



MASTER 1 - SCIENCES DE LA VIE DE LA SANTÉ

RAPPORT DE STAGE

Étude et modélisation qualitative d'une voie de synchronisation des horloges circadiennes périphériques par le noyau suprachiasmatique

Par Déborah BOYENVAL

Parcours Universitaire

Spécialité : Génétique, Immunité et Développement

Parcours : Biologie, Informatique et Mathématiques

Supervisée par

Messieurs *Gilles Bernot* et *Jean Paul Comet* - I3S
2000 Route des Lucioles
Sophia-Antipolis 06900 Valbonne

Et

Monsieur *Franck Delaunay*- iBV
28 Avenue de Valrose
Parc Valrose 06108 Nice

13 juin 2017

Résumé

Le système circadien est un système physiologique complexe et fortement intégré, assurant la coordination temporelle sur 24h de la physiologie et du comportement des organismes vivants. Le rythme de l'horloge centrale est principalement ajusté par la lumière. L'horloge centrale coordonne de manière tissu-spécifique les rythmes des différentes horloges périphériques. Le rythme de ces oscillateurs périphériques est ajusté *via* des signaux systémiques, relayant l'information temporelle.

Commune à une variété de voies métaboliques, la variation circadienne de température corporelle est considérée par la littérature du domaine comme un synchroniseur universel des horloges périphériques. À l'interface entre la biologie, l'informatique et les mathématiques, des méthodes de modélisation permettent de représenter le comportement d'un système complexe. Face à l'explosion combinatoire des paramètres, l'approche classique de modélisation par équations différentielles est peu adaptée à l'étude du système circadien dans son ensemble. Un nouveau cadre de modélisation hybride permet de représenter qualitativement la dynamique d'un système complexe, tout en considérant des informations chronométriques.

Ce rapport présente une étude de faisabilité de l'approche de modélisation hybride du mécanisme de synchronisation du système circadien. En s'appuyant sur des données biologiques expérimentales, des connaissances biologiques qualitatives et quantitatives ont été rassemblées et synthétisées, pour la première fois, dans un modèle hybride à 6 variables.

Dans le but d'analyser la dynamique d'un modèle hybride, un simulateur générique a été également conçu durant ce stage. À partir du modèle hybride à 6 variables, les résultats de simulation ont mis en évidence des phénomènes oscillatoires robustes, en condition LD 12 : 12 (12h de lumière puis 12 d'obscurité) et également en obscurité constante.

Mots-clés : Système circadien, synchronisation, modélisation hybride, simulateur générique.

Abstract

The circadian system is considered as a complex and highly integrated physiologic system, that daily coordinates the physiology and behaviour of living organisms. The central clock rhythm is mainly tuned by light and in turn, coordinates the rhythm of peripheral clocks in a tissue-specific manner. The rhythm of peripheral clocks themselves is tuned by systemic cues that relay temporal informations. Linked to a wide range of metabolic pathways, the circadian variation of core body temperature is considered as a universal resetting cue for mammalian peripheral oscillators.

At a crossroads between biology, computer science and mathematics, formal modelling allows the dynamic behaviour of a complex system to be represented. Facing the combinatorial explosion of parameters, the classical modelling approach based on differential equations is poorly suitable for the circadian system as a whole. A new hybrid modelling framework allows one to model the qualitative dynamic of such a complex system, while taking into account chronometrical informations.

This report provides a feasibility assessment on the hybrid approach, to modelize the circadian system synchronization. Based on experimental data, qualitative and quantitative biological knowledges were synthesized for the first time into a six-variables hybrid model.

We also designed a generic simulator in order to analyse the dynamic of any hybrid model. Simulation results from the six-variable model show robust oscillatory behaviours, both in LD 12 : 12 condition (12h of light then 12h of dark) and in constant dark condition.

Key-words : Circadian system, synchronization, hybrid modelling, generic simulator.

Sommaire

I	Introduction	1
1	La Biologie des Rythmes Circadiens	2
1.1	Le Système Circadien	2
1.2	Les Mécanismes Moléculaires de l'Horloge Circadienne des Mammifères	3
1.3	La Synchronisation du Système Circadien	5
2	Problématique et Objectifs du Stage	9
II	Matériels et Méthodes	10
1	La Démarche de Modélisation	11
2	Les Paramètres de Mesure d'un Rythme Biologique	12
3	Définition d'un Cadre Hybride de Modélisation de Réseaux de Régulation Biologique	13
3.1	Le Formalisme de René Thomas	13
3.2	Réseau Biologique Hybride	15
3.3	États Hybrides et Ressources	17
3.4	Construction des Trajectoires Hybrides de Réseaux de Régulations Génétiques	18
4	La Logique de Hoare Hybride Appliquée aux Contraintes de Phases	23
5	Matériels de Simulation de la Dynamique des Réseaux de Régulation Génétique	25
5.1	Le logiciel HyMBioNet	25
5.2	Construction d'un Simulateur de Dynamique en Python3	26
5.2.1	Conception du Projet de Simulateur	26
5.2.2	Fonctionnalités du Simulateur	26
III	Résultats et Discussions	27
6	Résultats	28
6.1	Un Modèle à 6 Variables de la Synchronisation du Système Circadien	28
6.2	Détermination des Données Biologiques Brutes de Référence	30
6.3	Deux Approches de Paramétrisation du Modèle	31
6.4	La Comparaison entre le Simulateur Prototype et HyMBioNet révèle un Subtile Décalage de Phase	33
6.5	Expérience in Silico : Condition Synchronisée par la Lumière versus Non Synchronisée	35
7	Discussions	38
7.1	Conséquences Méthodologiques de l'Abstraction des Connaissances dans un Modèle Hybride	38
7.2	Critère de Fiabilité des Données Biologiques Expérimentales	39

7.3	Test dos à dos du Simulateur HyMBioNet et du Simulateur Prototype	39
-----	---	----

IV	Conclusion et Perspectives	41
-----------	-----------------------------------	-----------

V	Annexes	45
----------	----------------	-----------

1.1	Logique Formelle et Tables de Vérité.	I
1.2	Données Expérimentales de Température Corporelle.	II
1.3	Graphe d'États de l'Horloge Circadienne Centrale et Périphérique Hépatique.	IV
1.4	Bilan des Compétences Acquises.	V

Liste des symboles

bHLH basic-Helix-Loop-Helix

BMAL1 Brain and Muscle ARN-t like protein 1

CBP CREB Binding Protein

CCG Clock-Controlled Gene

CLOCK Circadian Locomotor Output Kaput

CREB C-AMP Response Element-binding protein

CRY Cryptochrome

CT Circadian Time

CTL Computational Tree Logic

HAT Histone Acétyltransférase

HDAC Histone Désacétylase

HSF1 Heat Shock Factor 1

HSP Heat Shock Protein

NPAS2 Neuronal Period-ARNt-Sim 2

NSC Noyaux Suprachiasmatiques

PER Period

REV – ERB Reverse Viral Erythroblastis Oncogene Product

ROR Retinoic acid-related Orphan Receptor

ZT Zeitgeber Time

Remerciements

Tout d'abord, je tiens à remercier chaleureusement Gilles Bernot, Jean-Paul Comet et Franck Delaunay, mes trois co-encadrants de stage. Tous m'ont apporté d'innombrables conseils ainsi que de précieux encouragements.

Je remercie également l'équipe de Franck Delaunay au complet, pour leurs grandes aides dans les recherches de données biologiques.

Aussi je remercie particulièrement les deux doctorants Jonathan Behaegel et Emilien Cornillon, pour leurs patiences et leurs explications pédagogiques sur le formalisme hybride et la logique de Hoare hybride.

Enfin, je remercie l'ensemble des personnes que j'ai rencontrées durant ce stage, pour les conversations et expériences de pensées très intéressantes, qui n'auront de cesse de m'inspirer.

« Tout langage éclaire et obscurcit, révèle et cache. Tout langage est à la fois une source de connaissance et une source de questionnement. » Jean Claude Ameisen [1].

Première partie

Introduction

Chapitre 1

La Biologie des Rythmes Circadiens

La rotation de la Terre autour de son axe génère une alternance lumière-obscurité dont la période est de 24h. Tout au long du processus évolutif, les organismes vivants ont développé une horloge biologique interne leur permettant de s'adapter aux changements cycliques de l'environnement extérieur, en orchestrant dans le temps leurs processus physiologiques et comportementaux.

Un rythme est une répétition périodique d'une grande variété de phénomènes. Une horloge génératrice de rythmes biologiques dont la période est de 24h est appelée **horloge circadienne**. Le mot circadien prend racine étymologiquement du latin *circa* : "autour" et *diem* : "un jour".

1.1 Le Système Circadien

Les horloges circadiennes sont présentes dans toutes les cellules d'un organisme vivant donné et sont organisées en un **système circadien** hiérarchique intégré. Au sommet de la hiérarchie, se trouve une *horloge centrale* localisée dans les noyaux suprachiasmatiques (*NSC*) au niveau ventral de l'hypothalamus.

L'alternance lumière-obscurité (ou le cycle jour-nuit) est le signal de synchronisation environnemental le plus puissant de l'horloge centrale (**figure 1.1**). Ce signal a pour fonction d'ajuster la phase d'une horloge biologique par rapport à celle du synchroniseur environnemental. La période de l'horloge centrale est ainsi ajustée à exactement 24h. Ce mécanisme de forçage périodique est réalisé par l'intermédiaire du synchroniseur environnemental, appelé *zeitgeber* pour « donneur de temps » en allemand. La visualisation de la voie de régulation des NSC par la lumière, est disponible en **annexe II**.

Chez les mammifères, les horloges périphériques localisées dans de divers tissus périphériques, tels que le foie, les poumons ou encore les reins, sont soumises à des signaux de synchronisation émis directement ou indirectement par les noyaux suprachiasmatiques. Une caractéristique fondamentale de l'horloge circadienne est son caractère *endogène*. En effet, en absence de signaux environnementaux, les rythmes biologiques sont générés de manière résiduelle dans l'horloge centrale, mais également dans les horloges périphériques avec une période *approximative* de 24h. Les horloges périphériques sont qualifiées d'**autonomes** car elles ne dépendent pas de l'horloge centrale pour générer leurs propres rythmes biologiques circadiens (**Yoo et al.2004 [29]**).

La synchronisation du système circadien influence de nombreux processus biologiques chez les mammifères tels que le cycle veille-sommeil, la variation de température corporelle le métabolisme ou encore l'activité cardiovasculaire.

L'ensemble fonctionnel de ce système circadien permet la coordination et le fonctionnement optimisé de la physiologie et du comportement d'un organisme vivant par rapport à son environnement.

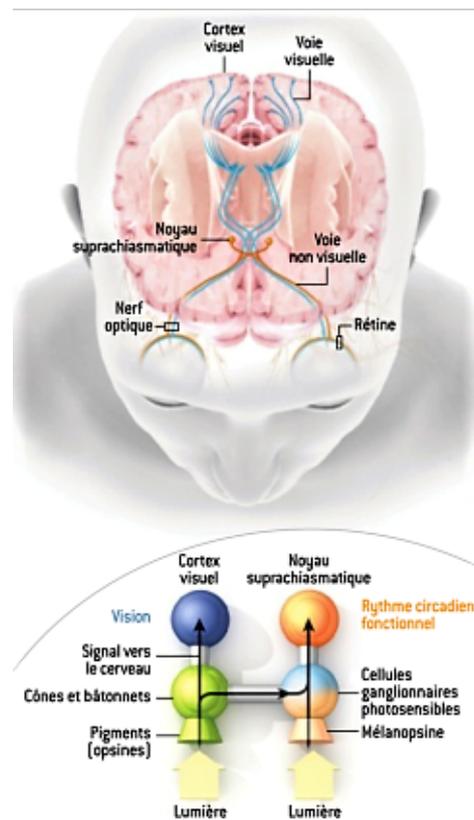


FIGURE 1.1 – Localisation des NSC - Voie visuelle et non visuelle.

La lumière est captée au niveau de la rétine par des cellules photosensibles appelées cellules ganglionnaires. Le signal lumineux est relayé par des photopigments (melanopsines) localisés à la surface des cellules ganglionnaires. L'information lumineuse est ensuite véhiculée le long du tractus rétino-hypothalamique (**en orange**) véhiculant lui-même les axones des cellules ganglionnaires. La voie empruntée par l'information lumineuse dans ce cas est qualifiée de *non visuelle* et aboutit aux deux noyaux suprachiasmatiques (*NSC*).

L'information lumineuse empruntant la voie visuelle est, quant à elle, véhiculée par les nerfs optiques (**en bleu**) vers le chiasma optique où les deux nerfs optiques se rassemblent pour aboutir au niveau du cortex visuel. Les voies visuelles et non visuelles sont donc à considérer indépendamment

Source : Ignacio Provencio. *Pour la Science*, 412, février 2012.

1.2 Les Mécanismes Moléculaires de l'Horloge Circadienne des Mammifères

La présence des horloges circadiennes dans une très grande variété d'organismes, allant des procaryotes photosynthétiques aux mammifères, a questionné **Seymour Benzer** sur la possible existence d'un oscillateur circadien intrinsèque déterminé génétiquement.

Les horloges circadiennes sont présentes dans pratiquement toutes les cellules d'un organisme et sont capables de générer des oscillations¹ moléculaires persistantes.

Le mécanisme moléculaire de l'horloge circadienne se base principalement sur l'interconnexion de 2 boucles de rétroaction négatives impliquant des gènes qualifiés de *gènes de l'horloge* (**figure 1.2**).

La première boucle, décrite historiquement, implique en premier lieu les gènes de l'horloge **Clock** et **Bmal1**. Ces gènes sont ensuite transcrits puis traduits en protéines *CLOCK* et *BMAL1* qui s'associent en hétérodimères *via* le domaine PAS (*Per Arnt Single-minded*). Le complexe hétérodimérique se lie à une séquence d'ADN appelée **E-box** (*Enhancer Box*), *via* le domaine *bHLH-PAS*, dans la

1. Une oscillation est définie comme étant une variation plus ou moins régulière autour d'une position d'équilibre.

région promotrice des gènes de l'horloge **Period** (Per 1 et Per 2) et **Cryptochrome** (Cry1 et Cry2). Cet événement entraîne une augmentation du taux de synthèse des protéines *PER* et *CRY* dans le cytoplasme. Les protéines PER et CRY s'associent ensuite en hétérodimères puis avec d'autres protéines, les Casein Kinases 1 ϵ et δ (**CK1 ϵ/δ**), pour former un complexe plus important avant d'être transloquées. Le rôle des CK1 ϵ/δ est de phosphoryler les protéines CLOCK et BMAL1 pour permettre leur translocation dans le noyau.

À l'intérieur du noyau, le complexe PER-CRY inhibe la transcription de ses propres gènes en empêchant la fixation de l'hétérodimère CLOCK-BMAL1 sur la séquence d'ADN E-box.

Dans le cytoplasme, l'hétérodimère PER-CRY est dégradé par le protéasome. Les protéines β **TcRP** et **FBLX3** polyubiquitinent respectivement PERs et CRYs afin de marquer ces protéines pour la dégradation protéasomale.

Ce phénomène de dégradation permet de freiner l'inhibition de l'hétérodimère CLOCK-BMAL1 par le complexe PER-CRY. Il est essentiel pour initier un nouveau cycle de transcription des gènes Per et Cry par le complexe CLOCK-BMAL1. Le taux de dégradation des protéines PER et CRY est une étape clé pour la mise en place de la période circadienne endogène de l'horloge moléculaire.

De manière additionnelle, vient s'interconnecter une deuxième boucle de rétroaction négative impliquant de nouveau l'hétérodimère CLOCK-BMAL1. Ce complexe protéique active la transcription des gènes **Ror $\alpha/\beta/\gamma$** et **Rev-erb α/β** codant pour des protéines de la famille des récepteurs nucléaires. Les protéines ROR $\alpha/\beta/\gamma$ et REV-ERB α/β entrent en compétition pour se fixer sur un élément de réponse RORE (*Retinoid acid-related Orphan Receptor response Elements*) situé en amont du promoteur du gène Bmal1.

ROR $\alpha/\beta/\gamma$ **activent** la transcription de Bmal1 tandis que REV-ERB α/β **l'inhibent**.

Ces boucles d'autorégulation transcriptionnelles et post-traductionnelles négatives sont considérées comme des *oscillateurs moléculaires*. Ils génèrent une rythmicité circadienne des gènes sous contrôle de l'horloge appelés **CCGs** pour *Clock Controlled Genes*. En effet, le complexe CLOCK-BMAL1 régule rythmiquement la transcription des CCGs. La régulation des CCGs par les protéines CLOCK et BMAL1 se réalise via le contrôle circadien :

- De la machinerie transcriptionnelle de la cellule dont le recrutement l'ARN polymérase II elle-même.
- Des modifications de chromatine par les protéines HATs (*histones acétyltransférases*) et HDACs (*histones désacétylases*) qui respectivement activent la transcription des gènes cibles en relâchant la chromatine et inhibent la transcription de ces gènes par compaction de la chromatine.

Plus précisément, les protéines CLOCK et BMAL1 se fixent aux séquences régulatrices des CCGs, et par la suite, les protéines **p300** et **CBP** sont recrutées respectivement par CLOCK et BMAL1. p300 et CBP sont des protéines co-activatrices dont la fonction est de relâcher la chromatine pour permettre à l'ARN polymérase II d'accéder aux gènes cibles et l'initiation de la transcription de ces gènes (**Koike et al. 2012** [18]).

Les oscillateurs moléculaires imposent un rythme circadien à une grande variété de fonctions biologiques de natures physiologiques et comportementales telles que : la sécrétion hormonale, le métabolisme glucidique et lipidique, le cycle veille-sommeil ou encore la variation de température corporelle.

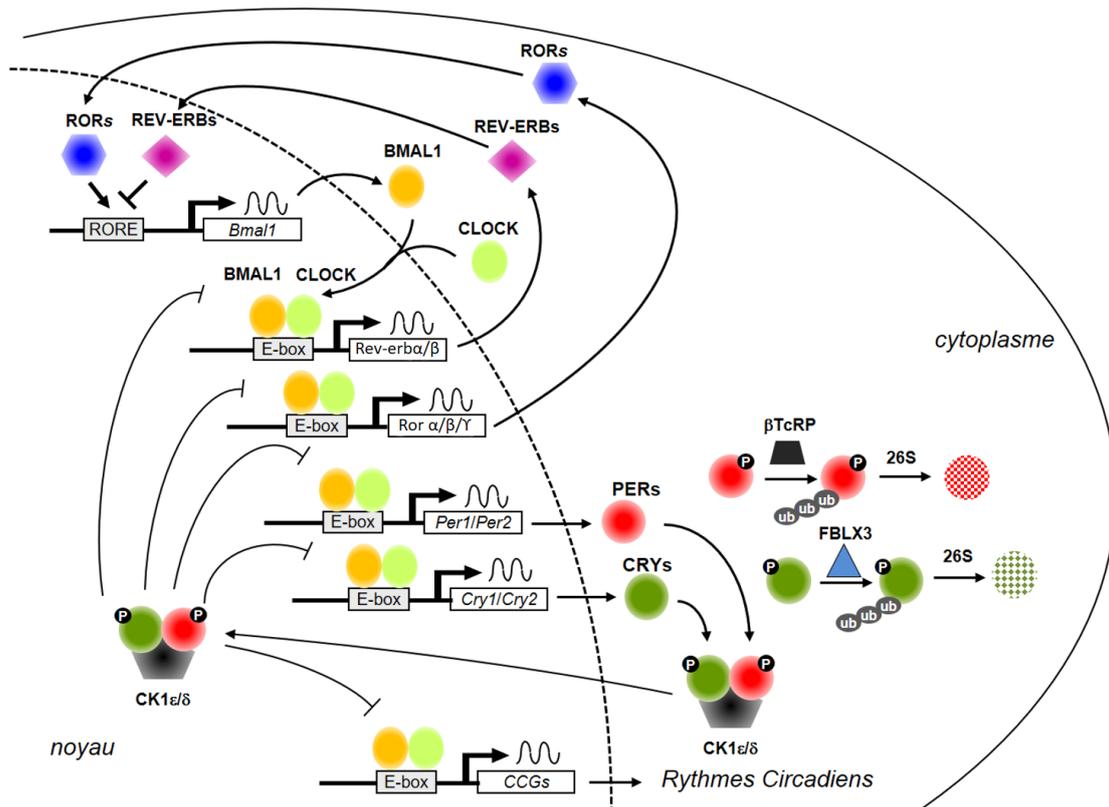


FIGURE 1.2 – Horloge Circadienne Moléculaire - Boucles d’Autorégulations Transcriptionnelles et Post-traductionnelles négatives.

Une boucle de rétroaction négative en biologie, implique par définition une ou plusieurs entités biologiques agissant en retour de manière négative sur la production de cette même entité. La première boucle de rétroaction négative implique les gènes Clock, Bmal1, Per et Cry. La deuxième boucle de rétroaction négative implique de nouveau les gènes Clock et Bmal1, ainsi que les gènes Ror $\alpha/\beta/\gamma$ et Rev-erb α/β . Les protéines PERs et CRYs sont dirigées vers le protéasome suite au marquage par une chaîne de polyubiquitine. Ce marquage sur les protéines PERs et CRYs est réalisé respectivement par la protéine β TcRP et la protéine FBLX3.

1.3 La Synchronisation du Système Circadien

Un *système biologique* est un ensemble d’entités biologiques² organisés en *réseaux d’interactions*. Le système circadien est composé d’une horloge centrale localisée dans les NSC ainsi que d’horloges périphériques localisées dans les cellules de divers tissus périphériques. Le fonctionnement intégré du système circadien nécessite la communication et la coordination des différentes composantes de ce système.

L’alternance lumière/obscurité est le principal zeitgeber de l’horloge centrale. Les NSC sont mis à jour par le signal rétinien et envoient à leur tour des **signaux systémiques** relayant l’information lumineuse temporelle auprès des horloges périphériques.

En 2004, Yoo SH et son équipe mirent en évidence que les horloges circadiennes périphériques³ ne requièrent pas de signaux de l’horloge centrale pour maintenir des oscillations circadiennes persistantes. L’horloge centrale (NSC) jouerait donc le rôle de *coordinateur* des horloges périphériques plutôt que de générateurs de rythmes circadiens en périphérie. (Yoo et al. 2004[29]).

2. Par exemple : gènes, protéines, métabolites, cellules ou encore des organismes.
 3. Tout oscillateur circadien non localisé dans les NSC, est considéré comme périphérique. Les horloges périphériques les plus étudiées chez la souris se localisent dans le foie, les poumons, les reins, la cornée ou encore le pancréas.

L'ajustement des horloges périphériques est relayé par une combinaison tissu-spécifique des signaux de synchronisation *systemiques*. Ce mécanisme de synchronisation est un moyen d'optimiser dans le temps l'expression des gènes de l'horloge selon les fonctions respectives des tissus.

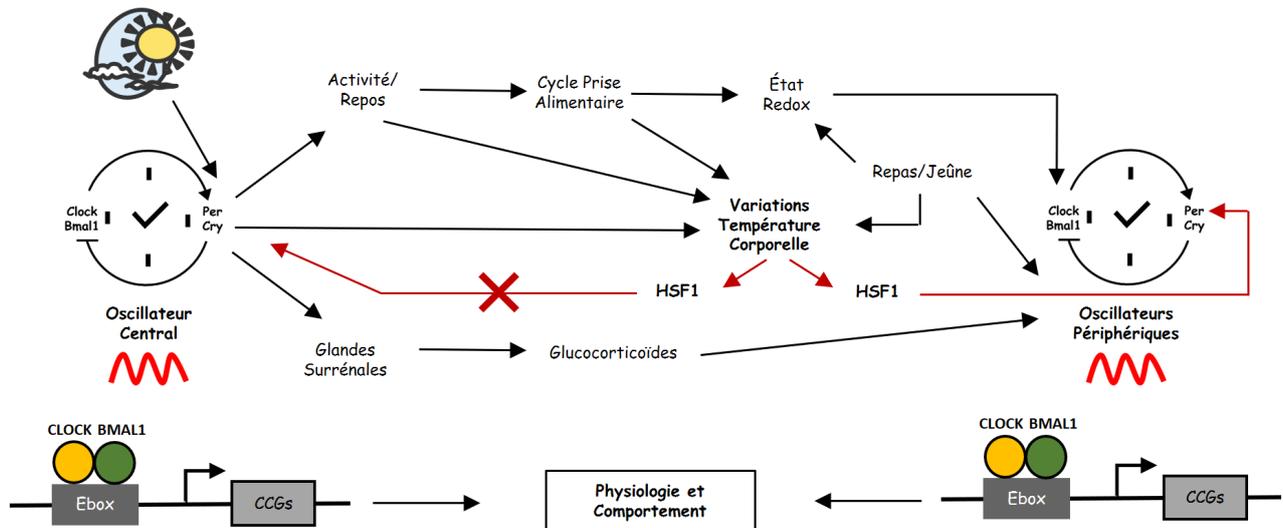


FIGURE 1.3 – Les Différentes Voies de Synchronisation des Horloges Périphériques.

Une multitude de signaux sont impliqués dans l'ajustement des horloges périphériques. Les phases de ces dernières sont entrainées par les NSC via des signaux directs ou indirects. (**Stratmann and Schibler 2006 [26]**). Les signaux humoraux (sang, LCR, sécrétions et muqueuses, etc.) et neuronaux émanant de l'horloge centrale participent à la synchronisation des horloges périphériques.

Parmi la multitude de signaux de synchronisation des horloges périphériques, nous allons recenser, de manière non exhaustive, quelques **signaux de synchronisation** étudiés dans la littérature spécialisée du sujet :

- Les **glucocorticoïdes** d'une part, sont synthétisés de manière circadienne via l'axe hypothalamo-hypophysio-surrénalien. Ils agissent ensuite sur la transcription du gène *Per1* et *Per2* pour finalement ajuster la phase de l'oscillateur périphérique. (**Ishida et al. 2005 [17]**, **So et al. 2009 [25]**).
- Le cycle de **prise alimentaire**, alternant repas/jeûne selon les périodes d'activité et de repos est capable d'entraîner la phase des oscillateurs périphériques. En outre, une restriction alimentaire ou un changement de rythme alimentaire peut provoquer un décalage de phase des horloges périphériques notamment dans le foie et le pancréas (**Damiola et al. 2000 [8]**).
- En 2001, Rutter et al. ont mis en évidence que l'affinité de l'hétérodimère CLOCK-BMAL1 pour le motif E-box (*enhancer box*) dépend de l'état d'oxydoréduction des cofacteurs NAD (*nicotinamide adénine dinucléotide*)⁴. Le NADH et NADPH (*nicotinamide adénine dinucléotide phosphate*) représentent la forme réduite du cofacteur NAD. Ce sont des porteurs d'électrons. Ils favorisent la fixation des hétérodimères CLOCK/BMAL1 sur la séquence d'ADN E-box. Inversement, le NAD⁺ et NADP⁺ représentent la forme oxydée du cofacteur NAD. Ce sont des capteurs d'électrons et interfèrent fortement avec le phénomène de fixation des hétérodimères sur la E-box (**Rutter et al. 2001 [23]**). Les auteurs émettent l'hypothèse que la prise alimentaire serait capable de synchroniser les oscillateurs périphériques par l'intermédiaire d'un changement d'état d'oxydoréduction de la cellule périphérique.

4. Les cofacteurs se comportent comme des transporteurs d'électrons. Ils participent à la catalyse de réactions d'oxydoréduction et de transfert de groupements.

- En 2010 l'équipe de Ethan Burh mit en évidence que la **variation de température corporelle** présente un fort effet de synchronisation de phase des horloges périphériques telles que la glande pituitaire et les poumons (**Buhr et al. 2010** [6]). Les auteurs mirent en évidence chez la souris qu'une variation **subtile** de température corporelle (2.5 ° C) suffit à entrainer la phase des oscillateurs périphériques, notamment dans le foie et les poumons.

Chez l'Homme, la température corporelle varie légèrement au cours de 24h entre 36 ° C et 37.5 ° C, avec la température la plus basse le matin et la plus haute en fin de journée.

Saini et son équipe mentionnèrent dans leurs travaux que le facteur de transcription HSF1 *Heat Shock Factor 1*, exprimé de manière température-dépendante, agirait comme un synchroniseur de phase **universel** des oscillateurs périphériques (**Saini et al. 2012**[24]). Le caractère universel du signal de synchronisation par les NSC via la variation de température émergea suite aux observations suivantes⁵ :

- Plusieurs signaux de synchronisation semblent converger vers la variation de température corporelle (**figure 1.3**). En effet, le métabolisme énergétique, dépendant de la prise alimentaire et du cycle repos/activité d'un organisme, influence la production cyclique de chaleur intra-corporelle et par extension la variation de température corporelle.
- En 2008, un criblage de facteurs de transcription dans le foie révèle que HSF1 présente un pattern d'expression circadien robuste [13]). Dans le foie la liaison de HSF1 sur les motifs HSEs en amont du gène *Per2* suit un rythme circadien).
- Des expériences de perte de fonction du gène HSF1, d'ARN interférence sur les transcrits HSF1 ainsi que de blocage pharmacologique de la protéine HSF1 par KNK437 montrent que des fibroblastes en culture ne sont plus synchronisés efficacement par un signal de synchronisation via la variation cyclique de température (**Saini et al. 2012** [24]).
- Une grande proportion de protéine HSF1 est transloquée dans le noyau de manière circadienne. Chez la souris, la translocation de HSF1 a lieu majoritairement la nuit, ce qui correspond bien à la période d'activité d'un rongeur nocturne (et par extension à une augmentation de température) (**Kornmann et al. 2007** [19]). En supplément, les auteurs montrèrent que l'ajustement de l'expression de *Per2*, par rapport à la variation de température corporelle, est bloqué dans des fibroblastes NIH3T3 HSF1-déficient.

Le **mécanisme de synchronisation** des oscillateurs périphériques *via* la variation de température corporelle, proposé dans la littérature scientifique de ces 7 dernières années, est le suivant :

L'horloge centrale, synchronisée principalement par la photopériode, induit une variation circadienne de température corporelle. En réponse à l'augmentation cyclique de température corporelle la concentration du facteur de transcription immédiat HSF1 augmente (**Reinke et al. 2008** [13]).

Suite à une élévation de température la dénaturation protéique augmente. Les protéines dénaturées entrent alors en compétition avec HSF1 pour la formation d'un hétérodimère avec la protéine HSP90 (*Heat Shock Protein 90*). Par conséquent, une proportion plus importante de HSF1 est libérée du complexe protéique. Les facteurs de transcription HSF1, sous forme libre, s'associent en homotrimères et se fixent sur une séquence promotrice appelée HSE (*Heat Shock Element*). Les HSEs se situent dans le promoteur de gènes cibles tels que *Per2*.

Par la suite, HSF1 se comporte en tant que messenger de phase en communiquant dans le système circadien des informations temporelles des NSC aux tissus périphériques. À l'échelle moléculaire, HSF1⁶ agit comme synchroniseur de phase **thermo-dépendant** des oscillateurs moléculaires périphériques en régulant rythmiquement l'expression du gène de l'horloge *Per2*.

5. Liste non exhaustive.

6. HSF1 est considéré, dans la littérature, comme le synchroniseur de phase thermo-dépendant majoritaire des oscillateurs périphériques. Néanmoins, dans un système aussi intégré que le système circadien, HSF1 opère très probablement en coopération avec d'autres régulateurs thermosensibles.

Enfin, il est important de souligner que le noyau suprachiasmatique est lui même **résistant** au signal de synchronisation via la variation de température corporelle. L'horloge centrale ne subit donc pas le propre signal synchronisation qu'elle génère.

Chapitre 2

Problématique et Objectifs du Stage

Les horloges circadiennes d'un organisme vivant sont soumises à l'influence d'une grande variété de synchroniseurs externes et intrinsèques. Par analogie, ces derniers agissent comme de véritables chefs d'orchestre des rythmes biologiques, en régulant leurs phases et amplitudes.

La problématique de stage s'articule autour d'un axe pluridisciplinaire. À l'interface entre biologie, informatique et mathématiques, *est-il possible de modéliser qualitativement un mécanisme de synchronisation du système circadien des mammifères par la variation de température corporelle ?*

Ce stage propose plusieurs objectifs pluridisciplinaires connexes, que la collaboration entre un laboratoire de biologie (**iBV**) et d'informatique (**I3S**) a permis de mettre en place :

- Rassembler et comprendre les connaissances biologiques sur les différents mécanismes de synchronisation du système circadien, saisir la complexité qu'est capable d'engendrer l'approche intégrée d'un tel système physiologique.
- En parallèle, développer et utiliser une méthode de modélisation qualitative hybride des réseaux de régulations génétiques. Ainsi, l'objectif est d'adapter un modèle hybride de l'*horloge circadienne* centrale des Mammifères, développé par **Cornillon et al [7]**, en un modèle hybride du *système circadien* synchronisé par la variation de température corporelle.
- Concevoir et développer un simulateur générique *fonctionnel*. Ce dernier a pour but de construire automatiquement la dynamique d'un réseau de régulation biologique. Cet objectif prend racine dans le besoin d'amélioration du prototype pré-existant : HyMBioNet.

In fine, l'objectif de fin de stage est d'être en mesure de confronter les résultats de simulation de la dynamique de chaque entité du modèle qualitatif d'une part ; avec leurs comportements oscillants, profils d'expression et d'activité issues d'expériences biologiques d'autre part.

Deuxième partie

Matériels et Méthodes

Chapitre 1

La Démarche de Modélisation

Plusieurs méthodes classiques de modélisation permettent de décrire et analyser le comportement d'un réseau de régulation génétique. L'approche la plus couramment utilisée est la modélisation par équations différentielles. Elle présente l'avantage de représenter la concentration des entités moléculaires par des valeurs variant de manière continue au cours du temps. L'approche stochastique consiste quant à elle, à formaliser l'évolution d'un système en ajoutant une probabilité à chaque transition.

Face à des systèmes biologiques complexes, l'approche différentielle est confrontée au problème d'explosion combinatoire des paramétrisations. L'approche stochastique nécessite, quant à elle, de passer par des méthodes d'approximations des solutions d'équations.

Une approche *logique* de modélisation permettrait de capturer de manière moins fastidieuse le comportement d'un réseau de régulation génétique.

Le processus de modélisation suivi dans le cadre de ce stage est le suivant :

1. En se basant sur une problématique biologique qualitative, il s'agit dans un premier temps d'établir l'inventaire des entités biologiques impliquées à partir des connaissances biologiques sous-jacentes. S'agissant d'une modélisation de réseau de régulation génétique les entités biologiques impliquées peuvent englober des gènes, des transcrits ou des protéines.
2. L'étape suivante consiste à déterminer les interactions entre ces différentes entités afin de construire le squelette du modèle sous forme de réseau de régulation génétique. Un modèle représente donc une abstraction des connaissances qui permet dans ce cas de mieux comprendre un système complexe.
3. Par la suite, la paramétrisation du modèle consiste attribuer des valeurs à chaque paramètre en se basant sur des données biologiques. Ces derniers permettent de construire la dynamique du modèle et de la confronter à des comportements biologiques connus.
4. Enfin, si le modèle n'est pas réfuté à la suite de cette confrontation avec les données biologiques expérimentales, le modélisateur est en mesure de suggérer des expériences *in silico*. Ces expériences nécessitent une réflexion préalable de sorte qu'elles aient le plus de chance de réfuter le modèle étudié. En effet, selon Karl Popper, il n'y a de certitude que négative car la science évolue par conjectures et réfutations [21].

Chapitre 2

Les Paramètres de Mesure d'un Rythme Biologique

Les mesures cinétiques de variables biologiques telles que des transcrits, protéines ou métabolites dans le cadre d'une étude des rythmes circadiens des mammifères, nécessite d'employer un vocabulaire descriptif précis. C'est pourquoi, en premier lieu, sera présenté une définition des différents paramètres utilisés pour décrire un **rythme** (**Figure 2.1**).

Par définition, un rythme est une variation répétitive d'un phénomène donné [12]. Il est décrit et quantifié par plusieurs paramètres :

- Une **période** : intervalle de temps séparant deux aspects identiques du phénomène. L'inverse de la période est la fréquence.
- Une **amplitude** : demi-distance verticale séparant les niveaux du creux et de la crête suivante (ou inversement). Une crête, ou *acrophase*, correspond à la valeur maximale prise par la variable dans l'échelle de temps considérée. *A contrario*, un creux, ou *batyphase*, correspond à la valeur minimale.
- Une **phase** : elle permet de situer la position de la variable dans une oscillation par rapport à une origine de référence. Le déplacement d'un rythme dans le temps est nommé *décalage de phase*.

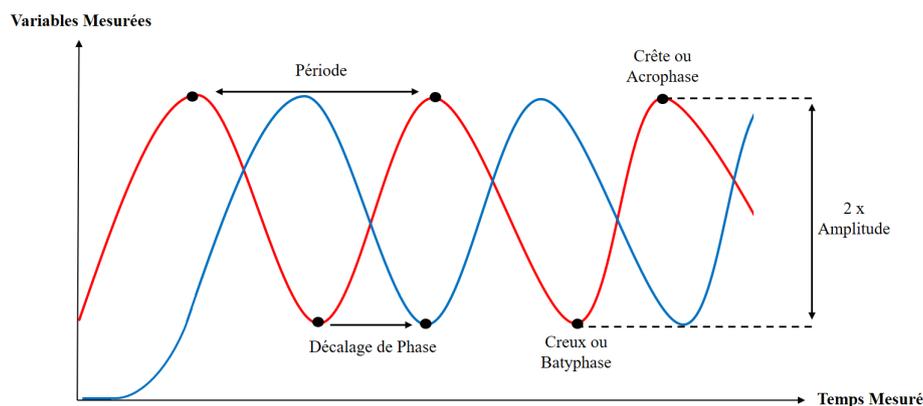


FIGURE 2.1 – Les Différents Paramètres d'un Rythme.

Les rythmes circadiens peuvent être ajustés ou synchronisés par des signaux environnementaux (*zeitgebers*). Des paramètres tels que la période, l'amplitude et la phase, permettent de refléter le niveau d'ajustement ou de synchronisation de la phase d'une entité par rapport à une autre.

Chapitre 3

Définition d'un Cadre Hybride de Modélisation de Réseaux de Régulation Biologique

3.1 Le Formalisme de René Thomas

La biologie systémique est une discipline s'intéressant à l'étude des systèmes biologiques **complexes** multi-échelles et dynamiques. Le New England Systems Institute (NECSI) propose une définition de la **complexité** d'un système : « *Un système complexe est un système composé de multiples entités, dont le comportement ne peut pas être déduit uniquement à partir du comportement de ses entités* ». ¹. Cette définition permet de placer la dynamique d'un système au coeur de la complexité, notamment si l'on considère que la dynamique des interactions entre entités biologiques participe à la formation d'une cohérence dans la réalisation d'une fonction commune. Par exemple, le système circadien est formé par plusieurs réseaux d'interactions ² et présente une dynamique qui génère des rythmes circadiens coordonnés temporellement entre plusieurs organes.

En termes de structures de données, un réseau d'interaction biologique peut être représenté sous la forme d'un **graphe d'interaction orienté**, où les sommets représentent les entités biologiques (ou une abstraction de plusieurs entités) et les arcs orientés représentent leurs interactions. Dans le cadre de ce mémoire de stage, nous nous focaliserons sur des interactions de type **régulation génique**. Lorsqu'un gène est exprimé, il produit, suite à des mécanismes de transcription et de traduction, une protéine. Cette dernière peut ensuite réguler positivement (activation) ou négativement (inhibition) sa propre expression ou l'expression d'autres gènes.

La dynamique des réseaux d'interactions peut être modélisée par un système d'équations différentielles ordinaires décrivant l'évolution des concentrations des entités biologiques impliquées. Selon les paramètres choisis pour décrire l'évolution de ces concentrations, l'analyse formelle de l'ensemble des modèles possibles met en exergue une explosion combinatoire des comportements d'un système donné. La résolution de telles équations différentielles devient rapidement coûteuse en terme de mémoire et de temps d'exécution lorsque la taille du réseau d'interaction augmente.

En 1973, le chercheur belge **René Thomas** introduit la notion de **modélisation discrète** des réseaux de régulation biologique. L'idée intuitive est la suivante : la concentration des entités biologiques impliquées dans le réseau de régulation n'est plus quantifiable dans un intervalle de concentration continu et homogène. Il est découpé en un nombre fini d'intervalles (**René Thomas, 1973 [27]**). Ces intervalles de concentrations sont qualifiés de **niveaux discrets**. Ainsi, à chaque entité est affecté un

1. <http://www.necsi.org> : une définition de la complexité adaptée à la discipline d'étude des systèmes biologiques. Site sûr et fiable.

2. Dans le cadre de cet exemple, un réseau d'interaction met en relation les différents composants de l'horloge moléculaire, définie dans la **partie 1.2 du chapitre 1**.

nombre fini de niveaux discret. Pour une entité et un nombre n de cible de régulation, on décompte $n+1$ intervalles de concentration, ou **états qualitatifs**.

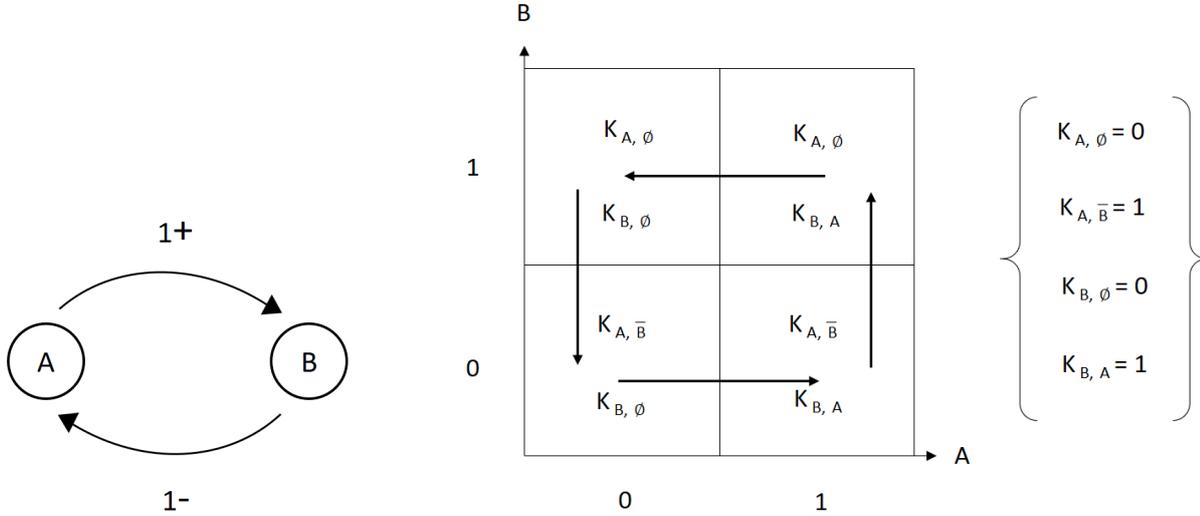


FIGURE 3.1 – **Graphe d’interactions avec seuils (à gauche) - Graphe d’état asynchrone et exemple de dynamique qualitative du réseau d’interaction entre A et B (à droite).**

Ce graphe d’interaction implique deux entités A et B. On affecte pour A et B, 2 niveaux de concentration selon l’intervalle $[[0,1]]$. Ainsi, A agit positivement sur B au seuil 1 et B agit négativement sur A au seuil 1. Les paramètres de chaque variable sont placés dans ces 4 états. Ils déterminent quelles ressources régulent la variable dans un état donné. Les valeurs des paramètres discrets sont déterminées à partir de connaissances biologiques. Considérons un exemple de paramétrisation dans cette figure. Les valeurs déterminent vers quel état la variable est attirée. Parce que la probabilité que deux variables passent leurs seuils en même temps est proche de 0, on décompose le changement de niveau discret de chaque variable. C’est pourquoi on qualifie ce graphe d’*asynchrone*.

On considère qu’une entité régulatrice est effective lorsque son niveau discret est supérieur ou égal au seuil d’action sur sa cible. Ainsi, on dira que le régulateur A agit sur l’entité B au seuil 1 et le régulateur B agit sur l’entité A au seuil 1.

Chaque entité biologique impliquée dans le graphe d’interaction peut être régulée par un ensemble de **ressources**. Il participe à la synthèse de la dite entité.

Selon le graphe d’interaction de la **figure 3.1**, l’ensemble des ressources affecté à l’entité A est $\{\emptyset, \bar{B}\}$ et celui affecté à l’entité B est $\{\emptyset, A\}$ ³.

En termes de ressources, *la présence d’un activateur est équivalente à l’absence d’un inhibiteur*. On considère que l’absence d’inhibition représente un apport positif pour l’entité régulée.

À partir de cet ensemble de ressources est déterminé l’ensemble \mathcal{K} des paramètres discrets de modélisation. Il se note $K_{v,\omega}$ avec

- v : une entité aussi appelée variable,
- ω une ressource de la variable.

Les valeurs des paramètres discrets permettent de construire la dynamique globale d’un réseau de régulation génétique et représentent le niveau discret vers lequel sont attirées les entités biologiques.

À droite de la **figure 3.1**, est représentée la dynamique du réseau d’interaction de gauche. Les transitions d’un niveau discret à un autre, pour chaque variable, sont représentées par des flèches indiquant

3. \bar{B} représente l’absence de l’inhibiteur B.

le changement de valeur de chaque variable séparément (**Bernot et al. 2004** [5]). L'ensemble des transitions forme une succession d'événements discrets et représente la **dynamique qualitative** du modèle.

Soit un état courant pour une variable v noté $\eta(v)$, on retrouve une transition $\eta(v) \rightarrow \eta(v) + 1$ *si et seulement si* :

- Les activateurs d'une variable donnée sont présents car ont passé leurs seuils.
- Les inhibiteurs d'une variable donnée sont absents car n'ont pas passé leurs seuils.
- Et $K_{v,\omega} > \eta(v)$.

Inversement, on retrouve une transition $\eta(v) \rightarrow \eta(v) - 1$ *si et seulement si* :

- Les activateurs d'une variable donnée sont absents car n'ont pas passé leurs seuils.
- Les inhibiteurs d'une variable donnée sont présents car ont passé leurs seuils.
- Et $K_{v,\omega} < \eta(v)$.

Ce formalisme discret de modélisation de réseaux de régulation biologique limite, de part la discrétisation de l'espace de concentration, l'explosion combinatoire des dynamiques possibles pour un graphe donné. À titre illustratif, pour l'exemple considéré dans cette section, il existe tout de même $2^2 \times 2^2$ soit 16 dynamiques possibles pour un réseau de régulation de cette petite taille.

Les vérifications formelles de modèles ou *model checker* permettent de réfuter un ou plusieurs modèles suite à sa confrontation avec des connaissances biologiques décrites sous forme de formules logiques (**Bernot et al. 2004** [5]).

L'avantage indéniable du formalisme de René Thomas, réside dans sa relative simplicité. Il permet de réduire le champ de recherche de paramétrisation, de savoir dans quel état tend le réseau de régulation, et de prédire le comportement qualitatif d'un système biologique pour une paramétrisation donnée.

En revanche, le point limitant de ce cadre discret de modélisation réside dans l'**abstraction du temps** entre les différents événements de régulation des entités biologiques du graphe. Les seuils d'activation ou d'inhibition dans le réseau de régulation génétique ne sont pas forcément franchis à intervalles de temps réguliers. Aussi, les intervalles de concentration ne sont pas de même taille pour chaque entité.

Dans le cadre de l'étude du comportement et de la **synchronisation** du système circadien, il est indispensable de prendre en compte des connaissances temporelles. En effet, le **délai** entre deux événements de régulation joue un rôle majeur dans la détermination d'un comportement donné. Ainsi, une extension du formalisme discret de René Thomas, prenant en compte des informations temporelles pour la construction de la dynamique qualitative, a été proposée par **Cornillon, Comet, Bernot et Enée** en 2016 [7]. Il s'agit d'un cadre **hybride** de modélisation de réseau de régulation biologique.

3.2 Réseau Biologique Hybride

Le formalisme hybride se base sur le formalisme de René Thomas enrichi en multiplexes (**Bernot, Comet and Khalis, 2008** [3]). Dans la section précédente, il est défini qu'un réseau génétique peut être représenté par un graphe orienté. Les entités biologiques, ou variables, sont représentées dans des cercles et les interaction entre entités sont représentées par des arcs labellés avec une valeur seuil.

Ce type de graphe peut être enrichi avec des **multiplexes** représentés dans des rectangles labellés avec le nom du multiplexe.

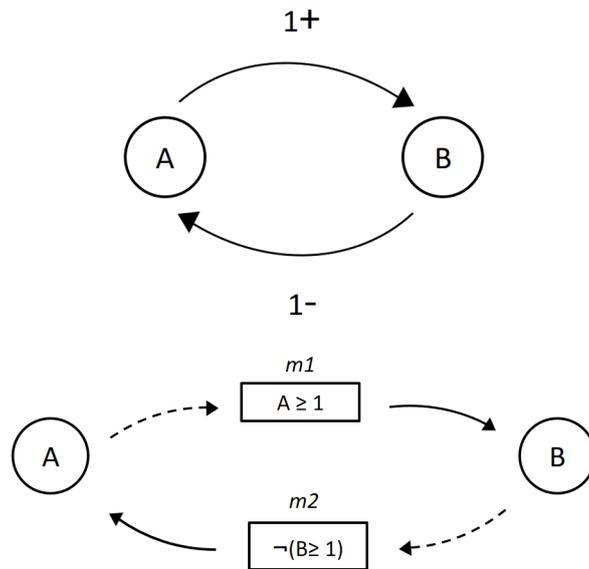


FIGURE 3.2 – **Grphe d’interactions avec seuils (haut) - Grphe d’interactions avec multiplexes (bas).**

Un multiplexe est composé d’une structure rectangulaire à deux parties. À l’intérieur se trouve une formule propositionnelle décrivant une information biologique. La partie haute du multiplexe contient le nom de ce dernier. Le nom est en général choisi de façon à être le plus représentatif de l’information biologique décrite à l’intérieur. Les flèches en pointillées représentent la variable, ou le groupe de variables en entrée, qui influencent ensuite une ou plusieurs variables en sortie (flèches pleines).

Dans cet exemple, la formule associée au multiplexe $m1$ signifie que l’entité A agit comme un activateur de B une fois son seuil 1 atteint. La formule associée au multiplexe $m2$ signifie que l’entité B agit sur A comme le contraire d’un activateur, à savoir un inhibiteur, une fois son seuil 1 atteint.

Les multiplexes contiennent des formules propositionnelles décrivant des phénomènes biologiques tels que :

- la formation de complexes moléculaires régulateurs agissant en coopération,
- la formation de complexes moléculaires régulateurs avec pour effet : une inhibition de l’action régulatrice des entités participant au complexe,
- la phosphorylation d’une protéine,
- la compétition de plusieurs ligands pour la fixation sur une entité réceptrice,
- ou encore la concentration requise d’une entité pour effectuer son action régulatrice.

Dans un état donné, un multiplexe est ressource de la variable si la formule propositionnelle est vraie. Les flèches en pointillées en entrée du multiplexe peuvent être négligées étant donné que les variables dans la formule propositionnelle impliquent forcément ces flèches en entrée.

L’introduction de multiplexes dans un graphe d’interaction, permet d’améliorer la précision du modèle du système biologique étudié dans la mesure où ils raffinent les dynamiques du réseau de régulation et réduisent le champ de paramétrisation possible.

Un réseau **hybride** de régulation génétique, et les notions sous-jacentes, ont été initialement définis par E. Cornillon, JP. Comet, G. Bernot et G. Enée en 2016 [7] :

Définition 1 (Réseau de régulation génétique hybride) *Un réseau de régulation génétique hybride est représenté par un tuple $R = (V, M, E, C)$ où :*

- V est un ensemble d'éléments appelés variables du réseau. À chaque variable $v \in V$ est associé un niveau de concentration discret maximum noté b_v (pour boundary) $\in \mathbb{N}^+$. Relatif à la structure du graphe, b_v représente le degrés sortant de v c'est à dire le nombre d'arcs sortants du sommet v .
- M est un ensemble d'éléments appelés multiplexes. À chaque multiplexe $m \in M$ est associé une formule propositionnelle φ_m appartenant au langage \mathcal{L} . Une proposition est un énoncé dont la valeur de vérité peut être vraie ou fausse. \mathcal{L} est un langage de logique. La logique au sens formel, cherche à trouver des règles générales de raisonnement en s'appuyant sur la forme des propositions et des liens qui peuvent exister entre elles, plutôt que sur leurs contenus. À l'intérieur de cette définition, le langage \mathcal{L} est défini de la sorte :
 - Si $v \in V$ et n , un entier tel que $1 \leq n \leq b_v$, alors $v \geq n$ est un atome de \mathcal{L} . Un atome est une proposition que l'on ne peut pas décomposer.
 - Si φ et ψ sont deux formules propositionnelles de \mathcal{L} , alors $\neg\varphi$, $(\varphi \vee \psi)$, $(\varphi \wedge \psi)$ et $(\varphi \Rightarrow \psi)$ appartiennent aussi à \mathcal{L} .
- E est un ensemble d'arcs de la forme $(m \rightarrow v) \in M \times V$, le produit cartésien de l'ensemble M des multiplexes et l'ensemble V des variables.
- $C = \{C_{v,\omega,n}\}$ est une famille de nombres réels, variant de manière continue, et indexés par le tuple (v,ω,n) où v , ω et n vérifient les trois conditions suivantes :
 1. $v \in V$.
 2. ω est un sous ensemble des prédécesseurs de v . Relatif à la structure d'un graphe, le prédécesseur d'une variable correspond à un arc entrant. Il s'agit de la ressource de la variable.
 3. n est un entier tel que $0 \leq n \leq b_v$.

$C_{v,\omega,n}$ est appelée la célérité de v ayant pour ressource ω au niveau de concentration n .

Les célérités $C_{v,\omega,n}$ fournissent la **direction** et la **vitesse d'évolution** de la concentration discrétisée des entités à l'intérieur de chaque état discret. Il existe une forte connexion entre les paramètres discrets $K_{v,\omega}$ du formalisme de René Thomas et les célérités $C_{v,\omega,n}$ du formalisme hybride :

- Si pour tout n , $C_{v,\omega,n}$ a un signe négatif, alors $K_{v,\omega} = 0$.
- Et si pour tout n , $C_{v,\omega,n}$ a un signe positif, alors $K_{v,\omega} = b_v$.
- Sinon, si il existe n_0 tel que $C_{v,\omega,n_0}=0$, alors $K_{v,\omega} = n_0$.

Les valeurs des paramètres discrets de Thomas nous permettent d'accéder à la dynamique globale du modèle étudié. Les valeurs des célérités, quant à elles, nous permettent d'utiliser des données biologiques temporelles afin de représenter la dynamique locale du modèle.

Cependant, notons que les valeurs des célérités ne sont pas proportionnelles à la vitesse de réaction en biologique. Tous les états discrets sont considérés avec la même taille de 1, alors que les intervalles de concentrations en biologie peuvent être de tailles très différentes.

3.3 États Hybrides et Ressources

Le cadre de modélisation défini dans cette section est qualifié de cadre **hybride** car on considère à l'intérieur de chaque état *discret* de longueur 1, une partie fractionnaire variant de manière *continue* dans l'intervalle $[0, 1]$.

Définition 2 (État hybride d'un réseau de régulation génétique) Soit $R = (V, M, E, C)$ un réseau hybride de régulation génétique. Un état hybride de R est un tuple $h = (\eta, \pi)$ où :

- η est appelé état discret de h . η est une fonction de V telle que pour tout $v \in V$, $0 \leq \eta(v) \leq b_v$. Il définit le niveau discret d'expression de la variable V .
- π est appelé la partie fractionnaire de h . π est une fonction de V et son domaine est l'intervalle de nombres réels $[0, 1]$. Il définit les positions continues à l'intérieur d'un état discret.

On appelle H l'ensemble des états hybrides de R .

Pour une variable v , un état hybride est composé d'un état discret $\eta(v)$ et d'une partie continue $\pi(v)$. Le tuple $(\eta(v), \pi(v))$ représente des coordonnées à l'intérieur d'un état qualitatif. Par ailleurs, la notion de ressource est primordiale car elle définit quelles célérités doivent être considérées dans chaque état discret.

Définissons une **ressource** dans le cadre hybride de modélisation :

Définition 3 (Ressources d'une variable d'un réseau hybride de régulation génétique) Soit $R = (V, M, E, C)$ un réseau hybride de régulation génétique, v une variable $\in V$, $h = (\eta, \pi)$ un état hybride de R et φ une formule de \mathcal{L} .

L'ensemble des ressources d'une variable v pour un état hybride h est défini par :

$$\rho(h, v) = \{m \in R^{-1}(v) \mid h \models \varphi_m\}$$

avec $R^{-1}(v)$ les prédécesseurs de v . Autrement dit, l'ensemble des ressources de la variable v dans un état hybride h ($\rho(h, v)$) est défini par les multiplexes appartenant aux prédécesseurs de v , dont la formule propositionnelle propre à ce multiplexe (φ_m) est satisfaite dans h .

Plus précisément, la relation de satisfaction $h \models \varphi$ (h satisfait φ) est définie par :

- Si φ est un atome $v \geq n$ avec $n \in [1, \dots, b_v]$, alors $h \models \varphi$ si et seulement si $\eta(v) \geq n$. Par exemple dans un multiplexe, si $\varphi_m = v \geq 1$, alors $h \models \varphi_m$ ssi $\eta(v) \geq 1$. Littéralement, si v est supérieur ou égal à 1 alors l'état hybride satisfera φ_m si et seulement si v se trouve dans un état supérieur ou égal à 1.
- Si φ est de la forme $\neg\psi$, alors $h \models \varphi$ ssi $h \not\models \psi$.
- Si φ est de la forme $\psi_1 \vee \psi_2$, alors $h \models \varphi$ ssi $h \models \psi_1$ ou $h \models \psi_2$.
- Si φ est de la forme $\psi_1 \wedge \psi_2$, alors $h \models \varphi$ ssi $h \models \neg(\neg\psi_1 \vee \neg\psi_2)$.
- Si φ est de la forme $\psi_1 \Rightarrow \psi_2$, alors $h \models \varphi$ ssi $h \models (\neg\psi_1 \vee \psi_2)$.

Si nécessaire, une explication des deux dernières lignes est disponible en **annexe 1.1**.

À l'intérieur d'un même état hybride, les célérités C_{v, ω, η_v} avec $\omega = \rho(v, \eta)$, sont les mêmes car les ressources y sont fixées *ad vitam eternum*. Cependant, lorsqu'il y a changement d'état hybride, la variable est régulée par un autre ensemble de ressources et la direction et la vitesse d'évolution du système change dans ce nouvel état.

À présent, définissons la méthode de construction de la dynamique du modèle.

3.4 Construction des Trajectoires Hybrides de Réseaux de Régulations Génétiques

Construire la dynamique d'un réseau de régulation génétique consiste à définir les trajectoires du modèle hybride à la fois dans les états discrets mais également entre deux états discrets.

Les trajectoires hybrides d'un réseau de régulation sont déterminées à partir des valeurs des célérités dans chaque état discret. Ces dernières déterminent à leurs tours, la direction et vitesse d'évolution de chaque variable dans un état donné. Les célérités prennent en compte des informations temporelles à savoir le temps écoulé entre deux événements de régulation.

On introduit ainsi la notion de **touch delay** d'une variable.

Définition 4 (Touch delay) Soit $R = (V, M, E, C)$ un réseau hybride de régulation génétique, v une variable $\in V$ et un état hybride $h = (\eta, \pi)$. On note $\delta_h(v)$ le touch delay mesurant le temps que prend une variable pour atteindre la bordure de l'état discret. Plus précisément, δ_h est une fonction de V variant dans $\mathbb{R}_+ \cup \{+\infty\}$ définie par :

- Si $C_{v,\rho(v,\eta),\eta(v)} = 0$ alors $\delta_h(v) = +\infty$. La variable n'évolue pas et le temps pour atteindre le bord de l'état discret est infini.
- Si $C_{v,\rho(v,\eta),\eta(v)} > 0$ alors $\delta_h(v) = \frac{1-\pi(v)}{C_{v,\rho(v,\eta),\eta(v)}}$.
- Si $C_{v,\rho(v,\eta),\eta(v)} < 0$ alors $\delta_h(v) = \frac{\pi(v)}{|C_{v,\rho(v,\eta),\eta(v)}|}$.

Expliquons cette définition par une illustration :

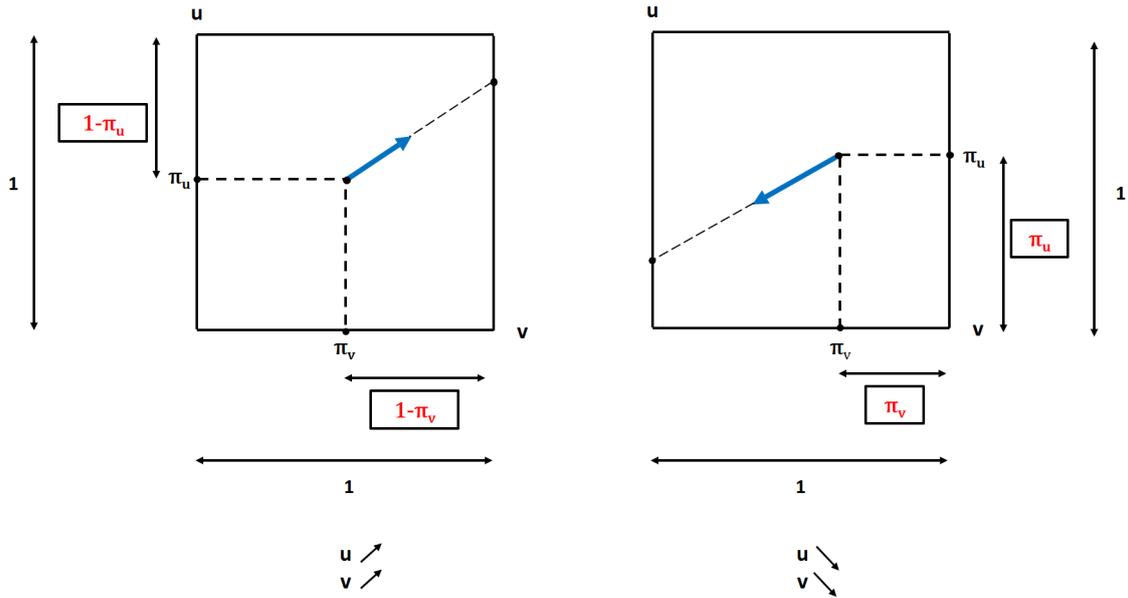


FIGURE 3.3 – Direction et vitesse d'évolution : les paramètres de la dynamique d'un système à l'intérieur d'un état discret.

La position précise d'une variable dans un état discret, est donnée par les coordonnées à l'intérieur du carré correspondant. Pour chaque variable v , on a une coordonnée $\pi(v)$ qui appartient à l'intervalle $[0, 1]$. À partir de ce point, chaque variable va évoluer suivant une vitesse donnée, et va atteindre l'un des deux bords (droit ou gauche) au bout d'un délai noté $\delta_h(v)$.

D'un point de vue mathématique, une trajectoire est un ensemble de positions successives au cours du temps. La direction et la vitesse d'évolution du système dans un état discret donné sont représentées par un vecteur vitesse à 2 composantes si l'on se trouve dans un espace à deux dimensions. Pour une variable v , sa **vitesse d'évolution** $C_{v,\omega,\eta(v)}$ dans un état discret $\eta(v)$ est défini par une variation de distance divisé par une durée.

En effet : $C_{v,\omega,\eta(v)} = \frac{1-\pi(v)}{\delta_h(v)}$ si $C_{v,\omega,\eta(v)} > 0$ (à gauche de la figure 3.3) ; et $C_{v,\omega,\eta(v)} = \frac{\pi(v)}{\delta_h(v)}$ si $C_{v,\omega,\eta(v)} < 0$ (à droite de la figure 3.3).

Une variable ayant atteint le bord de l'état discret n'est pas toujours en mesure d'effectuer une transition vers un nouvel état qualitatif. En effet, il peut ne pas exister d'état discret au delà de l'état courant, notamment dans le cas où le niveau qualitatif maximal (ou minimal) de la variable a été atteint.

On introduit alors la notion de *bords externes* et *internes* et le concept de *mur noir*.

Définition 5 (Bords externes, internes et mur noir) Soit $R = (V, M, E, C)$ un réseau hybride de régulation génétique, v une variable $\in V$ et un état hybride $h = (\eta, \pi)$.

1. Une variable v est en mesure de rencontrer un bord externe si :

- La variable v se trouve dans l'état $\eta_v = 0$ et sa célérité est négative.
- Ou la variable v se trouve dans l'état $\eta_v = b_v$ et sa célérité est positive.

Biologiquement, la variable rencontre un bord externe lorsqu'elle est totalement dégradée ou en concentration maximum. Autrement dit :

$$((C_{v,\rho(v,\eta),\eta(v)} < 0) \wedge (\eta_v = 0)) \vee ((C_{v,\rho(v,\eta),\eta(v)} > 0) \wedge (\eta_v = b_v)).$$

2. Soit $h' = (\eta', \pi')$ un autre état hybride tel que $\eta'_v = \eta_v + \text{sgn}(C_{v,\rho(v,\eta),\eta(v)})$ et $\eta'_u = \eta_u$ pour tout $u \neq v$. η et η' sont deux états voisins. La variable v est en mesure de rencontrer un bord interne si :

La célérité de v dans l'état η varie dans un sens et la célérité de v dans l'état η' varie dans l'autre sens. Autrement dit :

$$\text{sgn}(C_{v,\rho(v,\eta),\eta(v)}) \times \text{sgn}(C_{v,\rho(v,\eta'),\eta'(v)}) = -1$$

On définit de plus la notion de mur noir.

La variable v est considérée sur un mur noir si les deux conditions suivantes sont satisfaites :

1. $\delta_v = 0$: v a atteint son bord.
2. $\text{sgn}(C_{v,\rho(v,\eta),\eta(v)}) \times \text{sgn}(C_{v,\rho(v,\eta'),\eta'(v)}) = -1$: v dans son état η_v ne peut pas franchir le bord pour effectuer sa transition vers η'_v car la célérité de v dans cet état voisin est de signe opposé avec la célérité de v dans l'état voisin.

On note alors slide(v), le glissement de la variable v sur un mur noir ou un bord extérieur. On note slide(h) l'ensemble des variables subissant un glissement.

Les glissements permettent de capturer des informations biologiques supplémentaires et raffinent la dynamique d'un réseau de régulation génétique. Par exemple, le glissement d'une entité ayant une dynamique circadienne peut représenter le moyen qu'emprunte une cellule ou un type cellulaire pour recouvrir une dynamique stabilisée, à la suite d'une perturbation. Les informations biologiques de glissements apportent ainsi une certaine robustesse au modèle.

La probabilité que toutes les variables du modèle franchissent leurs seuils *exactement* en même temps est très faible.

On détermine ainsi l'ensemble $\text{first}(h)$. Pour un réseau de régulation génétique, et un état hybride h donné, il représente l'ensemble des variables changeant de niveau discret qualitatif η_v en **premier**.

Une variable v ne peut faire partie de l'ensemble $\text{first}(h)$ si elle vérifie une de ces deux conditions :

1. Si v se trouve sur un mur noir, alors elle ne peut pas franchir son bord. Elle ne peut pas évoluer tant que les autres variables n'évoluent pas.
2. Si la célérité de v dans l'état courant est nulle, la valeur qualitative de v ne peut pas changer car le touch delay $\delta_h(v)$ est infini.

L'évolution d'un réseau génétique se caractérise par deux types de transitions (illustrées en **figure 3.4**). D'une part, une transition continue **à l'intérieur** d'un état discret. Cette transition détermine quelle variable atteint le bord de l'état en premier et évalue si elle peut franchir ce bord pour changer de valeur qualitative. Et d'autre part, une transition discrète entre deux états discret. La variable v change de valeur qualitative et se trouve ainsi dans un nouvel état discret.

Finalement, on définit $\mathcal{R} = (H, T)$ un **espace d'états hybrides**.

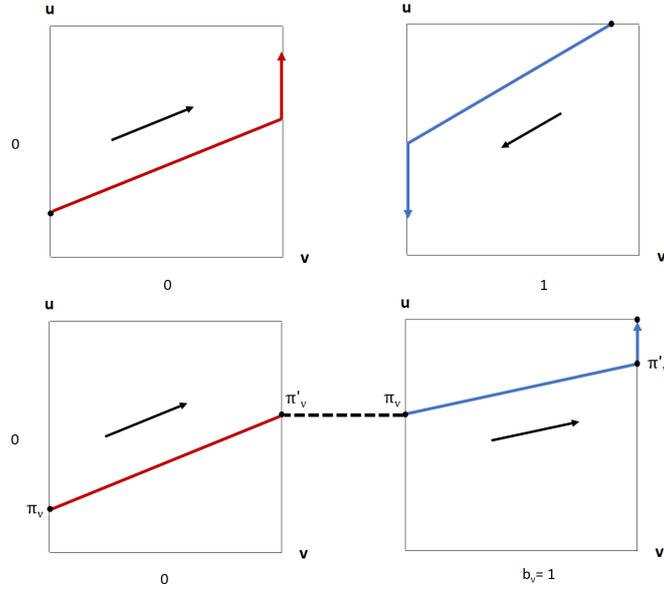


FIGURE 3.4 – Exemple de mur noir et de glissement sur un bord intérieur (haut) - Exemple de transition et de glissement sur un bord extérieur (bas).

Dans la partie supérieure gauche de la figure, la variable v ne peut pas franchir son bord droit car sur le bord gauche de l'état voisin, la célérité de v l'en empêche. On observe un glissement de la trajectoire le long du bord droit. La vitesse et la direction d'évolution de v ne dépend désormais que de la variable u .

Dans la partie inférieure gauche de la figure, la variable v est autorisée à effectuer une transition vers l'état voisin. Dans cet état voisin, v atteint le bord droit mais n'est pas autorisée à le franchir étant donné qu'il n'y a pas d'état voisin. La trajectoire glisse alors le long de ce bord, selon la variable u . $\pi(v)$ (resp. $\pi'(v)$) représente la coordonnée fractionnaire de v en entrée (resp. sortie) dans l'état discret considéré.

Définition 6 (Espace d'états hybrides) Soit $R = (V, M, E, C)$ un réseau hybride de régulation génétique, on nomme $\mathcal{R} = (H, T)$, l'espace d'états hybrides du réseau R où H est l'ensemble des états hybrides et T l'ensemble des transitions.

Il existe une transition d'un état hybride $h = (\eta, \pi)$ à un autre état hybride $h' = (\eta', \pi')$ ssi il existe une variable $v \in \text{first}(h)$ tel que :

1. Soit $\delta_h(v) \neq 0$ (la variable v n'a pas encore atteint son bord) et :

- $\eta' = \eta$: la variable est toujours dans son état courant et ne change pas de niveau qualitatif,
- la coordonnée fractionnaire de v en sortie notée $\pi'(v)$ est égale à 0 si $C_{v,\rho(v,\eta),\eta(v)} < 0$ et 1 si $C_{v,\rho(v,\eta),\eta(v)} > 0$,
- pour toute variable $u \neq v$, si u ne glisse pas sur un bord, alors la coordonnée fractionnaire de u en sortie notée $\pi'(u)$ est égale à :

$$\pi'(u) = \pi(u) + \delta_h(v) \times C_{u,\rho(u,\eta),\eta(u)}$$

$$\text{sachant que } C_{u,\rho(u,\eta),\eta(u)} = \frac{\pi'(u) - \pi(u)}{\delta_h(v)}.$$

Simon $\pi'(u) = \pi(u)$.

2. Ou $\delta_h(v) = 0$ (la variable v a atteint son bord) et :

- $\eta'(v) = \eta(v) + \text{sgn}(C_{v,\rho(v,\eta),\eta(v)})$: la variable v change d'état,
- la coordonnée fractionnaire de v en sortie notée $\pi'(v)$ est égale à 0 si $C_{v,\rho(v,\eta),\eta(v)} < 0$ et 1 si $C_{v,\rho(v,\eta),\eta(v)} > 0$,

- pour toute variable $u \neq v$, $\eta'(v) = \eta'(v)$ (la variable ne change pas de niveau qualitatif) et $\pi'(u) = \pi(u)$.

Ainsi, il est primordial de capturer le plus précisément les connaissances biologiques afin de construire la dynamique d'un modèle hybride à partir d'un jeu de valeurs de célérités.

Les contraintes biologiques temporelles et non-temporelles sur les célérités peuvent être déterminées par des méthodes formelles telles que la logique de Hoare. En utilisant des données biologiques représentées dans ce que l'on nomme un triplet de Hoare, il est possible de déterminer des contraintes sur les valeurs des célérités de sorte que le modèle se comporte selon une suite d'événements biologiques connus de la littérature scientifique.

Chapitre 4

La Logique de Hoare Hybride Appliquée aux Contraintes de Phases

Dans le cadre de modélisation hybride, la déduction des valeurs des paramètres à partir des données biologiques est essentielle pour la détermination du comportement d'un réseau génétique. Néanmoins, elle constitue une étape limitante dans la démarche de modélisation. Cette dernière décennie, une méthode d'identification formelle des paramètres de réseau de régulation génétique a été mise en place en se basant sur la logique de Hoare [4], [2].

Initialement définie par Charles Antony Richard Hoare en 1969 [15], la logique de Hoare permet de démontrer la validité d'un programme informatique. En s'adaptant à un contexte de modélisation hybride de réseaux génétiques, la logique de Hoare permet d'extraire des contraintes sur les paramètres d'un modèle à partir de connaissances biologiques.

Des données biologiques telles que les profils d'expression, d'activité ou encore la quantification d'entités biologiques (gènes, protéines, ARNs, métabolites ou encore des facteurs abiotiques) peuvent être exprimées sous la forme de formules logiques assez similaires à la syntaxe des programmes impératifs en informatique.

La logique de Hoare se base sur la construction d'un *triplet de Hoare* de la forme $\{Pre\} p \{Post\}$, détaillé en **figure 4.1** où :

- *Pre* et *Post* représentent des formules logiques propositionnelles respectivement appelées *pré-condition* et *post-condition*. La pré-condition et la post-condition caractérisent respectivement l'état initial et final du système.
- *p* représente le chemin biologique (*path*). Ce dernier retrace l'ordre et la durée des événements de régulation des entités biologiques étudiées. Plus précisément, il s'agit de déterminer l'ordre de passage des seuils d'activation ou d'inhibition de chaque variable du système et de déduire la durée entre ces différents passages de seuils.

En se basant sur une stratégie *backward*, la méthode consiste à parcourir à rebours le chemin discret de la post-condition à la pré-condition afin de déterminer des contraintes sur les paramètres dans chaque état du système.

C'est en se basant sur la logique de Hoare appliquée aux modèles de réseaux génétiques hybrides, que l'équipe MDSC construit en 2016 un **programme prototype** générant des contraintes sur les paramètres d'un réseau génétique hybride à partir d'un triplet de Hoare hybride en entrée. Un solveur de contrainte est ensuite utilisé afin de générer une ou plusieurs solutions satisfaisant l'ensemble des contraintes générées sur les paramètres d'un modèle hybride. À la sortie du solveur de contrainte, à chaque paramètre est associé une valeur ou un intervalle de valeurs de paramètres.

L'explication détaillée de la méthode d'hybridation de la logique de Hoare est disponible dans l'article de Behaegel et al. [2].

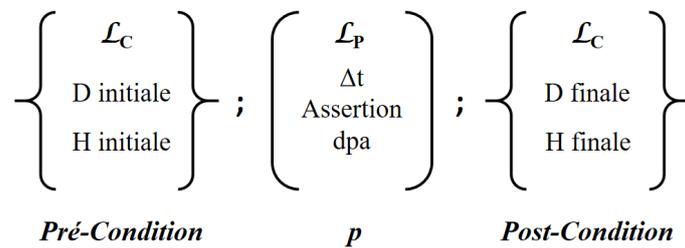


FIGURE 4.1 – **Structure d'un Triplet de Hoare Hybride.**

Un triplet de Hoare hybride est structuré en 3 parties.

La *pré-condition* définit l'état initial du système. Elle se compose de deux conditions initiales. D'une part, la condition discrète D c'est à dire le niveau discret de chaque variable et d'autre part, la condition hybride H c'est à dire la coordonnée fractionnaire de chaque variable dans un état qualitatif donné.

De la même manière que la pré-condition, la *post-condition* décrit l'état final du système par l'intermédiaire de la condition discrète et hybride.

La pré et post-condition définissent le langage de propriété nommé \mathcal{L}_C .

Le chemin p est décrit à l'aide de plusieurs termes. Tout d'abord Δt représente la durée de régulation d'une variable appartenant au système étudié. Ensuite le langage d'*assertion* permet de manipuler des connaissances biologiques telles que la saturation ou la dégradation totale d'une entité biologique.

Enfin l'atome du chemin discret (nommé *dpa* pour *discrete path atom*) représente l'évolution qualitative d'une variable v (par exemple $v+$ ou $v-$).

Chapitre 5

Matériels de Simulation de la Dynamique des Réseaux de Régulation Génétique

5.1 Le logiciel HyMBioNet

HyMBioNet (*Hybrid Modeling of Biological Networks*) est un simulateur de dynamique des réseaux de régulation génétique. Ce simulateur est accessible en ligne [9]. L'utilisation de ce simulateur se déroule en deux temps.

D'une part, une interface web permet de construire un réseau de régulation génétique sous forme de graphe d'interaction enrichi en multiplexes. L'interface web permet à l'utilisateur de fournir en entrée :

- la liste des entités impliquées dans le graphe d'interaction,
- la liste des *zeitgebers* définis dans ce cas comme une entité externe ne subissant l'influence d'aucune entité,
- ainsi que la liste des multiplexes composés d'un nom, d'une formule logique de premier ordre et d'une entité cible.

Ensuite, HyMBioNet génère le code de simulation en langage NetLogo, téléchargeable par l'utilisateur. NetLogo est un langage de programmation utilisé pour la modélisation et la simulation à base d'agents.

L'étape d'initialisation consiste à déterminer la valeurs des différents paramètres de simulation. Tout d'abord, l'ensemble des expressions littérales des célérités est déterminé automatiquement par HyMBioNet. L'interface Netlogo propose à l'utilisateur d'entrer les valeurs des délais plutôt que la valeur des célérités¹. L'utilisateur a ensuite la possibilité de déterminer la quantité initiale de chaque entité du réseau de régulation. Enfin, HyMBioNet propose un langage spécifiquement adapté à la description temporelle du comportement du *zeitgeber*.

L'interface NetLogo propose d'exporter les données numériques de simulation en format CSV afin de les analyser ultérieurement. La description détaillée étape par étape de la prise en main du simulateur ainsi qu'un exemple d'utilisation du simulateur HyMBioNet est disponible dans le chapitre *Hybrid Gene Networks : a new Framework and a Software Environment* [7].

L'intérêt qu'apporte NetLogo est indéniablement sa capacité à modéliser des systèmes complexes évoluant au cours du temps en proposant une interface homme-machine facile et intuitive d'utilisation.

1. La longueur de chaque état qualitatif est égale à 1, ainsi le délai de parcourt d'un état par une variable est égale à l'inverse de la célérité.

5.2 Construction d'un Simulateur de Dynamique en Python3

5.2.1 Conception du Projet de Simulateur

La gestion du pas de temps par HyMBioNet représente son principal inconvénient. Le temps de simulation est segmenté en pas de temps de durée régulière.

Cette gestion du temps par NetLogo est limitante en ce qui concerne l'atteignabilité du bord d'un état qualitatif. En effet d'après la définition du cadre hybride de modélisation, on considère une partie fractionnaire variant de manière *continue* à l'intérieur d'un état qualitatif. En admettant avoir choisi la granularité du temps² la plus fine possible, la probabilité que le système atteigne exactement le bord d'un état qualitatif est très faible.

Ainsi plus le temps de simulation est élevé plus la dynamique du système risque d'être décalée en fonction de l'erreur accumulée à chaque pas. Ce décalage de phase est problématique, dans la mesure où les résultats de simulation permettront par la suite de mesurer les décalages de phases entre deux systèmes oscillants.

Afin d'être en mesure d'obtenir des données de simulation au plus proche de la théorie sous-jacente, un projet de construction d'un nouveau simulateur de dynamique de modèle hybride a été mis en place. De la même manière qu'HyMBioNet, le nouveau simulateur se base sur la définition du cadre hybride de modélisation pour construire la dynamique d'un réseau de régulation génétique. La différence principale réside dans la gestion du délai d'atteinte du bord de l'état qualitatif *courant*. Ce dernier une fois calculé précisément, définit ensuite le délai sur lequel se base le calcul du nouvel état du système. Ce délai est calculé de nouveau lorsque le système change d'état.

Le pas de temps n'est pas homogène selon les états qualitatifs mais représente plus précisément la durée de régulation de chaque variable dans un état qualitatif donné.

5.2.2 Fonctionnalités du Simulateur

Le simulateur de dynamique de réseau de régulation a été codé en Python3. L'objectif du simulateur consiste à construire la dynamique du comportement d'un réseau de régulation génétique. Lors du processus d'initialisation de la simulation, l'utilisateur fournit en entrée le graphe d'interaction enrichi en multiplexe, le niveau discret maximal de chaque entité du graphe d'interaction ainsi que l'état initial du système.

Le simulateur fournit un fichier texte en sortie contenant les données numériques de simulation analysables directement par un outils de représentation graphique de données.

Le code source et la documentation du simulateur construit lors de ce stage, sont disponibles sur le site de l'I3S : <http://www.i3s.unice.fr/~comet/DOCUMENTS/hybridSimulator.zip>.

2. Durée d'un pas de temps la plus petite possible manipulable par des simulateurs à temps discret.

Troisième partie

Résultats et Discussions

Chapitre 6

Résultats

6.1 Un Modèle à 6 Variables de la Synchronisation du Système Circadien

La construction d'un modèle commence par un exercice de représentation synthétique des connaissances biologiques. Ces dernières résident dans la structure même du réseau de régulation.

Le réseau de régulation génétique de la **figure 6.1** est structuré en trois parties, reliées entre elles par des arcs représentant le mécanisme de synchronisation. Chaque partie est composée du même nombre d'éléments reliés par des configurations d'arcs identiques. Cette situation est qualifiée *d'isomorphisme* entre graphes. Chaque sous graphe est une représentation d'une boucle de rétroaction négative génératrice d'oscillations. En effet chaque élément a un effet inhibiteur indirect sur lui même ce qui conduit à une expression oscillante des entités du graphe d'interaction.

La synchronisation du système circadien des mammifères par la variation de température corporelle implique l'interaction entre plusieurs boucles d'autorégulations transcriptionnelles et post-traductionnelles négatives détaillées en **figure 1.2**. Le choix des éléments régulateurs du modèle se base sur l'objectif suivant : décrire à l'aide d'un **modèle simple**, le comportement oscillant des principales entités moléculaires des horloges circadiennes centrales et périphériques.

En 2016, l'équipe MDSC de l'I3S proposa un premier modèle hybride de l'horloge centrale des mammifères [7] en se basant sur la boucle de rétroaction négative composée :

- des transcrits de l'horloges Clock, Bmal1, Per, Cry, RORs, REV-ERBs ainsi que leurs protéines respectives.
- des protéines kinases $CK1\epsilon/\delta$.
- des protéines de marquage protéasomal β TcRP et FBLX3.

L'hétérodimère CLOCK-BMAL1 régulent positivement PER-CRY. Une fois dans le noyau, l'hétérodimère PER-CRY régule négativement CLOCK-BMAL1. *PER-CRY agit donc en retour sur sa propre transcription*. En outre, l'équipe de Choogon Lee déclare que l'abondance de la protéine mPER2 est *limitante* pour l'initiation de la boucle d'autorégulation négative transcriptionnelle et post-traductionnelle [20]. On émet ainsi *l'hypothèse* que l'entité Per2 (transcrit et protéine) est représentative du complexe protéique PER-CRY, formé par PER1, PER2, CRY1 et CRY2.

Par conséquent, nous proposons un modèle à deux variables de l'horloge centrale (en haut à droite de la **figure 6.1**), composé du transcrit mPer2 (**G**) et de la protéine mPER2 (**PC**). La régulation positive des gènes Per-Cry par les protéines CLOCK-BMAL1 et la régulation négative CLOCK-BMAL1 par les protéines PER-CRY sont toutes deux abstraites par transitivité¹.

1. Soit A, B et C. La transitivité désigne le processus par lequel : si $A \Rightarrow B$ et $B \Rightarrow C$ alors $A \Rightarrow C$.

Le système circadien est composé qualitativement de la même horloge moléculaire dans les différents tissus d'un organisme. C'est pourquoi, la structure du modèle de l'horloge périphérique hépatique (en bas à droite de la **figure 6.1**) est identique à celle de l'horloge centrale. Les variables **GF** et **PCF** représentent respectivement le transcrit *mPer2* et la protéine *mPER2* dans le foie.

D'autre part, le cycle jour nuit est représenté par la boucle de rétroaction impliquant la lumière **L** et son inhibiteur **X** représentant conceptuellement l'obscurité. L'action positive de L sur X (décrite par le multiplexe mX) favorise l'action négative de X sur L (décrite par le multiplexe mL). Cette boucle génère le comportement oscillant de la lumière et représente le zeitgeber de l'horloge centrale des mammifères.

Le mécanisme de synchronisation est représenté par deux arcs unilatéraux (de l'horloge centrale vers l'horloge périphérique) décrits par les multiplexes mZL et $mZGF$.

L'alternance lumière-obscurité représente le principal zeitgeber de l'horloge centrale². Le multiplexe mZL décrit l'activation de la transcription du gène *Per2* (**G**) par la lumière (**L**).

Les protéines de l'horloge centrale (en haut à droite de la **figure 6.1**), contrôlent en sortie la variation sur 24h de la température corporelle. Une élévation de la température entraîne une augmentation de la concentration de la protéine HSF1 agissant elle-même sur la transcription du gène *Per2* (**GF**) dans les tissus périphériques. Par transitivité, il est envisageable d'abstraire cette cascade de régulation linéaire en un arc unique $PC \rightarrow GF$ décrit par le mutliplexe $mZGF$.

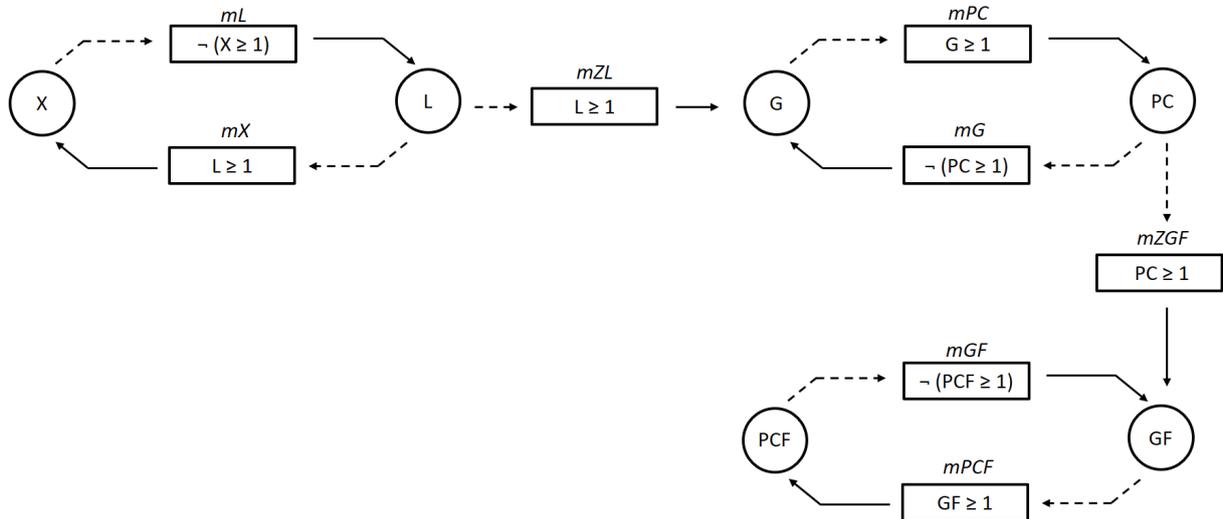


FIGURE 6.1 – Modèle à 6 variables de la synchronisation du système circadien des mammifères.

Le cycle jour-nuit, l'horloge circadienne centrale (NSC) et périphérique (foie) respectivement en haut à gauche, en haut à droite et en bas à droite. Un multiplexe (rectangle) dont la formule est $(A \geq 1)$ et qui a pour cible B, peut être interprété par : *A agit comme un activateur de B au seuil 1*. Le multiplexe régulant B et de formule $\neg(A \geq 1)$ s'interprète par : *A agit comme un inhibiteur de B au seuil 1*.

Le multiplexe mZL représente l'action de la lumière sur l'horloge centrale. Le multiplexe $mZGF$ représente par transitivité, l'action de la variation de température corporelle *via* HSF1 sur l'horloge circadienne périphérique.

2. La lumière active spécifiquement la transcription des gènes *Per*. Un pulse de lumière est à l'origine d'une cascade d'événements. Dans l'ordre : la libération de glutamate, l'activation des récepteurs NMDA, un influx d'ions calcium dans les neurones des NSC, la phosphorylation des protéines CREB et enfin la fixation de ces protéines sur une séquence d'activation transcriptionnelle dans la région promotrice des gènes *Per*

6.2 Détermination des Données Biologiques Brutes de Référence

La paramétrisation du modèle de la **figure 6.1**, nécessite la recherche de données expérimentales quantitatives. Une étude protéomique dans le foie de souris mâles C57B6 par **Wang et al. 2017** [28] fournit à la fois une quantification de GF et PCF (respectivement les transcrits mPer2 et les protéines nucléaires mPER2). L'aboutissement de recherches fastidieuses nous a permis d'avoir accès à la quantification de G (transcrit mPer2) mais également de PC (protéines nucléaires mPER2) dans les NSC de souris mâles C57B6 [11].

La partie (A) de la **figure 6.2** représente l'évolution sur 24h des concentrations des transcrits mPer2 et des protéines nucléaires mPER2, dans les NSC et le foie. Lors de la quantification, les souris furent placées en condition de synchronisation par la lumière selon un cycle de 12h d'obscurité suivi de 12h d'exposition à la lumière artificielle.

Chaque courbe en (A) est accompagnée d'une courbe lissée en pointillé, représentant un ajustement sinusoïdale par la méthode des moindres carrés **Cosinor**³.

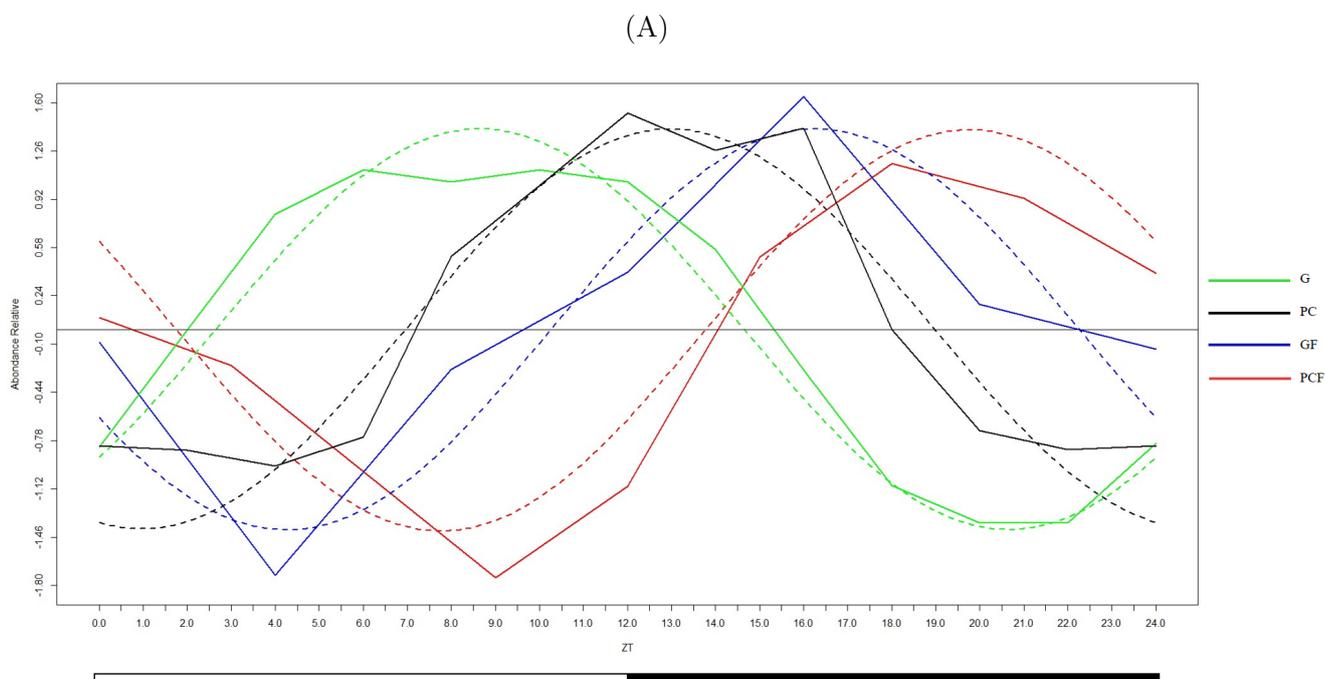
N'ayant pas accès à l'erreur standard des différentes mesures des oscillations, nous proposons en **figure 2.1 partie (B)**, une estimation des différents paramètres (avec erreur standard) des rythmes circadiens par méthode bootstrap, directement sur les données Cosinor.

Les transcrits et protéines issus du gène Per2 dans le foie et dans les NSC, oscillent de manière circadienne (variation de l'amplitude significativement circadienne : p-values < 0.01 évaluées par méthode bootstrap).

D'après les données expérimentales, les éléments de l'horloge G, PC, GF et PCF atteignent leurs acrophases respectives à **ZT 8.6**, **ZT 13**, **ZT 16.2** et **ZT 19.7**.

L'ensemble des résultats de la figure A et du tableau B est corroboré par l'article **Ripperger JA et al.** [22] représentant un important relais bibliographique auquel nous nous sommes référés.

3. Cette méthode de régression non linéaire se base sur l'équation de la fonction cosinus $y = A + B \times \cos(2\pi \times (t - C) / 24)$ où : A représente le niveau moyen de l'entité mesurée, B l'amplitude de son oscillation, C l'acrophase et t la mesure ponctuelle d'un temps (h).



(B)

Variable Mesurée	Paramètres de l'Analyse Cosinor des Données Brutes					
	NSC			Foie		
	Niveau Moyen	Amplitude	Acrophase (h)	Niveau Moyen	Amplitude	Acrophase (h)
mPer2 (ARNm)	290.71 ± 9.97	221.10 ± 14.18	8.65 ± 0.33	6.21 ± 0.26	1.50 ± 0.37	16.26 ± 0.89
mPER2 (Protéine nucléaire)	1.24 ± 0.02	0.25 ± 0.03	12.97 ± 0.43	$-0.28^1 \pm 0.13$	1.65 ± 0.21	19.76 ± 0.40

1 : Valeur non significative p-value>0.05 (p=0.087234).

FIGURE 6.2 – (A) Relation de phases entre transcrits mPer2 et protéines nucléaires mPER2 dans les NSC et le foie de souris mâles C57B6 et ajustement sinusoïdal par méthode cosinor en pointillés - (B) Analyses statistiques des données expérimentales par méthode cosinor et bootstrap.

(A) Profil d'expression sur 24h des transcrits mPer2 dans les NSC (G en vert) et dans le foie (GF en bleu) mesurés respectivement par hybridation *in situ* et RNA-seq. Abondances relatives sur 24h des protéines nucléaires PER2 dans les NSC (PC en noir) et dans le foie (PCF en rouge) mesurées respectivement par immunohistochimie et spectrométrie de masse. Les données expérimentales proviennent toutes de souris placées dans un environnement LD 12 : 12 (barre blanche : 12h de lumière - barre noire : 12h d'obscurité). Dans un soucis d'homogénéité de variances, les données représentées en A ont été centrées et réduites.

(B) Paramètres estimés par méthode Bootstrap. Le niveau moyen et l'amplitude sont quantifiés en unité arbitraire relative. Les valeurs sont toutes estimées avec une p-value <0.01 sauf mention explicite du contraire.

Sources : Les données brutes de la quantification des transcrits mPer2 et des protéines mPER2 des NSC : **Field et al. 2000** [11] Figure 1B; et du foie : **Wang et al. 2017**[28] Table S2.

6.3 Deux Approches de Paramétrisation du Modèle

La dynamique du modèle de la **figure 6.1** est construite à partir des valeurs des 32 célérités listées dans la **figure 6.4**. Le domaine de définition des valeurs de ces paramètres englobent l'ensemble des réels positifs et négatifs. Par conséquent, il existe une infinité de paramétrisations possibles. On

comprend ainsi la nécessité de définir au préalable des contraintes biologiques sur les valeurs de célérités.

La construction des contraintes biologiques inclut une hypothèse de modélisation. Le seuil d'activation de chaque entité du modèle est fixé au niveau de leur concentration *moyenne*. Étant donné que les données expérimentales sont centrées et réduites, le niveau moyen est égal à 0 pour toutes les entités⁴.

$$\left\{ \begin{array}{l} G = 0 \\ PC = 0 \\ GF = 0 \\ PCF = 1 \\ L = 1 \\ X = 0 \end{array} \right\} \left(\begin{array}{c} 0.74 \\ T \\ PCF- \end{array} \right); \left(\begin{array}{c} 1.23 \\ T \\ G+ \end{array} \right); \left(\begin{array}{c} 4.03 \\ Slide + (L) \\ X+ \end{array} \right); \left(\begin{array}{c} 1.13 \\ Slide + (G) \\ PC+ \end{array} \right); \left(\begin{array}{c} 2.47 \\ T \\ GF+ \end{array} \right); \left(\begin{array}{c} 2.4 \\ T \\ L- \end{array} \right); \left(\begin{array}{c} 2.12 \\ Slide + (GF) \\ PCF+ \end{array} \right);$$

$$\left(\begin{array}{c} 1.16 \\ Slide + (PC) \\ G- \end{array} \right); \left(\begin{array}{c} 2.68 \\ T \\ PC- \end{array} \right); \left(\begin{array}{c} 0.05 \\ Slide - (L) \\ X- \end{array} \right); \left(\begin{array}{c} 4.3 \\ Slide + (PCF) \\ GF- \end{array} \right); \left(\begin{array}{c} 1.69 \\ T \\ L+ \end{array} \right) \left\{ \begin{array}{l} G = 0 \\ PC = 0 \\ GF = 0 \\ PCF = 1 \\ L = 1 \\ X = 0 \end{array} \right\}$$

FIGURE 6.3 – **Triplet de Hoare Hybride.**

En partant d'un état initial, on retrouve l'ordre successif et la durée de chaque événement de régulation de la figure 6.2 A. Les *Slides +* représentent la saturation d'une entité.

Le triplet de Hoare proposé en **figure 6.3** capture les connaissances biologiques temporelles issues des données expérimentales de la **figure 6.2**. En partant de l'état initial de gauche, le triplet retrace chaque événement discret de régulation au cours du temps, à partir de l'ordre relatif des franchissements du seuil d'activation ($y = 0$) par les différentes courbes de la **figure 6.2 A**.

L'alpha test du prototype de générateur de contraintes automatiques, à partir du triplet de Hoare, n'a pas été concluant. Les valeurs des célérités fournies en sortie du solveur de contraintes fournissent des données inexploitable (données non présentées). En effet, cet alpha test a mis en évidence une erreur dans la théorie même de la logique de Hoare hybride. De manière synthétique, il s'est avéré que le programme ne générant pas de manière *exhaustive* l'ensemble des contraintes sur célérités. Les connaissances biologiques capturées dans le triplet de Hoare ne permettaient donc pas de contraindre suffisamment la dynamique du modèle, rendant inexploitable l'analyse des résultats.

De ce fait, une nouvelle stratégie de paramétrisation à été mise en place. L'idée de départ fut d'initialiser la paramétrisation du modèle hybride à partir d'un modèle de la synchronisation de l'horloge circadienne **centrale** déjà paramétré [7]. En utilisant le principe du *rasoir d'Ockham*⁵, l'astuce consiste à transposer cette paramétrisation à la fois au modèle de la synchronisation de l'horloge circadienne centrale mais aussi **périphérique**.

L'étape suivante consiste à utiliser le triplet de Hoare afin d'ajuster la paramétrisation précédemment initialisée. Le triplet défini en **figure 6.3** a permis de fractionner l'espace de concentration en 12 états distincts de régulation des entités biologiques. Cette étape permet de déterminer quelles célérités entrent en jeu dans chacun des ces états ainsi que la durée de chaque régulation.

L'étape finale a pour objectif de procéder à des *ajustements précis* de la valeur de chaque célérité dans chacun des 12 états, jusqu'à l'obtention d'une dynamique biologique proche des données expérimentales.

4. le seuil est représenté graphiquement par la droite d'équation $y = 0$ de la figure 6.2 A.

5. Aussi appelé principe de parcimonie, le principe du rasoir d'Ockham peut être résumé par la citation de Guillaume d'Ockham : « Une pluralité ne doit pas être posée sans nécessité ».

Célérités de G	Célérités de PC	Célérités de GF	Célérités de PCF	Célérités de L	Célérités de X
$C_{G,\emptyset,0} = -0.14$	$C_{PC,\emptyset,0} = -0.16$	$C_{GF,\emptyset,0} = -0.14$	$C_{PCF,\emptyset,0} = -0.13$	$C_{L,\emptyset,0} = -0.17$	$C_{X,\emptyset,0} = -0.17$
$C_{G,\emptyset,1} = -0.2$	$C_{PC,\emptyset,1} = -0.14$	$C_{GF,\emptyset,1} = -0.24$	$C_{PCF,\emptyset,1} = -0.14$	$C_{L,\emptyset,1} = -0.17$	$C_{X,\emptyset,1} = -0.17$
$C_{G,mG,0} = 0.18$	$C_{PC,mPC,0} = 0.2$	$C_{GF,mGF,0} = 0.14$	$C_{PCF,mPCF,0} = 0.21$	$C_{L,mL,0} = 0.17$	$C_{X,mX,0} = 0.17$
$C_{G,mG,1} = 0.36$	$C_{PC,mPC,1} = 0.29$	$C_{GF,mGF,1} = 0.48$	$C_{PCF,mPCF,1} = 0.32$	$C_{L,mL,1} = 0.17$	$C_{X,mX,1} = 0.17$
$C_{G,mZL,0} = -0.10$		$C_{GF,mZGF,0} = -0.14$			
$C_{G,mZL,1} = -0.10$		$C_{GF,mZGF,1} = -0.12$			
$C_{G,mG,mZL,0} = 0.22$		$C_{GF,mGF,mZGF,0} = 0.16$			
$C_{G,mG,mZL,1} = 0.36$		$C_{GF,mGF,mZGF,1} = 0.48$			

FIGURE 6.4 – Identification des Valeurs de Célérités.

L'identification des valeurs des célérités, fournit la vitesse d'évolution de l'ensemble des variables du modèle hybride, dans tous les états du systèmes.

6.4 La Comparaison entre le Simulateur Prototype et HyMBioNet révèle un Subtile Décalage de Phase

Le nouveau simulateur et HyMBioNet sont initialisés à partir des mêmes paramètres définis en **figure 6.4** de la section 2.2. À la différence du simulateur prototype, HyMbioNet manipule en entrée des délais calculés comme l'inverse des célérités.

L'intérêt du test dos à dos réside principalement dans la comparaison des deux résultats de simulation. Il ne s'agit pas d'expliciter les caractéristiques et les paramètres de chaque courbe par rapport aux données biologiques expérimentales, mais de comparer les deux dynamiques en sortie en se basant sur un critère comparatif.

La **figure 1.9** montre la dynamique du modèle construit avec HyMBionet en **A** et avec le simulateur prototype en **B**.

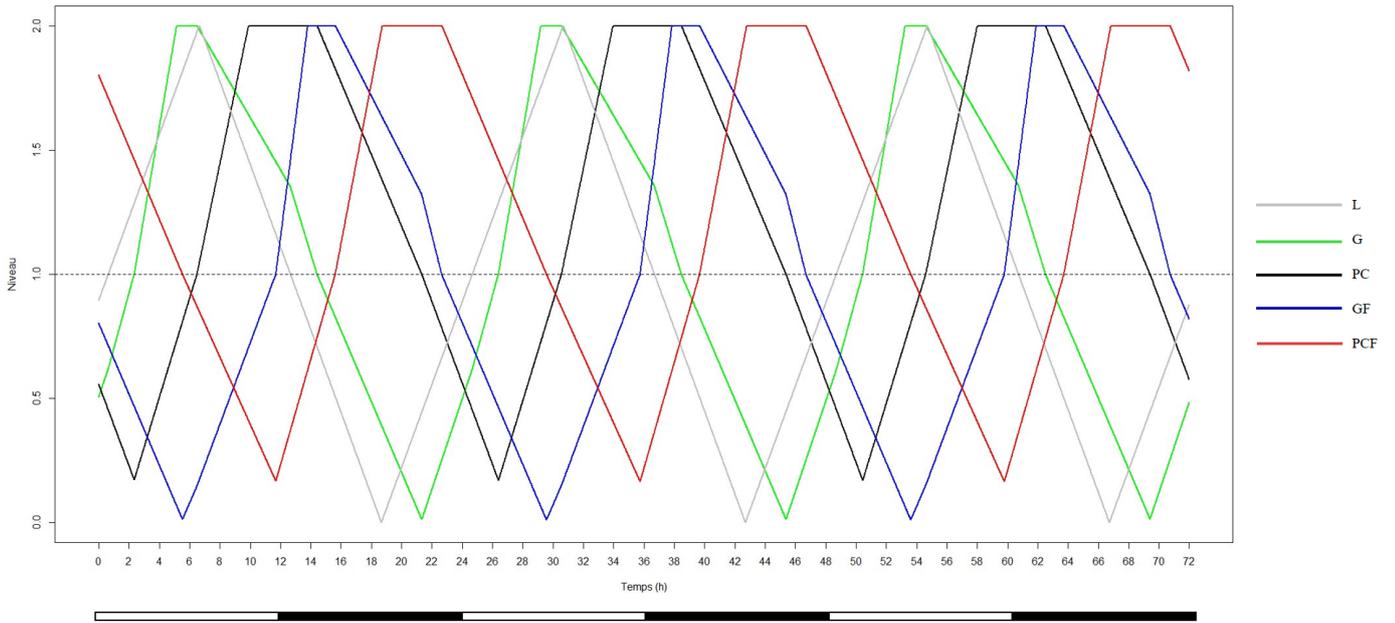
La comparaison des deux simulateurs se base sur l'oscillation déterministe de la lumière. En effet, le cycle jour-nuit a été paramétré de sorte que les durées du jour et de la nuit soient équiréparties sur 24h (LD 12 : 12). Le cycle de lumière est décrit par les variables L et X du modèles hybride. La lumière L est influencée négativement par la variable X et inversement, X est influencé positivement par L. Il existe 4 événements de régulation possibles : $L+$ (augmentation qualitative de L) ou $L-$ (diminution qualitative de L) pour L et $X+$ ou $X-$ pour X. Sur 24h, on fixe chacun des 4 délais de régulation à 6h pour obtenir une oscillation symétrique par rapport au seuil d'équation $y = 1$. Les valeurs de célérités de L et X sont donc égales à $\pm\frac{1}{6}$ selon la nature de la régulation (inhibition ou activation). C'est pourquoi on retrouve les valeurs ± 0.17 ($\frac{1}{6}$ arrondi aux centièmes) pour les célérités de X et L dans la **figure 6.4** de la partie *résultats et discussions*.

Le simulateur prototype en **B** montre bien une oscillation de la lumière symétrique par rapport au seuil d'activation. De 0 à 12h, la lumière est activée (L en gris est au dessus du seuil) alors que de 12h à 24h, la lumière est désactivée (L est en dessous du seuil). Ce simulateur capture des délais de régulation aussi *exacts* que le permet le calcul flottant en python.

HyMBioNet en **A** présente un décalage de phase d'environ 15 minutes, accumulé au cours des 560h de simulation. On retrouve un décalage de phase sur l'ensemble de la dynamique des 4 autres variables (G, GF, PC et PCF) par rapport au simulateur prototype.

Les résultats de simulations des deux simulateurs sont à premières vues très similaires. Cependant, en se basant sur l'oscillation déterministe de la lumière, il s'avère que le simulateur prototype fournit un résultat plus exact par rapport à HyMBioNet présentant un décalage de phase de 15 minutes.

(A) HyMBioNet



(B) Simulateur Prototype

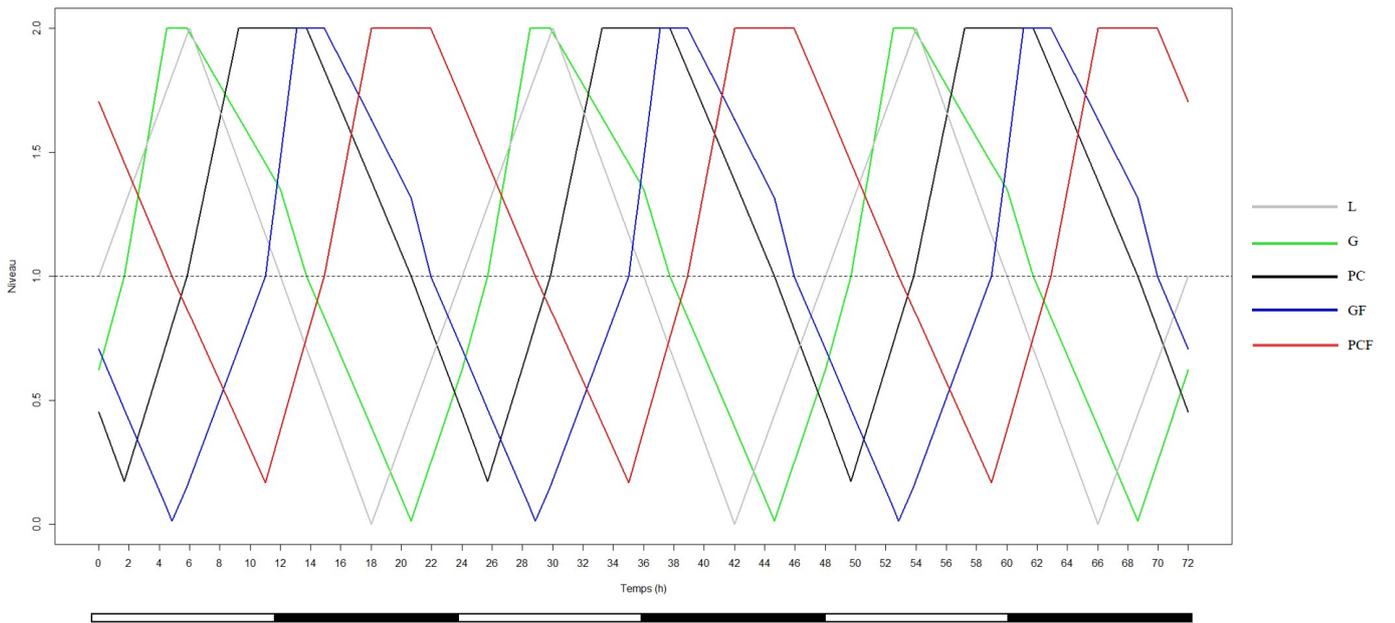


FIGURE 6.5 – **Comparaison des données de simulation par HyMBioNet et par le simulateur prototype.** Relation de phase sur 72h entre G, GF, PC et PCF et la lumière L. Le temps total des 2 simulations est de 560h. Les données représentées proviennent de l'intervalle de temps 480-552 (h). 20j, soit 480h, représentent la durée pour laquelle la période des 4 éléments du système circadien ne varie plus (variations de l'ordre des minutes).

G et PC représentent respectivement le gène mPer2 et la protéine mPER2 dans les NSC. De la même manière, GF et PCF représentent le gène mPer2 et la protéine mPER2 dans le foie.

En ordonnée est représentée la somme du niveau qualitatif et de la partie fractionnaire continue (cf. partie matériels et méthodes). Les barres blanches et noires représentent respectivement 12h de lumière et 12h d'obscurité.

6.5 Expérience *in Silico* : Condition Synchronisée par la Lumière versus Non Synchronisée

L'expérience *in silico* suivante consiste à confronter deux simulations initialisées dans deux conditions différentes.

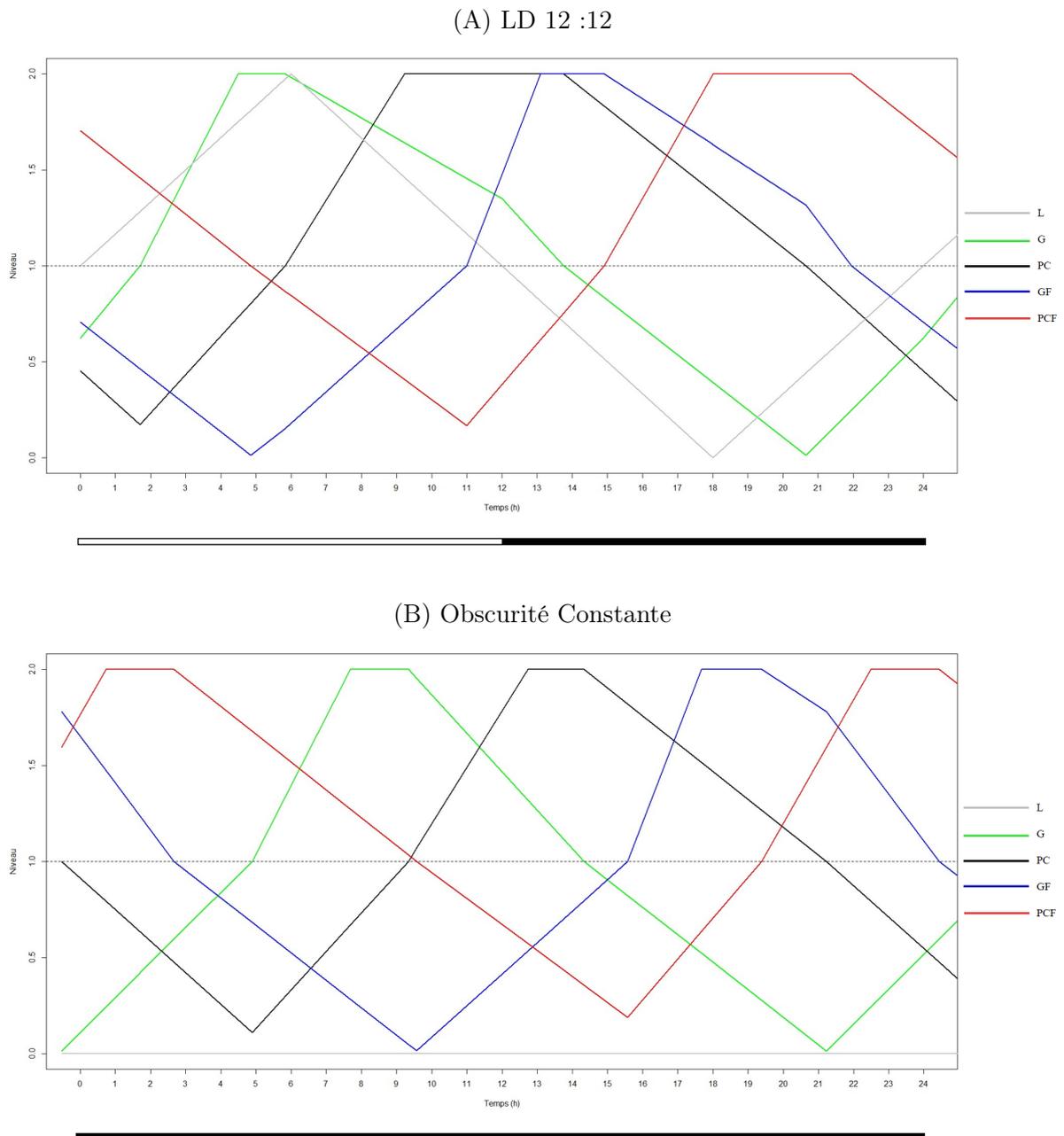


FIGURE 6.6 – (A et B) Simulation de la dynamique du modèle à 6 variables de la synchronisation du système circadien en conditions LD 12 : 12 et obscurité constante.

Le matériel utilisé est le simulateur développé en python, durant ce stage.

(A) Relation de phases sur 24h entre G, GF, PC et PCF et la lumière L. Simulation en condition LD 12 : 12 (alternance de 12h de lumière et 12h d'obscurité).

(B) Même relation de phases qu'en A. Simulation en condition LD 0 : 24 (obscurité constante).

Le temps total des 2 simulations est de 560h. Les données représentées proviennent de l'intervalle de temps 480-504 (h).

Une visualisation alternative des données, représentées dans un espace de phase, est disponible en **annexe 1.3**.

D'une part, le modèle du système circadien soumis à l'influence du zeitgeber de l'horloge centrale est initialisé en condition LD 12 : 12. Il s'agit de rendre effectif le cycle jour-nuit en le paramétrant d'après la **figure 6.4**. D'autre part, le modèle du système circadien *privé* de l'influence du zeitgeber de l'horloge centrale est initialisé en condition d'obscurité constante. Il s'agit cette fois, de rendre non fonctionnel le cycle jour-nuit, en fixant l'ensemble des célérités de L et X à 0.

La **figure 6.6 (A)** représente la simulation du modèle de la synchronisation du système circadien, en condition LD 12 : 12. Chaque variable du modèle oscille de manière circadienne. En obscurité constante, on retrouve en (B) des oscillations *persistantes et autonomes* des éléments de l'horloge centrale (NSC), mais également de ceux appartenant à l'horloge périphérique (foie). Ce résultat est cohérent si l'on se réfère à l'étude de **Yoo et al. 2004** [29]. Ces derniers mirent en évidence le caractère autonome des horloges centrales et périphériques.

La relation de phases en (A) montre que l'horloge centrale dans les NSC (G et PC) oscille en amont de l'horloge périphérique hépatique (GF et PCF). La synchronisation du système circadien est un mécanisme qui effectue la liaison entre ces deux systèmes. Il est donc cohérent que les éléments de l'oscillateur moléculaire central soient exprimés préalablement.

En considérant un seuil *moyen* de concentration, on retrouve une similitude dans l'ordre relatif des événements de régulation, par rapport aux données expérimentales.

Lorsque le modèle est soumis à un changement de régime de lumière en (B), on observe un décalage de phase de l'ensemble des variables. On remarque en outre, que le décalage de phase entre PC et GF est plus prononcé que les autres.

Cycle de lumière	G	PC	GF	PCF
LD 12 : 12	5.12	12.05	14.03	20.64
LD 0 : 24	8.79	13.80	18.82	25.28

Variable Mesurée	Intervalles de confiance des acrophases des données brutes					
	NSC			Foie		
	Acrophase (h)	IC (2.5%)	IC (97.5%)	Acrophase (h)	IC (2.5%)	IC (97.5%)
mPer2 (ARNm)	8.651	8.196	9.035	16.286	14.943	17.532
mPER2 (Protéine nucléaire)	12.951	12.212	13.719	19.763	19.168	20.424

1 : Valeur non significative p-value<0.05 (p=0.087234).

FIGURE 6.7 – Analyse des paramètres des rythmes circadiens générés par simulation.

Haut - Tableau des acrophases de chaque variable issues des données de simulation.

Bas - Tableau des acrophases issues des données expérimentales (en conditions LD 12 : 12) de la **figure 6.2**. Les intervalles de confiance des acrophases sont estimés par méthode bootstrap.

En haut de la **figure 6.7**, sont représentées les valeurs des acrophases des 4 entités moléculaires du modèle hybride. Dans une perspective d'analogie avec les données expérimentales, le tableau de la partie basse propose les valeurs *expérimentales* des acrophases en *conditions LD 12 : 12*. Ces résultats sont accompagnées d'une estimation des intervalles de confiance des acrophases.

D'après les données de simulation ; G, PC, GF et PCF atteignent leurs acrophases respectives ⁶ à **ZT 5.12**, **ZT 12.05**, **ZT 14.03** et **ZT 20.64**. Si l'on se réfère au tableau du bas de la **figure 6.7**,

6. Les dynamiques tracées en (A) et (B) présentent des « glissements ». Ces derniers représentent la saturation d'une entité moléculaire. L'acrophase est donc atteinte en plusieurs points. L'acrophase, dans le tableau (partie supérieure de la **figure 6.7**), est calculée comme la moyenne des points représentant le niveau maximal du niveau d'une variable.

les acrophases issues de la dynamique du modèle hybride n'appartiennent pas à leurs intervalles de confiance respectifs.

G, PC et GF sont en avance de phase par rapport aux données expérimentales tandis que PCF est en retard de phase.

D'autre part, lorsque le système circadien est synchronisé par la lumière (A), la période des oscillations atteint exactement $24h$. En obscurité constante (B), la période diminue jusqu'à $21.4h$. Aussi, la fenêtre de représentation des données est identique en (A) et (B) (de 480h à 504h) et la période des oscillations en (B) est plus petite. Par conséquent, le rythme des entités en (B) présente un **avancement drastique** de phase par rapport à la condition (A), face à un apparent retard de phase.

Ce résultat, certes quantitativement disproportionné, est *qualitativement* corroboré par **Field et al. 2000** [11] pour les NSC, et par **Lee et al. 2001** [20] pour le foie.

Chapitre 7

Discussions

7.1 Conséquences Méthodologiques de l'Abstraction des Connaissances dans un Modèle Hybride

Face à la multiplication des connaissances biologiques qualitatives et quantitatives, l'abstraction discrète des connaissances biologiques dans un modèle à 6 variables, facilite l'étude de la dynamique de la synchronisation du système circadien.

Ce modèle *simple* est capable de capturer le comportement périodique de l'oscillateur central, mais également de l'oscillateur périphérique. On est en mesure de qualifier ces oscillateurs d'*horloges*. En effet, ces derniers répondent aux critères suivants :

- Le modèle est capable de percevoir l'information temporelle par l'intermédiaire de l'information lumineuse (jour-nuit).
- Ces deux horloges sont capable de générer de manière endogène des oscillations non amorties, en présence ou en absence du zeitgeber *cycle jour-nuit*.
- En condition LD 12 :12, la période du système circadien modélisé vaut 24h. Cette période stable et persistante est exactement égale à celle du zeitgeber cycle jour-nuit, représenté par la variation circadienne de L. Les oscillations circadiennes sont *robustes* face au changement de régime lumineux proposé en **section 6.5**.
- La période de libre cours du système circadien, modélisé en obscurité constante, est fortement inférieure à 24h (*21.2h*). Ce résultat est *qualitativement* corroboré d'un point de vue expérimental. Néanmoins d'un point de vue quantitatif, des variations de période de cet ordre de grandeur ne sont pas retrouvées (période de libre cours de *23.45h* dans les travaux de **Lee et al.2001** [20] et de *23.5h* dans les travaux de **Yoo et al.2004** [29]).

Ce dernier point soulève une critique méthodologique de la méthode de modélisation employée au cours du stage. Le modèle à 6 variables présenté précédemment représente une *abstraction* du système circadien détaillé en **figure 1.2**. On constate une différence prononcée entre la période de libre cours expérimentale et celle issue de la simulation. De plus il persiste un léger décalage de phase entre données expérimentales et les données de simulation. Ces derniers peuvent être expliqués par deux hypothèses principales :

1. Le modèle proposé en **figure 6.1** n'est pas suffisamment détaillé pour obtenir une période de libre cours et une relation de phase en LD 12 : 12 proche des données expérimentales. La discrétisation de l'espace de concentration d'une (ou plusieurs) variable(s) en 2 niveaux discrets serait insuffisante. Il conviendrait alors de proposer plusieurs niveaux de concentration supplémentaires et d'en observer les conséquences sur la dynamique du modèle à plusieurs seuils.
2. La stratégie de paramétrisation utilisée en **section 6.3** ne permet pas de capturer les contraintes quantitatives exhaustives, à partir des données expérimentales. En effet, la paramétrisation par

contrainte de phase, est totalement dépendante des choix successifs d’ajustement des célérités. L’application exhaustive de contraintes de phase par logique de Hoare, représenterait une alternative intéressante à approfondir.

D’autre part, les résultats obtenus par modélisation qualitative mettent en évidence son indéniable capacité à capturer les caractéristiques rythmiques du système circadien. Il est ainsi intéressant d’envisager de construire et d’étudier la dynamique d’un modèle plus détaillé du système circadien. Ce modèle capturerait plus précisément en périphérie la régulation du gène *Per2* par *HSF1*. Il permettrait de capturer des comportements de synchronisation plus précis sur un modèle de plus grande taille qu’une méthode différentielle ou stochastique aurait probablement difficilement abordé.

7.2 Critère de Fiabilité des Données Biologiques Expérimentales

Forcé de constater que la paramétrisation du modèle est fortement dépendante des données biologiques de référence, on comprend l’importance de la caractérisation niveau de confiance imputé aux données expérimentales. En outre, le formalisme hybride construit de manière *déterministe* la dynamique d’un modèle hybride considérant une paramétrisation donnée. Une vigilance particulière se doit d’être portée aux analyses statistiques accompagnant la présentation des résultats expérimentaux quantitatifs.

Les données biologiques expérimentales, utilisées pour établir les contraintes biologiques sur paramètres de l’horloge hépatique [28], ne fournissent pas d’estimation de l’erreur standard de la moyenne. S’agissant d’une étude protéomique du foie de souris, les auteurs n’avaient pas pour objectif de produire des réplicats des mesures expérimentales, précisément sur mPer2 (GF) et mPER2 (PCF).

Étant donné l’ancienneté de l’article de référence, nous n’avons pas accès aux erreurs standard de la moyenne, calculées à partir des réplicats des données biologiques expérimentales brutes de mPer2 (G) et mPER2 (PC) dans les NSC [11].

La remise en question des données expérimentales de référence se base donc sur un croisement des données avec d’autres articles de la littérature [22].

Dans un soucis de rigueur, il serait intéressant par la suite, de générer des données biologiques expérimentales homogènes en termes de temps, de variabilité d’échantillonnage et de techniques expérimentales employées. En effet, les données cinétiques des transcrits G et GF d’une part et des protéines PC et PCF d’autre part, sont obtenues *via* 4 techniques de quantification différentes. En outre, les articles d’où sont issues les quantifications des éléments de l’horloge hépatique [28] et centrales, furent publiés à 17 ans d’intervalles.

Les perspectives d’améliorations les plus intéressantes, consisteraient à :

- Récupérer ou générer des données cinétiques expérimentales dans différents tissus, réalisées dans des conditions plus homogènes.
- Estimer l’erreur standard de la moyenne (SEM) des données expérimentales générées. Si les données se chevauchent, il serait envisageable de proposer plusieurs paramétrisations du modèle, en fonction des différents contraintes sur les valeurs de paramètres. La figure en **fig :tempe** propose un aperçu d’une analyse de donnée de la température corporelle avec SEM, généreusement fournies par la chercheuse Xiao-Mei Li.

7.3 Test dos à dos du Simulateur HyMBioNet et du Simulateur Prototype

Le test dos à dos (en anglais, *back to back testing*) permet de tester et vérifier le fonctionnement d’un programme informatique. Cette méthode permet d’éviter d’établir un ensemble exhaustif de tests

(méthode considérablement chronophage). Il s'agit plutôt de vérifier l'exactitude d'un nouveau programme en le confrontant à un prototype antérieur. Il s'agit ensuite de détecter les similitudes mais également les points de divergences entre les deux programmes.

Le *back to back test* a révélé plusieurs points de comparaison entre HyMBioNet et le simulateur prototype construit *ex nihilo*. Les traces biologiques obtenues par les deux programmes sont très similaires, à la différence qu'HyMBioNet génère d'infimes erreurs à chaque pas. Cette erreur accumulée au cours du temps est corrigée par le simulateur prototype.

Le nouveau simulateur conserve l'aspect générique d'HyMBioNet. En revanche, il présente l'avantage de construire une dynamique plus exacte en se basant au plus proche de la théorie de modélisation hybride.

Le simulateur prototype présente l'inconvénient majeur d'être non convivial à l'utilisation. Les perspectives d'améliorations consisteraient à mettre en place en priorité une interface homme-machine, afin de faciliter l'exécution du programme et la mise en place d'expériences *in silico* sur des modèles hybrides.

Quatrième partie

Conclusion et Perspectives

L'approche qualitative de modélisation des réseaux de régulation biologique, proposée par René Thomas en 1973, est communément utilisée en biologie systémique afin de construire et analyser le comportement dynamique d'un système complexe. Cette approche discrète offre un cadre formel permettant de mettre en évidence les propriétés dynamiques déterminantes d'un système complexe, organisé lui-même en sous-systèmes.

L'extension de cette approche discrète en un formalisme hybride de modélisation permet de prendre en compte additionnellement des informations chronométriques, indispensables dans l'étude de mécanismes de synchronisation. L'*étude de faisabilité* réalisée au cours du stage, a permis de modéliser avec succès le comportement oscillant d'un *premier modèle* de la synchronisation du système circadien.

En utilisant le principe de parcimonie, un modèle à 4 variables de la synchronisation de l'*horloge centrale* par la lumière, a été adapté en un modèle à 6 variables de la synchronisation du *système circadien*. L'étude du comportement dynamique du modèle hybride a été rendue possible *via* l'utilisation d'un simulateur générique fonctionnel également conçu et développé au cours de ce stage de 5 mois. Le modèle hybride du système circadien capture des caractéristiques statiques et dynamiques propres aux horloges circadiennes. Cependant, chaque modèle est dépendant des abstractions biologiques et des questions biologiques définies préalablement. Les conclusions ne peuvent en aucun cas s'étendre à des questions plus fines qui nécessiteront des modèles plus élaborés. De plus, constatant une pénurie de données expérimentales, il serait techniquement intéressant de mettre en *pratique* des protocoles expérimentaux permettant de générer des données quantitatives, accompagnées d'une estimation de l'erreur standard à la moyenne.

Dans la continuité du stage, il s'agirait de construire un modèle de *référence* du système circadien en condition non pathologique. Graduellement plus élaboré, le modèle inclurait dans un premier temps la régulation du gène *Per* *via* la variation de température dans un, voire plusieurs, organes périphériques. Aussi, la paramétrisation du nouveau modèle pourrait se fonder sur des données expérimentales de températures corporelles de souris déjà disponibles. La figure en **annexe 1.3** propose un aperçu de l'analyse de données expérimentales de la température corporelle¹.

L'étape de paramétrisation constitue l'étape limitante dans tout processus de modélisation. L'identification des célérités pourrait être facilitée par une automatisation de la construction des contraintes sur paramètres par logique de Hoare [2]. Un outil développé par l'équipe MDSC combiné avec un solveur de contraintes pourrait fournir les paramètres en entrée du simulateur, pour *in fine* analyser la dynamique générée.

Enfin le système circadien possède la particularité de distribuer en sortie un message rythmique à une multitude de voie de régulation de la physiologie et du comportement d'un individu. Des *disruptions* du système circadien (e.g décalage horaire chronique ou social) conduisent particulièrement à une hausse de l'incidence des troubles du sommeil, des maladies métaboliques [16] [14] mais également du développement des pathologies cancéreuses [10]. Les outils de modélisation hybride pourraient permettre, à long terme, de comprendre les effets des perturbations du cycle circadien sur ces troubles physiologiques et d'ajuster l'action d'une ou plusieurs entité(s) biologique(s) à des fins thérapeutiques.

1. Données aimablement fournies par la chercheuse Xiao-Mei Li (Inserm Villejuif). Ayant accès aux répliqués, nous avons estimé *l'erreur standard à la moyenne* à chaque mesure ponctuelle. Si les données expérimentales des entités du modèle se chevauchent, il serait envisageable de proposer plusieurs paramétrisations du modèle

Bibliographie

- [1] J. C. Ameisen, E. Klein, and D. Leglu. *Écrire la science*. Les Presses de l'ENSTA, 2010. Interview menée par Érick Castellan, Solène Charpentier et Théau Cosseron.
- [2] J. Behaegel, J. P. Comet, and M. Folschette. A hybridisation of the genetically modified hoare logic. *arXiv :1610.06715*, October 2016.
- [3] G. Bernot, J. P. Comet, and Z. Khalis. Gene regulatory networks with multiplexes. *European Simulation and Modelling Conference Proceedings*, pages 423–432, 2008.
- [4] G. Bernot, J. P. Comet, Z. Khalis, A. Richard, and O. Roux. A genetically modified hoare logic. *ArXiv : 1506.05887*, June 2015.
- [5] G. Bernot, J. P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks : extending thomas asynchronous logical approach with temporal logic. *J.theor.Biol*, 229 :339–347, 2004.
- [6] E. D. Buhr, S. H. Yoo, and J. S. Takahashi. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science*, 330 : 379-385, October 2010.
- [7] E. Cornillon, J. P. Comet, G. Bernot, and G. Énée. Hybrid gene networks : a new framework and a software environment. *Advances in systems and synthetic biology - Proceedings of the evry spring school*, pages 57–84, 2016.
- [8] F. Damiola, N. L. Minh, N. Preitner, B. Kornmann, F. Fleury-Olela, and U. Schibler. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes & Development*, 14 : 2950-2961, 2000.
- [9] G. Enée. Hymbionet. <http://www.i3s.unice.fr/~enee/HyMBioNet/>.
- [10] C. Feillet, G. T. J. van der Horst, F. Levy, D. A. Rand, and F. Delaunay. Coupling between the circadian clock and cell cycle oscillators : implication for healthy cells and malignant growth. *Frontiers in neurology*, page 6 : 96, May 2015.
- [11] M. Field, E. S. Maywood, J. A. O'Brien, D. R. Weaver, S. M. Reppert, and M. H. Hastings. Analysis of clock proteins in mouse scn demonstrates phylogenetic divergence of the circadian clockwork and resetting mechanisms. *Neuron*, 25 :437–447, February 2000.
- [12] A. Goldbeter. *Vie oscillatoire (La) : Au cœur des rythmes du vivant*. Editions Odile Jacob, 2010.
- [13] Hans Reinke, C. Saini, F. Fleury-Olena, C. Dibner, I. J. Benjamin, and U. Schibler. Differential display of dna-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes Dev*, 22 : 331-345, November 2008.
- [14] M. H. Hastings, A. B. reddy, and E. S. Maywood. A clockwork web : circadian timing in brain and periphery, in health and disease. *Nature Reviews Neuroscience*, 4 :649–661, August 2003.
- [15] C. A. R. Hoare. An axiomatic basis for computer programming. *Communications of the ACM*, pages 12 : 576–580, October 1969.

- [16] W. Huang, K. M. Ramsey, B. Marcheva, and J. Bass. Circadian rhythms, sleep, and metabolism. *J Clin Invest*, 121 :2133–2141, June 2011.
- [17] A. Ishida, T. Mutoh, T. Ueyama, H. Bando, S. Masubuchi, D. Nakahara, G. Tsujimoto, and H. Okamura. Light activates the adrenal gland : timing of gene expression and glucocorticoid release. *Cell Metab*, 2 : 297-307, November 2005.
- [18] N. Koike, S. H. Yoo, H. C. Huang, V. Kumar, C. Lee, T. K. Kim, and J. Takahashi. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Cell*, 338 : 349-354, October 2012.
- [19] B. Kornmann, O. Schaad, H. Bujard, J. S. Takahashi, and U. Schibler. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biology*, 5 : e34, January 2007.
- [20] C. Lee, J. P. Etchegaray, F. R. A. Cagampang, A. S. I. Loudon, and S. M. Reppert. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell*, 107 :855–867, December 2001.
- [21] K. R. Popper. *Conjectures et réfutations*. Payot, 2006.
- [22] J. A. Ripperger, C. Jud, and U. Albrecht. The daily rhythm of mice. *FEBS Letters*, 585 :1384–1392, May 2011.
- [23] J. Rutter, M. Reick, L. C. Wu, and S. L. McKnight. Regulation of clock and npas2 dna binding by the redox state of nad cofactors. *Science*, 293 :510-514, July 2001.
- [24] C. Saini, J. Morf, M. Stratmann, P. Gos, and U. Schibler. Simulated body temperature rhythms reveal the phase-shifting behavior and plasticity of mammalian circadian oscillators. *Genes Dev*, 26 : 567-580, February 2012.
- [25] A. Y. L. So, T. U. Bernal, M. Pillsbury, K. R. Yamamoto, and B. J. Fieldman. Glucocorticoid regulation of the circadian clock modulates glucose homeostasis. *Proc Natl Acad Sci U.S.A*, pages 106 : 17582–17587, October 2009.
- [26] M. Stratmann and U. Schibler. Properties, entrainment, and physiological functions of mammalian peripheral oscillators. *J.Biol.Rhythms*, 21 : 494-506, December 2006.
- [27] R. Thomas. Boolean formalization of genetic control circuits. *J.theor.Biol*, 42 :563–585, 1973.
- [28] J. Wang, D. Mauvoisin, E. Martin, F. Atger, A. N. nez Galindo, L. Dayon, F. Sizzano, A. Palini, M. Kussmann, P. Waridel, M. Quadroni, V. Dulic, F. naef, and F. Gachon. Nuclear proteomics uncovers diurnal regulatory landscapes in mouse liver. *Cell Metabolism*, 25 :102–117, Janvier 2017.
- [29] S. H. Yoo, S. Yamazaki, P. L. Lowrey, K. Shimomura, C. H. Ko, E. D. Buhr, S. M. Siepk, H. K. Hong, W. J. Oh, O. J. Yoo, M. Menaker, and J. S. Takahashi. Period2-luciferase real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc Natl Acad Sci U.S.A*, 101 : 5339-5346, April 2004.

Cinquième partie

Annexes

1.1 Logique Formelle et Tables de Vérité.

Soit p et q deux formules propositionnelles. Les relations logiques de base entre ces deux propositions sont les suivantes :

- **La négation** : $\neg p, \neg q$: non p , non q . Si p est fausse alors $\neg p$ est vraie.
- **La conjonction** : $p \wedge q$: p et q . $(p \wedge q)$ est vraie si et seulement si les deux propositions p et q sont vraies.
- **La disjonction** : $p \vee q$: p ou q . $(p \vee q)$ est vraie si au moins l'une des deux propositions est vraie.
- **L'implication** : $p \Rightarrow q$: p implique q .

Une table de vérité est un tableau utilisé en logique pour représenter la sémantique des expressions logiques. Il permet de donner la valeur de vérité (vraie ou fausse) d'une proposition. En outre, les tables de vérité permettent de réaliser des démonstrations d'équivalence entre deux expressions logiques. En effet, *si les tables de vérités de deux expressions logiques sont identiques*, alors on a prouvé l'équivalence logique entre ces deux expressions.

En utilisant des règles d'inférence, on peut déduire des propositions à partir d'autres propositions.

p	q	$p \wedge q$	$p \vee q$	$\neg p \vee \neg q$	$\neg(\neg p \vee \neg q)$
F	F	F	F	V	F
F	V	F	V	V	F
V	F	F	V	V	F
V	V	V	V	F	V

$$p \wedge q \Leftrightarrow \neg(\neg p \vee \neg q)$$

FIGURE 1.1 – Démonstration de $(p \wedge q) \Leftrightarrow (\neg(\neg p \vee \neg q))$

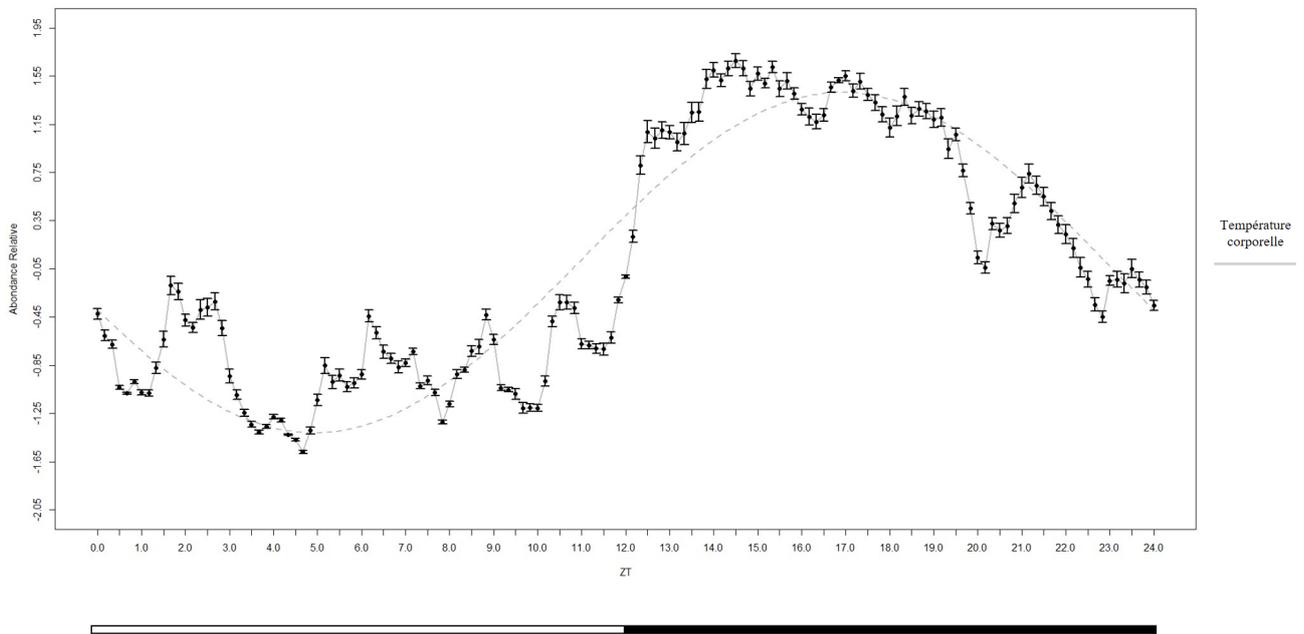
p	$\neg p$	q	$p \Rightarrow q$	$p \vee q$	$\neg p \vee q$
F	V	F	V	F	V
F	V	V	V	V	V
V	F	F	F	V	F
V	F	V	V	V	V

$$p \Rightarrow q \Leftrightarrow (\neg p \vee q)$$

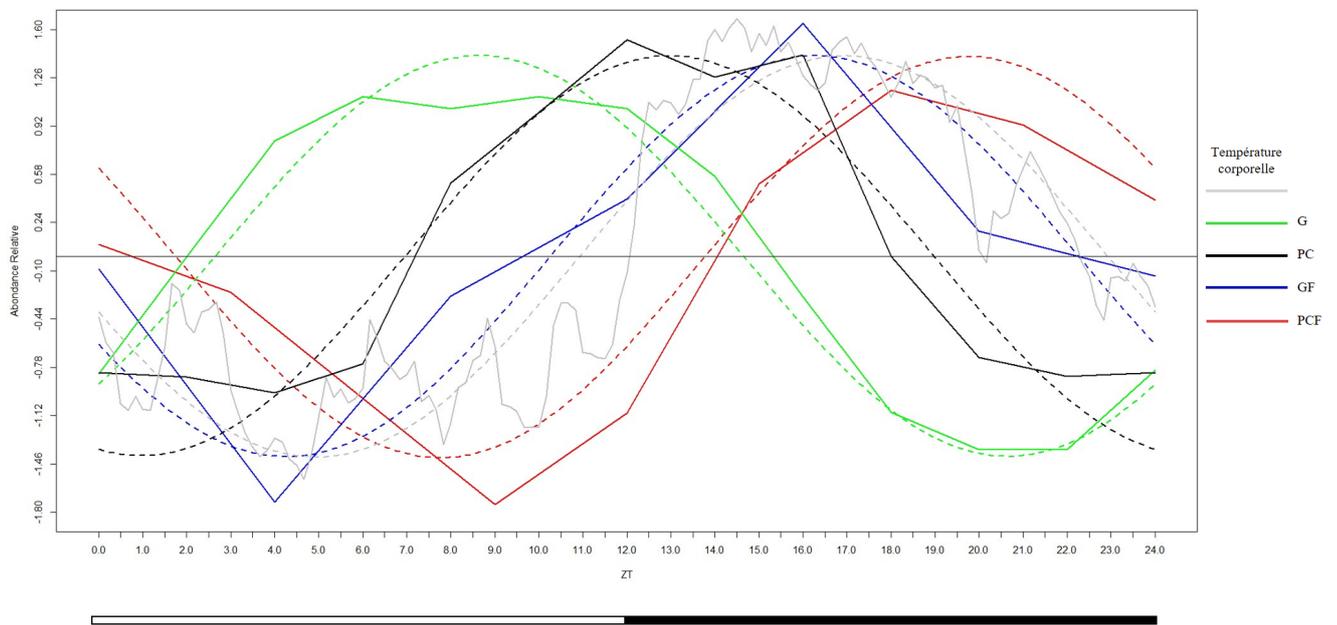
FIGURE 1.2 – Démonstration de $(p \Rightarrow q) \Leftrightarrow (\neg p \vee q)$

1.2 Données Expérimentales de Température Corporelle.

(A)



(B)



(C)

Variable Mesurée	Paramètres de l'Analyse Cosinor Centrée et Réduite		
	Niveau Moyen	Amplitude	Acrophase (h)
mPer2 (ARNm NSC)	290.71 ± 9.97	221.10 ± 14.18	8.65 ± 0.33
mPer2 (ARNm Foie)	6.21 ± 0.26	1.50 ± 0.37	16.26 ± 0.89
mPER2 (Protéine nucléaire NSC)	1.24 ± 0.02	0.25 ± 0.03	12.97 ± 0.43
mPER2 (Protéine nucléaire Foie)	-0.28 ¹ ± 0.13	1.65 ± 0.21	19.76 ± 0.40
Température corporelle	37.22 ± 0.02	0.93 ± 0.03	16.9 ± 0.14

¹ : Valeur non significative p-value<0.05 (p=0.087234).

FIGURE 1.3 – (A) Profil de température corporelle de souris sur 24h avec estimation de l'erreur standard à la moyenne - (B) Relation de phases entre la température corporelles de souris, ARNs mPer2 et protéines nucléaires mPER2 dans les NSC et le foie de souris mâles C57B6 accompagné de l'ajustement sinusoïdale des courbes - (C) Analyses statistiques des données expérimentales par méthode cosinor.

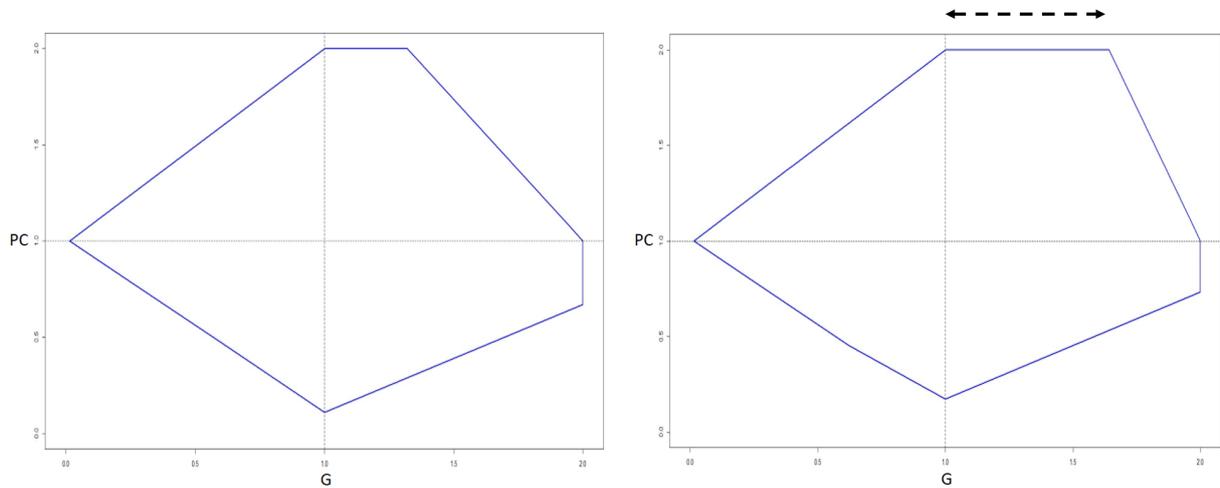
Profil d'expression sur 24h des transcrits Per2 dans les NSC (G en vert) et dans le foie (GF en bleu) mesurés respectivement par hybridation *in situ* et totale RNA-seq. Abondance relative sur 24h des protéines nucléaires PER2 dans les NSC (PC en noir) et dans le foie (PCF en rouge) mesurées respectivement par western blot et spectrométrie de masse. La température corporelle a été mesurée par télémétrie avec une résolution de 10 min sur une durée de 48h.

L'erreur standard à la moyenne (SEM) est représentée graphiquement en A par des barres d'erreurs. La SEM a été calculée à partir de 7 réplicats expérimentaux provenant de 7 souris C57B6 mâles différentes.

Les données expérimentales proviennent toutes de souris placées dans un environnement LD 12 : 12 (barre blanche : 12h de lumière - barre noire : 12h d'obscurité). Dans un soucis d'homogénéité de variances, les données représentées en A et B ont été centrées et réduites.

1.3 Graphe d'États de l'Horloge Circadienne Centrale et Périphérique Hépatique.

(A) NSC



(B) Foie

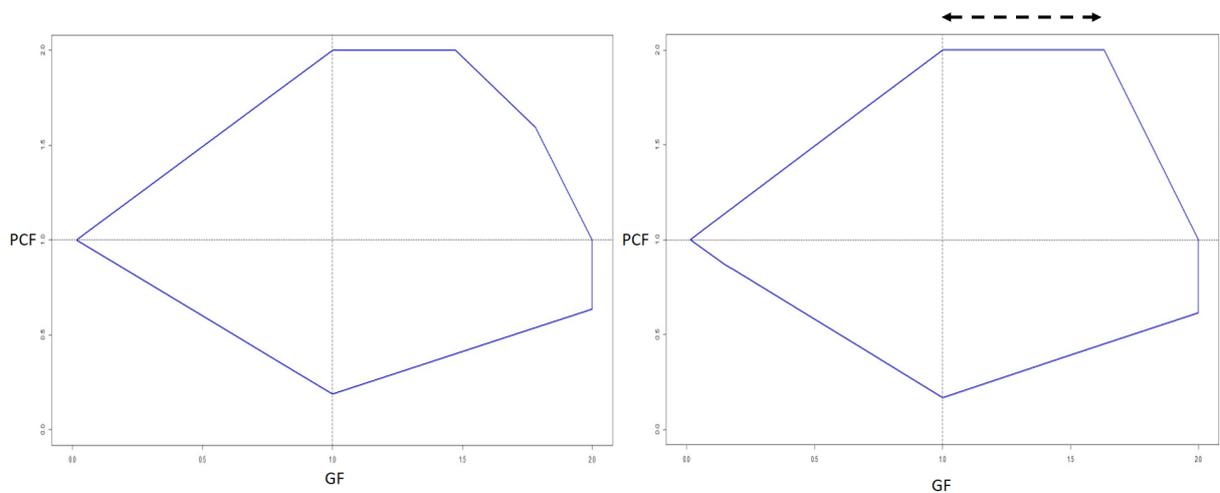


FIGURE 1.4 – Graphe d'états de l'horloge centrale et hépatique en obscurité constante (à gauche) et en conditions LD 12 : 12 (à droite)

Représentation de l'espace des phases avec le niveau qualitatif du transcrit *Per2* en abscisse et le niveau qualitatif de la protéine nucléaire *PER2* en ordonnée.

Ces graphes mettent en évidence dans chaque cadran, un cycle limite caractéristique d'un système oscillant non amorti. Lorsque le système modélisé est soumis à l'influence du zeitgeber de l'horloge centrale (alternance de 12h de lumière et 12h d'obscurité), la protéine *PC* dans les NSC et *PCF* dans le foie sont activées plus rapidement. Le glissement de *PC* (respectivement *PCF*) sur l'abscisse *G* (respectivement *GF*) est plus long en condition LD 12 : 12 (flèches en pointillés).

Cette simulation capture la connaissance biologique suivante : La lumière joue un rôle de synchroniseur de la phase des entités de l'horloge, en activant de surcroît la transcription du gène *Per2* dans les NSC (*G*). En réponse à cet événement, la protéine *PERCRY* subit par extension cette augmentation de production. Cet événement positif de régulation est impacté sur le foie par l'intermédiaire de l'arc synchronisant $PC \rightarrow GF$ (cf. figure 6.1).

1.4 Bilan des Compétences Acquisées.

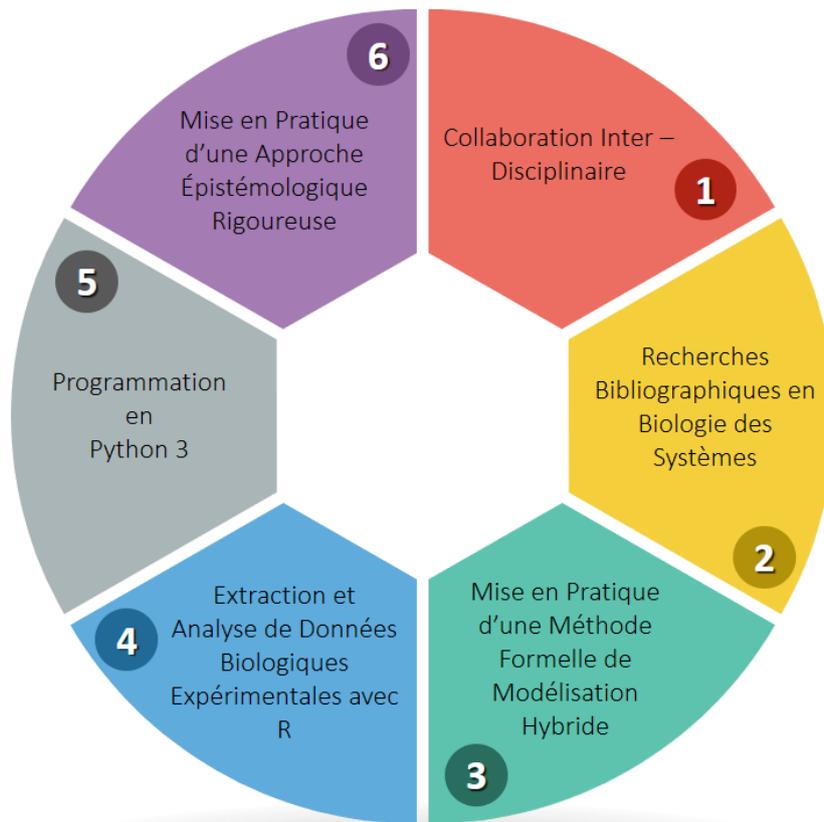


FIGURE 1.5 – **Bilan des compétences acquises et développées au cours du stage de M1.** Les compétences se résument en six points principaux. Les cinq premiers sont ordonnés chronologiquement alors que le dernier est à considérer en parallèle des 5 premiers.

1. Ce stage s'inscrit dans le cadre d'une **collaboration interdisciplinaire** entre les laboratoires iBV et I3S. Cette collaboration nécessita d'établir et entretenir un lien régulier et équilibré entre une équipe de biologistes des systèmes (iBV), et de modélisateurs de réseaux de régulation génétique (I3S).
2. La **recherche bibliographique** constitue une étape cruciale dans l'élaboration des connaissances sur lesquelles se basent les travaux de modélisation. La **biologie des systèmes** fait intervenir une grande variété de domaines en biologie. De ce fait, il fut primordial d'apprendre à s'orienter dans la complexité qu'engendre la recherche de telles quantités de connaissances biologiques.
3. La mise en pratique de la **modélisation hybride** de la dynamique du système circadien, nécessita la compréhension préalable des **cadres formels** rigoureux de raisonnement. Ces derniers englobent notamment les logiques propositionnelles, temporelles ou encore la logique de Hoare ; mais également l'approche booléenne de M. René Thomas sur laquelle se base la modélisation hybride de réseau de régulation génétique.
4. L'étape limitante dans le processus de modélisation hybride de la synchronisation du système circadien réside principalement dans l'identification de l'ensemble des paramètres du modèle. Ce problème exigea de mettre en place une **méthode de recherche, d'extraction et d'analyses des données expérimentales**, en se basant sur les diverses bases de données à disposition (PubMed, CircaDB, Array Express, GEO).

5. La conception d'un simulateur de dynamique de réseau de régulation génétique, impliqua au préalable d'acquérir des connaissances sur le langage **Python**, mais également « les bonnes pratiques de programmation ». Par exemple, réfléchir préalablement aux structures de données et à l'architecture du programme, définir des fonctions de petites tailles optimisées en terme de complexité et surtout de s'assurer de bien documenter son programme.
6. De manière plus globale, le travail de modélisation nécessita de porter constamment une attention particulière à la mise en pratique de **principes épistémologiques** tels que : la réfutabilité d'une théorie ou d'une hypothèse scientifique, la remise en question des méthodes et des données expérimentales ou encore la critique des analyses statistiques utilisées.