

2m11.2705.4

Université de Montréal

Élaboration d'un protocole pour le criblage fonctionnel
des gènes des dinoflagellés .

par

Hélène Chaput

Départements des Sciences Biologiques

Faculté des arts et sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de

Maître ès sciences (M. Sc.)
en sciences biologiques

février, 1999

©Hélène Chaput, 1999



8011-2702-V
QK
3
U54
1999
V.005

Université de Montréal

Élaboration d'un protocole pour le criblage génétique
des gènes des dinoflagellés

par

Hélène Clapar

Département des Sciences Biologiques

Faculté des arts et sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de

Maîtrise en sciences (M.Sc.)
en sciences biologiques

février, 1999

Hélène Clapar, 1999



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
Élaboration d'un protocole pour le criblage fonctionnel des gènes des dinoflagellés .

Présenté par :
Hélène Chaput

a été évalué par un jury composé des personnes suivantes :

Anctil, Michel	Président-rapporteur
Morse, David	Directeur de recherche
Cappadocia, Mario	Membre du jury

date: 10 mai 1999

SOMMAIRE

L'algue unicellulaire *Gonyaulax polyedra* est notre modèle pour l'étude des mécanismes reliant l'horloge interne aux rythmes circadiens observés. La bioluminescence, présente lors de la phase de nuit, ainsi que la division cellulaire observée au début de la phase de jour sont les deux rythmes sur lesquels porte la présente étude. Chez *Gonyaulax*, la présence des protéines LBP et luciférase, limitée à la phase obscure, permet la production de lumière. Une régulation au niveau de la traduction par une protéine inhibitrice CCTR se fixant en région 3' de l'ARNm LBP permet l'expression cyclique de la protéine LBP *in vivo* chez *Gonyaulax polyedra*. La construction d'une banque d'ADNc dans un vecteur d'expression constitutive (le p425 GPD) représente l'outil nécessaire pour l'isolation d'un ADNc codant pour la protéine CCTR. Cette isolation repose sur un mécanisme moléculaire d'interaction *in vivo* de la protéine CCTR avec la région régulatrice de l'ARNm lbp annexée à un gène létal inductible (le gène GSP1) dans la construction du plasmide pCS7. La transformation de levure possédant le pCS7 avec la banque d'ADNc permettrait la production de CCTR actif dans la levure et ainsi l'inhibition de la traduction de l'ARNm GSP1 par fixation du CCTR à la région régulatrice de lbp. Le criblage d'une banque d'ADNc comportant $2,8 \times 10^5$ clones a été effectué. La croissance de 127 colonies a été observée lors du premier criblage de la banque. Chaque colonie a été soumise à une extraction d'ADN plasmidique et les plasmides obtenus ont été utilisés pour une seconde transformation dans les levures porteuses du pCS7. Le second criblage nous a permis de constater qu'aucun de ces positifs ne peut produire une protéine CCTR fonctionnelle. Une contamination par des levures d'une autre souche ou par une source de carbone permettant la croissance (telle que le glucose), une présence de "révertants" ou le clonage d'un ADNc codant pour le facteur de sélection du pCS7 peuvent être à l'origine de ces "faux positifs" obtenus. La production de clones supplémentaires afin d'augmenter statistiquement les chances de représentation d'un gène impliqué dans la génération d'un rythme

circadien, ainsi que la transformation répétée de la banque dans les mêmes cellules pour permettre de réunir les différentes composantes qui pourraient être nécessaires à la formation d'un CCTR actif, pourraient permettre d'isoler un ADNc codant pour une protéine CCTR active.

Chez la majorité des dinoflagellés on observe la présence d'une phase S dans le cycle cellulaire. De plus, chez *Gonyaulax*, la division cellulaire est limitée à l'aurore. Il est donc probable que la protéine kinase $p34^{cdc2}$ qui régule la formation de l'hétérodimère MPF (impliqué dans l'initiation de la division cellulaire) soit également sujette à une régulation par l'horloge circadienne. L'isolation d'un homologue de la protéine kinase $p34^{cdc2}$ repose sur une technique de clonage par compensation d'une mutation. Les levures de la souche *cdc28* ne peuvent se diviser normalement à température restrictive de 34°C à cause d'une mutation thermosensible de la protéine kinase $p34^{cdc2}$. La transformation de la banque d'ADNc construite dans des levures de cette souche, puis l'incubation à température restrictive, permet le criblage pour un homologue de la $p34^{cdc2}$. Le premier criblage de la banque a permis la croissance de sept colonies qui ont toutes été soumises à une extraction d'ADN plasmidique puis à une retransformation dans les cellules de la souche *cdc28*. Aucun des positifs n'a franchi le seuil de ce second test. Le criblage de clones supplémentaires ou encore le criblage pour un homologue de la cycline (la seconde composante du MPF) pourrait permettre d'isoler directement ou indirectement l'homologue de la $p34^{cdc2}$.

TABLE DES MATIÈRES

SOMMAIRE	iii
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	ix
LISTE DES ABBRÉVIATIONS	x
REMERCIEMENTS	xiii
DÉDICACE	xiv
1. INTRODUCTION	1
1.1 La biologie des dinoflagellés	1
1.1.1 Biologie générale	1
1.1.2 Particularités du noyau	4
1.1.3 Le cycle cellulaire des dinoflagellés	5
1.1.4 Un organite bien particulier : le scintillon	7
1.2 Les dinoflagellés comme modèle pour l'étude des rythmes	8
1.2.1 Généralités sur les rythmes	8
1.2.2 Les rythmes observables chez <i>G. polyedra</i>	10
1.2.3 Les différents rythmes sont contrôlés par différentes horloges chez <i>G. polyedra</i>	15
1.3 Les horloges circadiennes : Étude comparée des mécanismes moléculaires	16
1.3.1 L'analyse génétique de la rythmicité	19
1.3.2 L'horloge génère un cycle par une boucle de rétroaction négative	21

1.3.3 Un zeitgeber essentiel : la lumière	24
1.3.4 La compensation aux variations de température	26
1.3.5 Régulation des "outputs"	27
1.4 Aperçu du projet	28
2. MATÉRIEL ET MÉTHODES	30
2.1 Culture de cellules	30
2.2 Acides nucléiques	30
2.2.1 Purification des plasmides	30
2.2.2 Purification de l'ADN génomique	31
2.2.3 Purification de l'ARN total	32
2.2.4 Purification des ARN messagers	33
2.2.5 Quantification des acides nucléiques	33
2.2.6 Digestions enzymatiques	33
2.2.7 Électrophorèse et électroélution	35
2.2.8 Transfert de types Southern et Northern	35
2.2.9 Marquage de l'ADN et hybridation	36
2.3 Banque d'ADNc	36
2.3.1 Production d'une banque d'ADNc dans les phages	36
2.3.2 Criblage d'une banque d'ADNc dans le vecteur Uni-Zap XR	37
2.3.3 Production d'une banque d'ADNc dans le vecteur p426 GPD	37
2.3.4 Criblage de la banque d'expression dans le vecteur p426 GPD chez la levure	40

2.4 Réaction en chaîne de polymérisation	40
2.5 Séquençage	44
2.6 Cellules compétentes et transformation	44
2.6.1 Chez <i>E. coli</i>	44
2.6.2 Chez la levure	45
3. RÉSULTATS	46
3.1 Contrôle circadien du rythme de la bioluminescence	46
3.1.1 L'isolation d'un ADNc codant pour la protéine CCTR	46
3.2 Contrôle circadien du rythme de la division cellulaire	52
3.2.1 Isolation d'un homologue de la kinase p34 ^{cdc2}	52
4. DISCUSSION	58
4.1 Isolation d'un ADNc codant pour la protéine CCTR	58
4.2 Isolation d'un ADNc homologue du gène <i>cdc2</i>	63
4.3 Conclusion	66
5. Bibliographie	67
ANNEXE : Article 1	xv

LISTE DES TABLEAUX

<u>Tableau I</u> : Les gènes impliqués dans la génération des rythmes	18
--	-----------

LISTE DES FIGURES

<u>Figure 1</u> :	Le dinoflagellé, <i>Gonyaulax polyedra</i>	3
<u>Figure 2</u> :	Deux rythmes observables chez <i>Gonyaulax polyedra</i>	12
<u>Figure 3</u> :	Les composantes d'une horloge circadienne	17
<u>Figure 4</u> :	Contrôle de la qualité des ARNm	34
<u>Figure 5</u> :	Contrôle de la qualité de la banque d'ADNc dans le vecteur Uni-Zap XR	38
<u>Figure 6</u> :	Production d'une banque d'ADNc dans le vecteur p426GPD	39
<u>figure 7</u> :	Contrôle de la qualité de la banque d'ADNc dans le vecteur p426GPD	41
<u>Figure 8</u> :	Construction du plasmide pCS7	42
<u>Figure 9</u> :	Interaction moléculaire <i>in vivo</i>	48
<u>Figure 10</u> :	Criblage pour isoler le CCTR	50
<u>Figure 11</u> :	Second criblage des positifs lors de l'isolation du CCTR	51
<u>Figure 12</u> :	Interaction moléculaire <i>in vivo</i> à 34°C	54
<u>Figure 13</u> :	Criblage pour l'isolation d'un homologue de la kinase p34 ^{cdc2}	56
<u>Figure 14</u> :	Second criblage des positifs lors de l'isolation de la p34 ^{cdc2}	57

LISTE DES ABRÉVIATIONS

ADN	acide désoxyribonucléique
ARN	acide ribonucléique
ARNm	acide ribonucléique messenger
ATP	adénosine triphosphate
bp	paire de base (s)
BSA	albumine de sérum de boeuf
C¹⁴	carbone 14
°C	degré Celsius
Ca²⁺	ion calcium
CCTR	" circadian controlled translational regulator "
clk	" clock "
CO₂	dioxyde de carbone
CT	temps circadien
CTAB	bromure de cetyltriméthylammonium
CTP	cytosine triphosphate
cyc	" cycle "
CYC	cycline
DEPC	diéthyl pyrocarbonate
EDTA	acide éthylènediamine tétraacétique
EGTA	acide éthylène glycol-O, O-bis (2-aminométhyle) -N, N, N, N-tétraacétique
frq	" frequency "
GAL	galactose
GIT	Guanidine isothiocyanate
H⁺	ion hydrogène
HOMeUra	5- hydroxyméthyluracile
IPTG	isopropyl thio-β-galactoside
kb	kilobase
kDa	kiloDalton

lbp	protéine de liaison à la luciférine
L/D	jour/nuit
Leu	leucine
LiAc	acétate de lithium
LiCl	chlorure de lithium
m	mètre
M	molaire
mM	millimolaire
Mg²⁺	ion magnésium
ml	millilitre
MPF	facteur promoteur de l'entrée en mitose
MTOC	centre d'organisation des microtubules
N	normal
NaCl	chlorure de sodium
Na₂HPO₄	phosphate de sodium dibasique
NaOAc	acétate de sodium
NaOH	hydroxide de sodium
ng	nanogramme
nm	nanomètre
ORF	cadre de lecture ouvert
³²P	phosphate radioactif (³² P)
PCP	péridinine- chlorophylle a- protéine
PCR	réaction en chaîne de polymérase
PEG	polyéthylèneglycol
per	" period "
pfu	unité de formation de plage de lyse
pH	mesure acide/base
pg	picogramme
RNAse	ribonucléase
RPM	tour par minute
s	seconde

ssDNA	ADN simple brin
SDS	sodium dodecyl sulfate
SSC	Citrate de sodium salin
TE	Tris- EDTA
thr	thréonine
tim	" timeless "
Tris-Cl	tris (hydroxyméthyle) aminométhane-chlore
tyr	tyrosine
µg	microgramme
µl	microlitre
µm	micromètre
Ura	uracile
UTR	région non-transcripte
wc	" white collar "
ZT	" zeitgeber time "

REMERCIEMENTS

J'aimerais d'abord et avant tout remercier avec un grand "R" mon directeur de recherche le docteur David Morse . Sa patience, sa passion pour la recherche, sa grande disponibilité, son support et son admirable sens de l'humour en font le directeur idéal pour moi !! Merci Dave !!

J'aimerais également remercier particulièrement ma douce compagne de laboratoire Nasha qui sait toujours quoi dire dans les moments plus difficiles. Merci à mes premiers confrères et guides Hien et Yosser qui m'ont généreusement transmis leur savoir. Merci à tous mes amis de l'institut pour les agréables cinq à sept, les victoires au scrabble (!!) et les promenades dans le merveilleux parc du Jardin Botanique . Merci au Dr. Daniel Matton ainsi qu' au Dr. Vincenzo de Luca et à leurs équipes pour les conseils et les "petits" prêts de matériel !! Merci à tous mes amis et à ma famille pour leur support et leurs encouragements . Lâchez pas !!

À mes parents

1. Introduction

1.1 La biologie des dinoflagellés

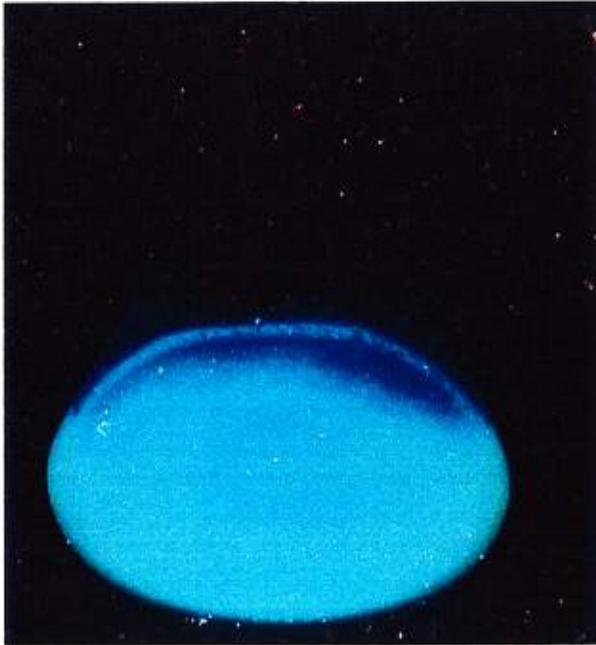
1.1.1 Biologie Générale

Les dinoflagellés ou Pyrrophytes (du Grec pyrros " pour de la couleur du feu ") forment une vaste division comprenant près de 1200 espèces différentes regroupées en 130 genres (Spector, 1984). Ces protozoaires sont des algues unicellulaires marines ou d'eau douce qui peuvent être phototrophes, phagotrophes ou mixotrophes étant donné qu'on les retrouve dans le phytoplancton ; elles sont donc à la base de nombreuses chaînes alimentaires dans les écosystèmes aquatiques. Leur taille varie de 0,01 à 2 mm de diamètre et elles possèdent toutes un stade mobile à l'intérieur du cycle de leur vie. Ce stade mobile est possible grâce à la présence de 2 flagelles (une longitudinale et une transversale) qui leur permettent de nager à des vitesses variant de 75 à 370 $\mu\text{m/s}$; et donc de parcourir de 20 à 30 mètres en 24 heures (Spector, 1984). Leur mode de reproduction est principalement asexué (par mitose) mais une reproduction sexuée a également été observée chez 18 espèces différentes réparties dans 12 genres (Spector, 1984). Certaines espèces marines sont reconnues pour leur capacité de produire de la lumière durant la nuit. Ce phénomène est appelé bioluminescence et il est possible de l'observer seulement lors de la nuit en nature. D'autres espèces marines ont une toute autre réputation puisqu'elles sont reconnues pour être responsables des marées rouges. Lors de ce phénomène la population d'algues est si dense que l'on peut dénombrer jusqu'à 10^7 cellules par litre d'eau (Homma, 1989a). La couleur alors observée est due aux pigments photosynthétiques de ces algues (la péricidine) qui est rouge. Les dinoflagellés

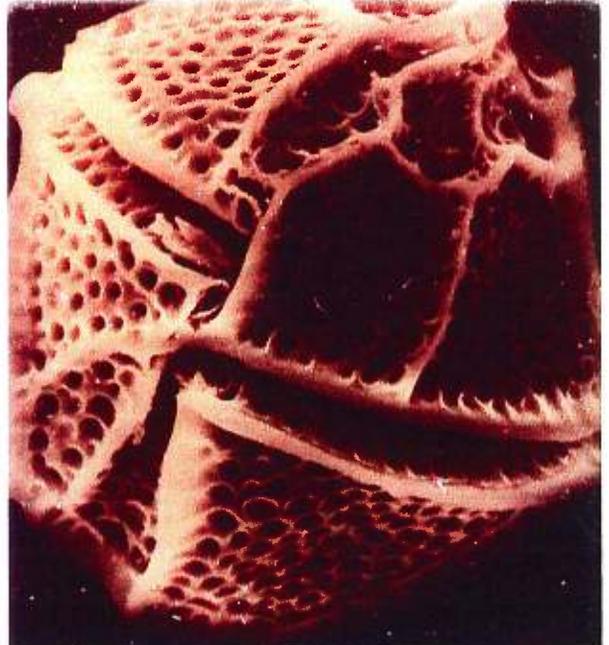
peuvent produire des neurotoxines qui, lors de ces marées rouges, se retrouvent en quantité suffisante pour entraîner la mort de nombreux organismes marins, bouleversant ainsi toute la chaîne alimentaire.

La grande majorité des dinoflagellés possèdent, à l'intérieur de la membrane plasmique, un complexe de vésicules contenant des plaques de cellulose que l'on nomme thèque. Cette thèque peut être très rigide et d'une grande complexité architecturale. Elle est fréquemment utilisée comme critère de classification. La classification des dinoflagellés est ardue et source de débats. On observe une telle variation de morphologie, d'habitat, de nutrition (certaines espèces sont autotrophes alors que d'autres sont hétérotrophes) à l'intérieur de cette division que les consensus sont difficiles. Certaines particularités des dinoflagellés telles que la présence de chromosomes condensés en permanence, l'absence d'histone (Rizzo, 1972) et un processus de mitose non-traditionnel (Leadbeater, 1967) en font un groupe à part. Dodge (1965) introduit le terme mésocaryotes afin de classifier les dinoflagellés comme un intermédiaire entre les procaryotes et les eucaryotes. Cependant, de nos jours, les dinoflagellés sont admis en tant qu'eucaryotes étant donné leur grand nombre de caractéristiques typiques des cellules eucaryotes telles que la présence d'un noyau et de tous les organites cytoplasmiques communs (appareil de Golgi, réticulum endoplasmique lisse et granuleux, mitochondrie et chloroplaste).

Le dinoflagellé qui est ici l'objet de notre étude est *Gonyaulax polyedra* (voir figure 1b). C'est un dinoflagellé marin mesurant en moyenne 35 x 45 µm, caractérisé par un noyau en forme de C contenant des chromosomes condensés, une thèque élaborée et la présence d'organites nommés scintillons à l'intérieur desquels se produit la bioluminescence (voir figure 1a).



A



B

Figure 1 : Le dinoflagellé *Gonyaulax polyedra*

A. Certaines espèces de dinoflagellés, dont *G. polyedra*, sont capables de produire de la lumière. Ce phénomène, observable ici dans un flacon de verre, se nomme bioluminescence. Photo : Dr. David S. Morse

B. Observation du dinoflagellé *Gonyaulax polyedra* par microscopie électronique à balayage. On observe ici la complexité architecturale de sa thèque. Grossissement 3000X. Photo : J. Dodge

1.1.2 Particularités du noyau

La plupart des événements cellulaires débutent dans le noyau ou via un message en provenance du noyau porté par une molécule d'acide ribonucléique (ARN). Le cytoplasme et la membrane nucléaire, perforée de nombreux pores d'environ 100 nm de diamètre, observés chez les dinoflagellés sont typiquement eucaryotes mais on retrouve plusieurs aspects nucléaires particuliers.

La forme du noyau des dinoflagellés varie énormément d'une espèce à l'autre. On retrouve entre autres une sphère typique chez *Cryptheconidium cohnii*, une forme pyramidale chez *G. dodgei* et une forme de croissant chez *Gonyaulax polyedra* (Spector, 1984). La quantité d'ADN génomique contenue varie également et est généralement assez considérable. Par exemple on retrouve 200 pg d'ADN génomique par noyau chez *Gonyaulax polyedra* (Tappan, 1980), soit environ 60 fois plus que dans un noyau haploïde humain. De cette abondance d'ADN génomique découle un grand nombre de chromosomes. On observe généralement plus de chromosomes chez les espèces autotrophes que chez les espèces hétérotrophes (Spector, 1984); et on estime à 250 le nombre de chromosomes de *Gonyaulax polyedra* (il y en a seulement 23 dans un gamète humain!) (David Morse, observations personnelles). Ces chromosomes sont condensés en permanence à travers tout le cycle cellulaire (Dodge, 1965) et attachés à l'enveloppe nucléaire (Rizzo, 1987); contrairement à ceux des eucaryotes qui sont condensés exclusivement en phase mitotique et libre dans le noyau.

Cette condensation des chromosomes reste un mystère puisqu'on observe une absence d'histone associée à l'ADN (Rizzo, 1972) et donc de nucléosome. L'absence de nucléosome est confirmée par examen microscopique d'étalement de la chromatine (Rizzo, 1981). La condensation de l'ADN des dinoflagellés pourrait donc être effectuée à l'aide de petites protéines basiques de 12 à 16 kDa dont la présence a été dépistée à l'intérieur du noyau ou avec des ions métalliques tels le Ca^{2+} ou le Mg^{2+} qui neutraliseraient les charges négatives des groupements

phosphates de l'ADN pour permettre sa condensation. De plus, la présence de protéines non-basiques (insolubles dans l'acide) a été dépistée dans le noyau (Rizzo, 1981), mais leur rôle dans la condensation de l'ADN reste inconnu. Les mécanismes permettant la transcription et/ou la réplication d'un ADN condensé n'ont pas encore été élucidés .

Des analyses biochimiques révèlent la présence d'une base azotée modifiée : la 5-hydroxyméthyluracile (HOMeUra) qui remplace 12 à 68 % des thymines chez les différentes espèces de dinoflagellés (Rae et Steele, 1978). Actuellement, les dinoflagellés semblent être les seuls eucaryotes à posséder cette base azotée modifiée. Étant donné que sa présence ne peut être responsable de la structure particulière des chromosomes des dinoflagellés, qu'elle est incompatible avec les principales fonctions des eucaryotes et qu'elle ne fait pas partie d'un système de modification couplé aux enzymes de restrictions chez les procaryotes, Rae et Steele (1978) proposent qu'il s'agit du vestige d'un mécanisme ancien.

1.1.3 Le cycle cellulaire des dinoflagellés

La présence de chromosomes condensés en permanence chez les dinoflagellés entraîne de nombreuses questions sur le mécanisme de réplication de l'ADN, le temps de réplication de l'ADN et la ségrégation des chromosomes lors de la mitose. Chez les procaryotes, la réplication de l'ADN est généralement un phénomène continu qui se produit sur un chromosome simple, non associé à des histones. Il est cependant intéressant de souligner qu'il semble exister quelques exceptions à cette règle. Par exemple, la synthèse d'ADN est observée pendant seulement 80 % du cycle cellulaire chez *M. faecalis* (Zusman et Rosenberg, 1970). Au contraire, chez les eucaryotes, un cycle cellulaire bien défini est universellement observé. La phase de réplication de l'ADN, nommée phase S, est intercalée entre deux phases de croissance et d'accomplissement des fonctions cellulaires, les phases G₁ et G₂ (Howard et Pelc, 1953). Le cycle cellulaire est bouclé par la phase de la mitose qui est le processus par lequel la cellule mère

distribue au moins une copie de son génome à chacune des cellules filles. Le cycle cellulaire comporte donc 4 étapes qui vont comme suit : G₁, S, G₂, M .

L'étude du cycle cellulaire des dinoflagellés fut et est encore, source de débats. Les premières données sur ce sujet sont publiées par Dodge (1966) qui observe une synthèse d'ADN continue chez *Prorocentrum* grâce à l'incorporation généralisée de nucléotides marqués par les cellules en culture. En 1971, Franker se penche sur l'étude du cycle cellulaire d'un autre dinoflagellé : *Anthopleura elegantissima*. Cette espèce est choisie parce que le synchronisme des cellules en culture est facilité par la présence d'une phase de cellules enkystées : les cellules non enkystées peuvent être éliminées de la culture en provoquant une lyse cellulaire par ajout d'eau distillée. Le profil de la synthèse d'ADN s'observe grâce à l'incorporation de C¹⁴ thymidine sur les cellules synchronisées. Il démontre la présence d'une phase S bien définie. Des études du même type ont été poursuivies par Franker et al. (1973) sur un autre dinoflagellé, *Cryptothecodinium cohnii*, démontrant également la présence d'une phase S bien distincte, couvrant environ 25% du cycle cellulaire.

En 1977, Filfilan et Sigee se penchent sur le cycle cellulaire de *P. micans* . Leurs observations du rythme d'incorporation de nucléotides marqués jumelées à des observations microscopiques du rythme de division cellulaire leur permettent de conclure que la synthèse d'ADN est principalement continue, avec un niveau de synthèse plus grand à l'interphase chez *P. micans* .

Plus récemment, des études menées par Homma et Hastings (1989) ont permis de démontrer que la synthèse d'ADN est clairement limitée à une phase S chez *Gonyaulax polyedra*. La synchronisation des cellules en culture est effectuée grâce à deux filtrations différentielles (Hastings et al., 1989). Selon ces auteurs, la majorité des dinoflagellés posséderaient un cycle cellulaire semblable à celui des eucaryotes typiques, incluant donc une phase S de réplication de l'ADN. Les divergences de résultats obtenus dans les études passées seraient probablement dues au manque de synchronisme des cellules en culture ainsi qu'à une imprécision des techniques utilisées.

Le cycle cellulaire des dinoflagellés est également remarquable par la présence d'une phase M (mitose) différente de celle observée chez les eucaryotes. En fait, la mitose des dinoflagellés est parfois même désignée par le terme dinomitose (Chatton, 1920). La présence de chromosomes condensés en permanence et attachés à l'enveloppe nucléaire ainsi que l'absence de destruction de cette enveloppe nucléaire et du nucléole (Rae, 1970) divergent de la prophase traditionnellement observée. De plus, la ségrégation des chromosomes est orientée à l'aide d'un fuseau mitotique extranucléaire, composé de microtubules parallèles, qui invaginent le noyau afin d'établir la polarité de la division des chromosomes. Les microtubules ne semblent pas pénétrer le noyau (Kubai et Ris, 1969) et leur centre d'origine est incertain puisque aucun centre d'organisation des microtubules (MTOC) n'est observé (Triemer et Fritz, 1984). En fait, la division des chromosomes semble s'effectuer grâce à une sorte de flux membranaire de l'enveloppe nucléaire, entraînant ainsi les chromosomes fixés aux deux pôles déterminés par le fuseau mitotique (Oakley et Dodge, 1974). Ce qui diverge grandement des métaphase, anaphase et télophase typiques des eucaryotes.

1.1.4 Un organe bien particulier : le scintillon

Plusieurs organismes, terrestres comme aquatiques, ont la capacité de produire de la lumière . Ce phénomène se nomme bioluminescence . Les fonctions de la bioluminescence sont variées : communication intraspécifique, mimétisme, camouflage ou éclairage (Herring, 1992). Plusieurs dinoflagellés marins sont bioluminescents. Par exemple les *Peridinium*, *Gonyaulax* et *Pyrocystis* peuvent tous produire de la lumière (Spector, 1984). L'étude de *Gonyaulax polyedra* a permis d'élucider quelques composantes et mécanismes de cette bioluminescence. La réaction de la bioluminescence repose principalement sur trois composantes : le substrat est un tétrapyrrole nommé luciférine (Nakamura et al., 1989), l'enzyme est une luciférase, un facteur qui lie le substrat nommé " luciferin-binding-protein " ou la protéine LBP (Fogel et Hastings, 1971). Cette bioluminescence se produit à l'intérieur de petits organites appelés scintillons (Desa et al., 1963). Ces

scintillons, situés à la périphérie mais à proximité de la membrane vacuolaire, confinent les luciférine, luciférase et LBP à l'intérieur de ses limites (Nicolas et al., 1991). En gros, la bioluminescence se produit en quelques étapes simples. La luciférine est initialement stabilisée par son lien avec la protéine LBP. L'arrivée d'un stimulus mécanique extérieur entraîne le déclenchement d'un potentiel d'action dans la membrane vacuolaire qui résulte en une entrée d'ions H^+ dans le cytoplasme de la région des scintillons. Cette baisse du pH permet un changement de conformation de LBP et donc la libération de la luciférine. La luciférine est alors oxydée par la luciférase, produisant ainsi de la lumière.

Le nombre de scintillons, tout comme la bioluminescence, varie selon une oscillation circadienne atteignant un maximum en début de nuit pour ensuite chuter drastiquement à l'aurore (Hasting et Sweeney, 1958 ; Fritz et al., 1990). Chez *Gonyaulax polyedra*, cette bioluminescence est facile à observer puisque les algues produisent des flashes lumineux spontanés et une lueur de faible intensité continue tout au long de la nuit (Hastings et Sweeney, 1958). Les protéines LBP et la luciférase sont dix fois plus abondantes la nuit, leur régulation implique donc une dégradation et synthèse de protéines (Sulzman et al., 1978 ; Dunlap et Hastings, 1981 ; Morse et al., 1989a). Ceci est en accord avec la destruction et synthèse *de novo* journalière des scintillons qui sont également dix fois plus abondants la nuit (Fritz et al., 1990 ; Morse et al., 1989a).

1.2 Les dinoflagellés comme modèle pour l'étude des rythmes

1.2.1 Généralités sur les rythmes

La plupart des organismes terrestres sont continuellement soumis à une alternance entre le jour et la nuit. Il n'est donc pas étonnant que certains organismes démontrent des rythmes biologiques avec une période de 24 heures qui aident à synchroniser leurs comportements avec les conditions extérieures. Un rythme est tout simplement la répétition d'un phénomène à intervalles réguliers. Il est

caractérisé par une période (temps requis pour effectuer un cycle complet), une phase (sa position dans le cycle) et une amplitude (son ampleur). Ces rythmes biologiques sont généralement sous le contrôle d'une horloge interne qui cause les cycles retrouvés dans les processus biologiques ou biochimiques. Deux exemples bien connus de rythmes circadiens chez l'humain sont le cycle éveil/sommeil et les variations de température corporelle. Le rôle de ces horloges internes est de procurer un avantage évolutif significatif à l'organisme qui peut ainsi coordonner ses processus physiologiques de façon à anticiper les changements environnementaux. Les rythmes circadiens possèdent plusieurs caractéristiques (Edmunds, 1988) :

- Le rythme est endogène (il opère sans indice environnemental) et possède une période précise (d'environ 24 heures) dans des conditions constantes .
- Le rythme repose sur une base génétique et possède une période précise . Des mutations dans la période sont hérissables entre les générations .
- La phase d'un rythme peut être avancée ou retardée par des signaux externes (généralement la lumière) . Des signaux réguliers et répétitifs vont permettre de synchroniser le rythme à leur période .
- La période du rythme est peu affectée par les variations de températures .

La troisième caractéristique nous permet de bien comprendre pourquoi les rythmes circadiens observés en nature ont une période de 24 heures. Malgré le fait qu'en conditions environnementales constantes, un rythme circadien possède habituellement une période légèrement différente de 24 heures. Cette période, propre à l'espèce et même à l'individu, se nomme " free-running period " (Moore-Ede et al., 1982).

Il semble important de mentionner qu'afin de faciliter l'étude des rythmes circadiens, le temps de la période est exprimé en temps circadien (CT). Conventionnellement, le temps circadien 0 (CT 0 ou CT 24) représente le début de la phase lumineuse. L'étude du fonctionnement des horloges endogènes est essentielle à l'avancement des connaissances des mécanismes responsables des troubles de fatigue, désordres du sommeil et tout simplement des symptômes du décalage horaire que nous ressentons lors des vols à long cours.

1.2.2 Les rythmes observables chez *Gonyaulax polyedra*

De nombreuses espèces procaryotes et eucaryotes, animales et végétales, manifestent des comportements que l'on peut qualifier de rythmes circadiens. Chez le dinoflagellé *Gonyaulax polyedra*, de nombreuses activités sont contrôlées par une horloge endogène. Étant donné que cet organisme est unicellulaire, les actions de l'horloge se doivent nécessairement d'être au niveau moléculaire et/ou biochimique. Sa simplicité en fait donc un excellent modèle pour l'étude des mécanismes de liaison des signaux environnementaux à l'horloge. De plus, les rythmes circadiens de *Gonyaulax polyedra* sont facilement observables. Parmi ceux-ci mentionnons un rythme de migration et d'activité photosynthétique. Durant la journée, les algues sont retrouvées à la surface de l'océan (Eppley et al., 1968) et l'activité photosynthétique (mesurée par la fixation du CO₂) est maximale (Hastings et al., 1961). Lors de la nuit, ces mêmes algues se retrouvent à une profondeur de 10 à 15 mètres.

Un rythme observé est celui de la bioluminescence (voir figure 2a,b). La bioluminescence nécessite la présence du substrat luciférine (Nakamura et al., 1989), de l'enzyme luciférase (Dunlap et Hastings, 1981) et de la protéine LBP qui empêche l'auto-oxydation du substrat, toutes enfermées à l'intérieur des scintillons (voir ci-dessus) (Morse et al., 1989a). Les différentes composantes du mécanisme de la bioluminescence oscillent en abondance (dix fois plus abondantes la nuit) permettant ainsi d'observer la bioluminescence lors de la nuit (Fritz et al.,

1990 ; Morse et al., 1989b). Le niveau de protéine LBP augmente drastiquement en début de phase obscure et cette augmentation est due à une activation de la synthèse protéique puisque le niveau d'ARN messenger LBP reste toujours élevé et constant sur un cycle de 24 heures (Morse et al., 1989b). Le niveau maximal de protéine LBP correspond à celui de la bioluminescence. La protéine LBP est détectée en abondance pour une durée d'environ six heures puis son niveau chute rapidement à l'aurore, suggérant la présence d'un mécanisme de dégradation protéique. On soupçonne donc que l'horloge agit à deux niveaux en ce qui concerne la présence de LBP. Un premier mécanisme viendrait activer la traduction de l'ARN messenger LBP en début de nuit ; et un second viendrait effectuer une modification post-traductionnelle afin d'accélérer la dégradation de LBP à l'aurore.

Le mécanisme d'activation de la traduction de LBP est l'objet de notre étude. En général, l'initiation de la traduction est influencée par cinq grands aspects de la structure de l'ARNm : (1) la présence du m7Gcap, (2) les séquences entourant le codon AUG, (3) la position du codon AUG, (4) la longueur de la région codante et (5) la structure secondaire de la région entourant le codon AUG (Kozak, 1991). De plus, la présence d'ouverture de cadre de lecture (ORF) additionnels en amont de la région codante peut affecter la traduction (Kozak, 1984). L'ARNm LBP possède quatre AUG en 5' dont un correspondant à un ORF situé en région 5' non-codante (5' UTR) qui affecte l'initiation de la traduction des AUGs en aval. En fait, on observe la production de deux principales formes de LBPs : LBP-A (synthétisé à partir des AUG 2 et 3) et LBP-B (synthétisé à partir d'AUG 4) (Mittag et al., 1997). Seule la première forme (LBP-A) oscille selon le rythme circadien mentionné plus haut.

Une séquence spécifique (UG répétés) située dans la région 3' non-traduite (3'UTR) de l'ARNm LBP semble être impliquée dans la régulation circadienne de la traduction. En effet, la présence d'une protéine capable de se lier à cette région a été dépistée dans les extraits de cellules de jour (et non dans celles de nuit) (Mittag et al., 1994). Cette protéine semblerait donc se lier à l'ARNm LBP de façon circadienne et elle est nommée CCTR pour " circadian controlled

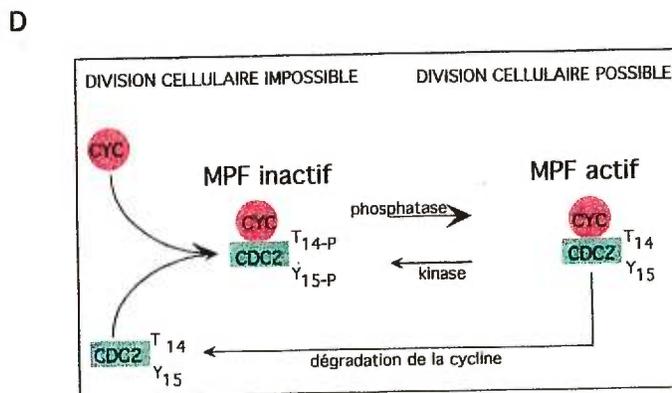
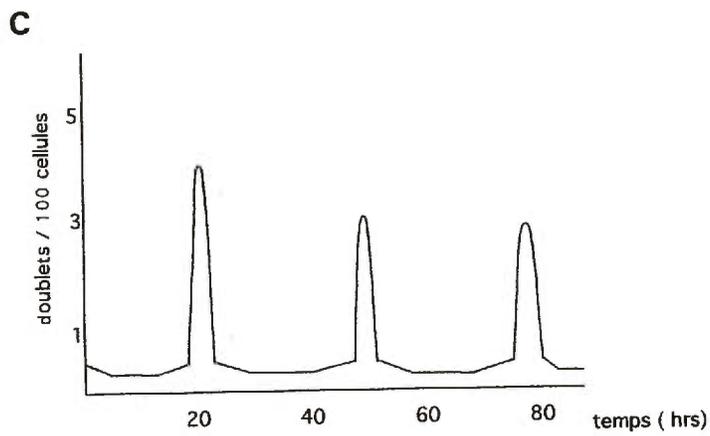
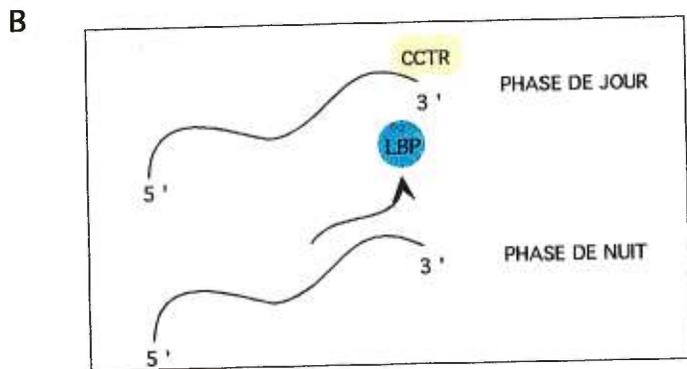
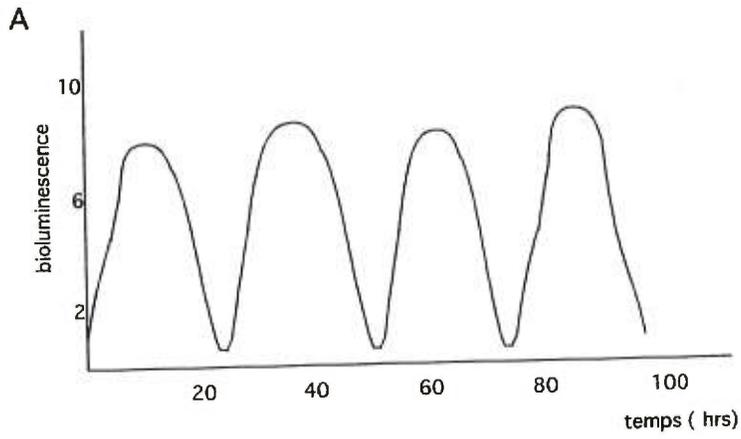


FIGURE 2

translational regulator " (Mittag et al., 1994). L'activité de liaison à l'ARNm LBP de la protéine CCTR est élevée lors de la phase lumineuse et décroît au début de la nuit, ce qui correspond à l'activation de la production de LBP-A, indiquant ainsi que la CCTR serait un répresseur de la traduction. Le mécanisme de cette répression reste encore mystérieux. La possibilité de formation d'un ARNm circulaire *in vivo* pourrait expliquer comment un inhibiteur fixé en 3' viendrait stopper la traduction initiée en 5' de l'ARNm.

Le second rythme qui est ici un objet de notre étude est celui de la division cellulaire (voir figure 2c,d). En effet, chez *Gonyaulax polyedra*, la mitose est limitée au début de la phase lumineuse (Nicolas et al., 1991). Le temps de génération typique de ces dinoflagellés est de 48 à 72 heures, mais même lorsque la croissance est terminée, la division n'est possible qu'à l'aurore (Homma et Hastings, 1989a et 1989b). Chez la majorité des eucaryotes, l'oscillation du cycle cellulaire est basée sur l'activité de deux protéines, une kinase $p34^{cdc2}$ et une cycline (Dunphy et al., 1988). Ces deux protéines s'associent pour former le complexe MPF " M-phase promoting factor " capable d'induire l'entrée en phase M dans un oocyte immature bloqué en phase G_2 (Dunphy et al., 1988). De plus, chez la levure fissipare (*Schizosaccharomyces pombe*), des expériences avec un mutant thermosensible du gène *cdc2* ont démontré que la présence de la protéine kinase $p34^{cdc2}$ est essentielle aux transitions entre les phases G_1 -S et G_2 -M. Ces deux protéines sont donc essentielles au bon déroulement du cycle cellulaire. La cycline s'accumule progressivement et continuellement au cours du cycle cellulaire. Lorsqu'elle atteint une concentration suffisante, la cycline active la $p34^{cdc2}$. La kinase activée catalyse ensuite plusieurs phosphorylations essentielles à la réorganisation cellulaire en prévision de la mitose, dont celle responsable de la dégradation de la cycline qui se produit en anaphase. La dégradation de la cycline entraîne la dissociation du MPF et donc la monomérisation de la $p34^{cdc2}$ (Evans et al., 1983). La forme inactive de la kinase $p34^{cdc2}$ comporte deux groupements phosphate jumelés à deux de ses acides aminés. La déphosphorylation de la kinase est nécessaire afin de permettre son activation (Murray, 1992). Une cellule de taille trop petite ou une algue *Gonyaulax polyedra* qui n'est pas à l'aurore pourrait

donc empêcher le déclenchement de la mitose en maintenant la kinase p34^{cdc2} phosphorylée. C'est pourquoi nous porterons une partie de nos études sur la recherche d'une kinase de type p34^{cdc2} dans l'algue *Gonyaulax polyedra* malgré les échecs des recherches traditionnelles déjà effectuées à cet effet (Salois et Morse, 1996).

1.2.3 Les différents rythmes sont contrôlés par différentes horloges chez *Gonyaulax polyedra*

Jusqu'à récemment, comme les différents rythmes observables retrouvés chez un seul organisme avaient tous la même période, on croyait qu'ils étaient tous contrôlés par la même horloge (McMurry et Hastings, 1972). Mais de récentes observations des rythmes circadiens chez certains organismes dans des conditions particulières ont permis de constater des dissociations de phases et de périodes.

Chez *Gonyaulax polyedra*, les rythmes circadiens des éclairs lumineux et de la production d'une plus faible lueur étaient traditionnellement considérés comme simultanés et donc jumelés ensemble en un seul rythme de la bioluminescence (Hastings et Sweeney, 1957). Ce rythme était donc considéré comme étant contrôlé par une seule horloge (Morse et al., 1990). Cette horloge était reconnue comme le seul et unique oscillateur, contrôlant également les rythmes de photosynthèse (Knoetzel et Rensing, 1990) de division cellulaire (Sweeney et Hastings, 1958) et de migration (Roennenberg et al., 1989).

Cependant, en 1992, Heyde et al. obtiennent des résultats révélant des périodes différentes pour la production de flashes lumineux et d'une lueur lors de l'exposition à différentes conditions extérieures. Ces résultats portent à croire qu'il existerait plusieurs oscillateurs capables de se dissocier en conditions particulières chez *G. polyedra*. Cette hypothèse est supportée par de plus amples résultats rapportés par Roennenberg et Morse (1993 et 1994). Sous certaines conditions particulières (par exemple de la lumière rouge de faible intensité), les rythmes de migration et de bioluminescence sont dissociés, démontrant ainsi la présence de deux oscillateurs différents dans une seule et même cellule. Le nombre

d'oscillateurs différents retrouvés chez *Gonyaulax polyedra* reste encore un sujet d'étude ; mais la possibilité de dénombrer un oscillateur différent pour chaque rythme observé oriente les hypothèses de recherches sur les horloges biologiques.

1.3 Les horloges circadiennes : Étude comparée des mécanismes moléculaires .

Les caractéristiques des différents rythmes nous suggèrent un modèle composé d'horloge centrale qui reçoit l'information des cycles de lumière extérieure (" input ") et règle la phase d'un (ou de plusieurs) rythme observable (" output ") (voir figure 3). Brièvement, les informations extérieures sont acheminées à l'horloge par un ensemble de voies physiologiques et/ou biochimiques. L'oscillateur circadien impose alors un rythme en coordination avec les informations reçues. Ce rythme est traduit par des comportements cellulaires et/ou physiologiques observables liés à l'horloge par une gamme de processus biochimiques. L'étude des horloges a adopté une approche génétique qui nous a permis d'identifier des gènes dans lesquels des mutations changent la période des rythmes chez *Neurospora crassa*, *Drosophila* et chez la souris.

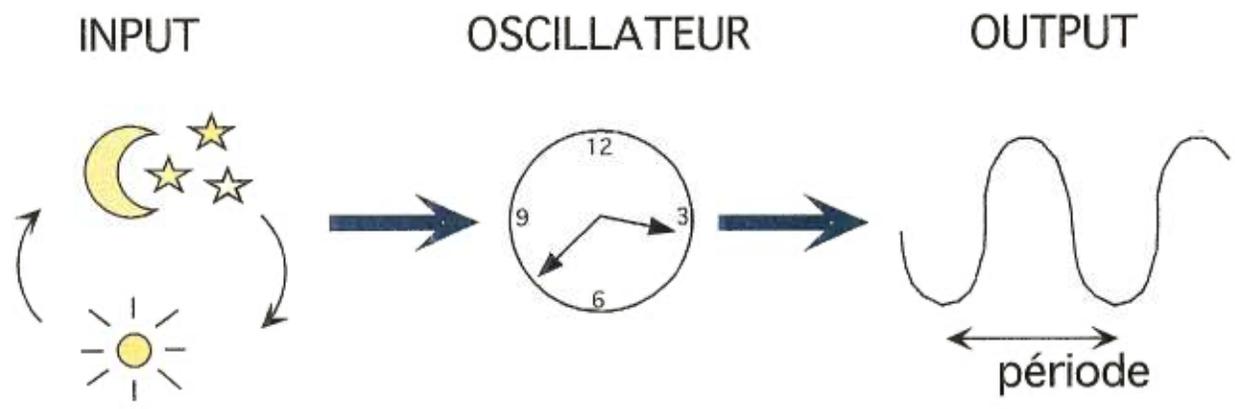


FIGURE 3: Les composantes d'une horloge circadienne .

Les horloges circadiennes sont composées de voies de transduction des signaux externes (input pathway) ainsi que d'un oscillateur central interprétant ces signaux externes et contrôlant le rythme observable par le biais de voies moléculaires et biochimiques (output pathway) .

Tableau I : Les gènes impliqués dans la génération des rythmes

locus	organisme	allèle mutant	particularité
frequency	<i>Neurospora</i>	frq ¹ → frq ⁹	
white collar	<i>Neurospora</i>	wc-1 et wc-2	GATA*-PAS*
period	<i>Drosophila</i>	per ^s , per ^L et per ⁰¹	PAS
timeless	<i>Drosophila</i>	tim	PAS
clock	<i>Drosophila</i>	clk	bHLH*-PAS
cycle	<i>Drosophila</i>	cyc	bHLH-PAS
clock	Souris	clk	bHLH-PAS

* GATA : domaine similaire aux facteurs de transcription

* bHLH : facteur de transcription

* PAS : domaine de liaison des protéines

1.3.1 L'analyse génétique de la rythmicité

Afin de faciliter l'étude moléculaire des horloges biologiques, de nombreux mutants ont été caractérisés chez *Neurospora crassa*, *Drosophila* et les souris (voir tableau I).

Par exemple, chez la moisissure *Neurospora crassa* il y a un rythme persistant d'une période d'environ 24 heures dans la formation des spores . Des mutations dans le gène *frequency* (*frq*) peuvent entraîner soit un changement de la période du rythme ou carrément l'élimination du rythme (Loros et al., 1989). Ainsi, les mutants de la souche *frq9* sont arythmiques et ce phénotype est récessif. La protéine FRQ semble être impliquée dans la génération du rythme car le mutant de la souche *frq9* contient un codon d'arrêt de la traduction au milieu du gène, ce qui produit une FRQ tronquée et non-fonctionnelle. Par contre, les mutants de la souche *frq7* présentent une période de la formation des conidies de 29 heures. Dans le même ordre d'idées, la mutation de *frq7* formerait une protéine avec une perte partielle de sa fonction. Une plus grande accumulation de message serait donc nécessaire pour obtenir le niveau d'activité protéique entraînant la rétro-inhibition, et la période est alors allongée. Cette mutation est à dominance incomplète, c'est à dire que son effet est additif. Par exemple, un hétérozygote *frq+/frq7* aurait une période intermédiaire entre un homozygote *frq7/frq7* et le type sauvage.

Chez la drosophile le gène *period* (*per*) est le plus étudié. Les mutants *per* ont une altération de la période dans le rythme de la locomotion. La protéine PER possède deux domaines PAS (Per-ARNT-SIM) responsables pour des interactions protéines-protéines et elle entre normalement dans le noyau de la cellule lors de son cycle. Une mutation située à la position 243 (dans le domaine PAS) de *per* entraîne une augmentation de la période (souche *perL*, période de 27 à 33 heures). Une hybridation *in situ* démontre que la localisation nucléaire de PER est retardée chez *perL* (Curtin et al., 1995).

Chez la souche *perS*, (mutation à la position 589 de PER) une réduction de la période est observée. Curieusement, les résultats de l'hybridation *in situ* nous permettent d'observer une localisation nucléaire de PERS au même temps que chez

le type sauvage. Cette entrée nucléaire de PER retardée chez *perL* et non-devancée chez *perS* semble indiquer que la localisation nucléaire n'est pas irrévocablement liée à la longueur de la période. L'entrée nucléaire serait donc plus une conséquence de l'activité de PER qu'un déterminant de la période. Le mécanisme entraînant une période de 19 heures malgré la localisation nucléaire à ZT 18 et 19 chez *perS* reste cependant un mystère à découvrir. Ces deux types de mutations (*perL* et *perS*) sont à dominance incomplète, les hétérozygotes possèdent donc des périodes intermédiaires.

Une autre mutation connue chez la drosophile est celle du gène *timeless* (*tim*). La perte de fonction de *tim* entraîne l'arythmicité (de la même façon qu'une exposition à la lumière, voir en section 1.3.4); TIM se lie à PER via les domaines PAS. Cette mutation est sans doute de type récessive (comme *frq9*), c'est-à-dire qu'un seul allèle de type sauvage suffit sûrement à la production de TIM et ainsi au bon fonctionnement de l'oscillateur.

Récemment, de nouveaux mutants ont été caractérisés chez la drosophile. Il s'agit des gènes *clock* (*clk*) et *cycle* (*cyc*). Les mutations de *clk* et de *cyc* entraînent la formation d'une protéine tronquée (Allada et al., 1998) (Rutila et al., 1998). Ces deux mutants sont de type semi-dominant. L'homozygote est arythmique et possède une faible transcription des gènes *per* et *tim* suggérant ainsi que CLK et/ ou CYC agissent comme des activateurs. L'hétérozygote *cyc*^{0/+} possède une altération de la période mais conserve une transcription massive et rythmique des gènes *per* et *tim*. L'allèle de type sauvage devra donc produire suffisamment d'ARNm *cyc* pour assurer le bon fonctionnement de l'oscillateur.

Chez la souris, des mutations du gène *clock* (*clk*) affectent le rythme d'activités. Chez l'homozygote, on observe une perte de la rythmicité et chez l'hétérozygote un allongement de la période (Vitaterna et al., 1994). Ces protéines CLK possèdent des domaines PAS, une région bHLH et une région riche en glutamine mais leur fonction est peu connue outre les parallèles possibles avec *clk* de la drosophile.

1.3.2 L'horloge génère un cycle par une boucle de rétroaction négative

Chez *Drosophila*, *Neurospora* et les mammifères, les gènes de l'horloge semblent être impliqués dans la formation d'une boucle de rétroaction négative. Chez *Neurospora*, le gène *frq* code pour une composante principale (la protéine FRQ) de sa propre rétro-inhibition au niveau de la transcription. Cela résulte en une oscillation journalière du niveau d'ARNm *frq* dont l'abondance maximale se situe à l'aurore. La protéine FRQ fluctue donc également avec un décalage par rapport à l'ARNm *frq* . L'inhibition par FRQ de sa propre transcription, qui constitue une boucle de rétroaction négative, a été démontrée par une surexpression de *frq* à l'aide d'un promoteur inductible par la quinine (construction qa-2pFRQ). L'induction de qa-2pFRQ avec la quinine entraîne une forte diminution de la quantité d'ARNm endogène (Aronson et al., 1994). FRQ semble donc être le responsable de sa propre rétro-inhibition au niveau de la transcription. Il est important de noter que si l'ARNm FRQ est maintenu à un niveau élevé grâce à ce promoteur inductible, les rythmes restent bloqués à un stade qui correspond à l'abondance maximale de l'ARNm *frq* (i.e. l'aurore).

Un problème à résoudre était comment cette boucle de rétroaction produisait une période de 24 heures, puisque les processus de transcription et de traduction sont normalement complétés dans un délai moins élevé. Une réponse à ce problème repose sur l'observation des modifications post-traductionnelles (phosphorylation à plusieurs sites) de FRQ. En effet, il est possible que seule la forme hautement phosphorylée de FRQ puisse pénétrer le noyau , comme c'est le cas pour la protéine PER (voir plus bas) (Edery et al., 1994a et 1994 b). Un délai de 12 heures est nécessaire pour phosphoryler complètement FRQ, ce qui contribue à la période de 24 heures. Il est à noter qu'un délai entre le début de la transcription et la formation de l'inhibiteur de la transcription est nécessaire et prédit par des considérations théoriques sur les systèmes oscillatoires. En effet, ce délai permet de conserver le déséquilibre nécessaire à la persistance de l'oscillation.

En plus d'une étape de rétroaction négative, un oscillateur se doit de contenir des éléments maintenant l'intensité de l'oscillation. Chez *Neurospora*

crassa, deux gènes : *white collar* 1 et 2 (*wc* 1 et *wc* 2) semblent assumer ce rôle car ils sont essentiels à une bonne expression du gène *frq* (Crosthwaite et al., 1997). Ces deux gènes agissent comme des éléments associés à l'horloge car leur présence permet de maintenir la production de FRQ suffisamment élevée pour conserver l'effet de la boucle de rétroaction négative. L'oscillation de *frq* est donc maintenue non seulement par la boucle de rétroaction négative mais également par des éléments activateurs agissant directement ou indirectement sur cette boucle.

Chez *Drosophila melanogaster* , les rythmes circadiens de la locomotion et de l'éclosion sont affectés par des mutations des gènes *per* et *tim*. Le clonage et l'analyse de l'expression de *per* ont permis de constater que l'ARNm *per* oscille avec une abondance maximale à la tombée de la nuit (à l'opposé de *frq*) (Zeng et al., 1994). La surexpression de *per* à l'aide d'un promoteur inductible par choc thermique entraîne la diminution de l'ARNm endogène ; suggérant ainsi que *per* encode son propre facteur d'inhibition transcriptionnel. La protéine PER oscille en masse et en abondance avec la même période mais une phase retardée d'environ 4 heures. La variation de masse est due principalement à une phosphorylation graduelle de la protéine PER (Ederly et al., 1994a). La phosphorylation de PER semble constituer un délai permettant d'éviter la diminution progressive de l'oscillation. Ce délai est nécessaire afin de conserver un déséquilibre entraînant la persistance de l'oscillation tel que mentionné pour l'oscillation de FRQ.

PER contient un domaine de dimérisation, le domaine PAS, qui permet la formation d'homodimères (PER-PER) et d'hétérodimères avec TIM (Huang, 1993 ; Seghal et al., 1994 ; Vosshall, 1994). Le gène *tim* est essentiel pour le fonctionnement de l'horloge. De plus, l'ARNm *tim* et TIM démontrent des rythmes d'oscillation semblables à l'ARNm *per* et PER respectivement. PER et TIM s'accumulent graduellement dans le cytoplasme de la cellule avec la même phase, période et amplitude et s'associent grâce au domaine PAS. L'association de PER à TIM permettrait de stabiliser et ainsi d'accumuler et de phosphoryler PER. En effet, l'abondance et la phosphorylation de PER sont faibles dans *timeless* (*tim* 0) (Gekakis, 1995 ; Zeng et al., 1994 ; Seghal et al., 1995). La formation de l'hétérodimère permettrait également le transport du complexe vers l'intérieur du

noyau (Vosshall, 1994). La localisation intranucléaire du dimère est suivie d'une baisse de la transcription de *per* et *tim*, suggérant une rétroinhibition négative au niveau de la transcription. Le mécanisme de cette rétroinhibition négative semble s'éclaircir depuis peu. Les récentes découvertes de Allada et al. (1998) nous ont fait connaître de nouveaux facteurs qui semblent éligibles en tant que lien entre l'hétérodimère PER-TIM et l'action directe sur les gènes *per* et *tim*. En effet, un gène homologue au gène *clock* des mammifères (*mclock*) a été isolé chez la drosophile (*dclock*). La présence de phénotype arythmique et d'une faible expression de *per* et *tim* chez le mutant *dclock* ont permis d'isoler ce gène situé sur le bras gauche du chromosome 3 à l'aide de marqueurs morphologiques. La protéine dCLOCK (1023 acides aminés) et la protéine mCLOCK (855 acides aminés) sont toutes deux riches en glutamine, possèdent une queue de polyglutamine et une région bHLH-PAS.

La mutation semi-dominante du gène *dclock* (*dclk*) est caractérisée par un codon stop prématuré, entraînant ainsi la formation d'une protéine tronquée. Cette protéine dCLOCK tronquée possède toujours sa région de fixation à l'ADN (la région bHLH-PAS) mais est dépourvue de sa queue de polyglutamine (région reconnue comme facteur d'activation de la transcription). Considérant que les gènes *per* et *tim* possèdent tous deux une boîte E (CACGTG) dans la région promotrice et que les boîtes E sont des sites connus pour l'activation et la transcription par des hétérodimères bHLH-PAS, nous pouvons soupçonner une activation de la transcription de *per* et *tim* par un hétérodimère bHLH-PAS comportant une protéine dCLOCK fonctionnelle. L'arrêt de l'expression cyclique de *per* et *tim* chez le mutant homozygote *dclk* et l'activation *in vitro* de l'expression de la luciférase dans une construction boîte E- luciférase par dCLOCK semblent indiquer que dCLOCK agit comme un activateur de transcription dont l'activité est inhibée par PER-TIM.

De plus, l'isolation de CYCLE (une deuxième protéine bHLH-PAS) dont les mutants homozygotes (*cyc-cyc*) sont arythmiques et présentent une faible production constitutive d'ARNm *per* et *tim* permet d'aller encore plus loin (Hall et al. 98). Les protéines CYC et CLK forment un hétérodimère qui se fixe sur la boîte E située dans les régions promotrices des gènes *per* et *tim*, activant ainsi la transcription de ces gènes via la queue de polyglutamine de CLK. Il est donc aisé de soupçonner que la forme nucléaire de l'hétérodimère

PER-TIM viendrait inhiber l'action de l'hétérodimère CYC-CLK afin de permettre un feedback négatif et donc une expression cyclique de *per* et *tim*. L'oscillateur inclurait donc à la fois des régulateurs négatifs (tels que PER-TIM ou FRQ) et des activateurs (tels que CLK-CYC ou WC1-WC2) ; l'interaction entre ces deux types de régulateurs reste cependant à démontrer.

1.3.3 Un zeitgeber essentiel : la lumière .

Tel que mentionné plus haut, en conditions environnementales constantes, un rythme circadien se manifeste avec une période légèrement différente de 24 heures, la " free-running period " qui varie selon les espèces, les individus et les conditions de température et luminosité. Cette " free-running period " est entraînée à une durée précise de 24 heures par un zeitgeber (donneur de temps) essentiel : la lumière. Cet effet de la lumière nécessite la présence de photorécepteurs ainsi qu'une transduction du signal qui doit être ensuite acheminé jusqu'à l'oscillateur (input).

Chez *Neurospora crassa* , on a d'abord observé qu'un flash lumineux entraînait une augmentation de la transcription de *frq*. En effet, le niveau d'ARNm *frq* augmente rapidement et massivement (proportionnellement à l'intensité de la lumière) suite à un flash lumineux à n'importe quel temps circadien de culture en obscurité constante (Crosthwaite et al.,1995). Donc un flash lumineux entraîne une avance de phase de l'oscillation normale de l'ARNm *frq* dans la période correspondant à l'augmentation naturelle du niveau d'ARNm *frq* et un délai de l'oscillation de l'ARNm *frq* dans la période correspondant à la diminution naturelle du niveau d'ARNm *frq*. Ces résultats expliquent de façon exemplaire les conséquences physiologiques d'un flash lumineux sur le rythme de la formation des conidies.

Une exposition constante à la lumière entraîne un arrêt du cycle de l'ARNm *frq* (comme dans le mutant *frq9*). Un retour à l'obscurité permet de ré-enclencher l'oscillation. Le mécanisme de cette augmentation de l'ARNm *frq* est différent de l'inhibition de la transcription par FRQ car la surexpression de l'ARNm *frq* à l'aide d'un promoteur inductible a permis de démontrer que l'induction de la transcription de *frq* par la lumière surpasse la rétroaction négative. En effet, un flash lumineux entraîne l'augmentation de l'abondance d'ARNm *frq* endogène malgré l'auto inhibition par qa-2pfrq (Loros et al.,

1989).

Les gènes nécessaires pour la transduction du message lumineux, *wc1* et *wc2*, sont des gènes identifiés comme étant des éléments activateurs dans la boucle de rétroaction mentionnée ci-haut (Crosthwaite et al., 1997). Les mutants *wc-1* et *wc-2* sont incapables de répondre à la lumière. Leur phénotype "aveugle" semble expliquer leur arythmicité apparente, les cultures étant incapables de synchroniser leur oscillateur. Plus précisément, *wc2* semble coder pour une composante de l'horloge impliquée dans la transduction du signal lumineux, favorisant l'induction de la transcription de *frq* et d'autres gènes contrôlés par la lumière. Le gène *wc1* serait plutôt un élément associé à l'horloge dont la présence favorise une production suffisamment élevée de FRQ et est absolument essentielle à l'induction de la transcription de *frq* suite à un flash lumineux. Étant donné que *wc1* et *wc2* possèdent tous deux un domaine PAS il est possible d'envisager la formation d'un hétérodimère (*wc1/wc2*) qui régulerait directement la transcription de *frq*.

Chez la drosophile la lumière entraîne la dégradation rapide de la protéine TIM (Myers et al., 1996). Tel que mentionné plus haut, *tim* est soumis à une boucle de rétro-inhibition entraînant le cycle de l'ARNm *tim* puis celui de la protéine TIM (avec environ quatre heures de retard). TIM s'associe avec PER pour former un hétérodimère et, suite à des modifications post-traductionnelles (la phosphorylation), ce complexe pénètre le noyau. Ce phénomène se déroule dans les cellules photoréceptrices et celles des neurones latéraux. La protéine TIM est nécessaire à la localisation nucléaire de PER et vice-versa (Hunter-Ensor et al., 1996) et l'action de la lumière sur TIM se répercute donc directement sur le cycle. En début de nuit, le flash produisant la dégradation de TIM entraîne un délai puisqu'il y a retard dans l'accumulation de TIM et la formation d'hétérodimères. En fin de nuit, le flash produit une avance de phase puisqu'il accélère le processus naturel de dégradation de TIM. La période correspondant à l'avant-midi et le début de l'après-midi forme une zone d'insensibilité à la lumière, la protéine TIM y étant virtuellement absente. La correspondance entre ces observations moléculaires et les réponses physiologiques observées suite à des flash lumineux (dans le cadre des rythmes circadiens) semble acceptable.

Une exposition constante à la lumière entraîne un arrêt du cycle et une réduction d'abondance et de la phosphorylation de TIM. La protéine PER semble également moins abondante et peu phosphorylée (Marrus et al., 1996).

La lumière semble donc agir au niveau post-traductionnel (contrairement à *Neurospora*) puisque l'abondance et l'oscillation des messages de *per* et *tim* semblent maintenues. Cependant une exposition prolongée à la lumière entraîne une perte de l'oscillation du message de *per* (Qiu et al., 1996). Une alternance L/D avec une noirceur minimale de 6 à 8 heures semble nécessaire pour maintenir le cycle de l'ARNm *per*. Nous pouvons donc supposer qu'une production et accumulation de TIM dans l'obscurité sont nécessaires au fonctionnement de l'oscillateur. La présence de TIM permettrait la stabilisation, l'accumulation et la phosphorylation de PER, ce qui entraînerait la localisation nucléaire ainsi que l'inhibition de la transcription par une action sur le complexe CLK-CYC.

1.3.4 La compensation aux variations de température .

Afin de conserver sa période, l'oscillateur est peu sensible aux variations de température. En général, pour une variation de 10°C, la vitesse des réactions enzymatiques composant l'horloge n'augmente pas ($Q_{10} = 1$).

Chez *Neurospora crassa* deux sites de début de la traduction de l'ARNm entraînent la formation de deux protéines FRQ de taille différente (l'initiation de la traduction peut se faire au premier ou au troisième codon AUG) (Garceau et al., 1997). En absence des deux formes de FRQ aucun rythme n'est observé. Par contre, avec les deux FRQ, le rythme persiste avec une période relativement constante de 18 à 30°C. Une mutation éliminant le premier AUG entraîne une perte de rythmicité à température élevée alors qu'une mutation dans le troisième AUG entraîne une perte de rythmicité à température basse. L'initiation de la traduction de *frq* varie selon la température, la plus grosse protéine FRQ étant stable aux hautes températures alors que la petite forme est stable aux températures plus basses (Liu et al., 1997). Chacune des deux formes est suffisante pour engendrer une oscillation à température moyenne et leur action combinée permet une oscillation de 18 à 30°C. Le mécanisme d'initiation de traduction différentiel selon la température reste cependant à éclaircir, mais il pourrait s'agir d'une conformation particulière qu'adopterait l'ARNm selon une température donnée, favorisant l'exposition de l'un ou l'autre des codons start.

Les grandes variations de température semblent être de robustes *zeitgebers* chez *Neurospora*. Une température de 4°C amène puis stoppe l'horloge à CT 0 et semble donc détruire toute trace de *frq* et/ou de *FRQ* puis bloquer la production de message. Cette remise de l'horloge permettrait l'arrêt de la sporulation en saison froide mais son mécanisme d'action est encore inconnu.

Chez la drosophile Huang et al. (1995) proposent un autre type de mécanisme pour la compensation de la température. Chez les souches de type sauvage, le rythme conserve sa période de 18 à 30°C. Avec l'augmentation de la température, deux interactions compétitives impliquant le domaine PAS sont accélérées. Ces deux interactions sont la formation d'hétérodimères PER-TIM (ce qui provoquerait une réduction de la période) et un repliement des monomères PER. Ce repliement serait dû à une interaction entre le domaine PAS et un domaine en C-terminal. Une augmentation des interactions PAS-domaine C avec la température permettrait de maintenir des niveaux normaux d'hétérodimères PER-TIM (et donc la période) malgré les variations de température. Le mutant *perL* présente une interaction du domaine PAS avec le domaine C-terminal plus accentuée que chez le type sauvage ; *perL* est donc plus sensible aux variations de températures que le type sauvage.

1.3.5 Régulation des "outputs"

À chaque horloge est associé un ensemble de voies de transduction qui acheminent les résultats de l'oscillation afin de réguler des processus biochimiques, cellulaires, physiologiques ou comportementaux observables. Chez *Neurospora crassa* on observe un rythme de formation de conidies. Les produits des gènes *frq* et *wc2* (constituants de l'oscillateur) et du gène *wc1* (élément associé à l'horloge) agissent directement ou indirectement sur les gènes contrôlés par l'horloge. Ces gènes entraînent la formation cyclique des spores. Chez la drosophile on observe un rythme de locomotion et d'éclosion. Les produits des gènes *per* et *tim* agissent de pair sur les gènes contrôlés par l'horloge de façon à engendrer ces rythmes observables. Les mécanismes de régulation des outputs

peuvent être directs où une partie de l'oscillateur agit sur le gène responsable du rythme observable. Par exemple, le dimère CLK-CYC peut agir comme un activateur des gènes possédant une boîte E dans la région promotrice. Malgré cette hypothèse il reste beaucoup de travail à accomplir à ce sujet.

1.4 Aperçu du projet

Le laboratoire du professeur D. Morse porte principalement ses recherches vers la compréhension des mécanismes de liaisons entre les différents rythmes circadiens observés et l'horloge qui les contrôle. Des approches des domaines de la biologie moléculaire et de la biochimie sont généralement utilisées afin d'élucider ces mécanismes.

Le présent projet comporte deux volets : (1) l'isolation d'un homologue de la protéine kinase p34^{cdc2} impliquée dans l'initiation de la division cellulaire et, (2) l'isolation et l'étude d'un ADNc codant pour la protéine CCTR impliquée dans la régulation de la traduction de l'ARNm LBP.

Afin d'atteindre notre premier but nous proposons de transformer des cellules de levure de la souche *cdc28* (mutant thermosensible de p34^{cdc2}) avec une banque d'ADNc de *Gonyaulax polyedra* insérée dans un vecteur d'expression constitutive chez la levure. Les cellules de cette souche ne peuvent se diviser normalement à température restrictive. Le criblage sera donc effectué par une simple recherche de cellules se divisant à cette température. Dans ce cas, nous pouvons suspecter que les cellules poussent parce que la banque permet la production d'un homologue de la p34^{cdc2}.

Le second volet implique également l'utilisation de la banque d'ADNc de *Gonyaulax polyedra*. La banque est transformée dans une souche de levure contenant déjà la construction pCS7. Le plasmide pCS7 est conçu de façon à contenir un gène létal inductible (GSP1) annexé à la séquence régulatrice (UG répétés) de l'ARNm LBP situé en région 3' non-codante (3' UTR). La séquence UG répétés est reconnue pour sa liaison à la protéine inhibitrice CCTR. La liaison

du CCTR à cette séquence entraîne l'arrêt de la traduction *in vivo* chez *Gonyaulax polyedra*. L'isolation d'un fragment d'ADNc codant pour la protéine CCTR sera possible par criblage de cellules en croissance malgré l'induction du gène létal Gsp1, étant donné que la présence du CCTR devrait stopper la traduction de Gsp1.

2. Matériel et Méthodes

2.1 Culture de cellules

Les cellules de *Gonyaulax polyedra* (souche 70, Provasoli-Guillard Culture Center, Booth Bay Harbor, Maine) sont cultivées dans 750 ml de milieu f/2 (Guillard, 1962). Les fioles en verre de 2,8 litres sont lavées, rincées au HCl (2 N) puis à l'eau distillée avant d'être remplies de f/2, autoclavées puis inoculées avec 75 ml d'une culture dense de *Gonyaulax polyedra*. La chambre de culture est gardée en permanence à 19°C et les cultures sont placées sous des tubes fluorescents à lumière blanche froide (150 microEinstein/ m² / s) contrôlés par un cycle de 12 heures de lumière et 12 heures d'obscurité (12:12 L: D). Les cellules sont recueillies deux heures après le début de la lumière (ZT 2) par filtration sous vide avec du papier Whatman 541, plongées dans l'azote liquide puis conservées à -80°C jusqu'à l'utilisation. Les cellules d'*Escherichia coli* (souche MRF', Stratagene) sont cultivée à 37°C dans le milieu LB (Sambrook et al., 1989) supplémenté au besoin de 100 µg/ml d'ampicilline et/ou 5 µg/ml de tétracycline pour assurer la sélection pour la présence d'un plasmide pBluescript et/ou le facteur F'.

Les cellules de *Saccharomyces cerevisiae* (souche SEY 6210 et YPH 499) sont cultivées à 30°C dans le milieu SC (0,67% de base d'azote de levures et bactéries sans acides aminés, 2% glucose et 0,2% "Drop-out mix" , (Rose et al. 1990). Des milieux sélectifs sans leucine et/ou uracile (SC-Leu, SC- Ura, SC- Leu - Ura) permettent de conserver des plasmides à l'intérieur des cellules lors de la culture.

2.2 Acides nucléiques

2.2.1 Purification des plasmides

L'ADN plasmidique d'*Escherichia coli* est purifié soit avec des minipréparations de lyse alcaline (" minipreps ") ou des maxipréparations de lyse alcaline (" maxipreps ") selon les protocoles de Sambrook et al.(1989). Afin d'accroître la pureté de l'ADN, il est parfois nécessaire de compléter la "maxiprep" avec une centrifugation sur gradient au chlorure de césium (Sambrook et al., 1989).

L'ADN plasmidique de la levure est purifié avec une minipréparation selon le protocole de Hieter et al. (1990) avec quelques modifications. Les cellules de levures

sont cultivées pour environ 12 heures à 30°C dans du milieu de culture approprié : un milieu complet (SC) sauf pour la sélection propre au plasmide (ex: SC-leu). Les cellules sont ensuite récupérées par une brève centrifugation puis resuspendues dans quatre volumes de mélange Triton (2 % Triton X-100, 1% SDS, 100mM NaCl, 10mM Tris-Cl pH 8, 1mM Na₂EDTA). À cette solution, un volume égal de phénol: chloroforme: alcool isoamylique (25:24:1) et 1/10 de volume de billes de verre préalablement lavées à l'acide sont ajoutées. Le tout est vortexé pendant deux minutes puis centrifugé cinq minutes à 14 000 rpm dans une microcentrifugeuse. Le surnageant est récupéré, traité au chloroforme : alcool isoamyl (24:1) puis l'ADN est précipité avec deux volumes d'éthanol 100% et 1/10 de volume de 3 M NaOAc pH 5.6 .

2.2.2 Purification de l'ADN génomique

L'ADN génomique est purifié selon un protocole maison, établi par P.Salois (1995) . Les cellules de *Gonyaulax polyedra* sont d'abord broyées dans de l'azote liquide à l'aide d'un mortier et d'un pilon pré-refroidis. La poudre ainsi obtenue est rapidement mélangée à deux volumes de tampon GIT (4M guanidium isothiocyanate, 50 mM Tris-HCl pH 8, 70 mM EDTA, 25 mM EGTA, 2% N-laurylsarcosine et 2% β-mercaptoethanol). La solution est ensuite soumise à deux extractions au phénol: chloroforme : alcool isoamyl (25:24:1) puis à une extraction au chloroforme: alcool isoamyl (24:1). Les acides nucléiques sont ensuite précipités par l'ajout de deux volumes d'éthanol 100% froid et 1/10 de volume de 3M NaOAc pH 5,6. Le nuage d'ADN ainsi formé est récupéré à l'aide d'une pipette Pasteur recourbée en crochet. L'ADN est ensuite lavé à l'éthanol 70%, asséché une dizaine de minutes à température de la pièce puis redissout dans du TE (10mM Tris-HCl pH 8, 1mM EDTA). Un traitement à la RNase A (25 µg/ml pour une heure à 37°C) puis un à la protéinase K (55 µg/ml pour une heure à 37°C) sont ensuite nécessaires pour se débarrasser des ARN et protéines présents (es) dans la solution. Les polysaccharides sont ensuite éliminés grâce à une extraction au CTAB (cetyltrimethylammonium bromide). Pour ce faire, la solution d'ADN doit être mélangée à un volume égal de solution au CTAB (55mM CTAB, 200mM Tris-HCl pH 8, 70 mM EDTA, 2 mM NaCl), préchauffée à 65°C. Le mélange est ensuite incubé dans un bain à 65°C pour 30 minutes avec agitation aux 5 minutes environ. On procède ensuite à une extraction au chloroforme: alcool isoamyl (24:1) puis le surnageant est prélevé et l'ADN qu'il contient précipité avec deux volumes d'éthanol à 100% (la solution contient suffisamment de sel pour permettre la précipitation). Le culot d'ADN est récupéré par centrifugation à 4°C (30 minutes à 8000 Xg) puis resuspendu dans du TE. L'extraction au CTAB doit être reproduite en totalité une seconde fois afin de se débarrasser efficacement des polysaccharides. Le

culot final d'ADN est lavé à l'éthanol 70%, redissout dans du TE puis conservé à -20°C jusqu'à l'utilisation.

2.2.3 Purification de l'ARN total

Plusieurs précautions doivent être prises afin de s'assurer de travailler dans un environnement exempt de toute RNase lors de l'extraction d'ARN. L'eau est traitée au DEPC 0,1% (diéthylpyrocarbonate) puis autoclavée et les instruments de verre ou de métal sont chauffés à 350°F pour environ 12 heures. Toutes les solutions nécessaires sont préparées avec de l'eau DEPC. Les billes de zirconium sont lavées au chloroforme puis conservées à -20°C. L'extraction est effectuée en permanence sur glace ou dans des appareils conservés à 4 °C. Les cellules gelées sont mélangées à un volume égal de billes de zirconium froides et à deux volumes de tampon GIT (voir recette en section 2.2.2). Les cellules sont ensuite brisées à l'aide d'un mini " bead beater " (Biospec Products) pendant deux minutes de forte agitation. Les billes et les débris non-solubles sont ensuite enlevés par centrifugation puis le surnageant est ajouté à un volume égal de phénol: chloroforme: alcool isoamyl (25:24:1). Le tout est vortexé pendant une minute puis centrifugé à 14000rpm dans une microcentrifugeuse à 4°C. Le surnageant est ensuite prélevé et l'extraction au phénol: chloroforme: alcool isoamyl est répétée jusqu'à ce qu'il n'y ait plus d'interface blanche entre la phase aqueuse et la phase organique. On procède ensuite à une extraction au chloroforme: alcool isoamyl seulement puis les acides nucléiques sont précipités par l'ajout de deux volumes d'éthanol 100% froid et de 1/10 de volume de 3M NaOAc pH 5,6 à la phase aqueuse. Le tout est conservé toute la nuit à -20°C. Les acides nucléiques sont récupérés par centrifugation à 14000 rpm à 4°C, puis le culot est resuspendu dans de l'eau DEPC. La solution d'acides nucléiques est ajustée à une concentration de 0,8 M NaCl puis mélangée à 1.1 volume de solution CTAB (voir recette en section 2.2.2). L'extraction au CTAB est effectuée telle que mentionné plus haut et les acides nucléiques sont précipités pour la nuit à -20°C. Le culot est redissout dans de l'eau DEPC puis les gros ARN sont précipités pendant cinq heures sur glace grâce à une addition de chlorure de lithium pour une concentration finale de 2 M. Les ARN sont finalement récupérés par centrifugation (30 minutes à 4°C à 27000 Xg), lavés à l'éthanol 70%, resuspendus dans de l'eau DEPC puis conservés à -80°C jusqu'à l'utilisation.

2.2.4 Purification des ARN messagers

Une colonne de plastique (séparateur d'ARNm ; Clontech) avec un ml d'oligodT (Pharmacia) est utilisée pour la purification des ARNm. La colonne de plastique est préalablement lavée avec une solution de lavage (0,1 M NaOH, 5 mM EDTA) puis rincée à l'eau DEPC jusqu'à ce que le pH de la solution sortante soit environ de 6. Cette même colonne est ensuite équilibrée avec cinq volumes de tampon de chargement (20 mM Tris-HCl pH 7.5 , 0,5 M LiCl, 0,5 mM EDTA et 0,1% SDS). La solution d'ARN totaux est chauffée cinq minutes à 65°C, refroidie sur glace, mélangée à un volume égal de solution de chargement 2X (même recette que plus haut mais avec deux fois la concentration des solutés) puis appliquée sur la colonne équilibrée. La solution d'ARN polyA⁻ sortante est récupérée, réchauffée à 65°C puis réappliquée sur la colonne. La colonne pleine d'ARN est ensuite lavée avec sept volumes de tampon de chargement 1X puis quatre volumes de tampon de chargement 1X 0,1 M LiCl. Les ARNm sont ensuite élués de la colonne par l'application de 500 µl de tampon d'éluion (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0,005% SDS). L'éluion des ARNm est répétée cinq à six fois. Les ARNm sont ensuite précipités par l'ajout de 2,2 volumes d'éthanol à 100% froid et de 1/10 de volume de 3 M NaOAc pH 5.6 puis conservés à -80°C jusqu'à l'utilisation. La figure 4 illustre le contrôle de la quantité et qualité (absence de dégradation, message enrichi) effectué par transfert de type Northern puis hybridation avec une sonde contrôle du gène PCP (Le, 1996).

2.2.5 Quantification des acides nucléiques

Les quantifications des acides nucléiques sont effectuées à l'aide d'un spectrophotomètre de type Beckman modèle Du65 (Beckman Instruments Inc.) ou par visualisation sur gel d'agarose 1% en comparaison avec des standards connus (digestion BstE II du phage lambda de New England Biolabs).

2.2.6 Digestions enzymatiques

Les digestions enzymatiques des acides nucléiques sont effectuées selon les conditions suggérées par le fabricant (Promega), sauf que le nombre d'unités d'enzyme utilisée est le double de la quantité recommandée et la digestion est effectuée toute la nuit à 37°C. La digestion est suivie d'une extraction au phénol : chloroforme : alcool isoamyl (25:24:1) et d'une précipitation de l'ADN afin d'éliminer l'enzyme de la solution. Le culot final d'ADN est resuspendu dans un petit volume de TE.



FIGURE 4 : Contrôle de la qualité des ARN messagers

Les ARN poly A⁺ sont extraits sur colonne d'oligos dT à partir d'une solution d'ARN totaux . Un microgramme des différents ARN (1: totaux, 2: poly A⁻, poly A⁺) est ensuite migré sur gel de formaldéhyde, transféré sur membrane puis hybridé avec la sonde contrôle PCP (Le, 1996) . Les résultats illustrés ici démontrent une bonne qualité (absence de dégradation et message enrichi) des ARNm .

2.2.7 Électrophorèse et électroélution

L'électrophorèse d'ADN sur gel d'agarose est effectuée telle que décrite par Sambrook et al. (1989). Les concentrations d'agarose varient généralement de 0,8% à 1,2% selon la taille des fragments d'ADN à séparer.

Les différents fragments d'ADN sont visualisés grâce à la fluorescence du bromure d'éthidium présent dans le gel (environ 0,5 µg/ml) sur une lampe à rayons ultraviolets (312 nm " transilluminator ", Fisher Scientific). Le marqueur de poids moléculaires utilisé est la digestion BstE II du phage lambda (New England Biolabs).

Les fragments d'ADN sont récupérés du gel par électroélution selon le protocole établi (Sambrook et al., 1989). Une brève inversion du courant électrique est cependant ajoutée à la fin de l'électroélution afin de s'assurer que l'ADN en solution ne reste pas collé au sac à dialyse. L'électrophorèse d'ARN est effectuée sur un gel de formaldéhyde selon le protocole établi (Sambrook et al., 1989). Le standard de poids moléculaire généralement utilisé est celui de Gibco (" 0,24-9,5 kb RNA ladder " de Gibco BRL). Les ARN sont visualisés sur une lampe à rayons ultraviolets grâce à une coloration au bromure d'éthidium effectuée après l'électrophorèse. L'électrophorèse de l'ARN peut également être effectuée sur gel d'agarose 1%. Pour ce faire, l'ARN doit être préalablement traité au méthyl-mercure (10 minutes dans du 0,01 M méthyl-mercure à température de la pièce).

2.2.8 Transfert de types Southern et Northern

Le transfert de l'ADN sur les membranes de nylon (Hybond-N⁺, Amersham) est effectué par la technique traditionnelle du buvard à l'hydroxyde de sodium (NaOH 0,4 M) (Southern, 1975).

Le transfert de l'ARN sur ces même membranes est également effectué par la technique du buvard. Le gel contenant l'ARN est d'abord plongé dans de l'eau pour cinq minutes, puis dans du 50 mM NaOH, 10 mM NaCl pour 30 minutes, dans du 0,1 M Tris-HCl pH7.5 pour 30 minutes et finalement dans du 20X SSC pour les 30 dernières minutes. Le transfert est ensuite effectué par buvard dans du 20X SSC.

Les acides nucléiques (ADN et ARN) sont ensuite fixés sur la membrane par une cuisson sous vide à 80°C pendant 2 heures.

2.2.9 Marquage d'ADN et hybridation

Les différents fragments d'ADN sont marqués avec de l'alpha³²P dCTP frais (ICN Biomedicals) et la trousse " Prime-a-Gene " de Promega selon les instructions du fabricant. Les nucléotides non-incorporés sont ensuite enlevés à l'aide d'une colonne de Biogel P60 (Bio-Rad) coulée à partir d'une suspension (20 mg/ml H₂O) dans une pipette Pasteur préalablement obturée par un bouchon de laine de verre. Environ 0,5 ng d'ADN radioactif par millilitre de solution au phosphate (0,5 M Na₂HPO₄, 7% SDS, 3% BSA, 1mM EDTA) est utilisé pour l'hybridation. Les pré-hybridation et hybridation des membranes sont effectuées dans des contenants en plastique hermétiques remplis de la solution au phosphate et plongés dans un bain-marie agitateur à 65°C. La pré-hybridation dans la solution au phosphate dure deux heures alors que l'hybridation, effectuée grâce à l'ajout de la sonde d'ADN dénaturée par la chaleur, se fait toute la nuit. Les membranes sont ensuite lavées 45 minutes à 65°C dans la solution de lavage I (5% SDS, 40 mM Na₂HPO₄) puis 30 minutes à 65°C dans la solution de lavage II (1% SDS, 40 mM Na₂HPO₄). Le second lavage peut être répété au besoin. L'autoradiographie est effectuée dans une cassette avec écran généralement à -80°C, pour une durée variable, avec des films X-OMAT AR de Kodak (Picker).

2.3 Banque d'ADNc

2.3.1 Production d'une banque d'ADNc dans les phages

Une banque d'ADNc a été construite à l'aide d'un kit de Stratagene . La synthèse d'ADNc est effectuée à l'aide du " cDNA synthesis kit " (Stratagene) en suivant les instructions du fabricant. La qualité (taille et abondance) des ADNc est vérifiée sur gel d'agarose avec transfert de Southern et hybridation avec une sonde contrôle du gène PCP (Le, 1996). L'ADNc de bonne taille est ensuite inséré dans une bactériophage lambda (Uni-Zap XR) et encapsulé grâce aux kits " Zap-cDNA synthesis kit " et " Zap-cDNA Gigapack III " selon les instructions de Stratagene. Le nombre de clones indépendants (1.4×10^6) et de recombinants contenant une insertion (93%) démontrent que la synthèse a très bien fonctionné. La qualité de la librairie amplifiée (présence, taille et diversité des insertions) à été vérifiée par une excision de masse effectuée selon les recommandations de Stratagene (double infection de XL1 Blue par des phages lambda recombinants et par des " helper phage " R408). Les colonies bactériennes individuelles sont utilisées pour des purifications d'ADN plasmidique de type miniprep et les insertions sont analysées par électrophorèse après une digestion avec

les enzymes de restriction Xho I et EcoR I, pour une analyse des insertions ; ainsi que pour une hybridation avec la sonde contrôle PCP (Le, 1996) (voir figure 5). Quelques insertions choisies au hasard ont été partiellement ou entièrement séquencées par un séquenceur automatique (UCDNA " Sequencing Services "). La séquence d'une insertion, identifiée par homologie comme étant la cytochrome oxydase sous-unité trois (*cox 3*), est présentée en annexe I.

2.3.2 Criblage d'une banque d'ADNc dans le vecteur Uni-Zap XR

Un protocole détaillé du criblage de la banque est inclus avec le kit de Stratagene. Il s'agit premièrement d'infecter les cellules hôtes XL1 Blue MRF' (Stratagene) avec un nombre variable de phages de la banque, mélanger le tout avec du NZY top agar (NZY liquide avec 0,7% d'agarose) refroidi à 48°C et étaler sur des pétris contenant du milieu NZY agar. Les boîtes sont incubées de huit à dix heures à 37°C afin de pouvoir observer les plages de lyse. L'ADN des plages de lyse est transféré sur des membranes de Nylon (Hybon-N⁺, Amersham) selon le protocole de Sambrook et al. (1989). Les acides nucléiques sont ensuite fixés sur la membrane par une cuisson sous vide de deux heures à 80°C. Les membranes sont ensuite pré-hybridées et hybridées dans le tampon phosphate tel que décrit plus haut. Différentes sondes d'ADN sont utilisées pour cribler la banque, toujours en suivant le même protocole de marquage (voir section 2.2.9). Les phages positifs sont utilisés pour une seconde et parfois une troisième infection avant d'être clairement purifiés. Les phages purs sont soumis à une excision effectuée à l'aide d'un " phage aide " (" helper phage R 408, Stratagene) en suivant les recommandations de Stratagene. Le plasmide pBluescript excisé sera ensuite purifié par un protocole de miniprep standard.

2.3.3 Production d'une banque d'ADNc dans le vecteur p426 GPD

Environ 1,5 µg d'ADNc a été produit grâce au kit " cDNA Synthesis kit " en suivant les recommandations de Stratagene. Le vecteur p426 GPD (Funk et al., 1995) est digéré avec les enzymes de restriction XhoI et EcoRI puis le plasmide linéaire est récupéré par électroélution (tel que décrit en sections 2.2.6 et 2.2.7) (voir figure 6). L'ADNc est inséré dans le plasmide p426 GPD par une ligation de 48 heures à 8°C avec de la T4 ADN ligase de Promega et un tampon de Promega additionné de rATP. Le ratio molaire de la ligation optimale (déterminé empiriquement) est de 1: 4 (vecteur : insertion). Le tout est utilisé pour une transformation dans des XL2 Blue MRF' ultracompetentes de Stratagene selon les instructions du fabricant. Environ 3 X 10⁵ clones indépendants ont été obtenus. La librairie primaire est ensuite amplifiée puis la

A	Uni-Zap XR ligué seul	Uni-Zap XR ligué avec ADNc
Nombre de plages de lyse par microgramme de vecteur Uni-Zap XR	1.6×10^3	1.4×10^6

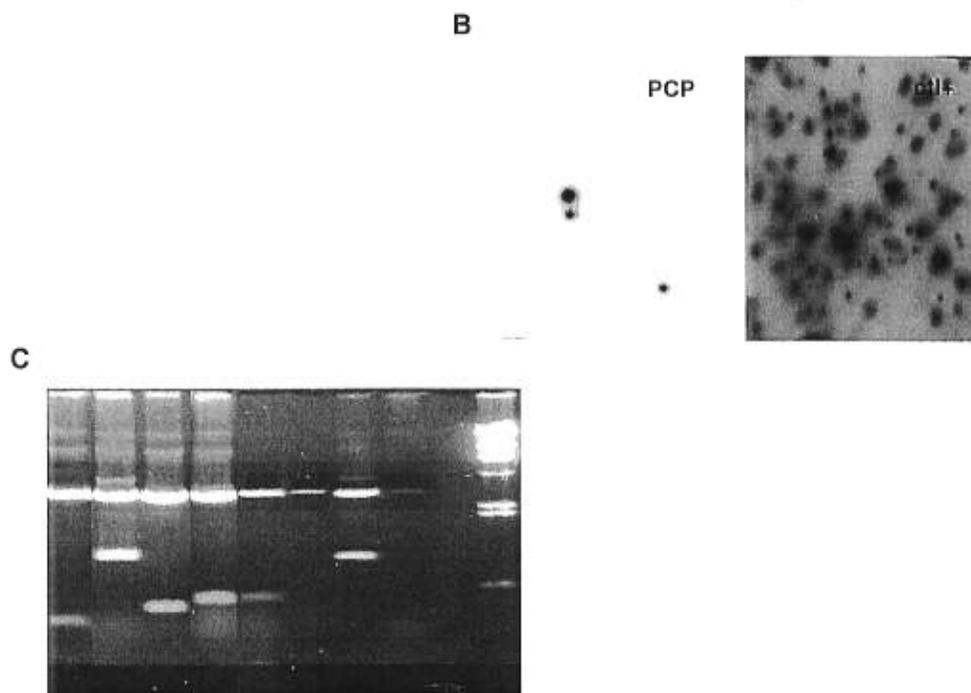


Figure 5 : Contrôle de la qualité de la banque d'ADNc dans le vecteur Uni-Zap XR

(A) Le nombre de clones obtenus dans la banque primaire est représentatif d'une bonne banque d'ADNc . (B) L'ADN des plages de lyse est transféré sur membrane et hybridé avec une sonde Uni-Zap XR (*ctl+*) et avec une sonde contrôle PCP (Le,1996) . Nous obtenons 1% des clones qui contiennent un fragment d'ADNc PCP, ce qui correspond aux résultats obtenus antérieurement (Le, 1996) . (C) Une excision de masse est effectuée sur la banque d'ADNc . Les colonies bactériennes ainsi obtenues sont soumises à une extraction d'ADN plasmidique de type miniprep . Les plasmides sont ensuite digérés avec les enzymes *EcoRI* et *Xho I* puis migrés sur gel d'agarose 1% . Les résultats obtenus démontrent la présence d'insertions de bonne taille dans 100% des extractions .

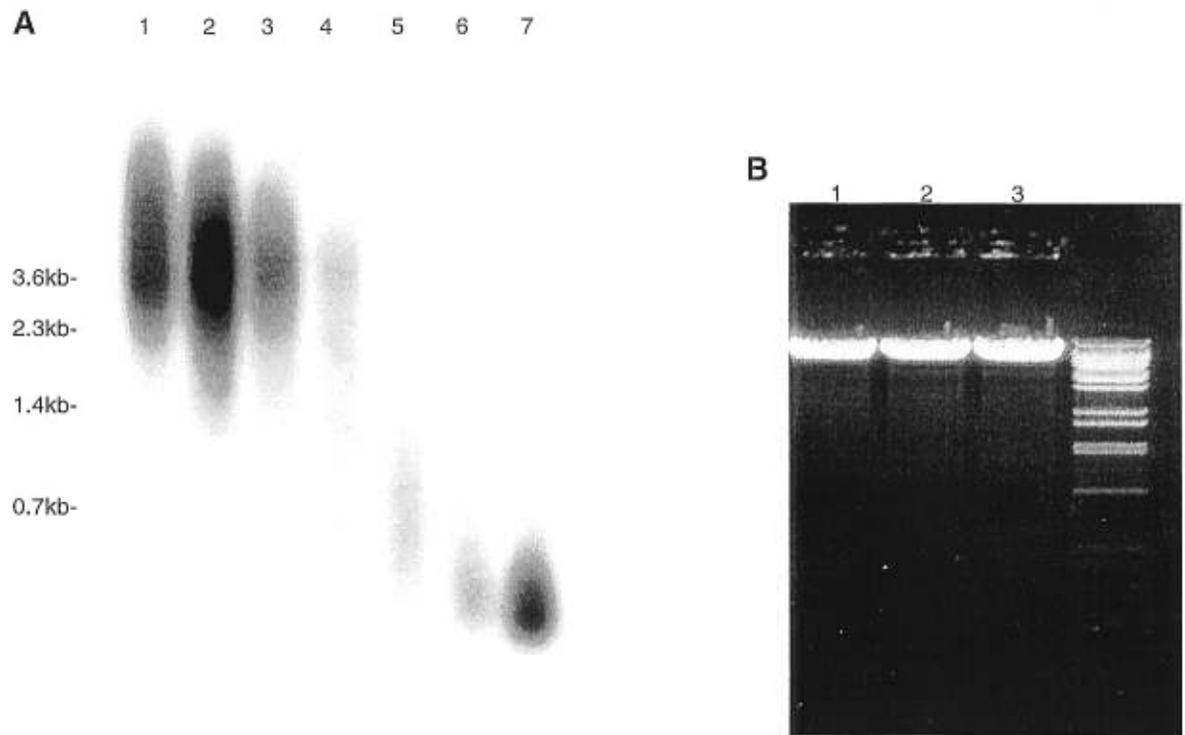


FIGURE 6 : Production d'une banque d'ADNc dans le vecteur p426 GPD

(A) Les ADNc sont produits à l'aide du kit de Stratagene puis les fractions de tailles différentes sont récupérées par migration sur colonne CL-2B (Stratagene) . Un échantillon radioactif de chaque fraction est migré sur gel d'agarose puis transféré sur membrane . Seules les trois premières fractions (contenant des ADNc de bonne taille) sont utilisées pour la production de la banque .
 (B) La qualité de la digestion du plasmide p426 GPD est testée par deux digestions simples avec les enzymes EcoRI (1) et Xho I (2) . Les résultats démontrent une digestion complète du plasmide par ces deux enzymes . Le plasmide doublement digéré par les enzymes Xho I et Eco RI (3) est utilisé pour l'insertion des ADNc .

qualité des insertions testée par hybridation avec une sonde contrôle du gène PCP (Le, 1996) et par miniprep (voir protocole à la section 2.2.1). Les résultats de ces analyses, illustrés à la figure 7, suggèrent que presque 100% des vecteurs possèdent une insertion de bonne taille.

2.3.4 Criblage de la banque d'expression dans le vecteur p426 GPD chez la levure .

Le criblage de notre banque d'expression fonctionnelle implique la capacité des cellules à résister à l'induction de l'expression du gène létal GSP1 (de P. Belhumeur, 1997, figure 7). Cette induction se fait grâce au promoteur inductible GAL qui a été joint au gène Gsp1 et à la substitution du glucose par du galactose ultrapur (Sigma) dans le milieu de culture final SC-Leu-Ura 2% galactose. Le plasmide pCS7 (construit par M. Mittag selon la stratégie démontrée en figure 8) contient également la séquence UG répétés qui confère une régulation au niveau traductionnel au gène GSP1. Le criblage de la banque s'effectue de par lui-même grâce au cumul des fonctions des deux plasmides (la transcription du gène létal de la construction pCS7 est induite et sa traduction est inhibée par la présence d'un agent inhibiteur lui-même produit par une insertion du p426 GPD). Les transformants positifs dont la croissance n'est pas stoppée par l'induction du gène létal sont obtenus après 48 heures d'incubation à 30°C sur pétris SC-leu-ura 2% galactose. Chacun de ces positifs est cultivé dans du milieu SC-leu-ura 2% galactose, soumis à une extraction d'ADN plasmidique par miniprep pour la levure (protocole en section 2.2.1) et utilisé pour transformer des bactéries. Les bactéries sont également soumises à une extraction d'ADN plasmidique par miniprep (voir protocoles en section 2.2.1), puis le mélange de plasmides obtenu est retransformé dans les cellules contenant la construction pCS7 afin de contrôler une seconde fois la croissance après induction du gène létal.

2.4 Réaction en chaîne de polymérisation (PCR)

Les réactions de PCR sont effectuées à l'aide d'un appareil Minicycler (MJ Research), avec du tampon de la Taq Polymérase de Promega et une Taq polymérase préparée selon le protocole du Pluthero (1993). Différents oligonucléotides ont été utilisés pour différentes réactions (50 pmol par réaction) ; soit les oligonucléotides T₃ et T₇ (qui permettent de vérifier les insertions du plasmide Bluescript), ou les oligonucléotides HC CYC 3' (cataactaattacatgatat) et HC gpd 5' (cttaaacttcttaaattcta) (qui permettent de vérifier les insertions du plasmide p426 GPD). Les températures de

A

	p 426 GPD digéré , ligué seul	p 426 GPD digéré, ligué avec l'ADNc	p 426 GPD surenroulé	p UC 18 surenroulé
nombre de colonies bactériennes par ug de vecteur	1×10^3	2.8×10^5	1×10^6	1×10^9

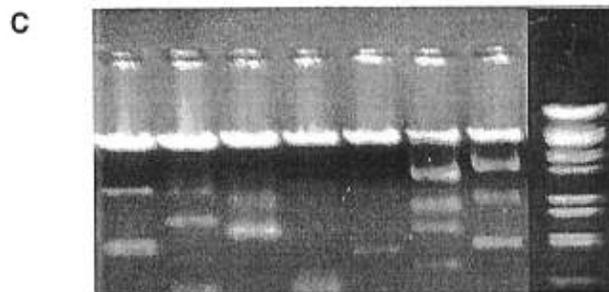
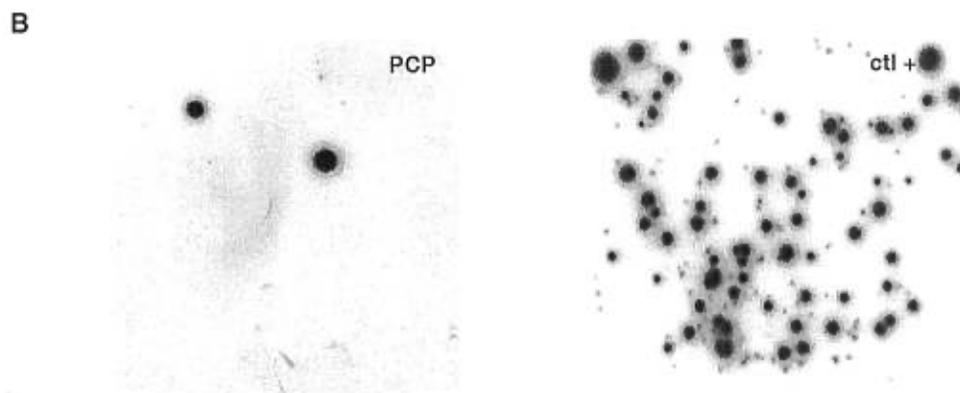
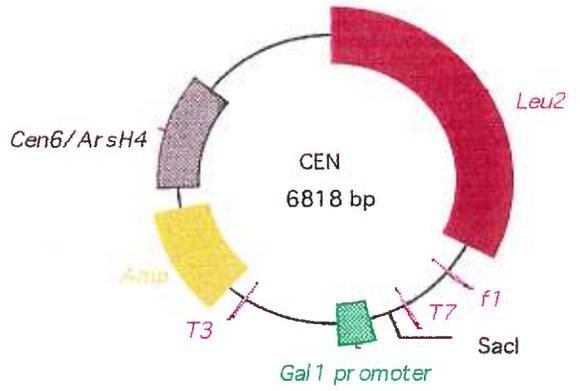
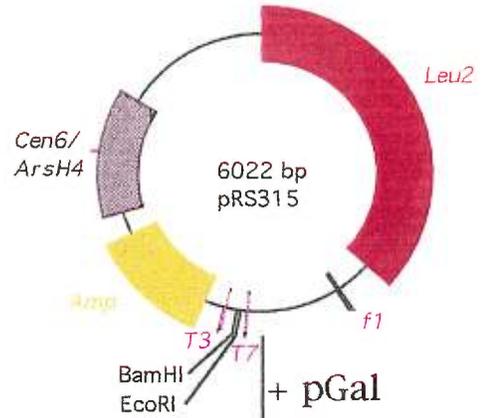
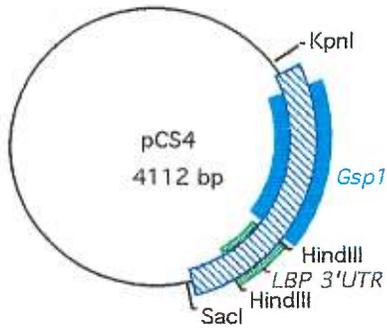
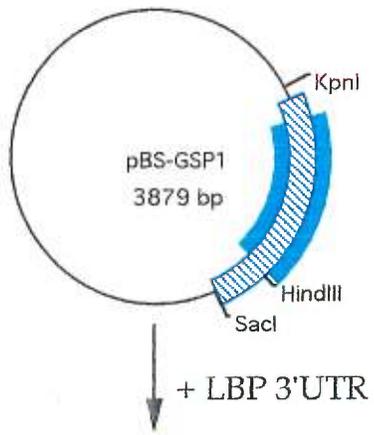
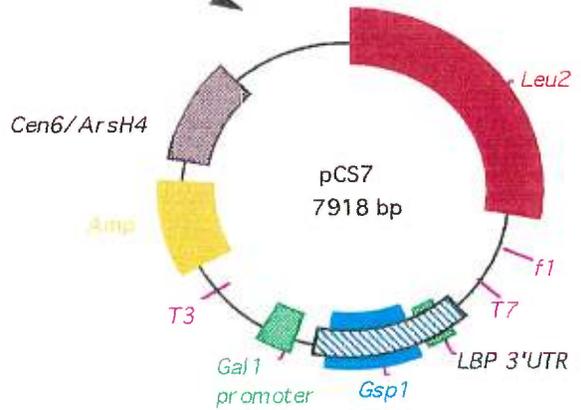


Figure 7 : Contrôle de la qualité de la banque d'ADNc obtenue dans le vecteur p426 GPD (A) Le nombre de clones obtenus dans la banque primaire semble représentatif d'une bonne banque d'ADNc . (B) L'ADN des colonies bactériennes est transféré sur membrane puis hybridé avec les sondes PCP et p 426 GPD (ctl+) . Nous obtenons 0.75% des clones qui contiennent un ADNc PCP, ce qui correspond aux résultats obtenus antérieurement (Le, 1996) . (C) Des colonies bactériennes choisies au hasard sont soumises à une extraction d'ADN plasmidique de type miniprep . Les plasmides sont ensuite digérés avec les enzymes XhoI et EcoRI puis migrés sur gel d'agarose 1% . Les résultats obtenus démontrent la présence d'insertions de bonne taille dans 100% des extractions .



clonage
avec extrémités
franches



dénaturation, hybridation et élongation varient selon les réactions de PCR mais le patron suivant fonctionne généralement bien avec les oligonucléotides mentionnés ci-haut. D'abord une première étape de 10 minutes à 94°C. Puis, 35 cycles comportant une phase d'une minute à 94°C, une phase d'une minute à 48°C et une phase d'une minute à 72°C. Une fois les cycles terminés, les échantillons sont soumis à un dernier 5 minutes à 72°C avant l'arrêt de l'appareil.

Pour la visualisation des résultats, les produits de PCR sont ensuite chargés sur un gel d'agarose et colorés au bromure d'éthidium tel que décrit plus haut.

2.5. Séquençage

Les produits à séquencer (ADN plasmidique double brin avec insertion) sont purifiés à l'aide de miniprep additionnée d'une précipitation au polyéthylène glycol (Sambrook et al., 1989). La pureté de la préparation de plasmide est ensuite vérifiée en chargeant un échantillon sur gel d'agarose. Un minimum de 5 µg de la préparation exempte de toute trace d'ARN est ensuite envoyée à l'Université de Calgary pour y être séquencée à l'aide d'un séquenceur automatique (UCDNA " Sequencing Services "). Les séquences d'ADN sont ensuite analysées avec le logiciel Geneworks pour Macintosh et confrontées à la base de données GENBANK à l'aide de l'algorithme BLAST (Altschul, 1990).

2.6 Cellules compétentes et transformation

2.6.1 Chez *Escherichia Coli*

Les cellules de la souche XL1Blue MRF' de Stratagene peuvent être rendues compétentes par un traitement au chlorure de calcium 0,1 M et transformées par des constructions plasmidiques de différentes tailles selon le protocole de Sambrook et al. (1989). Les recombinants sont étalés sur des boîtes de pétri contenant 12,5 µg/ml de tétracycline pour la sélection de l'épisome F' de la souche XL1 Blue MRF' et 100 µg/ ml d'ampicilline pour la sélection des transformants. Le rendement d'une telle méthode est cependant variable (de 10⁴ à 10⁷ colonies par microgramme de plasmide pBluescript surenroulé), et particulièrement faible pour des constructions plasmidiques de bonne taille (8000 pb et plus) telles que le p426 GPD (10³ à 10⁴ colonies par microgramme de plasmide).

Les cellules ultracompetentes XL2 Blue MRF' de Stratagene peuvent être utilisées pour obtenir de meilleurs rendements (jusqu'à 10⁹ colonies par microgramme de pBluescript et 10⁵ colonies par microgramme de p426 GPD). Dans ce cas la

transformation est effectuée avec une préparation d'ADN plasmidique purifié par gradient au chlorure de césium et selon les recommandations de Stratagene.

2.6.2 Chez la levure

Les cellules de la souche SEY 6210 et YPH 499 sont transformées selon la méthode LiAc (Gietz et al., 1995). Les cellules sont d'abord cultivées toute la nuit à 30°C dans un milieu riche complet YPAD (6 g de " yeast extract, 12 g de peptone, 12 g de glucose et 60 mg d'adénine hémissulfate sont dissout dans 600 ml d'eau distillée, le tout étant autoclavé). Les cellules sont cultivées jusqu'à ce que leur titre atteigne 2×10^7 cellules/ ml. Les cellules sont alors lavées avec une centrifugation (3000 Xg pour 5 minutes), un transfert dans de l'eau distillée et une recentrifugation. Ces cellules propres sont ensuite resuspendues deux fois dans du 100 mM LiAc pour atteindre un titre final de 2×10^9 cellules/ ml. Un volume de 50 μ l (1×10^8 cellules) est alors brièvement centrifugé et le culot est resuspendu dans 240 μ l de PEG (50% w/v) auquel 36 μ l de 1 M LiAc, 25 μ l de SS-DNA (2mg/ ml) et 50 μ l d'eau contenant le plasmide (0,1 à 10 μ g) sont ajoutés. Le tout est vigoureusement vortexé, incubé à 30°C pour 30 minutes, puis à 42°C pour 25 minutes, centrifugé à 3000 Xg pour 15 secondes puis le culot de cellules est resuspendu dans 1 ml d'eau stérile et la solution est étalée sur des pétris avec milieu sélectif SC-leu (4 g de " yeast nitrogen base " , 12 g de glucose, 0,4 g de mélange d'acides aminés sauf leucine et 10 g d'agar, ajoutés à 600 ml d'eau et le tout est autoclavé). Les pétris sont ensuite incubés environ 48 heures à 30°C pour obtenir les transformants. L'ajout d'un second plasmide aux transformants obtenus s'effectue selon le même protocole à quelques différences près : le milieu de culture initial est sélectif (SC-leu) afin de conserver la présence du premier plasmide. Le milieu de culture final, quant à lui, se doit de posséder deux types de sélections (SC-leu-ura).

3. Résultats

3.1 Contrôle circadien du rythme de la bioluminescence

3.1.1 L'isolation d'un ADNc codant pour la protéine CCTR

Le mécanisme de l'activation de la traduction de l'ARNm LBP se produit à la tombée de la nuit et est sous le contrôle d'une horloge circadienne (voir section 1.2.2). La protéine LBP produite à partir des AUG 2 et 3 de l'ARNm LBP oscille selon un rythme circadien (voir section 1.2.2). Une séquence UG répétée située en 3' non-traduite (3' UTR) de l'ARNm LBP agit comme un site de fixation pour une protéine inhibitrice (CCTR) capable de se lier à cette région durant la phase de jour (voir section 1.2.2). Le mécanisme de répression de la traduction de LBP par le CCTR reste encore inconnu mais la formation d'un ARNm circulaire *in vivo* pourrait expliquer comment une protéine fixée en 3' vient stopper la traduction initiée en 5' de l'ARNm.

Le protocole ici élaboré vise à permettre l'isolation et le clonage d'un ADNc codant pour la protéine CCTR impliquée dans la régulation de l'expression de *lbp*. Une banque d'ADNc produite avec un kit de Stratagene est insérée dans le vecteur p426 GPD (Funk et al., 1995). Ce vecteur d'expression constitutive peut être maintenu dans la levure grâce à une sélection pour la base azotée uracile. La banque ainsi produite est insérée dans des cellules de levure (souche SEY 6210) contenant la construction pCS7 (voir figure 8) selon le protocole de Gietz et al. (1995) (voir section 2.6.2). Le vecteur pCS7, maintenu dans la levure grâce à une sélection pour la leucine, contient le gène léta1 GSP1 (Belhumeur et al., 1993) placé sous le contrôle d'un promoteur inductible (le promoteur GAL) de façon à ce que, en présence de galactose, les cellules

expriment le gène létal et meurent. La construction pCS7 contient aussi la séquence régulatrice UG répétée insérée en région 3' du gène GSP1. Cela permettrait que la protéine CCTR de *Gonyaulax polyedra* puisse venir inhiber la traduction du gène létal selon le même mode d'action que pour l'ARNm LBP.

Les cellules de levure sont donc, dans un premier temps, transformées avec le plasmide pCS7, puis cultivées dans un milieu dépourvu de leucine (sélection pour le plasmide) et riche en glucose comme source de carbone (inhibant ainsi la transcription gène létal). Dans un deuxième temps, ces cellules sont transformées avec la banque d'ADNc contenue dans le plasmide p426 GPD, puis cultivées ultimement dans un milieu 2% galactose dépourvu de leucine et d'uracile. Cette concentration de galactose a été déterminée empiriquement comme étant nécessaire pour tuer toutes les cellules porteuses de la construction pCS7. Notre méthode de criblage repose donc sur la survie et la croissance de cellules possédant une protéine CCTR fonctionnelle capable de stopper la traduction du gène GSP1 en se fixant sur la séquence UG répétée de la construction pCS7 (voir figure 9).

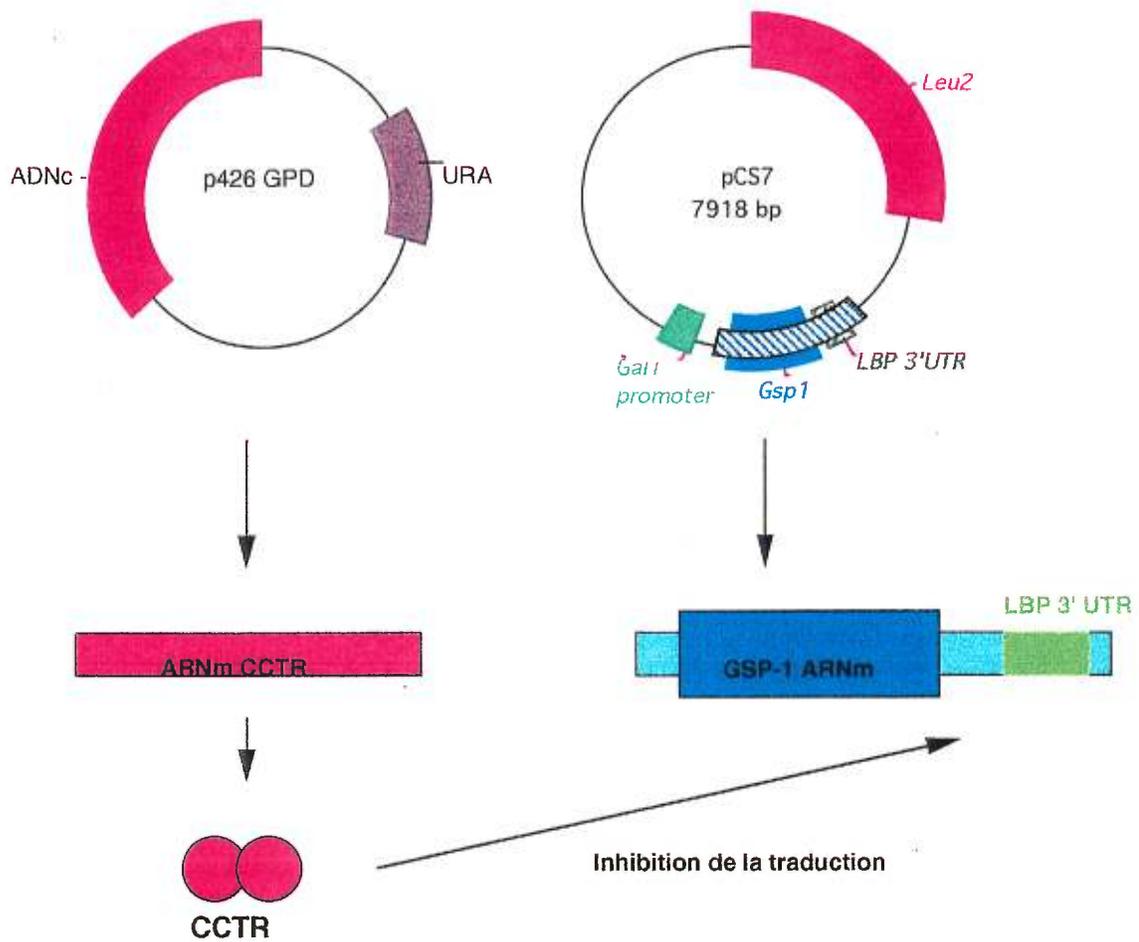


Figure 9 : Interaction moléculaire *in vivo* .

L'expression d'une protéine CCTR fonctionnelle par un ADNc contenu dans le plasmide p426GPD permettrait d'inhiber la traduction de l'ARNm GSP-1 suite à la fixation de la CCTR sur le 3' UTR LBP .

Les cellules transformées contenant les deux plasmides sont tout d'abord cultivées dans un milieu 2% glucose dépourvu de leucine et d'uracile pendant environ deux heures afin de leur permettre de « récupérer » du choc thermique de la transformation et de commencer à exprimer le contenu du plasmide p426 GPD. Elles sont ensuite étalées sur pétri contenant un milieu 2 % galactose sans leucine ni uracile et incubées à 30°C pour environ 48 heures. Outre une légère croissance secondaire, nous observons parfois la formation de quelques colonies suite à l'incubation. Un total de 127 colonies sont ainsi obtenues sur l'équivalent de $2,8 \times 10^5$ clones indépendants contenus dans la totalité de la banque d'ADNc (voir figure 10). Par comparaison, un criblage pour le gène *tim* utilisant le système " two-hybrid " de Statagene a permis d'isoler 360 positifs sur un total de 2×10^7 clones criblés (Gekakis, 1995). On observe donc une fréquence de positifs 25 fois plus élevée pour notre technique de criblage.

Chacun des positifs est repiqué puis cultivé dans un milieu 2% galactose sans leucine ni uracile pour environ 12 heures, puis une extraction d'ADN plasmidique de type miniprep est effectuée sur les cellules ainsi obtenues. Une partie de l'ADN plasmidique est chargé sur gel d'agarose 1%, soumis à une brève électrophorèse, coloré au bromure d'éthidium puis observé sur une lampe à rayons ultraviolets. Le patron de bandes observées semble démontrer seulement la présence du plasmide p426 GPD (voir figure 11a). Cela pourrait s'expliquer par le fait que le plasmide p426 GPD se retrouve à une abondance d'environ 30 copies par cellules alors que la construction pCS7 est normalement présente en une seule copie par cellule.

La solution d'ADN plasmidique de chacun des positifs obtenus est ensuite utilisée pour une seconde transformation dans des cellules de levure contenant la construction pCS7. La très faible croissance des transformants étalés sur un milieu 2% galactose moins leucine, moins uracile semble indiquer qu'aucun de ces positifs ne contient un ADNc capable de coder pour une protéine CCTR fonctionnelle (voir figure 11b).



Contrôle + : 2% glucose
- leucine
- uracile



Criblage : 2% galactose
- leucine
- uracile

Figure 10 : Criblage pour isoler le CCTR

Suite à la seconde transformation, les cellules de levure sont étalées sur des milieux moins leucine moins uracile, 2% glucose pour le contrôle positif permettant la croissance de tous les transformants et 2% galactose pour le criblage . Des colonies sont bien visibles sur le contrôle positif alors qu'une simple croissance secondaire est observée sur le milieu 2% galactose .

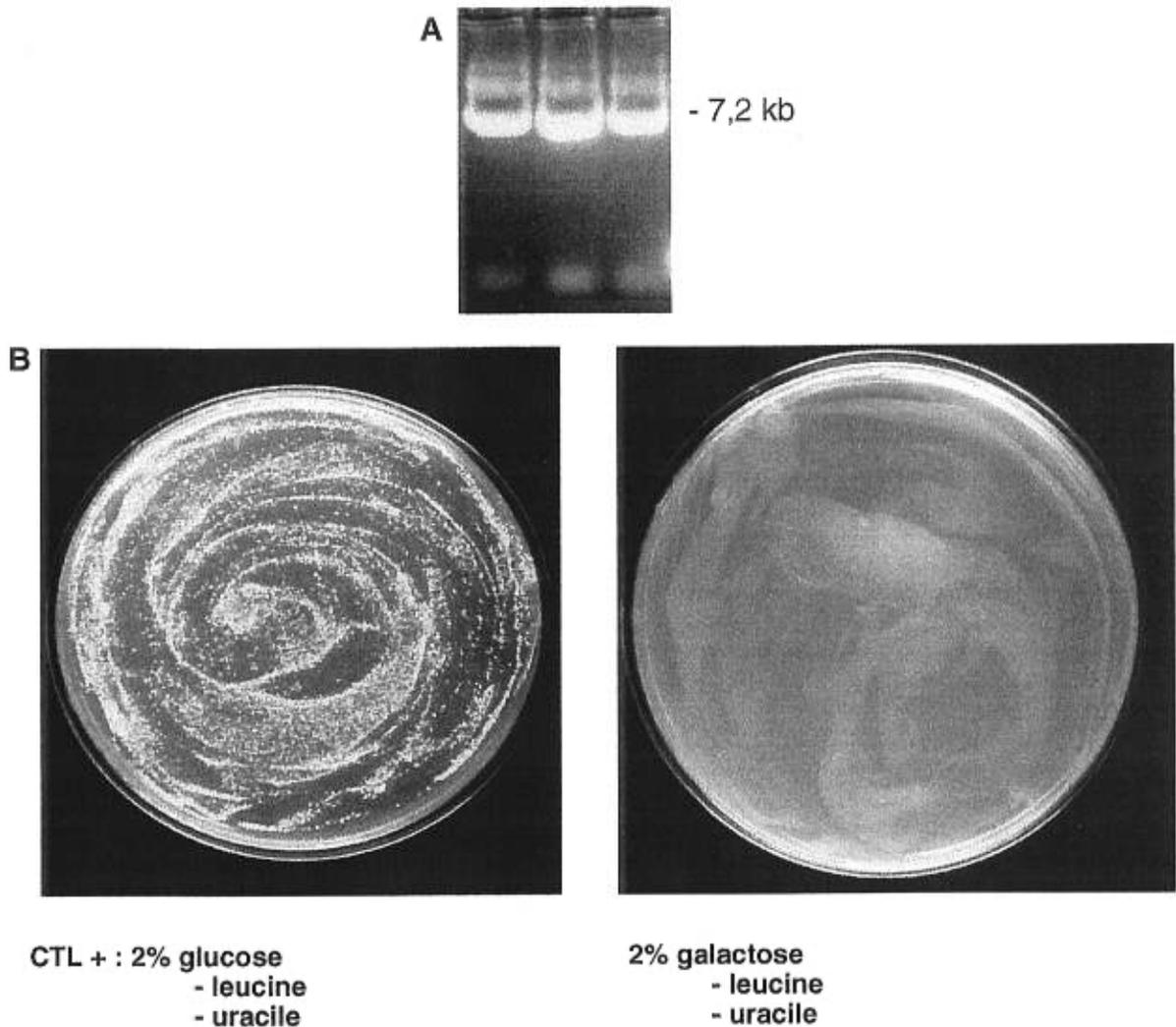


Figure 11 : Second criblage des positifs lors de l'isolation du CCTR.

A. Les positifs obtenus lors du premier criblage ont été cultivés puis soumis à une extraction d'ADN plasmidique de type miniprep. L'électrophorèse sur gel d'agarose 1% de cet ADN nous permet d'observer la présence de plasmide.

B. Une seconde transformation de l'ADN plasmidique isolé dans des levures contenant le pCS7 ne permet pas d'observer une bonne croissance.

3.2 Contrôle circadien du rythme de la division cellulaire

3.2.1 Isolation d'un homologue de la kinase p34^{cdc2}

Chez *Gonyaulax polyedra* la mitose est limitée au début de la phase lumineuse (Nicolas et al., 1991). Le temps de génération typique de ces algues est de 48 ou 72 heures (un multiple de 24) puisque la division cellulaire n'est possible qu'à l'aurore, même lorsque la croissance est terminée. En général, le cycle cellulaire des eucaryotes est régulé par un homologue de la kinase p34^{cdc2} de la levure. Cette kinase s'associe à la cycline lorsque : (i) la croissance cellulaire est terminée, (ii) la réplication de l'ADN a été effectuée et (iii) la cycline est retrouvée en bonne quantité (voir section 1.2.2) (Elledge, 1996). L'association de ces deux protéines forme le complexe MPF qui permet l'entrée en phase mitotique.

Le protocole présenté ici vise à permettre l'isolation et le clonage d'un homologue de la kinase p34^{cdc2} chez *Gonyaulax polyedra*. La banque d'ADNc produite avec les ARNm prélevés à CT2 est clonée dans le vecteur p426 GPD en aval du promoteur constitutif glycéraldéhyde-3-phosphate (voir section 2.3.3). La banque d'ADNc produite dans ce vecteur est transformée dans des levures de la souche thermosensible CMY 478. Ces cellules, qui poussent normalement à 28°C, possèdent une mutation thermosensible située dans le gène *cdc28*. Cette mutation rend inactive la protéine kinase p34^{cdc2} et produit donc l'arrêt de la croissance lorsque les cellules sont placées à une température restrictive de 34°C (Reed, 1980). Les levures transformées avec la banque sont donc tout simplement étalées sur boîtes de pétris et incubées à température restrictive. Notre méthode de criblage repose sur la croissance de cellules possédant un ADNc codant pour un homologue fonctionnel de la kinase p34^{cdc2}, permettant la formation d'un MPF actif et donc la division cellulaire (voir figure 12). Cette

technique a été utilisée avec succès lors du clonage de *cdc2* chez plusieurs organismes dont l'humain (Lee et Nurse, 1987).

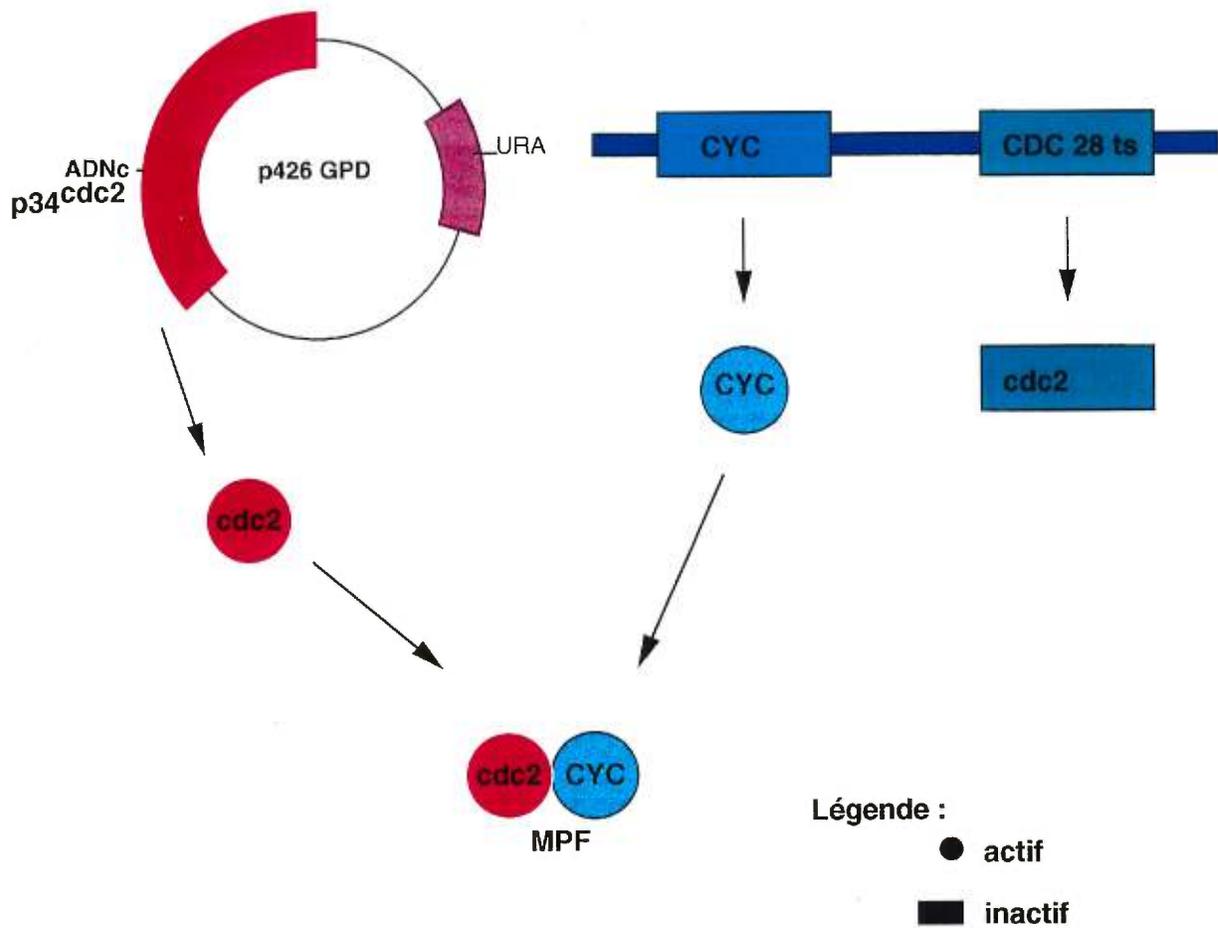


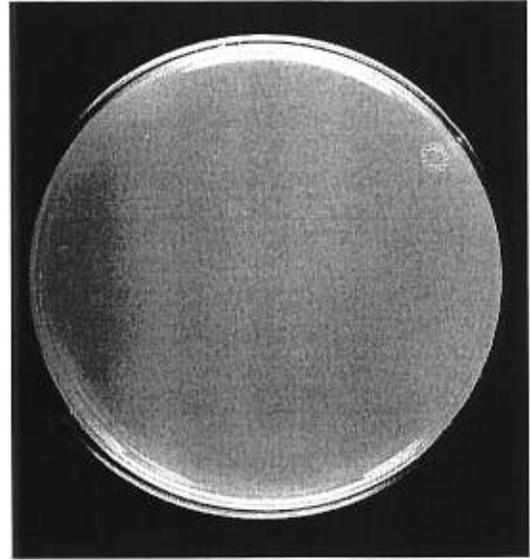
Figure 12 : Interaction moléculaire *in vivo* à 34°C.

L'expression d'un homologue de la p34^{cdc2} par un ADNc de *Gonyaulax polyedra* permettrait la formation d'un MPF actif et donc la division cellulaire à température restrictive de 34°C.

La banque d'ADNc d'environ 4×10^5 clones a été utilisée pour transformer les cellules de la souche CMY 478 et seulement 7 colonies ont été formées à température restrictive. Normalement, aucune croissance n'est observée à 34°C (voir figure 13). Les 7 colonies positives ont été repiquées puis cultivées dans un milieu sans uracile (sélection pour le plasmide). Les cellules ont été soumises à une extraction d'ADN plasmidique de type « miniprep ». L'ADN plasmidique est ensuite vérifié par électrophorèse sur gel d'agarose 1% et coloration au bromure d'éthidium (voir figure 14a) puis utilisé pour une seconde transformation dans des levures thermosensibles (voir figure 14b). Aucune croissance n'a été observée à température restrictive suite à la seconde transformation.



28 °C



34° C

Figure 13 : Criblage pour l'isolation d'un homologue de la kinase $p34^{cdc2}$.

Les transformants sont étalés sur des pétris et incubés respectivement : à 28°C pour le controle positif où on observe une bonne croissance, à 34°C pour le criblage où on observe généralement aucune croissance.

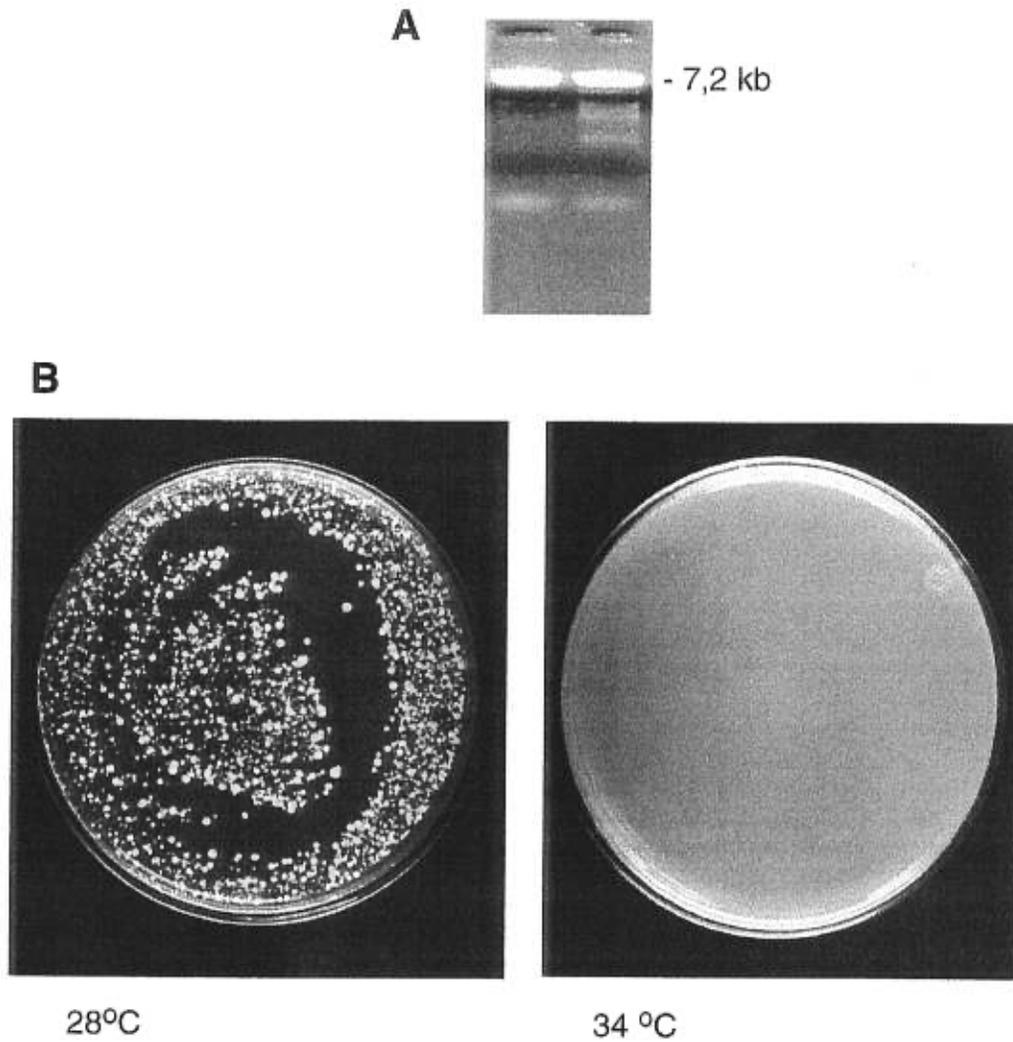


Figure 14 : Second criblage des positifs lors de l'isolation d'un homologue de la p34^{cdc2}

A. Les positifs obtenus lors du premier criblage ont été cultivés puis soumis à une extraction d'ADN plasmidique de type "miniprep". L'électrophorèse sur gel d'agarose 1% de cet ADN plasmidique nous permet d'observer une bande principale dont la taille correspond environ au vecteur p426 GPD non digéré.

B. Une seconde transformation de l'ADN plasmidique isolé, dans les levures de la souche thermosensible ne permet pas d'observer une croissance à température restrictive.

4. Discussion

4.1 Isolation d'un ADNc codant pour la protéine CCTR.

Les résultats négatifs obtenus lors du second criblage (voir figure 11B) indiquent que les ADNc étudiés ne peuvent produire une protéine CCTR fonctionnelle. La présente section expose les principales hypothèses permettant d'expliquer ces résultats.

Les cellules transformées contenant les deux plasmides (le pCS7 et le pGPD avec ADNc) sont d'abord cultivées deux heures en milieu liquide (2% glucose -leucine -uracile) puis étalées sur des pétris (2% galactose -leucine -uracile). L'absence de leucine et d'uracile dans ces deux milieux permet de conserver une sélection pour la présence des deux plasmides. La présence de galactose à une concentration de 2% (déterminée empiriquement) dans les milieux de culture solides permet d'activer la transcription du gène GSP1 et ainsi de tuer toutes les cellules porteuses du plasmide pCS7. Nous avons cependant obtenu 127 colonies suite au criblage de $2,8 \times 10^5$ clones. Ces positifs obtenus lors du premier criblage se sont tous avérés incapables de produire une protéine CCTR fonctionnelle lors du second criblage. Cependant, chacun d'entre eux contenait des plasmides d'environ 7 kb, observés lors de l'extraction de type miniprep effectuée entre les deux criblages. Les plasmides observés semblent tous de tailles semblables, correspondant à la taille attendue du plasmide pGPD avec ADNc inséré (voir figure 11A). Aucune trace du plasmide pCS7 (7,9 kb) n'a été observée dans ces mêmes extractions. Mais comme ce plasmide est normalement présent à une seule copie par cellule, il est possible que sa présence

n'ait tout simplement pas été visible sur le gel d'agarose coloré au bromure d'éthidium.

Dans un premier temps, nous pouvons soupçonner qu'une contamination par des cellules de levure d'une autre souche soit à l'origine de la présence des faux positifs observés. Chaque solution utilisée est d'abord stérilisée et le travail est effectué dans une hotte à flux laminaire mais les quantités de cellules compétentes utilisées sont astronomiques, le nombre de cellules étalées sur les pétris considérable et les périodes d'incubation plutôt longues (voir section 2.6.2). Des levures d'une autre souche, capables de produire leur propre leucine et possédant seulement le plasmide pGPD (sélection par l'uracile) pourraient croître en toute quiétude en présence du galactose contenu dans le milieu de culture final. Par contre, un faux positif dû à une contamination de ce type pourrait pousser s'il était repiqué sur un autre pétri.

La concentration de galactose dans le milieu de culture final a été déterminée empiriquement comme essentielle afin de tuer toutes les cellules contenant le pCS7. En effet, des tests effectués en milieux de culture liquide ont permis de constater qu'un milieu de culture 2% galactose avec une concentration aussi minime que 0,01% de glucose permettait la survie et la croissance de quelques cellules contenant le pCS7. Une contamination de glucose par les cellules de levure mortes environnant les colonies observées pourrait donc expliquer leur présence sur le milieu de culture final. Dans ce cas, on pourrait réduire le nombre de faux positifs en diluant la quantité de cellules mises sur la boîte initiale.

Chez *Saccharomyces cerevisiae* le gène GSP1 code pour un polypeptide de 219 acides aminés, apparenté aux GTPases monomériques de type ras . La mutation dans ce gène augmente la stabilité de la protéine GSP1 liée au GTP, provoquant ainsi son activation permanente et donc un phénotype létal dominant (Belhumeur et al., 1993). Il existe cependant des mutations de type "révertants"

(permettant de récupérer le phénotype initial ou regain de fonction) qui sont normalement observée à une fréquence environnant les 10^{-7} (Royer-Pokora, 1984). Un renversement de phénotype peut être produit par différents mécanismes tels que : i. mutation inverse de la défektivité génétique primaire; ii. seconde mutation compensatoire dans le même gène; iii. suppression de la copie défective du gène par un inhibiteur antisens; iv. augmentation de l'activité d'une voie métabolique secondaire. La fréquence d'observation de "révertants" est très faible mais sur un criblage de $2,8 \times 10^5$ clones il est possible qu'un regain de fonction soit à l'origine des quelques positifs observés lors du premier criblage. Afin de s'assurer que ces positifs ne sont pas des "révertants", une seconde transformation est nécessaire.

Le clonage d'un ADNc codant pour l'acide aminé leucine dans le plasmide pGPD pourrait également entraîner la présence de faux positifs. En effet, la présence de cet ADNc dans un vecteur d'expression constitutive comme le pGPD permettrait d'éliminer la pression de sélection du plasmide pCS7, en produisant suffisamment de leucine pour les besoins cellulaires. Sans pCS7, le galactose est tout à fait inoffensif pour les cellules de levure. Il est possible qu'une petite partie des cellules possédant un clone de leucine ait perdu leur pCS7 et ainsi formé les colonies positives du premier criblage. Lors d'un repiquage, ces clones pousseront mais ils ne permettront pas d'observer une croissance suite à une seconde transformation.

Nous devons également prendre en considération quelques failles plus techniques de la méthode de criblage utilisée. L'isolation du gène *timeless* chez la drosophile a été effectuée par une interaction protéine-protéine, à l'aide du gène *period* et d'un criblage effectué par le biais du système " two-hybrid " de Stratagene (Gekakis, 1995). Le criblage effectué chez la drosophile à été fait à partir d'une banque contenant 2×10^7 clones indépendants et a permis d'isoler 360 positifs par le test de la croissance en absence d'histidine. De plus, seulement 67

des 360 positifs ont également démontré des résultats positifs au second test plus stringent pour la présence d'activité enzymatique de type β -galactosidase (Gekakis,1995) et seulement 16 des 67 correspondaient au *timeless*. Nous pouvons donc déduire qu'environ un million de clones doivent être criblés avant d'avoir, statistiquement, une bonne chance d'isoler un gène impliqué dans la génération d'un rythme circadien. Notre banque d'ADNc de *Gonyaulax polyedra* contenait seulement $2,8 \times 10^5$ clones indépendants. La production d'un plus grand nombre de clones serait donc nécessaire afin de s'assurer, statistiquement, d'un criblage efficace.

L'interaction des protéines PER et TIM soulève un second problème technique qui pourrait s'appliquer à notre approche. TIM se lie à PER via les domaines PAS et la présence simultanée des deux protéines est essentielle pour le fonctionnement de l'horloge. En effet, c'est l'hétérodimère PER-TIM qui pénètre le noyau et y interagit avec l'hétérodimère CYC-CLK afin de permettre le "feedback" négatif et donc l'expression cyclique de gènes impliqués dans la génération d'un rythme circadien (voir section 1.3.2). Si le système de génération du rythme circadien de la bioluminescence de *Gonyaulax polyedra* ressemble à celui de la drosophile, la protéine CCTR pourrait être en fait un hétérodimère. Dans ce cas, il faudrait procéder à des transformations à haute concentration plasmidique afin d'inclure plusieurs plasmides pGPD avec ADNc dans une même cellule. Cette procédure de transformations répétées comporte toutefois un rendement si faible que l'obtention d'un hétérodimère fonctionnel semble presque impossible avec les moyens techniques actuels.

Chez *Neurospora crassa*, le gène *frq* code pour une composante principale de sa propre rétro-inhibition au niveau de la transcription (Aronson et al., 1994). La protéine FRQ doit cependant être modifiée post-traductionnellement par diverses phosphorylations à différents sites afin d'atteindre la forme active qui pénètre le noyau (voir section 1.3.2). Il est possible que la protéine CCTR nécessite également des modifications post-traductionnelles afin d'être activée. Si

tel est le cas, la levure hôte pourrait bien constituer un organisme trop différent de *Gonyaulax polyedra* pour reconnaître et activer la protéine CCTR.

De toutes ces hypothèses de travail, celle à envisager en priorité, considérant l'approche technique, est celle du manque de clones indépendants. La production de clones supplémentaires pourrait permettre d'atteindre le million souhaité et ainsi d'effectuer un criblage plus efficace. L'isolation d'un clone codant pour une protéine CCTR fonctionnelle permettrait d'étudier le lien entre cette protéine CCTR et la région de régulation de la traduction de LBP (en 3' UTR). La compréhension de ce mécanisme pourrait permettre de mieux saisir le fonctionnement de l'horloge régissant le rythme circadien de la bioluminescence chez *Gonyaulax polyedra*. L'ADNc ainsi isolé pourrait être cloné dans un vecteur d'expression de Qiagen (vecteur pQE " His-tagged ") permettant ainsi la purification de protéines CCTR nécessaires à la production d'anticorps spécifiques. À l'aide de cet anticorps, nous pourrions tester les variations d'abondance et de taille des protéines CCTR sur des gels de type Western effectués sur une période de 24 heures. Cette analyse permettrait de vérifier les changements circadiens de CCTR *in vivo*. Si des variations d'abondance sont dénotées, il faudrait alors mesurer les niveaux d'ARNm cctr sur un northern avec des échantillons étalés sur une période de 24 heures. Cela permettrait de déterminer si le contrôle de l'expression du gène *cctr* se situe au niveau traductionnel (comme pour le gène *lbp*) ou au niveau transcriptionnel (comme pour les gènes *per*, *tim* et *frq*).

Si des variations de taille de la protéine CCTR sont observées, nous pourrions alors suspecter que des modifications post-traductionnelles sont à l'origine des changements d'activité de cette protéine CCTR. Dans ce cas, une analyse des variations du point isoélectrique de la protéine CCTR serait effectuée sur un gel de protéines en deux dimensions, à l'aide d'un marquage avec l'anticorps spécifique produit. Une variation du point isoélectrique en fonction du

temps circadien confirmerait la présence de mécanismes de modifications post-traductionnelles de la protéine CCTR.

4.2 Isolation d'un ADNc homologue du gène *cdc2* de la levure .

Le gène *cdc2* joue un rôle essentiel dans le contrôle du cycle cellulaire de la levure *Schizosaccharomyces pombe* (voir section 1.2.2 . La banque d'ADNc construite dans le vecteur d'expression constitutive pGPD a été utilisée afin de tenter d'isoler un homologue du gène *cdc2*. La transformation de cette banque (contenant $2,8 \times 10^5$ clones) dans des cellules de la souche *cdc28* (mutant thermosensible de la $p34^{cdc2}$, Reed 1980) a permis d'observer la formation de sept colonies positives lors du premier criblage. En effet, les cellules de la souche *cdc28* ne peuvent se diviser normalement à température restrictive de 34°C, les sept colonies formées à cette température (voir figure 13) nous apparaissant donc comme porteuses potentielles d'un homologue de la $p34^{cdc2}$. Chaque positif a été cultivé puis soumis à une extraction d'ADN plasmidique de type " miniprep ". Un échantillon de chaque extraction a été placé sur gel d'agarose 1% et a permis d'observer la présence de plasmides d'environ 7 kb (voir figure 14a). Les solutions de plasmides ainsi obtenues ont été utilisées pour une seconde transformation dans des levures de la souche *cdc28*. Une incubation à 34°C ne nous a pas permis d'observer la croissance attendue (voir figure 14b). Donc, aucun des sept positifs ne semble capable de produire un homologue fonctionnel de la $p34^{cdc2}$.

Plusieurs hypothèses peuvent permettre d'expliquer ces résultats. Une fois de plus, la possibilité d'une contamination est à mentionner. Seules les cellules de la souche *cdc28* sont incapables de croître à une température de 34°C. La présence de levures d'une autre souche dans les cultures de cellules traitées au chlorure de césium pour devenir compétentes pourrait expliquer la croissance de

colonies sur un milieu sélectif pour le plasmide pGPD (moins uracile) incubé à 34°C.

De plus, le nombre de clones présents dans la banque ($2,8 \times 10^5$) semble insuffisant afin de s'assurer de la représentation de tous les gènes de *Gonyaulax polyedra*. Un plus grand nombre de clones permettrait, statistiquement, de s'assurer de la présence de l'homologue du gène *cdc2* lors du criblage.

Chez la majorité des eucaryotes, l'oscillation du cycle cellulaire est basée sur la formation de l'hétérodimère MPF (voir section 1.2.2). La kinase p34^{cdc2} doit être déphosphorylée sur la thréonine 14 et la tyrosine 15 pour s'associer à la cycline et ainsi former le MPF. La phosphatase *cdc25* est responsable de cette déphosphorylation. Chez *Gonyaulax polyedra* la mitose s'effectue d'une façon atypique (voir section 1.1.2). Cette division cellulaire non-conventionnelle porte à croire que le mécanisme moléculaire responsable de la mitose chez les dinoflagellés pourrait légèrement différer du mécanisme typique des eucaryotes. Dans ce cas, il est possible que les modifications post-traductionnelles de l'homologue de la kinase p34^{cdc2} de *Gonyaulax polyedra* ne puissent être effectuées par la phosphatase *cdc25* de la levure hôte. Sans activation de la p34^{cdc2} il n'y a pas de formation de MPF et donc aucune division mitotique possible. De plus, même si la kinase est activée, elle doit être capable de s'associer à la cycline de la cellule hôte. Si les deux protéines ne peuvent se reconnaître et s'associer, il n'y a pas de formation de MPF actif et donc pas de mitose. L'ensemble de ces mécanismes moléculaires nécessaires à la formation d'un MPF actif ainsi que les échecs lors de tentatives antérieures d'isolation d'un homologue de la p34^{cdc2} par des techniques plus conventionnelles chez *Gonyaulax polyedra* (Salois et Morse, 1995) peuvent inciter à croire que cette expérience est plutôt hasardeuse. La technique de clonage par compensation d'une mutation chez un organisme hétérologue tel que la levure a cependant fait ses preuves lors de l'isolation d'un homologue du gène *cdc2* chez l'humain (Lee et Nurse, 1987). De plus, la présence d'une période spécifique de synthèse d'ADN

(phase S) chez plusieurs dinoflagellés, dont *Gonyaulax polyedra* (Homma et Hastings, 1989b), ainsi que la détection immunologique, chez *Cryptothecodinium cohnii*, d'une protéine de 34 kDa capable de phosphoryler l'histone H1 *in vitro* (Rodriguez et al., 1993) sont deux indices qui portent à croire que *Gonyaulax* pourrait bel et bien posséder un homologue du gène *cdc2*.

Chez *Gonyaulax polyedra*, la mitose est limitée au début de la phase lumineuse. L'homologue du gène *cdc2* présent chez *Gonyaulax polyedra* pourrait donc jouer un rôle dans la formation du MPF et également être récepteur d'informations en provenance de l'horloge circadienne afin de limiter la mitose à cette portion de la journée. Il semble donc intéressant de comprendre le rôle de la kinase p34^{cdc2} dans le contrôle de l'horloge sur le cycle cellulaire. L'homologue du gène *cdc2* pourrait être utilisé pour fabriquer une sonde permettant d'observer les variations de l'ARNm *cdc2* sur une période de 24 heures à l'aide de Northern. Une variation d'abondance du message permettrait de conclure à une régulation au niveau de la transcription alors qu'un message constant indiquerait plutôt une régulation post-traductionnelle. L'isolation d'un homologue du *cdc2* permettrait également la production de kinases p34^{cdc2} via le vecteur d'expression de Qiagen susmentionné. La purification de p34^{cdc2} permettrait la production d'un anticorps spécifique à cette kinase. À l'aide de cet anticorps, les variations d'abondance et de taille de la kinase dans des échantillons étalés sur 24 heures pourront être observées sur des Westerns. Les variations de phosphorylation de cet homologue de la kinase p34^{cdc2} pourront être déterminées par marquage avec l'anticorps spécifique sur des gels en 2D produits avec des échantillons étalés sur une période de 24 heures. Chez la majorité des eucaryotes, une variation du point isoélectrique est observée lors de la déphosphorylation des deux acides aminés de la kinase. Chez *Gonyaulax polyedra*, un second changement du point isoélectrique pourrait être observé suite à une action des composantes de l'horloge circadienne. Finalement, l'obtention d'un homologue du gène *cdc2* permettrait d'obtenir également l'homologue de la seconde composante du MPF, la cycline, à

l'aide du système " two-hybrid " de Stratagene, utilisant les interactions protéines-protéines.

4.3 Conclusion

La production d'une banque d'ADNc dans un vecteur d'expression constitutive tel que le p426GPD est un outil fantastique pour l'isolation de différents clones. Malheureusement, les rendements de transformation de ce plasmide, d'abord chez *E. coli*, ensuite chez la levure sont très faibles. C'est pourquoi tous les efforts démontrés ont permis la production de seulement $2,8 \times 10^5$ clones. Les techniques de criblage utilisées sont basées sur : l'interaction protéine-ARNm pour l'isolation d'un clone codant pour la protéine CCTR et la compensation d'une mutation chez un organisme hétérologue pour l'isolation d'un homologue du gène *cdc2*. Ces criblages sont faciles et rapides à effectuer. C'est pourquoi des efforts supplémentaires ont été entrepris afin de répéter le criblage de tous les clones de la banque ainsi que la production de nouveaux clones. L'idéal serait de pouvoir cribler un million de clones indépendants. Un tel nombre de clones devrait permettre l'isolation d'un ADNc codant pour la protéine CCTR. Sinon, des transformations répétées pourront être entreprises afin de permettre la présence de plusieurs ADNc de *Gonyaulax* dans une seule et même cellule de levure, visant ainsi à réunir tous les constitutants nécessaires à la production d'une protéine CCTR fonctionnelle.

La technique de clonage par compensation d'une mutation, utilisée afin d'isoler un homologue du gène *cdc2* a déjà fait ses preuves avec d'autres organismes. Cependant, l'isolation d'un tel homologue chez *Gonyaulax polyedra* semble périlleuse et un "détour" via l'isolation d'un homologue de la cycline par le même principe technique sera peut-être nécessaire. L'isolation d'un des deux membres constituant le MPF devrait permettre assez aisément d'obtenir son partenaire à l'aide du système " two-hybrid " de Stratagene.

6. Bibliographie

Allada, R., White, N., So, W., Hall, J., and Rosbash, M. (1998). A mutant drosophila homolog of mammalian *Clock* disrupts circadian rhythm and transcription of *period* and *timeless* . *Cell* **93** : 791-804.

Altschul, S. F., Gish, W., Miller, E. W., and Lipman, D. J. (1990). Basic local alignment search tool . *J. Mol. Biol.* **215** : 403-410.

Antoch, P., Song, E.-J., Chang, A.-M., Vitaterna, M., Zhao, Y., Wilsbacher, D., Sangoram, M., King, P., Pinto, H., and Takahashi, J. (1997). Fonctionnal identification of the mouse circadian *Clock* gene by transgenic BAC rescue, *Cell* **89** : 655-667.

Aronson, B., Johnson, K., Loros, J., and Dunlap, J. (1994). Negative feedback defines a circadian clock : autoregulation of the clock gene frequency. *Science* **263** : 1578- 1584.

Balsalobre, A., Damiola, F., and Schibler, U. (1998). A serum shock induces circadian gene expression in mammalian tissue culture cells. *cell* **93** : 929-937.

Belhumeur, P. et al., (1993). GSP1 et GSP2, genetic suppressors of the prp20-1 mutant in *Saccharomyces cerevisiae* : GTP-binding proteins involved in the maintenance of nuclear organization . *Mol. Cell. Biol.* **13** : 2152- 2161.

Chatton, E. (1920). Les péridiniens parasites . Morphologie, reproduction, éthologie . *Arch. Zool. Exp. Gen.* **59** : 1-475.

Crosthwaite, S., Loros, J., and Dunlap, J. (1995). Light-induced phase resetting of a circadian clock is mediated by a rapid increase in frequency transcript. *Cell* **81** : 1003-1012.

Crosthwaite, S., Dunlap, J., and Loros, J. (1997). Neurospora *wc-1* and *wc-2* : Transcription, Photoresponses, and the origins of circadian rythmicity. *Science* **276** : 763-769.

Curtin, K., Huang, Z., and Rosbash, M. (1995). Temporally regulated nuclear entry of the *Drosophila* *period* protein contributes to the circadian clock. *Neuron* **14** : 365-372.

- Darlington, K., Wager-Smith, K., Ceriani, M., Staknis, D., Gekakis, N., Steeves, D., Weitz, C., Takahashi, J., and Kay, A. S. (1998).** Closing the circadian loop : CLOCK-induced transcription of its own inhibitors *per* and *tim* . *Science* **280** : 1599-1603.
- Desa, R. J., Hastings, W., and Vatter, A. E. (1963).** Luminescent « cristalline » particules : an organized subcellular bioluminescent system. *Science* **141** : 1269-1270.
- Dodge, J. D. (1965).** Chromosome structure in the dinoflagellates and the problem of the mesokaryotic cell. *Int. Cong. Ser. Excerpta. Med.* **91** : 264-265.
- Dodge, J. D. (1996).** The Dinophyceae. Dans : « The chromosomes of the algae » Goodward, M. B. E. ed., St-Martin's Press, New York, pp : 96-115.
- Dunlap, J. C., and Hastings, W. (1981).** The biological clock in *Gonyaulax* controls luciferine activity by regulating turnover . *J. Biol. Chem.* **256 (20)** : 10509-10518.
- Dunphy, W., Brizuela, L., Beach, D., and Newport, J. (1988).** The xenopus cdc2 protein is a component of MPF, a cytoplasmic regulator of mitosis. *Cell* **54** : 423-431.
- Ederly, I., Zwiebel, L., Dembinska, M., and Rosbash, M. (1994a).** Temporal phosphorylation of the drosophila period protein. *Proc. Natl. Acad. Sci. USA* **91** : 2260-2264.
- Ederly, I., Rutila, J., and Rosbash, M. (1994b).** Phase shifting of the circadian clock by induction of the Drosophila period protein. *Science* **263** : 237-240.
- Edmunds, L. E. Jr. (1988).** Cellular and molecular basis of biological clocks . Springer-Verlag. New York. 497pp.
- Eppley, R. W., Holm-Hansen, O., and Strickland, J. D. H. (1968).** Some observations on the vertical migration of dinoflagellates . *J. Phycol.* **4** : 333-340.
- Evans T., Rosenthal, E., Youngblom, J., Distel, D., and Hunt, T. (1983).** Cyclin : a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division . *Cell* **33** : 389-396.
- Fifilan, S. A., and Sigee, D. C. (1977).** Continuous DNA replication in the nucleus of the dinoflagellate *Protocestrum micans* . Ehrenberg . *J. Cell. Sci.* **27** : 81-90.
- Fogel, M., and Hastings, W. (1971).** A substrate-binding protein in the *Gonyaulax* bioluminescence reaction. *Arch. Biochem. Biophys.* **142** : 310-321.
- Franker, C. K. (1971).** Division synchrony in primary cultures of an endozoic dinoflagellate. *J. Phycol.* **7** : 165-169.

Franker, C. K., Smith, L. F., and Sakhrani, L. M. (1973). Morphotypic transition during synchronous growth of *Cryptothecodinium (Gyrodinium) cohnii* . Arch. Mikrobiol. **90** : 255-262.

Fritz, L., Morse, D., and Hastings, W. (1990). The circadian bioluminescence rhythm of *Gonyaulax polyedra* is related to daily variations in the number of light-emitting organelles. J. Cell. Sci. **95** : 321-328.

Garceau, N., Liu, Y., Loros, J., and Dunlap, J. (1997). Alternative initiation of translation and time specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY. Cell **89** : 469-476.

Gekakis, N., Saez, L., Delahaye-Brown, A.-M., Myers, M., Sehgal, A., Young, M., and Weitz, C. (1995). Isolation of timeless by PER protein interactions : defective interaction between timeless protein and long period mutant PERL. Science **270** : 811-815.

Gekakis, N., Staknis, D., Nguyen, B. H., Davis, C. F., Wilsbacher, D., King, P., Takahashi, J., Weitz, J. C. (1998). Role of the CLOCK protein in the mammalian circadian mechanism. Science **280** : 1564-1569.

Gietz, R., D., and Schiestl, R.,H. (1995). Transforming yeast with DNA . Methods in Molecular and Cellular Biology **5**: 255-269.

Guillard, R. R. L., and Ryther, J. H. (1962). Studies on marine planktonic diatoms : *Cyclotella nana* Hufstedt and *Denotula confervacea* (Cleve). Grand. Can. J. Microbiol. **8** : 229-239.

Hastings, W., and Sweeney, B. M. (1957). On the mecanism of temperature independance in a biological clock . Proc. Natl. Acad. Sci. USA **43** : 804-811.

Hastings, W., and Sweeney, B. M. (1958). A persistant diurnal rhythm of bioluminescence in *Gonyaulax polyedra* . Biological Bulletin **115** : 440-458.

Hastings, W., Astrachan, L., and Sweeney, B. M. (1961). A persistant daily rhythm in photosynthesis. J. Gen. Physiol. **45** : 69-76.

Herring, P. J. (1992). La luminescence des animaux. La Recherche **23** : 424-433.

Heyde, F. V. D., Wilkens, A., and Rensing, L. (1992). The effects of temperature on the circadian rhythms of flashing and glow in *Gonyaulax polyedra* : Are the two rhythms controlled by two oscillators ? J. Biol. Rhythms **7** : 115-123.

- Homma, K., and Hastings, W.** (1989a). Cell growth kinetics, division asymmetry, and volume control at division in the marine dinoflagellate *Gonyaulax polyedra* : a model of circadian clock control of the cell cycle . J. Cell Science **92** : 303-318.
- Homma, K., and Hastings, W.** (1989b). The S-phase is discrete and is controlled by the circadian rhythm in the dinoflagellate *Gonyaulax polyedra* . Exp. Cell. Res. **182** : 635-644.
- Howard, A., and Pelc, S. R.** (1953). Synthesis of deoxyribonucleic acid in normal and irradiated cells and its relation to chromosome breakage. Heredity **6** : 261-273.
- Huang, Z., Edery, I., and Rosbash, M.** (1993). PAS is a dimerization domain common to Drosophila Period and several transcription factors. Nature **364** : 259-262.
- Huang, Z., Curtin, K., and Rosbash, M.** (1995). PER protein interactions and temperature compensation of a circadian clock in Drosophila. Science **267** : 1169-1172.
- Hunter-Ensor, M., Ousley, A., and Sehgal, A.** (1996). Regulation of the Drosophila protein timeless suggests a mechanism for resetting the circadian clock by light. Cell **84** : 677-685.
- King, P., Zhao, Y., Sangoram, M., Wilsbacher, D., Tanaka, M., Antoch, P., Steeves, D., Vitaterna, M., Korhauser, M., Lowrey, L., Turek, W., and Takahashi, J.** (1997). Positional cloning of the mouse circadian *Clock* gene. Cell **89** : 641-653.
- Knoetzel, J., and Rensing (1990).** Characterisation of the photosynthetic apparatus from the marine dinoflagellate *Gonyaulax polyedra* : II. Circadian rhythmicity of photosynthesis and the supramolecular organisation of pigment-protein complexes. J. Plant Physiol. **136** : 280-288.
- Kozak, M.** (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs . Nucl. Acids Research **12** : 857-872.
- Kozak, M.** (1991). An analysis of vertebrate mRNA sequences : Initiations of translational control. J. Cell. Biol. **115** : 887-903.
- Kubai, D. F., and Ris, H.** (1969). Division in the dinoflagellate *Gyrodinium cohmi* Schiler. J. cell. Biol. **40** : 508-528.
- Leadbeater, B. et Dodge, J. D.** (1967). An electron microscopic study of nuclear and cell division in dinoflagellate. Ark. Microbiol. **57** : 239-254.
- Lee, M. G., and Nurse, P.** (1987). Complementation used to clone a human homologue of the fission yeast cell control gene *cdc2* . Nature **327** : 31-35.

- Liu, Y., Garceau, N., Loros, J., and Dunlap, J. (1997).** Thermally regulated translational control of FRQ mediates aspects of temperature responses in the *Neurospora* circadian clock. *Cell* **89** : 477-486.
- Loros, J., Denome, S. A., and Dunlap, J. (1989).** Molecular cloning of genes under control of the circadian clock in *Neurospora*. *Science* **243** : 385-388.
- Marrus, S., Zeng, H., and Rosbash, M. (1996).** Effect on constant light and circadian entrainment of pers flies : evidence for light mediated delay of the negative feedback loop in *Drosophila*. *EMBO J.* **15** : 6877-6886.
- McMurry, L., and Hastings, W. (1972).** No desynchronization among four different circadian rhythms in the unicellular algae, *Gonyaulax polyedra* . *Science***175** : 1137-1139.
- Mittag, M., Lee, D.-H., and Hastings, W. (1994).** Circadian expression of the luciferin-binding protein correlates with the binding of a protein to the 3' untranslated region of its mRNA. *Proc. Natl. Acad. Sci. USA* **91** : 5257-5261.
- Mittag, M., Eckerskorn, C., Strupat, K., and Hastings, W. (1997).** Differential translational initiation of *lbp* mRNA is caused by a 5' upstream open reading frame . *FEBS letters* **411(2-3)** : 245-250.
- Moore-Ede, M. C., Sulzman, F. M., and Fuller, C. A. (1982).** The clock that time us. Harvard University Press, Cambridge, Massachusetts .
- Morse, D., Pappenheimer, A. M., and Hastings, W. (1989a).** Role of a luciferin binding protein in the Circadian Bioluminescent Reaction of *Gonyaulax polyedra* . *J. Biol. Chem.* **264** : 11822-11826.
- Morse, D., Milos, P. M., Roux, E., and Hastings, W. (1989b).** Circadian regulation of bioluminescence in *Gonyaulax* involves translational control . *Proc. Natl. Acad. Sci. USA* **86** : 172-176.
- Morse, D., Fritz, L., and Hastings, W. (1990).** What is the clock ? Translational regulation of circadian bioluminescence . *TIBS* **15** : 262-265.
- Morse, D., Hastings, W., and Roenneberg, T. (1994).** Different phase responses of the two circadian oscillators in *Gonyaulax*. *J. Biol. Rhythms* **9** : 263-274.
- Mumberg, D., Müller, R., and Funk, M. (1995).** Yeast vector for the controlled expression of heterologous proteins in different genetic backgrounds, *Gene* **156**: 119-122.
- Murray, A. (1992).** Creative blocks : cell-cycle checkpoints and feedback controls . *Nature* **359** : 599-604.

- Myers, M., Wager-Smith, K., Rothenfluh-Hilfiker, A., and Young, M. (1996).** Light-induced entrainment of TIMELESS and entrainment of the *Drosophila* circadian clock. *Science* **271** : 1736-1740.
- Nakamura, H., Kishi, Y., Shimomura, O., Morse, D., and Hastings, W. (1989)** Structure of dinoflagellate luciferin and its enzymatic and non-enzymatic air-oxydation products. *J. Am. Chem. Soc.* **111** : 7607-7611.
- Nicolas, M. T., Morse, D., Bassot, J. M., and Hastings, W. (1991).** colocalisation of luciferin binding protein and luciferase to the scintillons of *Gonyaulax polyedra* revealed by double immunolabelling after fast-freeze fixation. *Protoplasma* **160** : 159-166.
- Oakley, B. R., and Dodge, J. D. (1974).** Kinetochores associated with the nuclear envelope in the mitosis of a dinoflagellate. *J. Cell. Biol.* **63** : 322-325.
- Pluthero, F. G. (1993).** Rapid purification of high-activity *Taq* DNA polymerase, *Nucleic Acids Research* **21:20**: 4850-4851.
- Qiu, J., and Hardin, P. (1996).** per mRNA cycling is locked to lights off under photoperiodic conditions that support circadian feedback loop function. *Mol. Cell. Biol.* **16** : 4182-4188.
- Rae, P. M. (1970)** The nature and processing of ribosomal ribonucleic acid in a dinoflagellate. *J. Cell. Biol.* **46** : 106-113.
- Rae, P. M., and Steele, R. E. (1978).** Modified bases in the DNAs of unicellular eukaryotes : an examination of distribution and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems* **10** : 37-53.
- Reed, S. I. (1980).** The selection of *S. cerevisiae* mutants defective in the start event of cell division. *Genetics* **95** : 561-577.
- Reppert, M. S., and Weaver, D. (1997).** Forward genetic approach strikes gold : cloning of a mammalian clock gene. *Cell* **89** : 487-490.
- Rizzo, P. J. et Nooden, L. D. (1972).** Chromosomal proteins in the dinoflagellate *Gyrodinium cohnii*, *Science* **176** : 796-797.
- Rizzo, P. J. (1981).** Comparative aspects of basic chromatin proteins in dinoflagellates. *Bio-Systems* **14** : 433-443.
- Rizzo, P. J. (1987).** Biochemistry of the Dinoflagellate nucleus. Dans « The biology of dinoflagellates » , Taylor, F. J. R. ed., Blackwell Scientific, Oxford, pp : 143-173.

- Rodriguez, M., Cho, J.W. et Rizzo, P.J.** (1993). Evidence for the presence of a *cdc2*-like protein kinase in the dinoflagellate *Cryptothecodinium cohnii*. J. Euk. Microbiol. **40** : 91-96.
- Roenneberg, T., Cofax, G. N., and Hastings, W.** (1989). A circadian rhythm of population behavior in *Gonyaulax polyedra* . J. Biol. Rhythms **4** : 201-216.
- Roenneberg, T., and Morse, D.** (1993). Two circadian oscillators in one cell. Nature **362** : 362-364.
- Rose, M., Winston, F., and Hieter, P.** (1990). Methods in yeast genetics, Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, p:126-130.
- Royer-Pokora, B. et Haseltine, A. W.** (1984). Isolation of a UV-resistant revertants from a xeroderma pigmentosum complementation group A cell line . Nature **311** : 390-392.
- Rutila, J., Suri, V., Le, M., So, W., Rosbash, M., and Hall, J.** (1998) CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of drosophila *period* and *timeless* . Cell **93** : 805-814.
- Salois, P., and Morse, D.** (1996). Do dinoflagellates contain a *cdc2*-like protein kinase? Mol. Mar. Biol. Biotech. **5** (**1**) : 52-61.
- Sambrook, J., Fritsch, E. F., and Maniatis, T.** (1989). Molecular cloning : A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sehgal, A., Price, J., Man, B., and Young, M.** (1994). Lost of circadian behavioral rhythms and per RNA oscillations in the Drosophila mutant *timeless*. Science **263** : 1603-1606.
- Sehgal, A., Rothenfluh-Hilfiker, A., Hunter-Ensor, M., Chen, Y., Myers, M., and Young, M.** (1995). Rhythmic expression of *timeless* : a basis for promoting circadian cycles in *period* gene autoregulation. Science **270** : 808-810.
- Southern, E. M.** (1975). Detection of specific sequences among DNA fragments separated by electrophoresis . J. Mol. Biol. **98** : 503-517.
- Spector, L. D.** (1984). Dinoflagellates. (Spector D.L. ed.), Academic Press Inc., Orlando, pp : 1-14, 107-147, 343-361.
- Sulzman, F. M., Krieger, N. R., Gooch, V. D., and Hastings, W.** (1978). A circadian rhythm of the luciferin binding protein from *Gonyaulax polyedra* . J. comp. Physiol. **128** : 251-257.

- Sweeney, B. M., and Hastings, W.** (1958). Rhythmic cell division in populations of *Gonyaulax polyedra*. *J. Protozool.* **5** : 217-224.
- Triemer, R. E., and Fritz, L.** (1984). Cell cycle and mitosis. Dans « *Dinoflagellates* » (Spector, D. L. ed.) Academic Press Inc., Orlando. pp.149-179.
- Vitaterna, M., King, D., Chang, A.-M., Kornhauser, J., Lowrey, P., McDonald, D., Dove, W., Pinto, L., Turek, F., and Takahashi, J.** (1994). Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. *Science* **264** : 719-725.
- Vosshal, L., Price, J., Sehgal, A., Saez, L., and Young, M.** (1994). Block in nuclear localization of period protein by a second clock mutant, *timeless*. *Science* **263** : 1606-1609.
- Zeng, H., Hardin, P., and Rosbash, M.** (1994). Constitutive overexpression of the *drosophila* period protein inhibits period mRNA cycling. *EMBO J.* **13** : 3590-3598.
- Zusman, D., and Rosenberg, E.** (1970). DNA cycle of *Myxococcus xanthus* . *J. Mol. Biol.* **49** : 609-619.

Annexe

Cette annexe présente une analyse d'un fragment d'ADN isolé lors de l'élaboration du protocole de criblage fonctionnel des gènes de *G. polyedra*. Ce fragment d'ADN a attiré notre attention puisqu'il semblait comporter des séquences de gènes mitochondriaux isolés pour la première fois chez *G. polyedra*.

Dinoflagellate mitochondrial transcripts contain random gene fragments and are polyadenylated at random sites

Hélène Chaput and David Morse

Biology Department, University of Montreal
4101 Sherbrooke est, Montreal, Canada H1X 2B2

Correspondence should be addressed to:

David Morse

Biology Department, University of Montreal
4101 Sherbrooke est, Montreal, Canada H1X 2B2

Tel: 514 872 9975

Fax: 514 872 9406

email: david.morse@umontreal.ca

Abstract

An AT-rich cDNA isolated from a *Gonyaulax* library was identified as a mitochondrial cytochrome oxidase subunit 3 (*cox3*) by sequence comparisons. Transcripts hybridizing to a *cox3* probe are expressed at high levels, and all appear to be polyadenylated. Three *cox3* cDNA sequences have been analyzed, and these sequences show two unusual features. First, while the length of the 5' end is relatively constant, the length at the 3' end varies by 1 kb between the longest and shortest clones. This suggests that poly(A) addition can occur at different sites, a hypothesis supported by the observation that the *cox3* probe hybridizes to transcripts with a wide range of sizes on Northern blots. Second, the longest cDNA contained a 350 base pair fragment of a *cox1* gene and a 600 bp fragment of a cytochrome b (*cob*) gene in the 3' extension. This suggests that the mitochondrial genome might contain numerous gene fragments, and this hypothesis is supported by both Northern and Southern hybridizations. Northern blots using the *cob* gene fragment as probe showed two size classes of transcript, one consistent with the sequence of the long cDNA and one smaller. Since the smaller form is long enough to encode a full length COB protein, we conclude that the *cob* gene fragment in the *cox3* cDNA is found in at least one other context in the mitochondrial genome. Southern blots using both *cox3* and *cob* sequences as probes of restriction enzyme digests which should not cut between them showed that numerous bands unique to each probe were present. We conclude that additional copies of both gene fragments are indeed present.

Introduction

Dinoflagellates are a diverse group of photosynthetic and autotrophic protists linked together by a number of unusual morphological traits. For example, dinoflagellates do not have histones [22], typically have chromosomes which stay condensed throughout the cell cycle [21], and their nuclear envelope does not break down during mitosis [27]. These dissimilarities with other algal groups has made phylogenetic studies based on morphology difficult, and the evolutionary position of the dinoflagellates is now best understood through molecular sequence analysis. Indeed, molecular phylogenies using several types of molecules are converging on a uniform picture of the evolutionary position of dinoflagellates. Trees based on nuclear rRNA sequences, for example, suggest that dinoflagellates are closely related to apicomplexans (a group of parasitic protists that includes *Plasmodium*) [13], and the sequence of two other nuclear encoded proteins (cAPK [23] and GAPDH [5]) is generally in good agreement with this evaluation, as is the phylogenetic analysis of a recently reported mitochondrial-encoded *cox1* gene [18]. Thus all the available data is consistent with a close relationship between the dinoflagellates and the apicomplexans.

In addition to the unusual morphological characteristics of the dinoflagellate nucleus, several features of nuclear gene expression are also unusual. For example, putative promoter sequences (up to 1 kb upstream of the mRNA start site) have no conserved sequences with any other eukaryotic promoters [12, 14], including a lack of what are considered to be necessary elements for indicating transcriptional start sites, such as the TATA box. As a caveat, however, it must be noted that no putative dinoflagellate promoter sequence has yet been shown to actually possess promoter activity, unlike the similarly unusual diatom *fcp* promoter which has been shown to drive reporter gene expression in transformed algae [2]. Nuclear encoded messages have a defined 3' terminal end, which includes a 50-150 base 3'UTR and as a minimum, a 15-30 base poly(A) tract [14, 16, 17]. However, there are no conserved consensus sequence elements for

polyadenylation in any of the dinoflagellate mRNAs analyzed to date. It is not known if there are specific sites for transcriptional termination or if the primary transcript is processed to defined lengths by endonucleolytic events.

Even less is known about organellar gene expression in dinoflagellates. There are no known plastid gene sequences available, and only one mitochondrial sequence, a cytochrome oxidase subunit 1 (*cox1*), has been reported to date [10, 18]. Analysis of this mitochondrial sequence [18] has suggested that dinoflagellate mitochondria use the universal genetic code, including initiation and termination codons, but has not provided any information on the generation or processing of the mitochondrial transcripts. Given the close phylogenetic relationship between dinoflagellates and apicomplexans, we were interested in examining mitochondrial transcripts for common features to mitochondrial gene expression in these two groups.

We have isolated several copies of a mitochondrial *cox3* transcript from a *Gonyaulax* cDNA library, and analysis of the sequence reveals some unexpected aspects to mitochondrial genes and their expression in dinoflagellates. First, dinoflagellate transcripts are polyadenylated, as are mitochondrial transcripts in other eukaryotes and some mitochondria and prokaryotes [25]. Second, *Gonyaulax* transcripts are of variable length, which we interpret as being a result of indiscriminate polyadenylation. This is unlike polyadenylation in other mitochondria, where while polyadenylation seems to occur at any 3' end, the 3' ends themselves are generated by precise endonucleolytic events [3, 7, 11]. Lastly, we have also found a *cox3* transcript with a 1 kb 3' extension. This cDNA contained portions of both *cox1* and *cytochrome b* (*cob*) genes, as though the mitochondrial genome downstream of the *cox3* sequence contained numerous gene fragments.

Methods

Cell culture

Gonyaulax polyedra (strain70) cultures, obtained from the Provasoli-Guillard Culture Collection (Booth Bay, Maine), were grown under a 12:12 L:D photoperiod (150 $\mu\text{mol m}^{-2} \text{s}^{-1}$) in a supplemented seawater medium [8] until cell densities of roughly 5,000 cells /mL were attained. Cells were collected by vacuum filtration on Whatman 541 paper and frozen until use.

Library construction and screening

RNA was extracted and purified from frozen *Gonyaulax* cells using a GIT buffer, LiCl precipitation and oligo dT chromatography essentially as described [23]. As the only modification, a cetyl triammonium bromide extraction was performed after precipitation by sodium acetate and prior to precipitation by LiCl. This extraction, to remove polysaccharides which inhibit the action of the reverse transcriptase, was carried out by resuspending the pellet of nucleic acids in 0.8 M NaCl. To this solution is added 1.1 volumes of hot (65°C) CTAB buffer (0.2 M Tris, pH 8 containing 70 mM EDTA, 2M NaCl and 55 mM CTAB). After a 30 minute incubation at 65°C, the solution is extracted once with 1 volume of chloroform. The aqueous phase is precipitated with sodium acetate and ethanol, and resuspended in DEPC-treated water before precipitation with LiCl. All RNA was prepared from cells harvested at LDT 2 (2 hours after lights on). A directionally cloned cDNA library was constructed in the lambda ZAP II vector (Stratagene, CA) following the manufacturer's instructions. The library contained 1.4×10^6 independent clones. Individual *cox3* cDNAs were isolated from the library by plaque screening [24] using hybridization to a *cox3* cDNA probe which had been obtained by sequencing of randomly isolated clones.

PCR and DNA Sequencing

PCR reactions were carried out in a 50 μL volume containing 100 pmol each oligonucleotide primer, 200 μM dNTP, Taq DNA polymerase buffer (Amersham) and 1.5

units Taq DNA polymerase (prepared as described [20]). For phage templates, roughly 10^5 pfu were used per reaction. The length of the different cDNAs were analyzed by amplification between the 5' directed 5'-caaagatcattgattggtccc-3' and the vector T3 oligonucleotides, or between the 3' directed 5'-tcttattaattgtttgtt-3' and the vector T7 oligonucleotides. Double stranded plasmid DNA was sequenced at the University of Calgary Core DNA Service facility (Calgary, Alberta), and the sequences were analyzed using the MacVector program (Oxford Molecular Group). Sequence comparisons to GENE BANK were carried out using the BLAST algorithm [1].

Southern and Northern hybridizations

RNA and DNA, both prepared using a CTAB precipitation step, were analyzed on agarose gels using standard procedures [24]. Hybridization was performed using the phosphate-SDS-BSA method [4]. Probes were prepared from different regions of the *cox3* cDNA by random primer labeling (Prime a gene, Promega Biotech) after restriction of the cDNA, purification of the different fragments by agarose gel electrophoresis, and electroelution of the DNA [24]. The *cox3* probe was prepared from a 730 bp EcoRI/EcoRV restriction fragment corresponding to the 5' end of the cDNA, while the cytochrome b probe was prepared from a 500 bp EcoRI/XhoI fragment from the 3' end of the cDNA.

Results

Isolation of a dinoflagellate *cox3* cDNA

An AT-rich cDNA was detected in a *Gonyaulax* cDNA library after sequencing of randomly selected dinoflagellate cDNAs. This 65% AT-rich sequence was substantially different from previously sequenced dinoflagellate cDNAs which are typically 65% GC-rich [14, 15, 17, 23]. It seemed possible that this cDNA represented a mitochondrial or plastid-encoded gene, and so further studies were undertaken to characterize this cDNA.

The sequence was first compared to known sequences in GENE BANK using the BLAST algorithm [1]. The highest scores were obtained with a cytochrome oxidase subunit 3 (*cox3*) sequence from various organisms, suggesting that the cDNA might encode a dinoflagellate *cox3*. This similarity is shown in a more detailed comparison of the deduced amino acid sequences (Fig. 1). The sequence most closely resembles *cox3* in the middle half (amino acids 82 through 219), while both N- and C-terminal regions are highly divergent. Indeed, while the total length of the deduced protein is consistent with that of COX3, this dinoflagellate sequence has a longer N-terminal and a shorter C-terminal than other COX3 sequences. A pairwise comparison of the amino acid identities of several COX3 sequences (Table I) shows that the similarity of the dinoflagellate sequence to that of other organisms, although low, is in the same range as the similarity between the *Plasmodium* COX3 and other COX3 sequences. Our cDNA either encodes a COX3 subunit, or encodes a *cox3* gene fragment which is found in an open reading frame coincidentally of the same size as COX3.

Polyadenylation stretches are added to the transcript at random sites

A mitochondrial protein encoded by the nucleus would be expected to contain a mitochondrial signal sequence. Since the presence of such a sequence would immediately rule out the possibility that the gene was encoded by the mitochondria, four additional clones were isolated and analyzed for the presence of a 5' extension. PCR analysis, using a 5' oligonucleotide paired with the T3 sequence in the vector, showed that the length of the

5' end was constant (Fig. 2), showing that there was no sequence corresponding to a mitochondrial signal peptide. We found no AUG initiation codon in the *cox3* sequence, suggesting either that this sequence was missing from even the longest clone, that translation may begin at the 5' end of transcript without an AUG as is the case for animal mitochondria [19], or that translation may begin at non-universal initiation codons as found in ciliates or apicomplexans [6]. Independent of the translation of the transcript, it appears as though the cDNAs arose from mRNAs which either had the same transcriptional start site or shared a common site for endonucleolytic processing.

Surprisingly, PCR analysis using a 3' directed oligonucleotide paired with the vector T7 sequence showed that the length of the 3' end was highly variable (Fig. 2). To determine the nature of these differences, the longest (*cox3.11*) and one of the shortest (*cox3.10*) cDNAs were sequenced and compared to the sequence of the original isolate (*cox3.2*) (Fig 3). The sequences were similar but not identical in the regions of overlap, and all had a poly(A) tail. Because the site of polyadenylation of the shortest sequence (*cox3.10*) coincided with an adenine rich region in the *cox3.2* sequence, it is possible that the tract of adenine residues at the 3' end of the short sequence might be an artifact of the oligo(dT) primer used for first strand cDNA synthesis instead of a true polyadenylation. However, no adenine residues are found in the *cox3.11* sequence at the putative polyadenylation site of the *cox3.2*, suggesting that this tract of adenine residues has been added post-transcriptionally.

To test the possibility that polyadenylation could occur at many different sites, the length of *cox3* containing transcripts was evaluated by Northern blots (Fig. 4). We observe transcripts hybridizing to the *cox3* probe in both total RNA (lane 3) and in poly(A) enriched RNA (lane 1) but not in poly(A) depleted RNA (lane 2), similar to transcripts hybridizing to the nuclear encoded *pcp* probe [12]. This indicates that all *cox3* containing transcripts are polyadenylated. We also observe that the size of the transcripts hybridizing to the *cox3* probe is more variable those hybridizing to the PCP probe. This is consistent

with a polyadenylation of different transcripts from the same gene at different sites, although we cannot rule out the possibility that the RNA contains many different *cox3* containing sequences all derived from different genomic contexts. The size of the transcripts in the region of the strongest hybridization signal is roughly 1.2 kb, similar to the *cox3.2* sequence, and is thus long enough to encode a COX3 subunit. Lastly, based on the fact that the *cox3* probe was exposed to film three times longer than the *pcp* probe, we conclude that the levels of the *cox3* transcript are one quarter to a third those of the *pcp* transcript (roughly 1% of the total message [12]). The *cox3* containing transcripts are thus relatively abundant.

The mitochondrial genome contains gene fragments

The hypothesis of random addition of poly(A) tails also suggests that some transcripts might terminate considerably downstream from the *cox3*. Thus, the long *cox3.11* cDNA might contain part of an open reading frame immediately downstream from the *cox3* sequence. To test this, BLAST searches were used to screen the gene bank data using sequence found in the 3' extension of the *cox3.11*. Significant levels of sequence similarity were found with the *cox1* gene of the dinoflagellate *Cryptothecodinium* (76 % sequence identity over 135 amino acids) and both the *cox1* and *cob* sequences of *Plasmodium* (62 % sequence identity over 113 amino acids, and 48 % sequence identity over 168 amino acids, respectively). Curiously, however, the *cox3.11* sequence did not contain full length copies of either sequence with only 45 % and 25 % of the *cox1* and *cob* sequences, respectively, being found. The size and order of the gene fragments found in the *cox3.11* cDNA are shown compared to the *Plasmodium* mitochondrial genome in Figure 5.

Several possibilities were considered to explain the presence of *cox1* and *cob* gene fragments. First, it was possible that these sequences represented unspliced exons, and that the other regions of the genes were found elsewhere in the mitochondrial genome. This seems unlikely, given that the *cox3.11* cDNAs were synthesized from a processed (at least

polyadenylated) transcript. Secondly, it was possible that these sequences were incomplete gene duplications, and that complete ORFs for the *cox1* and *cob* sequences would be found elsewhere in the genome. As one test of this hypothesis, Northern blots were probed with a *cob* sequence. If *cob* sequences were found only in the *cox3.11* context, only transcripts longer than 1.4 kb would be observed, whereas if *cob* sequences were found in other contexts we expected smaller transcripts to be present. The results (Fig. 4) show that, as for the *cox3* probe, the *cob* probe hybridizes to polyadenylated RNA with a wide range of transcript sizes. While the longest transcript (1.5 kb) is consistent with that expected based on the *cox3.11* sequence, the majority of the hybridization occurs with a polyadenylated RNA of 1 kb, consistent with the size of the *Plasmodium cob* transcript [26]. This transcript could be derived from the *cob* sequence found in *cox3.11* if a 5' end were generated by endonucleolytic cleavage just upstream of the *cob* sequence. However, in the absence of any recognizable motif for signaling the site of cleavage, it is more likely that this RNA is derived from transcription of a different *cob* gene or gene fragment elsewhere in the mitochondrial genome. This data supports the hypothesis that the *cob* sequence will be found in at least one additional context in the mitochondrial genome.

To confirm that *cob* sequences in addition to those adjacent to the *cox3* sequence were present in the genome, probes directed against both the *cox3* and *cob* sequences were tested separately on Southern blots. When enzymes without a restriction site in the *cox3.11* sequence were examined (Pst I and Pvu II), only one restriction fragment out of several in each digest was found to hybridize to both probes (Fig. 6, lanes Ps and Pv). The restriction fragment hybridizing to both probes presumably arises from DNA containing both *cox3* and *cob* sequences, as would be found in the *cox3.11* sequence. However, since both digests contain several other bands, we conclude that there are additional copies of both *cox3* and *cob* sequences in different contexts in the genome. Furthermore, when the restriction patterns obtained with both probes hybridized to an Eco RI digest are compared, we observe that one band apparently hybridizes to both probes but that there are several

other bands. Since the *cox3.11* restriction map (Fig. 6 top) predicts no restriction fragment will hybridize to both the *cox3* and *cob* probes, and only a single fragment of 1.15 kb will hybridize to the *cox3* probe, we conclude that *cox3* sequences as well are found in several other genomic contexts.

Discussion

We report here the sequence of a AT-rich dinoflagellate cDNA isolated from a library constructed from poly(A) enriched *Gonyaulax* RNA. This is the first report of a cDNA sequence from dinoflagellate mitochondria. The evidence supporting the identification of the sequence as a *cox3* cDNA includes the similarity to other *cox3* sequences (which, like the *Plasmodium cox3*, is nonetheless low; Table I), the number of times sequences with a common 5' end were isolated from the library (Fig. 2; 3) and the quantity and size (sufficient to encode a COX3 subunit) of the message by Northern analysis (Fig. 4). As a caveat, we note that our Northern blots cannot distinguish between hybridization to a sequence similar to that described here and hybridization to any other transcript of the same size which contains the *cox3* sequence used as a probe. It is thus possible that the *cox3* sequence reported here reflects expression of a gene fragment, similar to the gene fragments of *cox1* and *cob*, and that a different COX3 is expressed from a different *cox3* containing transcript. Although a definitive resolution to this issue will have to await the sequence of the entire dinoflagellate mitochondrial genome, we note that it would be unusual to maintain transcription of a sequence never destined to be translated.

The similarity of the dinoflagellate *cox3.11* sequence to the sequence of the *Plasmodium* mitochondrial genome is intriguing. First, the *cox3* sequence itself consists of an ORF with less sequence identity to other *cox3* genes than is typically found (Table I). Second, there is a similar order to the gene fragments present (i.e., *cox3* followed by *cox1* and then *cyb*). Lastly, the *cox1* sequence described in other dinoflagellates (a fragment only

in *cox3.11*) has been used to place dinoflagellates phylogenetically close to apicomplexans (including *Plasmodium*) [10, 18]. The *cox1* fragment in *cox3.11* is very similar to the genomic *cox1* sequence reported for the dinoflagellates *Prorocentrum* and *Cryptothecodinium* (95 and 85% sequence identity at the nucleic acid level, respectively), indicating that there is no editing of at least this region of the sequence. However, since both *Plasmodium* and *Gonyaulax* sequences are presumably highly derived, it is unlikely that the similarity in gene order will have much phylogenetic significance.

The presence of a full length *cox1* in the genomic DNA of two other dinoflagellates indicates that the *cox1* sequence in the *cox3.11* clone is a gene fragment and that a complete *cox1* gene should be found elsewhere in the mitochondrial genome. Hybridizations using the *cox1* fragment on Southern blots were unsuccessful, presumably due to the small size of the probe, but hybridizations with either *cox3* or *cob* sequences gave strong and reliable signals. Our analysis of the Southern blots (Figure 6) indicates that many different *cox3* and *cob* sequences are present in the mitochondrial genome. This conclusion is based principally on the observation that the two probes hybridize to several fragments of different sizes, even when the enzymes used do not have sites within the *cox3.11* sequence. We note that there is one band in each of the digests which appears to hybridize to both probes, consistent with the presence of the *cox3.11* sequence in the mitochondrial genome. This indicates that the mitochondrial genome will contain several different copies of these genes, including presumably a complete *cob* ORF. It is impossible to differentiate between a large genome with many different copies of these sequences and the presence of many smaller genomes distributed among the many mitochondria present. However, the finding of gene fragments in transcripts suggests that there is no system for splicing together gene fragments into complete gene copies.

The Northern blots (Fig. 4) show that both the *cox3* and *cob* probes hybridize to a larger range of transcript sizes than does the nuclear-encoded PCP gene probe. Furthermore, cDNAs, of lengths ranging from 0.7 to 1.7 kb have been isolated by

hybridization to a *cox3* probe. This suggests that dinoflagellate mitochondrial transcripts apparently contain no signal for generating a precise 3' end by transcriptional termination or endonucleolytic cleavage. The truncated cDNAs (*cox3.10*) may be similar to the truncated and polyadenylated *rpsO* transcripts of *E. coli* [9]. The possibility that the different cDNA lengths represent an artifact of the oligo(dT) primed cDNA synthesis and the fact that the mitochondrial genome is AT rich is ruled out by the polyadenylation at positions where there are no underlying tracts of adenine residues (compare *cox3.2* and *cox3.11* sequences in Fig. 3). Furthermore, all transcripts are apparently polyadenylated since little hybridization signal remains in poly(A) depleted RNA samples (Fig. 4). These dinoflagellate transcripts are thus unlike the mitochondrial transcripts of *Plasmodium*, which undergo precise processing from an original polycistronic message. Although we cannot determine how the 5' end is formed, the lack of precise endonucleolytic cleavage during generation of the 3' end suggests that there may be a defined origin of transcription in the mitochondria.

Acknowledgments

We thank Drs. M. Gray and F. Lang for critical reading of the manuscript. This work has been funded by an NSERC operating grant (DM) and research studentship (HC).

References

1. Altschul, SF, Gish, W, Miller, EW Lipman, DJ: Basic local alignment search tool. J Mol Biol 215: 403-410 (1990).
2. Apt, K, Kroth-Pancic, P Grossman, A: Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Mol. Gen. Genet. 252: 572-579 (1996).
3. Burger, G, Helmer-Citterich, M, Nelson, M, Werner, S Macino, G: RNA processing in *Neurospora crassa* mitochondria: transfer RNAs punctuate a large precursor transcript. EMBO J. 4: 197-204 (1985).
4. Church, GM Gilbert, W: Genomic Sequencing. Proc. Natl. Acad. Sci. USA 81: 1991-1995 (1984).
5. Fagin, T, Hastings, J Morse, D: Phylogeny of Glyceraldehyde-3-phosphate dehydrogenase indicates lateral gene transfer from cryptomonads to dinoflagellates. J. Mol. Evol. (in press): (1998).
6. Feagin, J: The extrachromosomal DNAs of apicomplexan parasites. Annu. Rev. Microbiol. 48: 81-104 (1994).
7. Gray, M Boer, P: Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. Philos. Trans. R. Soc. Lond. Biol. Sci. 319: 135-147 (1988).
8. Guillard, RRL Ryther, JH: Studies on marine planktonic diatoms: *Cyclotella nana* Hufstedt and *Denotula confervacea* (Cleve) Gran. Can. J. Microbiol. 8: 229-239 (1962).
9. Haugel-Neilsen, J, Hajnsdorf, E Régnier, P: EMBO J 15: 3144-52 (1996).
10. Inagaki, Y, Hayashi-Ishimaru, Y, Ehara, M, Igarishi, I Ohama, T: Algae or protozoa: phylogenetic position of euglenophytes and dinoflagellates as inferred from mitochondrial sequences. J. Mol. Evol. 45: 295-300 (1997).
11. Ji, Y, Mericle, B, Rehkopf, D, Anderson, J Feagin, J: The *Plasmodium falciparum* 6 kb element is polycistronically transcribed. Mol. Biochem. Parasitol. 81: 211-223 (1996).

12. Le, QH, Markovic, P, Jovine, R, Hastings, J Morse, D: Sequence and genomic organization of the peridinin-chlorophyll a-protein from *Gonyaulax polyedra*. *Mol. Gen. Genet.* 255: 595-604 (1997).
13. Lenaers, G, Scholin, C, Bhaud, Y, Saint-Hilaire, D Herzog, M: A molecular phylogeny of dinoflagellate protists (Pyrrhophyta) inferred from the sequence of 24S rRNA divergent domains D1 and D8. *J. Mol. Evol.* 31: 53-63 (1991).
14. Li, L Hastings, J: The structure and organization of the luciferase gene in the photosynthetic dinoflagellate *Gonyaulax polyedra*. *Plant Mol. Biol.* 36: 275-284 (1998).
15. Machabée, S, Wall, L Morse, D: Expression and genomic organization of a dinoflagellate gene family. *Plant Mol. Biol.* 25: 23-31 (1994).
16. Morse, D, Milos, PM, Roux, E Hastings, JW: Circadian regulation of bioluminescence in *Gonyaulax* involves translational control. *Proc. Natl. Acad. Sci. USA* 86: 172-176 (1989).
17. Morse, D, Salois, P, Markovic, P Hastings, JW: A nuclear encoded form II rubisco in dinoflagellates. *Science* 268: 1622-1624 (1995).
18. Norman, J Gray, M: The cytochrome oxidase subunit 1 gene (coxI) from the dinoflagellate *Cryptocodinium cohnii*. *FEBS letters* 413: 333-338 (1997).
19. Ojala, D, Montoya, J Attardi, G: tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290: 470-474 (1981).
20. Pluthero, F: Rapid purification of high activity Taq DNA polymerase. *Nucl. Acids Res.* 21: 4850-4851 (1993).
21. Rizzo, P: The enigma of the dinoflagellate chromosome. *J. Protozool.* 38: 246-252 (1991).
22. Rizzo, PJ: Comparative aspects of basic chromatin proteins in dinoflagellates. *Biosystems* 14: 433-443 (1981).
23. Salois, P Morse, D: Characterization and molecular phylogeny of a protein kinase cDNA from the dinoflagellate *Gonyaulax*. *J. Phycology* 33: 1063-1072 (1997).
24. Sambrook, J, Fritsch, EF Maniatis, T *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, (1989).

25. Sarkar, N: Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.* 66: 173-197 (1997).
26. Suplick, K, Morrisey, J Vaida, A: Complex transcription from the extrachromosomal DNA encoding mitochondrial function of *Plasmodium yoelii*. *Mol. Cell Biol.* 10: 6381-6388 (1990).
27. Triemer, R Fritz, L (1984) Cell cycle and mitosis

Gonyaulax 1 G I R H E D L C F L C F I E E I S A W R L V F W N S I H L E V A S D S F C F L L 40
Reclinomonas 1 M S Q T F V K K H P Y H I V D 15
fission yeast 1 M N T S T K F Q G H P Y H I V S 16

Gonyaulax 41 Y S S W L I V F V L C V F E H Y F S F I N L Y C W K G L H F S W N N F L I F I F 80
Reclinomonas 16 Q S P W P L L T S I G T L C S T F G G V - - - - M Y F H S Y P N G G F I A A L 50
fission yeast 17 A S P W P F F L S V V L F F N C L A A T - - - - L Y L H G Y K H S S V F F G I 51

Gonyaulax 81 I V T G L I F S F H F W F R D L L R E L A K K - - - Y E V L L I C L F L L F F L 117
Reclinomonas 51 G V S T I L F S L Y A W C R D I V R E G T Y Q G Q H T A A V Q N G L R I G M I L 90
fission yeast 52 S F L G L L A T M Y L W F R D M S T E A N I H G A H T K A V T K G L K M G F M L 91

Gonyaulax 118 F L A S E G L L F V S F F W A S F H S L S S P T L R V C L G E G F Y - - - - V 152
Reclinomonas 91 F I I S E V M F F V S F F W A F F H S S L S P T I E I G A V W P P Q - - G I E T 128
fission yeast 92 F T I S E T F L F A S I F W A F F H S S L S P T F E L G A V W P P V G M A D K T 131

Gonyaulax 153 P D P C E L T F A N T L L L S N A A I S L G N A F I S L E I S S - Q Y L I F F S 191
Reclinomonas 129 L N A W D V P F L N T V I L L M S G A T V T W S H H A M I H G N R T Q S I L G L 168
fission yeast 132 M D P L E V P T L N T V I L L T S G A S L T Y A H Y S L I A R N R E N A L K G L 171

Gonyaulax 192 L F S F I L A W T F I S L Q I K E F R I I S I I S G N F T K N N L W E P I N D 231
Reclinomonas 169 I F T I I L A V T F T G L Q V M E Y R E A S F S I A D G I Y G S T F Y M A T G F 208
fission yeast 172 Y M T I A L S F L F L G G Q A Y E Y W N A P F T I S D S V Y G A S F Y F A T G T 211

Gonyaulax 232 L C H E L Y F S S W K C C L G T S Y V Y C R F R N 256
Reclinomonas 209 H G F H V I V G T C M L S V C L V R E Y L Y H F T T T H H F G F E A S A W Y W H 248
fission yeast 212 H G I H I I V G T I T T T A A T Y N I Y T Y H L T N T H H N G F E C G I Y Y W H 251

Reclinomonas 249 F V D V V W L F L F T T I Y W W G G N 267
fission yeast 252 F C D V V W L F T Y L T I Y I W G S 269

Figure 1
Chaput and Morse

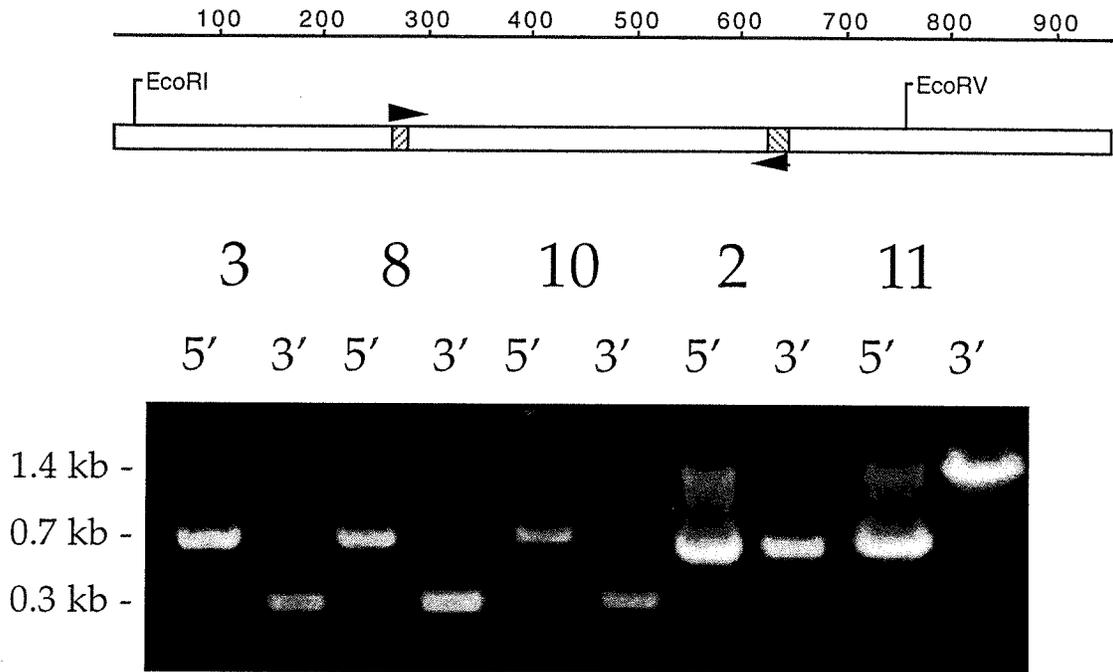


Figure 2
Chaput and Morse

cox3.10 1 GGCACGAACCTGGAGAAAGA CTTTGGTTTTCTTTGTTTTATTGAAGA A ATCTCTGCCTGGA 60
 cox3.2 1 CCTTGGGA 7
 cox3.11 1 CTTTGGTTTTCTTTGTTTTATTGAAGA G ATCTCTGCCTGGA 51

cox3.10 61 GACTTGTCTTTTGGAAATTCATTTCATCTGGAAGTTGCATCAG GTC CATTTTGTTTTCTTTG 120
 cox3.2 8 GACTTGTCTTTTGGAAATTCATTTCATCTGGAAGTTGCATCAGATTCAATTTTGTTCCTTC 67
 cox3.11 52 GACTTGTCTTTTGGAAATTCATTTCATCTGGAAGTTGCATCAGATTCAATTTTGTTCCTTC 111

cox3.10 121 TTTATAGTTCTTTGGTTAATAGTCTTTTGTTCCTATGCGTCTTTGAAACATTACTTCTCTTTTA 180
 cox3.2 68 TTTATAGTTCTTTGGTTAATAGTCTTTTGTTCCTATGCGTCTTTGAAACATTACTTCTCTTTTA 127
 cox3.11 112 TTTATAGTTCTTTGGTTAATAGTCTTTTGTTCCTATGCGTCTTTGAAACATTACTTCTCTTTTA 171

cox3.10 181 TTAATCTCTATTGCTGGAAAGGACTTCACTTCTCATGGAAATAAATTTTCTTATCTTTATCT 240
 cox3.2 128 TTAATCTCTATTGCTGGAAAGGACTTCACTTCTCATGGAAATAAATTTTCTTATCTTTATCT 187
 cox3.11 172 TTAATCTCTATTGCTGGAAAGGACTTCACTTCTCATGGAAATAAATTTTCTTATCTTTATCT 231

cox3.10 241 TTATTGTCAACAGGGCTTATCTTTAGTTTTCAATTTTGGTTTCAGAGATTTATTAAGAGAAAT 300
 cox3.2 188 TTATTGTCAACAGGGCTTATCTTTAGTTTTCAATTTTGGTTTCAGAGATTTATTAAGAGAAAT 247
 cox3.11 232 TTATTGTCAACAGGGCTTATCTTTAGTTTTCAATTTTGGTTTCAGAGATTTATTAAGAGAAAT 291

cox3.10 301 TAGCTAAAAAATATGAAGTCTTATTAATTTGTTTTGTTTTCTTCTTTTTTTTCTATTTCCTTG 360
 cox3.2 248 TAGCTAAAAAATATGAAGTCTTATTAATTTGTTTTGTTTTCTTCTTTTTTTTTCTATTTCCTTG 307
 cox3.11 292 TAGCTAAAAAATATGAAGTCTTATTAATTTGTTTTGTTTTCTTCTTTTTTTTTTCTATTTCCTTG 351

cox3.10 361 CTTCTGAAAGCTCTGTGTTATTTATATCCCTTCTTTTGGGCATCTTTTCATTCATTATCTTCTG 420
 cox3.2 308 CTTCTGAAAGGCTCTGTGTTATTTGTATCCCTTCTTTTGGGCATCTTTTCATTCATTATCTTCTC 367
 cox3.11 352 CTTCTGAAAGGCTCTGTGTTATTTGTATCCCTTCTTTTGGGCATCTTTTCATTCATTATCTTCTC 411

cox3.10 421 CAACTTTAAGGGTTTGTCTTTGGGAGAAGGTTTCTATGTTCCAGATCCTTGTGAATTAACCTT 480
 cox3.2 368 CAACTTTAAGGGTTTGTCTTTGGGAGAAGGTTTCTATGTTCCAGATCCTTGTGAATTAACCTT 427
 cox3.11 412 CAACTTTAAGGGTTTGTCTTTGGGAGAAGGTTTCTATGTTCCAGATCCTTGTGAATTAACCTT 471

cox3.10 481 TTGCTAATACACTTCTTTTATCTAATGCTGCTGTAICTTTAGGA GGTGCTTTTGTAAAGTT 540
 cox3.2 428 TTGCTAATACACTTCTTTTATCTAATGCTGCTGTAICTTTAGGAAATGCTTTTATAAGTT 487
 cox3.11 472 TTGCTAATACACTTCTTTTATCTAATGCTGCTATATCTTTAGGAAATGCTTTTATAAGTT 531

cox3.10 541 TAGAAATTTTCATCACAATATTTGATTTTCTTTTCATTATTCTCTTTCATACTAGCTTGGAA 600
 cox3.2 488 TAGAAATTTTCATCACAATATTTGATTTTCTTTTCATTATTCTCTTTCATACTAGCTTGGAA 547
 cox3.11 532 TAGAAATTTTCATCACAATATTTGATTTTCTTTTCATTATTCTCTTTCATACTAGCTTGGAA 591

cox3.10 601 CATTATATTAGTTTGGCAAATCAAAGAATTTCCGGATCATTAGCATAAATAATTTCTGGAGATT 660
 cox3.2 548 CATTATATTAGTTTGGCAAATCAAAGAATTTCCGGATCAATAGCATAAATAATTTCTGGGAATT 607
 cox3.11 592 CATTATATTAGTTTGGCAAATCAAAGAATTTCCGGATCATTAGCATAAATAATTTCTGGGAATT 651

cox3.10 661 TCACAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 681
 cox3.2 608 TCACAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 667
 cox3.11 652 TCACAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 711

cox3.2 668 CTTGGAAGTGTGTTTGGGGACATCATATGTATACTGTAGGTTTGAAGAACTGATACAAGA 727
 cox3.11 712 CTTGGAAGTGTGTTTGGGGACATCATATGTATACTGTAGGTTTGAAGAACTGATACAAGA 771

cox3.2 728 GCTTATTTTACAGGAGTTACAATCTTGATATCCTTACCGACAGGTACAATAAATCTTTAAAT 787
 cox3.11 772 GCTTATTTTACAGGAGTTACAATCTTGATATCCTTACCGACAGGTACAATAAATCTTTAAAT 831

cox3.2 788 TGGTTAAGTACATATCTTGGGAATCCGCCATTGTTACACCTCAAAACTACTTCAGCATTTC 847
 cox3.11 832 TGGTTAAGTACATATCTTGGGAATCCGCCATTGTTACACCTCAAAACTACTTCAGCATTTC 891

cox3.2 848 TTTTGGACTTCTCTTTTATTAATGTTTTCTATAGGTGGCTCAACAGGAGTAATTCTTTGGA 907
 cox3.11 892 TTTTGGACTTCTCTTTTATTAATGTTTTCTATAGGTGGCTCAACAGGAGTAATTCTTTGGA 951

cox3.2 908 AATGCTGCTGTGGATCTAGGTTTACATGAAATAAATAAATAAATAAATAAATAAATAAATAAATA 950
 cox3.11 952 AATGCTGCTGTGGATCTAGGTTTACATGAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1010

cox3.11 1011 TCCTATGAACAGACAGCTCTAATGTTAGAAATCATTCTCCTGGTCATTTGGTAGATTTTTT 1070
 cox3.11 1071 CTTTTTTCTAATTTAATGAGCTAGGAAGCAGTTTCTCTGTTGTTAATTTTCATTTCTGTGG 1130
 cox3.11 1131 CTGCAAAATAACAATAATGTTTACAGTCCAACATTGAAGAGGTTCTTTCTCTTTCAATTTCA 1190
 cox3.11 1191 ATTCCCATTCTTTCTTTGTTGTTCTAGAAATTCGTCATCTTTTTAATCGTCATTGTCAATC 1250
 cox3.11 1251 TTCTAGGAATCCTGTGAGGAATTC AACCAAAAAATAAATCCCATTTTTCCCTTTCAATTT 1310
 cox3.11 1311 TCAAAAAGATTTCTTTGGTTCTATTCTAATTCCTCGGCTATATGTTCTTCAAACCTCAATTT 1370
 cox3.11 1371 CGGGGTCTCTTTCTTTATCACATCCAGATAAATGCCCTTAGAAGCTTGTGGATTACTTACTCC 1430
 cox3.11 1431 TTTACATATAGTACCTGAATGGTATTTCTCTATGCCAATATGCTATGTTAAAAGCTGTGCC 1490
 cox3.11 1491 CAACAAAAATGCAGGATTCATAATCTTATTAACCTTCTATCTTCACTTTCTTTTTCTTGGG 1550
 cox3.11 1551 AGAAATTAGAAATCTAACAACCTTGTGTCTCTTACCTTTGGTCAATAATGGTTTTTCTCTT 1610
 cox3.11 1611 TCATTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1670
 cox3.11 1671 AGAAGTTTCTATCTTAATGCTCGTATCTTGACATTAATATTATTATTATTCTTTACTTATGTGT 1730
 cox3.11 1731 ATCTTATTTTTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1759

Figure 3

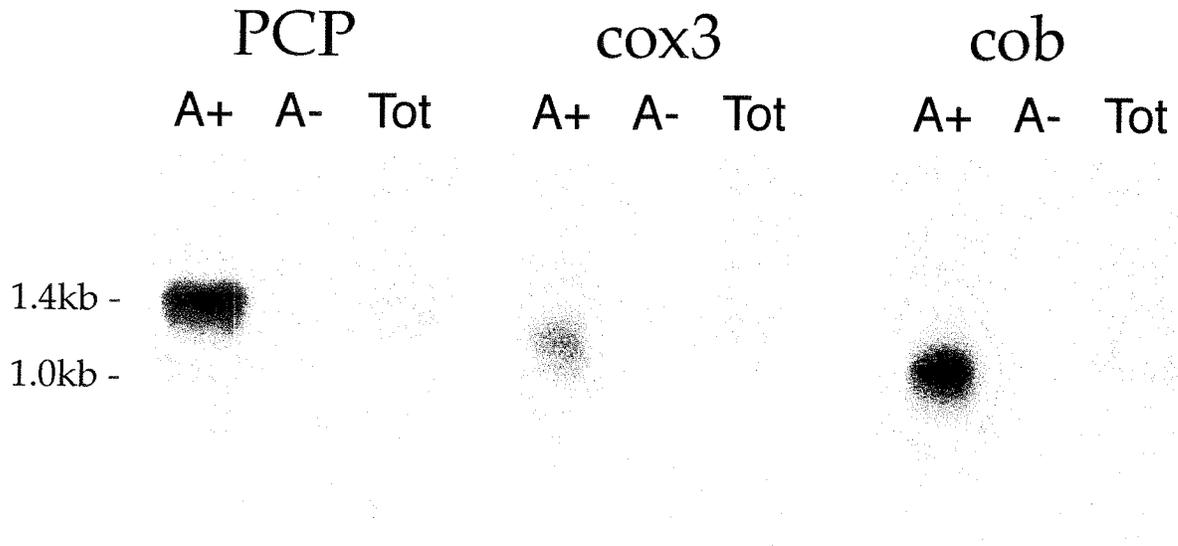


Figure 4
Chaput and Morse

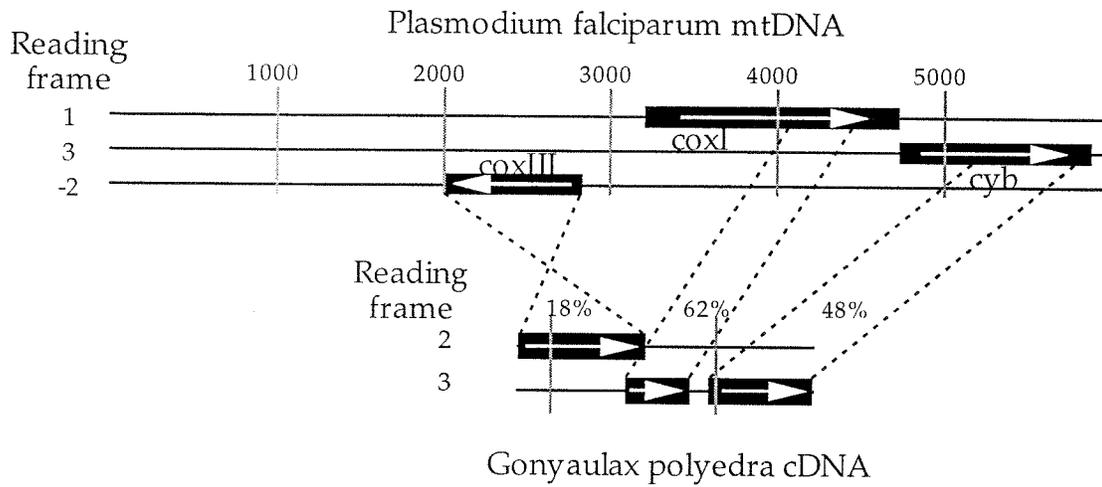


Figure 5
Chaput and Morse

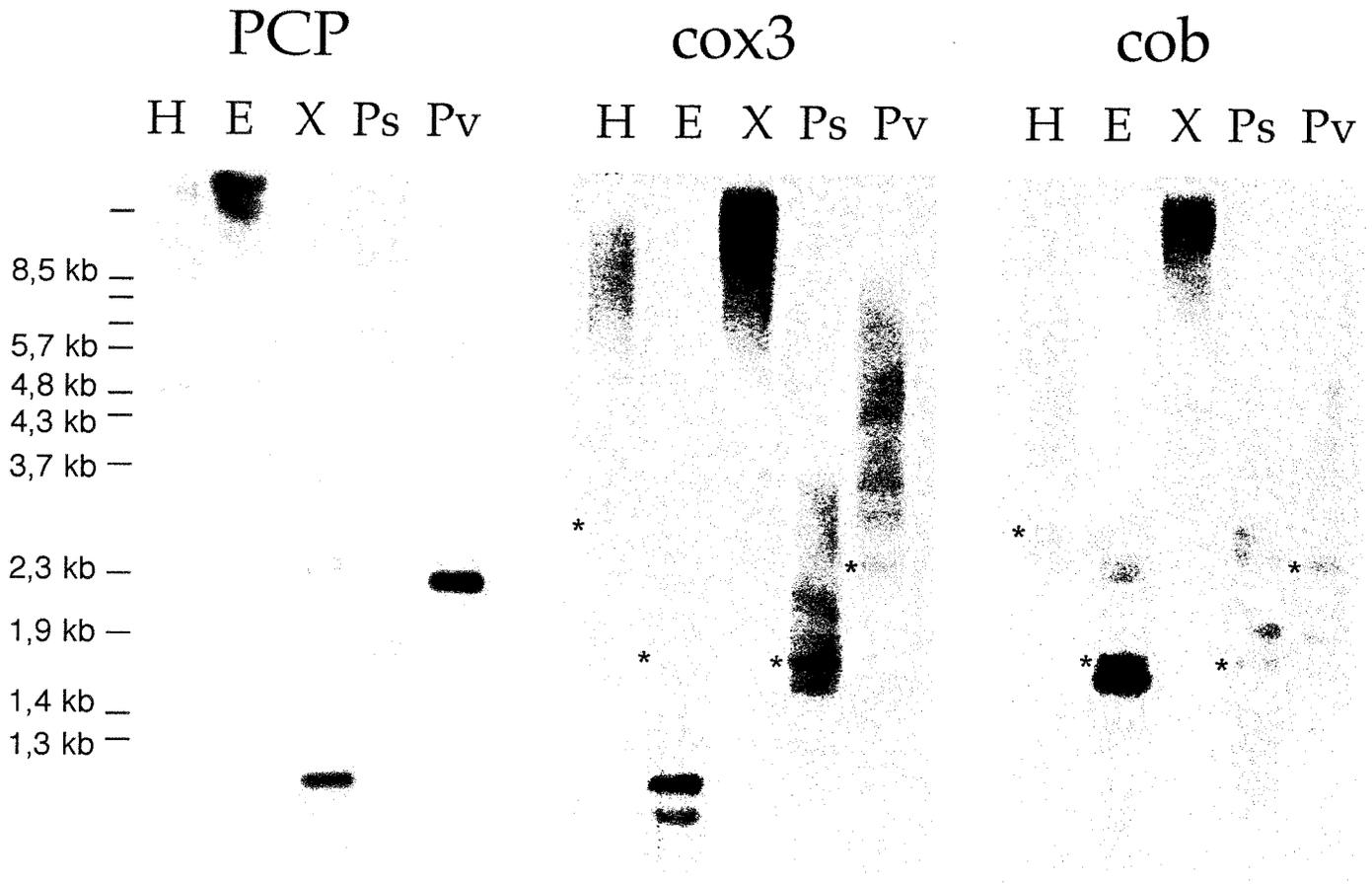
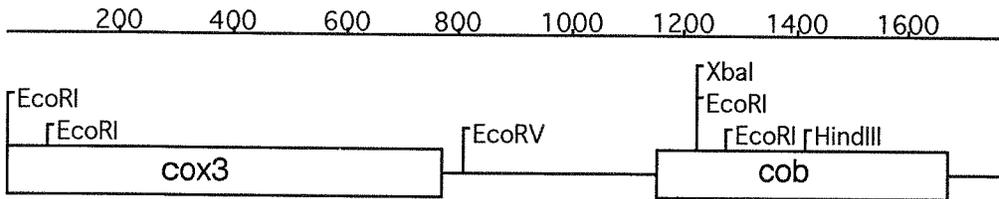


Figure 6
Chaput and Morse