

Une horloge circadienne minimale chez l'algue unicellulaire *Ostreococcus tauri*

P.-E. Morant^{1,2}, Q. Thommen^{1,2}, F. Corellou³, C. Vandermoère^{1,2}, F.-Y. Bouget³, & M. Lefranc^{1,2}

¹ Laboratoire de Physique des Lasers, Atomes, Molécules, UMR CNRS 8523, Université Lille 1, F-59655 Villeneuve d'Ascq Cedex, France

² Institut de Recherche Interdisciplinaire sur les Réseaux de Régulation Cellulaires, USR CNRS 3078, Université Lille 1, F-59655 Villeneuve d'Ascq Cedex, France

³ Laboratoire Modèles en Biologie Cellulaire et Evolutive, UMR CNRS-Paris 6 7628, Observatoire Océanologique de Banyuls sur mer, BP44, 66651 Banyuls sur Mer Cedex, France.

marc.lefranc@univ-lille1.fr

Résumé. Les horloges circadiennes sont des oscillateurs biologiques, constitués de réseaux de gènes et protéines se régulant de manière cyclique, qui permettent à de nombreux organismes de se synchroniser au cycle jour/nuit. Nous avons étudié les profils temporels d'activité de deux gènes putatifs de l'horloge de l'algue *Ostreococcus tauri*, reconstruits à partir de données expérimentales. Ces profils sont reproduits avec un accord excellent par le modèle mathématique d'un oscillateur minimal dans lequel ces deux gènes se régulent réciproquement. De manière contre-intuitive, le meilleur accord est obtenu pour un modèle non couplé à la lumière, ce qui ouvre des voies dans la compréhension de la structure de l'horloge.

Abstract. Circadian clocks are biological oscillators, made of networks of genes and proteins regulating each other cyclically, that allow many organisms to synchronize to the day/night cycle. We have studied temporal activity profiles of two putative genes of the clock of the *Ostreococcus tauri* alga, reconstructed from experimental data. These profiles are reproduced with excellent agreement by the mathematical model of a minimal oscillator, where these two genes regulate each other. Surprisingly, best agreement is obtained when the model is uncoupled to light, which gives insights into the structure of the clock.

1 Introduction

Chez de nombreux organismes, activité, température corporelle, alternance veille/sommeil, photosynthèse, et bien d'autres processus biologiques sont rythmés par l'horloge *circadienne*, un oscillateur biologique dont la période est d'environ 24 heures. En se synchronisant au cycle jour/nuit, cette horloge permet d'anticiper les modifications périodiques de l'environnement et de s'y adapter. C'est sa désynchronisation qui est à l'origine du décalage horaire ressenti à l'arrivée d'un vol long-courrier.

Les rouages de cette horloge sont constitués de réseaux biochimiques où interagissent gènes et protéines. L'activité des gènes, qui gouvernent la synthèse des protéines pour lesquels ils codent, est en effet régulé par d'autres protéines. Les boucles de rétroaction positive et/ou négative ainsi formées sont à l'origine d'oscillations robustes dans ces circuits, de période généralement légèrement différente de 24 heures lorsque l'organisme est placé en conditions constantes. C'est le couplage d'un ou plusieurs acteurs de l'horloge à la lumière du jour (par exemple la dégradation accélérée d'une protéine au jour) qui permet à l'oscillateur de se synchroniser exactement au cycle jour/nuit et de donner en continu une mesure de l'heure qu'il est.

Ces dernières années, les principaux composants des horloges circadiennes de plusieurs organismes modèles (*Arabidopsis*, *Neurospora*, *Drosophila* ...) ont pu être identifiés [1,2]. Cela a permis de constater que si les acteurs moléculaires concernés varient fortement entre deux organismes éloignés sur l'arbre de l'évolution, leurs réseaux d'interaction présentent souvent des structures similaires, probablement parce que fournissant des réponses différentes à un même problème dynamique. Cela a motivé de nombreux travaux de modélisation (par ex. [3,4]), qui ont dans certains cas suggéré des expériences clés menant à l'identification de nouveaux gènes de l'horloge [5]. Les rythmes circadiens constituent donc un champ d'exploration particulièrement prometteur pour la biologie systémique, qui vise non seulement à appréhender

les comportements collectifs résultant de l'interaction des acteurs élémentaires, mais également à favoriser les allers-retours entre expérience et théorie par une modélisation quantitative. Cependant, l'implication de nombreux acteurs moléculaires, l'existence fréquente de redondances entre gènes, ainsi que l'interconnexion des différents modules fonctionnels rend le plus souvent difficile un accord quantitatif. C'est pourquoi il est important d'identifier des organismes où certains circuits génétiques se prêtent particulièrement bien à une modélisation mathématique.

2 Les rythmes circadiens d'*Ostreococcus tauri*

Nous nous intéressons ici à la modélisation de l'horloge circadienne d'un nouvel organisme modèle, *Ostreococcus tauri*, découvert en 1994 dans l'étang de Thau, près de Sète (Fig. 1, à gauche). Cette algue verte unicellulaire microscopique est à ce jour le plus petit organisme eukaryote connu. Le séquençage récent de son génome a mis en évidence sa remarquable compaction (environ 11 millions de paires de bases, comparable à celui de la levure) et une très faible redondance génique. Cet organisme présente d'autant plus d'intérêt que l'ancêtre commun d'*O. tauri* et de la plante modèle *Arabidopsis thaliana* se situe vraisemblablement à la base de la lignée verte, d'où sont issus tous les organismes photosynthétiques. Or, de nombreuses études se sont intéressées à l'horloge circadienne d'*A. thaliana*, mettant en évidence plusieurs boucles de rétroaction imbriquées, aux fonctions complémentaires. Il serait donc particulièrement intéressant de comparer deux évolutions d'une même horloge circadienne.

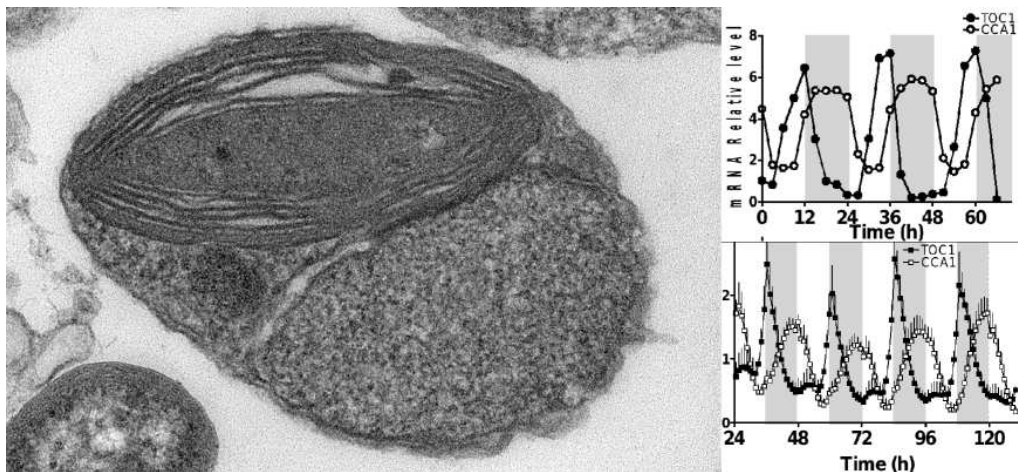


Fig. 1. A gauche : photographie au microscope électronique d'*Ostreococcus tauri*. A droite : données expérimentales concernant les gènes *TOC1* et *CCA1* [7] (en haut, données micropuces permettant de remonter aux variations de la quantité d'ARN en fonction du temps, en bas mesures de luminescence émises par des marqueurs bioluminescents permettant de remonter aux variations de la concentration en protéines).

Les rythmes circadiens d'*O. tauri* sont très marqués : les données d'expression de l'immense majorité des 8000 gènes varient de manière cyclique sur 24 heures, et les divisions cellulaires s'opèrent préférentiellement vers la tombée de la nuit, indiquant un contrôle du cycle cellulaire par l'horloge circadienne. Or, si environ 50 % des gènes d'*O. tauri* ont des homologues chez *A. Thaliana*, la recherche bioinformatique d'orthologues de la vingtaine de gènes impliqués dans l'horloge d'*Arabidopsis* n'identifie chez *O. Tauri* que deux gènes, apparentés aux gènes *CCA1* and *TOC1* d'*Arabidopsis* [7]. De plus, ce sont précisément ces deux gènes qui constituent l'ossature du modèle initialement proposé pour l'horloge d'*Arabidopsis* (*CCA1* réprime *TOC1* qui active *CCA1*) [6], jusqu'à ce qu'une analyse quantitative des données expérimentales ne démontre la nécessité de modèles plus complexes [4]. Si les pics d'activité de *TOC1* et *TOC1*-like sont relativement concordants chez les deux organismes, on observe que celui de *CCA1*-like intervient chez *Ostreococcus* au milieu de la nuit plutôt qu'au lever du jour. Or, c'est justement un intervalle de temps trop grand entre les pics qui invalide chez *Arabidopsis* le modèle de la boucle à

deux gènes. Tout cela conduit naturellement à se demander si ce modèle ne serait pas en fin de compte adapté à l'horloge d'*Ostreococcus* et à essayer de l'ajuster aux données expérimentales.

3 Données expérimentales

Nous utiliserons ici deux types de données expérimentales pour chacun des deux gènes. Les premières sont des mesures micropuces de la quantité d'ARN relativement à une référence, effectuées sur des extraits cellulaires prélevés toutes les trois heures (Fig. 1, en haut à droite) [7]. Ce sont des mesures difficilement reproductibles en raison de leur coût. Les deuxièmes nécessitent un travail important de constitution de lignées de cellules modifiées génétiquement, dans le génome desquelles ont été insérées un gène additionnel codant pour la protéine fonctionnelle (TOC1 ou CCA1) fusionnée à la luciférase, qui catalyse l'émission de photons par la luciférine. En fonction des constantes de temps des différents processus concernés, le signal de luminescence émis est proportionnel soit à la quantité de protéines nouvellement synthétisées, soit à la quantité totale de protéines, comme c'est le cas ici. Une fois les lignées stabilisées, le suivi automatisé de la variation de luminescence au cours d'une expérience permet donc d'obtenir des séries temporelles des variations de concentrations en protéines à intervalles d'une heure (Fig. 1, en bas à droite) [7]. D'autres lignées permettant de suivre l'activité transcriptionnelle des deux gènes au cours du temps ont également été construites mais ne sont pas utilisées ici [7].

Les données brutes ont été traitées afin de reconstruire des profils temporels en ARN et protéines qui puissent être ajustés par un modèle mathématique. Les données micropuces sont reproductibles avec une bonne précision, ce qui permet d'obtenir facilement un profil type, qui est ensuite interpolé avec des splines cubiques. Les données de luminescence présentent quant à elles d'importantes variations d'amplitude dans le temps, qui sont liées aux variations des conditions expérimentales, comme par exemple le nombre de cellules dans un puits ou leur arrangement. Nous avons ajusté les séries temporelles de luminescence par le produit d'une série de Fourier, représentant la composante périodique stationnaire que nous désirons étudier, par une fonction lentement variable correspondant aux variations expérimentales. Afin de se concentrer sur la dynamique à l'échelle de 24 heures de la boucle à deux gènes, et de laisser de côté les variations rapides qui correspondraient soit à des dynamiques transitoires soit tout simplement à du bruit, le nombre d'harmoniques de la série de Fourier est limité à 5.

4 Ajustement par un modèle minimal

Le modèle minimal de la boucle à deux gènes, où *CCA1* réprime *TOC1* qui active *CCA1*, est décrit par les quatre équations différentielles suivantes, où $M_{T,C}$ et $P_{T,C}$ représentent les concentrations en ARN et protéines pour chacun des deux gènes :

$$\dot{M}_T = \mu_T + \frac{\lambda_T}{1 + (P_C/P_{C_0})^{n_C}} - \delta_{M_T} \frac{K_{M_T} M_T}{K_{M_T} + M_T} \quad (1a)$$

$$\dot{P}_T = \beta_T M_T - \delta_{P_T} \frac{K_{P_T} P_T}{K_{P_T} + P_T} \quad (1b)$$

$$\dot{M}_C = \mu_C + \frac{\lambda_C (P_T/P_{T_0})^{n_T}}{1 + (P_T/P_{T_0})^{n_T}} - \delta_{M_C} \frac{K_{M_C} M_C}{K_{M_C} + M_C} \quad (1c)$$

$$\dot{P}_C = \beta_C M_C - \delta_{P_C} \frac{K_{P_C} P_C}{K_{P_C} + P_C} \quad (1d)$$

Les paramètres λ_X et μ_X représentent respectivement les fractions du taux de transcription du gène X régulée et non régulée par l'autre protéine, β_X est le taux de traduction de l'ARN en protéine, les P_{X_0} sont les seuils de protéines où les activations ou répressions transcriptionnelles sont à moitié effectives. Les termes négatifs décrivent une dégradation enzymatique classique de type Michaelis-Menten, où δ_Y représente le taux de dégradation à faible concentration et K_Y le seuil de saturation de la dégradation. La

présence d'une non-linéarité dans un ou plusieurs de ces termes de dégradation est essentielle pour l'existence d'oscillations auto-entretenues dans les équations (1), celles-ci disparaissant quand toutes constantes K_Y tendent simultanément vers l'infini.

Les profils temporels en ARN et protéines ne sont déterminés expérimentalement qu'à un facteur près. Deux jeux de paramètres différents qui mènent à des solutions se déduisant l'une de l'autre par une transformation d'échelle doivent donc être considérés comme équivalents du point de vue de l'ajustement. Nous sommes donc libres de fixer quatre paramètres arbitrairement, par exemple $\beta_C = \beta_T = 1$, $P_{C_0} = P_{T_0} = 100$, ce qui revient à fixer une échelle pour chacune des quatre variables des équations (1).

Afin que la procédure d'ajustement ne dépende pas outre mesure des détails de la méthode de reconstruction des profils temporels, nous avons défini la fonction de score comme étant l'erreur quadratique moyenne des temps de passage des quatre profils à 20 % et 80 % de leurs amplitudes maximales respectives, en supposant que le début de la montée et l'arrivée près du pic, ainsi que leurs équivalents à la descente, étaient des événements pertinents biologiquement. En pratique, l'ajustement sur ces temps de passage a en général mené à un bon ajustement sur toute la courbe.

Enfin, la procédure d'ajustement doit pouvoir faire varier un ou plusieurs paramètres des équations (1) entre le jour et la nuit (par exemple, une protéine peut voir sa dégradation s'accélérer au jour), car c'est précisément par ce mécanisme que l'horloge circadienne se synchronise au cycle jour/nuit. Il faut donc considérer successivement différents modes de couplage à la lumière, et les classer en fonction du meilleur score obtenu. On observe en pratique des solutions assez différentes selon les couplages retenus, et il est évidemment naturel d'espérer que l'on puisse par la modélisation mathématique obtenir des informations sur la voie d'entrée de la lumière.

5 Résultats

L'ajustement des quatre courbes expérimentales par le modèle (1) au moyen d'une exploration systématique de l'espace des paramètres mène à un accord surprenant entre données expérimentales et solutions numériques (Fig. 2, à gauche). On a supposé lors de cet ajustement que la période d'oscillation libre est de 24 heures, ce qui correspond à la moyenne des observations expérimentales. Cet accord est d'autant plus remarquable que les données ARN et protéine ont été acquises dans des expériences différentes, et par des techniques totalement distinctes. Les profils ARN sont reproduits avec une précision sans équivalent dans la littérature, et les caractéristiques principales des profils protéines sont également bien reproduites, notamment la position des pics de concentration. On note juste une difficulté pour le modèle théorique à reproduire le niveau bas de la protéine TOC1 (les simulations numériques le prédisent plus bas qu'il n'est observé expérimentalement) et le début de la montée de la protéine CCA1.

Il est difficile en l'état de pouvoir affirmer si ces désaccords mineurs sont dus à un modèle trop sommaire ou à des biais expérimentaux non pris en compte. Il serait naïf de croire que les équations (1) rendent compte de toute la réalité. Toujours est-il que si on ajuste les données ARN et protéines séparément, ce qui n'est pas illogique puisqu'il s'agit d'expériences différentes, on obtient des accords qui sont difficilement perfectibles compte tenu des difficultés d'obtention des données expérimentales (Fig. 2, à droite), même s'ils sont bien sûr obtenus pour des jeux de paramètres différents. Bien plus, on observe que l'ajustement simultané des données ARN et protéine, notamment pour ces dernières, s'améliore significativement si on suppose que le zéro des signaux de luminescence est biaisé et qu'on le considère comme paramètre libre de la procédure d'ajustement. Tout cela encourage à penser que l'accord de données expérimentales issues d'un organisme vivant, forcément complexe quelle que soit son relatif dépouillement, avec un modèle mathématique aussi caricatural ne peut être le fruit du hasard.

Mais il y a encore plus surprenant. En fait, les accords optimaux montrés à la figure 2 ont été obtenus pour des modèles non couplés à la lumière. Plus précisément, la procédure d'ajustement a, pour tous les modes de dépendance à l'éclairement possibles, systématiquement ramené la constante de couplage à 0 lors de l'optimisation. C'est un modèle d'oscillateur libre, décrit par les équations (1) sans aucune modulation des paramètres, qui reproduit le mieux des données expérimentales obtenues en réponse à une forte modulation de l'éclairement entre le jour et la nuit, tout au moins sur l'échelle de temps de 24

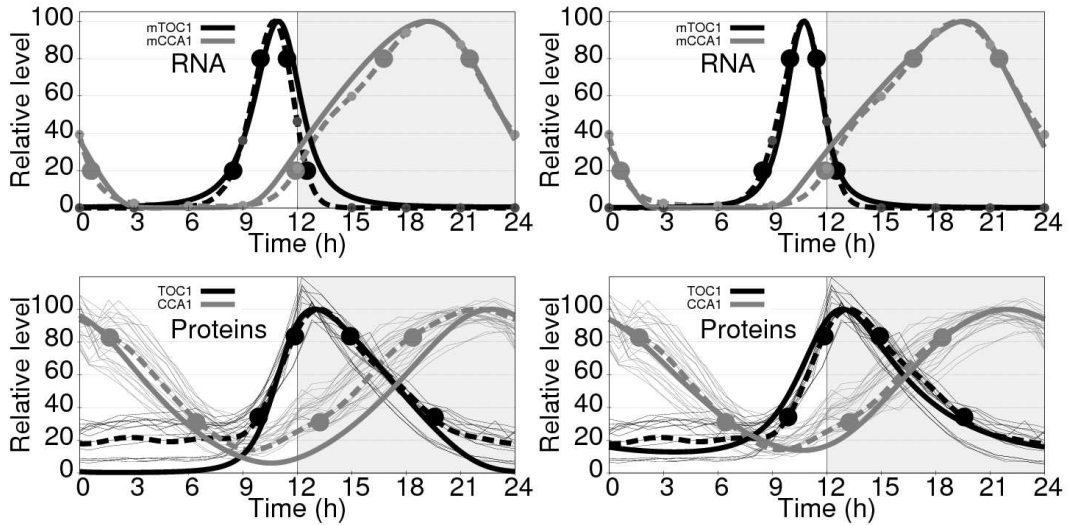


Fig. 2. A gauche : ajustement simultané des quatre courbes correspondant aux données ARN et protéine de la figure 1 après traitement ($n_c = 2$, $n_T = 2$, $1/\delta_{M_T} = 13.1$ mn, $1/\delta_{M_C} = 18.9$ mn, $1/\delta_{P_T} = 43.3$ mn, $1/\delta_{P_C} = 105.9$ mn, $K_{M_T} = 5.8nM$, $K_{M_C} = 0.61nM$, $K_{P_T} = 44nM$, $K_{P_C} = 800nM$, $\lambda_T = 25$ nM/h, $\lambda_C = 3.1$ nM/h, $\mu_T = 0.021$ nM/h, $\mu_C = 0.021$ nM/h). A droite : ajustement séparé des deux types de données (ARN d'un côté, protéines de l'autre), correspondant à deux expériences différentes. Les courbes cibles reconstruites à partir des données expérimentales sont indiquées en pointillés, les courbes issues des simulations numériques du modèle (1) en traits pleins. Dans le cas des données protéines, les lignes fines représentent les données brutes, normalisées par la fonction lentement variable obtenue lors du traitement.

heures à laquelle nous nous intéressons ici! Ce résultat est tout à fait contre-intuitif, et nous essaierons dans la prochaine section d'en tirer quelques pistes pour l'étude de l'horloge circadienne d'*Ostreococcus*.

Pour conclure, nous noterons que la qualité de l'ajustement ne dépend pas de la fréquence naturelle supposée. Nous avons effectué plusieurs ajustements pour des fréquences naturelles hors accrochage variant entre 23.8 et 25 heures et avons trouvé essentiellement les mêmes résultats, excepté bien sûr le fait qu'un couplage minimal est nécessaire à la synchronisation dès que période naturelle et période de forçage sont légèrement différentes. Les taux de couplages obtenus sont dans tous les cas très modestes, confirmant le résultat obtenu pour une fréquence naturelle supposée de 24 heures.

6 Discussion

Il nous faut d'abord insister sur le fait qu'il n'y a absolument rien d'évident au fait que le meilleur ajustement soit obtenu pour un couplage nul lorsque la période naturelle est supposée de 24 heures. Tout ce que permet l'égalité des périodes naturelle et de forçage, c'est qu'un verrouillage de fréquence puisse être obtenu dès un couplage infinitésimal. Mais il n'en reste pas moins que même dans ces conditions, le meilleur ajustement, c'est-à-dire non seulement le verrouillage en fréquence mais surtout le recouvrement optimal des signaux expérimentaux et numériques, doit se produire en général pour un couplage différent de zéro. Le fait que l'ajustement optimal soit obtenu pour un couplage nul révèle donc une propriété caractéristique de l'horloge circadienne d'*Ostreococcus tauri*.

Il faut ensuite souligner qu'il est impossible que l'horloge ne soit pas couplée au cycle jour/nuit, car on observe expérimentalement que la relation de phase entre l'horloge et le cycle de forçage met au plus deux à trois jours à s'établir lorsqu'on part d'une condition initiale quelconque. Il nous faut donc comprendre comment concilier le couplage nul de l'ajustement optimal et un couplage effectif bien réel.

Une première hypothèse de travail part de la remarque que tous les couplages qui ont été invalidés par la procédure d'ajustement agissent sur une échelle de temps de 24 heures. Ils sont en effet appliqués en imposant une certaine valeur d'un paramètre pendant les 12 heures du jour, puis une autre pendant

les 12 heures de nuit. Que l'absence de couplage soit préférable à un couplage de ce genre ne signifie pas qu'il n'y a aucun couplage, mais peut-être tout simplement que le couplage qui permet à l'horloge de se synchroniser n'est effectif que pendant de courts intervalles de temps. Un couplage qui n'agirait que pendant une petite fraction du cycle jour/nuit pourrait être quasiment invisible à l'échelle de 24 heures tout en contribuant efficacement à verrouiller la phase de l'horloge. Il est tentant de se demander si ces intervalles de couplage actif ne seraient pas localisés autour des transitions jour/nuit et nuit/jour, car c'est précisément à ces moments de la journée qu'un léger décalage de phase serait le plus facilement mesurable. Imaginons pour simplifier que les photorécepteurs ne soient allumés que pendant une heure de part et d'autre de la tombée de la nuit. Un décalage de 10 minutes fait passer le rapport jour/nuit de la période de mesure de 60 :60 à 70 :50 ou 50 :70, soit un changement facilement mesurable. Or, que représentent 10 minutes à l'échelle d'une journée? Dans le même temps, l'effet des corrections d'un tel couplage serait très faible et difficilement perceptible sur les signaux expérimentaux.

Mais cette idée d'un couplage localisé dans le temps n'est qu'un cas particulier d'une hypothèse plus générale. Les données expérimentales que nous avons utilisées pour ajuster le modèle théorique ont été enregistrées en régime permanent, alors que l'horloge est verrouillée en phase au cycle jour/nuit de manière stable. Peut-on imaginer que le couplage disparaît, ou tout au moins devient infime, dès que la relation de phase recherchée est atteinte? Cette hypothèse permettrait de concilier la relaxation rapide vers une relation de phase fixe que l'on observe expérimentalement avec l'ajustement à un oscillateur libre.

Un tel mécanisme dynamique présenterait un grand avantage. Dans la nature, l'éclairement reçu d'un jour sur l'autre peut varier de façon notable. Le signal que l'horloge reçoit en entrée et auquel elle doit se synchroniser est extrêmement régulier en phase, mais très bruyant en amplitude. A partir du moment où l'on a un couplage qui dépend de l'amplitude de ce signal (via l'intensité lumineuse), il n'est pas interdit de penser que la meilleure stratégie pour s'isoler de ce bruit soit précisément de débrancher le couplage lorsque la relation de phase recherchée est atteinte.

7 Conclusion

L'ajustement des données disponibles pour l'horloge circadienne de l'algue *Ostreococcus tauri* semble indiquer qu'un oscillateur génétique extrêmement simple, basé sur la régulation réciproque de deux gènes, est à l'œuvre au cœur de cette horloge. Il nous reste à expliciter les circuits qui permettent à cet oscillateur à se synchroniser au cycle jour/nuit et à comprendre pourquoi aucun couplage n'est apparent en régime synchronisé. S'agit-il d'une stratégie pour s'affranchir du bruit d'amplitude du cycle jour/nuit? Quoi qu'il en soit, *Ostreococcus* apparaît comme un organisme modèle particulièrement prometteur pour la biologie circadienne et la biologie systémique en général.

Références

1. J.C. DUNLAP, Molecular bases for circadian clocks, *Cell*, **96**, 271–290 (1999).
2. M.W. YOUNG & S. KAY, Time zones : a comparative genetics of circadian clocks, *Nature Genetics*, **2**, 702–715 (2001).
3. P. FRANÇOIS, A model for the neurospora circadian clock, *Biophysical Journal*, **88**, 2369–2383 (2005).
4. J. CW LOCKE *et al*, Extension of a genetic network model by iterative experimentation and mathematical analysis, *Molecular Systems Biology*, **88**, 4100018 (2005).
5. J. CW LOCKE *et al*, Experimental validation of a predicted feedback loop in the multi-oscillator clock of arabidopsis thaliana, *Molecular Systems Biology*, **88**, 59 (2006).
6. D. ALABADÍ, T. OYAMA, M. J. YANOVSKY, F. G. HARMON, P. MÁ S & S. A. KAY, Reciprocal Regulation Between *TOC1* and *LHY/CCA1* Within the Arabidopsis Circadian Clock, *Science*, **293**, 880–883 (2001).
7. F. CORELLOU, CH. SCHWARTZ, J.-P. MOTTA, F. SANCHEZ & F.-Y. BOUGET, A reduced version of the plant circadian clock relies on fine tuned diurnal and circadian regulations of the *TOC1/CCA1* feedback loop in the piccukaryote, *Ostreococcus*, soumis.