taux de production $\frac{\beta\rho_b}{\delta_r}$ et $\frac{\beta\rho_f}{\delta_r}$ sont supérieurs à ρ_A . La protéine A se dimérise avec la protéine B et ne régule plus le gène. De plus, le faible nombre de protéines A se dimérise avec la protéine B, ce qui maintient l’état à concentration $[B]$ élevée.

L’état bistable est dû à la répression transcriptionnelle du gène g_b par la protéine A, soit $\rho_0 > 1 > \rho_1$ (voir figure 1.8). Deux états stationnaires stables coexistent et le choix de l’état

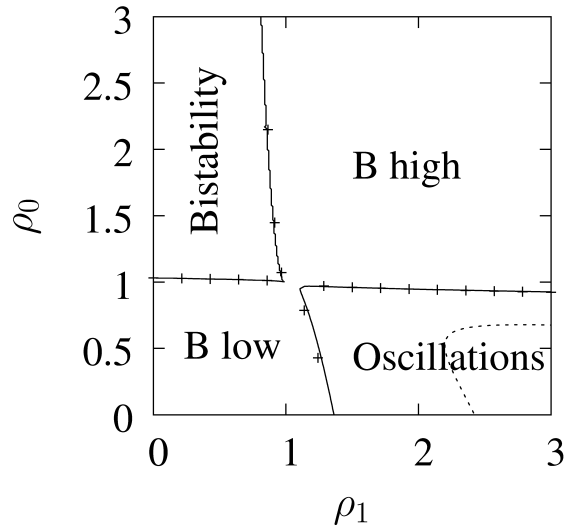


FIGURE 1.8 – Espace des paramètres de la MFL (déterminé par *François* [71]). Délimitation des frontières entre les 3 comportements dynamiques possibles. Il y a un état stable avec une seule solution unique : soit une concentration de B faible pour $\rho_0 < 1, \rho_1 < 1$ ou soit une concentration B haute pour $\rho_0 > 1, \rho_1 > 1$. Il y a un état bistable pour $\rho_0 > 1, \rho_1 < 1$ et un état oscillant pour $\rho_0 < 1, \rho_1 > 1$. Les inégalités sont approximatives. En ligne pleine : approximation analytique des frontières. Signes + : détermination numérique des frontières.

stationnaire dépend des conditions initiales (ou d’une éventuelle perturbation suffisante pour commuter entre les deux états). Prenons la situation où la concentration de A est faible au départ. Statistiquement, le gène g_b est dans l’état libre et est peu réprimé. Le taux de transcription normalisé $\frac{\beta\rho_f}{\delta_r}$ sera supérieur à ρ_A entraînant une production de la protéine B plus grande que celle de la protéine A. Le faible nombre de protéine A se dimérise rapidement en raison de la haute concentration de la protéine B. Cette complexation empêche la protéine A de réprimer le gène et maintient donc l’état de basse concentration de la protéine A et l’état de concentration élevée de B.

Si la condition initiale est une concentration de protéine A assez élevée, alors la protéine A réprime assez fortement la transcription du gène g_b . Le taux de production de la protéine A est plus élevé que celui de la protéine B. La protéine B est dimérisée par la protéine A et il se maintient une concentration de protéines A suffisante pour réprimer encore le gène. Cela stabilise l’état où la concentration de la protéine A est élevée et la concentration de la protéine B est faible.

Pour avoir l’état oscillant, la protéine A doit activer la transcription du gène g_b . La concentration de la protéine B s’élève et il y a une complexation avec la protéine A. Cette dernière complexation joue alors le rôle d’une régulation négative : la concentration de A diminue et donc n’active plus le gène. La concentration de la protéine B (qui continue à se complexer) diminue jusqu’à un seuil où cette fois la concentration de A réaugmente. Le comportement décrit correspond donc bien à des oscillations entretenues.

Un réseau génétique très simple peut avoir plusieurs comportements dynamiques, comme les oscillations ou la bistabilité. La présence des oscillations est vitale pour les organismes vivants et leur étude est importante pour comprendre de nombreux processus physiologiques.

La construction de réseaux génétiques artificiels peut nous permettre également de comprendre les oscillations dans une situation réelle, c'est-à-dire en présence de bruit absent dans l'étude précédente (1.2.7).

1.3.3 Réseaux artificiels oscillant : le Repressilator

Certains réseaux génétiques sont des oscillateurs génétiques parce que les concentrations et les activités des acteurs le constituant oscillent au cours du temps. Les oscillateurs génétiques sont relativement répandus car ils sont à l'origine de certains rythmes biologiques [73]. Ils sont également impliqués dans le cycle cellulaire [74, 75], ou dans le cycle du p53, un suppresseur de tumeur [76, 77]. La complexité de certains réseaux génétiques est telle qu'il n'est pas toujours évident d'identifier les gènes centraux et la nature des interactions impliquées dans l'apparition d'oscillations entretenues. Les réseaux artificiels nous permettent alors de comprendre quels sont les rôles précis des différents constituants d'un réseau génétique et dans quelles conditions nous avons des oscillations entretenues et robustes. Nous remarquons en particulier la présence d'une boucle de rétroaction négative [78] dans ces réseaux oscillants.

Leibler et al. ont réalisé un réseau génétique synthétique relativement simple à partir de trois gènes de la bactérie *Escherichia coli* [79] : les gènes λcl , $Lacl$ et $TetR$ (voir figure 1.9). Dans ce réseau, la protéine de λcl réprime le gène de $Lacl$, la protéine $Lacl$ le gène $TetR$ et la protéine $TetR$ le gène λcl . Cette boucle de rétroaction est négative car le nombre de régulations négatives est impair. Le gène $TetR$ réprime la transcription du gène rapporteur de la protéine rapportrice nommée GFP (Green Fluorescence Protein). Cette dernière émet de la lumière permettant une observation par microscopie fluorescente de l'activité transcriptionnelle au fil du temps.

La réalisation expérimentale a permis d'aboutir à des oscillations relativement bruitées (voir figure 1.11) dont l'amplitude moyenne s'élève à cause de l'accumulation de la GFP. La présence d'un tel bruit s'explique par la nature discrète des molécules et leurs interactions stochastiques (voir section 1.2.7 [40]), mais aussi par l'absence de processus spécifique de réduction du bruit dans leur réseau génétique. Les auteurs remarquent qu'une boucle de rétroaction avec une régulation négative et positive sont nécessaires pour réduire le bruit dans les circuits oscillants [80]. Ceci est d'ailleurs un argument pour la modélisation déterministe des réseaux génétiques sous-jacent des rythmes biologiques non synthétiques, relativement résistants à ce type de bruit [81] (sauf si on désire découvrir précisément les processus de résistance au bruit). La dynamique de ce réseau génétique est décrite par un système de 6 équations différentielles :

$$\frac{dM_i(t)}{dt} = -m_i + \frac{\alpha}{1 + (p_j)^n} + \alpha_0 \quad (1.31)$$

$$\frac{dp_i(t)}{dt} = -\beta(p_i - m_i) \quad (1.32)$$

avec $i = Lacl, TetR$ et λcl et $j = \lambda cl, Lacl$ et $TetR$. Les équations sont sans dimension. Le terme α_0 est le taux basal de production des protéines par le gène i en présence de la protéine P_j . En absence de protéine P_j , ce taux devient $\alpha_0 + \alpha$. Le paramètre β est le rapport entre le taux de dégradation de la protéine et de l'ARNm. Le paramètre n est le coefficient de Hill (dans les simulations, il est toujours supérieur ou égale à 1). Le temps est redimensionné pour que le taux de dégradation de l'ARN soit égal 1.

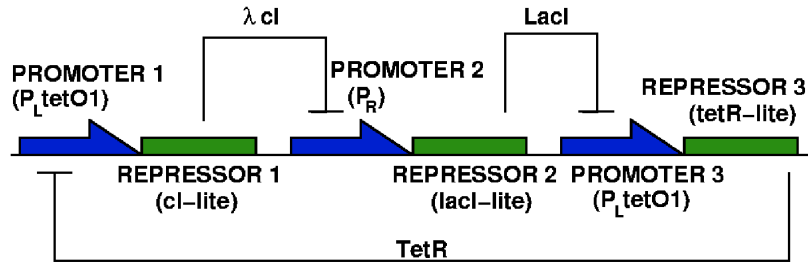


FIGURE 1.9 – Réseau de régulation génétique du repressilator (Image de Wikipedia). Trois gènes λcl , $Lacl$ et $TetR$ sont connectés entre eux et forment une boucle de rétroaction négative.

Les résultats des simulations numériques sont à la figure 1.10. Le modèle déterministe produit des oscillations avec des paramètres de l'ordre de grandeur des valeurs évaluées expérimentalement. D'autres simulations tiennent compte de la nature discrète des molécules et de la nature stochastique des réactions (voir section 1.2.7, simulations stochastiques à la figure 1.10). Les résultats s'interprètent grâce à une fonction d'autocorrélation qui diminue fortement pour des délais de plus de deux périodes, ce qui signifie que la durée et l'amplitude entre chaque cycle sont relativement fluctuantes pour une seule cellule.

Les auteurs déduisent que les oscillations en stochastique persistent avec une large variabilité. L'étude théorique aboutit aussi à la conclusion que les oscillations sont favorisées lorsque le coefficient de Hill du facteur de transcription est grand avec le terme α_0 faible [79].

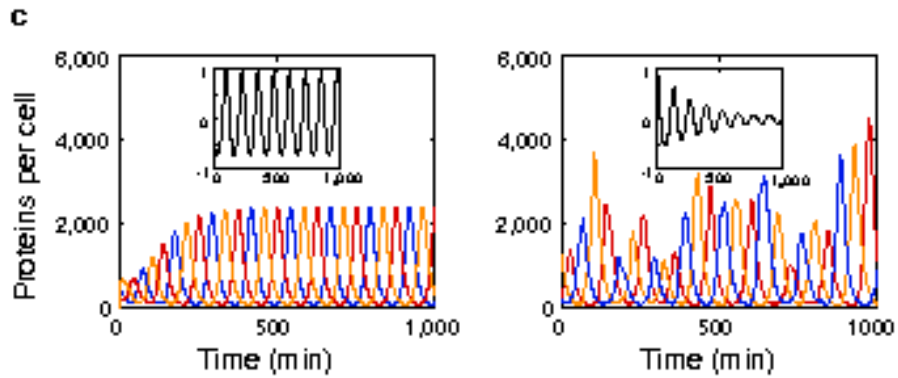


FIGURE 1.10 – Tirée de [79]. A droite : intégration numérique du modèle déterministe du repressilator. A gauche, simulation stochastique du même modèle. On montre la fonction d'autocorrélation des signaux dans le petit encadré (figures extraites de [79]).

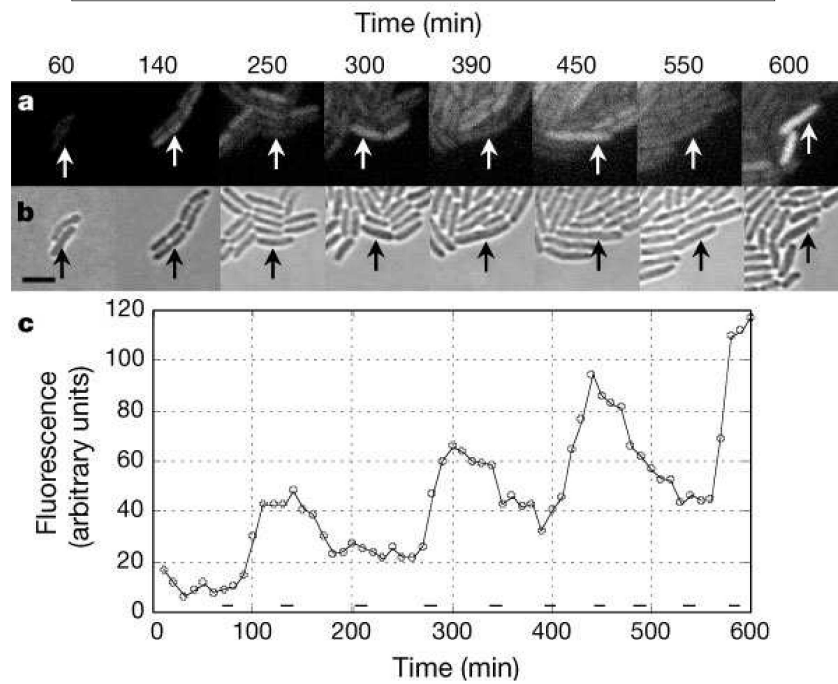


FIGURE 1.11 – Tirée de [79]. Figure du haut : évolution de la population de bactéries *Escheria coli* au cours du temps. Figure du bas : mesure de la fluorescence donnant la variation des quantités de protéines (figures extraites de [79]).

1.4 Horloges circadiennes

De nombreux processus physiologiques ont un rythme périodique, aux fluctuations près. Ils sont générés par des mécanismes biologiques sensibles à des stimuli extérieurs cycliques. Cependant, certains de ces rythmes persistent en absence de ces stimuli. Ces rythmes ont des propriétés intéressantes. Ils sont assez robustes pour résister aux fluctuations et sont capables d'être entraînés par une variation périodique de l'environnement (cycle jour/nuit, saison...). L'un des rythmes les plus remarquables est le rythme circadien.

1.4.1 Définition du rythme biologique

Les rythmes biologiques naissent de l'interaction d'acteurs biologiques (gène, protéine, ARNm ...) entraînant leur propre variation périodique dans le temps. Ces interactions biochimiques forment une horloge biologique. Ces horloges sont présentes au coeur des cellules, d'organes ou des tissus. Certaines d'entre eux contrôlent des fonctions d'ordre physiologique (sécrétions hormonale, rythme cardiaque), ou comportementales [24, 82, 83].

Les horloges possèdent de nombreuses voies d'entrée sensibles aux signalisations extérieures. Elles identifient les variations de l'environnement ou bien communiquent avec des horloges placées dans d'autres cellules ou tissus. Ainsi, une horloge biologique peut se synchroniser par couplage avec différents cycles naturels (cycle de lumière jour/nuit, cycle des saisons...), chimiques ou physiques (variation de la température). Ces derniers sont appelés également facteurs d'entraînement.

Certains rythmes biologiques n'existent que grâce au couplage avec des rythmes de l'environnement. Par exemple, la position des bancs de certains poissons dans l'océan dépend uniquement de stimuli de la luminosité du milieu [82].

Les horloges endogènes sont une autre classe d'horloge car leur existence ne dépend pas que des stimuli extérieurs. Ce rythme biologique est sensible à un cycle extérieur donné mais persiste en l'absence de cette dernière modulation. Cette horloge biologique est un réseau génétique possédant une dynamique non-linéaire. Ce type de rythme possède donc deux périodes. Lorsque l'horloge est isolée des stimuli, la période du rythme est dite en libre cours. Sinon, le rythme prend la période du cycle d'entraînement extérieur. Cela signifie que les facteurs d'entraînement ne créent pas le rythme mais en contrôlent la période et la phase.

Les rythmes biologiques couvrent une large gamme de période, qui peut être beaucoup plus grande ou petite que 24 heures. On dit que les rythmes sont :

- ultradiens si la période est inférieure à 24 heures ;
- circadiens si la période est d'environ 24 heures ;
- infradiens si la période est supérieure à 24 heures ;

Le tableau 1.12 donne des exemples de rythmes biologiques et leurs périodes. Nous remarquons que certains rythmes biologiques, comme le cycle mitotique, sont ultradiens ou circadiens. En fait, la durée de ce rythme dépend de l'espèce considérée.

Rythmes biologiques	Période (T)
Rythmes neuronaux	0.01-10 s
Rythmes cardiaques	1 s
Oscillations du calcium	1 s - 1min
Cycle mitotique	10 min - 24 h
Rythme circadien	20 - 28 h
Rythmes hormonaux	10 min - 3 h
Cycle ovarien	28 jours
Rythmes annuels	1 an

FIGURE 1.12 – Différents exemples de rythmes biologiques. Extraits de [1].

1.4.2 Définition du rythme circadien

On observe chez la plupart des organismes vivants des oscillations d'un certain nombre de grandeurs physiologiques sur 24 heures (températures corporelles, rythme veille-sommeil, état métabolique...) Ces oscillations ont pour but de permettre à l'organisme d'anticiper les variations périodiques de l'environnement dues à l'alternance jour nuit qui résulte de la rotation terrestre. On sait depuis quelques années que les rouages de ces horloges sont constitués de circuits de régulations génétiques [84] alors que le rythme circadien a été observé et décrit qualitativement depuis plusieurs siècles [85].

Le rythme circadien est contrôlé par l'horloge circadienne. En absence de modulation cyclique par ces facteurs d'entraînement, la période du rythme est d'environ 24 heures. Le principal facteur d'entraînement est la variation de l'intensité lumineuse. Ainsi, le cycle naturel d'alternance entre le jour et la nuit entraîne le rythme à une période de 24 heures. Lors d'expériences, on peut soumettre une cellule à une alternance de 12 heures de lumière et de 12 heures de nuit (appelé cycle Light/Day 12 :12 ou LD 12 :12). L'horloge circadienne se synchronise à une période de 24 heures exactement. La synchronisation de l'horloge peut se faire pour des périodes supérieures ou inférieures à 24 heures dans certaines limites. Mais au-delà, l'horloge ne se synchronise plus [86].

La température est un autre facteur d'entraînement remarquable de l'horloge circadienne. À différentes températures constantes, la période de l'horloge circadienne est indépendante de ces conditions. Cependant, la variation cyclique de la température entraîne l'horloge. Les facteurs d'entraînement de l'horloge circadienne sont appelés les Zeitgebers [87].

1.4.3 Ubiquité et topologie de l'horloge circadienne

Ce rythme a été mis en évidence dans de très nombreux organismes [73, 88, 84], comme les mammifères [4], les plantes [89], les insectes [90], les champignons [91] et des organismes uni-cellulaire [92, 93, 94] (voir tableau 1.13). Cette horloge est en général présente dans chaque tissu ou cellule d'un organisme.

Chez les insectes ou les plantes, les horloges des différents organes ou tissus sont indépendantes et se synchronisent directement à la lumière. En revanche, chez les mammifères, le rythme circadien est contrôlé par l'horloge des groupes de neurones appelés noyaux suprachiasmatiques (NSC). Cette horloge du cerveau synchronise les horloges périphériques présentes dans les autres organes pour les entraîner au rythme circadien. Lors de long voyage transméri dien, cette horloge centrale est rapidement remise à l'heure par les signaux lumineux afin d'ajuster notre cycle d'éveil et de sommeil au nouveau fuseau horaire. Cependant, les horloges des tissus périphériques passe par une phase de resynchronisation engendrant un certain déconfort [95]. En outre, une mutation du gène *PER2* chez l'être humain affecte fortement l'horloge circadienne en provoquant un important décalage du sommeil. Le malade ressent une grande fatigue très tôt le soir (entre 6 et 8 heures) et se réveille au milieu de la nuit (entre 1 et 3 heures du matin) [96]. Le rythme circadien contrôle également le cycle cellulaire de nombreux êtres vivants ou cellules. Le dérèglement de ce cycle, pouvant se transformer en cancer, est parfois dû à une défaillance du rythme circadien [16]. La compréhension de l'horloge circadienne peut donc être à l'origine d'un nouveau champ thérapeutique permettant de mieux cibler les causes de certains troubles ou maladies [95].

La particularité du réseau de toutes horloges est la présence d'une boucle à rétroaction négative. Il s'avère que cette boucle est primordiale pour l'obtention des oscillations [65]. Ce type de boucle peut agir sur les protéines grâce à des régulations post-traductionnelles [7]. Cependant, une grande part des horloges circadiennes sont à régulation transcriptionnelle : la rétroaction intervient en régulant la transcription des gènes de l'horloges [84].

Dans la partie suivante, nous présentons des similitudes entre l'oscillateur circadien et l'oscillateur auto-entretenu. Ensuite, nous montrons des exemples d'horloge circadienne possédant une boucle à rétroaction négative transcriptionnelle.



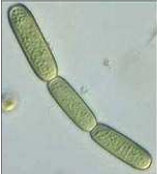



	Organisme	Rythmes circadiens	Principaux gènes de l'horloge
	<i>Neurospora</i> (champignon)	Conidiation Expression génétique	<i>frq, wc1, wc2</i>
	<i>Drosophila</i> (mouche)	Activité locomotrice Expression génétique	<i>per, tim, cyc, clk</i>
	<i>Synechococcus</i> (cyanobactérie)	Division cellulaire Photosynthèse Expression génétique	<i>kaiA, kaiB, kaiC</i>
	<i>Gonyaulax</i> (algue)	Division cellulaire Photosynthèse Bioluminescence	<i>lbp</i>
	<i>Euglena</i> (euglène)	Division cellulaire Photosynthèse Phototaxie	
	<i>Arabidopsis</i> (plante)	Mouvement des feuilles	<i>lhy, cca1, toc1</i>
	<i>Homme, souris, rat, hamster</i> (mammifères)	Activité locomotrice Système neuronal Expression génétique	<i>mper1,2,3, mtim, bmal1, clock (tau)</i>

FIGURE 1.13 – Récapitulatif de certains organismes possédant une horloge circadienne, les fonctions biologiques contrôlées par l'horloge et les gènes centraux de l'horloge (D'après Gonze, "les horloges biologiques", <http://student.ulb.ac.be/dgonze/CLOCKS/horloges.html>).

1.4.4 Analogie avec un oscillateur auto-entretenu

Nous décrivons certaines propriétés dynamiques de l'oscillateur auto-entretenu que possèdent également les oscillateurs circadiens. L'oscillateur de Van der Pol est pris comme exemple d'oscillateur auto-entretenu.

L'oscillateur auto-entretenu est un système possédant une source d'énergie continue et non rythmique. Ce système alterne entre deux statuts : la capacité à compenser les pertes suivie d'une incapacité à les compenser. La transition entre les deux situations, responsable des oscillations, se passe lorsque l'amplitude dépasse une valeur seuil.

L'oscillateur de Van der Pol est un modèle d'oscillateur auto-entretenu classique [97, 22, 98]. Il est obtenu en plaçant en parallèle une inductance, un condensateur et une diode à effet tunnel. Une alimentation continue est placée sur une des branches. Cette oscillateur est non-linéaire au niveau de son amortissement. Ce dernier a une influence fondamentale sur la limitation des amplitudes de l'oscillateur, lui permettant de passer d'un régime quasi-sinusoïdal vers un régime relaxatif (défini par la suite) grâce à la variation d'un seul paramètre lié à l'amortissement. L'oscillateur peut être soumis à un forçage externe périodique. L'équation de l'oscillateur, de variable x , est :

$$\ddot{x} + \omega_0^2 x + \epsilon(x^2 - 1)\dot{x} = F \cos(\Omega t) \Leftrightarrow \ddot{x} + \omega_0^2 x + \lambda(x)\dot{x} = F \cos(\Omega t) \quad (1.33)$$

où ω_0 est la pulsation propre, ϵ est le coefficient de friction ou de non-linéarité (il est toujours positif), F est l'amplitude du forçage, Ω est la pulsation de la fonction périodique forçant le système. Nous posons $\lambda(x) = \epsilon(x^2 - 1)$.

En absence de forçage, le comportement dépend du rapport entre la variable x et la valeur 1, qui est aussi une valeur seuil. Lorsque $x \ll 1$, alors, $\lambda(x) \rightarrow -\epsilon$. L'équation (1.33) est une équation linéaire donnant naissance à des oscillations croissantes exponentiellement. Lorsque $x = 1$, on passe par un oscillateur harmonique. Ensuite, x est supérieur à 1. Le terme non-linéaire $\lambda(x)$ devient négatif et prépondérant, le système perd de l'énergie et les oscillations décroissent.

L'amplitude des oscillations se stabilise à une valeur de compensation entre l'énergie dissipée dans le système et l'apport extérieur d'énergie. Cette valeur est différente du seuil car ce dernier permet la transition entre le cas où l'apport d'énergie est supérieur et le cas où les pertes l'emportent.

Nous visualisons la trajectoire de ce signal dans le plan (x, \dot{x}) appelé espace des phases (figure 1.14). Il est important de remarquer que la solution finale ne dépend pas des conditions initiales. La trajectoire du signal tend d'abord vers une orbite périodique pour former une trajectoire fermée et reste confinée sur cette orbite : c'est un cycle limite stable. L'indépendance de la solution finale par rapport aux conditions initiales est une caractéristique importante des oscillateurs non-linéaires dissipatifs. Cela est récurrent dans les réactions bio-chimiques du fait des systèmes de dégradations des protéines ou de l'ARNm.

Nous remarquons également que deux types de régimes sont présents :

- un régime quasi-sinusoïdal où les oscillations sont proches d'une sinusoïde (paramètre ϵ faible).
- un régime de relaxation où il apparaît deux échelles de temps distinctes (paramètre ϵ élevé), une échelle de temps très rapide au voisinage des extremums et une lente entre les extremums.

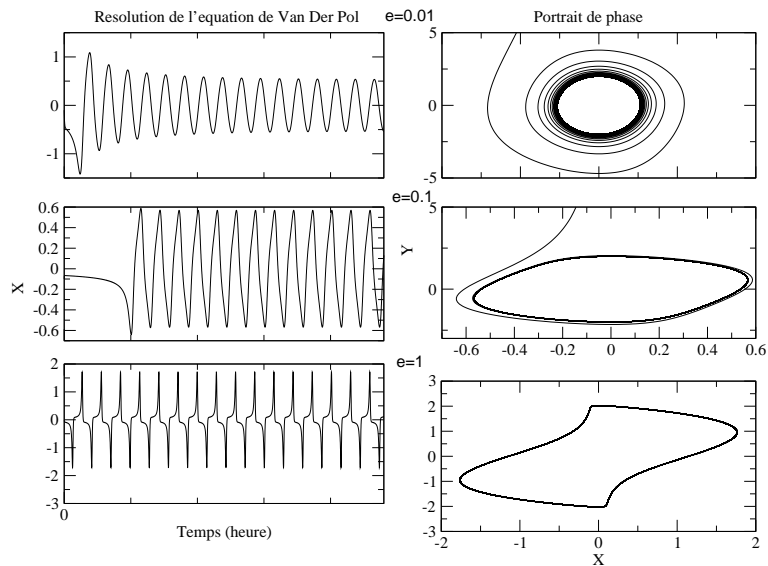


FIGURE 1.14 – Signaux de l'oscillateur de Van der Pol. Pour des conditions initiales hors de l'orbite périodique, la trajectoire converge vers le cycle limite. Plus le terme de friction ϵ est élevé, plus le signal de Van der Pol s'éloigne d'une forme sinusoïdale.

Mais ces régimes ne sont pas plus étudiés car nous ne comparons pas ces propriétés avec les oscillateurs circadiens⁴.

La présence du forçage F permet la synchronisation de plusieurs oscillateurs de Van der Pol. Nous considérons deux oscillateurs de Van der Pol ayant deux périodes naturelles proches. Si ces oscillateurs sont forcés périodiquement, alors il existe des valeurs de F tel que les oscillateurs soient entraînés à la période de forçage. Il y a un accrochage de phase, c'est-à-dire que la différence de phase entre les différents oscillateurs est constante et la période est égale à celle du forçage (figure 1.15). En général, plus la période en libre cours d'un oscillateur est proche de la période de forçage, moins l'amplitude F du forçage a besoin d'être grand pour synchroniser [97]. Nous comprenons que si la période libre de l'oscillateur d'un organisme est fortement différente de 24 heures, cela joue en défaveur de la survie de l'organisme car il est plus difficile de se synchroniser au rythme solaire.

Nous récapitulons maintenant les caractéristiques communes entre un oscillateur biologique circadien et un oscillateur auto-entretenu. L'oscillateur circadien est entraîné par le cycle d'alternance jour-nuit de 24 heures et il oscille exactement à la période de ce cycle. En absence de forçage, l'oscillation persiste avec une période proche de 24 heures. La solution oscillante est indépendante des conditions initiales.

4. Cependant, il existe une propriété commune entre les oscillateurs de relaxation et les oscillateurs circadiens [99].

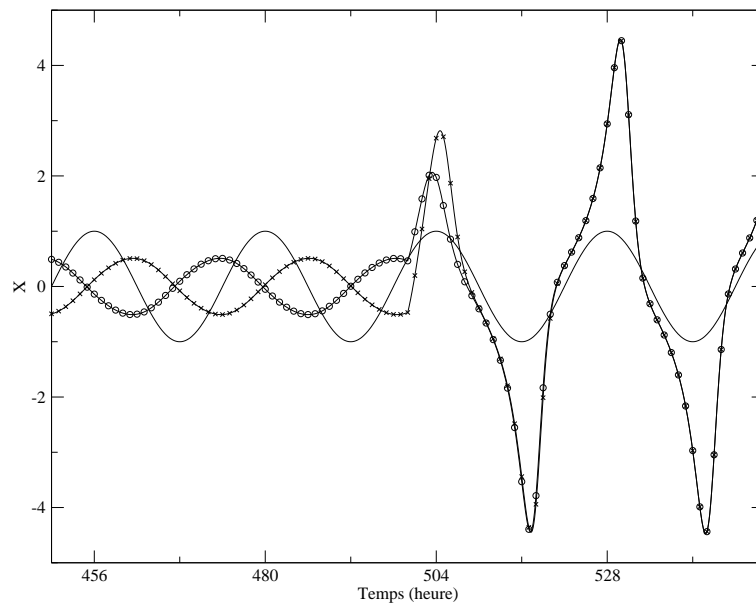


FIGURE 1.15 – Exemple de synchronisation de l'oscillateur de Van der Pol. En trait noir continu, l'oscillateur de forçage est un cosinus. La période de forçage est de 24 heures. Les deux oscillateurs de Van der Pol (représentés par des croix et des ronds) ne sont pas soumis à un forçage avant 500 heures. Leur période libre est de 25 heures et ils sont déphasés d'une demi-période. Après 500 heures, les deux oscillateurs de Van der Pol sont soumis au forçage périodique. Très rapidement, les deux oscillateurs se synchronisent et se superposent. La différence de phase est nulle et ils ont la même période de 24 heures.

1.4.5 Structures d'horloge circadienne de quelques organismes

1.4.5.1 Horloge circadienne de *Drosophila melanogaster*

La drosophile est un insecte également appelé mouche du vinaigre. C'est une espèce modèle dans la recherche en génétique car elle est très facile à élever. Son génome a été entièrement séquencé et il contient environ 15000 gènes. L'horloge circadienne est présente dans les différents tissus et dans les neurones de l'insecte. Nous récapitulons certains gènes essentiels de l'horloge au tableau 1.4.5.1.

Le tableau n'est pas complet mais contient les gènes essentiels au fonctionnement de l'horloge circadienne (voir figure 1.17). Les rôles de chaque gène dans le réseau génétique sont clairement identifiés. Certains gènes sont vitaux dans la régulation transcriptionnelle mais d'autres gènes participent activement à d'autres processus biochimiques décrits par la suite. Lors de la régulation transcriptionnelle, les deux gènes *clk* et *cyc* activent la transcription de deux autres gènes, *per* et *tim*. Ces deux derniers gènes, *per* et *tim*, répriment les deux gènes *clk* et *cyc*. L'ensemble de ces gènes participe à une boucle de rétroaction négative puisque *per* et *tim* répriment finalement leur propre expression en réprimant les gènes qui les activent. D'autres gènes, dont l'activité influe sur le fonctionnement de l'horloge, ne régulent pas la transcription. Ils ont un rôle important dans les processus biochimiques agissant sur les protéines de l'horloge comme le gène *dbt*, dont la protéine DOUBLE-TIME induit la dégradation de PER. Il peut être également important de connaître les gènes impliqués dans la réponse à la lumière de l'horloge circadienne. Le gène *cry* (chryptochrome) est responsable de la photoréception. Ce dernier gène est donc vital pour la synchronisation avec le cycle lumineux jour et nuit.

Mais le modèle mathématique ne doit pas forcément tenir en compte de tous les acteurs biologiques identifiés. Il est plus utile de se limiter aux boucles de rétroaction et les réactions bio-chimiques importantes. Leloup et al. ont écrit un modèle de l'horloge circadienne basé sur la boucle de rétroaction négative. Le complexe nucléaire *PER :TIM* réprime les gènes correspondants [38]. L'influence de la lumière est modélisée en modulant les constantes cinétiques liées à la dégradation de la protéine TIM. Les résultats numériques montrent l'existence d'un cycle limite. Les oscillations sont entretenues dans les conditions de noir continu et lors de l'entraînement au cycle jour-nuit. Les données numériques sont en bon accord avec les données expérimentales. Ils ont également montré que la dimérisation des deux protéines étaient un élément clef pour obtenir les oscillations. Ainsi, dans cette étude, la stratégie était de prendre en compte les gènes de la boucle de rétroaction et de montrer une certaine cohérence des résultats avec l'expérience.

Le modèle peut aussi servir à découvrir des gènes centraux de l'horloge qui n'ont pas été découverts expérimentalement, comme dans l'exemple suivant [15].

nom du gène	rythme	dimérisation avec	rôle de la protéine
period (<i>per</i>)	cycle protéine et ARNm	<i>TIM</i>	réprime <i>cyc</i> et <i>clk</i>
timeless (<i>tim</i>)	cycle protéine et ARNm	<i>PER</i> et le stabilise	réprime <i>cyc</i> et <i>clk</i>
double-time(<i>dbt</i>)	taux de transcription constant	<i>PER</i> et <i>PER :TIM</i>	favorise la dégradation de <i>TIM</i>
Clock(<i>clk</i>)	cycle protéine ARNm	<i>CYC</i>	active <i>per</i> , <i>tim</i> , et <i>vri</i>
cycle(<i>cyc</i>)	gène transcrit à un taux constant	<i>CLK</i>	active <i>per</i> , <i>tim</i> et <i>vri</i>
vrille(<i>vrl</i>)	gène transcrit à taux constant mais cycle protéine et ARNm	-	réprime l'accumulation de l'ARN <i>tim</i> et <i>per</i>
shaggy(<i>sgg</i>)	gène transcrit à un taux constant	-	Favorise la phosphorylation de TIM et la translocation nucléaire du dimère TOC :TIM
cryptochrome(<i>cry</i>)	cycle de l'ARNm	-	Photorécepteur circadien, favorise la dégradation photo-dépendante de TIM.

FIGURE 1.16 – Tableau récapitulant les caractéristiques des gènes identifiés de l'horloge circadienne de *Drosophila melanogaster*. La mutation de chaque gène entraîne une altération du rythme, à l'exception du gène *cry*. La mutation de ce dernier gène provoque un défaut lors du photo-entraînement. Tableau tiré de [84].

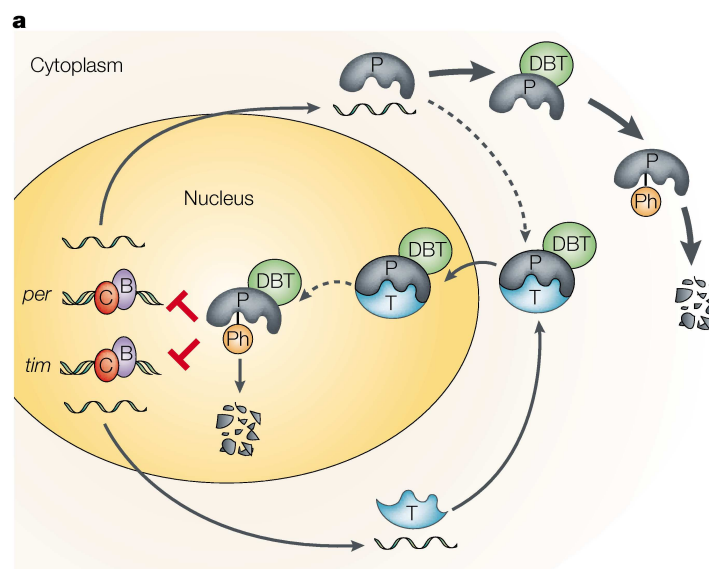


FIGURE 1.17 – Modèle biologique simplifié du réseau génétique de l'horloge circadienne de la drosophile. La cellule est séparée en deux parties, le noyau et le cytoplasme. La protéine DOUBLE-TIME (DBT) induit la dégradation de la protéine PER. La dimérisation de PER avec TIM provoque l'entrée nucléaire. La protéine PER réprime la transcription du gène *per* et *tim*. Le réseau génétique a une boucle de rétroaction négative. Schéma tiré de [84].

1.4.5.2 Modèle *Arabidopsis Thaliana*

Arabidopsis Thaliana est une plante verte originaire de l'Europe, d'Asie et de l'Afrique du nord-ouest. C'est un organisme modèle pour les plantes vertes en biologie moléculaire et en génétique. Son génome a été entièrement séquencé. Elle compte environ 27000 gènes codants pour environ 35000 protéines.

Cette plante a une horloge circadienne largement étudiée. L'étude de cette horloge est relativement complexe car de nombreux gènes participent au fonctionnement de l'horloge et sont redondants⁵. Les premiers modèles mathématiques ont été construits sur la base d'une boucle à trois gènes. Les gènes *cca1* et *lhy* sont des gènes redondants. Les protéines *CCA1* et *LHY* répriment l'expression du gène *toc1*. La protéine *TOC1* active la transcription des deux gènes *cca1* et *lhy*. Tout comme le modèle *Drosophila melanogaster*, le réseau est construit à partir d'une boucle de rétroaction négative. Cependant, les premiers modèles basés sur cette boucle n'étaient pas capables de retrouver correctement l'ensemble des données expérimentales [100]. Millar et al. utilisèrent la modélisation mathématique afin d'identifier des acteurs génétiques inconnus mais vitaux pour le réseau génétique de l'horloge circadienne [15]. La boucle de base fut modifiée en ajoutant deux hypothétiques gènes X et Y. Le gène X est un gène intermédiaire entre la régulation des gènes *lhy* et *cca* par *TOC*. Le gène Y est un gène activant le gène *toc* tandis que *TOC* réprime le gène Y. La présence du gène Y ajoute donc une boucle de rétroaction négative. Le modèle comprend par conséquent deux boucles à rétroaction négative emboîtées (voir figure 1.18). La réussite de ce modèle est double. D'abord, il a permis d'ajuster plutôt correctement les données expérimentales, reprises à la figure 1.19. Ensuite, le modèle suggéra l'existence de gènes non-identifiés qu'il resta à découvrir. Le gène *gi* fut un candidat potentiel pour être le gène Y. La figure 1.20 montre en effet une similarité entre les données expérimentales de l'ARN de ce gène avec le résultat de la simulation numérique [15].

Ainsi, la modélisation joue plusieurs rôles. Alors qu'elle valide un modèle biologique, elle peut également suggérer des expériences pour comprendre le système biologique.

5. Un gène redondant est un gène qui est la copie d'un autre gène dans le génome. Cela a pour effet que les gènes redondants ont parfois les mêmes fonctions biologiques.

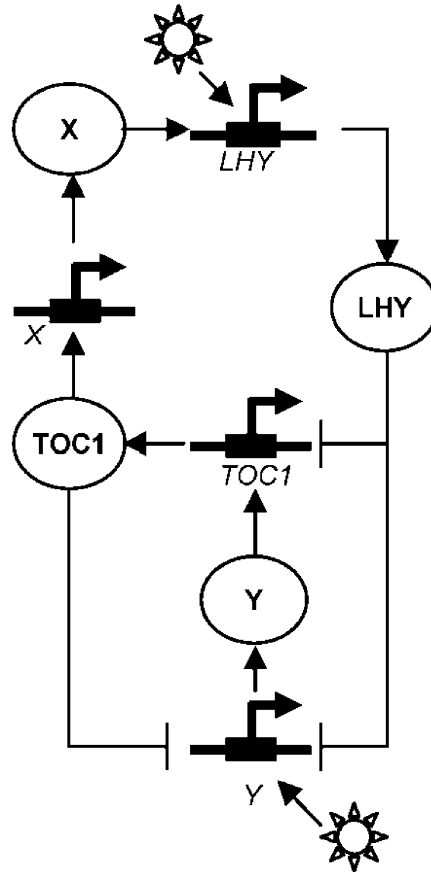


FIGURE 1.18 – Réseau de boucles de rétroaction emboîtées (Schéma tiré de [15]). Le gène *toc1* est activé par la lumière indirectement par le gène Y. Le gène Y active donc la transcription du gène *toc1*. Le gène *toc1* active le gène X et réprime le gène X. Le gène X active LHY. LHY réprime le gène *toc1* et le gène Y.

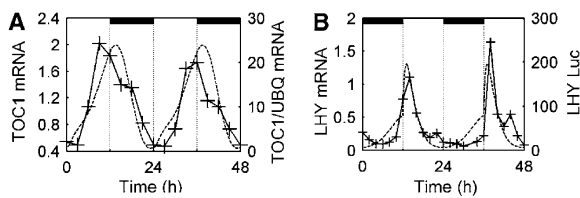


FIGURE 1.19 – Comparaison des simulations numériques du modèle des deux boucles emboîtées (traits pointillés) avec les données expérimentales (lignes avec croix) avec un entraînement par un cycle lumineux (12 h de jour et 12 h de nuit) (figures extraites de [15]) Figure A : niveau de l'ARN messenger de *toc1* sous un entraînement lumineux de 12h de jour et 12h de nuit. Figure B : niveau de l'ARN messenger de *lhy*.

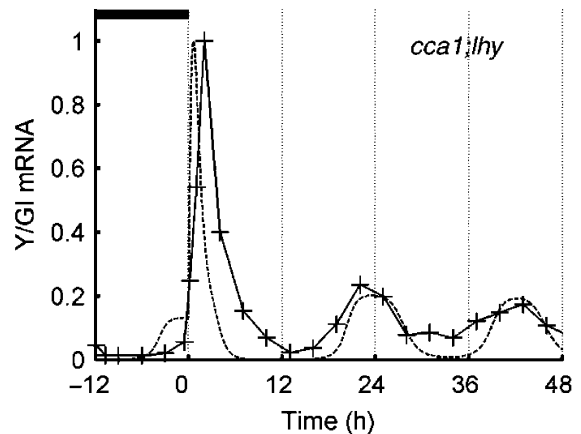


FIGURE 1.20 – Comparaison entre la simulation numérique de l'ARN du gène Y (ligne en pointillé) et du niveau d'ARN messenger du gène *gi* (ligne avec croix) (figures extraites de [15]).

1.5 Mesures quantitatives de l'expression génétique

L'élaboration des modèles dynamiques expliquant des processus biologiques doit son développement en grande partie aux outils de mesures quantitatives précis des grandeurs biologiques. En effet, la vérification expérimentale est un passage obligatoire de la validation d'un modèle. Les données expérimentales traitées dans cette thèse ont été obtenues à partir de deux techniques de mesures expliquées par la suite et nous donne accès aux quantités biologiques (ARNm, protéines ...) de l'expression génétique dans des populations de cellules.

1.5.1 Séquençage de l'ADN

Le séquençage de l'ADN consiste à déterminer dans l'ordre les bases nucléotides (adénine, guanine, cytosine et thymine) d'un fragment moléculaire de l'ADN [101]. Nous établissons ainsi une carte complète de l'ADN par l'identification des gènes, des régions *cis*-régulatrices des promoteurs ou tous autres domaines de l'ADN. Les technologies de séquençage sont vite devenues très performantes [102] et ont permis de décrypter le génome de nombreuses plantes et animaux [8, 103, 104]. Le séquençage de l'ADN est très vite devenu la base du travail d'un grand nombre de domaines de la biologie, comme la biotechnologie ou la biologie systémique.

Le séquençage de l'ADN est un outil de base pour la modélisation de réseaux génétiques car elle permet d'avoir une liste complète de gènes. Cependant, de nouvelles difficultés se révèlent puisque nous ne connaissons pas le rôle des gènes et la fonctionnalité de leur produits. Pour la modélisation circadienne, nous créons des cellules mutantes, c'est-à-dire des cellules comportant un défaut génétique et nous observons la conséquence sur le phénotype. Par exemple, la mutation du gène *TOC1* du modèle circadien *Arabidopsis Thaliana* modifie le phénotype de la plante en raccourcissant ou en allongeant la période du cycle circadien [89]. La base de ce travail a permis par la suite d'établir un modèle circadien plus complet mais long à établir [15, 100].

La bioinformatique compare les séquences de l'ADN entre elles afin d'identifier des séquences génétiques fortement similaires. Ces gènes sont alors des gènes homologues. Deux espèces différentes ayant des ancêtres communs peuvent avoir également des séquences génétiques similaires. Leurs gènes sont alors des orthologues si leurs protéines ont la même fonction. Parfois, des séquences sont similaires sans que les gènes soient homologues ou orthologues. Il peut s'agir de séquence de reconnaissance commune permettant le transport nucléaire ou autres.

Lors de l'étude d'un nouveau modèle biologique, nous comparons les orthologues d'une espèce voisine dont le modèle est connu. Cette stratégie fut appliquée pour la recherche des gènes de l'horloge circadienne d'*Ostreococcus Tauri* orthologues au gène de l'horloge circadienne d'*Arabidopsis Thaliana* (voir chapitre 3).

A partir de ces données, l'étude de l'expression des gènes s'effectue avec d'autres techniques permettant d'avoir des données de mesures de l'expression de l'ARNm, de la régulation transcriptionnelle (ou activité transcriptionnelle) et de protéines (ou activité traductionnelle).

1.5.2 Micropuces à ADN

La micropuce à ADN est un outil biotechnologique récent permettant d'analyser le niveau d'expression des gènes d'une cellule au fil du temps (avec une mesure toutes les trois heures par exemple). Le résultat de l'utilisation de cette technique est une intensité lumineuse révélatrice des quantités d'ARN exprimées par les gènes à un moment donné. Le principe est le suivant : les ARN totaux sont extraits des cellules, amplifiés et transformés en ADN complémentaires (ADNc)⁶. Les ADNc sont marqués par une sonde fluorescente. Les ADNc sont disposés sur une micropuce avec de l'ADN. L'hybridation avec les brins d'ADN provoque de l'émission lumineuse dont l'intensité est normalisée par rapport à un étalon et nous donne une mesure de la quantité relatif de l'ARNm.

Cette technologie est très reproductible et assure la robustesse des résultats expérimentaux [105]. En outre, l'environnement des cellules est modifiable au fil du temps (variation de paramètres chimiques, cycle lumineux) afin de voir comment l'expression génétique évolue. Cependant, le coût élevé des mesures rend difficile l'utilisation régulière de ce dispositif. C'est également une méthode destructive.

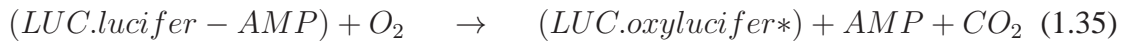
1.5.3 Gène rapporteur de la luciférase

Un gène rapporteur produit des protéines émettrices de lumière. Celles-ci sont repérées grâce à des caméras CCD (Charge Coupled Device) ou des photomultiplicateurs. Le gène rapporteur est inséré dans le génome des cellules étudiées à l'aide d'un vecteur. Le vecteur est une molécule d'ADN utilisée comme un véhicule de transfert de matériel génétique étranger dans les cellules. Le gène rapporteur peut être fusionné avec une séquence promotrice afin d'observer la régulation du promoteur (activité transcriptionnelle). Il peut être également fusionné à la séquence entière d'un gène pour produire une molécule de fusion entre la protéine luminescente et une autre protéine (activité traductionnelle) [81] (voir annexe B).

Un exemple de gène rapporteur est celui produisant la protéine fluorescente GFP (Green Fluorescence Protein) [106]. La protéine GFP absorbe la lumière dans le spectre bleu pour en émettre dans le spectre vert à 507 nm. Cette protéine est très stable avec une demi-vie proche de 24 heures. Cependant, la luminescence exogène peut causer des dommages toxiques ou perturber les cellules sensibles à la lumière, comme pour les organismes ayant des horloges circadiennes [107]. C'est une méthode de fluorescence qu'il convient de distinguer de la méthode de luminescence décrit ensuite.

Une autre classe de gène rapporteur correspond aux gènes de la luciférase qui nécessitent la présence d'un substrat pour émettre de la lumière [108]. La luciférase provenant de la luciole (voir figure 1.21) est une des enzymes les plus utilisées. Cette protéine catalyse l'oxydation d'un substrat protéique appelé la luciférine. Cette dernière émet ensuite de la lumière. Les réactions sont :

6. simple brin artificielle synthétisé à partir d'ARNm représentant la partie codante de la région du génome ayant été transcrit en cet ARNm



Le terme LUC désigne la luciférase, *lucifer* est la luciférine, l'ATP est l'Adénine Tri-Phosphate, AMP est l'Adénine Mono-Phosphate et PP est le pyrophosphate. Lors de la réaction (1.34), la catalyse par la luciférase est effectuée en présence d'ion magnésium ; l'ATP s'hydrolyse en AMP, produit de la PP, et forme un complexe de luciférase, de luciférine et d'AMP appelé luciférine-adénylate-enzyme. En présence d'oxygène O_2 dans la réaction (1.35), la luciférine-adénylate-enzyme s'oxyde puis se décarboxyle pour former de l'oxyluciférine dans un état excité. Ensuite, l'oxyluciférine se désexcite en émettant un photon pendant la réaction (1.36). La régénération de la luciférase s'effectue lors de la libération de la luciférase du complexe *LUC.oxyluciferin.AMP* et peut se révéler très lente lors de la réaction (1.37) [109]. De ce fait, la luciférase peut ne provoquer qu'une seule émission de photon avant d'être dégradée dans un organisme vivant. L'ajout ou la présence de coenzyme A dans le milieu améliore le rendement de l'émission de lumière par la luciférase en augmentant la libération de la luciférase lors de la réaction (1.37).

Le gène de la luciférase a été inséré dans le génome de très nombreux modèles vivants, comme certains mammifères [110], les insectes [90], les poissons [111], les bactéries, les plantes [112], indiquant que les techniques de fusion du gène rapporteur aux promoteurs ou à des gènes complets sont bien maîtrisées. Cette technique fut utilisée pour l'étude de l'horloge circadienne du nouveau modèle biologique qu'est l'*Ostreococcus Tauri* [93].

Cependant, de nombreuses conditions doivent être respectées pour interpréter correctement la luminescence mesurée. Il faut que les quantités des réactants soient en quantités suffisantes pour ne pas limiter les réaction (1.34) à (1.37). En général, les concentrations de l'ATP et l'oxygène ne sont pas des facteurs limitants [109]. En revanche, il faut fournir la luciférine en surabondance aux organismes ne produisant pas cette protéine [109]. Il a été vérifié que le transport de la luciférine à travers la membrane d'une cellule pouvait être très rapide devant les autres réactions [113].

Initialement utilisée pour détecter la présence d'ATP [114], la luciférase permet la détection des interactions de protéines [115], de suivre le mouvement de cellules *in vivo* [116, 117, 118], de visualiser les régulations transcriptionnelles de promoteur [119] et de suivre la quantité de protéines luminescentes [120]. Les deux derniers cas seront ceux utilisés pour l'étude de l'horloge circadienne d'*Ostreococcus Tauri*.

Pour visualiser la régulation transcriptionnelle d'un promoteur, le gène de la luciférase est fusionné à un promoteur et est inséré dans le génome d'une cellule. La mesure de la régulation se fait par la variation de l'intensité des protéines de luciférase traduites. Cette transformation est une fusion transcriptionnelle et reflète l'activité transcriptionnelle.

La deuxième transformation est la fusion d'un gène avec le gène rapporteur : cela produit une protéine fusionnée avec la luciférase. L'émission de lumière permet de repérer la protéine. Cependant, lorsque le photon est émis, la protéine devient invisible et il n'est pas possible de connaître l'évolution des concentrations uniquement à partir de ces données. L'interprétation de



FIGURE 1.21 – (Photographie de James Jordan) La luciole *photinus pyralis* est très répandue aux États-Unis et fait partie de la famille de Lampyradae. Elle possède des organes photogènes sur la face intérieure de leurs segments abdominaux leur permettant de produire elle-même de la lumière. Cette production de lumière s'effectue grâce à la synthèse d'une enzyme appelé la luciférase. La luciférase réagit avec la luciférine pour émettre un photon de lumière [121].

la luminescence obtenue dépend de constantes cinétiques liées à la luciférase (voir chapitre 4). Cette transformation est appelée une fusion traductionnelle et indique l'activité traductionnelle.

Comme la luciférase est censée être peu stable car elle émet rapidement un photon, cette protéine rapportrice est bien adaptée pour rendre compte de la dynamique rapide de la régulation des gènes et nous permet donc d'observer les rythmes biologiques liés aux réseaux génétiques tel que le rythme circadien. La luminescence de luciférase dans des cellules est représentée à la figure (1.22) [107]. Parfois, l'émission lumineuse peut être assez faible et la mesure de luminescence d'une petite zone (comme une cellule unique) nécessite un long temps d'exposition aux appareils de mesures, empêchant de voir la dynamique.

Les oscillations d'un réseau génétique sont dues à des processus chimiques clefs. Dans le chapitre suivant, nous étudions des réseaux génétiques simples dans lesquelles quelques processus favorisent l'apparition d'oscillations dans ces réseaux.

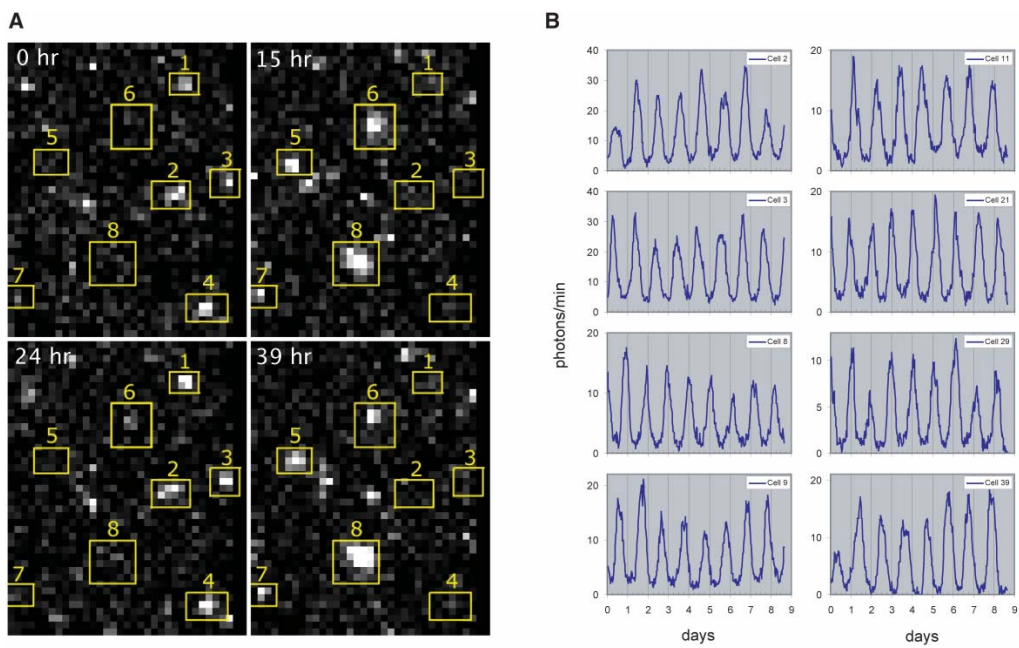


FIGURE 1.22 – Figure A : image de bioluminescence de cellules fibroblastes extraites de souris et montrant un rythme circadien par bioluminescence à travers l’activité du promoteur. Figure B : représentation du rythme circadien par bioluminescence des cellules fibroblastes sur une durée de 8 jours. Le numéro de cellule du graphe correspond au cellule lumineuse encadré aux figures gauches [122].

Chapitre 2

Réseau génétique oscillant à un seul gène

Dans ce chapitre, des modèles simples de réseaux génétiques, mais suffisamment complexes pour osciller, sont d'abord présentés. Nous récapitulons alors les propriétés mathématiques qui sont liées à des processus biologiques spécifiques responsables du déclenchement des oscillations classiques [37].

Dans ces modèles, les temps de commutations du gène sont supposés très courts par rapport à la période des oscillations. De ce fait, l'activité du gène est éliminée des équations dynamiques en l'exprimant en fonction de la concentration du facteur de transcription : le temps de réponse de l'activité du gène lors de la variation de la concentration du facteur de transcription est nul. Cependant, des travaux expérimentaux démontrent que le temps de réponse d'un gène peut être long [123, 124]. Une activation lente du gène est donc plausible et a un effet important sur la dynamique des oscillations [125] (voir annexe A). Nous décrivons alors un modèle de gène auto-régulé basé sur cette dernière hypothèse en donnant l'expression analytique du seuil d'oscillation [125] (voir annexe A).

2.1 Mécanismes permettant les oscillations dans un réseau génétique

Des oscillateurs biologiques sont générés par des interactions complexes entre les gènes et les protéines. L'apparition des oscillations nécessite des rétroactions négatives [18, 19, 37, 126, 127], des délais [128, 129, 130, 131, 132], de la non-linéarité [127, 133, 44, 43] et un équilibre entre les différentes échelles de temps des réactions chimiques de gain et de perte [37]. Nous rappelons ci-dessous une partie des résultats de Tyson and al.[37] donnant un lien entre ces différents effets.

2.1.1 Modèle de rétroaction négative à délai

Prenons le modèle du rythme circadien de la mouche à fruit dans lequel il a été montré qu'une boucle de rétroaction négative transcriptionnelle est au cœur du fonctionnement de l'horloge [134]. Le taux de synthèse des protéines à l'instant t dépend de la quantité au temps passé $t - \tau$. Ce délai est présent à cause de la durée de la transcription et de la traduction. Les

équations cinétiques sont :

$$\frac{dY(t)}{dt} = k_1 S \frac{K_d^p}{K_d^p + Y^p(t-\tau)} - k_2 E_T \frac{Y(t)}{K_m + Y(t)} \quad (2.1)$$

Mackey et Glas avaient déjà souligné le rôle déstabilisant d'un délai τ dans les systèmes physiologiques [135]. Le premier terme du membre de droite est le taux de synthèse des protéines et le second terme est le taux de dégradation. Le terme S est la concentration d'un facteur de transcription activant le gène tandis que le facteur $\frac{K_d^p}{K_d^p + Y^p(t-\tau)}$ exprime comment le gène est réprimé. Le paramètre K_d est la constante de dissociation de la protéine Y du gène. L'exposant p est la coopérativité de la protéine Y . k_1 est le taux de synthèse de Y lorsque la concentration de Y est faible. Le temps τ est le délai représentant le temps mis par la transcription du gène et la traduction de la protéine Y . Dans la deuxième moitié de l'équation, le terme E_T est la concentration totale de protéase dégradant la protéine, le terme k_2 est le taux de conversion du complexe enzyme-protéine en protéine dégradée et K_M est la constante de saturation.

2.1.2 Conditions d'oscillations

Les oscillations apparaissent pour certaines valeurs des constantes et du délai. Il y a des contraintes spécifiques sur les paramètres cinétiques suivants : le signal S , l'exposant p mesurant la non linéarité de la rétroaction, K_m mesurant la non linéarité de la dégradation, K_d et τ le délai. Pour des valeurs de p et de K_d/K_m fixées, il existe un délai minimal τ_m au-delà duquel le système oscille [132]. Les réactions chimiques doivent être suffisamment non-linéaires, c'est-à-dire que p et le rapport K_d/K_m doivent être assez grands pour faciliter l'apparition des oscillations. Tyson et al. ont tracé un diagramme de phases en fonction du délai et de la concentration S pour trouver les conditions sur les valeurs des paramètres (voir figure 2.1). Les quantités clefs sont le délai normalisé :

$$\frac{k_2 E_T \tau}{K_d} = \frac{\tau}{K_d / (k_2 E_T)} = \frac{\tau}{T_{degr}} = \frac{\text{délai}}{\text{échelle de temps de dégradation des protéines}} \quad (2.2)$$

et le signal S normalisé :

$$\frac{k_1 S}{k_2 E_T} = \frac{K_d / (k_2 E_T)}{K_d / (k_1 S)} = \frac{T_{degr}}{T_{syn}} = \frac{\text{échelle de temps de dégradation des protéines}}{\text{échelle de temps de synthèse des protéines}} \quad (2.3)$$

Le diagramme de phase établit clairement les zones de paramètres où le système oscille (voir figure 2.1). Sur ce diagramme, on trace en abscisse le signal S et en ordonnée le délai. Chaque courbe est tracée pour p et pour le rapport K_d/K_m fixés. Le système oscille pour tout point au-dessus d'une courbe donnée, à condition de respecter les valeurs fixées pour p et K_d/K_m . Une remarque est importante : en absence de rétroaction, le système n'oscille pas.

Ainsi, l'obtention des oscillations impose des conditions sur les valeurs des paramètres. En particulier sur la figure 2.1, une valeur du signal optimise l'apparition des oscillations et correspond au minimum de la courbe du délai en fonction de S . En outre, plus le rapport $\frac{K_d}{K_m}$ et l'exposant p sont grands et plus la zone des oscillations est grande.

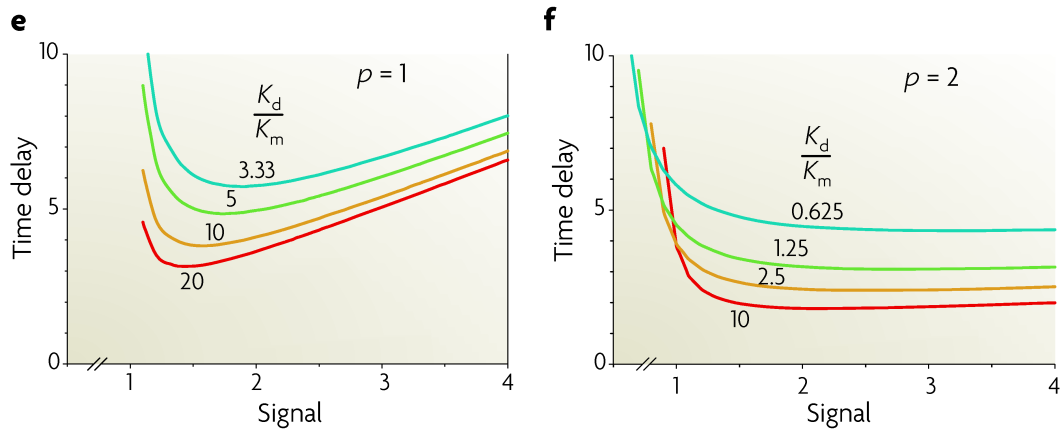


FIGURE 2.1 – Figure extrait [37] de Tyson et al. Figure e : A $p=1$, courbe délimitant les zones d'oscillation. Le système oscille pour tout couple de point (τ, S) au dessus de la courbe à un rapport $\frac{K_d}{K_m}$. f : diagramme de bifurcation pour $p=2$. On remarque que les conditions pour avoir des oscillations sont plus souples.

2.2 Gène auto-régulé à dynamique transcriptionnelle rapide

Nous présentons ici quelques exemples de modèles classiques d'un gène régulé par sa propre protéine. Le premier exemple est l'oscillateur de Goodwin [127] dans lequel la coopérativité des protéines est importante pour obtenir les oscillations. Le second modèle de gène auto-régulé nécessite un délai [130, 128, 136]. Le dernier modèle oscille également grâce à un délai apparaissant sous la forme d'un transport nucléaire et d'une succession de réactions chimiques, la multiphosphorylation [39, 1].

2.2.1 Exemple d'oscillateur à un seul gène : l'oscillateur de Goodwin

L'oscillateur de Goodwin est un oscillateur biochimique basé sur une boucle à rétroaction négative et une forte coopérativité de la protéine régulatrice. La figure 2.2 est un schéma du modèle. Le système d'équations correspondant est :

$$\frac{dM}{dt} = \frac{\lambda_0}{1 + (R/R_m)^n} - \delta_M M \quad (2.4)$$

$$\frac{dP}{dt} = \beta_0 M - \delta_P P \quad (2.5)$$

$$\frac{dR}{dt} = \nu_0 P - \delta_R R \quad (2.6)$$

$$(2.7)$$

Les variables M , P et R représentent respectivement la concentration d'ARNm, de l'enzyme et du produit de transformation d'un substrat avec l'enzyme P (le substrat est considéré en concentration constante). Le paramètre λ_0 est le taux de transcription à concentration R faible, le paramètre β_0 est le taux de traduction de l'ARN en protéine P et le paramètre ν_0 est le taux de catalyse de l'enzyme P donnant le produit final R . Les paramètres δ_M , δ_P et δ_R sont les constantes cinétiques de dégradations de l'ARNm M , de l'enzyme P et du produit final R . Le paramètre $1/R_m$ est le seuil de répression du gène par R . Le paramètre n mesure la coopérativité du répresseur R . Goodwin avait démontré que dans ce gène auto-réprimé, le degré de coopérativité doit excéder 8 ($n > 8$) pour observer des oscillations auto-entretenues. Ce modèle fut ensuite modifié par Bliss et al [137]. Ils supposèrent dans l'équation (2.6) une dégradation de type Michaelis-Menten tel que :

$$\frac{dR}{dt} = \nu_0 P - \delta_R \frac{R}{K + R} \quad (2.8)$$

Les oscillations auto-entretenues peuvent alors apparaître pour une coopérativité $n = 1$.

Pour résumer, il peut apparaître pour des systèmes relativement simples des oscillations auto-entretenues. Mais deux éléments jouent un rôle important : le degré de coopérativité de la fonction de transcription et la présence d'une dégradation de type Michaelis-Menten. La saturation favorise les oscillations, dont celles des modèles circadiens [36, 47].

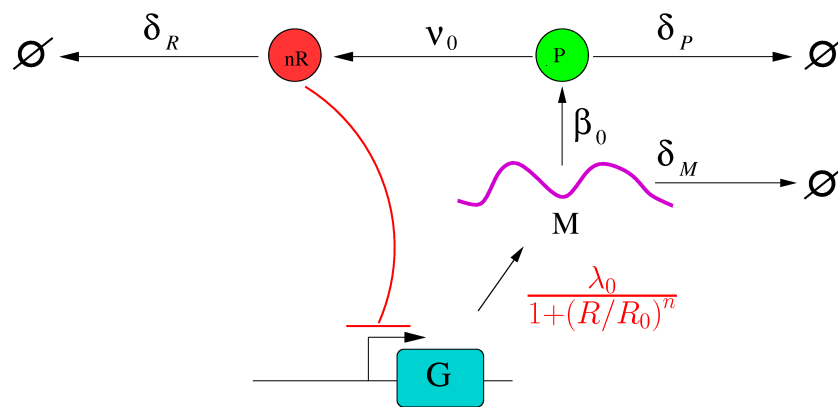


FIGURE 2.2 – Représentation schématique du modèle de Goodwin. La protéine P catalyse la formation de R. R vient réprimer par une rétroaction négative le gène G. G produit avec un taux de synthèse $\frac{\lambda_0}{1+(R/R_0)^n}$ l'ARN Messenger M.

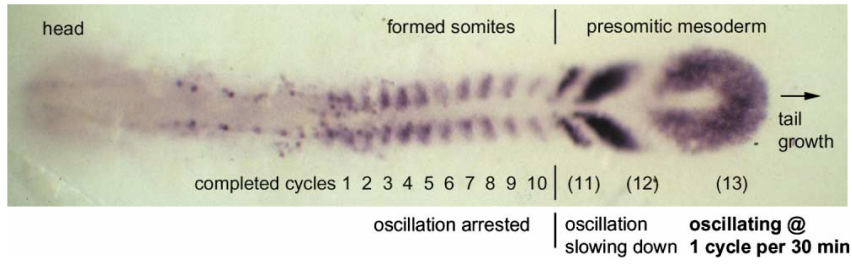


FIGURE 2.3 – Figure extraite de [136]. Exemple d’embryon du zebrafish. Les oscillations ont lieu dans la partie du presomitic mesoderm lors de la croissance de la queue et ralentissent pour s’arrêter. La durée du cycle est d’une demi-heure.

2.2.2 Exemple de gène auto-régulé contenant un délai temporel : la segmentation des somites

La segmentation des somites correspond à la formation des vertèbres de l’embryon d’un vertébré (voir figure 2.3 et 2.4). Il a été découvert dans certains organismes que ce processus était sous le contrôle d’un gène se régulant lui-même *via* à une boucle de rétroaction négative [138, 136]. La période de l’horloge somitique n’est pas assez grande comparée à la durée des processus de transcription et de traduction des protéines pour que ces derniers soient négligeables [130, 136, 128]. Ces délais peuvent donc être intégrés explicitement dans un modèle. Par exemple, celui de Monk [128] prend en compte un délai transcriptionnel explicite dans le système d’équations (2.9), afin de décrire le modèle biologique de Hiromi et al. [138] :

$$\frac{dM}{dt} = \frac{\lambda_0}{1 + (P(t - \tau_P)/P_0)^n} - \delta_M M(t) \quad (2.9a)$$

$$\frac{dP}{dt} = \beta_0 M(t) - \delta_P P(t) \quad (2.9b)$$

La protéine est représentée par P et l’ARNm par M . Le taux de traduction est β_0 , le paramètre λ_0 le taux de transcription à protéine faible et le paramètre P_0 le seuil de répression de la transcription du gène. L’exposant n est la coopérativité du facteur de transcription. Les paramètres δ_P et δ_M sont les constantes de dégradations de la protéine et de l’ARN messager. Le délai τ_P est le temps mis pour la transcription d’une protéine.

Les oscillations de M et de P issues de l’intégration numérique ont une période et une différence de phase conformes aux données expérimentales [138] (voir figure 2.4 et 2.5). Les valeurs des demi-vies de l’ARNm et de la protéine sont reprises de la littérature. Comme ces temps sont supérieurs au délai τ_P , une coopérativité élevée est nécessaire pour obtenir des oscillations soutenues ($n > 5$) d’après l’analyse mathématique. Mais des oscillations sont possibles pour $n = 2$ si le délai est plus grand que les demi-vies.

Ainsi, ces résultats montrent qu’un délai temporel peut avoir un impact significatif sur la dynamique du système lorsqu’il n’est pas négligeable devant les autres temps caractéristiques (comme la demi-vie des biomolécules). Ces études encouragent ainsi la recherche systématique de certains paramètres pertinents biologiquement pour les modèles ayant de faibles périodes et des rétroactions transcriptionnelles.

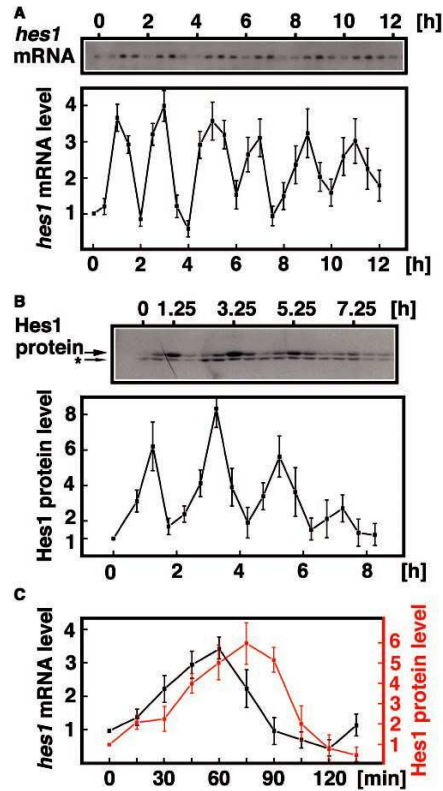


FIGURE 2.4 – Figures extraites de [138]. Oscillations des concentrations d’ARNm et de protéine du gène *Hes1* dans des cellules de souris. La durée du cycle est de deux heures.

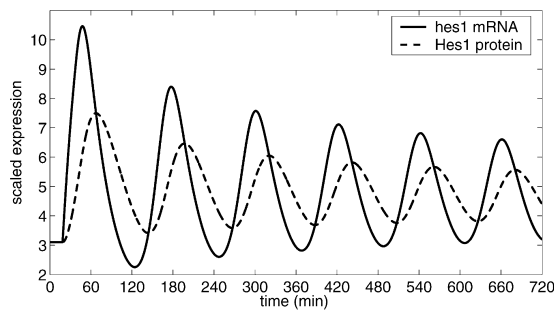


FIGURE 2.5 – Figure extraite de [128]. Intégration numérique du système d’équations 2.9. Les variables de l’ARN M et de la protéine P sont rééchellonnées par λ_0 et $\lambda_0\beta_0/\delta_P$ respectivement. Les paramètres normalisés sont : $\delta_M = 0.031/min$, $\delta_P = 0.03/min$, $\tau_P = 18.5min$, $P_0 = 100$ et $n = 5$. La durée du cycle est de deux heures.

2.2.3 Modèle de gène auto-régulé de la drosophile : délai par multiphosphorylation

2.2.3.1 Phosphorylation

La phosphorylation est l'addition d'un groupement de phosphate à une protéine [2]. Cette réaction est catalysée par d'autres protéines : les protéines kinases. Lorsque cette réaction chimique se produit plusieurs fois, elle crée des protéines multiphosphorylées. Dans la section suivante, nous exposons un modèle où la phosphorylation est déterminante pour le rythme circadien.

2.2.3.2 Modèle de Goldbeter

Il s'agit ici d'un modèle élaboré par Goldbeter [39, 38, 1, 91] pour expliquer à partir de données expérimentales [139] le rythme circadien de la drosophile contrôlé par l'oscillation de la protéine PER. Dans ce modèle de gène auto-régulé, la protéine PER réprime la transcription de son propre gène. Le délai n'est pas explicite mais est réactionnel, c'est-à-dire qu'il est dû à une succession de réactions chimiques. En l'occurrence, la protéine PER est phosphorylée de manière multiple [139] (voir figure 2.6). Les équations sont :

$$\dot{M} = v_s \frac{K_I^n}{K_I^n + P_N^n} - v_m \frac{M}{K_m + M} \quad (2.10a)$$

$$\dot{P}_0 = k_s M - V_1 \frac{P_0}{K_1 + P_0} + V_2 \frac{P_1}{K_2 + P_1} \quad (2.10b)$$

$$\dot{P}_1 = V_1 \frac{P_0}{K_1 + P_0} - V_2 \frac{P_1}{K_2 + P_1} - V_3 \frac{P_1}{K_3 + P_1} + V_4 \frac{P_2}{K_4 + P_2} \quad (2.10c)$$

$$\dot{P}_2 = V_3 \frac{P_1}{K_3 + P_1} - V_4 \frac{P_2}{K_4 + P_2} - k_1 P_2 + k_2 P_N - v_d \frac{P_2}{K_d + P_2} \quad (2.10d)$$

$$\dot{P}_N = k_1 P_2 - k_2 P_N \quad (2.10e)$$

L'ARNm de la protéine est symbolisé par M . Le gène est régulé négativement pour transcrire l'ARN selon la fonction $v_s \frac{K_I^n}{K_I^n + P_N^n}$, où n est la coopérativité du facteur de transcription, v_s le taux de transcription à protéine nulle. Le seuil de répression du gène est K_I^n . Les protéines P_0 , P_1 et P_2 sont respectivement la protéine PER dans l'état non-phosphorylé, phosphorylé une fois et complètement phosphorylé. Dans le dernier état, la protéine P_2 est dégradée [139] selon la fonction de Michaelis-Menten et traverse la membrane nucléaire [139] pour devenir une protéine nucléaire P_N . La constante cinétique de traversée nucléaire (du cytoplasme vers le noyau) est k_2 , le chemin inverse s'effectue par la protéine nucléaire avec la constante cinétique k_1 . Elle ne se dégrade pas dans le noyau. La traversée nucléaire fait également office d'un délai réactionnel [132]. La transition entre les états de phosphorylations s'effectue avec le taux $\text{signe} \times V_i \frac{P_j}{K_i + P_j}$ (le signe est + ou -, $i = 1, 2, 3$, $j = 0, 1, 2$). La réaction de transition enzymatique est une fonction de Michaelis-Menten. Les étapes de phosphorylations permettent au réseau de garder en mémoire les états précédents et sont équivalentes à un délai.

La présence de délais réactionnels (la traversée nucléaire [132] et la multiphosphorylation) permet d'obtenir des oscillations soutenues (voir figure 2.7) dans un domaine de l'espace des

paramètres grandissant avec la coopérativité du facteur de transcription. Il est à noter que la multiphosphorylation élève la période jusqu'à une valeur circadienne [39].

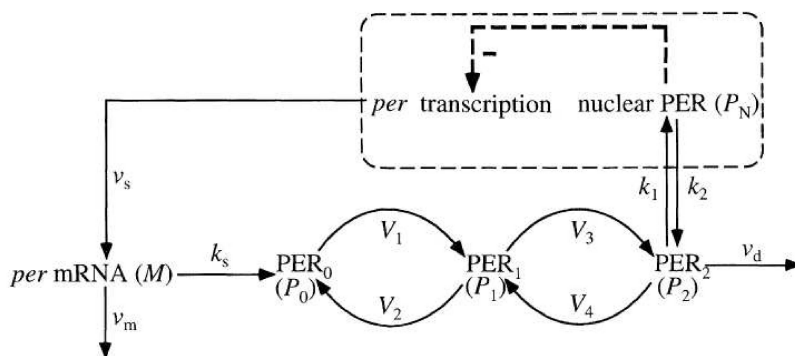


FIGURE 2.6 – Figure extraite de [39]. Modèle de l'horloge circadienne de la drosophile. Le gène *per* est régulé négativement par sa propre protéine. La protéine non phosphorylée PER_0 est phosphorylée deux fois de suite pour passer par deux autres états PER_1 et PER_2 . On remarque que la phosphorylation est une réaction réversible. Dans l'état de phosphorylation PER_2 , la protéine traverse la membrane nucléaire [140] et réprime le gène *per* [141]. Le gène *per* transcrit l'ARNm *per*. L'ARN se traduit ensuite en protéine PER_0 . L'ARN *per* se dégrade. La protéine PER ne se dégrade que dans l'état phosphorylé PER_2 .

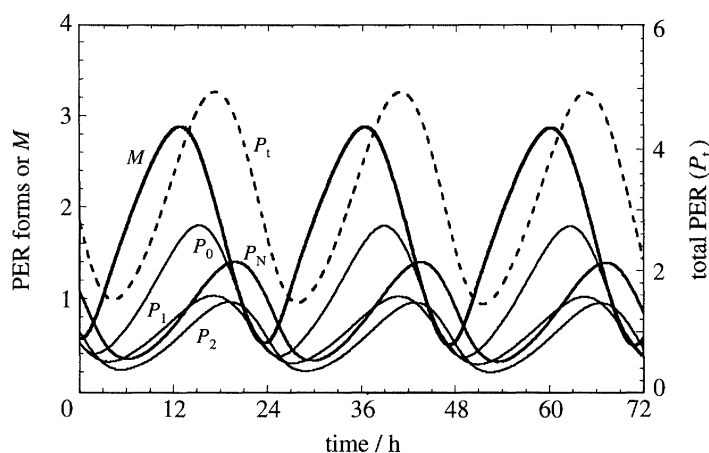


FIGURE 2.7 – Figure extraite de [39]. Oscillations entretenues obtenues à partir de l'intégration du modèle 2.10. La variable M est la concentration de l'ARNm, la variable P_t la somme de la concentration de chaque protéine PER sous différentes formes, P_N la protéine nucléaire, P_0 la protéine cytoplasmique non phosphorylée, P_1 phosphorylée une fois et P_2 phosphorylée deux fois. La période en libre cours est proche de 24 heures. Les oscillations sont soutenues pour $n=1$ ou 2 (n est le coefficient de Hill) mais le domaine de l'espace des paramètres correspondant aux oscillations est plus grand pour $n=4$.

2.2.4 Dynamique d'activation lente du gène

Dans la partie précédente, les modèles tenaient compte d'une boucle de rétroaction négative sur le gène régulé. Ces derniers modèles supposaient toujours que l'activité du gène s'ajustait immédiatement à la quantité de protéine : la variable du gène est alors éliminée des équations. Cependant, l'étape de transcription est faite d'un ensemble de réactions complexes [142] et leur mise en place nécessite du temps pour que le gène soit pleinement actif. Aussi, la dynamique du gène n'est pas négligeable si ce temps est comparable devant l'échelle de temps de la dynamique globale du système, comme devant une faible période des oscillations.

L'amélioration des techniques d'observation en temps réel de la transcription [143] et du nombre de transcrit [144] ont permis d'aboutir à des méthodes efficaces pour compter les transcrits (ou ARN) produits juste après leur transcription. La production de transcrits s'effectue par salves ou 'bursts' [123]. La salve de transcription est associée à un temps durant lequel de nombreux ARN sont produits et auquel succède une phase avec peu d'ARN fabriqués en moyenne. Chacune de ces phases est contrôlée par la régulation du gène. Les travaux de Golding et al. [123] mettent en évidence dans la bactérie *Escheria coli* la durée d'activité du gène. En moyenne, le temps d'activité est $\Delta_{on} = 6 \text{ min}$ et le temps d'inactivité $\Delta_{off} = 37 \text{ min}$ (voir figure 2.8). Ces valeurs sont comparables aux temps caractéristiques d'autres processus, comme la dégradation (durée de vie de l'ARN varie entre 1 et 100 minutes).

Des travaux ont déjà mis en évidence l'influence de l'activité du gène sur la dynamique globale d'un système bistable [145, 146]. Cependant, aucun travail n'indique l'effet d'une telle dynamique du gène dans les oscillateurs génétiques. Nous avons établi un modèle déterministe d'un gène régulant sa propre expression dans lequel nous cherchons les oscillations. Le temps de réponse du gène est variable afin de comprendre son importance pour obtenir les oscillations.

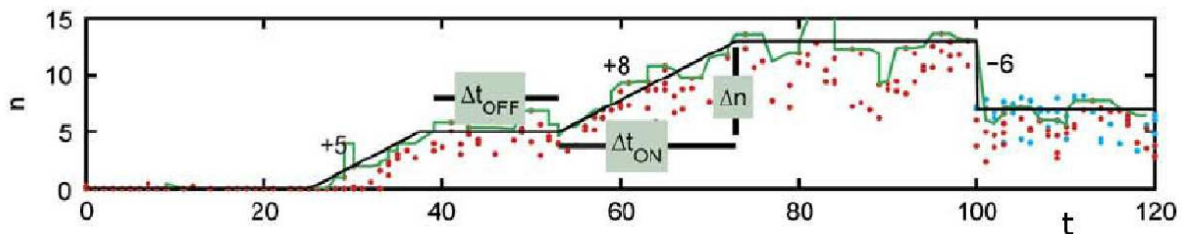


FIGURE 2.8 – Figure extrait de [123]. Dynamique transcriptionnelle d'un seul gène. Le nombre de transcrit n (point rouge) en fonction du temps t . Il y a une phase d'activité transcriptionnelle (Δ_{on}) d'une durée moyenne de 6 minutes et une phase d'inactivité transcriptionnelle (Δ_{off}) d'une durée moyenne de 37 minutes.

2.3 Dynamique d'un gène auto-régulé à temps de réponse variable du gène

Le réseau génétique est formé d'un gène réprimé par sa propre protéine (figure 2.9). Nous déterminons les conditions d'obtention des oscillations en considérant en particuliers l'influence de l'activité du gène sur la dynamique du réseau. Nous mettons en évidence la relation entre deux grandeurs, le temps de réponse du gène et un paramètre caractérisant les taux effectifs de dégradations. Cela aboutit à la découverte d'une résonance dans la dynamique : il existe une valeur du temps de réponse du gène permettant l'apparition d'oscillations sans une forte saturation de la dégradation. Ce paramètre de résonance se montre également pertinent lors de simulations stochastiques car les fluctuations des signaux de la quantité de protéines et d'ARNm sont diminués autour de ce temps.

2.3.1 Présentation du modèle

Nous supposons que l'activité du gène est une variable dynamique associée à la réaction cinétique décrite dans la section 1.2.2 :



où G est le gène non lié, nP un polymère formé de n protéines P se complexant et $G : P$ le gène lié au polymère. Le gène existe sous deux formes, lorsque le promoteur est lié selon la probabilité $G : P$ et lorsque le promoteur du gène est libre selon la probabilité G . Le complexe moléculaire nP se lie au promoteur G selon la constante cinétique α_0 et se libère selon la constante cinétique θ_0 . L'occupation maximale du gène, fixée à 1, correspond au nombre de copies du gène et conduit à la relation de conservation suivante :

$$1 = G + G : P \quad (2.12)$$

Cette relation permet d'éliminer la variable du gène lié, $G : P$, de l'équation dynamique de l'activité du gène. L'équation modélisant l'activité du gène par le polymère est :

$$\dot{G} = \theta_0(1 - G) - \alpha_0 C(P)G \quad (2.13)$$

Nous introduisons la fonction de protéine régulatrice $C(P)$ pour représenter l'interaction entre le gène et le polymère. Si la régulation se fait par un polymère de n protéines, alors $C(P) = P^n$, où n est le coefficient de Hill, un entier positif supérieur à 0. L'équation dynamique donnant la quantité d'ARNm est :

$$\dot{M} = \mu_0 + \lambda_0 G(t) - \delta_M H(M) \quad (2.14)$$

La fonction $\mu_0 + \lambda_0 G$ est le taux de transcription $\delta_M H(M)$. La fonction $H(M)$ est arbitraire, nulle en $M = 0$, monotone et la pente à l'origine est égale à 1. Pour de faibles valeurs de M , le taux de dégradation dépend de la valeur du paramètre δ_M . La dynamique de la protéine P est donnée par l'équation différentielle suivante :

$$\dot{P} = n\dot{G} + \beta_0 M - \delta_P F(P) \quad (2.15)$$

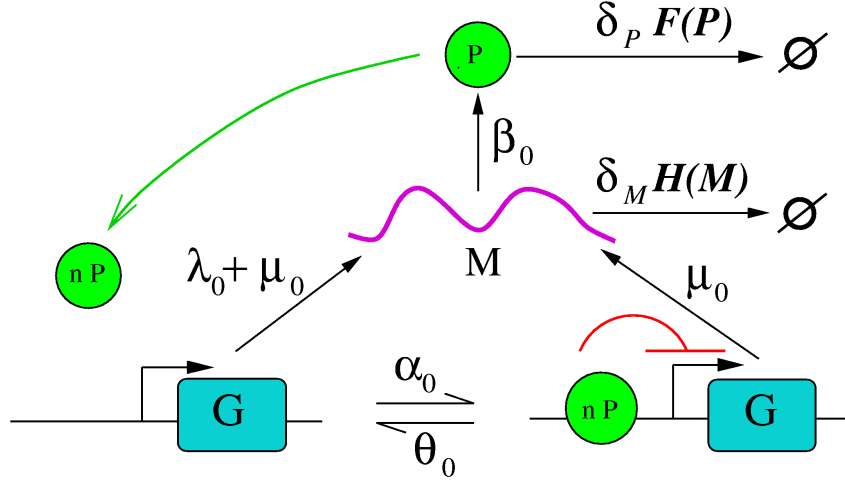


FIGURE 2.9 – Schéma du gène régulé par sa propre protéine. Le gène G est sous deux états, libre ou lié à la protéine nP (n entier si la protéine se polymérise). Le gène transcrit un ARNm noté M . L'ARN se traduit en protéine P .

L'équation (2.15) décrit la traduction des transcrits M en protéines. Le taux de dégradation est la fonction $\delta_P F(P)$. Comme pour l'ARN messager, la fonction $F(P)$ est une fonction arbitraire, nulle en $P = 0$, monotone et la pente de $F(P)$ est égale à 1 en $P = 0$. Ainsi, le taux de dégradation est donné par la valeur de δ_P lorsque le nombre de copies de protéines est faible.

2.3.2 Critère analytique d'apparition des oscillations

Un critère permet d'obtenir une relation à partir des paramètres déterminant le comportement dynamique du système. C'est une inégalité précisant si le comportement dynamique est oscillant. Auparavant, nous établirons les nouveaux paramètres du système pour éliminer les dimensions des équations. Nous calculerons les points fixes du système. Ces grandeurs sont nécessaires pour établir le critère.

2.3.2.1 Équations sans dimension et détermination des points fixes

La normalisation permet de réduire le nombre de paramètres apparents afin de simplifier l'étude analytique. De plus, la renormalisation du temps redéfinit librement l'échelle de temps du système en choisissant arbitrairement la période des oscillations du système.

Les nouveaux paramètres sont :

$$t = \frac{t'}{\delta_M}, G = g, P = pP_0 \quad (2.16a)$$

$$P_0 = C^{-1}\left(\frac{\theta_0}{\alpha_0}\right), M_0 = \frac{\delta_P P_0}{\beta_0}, \theta = \frac{\theta_0}{\delta_M} \quad (2.16b)$$

$$\alpha = \frac{\theta_0}{P_0 \delta_M}, \delta = \frac{\delta_P}{\delta_M}, \lambda = \frac{\lambda_0}{M_0 \delta_M}, \mu = \frac{\mu_0}{M_0 \delta_M} \quad (2.16c)$$

$$c(p) = \frac{C(P)}{C(P_0)}, f(P) = \frac{F(P)}{P_0}, h(M) = \frac{H(M)}{M_0} \quad (2.16d)$$

Le système d'équations se réécrit alors :

$$g' = \theta[1 - g(1 + c(p))] \quad (2.17a)$$

$$p' = n\alpha[1 - g(1 + c(p))] + \delta[m - f(p)] \quad (2.17b)$$

$$m' = \mu + \lambda g - h(m) \quad (2.17c)$$

en posant $x' = \frac{dx}{dt}$, où $x = g, p, m$. Les points fixes sont :

$$p^* = c^{-1} \left(\frac{1 - g^*}{g^*} \right), m^* = f(p^*), g^* = \frac{h(m^*) - \mu}{\lambda} \quad (2.18)$$

Pour que l'état d'équilibre soit unique, il faut que les fonctions f et h soient monotones et que $h[f(\infty)] > \mu$. Les pentes des fonctions de dégradation et de coopérativité au voisinage de l'état stationnaire s'écrivent :

$$s = \left. \frac{df(p)}{dp} \right|_{p=p^*}, u = \left. \frac{dh(m)}{dm} \right|_{m=m^*}, v = \left. \frac{dc(p)}{dp} \right|_{p=p^*} \quad (2.19)$$

Lorsque la valeur de la pente est faible ou négative, cela signifie que la fonction est fortement non-linéaire : pour une dégradation enzymatique, la pente nulle aux points fixes signifie que la dégradation est saturée [19, 43]. Ces paramètres permettent d'écrire le critère de stabilité du système.

2.3.2.2 Critère d'oscillation

La stabilité d'un système est déterminée par les valeurs propres λ de la matrice Jacobienne évaluée aux points fixes des équations dynamiques. Si le système est stable pour une valeur de paramètre μ , celui-ci peut devenir instable si μ passe par une valeur critique μ_c , le point de bifurcation. Selon le théorème de bifurcation de Hopf, si la linéarisation du système autour d'un point fixe donne deux valeurs propres complexes conjuguées, et que, dans le voisinage du point de bifurcation μ_c , la partie réelle des deux nombres complexes passe de négative à positive en passant par zéro, alors l'état stationnaire se déstabilise au point de bifurcation pour donner naissance à une solution de cycle limite [23, 147, 22, 18]. Nous utilisons ce théorème pour déterminer un critère analytique d'apparition des oscillations.

La forme générale du polynôme caractéristique est un polynôme du troisième degré :

$$p(\lambda) = \lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 \quad (2.20)$$

où les paramètres a_i sont une combinaison non triviale des constantes cinétiques de notre modèle (2.17). En supposant qu'on soit au point de bifurcation de Hopf, le polynôme s'écrit donc $p(\lambda) = (\lambda + i\omega)(\lambda - i\omega)(\lambda - \lambda_0)$ et les deux valeurs propres sont $\lambda = \pm i\omega$, où ω est la pulsation au point de bifurcation. En développant cette expression et en la comparant avec (2.20), on établit la condition suivante :

$$a_1a_2 - a_3 = 0 \quad (2.21)$$

Si les paramètres du système respecte ce critère, alors il y a une bifurcation de Hopf. On remarque que ce critère coïncide avec un des critères de Routh-Hurwitz [148, 18]. En fait, ce

dernier est décliné en plusieurs inégalités, certifiant, selon notre convention, la stabilité du système si les conditions suivantes sont respectées :

$$a_1 > 0 \quad (2.22a)$$

$$a_2 > 0 \quad (2.22b)$$

$$a_1 a_2 - a_3 > 0 \quad (2.22c)$$

Pour le système (2.17), a_1 et a_2 sont toujours positifs. Ainsi, nous pouvons nous focaliser sur les conditions pour que $a_1 a_2 - a_3$ soit négative. Nous notons cette quantité H qui s'exprime en fonction des paramètres du système par :

$$\begin{aligned} H = & ug^{*2} (\delta s + n\alpha v g^*) (\delta s + u + n\alpha v g^*) + \\ & g^* (-\lambda v \delta g^{*2} + n\alpha v g^* (\delta s + 2u) + (\delta s + u)^2) \theta + \\ & (\delta s + u) \theta^2 \end{aligned} \quad (2.23)$$

Les valeurs de m^* et p^* aux points fixes n'apparaissent pas car nous exprimons ces points fixes en fonction de g^* grâce aux égalités (2.18). La condition suffisante pour faire apparaître le cycle limite est $H < 0$.

Nous supposons une répression parfaite du gène $\mu = 0$ et un seuil à mi-répression P_0 très grand en faisant tendre le paramètre α vers 0. En factorisant par θ , la grandeur $\tau = \frac{g^*}{\theta}$ apparaît. Elle est appelée temps de réponse du gène tandis que le rapport $\frac{1}{\theta}$ est le temps d'occupation moyen du gène par le facteur de transcription. À partir des conditions sur α et μ posées juste avant, la somme et le produit des pentes de dégradations des protéines δs et de l'ARN u apparaissent également naturellement dans le critère (2.23). Ils sont notés respectivement σ et γ :

$$\sigma = \delta s + u \quad (2.24)$$

$$\gamma = \delta s u \quad (2.25)$$

À partir de ces hypothèses, nous posons une forme du critère plus simple à étudier :

$$H = \sigma + g_* (-\delta \lambda v g_*^2 + \sigma^2) \tau + \gamma g_*^2 \sigma \tau^2 < 0 \quad (2.26)$$

Le terme $-\delta \lambda v g_*^2$ est toujours négatif et favorise les oscillations s'il est grand. v étant lié à la coopérativité du système, il est plus grand lorsque la coopérativité augmente et facilite ainsi l'apparition d'oscillations. Mais cet aspect ne nous intéresse pas car de nombreuses études portent sur cela [18, 133, 126, 149]. Bien au contraire, il est intéressant de voir les conditions d'oscillations lorsque v est minimum, soit $v = 1$, pour nous placer dans les conditions les plus difficiles d'obtention des oscillations.

2.3.2.3 Condition nécessaire d'obtention des oscillations

À partir du critère (2.26), la condition nécessaire (mais non suffisante) pour observer des oscillations est :

$$\sigma^2 < \delta \lambda g^{*2} \quad (2.27)$$

Nous définissons le seuil critique σ_c :

$$\sigma_c = g^* \sqrt{\delta \lambda} \quad (2.28)$$

Ce paramètre apparaît ensuite dans la quantité H renormalisé.

2.3.2.4 Relation entre deux paramètres normalisés

Nous établissons une grandeur indiquant si les pentes de dégradations sont égales ou bien différentes. Par exemple, on peut avoir une dégradation saturée tandis que l'autre ne l'est pas. La mesure de la ressemblance entre les pentes est quantifiée par la paramètre ϵ . Cette relation est établie grâce à la somme σ et le produit γ des pentes de dégradation, présent naturellement dans le critère (2.26). La relation est donc du type $\epsilon = f(\gamma, \sigma)$, avec le paramètre ϵ une constante fixée arbitrairement dans un intervalle de valeur à déterminer. En outre, la détermination de cette fonction simplifie l'étude analytique du critère (2.26). Nous décrivons le mode de détermination de cette relation $\epsilon = f(\gamma, \sigma)$.

σ est fixé à une valeur donnée, avec u et s positifs. La valeur 0 est la valeur minimale que γ peut prendre et nous cherchons sa valeur maximale. En posant $\delta s = \sigma - u$, nous avons $\gamma = (\sigma - u)u$. Nous voulons la valeur de u pour que la dérivée de γ par rapport à u s'annule :

$$\frac{d\gamma}{du} = -2u + \sigma = 0 \Leftrightarrow u = \frac{\sigma}{2} \quad (2.29)$$

Le produit γ est maximal si $\delta s = \frac{\sigma}{2}$. Cette valeur γ est :

$$\gamma_{max} = \frac{\sigma^2}{4} \quad (2.30)$$

Le produit γ varie entre 0 et $\frac{\sigma^2}{4}$ avec σ fixé. Le paramètre ϵ est défini par :

$$\frac{\gamma}{\frac{\sigma^2}{4}} = \epsilon^2 \Leftrightarrow \epsilon = \frac{2\sqrt{\gamma}}{\sigma} \quad (2.31)$$

La grandeur ϵ appartient à l'intervalle $[0; 1]$ et indique si les taux de dégradations u et δs sont égaux ou bien totalement différents : le paramètre ϵ est un indicateur d'équilibre entre la dégradation de la protéine et de l'ARNm. Ce paramètre sera injecté dans le critère (2.26) pour aboutir à une forme simplifiée. Nous pourrions étudier le critère en fixant arbitrairement le paramètre ϵ afin de réduire l'exploration de l'espace des paramètres.

2.3.2.5 Forme simplifiée du critère

La définition du seuil critique σ_c et de l'indicateur d'équilibre ϵ nous permettent de réécrire (2.23) sous une forme plus simple. En redimensionnant :

$$\sigma = \sigma_c \Sigma, \gamma = \sigma_c^2 \left(\frac{\epsilon \Sigma}{2} \right)^2, \tau = \frac{T}{\sigma_c} \quad (2.32)$$

Le critère d'oscillation devient :

$$H_\epsilon(\Sigma, T) = \Sigma \left[\frac{\epsilon^2 \Sigma^2}{4} T^2 + \left(\Sigma - \frac{1}{\Sigma} \right) T + 1 \right] < 0 \quad (2.33)$$

Le paramètre Σ est la somme des pentes des fonctions de dégradations normalisées par rapport à l'inverse du seuil critique, et le paramètre T est la mise à l'échelle du temps de réponse du gène par rapport au seuil critique. Ce critère a l'avantage de permettre l'étude dans un plan à deux dimensions et d'obtenir plus de résultats analytiques pertinents lorsque nous considérons ϵ fixé.

2.4 Analyse du critère de stabilité

La présence d'oscillations est établie grâce au critère (2.33). Nous traçons un diagramme de phase du type $\Sigma = f(T)$. Nous cherchons l'ensemble des valeurs de Σ en fonction de T pour lesquelles le critère (2.32) s'annule. La forme d'une courbe du diagramme de phase permet de remarquer l'influence importante du temps de réponse du gène T car il existe une valeur T_{opt} élargissant le domaine des oscillations. Enfin, on effectue une recherche numérique de points oscillants avec différents types de fonctions de dégradations de l'ARNm et de la protéine. Ces simulations sont réalisées en ne posant ni α et ni μ égaux à zéro et servent à démontrer que le critère (2.33) reste vrai de manière plus générale, malgré qu'il soit établi à partir de ces simplifications (2.3.2.2).

2.4.1 Dynamique pour certains types de fonction de dégradation

Nous supposons que les dégradations de la protéine et de l'ARN messenger sont linéaires. Alors, $f(p) = p$ et $h(m) = m$ et donc $u = s = 1$. En plaçant ces valeurs dans le critère (2.23), nous avons :

$$H = (\delta + 1) + g_* [\delta (1 + g) + 1 + \delta^2] \tau + \delta g_*^2 (1 + \delta) \tau^2 > 0, \quad (2.34)$$

Le critère est toujours positif. Les oscillations n'apparaissent pas.

On étudie le cas où la saturation est saturées aux points fixes. La saturation est illustrée par une dégradation de type Michaelis-Menten (2.35) en guise d'exemple. La dérivée par rapport à x évaluée aux points fixes est (2.36) :

$$f(x) = \frac{Vx}{K + x} \quad (2.35)$$

$$\left. \frac{df(x)}{dx} \right|_{x=x^*} = \frac{VK}{(K + x^*)^2} \quad (2.36)$$

La dégradation est saturée lorsque le point fixe x^* est très grand devant la constante de saturation K . À l'état d'équilibre, le nombre de protéines est tellement grand par rapport aux enzymes de dégradation que les pertes sont égales à une constante. Ainsi, les dégradations saturées signifient que les pentes u et s sont nulles. La somme σ est nulle et le critère se réduit à :

$$H = -\delta\lambda\tau g^{*2} < 0 \quad (2.37)$$

Le système oscille toujours. Cependant, pour un gène auto-régulé, il est difficile d'atteindre une telle saturation. Pour avoir la pente s nulle, le point fixe p^* doit tendre vers l'infini ce qui oblige g^* à tendre vers zéro. Pour avoir la pente u nulle, le point m^* doit tendre vers l'infini, or cela n'est pas possible car m^* est égale à une fonction de saturation. Le réseau doit donc trouver un autre moyen pour bifurquer vers les oscillations. On verra d'ailleurs que cette bifurcation est rendue possible grâce à un temps de réponse du gène non nul.

Si la réponse du gène est immédiate, c'est-à-dire que le délai est nul, le paramètre θ est infini et τ tend vers 0. On a :

$$H = \sigma \quad (2.38)$$

Si les fonctions de dégradations sont monotones et croissantes pour toutes valeurs de points fixes, l'expression H est toujours positive.

Cependant, Tyson et *al.* ont remarqué qu'il existe des fonctions de dégradations non monotone pouvant avoir une pente négative sur un intervalle. Si le point fixe du système tombe sur cette intervalle, alors la somme des pentes des dégradation peut être négative aboutissant aux oscillations [43, 19].

2.4.2 Influence de la dynamique du gène sur l'apparition d'oscillations

Nous définissons une mesure de la non-linéarité grâce au degré de saturation (reliée ainsi au produit $\epsilon\Sigma$). Il existe une valeur de ϵ pour laquelle le système est le moins non linéaire. Nous détaillons alors les conditions d'obtention des oscillations pour des valeurs du temps de réponse du gène T en fonction de la somme des pentes de dégradations Σ . Le diagramme de phases obtenu à partir du critère sur $H_\epsilon(\Sigma, T)$ (2.33) indique les valeurs de T menant aux oscillations tout en étant le moins non linéaire (éloigné de la saturation). Il apparaît une valeur optimale du temps de réponse τ_{opt} correspondant à un phénomène de résonance facilitant les oscillations. Cette étude déterministe reste pertinente pour l'analyse stochastique du système (1.2.7).

2.4.2.1 Temps de réponse du gène et non-linéarité des équations

L'indice mesurant la non-linéarité des équations est la moyenne géométrique des pentes ν :

$$\nu = \sqrt{us} = \sqrt{\frac{\gamma}{\delta}} \quad (2.39)$$

Cette grandeur est reliée à l'indicateur d'équilibre ϵ par :

$$\nu = \frac{\epsilon\sigma}{2\sqrt{\delta}} \quad (2.40)$$

Une dégradation saturée, $u = 0$ ou $s = 0$ (soit $\epsilon = 0$) suffit à avoir $\nu = 0$. Si les deux dégradations sont linéaires, $\nu = 1$, alors les oscillations sont impossibles. Le cas extrême permettant d'obtenir des oscillations tout en minimisant la saturation des pentes de dégradation est en posant $\epsilon = 1$. En introduisant l'adimension (2.32), $\sigma = \sigma_c\Sigma$, et en remplaçant σ_c par la condition nécessaire d'oscillation (2.28), l'équation (2.40) devient :

$$\nu = \frac{g^*\sqrt{\lambda}}{2}\Sigma_1 \quad (2.41)$$

La valeur maximale de Σ_1 nous donne la valeur maximale que peut avoir le paramètre ν . Il faut ensuite déterminer le temps de réponse du gène T correspondant au maximum de Σ_1 (voir section 2.4.3.1) et qui sera appelé T_{opt} . C'est l'échelle de temps pour laquelle on peut observer des oscillations avec des non-linéarités de dégradation les plus simples.

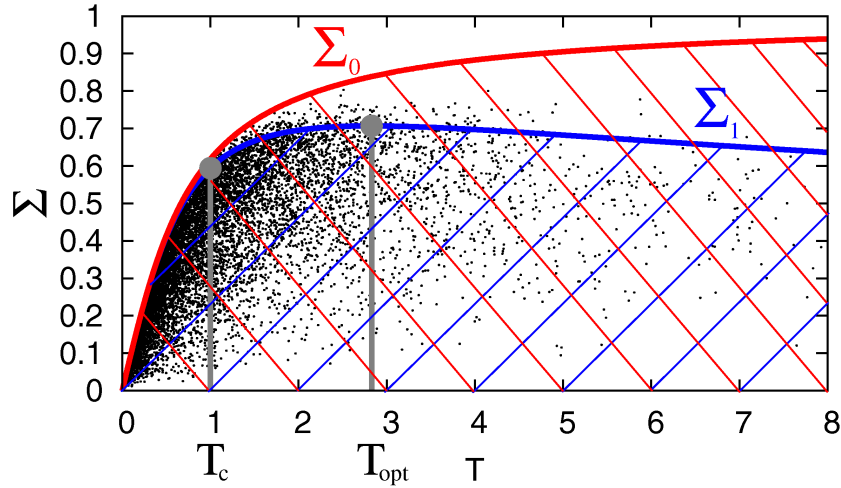


FIGURE 2.10 – Diagramme de phase du gène auto régulé obtenue à partir du critère (2.33). Les deux courbes limites sont Σ_0 en rouge et Σ_1 en bleu. Les points en noir correspondent à des ensembles de paramètres qui mènent à des oscillations en considérant une dégradation allostérique pour les protéines et une dégradation de Michaelis Menten pour l'ARNm. Nous tirons les paramètres au hasard dans les intervalles $\theta, \delta \in [10^{-1}, 10]$, $\theta/\alpha \in [10, 1000]$, $\lambda \in [1, 1000]$, $\lambda/\mu \in [10, 1000]$. Au-delà de la valeur $T_c = 1$, le temps de réponse du gène ne peut pas être négligé. $T_{opt} = 2\sqrt{2}$ correspond au maximum de Σ_1 .

2.4.2.2 Diagramme de phase

Le critère normalisé est :

$$H_\epsilon(\Sigma, T) = \Sigma \left[\frac{\epsilon^2 \Sigma^2}{4} T^2 + \left(\Sigma - \frac{1}{\Sigma} \right) T + 1 \right] < 0$$

La résolution numérique, visant à donner les zones d'oscillations, consiste à déterminer toutes valeurs du couple (Σ, T) tel que :

$$\frac{\epsilon^2 \Sigma^2}{4} T^2 + \left(\Sigma - \frac{1}{\Sigma} \right) T + 1 = 0 \quad (2.42)$$

Pour le cas trivial $\epsilon = 0$, la solution est :

$$\Sigma(T) = -1/2 + 1/2\sqrt{1 + 4T} \quad (2.43)$$

qui est une fonction monotone. Pour T grand par rapport à $1/4$, on a $\Sigma \approx \sqrt{T}$.

La courbe Σ_ϵ , tracée à ϵ fixé, est la frontière entre deux zones correspondantes à la présence ou l'absence d'oscillation. Il y a oscillation si $\Sigma \leq \Sigma_\epsilon(T)$, sinon le système est stable. Les deux courbes limites du critère (2.33) sont tracées pour $\epsilon = 0$ et pour $\epsilon = 1$. Pour tout ϵ compris entre 0 et 1, la courbe Σ_ϵ est comprise entre les deux courbes limites. Ainsi, tout ensemble de paramètre correspondant à un point au dessus de la courbe $\Sigma_0(T)$ n'entraîne pas d'oscillation tandis tout point en dessous de la courbes Σ_1 fait apparaître les oscillations. Nous traçons sur la figure 2.10 les courbes limites constituant le diagramme de phase.

2.4.2.3 Application avec une dégradation allostérique

Des ensembles de paramètres menant à des oscillations sont recherchés par tirage au sort numérique et correspondent à des points placés sur le diagramme de phase (2.10). Pour rester général, α et μ sont différents de 0 dans les simulations numériques montrant ainsi que le critère analytique reste valide dans un domaine des paramètres plus large, puisqu'il a été déterminé théoriquement avec $\alpha = \mu = 0$ ¹. La dégradation de la protéine est une dégradation allostérique [35] et la dégradation de l'ARN se fait par dégradation de Michaelis-Menten :

$$f(p) = \frac{p(a + p/\kappa)}{a + 2a(p/\kappa) + (p/\kappa)^2} \quad (2.44)$$

$$h(m) = \frac{\chi m}{\chi + m} \quad (2.45)$$

Sur le diagramme (2.10), nous posons les points obtenus. Il n'y a aucun point au-dessus de la courbe Σ_0 et le faible nombre de points au-dessus de Σ_1 ont un taux de dégradation proche de la saturation. Le critère (2.33) est donc une bonne approche analytique de la dynamique du réseau génétique.

2.4.2.4 Régions du diagramme de phase

L'expression $\Sigma(T)$ (2.42) est simplifiable selon les différentes régions du diagramme de phase 2.10. Dans la région T petit, le développement limité de la fonction $\Sigma_\epsilon(T)$ en T donne :

$$\Sigma_\epsilon(T) \approx -1/2 + 1/2\sqrt{1 + 4T} \approx T \quad (2.46)$$

Le critère est indépendant du paramètre ϵ car ce dernier est multiplié à T^2 et il ne peut pas être supérieur à 1. La condition d'obtention des oscillations (2.28) ($\sigma < g^*\sqrt{\delta\lambda}$) devient une condition suffisante pour avoir les oscillations. Dans la région $T > 1$, la valeur maximale de Σ_ϵ , correspondant au temps de réponse $T_m(\epsilon)$ est :

$$\Sigma_\epsilon = \frac{1}{\sqrt{1 + \epsilon}} \quad (2.47)$$

$$T_m(\epsilon) = 2\frac{\sqrt{\epsilon + 1}}{\epsilon} \quad (2.48)$$

La mesure de la non-linéarité ν s'écrit sous la forme :

$$\nu = \frac{g^*\sqrt{\lambda}}{2}\epsilon\frac{1}{\sqrt{1 + \epsilon}} \quad (2.49)$$

A chaque ϵ fixé, nous calculons un temps de réponse du gène $T_m(\epsilon)$ (soit une échelle de temps) favorisant les oscillations.

Dans la section suivante, nous déterminons le temps de réponse optimal du gène, c'est-à-dire le temps pour lequel ν est maximal. C'est la situation où la dégradation est la plus éloignée de la saturation.

1. Une approche alternative serait la recherche numérique des limites de bifurcation de l'espace des paramètres, en faisant varier continuellement α et μ . Mais il est probable qu'on obtienne une structure compliquée avec une ligne $\Sigma(T)$ pour chaque couple (α, μ) . Notre représentation est en fait une projection des points d'oscillations de l'espace des paramètres (Σ, T, α, μ) sur (Σ, T) .

2.4.3 Phénomène de résonance : facteur d'optimisation de l'apparition des oscillations

2.4.3.1 Temps de réponse optimale du gène

Nous donnons les valeurs de paramètres pour lesquelles la valeur du temps de réponse du gène optimise l'apparition des oscillations. Les dégradations sont le plus loin d'être saturées pour $\epsilon = 1$ (voir 2.3.2.4). Le temps de réponse optimal est :

$$T_{opt} = 2 \frac{\sqrt{\epsilon + 1}}{\epsilon} = 2\sqrt{2} \quad (2.50)$$

$$\Sigma_1 = \frac{1}{1 + \epsilon} = \frac{1}{\sqrt{2}} \quad (2.51)$$

Cela correspond à la valeur maximale de ν :

$$\nu = \frac{g^* \sqrt{\lambda}}{2\sqrt{2}} \quad (2.52)$$

Selon la relation (2.32) liant τ à T , nous établissons la valeur du temps de réponse du gène facilitant la bifurcation vers les oscillations :

$$\tau_{opt} = T_{opt}/\sigma_c = T_{opt}\tau_c = 2\sqrt{2} \frac{1}{g^* \sqrt{\lambda\delta}} \quad (2.53)$$

Nous écrivons le temps de réponse optimale du gène en fonction des paramètres avec dimension. Le paramètre $\tau_c = 1/\sigma_c$ est à comparer avec τ , l'échelle de temps de réponse du gène. Plus τ est petit devant τ_c , plus l'élimination adiabatique de la variable du gène est juste.

L'ordre de grandeurs des temps et de la période des oscillations sont obtenus pour les paramètres : $\lambda_0 = \beta_0 = 10 \text{ mn}^{-1}$, $P_0 = 100$ et $g^* = 0.5$. En remplaçant les paramètres sans dimensions par les paramètres avec dimensions, on a :

$$\tau_c = \delta_M \frac{1}{g^*} \sqrt{\frac{P_0}{\lambda_0 \beta_0}} \quad (2.54)$$

où τ_c donne l'échelle de temps caractéristique de la réponse du gène à partir de laquelle la régulation transcriptionnelle n'est pas instantanée. En redimensionnant à partir des égalités (2.16a), on a :

$$t_c = \frac{\tau_c}{\delta_M} = 2 \text{ min} \quad (2.55)$$

Sachant qu'à la bifurcation, la période des oscillations pour τ_{opt} est donnée par :

$$T_{osc} = 2\pi \sqrt{\frac{\tau_{opt}}{\sigma + \gamma\tau_{opt}}} = 4\pi \sqrt{\frac{2}{3}} \tau_c \quad (2.56)$$

Alors la période est d'environ 20 min à la résonance lors de la redimension tel que $t_c = \frac{\tau_c}{\delta_M} = 2 \text{ min}$.

Lorsque le temps de réponse du gène est inférieur d'au moins un ordre de grandeur par rapport à la période des oscillations, la variable du gène g est une mesure de l'activité moyenne

car de nombreuses fixations de la protéine sur le promoteur (et relargages) ont lieu durant un cycle. Mais comme la dynamique transcriptionnelle est constituée d'une succession de réactions discrètes [143], la pertinence d'une telle étude peut se poser lorsqu'on sort du domaine des oscillations déterministes. En fait, le temps de réponse du gène a-t-il un sens lors d'une étude stochastique du système ?

2.4.3.2 Intérêt de l'étude stochastique

Nous proposons une étude stochastique de notre modèle afin de voir si le temps de réponse optimale du gène τ_{opt} reste pertinent. La variable représentant le gène G est aléatoire en pouvant prendre 2 valeurs distinctes liées à la régulation (valeur 0 si le gène est éteint ou valeur 1 si le gène est actif). L'intégration numérique des équations (2.16) est réalisée grâce à l'algorithme de Gillespie [40] (voir section 1.2.7).

Le coefficient η détermine la régularité du temps entre chaque pic d'expression des protéines :

$$\eta = \frac{\sigma_{\Delta t}}{\langle \Delta t \rangle} \quad (2.57)$$

La grandeur $\langle \Delta t \rangle$ est l'intervalle de temps moyen entre chaque pic d'expression et $\sigma_{\Delta t}$ sa variance. Comme d'un cycle à l'autre l'amplitude du pic des protéines est fluctuante, l'intervalle de temps Δt entre deux pics est défini par la condition suivante : c'est le temps entre deux passages de la concentration de la protéine P à la valeur $P = 1.2P_{avg}$, séparé par au moins un passage à $P = 0.8P_{avg}$. La grandeur P_{avg} est la concentration moyenne de la protéine. Autrement dit, une variation d'amplitude d'au moins 40 % est considérée comme le passage d'un cycle d'intervalle de temps Δt . L'intégration stochastique du système (2.16) ne donne pas d'oscillations stochastiques car les fluctuations sur les niveaux de protéines et la valeur du coefficient de variation η sont trop grandes. Cependant, la valeur du coefficient η dépend de la valeur de τ . Nous traçons à la figure (2.11) plusieurs courbes de η en fonction de τ . Le minimum de η est atteint dans un voisinage de τ_{opt} pour n'importe quelle valeur du seuil de régulation. Donc dans ce voisinage, l'intervalle de temps Δt est plus régulé. C'est une preuve que notre étude déterministe est également pertinente en stochastique car les effets stochastiques sont minimisés autour d'une valeur déterminée par une étude en déterministe.

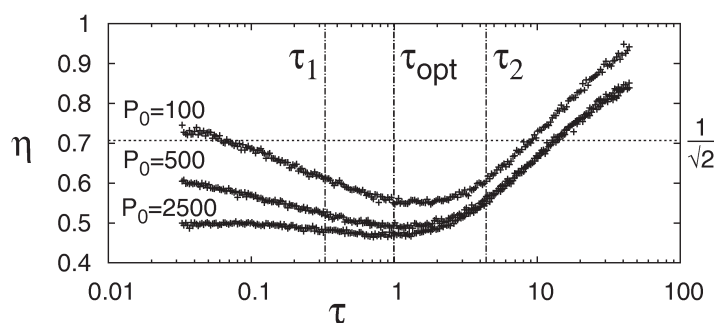


FIGURE 2.11 – Coefficient de variation η en fonction du temps de réponse du gène τ ou temps de résidence du facteur de transcription sur le gène. Les valeurs sont obtenues à partir des simulations stochastiques de 2.16. Valeurs des paramètres : $\delta = 1, \theta = 1, \alpha = \theta/P_0, \lambda = 21.54, \mu = 0.085, \chi = 95.5, \kappa = 21.68, a = 10^{-6}$. Les frontières délimitées par τ_1 et τ_2 correspondent à des domaines d'oscillation déterministe.

2.5 Conclusion

Dans nos travaux, nous avons montré que le temps de réponse du gène, le délai nécessaire pour que le gène ait une activité maximale, a une influence importante dans l'obtention des oscillations d'un gène régulé par sa propre protéine. Notre approche est originale car elle s'est démarquée de la modélisation habituelle des oscillations dans les réseaux monogénétiques, et qui se focalisait sur les effets provoqués soit par un délai, soit par des réactions fortement non-linéaires. Cependant, aucun travaux n'avait relié, grâce à une relation analytique, comment ces deux ingrédients interagissent ensemble pour générer des oscillations.

Le critère analytique permettant de déterminer les valeurs de paramètres conduisant aux oscillations nous a permis de prédire l'existence des oscillations en tenant compte d'un temps de réponse du gène variable. Lorsque le temps de réponse est nul, une forte non-linéarité au niveau de la dégradation est nécessaire pour obtenir des oscillations. La dégradation comporte soit une pente négative (une grande quantité de substrat défavorise sa propre dégradation), soit une saturation très importante. Nous avons retrouvé alors des résultats déjà connus. Cependant, une saturation importante est souvent difficile à atteindre pour des concentrations biologiques réalistes. Mais la dynamique transcriptionnelle lente du gène, dont l'existence est démontrée par l'expérience, permet de diminuer la saturation de la dégradation tout en conservant des oscillations. Il existe ainsi une valeur du temps de réponse pour laquelle la non-linéarité requise pour les oscillations est minimale. D'ailleurs, nous avons obtenu une expression analytique T_{opt} , le temps de réponse du gène optimisant l'apparition des oscillations.

Comme les interactions dans les réseaux génétiques sont par nature stochastiques, nous avons également réalisé une simulation stochastique de notre système, avec pour but de démontrer que notre approche déterministe reste pertinente. Nous avons alors remarqué que, dans un voisinage proche du temps de réponse T_{opt} , la régularité du temps entre les pics d'expression des protéines s'améliore grandement. Nous avons ainsi démontré que la dynamique transcriptionnelle lente du gène a une influence fondamentale dans le comportement de réseau génétique.

Chapitre 3

Nouveau modèle circadien d' *Ostreococcus Tauri*

3.1 Stratégie d'étude du rythme circadien

3.1.1 Interface biologie et mathématique

L'omniprésence et l'importance du rythme circadien dans les organismes vivants (voir section 1.4 page 34) nous incite à mieux comprendre ce phénomène. Grâce aux divers développements technologiques, nous disposons de données expérimentales fiables pour suivre les rythmes (voir section 1.5 page 45). L'analyse théorique de ces derniers met en oeuvre des outils mathématiques qui expliquent comment les oscillations apparaissent. Les mathématiques mènent également à de nouvelles approches permettant la découverte d'acteurs moléculaires vitaux pour le fonctionnement d'un réseau circadien [15]. Les mathématiques pourraient aussi permettre à la médecine de traiter des maladies liées au dérèglement des rythmes dans certains cancers [16].

Dans cette section, nous présentons la stratégie d'étude des rythmes circadiens d' *Ostreococcus Tauri*. Elle se sépare en deux parties : la détermination expérimentale des acteurs clés de l'horloge circadienne et la validation de ce résultat par le modèle mathématique, confirmant l'importance des gènes identifiés au coeur du réseau génétique circadien.

Nous verrons que l'horloge circadienne est construite sur la base d'un réseau génétique à rétroaction négative à deux gènes. La protéine *TOC* active la transcription du gène *cca* et la protéine *CCA* réprime la transcription du gène *toc*.

3.1.2 Modèle biologique

Nous décrivons ici les principaux résultats expérimentaux obtenus par Bouget et al. [92, 150, 151, 93]. Ils ont identifié deux gènes de l'horloge *toc1* et *cca1* à partir de la comparaison de séquences génomiques avec le modèle *Arabidopsis Thaliana* [93]. Ensuite, la présence d'un rythme circadien est vérifiée expérimentalement à partir des mesures temporelles de l'expression de ces gènes. Ces expressions sont fournies par des données d'ARNm, d'activités traductionnelles et transcriptionnelles de *toc* et de *cca*. Bouget et al. ont utilisé les micropuces à ADN pour prouver l'oscillation sur 24 heures des concentrations d'ARNm [93]. L'imagerie par la luciférase [89] a été utilisée pour obtenir les profils des protéines (activité traductionnelle)

et de l'activité transcriptionnelle (voir section 1.5.3). Bouget et al. ont créé plusieurs lignées d'*Ostreococcus Tauri* luminescentes. Le premier type de lignée donne accès à l'activité transcriptionnelle du promoteur : les signaux reflètent de la régulation du promoteur par des facteurs de transcription [92]. Le deuxième type de lignée donne l'activité traductionnelle : des photons sont émis par les protéines *TOC1* et *CCA1* fusionnées avec la luciférase [94, 81] (voir annexe B). Des lignées mutantes de surexpression ou d'antisens (qui supprime l'expression du gène) sont également construites afin de constater l'effet de l'altération de ces gènes sur l'horloge.

Ces expériences n'indiquent pas si les facteurs de transcription *TOC1* et *CCA1* interagissent directement avec les promoteurs *cca1* et *toc1* respectivement, aussi, Bouget et al. ont réalisé un test de mobilité sur gel pour connaître quelles sont les régulations directes des facteurs de transcription *TOC1* et *CCA1*. Ils ont alors découvert que *CCA1* se fixe sur le promoteur de *toc*.

3.1.3 Modèle mathématique

Le modèle mathématique que nous allons utiliser est celui d'un réseau à deux gènes basé sur une boucle transcriptionnelle à rétroaction négative : la protéine *TOC1* active la transcription du gène *cca1* tandis que la protéine *CCA1* réprime la transcription du gène *toc1*. C'est en effet l'hypothèse naturelle si l'on tient compte de la manière dont interagissent les homologues de ces deux gènes chez *Arabidopsis*. Le modèle est écrit à partir de la loi cinétique d'action de masse et est minimal car il ne contient que les réactions liées aux régulations, à la traduction et à la dégradation (voir section 1.2). Le travail consiste ensuite à déterminer les paramètres des équations donnant un bon accord entre les données expérimentales et les résultats numériques. Les données expérimentales sont d'abord traitées pour construire des profils expérimentaux cibles. Ils sont ensuite comparés avec les profils numériques, issus de l'intégration numérique des équations de l'horloge centrale. Nous verrons qu'un bon accord entre les profils numériques et les profils expérimentaux sera obtenu.

De manière surprenante, les profils s'ajustent mieux en l'absence de couplage à la lumière, c'est-à-dire que le couplage n'est pas apparent lorsque l'horloge est synchronisée au cycle jour-nuit. Thommen et al. ont exploité ce résultat et ont mis en évidence les propriétés de robustesse aux fluctuations lumineuses de l'horloge circadienne de l'organisme *Ostreococcus Tauri*. Ils ont montré que si le couplage à la lumière du réseau génétique se fait dans une fenêtre temporelle et spécifique, le réseau minimise l'impact des fluctuations lumineuses lorsque l'horloge est en phase avec le cycle jour/nuit. Cependant, si l'horloge est déphasée, la modulation paramétrique durant cette fenêtre de couplage à la lumière est suffisante pour provoquer la remise en phase avec le Zeitgeber (les signaux du cycle jour/nuit) [94].

En outre, l'ajustement simultané de profils expérimentaux de deux types d'expériences très différentes (micropuces et gène rapporteur) démontre également une autre forme de robustesse de l'horloge face aux variations environnementales et expérimentales.



FIGURE 3.1 – Photographie microscopique d'*Ostreococcus Tauri* tirée de [153]. Nous distinguons dans la partie supérieure le chloroplaste et dans la partie inférieure le noyau.

3.2 Rythme circadien d'*Ostreococcus Tauri*

3.2.1 Présentation de l'organisme

L'organisme *Ostreococcus Tauri* est une algue verte microscopique et unicellulaire. Découvert en 1991 dans l'étang de *Thau* situé au sud de la France, cet organisme est répandu dans les eaux salées du monde entier [152]. On pense qu'il s'agit du plus petit organisme eucaryote (c'est-à-dire que son ADN est contenu dans un noyau). Il est d'une taille comparable aux bactéries dépourvues de noyau.

Les propriétés de l'organisme *Ostreococcus Tauri* en font un modèle biologique de référence. Son organisation cellulaire est simple, le génome est relativement petit et compact avec peu de gènes redondants. La photosynthèse est assurée par des chloroplastes¹ [153], ce qui en fait un phytoplancton. Les espèces de picophytoplanctons eucaryotes sont répandues dans le monde entier et ont une importance fondamentale pour les cycles bio et géochimiques de la planète. La photosynthèse de ces organismes contribue à la production de carbone organique en assimilant le dioxyde de carbone des océans [154, 155]. Il apparaît donc important de comprendre les différentes caractéristiques physiologiques et les rythmes biologiques contrôlant cet organisme [103]. Parmi les rythmes biologiques intéressants, nous trouvons le cycle cellulaire contrôlé par la voie métabolique liée à la photosynthèse. Mais Bouget et al. ont montré que les principales protéines du cycle cellulaire étaient sous contrôle circadien [92]. Ainsi, la compréhension du rythme circadien nous renseignera sur le cycle du carbone.

1. Les chloroplastes sont des organites essentiels à la photosynthèse. Ils captent la lumière sur un spectre donné et convertissent l'énergie du photon en énergie de liaison chimique grâce à la chlorophylle qu'ils contiennent.

3.2.2 Identification de deux gènes de l'horloge : TOC et CCA

Le modèle circadien d'*Arabidopsis Thaliana* est une référence pour les plantes vertes [156] (voir section 1.4.5.2) et a des liens phylogénétiques avec *Ostreococcus Tauri*. La recherche, par Correllou et al., de gènes orthologues d'*Arabidopsis Thaliana* dans l'algue *Ostreococcus Tauri* est à la base de l'identification des gènes de l'horloge circadienne. Le tableau 3.2 indique des orthologues des deux espèces et le nombre d'homologues de chaque gène. Les gènes de la famille *cca/lhy* (Circadian Clock Associated/Late elongated Hypocotyl) et *toc* (Timing of CAB expression-1) sont des gènes centraux de l'horloge d'*Arabidopsis Thaliana*. Dans cette plante, la topologie du réseau génétique circadien est une boucle de régulation transcriptionnelle négative dans laquelle *TOC* active *lhy/cca* et *LHY/CCA* réprime *toc*. Le modèle de cette plante inclut au moins deux acteurs génétiques supplémentaires pour ajuster correctement les données expérimentales [15].

Comme les gènes *toc* et *cca* sont au coeur de l'horloge circadienne d'*Arabidopsis Thaliana*, leurs homologues peuvent aussi avoir un rôle important dans l'horloge d'*Ostreococcus Tauri*. De plus, la faible redondance génomique dans *Ostreococcus Tauri* devrait simplifier le rapport entre la théorie et l'expérience. Les gènes *toc* et *cca*, appelés respectivement *toc1* et *cca1*, sont en effet présents en un seul exemplaire (par abus de langage, nous noterons indifféremment *toc* ou *toc1*, idem pour *cca*). Les autres gènes du tableau 3.2 sont impliqués dans la perception de la lumière provoquant la synchronisation au cycle lumineux. Même si l'étude des mécanismes d'interactions entre les molécules et la lumière est intéressante d'un point de vue biologique [157] et dynamique (présence de cascade de phosphorylation [158]), la prise en compte détaillée de ces mécanismes moléculaires de photoréception n'est pas nécessaire pour démontrer l'importance des gènes *toc* et *cca* dans le fonctionnement de l'horloge. Au contraire même, l'ajout de photorécepteurs obscurcit la compréhension du modèle en le complexifiant.

Dans la section suivante, nous exposons les résultats expérimentaux obtenus avec les microcopes à ADN qui apportent une première information sur les profils d'expression génique.

		<i>Arabidopsis</i>	<i>Ostreococcus</i>
famille de gènes	rôle dans l'horloge circadienne	Nombre de copie	Nombre de copie
<i>cca1 lhy</i>	gènes centraux	11	1
<i>toc1</i>	gènes centraux	5	1
Cryptochromes (<i>cry</i>)	perception lumière bleu	3	2
Phototropins	perception de la lumière bleu	2	1

FIGURE 3.2 – Étude comparative de gènes impliqués dans l'horloge circadienne (voies d'entrée et oscillateurs centraux de l'horloge) d'*Ostreococcus* et d'*Arabidopsis Thaliana*. Les deux gènes centraux de l'horloge chez *Arabidopsis Thaliana* sont *toc* et *cca*. Le nombre de copie de ces gènes est assez élevé par rapport à *Ostreococcus Tauri*.

3.2.3 Données d'expression de l'ARN messenger par les micropuces à ADN

Les micropuces à ADN ont permis de fournir des données précises [105] sur l'expression d'un grand nombre de gènes d'*Ostreococcus Tauri*, lorsque les populations de cellules sont soumises à l'alternance de 12 heures de jour et de 12 heures de nuit [151]. Les profils de l'expression de l'ARNm des deux gènes de l'horloge (*toc* et *cca*) sont représentés sur la figure 3.3.

Les concentrations de l'ARNm des deux gènes oscillent au cours du temps. Ces oscillations sont nécessaires mais pas suffisantes pour démontrer une régulation circadienne. Par exemple, les oscillations de l'ARN sont-elles maintenues en condition d'éclairage constant ? Mais ces oscillations prouvent une régulation transcriptionnelle des gènes *toc1* et *cca1*, comme montré par Bouget et al. [92]. Il reste à confirmer cela par des expériences supplémentaires établissant l'importance de ces deux gènes dans l'horloge. Ces expériences ont été réalisées en utilisant des lignées transformées pour émettre un signal luminescent.

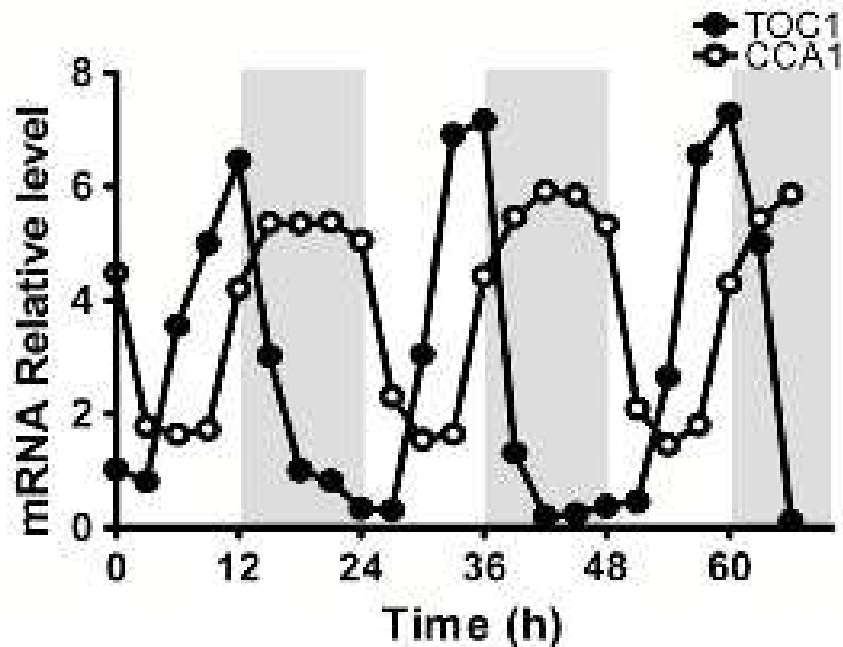


FIGURE 3.3 – Figure extraite de [93]. Résultat des micropuces à ADN donnant l'expression au cours du temps de l'ARN messenger *TOC* et *CCA*. Les trois cycles présentés ne se suivent pas dans le temps. À un instant $t = 0$ heure, trois échantillons sont extraits pour mesurer l'expression de l'ARN messenger. Les signaux obtenus sont mis à la suite. Les expériences sont réalisées avec une alternance de 12 heures de lumière (zone blanche) et 12 heures d'obscurité (zone grisée), soit un entraînement LD 12 : 12 (LD pour Light Dark). L'amplitude entre chaque pic des signaux fluctue faiblement signifiant une bonne reproductibilité des données.

3.2.4 Données de luminescences de l'activité génique et des concentrations de protéines

La luciférase est une protéine enzymatique catalysant l'émission de lumière par la luciférine (voir section 1.5.3). Cette dernière émet de la lumière verte. Cette propriété est utilisée pour suivre l'activité transcriptionnelle du gène ou la quantité de protéine [90]. La luciférine est injectée en quantité abondante dans le milieu au début des expériences de manière à ne pas être un facteur limitant.

3.2.4.1 Lignées de fusion transcriptionnelle

Un vecteur permet l'insertion dans le génome d'une fusion du promoteur d'intérêt et du gène codant la luciférase [93]. Selon la régulation du promoteur par les facteurs de transcription (voir section 1.1), l'ARN puis la protéine de luciférase sont synthétisés dans la cellule menant à une émission lumineuse. Dans notre modèle, un seul photon de lumière est émis grâce à la luciférase. Il n'y a pas de réactivation possible de cette dernière qui devient inactive [109]. Pour les modèles circadiens, l'intensité lumineuse mesurée oscille sur 24 heures.

Le décalage entre la véritable activité transcriptionnelle et le signal lumineux dépend du temps de synthèse de la luciférase ainsi que du temps d'émission de la lumière par cette dernière. Si ces temps sont courts, alors le déphasage entre le vrai pic d'activité du promoteur et le pic de lumière émise est faible. Cependant, cette durée n'influe pas sur l'interprétation qualitative des données de l'activité transcriptionnelle car une oscillation des signaux lumineux témoigne bien d'une régulation transcriptionnelle circadienne.

Bouget et al. ont créé deux types de lignées : un type *pTOC :LUC* et un type *pCCA :LUC*. Dans la lignée *pTOC :LUC* (resp. *pCCA :LUC*), le génome de la cellule contient un gène issu de la fusion du promoteur du gène *toc* (resp. *cca*) avec la partie codante du gène de la luciférase. À l'origine, chaque lignée est issue d'une seule cellule donnant naissance à une population après de multiples duplications.

Bouget et al. ont créé plusieurs lignées de *pTOC :LUC* et de *pCCA :LUC* afin de vérifier la reproductibilité expérimentale. Les différentes lignées *pTOC :LUC* (ou *pCCA :LUC*) sont susceptibles d'avoir des expressions génétiques différentes (peut-être un nouveau phénotype) car la valeur des paramètres dynamiques de la transcription dépend de la position du gène dans l'ADN [159, 160]. Or, lorsqu'un gène est inséré dans le génome, sa position n'est pas contrôlée. Les taux de transcription de chaque lignée peuvent donc différer. Comme les concentrations des acteurs du réseau génétique ont une dynamique non linéaire, une modification des taux basaux n'entraîne pas forcément une modification linéaire des concentrations moléculaires du réseau. De fortes altérations dans la forme de certains signaux lumineux issus de lignées différentes sont envisageables.

Ainsi, une lignée doit être issue d'une seule cellule modifiée génétiquement pour assurer un génome de même nature dans l'ensemble de la population. Le signal d'une population mélangeant des cellules venant de différentes expériences d'insertion génétique d'un type donné ne serait pas correctement interprétable.

3.2.4.2 Lignées de fusion traductionnelle

Une stratégie similaire est adoptée pour suivre l'activité traductionnelle, c'est-à-dire l'activité de synthèse de la protéine luminescente. Bouget et al. ont créé deux autres types de lignées appelés *TOC :LUC* et *CCA :LUC*. L'insertion dans le génome comprend maintenant le promoteur du gène *toc1* (resp. *cca1*) ainsi que les deux séquences codantes pour la protéine *toc1* (resp. *cca1*) et la luciférase. La transcription de ce gène, suivie de la traduction de l'ARNm, produit une nouvelle protéine dans la cellule. Cette protéine est la fusion de la protéine *TOC1* (ou *CCA1*) avec la protéine de la luciférase et s'appelle la protéine *TOC :LUC* (resp. *CCA :LUC*). La luciférase émet ensuite un seul photon de lumière sous l'action de *TOC :LUC* (resp. *CCA :LUC*). Cette dernière devient ensuite invisible.

L'interprétation du signal lumineux n'est pas immédiate car elle dépend de la durée de vie des protéines *TOC :LUC* (et pour *CCA :LUC*) par rapport au temps d'émission de lumière. Deux situations extrêmes sont possibles. Si la protéine *TOC :LUC* émet de la lumière juste après la traduction, le signal de lumière mesure le flux des protéines nouvellement créées. Dans ce cas, le temps d'émission de lumière est beaucoup plus faible que la durée de vie de la protéine *TOC :LUC*. Le signal luminescent est proportionnel (et presque en phase) avec le profil de concentration de l'ARNm de *TOC :LUC*. Si la durée de vie de la protéine est plus faible que le temps d'émission de lumière, ou que l'émission a lieu longtemps après la traduction, alors le signal lumineux est proportionnel à la concentration de protéine *TOC :LUC*. Donc, la détermination de ces deux durées est importante pour comprendre correctement les signaux lumineux.

Dans ce chapitre, les signaux de luminescence des lignées traductionnelles *TOC :LUC* (ou *CCA :LUC*) sont supposés directement assimilables aux concentrations de protéines *TOC* (ou resp. *CCA*). Dans le chapitre 4, nous montrerons pourquoi cette simplification est justifiée et modéliserons l'horloge circadienne en tenant compte de la cinétique de la luciférase.

3.2.4.3 Signaux d'activité transcriptionnelle et traductionnelle en cycle jour nuit

Bouget et al. ont vérifié la présence d'oscillations de l'activité transcriptionnelle et traductionnelle dans les conditions d'entraînement par un cycle de lumière. Chaque lignée est placée dans un puits individuel et est soumise à un cycle jour nuit (Light Dark : 12h/12h ou LD : 12/12) de 12 heures de jour et de 12 heures de nuit. La luminescence est mesurée par un photomultiplicateur. Les 4 courbes montrées à la figure 3.4 représentent la moyenne des lignées de chaque type de transformations génétiques. Nous montrerons à la section 3.3.3 les profils individuels des différentes lignées plutôt que leur moyenne.

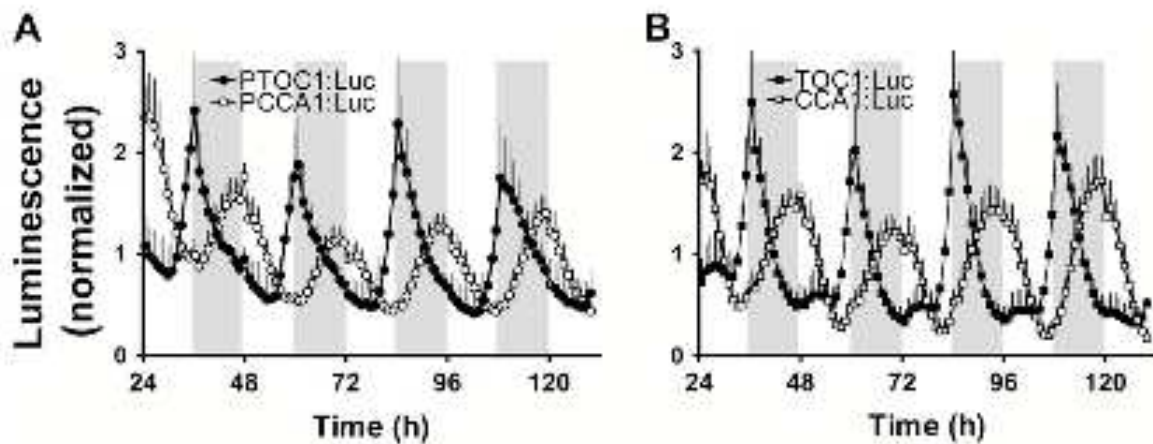


FIGURE 3.4 – Figures extraites de [93]. Figure A, carré noir : moyenne des signaux de luminescence des lignées transcriptionnelles *pTOC :LUC*. A, carré blanc : moyenne des signaux de luminescence des lignées transcriptionnelles *pCCA :LUC*. B : moyenne des signaux de luminescence des lignées *TOC :LUC* (carré noir) et des lignées *CCA :LUC* (carré blanc). Les populations de cellules sont soumises à 12 heures de jour (domaine blanc) et 12 heures de nuit (domaine gris).

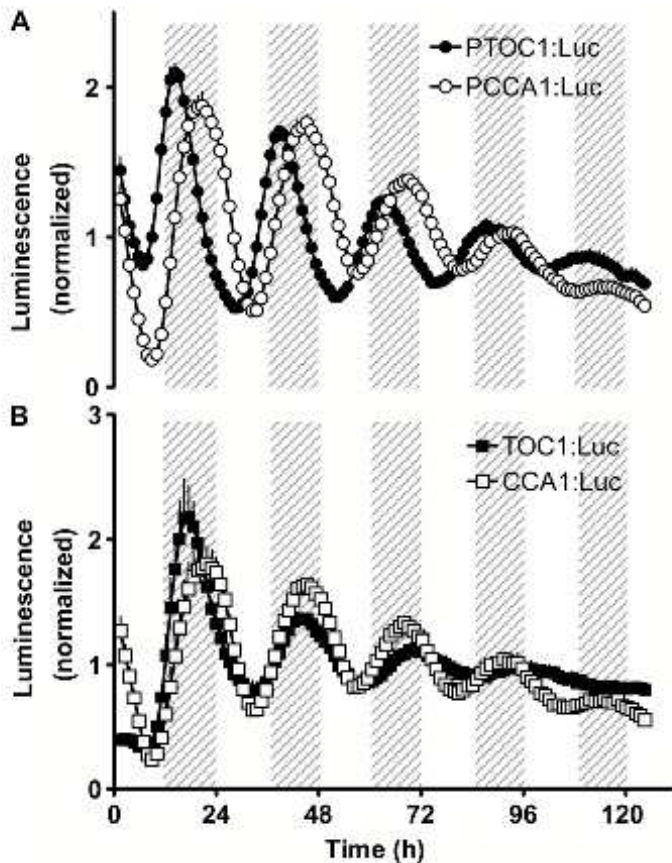


FIGURE 3.5 – Figures extraites de [93]. A : moyenne des signaux de luminescence des fusions transcriptionnelles de *pTOC :LUC* (points noirs) et de *pCCA :LUC* (points blancs). B : moyenne des signaux de luminescence des fusions traductionnelles de *TOC :LUC* et de *CCA :LUC*. Les expériences sont réalisées en jour continu.

3.2.4.4 Signaux d'activité transcriptionnelle et traductionnelle en jour continu

F-Y Bouget et al. ont relevé les signaux de luminescence des fusions transcriptionnelles et traductionnelles en condition de lumière constante (conditions LL, figure 3.5). Ces expériences montrent la persistance du rythme oscillant en absence de cycle lumineux et fournissent la période de l'oscillateur circadien en libre cours (*free running*).

L'amplitude des oscillations décroît pour disparaître au bout de 120 heures d'expériences. La diminution de l'amplitude des oscillations ne signifie pas nécessairement que chaque cellule agit comme un oscillateur faiblement amorti. Chaque cellule peut osciller en jour continu et se désynchroniser des autres car la variabilité stochastique fait que les cellules n'ont pas strictement toutes la même période [161]. Si la distribution des périodes est assez large, la somme des amplitudes s'atténue tandis que l'oscillation du rythme circadien des cellules individuelles est conservée. Nous admettons ici que les oscillations sont maintenues en jour continu et que l'amortissement est dû à une désynchronisation des cellules (voir section 3.3.3.2).

3.2.5 Effets de la surexpression et de la répression des gènes de l'horloge

Les expériences précédentes montrent des oscillations circadiennes de *toc1* et de *cca1* sans indications sur le rôle des protéines *TOC1* et *CCA1* en tant que facteur de transcription et sur leur influence dans la boucle transcriptionnelle de l'horloge. Ces informations sont fournies en sur ou sous-exprimant l'expression des gènes dans des lignées mutantes. A priori, la sur ou sous-expression d'un gène de l'horloge altère fortement l'expression des autres gènes du réseau génétique au point de casser les oscillations s'il est impliqué dans la boucle.

F-Y Bouget et al. ont construit de nouvelles lignées mutantes sur ou sous-exprimant l'expression du gène *toc1* ou du gène *cca1*. Pour la surexpression, Bouget et al. fusionnent un gène *toc1* (resp. *cca1*) avec un promoteur HAPT (High Affinity Promoter Transport). Ce promoteur rend l'activité du gène très forte. Ils ont ainsi créé deux types de vecteurs pouvant générer des lignées de surexpression du gène *toc1* et *cca1* appelées respectivement *TOC1-ox* et *CCA1-ox*. Un de ces gènes est ensuite inséré dans une lignée *TOC :LUC* ou *CCA :LUC*. Cela nous permet d'étudier l'effet de la surexpression de *cca1* sur *toc1* grâce à la luminescence de la protéine *TOC1 :LUC* par exemple (différentes combinaisons sont possibles).

La figure 3.6 présente également les signaux lumineux de *TOC1 :LUC* et de *CCA1 :LUC* pour une surexpression de *toc1* ou de *cca1*. Les expériences sont réalisées en jour continu (ou en LL). Une disparition très rapide des oscillations de *TOC1 :LUC* et de *CCA1 :LUC* est visible et conduit à une arythmie. La boucle de régulation transcriptionnelle est donc "cassée" lors de la surexpression d'un seul des deux gènes de l'horloge.

F-Y Bouget et al. ont créé des lignées où l'expression de *toc1* et de *cca1* sont supprimées. Ils insèrent donc dans le génome de la lignée *TOC1 :LUC* (resp. *CCA1 :LUC*) un gène né de la fusion du gène *cca1* (resp. *toc1*) et du promoteur HAPT. Mais le gène *cca1* (resp. *toc1*) est placé dans le sens de lecture inverse de l'ADN. Des ARN anti-sens de *cca1* sont produits et s'hybrident avec les ARNm de *cca1* pour empêcher la traduction en protéine. De telles lignées sont appelées *TOC1-as* et *CCA1-as*.

Une expérience (voir figure 3.6) montre la luminescence de *CCA1 :LUC* avec de l'antisens *TOC1-as*. L'oscillation de *CCA1 :LUC* disparaît très rapidement. Le facteur de transcription *TOC1* joue donc un rôle très important dans la boucle. Une autre expérience permet de suivre la luminescence de *TOC1 :LUC* avec de l'antisens *CCA1-as*. Dans cette expérience, le gène antisens *CCA1-as* ne détruit pas les oscillations de *TOC1 :LUC*. Bouget et al. en conclut que le gène *cca1* n'est pas primordial pour l'horloge.

Cependant, ils remarquent que la répression de l'expression de *cca1* dans les lignées *CCA-as* n'est peut-être pas assez efficace et qu'elle laisse un nombre faible de protéine *CCA1*. Il se peut que cette quantité soit suffisante pour réguler le gène *toc1*. En effet, cela est possible si le seuil de régulation du promoteur *toc1* par *CCA* est faible. D'ailleurs, cette dernière hypothèse est confirmée par les simulations numériques du modèle mathématique (voir sous-section 3.3.6 page 107).

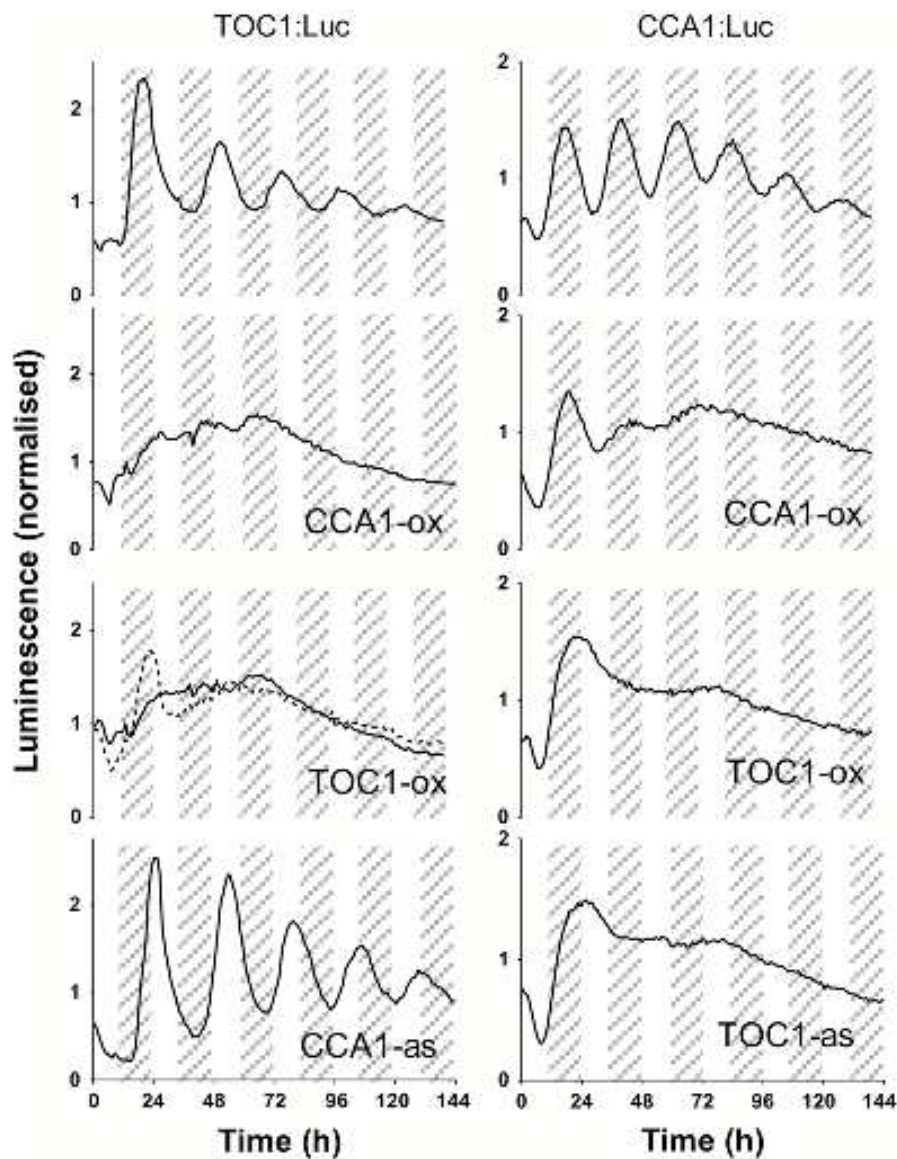


FIGURE 3.6 – Figures extraites de [93]. Exemples de signaux de luminescence de *TOC :Luc* et de *CCA :Luc* dans les différentes lignées de surexpression et antisens en condition de lumière continue. La nuit subjective est représentée par les domaines gris. La surexpression de *CCA* modifie le rythme de *TOC :Luc* (sur 23 lignées, 13 lignées n’oscillent plus), modifie le rythme de *CCA :Luc* (sur 22 lignées, 6 n’oscillent plus). Les lignées antisens de *CCA* n’ont aucun effet sur *TOC :Luc*. La surexpression de *TOC* modifie le rythme de *TOC :Luc* (sur 13 lignées, 7 n’oscillent plus) et de *CCA :Luc* (sur 42 lignées, 29 n’oscillent plus). Les lignées antisens de *TOC* peuvent détruire les oscillations des lignées *CCA :Luc* (sur 21 lignées, 9 n’oscillent plus).

3.2.6 Régulation directe de la protéine *CCA1* sur le promoteur du gène *toc1*

Bouget et al. ont également recherché des preuves d'interactions directes des facteurs de transcription sur les promoteurs de *toc1*.

Il existe dans le promoteur du gène *toc1* d'*Arabidopsis Thaliana* une séquence appelée Evening Element [162]. Cette séquence interagit avec la protéine *CCA1* et provoque la répression du gène [163]. FY Bouget et al. ont fait un test de mobilité sur gel par électrophorèse de la protéine *CCA1* avec l'Evening Element du promoteur de *toc1*. Celui-ci a révélé une interaction importante *in vitro*. Par la suite, FY Bouget et al. ont procédé à la mutation de quelques nucléotides de cette séquence afin d'éliminer l'interaction avec la protéine *CCA1*. Puis ils ont mesuré les signaux de luminescence, en lumière continue, des lignées mutées. Sur la figure 3.7, le rythme des signaux de luminescence est fortement altéré suggérant que l'Evening Element est une séquence clef dans les processus de régulation du gène de *toc1* par la protéine *CCA1*.

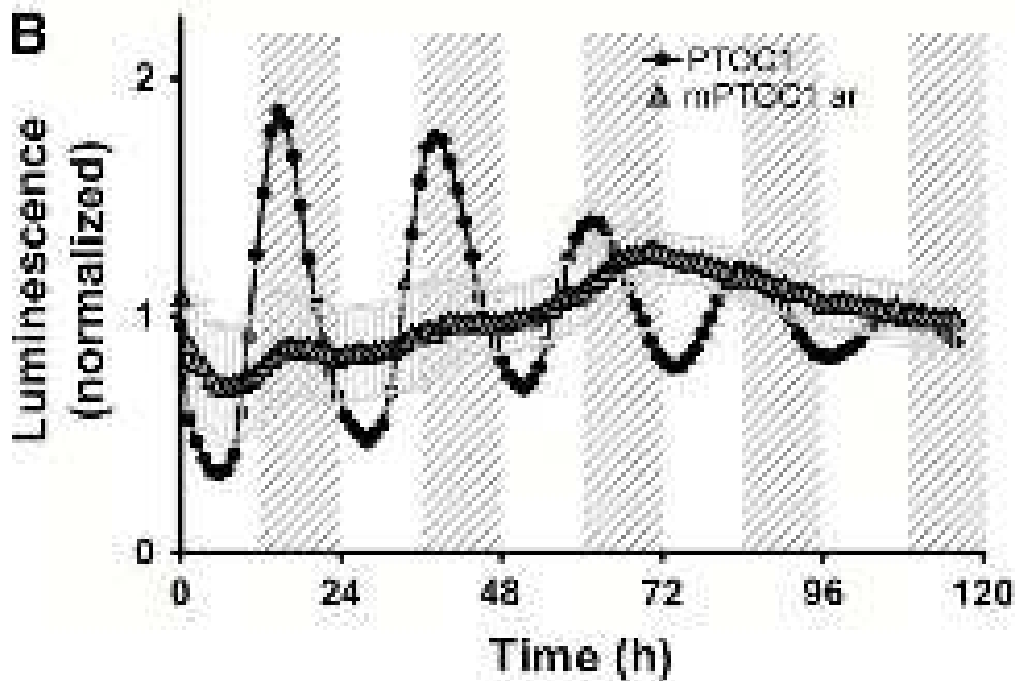


FIGURE 3.7 – D'après [93]. Signaux de luminescence de *pTOC :Luc*. Les signaux sont obtenus en condition de lumière constante après un entraînement de 12 heures de jour et de 12 heures de nuit. Points carrés : signaux de *pTOC :Luc* oscillant et servant de contrôle. Points triangles : lignées *pTOC :Luc* ayant subi une mutation de l'Evening Element. Le signal est une moyenne de 18 signaux dont le rythme est altéré. La mutation de l'Evening Element provoque systématiquement une perte du rythme du signal de *pTOC :Luc*.

3.2.7 Réseau génétique à deux gènes au coeur de l'horloge circadienne

Les lignées de sur ou sous-expression de *toc1* altèrent fortement le rythme de sa propre expression et de celle de *cca1* en condition de lumière constante. Le gène *toc1* participe à la régulation du gène *cca1* et indirectement à sa propre régulation. De plus, le niveau des signaux de luminescence des lignées *CCA1 :Luc* arythmiques à cause de la surexpression de *toc1* est plus élevé que celui des lignées de contrôle. Cela suggère que *TOC1* agit comme un activateur du gène *cca1*. Ainsi, le gène *toc1* est un gène vital de la boucle transcriptionnelle de l'horloge. Cependant, aucune interaction directe de *TOC1* sur une séquence du promoteur de *cca1* n'a été mise en évidence.

La surexpression du gène *cca1* altère le rythme de *TOC1* et son propre rythme. De plus, la surexpression de *cca1* diminue le niveau de luminescence des lignées de *TOC1 :Luc* arythmiques par rapport aux lignées de contrôle, signifiant une probable répression du gène *toc1*. Mais la lignée antisens *CCA1-as* ne casse pas le rythme de *TOC1 :LUC*, peut-être à cause d'une inefficacité de l'antisens. En revanche, une interaction directe entre la protéine *CCA1* et le gène *toc1* est observée.

A partir de ces résultats, il n'est pas possible de comprendre l'importance et le rôle exacte des deux gènes au sein de l'horloge. La vision d'une simple boucle à rétroaction négative impliquant les deux gènes est contredite par l'absence d'interaction directe entre *TOC1* et *cca1* et par l'absence d'arythmie dans les lignées *CCA1-as*.

Il faut prendre en compte de la difficulté des expériences réalisées et donc ne pas conclure trop rapidement sur le rôle des gènes *toc1* et *cca1*. D'abord, l'absence d'une démonstration de régulation directe peut être due aux nombres importants de séquences de reconnaissance du promoteur de *toc1*, rendant difficile l'identification de la bonne zone régulatrice. Ensuite, les ARNm anti-sens des lignées *CCA-as* ne sont peut-être pas assez nombreuses pour annihiler les ARNm originaux afin de casser la boucle transcriptionnelle. Le modèle mathématique peut aider à résoudre et comprendre ces problèmes. Ainsi, le modèle établi sur une boucle de rétroaction transcriptionnelle négative aura pour but de retrouver les résultats expérimentaux précédents tout en amenant plus de clarté sur la nature de la boucle transcriptionnelle. Un modèle retrouvant les données expérimentales encouragera l'hypothèse la plus simple ou bien apportera de nouvelles voies d'exploration.

3.3 Modèle de l'horloge circadienne

3.3.1 Introduction

Nous présentons ici un modèle de l'horloge circadienne de l'algue *Ostreococcus Tauri* qui nous permettra, par comparaisons avec les données expérimentales, de valider l'hypothèse que les deux gènes *toc1* et *cca1* forment une boucle de régulation transcriptionnelle négative où la protéine *TOC1* active la transcription du gène *cca1* et la protéine *CCA1* réprime l'expression du gène *toc1*. Le modèle suppose une régulation directe du promoteur du gène *toc1* par le facteur de transcription *CCA1*, démontrée par Bouget et al. [87], mais également une régulation directe du promoteur de *cca1* par le facteur de transcription *TOC1*, non établie expérimentalement.

Un algorithme conçu par Jensen et al. servira à montrer que les données expérimentales sont compatibles avec l'existence d'une boucle transcriptionnelle négative incluant *toc1* et *cca1* [65]. Ensuite, nous identifierons un biais présent dans les données expérimentales des lignées de luminescence. Les signaux sont traités pour créer des profils temporels de référence. Nous chercherons à minimiser la différence entre les profils numériques, provenant de l'intégration du modèle mathématique, et les profils expérimentaux, à travers une fonction de score. L'ajustement simultané des données issues de conditions expérimentales différentes (signaux de luminescence et micropuces) fournira une signature de la robustesse de l'horloge aux fluctuations expérimentales.

3.3.2 Équations de la boucle à deux gènes

3.3.2.1 Algorithme d'identification de la boucle de régulation

Le rythme d'une horloge biologique est observé à partir de la mesure des concentrations des protéines (ou de l'expression des gènes) [1]. La dynamique de ces grandeurs est déterminée par des équations différentielles [1, 38]. Un modèle peut être représenté par des équations différentielles ayant différentes caractéristiques (délais temporels [129, 130], effets spatiaux [164]). Chacune de ces caractéristiques peut être déterminante pour obtenir des oscillations. Pourtant, deux propriétés communes sont souvent présentes dans ces différents modèles. D'abord, un facteur de transcription ne change pas de nature régulatrice vis-à-vis d'un promoteur donné. Par exemple, une protéine activatrice ne devient pas un represseur d'un gène à une autre concentration. Ensuite, il faut au moins une rétroaction négative pour que les oscillations puissent apparaître [165].

En reprenant les travaux de Bouget et al., nous remarquons que le réseau génétique de l'horloge circadienne d'*Ostreococcus Tauri* possède ces deux propriétés décrites ci-dessus. La protéine *TOC* active la transcription du gène *cca*, sans qu'on sache si cette régulation est directe ou pas. La protéine *CCA* réprime directement la transcription du gène *toc*, sans que l'on puisse affirmer si sa répression est vitale pour l'horloge à cause de l'expression des lignées *CCA-as* (voir sous-section 3.2.5 page 82). Enfin, les deux gènes forment une boucle transcriptionnelle à rétroaction négative qui entraîne potentiellement des oscillations.

Un algorithme élaboré par Jensen et al. permet d'évaluer qualitativement la cohérence entre les données expérimentales de deux lignées traductionnelles (*TOC :LUC* et *CCA :LUC*) et l'hypothèse de la boucle à deux gènes constituée de *toc* et de *cca* en identifiant la nature régulatrice des protéines [65]. Nous choisissons ensuite de décrire cette boucle par le système d'équations

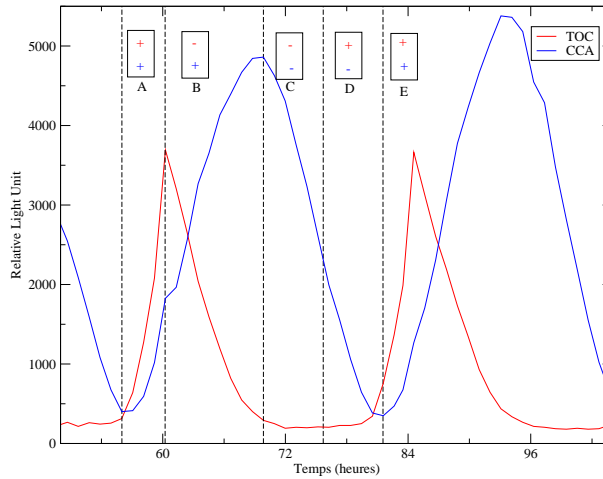


FIGURE 3.8 – Oscillations de l’horloge des signaux *TOC1 :LUC* et *CCA1 :LUC*.

différentielles le plus simple, en espérant que cela décrit fidèlement les données expérimentales. Par la suite, un tel modèle est confirmé ou complété par l’ajout d’acteurs moléculaires, par les prises en compte d’effets stochastiques ou spatio-temporels en fonction de la qualité de l’accord avec les données expérimentales.

L’algorithme permet de déduire à partir de l’alternance des extrêmes de différentes séries temporelles superposées comment les acteurs moléculaires se régulent. Nous appliquons cette méthode pour détecter un motif périodique. Ainsi, nous devons d’abord détecter et classer les minimums et maximums des séries temporelles. Sur la figure 3.8, nous superposons les signaux de luminescence des variables *TOC1* et *CCA1*. Nous découpons le motif périodique en intervalles délimités par l’extremum d’une variable jusqu’au prochain extremum d’une autre variable. Par exemple, le premier intervalle, délimité par deux lignes pleines verticales, débute au premier minimum de *TOC* et se termine au premier maximum de *CCA*. Dans chaque intervalle, les pentes de courbes doivent avoir un signe constant, sinon, cela signifie qu’il manque un acteur moléculaire (ou une variable). Par exemple dans le premier intervalle (figure 3.8), les variables *TOC* et *CCA* croissent de manière monotone. Nous pouvons alors associer à chaque intervalle un ensemble de symboles désignés par “+” ou “-”. L’ensemble contient autant de symboles que de variables. La nature du symbole “+” ou “-” est déterminée selon le signe de la pente de la variable. Dans le premier intervalle de la figure 3.8, les deux variables augmentent. On lui associe donc l’ensemble de symboles (+,+). Le motif périodique est alors converti en séquences discrètes dans l’esprit d’une “dynamique symbolique” [24].

La transition entre deux ensembles de symboles discrets est associé à la modification d’un seul symbole. Cela se traduit par l’activation ou la répression d’un acteur moléculaire par un autre. La signature d’une répression se détecte ainsi : si deux symboles sont les mêmes à un moment donné, un symbole change et l’acteur moléculaire associé à ce symbole est réprimé par l’autre dont le symbole est resté identique. De même, si au départ les deux symboles sont différents, alors cela indique une activation. Sur la figure 3.8, nous en déduisons que le gène *toc* est réprimé par *CCA*, tandis que *cca* est activé par *TOC*. Finalement, si le nombre de répresseur est pair, alors ce n’est pas une boucle à rétroaction négative. La topologie des signaux expérimentaux est cohérente avec un réseau génétique basé sur une boucle transcriptionnelle à rétroaction négative.

Nous n'incorporons dans le modèle que les processus de transcription, de traduction et de dégradation. Comme la période du rythme circadien est grande devant l'échelle de temps liée à l'activité du gène, la régulation s'ajuste instantanément à la modification de la concentration du facteur de transcription. Ainsi, la dynamique de transcription de l'ARNm est représentée par l'équation (1.19, page 22). La dynamique de traduction est décrite par l'équation (1.13, page 21). La dégradation est enzymatique et sa dynamique est décrite par une forme équivalente à l'équation (1.18, page 21). Le système d'équations différentielles représentant notre modèle de boucle à deux gènes est :

$$\frac{dM_T}{dt} = \mu_T + \frac{\lambda_T}{1 + (P_C/P_{C0})^{n_C}} - \frac{\delta_{M_T} K_{M_T} M_T}{K_{M_T} + M_T} \quad (3.1a)$$

$$\frac{dP_T}{dt} = \beta_T M_T - \delta_{P_T} \frac{K_{P_T} P_T}{K_{P_T} + P_T} \quad (3.1b)$$

$$\frac{dM_C}{dt} = \mu_C + \frac{\lambda_C (P_T/P_{T0})^{n_T}}{1 + (P_T/P_{T0})^{n_T}} - \frac{\delta_{M_C} K_{M_C} M_C}{K_{M_C} + M_C} \quad (3.1c)$$

$$\frac{dP_C}{dt} = \beta_C M_C - \delta_{P_C} \frac{K_{P_C} P_C}{K_{P_C} + P_C} \quad (3.1d)$$

Les variables M_T , P_T , M_C et P_C représentent respectivement les quantités de l'ARNm de *tocI*, de la protéine *TOCI*, de l'arn de *ccaI* et de la protéine *CCAI*. Les dégradations sont enzymatiques. δK est le taux de dégradation maximal et K est la constante de saturation de la dégradation. La traduction est proportionnelle à la quantité de l'ARNm où β est le taux de traduction. P_{C0} et P_{T0} sont respectivement les seuils d'activation et de répression de la transcription de *tocI* et de *ccaI*. Les paramètres μ et λ sont les taux basals de transcription.

Les équations (3.1) pourraient être complétées pour décrire des processus biologiques comme la compartimentation (distinction entre le noyau et le cytoplasme), la multiphosphorylation... Si l'ajustement sur les données expérimentales fonctionne avec ce modèle minimal, alors il pourra être complété pour retrouver les différentes propriétés circadiennes comme la robustesse de la période lors la variation des paramètres [166]. L'utilisation de ce modèle pour ajuster les données expérimentales fait l'hypothèse que les signaux de luminescence de *TOC :LUC* et de *CCA :LUC* sont directement assimilables aux concentrations des protéines *TOCI* et *CCAI*.

3.3.2.2 Voie d'entrée de la lumière dans le modèle

Les voies d'entrée de la lumière nécessitent l'intervention de nombreux acteurs moléculaires et de réactions assez complexes de manière à coupler l'horloge au cycle de lumière. Plutôt que de modéliser complètement ces voies biochimiques, quelques paramètres cinétiques sont modulés pour simuler directement l'influence de la lumière sur le réseau génétique par exemple. Des modèles circadiens modulent la constante de dégradation de la protéine *TOC* entre le jour et la nuit [157].

Un exemple de simple fonction de modulation est :

$$\vec{\mu}(t) = \vec{\mu}_{nuit} + L(t)(\vec{\mu}_{jour} - \vec{\mu}_{nuit}) \quad (3.2)$$

Les vecteurs $\vec{\mu}_{jour}$ et $\vec{\mu}_{nuit}$ rassemblent les valeurs de paramètres correspondantes respectivement au jour et à la nuit. La fonction $L(t)$ est la fonction porte définie par :

$$L(t) = 0 \quad \text{la nuit} \quad (3.3)$$

$$L(t) = 1 \quad \text{le jour} \quad (3.4)$$

Ce n'est qu'un exemple de fonction de modulation, une fonction gaussienne centrée au milieu du jour ou de la nuit est également imaginable...

Pour les simulations, il faut choisir quels paramètres moduler et avec quelles amplitudes. C'est une difficulté parfois majeure car de tels systèmes biochimiques contiennent au moins une dizaine de constantes cinétiques. Néanmoins, l'étude des différentes combinaisons de paramètres à moduler se révèle très utile pour expliquer certaines propriétés dynamiques de l'horloge circadienne [94].

3.3.3 Signaux de luminescence de l'activité transcriptionnelle et de l'activité traductionnelle

3.3.3.1 Signaux de luminescence en cycle jour nuit

Les données de luminescence de l'activité transcriptionnelle et de l'activité traductionnelle sont présentées à la figure 3.9. Il n'y aura pas d'ajustement des lignées d'activités transcriptionnelles dans ce chapitre qui seront utilisés au chapitre 4, lorsque le modèle tiendra compte de la cinétique de la luciférase.

Les lignées différentes sont injectées chacune dans un puits d'une plaquette de 96 puits, permettant ainsi d'effectuer un grand nombre de mesures dans les mêmes conditions expérimentales. Les signaux de luminescence utilisés sont montrés à la figure 3.9. Il y a ainsi plusieurs lignées de *TOC :LUC*, *CCA :LUC*, *pTOC :LUC* et *pCCA :LUC*. Deux expériences de luminescence sont faites pour une même lignée afin de vérifier leur reproductibilité expérimentale.

La concentration de cellules injectées est à peu près la même dans chaque puits. Les populations sont soumises à des cycles d'éclairement périodiques (12 heures de jour et 12 heures de nuit) sur plusieurs jours (environ une semaine). La mesure de luminescence est ensuite effectuée par un photomultiplicateur. Une mesure de luminescence est prise une fois toute les heures. La lecture de l'ensemble des puits dure environ 5 minutes. La perturbation lumineuse due à la lecture n'influe pas de manière majeure sur le rythme circadien des cellules. L'ensemble des données expérimentales étudiées est présenté à la figure 3.9. Après un régime transitoire, dont la durée est estimée à 48 heures, les signaux sont périodiques.

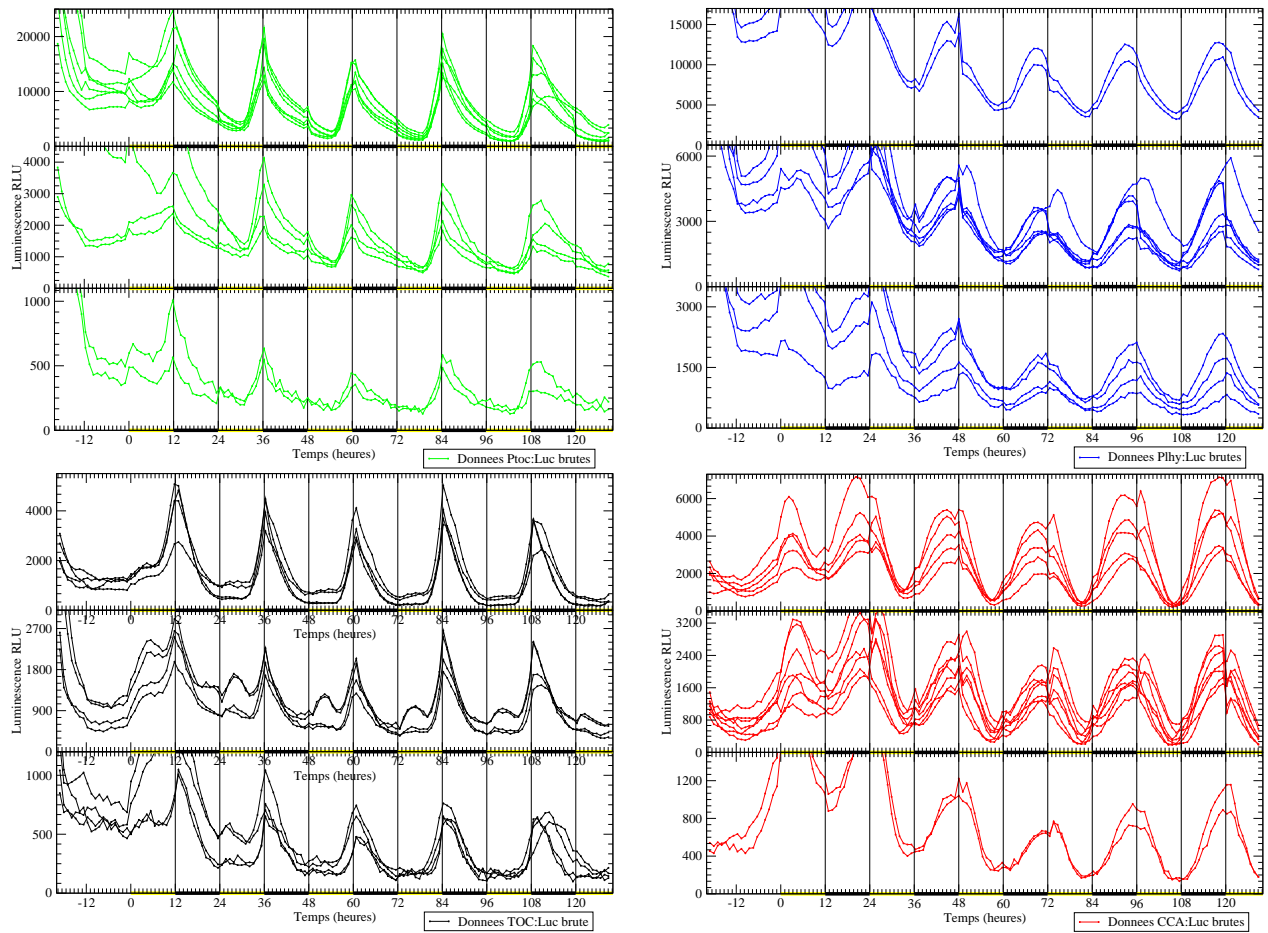


FIGURE 3.9 – Présentation des signaux de luminescence. Signaux de luminescence de l’activité transcriptionnelle des lignées $pTOC :LUC$ (en vert) et $pCCA :LUC$ (en bleu). Signaux de luminescence de l’activité traductionnelle des lignées $TOC :LUC$ (en noir) et $CCA :LUC$ (en rouge). Les courbes sont séparées en trois catégories en raison des amplitudes de l’expression fortement différentes.

3.3.3.2 Signaux de luminescence en jour continu

Les populations *TOCI :LUC* sont ici éclairées en lumière continue. Même si ce signal n'est pas directement utilisé lors de l'ajustement, l'étude de ce signal est pertinente pour montrer que l'amortissement du signal provient d'une désynchronisation des cellules [97, 167]. La période en libre cours des oscillations est calculée pour les lignées *TOCI :LUC*, *CCAI :LUC*, *pTOCI :LUC* et *pCCAI :LUC*. La période de la plupart des signaux en jour continu est proche de 24 heures en moyenne sauf pour les lignées *TOCI :LUC*.

Les signaux de luminescence de *TOCI :LUC* sont repris à la figure 3.10. Les oscillations persistent au moins sur trois cycles avant de s'amortir fortement. Si les cellules communiquent entre elles pour se synchroniser, elles doivent conserver leur rythme même en absence de forçage par la lumière. L'amortissement signifie donc l'absence de couplage entre les cellules.

Nous vérifions que les cellules sont couplées uniquement par la lumière. L'absence de ce couplage lumineux provoque une lente désynchronisation des cellules en raison de variations de la période libre au sein d'une même population. Cette désynchronisation des horloges des différentes cellules est alors responsable de l'amortissement des signaux de luminescence [97, 167].

La méthode la plus fiable pour constater la persistance de l'oscillation en condition d'éclairage continu est d'observer une cellule unique. Mais ce n'est pas encore techniquement possible car une cellule d'*Ostreococcus Tauri* est trop petite et n'émet pas assez de photons. Seul une méthode mathématique peut aider à résoudre ce problème. Pour expliquer l'amortissement des oscillations à cause d'une désynchronisation, nous ajustons les signaux (figure 3.10) par une somme de fonctions périodiques avec des périodes différentes. Un bon ajustement est une condition nécessaire, mais pas suffisante, pour démontrer notre hypothèse.

L'ajustement du profil expérimental de *TOCI :LUC* est fait par la fonction $FA(t)$:

$$FA(t) = P(t) \times [S(t) * G] \quad (3.5)$$

$$FA(t) = \left(\sum_{i=0}^{i=2} a_i t^i \right) \frac{1}{\sigma \sqrt{2\pi}} \int_{T_0 - \Delta T}^{T_0 + \Delta T} e^{-\frac{(T-T_0)^2}{2\sigma^2}} \left(\sum_{k=-\infty}^{+\infty} c_k e^{jk \frac{2\pi}{T} t} \right) dT \quad (3.6)$$

où

$$P(t) = \sum_{i=0}^{i=2} a_i t^i \quad (3.7)$$

Cette fonction est un polynôme de degré 2 où t est la variable du temps. Elle ajuste l'enveloppe du signal. Le degré est limité à 2 car, empiriquement, l'enveloppe lentement variable des signaux suit une parabole.

La fonction $S(t)$ est :

$$S(t) = \sum_{k=-2}^2 c_k e^{jk \frac{2\pi}{T} t} \quad (3.8)$$

C'est une série de Fourier. Elle est utilisée pour ajuster le signal en libre cours. Elle ne dépasse pas deux harmoniques sinon le nombre trop grand d'harmoniques perturbe l'ajustement. Les coefficients de Fourier c_n sont supposés identiques pour chaque cellule.

La distribution gaussienne de la période en libre cours est $G(T)$:

$$G(T) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(T-T_0)^2}{2\sigma^2}} \quad (3.9)$$

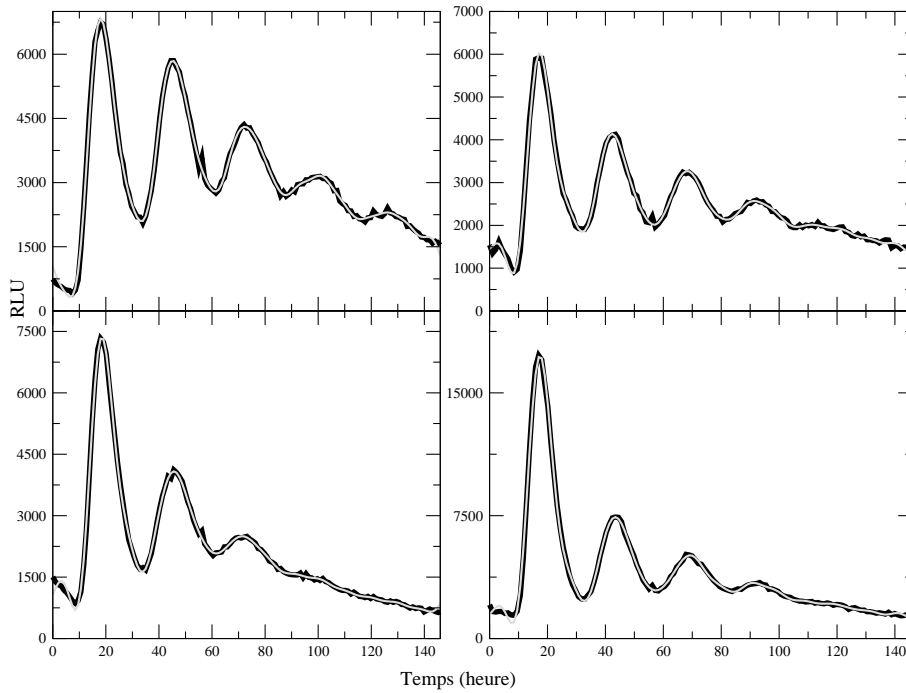


FIGURE 3.10 – Présentation d’expériences de luminescence de 4 lignées *TOC1 :LUC* en condition d’éclairage continu. En noir, signaux brutes. En gris, signaux de luminescence filtrés pour éliminer le bruit.

La distribution est centrée autour de la période moyenne T_0 des horloges de la population. En pratique, nous nous limitons aux périodes comprises dans un intervalle $[T_0 - \Delta_T; T_0 + \Delta_T]$ où Δ_T est déterminé numériquement de manière qu’il soit trois fois plus grand que l’écart-type σ de la distribution gaussienne. Les résultats des ajustements sont présentés à la figure 3.11. Les ajustements ne sont pas parfaits mais montrent que l’hypothèse sur la distribution des cellules est possible. Elles donnent accès aux valeurs de la période moyenne d’une population pour chaque signal de *TOC1 :LUC*. Ces périodes sont proches de celles trouvées par l’utilisation d’une Transformée de Fourier Rapide des signaux.

La moyenne des périodes des lignées *pTOC :LUC*, *pCCA :LUC* et *CCA :LUC* est très proche de 24 heures à l’exception de la moyenne de la période des lignées *TOC1 :LUC*. La période de ces dernières est en moyenne surélevée. Cela est expliqué par une plus grande sensibilité à la surexpression causée par l’insertion dans le génome d’une copie supplémentaire du gène *toc1*.

Chaque cellule peut se comporter comme un oscillateur individuel avec une dispersion des périodes amortissant le signal. L’observation en cellule unique, pouvant confirmer cette hypothèse se révèle malheureusement trop complexe à mettre en œuvre pour le moment.

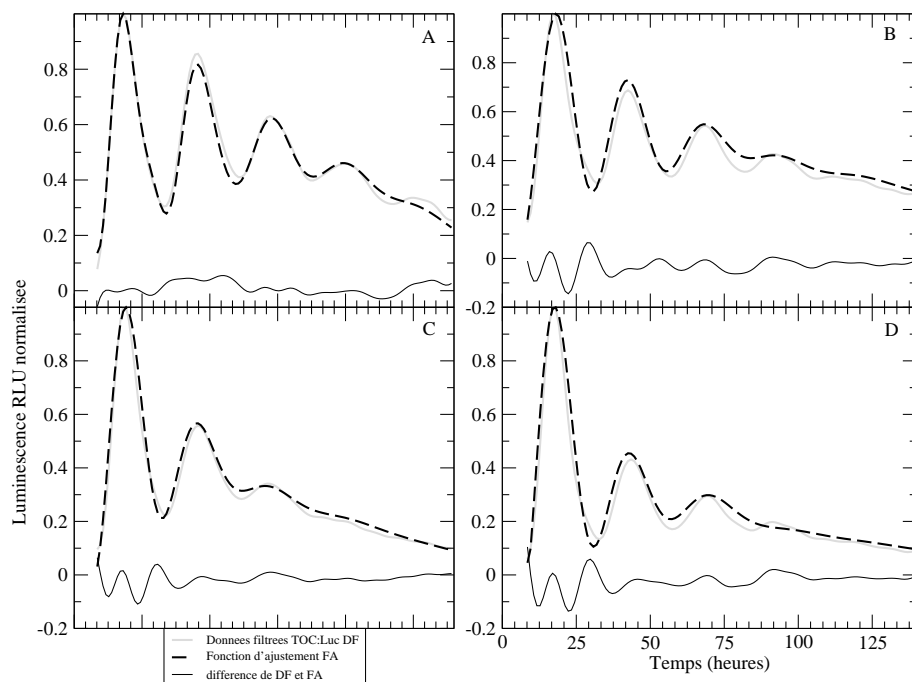


FIGURE 3.11 – Ajustement des profils de luminescence de quatre lignées traductionnelles *TOC1 :LUC* par une série de fourier dont la période suit une distribution gaussienne. La période et l'écart-type trouvé sont 27.4 ± 2.8 h (figure A), 25.4 ± 2.8 h (figure B), 26.21 ± 3.7 h (figure C), 25.87 ± 3.5 h (figure D). Nous les comparons avec la période trouvée par la Transformée de Fourier Rapide [168] des profils expérimentaux *TOC1 :LUC* : 27.27h (fig. A), 25.h (fig. B), 26.85h (fig. C) et 25.63h (fig. D).

3.3.4 Biais des signaux expérimentaux

L'étude des signaux expérimentaux révèle un biais dans les mesures de luminescence. Alors que le maximum de deux signaux provenant de la même lignée dans deux puits différents diffère d'un certain facteur, leur minimum coïncide. Le bruit est évalué dans deux puits vides afin de remarquer sa faible influence sur les signaux de luminescence.

3.3.4.1 Analyse du bruit de fond

Deux expériences de mesures de bruit de puits vides (entourés de puits pleins de cellules) sont montrées à la figure 3.12. Le spectre des signaux affichent clairement la présence d'un signal de période égale environ à 24 heures pour les deux expériences. Ce signal périodique est éliminé pour avoir uniquement les fluctuations dues au bruit. Les mesures de luminescence de chaque puits sont donc légèrement perturbées par les proches puits voisins. L'amplitude du bruit est au voisinage de 100 RLU plus ou moins 50 RLU pour les deux puits vides. L'intensité du minimum de luminescence des signaux au-delà de 1000 RLU est donc très peu perturbée par le bruit des appareils de mesures et l'influence des puits voisins est donc relativement faible.

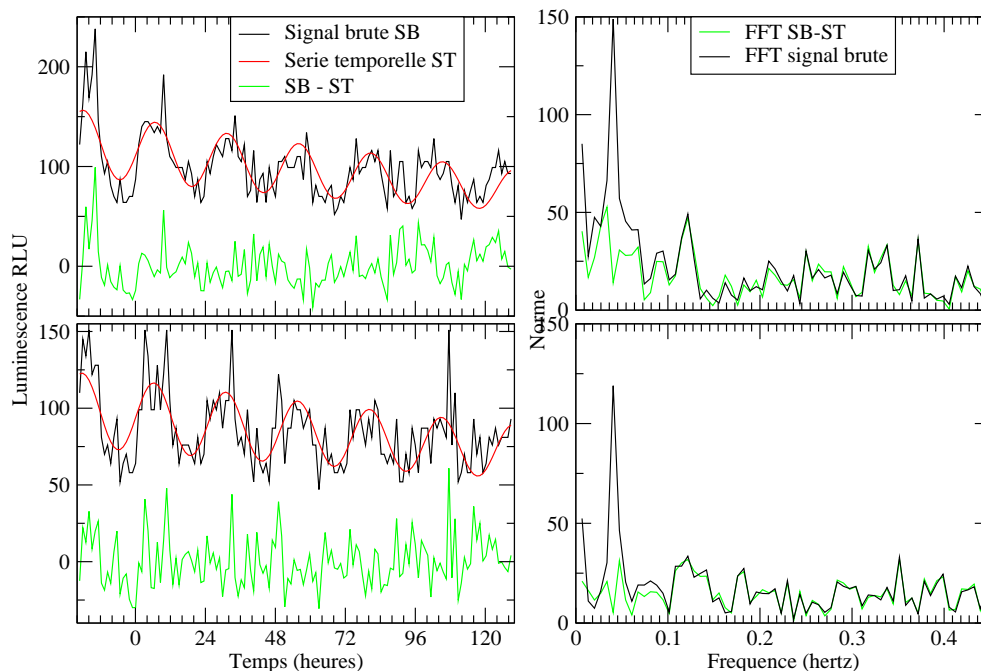


FIGURE 3.12 – A gauche, mesure de la luminescence de puits vides ajustée par une simple série de fourier. Un signal périodique est détecté et est dû aux émissions lumineuses des puits proches voisins, confirmé par la FFT de ces signaux. Ce dernier est retranché des signaux bruts ne laissant que les fluctuations de l'appareil de mesure.

3.3.4.2 Biais du plancher de luminescence

Il y a néanmoins une autre source de bruit que celle détectée précédemment. Pour chaque lignée, deux mesures de luminescence sont effectuées afin de vérifier la reproductibilité expérimentale de chaque lignée. Quelques exemples de signaux de luminescence sont fournis de la figure 3.14 à 3.17. Dans ces figures, chaque encadré contient deux signaux qui correspondent à la luminescence d'une même lignée dans deux puits de mesures différents. Des fluctuations d'amplitudes sont présentes entre ces signaux jumeaux. Cependant, la suppression du minimum de luminescence pour le mettre à zéro et la normalisation par rapport au maximum en supprimant l'enveloppe lentement variable mènent à une bonne superposition des deux signaux jumeaux, comme nous le voyons à la figure 3.13. Les signaux traités de cette manière démontrent la très bonne reproductibilité des expériences.

A priori, une variation du nombre de cellule dans un puits, même faible, suffit à créer deux signaux ne se superposant pas. Mais il doit y avoir dans tous les cas une relation entre les deux signaux jumeaux égale au rapport des concentrations de cellules. Mais cela n'est pas respecté pour les lignées traductionnelles *TOC :LUC* et *CCA :LUC*, révélant la présence d'un biais. En effet, les minimums des lignées jumelles coïncident à un niveau supérieur par rapport au bruit de plancher mesuré par le photomultiplicateur. Pour ces lignées, le minimum de luminescence mesuré correspond en fait à une concentration des protéines luminescentes proche de zéro dans les cellules.

Dans les figures 3.14 et 3.15, la différence entre les deux signaux jumeaux est tracée. Dans ces figures, cette différence est nulle lors des minimums des deux signaux jumeaux. Ce biais n'est pas directement mise en évidence dans les lignées transcriptionnelles. Aux deux figures 3.16 et 3.17 est tracée la différence des signaux jumeaux de l'activité transcriptionnelle. Il est ici plus difficile de conclure à la présence d'un biais.

En conclusion, le minimum de luminescence des lignées traductionnelles est biaisé : le minimum de luminescence, bien que non nul, correspond à une concentration de protéines proche de zéro. Ce biais n'apparaît pas directement dans les lignées transcriptionnelles.

Les simulations numériques concernant les lignées traductionnelles (*TOC :LUC* et *CCA :LUC*) effectuées par le modèle simple de l'horloge circadienne (section 3.3.6) montrent que l'ajustement est meilleur en mettant le minimum de luminescence des profils expérimentaux à zéro. En revanche, les résultats numériques concernant les lignées transcriptionnelles laisse la possibilité à ce que le minimum de luminescence soit dû à une activité basale ou une plus grande durée de vie des protéines de luciférase. Les causes du biais des signaux de fusion traductionnelle ne sont pas identifiées pour l'instant.

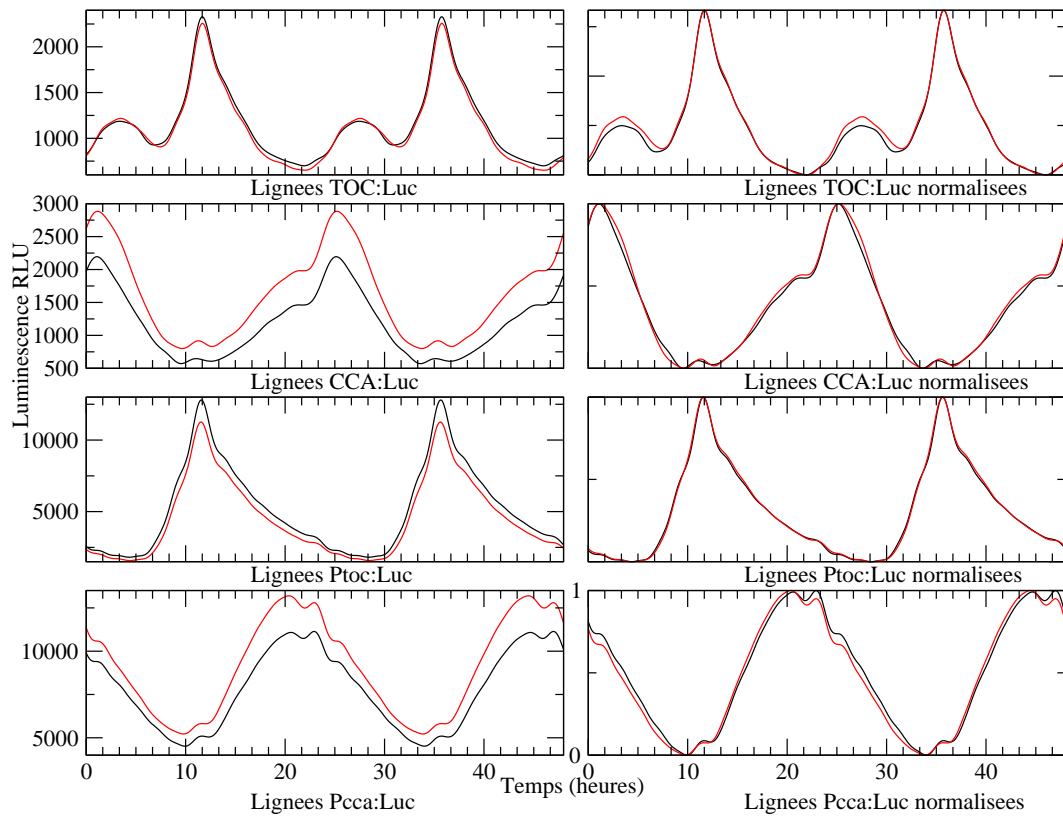


FIGURE 3.13 – A gauche : chaque figure contient les deux signaux bruts d’une population identique présente dans deux puits de mesures différents. A droite : normalisation par rapport au maximum et mise à zéro du minimum du signal de la figure de gauche. La superposition des deux signaux est très bonne. La normalisation a permis de faire abstraction des différences de concentrations cellulaires dans les puits et de relativiser l’importance des fluctuations de mesures liés au photomultiplicateur. La reproduction de certaines “fluctuations” entre les deux signaux signifie que ces variations sont propres à l’expression génétiques de la lignée.

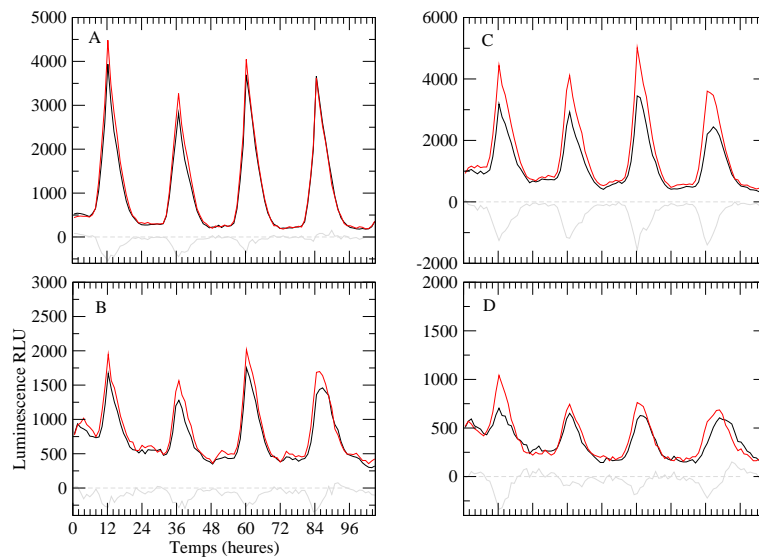


FIGURE 3.14 – Lignées de *TOC :LUC*. Figure A à D, comparaison des signaux lumineux de deux lignées identiques de type *TOC :LUC* mais mesurés dans deux puits différents (traits rouge et noir). En trait gris, différence entre les signaux. La différence des deux signaux montre une grande variation pic à pic alors que leur minimum tend vers zéro. En trait gris discontinu, ligne du zéro.

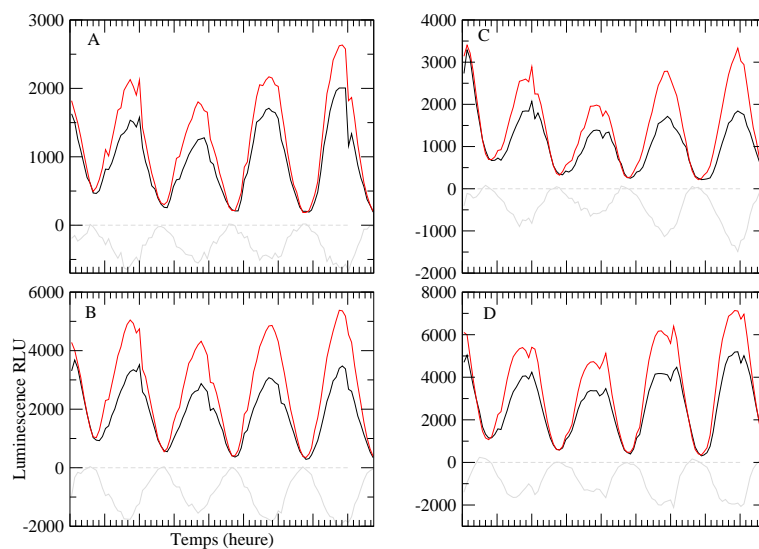


FIGURE 3.15 – De la figure A à D, en traits noirs et rouges, signaux jumeaux de différentes lignées *CCA :LUC*. La différence des signaux (trait gris) montre le minimum à zéro. En trait gris discontinu, ligne du zéro.

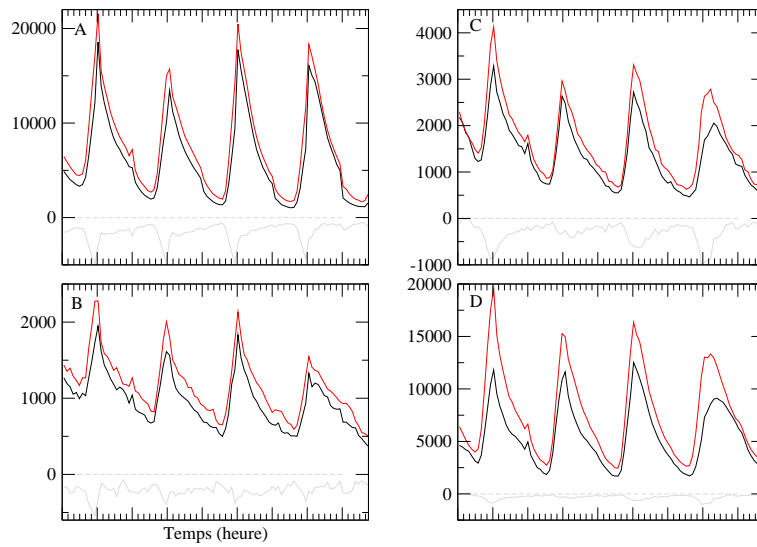


FIGURE 3.16 – Figure A à D : les deux signaux jumeaux (traits noirs et rouges) de chaque lignée *pTOC :LUC* mesurés dans deux puits différents. En trait gris, différence des signaux jumeaux. Le minimum ne descend pas à zéro.

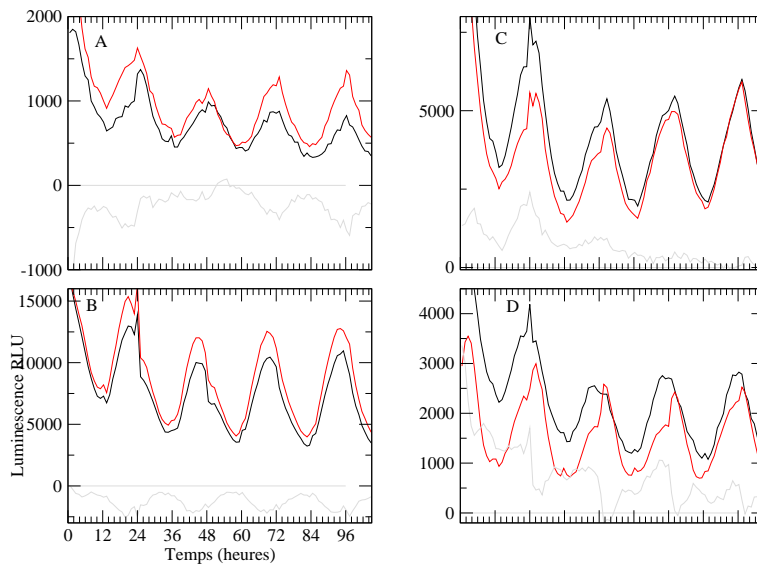


FIGURE 3.17 – Le trait noir et rouge sont les signaux de luminescence des deux signaux jumeaux de chaque lignée de *pCCA :LUC*, de la figure A à D. Leur différence en trait gris continu. En trait gris discontinu, ligne du zéro.

3.3.5 Traitement des signaux expérimentaux des puces à ADN et de luminescence

Pour valider le modèle mathématique de la boucle de rétroaction négative à deux gènes *toc1-cca1*, nous comparons maintenant les profils temporels du modèle et ceux des données expérimentales de l'ARN messager. Pour cela, nous déterminons un profil des concentrations d'ARN à partir de l'interpolation des données de micropuces à ADN. Ensuite, nous caractérisons les profils numériques et expérimentaux par les temps de passage à 20% et à 80% de l'amplitude maximale. L'accord entre la théorie et l'expérience est estimé par la différence sur le temps de passage.

3.3.5.1 Signaux des micropuces à ADN

Les données par micropuces à ADN ont une très bonne reproductibilité expérimentale [105]. L'utilisation des micropuces nécessite le prélèvement suivi du sacrifice d'une fraction de la population d'algues. Mais cette opération ne perturbe pas le fonctionnement de l'horloge des autres algues car elle est inhérente à chaque cellule et n'est pas couplée aux horloges des cellules voisines. Les organismes étudiés ne sont pas modifiés génétiquement, contrairement aux méthodes d'insertion de gène de luminescence dans le génome. Ainsi, les cellules ne subissent pas une modification majeure de leur phénotype.

Les signaux de l'ARN de *TOC1* et de *CCA1* nous sont fournis par les micropuces à ADN [92] (voir figure 3.18). Lors de ces expériences, les cellules sont soumises à un cycle de jour et d'obscurité de 12 heures chacun. L'expérience est réalisée en triplicat pour vérifier la reproductibilité expérimentale. Sur la figure 3.18, les données des micropuces de chaque expérience sont superposées. Celle-ci est plutôt bonne avec une variation très faible des niveaux maximaux.

Le pas d'échantillonnage, d'une mesure toutes les 3 heures est relativement grand et n'est pas optimal pour déterminer le profil. Par exemple, la position des vrais pics d'expression n'est pas précise. La stratégie employée dans la section suivante règle en partie ce problème en déterminant la position du maximum la plus probable par rapport aux données.

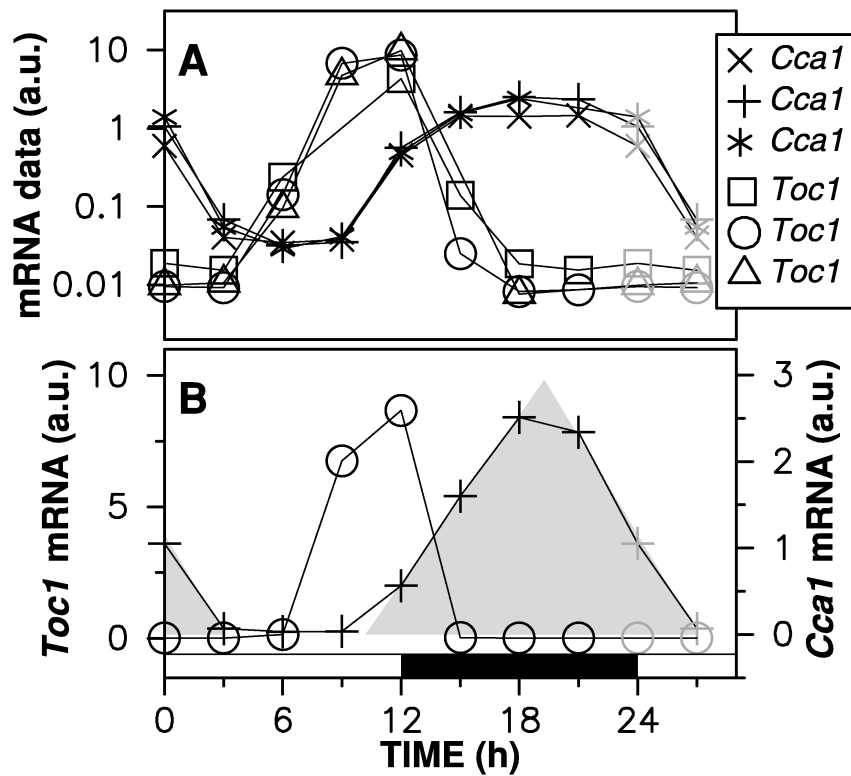


FIGURE 3.18 – (Figures extraites de [94]) A : Données de micropuces à ADN de *TOC1* et de *CCA1* pour des cellules soumises à l’alternance de 12 heures de lumière et d’obscurité. Les données expérimentales de l’ARN de *TOC1* et de *CCA1* sont affichées en une échelle logarithmique. Les points en gris, au temps 24 et 27 heures sont les répliqués des points aux heures 0 et 3. B : représentation en échelle linéaire des signaux de l’ARN de *TOC1* et *CCA1* sélectionnés pour créer un profil cible pour les simulations numériques. .

3.3.5.2 Détermination d'un profil cible de l'ARN à partir des micropuces à ADN

La création du profil à partir des données de micropuces à ADN est la première étape pour la comparaison avec les résultats numériques. Une fois le profil créé, nous déterminons le temps de passage aux valeurs des amplitudes des oscillations à 20% et 80 % de l'amplitude maximale. Pour connaître avec exactitude les temps de passage, il est a priori nécessaire d'avoir une bonne résolution temporelle des données expérimentales, ainsi qu'une durée d'expérience suffisante. Mais la mesure d'un point toutes les trois heures rend difficile une estimation précise des temps de passage. Pour améliorer cela, nous effectuons une interpolation des points de mesures. La figure 3.19 montre le profil issu d'une interpolation par des splines cubiques simples.

L'interpolation par une spline cubique consiste à faire passer par tous les points une fonction polynomiale de degré trois par morceaux de manière à ce que la fonction et ses deux premières dérivées soient continues. L'interpolation est effectuée sur les points expérimentaux en échelle logarithmique. Ensuite, les profils de l'ARN en échelle linéaire sont obtenus en prenant l'exponentielle de 2 de la courbe interpolée à l'échelle logarithmique. Il faut interpoler les données à l'échelle logarithmique car le passage entre deux points des données de *TOCI* à l'échelle linéaire est parfois très raide (entre les points à 6 et 9 heures par exemple sur la figure (b) du schéma de 3.19) et un très grand nombre de points sont proches du niveau zéro. À l'échelle linéaire, la courbe d'interpolation passerait par des valeurs négatives car le degré du polynôme d'interpolation n'est pas assez élevé². Dans ces conditions, la courbe d'interpolation prendrait des valeurs non biologiques.

2. Ce problème apparaît si on interpole par une fonction polynomiale de degré est n , entier positive, en ayant $n + 1$ points de mesures. De fortes instabilités sous forme d'oscillations apparaissent entre certain points de mesures.

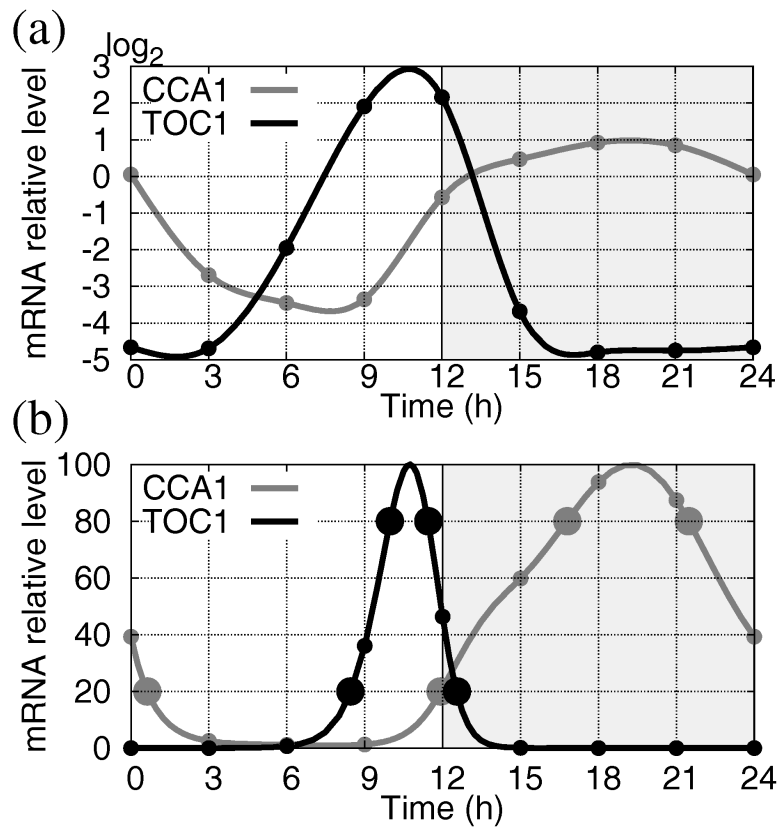


FIGURE 3.19 – D’après [81] (annexe B). Passage d’une cubic spline à travers une série de points correspondant aux résultats des données de puces à ADN. Figure a : les points expérimentaux originaux sont exprimés en base logarithmique de 2. Figure b : les cubic spline sont ensuite mis à la puissance 2 et normalisées à 100. Ces dernières courbes sont ensuite utilisées pour définir une fonction de score.

3.3.5.3 Détermination d'un profil cible à partir des lignées de fusion traductionnelle

Les signaux de fusion traductionnelle sont utilisés pour rapporter les concentrations de protéines au fil du temps. La création d'un profil pour les signaux de luminescence est plus compliquée pour l'ARNm. L'examen des données montre une variation importante des amplitudes de luminescence entre chaque pic (figure 3.9 page 90). Cette variation lente est peut être liée à un effet métabolique ou une irrégularité de la division cellulaire et il faut alors l'éliminer pour construire le profil cible.

Les signaux de luminescence ont des réponses vives aux transitions jour/nuit et nuit/jour pouvant refléter des mécanismes de réinitialisation de l'horloge. Il y a donc une dynamique sur une échelle de temps de 24 heures et une dynamique rapide confinée sur un intervalle de temps plus court. Pour l'instant, nous ne désirons pas tenir compte de ces réponses vives aux transitions de lumière car cela compliquerait le modèle sans apporter d'informations supplémentaires sur le rôle de *TOC1* et *CCA1*. La fonction de score utilisée ignore la présence de ces dynamiques rapides.

L'élévation du minimum (ou plancher) de luminescence dans les lignées traductionnelles est due à un biais expérimental (voir 3.3.4.2). Il est justifiable de l'éliminer du profil cible pour ne pas perturber l'ajustement. Pour confirmer que ce plancher de luminescence biaise bien les résultats, nous créons deux types de profils : un éliminant le plancher et l'autre le conservant. La qualité de l'ajustement devra confirmer ou infirmer la présence du biais. Les signaux de luminescence des fusions traductionnelles ont un transitoire se terminant au bout de 48 heures. Ce transitoire n'apparaît pas dans le profil cible.

Le profil expérimental est estimé grâce à un ajustement des signaux expérimentaux par la fonction $p(t)F(t) + b$ en utilisant une méthode de moindres carrés non linéaire [168]. La fonction $F(t)$ est une série de Fourier décrite par :

$$F(t) = \sum_{k=0}^N a_k \cos(2\pi kt/T + \phi_k) \quad (3.10)$$

où N est le nombre d'harmoniques, a_k les coefficients de Fourier de l'harmonique k , t le temps et ϕ_k la phase. Le paramètre T est fixé à 24 heures. La fonction $p(t)$ est une fonction polynomiale d'ordre 3 ou 4 égale au nombre de pics des signaux de luminescence ajustés. Si nous ajustons par un polynôme de degré plus élevé, l'ajustement ne prendrait pas en compte la série de Fourier. De ce fait, la fonction polynomiale $p(t)$ modélise la variation lente des signaux de luminescence (variations pic à pic). La série de Fourier $F(t)$ représente l'oscillation moyenne des signaux. Le nombre d'harmonique N de la série est fixé à 5, un nombre plus élevé n'améliorant pas l'ajustement. Enfin, le paramètre b ajuste le plancher de luminescence des signaux. La série de Fourier constitue le profil cible de chaque lignée de fusion traductionnelle. Cette stratégie réduit l'amplitude des réponses vives aux transitions de lumières. Le profil final est obtenu en faisant la moyenne des séries de fourier de chaque lignée. Le traitement des signaux de luminescence des fusions traductionnelles de *TOC1* et de *CCA1* est représenté à la figure 3.20.

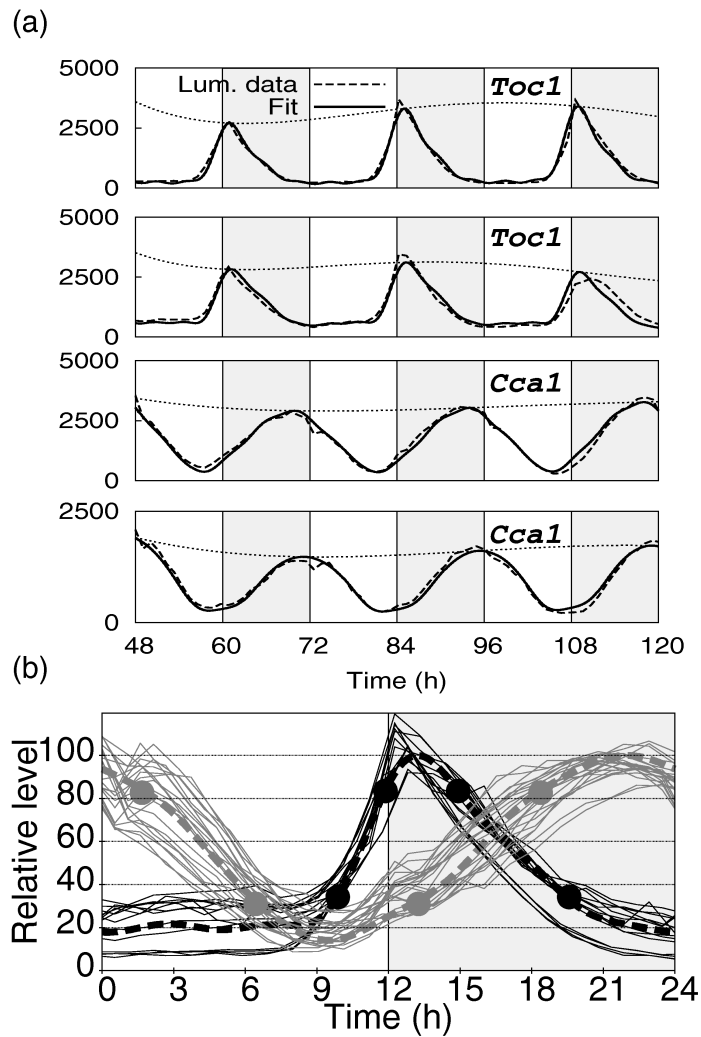


FIGURE 3.20 – (Figures extraites de [81]) Profils des protéines obtenus à partir de l’ajustement par la fonction $F(t)p(t)$ des signaux de luminescence des lignées de fusion transcriptionnelles de *TOC1* et de *CCA1*. La partie blanche (par exemple entre 0 et 12 heures sur figure (b)) correspond à la présence de lumière et la partie grisée (par exemple entre 12 et 24 heures sur figure (b)) correspond à l’obscurité. Figure (a) : quelques exemples de séries temporelles de luminescence des lignées de fusion traductionnelles (en lignes coupées) ajustées par la fonction $F(t)p(t)$ (trait plein). La fonction d’ajustement de période $T = 24h$ a 5 harmoniques et le degré du polynôme est de 4. Nous représentons aussi en ligne pointillées la fonction polynomiale représentant l’enveloppe lentement variable de la série temporelle. Figure (b) : en trait fin (gris pour *CCA1* et noir pour *TOC1*), séries de Fourier ajustant chaque signal de luminescence des fusions transcriptionnelles. En trait discontinu épais, série de Fourier moyenne représentant le profil des signaux de *TOC1* et de *CCA1*. Les gros points représentent les passages au niveau de 20% et de 80% de l’amplitude maximale du profil de *TOC1* et de *CCA1*. Ces points servent de cible pour l’ajustement par le système d’équations décrivant la boucle transcriptionnelle *TOC1-CCA1* 3.1.

3.3.5.4 Quantification de l'accord entre le profil expérimental et le profil numérique

La construction de profils des données expérimentales constitue la première étape pour quantifier la différence avec les profils numériques issus du modèle (3.1). Nous déterminons les temps de passage à 20% et à 80% de l'amplitude maximale des profils expérimentaux et numériques et nous calculons ensuite la somme des différences au carré des temps de passage

La combinaison dans une même fonction de score des temps de passage pour les profils de l'ARNm et des lignées traductionnelles permet de tenir compte de manière harmonieuse de deux types de données bien qu'elles aient des incertitudes de nature différentes en raison des contextes expérimentaux distincts.

Nous limitons le nombre de temps de passage car les résultats numériques montrent que cela suffit à obtenir de bons ajustements tout en évitant de donner un poids trop important aux fluctuations.

Les temps de passage à 20% et 80% de l'amplitude maximale sont des meilleurs mesures que les minimums et maximums du signal. La variabilité de l'expression est intrinsèque au système biologique [161], donc il ne faut pas baser l'ajustement sur les temps de passage très sensibles à cette variabilité. De longs plateaux autour des minimums de *TOC :LUC* rendent aléatoire la détermination du temps de passage. De plus, les temps de passage à 20% et 80% permettent de caractériser la largeur des pics.

Nous définissons maintenant les différentes fonctions de score utilisées pour les ajustements. Ces différentes définitions permettent d'affiner nos conclusions. Les conventions d'écriture dans chaque fonction de score sont les suivantes : $\Delta_{20\uparrow}^X$ (et respectivement $\Delta_{80\uparrow}^X$) est la différence des temps de passage à 20% (respectivement 80%) entre les maximums du profil numérique et expérimental, lorsque la pente des profils est positive. $\Delta_{20\downarrow}^X$ et $\Delta_{80\downarrow}^X$ sont leurs équivalents lorsque les profils sont décroissants. X représente l'espèce M_c , P_c , M_t ou P_t . Pour chaque espèce, la fonction d'erreur quadratique est la suivante :

$$Err(X) = (\Delta_{20\uparrow}^X)^2 + (\Delta_{20\downarrow}^X)^2 + (\Delta_{80\uparrow}^X)^2 + (\Delta_{80\downarrow}^X)^2 \quad (3.11)$$

La première fonction de score est S_{MP} . Elle est la combinaison des erreurs pour les quatre espèces moléculaires (protéines et ARNm) :

$$S_{MP} = \sqrt{1/16 \sum_{X \in M_c, P_c, M_t, P_t} Err(X)} \quad (3.12)$$

Selon les profils, le signal de plancher a été retiré ou pas.

La seconde fonction de score, S_{MPF} , permet de caractériser le fait que deux profils peuvent avoir des temps de passage proches mais différer par un niveau minimum, proche de zéro ou bien à une valeur finie, en raison du plancher de luminescence. L'erreur sur le niveau minimum est définie par la fonction :

$$\Phi_X = \frac{\min(\hat{X})}{\max(\hat{X})} - \frac{\min(X^*)}{\max(X^*)} \quad (3.13)$$

\hat{X} et X^* sont respectivement les profils numérique et expérimental de l'espèce X . La fonction Φ_X est pondérée tel que $\Phi_X = 0.05$ correspond à une erreur de temps de passage de 60 min. La fonction de score est :

$$S_{MPF} = \frac{1}{\sqrt{18}} \sqrt{16(S_{MP})^2 + \sum_{X \in P_c, P_t} (\Phi_X / \Phi_0)^2} \quad (3.14)$$

Le paramètre $\Phi_0 = 0.05/60 = 1/1200$ est l'erreur du niveau minimum de luminescence considéré équivalente à 1 min. Deux autres fonctions de scores considèrent les ajustements séparés des protéines et de l'ARN messenger.

Les deux dernières fonctions de score correspondent aux fonctions de score n'ajustant que les protéines ou les ARNm. La fonction de score ajustant les ARNm s'écrit :

$$S_M = \sqrt{1/8 \sum_{X \in M_c, M_t} Err(X)} \quad (3.15)$$

La fonction de score n'ajustant que les profils des protéines s'écrit :

$$S_P = \sqrt{\frac{1}{10} \sum_{X \in P_c, P_t} (Err(X) + (\Phi_X/(2\Phi_0))^2)} \quad (3.16)$$

Le terme Φ_X est normalisé par rapport au terme $2\Phi_0$ de façon que l'erreur relative est équivalente à la fonction de score (3.14).

La recherche des paramètres est fait par un algorithme de tirage aléatoire. Elle est ensuite suivie par une optimisation grâce à une méthode de moindre carrés non-linéaires (algorithme modifié de levenberg-marquardt, routine d'application de fortran « lmdif.f » [168]). L'incorporation d'une fonction de score dans l'algorithme pour désigner les meilleurs paramètres n'est bien sûr pas nouvelle. Pour l'étude du rythme circadien de la plante *Arabidopsis Thaliana*, Locke et *al.* ont également établi une fonction de score, plus complexe, et calculée à partir de la somme de la moyenne d'erreurs quadratiques [100, 15]. Leur fonction tient compte de l'erreur sur la période libre de leur oscillateur. Cependant, il est préférable de redimensionner le temps, comme cela est fait dans cette thèse. Ainsi, nous pouvons fixer arbitrairement la valeur de cette période, quels que soient les paramètres tirés au sort et amenant aux oscillations. L'exploration de l'espace des paramètres est ainsi plus rapide. Locke et *al.* incorporent l'erreur sur la phase des profils de *CCA* et *TOC* et sur leurs amplitudes. Malheureusement, ils ne traitent pas leurs données expérimentales, or il est statistiquement peu probable qu'un extrêmu d'une donnée expérimentale avec un pas large de mesure soit le véritable extrêmu. Il effectue alors des ajustements très précis sans tenir compte de l'incertitude lié à la méconnaissance de la position et l'amplitude des pics. Notre traitement de données nous permet de tenir compte de la phase et de l'amplitude des profils grâce au temps de passage à 20 et 80 % de l'amplitude. La fonction de score de Locke et *al.* contient également des termes phénoménologiques qui servent à retrouver certaines formes de profils spécifiques à leurs expériences, comme des signatures de la phase transitoire en réponse à des stimuli. Cependant, nous n'avons pas besoin de cela pour notre étude car nous ajustons avec des profils stationnaires et la forme du profil est aussi caractérisée par les temps de passage.

3.3.6 Résultats de l'ajustement sans couplage à la lumière

La figure (3.21) montre les profils numériques obtenus par l'ajustement des profils expérimentaux par le modèle (3.1). Les paramètres correspondants aux simulations sont donnés dans la table (3.22). Les ajustements sont effectués en enlevant le plancher de luminescence (à droite dans la figure 3.21) ou en le laissant (à gauche dans la figure 3.21). Avant de comparer ces deux cas, quelques remarques sont utiles.

Alors que le nombre de temps de passage pour construire les fonctions de score est relativement faible, la forme des profils numériques suit bien la forme des profils expérimentaux pendant toute la journée. Nous aurions pu nous attendre à avoir des profils complètement différents réussissant à passer par les points cibles des profils expérimentaux. Ainsi, le nombre minimal d'interactions chimiques modélisant la boucle à régulation transcriptionnelle *TOCI-CCAI* (régulation transcriptionnelle, dégradation enzymatique, traduction des protéines) est suffisant pour décrire la boucle. De plus, le degré de liberté de l'espace des paramètres est réduit à son minimum car il n'y a pas de modulation des paramètres. Même les ajustements simultanés (fonctions de score S_{MP} et S_{MPF}) des profils de protéines et d'ARN sont très bons alors qu'ils sont les plus contraints.

La qualité des ajustements ne dépend pas que de l'erreur résiduelle mais aussi de la cohérence des paramètres obtenus pour les différentes fonctions de score. Leur véritable signification est parfois plutôt dans leur combinaison que dans le paramètre isolé. Par exemple, lorsque la dégradation est saturée, le terme pertinent est $\delta_X K_X$: chaque terme δ_X ou K_X a parfois des valeurs très fluctuantes tout en donnant des profils numériques très similaires car le produit $\delta_X K_X$ reste à peu près constant. Nous définissons le terme D_X , le taux de dégradation effectif à la valeur moyenne de la dégradation :

$$D_X = \frac{\delta_X K_X}{K_X + \bar{X}} \quad (3.17)$$

Ce terme permet une comparaison des taux de dégradations obtenue par les ajustements dans les différentes conditions.

3.3.6.1 Ajustement des profils biaisés par le plancher de luminescence

Nous discutons ici les parties (a) à (d) de la figure 3.21. Lors de ces ajustements, le choix est de ne pas enlever le plancher de luminescence, malgré le biais démontré en 3.3.4.2. L'ajustement en utilisant cette fonction de score révèle que le profil numérique de *TOCI* n'arrive pas à ajuster le plancher, soutenant l'hypothèse du biais dont l'existence a été démontrée indépendamment du modèle. En revanche, la fonction de score ayant comme erreur la différence des planchers de luminescence des profils numérique et expérimentaux (fonction S_{MPF}) permet un meilleur ajustement des profils *TOCI :LUC*. Cependant, cela introduit une avance de phase du pic de la protéine et un léger niveau de base dans l'ARN de *TOCI*. L'ajustement effectué par la fonction de score S_P ne tient compte que des temps de passage des protéines. Cette ajustement échoue complètement à reproduire des profils de l'ARN.

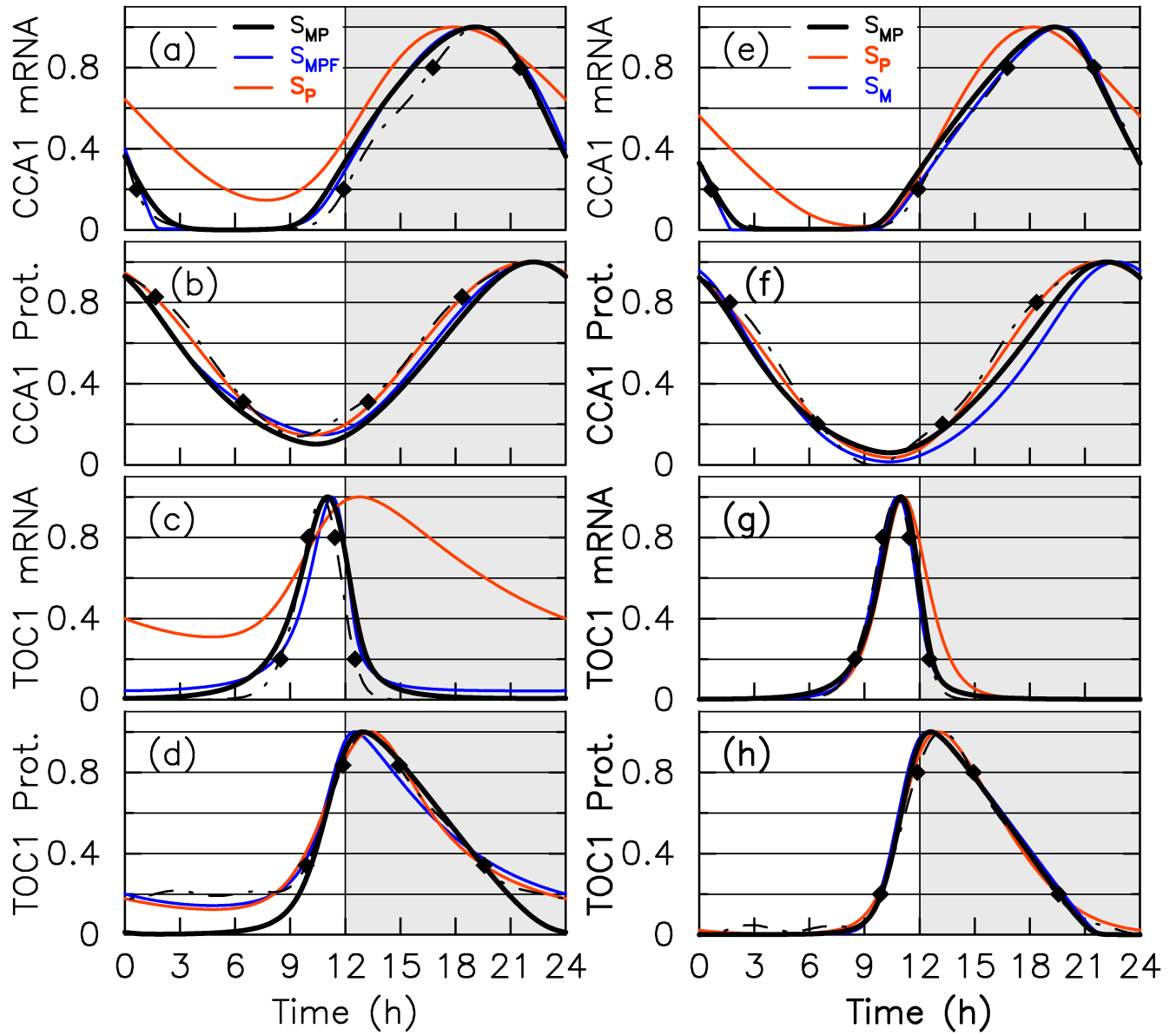


FIGURE 3.21 – Ajustements des profils expérimentaux (lignes noires pointillées) (figures extraites de [81]). Les courbes en couleurs sont les profils numériques générés par l’ajustement des profils expérimentaux par le modèle (3.1). Les figures (a) à (d) sont les ajustement des profils expérimentaux sans supprimer le biais lié au plancher de luminescence. Les fonctions de score relatives à ces ajustements sont S_{MP} (ligne noir), S_{MPF} (ligne bleu), et S_P (ligne orange). Les courbes en couleurs sur les figures (e) à (h) sont les ajustements en supprimant le plancher de luminescence des profils expérimentaux. On utilise les fonctions de score S_{MP} (ligne noire), S_P (ligne orange) et S_M (ligne bleu).

Symbole	S_{MP}	S_{MPF}	S_P	S_M	S_{MP}	S_P
Score (min)	37.7	32.8	12.9	4.2	22.7	11.3
Niveau de base	NR	NR	NR	-	R	R
n_C	2	2	1	2	2	2
n_T	2	2	2	2	2	2
μ_c (nM/h)	1.92×10^{-3}	1.21×10^{-3}	5.18×10^{-3}	2.97×10^{-4}	1.53×10^{-1}	1.46×10^{-1}
λ_c (nM/h)	3.64	6.61	6.56	3.32	3.11	3.31
P_{T_0} (nM)	31.6	27.0	91.6	18.8	18.7	50.0
β_C (h^{-1})	2.6	2.5	3.56	2.89	2.83	3.78
μ_T (nM/h)	0.0117	104	1.69	1.48	0.467	0.0270
λ_t (nM/h)	560	3130	67.0	272	487	233
P_{C_0} (nM)	7.29	5.65	8.74	3.33	4.51	2.73
β_T (h^{-1})	0.667	0.682	6.86	0.811	0.812	0.759
$1/\delta_{M_c}$ (h)	0.542	0.0223	0.831	0.0161	0.195	0.652
$1/\delta_{P_c}$ (h)	3.17	4.78	1.51	1.17	2.36	1.44
$1/\delta_{M_t}$ (h)	0.14	0.0215	4.97	0.151	0.129	0.736
$1/\delta_{P_t}$ (h)	0.559	4.2	0.0659	0.118	0.199	1.65
K_{M_c} (nM)	1.35	0.105	2.24	0.0315	0.407	0.842
K_{P_c} (nM)	133	746	77.5	23.6	75.9	72.2
K_{M_t} (nM)	37.7	12.2	128	60.5	28.3	157
K_{P_t} (nM)	7.77	311	45.7	1.52	2.76	47.6
D_{M_c} (h^{-1})	0.220	0.466	0.220	0.196	0.201	0.119
D_{P_c} (h^{-1})	0.180	0.185	0.290	0.164	0.183	0.292
D_{M_T} (h^{-1})	2.49	6.92	0.13	3.08	2.23	0.939
D_{P_T} (h^{-1})	0.129	0.180	4.76	0.127	0.135	0.196

FIGURE 3.22 – Valeurs des paramètres du modèle pour le score le plus bas des différentes fonctions de score. Nous supposons que la période de l'oscillateur sans forçage lumineux est de 24 heures. Les paramètres sont normalisés par rapport au maximum de 100 nM du profil des protéines et par rapport à la valeur maximale des profils de l'ARN messager (10 nM pour l'ARN de *CCAI* et 70 nM pour l'ARN de *TOCI*). La troisième ligne du tableau indique si le niveau du plancher de luminescence est enlevé noté (R) ou pas (NR). Les quatre dernières lignes sont les taux de dégradations D_X à la valeur moyenne de l'espèce $\bar{X} = (\max(X) - \min(X))/2$ tel que $D_X = \delta_X K_X / (K_X + \bar{X})$.

3.3.6.2 Ajustement des profils sans le plancher de luminescence

Les ajustements sont réalisés par rapport au modèle (3.1) en supprimant dans les profils expérimentaux le plancher de luminescence. Cela concerne les parties (e) à (h) de la figure (3.21). L'ajustement obtenu en utilisant la fonction de score S_{MP} (ajustement simultané des 4 profils cibles) donne un score bien meilleur comparé aux ajustements effectués précédemment et utilisant les scores S_{MP} et S_{MPF} . Seule la montée du profil numérique des protéines de *CCAI* n'est pas assez rapide par rapport au profil expérimental. L'ajustement ne tenant compte que des temps de passage des profils de l'ARN (fonction de score S_M) mène également à une bonne superposition des quatre profils numériques et expérimentaux (ARN et protéine), excepté lors de la montée du profil de protéines de *CCAI*. L'ajustement effectué uniquement sur les temps de passage des profils des protéines reproduit maintenant au mieux les profils de l'ARNm.

On regroupe au tableau 3.22 les valeurs des paramètres obtenues lors des simulations. Ces dernières valeurs sont cohérentes car il y a rarement des variations importantes entre les ajustements obtenus avec les différentes fonctions de score.

3.3.7 Ajustement avec couplage à la lumière

Pour vérifier que les résultats précédents ne dépendent pas de manière critique de la période de l'oscillateur en libre cours fixée à 24 heures, des ajustements sont effectués avec des périodes en libre cours égales à 23.5 et 25 heures. Les ajustements simultanés des ARN et des protéines (fonction de score S_{MP}) sont effectués en supprimant le plancher de luminescence. Un paramètre (P_{C_0} pour l'oscillateur de période à 25 h et δ_{PC} pour l'oscillateur de période 23.8 h) est modulé par une fonction porte afin de synchroniser avec le cycle solaire de 24 heures.

Les résultats des simulations numériques sont montrés à la figure 3.23 et les valeurs de paramètres sont données au tableau 3.24. Ces derniers ajustement sont corrects mais moins bons qu'en absence de couplage à la lumière. Nous en concluons que forcer la présence du couplage détériore l'ajustement car ce dernier ne laisse pas de signature dans les données lorsque l'horloge est synchronisée.

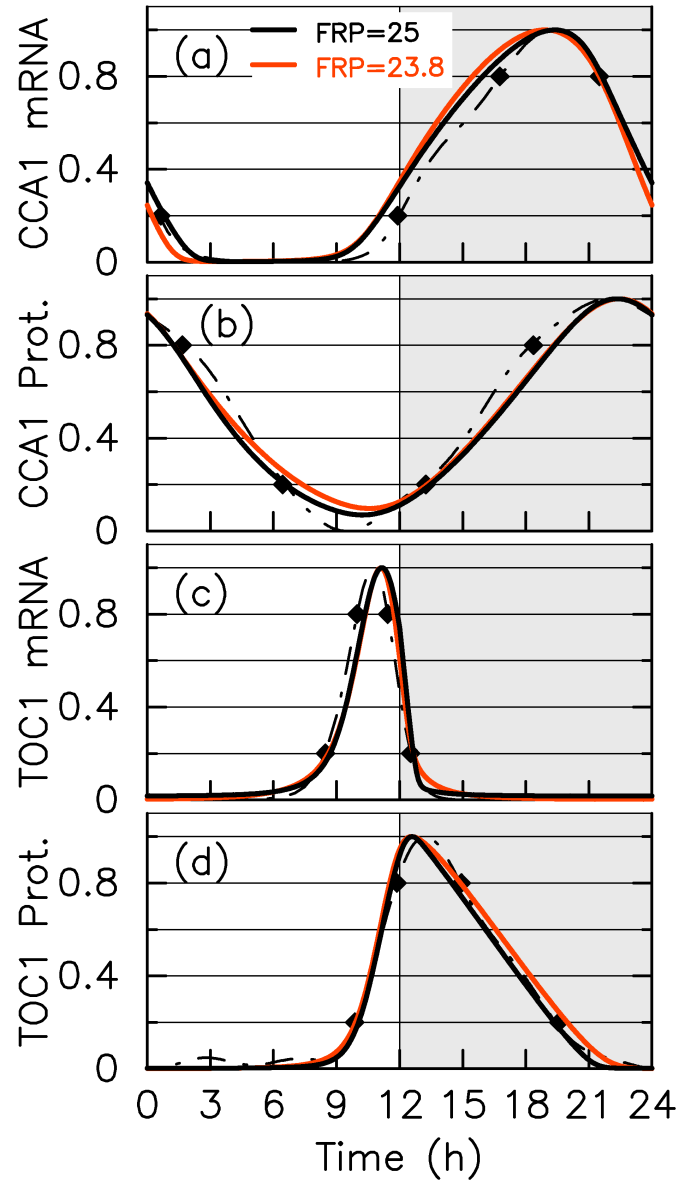


FIGURE 3.23 – Résultats de l’ajustement du modèle dépendant de la lumière tirée de [81]. La partie grisée des figures représente la nuit et la partie blanche représente le jour. Les points sont les temps de passage du profil expérimental (courbe en pointillée). Le plancher luminescence est enlevé. Nous ajustons par la fonction de score S_{MP} . La période de l’oscillateur autonome (en jour continu) est de 25 heures (respectivement 23.8 heures) et le profil y correspondant est représenté- en ligne noire (respectivement orange). La synchronisation du modèle de période libre à 25 heures (respectivement 23.8 h) est obtenue en modulant le paramètre P_{C_0} (respectivement δ_{PC}).

FRP (h)	25	23.8
Score (min)	27.8	34.3 .3
n_C	1	1
n_T	2	2
μ_c (nM/h)	6.06×10^{-4}	4.73×10^{-2}
λ_c (nM/h)	3.82	4.99
P_{T_0} (nM)	15.9	27.3
β_C (h^{-1})	2.71	2.34
μ_T (nM/h)	22.7	6.07×10^{-2}
λ_t (nM/h)	357	1130
P_{C_0} (nM)	5.94	6.15
$P_{C_0}^{night}$ (nM)	11.3	
β_T (h^{-1})	0.805	0.809
$1/\delta_{Mc}$ (h)	0.305	0.296
$1/\delta_{Pc}$ (h)	2.28	2.5
$1/\delta_{Pc}^{night}$ (h)		2.6
$1/\delta_{Mt}$ (h)	4.28×10^{-2}	4.51×10^{-2}
$1/\delta_{Pt}$ (h)	0.224	0.402
K_{Mc} (nM)	0.690	0.879
K_{Pc} (nM)	68.5	58.3
K_{Mt} (nM)	6.96	17.1
K_{Pt} (nM)	3.23	5.62

FIGURE 3.24 – Valeurs des paramètres du modèle 3.1. Ces derniers sont renormalisés comme dans le tableau 3.22.

3.4 Conclusion

L'étude de l'horloge circadienne d'une espèce aboutit souvent à l'identification d'un très grand nombre de gènes, généralement tous essentielles au bon fonctionnement de l'horloge. Ainsi, les nombreux travaux de modélisation mathématique des horloges circadiennes s'avèrent complexes ou manquent parfois de confrontation avec les données expérimentales.

Découvert dans les années 90, l'algue unicellulaire *Ostreococcus Tauri* est un organisme présentant une organisation cellulaire simple et un génome compact. La comparaison de séquences génomiques a permis de mettre en évidence un petit nombre de gènes impliqués dans son horloge circadienne par rapport au modèle de référence, *Arabidopsis Thaliana*. Nous avons alors modélisé le réseau génétique de cette plante à partir des acteurs génétiques identifiés et de données expérimentales.

Le modèle se réduit à deux gènes, *TOC* et *CCA*, se régulant mutuellement pour aboutir à une boucle de rétroaction négative, dans laquelle la protéine *TOC* active le gène *cca*, tandis que *CCA* réprime *toc*. Nous avons procédé à un traitement des données de micropuces à ADN et de données de luminescence des lignées transcriptionnelles et traductionnelles. La présence d'un biais, c'est-à-dire d'un plancher de luminescence, est alors constaté dans les signaux de fusion traductionnelle. Ensuite, l'ajustement est effectué avec une fonction de score simple, mais efficace, ne prenant en compte que la somme des erreurs quadratiques entre des points cibles. Ces points cibles sont les temps des profils lorsque l'amplitude atteint 20 et 80 % de l'amplitude maximale. Ces résultats numériques sont cohérents avec l'identification du biais car la fonction de score donne toujours un meilleur résultat pour des ajustements sans le plancher de luminescence.

Mais la particularité importante de nos résultats est la suivante : les scores sont les meilleurs lorsque l'ajustement des profils expérimentaux est effectué par un oscillateur libre (et dont la période est fixé à 24 heures). Cela signifie que la signature du couplage, lorsque l'horloge est déjà en phase, n'est pas apparente. Rappelons que ce phénomène a été associé, par Thommen et *al.*, à de la robustesse aux fluctuations d'éclairement [94]. D'autres simulations numériques ont été réalisées en présence de couplage, par modulation paramétrique, mais les résultats n'ont pas été améliorés, ceux qui démontrent que la qualité des ajustements ne dépend pas de manière critique de la période de l'oscillateur libre. De plus, les résultats des ajustements concernant deux types de profils de provenance très différentes (micropuces, luminescence) sont très bons. L'horloge est donc robuste à une forme de bruit causée par des conditions expérimentales distinctes. Ainsi, l'organisme *Ostreococcus Tauri* est peut-être l'organisme possédant l'horloge circadienne la plus simple.

Cependant, certaines approximations ont été faites en ceux qui concernent les données traductionnelles. Nous avons supposé que le signal lumineux correspondait exactement au profil des protéines. Comme la luciférase est inactivée quand elle émet un photon, il est en fait nécessaire de déterminer sa durée d'émission pour savoir si la lumière est émise peu de temps ou longtemps après la traduction. De plus, l'insertion d'un gène supplémentaire dans le génome perturbe l'expression, qui pourrait alors être fort différente de celle des cellules sauvages. Dans le chapitre suivant, le modèle de l'horloge circadienne est complété en tenant compte de la cinétique de la luciférase. D'abord, nous déterminons la durée de vie de la luciférase. Ensuite, plusieurs modèles de l'horloge circadienne tenant en compte la cinétique de la luciférase sont établis afin d'ajuster les lignées transcriptionnelles et traductionnelles. Mais quels que soient les résultats, les résultats précédents ne seront pas modifiés car ils sont corroborés par un ajustement.

tement sur les données de micropuces. Ces dernières données correspondent à l'expression des cellules sauvages.

Chapitre 4

Modèles de l'horloge circadienne prenant en compte la dynamique de la luciférase

4.1 Introduction

Les nouveaux modèles que nous présentons ici visent à confirmer que les signaux traductionnelles *TOC :LUC* et *CCA :LUC* fournissent une information pertinente sur les concentrations de protéines *TOC* et *CCA*. En particulier, nous essayons de mesurer l'altération du signal *TOC* et *CCA* causée par la surexpression due aux copies de gènes supplémentaires. Dans le cas idéal, nous pourrions reconstruire les profils *TOC* et *CCA* à partir des paramètres retrouvés par ces ajustements.

Nous discutons d'abord les équations décrivant la dynamique des gènes *toc*, *cca* et rapporteurs dans les lignées traductionnelles et transcriptionnelles. L'établissement du modèle des lignées traductionnelles nécessite de prendre en compte la compétition entre les différentes formes de la protéine pour la dégradation et la régulation transcriptionnelle.

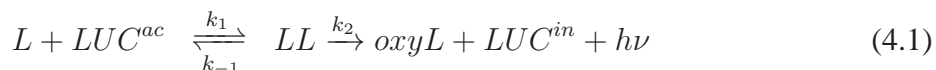
La création des profils expérimentaux diffère légèrement de celle utilisée dans le chapitre 3 et est basée sur la Transformée de Fourier Rapide. Une fonction de score permettant de juger la ressemblance entre les profils numériques et expérimentaux est défini. Nous procédons ensuite aux ajustements par des modèles tenant compte de la luciférase.

Nous déterminons la durée d'émission de lumière par la luciférase afin de démontrer que le signal est plus proche de la concentration en protéine qu'en ARNm. Ensuite, l'ajustement par les modèles décrivant les lignées transcriptionnelles fournit d'autres paramètres cinétiques liés à la luciférase. Ces simulations fournissent des profils des protéines originales *TOC :LUC* et *CCA :LUC*.

4.2 Nouveaux modèles de l'horloge circadienne

4.2.1 Cinétique de la luciférase

Le schéma réactionnel menant à l'émission d'un photon par la luciférine, catalysée par la luciférase, est la suivante [113] :



où L représente la luciférine. La luciférase existe sous deux formes : la forme active LUC^{ac} avant l'émission du photon et la forme inactive LUC^{in} après émission du photon. Le terme $h\nu$ représente le photon. Le complexe LL est le complexe intermédiaire $LUC-luciferin-AMP$, $oxyL$ symbolise l'oxyluciférine et les paramètres k_1 , k_{-1} et k_2 sont des constantes cinétiques. Le dioxygène O_2 , l'Adénine-Tri-Phosphate (ATP), l'Adénine Mono-Phosphate (AMP) et le pyrophosphate (PP) interviennent dans la réaction mais ne sont pas limitants et ne sont donc pas pris en compte. La luciférine, notée L_0 , est injectée en excès au début de l'expérience de manière que ce substrat soit maintenu en quantité quasi-constante. La formation du complexe LL étant très rapide, sa dynamique est en état stationnaire. Les équations donnant le flux de lumière ϕ sont :

$$\frac{dLL(t)}{dt} = k_1 L_0 LUC^{ac}(t) - (k_{-1} + k_2) LL(t) \approx 0 \quad (4.2)$$

$$\phi(t) = k_2 LL(t) \approx \frac{L_0 k_2 k_1}{k_{-1} + k_2} LUC^{ac}(t) \quad (4.3)$$

L'intensité I des signaux expérimentaux correspond à l'intégration de $\phi(t)$ par le photomultiplicateur pendant le temps de mesure. La mesure étant très rapide, la luciférase active est considérée en quantité constante (l'ensemble des 96 puits est lu en 5 minutes, soit moins de 3 secondes par puit). L'intensité est donc proportionnelle à la quantité de luciférase active au moment de la mesure :

$$I(t) \propto LUC^{ac}(t) \quad (4.4)$$

Ainsi, la concentration de luciférase active nous donne directement l'intensité lumineuse.

4.2.2 Modèle des lignées de fusion donnant l'activité promoteur

Nous présentons les deux systèmes d'équations différentielles décrivant les deux types de lignées transcriptionnelles. Dans ces lignées coexistent le gène original et le gène rapporteur. Le rapporteur inséré dans le génome oblige à modifier le modèle en ajoutant deux variables : l'ARNm de la luciférase contrôlé par le promoteur (toc ou cca) et la protéine de la luciférase. Le signal de lumière émis par la luciférase est uniquement une sortie sans influence sur l'horloge. Les équations présentées sont donc en fait très semblables à celles du modèle de l'horloge (3.1, page 88) avec deux équations dynamiques supplémentaires. Nous n'écrivons donc que les équations donnant la quantité d'ARNm et de protéine de la luciférase s'ajoutant au système (3.1). Pour $TOC :LUC$, les deux équations supplémentaires sont :

$$\frac{dMl}{dt} = \mu_{tl} + \frac{\lambda_{tl}}{(Pc/Pcl_0)^2} - \delta_{ml} \frac{\kappa_{ml} Ml}{\kappa_{ml} + Ml} \quad (4.5)$$

$$\frac{dL}{dt} = \beta_{ml} Ml - \delta_l L \quad (4.6)$$

tandis que pour les lignées *pCCA :LUC* :

$$\frac{dMl}{dt} = \mu_{cl} + \lambda_{cl} \frac{(Pt/Ptl_0)^2}{1 + (Pt/Ptl_0)^2} - \delta_{ml} \frac{\kappa_{ml} Ml}{\kappa_{ml} + Ml} \quad (4.7)$$

$$\frac{dL}{dt} = \beta_{ml} Ml - \delta_l L \quad (4.8)$$

Les variables Ml et L sont respectivement les concentrations en ARNm de la luciférase et en luciférase active. Les paramètres μ_* et λ_* sont les taux basals de transcription et dépendent de la nature du promoteur, mais également de la position du gène dans l'ADN. Le paramètre δ_{ml} est le taux de dégradation. Les paramètres κ_{ml} sont les constantes de saturation. Le paramètre β_{ml} est le taux de traduction. Le paramètre δ_l est le taux d'émission lumineuse.

4.2.3 Modèles décrivant des lignées traductionnelles

Les lignées *TOC :LUC* contiennent trois formes de protéines ayant le même rôle de facteur de transcription : *TOC*, *TOC :LUC* active et *TOC :LUC* inactive (idem pour les lignées *CCA :LUC*). Les cinétiques de transcription et de dégradations sont modifiées à cause d'une compétition entre les facteurs de transcription pour la régulation du gène et pour l'interaction avec les protéases. Les nouvelles expressions sont établies ci-dessous.

4.2.3.1 Compétition entre facteurs de transcription

Pour les lignées *TOC :LUC*, les trois facteurs de transcription sont indiscernables et nous supposons que leur interaction avec le promoteur n'est pas modifiée. C'est donc la somme totale de protéines qui agit en tant que régulateur. La nouvelle cinétique de transcription de synthèse de l'ARNm M_c s'écrit donc :

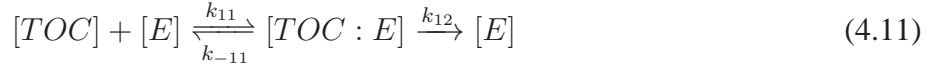
$$\frac{dM_c}{dt} = \mu_c + \lambda_c \frac{(Pt + Ptl + Ptl_i)^2}{1 + 1/Pt_0^2 (Pt + Ptl + Ptl_0)^2} \quad (4.9)$$

De même, la cinétique de transcription de l'ARN de M_t est :

$$\frac{dM_t}{dt} = \mu_t + \frac{\lambda_t}{1 + 1/Pc_0^2 (Pc + Pcl + Pcli)^2} \quad (4.10)$$

4.2.3.2 Compétition entre protéines pour l'enzyme de dégradation

Les processus de dégradation se produisent grâce aux enzymes de dégradation, les protéases. Les fonctions de dégradation dans le modèle de la boucle à deux gènes étaient écrites en considérant l'existence d'un seul type de protéine dégradé par l'enzyme. Dans le modèle décrivant la cinétique de la luciférase, plusieurs types de protéines sont dégradés par la même enzyme. Tout comme pour les sites de régulation des gènes *toc* et *cca*, il y a une compétition entre les protéines pour les sites des enzymes de dégradation. En prenant la lignée *TOC :Luc* comme exemple, le schéma réactionnel est le suivant :



De fait de la présence de la luciférase, l'affinité réactionnelle avec l'enzyme varie. C'est pourquoi les constantes cinétiques de la réaction (4.11) sont différentes de celles des réactions (4.12) et (4.13). La fonction de dégradation originelle s'écrivait :

$$\frac{dP_T}{dt} = -\delta_{P_T} \frac{K_{P_T} P_T}{K_{P_T} + P_T} \quad (4.14)$$

où P_T est la protéine *TOC*, δ_{P_T} le taux de dégradation et K_{P_T} la constante de saturation. Les nouvelles fonctions de dégradation sont :

$$\frac{dPt}{dt} = -\delta_{pt} \frac{Pt}{1 + \frac{Pt}{\kappa_{pt}} + \frac{Ptl}{\kappa_{ptl}} + \frac{Ptli}{\kappa_{Ptli}}} \quad (4.15)$$

$$\frac{dPtl}{dt} = -\delta_{ptl} \frac{Ptl}{1 + \frac{Pt}{\kappa_{pt}} + \frac{Ptl}{\kappa_{ptl}} + \frac{Ptli}{\kappa_{Ptli}}} \quad (4.16)$$

$$\frac{dPtli}{dt} = -\delta_{ptli} \frac{Ptli}{1 + \frac{Pt}{\kappa_{pt}} + \frac{Ptl}{\kappa_{ptl}} + \frac{Ptli}{\kappa_{Ptli}}} \quad (4.17)$$

Les variables Pt , Ptl et $Ptli$ sont respectivement la concentration des protéines *TOC*, *TOC : Luc* actives et *TOC : Luc* inactives. De même, les paramètres δ_* sont les taux de dégradation lors de la saturation. Les paramètres κ_* sont les constantes de saturation. Les expressions des taux de dégradation sont les mêmes pour l'ARN messager. L'annexe C.2 montre comment obtenir ces fonctions de dégradation.

4.2.3.3 Modèles de l'activité traductionnelle

La compétition pour les sites d'activation du gène *cca* et pour les sites de dégradation enzymatique (voir sections 4.2.3.1 et 4.2.3.2) conduisent aux équations du modèle suivant décrivant la cinétique de *TOC :LUC* :

$$\frac{dMt}{dt} = \mu_t + \lambda_t \frac{1}{1 + \left(\frac{Pc}{Pc_0}\right)^2} - \delta_{mt} \frac{Mt}{1 + \frac{Mt}{\kappa_{mt}} + \frac{Mtl}{\kappa_{mtl}}} \quad (4.18)$$

$$\frac{dMtl}{dt} = \mu_{tl} + \lambda_{tl} \frac{1}{1 + \left(\frac{Pc}{Pc_0}\right)^2} - \delta_{mt} \frac{Mtl}{1 + \frac{Mt}{\kappa_{mt}} + \frac{Mtl}{\kappa_{mtl}}} \quad (4.19)$$

$$\frac{dPt}{dt} = \beta_{mt}Mt - \delta_{pt} \frac{Pt}{1 + \frac{Pt}{\kappa_{pt}} + \frac{Ptl}{\kappa_{ptl}} + \frac{Ptli}{\kappa_{ptli}}} \quad (4.20)$$

$$\frac{dPtl}{dt} = \beta_{mtl}Mtl - \delta_{ptl} \frac{Ptl}{1 + \frac{Pt}{\kappa_{pt}} + \frac{Ptl}{\kappa_{ptl}} + \frac{Ptli}{\kappa_{ptli}}} - \delta_l Ptl \quad (4.21)$$

$$\frac{dPtli}{dt} = \delta_l Ptl - \delta_{ptli} \frac{Ptli}{1 + \frac{Pt}{\kappa_{pt}} + \frac{Ptl}{\kappa_{ptl}} + \frac{Ptli}{\kappa_{ptli}}} \quad (4.22)$$

$$\frac{dMc}{dt} = \mu_c + \lambda_c \frac{(Pt + Ptl + Ptli)^2}{1 + 1/Pt_0^2 (Pt + Ptl + Ptli_0)^2} - \delta_{mc} \frac{\kappa_{mc}Mc}{\kappa_{mc} + Mc} \quad (4.23)$$

$$\frac{dPc}{dt} = \beta_{mc}Mc - \delta_{pc} \kappa_{pc} \frac{Pc}{\kappa_{pc} + Pc} \quad (4.24)$$

Les variables *Mt*, *Mtl* représentent les concentrations de l'ARN messager de la protéine *TOC* et de la protéine *TOC : Luc*. *Pt*, *Ptl* et *Ptli* sont les concentrations de la protéine *TOC* et des protéines *TOC : Luc* active et inactive. Les variables *Mc* et *Pc* sont les concentrations de l'ARN messager et de la protéine *CCA*. Le paramètre δ_l est le taux d'émission lumineuse de la luciférase.

Dans le modèle décrivant la lignée *CCA :LUC*, les différentes variantes de la protéine *CCA* sont *CCA*, *CCA :LUC* active et *CCA :LUC* inactive. Les équations sont :

$$\frac{dMc(t)}{dt} = \mu_c + \lambda_c \frac{\left(\frac{Pt}{Pt_0}\right)^2}{1 + \left(\frac{Pt}{Pt_0}\right)^2} - \delta_{mc} \frac{Mc}{1 + \frac{Mc}{\kappa_{mc}} + \frac{Mcl}{\kappa_{pcl}}} \quad (4.25)$$

$$\frac{dMcl(t)}{dt} = \mu_{cl} + \lambda_{cl} \frac{\left(\frac{Pt}{Pt_0}\right)^2}{1 + \left(\frac{Pt}{Pt_0}\right)^2} - \delta_{mcl} \frac{Mcl}{1 + \frac{Mc}{\kappa_{mc}} + \frac{Mcl}{\kappa_{pcl}}} \quad (4.26)$$

$$\frac{dPc(t)}{dt} = \beta_{mc}Mc - \delta_{pc} \frac{Pc}{1 + \frac{Pc}{\kappa_{pc}} + \frac{Pcl}{\kappa_{pcl}} + \frac{Pcli}{\kappa_{pcli}}} \quad (4.27)$$

$$\frac{dPcl(t)}{dt} = \beta_{mcl}Mcl - \delta_{pcl} \frac{Pcl}{1 + \frac{Pc}{\kappa_{pc}} + \frac{Pcl}{\kappa_{pcl}} + \frac{Pcli}{\kappa_{pcli}}} - \delta_l Pcl \quad (4.28)$$

$$\frac{dPcli(t)}{dt} = \delta_l Pcl - \delta_{pcli} \frac{Pcli}{1 + \frac{Pc}{\kappa_{pc}} + \frac{Pcl}{\kappa_{pcl}} + \frac{Pcli}{\kappa_{pcli}}} \quad (4.29)$$

$$\frac{dMt}{dt} = \mu_t + \frac{\lambda_t}{1 + 1/Pc_0^2 (Pc + Pcl + Pcli)^2} - \delta_{mt} \frac{\kappa_{mt}Mt}{\kappa_{mt} + Mt} \quad (4.30)$$

Les variables Mcl , Pcl et $Pcli$ représentent les concentrations respectivement de l'ARNm de la protéine $CCA : Luc$ et des protéines $CCA : Luc$ active et inactive.

4.3 Stratégie d'ajustement des données par le modèle

4.3.1 Traitement des données par Transformée de Fourier Rapide

Dans la section 3.3.5.3, les profils expérimentaux ont été créés à partir de l'ajustement des données expérimentales par une série de Fourier multipliée par un polynôme lentement variable. L'ajustement était obtenu par une méthode de moindres carrés non-linéaire en imposant l'harmonique de base ($\omega = \frac{2\pi}{24}$) [168]. Cet ajustement permettait d'obtenir les coefficients de Fourier des données expérimentales utilisés ensuite pour engendrer les profils expérimentaux.

Nous utilisons ici la Transformée de Fourier Rapide des données expérimentales (ou FFT) [168] afin d'éliminer l'enveloppe lentement variable des données. La FFT est un algorithme rapide de calcul de transformée de Fourier discrète de signaux. La FFT décompose le signal dans le domaine fréquentiel en calculant pour chaque fréquence identifiée son coefficient de Fourier complexe. Les basses fréquences inférieures à $1/24 h^{-1}$ correspondent à la modulation lente des signaux et sont éliminées.

Les coefficients de Fourier des fréquences multiples exactes de $1/24$ sont également calculés. Ainsi, les coefficients obtenus permettent de tracer la série de Fourier $S_e(t)$ du signal :

$$S_e(t) = \sum_{n=-N}^N c_{e_n} e^{j\omega_n t} \quad (4.31)$$

L'entier N est le nombre d'harmoniques. Le terme c_{e_n} est le coefficient complexe lié à la pulsation harmonique $\omega_n = \frac{2n\pi}{24}$.

À la figure 4.1, quelques exemples de profils expérimentaux construits de cette manière sont montrés.

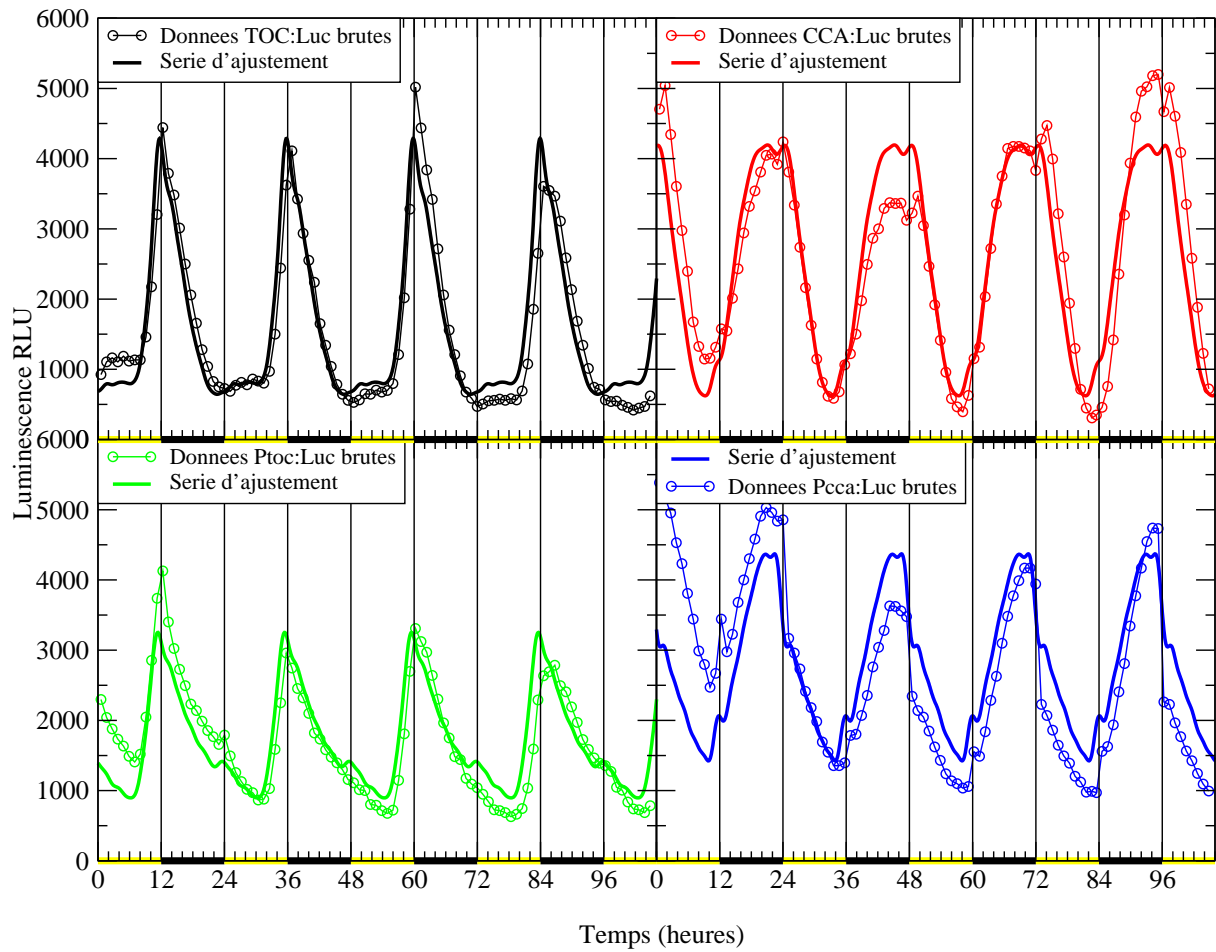


FIGURE 4.1 – Signaux expérimentaux brutes (lignes avec les ronds) superposés à la série de fourier reconstruite (traits plein) à partir des coefficients de fourier calculés par la FFT.

4.3.2 Fonction de score

Afin de quantifier l'ajustement des signaux, une fonction de score F_{sco} est définie comme à la section 3.3.5. Celle-ci est simplement la distance quadratique entre le profil expérimental $S_e(t)$ et le profil numérique $S_n(t)$ échantillonnés à intervalle réguliers :

$$F_{sco} = \sqrt{\sum_{i=0}^N (S_e(t_i) - S_n(t_i))^2} \quad (4.32)$$

S'il y a un ajustement sur plusieurs profils, alors le score final sera la somme du score de chaque profil.

Le choix des conditions initiales lors de l'intégration des équations non couplées à la lumière est déterminant pour avoir la bonne phase par rapport au profil expérimental. Comme le score (voir 4.3.2) dépend de la phase, c'est-à-dire des conditions initiales du cycle limite, de bon paramètres d'ajustement peuvent avoir un mauvais score à cause du décalage de la courbe. Il faut donc déterminer les conditions initiales donnant le meilleur score entre les profils expérimental et numérique. Nous procédons numériquement en déplaçant le profil théorique le long du profil expérimental. Le déplacement peut être optimisé en employant la méthode des moindres carrés non-linéaire [168] et fournit les conditions initiales minimisant le score.

4.3.3 Méthodes d'ajustement des courbes

Le principe est décrit dans la section 3. En utilisant les modèles établis dans la partie précédente, nous ajustons les profils expérimentaux avec les profils numériques en recherchant les paramètres qui minimisent la fonction de score. Cependant, au lieu de créer un seul profil expérimental résultant de la moyenne des séries de Fourier obtenus pour chaque lignée, nous ajustons les profils des lignes séparément. La recherche des paramètres lors de l'ajustement par le modèle est faite grâce à une méthode de moindres carrés non linéaire.

4.4 Influence des paramètres cinétiques sur l'interprétation des expériences

Dans le chapitre 3, les profils de luminescence $TOC :LUC$ (resp. $CCA :LUC$) sont directement assimilés aux concentrations de TOC (resp. CCA). Mais cette hypothèse n'est valable que si plusieurs conditions sur les paramètres sont respectées, comme nous allons le voir dans la partie suivante.

4.4.1 Variation de la durée d'émission de lumière par la luciférase

Nous prenons comme exemple la lignée traductionnelle $TOC :LUC$. D'après la section 4.2.3, les équations dynamiques gouvernant les concentrations des protéines s'écrivent :

$$\frac{dPt}{dt} = \beta Mt - \delta_p \kappa_p \frac{Pt}{\kappa_p + Pt + Ptl + Ptl_i} \quad (4.33)$$

$$\frac{dPtl}{dt} = \beta_{ptl} Mtl - \delta_{ptl} \kappa_{pt} \frac{Ptl}{\kappa_{pt} + Pt + Ptl + Ptl_i} - \delta_l Ptl \quad (4.34)$$

Les variables Mt , Mtl , Pt et Ptl correspondent respectivement aux concentrations de l'ARNm de TOC , de l'ARNm $TOC :LUC$, de la protéine TOC et de $TOC :LUC$ active.

Les paramètres δ_p , et δ_{ptl} sont les taux de dégradation de la protéine TOC et de la protéine $TOC :LUC$ active. Le paramètre κ_p est la constante de saturation. Les paramètres β et β_{ptl} sont respectivement les taux de traduction de l'ARN en protéine TOC et en protéine $TOC :LUC$.

Le paramètre δ_l correspond à l'inverse de la demi-vie d'émission du photon. La valeur du paramètre δ_l influence l'interprétation du signal de luminescence. Étudions en effet le comportement asymptotique des équations 4.33 lorsque δ_l est très grand devant les autres paramètres (application de la théorie de perturbation de Poincaré [169, 98] aux équations 4.33). Nous trouvons alors que la dynamique de la quantité Ptl à l'ordre principal, dans laquelle la dégradation n'intervient pas, est :

$$Ptl = \frac{\beta_{ptl}}{\delta_l} Mtl \quad (4.35)$$

L'égalité (4.35) indique que la quantité de protéine Ptl est proportionnelle à la quantité d'ARNm de $TOC :LUC$. Le résultat est cohérent avec le fait que : si δ_{luc} est très grand, une protéine Ptl traduite émet immédiatement un photon de lumière. Le signal lumineux est proportionnel au taux de synthèse des protéines et donc à la quantité d'ARN messenger, et ne donne pas d'information sur la quantité de la protéine au fil du temps.

Si nous supposons le paramètre δ_l petit, alors la dynamique de Ptl est donnée au premier ordre :

$$\frac{dPtl}{dt} = \beta_{ptl} Mtl - \delta_{ptl} \kappa_{pt} \frac{Ptl}{\kappa_{pt} + Pt + Ptl + Ptl_i} \quad (4.36)$$

Ainsi, Ptl sera très proche de Pt si les paramètres cinétiques des espèces supplémentaires, ARN et protéine $TOC :LUC$, tendent vers ceux du gène natif TOC . Cependant, l'insertion non contrôlée du gène rapporteur dans le génome peut modifier les taux basals de transcription. De plus, la présence de la luciférase peut altérer également le taux de traduction et l'interaction

avec la protéase. Rigoureusement, il n'est donc pas évident que le profil de concentration de la luciférase soit identique à celui de la protéine *TOC*. Mais la différence entre le profil de luminescence et de la protéine ne remet en cause ni la conclusion sur l'importance de *toc* et *cca* au cœur de la boucle, ni les résultats de Thommen [94] sur la robustesse aux fluctuations.

Les valeurs intermédiaires de la durée de vie δ_l fournissent un signal de luminescence intermédiaire entre l'ARNm et la protéine *TOC :LUC*. La détermination de ce paramètre est effectuée dans la partie 4.4.2.

4.4.2 Temps d'inactivation de la luciférase

Selon le temps d'émission d'un photon $1/\delta_l$, le signal de luminescence sera plutôt de la concentration de l'ARN ou de la protéine *TOC :LUC*.

La valeur du paramètre $1/\delta_l$ est déterminée à partir des lignées transcriptionnelles *pTOC :LUC*. Ce calcul est effectué en supposant que la décroissance du signal *pTOC :LUC* se passe sans traduction des protéines entre 15 et 29 heures (voir figure 4.2).

La dynamique de la luciférase active est modélisée par :

$$\frac{dLUC}{dt} = \beta M(t) - \delta_l LUC(t) \quad (4.37)$$

Pendant l'intervalle de temps considéré, l'équation se réduit à :

$$\frac{dLUC}{dt} = -\delta_l LUC(t) \Rightarrow LUC(t) = LUC(t = 15)e^{-\delta_l t} \quad (4.38)$$

À la figure 4.3, les résultats de l'ajustement et les valeurs trouvées de δ_l sont illustrés. La durée d'émission de la luciférase est d'environ 9.45 heures. Ce temps de vie est de très loin suffisamment grand pour que l'émission lumineuse ne se passe pas juste après la traduction. A priori, la dégradation de la protéine *TOC* ou *CCA* prédomine dans la dynamique. Ainsi, les profils de luminescence des lignées traductionnelles devraient être proches des profils de concentration de protéine.

Des travaux réalisés par F-Y Bouget et al. confirment le résultat obtenu. Ils ont réalisé des expériences en injectant dans des lignées de *TOC :LUC* de l'émétine dihydrochloride. Ce produit inhibe la traduction. Ainsi, la décroissance du signal ne découle que des processus de dégradation ou de désactivation de la protéine. Il y a deux processus en cours : l'émission de lumière par la luciférase et la dégradation de la protéine *TOC :LUC*. Ces deux dernières réactions s'additionnent et provoquent une décroissance plus rapide du signal de la protéine *TOC :LUC* par rapport au signal de décroissance de *pTOC :LUC*. Cependant, le signal ne décroît pas instantanément et indique que l'émission de lumière n'a pas lieu juste après la traduction. Le temps de relaxation donne une borne inférieure du temps de vie d'émission de lumière. Le fait que ce temps de vie soit important entraîne que les profils de *TOC :LUC* sont proches de ceux de *TOC*.

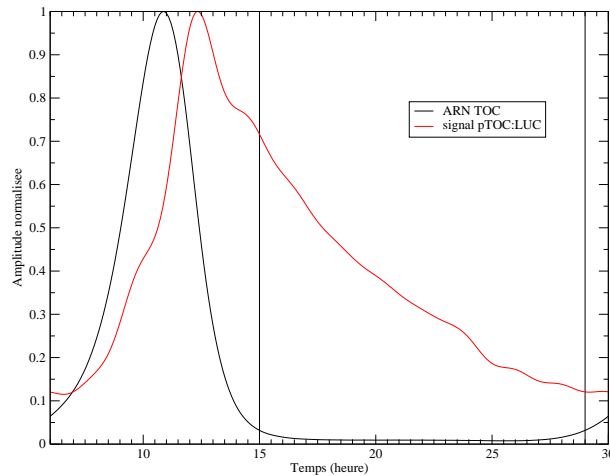


FIGURE 4.2 – Superposition des données de micropuces à ADN de *toc* (expression de l'ARN de TOC en noir) avec le signal *pTOC :LUC* (en rouge). Entre 15 heures et 29 heures, si le niveau de l'ARN de *LUC* est aussi faible que celle de l'ARN de *TOC*, la décroissance du signal lumineux de *LUC* est uniquement due à l'émission lumineuse de la luciférase car il n'y a pas de traduction de l'ARNm en protéine.

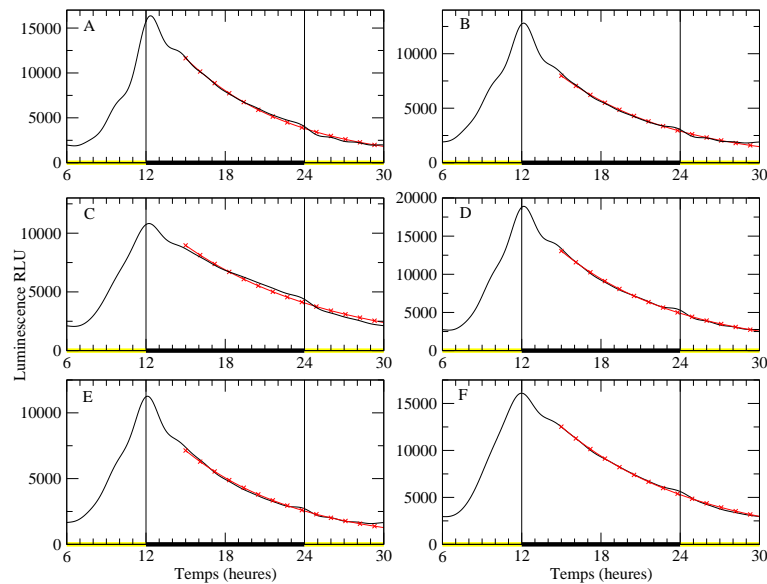


FIGURE 4.3 – Ajustement par une exponentielle décroissante des signaux brutes de ptoC. Les temps de vie sont respectivement pour chaque figure : A : $\delta_l = 0.12h^{-1}$, B : $\delta_l = 0.11h^{-1}$, C : $\delta_l = 0.09h^{-1}$, D : $\delta_l = 0.11h^{-1}$, E : $\delta_l = 0.11h^{-1}$, F : $\delta_l = 0.1h^{-1}$, donnant une moyenne de 0.11 .

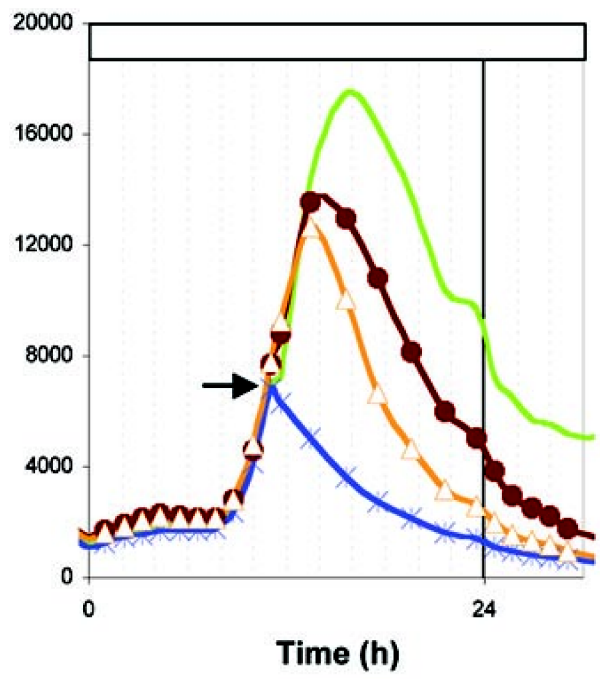


FIGURE 4.4 – Figure tirée de [87]. La courbe nous intéressant est en bleue. L’injection de l’émétine dihydrochloride provoque l’arrêt de la traduction. La décroissance du signal résulte de l’émission de lumière par la luciférase, mais elle est accéléré par la dégradation de la protéine *TOC :LUC*.

4.5 Résultat de l'ajustement par les modèles détaillés

Nous présentons les résultats de l'ajustement des données expérimentales par les modèles présentés dans la section (4.2).

4.5.1 Protocole des ajustements

L'algorithme recherche le jeu de paramètres croisant au mieux les simulations numériques des modèles mathématiques avec les données expérimentales, en se basant sur la fonction de score définie à la sous-section 4.3.2 à la page 122. La fonction de score utilise les données de micropuce à ADN de TOC et de CCA, ainsi que la série de Fourier du signal de luminescence. Les données de micropuce à ADN, malgré leur pas de mesure assez large (un point toutes les trois heures), sont des données fiables et communes aux expériences réalisées sur toutes les lignées. Les ajustements concernent 6 expériences de luminescence (une par lignée) pour chaque type de transformation génétique d'*Ostreococcus Tauri*, soit *TOC :LUC*, *CCA :LUC*, *pTOC :LUC* et *pCCA :LUC* (soit un total de 24 ajustements).

Entre les modèles de la boucle à deux gènes tenant en compte la cinétique de la luciférase et le modèle simple vu au chapitre 3, un certain nombre de paramètres sont communs puisque les nouveaux modèles ne font que compléter le modèle de base. Dans l'article de Thommen et al. [94], les ajustements ont été effectués uniquement sur les profils expérimentaux des micropuces à ADN. L'ajustement des ARN étant bon. Nous reprenons les paramètres de cet article pour notre modèle tenant compte de la luciférase afin de réduire l'espace d'exploration des paramètres.

Les ajustements numériques sont réalisés sans couplage à la lumière, c'est-à-dire sans changements des paramètres lors du passage entre le jour et la nuit (et inversement). Thommen et al. ont montré [94] que le couplage à la lumière n'est pas visible dans les signaux entraînés par les Zeitgebers.

4.5.2 Ajustement des lignées *pTOC :Luc* et *pCCA :LUC*

Les ajustements des lignées *pTOC :LUC* sont à la figure 4.5. Les ajustements sont corrects, le meilleur étant celui de la lignée 6. Cependant, les données expérimentales sont difficiles à reproduire à 9 et 13 heures environs, soit au voisinage du maximum, en raison de multiples fluctuations. Malgré cela, le niveau bas, la décroissance et la phase du signal sont retrouvés numériquement. L'ajustement du niveau bas s'explique en général par un taux basal de luminescence du gène *pTOC :LUC* (paramètre μ_{tl}) plus élevé que celui du gène *toc* (paramètre λ_t et μ_t , voir tableau en annexe E.1 page 166).

Les ajustements des signaux de *pCCA :Luc* sont présentés sur la figure 4.6. Les ajustements des lignées 1 et 3 sont corrects contrairement aux autres ajustements. La partie la plus difficile à ajuster concerne la montée des signaux *pLHY :LUC*. Le modèle ne peut pas reproduire le rebond se passant aux alentours de 22 heures et peu présent dans les lignées 1 et 3. Dans les ajustements par le modèle ne tenant pas compte de la cinétique de la luciférase (chapitre 3), il était déjà difficile de reproduire la montée du profil des protéines *CCA*, même si le profil résultait d'une moyenne. Seuls les paramètres des lignées 1 et 3 seront discutés.

4.5.3 Comparaison entre les lignées *pTOC :LUC* et *pCCA :LUC*

Certains paramètres cinétiques liés à la luciférase sont comparables entre les lignées *pTOC :LUC* et *pCCA :LUC* puisqu'ils concernent les mêmes molécules d'ARNm et protéine *LUC*. Il s'agit du paramètre de traduction de l'ARNm en protéine, du temps de vie de la luciférase et de la dégradation de l'ARNm de *LUC*. Les fonctions de transcription ne sont pas analysables car l'insertion aux différents sites de l'ADN modifient leurs valeurs.

Le taux de traduction β_{ml} des lignées *pTOC :LUC* est 0.35 h^{-1} (en moyenne) et celui des lignées 1 et 3 de *pLHY :LUC* sont 0.22 et 0.26 h^{-1} . Ces valeurs sont relativement cohérentes.

En moyenne, le temps de demi-vie de la luciférase est de 7.6 h pour les lignées *pTOC :LUC*. Ce dernier est du même ordre de grandeurs mais inférieur à celui trouvé dans la section 4.4.2 correspondant à l'ajustement direct du signal sur une la partie décroissante (9.45 h). Cette différence s'explique par le fait que l'ajustement par le modèle de *pTOC :LUC* ne redescend pas assez vite après le pic d'expression maximum. L'ARNm de *LUC* (donnée non montré) est plus stable impliquant un niveau bas plus élevé. Mais le signal de l'ARNm reste constant durant toute la décroissance du signal *LUC* à cause d'une saturation de la transcription et de la dégradation de l'ARNm. Il en est de même pour la lignée 1 de *pCCA :LUC*. Cependant, la dégradation de l'ARNm de la lignée 3 est principalement linéaire contrairement aux autres lignées empêchant un niveau constant d'ARNm. Malgré cette importante différence, cette lignée a le deuxième meilleur score parmi celles de *pTOC :LUC* et *pCCA :LUC*. Le modèle se révèle alors insuffisant à expliquer la stabilité de l'ARNm en raison d'un surajustement lié aux trop grands nombres de paramètres et à un manque de données à ajuster.

Pour les lignées transcriptionnelles, les profils numériques de *TOC* et de *CCA* sont similaires puisque l'ajustement des paramètres concernent uniquement les constantes cinétiques liées à la luciférase sans perturber la boucle de l'horloge centrale. L'ajustement des lignées traductionnelles, fait par la suite, n'aboutit pas toujours à des profils de protéines *TOC* et *CCA* semblables comme expliqué dans la section suivante.

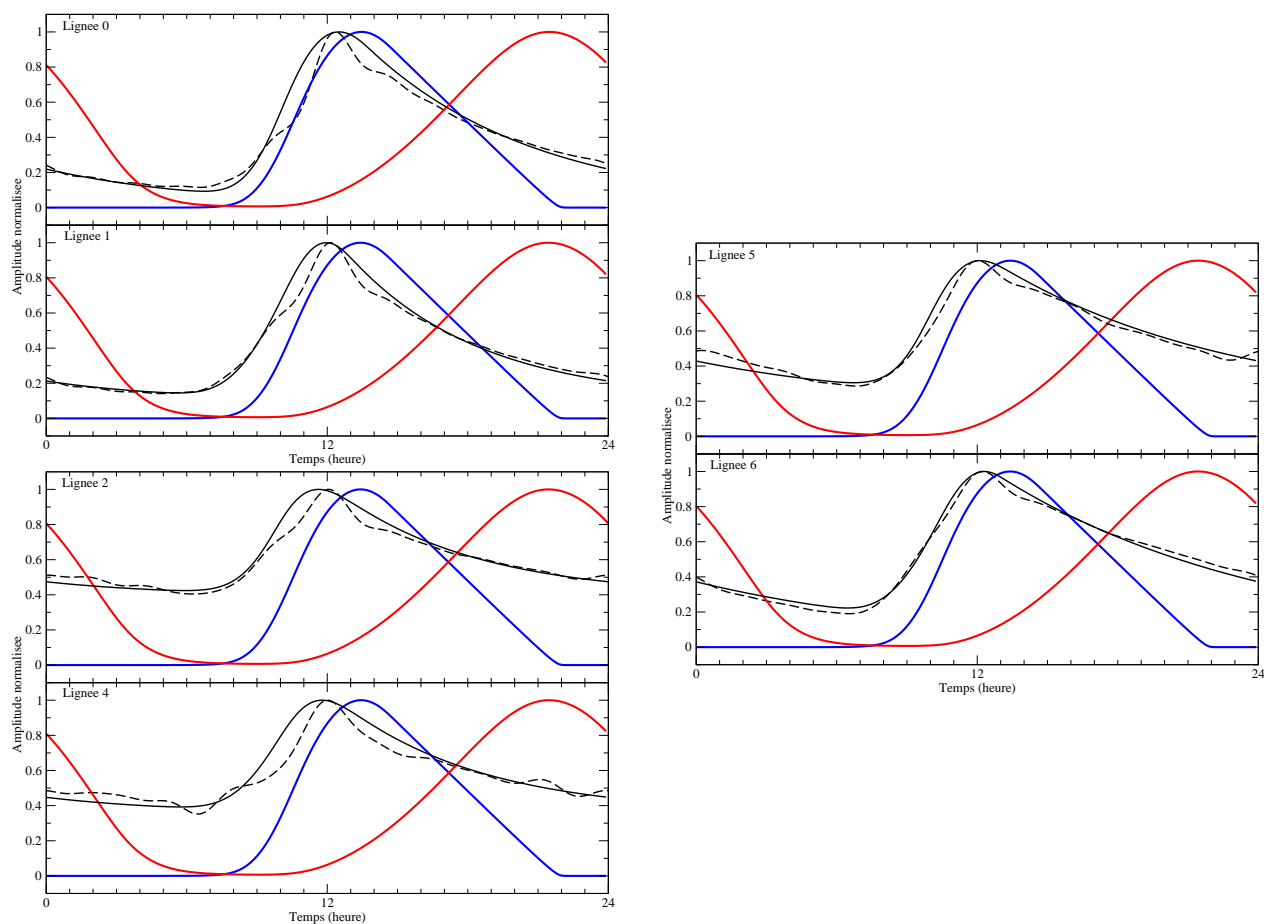


FIGURE 4.5 – Ajustement du modèle sans couplage à la lumière. Les profils expérimentaux de $pTOC :LUC$ sont en traits pointillés noirs, les profils numériques de $pTOC :LUC$ sont en traits noir plein. Les profils numériques de la protéine TOC et CCA sont respectivement en trait bleu et rouge. Les scores et les valeurs du taux δ_l pour chaque lignée sont : lignée 0 : $score = 0.61$, $\delta_l = 0.14 h^{-1}$; lignée 1 : $score = 0.51$, $\delta_l = 0.18 h^{-1}$; lignée 2 : $score = 0.51$, $\delta_l = 0.17 h^{-1}$; lignée 4 : $score = 0.6$, $\delta_l = 0.17 h^{-1}$; lignée 5 : $score = 0.49$, $\delta_l = 0.1 h^{-1}$; lignée 6 : $score = 0.45$, $\delta_l = 0.09 h^{-1}$;

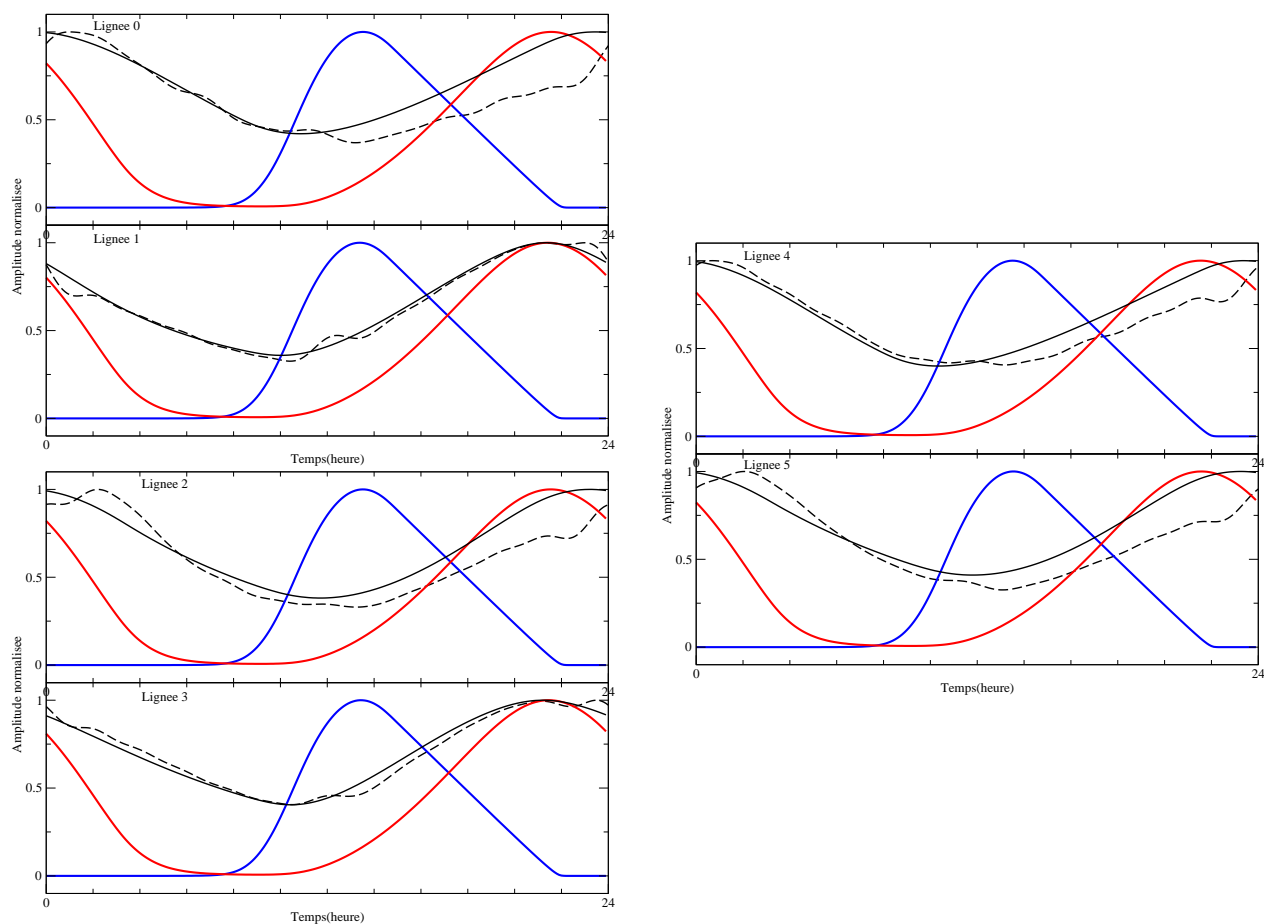


FIGURE 4.6 – Ajustement du modèle sans couplage à la lumière. En trait plein noir, simulation numérique de $pCCA :Luc$. En trait noir discontinu, profils expérimentaux de $pCCA :LUC$. Les profils numériques de la protéine TOC et CCA sont respectivement en trait bleu et rouge. Les scores et les valeurs du taux δ_l pour chaque lignée sont : lignée 0 : $score = 1.3$, $\delta_l = 0.16 h^{-1}$; lignée 1 : $score = 0.46$, $\delta_l = 0.17 h^{-1}$; lignée 2 : $score = 1.1$, $\delta_l = 0.11 h^{-1}$; lignée 3 : $score = 0.48$, $\delta_l = 0.22 h^{-1}$; lignée 4 : $score = 0.82$, $\delta_l = 0.16 h^{-1}$; lignée 5 : $score = 1.28$, $\delta_l = 0.1 h^{-1}$;

4.5.4 Indice de saturation et de régulation

Afin de comparer les paramètres des lignées traductionnelles, nous définissons deux indices. L'indice de saturation mesure le degré moyen de saturation pour la dégradation enzymatique de l'espèce X :

$$I_X = \frac{\bar{X}}{\kappa_X} \quad (4.39)$$

Le paramètre κ est la constante de saturation et \bar{X} est la moyenne de X sur une période circadienne. Si l'espèce est soumise à une dégradation compétitive (plusieurs espèces voisines sont dégradées par la même enzyme), alors l'indice devient :

$$I_X = \sum_{i=0}^N \frac{\bar{X}_i}{\kappa_{X_i}} \quad (4.40)$$

Le paramètre N est le nombre de protéines i soumises à la compétition de dégradation. L'indice de saturation est à comparer avec la valeur 1 pour mesurer la saturation. Un indice faible (resp. grand) devant 1 indique une dégradation linéaire (resp. saturée).

De même, nous établissons un indice de régulation lié au seuil de régulation de la transcription et défini par le rapport :

$$S_{M_X} = \frac{\bar{Y}}{P} \quad (4.41)$$

Le paramètre P est le seuil de régulation de la transcription et la variable Y la concentration du facteur de transcription. Lors d'une régulation compétitive par plusieurs facteurs de transcription, l'indice s'écrit alors :

$$S_{M_X} = \sum_{i=0}^N \frac{\bar{Y}_i}{P_i} \quad (4.42)$$

En comparant l'indice S à la valeur 1, on sait si le gène est susceptible de commuter plus ou moins facilement lorsque la concentration du facteur de transcription oscille entre ses valeurs minimum et maximum. Un indice S faible (resp. grand) devant 1 signifie que la commutation du gène est difficile (resp. facile) à atteindre.

4.5.5 Ajustement des lignées traductionnelles *TOC :LUC* et *CCA :LUC*

Les valeurs des paramètres communs entre le modèle du chapitre 3 et les modèles représentant les lignées traductionnelles sont celles trouvées par les travaux de Thommen et al. [94]. Cependant, une renormalisation de ces paramètres est nécessaire lors de l'ajustement car la période libre de l'oscillateur dépend maintenant des paramètres liés aux gènes insérés dans le génome. Cette renormalisation sert à ajuster la période libre de l'oscillateur à 24 heures. On procède ainsi car, dans les travaux de Thommen et al. [94], une telle remise à l'échelle de la période n'avait alors pas contrarié l'ajustement par l'oscillateur libre. En revanche, il ne serait pas étonnant que cette façon de faire puisse ne pas être adaptée pour les lignées *TOC :LUC* car leur période s'allonge fortement en condition LL (voir section 3.3.3.2).

Les résultats de l'ajustement des lignées *TOC :LUC* sont représentés à la figure 4.7. Les meilleurs ajustements sont obtenus pour les lignées 1 et 4 et sont pris comme référence pour la discussion.

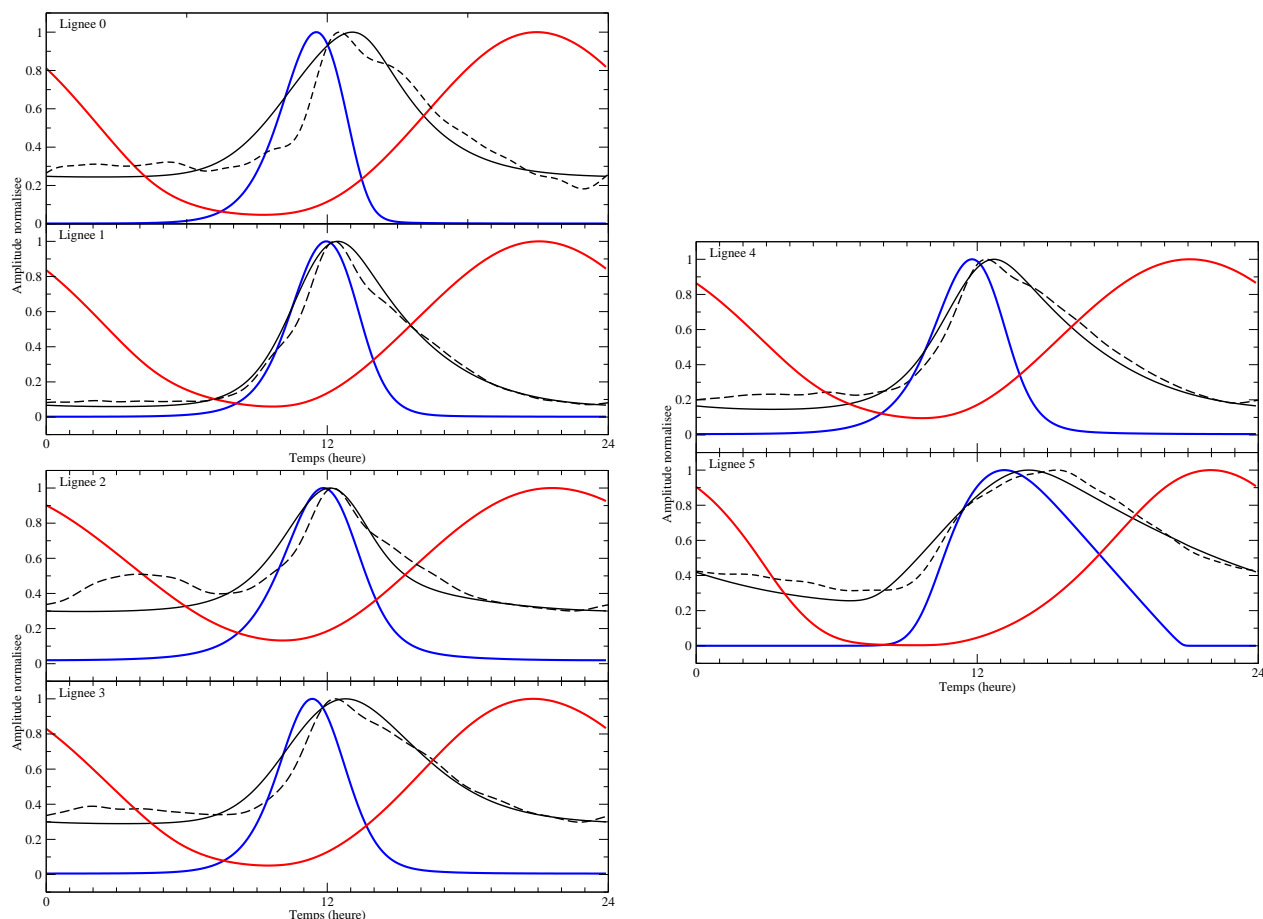


FIGURE 4.7 – Ajustement du modèle sans couplage à la lumière. En trait plein noir, simulation numérique de $TOC : Luc$. En trait bleu (rouge), simulation numérique de $TOC (CCA)$. L’amplitude maximale de TOC et CCA sont ramenées à 1 par souci de clarté mais ne correspond pas aux vrais maximums . En trait noir discontinu, série de fourier résultant donnant le profil expérimental. Scores et valeurs du taux d’émission lumineuse δ_l pour chaque lignée : lignée 0 : $score = 0.54$, $\delta_l = 1.66 h^{-1}$; lignée 1 : $score = 0.23$, $\delta_l = 0.29 h^{-1}$; lignée 2 : $score = 1.22$, $\delta_l = 0.9 h^{-1}$; lignée 3 : $score = 0.56$, $\delta_l = 3.95 h^{-1}$; lignée 4 : $score = 0.38$, $\delta_l = 0.07 h^{-1}$; lignée 5 : $score = 0.57$, $\delta_l = 23.8 h^{-1}$;

Les profils des protéines TOC et CCA sont fortement modifiés par rapport aux simulations numériques des lignées $pTOC : LUC$ et $pCCA : LUC$. Les différences dans le profil de la protéine TOC se situe au niveau du temps du maximum d’expression, avant 12 h, et de la décroissance du signal éloignée de la saturation.

Les ajustements des lignées $CCA : LUC$ sont montrés à la figure 4.8. Les deux ajustements pris comme référence sont ceux des lignées 4 et 5. Les profils des protéines TOC et CCA sont très similaires par rapport aux profils des lignées traductionnelles.

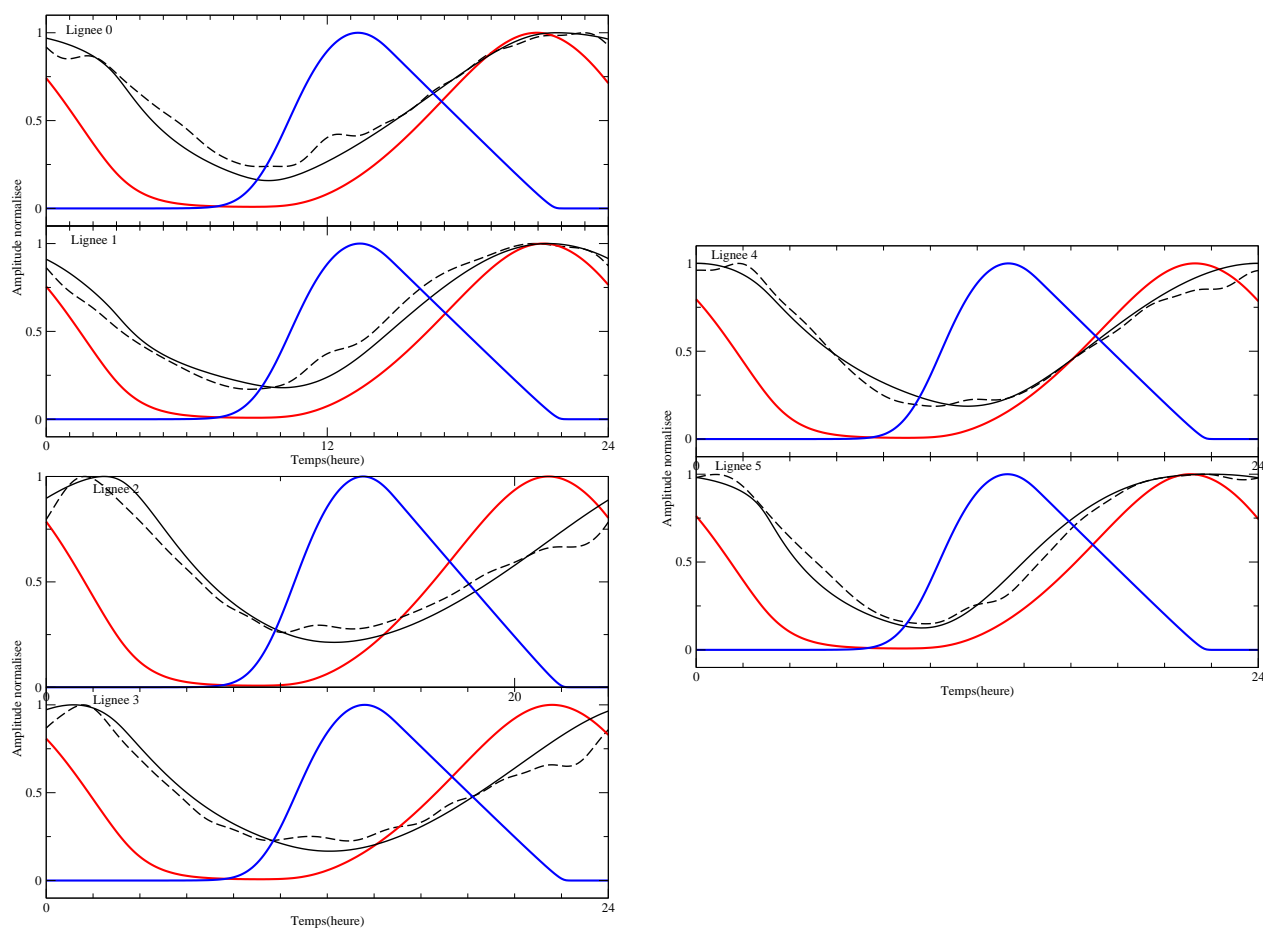


FIGURE 4.8 – Ajustement du modèle sans couplage à la lumière. En trait plein noir, simulation numérique de *CCA : Luc*. Les profils de protéine *CCA* en trait rouge et de *TOC* en trait bleu. Ces profils sont divisés pour avoir 1 comme maximum afin d’avoir une meilleure visibilité de l’ensemble des profils. En trait noir discontinu, profil expérimental de *CCA :LUC*. Scores et valeurs du taux d’émission lumineuse δ_l pour chaque lignée : lignée 0 : $score = 0.9$, $\delta_l = 33 h^{-1}$; lignée 1 : $score = 1.0$, $\delta_l = 14.2 h^{-1}$; lignée 2 : $score = 0.72$, $\delta_l = 12 h^{-1}$; lignée 3 : $score = 1.0$, $\delta_l = 0.18 h^{-1}$; lignée 4 : $score = 0.63$, $\delta_l = 0.01 h^{-1}$; lignée 5 : $score = 0.51$, $\delta_l = 17 h^{-1}$;

4.6 Discussions des paramètres des ajustements des lignées traductionnelles

L'insertion de copies supplémentaires de gènes dans les lignées traductionnelles entraîne une surexpression des protéines. Au-delà d'un certain seuil, le rythme est détruit (voir section 3.2.5 page 82). Mais l'ajout en faible nombre de copies n'altère pas l'existence du rythme même en sélectionnant les lignées ayant l'expression de luminescence la plus forte. Cependant, des effets sont introduits nécessairement et modifient la transcription et la dégradation. Par exemple, le seuil de régulation de la transcription est atteint plus rapidement et la saturation de la dégradation est favorisée. En principe, les simulations des modèles des lignées traductionnelles devraient montrer ces deux dernières tendances pour être cohérentes.

Cependant, les résultats numériques du modèle *TOC :LUC* ne respectent pas cela. L'indice de saturation de la protéine *TOC* diminue d'un ordre de grandeur par rapport à celui des lignées transcriptionnelles et de *CCA :LUC* (voir tableau 4.9). On tend alors à éloigner la dégradation de *TOC* de la saturation plutôt que de s'en rapprocher. En revanche, dans les simulations des lignées traductionnelles de *CCA :LUC*, la dégradation de la protéine *TOC* reste plutôt saturée.

L'indice du seuil de régulation S_{Mcca} des lignées *TOC :LUC* descend également d'un ordre de grandeur, signifiant que la commutation est plus difficile à réaliser, alors que la transcription de l'ARNm de *CCA* est soumise à une surexpression des protéines *TOC* (voir tableau 4.10). Les indices des lignées de *CCA :LUC* n'ont pas de tels écarts, ce qui permet en partie de retrouver des profils des protéines *TOC* et *CCA* similaires aux lignées transcriptionnelles.

Plusieurs explications peuvent être avancées pour comprendre cette déviation lors de l'ajustement dans les lignées *TOC :LUC*. Le choix de la période libre peut expliquer l'incohérence des paramètres des lignées *TOC :LUC*. Cette dernière est fixée à 24 heures alors que les expériences en jour continu prouvent un allongement de la période libre des lignées *TOC :LUC*. Peut-être n'est-il pas possible d'ajuster l'oscillateur en condition d'entraînement par un modèle d'oscillateur libre. Cependant, on ne peut pas effectuer l'ajustement en incorporant une modulation paramétrique pour simuler l'entraînement dans les lignées *TOC :LUC* tout en conservant le profil des micropuces à ADN. En effet, rien ne garantit que les profils ne sont pas également fortement altérés dans les lignées *TOC :LUC*.

Sans que cela soit parfaitement prouvé, il est également possible que le nombre de constantes soit trop important, et donc qu'il existe plusieurs zones dans l'espace des paramètres minimisant le score sans apporter un véritable sens pour le modèle biologique. Ce surajustement est détecté pour le modèle des lignées *CCA :LUC* du fait de la valeur du taux δ_l de la lignée 5 dont les simulations ajustent bien les profils expérimentaux. Alors que les valeurs de δ_l indiquent une émission stable, c'est-à-dire un profil éloigné de l'ARNm, le taux d'émission de photon de la lignée 5 (*CCA*) est très grand, rentrant en contradiction avec tous les autres ajustements.

	<i>pLHY :LUC</i>	<i>TOC :LUC</i>		<i>LHY :LUC</i>	
	<i>PTOC :LUC</i>	lignée 1	lignée 4	lignée 4	lignée 5
I_{Pc}	1.02	0.25	0.19	1.02	0.72
I_{Pt}	62.25	0.11	0.04	61.9	10.3

FIGURE 4.9 – Tableau regroupant les indices de saturation des lignées de fusion. En rouge, l'indice est fourni par la formule 4.40.

	<i>PLHY :LUC</i>	<i>TOC :LUC</i>		<i>LHY :LUC</i>	
	<i>PTOC :LUC</i>	lignée 1	lignée 4	lignée 4	lignée 5
S_{Mcca}	1.07	0.03	0.016	1.07	0.18
S_{Mtoc}	45.35	10.97	8.3	44.94	14.54
$S_{Mccaluc}$	-	-	-	0.32	1.27
$S_{Mtocluc}$	-	7.97	1.73	-	-

FIGURE 4.10 – Tableau regroupant les indices de seuil de régulation du gène.

4.7 Conclusion

La modification du génome par l'insertion de gène de la luciférase justifie l'établissement d'un modèle plus complexe tenant compte de la cinétique de la luciférase. L'insertion du gène a un double effet. La première influence, difficilement quantifiable par un modèle de large dimension, est l'altération des constantes cinétiques. Comme la zone d'insertion du gène n'est pas contrôlable, on doit s'attendre à des valeurs du taux de transcription fluctuantes entre les différentes lignées et ceux du gène natif. De plus, la présence d'une protéine de luciférase peut également modifier l'affinité chimique des facteurs de transcription. Ensuite, l'insertion d'une copie supplémentaire introduit nécessairement de la surexpression. Il faut alors pouvoir juger jusqu'à quel seuil cette surexpression modifie le phénotype.

Mais auparavant, il est important d'interpréter correctement les signaux de lumière en déterminant le temps moyen au bout duquel le photon est émis. Deux interprétations étaient possibles : soit ce temps est court et le signal représente un flux de traduction (soit le profil de l'ARN), soit ce temps est long et donc le signal est proportionnel aux profils de protéines totales. Nous avons alors déterminé par des ajustements que ce temps est suffisamment long pour conclure que la deuxième option était la bonne. Cela fut confirmé par des expériences de pharmacologie.

Grâce aux nouveaux systèmes d'équations incorporants la cinétique de la luciférase, nous avons ajusté les lignées transcriptionnelles et traductionnelles. L'ajustement des lignées transcriptionnelles concluent que le temps d'émission du photon est cohérent avec celui trouvé précédemment. De plus, les profils de gènes natifs (*TOC* et *CCA*) reconstruits sont semblables à ceux obtenus à partir des lignées traductionnelles *CCA :LUC*. Par contre, on observe une incohérence avec les profils reconstruits à partir des lignées *TOC :LUC*. En effet, on remarque une absence d'effets de surexpression dans les profils *TOC* et *CCA* qui devraient normalement être présents. Il est possible qu'il soit incorrect d'ajuster par un oscillateur libre ces dernières lignées. En effet, la période libre moyenne des lignées *TOC :LUC* est assez grande, environ de 26 heures, contrairement aux autres lignées dont la période est autour de 24 heures. Le fait de ramener la période à 24 heures a peut-être fait orienter l'algorithme à rechercher une absence de surexpression. Ainsi, il serait intéressant d'ajuster avec couplage.

En conclusion, nous avons bien démontré que le signal de luminescence est caractéristique d'une concentration de protéine et non pas d'ARNm. Mais le modèle ne permet pas de prédire avec certitude la forme des profils de protéines natifs *TOC* et *CCA* à cause de l'incohérence des ajustements par les lignées *TOC :LUC*.

Chapitre 5

Conclusion

L'analyse mathématique de systèmes biologiques révèle les différentes stratégies qu'ils emploient pour survivre. En particulier, la dynamique non linéaire de la régulation génique ou des processus de dégradation est astucieusement utilisée par la Nature pour optimiser les processus de réparation de l'ADN, de réduction des fluctuations de concentrations, ... [56, 53, 57, 51], et les rendre robustes.

Ne serait-ce que dans le plus simple des réseaux génétiques à rétroaction négative, plusieurs facteurs mènent indépendamment à l'apparition de rythmes. La saturation de la dégradation, la présence d'un délai temporel, ou une cascade de phosphorylation sont autant d'ingrédients pour l'apparition d'une dynamique non triviale. Dans ce travail, l'étude de la dynamique intrinsèque du gène (dans laquelle, l'activité génique n'est pas asservie mais devient une variable autonome) dévoile un nouveau facteur favorisant les oscillations, ou plus généralement la régularité du rythme. Pour être pertinente, cette dynamique doit être associée à un temps de réponse du gène adapté à la dynamique globale du réseau. Ce temps de réponse ne doit pas être trop court, car le gène est sinon asservi, et il ne doit pas être trop long sinon l'activité génique ne peut plus être décrite par une variable continue.

Nous avons montré qu'il existe un temps de réponse du gène optimisant l'apparition des oscillations en régime déterministe, et diminuant les fluctuations de période en régime stochastique. La présence d'un temps de réponse fini n'est pas suffisante en soi pour induire une dynamique complexe, mais il facilite son apparition en interaction avec d'autres processus. De ce fait, la multiplicité de facteurs favorisant les oscillations permet à un réseau génétique d'évoluer de manière flexible vers un comportement oscillant, en modifiant légèrement plusieurs réactions biochimiques plutôt qu'en variant brutalement un nombre restreint de paramètres biochimiques. Au final, la région de l'espace des paramètres où des oscillations peuvent être observées s'élargit notablement.

Cependant, un temps de réponse transcriptionnel non nul n'est qu'un processus parmi d'autres participant à l'expression du gène. Sont présents dans tous les cas des délais, de manière explicite ou via une étape réactionnelle (comme la traversée nucléaire). Ces derniers processus n'ont cependant pas été étudiés en combinaison avec un temps de réponse transcriptionnel. Les travaux actuels de J. Wang et al. ont pour but d'expliquer l'influence de ces différents types de réactions à délai sur la dynamique du gène autorégulé avec mémoire transcriptionnelle [132].

La complexité immédiatement supérieure à celle d'un réseau à un seul gène se trouve dans un réseau à deux gènes. Ce modèle a été utilisé pour comprendre l'horloge circadienne d'une

algue verte microscopique *Ostreococcus Tauri*. Expérimentalement, deux gènes centraux de l'horloge circadienne ont été mis en évidence par l'étude de leur comportement grâce aux données d'expression génique (luminescence émise par des lignées de fusion transcriptionnelle et traductionnelle, micropuce à ADN) réalisées dans différentes conditions expérimentales (cycle circadien, jour continu). L'hypothèse qu'il nous fallait confirmer était que l'horloge est basée sur un réseau génétique à rétroaction négative : la protéine *TOC* active la transcription du gène *cca* et la protéine *CCA* réprime la transcription du gène *toc* [93]. Ce modèle simple décrit ce qui semble être un oscillateur central de l'horloge, celle-ci étant probablement plus complexe, comme pour la plupart des modèles circadiens [84, 100].

Même si ce modèle à une boucle n'est pas complet, un ajustement des données expérimentales de *toc* et *cca* par le modèle mathématique correspondant au réseau justifie l'importance de ces gènes au coeur de l'horloge. Après traitement des données expérimentales obtenues par micropuces ADN ou mesure de la luminescence (révélant au passage un biais dans les séries temporelles), l'ajustement numérique simultané des données de deux types d'expériences différentes est excellent et montre que l'horloge circadienne est robuste par rapport aux modifications de l'environnement.

Cependant, ces résultats font par souci de simplicité l'hypothèse que les données de luminescence des lignées traductionnelles sont directement assimilables aux profils d'expression des protéines natives. Or, le gène rapporteur est non seulement rajouté au gène sauvage, mais il est inséré dans une zone non contrôlée de l'ADN, modifiant les taux de transcription et différents paramètres cinétiques. Par ailleurs, le temps d'émission de la luciférase influe sur l'interprétation des signaux de luminescence. Nous avons donc déterminé ce temps à partir des séries temporelles expérimentales. Nous trouvons que ce temps est long, de l'ordre de plusieurs heures, ce qui implique que le profil de luminescence est plus proche du profil temporel de la protéine que de celui de l'ARN. Afin de prendre en compte plus finement la cinétique de la luciférase, nous avons commencé à explorer des modèles détaillés, incorporant explicitement les gènes rapporteurs. Nous trouvons des résultats dans l'ensemble cohérents avec ceux obtenus précédemment, même si l'augmentation de la période libre de la lignée *TOC :LUC* due à la surexpression remet en cause la pertinence de l'ajustement par un oscillateur libre de période naturelle 24 heures.

En conclusion, nous espérons avoir illustré dans ce travail l'utilité de la dynamique non linéaire pour comprendre le fonctionnement des oscillateurs biologiques basés sur des circuits génétiques, et en dégager des principes généraux. Dans le cas d'un gène auto-réprimé, nous avons montré comment la nature peut jouer simultanément sur délai et dégradation pour engendrer des oscillations. Dans l'étude de l'horloge circadienne d'*Ostreococcus tauri*, nous avons contribué à isoler son oscillateur central et démontrer ses propriétés de robustesse par l'ajustement simultané d'expériences effectuées dans des conditions différentes.

Annexe A

Annexe au chapitre 2 : Oscillations in the Expression of a Self-Repressed Gene Induced by a Slow Transcriptional Dynamics

paru dans :

« Physical Review Letters, volume 102, numéro 6, page 068104 en Février 2009 »

Auteurs :

Pierre-Emmanuel Morant, Quentin Thommen, François Lemaire, Constant Vandermoëre, Benjamin Parent et Marc Lefranc

Oscillations in the Expression of a Self-Repressed Gene Induced by a Slow Transcriptional Dynamics

Pierre-Emmanuel Morant,^{1,2} Quentin Thommen,^{1,2} François Lemaire,^{3,4} Constant Vandermoère,^{1,2} Benjamin Parent,^{5,6,7} and Marc Lefranc^{1,2,7,*}

¹Université des Sciences et Technologies de Lille, PhLAM, F-59655 Villeneuve d'Ascq, France

²CNRS, UMR 8523, FR 2416, F-59655 Villeneuve d'Ascq, France

³Université des Sciences et Technologies de Lille, LIFL, 59655 Villeneuve d'Ascq, France

⁴CNRS, UMR 8022, F-59655 Villeneuve d'Ascq, France

⁵Université des Sciences et Technologies de Lille, UGSF, F-59655 Villeneuve d'Ascq, France

⁶CNRS, UMR 8576, F-59655 Villeneuve d'Ascq, France

⁷Interdisciplinary Research Institute, USR CNRS 3078, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq, France

(Received 7 March 2008; published 13 February 2009)

We revisit the dynamics of a gene repressed by its own protein in the case where the transcription rate does not adapt instantaneously to protein concentration but is a dynamical variable. We derive analytical criteria for the appearance of sustained oscillations and find that they require degradation mechanisms much less nonlinear than for infinitely fast regulation. Deterministic predictions are confirmed by stochastic simulations of this minimal genetic oscillator.

DOI: 10.1103/PhysRevLett.102.068104

PACS numbers: 87.18.Vf, 82.40.Bj, 87.16.Yc

Networks of genes interacting via regulatory proteins modulating their activities are highly nonlinear systems which display a variety of dynamical behaviors [1–3]. Their modeling has generally assumed that gene activation is fast compared to other processes so that transcription rate reacts instantaneously to protein concentration. However, transcription is a complex process [4]. In the last years, it has been increasingly recognized that gene activity fluctuations can be slow and that this can affect the behavior of gene regulatory networks. In particular, slow transcriptional bursting and transcriptional memory have been observed experimentally [5–7]. Theoretically, it has been shown that slow activation dynamics can lead to bursts in expression [8,9], induce bistability [10] or modify the flipping rate of a genetic switch [11,12].

In this Letter, we show that slow promoter dynamics can also lead to oscillations by investigating how transcriptional dynamics modifies the behavior of a single gene repressed by its own protein [13–20]. This old problem of theoretical biology has been recently revived by the study of the *Hes1* gene involved in the somite clock [21]. The usual view is that oscillations appear in this genetic circuit only when additional steps are inserted in the feedback loop [3,16]. In the Goodwin and Bliss oscillators [15–17], the gene protein catalyzes synthesis of the actual repressor. In early circadian models, transport of the repressor into the nucleus [13,14] is a key oscillatory ingredient. In fact, the mere introduction of a time delay in the one-gene circuit model (accounting for protein transport or more generally a cascade of intermediate steps [22]) can destabilize it [3,18–20]. As will be of particular interest here, oscillations may also be induced by strongly nonlinear degradation mechanisms [23].

For simplicity, we study the case of an elementary kinetic equation describing regulation through protein-DNA binding [24], which we, however, prefer to view as a minimal description of transcriptional memory in more complex mechanisms. We derive an analytical expression of the oscillation threshold, and show that when the gene response time is appropriately tuned, the one-gene circuit can be destabilized (and oscillations induced) by degradation mechanisms much less nonlinear than for infinitely fast regulation. This result provides new insights into the interplay of nonlinearity and time delay. Stochastic simulations confirm that the main results of our analysis carry over to low-copy-number situations.

Our study is based on the following three-variable model describing the genetic circuit represented in Fig. 1:

$$\dot{G} = \theta_0(1 - G) - \alpha_0 C(P)G, \quad (1a)$$

$$\dot{P} = n\dot{G} + \beta_0 M - \delta_P F(P), \quad (1b)$$

$$\dot{M} = \mu_0 + \lambda_0 G - \delta_M H(M), \quad (1c)$$

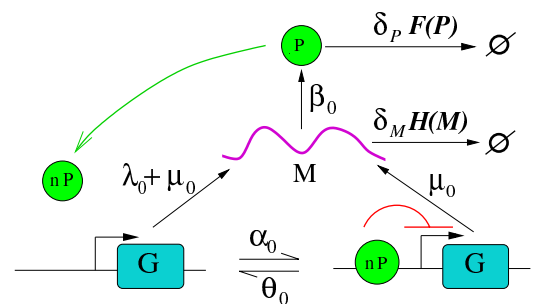


FIG. 1 (color online). Reaction diagram of the self-regulated gene circuit.

where G , P , and M represent gene activity, protein and RNA copy numbers. Equation (1a) formally describes the kinetics of protein-DNA binding at rate α_0 and unbinding at rate θ_0 [24]. Possible cooperativity effects are taken into account via the function $C(P)$ and the number n of proteins binding DNA. Single-protein regulation corresponds to $C(P) = P$, $n = 1$. More generally, Eq. (1a) is a minimal model for the dynamics of an effective gene activity G slowly relaxing towards an equilibrium value given by the gene regulation function $G = 1/[1 + C(P)/C(P_0)]$, with P_0 the half-expression threshold. Such a model can be obtained as the leading approximation of a mechanistic model of transcription including all processes concurring to gene expression (e.g., protein-DNA interaction, formation of open complex, polymerase recruitment, chromatin remodeling, ...) when there is a dominant limiting step. Equation (1a) causes gene activity G to lag fast changes in protein level, and plays a dynamical role similar to explicit time delays [18–20] or to transport equations [14]. In Eq. (1b), the three terms correspond to binding or unbinding, translation, and degradation. Equation (1c) describes transcription at rate $\mu_0 + \lambda_0 G$ and RNA degradation. In order to understand how oscillations can be induced by tuning protein and RNA degradation, we derive the oscillation criterion for arbitrary degradation functions $F(P)$ and $H(M)$ with unit derivative at zero, δ_P and δ_M being the low-copy-number degradation rates.

By renormalizing time, variables, parameters, cooperativity and degradation functions according to

$$t = \frac{t'}{\delta_M}, \quad G = g, \quad P = pP_0, \quad M = mM_0, \quad (2a)$$

$$P_0 = C^{-1}\left(\frac{\theta_0}{\alpha_0}\right), \quad M_0 = \frac{\delta_P P_0}{\beta_0}, \quad \theta = \frac{\theta_0}{\delta_M}, \quad (2b)$$

$$\alpha = \frac{\theta_0}{P_0 \delta_M}, \quad \delta = \frac{\delta_P}{\delta_M}, \quad \lambda = \frac{\lambda_0}{M_0 \delta_M}, \quad \mu = \frac{\mu_0}{M_0 \delta_M}, \quad (2c)$$

$$c(p) = \frac{C(P)}{C(P_0)}, \quad f(p) = \frac{F(P)}{P_0}, \quad h(m) = \frac{H(M)}{M_0}. \quad (2d)$$

Equations (1) can be rewritten in dimensionless form

$$g' = \theta[1 - g(1 + c(p))], \quad (3a)$$

$$p' = n\alpha[1 - g(1 + c(p))] + \delta[m - f(p)], \quad (3b)$$

$$m' = \mu + \lambda g - h(m), \quad (3c)$$

where $x' = dx/dt'$. When f and g are monotonic and $h[f(\infty)] > \mu$, model (3) has a single steady state (g_*, p_*, m_*) satisfying the fixed point equations

$$g_* = \frac{1}{1 + c(p_*)}, \quad m_* = f(p_*), \quad g_* = \frac{h(m_*) - \mu}{\lambda}. \quad (4)$$

Note that the steady state depends only on parameters λ and μ as well as on functions c , f and h , whereas param-

eters θ , α , δ control time scales. The behavior of the degradation and cooperativity functions in the neighborhood of the steady state is described by the slopes

$$s = \left. \frac{df(p)}{dp} \right|_{p=p_*}, \quad u = \left. \frac{dh(m)}{dm} \right|_{m=m_*}, \quad v = \left. \frac{dc(p)}{dp} \right|_{p=p_*}.$$

In the case of linear degradation [$f(p) = p$, $h(m) = m$], we have $u = s = 1$. Small or even negative values of the slopes s and u generally denote strongly nonlinear degradation mechanisms [2,23], including saturation.

To assess whether Eqs. (3) can display oscillations, we have searched for parameter values where the fixed point specified by (4) loses stability to a periodic solution via a Hopf bifurcation (i.e., a pair of conjugate eigenvalues of the linearized problem cross the imaginary axis). For simplicity, we assume perfect repression ($\mu = 0$) and a large threshold P_0 ($\alpha \sim 0$). Under this approximation, the Routh-Hürwitz stability criterion [25] indicates that a Hopf bifurcation occurs when the quantity

$$\mathcal{H} = \sigma + (-\delta\lambda v g_*^2 + \sigma^2)\tau + \gamma\sigma\tau^2 \quad (5)$$

crosses zero to become negative, where $\tau = g_*/\theta$ is the gene response time, and the sum $\sigma = \delta s + u$ and product $\gamma = \delta s u$ are symmetric functions of degradation rates δs and u . At bifurcation, the oscillation period is $\tau_{\text{osc}} = 2\pi\sqrt{\tau/(\sigma + \gamma\tau)}$ where σ and γ satisfy $\mathcal{H} = 0$. Cooperativity essentially change the feedback strength from $\delta\lambda$ to $\delta\lambda v$; thus, we assume for simplicity single-protein regulation [$c(p) = p$, $v = 1$] thereafter.

Equation (5) shows that a strong feedback destabilizes the system while high degradation rates (large σ and γ) tend to stabilize it. In the single-protein case, $\mathcal{H} > 0$ when both protein and RNA are linearly degraded and no oscillations occur. Conversely, when protein and RNA degradations are completely saturated ($s = u = \sigma = 0$), $\mathcal{H} = -\delta\lambda g_*^2 \tau < 0$, indicating that oscillations then appear systematically. The behavior in intermediate cases depends on the value of the response time τ .

In the classical case $\tau = 0$, $\mathcal{H} = \sigma$ and oscillations appear only for $\sigma < 0$. It is indeed known that negative effective degradation rates can lead to oscillations [2,23]. We thus restrict ourselves to showing that at finite τ , oscillations can occur for $u, s > 0$. More precisely, we want to understand how oscillations arise away from the saturated cases $u = 0$ or $s = 0$. To this end, we use the geometric slope average $\nu = \sqrt{us} = \sqrt{\gamma/\delta}$ as an index ($\nu = 1$ in the linear case), seeking to determine the maximum value of ν at which oscillations can be observed, and for which values of τ this extremum is achieved.

The quantities σ and γ play complementary roles. The total degradation rate σ controls instability onset for small to moderate τ . Moreover, Eq. (5) indicates that when $\sigma, \gamma > 0$, a necessary condition for oscillations is

$$\sigma < \sigma_c = g_* \sqrt{\delta \lambda} = \frac{g_* \sqrt{2}}{t_{sw}}, \quad (6)$$

where t_{sw} is the time during which a fully active gene synthesizes the amount of protein corresponding to half-repression threshold. The degradation rate product γ is relevant only for large τ , blocking oscillations if it is too large. In particular, $\gamma = 0$ guarantees the onset of oscillations for sufficiently large τ whenever (6) holds. For $u, s > 0$ and a given value of σ , γ can take any value between $\sigma^2/4$ and 0 depending on whether the two degradation rates are equal or completely unbalanced, one being equal to zero and the other to σ . We use below $\epsilon = 2\sqrt{\gamma}/\sigma \in [0, 1]$ as a balance indicator.

Remarkably, we note that under the rescaling

$$\sigma = \sigma_c \Sigma, \quad \gamma = \sigma_c^2 \left(\frac{\epsilon \Sigma}{2}\right)^2, \quad \tau = \frac{T}{\sigma_c}, \quad (7)$$

the oscillation condition can be rewritten without explicit parameter dependence

$$\mathcal{H}_\epsilon(\Sigma, T) = \Sigma \times \left[\frac{\epsilon^2 \Sigma^2}{4} T^2 + \left(\Sigma - \frac{1}{\Sigma} \right) T + 1 \right] < 0 \quad (8)$$

and defines a series of curves $\Sigma_\epsilon(T)$ such that given a balance index ϵ and a response time T , oscillations are found for $\Sigma \leq \Sigma_\epsilon(T)$. Figure 2 shows the limit curves $\Sigma_1(T)$ and $\Sigma_0(T)$ which are important to understand the bifurcation diagram: regardless of the value of ϵ , the circuit always (never) oscillates when $\Sigma < \Sigma_1(T)$ [$\Sigma > \Sigma_0(T)$]. To support our analysis, we have searched the parameter space of Eqs. (3) for oscillatory behavior for $\alpha, \mu \neq 0$, assuming for definiteness allosteric protein degradation and Michaelis-Menten RNA degradation

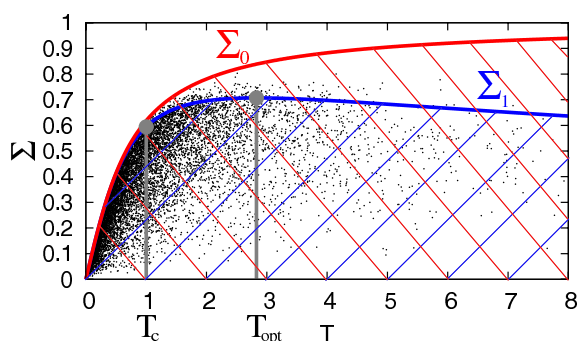


FIG. 2 (color online). Bifurcation diagram of Eqs. (3) in the (Σ, T) plane according to (8). A system with balance index ϵ oscillates for $\Sigma < \Sigma_\epsilon(T)$ (hatched areas). Black dots indicate oscillating parameter sets of (3), with $\theta, \delta \in [10^{-1}, 10]$, $\theta/\alpha \in [10, 1000]$, $\lambda \in [1, 10^3]$, $\lambda/\mu \in [10, 10^3]$, assuming the degradation mechanisms (9) with $a \in [10^{-6}, 1]$, $\kappa, \chi \in [1, 100]$. $T_c = 1$ is the time scale at which transcriptional dynamics cannot be neglected, $T_{opt} = 2\sqrt{2}$ is the location of the maximum of $\Sigma_1(T)$.

$$f(p) = \frac{p \times (a + p/\kappa)}{a + 2a(p/\kappa) + (p/\kappa)^2}, \quad h(m) = \frac{\chi m}{\chi + m}. \quad (9)$$

Points in the (Σ, T) plane associated with oscillating parameter sets are shown as black dots in Fig. 2. Agreement is excellent: all dots are below the $\Sigma_0(T)$ curve and the few significantly above $\Sigma_1(T)$ have one small degradation rate. We are thus confident that our analysis allows us to understand the behavior of Eqs. (3).

Two regions can be distinguished in Fig. 2. For $T < 1$, the instability threshold $\Sigma_\epsilon(T)$ is practically independent of ϵ and increases rapidly with T . For small T , $\Sigma_\epsilon(T) \sim T$ (thus, the oscillation criterion is $\sigma < \sigma_c^2 \tau$). In the $T > 1$ region, $\Sigma_\epsilon(T)$ reaches its maximum value $\Sigma_m(\epsilon) = 1/\sqrt{\epsilon + 1}$ at $T = T_m(\epsilon) = 2\sqrt{\epsilon + 1}/\epsilon$, and then decreases as T^{-1} for $T \rightarrow \infty$, except for $\epsilon = 0$ where it monotonically increases towards $\Sigma = 1$. At fixed Σ , oscillations are thus found in a finite range of T , which widens gradually, and is eventually infinite, as $\epsilon \rightarrow 0$.

In the oscillation region, the index $\nu \sim \epsilon \Sigma$ measuring distance from saturated degradation reaches its maximal value $\nu_{opt} = g_* \sqrt{\lambda/8}$ for $T = T_{opt} = 2\sqrt{2}$ and $\epsilon = 1$, at the maximum of the $\Sigma_1(T)$ curve. Our analysis thus unveils a resonance phenomenon in the dynamics of a self-regulated gene with dynamical transcription rate: this circuit bifurcates most easily to periodic behavior, or more generally is least stable, at a finite value of the gene relaxation time given by $\tau_{opt} = 2\sqrt{2}\tau_c$, where

$$\tau_c = \frac{1}{g_* \sqrt{\delta \lambda}} = \frac{1}{\sigma_c} = \frac{t_{sw}}{g_* \sqrt{2}} = \delta_M \times \frac{1}{g_*} \sqrt{\frac{P_0}{\lambda_0 \beta_0}}. \quad (10)$$

The quantity τ_c gives the time scale at which dynamical behavior departs from the fast regulation case. Because g_* is determined by (4), computing precise lower bounds on τ_c with (10) requires specifying the degradation mechanisms. Fixing $\lambda_0 = \beta_0 = 10 \text{ mn}^{-1}$, $P_0 = 100$, and $g_* = 0.5$, provides an estimate $t_c = \tau_c/\delta_M = 2 \text{ mn}$ which is not unrealistically larger than typical gene induction times. At resonance, the oscillation period is $\tau_{osc} = 4\pi\sqrt{2/3}\tau_c$ (corresponding to about 20 mn in the example above).

Since a common interpretation of the diagram of Fig. 1 is that there are only two gene states (bound or unbound) [26], one may wonder whether our deterministic analysis is relevant. If g is viewed as a temporal average of gene activity, our results are valid when the response time τ is small compared to the oscillation period so that there are many binding or unbinding events by cycle [27]. Moreover, transcription is a complex process involving a number of distinct steps [4], and Eq. (1a) is the simplest way to model memory effects arising from cooperativity in the transcription machinery [6].

Anyhow, we now show that even when G is a stochastic variable jumping between 0 and 1, our main result still

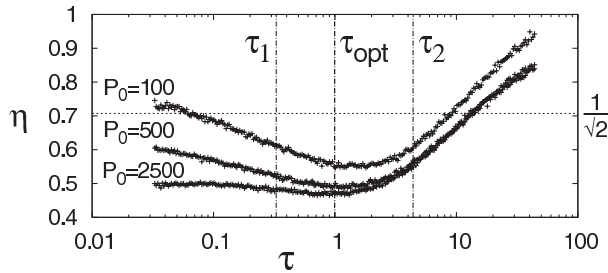


FIG. 3. Coefficient of variation of interpeak time intervals η vs mean residence time τ for a stochastic simulation of (3) and (9) rescaled to original variables G, P, M at three values of threshold P_0 (from top to bottom, $P_0 = 100, 500, 2500$). Parameter values correspond to $\delta = 1, \theta = 1, \alpha = \theta/P_0, \lambda = 21.54, \mu = 0.085, \chi = 95.5, \kappa = 21.68, a = 10^{-6}$. τ_1 and τ_2 are the boundaries of the deterministic oscillation domain.

holds: there is a time scale near τ_{opt} at which oscillations are enhanced. To this end we have carried out stochastic simulations of the reaction network of Fig. 1 using the Gillespie algorithm [28], varying the response time τ at fixed P_0 . Instead of regular oscillations, a sequence of irregularly spaced peaks in protein concentration is observed. A natural question is then whether protein peaks occur more regularly at parameter values where the deterministic model oscillates, in particular, when $\tau = \tau_{\text{opt}}$.

We define interpeak times as the time intervals Δt between two crossings of $P = 1.2P_{\text{avg}}$ separated by at least one crossing of $P = 0.8P_{\text{avg}}$, with P_{avg} the mean protein level (thus imposing a minimum amplitude of 40%). Their distribution is characterized by the coefficient of variation $\eta = \frac{\sigma_{\Delta t}}{\langle \Delta t \rangle}$. A typical variation of η with τ in our system is shown in Fig. 3. It definitely suggests that the deterministic analysis remains relevant in the stochastic regime, since there is clearly a time scale near τ_{opt} where interpeak time fluctuations are minimal.

In conclusion, we have shown that a nontrivial transcriptional dynamics can destabilize a self-regulated gene. Although it is known that strongly nonlinear degradation mechanisms can induce oscillations in this system, we observe a resonancelike effect such that a much weaker nonlinearity is required when the gene response time matches a characteristic time. Its expression can be computed analytically, which allows us to identify the parameter regions where this effect cannot be neglected. Stochastic simulations confirm the relevance of this time scale in the dynamics of the self-regulated gene. This shows that transcriptional dynamics is a possible source of oscillatory behavior besides other deterministic [1–3,13,14,18–20,23] and stochastic [29,30] effects.

A natural question is whether our conclusions remain valid when more detailed transcriptional mechanisms or multiple sources of delay are taken into account. It can be shown that adding an explicit delay to Eqs. (3) only further destabilizes the circuit and thus does not essentially interfere with the effect of a gene response delay. Future studies should thus focus on cases where the transcriptional dynamics is more complex and features more than one limiting step.

*marc.lefranc@univ-lille1.fr

- [1] A. Goldbeter, *Biochemical Oscillations and Cellular Rhythms* (Cambridge University Press, Cambridge, England, 1996).
- [2] C.P. Fall, E.S. Marland, J.M. Wagner, and J.J. Tyson, *Computational Cell Biology* (Springer, New York, 2002).
- [3] G. Tiana *et al.*, *Phys. Biol.* **4**, R1 (2007).
- [4] X. Darzacq *et al.*, *Nat. Struct. Mol. Biol.* **14**, 796 (2007).
- [5] I. Golding *et al.*, *Cell* **123**, 1025 (2005).
- [6] J.R. Chubb *et al.*, *Curr. Biol.* **16**, 1018 (2006).
- [7] A. Raj *et al.*, *PLoS Biol.* **4**, e309 (2006).
- [8] J.E.M. Hornos *et al.*, *Phys. Rev. E* **72**, 051907 (2005).
- [9] J.S. van Zon *et al.*, *Biophys. J.* **91**, 4350 (2006).
- [10] A. Lipshat *et al.*, *Phys. Rev. Lett.* **96**, 188101 (2006).
- [11] A.M. Walczak, J.N. Onuchic, and P.G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18926 (2005).
- [12] M.J. Morelli *et al.*, *Biophys. J.* **94**, 3413 (2008).
- [13] A. Goldbeter, *Proc. R. Soc. B* **261**, 319 (1995).
- [14] J.-C. Leloup, D. Gonze, and A. Goldbeter, *J. Biol. Rhythms* **14**, 433 (1999).
- [15] B.C. Goodwin, *Adv. Enz. Reg.* **3**, 425 (1965).
- [16] J.S. Griffith, *J. Theor. Biol.* **20**, 202 (1968).
- [17] R.D. Bliss, P.R. Painter, and A.G. Marr, *J. Theor. Biol.* **97**, 177 (1982).
- [18] M.H. Jensen, K. Sneppen, and G. Tiana, *FEBS Lett.* **541**, 176 (2003).
- [19] N.A.M. Monk, *Curr. Biol.* **13**, 1409 (2003).
- [20] J. Lewis, *Curr. Biol.* **13**, 1398 (2003).
- [21] H. Hirata *et al.*, *Science* **298**, 840 (2002).
- [22] L.G. Morelli and F. Jülicher, *Phys. Rev. Lett.* **98**, 228101 (2007).
- [23] J.J. Tyson *et al.*, *Biophys. J.* **77**, 2411 (1999).
- [24] P. François and V. Hakim, *Phys. Rev. E* **72**, 031908 (2005).
- [25] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic Press, San Diego, 2000).
- [26] T.B. Kepler and T.C. Elston, *Biophys. J.* **81**, 3116 (2001).
- [27] D.B. Forger and C.S. Peskin, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 321 (2005).
- [28] D.T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [29] A. Loinger and O. Biham, *Phys. Rev. E* **76**, 051917 (2007).
- [30] R. Bllossey, L. Cardelli, and A. Phillips, *Human Science Frontier Program J.* **2**, 17 (2008).

Annexe B

Annexe au chapitre 3 : A robust two-gene oscillator at the core of *Ostreococcus tauri* circadian clock

paru dans :

« Chaos : An Interdisciplinary Journal of Nonlinear Science, volume 20, numéro 4, page 12 en Novembre 2010 »

Auteurs :

Pierre-Emmanuel Morant, Quentin Thommen, Benjamin Pfeuty, Constant Vandermoëre, Florence Corellou, Francois-Yves Bouget et Marc Lefranc

A robust two-gene oscillator at the core of *Ostreococcus tauri* circadian clock

Pierre-Emmanuel Morant,¹ Quentin Thommen,¹ Benjamin Pfeuty,¹ Constant Vandermoere,¹ Florence Corellou,² François-Yves Bouget,² and Marc Lefranc^{1,a)}

¹Laboratoire de Physique des Lasers, Atomes, Molécules, Université Lille 1, CNRS, F-59655 Villeneuve d'Ascq, France and Institut de Recherche Interdisciplinaire, Université Lille 1, CNRS, F-59655 Villeneuve d'Ascq, France

²Laboratoire d'Océanographie Microbienne, Université Pierre et Marie Curie-Paris 6, CNRS, Observatoire Océanologique de Banyuls/Mer, 66650 Banyuls/Mer, France

(Received 15 November 2010; accepted 1 December 2010; published online 30 December 2010)

The microscopic green alga *Ostreococcus tauri* is rapidly emerging as a promising model organism in the green lineage. In particular, recent results by Corellou *et al.* [Plant Cell **21**, 3436 (2009)] and Thommen *et al.* [PLOS Comput. Biol. **6**, e1000990 (2010)] strongly suggest that its circadian clock is a simplified version of *Arabidopsis thaliana* clock, and that it is architected so as to be robust to natural daylight fluctuations. In this work, we analyze the time series data from luminescent reporters for the two central clock genes *TOC1* and *CCA1* and correlate them with microarray data previously analyzed. Our mathematical analysis strongly supports both the existence of a simple two-gene oscillator at the core of *Ostreococcus tauri* clock and the fact that its dynamics is not affected by light in normal entrainment conditions, a signature of its robustness. © 2010 American Institute of Physics. [doi:10.1063/1.3530118]

In order to anticipate periodic environmental changes induced by Earth rotation, many organisms have evolved a circadian clock, a genetic oscillator which generates biochemical rhythms with a period around 24 h. Exact synchronization with the day/night cycle requires that one or more clock components sense daylight (for example, a protein degrades faster in the light). However, daylight intensity can highly fluctuate from day to day, or during the day, due to environmental factors such as sky cover. The question then arises as to how circadian clocks can keep time without being continuously reset by the signal that should entrain them. A mathematical analysis of *Ostreococcus tauri* clock, whose molecular basis has been identified recently,¹ has unveiled a simple and elegant mechanism which exploits the dynamical properties of the core clock oscillator to make it robust to daylight fluctuations.² When the clock is on time, coupling to light is activated precisely when the oscillator does not respond to external perturbations, making it blind to light and its fluctuations. If the clock has drifted and needs resetting, however, light affects the oscillator in a different part of its cycle, where it reacts so as to recover the entrainment phase. In this work, we provide strong evidence for the presence of a robust two-gene oscillator at the core of *Ostreococcus* clock by showing that a minimal light-independent model can reproduce very accurately microarray data and luminescent reporter data recorded in different experiments.

I. INTRODUCTION

Biochemical oscillations are widespread biological phenomena involved in many important cellular processes such

as signaling, development, motility, or metabolism.^{3,4} Understanding how such a simple dynamical pattern has been implemented repeatedly to support a great diversity of biological functions is appealing for scientists seeking to uncover unifying principles.⁵ By identifying the molecular machinery behind cellular rhythms, molecular biology has first laid the groundwork for the development of strategies toward their comprehensive understanding.⁶ In a second stage, how a collective dynamics can emerge in networks of interacting molecular actors and how it can be harnessed robustly has been under focus. Besides the design of synthetic genetic circuits with specific oscillatory abilities,^{7,8} a common strategy to gain insight into this question has been to construct quantitative dynamical models constrained by experiments.^{3,9–11} Indeed, the nonlinear dynamical behaviors that underlie oscillations can only be fully captured through a mathematical description.

However, the ever-growing amount of experimental data obtained with genetic transformations and real-time monitoring presents extraordinary challenges to modelers. How to design relevant mathematical models that not only reproduce the data but give insight into the architecture of a biological system? How to avoid pitfalls such as overfitting which can ascribe biological meaning to experimental artifacts? As we hope to illustrate here, it is important to combine careful data analysis and minimal modeling, and to check the consistence of the different data sets at all stages of model building.

Circadian clocks are systems of choice for quantitative studies of the function and design of biochemical oscillators.¹² They operate in many species to keep track of the most regular environmental constraint, the alternation of daylight and darkness caused by Earth rotation, so as to finely control the cellular physiology accordingly.¹³ This basic function is vital in many organisms such as plants, which

^{a)}Electronic mail: marc.lefranc@univ-lille1.fr.

need to timely coordinate their photosynthesis, and thus their growth and division, to daily changes in light intensity.^{14–16} However, a precise entrainment of circadian rhythms to the diurnal cycle is potentially challenged by many sources of variability including molecular fluctuations^{17,18} and temperature variations,^{19,20} but also fluctuations of daylight intensity from day to day or during the day, due to environmental factors such as sky cover.^{2,21–23} To account for the remarkable ability of circadian oscillators to run autonomously and to be precisely and robustly entrained, experimental efforts aimed at unraveling their complex architecture^{24–26} have motivated studies trying to adjust mathematical models to experimental data,^{2,27–33} an approach that is increasingly necessary as models become more complex, featuring generally several feedback loops.

In this paper, we use such a quantitative modeling approach to investigate the dynamics of the circadian rhythms of the smallest free-living eukaryote known to date, *Ostreococcus tauri*. This microscopic green alga displays a very simple cellular organization and a small and compact genome.^{34–36} Very recently, the molecular basis of its circadian clock has been extensively characterized by Corellou *et al.*, who carried out an extensive work of genetic transformation, leading to transcriptional and translational fusion lines allowing one to monitor transcriptional activity and protein dynamics in living cells.^{1,37} Their results point to a core architecture comprising two genes, similar to *Arabidopsis* central clock genes *TOC1* and *CCA1*.³⁸ These two genes display rhythmic expression both under light/dark alternation and in constant light conditions. Thus, the unicellular green alga *Ostreococcus tauri* has emerged as a promising organism model to study the circadian rhythm in single photosynthetic eukaryotic cell combining experimental and modeling approaches.

The goal of this modeling study is to check carefully whether experimental data obtained through various channels support the hypothesis that the circadian clock of *Ostreococcus Tauri* contains a simple two-gene transcriptional loop serving as a core oscillator,^{1,2} and that the dynamics of this oscillator is not affected by light in normal entrainment conditions.² An important point of the present work is the simultaneous adjustment of microarray data and luminescent reporter data recorded in two different experiments^{1,15} under different conditions. While this may not be sensible for a general system whose parameters will typically change with experimental conditions, it is expected here that the core clock oscillator is robust to environmental changes. As advocated by Thommen *et al.*,² it should thus deliver similar signals in different experiments carried out with the same photoperiod. Therefore, being able to reproduce with the same model signals recorded with different techniques in different experiments provides strong evidence that the core *Ostreococcus* clock oscillator is indeed robust.

Before adjusting a minimal model to experimental data, a careful data analysis of both microarray and luminescence time series has been performed. It allowed us to detect and remove an experimental artifact that would otherwise have prevented an optimal adjustment and thus would have falsely called for a more complex model. The combined use of

microarray data and luminescence reporter data provided a unique opportunity to calibrate the latter with respect to the former, given that very little is known on luciferase dynamics in *Ostreococcus*. This is all the more important as an extensive collection of experimental data has been acquired with these reporters and will form the basis of future quantitative studies of *Ostreococcus* clock.

We not only found that a simple two-gene transcriptional feedback loop model reproduces perfectly the CCA1 and TOC1 transcript and protein profiles obtained from data analysis, but that it does so with no model parameter depending on light intensity whereas the data have been recorded under light/dark alternation. This confirms our previous observation based only on microarray data of limited time resolution.² As we have proposed previously, this counterintuitive finding indicates that the coupling to light does not influence the oscillator in normal entrainment conditions, shielding it from daylight fluctuations.² Besides confirming the two-gene loop hypothesis and the oscillator robustness, this work also identifies unambiguously the mechanistic origin of oscillation in the transcriptional negative feedback loop involving CCA1 and TOC1, delayed by the saturated degradation of CCA1 mRNA and TOC1 protein.

II. DATA ANALYSIS

In this section, we first present the experimental data and explain how we determine mRNA and protein concentrations from them. We then discuss the simplest model of the TOC1-CCA1 negative feedback loop, which consists of four differential equations describing the time evolution of the TOC1 and CCA1 mRNA and protein concentrations. To match the numerical time profiles generated by this model, we need to reconstruct target profiles from the experimental data. RNA profiles will be interpolated from microarray data whereas protein profiles will be extracted from luminescence time series in the form of a Fourier series describing the dynamics at long time scales. Both types of profile will be characterized by their times of passage at 20% and 80% of the oscillation amplitude, which provide a uniform characterization of the two types of data considered here. Model adjustment will be carried out by minimizing discrepancies between the experimental and numerical passage times.

A. Experimental data

The microarray data used here come from the study in Ref. 15. They were recorded at 3 h time intervals under 12:12 light/dark alternation. The mRNA time profiles that will serve as targets for model adjustment are the same as in our previous study,² where we showed that they could be accurately approached by solutions of a simple set of differential equations, an indication of the high quality of these data. The fact that microarray data reflect mRNA level without ambiguity was also checked by quantitative RT-PCR (reverse transcription polymerase chain reaction).

The luminescent reporter data used here are those presented by Corellou *et al.* in Ref. 1 to provide evidence of the central role of the TOC1 and CCA1 genes in *Ostreococcus*

clock. They had been recorded at 1 h intervals under a 12:12 light/dark cycle from two types of transgenic cell lines.

In transcriptional fusion lines pTOC1:luc and pCCA1:luc, the sequence inserted into the genome consists of the promoter of one of the two clock genes fused to the coding sequence of firefly luciferase,¹ so that TOC1 or CCA1 transcriptional activity drives luciferase expression. Luciferase catalyzes the transformation of the substrate luciferin into oxyluciferin, in which one photon is emitted.³⁹ The luminescence signal provides information about the quantity of luciferase synthesized and thus on the transcriptional activity of the promoters.

In translational fusion lines TOC1:luc and CCA1:luc, both the coding sequences of one of the clock proteins (TOC1 or CCA1) and of luciferase are fused to the promoter,¹ so that a fusion protein combining the clock protein with luciferase is synthesized by the additional gene with a transcriptional activity similar to the original clock gene. In this case, the luminescence signal also provides information about the protein dynamics, and in particular on its degradation kinetics. It is usually assumed that luciferase remains complexed and inactive after the reaction and does not recover its activity before being degraded. Thus, two limiting cases can be considered depending on whether the luciferase reaction time (defined as the average time from synthesis to photon emission) is much shorter or much longer than the protein lifetime.

In the former case, luciferase reacts immediately after being synthesized. Since only freshly made proteins then contribute to the luminescence signal, the latter is obviously proportional to the protein synthesis rate, hence to cytoplasmic fusion mRNA concentration. If we assume for simplicity that the kinetic constants (synthesis and degradation rates) of the fusion proteins TOC1:luc and CCA1:luc and of their mRNA are similar to that of their native counterparts, then the concentrations of the former are proportional to the concentrations of the latter. In this case, the luminescence signal tracks mRNA concentration.

In the opposite case where the probability of photon emission by luciferase is very low, and with the same assumption of identical constants for fusion and native molecular actors, it is easily seen that the luminescence signal is proportional to protein concentration. In this limit, indeed, the photon emission probability is constant throughout protein lifetime, thus luminescence intensity is proportional to the number of TOC1:luc or CCA1:luc fusion proteins, which is in turn proportional to the number of native clock proteins. In this case, the luminescence signal can be used as an indicator of protein concentration. In the general case, it should be intermediate between the RNA and protein time courses.

A time lag of at least 2 h between the maxima of RNA concentration as indicated by microarray data and the maxima of translational fusion lines can be observed, which indicates that photon emission probability is low and that luciferase lifetime is large compared to other protein lifetimes. This is consistent with recent experiments in which translation was blocked in *Ostreococcus* cell cultures with emetine dihydrochloride³⁷ and time evolution of TOC1:luc luminescence was monitored. It was observed that the signal

would slowly decay over more than 12 h, implying that the luciferase reaction time is at least of this order. Thus, we will use in the following translational fusion line signals as indicators of protein concentrations, which allows us to keep our mathematical model as simple as possible.

B. Model of the TOC1-CCA1 oscillator used for data adjustment

The minimal model for the TOC1-CCA1 transcriptional feedback loop, where TOC1 activates CCA1 and CCA1 represses TOC1, consists of the following four ordinary differential equations (ODEs):

$$\dot{M}_T = \mu_T + \frac{\lambda_T}{1 + (P_C/P_{C0})^{n_C}} - \delta_{M_T} \frac{K_{M_T} M_T}{K_{M_T} + M_T}, \quad (1a)$$

$$\dot{P}_T = \beta_T M_T - \delta_{P_T} \frac{K_{P_T} P_T}{K_{P_T} + P_T}, \quad (1b)$$

$$\dot{M}_C = \mu_C + \frac{\lambda_C (P_T/P_{T0})^{n_T}}{1 + (P_T/P_{T0})^{n_T}} - \delta_{M_C} \frac{K_{M_C} M_C}{K_{M_C} + M_C}, \quad (1c)$$

$$\dot{P}_C = \beta_C M_C - \delta_{P_C} \frac{K_{P_C} P_C}{K_{P_C} + P_C}. \quad (1d)$$

Equation (1) describes the time evolution of mRNA concentrations M_C and M_T and protein concentrations P_C and P_T for the CCA1 and TOC1 genes, respectively, as they result from mRNA synthesis regulated by the other protein, translation, and enzymatic degradation. TOC1 transcription rate varies between μ_T at infinite CCA1 concentration and $\mu_T + \lambda_T$ at zero CCA1 concentration according to the usual gene regulation function with threshold P_{C0} and cooperativity n_C . Similarly, CCA1 transcription rate is μ_C (respectively, $\mu_C + \lambda_C$) at zero (respectively, infinite) TOC1 concentration, with threshold P_{T0} and cooperativity n_T . TOC1 and CCA1 translation rates are β_T and β_C , respectively. For each species X , the Michaelis-Menten degradation term is written so that δ_X is the low-concentration degradation rate and K_X is the saturation threshold.

A more detailed model would take into account compartmentalization, with each actor having separate nuclear and cytoplasmic concentrations, as well as the fact that the luminescence signal is linked to a third gene artificially inserted in the genome, with the kinetic constants of its mRNA and protein possibly different from those of its native counterpart. With this model, it is the profile predicted for the total fusion protein concentration that would be adjusted to the experimental time series rather than the native protein profile. However, the important point is that such a biochemically detailed model taking into account all native and inserted molecular actors reduces to Eq. (1) in the limiting case where fusion proteins and mRNA have the same kinetic constants as the native molecules (e.g., TOC1:luc and TOC1 degradation rates are equal), luciferase reaction time is large, and nucleocytoplasmic transport is fast. When the minimal model already adjusts the data with excellent accuracy, as

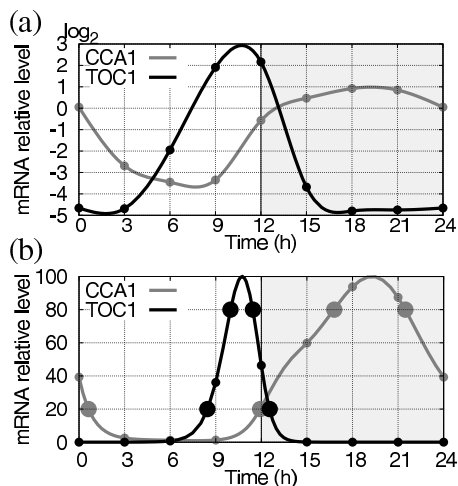


FIG. 1. Construction of RNA target profiles as functions of Zeitgeber time (ZT) describing phases within the dark/light cycle, with time ZT0 that corresponds to dawn and time ZT12 to dusk. Interpolating curves going through data points (dots) are represented. (a) Interpolated curves from microarray data using cubic splines. (b) The exponentials of the interpolating curves were computed to obtain a smooth approximation of mRNA profiles. Large dots indicate passages through levels corresponding to 20% and 80% of maximum amplitude and serve as target points for adjustment.

will be the case here, there is no point in using the sophisticated model, which can only fit the data better. In fact, its higher flexibility could lead to overfit the data and ascribe incorrectly biological meaning to experimental artifacts.

As in Ref. 2, the free-running period (FRP) of the uncoupled oscillator is chosen equal to 24 h, which was the mean value observed in experiments.¹ In this case, this oscillator is adjusted to experimental data without modulation since our goal in this work is to show that the average oscillation does not carry any signature of coupling to light. As a control, we will also consider adjustment to oscillators of FRP 23.8 and 25 h under day conditions. In these two cases, a small modulation of some parameters is required to achieve frequency locking. To keep the FRP fixed during adjustment, we rescale kinetic parameters after each optimization step so that the period of the free-running limit cycle matches the desired value. This subsequently applies to the parameter set which the adjustment converges to.

C. Reconstruction of target profiles and determination of passage times

Experimental RNA target profiles were estimated from the same microarray data points as used in Ref. 2. Due to the high quality of data, noise reduction was not required. However, the relatively low temporal resolution (one sample every 3 h) made it difficult to estimate accurately passage times and positions of expression peaks. Thus, the profiles were obtained by interpolating between data points. Time series in the logarithm of RNA concentration showed a very smooth behavior and were well approximated by simple cubic splines (Fig. 1). These interpolating curves were thus considered to provide an optimal approximation to the variations of the logarithm of mRNA concentration, and mRNA linear target curves were generated by exponentiation of the interpolating curves.

That the largest data point is less than half of the maximum of the reconstructed TOC1 mRNA profile should not be surprising. In fact, this profile is the most probable one given the data points if we assume sufficient smoothness, or equivalently that the same dynamics acts over the 24 h cycle. Any other curve would imply fast, transient processes not linked to the TOC1-CCA1 loop. Since RNA concentration rises from almost zero at ZT6 (Zeitgeber Time 6, i.e., 6 h after dawn) to almost the maximum level measured at ZT9 and decays from the same level at ZT12 to almost zero at ZT15, it is indeed obvious that the true maximum located between ZT9 and ZT12 must be much higher than the two largest data points. Incorrectly assuming that the largest data point is the profile maximum would in fact bias the analysis.

We checked that interpolating either the CCA1 mRNA concentrations or their logarithms leads to two almost superimposing profile curves, showing that both sequences of values provide the same information. However, the profile interpolated directly from TOC1 mRNA concentrations rather than from their logarithms was highly nonphysical, with interpolated concentrations becoming clearly negative on each side of the expression peak. This is a direct consequence of the fact that only two data points are well above the zero level. Implicit in the interpolation procedure is the reconstruction of successive time derivatives of a function from its values at different points. The long sequence of almost zero data points makes the correspondence between function values and time derivatives nearly singular, so that the reconstruction is highly unstable. During the time interval near zero, all time derivatives up to high order appear to be zero, then jump to high values at the beginning and the end of the expression peak. Since the interpolation error is proportional to the maximum value of the fourth time derivative in the interval, the resulting approximation is poor.

The series of logarithms of TOC1 mRNA concentrations does not have this problem and allows optimal reconstruction of the TOC1 mRNA profile. The fact that the logarithm of a time series with long intervals near zero provides more information about the dynamics than the original time series has also been noted in the analysis of chaotic time series.^{40,41} Furthermore, it should be noted that many statistical analyses of microarray data use the logarithms of mRNA concentrations because they are more evenly distributed and provide more information.

Compared to a more usual method based on least-squares-fitting the data points (as used in Ref. 2), this interpolation procedure can only make adjustment more difficult, since it constrains tightly the shape of the profile. However, it is fully consistent with the good agreement found with a low-dimensional ODE set in our previous study.² The passage times were then determined as the times at which the interpolated profile would cross the appropriate level (Fig. 1).

Protein profiles were reconstructed from translational fusion luminescence signals used as indicators of protein concentrations, as discussed above. The reconstruction was more involved than for mRNA because these time series display significant amplitude variations from peak to peak across the

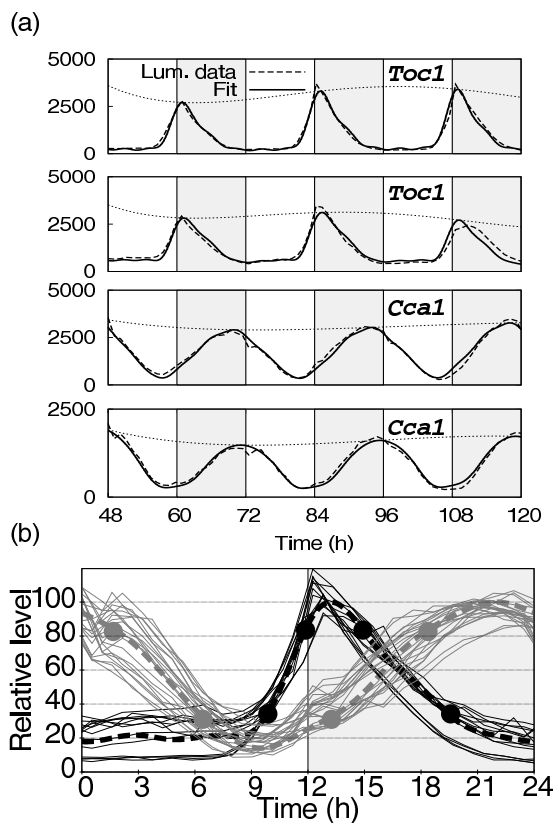


FIG. 2. Reconstruction of protein target profiles as functions of ZT describing phases within the dark/light cycle, with time ZT0 that corresponds to dawn and time ZT12 to dusk. (a) Luminescence time series (dashed lines) for individual wells are fitted by the product of a Fourier series of period 24 h with five harmonics by a slowly varying polynomial function of degree 4 (solid lines). The slowly varying envelope is also shown with a dotted line. From top to bottom, the first (last) two panels correspond to two TOC1:luc (CCA1:luc) translational fusion lines with different insertions in the genome. The different Fourier series are normalized to have the same maximum and are averaged. (b) The average Fourier series (dashed lines, black: TOC1, gray: CCA1) provides a smooth approximation to the cloud of individual line time profiles renormalized using the slowly varying polynomial function and wrapped around 24 h (thin solid lines). Large dots indicate passages through levels corresponding to 20% and 80% of maximum amplitude and serve as target points for adjustment.

experiment, which may be for instance linked to variations in the number of cells contributing to light emission or other experimental factors.

The first 48 recording hours, considered as transient, were discarded and traces were then least-squares fitted by the product $p(t)F(t)$ of a Fourier series $F(t) = \sum_{k=0}^N a_k \cos(2\pi kt/T + \phi_k)$, with $T=24$ h, by a low-order polynomial $p(t)$ accounting for slowly varying experimental conditions (Fig. 2). The order of the polynomial $p(t)$ was generally chosen to be 3 or 4, roughly equal to the number of periods in the fitted segment to avoid overfitting. The Fourier series $F(t)$ represents the average periodic biochemical oscillation associated with circadian oscillations while $p(t)$ models a slowly varying gain (due, for example, to a variable number of cells contributing to light emission).

Because we are interested in long lasting mechanisms sustaining the oscillation, the number N of harmonics of the Fourier series was fixed to 5 (an odd number, since the spectrum of the square wave forcing by light only contains odd

harmonics). We also tried slightly higher numbers of harmonics, with little difference in the results. Fourier fitting naturally smooths out acute responses at day/night and night/day transitions that can be seen in the raw signals and which may reflect rapid mechanisms involved in clock resetting. This separation of dynamical processes occurring on a 24 h time scale on the one hand, and confined to short time intervals on the other hand, must be understood with reference to our previous finding that the average dynamics of the TOC1-CCA1 oscillator is identical to that of a free-running oscillator, with coupling to light and resetting occurring transiently in specific time intervals.² Confirming this key observation with a careful analysis of the luminescence time series is one of the main goals of this paper. It will be reached if we find that the average protein profile does not carry any signature of a coupling to light.

The Fourier series obtained from the fits of all bioluminescence traces (possibly corresponding to different transgenic lines) recorded in the experiment were then averaged, and the average was used as surrogate for the protein temporal profile after suitable normalization. The resulting profile is shown in Fig. 2(b), along with individual luminescence traces from different wells, renormalized using the slowly varying polynomial function $p(t)$. The variability observed is partially due to the fact that a few different transgenic lines corresponding to different insertions in the genome were used. The times of passage at 20% and 80% of the amplitude were then determined, as well as the minimal level reached, for subsequent model adjustment.

We also tried fitting each luminescence trace to $p(t)F(t)+b$, where b represents a possible constant bias. Surprisingly, the fit was consistently better when b was equal to the minimum luminescence level. In other words, the Fourier fitting procedure tended to remove the floor level. Because this could indicate the existence of a bias in the luminescence level, we examined more closely the raw data, and found that indeed the zero luminescence level does not correspond to the zero protein level.

In Fig. 3, we show two individual traces monitoring luminescence intensities over time in two wells of the luminometer plate. These two wells contain genetically identical cell cultures monitored simultaneously in the same experiment. Besides an excellent overall reproducibility, the comparison of the two traces reveals the existence of a significant bias. Assuming that identical clocks run in different cells, and that there is a well-defined average number of fusion proteins per cell, the luminescence signals should be proportional to numbers of cells in each well. Therefore, the two signals should be approximately proportional to each other.

Contrary to this, we observe that both for TOC1:luc and CCA1:luc, the maxima of the two signals differ by a large factor, which remains roughly constant over time while minima almost coincide. If the two traces were proportional to each other, the minima should be in the same ratio as the maxima. The only simple explanation for the systematic coincidence of minima is that they correspond to a zero protein level, and that there is the same bias on the two time series. This bias can therefore be removed by subtracting the two traces, as shown in Fig. 3, where the difference curves pro-

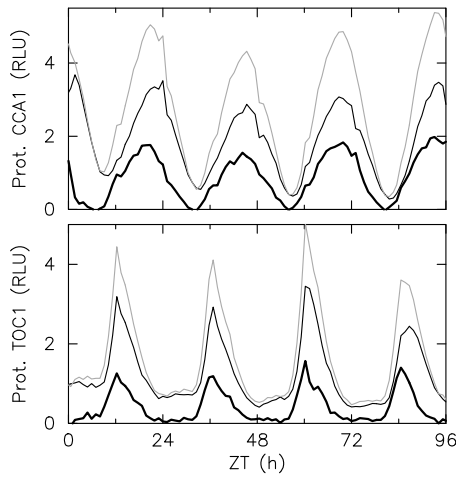


FIG. 3. Translational fusion lines data recorded under 12:12 LD alternation. Time zero corresponds to dawn. On the top panel (respectively, bottom panel), the time evolution of photon count (in relative luminescence units) of CCA1 (respectively, TOC1) translational fusion lines is drawn for two biological duplicates monitored at the same time in the same conditions (black and gray thin solid lines). Their difference is plotted as a thick black solid line.

vide a better estimation of the actual protein profiles than the original traces. Importantly, the CCA1 protein level is predicted to touch zero very shortly near ZT9, while the TOC1 protein level appears to stay near zero for a large interval of time between ZT21 and ZT8 (Fig. 3). Note that this bias is not constant in time but varies slowly so that the different minima correspond to different luminescence levels.

While the existence of an experimental bias in the luminescence level is strongly supported by the observations above, we have currently no explanation for it. In the following, we will therefore adjust two versions of the experimental profiles: (i) the Fourier series $F(t)$ determined directly from fitting $p(t)F(t)$ to the time series, as described above, as if there was no bias, and (ii) the same Fourier series with the floor level removed so that they touch zero, as with the difference curves in Fig. 3. More precisely, the target curves are then $F(t) - F_0$, where the $F(t)$ are obtained from fitting $p(t)F(t)$ and $F_0 = \min_t F(t)$. This will allow us to assess the influence of the bias on model fitting. Again, using a more sophisticated model of the floor level bias could only lead to a better adjustment, which will not be needed, and would only make sense with a better understanding of the physical origin of the bias.

D. Measuring goodness of fit by passage times

Adjusting experimental data to a mathematical model requires a score function quantifying the discrepancy between experimental and simulated profiles, to be minimized over parameter space. One difficulty is that microarray and luminescent reporter data are very different in nature and in particular have different uncertainties so that it does make little sense to combine their adjustment errors. Therefore, we used passage times as a unifying measure of fitting errors, determining for each temporal profile the times of passage at 20% and 80% of the interval between the minimal and the maximal values. We then compared passage times for the

experimental and numerical profiles, with the root mean square timing error serving as a goodness of fit. Another advantage of passage times is that they have a clear biological significance, with crossings of the 20% level bracketing the time interval with expression significantly above background level and the crossings with the 80% level bracketing the expression peak.

We denote by $\Delta_{20\uparrow}^X$ and $\Delta_{20\downarrow}^X$ (respectively, $\Delta_{80\uparrow}^X$ and $\Delta_{80\downarrow}^X$) the time passage error in minutes for the concentration of species X ($X = M_C, P_C, M_T, P_T$) at 20% (respectively, 80%) of the interval between minimal and maximal values when increasing and decreasing. For each species X , one can therefore define a quadratic error function equal to

$$Err(X) = (\Delta_{20\uparrow}^X)^2 + (\Delta_{80\uparrow}^X)^2 + (\Delta_{80\downarrow}^X)^2 + (\Delta_{20\downarrow}^X)^2. \quad (2)$$

The first and most natural score function is a rms error combining error for all species (mRNA and proteins),

$$S_{MP} = \sqrt{\frac{1}{16} \sum_{X \in M_C, P_C, M_T, P_T} Err(X)}. \quad (3)$$

In some cases, profiles with similar passage times differ by their floor level. We therefore introduce a floor level error $\Phi_X = \min(\hat{X}) / \max(\hat{X}) - \min(X^*) / \max(X^*)$, where \hat{X} is the numerical profile for species X and X^* is the experimental target for \hat{X} . We choose a ponderation such that a floor level error $\Phi_X = 0.05$ is equivalent to a passage time error of 60 min. The score function of this second scheme is

$$S_{MPF} = \frac{1}{\sqrt{18}} \sqrt{16(S_{MP})^2 + \sum_{X \in P_C, P_T} (\Phi_X / \Phi_0)^2}, \quad (4)$$

where $\Phi_0 = 0.05 / 60 = 1 / 1200$ is the relative floor level error considered equivalent to 1 min.

Finally, microarray and luminescence data were not only adjusted simultaneously but also separately in order to assess the coherence between the two types of data. To this end, we use the following score functions:

$$S_M = \sqrt{\frac{1}{8} \sum_{X \in M_C, M_T} Err(X)}, \quad (5)$$

$$S_P = \sqrt{\frac{1}{10} \sum_{X \in P_C, P_T} (Err(X) + (\Phi_X / 2\Phi_0)^2)}, \quad (6)$$

where the Φ_X error term is scaled to $2\Phi_0$ so that it has the same relative importance in the total rms error as in Eq. (4).

III. RESULTS: ADJUSTMENT OF MODEL TO RNA AND PROTEIN DATA

Figure 4 shows the numerical solutions of model (1) best adjusting the experimental data according to the score functions defined in Sec. II D, both without taking into account the floor level bias [Figs. 4(a)–4(d)] or with floor levels removed [Figs. 4(e)–4(h)]. The corresponding parameter values are given in Table I.

A first remark is that although a set of four passage times is a very crude description of a temporal profile, it turns out that numerical solutions and experimental profiles that have

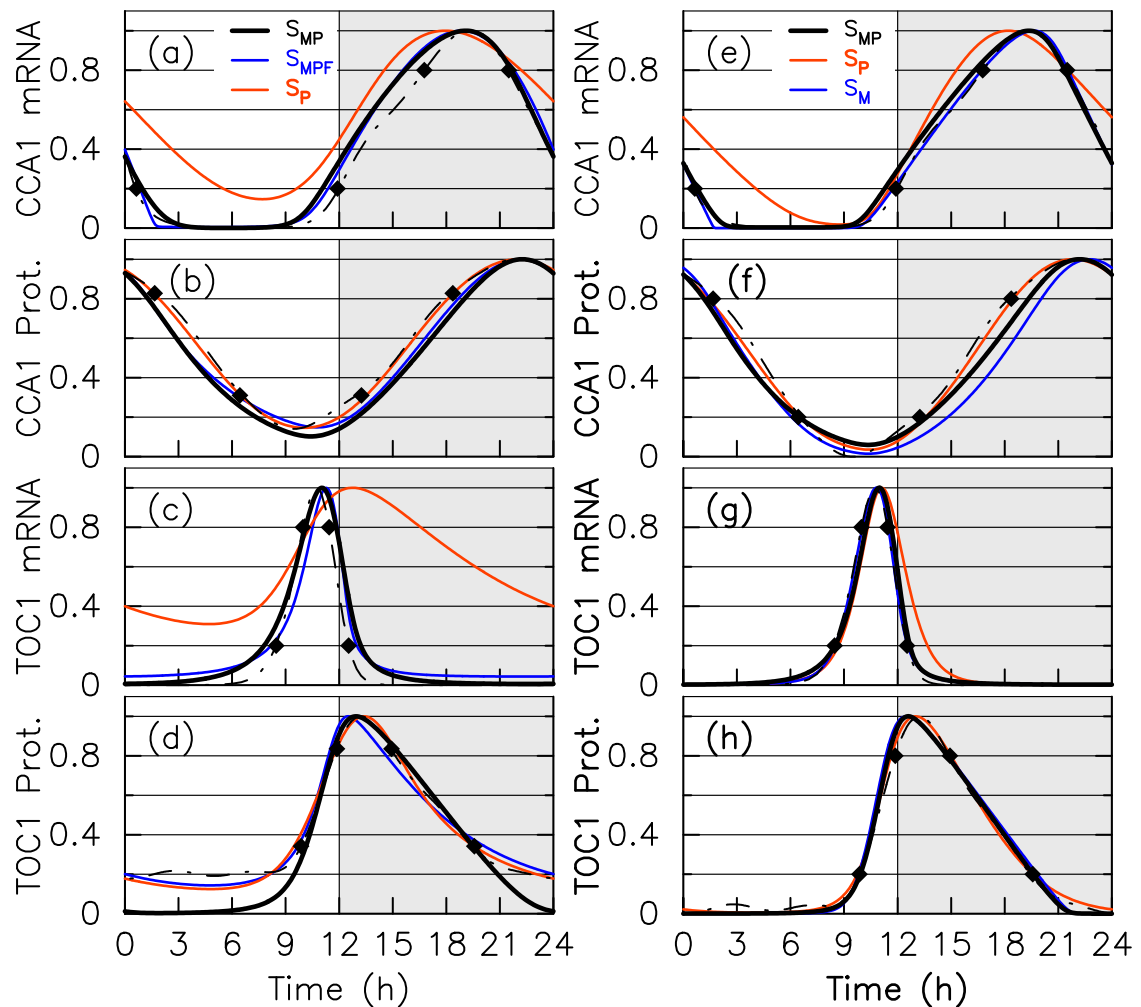


FIG. 4. (Color online) Adjustment of numerical solutions of the free-running model with FRP equal to 24 h (solid lines) to target curves (dot-dashed lines) using various score functions. Panels (a)–(d) in left column show the adjustment of mRNA and protein profiles without bias correction using score functions S_{MP} (black solid line), S_{MPF} (dark gray-blue thin solid line), and S_P (light gray or red thin solid line). See text for the definition of score functions. Panels (e)–(h) in right column show the adjustment of mRNA and protein profiles with protein floor level removed using score functions S_{MP} (black solid line), S_P (light gray-red thin solid line), and S_M (dark gray or blue thin solid line). [(a) and (e)] CCA1 mRNA. [(b) and (f)] CCA1 protein. [(c) and (g)] TOC1 mRNA. [(d) and (h)] TOC1 protein.

close passage times generally follow each other closely all over the day. This is consistent with the hypothesis that experimental data are well described by a low-dimensional ODE set. Second, the quality of the adjustment is globally very good given the simplicity of the model. Clearly, Fig. 4 shows that Eq. (1) can adjust simultaneously the protein profiles reconstructed from luminescence time series and the RNA profiles reconstructed from microarray data (score functions S_{MP} and S_{MPF}). Especially, removing the protein profile floor levels so as to correct for the detected experimental bias leads to impressive agreement between numerical and target curves, which almost superimpose onto each other.

It is interesting to look more closely at the main discrepancies in Figs. 4(a)–4(d). When only passage times are taken into account (score function S_{MP}), the nonzero floor level of the TOC1 protein is not reproduced. This issue can be addressed by using a score function that takes into account floor level error (score function S_{MPF}). In this case, the floor level agreement is improved at the cost of slightly advancing

the protein peak and of inducing a small floor level for TOC1 mRNA. Although small, the latter is definitely much higher than the value that can be accurately determined from microarray data (which estimate the logarithm of mRNA concentration). Last, the numerical solution obtained by adjusting the protein profiles alone (score function S_P) predicts mRNA profiles rather poorly (Fig. 4), which could suggest that the two types of data are inconsistent.

To an uninformed mind, these minor discrepancies would most probably suggest shortcomings of model (1), not surprisingly given its caricatural simplicity, or the fact that microarray and luminescence data have been recorded in different experiments. However, these discrepancies are clearly linked to errors in floor level, which are naturally explained by the protein floor level bias discussed in Sec. II C and evidenced in Fig. 3.

Let us now consider more closely the adjustment of model (1) to protein target profiles with floor levels removed, together with RNA target profiles, shown in Figs. 4(e)–4(h). Goodness of fit is clearly improved compared to Figs.

TABLE I. Model parameter values. Optimal parameter values for adjustment of model to data using various score functions, assuming a free-running period of 24 h. Parameters are rescaled so that the maximum value of protein profiles is 100 nM, and the maximum value of CCA1 (respectively, TOC1) mRNA profile is 10 nM (respectively, 70 nM). The TOC1 and CCA1 mRNA maximum values are chosen in the same proportion as in microarray data. The third row of the table indicates whether the floor levels of luminescence data are removed (R) or not (NR). The last part of the table gives the degradation rate D_X at the mean value $\bar{X}=(\max(X)+\min(X))/2$; $D_X=\delta_X K_X/(K_X+\bar{X})$ for each species.

	S_{MP}	S_{MPF}	S_P	S_M	S_{MP}	S_P
Score (min)	37.7	32.8	12.9	4.2	22.7	11.3
Floor level	NR	NR	NR	–	R	R
n_C	2	2	1	2	2	2
n_T	2	2	2	2	2	2
μ_C (nM h ⁻¹)	1.92×10^{-3}	1.21×10^{-3}	5.18×10^{-3}	2.97×10^{-4}	1.53×10^{-1}	1.46×10^{-1}
λ_C (nM h ⁻¹)	3.64	6.61	6.56	3.32	3.11	3.31
P_{T0} (nM)	31.6	27.0	91.6	18.8	18.7	50.0
β_C (h ⁻¹)	2.60	2.50	3.56	2.89	2.83	3.78
μ_T (nM h ⁻¹)	0.0117	104	1.69	1.48	0.467	0.0270
λ_T (nM h ⁻¹)	560	3130	67.0	272	487	233
P_{C0} (nM)	7.29	5.65	8.74	3.33	4.51	2.73
β_T (h ⁻¹)	0.667	0.682	6.86	0.811	0.812	0.759
$1/\delta_{M_C}$ (h)	0.542	0.0223	0.831	0.0161	0.195	0.652
$1/\delta_{P_C}$ (h)	3.17	4.78	1.51	1.17	2.36	1.44
$1/\delta_{M_T}$ (h)	0.140	0.0215	4.97	0.151	0.129	0.736
$1/\delta_{P_T}$ (h)	0.559	4.20	0.0659	0.118	0.199	1.65
K_{M_C} (nM)	1.35	0.105	2.24	0.0315	0.407	0.842
K_{P_C} (nM)	133	746	77.5	23.6	75.9	72.2
K_{M_T} (nM)	37.7	12.2	128	60.5	28.3	157
K_{P_T} (nM)	7.77	311	45.7	1.52	2.76	47.6
D_{M_C} (h ⁻¹)	0.220	0.466	0.220	0.196	0.201	0.119
D_{P_C} (h ⁻¹)	0.180	0.185	0.290	0.164	0.183	0.292
D_{M_T} (h ⁻¹)	2.49	6.92	0.130	3.08	2.23	0.939
D_{P_T} (h ⁻¹)	0.129	0.180	4.76	0.127	0.135	0.196

4(a)–4(d), where it was already good. The simultaneous adjustment of the four target curves (score function S_{MP}) is achieved with an accuracy rarely obtained for a genetic circuit, with all numerical curves superimposing perfectly on the experimental ones, except that the numerical CCA1 protein profile rises slightly more slowly than the experimental one between dusk and CCA1 expression peak. It is quite striking to note that the two curves begin to separate precisely in the time interval where a transient window of CCA1 stabilization was predicted to occur in Ref. 2. Remarkably, numerical profiles obtained by adjusting RNA data alone (score function S_M) reproduce all experimental data very well also. When protein data are adjusted separately (score function S_P), RNA profiles are much better reproduced than when the protein floor is not removed, with only the CCA1 mRNA profile showing noticeable discrepancies in certain parts of the day, but with the global shape of the profile preserved. Thus, we can conclude that removing floor level significantly improves adjustment and restores coherence between microarray and luminescence data. Given the simplicity of the model, it is very unlikely that this may have occurred by chance. As we will discuss in Sec. IV, it is moreover a remarkable finding that experimental data recorded with different techniques and in different experiments^{1,15} can be adjusted simultaneously with such high accuracy by a

minimal model. This certainly reveals an important property of the clock oscillator.

The parameter sets in the last three columns of Table I, which correspond to profiles with protein floors removed, are generally more consistent between themselves than those in the first three columns, where the bias has not been corrected, and where important variations can be seen (compare, for example, λ_T across the first three columns). This reflects the fact that when the floor level is not removed, it is more difficult to predict mRNA profiles from adjusting protein profiles alone and vice versa. To make meaningful comparisons, it should however be kept in mind that the significant quantities are often not parameters themselves but combinations of them. When degradation is saturated, for example, the relevant quantity is the product $\delta_X K_X$, and it may happen that δ_X and K_X fluctuate more between columns than their product. Therefore, we give at the bottom of Table I the effective degradation rates at a mean value of the concentration D_X , which appear to be much more consistent than the individual degradation parameters. Another example is that although the values of μ_i in the fourth and fifth columns are quite different, the minimum values taken by TOC1 transcription rate in both cases are in fact comparable (1.78 versus 1.45 nM/h), and result from different combinations of λ_T , P_{C0} , and of the minimal value reached by the P_C profile,

which is very close to the repression threshold P_{C0} . This illustrates the general fact that although collective fitting can provide well-constrained predictions (here the time profiles), individual parameters may be poorly constrained, a feature observed in many systems biology models.⁴²

Some key ingredients of the dynamics can still be extracted unambiguously from a careful examination of Table I. The first are the strongly saturated degradations of M_C and P_T , characterized by values of K_{M_C} and K_{P_T} so small that the concentrations of the respective molecular actors are well above them for almost all of the diurnal cycle. The signature of this behavior in the time profiles is a straight-line decay after the expression peak. A natural question is whether what appears as saturated degradation of a given actor may in fact result from an interaction with another actor not yet identified. In the case of TOC1, a post-transcriptional regulation acting after dusk has indeed been evidenced experimentally.³⁷ Other key features are a maximum transcription rate significantly higher for TOC1 than for CCA1 and a small threshold of repression P_{C0} , comparable to the minimum value of the CCA1 protein level. These are related since the latter implies that TOC1 is repressed most of the time except in a very narrow time interval around the minimum of CCA1 protein level (this correlates with the narrow TOC1 mRNA peak). This must be compensated for by a high TOC1 transcription rate. Interestingly, the small value for P_{C0} can account for the experimentally observed relative inefficiency of antisense strategies against CCA1.¹ If CCA1 level is reduced, even significantly, the small time interval where TOC1 is not repressed will extend only slightly, with the dynamical behavior at other times being mostly unchanged. Thus, the global perturbation will remain limited and arrhythmia will be difficult to induce. In contrast to this, the activation threshold P_{T0} above which TOC1 activates CCA1 is relatively low so that activation occurs over a rather large time interval of more than 10 h. Therefore, modifications of the activation threshold or TOC1 protein level will influence the dynamics for a long time and thus easily disrupt oscillations. This also explains why FRP is much more sensitive to overexpression of TOC1 induced by TOC1:luc insertion than of CCA1 induced by CCA1:luc.¹

Finally, we carried out data adjustment assuming free-running periods of 23.8 and 25 h, to verify that our results do not depend critically on FRP (Fig. 5, parameters given in Table II). In this case, we allowed day/night modulation of some parameters to ensure frequency locking with the diurnal cycle. The agreement remains very good, although it degrades noticeably compared to the case of a FRP of 24 h. In particular, a phase shift of the TOC1 mRNA profile is induced. The parameter modulation remains very small (for a FRP of 25 h, the repression threshold is modulated but remains below the minimum of the CCA1 temporal profile, so that the effect of the change is minimal). We furthermore observed that when the FRP was a freely adjustable parameter, it would systematically converge toward a value of 24 h. All this confirms that there is no day/night parameter modulation in the TOC1-CCA1 feedback loop and that coupling can only occur in specific time windows as proposed in Ref. 2.

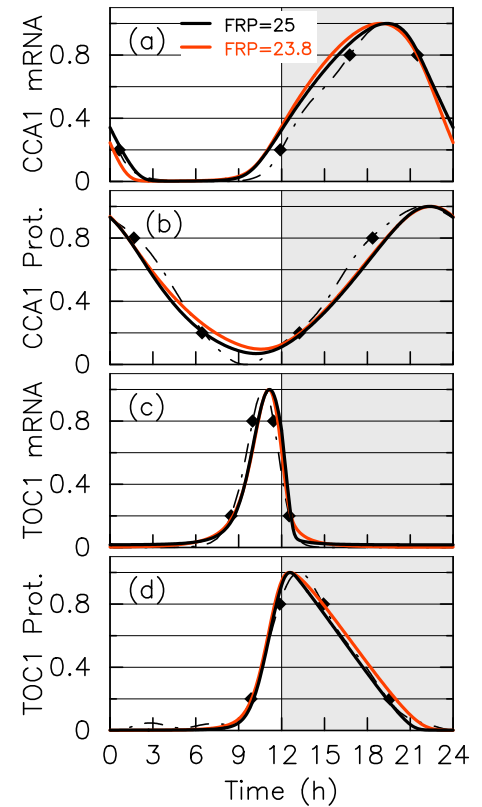


FIG. 5. (Color online) Adjustment of numerical solutions of the light-dependent model (solid lines) to target curves with floor levels removed (dot-dashed line) using the S_{MP} score function. Adjustment of (a) CCA1 mRNA, (b) CCA1 protein, (c) TOC1 mRNA, and (d) TOC1 protein without the floor level of proteins. The FRP of the autonomous oscillator is either of 25 h (black solid line) or 23.8 h (light gray or red solid line). Synchronization is obtained for the 25 h (respectively, 23.8 h) FRP model by assuming that the parameter P_{C0} (respectively, δ_{P_C}) takes a different value at day and at night.

TABLE II. Model parameter values. Optimal parameter values for adjustment of model to data using various score functions and assuming a free-running period of 25 or 23.8 h. Parameters are rescaled as in Table I.

FRP (h)	25	23.8
Score (min)	27.8	34.3
n_C	1	1
n_T	2	2
μ_C (nM h ⁻¹)	6.06×10^{-4}	4.73×10^{-2}
λ_C (nM h ⁻¹)	3.82	4.99
P_{T0} (nM)	15.9	27.3
β_C (h ⁻¹)	2.71	2.34
μ_T (nM h ⁻¹)	22.7	6.07×10^{-2}
λ_T (nM h ⁻¹)	357	1130
P_{C0} (nM)	5.94	6.15
P_{C0}^{night} (nM)	11.3	
β_T (h ⁻¹)	0.805	0.809
$1/\delta_{M_C}$ (h)	0.305	0.296
$1/\delta_{P_C}$ (h)	2.28	2.50
$1/\delta_{P_C}^{\text{night}}$ (h)		2.60
$1/\delta_{M_T}$ (h)	4.28×10^{-2}	4.51×10^{-2}
$1/\delta_{P_T}$ (h)	0.224	0.402
K_{M_C} (nM)	0.690	0.879
K_{P_C} (nM)	68.5	58.3
K_{M_T} (nM)	6.96	17.1
K_{P_T} (nM)	3.23	5.62

IV. DISCUSSION AND CONCLUSION

In this work, we have carried out a careful and detailed analysis of time series characterizing temporal variations of expression of the two central genes of *Ostreococcus* circadian clock, TOC1 and CCA1.¹ These time series had been previously obtained from microarray data and luminescent reporter data recorded in two different experiments.^{1,15} From these time series, we have extracted periodic temporal profiles for the RNA and protein concentrations of the two genes, assumed to represent the mean circadian oscillations in individual cells so as to optimize comparison with numerical profiles, which are inherently periodic. In particular, approximating the periodic component of luminescence time series by a Fourier series allowed us to separate the dynamics at work throughout the day from fast transient processes activated only near light/dark and dark/light transitions, and probably involved in occasional resetting of the clock (compare raw data and target profile in Fig. 2). The data processing also allowed us to detect a bias in the luminescence time series, which was confirmed by a direct comparison of individual signals from two genetically identical cell cultures (they were not proportional to each other). This bias manifests itself in a nonzero luminescence floor level. Taking it into account allowed us to show that both protein levels approach zero at some times of the day, which was essential for the quality of the subsequent model adjustment. This illustrates how important it is to exploit the redundancy of data by verifying that data that should provide the same information are indeed consistent.

Thanks to the careful data processing, we could evidence an extraordinarily good agreement between a minimal model of a two-gene transcriptional feedback loop and the reconstructed concentration profiles. This model describes activation of CCA1 by TOC1 and repression of TOC1 by CCA1. It describes only regulated transcription of the two genes, translation and degradation, and comprises only differential equations, from which the time evolution of the two mRNA and the two proteins can be computed and compared to the target profiles. Therefore, a biochemically detailed model taking into account compartmentalization or genomic insertions would have only served to adjust details without biological relevance, and could well have masked one of the main results, which is the good adjustment by a free-running oscillator model. It is also quite remarkable that using only four points of the reconstructed profiles sufficed to constrain the numerical profiles to follow their targets throughout the day. Here also, adding more points would have only forced the model to adjust to irrelevant details without gaining information. This suggests that the complexity of the model and the constraining data should be carefully matched.

The excellent agreement found between the model and the data supports unambiguously the existence of a two-gene oscillator at the core of *Ostreococcus* circadian clock. This not only builds a solid foundation on which future studies of this clock can rely but also provides what we believe is one of the clearest examples of a natural few-gene oscillator evidenced from experimental data. This is all the more important as the role of this oscillator in the circadian clock constrains it to be extremely robust to all kinds of fluctuations.

Understanding the dynamical ingredients besides transcriptional regulation that underlie its robustness will certainly be of high interest for the study of genetic oscillators in general. Two remarkable features of the TOC1-CCA1 oscillator have indeed emerged from our analysis. First, a strongly saturated degradation has been evidenced both for the CCA1 mRNA (already noted in our previous work²) and the TOC1 protein (detected in this work), which manifests itself by a straight-line decay at high concentrations after the expression peak. This behavior may be the signature of post-transcriptional and post-translational interactions, and is thus compatible with the experimental observations of Ref. 37. It also supports the putative role of saturated degradation mechanisms as an efficient mechanism to introduce effective delays along a negative feedback loop^{10,43–45} and the recent observation that this is a key ingredient to generate robust oscillations.^{8,46} The second remarkable feature predicted by the model is that the TOC1 gene is repressed by CCA1 during most of the day, except during a short time interval located 1 or 2 h before dusk, which is consistent with the very narrow peak of TOC1 mRNA expression observed in the experimental data. The small duration of TOC1 expression is compensated by a high transcription rate. One may wonder whether this design has an influence on the robustness to molecular fluctuations. In any case, it explains why functional studies of the oscillator showed that it was much more sensitive to perturbations in the TOC1 level than in the CCA1 level.¹

It has recently been proposed that circadian clocks should not only be robust to fluctuations in molecule numbers^{17,18} and to temperature variations^{19,20} but also to fluctuations in the daylight intensity pattern, which is crucial to synchronize the circadian clock to the day/night cycle.^{2,23} Our results demonstrate such robustness for the TOC1-CCA1 oscillator in two different ways. First, the experimental data are reproduced accurately by a free-running oscillator model. This had already been noted by Ref. 2, but the confirmation of this behavior using luminescence signals recorded every hour reinforces its plausibility significantly. As shown in Ref. 2, it can only be explained by assuming that coupling to light is scheduled so as to be active precisely when the oscillator does not respond to external perturbations. In this scheme, the oscillator is not affected by light in normal entrainment conditions, which naturally makes it blind to light fluctuations, and thus behaves as if it was free-running. When the oscillator drifts out of phase, light sensing occurs at a different time of its cycle, where it responds so as to recover the normal entrainment phase. The acute responses to light/dark and dark/light transitions that occur transiently in raw signals (Fig. 2) may correspond to such resetting.

The second strong evidence for the robustness of the TOC1-CCA1 oscillator comes from the fact that the profiles adjusted precisely and simultaneously by our simple model are reconstructed from time series recorded in different experiments by two different techniques, with totally different setups and in particular under very different lighting conditions. This strongly suggests that *Ostreococcus* clock is able to tick exactly in the same way in different conditions, and in particular in different levels of light, which is an obvious requisite for robustness to daylight fluctuations. If the clock

can cope with randomly varying daylight profiles without being perturbed, it can certainly accommodate constantly low or constantly high daylight levels and deliver similar biochemical signals in both cases. The fact that the model adjusted from RNA profiles predicts protein profiles and vice versa (although with less precision in the latter case) clearly illustrates the consistency between the two experiments. Again, we stress that this remarkable finding may have remained masked without our careful data processing.

It is important to mention that both the presence of saturated degradation kinetics and the absence of effective coupling to light when entrained are strong predictions that can be tested experimentally. While the impressive agreement between the model and experiments obtained here clearly shows that a TOC1-CCA1 oscillator underlies *Ostreococcus* clock, it should not make us forget that it is only a part of it. The free-running average behavior in entrainment conditions and the acute responses at day/night and night/day transitions can only be explained if the TOC1-CCA1 loop interacts with other molecular loops and feedback loops. Indeed, it was shown in Ref. 2 that robustness to daylight fluctuations involves a precisely timed coupling activation. This necessarily requires additional feedback loops designed so as to generate the signal that will drive optimal coupling. Furthermore, interactions with other molecular actors may well be responsible for the saturated degradation of CCA1 mRNA and TOC1 protein.

Thus, we have now to identify these additional feedback loops as well as the light input pathways, which should help us to understand how the clock can synchronize in spite of fluctuations to day/night cycles of variable duration across the year. Adapting to different photoperiods and adapting to different light levels are indeed unrelated evolutionary goals that must be simultaneously satisfied. In the end, we hope to build an accurate and comprehensive model of a simple and robust circadian clock. If the agreement between theory and experiment remains comparable to what was achieved in the present study, this may well provide us with new and deep insights into the design and function of circadian clocks. More generally, we believe our results promote *Ostreococcus*, whose low genomic redundancy³⁶ is probably crucial for allowing accurate quantitative approaches, as a very promising model for systems biology.

ACKNOWLEDGMENTS

This work has been supported by the ANR Grant No. 07BSYS004 for F.-Y.B. and M.L., by the CNRS interdisciplinary program “Interface Physique, Biologie et Chimie: Soutien à la prise de risque” for M.L., as well as by the French Ministry of Higher Education and Research, Nord-Pas de Calais Regional Council, and FEDER through Contrat de Projets Etat-Region (CPER) 2007–2013.

¹F. Corellou, C. Schwartz, J.-P. Motta, E. B. Djouani-Tahri, F. Sanchez, and F.-Y. Bouget, *Plant Cell* **21**, 3436 (2009).

²Q. Thommen, B. Pfeuty, P. Morant, F. Corellou, F. Bouget, and M. Lefranc, *PLOS Comput. Biol.* **6**, e1000990 (2010).

³A. Goldbeter, *Biochemical Oscillations and Cellular Rhythms* (Cambridge University Press, Cambridge, 1996).

⁴J. Hasty, A. Hoffmann, and S. Golden, *Curr. Opin. Genet. Dev.* **20**, 571 (2010).

⁵A. T. Winfree, *The Geometry of Biological Time*, 3rd ed. (Springer-Verlag, Berlin, 2001).

⁶*Molecular Genetics of Biological Rhythms*, edited by M. W. Young (Dekker, New York, 1993).

⁷M. Elowitz and S. Leibler, *Nature (London)* **403**, 335 (2000).

⁸J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, *Nature (London)* **456**, 516 (2008).

⁹L. Hartwell, J. Hopfield, S. Leibler, and A. Murray, *Nature (London)* **402**, C47 (1999).

¹⁰G. Tiana, S. Krishna, S. Pigolotti, M. H. Jensen, and K. Sneppen, *Phys. Biol.* **4**, R1 (2007).

¹¹B. Mengel, A. Hunziker, L. Pedersen, A. Trusina, M. H. Jensen, and S. Krishna, *Curr. Opin. Genet. Dev.* **20**, 656 (2010).

¹²D. A. Rand, B. V. Shulgin, D. Salazar, and A. J. Millar, *J. R. Soc., Interface* **1**, 119 (2004).

¹³C. S. Pittendrigh, *Cold Spring Harbor Symp. Quant. Biol.* **25**, 159 (1960).

¹⁴A. N. Dodd, N. Salathia, A. Hall, E. Kevei, R. Toth, F. Nagy, J. Hibberd, A. J. Millar, and A. A. Webb, *Science* **309**, 630 (2005).

¹⁵M. Moulager, A. Monnier, B. Jesson, R. Bouvet, J. Mosser, C. Schwartz, L. Gamier, F. Corellou, and F.-Y. Bouget, *Plant Physiol.* **144**, 1360 (2007).

¹⁶A. Monnier, S. Liverani, R. Bouvet, B. Jesson, J. Smith, J. Mosser, F. Corellou, and F. Bouget, *BMC Genomics* **11**, 192 (2010).

¹⁷D. Gonze, J. Halloy, and A. Goldbeter, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 673 (2002).

¹⁸N. Barkai and S. Leibler, *Nature (London)* **403**, 267 (2000).

¹⁹C. S. Pittendrigh, *Proc. Natl. Acad. Sci. U.S.A.* **40**, 1018 (1954).

²⁰L. Rensing and P. Ruoff, *Chronobiol. Int.* **19**, 807 (2002).

²¹M. Comas, D. Beersma, R. Hut, and S. Daan, *J. Biol. Rhythms* **23**, 425 (2008).

²²D. G. M. Beersma, S. Daan, and R. A. Hut, *J. Biol. Rhythms* **14**, 320 (1999).

²³C. Troein, J. C. W. Locke, M. S. Turner, and A. J. Millar, *Curr. Biol.* **19**, 1961 (2009).

²⁴J. C. Dunlap, *Cell* **96**, 271 (1999).

²⁵M. W. Young and S. Kay, *Nat. Genet.* **2**, 702 (2001).

²⁶S. Panda, J. B. Hogenesch, and S. A. Kay, *Nature (London)* **417**, 329 (2002).

²⁷D. B. Forger and C. S. Peskin, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14806 (2003).

²⁸D. B. Forger, D. A. Dean II, K. Gurdziel, J.-C. Leloup, C. Lee, C. Von Gall, J.-P. Eichegaray, R. E. Kronauer, A. Goldbeter, C. S. Peskin, M. E. Jewett, and D. R. Weaver, *OMICS* **7**, 387 (2003).

²⁹J. C. W. Locke, M. M. Southern, L. Kozma-Bognar, V. Hibberd, P. E. Brown, M. S. Turner, and A. J. Millar, *Mol. Syst. Biol.* **1**, 2005.0013 (2005).

³⁰J. C. W. Locke, L. Kozma-Bognar, P. D. Gould, B. Feher, E. Kevei, F. Nagy, M. S. Turner, A. Hall, and A. J. Millar, *Mol. Syst. Biol.* **2**, 59 (2006).

³¹M. N. Zeilinger, E. M. Farre, S. R. Taylor, and S. A. Kay, *Mol. Syst. Biol.* **2**, 58 (2006).

³²J. D. Salazar, T. Saithong, P. E. Brown, J. Foreman, J. C. W. Locke, K. J. Halliday, I. A. Carré, D. A. Rand, and A. J. Millar, *Cell* **139**, 1170 (2009).

³³P. François, *Biophys. J.* **88**, 2369 (2005).

³⁴C. Courties, A. Vaquer, M. Troussellier, J. Lautier, M. J. Chretiennot-Dinet, J. Neveux, M. C. Machado, and H. Claustré, *Nature (London)* **370**, 255 (1994).

³⁵M. J. Chretiennot-Dinet, C. Courties, A. Vaquer, J. Neveux, H. Claustré, J. Lautier, and M. C. Machado, *Phycologia* **34**, 285 (1995).

³⁶E. Derelle, C. Ferraz, S. Rombauts, P. Rouzé, A. Z. Worden, S. Robbins, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, Y. Saeyns, J. Wuyts, K. Jabbari, C. Bowler, O. Panaud, B. Piégu, S. G. Ball, J.-P. Ral, F.-Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, and H. Moreau, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11647 (2006).

³⁷E. Djouani-Tahri, J.-P. Motta, F.-Y. Bouget, and F. Corellou, *Plant Signal Behav.* **5**, 332 (2010).

³⁸D. Alabadi, T. Oyama, M. J. Yanovsky, F. G. Harmon, P. Mas, and S. A. Kay, *Science* **293**, 880 (2001).

³⁹B. Gates and M. DeLuca, *Arch. Biochem. Biophys.* **169**, 616 (1975).

⁴⁰M. Lefranc, D. Hennequin, and P. Glorieux, *Phys. Lett. A* **163**, 269 (1992).

⁴¹C. Letellier and L. A. Aguirre, *Phys. Rev. E* **82**, 016204 (2010).

⁴²R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, *PLoS Comput. Biol.* **3**, e189 (2007).

⁴³G. Kurosawa and Y. Isawa, *J. Biol. Rhythms* **17**, 568 (2002).

⁴⁴W. W. Wong, T. Y. Tsai, and J. C. Liao, *Mol. Syst. Biol.* **3**, 130 (2007).

⁴⁵P.-E. Morant, Q. Thommen, F. Lemaire, C. Vandermoere, B. Parent, and M. Lefranc, *Phys. Rev. Lett.* **102**, 068104 (2009).

⁴⁶W. Mather, M. R. Bennett, J. Hasty, and L. S. Tsimring, *Phys. Rev. Lett.* **102**, 068105 (2009).

Annexe C

Fonction de transcription et de dégradation

C.1 Fonction de transcription

Dans cette annexe, on détaille le calcul pour obtenir la fonction de transcription de répression ou d'activation dans le cas d'une compétition entre facteurs de transcription pour le même promoteur. On illustre le calcul en prenant deux facteurs de transcription appelé P_1 et P_2 régulant le gène G . Les réactions s'écrivent :



Le bilan de conservation du gène s'écrit :

$$N = [G] + [P_1 : G] + [P_2 : G] \quad (\text{C.5})$$

Le paramètre N est le nombre de copie du gène. Les équations dynamiques du gène dans ces différentes formes et de l'ARN s'écrivent :

$$\frac{dG(t)}{dt} = -G(t)(\alpha_1 P_1(t) + \alpha_2 P_2(t)) + \theta_1 P_1 : G(t) + \theta_2 P_2 : G(t) \quad (\text{C.6})$$

$$\frac{dP_1 : G(t)}{dt} = \alpha_1 P_1(t)G(t) + \theta_1 P_1 : G(t) \quad (\text{C.7})$$

$$\frac{dP_2 : G(t)}{dt} = \alpha_2 P_2(t)G(t) + \theta_2 P_2 : G(t) \quad (\text{C.8})$$

$$\frac{dM(t)}{dt} = \rho_1 G(t) + \rho_2 (P_1 : G(t) + P_2 : G(t)) \quad (\text{C.9})$$

On suppose un équilibre rapide du gène tel qu'on puisse éliminer adiabatiquement les variables $P_1 : G$ et $P_2 : G$:

$$P_1 : G(t) = \frac{\alpha_1}{\theta_1} P_1 G \quad (\text{C.10})$$

$$P_2 : G(t) = \frac{\alpha_2}{\theta_2} P_2 G \quad (\text{C.11})$$

À partir du bilan de conservation du gène C.5, la fonction donnant le gène au cours du temps peut s'écrire en fonction de P_1 et P_2 en substituant $P_1 : G$ et $P_2 : G$ tel que :

$$G(t) = \frac{N}{1 + \frac{\alpha_1}{\theta_1} P_1(t) + \frac{\alpha_2}{\theta_2} P_2(t)} \quad (\text{C.12})$$

On réécrit l'équation dynamique C.9 de l'ARN messenger en utilisant la loi de conservation du gène C.5 pour éliminer les complexes protéines-gène :

$$\frac{dM(t)}{dt} = \rho_2 N + (\rho_1 - \rho_2) G(t) \quad (\text{C.13})$$

La partie à droite de l'égalité de l'équation C.13 est la fonction de transcription. Deux cas sont à aborder. Le premier est le cas où ρ_2 est supérieur à ρ_1 . Cela signifie que le gène est activé car il produit plus d'ARNm lorsqu'il est régulé par la protéine. Le deuxième cas est la situation où ρ_1 est supérieur à ρ_2 . Le gène est alors réprimé par le facteur de transcription. On réécrit l'équation dynamique C.13 pour avoir une forme équivalente à celle que nous avons utilisé dans le manuscrit de thèse. On étudie d'abord la forme de la fonction de transcription pour une répression du gène. Le résultat est immédiat en posant $\rho_2 N = \mu$ et $\lambda = N(\rho_1 - \rho_2)$. La fonction s'écrit avec des paramètres positifs. On étudie maintenant le système dans le cas d'une activation du gène par le facteur de transcription. On injecte l'expression C.12 du gène dans l'équation C.13 puis on réduit au même dénominateur pour obtenir l'expression suivante :

$$\frac{dM(t)}{dt} = N \frac{\rho_1 + \rho_2 \left(\frac{\alpha_1}{\theta_1} P_1 + \frac{\alpha_2}{\theta_2} P_2 \right)}{1 + \frac{\alpha_1}{\theta_1} P_1 + \frac{\alpha_2}{\theta_2} P_2} = N \rho_1 + N \frac{(\rho_2 - \rho_1) \left(\frac{\alpha_1}{\theta_1} P_1 + \frac{\alpha_2}{\theta_2} P_2 \right)}{\left(1 + \frac{\alpha_1}{\theta_1} P_1 + \frac{\alpha_2}{\theta_2} P_2 \right)} \quad (\text{C.14})$$

On pose $N \rho_1 = \mu$ et $N(\rho_2 - \rho_1) = \lambda$ pour retrouver a forme étudié dansla thèse.

C.2 Fonction de dégradation

On décrit dans cette section comment établir la fonction de dégradation enzymatique lorsqu'il y a une compétition pour la même enzyme de dégradation entre plusieurs protéines. L'enzyme est nommée E et les deux protéines en compétition sont P_1 et P_2 . Les réactions enzymatiques s'écrivent :



Comme la quantité totale d'enzyme est conservé, on peut écrire le bilan de conservation de l'enzyme :

$$E_0 = [E] + [P_1 : E] + [P_2 : E] \quad (\text{C.19})$$

On écrit les équations dynamiques des variables $[P_1 : E]$, $[P_2 : E]$ et $[P_1]$ sont :

$$\frac{dP_1 : E(t)}{dt} = k_{11}P_1(t)E(t) - (k_{-11} + k_{12})P_1 : E(t) \quad (\text{C.20})$$

$$\frac{dP_2 : E(t)}{dt} = k_{21}P_2(t)E(t) - (k_{-21} + k_{22})P_2 : E(t) \quad (\text{C.21})$$

$$\frac{dP_1(t)}{dt} = -k_{11}P_1(t)E(t) - +k_{-11}P_1 : E(t) \quad (\text{C.22})$$

Les équations dynamiques C.20 et C.21 sont supposées en quasi-équilibre car les complexes enzyme-protéine sont très instables. Il vient alors :

$$P_1 : E(t) = \frac{k_{11}P_1(t)E(t)}{k_{-11} + k_{12}} \quad (\text{C.23})$$

$$P_2 : E(t) = \frac{k_{21}P_2(t)E(t)}{k_{-21} + k_{22}} \quad (\text{C.24})$$

On peut alors écrire la fonction donnant $E(t)$ en remplaçant $P_1 : E(t)$ et $P_2 : E(t)$ par leur expression C.23 et C.24 et en les injectant dans le bilan de conservation enzymatique C.19. Il vient :

$$E(t) = \frac{E_0}{1 + \frac{k_{11}P_1}{k_{-11} + k_{12}} + \frac{k_{21}P_2}{k_{-21} + k_{22}}} \quad (\text{C.25})$$

On remplace ensuite $E(t)$ par son expression C.25, $P_1 : E(t)$ par son expression C.23 et $P_2 : E(t)$ par son expression C.24 dans l'équation dynamique donnant la variable $P_1(t)$ (ou $P_2(t)$) pour avoir la dégradation enzymatique de la protéine :

$$\frac{dP_1(t)}{dt} = -\frac{\delta_1 P_1(t)}{1 + \frac{P_1(t)}{\kappa_1} + \frac{P_2(t)}{\kappa_2}} \quad (\text{C.26})$$

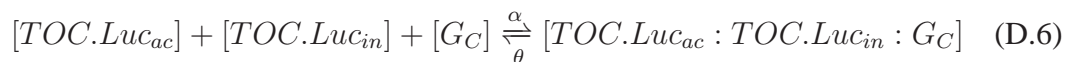
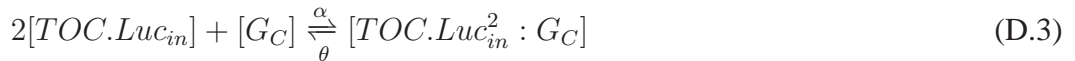
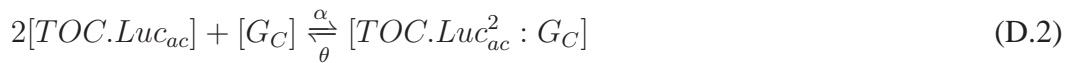
On pose dans l'expression C.26 : $\delta_1 = \frac{k_{11}k_{12}E_0}{k_{11} + k_{12}}$, $\kappa_1 = \frac{k_{11} + k_{12}}{k_{11}}$ et $\kappa_2 = \frac{k_{21} + k_{22}}{k_{21}}$.

Annexe D

Compétition entre facteurs de transcription

Le bilan des différentes formes de la protéine *TOC* sont la protéine *TOC* issue du gène originale du génôme, la protéine de fusion active *TOC.LUC_{ac}* (tant que la luciférase n'a pas provoqué d'émission de lumière) issue du gène ajouté et la protéine de fusion inactive *TOC.LUC_{in}*, résultant de l'émission d'un photon de la luciférase. La présence de ces différents acteurs provoque une compétition pour les sites d'activation (resp. de répression) du promoteur du gène de *CCA* (resp. *TOC*) et modifie la fonction de transcription.

Pour établir le modèle, un dimère de protéine *TOC* (resp. *CCA*) se forme toujours avant de réguler le gène *CCA* (resp. *TOC*) []. Les réactions chimiques correspondantes à la régulation du gène de la protéine *CCA*, appelé G_C , par les six différentes formes des dimères de protéine de la lignée *TOC :LUC* sont :



Le système d'équations différentielles correspondant à la formation du complexe dimère-promoteur s'établit à partir de la loi d'action de masse. Nous notons P_t pour *TOC*, P_{tl} pour *TOC.Luc_{ac}*

et $Ptli$ pour $TOC.Luc_{in}$:

$$\frac{d[Pt^2 : G_C]}{dt} = 1/2\alpha[Pt]^2[G_C] - \theta_1[Pt^2 : G_C] \approx 0 \quad (D.7)$$

$$\frac{d[Pt^2 : G_C]}{dt} = 1/2\alpha[Pt]^2[G_C] - \theta_1[Pt^2 : G_C] \approx 0 \quad (D.8)$$

$$\frac{d[Ptli^2 : G_C]}{dt} = 1/2\alpha[Ptli]^2[G_C] - \theta[Ptli^2 : G_C] \approx 0 \quad (D.9)$$

$$\frac{d[Pt : Ptl : G_C]}{dt} = \alpha[Pt][Ptl][G_C] - \theta[Pt : Ptl : G_C] \approx 0 \quad (D.10)$$

$$\frac{d[Pt : Ptli : G_C]}{dt} = \alpha[Pt][Ptli][G_C] - \theta[Pt : Ptli : G_C] \approx 0 \quad (D.11)$$

$$\frac{d[Ptli : Ptl : G_C]}{dt} = \alpha[Ptli][Ptl][G_C] - \theta[Ptli : Ptl : G_C] \approx 0 \quad (D.12)$$

Les équations D.7 à D.12 sont l'état de quasi-équilibre et donnent les expressions des gènes liés $[i : G_C]$, avec

$$i = \{Pt^2, Ptli^2, Pt : Ptl, Pt : Ptli, Ptl : Ptli\}$$

en fonction de $[G_C]$ et des différentes formes de dimères de TOC .

La loi de conservation du gène étant :

$$1 = [G] + \sum_i [i : G_C] \quad (D.13)$$

Dans l'égalité D.13, nous substituons les termes $[i : G_C]$ par leur expression en fonction de G_C et des différentes formes de dimère de TOC déduite de l'état de quasi-équilibre des équations D.7 à D.12. L'expression du gène est :

$$[G_C] = \frac{1}{1 + 1/Pt_0^2 ([Pt] + [Ptl] + [Ptli])^2} \quad (D.14)$$

Le paramètre Pt_0 est égale à $\frac{\theta}{\alpha}$. L'égalité D.14 est équivalente à l'équation 1.10 en remplaçant $P(t)$ par la somme des différentes forme de la protéine. Finalement, la nouvelle fonction de transcription de synthèse de l'ARNm Mc s'écrit :

$$\frac{dMc}{dt} = \mu_c + \lambda_c \frac{(Pt + Ptl + Ptli)^2}{1 + 1/Pt_0^2 (Pt + Ptl + Ptli)^2} \quad (D.15)$$

De même, la fonction de transcription de l'ARN de Mt est :

$$\frac{dMt}{dt} = \mu_t + \frac{\lambda_t}{1 + 1/Pt_0^2 (Pc + Pcl + Pcli)^2} \quad (D.16)$$

Annexe E

Tableaux des valeurs de paramètres du modèle tenant compte de la luciférase

Dans chacune des sections, nous regroupons les valeurs des paramètres déduit des ajustements des figures 4.5, 4.6, 4.7 et 4.8.

Le tableau suivant indique l'unité de chaque paramètre :

taux de transcription : μ, λ	nM/h
taux de traduction : β	$1/h$
taux de dégradation : δ	$1/h$
seuil de régulation : P	nM
constante de saturation : κ	nM

E.1 Lignée $pTOC :LUC$

numero de lignée	0	1	2	4	5	6
score	0.53	0.51	0.51	0.6	0.49	0.45
n_C	2	2	2	2	2	2
n_T	2	2	2	2	2	2
μ_c	0.32	0.32	0.32	0.32	0.32	0.32
μ_t	0.009	0.009	0.009	0.009	0.009	0.009
μ_{tl}	0.0	0.7	1.08	1.18	0.36	0.0
λ_c	0.36	0.36	0.36	0.36	0.36	0.36
λ_t	0.9	0.9	0.9	0.9	0.9	0.9
λ_{tl}	2.89	1.11	50.93	1.49	12.55	1.12
β_{mt}	0.43	0.43	0.43	0.43	0.43	0.43
β_{mc}	0.45	0.45	0.45	0.45	0.45	0.45
β_{ml}	0.4	0.4	0.35	0.34	0.32	0.31
P_{t_0}	0.47	0.47	0.47	0.46	0.46	0.47
P_{C_0}	0.01	0.01	0.01	0.01	0.01	0.01
$P_{C_{0l}}$	0.004	0.02	0.001	0.007	0.002	0.01
κ_{mc}	0.006	0.006	0.006	0.006	0.006	0.006
κ_{pc}	0.49	0.49	0.49	0.49	0.49	0.49
κ_{pt}	0.008	0.008	0.008	0.008	0.008	0.008
κ_{mt}	0.03	0.03	0.03	0.03	0.03	0.03
κ_{ml}	0.04	0.04	0.13	0.10	0.07	0.05
δ_{ml}	8.93	32.53	12.95	18.01	12.47	13.09
δ_{mt}	11.54	11.54	11.54	11.54	11.54	11.54
δ_{mc}	71.51	71.51	71.51	71.51	71.51	71.5
δ_{pt}	16.38	16.38	16.38	16.38	16.38	16.38
δ_{pc}	1.07	1.07	1.07	1.07	1.07	1.07
δ_l	0.14	0.18	0.17	0.17	0.1	0.09

E.2 Lignée $pLHY :LUC$

numero de lignée	0	1	2	3	4	5
score	1.3	0.46	1.1	0.48	0.82	1.28
n_C	2	2	2	2	2	2
n_T	2	2	2	2	2	2
μ_c	0.32	0.32	0.32	0.32	0.32	0.32
μ_t	0.009	0.009	0.009	0.009	0.009	0.009
μ_{cl}	0.0	0.54	0.11	0.025	0.0015	0.12
λ_c	0.36	0.36	0.36	0.36	0.36	0.36
λ_{cl}	0.18	0.31	0.28	0.23	0.2	0.29
λ_t	0.9	0.9	0.9	0.9	0.9	0.9
β_{mt}	0.43	0.43	0.43	0.43	0.43	0.43
β_{mc}	0.45	0.45	0.45	0.45	0.45	0.45
β_{ml}	0.21	0.22	0.17	0.26	0.21	0.16
Pt_0	0.46	0.46	0.46	0.47	0.47	0.47
Pc_0	0.01	0.01	0.01	0.01	0.01	0.01
Ptl_0	0.06	0.23	0.22	0.3	0.03	0.22
κ_{mc}	0.006	0.006	0.006	0.006	0.006	0.006
κ_{pc}	0.49	0.49	0.49	0.49	0.49	0.49
κ_{pt}	0.008	0.008	0.008	0.008	0.008	0.008
κ_{mt}	0.03	0.03	0.03	0.03	0.03	0.03
κ_{ml}	0.36	0.09	0.01	61738.7	0.32	0.013
δ_{ml}	0.49	9.55	21.2	0.18	0.64	21.26
δ_{mt}	11.54	11.54	11.54	11.54	11.54	11.54
δ_{mc}	71.5	71.51	71.51	71.51	71.51	71.51
δ_{pt}	16.38	16.38	16.38	16.38	16.38	16.38
δ_{pc}	1.07	1.07	1.07	1.07	1.07	1.07
δ_l	0.16	0.17	0.11	0.22	0.16	0.1

E.3 Lignées *TOC :LUC*

numero de lignée	0	1	2	3	4	5
score	0.54	0.23	1.22	0.56	0.38	0.57
n_C	2	2	2	2	2	2
n_T	2	2	2	2	2	2
μ_c	0.36	0.32	0.36	0.31	0.37	0.29
μ_t	0.02	0.02	0.1	0.04	0.04	0.01
μ_{tl}	0.19	1.44	2.32	0.35	4.05	0.18
λ_c	0.42	0.37	0.41	0.36	0.42	0.34
λ_{tl}	0.001	2.11	0.76	0.002	0.34	0.001
λ_t	2.39	2.73	10.26	3.89	3.9	1.18
β_{mt}	0.33	0.04	0.02	0.19	0.02	9186.92
β_{mc}	0.46	0.39	0.39	0.44	0.38	0.37
β_{mtl}	1.92	0.53	1.33	4.65	0.46	23.8
P_{t_0}	2.92	0.66	1.63	2.7	0.69	7129.95
P_{c_0}	0.04	0.05	0.06	0.03	0.07	0.005
$P_{c_{0l}}$	0.44	0.07	0.13	0.0004	0.32	0.64
$P_{t_{l0}}$	0.05	0.01	0.04	0.45	0.02	1172.93
κ_{mc}	0.01	0.01	0.02	0.01	0.02	0.004
κ_{pc}	1.76	2.13	2.48	1.42	2.92	0.21
κ_{pt}	0.05	0.01	0.03	0.05	0.01	122.87
κ_{mt}	0.13	0.19	0.75	0.21	0.30	0.03
κ_{mtl}	1.85	4.17	0.92	0.34	6.55	0.14
κ_{ptl}	3.61	0.33	0.87	3.04	0.35	3534.3
δ_{mtl}	0.76	15.	9.04	2.0	19.11	2.19
δ_{mt}	6.56	5.16	5.04	6.63	4.69	15.58
δ_{mc}	40.66	32.	31.2	41.09	29.06	96.51
δ_{pt}	9.31	7.33	7.15	9.41	6.66	22.11
δ_{pc}	0.61	0.48	0.47	0.62	0.44	1.45
δ_{ptl}	0.96	0.97	1.41	2.66	1.02	0.14
δ_l	1.66	0.15	0.9	3.95	0.07	23.80

E.4 Lignées *LHY :LUC*

numero de lignée	0	1	2	3	4	5
score	0.7	0.7	0.65	0.72	0.52	0.51
n_C	2	2	2	2	2	2
n_T	2	2	2	2	2	2
μ_c	0.33	0.33	0.32	0.32	0.32	0.33
μ_t	0.01	0.01	0.01	0.01	0.01	0.01
μ_{cl}	0.01	0.003	0.08	0.17	0.09	0.01
λ_c	0.39	0.38	0.37	0.37	0.37	0.38
λ_{cl}	0.08	1.50	0.003	0.18	0.63	0.34
λ_t	0.89	0.88	0.89	0.89	0.90	0.89
β_{mt}	0.42	0.42	0.43	0.43	0.44	0.43
β_{mc}	0.48	0.47	0.46	0.45	0.45	0.47
β_{mcl}	$3.6e^{-06}$	$1.3e^{-4}$	$1.5e^{-5}$	$7.3e^{-6}$	$2.4e^{-7}$	$9.8e^{-5}$
Pt_0	0.5	0.49	0.48	0.47	0.47	0.48
Pt_{0l}	0.23	3.35	0.05	2.28	1.56	1.3
Pc_0	0.01	0.01	0.01	0.01	0.01	0.01
κ_{mc}	0.01	0.01	0.01	0.01	0.01	0.01
κ_{pc}	0.61	0.57	0.54	0.51	0.5	0.55
κ_{pt}	0.01	0.01	0.01	0.01	0.01	0.01
κ_{mt}	0.03	0.03	0.03	0.03	0.03	0.03
κ_{mcl}	3.19	7.93	4.37	2.51	3.93	4.56
κ_{pcl}	0.002	0.01	0.01	0.07	0.03	0.03
δ_{mcl}	0.96	0.51	3.21	18.41	16.55	1.12
δ_{mt}	10.95	11.01	11.2	11.35	11.56	11.24
δ_{mc}	67.85	68.20	69.41	70.34	71.59	69.61
δ_{pt}	15.54	15.62	15.9	16.11	16.4	15.94
δ_{pc}	1.02	1.02	1.04	1.05	1.07	1.04
δ_{pcl}	1.33	6.76	0.15	0.17	0.08	6.71
δ_l	1.95	1.82	0.12	0.06	0.14	17.07

Bibliographie

- [1] Albert Goldbeter. *Biochemical Oscillations and Cellular Rhythms : The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, April 1997.
- [2] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular biology of the cell*. Garland Science, 4 edition, 2002.
- [3] Kurt W. Kohn. Molecular interaction map of the mammalian cell cycle control and dna repair systems. *Molecular Biology of the Cell*, 10(8) :2703–2734, 1999.
- [4] Reppert SM and Weaver DR. Molecular analysis of mammalian circadian rhythms. *Annu. Rev. Physiol.*, 63 :647–76, 2001.
- [5] L Mendoza, D Thieffry, and E R Alvarez-Buylla. Genetic control of flower morphogenesis in arabidopsis thaliana : a logical analysis. *Bioinformatics*, 15(7) :593–606, 1999.
- [6] M.R. Doyle, S.J. Davis, R.M. Bastow, H.G. McWatters, L. Kozma-Bognar, F. Nagy, A.J. Millar, and R.M. Amasino. Elf4, a gene controlling circadian rhythms and flowering time in arabidopsis thaliana. *Nature*, 419(6902) :74–77, 2002.
- [7] Irina Mihalcescu, Weihong Hsing, and Stanislas Leibler. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature*, 430(6995) :81–85, 2004.
- [8] J. Craig Venter and al. The sequence of the human genome. *Science*, 291(5507) :1304–1351, 2001.
- [9] Pandey A and Mann M. Proteomics to study genes and genomes. *Nature*, 405(6788) :837–46, 2000.
- [10] J. van Helden, A. Naim, R. Mancuso, M. Eldridge, L. Wernisch, D. Gilbert, and S. J. Wodak. Representing and analysing molecular and cellular function using the computer. *Biol Chem*, 381(9-10) :921–935, 2000.
- [11] Granjeaud S, Bertucci F, and Jordan BR. Expression profiling : Dna arrays in many guises. *Bioessays*, 21(9) :781–90, 1999.
- [12] Tucker CL, Gera JF, and Uetz P. Towards an understanding of complex protein networks. *Trends Cell Biol*, 11(3) :102–6, 2001.
- [13] Jeff Hasty, Farren Isaacs, Milos Dolnik, David McMillen, and J. J. Collins. Designer gene networks : Towards fundamental cellular control. *Chaos*, 11(1) :207–220, 2001.
- [14] Michael A. A. Savageau. Design principles for elementary gene circuits : Elements, methods, and examples. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 11(1) :142–159, March 2001.
- [15] Locke JC, Southern MM, Kozma-Bognar L, Hibberd V, Brown PE, Turner MS, and Millar AJ. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology*, 1 :2005.0013, 2005.

- [16] Jean Clairambault, Philippe Michel, and Benoît Perthame. Circadian rhythm and tumour growth. *C.R. Acad. Sci.*, 342(1) :17–22, 2006.
- [17] Johan Paulsson. Models of stochastic gene expression. *Physics of Life Reviews*, 2(2) :157 – 175, 2005.
- [18] J. D. Murray. *Mathematical Biology : I. An Introduction (Interdisciplinary Applied Mathematics) (Pt. 1)*. Interdisciplinary applied mathematics. Springer, 3rd edition, January 2007.
- [19] Christopher P. Fall, Eric S. Marland, and John M. Wagner. *Computational Cell Biology*. Springer, Berlin, February 2002.
- [20] R. Edwards, H. T. Siegelmann, K. Aziza, and L. Glass. Symbolic dynamics and computation in model gene networks. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 11(1) :160–169, 2001.
- [21] René Thomas. Regulatory networks seen as asynchronous automata : A logical description. *Journal of Theoretical Biology*, 153(1) :1 – 23, 1991.
- [22] E. Atlee Jackson. *Perspective of nonlinear dynamics*. Cambridge University Press, first edition edition, 1991.
- [23] J. Hale and H. Koça. *Dynamics and Bifurcations*. Springer-Verlag, New York, 1991.
- [24] Steven H. Strogatz. *Nonlinear Dynamics And Chaos : With Applications To Physics, Biology, Chemistry, And Engineering (Studies in Nonlinearity)*. Studies in nonlinearity. Perseus Books Group, 1 edition, January 1994.
- [25] James E. Ferrell Jr., Joseph R. Pomerening, Sun Young Kim, Nikki B. Trunnell, Wen Xiong, Chi-Ying Frederick Huang, and Eric M. Machleder. Simple, realistic models of complex biological processes : Positive feedback and bistability in a cell fate switch and a cell cycle oscillator. *FEBS Letters*, 583(24) :3999–4005, 2009.
- [26] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. i. structural conditions of multistationarity and other nontrivial behavior. *Chaos*, 11(1) :170–179, 2001.
- [27] Bela Novak, John J. Tyson, Bela Gyorffy, and Attila Csikasz-Nagy. Irreversible cell-cycle transitions are due to systems- level feedback. *Nat Cell Biol.*, 9(7) :724–728, 2007.
- [28] DAVID E. CLAPHAM. The g-protein nanomachine. *Nature*, 379 :297–299, 196.
- [29] JB Hurley, MI Simon, DB Teplow, JD Robishaw, and AG Gilman. Homologies between signal transducing g proteins and ras gene products. *Science*, 226(4676) :860–862, 1984.
- [30] D. Thieffry and H. De Jong. Modélisation, analyse et simulation des réseaux génétiques. *Médecine/Sciences*, 18(4) :492–502, 2002.
- [31] P. François. A model for the neurospora circadian clock. *Biophysical Journal*, 88(4) :2369–2383, 2005.
- [32] Paul François. *Réseaux génétiques : conception, modélisation et dynamique*. PhD thesis, Université Paris VII, 2005.
- [33] Ido Golding and Edward C. Cox. Eukaryotic transcription : What does it mean for a gene to be 'on' ? *Current Biology*, 16(10) :R371–R373, 2006.

- [34] Tom Chou. Ribosome recycling, diffusion, and mrna loop formation in translational regulation. *Biophys J.*, 85(2) :755–773, 2003.
- [35] Athel Cornish-Bowden. *Cinétique enzymatique*. Edp sciences, 2005.
- [36] Gen Kurosawa and Yoh Iwasa. Saturation of enzyme kinetics in circadian clock models. *Journal of Biological Rhythms*, 17(6) :568–577, 2002.
- [37] Béla Novák and John J. Tyson. Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol.*, 9(12) :981–991, 2008.
- [38] Jean-Christophe Leloup and Albert Goldbeter. A model for circadian rhythms in drosophila incorporating the formation of a complex between the per and tim proteins. *Journal of Biological Rhythms*, 13(1) :70–87, 1998.
- [39] A. Goldbeter. A model for circadian oscillations in the drosophila period protein (per). *Proc Biol Sci*, 261(1362) :319–324, 1995.
- [40] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25) :2340–2361, 1977.
- [41] NG Van Kampen. *Stochastic processes in physics and chemistry*. North Holland, 2007.
- [42] Michail Stamatakis and Nikos V Mantzaris. Comparison of deterministic and stochastic models of the lac operon genetic network. *Biophysical Journal*, 96(3) :887–906, 2009.
- [43] John J. Tyson, Christian I. Hong, C. Dennis Thron, and Bela Novak. A simple model of circadian rhythms based on dimerization and proteolysis of per and tim. *Biophysical Journal*, 77(4) :2411–2417, 1999.
- [44] Nicolas E. Buchler, Ulrich Gerland, and Terence Hwa. Nonlinear protein degradation and the function of genetic circuits. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27) :9559–9564, 2005.
- [45] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs : Simple building blocks of complex networks. *Science*, 298(5594) :824–827, 2002.
- [46] Uzman Akif. Genes and signals : Ptashne, m., gann, a. *Biochemistry and Molecular Biology Education*, 30(5) :340–341, 2002.
- [47] G Tiana, S Krishna, S Pigolotti, M H Jensen, and K Sneppen. Oscillations and temporal signalling in cells. *Physical Biology*, 4(2) :R1, 2007.
- [48] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794) :378–382, 2000.
- [49] Takashi Ito, Tomoko Chiba, Ritsuko Ozawa, Mikio Yoshida, Masahira Hattori, and Yoshiyuki Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences*, 98(8) :4569–4574, 2001.
- [50] P Uetz, L Giot, G Cagney, T A Manseld, R S Judson, J R Knight, D Lockshon, V Narayan, M Srinivasan, P Pochart, A Qureshi-Emili, Y Li, B Godwin, D Conover, T Kalbeisch, G Vijayadamodar, M Yang, M Johnston, S Fields, and J M Rothberg. A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae. *Nature*, 403(6770) :601–603, 2000.
- [51] Nitzan Rosenfeld, Michael B Elowitz, and Uri Alon. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology*, 323(5) :785 – 793, 2002.

- [52] M. A. Savageau. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*, 252(5484) :546–549, 1974.
- [53] Becskei A. and Serrano L. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786) :590–593, 2000.
- [54] D. Thieffry, A. M. Huerta, E. Pérez-Rueda, and J. Collado-Vides. From specific gene regulation to genomic networks : a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays*, 20(5) :433–440, 1998.
- [55] Tong Ihn Lee, Nicola J. Rinaldi, François Robert, Duncan T. Odom, Ziv Bar-Joseph, Georg K. Gerber, Nancy M. Hannett, Christopher T. Harbison, Craig M. Thompson, Itamar Simon, Julia Zeitlinger, Ezra G. Jennings, Heather L. Murray, D. Benjamin Gordon, Bing Ren, John J. Wyrick, Jean-Bosco Tagne, Thomas L. Volkert, Ernest Fraenkel, David K. Gifford, and Richard A. Young. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298(5594) :799–804, 2002.
- [56] Francisco M. Camas, Jesús Blázquez, and Juan F. Poyatos. Autogenous and nonautogenous control of response in a genetic network. *Proceedings of the National Academy of Sciences*, 103(34) :12718–12723, 2006.
- [57] Yann Dublanche, Konstantinos Michalodimitrakis, Nico Kümmerer, Mathilde Fogliarini, and Luis Serrano. Noise in transcription negative feedback loops : simulation and experimental analysis. *Molecular systems biology*, 2, 2006.
- [58] Kalir S., Mangan S., and Alon U. A coherent feed-forward loop with a sum input function prolongs flagella production in *Escherichia coli*. *Mol. Syst. Biol.*, 1, 2006.
- [59] Yusuke T. Maeda and Masaki Sano. Regulatory dynamics of synthetic gene networks with positive feedback. *Journal of Molecular Biology*, 359(4) :1107 – 1124, 2006.
- [60] Denise M Wolf and Adam P Arkin. Motifs, modules and games in bacteria. *Current Opinion in Microbiology*, 6(2) :125 – 134, 2003.
- [61] Becskei A., Seraphin B., and Serrano L. Positive feedback in eukaryotic gene networks : cell differentiation by graded to binary response conversion. *EMBO J.*, 20(10) :2528–2535, 2001.
- [62] Beat P. Kramer and Martin Fussenegger. Hysteresis in a synthetic mammalian gene network. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27) :9517–9522, 2005.
- [63] Smolen P, Baxter DA, and Byrne JH. Mathematical modeling of gene networks. *Neuron*, 26(3) :567–580, 2000.
- [64] Uri Alon. Network motifs : theory and experimental approaches. *Nat Rev Genet*, 8(6) :450–461, 2007.
- [65] Simone Pigolotti, Sandeep Krishna, and Mogens H. Jensen. Oscillation patterns in negative feedback loops. *Proceedings of the National Academy of Sciences*, 104(16) :6533–6537, 2007.
- [66] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet*, 31(1) :64–68, 2002.
- [67] Duncan T. Odom, Nora Zizlsperger, D. Benjamin Gordon, George W. Bell, Nicola J. Rinaldi, Heather L. Murray, Tom L. Volkert, Jörg Schreiber, P. Alexander Rolfe, David K.

- Gifford, Ernest Fraenkel, Graeme I. Bell, and Richard A. Young. Control of pancreas and liver gene expression by hnf transcription factors. *Science*, 303(5662) :1378–1381, 2004.
- [68] Louis A. Saddic, Bärbel Huvermann, Staver Bezhani, Yanhui Su, Cara M. Winter, Chang Seob Kwon, Richard P. Collum, and Doris Wagner. The leafy target *lmi1* is a meristem identity regulator and acts together with leafy to regulate expression of cauliflower. *Development*, 133(9) :1673–1682, 2006.
- [69] Boyer L.A. and et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6) :947–956, 2005.
- [70] Esti Yeger-Lotem, Shmuel Sattath, Nadav Kashtan, Shalev Itzkovitz, Ron Milo, Ron Y. Pinter, Uri Alon, and Hanah Margalit. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16) :5934–5939, 2004.
- [71] Paul François and Vincent Hakim. Core genetic module : The mixed feedback loop. *Phys. Rev. E*, 72(3) :031908, Sep 2005.
- [72] Paul François and Vincent Hakim. Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences of the United States of America*, 101(2) :580–585, 2004.
- [73] J. Dunlap. Molecular Bases for Circadian Clocks. *Cell*, 96(2) :271–290, January 1999.
- [74] J J Tyson. Modeling the cell division cycle : *cdc2* and cyclin interactions. *Proceedings of the National Academy of Sciences*, 88(16) :7328–7332, 1991.
- [75] Katherine C. Chen, Attila Csikasz-Nagy, Bela Gyorffy, John Val, Bela Novak, and John J. Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Molecular Biology of the Cell*, 11(1) :369–391, 2000.
- [76] Naama Geva-Zatorsky, Nitzan Rosenfeld, Shalev Itzkovitz, Ron Milo, Alex Sigal, Erez Dekel, Talia Yarnitzky, Yuvalal Liron, Paz Polak, Galit Lahav, and Uri Alon. Oscillations and variability in the p53 system. *Molecular Systems Biology*, 2(1), June 2006.
- [77] Galit Lahav, Nitzan Rosenfeld, Alex Sigal, Naama Geva-Zatorsky, Arnold J. Levine, Michael B. Elowitz, and Uri Alon. Dynamics of the p53-mdm2 feedback loop in individual cells. *Nat Genet*, 36(2) :147–150, 2004.
- [78] John J Tyson, Katherine C Chen, and Bela Novak. Sniffers, buzzers, toggles and blinkers : dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2) :221 – 231, 2003.
- [79] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403 :335–338, 2000.
- [80] N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, 403(6767) :267–268, January 2000.
- [81] Pierre-Emmanuel Morant, Quentin Thommen, Benjamin Pfeuty, Constant Vandermoere, Florence Corellou, Francois-Yves Bouget, and Marc Lefranc. A robust two-gene oscillator at the core of *ostreococcus tauri* circadian clock. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 20(4) :045108, 2010.
- [82] A.T. Winfree. *The geometry of biological time*. Interdisciplinary applied mathematics : Mathematical biology. Springer, 2001.

- [83] G.B. Ermentrout and J. Rinzel. Beyond a pacemaker's entrainment limit : phase walk-through. *Am J Physiol Regulatory Integrative Comp Physiol*, 246 :102–106, 1984.
- [84] Michael W. Young and Steve A. Kay. Time zone : a comparative genetics of circadian clocks. *Nature Genetics*, 2 :702–715, 2001.
- [85] Marcelo J. Yanovsky and Steve A. Kay. Living by the calendar : how plants know when to flower. *Nat Rev Mol Cell Biol*, 4(4) :265–275, 2003.
- [86] Bruce VG. The biological clock in *clamydomonas reinhardtii*. *J Protozool*, 17 :328–334, 1970.
- [87] El Batoul Djouani-Tahri, Jean-Paul Motta, François-Yves Bouget, and Florence Corellou. Insights into the regulation of the core clock component *toc1* in the green picoeukaryote *ostreococcus*. *Plant Signaling and Behaviour*, 5(3) :332–335, 2010.
- [88] Michael W. Young. Big ben rings in a lesson on biological clocks. *Neuron*, 36(6) :1001 – 1005, 2002.
- [89] AJ Millar, IA Carre, CA Strayer, NH Chua, and SA Kay. Circadian clock mutants in *arabidopsis* identified by luciferase imaging. *Science*, 267(5201) :1161–1163, 1995.
- [90] C. Brandes, J. D. Plautz, R. Stanewsky, C. F. Jamison, M. Straume, K. V. Wood, S. A. Kau, and J. C. Hall. Novel features of *drosophila* period transcription revealed by real-time luciferase reporting. *Neuron*, 16(4) :687–692, 1996.
- [91] Didier Gonze, Jean-Christophe Leloup, and Albert Goldbeter. Theoretical models for circadian rhythms in *neurospora* and *drosophila*. *C.R. Acad. Sci.*, 323(1) :57–67, 2000.
- [92] Mickael Moulager, Annabelle Monnier, Béline Jesson, Régis Bouvet, Jean Mosser, Christian Schwartz, Lionel Garnier, Florence Corellou, and François-Yves Bouget. Light-dependent regulation of cell division in *ostreococcus* : Evidence for a major transcriptional input. *Plant Physiology*, 144(3) :1360–1369, 2007.
- [93] Florence Corellou, Christian Schwartz, Jean-Paul Motta, El Batoul Djouani-Tahri, Frederic Sanchez, and Francois-Yves Bouget. Clocks in the green lineage : Comparative functional analysis of the circadian architecture of the picoeukaryote *ostreococcus*. *Plant Cell*, 21 :3436–3449, 2009.
- [94] Quentin Thommen, Benjamin Pfeuty, Pierre-Emmanuel Morant, Florence Corellou, François-Yves Bouget, and Marc Lefranc. Robustness of circadian clocks to daylight fluctuations : Hints from the picoeukaryote *ostreococcus tauri*. *PLoS Comput Biol*, 6(11) :e1000990, 11 2010.
- [95] Andrew C Liu, Warren G Lewis, and Steve A Kay. Mammalian circadian signaling networks and therapeutic targets. *Nat Chem Biol*, 3 :630–639, 2007.
- [96] Kong L. Toh, Christopher R. Jones, Yan He, Erik J. Eide, William A. Hinz, David M. Virshup, Louis J. Ptacek, and Ying-Hui Fu. An *hper2* phosphorylation site mutation in familial advanced sleep phase syndrome. *Science*, 291(5506) :1040–1043, 2001.
- [97] Arkady Pikovsky, Michael Rosenblum, and Jürgen Kurths. *Synchronization : A Universal Concept in Nonlinear Science*. Cambridge University Press, January 2002.
- [98] A.H. Nayfeh. *Introduction to perturbation techniques*. Wiley classics library. Wiley, 1981.

- [99] Tony Yu-Chen Tsai, Yoon Sup Choi, Wenzhe Ma, Joseph R. Pomeroy, Chao Tang, and James E. Ferrell. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science*, 321(5885) :126–129, 2008.
- [100] J.C.W. Locke, A.J. Millar, and M.S. Turner. Modelling genetic networks with noisy and varied experimental data : the circadian clock in arabidopsis thaliana. *J. Theor. Biol.*, 234(3) :383 – 393, 2005.
- [101] O Olsvik, J Wahlberg, B Petterson, M Uhlén, T Popovic, I K Wachsmuth, and P I Fields. Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit b in vibrio cholerae o1 strains. *Journal of Clinical Microbiology*, 31(1) :22–25, 1993.
- [102] Neil Hall. Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology*, 210(9) :1518–1525, 2007.
- [103] Evelyne Derelle, Conchita Ferraz, Stephane Rombauts, Pierre Rouzé, Alexandra Z. Worden, Steven Robbens, Frédéric Partensky, Sven Degroeve, Sophie Echeynié, Richard Cooke, Yvan Saeys, Jan Wuyts, Kamel Jabbari, Chris Bowler, Olivier Panaud, Benoît Piégu, Steven G. Ball, Jean-Philippe Ral, François-Yves Bouget, Gwenael Piganeau, Bernard De Baets, André Picard, Michel Delseny, Jacques Demaille, Yves Van de Peer, and Hervé Moreau. Genome analysis of the smallest free-living eukaryote ostreococcus tauri unveils many unique features. *Proceedings of the National Academy of Sciences*, 103(31) :11647–11652, 2006.
- [104] Analysis of the genome sequence of the flowering plant arabidopsis thaliana. *Nature*, 408(6814) :796–815, 2000.
- [105] Roger D. Canales, Yuling Luo, James C. Willey, Bradley Austermler, Catalin C. Barbacioru, Cecilie Boysen, Kathryn Hunkapiller, Roderick V. Jensen, Charles R. Knight, Kathleen Y. Lee, Yunqing Ma, Botoul Maqsoodi, Adam Papallo, Elizabeth Herness Peters, Karen Poulter, Patricia L. Ruppel, Raymond R. Samaha, Leming Shi, Wen Yang, Lu Zhang, and Federico M. Goodsaid. Evaluation of dna microarray results with quantitative gene expression platforms. *Nature Biotechnology*, 24(9) :1115–1122, 2006.
- [106] O. Shimomura, F. H. Johnson, and Y. Saiga. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of cellular and comparative physiology*, 59 :223–239, June 1962.
- [107] David K. Welsh and Steve A. Kay. Bioluminescence imaging in living organisms. *Current Opinion in Biotechnology*, 16(1) :73 – 78, 2005. Analytical biotechnology.
- [108] Greer L.F. III and Szalay AA. Imaging of light emission from the expression of luciferases in living cells and organisms : a review. *Luminescence*, 17(1) :43–74, 2002.
- [109] Wessel Van Leeuwen, Marc Hagendoorn, Tom Ruttink, Remco Van Poecke, Linus Van Der Plas, and Alexander Van Der Krol. The use of the luciferase reporter system for in planta gene expression studies. *Plant Molecular Biology Reporter*, 18(2) :143a–143t, 2000. 10.1007/BF02824024.
- [110] De Wet J.R., Wood K.V., Deluca M., Helinski D.R., and Subramani S. Firefly luciferase gene : structure and expression in mammalian cells. *Mol. Cell Biol.*, 7(2) :725–737, 1987.

- [111] R. Mayerhofer, K. Araki, and A. A. Szalay. Monitoring of spatial expression of firefly luciferase in transformed zebrafish. *Journal of Bioluminescence and Chemiluminescence*, 10(5) :271–275, 1995.
- [112] DAVID W. OW, JEFFREY R. DE WET, DONALD R. HELINSKI, STEPHEN H. HOWELL, KEITH V. WOOD, and MARLENE DELUCA. Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science*, 234(4778) :856–859, 1986.
- [113] Jolene M. Ignowski and David V. Schaffer. Kinetic analysis and modeling of firefly luciferase as a quantitative reporter gene in live mammalian cells. *Biotechnology and Bioengineering*, 86(7) :827–834, 2004.
- [114] Dumollard R., Marangos P., Fitzharris G., Swann K., Duchen M., and Carroll J. Sperm-triggered $[ca^{2+}]$ oscillations and ca^{2+} homeostasis in the mouse egg have an absolute requirement for mitochondrial atp production. *Development*, 131(13) :3057–3067, 2004.
- [115] Yao Xu, Akihito Kanauchi, Albrecht G. von Arnim, David W. Piston, and Carl Hirschie Johnson. [12] bioluminescence resonance energy transfer : Monitoring protein-protein interactions in living cells. In Ian Parker Gerard Marriott, editor, *Biophotonics, Part A*, volume 360 of *Methods in Enzymology*, pages 289 – 301. Academic Press, 2003.
- [116] Choy G., Choyke P., and Libutti S.K. Current advances in molecular imaging : noninvasive in vivo bioluminescent and fluorescent optical imaging in cancer research. *Mol Imaging*, 2(4) :303–312, 2003.
- [117] K. Shah, A. Jacobs, X.O. Breakefield, and R. Weissleder. Imaging of gene therapy for cancer. *Gene Ther*, 11(15) :1175–1187, 2004.
- [118] Timothy C. Doyle, Stacy M. Burns, and Christopher H. Contag. Technoreview : In vivo bioluminescence imaging for integrated studies of infection. *Cellular Microbiology*, 6(4) :303–317, 2004.
- [119] Ciana P., Raviscioni M., Mussi P. Vegeto E., Que I., Parker M.G., Lowik C., and Maggi A. In vivo imaging of transcriptionally active estrogen receptors. *Nat. Med.*, 9(1) :82–86, 2003.
- [120] Zhang G.J., Safran M., Wei W., Sorensen E., Lassota P., Zhelev N., Neuberger D.S., Shapiro G., and Kaelin W.G. Jr. Bioluminescent imaging of cdk2 inhibition in vivo. *Nat Med*, 10(6) :643–648, 2004.
- [121] J R de Wet, K V Wood, D R Helinski, and M DeLuca. Cloning of firefly luciferase cDNA and the expression of active luciferase in escherichia coli. *Proceedings of the National Academy of Sciences*, 82(23) :7870–7873, 1985.
- [122] Welsh David K., Yoo Seung-Hee, Liu Andrew C., Takahashi Joseph S., and Kay Steve A. Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Current biology*, 14(24) :2289–2295, 2004.
- [123] Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6) :1025–1036, 2005.
- [124] Chubb J.R., Treck T., Shenoy S.M., and Singer R.H. Transcriptional pulsing of a developmental gene. *Current Biology*, 16(10) :1018–1025, 2006.

- [125] Pierre-Emmanuel Morant, Quentin Thommen, François Lemaire, Constant Vandermoëre, Benjamin Parent, and Marc Lefranc. Oscillations in the expression of a self-repressed gene induced by a slow transcriptional dynamics. *Phys. Rev. Lett.*, 102(6) :068104, Feb 2009.
- [126] J. S. Griffith. Mathematics of cellular control processes i. negative feedback to one gene. *Journal of Theoretical Biology*, 20(2) :202 – 208, 1968.
- [127] Brian C. Goodwin. Oscillatory behavior of enzymatic control processes. *Advances in Enzyme Regulation*, 3 :425–428, 1965.
- [128] Nicholas A.M. Monk. Oscillatory expression of hes1, p53, and nf-[kappa]b driven by transcriptional time delays. *Current Biology*, 13(16) :1409 – 1413, 2003.
- [129] G. Tiana, M.H. Jensen, and K. Sneppen. Time delay as a key to apoptosis induction in the p53 network. *The European Physical Journal B - Condensed Matter and Complex Systems*, 29(1) :135–140, 2002. 10.1140/epjb/e2002-00271-1.
- [130] M.H Jensen, K Sneppen, and G Tiana. Sustained oscillations and time delays in gene expression of protein hes1. *FEBS Letters*, 541(1) :176–177, 2003.
- [131] Jensen M., Sneppen K., and Tiana G. Oscillating gene expressions in regulatory networks. In A. T. Skjeltorp and A. V. Belushkin, editors, *Forces, Growth and Form in Soft Condensed Matter : At the Interface between Physics and Biology*, volume 160 of *NATO Science Series II : Mathematics, Physics and Chemistry*, pages 195–202. Springer Netherlands, 2005.
- [132] Wang Jingkui, Thommen Quentin, and Lefranc Marc. Oscillations d’expression d’un gène auto-régulé : interaction d’un délai de transport avec la réponse transcriptionnelle. *Comptes-rendus de la 13e Rencontres du Non-Linéaire*, 13 :211, 2010.
- [133] François Boulter, Marc Lefranc, François Lemaire, Pierre-Emmanuel Morant, and Asli Ürgüplü. On proving the absence of oscillations in models of genetic circuits. In Hirokazu Anai, Katsuhisa Horimoto, and Temur Kutsia, editors, *Algebraic Biology*, volume 4545 of *Lecture Notes in Computer Science*, pages 66–80. Springer Berlin / Heidelberg, 2007.
- [134] P. E. Hardin, J. C. Hall, and M. Rosbash. Feedback of the drosophila period gene product on circadian cycling of its messenger rna levels. *Nature*, 343(6258) :536–540, 1990.
- [135] MC Mackey and L Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300) :287–289, 1977.
- [136] Julian Lewis. Autoinhibition with transcriptional delay : A simple mechanism for the zebrafish somitogenesis oscillator. *Current Biology*, 13(16) :1398 – 1408, 2003.
- [137] Richard D. Bliss, Page R. Painter, and Allen G. Marr. Role of feedback inhibition in stabilizing the classical operon. *Journal of Theoretical Biology*, 97(2) :177 – 193, 1982.
- [138] Hiromi Hirata, Shigeki Yoshiura, Toshiyuki Ohtsuka, Yasumasa Bessho, Takahiro Harada, Kenichi Yoshikawa, and Ryoichiro Kageyama. Oscillatory expression of the bhlh factor hes1 regulated by a negative feedback loop. *Science*, 298(5594) :840–843, 2002.
- [139] I Edery, L J Zwiebel, M E Dembinska, and M Rosbash. Temporal phosphorylation of the drosophila period protein. *Proceedings of the National Academy of Sciences*, 91(6) :2260–2264, 1994.

- [140] Curtin K.D., Huang Z.J., and Rosbash M. Temporally regulated nuclear entry of the *Drosophila* period protein contributes to the circadian clock. *Neuron*, 14(2) :365–372, 1995.
- [141] Huang Z.J., Edery I., and Rosbash M. Pas is a dimerization domain common to *Drosophila* period and several transcription factors. *Nature*, 364(6434) :259–262, 1993.
- [142] M. Kaern, T.C. Elston, W.J. Blake, and J.J. Collins. Stochasticity in gene expression : from theories to phenotypes. *Nat. Rev. Genet.*, 6(6) :451–464, 2005.
- [143] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer. In vivo dynamics of rna polymerase ii transcription. *Nat. Struct. Mol. Biol.*, 14(9) :796–806, 2007.
- [144] Andrea M. Femino, Fredric S. Fay, Kevin Fogarty, and Robert H. Singer. Visualization of single rna transcripts in situ. *Science*, 280(5363) :585–590, 1998.
- [145] Azi Lipshtat, Adiel Loinger, Nathalie Q. Balaban, and Ofer Biham. Genetic toggle switch without cooperative binding. *Phys. Rev. Lett.*, 96(18) :188101, May 2006.
- [146] M. J. Morelli, S. Tanase-Nicola, R. J. Allen, and P. R. ten Wolde. Reaction coordinates for the flipping of genetic switches. *Biophys. J.*, 94(9) :3413–3423, 2008.
- [147] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, second edition, August 1983.
- [148] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products, Fifth Edition*. Academic Press, 5th edition, January 1994.
- [149] J. S. Griffith. Mathematics of cellular control processes ii. positive feedback to one gene. *Journal of Theoretical Biology*, 20(2) :209–216, 1968.
- [150] Mickael Moulager, Florence Corellou, Valérie Vergé, Marie-Line Escande, and François-Yves Bouget. Integration of light signals by the retinoblastoma pathway in the control of s phase entry in the picophytoplanktonic cell *Ostreococcus*. *PLoS Genet*, 6(5) :e1000957, 05 2010.
- [151] Annabelle Monnier, Silvia Liverani, Régis Bouvet, Béline Jesson, Jim Q Smith, Jean Mosser, Florence Corellou, and François-Yves Bouget. Orchestrated transcription of biological processes in the marine picoeukaryote *ostreococcus* exposed to light/dark cycles. *BMC Genomics*, 11(192), 2010.
- [152] C. Courties, A. Vaquer, M. Troussellier, J. Lautier, M.J. Chretiennot-Dinet, J. Neveux, M.C. Machado, and H. Claustre. Smallest eukaryotic organism. *Nature*, 370(6487) :255–255, 1994.
- [153] M-J. Chretiennot-Dinet, C. Courties, A. Vaquer, J. Neveux, H. Claustre, J. Lautier, and M. C. Machado. A new marine pico eukaryote : *Ostreococcus tauri* gen. et sp. nov. (chlorophyta, prasinophyceae). *Phycologia*, 34(4) :285–292, 1995.
- [154] Fabrice Not, Mikel Latasa, Dominique Marie, Thierry Cariou, Daniel Vaultot, and Nathalie Simon. A single species, *micromonas pusilla* (prasinophyceae), dominates the eukaryotic picoplankton in the western english channel. *Appl. Environ. Microbiol.*, 70(7) :4064–4072, 2004.
- [155] Manon Viprey, Laure Guillou, Martial Ferréol, and Daniel Vaultot. Wide genetic diversity of picoplanktonic green algae (chloroplastida) in the mediterranean sea uncovered by a phylum-biased pcr approach. *Environmental Microbiology*, 10(7) :1804–1822, 2008.

- [156] Andrew J. Millar. Tansley review : Biological clocks in arabidopsis thaliana. *New Phytol.*, 141(2) :175–197, 1999.
- [157] H. G. McWatters, R. M. Bastow, A. Hall, and A. J. Millar. The elf3 zeitnehmer regulates light signalling to the circadian clock. *Nature*, 408(6813) :716–720, 2000.
- [158] Vijay Chickarmane, Boris N. Kholodenko, and Herbert M. Sauro. Oscillatory dynamics arising from competitive inhibition and multisite phosphorylation. *Journal of Theoretical Biology*, 244(1) :68 – 76, 2007.
- [159] Wessel van Leeuwen, Tom Ruttink, Antoinette W. M. Borst $\frac{3}{4}$ Vrenszen, Linus H. W. van der Plas, and Alexander R. van der Krol. Characterization of position $\frac{3}{4}$ induced spatial and temporal regulation of transgene promoter activity in plants. *Journal of Experimental Botany*, 52(358) :949–959, 2001.
- [160] Attila Becskei, Benjamin B Kaufmann, and Alexander van Oudenaarden. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat Genet*, 37(9) :937–944, 2005.
- [161] Mukund Thattai and Alexander van Oudenaarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167(1) :523–530, 2004.
- [162] David Alabadi, Tokitaka Oyama, Marcelo J. Yanovsky, Franklin G. Harmon, Paloma Más, and Steve A. Kay. Reciprocal regulation between toc1 and lhy/ccal within the arabidopsis circadian clock. *Science*, 293(5531) :880–883, 2001.
- [163] Stacey L. Harmer and Steve A. Kay. Positive and negative factors confer phase-specific circadian regulation of transcription in arabidopsis. *The Plant Cell Online*, 17(7) :1926–1940, 2005.
- [164] K.E. Gordon, I.M.M. van Leeuwen, S. Lain, and M.A.J. Chaplain. Spatio-temporal modelling of the p53-mdm2 oscillatory system. *Mathematical Modelling of Natural Phenomena*, 4 :97–116, 2009.
- [165] Thomas R. *Quantum Noise*, volume 9. Springer,Berlin, 2004.
- [166] Daniel E. Zak, Jörg Stelling, and Francis J. Doyle III. Sensitivity analysis of oscillatory (bio)chemical systems. *Computers and Chemical Engineering*, 29(3) :663 – 673, 2005. Computational Challenges in Biology.
- [167] Amanda-Jayne F. Carr and David Whitmore. Imaging of single light-responsive clock cells reveals fluctuating free-running periods. *Nat Cell Biol*, 7(3) :319–321, 2005.
- [168] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C (2nd ed.) : the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1992.
- [169] Mark H. Holmes. *Introduction to Perturbation Methods*. Number 20 in Texts in Applied Mathematics. Springer-Verlag, New York, NY, 1995.