

Thèse de Doctorat de l'Université du Maine
Faculté des Sciences et Techniques

Présentée par

Stéphanie Godet

Pour obtenir le grade de Docteur de l'Université du Maine
Spécialité Biophysiology des Organismes et des Populations

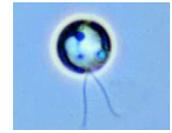


Isolement et caractérisation de trois gènes codant
une lipase et deux estérases hypothétiques
chez la microalgue marine *Isochrysis galbana*
(*Prymnesiophyceae*, *Haptophyta*)

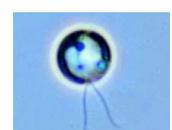
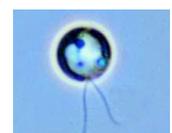


Soutenue le 26 mars 2008 devant le jury composé de :

M. Frédéric Carrière	DR, CNRS EIPL, Marseille	Rapporteur
M. Abdelghani Sghir	PR, Université d'Evry Val d'Essonne	Rapporteur
Mme Annick Manceau	PR, Université du Maine	Examineur
Mme Gwennola Ermel	PR, Université de Rennes I	Examineur
Mme Françoise Ergan	PR, Université du Maine	Directeur de thèse
Mme Josiane Hérault	MCF, Université du Maine	Codirecteur de thèse



La grandeur de la nature se révèle en chaque être vivant aussi microscopique soit-il.



~ Remerciements ~

L'ensemble des travaux présentés dans ce mémoire a été réalisé au sein du Laboratoire d'Applications des Lipases de l'Université du Maine (LALUM), dans le Département Génie Biologique de l'IUT de Laval.

Je remercie les Collectivités Locales Mayennaises pour avoir financé ces travaux de Recherche.

Je remercie Monsieur Frédéric Carrière, Directeur de Recherche au CNRS et Monsieur Abdelghani Sghir, Professeur à l'Université d'Evry Val d'Essonne, qui m'ont fait l'honneur d'accepter de juger ce travail et d'en être rapporteurs. Je remercie Mme Annick Manceau, Professeur à l'Université du Maine et Mme Gwennola Ermel, Professeur à l'Université de Rennes I, pour avoir accepté de participer à ce jury de thèse.

Je remercie Françoise Ergan, Professeur à l'Université du Maine, pour m'avoir accueillie dans son laboratoire de recherche et pour la confiance qu'elle m'a accordée pour mener ce projet.

J'exprime toute ma reconnaissance envers Josiane Héroult et Gaëlle Pencreac'h, Maîtres de Conférences à l'Université du Maine, pour avoir codirigé ce travail, pour leur confiance, leur disponibilité, leurs conseils et pour le temps consacré à la correction de ce manuscrit.

Je remercie Céline Loiseau, Maître de Conférences à l'Université du Maine, pour sa disponibilité constante, sa participation à ce projet et ses conseils en matière d'enseignement.

Je remercie également les membres de l'équipe du LALUM, Laurent Poisson et Samia Soultani, Maîtres de Conférences à l'Université du Maine.

Je remercie Isabelle Martin, Ingénieur d'Etude, pour sa gentillesse, sa disponibilité et son aide très précieuse dans la commande et la gestion des réactifs et du matériel.

Je remercie Rose-Marie Leroux, Technicienne, pour ses qualités humaines et sa participation enthousiaste aux expérimentations. Merci aussi à Frédérique Guéno.

J'exprime tout mon soutien et mes encouragements aux doctorants Cécile Lorentz et Freddy Guihemeuf. Je remercie Marie Devos pour m'avoir initiée à l'analyse des lipides de microalgues. Je remercie les stagiaires Mélanie Monnet, Annabelle Suisse et Pierre Goffi.

Je tiens à remercier tous les enseignants du Département Génie Biologique pour leur aide et leurs précieux conseils, plus particulièrement Laurent Gauvry et Vincent Blanckaert, Françoise Lourdaï, Gilles Braud, Sophie Ledru et Jean-Michel Launay.

Je remercie également Isabelle Brou et Jean Luc Hoarau pour l'accueil qu'ils nous réservent au secrétariat, leur bonne humeur quotidienne et leur efficacité professionnelle.

Je pense très fort à tous mes amis que je n'ai pas vus depuis longtemps, Sandra, Nicolas, Magali, Matthieu, Fabienne et Gérald, à leurs enfants et à ma petite filleule Maud. Mes amis préparez vous à fêter comme il se doit mon enterrement de vie de thésarde!

A ma grand mère, mon père et ma mère pour leur soutien constant

A mon frère Romain

A mon amoureux Sébastien

A mon chat qui est resté à mes côtés pendant ces mois douloureux de rédaction!

~ Table des matières ~

~ Liste des figures ~	1
~ Liste des tableaux ~	6
~ Liste des annexes ~	8
~ Liste des abréviations ~	9
~ Code génétique ~	11
~ Code des acides aminés ~	12
~ Liste des acides gras cités ~	13
INTRODUCTION GENERALE.....	14
SYNTHESE BIBLIOGRAPHIQUE.....	18
1. Lipases et carboxylestérases : enzymes lipolytiques.....	19
1.1. Définition	19
1.2. Les lipides	19
1.3. Réactions catalysées	21
1.4. Exemples de fonctions biologiques.....	22
1.5. Histoire moléculaire évolutive	22
1.6. Caractéristiques structurales.....	23
1.7. Mécanisme d'action	28
1.8. Spécificité de substrat.....	29
1.9. Classification des enzymes lipolytiques.....	30
1.10. Les enzymes lipolytiques de microalgues.....	37
2. Les microalgues marines	38
2.1. Les microalgues marines dans la Biosphère	38
2.2. L'histoire évolutive des microalgues	39
2.3. Position phylogénétique des microalgues	44
2.4. Microalgues et génomique	46
2.5. Lipides et acides gras de microalgues	48
2.6. La microalgue <i>Isochrysis galbana</i>	52
3. Applications industrielles des enzymes lipolytiques et valorisation des lipides de microalgues	56

3.1.	Applications industrielles des enzymes lipolytiques.....	56
3.2.	Valorisation commerciale des acides gras et lipides de microalgues.....	59
MATERIELS ET METHODES		63
1.	Matériel biologique	64
1.1.	<i>Isochrysis galbana</i> Parke	64
1.2.	Milieu de culture Provasoli	64
1.3.	Culture en Erlenmeyers	65
1.4.	Culture en photobioréacteur	65
1.5.	Suivi de la croissance et récolte de la biomasse.....	66
2.	Analyse des lipides.....	66
2.1.	Extraction des lipides totaux	66
2.2.	Analyse par chromatographie sur couche mince.....	66
3.	Mesure des activités lipolytiques	67
3.1.	Mise en évidence d'une activité lipolytique endogène chez <i>I. galbana</i>	67
3.2.	Mesure de l'activité lipolytique de <i>Saccharomyces cerevisiae</i> transformée avec le plasmide pYES2 recombiné	68
4.	Isolement des acides nucléiques.....	69
4.1.	Extraction et purification des ARN totaux.....	69
4.2.	Purification des ARN poly (A ⁺)	70
4.3.	Extraction et purification de l'ADN génomique	71
4.4.	Extraction et purification d'ADN de bactériophages.....	71
4.5.	Extraction et purification d'ADN plasmidique	72
5.	Analyses des acides nucléiques.....	73
5.1.	Analyse quantitative	73
5.2.	Analyse qualitative	74
6.	Techniques d'amplification des acides nucléiques	74
6.1.	Transcription inverse et réaction de polymérisation en chaîne	74
6.2.	Amplification rapide des extrémités des ADNc (RACE-PCR)	76
6.3.	Nested-PCR.....	79
6.4.	Conception des amorces.....	79
6.5.	Séquençage des produits amplifiés	85
7.	Construction et criblage d'une banque d'ADN complémentaires	85

7.1.	Construction d'une banque phagique d'ADNc	85
7.2.	Criblage de la banque d'ADNc	87
8.	Clonage et expression des séquences codantes	87
8.1.	Le vecteur de clonage et d'expression : le plasmide pYES2	87
8.2.	L'insert (ADNc)	88
8.3.	Digestion enzymatique du plasmide et de l'insert (ADNc)	89
8.4.	Ligation du plasmide et de l'insert	89
8.5.	Préparation de bactéries et de levures électrocompétentes	89
8.6.	Transformation par électroporation.....	90
8.7.	Sélection des transformants.....	90
8.8.	Analyse des clones positifs	90
8.9.	Induction de l'expression du gène cloné et analyse des transcrits	91
9.	Analyses bioinformatiques des séquences	92
9.1.	Bases de séquences.....	92
9.2.	Analyse des séquences nucléotidiques.....	94
9.3.	Profils physicochimiques des protéines	94
9.4.	Recherche d'homologies	95
9.5.	Alignements multiples.....	96
9.6.	Phylogénie.....	97
9.7.	Prédiction de la structure tertiaire des protéines	99
RESULTATS ET DISCUSSION.....		100
1.	Analyse de l'activité lipolytique endogène d' <i>Isochrysis galbana</i>	101
1.1.	Mise en évidence d'une activité lipolytique.....	101
1.2.	Localisation de l'activité lipolytique.....	102
1.3.	Conclusion.....	104
2.	Isolement de 3 gènes codant des enzymes lipolytiques chez <i>Isochrysis galbana</i>	104
2.1.	Construction et criblage d'une banque d'ADN complémentaires	104
2.2.	Exploitation de la banque EST d' <i>Isochrysis galbana</i> (Keeling, 2006).....	114
2.3.	Isolement de 3 ADNc codant des enzymes lipolytiques par la méthode de RACE-PCR	118
3.	Caractérisation préliminaire et analyses comparatives du gène <i>IgLip</i> codant une lipase hypothétique chez <i>Isochrysis galbana</i>	124

3.1.	Analyse et traduction de la séquence nucléotidique <i>IgLip</i>	124
3.2.	Prédiction des caractéristiques physicochimiques	126
3.3.	Prédiction de la fonction enzymatique	127
3.4.	Prédiction de l'organisation structurale.....	133
3.5.	Séquences homologues à <i>IgLip</i> chez les microalgues.....	138
3.6.	Classification d' <i>IgLip</i> et de ses homologues.....	141
3.7.	Structure intron/exon du gène <i>IgLip</i>	142
4.	Caractérisation préliminaire et analyses comparatives du gène <i>IgEst1</i> codant une estérase hypothétique chez <i>Isochrysis galbana</i>	143
4.1.	Analyse et traduction de la séquence nucléotidique <i>IgEst1</i>	143
4.2.	Prédiction des caractéristiques physicochimiques	144
4.3.	Prédiction de la fonction enzymatique	146
4.4.	Prédiction de l'organisation structurale.....	151
4.5.	Séquences homologues à <i>IgEst1</i> chez les microalgues	153
4.6.	Classification et analyse phylogénétique	158
5.	Caractérisation préliminaire et analyses comparatives du gène <i>IgEst2</i> codant une estérase hypothétique chez <i>Isochrysis galbana</i>	162
5.1.	Analyse et traduction de la séquence nucléotidique <i>IgEst2</i>	162
5.2.	Prédiction des caractéristiques physicochimiques	163
5.3.	Prédiction de la fonction enzymatique	165
5.4.	Prédiction de l'organisation structurale.....	168
5.5.	Séquences homologues à <i>IgEst2</i> chez les microalgues	170
5.6.	Classification.....	172
5.7.	Structure intron/exon du gène <i>IgEst2</i>	173
6.	Clonage et induction de l'expression des gènes.....	174
6.1.	Clonage de la séquence codante <i>IgLip</i>	174
6.2.	Clonage et expression de la séquence codante <i>IgEst1</i>	177
6.3.	Clonage et expression de la séquence codante <i>IgEst2</i>	181
7.	Discussion générale.....	187
7.1.	Confirmation d'une activité lipolytique endogène chez <i>Isochrysis galbana</i>	187
7.2.	Présence de 10 gènes susceptibles de coder des enzymes lipolytiques chez <i>Isochrysis galbana</i>	188

7.3. Isolement de 3 gènes codant 1 lipase et 2 estérases hypothétiques chez <i>Isochrysis galbana</i>	189
7.4. Présence d'isoenzymes chez <i>Isochrysis galbana</i>	190
7.5. Présence de gènes avec et sans introns chez <i>Isochrysis galbana</i>	191
CONCLUSION ET PERSPECTIVES	192
ANNEXES	195
BIBLIOGRAPHIE	221

~ Liste des figures ~

Figure 1 : Structure simplifiée de quelques acides gras.	20
Figure 2 : Structure simplifiée des triglycérides.	20
Figure 3 : Structure simplifiée des phospholipides.	21
Figure 4 : Hydrolyse des triglycérides.	21
Figure 5 : Diversité des réactions catalysées par les estérases.	22
Figure 6 : Représentation schématique de la structure secondaire d'une hydrolase à repliement α/β (Ollis <i>et al.</i> , 1992).	24
Figure 7 : Structure tridimensionnelle de la carboxylestérase monomérique du kiwi <i>Actinidia eriantha</i> (PDB 2o7r) (Ileperuma <i>et al.</i> , 2007).	24
Figure 8 : Représentation schématique des deux conformations d'une lipase.	26
Figure 9 : Représentation moléculaire des deux conformations de la lipase de <i>Thermomyces lanuginosa</i> d'après Svendsen, 2000.	27
Figure 10 : Schématisation de l'hydrolyse d'un ester par une enzyme lipolytique d'après Fischer, 2004.	28
Figure 11 : Arbre phylogénétique des hydrolases à repliement α/β (Fischer, 2004).	36
Figure 12 : L'origine endosymbiotique des microalgues d'après Delwiche, 1999.	42
Figure 13 : Transferts de gènes endosymbiotiques d'après Li <i>et al.</i> , 2006.	43
Figure 14 : Arbre phylogénétique des lignées eucaryotes d'après Keeling <i>et al.</i> , 2005.	45
Figure 15 : Effet de la température sur la composition en acide arachidonique des lipides neutres et polaires de <i>Parietochloris incisa</i> d'après Cohen <i>et al.</i> , 2000.	51
Figure 16 : <i>Isochrysis galbana</i> Parke, 1939.	53
Figure 17 : Arbre phylogénétique des haptophytes (Edwardsen <i>et al.</i> , 2000; Yoshida <i>et al.</i> , 2006).	54
Figure 18 : Synthèse et encadrement des ADNc, selon le mécanisme SMART TM (Switching Mechanism At 5' end of RNA Transcript) (d'après Clontech).	77
Figure 19 : Placement des couples d'amorces spécifiques et universelles utilisées en RACE PCR et Nested PCR (d'après Clontech).	78
Figure 20 : Alignement des iPLA ₂ hypothétiques des microalgues <i>Cyanidoschyzon merolae</i> (<i>CmiPLA2</i>) et <i>Thalassiosira pseudonana</i> (<i>TpiPLA2</i>) avec leur homologue la iPLA ₂ humaine (<i>HsiPLA2</i>).	80

Figure 21 : Conception des amorces sens plus ou moins dégénérées calquées sur les motifs conservés chez les iPLA ₂ hypothétiques des microalgues <i>Cyanidoschyzon merolae</i> et <i>Thalassiosira pseudonana</i>	81
Figure 22 : Le vecteur Zap express (Stratagène).....	86
Figure 23 : Le vecteur d'expression pYES2 (Invitrogen).....	88
Figure 24 : Chromatographie sur couche mince des lipides totaux d' <i>Isochrysis galbana</i> pour la mise en évidence de l'activité lipolytique endogène.....	102
Figure 25 : Chromatographie sur couche mince des lipides totaux d' <i>Isochrysis galbana</i> pour la localisation intracellulaire de l'activité lipolytique endogène.....	103
Figure 26 : Electrophorèses en gels d'agarose des ARN totaux et des ARNm d' <i>Isochrysis galbana</i>	105
Figure 27 : Electrophorèse en gel d'agarose des ADNc double brin fractionnés en 3 groupes de tailles différentes.	106
Figure 28 : Electrophorèse en gel d'agarose d'ADN extrait de phages recombinés prélevés aléatoirement dans les banques 1, 2 et 3, et amplifié par PCR avec les amorces T3S et T7AS.	107
Figure 29 : Electrophorèses en gels d'agarose d'ADNc amplifiés avec des amorces hétérologues non dégénérées.....	109
Figure 30 : Electrophorèses en gels d'agarose d'ADN extrait de phages recombinés (banques 2 et 3) et amplifié par PCR avec le couple d'amorces <i>CmPLM2S/T7AS</i>	110
Figure 31 : Alignement local de la séquence du fragment 2 amplifié par PCR avec les amorces <i>CmPLM2S</i> et <i>T7AS</i> dans le cadre du criblage de la banque d'ADNc 2.	111
Figure 32 : Electrophorèses en gels d'agarose d'ADN extrait de phages recombinés (banques 2 et 3) et amplifié par PCR avec le couple d'amorces <i>T3S/CmPLM2AS</i>	112
Figure 33 : Electrophorèses en gels d'agarose d'ADN extrait de phages recombinés (banques 2 et 3) et amplifié par PCR avec le couple d'amorces <i>CmLipMS/T7AS</i>	113
Figure 34 : Séquence nucléotidique du cluster ISL00001844.....	117
Figure 35 : Séquence nucléotidique du cluster ISL00008005.....	118
Figure 36 : Séquence nucléotidique du cluster ISL00001076.....	118
Figure 37 : Electrophorèses en gels d'agarose des régions terminales 5' et 3' des ADNc <i>IgLip</i> , <i>IgEst1</i> et <i>IgEst2</i> amplifiées par RACE PCR.	119
Figure 38 : Extrait du chromatogramme de séquençage de l'échantillon 3'-RACE <i>IgLip</i> amplifié par RACE PCR (MilleGen, France).	120

Figure 39 : Electrophorèses en gels d'agarose des régions terminales 3' et 5' des ADNc <i>IgLip</i> , <i>IgEst1</i> et <i>IgEst2</i> amplifiées par Nested PCR.	121
Figure 40 : Extrait du chromatogramme de séquençage de l'échantillon 3'-RACE <i>IgLip</i> amplifié par RACE et Nested PCR (MilleGen, France).	122
Figure 41 : Electrophorèses en gels d'agarose des séquences codantes complètes des ADNc <i>IgLip</i> , <i>IgEst1</i> et <i>IgEst2</i>	123
Figure 42 : Séquence nucléotidique de l'ADNc <i>IgLip</i> et traduction de la région codante.....	125
Figure 43 : Profils hydrophobe et hydrophile de la protéine <i>IgLip</i>	126
Figure 44 : Alignement de la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) avec trois de ses homologues chez les plantes supérieures <i>Arabidopsis thaliana</i> (<i>AtLip</i>), <i>Oryza sativa</i> (<i>OsLip</i>) et <i>Triticum aestivum</i> (<i>TaLip</i>).	128
Figure 45 : Alignement de la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) avec cinq de ses homologues chez les trypanosomes, <i>Leishmania braziliensis</i> (<i>LbLip</i>), <i>Leishmania infantum</i> (<i>LiLip</i>), <i>Leishmania major</i> (<i>LmLip</i>), <i>Trypanosoma cruzi</i> (<i>TcLip</i>) et <i>Trypanosoma brucei</i> (<i>TbLip</i>).	131
Figure 46 : Alignement de la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) avec la lipase de <i>Thermomyces lanuginosa</i> (<i>TlLip</i>) (Swiss-Prot O59952 ; code PDB 1dt3).	136
Figure 47 : Alignement de la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) avec la lipase de <i>Rhizomucor miehei</i> (<i>RmLip</i>) (Swiss-Prot P19515 ; code PDB 4tgl).	137
Figure 48 : Alignement de la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) avec des homologues de lipase chez les microalgues <i>Cyanidioschyzon merolae</i> (<i>CmLip</i>), <i>Thalassiosira pseudonana</i> (<i>TpLip1</i> et <i>TpLip2</i>), <i>Phaeodactylum tricornutum</i> (<i>PtLip1</i> , <i>PtLip2</i> et <i>PtLip3</i>), <i>Chlamydomonas reinhardtii</i> (<i>CrLip1</i> et <i>CrLip2</i>), <i>Ostreococcus lucimarinus</i> (<i>OlLip1</i> et <i>OlLip2</i>) et <i>Ostreococcus tauri</i> (<i>OtLip1</i> , <i>OtLip2</i> , <i>OtLip3</i> et <i>OtLip4</i>).	140
Figure 49 : Alignement de la région C-terminale de la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) avec le marqueur EG035443 de la microalgue <i>Emiliana huxleyi</i>	141
Figure 50 : Structure intron / exon du gène codant la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>).	142
Figure 51 : Séquence nucléotidique de l'ADNc <i>IgEst1</i> et traduction de la région codante...	143
Figure 52 : Prédiction de la présence d'un signal peptidique dans la région N-terminale d' <i>IgEst1</i> , avec le serveur SignalP.	145
Figure 53 : Profils hydrophobe et hydrophile de la protéine <i>IgEst1</i>	146

Figure 54 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) avec ses homologues chez les champignons <i>Aspergillus niger</i> (<i>AnPLip/CEst</i>) et <i>Aspergillus oryzae</i> (<i>AoPLip/CEst</i>).	147
Figure 55 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) avec ses homologues chez l'homme <i>Homo sapiens</i> (<i>HsLPLipI</i> et <i>II</i>), la souris <i>Mus musculus</i> (<i>MmLPLipI</i>), la drosophile <i>Drosophila melanogaster</i> (<i>DmPLip/CEst</i>) et les plantes <i>Arabidopsis thaliana</i> (<i>AtPLip/CEst</i>) et <i>Oryza sativa</i> (<i>OsPLip/CEst</i>).	149
Figure 56 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) avec la thioestérase humaine (<i>HsATEst</i>) (Swiss-Prot 075608 ; code PDB 1fj2).	152
Figure 57 : Modèle structural de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>).	153
Figure 58 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) avec des homologues de phospholipase et de carboxylestérase chez les microalgues <i>Ostreococcus lucimarinus</i> (<i>OiPLip/CEst1</i> , <i>OiPLip/CEst2</i>), <i>Ostreococcus tauri</i> (<i>OtPLip/CEst</i>), <i>Phaeodactylum tricornutum</i> (<i>PtPLip/CEst</i>), <i>Thalassiosira pseudonana</i> (<i>TpPLip/CEst1</i> , <i>TpPLip/CEst2</i>) et <i>Cyanidioschyzon merolae</i> (<i>CmLPLipII</i>).	155
Figure 59 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) avec deux de ses homologues chez les diatomées <i>Phaeodactylum tricornutum</i> (<i>PtPLip/CEst</i>) et <i>Thalassiosira pseudonana</i> (<i>TpPLip/CEst2</i>).	157
Figure 60 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) avec un homologue de microorganisme planctonique marin.	158
Figure 61 : Phylogénie de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) et de ses homologues, obtenue avec la méthode du Neighbor Joining sur les régions extraites par Gblocks.....	160
Figure 62 : Phylogénie de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) et de ses homologues, obtenue avec la méthode du Maximum de Parcimonie sur les régions extraites par Gblocks.	161
Figure 63 : Séquence nucléotidique de l'ADNc <i>IgEst2</i> et traduction de la région codante...	162
Figure 64 : Prédiction de la présence d'un signal peptidique dans la région N-terminale d' <i>IgEst2</i> , avec le serveur SignalP.....	163
Figure 65 : Profils hydrophobe et hydrophile de la protéine <i>IgEst2</i>	164
Figure 66 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst2</i>) avec six de ses homologues chez les bactéries <i>Flavobacterium johnsoniae</i> (<i>FjPLip/CEst</i>), <i>Chlorobium limicola</i> (<i>ClPLip/CEst</i>), <i>Chloroflexus aurantiacus</i> (<i>CaPLip/CEst</i>),	

<i>Roseiflexus</i> sp. (<i>RspPLip/CEst</i>), <i>Herpetosiphon aurantiacus</i> (<i>HaPLip/CEst</i>) et une cyanobactérie <i>Synechococcus</i> sp. (<i>SspPLip/CEst</i>).....	166
Figure 67 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst2</i>) avec son plus proche homologue identifié chez la bactérie <i>Herpetosiphon aurantiacus</i> (<i>HaPLip/CEst</i>).....	167
Figure 68 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst2</i>) avec la carboxylestérase de <i>Pseudomonas fluorescens</i> (<i>PfCEst</i>) (Swiss-Prot Q53547 ; code PDB 1auo).....	169
Figure 69 : Modèle structural de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst2</i>).	170
Figure 70 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst2</i>) avec sept de ses homologues chez les microalgues <i>Ostreococcus tauri</i> (<i>OtPLip/CEst</i>), <i>Ostreococcus lucimarinus</i> (<i>OlPLip/CEst1</i> et <i>OlPLip/CEst2</i>), <i>Phaeodactylum tricornutum</i> (<i>PtPLip/CEst</i>), <i>Thalassiosira pseudonana</i> (<i>TpPLip/CEst1</i> et <i>TpPLip/CEst2</i>) et <i>Cyanidioschyzon merolae</i> (<i>CmLPLipII</i>).....	171
Figure 71 : Alignement de l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst2</i>) avec les phospholipases A ₂ Ca ²⁺ -indépendantes des vertébrés <i>Oryctolagus cuniculus</i> (<i>OcPLA2</i>) et <i>Rattus norvegicus</i> (<i>RnPLA2</i>).....	173
Figure 72 : Construction génétique du vecteur pYES2/IgLip.....	175
Figure 73 : Electrophorèse en gel d'agarose du plasmide pYES2/IgLip digéré ou amplifié (clone 1).....	176
Figure 74 : Construction génétique du vecteur pYES2/IgEst1.	178
Figure 75 : Electrophorèse en gel d'agarose du plasmide pYES2/IgEst1 digéré ou amplifié (clone A).....	179
Figure 76 : Alignement <i>via</i> ClustalW de la séquence protéique déduite du transcrit isolé chez <i>Isochrysis galbana</i> (<i>IgEst1</i>) avec la séquence de l'insert cloné chez <i>Escherichia coli</i> (clone A).....	180
Figure 77 : Electrophorèse en gel d'agarose du transcrit <i>IgEst1</i> amplifié par PCR après induction de l'expression.	181
Figure 78 : Construction génétique du vecteur pYES2/IgEst2.	182
Figure 79 : Electrophorèse en gel d'agarose du plasmide pYES2/IgEst2 digéré ou amplifié (clone α).	183
Figure 80 : Electrophorèse en gel d'agarose du transcrit <i>IgEst2</i> amplifié par PCR après induction de l'expression.	184

~ Liste des tableaux ~

Tableau 1 : Comparaison entre les régions du volet amphiphile des isozymes de lipase du champignon <i>Candida rugosa</i> (Lotti <i>et al.</i> , 1994).....	27
Tableau 2 : Huit familles d'enzymes lipolytiques bactériennes d'après Arpigny et Jaeger, 1999.....	31
Tableau 3 : Liste des superfamilles et familles d'homologues figurant dans la Lipase Engineering Database (Fischer, 2004).	33
Tableau 4 : Quelques données statistiques relatives aux génomes nucléaires de microalgues.	47
Tableau 5 : Profil en acides gras des lipides d' <i>Isochrysis galbana</i> cultivée en bioréacteur de 300 L (Devos, 2005).....	55
Tableau 6 : Applications industrielles des enzymes lipolytiques d'après Jaeger et Eggert, 2002, Hasan <i>et al.</i> , 2006.	58
Tableau 7 : Teneur en lipides de quelques espèces de microalgues d'après Chisti, 2007.	59
Tableau 8 : Principales espèces de microalgues d'intérêt pour la production d'EPA et de DHA (Pencreac'h <i>et al.</i> , 2004).	61
Tableau 9 : Séquences des amorces T3 et T7 et des amorces non dégénérées spécifiques de la microalgue <i>Cyanidoschyzon merolae</i>	82
Tableau 10 : Liste des couples d'amorces spécifiques d' <i>Isochrysis galbana</i>	84
Tableau 11 : Echelle d'hydrophobicité de Kyte et Doolittle.....	95
Tableau 12 : Liste des espèces et des séquences utilisées pour l'analyse phylogénétique.	98
Tableau 13 : Liste des clusters identifiés dans la banque EST d' <i>Isochrysis galbana</i> présentant des homologies significatives avec des enzymes lipolytiques.	116
Tableau 14 : Pourcentages d'identité et de similitude entre la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) et ses homologues végétales chez <i>Arabidopsis thaliana</i> (<i>AtLip</i>), <i>Oryza sativa</i> (<i>OsLip</i>) et <i>Triticum aestivum</i> (<i>TaLip</i>).....	129
Tableau 15 : Conservation de la région du coude nucléophile chez la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) et les lipases homologues d' <i>Arabidopsis thaliana</i> (<i>AtLip</i>), <i>Oryza sativa</i> (<i>OsLip</i>) et <i>Triticum aestivum</i> (<i>TaLip</i>).	129
Tableau 16 : Pourcentages d'identité et de similitude entre la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) et ses homologues chez les trypanosomes, <i>Leishmania braziliensis</i>	

(<i>LbLip</i>), <i>Leishmania infantum</i> (<i>LiLip</i>), <i>Leishmania major</i> (<i>LmLip</i>), <i>Trypanosoma cruzi</i> (<i>TcLip</i>) et <i>Trypanosoma brucei</i> (<i>TbLip</i>).....	132
Tableau 17 : Conservation de la région du coude nucléophile de la lipase hypothétique d' <i>Isochrysis galbana</i> (<i>IgLip</i>) et des lipases homologues de trypanosomes <i>Leishmania braziliensis</i> (<i>LbLip</i>), <i>Leishmania infantum</i> (<i>LiLip</i>), <i>Leishmania major</i> (<i>LmLip</i>), <i>Trypanosoma cruzi</i> (<i>TcLip</i>) et <i>Trypanosoma brucei</i> (<i>TbLip</i>).....	132
Tableau 18 : Pourcentages d'identité et de similitude entre l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) et ses homologues fongiques chez <i>Aspergillus niger</i> (<i>AnPLip/CEst</i>) et <i>Aspergillus oryzae</i> (<i>AoPLip/CEst</i>).....	147
Tableau 19 : Pourcentages d'identité et de similitude entre l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) et ses homologues humaines (<i>HsLPLipI</i> et <i>II</i>), murine (<i>MmLPLipI</i>) et végétales (<i>AtPLip/CEst</i> , <i>OsPLip/CEst</i>).....	150
Tableau 20 : Pourcentages d'identité et de similitude entre l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst1</i>) et ses homologues microalgales chez <i>Ostreococcus lucimarinus</i> (<i>OiPLip/CEst1</i> , <i>OiPLip/CEst2</i>), <i>Ostreococcus tauri</i> (<i>OtPLip/CEst</i>), <i>Phaeodactylum tricornutum</i> (<i>PtPLip/CEst</i>), <i>Thalassiosira pseudonana</i> (<i>TpPLip/CEst1</i> , <i>TpPLip/CEst2</i>) et <i>Cyanidioschyzon merolae</i> (<i>CmLPLipII</i>).	156
Tableau 21 : Pourcentages de résidus identiques et similaires entre l'estérase hypothétique d' <i>Isochrysis galbana</i> (<i>IgEst2</i>) et ses homologues procaryotiques <i>FjPLip/CEst</i> , <i>ClPLip/CEst</i> , <i>CaPLip/CEst</i> , <i>RspPLip/CEst</i> , <i>HaPLip/CEst</i> et <i>SspPLip/CEst</i>	167
Tableau 22 : Pourcentages de résidus identiques et similaires entre l'estérase hypothétique d' <i>I. galbana</i> (<i>IgEst2</i>) et ses homologues microalgales chez <i>Ostreococcus tauri</i> (<i>OtPLip/CEst</i>), <i>Ostreococcus lucimarinus</i> (<i>OiPLip/CEst1</i> et <i>OiPLip/CEst2</i>), <i>Phaeodactylum tricornutum</i> (<i>PtPLip/CEst</i>), <i>Thalassiosira pseudonana</i> (<i>TpPLip/CEst1</i> et <i>TpPLip/CEst2</i>) et <i>Cyanidioschyzon merolae</i> (<i>CmLPLipII</i>).	172
Tableau 23 : Mesure de l'activité lipolytique des levures transformées avec le plasmide recombiné pYES2/ <i>IgEst2</i>	185

~ Liste des annexes ~

Annexe 1 : Principe de la réaction de polymérisation en chaîne (PCR)	196
Annexe 2 : Principe de la réaction 3'-RACE (Clontech)	197
Annexe 3 : Principe de la réaction 5'-RACE (Clontech)	198
Annexe 4 : Synthèse des ADNc pour la construction de la banque d'ADNc (Stratagène)....	199
Annexe 5 : Liste des bases de données et des outils informatiques	200
Annexe 6 : Séquences nucléotidiques des fragments amplifiés dans le cadre du criblage de la banque d'ADNc par PCR	201
Annexe 7 : Clusters d'enzymes lipolytiques indexés dans la Base EST d' <i>I. galbana</i> (Keeling, 2006).....	202
Annexe 8 : Profil des blocs A, B, C, D, E et F (format WebLogo) de la famille Lipase 3 indexée IPB005592 dans la base de données Blocks.	205
Annexe 9 : Alignement des séquences génomique et d'ADNc d' <i>IgLip</i>	207
Annexe 10 : Profil des blocs A et B (format WebLogo) de la famille Phospholipase / Carboxylestérase indexée IPB003140 dans la base de données Blocks.	209
Annexe 11 : Extraction de signaux phylogénétiques fiables avec Gblocks	210
Annexe 12 : Alignement des séquences génomique et d'ADNc d' <i>IgEst2</i>	216
Annexe 13 : Analyse du plasmide recombiné pYES2/ <i>IgLip</i> (clone 1) et séquençage de l'insert.....	217
Annexe 14 : Analyse du plasmide recombiné pYES2/ <i>IgEst1</i> (clone A) et séquençage de l'insert.....	219
Annexe 15 : Analyse du plasmide recombiné pYES2/ <i>IgEst2</i> (clone α) et séquençage de l'insert.....	220

~ Liste des abréviations ~

AA	Acide Arachidonique
AASp	Amorce Anti Sens spécifique
ADN	Acide DésoxyriboNucléique
ADNc	Acide DésoxyriboNucléique complémentaire
AFSSA	Agence Française de Sécurité Sanitaire des Aliments
AGPI	Acides Gras PolyInsaturés
AGPI-LC	Acides Gras PolyInsaturés à Longues Chaînes
AGPI- ω 3	Acides Gras PolyInsaturés Oméga 3
ARN	Acide RiboNucléique
ARNr	Acide RiboNucléique ribosomique
ARNm	Acide RiboNucléique messenger
ARN poly(A ⁺)	Acide RiboNucléique polyadénylé
ARNt	Acide RiboNucléique de transfert
ARNT	Acide RiboNucléique Total
ASp	Amorce Sens spécifique
AU	Amorce Universelle
BLAST	Basic Local Alignment Search Tool
CDS	CoDing Sequence
CEst	CarboxylEstérase
CCM	Chromatographie sur Couche Mince
Da	Dalton
dAS	Amorce dégénérée Anti Sens
DHA	DocosaHexaenoic Acid
dNTP	désoxyriboNucléotide TriPhosphate
dS	Amorce dégénérée Sens
EC	Enzyme Commission
EPA	EicosaPentaenoic Acid
EST	Expressed Sequence Tag
Est	Estérase
IgEst1	<i>Isochrysis galbana</i> Estérase 1
IgEst2	<i>Isochrysis galbana</i> Estérase 2
IgLip	<i>Isochrysis galbana</i> Lipase
LED	Lipase Engineering Database
Lip	Lipase
LPLip	LysoPhosphoLipase
NASp	Nested Amorce Sens spécifique
NAASp	Nested Amorce Anti Sens spécifique
NAU	Nested Amorce Universelle
ORF	Open Reading Frame
pb	paire de base
PCR	Polymerase Chain Reaction
PLA ₂	PhosphoLipase A ₂
cPLA ₂	PhosphoLipase A ₂ dépendante du calcium
iPLA ₂	PhosphoLipase A ₂ indépendante du calcium

PLip	PhosphoLipase
PS RT	PowerScript Reverse Transcriptase
PDB	Protein Data Bank
RACE PCR	Rapid Amplification of cDNA Ends Polymerase Chain Reaction
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SSU rRNA	Small SubUnit of ribosomal RiboNucleic Acid
ufp	unité formant une plage de lyse

~ Code génétique ~

Première position (extrémité 5')	Seconde position				Troisième position (extrémité 3')
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Code wobble		Code wobble	
R	A/G	B	C/G/T
Y	C/T	D	A/G/T
M	A/C	H	A/C/T
K	G/T	V	A/C/G
S	C/G	N	A/C/G/T
W	A/T		

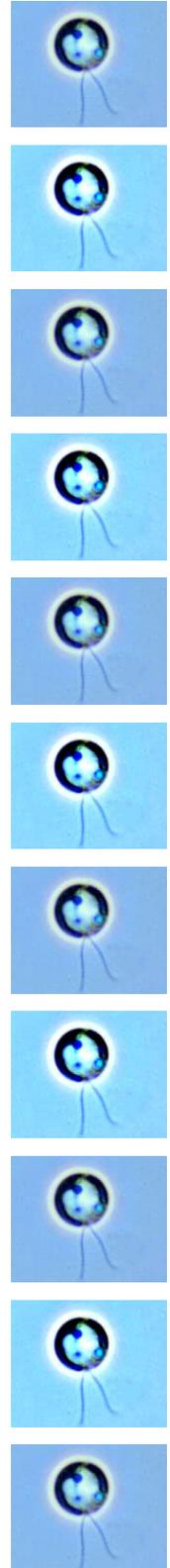
~ Code des acides aminés ~

Nom entier	Abréviation en trois lettres	Code à une lettre
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Leucine	Leu	L
Isoleucine	Ile	I
Sérine	Ser	S
Thréonine	Thr	T
Cystéine	Cys	C
Méthionine	Met	M
Acide Aspartique	Asp	D
Asparagine	Asn	N
Acide glutamique	Glu	E
Glutamine	Gln	Q
Lysine	Lys	K
Arginine	Arg	R
Histidine	His	H
Phénylalanine	Phe	F
Tyrosine	Tyr	Y
Tryptophane	Trp	W
Proline	Pro	P

~ Liste des acides gras cités ~

	Nom usuel	Nom systématique
C14 :0	Acide myristique	Acide tétradécanoïque
C16 :0	Acide palmitique	Acide hétéradécanoïque
C18 :1 ω9	Acide oléique	Acide cis-9 octadécénoïque
C18 :2 ω6	Acide linoléique	Acide cis-9, cis-12 octadécadiénoïque
C18 :3 ω3	Acide α-linolénique	Acide cis-9, cis-12, cis-15 octadécatriénoïque
C18 :3 ω6	Acide γ-linolénique	Acide cis-6, cis-9, cis-12 octadécatriénoïque
C18 :4 ω3	Acide stéaridonique	Acide cis-6, cis-9, cis-12, cis-15 octadécatétraénoïque
C20 :4 ω6	Acide arachidonique	Acide cis-5, cis-8, cis-11, cis-14 eicosatétraénoïque
C20 :5 ω3 EPA	Acide timnodonique	Acide cis-5, cis-8, cis-11, cis-14, cis-17 eicosapentaénoïque
C22 :4 ω3	-	Acide cis-10, cis-13, cis-16, cis-19 docosatétaénoïque
C22 :4 ω6	-	Acide cis-7, cis-10, cis-13, cis-16 docosatétaénoïque
C22 :5 ω3	-	Acide cis-7, cis-10, cis-13, cis-16, cis-19 docosapentaénoïque
C22 :5 ω6	-	Acide cis-4, cis-7, cis-10, cis-13, cis-16 docosapentaénoïque
C22 :6 ω3 DHA	Acide cervonique	Acide cis-4, cis-7, cis-10, cis-13, cis-16, cis-19 docosaénoïque
C24 :0	Acide lignocérique	Acide tétracosanoïque
C24 :1 ω9	Acide nervonique	Acide cis-15 tétracosénoïque

INTRODUCTION GENERALE



Les enzymes lipolytiques sont des estérases. Elles catalysent l'hydrolyse et la synthèse des lipides en agissant spécifiquement au niveau des liaisons esters. De nombreuses lipases et carboxylestérases d'origine microbienne, animale ou végétale ont été purifiées et caractérisées d'un point de vue structural et fonctionnel. En ce qui concerne les algues, la mise en évidence d'activités lipolytiques endogènes a été décrite à plusieurs reprises mais à notre connaissance, la littérature ne fournit, à ce jour, aucune donnée sur la caractérisation complète d'enzymes lipolytiques d'origine algale. Par conséquent, les scientifiques ignorent encore tout des propriétés et spécificités d'action des enzymes lipolytiques algales ainsi que des mécanismes de régulation de l'expression des gènes codant ces enzymes. Dans ce contexte scientifique encore peu exploré, le Laboratoire d'Applications des Lipases de l'Université du Maine (LALUM) a décidé d'axer une partie de ses recherches sur la caractérisation des enzymes lipolytiques de la microalgue marine *Isochrysis galbana*.



Les microalgues sont des organismes unicellulaires, eucaryotiques et photosynthétiques dont le profil lipidique est original, complexe et diversifié. Elles se singularisent en particulier par leur composition en acides gras. Certaines espèces présentent des teneurs très élevées en acides gras polyinsaturés à longue chaîne de la famille des oméga 3 comme l'acide docosahexaénoïque (C22 :6 ω 3, DHA) et l'acide eicosapentaénoïque (C20 :5 ω 3, EPA). Cette caractéristique permet donc d'envisager la présence, chez ces microorganismes, d'enzymes lipolytiques offrant des propriétés elles-mêmes originales et diversifiées.

D'autre part, certaines espèces de microalgues évoluent dans des environnements extrêmes en terme, par exemple, de température (eaux froides) et de salinité (eaux marines). Elles ont donc développé des stratégies d'adaptation à ces environnements. En particulier, elles ont la capacité de modifier rapidement la composition en acides gras de leur membrane cellulaire en fonction de la température ambiante. Une étude récente, portant sur la microalgue *Parietochloris incisa*, a suggéré que, pour cela, les triglycérides microalgaux constituaient un réservoir de groupements acyles pouvant être transférés aux phospholipides et glycolipides membranaires en cas de besoin. Ce processus adaptatif fait très certainement intervenir des enzymes lipolytiques et on peut donc envisager que l'expression des gènes codant ces enzymes serait notamment régulée par des facteurs environnementaux.

Les microalgues marines constituent donc un modèle d'étude particulièrement intéressant pour la caractérisation de nouvelles enzymes lipolytiques et l'étude de la régulation de l'expression de leurs gènes.



Ce travail de thèse s'inscrit également dans un contexte économique favorable.

En effet, les microalgues marines font depuis de nombreuses années l'objet de recherches dans un but de valorisation économique de leur biomasse et de diverses molécules à haute valeur ajoutée, comme des antioxydants ou bien encore des pigments et des polysaccharides. Leurs acides gras polyinsaturés oméga 3 bénéficient également d'une attention particulière. En effet, les vertus thérapeutiques et nutritionnelles, maintenant avérées, du DHA et de l'EPA ont considérablement fait croître la demande en oméga 3. Et les microalgues constituent aujourd'hui une source potentielle, complémentaire, et même alternative aux huiles des sous-produits de la pêche pour la production de ces oméga 3. Les microalgues présentent, pour cela, certains avantages car leur culture et leur composition en acides gras sont contrôlables *in vitro*.

La caractérisation de nouvelles enzymes lipolytiques ajouterait donc à l'intérêt économique déjà porté aux microalgues car ces enzymes, et principalement les lipases, constituent un outil biotechnologique important déjà utilisé pour la transformation ou la synthèse des lipides dans de nombreux procédés pharmaceutiques, agrochimiques ou agroalimentaires.



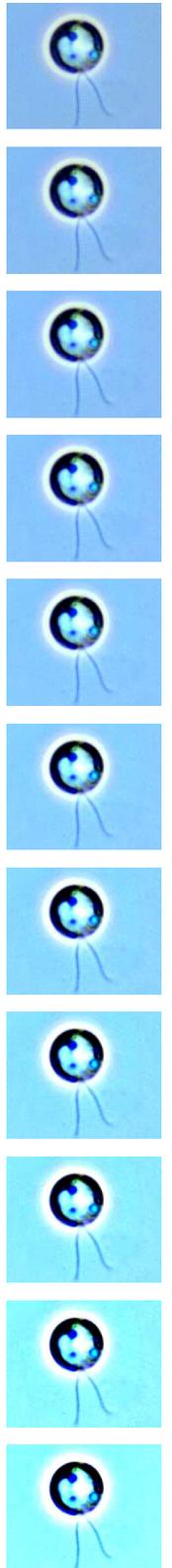
Une approche moléculaire a été adoptée dans le but de purifier et de caractériser les enzymes lipolytiques d'*Isochrysis galbana*. Cette approche, qui consiste à cloner et surexprimer les gènes codant les enzymes d'intérêt, nécessite, pour commencer, l'isolement de ces gènes. Ce manuscrit présente donc le travail qui a été accompli pour l'isolement, l'analyse et le clonage de 3 gènes codant une lipase et deux carboxylestérases hypothétiques chez la microalgue *Isochrysis galbana*.

La première partie de ce mémoire consiste en une synthèse bibliographique divisée en trois chapitres. Le premier porte sur la structure, le mécanisme d'action et la classification des lipases et des carboxylestérases. Le second chapitre concerne les microalgues marines en général et le modèle d'étude, choisi par le laboratoire, la microalgue *Isochrysis galbana*. Dans

le troisième chapitre, seront décrites les applications industrielles des enzymes lipolytiques et la valorisation des lipides et des acides gras de microalgues.

Après une seconde partie consacrée aux moyens, les résultats obtenus seront exposés dans une troisième partie divisée en 7 chapitres. Pour commencer, la mise en évidence et la localisation d'activités lipolytiques endogènes chez *Isochrysis galbana* seront décrites. Dans le second chapitre, seront présentées les deux stratégies successivement développées pour l'isolement des gènes, d'une part la construction et le criblage d'une banque d'ADNc par des sondes hétérologues, d'autre part l'exploitation de la banque EST d'*Isochrysis galbana* pour la recherche d'homologues de lipases et d'estérases et l'obtention de 3 ADNc pleine longueur par RACE et Nested PCR. Les 3 chapitres suivants seront consacrés aux analyses bioinformatiques des 3 séquences préalablement isolées. Dans le sixième chapitre seront décrits le clonage des séquences codantes dans un vecteur d'expression de levure, l'induction de l'expression des gènes et le résultat des premières analyses effectuées en vue de détecter l'activité lipolytique des enzymes recombinantes. Le chapitre sept clôture cette partie par une discussion générale.

SYNTHESE BIBLIOGRAPHIQUE



1. Lipases et carboxylestérases : enzymes lipolytiques

1.1. Définition

Les lipases et les carboxylestérases sont, au même titre que les phospholipases, des enzymes lipolytiques, c'est-à-dire des estérases qui catalysent l'hydrolyse et la synthèse de lipides. Les lipases et les carboxylestérases sont présentes chez tous les organismes qu'ils soient procaryotes ou eucaryotes. Impliquées dans le métabolisme des lipides, elles jouent un rôle essentiel dans de nombreux processus vitaux en participant à la libération et la mobilisation des acides gras.

1.2. Les lipides

1.2.1. Les acides gras

Les acides gras entrent dans la composition des lipides. Ce sont des acides carboxyliques (-COOH) à longue chaîne aliphatique (R) hydrocarbonée constituée d'au moins 4 atomes de carbone. Les acides gras sont amphiphiles, peu solubles dans l'eau mais solubles dans les solvants organiques. Leur solubilité dans l'eau est d'autant plus faible que la chaîne est longue et les doubles liaisons nombreuses. On distingue les acides gras saturés, sans double liaison, des acides gras insaturés, avec une (ou plusieurs) double(s) liaison(s), qui procure(nt) de la rigidité à la molécule. Les acides gras insaturés se classent en série selon la place de la première double liaison à partir de l'atome de carbone de l'extrémité méthyle (carbone ω). Par exemple, la première double liaison des acides gras de la série des oméga 3 (ω 3) est située à trois atomes de carbone de l'extrémité méthyle.

Sur la Figure 1 sont schématisés quelques acides gras, parmi lesquels figurent l'acide linoléique (C18 :2 ω 6) et l'acide α -linoléique (C18 :3 ω 3). Ces deux acides gras ne peuvent être synthétisés par l'homme. Ils doivent impérativement être apportés par l'alimentation et sont donc qualifiés d'acides gras indispensables. L'acide palmitique (C16 :0) est un des acides gras saturés les plus répandus. L'acide eicosapentaénoïque (C20 :5 ω 3, EPA) et l'acide docosahexaénoïque (C22 :6 ω 3, DHA) sont des acides gras polyinsaturés à longue chaîne (AGPI-LC) de la série des oméga 3.

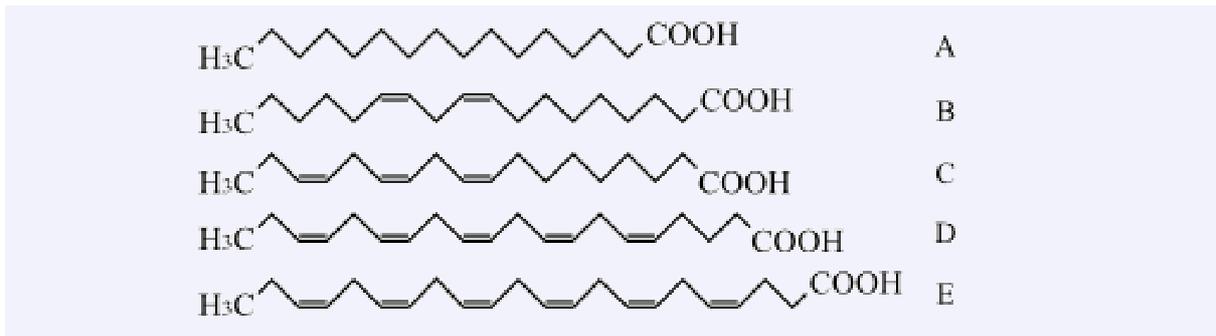


Figure 1 : Structure simplifiée de quelques acides gras.

A) Acide palmitique (C16 :0). B) Acide linoléique (C18 :2 ω6). C) Acide α-linolénique (C18 :3 ω3). D) Acide eicosapentaénoïque (C20 :5 ω3, EPA). E) Acide docosahexaénoïque (C22 :6 ω3, DHA).

1.2.2. Triglycérides et phospholipides

Ces lipides sont formés d'alcools et d'acides gras. Ils sont peu ou pas solubles dans l'eau et solubles dans les solvants organiques.

Les triglycérides sont des esters d'acides gras et de glycérol (Figure 2). Les triglycérides sont neutres et très hydrophobes car les polarités des groupements hydroxyle du glycérol et carboxyle des acides gras s'annulent dans les liaisons esters. Les triglycérides sont notamment impliqués dans le stockage intracellulaire et le transport des acides gras.

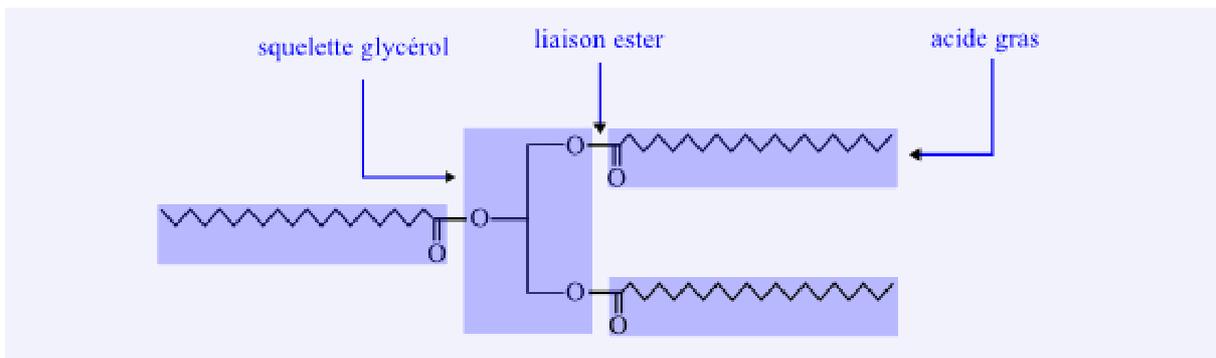


Figure 2 : Structure simplifiée des triglycérides.

La structure des phospholipides est présentée Figure 3A. Sur le squelette du glycérol, 2 acides gras sont estérifiés et la troisième position est occupée par un groupement phosphate lui-même lié à un alcool. Dans le cas des phosphatidylcholines, cet alcool est la choline (Figure 3B). Les phospholipides sont des molécules amphiphiles.

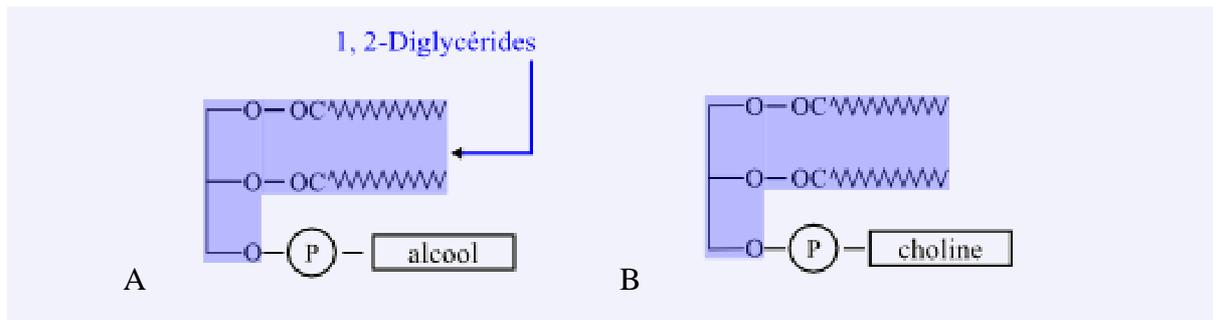


Figure 3 : Structure simplifiée des phospholipides.

A) Structure générale. B) Exemple de phospholipide : la phosphatidylcholine.

1.3. Réactions catalysées

Les lipases et les carboxylestérases sont des estérases. Elles se classent dans la superfamille des hydrolases à repliement α/β (EC 3) (Ollis *et al.*, 1992), qui rassemble un grand nombre d'enzymes lytiques actives sur des substrats très différents : acylhydrolases, protéases, peroxydases.... Lipases et carboxylestérases sont des acylhydrolases qui agissent spécifiquement au niveau des liaisons esters des lipides (Figure 4). Les lipases sont des hydrolases d'esters d'acylglycérols (EC 3.1.1.3). Elles sont actives sur les lipides à longues chaînes ($\geq C10$). Elles catalysent préférentiellement l'hydrolyse de triglycérides en diglycérides, monoglycérides, acides gras libres et glycérol (Jensen *et al.*, 1990). Cette spécificité de substrat les distingue des carboxylestérases (EC 3.1.1.1) qui, elles, sont actives sur les lipides à plus courtes chaînes.

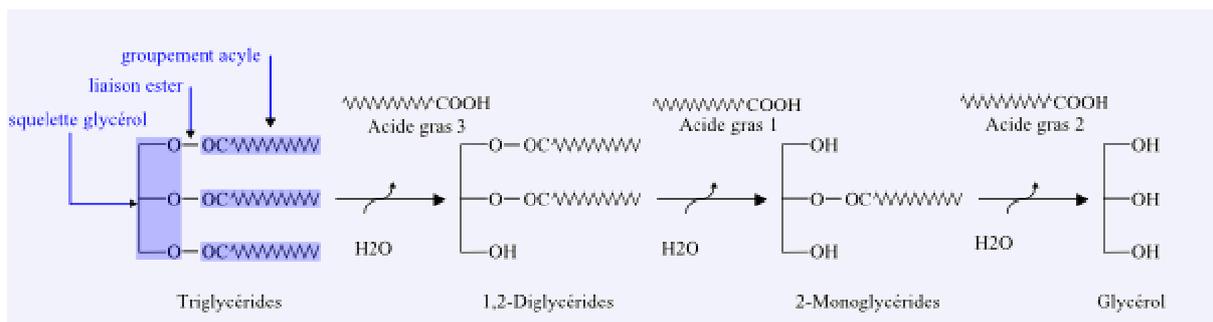


Figure 4 : Hydrolyse des triglycérides.

Les produits du catabolisme sont les acides gras et le glycérol.

Les réactions catalysées par les lipases et les carboxylestérases sur les substrats lipidiques sont l'hydrolyse, l'estérification, les transestérifications (acidolyse et alcoololyse) et l'interestérification (Figure 5). Les lipases peuvent donc être qualifiées d'acyltransférases

puisque les acides gras sont soit libérés soit transférés à un composé disposant d'un groupe hydroxyle libre.

Hydrolyse	$R_1\text{-COOR}_2 + \text{H}_2\text{O}$	\rightarrow	$R_1\text{-COOH} + R_2\text{-OH}$
Estérification	$R_1\text{-OH} + R_2\text{-COOH}$	\rightarrow	$\text{H}_2\text{O} + R_2\text{-COOR}_1$
Acidolyse	$R_1\text{-COOH} + R_2\text{-COOR}_3$	\rightarrow	$R_2\text{-COOH} + R_1\text{-COOR}_3$
Alcoololyse	$R_1\text{-OH} + R_2\text{-COOR}_3$	\rightarrow	$R_3\text{-OH} + R_2\text{-COOR}_1$
Interestérification	$R_1\text{-COOR}_2 + R_3\text{-COOR}_4$	\rightarrow	$R_1\text{-COOR}_4 + R_3\text{-COOR}_2$

Figure 5 : Diversité des réactions catalysées par les estérases.

1.4. Exemples de fonctions biologiques

Les enzymes lipolytiques sont impliquées dans la lyse et la mobilisation des réserves lipidiques. Une fois libérés, les acides gras constituent des substrats énergétiques, leur oxydation étant l'une des voies essentielles de production d'énergie dans les cellules. Les acides gras sont aussi les précurseurs de nombreux métabolites secondaires.

Chez les mammifères, par exemple, les lipases gastriques et pancréatiques participent à la digestion et l'absorption des triglycérides apportés par l'alimentation (Mu et Porsgaard, 2005). Certaines enzymes lipolytiques végétales interviennent dans des voies de signalisation cellulaire. Les phospholipases participent à la synthèse de composés impliqués dans les systèmes de défense chimique des plantes (Chandra *et al.*, 1996; Narvaez-Vasquez *et al.*, 1999). Chez les microorganismes pathogènes, les enzymes lipolytiques extracellulaires, par exemple, contribuent à l'infection cellulaire en hydrolysant les membranes plasmiques (Schaller *et al.*, 2005).

1.5. Histoire moléculaire évolutive

Les estérases sont des macromolécules de nature protéique donc codées par des gènes. Les lipases et les carboxylestérases partagent peu de similarités de séquences et n'agissent pas sur les mêmes substrats. Par contre, elles sont structurellement analogues et affichent un

repliement de type α/β . L'agencement du site catalytique est également strictement préservé (Ollis *et al.*, 1992; Nardini et Dijkstra, 1999).

Ces observations soutiennent l'hypothèse selon laquelle lipases et carboxylestérases seraient phylogénétiquement liées à la superfamille multigénique des hydrolases à repliement α/β (Myers *et al.*, 1988; Cygler *et al.*, 1993). Les enzymes de cette superfamille diffèrent par leurs séquences primaires d'acides aminés alors qu'elles possèdent une même conformation tridimensionnelle. Elles dérivent probablement d'une protéine ancestrale commune de conformation stable qui aurait évolué selon un processus d'adaptation moléculaire et d'optimisation fonctionnelle. En d'autres termes, ces enzymes seraient nées de la duplication puis de la divergence d'un gène ancestral. Ce processus évolutif leur aurait assuré l'acquisition de nouvelles fonctions grâce à des substitutions d'acides aminés, des insertions ou des délétions de boucles. Cette théorie expliquerait l'existence d'enzymes de même configuration spatiale, de mêmes fonctions catalytiques, avec des spécificités de substrat diverses et peu d'homologies de séquences (Holmquist, 2000; Wong et Schotz, 2002).

1.6. Caractéristiques structurales

1.6.1. Le repliement α/β

D'après les données biochimiques et cristallographiques, la structure des lipases et des carboxylestérases est identique à celle des hydrolases à repliement α/β . La Figure 6 illustre la structure type de ces hydrolases qui consiste en un noyau formé de feuilletts β parallèles hydrophobes entourés par des hélices α amphiphiles.

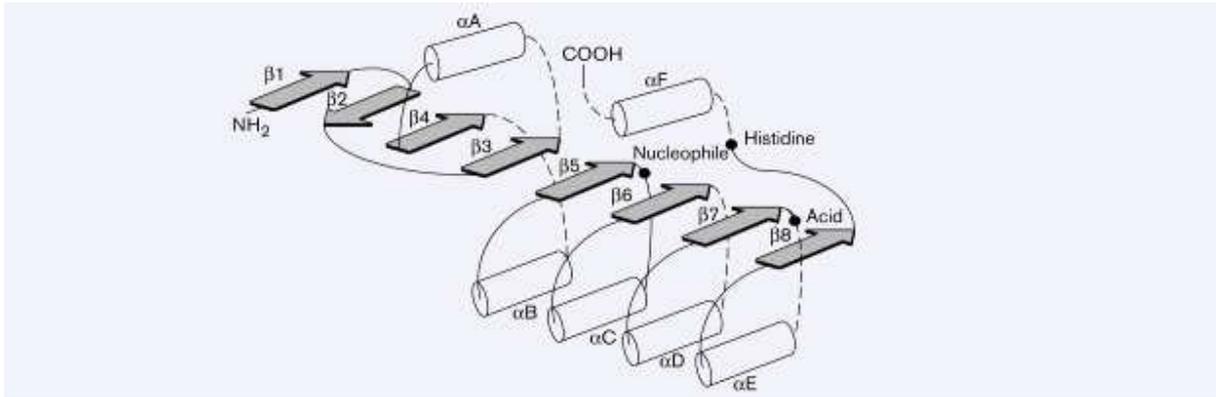


Figure 6 : Représentation schématique de la structure secondaire d'une hydrolase à repliement α/β (Ollis *et al.*, 1992).

Les feuillets β sont représentés par des flèches, les hélices α par des cylindres et les traits correspondent aux boucles.

Dans la Protein Data Bank (PDB), qui recense les protéines dont les structures tridimensionnelles ont été élucidées, les lipases et les carboxylestérases comptabilisent respectivement 67 et 28 entrées. Il s'agit principalement d'enzymes d'origine microbienne ou animale. Lipases et carboxylestérases sont formées d'une ou plusieurs chaînes polypeptidiques : elles peuvent être monomériques, dimériques ou tétramériques. La Figure 7 illustre la structure tridimensionnelle récemment résolue de la carboxylestérase du kiwi (Ileperuma *et al.*, 2007).

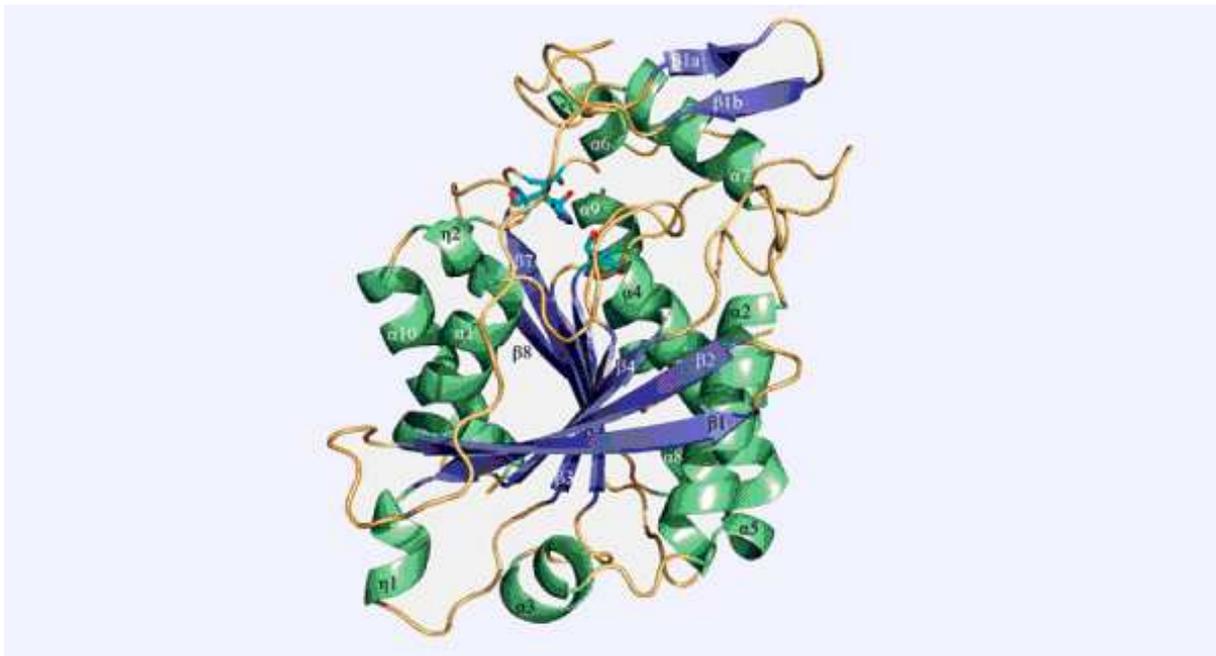


Figure 7 : Structure tridimensionnelle de la carboxylestérase monomérique du kiwi *Actinidia eriantha* (PDB 2o7r) (Ileperuma *et al.*, 2007).

Les feuillets β sont représentés par des flèches bleues et les hélices α par des formes hélicoïdales de couleur verte.

1.6.2. La triade catalytique

Le site catalytique apparaît toujours sous la forme d'une triade hautement conservée : sérine / acide aspartique / histidine (Ser/Asp/His ou S/D/H). Pour chaque sous-famille d'hydrolases, l'ordre d'apparition de ces résidus aminés dans la séquence peptidique est fixe : le résidu nucléophile, *i.e.* la sérine, précède les deux autres résidus et se positionne après le feuillet $\beta 5$. Le résidu acide, l'acide aspartique, apparaît très souvent après le feuillet $\beta 7$ et enfin, l'histidine se place toujours après le dernier feuillet (Nardini et Dijkstra, 1999). L'acide aspartique est parfois remplacé par un acide glutamique. C'est le cas pour les lipases de *Geotrichum candidum* (Schrag *et al.*, 1991) et de *Candida rugosa* (Grochulski *et al.*, 1993; Mancheno *et al.*, 2003).

L'emplacement des résidus de la triade catalytique est précisé sur la Figure 6 par des ronds noirs, avec, dans l'ordre, de l'extrémité amine à l'extrémité carboxylique, le résidu nucléophile (sérine), le résidu acide (acide aspartique ou glutamique) et le résidu basique, l'histidine.

1.6.3. Le pentapeptide G-X-S-X-G

La sérine catalytique se situe au centre d'un pentapeptide consensus Gly-X-Ser-X-Gly, où les résidus X correspondent à des acides aminés quelconques (Ollis *et al.*, 1992). Ce pentapeptide, toujours intercalé entre un feuillet β et une hélice α , se situe au niveau d'une jonction anguleuse formant un « coude nucléophile » très réactif, situé à peu près au milieu de la chaîne peptidique. La sérine s'y trouve dans une conformation énergétiquement défavorable (ϵ) lui permettant de faire pointer sa chaîne latérale vers l'extérieur. On parle de conformation β - ϵ Ser- α (Derewenda et Derewenda, 1991; Ollis *et al.*, 1992; Cygler *et al.*, 1993). Cet emplacement fait de la sérine un résidu accessible au substrat et aux molécules d'eau.

La région autour du pentapeptide présente une relative conservation. D'après la base de données PROSITE qui répertorie les motifs fonctionnels significatifs des protéines, la sérine catalytique (en gras souligné) des carboxylestérases est incluse dans la séquence consensus F-[GR]-G-X(4)-[LIVM]-X-[LIV]-X-G-X-**S**-[STAG]-G. Le même résidu chez les lipases figure au sein de la séquence consensus suivante : [LIV]-X-[LIVFY]-[LIVMST]-G-[HYWV]-**S**-X-G-[GSTAC].

1.6.4. Le volet amphiphile et l'activation interfaciale des lipases

Les lipases sont solubles en milieu aqueux mais exercent leur activité hydrolytique essentiellement sur des substrats peu ou non solubles dans l'eau. Cette propriété les différencie des carboxylestérases qui sont actives sur les substrats solubles. L'activité des lipases est beaucoup plus importante lorsque les substrats sont sous forme d'émulsion, de liposomes ou de micelles et non pas de monomères solubles. Elles agissent alors à l'interface eau/lipides et leur activité est liée à la présence de cette interface. On parle de phénomène d'activation interfaciale (Sarda et Desnuelle, 1958).

Contrairement aux carboxylestérases, le site actif des lipases peut être recouvert par un domaine mobile amphiphile (appelé aussi volet ou « lid »). Lorsque l'enzyme est au contact d'une interface eau/lipides, ce volet s'écarterait permettant la fixation du substrat et son hydrolyse (Brzozowski *et al.*, 1991; Brzozowski *et al.*, 2000). En absence d'interface, une très faible activité est observée. On suppose que le volet cacherait alors le site actif (Figure 8).

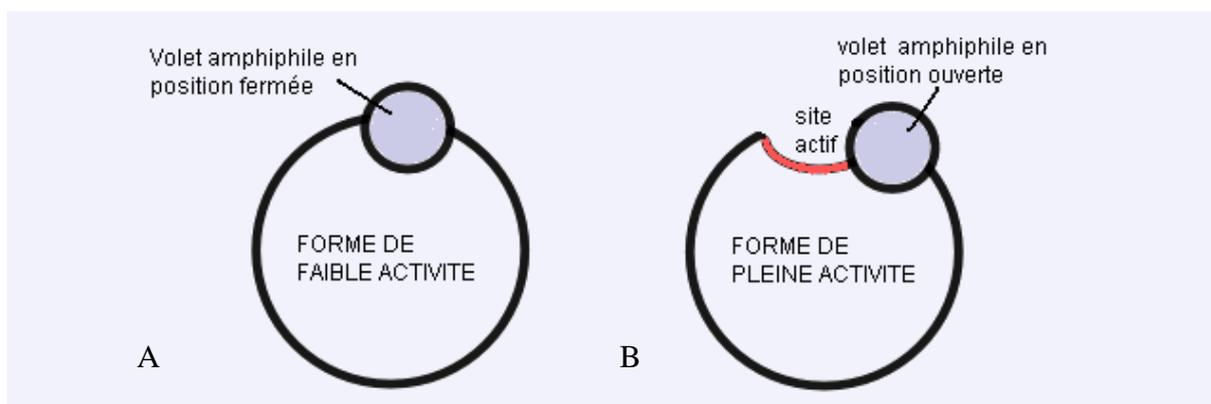


Figure 8 : Représentation schématique des deux conformations d'une lipase.

A) Forme de faible activité où le volet amphiphile serait en position fermée. B) Forme de pleine activité où le volet serait ouvert.

La présence d'un volet amphiphile a été démontrée par les études cristallographiques de plusieurs lipases (Figure 9) (Brady *et al.*, 1990; Winkler *et al.*, 1990; Derewenda *et al.*, 1992a; Derewenda *et al.*, 1992b). Cependant, certaines lipases ne suivent pas strictement cette règle. Par exemple, la lipase B de *Candida antarctica* possède en guise de volet une courte hélice très mobile et ne montre pas de phénomène d'activation interfaciale (Uppenberg *et al.*, 1994a; Uppenberg *et al.*, 1994b; Martinelle *et al.*, 1995). Le site catalytique de la cutinase de *Fusarium solani* est directement exposé aux solvants et l'enzyme ne présente pas de

phénomène d'activation interfaciale mais hydrolyse aussi bien des substrats peu solubles que solubles (Longhi et Cambillau, 1999).

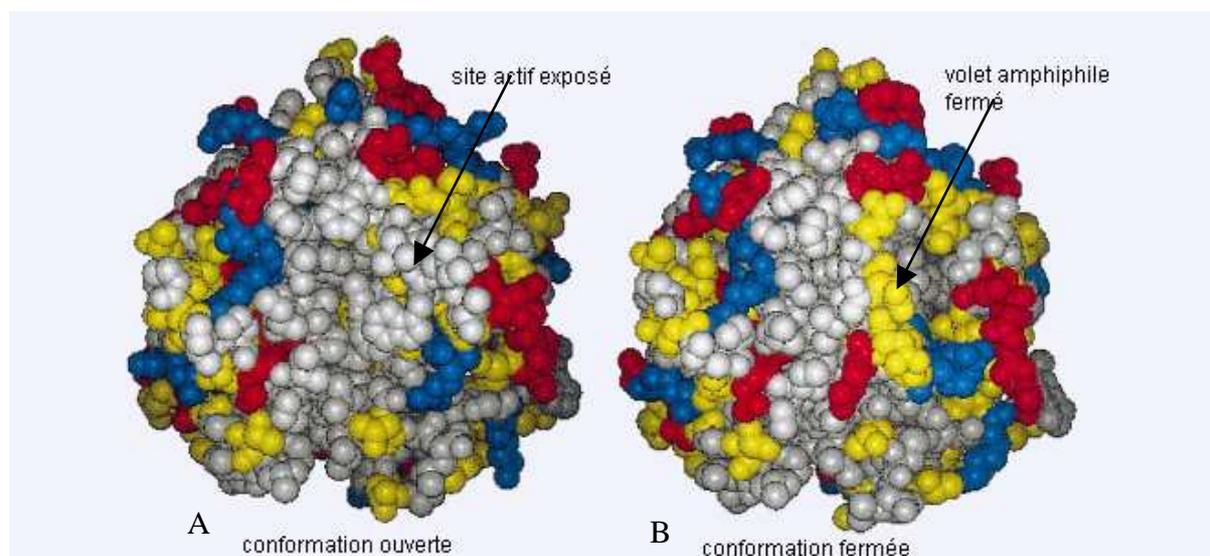


Figure 9 : Représentation moléculaire des deux conformations de la lipase de *Thermomyces lanuginosa* d'après Svendsen, 2000.

A) Conformation ouverte B) Conformation fermée. Les résidus hydrophobes sont colorés en blanc, les résidus hydrophiles en jaune, les résidus chargés positivement en bleu, ceux chargés négativement sont en rouge.

La composition en acides aminés du volet amphiphile est peu conservée parmi les lipases et même entre les isoformes d'une même espèce. Par exemple, les 5 isozymes de lipase identifiées chez *Candida rugosa* montrent des divergences de séquence au niveau du volet amphiphile (Tableau 1) (Lotti *et al.*, 1994). Ces variations seraient par ailleurs responsables des différentes activités et spécificités de ces isoformes (Akoh *et al.*, 2004).

Région/fonction	Position du résidu	Lipase 1	Lipase 2	Lipase 3	Lipase 4	Lipase 5
Volet amphiphile	66	E	M	E	L	E
	71	E	D	E	S	E
	76	A	N	T	A	V
	78	L	R	L	I	V
	80	L	L	L	S	L
	91	S	L	L	L	L

Tableau 1 : Comparaison entre les régions du volet amphiphile des isozymes de lipase du champignon *Candida rugosa* (Lotti *et al.*, 1994).

1.7. Mécanisme d'action

Le mécanisme d'action des lipases et des carboxylestérases est identique à celui des protéases à sérine qui a été étudié dès les années 60 sur différentes protéases telle que la chymotrypsine (Hubbard et Shoupe, 1977). Le mécanisme de catalyse fait appel à une sérine particulièrement réactive, située au sein du site actif. La sérine est activée par un relais de charge formé en amont par l'histidine et l'acide aspartique (ou l'acide glutamique) de la triade catalytique. La Figure 10 montre les différentes étapes de l'hydrolyse d'un ester par une enzyme lipolytique.

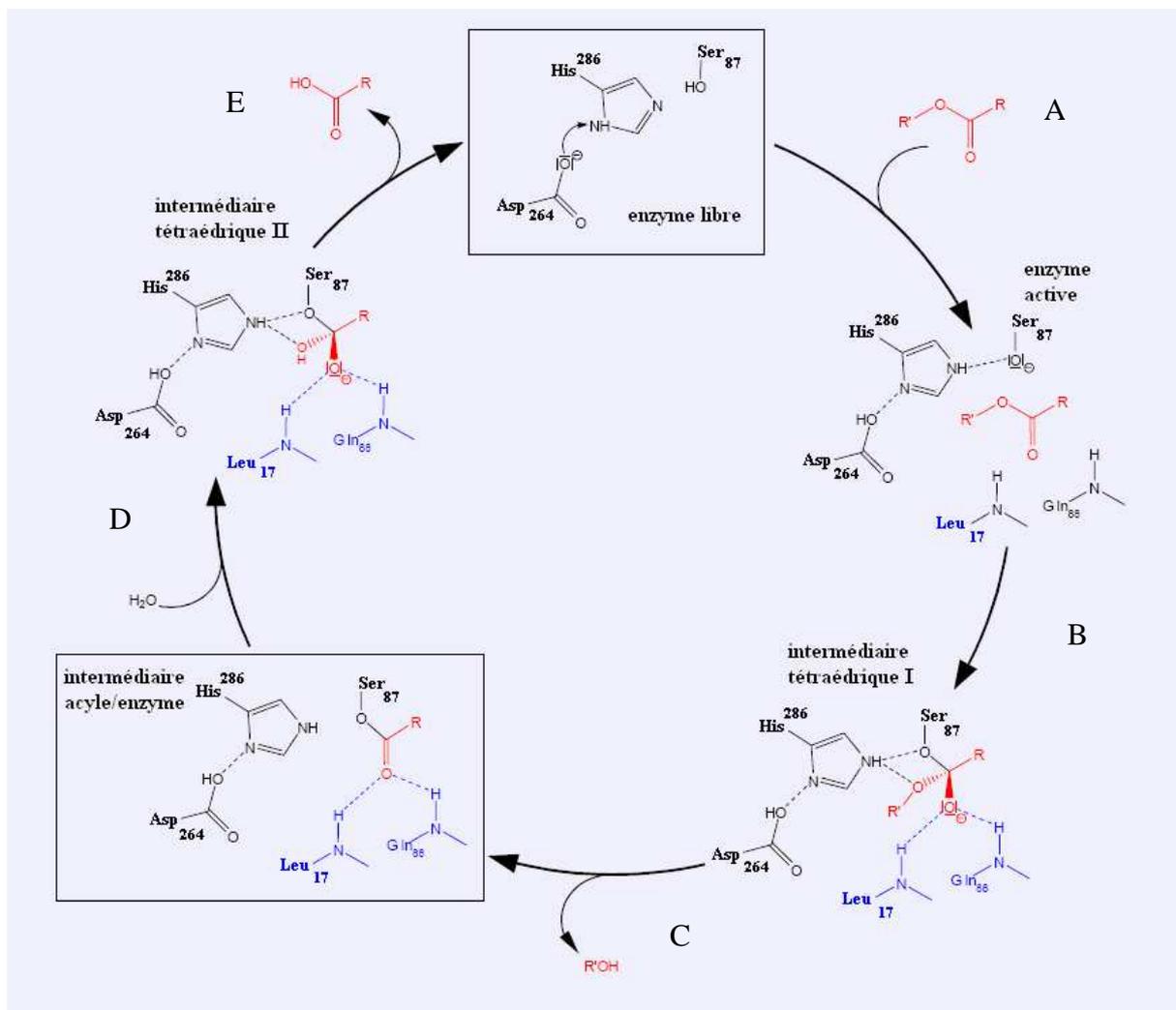


Figure 10 : Schématisation de l'hydrolyse d'un ester par une enzyme lipolytique d'après Fischer, 2004.

La position des résidus formant la triade catalytique correspond à celle qu'ils occupent au sein de la lipase de *Burkholderia cepacia*. Le substrat figure en rouge. Les acides aminés impliqués dans la composition du puits oxyanionique sont en bleu. Les acides aminés catalytiques sont en noir.

La première étape de la catalyse est l'approche de l'ester qui va subir une attaque nucléophile du résidu sérine (Figure 10A). Il se forme ensuite un intermédiaire tétraédrique portant une charge négative sur l'oxygène (Figure 10B). Cet intermédiaire, désigné également sous le terme d'oxyanion, très instable, est stabilisé au sein d'un puits oxyanionique (poche hydrophobe) par des ponts hydrogènes. Par retour de la charge portée par l'oxyanion, le groupement alcool est libéré (Figure 10C). Le groupement acyle resté lié à l'enzyme, formant l'intermédiaire acyle/enzyme, subit ensuite l'attaque de molécules d'eau (Figure 10D). Il y a formation d'un deuxième complexe tétraédrique qui se dissocie ensuite en libérant le produit carboxylique d'une part et l'enzyme sous sa forme native d'autre part (Figure 10E).

1.8. Spécificité de substrat

Les enzymes lipolytiques présentent une conformation et un modèle d'action quasi identiques, par contre elles affichent une grande diversité de spécificité de substrat. Les lipases, par exemple, montrent globalement une préférence pour les triglycérides à longues chaînes, tandis que les carboxylestérases sont définies comme agissant essentiellement sur des substrats à courtes chaînes. La spécificité de la réaction provient de différences subtiles, au niveau de la poche hydrophobe, qui permettent de positionner certains substrats et d'en exclure d'autres. Seuls quelques résidus formant cette poche déterminent la spécificité de l'interaction de l'enzyme avec son substrat.

Joerger et son équipe ont ainsi réussi à transformer la spécificité de substrat de la lipase du champignon *Rhizomucor delemar*, en modifiant, par mutagenèse dirigée, les résidus F95 et V209, positionnés respectivement à l'extrémité et au centre de la poche hydrophobe (Joerger et Haas, 1994). La mutation F95D accroît l'activité de la lipase sur la tricapryline. Le mutant V209T voit son activité diminuer à l'égard de la trioléine, tandis que le mutant V209W présente une activité quatre fois plus élevée pour la tributyrine. La substitution d'un seul acide aminé spécifique suffit pour altérer la spécificité de l'enzyme vis-à-vis de la longueur de la chaîne : le mutant A85F de la cutinase du champignon *Fusarium solani pisi* est plus actif pour les chaînes de longueur supérieure à C7 et moins actif pour les chaînes inférieures à C7 par rapport au type sauvage (Mannesse *et al.*, 1995; Svendsen, 2000). Autre exemple, la mutation du résidu D290A de la lipase extracellulaire de *Staphylococcus simulans* s'accompagne aussi d'un changement significatif de la spécificité de la longueur de chaîne des groupements acyles (Sayari *et al.*, 2007).

Les travaux de Alam, publiés en 2006, portent sur une lipase hépatothique humaine. L'auteur a étudié les incidences de la modification et de la suppression du domaine de type FLXLXXXn (où n est un résidu non polaire et X un acide aminé quelconque excepté la proline) (Alam *et al.*, 2006). La fonction de ce domaine n'a pas été clairement établie. Toutefois, il serait impliqué dans les interactions de l'enzyme avec son substrat (Au-Young et Fielding, 1992). Il a été identifié à plusieurs reprises chez les mammifères (Lehner et Verger, 1997). Chez cette lipase hépatothique humaine, ce domaine correspond à l'octapeptide FLDLIADV et se situe dans l'hélice α adjacente aux résidus catalytiques. La suppression de cette hélice α annule l'activité lipolytique de l'enzyme. Le remplacement du résidu phénylalanine 417 par une alanine entraîne une réduction de l'activité de 40 % de l'enzyme à l'égard des esters solubles et insolubles (Alam *et al.*, 2006).

1.9. Classification des enzymes lipolytiques

Les enzymes sont généralement classées en fonction de leurs propriétés biochimiques (spécificité de substrat, stéréosélectivité, pH optimal, température optimale, stabilité dans les solvants). L'inconvénient de cette approche est qu'elle ne permet pas de tenir compte de toutes les protéines figurant dans les bases de données, en raison du temps que requiert la validation expérimentale. Les enzymes peuvent aussi être classées d'après les homologies de séquences établies par des alignements et des analyses phylogénétiques. L'intérêt de cette seconde approche est qu'elle peut être appliquée à l'ensemble des données.

La classification des enzymes lipolytiques a toujours été complexe en raison des particularités de ces enzymes. En effet, les estérases se singularisent par leur diversité fonctionnelle, par l'étendue de leur spécificité de substrat et leur faible homologie de séquence.

Les lipases, par exemple, partagent quatre signes distinctifs : une triade de résidus formant le site actif, un mécanisme d'activation interfaciale, un volet amphiphile et la capacité de catalyser l'hydrolyse des acylglycérols à longues chaînes. Mais toutes les lipases ne possèdent pas ces quatre caractéristiques simultanément. De surcroît, aucune de ces caractéristiques ne constitue à elle seule un critère suffisant pour classer une enzyme dans la famille des lipases. D'autres enzymes, comme les protéases à sérine (Kraut, 1977; Craik *et al.*, 1987) et les cutinases (Martinez *et al.*, 1994; Longhi et Cambillau, 1999), possèdent également la triade catalytique S/D/H. La cutinase de *Fusarium solani* peut hydrolyser des triacylglycérols à courtes et longues chaînes (Mannesse *et al.*, 1995).

En 1998, Schmid et Verger suggèrent qu'il est plus approprié de classer une enzyme parmi les lipases selon qu'elle hydrolyse ou non des acylglycérols à longues chaînes (Schmid et Verger, 1998). La lipase de *Bacillus subtilis* est une petite enzyme d'environ 19,3 kDa, qui ne possède pas de volet amphiphile et ne montre pas d'activation interfaciale aux interfaces eau/lipides. Cette enzyme exerce un maximum d'activité hydrolytique sur les esters de glycérol et d'acides gras à chaîne de longueur moyenne (C8). Néanmoins cette enzyme se classe au sein de la famille des lipases car elle catalyse aussi l'hydrolyse d'esters de glycérol et d'acides gras à longues chaînes (van Pouderoyen *et al.*, 2001).

Toutes ces exceptions rendent la classification des lipases délicate et aujourd'hui, un certain nombre de classifications repose essentiellement sur les homologues de séquences. En 1999, Arpigny et Jaeger ont réparti en huit familles les enzymes lipolytiques bactériennes sur la base d'homologies de séquence (Tableau 2) (Arpigny et Jaeger, 1999).

Familles	Caractéristiques
I	Vraies lipases*
II	Lipases et estérases à motif G-D-S-L
III	Similaires (20%) aux isoformes plasmique et intracellulaire de la PAF-AH humaine
IV	Lipases hormonosensibles
V	Similaires (20-25 %) aux déhalogénases, aux époxyde hydrolases
VI	Petites estérases 23-26 kDa
VII	Grandes estérases 55 kDa
VIII	Homologues aux β -lactamases de classe C

Tableau 2 : Huit familles d'enzymes lipolytiques bactériennes d'après Arpigny et Jaeger, 1999.

*On définit par « vraies lipases » celles qui rassemblent toutes les caractéristiques (triade catalytique, activation interfaciale, volet et acylglycérols à longues chaînes comme substrat de prédilection).

Plus récemment, la Lipase Engineering Database (LED) a été conçue en vue de faciliter l'étude des relations séquence / structure / fonction des enzymes de la superfamille des hydrolases à repliement α/β (Pleiss *et al.*, 2000). La LED est une base de données où les enzymes lipolytiques sont classées en différentes familles. Sa conception repose sur la base d'homologies de séquences établies à partir d'alignements multiples globaux. La LED constitue un excellent instrument d'analyse des relations existant entre les séquences, structures et fonctions des lipases, carboxylestérases et protéines proches.

En 2004, la LED rassemblait 3148 séquences correspondant à 2313 protéines dont 35 % de protéines hypothétiques et moins de 5 % de protéines structurellement caractérisées (Fisher, 2004). Aujourd'hui la LED dénombre 6138 séquences pour 4322 protéines. Les séquences sont agencées en trois niveaux hiérarchiques qui sont : la séquence elle-même, la famille

d'homologues et la superfamille. Toutes les séquences montrant un haut degré de similitude ont été classées en 103 familles d'homologues. Les familles d'homologues présentant, entre elles, un faible mais significatif degré de similitude ont été regroupées en 37 superfamilles (Tableau 3). Les superfamilles ne présentent pas de similarité de séquences significatives entre elles (Fischer et Pleiss, 2003). Les alignements multiples sont disponibles et les résidus impliqués fonctionnellement sont annotés.

Klasse	Superfamille	Homologue Famille
GGGX	abH1 - Carboxylesterases	abH1.1 (Caenorhabditis elegans esterases I) abH1.2 (Mammalian carboxylesterases) abH1.3 (Mammalian bile salt activated lipase like) abH1.4 (Acetylcholinesterases) abH1.5 (Bacillus esterases) abH1.6 (Alpha esterases) abH1.7 (Juvenile hormone esterases) abH1.8 (Drosophila glutactin like) abH1.9 (Drosophila esterases) abH1.10 (Miscellaneous) abH1.11 (Caenorhabditis elegans esterases II) abH1.12 (Thyroglobulin like)
GGGX	abH2 - Yarrowia lipolytica lipase like	abH2.1 (Yarrowia lipolytica lipase like)
GGGX	abH3 - Candida rugosa lipase like	abH3.1 (Candida rugosa lipase like)
GGGX	abH4 - Moraxella lipase 2 like	abH4.1 (Moraxella lipase 2 like) abH4.2 (Acinetobacter esterases) abH4.3 (N-deacetylases) abH4.4 (Bacillus sphaericus lipase like)
GGGX	abH5 - Hormone sensitive lipases	abH5.1 (Hormone sensitive lipases)
GGGX	abH6 - Brefeldin A esterase like	abH6.1 (Brefeldin A esterase like)
GX	abH7 - Moraxella lipase 3 like	abH7.1 (Haemophilus influenzae lipase like) abH7.2 (Moraxella lipase 3 like)
GX	abH8 - Cytosolic Hydrolases	abH8.1 (soluble bacterial epoxide hydrolases I) abH8.2 (soluble bacterial epoxide hydrolases II) abH8.3 (soluble mammalian epoxide hydrolases) abH8.4 (soluble plant epoxide hydrolases) abH8.5 (soluble haloalkane dehalogenases) abH8.6 (miscellaneous) abH8.7 (soluble epoxide hydrolases (beta6)) abH8.8 (soluble meta cleavage compound hydrolases II) abH8.9 (soluble non-heme peroxidases) abH8.10 (soluble haloalkane dehalogenases (beta6)) abH8.11 (soluble meta cleavage compound hydrolases I) abH8.12 (Enol-lactone hydrolases) abH8.13 (soluble esterases / lipases / peptidases) abH8.14 (Ccg1/TafII250-interacting factor B like)
GX	abH9 - Microsomal Hydrolases	abH9.1 (microsomal epoxide hydrolases) abH9.2 (BioH protein like) abH9.3 (Proline iminopeptidases)
GX	abH10 - Uncultured crenarchaeote	abH10.1 (uncultured crenarchaeote)
GX	abH11 - Carboxylesterases	abH11.1 (Bacillus carboxylesterases) abH11.2 (Mycoplasma lipases) abH11.3 (Thermotoga maritima esterase like)
GX	abH12 - Hydroxynitrile lyases	abH12.1 (Hydroxynitrile lyases) abH12.2 (Arabidopsis thaliana esterases II) abH12.3 (Arabidopsis thaliana esterases I)
GX	abH13 - Bacterial esterases	abH13.1 (Bacterial esterase)
GX	abH14 - Gastric lipases	abH14.1 (Lysosomal acid lipases) abH14.2 (Gastric lipases)
GX	abH15 - Burkholderia lipases	abH15.1 (Staphylococcus aureus lipase like) abH15.2 (Burkholderia cepacia lipase like)

Tableau 3 : Liste des superfamilles et familles d'homologues figurant dans la Lipase Engineering Database (Fischer, 2004).

Nomenclature abHn.m où abH signifie alpha beta hydrolase, suivi par le numéro attribué à la superfamille n, séparé du numéro de la famille d'homologues m par un point.

		abH15.3 (Saccharomyces cerevisiae lipase 2 like)
GX	abH16 - Streptomyces lipases	abH16.1 (Streptomyces lipases)
GX	abH17 - Chloroflexus aurantiacus lipase like	abH17.1 (Chloroflexus aurantiacus lipase like)
GX	abH18 - Bacillus lipases	abH18.1 (Bacillus lipases)
GX	abH19 - Thioesterases	abH19.1 (Palmitoyl-protein thioesterase 1 like) abH19.2 (Palmitoyl-protein thioesterase 2 like)
GX	abH20 - Lipoprotein lipases	abH20.1 (Hepatic lipases) abH20.2 (Lipoprotein lipases) abH20.3 (Pancreatic lipases)
GX	abH21 - Bacterial esterases	abH21.1 (Chlorobium esterases) abH21.2 (Rickettsia conorii esterase like) abH21.3 (Pseudomonas esterases) abH21.4 (Carboxylesterases) abH21.5 (Nostoc sp. esterase like)
GX	abH22 - Lysophospholipase	abH22.1 (Carboxylesterases) abH22.2 (Microbulbifer degradans esterase like) abH22.3 (Lysophospholipase) abH22.4 (Arabidopsis thaliana lysophospholipases) abH22.5 (Arabidopsis thaliana proteins) abH22.6 Trypanosoma brucei lysophospholipase like) abH22.7 (Homo sapiens lysophospholipase like) abH22.8 (Schizosaccharomyces pombe lysophospholipase like) abH22.9 (Saccharomyces cerevisiae proteins)
GX	abH23 - Filamentous fungi lipases	abH23.1 (Rhizomucor mihei lipase like) abH23.2 (Saccharomyces lipase like)
GX	abH24 - Pseudomonas lipases	abH24.1 (Pseudomonas lipases)
GX	abH25 - Moraxella lipase 1 like	abH25.1 (Moraxella lipase 1 like)
GX	abH26 - Deacetylases	abH26.1 (Deacetylases)
Y	abH27 - Dipeptidyl peptidase IV like	abH27.1 (Dipeptidyl peptidase IV like)
Y	abH28 - Prolyl endopeptidases	abH28.1 (Propyl endopeptidases)
Y	abH29 - Dipeptidyl-peptidases	abH29.1 (Dipeptidyl-peptidases)
Y	abH30 - Cocaine esterases	abH30.1 (Cocaine esterases)
GX	abH31 - Dienlactone Hydrolases	abH31.1 (Xanthomonas dienelactone hydrolases) abH31.2 (Carboxymethylenebutenolidases) abH31.3 (Caulobacter dienelactone hydrolases) abH31.4 (Nostoc dienelactone hydrolases) abH31.5 (Salmonella carboxymethylenebutenolidase) abH31.6 (Agrobacterium carboxymethylenebutenolidase) abH31.7 (Rhodococcus dienelactone hydrolases)
GX	abH32 - Xylanase esterases	abH32.1 (Xylanase Z esterase domain) abH32.2 (Xylanase Y esterase domain)
GX	abH33 - Antigen 85	abH33.1 (Antigen 85-C) abH33.2 (Antigen 85-A) abH33.3 (Antigen 85-B)
GX	abH34 - Lysosomal protective protein like	abH34.1 (Lysosomal protective protein like) abH34.2 (Serine carboxypeptidase II like)
GX	abH35 - Acyl-transferases	abH35.1 (Acyl-transferases)
GX	abH36 - Cutinases	abH36.1 (Colletotrichum cutinases) abH36.2 (Botryotinia cutinases) abH36.3 (Fusarium cutinases) abH36.4 (Mycobacterium cutinases)
GX	abH37 - Candida antarctica lipase like	abH37.1 (Candida antarctica lipase B like)

Tableau 3 (suite)

Les protéines ont également été réparties en trois classes, GX, GGGX et Y, selon la séquence et la structure du puits oxyanionique (Pleiss *et al.*, 2000). La classe GX regroupe 27 superfamilles, 79 familles d'homologues, 3623 séquences et 349 structures. Le résidu X à l'origine du puits oxyanionique est bien conservé et toujours précédé d'une glycine strictement conservée. Cette classe comprend principalement des lipases de bactéries et de champignons, des lipases eucaryotes (lipases hépatiques, pancréatiques, gastriques, des lipoprotéines lipases et des lipases acides lysosomales), des cutinases, des phospholipases et quelques peroxydases. La classe GGGX comprend 6 superfamilles, 20 familles d'homologues, 1780 séquences et 167 structures. Le résidu responsable du puits oxyanionique fait partie d'un motif bien conservé GGGX où X est un acide aminé hydrophobe. La classe GGGX rassemble plutôt des estérases bactériennes, des α -estérases, des carboxylestérases, des lipases activées par les sels biliaires, des lipases hormones dépendantes, des acétylcholinestérases et des thioestérases. La classe Y (peptidases et cocaïne estérases) rassemble 4 superfamilles, 735 séquences et 58 structures où le puits oxyanionique est formé par le groupe hydroxyle de la tyrosine, résidu hautement conservé à travers les 4 superfamilles. La Figure 11 présente ces différentes classes sous forme d'un arbre phylogénétique proposé par Fisher en 2004. Cet arbre ne coïncide pas nécessairement avec les relations évolutives des organismes. En effet, d'une part, les séquences d'une même famille d'homologues appartiennent à des organismes très différents. D'autre part, les séquences de carboxylestérases ou de lipases provenant d'un même organisme sont dispersées à travers les superfamilles et les familles d'homologues.

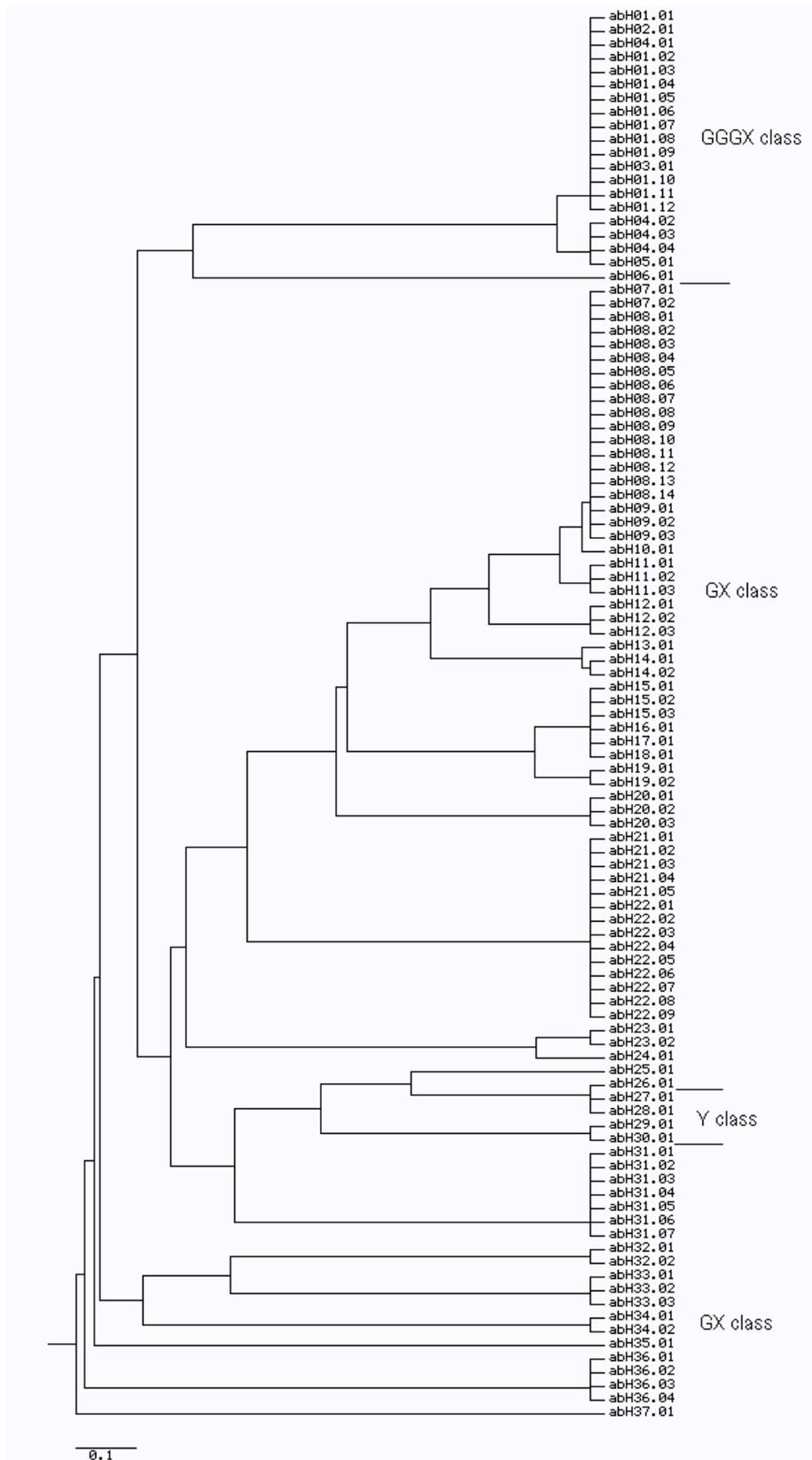


Figure 11 : Arbre phylogénétique des hydrolases à repliement α/β (Fischer, 2004).

Une autre base de données consacrée aux lipases et aux estérases microbiennes, la MELDB (Microbial Esterases Lipases Database), a vu le jour en 2006 (Kang *et al.*, 2006). Cette nouvelle classification utilise les relations entre les propriétés biochimiques et les régions conservées fonctionnelles. Dans cette étude, les enzymes ont été classées d'après des alignements locaux sur les régions conservées. Ces régions ont été divisées en deux groupes, les domaines connus et les régions conservées de fonction inconnue. Cette approche diffère de celles qui, traditionnellement, se fondent sur des alignements globaux et des calculs de distance. Elle permet de réduire les méprises dues aux mauvais alignements de régions éloignées ou différentes. La MELDB recense 402 estérases et 481 lipases, issues de 257 souches, et réparties respectivement en 20 et 30 familles. Les groupes générés par la MELDB correspondent globalement à ceux de la LED. Toutefois des différences significatives ont été observées dans la composition des sous-groupes (Kang *et al.*, 2006).

1.10. Les enzymes lipolytiques de microalgues

Les lipases et les carboxylestérases sont des enzymes très étudiées, tant d'un point de vue de la relation structure/fonction, que de la régulation de la synthèse. Mais cette affirmation n'est valable que si l'on considère les enzymes lipolytiques d'origine animale, végétale, bactérienne ou fongique. La connaissance des enzymes algales est extrêmement réduite et fragmentée car très peu de travaux leurs ont été consacrés. A ce jour, aucune enzyme lipolytique de microalgue n'a été purifiée et caractérisée d'un point de vue biochimique.

En revanche, la présence d'enzymes lipolytiques, principalement des phospholipases, a été établie à quelques reprises par des mesures d'activités sur des extraits d'algues bruts ou des extraits protéiques. Les travaux de Bilinski décrivent la détection biochimique d'une phospholipase C chez la microalgue marine *Monochrysis lutheri* (Bilinski *et al.*, 1968). L'activité de l'enzyme a été mesurée à partir d'extraits protéiques totaux. L'enzyme catalyse l'hydrolyse de la phosphatidylcholine marquée au ^{14}C et libère de la phosphorylcholine. Cette phospholipase C n'a pu être purifiée par manque de matériel algal. Cependant les extraits bruts utilisés ont montré une activité non négligeable et reproductible. L'activité d'une phospholipase D a également été détectée chez la microalgue rouge unicellulaire *Porphyridium cruentum*. L'activité enzymatique a ici été estimée sur des lysats cellulaires obtenus par sonication puis incubés avec de la lécithine marquée au ^{14}C . La libération de choline témoigne de l'action d'une phospholipase D (Antia *et al.*, 1970).

Des activités lipolytiques ont aussi été observées chez *Dunaliella salina* (Roessler, 1990) et *Skeletonema costatum* (Bergé, 1996). Chez la diatomée *Thalassiosira rotula*, c'est l'action d'une phospholipase A₂ qui a été détectée (Pohnert, 2002). Chez la microalgue *I. galbana*, une activité lipolytique endogène a, dans un premier temps, été suspectée suite à l'observation d'acides gras libres dans les lipides totaux (Poisson, 1999). Des travaux menés au laboratoire en 2005, ont confirmé cette hypothèse avec la mise en évidence d'activités glycolipasiques et phospholipasiques intracellulaires chez *I. galbana* (Devos, 2005).

2. Les microalgues marines

2.1. Les microalgues marines dans la Biosphère

Les microalgues font partie des premiers organismes vivants apparus sur notre planète. Ce sont des organismes photosynthétiques unicellulaires. Les microalgues jouent un rôle majeur dans les écosystèmes marins, en tant que producteurs primaires. Les microalgues forment le phytoplancton qui est à la base de nombreuses chaînes alimentaires. L'embranchement des Haptophyta, auquel se rattache la microalgue *Isochrysis galbana*, forme, avec celui des Ochrophyta (diatomées) et des Dinophyta (dinoflagellés), l'un des trois principaux groupes eucaryotiques phytoplanctoniques marin. Les haptophytes appartiennent majoritairement au nanoplancton.

A l'échelle de la planète, les algues microscopiques marines sont impliquées dans des cycles biogéochimiques fondamentaux comme ceux du carbone et du soufre et participent ainsi à la régulation du climat global (Simo, 2001). Mais la pollution d'origine anthropique fragilise cet équilibre et amplifie les phénomènes de prolifération des microalgues. Lorsque les conditions nutritionnelles et lumineuses sont réunies, il arrive que les microalgues se multiplient et atteignent une densité importante (blooms), allant jusqu'à créer des zones d'anoxie pour la faune aquatique. Ces floraisons phytoplanctoniques sont d'autant plus inquiétantes lorsque les algues incriminées sécrètent aussi des substances toxiques (Maso et Garces, 2006).

Malgré leur importance écologique majeure, les microalgues restent peu connues par manque de données biologiques. Devant la nécessité d'approfondir les connaissances sur les organismes vivants et leurs milieux, pour préserver la diversité biologique et les écosystèmes et pour développer la valorisation commerciale des microalgues, plusieurs programmes de recherche internationaux se sont mis en place. En Europe, par exemple, les efforts de

recherche se sont focalisés sur la biologie, l'écologie et la génomique des haptophytes avec les projets CODENET (Coccolithophorid Evolutionary Biodiversity and Ecology Network) et EHUX (*Emiliana huxleyi*) (Harris, 1996).

2.2. L'histoire évolutive des microalgues

Les microalgues possèdent un noyau, nettement délimité par une membrane, et des organites intracellulaires (plastides et mitochondries). L'évolution moléculaire de leur patrimoine génétique est unique. Illustrée dans ce paragraphe, nous verrons qu'elle se fonde notamment sur des transferts de gènes endosymbiotiques ayant abouti à la formation de génomes inédits.

2.2.1. L'origine symbiotique et procaryotique des plastides

Les microalgues partagent une origine évolutive commune basée sur la théorie de l'endosymbiose. Cette théorie évolutive s'appuie sur l'hypothèse de l'origine symbiotique et procaryotique des mitochondries et des plastides (Margulis, 1992). Dès la fin du XIX^{ème} siècle, quelques scientifiques suggèrent que les plastides découlent de bactéries qui auraient vécu en symbiose dans des cellules primitives après avoir été ingérées. Des années plus tard, les techniques d'analyses modernes et notamment les techniques de biologie moléculaire ont conforté ces observations en confirmant les liens de parenté entre ces organites et les procaryotes.

Les fondements de l'origine endosymbiotique et procaryotique des plastides s'appuient sur des ressemblances structurales, biochimiques et moléculaires entre les cyanobactéries et les plastides de cellules eucaryotes actuelles (Selosse et Loiseaux-de Goër, 1997). Les plastides sont des organites indépendants. Ils proviennent toujours de la division d'un plaste préexistant, à la différence des autres organites intracellulaires. Les plastides possèdent leur propre génome (Ris et Plaut, 1962). Cet ADN ressemble à celui des procaryotes : il consiste en plusieurs copies identiques d'une molécule circulaire. On a découvert aussi des gènes homologues de gènes bactériens dans le génome des plastides (Reith et Munholland, 1993). Les gènes de ces organites ont la même structure que ceux des procaryotes (même zone promotrice). Les ribosomes des plastides sont plus petits que ceux du cytoplasme. Ils ont en fait la taille et la composition en ARN des ribosomes de procaryotes. Les plastides sont délimités par une double membrane : une membrane interne dont la composition lipidique est proche de celle des

membranes bactériennes et une membrane externe issue de la membrane plasmique de l'hôte qui entoure la bactérie lors de sa pénétration dans la cellule.

L'endosymbiose n'est pas un phénomène rare. Actuellement, certaines espèces de bactéries et d'algues unicellulaires sont impliquées dans des associations symbiotiques et vivent dans le cytoplasme d'eucaryotes pluri ou unicellulaires. C'est souvent le cas chez les cnidaires comme les anémones et les coraux. L'anémone de mer *Condylactis gigantea*, par exemple, abrite dans ses cellules des algues photosynthétiques (des zooxanthelles) qui approvisionnent l'anémone en molécules organiques. Le dinoflagellé *Symbiodinium sp.* participe au développement des récifs coraliens en assurant le transfert du carbone fixé aux tissus coralliens (Selosse, 2006).

2.2.2. L'origine endosymbiotique des microalgues

Apparues, il y a plusieurs centaines de millions d'années, les microalgues seraient issues de 1 à 3 phénomènes successifs d'endosymbiose car leurs plastes sont entourés de plusieurs membranes (Keeling *et al.*, 2004). Les haptophytes, dont les plastes sont entourés de 4 membranes, auraient, au cours de leur évolution, subi 2 endosymbioses successives (Delwiche, 1999) (Figure 12).

Une première association symbiotique, entre une bactérie photosynthétique (une cyanobactérie) et un eucaryote primitif non photosynthétique, serait à l'origine de l'existence des algues rouges unicellulaires (Figure 12A). Les membranes photosynthétiques des plastes des algues rouges évoquent fortement les thylacoïdes de bactéries photosynthétiques. Les lipides caractéristiques des plastes se retrouvent également chez les cyanobactéries. Enfin, l'équipement pigmentaire des algues rouges rappelle celui des cyanobactéries : chlorophylle a et phycobiliprotéines.

Une deuxième endosymbiose, entre un eucaryote photosynthétique (une algue rouge) et un eucaryote non photosynthétique, aurait permis l'apparition de plusieurs lignées de microalgues, comme les cryptophytes, les hétérokontes et les haptophytes (Figure 12B). Les haptophytes possèdent, en effet, un ou deux plastes entourés de quatre membranes, qui, de l'extérieur vers l'intérieur, correspondraient à une membrane de séquestration de l'hôte secondaire, une membrane du symbiote secondaire (l'algue internalisée), une membrane de l'hôte primaire, une membrane du symbiote primaire (la cyanobactérie).

L'étude de la cryptophyte *Guillardia theta* a fortement contribué à l'approbation de l'origine symbiotique secondaire des plastes à quatre membranes des microalgues. *G. theta* est une petite algue unicellulaire, dont le plaste ne serait, en fait, qu'une algue rouge régressée (Douglas et Penny, 1999). Le plaste à 4 membranes de la cryptophyte contient un restant de reticulum d'algue rouge et l'espace entre la 2^{ème} et la 3^{ème} membrane comporte un organe contenant de l'ADN, le nucléomorphe. Le génome du nucléomorphe a été entièrement séquencé (Douglas *et al.*, 2001) ; il est constitué de trois petits chromosomes dont les gènes sont proches de ceux des algues rouges. *G. theta* contient donc quatre génomes distincts situés respectivement dans le noyau cellulaire, dans la mitochondrie, dans le noyau vestigial ou nucléomorphe de l'algue internalisée et dans le plaste de l'algue internalisée.

Des incertitudes persistent cependant, notamment sur l'ordre d'apparition et l'indépendance des événements d'endosymbiose. Les analyses phylogénétiques actuelles tendent à conforter l'idée que les plastes des haptophytes, des cryptophytes et des hétérokotes dériveraient d'endosymbioses secondaires indépendantes.

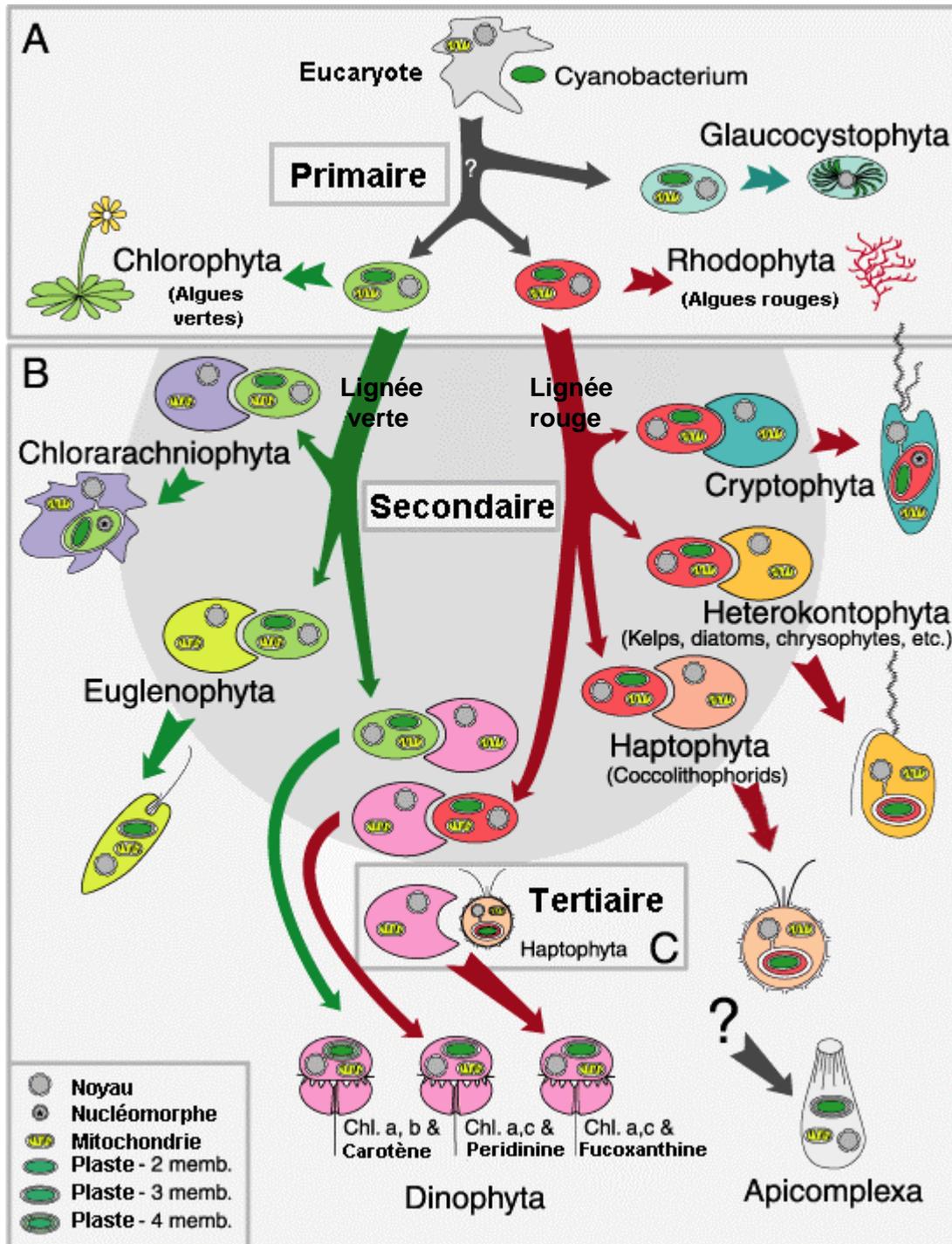


Figure 12 : L'origine endosymbiotique des microalgues d'après Delwiche, 1999.
 A) Endosymbiose primaire. B) Endosymbiose secondaire. C) Endosymbiose tertiaire.

2.2.3. Les transferts de gènes

La symbiose jouerait un rôle important dans l'évolution des organismes vivants en créant très rapidement des organismes chimériques à l'origine de lignées nouvelles (Selosse et Loiseaux-de Goër, 1997). En rapprochant des partenaires différents et complémentaires, elle favoriserait les transferts de gènes et créerait de nouveaux génomes (Figure 13). Les microalgues disposeraient donc d'un patrimoine génétique original. Leur génome serait le résultat de transferts de gènes qui se seraient produits, à chaque endosymbiose, entre les génomes nucléaires, chloroplastiques et mitochondriaux des hôtes et des endosymbiotes (Hackett *et al.*, 2004; Li *et al.*, 2006).

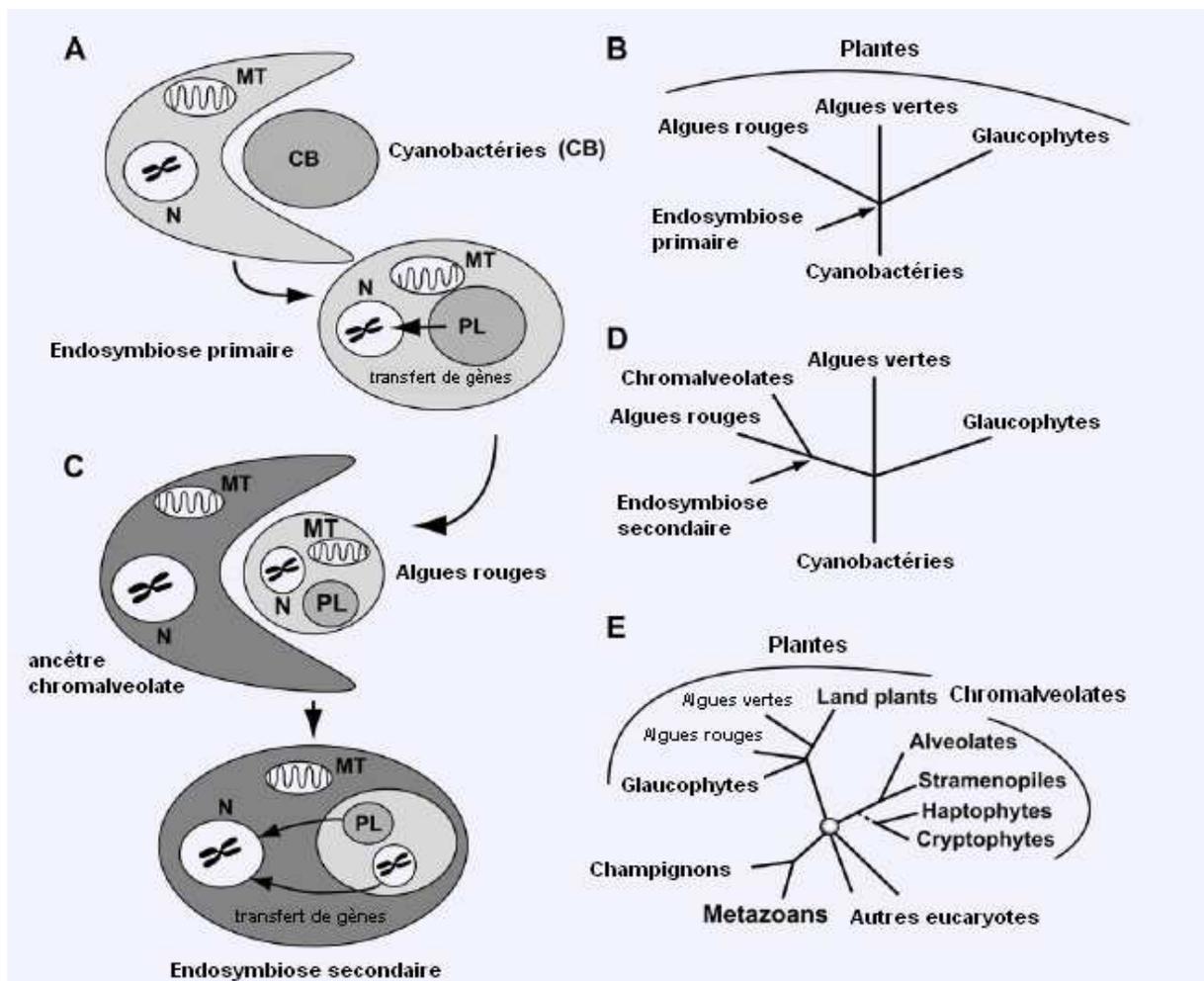


Figure 13 : Transferts de gènes endosymbiotiques d'après Li *et al.*, 2006.

Les transferts de gènes sont indiqués par des flèches noires. A) Endosymbiose primaire entre une cyanobactérie et un protiste hétérotrophe avec transfert de gènes de la bactérie vers le noyau du protiste. C) Endosymbiose secondaire d'une algue rouge. Le noyau de l'algue rouge endosymbiote a disparu chez tous les chromalvéolates excepté les cryptophytes. Tous les membres de la lignée des chromalvéolates possèdent des gènes issus de l'algue rouge internalisée. B, D et E) Arbres eucaryotiques simplifiés illustrant les principales lignées de microalgues. MT : mitochondrie, N : noyau, CB : cyanobactérie, PL : plaste.

D'un point de vue génétique, les microalgues sont donc très probablement des organismes sources d'innovations et de spécificités génétiques.

2.3. Position phylogénétique des microalgues

L'établissement des relations phylogénétiques entre les organismes a longtemps reposé sur des considérations morphologiques et biochimiques. Cette première approche avait permis, avec un certain succès, la différenciation de nombreuses lignées sans pour autant permettre de résoudre les liens existants entre ces lignées. A l'origine de nombreuses avancées scientifiques, la biologie moléculaire a également révolutionné la phylogénie, dans les années 1990. Au départ, les arbres étaient générés à partir d'un seul gène, qui, le plus souvent, était le gène codant pour la petite sous-unité de l'ARN ribosomique (SSU rRNA). Aujourd'hui la construction des arbres phylogénétiques combine une variété de données de toutes natures.

Les microalgues constituent un groupe d'organismes polyphylétique. Bien qu'elles soient des organismes photosynthétiques, leur histoire évolutive est fondamentalement différente de celle des végétaux terrestres. Cinq grands groupes caractérisent la diversité des eucaryotes (Figure 14). L'ordre d'apparition des divergences demeure encore incertain : c'est le cas lorsque plusieurs branches émergent simultanément.

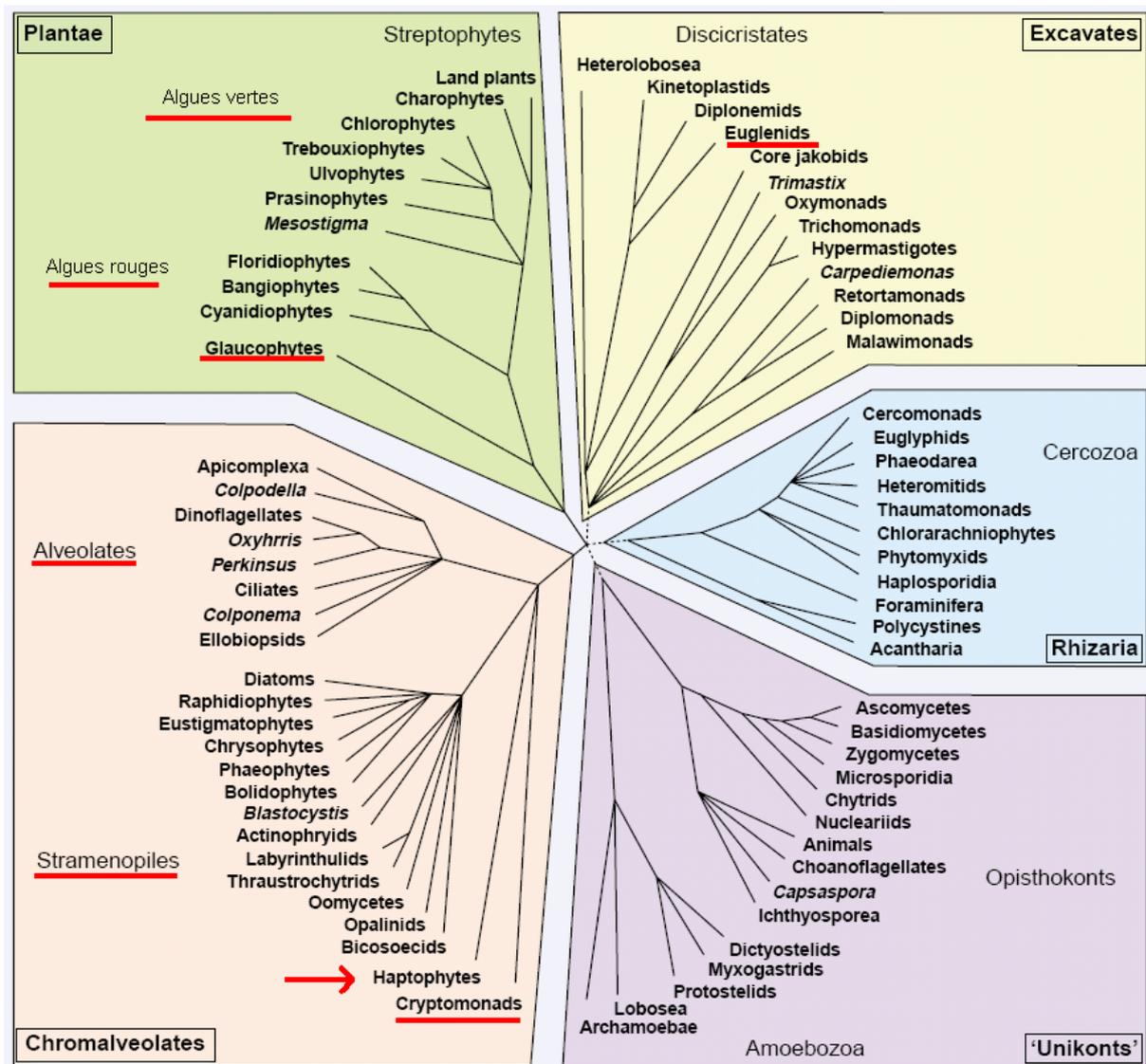


Figure 14 : Arbre phylogénétique des lignées eucaryotes d'après Keeling *et al.*, 2005.

Cet arbre a été construit à partir de données morphologiques, biochimiques et moléculaires. Les principaux groupes de microalgues sont soulignés en rouge et répartis dans trois ensembles différents : les alveolates, les stramenopiles, les haptophytes et les cryptomonades qui forment l'ensemble des Chromalveolates, les algues rouges, les algues vertes et les glaucophytes qui appartiennent à l'ensemble des Plantae et les euglènes. La flèche rouge indique la lignée des microalgues haptophytes au sein de laquelle se classe *Isochrysis galbana*.

Les chromalvéolates concentrent la grande diversité des microalgues parmi lesquelles figurent quelques groupes majeurs comme les diatomées, les dinoflagellés et les coccolithophores (haptophytes). A l'origine, la classification des organismes dans le groupe des chromalvéolates était principalement basée sur la présence de plastes, vestiges de la régression d'une algue rouge, après une endosymbiose secondaire (Cavalier-Smith, 1998). Depuis, les arbres, conçus, par exemple, à partir de la SSU rRNA ou de protéines cytosoliques codées par des gènes nucléaires, ont entériné la majorité des lignées de chromalvéolates (Van

de Peer et De Wachter, 1997; Baldauf *et al.*, 2000; Harper *et al.*, 2005). Au sein de ce groupe, les haptophytes forment une lignée eucaryotique distincte et relativement éloignée des algues stramenopiles (diatomées) et des alvéolates (dinoflagellés) en dépit des similitudes de l'ultrastructure des plastes et de la pigmentation.

2.4. Microalgues et génomique

2.4.1. Banques génomiques et collections d'ADN complémentaires (ADNc)

La maîtrise des techniques de biologie moléculaire et le développement de la bioinformatique ont favorisé l'isolement, le séquençage et l'annotation fonctionnelle des génomes ou des collections d'ADNc. On assiste aujourd'hui à une explosion de l'information génétique disponible. Depuis une décennie, les programmes de séquençage génomique de microalgues se développent (Grossman, 2005; Grossman, 2007).

Les génomes des organites intracellulaires et plus particulièrement les génomes des plastes sont aujourd'hui disponibles pour un grand nombre de microalgues, comme par exemple, la cryptophyte *Guillardia theta* (Douglas et Penny, 1999), la diatomée *Odontella sinensis* (Chu *et al.*, 2004), l'algue rouge *Cyanidium caldarium* (Glockner *et al.*, 2000) ou encore l'haptophyte *Emiliana huxleyi* dont les génomes de la mitochondrie (Sanchez Puerta *et al.*, 2004) et du chloroplaste (Sanchez Puerta *et al.*, 2005) ont été intégralement séquencés. Le génome mitochondrial d'*E. huxleyi* (29013 pb) code 21 protéines dont 14 composants de la chaîne respiratoire et 5 protéines ribosomales. Aucun de ces gènes ne contient d'introns. 27 gènes d'ARN sont également présents ; ils codent la grande et la petite sous-unité de l'ARN ribosomique et 25 ARNt. Le génome chloroplastique d'*E. huxleyi* (105309 pb) code 110 protéines impliquées dans la photosynthèse (sous-unités des photosystèmes I et II, phycobiliprotéines), le métabolisme énergétique (protéines de la chaîne du transport d'électrons, protéines impliquées dans la synthèse d'ATP, dans le métabolisme du pyruvate et de l'acétyl-CoA), la transcription (ARN polymérases, régulateurs), la traduction (protéines ribosomales, facteurs d'initiation, d'élongation), la biosynthèse d'acides aminés, d'acides gras, de la chlorophylle, de cofacteurs et de carbohydrates, le transport, la division cellulaire, le repliement des protéines, les voies de sécrétion, la détoxification et la dégradation. En plus de ces 110 gènes codant des protéines, 2 copies des gènes codant l'ARN 5S, les petite et grande sous-unités de l'ARN ribosomique sont présentes.

Les connaissances génomiques à l'égard des microalgues sont cependant encore très limitées car pour l'instant seulement six projets de séquençage du génome nucléaire de microalgues ont été menés à terme. C'est peu par rapport à la diversité et à l'importance écologique de ces organismes. Les projets achevés concernent l'algue rouge primitive *Cyanidioschyzon merolae* (Matsuzaki *et al.*, 2004; Misumi *et al.*, 2005), deux diatomées *Thalassiosira pseudonana* (Armbrust *et al.*, 2004) et *Phaeodactylum tricorutum* (Scala *et al.*, 2002), deux prasinophytes *Ostreococcus tauri* (Derelle *et al.*, 2006) et *Ostreococcus lucimarinus* (Palenik *et al.*, 2007) et la microalgue verte *Chlamydomonas reinhardtii* (Merchant *et al.*, 2007). Les principales caractéristiques de leurs génomes sont répertoriées sous forme statistique dans le Tableau 4. Ces microalgues appartiennent à des phylums distincts et parfois éloignés de celui des haptophytes. Mais le séquençage et l'annotation du génome nucléaire de l'haptophyte *Emiliana huxleyi*, espèce voisine d'*I. galbana*, sont, actuellement, sur le point d'être achevés (http://jgi.doe.gov/sequencing/DOE_microbes2003.html).

	<i>C. merolae</i>	<i>P. tricorutum</i>	<i>T. pseudonana</i>	<i>C. reinhardtii</i>	<i>O. tauri</i>	<i>O. lucimarinus</i>
Taille du génome en Mpb	16,52	31	34,27	121	12,56	13,2
Nombre de chromosomes	20	33	24	17	20	21
Nombre de gènes	5331	10681	11 242	15143	7892	7651
Exons par gène	NC	1,8	2,5	8,33	1,57	1,27
Gènes avec introns (%)	0,5	NC	NC	92	39	NC
Longueurs moyennes en pb ou en aa :						
Gène	1552	1572	1745	4312	1245	1284
Transcrit	1543	1465	1556	1580	1175	1234
Exon	NC	824	613	190	750	970
Intron	248	138	125	373	126	187
Protéine	NC	453	498	NC	NC	NC

Tableau 4 : Quelques données statistiques relatives aux génomes nucléaires de microalgues.

Ces informations proviennent soit de la littérature, soit des bases de données consultables en ligne. NC : données non communiquées.

Parallèlement aux banques génomiques, les banques d'ADNc, qui compilent des étiquettes de séquences transcrites (EST), se sont aussi multipliées. Les EST sont de courts fragments d'ADNc sélectionnés aléatoirement puis séquencés partiellement sur une longueur de 300 à 600 nucléotides, ce qui est suffisamment long pour caractériser les gènes sans ambiguïté. Le séquençage se fait à partir de l'une ou l'autre des extrémités, parfois au sein de l'ADNc. Les EST représentent les gènes exprimés dans une cellule à un temps donné. Les banques d'ADNc sont plus simples à concevoir et moins onéreuses que les banques génomiques, elles

sont donc plus nombreuses. A ce jour, quatre microalgues de l'embranchement des Haptophyta ont fait l'objet de ces projets de séquençage partiel : *Pavlova lutheri* (Keeling, 2006) de la classe des Pavlovophyceae et trois espèces de la classe des Prymnesiophyceae, la prymnesiale *Prymnesium parvum* (La Claire, 2006) et les isochrysidales *Emiliana huxleyi* (Wahlund *et al.*, 2004) et *Isochrysis galbana* (Keeling, 2006). Seuls 36 % des 1523 EST d'*E. huxleyi* analysés présentent des homologies significatives avec les séquences indexées dans GenBank. La majorité des séquences (64 %) montre peu ou pas de similarités de séquences significatives avec les données existantes ce qui permet d'envisager la présence de nouveaux gènes (Wahlund *et al.*, 2004).

2.4.2. Les apports de la génomique

La génomique comparative consiste à comparer les données de séquences. En évaluant les différences et les similitudes entre les microalgues et les autres eucaryotes, on obtient des indications sur l'origine, l'évolution et le fonctionnement de ces microorganismes. Connaître les gènes d'un organisme, c'est déjà connaître un peu son fonctionnement. Avec ces données de séquences, on commence également à comprendre l'organisation et le profil d'expression des gènes chez les algues.

Le séquençage complet du génome de la diatomée, *Thalassiosira pseudonana* a révélé que, d'un point de vue moléculaire, les diatomées partagent des homologies de séquences avec tous les types d'organismes. Pour la moitié des protéines de la diatomée, des scores d'alignements très élevés ont été retrouvés aussi bien avec la plante *Arabidopsis thaliana*, qu'avec la souris *Mus musculus* ou l'algue rouge *Cyanidioschyzon merolae* (Armbrust *et al.*, 2004).

2.5. Lipides et acides gras de microalgues

2.5.1. Les lipides

Les microalgues possèdent trois grandes classes de lipides : les lipides neutres constitués de triglycérides, d'acides gras libres, de stérols ; les glycolipides constitués essentiellement de mono- et di-galactosyldiglycérides et de sulfoquinovosyldiglycérides ; et enfin les phospholipides constitués surtout de phosphatidylcholines et phosphatidyléthanolamines (Alonso *et al.*, 1998).

Certaines algues sont capables d'accumuler jusqu'à 86 % de leur poids sec en lipides neutres, surtout en réponse à des modifications de leur environnement (concentration en azote, salinité, température) (Murphy *et al.*, 2001). Ces lipides de réserve sont stockés dans des particules subcellulaires, les corps lipidiques. Les lipides accumulés dans ces gouttelettes cytosoliques de 1 à 8 μm sont majoritairement des triglycérides dont la composition en acides gras est très diversifiée (Liu et Lin, 2001).

Si les triglycérides constituent une forme de stockage d'énergie, les glycolipides et phospholipides ont un rôle structural. Les glycolipides sont associés exclusivement aux membranes plastidiennes. Les plastes étant le siège de la photosynthèse, ces lipides se retrouvent spécifiquement chez tous les organismes photosynthétiques. Les phospholipides sont les constituants des membranes cellulaires.

2.5.2. Composition en acides gras des lipides algaux

Les lipides de microalgues se caractérisent par des profils en acides gras très riches et diversifiés (Alonso *et al.*, 1998). De nombreuses études portant sur la composition des lipides en acides gras ont révélé la présence, chez certaines microalgues, d'acides gras polyinsaturés (AGPI) en large quantité. Certains de ces AGPI appartiennent à la série des oméga 3 (AGPI- ω 3). C'est le cas de l'acide docosahexaénoïque (C22:6 ω 3, DHA) et de l'acide eicosapentaénoïque (C20:5 ω 3, EPA) dont les effets bénéfiques pour la santé sont aujourd'hui largement reconnus (Guesnet *et al.*, 2005).

Chez les haptophytes *Isochrysis sp.* et *Pavlova lutheri*, on observe une distribution, en pourcentage d'acides gras des lipides totaux, dominée par les types C14:0, C16:0, C16:1 ω 9, C18:1 ω 9 et C22:6 ω 3 (Liu et Lin, 2001). Plus élevée chez les espèces du genre *Isochrysis*, la teneur en AGPI de type C18:1 ω 9 et de type C18 :4 ω 3 varie de 15 à 32 % et de 15 à 24,9 % selon la souche. Par contre, l'EPA, peu présent chez *Isochrysis*, est trouvé en très grande quantité chez *P. lutheri* (21 %). L'abondance en AGPI varie donc d'une espèce à l'autre.

Dans la majorité des cas, ce sont les lipides polaires (glycolipides et phospholipides) qui sont riches en AGPI. Une étude a montré que chez *Nannochloropsis sp.*, l'acide gras majoritaire est l'EPA (29 % des acides gras des lipides totaux). Il représente 40 % des acides gras des glycolipides (Sukenic *et al.*, 1989). La même tendance a été observée chez *I. galbana* : le DHA, qui est souvent l'AGPI majoritaire de cette souche, représente jusqu'à 50 % des acides

gras des phospholipides (Devos *et al.*, 2006). Certaines souches de dinoflagellés affichent également cette aptitude : les glycolipides de ces souches sont, à quelques exceptions près, riches en C18:4 ω 3 et C18:5 ω 3 et leurs phospholipides sont riches en DHA (Leblond et Chapman, 2000).

2.5.3. Fonctions biochimique et biologique des lipides et acides gras algaux

Depuis longtemps, les lipides et les AGPI de microalgues représentent une véritable manne pour les professionnels de l'aquaculture, de la nutrition animale et de la diététique. Les microalgues ont donc fait l'objet d'études portant sur l'impact des facteurs abiotiques sur la synthèse des lipides et des acides gras et sur la composition des lipides en acides gras. Comme on va le voir, ces travaux constituent, finalement, un moyen d'appréhender, chez les microalgues, l'importance de la lipolyse et du rôle joué par les enzymes lipolytiques.

Défense chimique

Pour se défendre des prédateurs herbivores, les microalgues ont conçu des systèmes de défenses sophistiqués. Certaines espèces de diatomées synthétisent des aldéhydes toxiques qui influent négativement sur la reproduction des copépodes (Ianora *et al.*, 1996; Ban *et al.*, 1997). En effet, ces composés stoppent la division cellulaire et favorisent l'apoptose chez les embryons et les larves de copépodes (Adolph *et al.*, 2004).

Chez la diatomée *Thalassiosira rotula*, un stress mécanique, comme la sonication, génère une élévation remarquable du taux d'aldéhydes qui est nul lorsque les algues ne sont soumises à aucun stress. La sonication stimule aussi la production d'AGPI dans la cellule. Selon l'auteur, après une agression, la diatomée *Thalassiosira rotula* investit son énergie métabolique dans l'élaboration de phospholipides. Ces lipides polaires sont ensuite rapidement hydrolysés entraînant la libération d'AGPI, principalement des acides eicosanoïques. Or les acides eicosanoïques sont impliqués dans la voie de biosynthèse des aldéhydes en tant que précurseurs de ces métabolites défensifs. L'inhibition des lipases s'accompagne d'une forte diminution de la production d'aldéhydes (Pohnert, 2002).

Adaptation aux fluctuations environnementales

Des études récentes portant sur la microalgue verte d'eau douce *Parietochloris incisa* suggèrent que, chez certaines espèces de microalgues, les triglycérides (lipides neutres) ne

seraient pas seulement un produit final du métabolisme lipidique, mais constitueraient un réservoir d'AGPI pour la production rapide de phospholipides et de glycolipides membranaires (lipides polaires) (Cohen *et al.*, 2000; Khozin-Goldberg *et al.*, 2005).

Les triglycérides de la chlorophyte *Parietochloris incisa* sont naturellement riches en acide arachidonique (20 :4 ω6, AA). Cet acide gras représente 43 % des acides gras des lipides totaux de la microalgue. L'AA est majoritairement présent dans les triglycérides qui constituent la classe de lipides dominante. Les triglycérides contiennent 43 % de l'AA total en phase logarithmique de croissance de l'algue et 47 % en phase stationnaire (Bigogno *et al.*, 2002a; Bigogno *et al.*, 2002b). Une étude a permis d'établir une corrélation entre la diminution de la température du milieu de culture et le changement de la composition en AA des lipides neutres et polaires chez *P. incisa* (Cohen *et al.*, 2000). En suivant le parcours de l'AA marqué au ^{14}C , les auteurs de l'étude ont observé chez l'algue cultivée à 25 °C, un niveau de radioactivité relativement stable aussi bien chez les lipides neutres que polaires. A 12 °C, un rapide transfert de la radioactivité se produit des triglycérides vers les lipides polaires. A 4 °C le phénomène s'accroît (Figure 15). Au cours de cette expérience, le taux de radioactivité mesuré ne faiblit pas et l'AA demeure le seul acide gras radiomarqué à la fin de l'expérimentation prouvant que les acides gras ne subissent pas de transformations.

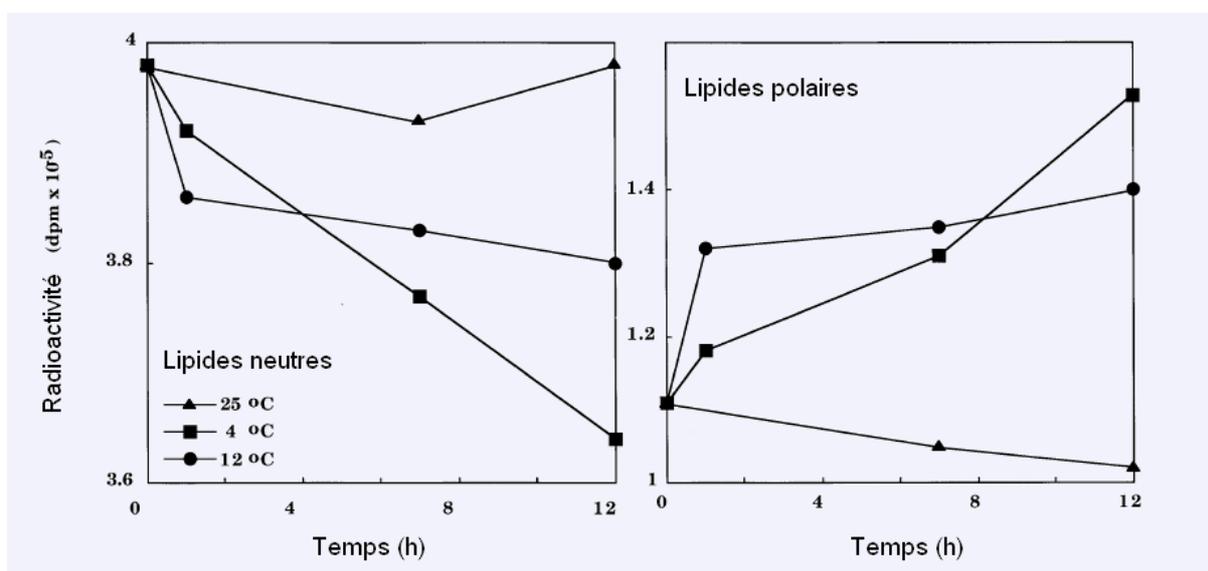


Figure 15 : Effet de la température sur la composition en acide arachidonique des lipides neutres et polaires de *Parietochloris incisa* d'après Cohen *et al.*, 2000.

Les algues ont été cultivées une première fois 12 h à 25 °C en présence d'acide arachidonique (20 :4 ω6, AA) marqué au ^{14}C , puis 12 h à 25 °C, à 12 °C et à 4 °C.

L'augmentation du degré d'insaturation des membranes liée à une baisse brutale de la température est un phénomène connu depuis longtemps (Patterson, 1970). L'adaptation à de soudaines chutes de températures inclut l'amélioration rapide du contenu membranaire en acides gras insaturés. Cohen et son équipe ont donc suggéré que chez les algues riches en AGPI-LC et dont l'habitat naturel est caractérisé par des changements rapides de températures, les triglycérides riches en AGPI formeraient un réservoir en AGPI pour la construction des membranes et pour améliorer le degré d'insaturation de ces membranes.

Les microalgues peuplent une large variété de niches écologiques. Certaines espèces ont développé des stratégies d'adaptation à leur milieu et réagissent aux fluctuations des facteurs abiotiques. Les variations de température, de salinité, de luminosité ou de teneur azotée influent notamment sur l'organisation de leur membrane cellulaire. Les triglycérides formant les corps lipidiques cytosoliques des microalgues serviraient alors de tampon interne. En fournissant rapidement des groupements acyles spécifiques, les triglycérides participeraient à la réorganisation et à la stabilisation des membranes dans un environnement fluctuant. Ces processus de transfert d'acides gras des lipides neutres vers les polaires nécessitent probablement l'intervention de lipases et de phospholipases.

2.6. La microalgue *Isochrysis galbana*

La microalgue *Isochrysis galbana* a été choisie comme modèle d'étude par le LALUM, principalement en raison de sa teneur élevée en DHA. Elle est au centre d'un projet relatif à la valorisation de ses lipides. Le laboratoire travaille actuellement au développement de procédés enzymatiques pour la production de phospholipides structurés riches en DHA à partir des lipides d'*I. galbana*.

2.6.1. Organisation structurale, taxonomie et phylogénie

Les représentants de la division des Haptophyta, comme l'espèce *I. galbana*, évoluent surtout dans les milieux océaniques oligotrophes. Les haptophytes sont unicellulaires et appartiennent majoritairement au nanophytoplancton. Toutes les espèces de ce groupe sont caractérisées par la présence d'un appendice multi-fonctionnel particulier, l'haptonème. Situé entre les deux flagelles, il aurait un rôle dans la détection de l'environnement extérieur et dans l'adhésion au substrat. L'haptonème aiderait également à la préhension des proies chez certaines espèces

mixotrophes (Kawachi *et al.*, 1991). L'haptophyte toxique *Prymnesium parvum*, par exemple, est une espèce à la fois capable de produire des molécules organiques par photosynthèse et capable d'incorporer par phagocytose des particules de grande taille et même des microorganismes planctoniques variés (bactéries, protistes) (Tillmann, 1998; Skovgaard *et al.*, 2003; Tillmann, 2003). L'embranchement des haptophytes compte environ 80 genres et près de 300 espèces réparties en deux classes : Pavlovophyceae et Prymnesiophyceae. La classe des Prymnesiophyceae décrite par Hibberd en 1976 contient quatre ordres différents : les Isochrysidales, les Coccolithales (Schwarz, 1932), les Prymnesiales (Papenfuss, 1955) et les Phaeocystales (Medlin, 2000) (Jordan et Chamberlain, 1997).

Le genre *Isochrysis* appartient à l'ordre des Isochrysidales, défini par Pascher en 1910. L'espèce *galbana* a été identifiée pour la première fois par Bruce, Knight et Parke en 1939 (Figure 16). C'est une algue mobile qui possède deux flagelles lisses, égaux, insérés apicalement et dont l'haptonème est réduit. La cellule est recouverte d'écailles organiques à la différence de l'espèce voisine *Emiliana huxleyi* dont les écailles organiques sont, en plus, recouvertes d'un dépôt calcaire constituant des structures particulières, les coccolithes.

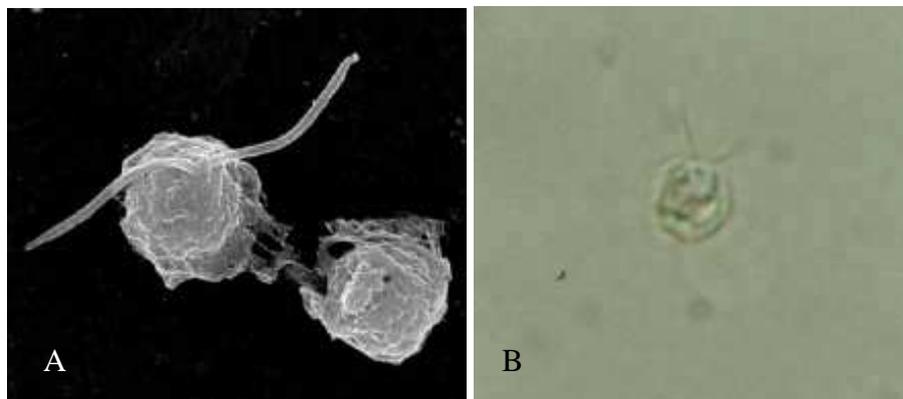


Figure 16 : *Isochrysis galbana* Parke, 1939.

A) Microscopie électronique à balayage, Observatoire Océanologique de Banyuls : photo M.J. Chrétiennot-Dinet (Robert *et al.*, 2004). B) Microscopie photonique (Poisson, 1999).

Les arbres phylogénétiques montrent les haptophytes comme un groupe monophylétique stable. L'arbre de la Figure 17 souligne qu'*I. galbana*, bien qu'elle ne soit pas une coccolithophore, est très proche de la coccolithophore *E. huxleyi*. Les Isochrysidales sont séparées des autres coccolithophores d'un point de vue génétique (gène codant pour la petite sous-unité ribosomale 18S (Edwardsen *et al.*, 2000), gène codant pour la grande sous-unité de la rubisco (Fujiwara *et al.*, 2001) mais aussi par des propriétés biochimiques et ultrastructurales (Jordan et Chamberlain, 1997; Andersen, 2004).

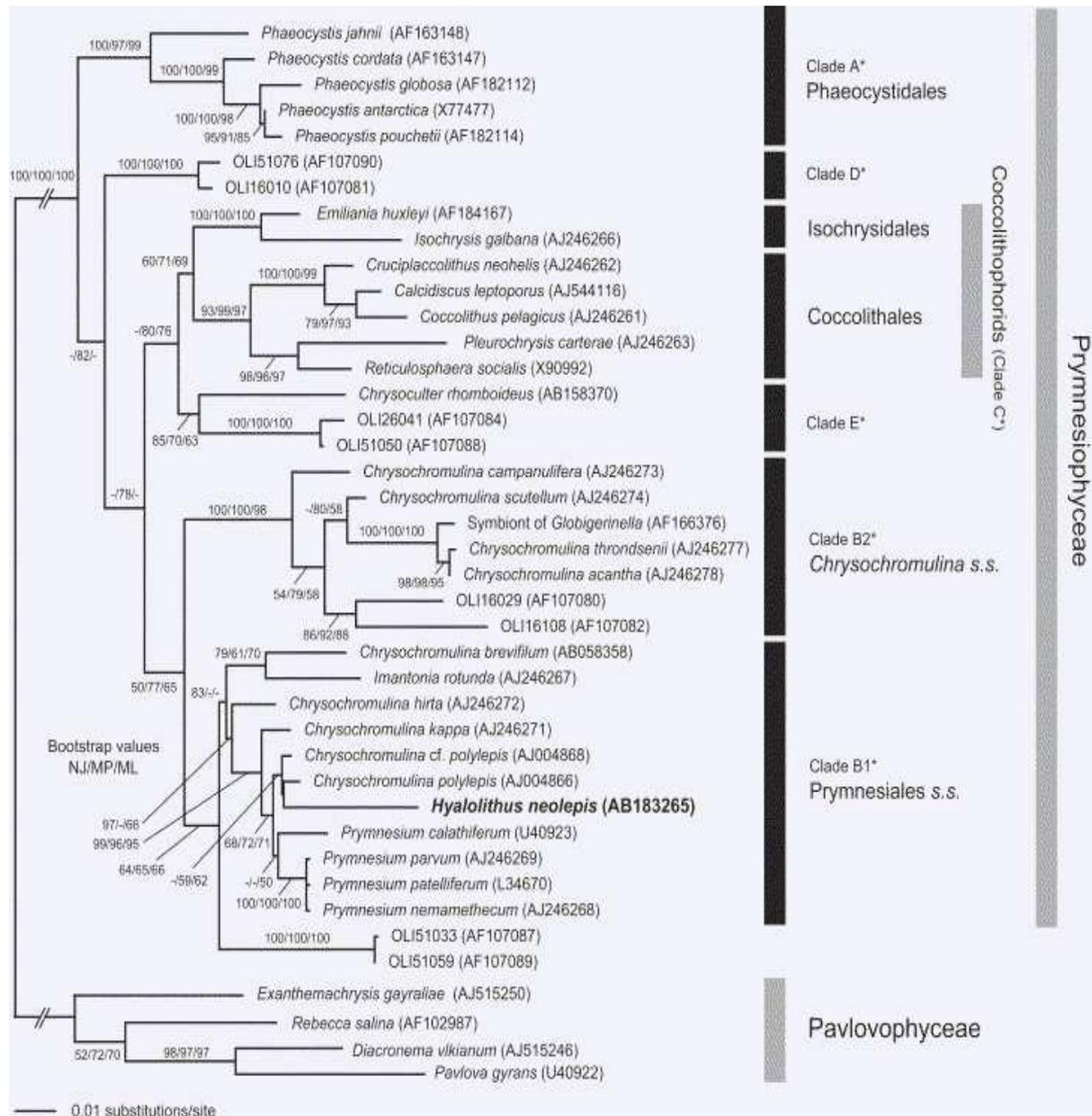


Figure 17 : Arbre phylogénétique des haptophytes (Edwardsen *et al.*, 2000; Yoshida *et al.*, 2006). L'arbre a été construit à partir des séquences de l'ADNr 18S et à partir des données morphologiques et ultrastructurales disponibles.

2.6.2. Lipides, composition en acides gras et activité lipolytique endogène

La présence de corps lipidiques chez *Isochrysis sp.* a été observée par microscopie optique et par microscopie électronique (Liu et Lin, 2001). Selon les auteurs, les corps lipidiques se forment au sein des chloroplastes et plus précisément dans les thylacoïdes. Les granules lipidiques apparaissent au cours de la phase de croissance d'une culture. Une partie des granules lipidiques accumulés diffuse ensuite dans le cytosol formant des gouttelettes lipidiques dont la taille varie de 0,5 à 3 μm , visibles uniquement en phase stationnaire.

Le profil en acides gras d'*I. galbana* est relativement complexe puisqu'une dizaine d'acides gras différents ont été détectés et recensés (Tableau 5).

	Lipides totaux	Lipides neutres	Glycolipides	Phospholipides
C14 :0	18 %	11,8 %	21,4 %	23,7 %
C16 :0	8,8 %	13,8 %	10,5 %	11,5 %
C18 :1 ω9	13,5 %	15,8 %	13,0 %	9,4 %
C18 :2 ω6	4,3 %	5,3 %	4,6 %	1,5 %
C18 :3 ω3	11,4 %	11,0 %	15,4 %	0,8 %
C18 :4 ω3	9,5 %	12,7 %	10,9 %	0,5 %
C20 :4 ω6	8,7 %	11,2 %	10,2 %	0,4 %
EPA	1,0 %	0,9 %	1,2 %	0,1 %
Cx :y	3,0 %	1,5 %	1,9 %	5,9 %
DHA	13,5 %	7,3 %	3,1 %	51,5 %

Tableau 5 : Profil en acides gras des lipides d'*Isochrysis galbana* cultivée en bioréacteur de 300 L (Devos, 2005).

L'acide gras noté Cx :y n'a pas été identifié. Selon les analyses effectuées, antérieurement au laboratoire, par chromatographie en phase gazeuse, cet acide gras serait un isomère de position du C24 :1 (Devos, 2005).

Parmi les acides gras détectés chez *I. galbana*, sept (C14:0, C16:0, C18:1 ω9, C18:3 ω3, C18:4 ω3, C20:4 ω6 et C22:6 ω3) sont présents dans des proportions importantes comprises entre 8,7 et 18 % des acides gras des lipides totaux (Devos *et al.*, 2006). Parmi les acides gras essentiels de la série des oméga 3, le DHA (C22:6 ω3) et l'EPA (C20:5 ω3) représentent respectivement 13,5 % et 1 % des acides gras des lipides totaux (Devos *et al.*, 2006). 75 % de la quantité totale de DHA sont localisés dans les phospholipides, le reste étant réparti équitablement entre les lipides neutres et les glycolipides. Chez *I. galbana*, le DHA est l'acide gras majoritaire des phospholipides qui constituent donc une fraction lipidique d'intérêt pour la valorisation de cet acide gras essentiel.

Chez *I. galbana*, des acides gras libres ont été observés parmi les lipides, suggérant une activité lipolytique endogène (Poisson, 1999). Des travaux plus récents ont montré, chez *I. galbana*, une activité lipolytique endogène dont l'action était spécifiquement dirigée contre les glycolipides et les phospholipides (Devos *et al.*, 2006). L'étude du profil en acides gras des phospholipides a révélé que le DHA, que l'on sait préférentiellement localisé en position sn-2 des phospholipides, était libéré en plus grande proportion au cours de cette activité enzymatique. Cette activité lipolytique inclut probablement une phospholipase de type A₂ (Devos *et al.*, 2006).

3. Applications industrielles des enzymes lipolytiques et valorisation des lipides de microalgues

Le thème développé dans ce troisième chapitre, illustre l'importance et les attraits des enzymes lipolytiques pour les industries biotechnologiques. De leur côté, les microalgues synthétisent un ensemble de molécules très variées, aux propriétés scientifiquement et industriellement reconnues. Les lipides algaux, par exemple, représentent aujourd'hui une opportunité intéressante de valorisation des algues, dans les domaines de la production de biocarburants ou de compléments alimentaires enrichis en AGPI essentiels.

3.1. Applications industrielles des enzymes lipolytiques

L'industrie enzymatique que nous connaissons aujourd'hui résulte de l'évolution rapide de la biotechnologie moderne des quatre dernières décennies. En permettant la production, à grande échelle, d'enzymes à partir de souches de microorganismes préalablement sélectionnées, le développement des procédés de fermentation a lancé l'introduction des enzymes dans les processus industriels. Aujourd'hui, avec l'apport du génie génétique, les enzymes industrielles sont produites par des souches de microorganismes modifiés génétiquement.

Les enzymes peuvent se substituer aux catalyseurs chimiques et répondre aux exigences du développement industriel durable. En effet, elles augmentent l'efficacité de production et réduisent, pour celles qui agissent à température ambiante, la consommation d'énergie. Beaucoup de procédés industriels fonctionnent à haute température ou haute pression ou dans des conditions hautement acides ou basiques. L'emploi des enzymes permet d'éviter ces conditions extrêmes ainsi que l'utilisation de réactifs corrosifs, car elles agissent à des températures modérées, à pression atmosphérique et dans des conditions de pH proche du neutre. Elles permettent aussi d'obtenir des produits sans déchets ou renfermant un minimum de déchets bio-dégradables.

Après les protéases et les carbohydrases, les enzymes lipolytiques sont le troisième grand groupe d'enzymes industrielles (Hasan *et al.*, 2006). La majorité des enzymes actuellement utilisées en industrie provient de bactéries et de champignons. Elles sont généralement plus stables que les lipases végétales ou animales (Hasan *et al.*, 2006). Les enzymes lipolytiques sont des biocatalyseurs d'intérêt car elles sont capables d'agir préférentiellement sur une

forme particulière de lipide (sélectivité de substrat), sur un acide gras déterminé (typosélectivité), ou sur une configuration spatiale donnée (stéréosélectivité). Ces enzymes ont également la faculté de distinguer les positions externes (1 et 3) et interne (2) du squelette carboné du glycérol : c'est la régiosélectivité.

En 1988, la société Novozymes lance la production de la première lipase recombinante commerciale : la lipolase®. C'est le premier produit enzymatique fabriqué à l'aide de microorganismes modifiés par manipulation génétique : dans ce cas, le gène codant pour la lipase du champignon ascomycète *Thermomyces lanuginosa* a été cloné et exprimé chez son cousin taxonomique *Aspergillus oryzae*.

Aujourd'hui, les lipases et les carboxylestérases sont utilisées dans de nombreux secteurs industriels. Leurs principales applications sont regroupées dans le Tableau 6.

Domaine industriel	Réaction	Exemples de produit ou d'application
Nettoyage, Détergents	Hydrolyse des graisses	Lessives, détachants
Industrie du papier	Hydrolyse	Dissolution enzymatique des pitches, blanchiment écologique de la pâte à papier
Textile	Hydrolyse	Blanchiment des tissus de type jean, modification des fibres synthétiques (polyester)
Cosmétique	Synthèse, transestérification	Soins cutanés, crèmes amaigrissantes, parfums
Oléochimie (huiles et graisses)	Transestérification, hydrolyse	Synthèse des esters, hydrolyse des graisses et de la lécithine
Chimie fine	Synthèse énantiosélective	Synthèse de substances organiques
Dégradation des hydrocarbures, Traitement des déchets	Hydrolyse	Nettoyage des eaux contaminées
Production de biodiesel*	Transestérification	Esters de méthyle
Tannage	Hydrolyse	Préparation des peaux, élimination des poils et de la graisse
Alimentation	Estérification	Synthèse de lipides structurés de meilleure qualité nutritionnelle, synthèse d'esters aromatiques
Applications médicales	Transestérification, hydrolyse	Lipides spécifiques, aides digestives, action bactéricide, diagnostic

Tableau 6 : Applications industrielles des enzymes lipolytiques d'après Jaeger et Eggert, 2002, Hasan *et al.*, 2006.

* Ce domaine est actuellement l'objet de travaux de recherche et n'a pas encore été transposé au stade industriel.

3.2. Valorisation commerciale des acides gras et lipides de microalgues

3.2.1. Algocarburants

L'idée d'utiliser les microalgues comme source d'agrocarburants n'est pas récente mais tend à devenir sérieusement considérée, avec l'épuisement à moyen terme des ressources pétrolières et avec les problèmes du réchauffement climatique lié aux combustibles fossiles (Chisti, 2007). Actuellement, le biodiesel est essentiellement produit à partir d'huiles d'origine végétale. Mais cette filière ne pourra couvrir à elle seule tous les besoins énergétiques, car pour produire les 49,35 millions de tonnes équivalent pétrole consommés annuellement par le transport en France, il faudrait cultiver 118 % de la surface totale du pays en tournesol (Danielo, 2005).

Comparées aux plantes, les microalgues réunissent un certain nombre d'avantages. Elles se multiplient rapidement (doublement de la biomasse en 24 heures), le rendement photosynthétique est beaucoup plus élevé et aucun apport phytosanitaire n'est nécessaire. Certaines souches de microalgues peuvent accumuler 50 à 80 % de leur poids sec en lipides, bien que des teneurs comprises entre 20 et 50% soient plus communes (Tableau 7) (Chisti, 2007).

Espèce	Teneur en lipides (en pourcentage de la matière sèche)
<i>Botryococcus braunii</i>	25-75
<i>Chlorella sp.</i>	28-32
<i>Cryptocodinium cohnii</i>	20
<i>Cylindrotheca sp.</i>	16-37
<i>Isochrysis sp.</i>	25-33
<i>Monallanthus salina</i>	>20
<i>Nannochloris sp.</i>	20-35
<i>Nannochloropsis sp.</i>	31-68
<i>Neochloris oleoabundans</i>	35-54
<i>Nitzschia sp.</i>	45-47
<i>Phaeodactylum tricorutum</i>	20-30
<i>Schizochytrium sp.</i>	50-77

Tableau 7 : Teneur en lipides de quelques espèces de microalgues d'après Chisti, 2007.

Suite au premier choc pétrolier intervenu en 1973, les Etats-Unis ont été les premiers à s'interroger sur la production de carburants à partir d'espèces de microalgues riches en triglycérides. En 1978, le Department of Energy's Office of Fuel Development fonde un programme de recherche, l'Aquatic Species Program (ASP), afin d'évaluer la faisabilité d'une production de biodiesel à partir de microalgues (Brown et Philippon, 1999). Aujourd'hui, différentes entreprises et laboratoires de recherche poursuivent ce travail, en cherchant à optimiser la bioproduction de la microalgue à la pompe et les techniques d'extraction. Par exemple, en France, le projet Shamash, qui a démarré en 2006 et regroupe 8 équipes et entreprises, a pour objectif de produire un biocarburant sous forme de monoester méthylique à partir de microalgues.

3.2.2. AGPI de microalgues, santé et alimentation

Les AGPI-LC de la série des oméga 3, comme l'EPA et le DHA, sont des acides gras essentiels pour l'homme. Contrairement aux algues marines qui sont capables de les synthétiser *de novo*, la capacité de synthèse des AGPI-LC, chez l'homme, est relativement faible et nécessite la présence de deux acides gras indispensables : l'acide linoléique (C18 :2 ω 6) et l'acide α -linoléique (C18 :3 ω 3). C'est pourquoi, les besoins humains en AGPI-LC doivent être complétés par l'alimentation.

Les effets bénéfiques du DHA et de l'EPA sur la santé sont, aujourd'hui, largement reconnus. Les AGPI-LC ont un rôle clé dans diverses fonctions biologiques : réponse inflammatoire, croissance et développement du fœtus, développement cérébral et rétinien post et pré-natal (Guesnet *et al.*, 2005; Zhao *et al.*, 2005; Chalon, 2006). Les acides gras oméga 3 figurent en effet parmi les nutriments essentiels qui interviennent dans la structure et la fonction du cerveau. Récemment, une relation a été prouvée entre la teneur membranaire cérébrale en DHA et les processus de neurotransmission (Vancassel, 2004; du Bois *et al.*, 2005). Une carence alimentaire en DHA conduirait à une modification de la libération des neurotransmetteurs et serait impliquée dans l'apparition de certaines maladies psychiatriques, dont l'hyperactivité chez l'enfant, la dépression, la schizophrénie ou la démence (Peet, 2003b; Peet, 2003a; Itomura *et al.*, 2005; McNamara et Carlson, 2006; Nieminen *et al.*, 2006). De nombreux travaux, montrant les effets des AGPI-LC de la série des oméga-3, en particulier ceux des huiles de poisson, sur des modèles animaux et chez l'homme, ont également été

publiés dans le cadre de la prévention de l'obésité, des maladies cardio-vasculaires, et de l'infarctus cérébral (Block *et al.*, 2007; von Schacky et Harris, 2007).

Selon un rapport de l'AFSSA de 2001 (données en cours de réactualisation), les apports alimentaires en acides gras oméga 3 de la population française sont insuffisants. L'agence qui préconise un apport de 120 mg de DHA par jour, estime que 100 % des français n'ont pas 2/3 des apports journaliers recommandés en oméga 3. Pour combler ce déséquilibre nutritionnel, l'AFSSA recommande la consommation d'aliments naturellement riches ou enrichis en acides gras oméga 3. Les industriels ont donc introduit sur le marché de nombreux compléments alimentaires riches en AGPI- ω 3, qui sont obtenus principalement à partir d'huiles végétales et d'huiles de poisson. Toutefois, ces sources sont aujourd'hui insuffisantes pour combler les attentes d'un marché en expansion. Des sources alternatives sont activement recherchées. Les huiles issues de microalgues diffèrent des huiles de végétaux supérieurs par leur concentration nettement supérieure en EPA et DHA. Des espèces de microalgues, dont la forte teneur en AGPI- ω 3 a été démontrée, sont donc de plus en plus considérées comme des possibles sources de DHA ou d'EPA (Tableau 8) (Pencreac'h *et al.*, 2004).

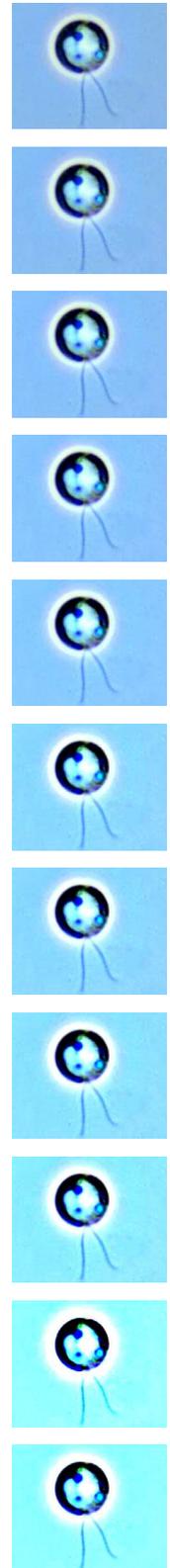
Espèce	AGPI- ω 3 produit	
	EPA	DHA
<i>Skeletonema costatum</i>	+	-
<i>Phaeodactylum tricornutum</i>	+	-
<i>Porphyridium purpureum</i>	+	-
<i>Monodus subterraneus</i>	+	-
<i>Nannochloropsis oculata</i>	+	-
<i>Chlorella minutissima</i>	+	-
<i>Cryptocodinium cohnii</i>	-	+
<i>Isochrysis galbana</i>	-	+
<i>Pavlova lutheri</i>	+	+

Tableau 8 : Principales espèces de microalgues d'intérêt pour la production d'EPA et de DHA (Pencreac'h *et al.*, 2004).

Actuellement, le DHA est le seul AGPI d'origine microalgale commercialisé. Un complément composé de triglycérides riches en DHA issus de *Cryptocodinium cohnii* est commercialisé par la société Martek Biosciences Corporation fondée en 1985 (Columbia, Etats-Unis), sous le nom de DHASCO®. Il est notamment préconisé pour la supplémentation des laits maternisés dans le cadre du développement cérébral postnatal. Le dinoflagellé *Cryptocodinium cohnii*

produit 30 % d'acides gras (en pourcentage de poids sec), le DHA pouvant atteindre 50 % des acides gras totaux (Jiang *et al.*, 1999). OmegaTech, aux Etats-Unis, exploite la microalgue *Schizochytrium* pour produire une huile commercialisée sous le nom DHA Gold®. Cette huile, riche en DHA, est utilisée, comme supplément alimentaire, en aquaculture, en alimentation animale et humaine (Spolaore *et al.*, 2006).

MATERIELS ET METHODES



1. Matériel biologique

1.1. *Isochrysis galbana* Parke

La souche axénique d'*Isochrysis galbana* (référence CCAP 927/1) provient de l'algothèque Culture Collection of Algae and Protozoae (Ambleside, Grande Bretagne) (Gachon *et al.*, 2007). Il s'agit de la souche *I. galbana* Parke qui fut identifiée, pour la première fois, par Mary Parke en 1939.

1.2. Milieu de culture Provasoli

I. galbana est cultivée au laboratoire en eau de mer stérilisée, enrichie en nutriments (milieu Provasoli ES). Le milieu est utilisé au tiers de sa concentration habituelle (Provasoli 1/3) (Robert, 1983). Les solutions à la base de sa préparation sont les suivantes :

Solution d'enrichissement :

NaNO ₃	350 mg
Na ₂ -Glycérophosphate	50 mg
Tris	500 mg
Solution A	2,5 mL
Solution B	25 mL
Eau déionisée	qsp 100 mL

Solution de vitamines :

Thiamine	50 mg
Vitamine B12 (0,2 g/L)	1 mL
Biotine (0,2 g/L)	5 mL
Eau déionisée	qsp 100 mL

Solution A :

Fe(NH ₄) ₂ (SO ₄) ₂ ·6H ₂ O	3,51 g
Na ₂ -EDTA	3,50 g
Eau déionisée	qsp 500 mL

Solution B :

H ₃ BO ₃	570 mg
FeCl ₃ , 6H ₂ O	24,5 mg
MnSO ₄ , 4H ₂ O	82 mg
ZnSO ₄ , 7H ₂ O	11 mg
CoSO ₄ , 7H ₂ O	2,4 mg
Na ₂ -EDTA	500 mg
Eau déionisée	qsp 500 mL

Solution mère de métasilicate* (100 mM) :

Na ₂ O ₃ Si, 9H ₂ O	14,2 g
Eau déionisée	qsp 500 mL

* Ce milieu de culture avait, au départ, été conçu pour la culture de diatomées, microalgues qui possèdent un exosquelette siliceux. La composition exacte du milieu a été conservée au laboratoire, bien que le métasilicate ne soit pas nécessaire à la culture d'*I. galbana*.

L'eau de mer est filtrée sur filtre de cellulose. La salinité de l'eau de mer est ajustée à 28 g/L. Pour une eau de mer de teneur moyenne en NaCl de 35 g/L, il convient d'ajouter 200 mL d'eau déionisée à 800 mL d'eau de mer. 7 mL de solution d'enrichissement et 1 mL de solution fille de métasilicate (100 μM) sont ajoutés par litre de milieu. Le pH est alors ajusté à 7,8 avec du HCl 1 N. Puis 2 mL de solution de vitamines ainsi que 80 mg d'hydrogénocarbonate de sodium (NaHCO_3) sont ajoutés par litre de milieu. Le milieu est enfin stérilisé à l'autoclave à 120 °C pendant 20 minutes.

1.3. Culture en Erlenmeyers

I. galbana est cultivée dans des Erlenmeyers de 500 mL contenant 200 mL de milieu de culture. Le milieu est inoculé stérilement avec 20 mL d'une pré-culture âgée de 7 jours. Les pré-cultures et les cultures sont placées à 20 °C en chambre d'incubation (Versatile Environmental test Chamber, Modèle MLR-350, Sanyo, Munich, Allemagne). Un éclairage de 40 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ est apporté grâce à 3 tubes fluorescents de 37 W.

1.4. Culture en photobioréacteur

I. galbana est également cultivée en photobioréacteur cylindrique de 2 L (volume utile : 1,5 L) couplé à une console de régulation et à une interface d'acquisition (Modèle Maestro, Pierre Guérin SA, Division Biolaffite, Mauzé sur le Mignon, France). Un volume de 1,3 L de milieu Provasoli ES est inoculé stérilement par une pré-culture de 200 mL âgée de 7 jours réalisée en Erlenmeyer de 500 mL. L'agitation dans le bioréacteur est assurée par une hélice dont la vitesse de rotation est fixée à 100 rpm. La température est réglée à 20 °C par un système de circulation d'eau dans une double paroi. Quatre tubes fluorescents de 37 W sont placés à 16 cm de la cuve pour assurer un éclairage de 60 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Le pH est régulé à une valeur de 8 grâce à des injections automatisées de CO_2 délivrées par une électrovanne. L'ouverture de l'électrovanne est commandée en fonction de la dérive du pH de la culture au dessus de la valeur consigne. Ces conditions de culture ont été mises au point, précédemment au laboratoire (Poisson et Ergon, 2001).

1.5. Suivi de la croissance et récolte de la biomasse

La densité cellulaire est déterminée par dénombrement sur cellule de Malassez ou estimée par spectrophotométrie à 520 nm. Avant dénombrement, la mobilité des cellules d'*I. galbana* est stoppée à l'aide de 1 à 3 µL de Lugol 5 % ajoutés aux 2 mL de prélèvement. La biomasse est récupérée par centrifugation à 2500 x g pendant 10 minutes à 4 °C (centrifugeuse Sigma 4K15, Fisher Bioblock Scientific, Illkirch, France). Les culots de cellules algales récoltés sont ensuite stockés à -70 °C.

2. Analyse des lipides

2.1. Extraction des lipides totaux

Les lipides totaux sont extraits d'après la méthode de Bligh et Dyer (Bligh et Dyer, 1959). Le protocole est donné pour une quantité de biomasse correspondant à une culture de 1,5 L en photobioréacteur. Après centrifugation (2500 x g, 10 min, 4 °C), les algues récoltées sont resuspendues dans du tampon phosphate 20 mM (pH 7,4) puis broyées dans un tube de Dounce. Le broyat est mélangé à 3 mL d'un mélange méthanol:chloroforme (1:2, v/v). Les échantillons sont agités vigoureusement au vortex pendant une vingtaine de secondes, puis centrifugés à 3000 x g pendant 5 min à température ambiante. La phase inférieure chloroformique, contenant les lipides, est prélevée et transférée dans un tube à hémolyse en verre. Pour concentrer les échantillons, le solvant est évaporé sous un flux d'azote.

2.2. Analyse par chromatographie sur couche mince

Les lipides totaux extraits sont analysés par chromatographie sur couche mince (CCM). Les échantillons lipidiques sont déposés à intervalles réguliers sur une plaque de silice (Silica gel 60, Merck, France) de dimensions 20 x 20 cm à l'aide d'une micropipette. La séparation des différentes classes de lipides est réalisée en deux temps (Henderson et Tocher, 1992). La première migration est effectuée dans le mélange de solvants suivant : acétate de méthyle: alcool isopropylique:chloroforme:alcool méthylique:solution aqueuse de KCl à 0,5 % (25:25:25:10:9 v/v/v/v/v). Dans ces conditions, les lipides polaires (phospholipides et glycolipides) sont séparés ; les lipides neutres migrant avec le front de solvant. La migration est arrêtée lorsque le front de solvant atteint environ la moitié de la plaque. La plaque est alors

placée dans une seconde cuve saturée avec le mélange suivant : hexane:éther éthylique:acide acétique glacial (80:20:2 v/v/v), et ceci pour permettre la séparation des lipides neutres et des acides gras libres. La migration est stoppée lorsque le front du solvant atteint le haut de la plaque. Après séchage de la plaque, les lipides sont révélés par pulvérisation d'une solution d' α -naphthol (0,25 g d' α -naphthol dans 50 mL d'éthanol et 50 mL de H₂SO₄ à 20 %) suivi d'un chauffage de la plaque dans une étuve à 100 °C pendant environ 15 min. Les fractions lipidiques obtenues sont constituées de spots spécifiques. Ils ont été identifiés, lors des travaux de Marie Devos, par séparation des classes de lipides puis par comparaison de leur rapport frontal avec ceux de standards commerciaux (Larodan Lipids, Malmö, Suède) chromatographiés dans les mêmes conditions (Devos, 2005).

3. Mesure des activités lipolytiques

3.1. Mise en évidence d'une activité lipolytique endogène chez *I. galbana*

3.1.1. Détection de l'activité lipolytique endogène d'*I. galbana*

Ce protocole a été mis au point antérieurement au laboratoire (Devos, 2005). Il consiste à récolter la biomasse d'une culture de 1,5 L en photobioréacteur. Cette biomasse est ensuite broyée au piston pendant une minute dans un tube de Dounce de 15 mL, en présence de 5 mL de tampon phosphate 20 mM (pH 7,4). Le broyat est récupéré puis transféré dans des microtubes à raison de 1 mL par tube. Les microtubes sont incubés sous agitation (800 rpm) à 40 °C dans un thermomixeur (Eppendorf, Le Pecq, France) pendant 24 heures afin de permettre l'hydrolyse des lipides endogènes par l'activité lipolytique endogène. Après réaction, les lipides sont extraits (§2.1.) et analysés en CCM (§2.2.) afin d'observer la disparition éventuelle de lipides sous l'action d'une activité lipolytique endogène. Pour certifier l'origine hydrolytique enzymatique des acides gras libérés, des échantillons témoins sont ébouillantés avant d'être incubés.

3.1.2. Localisation de l'activité lipolytique endogène d'*I. galbana*

Pour tenter de localiser l'activité enzymatique dans la cellule, les algues sont broyées dans un tube de Dounce puis centrifugées 30 minutes à 10000 x g de façon à ce que le cytosol soit séparé des résidus membranaires. Les fractions cytosolique et membranaire sont réparties

dans deux microtubes distincts de 1 mL. La fraction cytosolique est additionnée de 100 µL de lipides préalablement extraits de la microalgue. Les échantillons sont incubés 24 h à 40 °C sous agitation à 800 rpm. Après réaction, les lipides totaux sont extraits et analysés par CCM (§2.2.).

3.2. Mesure de l'activité lipolytique de *Saccharomyces cerevisiae* transformée avec le plasmide pYES2 recombiné

3.2.1. Test d'hydrolyse de substrat gélosé

L'activité lipolytique des levures modifiées a été testée sur des émulsions gélosées de tributyrine (C4), d'huile d'olive (la trioléine C18 étant l'espèce lipidique dominante) et de lécithine de soja (phospholipides). Les substrats sont préparés à 10 g/L dans du milieu SC-U (composition décrite au §8.7.) supplémenté de 1,5 % (w/v) d'agar et de 2 % (w/v) de galactose (sucre nécessaire à l'induction de l'expression de l'insert). Les solutions sont émulsionnées à l'aide d'un homogénéisateur puis chauffées jusqu'à ébullition avant d'être coulées en boîtes de Petri. Les levures transformées avec le plasmide recombiné sont étalées sur boîte de Petri puis incubées 48 h à 30 °C. Après croissance, la présence d'une activité lipolytique extracellulaire se traduit par l'apparition d'une zone claire résultant de l'hydrolyse des substrats émulsionnés. La visualisation de la région hydrolysée peut être accentuée en présence de bleu Victoria à 10 µg/mL (Yadav *et al.*, 1998; Abousalham et Verger, 2000). Des témoins positifs sont préalablement réalisés avec deux enzymes commerciales, la lipase F (Amano Enzyme Europe, Oxfordshire, Grande Bretagne), extraite de *Rhizopus oryzae* et la lécitase 10L (Novozymes, Bagsvaerd, Danemark) qui est une solution de phospholipase A₂ purifiée obtenue à partir de pancréas de porc. Les enzymes ont été déposées en surface ou dans des puits creusés dans la gélose. Des témoins négatifs sont réalisés à partir de levures transformées avec le plasmide non recombiné.

3.2.2. Test d'hydrolyse de substrat chromophore

Ce test d'activité consiste à mesurer l'hydrolyse de palmitate de para-nitrophényle à 16,5 mM préparé dans de l'isopropanol. Le mélange réactionnel est composé d'1 volume de cette solution de substrat et de 9 volumes de tampon Tris-HCl 50 mM (pH 8,0) contenant 0,4 % (v/v) de Triton X100 et 0,1 % (w/v) de gomme arabique (Pencreac'h et Baratti, 1996). Le

mélange réactionnel est placé à 37 °C. La réaction démarre avec l'ajout de la lipase F (contrôle positif) ou de l'échantillon à tester (broyat de levures transformées ou surnageant de culture de ces levures) préparés dans du tampon Tris-HCl 50 mM (pH 8,0). La formation de de p-nitrophényle est suivie par l'évolution de l'absorbance à 410 nm par spectrophotométrie.

3.2.3. Test d'hydrolyse des lipides totaux d'*I. galbana*

Les levures transformées avec le plasmide recombiné sont cultivées en Erlenmeyers de 500 mL, dans 100 mL de milieu sélectif SC-U complété en galactose. Après 24 h d'incubation à 30 °C, la biomasse et le surnageant de culture sont séparés par centrifugation 5 min à 5000 x g. La biomasse est broyée au piston pendant une minute dans un tube de Dounce de 15 mL, en présence de 5 mL de tampon phosphate 20 mM (pH 7,4). Le broyat de cellules et le surnageant de culture sont transférés dans des microtubes. Sont ajoutés 100 µL de lipides totaux préalablement extraits de la microalgue *I. galbana*. Les microtubes sont incubés à 40 °C sous agitation à 800 rpm dans un thermomixeur pendant 24 heures afin de permettre l'hydrolyse des lipides de la microalgue par l'enzyme recombinée. Après réaction, les lipides sont extraits (§2.1.) et analysés en CCM (§2.2.) afin d'observer la disparition éventuelle de lipides sous l'action d'une activité lipolytique. Pour certifier l'origine hydrolytique enzymatique des acides gras libérés, des échantillons témoins sont ébouillantés avant d'être incubés. Pour certifier que l'hydrolyse résulte de l'action de l'enzyme recombinante, la souche de levure transformée avec le plasmide non recombiné, cultivée dans un milieu SC-U en présence de galactose, est traitée en parallèle dans les mêmes conditions.

4. Isolement des acides nucléiques

4.1. Extraction et purification des ARN totaux

La méthode utilisée est fondée d'une part sur la solubilisation différentielle des acides nucléiques et des protéines entre 2 phases non miscibles (1 phase aqueuse et 1 phase phénolique) et d'autre part sur la précipitation différentielle de l'ADN et de l'ARN dans le chlorure de lithium (LiCl). Pour inactiver les RNases, l'extraction des ARN totaux doit être réalisée avec des solutions traitées à 0,1 % (v/v) de diéthylpyrocarbonate (DEPC). Après 12 h d'incubation à température ambiante, les solutions sont autoclavées de façon à inactiver le DEPC.

Les extractions d'ARN ont été réalisées à partir de cultures de microalgues en photobioréacteur de 1,5 L. Les culots d'algues, obtenus après centrifugation, sont broyés au mortier dans l'azote liquide jusqu'à l'obtention d'une fine poudre. Celle-ci est ensuite reprise dans 5 mL de tampon de lyse à pH 8 (1 volume de Tris-HCl 2 M, 2 volumes d'EDTA 0,5 M, 1 volume de SDS 20 %) et 5 mL d'un mélange de phénol:chloroforme:alcool isoamylique (25:24:1 v/v/v). Le tout est agité 30 minutes. Après centrifugation (15 min à 4500 x g, 16 °C), la phase aqueuse (supérieure) contenant les acides nucléiques est prélevée et transférée dans un tube contenant 5 mL de chloroforme. Après agitation, et centrifugation, la phase aqueuse est mélangée à un même volume de chlorure de lithium 6 M. L'ensemble est placé dans la glace, environ 16 heures, pour précipiter les ARN. Ces derniers sont récupérés le lendemain par centrifugation à 4 °C à 10000 x g pendant 20 minutes. Le culot est solubilisé dans 1 mL d'eau ultrapure traitée au DEPC. Les ARN sont ensuite précipités, en milieu acide, par l'ajout de 1/10^{ème} de volume d'acétate de sodium 3 M (pH 5,2) et de 3 volumes d'éthanol 100 %. Le mélange est centrifugé 20 min à 10000 x g à 16 °C. Le culot est dissout dans un volume approprié d'eau ultrapure traitée au DEPC. Les échantillons sont systématiquement contrôlés et traités avec une DNase avant d'être conservés à -70 °C.

4.2. Purification des ARN poly (A⁺)

Les ARN messagers polyadénylés (ARNm ou ARN poly (A⁺)) représentent moins de 10 % des ARN totaux produits. Les 90 % restants correspondent aux ARN impliqués dans les étapes transcriptionnelles ou traductionnelles (ARN ribosomaux (ARNr), ARN de transfert (ARNt), ARN nucléaires, ARN interférents). La purification des ARN messagers (poly(A⁺)) s'obtient par élimination des ARNr et des ARNt les plus abondants. Le kit GenEluteTM mRNA Miniprep de Sigma a été utilisé pour sélectionner les ARN poly (A⁺) selon le protocole recommandé par le fabricant. La méthode utilise la chromatographie d'affinité. Le segment polyadénylé situé à l'extrémité 3' des ARNm va s'hybrider avec des séquences oligo-(dT)30 liées de façon covalente à des billes de polystyrène de un micromètre. Les billes de polystyrène créent moins de liaisons non spécifiques que la cellulose et restent en suspension lors de l'hybridation avec les ARNm en l'absence d'agitation (Sene *et al.*, 1982). La phase aqueuse contenant les ARNr et ARNt est éliminée. La fraction enrichie en ARNm est lavée puis éluée avec 50 µL de Tris-HCl 10 mM pH 7,4.

4.3. Extraction et purification de l'ADN génomique

Plusieurs méthodes ont été utilisées pour l'extraction de l'ADN génomique d'*I. galbana*. La première méthode fait appel au protocole d'extraction et de purification des ARN totaux (§4.1.) qui permet d'isoler simultanément l'ADN génomique. Après précipitation des ARN au chlorure de lithium 6 M et récupération des ARN par centrifugation, le surnageant est additionné d'éthanol 100 %. L'ADN génomique précipite sous forme de filaments visibles qui sont récupérés à l'aide d'une pipette. L'ADN est ensuite resuspendu dans l'eau, lavé avant d'être à nouveau précipité, en milieu acide, par l'ajout de 1/10^{ème} de volume d'acétate de sodium 3 M (pH 5,2) et de 3 volumes d'éthanol 100 %. Le mélange est incubé à -70 °C pendant 30 min. L'ADN précipité est récupéré par centrifugation à 15000 x g pendant 10 min, lavé avec de l'éthanol à 70 % et dissout dans un volume approprié de Tris-EDTA (Tris-HCl 10 mM, EDTA 1 mM pH 8).

La seconde méthode est une extraction au phénol. Les algues cultivées en Erlenmeyers sont récoltées par centrifugation et lysées dans 2 mL d'un mélange (1:1 v/v) de TAE 1X (Tris/acétate 40 mM, EDTA 1 mM pH 8) et de SDS 1 %. Les ARN sont éliminés sous l'action de 4 µL de RNase à 10 mg/mL (37 °C, 30 min). L'ADN est ensuite extrait par l'ajout de 2 mL de phénol saturé en Tris HCl. Le tout est vigoureusement agité puis centrifugé 5 min à 6000 x g. La phase aqueuse (supérieure) est prélevée puis mélangée à un même volume de chloroforme:alcool isoamylique (24:1 v/v). Après centrifugation, l'ADN est précipité avec 2 volumes d'éthanol 100 %, 1/10^{ème} d'acétate de sodium à 3 M (pH 5,2). Le mélange est incubé à -70 °C pendant 90 min. L'ADN précipité est récupéré par centrifugation à 15000 x g pendant 10 min, lavé avec de l'éthanol à 70 % et dissout dans un volume de Tris-EDTA (Tris-HCl 10 mM, EDTA 1 mM pH 8).

4.4. Extraction et purification d'ADN de bactériophages

La méthode utilisée est décrite par Sambrook et Maniatis dans la seconde édition du manuel « Molecular Cloning » de 1989 (Sambrook *et al.*, 1989). Les phages sont multipliés sur un tapis bactérien de manière à obtenir des plages de lyse bien individualisées. Une plage de lyse isolée est prélevée et déposée dans 1 mL de tampon SM (sodium magnésium) (NaCl 10 mM, MgSO₄.7H₂O 8 mM, Tris-HCl 50 mM (pH 7,5), gélatine 0,01 % (w/v)) additionné d'une goutte de chloroforme. Après 2 h d'incubation à température ambiante, 50 µL de la

suspension de bactériophages sont mélangés à 100 μL de bactéries *E. coli* XL1-Blue MRF' (Stratagène) diluées à une D.O.₆₀₀ de 0,5 dans du MgSO_4 10 mM. La suspension est incubée 20 min à 37 °C, puis supplémentée de quelques mL de milieu NZY à 60 °C (NaCl 85 mM, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 8mM, extrait de levure 5 g/L, NZ amine 10 g/L et 0,7 % d'agarose (w/v)). Le tout est immédiatement déposé sur milieu solide NZY (bouillon NZY + agar 1,5 % (w/v)) en boîte de Petri et immobilisé 10 min pour assurer une prise en masse de la gélose. Les bactéries et les phages sont incubés une nuit à 37 °C jusqu'à ce que les plages de lyse recouvrent presque entièrement la surface. 3 mL de solution λ (Tris-HCl 10 mM (pH 7,5), MgSO_4 10 mM) sont déposés sur la surface gélosée. Après 2 h d'incubation à température ambiante sous agitation douce et constante, le surnageant est récolté puis centrifugé 10 min à 4000 x g à 4 °C. Le surnageant est réparti par mL en microtubes de 2 mL. Il est traité avec 1 μL de RNase et de DNase à 1 mg/mL. La mixture est incubée 15 minutes à 37 °C. Un même volume d'une solution de NaCl 3M et de polyéthylène glycol 6000 30 % (w/v) préparée dans une solution λ est ajouté à la mixture. Le tout est vortexé, incubé 1 h dans la glace. Les phages précipités sont récupérés par centrifugation à 10000 x g 10 min à 4 °C. Le culot de bactériophages est resuspendu dans 0,5 mL de TE pH 8,0. Les phages sont lysés en présence de SDS 10 %, 5 min à 68 °C. Sont ajoutés successivement 10 μL de NaCl 5 M, 700 μL d'un mélange phénol:chloroforme:alcool isoamylique (25:24:1 v/v/v). Après agitation puis centrifugation 15 min à 10000 x g à température ambiante, la phase aqueuse (supérieure) contenant les acides nucléiques est prélevée, transférée, lavée avec du chloroforme. L'ADN phagique est ensuite précipité, en milieu acide, par l'ajout de 1/10^{ème} de volume d'acétate de sodium 3 M (pH 5,2) et de 3 volumes d'éthanol 100 %. Le mélange est incubé à -70 °C pendant 15 min. L'ADN précipité est récupéré par centrifugation à 12000 x g pendant 15 min à 4 °C, lavé avec de l'éthanol à 70 % froid, séché quelques instants à température ambiante puis dissout dans 50 μL de Tris-EDTA pH 8.

4.5. Extraction et purification d'ADN plasmidique

La purification d'ADN plasmidique est réalisée avec le kit PureYield™ Plasmid Midiprep System (Promega), selon les indications fournies par le fabricant. Ce kit utilise la méthode de lyse alcaline des cellules initialement proposée par Birnboim et Doly (Birnboim et Doly, 1979) et reprise par Sambrook (Sambrook *et al.*, 1989). Après extraction, la solution contenant l'ADN plasmidique est transférée dans une colonne avec filtre de silice. L'ADN retenu est lavé puis élué, soit avec de l'eau, soit avec un tampon approprié. La préparation

d'ADN est enfin analysée par électrophorèse et dosée par spectrophotométrie d'absorption à 260 nm.

Une autre méthode d'extraction reposant sur l'électroextraction, a été utilisée pour extraire l'ADN plasmidique de la levure *Saccharomyces cerevisiae*. Cette méthode permet d'extraire rapidement de l'ADN à partir de très peu de matériel biologique. L'approche mise au point s'inspire de travaux portant sur l'extraction de protéines de cellules de levures soumises à des impulsions électriques (Suga et Hatakeyama, 2001). En biologie moléculaire, les techniques d'électroporation sont aujourd'hui utilisées pour le transfert d'exogènes dans la cellule. Les chocs électriques accroissent la perméabilité des membranes cellulaires rendant possible les échanges entre les macromolécules intra et extracellulaires. Le protocole a été mis au point avec un électroporateur 2510 de marque Eppendorf. L'extraction est effectuée à partir de 2 mL d'une culture de 100 mL incubée 24h à 30°C, dans du milieu YPD (extrait de levure 1 % (w/v), peptone 2 % (w/v), dextrose 2 % (w/v)). Après centrifugation 5 min à 1600 x g à 4 °C, le culot de cellules est rincé 3 fois avec 200 µL d'eau ultrapure stérile. L'électroporation est effectuée dans une cuve de 2 mm, dans 200 µL d'eau ultrapure stérile. Les cellules sont soumises à des impulsions électriques successives (de 1 à 5) de 2500 V chacune. Le mixe d'électroporation est complété à 1 mL avec 600 µL d'eau physiologique (NaCl 9 ‰). Le tout est centrifugé 20 min à 10000 x g à 4 °C. Les acides nucléiques expulsés des cellules se retrouvent dans le surnageant.

5. Analyses des acides nucléiques

5.1. Analyse quantitative

La quantification des acides nucléiques extraits est réalisée par spectrophotométrie. Le dosage spectrophotométrique s'effectue à la longueur d'onde de 260 nm car les bases puriques et pyrimidiques des acides nucléiques absorbent fortement dans l'ultraviolet à cette longueur d'onde. Il est admis qu'une unité d'absorbance à 260 nm (A_{260}) correspond à une solution d'ARN à 40 µg/mL ou à une solution d'ADN double brin à 50 µg/mL. Un spectre d'absorption est réalisé entre 210 et 320 nm pour déceler la présence d'éventuels carbohydrates et de protéines qui absorbent respectivement à 230 et 280 nm. L'estimation du degré de pureté de l'échantillon se fait par le calcul du rapport A_{260}/A_{280} . Un rapport compris entre 1,8 et 2,0 indique un acide nucléique pur et dépourvu de contamination protéique.

5.2. Analyse qualitative

Les acides nucléiques extraits sont soumis à une électrophorèse, en gel d'agarose non dénaturant, pour évaluer leur intégrité et rechercher la présence de contaminants d'ADN génomique ou d'ARN selon le type d'acide nucléique analysé. La qualité des fragments amplifiés par PCR est également examinée sur gel d'agarose. Les séparations d'ADN et d'ARN ont été réalisées 90 min à voltage constant (100 V) sur gels d'agarose de 1 à 2,5 % (w/v) en présence de tampon TAE 1X. Après la migration électrophorétique, les gels d'agarose sont incubés 15 min dans du TAE 1X additionné de bromure d'éthidium (BET) (concentration finale de 2 µg/mL). Le BET est un agent fluorescent qui s'intercale entre les bases azotées des acides nucléiques. Il fluoresce lorsqu'il est excité sous l'action d'un rayon ultraviolet. Les images sont capturées au moyen du système GelDoc 2000 (Bio Rad).

La taille des amplicons ou des fragments de digestion est évaluée par référence à des marqueurs de tailles. Sont utilisés comme échelle de taille, dans ces expérimentations :

- les produits de la digestion de l'ADN du phage λ par les enzymes *EcoRI* et *HindIII* (21226, 5148, 4973, 4268, 3530, 2027, 1904, 1584, 1375, 947, 831 et 564 pb)
- les produits de la digestion du plasmide pBR322 par l'enzyme *HaeIII* (587, 540, 504, 458, 434, 267, 234, 213, 192, 184, 124, 123, 104, 89 et 80 pb)
- les produits de la digestion du plasmide Puc18 par l'enzyme *HaeIII* (587, 458, 434, 298, 267, 257, 174, 102 et 80 pb)
- et trois autres marqueurs (1 kb, 100 pb, 2-Log) commercialisés par BioLabs.

6. Techniques d'amplification des acides nucléiques

6.1. Transcription inverse et réaction de polymérisation en chaîne

6.1.1. La transcription inverse

La transcription inverse est la réaction qui consiste à synthétiser une molécule d'ADN à partir d'une séquence d'ARN. Le fragment d'ADN généré est appelé ADN complémentaire (ADNc). La réaction de transcription inverse est catalysée par une transcriptase inverse (RT) d'origine virale. Elle est généralement suivie d'une amplification du fragment d'ADNc.

Dans ces travaux, les ARN totaux ou les ARNm ont été rétro-transcrits en ADNc en utilisant la transcriptase inverse eAMV-RT de Sigma. A 1 µg d'ARN totaux ou 100 ng d'ARNm sont ajoutés 500 ng d'amorces poly(dT)₂₃. Le volume réactionnel est ajusté à 10 µL avec de l'eau traitée au DEPC. Les ARN sont dénaturés 10 min à 70 °C pour éliminer les structures secondaires, puis immédiatement placés sur de la glace. Aux 10 µL précédents sont ajoutés 10 µL d'une solution contenant 0,5 mM de dNTP, 20 unités d'inhibiteur de RNase, 20 unités de eAMV-RT et du tampon de réaction 1X. Le milieu réactionnel est alors incubé 50 min à 45 °C. Les échantillons sont conservés à -20 °C jusqu'à utilisation.

6.1.2. La réaction de polymérisation en chaîne

La PCR, mise au point par Saiki et Mullis en 1986, est une technique qui permet de synthétiser et d'amplifier l'ADN. Cette technique, dont le principe est détaillé en Annexe 1, est basée sur des cycles d'amplification, composés chacun de trois étapes :

- une étape de dénaturation qui permet de séparer les doubles brins d'ADN matrice et de les linéariser
- une étape d'hybridation au cours de laquelle l'amorce va s'apparier à la séquence de l'ADN matrice qui lui est complémentaire et,
- une étape d'élongation pendant laquelle l'ADN polymérase va synthétiser l'ADN à partir de l'extrémité 3'OH de l'amorce dans le sens 5'P vers 3'OH.

La *Taq* ADN polymérase (Sigma) a été utilisée pour des amplifications de contrôle et la JumpStart™ *Taq* ADN polymérase (Sigma) pour l'amplification de fragments destinés au séquençage. Il est, en effet, préférable d'utiliser une ADN polymérase de haute fidélité, surtout si le fragment à amplifier est long (> 1000 pb), de façon à limiter les erreurs dans la séquence amplifiée. Le mélange réactionnel est réalisé dans un volume final de 25 µL comprenant 100 ng d'ADN matrice, 2,5 µL de tampon de réaction 10X (100 Mm Tris-HCl (pH 8,3), 500 Mm KCl), 200 mM de MgCl₂, 200 µM de dNTP, 50 pmol d'amorce sens, 50 pmol d'amorce anti sens, 2,5 unités d'ADN polymérase. Le volume final est ajusté avec de l'eau ultrapure stérilisée. Les échantillons sont ensuite placés dans le thermocycleur de type GeneAmp® PCR System 9700 (Applied Biosystems) et soumis à un programme comportant une première étape de dénaturation de 5 min à 95 °C, de 35 cycles comprenant une dénaturation d'1 min à 95 °C, une étape d'hybridation d'1 min entre 62 et 65 °C selon les

amorces et une élongation de 1 min 30 à 72 °C, puis une étape finale d'élongation à 72 °C de 10 min pour l'achèvement des brins incomplets.

6.2. Amplification rapide des extrémités des ADNc (RACE-PCR)

La RACE PCR est une méthode permettant d'isoler la séquence complète d'un ARNm ciblé parmi des milliers d'autres. L'application de cette méthode exige, au préalable, de connaître une séquence d'au moins 30 nucléotides, au sein du gène candidat (Chenchik *et al.*, 1996). Les séquences des ADNc pleine longueur ont été obtenues par des expériences de 5'- et 3'-RACE PCR, en utilisant le kit SMARTTM RACE cDNA Amplification, commercialisé par Clontech.

6.2.1. La transcription inverse

Les premiers brins d'ADNc, nouvellement synthétisés par transcription inverse, sont, ici, encadrés par deux courts oligonucléotides (Figure 18). L'oligo (dT)30 (Oligo A), qui sert d'amorce à la RT en s'hybridant sur la queue polyadénylée de l'ARNm, est ici additionné d'une séquence de 25 nucléotides. La méthode fait aussi appel à une reverse transcriptase spécifique. La PowerScript Reverse Transcriptase (PS RT) de Clontech est une variante de la transcriptase inverse du virus de la Maloney Murine Leucaemia, qui, lorsqu'elle atteint la fin d'une matrice d'ARN, exerce une activité terminale transférase, c'est-à-dire qu'elle ajoute 3 à 5 résidus (essentiellement des dC) à l'extrémité 3' du premier brin d'ADNc formé (Figure 18). Un deuxième oligonucléotide (Oligo B), fourni dans le kit, long de 30 pb et constitué d'une terminaison 3' de résidus G, s'hybride sur cette région polyC et sert de matrice pour la PowerScript RT qui synthétise la terminaison du premier brin d'ADNc. Tous les ADNc, synthétisés par la PS RT à partir des ARNm ou des ARN totaux, sont donc encadrés par deux courtes régions nucléotidiques de séquence connue. Ces deux zones serviront de matrice pour l'hybridation d'amorces, qualifiées d'universelles, dans les réactions PCR ultérieures (Figure 18).

La synthèse des ADNc se déroule dans un volume réactionnel final de 10 µL. A 1 µg d'ARN total ou 100 ng d'ARNm sont ajoutés les oligonucléotides A et/ou B (1 µL à 12 µM par réaction). Le mélange est incubé 10 min à 72 °C puis immédiatement transféré sur glace. Sont alors ajoutés 2 µL de tampon de réaction 5X, 1 µL de dithiothréitol 20 mM, 1 µL de dNTP

10 mM et 1 μ L de PS RT à 100 unités/ μ L. Le tout est mélangé puis incubé 1 heure 30 à 42 °C. Les ADNc nouvellement synthétisés sont conservés à -20°C jusqu'à 3 mois.

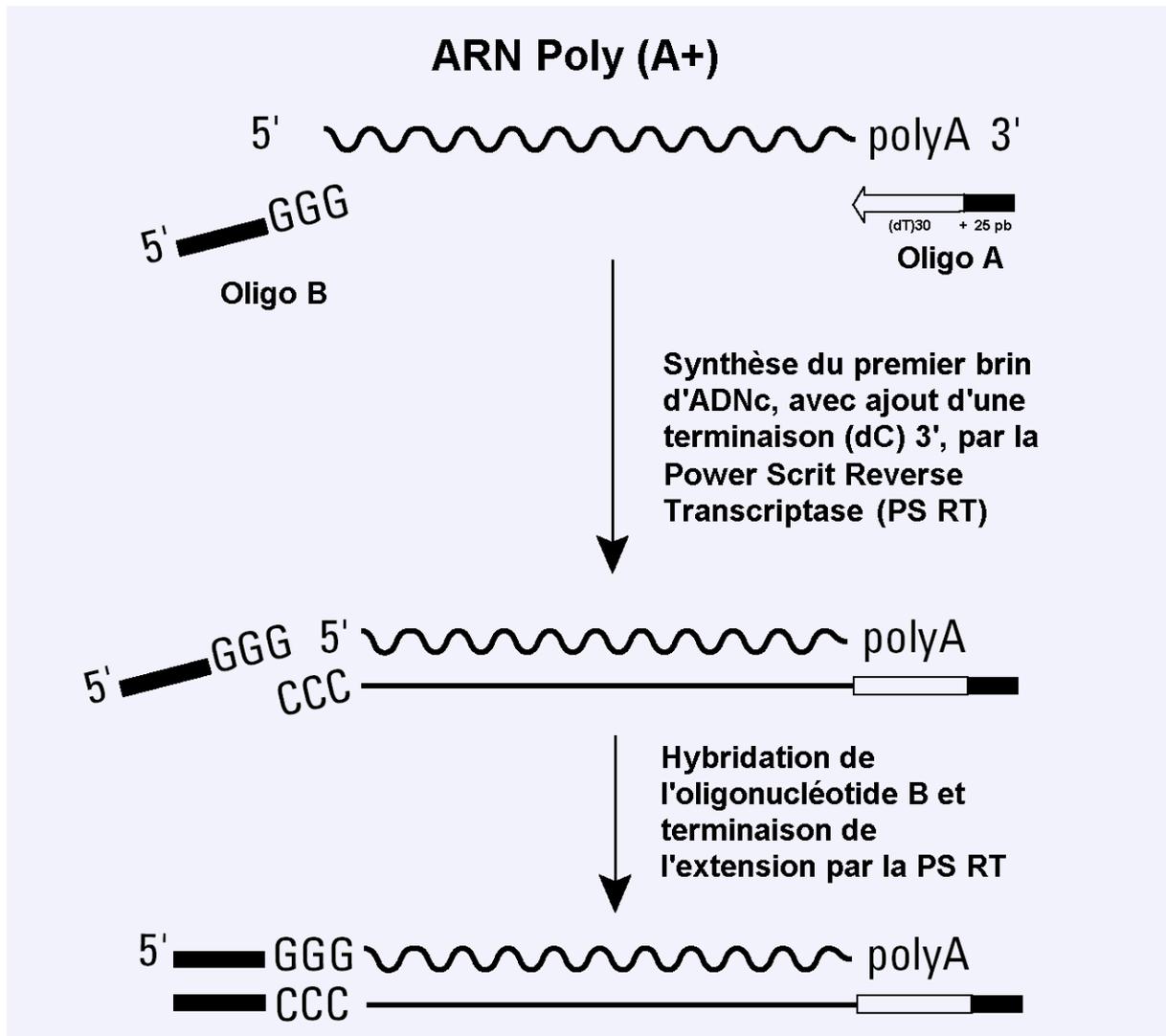


Figure 18 : Synthèse et encadrement des ADNc, selon le mécanisme SMARTTM (Switching Mechanism At 5' end of RNA Transcript) (d'après Clontech).

6.2.2. La réaction de polymérisation en chaîne

La connaissance d'une région du gène candidat permet de définir deux amorces spécifiques encadrant cette région : une amorce sens spécifique (*ASsp*) et une amorce anti sens spécifique (*AASsp*). Elles sont positionnées de telle sorte que la région connue soit amplifiée dans les deux sens puis séquencée deux fois (Figure 19). Pour l'amplification par PCR de la région 3' de l'ADNc, l'amorce sens spécifique *ASsp* est utilisée en couple avec un mixe d'amorces universelles (*AU*) fourni par le fabricant (Figure 19 et Annexe 2). Les amorces *AASsp* et *AU* assureront l'amplification de la région 5' (Figure 19 et Annexe 3). Après séquençage des deux

régions 5' et 3' amplifiées, la séquence complète peut être reconstituée par chevauchement des deux séquences obtenues.

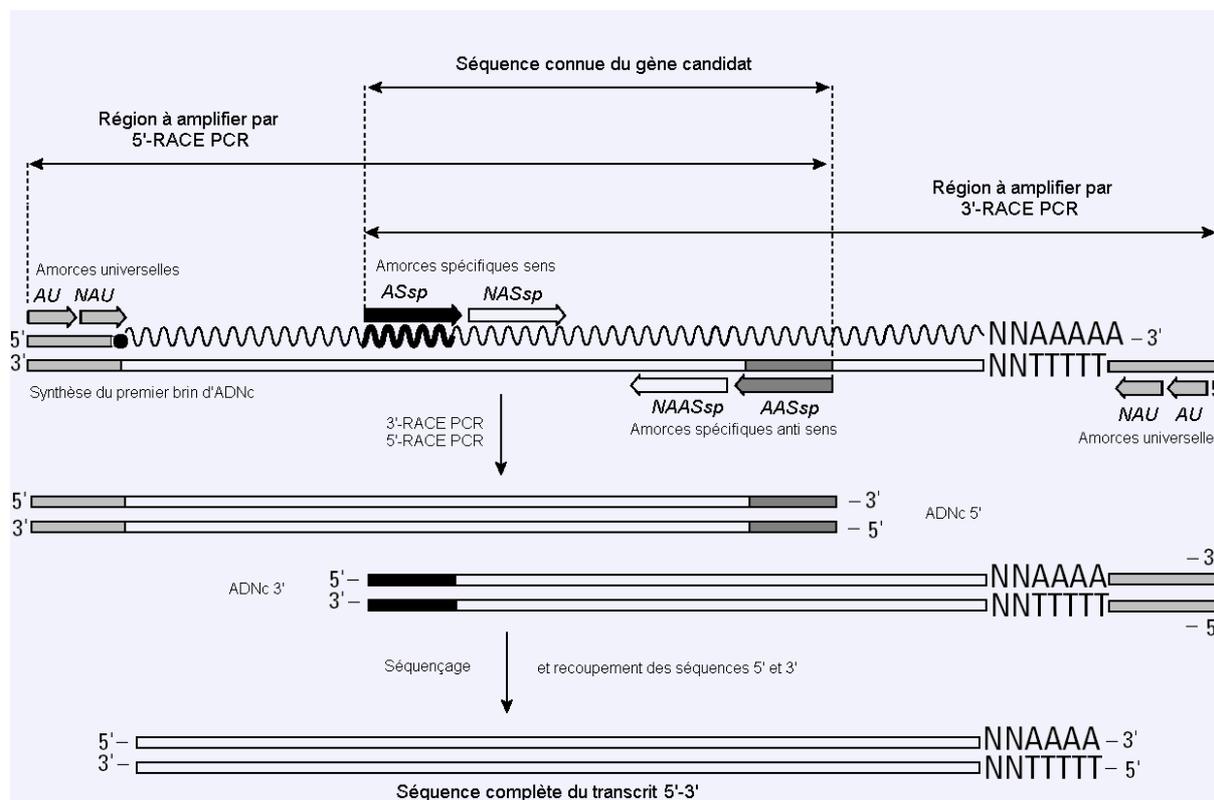


Figure 19 : Placement des couples d'amorces spécifiques et universelles utilisées en RACE PCR et Nested PCR (d'après Clontech).

(N)ASsp : (Nested) Amorce Sens Spécifique. (N)AASsp : (Nested) Amorce Antisens Spécifique. AU : Amorce Universelle. NAU : Nested Amorce Universelle.

L'amplification est réalisée avec les réactifs du kit PCR AdvantageTM de Clontech. La réaction PCR est réalisée dans un volume total de 50 μ L. Les ADNc précédemment synthétisés sont, à raison de 2,5 μ L, mélangés avec les couples appropriés d'amorces sens et anti sens (5 μ L du mélange d'amorces universelles 10X, et 2 μ L d'amorce spécifique 5 μ M), 1 μ L de dNTP 10 mM, 5 μ L de tampon de réaction 10X, et 1 μ L du mélange optimisé « Advantage 2 Polymerase » 50X. Le volume final est ajusté à 50 μ L avec de l'eau exempte de nucléases fournie par le fabricant. Le programme d'amplification se compose de 35 cycles comprenant chacun une étape de 30 s de dénaturation à 94 °C, une étape de 30 s d'hybridation à 68 °C et une étape de 3 min d'élongation à 72 °C. Une étape de dénaturation de 5 min à 94 °C précède le premier cycle, et une étape finale d'élongation de 10 min à 72 °C suit le dernier cycle.

6.3. Nested-PCR

La méthode dite de nested PCR est employée pour obtenir une amplification plus importante et plus spécifique que celle obtenue avec une PCR classique. La nested-PCR correspond à deux amplifications successives réalisées avec deux couples d'amorces différents. La première amplification (RACE-PCR) a été réalisée avec une amorce universelle et une amorce spécifique localisée à l'extrémité de la région connue du gène candidat. Les bandes majeures obtenues ont été découpées, extraites du gel d'agarose, purifiées puis utilisées comme matrice pour la seconde PCR. Les amorces spécifiques (*NAS_{sp}*, *NAAS_{sp}*) et universelles (*NAU*) utilisées pour cette seconde PCR sont internes par rapport aux précédentes, c'est-à-dire qu'elles vont s'hybrider à l'intérieur de la séquence résultant de la première amplification (Figure 19). Les conditions d'amplification appliquées en nested PCR sont identiques à celles de la RACE-PCR.

6.4. Conception des amorces

6.4.1. Amorces dégénérées

Dans le cadre de la recherche de sondes spécifiques, la définition d'amorces dégénérées a été entreprise. Pour cela des séquences nucléotidiques de gènes codant des phospholipases A₂ calcium indépendantes associées aux membranes (iPLA₂) ont été importées des banques de données de *Thalassiosira pseudonana* et *Cyanidioschyzon merolae*. L'alignement des séquences entre elles et avec d'autres iPLA₂ a révélé la présence de 3 motifs relativement bien conservés (Figure 20).

```

TpiPLA2      GQRGLRILSLDGGGTRGIAAVTSIRHIVEAMGGVEVCDADFMDIVGTSTGAI VAFVLVGLRR
HsiPLA2      KGRGIRILSIDGGGTRGVVALQTLRKLVE-LTQKPVHQLFDYICGVSTGAILAFMLGLFH
CmiPLA2      GQRGLRILALDGGGTRALMSFEILKHLTK-LTGCQLHEMFDIICGTSTGAI IAGSLGIRR
              **:***::*****.: :. :.:.: : : : * * * *.*****:* :*: .

TpiPLA2      ESAADARIRYDTLIKRI FVKS-LLKPIMLATTTASYDEANLMDVLQEILKDDGMLDSRAN
HsiPLA2      MPLDECEELYRKLGSDFVSNVIVGTVKMSWSHAFYDSQ TWENILKDRMG S ALMIETARN
CmiPLA2      RPVEEVESLYRELIGKIFAKK-LSSAPKMLLTRAYYD TDLFESILKREAGSLRMIDSTMD
              . : * * :* :. : . : : * ** .:*: . *::: :

TpiPLA2      PEVPLITAVSSKMSS--TPSQLCLLRNYNYGGGELNDSFCSFRVTQKIALRATTAAPTFF
HsiPLA2      PTCPKVAAVSTIVNRGITPKAF-VFRNYGHFPGINSHYLG GCQYKMWQAIRASSAAPGYF
CmiPLA2      RDMNYVFFVSSVMNR--RPHQLHLFRNYCHAPGQESRYPGTV DATLWQGM RASSAAP TFF
              : ** :. * : :*** : * . . . :***:*** :*

TpiPLA2      KPLLSFEELYVDGGIVASNPTAVAVHEARSVFPVPLELIVSVGTG
HsiPLA2      AEYALGNDLHQDGLLLNNPSALAMHECKCLWPDVPLECIVSLGTG
CmiPLA2      SEIVLNGLIHADGALVANNPAGVAAHEARRLFPNVP IELLVSVGTG
              :: **.: : .**.:* **.. :*.***: **:*

```

Figure 20 : Aligement des iPLA₂ hypothétiques des microalgues *Cyanidoschyzon merolae* (*CmiPLA2*) et *Thalassiosira pseudonana* (*TpiPLA2*) avec leur homologue la iPLA₂ humaine (*HsiPLA2*).

Organisme / numéro d'accèsion : *Cyanidoschyzon merolae* / CMR500C, *Thalassiosira pseudonana* / 98946 (version1) et *Homo sapiens* / BC032999. Les résidus strictement conservés sont indiqués par des étoiles, les résidus bien conservés par des points. Les 3 motifs conservés utilisés pour la conception des amorces nucléotidiques sont surlignés en gris.

Dans ces expérimentations, trois motifs, relativement bien conservés au cours de l'évolution chez les iPLA₂, ont été choisis pour la conception des amorces. Ce sont, de l'extrémité N- à l'extrémité C-terminale, les motifs D-G-G-G-T-R-A, G-T-S-T-G-A-I et R-A-S-S-A-A-P, respectivement nommés PLM1, PLM2 et PLM3. Le second motif, PLM2, inclut le pentapeptide consensus des enzymes lipolytiques avec la sérine catalytique. Les 2 motifs PLM1 et PLM3 n'entrent pas dans la constitution du site actif et leur fonction est inconnue. Pour s'assurer de leur spécificité phospholipasique, les motifs ont été alignés, séparément, avec Blastp dans GenBank. Les protéines homologues obtenues sont en majorité des iPLA₂.

Les amorces dégénérées correspondant aux 3 motifs sont définies sur la base du code génétique. Le degré de dégénérescence doit être le plus faible possible. L'extrémité 3' doit être la plus spécifique possible pour permettre un bon début de synthèse par la polymérase. Mais les contraintes imposées par le faible taux de conservation des enzymes et par la petite taille des motifs conservés n'ont pas toujours permis de respecter au mieux ces critères (Figure 21). Les amorces fortement dégénérées résultent des séquences nucléotidiques *TpiPLA2* et *CmiPLA2* et combinent toutes les possibilités nucléotidiques liées à la dégénérescence du code génétique. Les amorces faiblement dégénérées sont calquées sur les séquences *TpiPLA2* et *CmiPLA2* et ne tiennent pas compte de la dégénérescence du code génétique.

TpiPLA2 GQRGLRILSLDGGGTRGIAAVTSIRH
 HsiPLA2 KRGIRILSIDGGGTRGVVALQTLRK
 CmiPLA2 GQRGLRILALDGGGTRALMSFEILKH
 :*:*****: :. :. :

TpiPLA2 IVEAMGGVEVCDAFDMIVGTSTGAIIVAFVLGLRR
 HsiPLA2 LVE-LTQKPVHQLFDYICGVSTGAILAFMLGLFH
 CmiPLA2 LTK-LTGCQLHEMFDIICGTSTGAI IAGSLGIRR
 :.: : : : * * * * .*****:* :*: .

Acides amines conservés	D Asp	G Gly	G Gly	G Gly	T Thr	R Arg
<i>C. merolae</i>	GAC	GGA	GGT	GGC	ACT	CGG
<i>T. pseudonana</i>	GAT	GGG	GGC	GGA	ACG	AGA
Possibilités nucléotidiques de 5' vers 3'	GAT C	GGT G C A	GGT G C A	GGT G C A	ACT G C A	CGT C A G A
Amorce fortement dégénérée <i>PLM1dS</i>	GAY	GGN	GGN	GGN	ACN	MG
Amorce faiblement dégénérée	GAY	GGR	GGY	GGM	ACK	MG

Acides amines conservés	G Gly	T Thr	S Ser	T Thr	G Gly	A Ala	I Ile
<i>C. merolae</i>	GGA	ACC	TCG	ACG	GGG	GCG	ATT
<i>T. pseudonana</i>	GGA	ACG	TCA	ACA	GGG	GCA	ATT
Possibilités nucléotidiques de 5' vers 3'	GGT G C A	ACT G C A	TCA G AGT C	ACT G C A	GGT G C A	GCT G C A	ATT C A A
Amorce fortement dégénérée <i>PLM2dS</i>	GGN	ACN	TCN	ACN	GGN	GCN	AT
Amorce faiblement dégénérée	GGA	ACS	TCR	ACR	GGG	GCR	ATT

TpiPLA2 RNYNYGGGELNDSFCSFRVTQKIALRATTAAPTFF
 HsiPLA2 RNYGHFPGINSHYLGGCQYKMWQAIRASSAAPGYF
 CmiPLA2 RNYCHAPGQESRYPGTVDATLWQGM RASSAAPTFF
 *** : * . . . :***:*** :*

Acides amines conservés	R Arg	A Ala	S/T Ser/Thr	S/T Ser/Thr	A Ala	A Ala	P Pro	
<i>C. merolae</i>	CGG	GCC	TCG	TCG	GCG	GCT	CCG	
<i>T. pseudonana</i>	AGG	GCA	ACA	ACT	GCC	GCA	CCA	
Possibilités nucléotidiques de 5' vers 3'	CGT C A G A	GCT G C A	TCA G AGT C	ACT G C A	TCA G AGT C	ACT G C A	GCT G C A	CCT G C A
Amorce fortement dégénérée <i>PLM3dS</i>	MGR	GCN	TCN	TCN	GCN	GCN	CC	
Amorce faiblement dégénérée	MGG	GCM	WCR	WCK	GCS	GCW	CC	

Figure 21 : Conception des amorces sens plus ou moins dégénérées calquées sur les motifs conservés chez les iPLA₂ hypothétiques des microalgues *Cyanidoschyzon merolae* et *Thalassiosira pseudonana*.

6.4.2. Amorces non dégénérées et non spécifiques d'*Isochrysis galbana*

Des amorces ont également été définies à partir de séquences similaires à une iPLA₂ (CMR500C) et à une lipase (CMP157C) chez la microalgue rouge *C. merolae*. Les motifs conservés ont servi de base pour concevoir les amorces. Concernant la lipase hypothétique CMP157C, c'est le motif incluant la potentielle sérine catalytique G-H-S-L-G-A-A qui a servi de modèle pour concevoir l'amorce *CmLipS*. Ces amorces ont été utilisées dans des expérimentations d'amplifications pour la recherche de sondes homologues chez *I. galbana*. Certaines d'entre elles ont aussi été utilisées en couple avec les amorces universelles calquées sur les promoteurs bactériens T3 et T7 pour le criblage de la banque d'ADNc en vue de l'isolement d'un gène de PLA₂ et de lipase.

Amorces	Séquences 5'-3'	Taille en pb	Tm en °C
T3S	CGAAATTAACCCTCACTAAAGGGAACAAA	29	63, 21
T7AS	ATTGTAATACGACTCACTATAGGGCGAA	28	63, 15
<i>CmPLM1S</i>	GACGGAGGTGGCACTCGGGCG	21	72, 33
<i>CmPLM2S</i>	GGAACCTCGACGGGGGCGATT	21	68, 42
<i>CmPLM2AS</i>	AATCGCCCCGTCGAGGTTC	21	68, 42
<i>CmPLM3S</i>	CGGGCCTCGTCGGCGGCTCCG	21	76, 23
<i>CmPLM3AS</i>	CGGAGCCGCGACGAGGCCCG	21	76, 23
<i>CmLipS</i>	GGACATTGCTGGGTGCCGCG	21	70, 38
<i>CmLipAS</i>	CGCGCACCCAGCGAATGTCC	21	70, 38

Tableau 9 : Séquences des amorces T3 et T7 et des amorces non dégénérées spécifiques de la microalgue *Cyanidoschyzon merolae*.

6.4.3. Amorces spécifiques d'*Isochrysis galbana*

Dans le cadre de l'isolement des gènes par RACE PCR, des amorces spécifiques définies à partir des marqueurs de séquences exprimées référencés dans la banque EST d'*I. galbana* (Keeling, 2006) ont été utilisées pour l'amplification des marqueurs, en RACE et nested PCR et pour l'amplification des ADNc pleine-longueur. La liste exhaustive des amorces utilisées figure dans le Tableau 10.

La synthèse des amorces a été réalisée par Opéron Biotechnologies (Cologne, Allemagne).

Couples d'amorces	Séquences 5'-3'	Taille en pb	Taille des amplicons en pb	Tm en °C
ISL00001844 / IgLip				
RT-PCR Amplification du cluster ISL00001844				
Amorce sens <i>LipS</i>	ATGGCTGGGGTAGTGGCTACGACAT	25	461	67,86
Amorce anti sens <i>LipAS</i>	ATCGCGTGCATTGACCACACGTAGCAT	27		67,64
3'-RACE PCR				
Amorce sens <i>LipS</i>	ATGGCTGGGGTAGTGGCTACGACAT	25	1342	67,86
Mixe d'amorces universelles (Clontech)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	45		
	CTAATACGACTCACTATAGGGC	22		
5'-RACE PCR				
Amorce anti sens <i>LipAS</i>	ATCGCGTGCATTGACCACACGTAGCAT	27	1309	67,64
Mixe d'amorces universelles (Clontech)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	45		
	CTAATACGACTCACTATAGGGC	22		
3'-Nested PCR				
Amorce sens <i>NlipS</i>	CTCAATCATCCTCGGTAGCCAGCGAATCGT	30	1160	70,1
Amorce universelle nested (Clontech)	AAGCAGTGGTATCAACGCAGAGT	23		
5'-Nested PCR				
Amorce anti sens <i>NlipAS</i>	AAACACTCTCCAGGCGATGGAGCTGGA	28	1130	70,47
Amorce universelle nested (Clontech)	AAGCAGTGGTATCAACGCAGAGT	23		
RT-PCR pleine longueur				
Amorce sens <i>EcoRIIip</i>	GGAATTCCATGCGCAGGCGGCGGGCGAAAGAGG	35	1392	78,73
Amorce anti sens <i>LipXbaI</i>	GCTCTAGAGCTCATCCGTACGAGTATGTCCGGCACCCG	37		75,76
ISL00008005 / IgEst1				
RT-PCR Amplification du cluster ISL00008005				
Amorce sens <i>Est1S</i>	ATCCACGGGCTGGGCGACAGCAACATG	27	407	72,2
Amorce anti sens <i>Est1AS</i>	TTCACAGCCTCCGGCGCCAGTATGAAA	27		69,16
3'-RACE PCR				
Amorce sens <i>Est1S</i>	ATCCACGGGCTGGGCGACAGCAACATG	27	964	72,2
Mixe d'amorces universelles (Clontech)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	45		
	CTAATACGACTCACTATAGGGC	22		
5'-RACE PCR				
Amorce anti sens <i>Est1AS</i>	TTCACAGCCTCCGGCGCCAGTATGAAA	27	696	69,16

Mixte d'amorces universelles (Clontech)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGHCAGAGT	45		
	CTAATACGACTCACTATAGGGC	22		
3'-Nested PCR				
Amorce sens <i>NEst1S</i>	AACGGCGGAATGTCGATGCCAGTTGGTATGA	32	825	71,06
Amorce universelle nested (Clontech)	AAGCAGTGGTATCAACGCAGAGT	23		
5'-Nested PCR				
Amorce anti sens <i>NEst1AS</i>	AGGGTGTGTGAATACTGCAGCCCCTGTGAAGA	32	602	71,06
Amorce universelle nested (Clontech)	AAGCAGTGGTATCAACGCAGAGT	23		
RT-PCR pleine longueur				
Amorce sens <i>EcoRIEst1</i>	GGAATTCCATGGCGCGAGGGCTTGCCCGGTGCA	33	846	77,08
Amorce anti sens <i>Est1XbaI</i>	GCTCTAGAGCTCAGGCGTCTCTGGCAGTCTGGCCT	36		77,21
ISL00001076 / IgEst2				
RT-PCR Amplification du cluster ISL00001076				
Amorce sens <i>Est2S</i>	CGTCGTCAGATGCTGTACTTGCATTTATTG	30	471	66
Amorce anti sens <i>Est2AS</i>	GGAACCTCCTACTGATCACCCCTTGATGCTC	30		68,73
3'-RACE PCR				
Amorce sens <i>Est2S</i>	CGTCGTCAGATGCTGTACTTGCATTTATTG	30	828	66
Mixte d'amorces universelles (Clontech)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	45		
	CTAATACGACTCACTATAGGGC	22		
5'-RACE PCR				
Amorce anti sens <i>Est2AS</i>	GGAACCTCCTACTGATCACCCCTTGATGCTC	30	830	68,73
Mixte d'amorces universelles (Clontech)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	45		
	CTAATACGACTCACTATAGGGC	22		
3'-Nested PCR				
Amorce sens <i>NEst2S</i>	GAGCTAATGCAGGCTGGTACCCCACTTTAT	30	612	68,73
Amorce universelle nested (Clontech)	AAGCAGTGGTATCAACGCAGAGT	23		
5'-Nested PCR				
Amorce sens <i>NEst2AS</i>	TCCTCCTTATATATTAAGTCCGCACCCGGT	30	704	67,37
Amorce universelle nested (Clontech)	AAGCAGTGGTATCAACGCAGAGT	23		
RT-PCR pleine longueur				
Amorce sens <i>EcoRIEst2</i>	GGAATTCCATGAGGTACGCCGTGCGGGCCGCCACTA	36	723	77,21
Amorce anti sens <i>Est2XbaI</i>	GCTCTAGAGCCTACTGATCACCCCTTGATGCTCCGGA	37		73,55

Tableau 10 : Liste des couples d'amorces spécifiques d'*Isochrysis galbana*.

6.5. Séquençage des produits amplifiés

Les produits amplifiés ont été purifiés avec le kit d'extraction WIZARD® SV Gel and PCR Clean-Up System (Promega). Ce kit est adapté pour des fragments de taille comprise entre 100 pb et 10 kb. Les produits amplifiés sont purifiés soit directement après la PCR soit après électrophorèse sur gel d'agarose. Dans ce dernier cas, la zone du gel, contenant les fragments d'ADN d'intérêt, est excisée à l'aide de lames de rasoir stériles. L'ADN est ensuite extrait du gel d'agarose. Il s'agit de faire fondre l'agarose dans une solution d'isothiocyanate de guanidine 4,5 M et d'acétate de potassium 0,5 M pH 5, dont la force ionique permet la liaison de l'ADN sur une matrice de silice. Après lavages, l'ADN est élué des colonnes par 50 µL d'eau ultrapure et stocké à -20 °C. Un aliquote de cet ADN purifié est analysé par électrophorèse.

Les produits d'amplification purifiés ont ensuite été dosés par fluorimétrie (Fluorimètre Versa Fluor™, Bio Rad). Le fluorochrome PicoGreen® (λ_{exc} : 485 nm λ_{em} : 535 nm), commercialisé par Invitrogen a été choisi pour le dosage. Il assure un dosage environ 10 000 fois plus sensible qu'un dosage effectué par spectrophotométrie à 260 nm. Il permet de détecter jusqu'à 25 pg/mL d'ADN en solution. Les produits PCR sont quantifiés en moyennant les résultats de deux dosages fluorimétriques réalisés avec deux dilutions différentes. Le séquençage des fragments d'ADN a été réalisé par la société MilleGen (Labège, France). Les résultats sont obtenus sous 24 h sous format FASTA et sous forme de chromatogramme.

7. Construction et criblage d'une banque d'ADN complémentaires

7.1. Construction d'une banque phagique d'ADNc

La banque d'ADNc a été construite à l'aide des Kits Zap express® cDNA Synthesis et Zap express® cDNA Gigapack® III Gold Cloning de Stratagène (La Jolla, Etats Unis) selon les recommandations du fabricant. Le vecteur Zap express est un bactériophage λ .

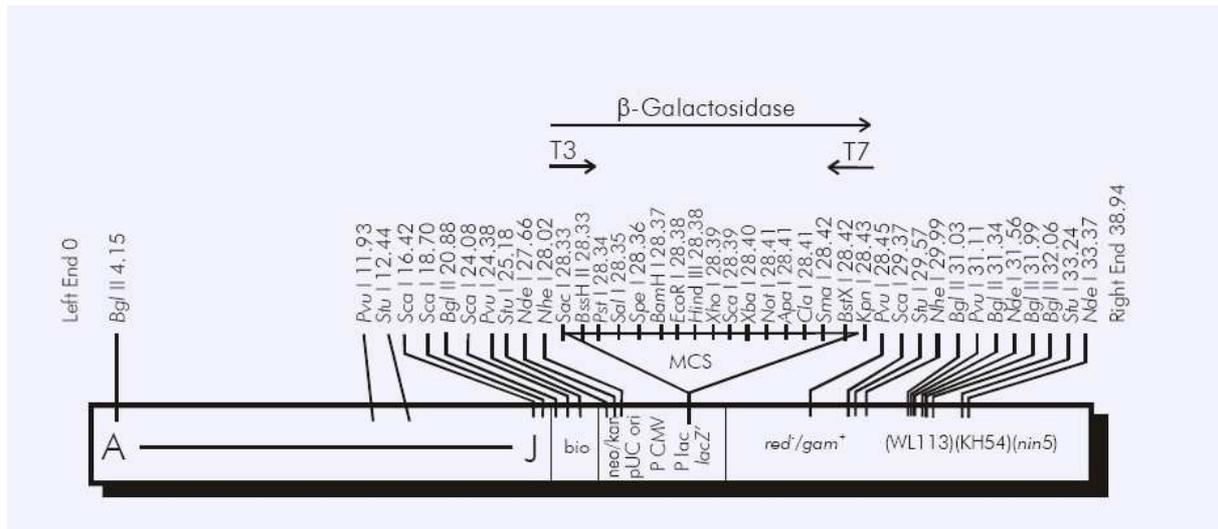


Figure 22 : Le vecteur Zap express (Stratagène).

La construction de la banque a nécessité de nombreuses étapes :

- Extraction des ARN totaux à partir de la biomasse microalgale d'une culture de 1,5 L en photobioréacteur.
- Purification des ARN polyadénylés.
- Copie de 5 µg d'ARNm en ADN simple brin complémentaire par l'action d'une transcriptase inverse (Annexe 4).
- Elimination spécifique des ARNm par la RNase H (Annexe 4).
- Synthèse du second brin d'ADNc par une ADN polymérase (Annexe 4)
- Ligation et digestion d'adaptateurs contenant les sites de restriction *EcoRI* et *XhoI* (Annexe 4).
- Fractionnement des ADNc.

Le protocole préconisé par le fabricant impliquait l'utilisation de dNTP[α - 32 P] lors de la synthèse du second brin d'ADNc. Le laboratoire n'étant pas habilité à manipuler des isotopes radioactifs, le fractionnement des ADNc a été réalisé par électrophorèse sur gel d'agarose. Pour cela, les ADNc fraîchement digérés ont été solubilisés dans 10 µL d'eau ultrapure stérile. La migration électrophorétique a été effectuée sur un gel d'agarose à 1 % (w/v). Le BET et les UV ayant un fort pouvoir mutagène, les ADNc (8 µL) destinés au fractionnement ont migré en parallèle d'ADNc (1 µL) colorés au BET, soumis aux UV et utilisés comme repère visuel pour l'excision. Trois zones correspondant à des ADNc de tailles différentes ont été découpées à l'aide de lames stériles. Les ADNc ont ensuite été extraits du gel d'agarose et purifiés avec le kit d'extraction WIZARD® SV Gel and PCR Clean-Up System (Promega). Après élution dans 50 µL d'eau ultrapure, les 3 fractions ont

été précipitées par ajout de 5 μ L d'acétate de sodium 3M (pH 5,2) et de 165 μ L d'éthanol 100 %, incubées une nuit à -20 °C. Les précipités récoltés par centrifugation (10000 x g, 4 °C, 30 min) ont été dissous dans 5 μ L d'eau ultrapure stérile.

- Ligation unidirectionnelle dans le vecteur de clonage Zap express (ADN du bactériophage λ), fourni prédigéré avec les enzymes *EcoRI* et *XhoI*.
- Incubation du vecteur en présence des protéines de la capsid du bactériophage λ .
- Infection de la souche *E. coli* XL1-Blue MRF' par les phages reconstitués et titration de la banque primaire. La banque a donc été divisée en 3 sous-groupes dont la conception repose sur le fractionnement des ADNc selon leur taille : ADNc \leq 564 pb (banque 1), 564 pb \leq ADNc \leq 1375 pb (banque 2), 1375 pb \leq ADNc \leq 4268 pb (banque 3).
- Amplification des banques d'ADNc primaires pour accroître et stabiliser les quantités stocks.

7.2. Criblage de la banque d'ADNc

Sur une boîte de Petri, les phages sont multipliés sur un tapis bactérien de manière à obtenir un maximum de plages de lyse isolées. Les phages sont ensuite récoltés puis lysés pour en extraire l'ADN de la façon décrite au §4.4. L'ADN phagique est ensuite criblé par amplification PCR avec les amorces appropriées.

8. Clonage et expression des séquences codantes

8.1. Le vecteur de clonage et d'expression : le plasmide pYES2

Le plasmide pYES2 (5,9 kb) est fourni par Invitrogen (Figure 23). Il est utilisé pour les expériences de clonage dans la souche TOP10F' d'*Escherichia coli* et permet la synthèse de protéines recombinantes dans la souche INVSc1 de la levure *Saccharomyces cerevisiae*. Il possède :

- Un site de clonage multiple situé en aval du promoteur.
- Le promoteur GAL1 inductible par le galactose qui contrôle l'expression du gène inséré.
- Un signal de terminaison de transcription CYC1 nécessaire à la stabilisation des ARNm.

- Un gène URA3 codant pour une enzyme impliquée dans la synthèse d'uracile pour la sélection des levures transformées.
- Un gène de résistance à l'ampicilline pour la sélection des transformants dans *E. coli*.
- Plusieurs origines de réplication pour maintenir le nombre de copies plasmidiques à un haut niveau. pUC ori est l'origine de réplication bactérienne et 2μ ori l'origine de réplication de levure.

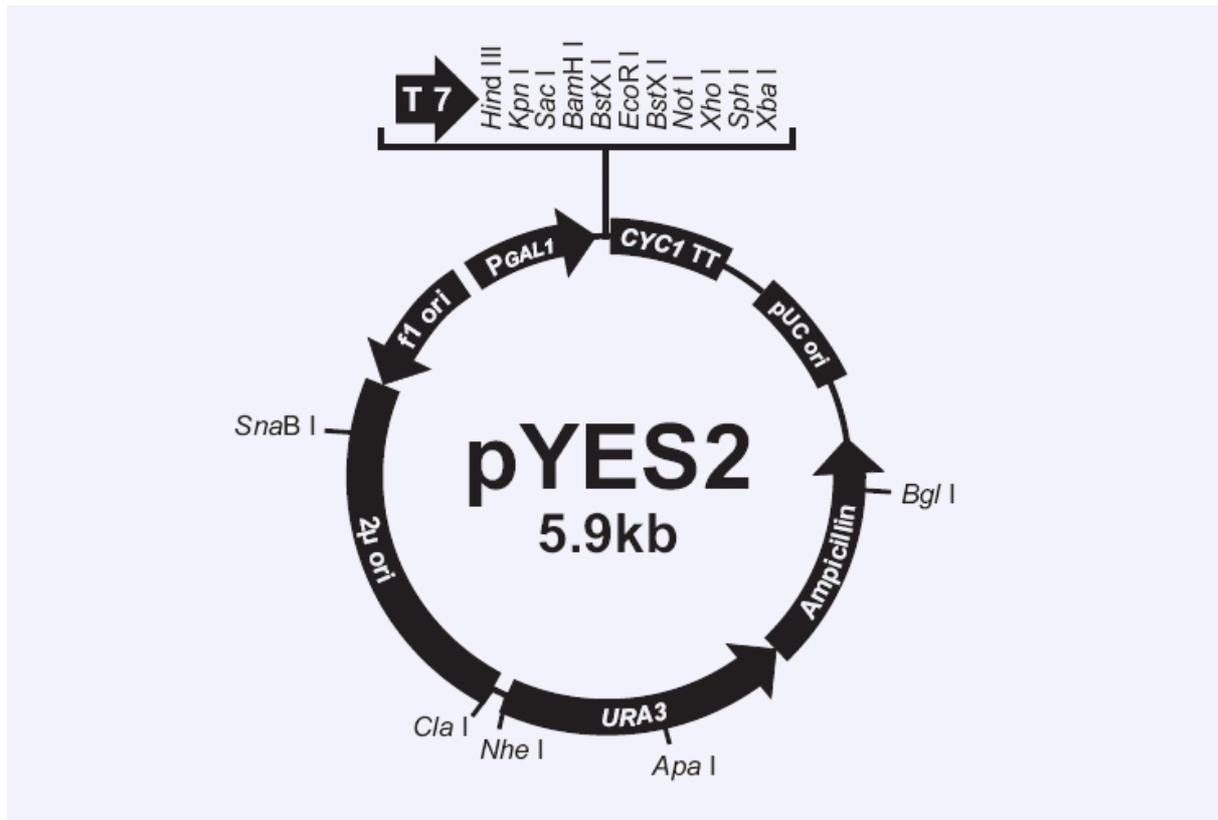


Figure 23 : Le vecteur d'expression pYES2 (Invitrogen).

8.2. L'insert (ADNc)

Les séquences nucléotidiques contenant les parties codantes des transcrits d'intérêt ont été amplifiées par PCR. Les amorces utilisées possèdent la même température d'hybridation (70 °C) et incorporent un site de restriction *EcoR* I ou *Xba* I. Les conditions d'amplification se composent d'une dénaturation initiale de 5 min à 95 °C, suivie de 35 cycles de 30 s à 95 °C, 30 s à 70 °C, de 3 ou 4 min, selon la taille du gène amplifié, à 72 °C, puis d'une étape finale d'élongation de 10 min à 72 °C.

8.3. Digestion enzymatique du plasmide et de l'insert (ADNc)

Le plasmide et l'insert sont digérés dans un volume final de 20 µL avec 2 µL de tampon de réaction 10X (Tris-HCl 20 mM, NaCl 100 mM, EDTA 0,1 mM, 2-mercaptoethanol 10 mM, Triton X-100 0,01 % (v/v), glycérol 50 % (v/v)) et avec 10 unités d'enzymes de restriction *EcoR* I et *Xba* I. Le milieu réactionnel est incubé 1 h à 37 °C. Les enzymes sont inactivées par 20 min d'incubation à 65 °C.

8.4. Ligation du plasmide et de l'insert

La ligation de l'insert avec le vecteur est réalisée avec la T4 ADN ligase (Promega). Le rapport molaire utilisé entre le plasmide et l'insert peut varier de 1/1 à 1/3. La réaction est effectuée dans un volume final de 10 µL en présence du tampon de ligation (1X final : Tris-HCl 30 mM (pH 7,8), MgCl₂ 10 mM, ATP 1 mM, DTT 10 mM). A 100 ng du vecteur pYES2 linéarisé et 100 ng d'insert, ont été ajoutées 3 unités de T4 ADN ligase. Le milieu réactionnel est incubé une nuit à 16 °C.

8.5. Préparation de bactéries et de levures électrocompétentes

Les bactéries sont mises en culture pendant une nuit à 37 °C, sous agitation, dans 5 mL de milieu LB (1 % NaCl, 1 % de tryptone, 0,5 % d'extrait de levure). Cette préculture est utilisée pour ensemercer 250 mL de milieu LB. Lorsque l'absorbance à 600 nm atteint une valeur comprise entre 0,6 et 0,8, la culture est centrifugée 15 min à 2600 x g et à 4 °C. Les culots de cellules sont lavés 3 fois (pour éliminer les sels) avec de l'eau ultrapure stérile. Les derniers lavages sont réalisés dans du glycérol à 10 %, préalablement refroidi dans la glace. Le culot cellulaire est finalement repris dans du glycérol 10 % (Sambrook *et al.*, 1989).

Les levures sont cultivées à 30 °C dans 250 mL de milieu YPD (1 % d'extrait de levure, 2 % de peptone, 2 % de dextrose), jusqu'à ce que l'absorbance à 600 nm atteigne une valeur comprise entre 0,6 et 0,8. Les cellules récoltées par centrifugation sont lavées à l'eau ultrapure stérile. Les cellules sont ensuite incubées 15 min dans 500 µL de dithiothréitol 25 mM (Suga et Hatakeyama, 2001), centrifugées puis resuspendues dans du sorbitol 1 M.

Les bactéries et les levures électrocompétentes ainsi préparées sont aliquotées par volume de 50 µL, congelées dans l'azote liquide et conservées à -70 °C.

8.6. Transformation par électroporation

A 50 μ L de bactéries ou de levures électrocompétentes sont ajoutés 1 à 5 μ L du produit de ligation. L'ensemble est placé dans une cuvette de 2 mm (préalablement refroidie dans la glace). La cuvette est ensuite placée dans l'appareil d'électroporation (Electroporator 2510, Eppendorf). Les conditions de voltage pour l'électroporation sont de 25 kV/cm pour les bactéries et de 15 kV/cm pour les levures. Après l'impulsion (la constante de temps τ doit être inférieure à 5 ms), les bactéries sont immédiatement reprises dans 1 mL de milieu SOC (bacto-tryptone 2 %, extrait de levure 0,5 %, NaCl 8,5 mM, KCl 2,5 mM, MgCl₂ 10 mM, glucose 20 mM) et placées à 37 °C pendant 1 h. Après une centrifugation d'1 min à 2500 x g, le culot bactérien est repris dans 150 μ L de milieu LB. Ensuite, la suspension de bactéries électroporées est étalée sur milieu solide sélectif.

8.7. Sélection des transformants

Les bactéries Amp^r transformées ont acquis le gène de résistance à l'ampicilline (Sambrook *et al.*, 1989). Elles sont sélectionnées sur boîte de LB agar (2 % w/v) à 50 μ g/mL d'ampicilline. Les levures transformées sont sélectionnées sur milieu SC-U dont voici la composition : 0,67 % (w/v) d'une base de levure azotée sans acides aminés, 2 % (w/v) d'une source de carbone (glucose ou galactose), 0,01 % (w/v) d'adénine, d'arginine, de cystéine, de leucine, de lysine, de thréonine, de tryptophane ; 0,005 % (w/v) d'acide aspartique, d'histidine, d'isoleucine, de méthionine, de phénylalanine, de proline, de sérine, de tyrosine, et de valine, plus ou moins 2 % (w/v) d'agar. L'absence d'uracile (U) est ici le facteur sélectif des levures transformées. La souche INVSc1 utilisée est auxotrophe pour l'uracile. Les levures transformées avec le plasmide pYES2 se montreront prototrophes à l'égard de l'uracile et seront à même de se développer sur un milieu déficient en uracile.

8.8. Analyse des clones positifs

La présence de l'insert est mise en évidence par digestion enzymatique et/ou PCR. Pour chacune des constructions génétiques, plusieurs clones bactériens ont été séquencés avant de réaliser l'expression du gène.

8.9. Induction de l'expression du gène cloné et analyse des transcrits

Les gènes clonés sont exprimés chez la levure *Saccharomyces cerevisiae* INVSc1. La transcription du gène est sous le contrôle du promoteur GAL1, inductible par le galactose et réprimé en présence de glucose. La présence ou l'absence du transcrit ont été contrôlées pour les levures transformées avec les inserts *IgEst1* et *IgEst2*. Pour cela, trois cultures ont été incubées en parallèle, à 30 °C dans un milieu SC :

- Une culture de levures transformées avec le plasmide non recombiné dans un milieu SC-U, servant de témoin négatif
- Une culture de levures transformées avec le plasmide recombiné dans un milieu SC-U en présence de glucose
- Une culture de levures transformées avec le plasmide recombiné dans un milieu SC-U en présence de galactose.

Lorsque la croissance atteint la phase stationnaire (densité stable), les cellules sont récoltées par centrifugation, broyées dans l'azote liquide en une fine poudre à partir de laquelle sont extraits les ARN totaux. S'agissant de déterminer la présence ou l'absence du transcrit, il est impératif d'éliminer toute trace d'ADN plasmidique des préparations d'ARN. Pour détecter la présence d'ADN, les ARN totaux traités par une DNase ont été soumis à deux types de vérification : une analyse qualitative par électrophorèse sur gel d'agarose et une PCR effectuée directement sur l'ARN (témoin négatif). La présence des transcrits a ensuite été déterminée par PCR avec des couples d'amorces spécifiques des séquences codantes insérées. Les conditions d'amplification se composent d'une dénaturation initiale de 5 minutes à 94 °C, suivie de 35 cycles de 30 s à 95 °C, 30 s à la température d'hybridation appropriée, 3 min à 72 °C, puis d'une étape finale d'élongation de 10 minutes à 72°C.

9. Analyses bioinformatiques des séquences

Les bases de données biologiques consultées et les outils bioinformatiques utilisés pour ces travaux de thèse sont listés, par ordre alphabétique, en Annexe 5, avec leurs adresses Internet respectives.

9.1. Bases de séquences

9.1.1. Les bases de séquences nucléiques et protéiques généralistes

Les bases de séquences publiques généralistes ont été interrogées pour appréhender la fonction des protéines codées par les 3 séquences nouvellement obtenues chez *I. galbana*. Les bases de séquences nucléiques sont au nombre de 3 : DDBJ (DNA Data Bank of Japan) (Tateno *et al.*, 2000), GenBank aux USA (Benson *et al.*, 2007) et EMBL (European Molecular Biology Laboratory nucleotide sequence database) (Kulikova *et al.*, 2007). Les 3 bases sont réunies au sein de l'International Nucleotide Sequence Database Collaboration. Interconnectées, elles échangent leurs données quotidiennement. Les nouvelles séquences ont été déposées sur EMBL à l'aide du formulaire interactif Webin. La base SWISS-PROT est une base de données protéiques annotées avec un faible niveau de redondance. SWISS-PROT a été créée en 1986 par Amos Bairoch au Département de Biochimie Médicale de l'Université de Genève (Bairoch *et al.*, 2004).

9.1.2. Les bases de séquences de microalgues

La base de données de type EST d'*I. galbana* est rattachée à une base de données globale, la TBestDB (Taxonomically Broad EST Database), qui centralise et organise les marqueurs de séquences exprimées d'eucaryotes unicellulaires. Ces informations de nature nucléotidique sont issues d'un programme de séquençage partiel d'ADNc. Ce projet de séquençage découle, entre autres, de la collaboration de six universités canadiennes, dont l'objectif est d'explorer la diversité des gènes exprimés chez les protistes unicellulaires. Plusieurs espèces de microalgues figurent parmi les organismes étudiés comme le dinoflagellé *Heterocapsa triquetra*, la cryptophyte *Guillardia theta*, ou encore l'haptophyte *Pavlova lutheri*.

Les homologues des séquences d'*I. galbana* ont été recherchés dans 6 banques de données publiques chez les algues unicellulaires. Les génomes complets de 6 microalgues sont connus.

Il s'agit de l'algue rouge *Cyanidioschyzon merolae* (Matsuzaki *et al.*, 2004), des diatomées *Thalassiosira pseudonana* (Armbrust *et al.*, 2004) et *Phaeodactylum tricornutum* (Scala *et al.*, 2002), de la microalgue verte *Chlamydomonas reinhardtii* (Merchant *et al.*, 2007) et des prasinophytes *Ostreococcus tauri* (Derelle *et al.*, 2006) et *Ostreococcus lucimarinus* (Palenik *et al.*, 2007).

9.1.3. Les bases de séquences spécialisées sur les enzymes lipolytiques

Pour classer les enzymes identifiées et déterminer leur famille d'appartenance, la Lipase Engineering Database (LED) (Pleiss *et al.*, 2000) et la Microbial Esterases Lipases Database (MELDB) (Kang *et al.*, 2006) ont été consultées. La LED concentre des données de séquences annotées relatives aux enzymes de la superfamille des hydrolases à repliement α/β . La MELDB est spécialisée dans le stockage d'informations sur les lipases et les estérases de microorganismes.

9.1.4. Les bases de signatures protéiques

Quand les séquences divergent excessivement (moins de 30 % d'identité de séquences), la fonction des protéines peut être appréhendée en recherchant la présence de signatures caractéristiques d'un domaine ou d'une fonction protéique. Ces signatures sont des motifs, des blocs, des profils ou des empreintes. Les motifs sont de courtes régions conservées ne contenant ni insertion ni délétion (gaps). Lorsqu'ils sont définis par des alignements multiples, les motifs protéiques correspondent à des blocs c'est-à-dire des sous séquences de régions conservées présentées sous forme d'alignements multiples sans gaps. Les profils sont des alignements multiples de motifs convertis en tables de fréquence qui reflètent la probabilité pour chaque acide aminé de se trouver à une position particulière du motif, tout en considérant les propriétés mutationnelles des acides aminés. Ces matrices de pondération transcrivent la complexité d'un motif. Les empreintes contiennent un ou plusieurs motifs avec ou sans gaps. Des bases de données recensent ces domaines et signatures. La base PROSITE (Sigrist *et al.*, 2002) répertorie motifs et profils protéiques de sites fonctionnels sous forme de séquences consensus. La base BLOCKS (Henikoff *et al.*, 2000) regroupe les motifs sans gaps, de familles de protéines présentes dans SWISS-PROT, sous forme d'alignements multiples (profils).

9.1.5. Les bases de structures tridimensionnelles

La Protein Data Bank (Bhat *et al.*, 2001) rassemble l'ensemble des structures tridimensionnelles expérimentales et théoriques des macromolécules biologiques. Les structures sont résolues par cristallographie par diffraction des rayons X ou par spectroscopie de Résonance Magnétique Nucléaire. Les modèles moléculaires sont établis à l'aide de logiciels spécifiques (§9.7.)

9.2. Analyse des séquences nucléotidiques

La recherche des phases ouvertes de lecture a été réalisée avec GeneMark (Borodovsky et McIninch, 1993), la traduction des séquences nucléotidiques avec SMS (The Sequence Manipulation Suite) (Stothard, 2000) et la structure intronique des gènes a été déduite à l'aide du programme Spidey (Wheelan *et al.*, 2001). Les cartes de restriction des inserts et des plasmides ont été établies avec les programmes SMS et pDRAW32.

9.3. Profils physicochimiques des protéines

La prédiction des caractéristiques physicochimiques des protéines apporte des informations nécessaires au traitement expérimental des protéines dans les domaines de l'extraction et de la purification. Par exemple, la connaissance du pH, pour lequel la protéine présente une charge nette nulle, est utile pour la séparation d'un mélange de protéines par migration électrophorétique dans un gel contenant des ampholines qui créent un gradient de pH.

La masse moléculaire et le pH du point isoélectrique des protéines ont été déterminés grâce aux programmes hébergés sur le site SMS (Stothard, 2000).

Le programme SignalP 3.0 (Bendtsen *et al.*, 2004) prédit la présence ou non de peptides signaux avec sites de clivage reconnu par la signal peptidase I (SPase) dans les régions N-terminales. Selon le type de données utilisées comme références (eucaryotes, bactéries Gram- et Gram+), le serveur annonce, après un calcul algorithmique de type markov caché, les sites de clivages probables. Les peptides signaux sont de courts segments peptidiques positionnés à l'extrémité N-terminale des protéines destinées à être sécrétées ou exportées dans un compartiment intracellulaire. Ces signaux sont responsables de l'adressage des protéines aux translocons.

Les profils hydrophobes et hydrophiles ont été générés à partir du programme TopPred (Claros et von Heijne, 1994). Un indice d'hydrophobicité (Tableau 11) est attribué à chaque acide aminé d'après une échelle conventionnelle élaborée, en 1982, par Kyte et Doolittle.

Acide aminé		Acide aminé		Acide aminé		Acide aminé	
Ile	+4,5	Met	+1,9	Ser	-0,8	Asp	-3,5
Val	+4,2	Ala	+1,8	Tyr	-1,3	Glu	-3,5
Leu	+3,8	Gly	-0,4	Pro	-1,6	Gln	-3,5
Phe	+2,8	Thr	-0,7	His	-3,2	Lys	-3,9
Cys	+2,5	Trp	-0,9	Asn	-3,5	Arg	-4,5

Tableau 11 : Echelle d'hydrophobicité de Kyte et Doolittle.

Les acides aminés sont classés par ordre décroissant d'indice hydrophobe. Les résidus hydrophiles ont un score < 0 et les résidus hydrophobes un score > 0.

Dans une protéine, la succession d'acides aminés hydrophobes forme des zones d'hydrophobicité qui définissent le profil d'hydrophobicité de la molécule. Ce profil permet de repérer par exemple des régions potentiellement transmembranaires, qui, compte tenu de la nature des membranes biologiques, sont nécessairement hydrophobes. Pour tracer le profil d'hydrophobicité d'une protéine, le score de chaque résidu est pris en compte dans un premier temps. Le profil est ensuite lissé en calculant le score moyen sur une fenêtre de résidus de taille variable, glissant le long de la séquence. Une fenêtre de 5 à 7 résidus est recommandée pour identifier les régions potentiellement exposées en surface. Une fenêtre de 19 à 21 résidus est préconisée pour la recherche de domaines transmembranaires. Si le score moyen dépasse 1,8 dans une zone, alors celle-ci est transmembranaire. Une hélice transmembranaire correspond à une zone transmembranaire de plus de 5 résidus consécutifs.

9.4. Recherche d'homologies

Pour identifier la fonction associée à une nouvelle séquence, la bioinformatique a développé de nombreux outils. La première méthode utilisée consiste à comparer la nouvelle séquence avec l'ensemble des séquences classées dans les bases de données. La seconde méthode permet, par la recherche de motifs caractéristiques, d'identifier une fonction ou une appartenance à une famille de protéines.

9.4.1. Par comparaison de séquences

La recherche de similitudes entre séquences dans les banques de données a été réalisée par le biais d'un BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990). La valeur E calculée par le programme pour chaque alignement permet d'estimer si l'alignement observé a été obtenu par hasard ou s'il a une réelle signification biologique. Plus la valeur E est proche de 0, plus l'alignement est statistiquement significatif.

9.4.2. Par recherche de motifs et de profils

Pour rechercher la présence de motifs protéiques conservés caractéristiques d'une famille de protéines, la méthode de comparaison par profil a été utilisée (Gribskov *et al.*, 1987). Cette méthode s'applique aux motifs définis sous forme d'alignements de séquences. Le programme BLOCKSEARCHER (Fuchs, 1993) recherche sur une séquence protéique l'ensemble des blocs protéiques, convertis en profil et contenus dans la base BLOCK. Les profils présentés dans cette étude sont en weblogo. Ce format traduit la fréquence de position des résidus alignés par la taille des lettres.

9.5. Alignements multiples

Les alignements multiples de séquences protéiques ont été réalisés en utilisant les programmes ClustalW (Thompson *et al.*, 1994) et MUSCLE (Edgar, 2004). L'alignement multiple consiste à aligner plusieurs séquences dans leur intégralité. Les séquences sont comparées deux à deux. ClustalW calcule les scores de similitude et construit une matrice de distances pour l'ensemble des séquences alignées deux à deux. Le programme génère ensuite une séquence consensus à partir de l'alignement multiple des deux séquences les plus proches. Ce consensus est aligné avec la troisième séquence la plus proche. L'alignement se construit progressivement par l'établissement de consensus successifs jusqu'à sa forme multiple finale. Les alignements présentés dans le manuscrit ont été édités avec la version 2.2 du programme ESPript (Gouet *et al.*, 1999).

9.6. Phylogénie

9.6.1. L'alignement

L'alignement multiple (§9.5.) des séquences à étudier constitue la première étape de l'analyse phylogénétique. L'alignement d'un nombre élevé de séquences (61 dans cette étude, Tableau 12, page 98) est difficile à réaliser, même avec les logiciels d'alignements automatiques. Il nécessite toujours d'être réajusté manuellement. Dans notre contexte, qui est celui de l'étude de séquences faiblement identiques, les arbres phylogénétiques n'ont pas été construits à partir des séquences complètes. Pour améliorer l'analyse, les alignements ont été traités avec la version 0.91 b du programme Gblocks (Castresana, 2000). Ce serveur permet d'extraire les régions pauvrement appariées, ambiguës et hypervariables qui fausseraient l'analyse et permet de générer des blocs de régions informatives c'est-à-dire des signaux phylogénétiquement exploitables. Les arbres présentés sont enracinés. Le groupe extérieur (outgroupe) choisi pour cela rassemble les protéines Yolk de 5 espèces de drosophiles (*Drosophila virilis* / AAW47625 ; *Drosophila montana* / ABU80343 ; *Drosophila americana* / ABE68792 ; *Drosophila ezoana* / AAW47631 ; *Drosophila littoralis* / ABU80342) (Hide *et al.*, 1992).

9.6.2. Méthodes de construction des arbres

Les arbres ont été générés selon 2 méthodes, le Neighbor-Joining (Saitou et Nei, 1987) et le maximum de parcimonie. Aucune des méthodes d'analyse phylogénétique n'incorpore tous les aspects des mécanismes biologiques de l'évolution des séquences. Il est donc convenu d'estimer que seules les branches retrouvées par plusieurs méthodes sont susceptibles d'être une bonne image de la réalité. Les analyses phylogénétiques en Neighbor Joining et parcimonie ont été réalisées à l'aide du logiciel PHYLO_WIN (Galtier *et al.*, 1996). Les arbres ont ensuite été édités dans TreeView (Page, 1996).

Espèces	Classes	Protéines	N° d'accension
<i>Anopheles gambiae</i>	Arthropoda	Protéine hypothétique	31239983
<i>Apis mellifera</i>	Arthropoda	Protéine hypothétique	48101936
<i>Arabidopsis thaliana</i> (1)	Plantae	Protéine inconnue	BAC42224
<i>Arabidopsis thaliana</i> (2)	Plantae	Phospholipase/Carboxylestérase	15235601
<i>Aspergillus fumigatus</i>	Ascomycota	Phospholipase	70984585
<i>Aspergillus niger</i>	Ascomycota		1395827
<i>Aspergillus oryzae</i>	Ascomycota	Phospholipase/Carboxylestérase	BAE62787
<i>Bos taurus</i>	Vertebrata	Protéine hypothétique	77736321
<i>Caenorhabditis briggsae</i>	Nematoda	Protéine hypothétique	CAE63703
<i>Caenorhabditis elegans</i>	Nematoda	Protéine hypothétique	T23324
<i>Candida albicans</i>	Ascomycota	Putative PLip/CEst	68470360
<i>Canis familiaris</i>	Vertebrata	Lysophospholipase II	73950621
<i>Danio rerio</i>	Vertebrata	Protéine hypothétique	41152185
<i>Debaryomyces hansenii</i>	Ascomycota	Protéine hypothétique	50420413
<i>Dictyostelium discoideum</i>	Amebozoa	Protéine hypothétique	EAL66424
<i>Drosophila melanogaster</i>	Arthropoda	Protéine hypothétique	AAG22322
<i>Emericella nidulans</i>	Ascomycota	Protéine hypothétique	EAA60541
<i>Filobasidiella neoformans</i>	Basidiomycota	Protéine hypothétique	EAL20149
<i>Gallus gallus</i> (1)	Vertebrata	Lysophospholipase I	50737613
<i>Gallus gallus</i> (2)	Vertebrata	Lysophospholipase II	50759892
<i>Gibberella zeae</i>	Ascomycota	Protéine hypothétique	EAA74608
<i>Homo sapiens</i> (1)	Vertebrata	Lysophospholipase I	O75608
<i>Homo sapiens</i> (2)	Vertebrata	Lysophospholipase II	AAC72844
<i>Isochrysis galgana</i>	Chromalveolata	Estérase putative	CAO02582
<i>Magnaporthe grisea</i>	Ascomycota	Protéine hypothétique	EAA47568
<i>Mus musculus</i> (1)	Vertebrata	Lysophospholipase I	P97823
<i>Mus musculus</i> (2)	Vertebrata	Lysophospholipase II	7242156
<i>Neurospora crassa</i>	Ascomycota	Putative lysophospholipase	CAC10084
<i>Oryctolagus cuniculus</i> (1)	Vertebrata	iPLA ₂ calcium indépendante isoforme 1	AAC63431
<i>Oryctolagus cuniculus</i> (2)	Vertebrata	iPLA ₂ calcium indépendante isoforme 2	O77821
<i>Oryza sativa</i> (1)	Plantae	Phospholipase/Carboxylestérase	CAE02817
<i>Oryza sativa</i> (2)	Plantae	Phospholipase/Carboxylestérase	50929517
<i>Oryza sativa</i> (3)	Plantae	Protéine inconnue	CAE02816
<i>Pan troglodytes</i>	Vertebrata	Lysophospholipase II	55586469
<i>Phaeodactylum tricorutum</i>	Chromalveolata	Putative PLip/CEst	21816
<i>Pongo pygmaeus</i>	Vertebrata	Protéine hypothétique	CAH90793
<i>Pseudomonas aeruginosa</i>	Bacteria	Probable carboxylestérase	A83163
<i>Pseudomonas fluorescens</i>	Bacteria	Carboxylestérase	Q51758
<i>Pseudomonas fluorescens</i>	Bacteria	Estérase prédite	23059179
<i>Pseudomonas putida</i>	Bacteria	Carboxylestérase	AAN66920
<i>Rattus norvegicus</i> (1)	Vertebrata	iPLA ₂ calcium indépendante	P70470
<i>Rattus norvegicus</i> (2)	Vertebrata	Lysophospholipase II	13786178
<i>Schistosoma japonicum</i>	Plathelminthes	Lysophospholipase	AAD52700
<i>Schistosoma mansoni</i>	Plathelminthes	Homologue lysophospholipase	AAC62254
<i>Silurana tropicalis</i>	Vertebrata	Protéine hypothétique	AAH64187
<i>Strongylocentrotus purpuratus</i>	Echinodermata	Lysophospholipase I	72050054
<i>Tetraodon nigroviridis</i> (1)	Vertebrata	Protéine hypothétique	CAG10223
<i>Tetraodon nigroviridis</i> (2)	Vertebrata	Protéine hypothétique	CAG02318
<i>Thalassiosira pseudonana</i> (1)	Chromalveolata	Putative PLip/CEst	6090
<i>Thalassiosira pseudonana</i> (2)	Chromalveolata	Putative PLip/CEst	18091
<i>Ustilago maydis</i>	Basidiomycota	Protéine hypothétique	71003193
<i>Xenopus laevis</i> (1)	Vertebrata	Protéine hypothétique	AAH73342
<i>Xenopus laevis</i> (2)	Vertebrata	Similaire lysophospholipase II	AAH44315
<i>Xenopus tropicalis</i> (2)	Vertebrata	Protéine hypothétique	45361419
<i>Xenopus tropicalis</i> (1)	Vertebrata	Lysophospholipase II	AAH75270
<i>Yarrowia lipolytica</i>	Ascomycota	Protéine hypothétique	50545321

Tableau 12 : Liste des espèces et des séquences utilisées pour l'analyse phylogénétique.

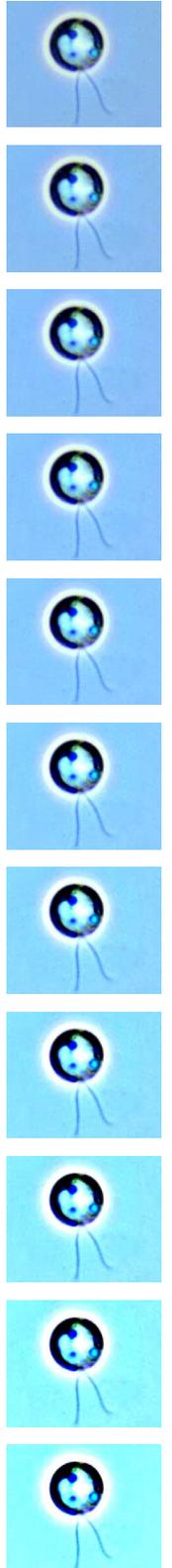
9.6.3. Estimation de la robustesse des arbres

La fiabilité des arbres doit être statistiquement établie. Pour cela, l'analyse de bootstrap est utilisée. C'est une méthode d'échantillonnage qui examine l'influence de la suppression de certaines informations sur la stabilité de l'arbre : n colonnes sont choisies arbitrairement à la place de l'alignement d'origine. Cette opération doit être réalisée entre 100 et 1000 fois. Un arbre est calculé pour chaque échantillonnage. Les différents arbres sont comparés puis compilés en un arbre consensus. Un pourcentage de robustesse (valeur de bootstrap) est attribué à chacun des nœuds de l'arbre consensus en fonction du nombre de fois où le nœud a été retrouvé à la position qu'il occupe au sein de l'arbre consensus. Plus la valeur de bootstrap est élevée, plus la fiabilité de la branche est importante. 500 valeurs de bootstrap ont été utilisées pour la construction des arbres présentés dans ce mémoire.

9.7. Prédiction de la structure tertiaire des protéines

Les modèles moléculaires tridimensionnels ont été construits avec le programme Geno3D (Combet *et al.*, 2002) par comparaison et alignement avec les plus proches homologues de conformation 3-D connue. Les protéines qui ont servi d'empreintes pour construire les modèles structuraux des protéines étudiées dans ce mémoire sont la thioestérase humaine (code PDB : 1fj2) (Devedjiev *et al.*, 2000) et la carboxylestérase de *Pseudomonas fluorescens* (code PDB : 1auo) (Kim *et al.*, 1997). Plus la similarité est grande, plus la modélisation est performante. Le programme Geno3D fonctionne selon une méthode géométrique qui consiste à mesurer des distances sur la protéine empreinte puis à les appliquer à la molécule à modéliser. Le logiciel détermine ensuite les conformations possibles qui, à la fois, satisfont les contraintes spatiales (distances) déduites de l'empreinte et présentent une optimisation énergétique mécanique et dynamique (liaisons covalentes, angles de valence, rayons de Van der Waals) (Combet *et al.*, 2002). Les images tri-dimensionnelles ont été éditées et annotées en utilisant le logiciel RasMol version 2.6 (Sayle et Milner-White, 1995) et le programme Protein Explorer (Martz, 2002).

RESULTATS ET DISCUSSION



1. Analyse de l'activité lipolytique endogène d'*Isochrysis galbana*

1.1. Mise en évidence d'une activité lipolytique

Initialement, la présence d'une activité lipolytique endogène chez *I. galbana* avait été soupçonnée, par l'observation d'acides gras libres dans les lipides totaux extraits de la microalgue (Poisson, 1999). Par la suite, les analyses lipidiques par CCM, effectuées au laboratoire par Marie Devos en 2005, ont révélé l'existence d'activités lipolytiques endogènes à l'encontre des glycolipides et des phospholipides (Devos, 2005). Afin de confirmer ce point et de s'assurer de la présence d'enzymes lipolytiques chez *I. galbana*, nous avons, pour commencer ces travaux de thèse, renouvelé cette expérience, en conservant le protocole déjà établi.

Les fractions lipidiques obtenues sont constituées de spots spécifiques (Figure 24). Ils ont été identifiés, lors des travaux de Marie Devos, par séparation des classes de lipides puis par comparaison de leur rapport frontal avec ceux de standards chromatographiés dans les mêmes conditions. Ainsi, on observe trois fractions de lipides distincts : les lipides neutres, les glycolipides et les phospholipides. Les lipides neutres contiennent des triglycérides, des stérols (Figure 24 : dépôt 1), parfois des acides gras libres (Figure 24 : dépôts 2 et 3) et des pigments verts et orangés. Dans la fraction des glycolipides, les spots observés correspondent à des monogalactosyldiglycérides, des digalactosyldiglycérides, des sulfoquinovosyldiglycérides et à un glycolipide non identifié. On retrouve, dans la fraction des phospholipides, la phosphatidylcholine et un autre phospholipide non identifié (Devos, 2005).

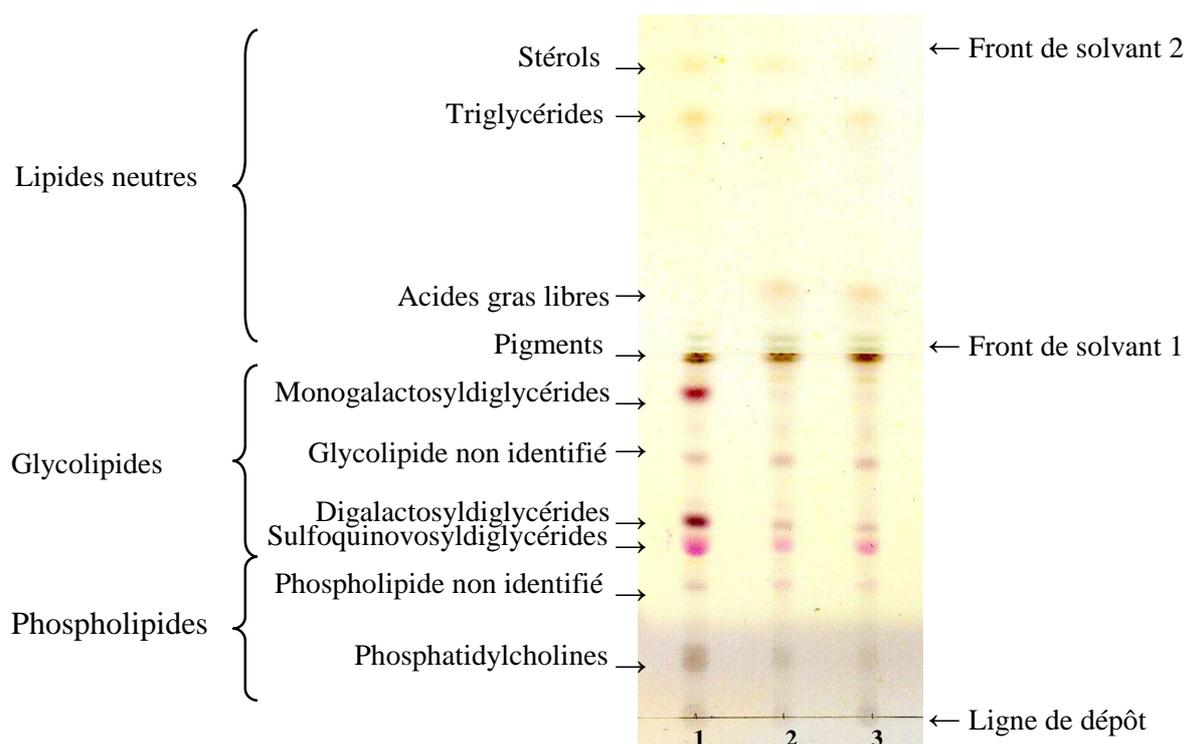


Figure 24 : Chromatographie sur couche mince des lipides totaux d'*Isochrysis galbana* pour la mise en évidence de l'activité lipolytique endogène.

Dépôt 1 : Echantillon témoin, ébouillanté avant incubation à 40 °C pendant 24 heures d'agitation. Dépôts 2 et 3 : Echantillons non ébouillantés, incubés à 40 °C pendant 24 heures d'agitation.

Après 24 heures d'incubation à 40 °C, on constate la disparition ou l'atténuation de certains spots de lipides dans les échantillons non ébouillantés (Figure 24 : dépôts 2 et 3). Ainsi, on observe, sur la plaque de silice, que le spot de monogalactosyldiglycérides a quasiment disparu et que les spots des sulfoquinovosyldiglycérides et digalactosyldiglycérides sont réduits. Les spots de phosphatidylcholines et du phospholipide non identifié apparaissent également atténués. Parallèlement, on note l'apparition d'un spot d'acides gras libres (Figure 24 : dépôts 2 et 3), absent de l'échantillon témoin ébouillanté (Figure 24 : dépôt 1). Les acides gras ont donc été libérés par l'action d'une activité lipolytique endogène, dirigée vers les glycolipides et les phospholipides.

1.2. Localisation de l'activité lipolytique

Dans un second temps, nous nous sommes interrogés sur la localisation intracellulaire de l'activité lipolytique mise en évidence chez *I. galbana*. Pour cela, une analyse lipidique par CCM a été réalisée à partir de fractions cytosoliques et membranaires préalablement séparées par broyage des cellules suivi d'une centrifugation.

Les résultats (Figure 25) montrent l'apparition de spots d'acides gras libres et l'atténuation des spots de glycolipides et de phospholipides, dans le broyat total et la fraction membranaire non ébouillantés (Figure 25 : dépôts 2 et 6). Aucun spot d'acide gras libre n'est détectable dans la fraction cytosolique (Figure 25 : dépôt 4). L'activité lipolytique endogène mise en évidence semble donc associée aux membranes de la microalgue. Les enzymes impliquées dans ces actions hydrolytiques, bien qu'elles soient probablement solubles comme toutes les enzymes lipolytiques, fonctionneraient étroitement associées aux membranes.

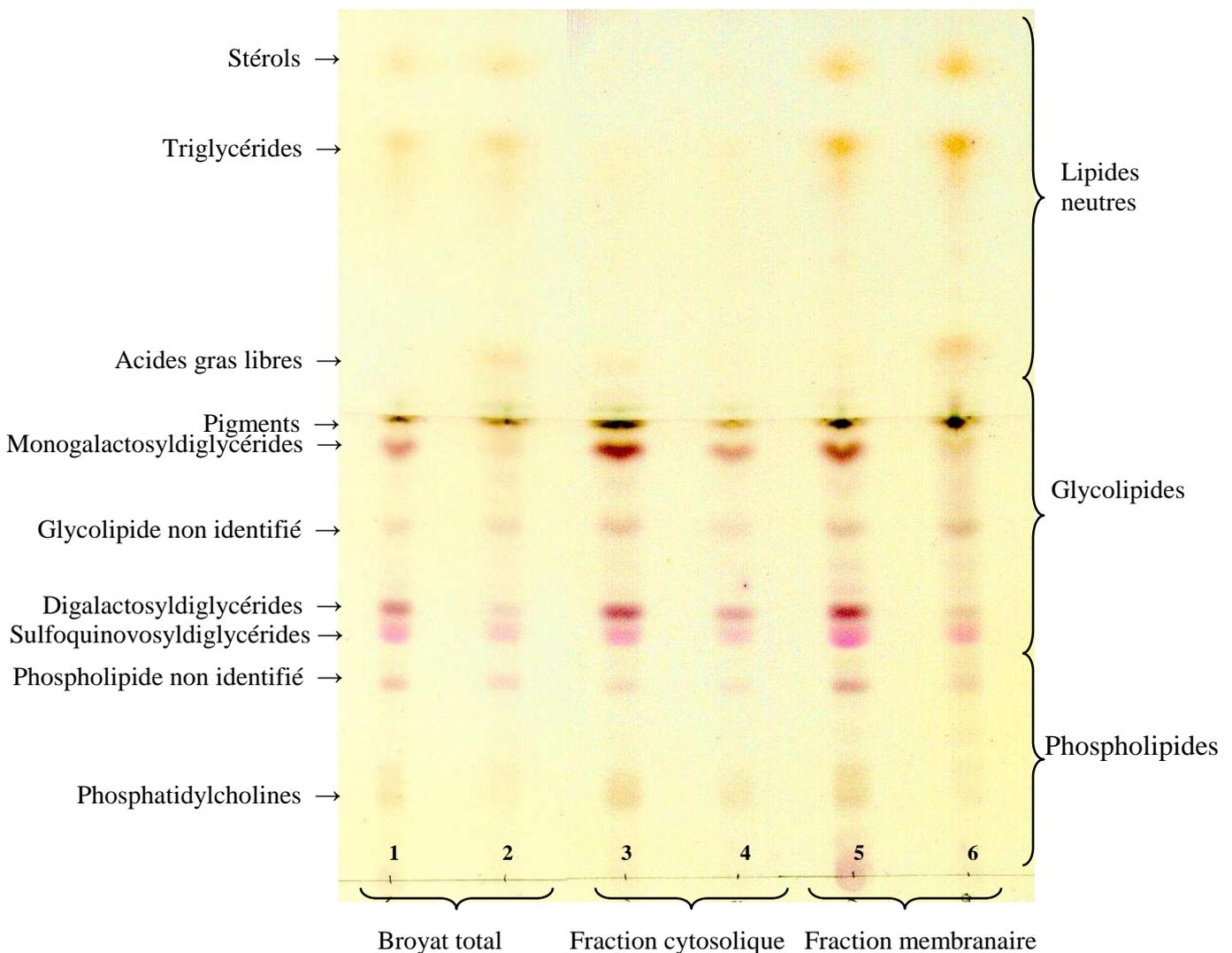


Figure 25 : Chromatographie sur couche mince des lipides totaux d'*Isochrysis galbana* pour la localisation intracellulaire de l'activité lipolytique endogène.

Dépôts 1 et 2 : Broyat total

Dépôts 3 et 4 : Fractions cytosoliques

Dépôts 5 et 6 : Fractions membranaires.

Les échantillons 1, 3 et 5 ont été ébouillantés pour servir de témoins négatifs.

1.3. Conclusion

La présence, chez *I. galbana*, d'activités lipolytiques endogènes vis-à-vis des glycolipides et des phospholipides, a été confirmée : la microalgue *I. galbana* possède des enzymes à activité lipolytique associées aux membranes.

2. Isolement de 3 gènes codant des enzymes lipolytiques chez *Isochrysis galbana*

L'approche moléculaire a été choisie pour purifier et caractériser les enzymes lipolytiques d'*I. galbana*. Cette approche consiste à cloner et surexprimer les gènes codant les enzymes recherchées. Elle nécessite, pour commencer, l'isolement des gènes codant les enzymes d'intérêt. Pour cela, deux stratégies ont été successivement appliquées. Dans un premier temps, la construction et le criblage d'une banque d'ADNc (§2.1.) ont été réalisés. Dans un second temps, la méthode d'amplification par RACE-PCR a été adoptée. Cette seconde stratégie a pu être développée après la publication du premier projet de séquençage partiel des ADNc relatif à la microalgue *I. galbana* et révélant la présence de marqueurs d'enzymes lipolytiques (§2.2.). La seconde méthode a permis l'obtention de trois ADNc pleine-longueur (§2.3.).

2.1. Construction et criblage d'une banque d'ADN complémentaires

2.1.1. Extraction des ARN totaux et purification des ARN messagers

La banque d'ADNc a été réalisée à partir d'ARN totaux extraits de la biomasse d'une culture de 1,5 L d'*I. galbana* en photobioréacteur. Les ARN messagers ont été purifiés puis rétrotranscrits en ADNc avant d'être utilisés pour construire la banque. La Figure 26A illustre les profils électrophorétiques des ARN totaux (ARNT) extraits. Les ARNm apparaissent en fond sous forme d'une trace. Les bandes épaisses caractérisent les ARN ribosomiques. Classiquement, les profils électrophorétiques des ARNr chez les eucaryotes montrent 4 bandes majeures correspondant, par ordre de taille décroissante, aux ARNr 25-28S, 17-19S, 5,8S et 5S. Le profil d'ARN d'*I. galbana* totalise plus de 4 bandes et semble, à première vue, plus complexe que celui de la levure *Saccharomyces cerevisiae*, par exemple (Figure 26B). Après purification, les ARNm apparaissent clairement sur le gel d'agarose (Figure 26C). La

zone maximale d'intensité lumineuse de la smear s'étale approximativement de 500 à 1500 pb. Il subsiste quelques traces d'ARNr.

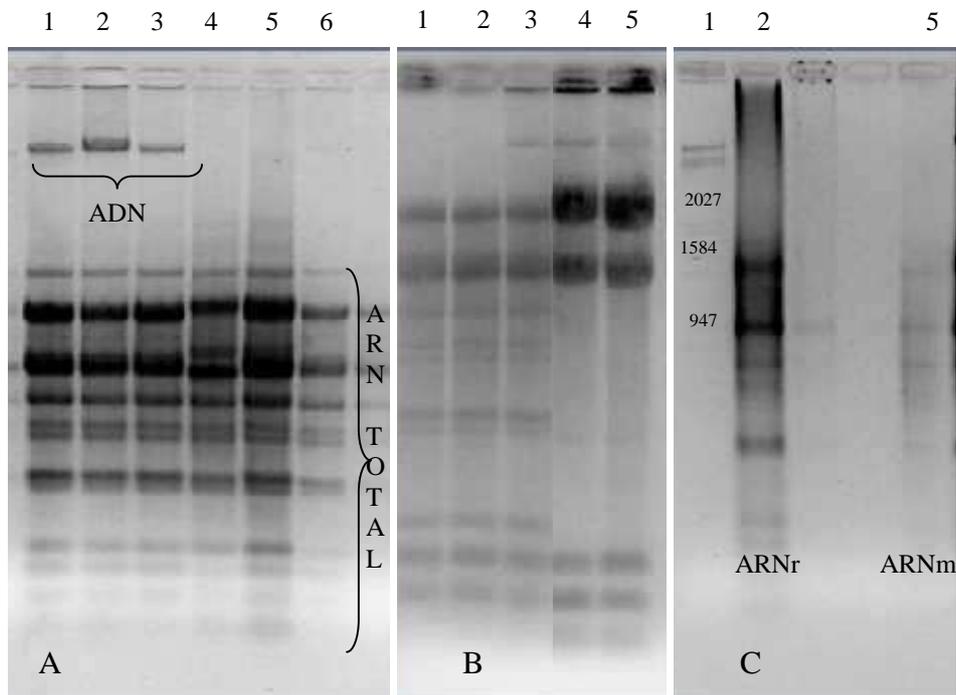


Figure 26 : Electrophorèses en gels d'agarose des ARN totaux et des ARNm d'*Isochrysis galbana*.

Les acides nucléiques ont été séparés sur gels à 2,5 % (w/v). Le marqueur de taille utilisé est le bactériophage λ digéré par les enzymes *EcoRI* et *HindIII*. L'échelle de taille est en paires de base. **A)** Isolement des ARN totaux d'*I. galbana*. Les dépôts 1 à 3 correspondent à des échantillons non traités à la DNase, à la différence de ceux déposés dans les puits 4, 5 et 6. **B)** Comparaison des profils d'ARN totaux d'*I. galbana* (pistes 1 à 3) et de *Saccharomyces cerevisiae* (pistes 4 et 5). **C)** Purification des ARNm d'*I. galbana*. La fraction enrichie en ARNm (piste 5) s'obtient par chromatographie d'affinité. La méthode assure l'élimination d'une proportion importante des ARNr (piste 2).

2.1.2. Fractionnement des ADNc et construction de la banque

Après rétrotranscription de la fraction enrichie en ARNm, les ADNc obtenus ont été fractionnés selon leur taille afin d'éviter que la banque d'ADNc ne soit surreprésentée en fragments de petites tailles. Lors du clonage, les ADNc de petites tailles ont, en effet, tendance à se liguer préférentiellement à l'ADN du phage, comparés aux ADNc de tailles plus importantes. Pour optimiser la fréquence des gènes de grandes tailles et obtenir un clonage homogène, les ADNc ont été fractionnés en 3 groupes, de tailles différentes : ADNc ≤ 564 pb (groupe 1), $564 \text{ pb} \leq \text{ADNc} \leq 1375$ pb (groupe 2), $1375 \text{ pb} \leq \text{ADNc} \leq 4268$ pb (groupe 3). Les tailles indiquées pour la constitution des groupes font référence à celles du marqueur de taille. Ces 3 fractions, représentées sur la Figure 27, ont été clonées dans l'ADN du phage λ et ont permis la réalisation de 3 banques d'ADNc différentes.

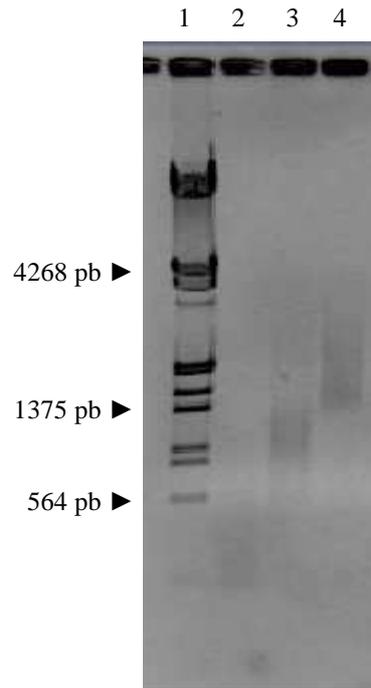


Figure 27 : Electrophorèse en gel d'agarose des ADNc double brin fractionnés en 3 groupes de tailles différentes.

Les acides nucléiques ont été séparés sur gel à 1 % (w/v). Piste 1 : Marqueur de taille (bactériophage λ digéré avec les enzymes *EcoRI* et *HindIII*). Piste 2 : ADNc \leq à 564 pb. Piste 3 : 564 pb \leq ADNc \leq 1375 pb. Piste 4 : 1375 \leq ADNc \leq 4268 pb.

2.1.3. Titration et analyse de la banque d'ADNc

Les phages, reconstitués avec l'ADN recombinant, ont été utilisés pour infecter une culture bactérienne. Chaque plaque de lyse contient les clones d'un même phage recombiné. Le nombre de plaques de lyses présentes sur les boîtes de Petri a permis de titrer les trois banques primaires : $1,7 \cdot 10^6$ ufp/mL (banque 1), $6,4 \cdot 10^6$ ufp/mL (banque 2), $4,9 \cdot 10^6$ ufp/mL (banque 3). Pour chacune des 3 banques, la présence et la taille des inserts ont été contrôlées. Pour cela, des plaques de lyse isolées ont été prélevées de façon aléatoire. Les phages récupérés ont été multipliés, récoltés et lysés afin d'en extraire l'ADN. Chaque clone a ensuite été contrôlé par amplification de l'insert avec le couple d'amorces calquées sur les promoteurs T3 et T7 qui encadrent le site de clonage du phage λ . La Figure 28 illustre les résultats du contrôle de deux clones, prélevés au hasard, pour chacune des 3 banques. La taille des inserts amplifiés concorde approximativement avec celle des fractions d'ADNc clonées.

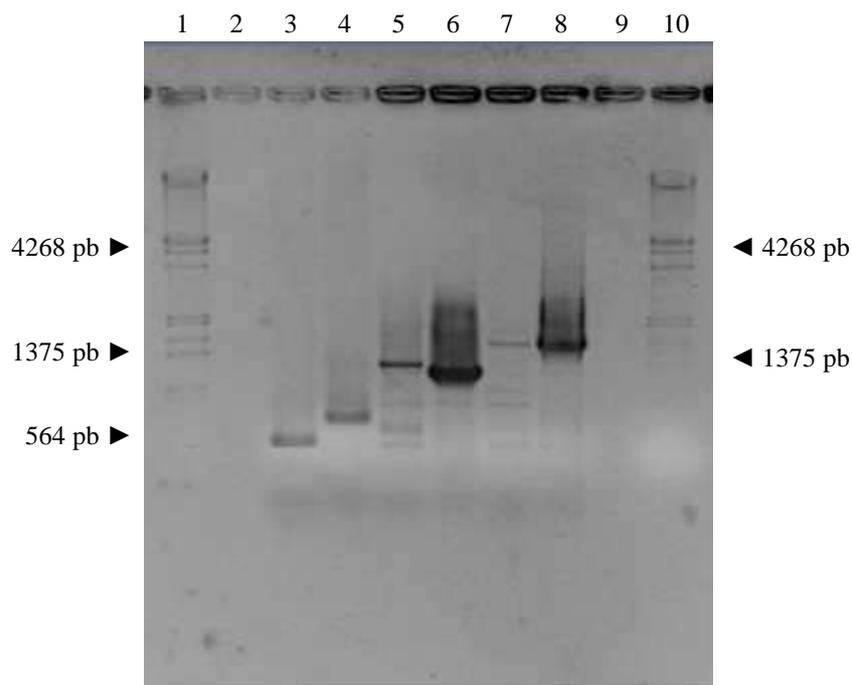


Figure 28 : Electrophorèse en gel d'agarose d'ADN extrait de phages recombinés prélevés aléatoirement dans les banques 1, 2 et 3, et amplifié par PCR avec les amorces T3S et T7AS.

Les acides nucléiques ont été séparés sur gel à 1 % (w/v). Le marqueur de taille utilisé est le bactériophage λ digéré par les enzymes *EcoRI* et *HindIII*. Les inserts ont été amplifiés par PCR avec le couple d'amorces T3S/T7AS. Pistes 1 et 10 : Marqueurs de taille. Pistes 3 et 4 : Clones 1 et 2 de la banque 1. Pistes 5 et 6 : Clones 1 et 2 de la banque 2. Pistes 7 et 8 : Clones 1 et 2 de la banque 3.

2.1.4. Recherche de sondes pour le criblage de la banque

Suite aux travaux préalablement effectués par M. Devos au sein du laboratoire, nous avons, pour commencer, focalisé notre recherche sur l'isolement d'un gène de phospholipase de type A₂ (PLA₂). En effet, après avoir révélé l'existence d'une activité phospholipasique chez *I. galbana*, il a été observé que le DHA, préférentiellement localisé en position sn-2 des phospholipides, était libéré en plus grande proportion par cette activité lipolytique. L'implication d'une PLA₂ avait donc été suggérée (Devos, 2005).

Lorsque ces travaux de thèse ont débuté, nous ne disposions pas de sondes spécifiques, correspondant aux gènes recherchés, pour cribler la banque d'ADNc. Effectivement, en 2005, la littérature était encore exempte d'informations sur les séquences des enzymes lipolytiques d'*I. galbana*. Pour commencer, il a été envisagé d'obtenir des sondes nucléotidiques spécifiques d'*I. galbana* par amplification de régions d'ADNc avec de courtes amorces hétérologues calquées sur des motifs de PLA₂ conservés chez d'autres organismes.

Deux séquences homologues de PLA₂ présentes dans les bases de données génomiques des microalgues *Cyanidoschyzon merolae* et *Thalassiosira pseudonana* ont été choisies et analysées pour la conception des amorces nucléotidiques. Elles présentent des similitudes significatives avec des PLA₂ de type cellulaire, indépendantes du calcium (iPLA₂) et associées aux membranes. Les iPLA₂ sont généralement des protéines de 85 à 88 kDa contenant un motif de fixation de l'ATP de type Gly-X-Gly-X-X-Gly, un motif consensus de lipase de type Gly-X-Ser-Thr-Gly, la triade catalytique Ser/Asp/His et plusieurs motifs ankyrine (Winstead *et al.*, 2000). Les répétitions ankyrines sont communément retrouvées chez plus de 400 protéines différentes. Généralement situés dans la région N-terminale, ces motifs sont impliqués dans les interactions intermoléculaires (Sedgwick et Smerdon, 1999). Les motifs, choisis pour la conception des amorces, sont bien conservés chez les iPLA₂. Ce sont, de l'extrémité N- à l'extrémité C-terminale, les motifs D-G-G-G-T-R-A, G-T-S-T-G-A-I et R-A-S-S-A-A-P, respectivement nommés PLM1, PLM2 et PLM3. Le second motif, PLM2, inclut le pentapeptide consensus des enzymes lipolytiques avec la sérine catalytique.

Dans un premier temps, des amorces dégénérées sens et antisens (*dS* et *dAS*) ont été conçues à partir de ces motifs. Les couples d'amorces *PLM1dS/PLM2dAS*, *PLM2dS/PLM3dAS* et *PLM1dS/PLM3dAS* ont ensuite été testés pour l'amplification de régions d'ADNc d'*I. galbana*. Ces amorces n'ont pas permis l'obtention d'amplicons spécifiques mais des smears, résultant probablement d'un degré de dégénérescence des amorces beaucoup trop important.

La conception de nouvelles amorces hétérologues affichant un degré de dégénérescence moins élevé s'est révélé problématique à cause du faible taux de conservation des séquences protéiques des phospholipases. Pour cette raison, des amorces non dégénérées sens et antisens ont été définies, dans un second temps, à partir des séquences nucléotidiques codant les motifs PLM1, PLM2 et PLM3 de la PLA₂, identifiée chez la microalgue *Cyanidoschyzon merolae* : *CmPLMS* et *CmPLMAS*. Les nouveaux couples d'amorces ont permis l'amplification de bandes nettes et distinctes (Figure 29A et B).

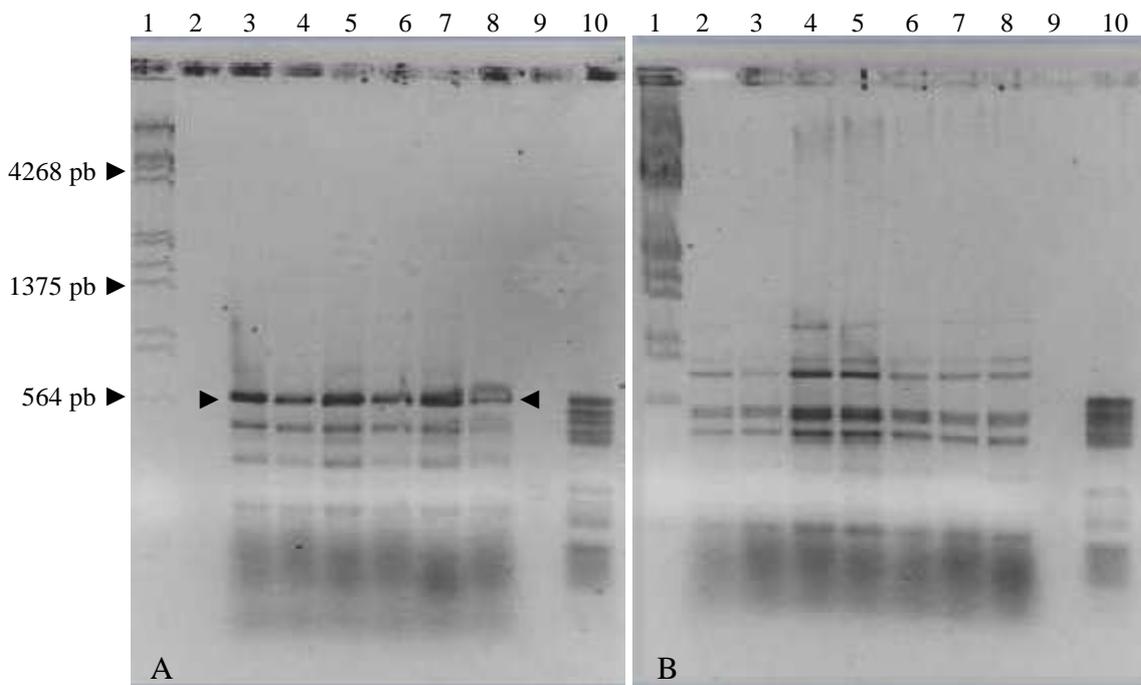


Figure 29 : Electrophorèses en gels d'agarose d'ADNc amplifiés avec des amorces hétérologues non dégénérées.

Les acides nucléiques ont été séparés sur gels à 2 % (w/v). Le bactériophage λ digéré par les enzymes *EcoRI* et *HindIII* (piste 1) et le plasmide pBr322 digéré par *HaeIII* (piste 10) sont utilisés comme marqueurs de taille. Les ADNc ont été amplifiés avec les couples d'amorces non dégénérées dessinées à partir de la séquence CMR500C de l'algue rouge *Cyanidioschyzon merolae*.

A) Amplifications réalisées avec le couple d'amorces *CmPLM1S/CmPLM3AS* (pistes 3 à 8).

B) Amplifications réalisées avec le couple d'amorces *CmPLM2S/CmPLM3AS* (pistes 2 à 8).

Toutes les bandes n'ont pu être amplifiées et extraites du gel d'agarose en quantité suffisante pour être séquencées. D'autre part, le séquençage, qui dépend fortement de la purification des produits PCR, n'a pas fonctionné dans tous les cas. La bande, indiquée par des triangles noirs sur la Figure 29A, a été séquencée (Annexe 6). Elle compte 594 pb. Des alignements locaux ont été effectués avec le programme blastx qui assure la comparaison de la séquence nucléotidique, traduite dans les 6 cadres de lecture, avec les séquences protéiques indexées dans les bases de données. Les résultats montrent qu'il s'agit d'un faux positif, sans aucune relation avec d'éventuelles iPLA₂.

2.1.5. Criblage de la banque d'ADNc par PCR

Parallèlement à la recherche de sondes spécifiques, la banque d'ADNc a été criblée, de façon non exhaustive, en vue d'isoler un gène de iPLA₂ et un gène de lipase, en utilisant les courtes séquences hétérologues codant des motifs conservés. Le laboratoire n'étant pas habilité pour la manipulation d'isotopes radioactifs, les techniques d'amplification (PCR et RT-PCR) ont été appliquées pour le criblage de la banque d'ADNc.

2.1.5.1. Recherche d'un gène codant une phospholipase A₂

La banque d'ADNc a été criblée par PCR en utilisant une amorce *T3S* ou *T7AS* couplée avec l'amorce du motif 2, non dégénérée, sens ou antisens (*CmPLM2S* ou *AS*). Le motif 2 a été choisi préférentiellement pour son implication reconnue dans la composition du site catalytique des phospholipases.

Pour commencer, les banques 2 et 3 ont été criblées avec le couple d'amorces *CmPLM2S* et *T7AS*. Comme le montre la Figure 30A, plusieurs bandes distinctes ont été amplifiées. Trois amplicons de tailles environ égales à 876, 439 et 180 pb (Figure 30A, piste 3) ont été obtenus après amplification d'ADN de phage recombiné extrait de la banque 2. Ces 3 bandes ont été amplifiées à nouveau (Figure 30B), isolées et purifiées (Figure 30C) afin d'être séquencées.

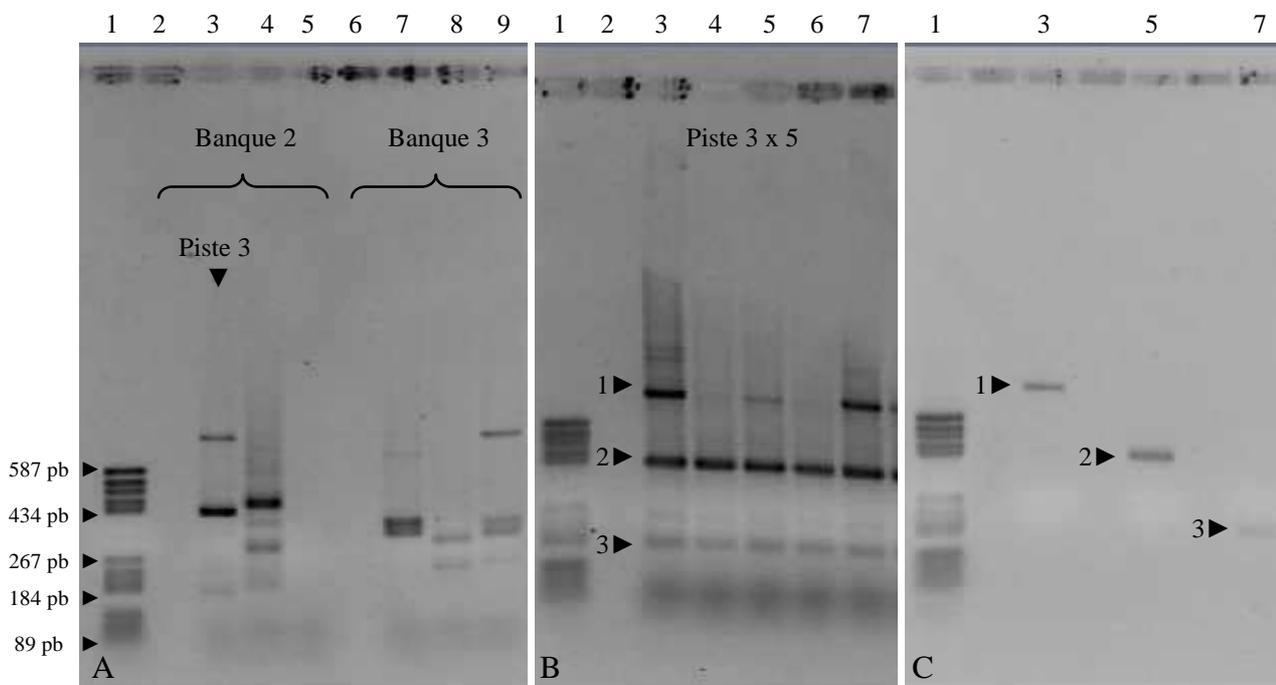


Figure 30 : Electrophorèses en gels d'agarose d'ADN extrait de phages recombinés (banques 2 et 3) et amplifié par PCR avec le couple d'amorces *CmPLM2S/T7AS*.

Les acides nucléiques ont été séparés sur gels à 1,5 % (w/v). Le marqueur de taille utilisé est le plasmide pBr322 digéré par *HaeIII*. **A)** Criblage des banques 2 (pistes 3, 4 et 5) et 3 (pistes 7, 8 et 9) par PCR avec le couple d'amorces *CmPLM2S/T7AS*. **B)** L'amplification du premier criblage de la banque 2 (Figure 30A, piste 3) a été reproduite 5 fois, avant l'excision des bandes, afin d'augmenter la quantité des amplicons pour le séquençage. **C)** Vérification de l'intégrité des 3 bandes après excision et extraction de l'agarose.

Seule la bande 2 théoriquement égale à 439 pb a pu être séquencée (Annexe 6) et analysée (Figure 31). La séquence comporte un codon stop et un signal de polyadénylation en 3'. Le polypeptide n'appartient pas à une iPLA₂ et présente des similitudes avec des protéines

hypothétiques identifiées chez la microalgue *Ostreococcus lucimarinus* (valeur E égale à 3^{-10}) et avec des ARN hélicases ATP-dépendantes dont celle de *Theileria parva* (valeur E égale à 4^{-08}).

```

protéine hypothétique [Ostreococcus lucimarinus CCE9901] XP_001420287
Score = 67.4 bits (163), valeur E = 3e-10
Identité = 34/62 (54%), Similarité = 48/62 (77%), Gaps = 0/62 (0%)
cadre de lecture = +2
Ig      8      LVLLLDYDKFELIKLLLKHRWKISVCTRLAQAQSEEEKASMLAEMA AVPQAVEVLDALAA 187
          LV +LDYDKF+LIK+LLK+R K+  CTRLA++ SEEE A++ A+M A P A  +L ++ A
Ol      321   LVSM LDYDKFDLIKVLLKNRLKVFWCTRLARSASEEETANIEAQM KADPAAAGILASMRA 380
Ig      188   SR   193
          +R
Ol      381   TR   382

ARN hélicase ATP-dépendante [Theileria parva strain Muguga] XP_762871
Score = 60.5 bits (145), valeur E = 4e-08
Identité = 30/59 (50%), Similarité = 41/59 (69%), Gaps = 0/59 (0%)
cadre de lecture = +2
Ig      8      LVLLLDYDKFELIKLLLKHRWKISVCTRLAQAQSEEEKASMLAEMA AVPQAVEVLDALA 184
          LVLLL Y+ F+ +KLLL++R+KI  CTRL QAQS++EK  +  EM+  Q  VL  L+
Tp      297   LVLLLKYENFDFVKLLLRNRYKIIYCTRLGQAQSQDEKNKIFDEMSKTAQGQLVLQELS 355

```

Figure 31 : Alignement local de la séquence du fragment 2 amplifié par PCR avec les amorces *CmPLM2S* et *T7AS* dans le cadre du criblage de la banque d'ADNc 2.

La séquence nucléotidique du fragment a été traduite dans les 6 cadres de lectures puis alignée avec les séquences protéiques répertoriées dans les bases de données publiques (blastx).

Les banques 2 et 3 ont également été criblées avec le couple d'amorces *T3S* et *CmPLM2AS*, de façon à amplifier des fragments qui ne possèdent pas de queues polyadénylées, lesquelles sont contraignantes lors du séquençage. Quatre bandes de tailles environ égales à 685, 600, 405 et 395 pb (Figure 32A, piste 4) ont été obtenues après amplification d'ADN de phage recombiné extrait de la banque 2. Ces fragments notés de 1 à 4 ont été amplifiés à nouveau (Figure 32B), isolés et purifiés afin d'être séquencés (Figure 32C).

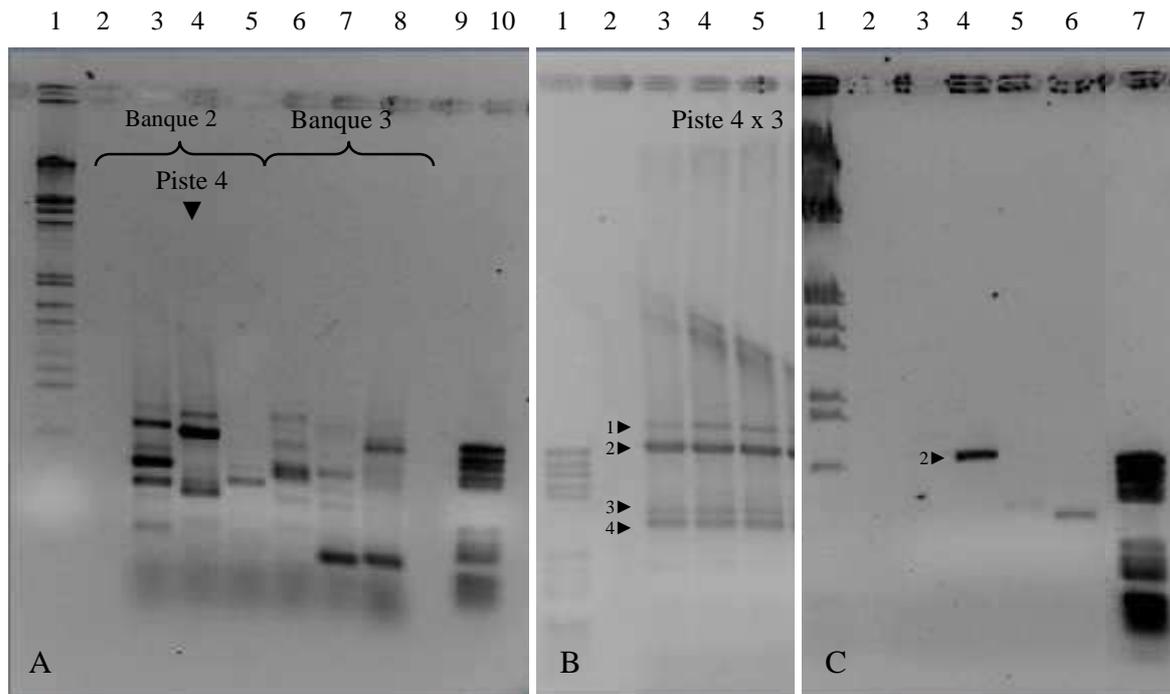


Figure 32 : Electrophorèses en gels d'agarose d'ADN extrait de phages recombinés (banques 2 et 3) et amplifié par PCR avec le couple d'amorces *T3S/CmPLM2AS*.

Les acides nucléiques ont été séparés sur gels à 1,5 et 2 % (w/v). Le bactériophage λ digéré par les enzymes *EcoRI* et *HindIII* et le plasmide pBr322 digéré par *HaeIII* ont été utilisés comme marqueurs de taille. **A)** Criblage des banques 2 (pistes 3, 4 et 5) et 3 (pistes 6, 7 et 8) par PCR avec le couple d'amorces *T3S/CmPLM2AS*. **B)** L'amplification du second criblage de la banque 2 (Figure 32A, piste 4) a été reproduit plusieurs fois, avant l'excision des bandes, afin d'augmenter la quantité des amplicons pour le séquençage. **C)** Vérification de l'intégrité des 4 bandes après excision et extraction de l'agarose.

Seule la bande, théoriquement égale à 600 pb, c'est-à-dire la bande 2 a pu être séquencée (Annexe 6) et analysée. La taille du fragment séquencé est de 584 pb. Les résultats d'alignements locaux ne montrent pas de similitudes significatives ni avec des iPLA₂ ni avec des protéines existantes.

2.1.5.2. Recherche d'un gène codant une lipase

La recherche a également été orientée vers les gènes de lipase. Les banques 2 et 3 ont été criblées par PCR. Le couple d'amorces, mis en jeu lors de ces réactions d'amplifications, est constitué de l'amorce antisens *T7AS* et d'une amorce sens de séquence identique à celle du pentapeptide consensus identifié chez une lipase hypothétique de l'algue *Cyanidoschyzon merolae* (*CmLipMS*). Quatre bandes de tailles environ égales à 876, 656, 461 et 200 pb ont été obtenues après amplification d'ADN de phage recombiné extrait de la banque 2 (Figure 33A, piste 4).

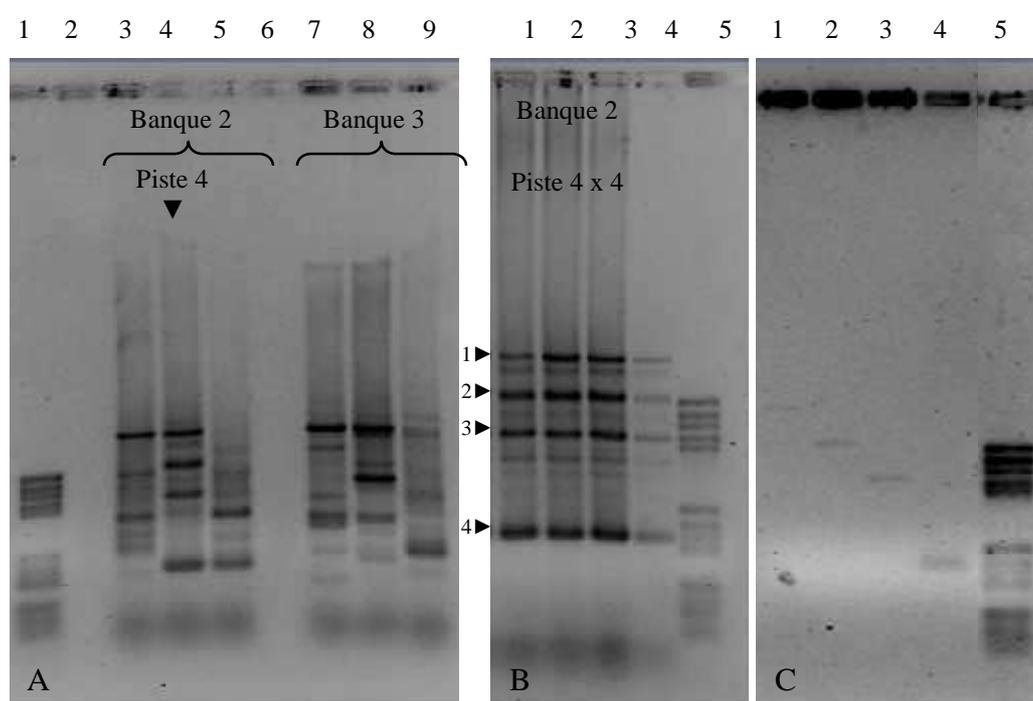


Figure 33 : Electrophorèses en gels d'agarose d'ADN extrait de phages recombinés (banques 2 et 3) et amplifié par PCR avec le couple d'amorces *CmLipMS/T7AS*.

Les acides nucléiques ont été séparés sur gels à 1,5 % (w/v). Le plasmide pBr322 digéré par *HaeIII* a été utilisé comme marqueur de taille. **A)** Criblage des banques 2 (pistes 3, 4 et 5) et 3 (pistes 7, 8 et 9) par PCR avec le couple d'amorces *CmLipMS/T7AS*. **B)** L'amplification du second criblage de la banque 2 (Figure 33A, piste 4) a été reproduit plusieurs fois, avant l'excision des bandes, afin d'augmenter la quantité des amplicons pour le séquençage. **C)** Vérification de l'intégrité des 4 bandes après excision et extraction de l'agarose.

Le fragment, numéroté 3 sur la Figure 33B, a été séquençé (Annexe 6). La séquence comporte un codon stop et le début d'un signal de polyadénylation en 3'. Les alignements obtenus avec blastx n'incluent pas d'homologues de lipase. Deux autres fragments de 517 et 240 pb obtenus avec les amorces *CmLipMS* et *T7AS* (photos des gels non représentées) ont également été séquençés (Annexe 6). Ils ne correspondent pas à des gènes de lipase.

2.1.6. Conclusion

La première stratégie envisagée pour l'isolement de gènes codant des enzymes lipolytiques chez *I. galbana* s'est révélée infructueuse pour de multiples raisons. La première raison est l'absence, à ce stade des travaux de thèse, de données spécifiques à l'organisme étudié. La seconde raison dépend de la nature même des enzymes étudiées dont le faible taux de conservation en acides aminés, combiné à la dégénérescence du code génétique, ne permet pas de concevoir des sondes hétérologues avec une homologie de séquence suffisante pour s'hybrider dans des conditions qui ne donnent pas de faux positifs. Enfin, la probabilité d'isoler un gène par criblage d'une banque est faible même lorsque les conditions de travail sont optimales (sonde radioactive spécifique, banque d'ADNc homogène et construite après induction de l'expression des gènes recherchés). Dans notre contexte, cette stratégie n'est donc pas la plus efficace pour isoler des gènes d'enzymes lipolytiques chez *I. galbana*.

2.2. Exploitation de la banque EST d'*Isochrysis galbana* (Keeling, 2006)

En 2006, sont publiés les résultats du premier projet de séquençage partiel des ADNc, relatif à la microalgue *Isochrysis galbana* (Keeling, 2006). L'analyse des marqueurs de séquences exprimées a permis d'entreprendre une étude à partir d'amorces et de sondes codant spécifiquement des enzymes lipolytiques chez *I. galbana*.

2.2.1. Recherche de marqueurs de séquences spécifiques codant des enzymes lipolytiques chez *Isochrysis galbana*

La base de données d'étiquettes de séquences transcrites (EST) d'*I. galbana* répertorie 12 clusters, sur un total de 6095, dont la traduction présente des similitudes statistiquement significatives avec des enzymes lipolytiques (Tableau 13, Annexe 7). Cela signifie que, dans les conditions de culture utilisées pour ce projet de séquençage, 12 gènes, susceptibles de coder pour des enzymes lipolytiques, se sont exprimés.

Les séquences nucléotidiques partielles des 12 gènes identifiés ont été à nouveau comparées avec l'ensemble des séquences éditées dans les bases de données publiques, en utilisant le programme d'alignement Blastx. Ce programme assure la comparaison de la séquence nucléotidique, traduite dans les 6 cadres de lecture, avec les séquences protéiques indexées dans les bases de données. Les résultats obtenus ont confirmé les annotations publiées dans la

banque EST : les 12 gènes présentent tous des similitudes avec une ou plusieurs familles d'enzymes impliquées dans le métabolisme lipolytique. Parmi les 12, 10 gènes (clusters surlignés dans le Tableau 13) présentent réellement un intérêt, par rapport à notre projet, en raison de l'homologie et du score d'alignement obtenus. Dans ces 10 candidats potentiellement intéressants, figurent un homologue de lipase, un homologue de phospholipase C et 8 homologues d'estérases. Aucun de ces gènes n'est disponible sur toute sa longueur car les signaux de début et de fin de traduction sont généralement absents. Cela implique que les séquences codantes complètes devront être déterminées par une démarche expérimentale.

Numéro de cluster	Nombre d'EST par cluster	Taille (pb)	Familles de protéines homologues	Meilleur alignement (Blastx : GenBank, PDB, SwissProt)	Valeur E du meilleur alignement local	Codon stop	Codon start + codon stop en amont
ISL00000940	2	655	Est/Lip/TEst, HF, acylT, PlacylT	acylT <i>Ostreococcus lucimarinus</i> (microalgue verte)	2,00 ^E -26	non	non
ISL00001076	3	679	PLip/CEst, LPLip	PLip/CEst <i>Synechococcus sp.</i> (cyanobactérie)	2,00 ^E -12	oui	non
ISL00001844	1	533	Lip, TAG Lip, Lip précurseur, Lip 3	Lipase putative <i>Leishmania major</i> (eucaryote unicellulaire)	3,00 ^E -04	non	non
ISL00002069	1	319	HF,Est/Lip/TEst, LPLip, MAG Lip, AG Lip, Plip	LPLip, Est/Lip <i>Vitis vinifera</i> (plante)	4,00 ^E -08	oui	non
ISL00002977	1	440	Est/Lip/TEst, HF, acylT, PlacylT	acylT <i>Ostreococcus lucimarinus</i> (microalgue verte)	2,00 ^E -18	non	non
ISL00003195	2	484	PLipC	PLipC <i>Nematostella vectensis</i> (cnidaire)	9,00 ^E -17	non	non
ISL00003216	2	598	Ank motif, facteur nucléaire, protéine associée à PLip	Protéine associée à une PLip <i>Serratia marcescens</i> (protéobactérie)	3,00 ^E -09	non	non
ISL00003782	1	413	LCacylT, LPLip, PLipA ₁ , PLipA ₂	LCacylT protéine <i>Caenorhabditis elegans</i> (nématode)	9,00 ^E -14	non	non

ISL00006775	1	422	Asparaginase, LPLip	Asparaginase <i>Dictyostelium discoideum</i> (eucaryote unicellulaire)	3,00 ^E -22	non	non
ISL00006895	1	350	LPLip, PLip/CEst, CEH, acyl TEst, PLA ₂ Ca ²⁺ i	PLip/CEst <i>Ostreococcus lucimarinus</i> (microalgue verte)	2,00E-05	non	non
ISL00007519	1	732	LPLip, PLip/CEst, AcylTEst	Carboxylestérase <i>Chromohalobacter salexigens</i> (protéobactérie)	1,00 ^E -12	non	non
ISL00008005	1	577	PLip/CEst, LPLip, AcylTEst, PLipA ₂ Ca ²⁺ i	PLip/CEst <i>Nasonia vitripennis</i> (insecte)	2,00 ^E -27	non	non

Tableau 13 : Liste des clusters identifiés dans la banque EST d'*Isochrysis galbana* présentant des homologies significatives avec des enzymes lipolytiques.

La numérotation indiquée est celle de la banque EST. HF : hydrolase α/β famille, Lip : lipase, AG Lip : acylglycérol lipase, MAG Lip : monoacylglycérol lipase, TAG Lip triacylglycérol lipase, Lip3 lipase de classe 3, Est/Lip/TEst estérase/lipase/thioestérase, Acyl Test : acyl thioestérase, PLip/CEst : phospholipase / carboxylestérase, PLipC : phospholipase C, PLipA₂ Ca²⁺i : phospholipase A₂ Ca²⁺ indépendante, PLipA₁ : phospholipase A₁, PLipA₂ : phospholipase A₂, LPLip : lysophospholipase, LCacylT : lécithine cholestérol acyltransférase, AcylT : acyltransférase, Pl AcylT : phospholipides acyltransférase. Les différents auteurs utilisent parfois des dénominations différentes pour désigner les mêmes types d'enzymes. N° d'accension des séquences dans GenBank : *N. vitripennis* XP_001599074, *D. discoideum* AY392437, *C. salexigens* YP_573162, *C. elegans* NP_492033, *O. lucimarinus* XP_001419459, *L. major* CAJ08439, *Vitis vinifera* CAO68548, *Serratia marcescens* AAM13979, *Synechococcus* sp. YP 479022, *N. vectensis* XP 001623656, *O. lucimarinus* XP_001418794, *H. sapiens* XP_001126680, *Oryza sativa* BAD07907, *B. pertussis* NP 880255.

2.2.2. Choix de trois gènes

Trois gènes, susceptibles de coder des enzymes intéressantes pour notre projet, ont été choisis pour continuer cette étude. Ces trois gènes, censés correspondre à des gènes de lipase et de carboxylestérase ou de phospholipase, sont indexés dans les clusters ISL00001844, ISL00008005 et ISL00001076 (clusters surlignés en rouge dans le Tableau 13). Les trois transcrits des clusters n° ISL00001844, ISL00008005 et ISL00001076 seront nommés respectivement *IgLip*, *IgEst1* et *IgEst2*, par analogie à la nomenclature déjà utilisée pour répertorier les séquences de lipases et d'estérases présentes dans les banques de données ou décrites dans la littérature : *IgLip* pour *Isochrysis galbana* Lipase, *IgEst1* et *IgEst2* pour *Isochrysis galbana* Esterase n°1 et n°2.

Le cluster ISL00001844 (Figure 34) comprend une seule étiquette de séquence transcrite, d'une taille équivalente à 533 nucléotides. Malgré la présence d'un codon start en amont de la séquence, l'analyse des alignements avec d'autres homologues conduit à penser que le transcrit est incomplet. En effet, d'après ces alignements et en assumant que la séquence d'*I. galbana* possède une région N-terminale similaire à ses homologues, il manquerait une centaine d'acides aminés. De plus, aucun codon stop situé sur la partie 5' non codante ne précède l'hypothétique codon start, ce qui ne nous permet pas d'affirmer qu'il s'agit du réel signal d'initiation de la traduction. Le codon de terminaison de traduction n'apparaît pas non plus en fin de séquence. Nous sommes donc en présence d'une région interne pour laquelle il manque les régions N- et C-terminales. Les alignements du cluster n° ISL00001844 montrent des homologies significatives principalement avec des lipases déjà caractérisées ou suspectées chez d'autres organismes. Le meilleur alignement est obtenu avec la lipase putative du protiste *Leishmania major*. La valeur E est suffisamment proche de zéro pour penser que cet alignement n'est pas le résultat du hasard.

```
5' -CTCCGGCTGGGTGCACCATGGCTGGGGTAGTGGCTACGACATGCTTCGCGCGCGCCTCATCACGGCTGTGCTGCTAGCGC
TTGAGGCTCGGCCGACGTACCGCCTCGTCTCGTACCGGCCACTCTGCGCGGGGCACTGGCCTCCATCTTTGTGGCCGAGCTC
GCCACCGTCCGAGCACTCAATCATCCTCGGTAGCCAGCGAATCGTCCGCGTACCGCGCCGAGGCCGGCTCTGACTCCAAAGG
CGAAGCTGACACGAAGGCTGCCCGGGCTGATGAGTCCGCCGCCGCCCTCCAGCTCCATCGCCTGGGAGAGTGTTCGGAGGTTG
CGCCGCTGCTCAGGCGACCGCTCGCGCCGCCCGCTGTACATCACCTTCGCCCGCCGATGGCCGGGAACCGCGAGTTTGGC
GCCGCCATCGATTCCGGCCTGGCCGCGCAGGGCAACGTGATGCTACGTGTGGTCAATGCACGCGATCCGGCCGTGCGTGCACC
CAACACGCTGAATGCGGGTGGCAGCGTCCAGATCCCCTGCCCAATGGGGGGCG-3'
```

Figure 34 : Séquence nucléotidique du cluster ISL00001844.

Les couples d'amorces utilisés sont surlignés en gris, avec de l'extrémité 5' à l'extrémité 3', l'amorce sens spécifique *ASsp* (dénommée *LipS*), la nested amorce sens spécifique *NASsp* (*NLipS*), la nested amorce anti sens spécifique *NAASsp* (*NLipAS*), et l'amorce anti sens spécifique *AASsp* (*LipAS*).

Le cluster n° ISL00008005 (Figure 35) est constitué d'un seul EST. La séquence de 577 pb est incomplète. Il n'y a pas de codon stop. La première méthionine est rencontrée au 17^{ème} codon. Il ne s'agit probablement pas du codon d'initiation de la traduction. Il n'est précédé d'aucun codon stop. De plus, les alignements réalisés avec des séquences homologues montrent la présence d'une suite de 4 à 5 résidus hautement conservés en amont de cette méthionine. La séquence protéique putative serait donc incomplète en N- et C-terminal. Elle présente des similitudes significatives avec des séquences de lysophospholipases, de phospholipases et de carboxylestérases.

5' -GCACACCGCGACGGTCATCGGCCCATCCACGGGCTGGGCGACAGCAACATGGGCTGGGCGGACGTGGCGATGCAGCTGC
 AGAGTGTGCATGCCGTATTGCAAGTTCATCCTCCCCAATGCGCCCGTCCGTCTGTCCACCTCAACGGCGGAATGTCGATGCC
 AGTTGGTATGACATCACGAGTTTGGATAAGCGCGAGTCCAGCCGTGCACGGGCATCGAAGAGTCTCGCAGGCGATGCTCGA
 CCTCATCTCCGCTGAGGTTGCGTCCGGCATCCCCCTCCCGCATCGCCGCTTCTCCAGGGAGGCGCCGTCGCC
 TCTTCACAGGGCTGTCAGTATTACACACCCCTGGCCGCGTGTATGCTTGTCCGGCTACTTGGCGGCGAGGAGCGTTTCAT
 ACTGGCGCCGGAGGCTGTGAACACTCCCGTCGCTCACTTTCACGGCTCGGATGATCAAACGGTCCAAATCAAATGGGCCCGCG
 GCTCTCAGGCCACCTGCGCGAGCTGGGCATCCGCACCTACGAGCTAAAAGAGTACTCGCCCTCGGCCACAGCGCTCGCA-
 3'

Figure 35 : Séquence nucléotidique du cluster ISL00008005.

Les couples d'amorces utilisés sont surlignés en gris, avec de l'extrémité 5' à l'extrémité 3', l'amorce sens spécifique *ASsp* (dénommée *EstIS*), la nested amorce sens spécifique *NASsp* (*NEstIS*), la nested amorce anti sens spécifique *NAASsp* (*NEstIAS*), et l'amorce anti sens spécifique *AASsp* (*EstIAS*).

Le troisième cluster numéroté ISL00001076 (Figure 36) est constitué de 3 EST. Le transcrit, de 679 pb, est incomplet. La première méthionine est rencontrée au 89^{ème} codon et il y a un codon stop TAG dans la région 5'. La séquence protéique putative de 226 acides aminés serait largement incomplète du côté N-terminal. En revanche, elle semble complète du côté C-terminal. Elle partage un taux d'identité et de similitude significatif avec des homologues d'estérases, de lysophospholipase et de phospholipase.

5' -aattctgcagatatttggagctttcgtTGGTTCGAGGGTATTCTTCGGCCCCAGAGCAGATTGCGCTCGACAAGACCG
 TCGGTTGTCGTCAGATGCTGTACTTGCATTTATTGAGGAAGCGCCTGCCAAGTTTGGGATCGATCCGAAGCGAGTCTTCCTG
 CTGGGATTTTCGACAGGGCGCCACACTCGTTTGGACGACCTCTTGTCCAGATGGTCGCGCCCCGGGCTCATCGCCGGCGGCT
 AACACTGTCTGGCCGGCTGTTTCTGAGCTAATGCAGGCTGGTACCCCACTTTATGCGCGTCTCGCCGACCCGACGCAGCTGA
 GGGATTGTGCCGCTTTGCTCACACGGCGCCCAAGACTCGGTTACCCCGGTAGCTATCGGGCGCTCGAATGAGCGGCTGTTT
 CGAGACTGGGCACCGGGTCCGGACTTAATATATAAGGAGGACCCTGTTGCAGGCCATGAGATATCGAAGGCCGCTGGCTGCTTC
 AATGTCCCAATGGTTCCGTCGGTATGTCTCCGGAGCATCAAAGGGTATCAGTAGGAGTTTCAGCGGCGAGCAgCACCCGTCC
 ATAGAGCAGGCTGTGGAAGCCAACCTTTGTGATGACGCGCCAAGATGCGTGGGCAAGCGTGGGTCACGCGCCCTCAAAC
 ATTCGAACGCTCGTCTCC-3'

Figure 36 : Séquence nucléotidique du cluster ISL00001076.

Les couples d'amorces utilisés sont surlignés en gris, avec de l'extrémité 5' à l'extrémité 3', l'amorce sens spécifique *ASsp* (dénommée *Est2S*), la nested amorce sens spécifique *NASsp* (*NEst2S*), la nested amorce anti sens spécifique (*NEst2ASsp*), et l'amorce anti sens spécifique *AASsp* (*Est2AS*).

2.3. Isolement de 3 ADNc codant des enzymes lipolytiques par la méthode de RACE-PCR

2.3.1. Amplification rapide des extrémités 3' et 5' des ADNc

Les séquences complètes des 3 gènes candidats ont été obtenues grâce à la technique d'amplification rapide des extrémités des ADNc (RACE).

Les extrémités 5' et 3' des ADNc *IgLip*, *IgEst1* et *IgEst2* ont été isolées par RACE PCR. Au total six fragments ont été amplifiés : 3'- et 5'- RACE *IgLip*, 3'- et 5'- RACE *IgEst1*, 3'- et 5'- RACE *IgEst2*. Le couple d'amorces utilisé en RACE PCR comprend une amorce universelle *AU* qui se fixe indifféremment sur tous les ADNc et une amorce spécifique du

gène candidat. C'est la région interne connue des gènes candidats qui a servi à l'élaboration des amorces spécifiques sens et antisens, respectivement destinées à l'amplification des extrémités 3' et 5'. Pour chaque gène, les deux amorces spécifiques ont été choisies de façon à ce que la région interne soit séquencée dans les deux sens. Cette mesure permettra de s'assurer, au moment de la reconstitution de la séquence complète, par chevauchement des séquences des fragments 3'- et 5'- RACE, que les extrémités 3' et 5' amplifiées appartiennent bien au même gène. La technique permet d'obtenir des bandes majeures qui sont généralement entourées de bandes parasites formant une smear importante (Figure 37).

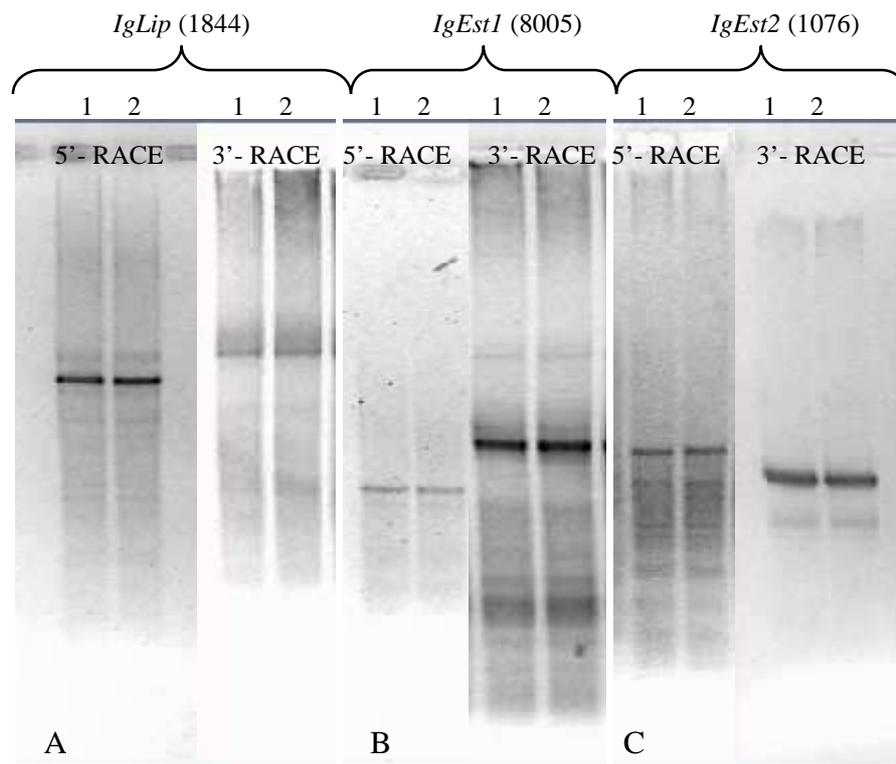


Figure 37 : Electrophorèses en gels d'agarose des régions terminales 5' et 3' des ADNc *IgLip*, *IgEst1* et *IgEst2* amplifiées par RACE PCR.

Les acides nucléiques ont été séparés sur gels à 1,5 % (w/v). **A)** Deux réplifications de l'amplification de l'extrémité 5' avec le couple d'amorces *AU/LipAS* : 5'-RACE *IgLip*. Deux réplifications de l'amplification de l'extrémité 3' avec le couple d'amorces *LipS/AU* : 3'-RACE *IgLip*. **B)** Deux réplifications de l'amplification de l'extrémité 5' avec le couple d'amorces *AU/Est1AS* : 5'-RACE *IgEst1*. Deux réplifications de l'amplification de l'extrémité 3' avec la couple d'amorces *Est1S/AU* : 3'-RACE *IgEst1*. **C)** Deux réplifications de l'amplification de l'extrémité 5' avec le couple d'amorces *AU/Est2AS* : 5'-RACE *IgEst2*. Deux réplifications de l'amplification de l'extrémité 3' avec le couple d'amorces *Est2S/AU* : 3'-RACE *IgEst2*.

A ce niveau, le séquençage des fragments amplifiés est impossible. En effet, le découpage manuel des bandes d'intérêt, par son imprécision, ne permet pas de s'affranchir des amplicons parasites. Les essais de séquençage tentés à ce stade ont quasiment tous échoué en raison d'un

important bruit de fond généré par les amplifications non spécifiques. Le chromatogramme, présenté en Figure 38, illustre l'illisibilité des résultats de séquençage.

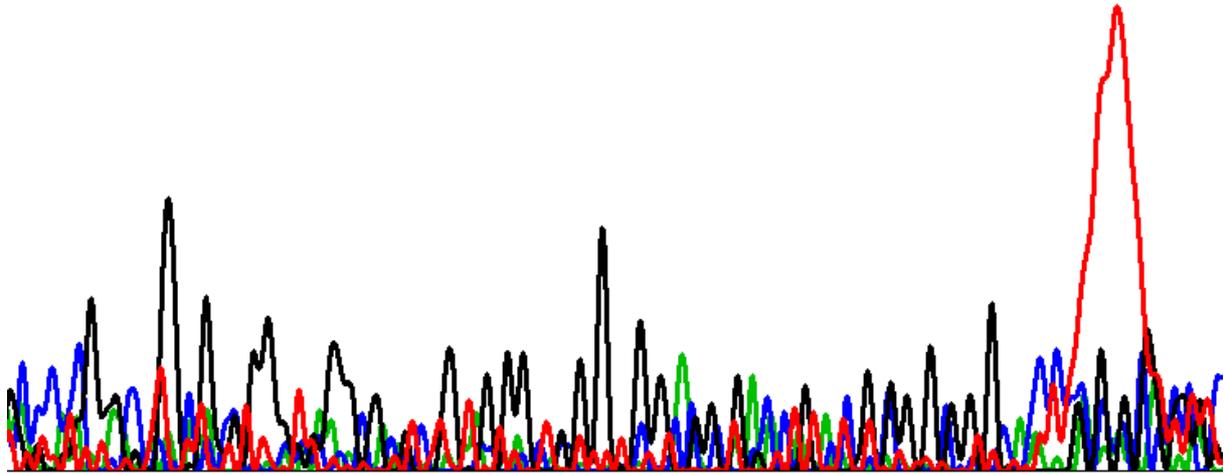


Figure 38 : Extrait du chromatogramme de séquençage de l'échantillon 3'-RACE *IgLip* amplifié par RACE PCR (MilleGen, France).

Aucune fluorescence spécifique n'a pu être enregistrée. Le signal présent correspond à un bruit de fond important. Les signaux obtenus se chevauchent traduisant la présence de bandes parasites. Rouge Thymine, vert Adénine, bleu Cytosine, noir Guanine.

2.3.2. Nested PCR

Afin de réduire les amplifications parasites et pour renforcer la spécificité de l'amplification, la technique de nested-PCR a été appliquée. Une première amplification, identique à la RACE-PCR, a été réalisée avec les amorces (spécifique et universelle) externes bordant la séquence à amplifier. Les bandes majeures obtenues ont été découpées, extraites du gel d'agarose, purifiées puis utilisées comme matrice pour la seconde PCR. Les amorces utilisées pour cette seconde PCR sont internes par rapport aux précédentes, c'est-à-dire qu'elles vont s'hybrider à l'intérieur de la séquence résultant de la première amplification. Les résultats obtenus montrent que les amplifications non spécifiques ont été significativement réduites (Figure 39). De plus, la quantité d'ADN amplifié a augmenté.

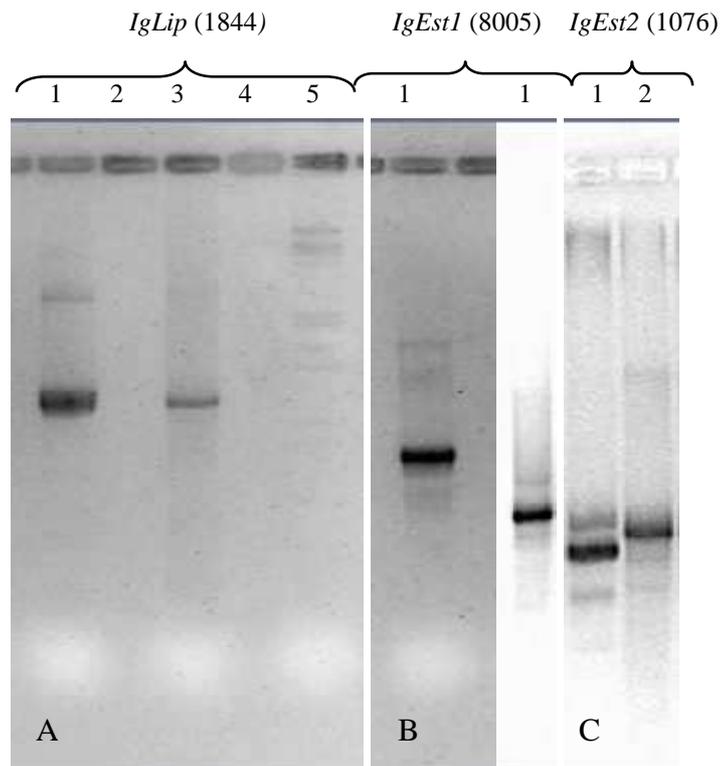


Figure 39 : Electrophorèses en gels d'agarose des régions terminales 3' et 5' des ADNc *IgLip*, *IgEst1* et *IgEst2* amplifiées par Nested PCR.

Les acides nucléiques ont été séparés sur gels à 2 % (w/v). **A)** Nested PCR des fragments 3'-RACE (piste 1) et 5'-RACE (piste 3) *IgLip*. **B)** Nested PCR des fragments 3'-RACE (gel 1, piste 1) et 5'-RACE (gel 2, piste 1) *IgEst1*. **C)** Nested PCR des fragments 3'-RACE (piste 1) et 5'-RACE (piste 2) *IgEst2*.

2.3.3. Séquençage des extrémités 3' et 5' des ADNc et reconstitution de la séquence complète

Les fragments d'intérêt obtenus par RACE et Nested PCR ont été découpés puis séparés de leur matrice d'agarose. Cette extraction a été contrôlée par électrophorèse en déposant 2 μ L d'échantillons extraits sur gel d'agarose 2 % (w/v). Le taille des fragments a été estimée graphiquement par comparaison avec la distance de migration parcourue par le marqueur de taille sur le gel. Les tailles des fragments 3'- et 5'-RACE de l'ADNc *IgLip*, par exemple, sont approximativement équivalentes à 1140 et 1075 pb. Les échantillons ont également été dosés par fluorimétrie. Au total, six fragments ont été séquencés par la société MilleGen. Comme le montre le chromatogramme de la Figure 40, les signaux parasites ont disparu.

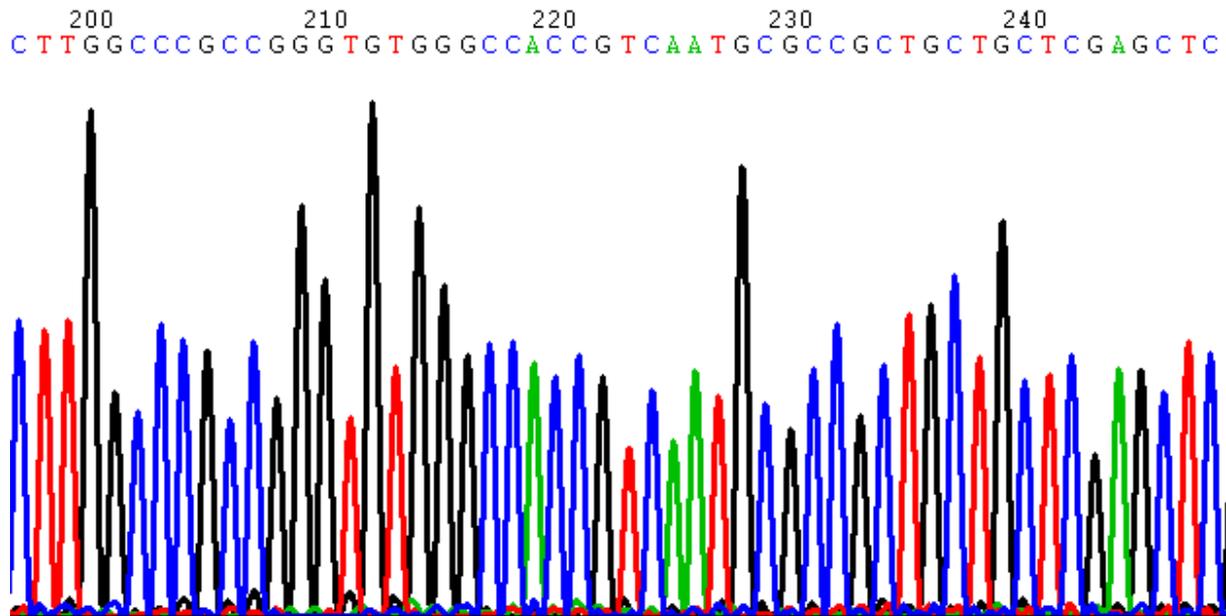


Figure 40 : Extrait du chromatogramme de séquençage de l'échantillon 3'-RACE *IgLip* amplifié par RACE et Nested PCR (MilleGen, France).
Rouge Thymine, vert Adénine, bleu Cytosine, noir Guanine.

La séquence pleine longueur des trois ADNc *IgLip*, *IgEst1*, *IgEst2* a ensuite été reconstituée par chevauchement des séquences des fragments 3' et 5'. Les tailles réelles des ADNc *IgLip*, *IgEst1* et *IgEst2* sont respectivement de 2089 pb, 1124 pb et 1058 pb.

2.3.4. Amplification des ADNc pleine-longueur

Après avoir reconstitué les séquences, les trois gènes ont été amplifiés et isolés entièrement c'est-à-dire sur toute leur longueur. Pour cela, de nouveaux couples d'amorces ont été conçus sur la base des extrémités des séquences codantes (CDS). Chaque amorce a été additionnée d'un site de coupure d'enzyme de restriction en vue d'une insertion plasmidique ultérieure : *EcoRI* pour l'amorce sens, *XbaI* pour l'amorce antisens.

Comme le montre la Figure 41, la taille des fragments amplifiés visibles sur le gel correspond à celle attendue pour chacune des trois séquences codantes ADNc : 1392 pb pour *IgLip*, 846 pb pour *IgEst1* et 723 pb pour *IgEst2*. Les trois gènes ont été séquencés par la société MilleGen. Les séquences obtenues correspondent parfaitement à celles reconstituées à partir des régions 3' et 5'.

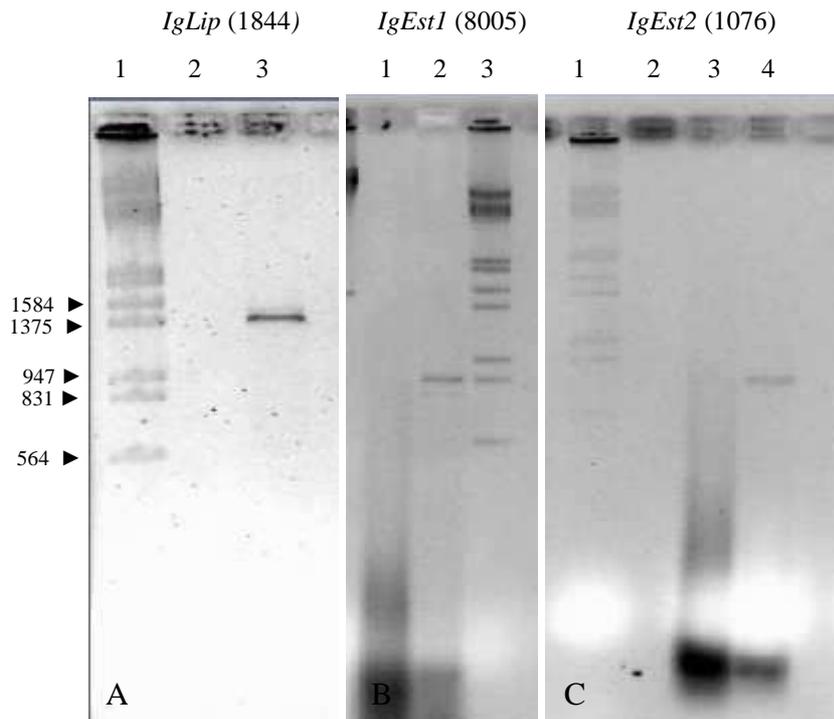


Figure 41 : Electrophorèses en gels d'agarose des séquences codantes complètes des ADNc *IgLip*, *IgEst1* et *IgEst2*.

Les acides nucléiques ont été séparés sur gels à 1 % (w/v). **A)** *IgLip* : PCR sur ARN (piste 2), PCR sur ADNc (piste 3). **B)** *IgEst1* : PCR sur ARN (piste 1), PCR sur ADNc (piste 2). **C)** *IgEst2* : PCR sur ARN (piste 3), PCR sur ADNc (piste 4). Les PCR réalisées à partir d'ARN constituent les témoins négatifs permettant de s'assurer de l'absence d'ADN dans les échantillons.

2.3.5. Conclusion

Les séquences nucléotidiques ont été déposées dans la base de données européenne EMBL. La séquence de l'ADNc du gène *IgLip* est référencée sous le numéro d'accèsion AM749809, celle du gène *IgEst1* sous le numéro d'accèsion AM748093 et celle du gène *IgEst2* sous le numéro d'accèsion AM711634.

La dernière méthode appliquée (RACE-PCR) a donc permis d'isoler trois ADNc, chez la microalgue *I. galbana*, susceptibles de coder des enzymes lipolytiques. Les analyses comparatives des trois séquences, dénommées *IgLip*, *IgEst1* et *IgEst2*, sont respectivement présentées et analysées dans les chapitres suivants 3, 4 et 5.

3. Caractérisation préliminaire et analyses comparatives du gène *IgLip* codant une lipase hypothétique chez *Isochrysis galbana*

3.1. Analyse et traduction de la séquence nucléotidique *IgLip*

L'assemblage des régions 5' et 3', obtenues par RACE PCR et séquencées, forme une séquence nucléotidique finale de 2062 paires de bases (Figure 42). La recherche de codons d'initiation et de terminaison a été effectuée dans les six cadres de lecture. L'ADNc pleine longueur se termine par une queue polyadénylée et contient un cadre ouvert de lecture (ORF), encadré par deux régions non codantes 5' et 3' de tailles respectivement équivalentes à 339 et 352 pb. Deux codons ATG sont susceptibles d'initier la traduction. Ils figurent en 340-342^{ème} et 562-564^{ème} positions. Le codon stop TGA conclut la séquence codante en 1713^{ème} position. En considérant le premier codon ATG comme le codon d'initiation, l'ORF est de 1374 pb et code une protéine putative de 457 acides aminés. Dix résidus cystéine ont été répertoriés dans la séquence. Deux sites potentiels de glycosylation NCS et NSS sont placés en position 146 et 357.

```

1  GGTATCAACGCAGAGTACGCGGGGATCGCGGGGCTGTGCGGGTTCCTCCGGCGCGCGTGGCGGTGGTGCATTTCGAG
53  GATGATTTTCGAGCCCTGGCCCAAGCGGCGCCGACAGCAGTCCAGACGGGGGCTGCCGCGAGTCACTCTAGCTGG
128  CATTGTCACTGTAGCTTTGCTCTGGGTGGGGCGTATCCAGTCCGCCTCACTTGCCGGTGAAGCGGTTCTACGTC
203  ATCACCAGGAAGAAATATTGGGTCTACGGAGTGGGCGAGTGAACCCGGTGCAGCTGCCTCAAGAGGCCAATCAGAT
1  TTTTTCGAGCACGCGGAGTGGATGCGCCATCGAGCATCGGAGCGGGACCAGGCGGCCAAACATGCGCAGGCGGC
278  R R R R R
6  R R K R L S W R A T P W R S D L D S D L R T F E T
353  GCGCGGAAAAGAGGCTGTCTGGCGAGCCACGCCATGGCGCAGCGATTTGGATTCCGACCTGCGTACTTTTCGAAA
31  W R S C R P A D N E A L I A E R L R L A R R V W A
428  CGTGGCGCTCTGCCGGCCGCGAGACAATGAGGCGTTGATCGCCGAGCGTCTCCGCTTGCCCGCGGGTGTGGG
56  T V N A P L L L E L A A A L L C P P Q M V V R W T
503  CCACCGTCAATGCGCCGCTGCTGCTCGAGCTCGCGCGGCTCTGCTGTGCCACCACAAATGCTGCTTCGCTGGA
81  C S R C G G T G L Q P D P Q V I I A D G G S F L A
578  CGTGTTCACGCTGCGGAGGGACTGGCTTGACGCCGATCCGCAAGTCATCATCGCAGACGGCGGCTCCTTCCTCG
106  L V G I N E A E Q W V V V A F R G S L D G N L G A
653  CGTAGTGGGTATCAATGAAGCCGAACAGTGGGTGGTGGTGCCTTTCCGCGCTCTCTCGACGGCAACCTAGGGG
131  W G R N L Q C W T D P R S F H N C S G W V H H G W
728  CTTGGGGTCGCAACTTGAATGCTGGACCGACCCGAGGAGCTTCCACAATTGCTCCGGCTGGGTGCACCATGGCT
156  G S G Y D M L R A R L I T A V L L A L E A R P T Y
803  GGGGTAGTGGCTACGACATGCTTCGCGCGCCTCATCACGGCTGTGCTGCTAGCGCTTGAGGCTCGGCCGACGT
181  R L V V T G H S A G G A L A S I F V A E L A T V G
878  ACCGCCTCGTCGTCACCGCCACTCTGCCGGCGGGGCACTGGCCTCCATCTTGTGGCCGAGCTCGCCACCGTCG
206  A A Q S S S V A S E S S A Y R A E A G S D S K G E
953  GAGCAGCTCAATCATCTCGGTAGCCAGCGAATCGTCCGCTACCGCGCCGAGGCCGGCTCTGACTCCAAAGGCG
231  A D T K A A A A D E S A A A S S I A W E S V S E
1028  AAGCTGACACGAAGGCTGCCGCGGCTGATGAGTCCGCCGCGCCCTCCAGCTCCATCGCCTGGGAGAGTGTTCGG
256  V A P L L R R P L A P P P L Y I T F A A P M A G N
1103  AGGTTGCGCCGCTGCTCAGGCGACCGCTCGCGCCGCCCGCTGTACATCACCTTCGCCGCGCCGATGGCCGGGA
281  R E F A A A I D S G L A A Q G N V M L R V V N A R
1178  ACCGCGAGTTTGGCGCCGCATCGATTCGGCCTGGCCGCGCAGGGCAACGTGATGCTACGTGTGGTCAATGCAC
306  D P A V R A P N T L N A G G S V Q I P C P M G G E
1253  GCGATCCGGCCGTGCGTGACCCCAACCGTGAATGCGGGTGGCAGCGTCCAGATCCCCCTGCCAATGGGGGCG
331  D Q F S H A G Q E V W L T N A A D A D M S K A Q F
1328  AGGACCAGTTTTTCACATGCGGGGAGGAAGTGTGGCTGACAAAACGCTGCCGACGCCGACATGAGCAAGGCGCAAT
356  C N S S G A F G D G D D A S M S D P N P L C S S S
1403  TCTGCAACAGCTCAGGCGGTTCCGGGGACGGCGATGACGCCAGCATGTCCGACCCGAACCCGCTGTGCTCCAGCT
381  L G E S S L T W V H H N S Y L G V D T G A C P D T
1478  CCCTCGGCGAGTCGACCTGACTTGGTCCACCACAACCTACACTCGGCGTGCACAGGCGCGTCCCGGATA
406  M D A N Q W E E Y I K Y T K S L H P Q L R H V G Y
1553  CGATGGATGCGAATCAGTGGGAAGAGTACATCAAGTATAACAGAGTCTCCACCCGAGTTGAGACACGTCCGGCT
431  G E D A P D G G F D S T V Q R S E A E R V P T Y S
1628  ATGGTGAGGACGCACCGGACGGCGGTTTCGACAGCACGGTGCAGCGGTCCGAGGCGGAGCGGGTCCGACATACT
456  Y G *
1703  CGTACGGATGATGTCGGATGGTGCAGGGTGCCTTTTAGTTTTATTATGGCTTGGAGCTCCGGGCGTGTTCACA
1778  ACGCAGTTCTGCTCGTGTGGGCGGCGCCAACGCACGCATCGGCGACATTCACGCACCCACGCTGGCGCAACC
1853  AGCACC GCGTGGCACTGGAGATGGGTAGGAGAAGCGTGCATTGCGGCATGTGCGGCGCTGGCTGCGATGGGAGG
1928  TTCACCAGGCCAGACTCGCACACTGAGACTGGGCACAGCTGTGATGAGGGCTCTTGGACTTATGAGCTGTGAGC
2003  TTTAAGGGCGGCTGCGATTTCGTACGAGGGGTGTGTGGGCACTGAACACTCGTTTGGCGAAAAAACAACAAAAA
2078  AAAAAAAAAA

```

Figure 42 : Séquence nucléotidique de l'ADNc *IgLip* et traduction de la région codante.

Les amorces dessinées à partir de l'étiquette de séquence transcrite n°ISL00001844 et utilisées en RACE et Nested PCR sont écrites et soulignées en rouge. Les codons d'initiation et de terminaison de traduction sont indiqués en blanc sur fond rouge. Les sites potentiels de N-glycosylation de type NXS/T sont colorés en mauve. Les résidus cystéine sont soulignés en noir. Les régions non traduites 5' et 3' sont écrites en gras.

3.2. Prédiction des caractéristiques physicochimiques

3.2.1. Signal peptidique

Aucun signal peptidique n'a été détecté, avec le serveur SignalP, dans la région N-terminale de la séquence *IgLip*.

3.2.2. Masse moléculaire et pH du point isoélectrique

La séquence codante d'*IgLip* code une protéine putative de 457 acides aminés, de masse moléculaire théorique de 49,06 kDa et de pI théorique de 5,65.

3.2.3. Profils d'hydrophobicité et d'hydrophilicité

Le profil d'hydrophobicité de la protéine *IgLip* est représenté en jaune sur la Figure 43. On observe deux courts enchaînements d'acides aminés avec un indice faiblement hydrophobe (>1). L'absence de zones à fort indice d'hydrophobicité ($>1,8$) sur le profil de la molécule suggère que la protéine *IgLip* ne comporte pas de régions transmembranaires.

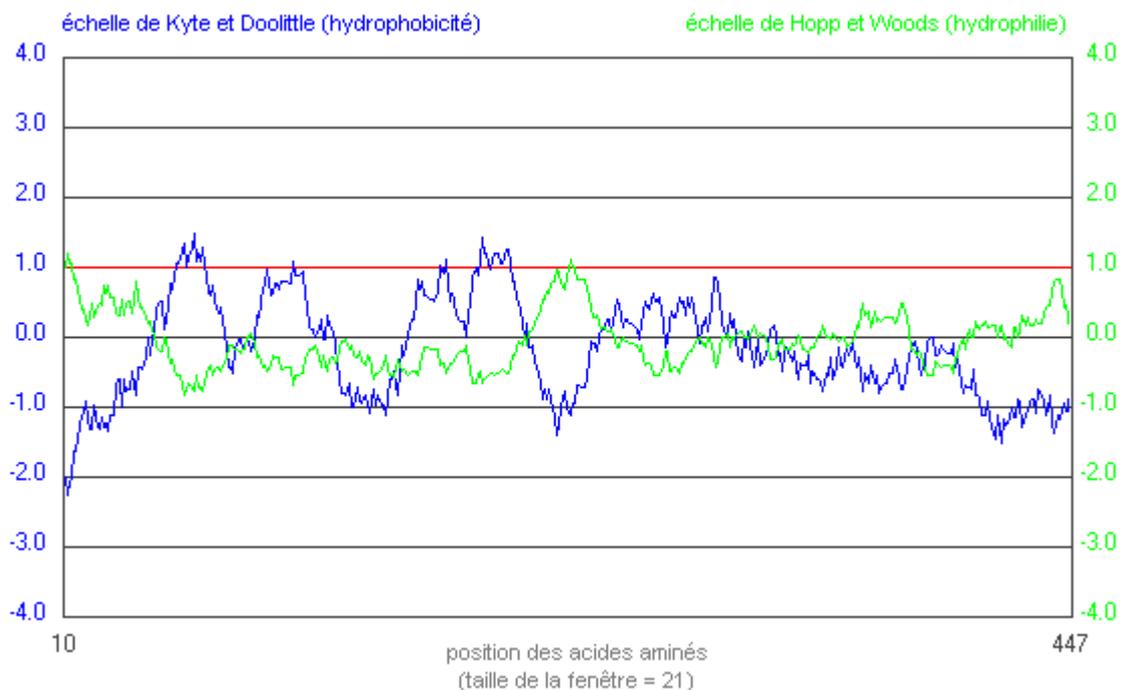


Figure 43 : Profils hydrophobe et hydrophile de la protéine *IgLip*.

Axe des ordonnées (bleu) : échelle de Kyte et Doolittle. Axe des ordonnées (vert) : échelle de Hopp et Woods. Axe des abscisses : position des acides aminés. La taille de la fenêtre est de 21 acides aminés.

3.3. Prédiction de la fonction enzymatique

3.3.1. Recherche de séquences homologues

Afin d'estimer la fonction de la protéine hypothétique d'*I. galbana*, la séquence *IgLip* a été utilisée pour identifier des séquences homologues déjà caractérisées chez d'autres organismes. Pour cela, *IgLip* a été comparé aux séquences répertoriées dans les banques de données publiques avec les programmes d'alignement Blastx et Blastp. Les résultats de cette analyse (non illustrés) montrent qu'*IgLip* s'aligne avec des lipases déjà caractérisées et avec un grand nombre de séquences annotées comme étant des homologues de lipase. Les meilleurs alignements obtenus impliquent des homologues de lipases isolés chez des végétaux supérieurs et chez des protistes trypanosomes. Afin de visualiser les régions identiques et similaires de la séquence protéique complète *IgLip* avec celles de ses plus proches homologues, des alignements multiples ont été réalisés avec le programme MUSCLE.

Homologie avec des lipases de plantes supérieures

L'alignement de la séquence protéique de la lipase hypothétique d'*I. galbana* avec trois séquences de lipases végétales est présentée en Figure 44. Les lipases examinées appartiennent au blé tendre, *Triticum aestivum* (*TaLip*), au riz, *Oryza sativa* (*OsLip*) et à l'arabette *Arabidopsis thaliana* (*AtLip*). Globalement, le taux de conservation observé est faible. Les régions N- et C- terminales, plus précisément les 50 premiers et derniers acides aminés, s'alignent difficilement. On constate, également, chez *I. galbana*, la présence d'une région qui ne s'apparie avec aucune de ses homologues végétales. Cette région, d'environ 60 résidus, est située au cœur de la protéine, dans la deuxième centaine de résidus.

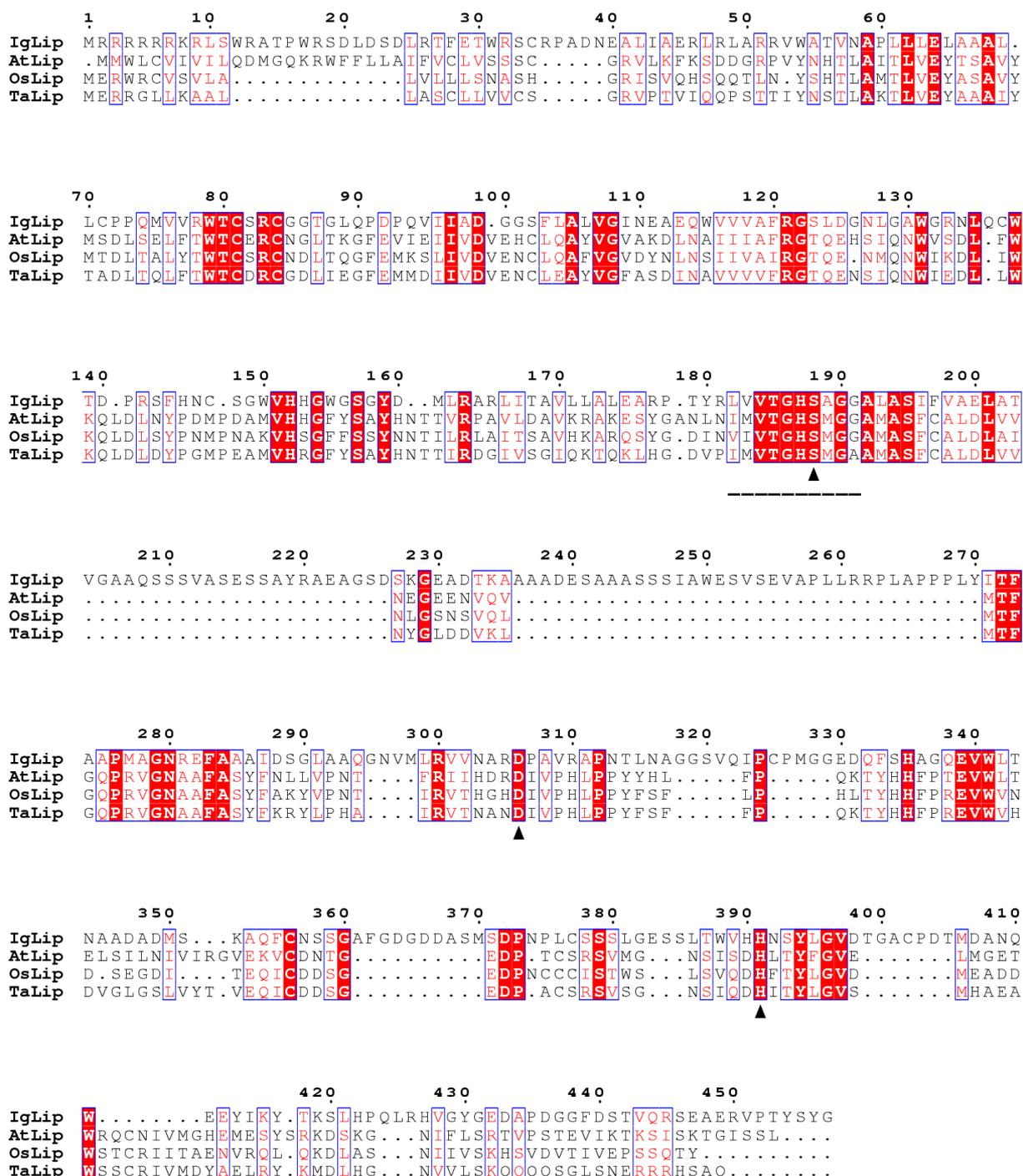


Figure 44 : Alignement de la lipase hypothétique d'*Isochrysis galbana* (IgLip) avec trois de ses homologues chez les plantes supérieures *Arabidopsis thaliana* (AtLip), *Oryza sativa* (OsLip) et *Triticum aestivum* (TaLip).

Organisme / numéro d'accèsion : *Triticum aestivum* / CAD32696 ; *Oryza sativa* / NP_001060581 ; *Arabidopsis thaliana* / NP_197365. Les résidus strictement conservés sont indiqués en blanc sur fond rouge, les résidus bien conservés en rouge et encadrés en bleu. Les triangles noirs indiquent l'emplacement des trois résidus Ser/Asp/His supposés constituer la triade catalytique. La région conservée autour de la sérine nucléophile est soulignée par un trait noir discontinu.

Le pourcentage d'identité, c'est-à-dire le pourcentage de résidus strictement conservés, atteint 18 % avec *AtLip*, 20 % avec *OsLip* et 19 % avec *TaLip* (Tableau 14). Le pourcentage de similitude correspond aux résidus similaires c'est-à-dire aux acides aminés de mêmes propriétés biochimiques. Le taux de similarité est plus élevé que le taux d'identité. Il varie entre 30 et 35 %. L'homologie de séquence est plus importante entre les trois séquences des végétaux. Le taux d'identité atteint, par exemple, 53 % entre la lipase d'*O. sativa* et celle de *T. aestivum*.

	<i>AtLip</i>	<i>OsLip</i>	<i>TaLip</i>
<i>IgLip</i>	18 % 32 %	20 % 35 %	19 % 30 %
<i>AtLip</i>		49 % 67 %	50 % 68 %
<i>OsLip</i>			53 % 69 %

Tableau 14 : Pourcentages d'identité et de similitude entre la lipase hypothétique d'*Isochrysis galbana* (*IgLip*) et ses homologues végétales chez *Arabidopsis thaliana* (*AtLip*), *Oryza sativa* (*OsLip*) et *Triticum aestivum* (*TaLip*).

Le pourcentage indiqué en blanc sur fond rouge correspond au pourcentage d'identité c'est-à-dire au pourcentage de résidus strictement conservés entre deux séquences. Les chiffres écrits en rouge indiquent les pourcentages de similitude, c'est-à-dire le pourcentage de résidus biochimiquement proches.

Bien que les quatre séquences protéiques soient, dans leur ensemble, peu conservées, elles possèdent toutes les résidus censés intervenir dans le mécanisme catalytique des lipases. La triade catalytique Ser/Asp/His (S/D/H) est indiquée par des triangles noirs sur la Figure 44. Dans la séquence *IgLip*, la région comprise entre les résidus L182 et G191 s'aligne très bien avec celles de ses homologues végétales. Cette région correspondrait au coude nucléophile car elle inclut le pentapeptide consensus G-X-S-X-G. La région est soulignée par des pointillés noirs sur la Figure 44. Chez *I. galbana* et les trois plantes, cette région concorde parfaitement avec la région consensus des lipases définie dans PROSITE (PS00120) [LIV]-X-[LIVFY]-[LIVMST]-G-[HYWV]-S-X-G-[GSTAC] (Tableau 15).

Motif consensus	[LIV]	X	[LIVFY]	[LIVMST]	G	[HYWV]	<u>S</u>	X	G	[GSTAC]
<i>IgLip</i>	L	V	V	T	G	H	S	A	G	G
<i>AtLip</i>	I	M	V	T	G	H	S	M	G	G
<i>OsLip</i>	V	I	V	T	G	H	S	M	G	G
<i>TaLip</i>	I	M	V	T	G	H	S	M	G	A

Tableau 15 : Conservation de la région du coude nucléophile chez la lipase hypothétique d'*Isochrysis galbana* (*IgLip*) et les lipases homologues d'*Arabidopsis thaliana* (*AtLip*), *Oryza sativa* (*OsLip*) et *Triticum aestivum* (*TaLip*).

Comparaison avec le motif consensus des lipases référencé PS00120 sous PROSITE.

Homologie avec des lipases de trypanosomes

La lipase hypothétique d'*I. galbana* s'apparie également avec des lipases de trypanosomes, identifiées, elles aussi, par analogie de séquence. Les trypanosomes sont des protistes de la famille des *Trypanosomatidae*. Ce sont, au même titre que les microalgues, des organismes eucaryotes unicellulaires. Ils constituent néanmoins un groupe de protistes, les excavates, phylogénétiquement distinct de celui des chromalvéolates (Keeling *et al.*, 2005). Contrairement aux microalgues, les trypanosomes ne sont pas photoautotrophes. Ils appartiennent à la famille des kinétoplastides et ce sont des organismes parasites, souvent pathogènes pour l'homme.

En dehors des extrémités N- et C- terminales et de la zone d'insertion d'environ 60 résidus présente chez *I. galbana*, l'alignement semble fiable (Figure 45). Il révèle la présence de cinq résidus cystéine conservés (indiqués par des points verts sur la Figure 45). Les régions, susceptibles de contenir les résidus impliqués dans l'action catalytique des lipases, apparaissent également conservées. Les motifs caractéristiques des lipases, à savoir la triade catalytique Ser/Asp/His et le coude nucléophile de type G-X-S-X-G, sont présents.

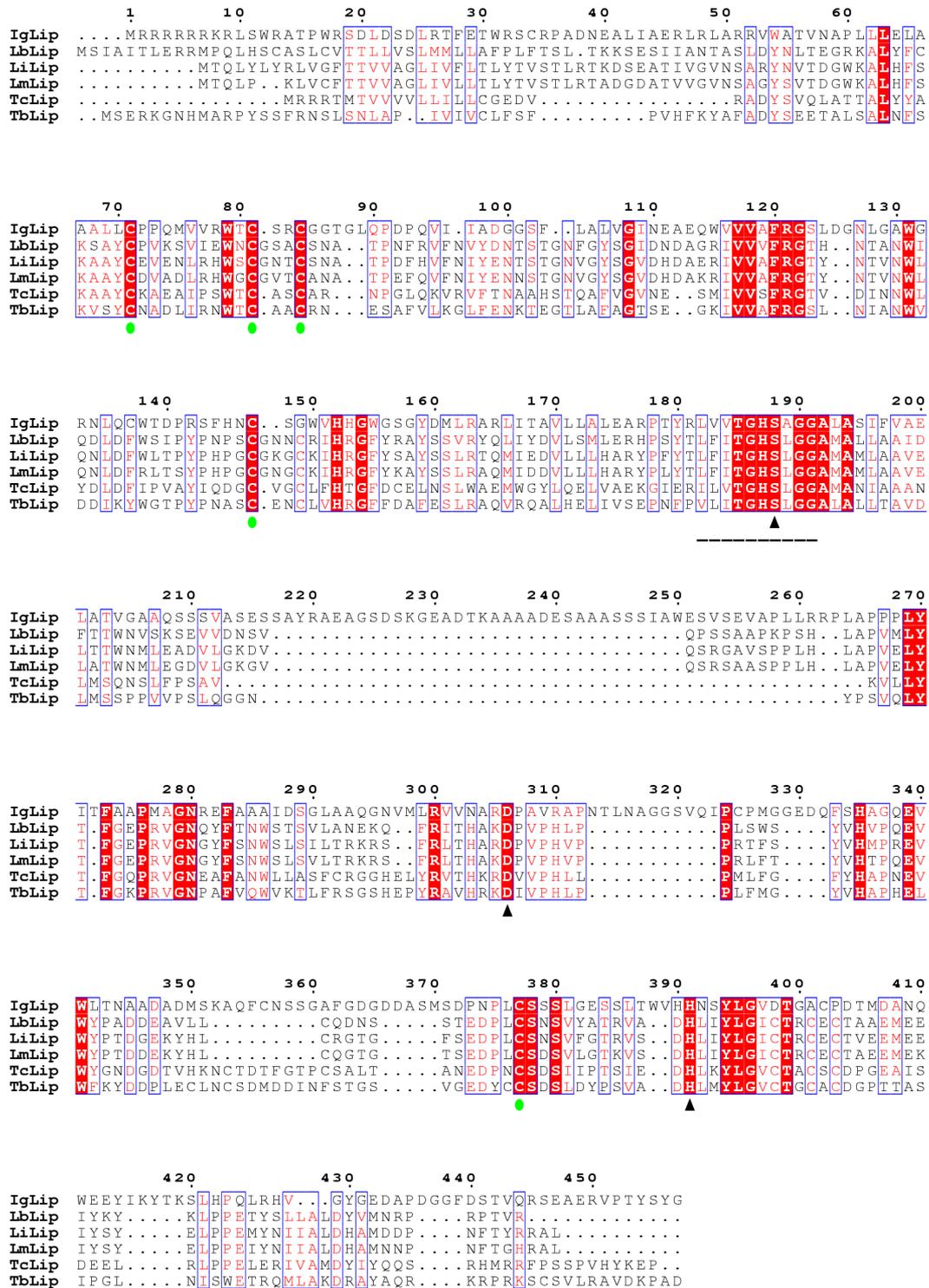


Figure 45 : Alignement de la lipase hypothétique d'*Isochrysis galbana* (IgLip) avec cinq de ses homologues chez les trypanosomes, *Leishmania braziliensis* (LbLip), *Leishmania infantum* (LiLip), *Leishmania major* (LmLip), *Trypanosoma cruzi* (TcLip) et *Trypanosoma brucei* (TbLip).

Organisme / numéro d'accèsion : *L. braziliensis* / XP_001567267 ; *L. infantum* / XP_001467571 ; *L. major* / CAJ08439 ; *T. cruzi* / XP_804459 ; *T. brucei* / XP_847573. Les résidus strictement conservés sont indiqués en blanc sur fond rouge, les résidus bien conservés en rouge et encadrés en bleu. Les triangles noirs indiquent l'emplacement des trois résidus Ser/Asp/His supposés constituer la triade catalytique. La région conservée autour de la sérine nucléophile est soulignée par un trait noir discontinu. Les résidus cystéine conservés sont signalés par un point vert.

Les meilleurs alignements ont été obtenus avec trois espèces du genre *Leishmania*. Les lipases des espèces *braziliensis* (*LbLip*), *infantum* (*LiLip*) et *major* (*LmLip*) présentent respectivement 21, 21 et 20 % d'identité avec *IgLip* et les taux de similarité varient de 29 à 35 % (Tableau 16). Il apparaît donc que le taux d'identité entre la séquence protéique d'*IgLip* et les différentes séquences homologues étudiées ne dépasse pas 21 %.

	<i>LbLip</i>	<i>LiLip</i>	<i>LmLip</i>	<i>TcLip</i>	<i>TbLip</i>
<i>IgLip</i>	21 % 34 %	21 % 35 % 74 %	20 % 33 % 57 % 72 % 85 % 92 %	18 % 29 % 28 % 45 % 32 % 45 % 33 % 46 %	17 % 30 % 30 % 47 % 29 % 45 % 30 % 44 % 35 % 51 %
<i>LbLip</i>					
<i>LiLip</i>					
<i>LmLip</i>					
<i>TcLip</i>					

Tableau 16 : Pourcentages d'identité et de similitude entre la lipase hypothétique d'*Isochrysis galbana* (*IgLip*) et ses homologues chez les trypanosomes, *Leishmania braziliensis* (*LbLip*), *Leishmania infantum* (*LiLip*), *Leishmania major* (*LmLip*), *Trypanosoma cruzi* (*TcLip*) et *Trypanosoma brucei* (*TbLip*).

Le pourcentage indiqué en blanc sur fond rouge correspond au pourcentage d'identité, c'est-à-dire au pourcentage de résidus strictement conservés entre deux séquences. Les chiffres écrits en rouge indiquent la proportion de résidus similaires, c'est-à-dire la proportion de résidus biochimiquement proches.

La région conservée autour de la sérine catalytique est conforme au motif caractéristique des lipases indexé PS00120 sous PROSITE (Tableau 17).

Motif consensus	[LIV]	X	[LIVFY]	[LIVMST]	G	[HYWV]	<u>S</u>	X	G	[GSTAC]
<i>IgLip</i>	L	V	V	T	G	H	S	A	G	G
<i>LbLip</i>	L	F	I	T	G	H	S	L	G	G
<i>LiLip</i>	L	F	I	T	G	H	S	L	G	G
<i>LmLip</i>	L	F	I	T	G	H	S	L	G	G
<i>TcLip</i>	I	L	V	T	G	H	S	L	G	G
<i>TbLip</i>	V	L	I	T	G	H	S	L	G	G

Tableau 17 : Conservation de la région du coude nucléophile de la lipase hypothétique d'*Isochrysis galbana* (*IgLip*) et des lipases homologues de trypanosomes *Leishmania braziliensis* (*LbLip*), *Leishmania infantum* (*LiLip*), *Leishmania major* (*LmLip*), *Trypanosoma cruzi* (*TcLip*) et *Trypanosoma brucei* (*TbLip*).

Comparaison avec le motif consensus des lipases référencé PS00120 sous PROSITE.

3.3.2. Recherche de motifs fonctionnels

Pour rechercher la présence de motifs protéiques conservés caractéristiques d'une famille de protéines, la méthode de comparaison par profil a été utilisée *via* le programme Blocksearcher. Les résultats obtenus montrent que la séquence protéique d'*IgLip* présente des similitudes significatives avec 1 des 6 blocs qui caractérisent la famille Lipase 3 indexée IPB005592 dans la base de données Blocks (Annexe 8). Cette famille de lipases regroupe, à ce jour, essentiellement des lipases de champignons. Le bloc conservé correspond au bloc IPB005592D. Il coïncide avec la région du coude nucléophile. Les 5 autres blocs sont un peu plus éloignés des profils définis et plus faiblement conservés en raison de la présence de gaps, c'est-à-dire d'insertions ou de délétions. Cependant, on retrouve, dans la séquence *IgLip*, les acides aminés les plus fréquemment identifiés et donc les plus fortement conservés de ces cinq blocs. *IgLip* contient, par exemple, les résidus F¹⁹-R-G-S²² du bloc IPB005592B de la position 120 à 123, les résidus R14, D20 et P26 du bloc IPB005592E en position 300, 306, 312 et les résidus H2 et Y5 du bloc IPB005592F en positions 391 et 394.

3.4. Prédiction de l'organisation structurale

L'alignement de la lipase d'*I. galbana* avec des homologues de lipases, elles-mêmes identifiées et annotées à partir d'alignements de séquences, fournit seulement une estimation générale de la fonction enzymatique. Pour affiner cette estimation, obtenir une prédiction fonctionnelle plus fiable et repérer les résidus jouant probablement un rôle fonctionnel et structural important, la séquence *IgLip* a été comparée à celles de lipases complètement caractérisées. Les lipases complètement caractérisées correspondent à des lipases dont les relations structure/fonction biochimique ont été élucidées expérimentalement par le clonage, la surexpression du gène et les mesures d'activité de l'enzyme purifiée, par des analyses de cristallographie de l'enzyme seule ou complexée avec un inhibiteur et par des études de mutagenèse dirigée.

Parmi les lipases caractérisées les plus proches de celle d'*I. galbana*, figurent les lipases des champignons filamenteux *Thermomyces lanuginosa* (Ascomycète) et *Rhizomucor miehei* (Zygomycète). Ces deux lipases ont fait l'objet de nombreuses études structurales et fonctionnelles (Brady *et al.*, 1990; Derewenda *et al.*, 1992b; Holmquist *et al.*, 1994; Derewenda *et al.*, 1994a; Martinelle *et al.*, 1996; Peters *et al.*, 1998; Brzozowski *et al.*, 2000).

Leur conformation tridimensionnelle est connue. Elles cumulent respectivement 20 et 4 entrées PDB. Les travaux cristallographiques ont permis de comprendre leur assemblage, de connaître les résidus impliqués dans le mécanisme catalytique et dans les interactions avec le substrat.

Les alignements ont été conçus avec MUSCLE et ClustalW. La comparaison de la séquence protéique *IgLip* avec celles des lipases de *Thermomyces lanuginosa* (*TLip*) (Figure 46, page 137) et de *Rhizomucor miehei* (*RmLip*) (Figure 47, page 138) suggère que l'organisation structurale nécessaire à la mise en place de l'activité catalytique est conservée chez *I. galbana* et que la microalgue posséderait des spécificités structurales.

La lipase *IgLip* présente environ 15 % d'identité et entre 24 et 26 % de similitude avec ces deux homologues fongiques. L'homologie n'est pas retrouvée sur toute la longueur de la séquence, surtout dans les régions N- et C- terminales et au niveau de la zone d'insertion, d'environ 60 acides aminés, présente chez *I. galbana*. Ce fragment d'une soixantaine de résidus n'a d'ailleurs été rencontré chez aucun autre organisme. L'homologie de séquence est retrouvée principalement autour des sites catalytiques. Aucun signal de sécrétion n'a été détecté dans la région N-terminale d'*IgLip*.

Les résultats confirment que les résidus Ser188, Asp306 et His391, seraient les résidus composant la triade catalytique d'*IgLip*. La triade potentielle d'*IgLip* s'aligne, en effet, avec les résidus catalytiques Ser168/Asp223/His280 de *TLip* et Ser238/Asp297/His351 de *RmLip*. Toutefois cette indication est à considérer avec prudence compte tenu de la taille de la zone d'insertion désappariée. L'éloignement des acides aminés Ser188 et Asp306 influence probablement le repliement de la protéine pour permettre leur rapprochement. Cette variabilité a probablement une incidence sur les spécificités fonctionnelles de l'enzyme hypothétique d'*I. galbana*.

La séquence consensus qui caractérise le coude nucléophile des lipases est également très conservée. Cette région comprend chez *I. galbana* les 10 acides aminés L-V-V-T-G-H-S¹⁸⁸-A-G-G (Figure 45). Elle s'aligne quasi parfaitement avec le coude nucléophile des lipases fongiques et n'en diffère que par 3 résidus. Les coudes des lipases *TLip* et *RmLip* correspondent respectivement aux séquences V-V-F-T-G-H-S¹⁶⁸-L-G-G (Figure 46) et V-A-V-T-G-H-S²³⁸-L-G-G (Figure 47). Parmi toutes les combinaisons possibles de la séquence consensus [LIV]-X-[LIVFY]-[LIVMST]-G-[HYWV]-S-X-G-[GSTAC], celle d'*IgLip* n'a

jamais été complètement retrouvée chez d'autres organismes, référencés à ce jour. Les motifs les plus proches ont été trouvés chez des lipases et des protéines non identifiées de bactéries, de plantes, de champignons et de microalgues. Ces motifs présentent au mieux 9 résidus strictement conservés sur 10.

Parmi les deux résidus connus comme étant impliqués dans le trou oxyanionique, c'est-à-dire dans la stabilisation de l'intermédiaire oxyanion, seule la sérine (S), positionnée en 105 chez *TLLip* et en 176 chez *RmLip*, est conservée chez *IgLip* en position 123. L'autre résidu, mis en jeu dans cette interaction, correspond, chez les deux champignons, à la leucine (L), située à droite de la sérine catalytique. La leucine est un résidu hydrophobe non polaire. Chez la microalgue, la leucine est remplacée par l'alanine qui présente les mêmes propriétés physicochimiques. Ces deux résidus diffèrent par leur chaîne latérale. L'alanine présente un encombrement stérique plus faible. Cette dissemblance est peut-être à l'origine d'une modification de la conformation spatiale de la poche abritant le site actif, ce qui pourrait affecter la spécificité de substrat.

Chez *T. lanuginosa*, les résidus G104, H167, P196, P229 et P230 (indiqués par des triangles verts sur la Figure 46) sont tous impliqués dans la fixation du substrat et de ses intermédiaires catalytiques. Ils sont tous conservés chez *I. galbana* et correspondent aux résidus G122, H187, P276, P312, P324 de la protéine. Ils sont également strictement conservés chez *R. miehei* et chez les homologues de plantes et de protistes étudiés précédemment.

Les études cristallographiques portant sur les lipases *TLLip* et *RmLip* ont mis en évidence l'existence d'un volet amphiphile constitué de 12 acides aminés pour *TLLip* et de 14 pour *RmLip*. Seuls deux acides aminés, le tryptophane (W111 *TLLip*, W182 *RmLip*) et la leucine (L115 *TLLip*, L186 *RmLip*) sont conservés chez *I. galbana* en position 131 et 134. Les lipases de champignons filamenteux ont, comme les lipases pancréatiques, un résidu tryptophane conservé au centre du volet amphiphile (Derewenda *et al.*, 1994b). L'importance du rôle joué par le tryptophane du volet de la lipase de *T. lanuginosa* a été démontrée par des études de mutagenèse dirigée (Martinelle *et al.*, 1996). La substitution du Trp par les résidus Phe, Leu, Gly ou Glu entraîne une modification de la spécificité de l'enzyme par rapport au substrat, et plus précisément par rapport à la longueur de chaîne des acides gras des triglycérides.

Les six résidus cystéine à l'origine de trois ponts disulfures chez *RmLip* s'alignent correctement avec six des dix cystéines observées chez *IgLip*.

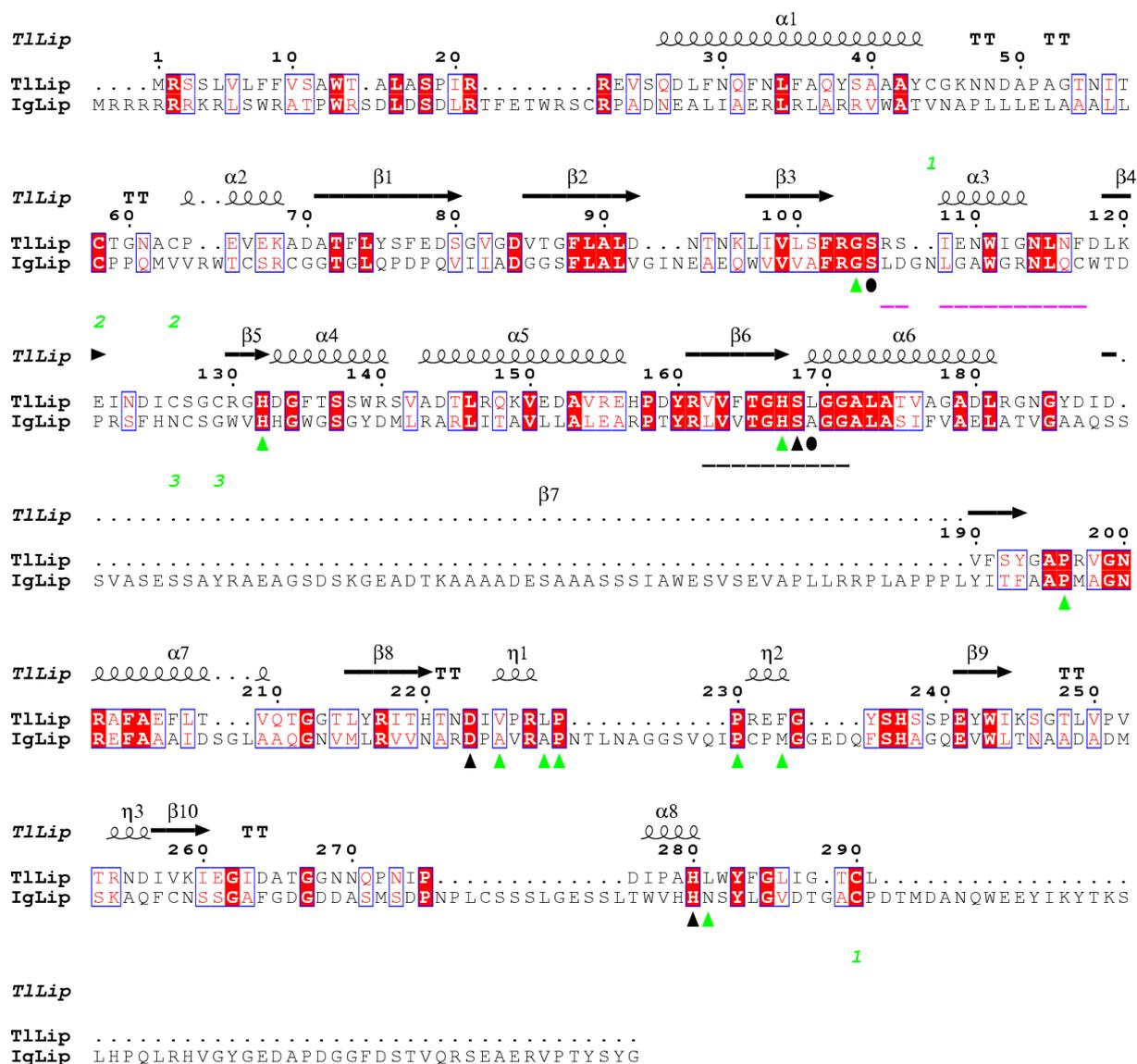


Figure 46 : Alignement de la lipase hypothétique d'*Isochrysis galbana* (*IgLip*) avec la lipase de *Thermomyces lanuginosa* (*TlLip*) (Swiss-Prot O59952 ; code PDB 1dt3).

Les résidus strictement conservés sont indiqués en blanc sur fond rouge, les résidus bien conservés en rouge et encadrés en bleu. La triade catalytique Ser/Asp/His est indiquée par des triangles noirs. La région du coude nucléophile incluant le pentapeptide consensus G-X-S-X-G est soulignée en noir. La région du volet amphiphile est soulignée en rose. Les triangles verts et les ronds noirs correspondent aux résidus impliqués dans la fixation du substrat et de ses intermédiaires. Les résidus cystéine et les ponts disulfure sont chiffrés en vert. La structure secondaire de *TlLip* figure au-dessus de l'alignement : les hélices et les feuillets sont schématisés respectivement par des hélices et des flèches.

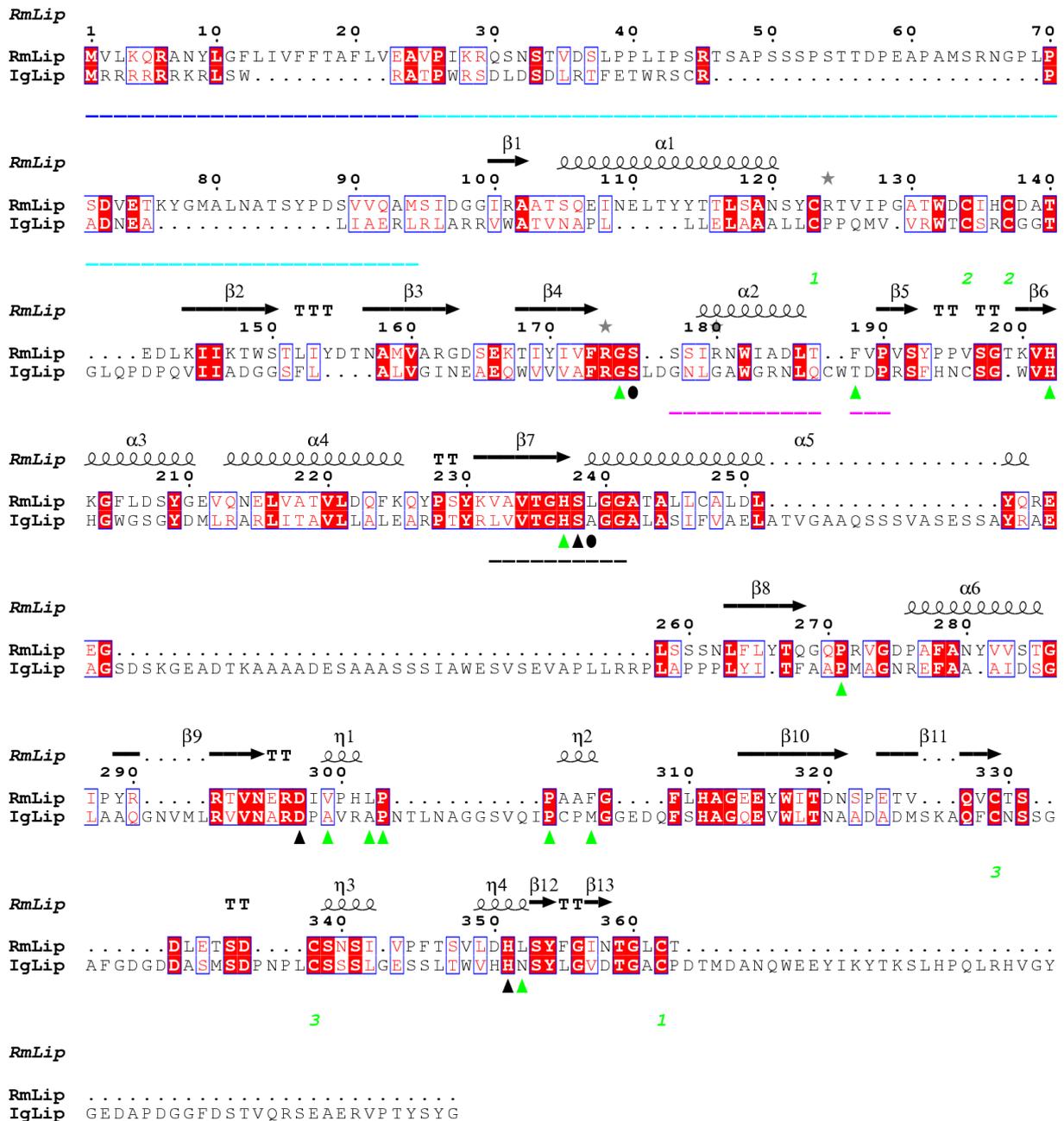


Figure 47 : Alignement de la lipase hypothétique d'*Isochrysis galbana* (*IgLip*) avec la lipase de *Rhizomucor miehei* (*RmLip*) (Swiss-Prot P19515 ; code PDB 4tgl).

Les résidus strictement conservés sont indiqués en blanc sur fond rouge, les résidus bien conservés en rouge et encadrés en bleu. La triade catalytique Ser/Asp/His est indiquée par des triangles noirs. La région du coude nucléophile incluant le pentapeptide consensus G-X-S-X-G est soulignée en noir. La région du volet amphiphile est soulignée en rose. Les triangles verts et les ronds noirs correspondent aux résidus impliqués dans la fixation du substrat et de ses intermédiaires. Les résidus cystéine et les ponts disulfure sont chiffrés en vert. La structure secondaire de *RmLip* figure au-dessus de l'alignement : les hélices et les feuillets sont schématisés respectivement par des hélices et des flèches.

3.5. Séquences homologues à *IgLip* chez les microalgues

La séquence d'*IgLip* a été utilisée pour rechercher des séquences homologues dans les banques de données publiques de microalgues dont le génome a été entièrement séquencé. Cette analyse a permis de répertorier 13 séquences homologues à *IgLip* chez les microalgues.

Dans le génome de l'algue rouge unicellulaire *Cyanidioschyzon merolae* (Matsuzaki *et al.*, 2004), un homologue de lipase, dont la séquence protéique est référencée CMH156C, a été identifié. La séquence nucléotidique de cet homologue, dénommé arbitrairement *CmLip*, contient un cadre ouvert de lecture, de taille équivalente à 1641 pb, et qui code pour une protéine de 577 acides aminés.

Quatre séquences homologues sont également référencées dans les banques génomiques de deux diatomées. Chez la diatomée *Thalassiosira pseudonana* (Armbrust *et al.*, 2004), deux homologues d'*IgLip* ont été identifiés et nommés *TpLip1* et *TpLip2*. Dans la banque, leurs identifiants protéiques (id.) sont respectivement numérotés 24111 et 7093. Les transcrits *TpLip1* et *TpLip2* sont constitués respectivement de 1827 et 2583 nucléotides. Le premier code pour un polypeptide de 568 acides aminés et le second pour 861 acides aminés. Le transcrit *TpLip2* résulte de l'assemblage de quatre exons de 563, 1002, 97 et 921 pb. Dans le génome de *Phaeodactylum tricornutum* (Scala *et al.*, 2002), il existe trois gènes homologues à des lipases (*PtLip1*, *PtLip2* et *PtLip3*). La taille du premier transcrit en paire de base est égale à 2676. La protéine correspondante est composée de 891 résidus et elle est indexée sous le numéro 41346. Le second gène de 1431 nucléotides contient un intron de 84 nucléotides et correspond donc à un transcrit de 1347 nucléotides. Sa traduction donne une protéine de 448 acides aminés (id. : 32834). Enfin le dernier gène, sans intron, dont la séquence codante est de 1080 pb, code pour une protéine de 360 acides aminés (id. : 37711).

Chez *Chlamydomonas reinhardtii* (Asamizu *et al.*, 1999), algue verte unicellulaire de la classe des Chlorophyceae, deux séquences similaires à des lipases ont été identifiées (*CrLip1* et *CrLip2*). Les deux protéines 190433 et 142444 sont constituées l'une de 571 et l'autre de 527 acides aminés. Elles correspondent chacune à la traduction de séquences codantes de 1671 et 1581 pb. Les deux gènes codant pour ces protéines sont constitués de nombreux introns.

Les banques génomiques de deux autres microalgues vertes du genre *Ostreococcus* et de la classe des *Prasinophyceae* ont également été consultées (Derelle *et al.*, 2006; Palenik *et al.*, 2007). Le génome d'*Ostreococcus lucimarinus* comporte deux gènes sans intron codant probablement pour des lipases. Leurs séquences codantes de 2019 et 2085 nucléotides peuvent être traduites en protéines de 673 et 695 acides aminés (id. : 33194 *OILip1* et 24393 *OILip2*). Chez *Ostreococcus tauri*, il existe quatre séquences homologues de lipase (*OtLip1* à *OtLip4*). Les quatre gènes codent pour des enzymes de 342 (id. : 36169), 290 (id. 30127), 689 (id. 10623) et 811 (id. 34430) acides aminés. Seule la protéine 34430 est codée par un gène constitué de 3 introns de tailles égales à 130, 158 et 64 pb.

L'alignement de séquences protéiques, généré avec MUSCLE, montre que ces homologues partagent peu d'identité et de similitude avec *IgLip* (Figure 48). Les meilleurs alignements d'*IgLip* ont été obtenus avec la séquence *PtLip3* de la microalgue *Phaeodactylum tricornutum* (16 % d'identité et 27 % de similitude) et avec la séquence *OtLip2* de la microalgue *Ostreococcus tauri* (18 % d'identité et 28 % de similitude). Bien que les séquences protéiques de microalgues partagent un faible taux de conservation des acides aminés, elles possèdent toutes les motifs catalytiques des lipases référencées dans les banques de séquences. Le pentapeptide hautement conservé chez les lipases G-X-S-X-G est souligné en noir sur la Figure 48. La région qui entoure ce pentapeptide consensus apparaît relativement bien conservée.

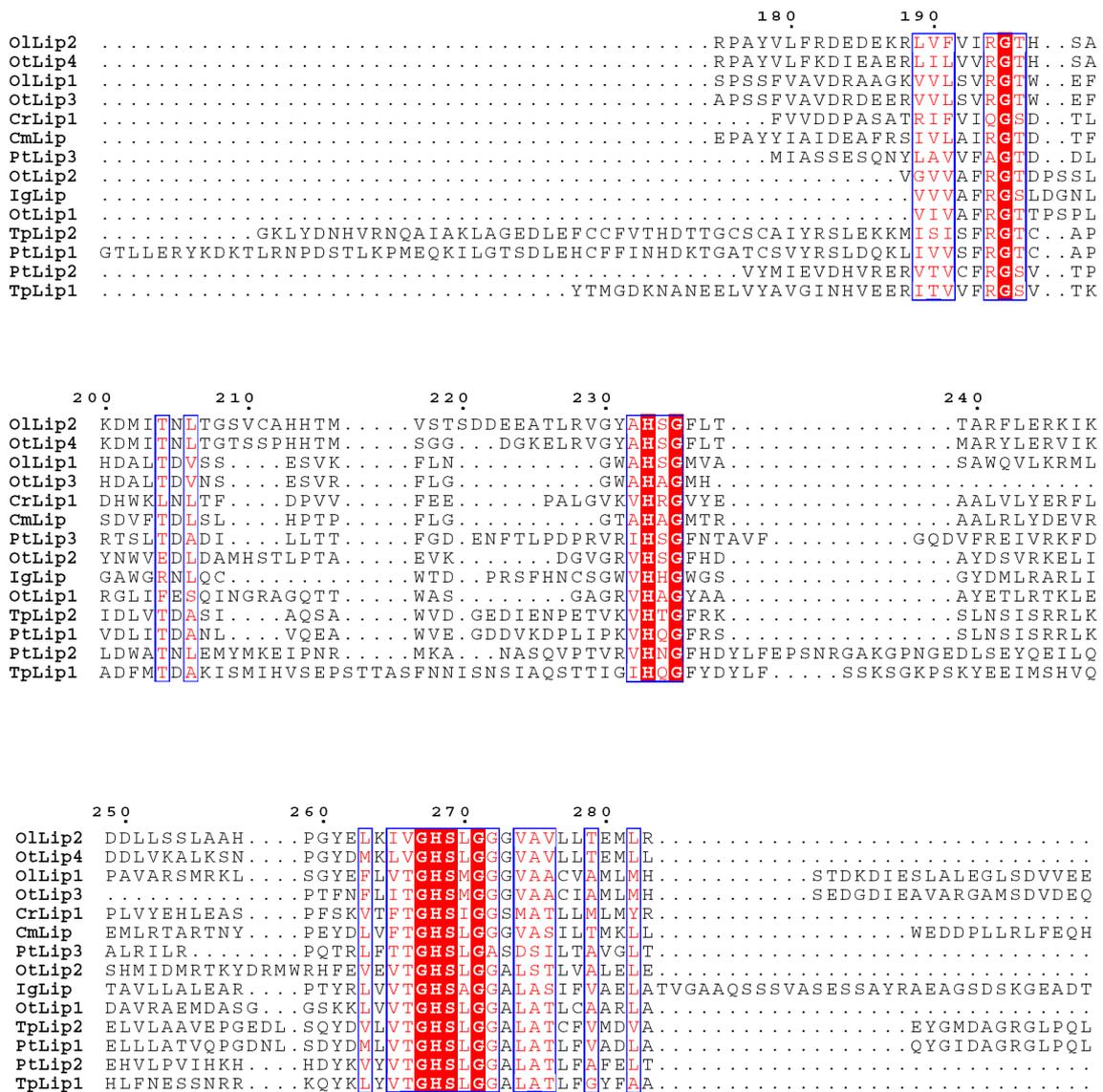


Figure 48 : Alignement de la lipase hypothétique d'*Isochrysis galbana* (IgLip) avec des homologues de lipase chez les microalgues *Cyanidioschyzon merolae* (CmLip), *Thalassiosira pseudonana* (TpLip1 et TpLip2), *Phaeodactylum tricornutum* (PtLip1, PtLip2 et PtLip3), *Chlamydomonas reinhardtii* (CrLip1 et CrLip2), *Ostreococcus lucimarinus* (Ollip1 et Ollip2) et *Ostreococcus tauri* (Otlip1, Otlip2, Otlip3 et Otlip4).

Organisme / numéro d'identifiant : *C. merolae* / CMH156C ; *T. pseudonana* / 24111 (TpLip1) / 7093 (TpLip2) ; *P. tricornutum* / 41346 (PtLip1) / 32834 (PtLip2) / 37711 (PtLip3) ; *C. reinhardtii* / 190433 (CrLip1) / 142444 (CrLip2) ; *O. lucimarinus* / 33194 (Ollip1) / 24393 (Ollip2) ; *O. tauri* / 36169 (Otlip1) / 30127 (Otlip2) / 10623 (Otlip3) / 34430 (Otlip4). Les résidus strictement conservés sont indiqués en blanc sur fond rouge, les résidus bien conservés en rouge et encadrés en bleu. Le triangle noir indique l'emplacement du résidu Ser supposé appartenir à la triade catalytique Ser/Asp/His. La région conservée autour de la sérine nucléophile est soulignée par un trait noir discontinu.

La banque EST de la microalgue *Emiliania huxleyi* a également été consultée pour la recherche de transcrits homologues à *IgLip*. *Emiliania huxleyi* est une microalgue haptophyte phylogénétiquement très proche d'*I. galbana*. Un transcrit de 111 pb référencé sous le n° d'accèsion EG035443, a été identifié. Compte tenu de sa taille, ce transcrit est incomplet. Il possède un signal de terminaison de traduction TAG au 108^{ème} codon. Il est incomplet du côté N-terminal. Le polypeptide obtenu après traduction du transcrit s'aligne relativement bien avec la région C-terminale d'*IgLip* (Figure 49). Jusqu'à présent, toutes les séquences homologues identifiées s'alignaient très mal dans leur région C-terminale avec celle d'*IgLip*. Cette fois ci le taux d'identité est plus élevé et atteint sur cette portion 39 %. Le résidu catalytique histidine est présent en position 94 et conservé chez *E. huxleyi*. Ces observations nous conduisent à penser que l'étiquette de séquence exprimée identifiée chez *E. huxleyi* coderait pour une lipase proche de celle d'*I. galbana*. Cependant, l'acide aspartique (position D9 sur la Figure 49) précédemment identifié comme étant un des trois résidus catalytiques chez *I. galbana* n'est pas conservé dans la séquence d'*E. huxleyi*. Une erreur de séquençage du transcrit EG035443 est envisageable compte tenu du taux d'erreur de séquençage important des banques EST.

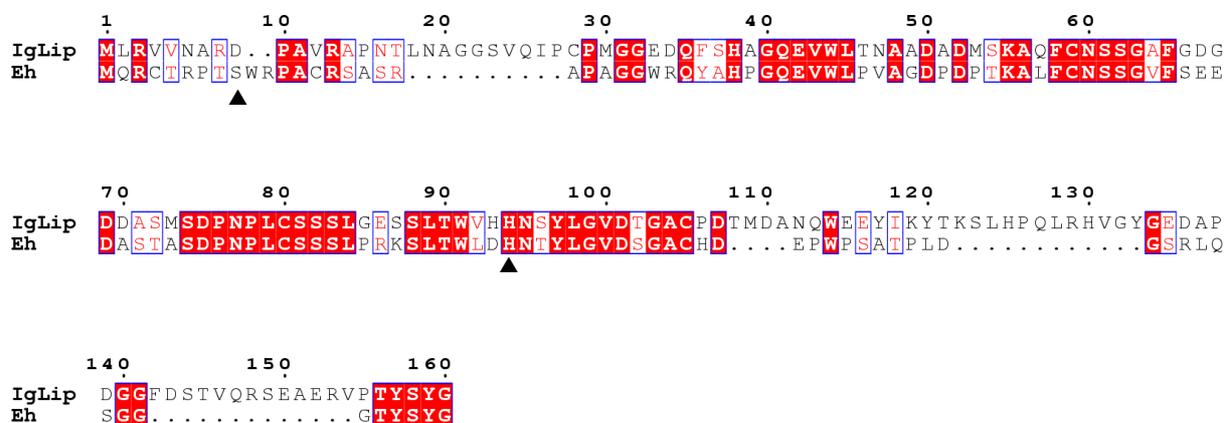


Figure 49 : Alignement de la région C-terminale de la lipase hypothétique d'*Isochrysis galbana* (*IgLip*) avec le marqueur EG035443 de la microalgue *Emiliania huxleyi*.

Les résidus strictement conservés sont indiqués en blanc sur fond rouge, les résidus bien conservés en rouge et encadrés en bleu. Les triangles noirs indiquent l'emplacement des résidus Asp et His supposés appartenir à la triade catalytique Ser/Asp/His.

3.6. Classification d'*IgLip* et de ses homologues

Pour compléter les analyses structurales et positionner les séquences nouvellement identifiées chez *I. galbana* et chez les autres microalgues, les séquences protéiques ont été comparées

aux séquences d'enzymes lipolytiques classées dans la Lipase Engineering Database (LED) de l'Université de Stuttgart. Les meilleurs alignements ont tous été obtenus avec les séquences classées dans la sous famille des lipases similaires à celle de *Rhizomucor miehei* (référence LED : abH23.01). Cette sous famille appartient à la superfamille des lipases des champignons filamenteux (abH23).

3.7. Structure intron/exon du gène *IgLip*

Pour déterminer l'organisation du gène, c'est-à-dire la présence d'éventuels introns au sein du gène, l'ADN génomique de la microalgue a été isolé, purifié puis soumis à une PCR avec le couple d'amorces *EcoRIIip/LipXbaI*. L'électrophorèse des produits PCR sur gel d'agarose 1 % (w/v) a permis la visualisation d'une bande de taille supérieure à celle de la séquence codante de l'ADNc. Après séquençage, l'alignement de la séquence d'ADN génomique avec celle du transcrit a confirmé la présence de 2 introns de tailles équivalentes à 173 pb et 121 pb (Figure 50 et Annexe 9). Le transcrit résulte donc de l'assemblage de trois exons de 341, 823 et 210 nucléotides. Les deux introns, à l'image de ceux des gènes eucaryotes, débutent par le couple de nucléotides GT. La jonction intron/exon, AG/GT, hautement conservée chez les eucaryotes, est également présente. La séquence génomique du gène *IgLip* présente une taille de 1668 pb.

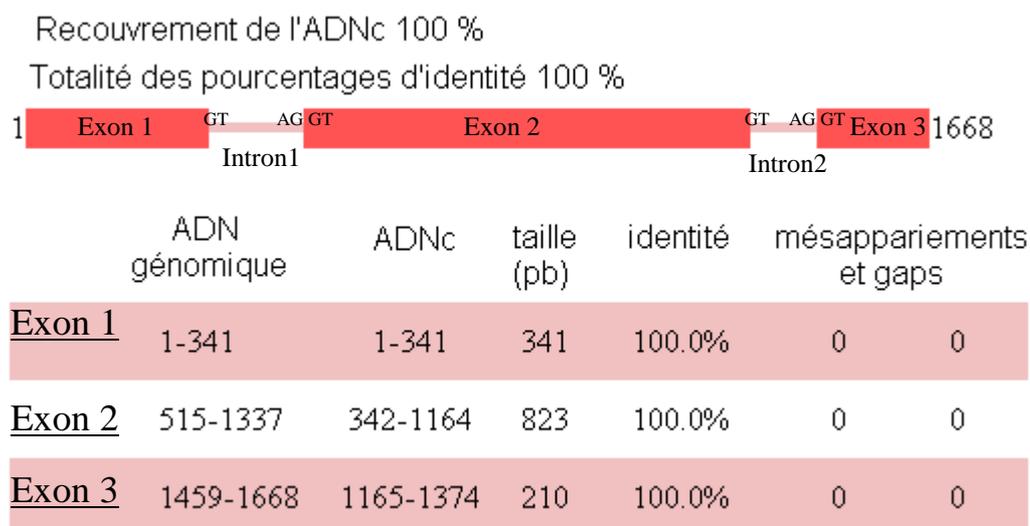


Figure 50 : Structure intron / exon du gène codant la lipase hypothétique d'*Isochrysis galbana* (*IgLip*).

La séquence génomique d'*IgLip* est présentée. Les blocs rouges et les segments roses représentent respectivement les exons et les introns. La comparaison de la séquence génomique avec celle de l'ADNc a été réalisée avec le programme Spidey.

4. Caractérisation préliminaire et analyses comparatives du gène *IgEst1* codant une estérase hypothétique chez *Isochrysis galbana*

4.1. Analyse et traduction de la séquence nucléotidique *IgEst1*

L'ADNc pleine longueur, obtenu par amplification rapide des extrémités 5' et 3', est constitué de 1124 nucléotides (Figure 51), jusqu'à la queue polyadénylée en 3' qui fait partie des éléments distinctifs des ARNm. L'analyse informatique de la séquence nucléotidique, dans les six cadres de lecture possibles, a révélé la présence d'un ORF encadré par des signaux de traduction et compris entre deux régions 5' et 3' non traduites. Le codon start ATG et le codon stop TGA susceptibles d'initier et de terminer la traduction figurent respectivement en positions 19-21 et 844-846. La séquence codante atteint une taille de 825 pb.

```

1           M A R G L A A C T R L A A A A
1  CAGAGTATGCGGGGATACCCTGATCGAGCCGATCGCGGAGGGCTTGCCGCGTGCACACGACTTGCAGCGGCGGC
16  A A V S A V A S S P L Q S R V S A V S M S T A A G
63  AGCGGCGGTCTCTGCTGTGCTTCCAGCCCGTGAATCTCGTGTCTCCGCGTCAGCATGTCGACCGCTGCCGG
41  A A P P R I T Q D R A R N N L I T I A P G E G M P
138 TGCGGCGCCCGCGCATCACGACGAGACCGCGCGCAACAACCTGATCACCATCGCGCCCGGGGAAGGCATGCC
66  H T A T V I G P I H G L G D S N M G W A D V A M Q
213 GCACACCGCGACGGTTCATCGGCCCCATCCACGGGCTGGGCGACGCAACATCGGGCTGGGCGGACGTGGCGATGCA
91  L Q S V M P Y C K F I L P N A P V R P V T L N G G
288 GCTGCAGAGTGTTCATGCCGTATTGCAAGTTCATCTCCCAATGCGCCCGTCCGTCCGTGTCACCCCTCACGGCGG
116 M S M P S W Y D I T S L D K R E S Q P C T G I E E
363 AATGTCGATGCCAGTTGGTATGACATCACGAGTTTGGATAAGCGCGAGTCCCAGCCGTGCACGGGCATCGAAGA
141 S R Q A M L D L I S A E V A S G I P P S R I A I A
438 GTCTCGCCAGGCGATGCTCGACCTCATCTCCGTTGAGGTTGCGTCCGGCATCCCCCCTCCCGATCGCCATCGC
166 G F S Q G G A V A L F T G L Q Y S H T L A G V L C
513 CGGCTTCTCCAGGGAGGCGCCGTGCCCCCTTCACAGGGCTGCAGTATTCACACACCCTGGCCGGCGTGCTATG
191 L S G Y L A A E E R F I L A P E A V N T P V A H F
588 CTTGTCCGGTACTTGGCGGCGGAGGAGCGTTTCATACTGGCGCGGAGGCTGTGAACTACTCCCGTCCGCTCACTT
216 H G S D D Q T V Q I K W A R G S Q A H L R E L G I
663 TCACGGCTCGGATGATCAAACGGTCCAAATCAAATGGGCCCGCGGCTCTCAGGCCACCTGCGCGAGCTGGGCAT
241 R T Y E L K E Y S P L G H S A S Q Q E I A D V L A
738 CCGCACCTACGAGCTAAAAGAGTACTCGCCCCCTCGGCCACAGCGCTCGCAGCAGGAGATCGCGGACGTGCTTGC
266 W L Q A R L P E D A *
813 ATGGCTTCAGCCAGACTGCCAGAGGACGCTGAGGATAGCCTGGAGCATATTTAGTATCTAGTGGGGAGCTGCC
888 CCTGACAAGCCACGCGTGGGGAAGGCTCGAATCCCAAGTCTGGATCCACATATATGCCACACTTGTATGGAGGC
963 CTCATCCACATCTGAGATCAACAAATGTGGGTGTGGCGGCGTCCGCATTTTTGTATACGCTTAGGCGCTTACACT
1038 ACGGATATTCGGATTATCTCTGGATATCTTCTTCTATCTTCTATGATCCCATGGGATTATCGATTATAATAT
1113 CATTATCGATTCAAAAAAAAAAAAAAAAAAAAAAAAAA

```

Figure 51 : Séquence nucléotidique de l'ADNc *IgEst1* et traduction de la région codante.

Les amorces dessinées à partir de l'étiquette de séquence transcrite n°ISL00008005 et utilisées en RACE et Nested PCR sont écrites et soulignées en vert. Les codons d'initiation et de terminaison de traduction sont indiqués en noir sur fond vert. Les résidus cystéine sont soulignés en noir. Les régions non traduites 5' et 3' sont écrites en gras.

La séquence protéique obtenue par déduction ne contient pas de sites potentiels de glycosylation de type NXS et NXT. Quatre résidus cystéine ont été repérés dans la séquence. Le marqueur de séquence ISL00008005 indexé dans la banque EST d'*I. galbana* présentait une erreur de séquençage : l'addition d'un résidu glycine entre les résidus glycine 552 et cytosine 553 entraînait un décalage du cadre de lecture, une mauvaise traduction et un alignement incorrect.

4.2. Prédiction des caractéristiques physicochimiques

4.2.1. Signal peptidique

La présence d'un peptide signal est fortement soupçonnée en début de séquence codante (Figure 52). Selon le serveur SignalP, il s'agirait d'un signal peptidique classique figurant parmi les signaux les plus communément rencontrés chez les organismes vivants. Riche en résidus hydrophobes, il se termine par un site de clivage spécifiquement reconnu par la signal peptidase I (SPase). Le peptide signal est un peptide d'adressage situé à l'extrémité N-terminale des protéines qui oriente le déplacement des protéines en direction d'un compartiment cellulaire (reticulum endoplasmique, appareil de Golgi, lysosomes, mitochondries, chloroplastes) ou dans le milieu extracellulaire. Le peptide signal assure, par sa richesse en acides aminés hydrophobes, l'incorporation des protéines dans les membranes.

Le modèle de prédiction a été conçu à partir de données eucaryotes. Il suppose la présence d'un signal de sécrétion long de 23 acides aminés avec une très forte probabilité de l'ordre de 0,999 et avec un site de clivage positionné entre les résidus S23 et S24. Le signal présente une structure classique composée d'une région de résidus chargés positivement (région-n en vert), suivie par une région de résidus hydrophobes (région-h en bleu foncé) et se terminant par une région de résidus neutres et polaires (région-c en bleu clair). Les résidus situés en position -3 (S23) et -1 (V21) par rapport au site de clivage sont petits et neutres, condition nécessaire pour que la coupure se produise correctement.

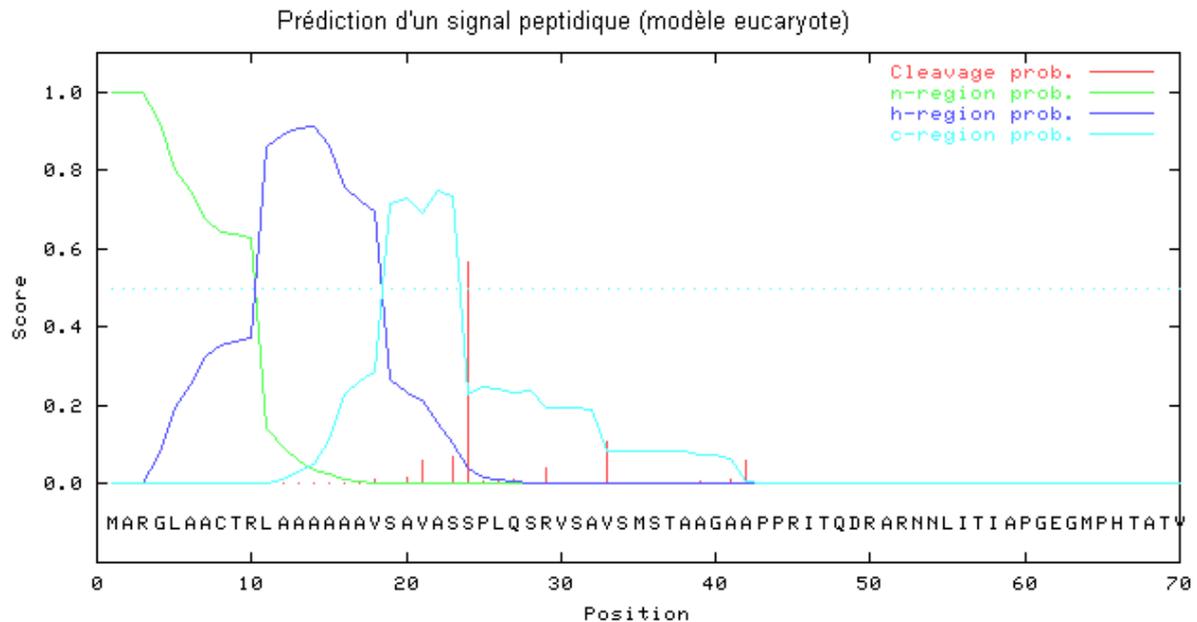


Figure 52 : Prédiction de la présence d'un signal peptidique dans la région N-terminale d'*IgEst1*, avec le serveur SignalP.

L'axe des abscisses indique la position des acides aminés. L'axe des ordonnées indique l'indice de probabilité. Les sites de clivage sont en rouge. Les résidus chargés sont en vert, les résidus hydrophobes en bleu foncé et les résidus hydrophyles en bleu clair.

4.2.2. Masse moléculaire et pH du point isoélectrique

Le cadre ouvert de lecture d'*IgEst1* code un polypeptide de 275 acides aminés. En tenant compte du signal prédit par le modèle eucaryote, le polypeptide mature serait de 252 résidus après clivage du peptide signal. Cela correspond à une petite protéine de poids moléculaire approximativement égal à 26,92 kDa et de pI théorique de 5,87.

4.2.3. Profils d'hydrophobicité et d'hydrophilicité

Le profil d'hydrophobicité de la protéine *IgEst1* est représenté sur la Figure 53. On observe trois zones avec un indice d'hydrophobicité > 1 . La première zone hydrophobe concorde avec le signal peptidique. L'absence de zones à fort indice d'hydrophobicité sur le profil de la molécule suggère que la protéine *IgEst1* n'est pas une molécule transmembranaire.



Figure 53 : Profils hydrophobe et hydrophile de la protéine *IgEst1*.

Axe des ordonnées (bleu) : échelle de Kyte et Doolittle. Axe des ordonnées (vert) : échelle de Hopp et Woods. Axe des abscisses : position des acides aminés. La taille de la fenêtre est de 21 acides aminés.

4.3. Prédiction de la fonction enzymatique

4.3.1. Recherche de séquences homologues

La recherche de séquences homologues dans les bases de données publiques a permis d'identifier des protéines homologues chez un grand nombre d'organismes. Les meilleurs scores d'alignement ont été obtenus avec des séquences de champignons Ascomycètes, identifiées elles-mêmes par comparaison de séquences et classées dans la famille des phospholipases et carboxylestérases (PLip/CEst). La Figure 54 représente l'alignement, réalisé avec le programme MUSCLE, de la séquence protéique d'*IgEst1* avec celles des champignons *Aspergillus niger* (*AnPLip/CEst*) et *Aspergillus oryzae* (*AoPLip/CEst*). L'alignement est déficient pour les 60 premiers acides aminés dans la région N-terminale d'*I. galbana*. Le reste de la séquence s'aligne correctement avec les homologues fongiques. On retrouve, de la 166^{ème} à la 170^{ème} position, le pentapeptide consensus G-X-S-X-G strictement conservé sous la forme G-F-S¹⁶⁸-Q-G chez tous les homologues identifiés. Les deux résidus Asp et His de la potentielle triade catalytique seraient situés respectivement en position 220 et 253 chez *I. galbana*.

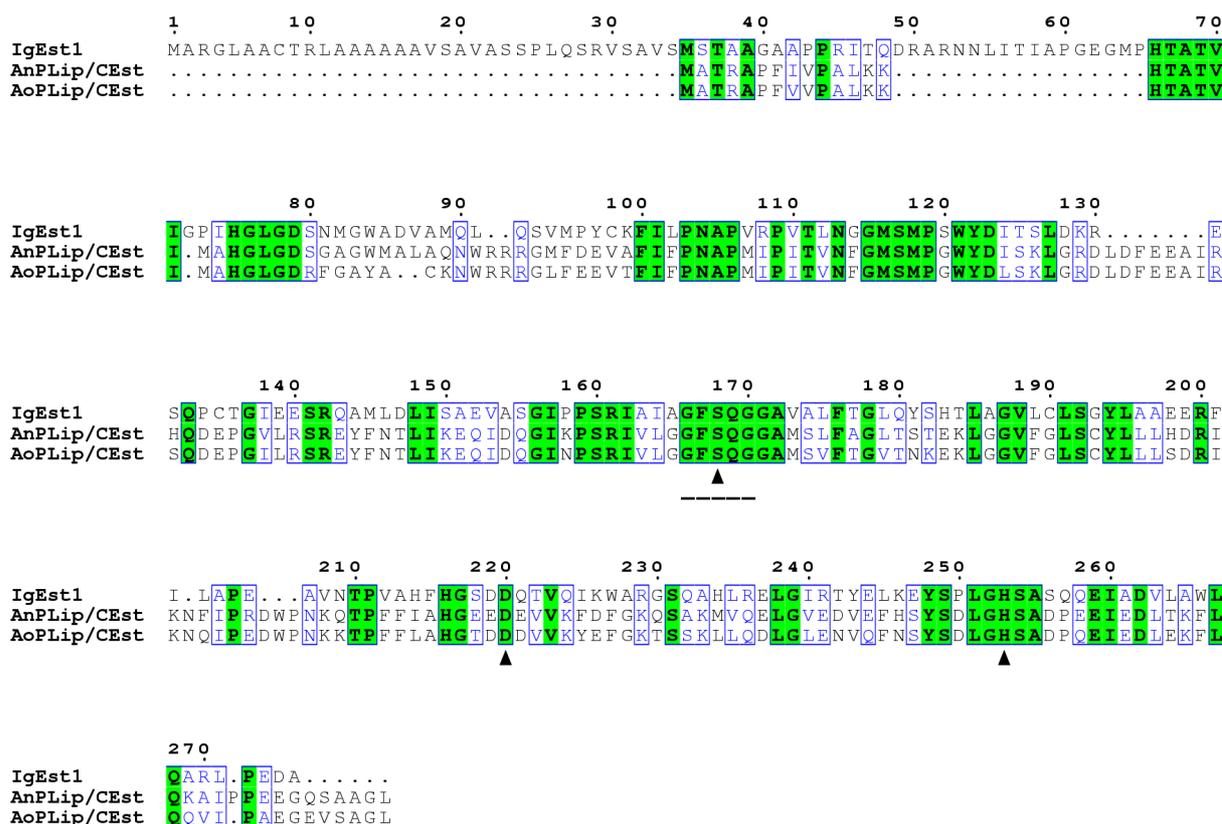


Figure 54 : Aligment de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) avec ses homologues chez les champignons *Aspergillus niger* (*AnPLip/CEst*) et *Aspergillus oryzae* (*AoPLip/CEst*).

Organisme / numéro d'accèsion : *Aspergillus niger* / XP 001395827 ; *Aspergillus oryzae* / BAE62787. Les résidus strictement conservés sont indiqués en noir sur fond vert, les résidus bien conservés écrits et encadrés en bleu. Les triangles noirs indiquent l'emplacement des trois résidus Ser/Asp/His supposés constituer la triade catalytique. Le pentapeptide consensus G-X-S-X-G est souligné par un trait noir discontinu.

La séquence isolée chez *I. galbana* partage respectivement 31 et 32 % de résidus identiques avec ses homologues fongiques *AoPLip/CEst* et *AnPLip/CEst* (Tableau 18).

	<i>AnPLip/CEst</i>	<i>AoPLip/CEst</i>
<i>IgEst1</i>	32 %	31 %
	44 %	44 %
<i>AnPLip/CEst</i>		79 %
		90 %

Tableau 18 : Pourcentages d'identité et de similitude entre l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) et ses homologues fongiques chez *Aspergillus niger* (*AnPLip/CEst*) et *Aspergillus oryzae* (*AoPLip/CEst*).

Le pourcentage indiqué en noir sur fond vert correspond au pourcentage d'identité, c'est-à-dire au pourcentage de résidus strictement conservés entre deux séquences. Les chiffres écrits en bleu indiquent la proportion de résidus similaires, c'est-à-dire la proportion de résidus biochimiquement proches, entre deux séquences.

D'autres homologues, présentant de bons scores d'alignement avec *IgEst1* ont été identifiés chez des organismes de diverses origines (Figure 55). *IgEst1* s'aligne correctement (excepté dans la région N-terminale) avec les lysophospholipases I et II humaines (*HsLPLipI* et *II*) et avec la lysophospholipase I de la souris *Mus musculus* (*MmLPLipI*). *IgEst1* partage aussi des similarités de séquence avec une protéine identifiée chez la drosophile, *Drosophila melanogaster* (*DmPLip/CEst*) et deux protéines de plantes supérieures *Arabidopsis thaliana* (*AtPLip/CEst*) et *Oryza sativa* (*OsPLip/CEst*). Ces trois protéines ont été classées, par homologie de séquence, dans la famille des phospholipases et carboxylestérases.

IgEst1 partage 28 % de résidus identiques avec ses homologues humaine et murine (Tableau 19).

	<i>HsLPLipI</i>	<i>HsLPLipII</i>	<i>MmLPLipI</i>	<i>DmPLip/CEst</i>	<i>AtPLip/CEst</i>	<i>OsPLip/CEst</i>
<i>IgEst1</i>	28 % 44 %	26 % 42 %	28 % 43 %	26 % 44 %	23 % 38 %	26 % 42 %
<i>HsLPLipI</i>		66 % 81 %	91 % 96 %	50 % 61 %	28 % 45 %	28 % 43 %
<i>HsLPLipII</i>			63 % 81 %	49 % 63 %	28 % 46 %	28 % 45 %
<i>MmLPLipI</i>				49 % 61 %	27 % 45 %	28 % 44 %
<i>DmPLip/CEst</i>					31 % 47 %	28 % 43 %
<i>AtPLip/CEst</i>						57 % 69 %

Tableau 19 : Pourcentages d'identité et de similitude entre l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) et ses homologues humaines (*HsLPLipI* et *II*), murine (*MmLPLipI*) et végétales (*AtPLip/CEst*, *OsPLip/CEst*).

Le pourcentage indiqué en noir sur fond vert correspond au pourcentage d'identité, c'est-à-dire au pourcentage de résidus strictement conservés entre deux séquences. Les chiffres écrits en bleu indiquent la proportion de résidus similaires, c'est-à-dire la proportion de résidus biochimiquement proches entre deux séquences.

4.3.2. Recherche de motifs fonctionnels

La séquence *IgEst1* a été comparée avec les motifs classés et stockés, sous forme de profils, dans la base de données Blocks. La séquence protéique d'*IgEst1* comporte 2 motifs caractéristiques des enzymes de la famille des phospholipases et carboxylestérases IPB003140. Les deux motifs (ou blocs) sont indexés IPB003140A et IPB003140B. Ils sont présentés en format WebLogo en Annexe 10. Les résidus les mieux conservés du bloc A, H6, G7 et G9 correspondent aux résidus H75, G76 et G78 chez *IgEst1* (Figure 55). Le bloc B inclut le pentapeptide consensus G-X-S-X-G majoritairement présent sous la forme G-F-S-Q-G, que l'on retrouve de la 166^{ème} à la 170^{ème} position sur la Figure 55.

4.4. Prédiction de l'organisation structurale

4.4.1. Alignement avec la thioestérase humaine

Afin de prédire au mieux son organisation structurale et les résidus impliqués dans son mécanisme catalytique, l'estérase d'*Isochrysis galbana IgEst1* a été comparée avec la plus proche estérase pleinement caractérisée c'est-à-dire dont la fonction et la structure tridimensionnelle sont connues (Figure 56). Cette estérase est la thioestérase humaine référencée sous le code PDB 1fj2 (Devedjiev *et al.*, 2000). Les séquences d'*IgEst1* et de la thioestérase humaine (*HsATEst*) partagent 29 % de résidus identiques et 45 % de résidus similaires. L'estérase 1 d'*I. galbana* contient la triade catalytique potentielle Ser168, Asp220, His253 parfaitement superposable à celle de l'estérase humaine Ser119, Asp174 et His208. Les deux résidus L30 et Q120 ont été identifiés chez l'estérase humaine comme étant des résidus impliqués dans le trou oxyanionique. Ils sont parfaitement conservés chez *IgEst1*. Parmi les quatre acides aminés connus pour leurs interactions avec le substrat chez *HsATEst*, seuls L63 et L78 sont présents chez *IgEst1*. Les résidus W145 et F181 sont respectivement remplacés par les résidus Y194 et W227 chez *IgEst1*. Ces 3 acides aminés sont des acides aminés aromatiques.

Cet alignement nous montre que l'organisation structurale nécessaire à la mise en place de l'activité catalytique est conservée chez l'estérase 1 d'*I. galbana*.

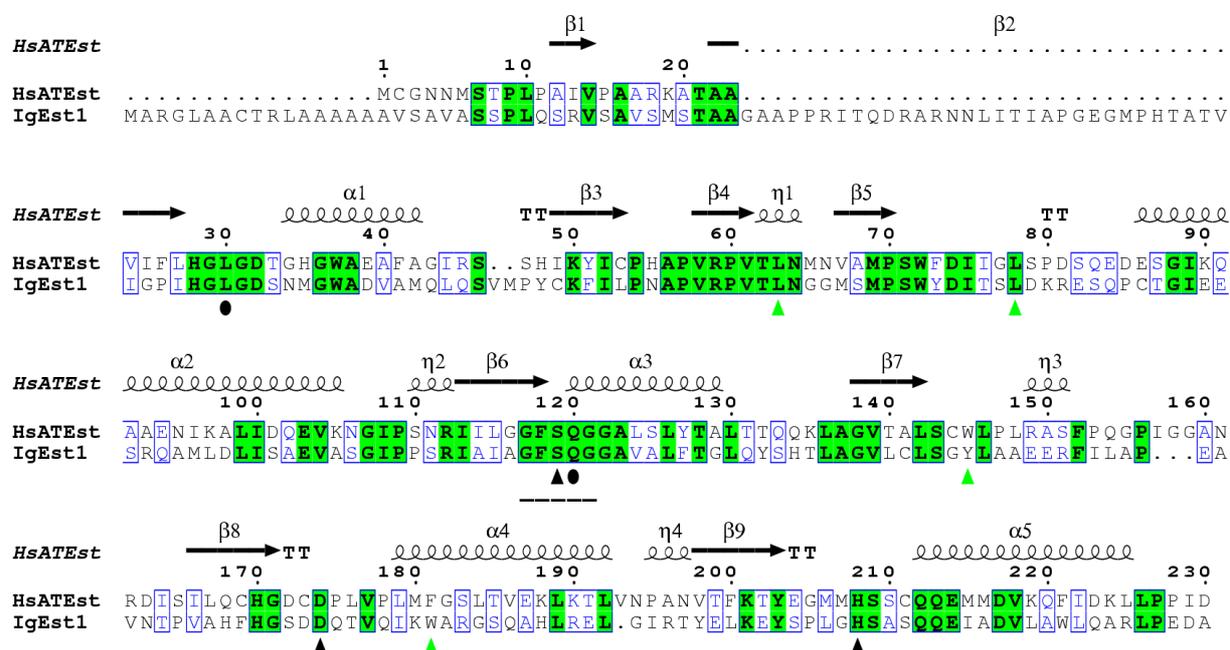


Figure 56 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) avec la thioestérase humaine (*HsATEst*) (Swiss-Prot 075608 ; code PDB 1fj2).

Les résidus strictement conservés sont indiqués en noir sur fond vert, les résidus bien conservés écrits et encadrés en bleu. La triade catalytique Ser/Asp/His est indiquée par des triangles noirs. Le pentapeptide consensus G-X-S-X-G est souligné en noir. Les triangles verts et les ronds noirs correspondent aux résidus impliqués dans la fixation du substrat et de ses intermédiaires. La structure secondaire de l'estérase humaine figure au dessus de l'alignement.

4.4.2. Modèle tridimensionnel

Un modèle tridimensionnel a été construit par comparaison avec la thioestérase humaine. Dans l'alignement de la séquence entre *IgEst1* et la thioestérase humaine, nous avons observé qu'il n'existe que peu de gaps, en dehors de la région N-terminale. Cela a permis d'obtenir un modèle tridimensionnel pour la protéine *IgEst1* à partir des données structurales de la thioestérase humaine (Figure 57). Ce modèle demeure hypothétique d'autant plus que 25 % d'identité est admis comme étant la limite inférieure de construction d'un modèle. Cependant l'analyse des séquences montre que la modélisation est fiable sur toute la longueur de la séquence d'*IgEst1* sauf pour les 70 premiers acides aminés. La protéine *IgEst1* posséderait, comme la thioestérase humaine, une structure globulaire avec repliement α/β caractéristique de la superfamille des α/β hydrolases, constitué de feuillets β entourés de part et d'autres par des hélices α .

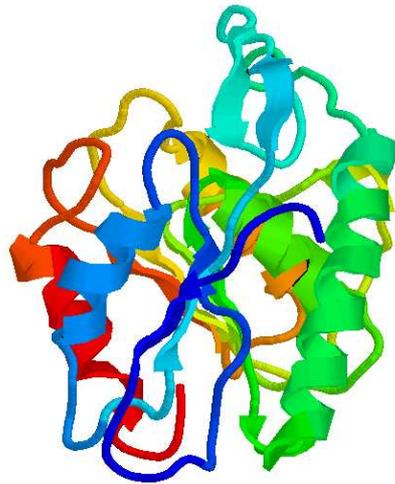


Figure 57 : Modèle structural de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*).

Le modèle 3-D d'*IgEst1* a été établi grâce au programme Geno3D, par comparaison avec les données structurales de la thioestérase humaine (code PDB 1fj2). La molécule a été éditée avec le programme Protein Explorer.

4.5. Séquences homologues à *IgEst1* chez les microalgues

La séquence d'*IgEst1* a été utilisée pour rechercher des séquences homologues dans les banques de données publiques de microalgues. Cette analyse a permis d'identifier 7 séquences homologues de microalgues (Figure 58).

Deux séquences homologues ont été identifiées chez la microalgue verte *Ostreococcus lucimarinus*. Les deux gènes identifiés ne comportent pas d'introns. Le premier transcrit compte 960 nucléotides et une séquence codante (CDS) de 681 pb codant une protéine de 226 acides aminés et de 24,08 kDa (n° d'identification de la protéine, id. : 32853). Le second transcrit, de 855 pb, code une protéine de 284 acides aminés et de 31,02 kDa (id. : 93362). Chez l'espèce voisine *Ostreococcus tauri*, l'homologue d'*IgEst1* est une protéine référencée 26019 constituée de 227 acides aminés. Il s'agit encore d'une petite protéine dont le poids moléculaire théorique est de 24,35 kDa. Elle est codée par un gène sans intron dont la taille en pb du transcrit et celle de la CDS atteignent respectivement 919 et 684. Ces trois protéines appartenant au genre *Ostreococcus* ont été annotées comme étant des enzymes similaires à des phospholipases et des carboxylestérases. Elles seront dans notre étude nommées *OlPLip/CEst1*, *OlPLip/CEst2* et *OtPLip/CEst*.

Dans le génome de l'algue rouge unicellulaire *Cyanidioschyzon merolae*, un gène homologue sans intron, de séquence protéique référencée CMF176C, a été identifié. Annotée par homologie de séquence, la protéine serait similaire à une lysophospholipase de type II (*CmLPLipII*). La traduction de la séquence codante nucléotidique de cet homologue correspond à une protéine de 281 acides aminés, de poids moléculaire potentiellement égal à 31,34 kDa.

Les dernières séquences homologues identifiées sont issues du génome des diatomées *Phaeodactylum tricornutum* (*PtPLip/CEst*) et *Thalassiosira pseudonana* (*TpPLip/CEst1*, *TpPLip/CEst2*). Chez *P. tricornutum*, le gène homologue identifié comporte un intron de 84 pb et deux exons de 291 et 766 pb. La séquence codante de 957 pb se traduit en une protéine de 318 résidus de poids moléculaire égal à 34,95 kDa. La protéine correspondante est indexée sous le numéro 21816.

Chez la microalgue *T. pseudonana*, la séquence protéique 6090 est similaire à celle d'*IgEst1*. Le gène comporte deux introns de 74 et 124 pb et trois exons de 343, 35 et 383 pb dont la somme égale à 761 pb indique la taille de la CDS. Le gène code une protéine de 253 résidus et de poids moléculaire théorique égal à 27,77 kDa. La seconde protéine similaire à *IgEst1* est indexée 1809. Elle est incomplète du côté N-terminal. Elle est codée par un gène avec deux introns de 137 et 84 pb.

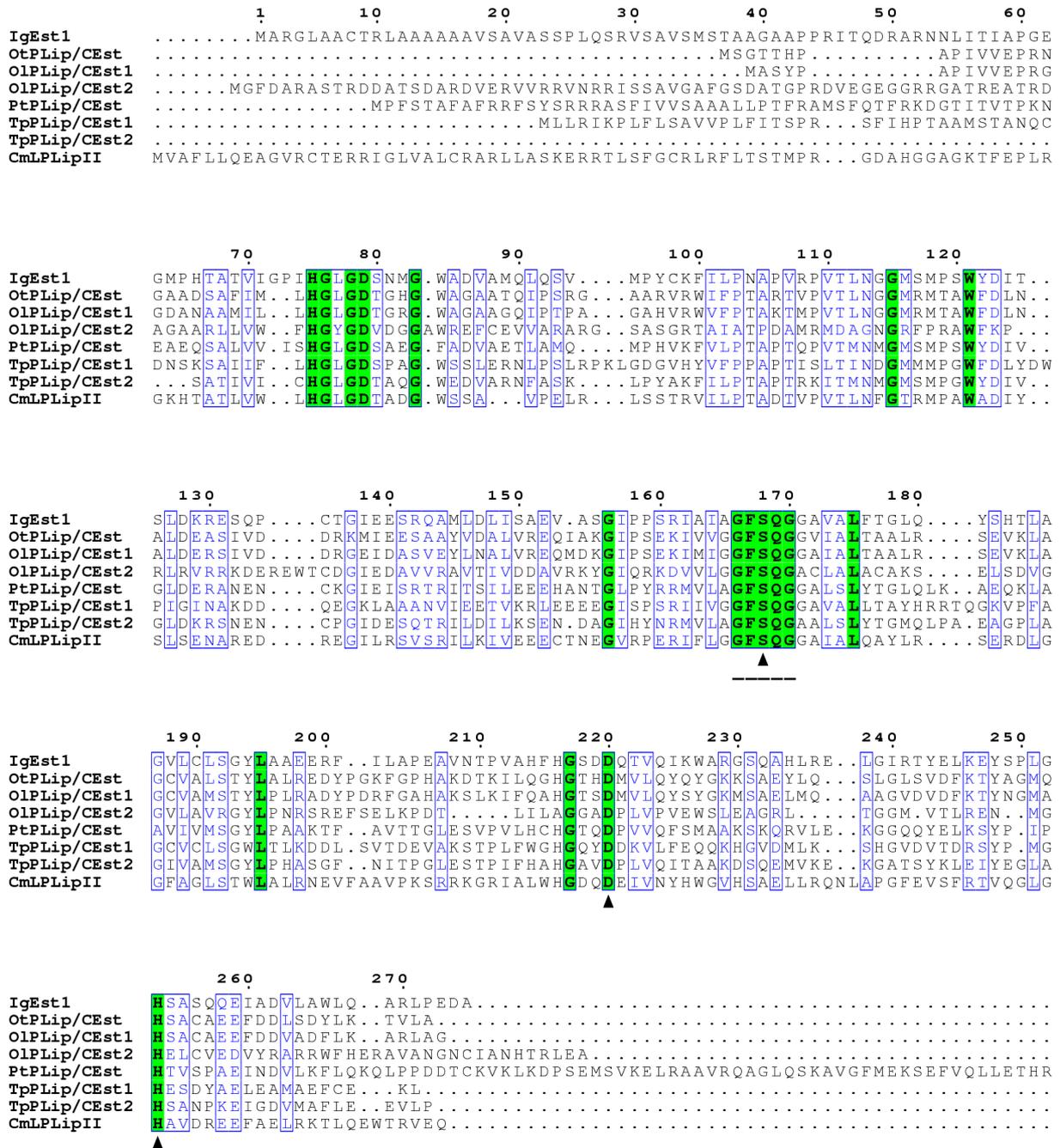


Figure 58 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (IgEst1) avec des homologues de phospholipase et de carboxylestérase chez les microalgues *Ostreococcus lucimarinus* (OiPLip/CEst1, OiPLip/CEst2), *Ostreococcus tauri* (OtPLip/CEst), *Phaeodactylum tricornutum* (PtPLip/CEst), *Thalassiosira pseudonana* (TpPLip/CEst1, TpPLip/CEst2) et *Cyanidioschyzon merolae* (CmLPLipII).

Organisme / numéro d'identifiant : *O. lucimarinus* / 32853 (OiPLip/CEst1) / 93362 (OiPLip/CEst2) ; *O. tauri* / 26019 ; *P. tricornutum* / 21816 ; *T. pseudonana* / 6090 (TpPLip/CEst1) / 18091 (TpPLip/CEst2) ; *C. merolae* / CMF176C. Les résidus strictement conservés sont indiqués en noir sur fond vert, les résidus bien conservés écrits et encadrés en bleu. Les triangles noirs indiquent l'emplacement des trois résidus Ser/Asp/His supposés constituer la triade catalytique. La région conservée G-X-S-X-G autour de la sérine nucléophile est soulignée par un trait noir discontinu.

Le meilleur alignement a été obtenu avec les séquences homologues identifiées chez les diatomées. *IgEst1* est identique aux séquences *PtPLip/CEst* et *TpPLip/CEst2*, avec 32 et 35 % d'identité (Tableau 20). La séquence *IgEst1* présente également 26 et 27 % d'identité avec les séquences identifiées chez les microalgues vertes *O. tauri* et *O. lucimarinus*. Enfin, 25 % de résidus identiques ont été comptabilisés avec l'algue rouge *C. merolae*.

	<i>OtPLip</i> / <i>CEst</i>	<i>OlPLip</i> / <i>CEst1</i>	<i>OlPLip</i> / <i>CEst2</i>	<i>PtPLip</i> / <i>CEst</i>	<i>TpPLip</i> / <i>CEst1</i>	<i>TpPLip</i> / <i>CEst2</i>	<i>CmLPLipII</i>
<i>IgEst1</i>	27 % 43 %	26 % 42 %	18 % 31 %	32 % 45 %	21 % 37 %	35 % 49 %	25 % 38 %
<i>OtPLip/CEst</i>		72 % 82 %	16 % 27 %	21 % 36 %	28 % 46 %	29 % 45 %	30 % 44 %
<i>OlPLip/CEst1</i>			17 % 28 %	22 % 38 %	29 % 48 %	31 % 50 %	30 % 44 %
<i>OlPLip/CEst2</i>				15 % 28 %	15 % 26 %	16 % 27 %	20 % 34 %
<i>PtPLip/CEst</i>					18 % 33 %	35 % 47 %	22 % 34 %
<i>TpPLip/CEst1</i>						24 % 39 %	25 % 42 %
<i>TpPLip/CEst2</i>							21 % 34 %

Tableau 20 : Pourcentages d'identité et de similitude entre l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) et ses homologues microalgales chez *Ostreococcus lucimarinus* (*OlPLip/CEst1*, *OlPLip/CEst2*), *Ostreococcus tauri* (*OtPLip/CEst*), *Phaeodactylum tricornutum* (*PtPLip/CEst*), *Thalassiosira pseudonana* (*TpPLip/CEst1*, *TpPLip/CEst2*) et *Cyanidioschyzon merolae* (*CmLPLipII*).

Le pourcentage indiqué en noir sur fond vert correspond au pourcentage d'identité, c'est-à-dire au pourcentage de résidus strictement conservés entre deux séquences. Les chiffres écrits en bleu indiquent la proportion de résidus similaires, c'est-à-dire la proportion de résidus biochimiquement proches entre deux séquences.

L'alignement multiple entre la séquence *IgEst1* et ses deux plus proches homologues, les séquences des diatomées *P. tricornutum* et *T. pseudonana*, est représenté en Figure 59. Cette comparaison permet de mieux visualiser les zones identiques et similaires. Excepté dans la région N-terminale, *IgEst1* s'apparie de façon homogène avec ses deux homologues. On distingue 5 blocs constitués d'au moins 4 résidus identiques consécutifs. Parmi ces blocs, on retrouve les 2 blocs signature de la famille IPB003140 qui regroupe phospholipases et carboxylestérases (§2.3.2).

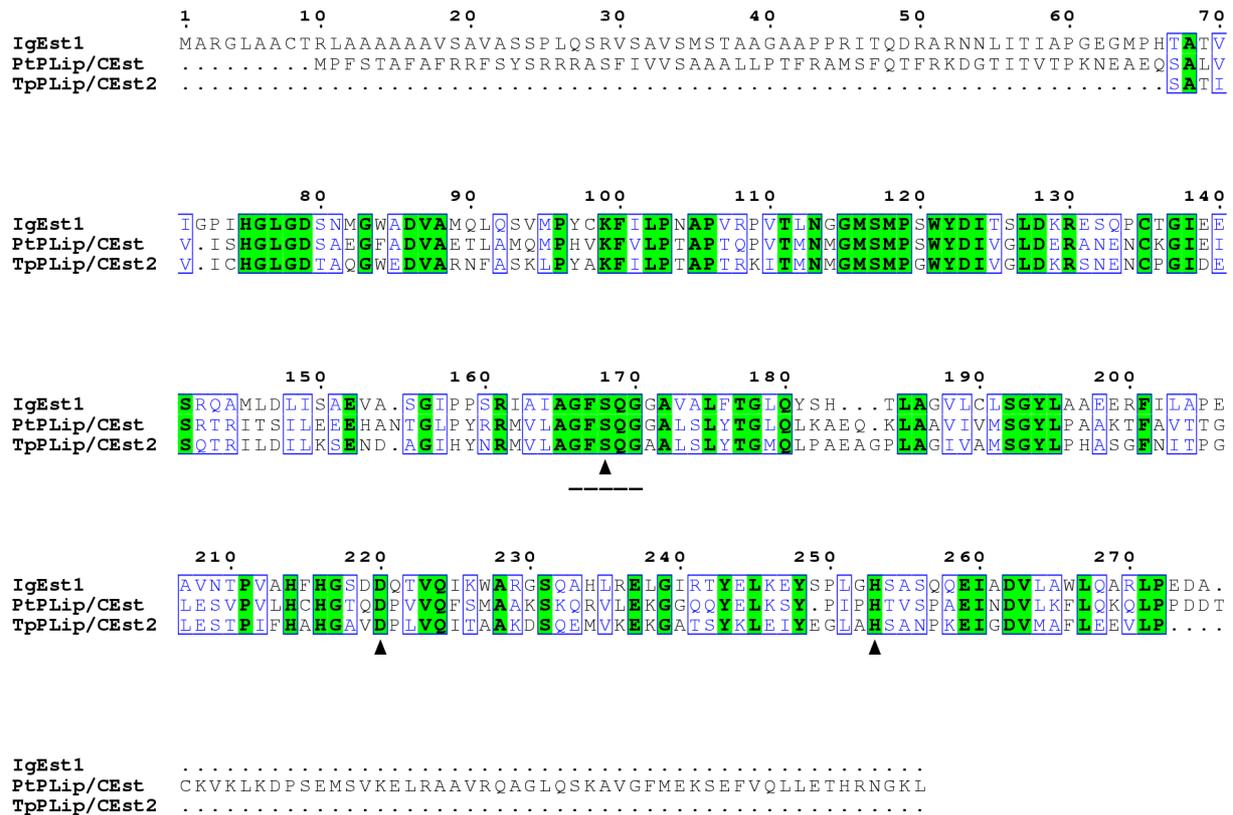


Figure 59 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) avec deux de ses homologues chez les diatomées *Phaeodactylum tricornutum* (*PtPLip/CEst*) et *Thalassiosira pseudonana* (*TpPLip/CEst2*).

Les résidus strictement conservés sont indiqués en noir sur fond vert, les résidus bien conservés écrits et encadrés en bleu. Les triangles noirs indiquent l'emplacement des trois résidus Ser/Asp/His supposés constituer la triade catalytique. Le pentapeptide consensus G-X-S-X-G est souligné par un trait noir discontinu.

Une séquence avec un fort taux d'identité (50 %) et de similitude (63 %) a été identifiée dans GenBank (Figure 60). Cette séquence provient d'échantillons de microorganismes prélevés dans le milieu marin (Rusch *et al.*, 2007). Ces échantillons environnementaux ont été récoltés au cours de l'expédition « Sorcerer II Global Ocean Sampling », dans le cadre d'un projet d'exploration des génomes de microorganismes planctoniques de surface en milieu océanique (Yooseph *et al.*, 2007).

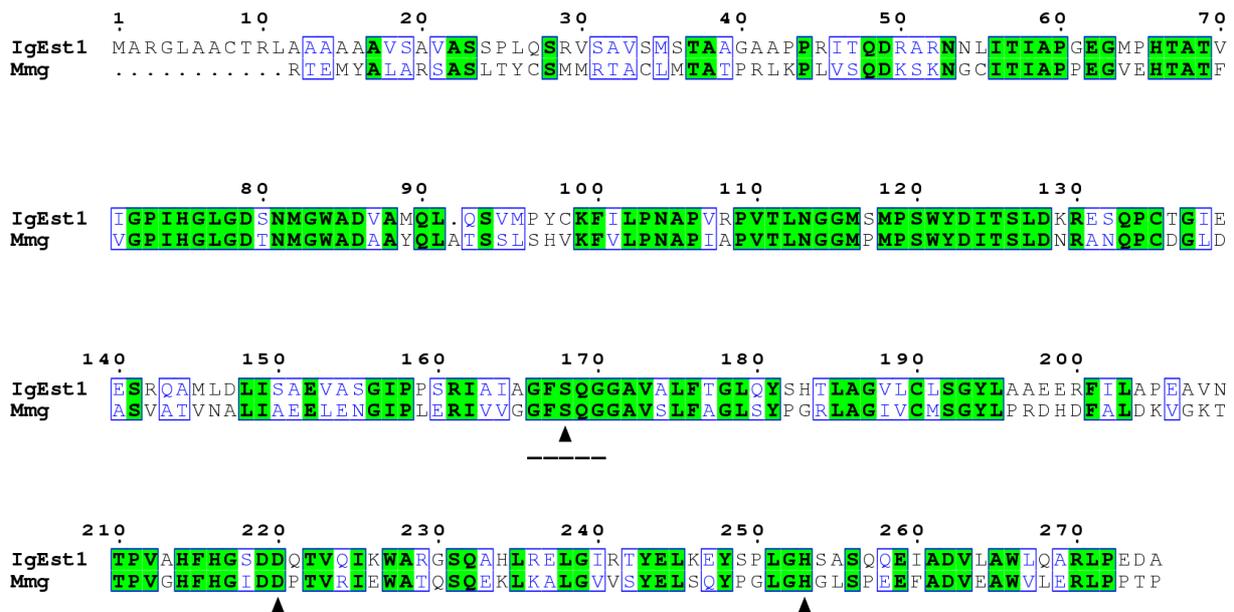


Figure 60 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) avec un homologue de microorganisme planctonique marin.

L'homologue, de numéro d'accèsion ECS53966, est dénommé Mmg pour « marine metagenomic ». Les résidus strictement conservés sont indiqués en noir sur fond vert, les résidus bien conservés écrits et encadrés en bleu. Les triangles noirs indiquent l'emplacement des trois résidus Ser/Asp/His supposés constituer la triade catalytique. Le pentapeptide consensus G-X-S-X-G est souligné par un trait noir discontinu.

4.6. Classification et analyse phylogénétique

Pour compléter les analyses comparatives fonctionnelles et structurales et pour positionner les séquences nouvellement identifiées chez *I. galbana*, *Phaeodactylum tricornutum* et *Thalassiosira pseudonana*, une analyse phylogénétique des lysophospholipases et des enzymes homologues classées dans la famille des phospholipases et carboxylestérases a été entreprise. Les trois séquences microalgales ont été comparées avec 61 séquences homologues provenant de bactéries, de champignons, de plantes, de nématodes, de plathelminthes, d'insectes et de vertébrés. Ces séquences sont disponibles dans les banques de données nucléotidiques et protéiques. Les séquences pleine-longueur ont été alignées avec le programme MUSCLE. Le logiciel Gblocks, version 0.91b, a été utilisé pour extraire de l'alignement les régions conservées (Annexe 11). Cette opération, préalable à la construction de l'arbre, permet d'éliminer les erreurs d'alignement qui pourraient fausser l'analyse phylogénétique.

L'analyse phylogénétique a été réalisée avec la méthode du Neighbor Joining et 500 valeurs de bootstraps sur les régions extraites par Gblocks. L'arbre a été enraciné avec un groupe de 5

protéines Yolk de différentes espèces de drosophiles. L'analyse phylogénétique a mené à l'identification de 90 sites informatifs pour la construction des arbres.

L'arbre consensus obtenu (Figure 61) montre une topologie caractérisée par la présence de plusieurs groupes séparés par des nœuds robustes. Un premier groupe inclut les carboxylestérases identifiées chez les plantes. Un second groupe rassemble les carboxylestérases et les phospholipases de bactéries (*Pseudomonas aeruginosa*, *Pseudomonas fluorescens* et *Pseudomonas putida*). Les lysophospholipases de type I et II des vertébrés sont respectivement réparties en 2 sous-groupes distincts. La séquence isolée chez *I. galbana* constitue, avec les séquences identifiées chez *Phaeodactylum tricornutum* (*PtPLip/CEst*) et *Thalassiosira pseudonana* (*TpPLip/CEst2*), un nouveau groupe. Cela s'explique par les proportions relativement importantes de résidus identiques et similaires observés précédemment entre la séquence de l'haptophyte et celles des diatomées (§4.5.). Ce nouveau groupe apparaît voisin de celui des champignons ascomycètes filamenteux (*Aspergillus fumigatus*, *Aspergillus niger*, *Aspergillus oryzae* et *Emericella nidulans*). Au sein de ce nouveau groupe, les séquences des 2 diatomées forment un sous-groupe soutenu par une forte valeur de bootstrap (100 %).

Pour valider cette analyse, la même étude a été réalisée avec une autre méthode de phylogénie : le Maximum de Parcimonie (Figure 62). L'arbre généré montre une topologie comparable avec le premier. L'arbre consensus obtenu confirme la distribution des enzymes dans leurs classes respectives. Il confirme également l'assignation des séquences *IgEst1*, *PtPLip/CEst* et *TpPLip/CEst2* dans un nouveau groupe.

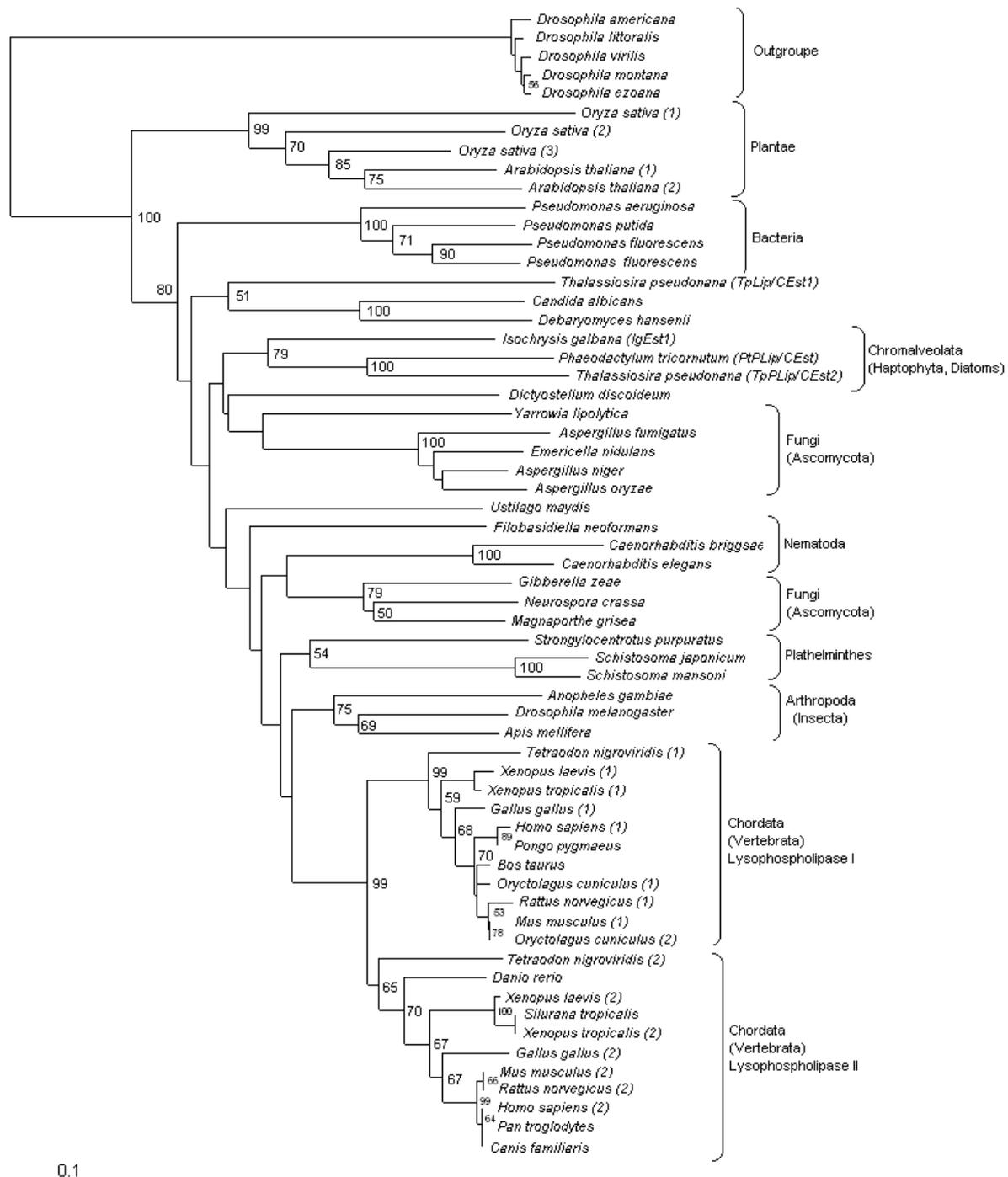


Figure 61 : Phylogénie de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst1*) et de ses homologues, obtenue avec la méthode du Neighbor Joining sur les régions extraites par Gblocks. L'analyse phylogénétique a été réalisée sur des membres des familles phospholipase, carboxylestérase et lysophospholipase, isolés ou identifiés chez différents organismes. Les groupes d'organismes et les différentes familles d'enzymes sont indiqués par des accolades. L'arbre consensus a été obtenu avec la méthode du Neighbor Joining et 500 valeurs de bootstraps sur les régions conservées. Seules les valeurs de bootstraps supérieures à 50 % sont indiquées. L'échelle donne une mesure relative des branches.

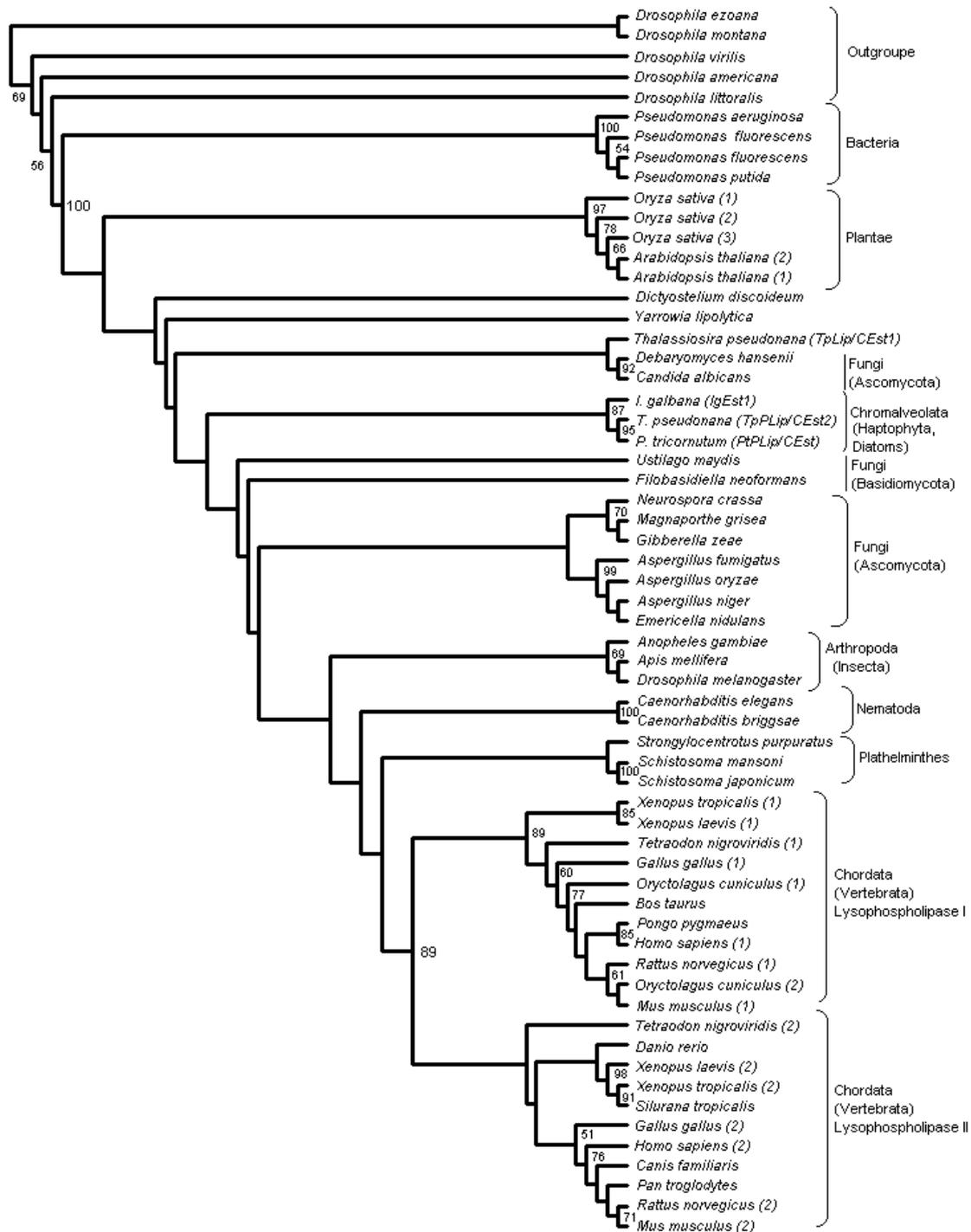


Figure 62 : Phylogénie de l'estérase hypothétique d'*Isochrysis galbana* (IgEst1) et de ses homologues, obtenue avec la méthode du Maximum de Parcimonie sur les régions extraites par Gblocks.

Les séquences utilisées sont les mêmes que celles de la Figure 61. Les groupes d'organismes et les différentes familles d'enzymes sont indiqués par des accolades. L'arbre consensus enraciné a été obtenu avec la méthode du Maximum de Parcimonie et 500 valeurs de bootstraps sur les régions conservées. Seules les valeurs de bootstraps supérieures à 50 % sont indiquées.

5. Caractérisation préliminaire et analyses comparatives du gène *IgEst2* codant une estérase hypothétique chez *Isochrysis galbana*

5.1. Analyse et traduction de la séquence nucléotidique *IgEst2*

La taille de la dernière séquence identifiée, *IgEst2*, est de 1058 pb, sans tenir compte de la queue polyadénylée (Figure 63). La région non codante située en amont du codon d'initiation ATG (positions 67-69) compte 66 nucléotides. La région située en aval du codon stop TAG (positions 769-771) est plus longue de 221 nucléotides. Les régions non traduites encadrent une séquence codante de 702 pb. Celle-ci code une protéine de 234 acides aminés. Un seul résidu cystéine figure dans la séquence protéique. Aucun site potentiel de glycosylation de type NXS/T n'a été détecté.

```

1 CAGTGGTATCAACGCAGAGTACGCGGGGATAGCTTCCGAGCCAATCCTCCCCTCGGCCGCCGCCGAGGGCCAGTC
1 M R Y A V R A A T T S F G A A P A L I
49 GCAGGGCTTGTCACCTCAATCAGGTACGCCGTGCGGGCCGCACTACGTCTTTCGGCGCAGCGCCCGCGCTGATC
20 L L L H G T G D N E Q G L L G A I G Q L A P P D A
124 CTGCTGCTGCATGGCACTGGGACAACGAGCAAGGCCCTTCTCGGCCGATAGGGCAGCTTGCGCCCGCTGACGCC
45 V V V S L R G P L Q A P F G G F R W F E G Y S S A
199 GTTGTGCTCTCGTTGCGCGCCCTCTGCAAGCACCATTGGAGGCTTTCGTTGGTTCGAGGGGTATTCTTCGGCC
70 P E Q I A L D K T V G S S S D A V L A F I E E A P
274 CCAGAGCAGATTGCGCTCGACAAGACCGTCGGTTATCCCTCAGATGCTGACTTGCATTTATTAGGAAGCGCCT
95 A K F G I D P K R V F L L G F S Q G A T L V W T T
349 GCCAAGTTTGGGATCGATCCGAAGCGAGTCTTCTGCTGGGATTTTCGCAGGGCGCCACACTCGTTTGGACGACC
120 L L S R W S R P G L I A G G L T L S G R L F P E L
424 CTCTTGTCCAGATGGTTCGCCCGGGCTCATCGCCGGCGGGCTAACACTGTCTGGCCGGCTGTTTCTTAGCTTA
145 M Q A G T P L Y A R L A D P T Q L R D C A V F A S
499 ATGCAGGCTGTAACCCCACTTATGCGCGTCTCGCCGACCCGACGCAGCTGAGGGATTGTGCCGCTTTTGCCCTCA
170 H G A Q D S V T P V A I G R S N E R L F R D W A P
574 CACGGCGCCCAAGACTCGGTTACCCCGGTAGCTATCGGGCGCTCGAATGAGCGGCTGTTTCGAGACTGGGCACCC
195 G A D L I Y K E D P V A G H E I S K A V A A S M S
649 GGTCCGCACTTAATATATAAGGAGGACCCCTGTTGCAGGCCATGAGATATCGAAGGCCGTGGCTGCTCAATGTCC
220 Q W F R R Y V S G A S K G D Q *
724 CAATGGTTCGGTATGTCTCCGTAAGCTCAAGGGTGTGATCAGTACGACTTCCAGCGGCGAGCAGCACCCGTC
799 CATAGAGCAGGCTGTGGAAGCCAACCTTTGTGATGACGCGCCAAGATGCGTGGGCAAGCGTGCAGTCCACGCGC
874 CCCTCAAACATTCGAACGCTCGTCTCCTCTGCTGTGGCCAGCAAACAGGGATGCATGCCTGTGATGTGTCCAC
949 ACGCTCCATGCCTCTCCTGCTACCCTAGCAGCAGCATGCTAGATGCATAACGCCTGGCGGGCCTGAGCGATTGT
1024 GCACGTGAGCATGAGTGCATACACCACGATGACACAAAAAAAAAAAAAAAAAAAA

```

Figure 63 : Séquence nucléotidique de l'ADNc *IgEst2* et traduction de la région codante

Les amorces dessinées à partir de l'étiquette de séquence transcrite n°ISL00001076 et utilisées en RACE et Nested PCR sont écrites et soulignées en jaune sur fond gris. Les codons d'initiation et de terminaison de traduction sont indiqués en jaune sur fond rouge. Le résidu cystéine est souligné en noir. Les régions non traduites 5' et 3' sont écrites en gras.

5.2. Prédiction des caractéristiques physicochimiques

5.2.1. Signal peptide

Le modèle de prédiction a été établi, à partir de données eucaryotes. Le serveur prédit, dans la région N-terminale d'*IgEst2*, l'existence d'un peptide signal de 27 acides aminés avec une probabilité de 0,962 (Figure 64). Le site de coupure reconnu par la peptidase signal I est prédit à 0,803 entre les résidus G26 et D27. On retrouve une région de résidus chargés positivement, suivie par des résidus hydrophobes puis par des résidus neutres et polaires. La protéine *IgEst2* posséderait, comme *IgEst1*, un signal d'adressage dans sa région N-terminale nécessaire à son éventuelle incorporation dans un compartiment cellulaire ou à son excrétion.

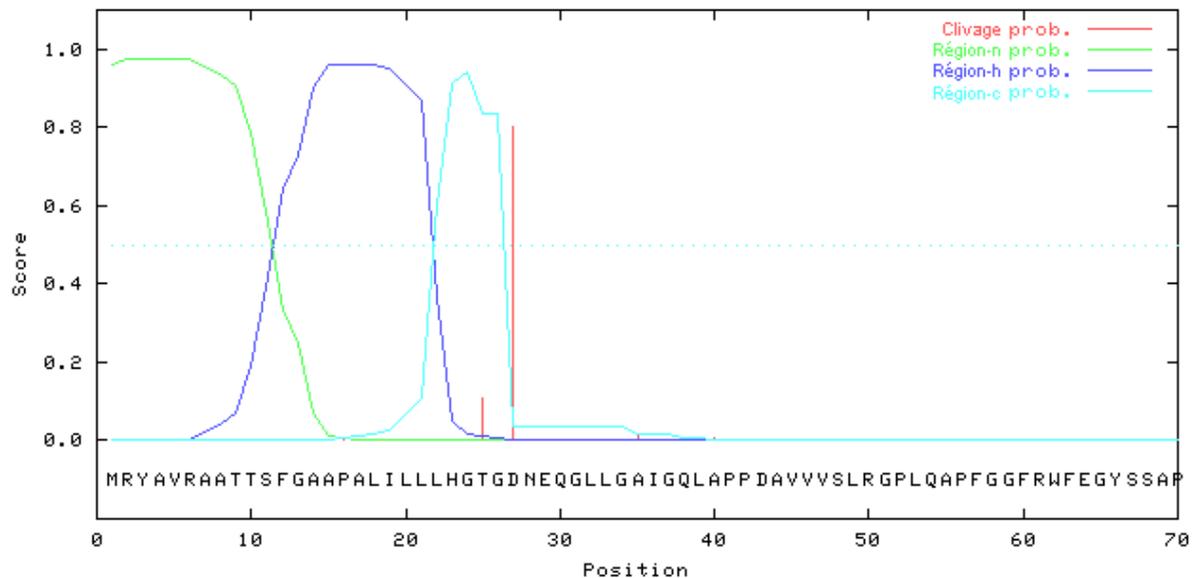


Figure 64 : Prédiction de la présence d'un signal peptidique dans la région N-terminale d'*IgEst2*, avec le serveur SignalP.

L'axe des abscisses indique la position des acides aminés. L'axe des ordonnées indique l'indice de probabilité. Les sites de clivage sont en rouge. Les résidus chargés sont en vert, les résidus hydrophobes en bleu foncé et les résidus hydrophiles en bleu clair.

5.2.2. Masse moléculaire et pH du point isoélectrique

Le cadre ouvert de lecture d'*IgEst2* code une protéine de 234 acides aminés. En tenant compte du signal prédit par le serveur SignalP, la protéine mature serait de 208 résidus après clivage du peptide signal. Cela correspond à une petite protéine de poids moléculaire approximativement égal à 22,24 kDa et de pI théorique égal à 5,80.

5.2.3. Profil d'hydrophobicité

En dehors de la région du peptide signal, aucune zone hydrophobe n'atteint un indice supérieur à 1 sur l'échelle de Kyte et Doolittle (Figure 65). La protéine *IgEst2* ne présente donc pas de régions transmembranaires.

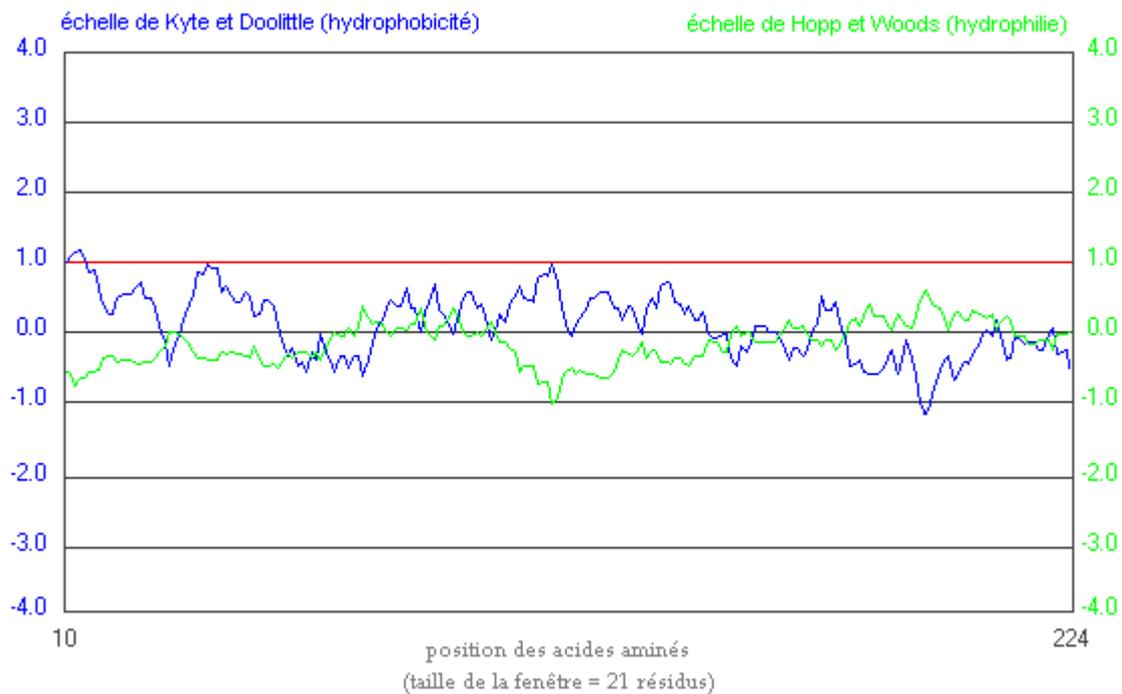


Figure 65 : Profils hydrophobe et hydrophile de la protéine *IgEst2*.

Axe des ordonnées (bleu) : échelle de Kyte et Doolittle. Axe des ordonnées (vert) : échelle de Hopp et Woods. Axe des abscisses : position des acides aminés. La taille de la fenêtre est de 21 acides aminés.

5.3. Prédiction de la fonction enzymatique

5.3.1. Recherche de séquences homologues

Comme pour les deux séquences précédentes, la fonction de la protéine *IgEst2* a été, dans un premier temps, estimée par la recherche de séquences homologues chez les autres organismes. Pour cela, des alignements locaux (blastx, blastp) ont été effectués avec *IgEst2* et les séquences répertoriées dans les banques de données publiques. Les meilleurs scores de ces alignements ont été obtenus avec des séquences d'origine bactérienne, identifiées elles aussi par homologie de séquences et annotées phospholipase/carboxylestérase (*PLip/CEst*).

Pour mettre en évidence les régions conservées, un alignement multiple de séquences protéiques a été réalisé avec le programme MUSCLE. Il compare la séquence *IgEst2* avec six séquences homologues de type phospholipase / carboxylestérase identifiées chez diverses bactéries (Figure 66). Les protéines alignées appartiennent à trois bactéries de la classe des Chloroflexi, *Roseiflexus sp.* (*RspPLip/CEst*), *Chloroflexus aurantiacus* (*CaPLip/CEst*) et *Herpetosiphon aurantiacus* (*HaPLip/CEst*), une cyanobactérie *Synechococcus sp.* (*SspPLip/CEst*), une bactérie de la classe des Flavobacteria *Flavobacterium johnsoniae* (*FjPLip/CEst*) et une bactérie de la classe des Chlorobia, *Chlorobium limicola* (*ClPLip/CEst*).

En dehors des régions N- et C-terminales qui ne s'alignent pas, on distingue plusieurs régions moyennement conservées. On retrouve une région de résidus qui pourrait correspondre à la région du coude nucléophile des lipases et des estérases. Le pentapeptide consensus G-X-S-X-G, propre aux enzymes lipolytiques, y est très conservé. Il apparaît ici sous la séquence G-F-S¹¹⁰-Q-G, séquence identique pour les sept protéines alignées. Les trois résidus Ser110, Asp174 et His208 sont strictement conservés et correspondent très probablement à la triade catalytique caractéristique des lipases, des estérases et des enzymes lipolytiques en général.

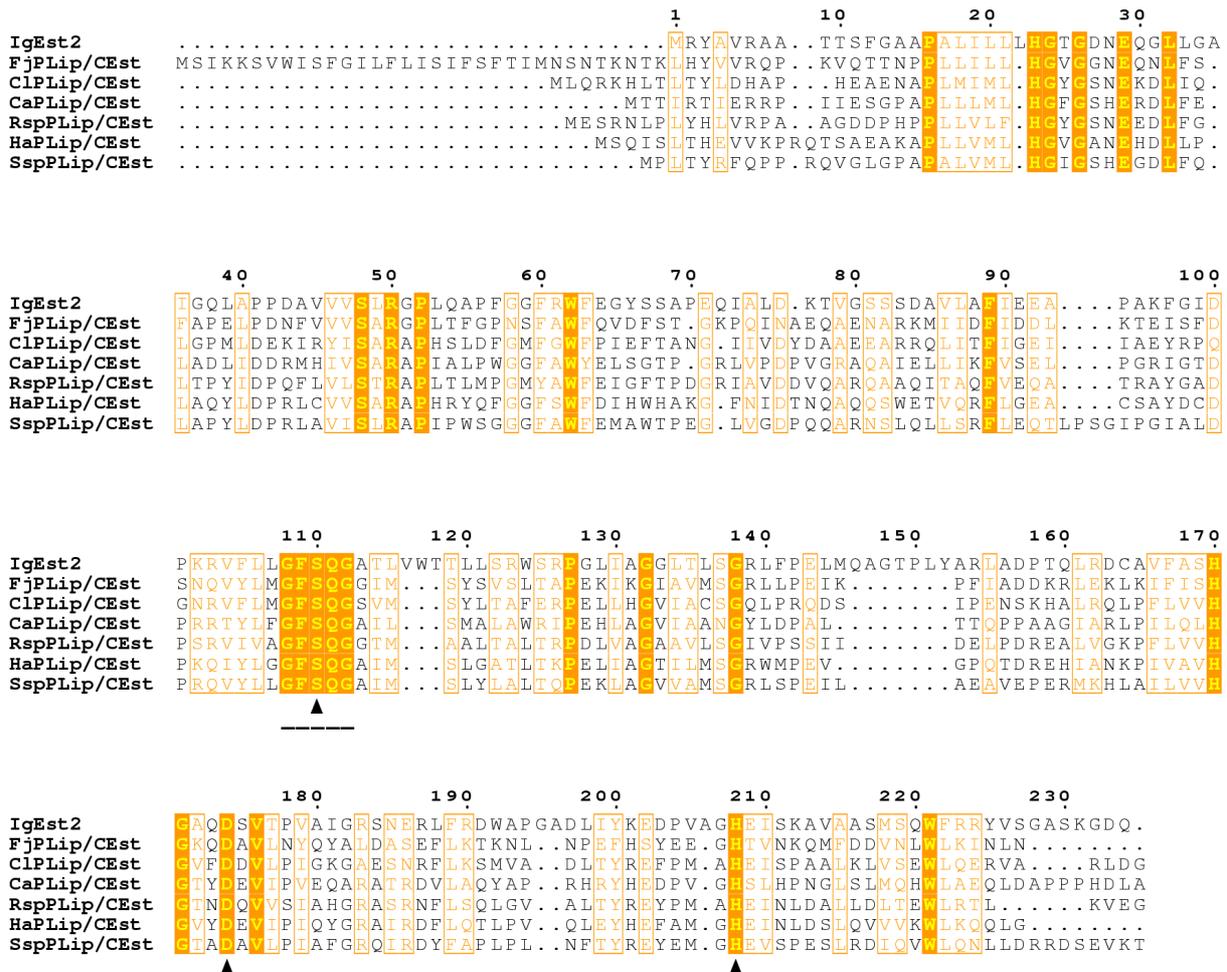


Figure 66 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (IgEst2) avec six de ses homologues chez les bactéries *Flavobacterium johnsoniae* (FjPLip/CEst), *Chlorobium limicola* (ClPLip/CEst), *Chloroflexus aurantiacus* (CaPLip/CEst), *Roseiflexus sp.* (RspPLip/CEst), *Herpetosiphon aurantiacus* (HaPLip/CEst) et une cyanobactérie *Synechococcus sp.* (SspPLip/CEst).

Organisme / numéro d'accèsion : *Roseiflexus sp.* / YP 001277249 ; *Chloroflexus aurantiacus* / ZP 00765825 ; *Herpetosiphon aurantiacus* / ZP 01427120 ; *Synechococcus sp.* / YP 479022 ; *Flavobacterium johnsoniae* / YP 001195579 ; *Chlorobium limicola* / ZP 00511958. Les résidus strictement conservés sont écrits en jaune sur fond orangé. Les résidus similaires sont écrits et encadrés en orange. Les résidus Ser/Asp/His supposés constituer la triade catalytique sont indiqués par des triangles noirs. Le pentapeptide consensus G-X-S-X-G est souligné par un trait noir discontinu.

Le pourcentage d'identité d'*IgEst2* avec ses homologues oscille majoritairement autour de 27 % et n'excède pas 30 % (Tableau 21). Le taux de résidus similaires est plus élevé et varie de 38 % pour *FjPLip/CEst* à 47 % pour *HaPLip/CEst* et *SspPLip/CEst*.

	<i>FjPLip/CEst</i>	<i>CIPLip/CEst</i>	<i>CaPLip/CEst</i>	<i>RspPLip/CEst</i>	<i>HaPLip/CEst</i>	<i>SspPLip/CEst</i>
<i>IgEst2</i>	22 %	27 %	27 %	27 %	29 %	27 %
<i>FjPLip/CEst</i>	38 %	43 %	44 %	44 %	47 %	47 %
<i>CIPLip/CEst</i>		26 %	23 %	26 %	30 %	29 %
		45 %	42 %	46 %	44 %	48 %
<i>CaPLip/CEst</i>			31 %	39 %	34 %	34 %
			51 %	56 %	56 %	51 %
<i>RspPLip/CEst</i>				29 %	35 %	40 %
				48 %	52 %	54 %
<i>HaPLip/CEst</i>					37 %	37 %
					55 %	57 %
						41 %
						59 %

Tableau 21 : Pourcentages de résidus identiques et similaires entre l'estérase hypothétique d'*Isochrysis galbana* (*IgEst2*) et ses homologues procaryotiques *FjPLip/CEst*, *CIPLip/CEst*, *CaPLip/CEst*, *RspPLip/CEst*, *HaPLip/CEst* et *SspPLip/CEst*.

Le pourcentage indiqué en jaune sur fond orange correspond au pourcentage d'identité, c'est-à-dire au pourcentage de résidus strictement conservés entre deux séquences. Les chiffres écrits en orange indiquent la proportion de résidus similaires, c'est-à-dire la proportion de résidus biochimiquement proches.

La Figure 67 illustre un des meilleurs alignements obtenus avec la bactérie *Herpetosiphon aurantiacus* : 29 % de résidus identiques, 50 % de résidus similaires et 11 % de gap.

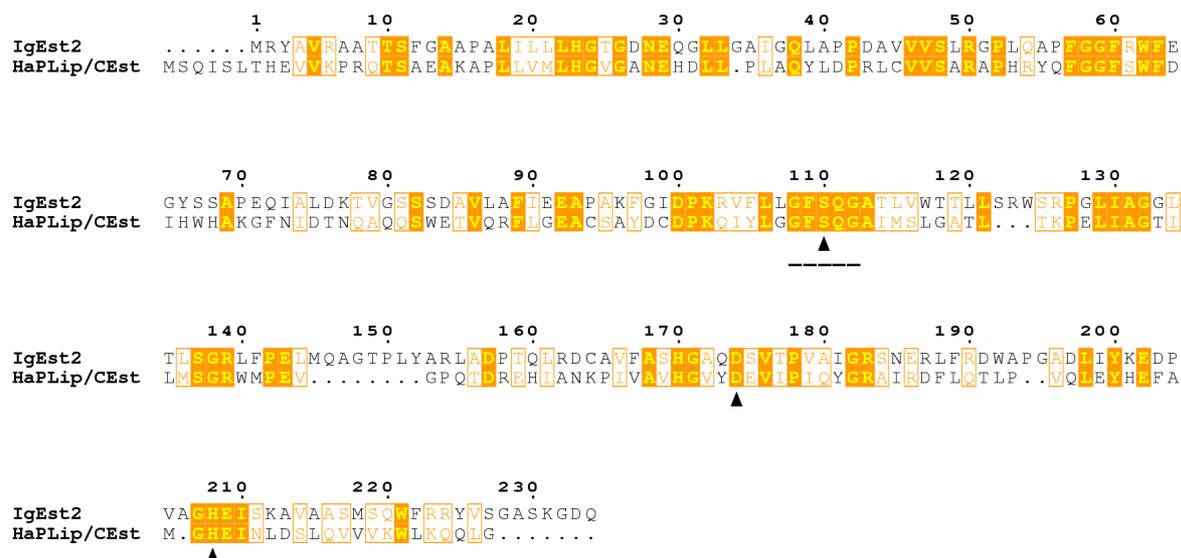


Figure 67 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst2*) avec son plus proche homologue identifié chez la bactérie *Herpetosiphon aurantiacus* (*HaPLip/CEst*).

Les résidus strictement conservés sont écrits en jaune sur fond orange. Les résidus similaires sont écrits et encadrés en orange. Les résidus Ser/Asp/His supposés constituer la triade catalytique sont indiqués par des triangles noirs. Le pentapeptide consensus G-X-S-X-G est souligné par un trait noir discontinu.

5.3.2. Recherche de motifs fonctionnels

La séquence *IgEst2* a été comparée avec les motifs classés et stockés, sous forme de profils, dans la base de données Blocks. A l'instar d'*IgEst1*, la séquence *IgEst2* comporte 2 motifs caractéristiques de la famille d'enzymes, référencée IPR003140 dans la base de données InterPro et qui regroupe des phospholipases et des carboxylestérases. Les deux motifs (ou blocs) sont indexés IPB003140A et IPB003140B (Annexe 10). Le bloc B inclut le pentapeptide consensus G-X-S-X-G majoritairement présent sous la forme G-F-S-Q-G.

5.4. Prédiction de l'organisation structurale

5.4.1. Alignement avec la carboxylestérase de *Pseudomonas fluorescens*

La séquence protéique *IgEst2* a été comparée à sa plus proche homologue, de structure et de fonction connues, la carboxylestérase de la protéobactérie *Pseudomonas fluorescens* (*PfCEst*). Cette carboxylestérase a été étudiée d'un point de vue fonctionnel et structural dans les années 90. Une étude, publiée en 1991, rapporte le clonage et l'expression du gène de cette carboxylestérase chez la bactérie *Escherichia coli* (Hong *et al.*, 1991). Le gène comporte un cadre ouvert de lecture de 654 pb. Il code une protéine de 218 acides aminés et de poids moléculaire théoriquement égal à 23,856 kDa. La séquence protéique contient le pentapeptide G-F-S¹¹⁴-Q-G qui concorde avec la séquence consensus caractéristique des sites actifs à sérine, de type G-X-S-X-G, observés chez un grand nombre d'estérases et de lipases à sérine. Dans cette étude, le gène, codant l'estérase supposée, a été cloné dans le plasmide pUE1. L'enzyme synthétisée par *E. coli* a été purifiée par chromatographie d'affinité sur colonne et caractérisée par mesure de son activité hydrolytique sur différents substrats (esters). Les résultats de mesure d'activité montrent que l'enzyme catalyse l'hydrolyse du phénylacétate et des esters de méthyle et de p-nitrophényl à courtes chaînes d'acides gras (C2-C6). L'estérase purifiée présente un maximum d'activité à 45 °C et à un pH optimal compris entre 6,0 et 9,0. Homodimère, elle est constituée de deux sous unités identiques.

La carboxylestérase de *P. fluorescens* a également fait l'objet d'une analyse structurale par cristallographie aux rayons X (Kim *et al.*, 1997). La structure de l'enzyme a été résolue sous sa forme native et sous la forme d'un complexe avec un inhibiteur. Les sous-unités composant le dimère sont globulaires. Elles sont chacune constituées de feuillets β formant une région centrale flanquée de 6 hélices α . L'étude de la structure a révélé la présence d'une triade

catalytique Ser114, Asp168 et His199. Deux résidus, Leu23 et Gln115, sont, d'après cette étude, impliqués dans la formation du trou oxyanionique. A la différence des lipases, le trou oxyanionique des estérases est préformé et continuellement présent, comme pour la cutinase (Martinez *et al.*, 1994). Le trou oxyanionique des lipases est absent dans la forme inactive et la conformation fermée. Il n'apparaît qu'au moment de l'activation interfaciale qui accompagne l'ouverture du volet amphiphile (Derewenda *et al.*, 1992). Les chaînes latérales de quatre résidus hydrophobes Ile58, Ile 70, Met73 et Val 170, constituent le site hydrophobe acceptant les chaînes d'acide gras.

L'alignement, réalisé avec MUSCLE, a permis d'identifier avec plus de précision les principaux résidus d'*IgEst2* susceptibles d'être impliqués dans le mécanisme catalytique (Figure 68). La triade catalytique apparaît strictement conservée et correspond, Chez *IgEst2*, aux résidus Ser110, Asp174 et His208. Les résidus Gln115 et Val170 impliqués dans la stabilisation du substrat, chez *P. fluorescens*, sont également conservés chez *I. galbana* en positions 111 et 176.

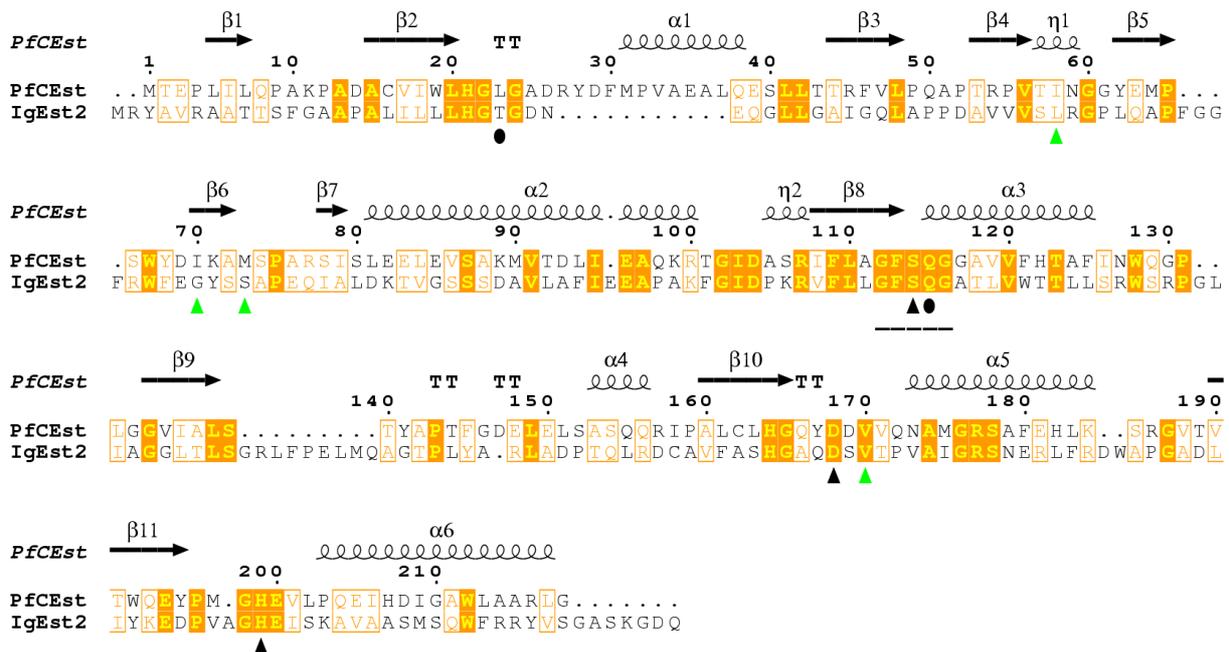


Figure 68 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst2*) avec la carboxylestérase de *Pseudomonas fluorescens* (*PfCEst*) (Swiss-Prot Q53547 ; code PDB 1auo).

Les résidus strictement conservés sont indiqués en jaune sur fond orangé, les résidus bien conservés écrits et encadrés en orange. La triade catalytique Ser/Asp/His est indiquée par des triangles noirs. Le pentapeptide consensus G-X-S-X-G est souligné en noir. Les triangles verts correspondent aux résidus impliqués dans la fixation du substrat et de ses intermédiaires. Les résidus indiqués par des ronds noirs interviennent dans la stabilisation de l'intermédiaire oxyanionique. La structure secondaire de la carboxylestérase de *P. fluorescens* figure au-dessus de l'alignement.

5.4.2. Modèle tridimensionnel

Afin d'avoir une idée de la structure tridimensionnelle de l'estérase 2 d'*Isochrysis galbana*, un modèle a été construit avec le programme Geno3D en utilisant comme référence la structure connue de la carboxylestérase de *P. fluorescens* (code PDB 1auo) (Figure 69). La prédiction doit être analysée avec prudence car le taux d'identité est faible (21 %). La protéine *IgEst2* posséderait, comme la carboxylestérase de *P. fluorescens*, une structure globulaire avec repliement α/β caractéristique de la superfamille des α/β hydrolases, constitué de feuillets β entourés de part et d'autre par des hélices α .

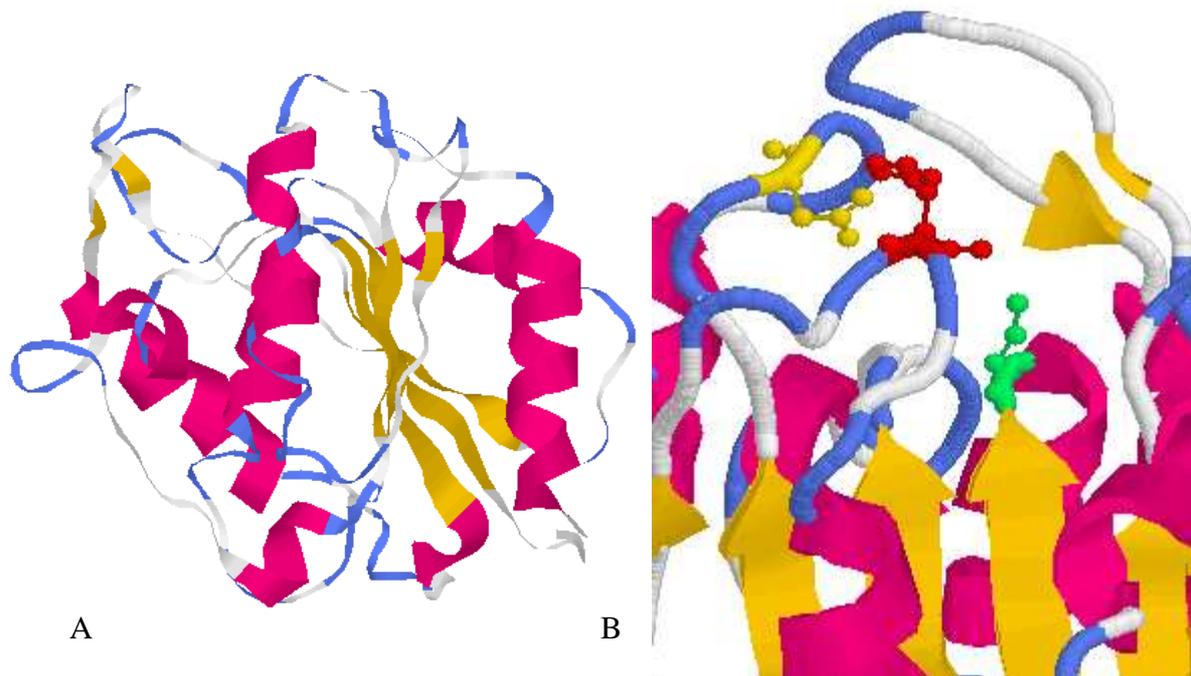


Figure 69 : Modèle structural de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst2*).

Le modèle 3-D d'*IgEst2* a été établi grâce au programme Geno3D, par comparaison avec les données structurales de la carboxylestérase de *Pseudomonas fluorescens* (code PDB 1auo). Les images tridimensionnelles ont été éditées avec le programme Protein Explorer. **A)** Structure globulaire. **B)** Les trois résidus de la triade catalytique sont représentés en vert (Ser110), jaune (Asp174) et rouge (His208).

5.5. Séquences homologues à *IgEst2* chez les microalgues

La séquence d'*IgEst2* a été utilisée pour rechercher des séquences homologues dans les banques de données publiques de microalgues. Cette analyse a permis d'identifier 7 séquences homologues chez les microalgues *Cyanidioschyzon merolae*, *Ostreococcus lucimarinus*, *Ostreococcus tauri*, *Phaeodactylum tricornutum* et *Thalassiosira pseudonana*. Les séquences protéiques identifiées correspondent à celles déjà identifiées pour *IgEst1* (Figure 70).

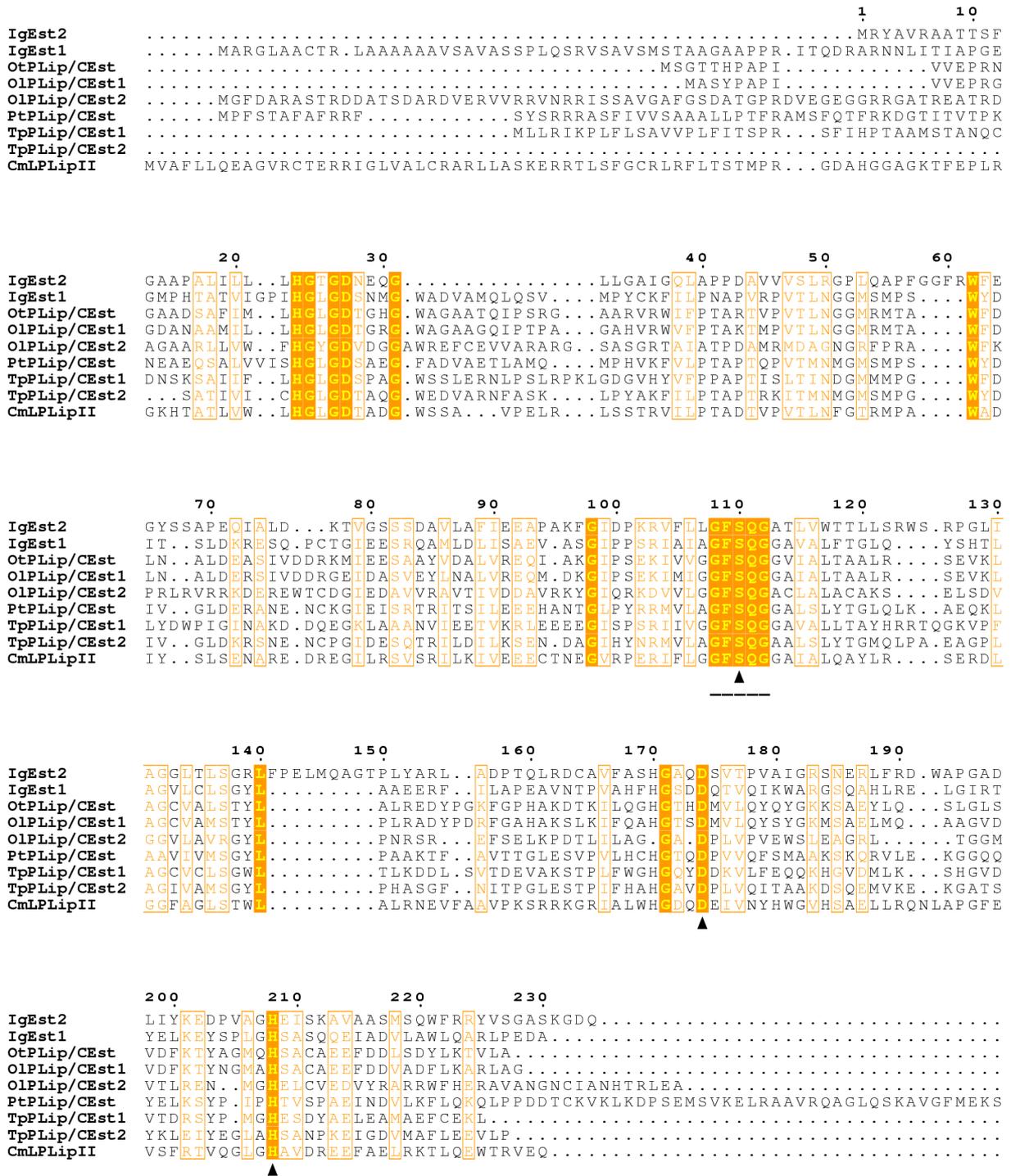


Figure 70 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (IgEst2) avec sept de ses homologues chez les microalgues *Ostreococcus tauri* (OtPLip/CEst), *Ostreococcus lucimarinus* (OlPLip/CEst1 et OlPLip/CEst2), *Phaeodactylum tricornerutum* (PtPLip/CEst), *Thalassiosira pseudonana* (TpPLip/CEst1 et TpPLip/CEst2) et *Cyanidioschyzon merolae* (CmLPLipII).

Organisme / numéro d'identifiant : *O. tauri* / 26019 ; *O. lucimarinus* / 32853 (OlPLip/CEst1) / 93362 (OlPLip/CEst2) ; *P. tricornerutum* / 21816 ; *T. pseudonana* / 6090 (TpPLip/CEst1) / 18091 (TpPLip/CEst2) ; *C. merolae* / CMF176C. Les résidus strictement conservés sont écrits en jaune sur fond orangé. Les résidus similaires sont écrits et encadrés en orange. Les résidus Ser/Asp/His supposés constituer la triade catalytique sont indiqués par des triangles noirs. Le pentapeptide consensus G-X-S-X-G est souligné par un trait noir discontinu.

Les pourcentages d'homologie d'*IgEst2* avec ses homologues varie de 14 à 18 % (Tableau 22).

	<i>IgEst1</i>	<i>OtPLip/CEst</i>	<i>OiPLip/CEst1</i>	<i>OiPLip/CEst2</i>	<i>PtPLip/CEst</i>	<i>TpPLip/CEst1</i>	<i>TpPLip/CEst2</i>	<i>CmPLipII</i>
<i>IgEst2</i>	16 %	17 %	18 %	16 %	14 %	16 %	16 %	18 %
	28 %	33 %	35 %	25 %	26 %	31 %	32 %	29 %
<i>IgEst1</i>		27 %	26 %	18 %	32 %	22 %	35 %	26 %
		44 %	43 %	34 %	45 %	37 %	49 %	40 %
<i>OtPLip/CEst</i>			72 %	17 %	20 %	27 %	28 %	31 %
			82 %	28 %	37 %	46 %	47 %	44 %
<i>OiPLip/CEst1</i>				16 %	22 %	27 %	30 %	30 %
				27 %	38 %	46 %	50 %	43 %
<i>OiPLip/CEst2</i>					15 %	15 %	16 %	20 %
					30 %	27 %	28 %	33 %
<i>PtPLip/CEst</i>						18 %	35 %	22 %
						33 %	47 %	34 %
<i>TpPLip/CEst1</i>							23 %	25 %
							40 %	42 %
<i>TpPLip/CEst2</i>								22 %
								34 %

Tableau 22 : Pourcentages de résidus identiques et similaires entre l'estérase hypothétique d'*I. galbana* (*IgEst2*) et ses homologues microalgales chez *Ostreococcus tauri* (*OtPLip/CEst*), *Ostreococcus lucimarinus* (*OiPLip/CEst1* et *OiPLip/CEst2*), *Phaeodactylum tricorutum* (*PtPLip/CEst*), *Thalassiosira pseudonana* (*TpPLip/CEst1* et *TpPLip/CEst2*) et *Cyanidioschyzon merolae* (*CmLPLipII*).

Le pourcentage indiqué en jaune sur fond orange correspond au pourcentage d'identité, c'est-à-dire au pourcentage de résidus strictement conservés entre deux séquences. Les chiffres écrits en orange indiquent la proportion de résidus similaires, c'est-à-dire la proportion de résidus biochimiquement proches.

5.6. Classification

IgEst2 est similaire aux estérases d'origine bactérienne classées dans la VI^{ème} famille parmi les huit familles décrites par Arpigny et Jaeger en 1999 et dont les caractéristiques sont indiquées dans le Tableau 2, page 30. Avec une taille moléculaire comprise entre les 23 et 26 kDa, les enzymes de cette famille figurent parmi les plus petites estérases connues. La carboxylestérase de *Pseudomonas fluorescens* appartient à cette classe d'enzyme. Active sous forme dimérique, elle hydrolyse préférentiellement des substrats de petite taille moléculaire avec une large gamme de spécificités. Elle ne montre aucune activité à l'égard des triglycérides à longues chaînes. On connaît peu de choses à propos des autres enzymes de cette catégorie. Leurs séquences protéiques proviennent de programmes de séquençage génomique, sauf pour l'estérase de *Spirulina platensis* qui a été clonée. Les enzymes de la

famille VI présentent approximativement 40 % de similitude avec des phospholipases A₂, Ca²⁺-indépendantes (iPLA₂), qui présentent aussi des activités lysophospholipasiques (Portilla *et al.*, 1998). Un alignement a donc été réalisé entre *IgEst2* et 2 iPLA₂ démontrant que la séquence d'*I. galbana* partage aussi des similarités de séquences avec ce type de phospholipase (Figure 71).

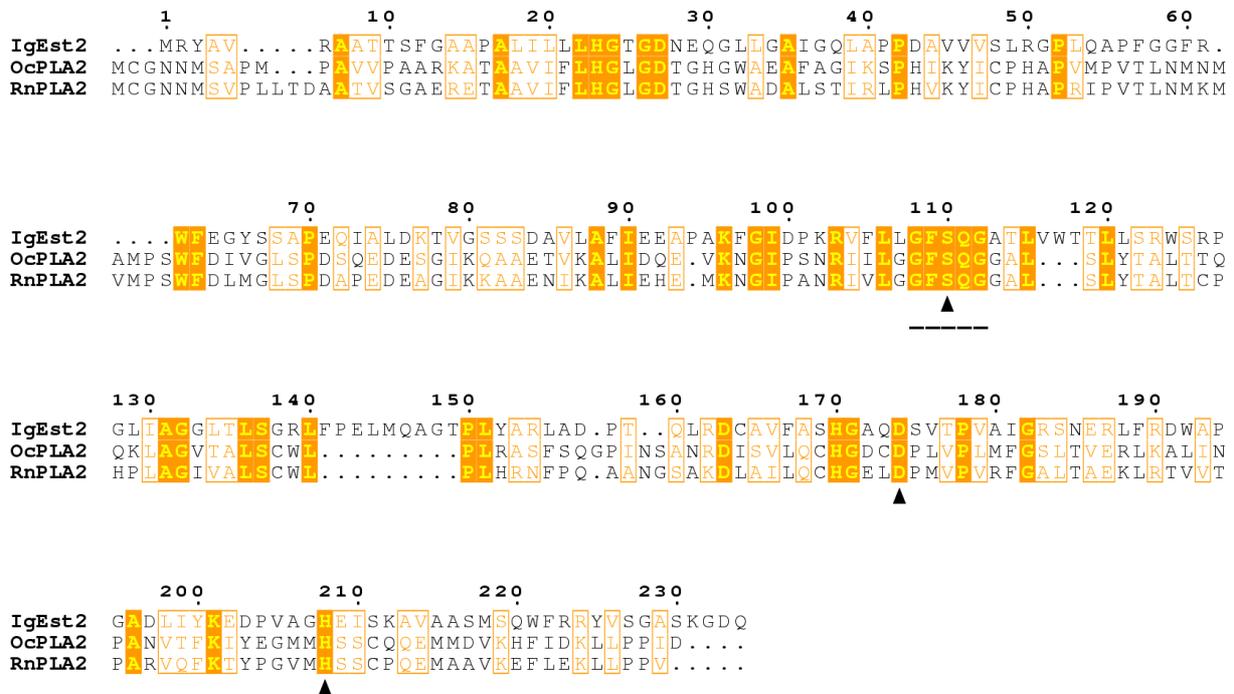


Figure 71 : Alignement de l'estérase hypothétique d'*Isochrysis galbana* (*IgEst2*) avec les phospholipases A₂ Ca²⁺-indépendantes des vertébrés *Oryctolagus cuniculus* (*OcPLA2*) et *Rattus norvegicus* (*RnPLA2*).

Organisme / numéro d'accèsion : *O. cuniculus* / NP 001103287 ; *R. norvegicus* / NP_112632. Les résidus strictement conservés sont écrits en jaune sur fond orange. Les résidus similaires sont écrits et encadrés en orange. Les résidus Ser/Asp/His supposés constituer la triade catalytique sont indiqués par des triangles noirs. Le pentapeptide consensus G-X-S-X-G est souligné par un trait noir discontinu.

5.7. Structure intron/exon du gène *IgEst2*

Pour mettre en évidence la présence éventuelle d'introns, une PCR sur l'ADN génomique avec les couples d'amorces *EcoRIEst2/Est2XbaI* a été réalisée. Après électrophorèse sur gel d'agarose, la taille de l'amplicon identique à celle du transcrit laissait supposer l'absence d'introns. L'alignement de la séquence d'ADN génomique obtenue avec celle du transcrit a confirmé cette absence (Annexe 12).

6. Clonage et induction de l'expression des gènes

6.1. Clonage de la séquence codante *IgLip*

6.1.1. Construction génétique pYES2/*IgLip*

La carte de restriction de la séquence codante *IgLip* a été établie en recherchant, avec le programme pDRAW32, la présence des sites de coupure des enzymes de restriction *HindIII*, *KpnI*, *SacI*, *BamHI*, *BstXI*, *EcoRI*, *NotI*, *XhoI*, *SphI* et *XbaI*. Ces dix sites de coupures composent le site de clonage du plasmide pYES2. Les sites de restriction des enzymes *XhoI*, *SacI*, *BstXI* et *NotI* sont présents dans la séquence codante (Figure 72A). Les deux sites de restriction *EcoRI* et *XbaI* ont été choisis après analyse de la carte pour le clonage du gène. Ils ont été rajoutés aux extrémités 5' et 3' de la séquence codante pour en permettre l'insertion unidirectionnelle dans le plasmide. La taille du plasmide recombiné avec l'insert *IgLip* est de 7194 pb (Figure 72B).

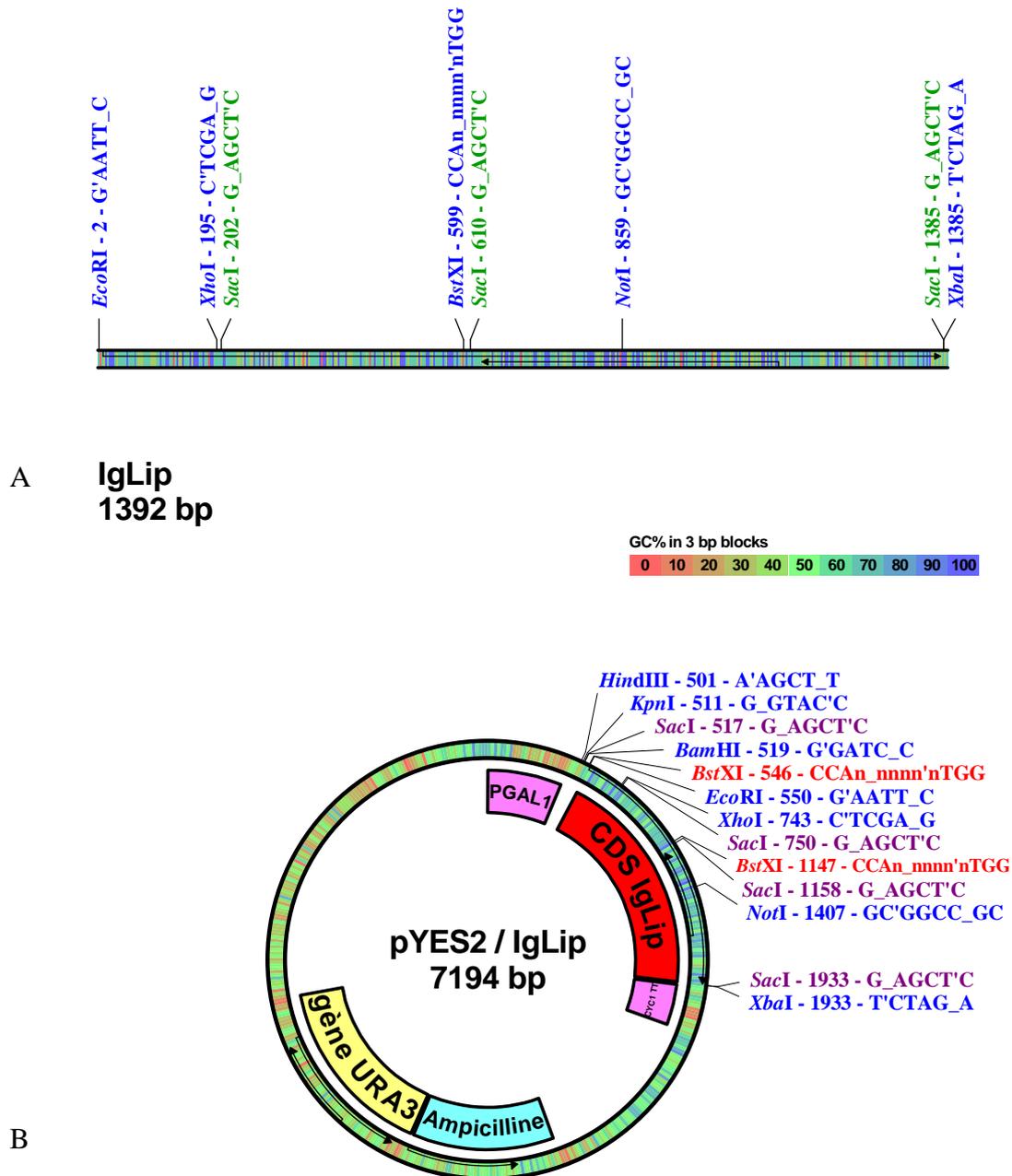


Figure 72 : Construction génétique du vecteur pYES2/IgLip.

Les cartes ont été réalisées et éditées avec pDRAW32. Le pourcentage de GC des séquences est signalé par un code de couleur. Les flèches indiquent la présence et le sens des ORF.

A) Carte de restriction de la séquence codante *IgLip* après ajout des adaptateurs *EcoRI* et *XbaI*.
B) Carte de restriction du plasmide pYES2 recombiné avec *IgLip*. Les promoteur et terminateur sont indiqués sur fond mauve, l'insert sur fond rouge. Le gène de résistance à l'ampicilline est en bleu. Le gène codant une enzyme impliquée dans la synthèse de l'uracile est mentionné en jaune.

6.1.2. Transformation d'*Escherichia coli* et analyse des clones positifs

Le plasmide recombiné a été cloné avec la souche d'*Escherichia coli* TOP10F'. Plusieurs clones positifs ont été isolés. L'ADN plasmidique du clone 1 a été extrait et des contrôles, destinés à vérifier la présence de l'insert, ont été réalisés (Figure 73). Le premier contrôle a été effectué par une double digestion enzymatique avec les enzymes *EcoRI* et *XbaI*. Deux bandes ont été obtenues (Figure 73A). Leurs tailles correspondent approximativement à celles attendues soit 5811 pb pour l'une et 1383 pb pour l'autre. La présence de l'insert a également été contrôlée par amplification avec des couples d'amorces spécifiques à celui-ci (Figure 73B). Quatre régions de 658, 652, 461 et 1392 pb ont donc été amplifiées avec les amorces appropriées. Les amplicons obtenus présentent une taille similaire. Parallèlement, des amplifications témoins ont été réalisées, avec les mêmes couples d'amorces spécifiques d'*IgLip* sur des bactéries transformées avec le plasmide non recombiné.

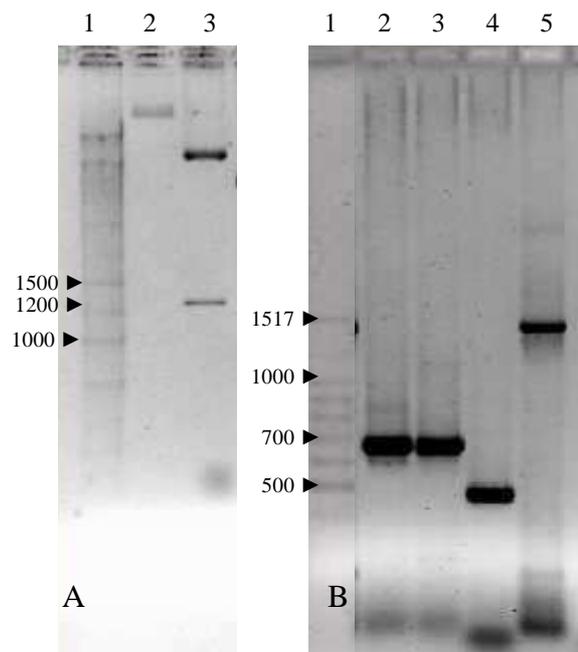


Figure 73 : Electrophorèse en gel d'agarose du plasmide pYES2/IgLip digéré ou amplifié (clone 1).

Les acides nucléiques ont été séparés sur gels à 2 et 1 % (w/v). **A)** Vérification de la présence de l'insert *IgLip* par digestion enzymatique : plasmide non digéré (piste 2), plasmide digéré avec *EcoRI* et *XbaI* (piste 3). **B)** Vérification de la présence de l'insert *IgLip* dans le clone 1 par PCR : les amplicons 1, 2, 3 et 4 ont été respectivement obtenus avec les couples d'amorces *EcoRI*Lip/*NLipAS*, *NLipS*/*LipXbaI*, *LipS*/*LipAS* et *EcoRI*Lip/*LipXbaI*. Les marqueurs de taille (ADN 2-Log et 100 pb) sont en paires de base.

6.1.3. Séquençage de l'insert *IgLip*

Le séquençage de l'insert *IgLip* a révélé la présence de plusieurs mutations dont une délétion du nucléotide T83 en début de séquence (Annexe 13). En décalant le cadre de lecture d'un nucléotide, cette délétion risque d'entraver la traduction et la synthèse d'une protéine fonctionnelle. Pour cette raison, l'étude de l'expression d'*IgLip* n'a pas été poursuivie. Cette mutation a été retrouvée chez tous les clones analysés. Pour éviter les désagréments liés à la dégénérescence génétique de la souche utilisée qui peut se produire au cours des repiquages successifs, le laboratoire devrait envisager la mise au point d'une méthode de cryopréservation de la microalgue *I. galbana* en testant différents agents de cryopréservation, et différentes stratégies de refroidissement. Cette technique permettrait de démarrer régulièrement des nouvelles cultures à partir de stocks axéniques et génétiquement stables. Une étude portant sur la cryopréservation de plusieurs espèces de microalgues a été conduite récemment (Rhodes *et al.*, 2006).

6.2. Clonage et expression de la séquence codante *IgEst1*

6.2.1. Construction génétique pYES2/*IgEst1*

La carte de restriction de la séquence codante *IgEst1* a été établie en recherchant la présence des sites de restriction des enzymes *HindIII*, *KpnI*, *SacI*, *BamHI*, *BstXI*, *EcoRI*, *NotI*, *XhoI*, *SphI* et *XbaI* (Figure 74A). Ces dix sites de coupures composent le site de clonage du plasmide pYES2. Les sites de restriction des enzymes *SacI* et *SphI* ont été reconnus dans la séquence. Ces deux enzymes ne pouvaient donc pas être utilisées pour le clonage du gène. Les deux sites de restriction *EcoRI* et *XbaI* ont été choisis après analyse de la carte pour le clonage. Ils ont été rajoutés aux extrémités 5' et 3' de la séquence codante pour en permettre l'insertion plasmidique unidirectionnelle. La taille du plasmide recombiné avec l'insert *IgEst1* est de 6648 pb (Figure 74B).

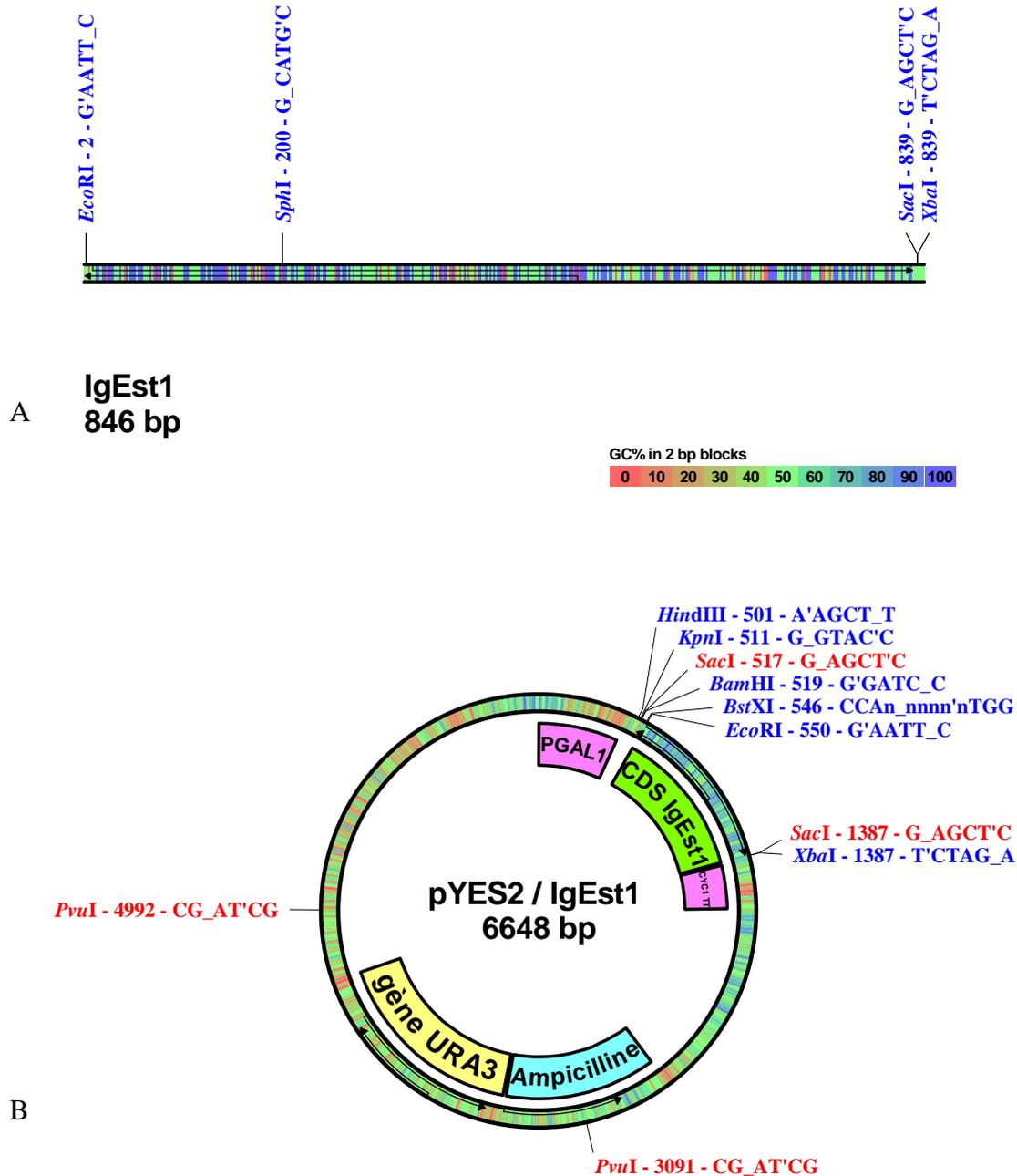


Figure 74 : Construction génétique du vecteur pYES2/IgEst1.

Les cartes ont été réalisées et éditées avec pDRAW32. Le pourcentage de GC des séquences est signalé par un code de couleur. Les flèches indiquent la présence et le sens des ORF.

A) Carte de restriction de la séquence codante *IgEst1* après ajout des adaptateurs *EcoRI* et *XbaI*.
B) Carte de restriction du plasmide pYES2 recombiné avec *IgEst1*. Les promoteur et terminateur sont indiqués sur fond mauve, l'insert sur fond vert. Le gène de résistance à l'ampicilline est en bleu. Le gène codant une enzyme impliquée dans la synthèse de l'uracile est mentionné en jaune.

6.2.2. Transformation d'*Escherichia coli* et analyse des clones positifs

Les bactéries recombinantes ont été sélectionnées sur milieu LB Amp⁺. Plusieurs clones positifs ont ensuite été isolés, cultivés et analysés. L'ADN plasmidique a été extrait du clone A. La présence de l'insert a été contrôlée par digestion enzymatique et par PCR (Figure 75). La double digestion du plasmide pYES2 avec les enzymes *EcoRI* et *XbaI* a produit un fragment de taille proche de celle de l'insert *IgEst1* (846 pb). Les digestions effectuées avec *SacI* et *PvuI* ont également confirmé la présence de l'insert. En effet, elles ont, chacune, sectionné le plasmide en deux fragments de tailles égales à celles attendues, soit 4747 et 1901 pb avec *PvuI* et 5778 et 870 pb avec *SacI* (Figure 75 A). La présence de l'insert *IgEst1* a aussi été confirmée par PCR en utilisant des amorces spécifiques à ce gène. La taille des amplicons obtenus a été déterminée graphiquement par référence au marqueur de taille. Les tailles correspondent approximativement à celles attendues qui sont pour les amplicons 1, 2, 3 et 4 égales à 846, 407, 394 et 315 pb respectivement (Figure 75 B). Parallèlement, des amplifications témoins ont été réalisées, avec les mêmes couples d'amorces spécifiques d'*IgEst1* mais sur des bactéries transformées avec le plasmide non recombiné.

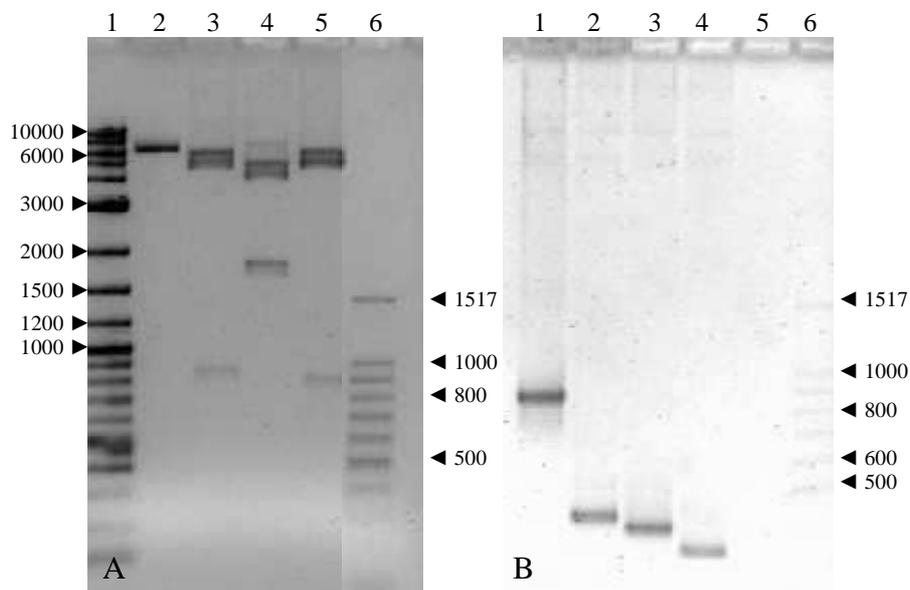


Figure 75 : Electrophorèse en gel d'agarose du plasmide pYES2/IgEst1 digéré ou amplifié (clone A).

Les acides nucléiques ont été séparés sur gels à 2 et 1 % (w/v). **A)** Vérification de la présence de l'insert *IgEst1* par digestion enzymatique : plasmide non digéré (piste 2), plasmide digéré avec *SacI* (piste 3), avec *PvuI* (piste 4) et double digestion avec *EcoRI* et *XbaI* (piste 5). **B)** Vérification de la présence de l'insert *IgEst1* dans le clone A par PCR : les amplicons 1, 2, 3 et 4 ont été respectivement obtenus avec les couples d'amorces *EcoRI**Est1*/*Est1**XbaI*, *Est1S*/*Est1AS*, *EcoRI**Est1*/*NEst1AS* et *NEst1S*/*Est1XbaI*. Les marqueurs de taille (ADN 2-Log et 100 pb) sont en paires de base.

6.2.3. Séquençage de l'insert *IgEst1* (clone A)

L'insert *IgEst1* du clone A a été séquençé afin de vérifier l'intégrité de la séquence. Comme nous pouvons le voir en Annexe 14, 2 mutations se sont produites : la substitution de la cytosine 309 par une adénine et la délétion en fin de séquence d'une adénine. La délétion en fin de séquence n'a probablement aucune incidence ni sur la transcription du gène ni sur la traduction et la fonctionnalité de la protéine. En effet, cette mutation génère la formation d'un autre codon stop (Figure 76). De plus, les derniers acides aminés ne jouent probablement aucun rôle dans le mécanisme catalytique de l'enzyme et interviennent peu dans le repliement de la protéine. Quant à la substitution, c'est une mutation muette car le nouveau codon (CCC) code pour le même acide aminé (proline) que le codon initial (CCA). La délétion cause un raccourcissement du polypeptide de 2 résidus en C-terminal. L'expression de l'insert du clone A sera donc étudiée.

```

CloneA      MARGLAACTRLAAAAA AVSAVASSPLQSRVSAVSMSTAAGAAPPRI TQDRARNNLIT IAP 60
IgEst1      MARGLAACTRLAAAAA AVSAVASSPLQSRVSAVSMSTAAGAAPPRI TQDRARNNLIT IAP 60
*****
CloneA      GEGMPHTATVIGPIHGLGDSNMGWADVAMQLQSVMPYCKFIL P NAPVVRPVTLNGGSMPS 120
IgEst1      GEGMPHTATVIGPIHGLGDSNMGWADVAMQLQSVMPYCKFIL P NAPVVRPVTLNGGSMPS 120
*****
CloneA      WYDITSLDKRESQPCTGIEESRQAMLDLISAEVASGIPPSRIAIAGFSQGGAVALFTGLQ 180
IgEst1      WYDITSLDKRESQPCTGIEESRQAMLDLISAEVASGIPPSRIAIAGFSQGGAVALFTGLQ 180
*****
CloneA      YSHTLAGVLC LSGYLAAEERFILAPEAVNTPVAHFHGSDDQTVQIKWARGSQAHLRELGI 240
IgEst1      YSHTLAGVLC LSGYLAAEERFILAPEAVNTPVAHFHGSDDQTVQIKWARGSQAHLRELGI 240
*****
CloneA      RTYELKEYSPLGHSASQQEIADVLAWLQARLP G TPEL 277
IgEst1      RTYELKEYSPLGHSASQQEIADVLAWLQARLP E DA-- 275
*****

```

Figure 76 : Alignement via ClustalW de la séquence protéique déduite du transcrit isolé chez *Isochrysis galbana* (*IgEst1*) avec la séquence de l'insert cloné chez *Escherichia coli* (clone A). Les mutations sont indiquées en rouge sur fond gris.

6.2.4. Induction de l'expression d'*IgEst1* dans *Saccharomyces cerevisiae*

L'expression du transcrit a été induite par la présence de galactose dans le milieu de culture SC-U des levures. La présence du transcrit a été détectée par amplification avec le couple d'amorces *EcoRIEst1/Est1XbaI* à partir de cultures de 12 heures. Pour s'assurer de l'absence d'ADN contaminant dans les échantillons d'ARN, des amplifications témoins ont été réalisées à partir d'ARN total. L'effet des sucres inhibiteur (glucose) et inducteur (galactose) ont été testés en parallèle. N'ayant pas utilisé de standard interne constitutif dans cette expérimentation, seules la présence ou l'absence des transcrits pourront être déterminées. Les

résultats montrent que le système d'expression fonctionne au niveau transcriptionnel. L'insert *IgEst1* est transcrit en présence de glucose et de galactose (Figure 77).

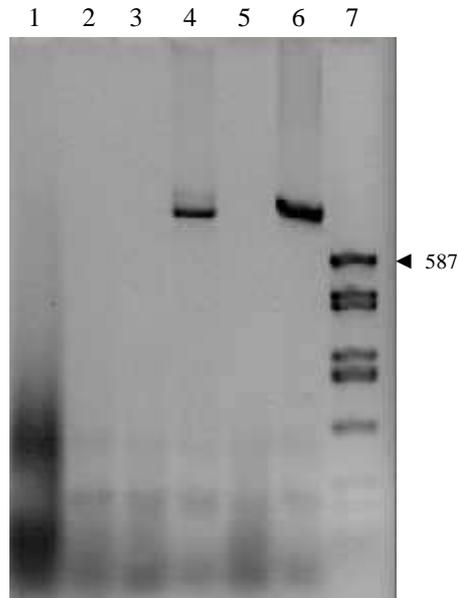


Figure 77 : Electrophorèse en gel d'agarose du transcrit *IgEst1* amplifié par PCR après induction de l'expression.

Les acides nucléiques ont été séparés sur gel à 1 % (w/v) d'agarose. Le marqueur de taille utilisé est le plasmide PUC18 digéré avec *HaeIII* (piste 7). Les amplifications ont été réalisées avec les couples d'amorces *EcoRIEst1/Est1XbaI*. Piste 1 : PCR sur ARN extrait de levures transformées avec le plasmide pYES2 non recombiné. Piste 2 : PCR sur ADNc de levures transformées avec le plasmide pYES2 non recombiné. Piste 3 : PCR sur ARN extrait de levures transformées avec le plasmide recombiné pYES2/*IgEst1* et cultivées en présence de glucose. Piste 4 : PCR sur ADNc de levures transformées avec le plasmide recombiné pYES2/*IgEst1* et cultivées en présence de glucose. Piste 5 : PCR sur ARN extrait de levures transformées avec le plasmide recombiné pYES2/*IgEst1* et cultivées en présence de galactose. Piste 6 : PCR sur ADNc de levures transformées avec le plasmide recombiné pYES2/*IgEst1* et cultivées en présence de galactose.

Les premières mesures d'activités enzymatiques réalisées n'ont pas permis de mettre en évidence une activité lipolytique significative.

6.3. Clonage et expression de la séquence codante *IgEst2*

6.3.1. Construction génétique pYES2/*IgEst2*

La carte de restriction révèle que seul le site de coupure de l'enzyme *KpnI* ne peut être utilisé comme adaptateur pour l'insertion et le clonage d'*IgEst2* (Figure 78A). Ce sont à nouveau les deux sites de restriction *EcoRI* et *XbaI* qui ont été choisis pour l'insertion dans pYES2. Ils ont été rajoutés aux extrémités 5' et 3' de la séquence codante pour en permettre l'insertion plasmidique unidirectionnelle. La taille du plasmide recombiné avec l'insert *IgEst2* est de 6525 pb (Figure 78B).

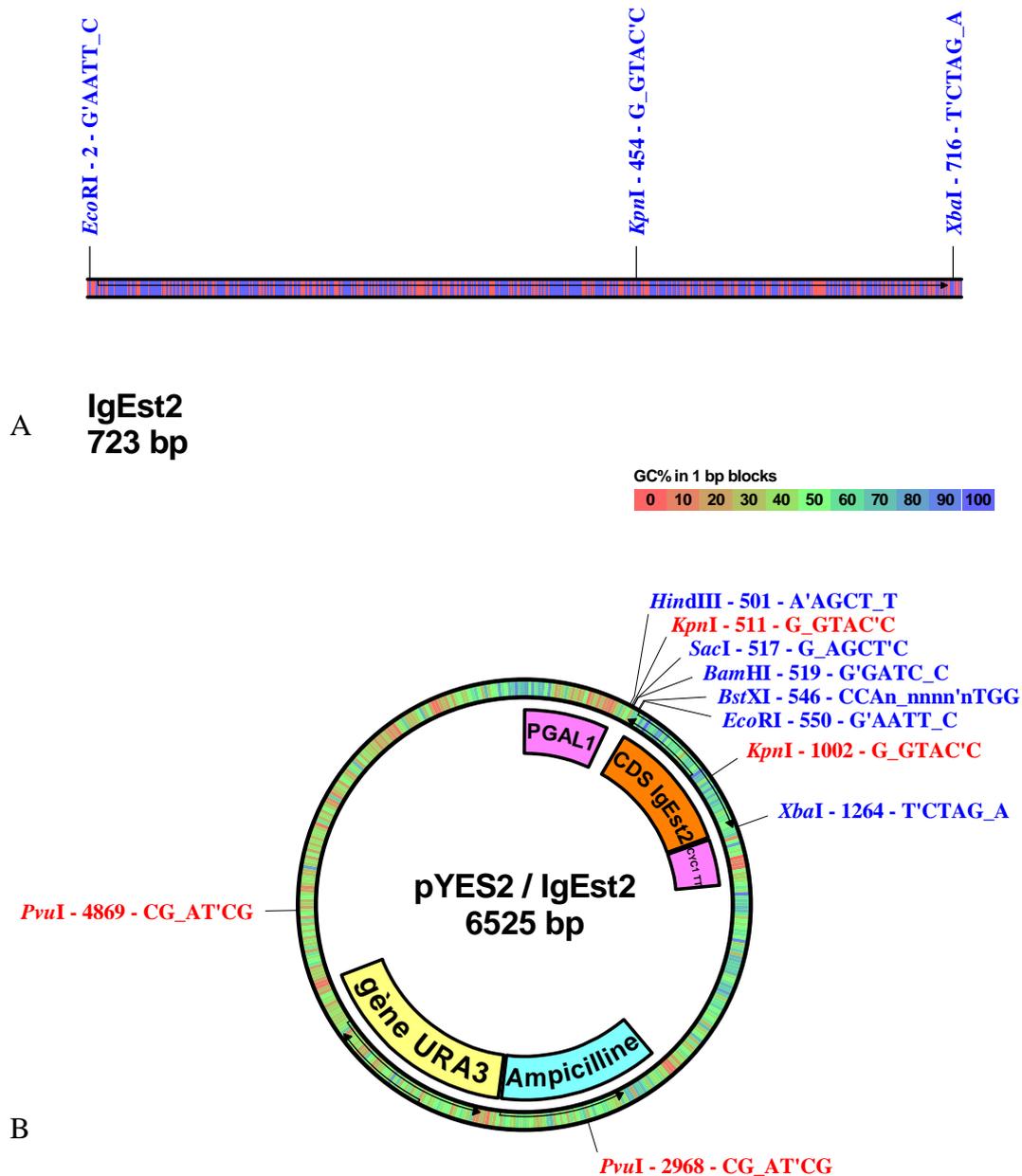


Figure 78 : Construction génétique du vecteur pYES2/IgEst2.

Les cartes ont été réalisées et éditées avec pDRAW32. Le pourcentage de GC des séquences est signalé par un code de couleur. Les flèches indiquent la présence et le sens des ORF.

A) Carte de restriction de la séquence codante *IgEst2* après ajout des adaptateurs *EcoRI* et *XbaI*.
B) Carte de restriction du plasmide pYES2 recombiné avec *IgEst2*. Les promoteur et terminateur sont indiqués sur fond mauve, l'insert sur fond orange. Le gène de résistance à l'ampicilline est en bleu. Le gène codant une enzyme impliquée dans la synthèse de l'uracile est mentionné en jaune.

6.3.2. Analyse des clones positifs

La Figure 79 ci-dessous présente les résultats obtenus pour le contrôle du clone α transformé avec le plasmide pYES2/IgEst2. Les digestions enzymatiques génèrent des fragments de tailles approximativement égales à celles attendues qui sont de 491 et 6034 pb avec *KpnI* et de 1901 et 4624 pb avec *PvuI*. La présence de l'insert *IgEst2* a aussi été confirmée par PCR en utilisant des amorces spécifiques à ce gène. Les tailles des amplicons correspondent à celles attendues qui sont pour les amplicons 1, 2, 3 et 4 respectivement égales à 723, 471, 289 et 616 pb.

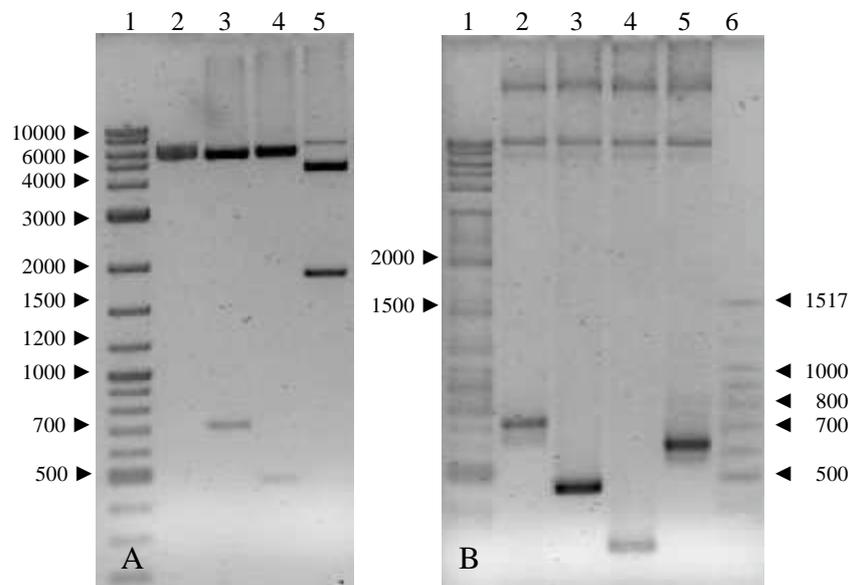


Figure 79 : Electrophorèse en gel d'agarose du plasmide pYES2/IgEst2 digéré ou amplifié (clone α).

Les acides nucléiques ont été séparés sur gels à 2 et 1 % (w/v). **A)** Vérification de la présence de l'insert *IgEst2* dans le clone α par digestion enzymatique : plasmide non digéré (piste 2), double digestion avec *EcoRI* et *XbaI* (piste 3), plasmide digéré avec *KpnI* (piste 4), avec *PvuI* (piste 5). **B)** Vérification de la présence de l'insert *IgEst2* dans le clone α par PCR : les amplicons 1, 2, 3 et 4 ont été obtenus avec les couples d'amorces *EcoRIEst2/Est2XbaI*, *Est2S/Est2AS*, *NEst2S/Est2XbaI* et *EcoRIEst2/NEst2AS* respectivement. Les marqueurs de taille utilisés correspondent aux marqueurs 2log et 100 pb.

6.3.3. Séquençage de l'insert *IgEst2* (clone α)

Le clone α a été séquençé afin de vérifier l'intégrité de l'insert *IgEst2*. Comme nous pouvons le voir en Annexe 15, la séquence de l'insert *IgEst2* est intègre et n'a pas subi de mutation.

6.3.4. Induction de l'expression d'IgEst2 dans *Saccharomyces cerevisiae*

L'expression du transcrit a été induite par la présence de galactose dans le milieu de culture SC-U destiné aux levures transformées. La présence du transcrit a été détectée par amplification avec le couple d'amorces *Est2S/Est2AS* à partir de cultures de 12 heures. Pour s'assurer de l'absence d'ADN contaminant dans les échantillons d'ARN des amplifications témoins ont été réalisées à partir d'ARN total. Des levures témoins transformées avec le plasmide non recombiné ont été testées en parallèle. Dans cette expérimentation seules les levures cultivées en présence du sucre inducteur (galactose) ont été analysées. Cette expérimentation permet de conclure sur la présence du transcrit et sur le bon fonctionnement transcriptionnel du système d'expression pYES2/*S. cerevisiae*.

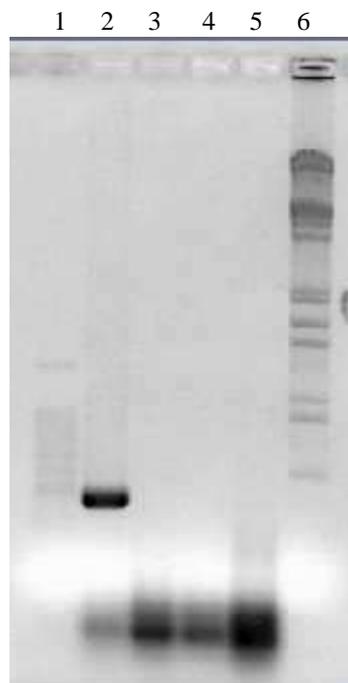


Figure 80 : Electrophorèse en gel d'agarose du transcrit *IgEst2* amplifié par PCR après induction de l'expression.

Les acides nucléiques ont été séparés sur gel à 1 % (w/v) d'agarose. Le marqueur de taille utilisé est le bactériophage λ digéré par les enzymes *EcoRI* et *HindIII* (piste 6). Les amplifications ont été réalisées avec le couple d'amorces *Est2S/Est2AS*. Piste 2 : PCR sur ADNc de levures transformées avec le plasmide recombiné pYES2/IgEst2 et cultivées en présence de galactose. Piste 3 : PCR sur ARN extrait de levures transformées avec le plasmide recombiné pYES2/IgEst2 et cultivées en présence de galactose. Piste 4 : PCR sur ADNc de levures transformées avec le plasmide pYES2 non recombiné. Piste 5 : PCR sur ARN extrait de levures transformées avec le plasmide pYES2 non recombiné.

6.3.5. Mesure de l'activité enzymatique

Les premières mesures d'activités enzymatiques réalisées n'ont pas permis de mettre en évidence une activité lipolytique significative.

Tests d'hydrolyse	Observations	Activité lipolytique	
		Levures transformées	Levures témoins
Emulsions gélosées d'huile d'olive, de tributyrine et de lécithine de soja	Absence de lyse	-	-
Palmitate de paranitrophényle	Formation de paranitrophénol	+	+
Lipides totaux d' <i>I. galbana</i> (24h, 40°C)	Apparition d'acides gras libres	+	+

Tableau 23 : Mesure de l'activité lipolytique des levures transformées avec le plasmide recombiné pYES2/IgEst2.

Pour détecter l'activité lipolytique des levures transformées avec le plasmide recombiné pYES2/IgEst2, différents tests d'hydrolyse ont été développés (Tableau 23). Au départ l'activité d'hydrolyse a été testée en cultivant la levure transformée sur des émulsions de substrats gélosés incluant soit de l'huile d'olive (trioléine) soit de la tributyrine soit un mélange de phospholipides, la lécithine de soja. Ces tests n'ont pas mis en évidence d'activités lipolytiques extracellulaires.

Ensuite des essais d'hydrolyse en milieu liquide ont été réalisés avec un ester d'acide gras à 16 carbones, le palmitate, couplé à une molécule chromogène, le paranytrophényle. La formation de paranitrophénol a pu être observée avec les levures transformées mais aussi avec les levures témoins. Cependant aucune cinétique n'a pu être tracée en raison de la formation simultanée d'un précipité faussant et déstabilisant les valeurs d'absorbance mesurées. Ce travail a été poursuivi par un test d'hydrolyse sur les lipides totaux d'*isochrysis galbana*. Pour cela des échantillons de biomasse de levures transformées et des levures témoins ont été broyées puis incubées 24 h à 40 °C en présence des lipides d'*Isochrysis galbana*. Les lipides extraits ont ensuite été analysés par CCM. Des tests d'hydrolyse ont été réalisés avec les surnageants de culture et avec les cellules de levures broyées. Une première observation peut

être faite par rapport au test d'hydrolyse réalisé avec les surnageants de culture. Nous n'observons ni l'apparition d'AG libres ni la disparition ou l'atténuation de spots lipidiques. Ces résultats confirment les résultats obtenus avec les tests d'hydrolyse sur substrat gélosé. Les levures transformées et témoins ne présentent pas d'activité lipolytique extracellulaires. Concernant les tests d'hydrolyse réalisés avec les cellules de levures, nous pouvons observer l'apparition d'AG libres aussi bien pour le test réalisé avec les levures transformées que pour les levures témoins. Par contre nous n'observons pas en parallèle la disparition ou l'atténuation significatives de spots lipidiques.

Ces expérimentations constituent des tests d'hydrolyse préliminaires qui devront être renouvelés avec des conditions définies avec précision.

7. Discussion générale

Le but du travail présenté était d'une part de confirmer la présence d'une activité lipolytique endogène chez la microalgue *Isochrysis galbana* et d'autre part d'isoler des gènes codant des enzymes homologues à des phospholipases, lipases ou estérases. Plusieurs gènes candidats ont été identifiés dans la base de séquences EST d'*I. galbana* et 3 gènes homologues à une lipase et deux estérases, que nous avons nommés *IgLip*, *IgEst1* et *IgEst2*, ont été isolés sur toute leur longueur, analysés et clonés. Ces résultats indiquent l'existence d'une homologue de lipase et de deux homologues d'estérases chez le modèle étudié ainsi que chez d'autres microalgues. Les gènes codant les estérases hypothétiques ont été exprimés dans un système hétérologue et les premières mesures d'activité lipolytique ont été entreprises.

7.1. Confirmation d'une activité lipolytique endogène chez *Isochrysis galbana*

L'étude de l'activité lipolytique endogène de la microalgue *I. galbana* a confirmé la présence d'activités glycolipasique et phospholipasique préalablement mises en évidence au laboratoire et révélé que ces activités sont associées aux membranes. L'action d'une phospholipase de type A₂ (PLA₂) avait été suggérée après avoir constaté que le DHA, majoritairement présent en position sn-2 des phospholipides, était libéré en plus grande proportion au cours de leur hydrolyse (Devos, 2005).

Ces différentes observations nous permettent d'énoncer l'existence probable chez *I. galbana* d'une phospholipase de type A₂ associée aux membranes. Les phospholipases A₂ constituent une vaste famille d'enzymes. Certaines PLA₂ présentent en effet la caractéristique d'être étroitement associées aux membranes tandis que d'autres sont soit cytosoliques soit sécrétées. Les PLA₂ intracellulaires ont été réparties en 2 groupes, les PLA₂ faiblement dépendantes du calcium (cPLA₂), cytoplasmiques, et les PLA₂ indépendantes du calcium (iPLA₂) (Ackermann et Dennis, 1995; Winstead *et al.*, 2000). Les iPLA₂ sont elles-mêmes divisées en 2 catégories : les iPLA₂ libres dans le cytoplasme (Six et Dennis, 2003) et les iPLA₂ associées aux membranes. Les iPLA₂ ont déjà été clonées et purifiées chez différentes espèces animales (Mancuso *et al.*, 2000; Tanaka *et al.*, 2000; Jenkins *et al.*, 2004). Plusieurs isoformes d'iPLA₂ ont été identifiées chez l'être humain, où elles se différencient par leur distribution tissulaire et leur localisation cellulaire. Un seul gène serait à l'origine des iPLA₂ humaines (Larsson

Forsell *et al.*, 1999). Le gène responsable serait transcrit sous différentes formes par un processus d'épissage alternatif (Larsson *et al.*, 1998).

La génomique comparative a révélé l'existence de gènes susceptibles de coder des iPLA₂ associées aux membranes chez les microalgues ayant fait l'objet d'un programme de séquençage à grande échelle. Chez l'algue rouge unicellulaire *Cyanidioschyzon merolae*, par exemple, 2 séquences codantes référencées CMR500C et CMT312C présentent des similitudes significatives avec de telles enzymes. La protéine indexée 23984 chez la diatomée *Thalassiosira pseudonana* est également une homologue d'iPLA₂ associée aux membranes.

7.2. Présence de 10 gènes susceptibles de coder des enzymes lipolytiques chez *Isochrysis galbana*

Après la détection d'une activité lipolytique endogène, l'identification de gènes constitue la seconde mise en évidence de la présence d'enzymes lipolytiques chez *I. galbana*. L'exploration de la banque EST d'*Isochrysis galbana* (Keeling, 2006) a révélé l'existence d'au moins 10 gènes d'enzymes lipolytiques différents. Ces gènes sont répertoriés dans la banque sous forme de courtes séquences. Les phases ouvertes de lecture sont incomplètes, tronquées à leurs extrémités 5' et 3'. La traduction des marqueurs montre des similitudes significatives avec des lipases, des estérases, des phospholipases et des lysophospholipases déjà identifiées et caractérisées chez d'autres organismes.

La diversité des enzymes lipolytiques présentes au sein de cette algue unicellulaire n'est pas surprenante. Les voies métaboliques lipidiques impliquent en effet l'action de nombreuses enzymes lipolytiques. De plus, sachant que les banques d'ADNc n'offrent qu'une vision réduite du réel potentiel génétique des organismes, l'existence de gènes d'enzymes lipolytiques supplémentaires n'est pas exclue.

Dans les conditions de culture utilisées pour le projet de séquençage partiel (Keeling, 2006), le niveau d'expression des gènes d'intérêt est faible si l'on se réfère au rapport du nombre d'EST par cluster. Sur les 12 clusters présentant des similitudes statistiquement significatives avec des enzymes lipolytiques, 8 ne dénombrent qu'un seul marqueur EST. Parmi les 4 autres clusters, 3 comptent 2 marqueurs EST et seul le cluster n° ISL00001076 en regroupe 3. Par comparaison, le cluster d'un gène constitutif fortement exprimé, comme celui codant pour l'ARNr 18S comptabilise 118 marqueurs EST. Les gènes qui nous intéressent ne sont pas

rare mais s'expriment probablement plus différemment dans certaines conditions sous l'effet de facteurs biotiques ou abiotiques.

7.3. Isolement de 3 gènes codant 1 lipase et 2 estérases hypothétiques chez *Isochrysis galbana*

Parmi les 12 gènes d'enzymes lipolytiques hypothétiques indexés, sous forme d'étiquettes, dans la banque EST d'*I. galbana*, 3 de ces gènes d'intérêt pour le laboratoire ont été isolés et entièrement séquencés au cours de cette étude. La première séquence, nommée *IgLip*, présente 20 % d'identité avec des lipases de champignons filamenteux. Les deux autres séquences identifiées, nommées *IgEst1* et *IgEst2*, montrent des similitudes avec des estérases, des phospholipases et des lysophospholipases du règne animal, végétal ou microbien. Les similarités observées lors des alignements de séquences sont faibles et les taux d'identité varient entre 20 et 30 %.

Ce faible taux de conservation des séquences protéiques est, en réalité, caractéristique des enzymes lipolytiques et plus généralement des enzymes de la superfamille des hydrolases à repliement α/β , pour lesquelles les structures sont beaucoup plus conservées que les séquences. En effet, seuls quelques résidus et une petite partie de la structure sont généralement responsables de l'activité catalytique des enzymes. Ces quelques acides aminés sont donc soumis à une forte pression de sélection et sont systématiquement retrouvés dans les séquences. Par contre, la nature des autres acides aminés n'est pas essentielle car ils sont impliqués dans le repliement global de la protéine qui a pour rôle de positionner correctement les résidus catalytiques dans l'espace. De nombreuses études de cristallisation des protéines ont démontré que des enzymes de séquences protéiques différentes pouvaient adopter une même conformation dans l'espace.

Malgré ce faible taux d'identité observé, les séquences obtenues chez *I. galbana* contiennent l'ensemble des sites catalytiques nécessaires à l'activité des enzymes lipolytiques : la triade catalytique Ser/Asp/His et le pentapeptide consensus Gly-X-Ser-X-Gly. A cela se rajoute la conservation de quelques résidus qui seraient potentiellement impliqués dans la fixation de l'enzyme au substrat.

Les séquences obtenues présentent quelques particularités susceptibles d'influencer la spécificité de substrat des enzymes hypothétiques nouvellement identifiées. Certains résidus

impliqués dans la fixation au substrat chez des enzymes de structure et de fonction connues ne sont pas conservés ce qui implique sans doute une modification de la conformation spatiale de la poche abritant le site actif. La séquence *IgLip* présente une zone d'insertion d'environ 60 résidus et qui éloigne les deux résidus Ser et Asp potentiellement impliqués dans le mécanisme catalytique.

7.4. Présence d'isoenzymes chez *Isochrysis galbana*

Les deux séquences *IgEst1* et *IgEst2* présentent une certaine similarité. Il est en effet ordinaire de rencontrer, au sein d'un même organisme, plusieurs lipases et estérases présentant un taux de similitude faible mais néanmoins significatif. Ces variabilités de séquences reflètent généralement des spécificités fonctionnelles ou des implications dans des voies de signalisation distinctes.

Les enzymes lipolytiques sont des enzymes appartenant à une famille multigénique. Certaines isoformes de lipase, par exemple, sont exprimées de façon constitutive dans l'organisme, d'autres sont spécifiquement induites lors de stress biotiques ou abiotiques, tandis que pour d'autres les profils d'expression de ces différents gènes sont souvent spécifiques d'une distribution tissulaire. Chez la drosophile *Drosophila melanogaster*, par exemple, les lipases homologues *DmLip1*, *DmLip2* et *DmLip3* constituent une famille de gènes qui s'expriment spécifiquement selon le tissu ou le stade de développement. Les gènes *DmLip1* et *DmLip3* sont transcrits dans différents tissus et à différents stades de développement, ce qui conduit à penser que les deux lipases auraient des fonctions différentes. Le gène *DmLip1* code pour une protéine similaire aux lipases digestives, s'exprime dans les ovaires, au stade embryonnaire et avec une autre taille de transcrit à tous les autres stades du développement. L'enzyme codée par le gène *DmLip3* s'apparente plutôt aux enzymes acides liposomales, elle n'est synthétisée que pendant la période larvaire. Le gène *DmLip2* semble non fonctionnel (Pistillo *et al.*, 1998).

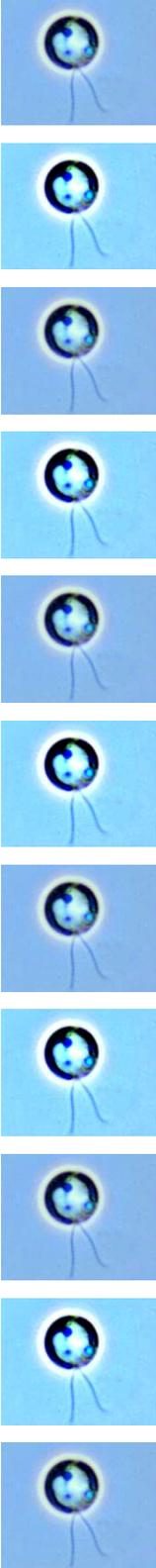
Le champignon pathogène pour l'homme, *Candida albicans*, possède une grande famille de gènes de lipases : 10 gènes de lipase ont été identifiés (*CaLip1-10*) dont l'expression varie au cours de la colonisation et de l'infection et selon le type de muqueuse infectée (Schofield *et al.*, 2005).

7.5. Présence de gènes avec et sans introns chez *Isochrysis galbana*

La présence de gènes avec et sans introns a été mise en évidence chez la microalgue *Isochrysis galbana*. Le gène *IgLip* est constitué de 2 introns de 173 pb et 121 pb. Le transcrit *IgLip* de 1374 pb résulte de l'assemblage de trois exons de 341, 823 et 210 pb. Le gène *IgEst2* ne comporte pas d'introns. Les tailles du gène et du transcrit sont donc équivalentes. Cette constatation atteste les informations publiées sur la structure des gènes de microalgues. On trouve, en effet chez les microalgues, aussi bien des gènes avec introns que des gènes sans introns. Ces derniers sont toutefois majoritaires. Par exemple, chez l'algue rouge unicellulaire *Cyanidioschyzon merolae*, 5331 gènes ont été répertoriés. Seuls 27 gènes, soit 0,5 % de la totalité des gènes de la microalgue, présentent des séquences introniques. Parmi les 27 gènes, 26 d'entre eux ne contiennent qu'un seul intron. Chez *Cyanidioschyzon merolae*, la longueur moyenne des gènes est égale à 1552 pb, celle des exons à 1540 pb et celle des introns à 248 pb (Matsuzaki, 2004).

Cette caractéristique semble commune à différentes lignées de microalgues, puisque la recherche d'homologues dans les bases de données génomiques existantes a révélé l'existence de gènes d'enzymes lipolytiques avec et sans introns.

CONCLUSION ET PERSPECTIVES



Ces travaux de thèse ont permis l'isolement de 3 ADNc pleine-longueur par amplification rapide de leurs extrémités 3' et 5'. Les 3 séquences d'ADN se sont révélées complémentaires de 3 ARN messagers possédant chacun un segment polyadénylé à leur extrémité 3' et une longue phase ouverte de lecture avec les signaux de traduction adéquats c'est-à-dire un codon d'initiation et un codon de terminaison. De part et d'autre de la séquence codante, ont été observées des régions non traduites situées aux extrémités 3' et 5', nécessaires à la régulation de la traduction de l'ARNm. Les 3 gènes codent des protéines de 457, 252 et 208 acides aminés, de poids moléculaires théoriques respectivement équivalents à 49,06 kDa, 26,92 kDa et 22,24 kDa.

Les fonctions associées aux nouvelles séquences ont été prédites en recherchant des séquences homologues dans les bases de données et les indices d'appartenance à des familles de protéines. Les 3 protéines identifiées chez *I. galbana* se classent dans la superfamille des hydrolases à repliement α/β . Elles possèdent la triade catalytique Ser/Asp/His et le pentapeptide consensus Gly-X-Ser-X-Gly, deux signatures hautement conservées chez les hydrolases α/β procaryotiques et eucaryotiques. La première séquence, nommée *IgLip*, présente 20 % d'identité avec des lipases de champignons filamenteux. Les deux autres séquences identifiées, nommées *IgEst1* et *IgEst2*, montrent des similitudes avec des estérases, des phospholipases et des lysophospholipases du règne animal, végétal ou microbien. Les similarités observées lors des alignements de séquences sont faibles et les taux d'identité varient entre 20 et 30 %. Ce faible taux de conservation des séquences protéiques est en fait communément observé dans la famille des enzymes lipolytiques pour lesquelles les structures sont beaucoup plus conservées que les séquences.

A la fin de cette étude, une démarche expérimentale a été entreprise pour confirmer les fonctions lipolytiques prédites. Pour cela, les séquences codantes *IgEst1* et *IgEst2* ont été clonées puis transférées dans un système d'expression hétérologue de levure. Les tests de mesure de l'activité lipolytique des protéines recombinantes n'ont pour l'instant pas permis de détecter une activité enzymatique significative agréant la fonction supposée des protéines.

Cette étude a donc permis d'isoler, d'analyser et de cloner 3 gènes codant une lipase et 2 estérases hypothétiques chez la microalgue *Isochrysis galbana* et d'identifier des séquences homologues dans les bases de données génomiques de microalgues. La prédiction fonctionnelle par analogie de séquences est fréquemment réalisée pour l'étude de nouvelles

lipases et estérases dans le monde animal, végétal et fongique. Une telle analyse n'avait pas encore été développée pour l'étude des enzymes lipolytiques de microalgues.



Dans la poursuite de ce travail, le gène *IgLip*, pour lequel la délétion d'une thymine en début de séquence a été détectée, devra être cloné à nouveau.

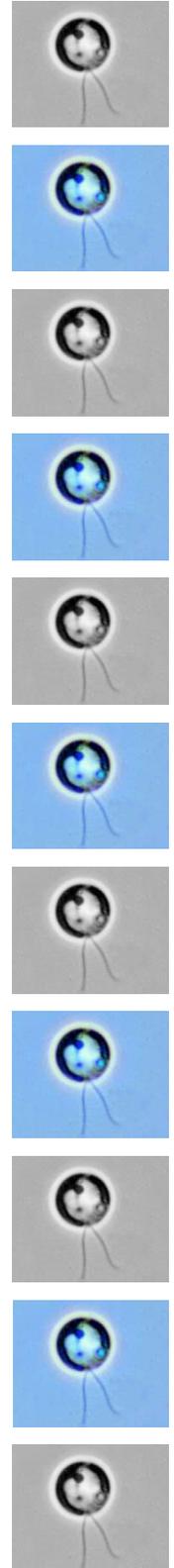
L'étude de l'expression des séquences codantes *IgEst1* et *IgEst2* pourra, quant à elle, être poursuivie. Le bon fonctionnement de la transcription des inserts, avec le système d'expression pYES2/*Saccharomyces cerevisiae*, ayant été démontré au cours de cette étude, il faudra donc vérifier, dans un premier temps, l'efficacité du système en matière de traduction des transcrits en utilisant, par exemple, des méthodes de séparation électrophorétique des protéines. Ensuite, la détection d'une activité lipolytique des levures modifiées pourra être entreprise à nouveau. Une gamme plus étendue de substrats et des techniques de mesure d'activité plus sensibles (fluorimétrie) ou plus spécifiques (zymogramme) seront à tester. Enfin, si les protéines recombinantes se révélaient trop faiblement synthétisées ou non fonctionnelles dans ce système, l'utilisation de vecteurs d'expression et de cellules hôtes différents seraient à prévoir. La levure *Pichia pastoris* et le système baculovirus / cellule d'insecte sont actuellement très utilisés.



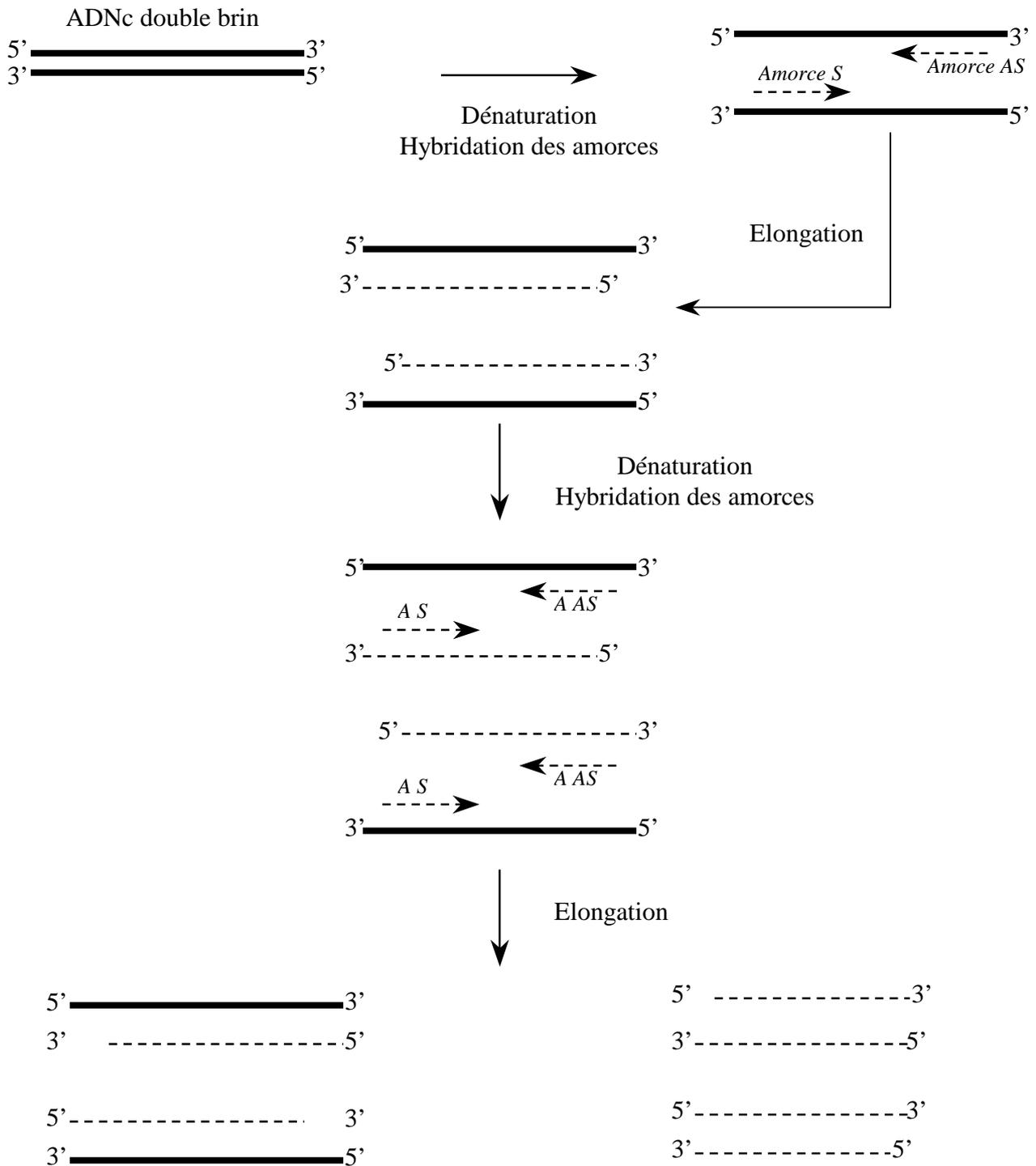
Lorsqu'un système fonctionnel de surexpression des gènes aura été trouvé, la purification et la caractérisation biochimique des enzymes recombinantes pourront être réalisées. Par la suite, le laboratoire développera des études portant sur la régulation de l'activité transcriptionnelle des gènes et les facteurs d'activation des voies de lipolyse afin d'analyser et de comprendre le rôle joué par les enzymes lipolytiques microalgales dans la mobilisation des triglycérides de réserve ou le remodelage des phospholipides membranaires.

A terme, le laboratoire souhaite appliquer cette même démarche aux autres gènes d'intérêts identifiés grâce aux marqueurs de la banques EST d'*Isochrysis galbana*.

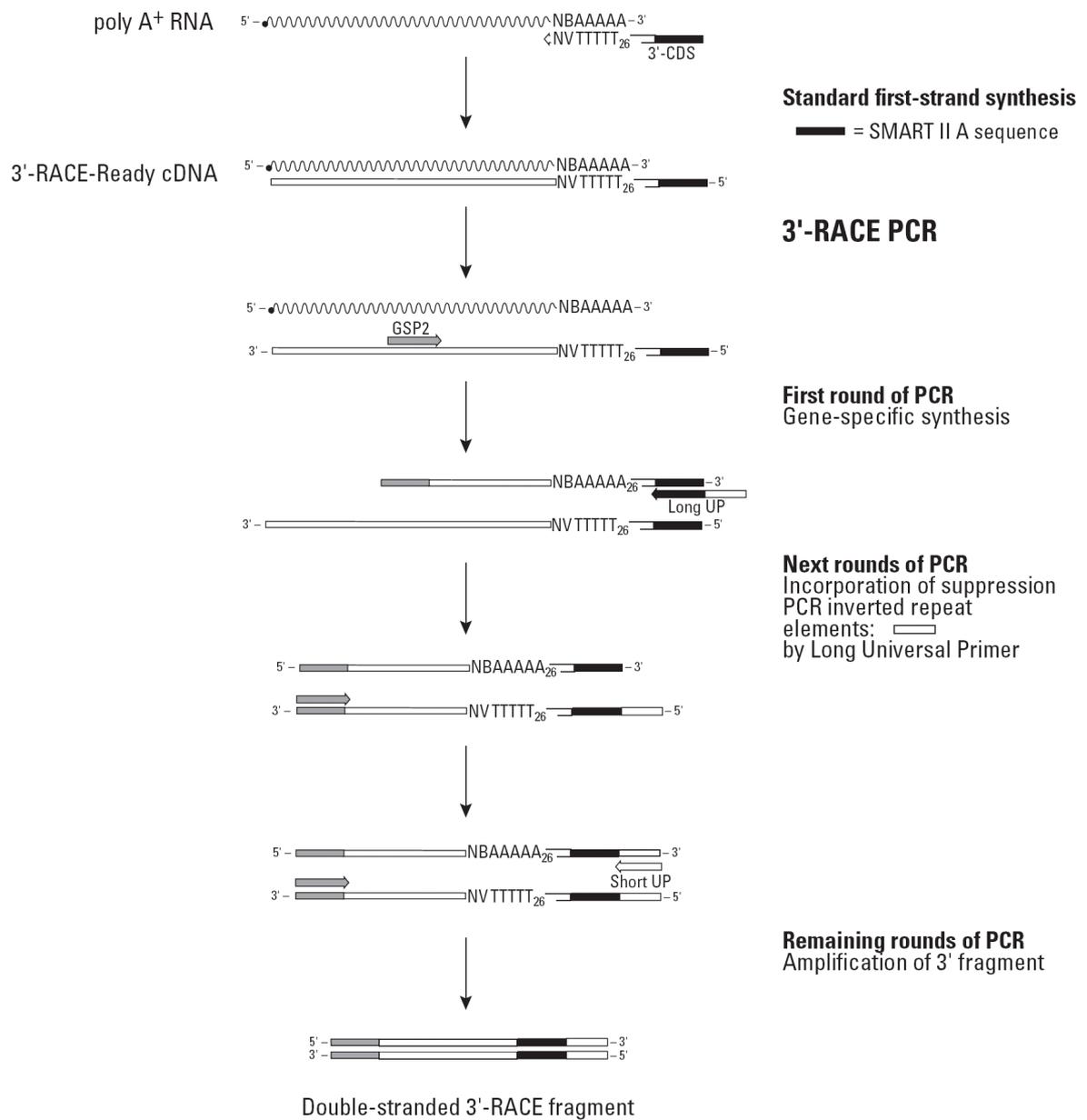
ANNEXES



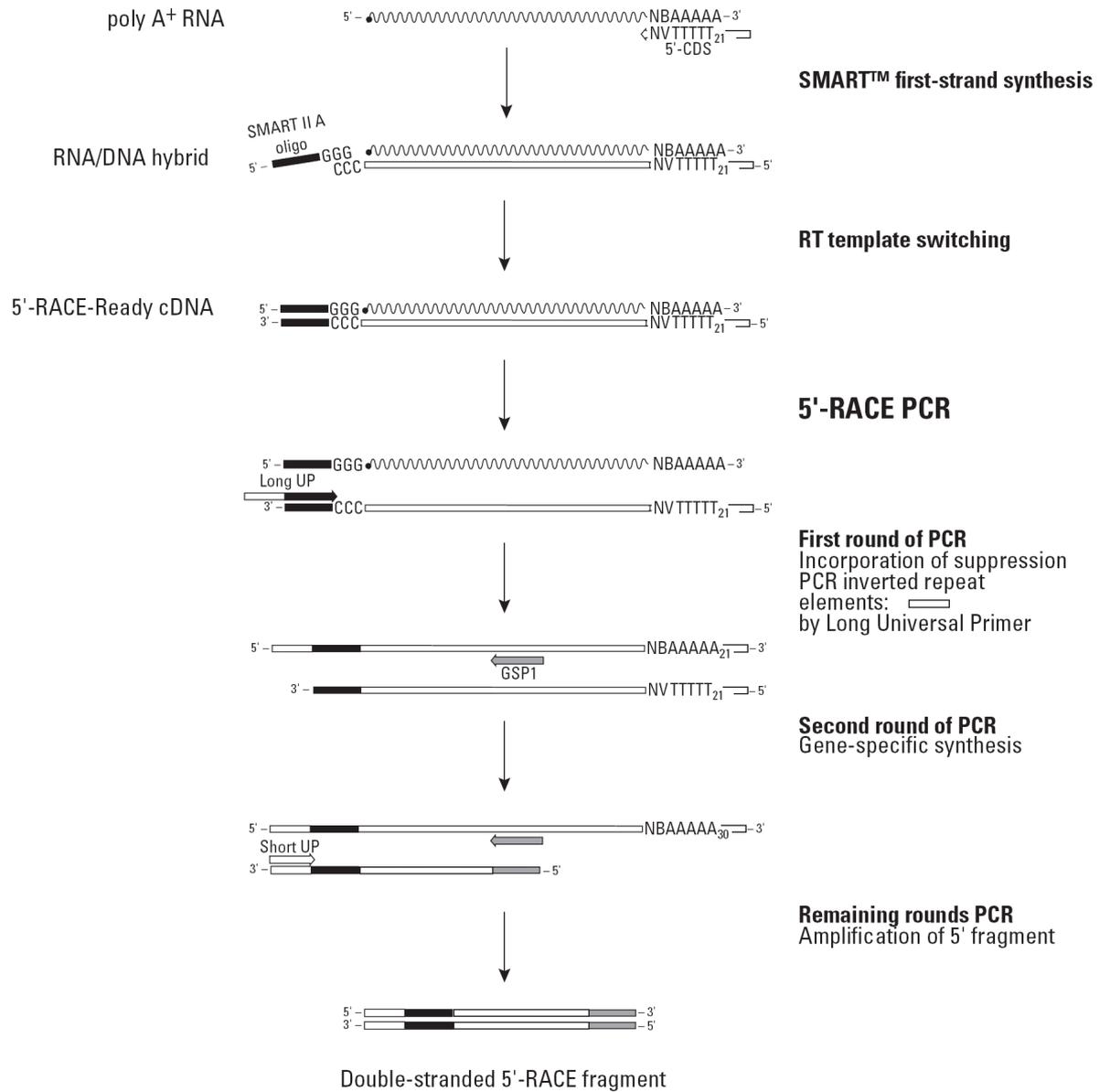
Annexe 1 : Principe de la réaction de polymérisation en chaîne (PCR)



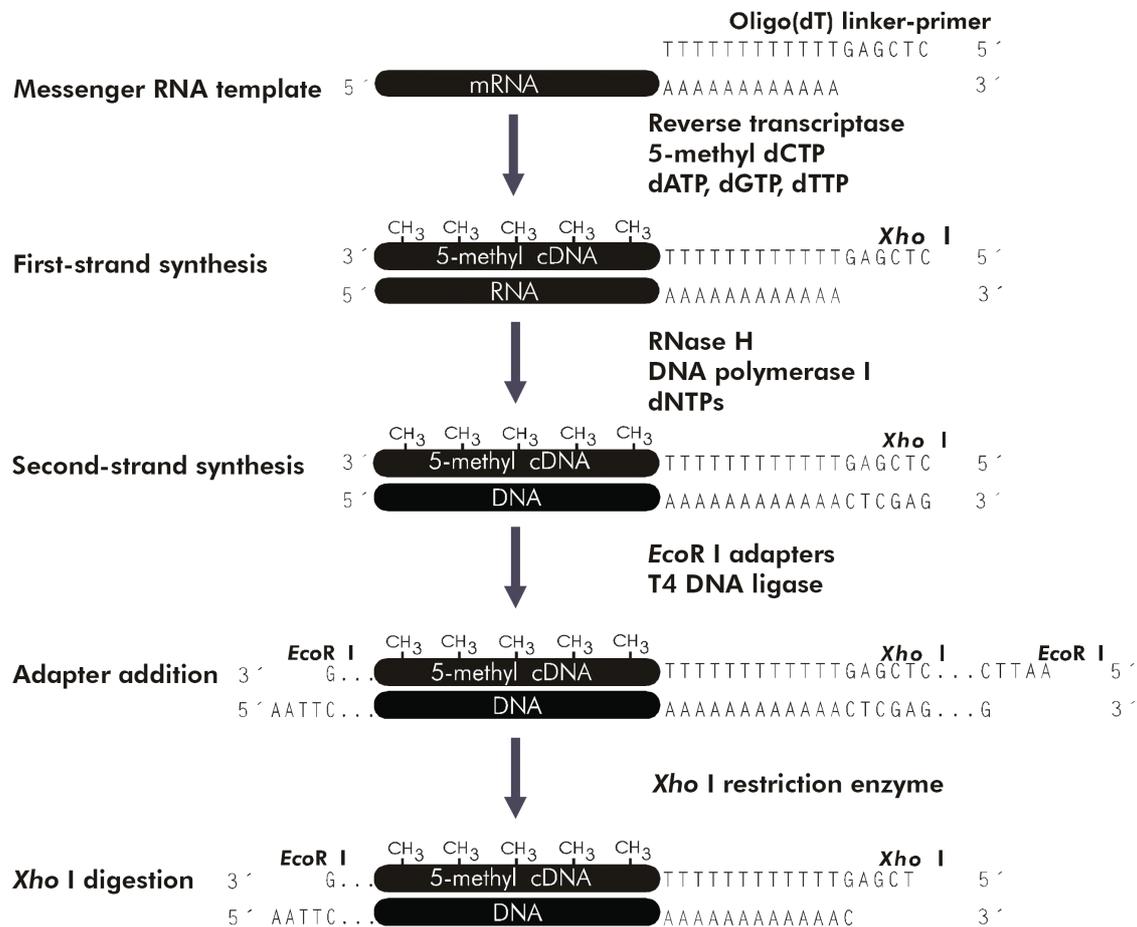
Annexe 2 : Principe de la réaction 3'-RACE (Clontech)



Annexe 3 : Principe de la réaction 5'-RACE (Clontech)



Annexe 4 : Synthèse des ADNc pour la construction de la banque d'ADNc (Stratagène)



Annexe 5 : Liste des bases de données et des outils informatiques

Bases de données

Blocks	http://blocks.fhcr.org/
DDJB	http://www.ddbj.nig.ac.jp/index-e.html
EMBL	http://www.ebi.ac.uk
GenBank	http://www.ncbi.nlm.nih.gov/
INSDC	http://www.insdc.org/page.php?page=home
LED	http://www.led.uni-stuttgart.de/
MELDB	http://www.gem.re.kr/melddb/
PROSITE	http://www.expasy.ch/prosite/
PDB	http://www.rcsb.org/pdb/home/home.do
Swiss-Prot	http://expasy.org/sprot/
TbestDB	http://tbestdb.bcm.umontreal.ca/searches/welcome.php
<i>I. galbana</i>	http://tbestdb.bcm.umontreal.ca/searches/organism.php?orgID=IS
<i>P. lutheri</i>	http://tbestdb.bcm.umontreal.ca/searches/organism.php?orgID=PL
<i>C. merolae</i>	http://merolae.biol.s.u-tokyo.ac.jp/
<i>T. pseudonana</i>	http://genome.jgi-psf.org/Thaps3/Thaps3.home.html
<i>P. tricorutum</i>	http://genome.jgi-psf.org/Phatr2/Phatr2.home.html
<i>C. reinhardtii</i>	http://genome.jgi-psf.org/Chlre3/Chlre3.home.html
<i>O. tauri</i>	http://genome.jgi-psf.org/Ostta4/Ostta4.home.html
<i>O. lucimarinus</i>	http://genome.jgi-psf.org/Ost9901_3/Ost9901_3.home.html

Outils bioinformatiques

BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
Block Searcher	http://bioinformatics.weizmann.ac.il/blocks/blocks_search.html
ClustalW	http://www.ebi.ac.uk/Tools/clustalw/index.html
ESPrpt	http://esprpt.ibcp.fr/ESPrpt/cgi-bin/ESPrpt.cgi
Gblocks	http://molevol.cmima.csic.es/castresana/Gblocks_server.html
GeneMark	http://opal.biology.gatech.edu/GeneMark/
Geno3D	http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html
MUSCLE	http://www.ebi.ac.uk/muscle/
pDRAW32	http://www.acaclone.com
Protein Explorer	http://librairiedemolecules.education.fr/outils/pevf/pe1_982/protexpl/frntdoor.htm#dl
SignalP 3.0	http://www.cbs.dtu.dk/services/SignalP/
SMS	http://www.ualberta.ca/~stothard/javascript/index.html
Spidey	http://www.ncbi.nlm.nih.gov/spidey/
TOPPRED	http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html
Webin	http://www.ebi.ac.uk/Submissions/

Annexe 7 : Clusters d'enzymes lipolytiques indexés dans la Base EST d'*I. galbana* (Keeling, 2006)

© 2007 TBestDB Group
tbestdb@bch.umontreal.ca

>Cluster Id : ISL00008005, AutoFACT Annotation : Acyl protein thioesterase 1 related cluster

5'-GCACACCGCGACGGTCATCGGCCCATCCACGGGCTGGGCGACAGCAACATGGGCTGGGCCGACGTGGCGATGCAGCTGCAGAGTGTATGCCGTATTGCAAGTTTATCCTCCCCAATGCGCCCCGTCCGTCTGTACCTCAACGGCGGAATGTTCGATGCCAGTTGGTATGACATCACGAGTTTGGATAAGCGCGAGTCCAGCCGTGCACGGGCATCGAAGAGTCTCGCCAGGCGATGTTCGACCTCATCTCCGCTGAGGTTGCGTCCGGCATCCCCCCTCCCGCATGCCATCGCCGGCTTCTCCAGGGAGGCGCCGTCGCCCTTTACACAGGGGTCAGTATTACACACACCTGGCCGGCGTGTATGCTTGTCCGGTACTTGGCGGGGAGGAGCGTTTCATACTGGCGCCGAGGCTGTGAACACTCCCGTCCGCTACTTTACGGCTCGGATGATCAAACGGTCAAATCAAATGGGCCCGCGCTCTCAGGCCACCTGCGCGAGCTGGGCATCCGCACCTACGAGCTAAAAGAGTACTCGCCCTCGGCCACAGCGCTCGCA-3'

>Cluster Id : ISL00006775, AutoFACT Annotation : Asparaginase related cluster

5'-GGGTGCACAGCGACCTGGACGCGAGGGTGACCGTGTATGCGGCTCTCACCAGGCTTCGACGACTCTGCCATCGCCGCATGATCGACAGCGCCGAGATGCGCGGGCTGGTGTGCTGTATGGCACAGGTAACGGCCCGTGCACAAAGACCGACTTCATACGGGCCATCGATATGGCTGTGAAGAAAGACATCGCGTGTGTCGAGTACGCGAGTGTATATCGGGGCGGTTAGACTCGAGACTTACGAGGTGGGCCGGGACTCCTCGCCCTCGGCGTTGTCTCGGCTGGAGATATGACCACGGAAGCTTGTGTGACCAAGCTCGCCTATTGAGCGGGCGGGTTTGGCGGGCGCTCGCTCGCAGAGGCCATGCGGACCAGCCTCGCGGGGAGCTTACCGAGCTGGTCCGGG-3'

>Cluster Id : ISL00007519 AutoFACT Annotation : Carboxylesterase related cluster

5'-GGGCTGGCATTCAAACGAGCGCTCCGGCCTCGCAGGTATTGCTCGGACGCATCCAGAATCGCCGCTTCTCCAAGTCGCAATCCTCGCTCTCGGACGGCAAAGAGGACAGCGAGACCGATTGGATGACGACTGCAGCGATGCCTCGATCGACGGATATCGGGGCCCCCTCAGGTCGGATTCCGGCAGCAGCGTCCGAGGATCCCGGTATCGTGATCATCGAGCCCTCTGGTCCACACACGACACCGCCATTCTATTGCACGGCATGTACTGCTCTCCGGAGAGCTCCGATACATTTGTCGGTCTACCTGCTGTCAAATCCTTTGGCCTGTGCGCAGGGATTAATAATGTCTTCCCGATGCGCCGCTCGGACCATCAGCTGGCCGACAGGCCCCGAGGCAAACGTTGCCTCCTGGTACAATTACTACACGCGCCGGGATGGCGAGCTCGAGCATGACGCTTGAATGAGGCGCACCTGGCTTCGCAAACCCGGCGGATCCACTCCATTGTGCGAGCGCGAGGCTGCGCTTCTTTGGGGCGACGCGCGCCGCATCATGCTCGGTGGGAGCTCACAAAGGGGACAGTCTCCATGCTACGATCAATTACCATCGCCCTCTCGGCGACTGCTTTGTCTTCGCTCATGCTCATCGATTCCGGTACGTTTCCGAGAGACAAGCGCTCCCCCGCCCGGCACACCCGCTTTCGTCTTCGCC-3'

>Cluster Id : ISL00003782, AutoFACT Annotation : Cluster related to UPI0000585F8E
PREDICTED: similar to 1-O-acylceramide synthase precursor (ACS)

5'-TGGGACATGGTGGTGCAGCTCGAGAGGGCCGGTACCAGCGGGGCAACAACCTATTCCGGCGTCCGCTACGACTGGCGGCTGCCGCGACCGAGAAAGCTGTGCGCCGACTTGGCGACGGTACTGCACCACATCACAAACGCGACGGGGCACCGGAAGGCGCTGCTGGTTCGGGCACTCGCTCGGCAACCTGCAGTGTCCACTGCCTCAACAACGCTTTTGGCCACGAGACGCTCTCCAAGATCAATCGCTCATGGCATTGCGGCGCCGCTCGCTGGCTTCCCCAGGTGGTGCCTGTTCTGTTCTCGGTGCCGAGATGGTATCCGCTCATCATCTCCGACCGGAGACCGGAGACTTTGCCCGCAGATGCCGGCCGCTACATGCTGCTGCCGGACGAGCGCAT-3'

>Cluster Id : ISL00000940, AutoFACT Annotation : Esterase/lipase/thioesterase-like protein related cluster

5'-cccctcgtcttcccccgctgctcgaagccactcgcccctcgcgcccgtgcccctaccggttgctggcaccgcccgcacctTTGCGGCTTCAACGCGACGCCGTCTTGCCGAAACCTTTCAGCTTCTCAAGCGCGCCAGACAGTTCTCTTCCCCGGAGGAGCGCGAGGTTTCAAGCGAAGGGTGAGGACTACACCCTTCTTGCCGGAGGAGCCGACTTGGTCCGCGTGGCCGCACGGCTCAATGCGACGCTCATCCCCTTCTCCGGCTGGGCGGCGACGAGTCGATGGCGATACTCATGGACTCACAGGAGCTGCTCGACGCGCCGCGCCGGCGACTTTTTCGCTCCCGCATCGCCGGCTGCCTAGCTTCGTCGACGGGACGCTTTCGTGCGCCCGTCCATCACGCCGAGCGGCACTATTCTCTTCGCCAGCCAATCTCCACGCGGACATCGACCCGGCCGACCGCCGAGCGTGGCAGGAGGCTACGGCCGGCTGCGCGCGGGTGGGGAGGGGTCGGACGGCTCCAGCGTGTGGGGACA GGGATCCGTATCGCGACTGCTGCCCCGACCCGCcTACGAGGcgccacCGGGCTCCGGCCCCCGCCTGCCGAG-3'

>Cluster Id : ISL00002977, AutoFACT Annotation : Esterase/lipase/thioesterase-like protein related cluster

5'-AAGAGGGGATGGCCGCCATCCTCATCGTGGGAATCATCAGTTGTACGGGTTTCAGGACCTGCCCTCCTGGTTCGAGGAGGTTACCGGTCCACGGGTGTGCAGTGTGCTCCCTGGCGCACCCGGTTCATCTTCTCGGGCCGCGCCGCCCGGACACTCAGGAGGCCCTAGCGCGGATGAGGATGTAAACGGCGGGGGCCACGGGTTGAGCCAGCCAGTCGGGCCGCGGGGGGGCCAGGGCAGCTTCGGCGCAATAGACTATTCCGGCTTCGGCGCGGTCCCTGTCTCGCCGCGCGCTGTTCAAGCTGCTCAAGCGGGGGAGCCGACGCTGCTTACCCGGCGCGGTGCGCGAGGCTTCAAGTCGACTAAGAAGGGGGAGGCCTACAAGCTCTTTTGGCCCCACGGACCGATTCCGGTGACTTCGCCCC-3'

>Cluster Id : ISL00001844, AutoFACT Annotation : Lipase, putative related cluster
5'-GCTCCGGCTGGGTGCACCATGGCTGGGGTAGTGGCTACGACATGCTTCGCGCGCGCTCATCACGGCTGTGCTGCTAGCG
CTTGAGGCTCGGCCGACGTACCGCTCGTCTGTCACCGGCACTCTGCCGGCGGGGCACTGGCCTCCATCTTTGTGGCCGAGCT
CGCCACCGTCCGAGCAGCTCAATCATCTCGGTAGCCAGCGAATCGTCCGCGTACCGCGCCGAGGCCGGCTCTGACTCCAAAG
GCGAAGCTGACACGAAGGCTGCCGCGGCTGATGAGTCCGCGCCGCTCCAGCTCCATCGCTGGGAGAGTGTTCGGAGGTT
GCGCCGCTGCTCAGGCGACCGCTCGCGCCGCGCGCTGTACATCACCTTCGCGCGCCGATGGCCGGGAACCGGAGTTTGC
GGCCGCCATCGATTCCGGCTGGCCGCGCAGGGCAACGTGATGCTACGTGTGGTCAATGCACGCGATCCGCGCGTGCCTGCAC
CCAACACGCTGAATGCGGGTGGCAGCTCCAGATCCCC-3'

>Cluster Id : ISL00002069, AutoFACT Annotation : Lysophospholipase homolog related
cluster
5'-CTGCCGAACCTCCGCTGCCCTTCTTCATCGCCACGGCGGCGCCGACACCGTGTGCTGCCGGCCGCTCGGCGCTGCT
GTACAGCGCAGCTCCACACCCGAAGAGAGCCGGGCGCGCAAGGTGTATGAAGCGGGCTTCATGAGATGCTCAACTCGGAGC
AGCGCGAGGCCGCTCATGCAAGACATGGTCCGGTGGATACGGTCCAGGCTCTGAAACAAAAACAACACTTCGAGCACCAGCCGT
GAAAAAAAACACTCGGGGAGCGGTCTGCACGGGTGGATCCGGACCCCGGGCTCCAAAAACAAGTCCGAG-3'

>Cluster Id : ISL00003216, AutoFACT Annotation : Phospholipase associated protein
related cluster
5'-CAGCCTGCCGTACGTGCGAACGCCGGTGGCCCGCCGAGCGTCCGCTTCACCCGGTTGCAGCGTCTCGTGGCCATGGA
GCGCCGTTTCGAGGCAGAGCGGCAGGCTCTGCTGGATCGACTCGCTGCGGAGGCGCACACAGCCGCGCGCCGGTGTCTCAA
AGCTGAAACAGCTGCGCGAGGTGCTGCTGGCTGCGCTTCCGCGCCCGACGAGAATGGTCTCTGCCGTCTACACCACGCGGTG
CTCGATTGCGATGCGGCGCGGGTCCGGTCCGCTGGTACGCGCTGGGGCTCAGACCGATGTCGCTGCGGGCGAGCCTGGTGTCT
CCCGTCCATCTGACCATTCTGGCGGGCGACAGGCGCGGGGCTGCCTGCAGGAGCTTCGACGGCGCGCCGATGTGCGACG
CCCCGACGCGAATGGCGTGACGGCGCTTCATGCAGCCTGCAACCTGAACATGCCGCGGCTCGCCTCGATCCTTCTTGGCGCT
GGCGCGGTTCCCGACGCTCCGACGCCGCTTCGGAGTGTCCCTCTTcagctgGCCGGCTGGGCAAACACTCGGCTGAGGCGGCcga
gCGATGCTGGGGGAGGTG-3'

>Cluster Id : ISL00001076, AutoFACT Annotation : Phospholipase/Carboxylesterase
related cluster
5'-aattctgcagatattggaggcttctcgtTGGTTCGAGGGTATTCTTCGGCCCCAGAGCAGATTGCGCTCGACAAGACCG
TCGGTTCGTCGTCAGATGCTGTACTTGCATTTATTGAGGAAGCGCCTGCCAAGTTTGGGATCGATCCGAAGCGAGTCTTCCTG
CTGGGATTTTCGCGAGGGCGCCACACTCGTTCGACGACCCCTCTTGTCCAGATGGTTCGCGCCCCGGGCTCATCGCCGGCGGGCT
AACACTGTCTGGCCGGCTGTTTCCAGCTAATGACGAGTGGTACCCACTTTATGCGCGTCTCGCCGACCCGACCGCAGCTGA
GGATTTGTGCCGCTTTGCTCACACGGCGCCCAAGACTCGGTTACCCCGGTAGCTATCGGGCGCTCGAATGAGCGGCTGTTT
CGAGACTGGGCACCGGGTGGGACTTAATATATAAGGAGGACCCTGTTGCAGGCCATGAGATATCGAAGCCGTGGCTGCTTC
AATGTCCAATGGTTCGTCGGTATGTCTCCGGAGCATCAAAGGGTGTATCAGTAGGAGTTCAGCGGCGAGCAgCACCCGTCC
ATAGAGCAGGCTGTGAAGCCAACCTTTGTGATGACGCGCAAGATGCGTGGGCAAGCGTGCAGTCCACGCGCCCTCAAAC
ATTCGAACGCTCGTCTCC-3'

>Cluster Id : ISL00003195, AutoFACT Annotation : Plcd1 phospholipase C, delta 1
[EC:3.1.4.11] [KO:K05857]
5'-cggcgatgcaacagtccatctgctcgggctgcccctggccacgcccggcggcatttcggctcggCGCACGgCCGCGATCGG
GACTCGGGATCATCGGGTGTGCCATCAACGCACACCACAGTCCGCGCCCGGGCGGCGCAACGCCGAGGCGGAGGCCGAGGC
GCGCTGGCCTTCGCGCGCACGCTCACGGTCCACGGCAGCTTCGATCCCGCGGACGAGCTGTGGGCGTGGTGGCCG
ACTCTGAGCTCAACGGCGCGGTGCACCCGCGGGAGAGCCGGTGTGGCAGGACATGAAGCAGCCGCTGGCGCACTACCTCATC
TTCTCCAGCCACAACACTTACCTCATCGGCGAGCAGGTGATGGGCGTGTGATGGCCACGGCCTACTCGCGCGCATCGAGAT
GGGCTGCCGGTGCCTCGAAATAGACTGTTGGGATGGGGGTGGGCGGACCACTACGAGCCCATAGTACTCA-3'

>Cluster Id : ISL00006895, AutoFACT Annotation : [R] COG0400 Predicted esterase
5'-TGGCAAGGGATGCGTGAGGGCGAAAATGAGAAGGGACCTCTCGAGTTTGGAGCCGATGGAGAGCCCGTGGACCGCTGCAGT
GATTTGGCTCCACGGCTTTGGCGATGAACCCGAGGCATGGGCCGAATTCATGGCGGAGATAGGCGAGCAAGGCCGCTCTGGG
CCTGGGTCTTCCCTCGGGCGAAGGTACTGCCGACGCTTGCTACGGTGGGCGAGGAGCTGACCGCATGGGCGAGTTTCTGGAC
GCTGAGGTATCCGCCCGGGTGGCGCTGATTACGACTCGAGCGCAACAACTTCGCCAAGACAGTGCCTCTCATCATCGCAG
CGTTGATGCGGTACGTGCAG-3'

>Cluster Id : ISL00006799, AutoFACT Annotation : trEST Lipase-like protein related
cluster
5'-GGAATTCTGCAGATAGGAATTGAAGAACGAGGAGGGCGAGGCGAGCTCTTCAATGCACGGTTAGGATCGGGATTGGCGCA
GCGGCCGCGTGTTCGCTGTTGGTTTCATTTATTTCTGTTGGTACAATGATCCTGATCGATGGAGGCAATGAGGGTGCAGCAA
CAGTGGCGAGGCGGTGCGAGTCCAGGCGCGGCGAGCTTGGAGCGGCGCGAGAGGTGCAGATGACATGGCAGGAGGTCCACAT
CCGTCCCATGCAATGTGTCACGTGGTGTATTGGGGTAACCGCTTTTTCGGTGTGGTCTTGGTATTGGGAGCCTCGGGCACGCCGTG
CGTGGAGCCCGTCTTCCGTTGCATGGGTGTACATTGAGCTGACATCGCTGTGGATCGGATGGATCGCCAGCCAGATCCGT
GGATCTCCGGAGGCAATGGCTCCAAATGCAATGCACAGGAATAGCTGGAGATCGCACCCGATGGAGACCCCGGATCGCTGGAT
CGCACGCTGGACCATCTGGTGAAGAAGGTTGAGCTCCCGTTATT-3'

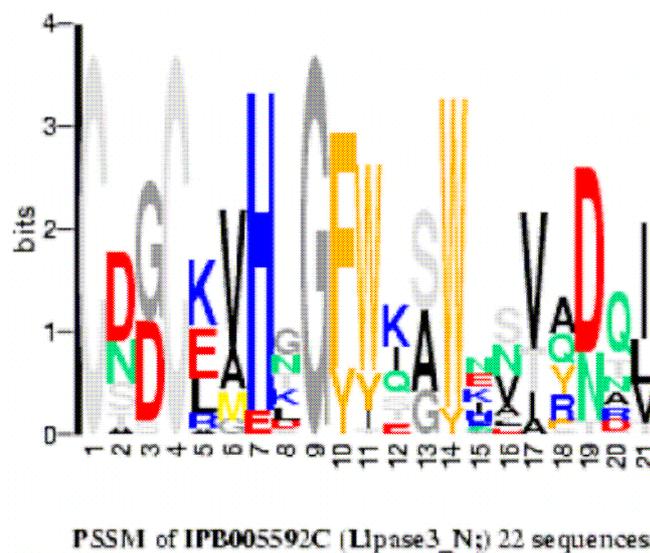
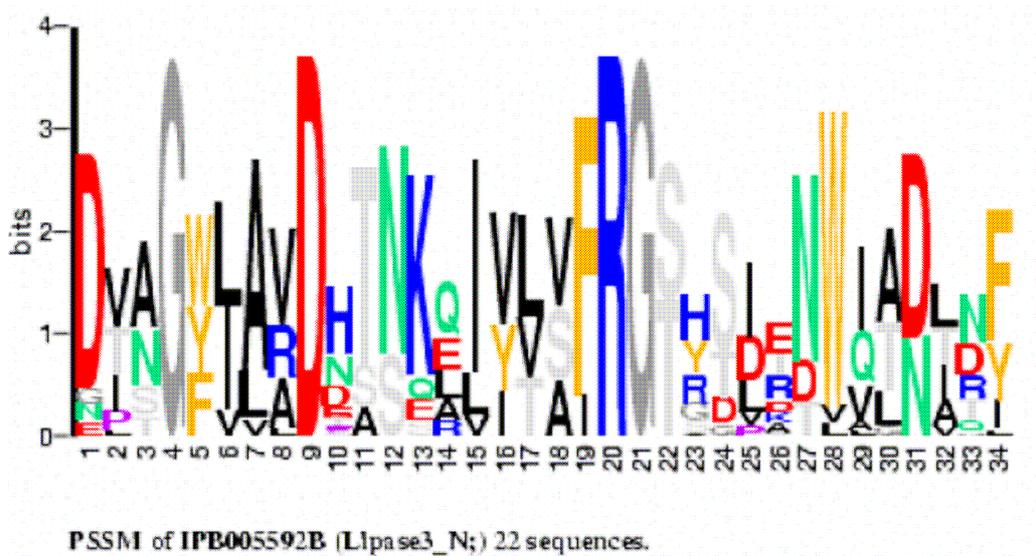
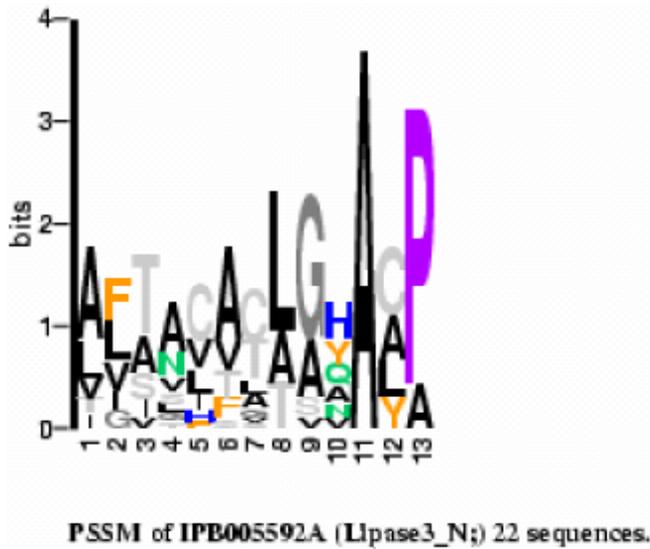
>Cluster Id : ISL00003084, AutoFACT Annotation : trEST Putative GDSL-motif
lipase/hydrolase protein related cluster

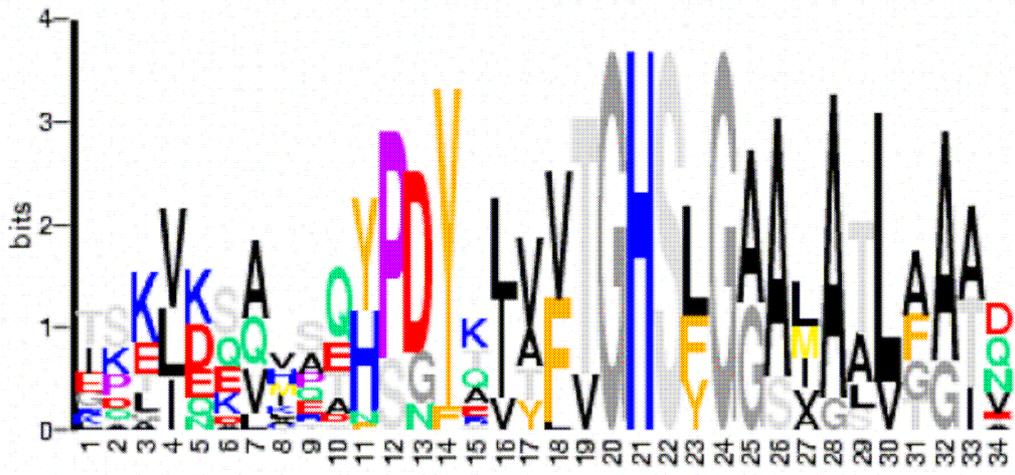
5'-ATCGCTCTCCATCGAGGGACTCGCACTGCTGTGGCCCCCTCGACTCGGCCAGCGCCACCCTGTTTGTAGCGGGCGGCTTT
GGTGCGGCCATTCCCCTCCAGTGCCCGTGCCTCTTCCACTGCAGGCAGCCTAGTGGACAGCACCCCTTCCACCGCCACCGC
TAGTGTCTCCAATATTGGGGCGGGCTGCCGGTTGGGCTCAAGAGTTGGCACGTGCGATCCTCCCTCGAGACCTGGGACAGC-
3'

>Cluster Id : ISL00002709, AutoFACT Annotation : trEST Putative GDSL-motif
lipase/hydrolase related cluster

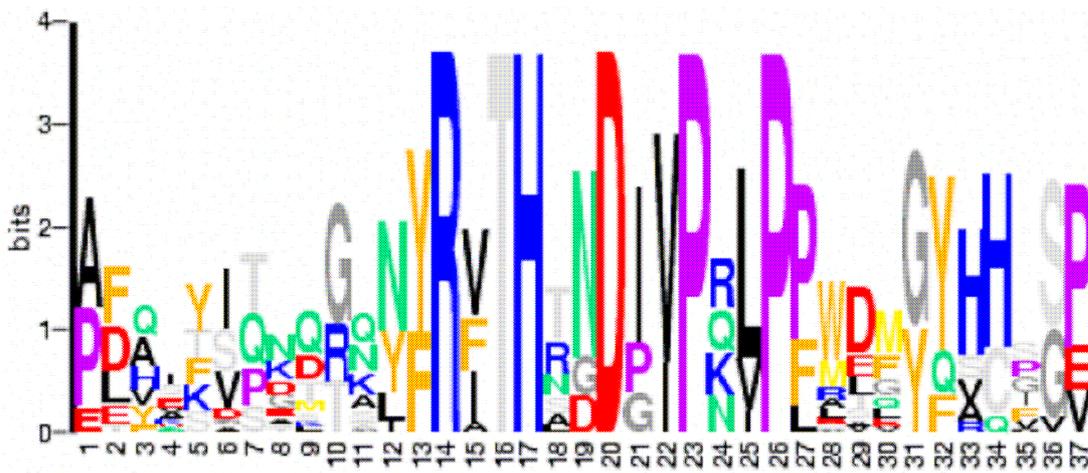
5'-ctggcagctgcbggttgcggggccgtcaagttccacgccaagggccgcaactctttaatgcbgctgtgccaactgcbgcbgcbg
TGTGCACGGGTGACGGGGCGACGCCGGTGCACATCATAGGTGAGCGAGGGCGTATGCGACAGCGATAGGTTGGAGCTGGTGC
GGAGCGATTGATGGGGCATTGCGCACATCTGGAGAACAGTGGATGCCGAAAATGAGTAGCCCTGAGCGACCCGGCTCGACCTC
TCTCTGACTGTAGGCATTCCAAAGGAGGGCTTCGAGATCACAGAGGGCGAGGACAAGGTGAAGTGGGCAATGGGTGCGGGCAA
CATGGAGCACATTTTCTGCACCGAGTGCGGCACCGCCCTGATGCGGCAGGGCGACACTTTCGCGGCGCTCTTCCCGCCAAACT
TCTGGCTGGAGACCCCTGACCCCTCCTCCAAGTGCGGTGTGAGCTGCAAGCTGCCCGCGAGCTGCTGCCCAACCGGCCTAC
AAGTATGAGAACCGCCACCAGGACTGGAACGACAATCTGCCCAAGTACATCAAATATGACGAGTCGCCCCGCGTCAAGAACAA
CGGCTTCGCCATCTTTACTACTGACCTAGAGTCTACAAAGGACGAATCGCAAGGGGACACGTCTCTGCTGCGTCATCTCATAGC
CACCGGAGCAGTGGCCAATGGATCTCGTAGGATCTCTGGGCTGGGCGTATCTAGATGCCACCCATTGCTTTTATGCTCTTT
TGTACGGGGCTCTCTGCGCACACGCACTACACACGCACGGCGGCACCgCAACCGCGTGATCACGGGCGGTGAGATTAa
GCAACTCAATAAAAAATAG-3'

Annexe 8 : Profil des blocs A, B, C, D, E et F (format WebLogo) de la famille Lipase 3 indexée IPB005592 dans la base de données Blocks.

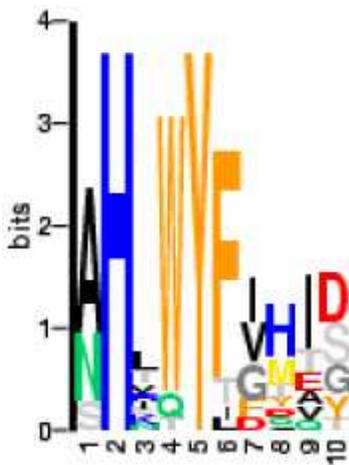




PSSM of IPB005592D (L1pase3_N); 22 sequences.



PSSM of IPB005592E (L1pase3_N); 22 sequences.



PSSM of IPB005592F (L1pase3_N); 22 sequences.

Annexe 9 : Aligement des séquences génomique et d'ADNc d'IgLip

CLUSTAL W 2.0 multiple sequence alignment

Les séquences introniques sont surlignées en gris.

```

IgLipADN      AGGCGGCCAAACATGCGCAGGCGGCGGCGGCGAAAGAGGCTGTCCTGGCGAGCCACGCCA 60
IgLipADNc    -----ATGCGCAGGCGGCGGCGGCGAAAGAGGCTGTCCTGGCGAGCCACGCCA 48
              *****
IgLipADN      TGGCGCAGCGATTGGATTCCGACCTGCGTACTTTTCGAAACGTGGCGCTCCTGCCGCGCG 120
IgLipADNc    TGGCGCAGCGATTGGATTCCGACCTGCGTACTTTTCGAAACGTGGCGCTCCTGCCGCGCG 108
              *****
IgLipADN      GCAGACAATGAGGCGTTGATCGCCGAGCGTCTCCGCTTGGCCCGCCGGGTGTGGGCCACC 180
IgLipADNc    GCAGACAATGAGGCGTTGATCGCCGAGCGTCTCCGCTTGGCCCGCCGGGTGTGGGCCACC 168
              *****
IgLipADN      GTCAATGCGCCGCTGCTGCTCGAGCTCGCGGCGGCTCTGCTGTGCCACCACAAATGGTC 240
IgLipADNc    GTCAATGCGCCGCTGCTGCTCGAGCTCGCGGCGGCTCTGCTGTGCCACCACAAATGGTC 228
              *****
IgLipADN      GTTCGCTGGACGTGTTACGCTGCGGAGGGACTGGCTTGCAGCCGATCCGCAAGTCATC 300
IgLipADNc    GTTCGCTGGACGTGTTACGCTGCGGAGGGACTGGCTTGCAGCCGATCCGCAAGTCATC 288
              *****
IgLipADN      ATCGCAGACGGCGGCTCCTTCCTCGCGCTAGTGGGTATCAATGAAGCCGAACAGTGC 360
IgLipADNc    ATCGCAGACGGCGGCTCCTTCCTCGCGCTAGTGGGTATCAATGAAGCCGAACAGTGC 344
              *****
IgLipADN      GCGCCCTGCCTGTTGCTCGCTGGCAGTGCCCTGTGCAAACATAAACGACTCGCTC 420
IgLipADNc    -----
              *****
IgLipADN      CACGGCACCAGCGCCACAGCTTGGCAACCAATGGTCAAGTTCGGCCCGGCTGCCTCAGCG 480
IgLipADNc    -----
              *****
IgLipADN      GTGTGCTCACACGTACTTGCTTTCTCTGCTGCGTCCCTGCCACAGGTGGGTGGTGGTGG 540
IgLipADNc    -----GGTGGTGGTGGTGG 355
              *****
IgLipADN      CCTTTCGCGGCTCTCTCGACGGCAACCTAGGGGCTTGGGGTCGCAACTTGCAATGCTGGA 600
IgLipADNc    CCTTTCGCGGCTCTCTCGACGGCAACCTAGGGGCTTGGGGTCGCAACTTGCAATGCTGGA 415
              *****
IgLipADN      CCGACCCGAGGAGCTTCCACAATTGCTCCGGCTGGGTGCACCATGGCTGGGGTAGTGGCT 660
IgLipADNc    CCGACCCGAGGAGCTTCCACAATTGCTCCGGCTGGGTGCACCATGGCTGGGGTAGTGGCT 475
              *****
IgLipADN      ACGACATGCTTCGCGCGCGCCTCATCACGGCTGTGCTGCTAGCGCTTGAGGCTCGGCCGA 720
IgLipADNc    ACGACATGCTTCGCGCGCGCCTCATCACGGCTGTGCTGCTAGCGCTTGAGGCTCGGCCGA 535
              *****
IgLipADN      CGTACCGCCTCGTCGTCACCGGCCACTCTGCCGCGGGGACCTGGCCTCCATCTTTGTGG 780
IgLipADNc    CGTACCGCCTCGTCGTCACCGGCCACTCTGCCGCGGGGACCTGGCCTCCATCTTTGTGG 595
              *****
IgLipADN      CCGAGCTCGCCACCGTCGGAGCAGCTCAATCATCTCGGTAGCCAGCGAATCGTCCGCGT 840
IgLipADNc    CCGAGCTCGCCACCGTCGGAGCAGCTCAATCATCTCGGTAGCCAGCGAATCGTCCGCGT 655
              *****
IgLipADN      ACCGCGCCGAGGCGGCTCTGACTCCAAAGGCGAAGCTGACACGAAGGCTGCCGCGGCTG 900
IgLipADNc    ACCGCGCCGAGGCGGCTCTGACTCCAAAGGCGAAGCTGACACGAAGGCTGCCGCGGCTG 715
              *****
IgLipADN      ATGAGTCCGCCCGCGCTCCAGCTCCATCGCTGGGAGAGTGTTCGGAGGTTGCGCCGC 960
IgLipADNc    ATGAGTCCGCCCGCGCTCCAGCTCCATCGCTGGGAGAGTGTTCGGAGGTTGCGCCGC 775
              *****
IgLipADN      TGCTCAGGCGACCGCTCGCGCCGCGCCGCTGTACATCACCTTCGCGCGCCGATGGCCG 1020
IgLipADNc    TGCTCAGGCGACCGCTCGCGCCGCGCCGCTGTACATCACCTTCGCGCGCCGATGGCCG 835
              *****
IgLipADN      GGAACCGCGAGTTTGGCGCCGCATCGATTCCGGCCTGGCCGCGCAGGGCAACGTGATGC 1080
IgLipADNc    GGAACCGCGAGTTTGGCGCCGCATCGATTCCGGCCTGGCCGCGCAGGGCAACGTGATGC 895
              *****
IgLipADN      TACGTGTGGTCAATGCACGCGATCCGGCCGTGCGTGCACCAACACGCTGAATGCGGGTG 1140
IgLipADNc    TACGTGTGGTCAATGCACGCGATCCGGCCGTGCGTGCACCAACACGCTGAATGCGGGTG 955
              *****
IgLipADN      GCAGCGTCCAGATCCCCTGCCCAATGGGGGGCGAGGACCAGTTTTACATGCGGGGCAGG 1200
IgLipADNc    GCAGCGTCCAGATCCCCTGCCCAATGGGGGGCGAGGACCAGTTTTACATGCGGGGCAGG 1015
              *****

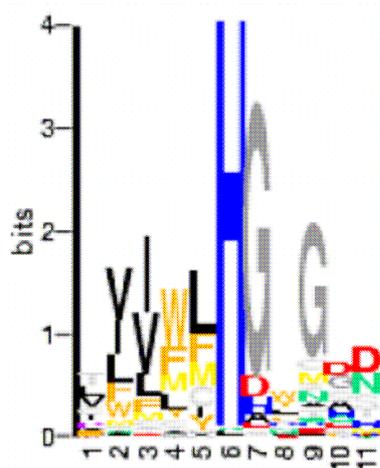
```

```

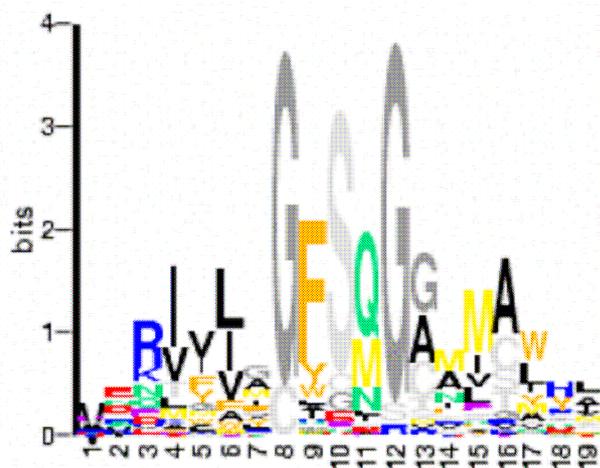
IgLipADN      AAGTGTGGCTGACAAACGCTGCCGACGCCGACATGAGCAAGGCGCAATTCTGCAACAGCT 1260
IgLipADNc    AAGTGTGGCTGACAAACGCTGCCGACGCCGACATGAGCAAGGCGCAATTCTGCAACAGCT 1075
*****
IgLipADN      CAGGCGCGTTCGGGGACGGCGATGACGCCAGCATGTCCGACCCGAACCCGCTGTGCTCCA 1320
IgLipADNc    CAGGCGCGTTCGGGGACGGCGATGACGCCAGCATGTCCGACCCGAACCCGCTGTGCTCCA 1135
*****
IgLipADN      GCTCCCTCGGCGAGTCGAGCCTGACTTGGGTACGCCGCAAAGCCGAAGCCGCTAAGAGC 1380
IgLipADNc    GCTCCCTCGGCGAGTCGAGCCTGACTTGGGT----- 1166
*****
IgLipADN      CACAATGGCCACCCCCCAAGACGCGCGCACACTGTTTCGCGATTAGCACTTGCGGATGCCC 1440
IgLipADNc    -----
*****
IgLipADN      GTCTCATTCCGCTCTGCGGCCTCGCGCTCAGGTCCACCACAACTCATACCTCGGCGTCGAC 1500
IgLipADNc    -----CCACCACAACTCATACCTCGGCGTCGAC 1194
*****
IgLipADN      ACAGGCGCGTGCCCCGATACGATGGATGCGAATCAGTGGGAAGAGTACATCAAGTATACC 1560
IgLipADNc    ACAGGCGCGTGCCCCGATACGATGGATGCGAATCAGTGGGAAGAGTACATCAAGTATACC 1254
*****
IgLipADN      AAGAGTCTCCACCCGAGTTGAGACACGTCGGCTATGGTGAGGACGCACCCGACGGCGGT 1620
IgLipADNc    AAGAGTCTCCACCCGAGTTGAGACACGTCGGCTATGGTGAGGACGCACCCGACGGCGGT 1314
*****
IgLipADN      TTCGACAGCACGGTGCAGCGGTCCGAGGCGGAGCGGGTGCCGACATACTCGTACGGATGA 1680
IgLipADNc    TTCGACAGCACGGTGCAGCGGTCCGAGGCGGAGCGGGTGCCGACATACTCGTACGGATGA 1374
*****
IgLipADN      TGTCCGATGGTGCA 1694
IgLipADNc    -----

```

Annexe 10 : Profil des blocs A et B (format WebLogo) de la famille Phospholipase / Carboxylestérase indexée IPB003140 dans la base de données Blocks.



PSSM of IPB003140A (PLP_Cesterase;) 570 sequences.



PSSM of IPB003140B (PLP_Cesterase;) 570 sequences.

Annexe 11 : Extraction de signaux phylogénétiques fiables avec Gblocks

Nombre de séquences : 60 / Type d'alignement : protéines / Positions sélectionnées :174

```

                250      260      270      280      290      300
=====+=====+=====+=====+=====+=====+
D.virilis AAW4 NAVSQALKPSQWLSV-----QELQAIPAVD---ELTL
D.montana ABU80 NTVNQALKPSQWLSV-----QELEAIPSD---ELTL
D.americana AB NTVSQALKPSQWLSV-----QELQSIPAVD---ELTL
D.ezoana AAW476 NTVSQALKPSQWLSV-----QELQAIPAVD---ELTL
D.littoralis AB NSVSQALKPSQWLSV-----QELQAIPAVD---ELTL
TpPLip/CEst1 PRSFHPTAAMSTAN-----QCDNSKSAII-FLHGLGD
O.sativa_509295 -----MEPATSRAP-----R----SRFVV-WLHGLGD
O.sativa_CAE028 -----MAA-----A--GGYGRVL-WLHGSQ
O.sativa_CAE028 -----MAAAAA-----RNRSFVL-WLHGLGD
A.thaliana_1523 MRTSRLKKPIVLLTIALSSVFFFAFFFLNKSDVSSTSALRNRDSMARTFII-WLHGLGD
A.thaliana_BAC4 MRTSRLKKPIVLLTIALSSVFFFAFFFLNKSDVSSTSALRNRDSMARTFII-WLHGLGD
P.aeruginosa_A8 -----MSEPLILDAP-----NADACII-WLHGLGA
P.fluorescens_Q -----MTEPLILQPA-----K--PADACVI-WLHGLGA
P.putida_AAN669 -----MTNPLILEPQ-----K--TADACVI-WLHGLGA
Igest1_CAO02582 MSTAAGAAPPRTIQD-----RARNNLITIAPEGEGMPHTATVIGPIHGLGD
PtPLip/CEst FQTRFRKDGITITVTPK-----NEAEQSALVV-ISHGLGD
TpPLip/CEst2 -----SATIV-ICHGLGD
D.discoideum_EA MIKTILKNSYIQKEI-----T--THSATVI-FSHGLGD
C.albicans_6847 TTESMSVSAIRIPAN-----GS-SAKAAVI-FLHGLGD
D.hansenii_5042 SKKMSQLIPAVRVAA-----TAKPAKSAAII-FVHGLGD
Y.lipolytica_50 ---MPPYPAVRIPAK-----A--AHTATVI-FLHGLGD
S.japonicum_AAD -MANKLLPAAVVASR-----S--KHTATLI-FLHGLGD
S.mansoni_AAC62 IDMANKLLPAAVVAS-----RS-KHSATLI-FLHGLGD
C.briggsae_CAE6 MSRIANGEPAVVRAR-----G--QHKGTII-FLHGLGD
C.elegans_T2332 MSAVSNGNPSIVSPR-----G--EHKGTII-FLHGLGD
S.purpuratus_72 YSSMAACTGPALLPS-----A--KHTATVI-FLHGLGD
T.nigroviridis_ MSVPLLANAVTVSGT-----E--KETAVVI-FLHGLGD
D.rerio_4115218 MSVPLLAEAVTVPGT-----E--KETAVVI-FLHGLGD
S.tropicalis_AA MSVPLLTDAVTVPAG-----E--RETGAVI-FLHGLGD
X.tropicalis_45 MSVPLLTDAVTVPAG-----E--RETGAVI-FLHGLGD
X.laevigata_AAH443 MSVPLLTDAVTVPAG-----E--RETGAVI-FLHGLGD
G.gallus_507598 MSVPLLDAAVTVSGA-----E--RETAIVI-FLHGLGD
P.troglodytes_5 MSVPLLTDAATVSGA-----E--RETAIVI-FLHGLGD
C.familiaris_73 MSVPLLTDAATVSGA-----E--RETAIVI-FLHGLGD
H.sapiens_AAC72 MSVPLLTDAATVSGA-----E--RETAIVI-FLHGLGD
M.musculus_7242 MSVPLLTDAATVSGA-----E--RETAIVI-FLHGLGD
R.norvegicus_13 MSVPLLTDAATVSGA-----E--RETAIVI-FLHGLGD
T.nigroviridis_ MSAPL---PAIVPAA-----R--RATAIVI-FLHGLGD
X.laevigata_AAH733 MSAPL---PAIVPAA-----R--KATAIVI-FLHGLGD
X.tropicalis_AA MSAPL---PAIVPAA-----R--KATAIVI-FLHGLGD
G.gallus_507376 MSAPL---PAIVPAA-----R--KATAIVI-FLHGLGD
M.musculus_P978 MSAPM---PAVVPAA-----R--KATAIVI-FLHGLGD
O.cuniculus_077 MSAPM---PAVVPAA-----R--KATAIVI-FLHGLGD
R.norvegicus_P7 MSAPM---PAVVPAA-----R--KATAIVI-FLHGLGD
B.taurus_777363 MSAPL---PAIVPAA-----R--KATAIVI-FLHGLGD
O.cuniculus_AAC -----ASVPAA-----R--KATAIVI-FLHGLGD
H.sapiens_07560 MSTPL---PAIVPAA-----R--KATAIVI-FLHGLGD
P.pygmaeus_CAH9 MSTPL---PAIVPAA-----R--KATAIVI-FLHGLGD
D.melanogaster_ -----MAAPVIVEAT-----V--KQTATLI-FMHGLGD
A.gambiae_31239 -----MLQLI-FLHGLGD
A.mellifera_481 ---MTTVSPVVI AAT-----A--RHTATLI-FFHGLGD
U.maydis_710031 ---MSVLKTLVINPR-----SGVKPTATLF-FLHGLGD
F.neoformans_EA ---MPTSLKHLKISP-----E--AHTATVI-FLHGLGD
E.nidulans_EAA6 ---MSRAPFIVPAL-----K--KHTATVI-MAHGLGD
A.fumigatus_709 ---MAPRAPYIVPAL-----K--KHTATVI-MAHGLGD
A.niger_1395827 ---MATRAPFIVPAL-----K--KHTATVI-MAHGLGD
A.oryzae_BAE627 ---MATRAPFVVPAL-----K--KHTATVI-MAHGLGD
N.crassa_CAC100 MASLARRPPLVPAV-----A--RHTATVI-FIHGLGD
G.zeae_EAA74608 ---MSSGRIAPLVFPAA-----S--QHTATVI-FVHGLGD
M.grisea_EAA475 ---MPNYPAPLRFPA-----G--RHTATVI-FIHGLGD
#####

```

```

          310          320          330          340          350          360
=====+=====+=====+=====+=====+=====+
D.virilis AAW4  ERLENMSLEKGAELLQQVY-----HLSQINHDVEPNFVPSNIQVYVP-----
D.montana ABU80 ERLENMSLEKGAELLQQVY-----HLSQINHDVEPNFVPSNIQVYVP-----
D.americana AB  ERLENMSLEKGAELLQQVY-----HLSQINHDVEPNFVPSNIQVYVP-----
D.ezoana AAW476 ERLENMSLEKGAELLQQVY-----HLSQINHDVEPNFVPSNIQVYVP-----
D.littoralis AB ERLENMSLEKGAELLQQVY-----HLSQINHDVEPNFVPSNIQVYVP-----
TpPLip/CEst1   SPAGWSSLERNLPSLRPKLG-----DGVHYVFPAPTISLTINDGMMMPGWFDLYDWP
O.sativa_509295 TGRANEFLADS-FPTTAAF-----ADARWAFPTAPTAPVTCNRGLMLPWSWFDIHDAP
O.sativa_CAEO28 TGEESRAQVAP-YFAAPEL-----ASVRFSPPTAPTSSIPCYGGEVITAWFAIPEVP
O.sativa_CAEO28 SGPANEP IRN--FFSAPEF-----RLTKWAFPSAPNSPVSCNHGAVMPSWFDIHELP
A.thaliana_1523 SGPANEP IKT--LFRSQEF-----RNTKWLFPSPAPPNPVSCNYGAVMPSWFDIPELP
A.thaliana_BAC4 SGPANEP IKT--LFRSQEF-----RNTKWLFPSPAPPNPVSCNYGAVMPSWFDIPELP
P.aeruginosa_A8 DRTDFKPVA---EALQMVLF-----PSTRFILPQAPSQAVTVNGGWVMPWSYDILAF-
P.fluorescens_Q DRYDFLPVA---EALQETL-----LSTRFVLPQAPTRPVTINGGYEMPSWYDIKAM-
P.putida_AAN669 DRYDFLPVA---EFMQERL-----LSTRFIMPQAPTRPVTINGGYAMPWSYDIKAM-
IgEst1_CA002582 SNMGWADV---MQLQSQVM-----PYCKFILPNAPVPRVPTLNGGMSMPWSYDITSL-
PtPLip/CEst    SAEGFADVA---ETLAMQM-----PHVKFVLPAPTQPVTMNMGMSMPWSYDIVGL-
TpPLip/CEst2   TAQGWEDVA---RNFASKL-----PYAKFILPTAPTRKITMNMGMSMPGWYDIVGL-
D.discoideum_EA SGAGWIEVME---EIQSRNN-----GHIRFICPNAPIQAVTLNGGFKMPWSYDIKSL-
C.albicans_6847 SGDGWSWLPQLVVSQSKLIN-----DPINYVFPNAPKIPVTTINNGFAMPWFDIYELG
P.hansenii_5042 SGGSGWS-----WFPQLAKQSNIIKNCDSINYVFPNAPLMPITANGGYVMPGWFDIYEFG
Y.lipolytica_50 SGAGWMFLAEE-ARKAQRLL-----NHVKFIFPEAPQQPVSLNFGMRMPSWYDIKELA
S.japonicum_AAD TGHGWS-----TLRQVVP-----NYFKVICPHANSIPVTLNGGCMCPAWYDIFAL-
S.mansoni_AAC62 TGHGWS-----ALKEYVP-----DYFKIICPHANSIPVTLNGGCMCPAWYDIYAL-
C.briggsae_CAEO6 QGTGWADA---FSTEARH-----DNIKAIICPHSAERSVTLNMGMRMPAWYDLFGL-
C.elegans_T2332 QGHGWADA---FKTEAKH-----DNIKFICPHSSERPVTLNMGMRMPAWYDLFGL-
S.purpuratus_72 QGHGWCS-----SFEEIKE-----PHIKYIFPNAPNPNVTLNMGVMPWSWFDIISL-
T.nigroviridis_ TGYGWAE-----TLTEIQP-----PYVKFICPHAPTIPVTLNKNMMPAWFDLIGL-
D.erio_4115218 TGHGWAD-----AMTIRL-----PYIKYICPHAPRIPVTLNLMKMTMPSWFDLMGL-
S.tropicalis_AA TGHGWAE-----ALSAILR-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
X.tropicalis_45 TGHGWAE-----ALSAILR-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
X.laevis_AAH443 TGHGWAE-----TLSAIKL-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
G.gallus_507598 TGHSWAE-----ALSSIPL-----PYVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
P.trogodytes_5  TGHSWAD-----ALSTIRL-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
C.familiaris_73 TGHSWAD-----ALSTIRL-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
H.sapiens_AAC72 TGHSWAD-----ALSTIRL-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
M.musculus_7242 TGHSWAD-----ALSTIRL-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
R.norvegicus_13 TGHSWAD-----ALSTIRL-----PHVKYICPHAPRIPVTLNMMKVMMPAWFDLMGL-
T.nigroviridis_ TGHSWAD-----TFAGLRL-----PHVKYICPHAPVMPVSLNMMNMSRWSWFDIHGL-
X.laevis_AAH733 TGHGWAE-----AMASIKS-----PHVKYICPHAPIMPVSLNMMNAMPWFDIIGL-
X.tropicalis_AA TGHGWAE-----AMASIKS-----PHVKYICPHAPIMPVSLNMMNAMPWFDIIGL-
G.gallus_507376 TGHGWSE-----ALAGIKS-----PHVKYICPHAPFMPVTLNMMNAMPWFDIIGL-
M.musculus_P978 TGHGWAE-----AFAGIKS-----PHIKYICPHAPVMPVTLNMMNAMPWFDIVGL-
O.cuniculus_O77 TGHGWAE-----AFAGIKS-----PHIKYICPHAPVMPVTLNMMNAMPWFDIVGL-
R.norvegicus_P7 TGHGWAE-----AFAGIKS-----SHIKYICPHAPVMPVTLNMMNAMPWFDIIGL-
B.taurus_777363 TGHGWAE-----AFAGIRS-----AHIKYICPHAPVMPVTLNMMNAMPWFDIIGL-
O.cuniculus_AAC TGHGWAE-----AFAGIRS-----SHIKYICPHAPVMPVTLNMMNAMPWFDIIGL-
H.sapiens_O7560 TGHGWAE-----AFAGIRS-----SHIKYICPHAPVMPVTLNMMNAMPWFDIIGL-
P.pygmaeus_CAH9 TGHGWAE-----AFAGIRS-----SHIKYICPHAPVMPVTLNMMNAMPWFDIIGL-
D.melanogaster_ TGHGWSS-----ALAAIRP-----PFMKVICPTAPTQPVSLNAGFRMPSWFDLKTLL-
A.gambiae_31239 TGHGWAT-----SMGALRT-----PDMKVICPTAPNMPVTMNGGFRLLNSWFDLKSII-
A.mellifera_481 TGHGWAS-----SMGAVRS-----PHIKVICPTASTMPVTLNAGFRMPSWFDLRSLS-
U.maydis_710031 SSAGWSDVAQM--LSQRPSL-----SHVRFVLPNAPIQPVTNMGMPMPWSWFDILALD
F.neoformans_EA SGHWLPLVA---KMLWSSF-----PNVKWILPHAPTIPVSLNHGMAMPWFDIRHLD
E.nidulans_EAA6 SGAGWVSLAHN-WRRRGLF-----EETVFIIPNAPMIPITVNFGMSMPGWYDITKLG
A.fumigatus_709 RMSLAQN-----WRRRGMF-----DEVAFIFPNAPMIPITVNFGMTMPGWHDLTKLG
A.niger_1395827 SGAGWMALAQN-WRRRGMF-----DEVAFIFPNAPMIPITVNFGMSMPGWYDISKLG
A.oryzae_BAE627 RFAGYACKN---WRRRGLF-----EETVFIIPNAPMIPITVNFGMSMPGWYDLSKLG
N.crassa_CAC100 TGHGWASAVEQ-WRRRQRL-----DEVKFILPHAPSIPITANWGMKMPGWYDIFAID
G.zeae_EAA74608 TGHGWASAVEN-WRRRQKM-----DEVKFILPHAPQIPISVNMGMRMPGWFDIKQLG
M.grisea_EAA475 SGHWAPAVEN-WRRRQKL-----DEVKFILPHAPTIPVTCNMGMMPGWYDIKSID
#####

```

```

          370          380          390          400          410          420
=====+=====+=====+=====+=====+=====+
D.virilis AAW4 -----KPNGDK----IVAPINEMIQLRKQKQ-N-FG-----QDEVTIIIVTGLPQT
D.montana ABU80 -----KPNGDK----IVAPINEMIQLRKQKQ-N-FG-----QDEVTIIIVTGLPQT
D.americana AB -----KPNGDK----IVAPISEMIQLRKQKQ-N-FG-----QDEVTIIIVTGLPQT
D.ezoana AAW476 -----KPNGDK----IVAPINEMIQLRKQKQ-N-FG-----QDEVTIIIVTGLPQT
D.littoralis AB -----KPNGDK----IVAPINEMIQLLKQKQ-N-FG-----QDEVTIIIVTGLPQT
TpPlip/Cest1 I----GINAKDDQEGKLAANVIEETVKRLE-EEEG-----ISPSRIIVGGFSQG
O.sativa_509295 I----TSVSVRDEEDVLRVQSVHAMIDREI-A-AG-----TNPQDVFVFGLSQG
O.sativa_CAE028 I----TARTARDEKEVLKAVERVHMLDGEV-A-AG-----TSPSNIFVCGLSQG
O.sativa_CAE028 M----SSGSQDDSGVLKAVENVHAMIDKEV-A-DG-----IPPENIFVCGFSQG
A.thaliana_1523 L----TAGSPKDESSLKAVKNVHAIIDKEI-A-GG-----INPENVIICGFSQG
A.thaliana_BAC4 L----TAGSPKDESSLKAVKNVHAIIDKEI-A-GG-----INPENVIICGFSQG
P.aeruginosa_A8 -----SPARAIDEDQLNASADQVIALIDEQR-A-KG-----IAAERIIILAGFSQG
P.fluorescens_Q -----SPARISLEELLETSAKTVTDLIETQQ-R-TG-----IDTSRIFLAGFSQG
P.putida_AAN669 -----TPARAIDEAQLEESAQVVALIKAEQ-A-KG-----IDLTRIFLAGFSQG
IgEst1_CAO02582 -----DKRESQPCTGIEESRQAMLDLISAEV-A-SG-----IPPSRIIAGFSQG
PtPlip/Cest -----DERANENCKGIEISRTRITSILEEEHAN-TG-----LPYRRMVLGFSQG
TpPlip/Cest2 -----DKRSNENCPGIDESQTRILDILKSEN-D-AG-----IHYNRMVLGFSQG
D.discoideum_EA -----SSRGDEDPAQVDESKNI IETIIKHEM-EEEK-----IPAERIIIGGFSQG
C.albicans_6847 -----NPHAKQDVTGFFKSCVLEKFEILEQH-NKFN-----IPLEKIIIGGFSQG
D.hansenii_5042 -----NPEAKQDIDGPHKSCETLKSILIKEQI-DNHD-----IPADKIIIGGFSQG
Y.luolaitica_50 -----NVNAAQDQEGILEESQVRLLESILKEET-D-AG-----VPANRIVIGGFSQG
S.japonicum_AAD -----SENAKQDEPGIKGASVELGKFVDAKI-K-AG-----IPVENIVIGGFSQG
S.mansoni_AAC62 -----SENAKQDEAGIKEASLELGKFVDAEI-K-AG-----VPIGNIVIGGFSQG
C.briggsae_CAE6 -----DATAREDADGIQAAAQVHHLIDEEI-N-AG-----IPADHIAVGGFSMG
C.elegans_T2332 -----DPNAQEDEQGINRATQYVHQLIDAEV-A-AG-----IPASRIAVGGFSMG
S.purpuratus_72 -----GAEGKEDKEGILKASANLLKMVAEEE-S-HG-----IAPNRIVIGGFSQG
T.nigroviridis_ -----SHDSPEDETGIKKAETIKALIEHEA-K-NG-----IPPHRIILGGFSQG
D.rerio_4115218 -----SPESPEDEAGIKKRAENIKAIIDHEV-K-NG-----IPSNRIVLGGFSQG
S.tropicalis_AA -----SPDAPEDEAGIKKAAESIKTIIIEHEV-K-NG-----IPANRIVLGGFSQG
X.tropicalis_45 -----SPDAPEDEAGIKKAAESIKTIIIEHEV-K-NG-----IPANRIVLGGFSQG
X.laevis_AA443 -----SPDAPEDEAGIKKAAESIKTIIIEHEV-K-NG-----IPANRIVLGGFSQG
G.gallus_507598 -----TPDAPEDEAGIKKAAESIKAIIEHEM-K-NG-----IPPNRILLGGFSQG
P.troglodytes_5 -----SPDAPEDEAGIKKAAENIKALIEHEM-K-NG-----IPANRIVLGGFSQG
C.familiaris_73 -----SPDAPEDEAGIKKAAENIKALIEHEM-K-NG-----IPANRIVLGGFSQG
H.sapiens_AAC72 -----SPDAPEDEAGIKKAAENIKALIEHEM-K-NG-----IPANRIVLGGFSQG
M.musculus_7242 -----SPDAPEDEAGIKKAAENIKALIEHEM-K-NG-----IPANRIVLGGFSQG
S.norvegicus_13 -----SPDAPEDEAGIKKAAENIKALIEHEM-K-NG-----IPANRIVLGGFSQG
T.nigroviridis_ -----SPDAAEDEPGIKQASENIKAMIDQEV-K-NG-----IPSHRIILGGFSQG
X.laevis_AA733 -----SPDAQEDEAGIKKAAENVKALIDQEV-K-NG-----IPSNRIILGGFSQG
X.tropicalis_AA -----SPDAQEDEAGIKKAAENVKALIDQEI-K-NG-----IPSNRIILGGFSQG
G.gallus_507376 -----SPDSQEDEVGIKQAENVKALIDQEV-K-NG-----IPSNRIILGGFSQG
M.musculus_P978 -----SPDSQEDESIGIKQAETVKALIDQEV-K-NG-----IPSNRIILGGFSQG
O.cuniculus_077 -----SPDSQEDESIGIKQAETVKALIDQEV-K-NG-----IPSNRIILGGFSQG
R.norvegicus_P7 -----SPDSQEDESIGIKQAETVKALIDQEV-K-NG-----IPSNRIILGGFSQG
B.taurus_777363 -----SPDSLEDEAGIKKAAENVKALIDQEV-K-NG-----IPSNRIILGGFSQG
O.cuniculus_AAC -----SPDSQEDEPGIKQAENVKALIDQEV-K-NG-----IPSNRIILGGFSQG
H.sapiens_07560 -----SPDSQEDESIGIKQAENIKALIDQEV-K-NG-----IPSNRIILGGFSQG
P.pygmaeus_CAH9 -----SPDSQEDESIGIKQAENIKALIDQEV-K-NG-----IPSNRIILGGFSQG
D.melanogaster_ -----DIGGPEDEPGIQSARDSVHGMIQKEI-S-AG-----IPANRIVLGGFSQG
A.gambiae_31239 -----SISDPEDEEGIKKATRYVHELIQSEM-K-AG-----ILSNRIMLGGFSQG
A.mellifera_481 -----EPSGPEDEEGIRRAEMVHSLIAEEV-A-AG-----IPTKRIVLGGFSQG
U.maydis_710031 -----DLSGAEDEAGLLKSTDEIKKLIKAEN--DGTAKDLDGHKIPSERIVVGGFSQG
F.neoformans_EA KL----DNSSENDEQGMLETLSVDELIIQAEV-D-SG-----IPENRIVLGGFSQG
E.nidulans_EAA6 RDLDFQEAVKNQDEAGILKSRDYFNLSILIKEQM-D-QG-----IKPSRIVLGGFSQG
A.fumigatus_709 RELDYESAIRHQDEPGVLRSDRYFNNTLIKEQI-D-KG-----IKPSRIVLGGFSQG
A.niger_1395827 RDLDFEAAIRHQDEPGVLRSDRYFNNTLIKEQI-D-QG-----IKPSRIVLGGFSQG
A.oryzae_BAE627 RDLDFEAAIRSQDEPGILRSREYFNNTLIKEQI-D-QG-----INPSRIVLGGFSQG
N.crassa_CAC100 GSAE--ALRRNEDEAGILTSQAYFHDLIQKEI-D-SG-----IPADRIVIGGFSQG
G.zeae_EAA74608 GDVD--SLIRNEDETEGIKLSQKYFHNLIQKEI-D-SG-----IVPERIVLGGFSQG
M.grisea_EAA475 GTPE--SLRKDEDEEGILLSQKYFHDLIQKEI-D-AG-----IPSERIVLGGFSQG
#####

```

```

          430          440          450          460          470          480
=====+=====+=====+=====+=====+=====+
D.virilis AAW4 S-----ETVKKATR---KLVQAYMQRYNLQQQQQAQKYNT----ESEEDNKRQRTSSEE
D.montana ABU80 S-----ETVKKATR---KLVQAYMQRYNLQQQQQAQKYNS----ESNEDNKRQRTSSEE
D.americana AB S-----ETVKKATR---KLVQAYMQRYNLQQQQQAQKYNS----ESEEDNKRQRTSSEE
D.ezoana AAW476 S-----ETVKKATR---KLVQAYMQRYNLQQQQQAQKYNS----ESEDNKRQRTSSEE
D.littoralis AB S-----ETVKKATR---KLVQAYMQRYNLQQQQQAQKYNS----ESEDNKRQRTSSEE
TpPlip/Cest1 GAVALLTAYHRRTOGKVPFAGCVCLSGWLTLLKDDLSVTDEV-----AKSTPLFWGHGQY
O.sativa_509295 GALGIASVLL-HPK---TLGGCAVFSGFLPFNSSFAVRVTA-----QAKKTPVLWIHGQA
O.sativa_CAE028 GALAIASVLL-YPM---TLGGCVVFSGSLPLSKTFAESIPS-----EARKTPVLWFWHGMA
O.sativa_CAE028 GALTLASVLL-YPK---TLGGGAVFSGWLPGSSVTERISP-----EARKTPILWWSHGIA
A.thaliana_1523 GALTLASVLL-YPK---TIGGGAVFSGWIPFNSSITNQFTE-----DAKKTPLWWSHGID
A.thaliana_BAC4 GALTLASVLL-YPK---TIGGGAVFSGWIPFNSSITNQFTE-----DAKKTPLWWSHGID
P.aeruginosa_A8 GAVVLHTAFRRYAQ---PLGGVLALSTYAPTFFDDLALDERH-----KRIPLVHLHGSQ
P.fluorescens_Q GAVVHTAFKKWEG---PLGGVIALSTYAPTFFDNDLQLSAS-----QQRIPTLCLHGQY
P.putida_AAN669 GAVVLHTAYIKWQE---ALGGVIALSTYAPTFFDNDLQLSAC-----QQRTPALCLHGQY
IgEstl_CAO02582 GAVALFTGLQ-YSH---TLAGVLCCLSGYLAEEERFILAPEA-----VNTPVAFHFGSD
PtPlip/Cest GALSlyTGLQ-LKAEQ-KLAAIVMMSGYLPAAKTFVAVTTGL-----ESVPVLHCHGTQ
TpPlip/Cest2 AALSlyTGMQ-LPAEAGPLAGIVAMSGYLPASGPNITPGL-----ESTPLFHAHGA
D.discoideum_EA AALSlyTFYSQTET---KLGCCIALSGYLPLATKFFVANSLN-----KEQPLLMIHGDC
C.albicans_6847 AAISLATLAL-LDT---KIGGCVALSGFCPVREITDRYNK---NPGVNFDTPIFQGHGTV
D.hansenii_5042 AAVSLATVAL-LDF---KVGGVVALSGFSPIKESLPIQIMNK-----ANLETFIFQGHGTA
Y.lipolytica_50 GALSLATGCL-TQT---KLGIVGLSGYVPIKDYILSQHNT-----TNQDTPMFLWWSHGIA
S.japonicum_AAD GSVPLYNALT-STL---RYGGIVAFNCWLPLHTKFMSSPTL---LTIPKDVPIFQCHGLD
S.mansoni_AAC62 GSVLYNALT-STL---QYGGVAVFSCWLPLHTKFMSSPTL---LTMKDVPIFQCHGLE
C.briggsae_CAE6 GALAIYAGLT-YPK---KLGAIVGLSFFLQRQKFPGSYTA-----NNATPIFLGHGSQ
C.elegans_T2332 GALAIYAGLT-YPQ---K2332VGLSFFLQRQKFPGSFTA-----NNATPIFLGHGTD
S.purpuratus_72 GAVSLYSALT-DDR---PYAGVLALSTWMLPHQTF-----
T.nigroviridis_ GALSlyTALT-CQY---QLAGVVALSCWLPLHKSFPVSKMHFLLWMTSTSPSCSVHGM
D. rerio_4115218 GALSlyTALT-SQQ---QLAGVVALSCWLPLHKTFFQAAGA-----SANKDTPILQCHGEM
S.tropicalis_AA GALSlyTALS-CQH---KLAGVIGLSCWLPLHKTFFQAASG-----VNKEISVLQCHGEA
X.tropicalis_45 GALSlyTALS-CQH---KLAGVIGLSCWLPLHKTFFQAASG-----VNKEISVLQCHGEA
X.laevis_AAH443 GALSmyTALT-CQH---KLAGVVALSCWLPLHKTFFQAASG-----VNKEISVMQCHGEA
G.gallus_507598 GALSlyTALT-CQH---QLAGIVALSCWLPLHKAFFQAASN-----GVNKDIALQCHGEM
P.troglodytes_5 GALSlyTALT-CPH---PLAGIVALSCWLPLHRAFPPQAANG-----SAKDLAILQCHGEL
C.familiaris_73 GALSlyTALT-CPH---PLAGIVALSCWLPLHRAFPPQAANG-----SAKDLAILQCHGEL
H.sapiens_AAC72 GALSlyTALT-CPH---PLAGIVALSCWLPLHRAFPPQAANG-----SAKDLAILQCHGEL
M.musculus_7242 GALSlyTALT-CPH---PLAGIVALSCWLPLHRNFPQAANG-----SAKDLAILQCHGEL
P.norvegicus_13 GALSlyTALT-CPH---PLAGIVALSCWLPLHRNFPQAANG-----SAKDLAILQCHGEL
T.nigroviridis_ GALSlyTALT-TQQ---KLAGVVALSCWLPLRKSFPQASAN-----SANKDLHVLQCHGDA
X.laevis_AAH733 GALSlyTALT-TQQ---KLGGVVALSCWLPLRSSFPQAAAN-----SANKDVAVLQCHGES
X.tropicalis_AA GALSlyTALT-TQQ---KLAGVVALSCWLPLRSSFPQAAAN-----SANKDVAVLQCHGES
G.gallus_507376 GALSlyTALT-THQ---KLAGVVALSCWLPLRFSFVQAVG-----VNKEIPVLQCHGDC
M.musculus_P978 GALSlyTALT-TQQ---KLAGVTALSCWLPLRASFPQGPIN-----SANRDISVLQCHGDC
O.cuniculus_077 GALSlyTALT-TQQ---KLAGVTALSCWLPLRASFSQGPIN-----SANRDISVLQCHGDC
R.norvegicus_P7 GALSlyTALT-TQQ---KLAGVTALSCWLPLRASFSQGPIN-----SANRDISVLQCHGDC
B.taurus_777363 GALSlyTALT-TQQ---KLAGVTALSCWLPLRASFPQGPIG---GVNRDISILQCHGDL
O.cuniculus_AAC GALSlyTALT-TQQ---KLAEVVALSCWLPLRASFPQGPIS---GANRDISILQCHGDC
H.sapiens_07560 GALSlyTALT-TQQ---KLAGVTALSCWLPLRASFPQGPIG---GANRDISILQCHGDC
P.pygmaeus_CAH9 GALSlyTALT-TQQ---KLAGVTALSCWLPLRASFPQGPIG---GANRDISILQCHGDC
D.melanogaster_ GALALYSALT-YDQ---PLAGVVALSCWLPLHKQFPQAKVN-----SDDVPIFQAHGDY
A.gambiae_31239 GALAIYAGLT-FAE---PLAGVVALSCWLPLHKSFPVSRKC-----PDTVPVLQCHGDC
A.mellifera_481 GALAIYSALT-FPE---PLAGIIVALSAWLPLHQKFPABEIG-----NKNTPLLQCHGDC
U.maydis_710031 GAISLLTGLT-NPT---PVAGVAALSTWVPLRAKIATLRTP-----TSKTLKVFQAHGDA
F.neoformans_EA GAISVLNMLT-TKR---KLAGVVALSTWVPLNKHIVQMMSE-----HAKDIPVFWGHGTN
E.nidulans_EAA6 GAMSLSFGIT-GQE---KLGGVFGLSCYMLLSDRIKNIPIE---NFPNKKTPFFLAHGE
A.fumigatus_709 AAISVFTGIT-CKE---KLGGVFGLSSYLVLSDKLNKIPIE---NWPNKKTPFFLAHGLE
A.niger_1395827 GAMSLSFAGLT-STE---KLGGVFGLSCYLLHSDRIKNIPIE---DWPNKKTPFFLAHGETD
A.oryzae_BAE627 GAMSVFTGVT-NKE---KLGGVFGLSCYLLHSDRIKNIPIE---DWPNKKTPFFLAHGETD
N.crassa_CAC100 GAMSLFAGLT-AK---KLAGIIVALSSYLLSLKFAELVPK---PEFNKETPIFMAHGDA
G.zeae_EAA74608 GAMSLLAGLT-CTS---KLGIVGLSSWLLSKTFADMVVK---TDANRQTPVMMFHGEE
M.grisea_EAA475 CVMSIFSGLT-AKV---KLAIVVAMSAVPLSLKFKELVAS---CEANKATPIWMHGTT
#####

```

```

          490          500          510          520          530          540
=====+=====+=====+=====+=====+=====+
D.virilis AAW4 DYSEQAKNANTNTGDII-----VIDL-----GSTLT-NYKRYAMLDIEKTGAKIGKW
D.montana ABU80 DYSEQAKNANTNTGDII-----VIDL-----GSTLT-NYKRYAMLDIEKTGAKIGKW
D.americana AB DYSEQAKNANTNTGDII-----VIDL-----GSTLT-NYKRYAMLDIEKTGAKIGKW
D.ezoana AAW476 DYSEQAKNANTNTGDII-----VIDL-----GSTLT-NYKRYAMLDIEKTGAKIGKW
D.littoralis AB DYSEQAKNANTNTGDII-----VIDL-----GSTLT-NYKRYAMLDIEKTGAKIGKW
TpPlip/Cest1 DDKVLFQKQKHGVDMLK-----SHGV-----DVTDR-SYP-MGHESDYAELEAMAEF
O.sativa_509295 GSLIPIKEGRDGIKFLR-----GLGM-----SCEFK-VYDRLGHSLEYEYELDYCQRW
O.sativa_CAE028 DGVVLFQKQKHGVDMLK-----EIGM-----HCEFKVAYPALGHTLVDEELQYFRQW
O.sativa_CAE028 DNVVLFQKQKHGVDMLK-----NAGF-----SCEFK-AYPGLGHSISKEELYSLESW
A.thaliana_1523 DKTVLFQKQKHGVDMLK-----QAGV-----TCEFK-V-----
A.thaliana_BAC4 DKTVLFQKQKHGVDMLK-----QAGV-----TCEFK-AYPGLGHSISNEELQYLESW
P.aeruginosa_A8 DDVVDPALGRAAHDAALQ-----AQGV-----EVGWH-DYP-MGHEVSLEEIHDIHIGAW
P.fluorescens_Q DEVVQFAGQAGPPFLQ-----GRGV-----TVTWQ-EYP-MGHEVLPQEIHDIGAW
P.putida_AAN669 DSVVIPSMPGRATAFEYLN-----TWGV-----AARWY-EYP-MEHEVVVEELNDIHDW
IgEstl_CAO02582 DQTVQKQKHGVDMLK-----ELGIR-----TYELK-EYSPLGHSASQOEIADVLAW
PtPlip/Cest DPLVQFAGQAGPPFLQ-----EKGQ-----QYELK-SYP-IPHTVSPAELNDVLFK
TpPlip/Cest2 DPLVQFAGQAGPPFLQ-----EKGAT-----SYKLE-IYEGLAHSANPKEIGDVMFAF
D.discoideum_EA DQVVRHQWGLSFDHLK-----SQGI-----NGEFI-TLKGLGHHSSPEEIDLMTKF
C.albicans_6847 DPVINVDYQKQKHGVDMLK-----QLGFK-----NLKFN-TYEGVAHSASEEELADVLFK
D.hansenii_5042 DPIVNFDFGKQTSSELYQ-----KLGFK-----NVKFN-TYEGVAHSASEEELADAMNF
Y.lipolytica_Q DQVVRHQWGLSFDHLK-----NKLK-----NVDWH-QYEGVTHSCGFEEISDILNW
S.japonicum_AAD DCMIPFAMGKLTHELLK-----NFQLS-----KCELK-CYPDLSSHSSCEQEMEDLRTF
S.mansoni_AAC62 DYTIPFAMGKLTHELLK-----TFQLS-----KCELN-CYPQLSSHSSCEKEMGDLRTF
C.briggsae_CAE6 DFLVPLQKQKHGVDMLK-----QFNP-----NVEMH-VYRGMQHSSCSEEMRDLKTF
C.elegans_T232 DFLVPLQKQKHGVDMLK-----KFN-----KVELH-TYRGMQHSSCSEEMRDLKTF
S.purpuratus_72 -----KVELH-TYRGMQHSSCSEEMRDLKTF-----KYAG-----RAKLW
T.nigroviridis_ DFMIPLRFGDMTSSKLIQ-----SIVDPQ-----MVAFK-SYAGVPHGSCPCQEMADVKEF
D. rerio_4115218 DPMIPVQFAGMATAEKLK-----TIVSPE-----NITFR-TYPLMHSSCPQEMSAVKDF
S.tropicalis_AA DPMIPVRFNGLTSEKLN-----SVLNPS-----KVQFK-SYPGMHSTNQEMMAVKDF
X.tropicalis_45 DPMIPVRFNGLTSEKLN-----SVLNPS-----KVQFK-SYPGMHSTNQEMMAVKDF
X.laevis_AA443 DPMIPVRFNGLTSEKLN-----SVLNPS-----KVQFK-SYPGMHSTNQEMMAVKDF
G.gallus_507598 DPMIPVRFNGLTSEKLN-----SVVTPA-----KVQFK-TYPGVMHNSCPQEMMAVKDF
P.trogodytes_5 DPMVP-----RTVWG-----
C.familiaris_73 DPMVPVRFNGLTSEKLN-----SVVTPA-----RVQFK-TYPGVMHNSCPQEMAAVKDF
H.sapiens_AAC72 DPMVPVRFNGLTSEKLN-----SVVTPA-----RVQFK-TYPGVMHNSCPQEMAAVKDF
M.musculus_7242 DPMVPVRFNGLTSEKLN-----TVVTPA-----RVQFK-TYPGVMHNSCPQEMAAVKDF
R.norvegicus_13 DPMVPVRFNGLTSEKLN-----TVVTPA-----RVQFK-TYPGVMHNSCPQEMAAVKDF
T.nigroviridis_ DPIVFPVFGTQTAEKMK-----SLVNPS-----HMSFK-TYRGLCHSACPEEMVDIKRF
X.laevis_AA733 DPLVPLMFGSLTSEKLN-----TIISPA-----NVKFK-TYSGLMHSSCNQEMTDIKQF
X.tropicalis_AA DPLVPLMFGSLTSEKLN-----TIISPA-----NINFK-TYSGLMHSSCNQEMTDIKQF
G.gallus_507376 DPLVPLMFGSLTSEKLN-----SMINPA-----NITFR-TYSGMMHSSCIEMMDIKQF
M.musculus_P978 DPLVPLMFGSLTSEKLN-----ALINPA-----NVTFK-IYEGMMHSSCQEMMDVKHF
O.cuniculus_077 DPLVPLMFGSLTSEKLN-----ALINPA-----NVTFK-IYEGMMHSSCQEMMDVKHF
R.norvegicus_P7 DPLVPLMFGSLTSEKLN-----GLVNPA-----NVTFK-VYEGMMHSSCQEMMDVKYF
B.taurus_777363 DPLVPLMFGSLTSEKLN-----TLVNPA-----NVTFR-TYAGMMHSSCQEMMDIKQF
O.cuniculus_AAC DPLVPLMFGSLTSEKLN-----TLVNPA-----NVTFK-TYEGMMHSSCQEMMDVKQF
H.sapiens_07560 DPLVPLMFGSLTSEKLN-----TLVNPA-----NVTFK-TYEGMMHSSCQEMMDVKQF
P.pygmaeus_CAH9 DPLVPLMFGSLTSEKLN-----TLVNPA-----NVTFK-TYEGMMHSSCQEMMDVKQF
D.melanogaster_ DPVVPYKFGQLSASLLK-----SFMK-----NVTFK-TYSGLSHSSDDEMDVVDI
A.gambiae_31239 DPIVFPVFGTQTAEKMK-----SFMK-----NSHFQ-TYQGLGHSACDAELSDMKKF
A.mellifera_481 DPIVFPVFGTQTAEKMK-----QFMT-----QTEFK-TYRGMHNSCPQEMRDMKKF
U.maydis_710031 DPVVPYKFGQLSASLLK-----ELALNDK-----DVEFH-TYRPMHNSCPQEMRDLAAF
F.neoformans_EA DPVVDYRFGQKQKHGVDMLK-----ELGLLE-----DVTFN-SYKLSHNSADPVEIEDLEKF
E.nidulans_EAA6 DDIVPHEFGKRSAAEMAK-----ELGLLE-----DVTFN-SYKLSHNSADPVEIEDLEKF
A.fumigatus_709 DEIVLDFDGLSAAKMK-----EIGLE-----DVTFK-SYPNLGHSADPVEIEDLARF
A.niger_1395827 DEVVLFQKQKHGVDMLK-----ELGVE-----DVEFH-SYSDLGHSADPVEIEDLTKF
A.oryzae_BAE627 DDVVKYEFGKTSKLLQ-----DLGLE-----NVQFN-SYSDLGHSADPVEIEDLEKF
N.crassa_CAC100 DPVVNYKLGTMTRDLLK-----EMGY-----NVKFT-TYPGMGHSACLEELDAIEDF
G.zeae_EAA74608 DPIVPCERGLSAELK-----GLGY-----DVAWK-TYPGMGHSADPVEIEDLEAF
M.grisea_EAA475 DLVVPVTLGMMSEALLK-----DEGY-----QVSMK-LYPGMGHSACPEELDEVEAF
#####

```

```

                    550      560      570      580      590      600
=====+=====+=====+=====+=====+=====+
D.virilis AAW4 IVELTSELQLPHETIHMIGQNVGAHVAGAAAANEFTRLTGHKL-----
D.montana ABU80 IVELTSQLQLPHETIHMIGQNVGAHVAGAAAANEFTRLTGHKL-----
D.americana AB IVELTSELQLPHETIHMIGQNVGAHVAGAAAANEFTRLTGHKL-----
D.ezoana AAW476 IVELTSELQLPHETIHMIGQNVGAHVAGAAAANEFTRLTGHKL-----
D.littoralis AB IMELTSELQLPHETIHMIGQNVGAHVAGAAAANEFTRLTGHKL-----
TpPlip/CEst1 CEKL-----
O.sativa_509295 VEKILHRSGREGLIRRVSRNIFLCSNLFNSS-----
O.sativa_CAE028 IKDRLSQGTGVPVPSLSKMDLQ-----
O.sativa_CAE028 IKNHLKASQEKEN-----
A.thaliana_1523 -----
A.thaliana_BAC4 LKQRMQSSSSSS-----
P.aeruginosa_A8 LRKRL-----
P.fluorescens_Q LAERLR-----
P.putida_AAN669 LSKQLQ-----
IgEstl_CAO02582 LQARLPEDA-----
PtPlip/CEst LQKQLPPDDTCKVKLKDPSSEMSVKELRAAVRQAGLQSKAVGFMEKSEFVQLLETHRNGKL
TpPlip/CEst2 LEEVLP-----
D.discoideum_EA ISKTLPK-----
C.albicans_6847 IKNIVEK-----
D.hansenii_5042 IDDLVKK-----
Y.lipolytica_50 LEENIKE-----
S.japonicum_AAD LARNIPGTQ-----
S.mansoni_AAC62 LSKNIPGTQ-----
C.briggsae_CAE6 LSNHIAK-----
C.elegans_T2332 LSAHIAH-----
S.purpuratus_72 NGLAIPV-----
T.nigroviridis_ IEKYLPR-----
D.rerio_4115218 IEKQLPRV-----
S.tropicalis_AA LEKVLPRV-----
X.tropicalis_45 LEKVLPRV-----
X.laevis_AAH443 LQKVLPRM-----
G.gallus_507598 IEKLLPRI-----
P.troglodytes_5 -----PDG-----
C.familiaris_73 LEKLLPPV-----
H.sapiens_AAC72 LEKLLPPV-----
M.musculus_7242 LEKLLPPV-----
R.norvegicus_13 LEKLLPPV-----
T.nigroviridis_ IEKQLPPL-----
X.laevis_AAH733 IDKQLPPVN-----
X.tropicalis_AA IDKQLPPVN-----
G.gallus_507376 IDKHLPPVD-----
M.musculus_P978 IDKLLPPID-----
O.cuniculus_077 IDKLLPPID-----
R.norvegicus_P7 IDKLLPPID-----
B.taurus_777363 IDKLLPPVD-----
O.cuniculus_AAC IDKLLPPID-----
H.sapiens_07560 IDKLLPPID-----
P.pygmaeus_CAH9 IDKLLPPID-----
D.melanogaster_ ISKWVN-----
A.gambiae_31239 IDDHVPRQ-----
A.mellifera_481 IEKVLKP-----
U.maydis_710031 LEKVIPAQ-----
F.neoformans_EA LMEALK-----
E.nidulans_EAA6 LDRVIPAENGGSL-----
A.fumigatus_709 LQKVIPPEDDQASAGL-----
A.niger_1395827 LQKAIPPEEGQSAAGL-----
A.oryzae_BAE627 LQQVIPAEGEVSAGL-----
N.crassa_CAC100 LTERLPKVADKSEKSEL-----
G.zeae_EAA74608 LRKQLPPKN-----
M.grisea_EAA475 LRKSLPPLGNTGSGVLHQSTSRNRPHIREEECDPGGKLAPRNLGKY-----
#####

```

Parameters used

```

Minimum Number Of Sequences For A Conserved Position: 31
Minimum Number Of Sequences For A Flanking Position: 31
Maximum Number Of Contiguous Nonconserved Positions: 8
Minimum Length Of A Block: 5
Allowed Gap Positions: With Half
Use Similarity Matrices: Yes

```

Flank positions of the 9 selected block(s)

```

Flanks: [249 254] [288 292] [294 306] [331 359] [367 391] [407 430] [438 455]
[469 497] [522 546]

```

Annexe 12 : Alignement des séquences génomique et d'ADNc d'*IgEst2*

CLUSTAL W 2.0 multiple sequence alignment

```

IgEst2ADN      CCTCGGCCGCCGCCGAGGGCCAGTCGCAGGGCTTGTACCTCAATGAGGTACGCCGTGCG 60
IgEst2ADNc    CCTCGGCCGCCGCCGAGGGCCAGTCGCAGGGCTTGTACCTCAATGAGGTACGCCGTGCG 60
*****
IgEst2ADN      GGCCGCCACTACGTCTTTTCGGCGCAGCGCCCGCGCTGATCCTGCTGCTGCATGGCACTGG 120
IgEst2ADNc    GGCCGCCACTACGTCTTTTCGGCGCAGCGCCCGCGCTGATCCTGCTGCTGCATGGCACTGG 120
*****
IgEst2ADN      GGACAACGAGCAAGGCCTTCTCGGCGCGATAGGGCAGCTTGCGCCGCCTGACGCCGTTGT 180
IgEst2ADNc    GGACAACGAGCAAGGCCTTCTCGGCGCGATAGGGCAGCTTGCGCCGCCTGACGCCGTTGT 180
*****
IgEst2ADN      CGTCTCGTTGCGCGGCCCTCTGCAAGCACCATTGGAGGCTTTCGTTGGTTCGAGGGGTA 240
IgEst2ADNc    CGTCTCGTTGCGCGGCCCTCTGCAAGCACCATTGGAGGCTTTCGTTGGTTCGAGGGGTA 240
*****
IgEst2ADN      TTCTTCGGCCCCAGAGCAGATTGCGCTCGACAAGACCGTCGGTTCGTCGTCAGATGCTGT 300
IgEst2ADNc    TTCTTCGGCCCCAGAGCAGATTGCGCTCGACAAGACCGTCGGTTCGTCGTCAGATGCTGT 300
*****
IgEst2ADN      ACTTGCATTTATTGAGGAAGCGCCTGCCAAGTTTGGGATCGATCCGAAGCGAGTCTTCCT 360
IgEst2ADNc    ACTTGCATTTATTGAGGAAGCGCCTGCCAAGTTTGGGATCGATCCGAAGCGAGTCTTCCT 360
*****
IgEst2ADN      GCTGGGATTTTCGAGGGCGCCACACTCGTTTGGACGACCCTCTTGTCCAGATGGTCGCG 420
IgEst2ADNc    GCTGGGATTTTCGAGGGCGCCACACTCGTTTGGACGACCCTCTTGTCCAGATGGTCGCG 420
*****
IgEst2ADN      CCCCGGGCTCATCGCCGGCGGGCTAACACTGTCTGGCCGGCTGTTTCCTGAGCTAATGCA 480
IgEst2ADNc    CCCCGGGCTCATCGCCGGCGGGCTAACACTGTCTGGCCGGCTGTTTCCTGAGCTAATGCA 480
*****
IgEst2ADN      GGCTGGTACCCCACTTTATGCGCGTCTCGCCGACCCGACGCAGCTGAGGGATTGTGCCGT 540
IgEst2ADNc    GGCTGGTACCCCACTTTATGCGCGTCTCGCCGACCCGACGCAGCTGAGGGATTGTGCCGT 540
*****
IgEst2ADN      CTTTGCCTCACACGCGCCCAAGACTCGGTTACCCCGGTAGCTATCGGGCGCTCGAATGA 600
IgEst2ADNc    CTTTGCCTCACACGCGCCCAAGACTCGGTTACCCCGGTAGCTATCGGGCGCTCGAATGA 600
*****
IgEst2ADN      GCGGCTGTTTCGAGACTGGGCACCGGGTGCAGACTTAATATATAAGGAGGACCCTGTTGC 660
IgEst2ADNc    GCGGCTGTTTCGAGACTGGGCACCGGGTGCAGACTTAATATATAAGGAGGACCCTGTTGC 660
*****
IgEst2ADN      AGCCATGAGATATCGAAGCCCGTGGCTGCTTCAATGTCCCAATGGTTCCGTCGGTATGT 720
IgEst2ADNc    AGCCATGAGATATCGAAGCCCGTGGCTGCTTCAATGTCCCAATGGTTCCGTCGGTATGT 720
*****
IgEst2ADN      CTCCGGAGCATCAAAGGGTGATCAGTAGGA 750
IgEst2ADNc    CTCCGGAGCATCAAAGGGTGATCAGTAGGA 750
*****

```

Annexe 13 : Analyse du plasmide recombiné pYES2/IgLip (clone 1) et séquençage de l'insert

CLUSTAL W 2.0 multiple sequence alignment

Les mutations sont indiquées en rouge sur fond gris.

```

Iglip      -----ATGCGCAGGCGGCGGCGGCGAAAGAGGCTGTC 32
Clone1    GTAACGGCCGCCAGTGTGCTGGAATTCCATGCGCAGGCGGCGGCGGCGAAAGAGGCTGTC 60
          *****

Iglip      CTGGCGAGCCACGCCATGGCGCAGCGATTTGGATTCCGACCTGCGTACTTTCGAAACGTG 92
Clone1    CTGGCGAGCCACGCCATGGCGCAGCGATTTGGATTCCGACCTGCGTACTTTCGAAACGTG 119
          *****

Iglip      GCGCTCCTGCCGGCCGGCAGACAATGAGGCGTTGATCGCCGAGCGTCTCCGCTTGGCCCG 152
Clone1    GCGCTCCTGCCGGCCGGCAGACAATGAGGCGTTGATCGCCGAGCGTCTCCGCTTGGCCCG 179
          *****

Iglip      CCGGTGTGGGCCACCGTCAATGCGCGCTGCTGCTCGAGCTCGCGGCGGCTCTGCTGTG 212
Clone1    CCGGTGTGGGCCACCGTCAATGCGCGCTGCTGCTCGAGCTCGCGGCGGCTCTGCTGTG 239
          *****

Iglip      CCCACCACAAATGGTTCGCTGGACGTGTTACGCTGCGGAGGGACTGGCTTGCAGCC 272
Clone1    CCCACCACAAATGGTTCGCTGGACGTGTTACGCTGCGGAGGGACTGGCTTGCAGCC 299
          *****

Iglip      CGATCCGCAAGTCATCATCGCAGACGGCGGCTCCTTCCTCGCGCTAGTGGGTATCAATGA 332
Clone1    CGATCCGCAAGTCATCATCGCAGACGGCGGCTCCTTCCTCGCGCTAGTGGGTATCAATGA 359
          *****

Iglip      AGCCGAACAGTGGGTGGTGGTTCGCGGCTCTCTCGACGGCAACCTAGGGGCTTG 392
Clone1    AGCCGAACAGTGGGTGGTGGTTCGCGGCTCTCTCGACGGCAACCTAGGGGCTTG 419
          *****

Iglip      GGGTCGCAACTTGCAATGCTGGACCGACCCGAGGAGCTTCCACAATTGCTCCGGCTGGGT 452
Clone1    GGGTCGCAACTTGCAATGCTGGACCGACCCGAGGAGCTTCCACAATTGCTCCGGCTGGGT 479
          *****

Iglip      GCACCATGGCTGGGGTAGTGGCTACGACATGCTTCGCGCGCGCCTCATCACGGCTGTGCT 512
Clone1    GCACCATGGCTGGGGTAGTGGCTACGACATGCTTCGCGCGCGCCTCATCACGGCTGTGCT 539
          *****

Iglip      GCTAGCGCTTGAGGCTCGGCCGACGTACCGCCTCGTCGTCACCGGCCACTCTGCCGGCGG 572
Clone1    GCTAGCGCTTGAGGCTCGGCCGACGTACCGCCTCGTCGTCACCGGCCACTCTGCCGGCGG 599
          *****

Iglip      GGCCTGGCCTCCATCTTTGTGGCCGAGCTCGCCACCGTTCGGAGCAGCTCAATCATCCTC 632
Clone1    GGCCTGGCCTCCATCTTTGTGGCCGAGCTCGCCACCGTTCGGAGCAGCTCAATCATCCTC 659
          *****

Iglip      GGTAGCCAGCAATCGTCCGCGTACCGCGCCGAGGCCGGCTCTGACTCCAAAGGCGAAGC 692
Clone1    GGTAGCCAGCAATCGTCCGCGTACCGCGCCGAGGCCGGCTCTGACTCCAAAGGCGAAGC 719
          *****

Iglip      TGACACGAAGGCTGCCGCGGCTGATGAGTCCGCCCGCCCTCCAGCTCCATCGCCTGGGA 752
Clone1    TGACACGAAGGCTGCCGCGGCTGATGAGTCCGCCCGCCCTCCAGCTCCATCGCCTGGGA 779
          *****

Iglip      GAGTGTTCGAGGTTGCGCCGCTGCTCAGGCGACCGCTCGCGCCGCCCGCTGTACAT 812
Clone1    GAGTGTTCGAGGTTGCGCCGCTGCTCAGGCGACCGCTCGCGCCGCCCGCTGTACAT 839
          *****

Iglip      CACCTTCGCCGCGCCGATGGCCGGGAACCGCGAGTTTTCGGCCGCCATCGATTCCGGCCT 872
Clone1    CACCTTCGCCGCGCCGATGGCCGGGAACCGCGAGTTTTCGGCCGCCATCGATTCCGGCCT 899
          *****

Iglip      GGCCGCGCAGGGCAACGTGATGCTACGTGTGGTCAATGCACGCGATCCGGCCGTGCGTGC 932
Clone1    GGCCGCGCAGGGCAACGTGATGCTACGTGTGGTCAATGCACGCGATCCGGCCGTGCGTGC 959
          *****

Iglip      ACCCAACACGCTGAATGCGGGTGGCAGCGTCCAGATCCCCTGCCCAATGGGGGGCGAGGA 992
Clone1    ACCCAACACGCTGAATGCGGGTGGCAGCGTCCAGATCCCCTGCCCAATGGGGGGCGAGGA 1019
          *****

Iglip      CCAGTTTTCACATGCGGGGCGAGGAAGTGTGGCTGACAAACGCTGCCGACGCCGACATGAG 1052
Clone1    CCAGTTTTCACATGCGGGGCGAGGAAGTGTGGCTGACAAACGCTGCCGACGCCGACATGAG 1079
          *****

Iglip      CAAGGCGCAATTCTGCAACAGCTCAGGCGGTTTCGGGGACGGCGATGACGCCAGCATGTC 1112
Clone1    CAAGGCGCAATTCTGCAACAGCTCAGGCGGTTTCGGGGACGGCGATGACGCCAGCATGTC 1139
          *****

Iglip      CGACCCGAACCCGCTGTGCTCCAGCTCCCTCGGCGAGTCGAGCCTGACTTGGGTCCACCA 1172
Clone1    CGACCCGAACCCGCTGTGCTCCAGCTCCCTCGGCGAGTCGAGCCTGACTTGGGTCCACCA 1199
          *****

```

```
IgIip      CAACTCATACTCGGCGTCGACACAGGCGCGTGCCCCGATACGATGGATGCGAATCAGTG 1232
Clone1    CAACTCATACTCGGCGTCGACACAGGCGCGTGCCCCGATACGATGGATGCGAATCAGTG 1259
          *****
IgIip      GGAAGAGTACATCAAGTATACCAAGAGTCTCCACCCGAGTTGAGACACGTCGGCTATGG 1292
Clone1    GGAAGAGTACATCAAGTATACCAAGAGTCTCCACCCGAGTTGAGACACGTCGGCTATGG 1319
          *****
IgIip      TGAGGACGCACCCGACGGCGGTTTCGACAGCACGGTGCAGCGGTCCGAGGCGGAGCGGGT 1352
Clone1    TGAGGACGCACCCGACGGCGGTTTCGACAGCACGGTGCAGCGGTCCGAGGCGGAGCGGGT 1379
          *****
IgIip      GCCGACATACTCGTACGGATGA----- 1374
Clone1    GCCGACATACTCGTACGGATGAGCTCTAGAGGGCCGCATCATGTAATTAGTTATGTCACG 1439
          *****
```

Annexe 14 : Analyse du plasmide recombiné pYES2/IgEst1 (clone A) et séquençage de l'insert

CLUSTAL W 2.0 multiple sequence alignment
Les mutations sont indiquées en rouge sur fond gris.

```

CloneA   TAGTAACGGCCGCCAGTGTGCTGGAATTCCATGGCGCGAGGGCTTGCCGCGTGACACGA 60
IgEst1   -----ATGGCGCGAGGGCTTGCCGCGTGACACGA 30
          *****

CloneA   CTTGCAGCGGGCGGAGCGGGCGGTCTCTGCTGTTGCTTCCAGCCGCTGCAATCTCGTGTC 120
IgEst1   CTTGCAGCGGGCGGAGCGGGCGGTCTCTGCTGTTGCTTCCAGCCGCTGCAATCTCGTGTC 90
          *****

CloneA   TCCGCCGTCAGCATGTGACCGCTGCCGGTGCGGCGCCGCCGCGCATCACGCAGGACCGC 180
IgEst1   TCCGCCGTCAGCATGTGACCGCTGCCGGTGCGGCGCCGCCGCGCATCACGCAGGACCGC 150
          *****

CloneA   GCGCGCAACAACCTGATCACCATCGCGCCCGGGGAAGGCATGCCGCACACCCGCGACGGTC 240
IgEst1   GCGCGCAACAACCTGATCACCATCGCGCCCGGGGAAGGCATGCCGCACACCCGCGACGGTC 210
          *****

CloneA   ATCGGCCCCATCCACGGGCTGGGCGACAGCAACATGGGCTGGGCGGACGTGGCGATGCAG 300
IgEst1   ATCGGCCCCATCCACGGGCTGGGCGACAGCAACATGGGCTGGGCGGACGTGGCGATGCAG 270
          *****

CloneA   CTGCAGAGTGTGTCATGCCGTATTGCAAGTTCATCCTCCCAAATGCGCCGTCCGTCCTGTC 360
IgEst1   CTGCAGAGTGTGTCATGCCGTATTGCAAGTTCATCCTCCCCAATGCGCCGTCCGTCCTGTC 330
          *****

CloneA   ACCCTCAACGGCGGAATGTGCGATGCCAGTTGGTATGACATCACGAGTTTGGATAAGCGC 420
IgEst1   ACCCTCAACGGCGGAATGTGCGATGCCAGTTGGTATGACATCACGAGTTTGGATAAGCGC 390
          *****

CloneA   GAGTCCCAGCCGTGCACGGGCATCGAAGAGTCTCGCCAGGCGATGCTCGACCTCATCTCC 480
IgEst1   GAGTCCCAGCCGTGCACGGGCATCGAAGAGTCTCGCCAGGCGATGCTCGACCTCATCTCC 450
          *****

CloneA   GCTGAGGTTGCGTCCGGCATCCCCCCTCCCGCATCGCCATCGCCGGCTTCTCCAGGGA 540
IgEst1   GCTGAGGTTGCGTCCGGCATCCCCCCTCCCGCATCGCCATCGCCGGCTTCTCCAGGGA 510
          *****

CloneA   GGCGCCGTGCGCCTCTTACAGGGCTGCAGTATTACACACCCTGGCCGGCGTGCTATGC 600
IgEst1   GGCGCCGTGCGCCTCTTACAGGGCTGCAGTATTACACACCCTGGCCGGCGTGCTATGC 570
          *****

CloneA   TTGTCCGGCTACTTGGCGGGCGGAGGAGCGTTTCATACTGGCGCCGGAGGCTGTGAACACT 660
IgEst1   TTGTCCGGCTACTTGGCGGGCGGAGGAGCGTTTCATACTGGCGCCGGAGGCTGTGAACACT 630
          *****

CloneA   CCCGTCGCTCACTTTCACGGCTCGGATGATCAAACGGTCCAAATCAAATGGGCCCGCGGC 720
IgEst1   CCCGTCGCTCACTTTCACGGCTCGGATGATCAAACGGTCCAAATCAAATGGGCCCGCGGC 690
          *****

CloneA   TCTCAGGCCACCTGCGCGAGCTGGGCATCCGCACCTACGAGCTAAAAGAGTACTCGCCC 780
IgEst1   TCTCAGGCCACCTGCGCGAGCTGGGCATCCGCACCTACGAGCTAAAAGAGTACTCGCCC 750
          *****

CloneA   CTCGGCCACAGCGCCTCGCAGCAGGAGATCGCGGACGTGCTTGCATGGCTTCAGGCCAGA 840
IgEst1   CTCGGCCACAGCGCCTCGCAGCAGGAGATCGCGGACGTGCTTGCATGGCTTCAGGCCAGA 810
          *****

CloneA   CTGCCAG-GGACGCCTGAGCTCTAGAGGGCCGCATCATGTAATTAGTTATGTC 892
IgEst1   CTGCCAGAGGACGCCTGA----- 828
          *****

```

Annexe 15 : Analyse du plasmide recombiné pYES2/IgEst2 (clone α) et séquençage de l'insert

CLUSTAL W 2.0 multiple sequence alignment

```

Clone  $\alpha$  NNNNNNNNNNNNNNTCGGNTCNCTAGTAACGGCCGCCAGTGTGCTGGAATTCCATGAGGTA 60
IgEst2  -----ATGAGGTA 8
                *****

Clone  $\alpha$  CGCCGTGCGGGCCGCCACTACGTCTTTTCGGCGCAGCGCCCGCGCTGATCCTGCTGCTGCA 120
IgEst2  CGCCGTGCGGGCCGCCACTACGTCTTTTCGGCGCAGCGCCCGCGCTGATCCTGCTGCTGCA 68
                *****

Clone  $\alpha$  TGGCACTGGGGACAACGAGCAAGGCCTTCTCGGCGCGATAGGGCAGCTTGCGCCGCCTGA 180
IgEst2  TGGCACTGGGGACAACGAGCAAGGCCTTCTCGGCGCGATAGGGCAGCTTGCGCCGCCTGA 128
                *****

Clone  $\alpha$  CGCCGTTGTCGTCTCGTTGCGCGGCCCTCTGCAAGCACCATTGAGGCTTTCGTTGGTT 240
IgEst2  CGCCGTTGTCGTCTCGTTGCGCGGCCCTCTGCAAGCACCATTGAGGCTTTCGTTGGTT 188
                *****

Clone  $\alpha$  CGAGGGGTATTCTTCGGCCCCAGAGCAGATTGCGCTCGACAAGACCGTCGGTTCGTCGTC 300
IgEst2  CGAGGGGTATTCTTCGGCCCCAGAGCAGATTGCGCTCGACAAGACCGTCGGTTCGTCGTC 248
                *****

Clone  $\alpha$  AGATGCTGTACTTGCATTTATTGAGGAAGCGCCTGCCAAGTTTGGGATCGATCCGAAGCG 360
IgEst2  AGATGCTGTACTTGCATTTATTGAGGAAGCGCCTGCCAAGTTTGGGATCGATCCGAAGCG 308
                *****

Clone  $\alpha$  AGTCTTCCTGCTGGGATTTTCGCAGGGCGCCACACTCGTTTGGACGACCCCTTGTCCAG 420
IgEst2  AGTCTTCCTGCTGGGATTTTCGCAGGGCGCCACACTCGTTTGGACGACCCCTTGTCCAG 368
                *****

Clone  $\alpha$  ATGGTCGCGCCCCGGGCTCATCGCCGGCGGGCTAACACTGTCTGGCCGGCTGTTTCCTGA 480
IgEst2  ATGGTCGCGCCCCGGGCTCATCGCCGGCGGGCTAACACTGTCTGGCCGGCTGTTTCCTGA 428
                *****

Clone  $\alpha$  GCTAATGCAGGCTGGTACCCCACTTTATGCGCGTCTCGCCGACCCGACGCAGCTGAGGGA 540
IgEst2  GCTAATGCAGGCTGGTACCCCACTTTATGCGCGTCTCGCCGACCCGACGCAGCTGAGGGA 488
                *****

Clone  $\alpha$  TTGTGCCGTCTTTGCCTCACACGGCGCCCAAGACTCGGTTACCCCGGTAGCTATCGGGCG 600
IgEst2  TTGTGCCGTCTTTGCCTCACACGGCGCCCAAGACTCGGTTACCCCGGTAGCTATCGGGCG 548
                *****

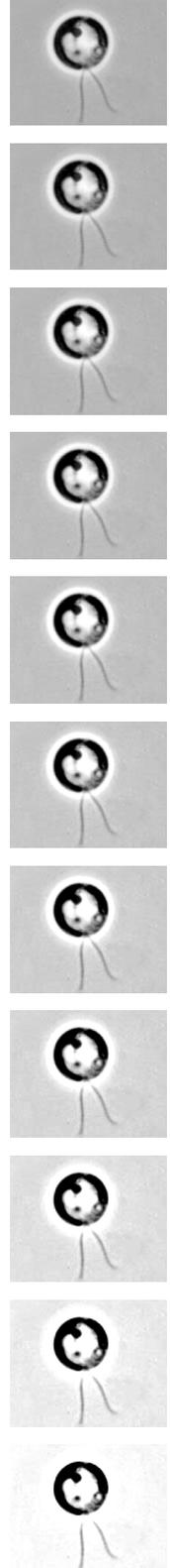
Clone  $\alpha$  CTCGAATGAGCGGCTGTTTCGAGACTGGGCACCGGGTGC GGACTTAATATATAAGGAGGA 660
IgEst2  CTCGAATGAGCGGCTGTTTCGAGACTGGGCACCGGGTGC GGACTTAATATATAAGGAGGA 608
                *****

Clone  $\alpha$  CCCTGTTGCAGGCCATGAGATATCGAAGGCCGTGGCTGCTTCAATGTCCAATGGTTCCG 720
IgEst2  CCCTGTTGCAGGCCATGAGATATCGAAGGCCGTGGCTGCTTCAATGTCCAATGGTTCCG 668
                *****

Clone  $\alpha$  TCGGTATGTCTCCGGAGCATCAAAGGGTGATCAGTAGGCTCTAGAGGGCCGCATCATGTA 780
IgEst2  TCGGTATGTCTCCGGAGCATCAAAGGGTGATCAGTAG----- 705
                *****

```

BIBLIOGRAPHIE



- Abousalham A. et Verger R.** (2000). Zymogram of pancreatic lipases. *Analytical Biochemistry*, 281(2): 234-6.
- Ackermann E.J. et Dennis E.A.** (1995). Mammalian calcium-independent phospholipase A₂. *Biochimica et Biophysica Acta*, 1259(2): 125-36.
- Adolph S., Bach S., Blondel M., Cueff A., Moreau M., Pohnert G., Poulet S.A., Wichard T. et Zuccaro A.** (2004). Cytotoxicity of diatom-derived oxylipins in organisms belonging to different phyla. *Journal of Experimental Biology*, 207: 2935-46.
- Akiba S. et Sato T.** (2004). Cellular function of calcium-independent phospholipase A₂. *Biological & Pharmaceutical Bulletin*, 27(8): 1174-8.
- Akoh C.C., Lee G.C. et Shaw J.F.** (2004). Protein engineering and applications of *Candida rugosa* lipase isoforms. *Lipids*, 39(6): 513-26.
- Alam M., Gilham D., Vance D.E. et Lehner R.** (2006). Mutation of F417 but not of L418 or L420 in the lipid binding domain decreases the activity of triacylglycerol hydrolase. *Journal of Lipid Research*, 47(2): 375-83.
- Alonso D.L., Belarbi E.-H., Rodriguez-Ruiz J., Segura C.I. et Gimenez A.** (1998). Acyl lipids of three microalgae. *Phytochemistry*, 47(8): 1473-81.
- Altschul S.F., Gish W., Miller W., Myers E.W. et Lipman D.J.** (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3): 403-10.
- Andersen R.A.** (2004). Biology and systematics of heterokont and haptophyte algae. *American Journal of Botany*, 91: 1508-22.
- Antia N.J., Bilinski E. et Lau Y.C.** (1970). Identification and characterization of phospholipase D in a unicellular red alga (*Porphyridium cruentum*). *Canadian Journal of biochemistry*, 48: 643-8.
- 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., Kröger N., Lau W.W.Y., 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., Ryneanson T.A., Saito M.A., Schwartz D.C., Thamtracoln 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.
- Arpigny J.L. et Jaeger K.E.** (1999). Bacterial lipolytic enzymes: classification and properties. *Biochemical Journal*, 343 Pt 1: 177-83.
- Asamizu E., Nakamura Y., Sato S., Fukuzawa H. et Tabata S.** (1999). A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. *DNA Research*, 6: 369-73.
- Au-Young J. et Fielding C.J.** (1992). Synthesis and secretion of wild-type and mutant human plasma cholesteryl ester transfer protein in baculovirus-transfected insect cells: the carboxyl-terminal region is required for both lipoprotein binding and catalysis of transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 89(9): 4094-8.
- Bairoch A., Boeckmann B., Ferro S. et Gasteiger E.** (2004). Swiss-Prot: juggling between evolution and stability. *Briefings in Bioinformatics*, 5(1): 39-55.

- Baldauf S.L., Roger A.J., Wenk-Siefert I. et Doolittle W.F.** (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, 290(5493): 972-7.
- Ban S., Burns C., Castel J., Chaudron Y., Christou E., Escribano R., Fonda Umani S., Gasparini S., Guerrero Ruiz F., Hoffmeyer M., Ianora A., Kang H.-K., Laabir M., Lacoste A., Miralto A., Ning X., Poulet S., Rodriguez V., Runge J., Shi J., Starr M., Uye S.-i. et Wang Y.** (1997). The paradox of diatom-copepod interactions. *Marine Ecology Progress Series*, 157: 287-93.
- Bendtsen J.D., Nielsen H., von Heijne G. et Brunak S.** (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*, 340(4): 783-95.
- Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J. et Wheeler D.L.** (2007). GenBank. *Nucleic Acids Research*, 35(Database issue): D21-5.
- Berezziat G.** (1996). [Diversity of phospholipases A₂ and their functions]. *Comptes Rendus des Séances de la Société de Biologie et des ses Filiales*, 190(4): 409-16.
- Bergé J.-P.** (1996). Optimisation de la production en masse de *Skeletonema costatum* (Grev.) Cleve. sur une eau souterraine salée en vue de sa valorisation: caractérisation des lipides et recherche de substances à activité biologique. Thèse de doctorat. *Université de Nantes*. 241.
- Bhat T.N., Bourne P., Feng Z., Gilliland G., Jain S., Ravichandran V., Schneider B., Schneider K., Thanki N., Weissig H., Westbrook J. et Berman H.M.** (2001). The PDB data uniformity project. *Nucleic Acids Research*, 29(1): 214-8.
- Bigogno C., Khozin-Goldberg I., Boussiba S., Vonshak A. et Cohen Z.** (2002a). Lipid and fatty acid composition of the green oleaginous alga *Parietochloris incisa*, the richest plant source of arachidonic acid. *Phytochemistry*, 60(5): 497-503.
- Bigogno C., Khozin-Goldberg I. et Cohen Z.** (2002b). Accumulation of arachidonic acid-rich triacylglycerols in the microalga *Parietochloris incisa* (Trebuxiophyceae, Chlorophyta). *Phytochemistry*, 60(2): 135-43.
- Bilinski E., Antia N.J. et Lau Y.C.** (1968). Characterization of phospholipase C from marine planktonic alga (*Monochrysis lutheri*). *Biochimica and Biophysica Acta*, 159: 496-502.
- Birnboim H.C. et Doly J.** (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6): 1513-23.
- Borodovsky M. et McIninch J.** (1993). GeneMark: parallel gene recognition for both DNA strands. *Computers & Chemistry*, 17(19): 123-33.
- Bligh E.G. et Dyer W.J.** (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37: 911-917.
- Block R.C., Harris W.S., Reid K.J., Sands S.A. et Spertus J.A.** (2007). EPA and DHA in blood cell membranes from acute coronary syndrome patients and controls. *Atherosclerosis*, In Press, Corrected Proof.
- Brady L., Brzozowski A.M., Derewenda Z.S., Dodson E., Dodson G., Tolley S., Turkenburg J.P., Christiansen L., Høge-Jensen B., Norskov L. et al.** (1990). A serine protease triad forms the catalytic centre of a triacylglycerol lipase. *Nature*, 343(6260): 767-70.
- Brown L.M. et Philippon P.** (1999). Microalgues, gazole et effet de serre. *Biofutur*, 1999(195): 28.

- Brzozowski A.M., Derewenda U., Derewenda Z.S., Dodson G.G., Lawson D.M., Turkenburg J.P., Bjorkling F., Huge-Jensen B., Patkar S.A. et Thim L.** (1991). A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature*, 351(6326): 491-4.
- Brzozowski A.M., Savage H., Verma C.S., Turkenburg J.P., Lawson D.M., Svendsen A. et Patkar S.** (2000). Structural origins of the interfacial activation in *Thermomyces (Humicola) lanuginosa* lipase. *Biochemistry*, 39(49): 15071-82.
- Castresana J.** (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular and Biological Evolution*, 17(4): 540-52.
- Castresana J.** (2007). Topological variation in single-gene phylogenetic trees. *Genome Biology*, 8(6): 216.
- Cavalier-Smith T.** (1998). A revised six-kingdom system of life. *Biological Reviews of The Cambridge Philosophical Society*, 73(3): 203-66.
- Chalon S.** (2006). Omega-3 fatty acids and monoamine neurotransmission. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 75(4-5): 259-69.
- Chandra S., Heinstejn P.F. et Low P.S.** (1996). Activation of Phospholipase A by Plant Defense Elicitors. *Plant Physiology*, 110: 979-86.
- Chenchik A., Diachenko L., Moqadam F., Tarabykin V., Lukyanov S. et Siebert P.D.** (1996). Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. *Biotechniques*, 21(3): 526-34.
- Chisti Y.** (2007). Biodiesel from microalgae. *Biotechnology Advances*, 25(3): 294-306.
- Chu K.H., Qi J., Yu Z.G. et Anh V.** (2004). Origin and phylogeny of chloroplasts revealed by a simple correlation analysis of complete genomes. *Molecular and Biological Evolution*, 21(1): 200-6.
- Claros M.G. et von Heijne G.** (1994). TopPred II: an improved software for membrane protein structure predictions. *Computer Applications in The Biosciences* 10(6): 685-6.
- Cohen Z., Khozin-Goldberg I., Adlerstein D. et Bigogno C.** (2000). The role of triacylglycerol as a reservoir of polyunsaturated fatty acids for the rapid production of chloroplastic lipids in certain microalgae. *Biochemical Society Transactions*, 28(6): 740-3.
- Combet C., Jambon M., Deleage G. et Geourjon C.** (2002). Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics*, 18(1): 213-4.
- Craik C.S., Rocznik S., Largman C. et Rutter W.J.** (1987). The catalytic role of the active site aspartic acid in serine proteases. *Science*, 237(4817): 909-13.
- Cygler M., Schrag J.D., Sussman J.L., Harel M., Silman I., Gentry M.K. et Doctor B.P.** (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Science*, 2(3): 366-82.
- Danielo O.** (2005). Fuel based on algal oil. *Biofutur*, 255: 33-6.
- Delwiche C.F.** (1999). Tracing the Thread of Plastid Diversity through the Tapestry of Life. *American Naturalist*, 154(S4): S164-77.
- Derelle E., Ferraz C., Rombauts S., Rouze P., Worden A.Z., Robbens S., Partensky F., Degroeve S., Echeynie S., Cooke R., Saeys Y., Wuyts J., Jabbari K., Bowler C., Panaud O., Piegu B., Ball S.G., Ral J.P., Bouget F.Y., Piganeau G., De Baets B., Picard A., Delseny M., Demaille J., Van de Peer Y. et Moreau H.** (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri*

- unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America*, 103(31): 11647-52.
- Derewenda U., Brzozowski A.M., Lawson D.M. et Derewenda Z.S.** (1992a). Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. *Biochemistry*, 31(5): 1532-41.
- Derewenda U., Swenson L., Green R., Wei Y., Yamaguchi S., Joerger R., Haas M.J. et Derewenda Z.S.** (1994a). Current progress in crystallographic studies of new lipases from filamentous fungi. *Protein Engineering*, 7(4): 551-7.
- Derewenda U., Swenson L., Wei Y., Green R., Kobos P.M., Joerger R., Haas M.J. et Derewenda Z.S.** (1994b). Conformational lability of lipases observed in the absence of an oil-water interface: crystallographic studies of enzymes from the fungi *Humicola lanuginosa* and *Rhizopus delemar*. *Journal of Lipid Research*, 35(3): 524-34.
- Derewenda Z.S. et Derewenda U.** (1991). Relationships among serine hydrolases: evidence for a common structural motif in triacylglyceride lipases and esterases. *Biochemistry and Cell Biology* 69(12): 842-51.
- Derewenda Z.S., Derewenda U. et Dodson G.G.** (1992b). The crystal and molecular structure of the *Rhizomucor miehei* triacylglyceride lipase at 1.9 Å resolution. *Journal of Molecular Biology*, 227(3): 818-39.
- Devedjiev Y., Dauter Z., Kuznetsov S.R., Jones T.L. et Derewenda Z.S.** (2000). Crystal structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 Å. *Structure*, 8(11): 1137-46.
- Devos M.** (2005). Applications des lipases pour la valorisation de l'acide docosahexaénoïque de la microalgue *Isochrysis galbana*. Thèse de doctorat. *Département Génie Biologique, IUT de Laval, Université du Maine*. 144.
- Devos M., Poisson L., Ergon F. et Pencreac'h G.** (2006). Enzymatic hydrolysis of phospholipids from *Isochrysis galbana* for docosahexaenoic acid enrichment. *Enzyme and Microbial Technology*, 39(4): 548-54.
- Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T., Wu X., Reith M., Cavalier-Smith T. et Maier U.G.** (2001). The highly reduced genome of an enslaved algal nucleus. *Nature*, 410(6832): 1091-6.
- Douglas S.E. et Penny S.L.** (1999). The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *Journal of Molecular Evolution*, 48(2): 236-44.
- du Bois T.M., Deng C. et Huang X.-F.** (2005). Membrane phospholipid composition, alterations in neurotransmitter systems and schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 29(6): 878-88.
- Edgar R.C.** (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5): 1792-7.
- Edwardsen B., Eikrem W., Green J.C., Andersen R.A., van der Stay S.Y.M. et Medlin L.K.** (2000). Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia*, 39: 19-35.

- El-Kouhen K., Blangy S., Ortiz E., Gardies A.-M., Ferte N. et Arondel V.** (2005). Identification and characterization of a triacylglycerol lipase in *Arabidopsis* homologous to mammalian acid lipases. *FEBS Letters*, 579(27): 6067-73.
- Fischer M.** (2004). Die Lipase Engineering Database systematische analyse familienspezifischer eigenschaften und der sequenz-struktur-funktionsbeziehung von α/β -hydrolasen. Thèse de doctorat. *Institut für Technische Biochemie der Universität Stuttgart*. 176.
- Fischer M. et Pleiss J.** (2003). The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Research*, 31(1): 319-21.
- Fuchs R.** (1993). Block searches on VAX and Alpha computer systems. *Computer Applications in The Biosciences* 9(5): 587-91.
- Fujiwara S., Tsuzuki M., Kawashi M., Minaka N. et Inouye I.** (2001). Molecular phylogeny of the haptophyta based on the rbcL gene and sequence variation in the spacer region of the rubisco operon. *Journal of Phycology*, 37: 121-29.
- Gachon C.M.M., Day J.G., Campbell C.N., Proschold T., Saxon R.J. et Kupper F.C.** (2007). The Culture Collection of Algae and Protozoa (CCAP): A biological resource for protistan genomics. *Gene*, 406(1-2): 51-7.
- Galtier N., Gouy M. et Gautier C.** (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Computer Applications in The Biosciences* 12(6): 543-8.
- Glockner G., Rosenthal A. et Valentin K.** (2000). The structure and gene repertoire of an ancient red algal plastid genome. *Journal of Molecular Evolution*, 51(4): 382-90.
- Gouet P., Courcelle E., Stuart D.I. et Metoz F.** (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, 15(4): 305-8.
- Gribnikov M., McLachlan A.D. et Eisenberg D.** (1987). Profile analysis: detection of distantly related proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 84(13): 4355-8.
- Grochulski P., Li Y., Schrag J.D., Bouthillier F., Smith P., Harrison D., Rubin B. et Cygler M.** (1993). Insights into interfacial activation from an open structure of *Candida rugosa* lipase. *Journal of Biological Chemistry*, 268(17): 12843-7.
- Grossman A.R.** (2005). Paths toward algal genomics. *Plant Physiology*, 137(2): 410-27.
- Grossman A.R.** (2007). In the grip of algal genomics. *Advances in Experimental Medicine and Biology*, 616: 54-76.
- Guesnet P., Alessandri J.-M., Vancassel S., Denis I. et Laviolle M.** (2005). Acides gras omega 3 et fonctions cerebrales. *Nutrition Clinique et Metabolisme*, 19(3): 131-4.
- Guindon S., Lethiec F., Duroux P. et Gascuel O.** (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research*, 33(Web Server issue): W557-9.
- Guschina I.A. et Harwood J.L.** (2006). Lipids and lipid metabolism in eukaryotic algae. *Progress in Lipid Research*, 45(2): 160-86.
- Hackett J.D., Yoon H.S., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Nosenko T. et Bhattacharya D.** (2004). Migration of the Plastid Genome to the Nucleus in a Peridinin Dinoflagellate. *Current Biology*, 14(3): 213-8.

- Harper J.T., Waanders E. et Keeling P.J.** (2005). On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 55(Pt 1): 487-96.
- Harris R.P.** (1996). Coccolithophorid dynamics: the european *Emiliana huxleyi* programme, EHUX. *Journal of Marine Systems*, 9: 1-11.
- Hasan F., Shah A.A. et Hameed A.** (2006). Industrial applications of microbial lipases. *Enzyme and Microbial Technology*, 39: 235-51.
- Henderson R.J. et Tocher D.T.** (1992). Thin-layer chromatography. Lipids analysis - A practical approach. *Hamilton R.J. et Hamilton S. Oxford, Oxford University Press*: 65-111.
- Henikoff J.G., Pietrokovski S., McCallum C.M. et Henikoff S.** (2000). Blocks-based methods for detecting protein homology. *Electrophoresis*, 21(9): 1700-6.
- Hide W.A., Chan L. et Li W.H.** (1992). Structure and evolution of the lipase superfamily. *Journal of Lipid Research*, 33(2): 167-78.
- Holmquist M.** (2000). Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Current Protein and Peptide Science*, 1(2): 209-35.
- Holmquist M., Martinelle M., Clausen I.G., Patkar S., Svendsen A. et Hult K.** (1994). Trp89 in the lid of *Humicola lanuginosa* lipase is important for efficient hydrolysis of tributyrin. *Lipids*, 29(9): 599-603.
- Hong K.H., Jang W.H., Choi K.D. et Yoo O.J.** (1991). Characterization of *Pseudomonas fluorescens* carboxylesterase: cloning and expression of the esterase gene in *Escherichia coli*. *Agricultural and Biological Chemistry*, 55(11): 2839-45.
- Hubbard C.D. et Shoupe T.S.** (1977). Mechanisms of acylation of chymotrypsin by phenyl esters of benzoic acid and acetic acid. *Journal of Biological Chemistry*, 252(5): 1633-8.
- Ianora A., Poulet S.A., Miralto A. et Grottoli R.** (1996). The diatom *Thalassiosira rotula* affects reproductive success in the copepod *Acartia clausi*. *Marine Biology*, 125(2): 279-86.
- Ileperuma N.R., Marshall S.D., Squire C.J., Baker H.M., Oakeshott J.G., Russell R.J., Plummer K.M., Newcomb R.D. et Baker E.N.** (2007). High-resolution crystal structure of plant carboxylesterase AeCXE1, from *Actinidia eriantha*, and its complex with a high-affinity inhibitor paraoxon. *Biochemistry*, 46(7): 1851-9.
- Itomura M., Hamazaki K., Sawazaki S., Kobayashi M., Terasawa K., Watanabe S. et Hamazaki T.** (2005). The effect of fish oil on physical aggression in schoolchildren -- a randomized, double-blind, placebo-controlled trial. *The Journal of Nutritional Biochemistry*, 16(3): 163-71.
- Jackson A., Jiao P.E., Ni I. et Fu G.K.** (2003). Agarose gel size fractionation of RNA for the cloning of full-length cDNAs. *Analytical Biochemistry*, 323: 252-5.
- Jenkins C.M., Mancuso D.J., Yan W., Sims H.F., Gibson B. et Gross R.W.** (2004). Identification, cloning, expression, and purification of three novel human calcium-independent phospholipase A₂ family members possessing triacylglycerol lipase and acylglycerol transacylase activities. *Journal of Biological Chemistry*, 279(47): 48968-75.
- Jensen R.G., Rubano Galluzzo D. et Bush V.J.** (1990). Selectivity is an important characteristic of lipases (acylglycerol hydrolases). *Biocatalysis*, 3: 307-316.

- Jiang Y., Chen F. et Liang S.-Z.** (1999). Production potential of docosahexaenoic acid by the heterotrophic marine dinoflagellate *Cryptocodinium cohnii*. *Process Biochemistry*, 34(6-7): 633-7.
- Joerger R.D. et Haas M.J.** (1994). Alteration of chain length selectivity of a *Rhizopus deleamar* lipase through site-directed mutagenesis. *Lipids*, 29(6): 377-84.
- Jordan R.W. et Chamberlain A.H.L.** (1997). Biodiversity among haptophyte algae. *Biodiversity and Conservation*, 6: 131-52.
- Kang H.Y., Kim J.F., Kim M.H., Park S.H., Oh T.K. et Hur C.G.** (2006). MELDB: a database for microbial esterases and lipases. *FEBS Letters*, 580(11): 2736-40.
- Kawachi M., Inouye I., Maeda O. et Chihara M.** (1991). The haptonema as a food-capturing device : observations on *Chrysochromulina hirta* (Prymnesiophyceae). *Phycologia*, 30(6): 563-73.
- Keeling P.J.** (2006). <http://tbestdb.bcm.umontreal.ca/searches/organism.php?orgID=IS>.
- Keeling P.J., Archibald J.M., Fast N.M. et Palmer J.D.** (2004). The evolution of modern eukaryotic phytoplankton. *Science*, 306.
- Keeling P.J., Burger G., Durnford D.G., Lang B.F., Lee R.W., Pearlman R.E., Roger A.J. et Gray M.W.** (2005). The tree of eukaryotes. *Trends in Ecology & Evolution*, 20(12): 670-6.
- Khozin-Goldberg I., Shrestha P. et Cohen Z.** (2005). Mobilization of arachidonyl moieties from triacylglycerols into chloroplastic lipids following recovery from nitrogen starvation of the microalga *Parietochloris incisa*. *Biochimica et Biophysica Acta*, 1738(1-3): 63-71.
- Kim K.K., Song H.K., Shin D.H., Hwang K.Y., Choe S., Yoo O.J. et Suh S.W.** (1997). Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an alpha/beta hydrolase with broad substrate specificity. *Structure*, 5(12): 1571-84.
- Kraut J.** (1977). Serine proteases: structure and mechanism of catalysis. *Annual Review of Biochemistry*, 46: 331-58.
- Kulikova T., Akhtar R., Aldebert P., Althorpe N., Andersson M., Baldwin A., Bates K., Bhattacharyya S., Bower L., Browne P., Castro M., Cochrane G., Duggan K., Eberhardt R., Faruque N., Hoad G., Kanz C., Lee C., Leinonen R., Lin Q., Lombard V., Lopez R., Lorenc D., McWilliam H., Mukherjee G., Nardone F., Pastor M.P., Plaister S., Sobhany S., Stoehr P., Vaughan R., Wu D., Zhu W. et Apweiler R.** (2007). EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Research*, 35(Database issue): D16-20.
- Kyte J. et Doolittle R.F.** (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1): 105-32.
- La Claire J.** (2006). Analysis of Expressed Sequence Tags from the Harmful Alga, *Prymnesium parvum* (Prymnesiophyceae, Haptophyta). *Marine Biotechnology*, 8(5): 534-46.
- Larsson Forsell P.K., Kennedy B.P. et Claesson H.E.** (1999). The human calcium-independent phospholipase A₂ gene multiple enzymes with distinct properties from a single gene. *European Journal of Biochemistry*, 262(2): 575-85.
- Larsson P.K., Claesson H.E. et Kennedy B.P.** (1998). Multiple splice variants of the human calcium-independent phospholipase A₂ and their effect on enzyme activity. *Journal of Biological Chemistry*, 273(1): 207-14.

- Leblond J.D. et Chapman P.J.** (2000). Lipid class distribution of highly unsaturated long chain fatty acids in marine dinoflagellates. *Journal of Phycology*, 36: 1103-1108.
- Lee H.Y., Bahn S.C., Kang Y.M., Lee K.H., Kim H.J., Noh E.K., Palta J.P., Shin J.S. et Ryu S.B.** (2003). Secretory low molecular weight phospholipase A₂ plays important roles in cell elongation and shoot gravitropism in *Arabidopsis*. *Plant Cell*, 15(9): 1990-2002.
- Lehner R. et Verger R.** (1997). Purification and characterization of a porcine liver microsomal triacylglycerol hydrolase. *Biochemistry*, 36(7): 1861-8.
- Li S., Nosenko T., Hackett J.D. et Bhattacharya D.** (2006). Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Molecular and Biological Evolution*, 23(3): 663-74.
- Liu C.-P. et Lin L.-P.** (2001). Ultrastructural study and lipid formation of *Isochrysis sp.* CCMP1324. *Botanical Bulletin of Academia Sinica*, 42: 207-14.
- Longhi S. et Cambillau C.** (1999). Structure-activity of cutinase, a small lipolytic enzyme. *Biochimica et Biophysica Acta*, 1441(2-3): 185-96.
- Lotti M., Tramontano A., Longhi S., Fusetti F., Brocca S., Pizzi E. et Alberghina L.** (1994). Variability within the *Candida rugosa* lipases family. *Protein Engineering*, 7(4): 531-5.
- Mancheno J.M., Pernas M.A., Martinez M.J., Ochoa B., Rua M.L. et Hermoso J.A.** (2003). Structural Insights into the Lipase/esterase Behavior in the *Candida rugosa* Lipases Family: Crystal Structure of the Lipase 2 Isoenzyme at 1.97 Å Resolution. *Journal of Molecular Biology*, 332(5): 1059-69.
- Mancuso D.J., Jenkins C.M. et Gross R.W.** (2000). The genomic organization, complete mRNA sequence, cloning, and expression of a novel human intracellular membrane-associated calcium-independent phospholipase A₂. *Journal of Biological Chemistry*, 275(14): 9937-45.
- Mannesse M.L., Cox R.C., Koops B.C., Verheij H.M., de Haas G.H., Egmond M.R., van der Hijden H.T. et de Vlieg J.** (1995). Cutinase from *Fusarium solani* pisi hydrolyzing triglyceride analogues. Effect of acyl chain length and position in the substrate molecule on activity and enantioselectivity. *Biochemistry*, 34(19): 6400-7.
- Margulis L.** (1992). Biodiversity: molecular biological domains, symbiosis and kingdom origins. *Biosystems*, 27(1): 39-51.
- Martinelle M., Holmquist M., Clausen I.G., Patkar S., Svendsen A. et Hult K.** (1996). The role of Glu87 and Trp89 in the lid of *Humicola lanuginosa* lipase. *Protein Engineering*, 9(6): 519-24.
- Martinelle M., Holmquist M. et Hult K.** (1995). On the interfacial activation of *Candida antarctica* lipase A and B as compared with *Humicola lanuginosa* lipase. *Biochimica et Biophysica Acta*, 1258(3): 272-6.
- Martinez C., Nicolas A., van Tilbeurgh H., Egloff M.P., Cudrey C., Verger R. et Cambillau C.** (1994). Cutinase, a lipolytic enzyme with a preformed oxyanion hole. *Biochemistry*, 33(1): 83-9.
- Martz E.** (2002). Protein Explorer: easy yet powerful macromolecular visualization. *Trends in Biochemical Sciences*, 27(2): 107-9.
- Maso M. et Garces E.** (2006). Harmful microalgae blooms (HAB); problematic and conditions that induce them. *Marine Pollution Bulletin*, 53(10-12): 620-30.
- Matsuzaki M., Misumi O., Shin-i T., Maruyama S., Takahara M., Miyagishima S.-Y., Mori T., Nishida K., Yagisawa F., Nishida K., Yoshida Y., Nishimura Y., Nakao S., Kobayashi T., Momoyama Y., Higashiyama T., Minoda A., Sano M., Nomoto H., Oishi K., Hayashi H., Ohta F., Nishizaka S.,**

- Haga S., Miura S., Morishita T., Kabeya Y., Terasawa K., Suzuki Y., Ishii Y., Asakawa S., Takano H., Ohta N., Kuroiwa H., Tanaka K., Shimizu N., Sugano S., Sato N., Nozaki H., Ogasawara N., Kohara Y. et Kuroiwa T. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10 D. *Nature Publishing Group*, 428: 653-7.
- McNamara R.K. et Carlson S.E. (2006). Role of omega-3 fatty acids in brain development and function: Potential implications for the pathogenesis and prevention of psychopathology. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 75(4-5): 329-49.
- 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(5848): 245-50.
- Misumi O., Matsuzaki M., Nozaki H., Miyagishima S.-Y., Mori T., Nishida K., Yagisawa F., Yoshida Y., Kuroiwa H. et Kuroiwa T. (2005). *Cyanidioschyzon merolae* genome. a tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. *Plant Physiology*, 137: 567-85.
- Mu H. et Porsgaard T. (2005). The metabolism of structured triacylglycerols. *Progress in Lipid Research*, 44(6): 430-48.
- Murphy D.J. (2001). The biogenesis and functions of lipid bodies in animals, plants and microorganisms. *Progress in Lipid Research*, 40(5): 325-438.
- Murphy J.C., Fox G.E. et Willson R.C. (2001). RNA isolation and fractionation with compaction agents. *Analytical Biochemistry*, 295: 143-8.
- Myers M., Richmond R.C. et Oakeshott J.G. (1988). On the origins of esterases. *Molecular and Biological Evolution*, 5(2): 113-9.
- Nagasaki K., Tarutani K. et Yamaguchi M. (1999). Growth characteristics of *Heterosigma akashiwo* virus and its possible use as a microbiological agent for red tide control. *Applied and Environmental Microbiology*, 65(3): 898-902.

- Nardini M. et Dijkstra B.W.** (1999). α/β Hydrolase fold enzymes: the family keeps growing. *Current Opinion in Biotechnology*, 9: 732-737.
- Narvaez-Vasquez J., Florin-Christensen J. et Ryan C.A.** (1999). Positional specificity of a phospholipase A activity induced by wounding, systemin, and oligosaccharide elicitors in tomato leaves. *Plant Cell*, 11(11): 2249-60.
- Nieminen L.R.G., Makino K.K., Mehta N., Virkkunen M., Kim H.Y. et Hibbeln J.R.** (2006). Relationship between omega-3 fatty acids and plasma neuroactive steroids in alcoholism, depression and controls. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 75(4-5): 309-14.
- Noble M.E., Cleasby A., Johnson L.N., Egmond M.R. et Frenken L.G.** (1993). The crystal structure of triacylglycerol lipase from *Pseudomonas glumae* reveals a partially redundant catalytic aspartate. *FEBS Letters*, 331(1-2): 123-8.
- Ollis D., Cheah E., Cygler M., Dijkstra B.W., Frolow F., Franken S., Harel M., Remington S., Silman L., Shrag J., Sussman J., Verschueren K. et Goldman A.** (1992). The α/β hydrolase fold. *Protein Engineering*, 5: 197-211.
- Page R.D.** (1996). TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in The Biosciences* 12(4): 357-8.
- Palenik B., Grimwood J., Aerts A., Rouze P., Salamov A., Putnam N., Dupont C., Jorgensen R., Derelle E., Rombauts S., Zhou K., Otilar R., Merchant S.S., Podell S., Gaasterland T., Napoli C., Gendler K., Manuell A., Tai V., Vallon O., Piganeau G., Jancek S., Heijde M., Jabbari K., Bowler C., Lohr M., Robbens S., Werner G., Dubchak I., Pazour G.J., Ren Q., Paulsen I., Delwiche C., Schmutz J., Rokhsar D., Van de Peer Y., Moreau H. et Grigoriev I.V.** (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18): 7705-10.
- Patterson G.W.** (1970). Effect of culture temperature on fatty acid composition of *Chlorella sorokiniana*. *Lipids*, 5(7): 597-600.
- Peet M.** (2003a). Eicosapentaenoic acid in the treatment of schizophrenia and depression: rationale and preliminary double-blind clinical trial results. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 69(6): 477-485.
- Peet M.** (2003b). Nutrition and schizophrenia: An epidemiological and clinical perspective. *Nutrition and Health*, 17(3): 211-219.
- Pencreac'h G. et Baratti J.C.** (1996). Hydrolysis of *p*-nitrophenyl palmitate in *n*-heptane by the *Pseudomonas cepacia* lipase: a simple test for the determination of lipase activity in organic media. *Enzyme and Microbial Technology*, 18: 417-22.
- Pencreac'h G., Devos M., Poisson L. et Ergon F.** (2004). Les microalgues marines: source alternative d'acide eicosapentaénoïque (EPA) et d'acide docosahexaénoïque (DHA). *Oléagineux, Corps Gras, Lipides*, 11: 118-22.
- Peters G.H., Svendsen A., Langberg H., Vind J., Patkar S.A., Toxvaerd S. et Kinnunen P.K.** (1998). Active serine involved in the stabilization of the active site loop in the *Humicola lanuginosa* lipase. *Biochemistry*, 37(36): 12375-83.

- Pistillo D., Manzi A., Tino A., Boyl P.P., Graziani F. et Malva C.** (1998). The *Drosophila melanogaster* lipase homologs: a gene family with tissue and developmental specific expression. *Journal of Molecular Biology*, 276(5): 877-85.
- Pleiss J., Fischer M., Peiker M., Thiele C. et Schmid R.D.** (2000). Lipase engineering database: Understanding and exploiting sequence-structure-function relationships. *Journal of Molecular Catalysis B: Enzymatic*, 10(5): 491-508.
- Pohnert G.** (2002). Phospholipase A₂ activity triggers the wound-activated chemical defense in the diatom *Thalassiosira rotula*. *Plant Physiology*, 129: 103-11.
- Poisson L.** (1999). Valorisation enzymatique de lipides: 1. Synthèse de cires à partir de la matière grasse laitière 2. Production et traitements d'acides gras polyinsaturés de microalgues. Thèse de doctorat. *Département Génie Biologique, IUT de Laval, Université du Maine*. 131.
- Poisson L. et Ergon F.** (2001). Docosahexaenoic acid ethyl esters from *Isochrysis galbana*. *Journal of Biotechnology*, 91: 75-81.
- Portilla D., Crew M.D., Grant D., Serrero G., Bates L.M., Dai G., Sasner M., Cheng J. et Buonanno A.** (1998). cDNA cloning and expression of a novel family of enzymes with calcium-independent phospholipase A₂ and lysophospholipase activities. *Journal of the American Society of Nephrology*, 9: 1178-86.
- Reith M. et Munholland J.** (1993). A High-Resolution Gene Map of the Chloroplast Genome of the Red Alga *Porphyra purpurea*. *Plant Cell*, 5(4): 465-75.
- Rhodes L., Smith J., Tervit R., Roberts R., Adamson J., Adams S. et Decker M.** (2006). Cryopreservation of economically valuable marine micro-algae in the classes *bacillariophyceae*, *chlorophyceae*, *cyanophyceae*, *dinophyceae*, *haptophyceae*, *prasinophyceae*, and *rhodophyceae*. *Cryobiology*, 52: 152-6.
- Ris H. et Plaut W.** (1962). Ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas*. *Journal of Cell Biology*, 13: 383-91.
- Robert J.-M.** (1983). Fertilité des eaux des claires ostréicoles et verdissement: utilisation de l'azote par les diatomées dominantes. Thèse de doctorat. *Université de Nantes*.
- Robert R., Connan J.P., Leroy B., Chrétiennot-Dinet M.J., Le Coz J.R., Moal J., Quéré C., Martin-Jézéquel V., Le Gourrierec G., Nicolas J.L., Bernard E., Kaas R. et Le Déan L.** (2004). Amélioration des productions phytoplanctoniques en éclosérie de mollusques : caractérisation des microalgues fourrage. Rapport. *Ifremer/DRV/RA/LPI Argenton, Laboratoire de Physiologie des Invertébrés Marins*. 144.
- Roessler P.G.** (1990). Environmental control of glycerolipid metabolism in microalgae: commercial implications and future research directions. *Journal of Phycology*, 26: 393-399.
- Rusch D.B., Halpern A.L., Sutton G., Heidelberg K.B., Williamson S., Yooseph S., Wu D., Eisen J.A., Hoffman J.M., Remington K., Beeson K., Tran B., Smith H., Baden-Tillson H., Stewart C., Thorpe J., Freeman J., Andrews-Pfannkoch C., Venter J.E., Li K., Kravitz S., Heidelberg J.F., Utterback T., Rogers Y.H., Falcon L.I., Souza V., Bonilla-Rosso G., Eguarte L.E., Karl D.M., Sathyendranath S., Platt T., Bermingham E., Gallardo V., Tamayo-Castillo G., Ferrari M.R., Strausberg R.L., Neilson K., Friedman R., Frazier M. et Venter J.C.** (2007). The Sorcerer II

- Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology*, 5(3): e77.
- Saitou N. et Nei M.** (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular and Biological Evolution*, 4(4): 406-25.
- Sambrook J., Fritsch E.F. et Maniatis T.** (1989). Molecular cloning : A laboratory Manual, *Cold Spring Harbor Laboratory Press*.
- Sanchez Puerta M.V., Bachvaroff T.R. et Delwiche C.F.** (2004). The complete mitochondrial genome sequence of the haptophyte *Emiliana huxleyi* and its relation to heterokonts (supplement). *DNA Research*, 11(1): 67-8.
- Sanchez Puerta M.V., Bachvaroff T.R. et Delwiche C.F.** (2005). The complete plastid genome sequence of the haptophyte *Emiliana huxleyi*: a comparison to other plastid genomes. *DNA Research*, 12(2): 151-6.
- Sarda L. et Desnuelle P.** (1958). Action de la lipase pancréatique sur les esters en émulsion. *Biochimica and Biophysica Acta*, 30: 513-521.
- Sayari A., Mosbah H. et Gargouri Y.** (2007). Importance of the residue Asp 290 on chain length selectivity and catalytic efficiency of recombinant *Staphylococcus simulans* lipase expressed in *E. coli*. *Molecular Biotechnology*, 36(1): 14-22.
- Sayle R.A. et Milner-White E.J.** (1995). RASMOL: biomolecular graphics for all. *Trends in Biochemical Sciences*, 20(9): 374-6.
- Scala S., Carels N., Falciatore A., Chiusano M.L. et Bowler C.** (2002). Genome properties of the diatom *Phaeodactylum tricornutum*. *Plant Physiology*, 129(3): 993-1002.
- Schaller M., Borelli C., Korting H.C. et Hube B.** (2005). Hydrolytic enzymes as virulence factors of *Candida albicans*. *Mycoses*, 48(6): 365-77.
- Schmid R.D. et Verger R.** (1998). Lipases: interfacial enzymes with attractive applications. *Angewandte Chemie International Edition*, 37: 1608-1633.
- Schofield D.A., Westwater C., Warner T. et Balish E.** (2005). Differential *Candida albicans* lipase gene expression during alimentary tract colonization and infection. *FEMS Microbiology Letters*, 244(2): 359-65.
- Schrag J.D., Li Y.G., Wu S. et Cygler M.** (1991). Ser-His-Glu triad forms the catalytic site of the lipase from *Geotrichum candidum*. *Nature*, 351(6329): 761-4.
- Sedgwick S.G. et Smerdon S.J.** (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends in Biochemical Sciences*, 24(8): 311-6.
- Selosse M.A.** (2006). Animal ou végétal : une distinction obsolète. *Pour La Science*, 350.
- Selosse M.A. et Loiseaux-de Goër S.** (1997). Evolution et endosymbiose. *La Recherche*, 296.
- Sene C., Girot P., Boschetti E., Plassat J.L., Bloch J. et Egly J.M.** (1982). Purification of polyadenylated mRNA on three oligo(dT)-substituted gels: a comparative study. *Journal of Chromatography*, 248(3): 441-5.
- Sigrist C.J., Cerutti L., Hulo N., Gattiker A., Falquet L., Pagni M., Bairoch A. et Bucher P.** (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings in Bioinformatics*, 3(3): 265-74.

- Simo R.** (2001). Production of atmospheric sulfur by oceanic plankton: biogeochemical, ecological and evolutionary links. *Trends in Ecology & Evolution*, 16(6): 287-94.
- Six D.A. et Dennis E.A.** (2003). Essential Ca²⁺-independent role of the group IVA cytosolic phospholipase A₂ C2 domain for interfacial activity. *Journal of Biological Chemistry*, 278(26): 23842-50.
- Skovgaard A., Legrand C., Hansen P.J. et Granéli E.** (2003). Effects of nutrient limitation on food uptake in the toxic haptophyte *Prymnesium parvum*. *Aquatic Microbial Ecology*, 31: 259–265.
- Spolaore P., Joannis-Cassan C., Duran E. et Isambert A.** (2006). Commercial applications of microalgae. *Journal of Bioscience and Bioengineering*, 101(2): 87-96.
- Stoesser G., Baker W., van den Broek A., Camon E., Garcia-Pastor M., Kanz C., Kulikova T., Lombard V., Lopez R., Parkinson H., Redaschi N., Sterk P., Stoehr P. et Tuli M.A.** (2001). The EMBL nucleotide sequence database. *Nucleic Acids Research*, 29(1): 17-21.
- Stothard P.** (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28(6): 1102, 1104.
- Suga M. et Hatakeyama T.** (2001). High efficiency transformation of *Schizosaccharomyces pombe* pretreated with thiol compounds by electroporation. *Yeast*, 18(11): 1015-21.
- Sukenik A., Carmeli Y. et Berner T.** (1989). Regulation of fatty acid composition by irradiance level in the eustigmatophyte *Nannochloropsis* sp. *Journal of Phycology*, 25: 686-92.
- Svendsen A.** (2000). Lipase protein engineering. *Biochimica et Biophysica Acta*, 1543(2): 223-38.
- Tanaka H., Takeya R. et Sumimoto H.** (2000). A novel intracellular membrane-bound calcium-independent phospholipase A₂. *Biochemical and Biophysical Research Communications*, 272(2): 320-6.
- Tateno Y., Miyazaki S., Ota M., Sugawara H. et Gojobori T.** (2000). DNA data bank of Japan (DDBJ) in collaboration with mass sequencing teams. *Nucleic Acids Research*, 28(1): 24-6.
- Thompson J.D., Higgins D.G. et Gibson T.J.** (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22): 4673-80.
- Tillmann U.** (1998). Phagotrophy by a plastidic haptophyte *Prymnesium patelliferum*. *Aquatic Microbial Ecology*, 14: 155-60.
- Tillmann U.** (2003). Kill and eat your predator: a winning strategy of the planktonic flagellate *Prymnesium parvum*. *Aquatic Microbial Ecology*, 32: 73–84.
- Tsai I.-H., Wang Y.-M., Chen Y.-H., Tsai T.-S. et Tu M.-C.** (2004). Venom phospholipases A₂ of bamboo viper (*Trimeresurus stejnegeri*): molecular characterization, geographic variations and evidence of multiple ancestries. *Biochemistry Journal*, 377: 215-23.
- Uppenberg J., Hansen M.T., Patkar S. et Jones T.A.** (1994a). The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica*. *Structure*, 2(4): 293-308.
- Uppenberg J., Patkar S., Bergfors T. et Jones T.A.** (1994b). Crystallization and preliminary X-ray studies of lipase B from *Candida antarctica*. *Journal of Molecular Biology*, 235(2): 790-2.
- Van de Peer Y. et De Wachter R.** (1997). Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *Journal of Molecular Evolution*, 45(6): 619-30.
- van Leeuwen W., Okresz L., Bogre L. et Munnik T.** (2004). Learning the lipid language of plant signalling. *Trends in Plant Science*, 9: 378-84.

- van Pouderoyen G., Eggert T., Jaeger K.E. et Dijkstra B.W.** (2001). The crystal structure of *Bacillus subtilis* lipase: a minimal alpha/beta hydrolase fold enzyme. *Journal of Molecular Biology*, 309(1): 215-26.
- Vancassel S.** (2004). Oméga 3 et neurotransmission cérébrale. *Oléagineux, Corps Gras, Lipides*, 11(1): 58-65.
- von Schacky C. et Harris W.S.** (2007). Cardiovascular benefits of omega-3 fatty acids. *Cardiovascular Research*, 73(2): 310-15.
- Wada M., Hara Y., Kato M., Yamada M. et Fujii T.** (1987). Diurnal appearance, fine structure, and chemical composition of fatty particles in *Heterosigma akashiwo* (Raphidophyceae). *Protoplasma*, 137(2): 134-9.
- Wahlund T.M., Hadaegh A.R., Clark R., Nguyen B., Fanelli M. et Read B.A.** (2004). Analysis of Expressed Sequence Tags from Calcifying Cells of Marine Coccolithophorid (*Emiliana huxleyi*). *Marine Biotechnology*, 6(3): 278-90.
- Wheelan S.J., Church D.M. et Ostell J.M.** (2001). Spidey: a tool for mRNA-to-genomic alignments. *Genome Research*, 11(11): 1952-7.
- Winkler F.K., D'Arcy A. et Hunziker W.** (1990). Structure of human pancreatic lipase. *Nature*, 343(6260): 771-4.
- Winstead M.V., Balsinde J. et Dennis E.A.** (2000). Calcium-independent phospholipase A₂ : structure and function. *Biochimica et Biophysica Acta*, 1488(1-2): 28-39.
- Wong H. et Schotz M.C.** (2002). The lipase gene family. *Journal of Lipid Research*, 43(7): 993-9.
- Yadav R.P., Saxena R.K., Gupta R. et Davidson W.S.** (1998). Rapid zymogram for lipase. *Biotechniques*, 24(5): 754-6.
- Yooseph S., Sutton G., Rusch D.B., Halpern A.L., Williamson S.J., Remington K., Eisen J.A., Heidelberg K.B., Manning G., Li W., Jaroszewski L., Cieplak P., Miller C.S., Li H., Mashiyama S.T., Joachimiak M.P., van Belle C., Chandonia J.M., Soergel D.A., Zhai Y., Natarajan K., Lee S., Raphael B.J., Bafna V., Friedman R., Brenner S.E., Godzik A., Eisenberg D., Dixon J.E., Taylor S.S., Strausberg R.L., Frazier M. et Venter J.C.** (2007). The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biology*, 5(3): e16.
- Yoshida M., Noël M.-H., Nakayama T., Naganuma T. et Inouye I.** (2006). A haptophyte bearing siliceous scales: ultrastructure and phylogenetic position of *Hyalolithus neolepis* gen. et sp. nov. (*Prymnesiophyceae, Haptophyta*). *Protist*, 157: 213-34.
- Zhao G., Etherton T.D., Martin K.R., Vanden Heuvel J.P., Gillies P.J., West S.G. et Kris-Etherton P.M.** (2005). Anti-inflammatory effects of polyunsaturated fatty acids in THP-1 cells. *Biochemical and Biophysical Research Communications*, 336(3): 909-17.