



HAL
open science

Réseaux génétiques : conception, modélisation et dynamique

Paul François

► **To cite this version:**

Paul François. Réseaux génétiques : conception, modélisation et dynamique. Physique [physics]. Université Paris-Diderot - Paris VII, 2005. Français. NNT : . tel-00010681

HAL Id: tel-00010681

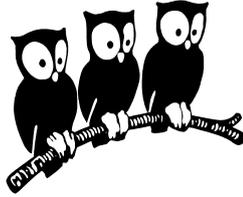
<https://theses.hal.science/tel-00010681>

Submitted on 18 Oct 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE NORMALE SUPÉRIEURE
Département de Physique
Laboratoire de Physique Statistique



THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS VII

présentée par

Paul FRANÇOIS

pour obtenir le titre de Docteur de l'Université Paris VII

Spécialité : Physique théorique

**Réseaux génétiques :
conception, modélisation et dynamique**

Thèse soutenue le 16 septembre 2005 devant le jury composé de :

Jean-Marc DI MEGLIO	Président
Vincent HAKIM	Directeur
François KÉPES	Examineur
Marc LEFRANC	Examineur
Denis THIEFFRY	Rapporteur
Massimo VERGASSOLA	Rapporteur

Remerciements

Ce manuscrit de thèse marque l'aboutissement de trois années de travail au Laboratoire de Physique Statistique du Département de Physique de l'École Normale Supérieure. Merci à son directeur Jacques Meunier pour m'y avoir accueilli.

Je remercie très chaleureusement Vincent Hakim d'avoir accepté de diriger ma thèse. J'ai bénéficié d'un encadrement de grande qualité et j'ai énormément appris à ses côtés. J'ai grandement apprécié ses qualités humaines durant nos nombreux échanges (y compris raquette en main!). J'ai été très impressionné par sa connaissance approfondie de nombreux domaines de la physique mais aussi de la biologie. J'ai beaucoup appris de sa méthode scientifique à la fois intuitive et rigoureuse, et espère avoir été un disciple à la hauteur du maître.

Un grand merci également à Kirone Mallick. Le stage d'option de l'X réalisé sous sa direction fut ma première expérience de recherche. J'en garde un excellent souvenir et c'est à cette époque que j'ai véritablement choisi la voie de la science. J'ai par la suite bénéficié de ses conseils avisés et de son soutien dans les inévitables moments de doute. J'ai beaucoup apprécié sa gentillesse et sa disponibilité, et je lui dois beaucoup.

Le Département de Physique en général, et le Laboratoire de Physique Statistique en particulier ne pourraient mener à bien leurs missions de recherche sans le dévouement de ses personnels. Je remercie donc nos secrétaires, Nora Sadaoui, Annie Ribaudeau et Carole Philippe-Barache ainsi que Rémy Portier et Daniel Le Moal pour le service informatique. Je tiens également à remercier Nicole Ribet, qui m'a suivi de loin tout au long de ces quatre années passées au Département de Physique et qui manquera beaucoup au DEA de Physique Théorique, ainsi qu'à ses anciens élèves.

Je remercie les membres du jury pour l'intérêt porté à mon travail. Merci aux examinateurs François Képes et Marc Lefranc, ainsi qu'aux rapporteurs Massimo Vergassola et Denis Thieffry. Merci enfin à Jean-Marc Di Meglio pour avoir accepté de présider mon jury.

Le LPS est un lieu très stimulant intellectuellement, où j'ai pu bénéficier d'un

environnement très favorable et propice aux discussions scientifiques (ou non) avec Arezki Boudaoud, Mokhtar Adda-Bedia, Xavier Leyronas, Frédéric Caupin, Hervé Henry, Agnès Huynh, Caroline Nore, François Pétrélis, Sébastien Aumaître, Jean Vannimendus, Bernard Derrida, Martine Ben Amar, Didier Chatenay, Rava Da Silveira. Merci également à toutes les équipes enseignantes avec lesquelles j'ai collaboré à l'Université Paris VII - Denis Diderot.

Je quitte également la fameuse Dc21, salle des bosons. Un grand merci tout d'abord à nos glorieux prédécesseurs qui nous ont transmis le flambeau à notre arrivée : Christophe Mora et Chi-Tuong Pham, dont les retours périodiques sont toujours fêtés et appréciés à leur juste valeur. Merci également à mes compagnons de route (et de foot!) tout au long de ma thèse Éric Sultan, Camille Énaud, ainsi qu'au boson délocalisé Jean-Marc Allain, merci à Cyril Cichowlas, à Giulia Tonini et merci au boson assimilé Yacine Dolivet. Enfin, merci et bonne chance à nos successeurs Hervé Rouault, Grégoire Deback, Damien Simon et Olaf Duemmer.

Je remercie toutes les personnes qui m'ont entouré pendant ma thèse, amis et famille. Une thèse représente l'aboutissement de longues années d'études. Mes parents m'ont toujours soutenu sur tous les plans durant toutes ces années et je leur en suis redevable et reconnaissant. Merci enfin à mon épouse Audrey, à la fois pour les discussions scientifiques, pour son soutien permanent et pour tout le reste. À mes parents et à Audrey, je dédie cette thèse.

Table des matières

Remerciements	5
Avant-propos	11
I Introduction aux réseaux génétiques	13
A Réseaux génétiques	13
B Structure des réseaux génétiques	14
B.1 Propriétés statistiques des réseaux génétiques	14
B.2 Structure modulaire	16
B.3 Des modules aux réseaux	17
C Modules artificiels	18
C.1 “Toggle switch”	18
C.2 Le répressilateur	23
C.3 Modules génétiques fabriqués aléatoirement	24
C.4 Optimisation d’un circuit génétique	26
D Conclusions	28
II Procédure d’évolution in silico	29
A Algorithme d’évolution	31
A.1 Modélisation des réseaux génétiques	31
A.2 Une étape de l’algorithme	32
B Commutateurs génétiques	36
B.1 Fonction de score	36
B.2 Exemples typiques des commutateurs obtenus	37
B.3 Autres commutateurs	37
C Oscillateurs	42
C.1 Fonction de score	42
C.2 Exemples d’oscillateurs obtenus	42
D D’autres applications possibles de l’algorithme	45

D.1	Rôle de l'ARN dans la dynamique des oscillateurs génétiques	45
D.2	Compensation de température	47
D.3	Vers la construction de motifs spatiaux	47
E	Conclusions préliminaires et motivations	49
E.1	Comparaison des réseaux obtenus avec des réseaux réels	49
E.2	Conclusions	51
F	Article	53
III Un module génétique prototype		61
A	Caractéristiques du modèle	61
A.1	Équations de base	61
A.2	Ordre de grandeur des constantes cinétiques et identification des petits paramètres	64
B	Aperçu de la dynamique dans les différents régimes de paramètres	65
C	Régime bistable	67
C.1	État stable avec A haut	70
C.2	État stable avec B haut	71
C.3	Bifurcation entre les régimes monostables et bistables	72
D	Régime oscillant	73
D.1	Analyse de stabilité linéaire	73
D.2	Étude du cycle limite	77
D.3	Phase I : concentration de A haute, concentration de B basse	77
D.4	Phase II : concentration de A basse, concentration de B haute	80
D.5	Condition de continuité pour la description du cycle limite	81
E	Raccordement des régimes à l'ordre $\sqrt{\delta}$ et phase intermédiaire	83
E.1	Transition de la phase I à la phase II	83
E.2	Transition de la phase II à la phase I	84
E.3	Corrections à l'ordre $\sqrt{\delta}$	85
F	Comparaison avec un modèle avec délai	86
G	Conclusions	87
H	Article	92
IV À propos des horloges circadiennes		107
A	Horloges circadiennes	107
A.1	Quelques définitions	108
A.2	Réseaux génétiques des horloges circadiennes	108
A.3	Un exemple : l'horloge de <i>Neurospora crassa</i>	109

B	Exemples de modèles d’horloges circadiennes	111
C	Résumé de l’article : Un modèle d’horloge circadienne de Neurospora	114
C.1	Modèle à une boucle	115
C.2	Mécanismes de l’oscillation	117
C.3	Contrôle par l’ARN	120
C.4	Modèle à deux boucles	121
C.5	Conclusions	125
D	Article	125
Perspectives		143
Conclusions		149
Appendices		153
A Compléments sur l’algorithme d’évolution		155
A.I	Exemples de commutateurs.	155
A.II	Simulations avec des dynamiques stochastiques	155
A.III	Exemples d’oscillateurs.	156
B Méthodes Informatiques		167
B.I	Classes utilisées dans l’algorithme	167
B.II	Exemple de code	169
C Modèle avec délai		177
C.I	Diagramme de phase	177
C.II	Mécanisme des oscillations	179
Bibliographie		181

Avant-propos

Cette thèse présente différents résultats théoriques concernant la modélisation des réseaux génétiques. Le chapitre I présente le contexte théorique et expérimental dans lequel ont été réalisés ces travaux, en mettant successivement l'accent sur les analyses statistiques des réseaux génétiques ainsi que sur les différents travaux de biologie synthétique. Le chapitre II présente la procédure d'évolution *in silico* visant à concevoir des réseaux génétiques remplissant une fonction spécifique donnée, publiée dans la référence [26], puis différents résultats inédits obtenus à partir de cette procédure. Le chapitre III contient une étude mathématique de certains des réseaux biologiquement pertinents obtenus à l'aide de la procédure d'évolution. Enfin, le chapitre IV présente une application des modèles suggérés au cas spécifique des horloges circadiennes, publiée dans la référence [27]. Nous concluons en élargissant cette approche à la formation de motifs spatiaux, et en proposant un modèle heuristique de formation des somites.

En annexe, on trouvera les informations complémentaires sur la procédure d'évolution *in silico*, différentes indications sur les implémentations informatiques utilisées ainsi qu'une étude succincte d'un modèle d'oscillateur génétique avec délai.

Les différents chapitres, même s'ils abordent des thématiques proches, peuvent être lus de manière largement indépendante.

Chapitre I

Introduction aux réseaux génétiques

A Réseaux génétiques

LA DÉCOUVERTE de la structure en double hélice de l'ADN au début des années 50 marque le début de la génétique moderne au sens large. Le déchiffrement du code génétique nous a ensuite révélé qu'en fait les séquences génétiques sont associées aux protéines présentes dans l'organisme. Rappelons le "dogme central" de la génétique dans sa version très simplifiée : l'ADN, support de l'information génétique, code les séquences d'acides aminés correspondant à des protéines données, véritables outils de base de la machinerie cellulaire. Cependant, l'ADN ne "produit" pas directement de protéines : la séquence d'ADN correspondant à un gène donné est d'abord copiée sous forme d'ARN messager (transcription), puis la séquence codée par cet ARN est lue par les ribosomes chargés de synthétiser la protéine (traduction). Les protéines vont ensuite se replier, adopter une structure tridimensionnelle bien définie et acquérir toutes leurs fonctions. Dans le "dogme central", un gène est donc associé à une protéine, via la production d'un ARN messager.

Depuis ces premières découvertes, des progrès considérables ont été réalisés amenant une connaissance approfondie des séquences génétiques. Il apparaît finalement que ces séquences génétiques héréditaires contiennent potentiellement une très grande partie de l'information biochimique nécessaire au bon fonctionnement d'un organisme. A très long terme, la mise au point d'outils pour comprendre l'organisation et la structure de ces séquences devrait donc peut-être permettre de comprendre la biologie à l'échelle d'un organisme.

En particulier, les différentes protéines ne sont pas produites de façon indépendante : des protéines peuvent interagir entre elles, activer ou réprimer la transcription d'autres gènes et donc la production d'autres protéines, interagir avec des ARN messagers ..., de telle sorte qu'il est commode de décrire l'ensemble de ces

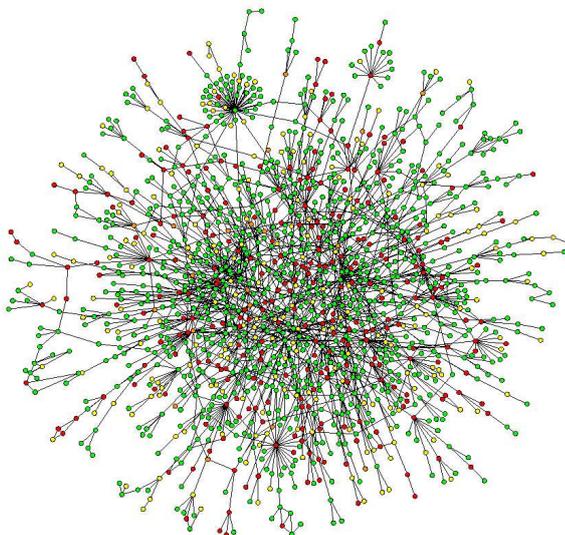


FIG. I.1: Un exemple de réseau génétique chez la levure. Chaque noeud correspond à une protéine, chaque lien à une interaction protéine-protéine. Illustration de Hawoong Jeong, université de Notre Dame.

interactions comme un “réseau génétique”. Une représentation synthétique d’un tel réseau serait un graphe, dans lequel chaque noeud correspond en fait à un triplet {gène,ARN,protéine} et chaque lien à une interaction. L’émergence de nouvelles méthodes expérimentales a permis d’explorer en profondeur ces réseaux : par exemple, l’utilisation de la technique du double-hybride chez la levure [42, 84] a permis d’obtenir de façon systématique de nombreuses données sur les interactions protéine-protéine. A l’aide de ces données, il est donc possible de représenter les réseaux d’interactions génétiques et d’analyser les circuits génétiques correspondant aux différents processus cellulaires (voir par exemple Figure I.1).

B Structure des réseaux génétiques

B.1 Propriétés statistiques des réseaux génétiques

La modélisation et l’étude de réseaux complexes repose en général sur la comparaison de ces réseaux avec des réseaux aléatoires bien connus. En particulier, l’étude et la comparaison des distributions de connectivité (i.e. la probabilité $P(k)$ pour un noeud d’avoir k voisins) est révélatrice de la structure, voire même des processus de formation des réseaux. Le premier modèle de référence de réseau complexe a été introduit par Erdős et Rényi [5, 21]. Dans ce premier modèle simple, on considère un réseau de taille N donné et on fixe le nombre n de liens dans ce réseau, chaque

lien étant choisi de façon équiprobable parmi les $\frac{N(N-1)}{2}$ liens possibles. Un modèle équivalent, le modèle “binomial”¹, revient à considérer N noeuds, chaque paire de noeuds étant reliée avec une probabilité p . Dans la limite N grand, on peut en particulier immédiatement calculer la distribution de connectivité du réseau, et cette distribution suit alors une loi de Poisson de paramètre $\langle k \rangle \sim pN$, si bien que cette distribution est piquée autour de $\langle k \rangle$. Ceci signifie qu’un tel réseau est relativement homogène : tous les noeuds du réseau ont *grosso modo* le même nombre de voisins et il existe très peu de noeuds fortement ou faiblement connectés.

À partir des données expérimentales, il est possible d’évaluer les distributions de connectivité des réseaux génétiques. Les premières études sur différents organismes ont montré que cette distribution semble en fait suivre une loi de puissance $P(k) = k^{-\gamma}$ à la fois par exemple pour les réseaux métaboliques [44] et pour les réseaux d’interactions protéine-protéine [55] si bien que les réseaux génétiques sont très différents des réseaux aléatoires type Erdős-Rényi. De tels réseaux sont qualifiés d’invariants d’échelle (“scale free”) et sont en fait très hétérogènes : leur topologie est dominée par quelques noeuds très connectés reliés à d’autres noeuds très peu connectés. Une façon de fabriquer numériquement de tels réseaux est d’ajouter un à un des noeuds en les connectant de façon préférentielle aux noeuds les plus connectés [4]. Ces lois de puissance suggèrent donc un mécanisme de développement des réseaux génétiques au cours de l’évolution.

Avec l’accumulation de données expérimentales, en particulier pour la levure, ce premier résultat a été affiné et Guelzim et al. ont montré par la suite que la distribution de connectivité est en fait plus complexe [36]. En tenant compte des orientations des liens, il semble en effet que les distributions de degré entrant et sortant suivent des lois différentes. La distribution de connectivité sortante (i.e. le nombre de gènes régulés par protéine régulatrice) semble suivre une loi de puissance ; en revanche, la distribution de connectivité entrante (i.e. le nombre de protéines régulatrices par gène régulé) semble plutôt suivre une loi exponentielle $P(k) \simeq e^{-k}$, ce qui signifie concrètement que la plupart des gènes sont individuellement régulés par peu de protéines. En examinant ensuite la corrélation entre les connectivités entrantes et sortantes, il apparaît que les gènes ayant peu de régulateurs tendent à avoir aussi peu de cibles. Comme ces gènes sont les plus nombreux, cela signifie que l’inactivation de tels gènes aura peu d’influence sur les autres gènes, tandis qu’au contraire, l’inactivation de gènes fortement connectés perturberait grandement le

¹Appelé ainsi car la connectivité d’un noeud suit alors une loi binômiale : si on appelle k_i le degré du noeud i (c’est-à-dire le nombre de noeuds reliés au noeud i), on voit immédiatement qu’on a $P(k_i = k) = C_{N-1}^k p^k (1-p)^{N-1-k}$

réseau d'interaction biochimique. Un tel réseau est donc robuste face aux dommages (suppression de noeuds choisis aléatoirement) mais sensible aux attaques ciblées [1].

B.2 Structure modulaire

Les réseaux génétiques sont donc très différents des réseaux aléatoires type Erdős-Rényi en ce qui concerne leur distribution de connectivité. Maintenant, connaissant la distribution de connectivité des réseaux génétiques, il est possible d'affiner l'étude des distributions de connectivité pour avoir des informations sur la topologie des réseaux génétiques. Ainsi, Uri Alon et ses collaborateurs ont développé une approche générale visant à caractériser un réseau complexe par ses sous-motifs typiques [60]. Ces modules caractéristiques sont par définition des sous-graphes présents dans le réseau de manière récurrente, et plus nombreux que dans des réseaux aléatoirement construits ayant la même distribution de connectivité. Cette stratégie a été appliquée au réseau transcriptionnel d'*Escherichia coli* et a permis de repérer trois réseaux significativement sur-représentés, représentés Figure I.2 [74].

Un des sous-réseaux particulièrement intéressant, près de huit fois plus représenté dans le réseau transcriptionnel d'*E. coli* que dans les réseaux aléatoires est la "Feedforward loop"(FFL). Ce module est composé de trois protéines X,Y,Z, X régulant Y, X et Y régulant Z par coordination. Selon l'effet des régulations (positives ou négatives), il existe 8 types de FFL. Le module le plus fréquent chez *E. coli* et la levure est le modèle avec trois activations [54]. Une simple modélisation permet alors de comprendre le rôle potentiel de la FFL en tant que "détecteur de persistance" . Si nous considérons la concentration de X comme un signal d'entrée et celle de Z comme un signal de sortie, pour induire la production de protéine Z, il faut maintenir une concentration de X élevée pendant une longue durée. Si la concentration de X n'est pas maintenue assez longtemps, X n'a pas le temps d'activer la production de Y et Z ne peut donc être produit².

L'accumulation des données sur les réseaux génétiques a également permis la recherche de motifs composites, i.e. des motifs caractéristiques incluant à la fois des régulations transcriptionnelles et des interactions post-transcriptionnelles. Là encore, des modules typiques ont été repérés [89,90] : ainsi, un des modules caractéristiques est la boucle de rétroaction mixte ("Mixed Feedback loop", voir aussi chapitre III), composée de deux protéines A et B, A régulant la transcription du gène *b*, A et B pouvant ensuite interagir. Des modules caractéristiques à 3 et 4 protéines

²Un équivalent de ce circuit est par ailleurs utilisé en électronique, en particulier pour les interrupteurs des téléphones cellulaires, où le signal d'entrée serait la pression à exercer.

ont également été détectés.

L'ensemble de ces travaux semblent donc montrer que la structure des réseaux génétiques, loin d'être aléatoire, est organisée en modules [38, 87]. Cette structure modulaire du génome évoque immédiatement une image mécaniste de l'organisation des réseaux génétiques. Un module génétique, à l'image d'un composant électronique pour un ordinateur, serait l'unité de base du vivant. Le traitement de nombreux signaux extérieurs se ferait au niveau génétique à l'aide de ces modules, conformément à l'intuition de Jacob et Monod il y a plus de 40 ans [62]. Dans cette image d'ingénieur du vivant, la modularité et la récurrence des motifs génétiques trouveraient une explication naturelle :

- la modularité aiderait à fractionner les tâches d'analyse des signaux/stimuli. Un module génétique serait alors composé d'une part de quelques gènes "entrées" et "sorties" interagissant avec les autres modules, et d'autre part de gènes internes au module interagissant entre eux, mais interagissant peu avec les gènes extérieurs aux modules afin de ne pas "parasiter" la fonction principale d'analyse du signal.
- la récurrence de ces motifs s'expliquerait par une pression évolutive ayant sélectionné ces motifs.

L'hypothèse de base est que la dynamique propre de ces modules sera conservée lors de l'assemblage de ces modules, et qu'on pourra alors en déduire des informations sur le fonctionnement global des réseaux génétiques.

B.3 Des modules aux réseaux

L'étape suivante dans l'étude des réseaux génétiques consiste donc à voir si les modules génétiques s'assemblent bien comme des petites briques biologiques indépendantes pour former un réseau fonctionnel plus grand. Ainsi Dobrin *et al.* [15] ont analysé la manière dont les modules transcriptionnels détectés chez *E.coli* se combinaient. L'analyse a alors révélé qu'en fait les différents modules étaient loin d'être indépendants et avaient au contraire des noeuds en commun. Par exemple, les 42 FFL détectées se regroupent en six agrégats de motifs. Il apparaît que les motifs détectés ont de nombreux noeuds en commun et forment un "supercluster" qui est en fait la structure de base de l'ensemble du réseau. Les motifs sont donc très inter connectés; les avantages de ces modules et leurs fonctions potentielles sont-ils néanmoins conservés ?

Afin de répondre à cette question, Mazurie *et al.* [58] ont étudié les différents modules trouvés chez la levure dans une perspective évolutionniste. En comparant

les données de *Saccharomyces cerevisiae* avec celles de quatre autres espèces de levure, ils ont tenté de détecter une éventuelle pression évolutive visant à favoriser et à préserver les motifs détectés. Une analyse phylogénétique révèle dans un premier temps qu’il ne semble pas y avoir de différences majeures entre les coévolutions des gènes en interaction dans les motifs et ceux en interaction simple (i.e. pas nécessairement dans un motif). Ce résultat semble donc infirmer l’hypothèse d’une évolution des motifs indépendamment du reste du génome. Puis ils ont étudié les agrégats de motifs et tenté de voir si les motifs jouaient un rôle fonctionnel déterminant en tant que tel dans les différents réseaux considérés. Sur les cinq exemples étudiés, il semble que les motifs mis en évidence ne jouent pas de rôles fonctionnels propres.

Ces différents travaux semblent donc nuancer l’idée simplifiée de motifs structuraux bien caractérisés jouant le rôle de briques de construction des réseaux génétiques. Cependant, les contre-exemples exhibés par Mazurie et al. sont éclairants dans la mesure où, s’il semble impossible d’isoler une fonction particulière générale pour les motifs, les agrégats de motifs ont tous une fonction très précise dans le réseaux génétique de la levure. Une approche alternative serait donc non pas de détecter des motifs donnés à partir des bases de données, mais plutôt de diviser les réseaux génétiques en ensemble de gènes réalisant en coordination une tâche donnée. Une même fonction pourrait par exemple être réalisée par plusieurs motifs plus ou moins compliqués. Il apparaît dans ce cas pertinent d’étudier les réseaux génétiques en se focalisant sur leur fonction, plutôt que sur leur structure. Une voie d’étude intéressante à ce titre est la synthèse et l’étude de réseaux artificiels [39], réalisant une fonction génétique donnée, voie que nous détaillons dans la suite.

C Modules artificiels

C.1 “Toggle switch”

Une des fonctions les plus importantes pour une cellule est la capacité à répondre à un stimulus (environnemental ou intracellulaire) par un signal de type “tout ou rien”. De tels processus sont très répandus dans toutes les étapes de différenciation cellulaire en particulier, et sont en général contrôlés au niveau génétique. Lorsque le stimulus initial disparaît, il peut être intéressant pour la cellule de maintenir le signal de sortie, conservant une mémoire de l’histoire passée de la cellule : le réseau génétique associé peut ainsi être bistable (voire même multistable) au sens dynamique. De tels modules présenteront de l’hystérésis et pourront être commutés d’un état “allumé” à un état “éteint” par un court stimulus.

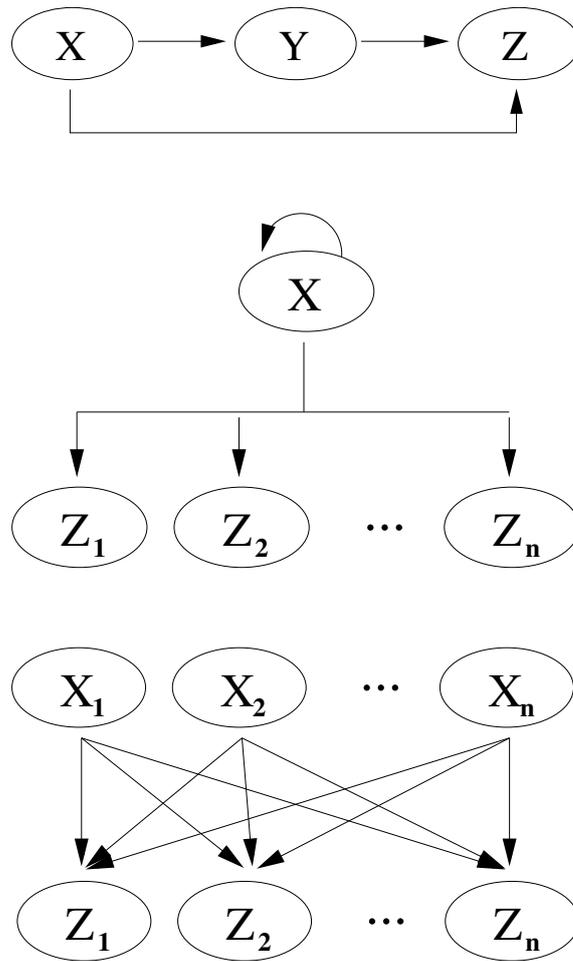


FIG. I.2: Les trois modules caractéristiques du réseau génétique d'*E. coli*. Haut : Détecteur de persistance (“feedforward loop”, ou FFL), Milieu : module à une seule entrée (“single input module”, ou SIM), Bas : régulons interrégulés (“dense overlapping regulons”, ou DOR)

Parmi les nombreux travaux mettant en évidence ces phénomènes de multistabilité dans les réseaux génétiques, citons deux exemples : la maturation de l’oeuf de *Xenopus* [23, 88], ou encore l’exemple classique de l’opéron lactose, dont les différentes propriétés d’hystérésis ont été récemment particulièrement bien mises en évidence expérimentalement [64]. Tous ces systèmes présentent des boucles de rétroaction (directes ou indirectes) positives, conformément à la conjecture de R.Thomas³ récemment prouvée mathématiquement [77]. Compte-tenu de l’importance de tels systèmes, la conception de systèmes multistables artificiels, pouvant être facilement modifiés et contrôlés, représente un enjeu considérable, à la fois sur le plan théorique, expérimental, ainsi que dans l’optique d’applications biotechnologiques [45].

Dans toute la suite, nous appellerons commutateur génétique (ou “switch”) tout module composé d’un couple de gènes/protéines vérifiant les propriétés suivantes :

- le module est bistable,
- les deux états stables du système sont définis comme suit : soient A et B les deux protéines composant ce module. Le premier état stable sera caractérisé par une haute concentration de A et une faible concentration de B. Le second état stable sera caractérisé par une haute concentration de B et une basse concentration de A.

Compte-tenu de la nature des états stables, l’idée naturelle est de concevoir un réseau génétique où les protéines A et B se répriment mutuellement⁴. C’est l’idée suivie par Gardner et al [31], qui ont développé une approche couplant théorie et expérience pour synthétiser plusieurs “toggle switch” chez *E.coli*.

C.1.a Théorie

Une étude complète du comportement du “toggle switch” peut être trouvée dans l’article de Cherry et Adler [9]. Le comportement du “toggle switch” est modélisé qualitativement par le système de deux équations différentielles suivant :

$$\begin{aligned}\frac{dA}{dt} &= \frac{\rho_A}{1 + (B/B_0)^\nu} - \delta_A A \\ \frac{dB}{dt} &= \frac{\rho_B}{1 + (A/A_0)^\mu} - \delta_B B\end{aligned}$$

A et B représentent ici les concentrations des protéines . Les deux répressions sont modélisées par des fonctions de Hill (représentant de façon effective le mé-

³René Thomas a conjecturé au début des années 80 que tout réseau multistable possède au moins un circuit de rétroaction positive.

⁴Notons que ce circuit génétique correspond globalement à une boucle de rétroaction positive par ajout en série de deux répressions.

canisme de répression de la transcription). ρ_A et ρ_B sont les taux de production effectifs des protéines A et B en l'absence de leur répresseur respectif. δ_A et δ_B sont les constantes de dégradation des protéines. ν et μ sont les coefficients de Hill associés aux repressions. D'un point de vue biochimique, ces exposants mesurent la coopérativité de ces réactions (i.e. concrètement le nombre de protéines qui doivent interagir au niveau des promoteurs pour réprimer la transcription).

Dans un premier temps, on voit immédiatement que le système comporte plusieurs points fixes uniquement si au moins un des paramètres μ et ν est plus grand que 1 : dans ce cas, les nullclines associées à cette équation ont une forme sigmoïdale, et si les taux de production des protéines sont du même ordre, il y a trois points d'intersection, correspondant à deux points fixes stables et un point fixe instable. Il est donc nécessaire d'avoir de la coopérativité dans le système, ainsi que des taux de production similaires pour chacune des deux protéines pour avoir bistabilité. Dans un second temps, on peut calculer le diagramme de bifurcation associé à ce système de deux équations : la taille de la région bistable augmente alors avec la coopérativité.

C.1.b Réalisation expérimentale

La Figure I.3 présente le schéma de principe des “toggle switch” réalisés par Gardner et al. Plusieurs types de plasmides ont été réalisés chez *E. coli*. La première paire de promoteur/répresseur est composée du répresseur Lac (*lacI*), en conjonction avec un promoteur *P_{trc2}*. Deux types de plasmides ont été ensuite synthétisés selon le deuxième couple de promoteur/répresseur désiré : soit le promoteur *P_{Ls1con}* modifié conjugué avec une protéine λ CI, impliqué dans le commutateur naturel du bactériophage λ (plasmide appelé pTAK), soit le répresseur Tet en conjugaison avec son promoteur associé (plasmide appelé pIKE). En aval de ce second répresseur est incluse la séquence correspondant à la GFP (“Green Fluorescent Protein”, protéine fluorescente verte) qui joue le rôle de rapporteur.

Un avantage de ces répresseurs est qu'ils agissent très coopérativement, ce qui augmente l'efficacité du commutateur. Cependant, les taux de transcription naturels associés à ces différents gènes sont assez différents, et les séquences génétiques ont dû être artificiellement modifiées afin d'avoir des taux de transcription du même ordre.

L'autre avantage de ce choix de promoteurs est que les repressions sont elles-mêmes contrôlées par des inducteurs spécifiques : l'IPTG (isopropyl- β -D-thiogalactopyranoside) lève la répression du Lac répresseur, l'aTc (anhydro-tétracycline)

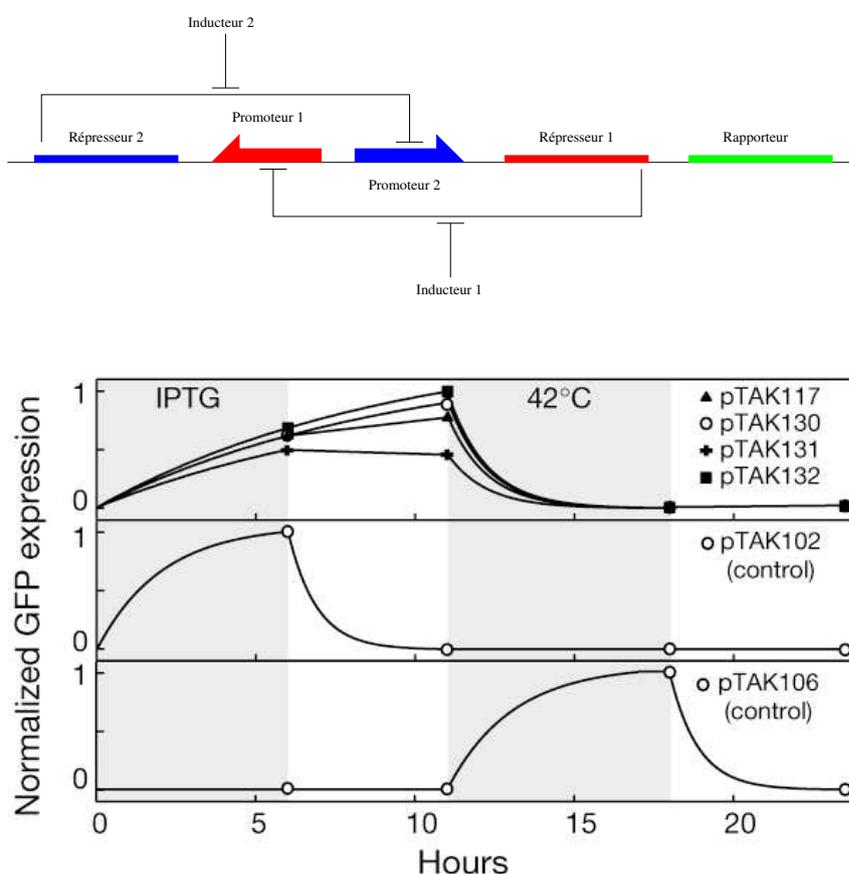


FIG. I.3: Schéma de principe et réalisation expérimentale du toggle switch.

la répression de Tet tandis que la répression exercée par λ CI est sensible à la température. Ainsi, on peut faire basculer (“toggle”) le module génétique d’un état à l’autre en produisant des pulses d’IPTG d’une part, de température ou d’aTC d’autre part, qui lèvent les différentes répressions.

Concrètement, si on prend l’exemple des plasmides pTAK, on vérifie la bistabilité en soumettant les bactéries à de l’IPTG pendant six heures. La répression par LacI est alors levée et λ CI peut s’accumuler, ainsi que la GFP. Ensuite, les cellules sont lavées et placées dans un milieu sans IPTG. On observe alors que la fluorescence se maintient, indiquant la persistance du premier état stable λ CI haut, lacI bas. Puis, les bactéries sont placées dans une étuve à 42°C pendant six autres heures, ce qui lève la répression par le répresseur λ CI. LacI peut alors s’accumuler et réprimer l’expression de l’autre gène, ce qui se traduit par une décroissance de la fluorescence. Les cellules sont ensuite remises à la température initiale (32°C) et la fluorescence reste basse, indiquant qu’on reste dans l’état stable LacI haut, répresseur λ CI bas.

C.2 Le répressilateur

Les oscillateurs génétiques constituent également d'importants exemples de modules, impliqués dans divers mécanismes de régulation cellulaires ou de synchronisation des cellules entre elles ou avec des stimuli extérieurs. Plus précisément, nous appellerons “oscillateur génétique” un système dans lequel les concentrations intracellulaires de certaines protéines oscillent au cours du temps. Citons à titre d'exemple les oscillateurs génétiques impliqués dans le cycle cellulaire [6, 81], les oscillateurs circadiens [13, 17, 92], ou encore l'oscillateur p53/Mdm2 [3, 46] (voir également les chapitres III et IV). Différents oscillateurs génétiques peuvent parfois être couplés entre eux, comme montré par Matsuo et al. dans le cas des oscillateurs circadiens et du cycle cellulaire [57]. En général, des boucles de rétroaction négatives sont à l'origine de ces oscillations [82]. L'idée de base pour concevoir un oscillateur génétique consiste donc à construire une boucle de rétroaction négative.

C.2.a Théorie

L'idée d'Elowitz et Leibler [18] est de construire une boucle de rétroaction globalement négative par mise en série de trois couples promoteur/répresseur, comme représenté Figure I.4. L'évolution des concentrations des ARN messagers (notés m) et de celles des protéines correspondantes (notées p) sont modélisées par les six équations adimensionnées symétriques suivantes :

$$\frac{dm_i}{dt} = \alpha_0 + \frac{\alpha}{(1 + p_j^n)} - m_i$$

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

avec $i = 1, 2, 3$, correspondant à $j = 3, 1, 2$ dans ces équations. α_0 représente le taux de production des protéines par cellules i en présence de la protéine j , $\alpha_0 + \alpha$ représente ce même taux en absence de cette protéine j , n est le coefficient de Hill correspondant à la répression. β représente le rapport entre le taux de dégradation des protéines et celui des ARN. Le temps est redimensionné pour que le taux de dégradation de l'ARN vale 1.

Dans ce modèle simplifié, il est relativement aisé de calculer la stabilité linéaire de ce système : comme dans le cas du toggle switch, un grand coefficient de Hill et un petit α_0 augmente le domaine dans lequel le système oscille.

C.2.b Réalisation expérimentale

Les trois couples promoteur/répresseur sont en fait des versions modifiées des trois couples déjà rencontrés (*lacI*, *tetR* et protéine λ CI) pour avoir des répressions fortes. Les séquences sont modifiées pour avoir un renouvellement rapide des protéines. Le plasmide est introduit chez *E. coli*, et la GFP est utilisée comme rapporteur. Comme on peut le voir sur la Figure I.4, on observe alors une oscillation de la fluorescence due à la production cyclique de GFP. Par ailleurs, les bactéries continuent de se reproduire si bien que la fluorescence moyenne de la culture augmente au cours du temps. Elowitz et Leibler ont par ailleurs vérifié qu'il n'y avait aucune corrélation entre la période de division et la période du cycle cellulaire, montrant ainsi la nature intrinsèque et indépendante de l'oscillation. Cependant, du fait de ces divisions en particulier et de l'absence de synchronisation, les oscillations sont très bruitées et on observe une grande variabilité de la phase et de l'amplitude des oscillations d'une bactérie à l'autre. Le répressilateur a par la suite été utilisé comme un mini-laboratoire cellulaire pour quantifier le bruit présent dans les réseaux génétiques [19].

C.3 Modules génétiques fabriqués aléatoirement

Les travaux présentés plus hauts ont mis en évidence la possibilité d'utiliser de véritables boîtes à outils génétiques (comprenant certains gènes et des promoteurs associés) afin de construire, dans une perspective d'ingénierie biotechnologique, des modules génétiques remplissant des fonctions données. Comme nous l'avons vu, le "toggle switch" comme le répressilateur ont été conçus à partir d'une idée théorique simple, et réalisés concrètement en jouant sur les différentes affinités, constantes chimiques, ... des protéines impliquées. Il est également frappant de constater qu'il est relativement facile de construire des modules fonctionnels à partir de gènes qui n'ont aucune interaction naturelle. Guet et al. [37] sont allés plus loin en fabriquant par combinatoire une véritable bibliothèque des fonctions génétiques pouvant être remplies par de tels modules. L'idée de base est de fabriquer de petits modules composés des trois gènes *lac*, *tet* et λcl et de cinq promoteurs régulés par les protéines associées : deux promoteurs sont réprimés par *LacI* avec des intensités différentes, un par *TetR*, et deux promoteurs régulés par λ CI, l'un positivement, l'autre négativement. On construit alors des plasmides comportant une seule copie de chaque gène, associé aléatoirement à l'un des cinq promoteurs mentionnés, créant ainsi une bibliothèque comprenant $5^3 = 125$ plasmides (voir schéma de principe Figure I.5).

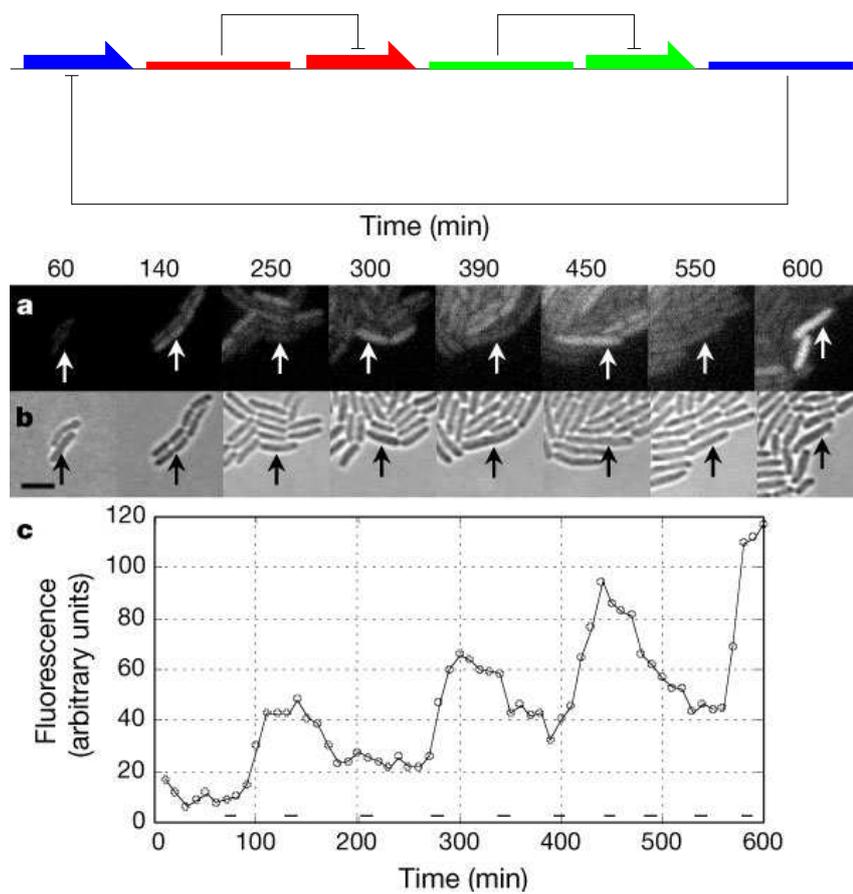


FIG. I.4: Schéma de principe et réalisation expérimentale du répressilateur. Haut : schéma du plasmide, Milieu : évolution de la population de bactéries au cours du temps, Bas : mesure de la fluorescence totale de la colonie.

Enfin, on ajoute le gène de la GFP contrôlé par P_λ (c'est-à-dire que la protéine λ CI réprime l'expression de la GFP).

Le plasmide est ensuite inséré dans une souche d' *E.coli*. L'idée est de fabriquer des portes logiques (par exemple ET, noté \wedge , OU noté \vee ...) prenant comme signaux d'entrées l'IPTG ou l'aTc, et comme signal de sortie le niveau de fluorescence des cellules. Les souches sont alors mises en culture dans quatre milieux nutritifs différents : pauvre en IPTG et en aTc, riche en IPTG et pauvre en aTc, riche en aTc et pauvre en IPTG et enfin riche en IPTG et en aTc. Les niveaux de fluorescence entre les différentes expériences sont assez distincts pour écrire une table de vérité booléenne permettant de décrire la fonction logique exercée. Les plasmides sont rétrospectivement séquencés pour connaître le module associé.

L'étude des différents modules obtenus montre que de nombreuses fonctions logiques peuvent être obtenues par cette méthode. De plus, on peut facilement changer la fonction logique d'un module en modifiant un seul lien dans le réseau génétique obtenu. Plusieurs topologies peuvent avoir la même fonction logique, mais on observe également que la même topologie peut se traduire par plusieurs fonctions différentes, si bien que certains modules présentent un fonctionnement différent de celui dicté par leur topologie (voir Figure I.5). Cela signifie que même pour un réseau transcriptionnel très simple, des mécanismes inconnus (ou ne pouvant pas être modélisés par une approche booléenne) jouent un rôle fondamental dans la régulation génétique.

C.4 Optimisation d'un circuit génétique

Une approche alternative pour concevoir des réseaux artificiels est d'utiliser la formidable capacité du vivant à évoluer et à s'adapter à une pression sélective. Plutôt que de concevoir des modules génétiques *ex nihilo*, Yokobayashi et al. [91] ont utilisé une stratégie d'évolution dirigée *in vivo* pour concevoir l'inverse d'une porte IMPLIQUE (\Rightarrow), appelée "inverseur" ("inverter"), voir Figure I.6⁵.

L'idée est de partir d'un circuit logique réalisant théoriquement la fonction voulue, mais qui, dans la pratique, fonctionnera mal pour diverses raisons biologiques inconnues, et de le modifier par évolution dirigée. Les deux signaux d'entrée choisis sont LacI et l'IPTG, le signal de sortie l'EYFP ("enhanced yellow fluorescent protein", protéine fluorescente jaune améliorée). Le promoteur P_{lac} réalise alors une porte IMPLIQUE naturellement : en effet, la transcription est réprimée uniquement en présence de LacI et en l'absence d'IPTG. En aval du promoteur est ensuite inséré

⁵la proposition logique $A \Rightarrow B$ est équivalente à la proposition $B \vee \neg A$. Ici, Yokobayashi et al. veulent construire une porte $\neg(A \Rightarrow B)$

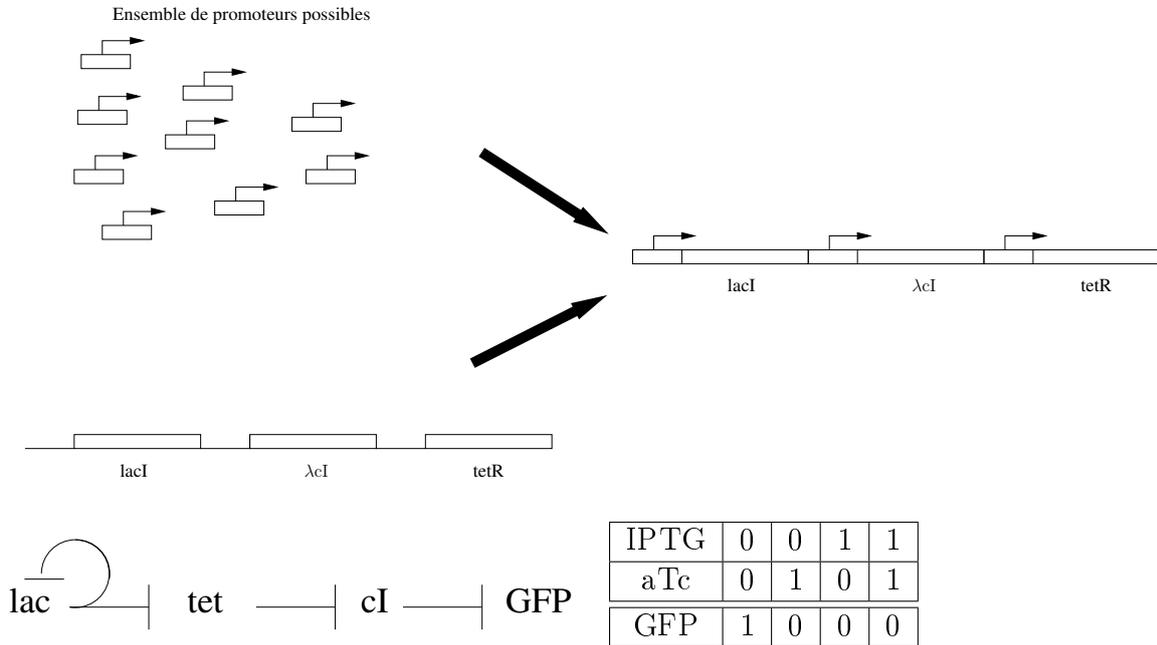


FIG. I.5: Conception de modules génétiques par combinatoire. Haut : schéma de principe. Bas : exemple de plasmide séquencé réalisant la fonction logique $\neg(\text{IPTG} \vee \text{aTc})$. On peut noter qu'a priori la fonction de ce module devrait plutôt être $\neg \text{aTc}$.

le gène λcI , qui sert de signal de sortie à cette première porte logique. La protéine λCI est donc produite dans tous les cas, sauf en présence de protéines LacI seules.

La transcription de l'EYFP est ensuite contrôlée un promoteur P_λ , qui joue le rôle de la négation (\neg). La transcription de l'EYFP n'a donc lieu qu'en l'absence de protéine λCI , c'est-à-dire uniquement en présence de LacI seule. En théorie, la juxtaposition de ces deux modules génétiques élémentaires, vues comme des portes logiques, devrait créer un inverseur. En réalité, cette porte logique ne fonctionne pas car le signal de sortie de la première porte logique (i.e. la concentration de λCI) est toujours interprété comme un signal "1" pour la seconde porte logique (i.e. dans tous les cas la concentration de CI est assez élevée pour réprimer l'expression du gène codant pour l'EYFP).

Pour résoudre cette difficulté, Yokobayashi et al. utilisent une procédure d'évolution dirigée pour d'ajuster les paramètres biochimiques afin d'ajuster l'interface entre les deux portes logiques. Pratiquement, les plasmides sont alors exposés à des mutagènes ciblant le gène λcI et son site de liaison des ribosome (RBS). Dès la première génération, de 2.5 à 5 % des souches sont alors capables de réaliser correctement la fonction logique "inverseur". Les plasmides sont alors séquencés pour examiner l'effet des mutations : la majeure partie de ces souches a en fait subi une

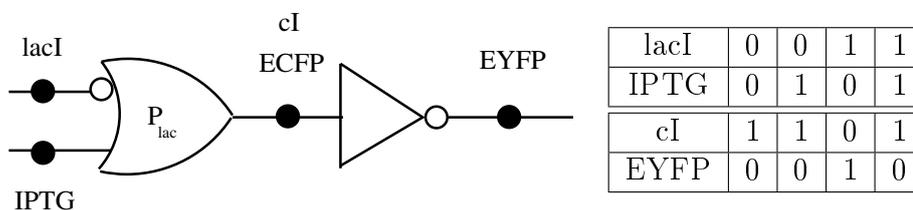


FIG. I.6: Porte logique optimisée avec la table de vérité associée.

sévère mutation de la séquence du gène λ cI, ayant pour effet une baisse très significative de la coopérativité au niveau du promoteur P_λ . Cette baisse de coopérativité entraîne une baisse de sensibilité au signal d'entrée et permet donc de réaliser la fonction désirée.

D Conclusions

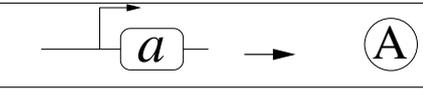
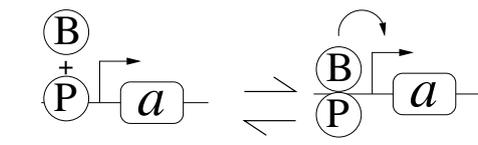
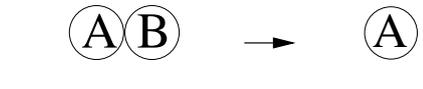
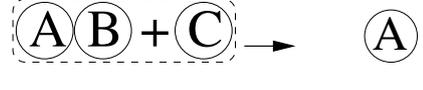
Nous avons donc vu dans cette partie que les réseaux génétiques ont une structure particulière. L'organisation fine des mécanismes vivants se traduit par une organisation particulière des interactions génétiques. Ainsi, les réseaux sont "invariants d'échelle". Des noeuds centraux très connectés contrôlent vraisemblablement de nombreux gènes très peu connectés. De plus, les gènes semblent être organisés en modules ou "supermodules". Il est ensuite possible que ces modules soient associés à des fonctions précises données. Les différents travaux de biologie synthétique présentés mettent en évidence la possibilité de fabriquer de véritables composants biologiques, chargés de traiter l'information à l'échelle de la cellule. Quelques interactions génétiques couplées peuvent se traduire par des comportements complexes (bistabilités, oscillations, intégration de fonction logiques...), à l'échelle de la cellule, voire de l'organisme. Le vivant est-il réellement organisé de façon si modulaire? Existe-t-il réellement une correspondance entre une fonction et un module génétique potentiel? Nous développons dans la suite des approches théoriques afin d'apporter quelques éléments de réponse à ces questions.

Chapitre II

Procédure d'Évolution *in silico*

LES MODULES GÉNÉTIQUES artificiels précédemment présentés ont été conçus dans une perspective d'ingénierie génétique, et présentent par exemple d'élégantes propriétés de symétries pour l'esprit humain. Cependant, les réseaux génétiques les plus efficaces réalisant certaines fonctions données n'ont aucune raison d'être très simples ou très symétriques. À titre d'exemple, examinons le cas du système bistable (ou switch), et de sa réalisation expérimentale, le "toggle switch". Nous avons vu dans la partie précédente que la réalisation biologique d'un tel système bistable nécessite de la coopérativité. Ceci signifie que le réseau d'interaction biochimique du "toggle switch" est plus compliqué que la première image "naïve" : de simples répressions transcriptionnelles ne suffisent pas, il faut des interactions supplémentaires pour expliquer la bistabilité (Voir Figure II.1 et référence [9]). On peut donc se demander s'il n'y a pas des circuits plus simples et peut-être moins artificiels réalisant la même fonction biochimique.

Afin d'apporter des éléments de réponse à ces différentes questions, nous avons mis au point une procédure d'évolution *in silico* pour concevoir de tels modules génétiques. L'idée de base est d'essayer de concevoir des modules génétiques réalisant des fonctions biochimiques données, sans aucune contrainte ou idée préconçue sur la nature des interactions nécessaires. Nous illustrons cette procédure en concevant des modules génétiques bistables (commutateurs ou "switchs") et des oscillateurs. Dans toute la suite, nous introduisons, et complétons nos résultats présentés dans la référence [26], reproduite ci-après.

#	REPRESENTATION	EQUATIONS
i)		$\frac{d}{dt}[A] = \tau_A[a] - \delta_A[A]$
ii)		$\frac{d}{dt}[a:P] = \theta[a:P:B] - \gamma[a:P][B]$ $\frac{d}{dt}[a:P:B] = \gamma[a:P][B] - \theta[a:P:B]$ $\frac{d}{dt}[A] = \tau_A[a:P] + \tau'_A[a:P:B]$
iii)		$\frac{d}{dt}[A] = -\tau_M[A]$ $\frac{d}{dt}[A^*] = \tau_M[A]$
iv)		$\frac{d}{dt}[A:B] = -\delta[A:B]$ $\frac{d}{dt}[A] = \delta[A:B]$
v)		$\frac{d}{dt}[A] = -\gamma[A][B]$ $\frac{d}{dt}[B] = -\gamma[A][B]$ $\frac{d}{dt}[A:B] = \gamma[A][B]$
vi)		$\frac{d}{dt}[B] = -\delta[A][B]$
vii)		$\frac{d}{dt}[A:B] = -\delta[A:B][C]$ $\frac{d}{dt}[C] = -\delta[A:B][C]$ $\frac{d}{dt}[A] = \delta[A:B][C]$

TAB. II.1: Liste de toutes les interactions possibles dans l'algorithme d'évolution, avec le schéma et les équations associées. Les constantes cinétiques sont représentées par des caractères grecs. A :B signifie ici le dimère formé par les protéines A et B, B :a représente la protéine B liée au promoteur du gène a. Les équations de la partie droite du tableau sont les termes associés à une réaction donnée : s'il y a plusieurs réactions, ces termes doivent être ajoutés entre eux. Par exemple, si une protéine A est à la fois produite et phosphorylée, en combinant les équations *i* et *iii*, l'équation d'évolution de la protéine A est $\frac{dA}{dt} = \tau_A[a] - \tau_M[A] - \delta_A[A]$

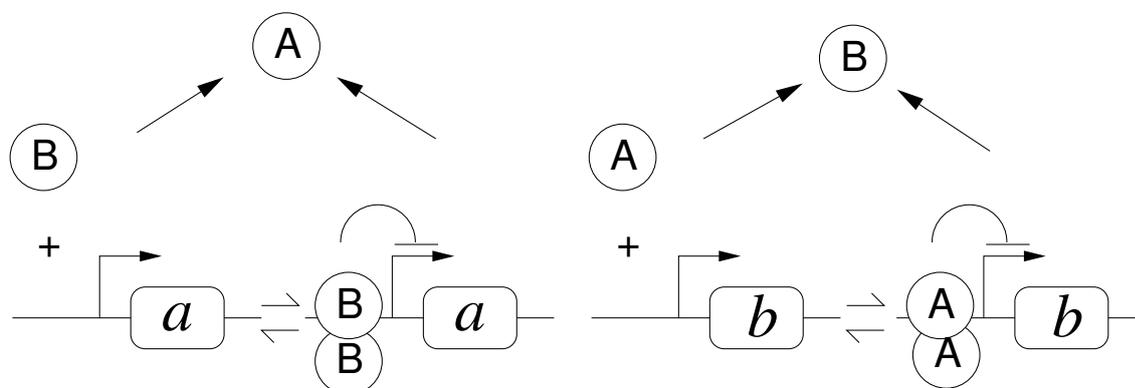


FIG. II.1: Exemple de réalisation réelle du toggle switch. Des interactions supplémentaires (par exemple des multimérisations) sont nécessaires pour expliquer le caractère bistable du toggle switch. Les flèches pointant vers les protéines A et B indiquent les réactions de production de ces protéines à partir des différents états du promoteur du gène.

A Algorithme d'évolution

A.1 Modélisation des réseaux génétiques

Un réseau génétique est modélisé par un ensemble de gènes et de protéines, ainsi que par un ensemble d'interactions biochimiques entre tous ces éléments. Toutes les réactions chimiques sont modélisées par des lois d'action de masse, avec des taux de réaction tirés au hasard puis modifiés dans la suite de l'algorithme. Les réactions possibles sont représentées dans la table II.1. Nous distinguerons deux types d'interactions génétiques :

- les interactions au niveau transcriptionnel. Ces interactions concernent toutes les régulations de la transcription d'un gène par une protéine. Ces réactions sont représentées par les figures *i*) et *ii*) dans le tableau II.1. En particulier, les mécanismes d'activation et de répression transcriptionnelles sont modélisés. Ainsi, dans l'exemple *ii*), si $\tau'_A > \tau_A$, la protéine B active la production de la protéine A, tandis qu'elle la réprime dans le cas contraire.
- les interactions au niveau post-transcriptionnel. Ces réactions sont représentées par les figures *iii*) à *vii*) dans le tableau II.1. À titre d'exemple, une protéine peut être phosphorylée (cas *iii*)), un dimère peut être dégradé en l'un de ses composants (cas *iv*)). Les autres interactions consistent en un ensemble de réactions protéine-protéine possibles : dimérisation (cas *v*)), où dégradations catalytiques (cas *vi*) et *vii*)).

Dans un premier temps, les étapes de transcription et de traduction sont condensées dans un seul terme, si bien que les protéines sont directement produites à partir

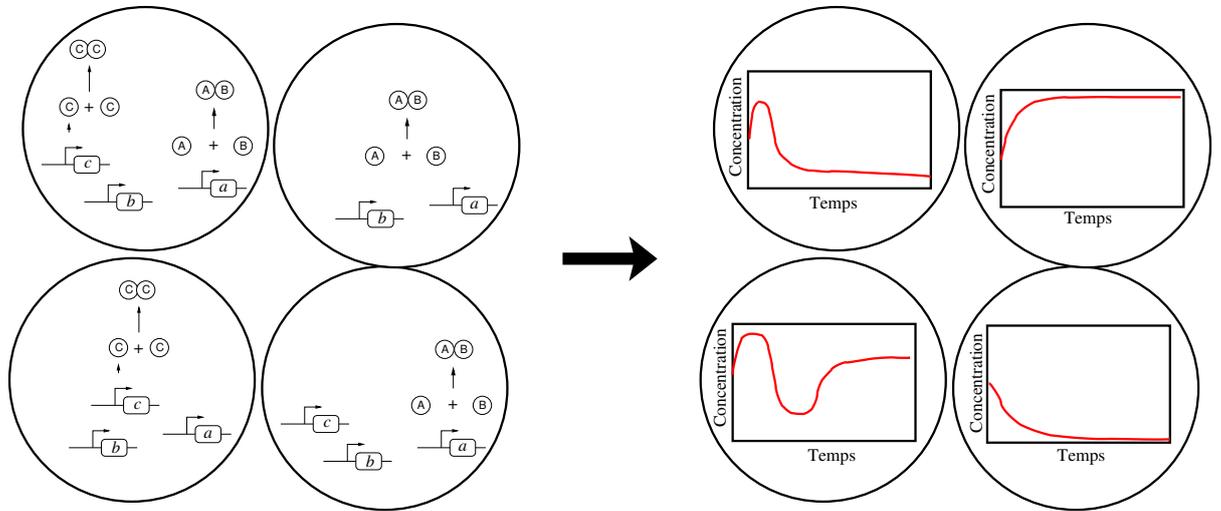


FIG. II.2: Algorithme : début de la phase de sélection puis intégration des équations différentielles. Supposons que dans cet exemple, nous voulions sélectionner des oscillateurs. Une fonction de score convenablement choisie nous indiquera alors que les deux cellules de gauche sont plus proches d’oscillateurs que les deux cellules de droite.

des gènes. Nous reviendrons sur cette hypothèse par la suite.

A.2 Une étape de l’algorithme

L’algorithme utilise une stratégie de croissance/sélection tout à fait générale. L’idée est d’avoir un ensemble de “cellules” (la population typique sera de 100 cellules), chacune avec un réseau génétique donné. À chaque étape de l’algorithme les cellules sont triées à l’aide d’une fonction de score permettant de mesurer l’adéquation des réseaux génétiques conçus à une “fonction” (typiquement commutateur ou oscillateur) donnée, puis la moitié des cellules ayant obtenu les plus bas scores est éliminée (phase de sélection, détaillée dans la partie A.2.a) . Les cellules restantes entrent alors dans une phase de croissance (détaillée dans la partie A.2.b) durant laquelle certaines cellules sont dupliquées afin de retrouver la population initiale et les réseaux génétiques des cellules nouvellement créées sont modifiés.

A.2.a Phase de sélection

Durant cette phase, les cellules sont évaluées. À chaque cellule correspond un circuit génétique et un ensemble d’équations différentielles associé. Ces équations sont intégrées avec différentes conditions initiales dépendant de la fonction désirée (Figure II.2).

À partir de ces évolutions dynamiques, une fonction de score est appliquée pour

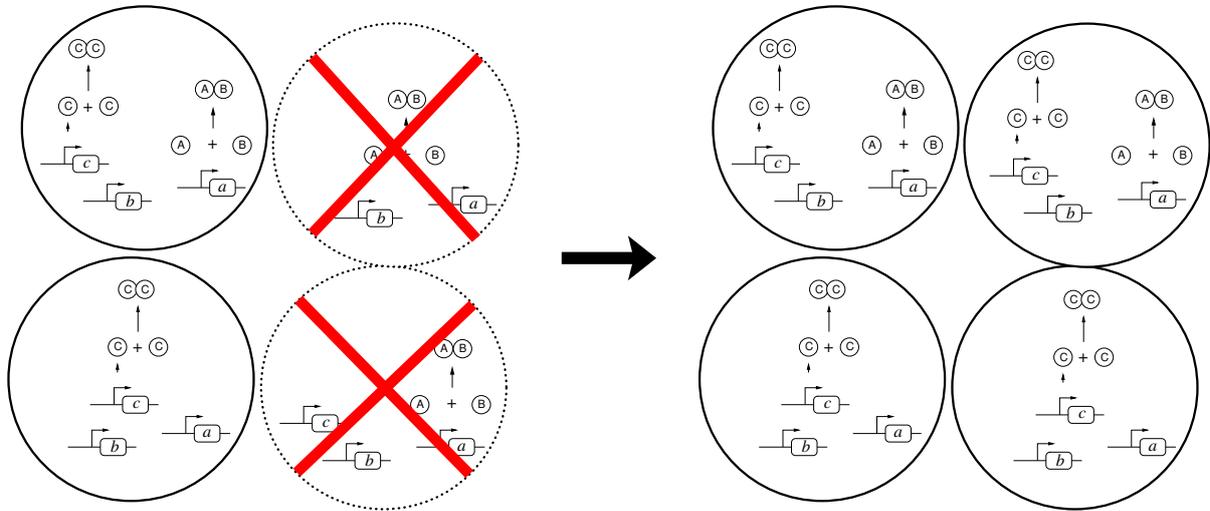


FIG. II.3: Algorithme : choix de la meilleure moitié. Cette moitié est ensuite dupliquée pour retrouver la population initiale.

évaluer quels sont les circuits génétiques les plus proches de la fonction désirée. Les fonctions spécifiques utilisées seront détaillées par la suite. Les cellules sont ensuite classées en fonction de leur score, puis la plus mauvaise moitié est éliminée et remplacée par la meilleure moitié (Figure II.3). On entre alors dans la phase de croissance.

A.2.b Phase de croissance

Durant la phase de croissance, des modifications sont faites dans les réseaux génétiques des cellules dupliquées. Les meilleurs résultats ont été obtenus avec deux modifications par étape de l'algorithme. Ces modifications sont tirées au hasard parmi les choix suivants :

- modification de la constante de dégradation d'une protéine.
- modification d'une constante cinétique d'une réaction.

Dans ce cas, on tire aléatoirement la constante à modifier, puis on tire un nombre uniformément entre 0 et 2 et on multiplie la constante choisie par ce nombre. Ces modifications ne modifient donc pas la "topologie" du réseau d'interaction.

Les autres "mutations" aléatoires modifient le réseau d'interaction :

- ajout d'un gène. On ajoute un nouveau gène et une nouvelle protéine. On tire au sort de nouvelles constantes cinétiques pour la transcription et la dégradation de la protéine.
- ajout d'une régulation transcriptionnelle. On tire au sort un promoteur (éventuellement déjà lié par une protéine) et une protéine. Il faut alors ajouter

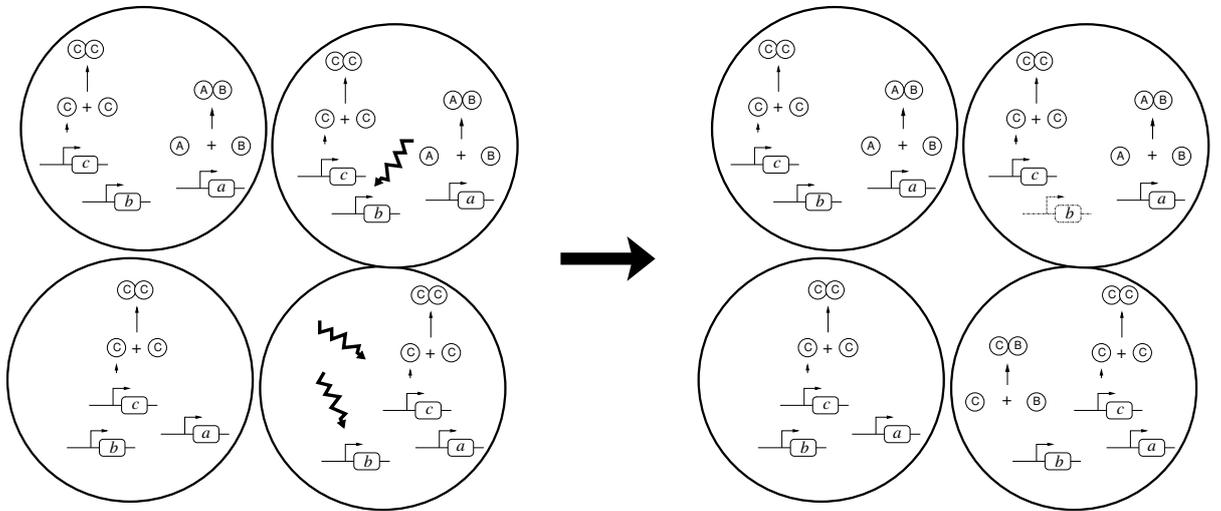


FIG. II.4: Algorithme : phase de croissance. Des mutations sont faites sur la moitié dupliquée. Par exemple, ici, pour la cellule en haut à droite, le taux de transcription de la protéine B est modifié, tandis qu'une nouvelle interaction protéine protéine est ajoutée pour la cellule en bas à droite. Puis on itère le processus avec cette nouvelle population.

trois nouvelles réactions au système, et tirer au sort les constantes cinétiques associées : l'attachement de la protéine au promoteur, le détachement de la protéine de ce promoteur et enfin le taux de transcription modifié dû à la présence de cette nouvelle protéine.

- ajout d'une réaction post-transcriptionnelle. On choisit d'abord aléatoirement si cette nouvelle réaction implique une ou deux protéines. Dans le premier cas, il y a deux possibilités : soit la protéine est modifiée et donne une nouvelle protéine ayant des propriétés différentes, soit un complexe peut se dissocier pour redonner une protéine existante. Dans le second cas, on peut avoir soit une dimérisation donnant un nouveau complexe, soit une dégradation catalytique donnant une protéine existante.

Précisons à ce stade que lorsqu'une nouvelle protéine ou un nouveau complexe est ajouté, on suppose que cette nouvelle espèce n'a aucune interaction avec les autres protéines dans le système. Ainsi, si la protéine A active la transcription d'un gène b, et forme ensuite un dimère avec la protéine B, la dimère A :B n'est a priori pas capable d'activer la transcription du gène b. Par la suite, les mutations aléatoires pourront éventuellement ajouter une régulation de l'expression du gène b par ce dimère A :B.

A.2.c États initiaux et finaux

Un phénomène très général dans les algorithmes d'optimisation évolutionnaire est l'apparition d'une multiplicité d'interactions non pertinentes pour résoudre un problème donné¹. Dans le cas présent, un premier problème est qu'il apparaît parfois des réactions qui ne jouent strictement aucun rôle dynamique, et il serait évidemment souhaitable de prévoir un mécanisme pour supprimer ces réactions. Cependant, des réactions inutiles à une étape de l'évolution peuvent s'avérer cruciales à une étape ultérieure lors de l'ajout d'une boucle de rétroaction, d'où la nécessité de trouver un équilibre entre la suppression et la conservation de réactions a priori inutiles. Un second problème est que l'algorithme d'évolution cherchant à optimiser une fonction de score, il peut apparaître des interactions spécifiques à l'optimisation de cette fonction de score mais qui ne sont pas nécessaires à la fonction génétique calculée. Par exemple, pour un oscillateur génétique, on peut supprimer certaines réactions, ce qui diminuera le score, tout en conservant le caractère oscillant du système. Par conséquent, nous avons inclus un terme de pénalisation dans la fonction de score pour empêcher la prolifération des interactions génétiques. A score égal, on privilégiera toujours le réseau génétique ayant le moins d'interactions. Concrètement, en appelant N le nombre de réactions, le score d'un réseau génétique est multiplié par $1 + N\varepsilon$ avec ε petit. Une fois l'algorithme terminé, nous simplifions les réseaux génétiques obtenus afin de conserver les seules réactions nécessaires au bon fonctionnement du réseau. Les interactions sont alors supprimées une à une dans l'ordre inverse de leur apparition et seules sont conservées les réactions qui ne modifient pas la fonction génétique désirée.

Précisons également qu'aucune condition n'est imposée sur les états initiaux des modules génétiques : pour tous les réseaux génétiques présentés dans la suite, les cellules de la première génération comportent deux gènes sans aucune interaction transcriptionnelle ou post-transcriptionnelle. L'algorithme se charge ensuite d'ajouter les interactions nécessaires à la réalisation de la fonction. Ainsi, aucune contrainte sur la nature des réseaux ou sur leur symétrie n'est imposée, contrairement à la plupart des circuits génétiques artificiels fabriqués *in vivo* tels que le "toggle switch" ou le "répressilateur".

Nous avons utilisé l'algorithme pour fabriquer deux types de fonctions : des commutateurs génétiques ("switchs") et des oscillateurs génétiques. Nous présentons dans la suite les fonctions de score associées ainsi que les circuits génétiques obtenus.

¹phénomène communément appelé "code bloat" [25]

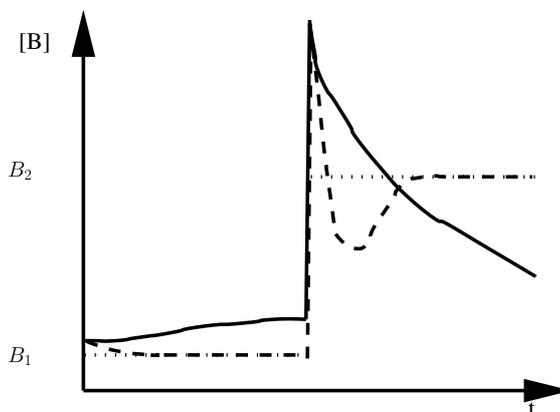


FIG. II.5: Fonction du score du commutateur génétique. Ligne pleine : évolution d'un circuit génétique non bistable. Tirets : système bistable. Pointillés : commutateur idéal.

B Commutateurs génétiques

B.1 Fonction de score

Les commutateurs génétiques que nous souhaitons fabriquer sont des systèmes bistables, avec au moins deux protéines A et B. Les deux états stables désirés sont similaires aux états stables du “toggle switch” : nous souhaitons obtenir un premier état stable avec une haute concentration de A et une basse concentration de B (couple $([A]_1, [B]_1)$, $[A]_1 \gg [B]_1$), et un second état stable avec une haute concentration de B et une basse concentration de A (couple $([A]_2, [B]_2)$, $[A]_2 \ll [B]_2$).

Afin d'évaluer la fonction de score des réseaux génétiques obtenus, nous choisissons tout d'abord des conditions initiales proches du premier état stable $([A]_1, [B]_1)$. Nous laissons ensuite évoluer le système pendant un temps T arbitraire et évaluons dans quelle mesure le système reste proche de cet état stable en calculant l'intégrale $S_1 = \int_0^T ([A](t) - [A]_1)^2 + ([B](t) - [B]_1)^2 dt$.

Ensuite, nous introduisons dans le système un pulse de protéines B, c'est-à-dire que nous montons soudainement la concentration de la protéine B à haut niveau. Le but est de perturber le système afin qu'il atteigne le second état stable. Nous laissons ensuite évoluer le système pendant le même temps T et évaluons de la même façon l'intégrale $S_2 = \int_T^{2T} ([A](t) - [A]_2)^2 + ([B](t) - [B]_2)^2 dt$.

Le score total du circuit génétique est donné par la somme $S = S_1 + S_2$ (voir Figure II.5).

Le temps T typique est de l'ordre de 100 minutes. L'amplitude du pulse de protéines B est du même ordre de grandeur que la concentration $[B]_2$.

B.2 Exemples typiques des commutateurs obtenus

De manière générale, les commutateurs obtenus sont assez différents des “toggle switches” classiques et incorporent à la fois des régulations transcriptionnelles et post-transcriptionnelles. Deux motifs en particulier ont été fréquemment trouvés et sont détaillés dans la suite.

Le premier motif de commutateur fréquemment trouvé par l’algorithme comporte uniquement deux interactions génétiques (voir Figure II.6). De manière identique au toggle switch, la protéine A réprime l’expression du gène b , mais la coopérativité n’est pas nécessaire pour cette répression. La seconde interaction est une dimérisation entre les protéines A et B, où le dimère est incapable de réprimer la transcription de B. Concrètement, lorsque A est présent à haute concentration, A réprime l’expression du gène b et le peu de protéines B produites sont titrées par les protéines A, si bien qu’on a un premier état stable avec beaucoup de A, mais peu de B. Le rôle de la dimérisation apparaît beaucoup plus clairement pour le second état stable : lorsqu’il y a beaucoup de B, les protéines B en excès titrent les protéines A présentes si bien qu’elles empêchent la répression. Si le taux production des protéines B est assez grand, on atteint alors le second état stable avec beaucoup de protéines B, quelques dimères A :B et peu de protéines A libres (voir Figure II.6 ; pour une analyse plus détaillée et plus quantitative, voir également le chapitre III).

Le second exemple typique de commutateur obtenu à l’aide de l’algorithme (Figure II.7) est lui aussi composé de deux interactions génétiques. Dans ce commutateur, il n’y a aucune répression au niveau transcriptionnel ; le seul mécanisme de régulation négative est la dimérisation entre les protéines A et B. La seconde régulation est une autoactivation du gène b par la protéine B. A haute concentration de B, le mécanisme d’autoactivation assure une production de B suffisante pour titrer les protéines A produites. On a donc un premier état stable avec beaucoup de B, quelques dimères A :B et peu de A libre. À basse concentration de B, les protéines A titrent les protéines B produites si bien que l’autoactivation de B ne peut avoir lieu et on reste dans un état stable où l’on a peu de B, beaucoup de A et peu de dimère A :B.

B.3 Autres commutateurs

Les deux commutateurs présentés ci-dessus sont représentatifs des commutateurs obtenus à l’aide de l’algorithme d’évolution. D’autres exemples de commutateurs sont présentés en annexe A. D’autres topologies ont été plus rarement obtenues. A

Réactions	Constantes	Stabilité
$a \rightarrow a+A$	0.20	0.9 -1.4
$A \rightarrow X$	0.0085	0.0-1.5
$b \rightarrow b+B$	0.37	0.7-1.3
$B \rightarrow X$	0.034	0.0-8.9
$A+B \rightarrow A :B$	0.72	0.1 - >10
$b+A \rightarrow b :A$	0.19	0.7-7.6
$b :A \rightarrow b+A$	0.42	0.2-1.5
$b :A \rightarrow b :A+B$	0.027	0.0-2.3

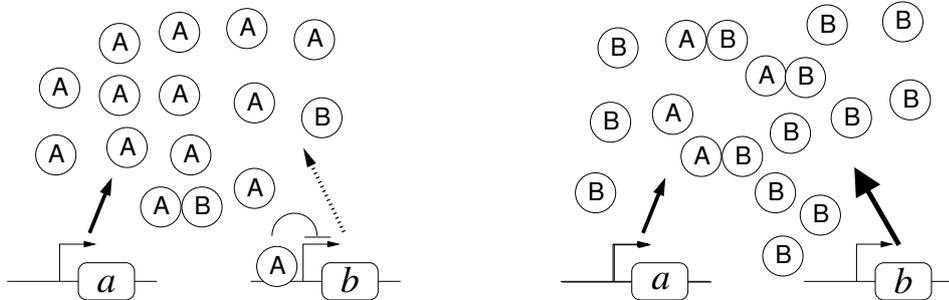
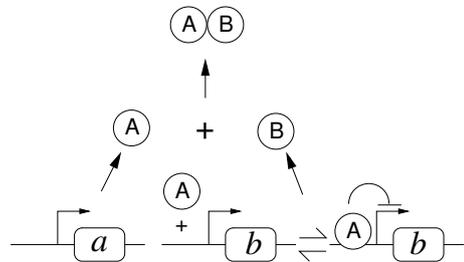


FIG. II.6: Premier exemple de commutateur typique obtenu avec l'algorithme. Haut : Schéma du commutateur, constantes cinétiques associées à chaque réaction et domaine de variation de chacun des paramètres conservant le caractère bistable du réseau. Bas : Représentation qualitative des états stables du commutateur.

Réactions	Constantes	Stabilité
$a \rightarrow a+A$	0.45	1.0-1.3
$A \rightarrow X$	0.024	0.0-1.0
$b \rightarrow b+B$	0.11	0.0-1.0
$B \rightarrow X$	0.030	0.8-3.1
$A+B \rightarrow A :B$	0.10	1.0-10<
$b+B \rightarrow b :B$	0.44	0.4-1.0
$b :B \rightarrow b+B$	0.57	1.0-2.5
$b :B \rightarrow b :B+B$	0.89	0.8-1.0

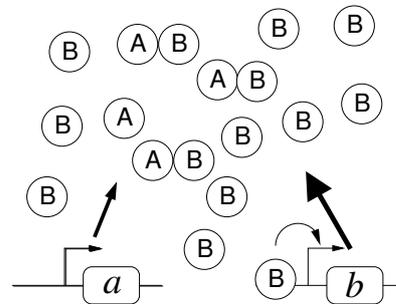
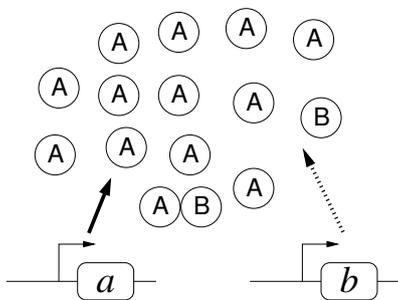
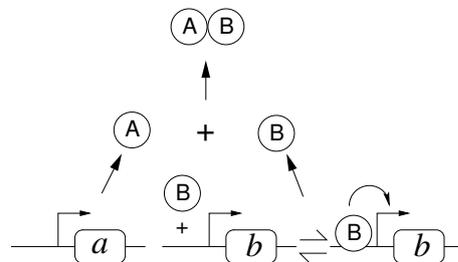


FIG. II.7: Second exemple de commutateur typique obtenu avec l'algorithme. Haut : Schéma du commutateur, constantes cinétiques associées à chaque réaction et domaine de variation de chacun des paramètres conservant le caractère bistable du réseau. Bas : Représentation qualitative des états stables du commutateur.

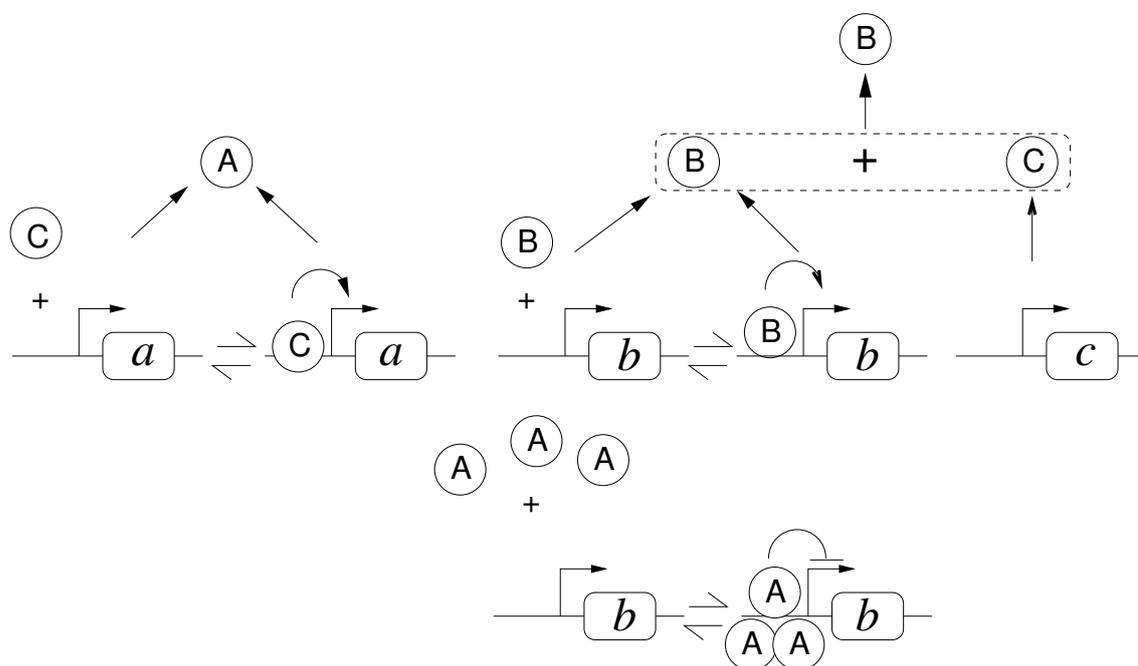


FIG. II.8: Un exemple de commutateur à trois gènes (les constantes pour ce commutateur sont données dans la table A.1, en annexe).

titre d'exemple, un commutateur à trois gènes est représenté en Figure II.8. Soulignons à ce stade que l'algorithme peut ajouter des gènes, et que ces gènes peuvent ensuite être utilisés pour clore des boucles de feedback. Nous avons également obtenu un autre exemple de commutateur à trois gènes présenté en annexe Figure A.5. Cependant, dans ce dernier exemple, la protéine B joue simplement le rôle de “rapporteur” et ne joue aucun rôle dynamique dans le commutateur.

Les commutateurs présentés ci-dessus reposent tous sur des interactions post-transcriptionnelles. Il est néanmoins possible d'obtenir des commutateurs reposant uniquement sur des interactions transcriptionnelles. De tels commutateurs ont été obtenus en diminuant considérablement la probabilité d'apparition d'une interaction protéine-protéine dans l'algorithme. Des exemples de type “toggle switch” sont présentés Figure II.9 et Figure A.1 en annexe. Conformément à l'étude théorique, une simple inhibition mutuelle ne suffit pas à assurer la bistabilité, et dans les circuits obtenus grâce à l'algorithme, on peut noter l'apparition de coopérativité et d'auto-activation, en plus des mécanismes d'inhibitions réciproques.

Reactions	Constantes
$a \rightarrow a+A$	0.092
$A \rightarrow \text{Nothing}$	0.021
$b \rightarrow b+B$	1.5
$B \rightarrow \text{Nothing}$	0.11
$a+A \rightarrow a:A$	0.50
$a:A \rightarrow a+A$	0.56
$a:A \rightarrow a:A+A$	0.38
$a+B \rightarrow a:B$	0.59
$a:B \rightarrow a+B$	0.082
$a:B \rightarrow a:B+A$	0.0085
$b+B \rightarrow b:B$	0.47
$b:B \rightarrow b+B$	0.27
$b:B \rightarrow b:B+B$	3.0
$b:B+A \rightarrow b:B:A$	0.64
$b:B:A \rightarrow b:B+A$	0.67
$b:B:A \rightarrow b:B:A+B$	0.88
$b:B:A+A \rightarrow b:B:A:A$	0.38
$b:B:A:A \rightarrow b:B:A+A$	0.13
$b:B:A:A \rightarrow b:B:A:A+B$	0.014

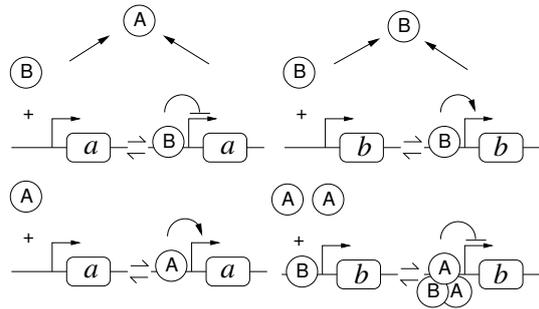


FIG. II.9: Un exemple du type “toggle switch”.

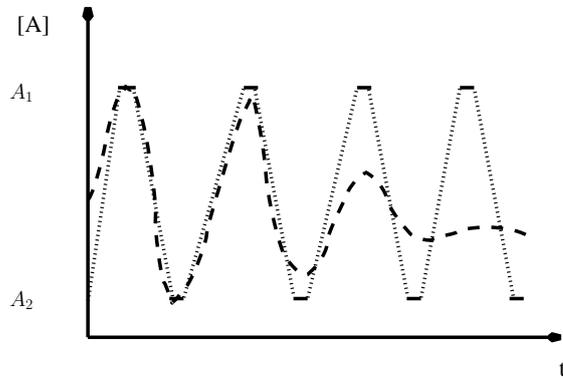


FIG. II.10: Fonction de score de l’oscillateur. Ligne pleine : évolution d’un circuit génétique non oscillant. Pointillés : oscillateur idéal.

C Oscillateurs

C.1 Fonction de score

Plusieurs fonctions de score ont été essayées pour fabriquer des oscillateurs génétiques (notamment en se basant sur l'analyse de Fourier par exemple). Cependant, une seule fonction de score nous a permis d'obtenir des oscillateurs génétiques : nous considérons donc l'évolution de la concentration d'une protéine A pendant un long intervalle de temps fixé, et nous souhaitons que la concentration de cette protéine oscille entre deux valeurs $[A]_1$ et $[A]_2$. Une période arbitraire T est choisie, ainsi qu'un petit intervalle de temps dT . Pour les demi-périodes $\tau = T/2, 3T/2, 5T/2, \dots$ nous calculons l'intégrale $\int_{\tau}^{\tau+dT} ([A](t) - [A]_1)^2 dt$. Pour les périodes $\tau = T, 2T, 3T, \dots$ nous calculons l'intégrale $\int_{\tau}^{\tau+dT} ([A](t) - [A]_2)^2 dt$. Nous ajoutons les valeurs de ces intégrales à la fonction de score. L'idée est de véritablement "apprendre" au circuit génétique à osciller entre les deux concentrations fixées (voir Figure II.10). Précisons également qu'afin de ne pas fixer la phase de l'oscillateur, les concentrations initiales des différentes protéines sont elles aussi soumises à l'évolution par l'algorithme et peuvent être aléatoirement modifiées au cours de la phase de croissance.

C.2 Exemples d'oscillateurs obtenus

La Figure II.11 présente l'évolution d'un réseau aboutissant à la création d'un oscillateur représenté Figure II.12. L'évolution dynamique de l'oscillateur final et de cinq de ses ancêtres est présenté. A la génération 260 les oscillations apparaissent et le score est ensuite optimisé rapidement dans les générations suivantes.

L'algorithme nous a permis de générer plusieurs types d'oscillateurs. Contrairement au cas des commutateurs, il ne semble pas possible de catégoriser précisément les différents oscillateurs. Ils existe néanmoins certains points communs entre plusieurs réseaux génétiques fabriqués. En particulier, dans la plupart des modules trouvés, une activation transcriptionnelle est associée à une répression post-transcriptionnelle par l'intermédiaire d'une dimérisation : le répresseur titre l'activateur transcriptionnel (voir Figures II.12 et II.13). Nous avons également obtenu certains oscillateurs basés purement sur des interactions protéine-protéine (Figure II.14)². Les constantes cinétiques associées à ces oscillateurs sont données en annexe A.

²Notons à ce stade que des oscillateurs purement biochimiques semblent effectivement exister dans la nature. On peut citer l'exemple du rythme circadien chez les cyanobactéries [63].

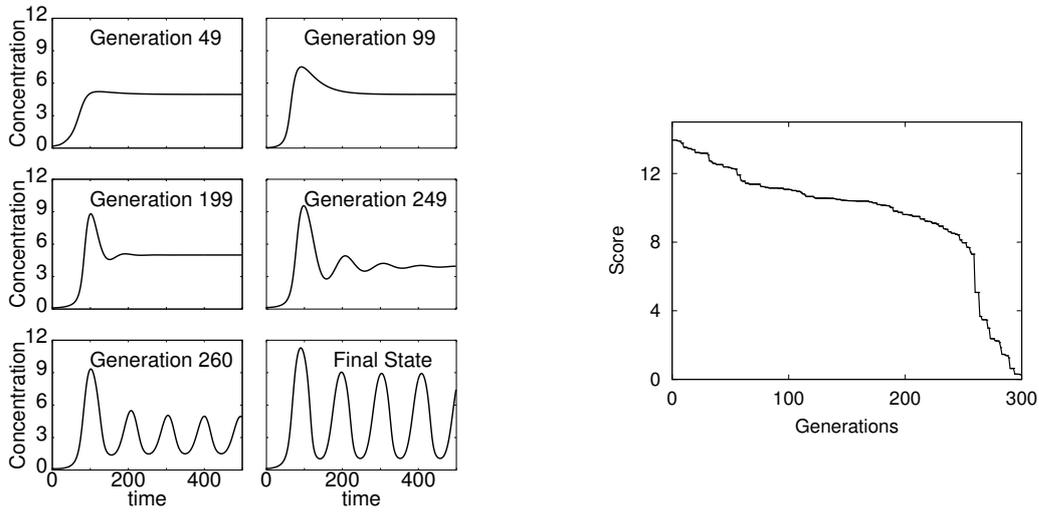


FIG. II.11: Évolution du score montré pour six générations différentes aboutissant à la création d'un oscillateur final correspondant à l'oscillateur de la Figure II.12.

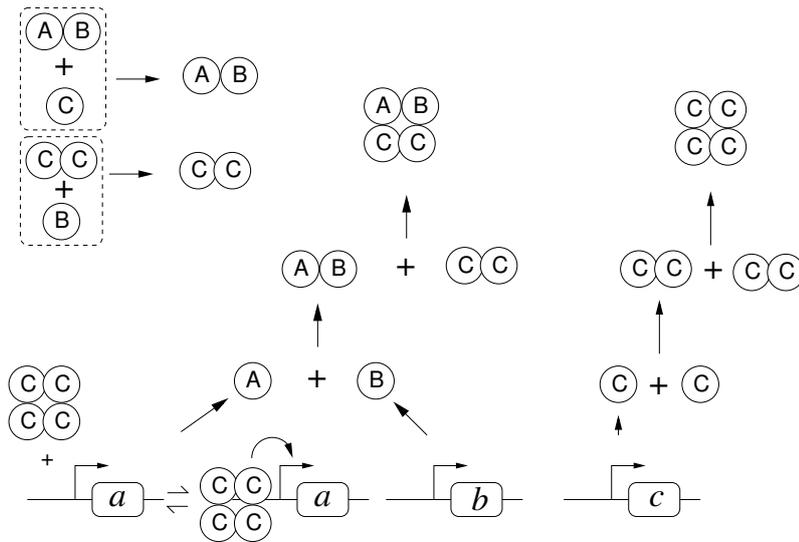


FIG. II.12: Un premier oscillateur obtenu à l'aide de l'algorithme. Noter l'activation du gène a par C puis la dimérisation entre $A :B$ et $C :C$.

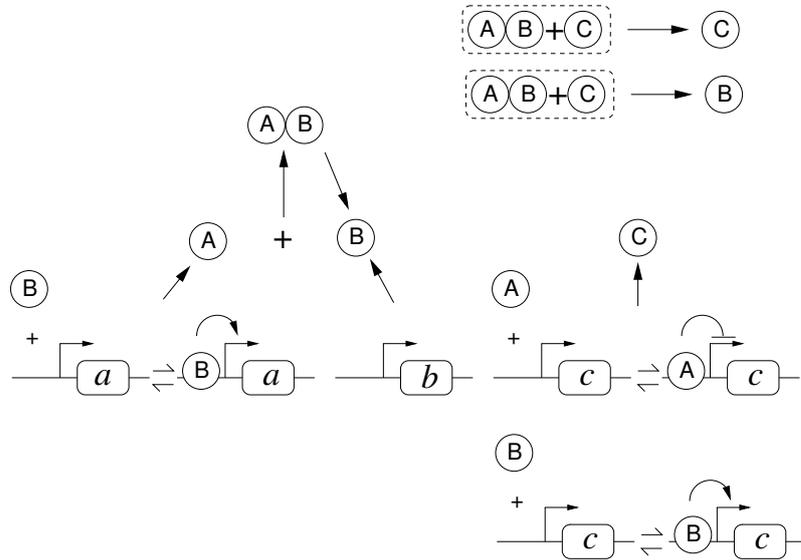


FIG. II.13: Un second oscillateur obtenu à l'aide de l'algorithme. Noter l'activation du gène a par B puis la dimérisation entre A et B .

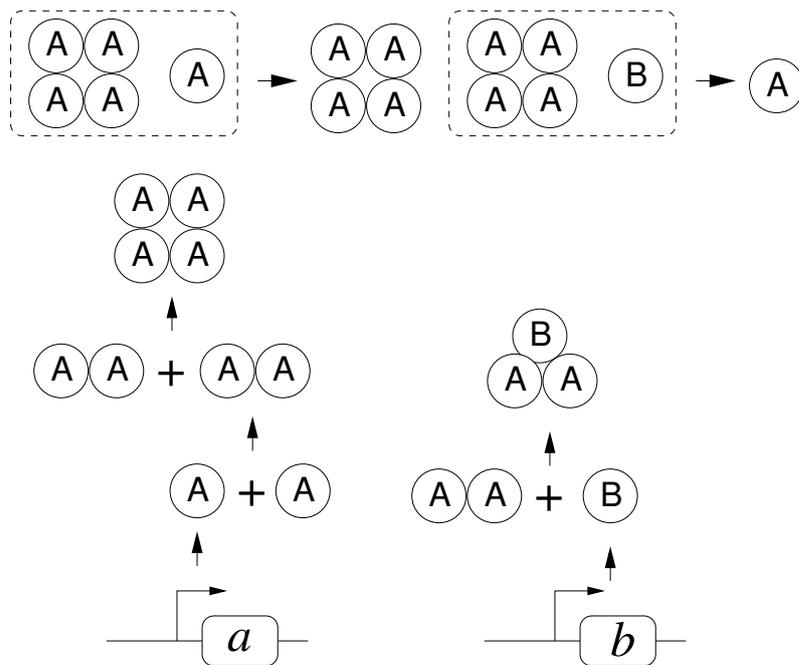


FIG. II.14: Un troisième oscillateur sans aucune régulation transcriptionnelle.

D D'autres applications possibles de l'algorithme

L'algorithme est assez souple pour être modifié et adapté afin de résoudre différentes questions posées par les régulations au niveau génétique. Nous examinons dans la suite trois exemples précis où l'algorithme fonctionne et peut potentiellement apporter un éclairage sur les problèmes de dynamique des réseaux génétiques.

D.1 Rôle de l'ARN dans la dynamique des oscillateurs génétiques

L'une des premières modifications apportées à l'algorithme est l'ajout de l'étape transcription/traduction négligée dans un premier temps. Le rôle de l'ARN sur les commutateurs est assez limité : nous obtenons alors les mêmes topologies que précédemment. En revanche, comme nous le verrons dans les parties suivantes, l'introduction de l'ARN semble pouvoir jouer un rôle crucial dans la dynamique des oscillateurs génétiques. Aussi avons-nous implémenté de nouveau l'algorithme pour trouver des oscillateurs en incluant cette étape de la transcription. Les résultats de l'algorithme d'évolution dans ce cas sont bien meilleurs : l'algorithme converge plus vite et les circuits génétiques obtenus sont plus simples. En particulier, l'algorithme parvient à fabriquer des oscillateurs génétiques à deux gènes, ce qui n'était en général pas le cas pour l'algorithme sans la transcription. Des exemples d'oscillateurs génétiques ainsi obtenus sont présentés Figures II.15 et II.16, ainsi qu'en appendice. Les oscillateurs nouvellement obtenus ont des topologies assez proches de celles des commutateurs. Dans le cas de l'oscillateur II.15, nous avons deux protéines A et B. La protéine A est cette fois capable de s'autoréprimer, tandis que B forme un dimère avec A et peut le dégrader catalytiquement, amenant la diminution de la concentration de A tout en empêchant l'autorépression, amenant donc l'augmentation de la concentration de A. La topologie de cet oscillateur est donc tout à fait proche de celle du commutateur II.7, où l'autoactivation serait remplacée par une autorépression. De la même façon, pour l'oscillateur II.16, la protéine B active la transcription du gène a et B forme un dimère avec la protéine A, empêchant cette activation. La topologie de cet oscillateur est donc analogue à celle du commutateur II.6, où la répression serait remplacée par une activation. Cette analogie entre commutateurs et oscillateurs sera développée et analysée plus longuement dans le chapitre III.

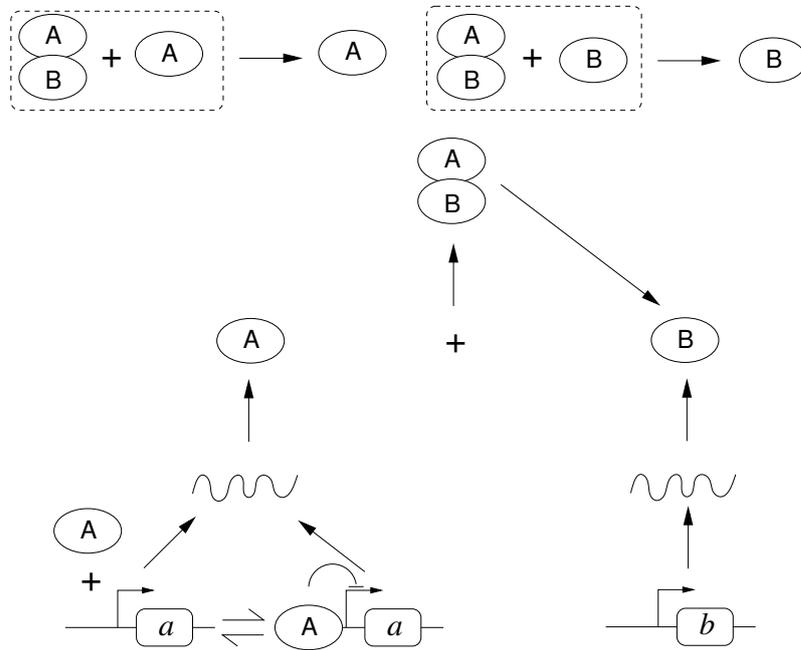


FIG. II.15: Un exemple d'oscillateur avec une topologie proche du commutateur de la Figure II.7. Noter l'autorépression du gène *a* par la protéine A.

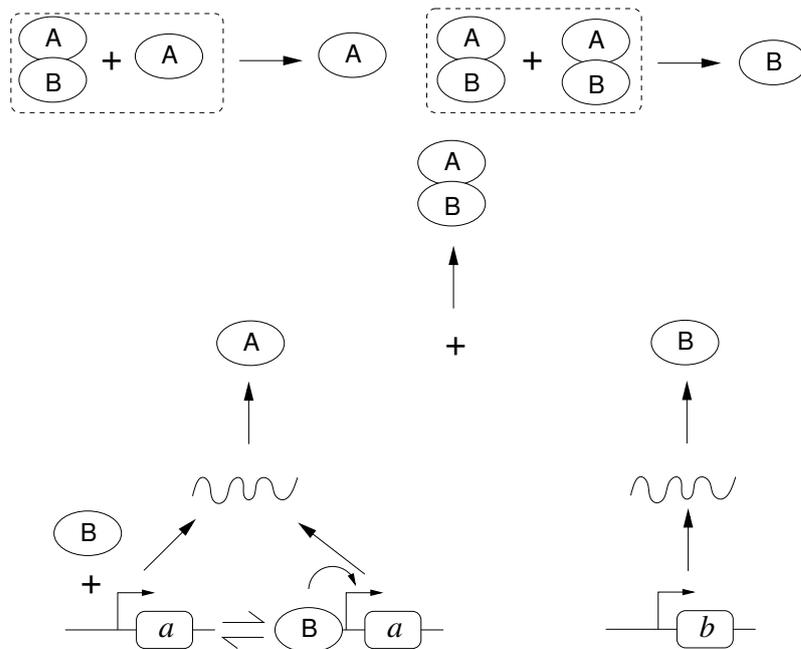


FIG. II.16: Un exemple d'oscillateur avec une topologie proche du commutateur de la Figure II.6. Noter l'activation du gène *a* par B et la dimérisation entre A et B.

D.2 Compensation de température

L'une des propriétés essentielles des horloges circadiennes est la compensation de température (voir chapitre IV) : les horloges circadiennes sont des oscillateurs génétiques qui, bien que sensibles à la température (en particulier l'amplitude des oscillations de protéines varient), gardent une période indépendante de la température. Il est possible d'inclure dans un modèle d'oscillateur génétique une dépendance en température par l'intermédiaire de la loi d'Arrhenius. Nous supposons donc qu'une constante chimique k suit la loi suivante de dépendance en température suivante :

$$k(T) = k_{\infty} e^{-E_a/RT}$$

où T représente la température et E_a l'énergie d'activation. À l'aide de l'algorithme et à partir d'un modèle d'horloge circadienne (présenté dans les chapitres III et IV), il est possible d'écrire une fonction de score qui optimise à la fois les paramètres et les énergies d'activation afin de choisir une période indépendante de la température. Plus précisément, nous partons d'un oscillateur circadien déjà oscillant³, et nous programmons l'algorithme de manière à ce qu'il optimise uniquement les énergies d'activation et les constantes cinétiques des circuits existants. Puis nous utilisons la même fonction de score que celle utilisée pour générer les oscillateurs génétiques pour l'oscillateur évoluant à deux températures différentes : 300 et 310 K. En particulier, nous contraignons l'amplitude et la période de la protéine A respectivement à 10 et à 23 unités arbitraires. L'algorithme génère alors des oscillateurs génétiques compensés en température. Un exemple est présenté Figure II.17, avec l'évolution de la protéine B en fonction de la température Figure II.18.

D.3 Vers la construction de motifs spatiaux

Enfin, une application de l'algorithme en cours de développement concerne l'ajout de la dimension spatiale dans la modélisation des réseaux génétiques, dans l'optique de traiter des problèmes de différenciation. En effet, l'algorithme peut aider à comprendre comment un même réseau génétique peut amener à différentes fonctions cellulaires. A titre d'exemple, la segmentation chez la drosophile est contrôlée par une protéine appelée Bicoid (Bcd). Bcd est présent dans l'oeuf de la drosophile et forme un gradient antéro-postérieur. Il a été ensuite montré que ce gradient mène à l'expression localisée de protéines, menant à la formation de bandes qui ensuite don-

³C'est le seul cas présenté dans ce chapitre où les modules génétiques n'ont pas été conçus entièrement par l'algorithme

Réactions	Constantes	E_a/R
$r_a \rightarrow X$	0.6	4827
$A \rightarrow X$	0.18	2189
$r_b \rightarrow X$	0.14	2721
$B \rightarrow X$	0.001	2138
$a \rightarrow a + r_a$	5	3313
$r_a \rightarrow r_a + A$	1	3447
$b \rightarrow b + r_b$	0.05	2202
$r_b \rightarrow r_b + B$	2.38	3029
$A+B \rightarrow A:B$	11.3	4889
$b+A \rightarrow b:A$	0.003	3108
$b:A \rightarrow b + A$	0.23	2449
$b:A \rightarrow b:A + r_b$	8.6	4854

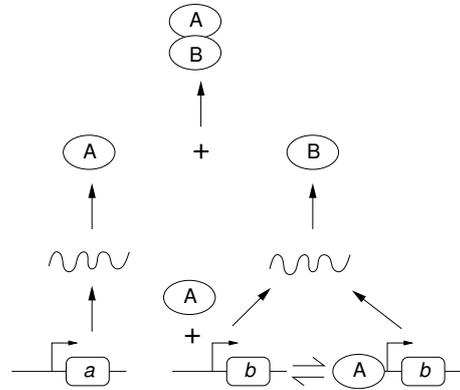


FIG. II.17: Schéma de l'oscillateur génétique optimisé pour avoir compensation de température avec les constantes et les énergies d'activation associées. Les constantes données sont les constantes à 300K.

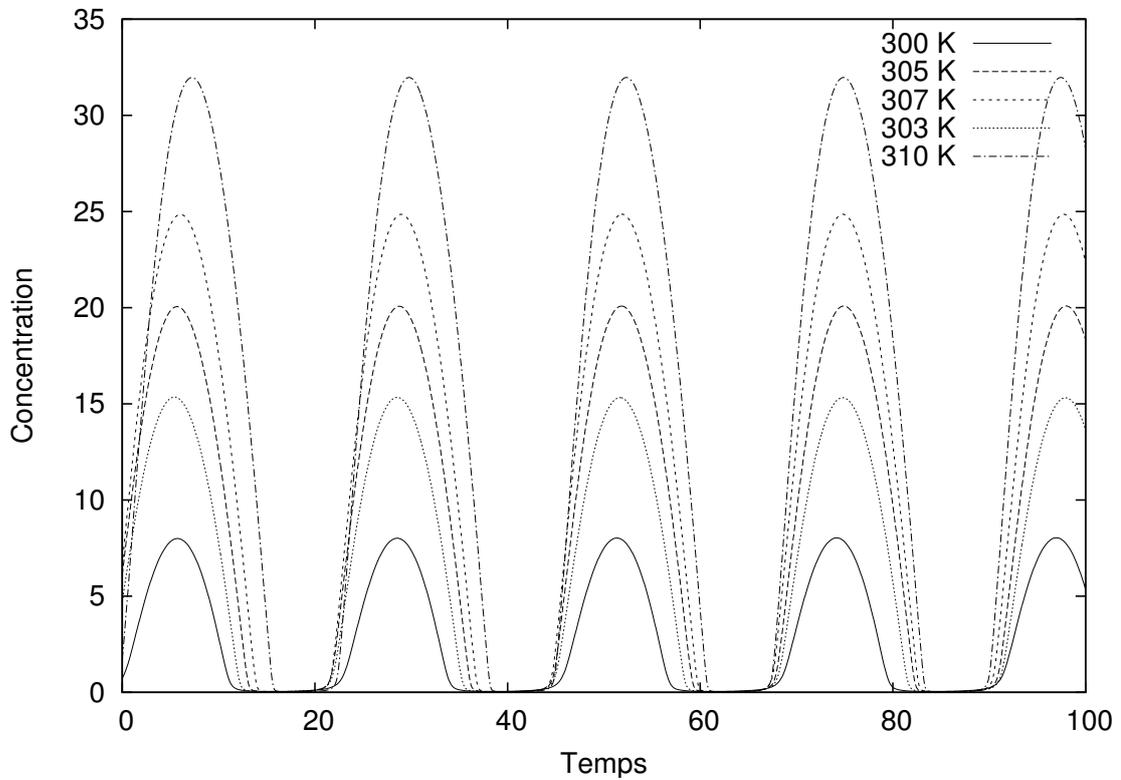


FIG. II.18: Un exemple d'oscillateur compensé en température obtenu à l'aide de l'algorithme. Il est possible d'avoir de fortes variations en amplitude sans modifier significativement la période de l'oscillateur.

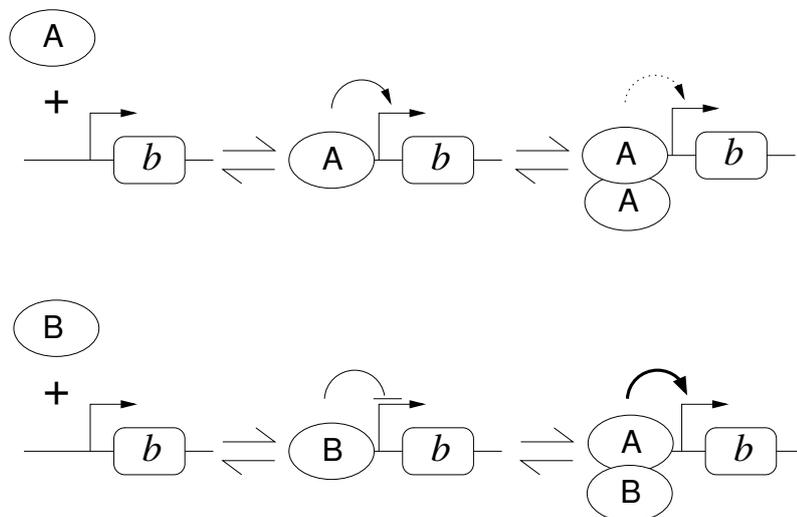


FIG. II.19: Interactions génétiques donnant lieu à une expression localisée de B. Plus une flèche est épaisse, plus le taux de production est important. Le jeu des interactions/répressions mène à une expression maximale de B pour une concentration moyenne de A : à haute concentration, deux A viennent diminuer le taux de transcription de B par rapport au taux maximum de production.

neront lieu à la segmentation et à la différenciation des cellules. Pour une revue des différents modèles proposés, on pourra consulter la revue de Thieffry et Sanchez [79].

Nous avons appliqué l'algorithme pour générer un cas très simplifié de différenciation : étant donné un gradient spatial (type Bcd), nous avons fabriqué des réseaux génétiques menant à l'expression localisée d'une protéine (en négligeant la diffusion dans un premier temps). Un exemple de réseau génétique de ce type est représenté Figure II.19, avec le profil d'expression de la protéine B en fonction du gradient de A Figure II.20.

E Conclusions préliminaires et motivations

E.1 Comparaison des réseaux obtenus avec des réseaux réels

Il est intéressant de comparer les modules obtenus aux modules existant réellement dans la nature. Le commutateur génétique présenté Figure II.6 est par exemple assez proche d'un mécanisme épigénétique dans l'opéron lactose et qui avait été déjà évoqué par Jacob et Monod [62]. La production de la protéine B peut être vue comme un raccourci du processus d'entrée d'allolactose dans la cellule : la transcription de l'opéron lactose (via la production de β -galactosidase et de perméase) est en effet suivie de l'entrée de l'allolactose dans la cellule sous l'action de la perméase. L'allolactose se lie ensuite au répresseur de l'opéron lactose (qui joue le rôle de la protéine

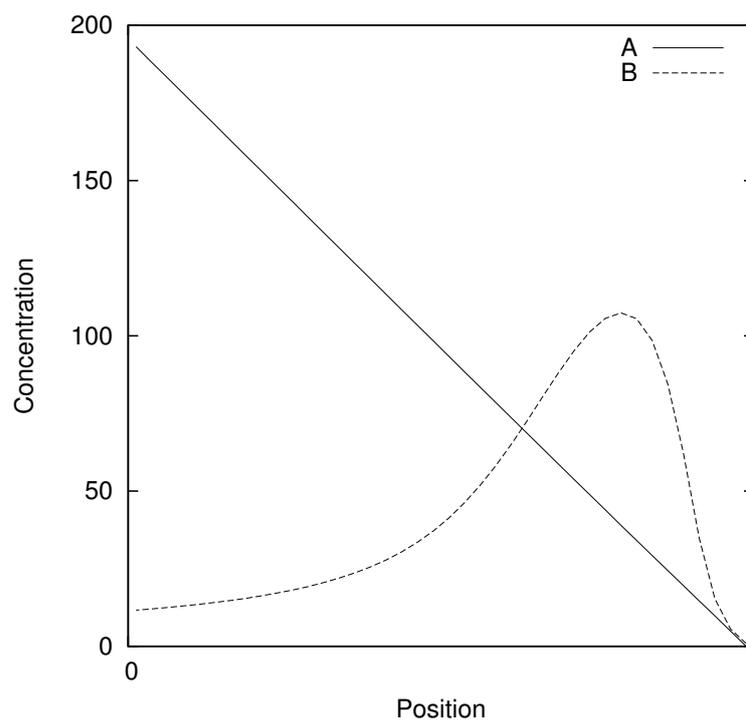


FIG. II.20: Concentration à l'équilibre de la protéine B (pointillés) en fonction du gradient de la protéine A (en traits pleins). L'expression de la protéine B est maximale pour une concentration de A aux alentours de 50, en unités arbitraires.

A) et empêche donc son action répressive.

Le principe de fonctionnement du commutateur présenté Figure II.7 est très semblable au commutateur basé sur une boucle de rétroaction positive impliquée dans le développement de la compétence chez *Bacillus Subtilis* [80]. Ce processus d'internalisation d'ADN exogène est sous le contrôle du gène *comK* qui peut activer sa propre transcription comme le gène *b*. Dans les cellules non compétentes, deux protéines, *MecA* et *CplC*, se lient à *ComK* et empêchent donc son action autoactivatrice, jouant le rôle de la protéine A.

Le principe de fonctionnement des oscillateurs obtenus (notamment les oscillateurs des Figures II.12, II.13, II.16) est très proche du principe de certains oscillateurs génétiques, en particulier les oscillateurs circadiens (voir chapitre IV) et l'oscillateur p53/Mdm2. p53 est une protéine impliquée dans la réponse au stress. Lorsque des cellules sont exposées à un mutagène, par exemple, la transcription de p53 est activée, mais à son tour, p53 active la transcription de Mdm2. Finalement, Mdm2 se lie à p53 et catalyse sa dégradation. En fonction du niveau de stress, ces différentes rétroactions se traduisent par des oscillations [3]. Cet oscillateur comme les oscillateurs circadiens et comme les oscillateurs obtenus à l'aide de l'algorithme est basé sur une boucle de rétroaction négative particulière : une protéine active l'expression d'une ou plusieurs autres protéines (niveau de la transcription), qui vont réprimer cette activation en interagissant avec la protéine activatrice (niveau post-transcription). Rappelons que le commutateur présenté Figure II.6 était lui aussi basé sur deux interactions de natures différentes entre deux protéines : l'une au niveau transcriptionnel, l'autre au niveau post-transcriptionnel. Plus généralement, ces modules basés sur deux interactions de type différent ("Mixed Feedback Loop", ou MFL) sont effectivement surreprésentés dans les réseaux génétiques [89, 90]. Des exemples de MFL sont reproduits dans la table II.2. La plupart de ces MFL sont impliquées dans les processus de différenciation cellulaire.

E.2 Conclusions

La procédure présentée ici fonctionne donc et semble pouvoir être validée par des exemples biologiques. Une caractéristique essentielle des différents réseaux obtenus est l'importance des interactions post-transcriptionnelles. Ainsi, les fonctions obtenues ne peuvent se réduire à des réseaux transcriptionnels, et les rétroactions dues aux interactions protéine-protéine sont centrales dans le fonctionnement des modules obtenus.

L'algorithme présenté doit être également vu comme une approche complémen-

Espèce A	Espèce B	Fonction biologique
Système bistable		
Lac Repressor	Allolactose	Métabolisme du lactose
Oscillateurs		
WC-1, WC-2	FRQ	Horloge circadienne de Neurospora
dCLK	PER, TIM	Horloge circadienne de la Drosophile
CLOCK, BMAL	PER, CRY	Horloge circadienne des mammifères
p53	Mdm2	Réponse au stress des mammifères
MFLs de la levure		
Swi6	Swi4	régulation de la transition G1/S
Gal4	Gal80	métabolisme du galactose
Gal4	Gal3	métabolisme du galactose
Gal4	Gal1	métabolisme du galactose
Ime1	Rim11	activation de la Méiose
Ume6	Ime1	activation de la Méiose
Ste12	Fus3	2 circuits, croissance pseudohyphale
Ste12	Far1	croissance pseudohyphale
Cbf1	Met28	métabolisme du soufre
Met4	Met28	métabolisme du soufre
Swi4	Clb2	Cycle cellulaire
Mbp1	Clb5	transition G1/S
Stb1	Cln1	inconnue
Stb1	Cln2	inconnue

TAB. II.2: Exemples biologiques de MFLs

taire des méthodes expérimentales existantes, présentées en introduction. D'un côté, l'approche *in silico* est certainement moins difficile à mettre en oeuvre qu'un véritable travail expérimental. Nous avons pu concevoir des réseaux plus compliqués que ceux obtenus par Guet et al. [37], et le processus d'évolution en lui-même est plus complexe et mieux contrôlé que celui obtenu par des expérimentations *in vivo* [91] ou par d'autres approches numériques [12]. De plus, les résultats et les différentes étapes de l'évolution *in silico* sont par nature plus accessibles et plus détaillées. D'un autre côté, il est évident que les approches *in silico* sont très simplifiées et il n'est pas sûr que les modules obtenus par l'algorithme d'évolution fonctionnent tels quels *in vivo*, un peu à l'image du circuit génétique présenté Figure I.5.

Enfin, soulignons que contrairement aux approches expérimentales classiques de fabrication de module [18, 31], aucune hypothèse sur la topologie des modules n'est faite. L'algorithme peut ainsi suggérer de nouveaux mécanismes, de nouveaux processus dynamiques ou mettre en exergue l'importance d'un processus donné dans les réseaux génétiques. Aussi avons nous développé différents modèles de modules génétiques, présentés dans les parties suivantes.

F Article

Nous reproduisons dans la suite l'article détaillant l'algorithme de base d'évolution, qui correspond à la référence [26].

Design of genetic networks with specified functions by evolution *in silico*

Paul François and Vincent Hakim*

Laboratoire de Physique Statistique, Centre National de la Recherche Scientifique–Unité Mixte de Recherche 8550, Ecole Normale Supérieure, 24, Rue Lhomond, 75231 Paris, France

Edited by Nancy J. Kopell, Boston University, Boston, MA, and approved November 19, 2003 (received for review July 19, 2003)

Recent studies have provided insights into the modular structure of genetic regulatory networks and emphasized the interest of quantitative functional descriptions. Here, to provide *a priori* knowledge of the structure of functional modules, we describe an evolutionary procedure *in silico* that creates small gene networks performing basic tasks. We used it to create networks functioning as bistable switches or oscillators. The obtained circuits provide a variety of functional designs, demonstrate the crucial role of posttranscriptional interactions, and highlight design principles also found in known biological networks. The procedure should prove helpful as a way to understand and create small functional modules with diverse functions as well as to analyze large networks.

Large gene networks are increasingly thought of as being built from smaller subnetworks (1–3) or “modules.” It is thus important to understand the structure and dynamics of small functional building blocks. Recently, this has been pursued in new ways by using both experimental (4–9) and computer approaches (10). For instance, statistical analyses have been used to determine recurring motifs of interactions (11) in transcriptional networks. However, both the direct problem of finding the function associated with a given motif of interactions and the inverse problem of finding gene networks performing given functions are not straightforwardly solved for several reasons. In most cases, the knowledge of only a partial subset of the existing interactions renders the determination of a module and its function difficult. Even when fully known, the geometry of the interactions between proteins and genes constrains but does not determine network behavior. Whether a network possesses a single functioning state, can be induced to switch between different states, or oscillates in time depends on the quantitative interactions between its components (12, 13) as, for instance, recently shown for the biologically important case of the NF- κ B module (14).

The inverse question can be illustrated by the design of a bistable switch, arguably one of the simplest functional elements. As classically conceived, this element is built out of reciprocal repression between two genes encoding transcription factors (15), as sketched in Fig. 1. In simple terms, when gene *a* is actively transcribed, the allied protein A is abundant and represses the transcription of gene *b*. B is thus expressed at a low level. A symmetric possibility is that gene *b* is active and B abundant, whereas *a* is repressed and A is present at a low concentration. A recently successful synthetic realization (5) demonstrates that indeed such a network of interactions can sustain two coexisting states. Mathematical analysis, however, shows that the laws of chemical interactions render it more complicated than naively thought (16): repression described by simple Michaelis–Menten kinetics is not sufficient to produce a working switch, and high-order Hill functions are required with, for instance, protein dimers or higher multimers interacting with DNA. When considering an existing gene network or the design of a new one, it would be useful to know whether a bistable switch can be made only out of two mutually repressing transcription factors or whether other interaction networks, less easily con-

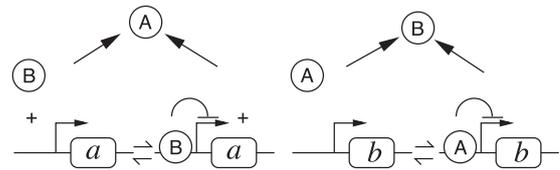


Fig. 1. Sketch of a bistable switch with reciprocal transcriptional repression between genes *a* and *b*.

ceived, could serve the same purpose, perhaps even in a better and more robust way.

To provide theoretical insight into this question, we wondered whether it is feasible to determine *a priori* the possible designs of a small genetic network performing a given basic function. To sample the variety of working possible schemes, we have designed an evolutionary procedure *in silico* that we describe below. Computer-simulated evolution was previously used to optimize the kinetic parameters of a chemical circuit of fixed topology (17). The more general goal of the present procedure is to obtain a network of genes and proteins implementing a chosen function without imposing *a priori* any particular design. Here, we illustrate the general procedure by using it to create bistable switches and oscillating networks.

The Evolutionary Procedure *in Silico*

The algorithm proceeds in successive rounds of growth and selection of a collection of independent networks (typically ≈ 100 “cells”), a general strategy in evolutionary computation (18). Growth consists of the enlargement of the network collection by the addition of mutated copies of the existing networks. Selection takes place as follows: first, a score is attributed to each network in the enlarged collection depending on how well its dynamics approaches what is required. Second, a fraction of the networks is chosen based on the score attribution, and the remaining ones are deleted to reduce the collection to its original size before the next growing phase. This general algorithm requires a number of specific choices, and we precisely describe those that we have made.

In our present implementation, a genetic network is defined by a number of genes and proteins[†] and chemical reactions described by deterministic rate equations. Possible chemical reactions are listed in Fig. 2 and consist of promotion or repression of gene transcription by proteins, along with posttranscriptional interactions. Posttranscriptional interactions have been included in two forms: two proteins can interact to produce another one (modeling formation of a complex catalytic

This paper was submitted directly (Track II) to the PNAS office.

*To whom correspondence should be addressed. E-mail: hakim@lps.ens.fr.

[†]For simplicity, in the present first implementation, mRNAs are not explicitly described because regulation of translation or of mRNA stability is known to exist (19) but is less widely documented than transcriptional control.

© 2004 by The National Academy of Sciences of the USA

#	REPRESENTATION	EQUATIONS
i)		$\frac{d}{dt}[A] = \tau_A[a] - \delta_A[A]$
ii)		$\frac{d}{dt}[a:P] = \theta[a:P:B] - \gamma[a:P][B]$ $\frac{d}{dt}[A] = \tau_A[a:P] + \tau'_A[a:P:B]$
iii)		$\frac{d}{dt}[A] = -\tau_M[A]$ $\frac{d}{dt}[A^*] = \tau_M[A]$
iv)		$\frac{d}{dt}[A:B] = -\delta[A:B]$ $\frac{d}{dt}[A] = \delta[A:B]$
v)		$\frac{d}{dt}[A] = -\gamma[A][B]$ $\frac{d}{dt}[B] = -\gamma[A][B]$ $\frac{d}{dt}[A:B] = \gamma[A][B]$
vi)		$\frac{d}{dt}[B] = -\delta[A][B]$
vii)		$\frac{d}{dt}[A:B] = -\delta[A:B][C]$ $\frac{d}{dt}[C] = -\delta[A:B][C]$ $\frac{d}{dt}[A] = \delta[A:B][C]$

Fig. 2. List of possible reactions. The schematic representations are used to represent the reactions in Figs. 3–5. In the allied rate equations, Greek letters (τ , δ , γ , θ) denote kinetic constants; A:B denotes the bound complex of protein A and B; and a:P denotes gene *a* with protein *P* bound on its promoter. Reaction *ii* is illustrated here only for the case of an already existing bound complex between a protein *P* and the promoter of a gene *a*. The same reaction is also possible between a protein and a “naked” promoter (i.e., without *P*). Only the term corresponding to the displayed reaction has been written on the right side of the equations, giving the evolution of protein concentrations. In a given network of reactions, all such terms should be added to obtain the evolution of a particular species (e.g., the evolution of a protein A produced from gene *a* and only undergoing a posttranscriptional modification would be obtained by combining *i* and *iii*: $d[A]/dt = \tau_A[a] - \delta_A[A] - \tau_M[A]$).

degradation), or one protein can be modified (modeling phosphorylation or other posttranslational modifications).

The algorithm starts with a collection of independent networks. These starting networks simply comprise two genes and their allied proteins, with randomly drawn production and degradation rates and no other reactions present. Then, alternate phases of growth and selection are performed to evolve the network collection and obtain networks achieving the specified function.

Growth Phase

The collection size is doubled by addition of a mutated copy of each network in the collection. The mutated copy differs by a fixed number of mutations from its parent network (two mutations gave satisfying results, and this is the choice made for the results presented below). The mutations are randomly drawn in succession from the five different following possibilities:

- (i) The degradation rate of a protein is modified.
- (ii) A kinetic constant of one reaction is modified.

In these first two cases, the constant is chosen at random among the existing ones and is multiplied by a random number uniformly drawn between 0 and 2.

The three other possibilities of mutations involve adding a new reaction from among those listed in Fig. 2:

(iii) A new gene is created, and the reaction corresponding to the production and degradation of the allied protein (Fig. 2*i*) is added to the network.

(iv) A new interaction between a protein and a gene promoter is introduced (Fig. 2*ii*). We randomly choose a protein and a gene or an existing gene/protein complex (bound complexes of gene promoters with proteins may have been previously created) and create a new entity describing their bound complex. Three reactions are added to the network: the binding of the protein to the gene promoter (or to the existing complex), the reverse reaction, and the modified production rate of the protein allied to the gene in the new complex state.

(v) Finally, a posttranscriptional reaction can be added. The choice of whether the reaction involves a single previously introduced protein or two proteins is made at random. In the single-protein case, a protein (or protein complex) is randomly drawn from among the existing ones. A new entity corresponding to a modified (e.g., phosphorylated) version of this protein is introduced (Fig. 2*iii*) together with the rate of the posttranscriptional modification and the degradation rate of the modified protein. When a protein complex is drawn, the reaction can alternatively be a partial degradation, with one protein of the complex chosen as the degradation product (Fig. 2*iv*). In the two-protein case, two proteins (or protein complexes) are drawn at random, and the reaction is chosen among possibilities *v*, *vi*, and *vii* of Fig. 2. It can be

- a dimerization (Fig. 2*v*). A new entity describing a bound complex of the two chosen proteins is introduced together with its rates of formation and degradation;
- catalytic degradation (Fig. 2*vi*) with the degraded species chosen between the two selected proteins with equal probability; or
- a partial catalytic degradation (Fig. 2*vii*) if one of the drawn entities is a protein complex. In this case, one of the proteins forming the complex is chosen as the reaction end product.

In cases *iii*–*v*, when a new reaction is added, its kinetic constants are randomly drawn. We assume the cellular volume to be unity so that concentrations and numbers of molecules are equivalent. The time unit is taken to be the minute. For simplicity, all kinetic constants are initially drawn randomly between 0 and 1 in these units, with uniform probability.[‡] Kinetic constants obtained for the final selected circuits depend also on the imposed scoring function and are within the physiological range for the two cases investigated below. Further knowledge of physiological constraints in given cases could be incorporated by restricting the range of possible initial values and of mutation changes (cases *i* and *ii*).

The respective probabilities p_a, \dots, p_e of the five mutation possibilities are fixed in each set of simulations. Good results were obtained with comparable probabilities for the different possibilities.[§]

Selection Phase

Once the network collection has been enlarged by the addition of mutated copies of existing networks, each network is evaluated. This step consists of integrating the set of coupled differential equations corresponding to the network reactions [using a Runge–Kutta algorithm (20)]. This dynamical evolution serves to evaluate a scoring function and allows for the ranking of the networks. The choice of the scoring function depends on the function to be selected and is detailed below for the two cases

[‡]This corresponds, for instance, to drawing initial protein degradation rates between 0 and 1 min^{-1} and initial forward rates of second-order reactions between 0 and $\approx 1 \text{ nm}^2/\text{min}$ for a typical bacterial volume.

[§]Switches like those of Fig. 3 A and B were obtained for a large set of different mutation rates. However, to obtain in the same set of simulations switches as those of Fig. 3 and switches based on reciprocal inhibition (Figs. 6 and 7), the probability p_e should be taken ≈ 20 times smaller than p_d .

investigated here. Based on the networks' ranking, the top-scoring half of the collection is kept and the other half deleted.[¶] At the end of the selection phase, the network collection has regained its original size and is ready for the next round of growth and selection.

For the cases investigated below, this evolution procedure produces networks achieving the required function within <100 generations for switches and a few hundred generations for oscillators.

Results and Discussion

Bistable Switches. In this case, the *in silico* evolution was directed toward obtaining a network with two possible stable states differing in particular in the concentrations [A], [B] of proteins A and B. In the first stable state, the desired concentration [A]₁ of protein A was high, and the concentration [B]₁ of protein B was low, whereas in the second one, [A]₂ was low and [B]₂ high. Each network was started at time 0 with initial concentrations corresponding to one of the desired states, for instance with [A]₁ and [B]₁. The network dynamical evolution was then computed from time 0 to time *T* and the concentrations of proteins A and B monitored to assess how close they stayed to their desired values. A pulse of B was then added at time *T* to try to switch the network to the second desired state [A]₂, [B]₂. The network evolution was then followed from *T* to 2*T* and the concentrations of proteins A and B monitored again. A score was then assigned depending on how well the system under consideration had approached the two desired states in the two dynamical periods [that is, depending on the proximity to zero of the sum of the integral from time 0 to time *T* of $([A] - [A]_1)^2 + ([B] - [B]_1)^2$ in the first phase and of the integral from time *T* to 2*T* of $([A] - [A]_2)^2 + ([B] - [B]_2)^2$ in the second phase]. The time interval *T* was typically taken to be ≈100 min. The procedure was also run for different values of the imposed concentrations, with “high” concentrations ([A]₁, [B]₂) going from a few tens to several hundreds of molecules and “low” ([A]₂, [B]₁) concentrations from one to a few tens of molecules. The magnitude of the switching pulse of B was taken to be of the same order as the desired stable high concentration of B ([B]₂) to avoid selecting networks with only a weakly stable first fixed point. Implementation of this procedure resulted in the creation of a variety of networks showing the required dynamical behavior. The networks were usually endowed with a multiplicity of accessory reactions that served to optimize the chosen scoring function but that also simply reflected evolutionary “trials” [this feature is related to the phenomenon of “code bloat” in evolving programs (18)]. To mitigate this phenomenon, a term depending on the number of reactions was added to the score function to penalize reaction multiplication and to direct evolution toward network simplification once a satisfying set of reactions had been found. The final networks were systematically pruned by deleting reactions from the most recently added to the oldest to leave a core set of reactions giving rise to the bistable behavior.

Three of these “core” networks are depicted in Fig. 3 together with the values of the kinetic constants. A sample of others is displayed in Figs. 6–12, which are published as supporting information on the PNAS web site. They show a number of noteworthy features. Most interesting, the obtained networks have various designs and are quite different from the “classical” one shown in Fig. 1. Fig. 3*A* displays a frequently found motif. Protein A represses gene *b* at the transcriptional level as in the

classical case. However, protein B simply acts through complexation with A, the complex AB being unable to repress gene *b*. Thus if A is high, *b* is repressed, but if B is high, all of the As are titrated and complexed, and B remains high. In contrast to the case of Fig. 1, simple rate equations are now sufficient to make the design work. Fig. 3*B* provides an equally simple scheme but quite strikingly, without any transcriptional repression. Activation by B of its own gene naturally provides the high B state. In this state, complexation of A with B leads at the same time to a low level of free (uncomplexed) A. However, another state is also possible when the concentration of B is low, because it is transcribed at a low basal level without autoactivation, and an increase in B concentration is prevented by complexation with the abundant free A. Fig. 3*C* provides an example of a more complicated scheme with three genes that again crucially involves posttranscriptional interactions. Of note, our evolutionary algorithm produced several variations on the classical switch of Fig. 1 with reciprocal repression at the transcription level (Figs. 6 and 7) but less frequently than alternative designs. This probably results from the larger number of elementary reactions needed to really make the design work.

Parameter values for the networks of Fig. 3*A* and *B* are given in Fig. 3 (values for the network of Fig. 3*C* are given in *Supporting Text*, which is published as supporting information on the PNAS web site) for imposed high concentrations ([A]₁, [B]₂) of only a few tens of proteins. Parameter values are provided in Figs. 8 and 9 for the same networks with several hundred proteins in the high concentration species. This indicates that the networks are able to perform the required function for a wide range of parameters. We further checked that the bistable behavior of the obtained networks did not depend on the precise kinetic parameters produced by the algorithm. A quantitative and simple measure of network sensitivity to parameter variation is obtained by varying one parameter at a time, maintaining all of the others fixed; the results are shown in Fig. 3. Evolution using deterministic dynamics for network evolution as used here tends to give some kinetic constants close to one end of their permitted interval of variation (marginally satisfying the constraints), but the bistable behavior holds for a range of kinetic parameters. Admissible variations range from 30% to 50% for the few very sensitive parameters, to >10-fold.

The importance of noise in biological circuits has been emphasized (21) and has been the subject of several recent experimental studies (22–24). A detailed modeling of stochastic steps in transcription and translation is beyond the scope of the present investigation. Nonetheless, to further assess the robustness of the selected networks and to test their resistance to noise, we simulated the switch networks of Fig. 3*A* and *B* with the stochastic dynamics corresponding to their rate equations, as detailed in *Supporting Text*. When the high-concentration species in the stable states have several hundred proteins, the finite particle number induces some fluctuations in the number of proteins but does not notably affect the networks' switch function (Fig. 13, which is published as supporting information on the PNAS web site). The networks still perform quite clearly as switches with only a few tens of particles in the high-concentration species despite much stronger fluctuations (Fig. 14, which is published as supporting information on the PNAS web site). So, the selected designs perform reassuringly well in the presence of noise. Under strong fluctuation conditions, the network of Fig. 3*A* switches spontaneously once every few hours between the two stable states (Fig. 14*A*). It can thus appear less robust to noise fluctuations than the design of Fig. 3*B*, which displays much rarer spontaneous switches (Fig. 14*B*). This is not an intrinsic weakness of the Fig. 3*A* scheme, however, because for other kinetic parameters, its stability is comparable to that of Fig. 3*B* (data not shown).

[¶]It might be thought that a milder selection procedure that would keep a proportion of bad-scoring circuits would be more efficient in some cases. This was tried by selecting the networks stochastically, based on the score attribution (high-scoring networks being retained with higher probability than low-scoring ones). For the particular examples studied here, this did not significantly change the convergence rate: the innovation rate is slow enough that significant improvements spread to all cells in a few generations.

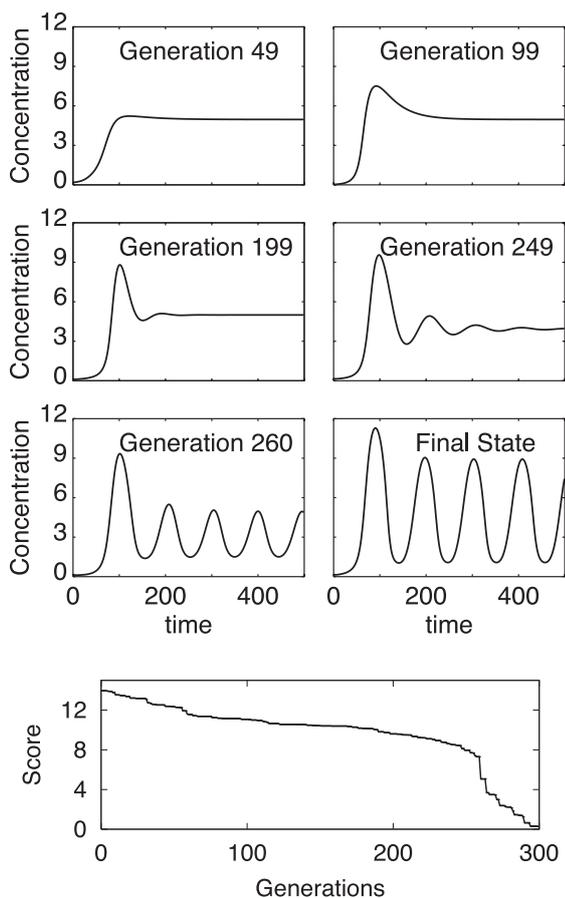


Fig. 4. Dynamics of six networks at different stages in the evolutionary process leading to the creation of an oscillatory network. The score evolution is shown (Bottom).

documented biological examples. The switch of Fig. 3A embodies a functioning principle that is, in a simplified way, quite analogous to that of the classic *lac* operon and was, in fact, also envisaged in ref. 15. The production of B by the gene *b* can be thought of as a shortcut version of the augmentation of allolactose (B) resulting from transcription of the *lac* operon (*b*) (via

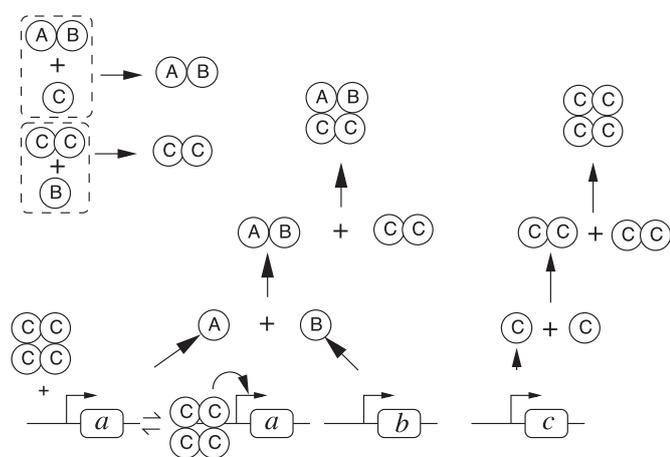


Fig. 5. The "core" oscillating network extracted from the evolutionary process shown in Fig. 4. The kinetic constants for this network are provided in Fig. 15. In Fig. 4, the rapid score decrease and emergence of oscillations at generation 260 are associated with the creation of the complex ABC.

production of β -galactosidase and a membrane permease). Allolactose in turn binds to an operon repressor protein, a role played by A in Fig. 3A, and lifts the repression of the operon transcription, analogous to the effect of complexation of B and A.

The principle of Fig. 3B appears to be used in the switch underlying *Xenopus* oocyte maturation from a G₂-arrested phase after exposure to progesterone. Conversion from the graded progesterone signal is thought to take place in the Mos/MEK1/p42 mitogen-activated protein (MAP) kinase (MAPK) cascade (25, 26). Mos can be thought of as playing the role of B in Fig. 3B, with positive feedback in the MAPK cascade arising via p42-induced Mos mRNA polyadenylation (27) and Mos stabilization (28), instead of the simple self promotion of *b* by protein B. The role of A appears to be played by the β -subunit of casein kinase II that binds to Mos and inhibits it (29).

The molecular switch controlling development of competence in *B. subtilis* under some starvation conditions seems to be even more directly related to the switch of Fig. 3B. Development of competence, that is, the ability to bind and internalize exogenous DNA, is under the control of a master competence gene *comK* that, like *b* in Fig. 3B, activates its own transcription. In non-competent cells, MecA binds to ComK and inhibits its activity, an inhibition that is further induced by formation of a ternary complex with ClpC. Thus, MecA and ClpC together play a role similar to that of A in Fig. 3B in the building of the competence switch (30).

The functioning principle of the oscillator of Fig. 5, where the complexation of the CC dimer with the heterodimer AB prevents activation of gene *a* by the C multimer, is quite reminiscent of that used by circadian oscillator gene networks. Using *Drosophila* network terminology as an example (31), this negative feedback loop acting on transcriptional activation via complexation can be compared to the transcriptional deactivation of period (*per*) and timeless (*tim*) genes after PER/TIM complexation with the Clock/Cycle (dCLK/CYC) transcriptional activating dimer. The circuit differs from that of any documented circadian network, but the analogy is close enough to suggest alternative models of circadian networks (P.F., unpublished data) which, contrary to some existing ones, do not make crucial use of high Hill coefficients (32) or delays coming from chains of phosphorylation (33).

The evolutionary algorithm described here is able to create a variety of small networks with prescribed behaviors that display both known and original designs. The selected networks make crucial use of posttranscriptional interactions and kinetics of interactions. Their functions could not be understood at all by focusing only on transcriptional interactions, a lesson of general value and a warning for future bioinformatics analyses.

Experimental procedures have recently been developed to reach goals related to that of the present work. In ref. 6, combinatorial synthesis was used to generate small networks of transcription encoding genes with various topologies. Subsequent screening allowed the extraction of networks performing a variety of simple tasks. As in the present study, no *a priori* design was imposed, and the same task was found to be realized by networks with different topologies. Ref. 7 used successive rounds of directed evolution to adjust kinetic rates in a rationally designed network and obtain a functional circuit. The evolved mutants present changes both in protein-DNA and protein-protein interactions. These studies appear quite complementary to the present one. On the one hand, the *in silico* route is certainly more flexible and less labor-intensive than the experimental ones. In the present investigation, we have, for instance, included without difficulty a much greater variety of interactions than in ref. 6 and evolutionary steps well beyond those accessible to ref. 7. *In silico* results are also easier to analyze than experimental ones, because a complete description of the cre-

Chapitre III

Une étude d'un module génétique prototype

Comme nous l'avons vu précédemment, le motif spécifique composé de deux protéines interagissant à la fois au niveau transcriptionnel et post-transcriptionnel semble constituer le motif de base des horloges circadiennes. Ce petit module, appelé "Mixed Feedback loop" semble également statistiquement surreprésenté chez la levure [90].

De plus, ce module a été obtenu comme structure fondamentale de plusieurs systèmes conçus par l'algorithme d'évolution *in silico*, présenté au chapitre II [26]. Nous proposons donc dans cette partie une étude des fonctions potentielles pouvant être réalisées par ce module. Dans un premier temps, nous proposons un modèle mathématique de ce module. Puis, à partir de ce modèle mathématique, nous mettons en évidence que ce module peut constituer soit un commutateur bistable, soit un oscillateur. Nous présentons ensuite les différentes propriétés de ces régimes.

A Caractéristiques du modèle

A.1 Équations de base

Un schéma du module étudié est représenté Figure III.1. La MFL se compose de deux protéines A et B, et des gènes associés g_a et g_b . A régule la transcription du gène g_b et interagit directement avec la protéine B. Notre but est d'étudier la dynamique de ce module dans les cas les plus simples : nous ne distinguerons donc pas les concentrations dans différents compartiments cellulaires et nous modélisons toutes les réactions par des lois d'action de masse. Nous prenons un volume cellulaire unitaire, si bien que la concentration d'une protéine A, notée $[A]$ représente de manière effective le nombre de protéines A présentes dans la cellule.

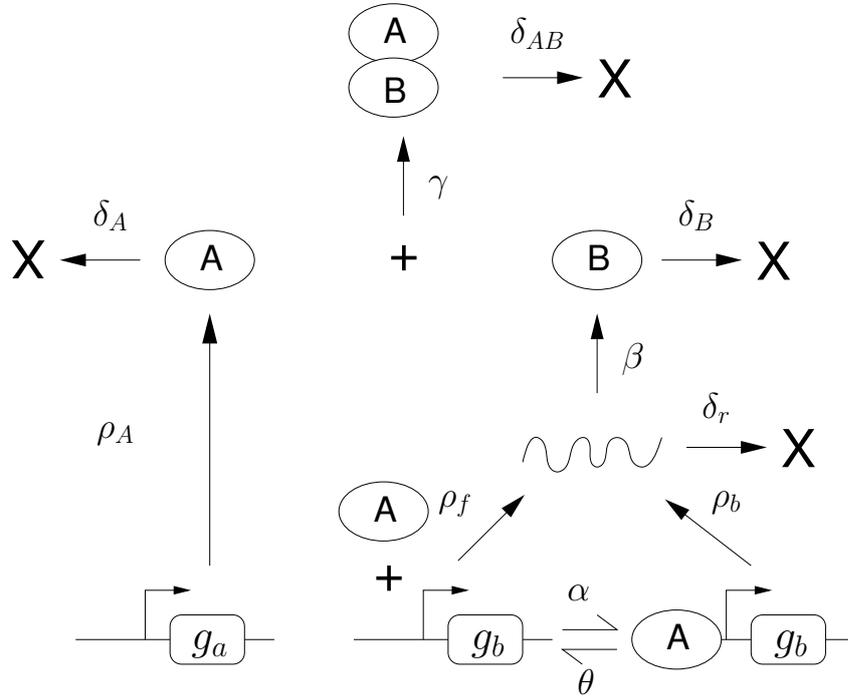


FIG. III.1: Représentation schématique du module

Les deux premières équations modélisent la régulation de la transcription du gène g_b par la protéine A :

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

$$\frac{d[r_b]}{dt} = \rho_f[g_b] + \rho_b[g_b : A] - \delta_r[r_b] \quad (\text{III.2})$$

Il est ici supposé que le gène g_b existe sous deux formes : une forme avec A lié à son promoteur, notée $[g_b : A]$, une forme sans A lié au promoteur, notée $[g_b]$. On suppose que $[g_b : A] + [g_b] = 1$, si bien que seule l'équation III.1 est nécessaire pour décrire les transitions entre les deux formes.

Les protéines A peuvent se lier au promoteur du gène b avec un taux α , et une fois liées elles peuvent être relâchées avec un taux θ . La régulation du taux de transcription du gène g_b par la protéine A est donnée par l'équation III.2, où r_b représente les transcrits de g_b . Lorsque A est lié au promoteur de g_b , la transcription a lieu avec un taux ρ_b , lorsque le promoteur est libre, avec un taux ρ_f . Par conséquent, si $\rho_b > \rho_f$, il y a activation transcriptionnelle tandis que le cas $\rho_f > \rho_b$ correspond à une répression transcriptionnelle. Nous supposons ensuite que les transcrits r_b se dégradent avec une constante de dégradation du premier ordre δ_r . Cette description

correspond au cas où il y a une seule copie du gène par cellule ; il est possible de se ramener au cas où il y a g_0 copies (par exemple $g_0 = 2$) en multipliant les taux ρ_b et ρ_f par g_0 .

La production des protéines A et B et leur interaction sont décrites dans les trois équations différentielles suivantes :

$$\begin{aligned} \frac{d[A]}{dt} = & \rho_A - \gamma[B].[A] - \delta_A[A] \\ & + \theta[g_b : A] - \alpha[g_b][A] \end{aligned} \quad (\text{III.3})$$

$$\frac{d[B]}{dt} = \beta[r_b] - \gamma[B].[A] - \delta_B[B] \quad (\text{III.4})$$

$$\frac{d[AB]}{dt} = \gamma[B].[A] - \delta_{AB}[AB] \quad (\text{III.5})$$

Dans ces équations, $[A]$, $[B]$, et $[AB]$ représentent respectivement les concentrations des protéines A, B et du dimère AB. Comme la régulation du gène g_a n'est pas considérée dans ce modèle simple, il n'est pas nécessaire de distinguer les étapes de transcription et de traduction pour la protéine A et on suppose juste que la protéine A est produite avec un taux de base ρ_A dans l'équation III.3. La seconde interaction cruciale dans ce module, l'interaction des protéines A et B est prise en compte en supposant que A et B dimérisent avec un taux γ . Il est également supposé que B ne peut pas interagir avec une protéine A déjà liée au promoteur du gène g_b . Nous supposons aussi que A se dégrade avec un taux de dégradation du premier ordre δ_A . Enfin, les deux derniers termes de l'équation III.3 prennent en compte la très petite diminution de la concentration de A due à l'attachement sur le promoteur du gène g_b .

L'équation III.4 décrit la traduction des transcrits $[r_b]$ avec un taux β . Les protéines B sont ensuite dégradées avec un taux δ_B et s'associent également aux protéines A.

Enfin l'équation III.5 décrit la dynamique du complexe formé par A et B. Nous supposons ici que ce dimère AB n'a aucune influence sur le reste du système. En particulier il ne peut ni activer la transcription, ni se dissocier en protéines A ou B¹. Nous n'étudierons donc pas cette équation dans la suite.

¹On peut tenir compte d'une dissociation du complexe. Cela revient au final à modifier la constante γ , et ne change donc pas les résultats présentés ci-après.

A.2 Ordre de grandeur des constantes cinétiques et identification des petits paramètres

Dans ce premier modèle simple, il y a déjà dix constantes cinétiques. Dans la suite nous discutons les échelles de temps typiques de chacun des processus impliqués en toute généralité et identifions les petits paramètres.

Les durées de vie des ARN messagers vont de plusieurs heures² à quelques minutes. Chez la levure, la durée de vie typique est de l'ordre de 20 minutes, on prendra donc $\delta_r = 0.03 \text{min}^{-1}$. Pour l'activation transcriptionnelle, le rapport $\theta/\alpha = [A]_0$ donne une idée de la concentration typique de protéines pour laquelle la transcription est activée. Une valeur typique serait d'environ 40 protéines par cellule [50]. θ donne une idée du temps caractéristique pendant lequel une protéine reste liée à un promoteur. Nous supposons que ce temps est de l'ordre de plusieurs minutes. Enfin, nous supposons que les taux de transcription typiques sont de l'ordre de quelques ARN messagers par minutes. Nous prendrons donc dans la suite $\alpha = \theta/40$, θ de l'ordre de δ_r , $\rho_f \simeq \text{min}^{-1}$, $\rho_b \simeq \text{min}^{-1}$.

Les durées de vie des protéines varient de quelques minutes à plusieurs heures. L'heure apparaît comme une durée de vie plausible. Nous prendrons donc dans la suite $\delta_A = \delta_B = 0.01 \text{min}^{-1}$, et plus généralement considérerons que les constantes de dégradation des protéines sont du même ordre que celles des ARN messagers. Pour les taux de traduction, $\beta = 3$ protéines par ARN messenger par minute est plausible. Si l'on combine ce taux avec les ordres de grandeur donnés précédemment, on trouve $\rho_A \simeq 100 \text{min}^{-1}$.

Nous considérerons finalement que les interactions entre protéines sont essentiellement limitées par la diffusion. Une constante de diffusion $D \simeq 2.5 \mu\text{m}^2 \cdot \text{s}^{-1}$ [20] donne un temps caractéristique de l'ordre de la minute pour une cellule de taille typique de $10 \mu\text{m}$. Nous prendrons donc $\gamma \simeq 1 \text{min}^{-1}$.

Ensuite, nous adimensionnons les équations. Nous normalisons la concentration des transcrits de g_b pour avoir un taux de production des protéines B égal à celui des protéines A, en définissant la concentration sans dimension $r = \frac{\beta[r_b]}{\rho_A}$. Nous normalisons ensuite la concentration des protéines par la concentration d'équilibre résultant de la production des protéines à un taux ρ_A et de la dimérisation à un taux γ , et définissons donc $A = \sqrt{\frac{\gamma}{\rho_A}}[A]$, $B = \sqrt{\frac{\gamma}{\rho_A}}[B]$. Nous redimensionnons

²Signalons ici que c'est en particulier le cas pour les horloges circadiennes. À titre d'exemple, il a été mesuré que la constante de dégradation des transcrits *frq* pour *Neurospora* était de l'ordre de 0.2h^{-1} , soit presque un ordre de grandeur au-dessous des constantes de dégradation typiques des ARN messagers. Ceci explique en partie la période très longue des horloges circadiennes comparée à celles des oscillateurs présentés ci-après.

également l'unité de temps et prenons comme temps caractéristique l'inverse du taux de dégradation des transcrits r_b , $1/\delta_r$. Les équations III.1 à III.4 se réécrivent de la façon suivante :

$$\frac{dg}{dt} = \tilde{\theta}(1 - g) - \tilde{\alpha}gA \quad (\text{III.6})$$

$$\frac{dr}{dt} = \rho_0g + \rho_1(1 - g) - r \quad (\text{III.7})$$

$$\delta \frac{dA}{dt} = 1 - AB + \mu\delta(\tilde{\theta}(1 - g) - \tilde{\alpha}gA) - \delta d_a A \quad (\text{III.8})$$

$$\delta \frac{dB}{dt} = r - AB - \delta d_b B \quad (\text{III.9})$$

où nous avons défini les paramètres suivants $\delta = \delta_r/\sqrt{\rho_A\gamma}$, $\rho_1 = \beta\rho_b/(\rho_A\delta_r)$, $\rho_0 = \beta\rho_f/(\rho_A\delta_r)$, $\tilde{\theta} = \theta/\delta_r$, $\tilde{\alpha} = \alpha\sqrt{\rho_A}/(\delta_r\sqrt{\gamma})$, $\mu = \sqrt{\gamma}/\rho_A$, $d_a = \delta_A/\delta_r$ et $d_b = \delta_B/\delta_r$. A ce stade, il reste néanmoins sept paramètres. Cependant, on peut voir dans cette expression que δ est un petit paramètre (de l'ordre de 3×10^{-3} avec les constantes données précédemment) comparé à tous les autres paramètres qui sont d'ordre 1. Cette propriété simplifie considérablement l'analyse comme nous le verrons dans la suite, et permet de réduire encore le nombre de paramètres pertinents dans ce système.

B Aperçu de la dynamique dans les différents régimes de paramètres

Le comportement dynamique du module dépend des deux taux de transcriptions du gène g_b (avec A lié ou non au promoteur) et plus précisément des taux de productions de la protéine B $\beta\rho_b/\delta_r$ et $\beta\rho_f/\delta_r$ comparés au taux de production de la protéine A. Il est donc commode d'étudier les différents comportements induits par les variations des deux taux précédemment définis $\rho_1 = \beta\rho_b/(\rho_A\delta_r)$ et $\rho_0 = \beta\rho_f/(\rho_A\delta_r)$.

Trois comportements qualitatifs différents peuvent être observés.

Le cas le plus simple apparaît lorsque le taux de production de A est toujours plus grand ou plus petit que le taux de production de B, que le promoteur de g_b soit lié ou non. Ceci signifie que $\beta\rho_b/\delta_r$ et $\beta\rho_f/\delta_r$ sont tous deux à la fois plus grands ou plus petits que ρ_A , ou encore que ρ_0 et ρ_1 sont tous deux plus grands ou plus petits que 1. Dans ce cas, le système a un seul état stable vers lequel il relaxe quelles que soient les conditions initiales.

Les cas plus intéressants sont les régimes de paramètres où le taux de production de A est un taux intermédiaire entre les taux de production de la protéine B, c'est-à-dire les cas où $\rho_0 > 1 > \rho_1$ ou $\rho_1 > 1 > \rho_0$.

Discutons tout d'abord le cas où A est un répresseur transcriptionnel, c'est-à-dire les cas où $\rho_0 > 1 > \rho_1$, qui est aussi le cas présenté dans la figure II.6. Si A n'est pas lié au promoteur, le taux de production de B est plus grand que celui de A, si bien que les protéines A sont immédiatement titrées par les protéines B. Ceci empêche les protéines A de venir réprimer l'expression du gène g_b et on atteint donc un premier état stable avec peu de protéines A et beaucoup de protéines B donné par :

$$[B]_1 \simeq \frac{(\rho_0 - 1)\rho_A}{\delta_B}, [A]_1 \simeq \frac{\rho_A}{\gamma[B]_1} \quad (\text{III.10})$$

Au contraire, s'il y a beaucoup de protéines A présentes dans le système, A va réprimer l'expression du gène g_b . Dans ce cas, le taux de production des protéines A est plus grand que le taux de production des protéines B. Les protéines B sont donc rapidement titrées, mais comme il y a plus de protéines A produites que de protéines B, il reste suffisamment de protéines A pour réprimer la transcription. Le système atteint alors un second état stable donné par les équations suivantes :

$$[A]_2 \simeq \frac{(1 - \rho_1)\rho_A}{\delta_A}, [B]_2 \simeq \frac{\rho_A\rho_1}{\gamma[A]_2} \quad (\text{III.11})$$

Le système est bistable dans le régime de paramètre où $\rho_0 > 1 > \rho_1$ uniquement dans la limite δ nul. Dans la limite δ très petit les frontières du domaine bistable peuvent être calculées explicitement (voir suite) et sont données par :

$$\rho_0 > 1 + 2\sqrt{\delta d_b(1 - \rho_1)/A_0}, \rho_1 \leq 1 \quad (\text{III.12})$$

$$\rho_1 < 1 - 2\sqrt{d_a A_0 \delta (\rho_0 - 1)}, \rho_0 \geq 1 \quad (\text{III.13})$$

avec $A_0 = \tilde{\theta}/\tilde{\alpha}$. Dans ce domaine, le système relaxe vers l'un ou l'autre de ces points fixes en fonction des conditions initiales. L'existence de ce domaine bistable se traduit par un phénomène d'hystérésis (voir Figure III.2) : en variant lentement le taux de production de la protéine A, le système peut sauter brutalement d'un état stable à l'autre.

Lorsque A joue le rôle d'activateur transcriptionnel, la complexation entre B et A joue le rôle de boucle de rétroaction négative. Le système se comporte alors comme un oscillateur dans la région où $\rho_1 > 1 > \rho_0$, voir par exemple Figure III.3.

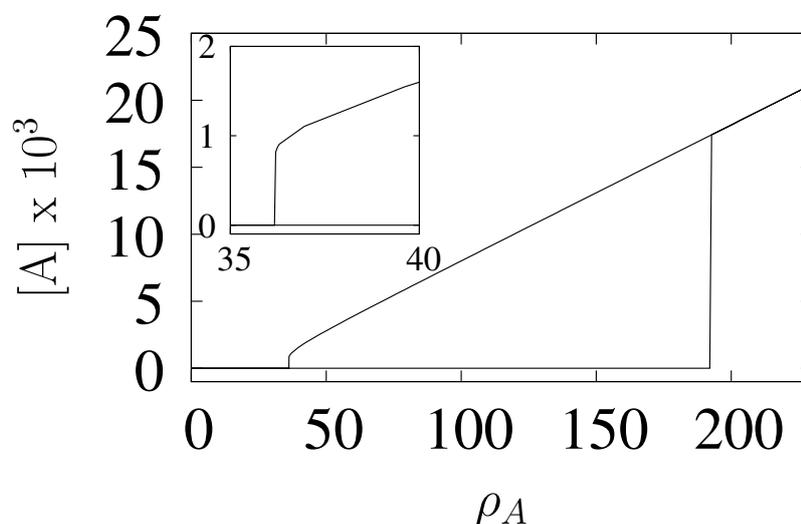


FIG. III.2: Hystérésis pour la MFL. Les valeurs des concentrations d'équilibre sont tracées en fonction de ρ_A . Les autres paramètres sont $\rho_f = 0.2 \text{ mol.min}^{-1}$, $\rho_b = 2 \text{ mol.min}^{-1}$, $\beta = 3 \text{ min}^{-1}$, $\theta = 0.04 \text{ min}^{-1}$, $\alpha = 0.001 \text{ mol}^{-1}.\text{min}^{-1}$, $\delta_A = 0.01 \text{ min}^{-1}$, $\delta_r = 0.03 \text{ min}^{-1}$, $\delta_B = 0.01 \text{ min}^{-1}$, $\gamma = 1 \text{ mol}^{-1}.\text{min}^{-1}$ où *mol* signifie molécules et *min* minutes.

Le comportement oscillant dépend essentiellement des paramètres ρ_0 et ρ_1 , et sont conservés pour des cinétiques très différentes de réaction entre les protéines et l'ADN, représentées par le paramètre $\tilde{\theta}$. Cependant, la taille du domaine d'oscillation dépend de ce paramètre, comme on peut le voir Figure III.4 ; en particulier, les oscillations disparaissent lorsque $\tilde{\theta}$ devient de l'ordre de $1/\delta$.

La figure III.4 montre la coexistence d'existence de ces deux domaines (oscillant et bistable) pour un jeu de paramètre où le taux de production de A est suffisamment grand. Dans toute la suite, nous détaillons l'étude analytique de ces différentes phases.

C Régime bistable

Dans un premier temps, nous posons $A_0 = \tilde{\theta}/\tilde{\alpha}$. A_0 représente la concentration adimensionnée critique à partir de laquelle la protéine A va se lier au promoteur du gène g_b . Nous calculons ensuite les concentrations à l'équilibre du gène, des transcrits et de la protéine B en fonction de la concentration de A :

$$g = \frac{A_0}{A + A_0} \quad (\text{III.14})$$

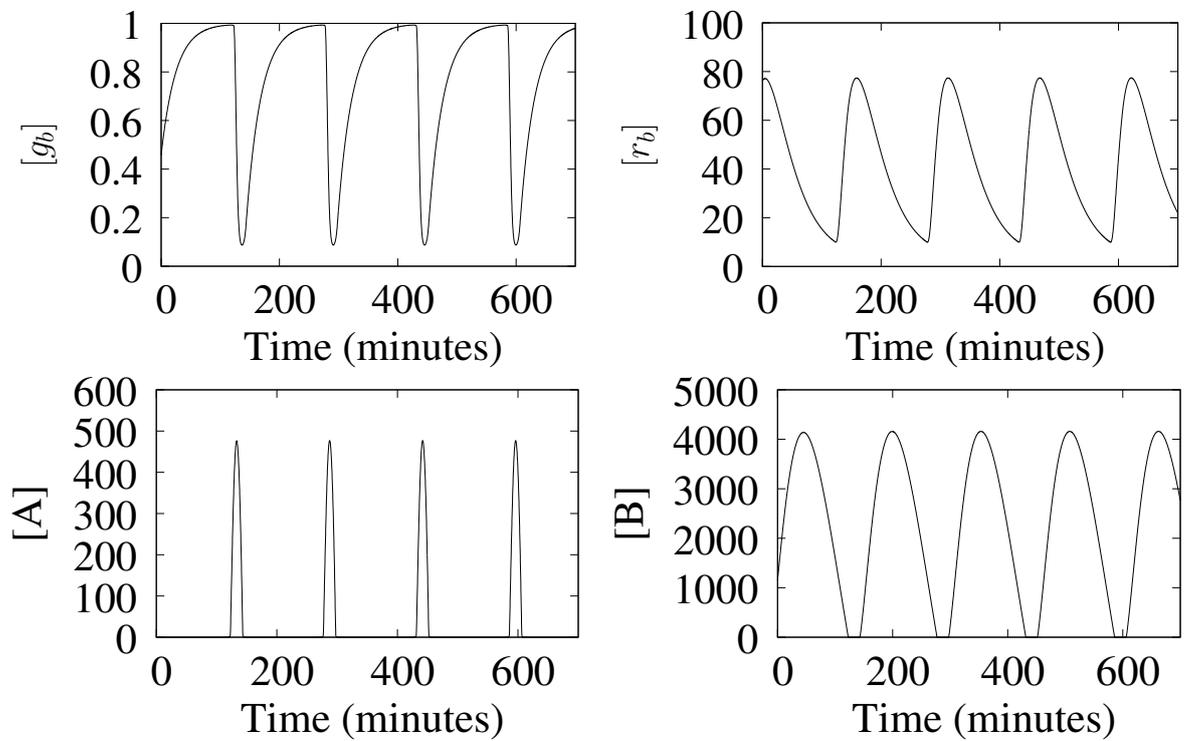


FIG. III.3: Oscillations pour la MFL. Les oscillations pour les différentes espèces sont représentées en fonction du temps. Les constantes sont les mêmes que dans la figure III.2 avec $\rho_A = 100 \text{ mol.min}^{-1}$, $\rho_f = 0.1 \text{ mol.min}^{-1}$, $\rho_b = 5 \text{ mol.min}^{-1}$.

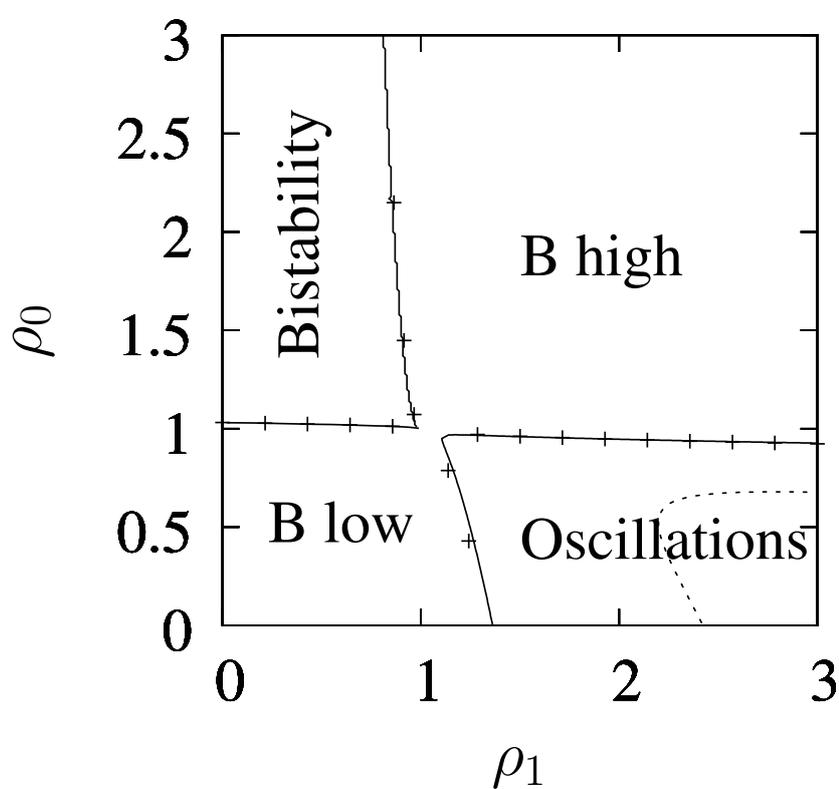


FIG. III.4: Diagramme de phase pour la MFL. Les paramètres sont $\delta = 0.003$, $d_a = d_b = 0.33$, $\tilde{\alpha} = 0.33$, $\mu = 0.31$, $\tilde{\theta} = 1.33$ + : détermination numérique de la frontière entre les différents domaines. Ligne pleine : approximation analytique des frontières (Eqs. III.27,III.29,III.34, III.39) . Ligne pointillée : frontière de stabilité du point fixe pour $\tilde{\theta}$ et $\tilde{\alpha}$ vingt fois plus grands.

$$r = \frac{\rho_1 A + \rho_0 A_0}{A + A_0} \quad (\text{III.15})$$

$$B = \frac{\rho_1 A + \rho_0 A_0}{(A + A_0)(A + \delta d_b)} \quad (\text{III.16})$$

ce qui donne l'équation suivante pour A :

$$1 = \delta d_a A + \frac{(\rho_1 A + \rho_0 A_0) A}{(A + A_0)(A + \delta d_b)} \quad (\text{III.17})$$

Dans la limite $\delta = 0$, on voit immédiatement qu'il n'y a qu'une seule solution à cette dernière équation, donnée par $A_2 = A_0 \frac{\rho_0 - 1}{1 - \rho_1}$. Cette solution n'existe donc que dans les régimes de paramètres où $\rho_0 < 1 < \rho_1$ et $\rho_1 < 1 < \rho_0$. Pour les petites valeurs de δ , deux autres points fixes apparaissent, un petit point fixe d'ordre δ , $A_1 \simeq \frac{\delta d_b}{\rho_0 - 1}$, et un grand point fixe d'ordre $1/\delta$, $A_3 \simeq \frac{1 - \rho_1}{\delta d_a}$. A_1 est un point fixe possible uniquement lorsque $\rho_0 > 1$ tandis que A_3 est un point fixe possible uniquement lorsque $\rho_1 < 1$.

L'équation III.17 ne peut donc avoir plus d'une solution uniquement dans le régime de paramètres où A est un répresseur de la transcription avec $\rho_1 < 1 < \rho_0$. Nous nous plaçons dans ce régime de paramètres dans toute cette sous-partie.

C.1 État stable avec A haut

Montrons que le point fixe noté A_3 est stable dans la limite des petits δ . Les équations III.8 et III.9 suggèrent que dans cette limite, il est possible de faire des approximations quasi-statiques car l'une ou l'autre des protéines doit atteindre un équilibre adiabatique après un temps d'ordre δ . Cependant, les deux protéines ne peuvent en général pas atteindre cet équilibre quasi-statique simultanément : par exemple, si $A \gg B$, seul B se trouve au quasi-équilibre tandis que A est alors d'ordre $1/\delta$. Lorsque la concentration de A est élevée, dans la limite des petits δ , nous posons $a = \delta A$, en supposant a d'ordre 1. D'après les équations III.6 et III.9, g et B atteignent alors des valeurs de quasi-équilibre après un temps d'ordre δ données par :

$$g = \tilde{\theta}/(\tilde{\theta} + \tilde{\alpha}/\delta a) \simeq \delta \tilde{\theta}/\tilde{\alpha} \cdot a \quad (\text{III.18})$$

$$B = r/(a/\delta) \simeq \delta r/a \quad (\text{III.19})$$

Dans ce cas, les équations de la dynamique de la MFL se réduisent aux deux équations suivantes dans la limite $\delta \rightarrow 0$:

$$\frac{dr}{dt} = \rho_1 - r \quad (\text{III.20})$$

$$\frac{da}{dt} = 1 - r - d_a a \quad (\text{III.21})$$

Ces deux équations montrent que dans ce cas, r et a convergent tous les deux vers des valeurs d'ordre 1, si bien que le point fixe est stable et est donné, à l'ordre le plus bas, par $r \simeq \rho_1$, $B \simeq g \simeq 0$, $a = (1 - \rho_1)/d_a$. Cet état stable n'est physique que dans le cas où $\rho_1 < 1$, ce qui signifie que le taux de production des protéines B lorsque A réprime l'expression du gène g_b n'est pas assez grand pour séquestrer les protéines A. Il reste donc toujours des protéines A dans le système et la répression ne peut donc être levée.

C.2 État stable avec B haut

De la même façon, dans la limite des petit δ , nous considérons cette fois la limite où beaucoup de B sont présents dans le système. Nous posons donc $b = \delta B$. Dans ce cas, A atteint rapidement un équilibre adiabatique donné par :

$$A \simeq \delta/b \quad (\text{III.22})$$

et les équations de la dynamique de la MFL se réduisent à :

$$\frac{dg}{dt} = \tilde{\theta}(1 - g) \quad (\text{III.23})$$

$$\frac{dr}{dt} = \rho_0 + (1 - \rho_1)(1 - g) - r \quad (\text{III.24})$$

$$\frac{db}{dt} = r - 1 - d_b b \quad (\text{III.25})$$

Les concentrations convergent alors aux temps longs vers le point fixe défini par $g \simeq 1$, $r \simeq \rho_0$, $b \simeq (\rho_0 - 1)/d_b$, $A \simeq 0$ à l'ordre le plus bas. Ce point fixe existe uniquement si $\rho_0 > 1$, ce qui signifie biologiquement que le taux de production des protéines B est suffisamment grand pour titrer toutes les protéines A, empêchant alors la répression de la transcription du gène g_b .

C.3 Bifurcation entre les régimes monostables et bistables

Dans cette section, nous analysons plus précisément la nature et la position des bifurcations entre les deux régimes, dans le domaine où $\rho_0 > 1 > \rho_1$ avec $\delta \ll 1$.

Considérons tout d'abord la frontière entre le régime bistable et le régime monostable avec une concentration de A élevée (partie en bas à gauche de la figure III.4). D'après l'expression des points fixes, A_1 et A_2 vont être du même ordre si $\rho_0 - 1 = O(\sqrt{\delta})$, et les deux concentrations sont alors d'ordre $\sqrt{\delta}$. Dans cette région, ces deux points fixes vont donc coalescer, puis disparaître le long d'une ligne dont nous déterminons maintenant l'équation. Supposons que $A \sim \sqrt{\delta}$, l'équation III.17 devient alors à l'ordre dominant :

$$(\rho_0 - 1) = (1 - \rho_1) \frac{A}{A_0} + \frac{\delta d_b}{A} \quad (\text{III.26})$$

ρ_1 étant fixé, le membre de droite atteint un minimum pour $A = \sqrt{\delta d_b A_0 / (1 - \rho_1)}$, qui est bien d'ordre $\sqrt{\delta}$, ce qui nous donne l'équation suivante pour ρ_0 :

$$\rho_0 = 1 + 2\sqrt{\frac{\delta d_b (1 - \rho_1)}{A_0}} \quad (\text{III.27})$$

En dessous de cette valeur de ρ_0 , l'équation III.26 n'a pas de solutions, et il n'y a donc qu'un seul point fixe, A_3 , dont la concentration est d'ordre $1/\delta$; au-dessus de cette valeur, l'équation a bien deux solutions, et il y a les trois points fixes. La bifurcation est donc une bifurcation noeud-col, dont la ligne de séparation est donnée par l'équation III.27. Remarquons que pour $\delta \rightarrow 0$, nous retrouvons la limite d'ordre 0, $\rho_0 = 1$.

Une analyse similaire peut être faite pour la transition entre le régime bistable et le régime monostable à faible concentration de A. De la même façon, A_2 et A_3 peuvent coalescer si $\rho_1 - 1 = O(\sqrt{\delta})$, et dans ce cas la concentration commune du point fixe A est d'ordre $1/\sqrt{\delta}$. Supposons que $A \sim 1/\sqrt{\delta}$, l'équation III.17 devient alors à l'ordre dominant :

$$(1 - \rho_1) = (\rho_0 - 1) \frac{A_0}{A} + \delta d_a A \quad (\text{III.28})$$

ρ_0 étant fixé, le membre de droite atteint un minimum pour $A = \sqrt{(\rho_0 - 1)A_0/\delta d_a}$, qui est bien d'ordre $1/\sqrt{\delta}$, ce qui nous donne l'équation suivante pour ρ_1 :

$$\rho_1 = 1 - 2\sqrt{d_a A_0 \delta (\rho_0 - 1)} \quad (\text{III.29})$$

Au-dessus de cette valeur de ρ_1 , l'équation III.28 n'a pas de solutions, et il n'y a donc qu'un seul point fixe, A_1 , dont la concentration est d'ordre δ ; au-dessus de cette valeur, l'équation a bien deux solutions, et il y a les trois points fixes. La bifurcation est donc une bifurcation noeud-col, dont la ligne de séparation est donnée par l'équation III.29. Remarquons que pour $\delta \rightarrow 0$, nous retrouvons la limite d'ordre 0, $\rho_1 = 1$.

Les approximations données dans les expressions III.27 et III.29 sont en accord avec les déterminations numériques de ces lignes comme on peut le voir Figure III.4.

D Régime oscillant

Dans cette partie, nous considérons maintenant que A est un activateur transcriptionnel, c'est-à-dire que $\rho_0 < \rho_1$. Dans ce cas, le membre de droite de l'équation III.17 est une fonction croissante de A, donc il y a un seul point fixe. Nous allons maintenant montrer que ce point fixe est instable dans un large domaine de paramètres, et que par conséquent, la MFL se comporte comme un oscillateur.

D.1 Analyse de stabilité linéaire

Nous analysons dans la suite la stabilité linéaire du point fixe (g^*, r^*, A^*, B^*) où A^* est l'unique solution de l'équation III.17 et où les valeurs de g^*, r^* , et B^* sont des fonctions de A^* données par les équations III.14 à III.16.

Après linéarisation de la dynamique de la MFL, le taux de croissance complexe X telle qu'une perturbation croît exponentiellement dans le temps comme $\exp(Xt)$ est donné par le polynôme caractéristique suivant :

$$\begin{aligned} (X + \tilde{\theta}(1 + \frac{A^*}{A_0}))(X + 1)((\delta(X + d_a) + B^*)(\delta(X + d_b) + A^*) - A^*B^*) \\ = \frac{A^*g^*}{A_0}(\rho_0 - \rho_1)\tilde{\theta} - \mu g^* \delta \tilde{\theta} / A_0 X(X + 1)(\delta(X + d_b) + A^*) \end{aligned} \quad (\text{III.30})$$

Dans la limite des petits δ , le calcul des racines de ce polynôme est simplifié. Comme on l'a dit, il y a un seul point fixe, appelé A_2 dans la section précédente dans le domaine de paramètres où $\rho_0 < 1 < \rho_1$. Nous avons donc, à l'ordre dominant $A^* \simeq \frac{\rho_0 - 1}{1 - \rho_1} A_0$, $B^* \simeq \frac{1 - \rho_1}{(\rho_0 - 1)A_0}$, $r^* \simeq 1$ et $g^* \simeq \frac{1 - \rho_1}{\rho_0 - \rho_1}$.

Dans le domaine où $\rho_0 < 1 < \rho_1$, les coordonnées du point fixe sont a priori d'ordre 1 comparé à δ . Supposons maintenant que X diverge dans la limite où δ tend vers 0. L'équation III.30 se réduit alors à l'ordre dominant à

$$X^3(\delta X + (A^* + B^*)) = \frac{A^* \tilde{\theta} g^*}{A_0 \delta} (\rho_0 - \rho_1) \quad (\text{III.31})$$

Cette équation a trois racines X_k , avec $k = 0, 1, 2$, d'ordre $\delta^{-1/3}$, proportionnelles aux racines cubiques de l'unité j^k :

$$X_k = -j^k \left[\frac{A^*}{A^* + B^*} \frac{\tilde{\theta} g^*}{A_0 \delta} (\rho_1 - \rho_0) \right]^{1/3}, \quad k = 0, 1, 2 \quad (\text{III.32})$$

La quatrième racine correspond à un mode stable de partie réelle $-(A^* + B^*)/\delta$. Les deux racines X_1 et X_2 sont complexes conjuguées et ont des parties réelles positives. Dans le domaine de paramètre où $\rho_0 < 1 < \rho_1$, à l'ordre le plus grand, le point fixe est donc instable et une petite perturbation croît exponentiellement en oscillant autour du point fixe. La dynamique tend alors vers un cycle limite stable qui sera analysé dans la partie suivante.

Lorsque ρ_1 ou ρ_0 tendent vers 1, A^* ou B^* explosent et l'instabilité du point fixe disparaît via une bifurcation de Hopf. Nous calculons dans la suite les équations de ces deux lignes de bifurcation.

D.1.a Bifurcation dans la région où A^* est grand

Si $\rho_1 \rightarrow 1$, A^* devient très grand et les parties réelles des valeurs propres deviennent d'ordre 1, il faut donc modifier notre approche. Nous négligeons alors $\delta(X + d_b)$ et B^* devant A^* et l'équation III.30 se réécrit, à l'ordre dominant :

$$(X + \tilde{\theta}(\frac{\rho_0 - \rho_1}{1 - \rho_1}))(X + 1)(X + d_a) = \frac{\tilde{\theta}(1 - \rho_1)}{\delta A_0} \quad (\text{III.33})$$

en remplaçant A^* et g^* par leurs expressions respectives. Lorsque $\frac{\tilde{\theta}(1 - \rho_1)}{\delta A_0}$ est la constante dominante, nous pouvons refaire le raisonnement précédent et nous voyons que nous avons de nouveaux trois valeurs propres proportionnelles aux racines de l'unité, dont deux ont des valeurs propres positives. Au contraire, lorsque cette constante est négligeable pour ρ_1 très proche de 1, il y a clairement trois racines de parties réelles négatives. À ce stade, on peut soit calculer les racines de cette équation perturbativement et annuler explicitement les parties réelles des racines complexes conjuguées, soit écrire directement les conditions de Routh-Hurwitz³. On

³les racines du polynôme $X^3 + b_1 X^2 + b_2 X + b_3$ ont des parties réelles négatives si et seulement si $b_1 b_2 > b_3$.

trouve alors à l'ordre dominant l'équation de la frontière du domaine où le point fixe est instable :

$$\rho_1 = 1 + \left(\delta \tilde{\theta} A_0 (d_a + 1) (\rho_1 - \rho_0)^2 \right)^{1/3} \quad (\text{III.34})$$

$\rho_1 - 1$ est donc d'ordre $\delta^{1/3}$ ce qui justifie a posteriori l'utilisation de l'équation approchée de A_2 dans ce calcul. Notons que dans la limite $\delta \rightarrow 0$, on retrouve l'équation de la frontière $\rho_1 = 1$.

D.1.b Bifurcation dans la région où A^* est petit

De la même façon, si $\rho_0 \rightarrow 1$, B^* devient d'ordre $1/(\rho_0 - 1)$ et on peut maintenant négliger $\delta(X + d_a)$ et A^* devant B^* et A^* .

L'équation III.30 se simplifie donc et on trouve à l'ordre dominant :

$$(X + \tilde{\theta})(X + 1)(X + d_b) = -E \quad (\text{III.35})$$

avec

$$E = \frac{A^* \tilde{\theta}}{A_0 \delta B^*} (\rho_1 - 1) \quad (\text{III.36})$$

De la même façon, quand E est très petit, les racines de l'équation III.35 ont toutes des parties réelles négatives, tandis que lorsque E est très grand, les racines sont proportionnelles aux trois racines de l'unité. Il y a donc une valeur critique de E , E_c telle que les parties réelles des deux racines complexes s'annulent. Cette valeur critique est directement donnée par les conditions de Routh-Hurwitz :

$$E_c = (d_b + 1 + \tilde{\theta})(\tilde{\theta} + d_b + d_b \tilde{\theta}) - d_b \tilde{\theta} \quad (\text{III.37})$$

La valeur de A^* associée se déduit de l'équation III.36 :

$$A^* = \sqrt{\frac{E_c A_0 \delta}{\tilde{\theta} (\rho_1 - 1)}} \quad (\text{III.38})$$

Si on réinjecte cette expression dans l'équation III.17, on trouve directement l'équation de la seconde frontière où le point fixe devient stable :

$$\rho_0 = 1 - (1 - d_b \tilde{\theta} / E_c) \sqrt{\frac{E_c \delta (\rho_1 - 1)}{A_0 \tilde{\theta}}} \quad (\text{III.39})$$

Notons que dans la limite $\delta \rightarrow 0$, on retrouve l'équation de la frontière $\rho_1 = 1$. La comparaison entre ces frontières calculées analytiquement et un calcul numérique de la zone de stabilité est représenté Figure III.4.

D.1.c Régime de variation de $\tilde{\theta}$

Dans la limite des petits δ , il est aussi possible de calculer une borne supérieure pour la constante $\tilde{\theta}$, qui est proportionnel à l'inverse du temps caractéristique de liaison de la protéine A sur le promoteur. Pour simplifier, nous considérerons dans cette partie que $\mu = 0$. Comme cette fois ρ_0 et ρ_1 sont arbitraires, il n'est plus possible de négliger δX ni devant A^* ni devant B^* dans l'équation III.30. Après simplification, à l'ordre le plus grand en $\tilde{\theta}$ et δ , l'équation III.30 peut se simplifier et se réécrit :

$$\begin{aligned} \delta^2 X^4 + (\delta^2 \tilde{\theta} (1 + \frac{A^*}{A_0}) + \delta(A^* + B^*)) X^3 + \delta \tilde{\theta} (1 + \frac{A^*}{A_0}) (A^* + B^*) X^2 \\ + \delta \tilde{\theta} (1 + \frac{A^*}{A_0}) ((d_a + 1)A^* + (d_b + 1)B^*) X = \frac{A^* g^*}{A_0} (\rho_0 - \rho_1) \tilde{\theta} \end{aligned} \quad (\text{III.40})$$

Les conditions de Routh-Hurwitz donnent directement la valeur de $\tilde{\theta}$ pour laquelle le point fixe devient instable, qui est donnée par la relation :

$$(\delta \tilde{\theta} u + v)w = \delta \tilde{\theta} u^2 v x \quad (\text{III.41})$$

en définissant $u = 1 + \frac{A^*}{A_0} \simeq \frac{\rho_0 - \rho_1}{1 - \rho_1}$, $v = A^* + B^* \simeq \frac{\rho_0 - 1}{1 - \rho_1} A_0 + \frac{1 - \rho_1}{A_0(\rho_0 - 1)}$, $w = \frac{A^* g^*}{A_0} (\rho_1 - \rho_0) \simeq 1 - \rho_0$, $x = (d_a + 1)A^* + (d_b + 1)B^*$. L'équation III.41 peut se réécrire :

$$\delta \tilde{\theta} = \frac{wv}{u^2 xv - uw} = f(u, v, w, x) \quad (\text{III.42})$$

Tous les autres paramètres étant libres, on peut montrer que $\delta \tilde{\theta}$ ne peut pas dépasser une valeur critique. En effet, il est possible avec un seul jeu de paramètres de maximiser la fonction f par rapport à toutes ses variables. Plus précisément, nous remarquons que $0 \leq w \leq 1$, $1 < u$, $2 \leq v \leq x$. De plus, f est une fonction croissante de w si bien que f est maximale pour $w = 1$, i.e. $\rho_0 = 0$. Dans ce régime de paramètres, f est aussi une fonction décroissante de u , si bien que f est aussi maximale pour $u \rightarrow 1$, i.e. pour $\rho_1 \rightarrow \infty$. Enfin, dans cette limite f est maximale lorsque $x - 1/v$ est minimal, ce qui donne le rapport optimum $A_0/(\rho_1 - 1)$. $x \geq v$, donc le minimum est atteint pour $x = v$, lorsque $d_a = d_b = 0$. Donc la fonction est maximale pour $v = 2$, la valeur minimale autorisée de v , qui correspond à $A_0 = \rho_1 - 1$ et la valeur maximale de $\tilde{\theta}$ pour laquelle le point fixe est instable est :

$$\tilde{\theta}_{max} = \frac{2}{3\delta}.$$

Dans le cas général u, v, w, x dépendent des différents paramètres et on peut donc calculer exactement la valeur maximale de $\tilde{\theta}$ en fonction de δ à partir de l'équation III.42. Il apparaît donc que $\tilde{\theta}_{max} \sim A/\delta$, avec $A < 2/3$.

D.2 Étude du cycle limite

Les oscillations de la MFL deviennent très non linéaires loin du seuil de bifurcation (ou lorsque la bifurcation est sous-critique). Comme il n'y a pas de boucles de rétroaction positives dans la MFL, un système réduit de deux équations aurait une divergence négative. Il serait donc impossible de décrire les oscillations avec un tel système à deux variables, obtenu par exemple en se focalisant uniquement sur la dynamique des protéines. Il faut donc utiliser une analyse spécifique qui est développée dans la suite.

Une période d'oscillation de la MFL peut se décomposer en deux phases distinctes : une première phase où $A \gg 1 \gg B$ (phase I), et une seconde phase où $B \gg 1 \gg A$ (phase II, voir Figure III.5)⁴. Nous décrivons explicitement ces deux phases dans la suite pour obtenir la description complète du cycle limite. La période d'oscillation est également calculée (à l'ordre le plus bas en δ). Nous utiliserons ensuite une technique de raccordement asymptotique pour relier les différents régimes et justifier a posteriori les conditions initiales choisies en première analyse.

D.3 Phase I : concentration de A haute, concentration de B basse

Nous définissons cette phase comme la partie du cycle limite où la concentration de la protéine A est plus grande que la concentration $[A]_0$, c'est-à-dire que nous considérons que $A \gg 1 \gg B$. La forme de l'équation III.8 nous amène à supposer que dans cette phase A est d'ordre $1/\delta$ et nous posons donc $a = \delta A$. La concentration des protéines A est suffisante pour s'attacher au promoteur du gène g_b ainsi que pour titrer toutes les protéines B. Dans la limite des petits δ , comme A est d'ordre $1/\delta$, la cinétique de ces deux réactions est alors très rapide comparée à la cinétique d'évolution des transcrits et de la protéine A. Par conséquent, g et B évoluent adiabatiquement, et on peut supposer que :

⁴Ces deux phases pourraient correspondre par exemple aux deux phases décrites expérimentalement pour l'oscillateur circadien de *Neurospora* [59], dont l'analyse est développée dans le chapitre IV

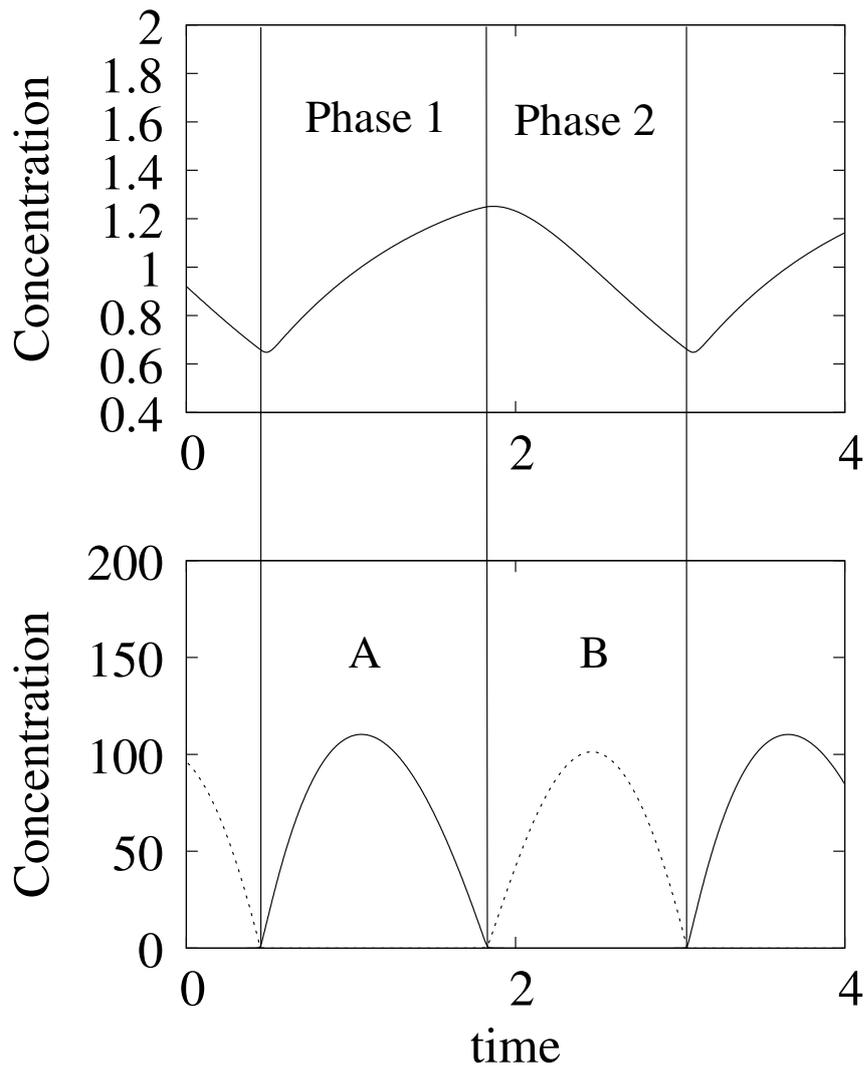


FIG. III.5: Distinction entre les deux phases Haut : oscillation de r . Bas : oscillations de A et de B . Les paramètres ont été ici choisis pour avoir des phases de durée similaire : $\delta = 0.001$, $\rho_1 = 1.5$, $\rho_0 = 0$, $\tilde{\theta} = 2$, $\tilde{\alpha} = 2$, $\mu = 1$, $d_a = d_b = 0$.

$$g = \tilde{\theta}/(\tilde{\theta} + \tilde{\alpha}/\delta a) \simeq \delta\tilde{\theta}/\tilde{\alpha}.a \quad (\text{III.43})$$

$$B \simeq \delta r/a \quad (\text{III.44})$$

Dans cette phase, la dynamique de la MFL se réduit donc, à l'ordre δ , aux deux équations suivantes :

$$\frac{dr}{dt} = \rho_1 - r \quad (\text{III.45})$$

$$\frac{da}{dt} = 1 - r - d_a a \quad (\text{III.46})$$

Le début de la phase I correspond à la fin de la phase II, où, comme nous le verrons, la concentration de protéines A est faible. Par continuité, nous supposons donc que a s'annule au début de la phase I. En appelant r_1 la valeur de r au début de la phase I, l'intégration des équations III.45 et III.46 est immédiate et on a :

$$r_I(t) = \rho_1 + (r_1 - \rho_1)e^{-t} \quad (\text{III.47})$$

$$a_I(t) = \frac{1 - \rho_1}{d_a}(1 - e^{-d_a t}) + \frac{r_1 - \rho_1}{d_a - 1}(e^{-d_a t} - e^{-t}) \quad (\text{III.48})$$

Les équations III.47 et III.48 sont valables uniquement dans la phase I. La phase I s'achève lorsque les approximations quasi-statiques faites précédemment ne sont plus valables, i.e. lorsque la concentration de A chute, après un temps t_1 tel que $a_I(t_1) = 0$. A la fin de cette phase, nous avons alors $r_I(t_1) = r_2$.

Le fait que la concentration de A doive monter pour ensuite redescendre impose des conditions particulières sur les paramètres. Ainsi, la concentration de A ne peut grimper que si au début de la phase I, il y a trop peu de B produits pour titrer les protéines A, ce qui impose en fait que $r_1 < 1$. Au contraire, A ne va chuter que lorsqu'il y a suffisamment de B produits pour titrer les protéines A, ce qui arrive uniquement s'il y a suffisamment de transcrits présents dans le système, ici $r > 1$. Ceci impose nécessairement la condition $\rho_1 > 1$.

À la fin de la phase I nous avons donc les équations suivantes :

$$r_I(t_1) = \rho_1 + (r_1 - \rho_1)e^{-t_1} \quad (\text{III.49})$$

$$0 = (1 - \rho_1)(1 - e^{-d_a t_1})/d_a - (\rho_1 - r_1)(e^{-d_a t_1} - e^{-t_1})/(d_a - 1) \quad (\text{III.50})$$

D.4 Phase II : concentration de A basse, concentration de B haute

Nous définissons cette seconde phase comme la phase où la concentration de A passe en-dessous de $[A]_0$, i.e. $B \gg 1 \gg A$. D'après la forme de l'équation III.9, on peut supposer que B est alors d'ordre $1/\delta$ et nous supposons alors $b = \delta B$.

Dans la limite des petits δ , la cinétique de titrage des protéines A est alors très rapide par rapport aux cinétiques d'évolution des autres variables et nous pouvons faire l'approximation adiabatique suivante :

$$A \simeq \delta/b \quad (\text{III.51})$$

et la dynamique de la MFL se réduit alors à l'ordre le plus bas aux trois équations suivantes :

$$\frac{dg}{dt} = \tilde{\theta}(1 - g) \quad (\text{III.52})$$

$$\frac{dr}{dt} = \rho_0 g + \rho_1(1 - g) - r \quad (\text{III.53})$$

$$\frac{db}{dt} = r - 1 - d_b b \quad (\text{III.54})$$

Par continuité de la fin de la phase I avec le début de la phase II, nous supposons maintenant qu'au début de la phase II, $b = 0$, $r = r_2$ et $g = 0$ - ces conditions initiales seront étudiées plus en détail par la suite. Les équations III.52 à III.54 peuvent alors être facilement intégrées et on trouve :

$$g_{II}(t) = 1 - e^{-\tilde{\theta}t} \quad (\text{III.55})$$

$$r_{II}(t) = \rho_0 + (r_2 - \rho_0 + (\rho_1 - \rho_0)/(\tilde{\theta} - 1))e^{-t} - (\rho_1 - \rho_0)/(\tilde{\theta} - 1)e^{-\tilde{\theta}t} \quad (\text{III.56})$$

$$b_{II}(t) = (r_2 - \rho_0 + (\rho_1 - \rho_0)/(\tilde{\theta} - 1))(e^{-d_b t} - e^{-t})/(1 - d_b) - (\rho_1 - \rho_0)/((\tilde{\theta} - d_b)(\tilde{\theta} - 1))(e^{-d_b t} - e^{-\tilde{\theta}t}) + (\rho_0 - 1)(1 - e^{-d_b t})/d_b \quad (\text{III.57})$$

Comme b est nul au début de la phase II, b doit tout d'abord augmenter avec le temps. Cela suppose que la production de B domine la complexation, et donc que r soit plus grand que 1 au début de la phase II, soit $r_2 > 1$. Aux temps longs, g tend

vers 1 et donc la concentration de r doit tendre vers ρ_0 . A la fin de la phase II, b doit décroître vers 0 par continuité avec la fin de la phase I. De la même façon que précédemment, cela impose qu'aux temps longs, on ait $r < 1$, et donc compte-tenu de la limite de r aux temps longs, il faut $\rho_0 < 1$. Avec ces conditions, la phase II s'achève après un temps t_2 et on a $b_{II}(t_2) = 0$. À la fin de la phase II nous avons donc les équations suivantes :

$$g_{II}(t_2) = 1 - e^{-\tilde{\theta}t_2} \quad (\text{III.58})$$

$$r_{II}(t_2) = \rho_0 + (r_2 - \rho_0 + (\rho_1 - \rho_0)/(\tilde{\theta} - 1))e^{-t_2} - (\rho_1 - \rho_0)/(\tilde{\theta} - 1)e^{-\tilde{\theta}t_2} \quad (\text{III.59})$$

$$0 = (r_2 - \rho_0 + (\rho_1 - \rho_0)/(\tilde{\theta} - 1))(e^{-d_b t_2} - e^{-t_2})/(1 - d_b) - (\rho_1 - \rho_0)/((\tilde{\theta} - d_b)(\tilde{\theta} - 1))(e^{-d_b t_2} - e^{-\tilde{\theta}t_2}) + (\rho_0 - 1)(1 - e^{-d_b t_2})/d_b \quad (\text{III.60})$$

D.5 Condition de continuité pour la description du cycle limite

Il reste donc maintenant quatre inconnues pour décrire complètement les oscillations. Les équations de continuité au début et à la fin de chaque phase nous donnent quatre équations :

$$r_I(t_1) = r_2 \quad , \quad a_I(t_1) = 0 \quad (\text{III.61})$$

$$r_{II}(t_2) = r_1 \quad , \quad b_{II}(t_2) = 0 \quad (\text{III.62})$$

Ces quatre équations peuvent être résolues numériquement pour calculer r_1, r_2, t_1, t_2 à partir des paramètres ρ_1, ρ_0, d_a, d_b et $\tilde{\theta}$, et permettent d'obtenir la période rescalée des oscillations

$$T_r = t_1 + t_2, \quad (\text{III.63})$$

la période dimensionnée de l'oscillateur étant donnée par $T = T_r/\delta_r$. La Figure III.6 présente la période mesurée de l'oscillateur en fonction de la période théorique pour différentes valeurs de δ : dans la limite $\delta \rightarrow 0$, la période T_r tend vers une fonction indépendante de δ et dépendant des cinq paramètres $\rho_1, \rho_0, d_a, d_b, \tilde{\theta}$.

Les équations III.61 et III.62 sont a priori difficiles à résoudre analytiquement, cependant, il est possible d'obtenir des valeurs analytiques de la période dans certains

cas limites. Par exemple, dans la limite $d_b \simeq 0$, et ρ_1 grand, l'équation pour $b_{II}(t_2) = 0$ devient, en faisant un développement de Taylor

$$0 = (\rho_0 - 1)t_2 + (r_2 - \rho_0 + (\rho_1 - \rho_0)/(\tilde{\theta} - 1))(1 - e^{-t_2}) - (\rho_1 - \rho_0)/((\tilde{\theta})(\tilde{\theta} - 1))(1 - e^{-\tilde{\theta}t_2}) \quad (\text{III.64})$$

Dans la limite des grands ρ_1 , à des termes exponentiellement petit près, nous trouvons donc :

$$t_2 \simeq (r_2 - \rho_0 + (\rho_1 - \rho_0)/\tilde{\theta})/(1 - \rho_0) \quad (\text{III.65})$$

et la condition de continuité à la fin de la phase II venant de l'équation III.62 devient :

$$r_1 \simeq \rho_0 \quad (\text{III.66})$$

De la même façon, à l'aide la condition de continuité à la fin de la phase I définie dans l'équation III.61 et de l'équation III.50, dans la limite des grands ρ_1 , t_1 sera petit devant 1. A l'aide d'un développement de Taylor de l'équation III.50 on trouve à l'ordre dominant :

$$t_1 \simeq \frac{2(1 - \rho_0)}{\rho_1} \quad (\text{III.67})$$

Maintenant, en utilisant la relation de continuité III.61 pour r_I à la fin de la phase I et l'équation III.49, on trouve immédiatement que

$$r_2 \simeq 2 - \rho_0 \quad (\text{III.68})$$

Cette dernière relation permet finalement d'avoir une valeur approchée de la période pour les grands ρ_1 dans la limite $d_b \simeq 0$:

$$T_r \simeq t_2 \simeq 2 + \frac{\rho_1 - \rho_0}{\tilde{\theta}(1 - \rho_0)} \quad (\text{III.69})$$

Cette période approchée est également représentée figure III.6 et est conforme à la période mesurée dans la limite des petits δ .

E Raccordement des régimes à l'ordre $\sqrt{\delta}$ et phase intermédiaire

Dans la suite, nous justifions certaines hypothèses faites précédemment pour raccorder les deux régimes. Nous étudions en particulier les transitions entre les deux phases. Cette étude à l'ordre $\sqrt{\delta}$ est également nécessaire pour justifier les conditions initiales prises sur la concentration du gène.

E.1 Transition de la phase I à la phase II

Nous considérons la transition entre la fin de la phase I avec A grand et B petit au début de la phase II où B est grand et A petit. Cette transition a lieu durant un temps d'ordre δ/δ_r . Durant cet intervalle de temps, ni la concentration de transcrits, ni la valeur de g_b n'a le temps de changer par rapport aux concentrations de protéines. Nous supposons donc dans les équations III.6 à III.9 que $r = r_2 > 1$ et redimensionnons le temps en posant $\tau = t/\delta$. Les équations III.8 et III.9 deviennent alors à l'ordre dominant :

$$\frac{dA}{dt_i} = 1 - A.B \quad (\text{III.70})$$

$$\frac{dB}{dt_i} = r_2 - A.B \quad (\text{III.71})$$

Les conditions aux limites pour raccorder les régimes sont :

$$A \rightarrow (1 - r_2)(\tau - \tau_1) + o(1), B \rightarrow 0 \quad \text{pour } \tau \rightarrow -\infty \quad (\text{III.72})$$

$$B \rightarrow (r_2 - 1)(\tau - \tau_2) + o(1), A \rightarrow 0 \quad \text{pour } \tau \rightarrow +\infty \quad (\text{III.73})$$

avec $r_2 > 1$ et τ_1 et τ_2 deux constantes à déterminer.

En faisant la différence entre les équations III.70 et III.71, on trouve immédiatement, en fixant l'origine des temps au moment où $A = B$, que

$$A - B = (1 - r_2)\tau \quad (\text{III.74})$$

Ce choix d'origine des temps implique immédiatement, d'après les équations III.72 et III.73 que $\tau_1 = \tau_2 = 0$. Les conditions asymptotiques III.72 et III.73 coïncident également avec les comportements respectifs de $a_I(t)/\delta$ à la fin de la phase I lorsque $t = t_1$ et de b_{II}/δ au début de la phase II lorsque $t = 0$ si l'on suppose que $a_I(t_1) = b_{II}(0) = 0$, ce qui justifie les hypothèses faites précédemment.

Les équations III.70 et III.71 donnent donc une approximation du comportement de $A(t)$ et de $B(t)$ pendant la transition de la phase I à la phase II. De plus, si l'on substitue l'équation III.74 dans l'équation III.70, nous pouvons remplacer les deux équations III.70 et III.71 par une seule équation de Riccati :

$$\frac{dA}{dt_i} = 1 - (A + (r_2 - 1)\tau)A \quad (\text{III.75})$$

Cependant, notons que l'approximation quasi-statique faite pour g dans la phase I (Équation III.43) diverge à la fin de la phase I lorsque A tend vers 0. Ceci signifie que l'approximation faite dans l'équation III.43 n'est plus valable, et il faut considérer une région intermédiaire plus grande pour raccorder les différents comportements de g . En fait, dans la région où $t_1 - t \sim \sqrt{\delta}$, l'équation pour g se réécrit :

$$\frac{dg}{dt} = \tilde{\theta} \left(1 - g \frac{a}{\delta A_0}\right) \quad (\text{III.76})$$

Or, à la fin de la phase I, nous avons $a = (1 - r_2)(t_1 - t) + o(\sqrt{\delta})$, si bien que tous les termes de l'équation III.76 sont bien du même ordre de grandeur dans la région où $t_1 - t \sim \sqrt{\delta}$ (car g lui-même est alors de l'ordre $\sqrt{\delta}$ dans l'approximation quasi-statique donnée par l'équation III.43). L'équation III.76 s'intègre explicitement, et on trouve :

$$g(t_1) - g(t) e^{\tilde{\alpha}(1-r_2)(t_1-t)^2/2\delta} = \int_{t-t_1}^0 \tilde{\theta} e^{\tilde{\alpha}(1-r_2)u^2/2\delta} du \quad (\text{III.77})$$

Ceci permet de calculer explicitement la valeur de g à la fin de la phase I et au début de la phase II donnée par la limite $t \rightarrow -\infty$ dans l'équation III.77 :

$$g_I(t_1) = g_2 = \tilde{\theta} \sqrt{\frac{\pi\delta}{2\tilde{\alpha}(r_2 - 1)}} \quad (\text{III.78})$$

L'équation III.78 donne une première correction à l'ordre $\sqrt{\delta}$ pour la période comme nous le verrons ci-dessous.

E.2 Transition de la phase II à la phase I

Nous considérons maintenant la transition entre la fin de la phase II où B est grand et A petit au début de la phase II où A est grand et B petit. Le traitement est tout à fait similaire au cas précédent. Cette transition a lieu durant un temps d'ordre

δ/δ_r . Durant cet intervalle de temps, ni la concentration de transcrits, ni la valeur de g_b n'a le temps de changer par rapport aux concentrations de protéines. Nous supposons donc dans les équations III.6 à III.9 que $r = r_1 < 1$ et redimensionnons le temps en posant $\tau = t/\delta$. Les conditions aux limites pour raccorder les régimes sont cette fois :

$$B \rightarrow (r_1 - 1)(\tau - \tau'_1) + o(1), A \rightarrow 0 \quad \text{pour} \quad \tau \rightarrow +\infty \quad (\text{III.79})$$

$$A \rightarrow (1 - r_1)(\tau - \tau'_2) + o(1), B \rightarrow 0 \quad \text{pour} \quad \tau \rightarrow -\infty \quad (\text{III.80})$$

De la même façon, $\tau'_1 = \tau'_2 = 0$. Ensuite, de nouveau, il faut raccorder les deux régimes plus précisément pour g : g est d'ordre 1 à la fin de la phase II et d'ordre δ durant la phase I. En fait, dans la région où $t - t_2 \sim \sqrt{\delta}$, l'attachement de la protéine A au promoteur de g_b n'est pas instantané et il faut que quelques protéines A s'accumulent avant d'avoir cette réaction. Au début de la phase I, nous avons cette fois $a = (1 - r_1)t + o(\sqrt{\delta})$. L'équation III.76 s'intègre de nouveau explicitement, et on trouve :

$$g_I(t) = \exp(\tilde{\alpha}(r_1 - 1)t^2/2\delta) \left[g_1 + \tilde{\theta} \int_0^t du e^{\tilde{\alpha}(1-r_1)u^2/2\delta} \right] \quad (\text{III.81})$$

L'équation III.81 décrit en fait la transition de $g_1 = 1 - \exp(\tilde{\theta}t_2)$, la valeur de g à la fin de la phase II, au régime quasi-statique donné par l'équation III.43, car le comportement asymptotique de l'équation III.81 est alors :

$$g(t) \sim \frac{A_0\delta}{(1 - r_1)t} \quad (\text{III.82})$$

E.3 Corrections à l'ordre $\sqrt{\delta}$

La forme des équations III.6 à III.9 laisse à penser que la période des oscillations peut se développer en puissance de δ . En fait, l'analyse des couches limites faite précédemment montre que la correction dominante est d'ordre $\sqrt{\delta}$. Plus précisément, l'équation III.55 doit en fait être remplacée, en vertu de l'analyse précédente, par l'équation suivante :

$$g_{II,c}(t) = 1 - (1 - g_2)e^{-\tilde{\theta}t} \quad (\text{III.83})$$

Il faut également corriger les expressions de r_{II} et de b_{II} de la façon suivante :

$$r_{II,c}(t) = \rho_0 + (r_2 - \rho_0 + (1 - g_2)(\rho_1 - \rho_0)/(\tilde{\theta} - 1))e^{-t} - (1 - g_2)(\rho_1 - \rho_0)/(\tilde{\theta} - 1)e^{-\tilde{\theta}t} \quad (\text{III.84})$$

$$b_{II,c}(t) = (r_2 - \rho_0 + (1 - g_2)(\rho_1 - \rho_0)/(\tilde{\theta} - 1))\frac{e^{-d_b t} - e^{-t}}{1 - d_b} - (1 - g_2)(\rho_1 - \rho_0)/((\tilde{\theta} - d_b)(\tilde{\theta} - 1))(e^{-d_b t} - e^{-\tilde{\theta}t}) + (\rho_0 - 1)(1 - e^{-d_b t})/d_b \quad (\text{III.85})$$

De la même façon, la correction calculée précédemment lors de la transition de la phase II à la phase I amène à modifier les équations III.47 et III.48. La forme de ces expressions ne change pas, mais il faut prendre une nouvelle valeur de r_1 donnée par :

$$r_{1,c} = (r_{II,c}(t_2) - \rho_1 - (\rho_1 - \rho_0)g_1\sqrt{\pi\delta/\tilde{\alpha}(2(1 - r_1))}) \quad (\text{III.86})$$

Enfin, il faut résoudre implicitement les équations :

$$r_{I,c}(t_1) = r_2 \quad , \quad a_{I,c}(t_1) = 0 \quad (\text{III.87})$$

en ajoutant également la condition $b_{II,c}(t_2) = 0$. La comparaison entre l'ordre $\sqrt{\delta}$ et l'intégration complète est donnée dans les figures III.7 et III.8 .

F Comparaison avec un modèle avec délai

Un modèle classique d'oscillateur génétique consiste en une protéine réprimant sa propre expression avec un délai [50,61]. Si l'on utilise une fonction de Hill pour modéliser cette répression, un simple modèle peut s'écrire :

$$\frac{dB(t)}{dt} = \frac{\rho}{B_0^n + B^n(t - \tau)} - cB \quad (\text{III.88})$$

où ρ est le taux de production de la protéine, c le taux de dégradation de la protéine et τ le délai phénoménologique de répression.

En posant $t = \theta\tau$, $\varepsilon = 1/(c\tau)$, $r = \rho/(cB_0^{n+1})$, $b(\theta) = B(t)/B_0$, l'équation adimensionnée pour la variable b s'écrit :

$$\varepsilon \frac{db}{d\theta} = \frac{r}{1 + b^n(\theta - 1)} - b \quad (\text{III.89})$$

Le diagramme de phase de cet oscillateur peut être calculé explicitement [34]. Pour $c\tau \gg 1$, les oscillations sont non-linéaires. La période d'un tel oscillateur est alors donnée par l'expression suivante :

$$T/\tau = 2 + \frac{\kappa}{c\tau} + \dots \quad (\text{III.90})$$

où κ est une constante proche de 2 pour $n = 2$ et ρ/c d'ordre 1 [50], voir aussi l'annexe C.

Il est possible de noter certaines analogies entre ce modèle simple avec délai et le présent modèle de MFL. Ainsi, si $\tau \gg 1/c$, la période du modèle avec délai est proportionnelle à τ tandis que celle de la MFL est proportionnelle à δ_r^{-1} . La demi-vie de l'ARN dans la MFL joue donc le même rôle que le délai phénoménologique dans l'équation III.88, ce qui est cohérent avec le rôle joué par la dynamique de l'ARN comme élément majeur déterminant l'oscillation. Le paramètre sans dimension ε dans l'équation III.89 joue un rôle tout à fait similaire au petit paramètre δ dans la MFL dans la mesure où, pour de grandes valeurs de ε , les oscillations disparaissent [34], et que pour les petites valeurs de ce paramètre, la période et la dynamique de l'oscillateur ne dépend pas de ε à l'ordre le plus grand [50], voir aussi l'annexe C. Dans le modèle avec délai, ε représente le rapport entre la durée de vie typique de la protéine sur le délai, alors que dans la MFL, δ représente le rapport entre le temps typique de production des protéines et de la séquestration sur la demi-vie de l'ARN. Ces deux analogies suggèrent que la constante de dégradation c de la protéine dans le modèle avec délai joue le même rôle que le paramètre $\sqrt{\rho_A \gamma}$ dans la MFL. Ceci signifie que la séquestration dans la MFL joue typiquement le même rôle que la dégradation dans le modèle avec délai. La cinétique rapide de ces deux processus comparée à celle d'un processus plus lent (délai τ , dynamique de l'ARN) permet l'alternance au cours de l'oscillation entre deux états (deux concentrations typiques pour la protéine B dans le modèle avec délai, deux phases dans la MFL).

G Conclusions

Nous avons présenté dans ce chapitre une étude mathématique d'un réseau génétique simple. Ce réseau présente la propriété intéressante de pouvoir se comporter à la fois comme un système bistable et à la fois comme un oscillateur, suivant les différents taux de transcription des protéines. La dimérisation joue un rôle crucial dans ce système en introduisant un véritable commutateur au niveau des protéines.

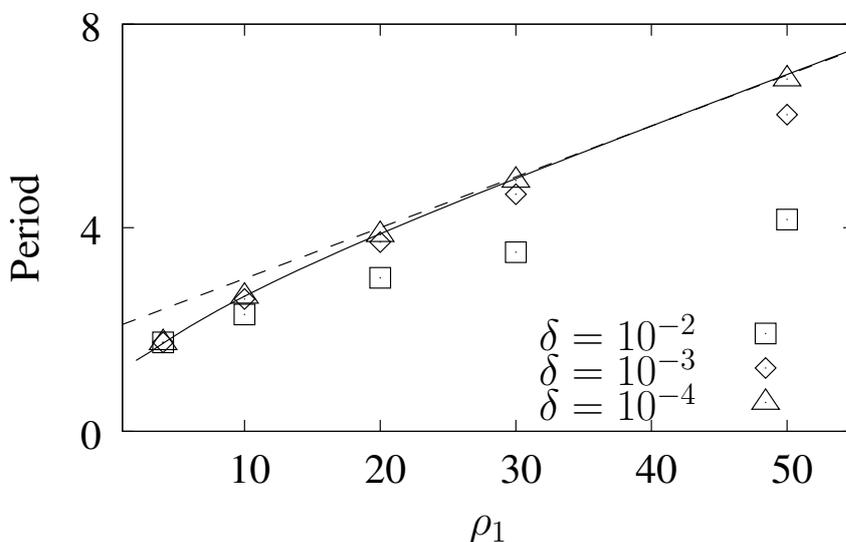


FIG. III.6: Comparaison entre la période analytique calculée à l'ordre le plus bas (trait plein), l'approximation de l'équation III.69 (en pointillés) et la période mesurée numériquement pour différentes valeurs de δ en fonction du paramètre ρ_1 . Les autres paramètres sont $\tilde{\theta} = 10$, $\tilde{\alpha} = \mu = 1$, $d_a = d_b = 0$.

Le système bistable est alors obtenu en couplant une répression transcriptionnelle à cette dimérisation ; l'oscillateur en lui couplant une activation transcriptionnelle.

La dynamique de l'ARN est également très importante pour l'oscillateur. Ce processus lent est fondamental pour la naissance des oscillations : les concentrations des deux protéines sont couplées à la concentration des transcrits de la protéine B. L'importance de cette étape de la transcription pour l'oscillation sera discutée plus en détail dans le modèle biologique présenté dans le chapitre IV.

Le rôle potentiel de ce motif en tant que commutateur génétique n'a été observé biologiquement, à notre connaissance, uniquement dans le cas de l'opéron lactose, comme nous l'avons vu à la fin du chapitre II. En fait, comme l'ont remarqué Mazurie et al [58], dans la plupart des réseaux génétiques où la MFL est impliquée, les différents motifs sont très enchevêtrés si bien qu'il paraît impossible de traiter ce motif isolément. En revanche, le motif oscillant semble au contraire relativement répandu, en particulier dans les horloges circadiennes. Nous avons donc choisi de modéliser l'horloge circadienne de *Neurospora crassa* par une MFL, et c'est cette approche que nous développons dans le chapitre IV.

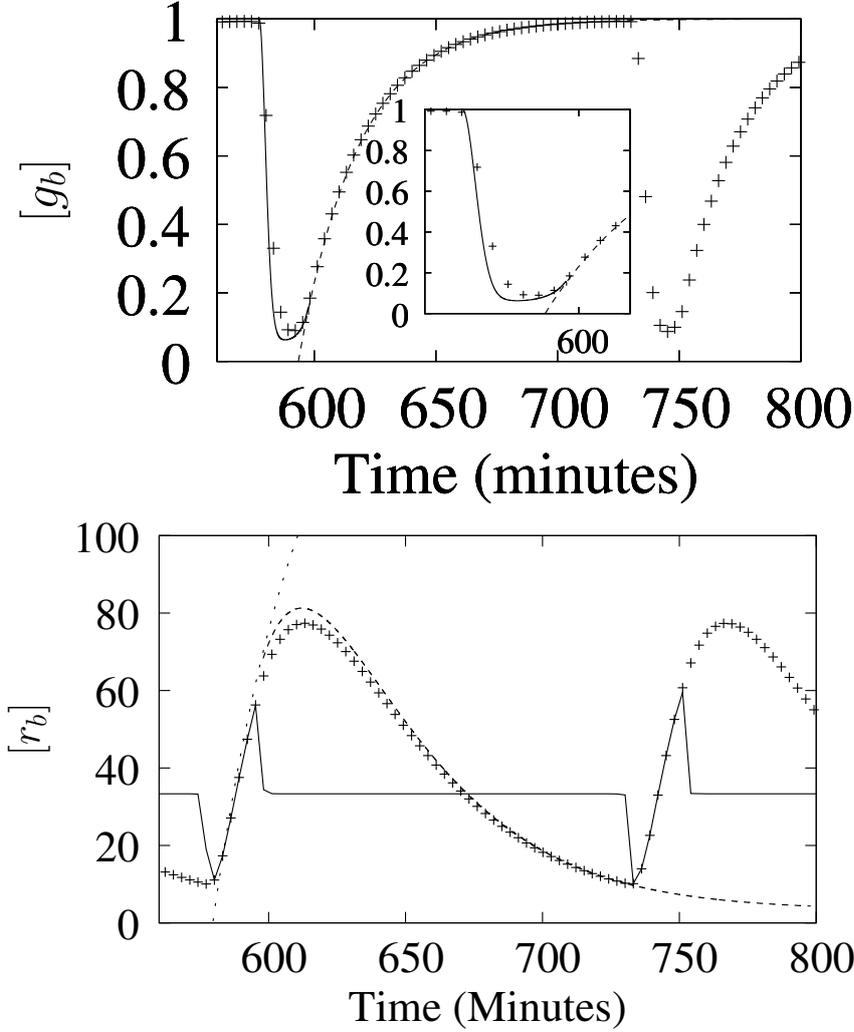


FIG. III.7: Comparaison entre l'approximation analytique à l'ordre $\sqrt{\delta}$ dans les deux phases (en pointillés) et le comportement réel intégré numériquement (symbole +). Sur le panneau du haut, nous avons tracé dans la phase I les différents comportements asymptotiques reliés à l'approximation quasi-stationnaire. Sur le panneau du bas, le produit AB/β est tracé en trait plein pour comparer avec l'approximation quasi-statique : si l'approximation quasi-statique est vraie, il vaut $[r_b]$ durant la phase I et ρ_A/β durant la phase II. Les paramètres choisis sont $\rho_A = 100 \text{ mol.min}^{-1}$, $\rho_f = 5 \text{ mol.min}^{-1}$, $\rho_b = 0.1 \text{ mol.min}^{-1}$, $\beta = 3 \text{ min}^{-1}$, $\theta = 0.04 \text{ min}^{-1}$, $\alpha = 0.001 \text{ mol}^{-1}.\text{min}^{-1}$, $\delta_A = 0.01 \text{ min}^{-1}$, $\delta_r = 0.03 \text{ min}^{-1}$, $\delta_B = 0.01 \text{ min}^{-1}$, $\gamma = 1 \text{ mol}^{-1}.\text{min}^{-1}$. Les paramètres rescalés correspondant sont $\rho_1 = 5$, $\rho_0 = 0.1$, $\delta = 0.003$, $\theta = 1.33$, $\alpha = 0.33$, $\mu = 0.31$, $d_a = d_b = 0.33$.

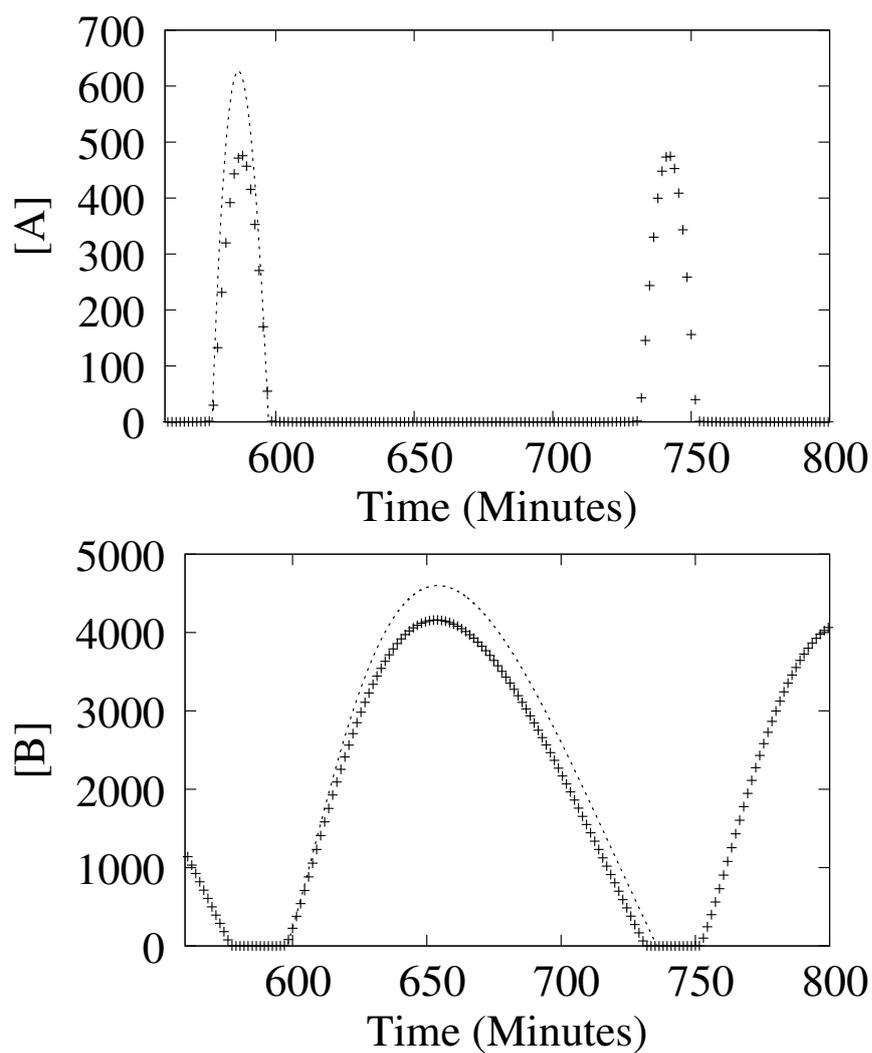


FIG. III.8: Comparaison entre l'approximation analytique à l'ordre $\sqrt{\delta}$ dans les deux phases (en pointillés) et le comportement réel intégré numériquement (symbole +), pour les protéines. Les paramètres sont les mêmes que pour la figure III.7.

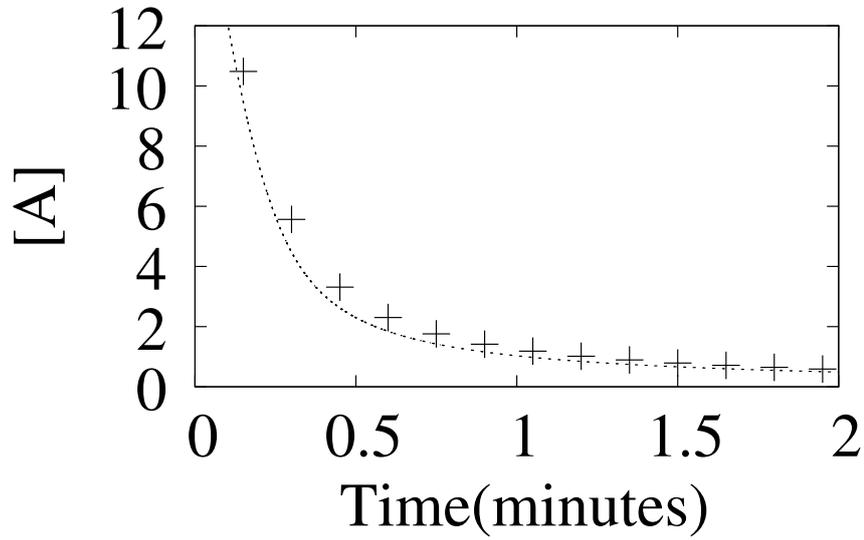


FIG. III.9: Comparaison entre le comportement réel et l'équation de Riccati durant la phase intermédiaire. Les paramètres sont les mêmes que dans la figure III.7.

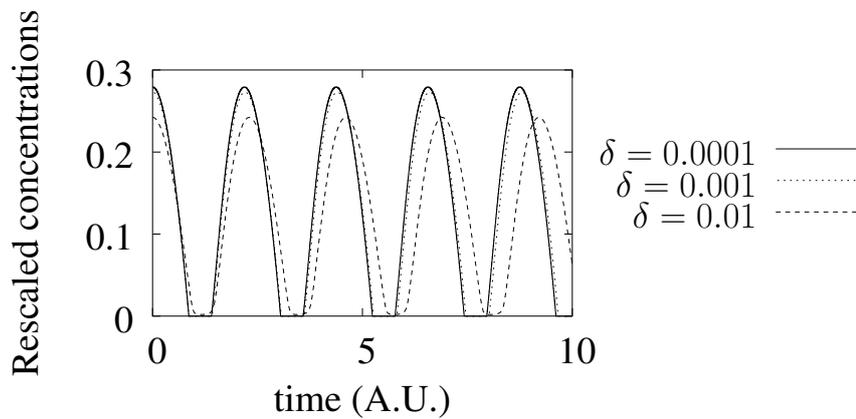


FIG. III.10: Amplitude et période des oscillations de A en fonction de δ . Nous traçons ici le produit $(\delta \cdot A)$ en fonction du temps t . Les paramètres considérés ici sont $\rho_1 = 3, \tilde{\theta} = 5, \tilde{\alpha} = 10, \mu = 1, d_a = d_b = 0$

H Article

Les résultats présentés dans cette partie ont été mis en forme dans un article publié dans *Phys. Rev. E*, reproduit ci-après, qui correspond à la référence [28].

Core genetic module: The mixed feedback loop

Paul François and Vincent Hakim

Laboratoire de Physique Statistique,* CNRS-UMR 8550, Ecole Normale Supérieure, 24, rue Lhomond 75231 Paris, France

(Received 27 May 2005; published 16 September 2005)

The so-called mixed feedback loop (MFL) is a small two-gene network where protein A regulates the transcription of protein B and the two proteins form a heterodimer. It has been found to be statistically over-represented in statistical analyses of gene and protein interaction databases and to lie at the core of several computer-generated genetic networks. Here, we propose and mathematically study a model of the MFL and show that, by itself, it can serve both as a bistable switch and as a clock (an oscillator) depending on kinetic parameters. The MFL phase diagram as well as a detailed description of the nonlinear oscillation regime are presented and some biological examples are discussed. The results emphasize the role of protein interactions in the function of genetic modules and the usefulness of modeling RNA dynamics explicitly.

DOI: 10.1103/PhysRevE.72.031908

PACS number(s): 87.17.Aa, 87.16.Yc, 82.40.Bj

I. INTRODUCTION

Biological cells rely on complex networks of biochemical interactions. Large scale statistical analyses have revealed that the transcriptional regulation networks and the networks of protein-protein interaction for different organisms are far from random and contain significantly recurring patterns [1] or motifs. Mathematical modeling is useful to determine if these motifs can by themselves fulfill useful functions and it has, for instance, helped to show that the common “feed forward loop” motif in transcriptional regulation [1] can act as a persistence detector [2]. More recently, a combined analysis of protein-protein interactions and transcriptional networks in yeast (*Saccharomyces cerevisiae*) has pointed out several motifs of mixed interactions [3,4]. The simplest such motif composed of both transcriptional and protein-protein interactions is the two-protein mixed feedback loop (MFL) depicted in Fig. 1. It is composed of a transcription factor A , produced from gene g_a , and of another protein B , produced from gene g_b . A regulates the transcription of gene g_b and also directly interacts with protein B . The MFL has independently been obtained as the core motif of several networks produced in a computer evolutionary search aiming at designing small functional genetic modules performing specific functions [5]. Here, to better understand the possible functions of this basic module, we propose a model of the MFL based on the simplest biochemical interactions. This mathematical model is described in Sec. II. This is used to show that there exist wide ranges of kinetic parameters where the MFL behaves either as a bistable switch or as an oscillator. For the convenience of the reader, an overview of the different dynamical regimes is provided in Sec. III. They are delimited in parameter space and their main characteristics are summarized. We then give detailed descriptions of the bistable regime in Sec. IV and of the nonlinear oscillations in Sec. V. A comparison is also made with a simple autoinhibitory gene model with delay. In the concluding section, the important role played by protein dimerization and

the usefulness of explicitly modeling mRNA dynamics are underlined and biological examples of the two proposed functions of the MFL are discussed.

II. A MATHEMATICAL MODEL OF THE MFL

A. Model formulation

As previously described, the MFL consists of two proteins A and B and their genes g_a and g_b , such that A regulates the transcription of gene g_b and also directly interacts with B . Our aim is to analyze the dynamics of this small genetic module and see what can be achieved in the simplest setting. Therefore, different cellular compartments and separate concentrations for the nucleus and cytoplasm are not considered and biochemical reactions are modeled by simple rate equations.

The proposed MFL model is represented schematically in Fig. 2 and consists of four equations that are described and explained below. The concentration of a chemical species X is denoted by square brackets and the cell volume is taken as volume unit. So, $[X]$ represents the effective number of X molecules present in the cell.

The first two equations model the transcriptional regulation of gene g_b by protein A

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

$$\frac{d[r_b]}{dt} = \rho_f[g_b] + \rho_b[g_b:A] - \delta_r[r_b], \quad (2)$$

where it is assumed that gene g_b exists under two forms, with A bound to its promoter with probability $[g_b:A]$ and without



FIG. 1. Schematic representation of the over-represented MFL motif [4]. The bold arrow represents a transcriptional regulation interaction and the dashed double arrow represents a protein-protein interaction.

*LPS is laboratoire associé aux universités Paris VI and VII.

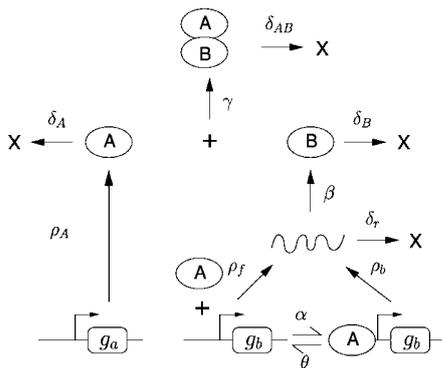


FIG. 2. The proposed model of the MFL motif. The Greek letters denote the model's different kinetic constants. The large crosses symbolize the degradation of the corresponding species.

A with probability $[g_b]$. Since $[g_b] + [g_b:A] = 1$, the single Eq. (1) is sufficient to describe the transition between the two forms. Specifically, A proteins bind to the g_b promoter at a rate α and when bound they are released at a rate θ . The regulation of transcription of gene g_b by protein A is described by Eq. (2) where r_b stands for g_b transcripts. When A is bound to the g_b promoter, the transcription is initiated at a rate ρ_b and, otherwise, it is initiated at a rate ρ_f . Thus, $\rho_b > \rho_f$ corresponds to transcriptional activation by A and $\rho_b < \rho_f$ to transcriptional repression. A first-order degradation for g_b mRNA at a constant rate δ_r has also been assumed. As given, the description is strictly valid for a single copy gene. However, it also applies for a gene with g_0 copy (e.g., $g_0 = 2$) provided that ρ_f and ρ_b are understood to be g_0 -fold greater than the corresponding elementary rates.

The production of the A and B proteins and their complexation are described by the following two equations that complete our description of the MFL module:

$$\frac{d[A]}{dt} = \rho_A - \gamma[B][A] - \delta_A[A] + \theta[g_b:A] - \alpha[g_b][A], \quad (3)$$

$$\frac{d[B]}{dt} = \beta[r_b] - \gamma[B][A] - \delta_B[B], \quad (4)$$

where $[A]$ and $[B]$, respectively, denote the concentration of proteins A and B . Since regulation of gene g_a is not considered, separate descriptions of its transcription and translation are not needed and it is simply assumed in Eq. (3) that protein A is produced at a given basal rate ρ_A . The second crucial interaction of the MFL, the direct interaction between protein A and B is taken into account by assuming that A and B associate at a rate γ . It is assumed that B does not interact with a protein A that is bound to the g_b promoter. In addition, Eq. (3) assumes a first-order degradation of protein A at a constant rate δ_A , and its last two terms come from the (small) contribution to the concentration of A in solution of the binding (unbinding) of A to (from) the g_b promoter. Equation (4) assumes that B proteins are produced from the transcripts of gene g_b at a rate β and are degraded at a rate δ_B in addition to their association with A proteins.

As described by Eqs. (3) and (4), the complexation between A and B proteins proceeds at a rate γ . For simplicity, we suppose that the AB complex does not interact with genes g_a and g_b , we neglect its dissociation [6] and simply assume that it is fully degraded at a rate δ_{AB} after its formation,

$$\frac{d[AB]}{dt} = \gamma[B][A] - \delta_{AB}[AB]. \quad (5)$$

Since the complex AB does not feed back on the dynamics of the other species, its concentration does not need to be monitored and Eq. (5) is not explicitly considered in the following.

B. Values of kinetic parameters and a small dimensionless parameter

Even in this simple model, ten kinetic constants should be specified. It is useful to consider the possible range of their values both to assess the biological relevance of the different dynamical regimes and to orient the model analysis.

Half-lives of mRNA range from a few minutes to several hours and are peaked around 20 min in yeast [7]. Therefore, $\delta_r = 0.03 \text{ min}^{-1}$ can be taken as a typical value.

For the transcription factor-gene promoter interaction, typical values appear to be a critical concentration $\theta/\alpha = [A]_0$ in the nanomolar range, a bound state lifetime of several minutes and activated transcription rates of a few mRNAs per minute. Therefore, we assume $\alpha = \theta/40$, that θ is of the same order as δ_r and ρ_f and ρ_b range from 0.1 min^{-1} to 10 min^{-1} .

Protein half-lives vary from a few minutes to several days [8]. The hour appears as a typical value. We choose $\delta_A = \delta_B = 0.01 \text{ min}^{-1}$ and more generally consider that the A and B protein half-lives are of the same order or longer than that of g_b mRNA. For protein production, $\beta = 3$ protein molecules per mRNA molecule per minute appears a plausible value for an eucaryotic cell [9]. This gives $\rho_A \approx 100 \text{ min}^{-1}$ when combined with the previous values for mRNA production.

We assume that the protein interaction is essentially limited by diffusion. A diffusion constant $D \approx 2.5 \mu\text{m}^2 \text{ s}^{-1}$ [10] gives a time s^2/D of about a minute for diffusion across a cell of a size $s = 10 \mu\text{m}$. Therefore, we choose $\gamma \approx 1 \text{ min}^{-1}$.

It is convenient to introduce dimensionless variables [Eqs. (1)–(4)] to decrease as far as possible the number of independent parameters.

We first normalize the g_b mRNA concentration by the concentration that gives a production of B protein equal to that of A . Thus, we define the dimensionless concentration $r = \beta[r_b]/\rho_A$. We normalize the protein concentrations by the equilibrium concentration resulting from production at a rate ρ_A and dimerization at a rate γ and define $A = \sqrt{\gamma/\rho_A}[A]$, $B = \sqrt{\gamma/\rho_A}[B]$. We also take as the time unit the inverse of g_b mRNA degradation rate $1/\delta_r$. With these substitutions, Eqs. (1)–(4) read,

$$\frac{dg}{dt} = \tilde{\theta} \left[(1-g) - g \frac{A}{A_0} \right], \quad (6)$$

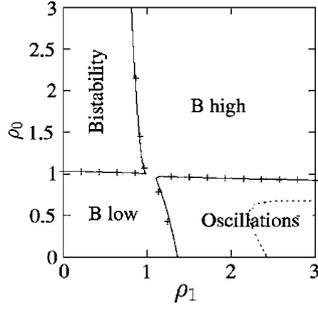


FIG. 3. Different dynamical regimes of the MFL as ρ_0 and ρ_1 are varied. The borders between different regimes are shown as computed from numerical solutions of Eqs. (18) and (31) for $\tilde{\theta} = 1.33$ (full lines) or $\tilde{\theta} = 26.6$ (dashed lines) as well as given by the asymptotics expansions [Eqs. (28), (30), (36), and (41)] for $\tilde{\theta} = 1.33$ (+ symbols). Here, and in all the following figures, except when explicitly specified, the other parameters are $[A]_0 = 40$ mol, $\beta = 3$ min $^{-1}$, $\delta_r = 0.03$ min $^{-1}$, $\delta_A = 0.01$ min $^{-1}$, $\delta_B = 0.01$ min $^{-1}$, $\rho_A = 100$ mol min $^{-1}$, $\gamma = 1$ mol $^{-1}$ min $^{-1}$ where mol stands for molecules and min for minutes. The corresponding dimensionless parameters are $\delta = 0.003$, $d_a = d_b = 0.33$, $A_0 = 4$, $\mu = 0.31$.

$$\frac{dr}{dt} = \rho_0 g + \rho_1 (1 - g) - r, \quad (7)$$

$$\frac{dB}{dt} = \frac{1}{\delta} (r - AB) - d_b B, \quad (8)$$

$$\frac{dA}{dt} = \frac{1}{\delta} (1 - AB) + \mu \tilde{\theta} \left[(1 - g) - g \frac{A}{A_0} \right] - d_a A, \quad (9)$$

where we have defined the following rescaled parameters $\delta = \delta_r / \sqrt{\rho_A \gamma}$, $\rho_1 = \beta \rho_b / (\rho_A \delta_r)$, $\rho_0 = \beta \rho_f / (\rho_A \delta_r)$, $\tilde{\theta} = \theta / \delta_r$, $\mu = \sqrt{\gamma / \rho_A}$, $d_a = \delta_A / \delta_r$ and $d_b = \delta_B / \delta_r$. We have also introduced the dimensionless critical binding concentration $A_0 = \sqrt{\gamma / \rho_A} \theta / \alpha$. The model still depends on seven parameters. In order to simplify its analysis, it is useful to note that δ is a small parameter (approximately equal to 3×10^{-3} with the previous estimations). The influence of three key parameters of order one in Eqs. (6)–(9) is particularly examined in the following. These are ρ_0 and ρ_1 which measure the strengths of the two possible states of B production (with or without A bound to gene g_b) as compared to that of A , and $\tilde{\theta}$ which compares the rates of A unbinding from DNA to that of mRNA degradation (α is supposed to vary with θ to maintain a fixed critical binding concentration A_0).

III. OVERVIEW OF THE DYNAMICS IN DIFFERENT PARAMETER REGIMES

We provide here an overview of the different dynamical regimes of the MFL, which are depicted in Fig. 3, and summarize their characteristics. The behavior of the MFL depends on whether protein A is a transcriptional activator or a transcriptional repressor and on the strengths of the two transcription rates of gene g_b (i.e., with A bound or not to its

promoter). More precisely, the key parameters are the strengths of B protein production, $\beta \rho_f / \delta_r$ and $\beta \rho_b / \delta_r$, as compared to the production rate ρ_A of protein A . Therefore, it is convenient to consider the previously introduced ratios $\rho_0 = \beta \rho_f / (\delta_r \rho_A)$ and $\rho_1 = \beta \rho_b / (\delta_r \rho_A)$.

We have observed that depending on the values of ρ_0 and ρ_1 the MFL can be monostable, exhibit bistability, or display oscillations. We qualitatively describe these three different cases in the following. Their biological relevance is further discussed in Sec. VI.

A. Monostable steady states

The simplest case occurs when the production rate of A is higher or lower than the production rate of B , irrespective of the state of the g_b promoter. That is when both $\beta \rho_f / \delta_r$ and $\beta \rho_b / \delta_r$ are either higher or lower than ρ_A . In this case, the MFL has a single stable state to which it relaxes starting from any initial conditions.

When both B production rates are higher than the production rate of A (i.e., $\rho_0 > 1$ and $\rho_1 > 1$), A proteins are quickly complexed by B and are unable to interact with the g_b promoter. The concentration $[A]$ of uncomplexed A proteins is, therefore, low and results from a simple balance between production and complexation. The high concentration of uncomplexed B proteins is the effective result from transcription at the free g_b promoter rate, complexation, and degradation,

$$[B] \approx \frac{1}{\delta_B} \left(\beta \frac{\rho_f}{\delta_r} - \rho_A \right), \quad [A] \approx \frac{\rho_A}{\gamma [B]}. \quad (10)$$

An equally simple but opposite result holds when both B production rates are smaller than the production rate of A ($\rho_0 < 1$ and $\rho_1 < 1$). Then, the concentration of uncomplexed A is high, the g_b promoter is occupied by A , and a low B concentration results from a balance between complexation and degradation,

$$[A] \approx \frac{1}{\delta_A} \left(\rho_A - \beta \frac{\rho_b}{\delta_r} \right), \quad [B] \approx \frac{\beta \rho_b}{\delta_r [A] \gamma}. \quad (11)$$

The dynamics of the MFL is richer when the production rate of A is intermediate between the two possible production rates of B , $\beta \rho_f / \delta_r$ and $\beta \rho_b / \delta_r$, i.e., when either $\rho_0 > 1 > \rho_1$ or $\rho_1 > 1 > \rho_0$. We consider these two cases in turn.

B. Transcriptional repression and bistability

We first discuss the case when A is a transcriptional repressor, $\rho_0 > 1 > \rho_1$. Then, two stable steady states can coexist. Let us suppose first that no A is bound to the g_b promoter. Then the production rate of B is larger than the production rate of A , and all produced A proteins are quickly complexed. This stably prevents the binding of A proteins to the g_b promoter and maintain a steady state with low A and high B concentrations approximately equal to

$$[B]_1 \approx \left(\beta \frac{\rho_f}{\delta_r} - \rho_A \right) \frac{1}{\delta_B}, \quad [A]_1 \approx \frac{\rho_A}{\gamma [B]_1}. \quad (12)$$

The second opposite possibility is that A is sufficiently abundant to repress the transcription of gene g_b . Then, since the

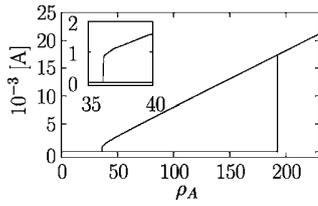


FIG. 4. In the bistable regime, a graded increase of ρ_A , the production of A , results in a jump in A concentration. The parameters are $\rho_f=0.2 \text{ mol min}^{-1}$, $\rho_b=2 \text{ mol min}^{-1}$, and $\theta=0.04 \text{ min}^{-1}$. The other parameters are as in Fig. 3.

production rate of A has been supposed to be higher than the production rate of B in the repressed state, B proteins are quickly complexed but uncomplexed A proteins are present to maintain the repression of the gene g_b transcription. This gives rise to a second stable state with high A and low B concentrations approximately equal to

$$[A]_2 \approx \left(\rho_A - \beta \frac{\rho_b}{\delta_r} \right) \frac{1}{\delta_A}, \quad [B]_2 \approx \frac{\beta \rho_b}{\gamma [A]_2 \delta_r}. \quad (13)$$

The bistability domain is only exactly given by the simple inequalities $\rho_0 > 1$ and $\rho_1 < 1$ when the ratio δ of g_b mRNA degradation rate over the effective protein dynamics rate is vanishingly small. As shown in Sec. IV and in Fig. 3, for small values of δ , it is more accurately given by

$$\begin{aligned} \rho_0 &> 1 + 2\sqrt{(1 - \rho_1)\delta_B/A_0}, \quad \rho_1 \leq 1, \\ \rho_1 &< 1 - 2\sqrt{(\rho_0 - 1)A_0\delta_A}, \quad \rho_0 \geq 1. \end{aligned} \quad (14)$$

In this intermediate range of production of A proteins, the network reaches one or the other fixed points, depending on its initial condition. The existence of this bistability domain can serve to convert a graded increase in A production in a much more abrupt switchlike response in A (and B) concentration as shown in Fig. 4. The usefulness of this general

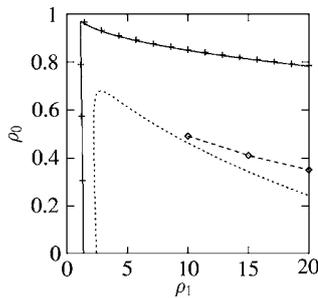


FIG. 5. The oscillation domain for the same parameters as in Fig. 3 but for a larger domain of ρ_1 , the ratio of activated production of B to that of A . As in Fig. 3, the short-dashed line marks the boundary of the domain where the steady state is unstable to oscillations for $\tilde{\theta}=26.6$. Numerical simulations (diamonds) show that the oscillating regime is stable up to the long-dashed line. Therefore, both the limit cycle and the fixed point are stable attractors for $\tilde{\theta}=26.6$ in the region between the short-dashed and long-dashed lines.

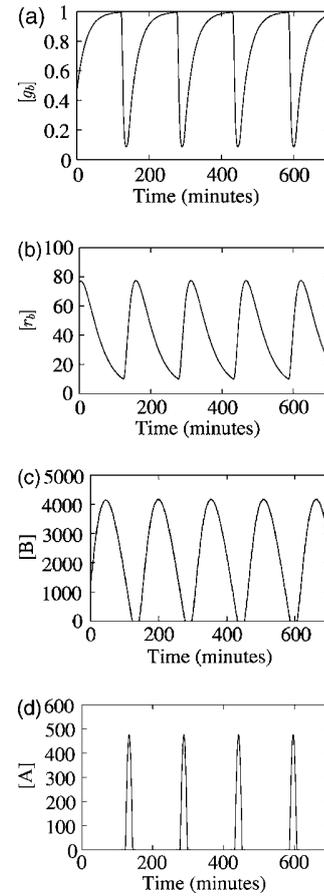


FIG. 6. MFL in the oscillatory regime. The concentrations of the different species are shown as a function of time and display sustained oscillations. Constants are the same as in Fig. 4 except that $\rho_A=100 \text{ mol min}^{-1}$, $\rho_f=0.1 \text{ mol min}^{-1}$, and $\rho_b=5 \text{ mol min}^{-1}$.

feature of multistability has been recently discussed in different contexts [11,12].

C. Transcriptional activation and oscillations

When A is a transcriptional activator, the complexation of B with A acts as a negative feedback and can serve to diminish the variation in B protein concentration when A varies. This leads also to oscillations when A production lies in the intermediate range $\rho_1 > 1 > \rho_0$. This oscillatory behavior mainly depends on the ratios of protein production ρ_0 , ρ_1 and exists for a large range of DNA-protein interaction kinetics. However, a faster kinetics leads to a smaller oscillatory domain as shown in Fig. 3. Oscillations cannot be sustained when $\tilde{\theta}$ becomes large and comparable to $1/\delta$ (with A_0 fixed) or equivalently when $\theta \sim \delta_r/\delta$. It can also be noted that for a sufficiently large activation of transcription by A , there exists a domain of coexistence between oscillations and steady behavior as shown in Fig. 5, i.e., depending on the initial conditions, concentrations will oscillate in time or reach steady levels.

For small δ , oscillations are nonlinear for most parameters. As can be seen in Fig. 6, an oscillation cycle comprises

two phases in succession, a first phase of duration T_1 when the concentration of protein A is high followed by a second phase of duration T_2 where the concentration of B is high. A full oscillation cycle of period $T=T_1+T_2$ proceeds as follows. Let us start with low concentrations of A and B proteins at the beginning of the first phase. When no A is bound to the g_b promoter, the B production rate is lower than the A production rate and complexation cannot prevent the increase of A concentration. When the concentration of A has reached a critical level, A starts to bind to the g_b promoter and activate transcription. This results in a higher production of B than A and the diminution of free A by complexation. Since the A concentration is high, the produced B s are quickly complexed and the concentration of uncomplexed B s remain low. Eventually, the A concentration drops below the binding level and no longer activates the g_b transcription. This marks the transition between the two parts of the oscillation cycle. B continues for a while to be produced from the transcripts, and since few A s are present, this now leads to a rise of the concentration of free B s. Finally, the concentrations of B transcripts and B proteins drop and phase II of the oscillation cycle terminates. A new cycle begins with low concentrations of A and B proteins.

One remarkable feature of the nonlinear oscillations coming from the smallness of the parameter δ is that the concentrations of A and B proteins in their respective high phase depend weakly on the complex association rate γ , as long as it is not too small for the oscillation to exist, and that the period of the oscillations is strikingly insensitive to the exact value of γ . For the parameters of Fig. 6, the oscillation period only changes by about 1% when γ is reduced from 1 min^{-1} to $2 \times 10^{-2} \text{ min}^{-1}$. This remains true even if the formation of the complex AB is not irreversible as is supposed in the present model. We have checked that the case when the complex AB forms with an association rate γ_a dissociates with a rate γ_d and is degraded with a rate δ_{AB} is well described by the present model when one takes the effective rate $\gamma = \gamma_a \delta_{AB} / (\gamma_d + \delta_{AB})$ for the irreversible formation of the complex (as obtained from a quasiequilibrium assumption).

In Secs. IV and V, we provide a more detailed analysis of the MFL different dynamical regimes and derivations of the results here summarized.

IV. THE SWITCH REGIME

We first determine the possible steady states for the MFL. The free gene, mRNA, and B protein concentrations are given in a steady state as a function of the A concentration by,

$$g = \frac{A_0}{A + A_0}, \quad (15)$$

$$r = \frac{\rho_1 A + \rho_0 A_0}{A + A_0}, \quad (16)$$

$$B = \frac{\rho_1 A + \rho_0 A_0}{(A + A_0)(A + \delta d_b)}. \quad (17)$$

The concentration of A itself satisfies the following equation:

$$1 = \delta d_a A + \frac{(\rho_1 A + \rho_0 A_0) A}{(A + A_0)(A + \delta d_b)}. \quad (18)$$

For $\delta=0$, the right-hand side (rhs) goes monotonically from ρ_0 at small A to ρ_1 at large A and a solution, $A_2 = A_0(\rho_0 - 1)/(1 - \rho_1)$ exists only when $\rho_0 < 1 < \rho_1$ or $\rho_1 < 1 < \rho_0$. For small δ , two other steady states are possible. A steady state with a small concentration of A , $A_1 \approx \delta d_b / (\rho_0 - 1)$, exists when $\rho_0 > 1$. Inversely a steady state with a large concentration of A , $A_3 \approx (1 - \rho_1) / (\delta d_a)$ exists when $\rho_1 < 1$. Therefore, Eq. (18) has multiple (i.e., three) fixed points only when A is a transcriptional repressor in the region $\rho_1 < 1 < \rho_0$. This is the parameter regime, which we examine in Secs. IV A–IV C.

A. The high A state

We show that the state with a high concentration of A proteins is stable. The forms of Eqs. (8) and (9) suggest simplifying the analysis by using the fact that the protein quickly reach a quasiequilibrium state. However, both proteins cannot be in quasiequilibrium at the same time. For instance, when proteins $A \gg B$, only B is in quasiequilibrium and A scales as $1/\delta$. When the A concentration is high, in the limit of small δ , we set $a = \delta A$. Substitution in Eqs. (6) and (8) shows that g and B reach on a fast time scale their quasiequilibrium concentration

$$g \approx \delta A_0 / a, \quad (19)$$

$$B \approx \delta r / a. \quad (20)$$

Therefore, the dynamics of the MFL reduces to the following two equations in the limit $\delta \rightarrow 0$:

$$\frac{dr}{dt} \approx \rho_1 - r, \quad (21)$$

$$\frac{da}{dt} \approx 1 - r - d_a a. \quad (22)$$

Equations (21) and (22) clearly show that the high A fixed point is stable and, as found above, satisfies $r \approx \rho_1$, $B \approx g \approx 0$, $a \approx (1 - \rho_1) / d_a$ at the lowest order. This steady state is possible only if $\rho_1 < 1$, that is when the B production rate is not high enough to titrate A and to prevent the repression by A .

B. The high B state

The high B state can be analyzed in a very similar way in the limit of small δ . When the B concentration is high, we define $b = \delta B$. In that case, A quickly reaches its quasiequilibrium concentration,

$$A \approx \delta / b, \quad (23)$$

and the MFL dynamics reduces to the following three equations:

$$\frac{dg}{dt} = \tilde{\theta}(1 - g), \quad (24)$$

$$\frac{dr}{dt} = \rho_0 g + \rho_1(1 - g) - r, \quad (25)$$

$$\frac{db}{dt} = r - 1 - d_b b. \quad (26)$$

The concentrations tend toward those of the high B fixed point $g \simeq 1$, $r \simeq \rho_0$, $b \simeq (\rho_0 - 1)/d_b$, $A \simeq 0$. This steady state exists only if $\rho_0 > 1$, when the production of B proteins is high enough to titrate A proteins and to prevent transcriptional repression by A .

C. Bifurcation between the monostable and bistable regimes

The previous analysis has shown that bistability exists in the domain $\rho_0 > 1 > \rho_1$ when $\delta \ll 1$. We analyze more precisely the nature and position of the bifurcations between the monostable and bistable regimes.

We consider first the transition between the bistable region and the monostable region with a high concentration of A (bottom-left corner in Fig. 3). It follows from Eq. (18) that the low (A_1) and middle (A_2) A concentration fixed points coalesce and disappear when $\rho_0 - 1 = O(\sqrt{\delta})$ and both A_1 and A_2 are of order $\sqrt{\delta}$. For $A \sim \sqrt{\delta}$, Eq. (18) becomes at the dominant order after expansion,

$$\rho_0 - 1 = (1 - \rho_1) \frac{A}{A_0} + \delta \frac{d_b}{A}. \quad (27)$$

For a given repression strength $\rho_1 < 1$, the rhs has a minimum for $A = \sqrt{\delta d_b A_0 / (1 - \rho_1)}$ (that is of order $\sqrt{\delta}$ as assumed) such that

$$\rho_0 = 1 + 2 \sqrt{\frac{\delta d_b (1 - \rho_1)}{A_0}} + O(\delta). \quad (28)$$

Above this value of ρ_0 , Eq. (27) has two solutions and below it has none. Therefore, Eq. (28) marks the saddle-node bifurcation line that separates the bistable domain from the monostable domain with a high A concentration fixed point. For $\delta \rightarrow 0$, the zeroth order boundary $\rho_0 = 1$ is, of course, retrieved.

The approximate expression (28) agrees well with an exact numerical determination of the bifurcation line as shown in Fig. 3.

The transition between the bistable regime and the low A concentration monostable regime (top-right corner of Fig. 3) can be similarly analyzed. The high (A_3) and middle (A_2) A concentration fixed points can coalesce when for both the concentration A is of order $1/\sqrt{\delta}$. Expansion of Eq. (18) for $A \sim 1/\sqrt{\delta}$ gives

$$1 - \rho_1 = (\rho_0 - 1) \frac{A_0}{A} + \delta d_a A. \quad (29)$$

For a given $\rho_0 > 1$, the rhs of Eq. (29) is the minimum for $A = \sqrt{A_0(\rho_0 - 1) / (\delta d_a)}$ and at this minimum, ρ_1 is equal to

$$\rho_1 = 1 - 2 \sqrt{d_a A_0 \delta (\rho_0 - 1)}. \quad (30)$$

For ρ_1 smaller than this value, Eq. (29) has two roots corresponding to A_2 and A_3 . The two roots merge and disappear

on the saddle-node bifurcation line (30) that marks the boundary between the bistable domain and the low A concentration monostable regime. Comparison between the approximate expression (30) and an exact numerical determination of the bifurcation line is shown in Fig. 3.

V. THE OSCILLATOR REGIME

We consider now the case when A is a transcriptional activator, that is when $\rho_0 < \rho_1$. In this case, the MFL has a single steady state since the rhs of Eq. (18) is a monotonically increasing function. However, as we show below, this steady state is unstable in a large domain of parameters and the MFL behaves as an oscillator.

A. The linear oscillatory instability

We begin by analyzing the linear stability of the fixed point (g^*, r^*, B^*, A^*) where A^* is the single solution of Eq. (18) and g^* , r^* , and B^* are given as functions of A^* by Eqs. (15)–(17). After linearization of the MFL dynamics [Eqs. (6)–(9)] around this fixed point, the complex growth rates σ of perturbations growing (or decreasing) exponentially in time like $\exp(\sigma t)$ are found to be the roots of the following characteristic polynomial:

$$\begin{aligned} & \left[\sigma + \tilde{\theta} \left(1 + \frac{A^*}{A_0} \right) \right] [\sigma + 1] \\ & \times \left[\left(\sigma + d_a + \frac{B^*}{\delta} \right) \left(\sigma + d_b + \frac{A^*}{\delta} \right) - \frac{A^* B^*}{\delta^2} \right] \\ & = \frac{g^* \tilde{\theta}}{A_0} \left[\frac{A^*}{\delta^2} (\rho_0 - \rho_1) - \mu \sigma (\sigma + 1) \left(\sigma + d_b + \frac{A^*}{\delta} \right) \right]. \end{aligned} \quad (31)$$

Again, the fact that δ is small simplifies the analysis. We consider first the case when the rescaled concentrations g^* , r^* , B^* , A^* are of order one. This corresponds to the fixed point A_2 of the previous part in the parameter regime $\rho_0 < 1 < \rho_1$, with $A^* \simeq A_0(1 - \rho_0)/(\rho_1 - 1)$, $B^* \simeq 1/A^*$, and $g^* \simeq (\rho_1 - 1)/(\rho_1 - \rho_0)$. Let us assume that the roots σ diverge as δ tends to zero. Then, Eq. (31) reduces at dominant order to

$$\sigma^3 (\delta \sigma + A^* + B^*) = \frac{A^* \tilde{\theta} g^*}{A_0 \delta} (\rho_0 - \rho_1). \quad (32)$$

Therefore, three roots σ_k , $k=0, 1, 2$, are of order $\delta^{-1/3}$ and proportional to the three cubic roots j^k of unity

$$\sigma_k = -j^k \left[\frac{A^*}{A^* + B^*} \frac{g^* \tilde{\theta}}{A_0 \delta} (\rho_1 - \rho_0) \right]^{1/3}, \quad k=0, 1, 2. \quad (33)$$

The fourth root σ_3 is of order $1/\delta$ and corresponds to a stable mode of real part $-(A^* + B^*)/\delta$. The two roots σ_1 and σ_2 are complex conjugates and have a positive real part. Thus, in this parameter domain, the MFL fixed point is oscillatory unstable. The dynamics tends toward an attractive limit cycle that we will analyze in Sec. V B. When A^* or B^* grows (i.e., when $\rho_1 \rightarrow 1$ or $\rho_0 \rightarrow 1$) the fixed point instability disappears

via a Hopf bifurcation. We analyze these two boundaries in turn.

1. A^* high

When $\rho_1 \rightarrow 1$, A^* becomes large and the real parts of the roots become of order one. Thus, Eq. (31) needs to be approximated in a different way. We neglect $\delta(\sigma + d_b)$ and B^* as compared to A^* in Eq. (31) and obtain

$$\left(\sigma + \frac{\tilde{\theta}(\rho_1 - \rho_0)}{\rho_1 - 1}\right)(\sigma + 1)(\sigma + d_a) = -\frac{\tilde{\theta}(\rho_1 - 1)}{\delta A_0}, \quad (34)$$

where we have replaced A^* and g^* by their expressions at the A_2 fixed point. When the rhs is the dominant constant in Eq. (34), its three roots are proportional as above to the three roots of (minus) unity and two of them have positive real parts. On the contrary, when ρ_1 becomes sufficiently close to 1, the rhs of Eq. (34) becomes negligible and Eq. (34) has obviously three real negative roots. Therefore, when $\rho_1 \rightarrow 1$, one traverses the boundary of the linearly unstable region. Two roots of Eq. (34) traverse the imaginary axis on this boundary in a Hopf bifurcation. Assuming (and checking afterward) that, on this boundary, these roots are small compared with $1/(\rho_1 - 1)$, their expressions are found perturbatively by expanding the first factor in Eq. (34),

$$\sigma_{\pm} = -\frac{1 + d_a}{2} + \frac{(\rho_1 - 1)^3}{2(\rho_1 - \rho_0)^2 A_0 \tilde{\theta} \delta} \pm \frac{i}{2} \sqrt{\frac{(\rho_1 - 1)^2}{A_0(\rho_1 - \rho_0) \delta}}. \quad (35)$$

This provides the location of the stability boundary,

$$\rho_1 = 1 + (\delta \tilde{\theta} A_0 (d_a + 1) (\rho_1 - \rho_0)^2)^{1/3}. \quad (36)$$

In the limit $\delta \rightarrow 0$ the zeroth order boundary $\rho_1 = 1$ is of course retrieved. It can be checked that the obtained result $\rho_1 - 1 \sim \delta^{1/3}$ justifies *a posteriori* the use of the A_2 δ -independent expression for the fixed point. This fixed point linear stability boundary (36) can also be directly obtained from the Routh-Hurwitz condition [14] on the third degree polynomial (34). Comparison of the small δ [Eq. (36)] with the exact linear stability boundary is provided in Fig. 3.

2. A^* small

We now consider the upper boundary ($\rho_0 \sim 1$) of the linearly unstable region. When ρ_0 tends toward 1, B^* grows as $1/(\rho_0 - 1)$ and becomes large compared to $\delta(\sigma + d_a)$ and A^* . Equation (31) then simplifies and reduces at leading order to

$$(\sigma + \tilde{\theta})(\sigma + 1)(\sigma + d_b) = -E \quad (37)$$

with

$$E = \frac{A^* \tilde{\theta}}{A_0 B^* \delta} (\rho_1 - 1). \quad (38)$$

Again when E is large the roots are proportional to the three roots of -1 and two have positive real parts. On the contrary,

the three roots are real and negative when E vanishes. There is a critical value E_c such that for $E < E_c$ the fixed point is linearly stable. This occurs when the real part of the two complex conjugate roots becomes negative at the value E_c given by the Routh-Hurwitz criterion [14],

$$E_c = (d_b + 1 + \tilde{\theta})(\tilde{\theta} + d_b + d_b \tilde{\theta}) - d_b \tilde{\theta}. \quad (39)$$

The allied value of A at the fixed point concentration is obtained from Eq. (38),

$$A^* = \sqrt{\frac{E_c A_0 \delta}{\tilde{\theta}(\rho_1 - 1)}}. \quad (40)$$

Using the fixed point Eq. (18), this in turn corresponds to the line in parameter space

$$\rho_0 = 1 - \sqrt{\frac{E_c \delta (\rho_1 - 1)}{A_0 \tilde{\theta}}} \left[1 - \frac{\tilde{\theta} d_b}{E_c} \right]. \quad (41)$$

Equation (41), of course, reduces to the zeroth order boundary $\rho_0 = 1$ when $\delta \rightarrow 0$. Comparison of the small- δ asymptotic expression (41) with the exact linear stability line is provided in Fig. 3.

B. A description of the nonlinear oscillations

The MFL oscillations quickly become strongly nonlinear away from threshold (or, of course, when the bifurcation is subcritical). As there are no autoregulatory direct or indirect positive feedback loops in the MFL, two-variable reductions have a negative divergence. It follows from the well-known Bendixson's criterion [13], that they cannot be used to describe the oscillatory regime. It is, for instance, not possible to properly describe the oscillations by only focusing on the protein dynamics. A specific analysis, therefore, is required and developed in this section.

A full period of the nonlinear oscillations can be split into two parts: a first phase with $A \gg 1$, and a second phase with $B \gg 1$ (see Fig. 7).

In the following, these two main phases of the limit cycle are described. We use simple continuity arguments to match the two phases and obtain a description of the whole limit cycle. The oscillation period is computed (at the lowest order in δ). A full justification of this simple matching procedure and a detailed study of the transition regimes between the two main phases are provided in the Appendix using matched asymptotic techniques.

1. Phase I: High A , Low B phase

This phase is defined as the fraction of the limit cycle where the concentration of protein A is larger than $[A]_0$, i.e., when $A \gg 1 \gg B$. Equation (9) leads us to assume that A is of the order $1/\delta$. So we define $a = \delta A$. In the limit of small δ , as A scales as $1/\delta$, both the binding to the g promoter and the dynamics of B resulting from its complexation with A are fast compared to that of mRNA and A . Consequently, g and B are in quasiequilibrium and obey at the lowest order in δ ,

$$g = \delta A_0 / a, \quad (42)$$

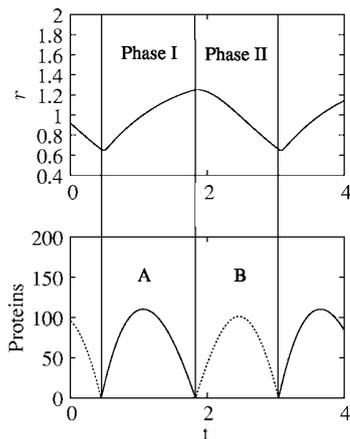


FIG. 7. Distinction between the two phases for the dimensionless equations [Eqs. (6)–(9)]. Top panel: oscillation of r . Bottom panel: oscillations for A (full line) and B (dotted line). In order to clearly depict phases I and II of an oscillation cycle, the parameters are here chosen so that the two phases have similar durations: $\delta = 0.001$, $\rho_1 = 1.5$, $\rho_0 = 0$, $\tilde{\theta} = 2$, $A_0 = 1$, $\mu = 1$, and $d_a = d_b = 0$.

$$B = \delta r/a. \quad (43)$$

The dynamics of the MFL, therefore, reduces to the following two equations at lowest order in δ :

$$\frac{dr}{dt} = \rho_1 - r, \quad (44)$$

$$\frac{da}{dt} = 1 - r - d_a a. \quad (45)$$

The beginning of phase I coincides with the end of phase II where, as explained below, A concentration is small. Therefore, continuity requires that a vanishes at the start of phase I (a detailed study of the transition region between the two phases is provided in the Appendix). Denoting by r_1 , the value of r at the start of phase I, an easy integration of the linear Eqs. (44) and (45) gives,

$$r_I(t) = \rho_1 + (r_1 - \rho_1)e^{-t}, \quad (46)$$

$$a_I(t) = \frac{1 - \rho_1}{d_a} [1 - e^{-d_a t}] + \frac{r_1 - \rho_1}{d_a - 1} [e^{-d_a t} - e^{-t}]. \quad (47)$$

Subscript I has been added to r and a in Eqs. (46) and (47) to emphasize that the corresponding expressions are valid during phase I only. Phase I ends with the fall of A concentration, after a time t_1 such that $a_I(t_1) = 0$. The mRNA concentration is then equal to $r_I(t_1) = r_2$.

One can note that the rise and fall of the concentration of protein A imposes restrictions on the parameters. The rise at the beginning of phase I is possible only if A production dominates over its complexation with B , that is, if $r_1 < 1$ [Eq. (45)]. The following fall requires the reverse which can only happen if B production becomes sufficiently important, namely $r > 1$. This requires $r_2 > 1$ and *a fortiori* $\rho_1 > 1$ [Eq. (44)].

2. Phase II: High B phase

This second phase is defined as the part of the limit cycle where the concentration of protein A falls below $[A]_0$, i.e., $B \gg 1 \gg A$. The form of Eq. (8) leads us to assume that B scales as $1/\delta$ and we define $b = \delta B$.

In the limit of small δ , the dynamics of A is fast compared to that of the other species and A is in quasiequilibrium. At the lowest order, its concentration reads,

$$A \approx \delta/b. \quad (48)$$

Thus, the dynamics of the MFL reduces, at the lowest order in δ , to the following three equations:

$$\frac{dg}{dt} = \tilde{\theta}(1 - g), \quad (49)$$

$$\frac{dr}{dt} = \rho_0 g + \rho_1(1 - g) - r, \quad (50)$$

$$\frac{db}{dt} = r - 1 - d_b b. \quad (51)$$

Continuity with the previous phase I leads us to require that at the beginning of phase II $b = 0$, $r = r_2$, and $g = 0$ (see the Appendix for a detailed justification). With these boundary conditions, the linear Eqs. (49)–(51) can readily be integrated to obtain

$$\begin{aligned} g_{II}(t) &= 1 - e^{-\tilde{\theta}t}, \\ r_{II}(t) &= \rho_0 + \left[r_2 - \rho_0 + \frac{\rho_1 - \rho_0}{\tilde{\theta} - 1} \right] e^{-t} - \frac{\rho_1 - \rho_0}{\tilde{\theta} - 1} e^{-\tilde{\theta}t}, \\ b_{II}(t) &= \frac{\rho_0 - 1}{d_b} [1 - e^{-d_b t}] - \frac{\rho_1 - \rho_0}{\tilde{\theta} - 1} \frac{e^{-d_b t} - e^{-\tilde{\theta}t}}{\tilde{\theta} - d_b} \\ &\quad + \left[r_2 - \rho_0 + \frac{\rho_1 - \rho_0}{\tilde{\theta} - 1} \right] \frac{e^{-d_b t} - e^{-t}}{1 - d_b}. \end{aligned} \quad (52)$$

Since b vanishes at the beginning of phase II, b should start by rising. This imposes that B production dominates over complexation and requires that the concentration r at the beginning of phase II is greater than 1, i.e., $r_2 > 1$ [see Eq. (51)]. Thus, r always remains larger than ρ_0 and would decrease toward ρ_0 for a long enough phase II [Eqs. (50) and (52)]. At the end of phase II, b should decrease to 0 to continuously match the beginning of phase I. This requires that the complexation then dominates over production of B , that is, $r < 1$ [Eq. (51)], and it is only possible when $\rho_0 < 1$. With these conditions met, phase II lasts a time t_2 and ends when $b_{II}(t_2) = 0$.

3. Matching of the two phases and period determination

In order to complete the description of the limit cycle, it remains to determine the four unknowns r_1 , r_2 , t_1 , and t_2 from the four conditions coming from the continuity of proteins and mRNA concentrations,

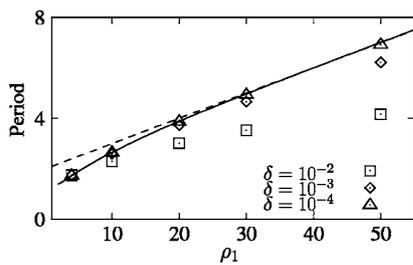


FIG. 8. Comparison between the lowest-order theoretical rescaled period T_r (bold line), the approximate expression for large ρ_1 [Eq. (61)] (dashed line) and numerically computed periods for different values of δ , as a function of parameter ρ_1 . Other parameters are $\tilde{\theta}=10$, $A_0=10$, $\mu=1$, and $d_a=d_b=0$.

$$r_I(t_1) = r_2, \quad a_I(t_1) = 0, \quad (53)$$

$$r_{II}(t_2) = r_1, \quad b_{II}(t_2) = 0. \quad (54)$$

One possibility to solve these equations is to use Eq. (53) to express r_1 and r_2 as a function of t_1 . It is then not difficult to see that the third equation [(54)] implicitly determines t_2 as a function of t_1 . Once r_1 , r_2 , and t_2 are obtained as functions of t_1 , the last equation $b(t_2)=0$ can be solved for t_1 by a one-dimensional root finding algorithm. In this way, we have determined r_1 , r_2 , t_1 , t_2 for different sets of kinetic constants and obtained the rescaled period of oscillation,

$$T_r = t_1 + t_2, \quad (55)$$

the dimensionful period being $T = T_r / \delta_r$. Results are shown in Fig. 8 and Fig. 9. The analytical period compares well to results obtained by direct numerical integration of Eqs. (6)–(9) for different values of δ with, of course, a closer agreement for smaller δ .

Equations (53) and (54) are difficult to solve analytically but analytic expressions can be obtained for various limiting cases. For instance, if the degradation of protein B is negligible ($d_b=0$), Eq. (52) simplifies to

$$b_{II}(t) = (\rho_0 - 1)t - \frac{\rho_1 - \rho_0}{(\tilde{\theta} - 1)\tilde{\theta}}(1 - e^{-\tilde{\theta}t}) + \left[r_2 - \rho_0 + \frac{\rho_1 - \rho_0}{\tilde{\theta} - 1} \right] (1 - e^{-t}). \quad (56)$$

If we further consider the limit of large ρ_1 , the condition $b_{II}(t_2)=0$ gives the following estimate for the duration of phase II, up to exponentially small terms,

$$t_2 \approx \frac{r_2 - \rho_0}{1 - \rho_0} + \frac{\rho_1 - \rho_0}{\tilde{\theta}(1 - \rho_0)}, \quad (57)$$

as well as the transcript concentration at the end of phase II

$$r_1 \approx \rho_0. \quad (58)$$

Similarly, the condition $a_I(t_1)=0$ together with Eq. (47) show that t_1 is small for large ρ_1 ,

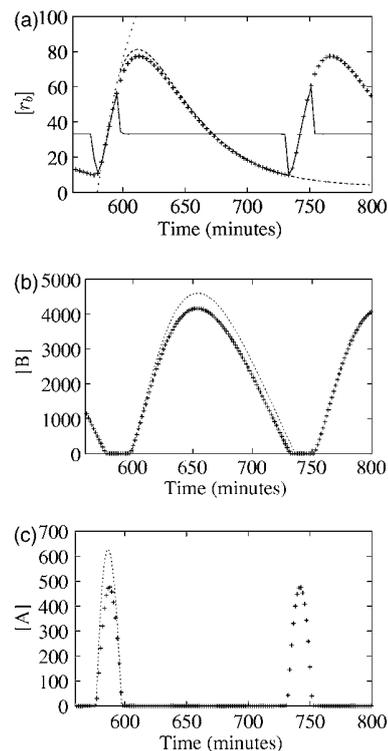


FIG. 9. Comparison between the full dynamics (+) symbols and the asymptotic description at order $\sqrt{\delta}$ (dashed and dotted lines) during one oscillation cycle. (a) Concentration of mRNA behavior. Phase I asymptotics (dotted lines) and phase II asymptotics (dashed lines) are shown separately. The product AB/β is also plotted (full line). The quasistatic assumption is seen to be well-satisfied away from the transition regions between the two phases. (b) Concentration of protein B . (c) Concentration of protein A . Parameters are as in Fig. 6.

$$t_1 \approx 2(1 - r_1)/\rho_1 \approx 2(1 - \rho_0)/\rho_1, \quad (59)$$

where expression (58) has been used in the second equality. Given the duration (59) of phase I, the concentration r_2 of the transcript at its end is directly obtained from Eqs. (46) and (58)

$$r_2 \approx 2 - \rho_0. \quad (60)$$

This finally provides the estimate of the period for large ρ_1 (with $d_b \approx 0$),

$$T_r \approx t_2 \approx 2 + \frac{\rho_1 - \rho_0}{\tilde{\theta}(1 - \rho_0)}. \quad (61)$$

A comparison between Eq. (61) and numerically determined oscillation periods is provided in Fig. 8

C. Comparison with a simple delayed negative feedback

Popular models for genetic oscillators consist in a protein repressing its own production with a phenomenological delay [15–17]. Using a Hill function to model this repression, the simplest model of this kind reads,

$$\frac{dB(t)}{dt} = \frac{\rho}{1 + [B(t - \tau)/B_0]^n} - cB, \quad (62)$$

where ρ is the protein production rate, c the protein degradation rate, and τ the phenomenological delay for repression. The phase diagram of this simple oscillator can be computed [18]. In the limit of long delays ($c\tau \gg 1$), the oscillations are nonlinear. In an expansion in the small parameter $\epsilon = 1/c\tau$, their period T is

$$T/\tau = 2 + \frac{\kappa}{c\tau} + \dots, \quad (63)$$

where κ is a constant which is approximately equal to 2 for $n=2$ and ρ/c of order one [19].

It is interesting to note some analogies between this simple model with delay and the previously studied MFL. When $\tau \gg 1/c$, the period of the delayed model scales as τ while the period of the MFL scales as δ_r^{-1} . The mRNA half-life in the MFL thus plays the role of the phenomenological delay in Eq. (62). This is in line with the mRNA being the major pacemaker of the MFL oscillator. The rescaled parameter ϵ in the simple model with delay plays a role similar to the rescaled parameter δ in the MFL in the sense that, first, for high values of ϵ , oscillations disappear [18], and for small values of this parameter, the period of the oscillations is independent of $1/c\tau$ at dominant order. In the delayed model, ϵ represents the ratio between the typical lifetime of protein over the delay, while in the MFL, δ represents the ratio between the typical time scale of protein production and sequestration over the typical lifetime of RNA. These analogies show that for the MFL $\sqrt{\rho_A \gamma}$ plays the role of the protein degradation constant c in the delayed model. Thus protein sequestration by complexation is the MFL analog of protein degradation in the simple model with delay. The fastness of both these processes relative to a slow mechanism (delay τ , RNA dynamics) ensures the flipping of both oscillators between two states (two concentrations of B for the delayed model, A high and B high for the MFL).

VI. DISCUSSION

We have proposed a simple model of the mixed feedback loop, an over-represented motif in different genomes. We have shown that by itself this motif can serve as a bistable switch or can generate oscillations.

A. Importance of dimerization and of RNA dynamics

The MFL dynamics crucially depends on the post-transcriptional interaction between A and B proteins. The dynamics of dimerization is fast and allows a high concentration of only one of the two proteins, thus effectively creating a dynamical switch between the two species. If, for instance, A proteins are present at a higher concentration than B proteins, all B s are quickly titrated and there only remains free A and AB dimers.

A bistable system is obtained by coupling this post-transcriptional mechanism to a transcriptional repression. The created bistable switch is quite different from the clas-

sical ‘‘toggle switch’’ [20], which is based on reciprocal transcriptional repressions between two genes. In the MFL model, the single transcriptional repression of gene g_b is sufficient for a working switch. Moreover, dimerization bypasses the need for cooperativity in the toggle switch [21] and may render the present module simpler to implement in a biological system.

When the dimerization between A and B is coupled to a transcriptional activation, the system behaves as an oscillator if the production of A proteins is intermediate between the two possible for B . Then, the level of g_b transcripts controls A protein dimerization and, of course, B production rates. At the beginning of the derepression phase (phase I), the low level of transcripts leads to a small production of B proteins and to a rise of A concentration. This first phase ends when the subsequent high production and accumulation of g_b transcripts give rise to a production of B proteins sufficient for the titration of all free A s. During the ensuing repression phase (phase II), the g_b transcription rate is low and the transcripts degradation produces a continuous decrease of their concentration. The parallel decrease of the B production rate ultimately leads to the end of this second phase when the B production rate is no longer sufficient for the titration of all produced free A s.

In summary, the core mechanism of the clock is built by coupling the rapid switch at the protein level to the slow mRNA dynamics. The protein switch in turn controls mRNA dynamics via transcriptional activation.

It is interesting to note that the MFL model in the simple form analyzed here provides a genetic network that oscillates without the need for several specific features previously considered in the literature for related networks. As already mentioned, oscillations require neither an additional positive feedback loop [22], nor the self-activation of gene A [23], nor a highly cooperative binding [24] of transcription factors to gene promoters. Similarly, an explicit delay is not needed although delays of various origins have been considered in models of oscillatory genetic networks in different contexts, as is further discussed below.

Besides the motif topology, one feature of the present MFL model that renders this possible is the explicit description of transcription and translation. The condensed representation of protein production by a single effective process that is often used, would necessitate supplementary interactions to make oscillations possible. The description of mRNA was for instance omitted in the first version of a previously proposed evolutionary algorithm [5]. Oscillatory networks with core MFL motifs were nonetheless produced but much less easily than in a refined second version that incorporates mRNA dynamics. This role of mRNA dynamics may well be worth bearing in mind when considering the elaborate regulation of translation that is presently being uncovered [25].

B. Biological MFL oscillators and switches

Given the proposed functions for the MFL motif, it is of primary interest to know whether cases of its use as an oscillator or a switch are already documented. The oscillator function seems the clearest. Circadian clocks are genetic os-

TABLE I. Some biological examples of MFL motifs. The yeast motifs are reproduced from Ref. [3]. An annotation taken from the SGD database [32] has been added.

A species	B species	Main biological functions
Bistable systems		
Lac Repressor	Allolactose	Lactose metabolism
Oscillators		
WC-1, WC-2	FRQ	Neurospora circadian clock
dCLK	PER, TIM	Drosophila circadian clock
CLOCK, BMAL	PER, CRY	Mammals circadian clock
p53	Mdm2	Stress response oscillators
Detected Yeast MFLs		
Swi6	Swi4	G1/S transition
Gal4	Gal80	Galactose metabolism
Gal4	Gal3	Galactose metabolism
Gal4	Gal1	Galactose metabolism
Ime1	Rim11	Meiose activation
Ume6	Ime1	Meiose activation
Ste12	Fus3	Pseudohyphal growth
Ste12	Far1	Pseudohyphal growth
Cbf1	Met28	Sulfur metabolism
Met4	Met28	Sulfur metabolism
Swi4	Clb2	Cell cycle
Mbp1	Clb5	G1/S transition
Stb1	Cln1	
Stb1	Cln2	

cillators which generate endogenous rhythms with a period close to 24 h and which are locked to an exact 24 h period by the alternation of day and night. A MFL motif is found at the core of all known eucaryotic circadian oscillators. Taking, for illustrative purposes, the fungus *Neurospora crassa*, one of the best studied model organisms, it has been established that a dimer of white-collar proteins (WCC) plays the role of *A* and activates the transcription of the *frequency* gene, the gene g_b in this example. In turn, the frequency (FRQ) protein interacts with WCC and prevents this activation [26]. Analogous motifs are found in the circadian genetic networks of flies and mammals, as shown in Table I. The p53/Mdm2 module [27] provides a different example. The p53 protein is a key tumor suppressor protein and an important role is played in its regulation by the Mdm2 protein. On the one hand, p53, similar to the *A* protein, binds to the Mdm2 gene and activates its transcription. On the other hand, the Mdm2 protein, as the *B* protein, binds to p53 and both blocks transcriptional activation by p53 and promotes its rapid proteolytic degradation. Cells exposed to stress have indeed been observed to present oscillations in both p53 and Mdm2 levels [27]. In both these examples, many other genetic interactions exist besides the above described MFL motifs and have been thought to provide delays necessary for oscillations. These have been postulated to come from chains of phosphorylations in circadian networks [28] or from a yet

unknown intermediate species [27] in the p53-Mdm2 system. The realization that the MFL can lead to oscillations without further interactions may be useful in suggesting alternative models of these particular genetic networks or in reassessing the role of known interactions. For instance, this has led one of us to formulate a model of the *Neurospora crassa* circadian clock [29] in which kinases and phosphorylation influence the cycle period by modifying protein degradation rates but in which phosphorylation is not needed to create key delays. The prevalence of the MFL motif probably arises from the usefulness of negative feedback, but it may also imply that oscillations in the genetic network are more common than is usually thought.

The evidence for uses of the MFL motif as a switch appears less clear cut at present. The classic case of the *E. coli* lactose operon [30] can be thought of as an effective version of a bistable MFL. The lac repressor represses the lac gene and plays the role of *A* in the MFL. The lac gene directs the production of a membrane permease, which itself drives the absorption of external lactose. Since allolactose binding to the lac repressor blocks its transcriptional activity, allolactose effectively plays the same role as *B* in the MFL. The search for a more direct switch example has led us to consider the different MFL motifs in yeast as reported in [3] and reproduced in Table I. It was noted in [3] that most of these MFL motifs were central modules in biochemical pathways. Most

of them take part in the cell cycle or in differentiation pathways, which certainly are processes where the cell switches from one function to another. However, we have found it difficult to disentangle the MFLs from numerous other known interactions and to confidently conclude that any of these motifs implements the proposed switch function. Similar difficulties have been recently emphasized for general motifs determined on purely statistical grounds and have led to the questioning of their functional significance [31]. In the present case, the difficulty of identifying biological cases of MFL switches, of course, arises from our own very partial knowledge of the networks listed in Table I, but it may also be due to the fact that the possible role of the MFL module as a switch was not fully realized in previous investigations. The present study will hopefully help and trigger direct experimental investigations of these questions.

APPENDIX: TRANSITIONS BETWEEN THE TWO PHASES OF AN OSCILLATION CYCLE AND $\sqrt{\delta}$ CORRECTION TO THE PERIOD

In the following, we analyze the transitions between phase I and phase II of an oscillation period and use matched asymptotics to precisely justify the assumptions made in Sec. V. This also provides the leading order ($\sqrt{\delta}$) correction to the zeroth order results of Sec. V.

1. From phase I to phase II

We consider the transition between the end of the high A /low B phase I and the start of the high B /low A phase II. This protein switch occurs on a time scale of order δ/δ_r . On this fast time scale the mRNA concentration does not have time to change. Thus, in Eqs. (6)–(9), we assume $r=r_2$ and introduce $\tau=t/\delta$. Equations (8) and (9) simply become at dominant order,

$$\frac{dB}{d\tau} = r_2 - AB, \tag{A1}$$

$$\frac{dA}{d\tau} = 1 - AB. \tag{A2}$$

The imposed boundary conditions are

$$A \rightarrow (1 - r_2)(\tau - \tau_1) + o(1), \quad B \rightarrow 0 \text{ at } \tau = -\infty, \tag{A3}$$

$$B \rightarrow (r_2 - 1)(\tau - \tau_2) + o(1), \quad A \rightarrow 0 \text{ at } \tau \rightarrow +\infty \tag{A4}$$

with the dimensionless transcript concentration $r_2 > 1$ (see Sec. V B 1) and τ_1 and τ_2 two constants to be determined. Equations (A1) and (A2) are autonomous in time and, therefore, invariant by time translation. For general nonintegrable equations, numerical integration would be required to obtain the difference $\tau_2 - \tau_1$. Here, however, the difference of protein concentrations, $A - B$, is easily integrated across the transition region. Fixing the time origin at the instant when $A = B$, one readily obtains the exact formula

$$A - B = (1 - r_2)\tau. \tag{A5}$$

Comparison with Eqs. (A3) and (A4) shows that with this choice of time origin $\tau_1 = \tau_2 = 0$ (and more generally $\tau_1 = \tau_2$).

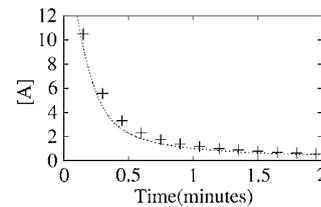


FIG. 10. The evolution of A given by the Riccati Eq. (A6) (+) is compared to that given by the complete MFL dynamics (dashed) during the transition from phase I to phase II. Parameters are as in Fig. 6.

The asymptotic conditions (A3) and (A4) coincide with the limiting behavior of $a_I(t)/\delta$ at the end of phase I near $t=t_1$, and with the limiting behavior of $b_{II}(t)$ at the beginning of phase II near $t=0$, provided that $a_I(t_1) = b_{II}(0) = 0$ as was required in the main text. Thus, Eqs. (A1) and (A2) provide a uniform approximation of $A(t)$ and $B(t)$ throughout the transition from phase I to phase II. Given the exact result (A5), the integration of Eqs. (A1) and (A2) can in fact be replaced to the integration of the single following Riccati equation:

$$\frac{dA}{d\tau} = 1 - [A + (r_2 - 1)\tau]A. \tag{A6}$$

A comparison between Eq. (A6) and full numerical evolution is shown in Fig. 10. There is one subtlety, however. The quasiequilibrium approximation for g [Eq. (42)] diverges at the end of phase I when $A \rightarrow 0$ and g becomes larger than 1 before the transition region (A1) and (A2) of order δ . This clearly signals the breakdown of the approximation (42) in a larger intermediate region at the end of phase I. In a region of size $t_1 - t \sim \sqrt{\delta}$, the evolution of g reduces to

$$\frac{dg}{dt} = \tilde{\theta} \left(1 - g \frac{a}{\delta A_0} \right). \tag{A7}$$

Indeed, taking $a = (1 - r_2)(t_1 - t) + o(\sqrt{\delta})$ shows that all three terms kept in Eq. (A7) are of the same magnitude when $g \sim t_1 - t \sim \sqrt{\delta}$ and that the additional term θg in Eq. (6) is negligible. Explicit integration in this intermediate region gives

$$g(t_1) - g(t) \exp[-\kappa_2(t_1 - t)^2] = \tilde{\theta} \int_{t-t_1}^0 du \exp(\kappa_2 u^2) \tag{A8}$$

where $\kappa_2 = \tilde{\theta}(1 - r_2)/(2A_0\delta)$. In particular, the limit $t \rightarrow -\infty$ determines the value of g at the end of phase I and beginning of phase II

$$g(t_1) = g_2 = \sqrt{\frac{\pi A_0 \tilde{\theta} \delta}{2(r_2 - 1)}}. \tag{A9}$$

Equation (A9) is the source of a correction of order $\sqrt{\delta}$ to the asymptotic result (55) as will be shown below.

2. From phase II to phase I

The transition from phase II to phase I can be analyzed quite similarly to the transition from phase I to phase II. In a time of order δ/δ_r at the end of phase II, r has no time to change and the transition is described by Eqs. (A1) and (A2) with r_2 replaced by $r_1 < 1$ (see Sec. V B 1) and the boundary conditions

$$B = (r_1 - 1)(\tau - \tau'_1) + o(1), \quad A \rightarrow 0 \text{ at } \tau = -\infty, \quad (\text{A10})$$

$$A = (1 - r_1)(\tau - \tau'_2) + o(1), \quad B \rightarrow 0 \text{ at } \tau = +\infty. \quad (\text{A11})$$

Again, $\tau'_1 = \tau'_2 = 0$ if the time origin is taken when $A = B$. This transition regime is followed by a longer transition on a time scale $t - t_2 \sim \sqrt{\delta}$ where g decreases from a value of order one to values of order δ characteristic of phase I. This reflects the fact that the binding of protein A to the promoter of gene g_b is not instantaneous and requires some accumulation of protein A at the beginning of phase I. Taking $A = (1 - r_1)t/\delta$, g evolution is described in this intermediate regime by

$$\frac{dg}{dt} = \tilde{\theta} \left[1 - t(1 - r_1) \frac{g}{A_0 \delta} \right]. \quad (\text{A12})$$

The corresponding evolution of g is

$$g(t) = \exp(-\kappa_1 t^2) \left[g_1 + \tilde{\theta} \int_0^t du \exp(\kappa_1 u^2) \right], \quad (\text{A13})$$

where $\kappa_1 = \tilde{\theta}(1 - r_1)/(2A_0\delta)$. It indeed describes the transition from $g_1 = 1 - \exp(-\tilde{\theta}t_2)$, the value of g at the end of phase II, to the quasiequilibrium regime since the asymptotic behavior of Eq. (A13) for $t \gg \sqrt{\delta}$ is

$$g(t) \sim \frac{A_0 \delta}{(1 - r_1)t}. \quad (\text{A14})$$

3. Dominant correction to the oscillation period

The form of Eqs. (6)–(9) could lead one to think that the oscillation period has an expansion in powers of the small parameter δ . However, here as often, boundary layers lead to more complicated expansions. The two transition regimes of duration $\sqrt{\delta}$ give rise to a dominant correction of order $\sqrt{\delta}$ to the oscillation period.

As pointed out above, due to the transition regime at the end of phase I, the starting value of g in phase II is g_2 [Eq. (A9)] of order $\sqrt{\delta}$. Therefore, the first Eq. (52) is replaced at order $\sqrt{\delta}$ by the corrected evolution,

$$g_{II,c}(t) = 1 - (1 - g_2)\exp(-\tilde{\theta}t). \quad (\text{A15})$$

This leads in turn to corrected evolutions $r_{II,c}(t)$ and $b_{II,c}(t)$ for the mRNA and B protein concentrations that are obtained by simply replacing $(\rho_1 - \rho_0)$ by $(\rho_1 - \rho_0)(1 - g_2)$ in the expressions (52) of the zeroth-order expressions $r_{II}(t)$ and $b_{II}(t)$.

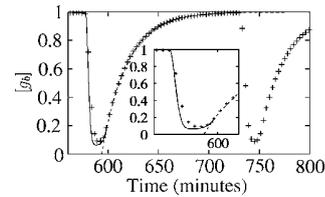


FIG. 11. Comparison of $[g_b]$ obtained from a full numerical integration (+) with the uniform approximation obtained by matching the different transition regimes (bold line). The exponential relaxation in phase II [Eq. (A15)] is also shown (dashed). An enlargement of phase I and the transition regimes is shown in the inset. Parameters are as in Fig. 6.

The other transition regime between phase II and phase I does not lead to modifications of the form of Eqs. (46) and (47) describing the evolutions of mRNA and A protein during the bulk of phase I. Instead, it results in a difference of order $\sqrt{\delta}$ between $r_{II,c}(t_2)$ and r_1 the effective concentrations of transcripts at the start of phase I. The transcript concentration depends on the promoter states at previous times. So, integrating Eq. (7) from the end of phase II, one obtains

$$r(t) = \rho_1 + [r_{II,c}(t_2) - \rho_1]\exp(-t) + (\rho_0 - \rho_1)\exp(-t) \int_0^t du \exp(u)g(u). \quad (\text{A16})$$

Replacing $g(t)$ by its expression (A12) during the transition regime at the start of phase I leads for $t \gg \sqrt{\delta}$ to

$$r(t) = \rho_1 + \left[r_{II,c}(t_2) - g_1 \sqrt{\frac{\pi}{4\kappa_1}} (\rho_1 - \rho_0) - \rho_1 \right] \exp(-t). \quad (\text{A17})$$

Namely, the previous Eq. (46) but with

$$r_1 = r_{II,c}(t_2) - g_1 \sqrt{\frac{\pi}{4\kappa_1}} (\rho_1 - \rho_0). \quad (\text{A18})$$

Note that only the g_1 term between the square brackets in Eq. (A12) contributes to Eq. (A17), the integral term giving a subdominant contribution.

Finally, the durations T_1 and T_2 of both phases and the total period $T_r = T_1 + T_2$ are obtained to $\sqrt{\delta}$ accuracy by solving as before

$$r_I(t_1) = r_2, \quad a_I(t_1) = 0 \quad (\text{A19})$$

together with the corrected version of Eq. (54), namely Eq. (A18) and $b_{II,c}(t_2) = 0$.

The asymptotic description of g_b evolution agrees well with numerical simulations of the full MFL model as shown in Fig. 11. The $\sqrt{\delta}$ correction terms lengthen the zeroth-order asymptotic estimation. For a moderate value of δ , a numerically comparable contribution, however, is coming from higher-order terms that tend to lengthen the period [coming, for instance, from the breakdown of the quasiequilibrium approximation for g_b Eq. (42)] as can be seen in Fig. 8.

- [1] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 763 (2002).
- [2] S. Mangan and U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11980 (2003).
- [3] E. Yeger-Lotem and H. Margalit, *Nucleic Acids Res.* **31**, 6053 (2003).
- [4] E. Yeger-Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R. Y. Pinter, U. Alon, and H. Margalit, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5934 (2004).
- [5] P. François and V. Hakim, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 580 (2004).
- [6] These assumptions simplify the analysis but do not play a crucial role. For instance, it was checked that dissociation of the complex with a rate γ_d can be accurately described with the present model by using an effective association rate γ as briefly discussed at the end of Sec. III C.
- [7] Y. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5860 (2002).
- [8] M. H. Glickman and A. Ciechanover, *Physiol. Rev.* **82**, 373 (2002).
- [9] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland Science, New York, 2004).
- [10] M. B. Elowitz, M. G. Surette, P. E. Wolf, J. B. Stock, and S. Leibler, *J. Bacteriol.* **181**, 197 (1999).
- [11] J. E. J. Ferrell, *Curr. Opin. Cell Biol.* **14**, 140 (2002).
- [12] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. V. Oudenaarden, *Nature (London)* **427**, 737 (2004).
- [13] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer-Verlag, New York, 1986).
- [14] All roots have a negative real part for the third degree polynomial equation with positive coefficients, $X^3 + b_1X^2 + b_2X + b_3 = 0$ if and only if $b_1b_2 - b_3 > 0$.
- [15] M. H. Jensen, K. Sneppen, and G. Tiana, *FEBS Lett.* **541**, 176 (2003).
- [16] J. Lewis, *Curr. Biol.* **13**, 1398 (2003).
- [17] N. A. Monk, *Curr. Biol.* **13**, 1409 (2003).
- [18] L. Glass and M. C. Mackey, *From Clocks to Chaos: The Rhythms of Life* (Princeton University Press, Princeton, NJ, 1999).
- [19] It is ingeniously argued in Ref. [16] that κ is exactly equal to 2, independently of n . The argument unfortunately appears erroneous. This can be explicitly checked in the limit $n \rightarrow +\infty$ where the repression is given by a simple step function and the limit cycle can be determined as well as $\kappa = \ln[r^2/(r-1)]$ with $r = \rho/(B_0c) > 1$.
- [20] T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature (London)* **403**, 339 (2000).
- [21] J. L. Cherry and F. R. Adler, *J. Theor. Biol.* **203**, 117 (2000).
- [22] J. J. Tyson, C. I. Hong, C. D. Thron, and B. Novak, *Biophys. J.* **77**, 2411 (1999).
- [23] J. M. Vilar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5988 (2002).
- [24] P. Ruoff and L. Rensing, *J. Theor. Biol.* **179**, 275 (1996).
- [25] J. D. Richter and N. Sonenberg, *Nature (London)* **477**, 477 (2005).
- [26] J. J. Loros and J. C. Dunlap, *Annu. Rev. Physiol.* **63**, 757 (2001).
- [27] R. L. Bar-Or *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11250 (2000).
- [28] J. C. Leloup and A. Goldbeter, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7051 (2003).
- [29] P. François, *Biophys. J.* **88**, 2369 (2005).
- [30] J. Monod and F. Jacob, *Cold Spring Harbor Symp. Quant. Biol.* **26**, 389 (1961).
- [31] A. Mazurie, S. Bottani, and M. Vergassola, *Genome Biology* **6**, R35 (2005).
- [32] K. Dolinski *et al.*, *Saccharomyces Genome Database*, <http://www.yeastgenome.org/> (2004).

Chapitre IV

À propos des horloges circadiennes

LES HORLOGES circadiennes sont des exemples classiques et fascinants d'oscillateurs génétiques [17, 86]. Dans cette partie, je présente différents résultats concernant les horloges circadiennes. Je présenterai tout d'abord quelques définitions et propriétés générales concernant ces horloges, puis présenterai quelques modèles d'horloges circadiennes. Enfin, je présenterai et résumerai nos résultats présentés dans la référence [27].

A Horloges circadiennes

De nombreux processus physiologiques sont organisés périodiquement dans le temps et sont couplés à différents rythmes naturels (rythme jour/nuit, rythme des saisons...). Ces rythmes physiologiques peuvent être la conséquence des stimuli périodiques, comme par exemple dans le cas des bancs de poissons dont la position dans l'océan varie en fonction de la luminosité [86]. Dans d'autres cas, ces rythmes sont au contraire endogènes, i.e. qu'ils persistent en l'absence de stimulus initial. Il existe donc des mécanismes biologiques pour générer de tels rythmes. D'un point de vue très général, les questions soulevées par ces rythmes biologiques endogènes sont fascinantes : d'une part ces processus doivent être suffisamment robustes pour "résister" aux différentes modifications de l'environnement auxquelles sont soumis tous les organismes vivants, mais d'autre part, ces processus doivent être suffisamment souples pour pouvoir être entraînés, voire contrôlés par les rythmes naturels auxquels ils sont couplés. L'un des exemples les plus connus est l'exemple du rythme circadien.

A.1 Quelques définitions

Dans la plupart des organismes eucaryotes¹, des oscillateurs spécifiques sont donc chargés de synchroniser l'organisme au rythme du jour et de la nuit. Ces oscillateurs sont appelés horloges "circadiennes" (du latin "circum die", autour du jour). De telles horloges ont été observées dès le XVIIIème siècle. Ainsi, en 1723, Jean-Jacques Ortoùs de Mairan est l'un des premiers à avoir réalisé une expérience désormais classique. Intrigué par l'ouverture périodique d'une plante héliotrope, la sensitive (*mimosa obscura*), il place cette plante en obscurité constante et observe que le rythme d'ouverture et de fermeture se poursuit à l'identique, mettant ainsi en évidence le caractère endogène de l'oscillation.

Les horloges circadiennes ont ainsi comme propriétés remarquables [86] :

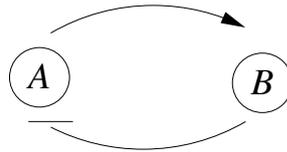
- Les horloges circadiennes ont des périodes d'environ 24 h. Ces horloges sont autonomes dans la mesure où les oscillations circadiennes persistent dans des conditions de température et d'obscurité constantes.
- La période de l'horloge circadienne ne dépend pas (ou très peu) de la température tant que la température demeure constante.
- Le rythme circadien est présent et persiste dans la plus petite unité biologique : la cellule
- Le rythme circadien persiste même lorsque la cellule est soumise à des perturbations chimiques importantes.
- Les horloges circadiennes peuvent être entraînées par des stimuli extérieurs. En particulier, elles peuvent être entraînées par la lumière. Notons qu'en général, les organismes sont sensibles à une longueur d'onde donnée dans le visible (par exemple le bleu pour *Neurospora*)
- Enfin, l'horloge en elle-même est malgré tout sensible à la température dans la mesure où elle peut-être entraînée par des variations de température. Il existe donc des phénomènes de "compensation" à l'intérieur de l'horloge : la période de l'horloge est indépendante de la température mais l'horloge en elle-même est sensible à la température. Cette propriété est appelée "compensation de température".

A.2 Réseaux génétiques des horloges circadiennes

Les progrès récents dans la connaissance des réseaux génétiques ont permis de mettre en évidence les mécanismes à la base des horloges circadiennes dans des

¹et en fait même chez certains procaryotes comme *Synechococcus elongatus* par exemple.

Organisme	Activateurs A	Répresseurs B
Neurospora Crassa	WC-1, WC-2	FRQ
Drosophile	dCLK	PER, TIM
Mammifères	CLOCK, BMAL	PER, CRY



TAB. IV.1: : Protéines impliquées dans différentes horloges circadiennes

organismes très différents (champignons, drosophile, mammifères) [17, 68, 69, 92]. Il se trouve que le principe de fonctionnement de base de ces horloges est identique dans la plupart des horloges connues : des modules génétiques particuliers, impliquant parfois des protéines homologues, sont à l'origine de ces oscillations au niveau cellulaire.

Ainsi, dans tous ces oscillateurs génétiques, on distingue deux groupes de protéines, impliquées dans une seule boucle de rétroaction principale. Le premier groupe est un groupe de protéines activatrices. Ces protéines sont des facteurs de transcription qui vont se lier à l'ADN pour activer la transcription des protéines du second groupe. Ce second groupe de protéines est un groupe de protéines répressives pour le premier groupe de protéines. Cette répression est toujours un mécanisme post-transcriptionnel : les protéines répressives vont réagir dans le noyau avec les protéines du premier groupe pour former des multimères incapables d'activer la transcription. Le circuit génétique de base des horloges circadiennes est donc tout à fait similaire au circuit modélisé dans le chapitre III.

A côté de cette boucle de rétroaction principale, d'autres boucles sont présentes, dont le rôle précis n'est pas toujours bien défini. À titre d'exemple, nous examinons maintenant l'exemple de l'horloge circadienne de *Neurospora*.

A.3 Un exemple : l'horloge de *Neurospora crassa*

Chez *Neurospora crassa*, il a été démontré que le rythme de sporulation asexuée lors de la croissance végétative est en fait dirigé par une telle horloge circadienne [66]. Au moins trois protéines sont impliquées dans ce rythme circadien : la protéine FREQUENCY (FRQ), qui est la protéine répressive, et les protéines White Collar (WC1 et WC2), protéines activatrice. Je résume ici les principaux faits expérimentaux concernant cet oscillateur, pour une revue détaillée, voir [53].

frq est le premier gène à avoir été identifié dans cette horloge circadienne [22]. Garceau et al. [30] ont décrit les oscillations de la protéine FRQ et des transcrits *frq*, ainsi que le rôle de certaines régulations post-transcriptionnelles. En obscurité constante, la protéine FRQ et l'ARN associé ont des dynamiques déphasées. Le pic de concentration d'ARN précède de quatre à six heures le maximum de concentration de protéines FRQ. En fait, Merrow et al. ont par la suite démontré que l'oscillation de FRQ peut-être divisée en deux phases, apparemment décorréllées [59]. La première phase est la phase de répression proprement dite durant laquelle FRQ réprime sa propre transcription, si bien que la concentration de transcrits *frq* baisse régulièrement (voir plus bas). Cette phase est la plus longue, de 14 à 18 heures chez les souches sauvages et dure aussi longtemps que la protéine FRQ est à niveau élevé. Lorsque la concentration de la protéine FRQ atteint un niveau suffisamment bas, le processus d'activation de la protéine FRQ est relancé et on entre dans la phase dite de "dérépression" durant laquelle le niveau d'ARN de FRQ croît continuellement. Cette phase dure environ 4 heures.

Par la suite, il a été démontré que deux protéines FRQ peuvent interagir entre elles, et que cette interaction est nécessaire aux oscillations circadiennes [7]. Cette interaction semble précéder la dimérisation entre FRQ et WCC, mais son rôle précis est encore inconnu à ce jour.

La dynamique des facteurs de transcription White Collar (WC-1 et WC-2) est moins bien connue. WC-2 est une protéine exprimée en large excès par rapport à WC-1 et FRQ, et dont la concentration semble rester constante tout au long du cycle [14]. WC-1 au contraire est régulée par FRQ (voir ci-dessous). WC-1 et WC-2 peuvent interagir pour former un complexe, appelé WCC (White Collar Complex). C'est ce complexe qui joue le rôle d'activateur dans le cycle circadien. Ce complexe se lie au promoteur du gène *frq* activant la transcription. Cette interaction protéine-ADN est cyclique et l'interaction entre WCC et le promoteur est maximale environ six heures avant le maximum de concentration de FRQ. Puis la protéine FRQ interagit avec le dimère WCC dans le noyau, formant un multimère inactif. Cependant, FRQ semble interagir uniquement avec le WCC libre [29]. La protéine FRQ exerce donc la répression en empêchant son activateur transcriptionnel de se lier à l'ADN. De plus, FRQ n'a aucune influence sur le taux de dégradation de la protéine WCC : il a été démontré que la stabilité de WCC est identique en présence et en absence de FRQ [47].

De récents travaux ont montré que le niveau des protéines WC-1 oscille également, mais en opposition de phase par rapport aux oscillations de FRQ [47]. Il a

été par la suite établi qu'en réalité, la production de la protéine WC-1 est favorisée par la protéine FRQ. Ainsi, si on induit artificiellement la production de FRQ dans des souches mutantes où le gène *frq* est inactif, de nouvelles protéines WC-1 sont produites après un délai de six à huit heures. Cette régulation est purement post-transcriptionnelle, car le niveau de transcrits *wc-1* reste constant tout au long du cycle. Le mécanisme exact de cette activation reste cependant inconnu. WC-2 semble également activé par FRQ, mais d'une manière non rythmique.

Les autres boucles de rétroaction impliquées dans l'horloge circadienne de *Neurospora* sont moins bien connues. Il semble néanmoins exister une boucle de rétroaction impliquée dans la réponse à la lumière. L'effet de la lumière est de lever l'autorépression de FRQ [11]. Un premier pulse de lumière active la production de transcrits de *frq* (de 4 à 25 fois le niveau moyen pendant une période). Ces transcrits sont ensuite très rapidement dégradés (avec des demi-vies de l'ordre d'une heure). En fait, il a été ensuite montré qu'il existe deux sites de liaison spécifique sensibles à la lumière dans le promoteur de *frq* appelés "Light Response Elements" (LREs) [29]. De plus, un gène (*vivid*) au moins a été identifié dans cette boucle de rétroaction. VIVID semble réguler sans toutefois contrôler la réponse à la lumière, vraisemblablement via une hyperphosphorylation de WC-1 [41].

B Exemples de modèles d'horloges circadiennes

De nombreux modèles d'horloges circadiennes ont été proposés. Nous faisons ici une revue succincte et non exhaustive de certains modèles parmi les plus intéressants, dans un ordre croissant de complexité et de réalisme :

- Une application du modèle de Goodwin par Ruoff et al.

Le modèle de Goodwin est un modèle simple d'oscillateur génétique basé sur une boucle de rétroaction négative, modélisé par trois équations différentielles couplées :

$$\begin{aligned}\frac{dX}{dt} &= \frac{k_1}{Z^n + 1} - k_4 X \\ \frac{dY}{dt} &= k_2 X - k_5 Y \\ \frac{dZ}{dt} &= k_3 Y - k_6 Z\end{aligned}$$

En terme biologiques, ce modèle peut se voir comme un modèle phénoménologique de production d'ARN (espèce X), codant pour une protéine Y qui

catalyse ensuite une réaction dont le produit final Z inhibe la production de l'ARN (fonction de Hill, avec le coefficient de Hill n). Concrètement, dans ce modèle, il faut prendre un coefficient de Hill très grand (9) pour avoir des oscillations.

Ruoff et al. [70–72] ont eu l'idée d'appliquer ce modèle à l'horloge circadienne de *Neurospora crassa* et en particulier ont essayé de voir si ce modèle était assez général pour faire des prédictions expérimentales sur cet oscillateur. Dans la terminologie de *Neurospora*, X représenterait l'ARN codant pour FRQ, Y la protéine FRQ et Z un inhibiteur hypothétique de FRQ. Ruoff et al. ont en particulier cherché à évaluer l'influence des variations de paramètres sur l'oscillateur dans deux cas : modifications dues à des variations de température, ou modifications liées à des mutations. Dans le premier cas [70], Ruoff et al. ont supposé que les variations des paramètres avec la température suivent une loi d'Arrhenius. Ils utilisent ensuite un Ansatz pour trouver une approximation de la période P en fonction des paramètres². Il leur est alors possible de trouver une relation entre les énergies d'activation des différentes réactions pour annuler la dépendance en température de l'oscillateur. De plus cet oscillateur compensé en température peut être malgré tout entraîné par des variations de températures. Dans un second temps, leur analyse montre également que la période de l'oscillation dépend assez fortement des constantes de dégradation des différentes espèces [71]. En particulier, si le taux de dégradation des protéines est divisé par deux, la période passe d'environ 22 heures à environ 29 heures. Cette dépendance de la période a ensuite été observée expérimentalement chez le mutant *frq⁷* [51].

- Les modèles avec délai de Smolen et al.

Smolen et al. [75, 76] ont proposé différents modèles d'horloges circadiennes avec délai. Ces modèles visent à modéliser explicitement et en détail les oscillations des seules protéines, si bien que les oscillations de l'ARN par exemple ne sont pas prises en compte. Ainsi, tout le processus d'activation transcriptionnelle et de traduction est modélisé par un terme phénoménologique. Par exemple, dans le modèle pour *Neurospora*, l'activation de FRQ par WCC est modélisée par une fonction de Michaelis-Menten avec délai. En revanche, toutes les interactions protéine-protéine et les phosphorylations sont modélisées explicitement. Dans ces modèles, les autres boucles de rétroaction positives obser-

²Concrètement, ils supposent que $P = K \prod_{n=1}^6 (k_i)^{\alpha_i}$, puis ils déterminent numériquement les α_i .

vées dans horloges circadiennes sont également simulées par des équations différentielles avec délai. Parmi leurs résultats principaux, notons que les boucles de rétroaction positives ne semblent pas nécessaires aux oscillations. En ce qui concerne la boucle centrale de rétroaction négative, il est en revanche nécessaire d'inclure des délais assez longs (de l'ordre de 7 heures) pour rendre compte des oscillations observées.

- Les modèles de Goldbeter et al..

Les modèles de Goldbeter et al. [35, 48, 49] sont des modèles visant à modéliser explicitement les comportements de toutes les espèces intervenant dans les oscillateurs circadiens, et en particulier les ARN et les protéines, afin de comprendre complètement l'origine des délais dans les oscillations circadiennes. Les activations et répressions transcriptionnelles sont en général modélisées par des fonctions de Hill. Les différents compartiments cellulaires ainsi que les phosphorylations successives des protéines sont explicitement modélisés, et sont à l'origine des délais nécessaires à la naissance des oscillations. De plus, dans le cas de la drosophile et des mammifères, il est possible d'inclure l'influence de la lumière dans ce modèle. En particulier, un modèle complet de l'horloge circadienne des mammifères permet d'expliquer les défauts d'entraînement par la lumière dans certaines pathologies humaines (“non 24 h sleep-wake syndrome”) [49].

- Le modèle de Forger et Peskin.

Forger et Peskin [24] ont proposé un modèle très détaillé d'horloges circadiennes des mammifères. Leur but est de modéliser explicitement par des lois d'action de masse toutes les réactions connues dans l'horloge circadienne des mammifères, ce qui représente 36 réactions agissant sur 74 variables. Cependant, la transcription n'est pas modélisée par une loi d'action de masse mais par un terme non-linéaire modélisant une probabilité d'activation du gène. Ce modèle a priori très complet permet d'expliquer les oscillations et les effets de certaines mutations.

D'un point de vue plus synthétique, deux types de modèles peuvent être distingués. Dans le premier type de modèles, une seule boucle de rétroaction est nécessaire pour expliquer les oscillations. Cependant, il est alors nécessaire de supposer l'existence d'étapes intermédiaires ou de délais dans le système. C'est le cas par exemple des modèles de Smolen et al. [75, 76], et des modèles de Goldbeter et collaborateurs [35, 48, 49]. Smolen et al. introduisent des délais explicites dans les équations différentielles pour expliquer les oscillations, tandis que Goldbeter et al. supposent

que des transports cellulaires ou des phosphorylations induisent ces délais. D'autres modèles n'introduisent pas de délais, mais reposent sur des hypothèses très spécifiques pour expliquer les oscillations. C'est le cas par exemple des modèles de Ruoff et al. [70–72] qui nécessitent une très grande coopérativité lors de la répression. On peut également citer le modèle de Tyson et al. [83] qui explique les oscillations par l'introduction d'une seconde boucle de rétroaction positive. Dans la plupart des modèles existants, il est donc nécessaire pour expliquer les oscillations soit d'introduire des délais plus ou moins explicites dans les équations soit de supposer des interactions spécifiques comme une coopérativité particulière ou une autre boucle de rétroaction. L'une des questions est donc de savoir s'il est possible de modéliser le rythme circadien sans introduire ni délais dans les équations, ni d'hypothèses particulières sur les réseaux d'interaction génétiques impliqués dans ce mécanisme.

C Résumé de l'article : Un modèle d'horloge circadienne de *Neurospora*

Dans la suite, nous introduisons et résumons nos résultats présentés en détail dans la référence [27], reproduite ci-après. Les modèles proposés ont pour vocation d'être à la fois simples, mais néanmoins suffisamment réalistes d'un point de vue biochimique. Nous nous sommes donc attachés à modéliser explicitement les différentes réactions à l'aide de lois d'action de masse et nous n'avons pas supposé de dynamiques plus compliquées telles que les dynamiques de Michaelis-Menten ou de Hill par exemple. De la même façon, nous nous sommes attachés à décrire explicitement dans les équations les dynamiques de chaque espèce régulée (gène, ARN, protéines).

Dans un premier temps, nous proposons un modèle pour lequel seule la boucle de rétroaction négative (répression de FRQ par WCC puis complexation entre FRQ et WCC) est nécessaire pour avoir des oscillations. Ce modèle à une boucle est relativement simple et a été étudié mathématiquement dans le chapitre III. Nous reprenons ici cette étude en nous focalisant davantage sur les processus biochimiques. L'intérêt de ce modèle simplifié est qu'il décrit qualitativement de nombreux aspects de l'oscillation circadienne de *Neurospora*. Dans un second temps, nous tenons compte de la seconde boucle de rétroaction positive (activation de WCC par FRQ) et proposons en particulier un mécanisme pour expliquer les observations expérimentales.

Ces modèles sont supposés représenter les oscillations dans un seul noyau et nous ne distinguerons pas les différents compartiments cellulaires. Toutes les concentra-

tions sont rapportées au volume du noyau, si bien qu'une concentration représente le nombre d'espèces présentes dans le noyau. Notons que ces concentrations sont expérimentalement très basses : il a été montré ainsi que l'ordre de grandeur de la concentration des protéines FRQ est de 10 à 50 copies par noyau. Cependant, il faut garder à l'esprit que durant son cycle asexué, *Neurospora* est en fait un syncytium, si bien qu'en fait les oscillations observées sont des oscillations moyennées sur plusieurs noyaux.

C.1 Modèle à une boucle

Rappelons les équations du modèle à une boucle avec la terminologie de *Neurospora*, correspondant au modèle de MFL développé précédemment. Les deux premières équations modélisent la régulation de la transcription de frequency :

$$\frac{d[frq]}{dt} = \theta[frq : WCC] - \alpha[frq][WCC] \quad (IV.1)$$

$$\frac{d[RNA]}{dt} = \rho_{FRQ}[frq : WCC] - \delta_{RNA}[RNA] \quad (IV.2)$$

Dans ces équations, *frq* représente le gène frequency libre (sans WCC lié) , *frq : WCC* représente gène auquel est lié le complexe WCC. RNA représente la concentration des transcrits de *frq*. Dans toute la suite, comme *Neurospora* est haploïde, on supposera que $[frq] + [frq : WCC] = 1$ gène par noyau. Une seule équation est donc nécessaire pour décrire la régulation de la promotion. Une protéine WCC peut se lier au promoteur du gène *frq* avec un taux α et la protéine peut se décrocher de ce promoteur avec un taux θ (Eq. IV.1). Lorsque WCC est lié au promoteur du gène frequency, la transcription est activée avec un taux ρ_{FRQ} . On suppose que l'ARN se dégrade exponentiellement avec une constante de dégradation δ_{RNA} .

Les équations suivantes modélisent la production des protéines et leurs interactions :

$$\frac{d[FRQ]}{dt} = \beta[RNA] - \gamma[FRQ].[WCC] - \delta_{FRQ}[FRQ] \quad (IV.3)$$

$$\begin{aligned} \frac{d[WCC]}{dt} = & \rho_{WCC} - \gamma[FRQ].[WCC] \\ & + \theta[frq : WCC] - \alpha[frq][WCC] - \delta_{WCC}[WCC] \end{aligned} \quad (IV.4)$$

$$\frac{d[T]}{dt} = \gamma[FRQ].[WCC] - \delta_T[T] \quad (IV.5)$$

biochimiques tels que la transcription ou la traduction sont négligés, car ces délais sont généralement courts comparés à la période d'une horloge circadienne. Les délais inhérents aux processus de transports cellulaires sont également négligés. Nous supposons également que le seul rôle des phosphorylations de FRQ est de fixer la constante de dégradation δ_{FRQ} .

C.2 Mécanismes de l'oscillation

Le modèle à une boucle proposé ci-dessus permet de simuler les oscillations du niveau des transcrits de *frq* et des protéines FRQ, comme on peut le voir Figure IV.2. Le délai entre le pic de transcrits *frq* et des protéines FRQ est d'environ six heures. La dégradation des transcrits *frq* demande environ 18 heures et est exponentielle, en accord avec les observations expérimentales. La phase de "dérépression" durant laquelle les transcrits de *frq* remontent à leur niveau maximum est d'environ 4 heures. Détaillons maintenant qualitativement le mécanisme des oscillations, en utilisant la nomenclature de la référence [59].

C.2.a Phase de dérépression

Cette phase correspond à la phase I du cycle de la MFL. Durant cette phase, la concentration de WCC libre est tout d'abord haute tandis que celle de FRQ est très basse. WCC active donc dans un premier temps la transcription de *frq* (voir Figure IV.3, panneau de gauche). Les transcrits de *frq* peuvent donc s'accumuler, puis les protéines FRQ sont ensuite traduites. Les protéines FRQ interagissent ensuite immédiatement avec le WCC libre qui est en large excès (voir Figure IV.3, panneau de droite). Lorsqu'il y a suffisamment de protéines FRQ, tous les complexes WCC ont été titrés. Une fois que le complexe WCC lié au promoteur FRQ est libéré, l'activation transcriptionnelle est terminée et l'oscillateur rentre dans la phase de répression.

C.2.b Phase de répression

Cette phase correspond à la phase II du cycle de la MFL. Au début de cette phase, l'ARN accumulé permet la traduction des protéines FRQ (Figure IV.4, panneau de gauche). Une partie des protéines FRQ produites titrent les protéines WCC, puis s'il y a suffisamment d'ARN, des protéines FRQ libres s'accumulent. Ensuite, l'ARN de FRQ disparaît peu à peu. La concentration des protéines FRQ décroît alors peu à peu, jusqu'à ce qu'il n'y ait plus assez de FRQ produit pour titrer les

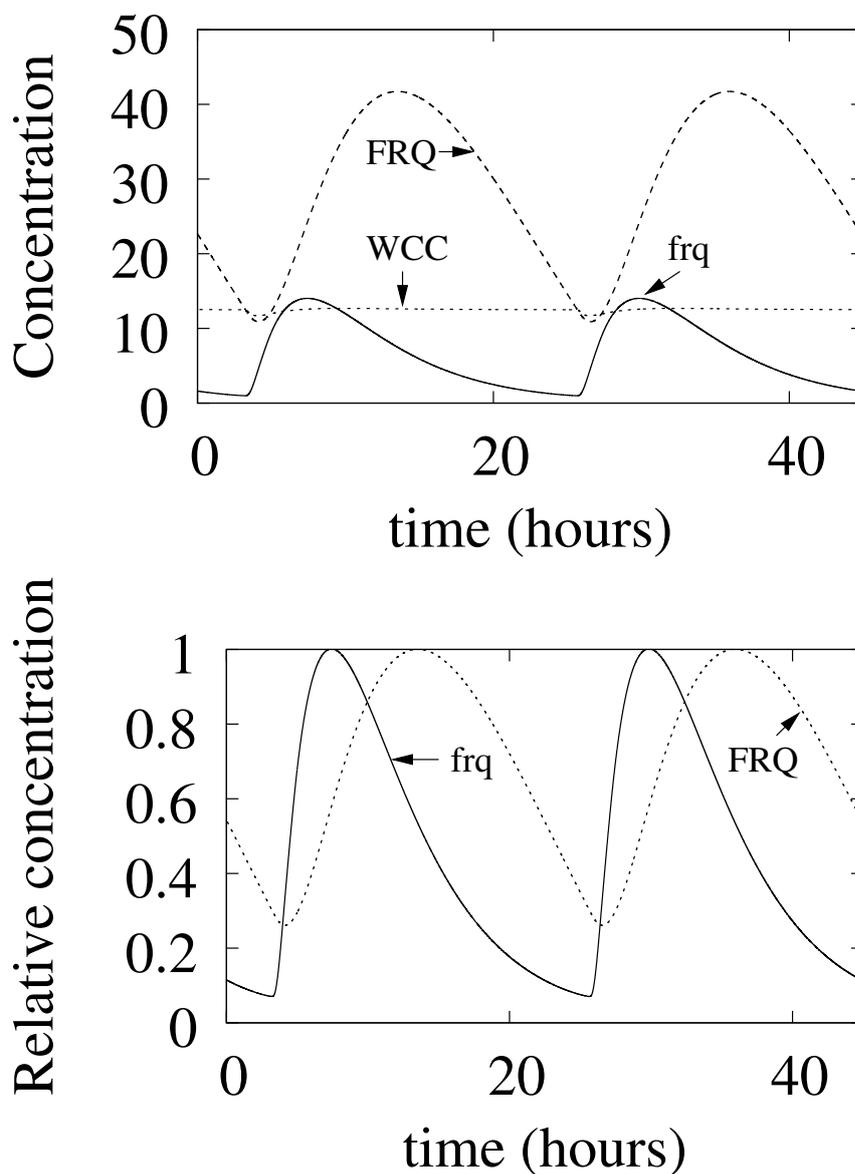


FIG. IV.2: Simulations numériques du modèle à une boucle. Haut : concentrations absolues, Bas : concentrations relatives. Les constantes pour ce modèle sont $\rho_{WCC} = 3.75 \text{ mol.h}^{-1}$, $\rho_{FRQ} = 7.5 \text{ mol.h}^{-1}$, $\beta = 0.7 \text{ h}^{-1}$, $\theta = 0.35 \text{ h}^{-1}$, $\alpha = 10 \text{ mol}^{-1}.\text{h}^{-1}$, $\delta_{WCC} = 0.3 \text{ h}^{-1}$, $\delta_{RNA} = 0.2 \text{ h}^{-1}$, $\delta_{FRQ} = 0.05 \text{ h}^{-1}$, $\gamma = 2000 \text{ mol}^{-1}.\text{h}^{-1}$, où *mol* signifie molécules et *h* heures. La constante de dégradation de l'hétérodimère WCC:FRQ est la même que celle de WC-1.

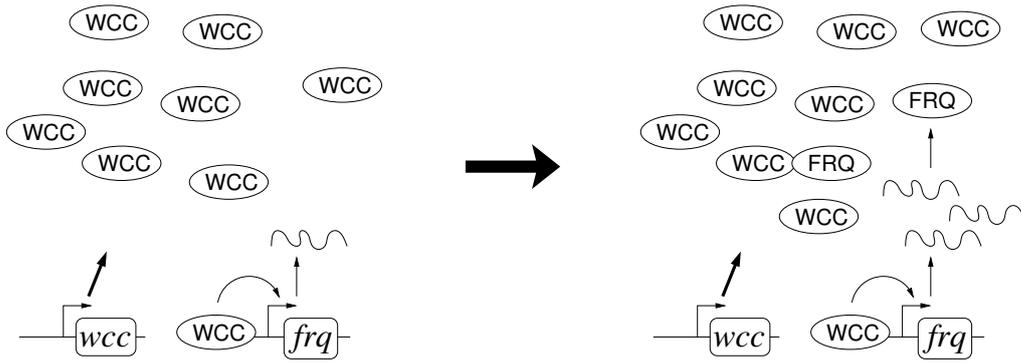


FIG. IV.3: Phase de dérèpression.

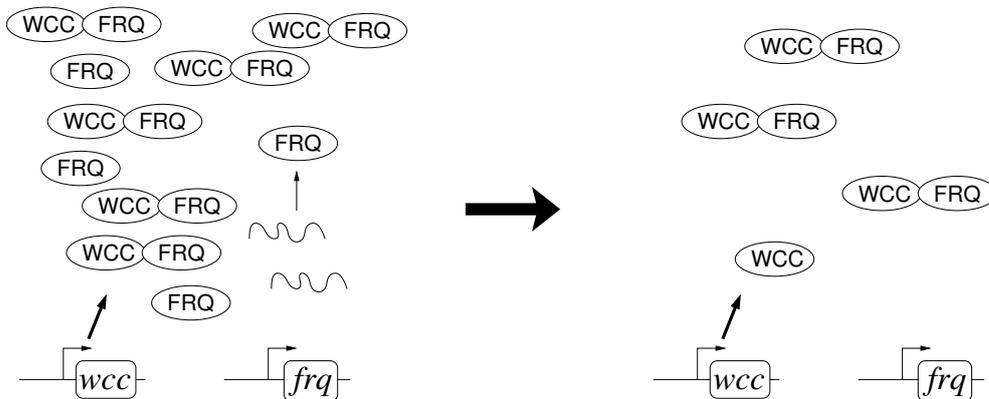


FIG. IV.4: Phase de répression.

protéines WCC (Figure IV.4, panneau de droite). Les protéines WCC peuvent alors s'accumuler de nouveau et un nouveau cycle peut commencer.

C.3 Contrôle par l'ARN

L'analyse mathématique précise de cet oscillateur simplifié est présentée dans le chapitre III. Comme nous l'avons vu, si la durée de vie de l'ARN est assez longue comparée aux temps typiques d'interaction protéine-protéine et aux temps typiques de production des protéines, la protéine ayant la plus basse concentration atteint un équilibre quasi-statique et sa dynamique est dirigée par celle de la protéine à haute concentration.

En résumé, lorsque la concentration de WCC est supérieure à celle de FRQ, les protéines WCC titrent toutes les protéines FRQ et après un temps très court, il reste uniquement des protéines WCC libres avec des dimères de WCC et FRQ. Au contraire, lorsque la concentration de FRQ est plus grande que celle de WCC, les protéines FRQ titrent les protéines WCC et après un temps très court, il reste uniquement des protéines FRQ libres avec des dimères de FRQ et de WCC. Les protéines ne peuvent donc être simultanément libres à des concentrations comparables. Une première conséquence de cette dimérisation est donc que la dynamique des deux protéines est contrôlée par la concentration des transcrits frq :

- lorsque FRQ est en excès, le taux de production de FRQ est proportionnel à la concentration des transcrits de frq , qui contrôle donc aussi la concentration de WCC à cause de l'hétérodimérisation,
- lorsque WCC est en excès, les protéines FRQ produites dimérisent immédiatement avec WCC, si bien que le taux de séquestration du WCC libre est proportionnel à la concentration des transcrits de frq .

Nous voyons donc sur ce modèle simple que la distinction explicite entre la traduction et la transcription est fondamentale pour expliquer les oscillations. En fait, dans ce modèle simple, modéliser la production des protéines FRQ par une simple étape effective (type Michaelis-Menten par exemple) détruit les oscillations. L'analyse qualitative permet de comprendre pourquoi il est nécessaire de tenir compte de cette étape de transcription : il est par exemple nécessaire d'avoir un réservoir suffisant d'ARN à la fin de la phase de dérépression pour avoir une accumulation suffisante de FRQ. Notons qu'expérimentalement le comportement qualitatif des protéines FRQ et des transcrits frq est assez différent (voir par exemple la référence [30]) : la dynamique des protéines FRQ ne reproduit pas simplement celle des transcrits frq avec un délai. Il est donc nécessaire de distinguer ces deux espèces dans

un modèle, et cette distinction permet effectivement d'expliquer le mécanisme des oscillations.

C.4 Modèle à deux boucles

Une boucle de rétroaction positive a été également observée chez *Neurospora* : FRQ active la production de WC-1 à travers un mécanisme post-transcriptionnel. Cependant, le mécanisme précis d'activation demeure inconnu. De plus, les oscillations de FRQ et de WC-1 sont alors en opposition de phase.

Nous avons cherché plusieurs modèles pour expliquer cette opposition de phase. Dans un premier temps, nous avons supposé que FRQ activait la traduction de WC-1 en se liant aux transcrits de *wc-1*, et que cette activation était catalysée par une enzyme intermédiaire. Ce mécanisme permettait d'expliquer l'opposition de phase des oscillations, mais ne permettait pas d'expliquer l'amplitude des oscillations de WC-1. Dans un second essai, comme FRQ peut former des homodimères [7], nous avons supposé que cette activation était dirigée par les dimères de FRQ. Cependant, dans ce cas, WCC était produit essentiellement à haute concentration de FRQ, si bien que le pic de WCC était simplement retardé mais n'était pas en opposition de phase. Après analyse de ces exemples, il apparaissait qu'une façon simple d'expliquer l'opposition de phase était d'avoir activation uniquement à des concentrations moyennes de FRQ, mais répression à haute concentration de FRQ. Un mécanisme particulier simple réalisant ce processus nous a alors été suggéré par l'algorithme d'évolution (voir Figure II.19 dans le chapitre II) et nous l'avons appliqué à l'horloge de *Neurospora*.

L'hypothèse de base dans ce second modèle est que les protéines FRQ peuvent interagir avec les transcrits de *wc-1* afin de modifier leur taux de traduction. Dans la suite, nous tiendrons donc explicitement compte des processus de transcription/traduction pour la protéine WC-1. En revanche, nous supposerons toujours que WC-2 est en large excès si bien que la dimérisation entre WC-1 et WC-2 est immédiate et qu'on assimile donc la concentration de WC-1 à celle de WCC. Dans ce second modèle à deux boucles, nous supposons que les protéines FRQ interagissent avec les transcrits *wc-1*. Le complexe protéine/ARN est directement traduit sans aucun délai. Dans ce modèle, nous tenons cette fois compte de l'homodimérisation de FRQ. Nous supposons qu'une deuxième protéine FRQ peut interagir avec la première protéine FRQ liée au transcrit de *wc-1*, et que cette fois ce nouveau complexe protéines/ARN ne peut plus être traduit. Les nouvelles équations pour la régulation des transcrits *wc-1* sont donc :

$$\frac{d[\text{RNA}_W]}{dt} = \rho_{\text{WCC}} - \delta_{\text{RNA}_W}[\text{RNA}_W] - \nu[\text{RNA}_W][\text{FRQ}] + \mu[\text{RNA}_{W+}] \quad (\text{IV.6})$$

$$\begin{aligned} \frac{d[\text{RNA}_{W+}]}{dt} &= \nu[\text{RNA}_W][\text{FRQ}] - (\delta_{\text{RNA}_W} + \mu)[\text{RNA}_{W+}] \\ &\quad - \eta[\text{RNA}_{W+}][\text{FRQ}] + \kappa[\text{RNA}_{W-}] \end{aligned} \quad (\text{IV.7})$$

$$\frac{d[\text{RNA}_{W-}]}{dt} = \eta[\text{RNA}_{W+}][\text{FRQ}] - (\delta_{\text{RNA}_W} + \kappa)[\text{RNA}_{W-}] \quad (\text{IV.8})$$

RNA_W représente la forme normale des transcrits *wc-1*, RNA_{W+} la forme complexée une fois avec FRQ, RNA_{W-} la forme complexée deux fois avec FRQ. On suppose toujours que les transcrits *wc-1* sont produits avec un taux fixe ρ_{WCC} , et que ces transcrits peuvent interagir avec FRQ pour former un complexe ARN/protéine avec un taux ν . La dissociation de ce complexe est possible avec un taux μ . Enfin, la forme activée des transcrits peut interagir une seconde fois avec FRQ.

Pour les régulations post-transcriptionnelles, on suppose que les protéines FRQ forment des homodimères avec un taux η et que cet homodimère peut se dissocier avec un taux κ . Enfin, l'homodimère de FRQ (FRQ_2) interagit avec WCC pour former un multimère $\text{FRQ}_2 : \text{WCC}$ avec un taux γ . Les équations pour FRQ, WCC, FRQ_2 et $\text{FRQ}_2 : \text{WCC}$ sont donc :

$$\begin{aligned} \frac{d[\text{FRQ}]}{dt} &= \beta[\text{RNA}] - \delta_{\text{FRQ}}[\text{FRQ}] \\ &\quad - \nu[\text{RNA}_W][\text{FRQ}] + \mu[\text{RNA}_{W+}] \\ &\quad - \eta[\text{RNA}_{W+}][\text{FRQ}] + \kappa[\text{RNA}_{W-}] \\ &\quad - 2\eta[\text{FRQ}]^2 + 2\kappa[\text{FRQ}_2] \end{aligned} \quad (\text{IV.9})$$

$$\begin{aligned} \frac{d[\text{FRQ}_2]}{dt} &= \eta[\text{FRQ}]^2 - \kappa[\text{FRQ}_2] \\ &\quad - \gamma[\text{FRQ}_2] \cdot [\text{WCC}] - \delta_{\text{FRQ}}[\text{FRQ}_2] \end{aligned} \quad (\text{IV.10})$$

$$\begin{aligned} \frac{d[\text{WCC}]}{dt} &= \beta_-[\text{RNA}_W] + \beta_+[\text{RNA}_{W+}] - \gamma[\text{FRQ}_2][\text{WCC}] \\ &\quad + \theta[\text{frq} : \text{WCC}] - \alpha[\text{frq}][\text{WCC}] - \delta_{\text{WCC}}[\text{WCC}] \end{aligned} \quad (\text{IV.11})$$

$$\frac{d[\text{FRQ}_2 : \text{WCC}]}{dt} = \gamma[\text{FRQ}_2][\text{WCC}] - \delta_T[\text{FRQ}_2 : \text{WCC}] \quad (\text{IV.12})$$

Une représentation schématique de ce modèle à deux boucles est représenté Figure IV.5, les simulations correspondantes sont présentées Figure IV.6.

C.4.a Rôle possible de la rétroaction positive

Expérimentalement, il est possible d'avoir de grandes variations de l'amplitude de l'oscillateur tout en conservant une période constante [8, 52]. Cheng et al. [8] ont

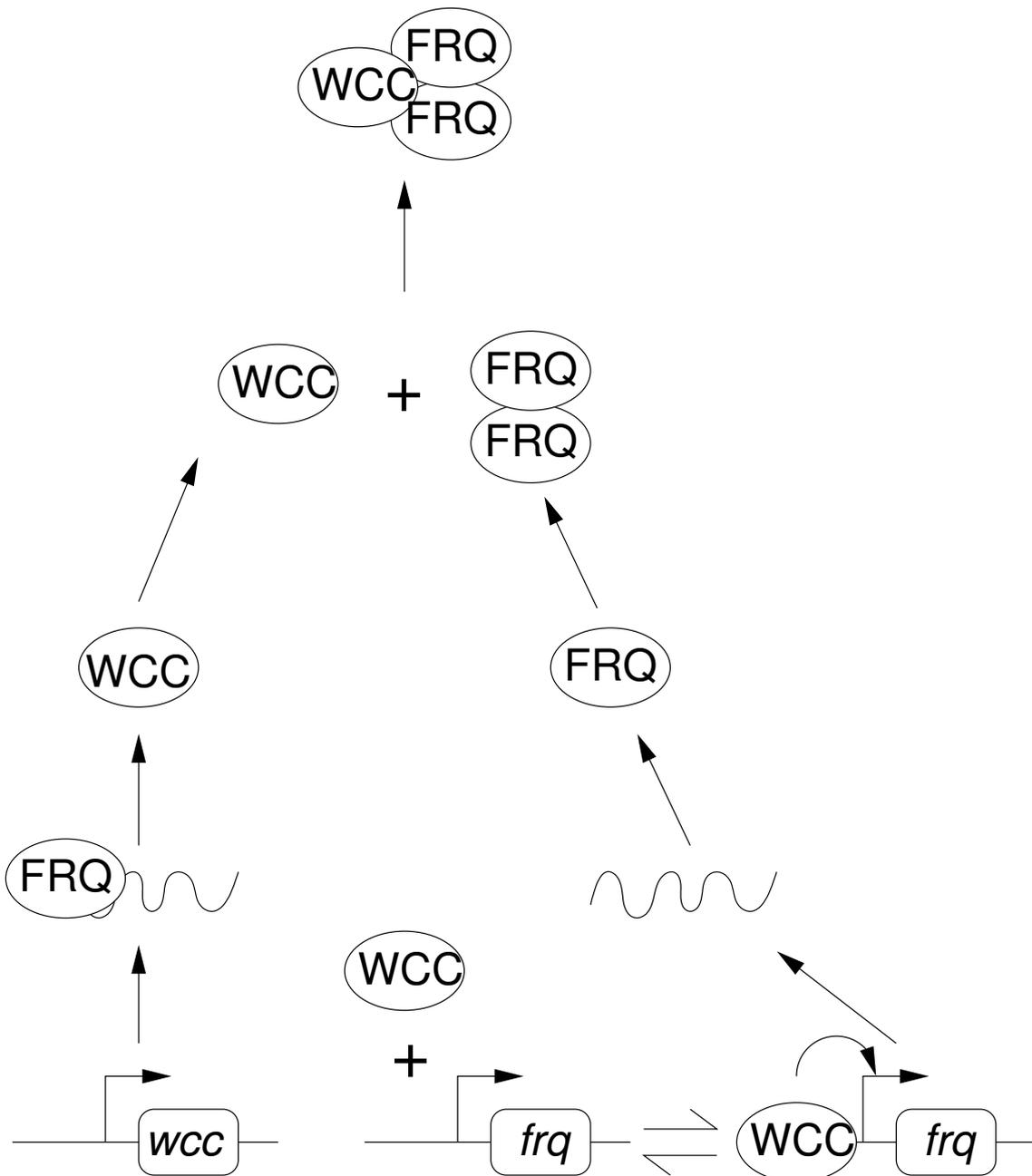


FIG. IV.5: Représentation schématique du modèle à deux boucles.

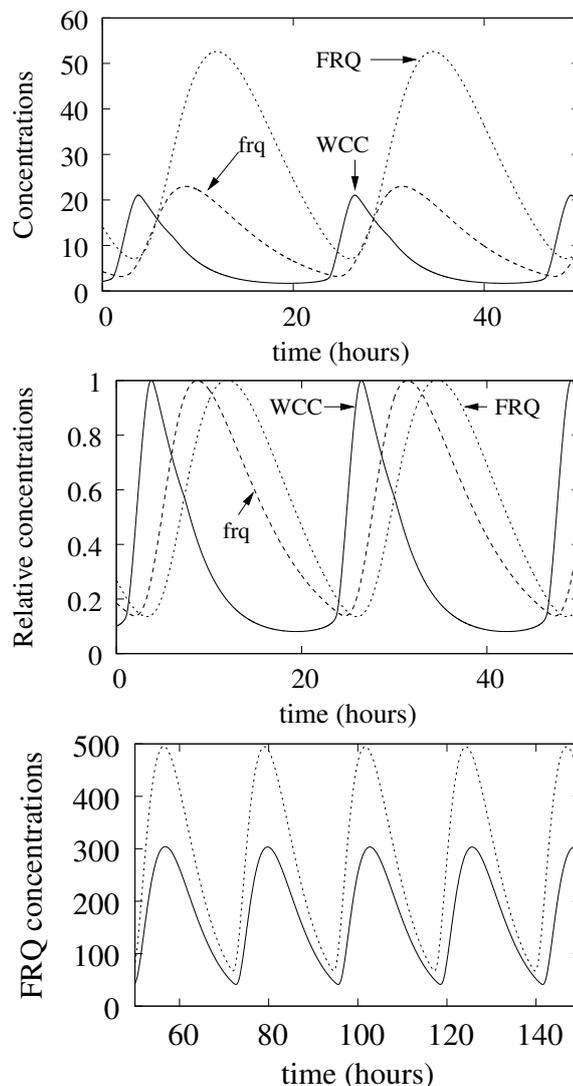


FIG. IV.6: Simulations du modèle à deux boucles avec activation post-transcriptionnelle de WCC par FRQ. Haut : concentrations absolues, Milieu : concentrations relatives, Bas : influence du taux de transcription de WCC sur les oscillations de FRQ, ligne pleine, $\rho_{WCC} = 2.5 \text{ mol.h}^{-1}$, en pointillés, $\rho_{WCC} = 5 \text{ mol.h}^{-1}$. Les constantes pour ce modèle sont $\rho_{WCC} = 2.5 \text{ mol.h}^{-1}$, $\rho_{FRQ} = 75 \text{ mol.h}^{-1}$, $\beta = 1 \text{ h}^{-1}$, $\theta = 0.25 \text{ h}^{-1}$, $\alpha = 0.003 \text{ mol}^{-1}.\text{h}^{-1}$, $\delta_{WCC} = 0.3 \text{ h}^{-1}$, $\delta_{RNA} = 0.2 \text{ h}^{-1}$, $\delta_{FRQ} = 0.3 \text{ h}^{-1}$, $\gamma = 1600 \text{ mol}^{-1}.\text{h}^{-1}$, $\delta_T = 0.3 \text{ h}^{-1}$, $\delta_{RNAW} = 1 \text{ h}^{-1}$, $\nu = 8000 \text{ mol}^{-1}.\text{h}^{-1}$, $\mu = 0.1 \text{ h}^{-1}$, $\beta^- = 0.001 \text{ h}^{-1}$, $\beta^+ = 10 \text{ h}^{-1}$, $\eta = 3000 \text{ mol.h}^{-1}$, $\kappa = 10 \text{ h}^{-1}$.

justement proposé qu'un rôle possible de la seconde boucle de rétroaction positive était d'ajuster les taux de production des différentes protéines pour garder une période constante. Le modèle à deux boucle décrit ici est compatible avec cette suggestion : lorsque les taux de production des protéines sont modifiés, les oscillations s'ajustent pour garder une période constante, comme on peut le voir sur le panneau du bas de la figure IV.6. Si l'on multiplie le paramètre ρ_{WCC} par quatre, la période est modifiée d'environ 2 % par rapport à la période de référence alors que l'amplitude est multipliée par trois, si l'on multiplie ce paramètre par huit, la modification de la période est d'environ 7 % et l'amplitude est environ cinq fois plus grande. Nous avons également comparé la résistance de l'amplitude et de la période des oscillations à de fortes variations de paramètres. Pour chaque modèle, nous avons multiplié puis divisé chaque paramètre par deux et comparé l'amplitude et la période par rapport au modèle de référence. Le résultat de ces simulations est présenté dans la Figure IV.7. Il apparaît que le modèle à deux boucles est beaucoup plus robuste que le modèle à une boucle face à la variation des paramètres.

C.5 Conclusions

Nous avons donc proposé ici un modèle simple de l'oscillateur circadien de *Neurospora crassa*, basé sur des lois d'action de masse. Ce modèle répond à un double objectif : être à la fois réaliste et suffisamment simple. Le modèle à une boucle présente l'intérêt d'être complètement étudiable mathématiquement. En particulier, il est possible de calculer beaucoup de propriétés de cet oscillateur comme l'amplitude et la période des oscillations, comme nous l'avons fait dans le chapitre III. Ce modèle montre également qu'il n'y a a priori pas besoin de faire des hypothèses supplémentaires sur le rôle des délais, sur le rôle des phosphorylations ou sur le rôle de la coopérativité dans la répression pour expliquer les oscillations. Ce modèle peut de plus être amélioré pour prendre en compte par exemple la seconde boucle de rétroaction positive présente chez *Neurospora*. Cette boucle de rétroaction semble en particulier favoriser la robustesse des oscillations face à la variation des paramètres.

D Article

Nous reproduisons dans la suite l'article détaillant notre modèle d'horloge circadienne, qui correspond à la référence [27].

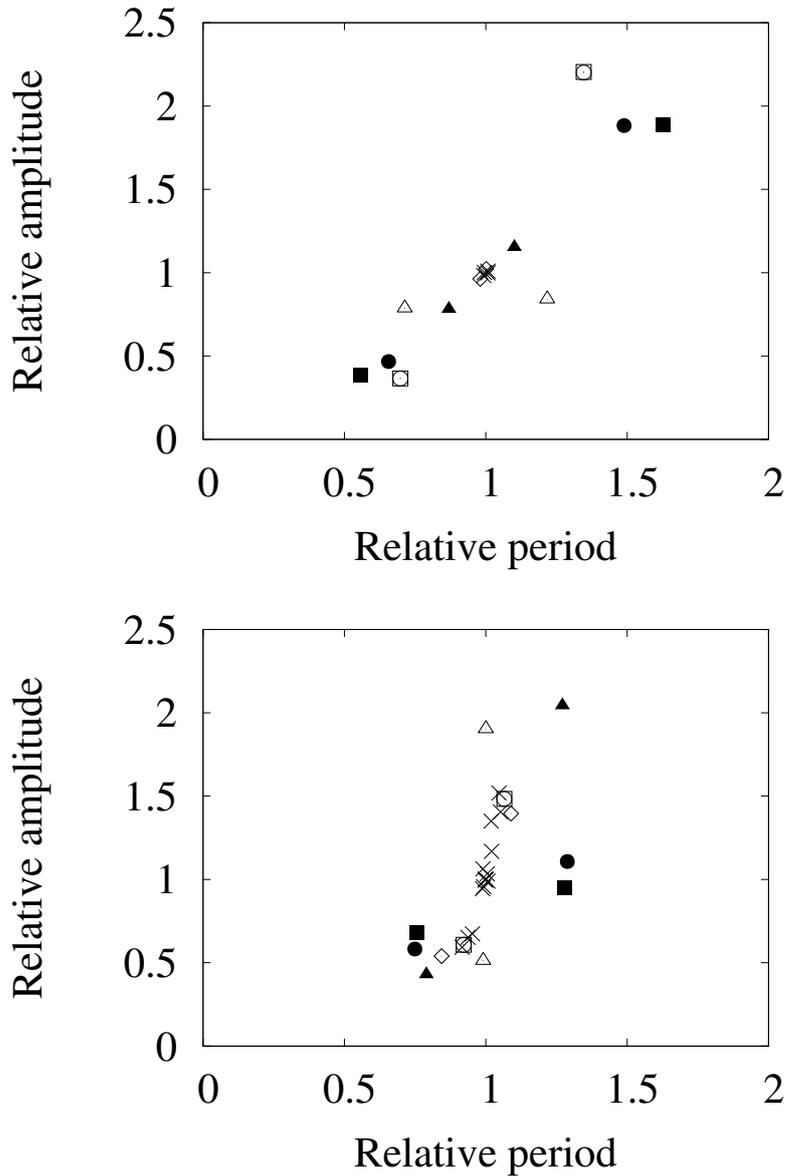


FIG. IV.7: Effet de fortes variations des paramètres sur les oscillations. Chaque point représente une simulation dans laquelle un paramètre est doublé ou divisé par deux en conservant les autres paramètres constants. Les amplitudes et les périodes sont mesurées relativement à l'amplitude et à la période du contrôle. L'amplitude mesurée est celle de l'oscillation de FRQ. Carré plein δ_{RNA} , cercle plein θ , triangle plein δ_{FRQ} , triangle ouvert ρ_{WCC} , carré ouvert ρ_{FRQ} , cercle ouvert β , losange δ_{WCC} , autres paramètres \times . Haut : modèle à une boucle, Bas, modèle à deux boucles

A Model for the *Neurospora* Circadian Clock

Paul François

Laboratoire de Physique Statistique, Centre National de la Recherche Scientifique, UMR 8550, Ecole Normale Supérieure, Paris, France

ABSTRACT Circadian clocks are important biological oscillators that generally involve two feedback loops. Here, we propose a new model for the *Neurospora crassa* circadian clock. First, we model its main negative feedback loop, including only experimentally well-documented reactions, the transcriptional activation of frequency (*frq*) by the white-collar complex (WCC), and the post-transcriptional dimerization of FRQ with WCC. This main loop is sufficient for oscillations and a similar one lies at the core of almost all known circadian clocks. Second, the model is refined to include the less characterized enhancement of white-collar 1 (WC-1) protein synthesis by FRQ, the positive second feedback loop. Numerical testing of different hypotheses led us to propose that the synthesis of WC-1 is enhanced by FRQ monomers and repressed by FRQ dimers. We demonstrate that this second loop contributes significantly to the robustness of the oscillator period against parameter variation. A phase response curve to light pulses is also computed and agrees well with experiments. On a general level, our results show that explicit time delays are not required for sustained oscillations but that it is crucial to take into account mRNA dynamics and protein-protein interactions.

INTRODUCTION

Circadian clocks are important examples of genetic oscillators used to synchronize organisms to the daily cycle of light and dark. Circadian rhythms have been widely studied for many years (Daan and Pittendrigh, 1976), and recent works have unveiled the detailed mechanisms of this internal timing in several organisms (Young, 2002; Reppert and Weaver, 2001, 2002). Clocks from different organisms appear to share common features. Their core component relies on one feedback loop including at least two genetic interactions, a positive and a negative one. At least two proteins or groups of proteins are involved in these genetic interactions. The first group of proteins, the activating proteins, interacts with the DNA and activates the transcription of genes corresponding to the second group of proteins. In coordination with some post-transcriptional modifications, this second group of proteins usually interacts in the cytoplasm and in the nucleus with the activating proteins, forming multimers unable to activate transcription. These proteins are consequently repressing proteins. The aim of this article is to describe and to study these interactions in a simple system where the core components of this main feedback loop have been well described, the *Neurospora crassa* circadian clock (Loros and Dunlap, 2001), and to compare it to the experiments. For this fungus, circadian rhythmic growth patterns were described 50 years ago (Pittendrigh et al., 1959). With advances in molecular biology, understanding of its circadian clock has improved, and main components of this clock have been determined and in some cases described in a quantitative way (Garceau et al., 1997; Ballario et al., 1998; Mellow et al., 1997; Lee et al., 2000; Froehlich et al., 2003).

In the following, a model of the *Neurospora* circadian clock main loop is first proposed and compared to available experimental data. Biological interactions are modeled with mass-action laws so that the necessary delays in the clock are the consequence of the well-described chemical reactions. The model of the core loop appears to correctly describe oscillations of *frq* transcripts and FRQ proteins, but does not account for the observed WC-1 oscillations. To describe them, a second positive loop involving the enhancement of WC-1 synthesis by FRQ (Lee et al., 2000) needs to be taken into account. Refined models are proposed and tested. This leads to the specific proposal that WC-1 translation is enhanced by FRQ monomers and suppressed by FRQ homodimers. The positive feedback loop is found to enhance robustness of the clock to parameter variations. Light response of this model is also computed, and is found to be in good agreement with the experiments.

Some experimental results about the *Neurospora* circadian clock

The *Neurospora* circadian clock is based on an autoregulatory negative-feedback loop with three proteins: the FREQUENCY protein, FRQ, the repressing protein; and white-collar proteins WC-1 and WC-2, the activating proteins. Here we summarize the main experimental findings (for detailed reviews, see Loros and Dunlap, 2001; Dunlap et al., 2004).

The gene *frq* is historically one of the first to have been identified as a part of the core *Neurospora*'s circadian clock (Feldman and Hoyle, 1973). In constant darkness, *frq* RNA and FRQ protein concentrations oscillate. The peak of *frq* transcript is followed after 4–6 h by a somewhat larger peak of FRQ proteins (Fig. 1 A redrawn from Garceau et al., 1997). The circadian cycle can be divided in two precisely

Submitted October 4, 2004, and accepted for publication December 28, 2004.

Address reprint requests to Paul François, Ecole Normale Supérieure, Physics, 24 rue Lhomond, Paris 75231, France. Tel.: 33-1-44-32-3763; E-mail: francois@toursol.lps.ens.fr.

© 2005 by the Biophysical Society

0006-3495/05/04/2369/15 \$2.00

doi: 10.1529/biophysj.104.053975

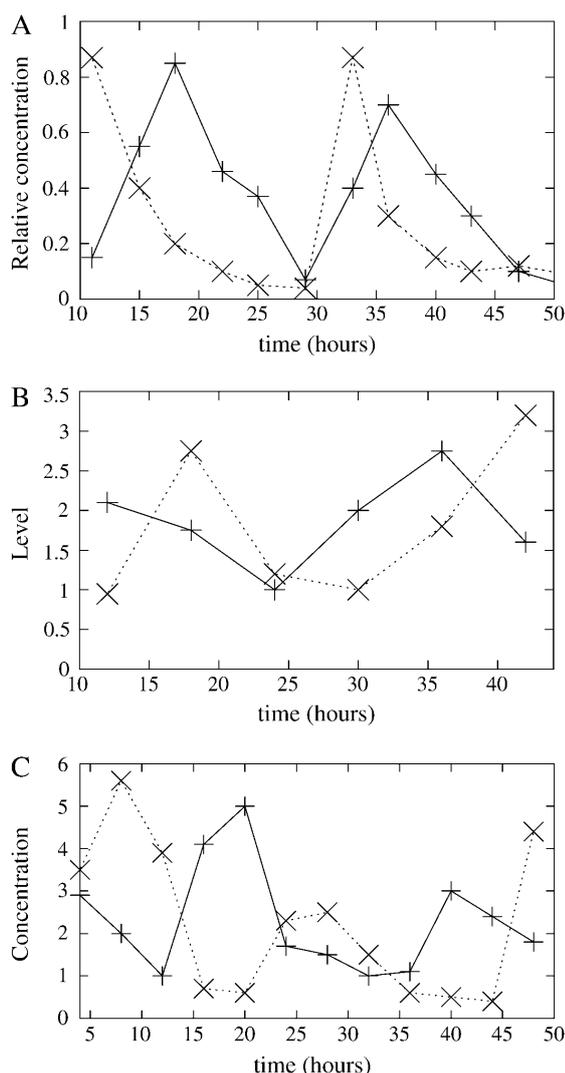


FIGURE 1 (A) Redrawn from experimental data from Garceau et al. (1997). The symbol \times is the relative concentration of *frq* RNA; and the symbol $+$ is the relative concentration of FRQ protein. (B) Redrawn from experimental data from Froehlich et al. (2003) showing rhythmic binding of WCC to the FRQ promoter. The symbol $+$ is the densitometry of the C-box-bound complex; and the symbol \times is the densitometry of FRQ protein. (C) Redrawn from experimental curves from Lee et al. (2000), showing antiphase oscillations of FRQ and WC-1. The symbol \times is the WC-1 protein level; and the symbol $+$ is the FRQ protein level.

defined phases (Merrow et al., 1997). The first phase is the negative feedback itself (repression) in which FRQ represses its own transcription. This phase is ~ 14 – 18 h long. The second phase (de-repression) is simply the recovery from this repression when *frq* transcript level returns to high concentrations. This step is 4 h long.

White-collar proteins (WC-1 and WC-2) are transcription factors. WC-2 is an abundant constitutive protein (Denault et al., 2001). White-collar proteins interact to form an heterodimer of WC-1 and WC-2, the white-collar complex (WCC). This complex rhythmically binds to the promoter of

the *frq* gene and enhances transcription, as shown in Fig. 1 B, redrawn from Froehlich et al. (2003). After splicing, FRQ protein is produced and interacts in the nucleus with WCC, preventing WCC's interaction with *frq* promoter. It is generally supposed that this mechanism of repression is due to the sequestration of WCC by FRQ (Denault et al., 2001; Froehlich et al., 2003).

The coiled-coil-domain-mediated FRQ-FRQ interaction is also necessary to circadian oscillations (Cheng et al., 2001a). It seems necessary for the interaction between FRQ and WCC, but its precise role in the circadian oscillation is still unknown.

Recent work showed that the level of WC-1 proteins also oscillates, in phase opposition to FRQ proteins, as shown in Fig. 1 C redrawn from Lee et al. (2000). However, the *wc-1* transcript level does not vary throughout the day. Dunlap and co-workers established that this mechanism is triggered by the presence of FRQ proteins (Lee et al., 2000). Also, when the *frq* gene is knocked-out, WC-1 level is very low compared to the level in wild-type cells. These experimental results all suggest an enhancement by FRQ of the production of new WC-1 proteins from the existing transcripts. However, the detailed mechanism of this enhancement is still unknown.

It was finally shown that transcription of WC-2 is also activated by FRQ, but in a nonrhythmic way (Cheng et al., 2001b). In this work, Cheng and co-workers engineered quinic acid (QA)-controlled strains of *Neurospora* and also observed that despite considerable changes in the levels of WC and FRQ proteins due to induction by QA, the period of the clock changed only slightly. The amplitude of the clock can therefore vary whereas its period remains constant.

METHODS

Model characteristics

For simplicity and because of lack of precise experimental data, different cellular compartments and separate concentrations for the nucleus and cytoplasm are not considered. All concentrations are referenced to the cell volume, so that concentrations represent the effective number of molecules present in the cells. As WC-2 is expressed in large excess compared to WC-1, we assume that WC-2 quickly interacts with WC-1 to form WCC.

The one-loop model

The first two equations model the transcriptional regulation of *frq* transcripts by WCC:

$$\frac{d[\text{frq}]}{dt} = \theta[\text{frq} : \text{WCC}] - \alpha[\text{frq}][\text{WCC}], \quad (1)$$

$$\frac{d[\text{RNA}]}{dt} = \rho_{\text{FRQ}}[\text{frq} : \text{WCC}] - \delta_{\text{RNA}}[\text{RNA}]. \quad (2)$$

Note that *frq* represents the frequency gene without WCC bound to its promoter; *frq*:WCC denotes the gene bound to WCC; and RNA stands for *frq* transcripts. As there is only one copy of each gene per cell during vegetative growth, $[\text{frq}] + [\text{frq} : \text{WCC}] = 1$ gene per cell, so that only one equation is necessary to describe regulation of the promotion. In details,

WCC proteins can bind to the *frq* gene with a rate α and the bound protein is released with a rate θ (Eq. 1). When WCC is bound to the frequency gene, transcription is initiated with a rate ρ_{FRQ} . We assume a first-order degradation for RNA with a constant degradation rate δ_{RNA} .

The following differential equations stand for protein productions and regulations:

$$\frac{d[FRQ]}{dt} = \beta[RNA] - \gamma[FRQ][WCC] - \delta_{FRQ}[FRQ], \quad (3)$$

$$\frac{d[WCC]}{dt} = \rho_{WCC} - \gamma[FRQ][WCC] + \theta[frq : WCC] - \alpha[frq][WCC] - \delta_{WCC}[WCC], \quad (4)$$

$$\frac{d[T]}{dt} = \gamma[FRQ][WCC] - \delta_T[T]. \quad (5)$$

In these equations, [FRQ], [WCC], and [T] denote, respectively, the concentration of FRQ protein, of WCC, and of the multimer formed by FRQ and WCC.

FRQ is translated from the transcripts with a rate β . The complex WCC is assumed to be produced with a fixed rate ρ_{WCC} . FRQ and WCC proteins can form a multimer T with a rate γ . It is supposed that FRQ only binds to the free form of WCC and does not bind to the WCC protein bound to the *frq* promoter. Additionally, the complex formed by WCC and FRQ is not able to promote transcription. Finally, the dissociation of the complex is neglected. A schematic representation of the one-loop model is presented in Fig. 2 A.

In this one-loop model, all time-delays inherent to the biological machinery, like transcription and translation delays, are neglected, since they are generally short compared to the period of the circadian clock. Delays induced by cellular transports are not included. We also suppose that the only role of phosphorylations is to fix the degradation rate δ_{FRQ} .

Models with another positive feedback loop

A positive feedback loop has been observed in *Neurospora*: thanks to a post-transcriptional mechanism, FRQ enhances WC-1 production. However, the precise mechanism mediating this enhancement is still unknown. To provide some indication on the type of possible biological mechanism, we study two additional models. Both of these models rely on the fact that FRQ might interact with the *wc-1* transcripts and enhance their transcription. In the following, WC-2 is still supposed to be in large excess and to quickly form a dimer with WC-1.

First two-loop model

In this model, we suppose that FRQ proteins interact with *wc-1* transcripts, forming a complex. The complex between FRQ and *wc-1* transcripts is translated with a delay τ after its formation. So additional equations describe the dynamics of *wc-1* transcripts, a species not explicitly taken into account in the previous one-loop model:

$$\frac{d[RNA_w]}{dt} = \rho_{WCC} - \delta_{RNA_w}[RNA_w] - \nu[RNA_w][FRQ] + \mu[RNA_{w+}], \quad (6)$$

$$\frac{d[RNA_{w+}]}{dt} = \nu[RNA_w][FRQ] - (\delta_{RNA_w} + \mu)[RNA_{w+}]. \quad (7)$$

The value RNA_w stands for the normal form of the *wc-1* transcripts, and RNA_{w+} stands for its enhanced form. The *wc-1* transcripts are supposed to be produced with a fixed rate ρ_{WCC} , and their normal form to interact with FRQ to form a complex with a second-order reaction rate ν . The reverse reaction is possible with a rate μ . Note that after some time, the total

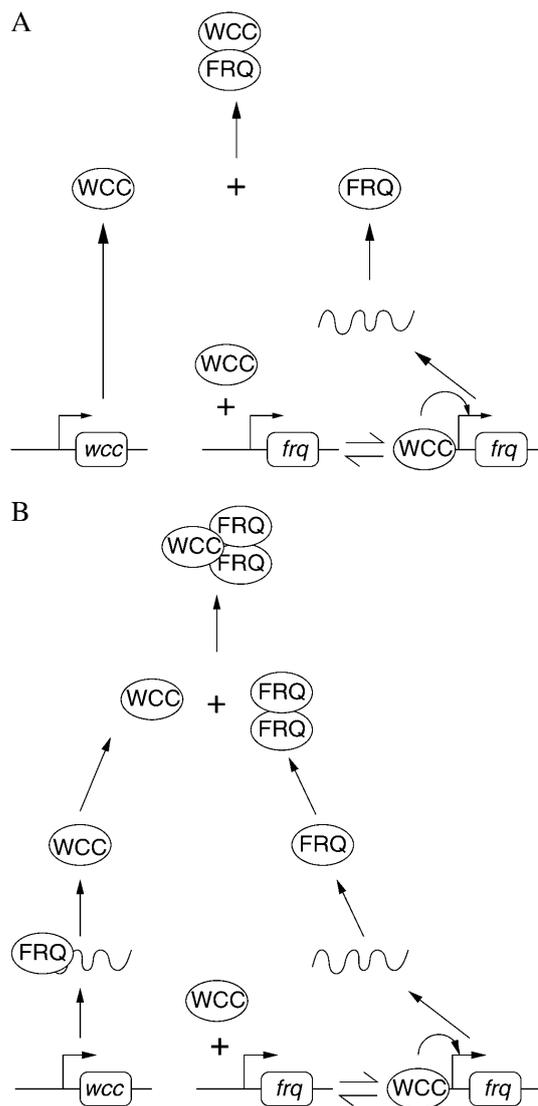


FIGURE 2 (A) Schematic representation of the one-loop model; (B) schematic representation of the second two-loop model.

concentration of the *wc-1* transcripts is constant. The equation for FRQ must be modified as

$$\frac{d[FRQ]}{dt} = \beta[RNA] - \gamma[FRQ][WCC] - \delta_{FRQ}[FRQ] - \nu[RNA_w][FRQ] + \mu[RNA_{w+}]. \quad (8)$$

Finally, the enhancement of translation is simply supposed to appear after a delay (which phenomenologically accounts for undescribed biological processes),

$$\frac{d[WCC]}{dt} = \beta_-[RNA_w] + \beta_+[RNA_{w+}]_\tau - \gamma[FRQ][WCC] + \theta[frq : WCC] - \alpha[frq][WCC] - \delta_{WCC}[WCC]. \quad (9)$$

There are two different translation rates in this last equation: β_- stands for the translation rate of the normal form and β_+ stands for the translation rate

of the complexed form. $[RNA_{W+}]_t$ stands for $[RNA_{W+}](t - \tau)$, where τ is the delay in translation of the second type of RNA.

Second two-loop model

As will be seen, the delay τ needs to be quite long to reproduce experimental data. This delay should result from well-defined biochemical interactions and several hypothetical interactions were tested to see which model could agree with the experimental data, as there is no precise description of the activation of WCC by FRQ yet.

This led us to propose a second two-loop model without any explicit delay. In this second two-loop model, FRQ proteins are still supposed to directly interact with *wc-1* transcripts, and form a complex. The complex between FRQ and *wc-1* transcript is translated after its formation without any delay. In this model we take into account the homodimerization of FRQ. It is hypothesized that another FRQ protein can interact with the FRQ protein bound to the *wc-1* transcript, and that the complex formed by this FRQ dimer cannot be translated. The new equations for the *wc-1* transcript are

$$\frac{d[RNA_W]}{dt} = \rho_{WCC} - \delta_{RNA_W}[RNA_W] - \nu[RNA_W][FRQ] + \mu[RNA_{W+}], \quad (10)$$

$$\frac{d[RNA_{W+}]}{dt} = \nu[RNA_W][FRQ] - (\delta_{RNA_W} + \mu)[RNA_{W+}] - \eta[RNA_{W+}][FRQ] + \kappa[RNA_{W-}], \quad (11)$$

$$\frac{d[RNA_{W-}]}{dt} = \eta[RNA_{W+}][FRQ] - (\delta_{RNA_W} + \kappa)[RNA_{W-}]. \quad (12)$$

The value RNA_W stands for the normal form of *wc-1* transcripts, RNA_{W+} stands for the enhanced form complexed with FRQ, and RNA_{W-} for the form complexed with two FRQ proteins. We suppose that *wc-1* transcripts are produced with a fixed rate ρ_{WCC} , and that the *wc-1* transcripts' normal form can interact with FRQ to form a RNA-protein complex with a second-order reaction rate ν . The reverse reaction is possible with a rate μ . The enhanced form can interact once again with FRQ.

For protein-protein interactions, we suppose that FRQ homodimerizes with second-order rate η and that homodimers dissociate into two FRQ proteins with rate κ . Then FRQ dimer (FRQ₂) can interact with WCC to form the multimer FRQ₂:WCC. Equations for FRQ, WCC, FRQ₂, and FRQ₂:WCC consequently are

$$\frac{d[FRQ]}{dt} = \beta[RNA] - \delta_{FRQ}[FRQ] - \nu[RNA_W][FRQ] + \mu[RNA_{W+}] - \eta[RNA_{W+}][FRQ] + \kappa[RNA_{W-}] - 2\eta[FRQ]^2 + 2\kappa[FRQ_2], \quad (13)$$

$$\frac{d[FRQ_2]}{dt} = \eta[FRQ]^2 - \kappa[FRQ_2] - \gamma[FRQ_2][WCC] - \delta_{FRQ}[FRQ_2], \quad (14)$$

$$\frac{d[WCC]}{dt} = \beta_-[RNA_W] + \beta_+[RNA_{W+}] - \gamma[FRQ_2][WCC] + \theta[frq:WCC] - \alpha[frq][WCC] - \delta_{WCC}[WCC], \quad (15)$$

$$\frac{d[FRQ_2:WCC]}{dt} = \gamma[FRQ_2][WCC] - \delta_T[FRQ_2:WCC]. \quad (16)$$

A schematic representation of the second two-loop model is presented in Fig. 2 B.

Parameters choice

A logarithmic plot of the *frq* transcript concentration from Garceau et al. (1997; not shown) shows that its decay is exponential. The first-order degradation rate is of the order of 0.2 h^{-1} . This value was already proposed by Ruoff et al. (1999) with a fit of the behavior predicted by the Goodwin model. However, it is not possible in the present model to know if this exponential decay is due to the real degradation rate of RNA or simply to the detachment constant of WCC from the *frq* promoter. Mathematical analysis of the one-loop model (P. François, unpublished) shows that the equation for RNA decay is $r(t) = A \exp(-\theta t) + B \exp(-\delta_{RNA} t)$, where A and B are two constants. If $\theta > \delta_{RNA}$, the RNA decay is mainly directed by its own degradation (parameter δ_{RNA}), since, after a short time, $\exp(-\delta_{RNA} t) \gg \exp(-\theta t)$. If $\theta < \delta_{RNA}$, it is directed by the dynamics of the detachment of WCC from the FRQ promoter (parameter θ). We therefore simulated both behaviors and saw no major qualitative difference between the models.

Experiments from Lee et al. (2000) provide the WC-1 degradation rate, and show that in the presence of FRQ, the WC-1 degradation rate is not affected by FRQ. Therefore, the degradation rate of the multimer WCC-FRQ δ_T is almost the same as the WC-1 degradation rate δ_{WCC} . Examination of the Western blots provides an approximate value of this rate. The WC-1 concentration is divided by ~ 3 within 4 h. This gives $\delta_{WCC} \approx 0.3 \text{ h}^{-1}$.

For the FRQ degradation rate, several parameters were tested, with no qualitative differences. Actually, it seems difficult to find its precise value from experimental data, as in the present models, the FRQ concentration decrease is directed by at least three parameters: θ , δ_{RNA} , and δ_{FRQ} . We therefore tested two set of parameters for the one-loop model: one with $\delta_{FRQ} = 0.05 \text{ h}^{-1}$, the other with $\delta_{FRQ} = 0.25 \text{ h}^{-1}$. The behaviors of the transcripts and the proteins for these two sets of parameters are similar. For the first two-loop model, $\delta_{FRQ} = 0.05 \text{ h}^{-1}$ gave the best results. For the second two-loop model, we chose $\delta_{FRQ} = 0.3 \text{ h}^{-1}$, a value similar to the WCC degradation rate and close to the value proposed by Ruoff et al. (1999) from fits of the Goodwin model.

It is not possible to deduce the values of the other parameters from curves with linear rise and decay without knowing absolute concentrations. Actually, in the models, it is possible to rescale parameters to obtain almost any absolute concentrations. For instance, in the one-loop model, if we multiply both ρ_{WCC} and β by a same constant c and divide γ and α by the same c , the qualitative dynamics of the system will not be changed, despite the change of absolute concentrations. We therefore chose parameters so that both kinetic constants and proteins concentration seem in a physiological range, and fit the oscillator period and the experimental curves. Similar oscillations occur for a very large set of parameters, so that the qualitative behavior of the oscillator is mostly independent of the choice of parameters.

Numerical methods

Integration of differential equations was performed with a Runge-Kutta algorithm. The time step was reduced until no significant difference in simulations appeared after further reduction. To ensure that the real asymptotic limit-cycle was observed, simulations were performed until no difference appeared between successive oscillations.

All programs were written in C++.

RESULTS

The one-loop model is considered and studied (Eqs. 1–5) in the sections “A simple model with only mass-action laws can simulate the *Neurospora crassa* circadian oscillator” and “RNA control and mechanism of repression”. The section “Models with a positive feedback loop give different qualitative behaviors for WCC” is devoted to the analysis of

the two-loop models. The section ‘‘Parameters variation, period dependence, and compensation’’ is devoted to an analysis of parameter dependence and to a comparison between models. The section ‘‘Phase response curve’’ studies the light response of the second two-loop model.

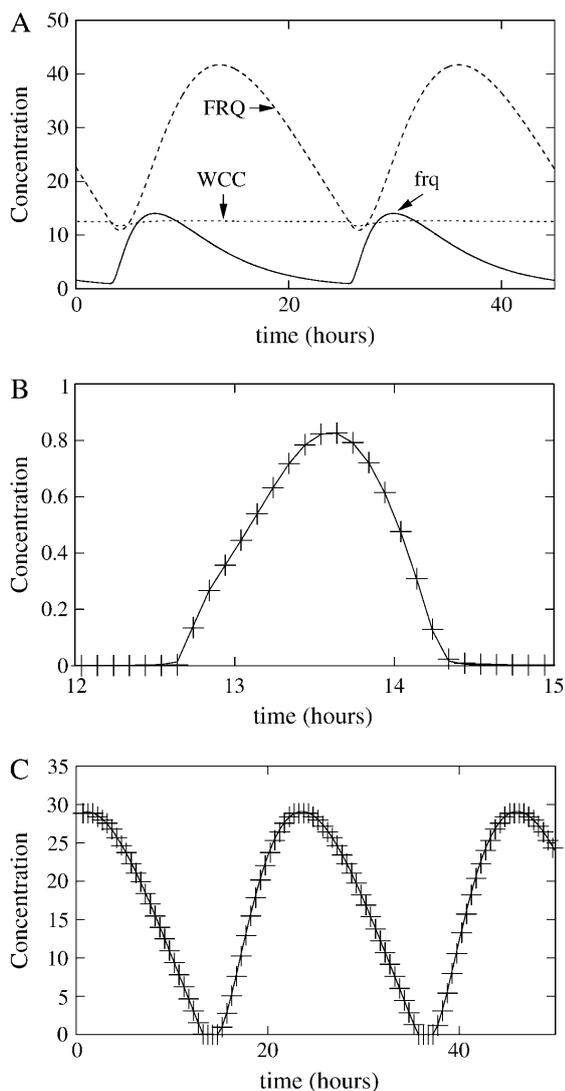


FIGURE 3 Simulation of the one-loop model of *Neurospora* circadian clock, with $\delta_{RNA} < \theta$. (A) Absolute concentration of proteins and transcripts. (B, solid line) Free WCC, + symbol: plot of $(\beta[RNA] - \delta_{FRQ}[FRQ]) / (\gamma[FRQ])$, confirming that we can make a quasistatic assumption to relate $[FRQ]$ to $[WCC]$. (C, solid line) Free FRQ (absolute concentration), + symbol: plot of $(\rho_{WCC} - \delta_{WCC}[WCC]) / (\gamma[WCC])$, confirming that we can make a quasistatic assumption to relate $[WCC]$ to $[FRQ]$. Constants for this first model are $\rho_{WCC} = 3.75 \text{ mol h}^{-1}$, $\rho_{FRQ} = 7.5 \text{ mol h}^{-1}$, $\beta = 0.7 \text{ h}^{-1}$, $\theta = 0.35 \text{ h}^{-1}$, $\alpha = 10 \text{ mol}^{-1} \text{ h}^{-1}$, $\delta_{WCC} = 0.3 \text{ h}^{-1}$, $\delta_{RNA} = 0.2 \text{ h}^{-1}$, $\delta_{FRQ} = 0.05 \text{ h}^{-1}$, and $\gamma = 2000 \text{ mol}^{-1} \text{ h}^{-1}$, where *mol* stands for molecules and *h* for hours. Degradation constant of the complex is the same as WC-1. Parameters were chosen to have oscillations qualitatively similar to the experimental curves presented in Fig. 1 A.

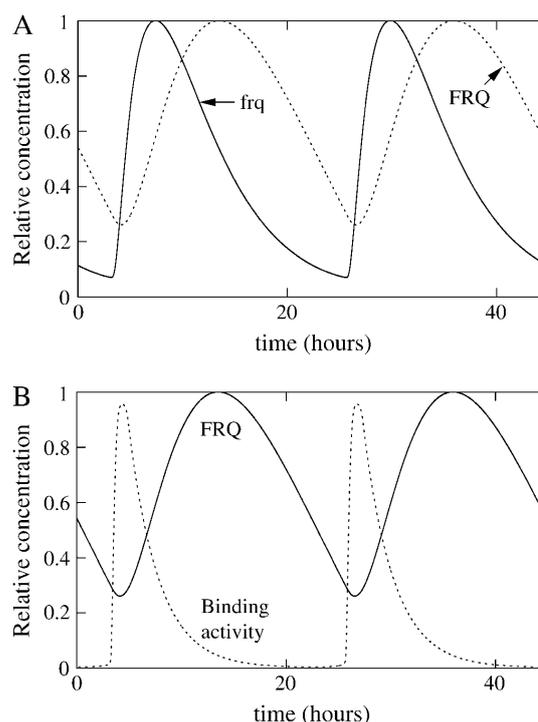


FIGURE 4 Simulation of the one-loop model of *Neurospora* circadian clock, with parameters of Fig. 3. Concentrations on A and B are rescaled by their maximum values for comparisons with experimental curves. (A, dotted line) Total FRQ (free + complexed). (A, solid line) FRQ RNA. (B, solid line) Total FRQ. (B, dotted line) Binding activity on *frq* promoter.

A simple model with only mass-action laws can simulate the *Neurospora crassa* circadian oscillator

The *Neurospora* one-loop model simulates oscillations of the levels of *frq* transcripts and FRQ proteins (Fig. 3 A and Fig. 4 A). The delay between the *frq* transcripts and the FRQ proteins peaks is ~ 6 h, in agreement with experimental observation. The decay of *frq* transcripts is exponential and requires 18 h in agreement with the experiments. The de-repression phase, when *frq* transcripts rise to their peak level, is ~ 4 h long. The behavior of the concentration of FRQ proteins is clearly not a simple exponential. This seems also to be the case for the experimental curves. Finally, WCC is observed to rhythmically bind to *frq* promoter (Fig. 4 B).

The oscillator dynamics can be separated between two phases, following the denomination of Merrow et al. (1997).

De-repression phase

During this phase, free (not complexed with WCC) FRQ concentration is low and free WCC concentration is high. The newly formed FRQ quickly interacts with the free WCC, which is in large excess. A consequence of this interaction is that the concentration of free FRQ is proportional to the

concentration of *frq* transcripts, and inversely proportional to free WCC concentration, as shown by the comparison between WCC concentration and this quasistatic assumption provided in Fig. 3 B.

At the same time, the free WCC interacts with the *frq* gene promoter. The concentration of *frq* transcripts consequently rises exponentially. When *frq* transcripts reach high concentration, almost all free WCC disappears and FRQ concentration can rise again. At the end of this phase, *frq* transcripts are near their maximum level.

Repression phase

FRQ free concentration is now high whereas WCC free concentration is low. This time, the newly formed WCC immediately interacts with the free FRQ in excess, produced with a high rate because of the high concentration of *frq* transcripts. A consequence of this interaction is that concentration of free WCC is inversely proportional to free FRQ concentration, as shown by the comparison between FRQ concentration and the quasistatic assumption provided in Fig. 3 C.

At the same time, bound WCC is released in an exponential way from the *frq* promoter. The *frq* transcription rate decays in the same way, and transcripts are degraded in an exponential way. As the FRQ production rate is proportional to transcript concentration, when the transcript level is low, the production rate becomes too low, and finally all the free FRQ is degraded. WCC can once again accumulate, and a new cycle begins.

For comparison with the experimental curves, the total (free + complexed) concentration of proteins should be taken into account. In the one-loop model (Eqs. 1–5), this concentration is almost constant for total WCC, because the degradation constant of free WCC is the same as the degradation constant of the complex (Fig. 3 A). The FRQ curves in the model are similar to the experimental ones (compare Fig. 1 A with Fig. 4 A).

RNA control and mechanism of repression

If the dynamics of WCC production and of the dimerization is fast enough compared to the characteristic time constants of *frq* transcripts, there is, effectively, a dynamical switch between WCC and FRQ. The low concentration species is in quasiequilibrium, and its dynamics is slaved to the high concentration species.

In simple terms, when WCC concentration is higher than FRQ concentration, WCC proteins titrate all free FRQ and after a very short time, only the free WCC, with dimers of WCC and FRQ, remains. Inversely, when FRQ concentration is higher than WCC concentration, FRQ proteins titrate all the WCC and after a very short time, the free FRQ, with dimers of WCC and FRQ, remains. Consequently, both proteins cannot be simultaneously present in uncomplexed form in the cell with comparable concentrations, and part of

the protein in excess and all the low concentration proteins are stored in the complex. A first consequence of the dimerization is, therefore, that the dynamics of *both* free proteins is controlled by the concentration of *frq* transcripts: When FRQ is in excess, free FRQ concentration is controlled by the concentration of the *frq* transcripts, and controls WCC free concentration thanks to the dimerization; and when WCC concentration is high, the FRQ production rate is proportional to the concentration of the *frq* transcripts. The produced FRQ proteins quickly dimerize with free WCC so that the free WCC sequestration rate actually is proportional to the transcript concentration.

Finally dimerization explains the repression mechanism: when FRQ is present at high concentration, it titrates WCC and therefore prevents its binding to the *frq* promoter.

To sum up, the core mechanism of the clock can be reduced to two coupled mechanisms—a slow dynamical process composed by all the transcription machinery, mathematically described by Eqs. 1 and 2, coupled to a rapid switch at the protein level, mathematically described by Eqs. 3 and 4. This switch, in turn, controls the slow process through the transcriptional activation.

Models with a positive feedback loop give different qualitative behaviors for WCC

The one-loop model does not take into account the regulation of WC-1 by FRQ, so that the WC-1 level is almost constant. Introduction of the experimentally observed second feedback loop is needed to explain WC-1 oscillations. It has been established that the activation of WC-1 by FRQ is mediated through a post-transcriptional mechanism (Lee et al., 2000). Here, we supposed that the translation of *wc-1* transcripts is enhanced by FRQ.

In experiments, the WC-1 concentration peak is out of phase with the FRQ concentration peak, so that this activation seems delayed. One simple way of modeling such a phase shift is to introduce a phenomenological delay in the equations, which effectively describes some biological processes such as cellular transport, for instance.

This strategy is followed in our first two-loop model in Eqs. 6–9 (see also Fig. 5). It gives realistic amplitudes for both FRQ and WCC oscillations. However, the delay length is critical, to qualitatively match the WCC behavior seen in experiments and to quantitatively account for the delays between peaks and for the period length. In the present model, to obtain a realistic behavior, the delay needed to be set to 7 h. As this delay is quite long, we therefore tried to find possible mechanisms explaining it.

Several attempts were necessary. We first supposed that FRQ enhanced translation of WC-1 by binding to *wc-1* transcripts, and that the enhancement was due to a low concentration enzyme mediating the binding of FRQ to *wc-1* transcripts. With the help of this enzyme, WCC production was delayed so that WCC levels oscillated with the right

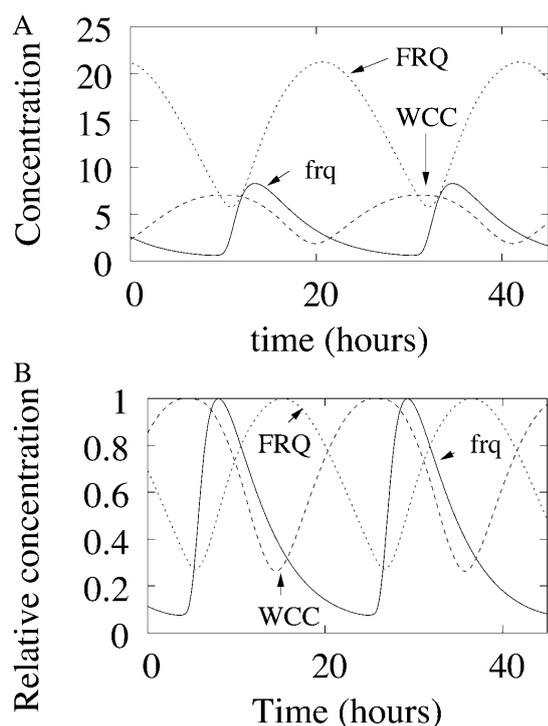


FIGURE 5 (A) Simulation of the first two-loop model with a post-transcriptional activation of WC-1 by FRQ. Oscillations of RNA (solid line), total WCC (dashed line), and total FRQ (dotted line). (B) Rescaled concentrations. Constants for this model are $\rho_{WCC} = 0.3 \text{ mol h}^{-1}$, $\rho_{FRQ} = 10 \text{ mol h}^{-1}$, $\beta = 0.6 \text{ h}^{-1}$, $\theta = 0.6 \text{ h}^{-1}$, $\alpha = 1 \text{ mol}^{-1} \text{ h}^{-1}$, $\delta_{WCC} = 0.3 \text{ h}^{-1}$, $\delta_{RNAw} = 3 \text{ h}^{-1}$, $\delta_{RNA} = 0.2 \text{ h}^{-1}$, $\delta_{FRQ} = 0.05 \text{ h}^{-1}$, $\gamma = 100 \text{ mol}^{-1} \text{ h}^{-1}$, $\delta_T = 0.3 \text{ h}^{-1}$, $\nu = 0.2 \text{ mol}^{-1} \text{ h}^{-1}$, $\mu = 0.01 \text{ h}^{-1}$, $\beta^- = 1 \text{ h}^{-1}$, $\beta^+ = 40 \text{ h}^{-1}$, and $\tau = 7 \text{ h}$, where *mol* stands for molecules and *h* for hours. Degradation constant of the complex is the same as WC-1. Parameters were chosen to have oscillations qualitatively similar to the experimental curves presented in the top and third panels of Fig. 1.

phase, but the amplitude of the oscillation was far too low (data not shown). Second, as FRQ is known to form homodimers (Cheng et al., 2001a), the enhancement was supposed to be mediated only by FRQ dimers. As a consequence, WCC was produced essentially when FRQ levels were high, so that WCC peak was only slightly delayed, and FRQ and WCC reached low values simultaneously (data not shown). From the analyses of these examples, it appeared that the simplest way to have out-of-phase oscillations between FRQ and WCC was to suppose that FRQ activation was possible at low concentrations of FRQ, but not at high concentrations of FRQ.

The second two-loop model, shown in Eqs. 10–16, is an attempt to model such a mechanism. In this model, there are no explicit delays. Instead, we suppose that FRQ monomers bind to *wc-1* transcripts to enhance their translation and that, on the contrary, binding of FRQ dimers to *wc-1* transcripts represses their translation. Consequently, at high concentration, FRQ homodimerization prevents WCC translation.

At low concentrations of FRQ, FRQ proteins essentially exist as monomers and can bind to *wc-1* transcripts to strongly enhance their translation, so that, as shown in Fig. 6, WCC peak occurs just after FRQ minimum, which explains the observed out-of-phase relationship between FRQ and WCC. If this homodimerization is switched off, the oscillations disappear and WC-1 is overexpressed (data not shown).

Parameters variation, period dependence, and compensation

Dimensionless one-loop system

All our proposed models are basically refinements of our one-loop model. A precise study of the one-loop model is therefore useful to understand some basic properties of these models. We summarize in the following the main results of a mathematical analysis of this model. The complete study will be published elsewhere.

To gain a better understanding of the model parameter dependence it is first useful to rescale variables. In the following, we set first $F = \sqrt{\gamma/\rho_{WCC}}[\text{FRQ}]$, $W = \sqrt{\gamma/\rho_{WCC}}[\text{WCC}]$, and $r = (\beta[\text{RNA}])/(\rho_{WCC})$, and $g = [frq]$.

Second, we define the following rescaled parameters: let be $\delta = \delta_{RNA}/\sqrt{\rho_{WCC}\gamma}$, $a = \beta\rho_{FRQ}/(\rho_{WCC}\delta_{RNA})$, $b = \theta/\delta_{RNA}$, $c = \alpha\sqrt{\rho_{WCC}}/(\delta_{RNA}\sqrt{\gamma})$, $d = \sqrt{\gamma/\rho_{WCC}}$, $\delta_F = \delta_{FRQ}/\sqrt{\rho_{WCC}\gamma}$, and $\delta_W = \delta_{WCC}/\sqrt{\rho_{WCC}\gamma}$; and we rescale time by taking a new time unit $t_1 = \delta_{RNA}t$. Taking time t_1 as new time t , the new ODEs for the one-loop model are

$$\frac{dg}{dt} = b(1 - g) - cgW, \quad (17)$$

$$\frac{dr}{dt} = a(1 - g) - r, \quad (18)$$

$$\delta \frac{dF}{dt} = r - F \cdot W - \delta_F F, \quad (19)$$

$$\delta \frac{dW}{dt} = 1 - F \cdot W + d\delta(b(1 - g) - cgW) - \delta_W W. \quad (20)$$

As an example, parameters used for the model in Fig. 3 give $\delta = 2.3 \times 10^{-3}$, $a = 7$, $b = 1.75$, $c = 2.2$, $\delta_F = 5.8 \times 10^{-4}$, $d = 23$, and $\delta_W = 3.5 \times 10^{-3}$.

For the one-loop model, all parameters can be varied over one order of magnitude without destroying oscillations (data not shown). Noteworthy is that, if we vary the parameters of the one-loop model, there are several subcritical Hopf bifurcations so the oscillations often appear with finite amplitude. This kind of hysteretical transition has previously been proposed as a possible mechanism for noise resistance in genetic oscillators (Barkai and Leibler, 1999). A consequence of these subcritical Hopf bifurcations is that, depending on the initial conditions, a stable limit-cycle can coexist with a stable fixed point. This phenomenon is illustrated in Fig. 7 for the full one-loop model.

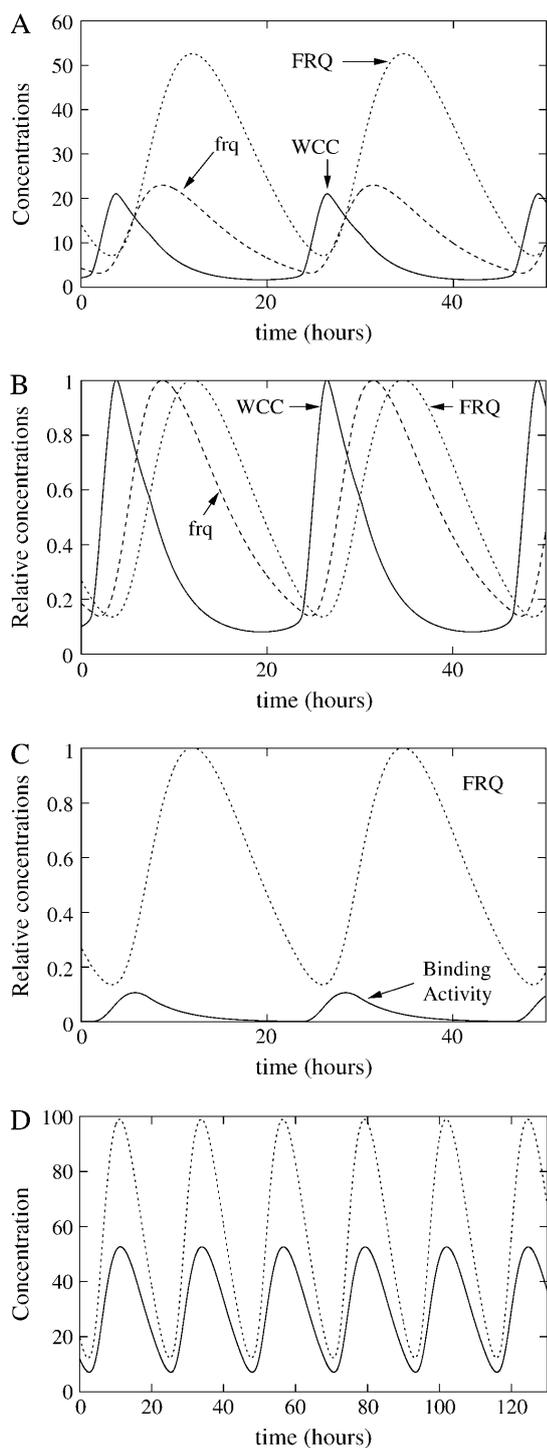


FIGURE 6 Simulation of the second two-loop model with a post-transcriptional activation of WC-1 by FRQ. (A) Absolute concentrations of total FRQ (including all complexes and dimers), total WCC, and *frq* transcripts. (B) Rescaled oscillations of RNA, WCC, and FRQ. (C) Binding of WCC of frequency promoter (solid line): binding activity (dashed line): FRQ concentration. (D) Influence of WCC transcription rate over FRQ oscillations (solid line): $\rho_{WCC} = 2.5 \text{ mol h}^{-1}$, and (dashed line): $\rho_{WCC} = 5 \text{ mol h}^{-1}$. Constants for this model are $\rho_{WCC} = 2.5 \text{ mol h}^{-1}$, $\rho_{FRQ} = 75 \text{ mol h}^{-1}$, $\beta = 1 \text{ h}^{-1}$, $\theta = 0.25 \text{ h}^{-1}$, $\alpha = 0.003 \text{ mol}^{-1} \text{ h}^{-1}$, $\delta_{WCC} = 0.3 \text{ h}^{-1}$, $\delta_{RNA} = 0.2 \text{ h}^{-1}$, $\delta_{FRQ} = 0.3 \text{ h}^{-1}$, $\gamma = 1600 \text{ mol}^{-1} \text{ h}^{-1}$, $\delta_T = 0.3 \text{ h}^{-1}$,

The smallness of the parameters δ , δ_F , and δ_W makes it possible to use matched asymptotic methods and to compute theoretically several properties of this system, as will be reported elsewhere. We summarize the main results of this analysis for the period and amplitude of the oscillator:

1. *Period of the oscillation.* At the lowest order in the small parameter δ , five parameters (δ_{RNA} , a , b , d_w , and d_f , with $d_f = \delta_f/\delta$ and $d_w = \delta_w/\delta$) are crucial for period determination. The period is given by $T = \delta_{RNA}^{-1} T_r(a, b, d_f, \text{ and } d_w)$. The value a corresponds to the ratio of the WCC protein production rate over the FRQ protein production rate, taking into account both transcription and translation (ρ_{WCC} is an effective rate that can actually be seen as: *WCC transcription rate* \times *WCC translation rate* \div *Degradation rates of the wc-1 transcripts*). The value b is the ratio between the release constant of WCC by the *frq* promoter and the degradation rates of the *frq* transcripts. The values d_f and d_w are the respective ratios of FRQ and WCC degradation rates over the degradation rates of the *frq* transcripts.
2. *Amplitude of the oscillation.* In the limit of small δ , the amplitude of the oscillations of free proteins scales as $1/\delta$ (Fig. 8).

Temperature dependence

A remarkable property of circadian clocks is their period-independence to temperature, together with the fact that they can still be reset or entrained by temperature pulses. The previously reported study can help in understanding this temperature compensation. Following Ruoff and Rensing (1996), temperature can be introduced into the one-loop model by Arrhenius equations. First, the important parameter a does not depend on temperature if the activation energies of transcription and translation do not depend on the nature of proteins. In this case, the only temperature-dependent parameters involved in the period determination are δ_{RNA} and b , and d_w , d_f if the WCC and FRQ degradation constant are not negligible. Supposing that

$$\begin{aligned} \delta_{RNA} &= \delta_{\infty} \exp(-E_r/RT) \\ \theta &= \theta_{\infty} \exp(-E_{\theta}/RT), \end{aligned}$$

we find $b = \delta_{\infty}/\theta_{\infty} \exp(-(E_r - E_{\theta})/RT)$, so that b remains constant if $E_r = E_{\theta} = E$. If we impose that the period does not vary $>5\%$ when temperature varies between $T_0 = 301 \text{ K}$ and $T_1 = 311 \text{ K}$, we find $E < 4000 \text{ J/mol}$, which seems a reasonable assumption and in the range of values proposed by Ruoff and Rensing (1996). Taking into account the

$\delta_{RNAw} = 1 \text{ h}^{-1}$, $\nu = 8000 \text{ mol}^{-1} \text{ h}^{-1}$, $\mu = 0.1 \text{ h}^{-1}$, $\beta^- = 0.001 \text{ h}^{-1}$, $\beta^+ = 10 \text{ h}^{-1}$, $\eta = 3000 \text{ mol h}^{-1}$, and $\kappa = 10 \text{ h}^{-1}$, where *mol* stands for molecules and *h* for hours. Degradation constant of the complex is the same as WC-1. Parameters were chosen to have oscillations qualitatively similar to the experimental curves presented in the top panel of Fig. 1.

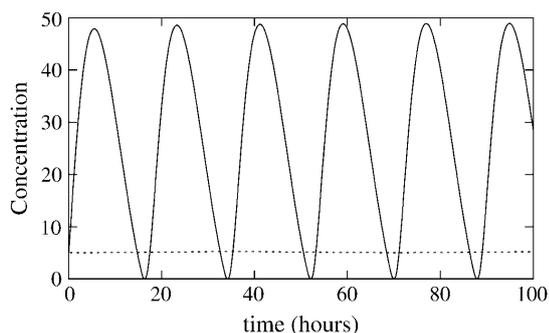


FIGURE 7 Coexistence of a stable limit cycle and a stable fixed point. (Solid line) Free FRQ for a first set of initial conditions. (Dotted line) Free FRQ for a second set of initial conditions. First set of initial conditions is $[WCC] = 0.0017$, $[RNA] = 10$, $[FRQ] = 5$, and $[frq] = 0.54$. Second set of initial conditions is $[WCC] = 0.0022$, $[RNA] = 4.44$, $[FRQ] = 5$, and $[frq] = 0.9$. Units are molecules per cell. Constants for this model are $\rho_{WCC} = 9 \text{ mol h}^{-1}$, $\rho_{FRQ} = 9 \text{ mol h}^{-1}$, $\beta = 2.25 \text{ h}^{-1}$, $\theta = 0.4 \text{ h}^{-1}$, $\alpha = 20 \text{ mol}^{-1} \text{ h}^{-1}$, $\delta_{WCC} = 0.3 \text{ h}^{-1}$, $\delta_{RNA} = 0.2 \text{ h}^{-1}$, $\delta_{FRQ} = 0.2 \text{ h}^{-1}$, and $\gamma = 800 \text{ mol}^{-1} \text{ h}^{-1}$, where mol stands for molecules and h for hours. Degradation constant of the complex is the same as WC-1.

degradation of the proteins (if they are not negligible) imposes a similar condition: the activation energy of the degradation rate must be of the same order as the activation energy of RNA degradation rate.

Another possibility is to consider that temperature has very little influence on dissociation of WCC from *frq* gene and over the degradation rates. This assumption is discussed below.

Finally, if production rates are modified while keeping the parameter a constant, the phase of the clock shifts, so that the oscillator can be entrained to a 24-h period (data not shown). So, importantly, the clock can be temperature-compensated and still entrained by changes in temperature.

Comparison between models: robustness to parameter variation

To gain insight into parameter dependence and to compare models, we doubled and halved constant rates for each reaction, one at a time, keeping the others constant. Results

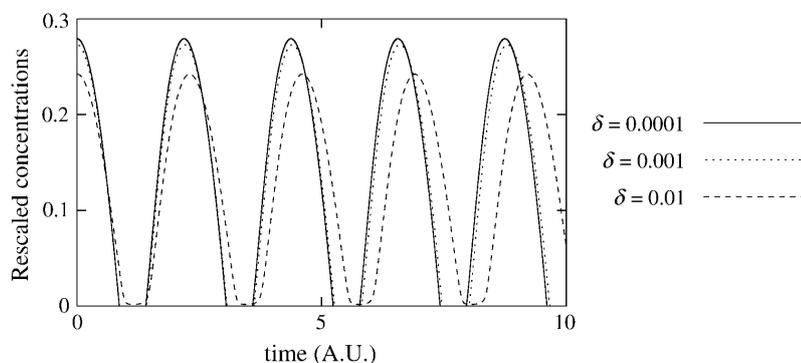


FIGURE 8 Scaling of the amplitude of the oscillator as a function of δ for the adimensioned system. Amplitude of (δW) is plotted as a function of time t (not rescaled). The amplitude of δW does not depend on δ in the limit of small δ , this shows that the amplitude of W scales as $1/\delta$. Parameters are $a = 3$, $b = 5$, $c = 10$, $d = 1$, and $\delta_w = \delta_f = 0$.

of this computation are given in Fig. 9 for two typical sets of parameters of a one-loop model and for both of the two-loop models.

For the first set of parameters for the one-loop model, the most sensitive parameters are transcription and translation rates (ρ_{WCC} , ρ_{FRQ} , β), as well as the degradation rate of *frq* transcripts and of FRQ proteins and the detachment rate θ of WCC from the *frq* promoter, with period variations from 13% to 63%. For other parameters, the period never varied by $>3\%$ from control. This dependence is explained by the mathematical analysis summarized before. For the first one-loop model, amplitude variations are correlated to period variations: in most cases, period and amplitude of the oscillations vary with the same order of magnitude.

We also tested a set of parameters with a higher FRQ degradation constant. The main dynamical consequence of this second choice of parameters is that the binding of WCC to FRQ promoter is much weaker (i.e., no quasistatic in the activation phase) than in the first set of parameters. With this second set of parameters, WCC proteins do not saturate the FRQ promoter. As can be seen on Fig. 9B, this enhances the global robustness to parameter variations, and the most sensitive parameters are δ_{RNA} , θ , and δ_{FRQ} .

The first two-loop model is not very robust to parameter variations. The oscillations disappear after modifications of ρ_{WCC} , *wc-1* transcription rate, and of four new parameters: the delay τ for the activation of translation, the *wc-1* transcript degradation rate δ_{RNAw} , the enhanced translation rate β_+ of these transcripts, and the rate of interaction between FRQ proteins and *wc-1* transcripts. All these parameters are implicated in the delaying processes in the positive feedback-loop. The other sensitive parameters are the same as in the one-loop model.

The second two-loop model is far more robust to parameter variations. The three most sensitive parameters for the period are θ , the rate for the release of WCC from the *frq* promoter; and the degradation rates of *frq* transcripts and FRQ dimers, with variations of the period from 20% to 27%. Doubling the WCC proteins' degradation rate also gives a period that is 15% shorter. For all other parameter variations, the period does not vary by $>9\%$, as can be

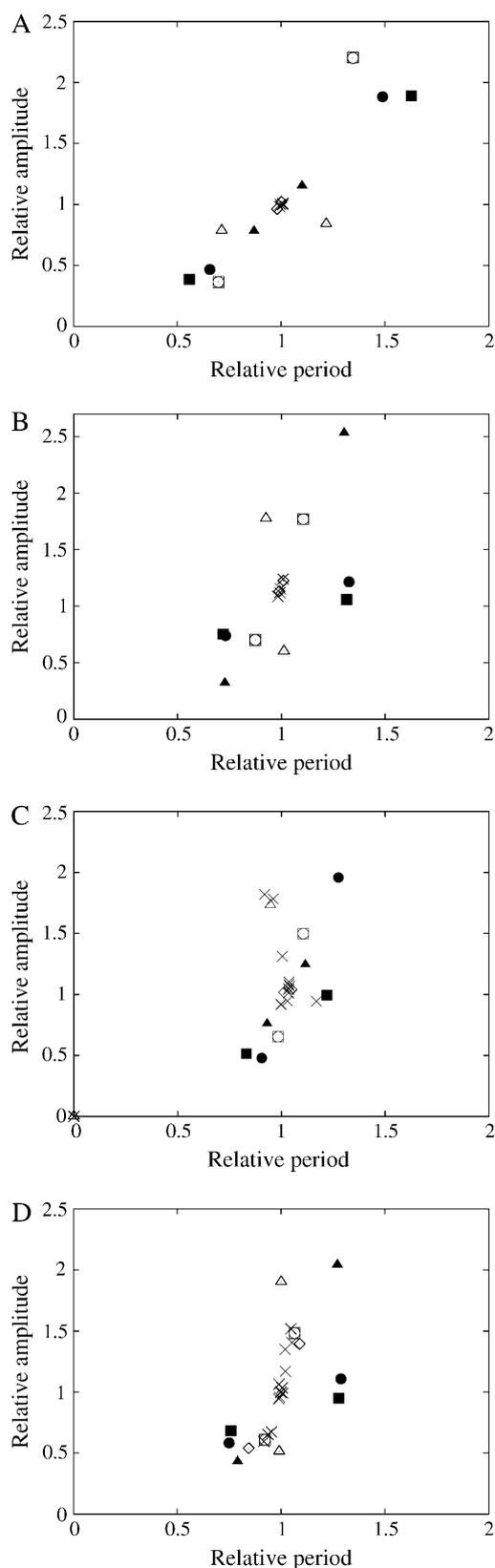


FIGURE 9 Effect of strong parameter variations on oscillations. Each point represents a simulation where one parameter has been doubled or

seen on Fig. 9 *D*. Contrary to the one-loop model, it is possible to have large amplitude variations without modifying the period of the clock. The amplitude of the second two-loop model still depends on synthesis rates, but its period is much less sensitive (compare parts *A* and *D* of Fig. 9). For instance, a doubling of the WCC transcription rate modifies the amplitude of the oscillations without modifying the period, as can be seen in Fig. 6 *D*.

Possible role of the positive feedback loop

Experiments show that in *Neurospora*, it is possible to have large variations of the amplitude of the oscillations while keeping the period constant (Liu et al., 1998; Cheng et al., 2001b). According to the previous analysis in the one-loop model, if transcription or translation rates for both proteins are multiplied by the same factor f , a remains constant, and the period does not change; but from the expression of δ , the amplitude of the adimensioned variable is multiplied by \sqrt{f} , and the real amplitude of the protein by f . Cheng et al. (2001b) proposed that one possible role of the positive feedback loop was to precisely adjust protein production rates to keep the period constant. The second two-loop model supports this suggestion: when synthesis rates of proteins are modified, the oscillator adjusts itself to keep the period constant, as can be seen in Fig. 6 *D* and Fig. 9 *D*. When the parameter ρ_{WCC} is multiplied by 4, the diminution of the period is $<2\%$ of the reference period, whereas the amplitude is more than three times higher; when it is multiplied by eight, the diminution of the period is $<7\%$ and the amplitude is approximately five times higher (data not shown). For comparison, for the second set of parameters for the one-loop model, when the same parameter is multiplied by 4, the period is 25% lower than the reference period, and when it is multiplied by 8, the period is 50% lower than the reference period. The second two-loop model consequently shows that specific biochemical mechanisms in the positive feedback loop could help in keeping the period constant, despite changes in the amplitude.

Phase response curve

The precise biochemical processes mediating the response of *Neurospora* circadian clock to light pulses are still not

halved, keeping the other parameters constant. Amplitude and period are plotted relative to the amplitude and the period of the control. Measured amplitude is the amplitude of total FRQ oscillation. (solid triangle, δ_{RNA} ; solid circle, θ ; solid square, δ_{FRQ} ; open triangle, ρ_{WCC} ; open square, ρ_{FRQ} ; open circle, β ; open diamond, δ_{WCC} ; and \times symbol, other parameters.) (A) One-loop model (parameters of Fig. 3). (B) One-loop model (with $\delta_{\text{FRQ}} = 0.25 \text{ h}^{-1}$, $\rho_{\text{FRQ}} = 20 \text{ mol h}^{-1}$, $\theta = 0.23 \text{ h}^{-1}$, and $\alpha = 4 \text{ mol}^{-1} \text{ h}^{-1}$, with other parameters the same as Fig. 3). (C) First two-loop model (parameters of Fig. 5). Five parameters destroy oscillations. (D) Second two-loop model (parameters of Fig. 6).

completely known. It was shown first that the effect of light pulses is to switch off the negative feedback loop (Crosthwaite et al., 1995). A light pulse first greatly increases the production of the *frq* transcripts (fourfold to 25-fold as compared to the average level during one cycle). Then, these newly formed transcripts are quickly degraded compared to normal transcripts (with half-lives of the order of 1 h). It was more recently shown that there are two specific binding sites for light response in the *frq* promoter called light-response elements (Froehlich et al., 2002). And finally, another negative feedback loop, implicating at least one gene, called *vivid*, has been discovered. VIVID seems to negatively regulate (but not to fully control) the gating of light input, probably via hyperphosphorylation of WC-1 (Heintzen et al., 2001).

A precise model of this feedback loop is not possible yet because of the lack of more precise experimental data. It is possible, however, to test some light response properties of the *Neurospora* circadian clock. Dunlap and co-workers hypothesized that the phase of the clock was given by the concentration of the *frq* transcripts (Crosthwaite et al., 1995; Loros and Dunlap, 2001), and that the effect of light was to switch the concentration of the *frq* transcripts to its maximum. We tested this heuristic model by suddenly raising the concentration of the *frq* transcripts to its maximum at different times of the cycle. We also introduced a supplementary effect due to the other negative feedback loop: we supposed that one role of *vivid* was to trigger the degradation of WC-1 (as proposed by Heintzen et al., 2001) and set WCC total concentration to zero, also including a degradation of the WCC bound to the *frq* promoter. Light pulses at different times of the cycle delay (negative-phase shift) or advance (positive-phase shift) the oscillations by different amounts. Phase response curves (PRCs) show the phase shift that corresponds to light pulses at different times of the cycle.

The PRC was computed for the one-loop model and for the second two-loop model. These PRCs are very similar, and only the PRC for the two-loop model is shown in Fig. 10.

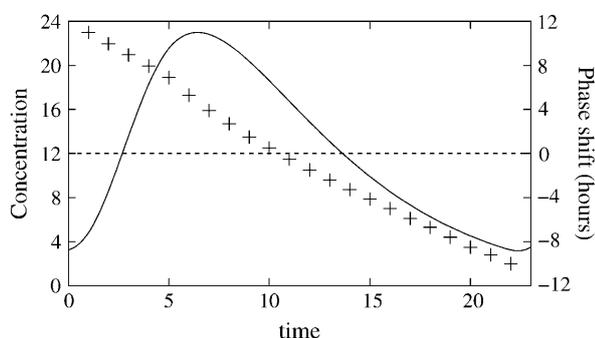


FIGURE 10 Phase response curve. (+ symbols) Response due to a sudden rise of RNA concentration to its maximum level with a simultaneous degradation of all WCCs at different times in the cycle. (Solid line) The variation of *frq* transcripts during a circadian cycle (without light pulses) is also shown for reference.

This PRC agrees qualitatively with the experimental observations (Crosthwaite et al., 1995) and supports the role of *frq* transcripts as a major determinant of the phase of the clock. When the level of the *frq* transcripts is rising, a light pulse produces a phase advance of the clock, as in this case the transcripts' peak is advanced by light. On the contrary, when the level of the *frq* transcripts is decaying, the qualitative behavior is a phase delay of the clock, as in this case the light-induced peak of the *frq* transcripts occurs after the normal peak in the cycle, and light makes the clock shift to the phase when the level of the *frq* transcripts is at maximum. This light response can be well explained by the previous analysis of the limit-cycle: the rapid degradation of WC-1 and rise of *frq* transcripts quickly shifts the clock to the beginning of the repression phase, when WC-1 no longer activates *frq* transcription and when there are enough *frq* transcripts to produce FRQ proteins in excess compared to WC-1 proteins. FRQ proteins then sequester the newly produced WC-1, so that transcription is repressed similarly to what happens in the limit cycle.

DISCUSSION

The model explains oscillations, and can be improved to have robustness

The one-loop model (Eqs. 1–5) describes basic features of *Neurospora* circadian oscillations. Both qualitative and quantitative behaviors of *frq* transcripts and FRQ protein are reproduced by this model. Reactions at the level of the transcription were supposed to be slower than protein-protein reactions, which completely explain phase shifts and qualitative behaviors. The two phases described in Mellow et al. (1997) clearly appear: the repression phase corresponds to the phase where FRQ protein concentration is high, and the de-repression phase corresponds to the phase where free WCC protein concentration is high. Rhythmic binding is a natural consequence of this switch at the protein level.

The explicit distinction in the equations between transcription and translation of the RNA is necessary to explain experimental curves. Modeling protein production by a single effective step leads to the destruction of the oscillations, in the present one-loop model. Actually, the FRQ protein and the *frq* transcripts are seen experimentally to have different behaviors, since FRQ concentration variation does not simply reproduce messenger RNA variation after a time delay (Fig. 1 A). This experimental result clearly stresses the need for separate modeling of transcription and translation. The level of RNA has an unexpected consequence on the protein concentration: for instance, in the de-repression phase, the rate at which free WCC proteins are sequestered is controlled by the concentration of the *frq* transcripts.

Protein-protein interactions are at the core of the system, and should be taken into account explicitly. Such reactions

are not equivalent to simple repression at the transcriptional level: heterodimerization is essential to couple RNA and proteins dynamics, and therefore plays a dynamical role different from a simple repression.

The present one-loop model also helps us to understand the clock temperature compensation as described in Temperature Dependence (above); when the production rates of proteins are varied the same way, in the limit of small δ , the period does not change despite a change of the amplitude of oscillation.

The fact that FRQ oscillations are observed for the one-loop model shows that the second positive feedback loop is not necessary for the occurrence of oscillations, confirming a previous study (Smolen et al., 2002). To take into account WC-1 oscillations, another feedback loop is required. We tested the hypothesis of a direct activation at the post-transcriptional level, without any delays, mediated, for instance, by some intermediate enzymes (models not shown). However, it was impossible to obtain realistic out-of-phase oscillations for FRQ and WC-1 with such models. To obtain out-of-phase relationships between FRQ and WCC, it was necessary to suppose that FRQ represses WCC production at high concentration, and activates it at low concentrations. This was modeled by taking into account the homodimerization of FRQ (Cheng et al., 2001a). This hypothesis gives behaviors in agreement with experimental observations. Besides, if the transcription rate of WCC is raised, the oscillation amplitudes are higher but the period changes only slightly (Fig. 6 D). In this model, one of the roles of the positive feedback loop could therefore be to adjust protein production rates to keep the period constant, despite amplitude changes, as proposed by Cheng et al. (2001b).

A light-PRC was also computed. It was shown that taking into account both the production of the *frq* transcripts and the hyperphosphorylation of WC-1 proteins explains the shifts of the clocks. This model of light influence confirms Crosthwaite et al. (1995), and the role of *frq* transcripts in the phase determination of the clock.

Comparison with other works

Even if major components of circadian clocks have been well-described experimentally, the dynamical origin of the oscillations remains quite unclear. Actually, most models of circadian clocks can be classified in two categories: models where delays are necessary to oscillations; and models where oscillations only depend on the specific assumptions made about the genetic interactions.

Examples of models with delays have been proposed by Smolen et al. (2001, 2002) and Goldbeter and co-workers (Gonze et al., 2000; Leloup and Goldbeter, 1998, 2003).

Smolen et al. (2001) hypothesize that the delays observed in circadian clocks are consequences of slow biological

processes (due to transcription, translation, or cellular transport, for instance) and use delayed differential equations to model circadian oscillations with phenomenological delays accounting for these mechanisms. The main conclusion of their models is that a positive feedback loop is not necessary to have oscillations, but that long time-delays (7 h in the *Neurospora* case) are necessary to account for oscillations.

In the present one-loop model, we explicitly modeled transcription and translation. This one-loop model shows that explicit delays in the equations are not necessary to produce oscillations. Then we took into account the second feedback loop to better explain the biological data. Two models were formulated: a model with explicit delays and a model without delays but with supplementary biophysical interactions. These two models present the same qualitative behavior. However, their properties are different. We showed, for instance, that our second two-loop model is far more robust to parameter variations than our first two-loop model. This means that it may not be possible to reduce the second two-loop model to a simplified version such as the first two-loop model with delays, without destroying some important properties of the model.

Goldbeter and co-workers (Gonze et al., 2000; Leloup and Goldbeter, 1998, 2003) have also intended to explicitly model the delaying mechanisms. In the *Neurospora* case, WCC activity has not been considered, but in the mammalian case, the corresponding proteins dynamics (BMAL, CLOCK) has been modeled. Goldbeter and co-workers have hypothesized that nuclear transports and successive phosphorylations observed in most of circadian clocks are at the origin of delays and are necessary for the oscillation. Only the hyperphosphorylated form of the proteins has been supposed to form heterodimers to repress transcription. Actually, experimental studies showed that FRQ is quickly phosphorylated (Garceau et al., 1997) and that its phosphorylation rate determines its degradation rate (Liu et al., 2000). Also, hypophosphorylated FRQ is also known to be able to bind to WCC (Yang et al., 2002). Therefore, in the present model, phosphorylation has been supposed to fix the degradation rate of FRQ, which is one of the most important parameters for the determination of the period length.

Some other models do not introduce slow processes, and these models suppose that specific interactions in the genetic network help in destabilizing the fixed point.

The Ruoff-Rensing model (Ruoff and Rensing, 1996) is essentially based on the Goodwin model (Goodwin, 1965). Transcription and translation are explicitly modeled. As in the present model, a slow amplification process (transcription and translation) is coupled with a rapid switch, at the transcriptional level. Dynamics of repression is modeled by a Hill function accounting for fast kinetics. For the system to oscillate, a high Hill exponent (>9) is needed, which implies a very high cooperativity. This acts as a phenomenological switch, accounting for possible mechanisms of undescribed origin. The dynamics of the present model is close in spirit to

the mechanism suggested by the Ruoff-Rensing model, with a slow accumulation of RNA and proteins coupled to a rapid repression. For instance, the influence of degradation rates predicted by the present model is very similar to what was proposed before for the Goodwin oscillator (Ruoff et al., 1999) and was confirmed experimentally by Liu et al. (2000). However, the present model bypasses the need of high cooperativity by taking into account the interaction between FRQ and WCC at the post-transcriptional level.

Another interesting model was proposed for the *Drosophila* circadian clock (Tyson et al., 1999) with a goal similar to that of the present model: to provide a minimal model, simple to analyze and to improve. As in the present model, dimerization played an essential role, but in Tyson et al. (1999), the crucial positive feedback loop was a consequence of stabilization of PER induced by this dimerization. However, the two models are quite different: Tyson and co-workers concluded that a positive feedback loop was required to explain oscillations. Such a positive feedback loop is not needed when one does not make any quasiequilibrium assumptions on the dynamics of the proteins as shown by the one-loop model of the present article. We propose that the role of the positive feedback loop is, rather, to improve robustness to variations in parameters.

Finally, in the previously described models of circadian clocks, regulation of transcription was modeled by Hill or Michaelis-Menten kinetics, modeling fast binding between DNA and protein, and in most of the models, quasiequilibrium is also assumed for RNA dynamics. Kinetics at the level of transcription is therefore supposed to be very quick. This focus on proteins requires us to make specific assumptions on the dynamics of the networks to have oscillations. The present models show that if one models transcription and translation and does not make any quasiequilibrium assumptions, both oscillations and biological delays can be explained without any further hypothesis.

The models raise experimental questions

Testing the model

First, the protein-protein reaction between FRQ and WCC plays a crucial role in the dynamics of the system. This reaction should be fast. Irreversibility is not necessarily needed, but multimerization should be greatly favored. For instance, for the parameters of Fig. 3, the dissociation rate of the complex must not be higher than $\sim 3.4 \text{ h}^{-1}$ to have sustained oscillations (data not shown). A possible experimental indication of this fact would be to measure the ratio between complexed and total proteins. For the protein with the lower concentration, this ratio should be close to 1.

Second, a consequence of this dimerization is the influence of *frq* transcripts on the dynamics of the system. The decay of messenger RNA plays a major role in the period determination. Raising or lowering the degradation

time-constant of *frq* messenger RNA significantly changes the repression phase length and the qualitative behavior of the proteins. An alternative possibility is to modify the dynamics of binding of WCC to *frq* promoter. Mathematical analysis reveals that a lower detachment rate θ should lengthen the cycle if this rate is lower than the *frq* transcript's degradation rate, whereas a higher detachment rate should shorten the repression phase if this rate is higher than the *frq* transcript's degradation rate. The influence of θ seems difficult to test experimentally. However, the influence of transcript degradation rate could be easily tested, since it is, in principle, possible to alter the stability of *frq* transcripts by polyadenylation. One could thus test the correlation between the transcript's degradation rate and the period. One could also, for instance, imagine restoring the function of short period mutants (such as *frq*¹ or *frq*²; Feldman and Hoyle, 1973) by raising the transcript's stability. The present models also give an indication on the influence of FRQ degradation rate on the period of the clock: a lower degradation rate produces a longer period. This was already predicted and confirmed experimentally for the Goodwin oscillator applied to model the *frq*⁷ mutant (Ruoff et al., 1999; Liu et al., 2000). For the second two-loop model, dividing the degradation rate of FRQ dimers by 2 changes the period from 22 h to 29 h, as in *frq*⁷ strains, and *frq* transcript levels are also $\sim 32\%$ higher than reference, qualitatively in agreement with experiments (Aronson et al., 1994) (data not shown).

Finally, to explain the phase shift between proteins, we proposed that only the monomer form of FRQ is active to enhance WC-1 translation. A consequence of this hypothesis is that in mutants where this homodimerization is switched off, constitutive levels of WC-1 should be very high, and even higher than the maximum level of WC-1 in normal cells. A possible test would be to vary FRQ levels and see that the enhancement of WC-1 translation does not vary monotonically with FRQ total concentration. For a low production rate, when FRQ mostly exists as a monomer, WC-1 translation rate should be high. For a high FRQ production rate, when FRQ mostly exists as a dimer, levels of WC-1 should be lower. This hypothesis could be tested experimentally. Strains have been artificially designed, where *frq* promoter is under the control of quinic acid (QA) (Aronson et al., 1994), and it is therefore possible to continuously vary FRQ protein production rate, while evaluating WC-1 concentration. One should observe a high WC-1 response only for a medium concentration of QA.

Improving the model

The measure of absolute concentrations would provide important data to refine the modeling. As some evolutions seem more or less linear, the present experimental data is not sufficient to fit the parameter values without this important information. This is also very important for understanding the mechanism of repression: if the repression mechanism is

based on the heterodimerization which sequesters the WCC (Denault et al., 2001; Froehlich et al., 2003), the stoichiometry imposes constraints on the relative concentration of WCC and FRQ. In the present two-loop model, the WCC peak is approximately two-and-one-half times lower than FRQ peak. However, global extract (nucleus+cytosol) seems to show that FRQ and WCC peaks are approximately of the same order of magnitude (Denault et al., 2001).

To our knowledge, no nuclear extracts have been measured to evaluate this precise stoichiometry. Further hypotheses are therefore needed to explain this observed ratio. First, there could be specific different nuclear localization for the proteins, explaining a different ratio within the nucleus. Second, there could also be other negative feedback regulating the WCC level. Third, the stoichiometry of heterodimerization could be different from what is generally supposed: there is still no experimental evidence that FRQ dimers bind to only one WCC complex. Basically, if WCC binds to a FRQ dimer, twice as much FRQ protein is required for repression than if, for instance, two WCCs interact with an FRQ dimer. This phenomenon is illustrated in Fig. 11: we computed a model where twice as much WCC protein was produced, but where FRQ dimers interact with two WCC complexes. As can be seen, WCC maximum concentration is twice the WCC maximum concentration of the model computed in Fig. 6.

Another question is the temperature influence over the system. As explained before, it is possible to reduce the temperature dependence of the clock if we suppose that activation energies of transcription and translation are of the same order for different proteins. If the network functions in the parameter regime where δ is small, sensitivity of the oscillator depends on the degradation rates of all species and on the kinetics of the detachment of WCC from the promoter. The oscillator is temperature-compensated if these constants are not temperature-dependent, similar to the Ruoff-Rensing model (Ruoff and Rensing, 1996). We know that in the *Neurospora* circadian clock, as well as in other circadian

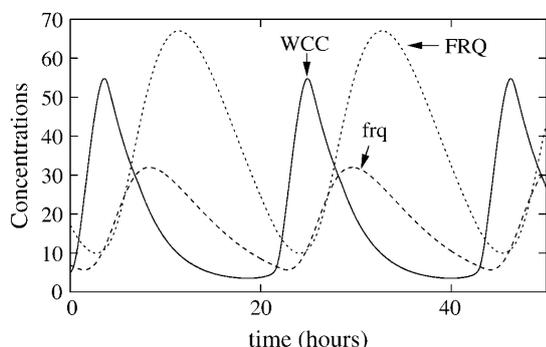


FIGURE 11 Effect of stoichiometry of the heterodimerization on the relative concentration of proteins. For this simulation, FRQ dimer is supposed to bind to two WCC proteins. Parameters are the same as in Fig. 6, with $\rho_{\text{WCC}} = 5 \text{ h}^{-1}$ and $\alpha = 0.0015 \text{ mol}^{-1} \text{ h}^{-1}$.

clocks, several species of RNA and proteins are produced from each gene, with various stability and chemical properties (Liu et al., 1997). Temperature compensation could be achieved by adjusting the different synthesis rates and varying ratios of the different species at different temperatures so as to achieve a temperature-independent, mean degradation rate, for each particular species. For instance, the most stable form of a protein could be produced in higher quantity at high temperature so that the mean degradation constant would be temperature-independent.

Global versus individual behaviors

An important assumption of the present and previous modeling studies is that the experimental curves (which at present mostly come from the average over many cells) represent faithfully the single cell oscillation. Recent experiments in mammalian fibroblasts stress the danger of this assumption: the dephasing of circadian oscillators between different cells results in the loss of global oscillations, even though individual cells oscillate (Nagoshi et al., 2004). A comparison of the *Neurospora* core feedback-loop with similar two-gene modules such as the p53/Mdm2 module also reveals that a globally coherent behavior can actually emerge from very different behaviors at the individual cell level (Bar-Or et al., 2000; Lahav et al., 2004). Behavior of individual cells is more pulsatile (Lahav et al., 2004), and *digital* in the sense that stress levels induce, at the individual scale, differences in the number of oscillations, but not in their amplitude or frequency. These different individual behaviors give rise to damped oscillations at the global scale, with various amplitudes and frequencies.

Such effects could also be at work in *Neurospora crassa* during its vegetative growth: the possible consequence of averaging over several nuclei could be to smooth out experimental curves, and render them more sinusoidal. Some nuclei could perhaps not oscillate at all. If, in some nuclei, a protein is produced at a given basal level—whereas its level oscillates in others—the amplitude of the average oscillation would be smaller than the real maximum amplitude. Intrinsic noise and nuclei-to-nuclei phase shift can also spread individual pulsatile behavior. Monitoring oscillations at the level of individual nuclei would thus be another important experimental step.

I thank V. Hakim for his help and advice, and for very useful discussions; J. C. Dunlap, J. Loros, and their group for useful discussions; B. Sagot and A. Moores for useful comments on the manuscript; and the referees for their comments and suggestions.

REFERENCES

- Aronson, B. D., K. A. Johnson, J. J. Loros, and J. C. Dunlap. 1994. Negative feedback defining a circadian clock: autoregulation of the clock gene frequency. *Science*. 263:1578–1584.

- Ballario, P., C. Talora, D. Galli, H. Linden, and G. Macino. 1998. Roles in dimerization and blue light photoresponse of the PAS and LOV domains of *Neurospora crassa* white-collar proteins. *Mol. Microbiol.* 29:719–729.
- Bar-Or, R. L., R. Maya, L. A. Segel, U. Alon, A. J. Levine, and M. Oren. 2000. Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proc. Natl. Acad. Sci. USA.* 97:11250–11255.
- Barkai, N., and S. Leibler. 1999. Biological rhythms: circadian clocks limited by noise. *Nature.* 403:267–268.
- Cheng, P., Y. Yang, C. Heintzen, and Y. Liu. 2001a. Coiled-coil domain mediated *frq-frq* interaction is essential for its circadian clock function in *Neurospora*. *EMBO J.* 20:101–108.
- Cheng, P., Y. Yang, and Y. Liu. 2001b. Interlocked feedback loops contribute to the robustness of the *Neurospora* circadian clock. *Proc. Natl. Acad. Sci. USA.* 98:7408–7413.
- Crosthwaite, S. K., J. J. Loros, and J. C. Dunlap. 1995. Light-induced resetting of a circadian clock is mediated by a rapid increase in frequency transcript. *Cell.* 81:1003–1012.
- Daan, S., and C. S. Pittendrigh. 1976. A functional analysis of circadian pacemakers in nocturnal rodents. *J. Comp. Physiol.* 106:291–331.
- Denault, D. L., J. J. Loros, and J. C. Dunlap. 2001. WC2 mediates WC1-*frq* interaction within the PAS protein-linked circadian feedback loop of *Neurospora*. *EMBO J.* 20:109–117.
- Dunlap, J. C., J. J. Loros, D. Denault, K. Lee, A. Froehlich, H. Colot, M. Shi, and A. Pogue. 2004. Genetics and molecular biology of circadian rhythms. In *The Mycota III, Biochemistry and Molecular Biology*. R. Brambl and G. Marzluf, editors. Springer-Verlag, Berlin and Heidelberg, Germany.
- Feldman, J. F., and M. N. Hoyle. 1973. Isolation of circadian clock mutants of *Neurospora crassa*. *Genetics.* 75:605–613.
- Froehlich, A. C., Y. Liu, J. J. Loros, and J. C. Dunlap. 2002. White-collar-1, a circadian blue light photoreceptor, binding to the *frequency* promoter. *Science.* 297:815–819.
- Froehlich, A. C., J. J. Loros, and J. C. Dunlap. 2003. Rhythmic binding of a white-collar-containing complex to the frequency promoter is inhibited by FREQUENCY. *Proc. Natl. Acad. Sci. USA.* 100:5914–5919.
- Garceau, N. Y., Y. Liu, J. J. Loros, and J. C. Dunlap. 1997. Alternative initiation of translation and time-specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY. *Cell.* 89:469–476.
- Gonze, D., J. C. Leloup, and A. Goldbeter. 2000. Theoretical models for circadian rhythms in *Neurospora* and *Drosophila*. *C. R. Acad. Sci. III.* 323:57–67.
- Goodwin, B. C. 1965. Oscillatory behavior in enzymatic control process. In *Advances in Enzyme Regulation*. G. Weber, editor. Pergamon Press, Oxford, UK. 425–438.
- Heintzen, C., J. J. Loros, and J. C. Dunlap. 2001. The PAS protein VIVID defines a clock-associated feedback loop that represses light input, modulates gating, and regulates clock resetting. *Cell.* 104:453–464.
- Lahav, G., N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. Elowitz, and U. Alon. 2004. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.* 36:147–150.
- Lee, K., J. J. Loros, and J. C. Dunlap. 2000. Interconnected feedback loops in the *Neurospora* circadian systems. *Science.* 289:107–110.
- Leloup, J. C., and A. Goldbeter. 1998. A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J. Biol. Rhythms.* 13:70–87.
- Leloup, J. C., and A. Goldbeter. 2003. Toward a detailed computational model for the mammalian circadian clock. *Proc. Natl. Acad. Sci. USA.* 100:7051–7056.
- Liu, Y., N. Y. Garceau, J. J. Loros, and J. C. Dunlap. 1997. Thermally regulated translational control of FRQ mediates aspects of temperature responses in the *Neurospora* circadian clock. *Cell.* 89:477–486.
- Liu, Y., J. J. Loros, and J. C. Dunlap. 2000. Phosphorylation of the *Neurospora* clock protein FREQUENCY determines its degradation rate and strongly influences the period length of the circadian clock. *Proc. Natl. Acad. Sci. USA.* 97:234–239.
- Liu, Y., M. W. Mellow, J. J. Loros, and J. C. Dunlap. 1998. How temperature changes reset a circadian oscillator. *Science.* 281:825–829.
- Loros, J. J., and J. C. Dunlap. 2001. Genetic and molecular analysis of circadian rhythms in *Neurospora*. *Annu. Rev. Physiol.* 63:757–794.
- Mellow, M. W., N. Y. Garceau, and J. C. Dunlap. 1997. Dissection of a circadian oscillation into discrete domains. *Proc. Natl. Acad. Sci. USA.* 94:3877–3882.
- Nagoshi, E., C. Saini, C. Bauer, T. Laroche, F. Naef, and U. Schibler. 2004. Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell.* 119:693–705.
- Pittendrigh, C. S., V. G. Bruce, N. S. Rosenzweig, and M. L. Rubin. 1959. A biological clock in *Neurospora*. *Nature.* 184:169–170.
- Reppert, S. M., and D. R. Weaver. 2001. Molecular analysis of mammalian circadian rhythms. *Annu. Rev. Physiol.* 63:647–676.
- Reppert, S. M., and D. R. Weaver. 2002. Coordination of circadian timing in mammals. *Nature.* 418:935–941.
- Ruoff, P., and L. Rensing. 1996. The temperature-compensated Goodwin model simulates many circadian clock properties. *J. Theor. Biol.* 179:275–285.
- Ruoff, P., M. Vinsjevnik, C. Monnerjahn, and L. Rensing. 1999. The Goodwin oscillator: on the importance of degradation reactions in the circadian clock. *J. Biol. Rhythms.* 14:469–479.
- Smolen, P., D. A. Baxter, and J. H. Byrne. 2001. Modeling circadian oscillations with interlocking positive and negative feedback loops. *J. Neurosci.* 21:6644–6656.
- Smolen, P., D. A. Baxter, and J. H. Byrne. 2002. A reduced model clarifies the role of feedback loops and time delays in the *Drosophila* circadian oscillator. *Biophys. J.* 83:2349–2359.
- Tyson, J. J., C. I. Hong, C. D. Thron, and B. Novak. 1999. A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM. *Biophys. J.* 77:2411–2417.
- Yang, Y., P. Cheng, and Y. Liu. 2002. Regulation of the *Neurospora* circadian clock by casein kinase II. *Genes Dev.* 16:994–1006.
- Young, M. W. 2002. Big Ben rings in a lesson on biological clocks. *Neuron.* 36:1001–1005.

Perspectives

En guise de perspective, évoquons maintenant une direction possible de recherche à partir des travaux présentés dans cette thèse. Nous avons déjà évoqué la possibilité de générer des motifs spatiaux à l'aide de l'algorithme. Plus généralement, une voie fructueuse pourrait être d'utiliser l'algorithme afin de trouver et de modéliser des réseaux génétiques associés à des processus de différenciation cellulaire (par exemple la formation de structures voire d'organes spécialisés). Ainsi, plusieurs techniques d'ingénierie inverse et d'analyse logique ont déjà permis d'analyser de tels systèmes. Par exemple, le mécanisme de segmentation de la drosophile a été analysé à la fois par des méthodes basées sur des fits à l'aide d'équations différentielles ordinaires [43,85], mais également par des méthodes basées sur des équations logiques, à variables booléennes [73] (pour une revue, on pourra consulter la référence [79]). Ces diverses méthodes ont permis en particulier de prédire certaines interactions ou de souligner l'importance de certaines boucles de rétroaction. L'algorithme d'évolution présenté dans cette thèse pourrait lui aussi suggérer des interactions en sélectionnant des réseaux génétiques à partir de la connaissance du comportement biologique.

L'un des autres exemples biologiques en cours d'études actuellement est le réseau génétique associé à la somitogénèse. Ce processus consiste en la formation de segments répétitifs le long de l'axe antéro-postérieur de l'embryon (appelés somites), qui donneront ensuite les vertèbres. Cooke et Zeeman [10] ont proposé que cette segmentation périodique spatialement est en fait la trace d'une oscillation temporelle couplée à l'avancée d'un front au moment de l'embryogénèse (processus appelé "clock and wavefront"). Effectivement, il a été ensuite confirmé expérimentalement que certaines concentrations de protéines oscillaient dans les embryons des vertébrés précisément à l'endroit où se forment les somites [65,67]. Il semble maintenant qu'un petit module génétique soit à l'origine de ces oscillations (pour une liste des différents "candidats", on pourra consulter la revue de Giudicelli et Lewis [33]). L'un des faits expérimentaux cruciaux dans la somitogénèse est que l'oscillateur génétique impliqué est en fait contrôlé par un gradient de morphogène (parmi les protéines impliquées, citons FGF8 [16] et Wnt3a [2]). Ces protéines sont très concentrées au

niveau du mésoderme présomitique, et leur concentration décroît le long de l'axe de formation des somites. Lorsque la concentration passe sous un seuil donné, l'oscillation à l'origine des somites s'arrête et selon la phase de l'oscillation, un somite se forme ou non.

L'un des problèmes théoriques posé par cette somitogénèse est de savoir comment, par l'intermédiaire d'un réseau génétique, on peut passer d'un état oscillant à un état différencié. Comme il existe deux états finaux possibles (dans ce cas avant ou arrière du somite), l'état différencié doit en fait se caractériser par deux états stables d'un point de vue dynamique. On est donc *a priori* en présence d'une ou plusieurs bifurcations contrôlées par la concentration de morphogène. Le mécanisme le plus simple permettant d'expliquer une telle somitogénèse serait alors le passage d'une phase oscillante à une phase bistable. La concentration de morphogène joue alors le rôle d'un paramètre extérieur au système, contrôlant la dynamique de celui-ci. À haute concentration, la présence du morphogène induirait une déstabilisation du point fixe et donc l'apparition d'un cycle limite. À basse concentration, on observerait au contraire l'apparition de deux points fixes stables et le système gagnerait l'un ou l'autre des points fixes. Lorsqu'on passe de haute concentration de morphogène à basse concentration de morphogène, le système passe d'un état oscillant à un état bistable. Selon la phase de l'oscillation au moment du changement de concentration de morphogène, l'état final serait l'un ou l'autre des deux états stables.

Or nous avons étudié dans cette thèse un réseau génétique typique pouvant se comporter comme un oscillateur ou comme un système bistable : il s'agit de la Mixed Feedack Loop. Pour passer d'un système oscillant à un système bistable, il "suffit" de convertir une activation transcriptionnelle en répression transcriptionnelle. Ce rôle serait alors joué par la protéine du morphogène. Nous avons donc imaginé le modèle suivant, décrit Figure IV.8 : nous appelons C la protéine du gradient de morphogène. Lorsque A et C sont absents, B est produit avec un taux de base élevé. En présence de A ou de C, B est réprimé. En revanche, lorsque A et C sont présents ensemble, ils agissent coopérativement pour activer la production de la protéine B. En présence de C, nous avons donc bien une MFL oscillante, tandis qu'en son absence nous avons une MFL bistable ³.

L'une des protéines oscillantes impliquées dans la somitogénèse, Axin2 [2] est

³Signalons à ce stade qu'il est également tout à fait possible d'obtenir un tel comportement - comme on peut le voir Figure IV.9- à l'aide de l'autre oscillateur génétique pouvant également se comporter comme un système bistable obtenu par l'algorithme d'évolution Figures II.15 et II.7. En fait, il n'est pas très difficile de voir analytiquement que le comportement de ce réseau est très similaire à celui du MFL.

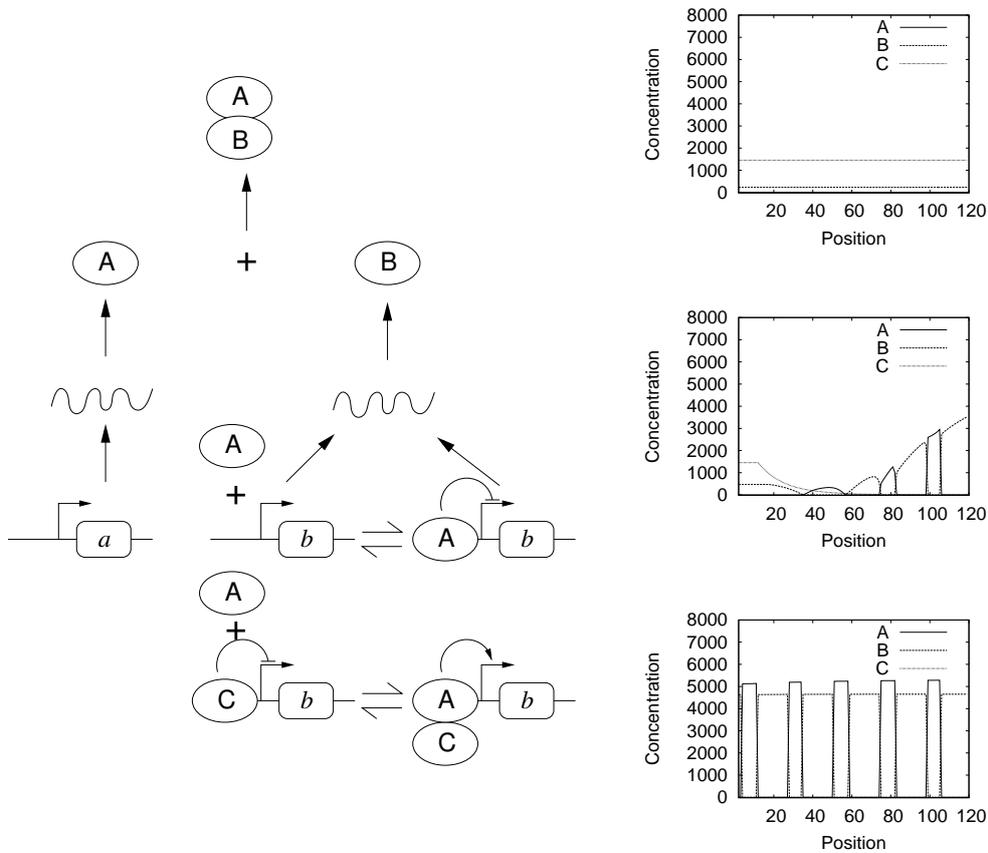


FIG. IV.8: Un réseau génétique pouvant donner lieu à la formation de motifs spatiaux, inspiré du réseau de la MFL. Dans le panneau de droite est présentée la concentration des protéines en fonction de la position pour différents temps. Les protéines A et B oscillent de manière synchronisée dans l'espace tant que la concentration de protéines C est haute (haut). Puis, lorsque la concentration de C décroît, on passe de l'état bistable à l'état oscillant et les motifs apparaissent (milieu et bas).

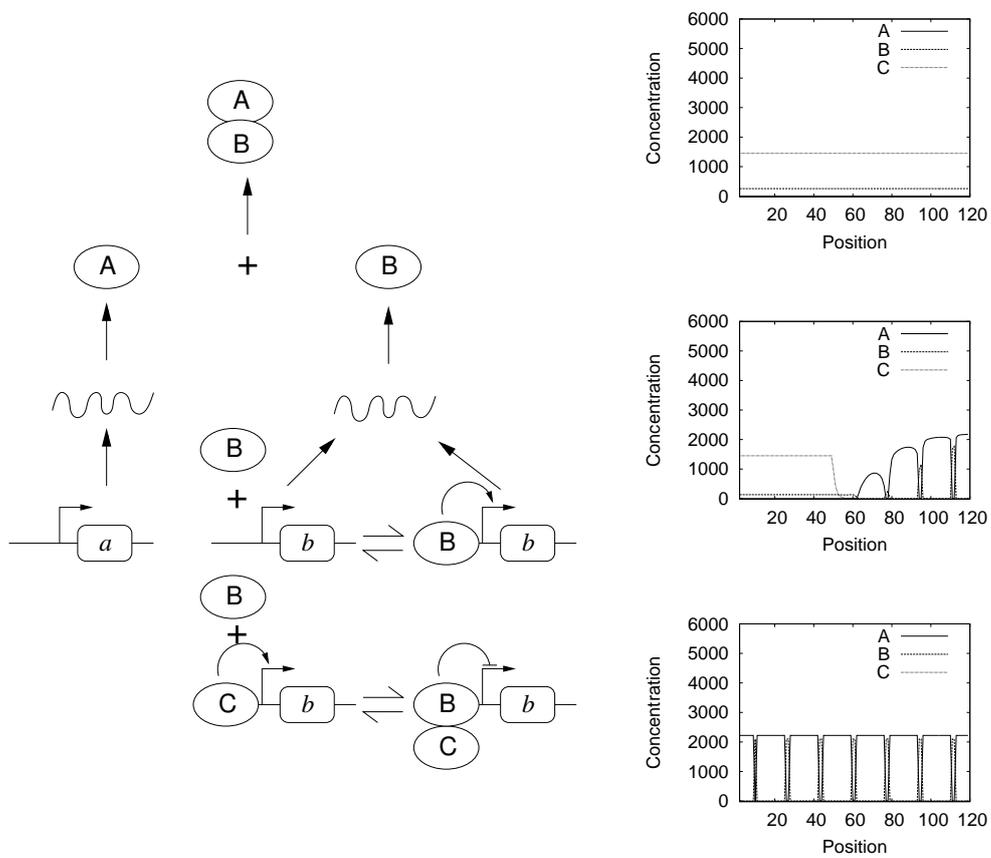


FIG. IV.9: Un autre réseau génétique pouvant donner lieu à la formation de motifs spatiaux, inspiré des réseaux présentés Figures II.15 et II.7. Dans le panneau de droite est présentée la concentration des protéines en fonction de la position pour différents temps

présente dans un module génétique tout à fait similaire à une MFL : ainsi, Axin2 interagit avec β -catenin et la phosphoryle, tandis que β -catenin active la transcription d'Axin2. Un prolongement naturel des travaux présentés dans cette thèse serait donc de suggérer des mécanismes afin de mieux comprendre de tels processus de morphogénèse.

Conclusions

Cette thèse nous a permis de considérer différentes questions théoriques liées à l'organisation et au fonctionnement des réseaux génétiques. L'organisation des réseaux génétiques en modules fonctionnels nous a conduits à développer une procédure d'évolution *in silico* permettant de concevoir des modules génétiques remplissant une fonction donnée. Cette approche s'est révélée fructueuse et a permis de concevoir de nombreux exemples de commutateurs et d'oscillateurs génétiques. Parmi les différents modules obtenus, la "Mixed Feedback Loop" nous a semblé être la plus intéressante d'un point de vue biologique et a été étudiée en détail analytiquement. Nous avons également suggéré différents modèles d'horloges circadiennes. Ces différents modèles et études ont permis de mettre en évidence l'importance dynamique de plusieurs processus biochimiques :

- les interactions post-transcriptionnelles, et en particulier les dimérisations sont fondamentales dans tous les modules étudiés. Ces dimérisations sont très importantes d'un point de vue dynamique car dans les cas que nous avons considérés, elles répriment l'activité des deux protéines, simultanément ou successivement. Lorsqu'elles sont couplées à une interaction transcriptionnelle, ces différentes rétroactions permettent d'avoir des comportements riches et complexes, à l'image de ce qui se passe pour la "Mixed Feedback Loop",
- la prise en compte de la dynamique de l'ARN est fondamentale pour comprendre les oscillations. La production des protéines n'est pas directe et passe par l'étape intermédiaire de la transcription. Ceci ajoute un degré de liberté au système, fondamental pour expliquer certaines oscillations. Ainsi, dans les modèles de MFL et d'horloges circadiennes, la dynamique de l'oscillation est fondamentalement dirigée par la dynamique des transcrits. Il est donc nécessaire de tenir compte et de modéliser explicitement cette étape.

Certaines questions restent en suspens. La pertinence et la réalité d'une vision modulaire du génome⁴ fait débat. Les implémentations de l'algorithme présentées

⁴dans le sens où des motifs spécifiques ayant une fonction donnée seraient sur-représentés

dans le chapitre II montrent qu'il n'y a aucune unicité des réseaux pouvant être associés à une fonction donnée. A priori, une structure modulaire des génomes suggérerait, selon l'approche d'Uri Alon, qu'il y a des mécanismes d'évolution qui sélectionnent spécifiquement des modules donnés et qu'à une fonction donnée, un seul motif est associé dans les réseaux génétiques. Or, de nouveaux travaux ont remis en question cette correspondance univoque entre module et fonction [58] : en particulier, il n'y a pas d'identification réciproque dans les réseaux génétiques entre un module détecté par l'analyse statistique et une fonction donnée. Au contraire, les modules repérés semblent trop imbriqués pour être fonctionnels par eux-mêmes et semblent plutôt être au coeur de clusters de modules, ou "supermodules". Cependant, des modules apparemment très différents, ou très complexes, peuvent avoir en fait des mécanismes de fonctionnement et des dynamiques très similaires. Par exemple, le fonctionnement des commutateurs II.6 et II.7 est très semblable, même si les topologies des réseaux génétiques sont différentes. Cela signifie qu'une fonction donnée (par exemple "commutateur génétique") pourrait se traduire par une dynamique préférentielle (par exemple une régulation transcriptionnelle pour réguler les taux de production, associée à une dimérisation pour commuter d'un état à l'autre), qui peut elle-même éventuellement se traduire par plusieurs réseaux (les commutateurs II.6 et II.7). Néanmoins, il est frappant de constater qu'en ce qui concerne les oscillateurs génétiques, la "Mixed feedback Loop" semble émerger comme module au sens donné par Uri Alon. Dans ce cas, les contraintes purement biologiques sont peut-être si fortes qu'un seul type de réseau est possible et émerge naturellement pour les réaliser.

Plus généralement, l'un des défis futurs posés aux modélisateurs sera de comprendre à la fois qualitativement et quantitativement l'évolution des différentes fonctions biologiques d'une espèce à l'autre du point de vue des séquences génétiques. Par exemple, nous avons évoqué une vision fonctionnelle des modules génétiques, et il est donc naturel de se demander comment les différentes pressions évolutives amènent l'émergence de ces sous-réseaux fonctionnels. L'évolution procède-t-elle en ajustant progressivement les forces des interactions entre les différents gènes dans des réseaux déjà existants, ou procède-t-elle en créant ou en réinventant soudainement des structures génétiques permettant de résoudre un problème donné ? De la même façon, on peut s'interroger sur le caractère optimal des réseaux obtenus. Un modélisateur pourrait penser que l'évolution va optimiser la structure du réseau pour répondre parfaitement à une fonction donnée. Or, l'évolution pourrait tout aussi bien privilégier des structures beaucoup moins efficaces mais suffisamment simples

pour apparaître rapidement, plutôt que de rechercher des structures fonctionnant mieux, mais nécessitant beaucoup d'étapes intermédiaires au cours de l'évolution. La définition d'optimum en terme biologique n'est donc pas nécessairement la même que celle des ingénieurs. Il apparaît également nécessaire de comprendre à la fois les contraintes imposées et les libertés laissées par les différents mécanismes biochimiques préexistants sur l'évolution des séquences. Ainsi, a-t-on réellement saisi à la fois tout l'intérêt et toutes les contraintes liées à l'organisation de l'information sous la forme du triplet gène-ARN-protéine ? Nous avons par exemple suggéré que l'étape de la transcription permettait d'avoir plus facilement des oscillateurs génétiques : en guise de conclusion, renversons cette perspective et demandons-nous pourquoi l'évolution a sélectionné les mécanismes du vivant tels que nous les connaissons. Peut-être la plus importante fonction optimisée par l'évolution est-elle tout simplement la capacité même à évoluer, comme le suggèrent certains travaux expérimentaux menés par Radman et collaborateurs (on pourra par exemple consulter la revue citée en référence [56], ainsi qu'une approche théorique associée [78]. . .).

Appendices

Appendice A

Compléments sur l’algorithme d’évolution

A.I Exemples de commutateurs.

L’algorithme a permis de créer plusieurs commutateurs de type “toggle switch” (Figure II.1) basés sur une inhibition réciproque. Un exemple est présenté dans le texte principal Figure II.9 et un autre exemple est présenté ici Figure A.1.

Les commutateurs des Figures II.6 et II.7 ont été obtenus pour une large variété de concentrations d’équilibre pour les points fixes. Les Figures A.2 et A.3 donnent des constantes cinétiques obtenues pour des concentrations de plusieurs centaines de protéines par cellule.

D’autres commutateurs basés sur les mêmes principes que les commutateurs des Figures II.6 et II.7 mais avec des interactions un peu plus complexes ont été obtenus. Une version plus compliquée du commutateur de la Figure II.7 est présentée Figure A.4.

Dans la Figure A.5 est présenté un commutateur dans lequel le gène b est en fait un gène reporteur d’un commutateur à deux gènes entre a et c ¹. Ce commutateur à deux gènes est très semblable au commutateur de la figure II.7.

Enfin, on trouvera dans la table A.1 les constantes cinétiques du commutateur II.8 qui repose véritablement sur l’interaction de 3 gènes.

A.II Simulations avec des dynamiques stochastiques

Afin de tester la résistance au bruit des commutateurs fabriqués, nous avons simulé la dynamique des commutateurs présentés Figures II.6 et II.7 (dont les réseaux

¹Pour des raisons de clarté du schéma, les gènes a et b sur lesquels la sélection opérerait ont été inversés et renommés

d'interaction sont identiques aux commutateurs présentés Figures A.2 et A.3) avec un algorithme stochastique [32]. Pour des jeux de paramètres correspondant à de hautes concentrations de protéines de plusieurs centaines de protéines par cellule (constantes cinétiques de la Figure A.2), des fluctuations de concentration sont visibles, mais la fonction “switch” du réseau ne semble pas affectée et la bistabilité est maintenue, comme le montre le Figure A.6. Pour des régimes de paramètres correspondant à des concentrations d'équilibre de l'ordre de la dizaine de protéines par cellule, les fluctuations sont plus importantes, mais le réseau génétique se comporte toujours comme un commutateur (Figure A.7). Dans ces conditions de forte fluctuation, le commutateur de la Figure II.6 “saute” spontanément entre les deux états avec une constante de temps de l'ordre de plusieurs heures. Un tel saut est montré Figure A.7A après environ 400 minutes d'évolution libre. Le commutateur de la Figure II.7 semble sauter spontanément moins souvent (Figure A.7B). Cependant, cette différence ne semble pas intrinsèque à la topologie des modules car pour d'autres jeux de paramètres, les résistances aux fluctuations sont comparables.

A.III Exemples d'oscillateurs.

Nous précisons les constantes cinétiques des oscillateurs donnés dans le texte principal dans les Figures A.8, A.9, A.10, A.11 et A.12.

Reactions	Constants
$a \rightarrow a+A$	0.0048
$A \rightarrow X$	0.0093
$b \rightarrow b+B$	1.3
$B \rightarrow X$	0.058
$c \rightarrow c+C$	0.75
$C \rightarrow X$	0.091
$B+C \rightarrow B$	0.81
$a+C \rightarrow a :C$	0.22
$a :C \rightarrow a+C$	0.95
$a :C \rightarrow a :C+A$	0.57
$b+B \rightarrow b :B$	0.39
$b :B \rightarrow b+B$	0.40
$b :B \rightarrow b :B+B$	1.6
$b+A \rightarrow b :A$	1.2
$b :A \rightarrow b+A$	0.6
$b :A \rightarrow b :A+B$	0.45
$b :A+A \rightarrow b :A :A$	0.25
$b :A :A \rightarrow b :A+A$	1.1
$b :A :A \rightarrow b :A :A+B$	0.060
$b :A :A+A \rightarrow b :A :A :A$	0.32
$b :A :A :A \rightarrow b :A :A+A$	0.10
$b :A :A :A \rightarrow b :A :A :A+B$	0.032

TAB. A.1: Constantes pour le commutateur à trois gènes présenté Figure II.8

Réactions	Constantes
$a \rightarrow a+A$	0.39
$A \rightarrow X$	0.00039
$b \rightarrow b+B$	0.26
$B \rightarrow X$	0.0033
$A+B \rightarrow A :B$	0.015
$A :B \rightarrow X$	0.34
$b+B \rightarrow b :B$	0.30
$b :B \rightarrow b+B$	1.0
$b :B \rightarrow b :B+B$	0.71

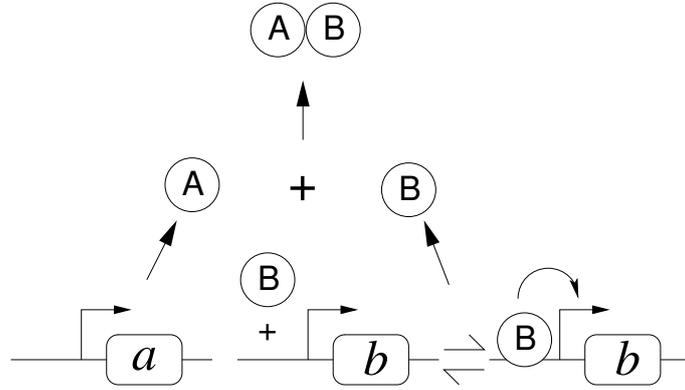


FIG. A.3: Commutateur avec la même topologie que dans la Figure II.7 mais avec des concentrations de plusieurs centaines de protéines par cellule.

Réactions	Constantes
$a \rightarrow a+A$	0.79
$A \rightarrow X$	0.046
$b \rightarrow b+B$	0.070
$B \rightarrow X$	0.0039
$A+B \rightarrow A :B$	0.28
$A :B \rightarrow X$	0.027
$a+A \rightarrow a :A$	0.78
$a :A \rightarrow a+A$	0.020
$a :A \rightarrow a :A+A$	0.75
$b+B \rightarrow b :B$	0.26
$b :B \rightarrow b+B$	0.95
$b :B \rightarrow b :B+B$	0.56
$a :A+B \rightarrow a :A :B$	0.087
$a :A :B \rightarrow a :A+B$	0.11
$a :A :B \rightarrow a :A :B+A$	0.30

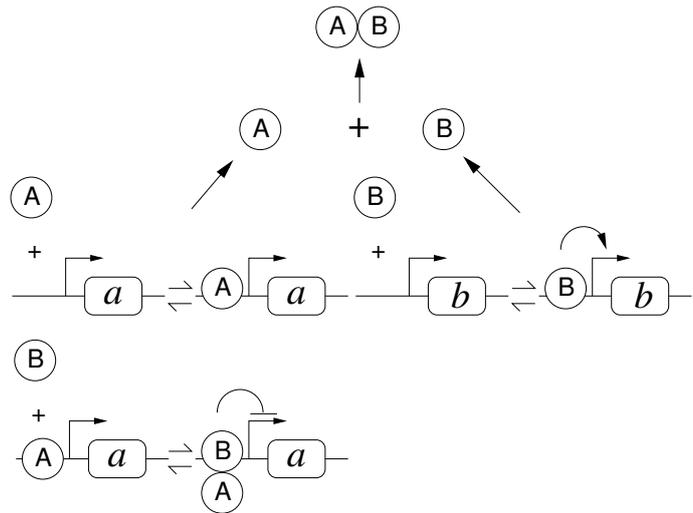


FIG. A.4: Commutateur reposant sur le même principe que celui de la Figure II.7 mais avec des interactions un peu plus compliquées.

Réactions	Constantes
$a \rightarrow a+A$	0.17
$A \rightarrow X$	0.038
$b \rightarrow b+B$	0.88
$B \rightarrow X$	0.060
$c \rightarrow c+C$	0.42
$C \rightarrow X$	0.041
$a+A \rightarrow a:A$	0.62
$a:A \rightarrow a+A$	0.70
$a:A \rightarrow a:A+A$	0.58
$b+C \rightarrow b:C$	0.98
$b:C \rightarrow b+C$	0.15
$b:C \rightarrow b:C+B$	0.95
$a:A+A \rightarrow a:A:A$	0.70
$a:A:A \rightarrow a:A+A$	0.83
$a:A:A \rightarrow a:A:A+A$	0.93
$b:C+A \rightarrow b:C:A$	0.94
$b:C:A \rightarrow b:C+A$	0.59
$b:C:A \rightarrow b:C:A+B$	0.15
$A+C \rightarrow A:C$	0.61

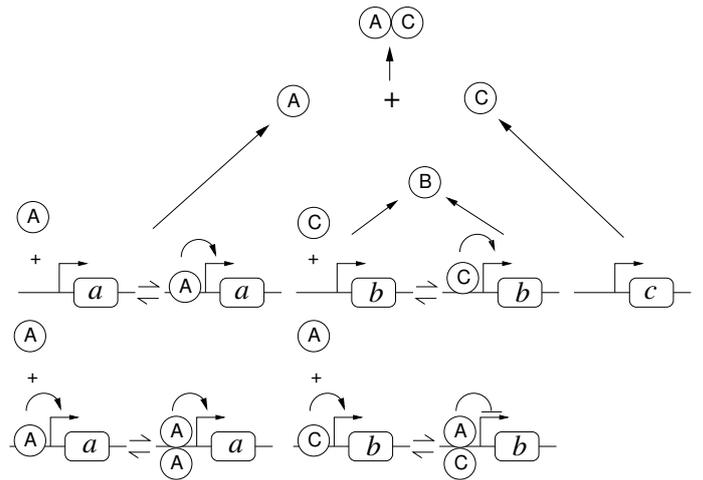


FIG. A.5: Commutateur avec trois gènes mais dans lequel b est un simple gène reporteur (la protéine B ne régule aucun des gènes a ou c et n'interagit avec aucune protéine).

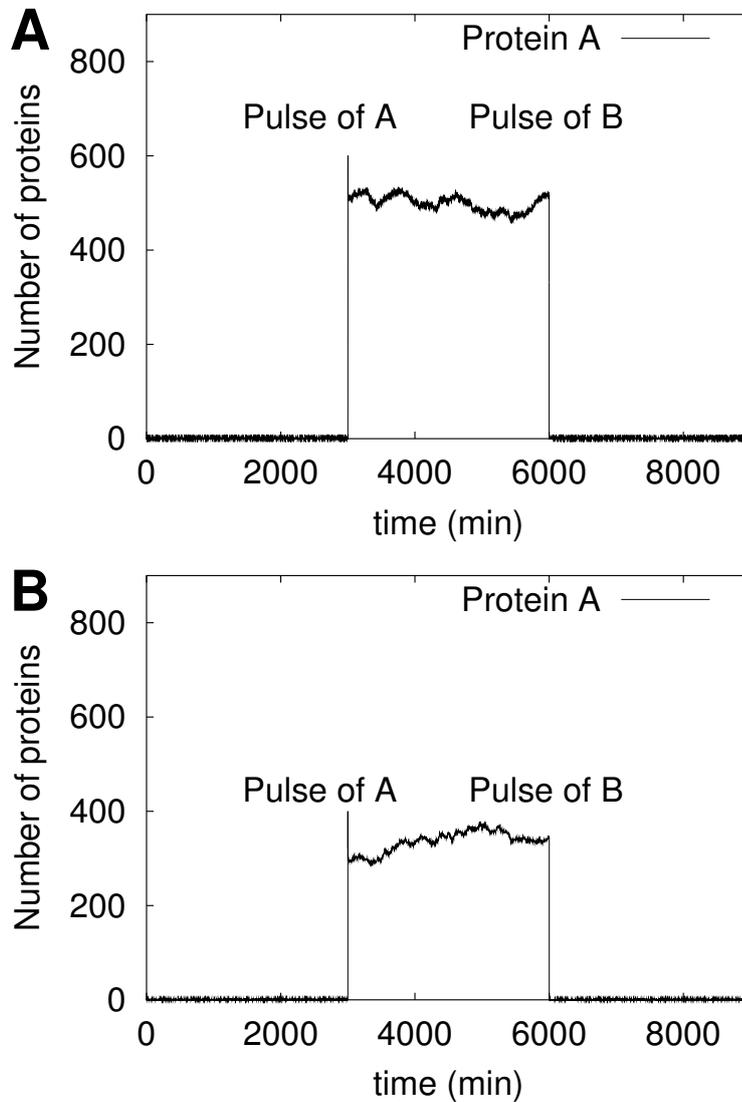


FIG. A.6: **A** : Évolution stochastique du commutateur présenté Figure A.2, version à haute concentration du commutateur de la Figure II.6. Pour ces valeurs de paramètres, les fluctuations stochastiques ne modifient pas significativement le caractère bistable du commutateur. L'état stable avec une basse concentration de A évolue librement et reste stable pendant 3000 minutes. Ensuite, on introduit un pulse de protéines A qui provoque le passage dans le second état stable avec une haute concentration de A. Ce nouvel état persiste pendant 3000 minutes, après lesquelles un pulse de B fait retourner le système à son état initial. **B** : Évolution similaire du commutateur présenté Figure A.3, version à haute concentration du commutateur de la Figure II.7 .

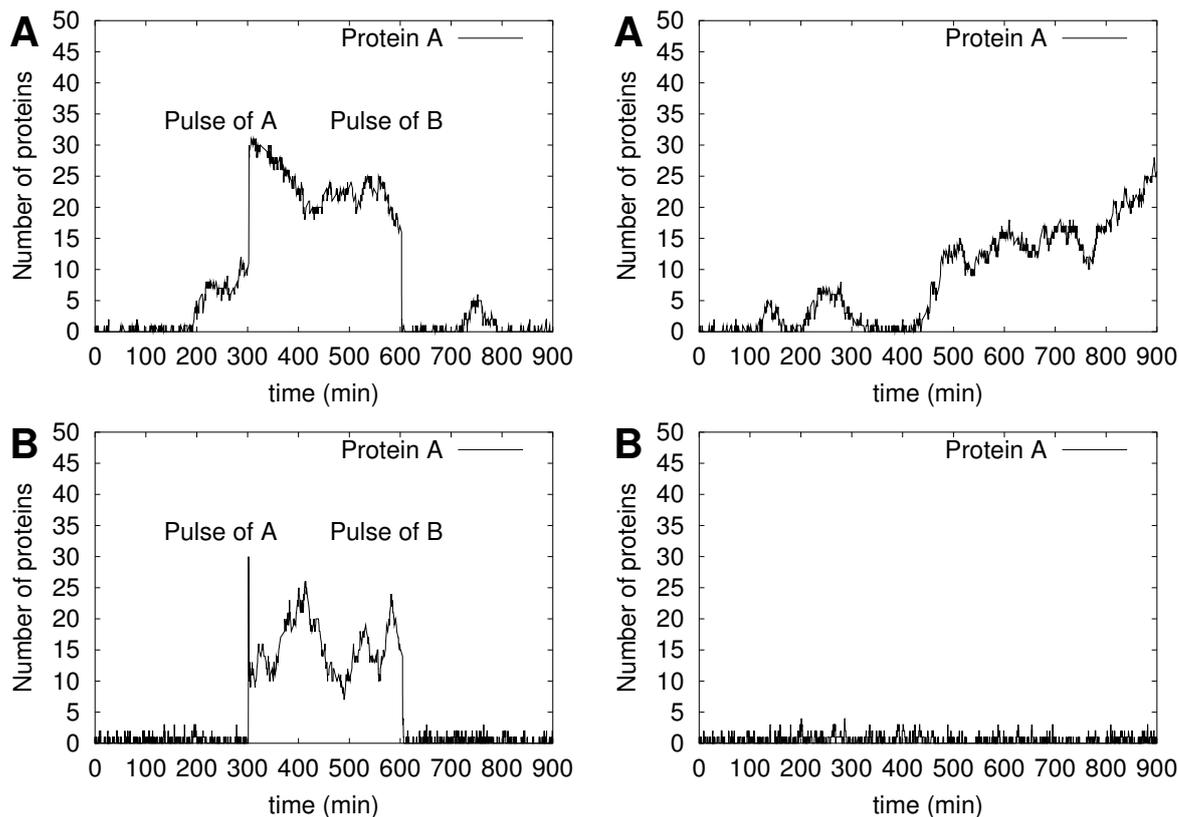


FIG. A.7: **A** : Évolution stochastique pour le commutateur de la Figure II.6 avec les constantes cinétiques données dans le texte principal. Pour ces paramètres, les concentrations maximales sont de l'ordre de la dizaine de protéines et les fluctuations sont importantes. Cependant, le caractère bistable du module est conservé. Gauche : Le passage à l'état stable de haute concentration pour A est induit au temps $t=300$ minutes. Un pulse de B remet le système dans son état initial au temps $t=600$ minutes. . Noter qu'au moment du premier pulse le réseau "saute" déjà spontanément à son deuxième état. Droite : Évolution libre du module pendant 900 minutes. Autour de $t=450$ minutes, le système "saute" de l'état A bas à l'état A haut. **B** : Évolution similaire pour le commutateur de la Figure II.7. Gauche : Le passage d'un état à l'autre est induit par des pulses de protéines. Droite : Évolution libre pendant 900 minutes. Pour ce jeu de paramètres, le commutateur II.7 est plus stable que le commutateur II.6.

Réactions		Constantes
a	$\rightarrow a+A$	0
A	$\rightarrow X$	0.00019
b	$\rightarrow b+B$	0.43
B	$\rightarrow X$	0.077
c	$\rightarrow c+C$	0.57
C	$\rightarrow X$	0.023
$A+B$	$\rightarrow A :B$	0.066
$A :B$	$\rightarrow X$	0.057
$C+C$	$\rightarrow C :C$	0.14
$C :C$	$\rightarrow X$	0.0014
$C :C + B$	$\rightarrow C :C$	0.21
$C :C + C :C$	$\rightarrow C :C :C :C$	0.37
$C :C :C :C$	$\rightarrow X$	0.24
$C+A :B$	$\rightarrow A :B$	1.5
$a + C :C :C :C$	$\rightarrow a :C :C :C :C$	0.5
$a :C :C :C :C$	$\rightarrow a + C :C :C :C$	0.33
$a :C :C :C :C$	$\rightarrow a :C :C :C :C + A$	0.56
$A :B + C :C$	$\rightarrow A :B :C :C$	0.74

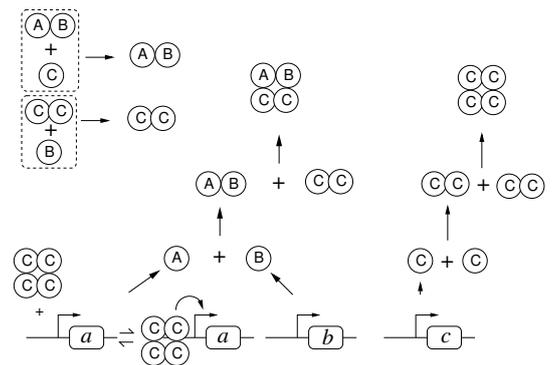


FIG. A.8: Constantes pour l'oscillateur de la Figure II.12.

Réactions	Constantes
$a \rightarrow a+A$	1.2
$A \rightarrow X$	0.000075
$b \rightarrow b+B$	2.0
$B \rightarrow X$	0.23
$c \rightarrow c+C$	1.2
$C \rightarrow X$	0.0045
$A+B \rightarrow A :B$	2.7
$A :B \rightarrow X$	0
$a+B \rightarrow a :B$	0.50
$a :B \rightarrow a+B$	0.11
$a :B \rightarrow a :B+A$	21
$c+B \rightarrow c :B$	2.9
$c :B \rightarrow c+B$	0.56
$c :B \rightarrow c :B+C$	11
$c+A \rightarrow c :A$	0.62
$c :A \rightarrow c+A$	0.90
$c :A \rightarrow c :A+C$	0.85
$A :B+C \rightarrow C$	0.47
$A :B+C \rightarrow B$	0.99
$A :B \rightarrow B$	0.20

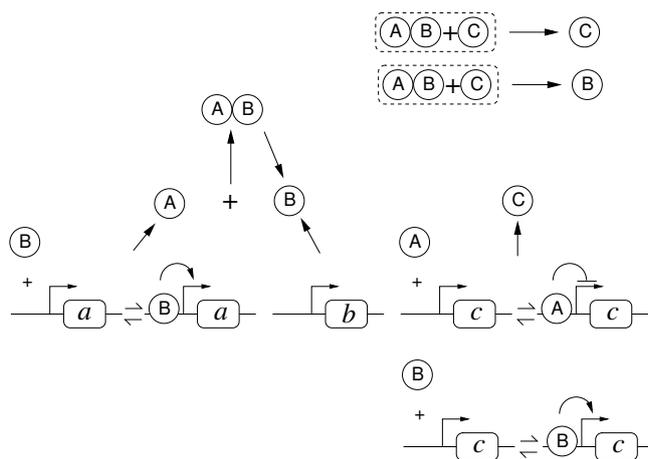


FIG. A.9: Constantes pour l'oscillateur de la Figure II.13.

Réactions	Constantes
$a \rightarrow a+A$	1.3
$A \rightarrow X$	0
$b \rightarrow b+B$	0.17
$B \rightarrow X$	0.00017
$A+A \rightarrow A :A$	0.012
$A :A \rightarrow X$	0.000047
$A :A+B \rightarrow A :A :B$	8.3
$A :A+A :A \rightarrow A :A :A :A$	0.23
$A :A :A :A \rightarrow X$	0.00018
$A :A :A :A+A \rightarrow A :A :A :A$	15
$A :A :A :A+B \rightarrow A$	0.39

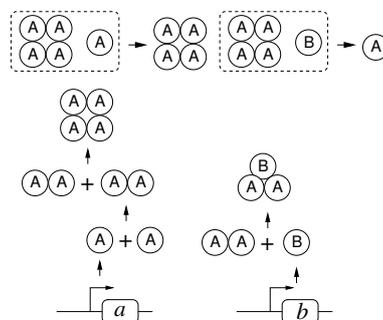


FIG. A.10: Constantes pour l'oscillateur de la Figure II.14.

Réactions	Constantes
$r_a \rightarrow X$	0.1
$A \rightarrow X$	0.3
$r_b \rightarrow X$	0.19
$B \rightarrow X$	0
$A :B \rightarrow X$	0.000031
$a \rightarrow a+r_a$	0.84
$r_a \rightarrow r_a + A$	2.35
$b \rightarrow b+r_b$	1.28
$r_b \rightarrow r_b + B$	0.84
$A + B \rightarrow A :B$	28.29
$A :B + B \rightarrow B$	2.7
$a + A \rightarrow a :A$	0.18
$a :A \rightarrow a + A$	0.29
$a :A \rightarrow r_a + a :A$	0.05
$A + A :B \rightarrow A$	21.78
$A :B \rightarrow B$	12.453150

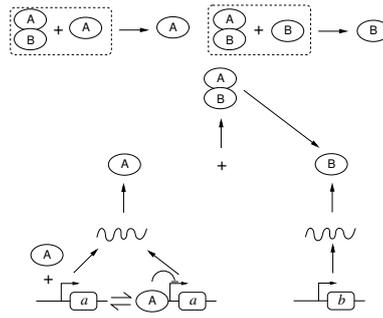


FIG. A.11: Constantes pour l'oscillateur de la Figure II.15.

Réactions	Constantes
$r_a \rightarrow X$	0.18
$A \rightarrow X$	0.16
$r_b \rightarrow X$	0.065
$B \rightarrow X$	0.000052
$A :B \rightarrow X$	0.0013
$a \rightarrow a+r_a$	0.0071
$r_a \rightarrow r_a + A$	1.12
$b \rightarrow b+r_b$	2.78
$r_b \rightarrow r_b + B$	0.098
$A + B \rightarrow A :B$	3.77
$a + B \rightarrow a :B$	0.58
$a :B \rightarrow a + B$	0.54
$a :B \rightarrow a :B + r_a$	2.16
$A :B + A :B \rightarrow B$	0.18
$A :B + A \rightarrow A$	0.30

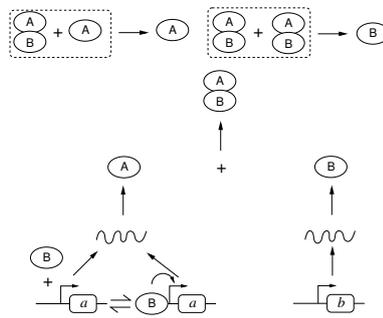


FIG. A.12: Constantes pour l'oscillateur de la Figure II.16.

Appendice B

Méthodes Informatiques

B.I Classes utilisées dans l'algorithme

Je présente ici les classes utilisées dans la plupart des simulations numériques pour modéliser les processus physico-chimiques, en particulier dans l'algorithme d'évolution *in silico*. Tous les codes sont rédigés en C++ : le choix de ce langage orienté objet permettait de définir facilement des classes et des outils associés correspondant à différents niveaux d'organisation imbriqués.

```
class Espece{
public:
    DP Constante;
    DP Qte;
};
```

Cette classe implémente toutes les espèces interagissantes. Une espèce chimique est définie par deux réels positifs. Le champ Constante contient la constante de dégradation de l'espèce. Cette constante peut être nulle (par exemple pour un gène). Toutes les dégradations sont supposées être exponentielles. Le champ Qte correspond à la quantité de réactifs présents dans la cellule. Ces espèces seront classées et numérotées dans la classe Expérience.

```
class Reaction
{
public:
    DP Constante;
    int Reactif1;
    int Reactif2;
    int Produit1;
    int Produit2;
};
```

Cette classe définit les réactions chimiques possibles dans une Expérience. Une réaction chimique est définie par une constante, deux réactifs (au plus) et deux produits (au plus). Ces réactifs et produits sont définis par des entiers qui correspondent aux numéros des espèces définies dans le tableau Réactifs de la classe Expérience. Il y a au moins un réactif et un produit. S'il y a un seul réactif (resp. un seul produit), l'entier Réactif2 (resp. Produit2) est supposé égal à -1.

```
class Experience{
public:
    DP Score;
    int NombreReactions;
```

```

    int NombreReactifs;
    int NombreGenes;
    int NombrePromus;
    Espece Reactifs[NReactifsMax+NGenesMax+NPromusMax];
    Reaction Reactions[NReactionsMax];
    //Méthodes de la classe Espèce
    void init();
    void AjoutEspece(DP,DP);
    void AjoutReaction(int,int,int,int,DP);
    void AjoutGene(DP,DP);
    void AjoutPromotion(int,int,DP,DP,DP);
    void AjoutTrans(int,DP, DP);
    void Degradation(int );
    void Etape(int);
    void proba(Vec_O_DP &);
    void ModifEspece(int,int &);
    void ModifReaction(int,int &);
    void RetireReaction(int );
    void ModifProm(int,int &);
    void Evolution(int &);
    void derivs(const DP , Vec_I_DP &, Vec_O_DP & );
    void rk4(Vec_I_DP &, Vec_I_DP &, const DP , const DP ,Vec_O_DP &);
    void rk dumb(Vec_IO_DP &,Mat_IO_DP &, const DP , const DP );
    DP CalculScore(Vec_I_DP &,Mat_I_DP &,int);
    int CtoV(int);
    int VtoC(int);
};

```

La classe *Experience* est la classe de base de tous les algorithmes. Elle correspond à une expérience ou une cellule contenant à la fois des réactifs et des réactions, et des outils associés. Le champ *Score* contient le score de la cellule et sert à classer les cellules dans les algorithmes d'évolution. Les indices *NombreReactifs*, *NombreReactions*, *NombreGenes*, *NombrePromus* sont des entiers permettant de connaître les nombres respectifs de réactifs (uniquement protéines et ARN), de réactions, de gènes et de complexes protéines ADN possibles dans la cellule. Le

tableau `Reactifs` est un tableau d'`Espece`, répertoriant toutes les espèces biochimiques présentes dans la cellule. Ce tableau est organisé comme suit : les gènes seront répertoriés entre les indices 0 et `NombreGenes-1`, les complexes protéines-ADN entre les indices `NGenesMax` et `NGenesMax+NombrePromus-1`, les autres espèces (protéines et éventuellement ARN) entre les indices `NGenesMax+NPromusMax` et `NGenesMax+NPromusMax+NombreReactifs-1`.

Le tableau `Reaction` contient simplement la liste des réactions possibles dans la cellule. Les numéros de réactif indiqués dans les réactions correspondent aux indices du tableau `Reactifs`.

```
class Milieu{
public:
    Experience Cellule[TailleMilieu]; // Cette classe sert à définir des collections
    //Méthodes de la classe Milieu de cellules, en particulier pour l'algo-
    void sort(); rithme d'évolution. Les cellules sont conte-
    void Selection(int &); nues dans un tableau.
    void Generation(int &,Mat_DP &);
};
```

B.II Exemple de code

Nous reproduisons dans la suite un exemple de code complet ayant servi à générer des oscillateurs avec ARN, en utilisant les types prédéfinis de numerical recipes.

Evolution.cc

```

#include "nr.h"
#include "nrutil_nr.h"
#include <cmath>
#include <stdio.h>
#include <ctype.h>
#include <stdlib.h>
#include <string.h>

const int NReactifsMax=10;
const int NPromusMax=10;
const int NReactionsMax=100;
const int NGenesMax=4;

const int TailleMilieu=100;
const int Ngen=500;
const int d=NPromusMax*NGenesMax; //debut des proteines dans le tableau

DP ranl(int &idum)
{
    const int IA=16807, IM=2147483647, IQ=127773, IR=2836, NTAB=32;
    const int NDIV=(1+IM-1)/NTAB;
    const DP EPS=3.0e-16, AM=1.0/IM, RNMX=(1.0-EPS);
    static int iy=0;
    static Vec_INT iv(NTAB);
    int j,k;
    DP temp;

    if (idum <= 0 || !iy) {
        if (-idum < 1) idum=1;
        else idum=-idum;
        for (j=NTAB+7; j>=0; j--) {
            k=idum/IQ;
            idum=IA*(idum-k*IQ)-IR*k;
            if (idum<0) idum +=IM;
            if (j<NTAB) iv[j]=idum;
        }
        iy=iv[0];
    }
    k=idum/IQ;
    idum=IA*(idum-k*IQ)-IR*k;
    if (idum<0) idum +=IM;
    j=iy/NDIV;
    iy=iv[j];
    iv[j]=idum;
    if ((temp=AM*iy) > RNMX) return RNMX;
    else return temp;
}

class Espece{//type Espece chimique: un nombre et une constante de degradation,un ta
bleau pour indiquer les composants
public:
    DP Constante;
    DP Qte;
    int Composants[NReactifsMax];
};

class Reaction//type Reaction chimique:Une constante, deux reactifs,deux produits; -
1 si on veut des reactifs ou des produits en moins

```

Evolution.cc

```

public:
    DP Constante;
    int Reactif1;
    int Reactif2;
    int Produit1;
    int Produit2;
};

class Experience{//Une experience:des reactifs et des reactions
public:
    DP Score;
    int NombreReactions;
    int NombreReactifs;
    int NombreGenes;
    int NombrePromus;
    Espece Reactifs[NReactifsMax+NGenesMax+NPromusMax]);//Place pour les genes +genesPr
omus+Proteines Ordre
    //Les places dans le tableau sont importantes pour respecter les types de reaction
    Reaction Reactions[NReactionsMax];
    int ARN[NGenesMax]; // Liste des ARN

    void init();
    void AjoutEspece(DP,DP);
    void AjoutDimere(DP,DP, int, int);
    void Fusion(int, int, int);
    void AjoutReaction(int, int, int, int, DP);
    void AjoutGene(DP, DP, DP, DP);
    void AjoutPromotion(int, int, DP, DP, DP);
    void AjoutBS(int, int, DP, DP, DP);
    void Degradation(int);
    void Etape(int);
    void proba(Vec_O_DP &);
    void ModifEspece(int, int &);
    void ModifReaction(int, int &);
    void ModifProm(int, int &);
    void AjoutPostTrans(int &);
    void RetireReaction(int);

    void Evolution(int &);
    void derives(const DP, Vec_I_DP &, Vec_I_DP &, Vec_O_DP &);
    void rk4(Vec_I_DP &, Vec_I_DP &, Vec_I_DP &, const DP, const DP, Vec_O_DP &);
    void rkdumb(Vec_IO_DP &, Mat_IO_DP &, const DP, const DP, const DP);
    DP CalculScore(Vec_I_DP &, Mat_I_DP &, int);
    int CtoV(int);
    int VtoC(int);
    int TestARN(int);

    void print();
    int TestComposant(int, int, int);
    int Testtranscrit(int, int, int);
    int Retourneproteine(int);
};

class Milieu{
public:
    Experience Celluie[TailleMilieu];
    void sort();
    void Selection();
    void Generation(int &, Mat_DP &);
};

void Experience::init(){//initialisation de l'experience
    Score=0.0;
    NombreReactions=0;
    NombreReactifs=0;
    NombreGenes=0;
    NombrePromus=0;
}

void Experience::AjoutEspece(DP Q,DP c){//Rajoute une proteine exterieure dans le mi

```

```

lieu
if (NombreReactifs<NReactifsMax)
{
    Reactifs[NombreReactifs+nGenesMax+NPromusMax].Composants[0]=NombreReactifs+nGe
nesMax+NPromusMax;
    for (int i=1;i<NReactifsMax;i++){Reactifs[NombreReactifs+nGenesMax+NPromusMax]
.Composants[i]=0;}
    Reactifs[NombreReactifs+nGenesMax+NPromusMax].Qte=0;//Q: quantite de reactif a
joute
    Reactifs[NombreReactifs+nGenesMax+NPromusMax].Constante=c;//c constante de rea
ction
    NombreReactifs++;
}

void Experience::Fusion(int i,int j,int k){// met les composants de i et j dans les
composants de k

    int Composant[NReactifsMax];
    int cont=0;
    //printf(" Entree Fusion \n");
    for (cont:=(cont<NReactifsMax)&&(Reactifs[i].Composants[cont]!=0);cont++){
        Composant[cont]=Reactifs[i].Composants[cont];
    }

    if (j!=-1){
        for (int indexj=0;((Reactifs[j].Composants[indexj]!=0)&&(indexj<NReactifsMax));in
dexj++){
            int test=0;
            for (int indexk=0;(indexk<cont)&&(test==0);indexk++){
                if (Composant[indexk]==Reactifs[j].Composants[indexj]) {test=1;} //teste s'il
y a des composants communs
            }
            if (test==0) {Composant[cont]=Reactifs[j].Composants[indexj];cont++;}
        }
    }

    Composant[cont]=k;
    for (int indexi=0;indexi<cont+1;indexi++){
        Reactifs[k].Composants[indexi]=Composant[indexi];
    }
}

void Experience::AJoutDimere(DP Q,DP c,int i,int j){//Rajoute une proteine exterieur
e dans le milieu
    if (NombreReactifs<NReactifsMax)
    {
        for (int co=0;co<NReactifsMax;co++)
        {Reactifs[NombreReactifs+nGenesMax+NPromusMax].Composants[co]=0;}
    }
    Fusion(i,j,NombreReactifs+nGenesMax+NPromusMax);
    Reactifs[NombreReactifs+nGenesMax+NPromusMax].Qte=Q;//Q: quantite de reactif ajo
ute
    Reactifs[NombreReactifs+nGenesMax+NPromusMax].Constante=c;//c constante de react
ion
    NombreReactifs++;
}

```

```

void Experience::ModifEspece(int i,int &a){//Modifie le taux de degradation d'une pr
oteine quelconque
    if (i<NombreReactifs){int k=i+nGenesMax+NPromusMax;
    DP c=Reactifs[k].Constante;
    DP r=rantl(a);
    c=r**2;
    if (c<0.00001) c=rantl(a);
    Reactifs[k].Constante=c;
    }
}

void Experience::ModifReaction(int i,int &a){
    if (i<NombreReactions){
        DP c=Reactions[i].Constante;
        DP r=rantl(a);
        c*=r**2;
        if (c<0.00001) c=rantl(a);
        Reactions[i].Constante=c;
    }
}

void Experience::RetireReaction(int k){
    if (NombreReactions>0){
        for(int i=k;i<NombreReactions-1;i++){
            Reactions[i]=Reactions[i+1];
            NombreReactions--;
        }
    }
}

bool Comp(int i,int k,int l,int m){
    if ((i==l)&&(k==m)) return false;
    if ((i==m)&&(k==l)) return false;
    return true;
}

void Experience::AJoutReaction(int i,int k,int l,int m,DP Q){//les entiers codent po
ur les cases du tableau Reactifs, l'ordre est Reactif1, Reactif2, Produit1, Produit2
    if ((NombreReactions<NReactionsMax)&&(i==1)&&(l==1)&&(Comp(i,k,l,m)) //si l'un
des entiers est -1, le reactif correspondant est NULL; i et l ne doivent pas prendre
la valeur -1
    {
        int j=NombreReactions;
        Reactions[j].Constante=Q;

        Reactions[j].Reactif1=i;
        Reactions[j].Reactif2=k;
        Reactions[j].Produit1=l;
        Reactions[j].Produit2=m;
        NombreReactions++;
        //printf(" reaction ajoutee %i %i %i \n", i,k,l,m);
    }
}

void Experience::AJoutGene(DP transcri,DP transla,DP dR,DP dP){
    if ((NombreGenes<NGenesMax)&&(NombreReactions<NReactionsMax-2)&&(NombreReactions<NR
eactionsMax-2)){
        int i=NombreGenes;
        int j=NombreReactions+nGenesMax+NPromusMax;//numero du gene et de la proteine
        Reactifs[i].Qte=1.0;//Ajoute le gene
        Reactions[i].Constante=0;
    }
}

```

```

AjouteEspece(5.0,dR); //Ajoute l'ARN
ARN[i]=j;
AjouteReaction(i,-1,i,j,transcrit); //Ajoute la transcription
AjouteEspece(5.0,dP); //Ajoute la proteine
AjouteReaction(j,-1,j,j+,transla); //Ajoute la traduction
NombreGenes++;
}
}

void Experience::AjoutPromotion(int G,int Prot,DP c,DP e,DP q){ //le gene est l'entier
er G, la proteine l'entier Prot
if ((NombreGenes>0)&&(NombrePromus<NPromusMax)&&(NombreReactions<NReactionsMax-3))
{
int i=NombrePromus+NGenesMax;
NombrePromus++;
Reacts[i].Qte=0;
Reacts[i].Constante=0;
int p=0; //entier de la proteine produite par le gene promu
for (int k=0;k<NombreReactions;k++)
{ if ((Reactions[k].Reactif1==G)&&(Reactions[k].Produit2>=NPromusMax+NGenesMax
)) p= Reactions[k].Produit2;
AjouteReaction(i,-1,p,i,c); //Ajoute la transcription promue
AjouteReaction(G,Prot+NGenesMax+NPromusMax,i,-1,e); //Ajoute la complexation
AjouteReaction(i,-1,G,Prot+NGenesMax+NPromusMax,q); //Fin de la complexation
}
}

void Experience::AjoutBS(int Prom,int Prot,DP c,DP e,DP q){ //le gene est l'entier G
, la proteine l'entier Prot
if ((NombreGenes>0)&&(NombrePromus<NPromusMax)&&(NombreReactions<NReactionsMax-3))
{
int i=NombrePromus+NGenesMax;
NombrePromus++;
Reacts[i].Qte=0;
Reacts[i].Constante=0;
int p=0; //entier de la proteine produite par le gene promu
for (int k=0;k<NombreReactions;k++)
{ if ((Reactions[k].Reactif1==Prom)&&(Reactions[k].Produit1>=NPromusMax+NGenes
Max)) p= Reactions[k].Produit1;
AjouteReaction(i,-1,p,i,c); //Ajoute la transcription promue
AjouteReaction(Prom,Prot+NGenesMax+NPromusMax,i,-1,e); //Ajoute la complexation
AjouteReaction(i,-1,Prom,Prot+NGenesMax+NPromusMax,q); //Fin de la complexation
}
}

int Experience::TestARN(int a){
int s=0;
int p=0;
if (a!=1){
do
{
for (int i=0;(i<NReactsMax)&&(Reacts[i].Composants[i]!=0);i++){
int b=Reacts[i].Composants[i];
if (ARN[p]==b) {s=1;return s;}
p++;
} while ((p<NGenesMax)&&(ARN[p]!=0));
}
return s;
}
}

int Experience::TestComposant(int i,int k,int l){ //regarde si l est un composant de
i et de k

```

```

int s=0;
for (int indexi=0;((indexi<NReactsMax)&&(Reacts[i].Composants[indexi])!=0);ind
exi++){
if (Reacts[i].Composants[indexi]==l) {return 1;}
}

if (k!=1){ for (int indexk=0;((indexk<NReactsMax)&&(Reacts[k].Composants[inde
xk])!=0);indexk++){
if (Reacts[k].Composants[indexk]==l) {return 1;}}
}
return s;
}

int Experience::TestTranscrit(int i,int k,int l){ //regarde si l peut etre transcrit
a partir de i ou de k
if (TestARN(i)==1){
if (l==Retourneproteine(i)) return 1;
}
}

if (TestARN(k)==1){
if (l==Retourneproteine(k)) return 1;
}
}

return 0;
}

int Experience::Retourneproteine(int i){ //retourne la proteine produite par un trans
crit modifie i
for (int p=0;(p<NGenesMax)&&(ARN[p]!=0);p++){
int b=ARN[p];
for (int l=0;((l<NReactsMax)&&(Reacts[i].Composants[l]==0);l++){
if (Reacts[i].Composants[l]==b) return b+1;
}
}

void Experience::print(){
printf(" Es.NombreGenes=%i; \n",NombreGenes);
printf(" Es.NombrePromus=%i; \n",NombrePromus);
for (int k=NGenesMax+NPromusMax;k<NombreReacts+NPromusMax+NGenesMax;k++){
printf(" Es.AjouteEspece(%f, %f); \n",Reacts[k].Qte,Reacts[k].Constante);
}

for (int k=0;k<NombreReactions;k++){
printf(" Es.AjouteReaction(%i,%i,%i, %i, %f); \n", Reactions[k].Reactif1,Reaction
s[k].Reactif2, Reactions[k].Produit1,Reactions[k].Produit2,Reactions[k].Constante);
}
}

```

```

for (int j=d;j<d+NombreReactifs;j++)
{
    printf(" \n Les composants de %i sont ", j);
    for (int indexj=0;((Reactifs[j].Composants[indexj]!=0)&&(indexj<NReactifsMax));
        indexj++) printf(" %i ", Reactifs[j].Composants[indexj]);
    printf(" \n");
}
}

void Experience::AjoutPostTrans(int &a) {
    int i;
    do {
        i = int(ranl(a)*(NombreReactifs)+NGenesMax+NPromusMax); while (TestARN(i)==1);
        int k;
        do {
            k = (int(ranl(a)*(NombreReactifs+1))-1);
            while ((k!=i)&&(TestARN(k+NGenesMax+NPromusMax)==1));
            if (k==i) k+=NGenesMax+NPromusMax;
            DP r=ranl(a);
            int l;
            int m;
            if (k==i) {
                if ((TestARN(i)==1)&&(ranl(a)>0.5))
                { l=Retourneproteine(i);
                  m=i;
                } // Ajoute une transcription
                else {
                    if ((r>1.0/NombreReactifs)|| (NombreReactifs==NReactifsMax))
                    {
                        do { l = int(ranl(a)*(NombreReactifs)+NGenesMax+NPromusMax); while ((Te
                            stComposant(i,k,l)==0)&&(TestTranscrit(i,k,l)==0));
                            if (TestTranscrit(i,k,l)!=0)
                            { m=i;
                              } else { m=-1; } // Ajoute une dégradation d'un dimere
                        }
                    } else {
                        AjoutDimere(5.0, ranl(a), i, k);
                        l=NombreReactifs-1+NGenesMax+NPromusMax;
                        m=-1; // Ajoute une modification post trans
                    }
                }
            } else
            { // on rentre dans le cas ou on a deux reactifs
                m=-1;
                r=ranl(a);
                if ((r>1.0/NombreReactifs)|| (NombreReactifs==NReactifsMax))
                {
                    do { l = int(ranl(a)*(NombreReactifs)+NGenesMax+NPromusMax); while
                        e ((TestComposant(i,k,l)==0)); // redonne un composant de depart
                    }
                    else {
                        AjoutDimere(5.0, ranl(a), i, k);
                        l=NombreReactifs-1+NGenesMax+NPromusMax; //ajout d'un dimere
                    }
                }
            }
        }
    }
}

```

```

AjoutReaction(i,k,l,m,ranl(a));
}

void Experience::Evolution(int &a) { //procedure d'evolution
    DP tauxNewR=1; //Taux relatif d'apparition d'une reaction
    DP tauxNewE=0; //Taux relatif d'apparition d'une espece
    DP tauxmodifR=5; //Taux relatif de modification d'une reaction
    DP tauxNewG=0.1; //Taux relatif de modification d'une espece
    DP tauxNewP=0.1; //Taux relatif d'apparition d'un gene
    DP tauxmodifQteR=1;

    //a variable muette pour le tirage au sort
    Vec_DP p(8);
    p[0]=tauxNewR;
    p[1]=tauxNewE;
    p[2]=tauxmodifR;
    p[3]=tauxmodifE;
    p[4]=tauxNewG;
    p[5]=tauxNewP;
    p[6]= tauxmodifQteR;
    p[7]=tauxNewR+tauxNewE+tauxmodifR+tauxmodifE+tauxNewG+tauxNewP+tauxmodifQteR;
    DP r=ranl(a)*p[7];
    DP inf=0.0;
    int i=0;
    do {
        inf +=p[i];
        i+=1;
    } while ((r>inf) );
    int Mu= i-1; //tire au sort la modification a effectuer

    if (Mu==0) {
        AjoutPostTrans(a);
    } //ajouter une reaction
    if (Mu==1) { AjoutEspece(5.0,ranl(a)); } //ajouter une espece
    if (Mu==2) {
        if (NombreReactions>0) { DP r=ranl(a);
            int j=int(ranl(a)*(NombreReactifs));
            ModifEspece(j,a);
        }
    } //modifier une reaction
    if (Mu==3) {
        if (NombreReactifs>0) {
            int j=int(ranl(a)*(NombreReactifs));
            ModifEspece(j,a);
        }
    } //MODIFIER une espece
    if (Mu==4) {
        AjoutGene(ranl(a), ranl(a), ranl(a), ranl(a));
    } //ajoute un gene
    if (Mu==5) {
        int j;
        do {
            j = int(ranl(a)*(NombreReactifs));
        } while (TestARN(j+d)==1);
        int k = int(ranl(a)*(NombreGenes+NombrePromus));
        if (k<NombreGenes)
            AjoutPromotion(k, j, ranl(a), ranl(a), ranl(a));
        else
            AjoutES(k-NombreGenes+NGenesMax, j, ranl(a), ranl(a), ranl(a));
    } //ajoute une promotion
    if (Mu==6) {

```

```

int j= int( rani(a)*(NombreReactifs));
Reactifs[j+d].Qte*=2*rani(a);
if (Reactifs[j+d].Qte<0.01) Reactifs[j+d].Qte=1;
}
}

int Experience::VtoC(int j)//ProcEDURE permettant de passer du vecteur a la cellule
{if (j<NombreGenes) return j;
if (j<NombreGenes+NombrePromus) return j-NombreGenes+NGenesMax;
return j-NombreGenes-NombrePromus+NGenesMax+NPromusMax;
}

int Experience::CtoV(int j)//ProcEDURE permettant de passer de la cellule au vecteur
{if (j<NGenesMax) return j;
if (j<NGenesMax+NPromusMax) return j+NombreGenes+NGenesMax;
return j+NombreGenes+NombrePromus-NGenesMax-NPromusMax;
}

void Experience::derivs(const DP x, Vec_I_DP &y, Vec_O_DP &dy){//Calcul des derivees
int nvar=NombreGenes+NombrePromus+NombreReactifs;
for (int j=0;j<nvar;j++) dy[j]=0; //initialisation

for (int j=0;j<NombreReactifs;j++)
{dy[j+NombreGenes+NombrePromus]=-y[j+NombreGenes+NombrePromus]*Reactifs[j+NGenes
Max+NPromusMax].Constante; //degradation de la proteine
for (int j=0;j<NombreReactions;j++)
{
int R1=Reactions[j].Reactif1;int r3= Reactions[j].Produit1;
int R2=Reactions[j].Reactif2;int r4=Reactions[j].Produit2;
DP c=Reactions[j].Constante;
R1=CtoV(R1);
R2=CtoV(R2);
r3= CtoV(r3);
r4=CtoV(r4);
if (R2==1) {
dy[R1]-=c * y[R1];
dy[r3]+=c * y[R1];
if (r4!=1) dy[r4]+=c * y[R1];
}
}
}
}

void Experience::rk4(Vec_I_DP &y, Vec_I_DP &xx, const DP x, const DP h, Vec_O_DP &y
out){
int i;
DP xh,hh,h6;
int n=y.size();
Vec_DP dym(n),dyt(n),yt(n);
hh=h*0.5;
h6=h/6.0;
xh=x+hh;
for (i=0;i<n;i++) yt[i]=y[i]+hh*dydx[i];
derivs(xh,yt,dyt);
for (i=0;i<n;i++) yt[i]=y[i]+hh*dyt[i];
derivs(xh,yt,dym);
for (i=0;i<n;i++) {
yt[i]=y[i]+h*dym[i];
dym[i]=dyt[i];
}
}

```

```

derivs(x+h,yt,dyt);
for (i=0;i<n;i++){
yout[i]=y[i]+h6*(dydx[i]+dyt[i]+2.0*dym[i]);
}
}

void Experience::rkdumb(Vec_IO_DP &xx,Mat_IO_DP &y, const DP x1, const DP x2){
int i,k;
int nvar=y.nrows();
int nstep=y.ncols()-1;
Vec_DP V(nvar), vout(nvar), dv(nvar);
for (int w=0;w<NombreGenes;w++) { v[w]=Reactifs[w].Qte;
y[w][0]=v[w];}

for (int w=0;w<NombrePromus;w++) {
v[w+NombreGenes]=Reactifs[w+NGenesMax].Qte;
y[w+NombreGenes][0]= v[w+NombreGenes];
}

for (int w=0;w<NombreReactifs;w++) {
v[w+NombreGenes+NombrePromus]=Reactifs[w+NGenesMax+NPromusMax].Qte;
y[w+NombreGenes+NombrePromus][0]= v[w+NombreGenes+NombrePromus];}

xx[0]=x1;
x=x1;
h=(x2-x1)/nstep;
for (k=0;k<nstep;k++){
derivs(x,v,dv);
rk4(v,dv,x,h,vout);
if (x+h ==x){
printf("Step size too small in routine rkdumb");
break;}
x+=h;
xx[k+1]=x;
for (i=0;i<nvar;i++) {
v[i]=vout[i];
y[i][k+1]=v[i];
}
}
}

DP Experience::CalculScore(Vec_I_DP &xx, Mat_I_DP &y,int ng){//fonction de score pou
r les oscillateurs
int N=NombreReactions;
int nvar=y.nrows();
int rep=NombreGenes+NombrePromus;
DP s=0;
int nstep=xx.size();
for (int i=0;i<100;i++){
int v=rep+1;
if (ng==1)
s+= ((y[v][i])*(y[v][i]));
else
s+= ((y[v][i]-10)*(y[v][i]-10));
}
}

```

```

    }
    //selon les cond initiales impose deux etats differents

    s*=exp(N*log(1.0001))/100.0;

    if (s<=le10) return s; //filtre les gros scores
    else return le10;
}

void SWAPM(Experience &a, Experience &b){
    Experience c;
    c.init();
    c=b;
    b=a;
    a=c;
}

void Milieu::sort()
{ const int M=7, NSTACK=50;
  int i,ir,j,k,jstack=-1,l=0;
  Experience a;
  a.init();
  Vec INT istack(NSTACK);
  int n=TailleMilieu;
  ir=n-1;
  for ( ;; ) {
    if (ir-1 < M) {
      for (j=l+1;j<=ir;j++) {
        a=Cellule[j];
        for (i=j-1;i>=l;i--) {
          if (Cellule[i].Score <= a.Score) break;
          Cellule[i+1]=Cellule[i];
          Cellule[i+1]=a;
        }
        if (jstack < 0) break;
        ir=istack[jstack--];
        l=istack[jstack--];
      } else {
        k=(l+ir) >> 1;
        Cellule[k]=Cellule[l+1];
        if (Cellule[l].Score > Cellule[ir].Score) {
          SWAPM(Cellule[l],Cellule[ir]);
        }
        if (Cellule[l+1].Score > Cellule[ir].Score) {
          SWAPM(Cellule[l+1],Cellule[ir]);
        }
        if (Cellule[l].Score > Cellule[l+1].Score) {
          SWAPM(Cellule[l],Cellule[l+1]);
        }
        l=l+1;
        j=ir;
        a=Cellule[l+1];
        for ( ;; ) {
          do i++; while (Cellule[i].Score < a.Score);
          do j--; while (Cellule[j].Score > a.Score);
          if (j < i) break;
          SWAPM(Cellule[i],Cellule[j]);
        }
      }
    }
  }
}

```

```

    Cellule[l+1]=Cellule[j];
    Cellule[j]=a;
    jstack += 2;
  }
  if (jstack >= NSTACK) exit(1);
  if (ir-i+1 >= j-1) {
    istack[jstack]=ir;
    istack[jstack-1]=i;
    ir=j-1;
  } else {
    istack[jstack]=j-1;
    istack[jstack-1]=l;
    l=i;
  }
}

void Milieu::Selection() {
  int nstep=2000;
  Vec_DP xx_p(nstep+1);
  for (int i=0;i<TailleMilieu;i++) {
    int nvar=Cellule[i].NombreReactifs+Cellule[i].NombreGenes+Cellule[i].NombrePromus;
    Mat_DP Y_p(nvar,nstep+1);
    Experience Es=Cellule[i];
    //Y_P contient les concentrations dans l'ordre : Genes, Promus, Proteines
    //printf(" Score%i=%f\n",i,Cellule[i].Score);
    int ng=1;
    DP s=0;
    for (int k=0;k<8;k++) {
      Es.rkdumb(xx_p,Y_p,k*11,(k+1)*11);
      s+=Es.CalculScore(xx_p,Y_p,ng);
      for (int w=0;w<Cellule[i].NombreGenes;w++) { Es.Reactifs[w].Qte=Y_p[w][nstep];
      }
      for (int w=0;w<Es.NombrePromus;w++) { Es.Reactifs[w+NGenesMax].Qte=Y_p[w+Es.NombreGenes][nstep];
      }
      for (int w=0;w<Es.NombreReactifs;w++) { Es.Reactifs[w+d].Qte=Y_p[w+Es.NombreGenes+Es.NombrePromus][nstep];
      }
      ng=1-ng;
    }
    Cellule[i].Score=s;
  }
  sort();
}

void Milieu::Generation(int &a,Mat_IO_DP &m) {
  for (int i=0;i<TailleMilieu;i++)

```

```

Cellule[i].init();

Cellule[i].AjoutGene(ranl(a), ranl(a), ranl(a), ranl(a));
Cellule[i].AjoutGene(ranl(a), ranl(a), ranl(a), ranl(a));

for (int w=0;w<Cellule[i].NombreGenes;w++) {Cellule[i].Reactifs[w].Qte=L.0;
    Cellule[i].Reactifs[1].Qte=L-Cellule[i].Reactifs[NGenesMax].Qte;
    for (int k=0;k<2;k++) {
        Cellule[i].Evolution(a);
    }
}
for (int j=0;j<Ngen;j++){
    for (int i=TailleMilieu/2;i<TailleMilieu;i++){
        Cellule[i]=Cellule[int(i-TailleMilieu/2)];
        for (int k=0;k<2;k++) {Cellule[i].Evolution(a);}
    }
}

```

```

    printf("Generation %i \n", j);
    printf(" Avant %f \n", Cellule[0].Score);

    m[0][j]=j;
    m[1][j]=Cellule[0].Score;
    Experience Es=Cellule[0];
    for (int k=0;k<NGenesMax;k++) printf(" %i ", Es.ARN[k]);
    int nvar=NReactifsMax+NGenesMax+NPromusMax;
    printf(" Es.NombreGenes=%i; \n", Es.NombreGenes);
    printf(" Es.NombrePromus=%i; \n", Es.NombrePromus);
    for (int k=NGenesMax+NPromusMax;k<Es.NombreReactifs+NPromusMax+NGenesMax;k++){
        printf(" Es.AjoutEspece(%f, %f):\n", Es.Reactifs[k].Qte, Es.Reactifs[k].Constance);
    }
}
}
}
}

```

```

for (int k=0;k<Es.NombreReactions;k++){
    printf ("Es.AjoutReaction(%i,%i,%i,%f); \n", Es.Reactions[k].Reactif1,Es.Reactions[k].Reactif2, Es.Reactions[k].Produit1,Es.Reactions[k].Produit2,Es.Reactions[k].Constance);}
Selection();
}
}
}

```

```

int main(){
    Milieu m;Mat_DP s(2,Ngen);
    int a;
    for (int b=-30;b>=60;b--) {
        a=b;
    }
    m.Generation(a,s);

    FILE *File;
    char titre[50];
    sprintf(titre, "Scoresb=%i", b);
    File=fopen(titre, "w");
    for (int l=0;l<Ngen;l++){
        fprintf(File, "%i %f \n", (int) s[0][l], s[1][l]);
        printf(" %i %i f=%f", s[0][l], s[1][l]);
    }
    fclose(File);
    FILE *fileprt;
    sprintf(titre, "ResultatsTotauxb=%i", b);
    fileprt=fopen(titre, "w");
    fprintf(fileprt, "Algorithme RNA, Parametres NReactifsMax=%i, NPromusMax=%i, NReactionsMax=%i, NGenesMax=%i,TailleMilieu=%i, %i Generations \n",NReactifsMax,NPromusMax,NReactionsMax,NGenesMax,TailleMilieu,Ngen);
}

```

```

for (int l=0;l<TailleMilieu;l++){
    Experience Es=m.Cellule[l];

    for (int k=0;k<NGenesMax;k++) printf(" %i ", Es.ARN[k]);
    int nvar=NReactifsMax+NGenesMax+NPromusMax;
    fprintf(fileprt, " Es.NombreGenes=%i; \n", Es.NombreGenes);
    fprintf(fileprt, " Es.NombrePromus=%i; \n", Es.NombrePromus);
    for (int k=NGenesMax+NPromusMax;k<Es.NombreReactifs+NPromusMax+NGenesMax;k++){
        fprintf(fileprt, " Es.AjoutEspece(%f, %f):\n", Es.Reactifs[k].Qte, Es.Reactifs[k].Constance);}
}
}
}
}

```

```

for (int k=0;k<Es.NombreReactions;k++){
    fprintf (fileprt, "Es.AjoutReaction(%i,%i,%i,%f); \n", Es.Reactions[k].Reactif1,Es.Reactions[k].Reactif2, Es.Reactions[k].Produit1,Es.Reactions[k].Produit2,Es.Reactions[k].Constance);}
}
}
}
}

```

Appendice C

Modèle avec délai

Nous donnons dans cette partie quelques indications sur l'étude de la version adimensionnée du modèle d'oscillateur avec délai III.88. Plus généralement, nous considérons le modèle adimensionné défini par l'équation suivante :

$$\varepsilon \frac{db}{d\theta} = \frac{r}{1 + b^n(\theta - 1)} - b = f(b(\theta - 1)) - b, \quad (\text{C.1})$$

qui correspond au modèle présenté dans l'équation III.89.

C.I Diagramme de phase

L'étude complète de la stabilité du point fixe pour des équations différentielles avec délai telles que celle présentée en C.1 peut être trouvée dans la référence [40]. Ces différentes conclusions sont également reprises dans le livre de Glass et Mackey cité en référence [34]. Dans cette section, nous rappelons sommairement quelques-uns de ces résultats. Nous considérons le cas où $f(b) = \frac{r}{1+b^n}$, mais les résultats peuvent être facilement généralisés à des fonctions plus compliquées. Soit b^* le point fixe. Nous allons retrouver les conditions pour avoir un point fixe instable en calculant explicitement des conditions sur le taux de croissance complexe $\mu + i\omega$ tel qu'une perturbation croisse exponentiellement comme $\exp((\mu + i\omega)\theta)$. Nous utilisons donc l'Ansatz suivant : $b = b^* + \rho e^{(\mu + i\omega)\theta}$.

L'équation C.1 devient alors :

$$\varepsilon(\mu + i\omega)\rho e^{(\mu + i\omega)\theta} = -\frac{nr b^{*(n-1)}}{(1 + b^{*n})^2} \rho e^{(\mu + i\omega)(\theta-1)} - \rho e^{(\mu + i\omega)\theta} \quad (\text{C.2})$$

soit

$$\varepsilon(\mu + i\omega) = -\frac{nr b^{*(n-1)}}{(1 + b^{*n})^2} e^{-(\mu + i\omega)} - 1 \quad (\text{C.3})$$

Identifiant les parties réelles et imaginaires, on trouve :

$$\varepsilon\mu = -\frac{nr b^{*(n-1)}}{(1+b^{*n})^2} e^{-\mu} \cos \omega - 1 \quad (\text{C.4})$$

$$\varepsilon\omega = \frac{nr b^{*(n-1)}}{(1+b^{*n})^2} e^{-\mu} \sin \omega \quad (\text{C.5})$$

ou en remplaçant dans la seconde équation :

$$\tan \omega = -\frac{\varepsilon\omega}{1 + \varepsilon\mu} \quad (\text{C.6})$$

À la bifurcation, la partie réelle du taux d'accroissement s'annule et l'équation C.4 impose que $\cos \omega = -\frac{(1+b^{*n})^2}{nr b^{*(n-1)}}$, soit $\omega = \cos^{-1} \left(-\frac{(1+b^{*n})^2}{nr b^{*(n-1)}} \right)$. L'équation C.6 impose alors la valeur de ε à la bifurcation pour $\omega \in [0, \pi]$. Cette valeur critique de ε est donnée par :

$$\varepsilon_c = \frac{\left(\left(\frac{nr b^{*(n-1)}}{(1+b^{*n})^2} \right)^2 - 1 \right)^{1/2}}{\cos^{-1} \left(-\frac{(1+b^{*n})^2}{nr b^{*(n-1)}} \right)} \quad (\text{C.7})$$

On a donc une première condition nécessaire $\frac{(1+b^{*n})^2}{nr b^{*(n-1)}} < 1$, soit après transformation en utilisant l'équation sur b^* $(n-1)r > nb^*$. On voit tout de suite qu'il faut alors nécessairement $n > 1$ pour avoir un point fixe instable.

Maintenant, les oscillations ne sont possibles que pour $\varepsilon < \varepsilon_c$. En effet, une simple résolution graphique montre que les solutions de l'équation $\varepsilon\omega = -\tan \omega$ tendent vers $\pi/2^+$ lorsque ε est grand. Dans ce cas, $\cos(\omega)$ devient petit et μ est alors négatif, donc le point fixe est stable. Le point fixe sera donc instable si :

$$\frac{(1+b^{*n})^2}{nr b^{*(n-1)}} < 1 \quad \text{et} \quad \varepsilon < \varepsilon_c \quad (\text{C.8})$$

A n fixé, on peut en tirer ensuite une condition sur r : par exemple, pour $n = 2$, cette condition impose que $r > 2b^*$, soit $b^3 - b^* > 0$, soit $b^* > 1$, ce qui n'est possible que si $r > 2$.

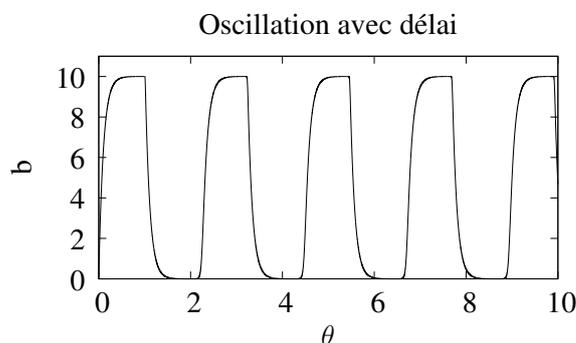


FIG. C.1: Oscillation de l'oscillateur défini dans l'équation C.1, avec $f(b) = 10/(1 + b^5)$ et $\varepsilon = 0.1$.

C.II Mécanisme des oscillations

Dans la limite des petits ε , il est possible de faire une approximation quasi-statique dans l'équation C.1 qui nous donne immédiatement l'équation suivante pour b :

$$b = f(b(\theta - 1)) \quad (\text{C.9})$$

Le comportement du cycle dépendra alors de la forme de f . En général, il existe deux points b_1 et b_2 tels que $b_1 = f(b_2)$ et $b_2 = f(b_1)$. L'oscillateur restera alors successivement pendant un temps d'ordre 1 à des concentrations proches de b_1 puis de b_2 . Il "sautera" littéralement d'une concentration à l'autre pendant un temps très court d'ordre ε (voir Figure C.1).

L'analyse complète du cycle limite à l'ordre ε est beaucoup plus complexe et n'a pas été complètement résolue à notre connaissance. Cependant, nous proposons ici une analyse dans un cas particulier de la fonction f où les équations sont intégrables explicitement, qui correspond à la limite n infini du cas où $f(b) = \frac{r}{1+b^n}$. En effet, pour n très grand, f vaut r dès que $b < 1$ et f vaut 0 sinon. On a donc en fait $f = r\Theta(1-b)$ où Θ est la fonction de Heavyside ($\Theta(x) = 1$ si $x > 0$, sinon $\Theta(x) = 0$). Dans ce cas, la concentration b oscille entre 0 et r . Plus précisément, prenons comme origine des temps dans le cycle limite un moment où $b = 1$ en décroissant. Si l'on suppose que la période pendant laquelle b reste supérieur à 1 est plus grande que 1, $\Theta(b(\theta - 1))$ est nul entre 0 et 1 et l'équation pour b pour $\theta \in [0, 1]$ est alors :

$$b(\theta) = e^{-\theta/\varepsilon} \quad (\text{C.10})$$

Ensuite, à $t = 1$, $b(\theta - 1)$ devient plus petit que 1, donc $\Theta(1 - b(\theta - 1))$ prend la valeur 1 tant que b ne repasse pas sous 1. L'équation pour b est alors

$$b(\theta) = r + (b_1 - r)e^{-(\theta-1)/\varepsilon} \quad (\text{C.11})$$

avec $b_1 = e^{-1/\varepsilon}$. Cette équation est valable jusqu'à ce que $b(\theta_1 - 1) = 1$. On trouve alors la valeur de θ_1 , en faisant un développement de Taylor de l'équation C.11 en θ à l'ordre ε :

$$\theta_1 = 2 + \varepsilon \ln\left(\frac{r}{r-1}\right) \quad (\text{C.12})$$

Ensuite, b décroît jusqu'à atteindre 1 en $\theta = \theta_2$ pour clore le cycle. L'équation pour b entre θ_1 et θ_2 est alors :

$$b(\theta) = b_2 e^{-(\theta-\theta_1)/\varepsilon} \quad (\text{C.13})$$

avec $b_2 = b(\theta_1)$. En faisant un développement de Taylor, on trouve immédiatement la valeur de la période :

$$\theta_2 = \theta_1 + \varepsilon \ln(r) = 2 + \varepsilon \ln\left(\frac{r^2}{r-1}\right) \quad (\text{C.14})$$

Bibliographie

- [1] R Albert, H Jeong, and A L Barabási, *Error and attack tolerance of complex networks*, Nature **406** (2000), 378–382.
- [2] A Aulehla, C Wehrle, B Brand-Saberi, R Kemler, A Gossler, B Kanzler, and B G Herrmann, *Wnt3a plays a major role in the segmentation clock controlling somitogenesis*, Dev. Cell **4** (2003), 395–406.
- [3] R Lev Bar-Or, R Maya, L A Segel, U Alon, A J Levine, and M Oren, *Generation of oscillations by the p53-Mdm2 feedback loop: A theoretical and experimental study*, Proc. Nat. Acad. Sci. U.S.A. **97** (2000), 11250–11255.
- [4] A L Barabási and R Albert, *Emergence of scaling in random networks*, Science **286** (1999), 509–512.
- [5] B Bollobás, *Random Graphs*, Academic, Londres, 1985.
- [6] K Chen, A Csikasz-Nagy, B Gyorffy, J Val, B Novák, and J J Tyson, *Kinetic analysis of a molecular model of the budding yeast cell cycle*, Mol. Biol. Cell **11** (1999), 369–391.
- [7] P Cheng, Y Yang, C Heintzen, and Y Liu, *Coiled-coil domain mediated frq-frq interaction is essential for its circadian clock function in Neurospora*, EMBO J. **20** (2001), 101–108.
- [8] P Cheng, Y Yang, and Y Liu, *Interlocked feedback loops contribute to the robustness of the Neurospora circadian clock*, Proc. Nat. Acad. Sci. U.S.A. **98** (2001), 7408–7413.
- [9] J L Cherry and F R Adler, *How to make a biological switch*, J.Theor.Biol. **203** (2000), 117–133.
- [10] J Cooke and E C Zeeman, *A clock and wavefront model for control of the number of repeated structures during animal morphogenesis*, J.Theor.Biol. **58** (1976), 455–476.

- [11] S K Crosthwaite, J J Loros, and J C Dunlap, *Light-induced resetting of a circadian clock is mediated by a rapid increase in frequency transcript*, Cell **81** (1995), 1003–1012.
- [12] Bray D and Lay S, *Computer simulated evolution of a network of cell-signaling molecules*, Biophys. J. **66** (1994), 972–977.
- [13] S Daan and C S Pittendrigh, *A Functional Analysis of Circadian Pacemakers in Nocturnal Rodents*, J.Comp.Physiol. **106** (1976), 291–331.
- [14] D L Denault, J J Loros, and J C Dunlap, *Wc-2 mediates wc-1-frq interaction within the pas protein-linked circadian feedback loop of Neurospora*, EMBO J. **20** (2001), 109–117.
- [15] R Dobrin, Q K Beg, A L Barabasi, and Z N Oltvai, *Aggregation of topological motifs in the escherichia coli transcriptional regulatory network*, BMC Bioinformatics. **5** (2004), 10.
- [16] J Dubrulle, M J McGrew, and O Pourquie, *Fgf signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal hox gene activation*, Cell **106** (2001), 219–232.
- [17] J C Dunlap, *Molecular bases for circadian clocks*, Cell **96** (1998), 271–290.
- [18] M B Elowitz and S Leibler, *A synthetic oscillatory network of transcriptional regulators*, Nature. **403** (2000), 335–338.
- [19] M B Elowitz, A J Levine, E D Siggia, and P S Swain, *Stochastic gene expression in a single cell*, Science **297** (2002), 1183–5.
- [20] M B Elowitz, M G Surette, P E Wolf, J B Stock, and S Leibler, *Protein mobility in the cytoplasm of escherichia coli*, J. Bacteriol. **181** (1999), 197–203.
- [21] P Erdős and A Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci. **5** (1960), 17–61.
- [22] J F Feldman and M N Hoyle, *Isolation of circadian clock mutants of Neurospora crassa*, Genetics. **75** (1973), 605–13.
- [23] J E Ferrell, *Building a cellular switch: more lessons from a good egg*, BioEssays **21** (1999), 866–870.
- [24] D B Forger and C S Peskin, *A detailed predictive model of the mammalian circadian clock*, Proc. Nat. Acad. Sci. U.S.A. **100** (2003), 14806–14811.
- [25] J A Foster, *Evolutionary computation*, Nat. Rev. Genet. **2** (2001), 428–436.
- [26] P François and V Hakim, *Design of genetic networks with specified functions by evolution in silico*, Proc.Nat. Acad.Sci.U.S.A. **101** (2004), 580–585.

- [27] P François, *A model for the neurospora circadian clock*, Biophys. J. **88** (2005), 2369–2383.
- [28] P François and V Hakim, *Core genetic network : The mixed feedback loop*, Phys. Rev. E **72** (2005), 031908.
- [29] A C Froehlich, Y Liu, J J Loros, and J C Dunlap, *White collar-1, a Circadian Blue Light Photoreceptor, Binding to the frequency promoter*, Science. **297** (2002), 815–819.
- [30] N Y Garceau, Y Liu, J J Loros, and J C Dunlap, *Alternative Initiation of Translation and Time-Specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY*, Cell. **89** (1997), 469–476.
- [31] T S Gardner, C R Cantor, and J J Collins, *Construction of a genetic toggle switch in escherichia coli*, Nature **403** (2000), 339–342.
- [32] D T Gillespie, J.Phys.Chem. **81** (1977), 2340–2361.
- [33] F Giudicelli and J Lewis, *The vertebrate segmentation clock*, Curr. Opinion Gen. Dev. **14** (2004), 407–414.
- [34] L Glass and M C Mackey, *From Clocks to Chaos : The Rhythms of Life*, Princeton, NJ : Princeton University Press, 1999.
- [35] D Gonze, J C Leloup, and A Goldbeter, *Theoretical models for circadian rhythms in Neurospora and Drosophila*, C.R. Acad.Sci. III. **323** (2000), 57–67.
- [36] N Guelzim, S Bottani, P Bourguine, and F Képès, *Topological and causal structure of the yeast transcriptional regulatory network*, Nature Genetics **31** (2002), 60–63.
- [37] C C Guet, M B Elowitz, W Hsing, and S Leibler, *Combinatorial synthesis of genetic networks*, Science. **296** (2002), 1466–1470.
- [38] L H Hartwell, J J Hopfield, S Leibler, and A W Murray, *From molecular to modular cell biology*, Nature **402** (1999), C47–C52.
- [39] J Hasty, D McMillen, and J J Collins, *Engineered gene circuits*, Nature. **420** (2002), 224–230.
- [40] N D Hayes, *Roots of the transcendental equations associated with a certain difference-differential equation*, J.Lond.Math.Soc. **25** (1950), 226–232.
- [41] C Heintzen, J J Loros, and J C Dunlap, *The pas protein vivid defines a clock associated feedback loop that represses light input, modulates gating and regulates clock resetting*, Cell. **104** (2001), 453–464.

- [42] T Ito, T Chiba, R Ozawa, M Yoshida, M Hattori, and Y Sakaki, *A comprehensive two-hybrid analysis to explore the yeast protein interactome*, Proc. Nat. Acad. Sci. U.S.A. **98** (2001), 4569–4574.
- [43] J Jaeger, M Blagov, D Kosman, K N Kozlov, Manu, E Myasnikova, S Surkova, C E Vanario-Alonso, M Samsonova, D H Sharp, and J Reinitz, *Dynamical analysis of regulatory interactions in the gap gene system of drosophila melanogaster*, Genetics **167** (2004), 1721–1737.
- [44] H Jeong, B Tombor, R Albert, Z N Oltvai, and A L Barabási, *The large-scale organization of metabolic networks*, Nature **407** (2000), 651–654.
- [45] H Kobayashi, M Kaerns, M Araki, K Chung, T S Gardner, C R Cantor, and J J Collins, *Programmable cells : Interfacing natural and engineered gene networks*, Proc. Nat. Acad. Sci. U.S.A. **101** (2004), 8414–8419.
- [46] G Lahav, N Rosenfeld, A Sigal, N Geva-Zatorsky, A J Levine, M Elowitz, and U Alon, *Dynamics of the p53-Mdm2 feedback loop in individual cells*, Nat. Genetics. **36** (2004), 147–150.
- [47] K Lee, J J Loros, and J C Dunlap, *Interconnected Feedback Loops in the Neurospora Circadian Systems*, Science. **289** (2000), 107–110.
- [48] J C Leloup and A Goldbeter, *A Model for Circadian Rhythms in Drosophila incorporating the formation of a complex between the PER and TIM proteins*, J. Biol.Rhythms. **13** (1998), 70–87.
- [49] ———, *Toward a detailed computational model for the mammalian circadian clock*, Proc. Nat. Acad. Sci. U.S.A. **100** (2003), 7051–56.
- [50] J Lewis, *Autoinhibition with Transcriptional Delay: A simple Mechanism for the Zebrafish Somitogenesis Oscillator*, Curr. Biol. **13** (2003), 1398–1408.
- [51] Y Liu, J J Loros, and J C Dunlap, *Phosphorylation of the Neurospora clock protein FREQUENCY determines its degradation rate and strongly influences the period length of the circadian clock*, Proc. Nat. Acad. Sci. U.S.A. **97** (2000), 234–239.
- [52] Y Liu, M W Merrow, J J Loros, and J C Dunlap, *How temperature changes reset a circadian oscillator*, Science **281** (1998), 825–829.
- [53] J J Loros and J C Dunlap, *Genetic and Molecular Analysis of circadian rhythms in Neurospora*, Annu.Rev.Physiol. **63** (2001), 757–794.
- [54] S Mangan and U Alon, *Structure and function of the feed-forward loop network motif*, Proc. Nat. Acad. Sci. U.S.A. **100** (2003), 11980–5.

- [55] S Maslov and K Sneppen, *Specificity and stability in topology of protein networks*, Science **296** (2002), 910–913.
- [56] I Matic, F Taddei, and M Radman, *Survival versus maintenance of genetic stability: a conflict of priorities during stress*, Research Microbiol. **155** (2004), 337–341.
- [57] T Matsuo, S Yamaguchi, S Mitsui, A Emi, F Shimoda, and H Okamura, *Control Mechanism of the Circadian Clock for Timing of Cell Division in Vivo*, Science. **302** (2003), 255–259.
- [58] A Mazurie, S Bottani, and M Vergassola, *An evolutionary and functional assessment of regulatory network motifs*, Genome Biol. **6** (2005), R35.
- [59] M W Mellow, N Y Garceau, and J C Dunlap, *Dissection of a circadian oscillation into discrete domains*, Proc. Nat. Acad. Sci. U.S.A. **94** (1997), 3877–3882.
- [60] R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, and U Alon, *Network motifs: simple building blocks of complex networks*, Science **298** (2002), 763–4.
- [61] N A Monk, *Oscillatory expression of hes1, p53, and nf-kappab driven by transcriptional time delays*, Curr. Biol. **13** (2003), 1409–1413.
- [62] J Monod and F Jacob, *Teleonomic mechanisms in cellular metabolism, growth, and differentiation*, Cold Spring Harbor Symp. Quant. Biol. **26** (1961), 389–401.
- [63] M Nakajima, K Imai, H Ito, T Nishiwaki, Y Murayama, H Iwasaki, T Oyama, and T Kondo, *Reconstitution of circadian oscillation of cyanobacterial kaic phosphorylation in vitro*, Science **308** (2005), 414–5.
- [64] E M Ozbudak, M Thattai, H N Lim, B I Shraiman, and A van Oudenaarden, *Multistability in the lactose utilization network of escherichia coli*, Nature **427** (2004), 737–740.
- [65] I Palmeirim, D Henrique, D Ish-Horowicz, and O Pourquie, *Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis*, Cell **91** (1997), 639–648.
- [66] C S Pittendrigh, V G Bruce, N S Rosenzweig, and M L Rubin, *A Biological clock in Neurospora*, Nature. **184** (1959), 169–170.
- [67] O Pourquie, *The segmentation clock : Converting embryonic time into spatial pattern*, Science **301** (2003), 328–330.
- [68] S M Reppert and D R Weaver, *Molecular Analysis of Mammalian Circadian Rhythms*, Annu. Rev. Physiol. **63** (2001), 647–76.

- [69] ———, *Coordination of circadian timing in mammals*, Nature. **418** (2002), 935–941.
- [70] P Ruoff and L Rensing, *The Temperature-Compensated Goodwin Model Simulates Many Circadian Clock Properties*, J.Theor.Biol. **179** (1996), 275–285.
- [71] P Ruoff, M Vinsjevik, C Monnerjahn, and L Rensing, *The Goodwin oscillator : On the Importance of Degradation Reactions in the Circadian Clock*, J.Biol.Rhythms **14** (1999), 469–479.
- [72] ———, *The Goodwin model : Simulating the effect of light pulses on the circadian sporulation rhythm of neurospora crassa*, J.theor.Biol. **209** (2001), 29–42.
- [73] L Sánchez and D Thieffry, *Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module*, J.Theor.Biol. **224** (2003), 517–537.
- [74] S S Shen-Orr, R Milo, S Mangan, and U Alon, *Network motifs in the transcriptional regulation network of escherichia coli*, Nature Genetics **31** (2002), 64–68.
- [75] P Smolen, D A Baxter, and J H Byrne, *Modeling Circadian Oscillations with Interlocking Positive and Negative Feedback Loops*, J. Neurosci. **21** (2001), 6644–6656.
- [76] ———, *A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator*, Biophys. J. **83** (2002), 2349–2359.
- [77] C Soulé, *Graphic requirements for multistationnarity*, ComPlexUs **1** (2003), 123–133.
- [78] F Taddei, M Radman, J Maynard-Smith, B Toupance, P H Gouyon, and B Godelle, *Role of mutator alleles in adaptive evolution*, Nature **387** (1997), 700–702.
- [79] D Thieffry and L Sánchez, *Dynamical modelling of pattern formation during embryonic development*, Curr. Opin. Gen. Dev. **13** (2003), 326–330.
- [80] K Turgay, L W Hamoen, G Venema, and D Dubnau, *Biochemical characterization of a molecular switch involving the heat shock protein clpc, which controls the activity of comk, the competence transcription factor of bacillus subtilis*, Genes Dev. **11** (1997), 119–128.
- [81] J J Tyson, *Modelling the cell division cycle : cdc2 and cyclin interactions*, Proc. Nat. Acad. Sci. U.S.A **88** (1991), 7328–7332.
- [82] J J Tyson, K C Chen, and B Novak, *Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell*, Current Opinion in Cell Biology **15** (2003), 221–231.

- [83] J J Tyson, C I Hong, C D Thron, and B Novak, *A simple Model of Circadian Rhythms Based on Dimerization and Proteolysis of PER and TIM*, *Biophys.J.* **77** (1999), 2411–2417.
- [84] P Uetz, L Giot, G Cagney, T A Mansfield, R S Judson, J R Knight, D Lockshon, V Narayan, M Srinivasan, P Pochart, A Qureshi-Emili, Y Li, B Godwin, D Conover, T Kalbfleisch, G Vijayadamodar, M Yang, M Johnston, S Fields, and J M Rothberg, *A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae*, *Nature* **403** (2000), 601–603.
- [85] G von Dassow, E Meir, E M Munro, and G M Odell, *The segment polarity network is a robust developmental module.*, *Nature* **406** (2000), 188–192.
- [86] A T Winfree, *The geometry of biological time*, *Biomathematics*, vol. 8, Springer-Verlag, New York, Heidelberg, Berlin, 1980.
- [87] D M Wolf and A P Arkin, *Motifs, modules and games in bacteria*, *Current Opinion in Microbiology* **6** (2003), 125–134.
- [88] W Xiong and J E Ferrell, *A positive-feedback-based bistable 'memory module' that governs a cell fate decision*, *Nature* **426** (2003), 460–465.
- [89] E Yeger-Lotem and H Margalit, *Detection of regulatory circuits by integrating the cellular networks of protein-protein interactions and transcription regulation*, *Nucleic Acids Res.* **31** (2003), 6053–6061.
- [90] E Yeger-Lotem, S Sattath, N Kashtan, S Itzkovitz, R Milo, R Y Pinter, U Alon, and H Margalit, *Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction*, *Proc. Nat. Acad. Sci. U.S.A.* **101** (2004), 5934–9.
- [91] Y Yokobayashi, R Weiss, and F H Arnold, *Directed evolution of a genetic circuit*, *Proc. Nat. Acad.Sci. U.S.A.* **99** (2002), 16587–16591.
- [92] M W Young, *Big ben rings in a lesson on biological clocks*, *Neuron.* **36** (2002), 1001–5.

ABSTRACT

This thesis presents a series of works all dealing with the structure and the dynamics of genetic networks.

After a first part introducing various recent works in the field, we propose a method to design genetic networks by evolution *in silico*. This procedure aims at designing genetic motifs performing a given function without any constraints on their structure. With the help of the algorithm, we found both new and well-known motifs. We also compare the obtained modules to real ones and propose some possible applications of the algorithm. In a second part, we analyze the properties of one of the modules found with the help of the algorithm which happens to be overrepresented in real genetic networks. The so-called “Mixed Feedback Loop” consists of two proteins A and B. Protein A regulates transcription of gene b, and the two proteins can form an inactive dimer. A detailed mathematical analysis is proposed. We show that this module can act either as a genetic switch, either as an oscillator. A phase diagram and a precise description of the limit-cycle is also derived analytically. Finally, in the third part, we focus on circadian clocks, which are important examples of genetic oscillators. The general properties of these oscillators are recalled and a model for the *Neurospora crassa* circadian clock is proposed. We especially show the importance of a second feedback loop to explain the robustness of the period of the oscillations with the variation of parameters.

KEY-WORDS : Genetic networks – Evolution in silico – Genetic switch
- Genetic oscillators – Modelling – Circadian clocks

RÉSUMÉ

Cette thèse regroupe une série de travaux concernant la structure et la dynamique des réseaux génétiques.

Après une première partie rappelant différents travaux récents dans ce domaine, nous proposons une méthode de conception de réseaux génétiques par évolution *in silico*. Cette procédure vise à concevoir des motifs génétiques réalisant une fonction donnée sans aucune contrainte sur leurs structures *a priori*. L'algorithme a permis de trouver de nombreux motifs originaux et connus. Une comparaison avec les réseaux existants ainsi que diverses extensions possibles du champ d'application de l'algorithme sont proposées. Nous étudions dans une seconde partie les propriétés d'un motif fréquemment trouvé par cet algorithme et sur représenté dans les réseaux naturels. Ce module de rétroaction mixte ("Mixed Feedback Loop", MFL) est constitué de deux protéines A et B, la protéine A régulant la transcription du gène b, les deux protéines formant un dimère inactif. Une analyse mathématique détaillée de ce module est proposée. En particulier, nous montrons que ce module peut se comporter soit comme un commutateur génétique bistable, soit comme un oscillateur. Un diagramme de phase ainsi qu'une description mathématique précise du cycle de l'oscillateur sont dérivés analytiquement. Enfin, dans une troisième partie, nous nous intéressons plus particulièrement à des exemples biologiques d'oscillateurs génétiques : les horloges circadiennes. Les propriétés générales de ces oscillateurs sont rappelées et un modèle pour l'horloge circadienne de *Neurospora crassa*, basé sur le réseau MFL, est proposé. En particulier, nous montrons l'importance d'une seconde boucle de rétroaction pour la robustesse des oscillations face à la variation des paramètres.

MOT-CLÉS : Réseaux génétiques – Evolution in silico – Commutateurs génétiques – Oscillateurs génétiques – Modélisation – Horloges circadiennes