

Université de Montréal

**Régulation transcriptionnelle du gène de la protéine
de liaison de la chlorophylle-a et de la péricidine chez
le dinoflagellé *Lingulodinium polyedrum***

Par

Mathieu Beauchemin

Département de Sciences Biologiques

Faculté des Arts et des Sciences

Mémoire présenté à la faculté des Arts et des Sciences en vue de l'obtention du grade
de Maîtrise en Sciences Biologiques

Octobre 2012

© Mathieu Beauchemin, 2012

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Régulation transcriptionnelle du gène de la protéine de liaison de la chlorophylle-a et
de la péridinine chez le dinoflagellé *Lingulodinium polyedrum*

Présenté par :
Mathieu Beauchemin

A été évalué par un jury composé des personnes suivantes :

Mario Cappadocia, président-rapporteur
David Morse, directeur de recherche
Normand Brisson, membre du jury

Résumé

Les dinoflagellés jouent un rôle très important dans l'écologie des océans en y réalisant une grande partie de la production primaire, en formant une association symbiotique avec les coraux et en ayant la capacité de produire des fleurs d'algues potentiellement toxiques pour les communautés côtières humaines et animales. Malgré tout, la biologie moléculaire des dinoflagellés n'a que très peu été étudiée dans les dernières années, les connaissances de processus de base comme la régulation de la transcription y étant fortement limitées. Une tentative pour élucider ce mécanisme a été réalisée chez les dinoflagellés photosynthétiques *Lingulodinium polyedrum* et *Amphidinium carterae*. Une expérience d'induction de la transcription du gène de la *Peridinin chlorophyll-a binding protein*, le complexe majeur de collecte de lumière, a été réalisée par une baisse de l'intensité lumineuse et a montré une faible augmentation (moins de 2 fois) du transcrit à court et long terme. Des expériences de simple-hybride et de retard sur gel (EMSA) ont été faits pour identifier de potentielles interactions protéine-ADN dans la région intergénique du gène PCP organisé en tandem. Ces essais ont été infructueux pour identifier de telles protéines. Une analyse du transcriptome de *L. polyedrum* a été effectuée, montrant une importante sous-représentation de domaines de liaison à l'ADN classique (comme *Heat-shock factor*, bZIP ou Myb) et une surreprésentation du domaine d'origine bactérienne *Cold shock* en comparaison avec d'autres eucaryotes unicellulaires. Ce travail suggère que les mécanismes de régulation transcriptionnelle des dinoflagellés pourraient différer substantiellement de ceux des autres eucaryotes.

Mots-clés : Dinoflagellés, *Lingulodinium polyedrum*, *Amphidinium carterae*,
régulation transcriptionnelle, *Peridinin chlorophyll-a binding protein*, transcriptome,
domaine de liaison à l'ADN

Abstract

Dinoflagellates are an important part of the ocean's ecology due to their large contribution to global carbon fixation, the symbiotic association they can make with corals and by their ability to form algal blooms potentially toxic for humans and animals in coastal communities. However, the molecular biology of dinoflagellates has been poorly studied in the past. Basic knowledge, such as regulation of gene expression, is severely limited. An attempt at deciphering basic gene regulation has been undertaken in the photosynthetic dinoflagellate *Lingulodinium polyedrum* and *Amphidinium carterae* using a reduction in available light intensity to induce the expression of the peridinin chlorophyll-a binding gene encoding the major light harvesting complex protein. A small increase in transcript abundance (less than 2 fold) was found in both short and long term experiments, yet neither yeast one-hybrid assays nor electrophoretic mobility shift assays (EMSA) showed any potential protein interactions with sequence derived from the intergenic spacer of the PCP tandem gene array. Interestingly, an analysis of the recently sequenced *L. polyedrum* transcriptome revealed an important under-representation of classic DNA-binding domains (such as Heat-shock factor, bZIP and Myb) and an over-representation of the bacterial cold-shock DNA-binding domain. This suggested that components of the transcription regulation machinery may be at least partially different in dinoflagellates.

Key words: Dinoflagellates, *Lingulodinium polyedrum*, *Amphidinium carterae*, transcriptional regulation, Peridinin Chlorophyll-a binding protein, transcriptome. DNA-binding domain

Table des matières

Résumé	I
Abstract.....	III
Table des matières	IV
Liste des figures.....	VI
Listes des tableaux	VII
Abréviations.....	VIII
1. Introduction.....	1
1.1 Photosynthèse chez les dinoflagellés.....	2
1.2 Régulation transcriptionnelle.....	6
1.3 Régulation traductionnelle.....	16
1.4 Régulation post-traductionnelle.....	18
1.5 Problématique et objectifs de l'étude	20
2. Matériels et méthodes	23
2.1 Culture de <i>Lingulodinium polyedrum</i> et d' <i>Amphidinium carterae</i>	23
2.2 Induction de PCP	23
2.2.1 Conditions de culture.....	23
2.2.2 Extraction et migration d'ARN	24
2.2.3 Analyse de l'expression de PCP par Northern blot	25
2.3 Simple hybride.....	26
2.3.1 Clonage des séquences promotrices de PCP et du gène rapporteur HIS3	26
2.3.2 Transformation des levures et criblage de la banque pADGal4	29
2.4 Essais de retard sur gel	31
2.5 Analyse des domaines de liaisons à l'ADN dans le transcriptome de <i>L. polyedrum</i>	33
3. Résultats.....	35
3.1 Induction de PCP	35
3.1.1 Induction à long terme chez <i>A. carterae</i>	35
3.1.2 Induction à long terme chez <i>L. polyedrum</i>	37

3.1.3	Induction à court terme chez <i>L. polyedrum</i>	38
3.2	Simple hybride	39
3.3	Essais de retard sur gel chez <i>L. polyedrum</i>	42
3.3.1	Sondes courtes	43
3.3.2	Sondes longues	44
3.4	Analyse des protéines de liaison à l'ADN	46
4.	Discussion	52
4.1	Induction de PCP en condition de faible intensité lumineuse	52
4.2	Essais d'interaction protéine-ADN	56
4.3	Protéines de liaison à l'ADN chez <i>L. polyedrum</i>	66
4.4	Perspective sur l'étude de la régulation transcriptionnelle chez les dinoflagellés	68
	Bibliographie	72

Liste des figures

Figure 1 : Organisation en tandem du gène PCP dans le génome de <i>L. polyedrum</i>	5
Figure 2 : Le niveau d'ARNm de PCP chez <i>A. carterae</i> varie faiblement à long terme lors d'une diminution de l'apport lumineux.	36
Figure 3 : Le niveau d'ARNm de PCP chez <i>L. polyedrum</i> augmente faiblement à long terme en condition d'intensité lumineuse basse.	37
Figure 4 : Induction de PCP chez <i>L. polyedrum</i> à court terme en condition d'intensité lumineuse basse.	39
Figure 5 : Résumé des différents fragments de l'unité PCP utilisés pour le simple hybride.	40
Figure 6 : Schéma des différentes sondes utilisées pour les essais de retards sur gel.	43
Figure 7 : Essai de retard sur gel (EMSA) avec 4 sondes courtes se situant près du site de début de la transcription de PCP.	44
Figure 8 : Essai de retard sur gel (EMSA) pour la sonde PCP497.	46
Figure 9 : Abondance relative de certains domaines de liaison à l'ADN dans le transcriptome de <i>L. polyedrum</i>	50

Listes des tableaux

Tableau 1 : Amorces utilisées pour amplifier les séquences de PCP et HIS3 par PCR	27
Tableau 2 : Amorces utilisées pour amplifier par PCR les sondes pour les essais de retard sur gel	32
Tableau 3 : Nombre de séquences comprenant certains domaines de liaison à l'ADN présent dans le transcriptome de <i>L. polyedrum</i> et dans les génomes de <i>Paramecium tetraurelia</i> , <i>Thalassiosira pseudonana</i> et <i>Chlamydomonas</i> <i>reinhardtii</i>	48

Abréviations

3-AT : 3 amino-triazole

ADN : Acide désoxyribonuléique

ADNc : Acide désoxyribonucléique complémentaire

ARN : Acide ribonucléique

ARNm : Acide ribonucléique messenger

ATP : Adénosine tri-phosphate

BSA : Albumine de sérum bovin

Dinap1 : *Dinoflagellate nuclear associated protein 1*

DTT : Dithiothréitol

EDTA : Acide éthylène diamine tétraacétique

FCP : *Fucoxanthin chlorophyll a/c binding protein*

HEPES : Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique

HIS : Histidine

HIS3: Imidazoleglycérol-phosphate déhydratase

LB : Luria-Bertoni

LBP : *Luciferin binding protein*

LCF : *Luciferase*

LD : *Light/Dark*

LEU: Leucine

MOPS : Acide 3-(N-morpholino) propane-sulfonique

PCP : *Peridinin Chlorophyll-a binding protein*

PCR: Réaction de polymérisation en chaîne

PEG : Glycol polyéthylénique

PFAM : *Protein families*

PMSF : Fluorure de phénylméthanesulfonyle

RubisCO : Ribulose 1,5 bisphosphate carboxylase/oxygénase

SC : *Synthetic complete*

SDS : Sodium dodécyl sulfate

Tris : 2-amino-2-hydroxyméthyl-1,3-propanediol

TBP : *TATA-box binding-protein*

TLF : *TATA-box like factor*

TRP : *TATA-box related proteins*

TRP: Tryptophane

URA : Uracile

1. Introduction

Les dinoflagellés constituent un groupe d'eucaryotes unicellulaires plutôt particulier faisant partie de la division des alvéolés, et sont donc de proches parents des ciliés et des apicomplexes. Ils se retrouvent généralement en milieu marin et comprennent tout aussi bien des hétérotrophes que des autotrophes photosynthétiques. Ils sont surtout connus pour leur association symbiotique avec les coraux (Gordon et Leggat, 2010) ainsi que pour la formation de marées rouges (*blooms*) et la production subséquente de phycotoxines contaminant les milieux marins et rendant impropres à la consommation humaine et animale les poissons et crustacés (Glibert et al., 2005). Leur rôle fondamental dans la production primaire des océans est par contre moins bien connu. En effet, les dinoflagellés sont, avec les diatomées, responsables de plus de la moitié de la fixation de carbone dans les océans, qui correspond elle-même à environ la moitié de la fixation de carbone totale sur la Terre (Field et al., 1998). Une espèce de dinoflagellé, *Lingulodinium polyedrum*, sert également, depuis plus de 50 ans, de modèle en biologie circadienne pour ses rythmes de bioluminescence et de photosynthèse et sa particularité à utiliser la régulation traductionnelle en lieu et place de la régulation transcriptionnelle comme principal moyen de régulation des rythmes biologiques (Hastings, 2007). Malgré les énormes implications que peuvent avoir les dinoflagellés sur la vie humaine, très peu d'éléments sont connus sur les mécanismes fondamentaux de la biologie des dinoflagellés. En ce sens, les mécanismes moléculaire de régulation d'expression des gènes, que ce soit la régulation transcriptionnelle,

traductionnelle ou post-traductionnelle, ont été très peu étudiés. Également, peu d'études ont été réalisées sur la réponse des dinoflagellés à différentes conditions environnementales, notamment au niveau de la photosynthèse.

1.1 Photosynthèse chez les dinoflagellés

Les dinoflagellés ont plusieurs caractéristiques uniques quant à la photosynthèse qui s'expliquent en partie par l'acquisition par endosymbiose secondaire d'un chloroplaste provenant d'une algue rouge (Fast et al., 2001). Ce chloroplaste est bordé par trois membranes et possède un mécanisme d'import protéique nécessitant le passage des protéines chloroplastiques à travers le réticulum endoplasmique et le Golgi par le biais d'un peptide signal « bipartite », d'un concept semblable à celui des euglènes, dont la première partie permet le passage dans le réticulum endoplasmique (Nassoury et al., 2003). Le passage des deux membranes restantes semble se faire de façon classique vu que la deuxième partie du peptide signal des dinoflagellés est suffisante pour cibler des protéines au chloroplaste des plantes supérieures (Nassoury et al., 2003). Ce chloroplaste possède également le génome le plus fortement réduit parmi tous les eucaryotes photosynthétiques. Contrairement au génome chloroplastique des plantes et des algues vertes, qui code pour environ 120 protéines sur un génome circulaire de 120 à 200 kb, celui des dinoflagellés code pour moins d'une vingtaine de protéines et possède une organisation tout à fait différente (Howe et al., 2008). Ces gènes codent principalement pour des sous-unités des photosystèmes I et II et du complexe cytochrome b_6f (Koumandou et al., 2004). Ils sont encodés sur des molécules d'ADN circulaire ressemblant à des plasmides et appelées « *minicircle* » (Zhang et al., 1999) qui comptent généralement un à trois gènes, toujours orientés de la

même façon par rapport à une séquence assez bien conservée dans chaque mini-cercle et qui pourrait peut-être servir de promoteur spécifique au chloroplaste (Howe et al., 2008). Les transcrits provenant du chloroplaste sont également pourvus d'une queue poly-U au lieu de la queue poly-A retrouvée dans les transcrits encodés dans le noyau (Wang et Morse, 2006).

La machinerie enzymatique responsable de la photosynthèse chez les dinoflagellés se distingue de plusieurs façons de celle des autres organismes photosynthétiques. Les dinoflagellés possèdent un pigment unique responsable de leur coloration rouge, la péricidine, une xanthophylle qui est liée par le principal complexe collecteur de lumière, la protéine de liaison de la chlorophylle-a et de la péricidine (*peridinin chlorophyll-a binding protein* (PCP)). La stratégie de récolte de lumière par les caroténoïdes est assez fréquente chez les algues. Elle permet de déplacer et d'élargir le spectre d'absorption pour capter la lumière bleue-verte, plus abondante sous l'eau que la lumière rouge, et ainsi de maximiser la quantité d'énergie lumineuse captée et dirigée vers les centres réactionnels (Polivka et al., 2006). PCP lie huit molécules de péricidine qui captent l'énergie lumineuse et la transmettent à deux molécules de chlorophylle-a. Le ratio de quatre péricidines pour une chlorophylle-a est toujours conservé pour PCP chez les différentes espèces de dinoflagellés et est nécessaire à la stabilité de la protéine, qui va ensuite former un homo-trimère (Schulte et al., 2010). Elle rend par contre la modulation de la capacité photosynthétique par la variation de la quantité et du ratio de pigment dans l'antenne beaucoup plus limitée. Une étude effectuée chez trois espèces du genre *Symbiodinium* a d'ailleurs montré que les ratios

entre les différents pigments (chlorophylle a, chlorophylle c et péricidine) restaient presque identiques sous deux intensités lumineuses différentes (Iglesias-Prieto et Trench, 1994).

Chez *L. polyedrum*, PCP est encodé par environ 5000 copies du gène organisées en tandem. Chaque unité comprend une séquence codante d'environ 1,3 kb précédée par une séquence intergénique de 1 kb, tel que présenté à la figure 1 (Le et al., 1997). La séquence protéique ne montre aucune similarité avec les complexes collecteurs de lumière classiques (LHC) présents chez à peu près tous les autres eucaryotes photosynthétiques. Ces complexes sont par contre également présents chez *L. polyedrum* et sont phosphorylés de façon similaire aux plantes en présence de lumière (Liu et al., 2012). Des analyses réalisées par PCR et par *Southern blot* indiquent que plusieurs copies de l'unité PCP sont assemblées bout à bout et que leurs séquences semblent être très bien conservées (Le et al., 1997). On ignore toutefois si les 5000 copies se présentent en un seul bloc ou bien si elles sont fragmentées en plusieurs blocs comprenant un nombre variable de copies réparties sur plusieurs chromosomes. Une autre espèce de dinoflagellé, *Amphidinium carterae*, contient, en plus de PCP, un complexe FCP (*Fucoxanthin chlorophyll a/c binding protein*) phylogénétiquement beaucoup plus proche des complexes collecteurs de lumière classiques comme ceux des algues vertes et rouges (Grossman et al., 1995; Hiller et al., 1993). Cependant, très peu de détails sur ce complexe et son implication dans la collecte de lumière sont connus chez les dinoflagellés.



Figure 1 : Organisation en tandem du gène PCP dans le génome de *L. polyedrum*.

Les rectangles blancs représentent la séquence codant l'ARNm et les flèches le site de début de la transcription de chaque unité. Les barres noires représentent les séquences intergéniques.

Les dinoflagellés sont également les seuls eucaryotes à posséder une RubisCO de type II provenant d'une α -protéobactérie et encodée dans le noyau (Morse et al., 1995; Rowan et al., 1996). Malgré une fonction enzymatique identique, cette enzyme ne possède que 25 % de similarité avec la RubisCO de type I présente chez les algues vertes, les plantes supérieures et la vaste majorité des autres eukaryotes photosynthétiques (Tabita et al., 2008). Elle possède, comme la type I, la capacité de lier et fixer le CO_2 et l' O_2 , mais paradoxalement pour un organisme produisant de l' O_2 par le biais de la photosynthèse et vivant dans un milieu marin qui est pauvre en CO_2 , sa spécificité pour ce dernier est plus faible que celle de type I, réduisant ainsi le rendement photosynthétique. Ceci suggère la présence d'un mécanisme de concentration du carbone très efficace vu la faible concentration et la lente diffusion du CO_2 en milieu aquatique (Lapointe et al., 2008). Ce mécanisme serait également potentiellement impliqué dans le rythme circadien de la photosynthèse en fournissant du CO_2 qui agirait comme accepteur final du pouvoir réducteur produit par le photosystème II (Mackenzie et Morse, 2011). Comme chez beaucoup d'autres algues, la RubisCO peut être concentrée dans une région du chloroplaste spécialisée dans la

fixation du carbone, le pyrénocyste (Badger et al., 1998). Ce micro-compartiment comprend généralement des anhydrases carboniques, des enzymes capables de convertir le bicarbonate en dioxyde de carbone, afin de concentrer ce dernier à proximité de la RubisCO et ainsi favoriser les réactions de fixation du carbone (Giordano et al., 2005). Une variation circadienne dans la formation et l'enrichissement en RubisCO du pyrénocyste serait en grande partie responsable de la capacité de *L. polyedrum* à fixer efficacement le carbone pendant le jour (Nassoury et al., 2001).

1.2 Régulation transcriptionnelle

La régulation transcriptionnelle a longtemps été perçue comme la clé de voute par laquelle un organisme contrôle ses réponses à différents stimuli environnementaux et son développement. Bien que cette vision ait été de beaucoup tempérée au profit de la régulation traductionnelle et post-traductionnelle, la régulation de la transcription reste un élément majeur du contrôle de l'expression génique d'un organisme. Elle passe typiquement par la liaison d'un facteur de transcription à une séquence précise en amont d'un gène. Cette liaison induit des changements dans la structure de la chromatine et/ou permet le recrutement des facteurs généraux de transcription et du complexe de l'ARN polymérase qui peut alors commencer à transcrire le gène. Une quantité très importante de données sur les mécanismes de régulation transcriptionnelle sont connues chez les organismes modèles. Par contre, les connaissances que l'on en a chez d'autres organismes comme les dinoflagellés sont très limitées.

Les noyaux et l'organisation génomique des dinoflagellés ont des caractéristiques uniques qui font d'eux un groupe tout à fait à part, repoussant les règles généralement bien définies des eucaryotes. Par exemple, malgré un gigantesque génome (environ 60 fois la taille de celui de l'être humain) toujours condensé tout au long du cycle cellulaire pour *L. polyedrum* (Costas et Goyanes, 2005), des histones classiques n'ont pas été détectées. Les dinoflagellés possèdent plutôt des protéines *histone-like* d'origine bactérienne présentes en un faible ratio de un pour dix par rapport à l'ADN en comparaison avec un ratio de un pour un pour les histones chez les autres eucaryotes (Moreno Diaz de la Espina et al., 2005). Ces protéines semblent être surtout présentes dans les zones transcriptionnellement actives (les zones périchromosomales) et absentes des chromosomes condensés, suggérant un rôle dans la régulation transcriptionnelle, probablement par une variation dans leur degré d'acétylation (Chudnovsky et al., 2002). L'absence d'histones classiques a longtemps été acceptée chez les dinoflagellés, mais l'avènement du séquençage à haut débit a modifié cette vision. En effet, la présence de plusieurs variantes des histones a été récemment détectée dans des échantillons environnementaux (Lin et al., 2010) ainsi qu'un répertoire complet de protéines de modification des histones chez *L. polyedrum* (Roy et Morse, 2012). Par contre, l'expression des histones est très faible et la détection de celles-ci au niveau protéique demeure pour le moment impossible (Roy et Morse, 2012). Ces données suggèrent donc un rôle fortement limité dans la régulation transcriptionnelle pour les histones classiques chez les dinoflagellés. Une seule étude a permis le séquençage d'un fragment relativement important (230 000 paires de bases) du génome du dinoflagellé *Heterocapsa triquetra*, révélant une faible densité en gènes

et la présence de séquences non-répétées sans similarité avec des séquences connues (McEwan et al., 2008).

Quant à eux, plutôt que de passer par un empaquetage protéique, les chromosomes seraient alors stabilisés par des ARN et des cations métalliques afin de les maintenir dans un état de cristal liquide (Chow et al., 2010). Ils possèdent aussi des télomères classiques, composés de la même séquence que chez les plantes (Fojtova et al., 2010). Le génome, très riche en GC, contient aussi une grande quantité de nucléotides modifiés comme la substitution des thymines par des 5-hydroxyméthyluracils, des cytosines par des 5-méthyl-cytosines et des adénines par des N⁶-méthyladénines (Rae, 1973; Steele et Rae, 1980). Ces bases modifiées étant distribuées de façon non-aléatoire dans le génome (Steele et Rae, 1980) et pouvant avoir une capacité d'interaction différente avec les protéines suggèrent qu'elles pourraient jouer un rôle non négligeable dans la régulation de l'expression des gènes.

Le faible nombre de séquences promotrices ayant été séquencées (en grande partie limité par l'absence de séquence génomique) rend l'analyse des promoteurs des dinoflagellés plutôt difficile. Les séquences potentiellement promotrices pour seulement trois gènes chez *L. polyedrum* sont connues en totalité pour PCP (Le et al., 1997) et la luciférase (LCF, (Li et Hastings, 1998)), deux gènes connus pour leur organisation en tandem, ou en partie pour la *Luciferin binding protein* (LBP, (Kim et al., 2000)). Ces séquences présentent plusieurs similitudes. Tout d'abord, aucune des

trois ne contient une boîte TATA en amont du site de départ de la transcription. Les séquences précédant ce site sont plutôt riches en GC, ce qui peut laisser croire que des protéines différentes des autres eucaryotes permettent l'initiation de la transcription (Kim et al., 2000). Il est intéressant de noter qu'un groupe a réussi à isoler une nouvelle classe de protéine d'initiation de la transcription, un intermédiaire entre les *TATA box binding protein* (TBP) et les *TATA-box like factor* (TLF)/*TATA-box related proteins* (TRP) chez le dinoflagellé hétérotrophe *C. cohnii* (Guillebault et al., 2002). Cette protéine a l'architecture des TBP et un site de liaison à l'ADN très semblable au TLF/TRP. Des études *in vitro* ont démontré qu'elle est capable de lier très faiblement une boîte TATA classique mais lie très bien une boîte TTTT (Guillebault et al., 2002). Par contre, aucune des séquences potentiellement promotrices ne contient un motif TTTT en amont du site de début de la transcription. L'implication biologique de cette protéine dans l'initiation de la transcription est inconnue, même s'il est possible pour les TBP de jouer leur rôle de recrutement de la machinerie transcriptionnelle sans nécessairement passer par la liaison d'une boîte TATA.

Les séquences en amont des régions codantes chez *L. polyedrum* semblent également être dépourvues d'autres séquences régulatrices de bases présentes chez les eucaryotes comme l'initiateur (Inr), présent au site de départ de la transcription, et des motifs en aval de ce site (*Downstream promoter element* (DPE) et *Downstream core element* (DCE)). Ces séquences servent normalement à lier des facteurs généraux de transcription et des protéines de la machinerie transcriptionnelle de base (Juven-Gershon et al., 2008). L'absence de ces séquences dans les promoteurs présumés des

dinoflagellés suggère des modifications assez importantes de ces facteurs afin qu'ils reconnaissent des séquences différentes de celles des autres eucaryotes. Il y a également une réduction importante du nombre et de la diversité des facteurs généraux de la transcription chez les alvéolés, rendant la machinerie transcriptionnelle beaucoup plus simple que chez les eucaryotes supérieurs (Meissner et Soldati, 2005).

Par ailleurs, deux motifs conservés dans les régions intergéniques de PCP et LCF ont été observés. Une séquence de 13 nucléotides est parfaitement conservée entre ces deux gènes et a été proposée comme étant un potentiel promoteur des dinoflagellés (Li et Hastings, 1998). Par contre, cette séquence est à un endroit assez différent dans chacune des séquences intergéniques, ce qui pourrait laisser croire à des rôles différents pour ces deux gènes. Une autre séquence d'une douzaine de nucléotides est conservée à 75% entre les deux gènes et est située dans les deux cas à environ 150 paires de bases en aval du site du début de la transcription, mais aucune description de protéines pouvant lier ces séquences n'est présente dans la littérature. Dernièrement, une quinzaine d'autres séquences intergéniques de gènes organisés en tandem ont été obtenues chez *A. carterae* (Bachvaroff et Place, 2008) et aucune ne semble contenir de boîte TATA ni d'autres éléments classiques reliés à l'initiation de la transcription. Une analyse bioinformatique de ces séquences n'a pas permis de dégager des motifs consensus ni de retrouver les deux séquences possiblement conservées dans PCP et LCF. Cette étude dégage cependant un patron où les gènes ayant des transcrits abondants sont à peu près tous organisés en tandem et présents en plusieurs copies, ce qui suggère que le nombre de copies d'un gène pourrait être le déterminant majeur

dans le niveau d'expression des gènes chez les dinoflagellés. Cette hypothèse n'est par contre pas confirmée par l'analyse du transcriptome de *L. polyedrum*, où les niveaux d'expression pour les gènes organisés en tandem ne semblent pas fortement corrélés au nombre de copies présent dans le génome (Beauchemin et al., 2012).

Tout comme pour les possibles séquences régulatrices, les informations quant aux facteurs de transcription chez les dinoflagellés sont très limitées. En effet, seulement deux études traitent de l'identification de protéines pouvant moduler globalement la transcription. L'identification de celles-ci a été effectuée en prenant des extraits nucléaires à différents moments du cycle cellulaire chez *C. cohnii*. En premier lieu, la *Dinoflagellate nuclear associated protein 1* (Dinap1) est une protéine présente en G1, capable d'interagir avec l'ADN génomique et préférentiellement associée aux zones transcriptionnellement actives en bordure du noyau (Bhaud et al., 1999). Cette protéine possède deux domaines doigt de zinc C2H2 et un domaine d'interaction protéique WW, suggérant l'organisation classique d'un facteur de transcription capable de lier une séquence cible et de recruter d'autres protéines. Une caractérisation ultérieure de cette protéine a révélé son interaction avec Dip1 et DapC, deux protéines ayant une certaine homologie avec des facteurs d'épissage (Guillebault et al., 2001). De nouveaux essais d'interactions avec l'ADN n'ont par contre pas donné de résultats positifs malgré la présence de domaines doigt de zinc pour Dinap1 et Dip1. Les auteurs suggèrent que les bases modifiées, qui n'ont pas été reproduites dans les essais *in vitro*, sont peut-être essentielles pour permettre l'interaction protéine-ADN (Guillebault et al., 2001). Des essais de transcription *in vitro* ont par ailleurs montré que ces protéines

étaient capables de moduler la transcription lorsqu'ajoutées en excès à des extraits nucléiques. Les auteurs font aussi l'hypothèse que ces trois protéines feraient partie du complexe d'épissage et pourraient également interagir avec le complexe de l'ARN polymérase II (Guillebault et al., 2001).

Une découverte importante quant à la régulation transcriptionnelle chez les dinoflagellés a été effectuée lorsqu'une équipe a découvert que chaque transcrit provenant du noyau recevait, par un mécanisme d'épissage en trans, une séquence de 22 nucléotides en amont du 5' UTR (Zhang et al., 2007). Cette séquence, appelée *Spliced leader* (SL), est identique chez tous les dinoflagellés, ce qui suggère que le mécanisme a été acquis tôt dans l'évolution des dinoflagellés (Wisecaver et Hackett, 2011). De façon très intéressante, cette séquence se retrouve tout aussi bien sur les transcrits issus de gènes codés par une copie que ceux codés par de multiples copies organisés en tandem (Bachvaroff et Place, 2008). Elle ouvre donc la porte à la transcription de polycistrons comprenant de multiples copies des gènes organisés en tandem. En effet, chez les trypanosomes, plusieurs gènes différents sont transcrits l'un à la suite de l'autre en de longs ARN polycistroniques. Ceux-ci sont ensuite clivés et additionnés d'une séquence SL permettant de délimiter le début de chaque séquence codante (Hastings, 2005). Basé sur ce modèle, il a été suggéré que l'addition de la séquence SL en 5' permet de délimiter les différentes copies codantes des séquences intergéniques lorsque celles-ci sont organisées en tandem chez les dinoflagellés (Lukes et al., 2009; Zhang et al., 2007). Ceci implique donc que la transcription de ces gènes soit régulée par des séquences placées en amont d'un bloc de copies du même gène

capable d'activer la transcription de plusieurs dizaines, voire plusieurs centaines de copies à la fois. Par contre, les données transcriptomiques obtenues pour *L. polyedrum* tendent à rejeter ce modèle dû au fait de l'absence quasi-totale de séquences intergéniques de la luciférase et de PCP, où le niveau d'expression de ces séquences est réduit de 5000 à 36000 fois, respectivement, par rapport aux séquences codantes. Au contraire, les introns ayant été transcrits et épissés sont facilement détectables dans le transcriptome de *S. pombe* et *H. sapiens*, en représentant entre 1% et 10% du niveau de la séquence codante (Ameur et al., 2011; Wilhelm et al., 2008). La présence de transcrits polycistroniques chez les dinoflagellés reste donc bien hypothétique.

Quelques études ont été réalisées dans les dernières années sur les changements transcriptionnels chez les dinoflagellés, en particulier pour des conditions de stress comme des changements de salinité (Lowe et al., 2011), des changements dans la concentration de nutriments comme le phosphate et le nitrate (Erdner et Anderson, 2006; Morey et al., 2011; Moustafa et al., 2010), des changements de densité cellulaire lors de *bloom* (Johnson et al., 2012) et des changements tout au long du cycle circadien (Okamoto et Hastings, 2003). Ces études ont produit des résultats plus ou moins cohérents lorsque deux ou plusieurs variables ont été combinées. Par exemple, des conditions de carence en nitrate et en phosphate, qui ont subséquentement été réintroduits dans le milieu de culture, ont induit d'importants changements transcriptionnels chez *Karenia brevis* pour l'ajout de nitrate mais pas pour l'ajout de phosphate (Morey et al., 2011). En revanche, une carence en phosphate chez *Alexandrium tamarense* induit de plus gros changements transcriptionnels qu'une

carence en nitrate (Moustafa et al., 2010), représentant moins de 5% des séquences analysées. Une étude semblable chez *Alexandrium fundyense* montre qu'au moins 10% des transcrits voient leur abondance changer dans ces deux mêmes conditions. Finalement, les changements circadiens semblent très peu influencer le changement des niveaux de transcrits chez *Pyrocystis lunula*. Environ 3% des séquences analysées par des expériences de micro-puces à ADN varie en fonction du temps circadien mais aucune ne voit son niveau augmenter ou diminuer de plus de 3 fois (Okamoto et Hastings, 2003), ce qui est en accord avec l'hypothèse que les rythmes circadiens des dinoflagellés sont surtout régulés aux niveaux traductionnel et post-traductionnel. Le peu d'études réalisées jusqu'à maintenant chez les dinoflagellés ne permet donc pas de déterminer de modèle clair quant à l'étendue des modifications transcriptionnelles lors de changements importants dans les conditions de cultures. L'arrivée du séquençage à très haut débit va certainement permettre de tester plusieurs autres conditions chez une multitude d'espèces de dinoflagellés dans les prochaines années.

La régulation transcriptionnelle des différentes protéines impliquées dans la photosynthèse chez les plantes est généralement associée à une acclimatation à plus ou moins long terme à des conditions environnementales changeantes comme la sécheresse ou une variation importante de la quantité de lumière. Cette acclimatation passe en général par la modulation de la quantité de pigments et d'antennes ainsi que par un changement dans la stœchiométrie des photosystèmes I et II (Eberhard et al., 2008). Quelques détails sont connus chez les dinoflagellés quant à la régulation transcriptionnelle de PCP lors de l'acclimatation à différentes intensités lumineuses.

Deux équipes ont démontré que le niveau d'ARNm de PCP augmentait lorsque l'intensité lumineuse était réduite. Tout d'abord, chez *Glenodinium*, la quantité de PCP obtenue après un essai de traduction *in vitro* est deux fois plus importante lorsque des ARNm extraits d'algues sous basses intensités lumineuses sont utilisés que lorsque la traduction est effectuée avec des ARNm extraits d'une culture mise sous forte intensité lumineuse (Roman et al., 1988). Le niveau d'ARNm semble donc plus élevé en conditions de faible intensité lumineuse, mais une quantification précise de la quantité d'ARN était impossible à l'époque puisque la séquence nucléotidique de PCP était inconnue. Une mesure plus précise du niveau d'ARNm de PCP sous différentes intensités lumineuses a été effectuée chez *Amphidinium carterae*. En basse lumière, le niveau de transcrite de PCP augmente de façon spectaculaire. En effet, celui-ci augmente de 33 et de 86 fois sous des conditions d'illumination moyenne et basse, respectivement (ten Lohuis et Miller, 1998). La hausse de transcription est corrélée avec une réduction de la méthylation des cytosines, laissant croire que le mécanisme de transcription des dinoflagellés est sensible à la méthylation des îlots CpG. Par contre, l'inhibition de la méthylation ne permet pas à elle seule la transcription de PCP, suggérant que l'induction de la transcription nécessite d'autres éléments classiques comme la liaison d'un ou plusieurs facteurs de transcription (ten Lohuis et Miller, 1998). La régulation transcriptionnelle des autres protéines impliquées dans la photosynthèse n'a pas été étudiée chez les dinoflagellés.

1.3 Régulation traductionnelle

La régulation traductionnelle a pris de plus en plus d'importance au cours des dernières années comme une façon efficace pour une cellule ou un organisme de réguler finement l'abondance d'une protéine et ainsi contrôler plusieurs processus. Deux types de régulation sont possibles, le contrôle global et le contrôle spécifique. Le premier est la régulation d'une grande population d'ARNm par le biais de la liaison de facteurs d'initiation de la traduction permettant le recrutement des ribosomes. L'abondance et les modifications post-traductionnelles de ces protéines influent sur l'initiation globale de la traduction. Le deuxième type est généralement limité à un faible nombre d'ARNm et comporte des protéines de liaison à l'ARN spécifiques pour une séquence ou une conformation des régions non-transcrites de l'ARNm (5' et 3'UTR). Une autre possibilité quant au contrôle spécifique est la liaison de petits ARN (siRNA et miRNA) sur l'ARNm, inhibant ainsi la traduction ou favorisant la dégradation de certains ARNm.

Le dinoflagellé *L. polyedrum* est un exemple classique en biologie circadienne où la régulation des rythmes circadiens passe par la traduction des gènes à des temps précis plutôt que par l'accumulation et la dégradation des ARNm à des temps précis (Hastings, 2007). Par exemple, la régulation de LBP passe par sa traduction exclusive lors de la période de nuit alors que la quantité d'ARNm est pratiquement constante pour toute la durée de la journée (Morse et al., 1989). La production de LBP enclenche la formation de scintillons, les organelles renfermant la luciferin et la luciférase et permettant à l'organisme de produire de la bioluminescence en phase de nuit. Plusieurs

autres protéines chez *L. polyedrum* voient aussi leur synthèse se dérouler à des moments précis de la journée, sans toutefois que la quantité totale de ces protéines ne change beaucoup (Markovic et al., 1996). Peu de détails sont connus sur les mécanismes permettant de réguler de façon circadienne la synthèse protéique chez *L. polyedrum*. Une protéine pouvant lier le 3'-UTR de l'ARNm de LBP dans la journée mais pas pendant la nuit a été identifiée (Mittag et al., 1994) sans que ces résultats ne puissent faire l'objet d'une confirmation et d'une étude plus approfondie (Lapointe et Morse, 2008).

Le contrôle de la synthèse des protéines photosynthétiques chez les dinoflagellés a fait l'objet de peu d'études. Par exemple, chez *Glenodinium*, la hausse du niveau d'ARNm se traduit par une hausse du niveau protéique de PCP. En effet, le niveau protéique de PCP augmente environ deux fois lorsque l'intensité lumineuse est diminuée de 4 fois (Roman et al., 1988). Par contre, de récentes études effectuées au laboratoire quant à la quantité de PCP chez *L. polyedrum* ne démontrent pas clairement une hausse de la quantité de cette antenne en condition de faible intensité lumineuse (Zampini et Morse, non-publié). L'abondance protéique de d'autres composantes de l'appareil photosynthétique sous différentes conditions lumineuses n'a pas été étudiée chez les dinoflagellés. Il est donc impossible de savoir si une séquence ou une structure quelconque dans l'ARNm de ces gènes pourrait réguler la traduction d'un groupe de protéines lors des changements de conditions lumineuses. Il est également possible que cette régulation traductionnelle ne passe que par le biais de la régulation globale et une variation de la quantité d'ARNm, comme observé chez *Glenodinium* (Roman et al.,

1988). Un aspect intéressant de la traduction de PCP est qu'elle est soumise à un contrôle circadien. En effet, sa synthèse ne se fait que lors des périodes subjectives de jour malgré une quantité à peu près constante de l'ARNm (Le et al., 2001). Ceci suggère qu'une ou des protéines peuvent s'associer à l'ARNm de PCP de façon à activer ou réprimer la traduction de celui-ci sous le contrôle de l'horloge circadienne.

1.4 Régulation post-traductionnelle

La phosphorylation des protéines a surtout été étudiée en lien avec les rythmes circadiens chez les dinoflagellés. Quelques études ayant testé l'effet d'inhibiteurs de la phosphorylation ainsi que d'inhibiteur de phosphatase ont montré un grand effet sur les rythmes de *L. polyedrum*. Par exemple, l'addition de 6-diméthylaminopurine, un inhibiteur de phosphorylation, aux cultures permet d'allonger la période circadienne de façon dose-dépendante et de stopper toute oscillation à forte dose (Comolli et al., 1994). Une autre étude réalisée avec trois inhibiteurs de phosphatase des sérines/thréonines a également montré un allongement de la période circadienne si les cultures étaient traitées avec l'acide okadaïque tandis que la cantharidin et la calyculin A ont causé des délais dans la phase du rythme de bioluminescence (Comolli et al., 1996). Une protéine phosphatase de type un a également été identifiée et caractérisée chez *L. polyedrum*. Malgré les essais précédents montrant un lien entre l'horloge circadienne et l'activité phosphatase de la cellule, cette protéine spécifique n'a pas de variation jour/nuit significative dans son activité (Comolli et al., 2003).

Plus récemment, une étude à grande échelle du phosphoprotéome de *L. polyedrum* a permis d'identifier quelques protéines pour lesquelles leur niveau de phosphorylation variait entre le milieu du jour et le milieu de la nuit (Liu et al., 2012). Par exemple, la protéine RAD24 a été identifiée comme étant phosphorylée durant la nuit mais pas durant le jour par une coloration de gel 2-D spécifique pour les phosphoprotéines. Cette protéine est associée à la réparation de l'ADN, et parce que la réplication du génome se déroule au milieu de la nuit, il semble que la phosphorylation soit contrôlée en lien avec le cycle cellulaire, qui lui-même est contrôlé par l'horloge circadienne (Dagenais-Bellefeuille et al., 2008). Une comparaison de l'enrichissement de phosphopeptides entre LD6 et LD18 a également pu identifier trois protéines ayant un domaine de liaison à l'ARN, en lien avec la régulation traductionnelle, et une polyketide synthase, qui pourrait être impliquée dans la synthèse de toxines (Liu et al., 2012). La phosphorylation semble donc jouer un rôle assez important dans le contrôle de certains processus liés à l'horloge circadienne chez *L. polyedrum*. D'autres types de modifications post-traductionnelles importantes, tel que l'ubiquitination, la glycosylation et l'acétylation et leurs effets associés n'ont malheureusement pas encore été examinés chez les dinoflagellés.

La régulation post-traductionnel des protéines photosynthétiques a fait l'objet de quelques études chez les dinoflagellés. Par exemple, la localisation, mais pas l'abondance protéique, de l'enzyme RubisCO change durant la période circadienne chez *L. polyedrum*. Celle-ci est concentrée dans le pyrénôïde lors de la phase de jour, permettant ainsi une fixation efficace du carbone et est dispersée dans le chloroplaste

durant la nuit (Nassoury et al., 2001). Comme la localisation de la RubisCO au pyrénéoïde est corrélée avec l'import protéique au chloroplaste, il est donc fort possible qu'une modification de l'enzyme ou bien son interaction avec un autre facteur permette le changement de localisation nycthéral (Nassoury et al., 2005). Il est également important de noter que les dinoflagellés ne possèdent pas de RubisCO activase classique (Portis et al., 2008) et que l'activation ou l'inhibition de la RubisCO par différentes conditions n'a toujours pas été étudiée. Ce changement de localisation est accompagné d'un changement dans la morphologie des chloroplastes qui deviennent allongés lors de la phase de jour, permettant ainsi aux protéines de collecte de lumière comme PCP d'être près de la surface de la cellule et ainsi capter le plus de lumière que possible (Nassoury et al., 2001).

1.5 Problématique et objectifs de l'étude

L'objectif principal du projet de recherche est d'identifier et de caractériser une ou des protéines de liaison à l'ADN chez le dinoflagellé *L. polyedrum*. Le choix des gènes à étudier pour des études de régulation transcriptionnelle et l'identification de protéines de liaison à l'ADN est fortement limité chez *L. polyedrum* par le nombre de séquences promotrices potentielles connues. En effet, seulement trois gènes ont une partie de ces séquences connues. Les gènes codant pour la luciférase et pour PCP étant organisés en tandem, les séquences intergéniques séparant les diverses unités codantes sont connues (Le et al., 1997; Li et Hastings, 1998). Pour LBP, une petite partie d'environ 100 paires de bases située en amont du site de départ de la transcription est

connu (Kim et al., 2000). La régulation circadienne de ces gènes passe par des mécanismes de régulation traductionnelle et post-traductionnelle, ce qui rend l'étude de leur régulation transcriptionnelle peu intéressante. Par contre, une condition étudiée chez un autre dinoflagellé, *A. carterae*, ayant permis une forte induction du gène codant pour PCP lorsque la luminosité est fortement réduite, tout en étant accompagné de changements dans la méthylation de l'ADN près de ce gène (ten Lohuis et Miller, 1998), semblait très intéressante comme point de départ pour l'étude de la régulation transcriptionnelle de PCP chez *L. polyedrum*. De plus, comme PCP est l'un des gènes les plus fortement exprimé chez les dinoflagellés, l'étude de potentiels éléments régulateurs pourrait permettre de trouver un promoteur fort permettant la manipulation génétique de l'organisme. Comme il est présentement impossible d'effectuer la transformation des dinoflagellés avec des gènes exogènes, cette découverte pourrait favoriser la création de vecteurs d'expression et ouvrir plusieurs nouvelles avenues quant à l'utilisation de techniques classiques de biologie moléculaire chez les dinoflagellés.

Afin d'étudier le mécanisme de régulation transcriptionnelle, des études d'induction du gène de PCP ont tout d'abord été effectuées en réduisant l'apport lumineux aux cultures d'*A. carterae* et de *L. polyedrum* et en mesurant le niveau d'expression du gène par *Northern blot*. Ces expériences ont été suivies par deux types d'essais d'interaction protéine-ADN afin de déterminer quelle région de l'unité PCP pourrait contenir un ou des motifs relié à la régulation transcripitonnelle. Pour ce faire, des études de simple-hybride, réalisées en criblant une banque d'ADNc dans la levure

et des essais de retard sur gel (EMSA) ont été effectuées. Finalement, les données transcriptomiques de *L. polyedrum* ont été analysées quant à leur contenu en protéines de liaison à l'ADN.

2. Matériels et méthodes

2.1 Culture de *Lingulodinium polyedrum* et d'*Amphidinium carterae*

Les dinoflagellés *Lingulodinium polyedrum*, souche CCMP1936 (précédemment connu sous le nom de *Gonyaulax polyedra*) et *Amphidinium carterae* ont été obtenus du *National Center for Marine algae* (Boothbay harbor, Maine, États-Unis). Ils ont été cultivés dans un milieu f/2 (Guillard et Ryther, 1962), sans agitation et sous un régime lumineux de 12 heures de lumière (environ 100 μmol de photons $\text{m}^{-2} \text{s}^{-1}$ fourni par des néons « *Cool white* » de Sylvania) et 12h de noirceur à une température de $18^{\circ}\text{C} \pm 1^{\circ}\text{C}$. Les temps appelés LD0 et LD12 représentent respectivement le début de la phase de lumière et le début de la phase de noirceur.

2.2 Induction de PCP

2.2.1 Conditions de culture

Des cultures de *L. polyedrum* et d'*A. carterae* ont été soumises à des conditions de luminosité réduite pendant des intervalles de quelques heures à quelques jours. La réduction de l'intensité lumineuse a été réalisée en couvrant les cultures avec un (intensité moyenne, environ 40 μmol de photons par m^2 par seconde) ou deux (intensité basse environ 12 μmol de photons par m^2 par seconde) filtres Rosco # 398 (gris neutre, transmission de 40%). La mesure des intensités lumineuses a été effectuée avec un luminomètre à une distance égale par rapport à la source lumineuse à celle où

les cultures ont été placées. Pour les essais à court terme, les cultures ont été soumises aux changements lumineux à partir de LD1 pour une durée de douze heures. Les essais à long terme ont quant à eux été réalisés à partir de LD6 et ce, sur une période de quatre jours pour *L. polyedrum* et de 12 jours pour *A. carterae*. Deux cents ml de cellules ont été récoltées pour chaque temps d'échantillonnage, par filtration (sur papier-filtre Whatman 541) pour les cultures de *L. polyedrum* et par centrifugation pendant 10 minutes à 4000g pour les cultures d'*A. carterae*. Les cellules ainsi récoltées ont ensuite été congelées extemporanément dans l'azote liquide et placées à -80 °C jusqu'à l'extraction d'ARN.

2.2.2 Extraction et migration d'ARN

L'ARN total a été extrait à l'aide du Trizol (Invitrogen) selon les instructions du fabricant et quantifié par spectrophotométrie. Dix µg de chaque échantillon ont ensuite été dénaturés à 65°C pendant 10 minutes dans un tampon comprenant 10% de formaldéhyde et 50% de formamide. Les extraits ont ensuite été séparés par électrophorèse en gel d'agarose dénaturant (tampon MOPS (40 mM MOPS pH 7,0, 10 mM acétate de sodium, 1 mM EDTA) 5% formaldéhyde, 1% agarose). Après migration, les gels ont été lavés deux fois pendant 15 minutes avec du SSC 10X (300 mM citrate de sodium, 1 M NaCl) et les ARN ont été transférés sur membrane de nylon chargée par transfert par capillarité pendant une nuit. Les membranes ont ensuite été traitées aux U.V. pour y fixer les acides nucléiques et colorées au bleu de

méthylène pour s'assurer de la qualité du transfert et de la quantité relative d'ARN dans chaque puits.

2.2.3 Analyse de l'expression de PCP par Northern blot

Des sondes radioactives spécifiques à PCP ont été générées par le système d'hybridation aléatoire d'hexamères (Promega) en présence de 50 μ Ci de [α - 32 P] dATP radioactif (Perkin Elmer) et 5 unités du grand fragment de l'ADN polymérase (Klenow) en utilisant 30 ng d'une matrice d'ADN. Celles-ci ont été amplifiées par PCR à partir d'un plasmide contenant le gène de PCP pour *L. polyedrum* (amorces 5' PCP : 5'-ATGCGCCTGTCTGTCGC-3' et 3'PCP : 5'-GGGTTTTCCAACCTCGAGAC-3') ou de l'ADN génomique pour *A. carterae* (amorces *A. carterae* PCPF : 5'-ATGGTTCGCAGTGGTAAGAAGGC-3' et *A. carterae* PCPR : 5'-GTTGAGCGGCTTGCACGACATCCTTG-3'). Les sondes ainsi générées ont été purifiées par Biogel P60 (Biorad) pour enlever les nucléotides non-incorporés et dénaturées 10 minutes à 95°C. Ces sondes ont été hybridées avec les membranes (préalablement incubées au moins une heure avant l'ajout de la sonde avec la solution d'hybridation) pour une vingtaine d'heures dans 50 ml de solution d'hybridation (0,5 M NaHPO₄, 7% SDS, 1% BSA, 1 mM EDTA) à 65°C sous rotation. Les membranes ont ensuite été lavées deux fois vingt minutes avec 50 ml d'une solution de lavage stringente (40 mM NaHPO₄, 5% SDS, 1 mM EDTA) et deux fois vingt minutes avec 50 ml d'une solution de lavage douce (40 mM NaHPO₄, 1% SDS), toujours à 65°C. Les membranes ont finalement été exposées pendant 24 à 72 heures

sur un écran au phosphore et la visualisation a été effectuée avec le système Typhoon 9200 (GE Healthcare). La quantification de l'intensité du signal obtenu pour chaque échantillon a été effectuée avec le logiciel Imagequant (GE Healthcare).

2.3 Simple hybride

2.3.1 Clonage des séquences promotrices de PCP et du gène rapporteur HIS3

Deux ensembles de vecteurs comprenant des régions d'environ 250 paires de bases de PCP, provenant de la séquence en amont du site du début de la transcription, ont été créés pour réaliser une expérience de simple hybride. La première série (PCP 500 et PCP250) comporte deux séquences différentes de PCP fusionnées avec le gène HIS3 sans promoteur basal et clonés dans le vecteur pYES2 (Invitrogen) tandis que la deuxième série (PCP 1083-1313, PCP 883-1133, PCP 683-912 et PCP 483-733) comporte quatre séquences différentes de PCP fusionnées au gène HIS3 comprenant une partie du promoteur basal de ce gène et clonés dans le vecteur p414CYC1 (Mumberg et al., 1995). Les différentes séquences de PCP ont été amplifiées à partir de l'ADN génomique de *L. polyedrum* (avec les amorces indiquées au tableau 1 pour chacune des constructions) et les deux versions du gène HIS3 ont été amplifiées à partir du plasmide p413GPD (version sans promoteur basal avec les amorces p413GPDHIS3F et p413GPDHIS3R et version avec promoteur basal avec les amorces p413HIS3PromF et p413GPDHIS3R (Mumberg et al., 1995)). Les réactions PCR ont été effectuées avec l'enzyme Taq Phusion (New England Biolabs) dans un volume de

50 µl selon les conditions suivantes : 0,5 unité de Taq Phusion, 0,2 mM de dNTP, 1µM d'amorce sens et 1µM d'amorce antisens et 25 ng de matrice. Elles ont été effectuées avec un thermocycleur PTC 200 (MJ Research) selon les conditions suivantes : 1 min 30s de dénaturation initiale à 98°C, 30 à 35 cycles comprenant 15s de dénaturation à 98°C, 20 sec d'hybridation à 68°C et 1 min d'élongation à 72°C et un cycle d'élongation finale de 5 min à 72°C. Les produits PCR ont ensuite été analysés par électrophorèse sur gel d'agarose (TAE 1X (40 mM Tris, 20 mM acide acétique et 1 mM EDTA) et 1% d'agarose) et purifiés par extraction au phénol/chloroforme et précipitation à l'éthanol.

Tableau 1 : Amorces utilisées pour amplifier les séquences de PCP et HIS3 par PCR

Nom de la construction	Nom de l'amorce	Séquence (5' vers 3')
PCP500	PCP500-F	GTACTGACTAGTTGCATTCTAATATGCC ATACC
	PCP500-R	ATCCGCGAACACGGCAG
PCP250	PCP-250F	GTACTGACTAGTGTTCGCTCAATCCGTG TG
	PCP-250R	AACACCACCACGCCAGTG
PCP 1083- 1313	PCP 1083-1102	ATGGGCCGCTCTCGTACC
	PCP 1333-1313	GTTCCAGTCAACCTCCTTCG
PCP 883- 1133	PCP 883-902	TCGCGGCCGCCTAAGATTCC
	PCP250orienR	AACACCACCACGCCAGTG
PCP 683 - 912	PCP 683-702	AGCACTAATTTCCCCTGCG
	PCP 933-912	GATGTTGGTTACGCAAGACTC
PCP 483 - 733	PCP 483-504	GGGTTTGTGGTGTGAGCTTCG
	PCP 733-714	ATCGCAGCGATTGCCATCG
	p413GPDHIS3F	CGTTACGACGTCATGACAGAGCAGAAA GCCCT
	P413GPDHIS3 R	CGTTACGACGTCCTACATAAGAACACCT TTGGTGG
	P413His3PromF	GCGATTGGCATTATCACATAATG

Les différentes régions de PCP ont ensuite été liguées avec le gène HIS3, comprenant ou non une centaine de paires de base du promoteur basal en utilisant des concentrations équimolaires pour chacun des produits PCR. Les ligations ont été faites avec l'enzyme T4 DNA ligase (New England Biolabs) à température pièce pendant 4 heures. Les produits de ces ligations ont été ré-amplifiés par PCR en utilisant des amorces situées aux extrémités des séquences (amorces sens de PCP pour chacune des constructions, amorce antisens p413GPDHIS3R, voir tableau 1) avec l'enzyme Taq Phusion (New England Biolabs) selon les conditions énumérées ci-haut. Les produits PCR ont ensuite été purifiés par extraction de gel d'agarose après électrophorèse avec les colonnes de silice EZ-10 (Biobasic). Les produits PCR destinés à être clonés dans le vecteur pYES2 ont été digérés par enzymes de restriction. Typiquement, 10 µl (environ 250ng) du produit PCR purifié a été digéré par 2 unités de l'enzyme XbaI et 2 unités de l'enzyme SpeI (New England Biolabs) pendant 1 heure à 37 °C et re-purifiés par extraction au phénol/chloroforme et précipitation à l'éthanol.

Les vecteurs pYES2 et p414CYC1 ont été préparés de façon à éliminer leur promoteur constitutif respectif placé en amont du site de clonage multiple. Deux µg du plasmide pYES2, préalablement extrait par lyse alcaline et purifié sur colonne de silice EZ-10 (Biobasic) ont été digérés avec 2 unités d'enzymes SpeI et 2 unités d'enzyme XbaI (New England Biolabs) à 37 °C pendant 2 heures. La même procédure a été appliquée au plasmide p414CYC1 en le digérant par les enzymes SacI et SmaI (New England Biolabs). Les deux vecteurs ont été purifiés sur gel d'agarose et les extrémités du vecteur p414 ont été rendus franches par un traitement avec la T4 DNA polymérase

selon les instructions du fabricant (New England Biolabs). Les ligations entre plasmide et insert ont finalement été effectuées avec un ratio molaire de 3:1 en présence de 10 unités de T4 DNA ligase (New England Biolabs) à 16°C pendant 4 heures.

La transformation a été effectuée par électroporation dans des bactéries *E. coli* de souche X11-Blue (Stratagene) qui ont été mises en culture pendant 16h à 37°C sur un milieu LB (composition par litre : 10 g de tryptone, 5 g d'extrait de levures et 10 g de NaCl) contenant 100 µg/ml d'ampiciline et 5 µg/ml de tétracycline. Les colonies ainsi obtenues ont ensuite été testées par PCR sur colonie et par digestion des plasmides pour s'assurer de la présence de l'insert et de son orientation. Les clones positifs ont finalement été séquencés par le service de séquençage Sanger de l'Institut de recherche en Immunologie et Cancérologie (IRIC) de l'Université de Montréal pour s'assurer de l'exactitude des séquences clonées.

2.3.2 Transformation des levures et criblage de la banque pADGal4

La souche de levure *Sacharomyces cerevisiae* YPH499 (génotype : *MATa*, *ura3-52*, *lys2-801_amber*, *ade2-101_ochre*, *trp1-Δ63*, *his3-Δ200*, *leu2-Δ1*) a été utilisée pour réaliser le simple hybride en criblant une banque d'ADNc cloné dans le vecteur pADGal4 (Stratagene). Cette banque avait préalablement été préparée par un ancien étudiant du laboratoire avec des ARNm de *L. polyedrum* extraient une heure après un cycle de 24 heures en condition constante, purifiés sur colonne poly-A avant d'être

rétrotranscrits en ADNc et clonés dans le vecteur (Thierry Bertomeu, communication personnelle).

Les levures ont d'abord été transformées avec les plasmides PCP-His-pYes2 et PCP-His-p414. Une colonie de levure a tout d'abord été placée dans 3 ml de milieu YPAD liquide (composition par litre : 6g d'extrait de levures, 12 g de peptone, 12 g de glucose et 60 mg d'adénine hémi-sulfate) et mis sous agitation à 30°C jusqu'à ce que la culture atteigne une densité optique de 0,5 à 0,6. Les levures ont ensuite été centrifugées à 8000g pendant 2 min et re-suspendues au vortex en présence de 10 µl d'ADN de sperme de saumon (10 mg/L, préalablement bouilli) et 1 µg de plasmide. Cinq cents µL de solution de transformation (40% PEG 3350, 100 mM acétate de lithium, 10 mM Tris-HCl pH 7,5 et 0,4 mM EDTA) ont alors été ajoutées aux levures. Celles-ci sont re-suspendues par inversion, laissées 30 minutes à température pièce et placées dans un bain à 42°C pendant 30 min. Les levures sont finalement centrifugées 2 minutes à 8000g, re-suspendues dans 200 µL d'eau stérile et mis sur un milieu de sélection approprié (SC -URA pour les constructions dans pYES2, SC -TRP pour les constructions dans p414) pour quelques jours à 30°C. La présence du plasmide dans les levures a été vérifiée par PCR avec les amorces permettant l'amplification des différentes séquences de PCP.

La transformation des levures pour le criblage à grande échelle de la banque d'ADNc pADGal4 a été effectuée avec le protocole à haute efficacité décrit par Gietz

et Woods (2002). Brièvement, une colonie de levures a été mise en culture dans 5ml de YPAD 2X et placé à 30 °C sous agitation pour 24 à 48 heures, jusqu'à ce que la culture atteigne une densité optique de 1, soit environ 2×10^7 cellules par ml. La culture a alors été ajoutée à 50 ml de YPAD 2X et incubé au moins 5 heures à 30 °C pour laisser les levures se diviser deux fois. À ce moment, la culture a été centrifugée à 3000g, lavée une fois avec 25 ml d'eau et re-suspendue dans 1 ml d'eau. Ensuite, 3,6 ml de solution de transformation (33% PEG3350, 100 mM acétate de lithium, 0,28 mg/ml d'ADN de sperme de saumon et 10 µg de la banque pADGal4) ont alors été ajoutés aux levures et celles-ci ont été vortexées vigoureusement pendant une minute et incubées pendant 60 minutes à 42 °C. Finalement, les levures ont été centrifugées pour enlever la solution de transformation, re-suspendues dans 1 ml d'eau et mis à 30 °C pendant 5 à 7 jours sur différents milieux sélectifs (SC -URA -LEU (construction dans pYES2) ou Sc -TRP -LEU (construction dans p414) pour tester l'efficacité de transformation et SC- URA (ou-TRP) -LEU -HIS avec 2 mM de 3-AT pour sélectionner de possibles clones positifs).

2.4 Essais de retard sur gel

Les sondes radioactives utilisées pour réaliser les essais de retard sur gel ont été amplifiées par PCR avec les amorces présentées au tableau 2. Environ 200 ng des produit PCR ont alors été marqués en 5' avec du γ -dATP³² par l'enzyme T4 polynucleotide kinase (New England Biolabs) selon les instructions du fabricant. Les sondes ont été purifiées par précipitation à l'éthanol pour enlever les nucléotides non-

incorporés et conservées à -20°C jusqu'à utilisation. L'activité spécifique de 1 µL de sonde a été mesurée au compteur à scintillation liquide (Tricarb, Perkin Elmer).

Tableau 2 : Amorces utilisées pour amplifier par PCR les sondes pour les essais de retard sur gel

Sonde	Amorce	Séquence (5' vers 3')
C	PCP intergen 15	CCTGCGATTGAGGTTACTG
	PCP intergen 16	CGGATTGAGCGAACACGC
D	PCP intergen 17	GCCGATGAGGGATGTAGT
	PCP intergen 19	TCAGTATGGAATTTGATGTTG
E	PCP intergen 20	GATTCCAACCTCGAACCGAGTC
	PCP intergen 21	CTCGTGCCTGTTGCCCTGTG
F	PCP intergen 22	CGCAGAGGCATGAGATGATGG
	PCP intergen 23	AGCGCGTGGCCGCCAACACC
PCP1	PCP intergen 1	CAACCTCCCGACATGAAG
	PCP intergen 6	CTGCTATCGGTGATCGGG
PCP497	PCP intergen 5	TTCTGGCACCTTGTGATC
	PCP intergen 10	GCGCGATTGCAATCGCAG
PCP997	PCP intergen 9	CAGTGGTCTTGAAAAGCC
	PCP1560-1540	GGCACGGACGCGATGACACG
PCP1497	PCP1501-1522	GTCCAGGGCTGACTGGGACAGC
	PCP2061-2040	ACCATCTTCTTCGGCACGGAGG
PCP1997	PCP1997-2019	GCTGACTACTCGGCCGTCAACG
	3'PCP (2360-2343)	GGGTTTTCCAACCTCGAGAC

Des cultures de *L. polyedrum* ont été récoltées par filtration au début de la phase de jour (LD2) ou au début de la phase de nuit (LD14) et les protéines totales ont été extraites dans le tampon Mittag (40 mM NaCl, 10 mM Tris HCl pH 7,4, 0,1 mM EDTA, 2 mM DTT, 10% glycérol, 300 µM PMSF, 3 µg/ml leupeptin, 25 mM NaF, 80 mM β-glycerophosphate (Mittag et al., 1994)) ou le tampon HEPES (20 mM HEPES-NaOH pH 7,9, 100 mM KCl, 1 mM EDTA, 1 mM DTT, 20% glycérol, 300 µM PMSF, 3 µg/ml leupeptin, 25 mM NaF, 80 mM β-glycerophosphate) en lysant les cellules au

Bead Beater (Biospec) pendant 3 minutes à 4°C. Les extraits ont ensuite été centrifugés à 16000g à 4°C pendant 15 minutes pour éliminer les billes et les débris cellulaires. Un dosage de l'extrait protéique a alors été effectué avec le réactif de Bradford (Biorad; (Bradford, 1976)) et 50 µg de protéines ont été ajoutés à la sonde radioactive (2000 coups par minute) et incubés à température pièce pendant 15 minutes. Ces réactions ont alors été analysées par électrophorèse non-dénaturante sur gel de polyacrylamide (0,5X TBE (50 mM Tris, 50 mM acide borique, 1 mM EDTA, pH 8,0), 10% glycérol, 4% polyacrylamide pour les sondes longues, 6% pour les sondes courtes) pendant environ une heure à 70 volts dans une chambre à 4°C. Après la migration, les gels ont été séchés à 65°C pendant une heure sur papier filtre à l'aide d'un séchoir à gel (Biorad), exposés pendant 24 heures sur un écran de phosphore et l'imagerie a été effectuée avec le système Typhoon 9200 (GE healthcare).

2.5 Analyse des domaines de liaisons à l'ADN dans le transcriptome de *L. polyedrum*

Le transcriptome de *L. polyedrum* a été séquencé à partir d'échantillons d'ARNm récoltés à LD6, LD18, LL18 et LD18 sans nitrate à l'aide d'un séquenceur Genome Analyzer IIX au Centre d'innovation Génome Québec et Université McGill (Montréal) comme décrit dans (Beauchemin et al., 2012). Au total, 312 millions de séquences de 76 paires de bases ont été obtenues, ce qui représente plus de 24 milliards de paires de bases. L'assemblage a été effectué avec Velvet (Zerbino et Birney, 2008) et Oases (Schulz et al., 2012) et a permis de générer 74337 séquences uniques de plus de 300 paires de bases. À celles-ci, 318 séquences ont été ajoutées à partir d'un

assemblage fait dans le logiciel Geneious (Drummond et al., 2011), pour un total de 74655 séquences. Celles-ci ont par la suite été analysées à l'aide du logiciel Blast2GO (Conesa et al., 2005) afin de les associer avec les domaines PFAM (Finn et al., 2010). Une comparaison des domaines PFAM reliés à la liaison avec l'ADN a été effectuée avec le cilié *Paramecium tetraurelia* (Aury et al., 2006), la diatomée *Thalassiosira pseudonana* (Armbrust et al., 2004), et l'algue verte *Chlamydomonas reinhardtii* (Merchant et al., 2007) en utilisant les modèles de gènes provenant du génome de ces organismes.

3. Résultats

3.1 Induction de PCP

En guise de travail préliminaire à la réalisation d'essais de retard sur gel, des essais d'induction de la transcription de PCP ont été effectués dans le but de trouver une condition permettant d'augmenter substantiellement la transcription de PCP. Ces essais sont basés sur une diminution de la quantité de lumière fournie aux cultures, et ce pendant de courtes ou de longues périodes. Des travaux précédents, effectués sur une autre espèce de dinoflagellé, *A. carterae*, ont montré une forte induction de PCP si les cultures étaient soumises à un éclairage faible pour une longue période de temps (ten Lohuis et Miller, 1998). Les expériences réalisées ont tenté de reproduire ces résultats chez *A. carterae*, tout en les appliquant à *L. polyedrum*.

3.1.1 Induction à long terme chez *A. carterae*

Des expériences d'induction à long terme ont d'abord été réalisées avec des conditions d'éclairage réduit chez *A. carterae*. Les cellules ont été cultivées pendant douze jours sous trois régimes lumineux différents (intensité haute (100 μmol de photons par m^2 par seconde), moyenne (40 μmol de photons par m^2 par seconde) et basse (12 μmol de photons par m^2 par seconde)) et elles ont été récoltées à quatre moments différents pendant ce laps de temps. L'analyse Northern présentée à la figure 2, permet de voir que la quantité de l'ARNm de PCP diminue de façon assez importante après 12 jours dans tous les échantillons. À court terme, après 2 jours, la

quantité du transcrit de PCP augmente d'une fois et demi en condition de plus faible éclairage par rapport à la condition normale. À plus long terme, la condition de luminosité la plus faible a un effet inverse à celui escompté. En effet, il y a une baisse assez importante de la quantité de PCP 4 et 6 jours après le début de l'expérience où celui-ci devient environ 2 fois moins abondant qu'en condition d'éclairage normal. Les échantillons sous intensité lumineuse moyenne voient quant à eux leur niveau de transcrits de PCP rester très semblable en comparaison avec celui du contrôle, avec des différences se situant entre 0,8 et 1,2 fois le niveau du témoin.

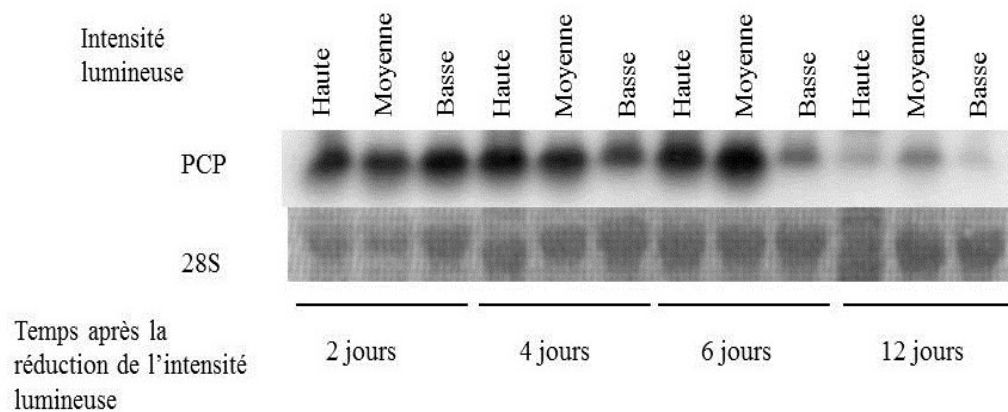


Figure 2 : Le niveau d'ARNm de PCP chez *A. carterae* varie faiblement à long terme lors d'une diminution de l'apport lumineux.

Les cultures d'*A. carterae* ont été placées pendant 2, 4, 6 et 12 jours sous différentes conditions lumineuses et le niveau d'ARNm de PCP a été mesuré par Northern blot. Le niveau de l'ARN ribosomal 28S, coloré au bleu de méthylène, sert de contrôle quant à la quantité d'ARN total mis sur gel. Intensité lumineuse haute : 100 μmol de photons par m^2 par seconde; intensité lumineuse moyenne : 40 μmol de photons par m^2 par seconde; intensité lumineuse faible : 12 μmol de photons par m^2 par seconde

3.1.2 Induction à long terme chez *L. polyedrum*

Une expérience de réduction de l'intensité lumineuse semblable à celle réalisée chez *A. carterae* a aussi été réalisée chez *L. polyedrum*. Le temps de culture sous faible intensité lumineuse a été limité à 4 jours vu que les cultures de *L. polyedrum* sont beaucoup moins robustes quant à leur survie à haute densité cellulaire sur une longue période de temps. Les résultats présentés à la figure 3 montrent une légère augmentation (environ 1,5 fois) de la quantité de PCP après 4 jours dans l'échantillon sous la plus faible intensité lumineuse. Quant à elle, l'intensité intermédiaire présente une quantité équivalente (1,1 fois) de transcrit de PCP par rapport au témoin placé sous forte intensité lumineuse.

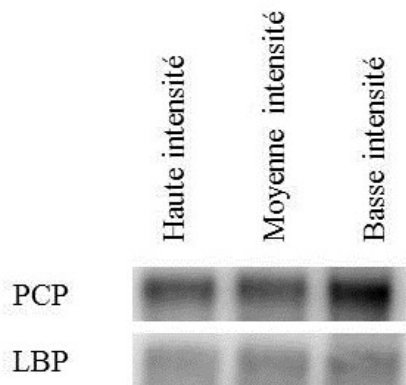


Figure 3 : Le niveau d'ARNm de PCP chez *L. polyedrum* augmente faiblement à long terme en condition d'intensité lumineuse basse.

Les cultures de *L. polyedrum* ont été placées pendant 4 jours sous différentes conditions lumineuses et le niveau d'ARNm de PCP a été mesuré par Northern blot. Le niveau d'ARNm de LBP sert de contrôle quant à la quantité d'ARN total mis sur gel. Intensité lumineuse haute : 100 μmol de photons par m^2 par seconde; intensité lumineuse moyenne : 40 μmol de photons par m^2 par seconde; intensité lumineuse faible : 12 μmol de photons par m^2 par seconde

3.1.3 Induction à court terme chez *L. polyedrum*

Un autre cadre temporel a été testé pour induire PCP, cette fois-ci à plus court terme. Les cultures ont vu leur apport lumineux être réduit d'environ 8 fois (condition de basse lumière lors de l'expérience à long terme) au temps LD1 et ce pendant 12 heures. Un échantillonnage assez serré dans les premières heures du changement de condition a été effectué et s'est poursuivi pendant 12 heures. La figure 4 présente les résultats de l'expérience de Northern effectuée à partir des échantillons d'ARN extrait des cultures récoltées à différents temps. Il semble y avoir une réponse en deux temps dans la quantité de transcrits de PCP. En effet, il y a une légère hausse dans les 2 premières heures après la mise en place des filtres, moment où le niveau du transcrit a augmenté d'environ 1,5 fois par rapport au temps 0. Une petite baisse dans le niveau de PCP a alors lieu après 3 et 4 heures d'induction. Finalement, il y a une hausse de près de deux fois du transcrit par rapport au contrôle après 5 heures d'induction. Le niveau du transcrits reste alors stable pour les points à 6 et 12 heures après l'induction.

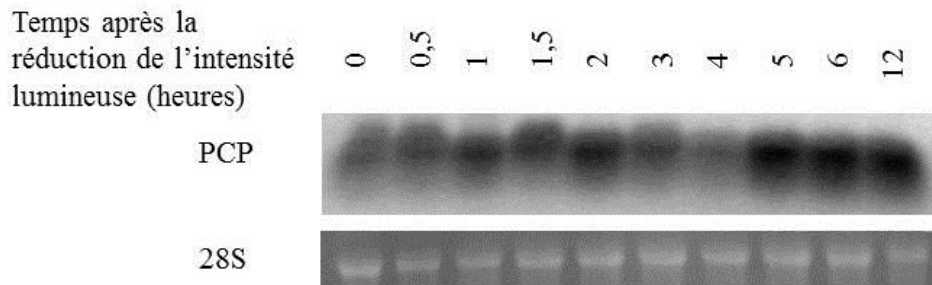


Figure 4 : Induction de PCP chez *L. polyedrum* à court terme en condition d'intensité lumineuse basse.

Les cultures de *L. polyedrum* ont été placées sous faible intensité lumineuse à LD1 et des échantillons ont été récoltés sur plusieurs heures après la réduction de l'intensité lumineuse à LD1. La quantité d'ARNm de PCP a été analysée par Northern blot. Le niveau de l'ARN ribosomal 28S, coloré au bromure d'éthidium, est montré pour indiquer le niveau d'ARN total présent dans chaque puits.

3.2 Simple hybride

La première technique utilisée pour tenter d'identifier des protéines de liaison à l'ADN pouvant potentiellement réguler la transcription de PCP a été le criblage d'une banque d'ADNc à l'aide de la technique de simple hybride. Cette technique permet d'identifier, par l'activation d'un gène rapporteur, des protéines pouvant se lier à une séquence d'ADN placée en avant de celui-ci. La première série de vecteurs comprend deux séquences de 250 paires de bases de PCP situées près du site de départ de la transcription de ce gène, tel que schématisé à la figure 5, partie A. Cette série contient un gène rapporteur HIS3 qui ne contient pas de séquence promotrice basale chez la levure. L'auto-activation du gène HIS3 a tout d'abord été testée après la transformation des levures avec les constructions PCP-HIS3-pYES2. Celles-ci ont été mises en culture sur un milieu SC -URA -HIS avec diverses concentrations de 3-AT, un inhibiteur

compétitif pour l'enzyme codé par le gène HIS3. Des concentrations de 0mM, 1mM, 2mM, 5mM et 10mM ont été testées de façon à trouver la plus faible concentration à laquelle la croissance des levures était stoppée sur milieu sélectif. Après 5 jours de croissance, les levures montraient une faible croissance sur les milieux comprenant 0mM et 1mM de 3-AT. En revanche, les milieux comprenant 2mM, 5mM et 10mM de 3-AT ne montraient aucune croissance des levures. Une concentration de 2mM de 3-AT a donc été choisie pour procéder au criblage de la banque pADGal4.

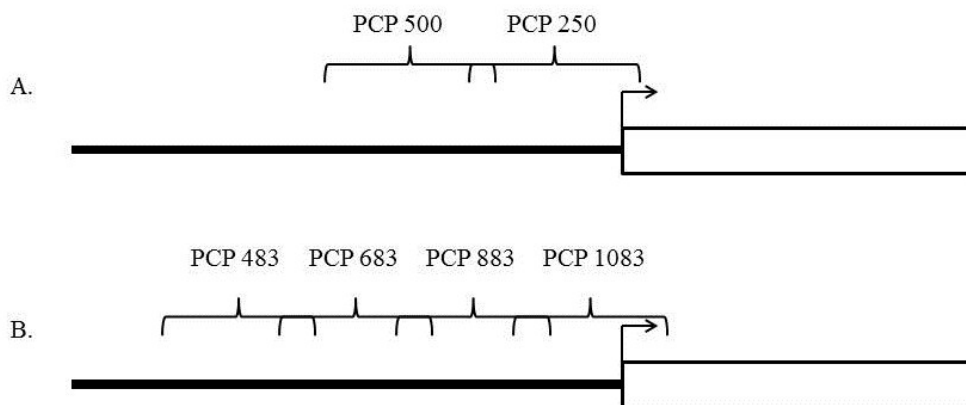


Figure 5 : Résumé des différents fragments de l'unité PCP utilisés pour le simple hybride.

A. Position relative des fragments PCP 500 et PCP 250 utilisés avec le vecteur pYes2 et le gène HIS3 sans promoteur basal. B. Position relative des fragments PCP 483, PCP 683, PCP 883 et PCP 1083 utilisés avec le vecteur p414 et le gène rapporteur HIS3 avec promoteur basal.

Deux séries de trente transformations avec la banque pADGal4 ont été effectuées pour chacune des deux constructions, permettant de cribler environ un million de clones pour chacune de celle-ci, ce qui représente plus de 75% de tous les

clones individuels de la banque d'ADNc. Les levures ont été placées en culture sur milieu SC –URA –LEU –HIS pendant une dizaine de jours et aucune colonies n'a émergé de ce crible, et ce pour les deux constructions. L'expérience a été répétée à nouveau et a généré les mêmes résultats négatifs.

Une autre série de vecteur a été construite, en incluant cette fois-ci une centaine de paires de bases du promoteur basal du gène HIS3. Celui-ci est généralement utilisé dans les essais simple-hybride afin de donner un cadre propice à la liaison des facteurs généraux de la transcription chez la levure. Pour ces constructions, quatre séquences de PCP, dont les positions sont schématisées à la partie B de la figure 5, ont été clonées en amont du gène HIS3 avec promoteur basal dans le vecteur p414. Ce nouveau vecteur a été utilisé vu les difficultés encourues lors du clonage de la première série de construction dans le plasmide pYES2. Un contrôle négatif ne comprenant que le gène HIS3 avec promoteur basal a également été généré afin de tester l'effet de celui-ci sur l'induction du gène rapporteur. Comme précédemment, les différentes constructions ont été testées pour l'auto-activation du gène rapporteur. Une fois transformé, les levures ont été mises en cultures sur milieu sélectif SC –TRP –HIS additionné de 0mM, 1mM, 2mM, 5mM, 10mM, 25mM et 50mM de 3-AT, l'inhibiteur du produit du gène HIS3. Le contrôle négatif, consistant au gène HIS3 sans séquence de PCP, n'a montré qu'une faible croissance sur milieu sélectif, croissance qui a pu être inhibée avec seulement 1mM de 3-AT. Par contre, les quatre autres constructions ont toutes démontrées une très forte auto-activation sur les milieux sélectifs additionnés de différentes concentrations de 3-AT. À la plus forte concentration testée, 50mM, les

levures montraient une très légère diminution dans leur vitesse de croissance. Cette diminution n'était malheureusement pas suffisante pour permettre d'identifier avec confiance des colonies de levures issues de l'auto-activation du gène rapporteur versus des colonies issues d'une activation du gène rapporteur par la liaison d'un facteur provenant de la banque pADGal4. Afin de s'assurer que les résultats de l'auto-activation élevée étaient bel et bien dus à la présence du plasmide et non pas un autre évènement, ceux-ci ont été transformés de nouveau dans la souche YPH499 et ont été soumis au même test d'auto-activation. Des résultats équivalents aux précédents ont été obtenus, ce qui a mené à l'abandon de cette série de vecteurs avant de procéder au criblage de la banque d'ADNc pADGal4.

3.3 Essais de retard sur gel chez *L. polyedrum*

La deuxième technique utilisée pour identifier des protéines de liaison à l'ADN a été les essais de retard sur gel (EMSA). Ces essais sont basés sur le principe qu'un fragment d'ADN, marqué à l'isotope P³², va migrer plus lentement dans une matrice de polyacrylamide s'il est lié par une ou des protéines que s'il est libre. La bande ainsi retardée indique une interaction protéine-ADN et permet subséquemment d'identifier la ou les protéines d'intérêt. Deux stratégies différentes ont été testées, une comportant de courtes sondes d'environ 150 paires de bases situées près du site de départ de la transcription de PCP et une autre avec de longues sondes d'environ 500 paires de bases permettant de couvrir l'entièreté de l'unité PCP.

3.3.1 Sondes courtes

Un schéma de la position des différentes sondes utilisées par rapport au locus PCP est présenté à la partie A de la figure 6. Ces sondes se retrouvent en amont ou près du site de départ de la transcription.

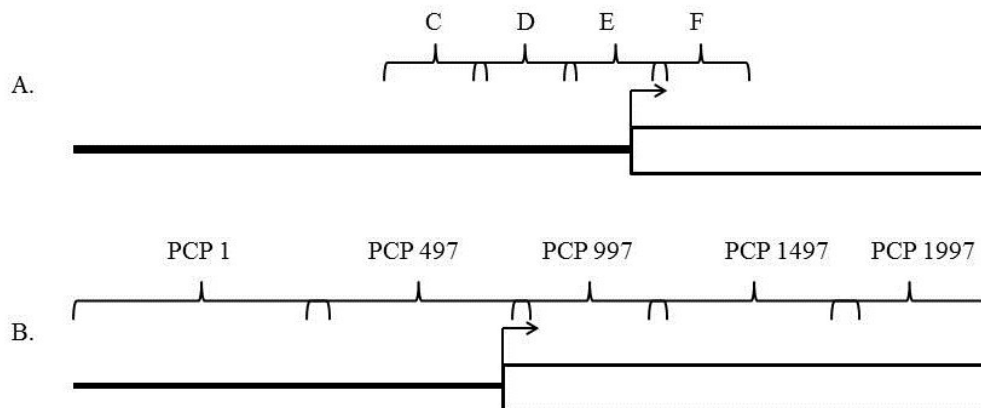


Figure 6 : Schéma des différentes sondes utilisées pour les essais de retards sur gel.

A. Sondes courtes, environ 150 paires de bases. B. Sondes longues (environ 500 paires de bases).

Ces sondes ont été utilisées pour deux séries d'expérience de retard sur gel, une avec des extraits protéiques totaux provenant de cultures récoltées au début de la phase de jour, à LD2, et une avec des extraits protéiques totaux provenant de cultures récoltées au début de la phase de nuit, à LD14. La figure 7 présente les résultats de ces essais. Pour les 2 conditions testées, et pour les 4 sondes utilisées, aucune différence entre les échantillons sans protéines et les échantillons avec protéines ne peut être observée. L'intensité de la sonde radioactive libre reste la même qu'il y ait ou pas de protéines dans l'échantillon, indiquant qu'il n'y a pas d'interaction protéine-ADN lors

de la réalisation de ces essais. Pour la sonde E, l'intensité réduite est due à un mauvais comptage du nombre de coups par minutes émis par la sonde radioactive.

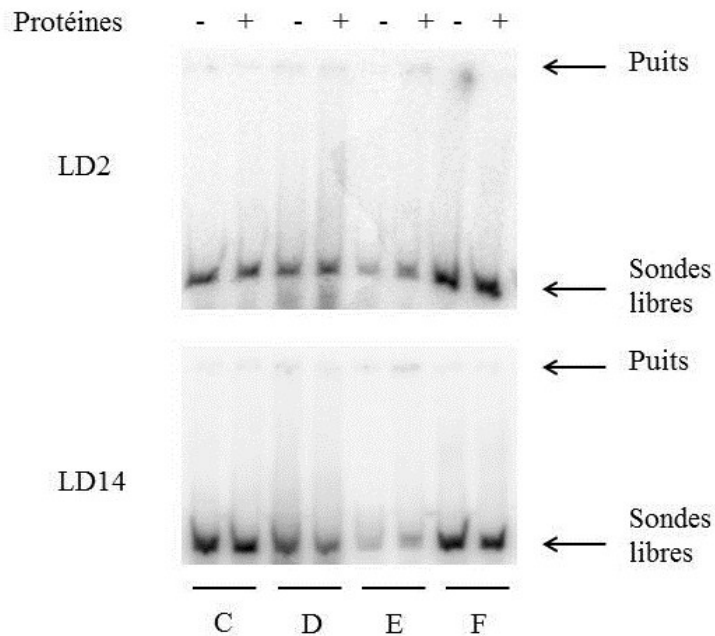


Figure 7 : Essai de retard sur gel (EMSA) avec 4 sondes courtes se situant près du site de début de la transcription de PCP.

Les essais ont été réalisés avec des cellules récoltées à LD2 (panneau du haut) et LD14 (panneau du bas). La présence d'un + indique que 50 µg de protéines ont été ajoutés à la sonde avant la migration.

3.3.2 Sondes longues

Un schéma de la position, en relation avec l'unité PCP, des 5 sondes longues utilisées pour les essais de retard sur gel est présenté à la figure 6, partie B. Comme pour les sondes courtes, des extraits protéiques de cellules récoltées à LD2 et à LD14 ont été utilisés pour la réalisation de l'expérience. De plus, des cultures récoltées une heure après la réduction de l'intensité lumineuse ont également été utilisées afin de tester un potentiel changement dans la liaison de protéines à l'ADN vu la faible

induction de PCP présentée à la figure 4. Cette figure présente les résultats pour la sonde PCP 497, qui comprend le site de départ de la transcription vers l'extrémité 3'. Les résultats obtenus sont très semblable pour les 3 échantillons protéiques utilisés. On peut tout d'abord constater que dans l'échantillon sans protéines, il y a une bande faible qui a migré moins rapidement que la majorité de la sonde. Les puits ayant reçu des protéines bouillies ne présentent pas de patron très différent du puits sans protéines, indiquant que le tampon utilisé pour extraire les protéines permet la migration normale de la sonde. Les 3 différentes quantités de protéines ne montrent pas de changements importants entre elles et présente un patron semblable au deux contrôles négatifs, ce qui indique que l'essai n'a pas permis de produire des interactions protéine-ADN. Les expériences effectuées pour les 4 autres sondes ont donné des résultats négatifs très similaires à la sonde PCP497 et ne sont pas présentés ici.

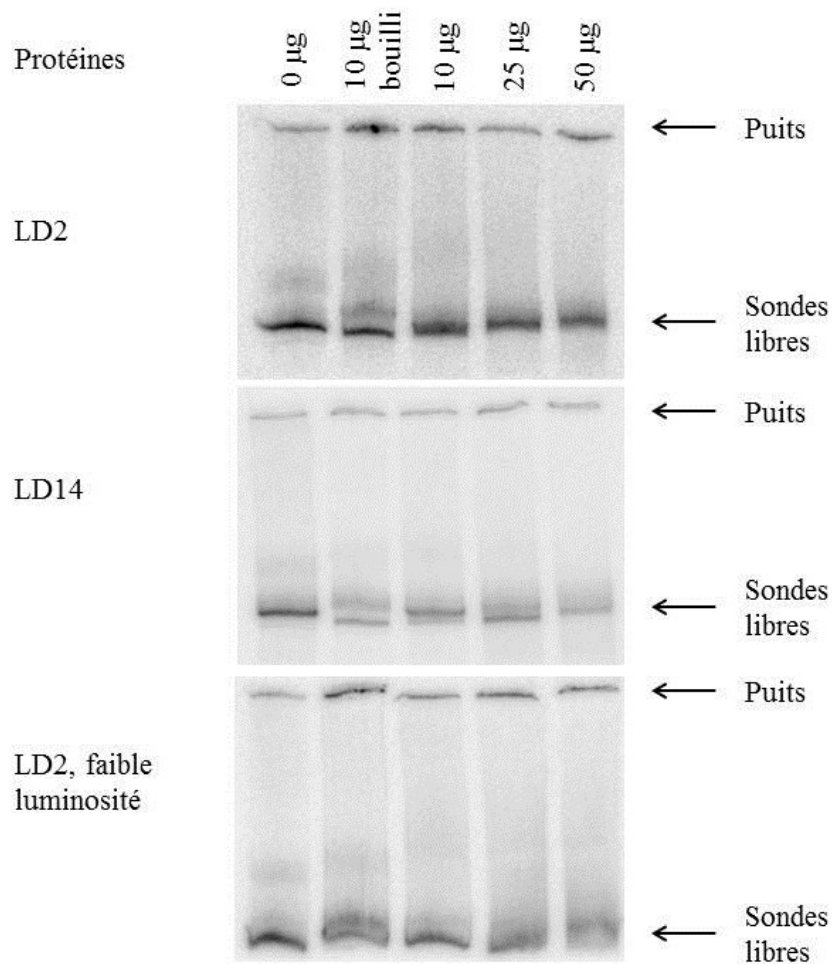


Figure 8 : Essai de retard sur gel (EMSA) pour la sonde PCP497.

L'essai a été effectué avec des protéines extraites à LD2 (panneau du haut), LD14 (panneau du milieu) et LD2 sous faible luminosité (panneau du bas). La quantité de protéines utilisée pour chaque puits est indiquée au-dessus du panneau du haut.

3.4 Analyse des protéines de liaison à l'ADN

Le séquençage à très haut débit du transcriptome de *L. polyedrum* (Beauchemin et al., 2012) a permis d'obtenir un aperçu des différentes familles de protéines de liaisons à l'ADN présentes chez cet organisme. Le tableau 3 présente le nombre de séquences identifiées comme possédant certains types de domaines de liaison à l'ADN.

Tout d'abord, on peut noter que le pourcentage total de ces séquences est assez semblable en comparaison avec le cilié *Paramecium tetraurelia*, avec environ 0,2% des séquences totales présentant ce type de domaine. Ce nombre est par contre plutôt faible en comparaison avec les diatomées et les algues vertes. Celles-ci possèdent en effet 1,8% et 0,8% de leur génome respectif qui code pour des protéines ayant un domaine de liaison à l'ADN. On peut également remarquer que la diversité de ces domaines est fortement réduite chez *L. polyedrum*. Plus de la moitié des domaines présentés au tableau 3 ne se retrouvent pas chez le dinoflagellé, une proportion semblable à celle du cilié *Paramecium tetraurelia*.

Tableau 3 : Nombre de séquences comprenant certains domaines de liaison à l'ADN présent dans le transcriptome de *L. polyedrum* et dans les génomes de *Paramecium tetraurelia*, *Thalassiosira pseudonana* et *Chlamydomonas reinhardtii*

Nombre de séquences (pourcentage du total)					
PFAM ID	Nom du domaine	<i>L. polyedrum</i>	<i>P. tetraurelia</i>	<i>T. pseudonana</i>	<i>C. reinhardtii</i>
PF00847	AP2	10 (0,013)	0 (0)	5 (0,044)	12 (0,079)
PF00170	bZIP	1 (0,001)	4 (0,010)	12 (0,105)	6 (0,039)
PF00313	Cold-shock	102 (0,137)	0 (0)	4 (0,035)	1 (0,007)
PF00125	Histones H2A/H2B/H3/H4	4 (0,005)	13 (0,033)	12 (0,105)	38 (0,249)
PF02701	Domaine à doigt de zinc Dof	0 (0)	0 (0)	0 (0)	1 (0,007)
PF02319	E2F/DP winged- helix	0 (0)	4 (0,010)	3 (0,026)	3 (0,019)
PF00320	Doigt de zinc GATA	0 (0)	0 (0)	0 (0)	8 (0,052)
PF00633	Hélice-épingle- hélice	0 (0)	0 (0)	2 (0,018)	1 (0,007)
PF00010	Hélice-boucle- hélice	0 (0)	0 (0)	3 (0,026)	4 (0,026)
PF01381	Hélice-tour- hélice	1 (0,001)	0 (0)	2 (0,018)	7 (0,046)
PF00505	Boîte HMG	13 (0,017)	3 (0,008)	19 (0,167)	7 (0,046)
PF05920	Homeobox KN domain	0 (0)	1 (0,003)	9 (0,079)	1 (0,007)
PF00447	Heat shock factor	0 (0)	20 (0,050)	90 (0,790)	2 (0,013)
PF00249	Myb-like	3 (0,004)	20 (0,050)	27 (0,237)	26 (0,170)
PF04542	Sigma-70	0 (0)	0 (0)	7 (0,061)	4 (0,026)
PF02135	Doigt de zinc TAZ	0 (0)	0 (0)	7 (0,061)	1 (0,007)
PF03638	Tesmin/TSO1- like CXC	0 (0)	6 (0,015)	6 (0,053)	2 (0,013)
PF03106	Domaine WRKY	0 (0)	0 (0)	0 (0)	1 (0,007)
PF00096	Doigt de zinc C2H2	1 (0,001)	1 (0,003)	1 (0,009)	3 (0,019)
PF00569	Doigt de zinc ZZ	14 (0,019)	2 (0,005)	4 (0,035)	4 (0,026)
	Total	149 (0,199)	74 (0,187)	213 (1,870)	132 (0,865)

La figure 9 présente le pourcentage relatif de certains domaines de liaison à l'ADN dans le transcriptome de *L. polyedrum* et en comparaison avec le cilié *Paramecium tetraurelia*, la diatomée *Thalassiosira pseudonana* et l'algue verte *Chlamydomonas reinhardtii*. Ce graphique met en évidence le fait que le domaine de liaison à l'ADN et à l'ARN de type Cold shock est fortement sur-représenté dans le transcriptome de *L. polyedrum*. Ce domaine est en effet quasi-absent chez les autres organismes utilisés aux fins de comparaison tout en représentant près de 70% de tous les domaines de liaisons à l'ADN présent chez *L. polyedrum*. Ce domaine d'origine bactérienne a tout d'abord été identifié comme pouvant lier l'ADN chez les bactéries mais son rôle chez les eucaryotes semble beaucoup plus large puisqu'il est également capable de lier l'ARN et de réguler la traduction (Mihailovich et al., 2010). Il est de façon assez surprenante totalement absent chez les ciliés, les plus proches parents des dinoflagellés parmi les espèces analysées. Il est donc fort possible que l'acquisition de ces gènes se soit fait après la divergence entre ciliés et dinoflagellés.

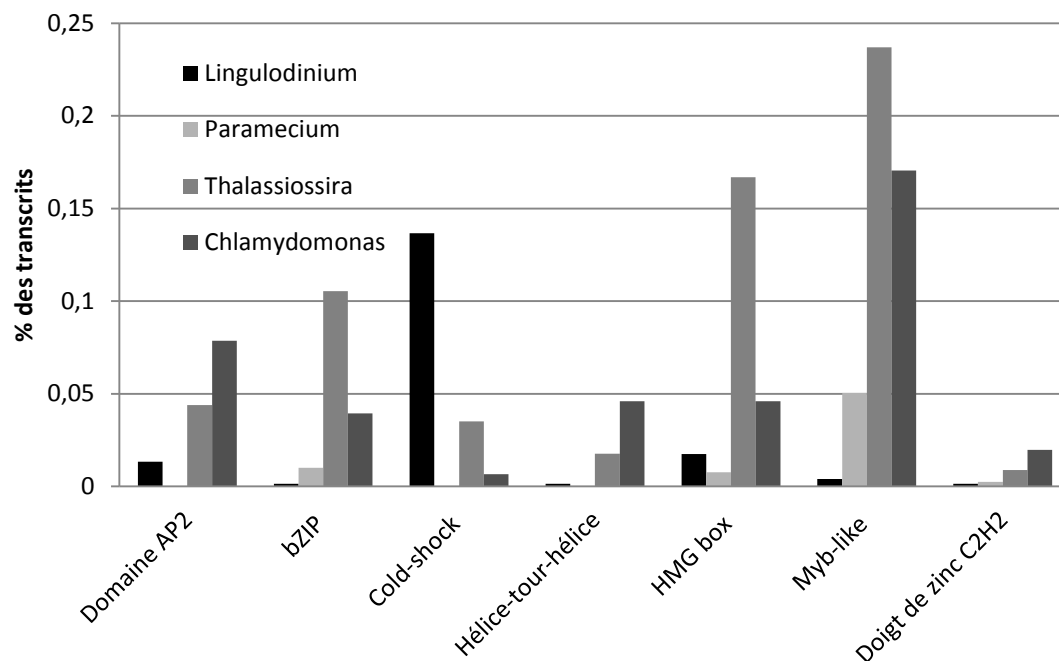


Figure 9 : Abondance relative de certains domaines de liaison à l'ADN dans le transcriptome de *L. polyedrum*

Les données sont présentées en pourcentage du nombre total de séquence codante dans le transcriptome de *L. polyedrum* et dans les génomes de *Paramecium tetraurelia*, *Thalassiossira pseudonana* et *Chlamydomonas reinhardtii*.

On peut également remarquer la présence de séquences codant pour des histones chez le dinoflagellé au tableau 3. Jusqu'à tout récemment, il était postulé que les dinoflagellés ne possédaient pas d'histones. De récents travaux ont réfuté cette conception en montrant que les dinoflagellés possèdent des séquences codants pour chacune de 4 histones de bases (Lin et al., 2010; Roy et Morse, 2012), sans toutefois que celles-ci semblent avoir une expression protéique suffisamment grande pour jouer un rôle important dans la régulation transcriptionnelle (Roy et Morse, 2012). En ce sens, il est normal de retrouver quelques séquences codantes pour des histones chez *L. polyedrum*, tout en remarquant qu'il y a un nombre beaucoup plus élevé de ces

séquences dans le génome des autres organismes analysés. Finalement, il est aussi plutôt surprenant de constater que le domaine *Heat shock factor* (tableau 3), qui se retrouve en grande quantité chez les diatomées (Rayko et al., 2010) et les ciliés, est totalement absent chez les dinoflagellés.

4. Discussion

4.1 Induction de PCP en condition de faible intensité lumineuse

Les expériences d'induction de PCP avaient pour but de permettre de trouver une condition à laquelle la transcription de ce gène était fortement activée. Une étude précédente avait montré que le dinoflagellé *A. carterae*, lorsque soumis à des conditions de faible et très faible luminosité, voyait la quantité d'ARNm de PCP augmenter de 33 à 86 fois, respectivement, sur une période de 14 jours (ten Lohuis et Miller, 1998). Ce changement dans la quantité de l'ARNm de PCP était également corrélé avec un changement dans la méthylation de l'ADN avoisinant le locus, suggérant que la hausse est au moins en partie due à une augmentation de la transcription et non seulement à un changement dans la stabilité de l'ARNm (ten Lohuis et Miller, 1998). Cette condition semblait donc un bon point de départ pour induire PCP chez les dinoflagellés et ainsi étudier la régulation transcriptionnelle de ce gène.

Ayant la chance d'avoir au laboratoire la même espèce que dans l'étude précédemment citée, j'ai tenté d'y reproduire les résultats en diminuant l'apport lumineux des cultures d'*A. carterae*. Trois intensités lumineuses ont été testées, la plus haute correspondant au niveau normal et la plus basse correspondant à l'intensité faible de l'étude précédente. Une intensité intermédiaire a également été ajoutée dans la réalisation de l'expérience. Les résultats de l'analyse Northern de l'expression de PCP

chez *A. carterae* présentés à la figure 2 sont très différents des résultats précédemment publié par ten Lohuis et Miller (1998). En effet, il n'y a qu'une faible augmentation du niveau du transcrit de PCP (environ 1,5 fois) dans la condition de plus faible lumière après 2 jours de réduction lumineuse et une diminution du transcrit après 4 et 6 jours. Après 12 jours, il y a une diminution du niveau du transcrit de PCP pour toutes les cultures. Il est possible qu'à ce moment, les cultures, approchant le point de saturation, réduisent leur métabolisme et, par le fait même, réduise la transcription de certains gènes. Ces données se comparent très mal à ceux précédemment publiées qui montraient une augmentation marquée du niveau du transcrit de PCP après 14 jours sous une intensité lumineuse comparable. Bien que cette mesure ait été prise après 14 jours au lieu des temps présentés ici (soit 2, 4, 6 et 12 jours), il semble très peu probable que les niveaux de transcrits se soient mis à augmenter dans les cultures privées de lumière vu la situation qui prévalait après 12 jours.

Malgré tout, des expériences semblables ont été réalisées chez *L. polyedrum*. Une expérience à long terme, sur 4 jours, présentée à la figure 3, a permis d'augmenter d'environ 1,5 fois le niveau de PCP en condition de faible luminosité par rapport à la culture témoin et à la culture ayant une luminosité intermédiaire. Cette augmentation est semblable à celle obtenue à long terme chez *A. carterae* et suggère que la réponse à long terme de l'organisme en condition de faible lumière ne passe pas par une augmentation de la transcription de PCP pour pouvoir au final produire plus de protéines de collecte de lumière et récolter le plus d'énergie lumineuse possible pour la photosynthèse. Cette explication peut être mise en relation avec de précédentes

expériences réalisées au laboratoire, où l'abondance protéique de PCP ne semblait pas changer de façon significative lors d'une diminution de l'intensité lumineuse (Zampini et Morse, non-publié).

Une expérience à court terme, présentée à la figure 4, a également été réalisée afin de déterminer si l'induction de PCP pouvait se faire dans les heures suivant l'application de la réduction du niveau de luminosité. L'abondance du transcrite de PCP a augmenté de façon similaire aux deux autres expériences présentées précédemment, avec des augmentations d'une fois et demi et de deux fois par rapport au contrôle, respectivement deux heures et cinq heures après la réduction de l'intensité lumineuse. Les variations obtenues pour cette expérience sont très semblables à ceux précédemment publiés pour *L. polyedrum*. En effet, Le et al. (2001) avaient montré, par analyse *Northern blot*, une variation dans la quantité d'ARNm de PCP qui se chiffrait à environ 2 fois tout au long de la période circadienne. Le séquençage à très haut débit du transcriptome de *L. polyedrum* a également montré des résultats semblables, où la différence maximale dans l'abondance du transcrite de PCP est de 1,6 fois entre un échantillon à LD0 et un à LD12 (Roy et al. non-publié).

Ces expériences d'induction visaient à trouver une condition dans laquelle la transcription de PCP était fortement augmentée afin d'en tirer profit pour identifier de potentiels régulateurs positifs de la transcription. Dans cette optique, les résultats obtenus sont plutôt décevants. Les changements dans l'abondance du transcrite de PCP

sont faibles pour les différentes conditions de luminosité et les différents lapses de temps testés et correspondent très bien avec les faibles variations circadiennes observées précédemment par *Northern blot* (Le et al., 2001) et par séquençage d'ARN à très haut débit (Roy et al. non-publié). Les faibles différences observées entre les différents échantillons sont donc très probablement non-significatives pour la biologie de l'organisme. Néanmoins, les faibles augmentations observées dans les différentes expériences ont été mises à profit lors des expériences de retards sur gel, où des échantillons provenant de cultures ayant vu leur intensité lumineuse réduite à court terme ont été utilisés à des fins de comparaison avec des cultures non-stressées.

En rétrospective, il apparaît plus ou moins probable que le niveau de PCP puisse augmenter de façon aussi spectaculaire que dans l'étude de ten Lohuis et Miller (1998). En effet, comme PCP est un des transcrits les plus abondants chez les dinoflagellés, représentant environ 1% de tous les ARNm (Le et al., 1997), il serait plutôt surprenant que son niveau puisse augmenter de 33 ou même de 86 fois, suggérant ainsi que plus d'un transcrit sur deux dans la cellule coderait pour ce complexe collecteur de lumière. Les résultats obtenus avec les expériences d'induction chez *A. carterae* et chez *L. polyedrum*, montrant une augmentation d'une fois et demie à deux fois du niveau de PCP tant à court terme qu'à long terme, semblent biologiquement beaucoup plus plausibles. Ces variations sont également près de ce que d'autres équipes ont obtenues, tant au niveau de la variation de la quantité d'ARNm de PCP tout au long de la journée chez *L. polyedrum* (Le et al., 2001) que de l'abondance relative de ce transcrit en condition de luminosité réduite chez *Glenodinium sp.*

(Roman et al., 1988). Il est également fort possible que l'acclimatation à des conditions de basses luminosités s'effectue par d'autres moyens qu'une augmentation de la quantité de protéines collectrices de lumière. Par exemple, chez l'algue verte *C. reinhardtii*, la réduction de la quantité de lumière n'induit pas une hausse de la quantité des différentes antennes mais un changement dans la stœchiométrie des photosystèmes I et II (Bonente et al., 2012). Ce mode d'acclimatation est en partie différent de celui des plantes supérieures (Eberhard et al., 2008) et il ne serait pas surprenant que les dinoflagellés adoptent une stratégie différente pour s'acclimater aux basses intensités lumineuses vu leurs particularités photosynthétiques. Cette acclimatation, et le degré possible de celle-ci, pourrait également varier d'espèce en espèce chez les dinoflagellés, comme suggéré par Igesias-Prieto et Trench (1994) lors d'une étude effectuée chez le genre *Symbiodinium*.

4.2 Essais d'interaction protéine-ADN

Afin de trouver des protéines capables de lier l'ADN, nous avons exploité la séquence intergénique connue du gène PCP qui semble être la séquence la plus logique pour réguler l'expression de chacune des copies organisées en tandem. Pour ce faire, des expériences de simple hybride et de retard sur gel ont été réalisées. La première technique, le simple hybride, permet d'isoler des clones d'une banque d'ADNc selon leur capacité à interagir avec une séquence précise de nucléotides placée en amont d'un gène rapporteur, dont la transcription est alors activée, permettant ainsi à la levure de compléter une déficience dans le métabolisme de certains acides aminés. Deux

stratégies ont successivement été essayées afin de trouver des protéines capables de lier des séquences issues de la séquence intergénique du gène PCP. La première tentative a été effectuée avec deux séquences de 250 paires de bases (vois figure 2, parti A) qui ont été placées en amont du gène rapporteur HIS3 amputé de son promoteur basal. Ces essais n'ont pas permis de d'identifier des clones permettant la croissance des levures sur milieu sélectif après criblage de plus d'un million de clones de la banque d'ADNc pADGal4 pour chacune des deux constructions. Comme cette banque comprend environ 1,3 millions de clones, elle n'a pas été complètement criblée lors de l'expérience de simple hybride. Il est donc possible que des séquences codant pour des protéines de liaisons à l'ADN aient été manquées par le criblage partiel de la banque d'ADNc.

Une autre possibilité pour l'échec de cette technique est le fait que les séquences de PCP placées en amont du gène rapporteur HIS3 ne contiennent pas les éléments nécessaires au recrutement des facteurs généraux de transcription permettant l'activation de la transcription chez la levure. Une deuxième tentative a donc été effectuée en utilisant une centaine de paires de base du promoteur basal du gène HIS3. Cette courte séquence contient les éléments nécessaires pour le recrutement du complexe de pré-initiation de la transcription et elle est utilisée dans certains vecteurs de simple hybride (Deplancke et al., 2004; Meijer et al., 1998). La présence de ce promoteur basal ne permet pas l'induction du gène rapporteur, comme en fait foi le contrôle négatif effectué avec celui-ci, qui ne requiert que 2 mM de 3-AT pour stopper toute croissance sur milieu sélectif. Malheureusement, les 4 séquences de PCP testées

ont provoqué une forte auto-activation du gène rapporteur menant à l'abandon de cette série de constructions avant le criblage de la banque d'ADNc. En effet, une concentration de 50mM de 3-AT n'a pas été suffisante pour stopper de façon importante la croissance des levures, ce qui aurait rendu le criblage de la banque presque impossible en générant une quantité très importante de faux positifs (Ouwerkerk et Meijer, 2001). Il semble donc que la présence des séquences de PCP semble suffisante pour recruter des protéines capables d'activer la transcription chez les levures. Une recherche pour des motifs de régulation transcriptionnelle présents dans les séquences clonées de PCP en relation avec des séquences présentes chez la levure a été effectuée avec YeastTract (Abdulrehman et al., 2011). Certains motifs de liaison pour des facteurs de transcription présents chez la levure ont été identifiés. Ceux-ci semblent permettre une réponse à certain stress environnementaux, comme Cup2p pour le cuivre et Stb5p pour le zinc ou bien induire des changements de développement chez la levure, comme Ash1p, qui est relié à l'expression des gènes de conjugaison. Bien que ces différentes conditions de stress ne se retrouvent pas dans l'expérience effectuée, il est possible que certaines séquences aient permis un contexte favorable à la liaison de protéines de liaison à l'ADN qui aurait pu à leur tour recruter les facteurs nécessaires à l'initiation de la transcription à l'aide du promoteur basal. Le fait que des séquences relativement longues (plus de 200 paires de bases) aient été utilisées pendant l'expérience a pu également favoriser ce phénomène.

Une des difficultés inhérente au criblage d'une banque d'ADNc se trouve dans la qualité des séquences composant celle-ci. Tout d'abord, le fait que la majorité des

séquences peuvent être dans le mauvais cadre de lecture (en théorie, environ 2 séquences sur 3 pour une banque clonée de façon directionnelle), réduit ainsi de beaucoup le nombre de clones fonctionnels exprimés dans la levure. De plus, la banque utilisée ayant été générée il y a quelques années, alors que les techniques de rétrotranscription étaient moins efficaces qu'aujourd'hui, elle contient probablement plusieurs séquences tronquées en 5', ce qui peut potentiellement éliminer des domaines fonctionnels important pour la fonction des protéines. Ce problème est évidemment plus fréquent et plus prononcé pour les longs ADNc. Finalement, les banques d'ADNc sont généralement enrichies en séquences fortement exprimées. Les facteurs de transcription étant en général assez faiblement exprimés, il devient encore plus difficile d'y trouver des clones complets clonés dans le bon cadre de lecture (Yang, 1998). Ces limitations sont donc à prendre en compte lorsque des résultats négatifs, comme ceux obtenues dans le premier crible, surviennent. Il est possible que la création d'une nouvelle banque d'ADNc, comprenant plus de clones et des clones plus complets pourraient permettre de trouver des interactions protéine-ADN à l'aide de la technique de simple-hybride, à condition de pouvoir régler les problèmes d'auto-activation.

Une autre technique classique d'identification de protéines de liaison à l'ADN a également été utilisée dans l'étude. Les essais de retard sur gel font partie des expériences les plus couramment utilisées pour identifier des protéines capables de lier des acides nucléiques. Elles sont basées sur le fait que les complexes protéines-ADN ont une masse plus grande que de l'ADN seule et prennent donc du retard par rapport à l'ADN libre de toute protéine lorsque ceux-ci subissent une électrophorèse (Hellman et

Fried, 2007; Laniel et al., 2001). Les complexes ainsi retardés peuvent ensuite être isolés et analysés par spectrométrie de masse afin de déterminer la ou les protéines ayant lié l'ADN. La technique permet également d'analyser finement des facteurs de transcription quant à leur spécificité pour la liaison de différents motifs d'ADN et la coopérativité entre différentes protéines permettant la liaison à l'ADN (Savery et Busby, 1998).

Tout comme les expériences de simple-hybride, les expériences de retard sur gel effectués dans cette étude n'ont pas permis d'identifier de protéine de liaison à l'ADN. Des sondes de deux tailles différentes, tel que présenté à la figure 6, ont été utilisées pour les essais. Les sondes plus courtes ont tout d'abord été utilisées parce qu'elles permettent de visualiser plus facilement un complexe protéine-ADN vu que la différence de taille entre les sondes libres et celles liées par une protéine est plus importante pour les sondes courtes que pour les sondes longues. Ces sondes ont été générées pour inclure le site de départ de la transcription de PCP ainsi que les séquences proches de ce site et qui seraient donc les plus à même d'être liées par des facteurs de transcription. L'utilisation d'extraits protéiques à deux temps circadiens différents (LD2 et LD14) n'a pas permis de visualiser de retard dans la migration des 4 sondes utilisées, comme montré à la figure 7. Des essais semblables ont été faits avec un groupe de 5 sondes plus longues. Ces sondes permettent de couvrir l'entièreté de l'unité PCP, tout aussi bien la séquence intergénique que la séquence codante. Tout comme pour les essais précédents, deux temps circadiens ont été testés, sans que cela ne permette d'obtenir des résultats positifs, et cela pour toutes les sondes testées. De

plus, une condition de faible intensité lumineuse a été utilisée. Comme celle-ci semblait induire faiblement la transcription de PCP à la figure 4, des extraits protéiques ont été récoltés à LD2, une heure après une baisse de l'intensité lumineuse. Les essais réalisés avec ces extraits n'ont pas permis d'obtenir de résultats différents par rapport aux essais faits à LD2 en condition de haute intensité lumineuse.

Plusieurs possibilités peuvent être évoquées pour expliquer les résultats négatifs obtenus lors des différentes expériences. La première possibilité est que les séquences utilisées pour réaliser ces essais ne contiennent pas de motifs de régulation de la transcription. L'hypothèse posée pour cette série d'expérience stipule que les séquences intergéniques des gènes organisés en tandem contiennent les éléments nécessaires à l'initiation et à la régulation de la transcription. Cette hypothèse est appuyée par le fait que les différentes copies de ces gènes sont toutes très semblables au niveau de l'ADN génomique (Le et al., 1997) et de l'ARNm (Beauchemin et al., 2012) et pourraient ainsi toutes être transcrites sans que de multiples mutations ne nuisent aux fonctions des protéines. Une corrélation entre le nombre de copies des gènes et l'abondance de leur ARNm laisse également croire que la majorité des copies sont transcrites (Bachvaroff et Place, 2008). Par contre, la découverte que les dinoflagellés emploient l'épissage en *trans* a permis de poser l'hypothèse que les gènes organisés en tandem pourraient être transcrits en un long ARN polycistronique dont les séquences intergéniques seraient éliminées lors du processus de maturation de l'ARN (Lukes et al., 2009; Zhang et al., 2007). Cette hypothèse impliquerait donc que l'initiation et la régulation de la transcription se fasse en amont de la première

séquence codante d'un groupe de gènes organisé en tandem et que les séquences intergéniques ne serviraient pas à réguler ces processus. L'utilisation du séquençage à très haut débit a par contre démontré que les séquences intergéniques de PCP et de la Luciférase ne sont pas transcrites chez *L. polyedrum*, suggérant que les ARN ne sont pas polycistroniques chez cette espèce (Beauchemin et al., 2012). Les séquences intergéniques pourraient donc être réellement impliquées dans la régulation transcriptionnelle. Les données obtenues dans ce travail ne peuvent malheureusement pas corroborer cette hypothèse vu qu'aucune protéine potentiellement capable de se lier à ces séquences n'a pu être identifiée. Il est important de noter que ces résultats négatifs ne peuvent pas prouver que les séquences intergéniques ne sont pas régulatrices de la transcription, et il reste possible que ces séquences puissent permettre de réguler la transcription d'une seule séquence codante de PCP. Cette hypothèse permettrait également d'expliquer le très haut niveau de conservation observé dans l'ARNm de la séquence codante et l'absence de transcription des séquences intergéniques. L'avènement du séquençage d'un génome de dinoflagellé dans les prochaines années devrait permettre d'avoir une bien meilleure idée de l'organisation des séquences promotrices des gènes organisés en tandem chez ces organismes.

Diverses autres explications sont possibles pour expliquer les résultats négatifs obtenus lors de l'étude. Tout d'abord, les essais de retard sur gel ont été faits en présumant que la transcription de PCP se faisait au début de la journée (LD2) ou de la nuit (LD14). Or, les données quant au moment où s'effectue la transcription de l'ARNm de PCP pendant le cycle circadien ne sont pas présentes dans la littérature. Il

est possible que le gène de PCP, malgré un niveau à peu près stable tout au long de la journée, ne soit transcrit qu'à certains moments de la journée. Une régulation de ce type se produit pour la traduction de PCP, où la synthèse se déroule presque exclusivement dans les premières heures de la journée (Le et al., 2001). Il est donc possible que la transcription de PCP se déroule à un moment qui n'a pas été testé dans les expériences présentées. Par exemple, elle pourrait être effectuée dans les heures qui précèdent la transition nuit-jour afin de générer une plus grande quantité de transcrits permettant ainsi la traduction dans les premières heures de la journée. À l'inverse, il est possible que la transcription de PCP soit à peu près égale toute la journée et que les petites différences observées dans la quantité du transcrit soient uniquement dues à un changement dans la vitesse de dégradation de l'ARNm. Des études s'intéressant au moment et à la vitesse où s'effectue la transcription des gènes n'ont jamais été publiées chez les dinoflagellés. Il serait également très intéressant de déterminer si la régulation transcriptionnelle chez *L. polyedrum* est contrôlée par l'horloge circadienne, comme le sont de multiples autres processus biologiques, et ce malgré une abondance constante de la très grande majorité des transcrits tout au long de la journée (Roy et al., non-publié).

Une difficulté technique qui n'a pu être éliminée lors de la réalisation des essais de retard sur gel est le fait que les extraits protéiques utilisés étaient des extraits totaux et non des extraits provenant de noyaux isolés. Comme les facteurs de transcription sont peu abondants et ne forment qu'une petite partie du protéome total d'une cellule (Yang, 1998), il est préférable d'isoler les noyaux des cellules avant d'extraire les

protéines pour réaliser l'expérience. Bien qu'il soit possible d'extraire les noyaux de dinoflagellés qui ne possèdent pas de thèque (Rizzo et al., 1982), il n'existe malheureusement pas de protocole d'extraction de noyaux pour les dinoflagellés ayant une thèque de cellulose. La présence de celle-ci rend très difficile le bris de la cellule par un moyen mécanique. La pression requise, générée par exemple par une presse de French, pour briser la thèque est assez grande pour faire éclater tous les organelles de la cellule, y compris le noyau. Il serait par contre envisageable de tenter de digérer la thèque de cellulose avec de la cellulase pour réduire la pression requise pour briser la thèque, comme précédemment réalisé lors de la purification de chloroplastes (Wang et al., 2005). Une cellulase provenant du dinoflagellé hétérotrophe *Cryptocodinium cohnii* a d'ailleurs été récemment caractérisée et pourrait être utilisée à cette fin chez *L. polyedrum* (Kwok et Wong, 2010). Une autre possibilité pour enrichir des facteurs de transcription aurait été d'utiliser des colonnes d'affinité pour les protéines de liaison à l'ADN, comme de l'héparine ou de la phospho-cellulose, afin d'enrichir les extraits en ces protéines (Gadgil et al., 2001). Ceux-ci auraient ensuite été utilisés pour réaliser les expériences de retard sur gel. La chromatographie d'affinité aurait également pu être réalisée de façon séquence-spécifique avec le locus de PCP pour tenter de purifier des facteurs de transcription se liant à ces séquences, permettant ainsi d'enrichir de plus de 500 fois les protéines d'intérêts (Kadonaga et Tjian, 1986).

Une autre difficulté technique est reliée au fait qu'il n'y a pas de séquences régulatrices et de facteurs de transcription associés décrits chez *L. polyedrum*, ce qui limite les contrôles positifs possiblement utilisables et rend plus difficile l'interprétation des résultats obtenus. Par exemple, si une séquence régulatrice était connue, elle aurait pu être utilisée lors de l'essai de simple hybride afin de tester si la banque permet d'identifier des protéines de liaison à l'ADN et si la séquence du gène rapporteur HIS3 sans promoteur basal est capable de permettre la transcription. Cette même séquence aurait aussi pu être utilisée lors des essais de retard sur gel et aurait permis de déterminer si les conditions d'extraction et de migration étaient suffisantes pour identifier une interaction protéine-ADN.

Finalement, les dinoflagellés sont connus pour la présence en grande proportion de plusieurs bases modifiées dans leur ADN génomique, comme le remplacement des thymines par des 5-hydroxy-méthyl-uraciles et des cytosines par des 5-méthyl-cytosines (Rae, 1973; Steele et Rae, 1980). Comme ces bases sont disposées de façon non-aléatoire, il est fort possible qu'elles puissent influencer la reconnaissance de l'ADN par différentes protéines, permettant ainsi une régulation différentielle des gènes. Chez le dinoflagellé *A. carterae*, il a été montré par *Southern blot* que le niveau de méthylation de l'ADN avoisinant le gène PCP était corrélé avec le niveau d'expression de celui-ci (ten Lohuis et Miller, 1998). Ces données suggèrent donc que la régulation de ce gène passe en partie par une modification du contexte de liaison des protéines au promoteur du gène. Comme les bases chez les levures lors de l'expérience de simple hybride et les sondes générées par PCR pour le retard sur gel ne

comprennent pas ces bases modifiées, il est possible que les protéines ne soient pas capables de reconnaître et de lier les séquences sans bases non-modifiées, tel que suggéré précédemment par Guillebault et al. (2001) lors de l'étude d'une protéine de liaison à l'ADN chez le dinoflagellé *C. conhi*. En ce cas, il serait bien évidemment très difficile de pouvoir générer des interactions protéines-ADN. De futures expériences pourraient être tentées avec les techniques de purification par affinité évoquées plus haut en se servant de l'ADN génomique fortement modifié provenant directement de *L. polyedrum* pour repêcher des protéines de liaison à l'ADN spécifiques à ces bases modifiées. L'obtention d'une grande quantité de séquences avec le séquençage complet du transcriptome permettrait également une identification plus facile de ces protéines par le biais d'analyse de spectrométrie de masse.

4.3 Protéines de liaison à l'ADN chez *L. polyedrum*

Le séquençage à très haut débit du transcriptome de *L. polyedrum* a permis d'avoir le premier portrait génétique d'un dinoflagellé. L'analyse de ce transcriptome a montré une sous-représentation en protéines de liaison à l'ADN en comparaison avec des eucaryotes unicellulaires (Beauchemin et al., 2012). En effet, le transcriptome contient environ 0,2% de protéines de liaison à l'ADN, un chiffre similaire à *P. tetraurelia* mais nettement plus faible que chez une diatomée et une algue verte (1,8% et 0,8%, respectivement) ainsi que d'autres espèces d'algues (Rayko et al., 2010). Une analyse des protéines possédant un domaine de liaison à l'ADN a montré une sous-représentation de la plupart des domaines présents chez les autres organismes, en

particulier pour les domaines *Heat shock factor* et *Myb-like*. Pour le premier, il est assez surprenant de voir une absence totale de ce domaine dans le transcriptome étant donné qu'il est présent en très grande quantité chez les ciliés et encore plus chez les diatomées (Rayko et al., 2010). Ce domaine, responsable de la réponse des organismes à des hautes températures (Sakurai et Enoki, 2010) est généralement présent en faible nombre chez les organismes unicellulaires (entre une et trois copies (Rayko et al., 2010)), suggérant que les dinoflagellés ont divergé des diatomées avant que le nombre de séquences possédant ce domaine explose chez ces derniers. Le domaine *Myb-like* est également présent en un faible nombre en comparaison avec tous les organismes analysés par Rayko et al. (2010). En effet, les domaines *Myb* sont, en particulier dans les organismes photosynthétiques, un des domaines de liaison à l'ADN les plus abondants (Riano-Pachon et al., 2008) et prennent part à la régulation transcriptionnelle d'une multitude de processus biologiques (Yanhui et al., 2006). Vu le nombre particulièrement élevée de séquences partout ailleurs, y compris chez de proches parents, il est surprenant que seulement 3 transcrits codent pour une protéine ayant ce domaine chez *L. polyedrum*. En général, on peut constater une diminution, et de la diversité et du nombre de séquences ayant un domaine de liaison à l'ADN chez *L. polyedrum* (Beauchemin et al., 2012) et un autre dinoflagellé, *Symbiodinium* (Bayer et al., 2012), en comparaison avec les autres organismes.

Un autre point important de l'analyse du transcriptome de *L. polyedrum* est la présence d'une grande quantité de séquences ayant un domaine de liaison à l'ADN de type *Cold shock*. En effet, il est présent dans 102 transcrits, formant presque 70% de

tous les transcrits possédant un domaine de liaison à l'ADN alors qu'il est presque absent des autres organismes analysés (tableau 3). Ce domaine d'origine bactérienne est, comme son nom l'indique, un domaine de réponse à une diminution de la température et a été bien étudié chez *E. coli* (Phadtare et Severinov, 2010). Bien qu'il ait à l'origine été décrit comme capable de lier l'ADN et de réguler la transcription, son rôle le mieux étudié est sa capacité à servir de chaperonne pour certains ARNm à basse température, permettant ainsi de protéger ces ARN des nucléases et de moduler la traduction. Le fait que ces domaines pourraient lier tout autant l'ADN et l'ARN est très intéressant pour les dinoflagellés étant donné l'hypothèse que la régulation de plusieurs processus semble préférentiellement passer par la régulation traductionnelle et post-traductionnelle. Pris ensemble, le faible nombre de protéines de liaison à l'ADN ainsi que la faible diversité des domaines correspondant renforcent l'idée que la régulation transcriptionnelle chez les dinoflagellés occupe une place moins importante que chez les autres eucaryotes, en particulier pour les rythmes circadiens (Hastings, 2007), malgré le fait que les génomes des dinoflagellés sont immenses et semblent contenir plusieurs gènes présents en multiples copies.

4.4 Perspective sur l'étude de la régulation transcriptionnelle chez les dinoflagellés

L'avènement du séquençage d'ARN à très haut débit va permettre d'étudier beaucoup plus facilement la régulation transcriptionnelle chez les dinoflagellés dans les prochaines années en contournant des problèmes comme l'absence de puces d'ADN

commerciales ainsi que le très faible nombre de séquences connues. Ces données permettront ainsi de caractériser la réponse transcriptionnelle en fonction de divers stress et changements environnementaux. À terme, la multiplication de ce type d'études pourrait permettre d'identifier des groupes de gènes régulés de façon semblable, ce qui faciliterait l'identification de séquences régulatrices et les facteurs de transcription qui y sont associées.

De plus, comme montré plus haut au tableau 3 et à la figure 9, les séquences issues des données transcriptomiques permettent d'identifier de possibles facteurs de transcription. L'étude de ceux-ci pourrait être alors grandement facilitée en ce sens que les protéines pourraient être étudiées à rebours, celles-ci pouvant être produites et analysées *in vitro*. La porte serait ainsi ouverte à d'autres types d'analyses qui n'ont jamais été utilisées chez les dinoflagellés, comme des essais d'immunoprécipitation de la chromatine (Collas, 2010). Il sera par contre plutôt difficile d'avoir une bonne idée des séquences régulatrices et des gènes qui leurs sont associées chez les dinoflagellés avant le séquençage complet d'au moins un génome. Ceci pourrait également permettre de déterminer si les séquences placées en amont des séquences codantes organisées en tandem peuvent potentiellement être impliquées dans la régulation de la transcription et si de potentiels séquences promotrices placées en amont des gènes possèdent de motifs connus de régulation génique.

Les difficultés encourues pour l'étude de la régulation transcriptionnelle chez les dinoflagellés pourraient se révéler très semblables à la situation qui prévaut chez leur plus proche parent, le groupe des apicomplexes dont fait partie l'agent de la malaria, *Plasmodium falciparum*. En effet, même si le séquençage complet du génome (très A/T riche, à l'opposé des dinoflagellés) de cet organisme a été complété il y a une dizaine d'années (Gardner et al., 2002), l'identification de motifs permettant la régulation de la transcription (Young et al., 2008) et de facteurs de transcription (Coulson et al., 2004) a été longue et ardue, malgré le fait qu'une part importante (plus de 40%) des transcrits (Le Roch et al., 2003) et des protéines (Florens et al., 2002) voient leur abondance changer selon les différents stades de vie du parasite. En particulier, le nombre de protéines associées à la transcription est environ trois fois plus faible que chez d'autres organismes de complexité semblable (Coulson et al., 2004), ce qui est une proportion semblable à ce qui est retrouvé chez *L. polyedrum* (Beauchemin et al., 2012). Également, bien que des gènes ayant le même patron de régulation se retrouvent sur la même partie d'un chromosome (Florens et al., 2002), les transcrits semblent être monocistroniques (Lanzer et al., 1992), tout comme chez les dinoflagellés. Par contre, la modulation de la transcription par la modification de la structure de la chromatine semble jouer un rôle assez important lors des changements de stades de développement (Miao et al., 2006), ce qui serait peu probable chez les dinoflagellés vu la très faible expression des histones (Roy et Morse, 2012). Finalement, l'identification de séquences régulatrices a été possible en corrélant les patrons d'expressions de groupes de gènes fonctionnellement semblables avec différents motifs présent en amont de ceux-ci (Young et al., 2008), suggérant une

architecture de régulation transcriptionnelle somme toute assez classique qui se retrouve peut-être chez les dinoflagellés.

En conclusion, les prochaines années, avec le séquençage prévu de quelques génomes et la multiplication d'études d'expression des gènes à grande échelle, devraient permettre de faire de grands pas dans la connaissance des mécanismes de régulation transcriptionnelle chez les dinoflagellés. Ces connaissances permettront d'avoir une meilleure idée de l'organisation des génomes gigantesques des dinoflagellés et des mécanismes permettant à la cellule d'y orchestrer une régulation précise des gènes.

Bibliographie

Abdulrehman, D., Monteiro, P.T., Teixeira, M.C., Mira, N.P., Lourenco, A.B., dos Santos, S.C., Cabrito, T.R., Francisco, A.P., Madeira, S.C., Aires, R.S., Oliveira, A.L., Sa-Correia, I., et Freitas, A.T. (2011). YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in *Saccharomyces cerevisiae* through a web services interface. *Nucleic acids research* *39*, D136-140.

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., et Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature structural & molecular biology* *18*, 1435-1440.

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., Bechner, M., Brzezinski, M.A., Chaal, B.K., Chiovitti, A., Davis, A.K., Demarest, M.S., Detter, J.C., Glavina, T., Goodstein, D., Hadi, M.Z., Hellsten, U., Hildebrand, M., Jenkins, B.D., Jurka, J., Kapitonov, V.V., Kroger, N., Lau, W.W., Lane, T.W., Larimer, F.W., Lippmeier, J.C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M.S., Palenik, B., Pazour, G.J., Richardson, P.M., Rynearson, T.A., Saito, M.A., Schwartz, D.C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F.P., et Rokhsar, D.S. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* *306*, 79-86.

Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J., et Wincker, P. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* *444*, 171-178.

Bachvaroff, T.R., et Place, A.R. (2008). From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS one* *3*, e2929.

Badger, M.R., Andrews, T.J., Whitney, S.M., Ludwig, M., Yellowlees, D.C., Leggat, W., et Price, G.D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO₂-concentrating mechanisms in algae. *Canadian Journal of Botany* *76*, 1052-1071.

Bayer, T., Aranda, M., Sunagawa, S., Yum, L.K., Desalvo, M.K., Lindquist, E., Coffroth, M.A., Voolstra, C.R., et Medina, M. (2012). Symbiodinium transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS one* *7*, e35269.

Beauchemin, M., Roy, S., Daoust, P., Dagenais-Bellefeuille, S., Bertomeu, T., Letourneau, L., Lang, B.F., et Morse, D. (2012). Dinoflagellate tandem array gene

transcripts are highly conserved and not polycistronic. *Proceedings of the National Academy of Sciences of the United States of America* *109*, 15793-15798.

Bhaud, Y., Geraud, M.L., Ausseil, J., Soyer-Gobillard, M.O., et Moreau, H. (1999). Cyclic expression of a nuclear protein in a dinoflagellate. *The Journal of eukaryotic microbiology* *46*, 259-267.

Bonente, G., Pippa, S., Castellano, S., Bassi, R., et Ballottari, M. (2012). Acclimation of *Chlamydomonas reinhardtii* to different growth irradiances. *The Journal of biological chemistry* *287*, 5833-5847.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* *72*, 248-254.

Chow, M.H., Yan, K.T., Bennett, M.J., et Wong, J.T. (2010). Birefringence and DNA condensation of liquid crystalline chromosomes. *Eukaryotic cell* *9*, 1577-1587.

Chudnovsky, Y., Li, J.F., Rizzo, P.J., Hastings, J.W., et Fagan, T.F. (2002). Cloning, Expression and characterization of a histone-like protein from the marine dinoflagellate *Lingulodinium polyedrum* (Dinophyceae). *Journal of Phycology* *38*, 543-550.

Collas, P. (2010). The current state of chromatin immunoprecipitation. *Molecular biotechnology* *45*, 87-100.

Comolli, J., Taylor, W., et Hastings, J.W. (1994). An inhibitor of protein phosphorylation stops the circadian oscillator and blocks light-induced phase shifting in *Gonyaulax polyedra*. *Journal of biological rhythms* *9*, 13-26.

Comolli, J., Taylor, W., Rehman, J., et Hastings, J.W. (1996). Inhibitors of serine/threonine phosphoprotein phosphatases alter circadian properties in *Gonyaulax polyedra*. *Plant physiology* *111*, 285-291.

Comolli, J.C., Fagan, T., et Hastings, J.W. (2003). A type-1 phosphoprotein phosphatase from a dinoflagellate as a possible component of the circadian mechanism. *Journal of biological rhythms* *18*, 367-376.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., et Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* *21*, 3674-3676.

Costas, E., et Goyanes, V. (2005). Architecture and evolution of dinoflagellate chromosomes: an enigmatic origin. *Cytogenetic and genome research* *109*, 268-275.

Coulson, R.M., Hall, N., et Ouzounis, C.A. (2004). Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome research* *14*, 1548-1554.

Dagenais-Bellefeuille, S., Bertomeu, T., et Morse, D. (2008). S-phase and M-phase timing are under independent circadian control in the dinoflagellate *Lingulodinium*. *Journal of biological rhythms* *23*, 400-408.

Deplancke, B., Dupuy, D., Vidal, M., et Walhout, A.J. (2004). A gateway-compatible yeast one-hybrid system. *Genome research* *14*, 2093-2101.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., et Wilson, A. (2011). Geneious v5.4.

Eberhard, S., Finazzi, G., et Wollman, F.A. (2008). The dynamics of photosynthesis. *Annu Rev Genet* *42*, 463-515.

- Erdner, D.L., et Anderson, D.M.** (2006). Global transcriptional profiling of the toxic dinoflagellate *Alexandrium fundyense* using Massively Parallel Signature Sequencing. *BMC genomics* 7, 88.
- Fast, N.M., Kissinger, J.C., Roos, D.S., et Keeling, P.J.** (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 18, 418-426.
- Field, C.B., Behrenfeld, M.J., Randerson, J.T., et Falkowski, P.** (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237-240.
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., et Bateman, A.** (2010). The Pfam protein families database. *Nucleic acids research* 38, D211-222.
- Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., Witney, A.A., Wolters, D., Wu, Y., Gardner, M.J., Holder, A.A., Sinden, R.E., Yates, J.R., et Carucci, D.J.** (2002). A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 419, 520-526.
- Fojtova, M., Wong, J.T., Dvorackova, M., Yan, K.T., Sykorova, E., et Fajkus, J.** (2010). Telomere maintenance in liquid crystalline chromosomes of dinoflagellates. *Chromosoma* 119, 485-493.
- Gadgil, H., Oak, S.A., et Jarrett, H.W.** (2001). Affinity purification of DNA-binding proteins. *Journal of biochemical and biophysical methods* 49, 607-624.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M., et Barrell, B.** (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498-511.
- Gietz, R.D., et Woods, R.A.** (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in enzymology* 350, 87-96.
- Giordano, M., Beardall, J., et Raven, J.A.** (2005). CO₂ concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu Rev Plant Biol* 56, 99-131.
- Glibert, P.M., Anderson, D.M., Gentien, P., Granéli, E., et Sellner, K.G.** (2005). The global, complex phenomena of harmful algal blooms. *Oceanography* 18, 136-147.
- Gordon, B.R., et Leggat, W.** (2010). Symbiodinium-invertebrate symbioses and the role of metabolomics. *Marine drugs* 8, 2546-2568.
- Grossman, A.R., Bhaya, D., Apt, K.E., et Kehoe, D.M.** (1995). Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annu Rev Genet* 29, 231-288.

- Guillard, R.R., et Ryther, J.H.** (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Grun. Canadian journal of microbiology 8, 229-239.
- Guillebault, D., Derelle, E., Bhaud, Y., et Moreau, H.** (2001). Role of nuclear WW domains and proline-rich proteins in dinoflagellate transcription. *Protist* 152, 127-138.
- Guillebault, D., Sasorith, S., Derelle, E., Wurtz, J.M., Lozano, J.C., Bingham, S., Tora, L., et Moreau, H.** (2002). A new class of transcription initiation factors, intermediate between TATA box-binding proteins (TBPs) and TBP-like factors (TLFs), is present in the marine unicellular organism, the dinoflagellate *Cryptothecodinium cohnii*. *The Journal of biological chemistry* 277, 40881-40886.
- Hastings, J.W.** (2007). The *Gonyaulax* clock at 50: translational control of circadian expression. *Cold Spring Harb Symp Quant Biol* 72, 141-144.
- Hastings, K.E.** (2005). SL trans-splicing: easy come or easy go? *Trends in genetics : TIG* 21, 240-247.
- Hellman, L.M., et Fried, M.G.** (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols* 2, 1849-1861.
- Hiller, R.G., Wrench, P.M., Gooley, A.P., Shoebridge, G., et Breton, J.** (1993). The major intrinsic light-harvesting protein of *Amphidinium*: characterization and relation to other light-harvesting proteins. *Photochem Photobiol* 57, 125-131.
- Howe, C.J., Nisbet, R.E., et Barbrook, A.C.** (2008). The remarkable chloroplast genome of dinoflagellates. *Journal of experimental botany* 59, 1035-1045.
- Igesias-Prieto, R., et Trench, R.K.** (1994). Acclimation and adaptation to irradiance in symbiotic dinoflagellates. I. Responses of the photosynthetic unit to changes in photon flux density *Marine Ecology Progress Series* 113, 147-162.
- Johnson, J.G., Morey, J.S., Neely, M.G., Ryan, J.C., et Van Dolah, F.M.** (2012). Transcriptome remodeling associated with chronological aging in the dinoflagellate, *Karenia brevis*. *Marine genomics* 5, 15-25.
- Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., et Kadonaga, J.T.** (2008). The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* 20, 253-259.
- Kadonaga, J.T., et Tjian, R.** (1986). Affinity purification of sequence-specific DNA binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* 83, 5889-5893.
- Kim, S., Lee, H.-G., et Lee, D.-H.** (2000). Isolation and sequence analysis of promoter DNA fragments of the Luciferin-Binding Protein gene in *Gonyaulax polyedra*. *Journal of plant biology* 43, 88-93.
- Koumandou, V.L., Nisbet, R.E., Barbrook, A.C., et Howe, C.J.** (2004). Dinoflagellate chloroplasts--where have all the genes gone? *Trends in genetics : TIG* 20, 261-267.
- Kwok, A.C., et Wong, J.T.** (2010). The activity of a wall-bound cellulase is required for and is coupled to cell cycle progression in the dinoflagellate *Cryptothecodinium cohnii*. *The Plant cell* 22, 1281-1298.
- Laniel, M.A., Beliveau, A., et Guerin, S.L.** (2001). Electrophoretic mobility shift assays for the analysis of DNA-protein interactions. *Methods in molecular biology* 148, 13-30.

- Lanzer, M., de Bruin, D., et Ravetch, J.V.** (1992). Transcription mapping of a 100 kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates. *The EMBO journal* *11*, 1949-1955.
- Lapointe, M., Mackenzie, T.D., et Morse, D.** (2008). An external delta-carbonic anhydrase in a free-living marine dinoflagellate may circumvent diffusion-limited carbon acquisition. *Plant physiology* *147*, 1427-1436.
- Lapointe, M., et Morse, D.** (2008). Reassessing the role of a 3'-UTR-binding translational inhibitor in regulation of circadian bioluminescence rhythm in the dinoflagellate *Gonyaulax*. *Biological chemistry* *389*, 13-19.
- Le, Q.H., Jovine, R.V., Markovic, P., et Morse, D.** (2001). Peridinin-Chlorophyll a-Protein is not implicated in the photosynthesis rhythm of the dinoflagellate *Gonyaulax* despite circadian regulation of its translation. *Biological Rhythm Research* *32*, 579-594.
- Le, Q.H., Markovic, P., Hastings, J.W., Jovine, R.V., et Morse, D.** (1997). Structure and organization of the peridinin-chlorophyll a-binding protein gene in *Gonyaulax polyedra*. *Mol Gen Genet* *255*, 595-604.
- Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., et Winzeler, E.A.** (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* *301*, 1503-1508.
- Li, L., et Hastings, J.W.** (1998). The structure and organization of the luciferase gene in the photosynthetic dinoflagellate *Gonyaulax polyedra*. *Plant molecular biology* *36*, 275-284.
- Lin, S., Zhang, H., Zhuang, Y., Tran, B., et Gill, J.** (2010). Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 20033-20038.
- Liu, B., Lo, S.C.-L., Matton, D.P., Lang, B.F., et Morse, D.** (2012). Daily Changes in the Phosphoproteome of the Dinoflagellate *Lingulodinium*. *Protist* *163*, 746-754.
- Lowe, C.D., Mello, L.V., Samatar, N., Martin, L.E., Montagnes, D.J., et Watts, P.C.** (2011). The transcriptome of the novel dinoflagellate *Oxyrrhis marina* (Alveolata: Dinophyceae): response to salinity examined by 454 sequencing. *BMC genomics* *12*, 519.
- Lukes, J., Leander, B.S., et Keeling, P.J.** (2009). Cascades of convergent evolution: the corresponding evolutionary histories of euglenozoans and dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* *106 Suppl 1*, 9963-9970.
- Mackenzie, T.D., et Morse, D.** (2011). Circadian photosynthetic reductant flow in the dinoflagellate *Lingulodinium* is limited by carbon availability. *Plant, cell & environment* *34*, 669-680.
- Markovic, P., Roenneberg, T., et Morse, D.** (1996). Phased protein synthesis at several circadian times does not change protein levels in *Gonyaulax*. *Journal of biological rhythms* *11*, 57-67.
- McEwan, M., Humayun, R., Slamovits, C.H., et Keeling, P.J.** (2008). Nuclear genome sequence survey of the Dinoflagellate *Heterocapsa triquetra*. *The Journal of eukaryotic microbiology* *55*, 530-535.

- Meijer, A.H., Ouwerkerk, P.B., et Hoge, J.H.** (1998). Vectors for transcription factor cloning and target site identification by means of genetic selection in yeast. *Yeast* *14*, 1407-1415.
- Meissner, M., et Soldati, D.** (2005). The transcription machinery and the molecular toolbox to control gene expression in *Toxoplasma gondii* and other protozoan parasites. *Microbes and infection / Institut Pasteur* *7*, 1376-1384.
- Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L., Marshall, W.F., Qu, L.H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A.V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riano-Pachon, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Martinez, D., Ngau, W.C., Otilar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I.V., Rokhsar, D.S., et Grossman, A.R.** (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* *318*, 245-250.
- Miao, J., Fan, Q., Cui, L., Li, J., Li, J., et Cui, L.** (2006). The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene* *369*, 53-65.
- Mihailovich, M., Militti, C., Gabaldon, T., et Gebauer, F.** (2010). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays : news and reviews in molecular, cellular and developmental biology* *32*, 109-118.
- Mittag, M., Lee, D.H., et Hastings, J.W.** (1994). Circadian expression of the luciferin-binding protein correlates with the binding of a protein to the 3' untranslated region of its mRNA. *Proceedings of the National Academy of Sciences of the United States of America* *91*, 5257-5261.
- Moreno Diaz de la Espina, S., Alverca, E., Cuadrado, A., et Franca, S.** (2005). Organization of the genome and gene expression in a nuclear environment lacking histones and nucleosomes: the amazing dinoflagellates. *European journal of cell biology* *84*, 137-149.
- Morey, J.S., Monroe, E.A., Kinney, A.L., Beal, M., Johnson, J.G., Hitchcock, G.L., et Van Dolah, F.M.** (2011). Transcriptomic response of the red tide dinoflagellate, *Karenia brevis*, to nitrogen and phosphorus depletion and addition. *BMC genomics* *12*, 346.

- Morse, D., Milos, P.M., Roux, E., et Hastings, J.W.** (1989). Circadian regulation of bioluminescence in *Gonyaulax* involves translational control. *Proceedings of the National Academy of Sciences of the United States of America* *86*, 172-176.
- Morse, D., Salois, P., Markovic, P., et Hastings, J.W.** (1995). A nuclear-encoded form II RuBisCO in dinoflagellates. *Science* *268*, 1622-1624.
- Moustafa, A., Evans, A.N., Kulis, D.M., Hackett, J.D., Erdner, D.L., Anderson, D.M., et Bhattacharya, D.** (2010). Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. *PloS one* *5*, e9688.
- Mumberg, D., Muller, R., et Funk, M.** (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* *156*, 119-122.
- Nassoury, N., Cappadocia, M., et Morse, D.** (2003). Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J Cell Sci* *116*, 2867-2874.
- Nassoury, N., Fritz, L., et Morse, D.** (2001). Circadian changes in ribulose-1,5-bisphosphate carboxylase/oxygenase distribution inside individual chloroplasts can account for the rhythm in dinoflagellate carbon fixation. *The Plant cell* *13*, 923-934.
- Nassoury, N., Wang, Y., et Morse, D.** (2005). Brefeldin A inhibits circadian remodeling of chloroplast structure in the dinoflagellate *Gonyaulax*. *Traffic* *6*, 548-561.
- Okamoto, O.K., et Hastings, J.W.** (2003). Novel dinoflagellate clock-related genes identified through microarray analysis. *Journal of Phycology* *39*, 519-526.
- Ouwerkerk, P.B., et Meijer, A.H.** (2001). Yeast one-hybrid screening for DNA-protein interactions. *Current protocols in molecular biology* / edited by Frederick M Ausubel [et al] *Chapter 12*, Unit 12 12.
- Phadtare, S., et Severinov, K.** (2010). RNA remodeling and gene regulation by cold shock proteins. *RNA biology* *7*, 788-795.
- Polivka, T., van Stokkum, I.H., Zigmantas, D., van Grondelle, R., Sundstrom, V., et Hiller, R.G.** (2006). Energy transfer in the major intrinsic light-harvesting complex from *Amphidinium carterae*. *Biochemistry* *45*, 8516-8526.
- Portis, A.R., Jr., Li, C., Wang, D., et Salvucci, M.E.** (2008). Regulation of Rubisco activase and its interaction with Rubisco. *Journal of experimental botany* *59*, 1597-1604.
- Rae, P.M.M.** (1973). 5-Hydroxymethyluracil in the DNA of a Dinoflagellate. *PNAS* *70*, 1141-1145.
- Rayko, E., Maumus, F., Maheswari, U., Jabbari, K., et Bowler, C.** (2010). Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *The New phytologist* *188*, 52-66.
- Riano-Pachon, D.M., Correa, L.G., Trejos-Espinosa, R., et Mueller-Roeber, B.** (2008). Green transcription factors: a chlamydomonas overview. *Genetics* *179*, 31-39.
- Rizzo, P.J., Jones, M., et Ray, S.M.** (1982). Isolation and properties of isolated nuclei from the Florida red tide dinoflagellate *Gymnodinium breve* (Davis). *The Journal of protozoology* *29*, 217-222.
- Roman, S.J., Govind, N.S., Triplett, E.L., et Prezelin, B.B.** (1988). Light Regulation of Peridinin-Chlorophyll a-Protein (PCP) Complexes in the Dinoflagellate, *Glenodinium* sp. : Use of Anti-Pcp Antibodies to Detect Pcp Gene Products in Cells Grown in Different Light Conditions. *Plant physiology* *88*, 594-599.

Rowan, R., Whitney, S.M., Fowler, A., et Yellowlees, D. (1996). Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *The Plant cell* 8, 539-553.

Roy, S., et Morse, D. (2012). A full suite of histone and histone modifying genes are transcribed in the dinoflagellate *Lingulodinium*. *PloS one* 7, e34340.

Sakurai, H., et Enoki, Y. (2010). Novel aspects of heat shock factors: DNA recognition, chromatin modulation and gene expression. *The FEBS journal* 277, 4140-4149.

Savery, N.J., et Busby, S.J. (1998). Mobility Shift Assays. In, pp. 121-129.

Schulte, T., Johannig, S., et Hofmann, E. (2010). Structure and function of native and refolded peridinin-chlorophyll-proteins from dinoflagellates. *European journal of cell biology* 89, 990-997.

Schulz, M.H., Zerbino, D.R., Vingron, M., et Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086-1092.

Steele, R.E., et Rae, P.M. (1980). Ordered distribution of modified bases in the DNA of a dinoflagellate. *Nucleic acids research* 24, 4709-4725.

Tabita, F.R., Hanson, T.E., Satagopan, S., Witte, B.H., et Kreef, N.E. (2008). Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 363, 2629-2640.

ten Lohuis, M.R., et Miller, D.J. (1998). Light-regulated transcription of genes encoding peridinin chlorophyll a proteins and the major intrinsic light-harvesting complex proteins in the dinoflagellate *amphidinium carterae* hultburt (Dinophyceae). Changes in cytosine methylation accompany photoadaptation. *Plant physiology* 117, 189-196.

Wang, Y., MacKenzie, T., et Morse, D. (2005). Purification of plastids from the dinoflagellate *Lingulodinium*. *Marine biotechnology* 7, 659-668.

Wang, Y., et Morse, D. (2006). Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*. *Nucleic acids research* 34, 613-619.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., et Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239-1243.

Wisecaver, J.H., et Hackett, J.D. (2011). Dinoflagellate genome evolution. *Annual review of microbiology* 65, 369-387.

Yang, V.W. (1998). Eukaryotic transcription factors: identification, characterization and functions. *The Journal of nutrition* 128, 2045-2051.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W., Xiaoming, Q., Yunping, S., Li, Z., Xiaohui, D., Jingchu, L., Xing-Wang, D., Zhangliang, C., Hongya, G., et Li-Jia, Q. (2006). The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant molecular biology* 60, 107-124.

- Young, J.A., Johnson, J.R., Benner, C., Yan, S.F., Chen, K., Le Roch, K.G., Zhou, Y., et Winzeler, E.A.** (2008). In silico discovery of transcription regulatory elements in *Plasmodium falciparum*. *BMC genomics* *9*, 70.
- Zerbino, D.R., et Birney, E.** (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* *18*, 821-829.
- Zhang, H., Hou, Y., Miranda, L., Campbell, D.A., Sturm, N.R., Gaasterland, T., et Lin, S.** (2007). Spliced leader RNA trans-splicing in dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 4618-4623.
- Zhang, Z., Green, B.R., et Cavalier-Smith, T.** (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature* *400*, 155-159.

