



HAL
open science

Etude de la morphogenèse et de la voie de N-glycosylation chez la diatomée *P. tricornutum*

Clément Ovide

► **To cite this version:**

Clément Ovide. Etude de la morphogenèse et de la voie de N-glycosylation chez la diatomée *P. tricornutum*. Biochimie, Biologie Moléculaire. Université de Rouen Normandie, 2016. Français. NNT: . tel-02301887

HAL Id: tel-02301887

<https://normandie-univ.hal.science/tel-02301887>

Submitted on 30 Sep 2019

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

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



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité ASPECTS MOLECULAIRES ET CELLULAIRES DE LA BIOLOGIE (4200006)

Préparée au sein de l'Université de Rouen Normandie

Etude de la morphogenèse et de la voie de *N*-glycosylation chez la diatomée *P. tricornutum*

Présentée et soutenue par
Clément OVIDE

Thèse soutenue publiquement le 16 Décembre 2016 devant le jury composé de

Mr Benoît SCHOEFS	Professeur, Laboratoire Mer, Molécules, Santé (EA2160) IUML – FR 3473 CNRS, Université du Maine à Le Mans	Rapporteur
Mme Fayza DABOUSSI	Directeur de recherche, Laboratoire LISBP INSA/CNRS 5504 - UMR INSA/INRA 792, INSA de Toulouse	Rapporteur
Mr Bruno SAINT-JEAN	Chargé de recherche, Laboratoire Physiologie et Biotechnologie des Algues (PDG-RBE-BRM-PBA), Centre Atlantique, Nantes	Examineur
Mme Angela FALCIATORE	Directeur de recherche, Laboratoire de Biologie Computationnelle et Quantitative UMR 7238 CNRS, Université Pierre et Marie Curie, Paris	Examineur
Mr Patrice LEROUGE	Professeur, Laboratoire de Glycobiologie et Matrice Extracellulaire végétale EA 4358, Université de Rouen Normandie	Examineur
Mme Hélène DAUCHEL	Maître de conférences, Laboratoire LITIS EA 4108, Université de Rouen Normandie	Examineur
Mme Marie-Christine KIEFER-MEYER	Ingénieur de recherche, Laboratoire de Glycobiologie et Matrice Extracellulaire végétale EA 4358, Université de Rouen Normandie	Examineur
Mme Muriel BARDOR	Professeur, Laboratoire de Glycobiologie et Matrice Extracellulaire végétale EA 4358, Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par Mme Muriel BARDOR, Professeur, Laboratoire de glycobiologie et Matrice Extracellulaire Végétale, Université de Rouen Normandie (EA4358 Glyco-MEV)





Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité ASPECTS MOLECULAIRES ET CELLULAIRES DE LA BIOLOGIE (4200006)

Préparée au sein de l'Université de Rouen Normandie

Etude de la morphogenèse et de la voie de *N*-glycosylation chez la diatomée *P. tricornutum*

Présentée et soutenue par
Clément OVIDE

Thèse soutenue publiquement le 16 Décembre 2016 devant le jury composé de

Mr Benoît SCHOEFS	Professeur, Laboratoire Mer, Molécules, Santé (EA2160) IUML – FR 3473 CNRS, Université du Maine à Le Mans	Rapporteur
Mme Fayza DABOUSSI	Directeur de recherche, Laboratoire LISBP INSA/CNRS 5504 - UMR INSA/INRA 792, INSA de Toulouse	Rapporteur
Mr Bruno SAINT-JEAN	Chargé de recherche, Laboratoire Physiologie et Biotechnologie des Algues (PDG-RBE-BRM-PBA), Centre Atlantique, Nantes	Examineur
Mme Angela FALCIATORE	Directeur de recherche, Laboratoire de Biologie Computationnelle et Quantitative UMR 7238 CNRS, Université Pierre et Marie Curie, Paris	Examineur
Mr Patrice LEROUGE	Professeur, Laboratoire de Glycobiologie et Matrice Extracellulaire végétale EA 4358, Université de Rouen Normandie	Examineur
Mme Hélène DAUCHEL	Maître de conférences, Laboratoire LITIS EA 4108, Université de Rouen Normandie	Examineur
Mme Marie-Christine KIEFER-MEYER	Ingénieur de recherche, Laboratoire de Glycobiologie et Matrice Extracellulaire végétale EA 4358, Université de Rouen Normandie	Examineur
Mme Muriel BARDOR	Professeur, Laboratoire de Glycobiologie et Matrice Extracellulaire végétale EA 4358, Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par Mme Muriel BARDOR, Professeur, Laboratoire de glycobiologie et Matrice Extracellulaire Végétale, Université de Rouen Normandie (EA4358 Glyco-MEV)



Remerciements

Je remercie en premier lieu le professeur Patrice LEROUGE, Directeur du Laboratoire Glycobiologie et Matrice Extracellulaire Végétale (Glyco-MEV, EA4358) de l'Université de Rouen, de m'avoir accueilli au sein de son équipe et d'avoir accepté de participer au jury de cette thèse.

Je remercie Madame Fayza DABOUSSI, Directrice de Recherche au sein du Laboratoire LISBP de l'INSA de Toulouse, et Monsieur Benoît SCHOEFS, Professeur à l'Université du Maine, de m'avoir fait l'honneur de rapporter sur ce travail.

Je remercie Madame Angela FALCIATORE, Directrice de Recherche au laboratoire LCQB de l'Université Pierre et Marie CURIE de Paris, ainsi que Monsieur Bruno SAINT-JEAN, Chargé de Recherche au Laboratoire PBA de Nantes, d'avoir accepté de participer au jury de cette thèse.

Je remercie tout particulièrement ma directrice de thèse, Muriel BARDOR, Professeur à l'université de Rouen, pour la confiance qu'elle a su m'accorder, ses conseils toujours avisés et son écoute attentive. J'ai eu le sentiment agréable d'avoir été encouragé et porté vers l'avant tout au long de cette thèse. Je vous remercie pour votre engagement, avec tous ces week-ends et ces nuits sacrifiées pour l'avancement de ce projet. J'ai vraiment apprécié votre encadrement qui a été très formateur, bien au-delà du plan scientifique.

Merci au Docteur Marie-Christine KIEFER-MEYER, Ingénieur de Recherche au Laboratoire Glyco-MEV, compagne de RNA-seq, pour les nombreuses heures passées ensemble cachés derrière sa muraille colorée à se questionner, à découvrir et à me prodiguer de précieux conseils sur la biologie moléculaire. J'ai énormément appris à vos côtés.

Je remercie le Docteur Hélène DAUCHEL, Maître de Conférences à l'Université de Rouen, d'avoir accepté de participer au jury de cette thèse. Merci pour votre collaboration à ce projet et vos conseils éclairés qui m'ont fait découvrir le monde passionnant de la bio-informatique.

Je remercie le Docteur Franziska HEMPEL, le Professeur Uwe MAIER, et toute l'équipe de doctorants du laboratoire Synmikro de l'Université de Marbourg pour m'avoir accueilli dans leur laboratoire, d'avoir mis à ma disposition tous les moyens dont j'avais besoin,

et surtout pour l'intégration dans leur joyeuse équipe qui m'a permis de me sentir bien plus qu'un simple visiteur.

Je remercie Carole BUREL, Carole PLASSON, Jessica VANIER et Gaëlle LUCAS du Laboratoire Glyco-MEV qui ont été d'une aide cruciale et qui n'ont jamais été avares de leur temps.

Je souhaite remercier le Docteur Sophie BERNARD du Laboratoire Glyco-MEV, ainsi que le Docteur Isabelle TOURNIER de la plateforme génomique de l'IRIB pour leur collaboration à ce travail et pour toute l'aide apportée.

Je remercie le Docteur Bérengère BAIET-HAMONIC qui m'a encadrée pendant mes stages de Master 1 et 2, qui a su m'intégrer au sein du laboratoire, et qui m'a donnée toutes les clés pour réaliser cette thèse. Toujours de bonne humeur, tu m'as fait découvrir ce que la chanson avait de meilleur... mais aussi de pire :). J'espère que tu t'éclates dans ta nouvelle vie.

Je remercie également les Docteurs Caroline BERARD, Nicolas VERGNE et Thierry LECROQ du laboratoire LITIS pour les longues heures passées à marier nos deux univers et pour votre contribution cruciale à ce projet.

Je tiens à remercier le Docteur Bruno GÜGI, Maître de Conférences à l'Université de Rouen, pour les discussions passionnantes que nous avons pu avoir, ainsi que pour la relecture de ce manuscrit.

Un énorme merci à toute la joyeuse bande du Glyco-MEV qui permet de venir travailler tous les matins avec la banane. Tout le monde a contribué de plus ou moins loin à cette thèse, ne serait-ce que par une discussion, un milieu autoclavé, un encouragement, ou même un coup de balai dans le bureau. Merci aux doctorants, passés ou présents, drôles (ou qui voudraient l'être), aux chauves, aux désagréables, aux libanais, aux champagnols, aux survivants, aux marins, aux belles rousses, aux brayonnes, aux coolest, à tous ceux qui ont partagés ce moment avec moi et m'ont permis d'encaisser avec plaisir toutes ces heures de travail.

Le remerciement ultime va à ma famille pour leur soutien indéfectible et leur présence tout au long de ces années, en particulier à mes parents qui ont toujours su me donner les moyens de réaliser mes projets.

Clément OVIDE

Présentation du candidat

DONNEES PERSONNELLES

Date de naissance : 28/04/1987

Nationalité française

PARCOURS UNIVERSITAIRE

• Formation actuelle depuis octobre 2012 :

Doctorat de Biologie cellulaire "Etude de la N-glycosylation et de la morphogenèse chez la diatomée modèle *Phaeodactylum tricornutum*", Université de Rouen Normandie (Laboratoire Glyco-MEV EA4358), Allocation de recherche ministérielle

• 2015-2016

Attaché temporaire d'enseignement et de recherche (ATER) en section 64 : Biochimie et biologie moléculaire, Université de Rouen

• 2010-2012

Master de Biologie Cellulaire parcours Sciences Végétales, Université de Rouen, Mention Bien, 7^{ème}/52

• 2008-2010

Licence de Biochimie et Biologie Moléculaire, Université de Rouen, Mention Assez Bien

• 2006-2008

Première année de médecine (PCEM1), Université de Rouen

ACTIVITES D'ENSEIGNEMENT

• 2015-2016

ATER en section 64, 72h d'enseignement équivalent TD, heures réparties en TP et TD de Biochimie assurés auprès d'étudiants inscrits en L1 et L3 de Licence SVTE.

• 2013-2015

Monitorat de Biochimie, 128h d'enseignement équivalent TD, heures réparties en TP et TD de Biochimie assurés auprès d'étudiants inscrits en L1, L2 et L3 de Licence SVTE, et 3^{ème} année d'école d'ingénieur

SEJOUR DE RECHERCHE A L'ETRANGER

Stage de 3 mois au sein du laboratoire Synmikro de l'université de Marburg (Allemagne) en vue de réaliser la construction de vecteurs contenant différents gènes de glycosyltransférases. Ces derniers ont été couplés à des tags afin de réaliser une localisation subcellulaire de ces enzymes chez la diatomée *Phaeodactylum tricornutum*. Pour cela les cellules ont été transformées par biolistique puis les recombinants ont été observés au microscope confocal.

PUBLICATIONS

• Publication en préparation : **Clément Ovide**, Marie-Christine Kiefer-Meyer, Caroline Bérard, Nicolas Vergne, Thierry Lecroq, Carole Plasson, Carole Burel, Sophie Bernard, Patrice Lerouge, Isabelle Tournier, Hélène Dauchel and Muriel Bardor. **Comparative transcriptome analysis of *P. tricornutum*'s morphotypes using in depth RNA sequencing reveals the oval morphotype features, especially that protein degradation and secretion is up-regulated in the oval morphotype.**

• Publication en preparation : **Clément Ovide**, Carole Plasson, Muriel Bardor and Marie-Christine Kiefer-Meyer **Identification of reference genes for quantitative real-time PCR analysis in *Phaeodactylum tricornutum* morphotypes**

• Elodie Mathieu-Rivet, Marie-Christine Kiefer-Meyer, Gaëtan Vanier, **Clément Ovide**, Carole Burel, Patrice Lerouge and Muriel Bardor (2014) **Protein N-glycosylation in eukaryotic microalgae and its impact on the production of microalgae-made biopharmaceuticals**. *Frontiers in plant science*, 5:359.

COMMUNICATIONS PAR AFFICHE

Le nom de la personne ayant réalisé la communication est souligné.

Congrès internationaux

• Kiefer-Meyer M.-C., **Ovide C.**, Tair W., Bérard C., Vergne N., Lecroq T., Lerouge P., Tournier I., Dauchel H. and Bardor M. Towards a better understanding of *Phaeodactylum tricornutum*'s morphogenesis. **Plant Genomics Congress, Londres 13-14 Mai 2013.**

• Ovide C., Kiefer-Meyer M.-C., Tair W., Bérard C., Vergne N., Lecroq T., Lerouge P., Tournier I., Dauchel H. and Bardor M. RNA-seq: a new tool to study *Phaeodactylum tricornutum*'s morphogenesis. **EMBO Workshop – The molecular life of diatoms, Paris 25-28 Juin 2013.**

• **Ovide C.***, Baïet B.*, Mathieu-Rivet E., Dardelle F., Burel C., Saint-Jean B., Kiefer-Meyer M.-C., Cadoret J.-P., Lerouge P. and Bardor M. Deciphering the N-glycosylation pathway in microalgae using the diatom *Phaeodactylum tricornutum* as a model. **EMBO Workshop – The molecular Life of diatoms, Paris 25-28 Juin 2013.** * These two co-authors equally contributed to this work

• Ovide C.*, Baïet B.*, Burel C., Lerouge P., Kiefer-Meyer M.-C. and Bardor M. Functional characterization of the Golgi Glycosyltransferases from the diatom *Phaeodactylum tricornutum*. **GLYCO-T, Porto 18-21 Juin 2014.**

* These two co-authors equally contributed to this work

• **Ovide C.**, Vanier G., Mathieu-Rivet E., Burel C., Lerouge P., Kiefer-Meyer M.-C. and Bardor M. Microalgae glycobiology could help deciphering the evolution of N-glycosylation pathways. **Gordon Research Conferences “Glycobiology”: Glycans as Mediators of Interactions Between Molecules, Cells and Organisms, Barga, Italie, 1-6 Mars 2015**

Congrès nationaux

• Ovide C., Kiefer-Meyer M.-C., Tair W., Bérard C., Vergne N., Lecroq T., Lerouge P., Tournier I., Dauchel H. and Bardor M. RNA-seq: a new tool to study *Phaeodactylum tricornutum*'s morphogenesis. **Scientific meeting of IRIB, Rouen 21 Juin 2013.**

• Ovide C. *, Kiefer-Meyer M.-C. *, Tair W., Bérard C., Vergne N., Lecroq T., Lerouge P., Tournier I., Dauchel H. and Bardor M. Elucidation of the morphogenesis mechanisms in the diatom *P.tricornutum* using RNA Sequencing technology. **Journées de l'école doctorale de l'EdNBISE, Le Havre 10-11 Avril 2014.**

* These two co-authors equally contributed to this work

• Ovide C.*, Baïet B.*, Hempel F., Lerouge P., Maier U., Kiefer-Meyer M.-C and Bardor M. Functional characterization of the Golgi Glycosyltransferases from the diatom *Phaeodactylum tricornutum*. **Journées de l'école doctorale de l'EdNBISE, Rouen 26-27 Mars 2015.**

* These two co-authors equally contributed to this work

COMMUNICATIONS ORALES

Congrès international

• **Ovide C.***, Baïet B.*, Mathieu-Rivet E., Dardelle F., Burel C., Saint-Jean B., Kiefer-Meyer M.-C., Cadoret J.-P., Lerouge P. and Bardor M. Deciphering the N-glycosylation pathway in microalgae using the diatom *Phaeodactylum tricornutum* as a model. **EMBO Workshop – The molecular Life of diatoms, Paris 25-28 Juin 2013.** * These two co-authors equally contributed to this work.

Congrès nationaux

• Mathieu-Rivet E., **Ovide C.**, Baïet B., Kiefer-Meyer M.-C., Burel C., Plasson C., Lerouge P. and Bardor M. Deciphering microalgae glycobiology: *Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum* as models to understand microalgae N-glycosylation pathways. **2^{ème} journée de l'IRIB, Mont-Saint Aignan 21 juin 2013.**

• **Ovide C.** *, Kiefer-Meyer M.-C. *, Tair W., Bérard C., Vergne N., Lecroq T., Lerouge P., Tournier I., Dauchel H. and Bardor M. Elucidation of the morphogenesis mechanisms in the diatom *P.tricornutum* using RNA Sequencing technology. **Journées de l'école doctorale de l'EdNBISE, Le Havre 10-11 Avril 2014.**
* These two co-authors equally contributed to this work

• **Ovide C.** *, Kiefer-Meyer M.-C. *, Bérard C., Vergne N., Lecroq T., Lerouge P., Tournier I., Dauchel H. and Bardor M. RNA Sequencing technology will help deciphering the morphogenesis mechanisms in the model diatom *Phaeodactylum tricornutum*. **4^{ème} journée de l'IRIB, Rouen 5 juin 2015.**
* These two co-authors equally contributed to this work

• **Ovide C.** *, Kiefer-Meyer M.-C. *, Bérard C., Vergne N., Lecroq T., Lerouge P., Tournier I., Dauchel H. and Bardor M. Etude des mécanismes de morphogenèse chez la diatomée modèle *Phaeodactylum tricornutum* par séquençage d'ARN à haut débit (RNA-Seq). **Journées phycologiques de France, Vannes 23-24 Septembre 2015.**
* These two co-authors equally contributed to this work

FORMATIONS COMPLÉMENTAIRES

Workshops

- 1st Wallnet Interreg Training Course: How my cell wall looks like?
- 2nd Wallnet Interreg Training Course
- 3rd Wallnet Interreg Training Course
- Doctoriales 2015

Formations de l'EdNBISE

- Mener sa thèse comme un projet
- Manager une équipe projet
- Manager l'innovation dans les organisations
- Propriété intellectuelle et ses enjeux
- Animer et conduire une réunion
- Maîtriser sa communication professionnelle
- Monter et mener un projet de recherche
- Anglais
- Biostatistiques
- Formations "mission enseignement"

Responsabilités

- Représentant élu des doctorants du laboratoire Glyco-MEV EA 4358 au conseil de laboratoire
- Représentant nommé des doctorants pour l'évaluation HCERES de l'école doctorale EDNBISE

Sommaire

Remerciements.....	3
Présentation du candidat.....	7
Sommaire.....	11
Liste des abréviations.....	17
Liste des figures.....	23
Liste des tableaux.....	31
Chapitre I : Synthèse bibliographique.....	35
1 Synthèse bibliographique.....	37
1.1 Microalgues.....	37
1.2 Diatomées.....	39
1.2.1 Structure Globale.....	39
1.2.2 Ultrastructure et composition pariétale.....	41
1.2.3 Reproduction.....	45
1.2.4 Sexualité.....	46
1.2.5 Division cellulaire.....	47
1.3 La diatomée <i>Phaeodactylum tricornutum</i>	49
1.3.1 Histoire.....	49
1.3.2 Particularité de sa plasticité morphologique.....	50
1.3.3 Structure.....	52
1.3.4 Morphogenèse.....	56
1.3.5 <i>P. tricornutum</i> : un organisme de référence.....	63
1.4 Processus de N-glycosylation dans les systèmes eucaryotes.....	66
1.4.1 Voie de biosynthèse du réticulum endoplasmique.....	67
1.4.2 Contrôle qualité des glycoprotéines.....	72
1.5 Voie de biosynthèse de l'appareil de Golgi.....	75
1.6 Localisation subcellulaire des glycosyltransférases golgiennes.....	80
1.7 Fucosylation des N-glycannes.....	82
1.7.1 Fucosyltransférases.....	82
1.7.2 Implication biologique.....	85
1.8 <i>Phaeodactylum tricornutum</i> , un modèle d'intérêt pour l'étude de la N-glycosylation.....	87
1.9 Objectifs de la thèse.....	101

Chapitre II : Matériel et méthodes.....	105
2 Matériel	107
2.1 Matériel végétal : <i>Phaeodactylum tricornutum</i>	107
2.2 Matériel animal : Cellules d'ovaire de hamster chinois (CHO)	107
2.3 Vecteurs plasmidiques utilisés	107
2.3.1 pUC57+FuT54599	107
2.3.2 pUC18+GalT.....	108
2.3.3 pcDNA3.1/V5-His-TOPO®.....	108
2.3.4 pJET1.2/blunt	109
2.3.5 pPha-NR	110
2.4 Amorces PCR	110
3 Méthodes.....	113
3.1 Culture de <i>P. tricornutum</i>	113
3.2 Culture des lignées cellulaires CHO.....	113
3.3 Analyses bioinformatiques	115
3.3.1 Analyse des séquençages	115
3.3.2 Analyse des séquences nucléotidiques	115
3.3.3 Analyse des séquences protéiques	115
3.4 Techniques de biologie moléculaire	116
3.4.1 Analyse des acides nucléiques	116
3.4.2 Clonage et expression des gènes d'intérêt.....	117
3.4.3 Analyse différentielle du transcriptome des trois morphotypes de <i>P. tricornutum</i> par RNA sequencing (RNA-Seq)	127
3.5 Localisation subcellulaire des glycosyltransférases de fusion par microscopie confocale	140
3.5.1 Préparation des échantillons	140
3.5.2 Paramètres d'observation.....	141
3.6 Localisation subcellulaire des glycosyltransférases de fusion par Microscopie Electronique à Transmission.....	141
3.6.1 Cryo-immobilisation haute pression	141
3.6.2 Réalisation des coupes ultrafines	143
3.6.3 Immunocytochimie	143
3.7 Techniques de biochimie.....	144
3.7.1 Extraction des protéines de <i>P. tricornutum</i> pour Western-blot.....	144
3.7.2 Western-blot.....	145

Chapitre III : Comparative transcriptome analysis of *P. tricornutum*'s morphotypes.149

4	Comparative transcriptome analysis of <i>P. tricornutum</i> 's morphotypes using in depth RNA sequencing reveals the oval morphotype features, especially that protein degradation and secretion is up-regulated in the oval morphotype.....	151
4.1	ABSTRACT.....	152
4.2	BACKGROUND.....	153
4.3	RESULTS.....	155
4.3.1	Experimental Design and mapping	155
4.3.2	Gene differential expression analysis.....	162
4.3.3	Functional profiling analysis.....	164
4.3.4	Intersection analysis reveals specific pathways which are up-regulated in the oval morphotype.....	172
4.4	DISCUSSION.....	180
4.5	MATERIELS AND METHODS.....	184
4.5.1	Culture and growth conditions.....	184
4.5.2	RNA extraction	184
4.5.3	Library preparation and RNA-sequencing.....	185
4.5.4	Bioinformatic analysis	185
4.5.5	qRT-PCR Validation.....	187
4.5.6	TEM microscopy for <i>P. tricornutum</i> morphotype characterization.....	187

Chapitre IV : Identification of reference genes for quantitative real-time PCR analysis in *Phaeodactylum tricornutum* morphotypes.....191

5	Identification of reference genes for quantitative real-time PCR analysis in <i>Phaeodactylum tricornutum</i> morphotypes	193
5.1	Abstract	193
5.2	Introduction/Background	194
5.2.1	qRT-PCR	194
5.2.2	Validation of reference genes for qRT-PCR data normalization.....	194
5.2.3	Identification of reference genes from transcriptomic datasets (microarray data and RNA-sequencing).....	195
5.2.4	Housekeeping genes of <i>P. tricornutum</i>	196
5.3	Main objectives	196
5.4	Material and Methods.....	197
5.4.1	Diatom cultures	197
5.4.2	RNA sequencing	198
5.4.3	qRT-PCR analysis	199
5.5	Results.....	202

5.5.1	Selection of candidate reference genes.....	202
5.5.2	qRT-PCR primer set validation	204
5.5.3	Expression stability of the tested reference genes and validation of reference genes	206
5.6	Discussion.....	209
Chapitre V : Caractérisation des fucosyltransférases putatives de <i>P. tricornutum</i>		213
6	Caractérisation des fucosyltransférases de <i>P. tricornutum</i>	215
6.1	Analyse bio-informatique des gènes PHATRDRRAFT_46109, PHATRDRRAFT_46110 et PHATRDRRAFT_54599 de <i>P. tricornutum</i>	216
6.1.1	Analyse du gène PHATRDRRAFT_54599.....	216
6.1.2	Analyse du gène PHATRDRRAFT_46109.....	217
6.1.3	Analyse du gène PHATRDRRAFT_46110.....	219
6.2	Comparaison des protéines putatives FuT46109NM, FuT46110NM et FuT54599 222	
6.3	Clonage des gènes putatifs codant potentiellement pour des α 1,3-fucosyltransférases.....	225
6.3.1	Clonage des gènes dans le vecteur de mammifère pcDNA3.1/V5.....	225
6.3.2	Clonage des gènes dans le vecteur de diatomée pPha-NR	226
6.3.3	Analyse des transformants obtenus	229
6.4	Localisation subcellulaire des enzymes GnT I et FuT54599 par microscopie électronique à transmission.....	238
7	Discussion et perspectives de la caractérisation des fucosyltransférases de <i>P. tricornutum</i>	242
Conclusion générale.....		251
Références bibliographiques.....		257
Annexe 1.....		285
Annexe 2.....		289
Annexe 3.....		327

Liste des abréviations

μL = Microlitre

μm = Micromètre

μM = Micromolaire

ADCC = Antibody-dependent cellular cytotoxicity

ADN = Acide désoxyribonucléique

ADNc = ADN complémentaire

ADNg = ADN génomique

ALG = Asparagine linked glycosylation

Arg = Arginine

ARN = Acide ribonucléique

ARNm = ARN messenger

Asn = Asparagine

ATP = Adénosine triphosphate

BP = Biological Process

BSA = Bovine serum albumin

Ca(II) = Ion calcium

CHO = Chinese hamster ovary

Co(II) = Ion cobalt

CO₂ = Dioxyde de carbone

CRDs = Carbohydrate-recognition domains

Ct = Cycles seuil

CTS = Cytoplasmic-transmembrane-stem

Cu(II) = Ion cuivre

Cys = Cystéine

DEG = Differentially Expressed Gene

Dol-P = Dolichol pyrophosphate

EDEM = ER degradation-enhancing α -mannosidase

EDTA = Éthylène diamine tétra-acétique

EGF = Epidermal growth factor

EPS = Substance exopolymérique

ERAD = Endoplasmic-reticulum-associated protein degradation

ERGIC = ER-Golgi intermediate compartment

EST = Expressed sequence tag
FBS = Fetal bovine serum
FDA = Food and drug administration
Fuc = Fucose
FucT = Fucosyltransférase
Fus = Fusiforme
Gal = Galactose
GalT = Galactosyltransférase
GDP = Guanosine diphosphate
GFP = Green fluorescent protein
Glc = Glucose
GlcNAc = N-acétylglucosamine
Glu = Glutamate
GnT I = N-acétylglucosaminyltransférase I
GnT II = N-acétylglucosaminyltransférase II
GO = Gene ontology
GOI = Gene of interest
GRAS = Generally recognized as safe
HA = Hémagglutinine
HGF = Hepatocyte growth factor
HRP = Peroxydase de raifort
IgG = Immunoglobuline G
IgM = Immunoglobuline M
ITS2 = Internal transcribed spacer 2
kb = Kilobase
KEGG = Kyoto Encyclopedia of Genes and Genomes
kV = Kilovolt
LBS = Lysis buffer solution
LCPAs = Longues chaînes de polyamines
Lys = Lysine
Man = Mannose
MET = Microscope électronique à transmission
min = Minute
mL = Millilitre

mM = Millimolaire

Mn(II) = Ion manganèse

Na⁺ = Ion sodium

ng = Nanogramme

NGS = Normal goat serum

NH₄⁺ = Ammonium

Ni(II) = Ion nickel

nm = Nanomètre

nM = Nanomolaire

NO₃⁻ = Nitrate

NR = Nitrate réductase

OST = Oligosaccharyltransférase

Ov = Ovale

P. tricornutum = *Phaeodactylum tricornutum*

Pb = Paire de base

PCR = Réaction en chaîne par polymérase

PDMPO = 2-(4-pyridyl)-5-((4-(2-diméthylaminoéthylaminocarbamoyl)méthoxy)phényl)oxazole

pM = Picomolaire

polyA = Polyadénylé

qPCR = PCR quantitative

QSP = Quantité suffisante pour

RE = Réticulum endoplasmique

RFP = Red fluorescent protein

RNAi = RNA interference

RNAse = Ribonucléase

RNA-Seq = RNA sequencing

rpm = Rotation par minute

RT-qPCR = Reverse transcription qPCR

SDS = Dodécylsulfate de sodium

SDS-PAGE = sodium dodecyl sulfate polyacrylamide gel electrophoresis

SDVs = Silica deposition vesicles

Ser = Sérine

SIT = Silicon transporter

TBE = Tris-Borate-EDTA

TBS = Tris-buffered saline

TBST = TBS + Tween 20

TGN = Trans Golgi Network

Thr = Thréonine

T_m = Melting temperature

TMD = Domaine transmembranaire

TMP = Thymidine monophosphate

Tri = Triradié

Trp = Tryptophane

Tyr = Tyrosine

UDP = Uridine diphosphate

UGGT = Glycoprotéine glucosyltransférase

UMP = Uridine monophosphate

XylT = Xylosyltransférase

YFP = Yellow fluorescent protein

Zn(II) = Ion zinc

Liste des figures

<i>Figure 1: Arbre phylogénétique des lignées eucaryotes d'après Keeling, 2013.</i>	36
<i>Figure 2: Origine endosymbiotique des diatomées d'après Keeling, 2013.</i>	37
<i>Figure 3: Structure du frustule des diatomées.</i>	38
<i>Figure 4: Diversité de la structure pariétale des diatomées.</i>	40
<i>Figure 5: Classement structural des diatomées.</i>	40
<i>Figure 6: Organisation ultrastructurale de la paroi des diatomées.</i>	41
<i>Figure 7: Structure des LCPAs retrouvées chez les diatomées.</i>	42
<i>Figure 8: Structure putative des polysaccharides pariétaux retrouvés chez P. tricornutum (Gügi et al., 2015).</i>	44
<i>Figure 9: Représentation des structures hypothétiques rencontrées chez Pinnularia viridis (Chiovitti et al., 2005).</i>	44
<i>Figure 10: Synthèse des hypothèques lors de la division cellulaire des diatomées.</i>	45
<i>Figure 11: Cycle de vie de la diatomée Seminavis robusta (Moeys et al., 2016).</i>	46
<i>Figure 12: Biogenèse de la paroi cellulaire des diatomées d'après Zurzolo and Bowler, 2001 ; Kröger and Poulsen 2008.</i>	48
<i>Figure 13: Présentation des trois morphotypes de P. tricornutum.</i>	49
<i>Figure 14: Prédilection morphologique des 10 souches de P. tricornutum.</i>	51
<i>Figure 15: Ultrastructure des trois morphotypes de P. tricornutum.</i>	52
<i>Figure 16: Structure des trois morphotypes de P. tricornutum.</i>	53
<i>Figure 17: Transition du morphotype fusiforme vers le morphotype ovale.</i>	55
<i>Figure 18: Transition du morphotype triradié vers le morphotype ovale.</i>	55
<i>Figure 19: Schéma des mitoses acytocinétiques impliquées dans la formation du morphotype ovale (Tesson et al., 2009).</i>	57
<i>Figure 20: Transformation du morphotype fusiforme en morphotype ovale (Tesson et al., 2009).</i>	57
<i>Figure 21: Conversion du morphotype ovale en morphotype triradié et fusiforme d'après Tesson et al., 2009 ; De Martino et al., 2011.</i>	59
<i>Figure 22: Etapes de la conversion du morphotype triradié en morphotype fusiforme d'après De Martino et al., 2007; De Martino et al., 2011.</i>	59
<i>Figure 23: Résumé des relations intermorphotypes de P. tricornutum d'après De Martino et al., 2011.</i>	60
<i>Figure 24: Nombre de publications par années avec le mot clé « Diatom ».</i>	63

<i>Figure 25: Voie de biosynthèse du précurseur oligosaccharidique Glc3Man9GlcNAc2.....</i>	<i>68</i>
<i>Figure 26: Transfert de l'oligosaccharide précurseur Glc3Man9GlcNAc2 sur un polypeptide en cours d'élongation par l'oligosaccharyltransférase (OST) d'après Schwarz and Aebi, 2011.....</i>	<i>70</i>
<i>Figure 27: Repliement et cycle de contrôle qualité des glycoprotéines N-liées dans le réticulum endoplasmique (RE) d'après Williams, 2006 ; Lamriben et al., 2015.....</i>	<i>73</i>
<i>Figure 28: Structure d'une glycosyltransférase golgienne, protéine membranaire de type II (Czlapinski and Bertozzi, 2006).</i>	<i>76</i>
<i>Figure 29: Biosynthèse et maturation des N-glycannes dans le réticulum endoplasmique et l'appareil de Golgi.</i>	<i>77</i>
<i>Figure 30: Exemples de structures N-glycanniques complexes rencontrées chez différents eucaryotes, d'après Schiller et al., 2012.</i>	<i>79</i>
<i>Figure 31: Localisation subcellulaire des glycosyltransférases et glycosidases golgiennes impliquées dans la maturation des N-glycannes complexes d'après Wildt and Gerngross, 2005 ; Schoberer et al., 2013.</i>	<i>81</i>
<i>Figure 32: Représentation schématique des séquences des fucosyltransférases eucaryotiques GT10 montrant les positions relatives des motifs conservés décrits dans la littérature selon Mollicone et al. 2009 (Bleu) et Both et al., 2011 (jaune).</i>	<i>83</i>
<i>Figure 33: Cristallographie de l'interaction entre l'α-1,3-FucT d'H. pylori et son substrat donneur, le GDP-fucose, d'après Sun et al., 2007.....</i>	<i>84</i>
<i>Figure 34: Carte du vecteur pUC57+FuT54599.....</i>	<i>107</i>
<i>Figure 35: Carte du vecteur pUC57+FuT54599.....</i>	<i>108</i>
<i>Figure 36: Carte du vecteur pcDNA3.1/V5-His-TOPO®.....</i>	<i>109</i>
<i>Figure 37: Carte du vecteur pJET1.2/blunt.....</i>	<i>109</i>
<i>Figure 38: Carte du vecteur pPhaNR.....</i>	<i>110</i>
<i>Figure 39: Cinétique de croissance des trois morphotypes de P. tricornutum.....</i>	<i>113</i>
<i>Figure 40: Schéma représentant la stratégie de clonage dans le vecteur pPha-NR.....</i>	<i>122</i>
<i>Figure 41: Schéma présentant le processus de préparation des échantillons pour la constitution d'une librairie RNA-seq (TruSeq Stranded mRNA kit ; Illumina)</i>	<i>128</i>
<i>Figure 42: Schéma présentant le protocole de purification et de fragmentation des ARNm</i>	<i>129</i>
<i>Figure 43: Schéma du « pipeline » bioinformatique utilisé pour réaliser l'analyse fonctionnelle du transcriptome des trois morphotypes de P. tricornutum.....</i>	<i>133</i>
<i>Figure 44: Programme du thermocycleur utilisé pour la qPCR</i>	<i>138</i>
<i>Figure 45: Les différentes formes de l'eau dans une cellule.</i>	<i>142</i>

Figure 46: Marqueur de taille protéique BP3603 Series (Fisher)	145
Figure 47: Micrographs of <i>P. tricornutum</i> dominant morphotypes present in the 3 enriched cultures of the Pt3 strain.	155
Figure 48: Cells and morphotypes count in the cultures of <i>P. tricornutum</i> over a 12 days period.	156
Figure 49: Summary of the experimental design and analytical pipeline used in this study.....	157
Figure 50: Histogram representing the percentages of read alignment with TopHat for the 4 biological replicates of the cultures enriched in the fusiform, triradiate and oval morphotype, respectively.....	158
Figure 51: Correlations between biological replicates and normalization performance with DEseq2.	159
Figure 52: Heatmap illustrating the correlation between the 4 biological replicates for each culture specifically enriched in either the fusiform (F), triradiate (T) or oval (O) morphotype.....	160
Figure 53: Validation of the RNA-Seq data using qRT-PCR analysis.....	161
Figure 54: Vulcano Plots of pairwise comparisons of (A) T versus F (TF); (B) O versus F (OF); (C) O versus T (OT).....	162
Figure 55: Graph representing the distribution of the significantly differentially expressed genes (DEG) in relation to their log ₂ fold change values.....	163
Figure 56: Pie charts representing the biological processes which are alliterated in the TF pairwise comparison.	165
Figure 57: Pie charts representing the molecular functions which are alliterated in the TF pairwise comparison.	166
Figure 58: Pie charts representing the biological processes which are alliterated in the OF pairwise comparison.	167
Figure 59: Pie charts representing the molecular functions which are alliterated in the OF pairwise comparison.	168
Figure 60: Pie charts representing the biological processes which are alliterated in the OT pairwise comparison.	170
Figure 61: Pie charts representing the molecular functions which are alliterated in the OT pairwise comparison.	171
Figure 62: Venn diagram showing the distribution of the DE genes between the three pairwise comparisons.....	172

<i>Figure 63: Pie charts representing the biological processes and molecular functions which are alliterated in the 52 common genes between TF, OF and OT.....</i>	<i>173</i>
<i>Figure 64: Pie charts representing the biological processes and molecular functions which are alliterated in the 302 genes included exclusively in OF.....</i>	<i>174</i>
<i>Figure 65: Mapping of the DE genes exclusive to the oval morphotype on metabolic pathways.</i>	<i>177</i>
<i>Figure 66: Mapping of the DE genes exclusive to the oval morphotype on the SNARE interactions in vesicular transport pathway.....</i>	<i>179</i>
<i>Figure 67: FastQC report depicting per base sequence quality for the 4 biological replicates of the respective enriched cultures for either the fusiform, triradiate or oval morphotype.</i>	<i>185</i>
<i>Figure 68: Stability ranking of the reference genes with RefFinder.....</i>	<i>207</i>
<i>Figure 69: Histogram representing the ratio of normalized gene expression for genes belonging to the N-glycosylation pathway comparing the oval and fusiform morphotypes and the triradiate and fusiform morphotypes respectively.....</i>	<i>208</i>
<i>Figure 70: Prédiction de domaines transmembranaires sur la séquence protéique FuT54599 par les programmes TMHMM (A), TMPred (B) et DAS (C).....</i>	<i>216</i>
<i>Figure 71: Prédiction de domaines transmembranaires sur la séquence protéique de l'ancienne version de la FuT46109 (FuT46109OM) par les programmes TMHMM (A), TMPred (B) et DAS (C).</i>	<i>217</i>
<i>Figure 72: Comparaison des portions N-terminales des séquences protéiques de la fucosyltransférase 46109 entre la nouvelle prédiction (FuT46109NM) et l'ancien modèle (FuT46109OM) par alignement avec le programme Clustal Omega.....</i>	<i>218</i>
<i>Figure 73: Prédiction de domaines transmembranaires sur la séquence protéique de la nouvelle version de la FuT46109 (46109NM) par les programmes TMHMM (A), TMPred (B) et DAS (C).</i>	<i>219</i>
<i>Figure 74: Prédiction de domaines transmembranaires sur la séquence protéique de l'ancienne version de la FuT46110 (FuT46110OM) par les programmes TMHMM (A), TMPred (B) et DAS (C).</i>	<i>220</i>
<i>Figure 75: Contrôle de l'épissage des introns de la séquence 46110OM par PCR.....</i>	<i>221</i>
<i>Figure 76: Comparaison des portions N-terminales des séquences protéiques de la fucosyltransférase 46110 entre la nouvelle prédiction (FuT46110NM) et l'ancien modèle (FuT46110OM) par alignement avec le programme Clustal Omega.....</i>	<i>221</i>

<i>Figure 77: Prédiction de domaines transmembranaires sur la séquence protéique de la nouvelle version de la FuT46110 (FuT46110NM) par les programmes TMHMM (A), TMpred (B) et DAS (C).</i>	222
<i>Figure 78: Représentations schématiques réalisées d'après l'outil de prédiction Pfam des trois fucosyltransférases putatives de P. tricornutum</i>	223
<i>Figure 79: Alignement des séquences protéiques correspondant aux domaines Pfam GT10 d'α1,3-fucosyltransférases caractérisées, ainsi que des séquences FuT46109NM, FuT46110NM et FuT54599 réalisé avec T-COFFEE v11</i>	224
<i>Figure 80: Représentation schématique des constructions génétiques FuT46109NM+pcDNA, FuT46110NM+pcDNA et FuT54599+pcDNA</i>	225
<i>Figure 81: Représentation schématique des constructions génétiques FuT46109NM-GFP et -V5, FuT46109OM-GFP, FuT46110NM-GFP et -V5, FuT46110OM-GFP, FuT54599-GFP et -V5, GnTI-GFP et -V5, GalT-GFP et -HA, ainsi que XylT-RFP et -cmcy dans le vecteur pPha-NR</i>	227
<i>Figure 82: Criblage PCR des co-transformants FuT54599-V5 et XylT-cmcy</i>	230
<i>Figure 83: Criblage des transformants GnT I-V5 par PCR et western-blot</i>	231
<i>Figure 84: Criblage par western-blot des co-transformants FuT54599-V5 et XylT-cmcy</i>	232
<i>Figure 85: Criblage des co-transformants GnT I-V5 et XylT-cmcy par PCR et western-blot</i>	232
<i>Figure 86: Observation en microscopie confocale des cultures de P. tricornutum transfectées avec les constructions GnT I-GFP, GalT-GFP et XylT-RFP</i>	233
<i>Figure 87: Observation en microscopie confocale de la culture de P. tricornutum co transfectée avec les constructions GnT I-GFP et XylT-RFP</i>	234
<i>Figure 88: Observation en microscopie confocale des cultures de P. tricornutum transformées avec les constructions FuT54599-GFP, FuT46110NM-GFP, FuT46109NM-GFP et FuT46109OM-GFP</i>	235
<i>Figure 89: Contrôle de la transgenèse dans les lignées de P. tricornutum transformées avec la construction FuT46110OM-GFP</i>	236
<i>Figure 90: Traduction prédite à partir de la séquence nucléotidique non épissée de l'ancienne version du gène 46110 (FuT46110OM)</i>	237
<i>Figure 91: Observation en microscopie confocale de la culture de P. tricornutum co-transfectée avec les constructions FuT54599-GFP et XylT-RFP</i>	238
<i>Figure 92: Observation en microscopie électronique à transmission de l'organisation ultrastructurale de la microalgue Phaeodactylum tricornutum Pt1.8.6</i>	239

<i>Figure 93: Observation en microscopie électronique à transmission de la localisation golgienne de la GnT I-V5 après immunomarquage avec un anticorps primaire dirigé contre l'épitope V5, et un anticorps secondaire couplé à une particule d'or de 10nm.....</i>	<i>240</i>
<i>Figure 94: Observation en microscopie électronique à transmission de la localisation golgienne de la FuT54599-V5 après immunomarquage avec un anticorps primaire dirigé contre l'épitope V5, et un anticorps secondaire couplé à une particule d'or de 10nm.....</i>	<i>241</i>
<i>Figure 95: Visualisation avec le logiciel IGV du profil d'expression du gène PHATRDRAFT_46109 issu de l'analyse RNA-seq effectuée sur les trois morphotypes de P. tricornutum.....</i>	<i>243</i>
<i>Figure 96: Visualisation avec le logiciel IGV du profil d'expression du gène PHATRDRAFT_46110 issu de l'analyse RNA-seq effectuée sur les trois morphotypes de P. tricornutum.....</i>	<i>244</i>
<i>Figure 97: Schéma représentant la stratégie de complémentation fonctionnelle de cellules CHO FuT8-/- avec les fucosyltransférases de P. tricornutum.....</i>	<i>248</i>

Liste des tableaux

<i>Table 1: Liste des amorces utilisées pour effectuer les amplifications par PCR et les réactions de séquençage</i>	111
<i>Table 2: Liste des amorces utilisées pour la validation des données de RNA-seq par qPCR</i>	112
<i>Table 3: Composition des milieux de culture de P. tricornutum</i>	114
<i>Table 4: Composition du milieu Conway en fonction du type de cellule cultivé</i>	114
<i>Table 5: Composition des mélanges réactionnels de PCR</i>	118
<i>Table 6: Conditions de PCR en fonction de la polymérase utilisée</i>	119
<i>Table 7: Mélange de ligation réalisé pour le clonage dans le vecteur pJET1.2/blunt</i>	120
<i>Table 8: Mélange réactionnel utilisé pour la digestion des fragments d'ADN par les enzymes de restriction</i>	120
<i>Table 9: Couples d'enzymes de restrictions utilisées pour extraire le fragment d'ADN considéré</i>	121
<i>Table 10: Composition des solutions utilisées dans le protocole d'extraction plasmidique « maison »</i>	123
<i>Table 11: Composition du mélange de restriction utilisé pour le criblage des vecteurs pcDNA3.1</i>	124
<i>Table 12: Composition du mélange de restriction utilisé pour la linéarisation des vecteurs pcDNA3.1</i>	125
<i>Table 13: Plan de charge des échantillons indexés sur la flowcell</i>	130
<i>Table 14: Amorces utilisées pour quantifier la librairie par qPCR</i>	131
<i>Table 15: Programme du thermocycleur utilisé pour quantifier la librairie par qPCR</i>	132
<i>Table 16: Mélanges réactionnels utilisés pour la reverse transcription</i>	134
<i>Table 17: Programme du thermocycleur utilisé pour la reverse transcription</i>	135
<i>Table 18: Milieu réactionnel réalisé pour le contrôle qualité des extraits d'ARN par amplification PCR</i>	135
<i>Table 19: Programme du thermocycleur utilisé pour le contrôle qualité des extraits d'ARN par amplification PCR</i>	136
<i>Table 20: Séquences des amorces utilisées pour le contrôle qualité des extraits d'ARN par amplification PCR</i>	136
<i>Table 21: Mix de réaction utilisé pour la validation des amorces par qPCR</i>	137
<i>Table 22: Composition des gels de concentration et de séparation pour SDS-PAGE</i>	145

<i>Table 23: Metric of mapping from TopHat for the 4 biological replicates of the cultures enriched in the fusiform, triradiate and oval morphotype, respectively.</i>	<i>158</i>
<i>Table 24: Results of the matrix distance computation performed after the normalization with DEseq2.</i>	<i>160</i>
<i>Table 25: Comparative table highlighting the RNA-Seq log2 fold change and the qPCR log2 ratio for the 22 genes randomly selected among the DGE list obtained from the TF; OF; OT pairwise comparison of the RNA-Seq analysis.</i>	<i>161</i>
<i>Table 26: List of the genes significantly differentially expressed (adjusted p-value below 0.01) in the OF, OT and TF pairwise comparisons.</i>	<i>162</i>
<i>Table 27: Lists of the significantly differentially expressed genes associated with GO terms which have been obtained from the Venn diagram as a result of the intersection analysis.</i>	<i>172</i>
<i>Table 28: List of the candidate housekeeping genes, with the primer sequences used in qRT-PCR analysis and the amplicon size.</i>	<i>203</i>
<i>Table 29: List of the candidate new reference genes selected from a RNA-Seq data analysis, with the primer sequences used in qRT-PCR analysis and the amplicon size.</i>	<i>204</i>
<i>Table 30: List of the candidate reference genes tested for the normalization of qRT-PCR expression studies in the 3 morphotypes of Phaeodactylum tricornutum.</i>	<i>205</i>

Chapitre I : Synthèse Bibliographique

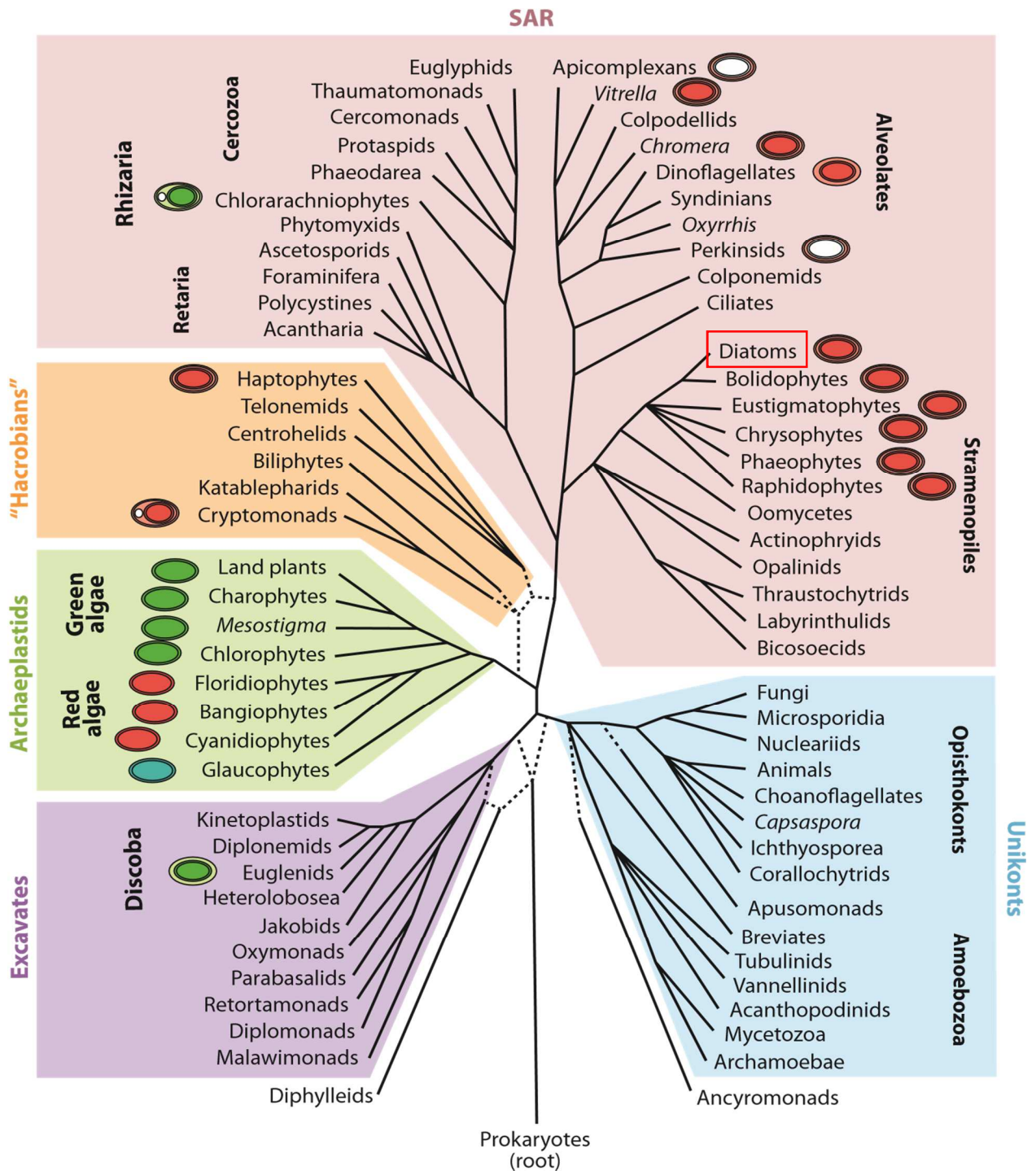


Figure 1 : Arbre phylogénétique des lignées eucaryotes d'après Keeling, 2013.

Les relations entre les différentes lignées eucaryotes ont été établies après synthèse de différentes études basées sur des analyses phylogénomiques à grande échelle, ou des caractéristiques génomiques rares. Les lignes pointillées représentent les relations particulièrement litigieuses au sein de la communauté scientifique. Les différentes possibilités faisant débat ont été tracées. La lignée des diatomées à laquelle appartient *Phaeodactylum tricorutum* a été encadrée en rouge.

1 Synthèse bibliographique

1.1 Microalgues

Les microalgues forment un groupe polyphylétique d'organismes microscopiques et photosynthétiques. Ce terme regroupe des organismes très variés, aussi bien unicellulaires que pluricellulaires, répartis sur des phylums très différents allant de la cyanobactérie (procaryote) aux microalgues eucaryotes. Apparues il y a environ 3,5 milliard d'années sous la forme de cyanobactéries, elles ont ensuite évolué au gré des endocytoses, apparaissant puis se répartissant dans les différents embranchements de l'arbre phylogénétique des eucaryotes. Actuellement, seul le supergroupe des Unikontes, auquel appartiennent les animaux, est exempt de microalgues (**figure 1** ; Keeling, 2013).

Les diatomées ou Bacillariophycées, sont des microalgues eucaryotes unicellulaires appartenant au groupe des Straménopiles. Il s'agit d'un des groupes phytoplanctoniques les plus riches comprenant une population estimée entre 30 000 et 100 000 espèces (Mann and Vanormelingen, 2013). Les diatomées sont apparues il y a environ 280 millions d'années à la suite d'un événement endosymbiotique (**figure 2** ; Medlin, 2000; Moustafa et al., 2009).

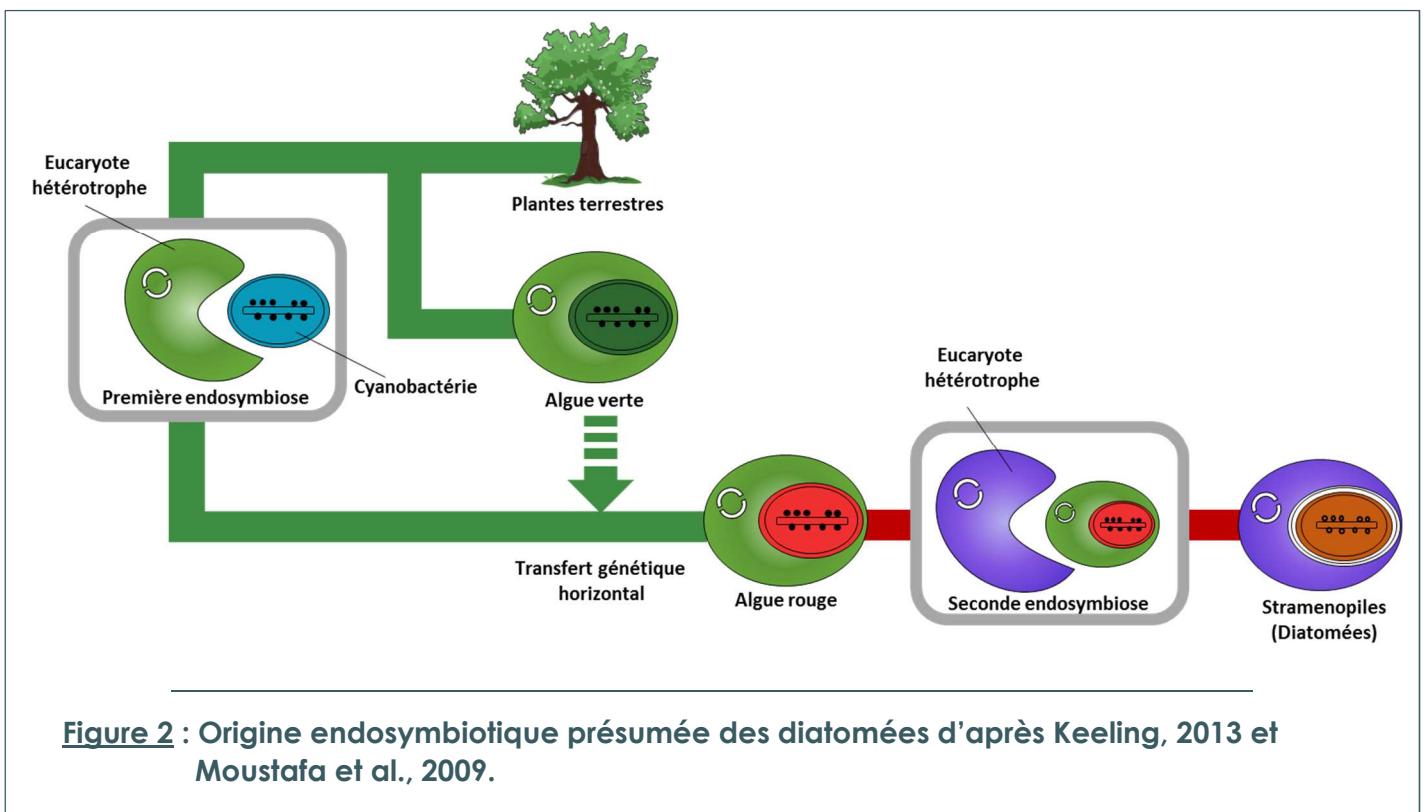
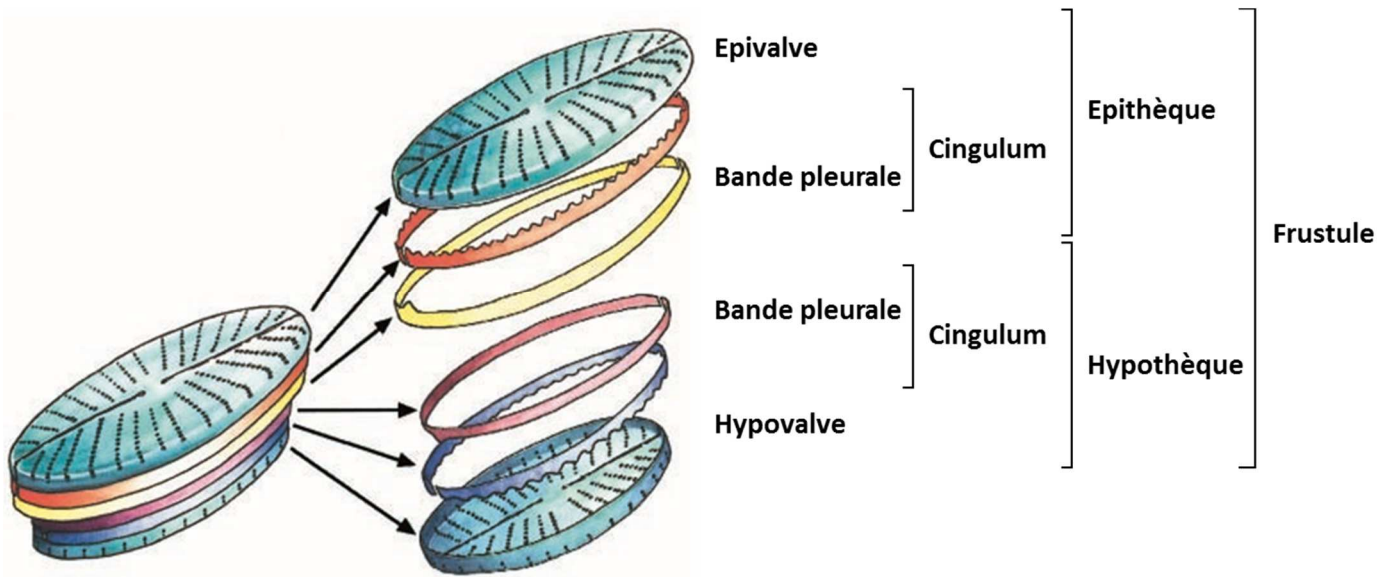


Figure 2 : Origine endosymbiotique présumée des diatomées d'après Keeling, 2013 et Moustafa et al., 2009.

A



B

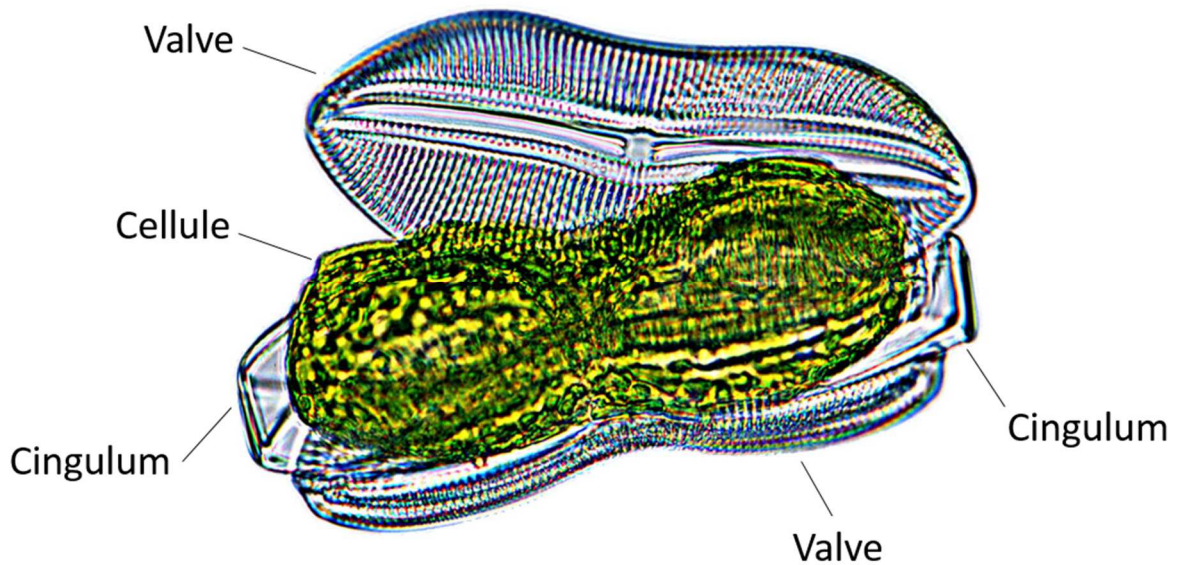


Figure 3 : Structure du frustule des diatomées.

A = Schéma représentant la structure pariétale des diatomées d'après Zurzolo and Bowler, 2001.

B = Photographie en microscopie optique de la diatomée *Diploneis beyrichiana* dont le frustule s'est ouvert accidentellement. Image extraite du site <http://www.diatomloir.eu/Site%20Diatom/>

De manière intéressante, les analyses génétiques réalisées chez les diatomées ont révélé une proximité inattendue avec les Metazoa (animaux) (Zurzolo and Bowler, 2001). Ainsi, d'après les analyses de données phylogénomiques et ultrastructurales, les diatomées sont placées dans un phylum proche de celui des Alveolates et éloignées des lignées vertes et rouges (Baldauf et al., 2000; Keeling, 2013). On peut ainsi pratiquement considérer les diatomées comme des animaux photosynthétiques, plutôt que des plantes unicellulaires (Zurzolo and Bowler, 2001). Leur position phylogénétique, à mi-chemin entre les plantes terrestres et les animaux, en font des organismes très intéressants pour l'étude de l'évolution des différentes voies métaboliques (Mathieu-Rivet et al., 2014).

1.2 Diatomées

1.2.1 Structure Globale

Selon l'espèce considérée, les diatomées peuvent présenter une taille cellulaire allant de 5µm à 5mm. Certaines de ces algues sont aussi capables de s'associer en formant des chaînes ou des colonies (Lee, 1999). A l'exception de quelques lignées, seules les gamètes mâles sont flagellés chez les diatomées. Chaque cellule est entourée par une paroi appelée frustule. Ce dernier est caractérisé par un squelette de silice amorphe ($\text{SiO}_2\text{nH}_2\text{O}$) (Martin-Jezequel et al., 2000; Wetherbee, 2002) associé à une enveloppe organique essentiellement composée de polysaccharides et de protéines (Hecky et al., 1973). Ce frustule est composé de deux parties imbriquées l'une dans l'autre : une plus grande, l'épithèque, qui chevauche une plus petite, l'hypothèque, à la manière d'une boîte de Pétri. Chaque thèque est ainsi composée à la fois d'une valve (épivalve ou hypovalve), qui correspond au fond de la boîte, et du cingulum, formé d'un ou plusieurs anneaux de silice, qui correspond à la paroi verticale (**figure 3**). La forme du frustule est très variable en fonction de l'espèce considérée. Doté de nanostructures génétiquement déterminées (Kröger, 2007), le frustule est devenu un critère pour l'identification et la classification des espèces (**figure 4**).

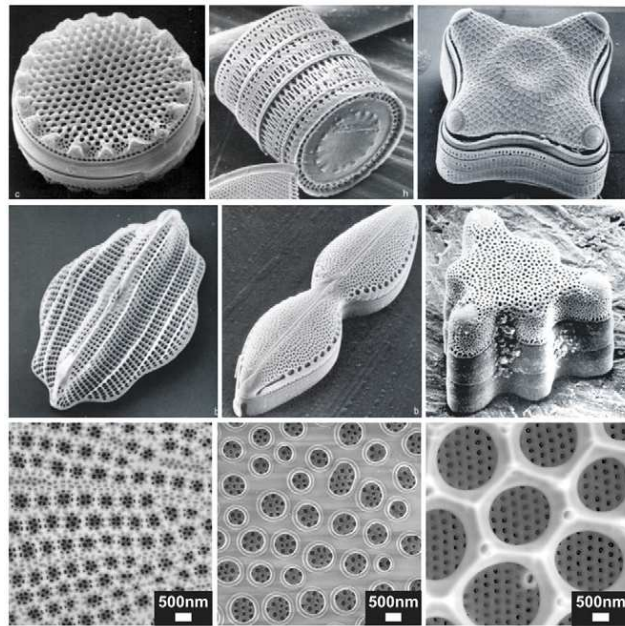


Figure 4 : Diversité de la structure pariétale des diatomées.

Prises de vue globales et ultrastructurales du frustule de différentes espèces de diatomées en microscopie électronique à balayage (Kröger and Poulsen, 2008).

Les diatomées ont ainsi été séparées en deux grands groupes facilement distinguables : d'un côté, les diatomées centriques qui possèdent une symétrie radiale, et d'un autre côté, les diatomées pennées qui possèdent une symétrie bilatérale. Les organismes modèles utilisés pour représenter ces grandes classes sont respectivement *Thalassiosira pseudonana* et *Phaeodactylum tricornutum*. Les classifications centrique et pennée sont elles-mêmes subdivisées en deux groupes : les diatomées centriques radiales et polaires, et les diatomées pennées raphidées et araphidées (**Figure 5** ; Kröger and Poulsen, 2008).

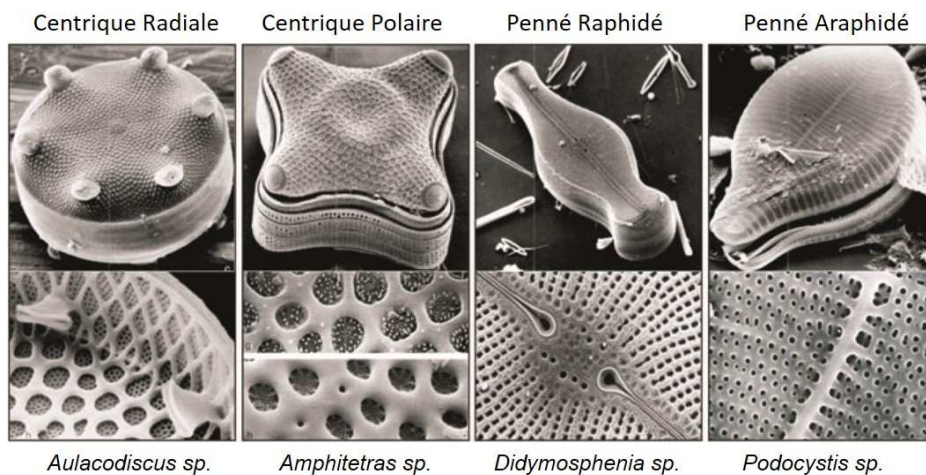


Figure 5 : Classement structural des diatomées.

Photographies en microscopie électronique à balayage (Kröger and Poulsen, 2008).

1.2.2 Ultrastructure et composition pariétale

La paroi cellulaire des diatomées observée au microscope électronique présente trois couches : une couche externe organique ondulée enrobée de mucilage sécrété, une couche intermédiaire minéralisée silicifiée contenant de la matière organique, et une dernière couche organique interne : le diatotepum (**figure 6** ; Borowitzka and Volcani, 1978).

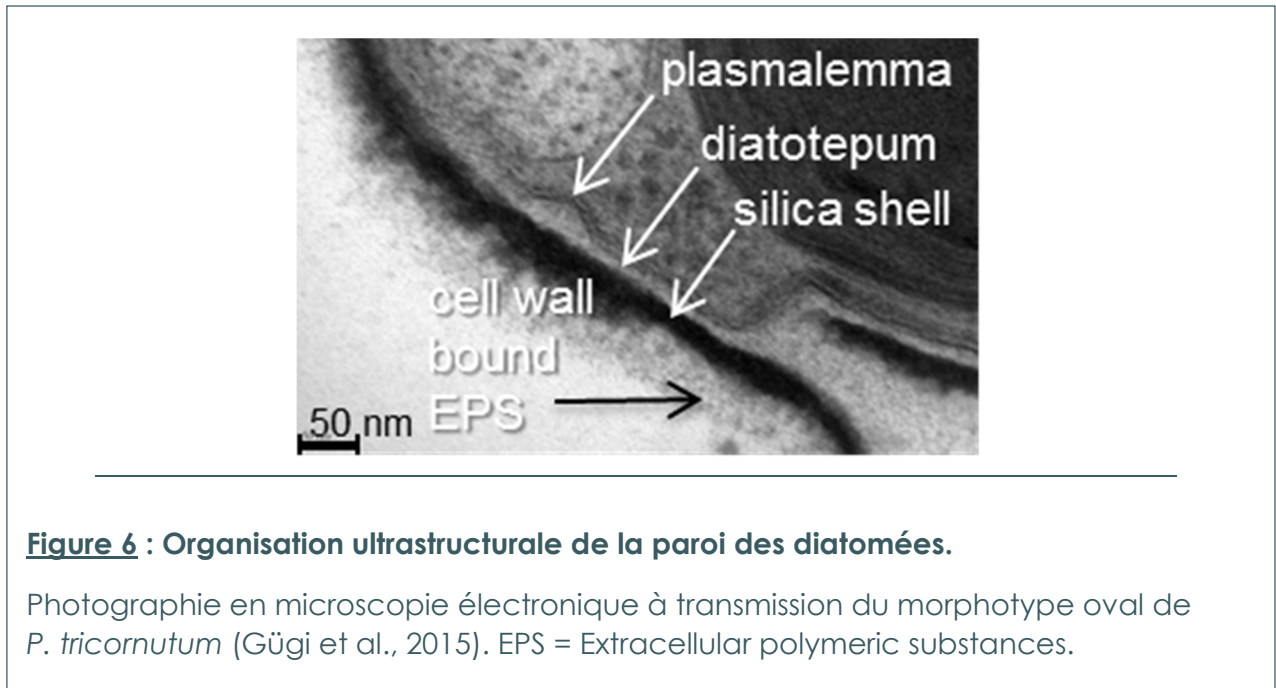


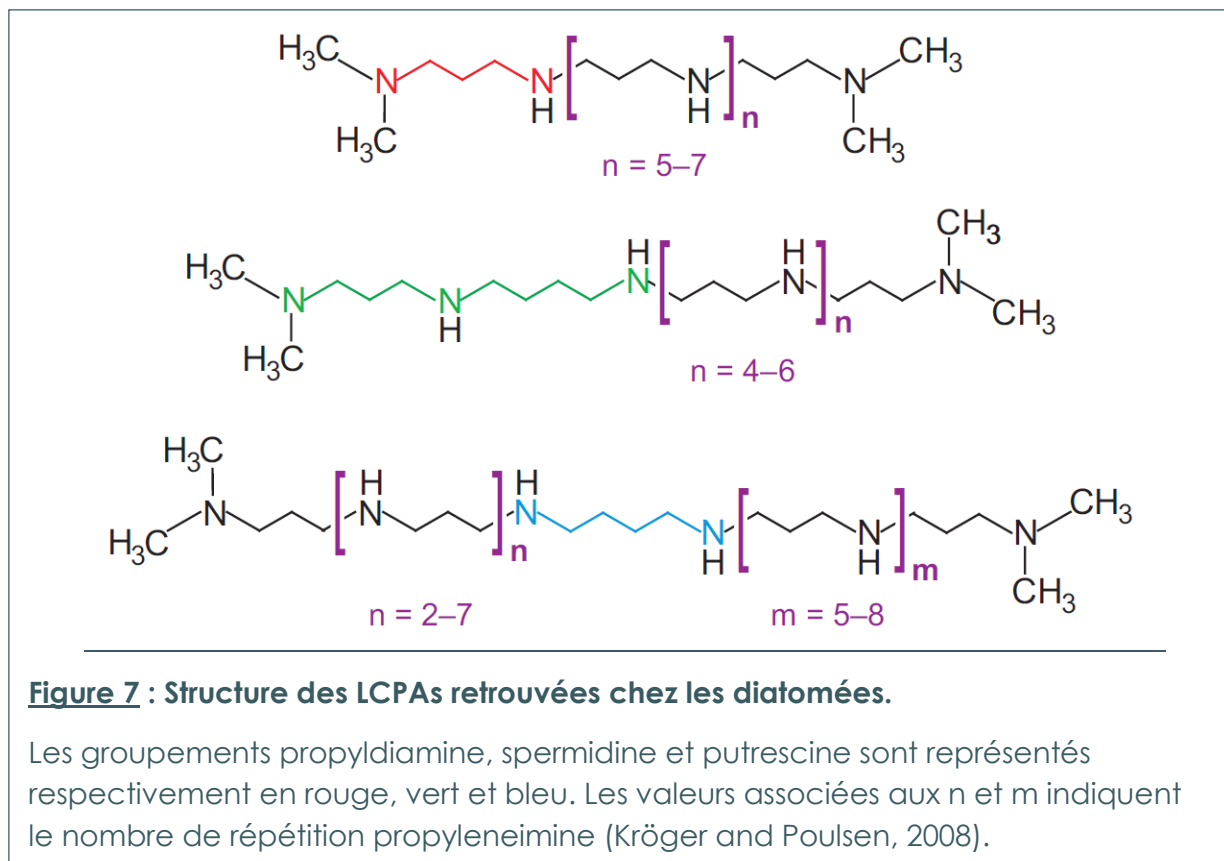
Figure 6 : Organisation ultrastructurale de la paroi des diatomées.

Photographie en microscopie électronique à transmission du morphotype oval de *P. tricornutum* (Gügi et al., 2015). EPS = Extracellular polymeric substances.

Cependant, bien que ces couches apparaissent bien distinctes, il semble très difficile de les séparer par traitement physique ou chimique (Borowitzka and Volcani, 1978). La couche organique externe est la première à apparaître pendant la formation pariétale, puis les deux autres vont venir se déposer en dessous. La couche du milieu contenant la silice semble ainsi être la paroi mature, tandis que le diatotepum, se situant au contact de la membrane plasmique, a été décrit comme étant composé de matériel fibrillaire avec une matrice amorphe, correspondant vraisemblablement à la zone de dépôt de la paroi nouvellement synthétisée (Borowitzka and Volcani 1978). Différentes études ont été menées afin de connaître la composition pariétale des diatomées. Cependant, les interactions fortes présentes entre les différentes couches rendent difficile l'attribution spécifique d'un composé à l'une ou l'autre des couches. Ainsi, nous envisagerons pour la suite du descriptif la paroi dans sa globalité.

Afin de réaliser la couche de biosilice qui compose leur frustule, les diatomées importent de l'acide monosilicique $\text{Si}(\text{OH})_4$ présent naturellement dans les océans (Tréguer et al., 1995) via des transporteurs Na^+ -dépendant (SIT = Silicon Transporter) (Hildebrand et al., 1997, 1998). La formation de la biosilice se déroule ensuite dans des vésicules intracellulaires spécifiques, les SDVs (Silica Deposition Vesicles), où elle va précipiter. Cependant, les mécanismes reliant le transport de

la silice à sa polymérisation, à savoir le stockage de l'acide silicique soluble ainsi que son mode de transport dans les SDVs, restent encore méconnus. Les SDVs vont ensuite aller fusionner avec la membrane cellulaire afin de libérer leur contenu au niveau de la paroi (Sumper and Kröger, 2004). La paroi des diatomées n'est pas uniquement constituée de silice, on y retrouve aussi des protéines et des polysaccharides. Trois principales familles de protéines ont pu être identifiées : les frustulines, les pleuralines et les silaffines. En plus de ces dernières, de très longues chaînes de polyamines (LCPAs) ont été mises en évidence (Sumper and Kröger, 2004). Les silaffines et les polyamines exercent une grande influence sur la formation de la silice *in vitro* en catalysant sa polymérisation sous forme de billes (Wetherbee, 2002), elles apparaissent ainsi directement impliquées dans la biogenèse de la silice. Les silaffines et les polyamines sont étroitement associées avec la silice et ne peuvent être resolubilisées qu'en éliminant complètement la silice (Sumper and Kröger, 2004). Ces polyamines pariétales sont les plus longues trouvées dans la nature, et chaque espèce de diatomée en possède une collection spécifique variant par la longueur de leur chaîne ou par leurs motifs de méthylation. Les LCPAs sont ainsi constituées de longues chaînes linéaires d'oligopropylèneimines liées à des propylendiamines, putrescines, ou spermidines. Leurs atomes d'azote peuvent présenter différents degrés de méthylation, permettant dans certains cas la création d'azote quaternaire (**figure 7** ; Kröger, 2007).



La reproduction fidèle de l'architecture des frustules entre les cellules d'une même espèce semble indiquer une transmission génétique. Chaque espèce de diatomée possédant son propre jeu de polyamines, et sachant que la longueur des chaînes de polyamine est liée à la taille des billes de silice polymérisée (Kröger et al., 2000), les variations de structure des LCPAs entre les différentes espèces seraient probablement un des liens entre la génétique et la formation des motifs pariétaux espèce-spécifiques (Zurzolo and Bowler, 2001). Des études menées par microscopie électronique à transmission ont montré que les frustulines (Poll et al., 1999) et les pleuralines (Kröger and Wetherbee, 2000) se retrouvent associées à la biosilice seulement après sa formation, ce qui laisse penser qu'elles ne sont pas impliquées dans sa genèse. Elles auraient plutôt un rôle structural, ainsi qu'une implication dans le développement du frustule.

Chez les diatomées, la partie organique de la paroi est aussi composée de polysaccharides polyanioniques (Chiovitti et al., 2005; McConville et al., 1999; Round et al., 1990). La composition monosaccharidique pariétale a été étudiée chez différentes diatomées montrant que les frustules contiennent une gamme assez large de monosaccharides, avec des sucres neutres (comme du glucose, mannose, galactose, xylose, arabinose, rhamnose, fucose), des sucres acides (acide glucuronique, acide galacturonique, acide 2-O-méthylglucuronique, acide mannuronique) et des sucres neutres O-méthylés (2-O-méthylrhamnose, 3-O-méthylrhamnose, 2,3-di-O-méthylrhamnose, 3-O-méthylxylose, 4-O-méthylxylose) (Abdullahi et al., 2006; Chiovitti et al., 2005). Ces études montrent que le mannose est le sucre majoritaire, sa fraction molaire pouvant dépasser parfois les 60%, supposant ainsi que le polysaccharide pariétal principal serait un mannane (Chiovitti et al., 2005; Ford and Percival, 1965). Il a aussi été mis en évidence la présence de chitine (poly-N-acétyl-D-glucosamine) chez *Cyclotella* et *Thalassiosira* (Blackwell et al., 1967). Cependant, il y a actuellement très peu d'informations disponibles quant à la structure détaillée des polysaccharides hydrolysés pour obtenir ces compositions (Gügi et al., 2015).

Beaucoup de données concernant la composition pariétale sont disponibles pour *P. tricornutum* qui posséderait dans sa paroi de l'arabinose, du fucose, du galactose, du glucose, du rhamnose, du xylose, de l'acide mannuronique et de l'acide galacturonique (Gügi et al., 2015). Cependant, Ford et Percival (1965) ont montré que la paroi de *P. tricornutum* est principalement composée de mannose, d'acide glucuronique et de groupements sulfate, formant ainsi du glucuronomannane sulfaté. Ce polysaccharide serait ainsi composé d'une chaîne principale de mannopyranoses liés en 1,3 ramifiée par des chaînes d'acide glucopyranosyluronic (1-3) mannopyranose, ou d'acide glucopyranosyluronic (1-3) mannopyranosyl (1-2) mannopyranose (**figure 8**).

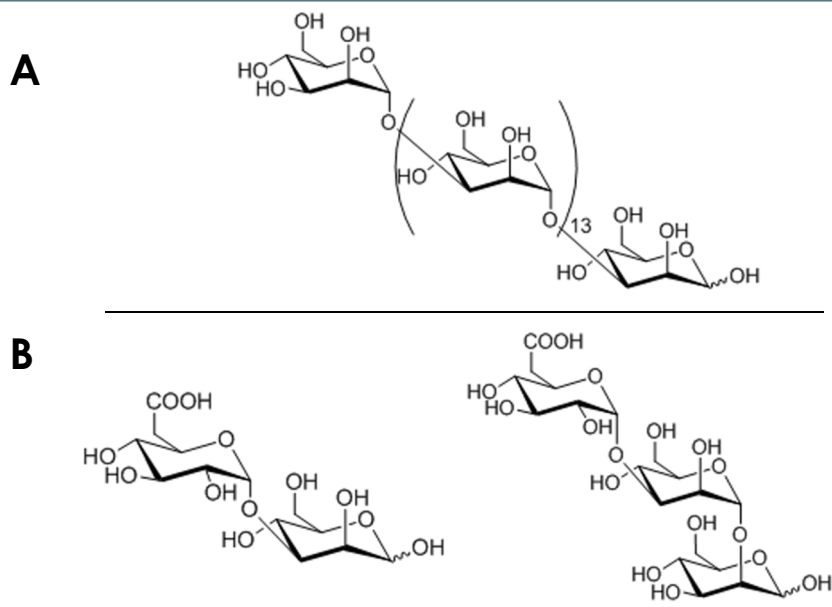


Figure 8 : Structure putative des polysaccharides pariétaux retrouvés chez *P. tricornutum* (Gügi et al., 2015).

Bien que les liaisons entre monosaccharides soient représentées en α , l'anométrie n'est pas encore déterminée. A = représentation de la chaîne principale d'1,3-mannopyranose ; B = ramifications d'acide glucopyranosyluronique (1-3) mannopyranose, et d'acide glucopyranosyluronique (1-3) mannopyranosyl (1-2) mannopyranose.

Même si cela n'est pas encore complètement confirmé, il semblerait que ces chaînes latérales puissent être liées au niveau du carbone 2 de certains mannoses de la chaîne principale, tout comme les groupements sulfates (Abdullahi et al., 2006). Cette hypothèse peut être appuyée par des travaux réalisés sur le glucuronomannane d'une autre diatomée, *Pinnularia viridis*, qui montrent que cette diatomée possède des chaînes latérales liées sur le carbone 2, et des sulfatations sur les carbones 2 et/ou 4 de certains mannoses de sa chaîne principale (**figure 9** ; (Chiovitti et al., 2005)).

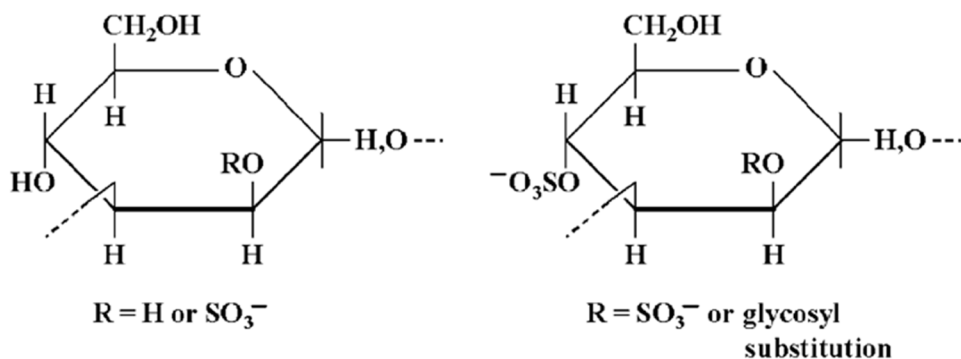


Figure 9 : Représentation des structures hypothétiques rencontrées chez *Pinnularia viridis* (Chiovitti et al., 2005).

1.2.3 Reproduction

La division cellulaire des diatomées se produit le plus souvent selon un processus de division mitotique asexuel. Cependant, à cause de la rigidité de leur paroi silicieuse s'opposant à l'expansion cellulaire, les deux cellules filles vont devoir se former à l'intérieur de la cellule mère (Hoek et al., 1995; Pickett-Heaps et al., 1984). Chez les diatomées, la division cellulaire est donc semi-conservative. Chaque cellule fille va hériter d'une des thèques parentales et former une seule nouvelle thèque. Ainsi, une des cellules filles va réutiliser l'épithèque de la cellule mère et synthétiser une nouvelle hypothèque correspondante, et l'autre va réutiliser l'hypothèque de la cellule mère en épithèque pour synthétiser une nouvelle hypothèque (**figure 10**). Ce processus de division implique donc que la taille d'une des deux cellules soit plus petite que celle de la cellule mère, et donc que la taille de la population décroisse. Dès lors, l'hétérogénéité et la variance de la population vont augmenter au fur et à mesure des divisions cellulaires. Ainsi, lorsque la taille atteint un seuil critique d'environ 30-40% de la taille maximale, la majorité des espèces vont procéder à une reproduction sexuée afin de restaurer la taille d'origine (Hoek et al., 1995; Mann, 1993).

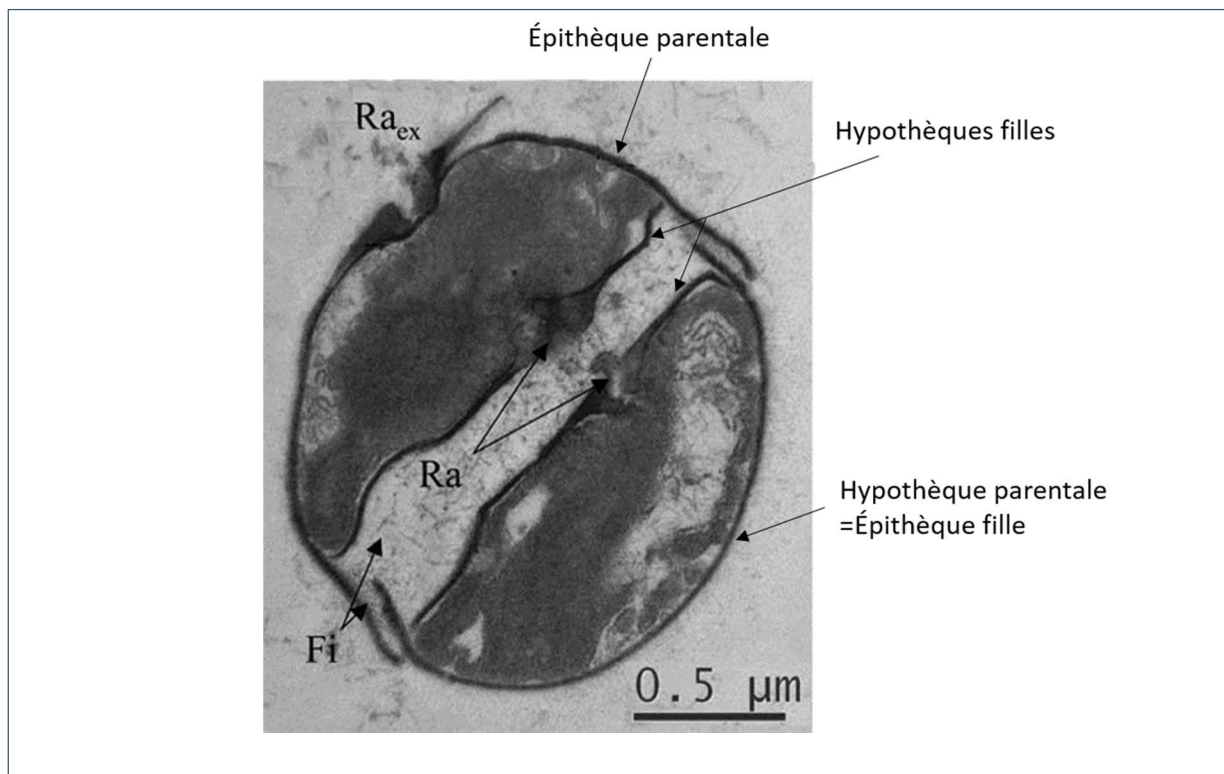


Figure 10 : Synthèse des hypothèques lors de la division cellulaire des diatomées.

Photographie prise en microscopie électronique à transmission du morphotype fusiforme de *Phaeodactylum tricornutum*, d'après Tesson et al., 2009. Ra = Raphé, Fi = matériel fibrillaire.

1.2.4 Sexualité

Le fait d'atteindre la taille critique n'est pas le seul élément déclencheur de la sexualité chez les diatomées. Différents signaux environnementaux peuvent en être aussi à l'origine comme la lumière, la salinité, des interactions biotiques avec des bactéries (Nagai and Imai, 1998), et dans quelques cas le stress nutritif (Chepurinov et al., 2004). L'intervention de phéromones a été récemment mise en évidence dans la sexualité des diatomées (Sato et al., 2011). Elles permettent ainsi de stimuler la sexualisation et jouent le rôle de chimioattractant, favorisant de fait la possibilité de rencontre entre deux gamètes (**figure 11**). La première structure de phéromone de diatomée a été mise en évidence à partir d'une culture de la diatomée pennée *Seminavis robusta*, et correspond à de la diproline (Gillard et al., 2013; Moeys et al., 2016). Peu de choses sont actuellement connues concernant son origine et son mécanisme d'action. Elle provient probablement d'une conversion à partir de glutamate puis serait relarguée dans le milieu par l'algue.

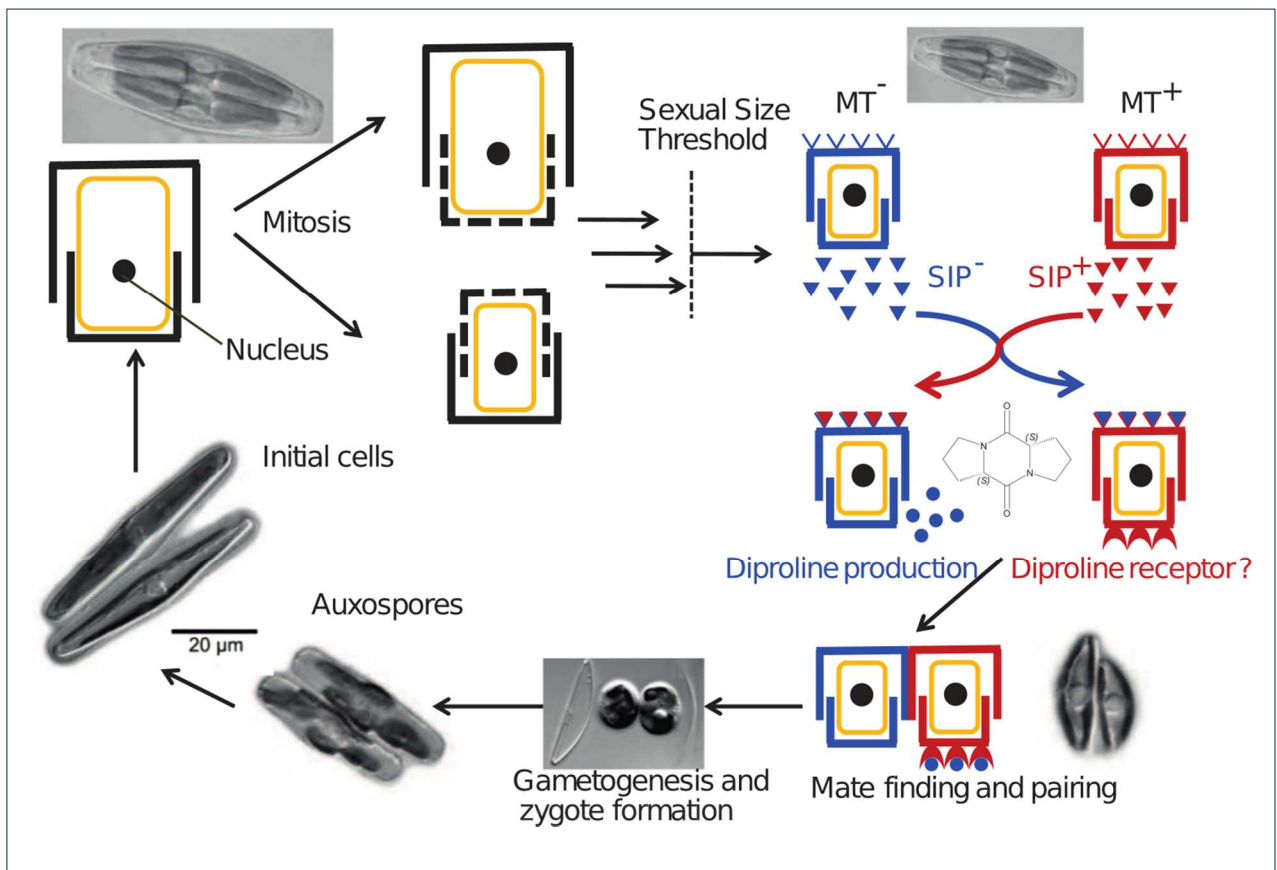


Figure 11 : Cycle de vie de la diatomée *Seminavis robusta* (Moeys et al., 2016).

Lorsque la taille des cellules passe en dessous d'un seuil critique, les cellules de type sexuel MT^- et MT^+ vont débiter la production de phéromones induisant la sexualité, respectivement SIP^- et SIP^+ . Sous l'influence de ces phéromones, les MT^- vont sécréter comme phéromone attractive de la diproline, et les MT^+ expriment probablement un récepteur à cette même phéromone. Ce mécanisme va permettre la rencontre des cellules de type sexuel opposé afin de générer un zygote.

La sexualité semble différer entre les diatomées centriques et les diatomées pennées. Ainsi, toutes les centriques sont oogames (elles produisent un ovule et un spermatozoïde), ou ont abandonné l'allogamie (zygote issu des gamètes de deux individus) pour des processus de reproduction asexuée comme l'automixie (zygote issu de la fusion de deux cellules haploïdes provenant d'un unique individu femelle) ou l'apomixie (clonage) (Mann, 1993). *A contrario*, les diatomées pennées sont généralement isogames (gamètes identiques) et non flagellées (Sato et al., 2011). Pour certains scientifiques, la présence de gamète mâle flagellé est même devenue un critère de distinction des diatomées centriques vis-à-vis des diatomées pennées (Hasle, 1983). De façon étonnante, les spermatozoïdes des diatomées centriques ne possèdent qu'un seul flagelle dépourvu de la paire de microtubules centraux habituellement rencontrée dans les axonèmes (Jensen et al., 2003; Manton, 1966). Les deux gamétanges produits vont ainsi s'accoupler, libérant chacun deux gamètes qui vont fusionner pour former deux zygotes qui vont croître à l'intérieur d'une auxospore (**Figure 11** ; Mann, 1993). Contrairement aux diatomées centriques, les diatomées pennées possèdent une grande variété d'auxosporulation catégorisées en 4 grandes classes basées sur le nombre d'auxospores produites par la cellule mère (Mann, 1993). Après germination, cette auxospore va former une cellule initiale qui va ensuite se diviser par mitose afin de produire une nouvelle génération de cellules végétatives.

1.2.5 Division cellulaire

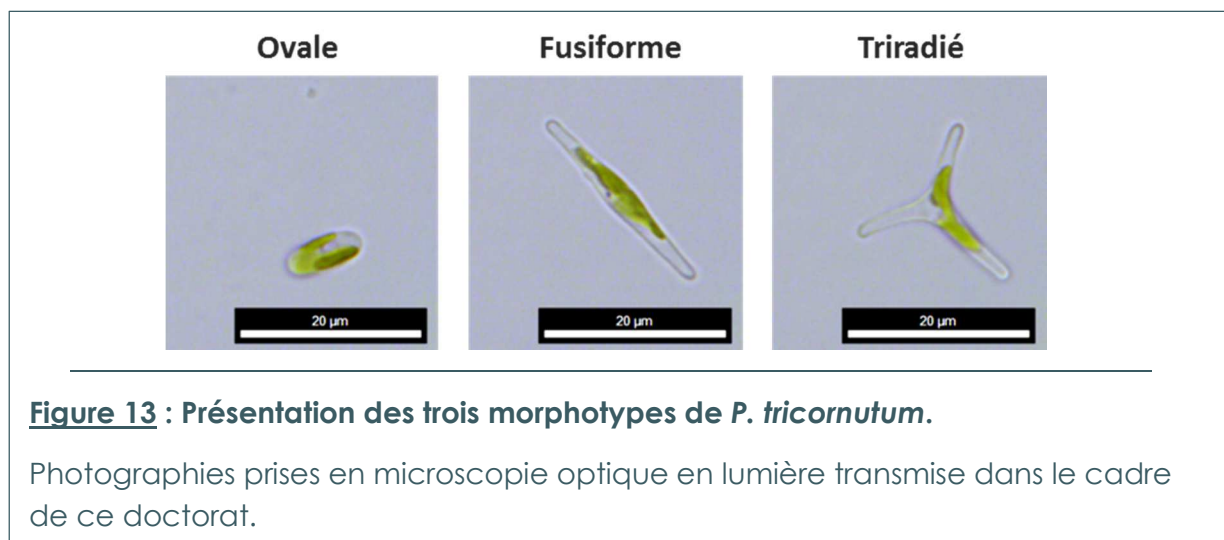
Ce processus a été observé au niveau ultrastructural chez de nombreuses espèces de diatomées, les étapes clés pouvant être résumées selon le schéma présenté sur la **Figure 12**. Les noyaux des cellules filles vont migrer chacun vers les extrémités de la cellule, puis un centrosome va venir se positionner entre le noyau et l'hypothèque nouvellement formée. Une SDV va venir se former entre le plasmalemme et le centrosome, puis elle va s'allonger de manière tubulaire. Elle va ensuite s'étendre perpendiculairement pour former une énorme vésicule du côté de l'hypothèque. La nouvelle valve va ainsi se former dans la SDV via le transport ciblé de silice, de protéines et de polysaccharides (Zurzolo and Bowler, 2001). Durant ce processus, la lumière de la SDV va s'acidifier pour permettre la polymérisation de la silice (Vrieling et al., 1999). Une partie du matériel organique va venir recouvrir la structure silicieuse, tandis qu'une autre partie va être impliquée dans le dépôt de la silice (Zurzolo and Bowler, 2001). Une fois la synthèse de la valve terminée, cette dernière sera exocytée par fusion de la membrane plasmique avec la membrane de la SDV, appelée aussi silicalemme. Ainsi, la face interne du silicalemme va devenir la nouvelle membrane plasmique. Les cellules posséderont ainsi sur cette valve naissante un nouveau raphé qui sera

Suite à ces événements, les deux cellules filles vont ensuite se séparer, puis elles vont croître de manière unidirectionnelle selon l'axe de division cellulaire. Comme le frustule est très rigide, la croissance cellulaire durant l'interphase n'est possible qu'en augmentant la distance entre l'épithèque et l'hypothèque. Ainsi, au fur et à mesure de la croissance cellulaire, de nouvelles bandes de silice vont être synthétisées au niveau du cingulum grâce au processus de SDVs décrit précédemment. Ces vésicules vont fusionner avec la membrane l'une après l'autre afin d'éviter la formation d'ouverture au niveau latéral de la paroi cellulaire, seule partie extensible du frustule (Sumper and Kröger, 2004). Une fois que la cellule a atteint la taille requise, un nouveau cycle de division cellulaire peut recommencer.

1.3 La diatomée *Phaeodactylum tricornutum*

1.3.1 Histoire

Parmi toutes les diatomées recensées jusqu'alors, *P. tricornutum* est une diatomée atypique du fait de son pléomorphisme unique. Cette capacité est liée en partie à sa paroi faiblement silicifiée par rapport aux autres diatomées (Lewin et al., 1958) lui permettant très certainement une plus grande plasticité morphologique (Borowitzka and Volcani, 1978). De ce fait, elle a la particularité de présenter naturellement trois morphotypes différents : ovale, fusiforme et triradié (**Figure 13**).



Ces trois morphotypes ont la même ultrastructure cellulaire à l'exception de leur paroi et de leur organisation vacuolaire (Borowitzka and Volcani, 1978). Le morphotype triradié fut le premier à avoir été isolé par Bohlin en 1897 à partir d'échantillons provenant de Plymouth (Grande-Bretagne) et des vasques naturelles de la mer Baltique, probablement grâce à sa forme originale (De Martino et al., 2007). Allen and Nelson (1910) isolèrent ensuite le morphotype fusiforme, qui de par sa morphologie et sa ressemblance avec le genre *Nitzschia*, fut alors qualifié comme étant

Nitzschia closterium f. minutissima. Wilson a le premier, en 1946, décrit la nature polymorphique de *P. tricornutum* (sous le nom de *Nitzschia closterium f. minutissima*), et révéla l'existence de trois morphotypes appartenant à la même espèce. Par la même occasion, il attira l'attention sur la grande ressemblance entre ses observations du morphotype triradié et les études précédentes de *P. tricornutum*. Ce n'est qu'un peu plus tard en 1954 qu'Hendey fit le rapprochement entre les deux et démontra que les études précédentes étaient basées en fait sur une espèce unique : *P. tricornutum* (Hendey et al., 1954). Cependant, la présence de leucosine, l'absence de motilité observée alors, et la structure pariétale silicifiée de *P. tricornutum*, ont mené à la conclusion que cette algue ne pouvait pas appartenir à la classe des diatomées (Hendey et al., 1954), mais plutôt à celle des Chrysophyceae. Ce n'est qu'après observation d'un raphé silicifié dans la forme ovale que Lewin (Lewin, 1958) a assigné *Phaeodactylum* aux Bacillariophyceae. *P. tricornutum* est actuellement la seule représentante du sous-ordre Phaeodactylinae, Famille Phaeodactylacea, genus *Phaeodactylum* (Lewin, 1958; Lewin et al., 1958). Cette algue ne semble pas très abondante dans la nature, bien que cosmopolite. C'est une espèce côtière essentiellement retrouvée au niveau des estuaires et des vasques naturelles (De Martino et al., 2007). En arrivant à tirer parti de ces environnements très fluctuants, aussi bien en terme de température que de salinité, *P. tricornutum* fait preuve d'une grande adaptabilité probablement liée à son polymorphisme (De Martino et al., 2007). Depuis le siècle dernier, *P. tricornutum* a été isolée à partir de différentes zones géographiques à travers le monde, et 10 souches axéniques (Pt1-Pt10) ont été caractérisées par De Martino et al. (2007), qui par séquençage de la région ITS2 ont déterminé quatre génotypes : A, B, C et D, définis selon le polymorphisme génétique de cette région chromosomique. Historiquement, les souches Pt1.8.6 et Pt3 ont été isolées à partir d'échantillons provenant respectivement de Blackpool et Plymouth (Grande-Bretagne) (De Martino et al., 2007). La souche Pt3 dérive de la souche Pt2 isolée aux environs de 1910 par E. J. Allen. Ainsi, Pt3 a été obtenue dans les années 1930 par un isolement clonal de Pt2 capable de pousser dans de l'eau douce. La souche Pt1.8.6 est quant à elle une des neuf cultures monoclonales réalisées à partir de la souche Pt1 isolée en 1956 par S. Coughlan.

1.3.2 Particularité de sa plasticité morphologique

La plasticité morphologique de *P. tricornutum* lui permet de pouvoir survivre en l'absence de silice dans le milieu (De Martino et al., 2007). Peu de publications documentent les étapes de la transformation/passage d'un morphotype à un autre (De Martino et al., 2007, 2011; Gutenbrunner et al., 1994; Lewin et al., 1958; Tesson et al., 2009; Wilson, 1946). Ainsi, *P. tricornutum* présente 3 morphotypes majoritaires qui sont les morphotypes ovale, fusiforme et triradié. Plus rarement,

trois autres morphotypes peuvent être observés : les morphotypes cruciforme (He et al., 2014; Wilson, 1946), boomerang et rond (De Martino et al., 2011). Les morphotypes cruciforme et boomerang semblent être plutôt des états transitoires entre deux des morphotypes principaux, tandis que le morphotype rond possède les caractéristiques d'une cellule en dormance (paroi épaisse, corps lipidiques, chloroplaste désorganisé). Cette variabilité morphologique ne semble pas dépendre du génotype, bien qu'il semble que certaines souches présentent des prédispositions pour un morphotype particulier (**figure 14** ; (De Martino et al., 2007). Ainsi, parmi les 10 souches de *P. tricornutum*, les souches Pt3 et Pt9 semblent posséder des prédispositions pour le morphotype ovale, la souche Pt8 plutôt pour le morphotype triradié, et les autres pour le morphotype fusiforme. Cependant, le morphotype fusiforme a déjà été observé dans les 10 souches de *P. tricornutum* appartenant aux 4 génotypes différents (De Martino et al., 2007). Malgré le flou qui entoure encore la sexualité de *Phaeodactylum*, la morphogenèse ne semble pas provenir d'une alternance entre un stade haploïde et diploïde de la microalgue (Darley, 1968). Les trois morphotypes majoritaires possèdent la même ultrastructure à l'exception de leur paroi cellulaire ainsi que de leur organisation vacuolaire (Borowitzka and Volcani, 1978), ce qui leur permet d'avoir une flottabilité différente (De Martino et al., 2011; Lewin et al., 1958; Wilson, 1946).














Accession name	Genotype	Morphology
Fusiform morphotype		
Pt1	A	
Pt2	A	
Pt3	A	
Pt4	B	
Pt5	C	
Pt6	D	
Pt7	D	
Pt8	D	
Pt9	A	
Pt10	C	
Oval morphotype		
Pt3	A	
Pt9	A	
Triradiate morphotype		
Pt8	D	

Figure 14 : Prédiposition morphologique des 10 souches de *P. tricornutum*.
D'après De Martino et al., 2007.

1.3.3 Structure

1.3.3.1 Générale

Phaeodactylum possède des organites dont l'ultrastructure est similaire à celles des autres diatomées. Ainsi, *Phaeodactylum* possède un unique chloroplaste central situé du côté de l'épivalve (**figure 15**). Ce plaste présente un seul pyrénoïde traversé par deux thylakoïdes. Le noyau se retrouve habituellement au centre à côté du chloroplaste. Son enveloppe nucléaire est reliée au réticulum endoplasmique qui entoure le chloroplaste. A l'interphase, selon les coupes observées, la microalgue peut présenter de un à quatre appareil(s) de Golgi toujours proche(s) de la membrane nucléaire, ainsi qu'une à plusieurs mitochondries. Les extrémités distales de la cellule sont occupées par de larges vacuoles qui comblent la différence de volume entre les différents morphotypes (Borowitzka and Volcani, 1978).

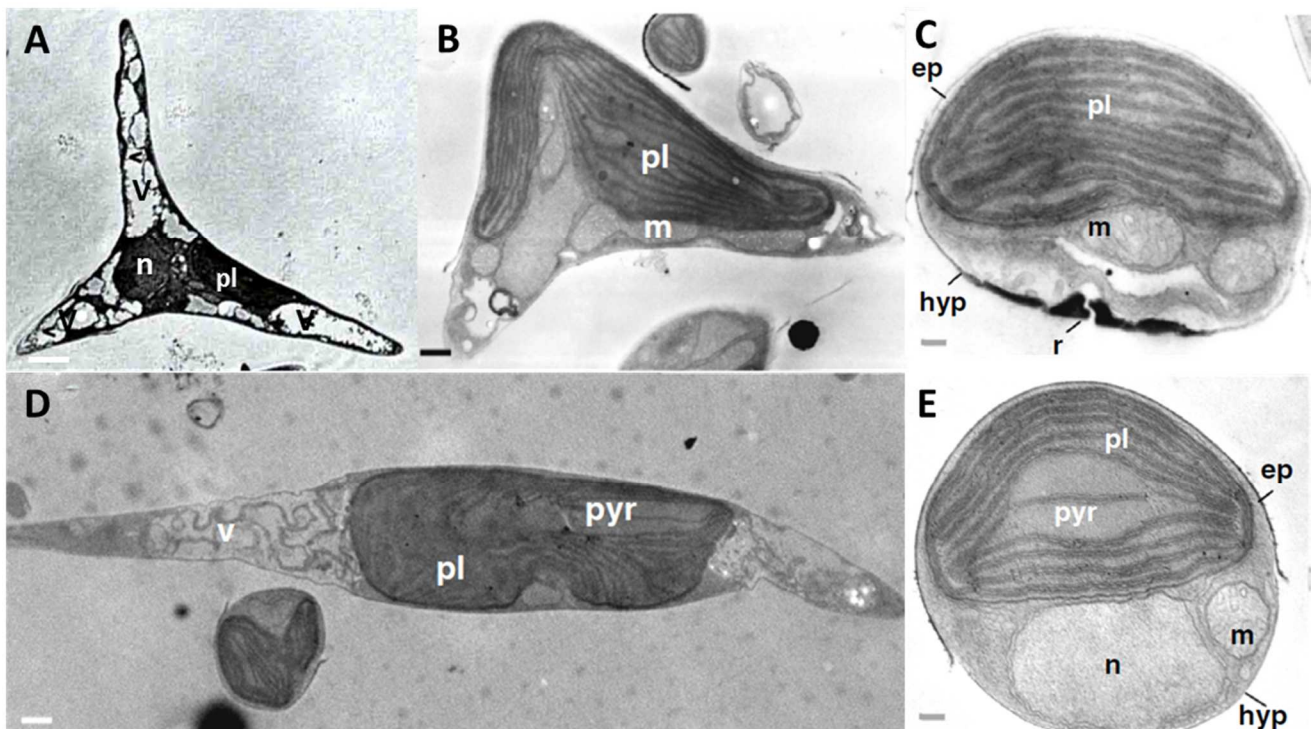


Figure 15 : Ultrastructure des trois morphotypes de *P. tricornutum*.

Photographies prises en microscopie électronique à transmission d'après De Martino et al., 2011 ; Borowitzka et Volcani, 1978. A et B = morphotype Triradié ; C = morphotype Ovale ; D et E = morphotype Fusiforme respectivement en section longitudinale et transversale. Pl = plastide ; pyr = pyrenoid ; n = noyau ; m = mitochondrie, ep = épivalve ; hyp = hypovalve. Echelle noire = 1,5 μ m (Photo B) ; échelle blanche = 1 μ m (Photos A et D) ; échelle grise = 0,3 μ m (Photos C et E).

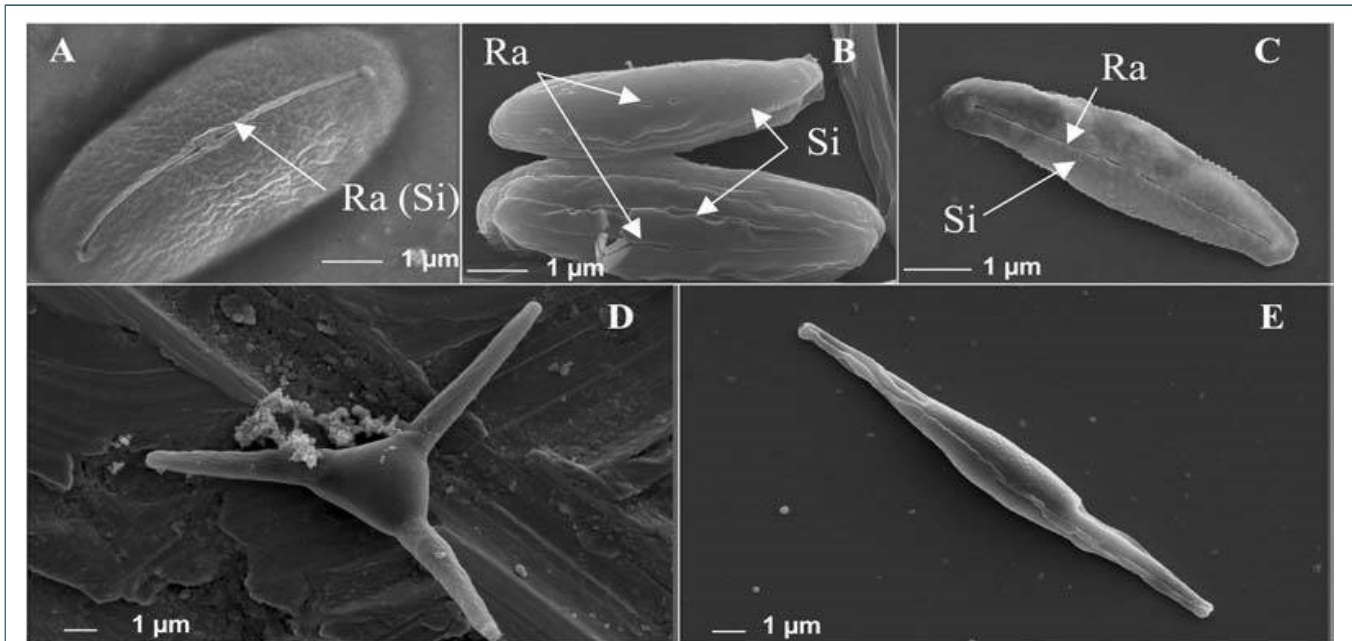


Figure 16 : Structure des trois morphotypes de *P. tricornutum*.

Photographies prises en microscopie électronique à balayage, Tesson et al., 2009.

A-C = morphotype Oval avec différents degrés de silicification : (A) seulement le raphé est silicifié ; (C) la valve entière est silicifiée ; (B) silicification intermédiaire entre (A) et (C). D = morphotype triradié. E = morphotype fusiforme. Ra = Raphé ; Si = structure silicifiée.

1.3.3.2 Fusiforme

Le morphotype fusiforme possède une forme de fuseau légèrement incurvé mesurant généralement de 25 à 35µm (**Figure 16**). Parfois possédant une taille plus petite, il est cependant rare d'en observer au-delà de 35µm. Le tiers du milieu de la cellule est bombé, puis le diamètre cellulaire diminue en allant vers les extrémités distales, formant ainsi deux prolongations appelées aussi « bras » de la cellule. On pourra noter la forme arrondie des extrémités de l'algue. La partie bombée de la cellule va contenir le noyau, le chloroplaste, ainsi que la majorité des organites, alors que les bras sont comblés par des vacuoles (Wilson, 1946). Le morphotype fusiforme a été observé comme étant la forme la plus stable (Volcani, 1981). Ce morphotype est effectivement la forme majoritairement isolée dans la nature, ainsi qu'observée dans les cultures en laboratoire, représentant plus de 95% de la population pour 7 des 10 souches (De Martino et al., 2007).

1.3.3.3 Ovale

Comme le suggère son nom, le morphotype ovale présente une forme ovoïde qui ne possède pas de bras (**figure 16**). De ce fait, contrairement aux autres morphotypes qui possèdent des extrémités comblées par des vacuoles, le volume de ces dernières est très restreint donnant l'impression que la majorité du volume intracellulaire est occupé par le chloroplaste (Wilson, 1946). *P. tricornutum* est classée comme une diatomée pennée raphidée, mais en réalité, seule sa forme ovale possède un raphé (Lewin et al., 1958). Ce morphotype est celui qui se rapproche le plus des autres diatomées. Il est le seul à posséder une véritable coque de silice située généralement sur l'hypothèque où se trouve le raphé de *P. tricornutum* (De Martino et al., 2007; Tesson et al., 2009). D'ailleurs, une des particularités de la forme ovale de *P. tricornutum* est de ne posséder qu'une seule des deux valves silicifiée (Lewin et al., 1958). Bien que de manière surprenante les morphotypes ovale et fusiforme contiennent la même quantité de silice (0,4-0,5 % de la masse sèche ; Lewin et al., 1958), elle est, chez l'ovale, utilisée pour la création de structures comme le raphé. Ceci peut expliquer la sensibilité de ce morphotype à l'absence de silice dans le milieu contrairement aux deux autres morphotypes. En effet, il a été observé chez le morphotype ovale, une diminution du taux de croissance de 50% lorsqu'il est cultivé dans un milieu exempt de silice (De Martino et al., 2007). Le degré de silicification chez ce morphotype est variable, il peut aller d'un simple raphé à une valve entière. Comme chez les autres diatomées, la valve est composée d'une partie interne organique située à l'extérieur de la membrane cellulaire, attachée à une portion externe silicifiée contenant le raphé placé en son centre. Cette couche organique correspond vraisemblablement à la paroi cellulaire observée chez les morphotypes fusiforme et triradié. La présence de silice au niveau de sa valve permet ainsi à la forme ovale d'avoir une plus grande résistance mécanique que les deux autres morphotypes (Francius et al., 2008). De plus, la présence du raphé chez ce morphotype lui permet de sécréter plus facilement une substance exopolymérique (EPS) qui lui permet de se mouvoir en glissant sur les surfaces (Dugdale et al., 2006; Higgins et al., 2003). Similairement aux autres diatomées, cette substance confère au morphotype ovale des propriétés d'adhésion, favorisant ainsi la formation de colonies par agrégation cellulaire (Stanley and Callow, 2007). Les cellules entourées de ce mucilage peuvent alors adhérer aux surfaces comme par exemple la paroi des flasques de culture (De Martino et al., 2007; Wilson, 1946). Le morphotype ovale possède une plus grande vitesse de sédimentation que les deux autres morphotypes, phénomène amplifié par l'agrégation cellulaire (De Martino et al., 2011; Lewin et al., 1958). Ce morphotype est donc préférentiellement benthique, alors que les morphotypes fusiforme et triradié sont retrouvés plus fréquemment avec le plancton (De Martino et al., 2007; Round et al., 1990).

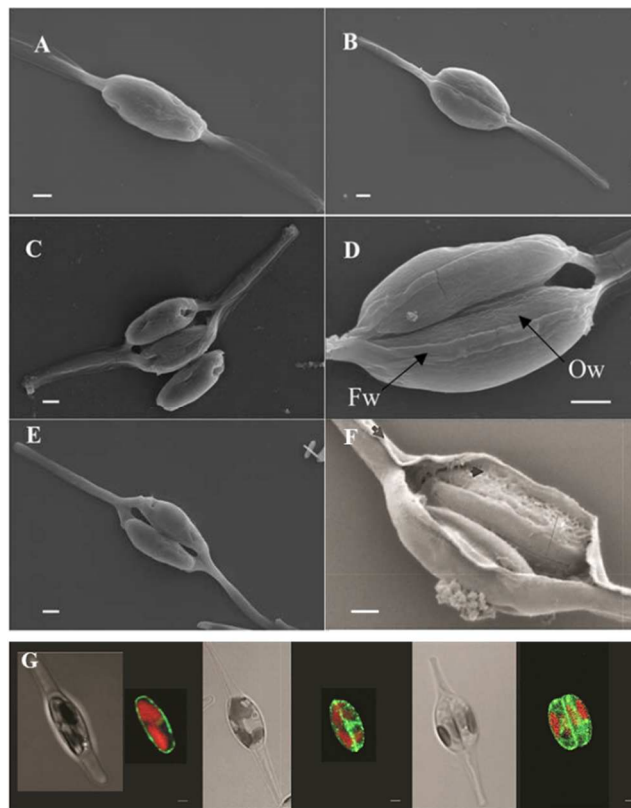


Figure 17 : Transition du morphotype fusiforme vers le morphotype ovale.

Photographies prises en microscopie électronique à balayage (A-F) et microscopie confocale (G) de différentes étapes de la transition. Les photographies de l'image (G) ont été réalisées sur une souche transgénique exprimant un marqueur de la voie de sécrétion : Sec4-YFP. Echelle = 1 μ m. D'après De Martino et al., 2011 ; Tesson et al., 2009.

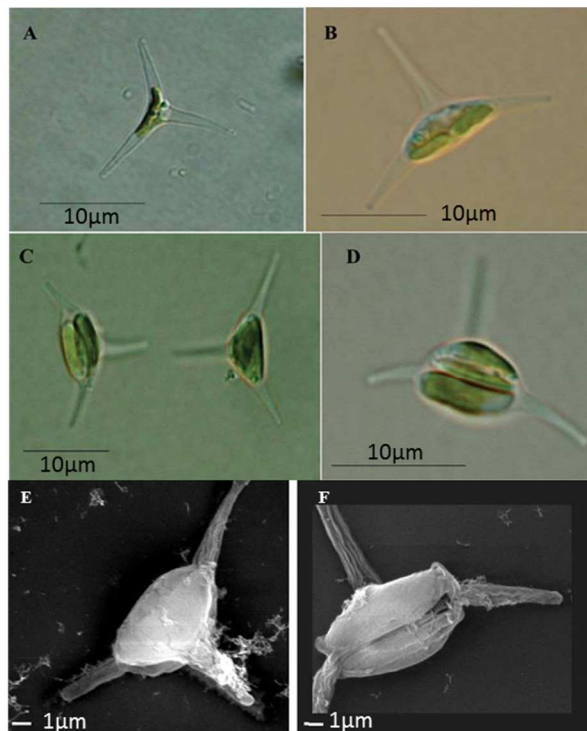


Figure 18 : Transition du morphotype triradié vers le morphotype ovale.

Photographies prises en microscopie optique à lumière transmise (A-D) et microscopie électronique à balayage (E et F) de différentes étapes de la transition.

D'après De Martino et al., 2011 ; Tesson et al., 2009.

Les échantillonnages effectués lors d'analyses environnementales étant majoritairement prélevés dans la colonne d'eau, il est probable que la quantité et la répartition de *Phaeodactylum* recensées dans les océans soient sous-estimées (De Martino et al., 2007).

1.3.3.4 *Triradié*

Bien que plus faiblement représenté, le morphotype triradié a pu être observé aussi bien dans des échantillons prélevés sur le terrain (Bohlin, 1897; De Martino et al., 2007; Johansen, 1991; Wilson, 1946), que dans des cultures en laboratoire (Borowitzka and Volcani, 1978; De Martino et al., 2007; Wilson, 1946). Ce morphotype a une organisation très proche du morphotype fusiforme, exception faite qu'il possède trois bras au lieu de deux (**figure 16**). De plus, la longueur des bras est plus petite que chez le fusiforme, mesurant entre 6 à 8µm chez le triradié (parfois de longueur inégale), contre 10µm chez le fusiforme. Les trois angles créés par les bras sont rarement égaux et sont très variables en fonction des individus (Wilson, 1946). Le morphotype triradié, tout comme le fusiforme, ne présente pas de coque de silice. Leur paroi est essentiellement organique à l'exception de quelques bandes de silice qui sont enchâssées dans le matériel organique. Le nombre de ces bandes peut aller de 0 à 10. Situées habituellement au niveau de la ceinture de l'algue, il est possible d'en retrouver au niveau de l'hypothèque des cellules les plus vieilles (Borowitzka and Volcani, 1978).

1.3.4 Morphogenèse

1.3.4.1 *Transition*

Par le passé, différents travaux ont tenté de comprendre le polymorphisme de cette microalgue et les conditions menant à l'apparition de chaque morphotype. Bien que les mécanismes moléculaires régissant la morphogenèse restent encore méconnus, des observations de ces interconversions ont été relatées dans différentes études. Ce processus semble difficile à suivre, et seules certaines étapes de cette morphogenèse ont pu être précisément observées.

1.3.4.1.1 *Transition des morphotypes fusiforme et triradié vers le morphotype ovale*

Lors de la transition vers le morphotype ovale, la cellule mère, qu'elle soit fusiforme ou triradiée, va enfler en son centre afin de pouvoir produire dans cette région la nouvelle cellule

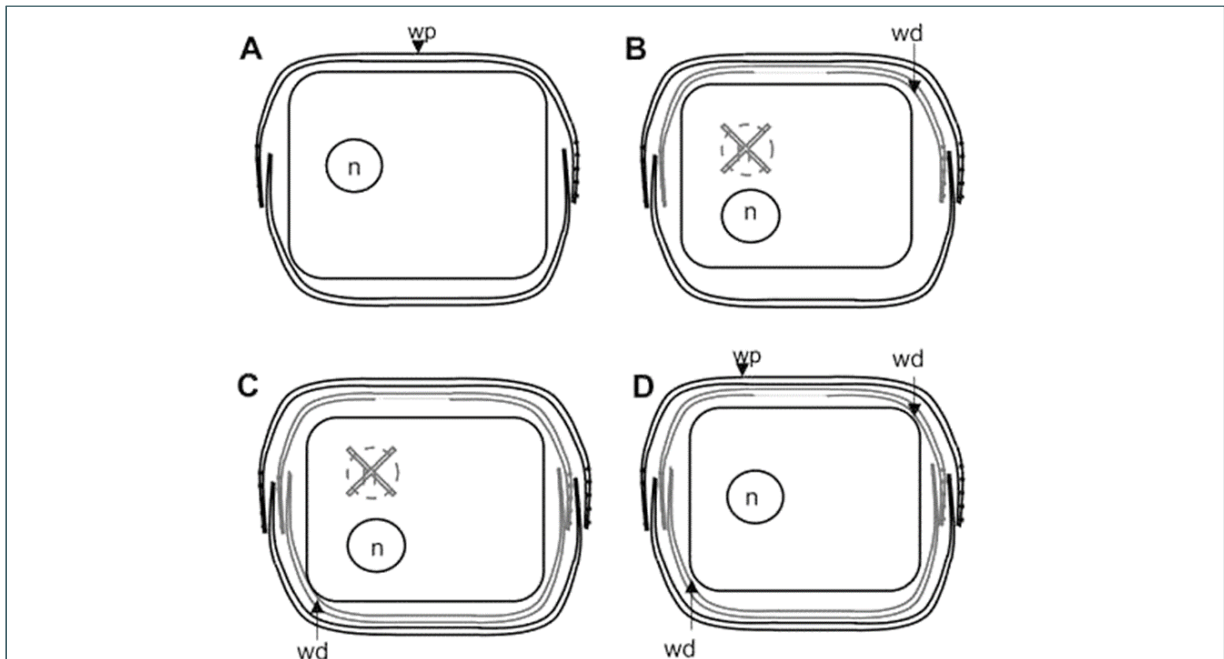


Figure 19 : Schéma des mitoses acytocinétiques impliquées dans la formation du morphotype ovale (Tesson et al., 2009).

A = cellule parentale ; B = première mitose acytocinétique, un noyau dégénère et une nouvelle valve est synthétisée ; C = seconde mitose acytocinétique, un noyau dégénère et une seconde valve est synthétisée ; D = nouvelle cellule ovale dans la cellule parentale. n = noyau ; wd = paroi de la cellule ovale ; wp = paroi cellulaire parentale.

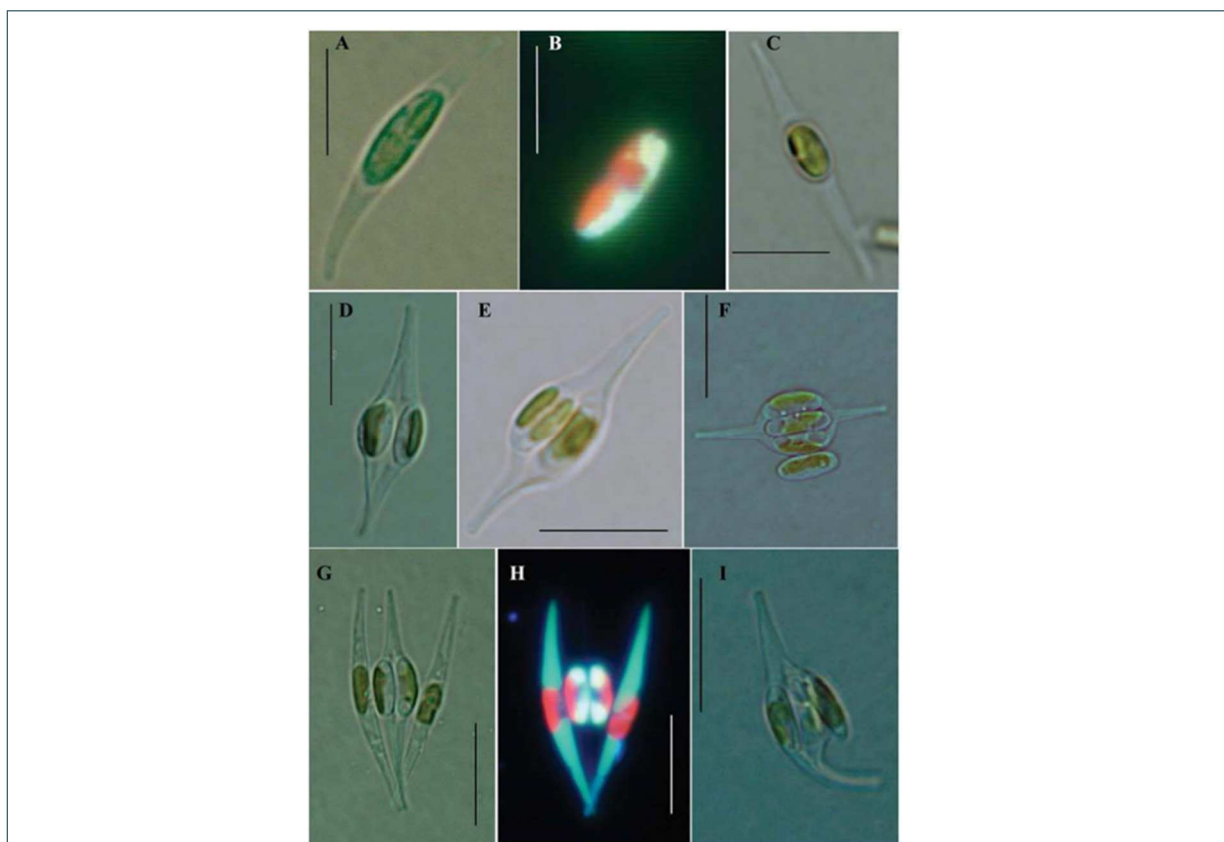


Figure 20 : Transformation du morphotype fusiforme en morphotype ovale (Tesson et al., 2009).

Photographies prises en microscopie optique en lumière transmise (A, C, D-G et I) et en épifluorescence (B et H). Les images de fluorescence ont été réalisées sur des cellules marquées au PDMPO. Echelle = 10 μ m.

ovale (**figures 17 et 18**). Cette dernière va ensuite se diviser au sein même de la cellule mère aboutissant à une paire de cellules ovales contenues dans le frustule parental (De Martino et al., 2007; Gutenbrunner et al., 1994; Lewin et al., 1958). Il est actuellement supposé que la formation de la première cellule ovale viendrait d'une succession de deux mitoses acytocinétiques (**figure 19** ; (Gutenbrunner et al., 1994). Les cellules filles vont ensuite être relâchées dans le milieu par rupture de la paroi parentale au niveau de la jonction des deux valves (**figure 17-F, figure 20-I**). Cependant, avant la libération, plusieurs divisions successives des cellules ovales peuvent se produire au sein de la cellule mère (**figure 20-D à F** ; De Martino et al., 2011). Afin de pouvoir aboutir à la formation de cellules ovales à partir des morphotypes fusiforme et triradié, le protoplasme de la cellule parentale va se rétracter à partir des apex pour s'invaginer jusqu'à la portion médiane de la cellule. L'implication de la vacuole a notamment été démontrée dans ce processus de différenciation (**figure 18-B, figure 20-A, B, C, G, H** ; (De Martino et al., 2007). De manière coordonnée, l'espace libéré entre la paroi et le plasmalemme va être comblé de matériel fibrillaire, qui va finir par enrober la nouvelle cellule ovale. La synthèse de cette substance fibrillaire pourrait d'un côté participer à la rétraction du protoplaste, tout en stabilisant la nouvelle cellule au centre de la cellule parentale (De Martino et al., 2011). Cette même substance peut être aussi observée lors de la division mitotique « normale » de *Phaeodactylum* (Borowitzka and Volcani, 1978), cependant l'origine de ce matériel fibrillaire reste encore méconnue. Au moins deux hypothèses ont été formulées. Selon la première, ce matériel fibrillaire pourrait provenir d'une excrétion de la nouvelle cellule ovale, à l'instar de ce qui a déjà été observé pour les EPS. A l'inverse, la deuxième explication pourrait être une désorganisation du diatopetum de la paroi mère, qui est composé de matériel fibrillaire (Tesson et al., 2009). Cette substance fibrillaire n'adhère cependant pas aux cellules ovales, puisque ces dernières seront relâchées dans le milieu rapidement après la rupture du frustule de la cellule mère.

1.3.4.1.2 Transition du morphotype ovale vers les morphotypes fusiforme et triradié

Le processus inverse de transformation du morphotype ovale vers les deux autres morphotypes ne semble pas utiliser les mêmes mécanismes. Ainsi, la morphogénèse ne sera pas générée suite à une mitose, mais plus simplement par élongation progressive de bras à partir de la cellule ovale (**figure 21** ; Wilson, 1946).



Figure 21 : Conversion du morphotype ovale en morphotype triradié et fusiforme d'après Tesson et al., 2009 ; De Martino et al., 2011.

Photographies prises en microscopie optique en lumière transmise. A = conversion vers le morphotype triradié ; B = conversion vers le morphotype fusiforme. Echelle grise = 10µm ; échelle blanche 2µm.

1.3.4.1.3 Transition du morphotype triradié vers le morphotype fusiforme

La transformation du morphotype triradié vers le morphotype fusiforme, ou inversement, va passer par différentes formes intermédiaires asymétriques (**figure 22**). La forme triradiée va ainsi commencer par rétracter progressivement un de ses bras jusqu'à le résorber complètement menant ainsi à la formation d'un morphotype intermédiaire en forme de boomerang (Borowitzka and Volcani, 1978; De Martino et al., 2007; Wilson, 1946). La cellule va ensuite se redresser pour finalement aboutir au morphotype fusiforme. Les relations entre les morphotypes peuvent donc se résumer selon le schéma suivant (**figure 23**).

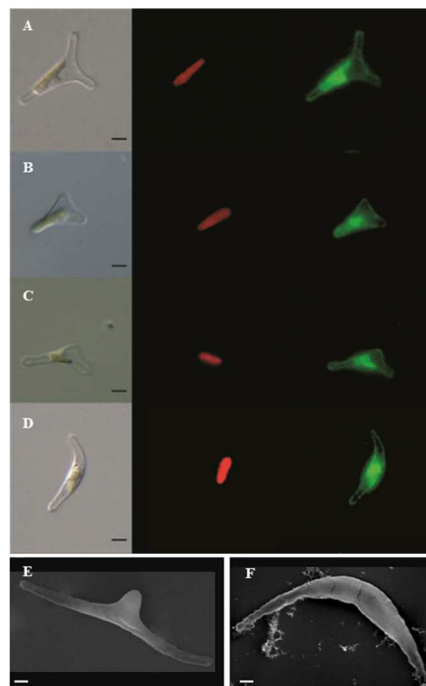


Figure 22 : Etapes de la conversion du morphotype triradié en morphotype fusiforme d'après De Martino et al., 2007; De Martino et al., 2011.

A-D = Photographies de cellules exprimant une eYFP cytosolique : à gauche = bright-field ; au milieu = autofluorescence chloroplastique ; à droite = eYFP. Echelle = 3µm. E-F = Photographies en microscopie électronique à balayage. Echelle = 1µm.

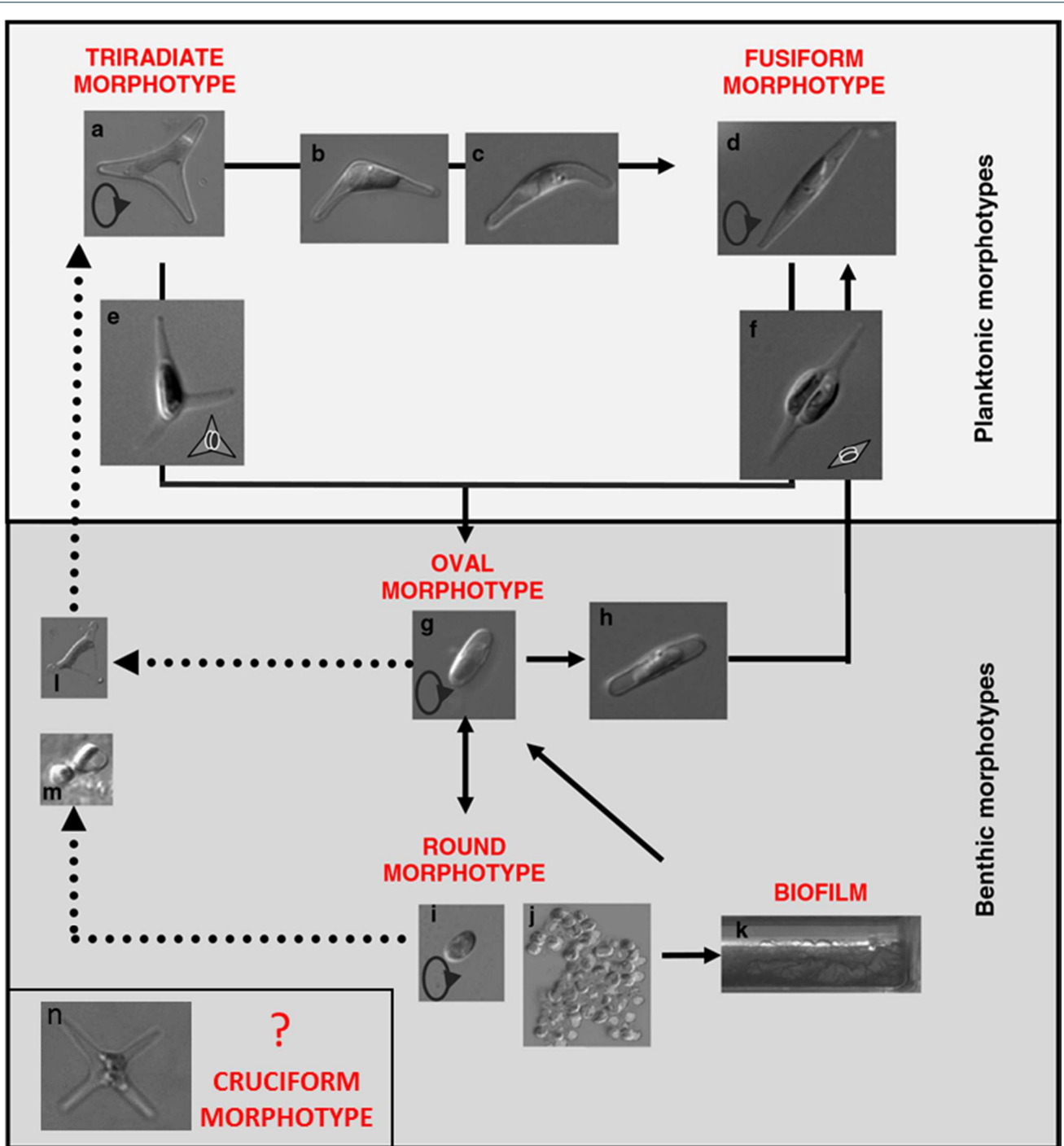


Figure 23 : Résumé des relations intermorphotypes de *P. tricornutum* d'après De Martino et al., 2011.

P. tricornutum présente trois morphotypes principaux : ovale (g), fusiforme (d) et triradié (a). Le morphotype triradié peut se transformer en morphotype fusiforme en adoptant des formes intermédiaires (boomerang ; b, c). En conditions de stress, les morphotypes fusiforme et triradié se convertissent en ovale par division cellulaire acytocinétique (e, f). Le processus inverse s'effectue par un allongement de la cellule (h, i). Lors d'un stress prolongé, le morphotype ovale peut adopter un morphotype rond (i) et former des agrégats (j) qui peuvent générer des biofilms (k). Un morphotype plus rare peut être observé, le morphotype cruciforme (n), qui ne possède pas de position clairement établie dans ce schéma.

1.3.4.2 Physiologie

1.3.4.2.1 Fusiforme

Etant la forme prédominante retrouvée aussi bien dans la nature qu'en culture, le morphotype fusiforme a été défini comme étant le morphotype "normal" de *Phaeodactylum* (Volcani, 1981; Wilson, 1946). C'est donc sur ce principe que le morphotype fusiforme cultivé dans des conditions standard a servi de référence pour différentes études. En utilisant des approches biochimiques, certains changements structurels ont pu être mis en évidence par rapport au morphotype ovale. Ainsi, il a été noté qu'après coloration au bleu de toluidine, qui colore les structures chargées négativement comme les acides nucléiques, les protéoglycannes, ou encore les polysaccharides sulfatés ou phosphorylés, le morphotype fusiforme présente une plus faible coloration que le morphotype ovale. Ce résultat corrobore les précédentes observations montrant, chez le morphotype ovale, une plus grande excrétion d'EPS qui possèderaient donc des composés négativement chargés. De plus, il a été montré par électrophorèse SDS-PAGE que ces deux morphotypes n'ont pas le même profil protéique, et que lorsqu'une lignée ovale se trouvait convertie en fusiforme, cette dernière adoptait le profil protéique caractéristique des fusiformes. Cette étude montre ainsi que le protéome de la microalgue subit de grandes modifications lors de cette transition d'un morphotype à l'autre (Gutenbrunner et al., 1994).

1.3.4.2.2 Ovale

La présence du morphotype ovale dans les cultures semble être corrélée avec l'apparition de conditions défavorables pour *Phaeodactylum* (De Martino et al., 2007; Johansen, 1991). En effet, la transformation des cellules ovales est souvent observable dans des cultures non agitées, sur milieu solide, ou encore en phase stationnaire de culture (Cooksey and Cooksey, 1974; Gutenbrunner et al., 1994). La température à laquelle sont cultivées les cellules semble aussi être un facteur influençant la morphogénèse. Ainsi, lorsque les cultures sont placées en deçà des températures de culture recommandées (15-25°C ; Kudo et al., 2000), il a été observé une augmentation de la population ovale à la défaveur des autres morphotypes. Ce même phénomène a aussi été observé lorsque *Phaeodactylum* est cultivée dans un milieu hyposalin, favorisant une nouvelle fois la formation du morphotype ovale (De Martino et al., 2011). Ces observations laissent donc penser que la forme ovale de *Phaeodactylum* serait une forme résistante de l'algue, plus tolérante au stress, et donc plus à même de survivre à des conditions de culture défavorables. Ceci est corroboré par le fait que le morphotype ovale est capable de survivre à un faible éclairage, ainsi qu'à un milieu pauvre en nutriments, contrairement aux autres morphotypes (De Martino et al., 2007). Ces

particularités, associées à une vitesse de sédimentation plus élevée, rendent ce morphotype particulièrement adapté à la vie benthique.

Des études transcriptomiques basées sur des EST ont été réalisées sur les trois morphotypes à partir de trois souches différentes enrichies en chaque morphotype : Pt3 = 80% ovale ; Pt8 = 65% triradié ; Pt1.8.6 > 95% fusiforme (Maheswari et al., 2009, 2010). Le morphotype ovale chez Pt3 ayant été obtenu en cultivant les cellules en milieu hyposalin. Chacune des bibliothèques a été comparée aux deux autres ainsi qu'à 13 autres bibliothèques générées à partir de la souche Pt1.8.6 cultivée dans différentes conditions (De Martino et al., 2011). Ainsi, parmi les gènes différentiellement exprimés chez l'ovale, il a été logiquement mis en évidence la présence de gènes impliqués dans le stress salin chez les plantes. Ces derniers agissent notamment sur le remodelage membranaire, la régulation osmotique, la signalisation, les canaux et transporteurs ioniques, ainsi que la photoprotection. On retrouve aussi certains gènes qui coderaient pour des enzymes potentiellement impliquées dans la fixation du carbone en C₄, ainsi que dans le métabolisme de l'ammonium. Les termes GO les plus sur-représentés sont ceux appartenant à la catégorie du métabolisme des lipides, tandis que le domaine InterPro le plus observé correspond à la famille Na⁺/H⁺ antiporter NhaC impliquée dans l'homéostasie du pH, ainsi que l'extrusion du sodium (De Martino et al., 2011).

1.3.4.2.3 Triradié

Par opposition au morphotype ovale, l'apparition du morphotype triradié semble être favorisée par des conditions hypersalines et alcalines, aussi bien en culture (Bartual et al., 2008) que dans la nature (Johansen, 1991). Cependant, les conditions d'obtention de ce morphotype restent peu déterminées. Il est d'ailleurs admis que ce morphotype est le plus difficile à maintenir en culture (Lewin et al., 1958). Néanmoins, il a été observé que le morphotype triradié est le plus sensible à l'hyposalinité, et que la plupart des cellules meurent dans ces conditions. La salinité semble donc pouvoir être un facteur de la morphogénèse, qui en fonction de sa concentration, favorise soit la formation du morphotype ovale, soit la formation du morphotype triradié (De Martino et al., 2011). Cette dernière étude s'oppose donc aux observations précédentes qui n'ont pas su clairement mettre en évidence de lien entre morphogénèse et salinité (Abdullahi et al., 2006; Borowitzka and Volcani, 1978; Gutenbrunner et al., 1994). Ces contradictions laissent penser que d'autres facteurs puissent également entrer en jeu dans la morphogénèse.

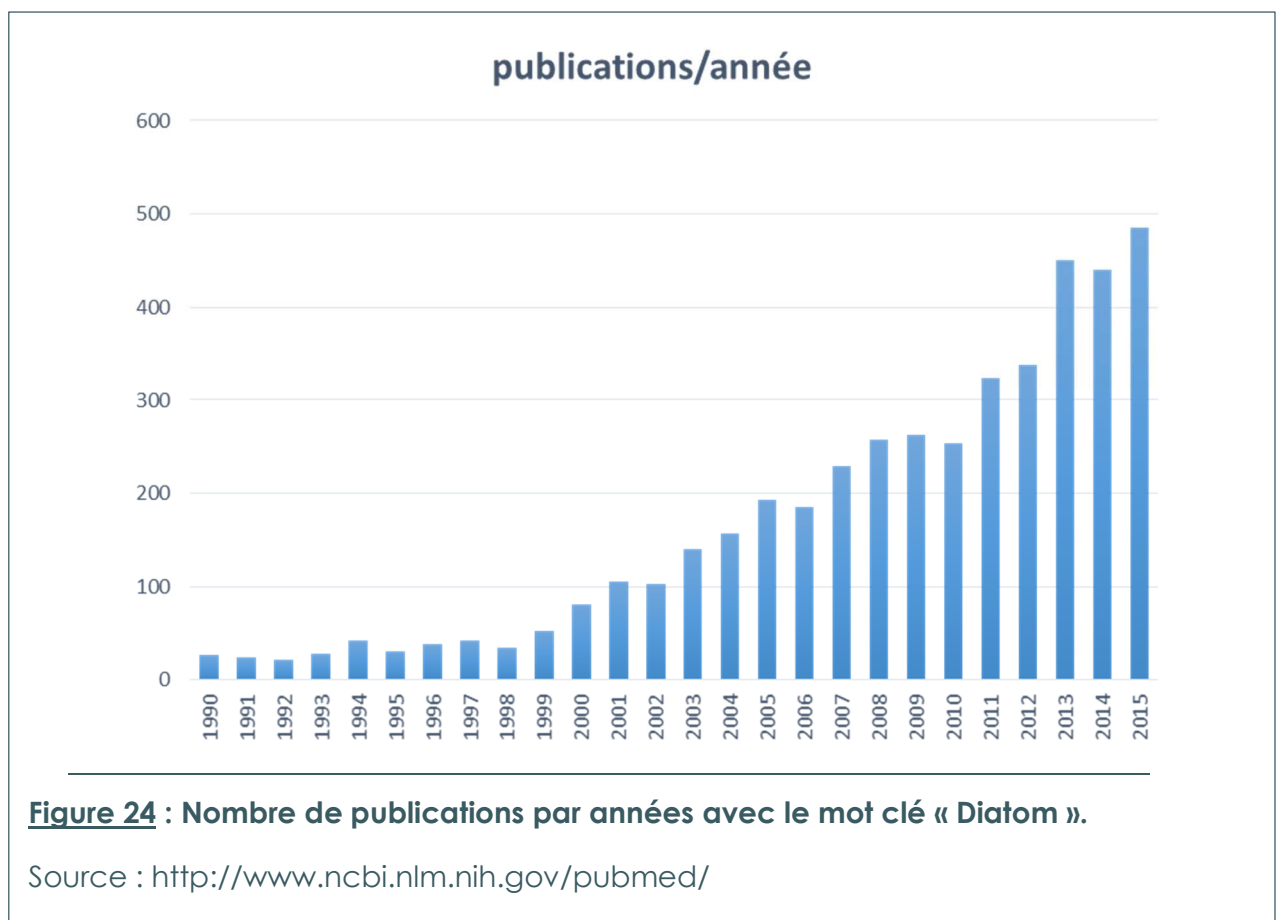
Le morphotype triradié est le morphotype qui présente la plus grande flottabilité, probablement dû à son volume vacuolaire plus important. Ainsi, tout comme le morphotype fusiforme, le morphotype triradié semble plus adapté à la zone pélagique. Les morphotypes triradié

et ovale apparaissent donc être les formes les plus éloignées d'un point de vue biologique, et seraient donc des formes acclimatées à des conditions environnementales opposées (De Martino et al., 2011). Les analyses d'EST chez le triradié montrent une sur-représentation des termes GO et domaines InterPro relatifs aux transports membranaires, ainsi qu'à des régulateurs de transcription impliqués dans le choc thermique, et des protéines impliquées dans des processus liés à l'ADN.

Ainsi, en dépit de ces différentes analyses réalisées sur les trois morphotypes de *Phaeodactylum*, les connaissances sur l'induction du polymorphisme de cette diatomée restent limitées. Des études plus approfondies seraient donc nécessaires afin de mieux cerner les mécanismes moléculaires impliqués dans la différenciation cellulaire et les mécanismes de régulations spécifiques aboutissant à chaque morphotype.

1.3.5 *P. tricornutum* : un organisme de référence

Les microalgues, et les diatomées en particulier, suscitent depuis 2000 un intérêt grandissant pour la communauté scientifique (**figure 24**).



Elles font l'objet d'études non seulement pour mieux comprendre leur rôle dans l'écosystème, mais aussi pour leur potentiel dans le domaine des biotechnologies. La grande capacité d'adaptation des diatomées face aux variations environnementales leur a permis de peupler un grand nombre de niches écologiques différentes. Elles ont ainsi conquis la plupart des habitats aquatiques, et sont une des espèces dominantes du phytoplancton océanique. Les diatomées sont non seulement un des composants majeurs du réseau trophique océanique (Worden et al., 2015), mais elles contribuent aussi jusqu'à 45% à la production primaire océanique, soit environ 20% de la production primaire globale (Field et al., 1998; Mann, 1999). Leur implication écologique ne s'arrête pas là, puisqu'en plus d'être impliquées dans le cycle du carbone, elles interviennent aussi dans le cycle biogéochimique de la silice (Yool and Tyrrell, 2003), et peuvent être utilisées pour de la phytoremédiation (Bozarth et al., 2009).

De plus, les diatomées présentent un grand potentiel pour différents domaines industriels. En effet, ce sont des sources naturelles d'antioxydants, de lipides insaturés, de vitamines et de pigments, ce qui attire particulièrement l'attention des industries cosmétiques et agroalimentaires (Becker, 2007; Kuczynska et al., 2015; Moreno et al., 1979). La composition de leur frustule silicifié intéresse également l'industrie pharmaceutique, en tant que vecteur de médicament (Zhang et al., 2013), et l'industrie des nanotechnologies, comme dispositif microélectronique ou biosensor (Bozarth et al., 2009; Parkinson and Gordon, 1999). À la recherche de nouvelles énergies renouvelables, le domaine de l'énergétique se tourne aussi vers les diatomées. L'intérêt porté à ces organismes provient de leur haute teneur naturelle en huile, propice à la fabrication de biodiesel, mais aussi en sucres permettant la synthèse de bioéthanol, ou encore en protéines pour de la biométhanisation (Bozarth et al., 2009; Dillschneider et al., 2013; Mata et al., 2010).

En plus de leur composition naturellement avantageuse, les diatomées présentent un grand potentiel comme usine cellulaire. En effet, grâce au génie génétique, il est aujourd'hui possible de faire produire à ces cellules des bioplastiques ou encore des composés à haute valeur ajoutée comme des anticorps ou des vaccins (Hempel et al., 2011a, 2011b; Vanier et al., 2015). Ces premiers résultats très prometteurs laissent supposer une ouverture à la production d'autres molécules économiquement intéressantes, comme des hormones ou des enzymes, qui sont actuellement produites par des cellules bactériennes, de levures, de mammifères, ou encore de plantes. La production de ces molécules chez les diatomées présente de nombreux avantages par rapport aux systèmes de production existants. En effet, les diatomées associent la facilité de culture et un taux de croissance élevé, qui peuvent être retrouvés chez les bactéries, à la capacité d'ajouter des modifications post-traductionnelles essentielles à l'activité des protéines les plus

complexes, que l'on retrouve chez les eucaryotes. De plus, les diatomées sont aussi moins coûteuses à produire que les cellules de mammifères et peuvent être considérées comme des organismes GRAS (Generally Recognized As Safe), ce qui permet de s'affranchir des contrôles qualité onéreux associés à la contamination croisée existant entre les pathogènes de cellules de mammifères et les cellules humaines. Les diatomées peuvent être cultivées dans différents environnements y compris dans les déserts (Winckelmann et al., 2015), ce qui évite d'empiéter sur les terres agricoles, et ainsi de rentrer en compétition avec l'industrie agroalimentaire. Elles présentent également l'avantage de pouvoir être cultivées dans des photobioréacteurs, ce qui rend possible le contrôle de la dissémination des algues transgéniques dans l'environnement, et permet d'occuper moins d'espace que les plantes transgéniques qui nécessitent un confinement dans d'énormes serres pour un même niveau de sécurité. Cependant, malgré les avantages évidents que proposent les diatomées, il existe actuellement plusieurs verrous scientifiques qui limitent la généralisation de leur utilisation pour la production de protéines recombinantes à l'échelle industrielle. On peut noter par exemple le faible niveau d'expression des protéines recombinantes par rapport aux systèmes mammifères plus optimisés, ou encore la méconnaissance des mécanismes de modifications post traductionnelles, comme la *N*-glycosylation, qui peuvent avoir un impact sur la demi-vie et l'activité biologique des protéines à intérêt thérapeutique (Lingg et al., 2012), ou encore leur immunogénicité (van Beers and Bardor, 2012).

Ainsi, afin d'approfondir les connaissances sur la biologie des diatomées, la communauté scientifique a dû développer de nouveaux outils. Dans ce contexte, *P. tricornutum* s'est naturellement imposée comme organisme modèle pour l'étude des diatomées grâce à sa facilité de culture, sa capacité à être transformée (Apt et al., 1996), et son petit génome (27,5Mb) totalement séquencé et disponible sur le site du Joint Genome Institute (Bowler et al., 2008). Selon la dernière version annotée du génome (version 3), *P. tricornutum* posséderait 12 178 gènes codants et 177 gènes non codants répartis sur 34 chromosomes. Plusieurs outils sont maintenant disponibles pour cette microalgue, dont notamment une base de données numérique de l'expression des gènes de *P. tricornutum*. Elle a été générée à partir de plus de 130 000 EST (Expressed Sequence Tag) obtenues à partir de cellules cultivées dans 16 conditions différentes (Maheswari et al., 2009, 2010). Des approches de séquençage haut débit (RNA-Seq) ont aussi été utilisées lors d'études menées sur des cultures déplétées en nutriments afin d'observer d'une part les différences d'expression des ARN messagers en fonction des conditions de culture (Valenzuela et al., 2012), et ont permis d'identifier d'autre part des microRNAs spécifiques de *P. tricornutum* (Huang et al., 2011).

Afin de mieux comprendre les mécanismes moléculaires et cellulaires, des vecteurs de sur-expression et de répression des gènes par RNAi ont été élaborés (De Riso et al., 2009) avec un choix de résistance à plusieurs antibiotiques de sélection (Nourseothricin, Puromycin, Chloramphenicol, Erythromycin, Zeocin, Phleomycin ; Apt et al., 1996; Falciatore et al., 1999), et la possibilité d'induire spécifiquement l'expression du transgène (Hempel et al., 2011a). De plus, trois méthodes de transformation sont actuellement décrites pour *P. tricornutum*. Parmi celles-ci, la biolistique est la méthode de transformation la plus répandue pour *P. tricornutum* (Apt et al., 1996; Falciatore et al., 1999), quelques articles font état d'une transformation par électroporation (Miyahara et al., 2013; Niu et al., 2012; Zhang and Hu, 2013) et plus récemment, une méthode très prometteuse de transfection par conjugaison bactérienne a été mise au point (Karas et al., 2015).

En plus de l'accessibilité de son génome et de la disponibilité d'outils de biologie moléculaire, les caractéristiques particulières de *P. tricornutum* en font un organisme de choix pour de multiples domaines d'étude. En effet, cette diatomée est la seule à posséder une silicification conditionnelle de sa paroi, ce qui la rend très intéressante pour l'étude des mécanismes liés à la synthèse du frustule. Elle possède une teneur et une productivité élevée en lipides (Mata et al., 2010), ce qui intéresse particulièrement les recherches sur la production de biodiesel. Et pour finir, sa grande capacité d'adaptation et son pléomorphisme sont des sujets d'étude intéressants pour mieux comprendre les processus de morphogénèse et d'acclimatation des diatomées face à leur environnement ainsi que les processus de biologie cellulaire comme la *N*-glycosylation.

1.4 Processus de *N*-glycosylation dans les systèmes eucaryotes

La *N*-glycosylation est une modification co- et post-traductionnelle acquise par les protéines qui empruntent le système endomembranaire de sécrétion. Cette modification est une étape clé dans la maturation de certaines protéines. Elle consiste en l'ajout de structures glucidiques qui permettront aux protéines d'acquérir leur conformation fonctionnelle, ainsi que des fonctions supplémentaires. L'avantage apporté par cette modification a fait qu'au cours de l'évolution eucaryotique, la *N*-glycosylation est devenue la modification protéique la plus répandue (Aebi, 2013). A titre d'exemple, il existe dans le glycoprotéome d'*A. thaliana* 2 186 sites de *N*-glycosylation, 2 229 sites chez *D. melanogaster*, et il est possible de retrouver dans le glycoprotéome de la souris plus de 2 300 protéines *N*-glycosylées avec au-delà de 5 000 sites de *N*-glycosylation (Zielinska et al., 2010, 2012).

La *N*-glycosylation se déroule spécifiquement dans deux compartiments cellulaires : le réticulum endoplasmique (RE) et l'appareil de Golgi. Ce processus est initié dans le RE où les protéines transloquées co-traductionnellement vont pouvoir subir les premières étapes de la *N*-glycosylation selon un procédé très conservé chez les eucaryotes. Les glycoprotéines vont ensuite être adressées à l'appareil de Golgi où elles subiront éventuellement une maturation de leurs *N*-glycannes. Ce processus métabolique fait intervenir de nombreuses glycosyltransférases et glycosidases qui vont venir agir de manière séquentielle, permettant d'aboutir à la synthèse d'un *N*-glycane mature dit complexe.

1.4.1 Voie de biosynthèse du réticulum endoplasmique

1.4.1.1 Synthèse du précurseur

1.4.1.1.1 Etapes cytosoliques

Le point de départ de la *N*-glycosylation prend place sur la face cytosolique de la membrane du RE, où un précurseur oligosaccharidique va être synthétisé sur une ancre lipidique, le dolichol pyrophosphate (Dol-P) (**figure 25** ; Aebi, 2013). Ce précurseur est assemblé par l'ajout successif de trois monosaccharides différents que sont la *N*-acétylglucosamine (GlcNAc), le mannose (Man) et le Glucose (Glc). Ces trois résidus proviennent du métabolisme primaire sous forme activée de nucléotides-sucres (Aebi et al., 2010). Les UDP-GlcNAc, GDP-Man, et UDP-Glc vont ainsi servir de substrat à des glycosyltransférases spécifiques appelées ALG (Asparagine Linked Glycosylation), qui sont retrouvées à la surface de la membrane du RE.

La voie de synthèse du précurseur oligosaccharidique va débiter par le transfert d'un résidu GlcNAc-P sur un Dol-P transmembranaire, catalysé par l'action d'une *N*-acétylglucosaminyl-phosphate-transférase (ALG7), aboutissant à la formation d'un dolichol pyrophosphate-GlcNAc (**figure 25** ; Aebi, 2013). Un second résidu GlcNAc va être transféré par l'action de deux protéines, les ALG13 et ALG14 (Bickel et al., 2005; Wang et al., 2008). Ces deux protéines permettant le transfert de GlcNAc sont associées sous forme d'un complexe protéique où l'ALG14 joue le rôle de sous-unité organisatrice (Lu et al., 2012; Noffz et al., 2009).

L'ajout des cinq résidus mannose suivants va se faire à l'aide de trois enzymes utilisant du GDP-Man comme substrat. De la même manière que précédemment, ces trois enzymes vont elles aussi former un complexe protéique (Gao et al., 2004). Le premier Man va être ajouté au GlcNAc₂-PP-Dol par la β -1,4-mannosyltransférase ALG1 (Couto et al., 1984), qui va servir d'accepteur pour l'ajout de deux autres résidus Man. Ces deux résidus vont être greffés de

manière successive, en α -1,3, puis en α -1,6 respectivement par la même enzyme : l'ALG2 (Kampf et al., 2009; O'Reilly et al., 2006). L'élongation de l'oligosaccharide va ensuite se poursuivre par l'addition successive de deux nouveaux résidus Man par l'ALG11, ce qui permettra d'aboutir à la formation du $\text{Man}_5\text{GlcNAc}_2\text{-PP-Dol}$, produit final de la biosynthèse du précurseur au niveau cytosolique (Cipollo et al., 2001).

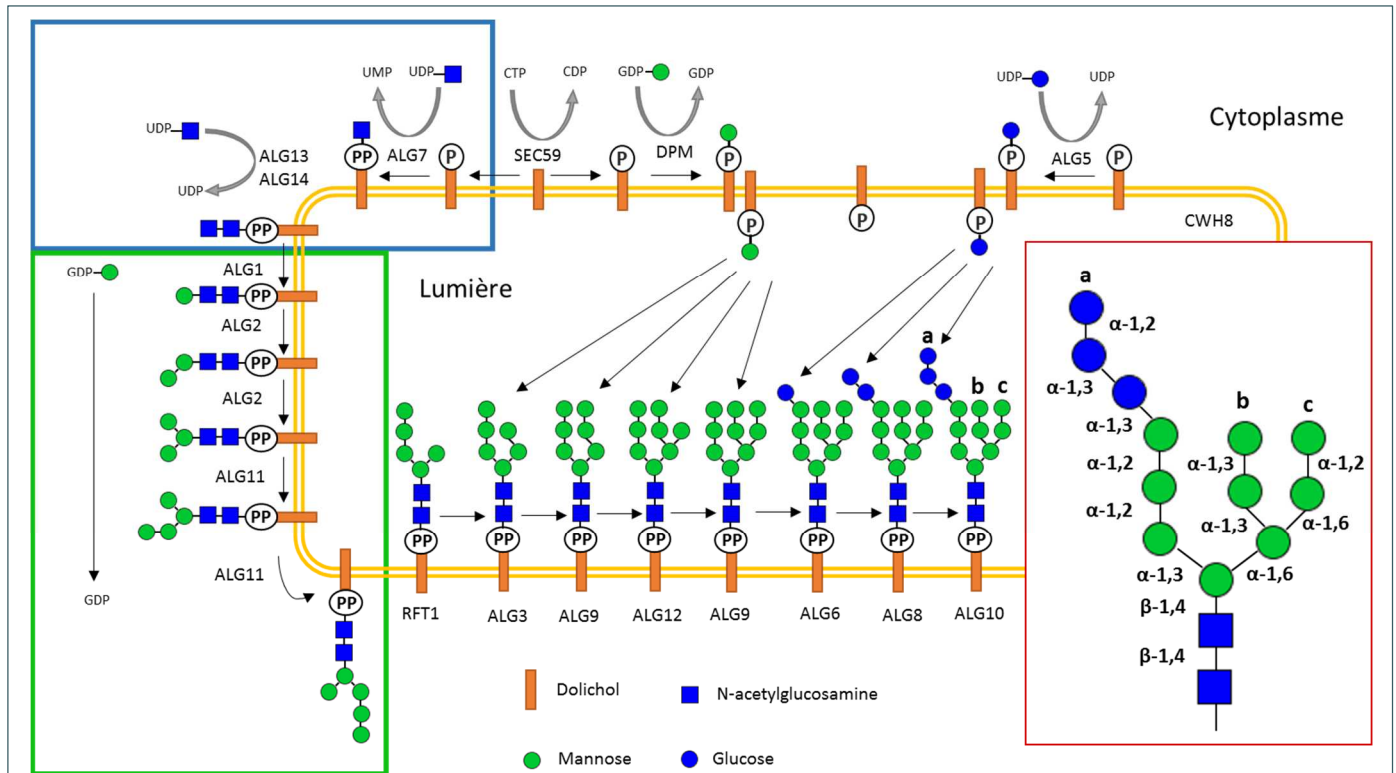


Figure 25 : Voie de biosynthèse du précurseur oligosaccharidique $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ d'après Aebi ,2013.

Ce processus est catalysé par différentes glycosyltransférases spécifiques nommées ALG (Asparagine Linked Glycosylation) indiquées sur le schéma. Les glycosyltransférases actives sur la face cytosolique de la membrane du Réticulum Endoplasmique (RE) qui utilisent l'UDP-GlcNAc et le GDP-Man forment deux complexes enzymatiques représentés chacun par des rectangles respectivement bleu et vert. Les lettres a, b et c désignent les trois antennes du précurseur oligosaccharidique final $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$. Un agrandissement indiquant les différentes liaisons de ce composé est représenté dans le cadre rouge.

1.4.1.1.2 Etapes lumineales

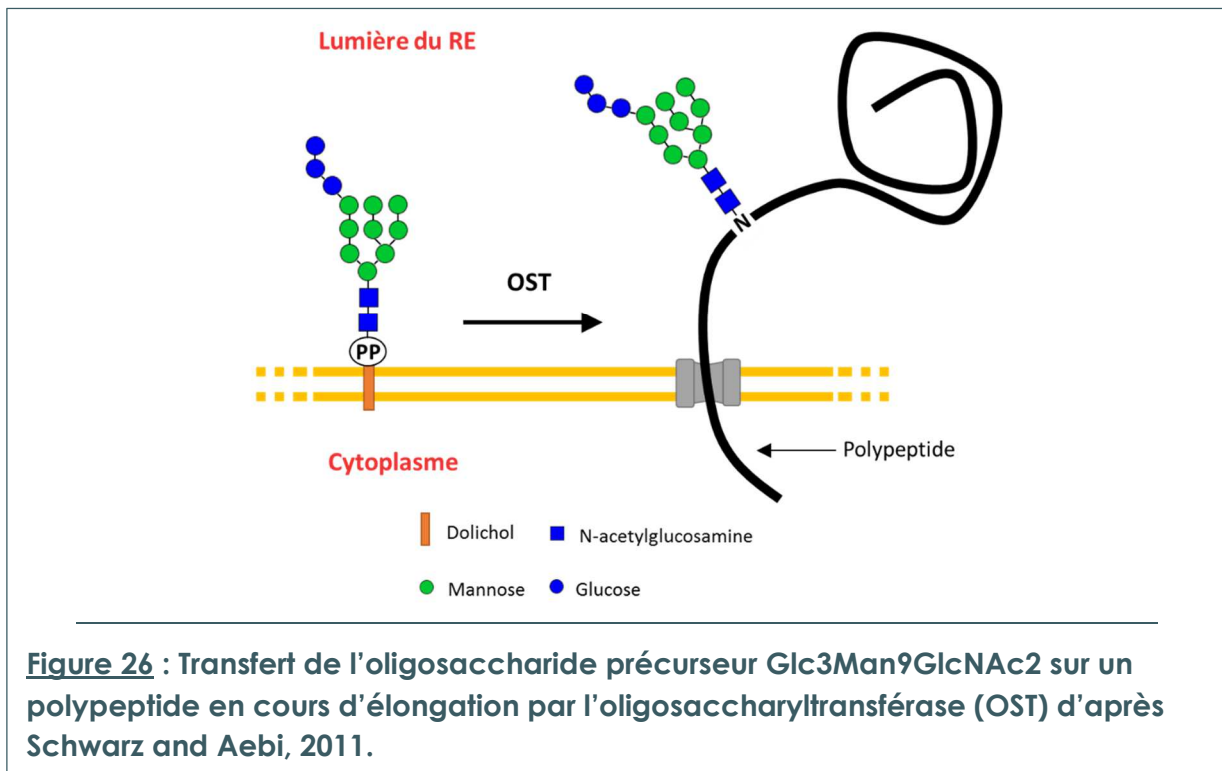
Les étapes suivantes de la biosynthèse du précurseur oligosaccharidique vont se dérouler sur la face luminale de la membrane du RE. Ceci va nécessiter un changement d'orientation du glycolipide, qui va ainsi être transloqué par un mouvement de flip-flop résultant de l'action de la flippase RFT1 (Helenius et al., 2002; Snider and Rogers, 1984). La maturation luminale du

précurseur consiste à ajouter quatre résidus mannose et trois résidus glucose supplémentaires. De manière intéressante, bien que la lumière du RE ne soit pas dépourvue de nucléotides-sucres, les glycosyltransférases catalysant ces réactions vont utiliser comme substrat respectivement du Dol-P-Man et du Dol-P-Glc. Cette particularité reflète probablement l'origine procaryotique de cette voie biologique (Schwarz and Aebi, 2011). Ces deux composés sont initialement synthétisés sur la face cytosolique de la membrane du RE, puis vont ensuite être transloqués à travers la membrane afin d'être utilisés comme substrat dans la lumière du RE (Rush and Waechter, 1995; Rush et al., 1998; Sanyal and Menon, 2010). Le Dol-P-Man est ainsi le produit d'une réaction catalysée par la Dol-P-Man synthase DPM1, au cours de laquelle un résidu mannose issu d'un GDP-Man est ajouté sur un Dol-P (Orlean et al., 1988). De manière analogue, le Dol-P-Glc va être synthétisé par ajout d'un résidu glucose provenant d'un UDP-Glc sur un Dol-P grâce à la Dol-P-Glc synthase ALG5 (Heesen et al., 1994).

Une fois l'oligosaccharide transféré sur la partie luminale de la membrane du RE, un premier mannose va être ajouté sur le précurseur via l'action de l' α -1,3-mannosyltransférase ALG3, aboutissant à la synthèse de $\text{Man}_6\text{GlcNAc}_2\text{-PP-Dol}$ (Aebi et al., 1996). Un second mannose va être greffé en α -1,2 sur le précédent par l'action de l' α -1,2-mannosyltransférase ALG9, formant ainsi l'antenne « b » de l'oligosaccharide (Frank and Aebi, 2005). L'élongation de l'antenne « c » va pouvoir ensuite être initiée par l'action de l' α -1,6-mannosyltransférase ALG12. La mannosylation du précurseur va se terminer par l'intervention de l' α -1,2-mannosyltransférase qui va ajouter le dernier mannose de la chaîne « c », aboutissant à la formation du glycoterpène $\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$ (Burda et al., 1999; Frank and Aebi, 2005). Suite à cela, la maturation du précurseur va pouvoir se terminer par une dernière étape de glucosylation assurée par trois enzymes différentes. La première à intervenir est l'ALG6 qui va permettre d'initier la glucosylation de l'antenne « a » de l'oligosaccharide (Reiss et al., 1996). Un second résidu Glucose va pouvoir ensuite être greffé en α -1,3 sur le précédent sous l'action de la glucosyltransférase ALG8 (Stagljar et al., 1994). Et enfin, la dernière enzyme à intervenir est l'ALG10 qui va permettre de catalyser l'ajout de glucose en α -1,2 sur le précédent résidu, finalisant ainsi la synthèse du précurseur sous la forme $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$ (Burda and Aebi, 1998). Ce dernier glucose aura une importance particulière pour la reconnaissance de l'oligosaccharide par l'OST lors du transfert sur la protéine (Aebi, 2013).

1.4.1.2 Transfert du précurseur par l'OST

Une fois le précurseur oligosaccharidique synthétisé, ce dernier va être transféré en bloc, de manière co-translationnelle, par l'oligosaccharyltransférase (OST) sur les sites consensus de *N*-glycosylation des protéines en cours de synthèse (**figure 26**). Cette enzyme forme un complexe multimérique composé d'enzymes transmembranaires situées au niveau du RE (Mohorko et al., 2011; Yan and Lennarz, 1999). Les sous-unités du complexe protéique peuvent varier en fonction des organismes, mais la sous-unité STT3, responsable de l'activité catalytique, est hautement conservée chez tous les eucaryotes (Schwarz and Aebi, 2011). L'OST va utiliser le précurseur $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$ comme substrat donneur, et les protéines contenant un site consensus de *N*-glycosylation comme substrat accepteur (Bause, 1983; Breton and Imberty, 1999; Helenius and Aebi, 2004). L'OST va ainsi catalyser la formation d'une liaison *N*-osidique entre le précurseur oligosaccharidique et la fonction amide de la chaîne latérale d'une asparagine appartenant à une séquence consensus Asn-X-Ser/Thr/Cys (où X peut être n'importe quel acide aminé, excepté la proline) d'un polypeptide (Imperiali and Hendrickson, 1995; Zielinska et al., 2010).



L'OST possède une spécificité de substrat très élevée pour le précurseur oligosaccharidique (Shrimal et al., 2013), dont la structure est très conservée chez les eucaryotes. Il a cependant été déterminé *in vitro* que le motif minimal du précurseur pouvant être accepté par l'OST est de la forme $\text{GlcNAc}_2\text{-PP-Dol}$ (Tai and Imperiali, 2001), bien qu'*in*

in vivo la préférence de l'enzyme va nettement pour l'oligosaccharide entièrement assemblé (Aebi, 2013; Burda et al., 1999). Il a en effet été démontré que le résidu α -1,2-glucose terminal de la chaîne « a » était un élément important pour la reconnaissance du substrat par l'OST, alors que les résidus mannose des chaînes « b » et « c » sont d'une importance bien plus limitée (Aebi, 2013; Burda et al., 1999). Il a cependant été démontré que l'OST de certains protistes a la capacité de transférer des précurseurs non glycosylés, voir dépourvus de certains mannoses (Kelleher et al., 2007).

Contrairement à sa haute spécificité de reconnaissance avec le précurseur oligosaccharidique, l'OST possède une spécificité plus souple pour le substrat accepteur. Cette propriété lui permet ainsi de s'adapter à la grande variété de protéines empruntant le système endomembranaire de sécrétion. Cependant, malgré cette capacité d'adaptation à des molécules aux propriétés physico-chimiques différentes, l'OST n'est capable d'ajouter un oligosaccharide qu'aux sites de *N*-glycosylation définis par la séquence Asn-X-Ser/Thr, et plus rarement par Asn-X-Cys, malgré la disponibilité d'une grande diversité de séquences polypeptidiques (Breuer et al., 2001). Des études menées sur les sites d'occupation des glycoprotéines ont néanmoins observé une préférence de l'OST pour les sites Asn-X-Thr plutôt que pour les sites Asn-X-Ser (Chen et al., 2007; Gavel and von Heijne, 1990). Cette spécificité de séquence peut être expliquée au niveau moléculaire par la nécessité d'une torsion à 180°C du peptide, avec un acide aminé hydroxylé à la position +2 du site accepteur, permettant au peptide d'être reconnu par le site de liaison de l'OST situé dans une poche formée par l'enzyme. Cette région située sur la sous-unité STT3 de l'OST est formée par un motif WWDYG hautement conservé (Lizak et al., 2011; Yan and Lennarz, 2002).

En complément, il est important de noter qu'*in vivo*, tous les sites potentiels de *N*-glycosylation ne sont pas forcément glycosylés. En effet, il a été estimé que seulement 2/3 de ces sites sont utilisés comme substrat par l'OST (Apweiler et al., 1999; Zielinska et al., 2010). Ces variations de *N*-glycosylation seraient influencées par les séquences protéiques entourant les sites consensus (Petrescu et al., 2003), la flexibilité de la région (Lizak et al., 2011; Whitley et al., 1996), par la proximité de domaine transmembranaire ou d'un autre site de *N*-glycosylation (Nilsson and Von Heijne, 1993).

Le transfert du précurseur oligosaccharidique par l'OST va ainsi permettre de donner naissance à une glycoprotéine, modifiant les propriétés biophysico-chimiques de la protéine cible, et affectant par conséquence son repliement de façon directe (Culyba et al., 2011; Hanson et al., 2009; Shental-Bechor and Levy, 2008) ou indirecte (Helenius and Aebi, 2004).

1.4.2 Contrôle qualité des glycoprotéines

Après son transfert sur la protéine, le *N*-glycane va être rapidement restructuré par des glycosidases résidentes du RE. Dans un premier temps, l'oligosaccharide $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ va subir l'élimination de son glucose distal par l'action de l' α -glucosidase I (**figure 27** ; Barker and Rose, 2013; Hettkamp et al., 1984; Shailubhai et al., 1991). Cette α -1,2-exoglucosidase est une protéine membranaire de type II associée au complexe de translocation, à proximité de l'OST (Deprez et al., 2005). Tout comme la forme triglucosylée du précurseur, l'oligosaccharide $\text{Glc}_2\text{Man}_9\text{GlcNAc}_2$ fixé sur la protéine possède une demi-vie très courte. En effet, après l'élimination du glucose terminal, le glycane di-glucosylé va rapidement servir de substrat pour l' α -glucosidase II qui va hydrolyser le second glucose (Brada and Dubach, 1984; Kaewmuangmoon et al., 2013). Cette α -1,3-exoglucosidase forme un hétérodimère soluble composé d'une sous-unité α catalytique, et d'une sous-unité β régulatrice (Trombetta et al., 1996; Watanabe et al., 2013). Cependant, bien que cette dernière sous-unité ne soit pas indispensable à l'activité catalytique, elle est nécessaire pour la maturation de la protéine, sa solubilité et sa stabilité, et contient le signal de rétention dans le RE (Lamriben et al., 2015).

L'oligosaccharide ainsi modifié va servir de ticket d'entrée à la protéine pour le cycle de repliement et de contrôle qualité. En effet, les *N*-glycannes jouent un rôle essentiel dans ces processus. Bien que toutes les protéines ne soient pas dépendantes de la *N*-glycosylation pour adopter leur conformation native, il a été observé *in vitro* un effet direct des *N*-glycannes sur le repliement de protéines glycosylées (Imperiali and O'Connor, 1999; Kern et al., 1993). Il a aussi été montré *in vivo* qu'après traitement de cellules par la tunicamycine (inhibiteur de la *N*-glycosylation), des protéines s'agrégeaient dans le RE et que certaines protéines qui arrivaient à être excrétées perdaient leur fonction (Helenius, 1994; Paulson, 1989). Ainsi, les *N*-glycannes permettent aux protéines d'adopter leur conformation fonctionnelle, notamment grâce à leur forme monoglucosylée qui va être reconnue par des lectines chaperonnes localisées dans le RE (Hahn et al., 1998; Molinari et al., 2004).

La forme $\text{GlcMan}_9\text{GlcNAc}_2$ résultant de l'action des α -glucosidases I et II va ainsi être reconnue par deux chaperonnes, la calnexine et la calréticuline (**figure 27**), qui associées chacune à une co-chaperonne ERp57, vont favoriser le repliement des glycoprotéines (High et al., 2000). Ces deux chaperonnes présentent la même organisation générale, exception faite que la calnexine est une protéine transmembranaire de type I, ce qui lui permettra d'interagir plutôt avec les glycoprotéines proches de la membrane, alors que la calréticuline est une

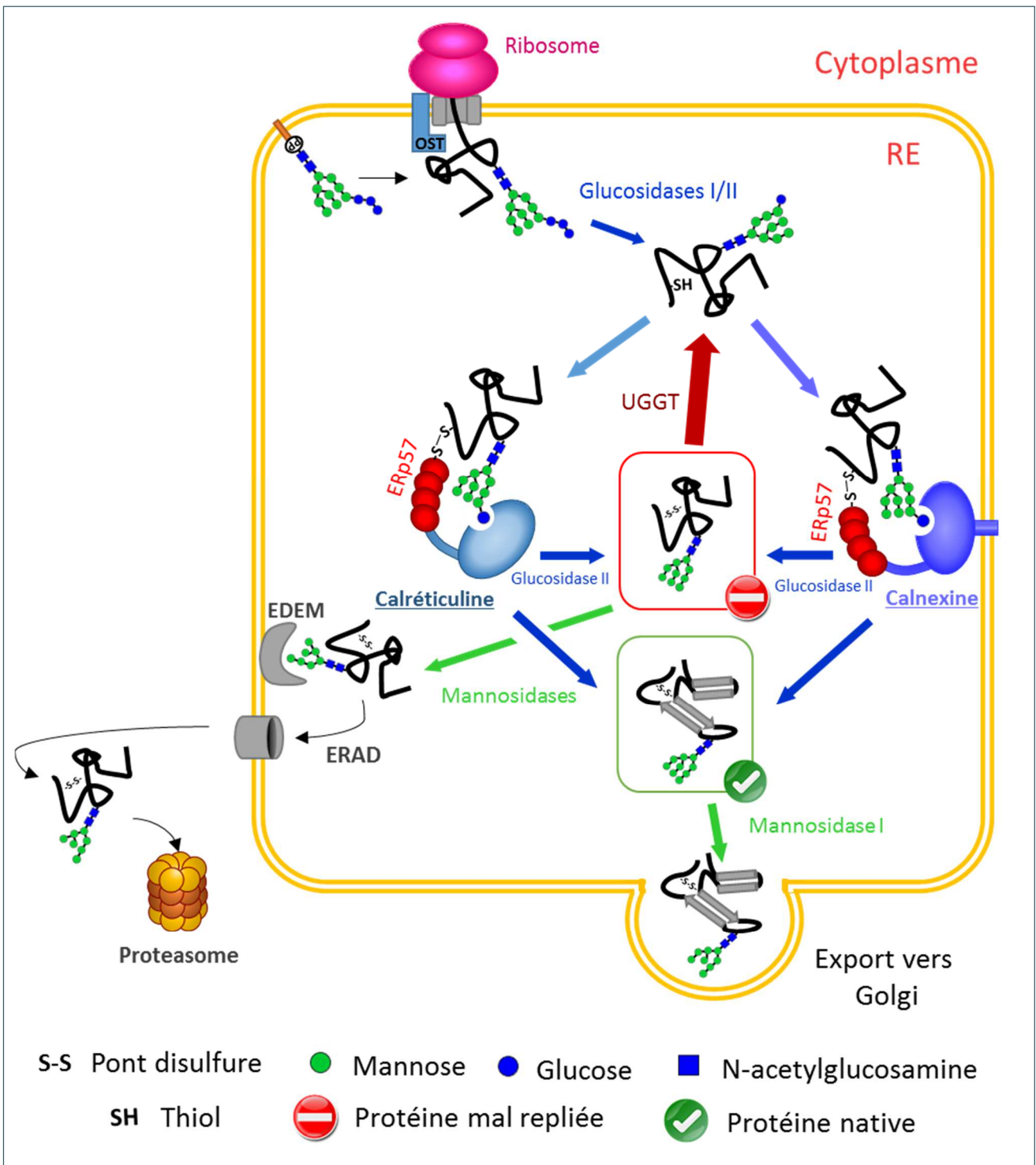


Figure 27 : Repliement et cycle de contrôle qualité des glycoprotéines N-liées dans le réticulum endoplasmique (RE) d'après Williams, 2006 ; Lamriben et al., 2015.

Ces processus impliquent deux lectines chaperonnes que sont la calnexine et la calréticuline, ainsi que la co-chaperonne ERp57. Cette voie est initiée par la déglucosylation de la glycoprotéine par les Glucosidases I et II. Les protéines pourront ainsi entrer dans le cycle de la calnexine (lectine transmembranaire) et de la calréticuline (lectine soluble) impliquant l'action de la glucosidase II et de l'UDP-Glucose : Glycoprotéine Glucosyltransférase (UGGT). Les glycoprotéines mal repliées seront dégradées par l'intermédiaire du système EDEM/ERAD/Protéasome. Les glycoprotéines natives pourront transiter vers l'appareil de Gogi grâce à l'intervention de la mannosidase I du RE.

protéine soluble, qui va plutôt venir s'associer avec des glycoprotéines luminales (Lamriben et al., 2015; Pieren et al., 2005; Wada et al., 1995). Ces deux protéines comprennent chacune trois domaines : un domaine globulaire *N*-terminal, un domaine *C*-terminal, et un domaine central riche en proline. Le domaine *N*-terminal est celui qui possède l'activité lectine permettant la liaison au *N*-glycane mono-glucosylé (Schrag et al., 2001). Les sites de liaison de ce domaine seraient notamment régulés par des co-facteurs comme l'ATP et le calcium (Li et al., 2001; Wijeyesakere et al., 2015). Les régions *C*-terminales des deux protéines divergent mais contiennent toutes deux le signal de rétention dans le RE (Lamriben et al., 2015). Et enfin, le domaine riche en proline situé au centre du polypeptide va former le « bras » de la protéine. Ce domaine est constitué d'une série de motifs séquentiels permettant l'interaction avec l'oxidoréductase ERp57, impliquée notamment dans la formation des ponts disulfure sur les peptides en cours de synthèse (Frickel et al., 2002; Kozlov et al., 2006).

L'interaction entre le polypeptide et la calnexine, ou calréticuline, va se terminer avec le clivage du dernier résidu glucose par l' α -glucosidase II qui va générer la structure Man₉GlcNAc₂ (Ellgaard and Helenius, 2001). Les protéines ayant acquis leur conformation native vont pouvoir être hydrolysées par une α -mannosidase du RE qui va retirer un α -1,2-mannose préférentiellement sur l'antenne « b » de l'oligosaccharide (Mast and Moremen, 2006; Wilson, 2012). Après avoir reçu cette approbation du système de contrôle qualité, les protéines correctement repliées vont pouvoir ensuite être transportées vers l'appareil de Golgi.

À l'inverse, les glycoprotéines n'ayant pas acquis leur conformation native vont être reconnues par l'UDP-glucose : glycoprotéine glucosyltransférase (UGGT) qui a pour fonction de remettre un glucose à la place de celui hydrolysé par l' α -glucosidase II (Sousa and Parodi, 1995). Cette protéine soluble possède à son extrémité *C*-terminale le domaine catalytique glucosyltransférase, et à son extrémité *N*-terminale un grand domaine qui lui permet de détecter les protéines mal repliées (Aebi et al., 2010; Guerin and Parodi, 2003; Tessier et al., 2000). *In vitro*, il a été montré que l'un des facteurs de reconnaissance principal de l'UGGT était l'exposition de clusters d'acides aminés hydrophobes sur la glycoprotéine mal repliée (Caramelo et al., 2003). L'UGGT est ainsi capable de détecter des défauts de repliement hautement localisés, et par ce fait, de re-glucosyler uniquement les chaînes de *N*-glycannes présentes dans les régions mal repliées (Ritter and Helenius, 2000). Une fois re-glucosylées, les protéines non conformes vont pouvoir être réintégrées dans un nouveau cycle de contrôle qualité grâce à la reconnaissance du précurseur monoglucosylé par les calnexine et calréticuline. Ce nouveau contact avec les chaperonnes donne à la glycoprotéine une nouvelle

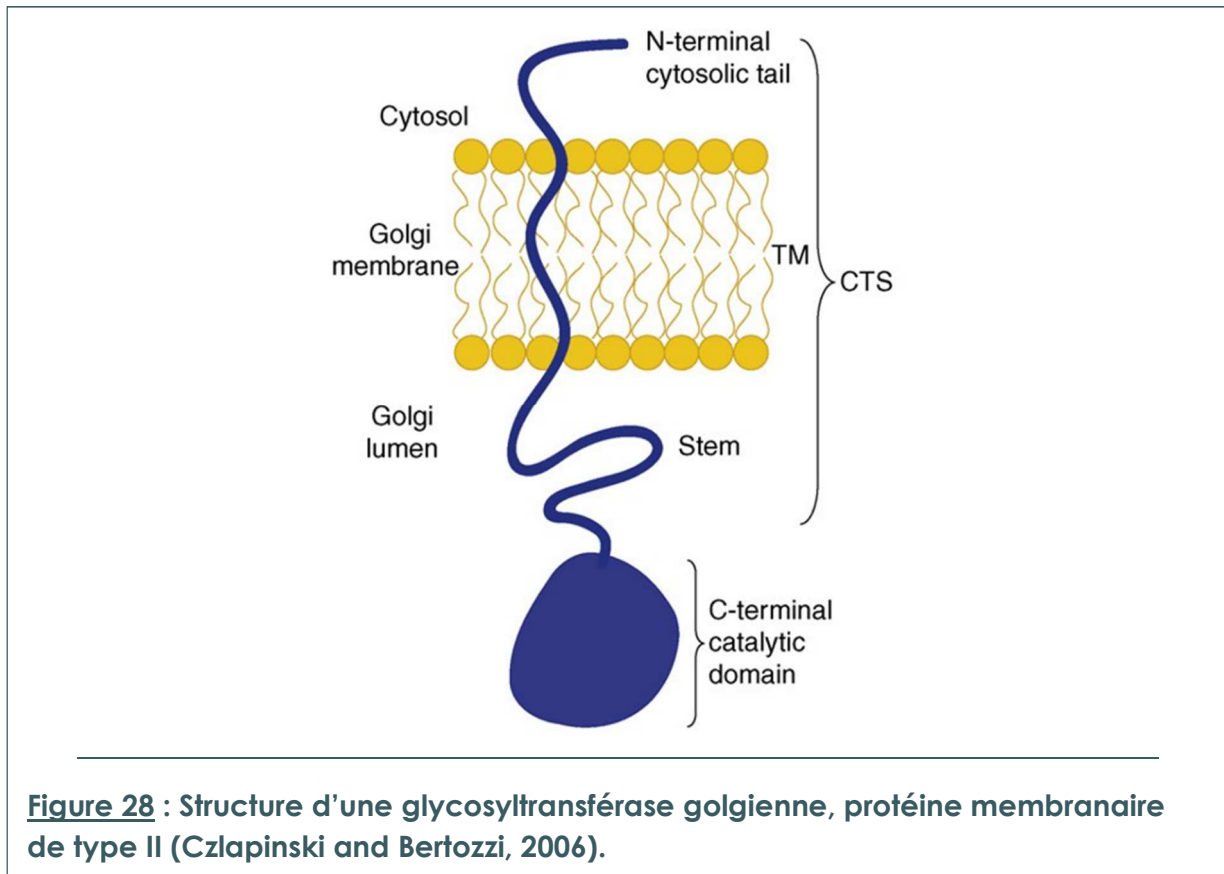
chance d'adopter sa conformation fonctionnelle. Les glycoprotéines vont ensuite être de nouveau libérées du complexe par l' α -glucosidase II, qui par son action opposée entre en compétition avec l'UGGT. Ainsi, les protéines mal repliées ne pourront sortir de ce cycle de contrôle que si elles ne sont pas reglucosylées par l'UGGT. Cette enzyme a donc un rôle décisif dans le RE puisqu'elle va permettre soit d'orienter les glycoprotéines vers l'appareil de Golgi, soit de les renvoyer dans le cycle de contrôle qualité. L'action de l'UGGT, associée à la rétention exercée par les chaperonnes, permet ainsi d'éviter la circulation de protéines non fonctionnelles, et potentiellement néfastes, en dehors du RE (Ellgaard and Helenius, 2003).

Les protéines mal repliées rejetées par l'UGGT subiront l'hydrolyse de deux mannoses par l' α -1-2-mannosidase I (Aikawa et al., 2014), ce qui va leur permettre d'être reconnues par la protéine EDEM (ER Degradation-Enhancing Mannosidases), puis d'être transloquées à travers la membrane pour être dégradées par le système ERAD (ER Associated protein Degradation) couplé aux protéasomes dans le cytoplasme (Määttä et al., 2010; Meusser et al., 2005; Olivari et al., 2006; Parodi, 2000).

1.5 Voie de biosynthèse de l'appareil de Golgi

Les glycoprotéines ayant acquis leur conformation vont pouvoir sortir du RE pour être amenées dans l'appareil de Golgi via le compartiment intermédiaire RE-Golgi (ER-Golgi intermediate compartment; ERGIC) par transport vésiculaire (Sato et al., 2014). L'acheminement des glycoprotéines à l'intérieur des vésicules de transport est réalisé grâce à l'intervention de deux lectines réceptrices, ERGIC-53 et VIP36, qui capturent les protéines via leurs chaînes glycaniques grâce à leurs domaines de reconnaissance lectiniques (Carbohydrate-Recognition Domains; CRDs) structurellement semblables à celui de la concanavaleine A (Fiedler et al., 1994; Hauri et al., 2000). Les glycoprotéines ayant atteint l'appareil de Golgi vont poursuivre la maturation de leurs *N*-glycannes, sous réserve d'accessibilité de ceux-ci par les enzymes de maturation golgiennes (Czlapinski and Bertozzi, 2006; Faye et al., 1986; Lu and Wold, 1991).

Ces enzymes de maturation golgiennes sont des protéines transmembranaires de type II possédant une courte queue cytosolique en région *N*-terminale, un domaine transmembranaire et une tige luminale en *C*-terminal (**figure 28**). Ces différentes parties forment la région CTS (Cytoplasmic-Transmembrane-Stem) impliquée notamment dans l'adressage de ces protéines à l'appareil de Golgi (Saint-Jore-Dupas et al., 2006). La région *C*-terminale contient quant à elle le domaine catalytique de l'enzyme (Breton, 2005; Breton and Imberty, 1999; Schoberer et al., 2013).



Par opposition aux mécanismes très conservés du RE, la maturation golgienne des *N*-glycannes a divergé au cours de l'évolution, permettant ainsi la création de structures espèce-spécifiques (Paulson, 1989). Cette diversité glycanique provient de l'acquisition et de l'expression d'un répertoire bien spécifique de glycosyltransférases et glycosidases propres à chaque organisme. L'intervention de ces différentes enzymes va se faire de manière séquentielle en empruntant deux voies possibles selon les espèces : une voie *N*-acétylglucosaminyltransférase I (GnT I) dépendante, ou une voie GnT I indépendante (**figure 29**). La deuxième voie, plus rarement observée, aboutit par exemple à la formation de *N*-glycannes oligomannosidiques pouvant posséder différents motifs glucidiques et groupements méthyl, comme chez la microalgue *C. reinhardtii* (**figure 29** ; Mathieu-Rivet et al., 2013), ou encore à des *N*-glycannes hyperglycosylés, comme chez la levure *S. cerevisiae*, qui peuvent posséder jusqu'à une centaine de résidus mannose par chaîne oligosaccharidique (De Pourcq et al., 2010). Cependant, la majorité des organismes eucaryotes étudiés jusqu'à présent, comme les mammifères, les plantes ou les invertébrés, possèdent une voie de maturation golgienne des *N*-glycannes qui dépend de la GnT I. Ainsi, ces organismes vont faire intervenir une α -mannosidase I golgienne (Lal et al., 1994 ; Kajiura et al., 2010). Cette enzyme va permettre

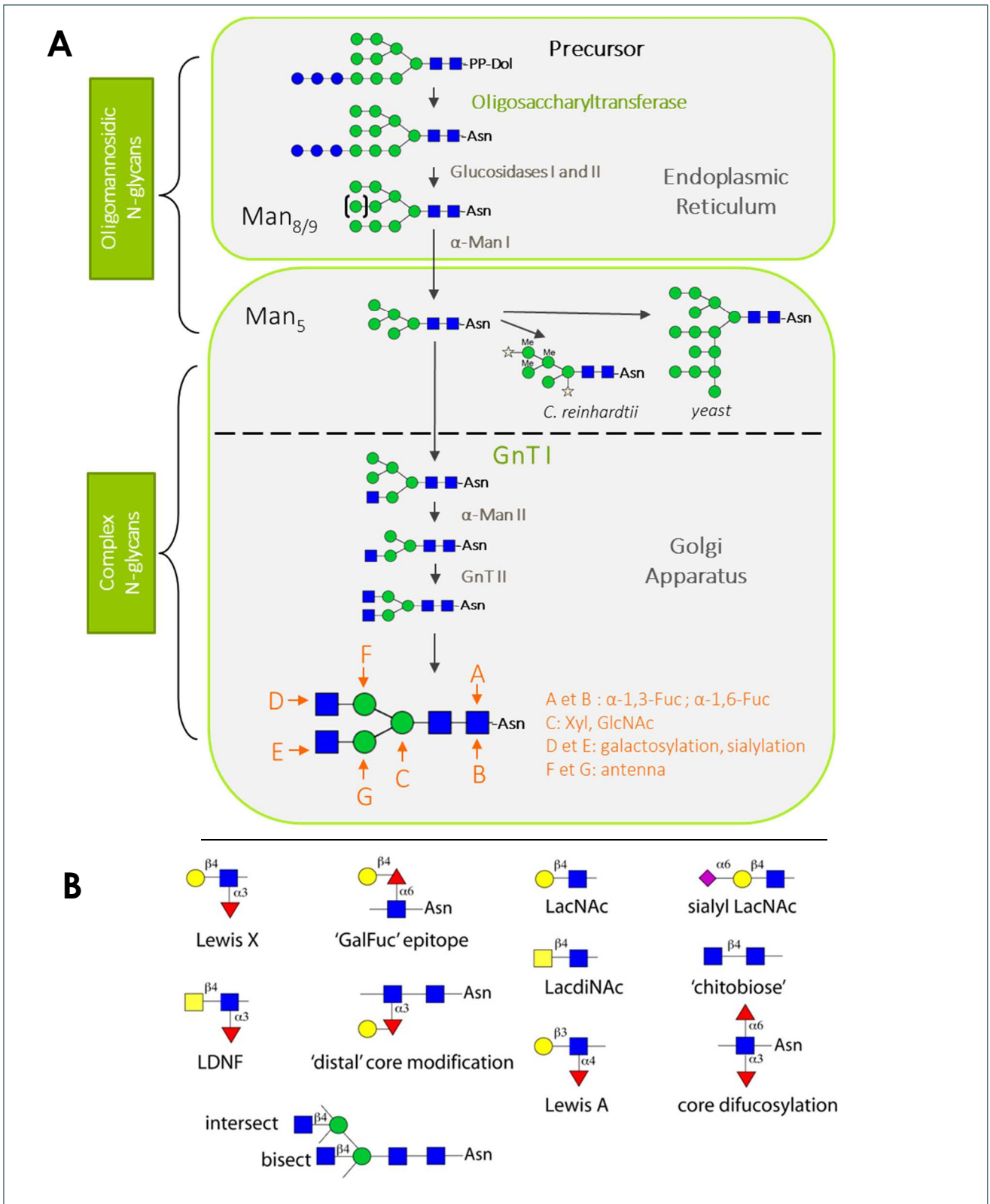


Figure 29 : Biosynthèse et maturation des N-glycannes dans le réticulum endoplasmique et l'appareil de Golgi.

A = Schémas de la voie de biosynthèse des N-glycannes. La ligne pointillée représente l'entrée dans la voie de maturation GnT I dépendante des N-glycannes complexes.

B = Représentation de différents épitopes pouvant être rencontrés sur des N-glycannes complexes.

Légende =
 ● Galactose ■ N-acetylgalactosamine ▲ Fucose ☆ Xylose
 ● Mannose ■ N-acetylglucosamine ◆ Sialic acid (N-acetylneuraminic acid)

Asn Asparagine PP-Dol Dolichol pyrophosphate

l'hydrolyse de un à quatre résidus mannose de l'oligosaccharide Man-8/9, permettant d'aboutir à la formation de *N*-glycannes oligomannosidiques possédant de 8 à 5 résidus mannose (Sturm et al., 1987; Szumilo et al., 1986). L'oligomannoside final à 5 résidus mannose va ainsi pouvoir être utilisé comme substrat par la GnT I, enzyme clé permettant d'initier la voie de biosynthèse des *N*-glycannes complexes. En effet, chez les eucaryotes, l'ajout d'un résidu GlcNAc en β -1,2 sur le mannose α -1,3 du Man-5 (Tezuka et al., 1992; Vella et al., 1984) est un pré-requis pour l'action ultérieure de toutes les enzymes golgiennes intervenant dans la maturation des *N*-glycannes complexes (Johnson and Chrispeels, 1987). La GnT I a ainsi une importance toute particulière, car contrairement aux *N*-glycannes oligomannosidiques, dont le rôle est essentiellement structurel, les *N*-glycannes complexes vont permettre de conférer aux protéines des fonctions biologiques supplémentaires. Ainsi chez les mammifères, tout comme l'absence de *N*-glycosylation, la mutation de la GnT I endogène est embryo-létale (Marek et al., 1999; Metzler et al., 1994). De plus, il est admis que les *N*-glycannes complexes sont impliqués en tant que ligands dans de nombreux processus d'interaction protéine/sucre intervenant par exemple dans le ciblage, l'adhésion cellulaire ou encore la signalisation (Varki, 1993).

Suite à l'action de la GnT I, l' α -mannosidase II va hydrolyser successivement les deux résidus mannose liés à l' α -1,6-mannose du Man-5 (Altmann and März, 1995; Kaushal et al., 1990; Strasser et al., 2006). Ils seront remplacés par un deuxième résidu GlcNAc ajouté en β -1,2 par la N-acétylglucosaminyltransférase II (GnT II), aboutissant à la formation du core glycanique de structure $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ (Johnson and Chrispeels, 1987; Strasser et al., 1999). Cet oligosaccharide est la dernière structure glycanique commune rencontrée chez les plantes et les mammifères (Lerouge et al., 1998).

Différentes glycosidases et glycosyltransférases propres à chaque organisme et exprimées de manière spécifique dans chaque tissu, vont ensuite intervenir. L'action de ces enzymes permet d'aboutir à des *N*-glycannes complexes présentant un core glycanique bi-antenné décoré de motifs glucidiques spécifiques. La grande biodiversité des *N*-glycannes complexes est ainsi liée d'une part à l'ajout de ces résidus glucidiques par une sélection de glycosyltransférases, telles que les fucosyltransférases (FucT), les galactosyltransférases (GalT), les xylosyltransférases (XylT) ou les sialyltransférases, mais aussi à la suppression de résidus catalysé par des glycosidases comme les hexosaminidases qui peuvent hydrolyser les résidus GlcNAc terminaux. Afin d'illustrer ce processus nous allons nous intéresser aux *N*-glycannes complexes produits par les plantes et les mammifères qui sont parmi les structures les mieux caractérisées actuellement (**figure 30** ; Schiller et al., 2012).

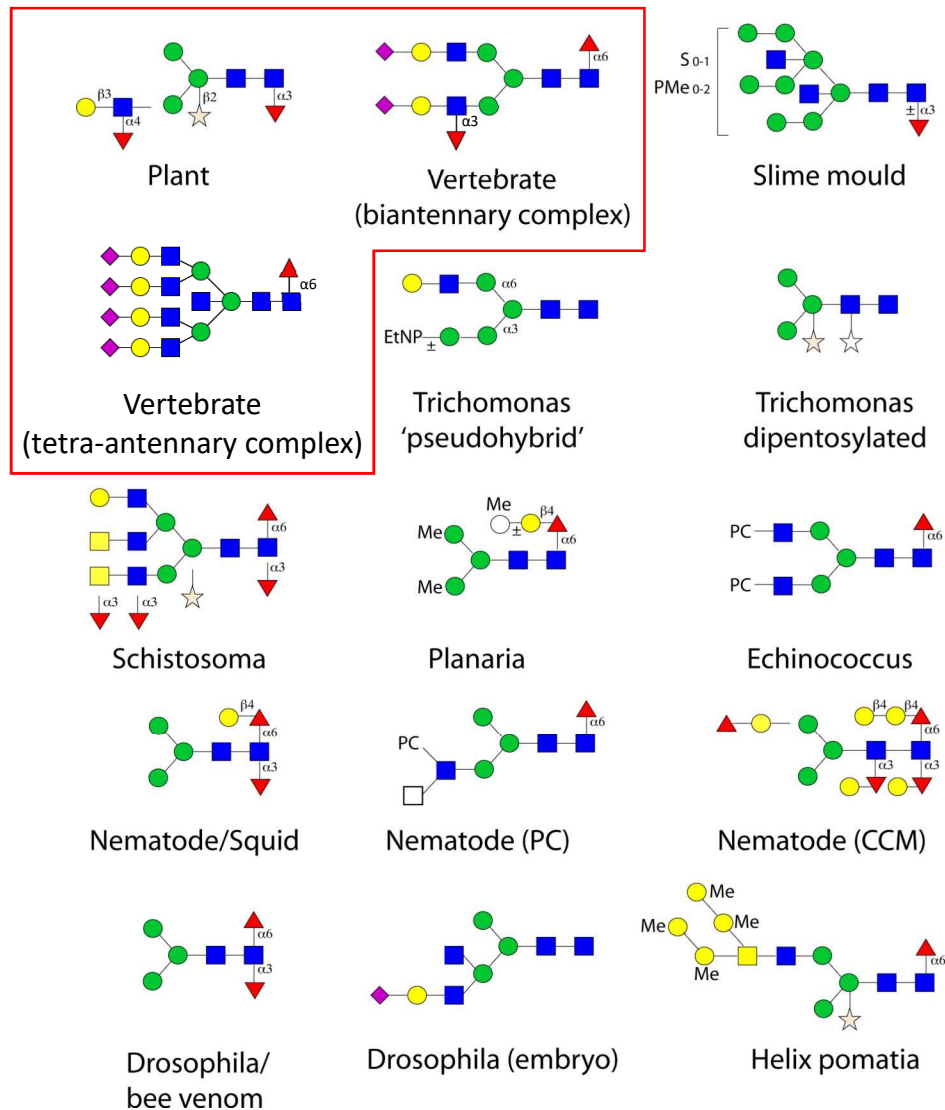


Figure 30 : Exemples de structures N-glycanniques complexes rencontrées chez différents eucaryotes, d'après Schiller et al., 2012.

Les N-glycannes de mammifères et de plantes cités en exemple ont été encadrés en rouge.

Légende : CCM = core chitobiose modification, EtNP = ethanolamine phosphate,

Me = methyl, PC = phosphorylcholine, PMe = methylphosphate, S = sulphate.

- | | | | |
|-------------|-------------------------|---|----------|
| ● Galactose | ■ N-acetylgalactosamine | ▲ Fucose | ☆ Xylose |
| ● Mannose | ■ N-acetylglucosamine | ◆ Sialic acid (N-acetylneuraminic acid) | |

Chez les plantes, la maturation du core glycanne débute par l'intervention de l' α -1,3-fucosyltransférase et de la β -1,2-xylosyltransférase qui catalysent respectivement l'ajout d'un résidu fucose en α -1,3 sur la GlcNAc proximale et d'un résidu xylose en β -1,2 sur le premier mannose du core (Breton et al., 1998a; Johnson and Chrispeels, 1987; Lerouge et al., 1998; Wilson, 2002). Des maturations tardives faisant intervenir une β -1,3-galactosyltransférase et une α -1,4-fucosyltransférase vont ensuite pouvoir être réalisées, permettant d'aboutir à la

formation d'épitope Lewis^a (Breton et al., 1998a; Fitchette-Lainé et al., 1997; Melo et al., 1997; Strasser et al., 2007).

Concernant les mammifères, leurs *N*-glycannes matures font preuve d'une grande variété, pouvant présenter des structures possédant de deux à quatre antennes et/ou une GlcNAc bissectrice liée en β -1,4 au premier mannose du core (Takahashi et al., 2009). Ces différentes structures sont biosynthétisées grâce à l'intervention des *N*-acétylglucosaminyltransférases III à VI sur l'oligosaccharide GlcNAc₂Man₃GlcNAc₂ (Brockhausen and Schachter, 1996). Sur ces différentes antennes vont pouvoir venir se greffer différents résidus comme du galactose qui peut être ajouté en β -1,4 aux GlcNAc distales, contrairement aux plantes où ce résidu est retrouvé lié en β -1,3 (Breton et al., 1998b; Imberty et al., 1999). Ces résidus galactose sont généralement modifiés par l'ajout d'acides sialique en α -2,3- ou α -2,6 catalysé par des sialyltransférases (Audry et al., 2011; Choi et al., 2014; Jeanneau, 2004). De plus, contrairement aux plantes qui présentent un fucose lié en α -1,3 sur la GlcNAc proximale, la fucosylation du core glycanique chez les mammifères est catalysée par une α -1,6-fucosyltransférase qui permet l'ajout d'un résidu fucose lié en α -1,6 (Breton et al., 1998a; Voynow et al., 1991; Wilson et al., 1976). De même, une fucosylation des GlcNAc distales peut avoir lieu en α -1,3.

1.6 Localisation subcellulaire des glycosyltransférases golgiennes

L'action séquentielle des glycosyltransférases et glycosidases golgiennes peut être expliquée d'une part par leur spécificité d'accepteur et de substrat, mais aussi par leur distribution golgienne qui est organisée telle une chaîne de montage. En effet, ces enzymes sont compartimentées dans les différentes saccules golgiennes créant ainsi un gradient partant de la saccule *cis*, où se situent les étapes les plus précoces de la maturation, jusqu'au TGN (Trans Golgi Network) qui contient les *N*-glycannes les plus matures (**figure 31** ; Schoberer et al., 2013; Wildt and Gerngross, 2005). Ainsi, les études de localisations subcellulaires, menées principalement via des techniques de surexpression de gènes en fusion avec des étiquettes, suggèrent que les enzymes intervenant en amont de la voie de biosynthèse comme l' α -mannosidase I golgienne, la GnT I, l' α -mannosidase II, la GnT II, les FucT, ainsi que la XylT chez les plantes et les GnT III à V chez les mammifères, seraient plutôt localisées dans les citernes *cis* et médian de l'appareil de Golgi (Pagny et al., 2003; Schoberer and Strasser, 2011; Strasser, 2016). Plus tardivement vont ensuite intervenir les sialyltransférases permettant la sialylation des *N*-glycannes de mammifère ou bien les β -1,3-galactosyltransférases et les α -1,3/ α -1,4-fucosyltransférases responsables de la formation de l'épitope Lewis^a chez les

plantes. Ces dernières enzymes semblent plutôt localisées au niveau des citernes *trans*-golgiennes, voire dans le TGN (Fitchette et al., 1999; Schaub et al., 2008; Schoberer and Strasser, 2011; Wildt and Gerngross, 2005). Des maturations post-golgiennes peuvent ensuite se dérouler dans la vacuole des plantes, ou encore au niveau de la membrane plasmique (Castilho et al., 2014; Liebming et al., 2011; Strasser et al., 2007). Ainsi, il a été observé chez *A. thaliana* la présence de β -*N*-acetylhexosaminidases catalysant le clivage des GlcNAc terminales, aussi bien dans la vacuole qu'au niveau de la membrane plasmique (Liebming et al., 2011; Strasser et al., 2007).

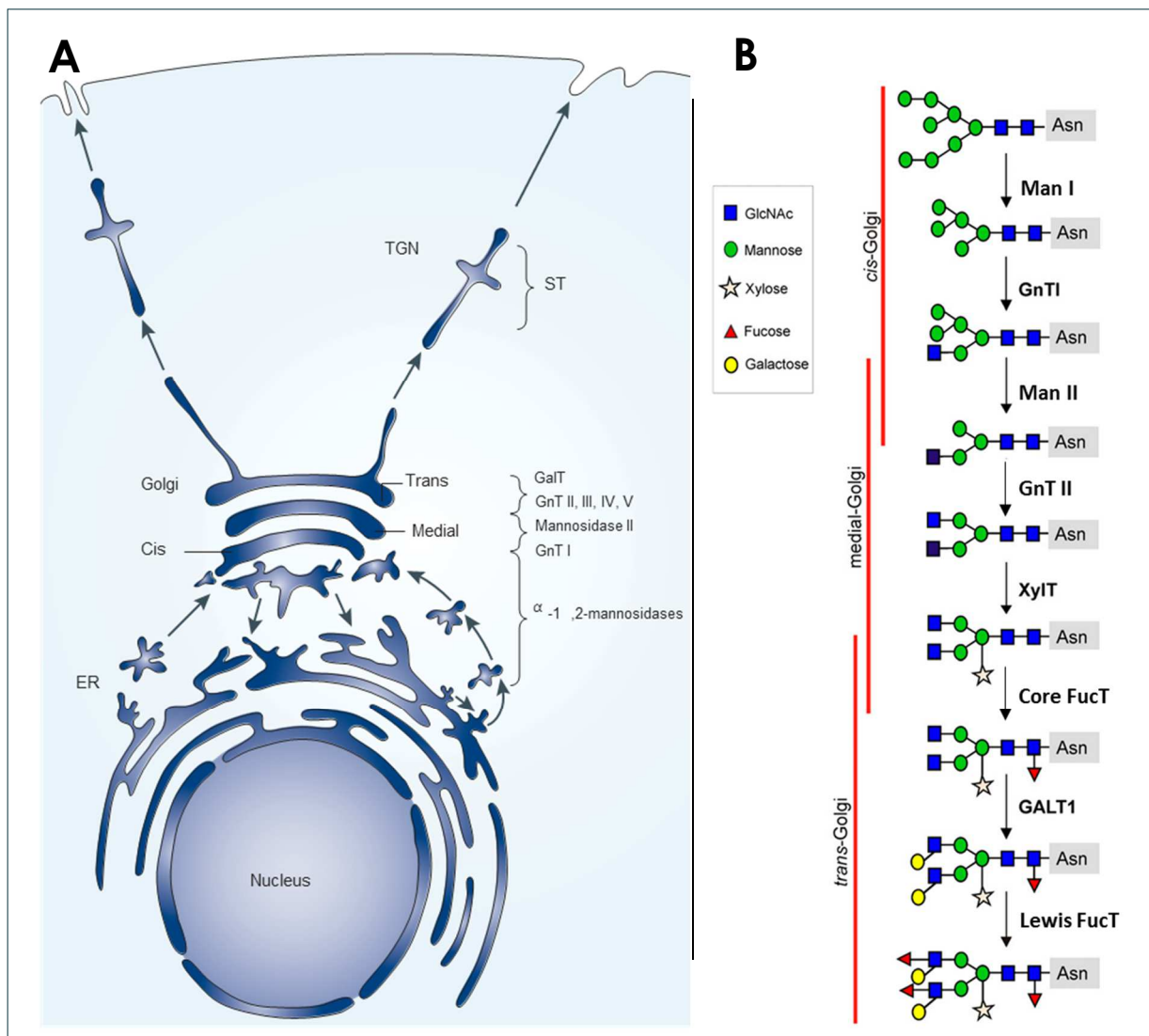


Figure 31 : Localisation subcellulaire des glycosyltransférases et glycosidases golgiennes impliquées dans la maturation des *N*-glycannes complexes d'après Wildt and Gerngross, 2005 ; Schoberer et al., 2013.

A = Répartition des enzymes golgiennes au sein du système sécrétoire de mammifère.

B = Répartition des enzymes golgiennes dans les différents compartiments de l'appareil de Golgi des plantes.

Les mécanismes d'adressage et de rétention de ces différentes glycosyltransférases dans les différents compartiments golgiens, et ce malgré l'important trafic vésiculaire bidirectionnel, restent encore méconnus. Cependant, certaines pistes sont investiguées, notamment celles liées à l'implication éventuelle de la région CTS de ces protéines transmembranaires. En effet, il a été montré chez les plantes qu'une augmentation de la longueur du domaine transmembranaire (TMD) de l' α -mannosidase I permettait de relocaliser l'enzyme dans la saccule *trans* de l'appareil de Golgi (Saint-Jore-Dupas et al., 2006). De plus, il existe chez les champignons et les mammifères des différences en longueur et en composition des TMD de glycosyltransférases qui dépendent du compartiment dans lequel elles sont localisées (Sharpe et al., 2010). La composition du TMD ne suffit pas à elle seule à expliquer la compartimentation des enzymes de *N*-glycosylation. En effet, la queue cytosolique influence aussi le positionnement de ces protéines le long de la voie de sécrétion, et une enzyme tronquée sur cette portion verra sa localisation affectée (Saint-Jore-Dupas et al., 2006). De plus, la composition lipidique de la membrane semble également influencer l'organisation des enzymes de *N*-glycosylation résidentes de l'appareil de Golgi. En effet, il a été observé, chez les cellules de mammifère, qu'un changement de la composition membranaire golgienne en phosphatidylinositol-4-phosphate induit une mauvaise localisation de la GnT I ainsi que de l' α -mannosidase II (Cheong et al., 2010). Ces résultats montrent qu'en plus de la structure des enzymes, la régulation spatiale des phospholipides semble être un facteur clé dans l'organisation golgienne. Ainsi, la concentration localisée des enzymes de *N*-glycosylation semble impliquer de multiples mécanismes qui nécessitent encore d'être complètement élucidés.

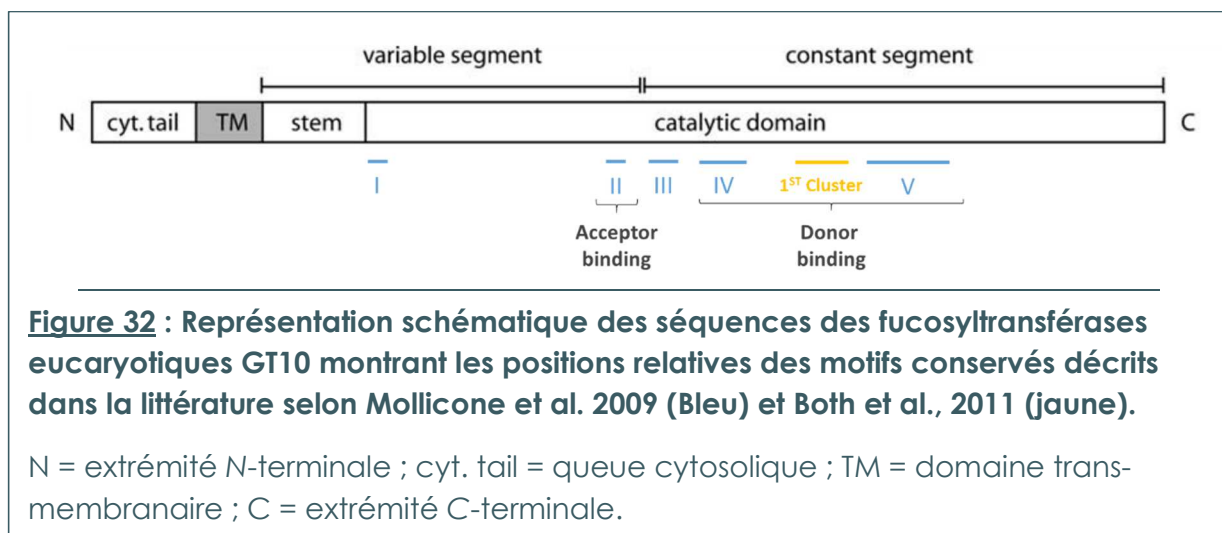
1.7 Fucosylation des *N*-glycannes

1.7.1 Fucosyltransférases

La fucosylation est l'une des principales étapes de maturation à l'origine de la diversité des *N*-glycannes complexes. Elle consiste en l'ajout d'un résidu L-fucopyranose (6-deoxy-L-galactose), à partir de GDP-fucose, sur une GlcNAc du *N*-glycanne. Cette modification est réalisée grâce à l'action de fucosyltransférases. Deux types de fucosylation ont été observées jusqu'à aujourd'hui sur les *N*-glycannes complexes. L'ajout du résidu fucose peut être effectué sur les GlcNAc terminales en α -1,3 chez les mammifères (de Vries et al., 2001) et chez certains insectes (Kubelka et al., 1993; Kurz et al., 2016), ou en α -1,4 chez les plantes (Lerouge et al., 1998). Cette fucosylation sur les antennes du *N*-glycanne est appelée « Lewis type », avec

l'appellation « Lewis A » pour un fucose lié en α -1,4, et « Lewis X » pour un fucose lié en α -1,3. Un second type de fucosylation peut être observé au niveau des GlcNAc proximales du core glycanique, appelé de fait « core type ». Ce fucose est alors lié en α -1,3 chez les plantes (Johnson and Chrispeels, 1987), en α -1,6 chez les animaux (Miyoshi et al., 1999), et en α -1,3 et α -1,6 chez certains invertébrés, présentant ainsi une bifucosylation de leurs *N*-glycannes (Staudacher and März, 1998). La fixation du fucose en α -1,6 est catalysée par une α -1,6-fucosyltransférase appartenant à la famille des GT23. L' α -1,3-fucosyltransférase, responsable de la fucosylation du core en α -1,3, et les α -1,3/4-fucosyltransférases fixant le fucose sur les branches terminales du *N*-glycanne appartiennent à la même famille des glycosyltransférases GT10. Ces dénominations « GT » désignent des ensembles de glycosyltransférases regroupées sous forme de familles basées sur des critères de structures et de domaines fonctionnels communs. Les différentes familles existantes peuvent être retrouvées sur la base de données CAZy (<http://www.cazy.org>) qui regroupe et classe l'ensemble des enzymes relatives à la maturation de polysaccharides.

Structurellement, les GT10 eucaryotiques sont des protéines transmembranaires de type II avec une topologie typique des glycosyltransférases golgiennes. Les comparaisons de séquences protéiques des enzymes appartenant à cette famille ont montré la présence de cinq motifs conservés (**figure 32** ; Both et al., 2011; Mollicone et al., 2009).



En effet, il a par exemple été montré chez la FucT III humaine, que le motif II est impliqué dans la spécificité de l'enzyme (Dupuy et al., 1999). Ainsi, lorsque le résidu Trp111 de sa séquence protéique est remplacé par une Arg, l'enzyme va catalyser la formation d'un épitope Lewis X au lieu d'un épitope Lewis A. Ce changement d'acide aminé implique ainsi non seulement une modification de la spécificité vis-à-vis du substrat accepteur (nécessite un

β -1,4-Gal au lieu d'un β -1,3-Gal lié à la GlcNAc), mais aussi un changement de la liaison du fucose ajouté (α -1,3 au lieu de α -1,4). De plus, la spécificité de substrat accepteur pour un même type de fucosyltransférase peut varier en fonction des organismes. En effet, il a été observé pour deux α -1,3-fucosyltransférases du core glycanne, la FucT-1 de *C. elegans* et la FucT-A d'*A. thaliana*, que la première accepte les *N*-glycannes paucimannosidiques, alors que la seconde nécessite au moins un résidu GlcNAc terminal (Paschinger et al., 2004; Wilson et al., 2001). Le motif α -1,3-FucT retrouvé dans les α -1,3/4-fucosyltransférases, correspondant au motif conservé V, possède différents degrés de conservation entre les espèces. Il est constitué de 19 acides aminés, dont 10 sont conservés chez tous les membres de la famille GT10 (Jost, 2004; Oriol et al., 1999). Certains de ces résidus sont impliqués directement dans la liaison avec le substrat donneur en formant des liaisons hydrogène avec le GDP-Fuc. C'est le cas notamment des résidus Tyr246, Glu249 et Lys250, observés chez la FucT d'*H. pylori* cristallisée (Sun et al., 2007). Bien que n'étant pas un des motifs conservés chez les GT10, l'extrémité C-terminale des α -1,3-FucT du core est très importante pour l'activité enzymatique. En effet, la délétion des 113 derniers résidus de la FucT-A, ou des 10 derniers de la FucT-1, cause une diminution supérieure à 99% de leur activité (Both et al., 2011). Structuellement, les α -1,3-fucosyltransférases eucaryotiques présentent un repliement GT-B, domaine protéique composé de deux sous-domaines Rossmann impliqués dans la liaison aux nucléotides et qui ne présente aucun signe évident de liaison à un ion métallique (**figure 33** ; (Breton, 2005).

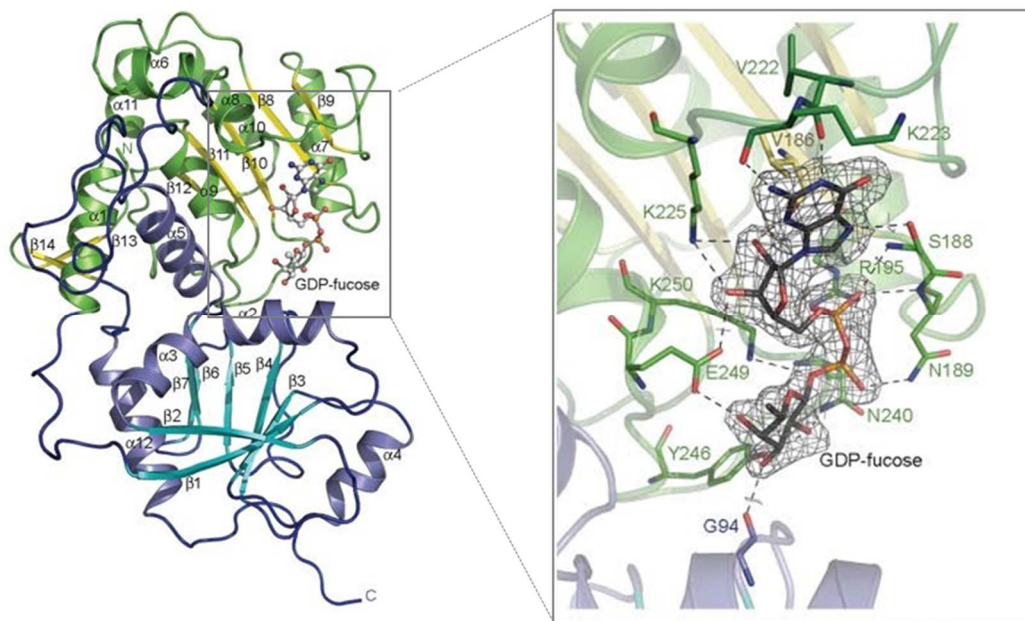


Figure 33 : Cristallographie de l'interaction entre l' α -1,3-FucT d'*H. pylori* et son substrat donneur, le GDP-fucose, d'après Sun et al., 2007.

Le complexe est représenté sous forme de diagramme en ruban, avec les feuillets β des domaines Rossmann représentés en cyan pour celui de l'extrémité N-terminale, et en jaune pour l'extrémité C-terminale. Les liaisons hydrogènes entre le GDP-fucose (représenté en boule et bâtons) et la protéine sont illustrées en lignes pointillées.

Malgré cela, il a été montré que l'ajout d'ion divalent augmente l'activité de l'enzyme, et accroît l'affinité pour le récepteur oligosaccharidique (Palma et al., 2004). Différents cofacteurs ont ainsi été identifiés comme le Mn(II), Ca(II), Ni(II) et le Co(II), ainsi que des inhibiteurs comme le Cu(II) et le Zn(II) (Both et al., 2011; Léonard et al., 2005; Murray et al., 1997). L'activité des GT10 est aussi impactée par la présence de glycosylations sur sa partie protéique. En effet, comme beaucoup de protéines golgiennes, les fucosyltransférases eucaryotiques sont *N*-glycosylées, pouvant posséder de 2 à 5 sites de *N*-glycosylation (Christensen et al., 2000; de Vries et al., 2001). Bien que la présence de *N*-glycannes ne soit pas essentielle, elle possède un grand impact sur l'activité des GT10. En effet, une étude menée sur la FucT-A, α -1,3-FucT possédant trois sites potentiels de *N*-glycosylation, a montré que la mutation d'un seul de ses sites de *N*-glycosylation pouvait causer jusqu'à 80% de réduction d'activité. Cependant, le triple mutant complètement déglycosylé conserve tout de même une activité résiduelle de l'ordre de 5% (Both et al., 2011). Ces fucosyltransférases impliquées dans la maturation des *N*-glycannes complexes font ainsi l'objet de différentes études, motivées en partie par le rôle de la fucosylation dans différents processus biologiques.

1.7.2 Implication biologique

Les *N*-glycannes complexes ont un rôle moins vital pour la cellule que les *N*-glycannes oligomannosidiques, cependant ils prennent toute leur importance lorsque l'on s'intéresse aux organismes multicellulaires. En effet, contrairement aux *N*-glycannes oligomannosidiques qui ont un rôle cellulaire crucial dans l'acquisition de la conformation native des glycoprotéines, les *N*-glycannes complexes sont surtout impliqués dans les interactions protéine/protéine, et donc cellule/cellule. Ainsi, bien que les lignées cellulaires mammifères dépourvues de glycannes fucosylés soient viables et ne présentent pas de perturbation de leur division (Malphettes et al., 2010), un défaut de fucosylation dans un organisme peut causer, ou être le résultat, de graves dysfonctionnements. En effet, chez les mammifères, une augmentation de la fucosylation, aussi bien au niveau distal que du core glycanique, a été observée dans différents tissus cancéreux : ovarien, hépatique, prostatique, ou encore colo-rectal (Holst et al., 2016; Miyoshi et al., 1999, 2007; Vermassen et al., 2015). La fucosylation des *N*-glycannes est ainsi devenue un biomarqueur pour certains cancers. Par exemple, l' α -foetoprotéine fucosylée au niveau du core de ses *N*-glycannes est reconnue par la FDA (Food and Drug Administration) comme étant un marqueur très spécifique du carcinome hépatocellulaire (Miyoshi et al., 2007). Cependant, le cancer n'est pas le seul processus dans lequel la fucosylation du core est

impliquée. En effet, les souris FuT8^{-/-}, mutées sur la fucosyltransférase permettant l'ajout de fucose en α -1,6 chez les mammifères, possèdent un taux de mortalité de 70% trois jours après la naissance, les survivantes présentant de sévères retards de croissance (Wang, 2005). Ces retards sont probablement causés par l'inhibition de différentes voies de signalisations impliquant des facteurs de croissance comme l'EGF (Epidermal Growth Factor) ou encore l'HGF (Hepatocyte Growth Factor). En effet, il a été démontré que la fucosylation au niveau du core des *N*-glycannes était nécessaire pour la liaison de l'EGF à son récepteur (Wang, 2005). C'est ainsi que les souris FuT8^{-/-} vont présenter un défaut de régénération du foie après hépatectomie, montrant ainsi l'implication de la fucosylation du core dans la prolifération cellulaire (Wang et al., 2015). Cependant, les facteurs de croissance ne sont pas les seules voies de signalisation affectées chez les souris FuT8^{-/-}. En effet, la fucosylation du core est aussi cruciale pour l'activité du récepteur TGF- β 1, qui est impliqué dans le contrôle de l'homéostasie de la matrice extracellulaire, provoquant ainsi chez la souris FuT8^{-/-} le développement d'emphysème (Wang, 2005). La fucosylation joue aussi un rôle important au niveau cérébral, où la FuT8 est très exprimée, provoquant ainsi chez la souris FuT8^{-/-} un phénotype schizophrène, impliquant très probablement un processus lié à la balance entre sérotonine et dopamine (Fukuda et al., 2011). Un des rôles de la fucosylation qui intéresse le plus les bio-industries, dont notamment le domaine pharmaceutique, est l'implication de la fucosylation du core dans le mécanisme ADCC (Antibody-Dependent Cellular Cytotoxicity). En effet, cette attaque lytique réalisée sur les cellules cibles par des anticorps est utilisée pour de multiples actions thérapeutiques, en particulier en oncologie (Clynes et al., 2000; Lewis et al., 1993). La présence de fucose au niveau du core des *N*-glycannes portés par les chaînes lourdes de l'anticorps va modifier la conformation du fragment Fc réduisant considérablement sa capacité à recruter les lymphocytes NK (Iida, 2006; Masuda et al., 2007). En effet, un anticorps dépourvu de fucose sera 50 fois plus efficace pour initier l'ADCC que son homologue fucosylé (Shinkawa et al., 2003). C'est pour cela que chez l'humain, 95% des IgGs circulants sont fucosylés au niveau du core (Pucic et al., 2011), permettant ainsi de prévenir l'élicitation de réponse ADCC destructive. Une dérégulation de ce système peut amener à une maladie auto-immune si l'ADCC est excessive, ou à l'apparition de cancer si l'ADCC est insuffisante (Lauc et al., 2014; Scanlan et al., 2008). L'activité biologique de la fucosylation en α -1,6 n'est pas la seule à être étudiée. En effet, les α -1,3-fucosyltransférases du core font aussi l'objet d'un intérêt particulier, puisqu'elles sont absentes chez les mammifères, ces derniers ne possèdent pas de *N*-glycannes α -1,3-fucosylés au niveau du core. De fait, ce résidu est reconnu comme épitope par le système immunitaire des mammifères, signalant ainsi la molécule comme un élément étranger. Cette

reconnaissance est notamment impliquée dans les phénomènes d'allergie au pollen, à la nourriture, ou plus généralement vis-à-vis des glycoprotéines produites chez les plantes ou les insectes, ce qui représente un enjeu primordial dans un contexte biotechnologique de production de protéines recombinantes à intérêt thérapeutique dans ces systèmes hétérologues (Bardor et al., 2003; Hemmer et al., 2001; Wilson et al., 1998).

1.8 *Phaeodactylum tricornutum*, un modèle d'intérêt pour l'étude de la N-glycosylation

Les microalgues comme *P. tricornutum* se positionnent comme une alternative très prometteuse pour la production de protéines recombinantes à intérêt thérapeutique. Ce système émergent de production offre des perspectives de réduction des coûts de production très intéressantes, couplées à une sécurité sanitaire accrue. Cependant, plusieurs questions doivent être résolues avant de pouvoir produire à l'échelle industrielle des biomédicaments obtenus à partir de microalgues. Parmi ces interrogations se trouve la question de la structure de *N*-glycannes présents chez les microalgues. En effet, actuellement, la majorité des protéines recombinantes à intérêt thérapeutique commercialisées sont glycosylées. La présence de *N*-glycannes étant nécessaire pour la fonctionnalité des glycoprotéines, et leur structure pouvant impacter leur demi-vie et générer des réactions allergiques, il est donc nécessaire de comprendre comment sont synthétisés les *N*-glycannes chez les microalgues. Avec sa facilité de culture, l'émergence de nouveaux outils de biologie moléculaire et sa capacité à produire des protéines recombinantes glycosylées, *P. tricornutum* se présente comme un sérieux modèle pour l'étude de la voie de *N*-glycosylation chez les microalgues. De plus, cette diatomée se situe dans une région phylogénétique où les voies de glycosylation des organismes environnants sont encore méconnues. Les motifs *N*-glycanniques étant espèce-spécifiques, la position phylogénétique de cette microalgue, à mi-chemin entre les organismes dont les voies de *N*-glycosylation sont les mieux connues, que sont les mammifères et les plantes, en fait une candidate particulièrement intéressante pour l'étude évolutive de la voie de *N*-glycosylation.

Les informations disponibles sur la voie de *N*-glycosylation de *P. tricornutum* au début de ma thèse ont été publiées dans une revue synthétisant l'ensemble des connaissances sur les voies de biosynthèse des *N*-glycannes chez les microalgues, et chez *P. tricornutum* en particulier. Cette revue est présentée ci-après.



Protein *N*-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals

Elodie Mathieu-Rivet^{1†}, Marie-Christine Kiefer-Meyer^{1†}, Gaëtan Vanier¹, Clément Ovide¹, Carole Burel¹, Patrice Lerouge¹ and Muriel Bardor^{1,2*}

¹ Laboratoire Glyco-MEV, Faculté des Sciences et Techniques, UPRES EA 4358, Normandie Université, IRIB, VASI, Mont-Saint-Aignan, France

² Institut Universitaire de France, Paris, France

Edited by:

Els J. M. Van Damme, Ghent University, Belgium

Reviewed by:

Richard Strasser, University of Natural Resources and Life Sciences, Austria

Eva L. Decker, University of Freiburg, Germany

*Correspondence:

Muriel Bardor, Laboratoire Glyco-MEV (Glycobiologie et Matrice Extracellulaire Végétale), UPRES EA 4358, UFR des Sciences et Techniques, Normandie Université, IRIB, VASI, Bâtiment Gadeau de Kerville, 1 rue Tesnière, 76821 Mont Saint Aignan Cedex, France
e-mail: muriel.bardor@univ-rouen.fr

[†] The two first authors equally contributed to the work.

Microalgae are currently used for the production of food compounds. Recently, few microalgae species have been investigated as potential biofactories for the production of biopharmaceuticals. Indeed in this context, microalgae are cheap, classified as Generally Recognized As Safe (GRAS) organisms and can be grown easily. However, problems remain to be solved before any industrial production of microalgae-made biopharmaceuticals. Among them, post-translational modifications of the proteins need to be considered. Especially, *N*-glycosylation acquired by the secreted recombinant proteins is of major concern since most of the biopharmaceuticals are *N*-glycosylated and it is well recognized that glycosylation represent one of their critical quality attribute. Therefore, the evaluation of microalgae as alternative cell factory for biopharmaceutical productions thus requires to investigate their *N*-glycosylation capability in order to determine to what extent it differs from their human counterpart and to determine appropriate strategies for remodeling the microalgae glycosylation into human-compatible oligosaccharides. Here, we review the secreted recombinant proteins which have been successfully produced in microalgae. We also report on recent bioinformatics and biochemical data concerning the structure of glycans *N*-linked to proteins from various microalgae phyla and comment the consequences on the glycan engineering strategies that may be necessary to render those microalgae-made biopharmaceuticals compatible with human therapy.

Keywords: microalgae, biopharmaceuticals, glycosylation pathway, glycan, *Chlamydomonas reinhardtii*, *Phaeodactylum tricornutum*, endoplasmic reticulum, Golgi apparatus

INTRODUCTION

Nowadays, biopharmaceuticals on the market consists of 200 products yielding overall global revenue greater than US\$100 billion (Walsh, 2010). Over the past 5 years, 140 biopharmaceuticals were approved in the European Union (EU) and the United States (US) markets (Walsh, 2010). These biopharmaceuticals are produced in various systems ranging from bacteria to mammalian cell cultures (Wong, 2005; Demain and Vaishnav, 2009; Huang et al., 2012). Among those, the Chinese Hamster Ovary (CHO) cells are currently the predominant industrial cell lines used for producing those drugs (Hossler et al., 2009), covering about 50% of the market (Demain and Vaishnav, 2009). However, the constant increasing needs for large amount of such therapeutic proteins, their high production cost in conventional expression systems and complicating factors related to potential virus contamination have driven scientists to explore new alternative production systems. In this context, various plant expression systems have emerged, including whole plants (*Nicotiana*, *Alfalfa*, *Maize* for examples) and *in vitro* culture systems such as plant cell suspensions (*Nicotiana tabacum*, *Lemna minor*, *Physcomitrella patens*) and hairy roots

(Stoger et al., 2005; Drake et al., 2009; Colgan et al., 2010; Decker and Reski, 2012; Parsons et al., 2012; Xu et al., 2012; Schillberg et al., 2013; Twyman et al., 2013; Buyel and Fischer, 2014). Several plant-made biopharmaceuticals have been successfully produced (De Muynck et al., 2010). This includes the Cerezyme's biosimilar (glucocerebrosidase) which has been produced in carrot cells (Shaaltiel et al., 2007) and was approved on May 2012 by the US Food and Drug Administration (Maxmen, 2012). Despite those successes, there is an increasing interest to use microalgae for biopharmaceutical production. Microalgae are unicellular photosynthetic organisms which encompass between 40,000 and probably several billion species (Cadoret and Bernard, 2008; Mata et al., 2010). As plants, microalgae are classified in Generally Recognized As Safe (GRAS) organisms. Moreover, they are cheap and easy to grow, making them potentially attractive cell factories for the large-scale production of recombinant proteins.

To date, microalgae have been mainly used for the production of food compounds or high-value added compounds like carotenoids (Spolaore et al., 2006; Sasso et al., 2012). In addition, as photosynthetic organisms, microalgae are very

efficient in converting sunlight into chemical energy, making them attractive for the production of carbohydrates, lipids, and hydrogen. Therefore, algal biomass represents a great potential for generating new sources of bioenergy such as biofuels (Beer et al., 2009; Lam and Lee, 2012; Merchant et al., 2012) and biomaterials (Hempel et al., 2011a). Several microalgae species have also been evaluated for their potential to express recombinant proteins. Among them, the model Chlorophyceae *Chlamydomonas reinhardtii* is currently the most investigated one for such a biotechnological application, due to the availability of genomic data and the existence of a powerful molecular toolkit including vectors allowing nuclear or chloroplastic transformation (Merchant et al., 2007; Harris, 2009). Additionally, the diatom *Phaeodactylum tricornerutum* is also considered as an emerging system for such an application (Hempel et al., 2011b; Hempel and Maier, 2012) as it can be genetically modified and grown quite easily. Other species are also good candidates for large-scale production of recombinant proteins, especially due to their ease to obtain algal biomass and their growth rate. For example, a marine green microalga from the class of the Chlorophyceae, *Dunaliella salina* (Geng et al., 2003), but also species from the genus *Chlorella* [i.e., *Chlorella vulgaris*, recently renamed *Coccomyxa* sp. C-169 *subellipsoidea* (Blanc et al., 2010), *Chlorella ellipsoidea*], which belongs to the Trebouxiophyceae, are promising bio-factories for large-scale production of high-value added proteins (Hawkins and Nakamura, 1999; Chen et al., 2001; Kim et al., 2002; Bai et al., 2013). However, molecular biology tools remain poorly developed for most of those microalgae even if genomic data became available (<http://genome.jgi.doe.gov/>; <http://www.phytozome.net/>; <http://www.ncbi.nlm.nih.gov/genome/>).

Whatever the considered species, many problems have still to be solved before any industrial production and commercialization of microalgae-made biopharmaceuticals. Among them, increasing the yield and secretion of recombinant protein represents a crucial issue to make these organisms competitive with traditionally used expression systems such as the CHO cell lines. Furthermore, post-translational modifications need to be considered. Especially, *N*-glycosylation acquired by the secreted recombinant proteins is crucial for biopharmaceuticals since more than one third of the approved ones are glycosylated (Gomord et al., 2010) and it represents a critical quality attribute for them (Lingg et al., 2012). Indeed, the presence and structures of the *N*-glycans are required for their biological activity, stability and half-life (Lingg et al., 2012). The evaluation of microalgae as alternative cell factory for biopharmaceutical production thus requires investigating their *N*-glycosylation capability in order to determine to what extent their *N*-glycans differ from their human counterpart. In this paper, we report on bio-informatic and biochemical data concerning the structures of glycan *N*-linked to endogenous proteins from various microalgae phyla. Based on these recent findings, strategies for the engineering of the glycosylation pathways in these new expression systems are proposed to obtain microalgae-made biopharmaceuticals that would carry oligosaccharides compatibles with human therapies.

MICROALGAE AS ALTERNATIVE SYSTEMS FOR PRODUCTION OF RECOMBINANT PROTEINS

AVAILABLE TOOLS FOR NUCLEAR TRANSFORMATION IN MICROALGAE *Chlamydomonas reinhardtii*

Currently, most of the studies reporting the expression of recombinant proteins in microalgae have been performed in the green alga *Chlamydomonas reinhardtii*. The majority of such productions has been targeted to the chloroplast. This concerned biopharmaceuticals such as the single-chain antibody directed against the glycoprotein D of the herpes simplex virus (Mayfield et al., 2003), but also the heavy and light chains of the antibody 83K7C, derived from a human IgG1 (Tran et al., 2009), human erythropoietin, domains 10 and 14 of human fibronectin, interferon β 1, proinsulin, vascular endothelial growth factor (VEGF), high mobility group protein B1(HMGB1) (Rasala et al., 2010) and two different immunotoxin proteins (Tran et al., 2012). Recently, an attenuated form of the E7 oncoprotein of the human papillomavirus (HPV) has also been produced in the same organelle (Demurtas et al., 2013). Indeed, this strategy is the most promising one to get high protein yields since *C. reinhardtii* possesses a large single chloroplast, representing about 40% of the total cell volume, and the proteins expressed in this organelle have been shown to represent 2–20% of total soluble proteins (Rasala and Mayfield, 2011). Moreover, three proteins have been shown to be expressed at levels which are sufficient for commercial production (Rasala et al., 2010). In contrast, secreted proteins produced from nuclear transformation in *C. reinhardtii* generally failed to accumulate to an equivalent level as the one observed in chloroplasts (Fuhrmann et al., 1999; Schroda et al., 2000; Specht et al., 2010). The low proteolysis in chloroplast could explain the high yields of recombinant proteins reached in this organelle (Surzycki et al., 2009; Potvin and Zhang, 2010). However, as far as glycosylated proteins such as biopharmaceuticals are concerned, chloroplast lacks the enzymatic machinery required for *N*-glycosylation. Therefore, for biopharmaceutical production, both nuclear expression and protein transport through the secretory pathway of microalgae are required.

Despite recent progress, nuclear expression remains challenging in *C. reinhardtii* (Specht et al., 2010). Indeed, low levels of expression usually observed could result from transgene silencing (Cerutti et al., 1997; Shaver et al., 2010; Rasala et al., 2012). Its GC-rich genome (Merchant et al., 2007) has been also thought to be a hindrance to the expression of foreign genes since it introduces a codon usage bias. Several strategies have been developed during the last decade to circumvent these problems. Among them, the necessity to optimize the coding sequence of the gene of interest appears to be essential for improving the nuclear expression of the foreign protein (Fuhrmann et al., 1999). With regards to promoter sequences, a few assays have been performed to express reporter genes under the control of the plant constitutive CaMV35S promoter (Ruecker et al., 2008; Díaz-Santos et al., 2013) or of other viral sequences (Ruecker et al., 2008). Recently, a high light-inducible promoter from *Dunaliella* has also been investigated in *C. reinhardtii*. This promoter has been shown to drive efficiently the expression of the luciferase reporter gene but it has not been used for expression of any biopharmaceutical yet

(Park et al., 2013). Currently, the major part of nuclear transgenes are expressed through a hybrid promoter, resulting from the fusion of the photosystem I complex (PSAD) (Fischer and Rochaix, 2001) or the Ribulose Bisphosphate Carboxylase Small Subunit (RBCS2) (Kindle, 1998) promoters with the HSP70A promoter which allows increasing the expression of the transgene (Schroda et al., 2000). In addition, several regulatory sequences have been included in the transgene coding sequence to enhance its expression. Thus, the insertion of the first intron of RBCS2 (Lumbreras et al., 1998), but also of the second and third introns (Eichler-Stahlberg et al., 2009) are required to increase the efficacy of the HSP70A/RBCS2 promoter. Compared to the CaMV35S promoter, these chimeric promoters remain actually the most efficient to drive the constitutive expression of nuclear transgenes in *C. reinhardtii* (Ruecker et al., 2008; Eichler-Stahlberg et al., 2009; Neupert et al., 2009; Rasala et al., 2012; Kumar et al., 2013). However, with a yield of secreted EPO estimated to 0.1 mg per liter of culture medium (Eichler-Stahlberg et al., 2009), the expression level of EPO under the control of the HSP70A/RBCS2A promoter remains too low for large scale production.

Rasala et al. (2012) developed a new vector in which the gene of interest is fused to the *BLE* gene of selection (which confers the resistance to bleomycin) *via* the nucleotide sequence encoding the foot-and-mouth-disease-virus 2A self-cleavage peptide (FMDV 2A) under the control of the HSP70A/RBCS2 promoter. The resulting translated product is processed into two proteins, with the 2A peptide fused to the C-terminal end of the first protein as already described for other expressions using a 2A self-cleavage peptide (Ho et al., 2013). This tandem expression allows the selection of transformants exhibiting a higher transgene expression. Finally, Neupert et al. (2009) used UV mutagenesis to generate new *C. reinhardtii* strains presenting the advantage to increase the expression of nuclear transgenes. Indeed, using one of these strains (UVM4) transformed with the specific vector pcCagLUC allowing the expression of the luciferase protein fused to the predicted signal peptide of the extracellular carbonic anhydrase 1 (CAH1), Lauersen and coworkers demonstrated that the amount of secreted recombinant proteins could be significantly improved, reaching up to 10 mg per liter of culture (Lauersen et al., 2013a). However, despite these significant achievements, efforts are still necessary to make *C. reinhardtii* competitive with CHO cell lines, for which yield between 5 and 10 g/L of recombinant protein is currently obtained (Demain and Vaishnav, 2009).

Other microalgae

Recently, other microalgae species have also been shown to be of interest as new expression systems. For example, two recent studies reported the expression of a monoclonal human IgG antibody against the Hepatitis B and its respective antigen in the diatom *P. tricornutum*. This antibody was either secreted in the culture medium or expressed in fusion to a DDEL (instead of KDEL which is used as a standard ER retention signal in other eukaryotes) sequence allowing its retention within the endoplasmic reticulum (ER) of *Phaeodactylum tricornutum* (Hempel et al., 2011b; Hempel and Maier, 2012). In contrast to *C. reinhardtii*, the codon usage in *P. tricornutum* is much closer to that of

human (Heitzer et al., 2007), which could give advantage to this diatom for the production of biopharmaceuticals. In *P. tricornutum*, nuclear expression of transgenes is usually mediated through the promoter of the *FCCA* gene which encodes the fucoxanthin chlorophyll a/c binding protein (Apt et al., 1996; Zaslavskaja et al., 2000) or through the nitrate reductase promoter which is induced when ammonium is replaced by nitrates as nitrogen source in the culture medium (Poulsen and Kröger, 2005; Gonzalez et al., 2011; Hempel et al., 2011b; Hempel and Maier, 2012; Stork et al., 2012). This specific nitrate reductase inducible promoter was used to express the fully-assembled human IgG antibody against Hepatitis B leading to 1.5–2.5 mg of recombinant antibody per liter of culture medium depending of the different clones (Hempel and Maier, 2012) and 21 mg per gram of algal dry weight for the ER-retained form of the antibody (Hempel et al., 2011b).

Expression of commercially relevant proteins in other species of microalgae remains less reported, since tools for nuclear transformation are still missing. As for *C. reinhardtii*, the CaMV35S promoter has been shown to be efficient in several other green microalgae, including *C. ellipsoidea* (Jarvis and Brown, 1991; Chen et al., 2001; Kim et al., 2002), *C. vulgaris* (Hawkins and Nakamura, 1999), *Haematococcus pluvialis* (Kathiresan et al., 2009) and *D. salina* (Geng et al., 2003; Feng et al., 2009; Chai et al., 2013) but in most of the cases, the expression levels achieved remained low. Some studies reported the use of other foreign promoters to drive the nuclear expression of heterologous proteins in *Chlorella* sp., such as the *C. reinhardtii* RBCS2 promoter which has been used successfully in *Chlorella ellipsoidea* for transient expression of recombinant protein or resistance expression (Hawkins and Nakamura, 1999; Kim et al., 2002). In contrast, the ubiquitin maize promoter seems to be promising in *Chlorella* sp., since it allows the stable expression of a rabbit gene encoding for an α -defensin (Bai et al., 2013). In some cases, the omega element of Tobacco Mosaic Virus (TMV) which is part of the 5'UTR has been used to enhance translation efficiency (Chen et al., 2001), leading to a yield of about 11 mg/L of recombinant protein (Bai et al., 2013). This ubiquitin- Ω promoter was also successfully used in *D. salina* to stably express a gene encoding the hepatitis B surface antigen (Geng et al., 2003).

EXPRESSION OF SECRETED THERAPEUTIC RECOMBINANT PROTEINS IN MICROALGAE

As mentioned previously, most of the recombinant proteins that have been produced in *C. reinhardtii* so far were expressed in the chloroplast because of the high level of protein accumulation reached in this organelle. Thus, among the 20 recombinant proteins of industrial interest expressed in *C. reinhardtii* (for a recent review, see Rasala and Mayfield, 2014), only 3 of them have been expressed successfully through the nuclear genome: a xylanase (Rasala et al., 2012), an ice binding-protein (Lauersen et al., 2013b) and the secreted EPO (Eichler-Stahlberg et al., 2009). As the xylanase and ice binding-protein are not therapeutical proteins, we will not discuss them further in this review. In addition, by using the signal peptide of the *C. reinhardtii* gene *ARS2* encoding a periplasmic arylsulfatase, the EPO has been produced in the culture medium of *C. reinhardtii* (Eichler-Stahlberg et al., 2009; **Table 1**). Furthermore, as previously indicated, Hempel and

Table 1 | Post-translational modifications of biopharmaceuticals expressed in the secretory system of microalgae.

Expressed protein	Promoter	N-glycosylation site ^a	Glycosylation ^b	References
IN C. REINHARDTII				
Human EPO	Hsp70A/RbcS2	Yes	Yes	Eichler-Stahlberg et al., 2009
IN P. TRICORNUTUM				
Monoclonal human IgG against the Hepatitis B surface antigen	Nitrate reductase	Yes	Yes	Hempel et al., 2011b; Hempel and Maier, 2012
Hepatitis B surface antigen	Nitrate reductase	Yes	No	Hempel et al., 2011b
IN D. SALINA				
Hepatitis B surface antigen	Maize Ubiquitin + Ω TMV enhancer	Yes	No	Geng et al., 2003

^aPredicted N-glycosylation sites by bio-informatic analysis of the protein sequence.

^bExperimental evidence of the presence of N-glycans attached to the N-glycosylation site.

Maier (2012) demonstrated the capability of *P. tricornutum* to synthesize and secrete a full length functional human IgG antibody against the Hepatitis B virus surface antigen. This study clearly demonstrates that diatoms are able to produce and correctly assemble complex proteins without affecting their biological activity (Table 1).

Little information is available on the post-translational modifications acquired by these microalgae-made therapeutic proteins in the secretory pathway. For example, the recombinant EPO expressed in *C. reinhardtii* exhibited a molecular mass of about 33 kDa suggesting the presence of post-translational modifications such as the addition of glycans on the microalgae-made EPO. This is consistent with the fact that the EPO is known to possess 3 N-glycosylation and one O-glycosylation sites (Table 1) (Lingg et al., 2012). Moreover, affino blotting with concanavalin A, a lectin specific for oligomannoside structures (Fitchette et al., 2007), of the ER-resident recombinant IgG expressed in *P. tricornutum* also suggested that this microalgae-made antibody is glycosylated (Table 1) (Hempel et al., 2011b). However, no structural detailed analyses are reported regarding the glycans attached to those microalgae-made biopharmaceuticals. Since the N-glycosylation of biopharmaceuticals is critical for their half-life, stability and biological activity (for a recent review, see Lingg et al., 2012), it is therefore essential to characterize the N-glycosylation process of secreted protein in microalgae.

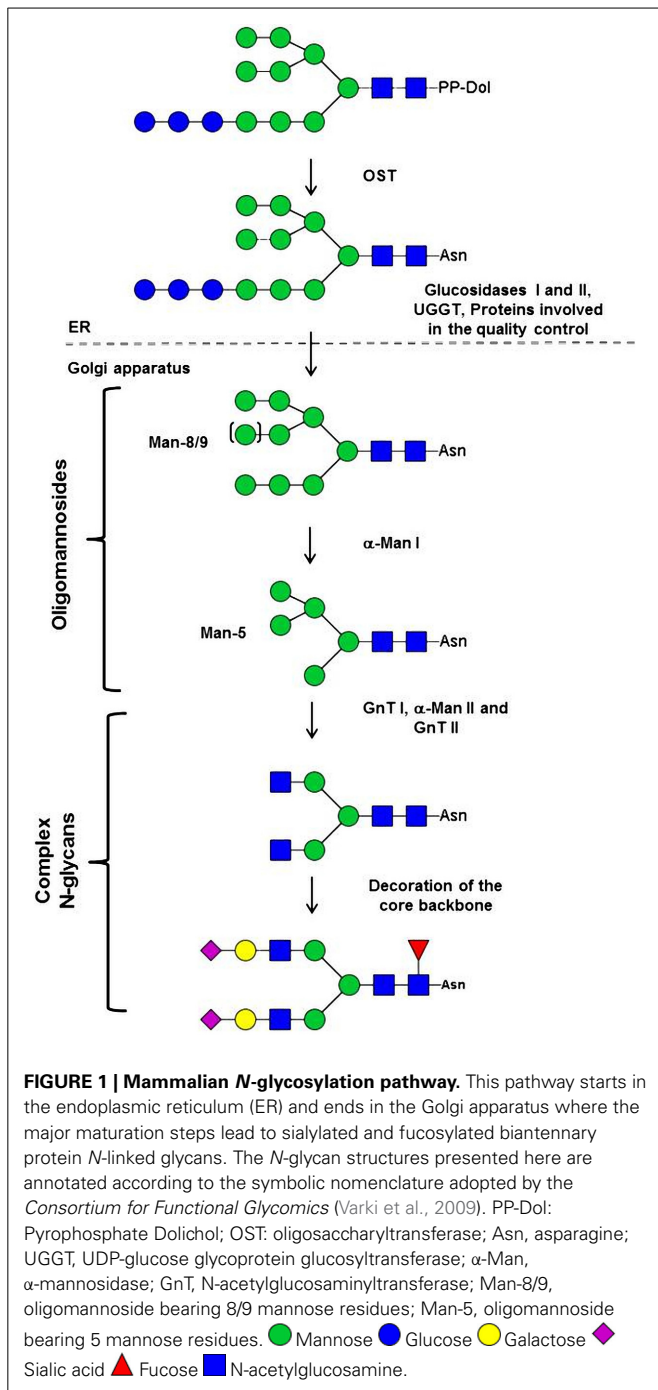
N-GLYCOSYLATION PATHWAYS IN MICROALGAE

GENERAL ASPECTS OF N-GLYCOSYLATION IN EUKARYOTES

N-glycosylation is a major co- and post-translational modification of proteins in eukaryotes occurring in both the ER and the Golgi apparatus (Figure 1). In this process, a lipid-linked oligosaccharide is first assembled by the stepwise addition of monosaccharides on a dolichol pyrophosphate on the cytosolic face and then in the lumen of the ER (Burda and Aebi, 1999). This precursor is then transferred by the oligosaccharyltransferase (OST) complex onto the asparagine residues of the consensus Asn-X-Ser/Thr sequences of a protein (Burda and Aebi, 1999). In 3.5% of the cases, other sequences like Asn-X-Cys, Asn-X-Val have been proven to be glycosylated on endogenous or recombinant protein produced both in mammals or plant cells (Gil et al., 2009; Zielinska et al., 2010; Matsui et al., 2011). The precursor is deglycosylated by the α -glucosidases I

and II and then reglycosylated by an UDP-glucose: glycoprotein glucosyltransferase (UGGT) to ensure the proper folding of the nascent protein through its interaction with ER-resident chaperones, such as calnexin and calreticulin. These ER events are conserved in eukaryotes because they are crucial for efficient protein folding and oligomerization (Helenius and Aebi, 2001). In contrast, evolutionary adaptation of N-glycan processing in the Golgi apparatus has given rise to a variety of organism-specific complex structures (Varki, 2011). First, α -mannosidases degrade the oligosaccharide precursor into oligomannosides ranging from Man₉GlcNAc₂ to Man₅GlcNAc₂ (Man-9 to Man-5). N-acetylglucosaminyltransferase I (GnT I) then transfers a first N-acetylglucosaminyl (GlcNAc) residue on Man-5 and initiates the synthesis of a large variety of structurally different complex-type N-glycans. In this GnT I-dependent N-glycan maturation, the processing continues by the removal of two mannosyl residues and the transfer of a second terminal GlcNAc residue, thus resulting in the synthesis of a core GlcNAc₂Man₃GlcNAc₂ which are common to mammals and all land plants studied so far (Lerouge et al., 1998; Wilson et al., 2001; Gomord et al., 2010; Varki, 2011) (Figure 1). This core is then decorated by the action of a specific repertoire of glycosyltransferases that differ from one organism to another. As a consequence, mature proteins leaving the secretory pathway carry organism-specific complex N-glycans allowing the protein to acquire a set of glycan-mediated biological functions (Varki, 1993; Gagneux and Varki, 1999). For instance, in mammals, most secreted proteins in blood circulation carry biantennary N-glycans decorated with α (1,6)-fucose residues and terminal sialic acids that impact either the protein activity or protein half-life (Figure 1; Lingg et al., 2012). In contrast, plant N-glycans are mainly of biantennary complex type N-glycan carrying a core- β (1,2)-xylose; a core α (1,3)-fucose and eventually terminal Lewis a antennae (Lerouge et al., 1998; Wilson et al., 2001).

By comparison with data available in pluricellular eukaryotes, information concerning protein N-glycosylation in microalgae remains very limited. A few studies using lectin blot analysis or enzymatic sequencing suggested that proteins secreted by green microalgae carry mainly oligomannosides or complex N-glycans having a core xylose residue (Balshüsemann and Jaenicke, 1990; Grunow et al., 1993; Gödel et al., 2000). More recently, a cell wall glycoprotein from the red microalgae *Porphyridium* sp. was



found to carry Man-8 and Man-9 oligomannosides containing 6-O-methyl mannoses and substituted by one or two xylose residues (Levy-Ontman et al., 2011).

STRUCTURAL INVESTIGATION OF GLYCAN *N*-LINKED TO MICROALGAE PROTEINS OF *CHLAMYDOMONAS REINHARDTII* AND *PHAEODACTYLUM TRICORNUTUM*

Deeper insights into the structure of glycans *N*-linked to proteins secreted by *C. reinhardtii* (Mathieu-Rivet et al., 2013) and *P. tricornutum* (Baïet et al., 2011) have been recently reported and *N*-glycosylation pathways in these two microalgae dedicated

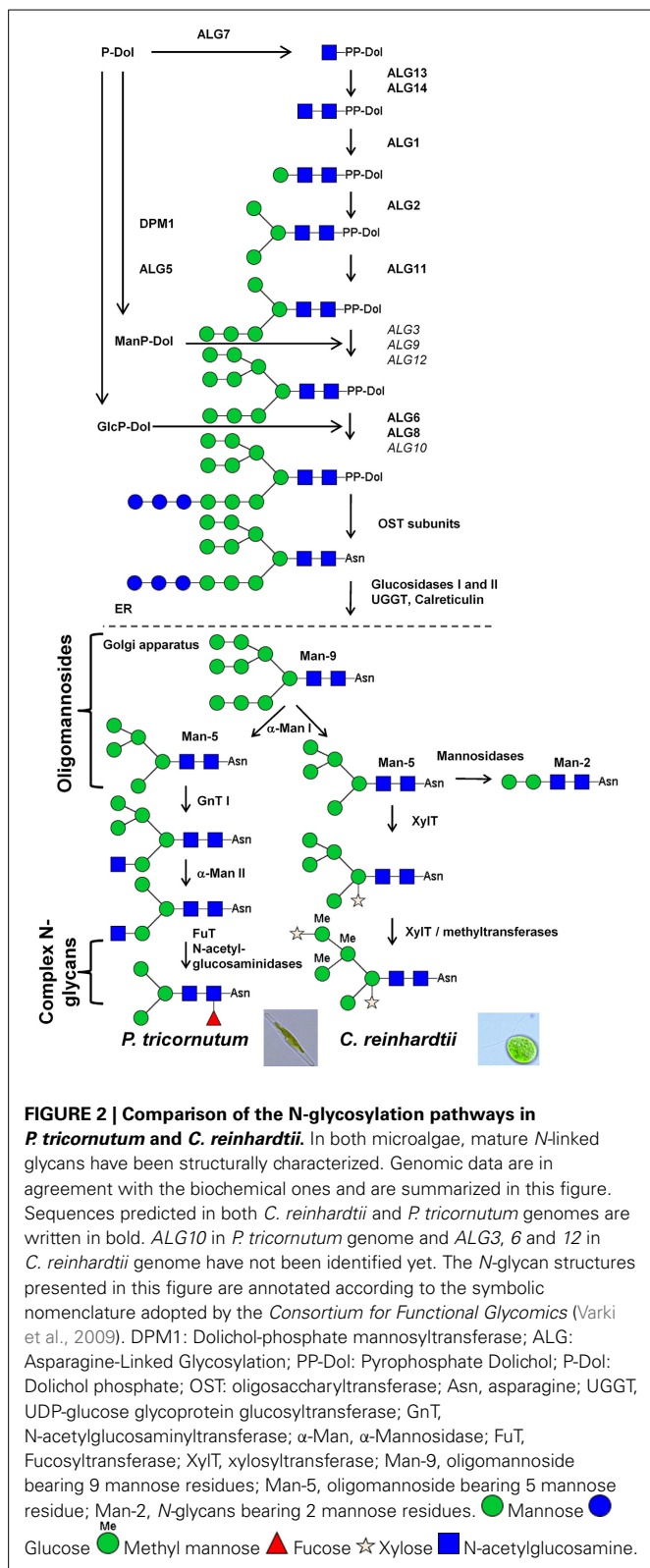
to biotechnology applications have been proposed. These studies combined data resulting from the search for genes encoding putative enzymes involved in the *N*-glycosylation pathway in the genomic databases and data resulting from biochemical analysis of glycan *N*-linked to secreted proteins.

Bio-informatic analysis

In *C. reinhardtii* and *P. tricornutum* genomes, most of the genes encoding enzymes involved in the biosynthesis of the dolichol pyrophosphate-linked oligosaccharide on the cytosolic face and in the lumen of the ER, named Asparagine-Linked Glycosylation (ALG) (Weerapana and Imperiali, 2006), are predicted (Figure 2). Although some ALG were not clearly identified in the genomes, large oligomannosides up to Man-9 were found in both *C. reinhardtii* (Mathieu-Rivet et al., 2013) and *P. tricornutum* (Baïet et al., 2011) suggesting that the synthesis of the oligosaccharide precursor is similar to the one described in other eukaryotes. In addition to ALG, genes encoding subunits of the oligosaccharyltransferase were also identified in both genomes (Figure 2). Glucosidases I and II, as well as ER-resident UGGT and chaperones such as calreticulin, are also predicted. These proteins are key elements of the quality control of proteins occurring in the ER and are crucial for acquisition of the proper folding of the nascent glycoprotein (Figure 2).

After the transfer of the glycoprotein into the Golgi apparatus, maturation of *N*-linked glycans starts with the trimming of Man-9/-8 by α -mannosidases I (α -Man I) that generate oligomannosides ranging from Man-8 to Man-5. Genes encoding α -Man I are predicted in *C. reinhardtii* and *P. tricornutum* genomes (Figure 2) suggesting that Golgi mediated trimming of oligomannosides also occurs in both microalgae. In higher eukaryotes including animals, insects and land plants, *N*-glycans are then processed by GnT I that transfers a first *N*-acetylglucosaminyl (GlcNAc) residue on the α (1,3)-mannose arm of Man-5 (Figure 1). Thus, GnT I is a key enzyme since it is the first glycosyltransferase occurring in the Golgi apparatus in the GnT I-dependent pathway giving rise to complex *N*-glycans that are required for normal morphogenesis in pluricellular organisms (Ioffe and Stanley, 1994; Metzler et al., 1994). A sequence encoding for a GnT I is predicted in *P. tricornutum* (Baïet et al., 2011) but not in *C. reinhardtii* (Mathieu-Rivet et al., 2013) suggesting that *N*-linked glycans from these two microalgae could be processed in the Golgi apparatus according to two different pathways, referred to as GnT I-dependent and GnT I-independent pathways (Zhu et al., 2004; Crispin et al., 2006; Grass et al., 2011).

In both GnT I-dependent and independent pathways, the next steps mainly consist in the transfer of monosaccharides on the *N*-glycan core to synthesize highly diverse complex *N*-glycans. Among residues that are added to the core, fucosyl residues either α (1,3)- or α (1,6)-linked to the proximal GlcNAc are commonly observed. In both microalgae, sequences encoding for putative α (1,3)-fucosyltransferases were identified in the genomes (Baïet et al., 2011; Mathieu-Rivet et al., 2013). Search for other genes encoding for putative *N*-glycan processing enzymes revealed sequences exhibiting homologies with α -Man II and GnT II in *P. tricornutum* but functional characterizations remain necessary to assess their enzymatic activities.



Biochemical analysis

A detailed glycomic analysis of *C. reinhardtii* proteins indicated that both secreted and membrane-bound proteins carry oligomannosides ranging from Man-2 to Man-5 and also complex

N-glycans containing 6-O-methyl mannoses and substituted by one or two xylose residues (Figure 2) (Mathieu-Rivet et al., 2013). Similar complex structures have been previously found in a cell wall glycoprotein isolated from the red microalgae *Porphyridium sp.* (Levy-Ontman et al., 2011). However, the location of the xylose residue on the core N-glycan differs in these two microalgae, since it is described to be linked to the chitobiose unit in *Porphyridium* rather than to C2 of the β -mannose as described in *C. reinhardtii* and as previously reported in land plants (Figure 2) (Lerouge et al., 1998). In agreement with bio-informatics data suggesting the absence of GnT I, these complex N-glycans are likely to result from the transfer of xylose residues onto Man-5 oligomannosides in a GnT I-independent manner. Then, methylation of mannose residues is thought to occur after xylosylation of complex N-glycans in *C. reinhardtii* as proposed in Figure 2. Alternatively, Man-5 can be trimmed by mannosidases down to Man-2 (Figure 2). Although O-methyl mannose residues was previously reported in some eukaryotes (Staudacher, 2012), complex N-glycans from *C. reinhardtii* highly differ from glycans described in animals and land plants. We postulate that xylosylation and methylation of oligomannosides in these microalgae may protect glycoproteins against deglycosylating enzymes such as endoglycosidases or peptide N-glycosidases.

In contrast to complex N-glycans from *C. reinhardtii*, glycans N-linked to proteins secreted by the diatom *P. tricornutum* can be processed through a GnT I-dependent pathway into partially fucosylated Man-3 (Figure 2) (Baïet et al., 2011). The GnT I gene predicted in the *P. tricornutum* genome was demonstrated to encode an active enzyme able to restore the maturation of N-linked glycans into complex-type N-glycans in the CHO *Lec1* mutant, defective in its endogenous GnT I (Baïet et al., 2011). The authors have proposed that this fucosylated N-linked glycan results from the addition of a terminal GlcNAc residue by GnT I on Man-5 followed by removal of two Man residues by an α -Man II and the transfer of a fucose residue by the predicted $\alpha(1,3)$ -FuT (Baïet et al., 2011). As illustrated in the proposed pathway depicted in Figure 2, the terminal GlcNAc introduced in the Golgi apparatus by the *P. tricornutum* GnT I is then likely removed in the secretory pathway by glucosaminidases as previously described in land plants and insect (Vitale and Chrispeels, 1984; Altmann et al., 1995). Efficient complementation of the CHO *Lec1* mutant is the main argument supporting that this diatom processes its N-glycans through a GnT I-dependent pathway. This result also indicates that the microalgae transferase is properly targeted to the Golgi apparatus when expressed in mammalian cells.

In these two studies performed on *P. tricornutum* and *C. reinhardtii* N-glycan processing and as described for land plants (Séveno et al., 2004; Zeleny et al., 2006), no sialic acid residues were identified (Baïet et al., 2011; Mathieu-Rivet et al., 2013). Sialic acids are terminal residues of O- and N-glycans that are specifically involved in many biological functions in mammals, such as the half-life of blood proteins and cell-cell adhesion processes (Figure 1) (Varki, 1993; Gagneux and Varki, 1999).

GLYCAN-REMODELING STRATEGIES IN MICROALGAE: PERSPECTIVES

Differences in specificity of Golgi transferases and glycosidases between eukaryotes give rise to glycosylation profiles that differ between mammals and other eukaryotic host cells used as cell factories. As a consequence, glycans *N*-linked to recombinant proteins produced in plants, yeast or even in animal cells differ from the original therapeutic proteins. This may result in either a decrease or absence of biological activity. Furthermore, unsuitable *N*-glycan structures introduced by the expression system can induce immune responses in humans and generate adverse reactions (Van Beers and Bardor, 2012), as reported for $\alpha(1,3)$ -Gal epitope or Neu5Gc on therapeutic drugs (Chung et al., 2008; Padler-Karavani et al., 2008). Similarly, plants synthesize *N*-glycans carrying immunogenic core-xylose and core- $\alpha(1,3)$ -fucose that may induce immune responses in human treated with plant-made biopharmaceuticals (Bardor et al., 2003). In consequence, whatever the expression system that is considered, strategies have been carried out for the *in vivo* remodeling of the protein *N*-glycans in order to obtain structures that meet pharmaceutical requirements. For instance, knock-out of endogenous genes involved in the transfer of core immunogenic epitopes have been carried out to engineer oligosaccharides synthesized by plant cell into human-compatible structures (Cox et al., 2006; Schähs et al., 2007; Strasser et al., 2008). Furthermore, knock-in methodologies based on the expression in the host cells of mammalian enzyme have been developed to *in vivo* introduce missing glycan sequences. In plants, these efforts resulted in the production of plant-derived therapeutic proteins carrying *N*-glycans similar to those found on human counterpart (Palacpac et al., 1999; Bakker et al., 2001, 2006; Misaki et al., 2006; Paccalet et al., 2007; Rouwendal et al., 2007; Castilho et al., 2008, 2010, 2013).

As reported in section Microalgae as Alternative Systems for Production of Recombinant Proteins, *C. reinhardtii* and *P. tricornutum* have been evaluated for their capacity to express therapeutic proteins. The recent characterization of glycans *N*-linked to their secreted proteins allows the design of strategies to engineer their *N*-glycan pathways for the production of human-compatible therapeutic proteins.

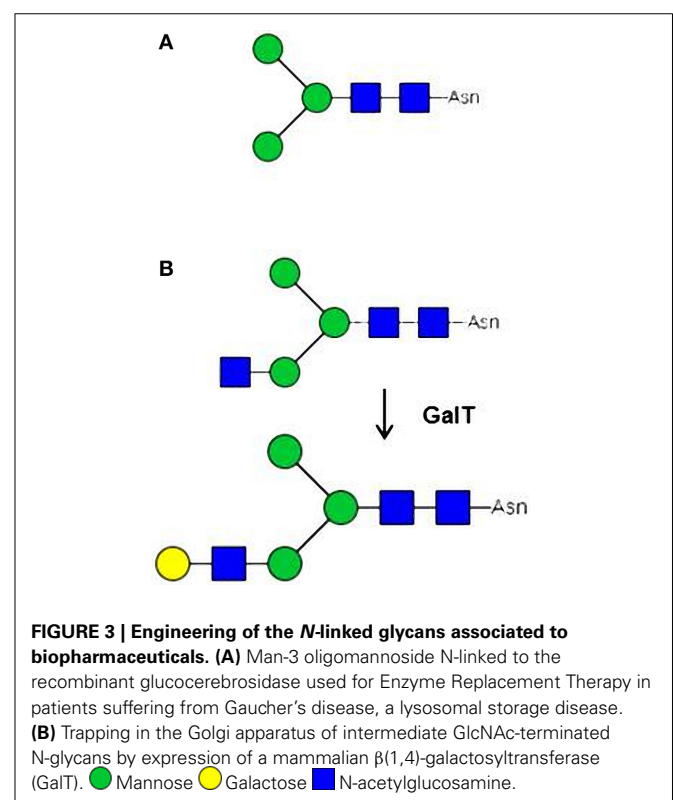
MICROALGAE FOR THE PRODUCTION OF LYSOSOMAL PROTEINS

Although complex *N*-glycans have been identified on proteins secreted by *C. reinhardtii* and *P. tricornutum*, most abundant oligosaccharides *N*-linked to proteins are oligomannosides ranging from Man-2 to Man-5 in *C. reinhardtii* (Mathieu-Rivet et al., 2013) and from Man-5 to Man-9 in *P. tricornutum* (Baïet et al., 2011) (Figure 2). These *N*-linked glycans are appropriate for the production of lysosomal therapeutics. For example, glucocerebrosidase is a glycoprotein drug administered intravenously into patients suffering from Gaucher's disease, a lysosomal storage disease. The effective targeting and internalization of this therapeutic drug into macrophages depend on terminal mannose residues of its *N*-glycans which are recognized by macrophage cell surface mannose receptors (Van Patten et al., 2007). For treatment, exogenous glucocerebrosidase is administered intravenously into patients. The preparation of the current

glucocerebrosidase expressed in CHO cells (Cerezyme®) requires *in vitro* post-purification and exoglycosidase digestions to expose the trimannose core (Man₃GlcNAc₂, Man-3) using a combination of at least three enzymes namely sialidase, galactosidase and *N*-acetylglucosaminidase (Figure 3A). These *in vitro* glyco-engineering steps increase considerably the production costs (Weinreb, 2008). Therefore, alternative expression systems that are capable of producing mannose terminated *N*-glycans have been developed as safe and cost effective production methods. These include cultured carrot cells (Shaaltiel et al., 2007) and *A. thaliana cgl* transgenic plants (He et al., 2012). However, the analysis of the *N*-glycan profiles of these glucocerebrosidase bio-similars revealed Man-5 with variable amount of Man-6 to Man-9 structures (for example low amount of Man 6 to Man-8 on glucocerebrosidase expressed in the Arabidopsis *cgl* mutant). The presence of such structures required careful biosafety studies to evaluate the possible binding to serum mannose binding lectin (MBL) and immunogenicity (Van Patten et al., 2007; Grabowski et al., 2014). The results described for *C. reinhardtii* (Mathieu-Rivet et al., 2013) suggest that this green microalgae could become in the future an appropriate platform for the production of a glucocerebrosidase bio-better which would naturally carry Man-2 to Man-5 glycan structures without any *in vitro* trimming of mannose residues (Figure 2).

MICROALGAE FOR THE PRODUCTION OF PROTEINS CARRYING HUMAN-COMPATIBLE COMPLEX-TYPE *N*-GLYCANS

Based on data reported in section *N*-glycosylation Pathways in Microalgae and Figure 2, the production in microalgae of



therapeutic proteins carrying human-like complex glycans will require intensive remodeling of their glycan pathways. Although no information is available on their immunogenicity, glycan epitopes identified on microalgae complex *N*-glycans will likely induce adverse effects in humans. Removal of these putative immunogenic glyco-epitopes, such as xylose and fucose residues, as well as methyl groups on mannose residues, will first require to identify gene encoding the corresponding glycosyltransferases and methyltransferases and then to carry out appropriate knock-out strategies such as RNA interference; artificial microRNA which have already been developed in microalgae (Zhao et al., 2009; Cerutti et al., 2011). If putative FuT genes have already been identified because of their high homologies with plant $\alpha(1,3)$ -FuT, the identification of sequences coding for xylosyltransferases (XylT) will be more tricky. Only the $\beta(1,2)$ -XylT family involved in complex *N*-glycan biosynthesis in land plants have been biochemically characterized so far (Strasser et al., 2000). To date, no specific domain for XylT activity has been clearly identified and so, search by homology for XylT in microalgae genomes failed to clearly identify putative candidates.

As previously reported in land plants, “missing” residues on glycan *N*-linked to therapeutic proteins can be introduced by co-expression with appropriate glycosyltransferases. Results regarding the complementation of CHO *Lec1* mutant by *P. tricornutum* GnT I (Baïet et al., 2011) showed that the targeting of Golgi glycosyltransferases is conserved between microalgae and higher eukaryotes. As a consequence, we can expect that the expression in microalgae of heterologous glycosyltransferases would allow the *in vivo* remodeling of their *N*-glycan pathways. This heterologous expression of glycosyltransferases should be particularly suitable for the remodeling of *N*-glycans in *P. tricornutum* because this diatom possesses a functional GnT I making its *N*-glycosylation pathway closer to the mammalian one (Figures 1, 2). The terminal GlcNAc introduced by GnT I constitute the starting point for the building of antennae in mammalian complex *N*-glycans (Figure 1). We postulated that in *P. tricornutum* this terminal GlcNAc is lost during the trafficking of the glycoprotein by action of *N*-acetylglucosaminidases (Figure 2). However, the expression of a mammalian $\beta(1,4)$ -galactosyltransferase in *P. tricornutum* would allow the transfer in the Golgi apparatus of a terminal Gal on the terminal GlcNAc, thus giving rise to a Gal $\beta(1,4)$ GlcNAc extension that will protect this *N*-acetylglucosamine residue introduced by GnT I from degradation by *N*-acetylglucosaminidases downstream in the secretory pathway (Figure 3B). In addition, this extension corresponds to glycosidic motif found on some therapeutic proteins such as IgG. This strategy based on the trapping of intermediate GlcNAc-terminated *N*-glycans has already been successfully carried out in land plants (Bakker et al., 2001; Huether et al., 2005; Vézina et al., 2009) and insect cells (Hollister et al., 1998).

SEARCH FOR MICROALGAE HAVING APPROPRIATE *N*-GLYCOSYLATION PATHWAYS

As *in vivo* engineering of the *N*-glycosylation pathways in *C. reinhardtii* and *P. tricornutum* may represent a tricky work, exploration for a microalga that exhibits a more appropriate protein glycosylation has to be considered. A first overview of protein

N-glycosylation in various microalgae species can be drawn on the basis of public genomic databases. In addition to *C. reinhardtii* and *P. tricornutum*, more than 15 genomes from other microalgae are now available (e.g., *Thalassiosira pseudonana*; *Cyanidioschyzon merolae*; *Ostreococcus tauri* and *Ostreococcus lucimarinus*; *Micromonas pusilla*; *Chlorella vulgaris*; *Coccomyxa subellipsoidea*; *Nannochloropsis gaditana*; *Monoraphidium neglectum*; Table 2) (Ambrust et al., 2004; Misumi et al., 2005; Palenik et al., 2007; Worden et al., 2009; Blanc et al., 2010, 2012; Radakovits et al., 2012; Vieler et al., 2012; Bogen et al., 2013). Searches for ALG genes and other sequences encoding proteins involved in the ER protein quality control indicated that enzymes of the ER machinery are predicted in all genomes suggesting that ER steps of the *N*-glycan pathway are conserved over the microalgae phyla as discussed recently in Levy-Ontman et al. (2014). With regard to Golgi events, Table 2 summarizes and reports the genes predicted to encode enzymes involved in the maturation of the *N*-glycans. As observed for *C. reinhardtii* and *P. tricornutum*, α -mannosidases (CAZy GH 47) and $\alpha(1,3)$ -fucosyltransferases (CAZy GT10) are predicted in most microalgae genomes. In contrast, genes encoding GnT I are not predicted in all genomes (Table 2; Baïet et al., 2011). For instance, as for *C. reinhardtii*, no GnT I was found in *Volvox* and *Ostreococcus* species whereas this key transferase is predicted in most other microalgae, including haptophytes and cryptophytes (Table 2; Baïet et al., 2011). As demonstrated in *C. reinhardtii* and *P. tricornutum*, this indicates that the Golgi maturation of protein *N*-linked glycans into complex oligosaccharides could either occur through a GnT I-dependent pathway or a GnT I-independent pathway in microalgae. In the context of microalgae-made biopharmaceutical production, this feature must be considered since the engineering of microalgae *N*-glycans into human compatible oligosaccharides would be facilitated if a GnT I-dependent machinery already exists in the strains selected as an expression system.

CONCLUSIONS

In conclusion, it can be considered that remodeling *N*-glycans as depicted in Figure 2 into human-compatible oligosaccharides would be a difficult challenge to attend. However, recent production in plants of therapeutic proteins carrying sialylated biantennary *N*-glycans demonstrates that such a glycan engineering can be achieved in a plant expression system (Castilho et al., 2013). It should also be noticed that only a few microalgae species have been investigated and so, data reported in this review should be considered as a starting point regarding protein glycosylation mechanisms occurring in these unicellular organisms and will serve as foundation for further glycobiology works. Efforts have now to be carried out to get more information concerning protein *N*-glycan structures and their pathways in different phyla to identify microalgae species that are more appropriate for glycan remodeling strategies into human-like structures. This will also necessitate the characterization of the specificity of Golgi glycosyltransferases, such as α -mannosidases and $\alpha(1,3)$ -fucosyltransferases predicted in the microalgae genomes, to evaluate the efficiency in microalgae of heterologous expression of glycosyltransferases as well as to identify genes encoding transferases (methyltransferases and xylosyltransferases) involved in

Table 2 | Genes predicted in microalgae genomes encoding proteins involved in *N*-glycan Golgi maturation such as α -Man I, mannosyl oligosaccharide alpha-1,2-mannosidase; GnT I, alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase; α (1,3)-FucT, core α (1,3)-fucosyltransferase.

Microalgae	α -Man I (K01230)	GnT I (K00726)	α (1,3)-FucT (K00753)	
Green algae	<i>Chlamydomonas reinhardtii</i>	XP_001700094.1	No	XP_001695259.1
	<i>Volvox carteri f. nagariensis</i>	XP_002957696.1	No	XP_002947231.1
	<i>Ostreococcus lucimarinus</i>	XP_001421581.1	No	XP_001420919.1 XP_001420471.1
	<i>Ostreococcus tauri</i>	XP_003083553.1	No	XP_003082568.1
	<i>Micromonas sp RCC299</i>	XP_002505356.1	XP_002507699.1	XP_002509370.1
		XP_002499658.1	XP_002506141.1	
	<i>Micromonas pusilla</i>	XP_003058165.1	XP_003056751.1	No
		XP_003060259.1		
	<i>Coccomyxa subellipsoidea C-169</i>	XP_005648743.1	XP_005642764.1	XP_005645422.1 XP_005645113.1 XP_005645114.1 XP_005646358.1 XP_005848640.1
	<i>Chlorella variabilis NC64A</i>	XP_005852238.1	XP_005843911.1	
Red algae	<i>Cyanidioschyzon merolae</i>	XP_005535714.1	No	No
	<i>Galdieria sulphuraria</i>	NO	XP_005709370.1 and XP_005709369.1	XP_005708906.1
Stramenopile	<i>Phaeodactylum tricornutum</i>	XP_002176357.1	XP_002182611.1	XP_002180610.1 XP_002180609.1 XP_002180883.1 XP_002290743.1
	<i>Thalassiosira pseudonana</i>	XP_002289677.1 XP_002291430.1 XP_002289678.1	XP_002286885.1	
	<i>Fragilariopsis cylindrus</i>	JGI Protein Id: 168118 JGI Protein Id: 261302	JGI Protein ID: 189180	JGI Protein ID: 169403
	<i>Aureococcus anophagefferens</i>	EGB10338.1 EGB09525.1 EGB06148.1 EGB09718.1	AC236110.1 EGB04621.1 EGB04867.1	EGB09791.1 EGB12497.1 EGB10167.1
	<i>Nannochloropsis gaditana CCMP526</i>	XP_005853502.1 XP_005853501.1	No	No
Hapto-phyte	<i>Emiliana huxley</i>	XP_005786216.1 XP_005777157.1 XP_005762577.1 XP_005758261.1 EOD18227.1 XP_005771362.1 XP_005771219.1	XP_005769752.1 XP_005762959.1 XP_005788156.1	XP_005775869.1
Crypto-phyte	<i>Guillardia theta</i>	XP_005835818.1 XP_005827979.1	XP_005818934.1	XP_005818444.1 XP_005834963.1 XP_005839924.1 XP_005827855.1 XP_005835841.1 XP_005829578.1

Sequences were retrieved using the KEGG orthology or by search with the protein's name (as defined in KEGG) in JGI, Phytozome or Genome at NCBI databanks. In addition, search for candidates was carried out by BLASTP or TBLASTN using human alpha-1,2-mannosidase protein sequences, rabbit GnT1 (P27115) and *Arabidopsis thaliana* α (1,3)-FucT (Q9LJK1 and Q9FX97) as query sequences.

the synthesis of putative immunogenic epitopes with the final goal to selectively inactivate their expression.

AUTHOR CONTRIBUTION

Elodie Mathieu-Rivet and Marie-Christine Kiefer-Meyer equally contributed to this work by drafting the manuscript and performing the genomic analyses of the microalgae genomes which are summarized in the **Table 2** of the manuscript. Gaëtan Vanier, Clément Ovide and Carole Burel participated to the writing and took care of the figure drawing. Patrice Lerouge and Muriel Bardor came out with the idea of this review, participate to the writing and coordinated the efforts from the team prior to take care of the submission of the manuscript.

ACKNOWLEDGMENTS

We apologize to colleagues whose work is not cited due to space limitations. Part of the work described in this review was supported financially by the University of Rouen, France as well as the French National Agency for research, ANR, within the PLANT-KBBE 2008 program in the frame of the ALGALGLYCO project. Finally, authors especially thank Dr Zhang Peiqing, Bioprocessing Technology Institute, Singapore and Prof. Driouch Azeddine, Laboratoire Glyco-MEV, University of Rouen, France for critical reading of this manuscript.

REFERENCES

- Altmann, F., Schwihla, H., Staudacher, E., Glössl, J., and März, L. (1995). Insect cells contain an unusual, membrane-bound-N-acetylglucosaminidase probably involved in the processing of protein N-glycans. *J. Biol. Chem.* 270, 17344–17349. doi: 10.1074/jbc.270.29.17344
- Ambrust, E. V., Berges, J., Bowler, C., Green, B., Martinez, D., Putnam, N., et al. (2004). *The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism*. Lawrence Berkeley National Laboratory. Available online at: <http://escholarship.org/uc/item/03j415ph> [Accessed June 16, 2014].
- Apt, K. E., Grossman, A. R., and Kroth-Pancic, P. G. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* 252, 572–579.
- Baïet, B., Burel, C., Saint-Jean, B., Louvet, R., Menu-Bouaouiche, L., Kiefer-Meyer, M.-C., et al. (2011). N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *J. Biol. Chem.* 286, 6152–6164. doi: 10.1074/jbc.M110.175711
- Bai, L.-L., Yin, W.-B., Chen, Y.-H., Niu, L.-L., Sun, Y.-R., Zhao, S.-M., et al. (2013). A new strategy to produce a defensin: stable production of mutated NP-1 in nitrate reductase-deficient *Chlorella ellipsoidea*. *PLoS ONE* 8:e54966. doi: 10.1371/journal.pone.0054966
- Bakker, H., Bardor, M., Molthoff, J. W., Gomord, V., Elbers, I., Stevens, L. H., et al. (2001). Galactose-extended glycans of antibodies produced by transgenic plants. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2899–2904. doi: 10.1073/pnas.031419998
- Bakker, H., Rouwendal, G. J., Karnoup, A. S., Florack, D. E., Stoopen, G. M., Helsper, J. P., et al. (2006). An antibody produced in tobacco expressing a hybrid β -1, 4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7577–7582. doi: 10.1073/pnas.0600879103
- Balshüsemann, D., and Jaenicke, L. (1990). The oligosaccharides of the glycoprotein pheromone of *Volvox carteri* f. *naragensis* Iyengar (Chlorophyceae). *Eur. J. Biochem.* 192, 231–237. doi: 10.1111/j.1432-1033.1990.tb19220.x
- Bardor, M., Faveeuw, C., Fitchette, A.-C., Gilbert, D., Galas, L., Trottein, F., et al. (2003). Immunoreactivity in mammals of two typical plant glyco-epitopes, core α (1,3)-fucose and core xylose. *Glycobiology* 13, 427–434. doi: 10.1093/glycob/cwg024
- Beer, L. L., Boyd, E. S., Peters, J. W., and Posewitz, M. C. (2009). Engineering algae for biohydrogen and biofuel production. *Curr. Opin. Biotechnol.* 20, 264–271. doi: 10.1016/j.copbio.2009.06.002
- Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D. D., et al. (2012). The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol.* 13:R39. doi: 10.1186/gb-2012-13-5-r39
- Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., et al. (2010). The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell Online* 22, 2943–2955. doi: 10.1105/tpc.110.076406
- Bogen, C., Al-Dilaimi, A., Albersmeier, A., Wichmann, J., Grundmann, M., Rupp, O., et al. (2013). Reconstruction of the lipid metabolism for the microalga *Monoraphidium neglectum* from its genome sequence reveals characteristics suitable for biofuel production. *BMC Genomics* 14:926. doi: 10.1186/1471-2164-14-926
- Burda, P., and Aebi, M. (1999). The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta* 1426, 239–257. doi: 10.1016/S0304-4165(98)00127-5
- Buyel, J. F., and Fischer, R. (2014). Downstream processing of biopharmaceutical proteins produced in plants: the pros and cons of flocculants. *Bioengineered* 5, 138–142. doi: 10.4161/bioe.28061
- Cadoret, J.-P., and Bernard, O. (2008). La production de biocarburant lipidique avec des microalgues: promesses et défis. *J. Soc. Biol.* 202, 201–211. doi: 10.1051/jbio:2008022
- Castilho, A., Neumann, L., Gattinger, P., Strasser, R., Vorauer-Uhl, K., Sterovsky, T., et al. (2013). Generation of biologically active multi-sialylated recombinant human EPOFc in plants. *PLoS ONE* 8:e54836. doi: 10.1371/journal.pone.0054836
- Castilho, A., Pabst, M., Leonard, R., Veit, C., Altmann, F., Mach, L., et al. (2008). Construction of a functional CMP-sialic acid biosynthesis pathway in *Arabidopsis*. *Plant Physiol.* 147, 331–339. doi: 10.1104/pp.108.117572
- Castilho, A., Strasser, R., Stadlmann, J., Grass, J., Jez, J., Gattinger, P., et al. (2010). In planta protein sialylation through overexpression of the respective mammalian pathway. *J. Biol. Chem.* 285, 15923–15930. doi: 10.1074/jbc.M109.088401
- Cerutti, H., Johnson, A. M., Gillham, N. W., and Boynton, J. E. (1997). Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. *Plant Cell* 9, 925–945. doi: 10.1105/tpc.9.6.925
- Cerutti, H., Ma, X., Msanne, J., and Repas, T. (2011). RNA-mediated silencing in algae: biological roles and tools for analysis of gene function. *Eukaryot Cell* 10, 1164–1172. doi: 10.1128/EC.05106-11
- Chai, X.-J., Chen, H.-X., Xu, W.-Q., and Xu, Y.-W. (2013). Expression of soybean Kunitz trypsin inhibitor gene SKTI in *Dunaliella salina*. *J. Appl. Phycol.* 25, 139–144. doi: 10.1007/s10811-012-9847-8
- Chen, Y., Wang, Y., Sun, Y., Zhang, L., and Li, W. (2001). Highly efficient expression of rabbit neutrophil peptide-1 gene in *Chlorella ellipsoidea* cells. *Curr. Genet.* 39, 365–370. doi: 10.1007/s002940100205
- Chung, C. H., Mirakhor, B., Chan, E., Le, Q.-T., Berlin, J., Morse, M., et al. (2008). Cetuximab-induced anaphylaxis and IgE specific for galactose- α -1, 3-galactose. *New Engl. J. Med.* 358, 1109–1117. doi: 10.1056/NEJMoa074943
- Colgan, R., Atkinson, C. J., Paul, M., Hassan, S., Drake, P. M. W., Sexton, A. L., et al. (2010). Optimisation of contained *Nicotiana tabacum* cultivation for the production of recombinant protein pharmaceuticals. *Transgenic Res.* 19, 241–256. doi: 10.1007/s11248-009-9303-y
- Cox, K. M., Sterling, J. D., Regan, J. T., Gasdaska, J. R., Frantz, K. K., Peele, C. G., et al. (2006). Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat. Biotechnol.* 24, 1591–1597. doi: 10.1038/nbt1260
- Crispin, M., Harvey, D. J., Chang, V. T., Yu, C., Aricescu, A. R., Jones, E. Y., et al. (2006). Inhibition of hybrid- and complex-type glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. *Glycobiology* 16, 748–756. doi: 10.1093/glycob/cwj119
- Decker, E. L., and Reski, R. (2012). Glycoprotein production in moss bioreactors. *Plant Cell Rep.* 31, 453–460. doi: 10.1007/s00299-011-1152-5
- Demain, A. L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 27, 297–306. doi: 10.1016/j.biotechadv.2009.01.008
- Demurtas, O. C., Massa, S., Ferrante, P., Venuti, A., Franconi, R., and Giuliano, G. (2013). A *Chlamydomonas*-derived human papillomavirus 16 E7 vaccine induces specific tumor protection. *PLoS ONE* 8:e61473. doi: 10.1371/journal.pone.0061473

- De Muynck, B., Navarre, C., and Boutry, M. (2010). Production of antibodies in plants: status after twenty years. *Plant Biotechnol. J.* 8, 529–563. doi: 10.1111/j.1467-7652.2009.00494.x
- Díaz-Santos, E., Vega, M., Vila, M., Vígara, J., and León, R. (2013). *Efficiency of Different Heterologous Promoters in the Unicellular Microalga Chlamydomonas Reinhardtii*. Available online at: <http://onlinelibrary.wiley.com/doi/10.1002/btpr.1690/full> [Accessed December 1, 2013].
- Drake, P. M. W., Barbi, T., Sexton, A., McGowan, E., Stadlmann, J., Navarre, C., et al. (2009). Development of rhizosecretion as a production system for recombinant proteins from hydroponic cultivated tobacco. *FASEB J.* 23, 3581–3589. doi: 10.1096/fj.09-131771
- Eichler-Stahlberg, A., Weisheit, W., Ruecker, O., and Heitzer, M. (2009). Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta* 229, 873–883. doi: 10.1007/s00425-008-0879-x
- Feng, S., Xue, L., Liu, H., and Lu, P. (2009). Improvement of efficiency of genetic transformation for *Dunaliella salina* by glass beads method. *Mol. Biol. Rep.* 36, 1433–1439. doi: 10.1007/s11033-008-9333-1
- Fischer, N., and Rochaix, J.-D. (2001). The flanking regions of *PsaD* drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics* 265, 888–894. doi: 10.1007/s004380100485
- Fitchette, A.-C., Dinh, O. T., Faye, L., and Bardor, M. (2007). “Plant Proteomics and Glycosylation,” in *Plant Proteomics Methods in Molecular Biology*, eds H. Thiellement, M. Zivy, C. Damerval, and V. Méchin (Humana Press), 317–342. Available online at: <http://link.springer.com/protocol/10.1385/1-59745-227-0%3A317> [Accessed June 16, 2014].
- Fuhrmann, M., Oertel, W., and Hegemann, P. (1999). A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. *Plant J.* 19, 353–361. doi: 10.1046/j.1365-313X.1999.00526.x
- Gagneux, P., and Varki, A. (1999). Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* 9, 747–755. doi: 10.1093/glycob/9.8.747
- Geng, D., Wang, Y., Wang, P., Li, W., and Sun, Y. (2003). Stable expression of hepatitis B surface antigen gene in *Dunaliella salina* (Chlorophyta). *J. Appl. Phycol.* 15, 451–456. doi: 10.1023/B:JAPH.0000004298.89183.e5
- Gil, G.-C., Velander, W. H., and Van Cott, K. E. (2009). N-glycosylation microheterogeneity and site occupancy of an Asn-X-Cys sequon in plasma-derived and recombinant Protein C. *Proteomics* 9, 2555–2567. doi: 10.1002/pmic.200800775
- Gödel, S., Becker, B., and Melkonian, M. (2000). Flagellar membrane proteins of *Tetraselmis striata* butcher (Chlorophyta). *Protist* 151, 147–159. doi: 10.1078/1434-4610-00015
- Gomord, V., Fitchette, A.-C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., et al. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 8, 564–587. doi: 10.1111/j.1467-7652.2009.00497.x
- Gonzalez, N. H., Felsner, G., Schramm, F. D., Klingl, A., Maier, U.-G., and Bolte, K. (2011). A single peroxisomal targeting signal mediates matrix protein import in diatoms. *PLoS ONE* 6:e25316. doi: 10.1371/journal.pone.0025316
- Grabowski, G. A., Golemba, M., and Shaaltiel, Y. (2014). Taliglucerase alfa: an enzyme replacement therapy using plant cell expression technology. *Mol. Genet. Metab.* 112, 1–8. doi: 10.1016/j.jymgme.2014.02.011
- Grass, J., Pabst, M., Kolarich, D., Polt, G., Leonard, R., Brecker, L., et al. (2011). Discovery and structural characterization of fucosylated oligomannosidic N-glycans in mushrooms. *J. Biol. Chem.* 286, 5977–5984. doi: 10.1074/jbc.M110.191304
- Grunow, A., Becker, B., and Melkonian, M. (1993). Isolation and characterization of the Golgi apparatus of a flagellate scaly green alga. *Eur. J. Cell Biol.* 61, 10–10.
- Harris, E. H. (2009). *The Chlamydomonas Sourcebook: Introduction to Chlamydomonas and Its Laboratory Use*. Oxford: Academic Press.
- Hawkins, R. L., and Nakamura, M. (1999). Expression of human growth hormone by the eukaryotic alga, *Chlorella*. *Curr. Microbiol.* 38, 335–341. doi: 10.1007/PL00006813
- He, X., Galpin, J. D., Tropak, M. B., Mahuran, D., Haselhorst, T., von Itzstein, M., et al. (2012). Production of active human glucocerebrosidase in seeds of *Arabidopsis thaliana* complex-glycan-deficient (cgl) plants. *Glycobiology* 22, 492–503. doi: 10.1093/glycob/cwr157
- Heitzer, M., Eckert, A., Fuhrmann, M., and Griesbeck, C. (2007). “Influence of codon bias on the expression of foreign genes in microalgae,” in *Transgenic Microalgae as Green Cell Factories* (Springer), 46–53. Available online at: http://link.springer.com/chapter/10.1007/978-0-387-75532-8_5 [Accessed December 1, 2013].
- Helenius, A., and Aebi, M. (2001). Intracellular functions of N-linked glycans. *Science* 291, 2364–2369. doi: 10.1126/science.291.5512.2364
- Hempel, F., Bozarth, A. S., Lindenkamp, N., Klingl, A., Zauner, S., Linne, U., et al. (2011a). Microalgae as bioreactors for bioplastic production. *Microb. Cell Fact.* 10:81. doi: 10.1186/1475-2859-10-81
- Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011b). Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS ONE* 6:e28424. doi: 10.1371/journal.pone.0028424
- Hempel, F., and Maier, U. G. (2012). An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microb. Cell Fact.* 11:126. doi: 10.1186/1475-2859-11-126
- Ho, S. C. L., Bardor, M., Li, B., Lee, J. J., Song, Z., Tong, Y. W., et al. (2013). Comparison of internal ribosome entry site (IRES) and Furin-2A (F2A) for monoclonal antibody expression level and quality in CHO cells. *PLoS ONE* 8:e63247. doi: 10.1371/journal.pone.0063247
- Hollister, J. R., Shaper, J. H., and Jarvis, D. L. (1998). Stable expression of mammalian beta 1,4-galactosyltransferase extends the N-glycosylation pathway in insect cells. *Glycobiology* 8, 473–480. doi: 10.1093/glycob/8.5.473
- Hossler, P., Khattak, S. F., and Li, Z. J. (2009). Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 19, 936–949. doi: 10.1093/glycob/cwp079
- Huang, C.-J., Lin, H., and Yang, X. (2012). Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J. Ind. Microbiol. Biotechnol.* 39, 383–399. doi: 10.1007/s10295-011-1082-9
- Huether, C. M., Lienhart, O., Baur, A., Stemmer, C., Gorr, G., Reski, R., et al. (2005). Glyco-Engineering of Moss Lacking Plant-Specific Sugar Residues. *Plant Biol.* 7, 292–299. doi: 10.1055/s-2005-837653
- Ioffe, E., and Stanley, P. (1994). Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc. Natl. Acad. Sci. U.S.A.* 91, 728–732. doi: 10.1073/pnas.91.2.728
- Jarvis, E. E., and Brown, L. M. (1991). Transient expression of firefly luciferase in protoplasts of the green alga *Chlorella ellipsoidea*. *Curr. Genet.* 19, 317–321. doi: 10.1007/BF00355062
- Kathiresan, S., Chandrashekar, A., Ravishankar, G. A., and Sarada, R. (2009). *Agrobacterium*-mediated transformation in the green alga *Haematococcus pluvialis* (Chlorophyceae, Volvocales) 1. *J. Phycol.* 45, 642–649. doi: 10.1111/j.1529-8817.2009.00688.x
- Kim, D.-H., Kim, Y. T., Cho, J. J., Bae, J.-H., Hur, S.-B., Hwang, I., et al. (2002). Stable integration and functional expression of flounder growth hormone gene in transformed microalga, *Chlorella ellipsoidea*. *Mar. Biotechnol.* 4, 63–73. doi: 10.1007/s1012601-0070-x
- Kindle, K. L. (1998). “High-frequency nuclear transformation of *Chlamydomonas reinhardtii*,” in *Methods in Enzymology Photosynthesis: Molecular Biology of Energy Capture*, ed L. McIntosh (Academic Press), 27–38. Available online at: <http://www.sciencedirect.com/science/article/pii/S0076687998970057> [Accessed January 20, 2014].
- Kumar, A., Falcao, V. R., and Sayre, R. T. (2013). Evaluating nuclear transgene expression systems in *Chlamydomonas reinhardtii*. *Algal Res.* 2, 321–332. doi: 10.1016/j.algal.2013.09.002
- Lam, M. K., and Lee, K. T. (2012). Microalgae biofuels: a critical review of issues, problems and the way forward. *Biotechnol. Adv.* 30, 673–690. doi: 10.1016/j.biotechadv.2011.11.008
- Lauersen, K. J., Berger, H., Mussgnug, J. H., and Kruse, O. (2013a). Efficient recombinant protein production and secretion from nuclear transgenes in *Chlamydomonas reinhardtii*. *J. Biotechnol.* 167, 101–110. doi: 10.1016/j.jbiotec.2012.10.010
- Lauersen, K. J., Vanderveer, T. L., Berger, H., Kaluza, I., Mussgnug, J. H., Walker, V. K., et al. (2013b). Ice recrystallization inhibition mediated by a nuclear-expressed and -secreted recombinant ice-binding protein in the microalga *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.* 97, 9763–9772. doi: 10.1007/s00253-013-5226-x
- Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Lainé, A.-C., Gomord, V., and Faye, L. (1998). “N-glycoprotein biosynthesis in plants: recent developments and future trends,” in *Protein Trafficking in*

- Plant Cells* (Springer), 31–48. Available online at: http://link.springer.com/chapter/10.1007/978-94-011-5298-3_2 [Accessed December 1, 2013].
- Levy-Ontman, O., Arad, S. M., Harvey, D. J., Parsons, T. B., Fairbanks, A., and Tekoah, Y. (2011). Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga *Porphyridium* sp. *J. Biol. Chem.* 286, 21340–21352. doi: 10.1074/jbc.M110.175042
- Levy-Ontman, O., Fisher, M., Shotland, Y., Weinstein, Y., Tekoah, Y., and Arad, S. M. (2014). Genes involved in the endoplasmic reticulum N-glycosylation pathway of the red microalga *Porphyridium* sp.: a bioinformatic study. *Int. J. Mol. Sci.* 15, 2305–2326. doi: 10.3390/ijms15022305
- Lingg, N., Zhang, P., Song, Z., and Bardor, M. (2012). The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol. J.* 7, 1462–1472. doi: 10.1002/biot.201200078
- Lumbreras, V., Stevens, D. R., and Purton, S. (1998). Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. *Plant J.* 14, 441–447. doi: 10.1046/j.1365-313X.1998.00145.x
- Mata, T. M., Martins, A. A., and Caetano, N. S. (2010). Microalgae for biodiesel production and other applications: a review. *Renew. Sust. Energ. Rev.* 14, 217–232. doi: 10.1016/j.rser.2009.07.020
- Mathieu-Rivet, E., Scholz, M., Arias, C., Dardelle, F., Schulze, S., Le Mauff, F., et al. (2013). Exploring the N-glycosylation pathway in *chlamydomonas reinhardtii* unravels novel complex structures. *Mol. Cell. Proteomics* 12, 3160–3183. doi: 10.1074/mcp.M113.028191
- Matsui, T., Takita, E., Sato, T., Kinjo, S., Aizawa, M., Sugiura, Y., et al. (2011). N-glycosylation at noncanonical Asn-X-Cys sequences in plant cells. *Glycobiology* 21, 994–999. doi: 10.1093/glycob/cwq198
- Maxmen, A. (2012). Drug-making plant blooms. *Nature* 485, 160–160. doi: 10.1038/485160a
- Mayfield, S. P., Franklin, S. E., and Lerner, R. A. (2003). Expression and assembly of a fully active antibody in algae. *Proc. Natl. Acad. Sci.* 100, 438–442. doi: 10.1073/pnas.0237108100
- Merchant, S. S., Kropat, J., Liu, B., Shaw, J., and Warakanont, J. (2012). TAG, You're it! *Chlamydomonas* as a reference organism for understanding algal triacylglycerol accumulation. *Curr. Opin. Biotechnol.* 23, 352–363. doi: 10.1016/j.copbio.2011.12.001
- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318, 245–250. doi: 10.1126/science.1143609
- Metzler, M., Gertz, A., Sarkar, M., Schachter, H., Schrader, J. W., and Marth, J. D. (1994). Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. *EMBO J.* 13, 2056–2065.
- Misaki, R., Fujiyama, K., and Seki, T. (2006). Expression of human CMP-N-acetylneuraminic acid synthetase and CMP-sialic acid transporter in tobacco suspension-cultured cell. *Biochem. Biophys. Res. Commun.* 339, 1184–1189. doi: 10.1016/j.bbrc.2005.11.130
- Misumi, O., Matsuzaki, M., Nozaki, H., Miyagishima, S., Mori, T., Nishida, K., et al. (2005). Cyanidioschyzon merolae genome. A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. *Plant Physiol.* 137, 567–585. doi: 10.1104/pp.104.053991
- Neupert, J., Karcher, D., and Bock, R. (2009). Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J.* 57, 1140–1150. doi: 10.1111/j.1365-313X.2008.03746.x
- Paccalet, T., Bardor, M., Rihouey, C., Delmas, F., Chevalier, C., D'Aoust, M.-A., et al. (2007). Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthetizing enzymes. *Plant Biotechnol. J.* 5, 16–25. doi: 10.1111/j.1467-7652.2006.00211.x
- Padler-Karavani, V., Yu, H., Cao, H., Chokhawala, H., Karp, F., Varki, N., et al. (2008). Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: potential implications for disease. *Glycobiology* 18, 818–830. doi: 10.1093/glycob/cwn072
- Palacpac, N. Q., Yoshida, S., Sakai, H., Kimura, Y., Fujiyama, K., Yoshida, T., et al. (1999). Stable expression of human β 1, 4-galactosyltransferase in plant cells modifies N-linked glycosylation patterns. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4692–4697. doi: 10.1073/pnas.96.8.4692
- Palenik, B., Grimwood, J., Aerts, A., Rouz e, P., Salamov, A., Putnam, N., et al. (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7705–7710. doi: 10.1073/pnas.0611046104
- Park, S., Lee, Y., Lee, J.-H., and Jin, E. (2013). Expression of the high light-inducible Dunaliella LIP promoter in *Chlamydomonas reinhardtii*. *Planta* 238, 1147–1156. doi: 10.1007/s00425-013-1955-4
- Parsons, J., Altmann, F., Arrenberg, C. K., Koprivova, A., Beike, A. K., Stemmer, C., et al. (2012). Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. *Plant Biotechnol. J.* 10, 851–861. doi: 10.1111/j.1467-7652.2012.00704.x
- Potvin, G., and Zhang, Z. (2010). Strategies for high-level recombinant protein expression in transgenic microalgae: a review. *Biotechnol. Adv.* 28, 910–918. doi: 10.1016/j.biotechadv.2010.08.006
- Poulsen, N., and Kr oger, N. (2005). A new molecular tool for transgenic diatoms. *Febs J.* 272, 3413–3423. doi: 10.1111/j.1742-4658.2005.04760.x
- Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., Settlage, R. E., Boore, J. L., et al. (2012). Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nat. Commun.* 3:686. doi: 10.1038/ncomms1688
- Rasala, B. A., Lee, P. A., Shen, Z., Briggs, S. P., Mendez, M., and Mayfield, S. P. (2012). Robust expression and secretion of Xylanase1 in *chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS ONE* 7:e43349. doi: 10.1371/journal.pone.0043349
- Rasala, B. A., and Mayfield, S. P. (2011). The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioengineered* 2, 50–54. doi: 10.4161/bbug.2.1.13423
- Rasala, B. A., and Mayfield, S. P. (2014). Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosyn. Res.* doi: 10.1007/s11120-014-9994-7. [Epub ahead of print].
- Rasala, B. A., Muto, M., Lee, P. A., Jager, M., Cardoso, R. M. F., Behnke, C. A., et al. (2010). Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*: production of therapeutic proteins in algae. *Plant Biotechnol. J.* 8, 719–733. doi: 10.1111/j.1467-7652.2010.00503.x
- Rouwental, G. J., Wuhler, M., Florack, D. E., Koeleman, C. A., Deelder, A. M., Bakker, H., et al. (2007). Efficient introduction of a bisecting GlcNAc residue in tobacco N-glycans by expression of the gene encoding human N-acetylglucosaminyltransferase III. *Glycobiology* 17, 334–344. doi: 10.1093/glycob/cwl078
- Ruecker, O., Zillner, K., Groebner-Ferreira, R., and Heitzer, M. (2008). Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics* 280, 153–162. doi: 10.1007/s00438-008-0352-3
- Sasso, S., Pohnert, G., Lohr, M., Mittag, M., and Hertweck, C. (2012). Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiol. Rev.* 36, 761–785. doi: 10.1111/j.1574-6976.2011.00304.x
- Sch ahs, M., Strasser, R., Stadlmann, J., Kunert, R., Rademacher, T., and Steinkellner, H. (2007). Production of a monoclonal antibody in plants with a humanized N-glycosylation pattern. *Plant Biotechnol. J.* 5, 657–663. doi: 10.1111/j.1467-7652.2007.00273.x
- Schillberg, S., Raven, N., Fischer, R., Twyman, R., and Schiermeyer, A. (2013). Molecular farming of pharmaceutical proteins using plant suspension cell and tissue cultures. *Curr. Pharm. Des.* 19, 5531–5542. doi: 10.2174/1381612811319310008
- Schroda, M., Bl ocker, D., and Beck, C. F. (2000). The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant J.* 21, 121–131. doi: 10.1046/j.1365-313x.2000.00652.x
- S veno, M., Bardor, M., Paccalet, T., Gomord, V., Lerouge, P., and Faye, L. (2004). Glycoprotein sialylation in plants? *Nat. Biotechnol.* 22, 1351–1352. doi: 10.1038/nbt1104-1351 doi: 10.1038/nbt1104-1351
- Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., et al. (2007). Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol. J.* 5, 579–590. doi: 10.1111/j.1467-7652.2007.00263.x
- Shaver, S., Casas-Mollano, J. A., Cerny, R. L., and Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*. *Epigenetics* 5, 301–312. doi: 10.4161/epi.5.4.11608
- Specht, E., Miyake-Stoner, S., and Mayfield, S. (2010). Micro-algae come of age as a platform for recombinant protein production. *Biotechnol. Lett.* 32, 1373–1383. doi: 10.1007/s10529-010-0326-5

- Spolaore, P., Joannis-Cassan, C., Duran, E., and Isambert, A. (2006). Commercial applications of microalgae. *J. Biosci. Bioeng.* 101, 87–96. doi: 10.1263/jbb.101.87
- Staudacher, E. (2012). Methylation—an uncommon modification of glycans. *Biol. Chem.* 393, 675–685. doi: 10.1515/hsz-2012-0132
- Stoger, E., Ma, J. K.-C., Fischer, R., and Christou, P. (2005). Sowing the seeds of success: pharmaceutical proteins from plants. *Curr. Opin. Biotechnol.* 16, 167–173. doi: 10.1016/j.copbio.2005.01.005
- Stork, S., Moog, D., Przyborski, J. M., Wilhelmi, I., Zauner, S., and Maier, U. G. (2012). Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryotic Cell* 11, 1472–1481. doi: 10.1128/EC.00183-12
- Strasser, R., Mucha, J., Mach, L., Altmann, F., Wilson, I. B., Glössl, J., et al. (2000). Molecular cloning and functional expression of beta1, 2-xylosyltransferase cDNA from *Arabidopsis thaliana*. *FEBS Lett.* 472, 105–108. doi: 10.1016/S0014-5793(00)01443-5
- Strasser, R., Stadlmann, J., Schähs, M., Stiegler, G., Quendler, H., Mach, L., et al. (2008). Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnol. J.* 6, 392–402. doi: 10.1111/j.1467-7652.2008.00330.x
- Surzycycki, R., Greenham, K., Kitayama, K., Dibal, F., Wagner, R., Rochaix, J.-D., et al. (2009). Factors effecting expression of vaccines in microalgae. *Biologicals* 37, 133–138. doi: 10.1016/j.biologics.2009.02.005
- Tran, M., Van, C., Barrera, D. J., Pettersson, P. L., Peinado, C. D., Bui, J., et al. (2012). Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* 110, E15–E22. doi: 10.1073/pnas.1214638110
- Tran, M., Zhou, B., Pettersson, P. L., Gonzalez, M. J., and Mayfield, S. P. (2009). Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol. Bioeng.* 104, 663–673. doi: 10.1002/bit.22446
- Twyman, R., Schillberg, S., and Fischer, R. (2013). Optimizing the yield of recombinant pharmaceutical proteins in plants. *Curr. Pharm. Des.* 19, 5486–5494. doi: 10.2174/1381612811319310004
- Van Beers, M. M. C., and Bardor, M. (2012). Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. *Biotechnol. J.* 7, 1473–1484. doi: 10.1002/biot.201200065
- Van Patten, S. M., Hughes, H., Huff, M. R., Piepenhagen, P. A., Waire, J., Qiu, H., et al. (2007). Effect of mannose chain length on targeting of glucocerebrosidase for enzyme replacement therapy of Gaucher disease. *Glycobiology* 17, 467–478. doi: 10.1093/glycob/cwm008
- Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 3, 97–130. doi: 10.1093/glycob/3.2.97
- Varki, A. (2011). Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harb. Perspect. Biol.* 3:a005462. doi: 10.1101/cshperspect.a005462
- Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth, J. D., et al. (2009). Symbol nomenclature for glycan representation. *Proteomics* 9, 5398–5399. doi: 10.1002/pmic.200900708
- Vézina, L.-P., Faye, L., Lerouge, P., D'Aoust, M.-A., Marquet-Blouin, E., Burel, C., et al. (2009). Transient co-expression for fast and high-yield production of antibodies with human-like N-glycans in plants. *Plant Biotechnol. J.* 7, 442–455. doi: 10.1111/j.1467-7652.2009.00414.x
- Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., Cornish, A. J., Harvey, C., et al. (2012). Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *nannochloropsis oceanica* CCMP1779. *PLoS Genet.* 8:e1003064. doi: 10.1371/journal.pgen.1003064
- Vitale, A., and Chrispeels, M. J. (1984). Transient N-acetylglucosamine in the biosynthesis of phytohemagglutinin: attachment in the Golgi apparatus and removal in protein bodies. *J. Cell Biol.* 99, 133–140. doi: 10.1083/jcb.99.1.133
- Walsh, G. (2010). Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* 28, 917–924. doi: 10.1038/nbt0910-917
- Weerapana, E., and Imperiali, B. (2006). Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16, 91R–101R. doi: 10.1093/glycob/cwj099
- Weinreb, N. J. (2008). *Imiglucerase and its Use for the Treatment of Gaucher's Disease*. Available online at: <http://informahealthcare.com/doi/abs/10.1517/14656566.9.11.1987> [Accessed December 1, 2013].
- Wilson, I. B. H., Zeleny, R., Kolarich, D., Staudacher, E., Stroop, C. J. M., Kamerling, J. P., et al. (2001). Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core α 1,3–linked fucose and xylose substitutions. *Glycobiology* 11, 261–274. doi: 10.1093/glycob/11.4.261
- Wong, C.-H. (2005). Protein glycosylation: new challenges and opportunities. *J. Org. Chem.* 70, 4219–4225. doi: 10.1021/jo050278f
- Worden, A. Z., Lee, J.-H., Mock, T., Rouze, P., Simmons, M. P., Aerts, A. L., et al. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *micromonas*. *Science* 324, 268–272. doi: 10.1126/science.1167222
- Xu, J., Dolan, M. C., Medrano, G., Cramer, C. L., and Weathers, P. J. (2012). Green factory: plants as bioproduction platforms for recombinant proteins. *Biotechnol. Adv.* 30, 1171–1184. doi: 10.1016/j.biotechadv.2011.08.020
- Zaslavskaya, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R., and Apt, K. E. (2000). Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* 36, 379–386. doi: 10.1046/j.1529-8817.2000.99164.x
- Zeleny, R., Kolarich, D., Strasser, R., and Altmann, F. (2006). Sialic acid concentrations in plants are in the range of inadvertent contamination. *Planta* 224, 222–227. doi: 10.1007/s00425-005-0206-8
- Zhao, T., Wang, W., Bai, X., and Qi, Y. (2009). Gene silencing by artificial microRNAs in *Chlamydomonas*. *Plant J.* 58, 157–164. doi: 10.1111/j.1365-313X.2008.03758.x
- Zhu, S., Hanneman, A., Reinhold, V. N., Spence, A. M., and Schachter, H. (2004). *Caenorhabditis elegans* triple null mutant lacking UDP-N-acetyl-D-glucosamine: α -3-D-mannoside β 1,2-N-acetylglucosaminyltransferase I. *Biochem. J.* 382, 995–1001. doi: 10.1042/BJ20040793
- Zielinska, D. F., Gnad, F., Wiśniewski, J. R., and Mann, M. (2010). Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141, 897–907. doi: 10.1016/j.cell.2010.04.012

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 May 2014; accepted: 04 July 2014; published online: 28 July 2014.

Citation: Mathieu-Rivet E, Kiefer-Meyer M-C, Vanier G, Ovide C, Burel C, Lerouge P and Bardor M (2014) Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front. Plant Sci.* 5:359. doi: 10.3389/fpls.2014.00359

This article was submitted to *Plant Physiology*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2014 Mathieu-Rivet, Kiefer-Meyer, Vanier, Ovide, Burel, Lerouge and Bardor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

1.9 Objectifs de la thèse

Au cours du résumé bibliographique présenté dans cette introduction, nous avons pu voir l'importance physiologique majeure du processus de *N*-glycosylation. Cette voie de biosynthèse peut ainsi être séparée en deux grandes parties. La première, située dans le RE et très conservée chez les eucaryotes, permet d'aboutir à la synthèse de *N*-glycannes oligomannosidiques impliqués dans l'acquisition de la conformation des glycoprotéines. La seconde partie du processus prend place au sein de l'appareil de Golgi, où cette fois-ci les mécanismes de maturation des *N*-glycannes ont divergé au cours de l'évolution, permettant d'aboutir à la formation de *N*-glycannes complexes organisme-spécifiques impliqués dans de nombreux processus. Les mécanismes de la *N*-glycosylation ont été très largement étudiés chez les levures, les plantes, les invertébrés et les mammifères, qui sont tous répartis dans deux grands phylums que sont les archaeplastids et les unikonts. Cependant, actuellement très peu de données sont disponibles concernant les organismes appartenant à d'autres phylums, comme les diatomées. L'étude des mécanismes de la *N*-glycosylation chez ces microalgues permettra ainsi de mieux comprendre l'origine évolutive de ce processus biologique. De plus, l'étude de la voie de *N*-glycosylation chez les diatomées présente un grand intérêt économique dans le cadre de la production de protéines thérapeutiques recombinantes par des systèmes hétérologues. En effet, il a été récemment montré que *P. tricornutum* présente différents avantages aussi bien sur le plan technique que biologique, et semble très prometteuse en tant qu'usine cellulaire. Cette représentante des diatomées pennées a la particularité de posséder trois morphotypes inter-convertibles, et bien que *P. tricornutum* fasse l'objet de nombreuses études, très peu d'informations sont disponibles quant à ces processus de *N*-glycosylation et aux mécanismes qui régissent sa morphogénèse.

C'est donc dans ce contexte scientifique que s'inscrivent mes travaux de thèse visant à étudier d'une part les morphotypes de *P. tricornutum*, et d'autre part les processus de la voie de biosynthèse des *N*-glycannes chez cette diatomée modèle.

L'étude des morphotypes de *P. tricornutum* a été abordée par une approche de transcriptomique à haut débit (RNA-seq) afin d'apporter de nouvelles informations et une vision globale sur les processus métaboliques pouvant être impliqués. De plus, les données générées lors de cette étude ont permis d'identifier de nouveaux gènes de référence pour l'étude des morphotypes de *P. tricornutum* par qRT-PCR. Nous avons illustré l'utilité de ces outils en étudiant le profil d'expression de différentes protéines clé intervenant dans la voie de *N*-glycosylation chez les trois morphotypes.

Dans le but de compléter la caractérisation de la voie de *N*-glycosylation de cette microalgue, la localisation subcellulaire de différentes glycosyltransférases a été réalisée en utilisant une stratégie de surexpression de gènes fusionnés à des étiquettes. Conjointement, la caractérisation des trois α 1,3-fucosyltransférases putatives de *P. tricornutum* a été initiée en utilisant une stratégie de complémentation fonctionnelle dans un système hétérologue mammifère dépourvu de la fucosyltransférase 8 endogène.

L'étude par transcriptomique des trois morphotypes de *P. tricornutum* est présentée sous la forme d'un article scientifique rédigé en langue anglaise dans le chapitre trois. Les résultats portant sur l'identification de gènes de référence pour l'étude des trois morphotypes de *P. tricornutum* en qRT-PCR sont présentés dans un quatrième chapitre, une nouvelle fois sous forme d'article scientifique rédigé en langue anglaise. Et enfin, le cinquième chapitre rédigé en français est consacré à l'étude des trois α 1,3-fucosyltransférases putatives de *P. tricornutum*. Les conclusions et discussions relatives aux résultats obtenus sont directement intégrées à la fin de chaque chapitre. Pour terminer, une conclusion générale sur l'ensemble des travaux vient clore ce manuscrit.

Chapitre II : Matériel et méthodes

2 Matériel

2.1 Matériel végétal : *Phaeodactylum tricornutum*

Les travaux décrits dans ce manuscrit ont été réalisés à partir de deux souches de *P. tricornutum*. Les cellules issues de la souche séquencée Pt.1.8.6 (CCAP1055/1) ont été utilisées pour les aspects de clonage et surexpression des glycosyltransférases, et la souche Pt3 (CCAP1052/1B) a été utilisée pour l'étude de la morphogenèse par analyse transcriptomique à haut débit. Respectivement, ces microalgues ont été initialement fournies par le laboratoire de Physiologie et Biotechnologie des Algues (PBA) de l'IFREMER de Nantes et par le laboratoire du Dr C. Bowler de l'Ecole Normale Supérieure de Paris.

2.2 Matériel animal : Cellules d'ovaire de hamster chinois (CHO)

La mise en évidence de l'activité des fucosyltransférases de *P. tricornutum* a été réalisée en utilisant des cellules de mammifère CHO FuT8^{-/-}, mutées sur le gène de l' α -1,6-fucosyltransférase endogène, issues de la souche sauvage CHO K1. Ces deux lignées ont été généreusement fournies par Gregory J. Cost de Sangamo Biosciences en Californie, USA (Malphettes et al., 2010). Les cellules ont été cultivées dans du milieu F-12 (Gibco) complémenté avec 10% de sérum fœtal bovin (FBS, Gibco), puis placées dans un incubateur à 37°C contenant une atmosphère contrôlée de CO₂ fixée à 5%.

2.3 Vecteurs plasmidiques utilisés

2.3.1 pUC57+FuT54599

Le gène codant potentiellement pour la fucosyltransférase 54599 (Phatr3_J54599) a été synthétisé par la société Genecust (www.genecust.com) et inséré dans le vecteur plasmidique pUC57. La carte du plasmide pUC57+FuT54599 est représentée sur la **figure 34**.

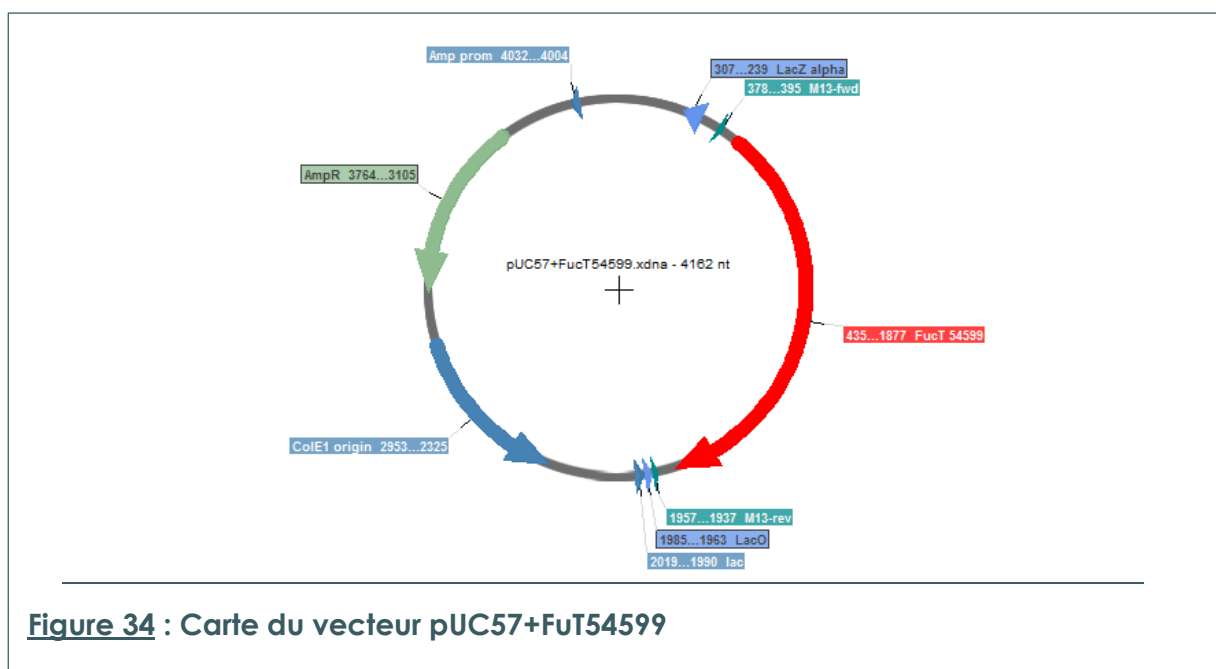
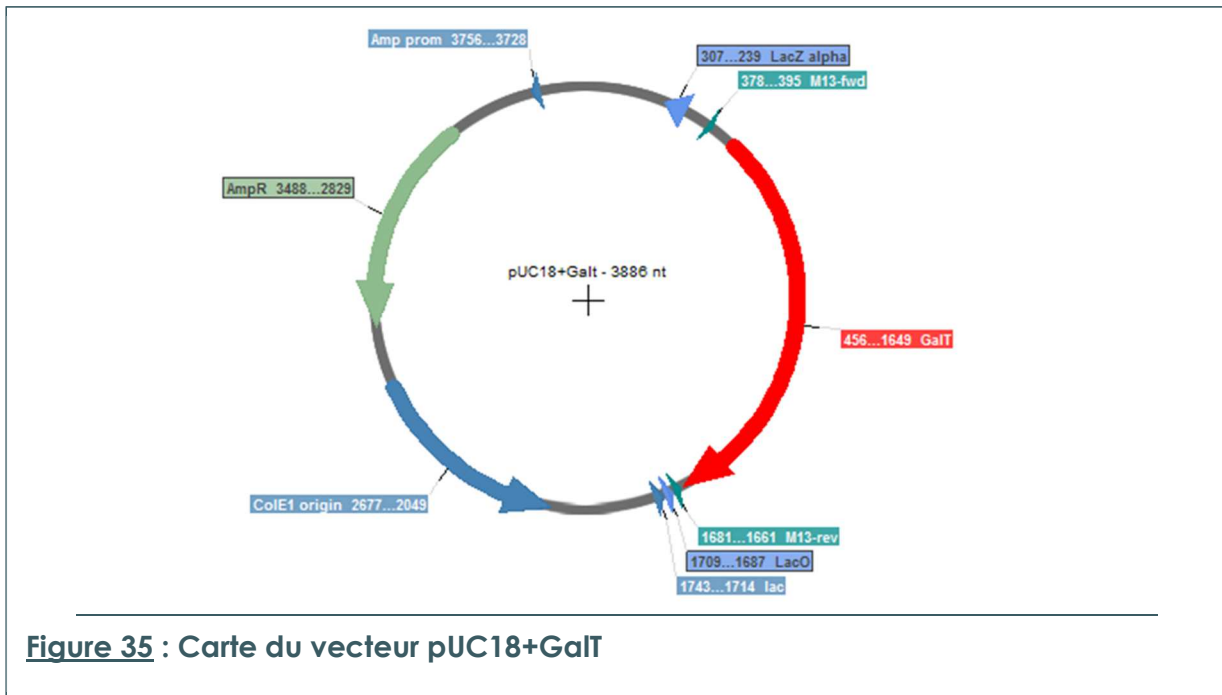


Figure 34 : Carte du vecteur pUC57+FuT54599

2.3.2 pUC18+GalT

Le gène codant pour la β -1,4-galactosyltransférase humaine utilisé lors de ces travaux est issu du plasmide pUC18+GalT décrit dans Bakker et al., 2001. La carte de ce plasmide est représentée sur la **figure 35**.



2.3.3 pcDNA3.1/V5-His-TOPO®

Le vecteur linéaire pcDNA3.1/V5-His-TOPO® (Invitrogen, K4800-01) est à la fois un vecteur de clonage utilisé pour transformer des cellules d'*E. coli* et un vecteur d'expression utilisé lors de la transfection de cellules CHO. Il permet un clonage par « T/A cloning », c'est-à-dire une ligation du produit PCR par hybridation de ses extrémités adénosines avec les extrémités thymidines du vecteur. Ce vecteur permet l'ajout de deux étiquettes en C-terminal de la protéine exprimée grâce à deux séquences placées en 3' de l'insert et codant pour un épitope V5 et une queue hexahistidine. Deux gènes de sélection sont aussi présents sur ce vecteur : un gène de résistance à l'ampicilline qui est utilisé lors du clonage dans les bactéries et un gène de résistance à la néomycine permettant la sélection des cellules de mammifère transfectées avec le transgène. La carte du vecteur est détaillée sur la **figure 36**.

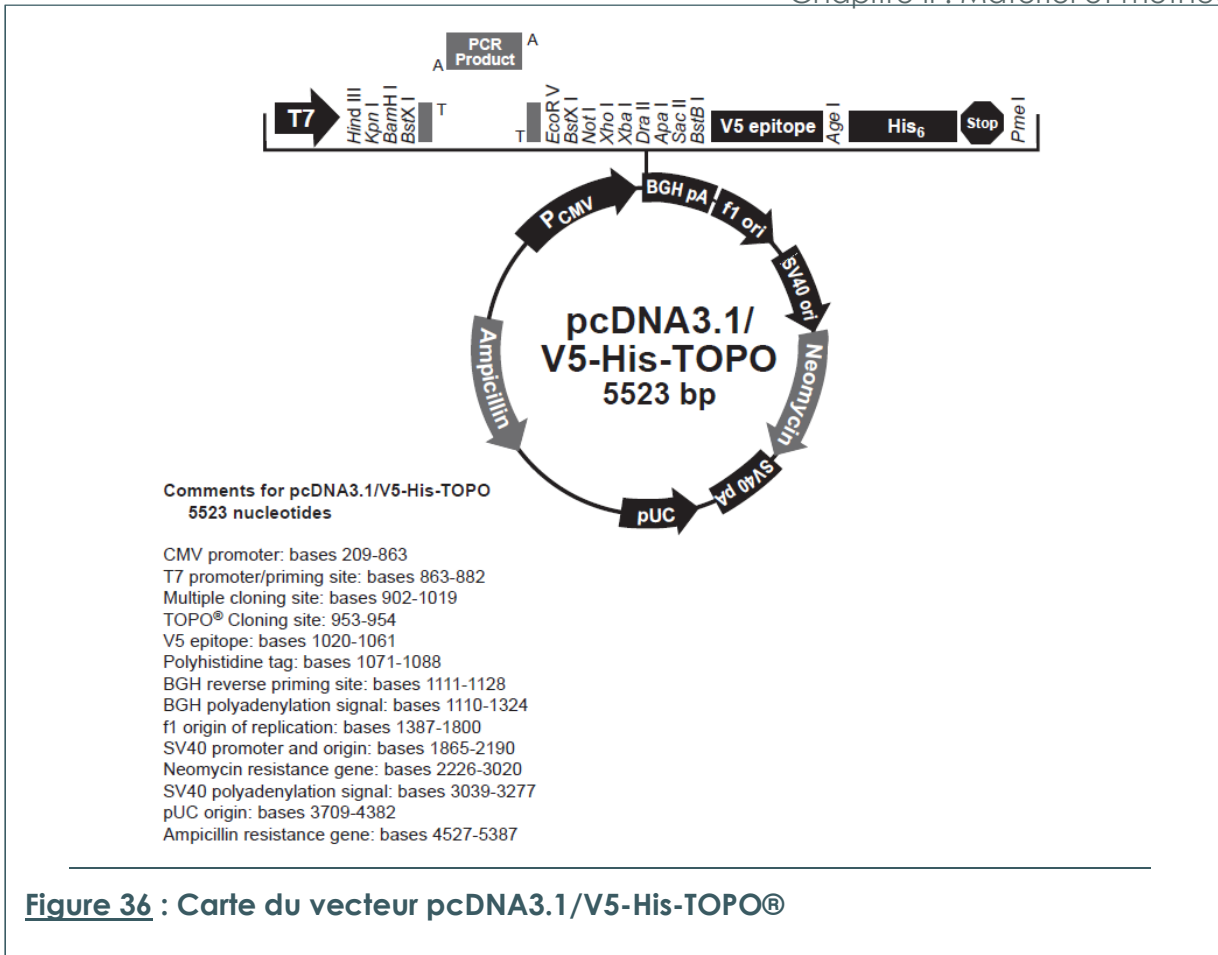


Figure 36 : Carte du vecteur pcDNA3.1/V5-His-TOPO®

2.3.4 pJET1.2/blunt

Le vecteur linéarisé pJET1.2/blunt (Thermo Scientific, K1231) est un vecteur de clonage utilisé pour la transfection des bactéries *E. coli*. Son site de clonage (extrémités franches) est situé au milieu d'un gène létal qui se trouve inactivé par la ligation de l'insert, permettant ainsi une sélection positive des cellules ayant intégré le plasmide recombinant. Ce plasmide possède en plus un gène de résistance à l'ampicilline permettant de sélectionner les bactéries transformées par le vecteur. La carte du vecteur est détaillée sur la **figure 37**.

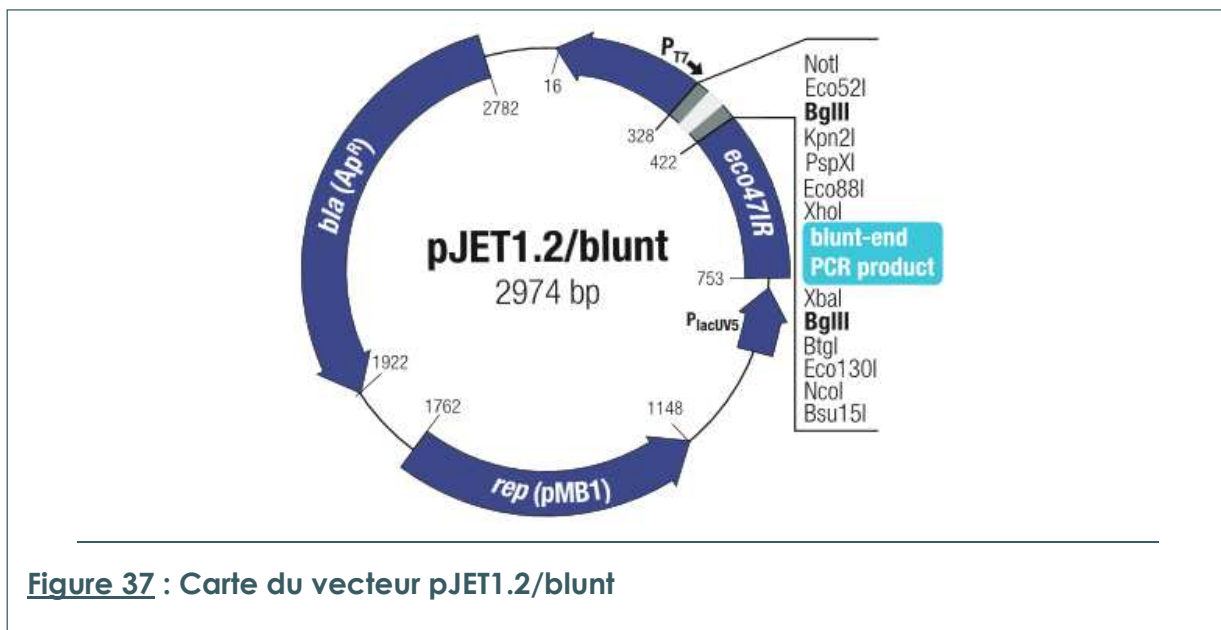
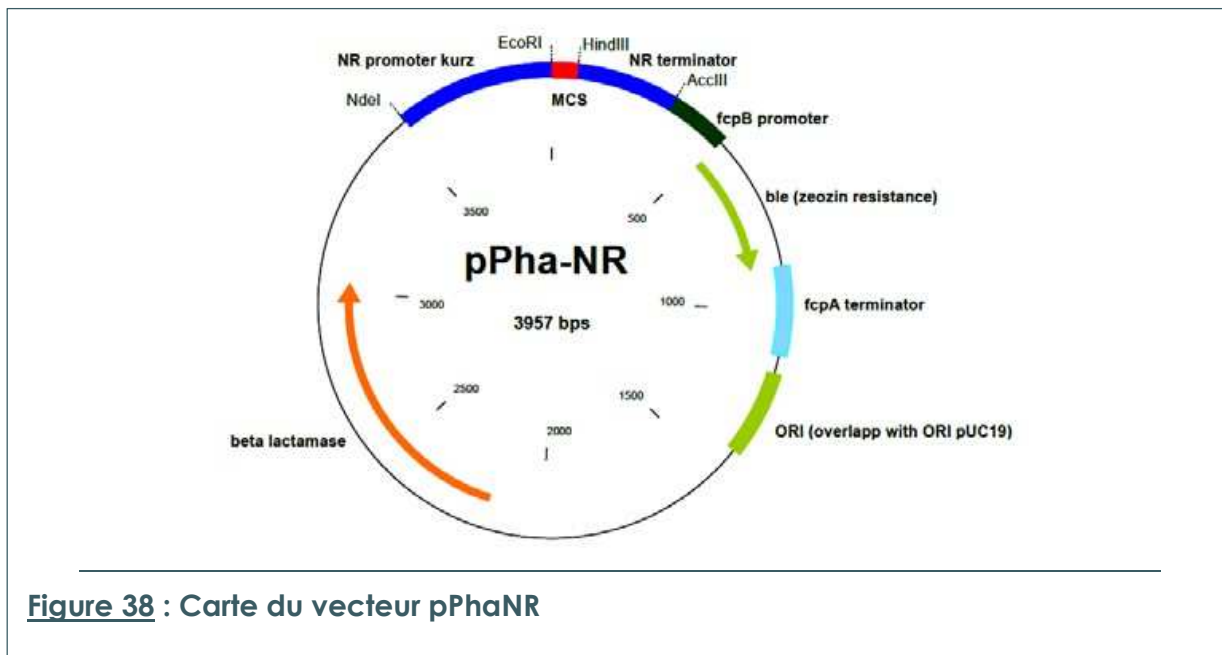


Figure 37 : Carte du vecteur pJET1.2/blunt

2.3.5 pPha-NR

Le vecteur pPha-NR est un vecteur d'expression utilisé chez *P. tricornutum*. La séquence de ce vecteur est disponible dans GenBank (Accession : JN180663.1). Ce vecteur possède un promoteur NR (Nitrate Réductase) inductible au nitrate permettant de contrôler l'expression de la protéine recombinante. Deux gènes de sélection sont présents sur ce vecteur : un gène de résistance à l'ampicilline qui est utilisé lors du clonage dans les bactéries et un gène de résistance à la zéocine permettant la sélection des diatomées possédant le transgène. La carte du vecteur est détaillée sur la **figure 38**. Ce vecteur nous a été généreusement fourni par le Dr F. Hempel et le Pr U. Maier du laboratoire Synmikro de l'Université de Marbourg (Allemagne).



2.4 Amorces PCR

Les séquences des amorces utilisées pour les expériences de clonage, séquençage et analyses par RT-qPCR (Reverse transcription qPCR) ont été définies à l'aide du logiciel sequencher 5.0 (Gene codes), ou du programme « Primer-BLAST » (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) en utilisant les paramètres par défaut, à l'exception des paramètres qui ont été optimisés de la façon suivante : Product size = 70-150bp ; Tm= 57°C-63°C ; Organisme = *Phaeodactylum tricornutum*. Les amorces déssalées ont été commandées chez Eurogentec (<http://www.eurogentec.com>) ou Sigma-Aldrich (<http://www.sigmaaldrich.com/life-science/custom-oligos.html>). La liste des amorces est présentée dans les **tableaux 1 et 2**.

Gène cible	Nom de l'amorce	Séquence nucléotidique 5'-3'	Tm (°C)	Utilisations
eGFP	5_eGFP_KpnI	GGTACCCCGGAGGAGGAGGA	73	Clonage
	3_eGFP_HindIII	AAGCTTTTACTTGTACAGCTCG	60	Clonage
FuT46109	5_46109_SacI	GAGCTCATGCAGAGTTTGAAT	61	Clonage
	3_46109V5_Sall	GTCGACTTACGTAGAATCGAG	58	Clonage
	3_46109_KpnI	GGTACCTGACGATACGTAAAAT	59	Clonage
FuT46110	5_46110_SacI	GAGCTCATGATTTCGCATACCA	65	Clonage
	3_46110_BamHI	GGATCCCACGAAGGTTTCCAC	69	Clonage
	3_46110V5_NheI	GCTAGCTTACGTAGAATCGAGACCGAGGAGGGTTAGGG	79	Clonage
	3_46110_int	TCCAACGGCATGATC	58	Séquençage
FuT54599	5_54599_SacI	GAGCTCATGCACTTCGCAAG	65	Clonage / criblage
	3_54599V5_HindIII	AAGCTTACGTAGAATCGAGACCGAGGAGA	72	Clonage / criblage
	3_54599_BglIII	AGATCTCGGATCGAACTCCA	65	Clonage
	FuT54599 5F	ATGTCACCTCGCAAGCGTGGTG	56	Clonage
	FuT54599 3R	CGGATCGAACTTCCACGAGG	56	Clonage
FuT46109NM	5_46109NMV5_KpnI	ATTTTACGTATCGTCAGGTACCAATTCT	65	Clonage
	5_46109_NM_SacI	GAGCTCATGGTATCAATCCGC	66	Clonage
	N5_46109NMV5_KpnI	AGAATTGGTACCTGACGATACGTAATAAAC	66	Clonage
	46109_LP1	TCCAATGCAACCGGTAGACT	65	Criblage
	46109_RP2	ATCCAAGGACCACGACTTGT	64	Criblage
	5'_46109NM_Kozak	GCCACCATGGTATCAATCCG	61	Clonage / criblage
	3_46109NMpcDNA	TGACGATACGTAATAAAGTTCGAGTGG	62	Clonage / criblage
	Pt46109_SF1	GCACTGGAAGACTCAGAAC	55	Séquençage
FuT46110NM	5_46110_NM2_SacI	GAGCTCATGCGCGTGTCTTTGC	73	Clonage
	5_46110NMV5_BamHI	GTGGAACCTTCGTGGGATCCAATTC	73	Clonage
	5'_46110NM_Kozak	GCCACCATGCGCGTGTCTTTG	69	Clonage / criblage
	3_46110NMpcDNA	CACGAAGGTTTCCACAGGATCCAAAAATC	70	Clonage / criblage
GnT I	5_GnTI_MunI	CAATTGATGCGGTTGTGGAAACG	72	Clonage / criblage
	3_GnTI_BamHI	GGATCCTCTTTTCGGTGACGGAA	72	Clonage
	3_GnTIV5_HindIII	AAGCTTACGTAGAATCGAGACC	61	Clonage / criblage
β-1,4-galactosyltransférase	3_GalT_BamHI	GGATCCGCTCGGTGTCCCGAT	75	Clonage
	5_GalT_EcoRI2	GAATTCATGAGGCTTCGGGAG	67	Clonage / criblage
Hémagglutinine	N5_HA_BGLII	ATCGTATGGGTACATAGATCTGCTCGGTGTCCCGATG	80	Clonage
	3_HA_HindIII	AAGGTTCTAGGCGTAATCTGGAA	64	Clonage / criblage
Xylosyltransférase	5_XylIT	ATGGCGTTTTTGCCGAATC	67	criblage
	3_XylIT_cmyc	AGAGATCTTCTCAGAAATTAATTTTGTTC	64	criblage
	NR-Rev	AACAGTGATCCTCATCG	50	Séquençage / criblage
	NR-For	AGCAACCATTCATTC	48	Séquençage / criblage
	pJET1.2 forward	CGACTCACTATAGGGAGAGCGGC	69	Séquençage
	pJET1.2 reverse	AAGAACATCGATTTTCCATGGCAG	69	Séquençage
	T7	TAATACGACTCACTATAGGG	52	Séquençage
	BGH Reverse	TAGAAGGCACAGTCGAGG	60	Séquençage

Tableau 1 : Liste des amorces utilisées pour effectuer les amplifications par PCR et les réactions de séquençage

Gène cible	Nom de l'amorce	Séquence nucléotidique 5'-3'	Tm (°C)	Utilisations
Phatr3_J47374	Phatr47374_F1	AGGGCGTGTAGAAACGACAG	62	Gene de référence
	Phatr47374_R1	CCCCGTTCGAATCAAGTAAACC	49,7	Gene de référence
Phatr3_J19761	Phatr19761_F1	CCCGAATTC AACCGAAAGGC	62	Gene de référence
	Phatr19761_R1	ATGGCAGCCTTGAGAGAGC	60	Gene de référence
Phatr3_J38708	Phatr38708_For1	TTTGTGGCTGGCAGGAGA	60	Validation RNA-seq
	Phatr38708_Rev1	AGCTGTCCGTTACAGTAGC	62	Validation RNA-seq
Phatr3_J47653	Phatr47653_For1	TTGGCGTCTTTGGGGATT	60	Validation RNA-seq
	Phatr47653_Rev1	GCATTGACAACCCACGCAAT	60	Validation RNA-seq
Phatr3_EG00182	Phatr3_EG00182_For1	GCTACGGAGAGCCTGATGTC	64	Validation RNA-seq
	Phatr3_EG00182_Rev1	CTTTTGTGGCCCTTCCACG	62	Validation RNA-seq
Phatr3_J54574	Phatr3_J54574_For1	AGTACAGGGTTCCGGAAGCG	62	Validation RNA-seq
	Phatr3_J54574_Rev1	AGGGATTCTCCGTCCCAA	62	Validation RNA-seq
Phatr3_J12405	Phatr3_J12405_For1	ATTGGGAGGTGTGCTGAGTG	62	Validation RNA-seq
	Phatr3_J12405_Rev1	TGCATCAGCGACTTGACTGT	60	Validation RNA-seq
Phatr3_J44251	Phatr3_J44251_For1	GGCTCCGATCCATTGATTTG	49,7	Validation RNA-seq
	Phatr3_J44251_Rev1	ACCGAGTTGAAAAACGTTGGG	47,3	Validation RNA-seq
Phatr3_EG02507	Phatr3_EG02507_For1	AATACGCCGGTGTGGACATT	60	Validation RNA-seq
	Phatr3_EG02507_Rev1	ATCCGTAACGCTGCTCGAAT	60	Validation RNA-seq
Phatr3_EG01990	Phatr3_EG01990_For1	TGCGATTCATAGTGGAGCCC	62	Validation RNA-seq
	Phatr3_EG01990_Rev1	TGATCTTGGCGACTGCATT	60	Validation RNA-seq
Phatrdraft1583	Phatrdraft1583_For1	TAGGAGGTTC AACACAGCC	62	Validation RNA-seq
	Phatrdraft1583_Rev1	AATGGTTCGAGAGTGGCAAG	62	Validation RNA-seq
Phatr3_J48834	Phatr3_J48834_For1	GTGGAATGACTCCGCTCCAA	62	Validation RNA-seq
	Phatr3_J48834_Rev1	AAAGACTCGAGGACTGCGTG	62	Validation RNA-seq
Phatr3_J49336	Phatr49336_For1	TGTTGAGCCAGCTGAAGGAG	62	Validation RNA-seq
	Phatr49336_Rev1	GAGCGCTTGGTTTGATCGAC	62	Validation RNA-seq
Phatr3_J35771	Phatr35771_For1	TGTCGGGGGAGTGACTACTT	62	Validation RNA-seq
	Phatr35771_Rev1	ACGGAACACATCAACCTCCC	62	Validation RNA-seq
Phatr3_J33783	Phatr33783_For1	TGTCGCTTCTCCGAGGTAT	62	Validation RNA-seq
	Phatr33783_Rev1	TTGCGACTGTACCAGCCAAT	60	Validation RNA-seq
Phatr3_J48356	Phatr48356_For1	CTGCGGAAACCAAAGTCGTG	62	Validation RNA-seq
	Phatr48356_Rev1	CGGAGCCAAAGTCATGGTCT	62	Validation RNA-seq
Phatr3_J45621	Phatr45621_For1	CATGGAAGCGATTGGCGAAG	62	Validation RNA-seq
	Phatr45621_Rev1	AGCCATGACAGCTTCACTCC	62	Validation RNA-seq
Phatr3_EG00085	Phatr3_EG00085_For1	CCGAAAGTTCACTTGAGAGTGC	49,7	Validation RNA-seq
	Phatr3_EG00085_Rev1	AAGCGTGTCCGACCAACACA	60	Validation RNA-seq
Phatr3_J41599	Phatr41599_For1	GAAACGCTGCTTCCGCTTCA	60	Validation RNA-seq
	Phatr41599_Rev1	ATTGCCGTGACTTTTGTGCG	60	Validation RNA-seq
Phatr3_J32459	Phatr32459_For1	CCAACCATCCCCACCATTCA	62	Validation RNA-seq
	Phatr32459_Rev1	GGGCTCCGCTGTAGTAGT	62	Validation RNA-seq
Phatr3_J33840	Phatr33840_For1	AGAAGAGGTGCGAGAAAGCC	62	Validation RNA-seq
	Phatr33840_Rev1	GCTCTCCAGCTCAGCAAGAA	62	Validation RNA-seq
Phatr3_J39760	Phatr39760_For1	AGTTAAGGCAGAAAGCACCCC	62	Validation RNA-seq
	Phatr39760_Rev1	GCAAGTCAAAGCGTCTTGG	62	Validation RNA-seq
Phatr3_J44524	Phatr44524_For1	GATATGGGCTGCTTGACGA	62	Validation RNA-seq
	Phatr44524_Rev1	TTGCGGAGATTCCACTCGTT	60	Validation RNA-seq
Phatr3_J50304	Phatr50304_For1	TTCGGTAGCTCGAGATGGA	62	Validation RNA-seq
	Phatr50304_Rev1	GGCAGTTCGAGAGATCCG	64	Validation RNA-seq
GDP-Fuc Transporteur	GDPfucTptFOR1	CTCTGATGGCGTCTTGGT	62	Validation RNA-seq
	GDPfucTptREV1	CGCTTGTGTTCCGTTTTTGG	47,3	Validation RNA-seq
GDP-Fuc Transporteur	GDPfucTptFOR2	GCAAAAACCGAACCAAGCG	47,3	Validation RNA-seq
	GDPfucTptREV2	ACGAATCCAGGCACGTAG	62	Validation RNA-seq
Phatr3_J46110	Phatr3_J46110_FOR7	ATGTCTGGCAGTACCGATGG	62	Validation RNA-seq
	Phatr3_J46110_REV7	CAAACGCTTCGGACAGTACC	62	Validation RNA-seq
Phatr3_J49563	Phatr3_J49563_FOR3	ATGGAATCTGCTCGCTCTCG	62	Validation RNA-seq
	Phatr3_J49563_REV3	CAGAATCCGGAAAGGCAGGT	62	Validation RNA-seq
Phatr3_J45073	Phatr45073_FOR1	GGATGAAGTCAACATGGCCG	62	Validation RNA-seq
	Phatr45073_REV1	CGAACCATTTGCTTTTCGGG	62	Validation RNA-seq
Phatr3_J54599	Pt54599FOR1	GTGGCAATTTCCGACCCTA	62	Validation RNA-seq
	Pt54599REV1	CGTGGACTACCGTAATCGGG	64	Validation RNA-seq
Phatr3_J54844	Phatr3_J54844_FOR4	ATCTCCGCACCTACCACTT	62	Validation RNA-seq
	Phatr3_J54844_REV4	TTGGGCTTCCAGATCCACTG	62	Validation RNA-seq
Phatr3_J46109	Phatr46109_FOR3	ATCAGTAGCGCCGCAATACA	60	Validation RNA-seq
	Phatr46109_REV3	ATTCCTTCTGCGGACTGTG	62	Validation RNA-seq

Tableau 2 : Liste des amorces utilisées pour la validation des données de RNA-seq par qPCR

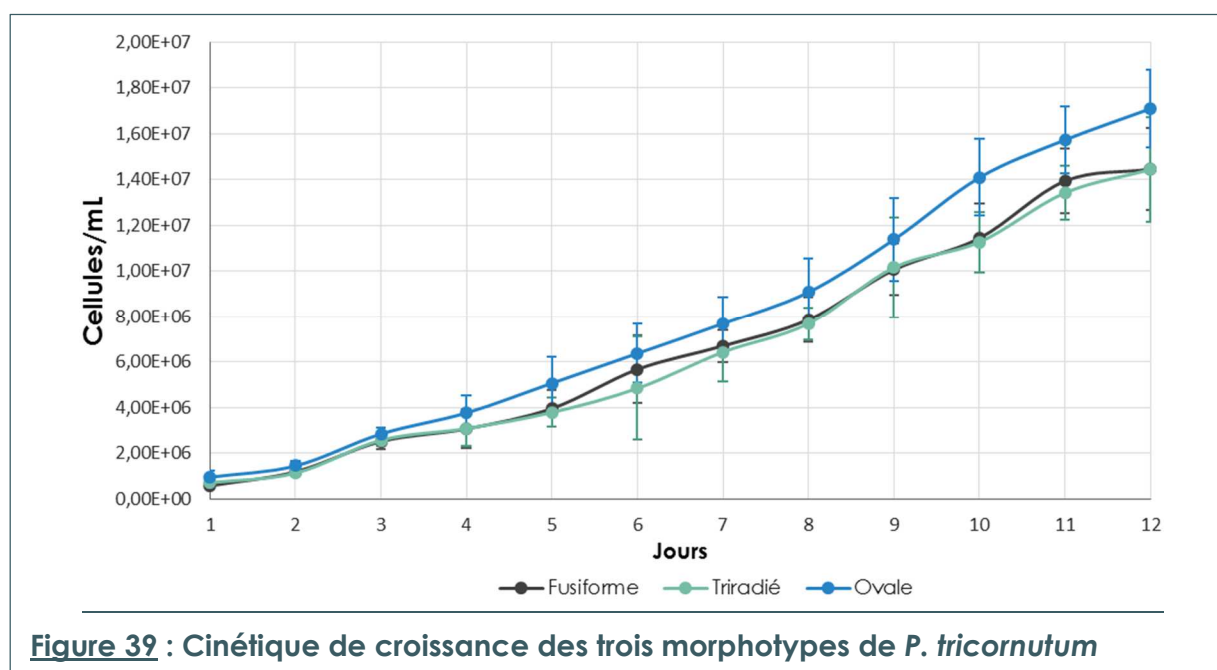
■ : amorces validées avec 95%-100% d'efficacité

■ : amorces validées avec 92%-95% d'efficacité

3 Méthodes

3.1 Culture de *P. tricornutum*

Les cultures de diatomées sont entretenues par repiquage de 25 mL de la culture mère dans 200 mL de milieu frais, préalablement filtré (taille des pores égale à 0,45 μm) et autoclavé pendant 20 minutes à 1 bar. La composition du milieu en fonction du morphotype cellulaire cultivé est donnée dans les **Tableaux 3 et 4**. Les cultures sont ensuite placées dans une armoire de culture, à 19°C, sous une irradiation de 50 $\mu\text{Einstein/m}^2/\text{s}$, avec une alternance jour/nuit de 16 heures/8 heures. Les différentes extractions réalisées au cours de ces travaux ont été effectuées à partir de cultures de cellules âgées de 8 jours, ce qui correspond au milieu de la phase exponentielle de croissance dans notre système de culture avec une densité d'environ 8×10^6 cellules/mL (**figure 39**).



3.2 Culture des lignées cellulaires CHO

Les lignées cellulaires de CHO sont entretenues en passant les cellules lorsqu'elles atteignent une confluence de 80%, soit environ deux fois par semaine. Pour cela le milieu de culture est enlevé, puis les flasques sont rincées avec 10mL de DPBS 1X (Dulbecco's Phosphate Buffered Saline ; Fisher ; E3468W). Les cellules CHO étant des cellules adhérentes, 1mL de trypsine 0,05% (Gibco ; 11590626) est ensuite ajouté pour mettre en suspension les cellules. Les flasques sont placées pendant 3 minutes dans l'incubateur thermostaté à 37°C, puis sont ajoutés ensuite 5mL de milieu F-12 (Gibco) complété avec 10% de sérum foetal bovin (FBS, Gibco) afin d'inactiver la trypsine. La suspension est ensuite centrifugée à 300g pendant

	Culture standard	Culture morphotype Ovale	Culture cellules transfectées
Sea salt (Instant Ocean®)	33,3g/L	3,3g/L	33,3g/L
Na ₂ O ₃ Si	0,08g/L	0,08g/L	0,08g/L
Milieu Conway	1 mL	1 mL	1 mL

Tableau 3 : Composition des milieux de culture de *P. tricornutum*

	Conposés	Conway standard	Conway pour cellules transfectées
Solution principale	Na ₂ EDTA	45g	45g
	NaNO ₃	100g	∅
	NH ₄ Cl	∅	80,25g
	H ₃ BO ₃	33,6	33,6
	NaH ₂ PO ₄	26g	26g
	MnCl ₂	0,36g	0,36g
	FeCl ₃	1,28g	1,28g
	Eau distillée	1L	1L
Solution traces de métaux	ZnCl ₂	0,021g	0,021g
	CoCl ₂	0,02g	0,02g
	(NH ₄) ₆ Mo ₇ O ₂₄	0,009g	0,009g
	CuSO ₄	0,02g	0,02g
	Eau distillée	1 mL	1 mL
Solution vitaminique	Vitamine B1 thiamine hydrochloride	0,1g	0,1g
	Vitamine B12 cyanocobalamine	0,005g	0,005g
	Eau distillée	50mL	50mL

Tableau 4 : Composition du milieu Conway en fonction du type de cellule cultivé

3 minutes, puis le culot cellulaire est repris avec 1mL de milieu complété. 100 μ L de cette suspension cellulaire sont ensuite déposés dans une nouvelle flasque contenant 12,5mL de milieu F-12 complété.

3.3 Analyses bioinformatiques

3.3.1 Analyse des séquençages

Afin de vérifier d'éventuelles erreurs lors de l'amplification, les séquences nucléotidiques obtenues à l'issue du séquençage des différentes constructions plasmidiques ont été alignées avec la séquence théorique via le programme Clustal Omega du serveur EMBL-EBI (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) en utilisant les paramètres par défaut, ainsi que le logiciel Sequencher 5.0. La séquence nucléotidique obtenue a ensuite été traduite en séquence protéique, puis alignée avec la séquence protéique théorique à l'aide de ces deux programmes en utilisant les paramètres par défaut. Le meilleur clone obtenu, possédant le minimum de différence avec la séquence théorique codée, a été sélectionné pour chaque construction plasmidique et utilisé pour réaliser les transfections.

3.3.2 Analyse des séquences nucléotidiques

La génération des cartes de restriction des inserts et des plasmides a été effectuée à l'aide des logiciels Sequencher 5.0 (Gene codes) ou SerialCloner 2.6.1 (Serial Basics). La traduction en séquences protéiques ont été réalisées avec le logiciel Sequencher 5.0 (Gene codes) ou le programme Expasy Translate tool situé sur le site <http://web.expasy.org/translate/>.

3.3.3 Analyse des séquences protéiques

Les prédictions de domaine transmembranaire ont été réalisées sur les serveurs TMHMM v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>), TM pred (http://www.ch.embnet.org/software/TMPRED_form.html) et DAS (<http://www.sbc.su.se/~miklos/DAS/>). Les prédictions de domaines fonctionnels ont été effectuées avec le programme Pfam (<http://pfam.xfam.org>) du serveur EMBL-EBI.

Les alignements de séquences protéiques ont été réalisés avec le programme Clustal Omega du serveur EMBL-EBI, ou avec le programme T-Coffee (<http://tcoffee.crg.cat>).

3.4 Techniques de biologie moléculaire

3.4.1 Analyse des acides nucléiques

3.4.1.1 Analyse quantitative par spectrophotométrie

La quantification des acides nucléiques a été réalisée à l'aide d'un Nanodrop 2000c (Thermo scientific) qui permet de doser par spectrophotométrie la quantité d'acides nucléiques présents dans l'échantillon en mesurant leur absorbance à 260nm. Les composés phénoliques absorbant à 230nm et les protéines à 280nm, le degré de pureté de l'échantillon est évalué par le calcul des rapports A260/A230 et A260/A280 dont les valeurs doivent être comprises entre 1,8 et 2,0 pour une qualité optimale. Les analyses ont été réalisées à partir d'1µL d'échantillon en utilisant un blanc réalisé avec le tampon dans lequel ont été repris les acides nucléiques.

3.4.1.2 Analyse qualitative par électrophorèse en gel d'agarose

La séparation des fragments d'ADN a été réalisée par électrophorèse en gel d'agarose. Cette technique permet à la fois de séparer les fragments d'acides nucléiques en fonction de leur taille, mais aussi d'évaluer la pureté et la qualité des échantillons. La migration des fragments est basée sur les charges négatives portées par les acides nucléiques à pH=8,3 qui sont attirés par l'anode de la cuve d'électrophorèse avec une vitesse de migration dépendante de leur taille et de la concentration en agarose du gel. Afin de séparer au mieux les fragments, des concentrations en agarose de 1% (p/v) ont été utilisées pour analyser les fragments de 0,2 à 5kb, et des concentrations de 2% pour les fragments plus petits, inférieurs à 0,2kb.

Les gels ont été préparés en utilisant un tampon Tris-Borate-EDTA 1X (TBE ; 80mM Tris Base, 89mM d'acide borique, 2mM EDTA, pH=8,3) ou un tampon Tris-Acetate-EDTA 0,5X (TAE ; 20 mM de Tris-Acetate, 0,5mM EDTA, pH=8,3), auquel a été ajouté de l'agarose en poudre (Fisher) à la concentration souhaitée ainsi que 3,6µL de Roti®-Safe (Carl Roth) pour 100mL de tampon, ou 5µL de SafeView™ (Gentaur) pour 100mL de tampon. Ces deux agents fluorescents vont s'intercaler entre les bases azotées de l'ADN, ce qui permettra lorsqu'ils seront excités sous l'action d'un rayon ultraviolet de détecter les fragments. Les migrations ont été réalisées dans des cuves à électrophorèse i-Mupid ou Mupid-ex (Advance) remplies du tampon qui a servi pour préparer le gel. Du tampon de charge (composé de quelques gouttes de bleu de bromophénol 0,01%, de 600µL de glycérol 50% et de 400µL de TE 1X (10mM Tris-HCl, pH=7,5 ; 1mM EDTA)) a été ajouté aux échantillons d'acides nucléiques avant dépôt des échantillons dans les puits du gel. Parallèlement, 4µL de marqueur de taille Smart Ladder 0,2kb-10kb (Eurogentec), ou 1kb DNA Ladder RTU (GeneDirex) ont été chargés dans le gel afin de pouvoir, après la migration, estimer la taille des fragments d'intérêt par rapport aux marqueurs.

La migration a été réalisée sous une tension constante de 100V et la visualisation des acides nucléiques a été effectuée sous ultraviolet (302nm) à l'aide d'un appareil Imager (Bio-capt, Vilber Lourmat).

3.4.2 Clonage et expression des gènes d'intérêt

3.4.2.1 Extraction de l'ADN génomique

Dans le but d'amplifier les gènes de la FuT46110 et de la FuT46109 putatives, une extraction d'ADN génomique a été réalisée à partir de la souche Pt1.8.6 de *P. tricornutum* en utilisant le kit « Nucleospin Plant II » (Macherey-Nagel). Pour cela, 225mL de microalgues cultivées en erlenmeyer à une densité de 8.10^6 cellules/mL ont été récoltées par centrifugation à 4 000g pendant 3 minutes. Le culot cellulaire a alors été repris dans 400 μ L de solution de lyse puis l'échantillon a été déposé dans un tube « lysing matrix tube E » (MP Biomedicals) contenant des billes en céramique de 1,4 mm, des billes en silice de 0,1mm et une bille en verre de 4mm. Le tube a été placé dans un agitateur FastPrep-24 (MP Biomedicals) et les cellules ont été broyées par 8 cycles de 1 minute à une vitesse de rotation de 6,5 m/s. L'échantillon a ensuite été centrifugé à 1 200g pendant 15 minutes dans le but de culoter les billes et les débris cellulaires. Le surnageant a été récupéré puis l'extraction d'ADN génomique a été effectuée en suivant les instructions du kit commercialisé par Macherey-Nagel. L'ADN a finalement été élué avec 100 μ L d'eau stérile.

Afin de cribler les colonies de *P. tricornutum* transformées avec un gène possédant une étiquette V5, un premier criblage par amplification PCR a été réalisé pour vérifier l'intégration du gène avant de procéder à un criblage par western-blot. Cette PCR ne nécessitant pas un ADN matrice d'une grande pureté, un protocole d'extraction rapide d'ADN de *P. tricornutum* a été utilisé. Pour cela, la colonie a été grattée avec une anse puis resuspendue dans un tube contenant 1,5mL d'H₂O stérile et centrifugé pendant 2 minutes à 7 000g. Le culot a été ensuite lavé avec 1mL d'eau stérile milli-Q pour enlever le sel puis centrifugé une nouvelle fois durant 2 minutes à 7 000g. Le culot a été repris avec 30 μ L d'eau stérile milli-Q puis incubé à 95°C pendant 10min. Les tubes ont été centrifugés pendant 2 minutes à 7 000g et le surnageant contenant l'ADN a été prélevé et déposé dans un nouveau tube propre.

3.4.2.2 Amplification par Polymérase Chain Reaction (PCR)

La PCR est une technique qui permet de synthétiser et d'amplifier de l'ADN *in vitro*. Elle est basée sur l'utilisation d'une enzyme, l'ADN polymérase, qui permet la réplication d'une séquence d'ADN spécifique, déterminée par des amorces oligonucléotidiques qui vont initier la synthèse de la région sélectionnée. L'amplification se fait selon des cycles composés de trois étapes :

- une étape de dénaturation qui permet de séparer les deux brins d'ADN matrice,
- 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 dernière étape d'élongation pendant laquelle une ADN polymérase va synthétiser l'ADN complémentaire à partir de l'extrémité 3'-OH de l'amorce dans le sens 5'-P vers 3'-OH du brin matrice.

L'ADN polymérase sélectionnée pour effectuer les PCR varie en fonction de l'utilisation qui est faite du fragment amplifié. Ainsi, pour réaliser un criblage de transformants, la polymérase GoTaq® G2 (Promega) a été utilisée. Cependant, pour réaliser les clonages, l'apparition de mutations au cours de la réaction doit être limitée. Dans ce but, ont été utilisées des ADN polymérases de haute-fidélité comme la Phusion® High-Fidelity DNA Polymerase (Finnzymes) pour le clonage dans pJet1.2 et la Extensor Hi-Fidelity (Thermo Scientific, AB-0720) pour le clonage dans pcDNA3.1. Le mélange réactionnel de PCR préparé respectivement pour chacune des polymérases et expériences est indiqué dans les **tableaux 5A et 5B**.

A	1 µL	ADN matrice	
	2 µL	Amorce 1 5 µM	0,4 µM
	2 µL	Amorce 2 5 µM	0,4 µM
	5 µL	Tampon HF 5x	1 X
	1 µL	MgCl ₂ 50mM	2mM
	1 µL	dNTP mix 10mM (soit 2,5mM de chaque dNTP)	0,4mM
	0,3 µL	Phusion polymérase (2 U/µL)	0,024U/µL
	12,7 µL	Eau nucléase free	
	25 µL	Total	Conc. finale
B	1 µL	ADN matrice	
	4 µL	Amorce 1 10 µM	0,8 µM
	4 µL	Amorce 2 10 µM	0,8 µM
	5 µL	Tampon 10X High fidelity PCR	1 X
	0,5 µL	dNTP mix 25mM (soit 6,25mM de chaque dNTP)	0,25mM
	0,3 µL	Extensor Hi-Fidelity (2 U/µL)	0,012U/µL
	35,2 µL	Eau nucléase free	
	50 µL	Total	Conc. finale

Tableau 5A : Composition des mélanges réactionnels de PCR.
A : mélange réactionnel pour la polymérase Phusion
B : mélange réactionnel pour la polymérase Extensor Hi-Fidelity

1 µl/2 µL	ADN matrice /ou Culture cellulaire	
2 µL	Amorce 1 5 µM	0,4 µM
2 µL	Amorce 2 5 µM	0,4 µM
5 µL	5X Green GoTaq® Reaction Buffer	1 X
0,25 µL	dNTP mix 25mM (soit 6,25mM de chaque dNTP)	0,25mM
0,12 µL	GoTaq® G2 DNA Polymerase (5 U/µL)	0,024U/µL
14,63 µL/13,63 µL	Eau nucléase free	
25 µL	Total	Conc. finale

Tableau 5B : Composition du mélange réactionnel de PCR pour la polymérase GoTaq®G2.

Ce mélange a été déposé dans des microtubes de 0,2mL, puis homogénéisé avant d'être placé dans un thermocycleur Labcycler (Sensoquest) ou Mastercycler personal (Eppendorf). Les conditions de PCR utilisées sont résumées dans le **tableau 6**. Les témoins négatifs ont été réalisés en remplaçant le volume d'ADN matrice par un volume équivalent d'eau stérile.

Etape	Température (°C)	Durée	Nombre de cycles
Dénaturation initiale	95 / 98 / 94	2 min / 3 min / 2min	1
Dénaturation	95 / 98 / 94	1 min / 30 sec / 30 sec	35
Hybridation	X	1 min / 20 sec / 30 sec	
Extension	72 / 72 / 68	3 min / 1 min 40 / 2min	
Extension finale	72 / 72 / 68	5 min / 4 min / 7min	1

Tableau 6 : Conditions de PCR en fonction de la polymérase utilisée. GoTaq®G2 ; Phusion High-Fidelity ; Extensor Hi-Fidelity

3.4.2.3 Purification du produit PCR

La purification des produits PCR a été réalisée directement en solution ou après électrophorèse à partir d'un gel d'agarose en utilisant les kits «Wizard SV Gel and PCR Clean-Up System » (Promega, A9282) ou «JETSORB Gel Extraction Kit » (Genomed, 110600). Dans les deux cas, l'utilisation de ces kits a été effectuée en suivant les instructions des fabricants respectifs.

3.4.2.4 Ligation et transformation bactérienne

La ligation du produit PCR dans le vecteur pcDNA3.1/V5-His-TOPO® et sa transformation dans des cellules thermocompétentes « one shot top 10 chemically competent *E. coli* » ont été réalisées avec le kit « pcDNA3.1/V5-His-TOPO TA expression kit » (Invitrogen, K800-01) en suivant le protocole du fournisseur. Le mélange de ligation utilisé pour chaque construction génétique contient 4µL de produit PCR pour 1µL (10ng) de vecteur comme préconisé dans la fiche technique du fournisseur.

Concernant la ligation du produit PCR dans le vecteur pJET1.2/blunt, elle a été réalisée avec le kit « CloneJET PCR Cloning Kit » (Thermo Scientific, K1231). Le mélange de ligation réalisé est présenté dans le **tableau 7**. Le mélange a ensuite été incubé pendant 1 heure à température ambiante. La transformation de la construction a été réalisée dans des cellules d'*E. coli* thermocompétentes Top 10 one shot (Thermofisher ; 11573117) ou Library efficiency DH5α (Thermofisher ; 10666493) en suivant le protocole du fournisseur.

Volume	Solution
10µL	Produit PCR purifié
0,5µL	pJET1.2/blunt Cloning Vector (50 ng/µL)
0,5µL	T4 DNA Ligase (5 U/µL)
1,5µL	Tampon réactionnel 10X
2,5µL	Eau nucléase free
15µL	Total

Tableau 7 : Mélange de ligation réalisé pour le clonage dans le vecteur pJET1.2/blunt

La ligation de fragments d'ADN dans le vecteur pPhaNR est réalisée grâce à l'hybridation d'extrémités cohésives. Ces dernières sont générées par des digestions enzymatiques en utilisant les enzymes de restriction (Thermo Scientific) listées dans le **tableau 9** en fonction du fragment excisé considéré, et dans les conditions présentées dans le **tableau 8**.

Volume	Solution
2µL	ADN plasmidique
0,4µL	Enzyme de restriction 1 (Thermo Scientific)
0,4µL	Enzyme de restriction 2 (Thermo Scientific)
1µL ou 2µL (1X ou 2X)	Tampon Tango 10X (Thermo Scientific)
QSP 10µL	Eau nucléase free

Tableau 8 : Mélange réactionnel utilisé pour la digestion des fragments d'ADN par les enzymes de restriction.

Fragment	Enzymes	Tampon Tango
FuT54599-V5	<i>SacI</i> / <i>HindIII</i>	Tango 1X
FuT54599 pour GFP	<i>SacI</i> / <i>BglII</i>	Tango 2X
FuT46109NM-V5	<i>SacI</i> / <i>Sall</i>	Tango 2X
FuT46109NM pour GFP	<i>SacI</i> / <i>HindIII</i>	Tango 1X
FuT46109OM pour GFP	<i>SacI</i> / <i>KpnI</i>	Tango 1X
FuT46110NM-V5	<i>SacI</i> / <i>KpnI</i>	Tango 1X
FuT46110NM pour GFP	<i>SacI</i> / <i>BamHI</i>	Tango 1X
FuT46110OM pour GFP	<i>SacI</i> / <i>BamHI</i>	Tango 1X
GnTI-V5	<i>MunI</i> / <i>HindIII</i>	Tango 1X
GnTI pour GFP	<i>MunI</i> / <i>BamHI</i>	Tango 1X
GalT-HA	<i>EcoRI</i> / <i>HindIII</i>	Tango 1X
GalT pour GFP	<i>EcoRI</i> / <i>BamHI</i>	Tango 2X
eGFP	<i>BamHI</i> / <i>HindIII</i>	Tango 1X
eGFP	<i>KpnI</i> / <i>HindIII</i>	Tango 1X
pPhaNR	<i>SacI</i> / <i>HindIII</i>	Tango 1X
pPha-NR	<i>EcoRI</i> / <i>HindIII</i>	Tango 1X
pPha-NR	<i>SacI</i> / <i>Sall</i>	Tango 2X
pPha-NR	<i>SacI</i> / <i>HindIII</i>	Tango 1X
pPha-NR	<i>SacI</i> / <i>XbaI</i>	Tango 1X
pPha-NR	<i>EcoRI</i> / <i>HindIII</i>	Tango 1X

Tableau 9 : Couples d'enzymes de restrictions utilisées pour extraire le fragment d'ADN considéré.

Ainsi dans un premier temps, les fragments d'intérêt ont été amplifiés par PCR avec des amorces permettant l'insertion de sites de restriction aux extrémités du fragment. Le choix des sites de restriction s'est basé sur leur disponibilité dans la cassette de clonage du vecteur, ainsi que sur leur absence dans la séquence de l'insert considéré. Les fragments ont ensuite été clonés dans le vecteur pJET1.2/blunt. Ce clonage préalable permet de faciliter ensuite la digestion par les enzymes de restriction qui sont des endonucléases. Un schéma présentant la stratégie employée est présenté sur la **figure 40**.

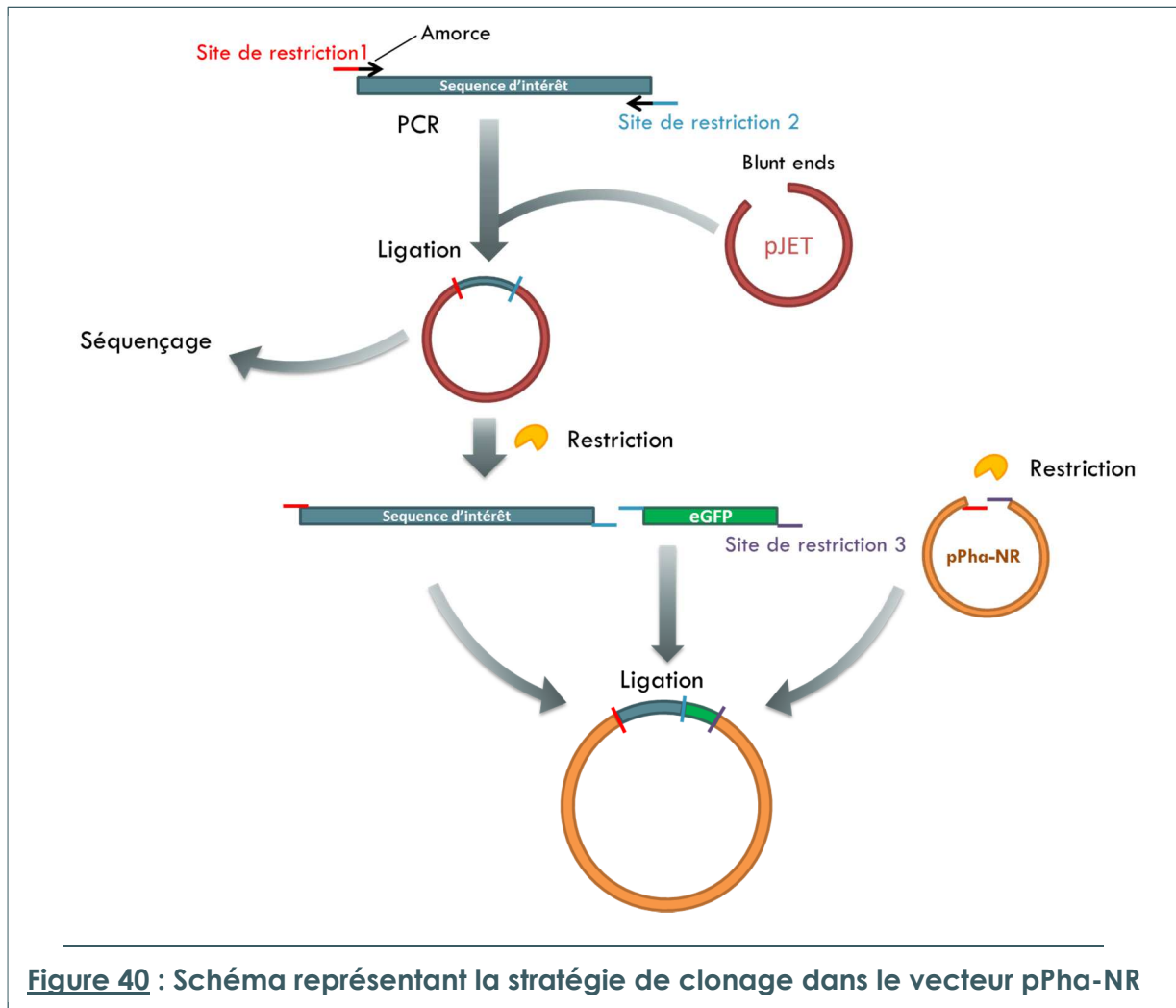


Figure 40 : Schéma représentant la stratégie de clonage dans le vecteur pPha-NR

Les digestions ont été réalisées pendant 4h à 37°C. La ligation des fragments digérés dans le vecteur pPhaNR a été réalisée en incubant 1 heure à température ambiante un milieu réactionnel composé de 10µL de fragments digérés et purifiés de l'insert et du vecteur avec 0,5µL de ligase T4 (5U/µL ; Thermo Scientific ; EL0014) et 1µL de tampon 10X (Thermo Scientific ; B69). Le produit de ligation a ensuite été transfecté dans des cellules thermocompétentes d'*E. coli* Top 10.

Après transformation, les bactéries ont été étalées sur une boîte de Pétri contenant du milieu LB gélosé (Sigma) auquel ont été ajoutés 50µg/mL d'ampicilline (Duchefa). La boîte a été incubée à 37°C pendant une nuit.

3.4.2.5 Extraction d'ADN plasmidique

Plusieurs colonies isolées obtenues suite à la transformation bactérienne ont été sélectionnées, puis repiquées dans 5mL de milieu LB liquide (Sigma) contenant 50µg/mL d'ampicilline (Duchefa) et placées à 37°C sous agitation à 180rpm. L'extraction et la

purification d'ADN plasmidique a été réalisée à partir de 3mL de préculture avec le kit « Pure Yield™ Plasmid Midiprep System » (Promega, A1222) ou avec le protocole « maison » décrit ci-après. Ces deux protocoles de purification reposent tous deux sur la méthode de la lyse alcaline des cellules, seule la méthode de purification varie. Ainsi, le kit propose une purification de l'ADN sur colonne de silice alors que le protocole « maison » utilise une précipitation alcoolique.

Afin de réaliser le protocole « maison », trois solutions ont été préparées comme présenté dans le **tableau 10**. Chaque culture a été répartie dans deux tubes de 1,5mL centrifugés ensuite pendant 1min à 16 100g pour culoter les cellules. Le culot a ensuite été resuspendu avec 200µL de P1, puis incubé durant 1min à température ambiante. 200µL de P2 ont ensuite été ajoutés, puis les tubes ont été homogénéisés par retournements et incubés pendant 5min à température ambiante. Suite à cela, 200µL de P3 ainsi que 25µL de chloroforme ont été ajoutés, puis le mélange a été homogénéisé par retournements et incubé pendant 5min sur de la glace. Les tubes ont été centrifugés à 16 100g pendant 10min à 4°C. Le surnageant a ensuite été déposé dans un nouveau tube avec 400µL d'isopropanol. Les tubes ont été mélangés, puis centrifugés à 16 100g pendant 20min à 4°C. Le surnageant a été jeté et le culot a été lavé avec 1mL d'éthanol 70%. Les tubes ont été une nouvelle fois centrifugés à 16 100g pendant 5min à 4°C puis le surnageant a été jeté et les culots séchés sous vide. Les plasmides ainsi purifiés ont été resuspendus avec 50µL d'eau « nuclease-free ».

Tampon P1 (resuspension)	Tris HCl (pH=8.0) EDTA RNase A	50mM 10mM 100µg/mL
Tampon P2 (lyse)	NaOH SDS	200mM 1% (w/v)
Tampon P3 (neutralisation)	KAc (pH=5,5)	3M

Tableau 10 : Composition des solutions utilisées dans le protocole d'extraction plasmidique « maison »

Lorsque des grandes quantités de plasmides purifiés sont nécessaires, 200µL de préculture bactérienne ont été inoculés dans 100mL de milieu LB liquide contenant 50µg/mL d'ampicilline et placées à 37°C sous agitation à 180rpm. L'extraction et la purification d'ADN plasmidique ont été réalisées sur l'ensemble de la culture avec le kit NucleoBond® Xtra Midi Plus (Macherey-Nagel, 740412). Ce kit utilise aussi la méthode de la lyse alcaline des cellules associée à une technique de purification par filtration, puis par chromatographie sur colonne de silice. Les plasmides ainsi purifiés ont été resuspendus avec 400µL d'eau « nuclease-free ».

3.4.2.6 Contrôle des clones sélectionnés

3.4.2.6.1 Sélection des clones positifs par digestion enzymatique

Une digestion enzymatique a été réalisée sur les extraits d'ADN plasmidique afin de sélectionner les clones positifs contenant le gène d'intérêt. Cette étape est d'autant plus importante pour le clonage dans pcDNA3.1 qui n'est pas orienté. En effet, le fragment d'intérêt pouvant s'insérer dans le vecteur dans les deux sens, la digestion permet de vérifier l'orientation correcte de l'insert. Pour cela, les constructions réalisées pcDNA3.1 + FuT46109 et pcDNA3.1 + FuT46110 ont été digérées avec l'enzyme de restriction *HindIII*, alors que la construction pcDNA3.1 + FuT54599 a été digérée avec *KpnI*. La composition du mélange réactionnel utilisé pour ces digestions est indiquée dans le **tableau 11**. En ce qui concerne les digestions contrôles des constructions avec les vecteurs pJET et pPha-NR, les digestions ont été réalisées comme présenté dans les **tableaux 8 et 9**. Une fois le mélange réalisé, les échantillons ont été délicatement homogénéisés puis incubés à 37°C pendant 4 heures. Le résultat des digestions enzymatiques a ensuite été visualisé par migration électrophorétique en gel d'agarose selon les conditions décrites précédemment.

Volume	Solution
2µL	ADN plasmidique
2µL	Tampon 10x E ou J (Promega)
0,2µL	BSA acétylée (10µg/µL)
0,5µL	Enzyme de restriction <i>HindIII</i> ou <i>KpnI</i> (10U/µL)
15,3µL	Eau nucléase free
20µL	Total

Tableau 11 : Composition du mélange de restriction utilisé pour le criblage des vecteurs pcDNA3.1

3.4.2.6.2 Séquençage

Dans le but de vérifier l'absence de mutations dans les inserts clonés dans les plasmides recombinants, ainsi que la spécificité des produits d'amplification PCR (expériences de qRT-PCR), les échantillons d'ADN ont été séquencés. Pour cela, nous avons fait appel à un prestataire de service, GATC Biotech. Les échantillons (ADN plasmidique ou produit d'amplification PCR) ont ainsi été préparés et envoyés selon les recommandations du prestataire, puis le séquençage des échantillons d'acides nucléiques a été réalisé selon la

méthode de Sanger par la société GATC Biotech (<https://www.gatc-biotech.com/fr/index.html>). Pour cela, des amorces flanquant la séquence d'intérêt ont été utilisées, ainsi qu'éventuellement une (ou des) amorce(s) interne(s) pour les séquences les plus longues.

3.4.2.7 Linéarisation du vecteur recombinant

Afin d'optimiser l'intégration des vecteurs recombinants dans le génome de la cellule de mammifère, une linéarisation préalable des plasmides a été réalisée. Pour cela, les différents plasmides purifiés ont été digérés par une enzyme de restriction possédant un site de coupure unique dans le plasmide: *Hind*III (Promega) pour les constructions pcDNA3.1 + FuT46109 et pcDNA3.1 + FuT46110, ou *Kpn*I (Promega) pour la construction pcDNA3.1 + FuT54599, dans les conditions présentées dans le **tableau 12**. Le mélange réactionnel a été incubé à 37°C pendant 4 heures. La linéarisation des vecteurs a été vérifiée par migration d'une aliquote du produit de digestion en gel d'agarose. La purification des vecteurs linéarisés a ensuite été réalisée avec le kit « Wizard SV Gel and PCR Clean-Up System » (Promega, A9282) en suivant les consignes du fournisseur.

Construction	ADN plasmidique	Tampon (promega)	Enzyme (10U/μL)	Eau stérile
pcDNA3.1+46109NM-V5	30μg	20μL tampon K	10μL Sca I	QSP 200μL
pcDNA3.1+46110NM-V5	30μg	20μL tampon D	10μL Bgl II	QSP 200μL
pcDNA3.1+54599-V5	30μg	20μL tampon D	10μL Bgl II	QSP 200μL

Tableau 12 : Composition du mélange de restriction utilisé pour la linéarisation des vecteurs pcDNA3.1

3.4.2.8 Transfection des cellules P.t1.8.6

La transfection stable des différents vecteurs pPha-NR recombinants dans les cellules de la souche P.t1.8.6 de *P. tricornutum* a été réalisée par biolistique avec un PDS-1000/He™ (Bio-Rad) utilisant l'hélium comme gaz vecteur. La première étape de ce procédé consiste à enrober des billes de tungstène avec le vecteur recombinant à transférer. Pour cela, 50µL de suspension de billes M10 de 0,7µm (Bio-Rad) ont été mélangées avec 5µg d'ADN plasmidique, 50µL de CaCl₂ 2,5M et 20µL de spermidine 0,1M (Acros Organics ; 132740010). Le mélange a été homogénéisé au vortex puis incubé pendant 10 minutes à température ambiante. Le culot de billes sédimentées a été repris ensuite avec 250µL d'éthanol 100% puis mélangé au vortex 30sec et incubé durant 10min à température ambiante. Le culot de billes a ensuite été repris avec 50µL d'éthanol 100%. Les particules ainsi préparées ont été utilisées dans l'heure pour la transfection afin d'éviter toute détérioration de l'enrobage.

Au préalable de la transfection, 24 heures à l'avance ont été étalées l'équivalent de 1×10^8 cellules de P.t1.8.6 par boîte de milieu gélosé. Trois boîtes ont été préparées par construction afin de pouvoir réaliser trois bombardements. Pour chaque boîte, 10µL de suspension de particules enrobées ont été utilisés. Après bombardement, les boîtes de Pétri ont été placées à 19-20°C pendant 24h, soit sous lumière continue à une intensité de 80µmol/m²/s, soit avec une photopériode 16h-8h à une intensité de 50µmol/m²/s. Les étalements ont ensuite été repris chacun avec 1mL de milieu de culture puis répartis sur trois nouvelles boîtes de Pétri contenant du milieu gélosé avec de la zéocine à 75µg/mL (Invitrogen ; R25001), ainsi que du NH₄⁺ au lieu du NO₃⁻ comme source d'azote pour ne pas induire le promoteur nitrate réductase. Les boîtes de Pétri ont été placées en chambre de culture jusqu'à l'apparition de colonies, soit environ 3 semaines. Ces dernières ont ensuite été repiquées sur du milieu contenant de la zéocine et du NH₄⁺ pour la conservation, et en parallèle sur du milieu contenant de la zéocine et du NO₃⁻ pour pouvoir exprimer la protéine recombinante et cribler les clones positifs par des approches de microscopie confocale ou de western-blot en fonction de l'étiquette portée (protéine fluorescente ou étiquette peptidique).

3.4.3 Analyse différentielle du transcriptome des trois morphotypes de *P. tricornutum* par RNA sequencing (RNA-Seq)

3.4.3.1 Extraction des ARN totaux

Afin de procéder à l'étude comparative des trois morphotypes de *P. tricornutum* par analyse transcriptomique à haut débit, quatre réplicats biologiques ont été réalisés pour chacun des trois morphotypes : fusiforme, ovale et triradié. Pour réaliser l'extraction des ARN totaux à partir de chaque culture, 100mL de suspension cellulaire ont été prélevées en milieu de phase exponentielle (8 jours) à une densité d'environ $8 \cdot 10^6$ cellules/mL, puis centrifugées à 10 000g pendant 3min à 19°C, de sorte à récolter les microalgues le plus vite possible dans le but de limiter l'expression de gènes liés au stress de la centrifugation. Les culots cellulaires ont été immédiatement repris dans 1mL de Trizol (Life Technologies) puis déposés dans des tubes de «lysing matrix E » (MP Biomedicals, référence n° 116914050) et congelés dans de l'azote liquide. Les échantillons ainsi préparés ont été conservés à -80°C en attendant de procéder à l'extraction.

Pour lyser les cellules, les échantillons ont ensuite été placés dans l'appareil FastPrep-24 (MP Biomedicals) afin de subir 4 cycles de 30sec à 6,5m/s. Entre chaque cycle, les tubes ont été conservés dans de la glace pour limiter les échauffements. Les tubes ont ensuite été incubés à température ambiante pendant 5min, puis centrifugés pendant 5min à 14 000g afin de culoter les billes et les débris. Les surnageants ont ensuite été déposés dans des nouveaux microtubes de 2mL « RNase-free ». 200µL de chloroforme ont été ajoutés dans chaque tube, puis les échantillons ont été mélangés par retournements. Les tubes ont été incubés à température ambiante pendant 3min, puis centrifugés à 12 000g pendant 15min à 4°C. Les phases aqueuses supérieures ont été ensuite déposées dans un nouveau tube « RNase-free » contenant le même volume d'éthanol 70% (400µL). Les tubes ont été homogénéisés puis l'échantillon a été déposé sur une colonne NucleoSpin RNA du kit NucleoSpin® RNA II (Macherey-Nagel ; 740955.50). La purification des ARN a ensuite été réalisée selon le protocole du kit, puis les ARN totaux ont été élués avec 60µL d'eau « RNase-free ».

3.4.3.2 Analyse qualitative par Bioanalyseur

La qualité des échantillons d'ARN destinés à l'analyse transcriptomique a été évaluée à l'aide d'un instrument Bioanalyzer 2100 (Agilent Technologies). Ce système microfluidique permet de réaliser une électrophorèse en gel à travers un fin capillaire, ce qui permet à l'aide d'une sonde fluorescente d'analyser de très faibles quantités d'échantillon. Cet outil permet de

générer un électrophérogramme et ainsi de visualiser l'état de dégradation des ARN. L'analyse de nos échantillons a été réalisée avec le kit RNA 6000 Pico (Agilent Technologies) selon le protocole du fournisseur.

3.4.3.3 Construction de la librairie d'ADNc

La librairie a été préparée selon le protocole du kit "TruSeq Stranded mRNA Sample" (Illumina ; RS-122-2101) dont les différentes étapes sont résumées dans le schéma de la **figure 41**. Le principe général est présenté si après avec des détails sur les étapes sujettes à variation. Le protocole complet est joint en annexe 2 de ce manuscrit.

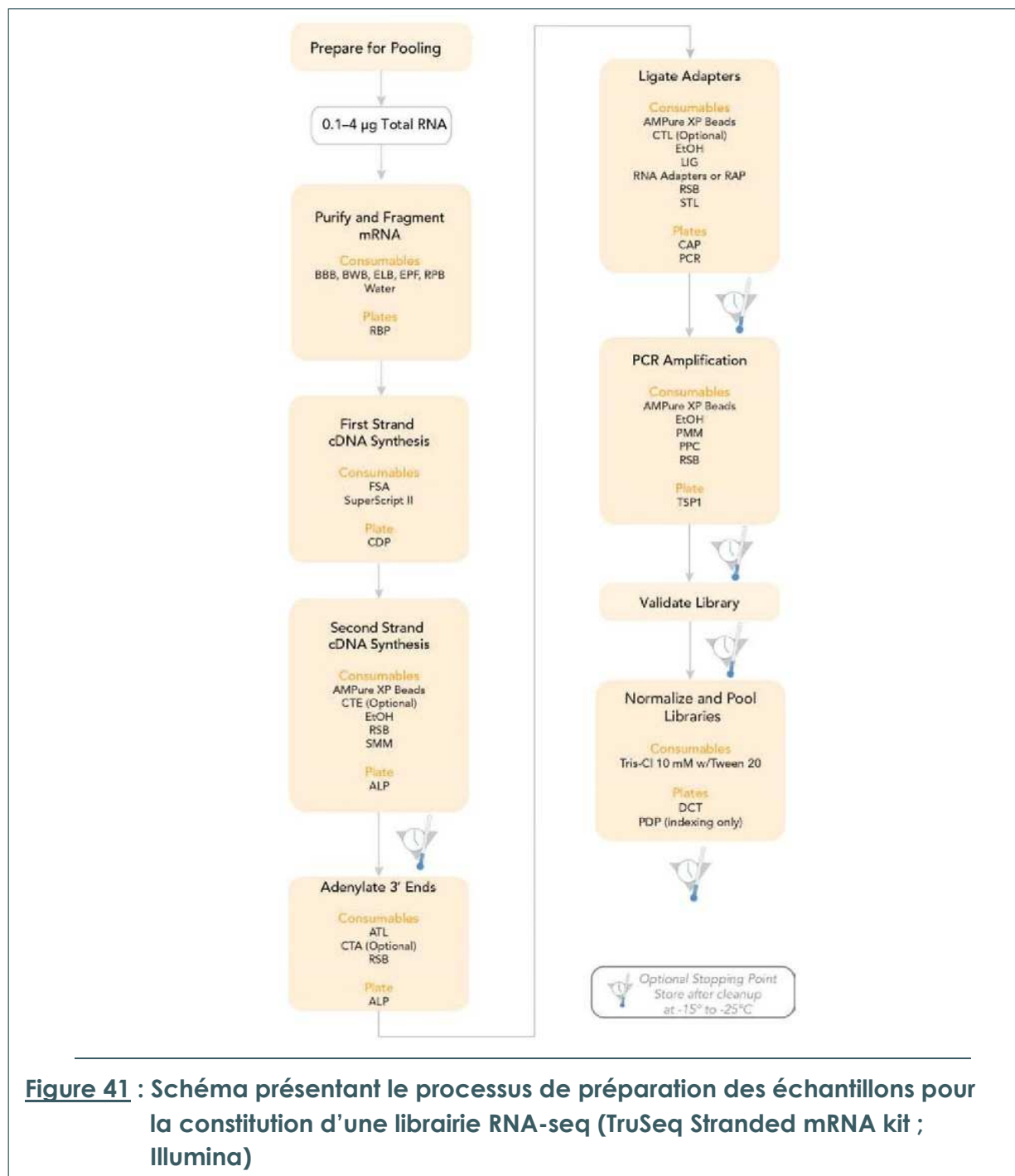
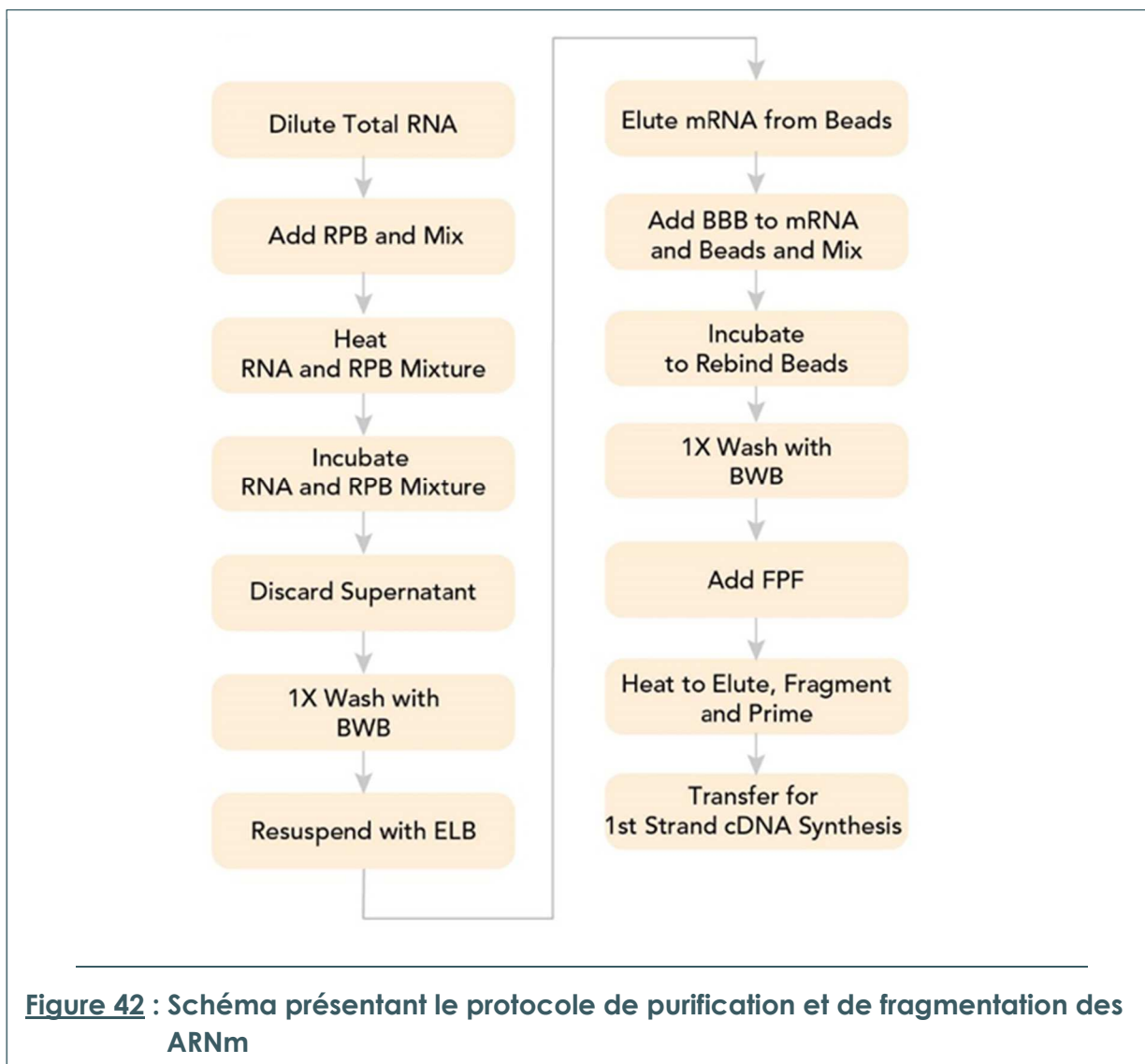


Figure 41 : Schéma présentant le processus de préparation des échantillons pour la constitution d'une librairie RNA-seq (TruSeq Stranded mRNA kit ; Illumina)

3.4.3.3.1 Purification et fragmentation des ARNm

Les extraits d'ARN totaux ont été dilués dans de l'eau « RNase-free » de façon à obtenir des échantillons de 50µL contenant chacun 4µg d'ARN. Les ARN messagers ont ensuite été purifiés et fragmentés en suivant le protocole présenté dans **la figure 42**. La première de ces étapes consiste à ajouter des billes magnétiques enrobées d'oligo dT afin de capturer les ARNm par leur queue polyA. Les échantillons ont ensuite été placés dans un thermocycleur afin de dénaturer les ARN et de faciliter la liaison des ARN polyA aux billes (incubation à 65°C pendant 5min). Les billes ayant capturé les ARNm ont été lavées deux fois avec la solution BWB (Bead Washing Buffer) à l'aide d'un portoir magnétique afin d'éliminer les ARN ribosomaux ainsi que les autres ARN non messagers. Les ARNm purifiés ont été ensuite mélangés avec la solution FPF (Fragment, Prime, Finish Mix) afin de les fragmenter (par incubation à 94°C pendant 8min) et d'ajouter des amorces « random hexamers » qui serviront lors de l'étape de « reverse- transcription » qui suivra.



3.4.3.3.2 Synthèse du premier brin d'ADN complémentaire

La synthèse du premier brin d'ADNc a été réalisée en ajoutant à 17µL d'échantillon d'ARNm fragmenté résultant de l'étape précédente 8µL d'un mélange réalisé avec un ratio de 1 volume de SuperScript II Reverse Transcriptase (Invitrogen) pour 9 volumes de "First Strand Synthesis Act D Mix". Les échantillons ont ensuite été déposés dans le thermocycleur et incubés selon le programme décrit dans le protocole du fournisseur.

3.4.3.3.3 Synthèse du second brin d'ADN complémentaire

Cette étape va permettre de générer un ADNc double brin en éliminant le brin d'ARN matrice. Ce dernier sera remplacé par un brin d'ADNc qui aura incorporé de l'UMP à la place de TMP. Ce procédé permettra de réaliser un séquençage brin spécifique et ainsi de savoir à partir de quel brin d'ADN provient l'ARN transcrit.

3.4.3.3.4 Adénylation des extrémités 3'

Suite à la synthèse du second brin d'ADNc, les fragments sont pourvus d'extrémités franches. Une étape d'adénylation à l'extrémité 3' des fragments va permettre non seulement l'hybridation subséquente des adaptateurs via leur extrémité complémentaire T dépassante, mais également d'éviter que les fragments se lient entre eux lors de la réaction de ligation des adaptateurs.

3.4.3.3.5 Ligation des adaptateurs

L'ajout d'adaptateurs va permettre aux fragments d'ADNc de s'hybrider sur la flowcell afin de créer des clusters lors du séquençage. La ligation des adaptateurs est obtenue par hybridation A/T comme présenté précédemment. Les adaptateurs sont munis d'index, composés d'une courte séquence nucléotidique, qui vont permettre le multiplexage des échantillons. Les index utilisés pour chaque échantillon, ainsi que le plan de charge sur la flowcell sont représentés dans le **tableau 13**.

Pistes de la Flowcell	2	3	4	5	6	7
Index 2 [AR002 ; seq : CGATGT(A)]	F1	O1	T1	O2	T2	F2
Index 4 [AR004 ; seq : TGACCA(A)]	O3	T3	F3	F4	O4	T4

Tableau 13 : Plan de charge des échantillons indexés sur la flowcell

F = Fusiforme ; O = Ovale ; T = Triradié ; n = numéro de réplicat biologique

3.4.3.3.6 Enrichissement par PCR

Cette PCR permet d'augmenter la quantité d'ADN de la librairie en amplifiant sélectivement les fragments d'ADNc possédant des adaptateurs fixés à chacune de leurs extrémités via des amorces complémentaires de la séquence nucléotidique des adaptateurs. Le nombre de cycles PCR doit être limité à cette étape afin d'éviter de biaiser la représentation des fragments dans la librairie. Les conditions de PCR utilisées ont été réalisées selon les préconisations du fournisseur (15 cycles).

3.4.3.3.7 Validation de la librairie

3.4.3.3.7.1 Contrôle qualité par Bioanalyser 2200 TapeStation

L'analyse des échantillons par TapeStation 2200 (Agilent Technologies) permet de s'assurer que la librairie est composée d'une seule population de fragments aux alentours de 250-260pb. De plus, elle permet de donner un aperçu quantitatif. L'analyse des échantillons a été réalisée selon le protocole du fournisseur.

3.4.3.3.7.2 Quantification par qPCR

Cette étape permet de réaliser un dosage plus précis des librairies afin de déposer chaque échantillon de façon équimolaire sur la flowcell, sachant qu'une librairie doit avoir au moins une concentration de 10nM pour pouvoir être ultérieurement séquencée efficacement. Les mesures ont été réalisées sur deux réplicats dilués au 1/10 000^{ème} (une première dilution au 1/100^{ème} [1µL échantillon + 99µL H₂O] qui est diluée à son tour au 1/100^{ème} [10µL échantillon + 990µL H₂O]), puis sur deux réplicats dilués au 1/20 000^{ème} (dilution au 1/200^{ème} à partir de la dilution au 1/100^{ème} [5µL échantillon + 995µL H₂O]). La PCR quantitative en temps réel a été réalisée avec le kit « KAPA SYBR® FAST qPCR » (Kapa Biosystems ; KK4601) selon le mode opératoire utilisé à la plate-forme génomique de l'IRIB. Les différents échantillons, le blanc sans ADN matrice, et les six standards ont été analysés en duplicats. Pour cela, un mélange réactionnel a été préparé avec l'équivalent par puits de 10µL de KAPA SYBR FAST qPCR, 2µL de Premix Primer Illumina, 0,4µL de Rox Reference Dye High, et 3,6µL d'eau « nuclease-free ». Les amorces utilisées, correspondant aux séquences des adaptateurs, sont présentées dans le **tableau 14**.

Nom de l'amorce	Séquence nucléotidique 5'-3'
Primer P1	AATGATACGGCGACCACCGA
Primer P2	CAAGCAGAAGACGGCATACGA

Tableau 14 : Amorces utilisées pour quantifier la librairie par qPCR

Un volume de 16µL de mélange réactionnel a ensuite été déposé dans chaque puits avec 4µL d'échantillon ou de standard. La plaque a ensuite été scellée avec un « Optical Adhesive Film », puis centrifugée pendant 1min à 12 000rpm. Le dosage de la plaque a été effectué avec un appareil 7300 Fast Real-Time PCR System (Applied Biosystems) programmé comme présenté dans le **tableau 15**.

Etape	Température (°C)	Durée	Nombre de cycles
Dénaturation initiale	95	5 min	1
Dénaturation	95	30 sec	40
Hybridation/Extension	60	45 sec	
Dissociation	95	15 sec	1
	60	30 sec	
	95	15 sec	

Tableau 15 : Programme du thermocycleur utilisé pour quantifier la librairie par qPCR

Afin d'estimer la quantité d'ADNc présente dans les librairies, la concentration a dans un premier temps été calculée en nM avec la formule suivante :

$$\frac{\text{quantité de produit en pM } 1^{\text{ère}} + 2^{\text{ème}} \text{ mesure}}{2} \times (452/300) \times 10\,000 \div 1000$$

Avec

- 452/300 correspondant à un ajustement lié à la différence de taille entre les ADN Standards par rapport à la taille des ADN dosés.
- X10 000 car dilution au 1/10 000^{ème}
- ÷ 1000 pour passer de pM en nM

La concentration en ng/µL a ensuite pu être calculée selon la formule :

$$\frac{(300 \times 660) \times \text{concentration en nM}}{1\,000\,000}$$

Avec

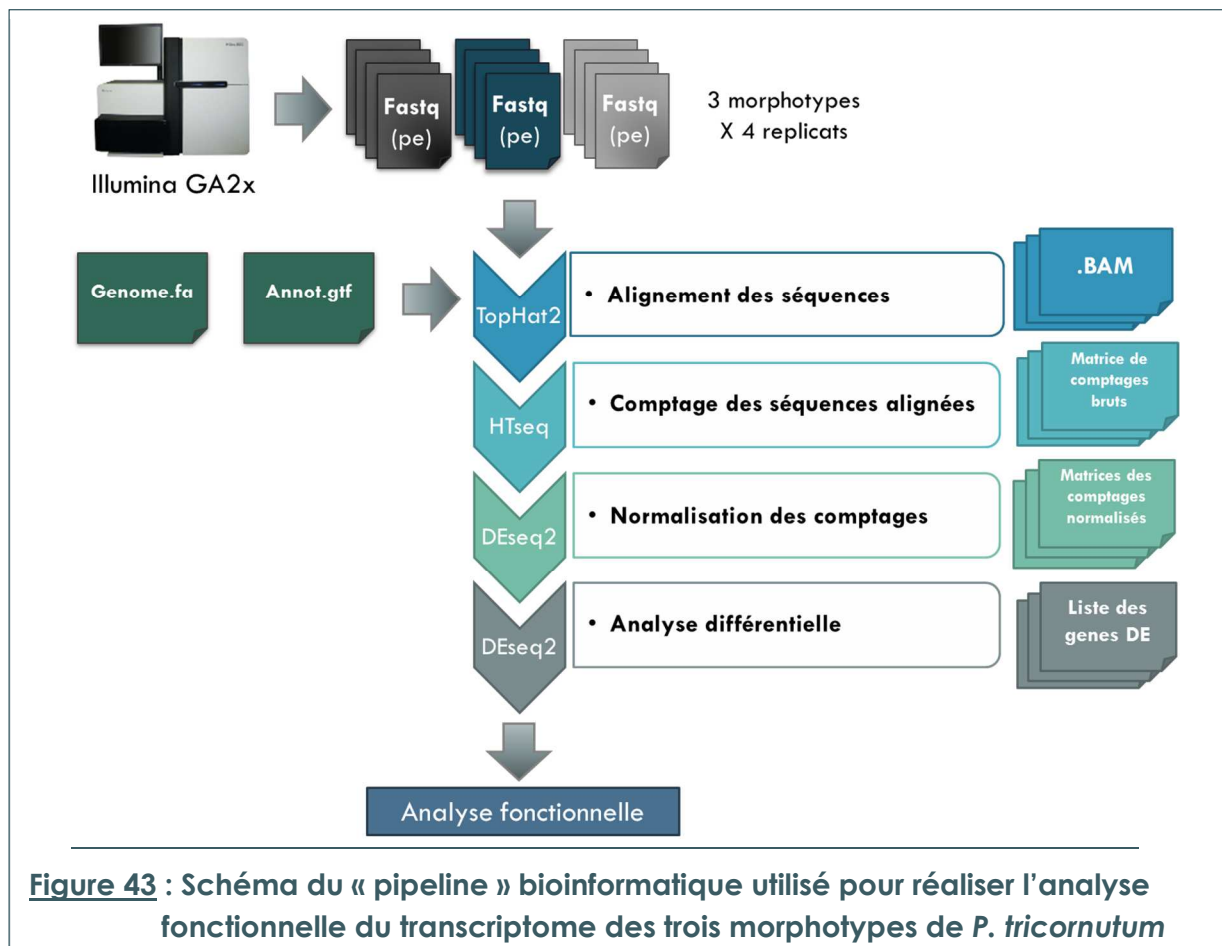
- la valeur 300x660 correspondant au nombre de nucléotides des fragments multiplié par la masse molaire d'une paire de nucléotide (2x330g/mol).
- ÷ 1 000 000 pour passer du L au µL

3.4.3.3.7.3 Séquençage haut-débit des bibliothèques

Les bibliothèques indexées ont été normalisées et rassemblées deux par deux selon le **tableau 13** présenté précédemment. Les échantillons ainsi déposés sur la flowcell ont été séquençés en pair-end de 75pb avec un séquenceur Illumina GAIIx selon le protocole du fournisseur. Les fichiers de sortie bruts ont été traités avec le logiciel « Illumina Real Time Analysis » (RTA 1.9). Le démultiplexage a été réalisé en utilisant CASAVA 1.8.2 (Illumina).

3.4.3.3.7.4 Analyse différentielle des séquences

Afin de comparer entre elles les données d'expression des trois morphotypes de *P. tricornutum*, un « pipeline » bioinformatique a été élaboré comme présenté sur la **figure 43**. Ainsi, les différentes séquences générées ont tout d'abord été alignées sur le génome séquencé (Bowler et al., 2008) et annoté de *P. tricornutum* (http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index) en utilisant le programme TopHat2 (Kim et al., 2013 ; <https://ccb.jhu.edu/software/tophat/index.shtml>), puis les séquences ainsi placées ont été comptées en fonction de leur position sur le génome avec HTSeq-count (Anders et al., 2015 ; <http://www-huber.embl.de/users/anders/HTSeq/doc/count.html>). Les comptages bruts ont ensuite été normalisés, puis comparés, avec le programme DESeq2 (Love et al., 2014 ; <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>).



3.4.3.3.7.5 Analyse fonctionnelle

Dans le but de réaliser l'analyse fonctionnelle des gènes différentiellement exprimés entre les trois morphotypes, un enrichissement fonctionnel a été réalisé. Pour cela, la séquence protéique de chaque gène différentiellement exprimé a été récupérée via l'outil BioMart du site EnsemblProtists (<http://protists.ensembl.org/index.html>). Les séquences ont ensuite été traitées avec le logiciel Blast2GO (BioBam Bioinformatics) en utilisant les paramètres standards. Ce logiciel permet d'attribuer des gene ontology (GO) aux protéines en comparant leur séquence à celles de différentes bases de données.

3.4.3.4 Validation des données de RNA-Seq par PCR quantitative

3.4.3.4.1 Reverse Transcription qPCR (RT-qPCR)

Les données d'expression générées par l'approche RNA-Seq ont été validées en utilisant une approche différente de type RT-qPCR. Dans ce but, une sélection de gènes a été analysée par qPCR afin de vérifier les niveaux d'expression obtenus en RNA-Seq. Pour cela, les mêmes extraits d'ARN que ceux ayant servi pour préparer les banques d'ADNc pour l'analyse en RNA-Seq ont été rétro-transcrits en ADNc en utilisant le kit « Highcap cDNA reverse transcription » (Applied Biosystems ; 4368814). La réaction de reverse transcription a été réalisée à partir de 1 µg d'ARNs totaux. Les échantillons d'ARN ont été dilués avec de l'eau « nuclease-free » pour atteindre un volume final de 10µL. Ces préparations ont ensuite été mélangées chacune avec 10µL de master mix préparé comme présenté dans le **tableau 16**, puis les mélanges réactionnels ont été placés dans un thermocycleur (Veriti, Applied Biosystems) programmé comme indiqué dans le **tableau 17**.

Solutions	Echantillons	Témoins négatifs
10x RT buffer	2µL	2µL
25X dNTP Mix (100 mM)	0,8µL	0,8µL
10X RT Random Primers	2µL	2µL
MultiScribe™ Reverse Transcriptase, 50 U/µL	1µL	0µL
RNase Inhibitor	1µL	1µL
Nuclease-free H2O	3,2µL	4,2µL
Total	10µL	10µL

Tableau 16 : Mélanges réactionnels utilisés pour la reverse transcription

Programmation	Etape 1	Etape 2	Etape 3
Température	25°C	37°C	85°C
Temps	10 minutes	120 minutes	5 minutes

Tableau 17 : Programme du thermocycleur utilisé pour la reverse transcription

3.4.3.4.2 Contrôle qualité par amplification PCR

Une amplification par PCR des témoins négatifs de reverse-transcription a été réalisée afin de vérifier l'absence de contamination des échantillons par de l'ADN génomique. Pour cela, l'amplification d'un gène domestique, celui codant pour l'actine 1, a été réalisé. Le mélange réactionnel a été préparé dans des microtubes comme indiqué dans le **tableau 18**, puis placé dans un thermocycleur (Veriti, Applied Biosystems). Les conditions de PCR utilisées sont indiquées dans le **tableau 19**. La PCR a été réalisée avec des amorces dont les séquences sont présentées dans le **tableau 20**, permettant d'amplifier spécifiquement le gène de l'actine 1. Au cours de cette amplification un témoin positif a été réalisé en utilisant comme matrice de l'ADN génomique de *P. tricornutum* et un témoin négatif a été réalisé avec de l'eau. Les échantillons ont finalement été analysés par électrophorèse en gel d'agarose 2%.

Volume	Solution
2µL	ADN matrice
0,4µL	Amorce Forward 10µM
0,4µL	Amorce Reverse 10µM
4µL	5X Green GoTaq® Reaction Buffer
1,6µL	dNTP mix 25mM (soit 6,25mM de chaque dNTP)
0,1µL	GoTaq® G2 DNA Polymerase (5 U/µL)
11,5µL	Eau nucléase free
20µL	Total

Tableau 18 : Milieu réactionnel réalisé pour le contrôle qualité des extraits d'ARN par amplification PCR

Etape	Température (°C)	Durée	Nombre de cycles
Dénaturation initiale	95	2 min	1
Dénaturation	95	30 sec	30
Hybridation	60	30 sec	
Extension	72	30 sec	
Extension finale	72	5 min	1

Tableau 19 : Programme du thermocycleur utilisé pour le contrôle qualité des extraits d'ARN par amplification PCR

Amorce	Séquence 5'→3'
Actin1 F1	GCC-TTA-CAC-CGC-CGT-ATT-CT
Actin1 R1	TCA-ATG-ACA-AGG-GCA-GCG-AT

Tableau 20 : Séquences des amorces utilisées pour le contrôle qualité des extraits d'ARN par amplification PCR

3.4.3.4.3 Traitement des ARN à la DNase

Les extraits d'ARN présentant des contaminations avec de l'ADN génomique ont été traités à la DNase avec le kit « TURBO DNA-free™ » (Ambion ; AM1907). Pour cela, l'équivalent de 2-3µg d'ARN ont été prélevés puis mélangés avec 4µL de Turbo DNase (8U), 4µL de tampon 10X, et complétés à 40µL avec de l'eau « nuclease-free ». Les échantillons ont été incubés pendant 30 minutes à 37°C, puis 4µL de solution stop (DNase inactivation reagent) ont été ajoutés et les échantillons ont de nouveau été incubés pendant 5 minutes à température ambiante. Les tubes ont été centrifugés à 10 000g pendant 1 minute 30 secondes et le surnageant contenant les ARN purifiés a été déposé dans un nouveau tube. Une « reverse-transcription » a été réalisée sur ces échantillons comme détaillé précédemment.

3.4.3.4.4 Validation des couples d'amorces

La validation d'amorces permet de vérifier que le gène d'intérêt et le gène de référence ont la même efficacité au cours de la PCR. Le principe de la validation est de faire une gamme standard à partir d'ADNc d'un échantillon dont on soupçonne la présence des gènes recherchés. Ensuite, pour chaque couple d'amorces, on pourra comparer l'écart de cycles entre chaque dilution pour déterminer leur efficacité. En plus de cela, la spécificité des amorces est aussi analysée lors de la validation. Ainsi, dans un premier temps une amplification PCR à partir

d'ADNg et d'ADNc a été réalisée avec chaque couple d'amorces, puis le produit d'amplification PCR obtenu a été purifié et séquencé. Les amorces ayant passé ce contrôle qualité sont ensuite analysées par PCR quantitative pour s'assurer qu'un seul produit PCR soit obtenu lors de la réaction d'amplification grâce aux courbes de fusion. Les amorces ont été considérées « validées » après avoir passé trois fois le contrôle qualité par qPCR. Les expériences de PCR quantitative ont été réalisées sur la Plate-forme PRIMACEN de l'Université de Rouen en utilisant le robot de pipetage Bravo (Agilent Technologies) et le thermocycleur QuantStudio™ 12K Flex (Thermo Scientific)

3.4.3.4.4.1 Préparation des échantillons pour la validation en qPCR

Les ADNc ont été dilués afin d'obtenir une concentration d'environ 800ng/μL, puis à partir de cet échantillon une dilution en cascade au 1/3, 1/9, 1/27 et 1/81 est ensuite réalisée. En parallèle, un milieu de réaction est préparé en utilisant le mélange «Fast SYBR® Green Master » (Applied Biosystem), comme présenté dans le **tableau 21**. Sur une plaque 96 puits sont ainsi déposés 3μL d'ADNc de chaque dilution, 3μL de témoin négatif de RT (ARNs non rétro-transcrits) et 3μL d'eau utilisée comme témoin négatif de PCR. Sont ensuite ajoutés dans chaque puit 10μL de mélange contenant les réactifs (**tableau 21**). Ces différents dépôts ont été réalisés en duplicat. La plaque 96 puits est ensuite scellée avec un film puis centrifugée pendant 2 minutes à 2800rpm.

Réactifs	Volume/puit
Power SYBR® Green PCR Master Mix	6,5 μL
Forward primer (10μM)	0,13 μL
Reverse Primer (10μM)	0,13 μL
H ₂ O	3,24 μL
Volume total	= 10μL

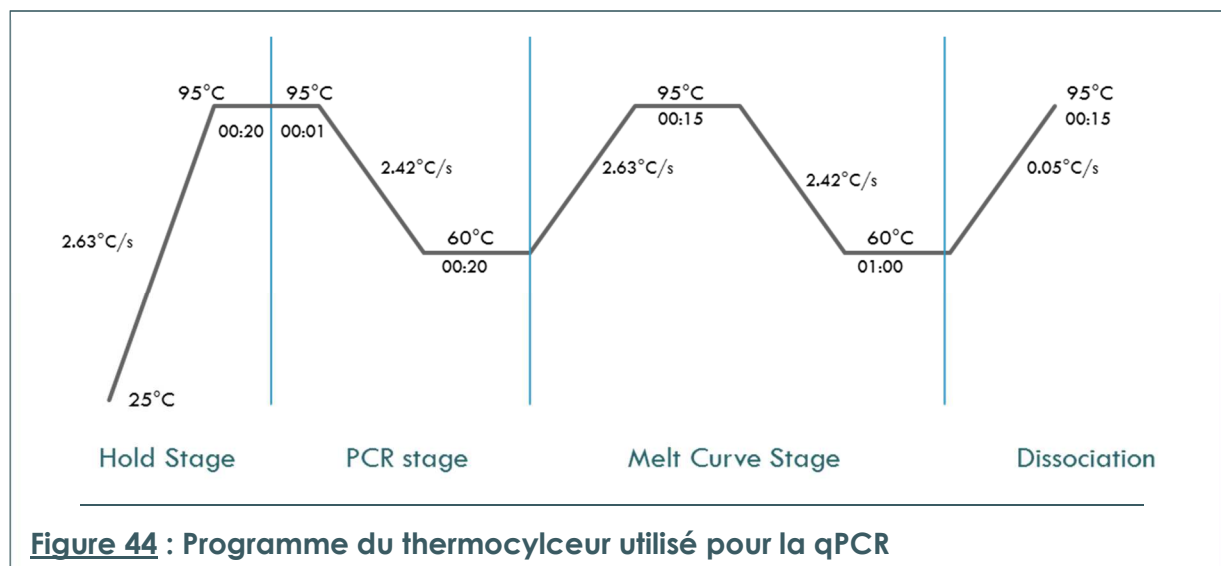
Tableau 21 : Mix de réaction utilisé pour la validation des amorces par qPCR

3.4.3.4.4.2 Analyse des résultats

Afin de valider un couple d'amorces, ce dernier doit posséder une efficacité proche de 100%, nous avons ainsi conservé uniquement les amorces possédant une efficacité comprise entre 92% et 100%. De plus, afin de vérifier la spécificité du couple d'amorce il a été vérifié que la courbe de fusion présentait un seul pic, signe qu'un seul produit d'amplification a été généré au cours de la PCR.

3.4.3.4.5 PCR quantitative

La PCR quantitative a été réalisée en duplicat pour chaque condition, c'est-à-dire pour chaque réplicat biologique provenant de chacun des trois morphotypes, afin d'obtenir un réplicat technique. Pour réaliser la mesure d'expression de chaque gène testé, un milieu réactionnel a été réalisé avec 10 μL du mélange décrit dans le **tableau 21** avec 6,5 μL de « Fast SYBR® Green PCR Master Mix » (Applied Biosystems), 0,13 μL de chaque amorce spécifique utilisée à une concentration de 10 μM , 3,24 μL d'eau « nuclease-free » auquel sont ajoutés 3 μL d'ADNc dilués au 1/5^{ème}. Les séquences des amorces utilisées sont référencées dans le **tableau 2** présenté précédemment. Les mélanges ont été préparés par une station robotisée Bravo (Agilent Technologies) puis les mesures ont été effectuées avec l'appareil « QuantStudio 12K Flex Real-Time PCR System » (Applied Biosystems). La PCR a été réalisée en utilisant le programme présenté sur la **figure 44**. Les gènes de référence utilisés dans cette étude ont été choisis à partir de données bibliographique (Siaut et al., 2007), ou des résultats de RNA-Seq en sélectionnant un panel de gènes parmi les moins différentiellement exprimés entre les morphotypes. Ces gènes ont été testés par qPCR et les meilleurs d'entre eux ont ensuite été sélectionnés grâce au logiciel NormFinder (<http://moma.dk/normfinder-software>) en utilisant les paramètres par défaut, ce qui a permis de déterminer ceux sont le plus stablement exprimés dans nos conditions.



3.4.3.4.6 Analyse des données de qRT-PCR

L'analyse de l'expression relative des différents gènes sélectionnés a été réalisée en utilisant les méthodes décrites par Livak and Schmittgen (2001) et Vandesompele et al. (2002). Dans cette approche, la méthode comparative des cycles seuil (Ct ; Livak and Schmittgen, 2001) et la normalisation par moyenne géométrique (Vandesompele et al., 2002) ont été combinées pour calculer le ratio entre l'expression du gène cible et les gènes de référence selon les calculs suivants :

- dans un premier temps, une moyenne arithmétique des Ct a été réalisée avec les deux réplicats techniques de chaque réplicat biologique :

$$\overline{Ct} = \frac{Ct1 + Ct2}{2}$$

- la quantité initiale de séquence cible présente dans un échantillon donné (N_0 , qui équivaut à l'intensité d'expression du gène cible) qui tient compte de l'efficacité d'amplification spécifique de chaque gène est donnée par la formule suivante :

$$N_{ct} = N_0 \times (1 + E_C)^{\overline{Ct}}$$

ce qui revient à :

$$N_0 = N_{ct} \times (1 + E_C)^{-\overline{Ct}}$$

- l'intensité d'expression peut s'exprimer pour chaque gène cible et chaque gène référence des quatre réplicats biologiques (k) de chaque morphotype (m) :

$$N_{0C_{m,k}} = N_{ct_C} \times (1 + E_C)^{-\overline{Ct_C}}$$

$$N_{0R_{m,k}} = N_{ct_R} (1 + E_R)^{-\overline{Ct_R}}$$

- un facteur de normalisation (NF) correspondant à la moyenne géométrique des N_0 a été calculé pour chaque réplicat biologique avec les deux gènes de référence :

$$NF = (N_{0R1_{m,k}} \times N_{0R2_{m,k}})^{1/2}$$

- la quantité relative normalisée (NRQ) pour chaque gène cible et pour chaque réplicat biologique correspond à:

$$NRQ_{m,k} = \frac{N_{0C_{m,k}}}{NF_{m,k}} = \frac{N_{ct_C}}{(N_{ct_{R1}} \times N_{ct_{R2}})^{1/2}} \times \frac{(1 + E_C)^{-\overline{Ct_C}}}{[(1 + E_{R1})^{-\overline{Ct_{R1}}} \times (1 + E_{R2})^{-\overline{Ct_{R2}}}]^{1/2}}$$

- la moyenne arithmétique des valeurs de NRQ des quatre réplicats biologiques de chaque morphotype est donnée par la formule :

$$\overline{NRQ_{m,k}} = \frac{\sum_{k=1}^4 NRQ_{m,k}}{4}$$

- les comparaisons de l'expression moyenne normalisée entre les morphotypes ont été réalisées deux à deux en effectuant un ratio avec les moyennes des NRQ :

$$\frac{\overline{NRQ_{mx}}}{\overline{NRQ_{my}}}$$

où mx et my représentent deux morphotypes considérés

- puis, afin de comparer les résultats de qRT-PCR avec ceux du RNAseq, le log2 de chaque comparaison a été calculé :

$$\log_2\left(\frac{\overline{NRQ_{mx}}}{\overline{NRQ_{my}}}\right)$$

Avec

- E_C = efficacité d'amplification du gène cible
- E_R = efficacité d'amplification du gène de référence (ici E_{R1} pour le gène de référence 1 et E_{R2} pour le gène de référence 2)
- $\overline{Ct_C}$ = moyenne des cycles seuils pour les réplicats techniques du gène cible
- $\overline{Ct_R}$ = moyenne des cycles seuils pour les réplicats techniques du gène de référence
- m = morphotype considéré
- N_0 = nombre initial de séquences cibles
- N_{ct} = nombre de séquences cibles au cycle seuil Ct

3.5 Localisation subcellulaire des glycosyltransférases de fusion par microscopie confocale

3.5.1 Préparation des échantillons

Les cultures de microalgues transfectées avec les différentes glycosyltransférases exprimées en fusion avec la eGFP (enhanced Green Fluorescent Protein) ont été induites durant 48h en amont de l'observation en déposant dans le milieu du NaNO_3 à une concentration finale de $890 \cdot 10^{-6}$ M. Les cultures de cellules exprimant ainsi ces protéines recombinantes fusionnées à la eGFP ont été déposées directement entre une lame de verre et une lamelle couvre-objet pour observation en microscopie confocale.

3.5.2 Paramètres d'observation

Les observations ont été réalisées sur un microscope confocal inversé Leica TCS SP5. Les réglages du gain (limite haute du PMT) et de l'offset (limite basse du PMT) ont été effectués pour chaque observation afin de diminuer au maximum les signaux de fluorescence non spécifiques. Les images ont été acquises avec deux configurations différentes qui sont respectivement détaillées ci-après. Les images ont ainsi été réalisées avec un frame-average de 2, un line-average de 2, au format 1024x1024 pixels avec une vitesse de scan de 400Hz et un laser à 20% sous un objectif HCX PL APO CS 63.0x1.40 OIL UV (toutes les images sauf celles concernant la GnT I, la XylT et la FuT54599); ou avec un frame-average de 1, un line-average de 2, au format 512x512 pixels avec une vitesse de scan de 400Hz et un laser à 56% sous un objectif HCX PL APO 63.0x1.32 OIL PH3 U (GnT I, XylT et FuT54599). Les observations de l'eGFP ont été réalisées avec un laser à argon réglé à 488nm, et les observations de la mRFP ont été réalisées avec un laser HeNe réglé à 543nm. L'acquisition des signaux de fluorescence a été réalisée sur trois canaux différents : le premier possédant une gamme de 498 à 530nm, correspondant aux longueurs d'ondes émises par la eGFP ; le second possédant une gamme de 680 à 710nm, correspondant à l'autofluorescence chloroplastique ; et le troisième sur une gamme de 580 à 600nm pour les longueurs d'onde émises par la mRFP.

3.6 Localisation subcellulaire des glycosyltransférases de fusion par Microscopie Electronique à Transmission

3.6.1 Cryo-immobilisation haute pression

3.6.1.1 Préparations des échantillons

Les suspensions de *P. tricornutum* transformées, âgées de 8 jours, sont induites en présence de NaNO_3 à une concentration finale de $890 \cdot 10^{-6}$ M pendant 48h. Les cultures de microalgues sont ensuite concentrées par centrifugation à 2 500g durant 15min dans un tube conique de 50mL. Les cellules sont rincées une fois avec du milieu de culture frais, puis le culot cellulaire est récupéré et resuspendu dans du milieu de culture contenant du mannitol (100mM) dont le rôle est de cryo-protéger les microalgues. Les cellules sont incubées pendant 2 heures dans la solution cryo-protectante. Les cellules sont ensuite de nouveau concentrées par centrifugation à 2 500g pendant 15min. Le surnageant est éliminé afin de ne garder qu'un culot épais. Sous stéréo-microscope, 2 μ L de cette suspension épaisse sont rapidement transférés dans une cupule d'aluminium (diamètre intérieur de 2mm ; profondeur cavité de 300 μ m; Leica), puis une deuxième cupule est déposée pour former un sandwich. Les cupules ont été préalablement

enduites d'une solution de lécithine de soja (100mM dans du chloroforme) pour faciliter le démoulage ultérieur de l'échantillon.

3.6.1.2 Cryo-fixation des échantillons sous haute pression

La cryo-fixation sous haute pression permet une fixation physique de l'échantillon, évitant ainsi les artefacts de préparation dus à l'utilisation de fixateurs chimiques. De plus, l'échantillon étant congelé très rapidement et à très basse température, il se retrouve congelé, non pas sous forme de « glace cristalline », mais sous forme de « glace vitreuse » qui n'altère pas son ultrastructure et préserve au mieux son état natif, comme présenté sur la **Figure 45**.

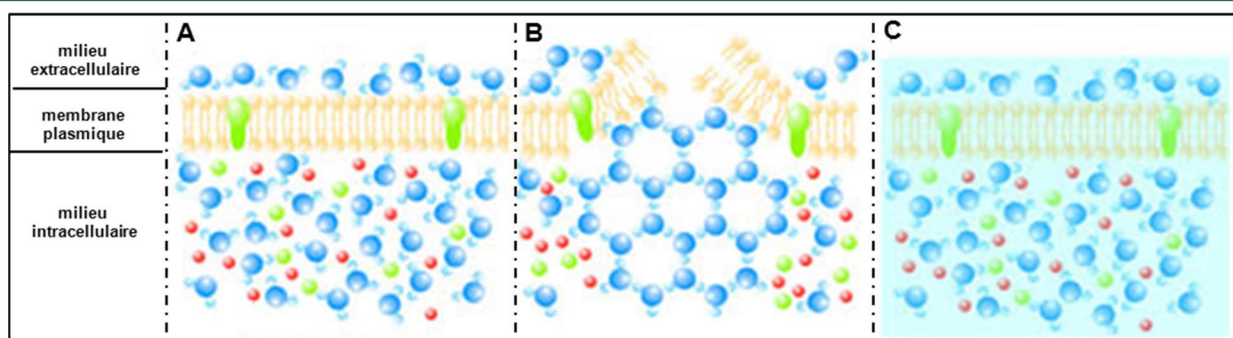


Figure 45 : Les différentes formes de l'eau dans une cellule.

A- Echantillon à température ambiante avec un milieu intracellulaire sous forme liquide,
B- Echantillon congelé et endommagé par la glace cristalline formée,
C- Echantillon congelé sous forme de glace vitreuse et non endommagé.

La congélation sous haute pression a été réalisée avec l'appareil HPM100 (Leica). Brièvement, la pression est générée par un système hydraulique faisant intervenir l'azote liquide comme liquide de compression. Le cryogène utilisé est l'azote liquide (-196°C). La pression (2100 bars), appliquée à l'intérieur de l'échantillon est synchronisée avec le refroidissement (jet d'azote liquide) appliqué à l'extérieur de l'échantillon entraînant une congélation ultrarapide de l'échantillon. Des taux de refroidissement variant de 16 000°C/sec à 20 000°C/sec sont obtenus. Lorsque l'échantillon est congelé, il tombe automatiquement dans un réservoir d'azote où il est conservé en attendant son transfert dans l'automate de substitution.

3.6.1.3 Cryo-substitution et enrésinement

Après congélation sous haute pression, les échantillons sont transférés dans l'automate de substitution AFS2 (Leica) pré-refroidi à -110°C. Les échantillons sont substitués dans l'acétone anhydre contenant 0,5% d'acétate d'uranyle et 0,5% de tétr oxyde d'osmium à -90°C pendant 96h. La température est graduellement augmentée de 2°C/h jusqu'à -60°C et maintenue

à cette température pendant 12h. La température est de nouveau remontée à -30°C (2°C/h). Après un palier de 12h à -30°C , la température est remontée (2°C/h) jusqu'à -15°C .

Les échantillons sont rincés deux fois 15 min dans de l'acétone anhydre puis enrésinés progressivement dans un mélange de résine LR White (Electron Microscopy Sciences) / Ethanol anhydre à -15°C de la façon suivante : résine 25% / 6h ; résine 50% / 16h ; résine 75% / 8h ; résine 100% / 16h ; résine 100% / 24h ; résine 100% + catalyseur /24h. Les échantillons sont délicatement ôtés des cupules sous loupe binoculaire et placés dans des moules avec adaptateurs dans la résine 100% contenant le catalyseur. L'ensemble est mis à polymériser pendant 48h sous UV à -15°C .

3.6.2 Réalisation des coupes ultrafines

Les sections ultra-fines de 80 à 90nm ont été obtenues sur un ultramicrotome UCT (Leica) avec un couteau de diamant Diatome 45° . Les sections ont été recueillies sur des grilles de nickel de 100 mesh à trames hexagonales recouvertes d'une membrane formvar-carbone (Leica).

3.6.3 Immunocytochimie

3.6.3.1 Immunomarquage des protéines

L'anticorps primaire utilisé est un anticorps produit chez la souris, qui reconnaît l'épitope V5 (mouse anti-V5 tag, Invitrogen). L'anticorps secondaire utilisé est un anticorps anti-souris conjugué à une particule d'or de 10nm (EM-goat anti mouse IgG+IgM10 nm gold particle, BBI solution).

Afin de saturer les sites non spécifiques, les sections ultra-fines sont incubées durant 15 min dans un tampon TBS (Tris 20mM ; NaCl 15mM) contenant de la BSA 0,2% et du Tween 20 0,05%, avant d'être rincées pendant 2 min dans un tampon (TBS contenant de la BSA 0,2%). Les coupes sont ensuite incubées dans une solution de TBS/BSA 0,2% contenant du sérum Normal Goat Serum (NGS) dilué au $1/30^{\text{ème}}$ et l'anticorps primaire dilué au $1/20^{\text{ème}}$. Cette incubation est réalisée durant 2h30 à 25°C , avant que les coupes soient rincées 6x5 min dans le tampon TBS /BSA 0,2%. L'incubation dans l'anticorps secondaire dilué au $1/20^{\text{ème}}$ dans une solution de (TBS + BSA 0,2% + NGS $1/30$) s'effectue pendant 1h à 25°C . Les rinçages sont effectués 6x5 min avec du tampon TBS /BSA 0,2%, puis 5 min dans du TBS seul. Afin de fixer les complexes immuns, une incubation de 10 min a été réalisée dans une solution de TBS / glutaraldéhyde 2 %. Les coupes sont ensuite rincées 1x5 min dans du TBS, puis 2x5 min dans l'eau milli-Q, avant d'être séchées à l'air.

3.6.3.2 Coloration et observation des coupes ultra-fines

Les échantillons ainsi préparés sont transparents aux électrons, il faut donc les contraster pour pouvoir les observer en MET. Pour cela, les grilles sont incubées pendant 3 min à température ambiante dans une solution Uranyl-less (Delta Microscopie France) puis 6 rinçages rapides dans de l'eau milli-Q sont réalisés. A la suite, les grilles sont plongées pendant 3 min dans une solution de citrate de plomb à 3% (Reynolds lead citrate, Chromalys SAS) et 6 lavages dans de l'eau milli-Q sont effectués. Les grilles sont ensuite séchées à l'air libre.

Dans le cas de grilles non immuno-marquées, le contraste est précédé d'une oxydation par une solution acide de permanganate de potassium dont le but est de l'augmenter (Sawaguchi et al., 2001) en procédant de la façon suivante : les grilles sont incubées durant 1 min à température ambiante dans une solution contenant 0,1% de KMnO_4 préparée dans une solution de H_2SO_4 0,05M. Enfin, les coupes ultra-fines sont rincées 6 fois dans de l'eau milli-Q avant de pouvoir être observées.

Les observations des coupes ont été réalisées avec le microscope électronique à transmission TECNAI 12 Biotwin (FEI company, Eindhoven, NL) sous une tension d'accélération de 80kV. Les images ont été acquises avec une caméra CCD Erlangshen ES500W (GATAN) pilotée par le logiciel Digital Micrograph (GATAN).

3.7 Techniques de biochimie

3.7.1 Extraction des protéines de *P. tricornutum* pour Western-blot

Dans le but de réaliser un criblage des transformants de *P. tricornutum* par Western-blot, une extraction protéique a été réalisée pour chaque clone. L'équivalent de 10mL de culture cellulaire induite 24h à l'avance a été récolté par centrifugation à 3 500g pendant 20 minutes. Les cellules ont ensuite été resuspendues dans du tampon phosphate (20mM K_2HPO_4 ; 20mM KH_2PO_4 ; pH=7,4). Les suspensions cellulaires ont été transférées dans un tube de «lysing matrix E » (MP Biomedicals) puis placées dans un broyeur Fastprep 24 (MP Biomedicals) où les cellules ont été lysées par 5 cycles de 30 sec à une vitesse de 6,5m/s. Les tubes ont ensuite été centrifugés à 300g pendant 3min. Les surnageants ont été de nouveau centrifugés à 20 000g pendant 30 minutes pour culoter la fraction plastidiale. Les surnageants de cette centrifugation sont récupérés puis centrifugés à 100 000g pour culoter la fraction microsomale. Les culots ont été repris avec 100 μL de LBS 1X pour être ultérieurement chargés sur gel SDS-PAGE.

3.7.2 Western-blot

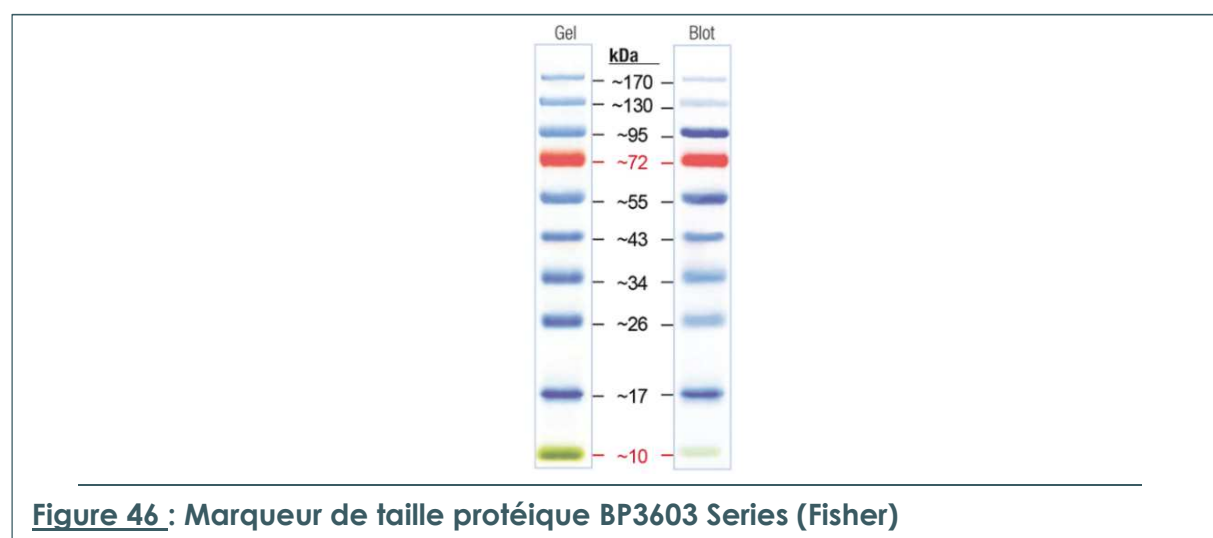
3.7.2.1 Séparation des protéines par gel d'électrophorèse SDS-PAGE

Cette technique permet de séparer les protéines selon leur taille et d'estimer ainsi leur masse moléculaire par comparaison avec la mobilité électrophorétique de protéines de référence. Dans cette étude, les protéines ont été séparées par migration sur un gel de polyacrylamide 12%. La composition des gels de concentration et de séparation est mentionnée dans le **tableau 22**.

Solutions	Gel de concentration	Gel de séparation
H ₂ O déionisée	0,73mL	2,15mL
Acrylamide:Bis-acrylamide (29:1) 40%	0,17mL	1,5mL
Tris-HCl 1,5M pH=8,8	-	1,25mL
Tris-HCl 1,5M pH=6,8	0,08mL	-
Dodécylsulfate de sodium (SDS; 10%)	10µL	50µL
Persulfate d'ammonium (10%)	10µL	50µL
TEMED	1µL	2µL

Tableau 22 : Composition des gels de concentration et de séparation pour SDS-PAGE

Pour chaque échantillon à analyser, une dénaturation a été réalisée pendant 10 minutes à 95°C dans le tampon de charge (Tris-HCl 62,5mM, pH=6,8 ; SDS 2% ; glycérol 10% ; β-mercaptoéthanol 2,5% ; bleu de bromophénol 0,005%), avant de le déposer dans les puits du gel de polyacrylamide. Cinq microlitres de marqueur de taille BP3603 Series (Fisher) ont été déposés en parallèle (**figure 46**). La migration a été réalisée sous une tension de 160V dans un tampon contenant 25mM de Tris-HCl, 1% de SDS et 192 mM de glycine jusqu'à ce que la séparation des protéines soit satisfaisante.



3.7.2.2 *Transfert électrophorétique de protéines sur membrane de nitrocellulose*

Afin de transférer les protéines du gel d'électrophorèse sur une membrane de nitrocellulose, cette dernière a été hydratée dans du tampon de transfert (200mM glycine, 25mM Tris, 0.1% SDS) à 4°C. Après migration, le gel SDS-PAGE va être mis en contact de la membrane de nitrocellulose, le transfert des protéines sur la membrane de nitrocellulose a été obtenu en 1h30 sous une tension électrique de 12V à l'aide d'un « Trans-Blot® SD Semi-Dry Transfer Cell » (Bio-Rad). L'efficacité du transfert a ensuite été contrôlée par une coloration réversible des protéines au rouge Ponceau (1% dans de l'acide trichloroacétique 3%). La décoloration de la membrane, après ce traitement, a été obtenue par un lavage à l'eau distillée, puis deux rinçages de 5 minutes avec un tampon TBS (10mM Tris-HCl, 500mM NaCl, pH=7,5). Toutes ces étapes de traitement de la membrane de nitrocellulose, après le transfert, ont été effectuées sous agitation à température ambiante. Les protéines ainsi immobilisées sur la membrane de nitrocellulose ont été analysées par immunodétection.

3.7.2.3 *Immuno-marquage sur empreinte*

La membrane de nitrocellulose a été saturée pendant une nuit sous agitation à température ambiante, dans une solution de TBST (tampon TBS complété avec 0,1% de Tween 20) contenant 5% de lait écrémé. La membrane a ensuite été incubée pendant 2 heures avec une solution de TBST 0,1% complété avec 1% de lait écrémé et l'anticorps primaire anti-V5 (Invitrogen ; 46-0705), ou anti-c-Myc (Sigma ; M4439) dilué au 1/5000^{ème}. Ces anticorps monoclonaux produits chez la souris permettent de reconnaître respectivement les épitopes V5 et c-Myc placés à l'extrémité C-terminale des protéines recombinantes. A l'issue de ce marquage, les anticorps non fixés ont été éliminés par une série de 6 lavages de 5 minutes dans le tampon TBST. La membrane a ensuite été incubée pendant 1 heure avec un anticorps secondaire anti-souris couplé à la peroxydase de raifort (HRP) dilué au 1/20 000^{ème} dans une solution de TBST 0,1% contenant 1% de lait écrémé. Suite à ce deuxième marquage, une série de 6 lavages de 5 minutes avec une solution de TBST 0,1% + 1% lait écrémé, suivie de 3 lavages de 10 minutes dans du tampon TBS a été réalisée.

3.7.2.4 *Révélation par électrochimiluminescence*

Les membranes de nitrocellulose marquées avec l'anticorps secondaire couplé à la HRP ont été révélées par électrochimiluminescence en utilisant le kit « Pierce™ ECL Western Blotting Substrate » (Thermo Scientific ; 32106) selon le protocole du fournisseur. Après élimination du mélange réactionnel, la membrane a été recouverte d'un film plastique et placée dans une cassette de révélation. Un film photographique a alors été exposé pendant 30 secondes

au contact de la membrane à l'intérieur de la cassette, dans une chambre noire. Ce film a ensuite été placé dans une solution de révélation (Kodak) diluée au $\frac{1}{4}$ dans de l'eau, puis rincé à l'eau, avant d'être déposé dans une solution de fixation (Kodak) diluée au $\frac{1}{5}$ dans de l'eau. Lorsque le film est devenu totalement transparent, il a alors été rincé à nouveau dans l'eau puis séché.

Chapitre III:

Comparative transcriptome analysis of *P. tricornutum*'s morphotypes

4 Comparative transcriptome analysis of *P. tricornutum*'s morphotypes using in depth RNA sequencing reveals the oval morphotype features, especially that protein degradation and secretion is up-regulated in the oval morphotype.

Clément Ovide^{1*}, Marie-Christine Kiefer-Meyer^{1*}, Caroline Bérard², Nicolas Vergne³, Thierry Lecroq², Carole Plasson¹, Carole Burel¹, Sophie Bernard^{1,4}, Patrice Lerouge¹, Isabelle Tournier⁵, Hélène Dauchel^{2*#} and Muriel Bardor^{1, 6*#}

** These authors equally contributed to the work.*

¹ Normandie Univ, UNIROUEN, laboratoire Glyco-MEV 76000 Rouen, France.

² Normandie Univ, UNIROUEN, laboratoire LITIS EA 4108, 76000 Rouen, France.

³ Normandie Univ, UNIROUEN, LMRS UMR 6085 CNRS, 76000 Rouen, France.

⁴ Normandie Univ, UNIROUEN, LMRS UMR 6085 CNRS, 76000 Rouen, France.

⁵ Normandie Univ, UNIROUEN, Inserm U1079, 76000 Rouen, France

⁶ Institut Universitaire de France (IUF), Paris, France.

#correspondence to:

Muriel BARDOR,

Laboratoire Glyco-MEV EA 4358, Université de Rouen, UFR des Sciences et Techniques, Bâtiment Henri Gadeau de Kerville, 1 rue Tesnières, 76821 Mont-Saint-Aignan Cedex, France. Tel: +33 2 35 14 67 51; e-mail: muriel.bardor@univ-rouen.fr

Hélène Dauchel,

LITIS EA 4108 Computer science, information processing and systems laboratory, Normandie Université, Université de Rouen, IRIB, 76821 Mont Saint Aignan Cedex, France. Tel: +33 2 35 14 63 89; e-mail: helene.dauchel@univ-rouen.fr

Short title: RNA-Seq analysis of *P. tricornutum*'s morphotypes

Key words: algae, transcriptome, morphogenesis, diatom, *Phaeodactylum tricornutum*

4.1 ABSTRACT

Background: *Phaeodactylum tricornutum* is the most studied diatom. It is a cosmopolitan diatom not widely distributed in nature, but is encountered principally in coastal unstable environments. It has been already hypothesized that the great adaptation capability of *P. tricornutum* is due to its pleomorphism. Indeed, *P. tricornutum* is an atypical diatom since it can display three major morphotypes: fusiform, triradiate and oval. Currently, little information is available regarding the physiological significance of this morphogenesis and the specificities of each morphotype.

Results: In this study, we adapted *P. tricornutum* Pt3 strain in order to obtain algal culture particularly enriched in one dominant morphotype: fusiform, triradiate or oval. These cultures were used to run high-throughput RNA-Sequencing. The whole transcriptome of each morphotype was determined. Pairwise comparison of the morphotypes allow us to highlight biological processes and molecular functions which are up- and down-regulated. Finally, intersection analysis allow us to identify the specific features from the oval morphotype which is of particular interest as it is known to be more resistant to stresses.

Conclusions: This study is the first transcriptome wide characterization of the three morphotypes from *P. tricornutum* performed on cultures specifically enriched for one or another morphotype issued from the same Pt3 strain. This work represents an important step for the understanding of the morphogenesis in *P. tricornutum* and highlight the particular features of the oval morphotype, often recognized as a stress morphotype.

4.2 BACKGROUND

Diatoms are a major component of phytoplankton communities. They belong to a group of unicellular heterokont microalgae expected to include at least 200 000 species (Armbrust, 2009). Their genomes contain unique combinations of nutrient assimilation and metabolic pathway which have been attributed to their origin (Keeling and Palmer, 2008). Indeed, diatoms derived from a serial secondary endosymbiotic event in which a green algae and subsequently a red one were engulfed by a heterotrophic eukaryote (Bowler et al., 2010; Moustafa et al., 2009; Prihoda et al., 2012). Among the diatom species, *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* have been chosen by the scientific community as a representative for the centric (Armbrust et al., 2004) and the pennate diatoms (Bowler et al., 2008), respectively. The comparative analysis of *T. pseudonana* and *P. tricornutum* genomes highlights the presence of 10% of diatom specific genes which has no homologs in other eukaryotes (Armbrust et al., 2004; Bowler et al., 2008; Montsant et al., 2005). In addition, it has been demonstrated that *P. tricornutum* and *T. pseudonana* possess 6.4% and 2% of transposable elements, respectively (Armbrust et al., 2004; Bowler et al., 2008; Maumus et al., 2009) as well as microRNAs (Huang et al., 2011; Norden-Krichmar et al., 2011).

P. tricornutum is probably the widest studied diatom and its genome has been sequenced (Bowler et al., 2008; Montsant et al., 2005). It is a cosmopolitan diatom not widely distributed in nature, but is encountered principally in coastal unstable environments such as estuaries or rock pools (De Martino et al., 2007; Rushforth et al., 1988). Moreover, *P. tricornutum* is able to proliferate in salinity and temperature caused by tidal effects, solar irradiation and human activities (De Martino et al., 2007). These observations highlight the great adaptation capability of this diatom. As previously hypothesized, the adaptation of *P. tricornutum* to a fluctuating environment is probably due to its pleomorphism (Borowitzka and Volcani, 1978; De Martino et al., 2007; Gutenbrunner et al., 1994; Wilson, 1946). Indeed, *P. tricornutum* is an atypical diatom since it can display three major morphotypes: fusiform, oval and triradiate (He et al., 2014; Lewin et al., 1958). As compared to the other diatoms, *P. tricornutum* present a high plasticity since it contains a poorly silicified cell wall (Borowitzka and Volcani, 1978; De Martino et al., 2007). Currently, little information is available regarding the molecular and cellular biology of *P. tricornutum* and especially the physiological significance of its morphogenesis as well as its mechanism of regulation.

Studies already suggested that the morphogenesis in *P. tricornutum* is not dependent from its genotype but is induced by environmental factors (De Martino et al., 2007). For example, it has been observed that the fusiform morphotype which is the most widespread one and the triradiate are

avored in unstressed planktonic environment (De Martino et al., 2007). It was also noticed that the triradiate morphotype preferentially develops with alkaline condition (Bartual et al., 2008; Johansen, 1991). In contrast, the oval morphotype seems to increase in unfavorable conditions such as hyposaline conditions, low temperature, low light (De Martino et al., 2011), unshaken liquid culture (Gutenbrunner et al., 1994) or solid substrate (Borowitzka and Volcani, 1978). Comparative transcriptomic analysis based on EST showed that the oval morphotype presents up-regulated genes encoding proteins involved in hyposalinity and/or cold stress responses, with putative roles in signaling, membrane remodeling, cryoprotection, osmoregulation, and protein degradation (De Martino et al., 2007). Furthermore, this form demonstrate an ability to survive with limited nutrient availability which suggests that the oval morphotype could be a resistance form to stresses. More generally, it has been proposed that osmolarity changes could act as a trigger for changes in morphology (De Martino et al., 2011).

In this paper, we made use of the power that RNA sequencing (RNA-seq) is offering to identify the whole transcriptome from the fusiform, oval and triradiate morphotypes generated from the same *P. tricornutum* Pt3 strain. The objective of the study was to gain insight into the diatom morphogenesis by identifying genes that are differentially expressed and bringing out specific features which are representatives of each morphotype. A particular attention has been paid to the oval morphotype as it is recognized to be highly resistant to stresses (De Martino et al., 2007, 2011).

4.3 RESULTS

4.3.1 Experimental Design and mapping

In order to evaluate transcriptome changes in *P. tricornutum*'s morphotypes, cultures of *P. tricornutum* Pt3 strain enriched in one specific and dominant morphotype, either fusiform (91%), triradiate (81%) or oval (94.5%), were generated and followed over a 12 days continuous culture without any change on the morphotype repartition (**Figure 47 A to F; Figure 48**).

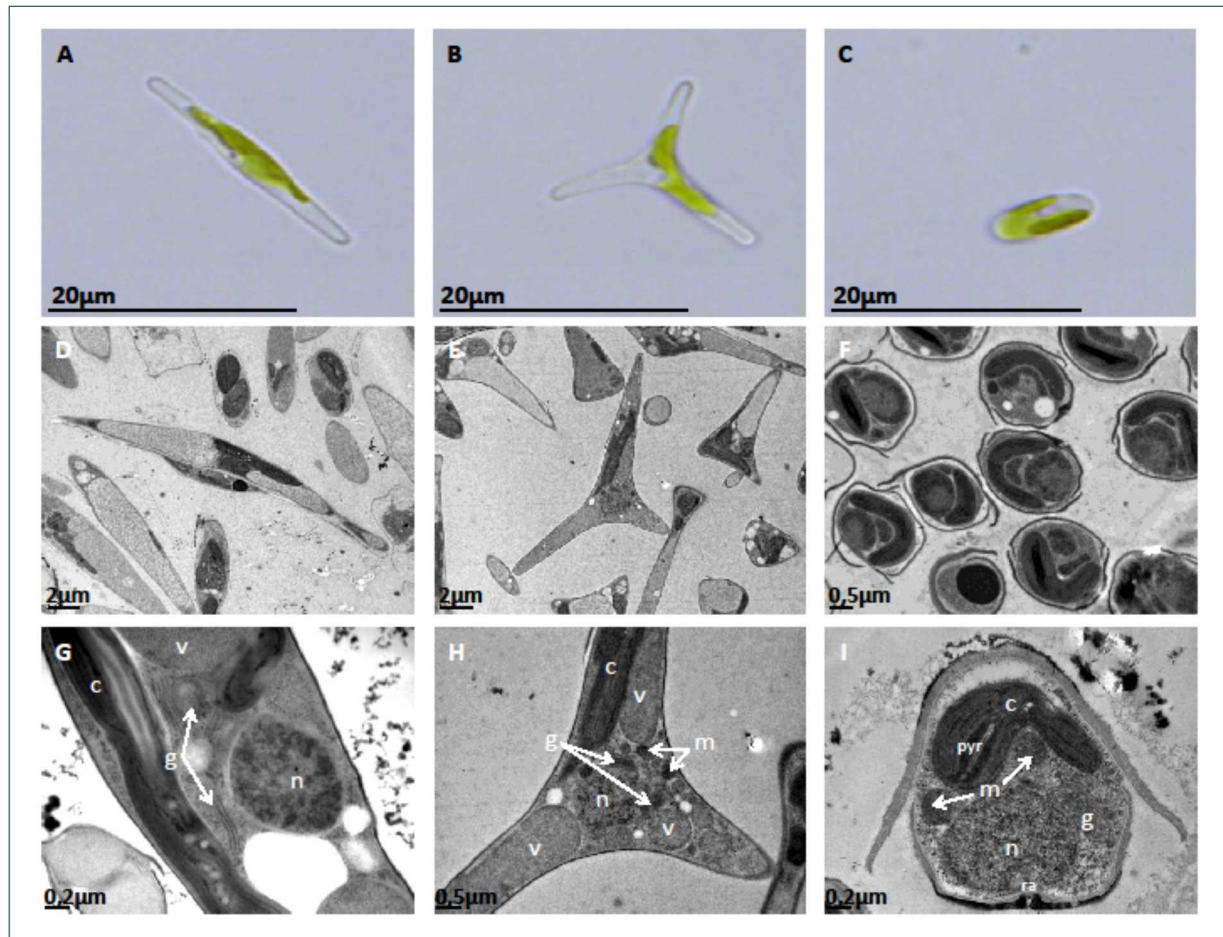
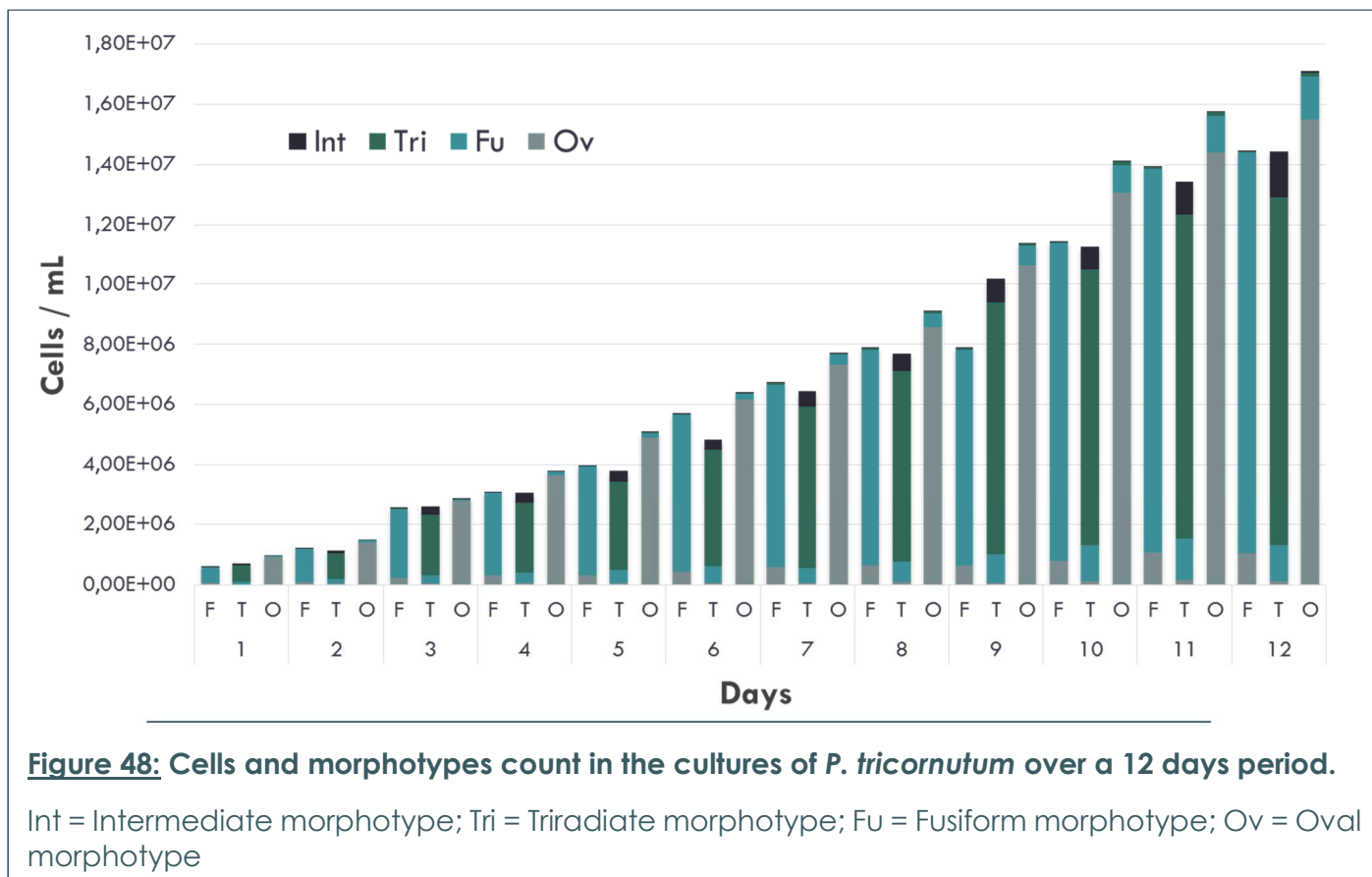


Figure 47: Micrographs of *P. tricornutum* dominant morphotypes present in the 3 enriched cultures of the Pt3 strain.

A, B, C: Micrographs obtained by light microscopy of *P. tricornutum* cells alive. A: fusiform morphotype; B: triradiate morphotype and C: Oval morphotype.

D, E, F: Transmission electron microscopy (TEM) micrographs of the three morphotypes. Overview of the cells which were embedded in LRW resin with 0.5% uranyl acetate in a methanol/Reynold's lead citrate solution. D: fusiform morphotype; E: triradiate morphotype and F: Oval morphotype.

G, H, I: Enlarge views of the TEM micrographs showing general cellular distribution of organelles in the fusiform cells (G), in the triradiate one (H) and in the oval cell type (I). n: nucleus; g: Golgi apparatus; v: vacuole; m: mitochondria; pyr: pyrenoid;; c: chloroplast; ra: raphe.



The diatom cells were grown as described in the experimental section and follow similar growth phase (**Figure 48 and 39**). The culture conditions used in this study allow us to maintain the specificity of each culture batch (**Figure 48**) as well as the diatom cell integrity and organization for each morphotype as illustrated by the TEM micrographs (**Figure 47 G to I**) and according to the literature (Gügi et al., 2015; Tesson et al., 2009).

In order to study differential gene expression between *P. tricornutum*'s morphotypes and assign specificities for each morphotype, gene transcription profiling was investigated using high-throughput RNA-Sequencing (RNA-Seq) following the experimental design which is summarized in the **Figure 49**.

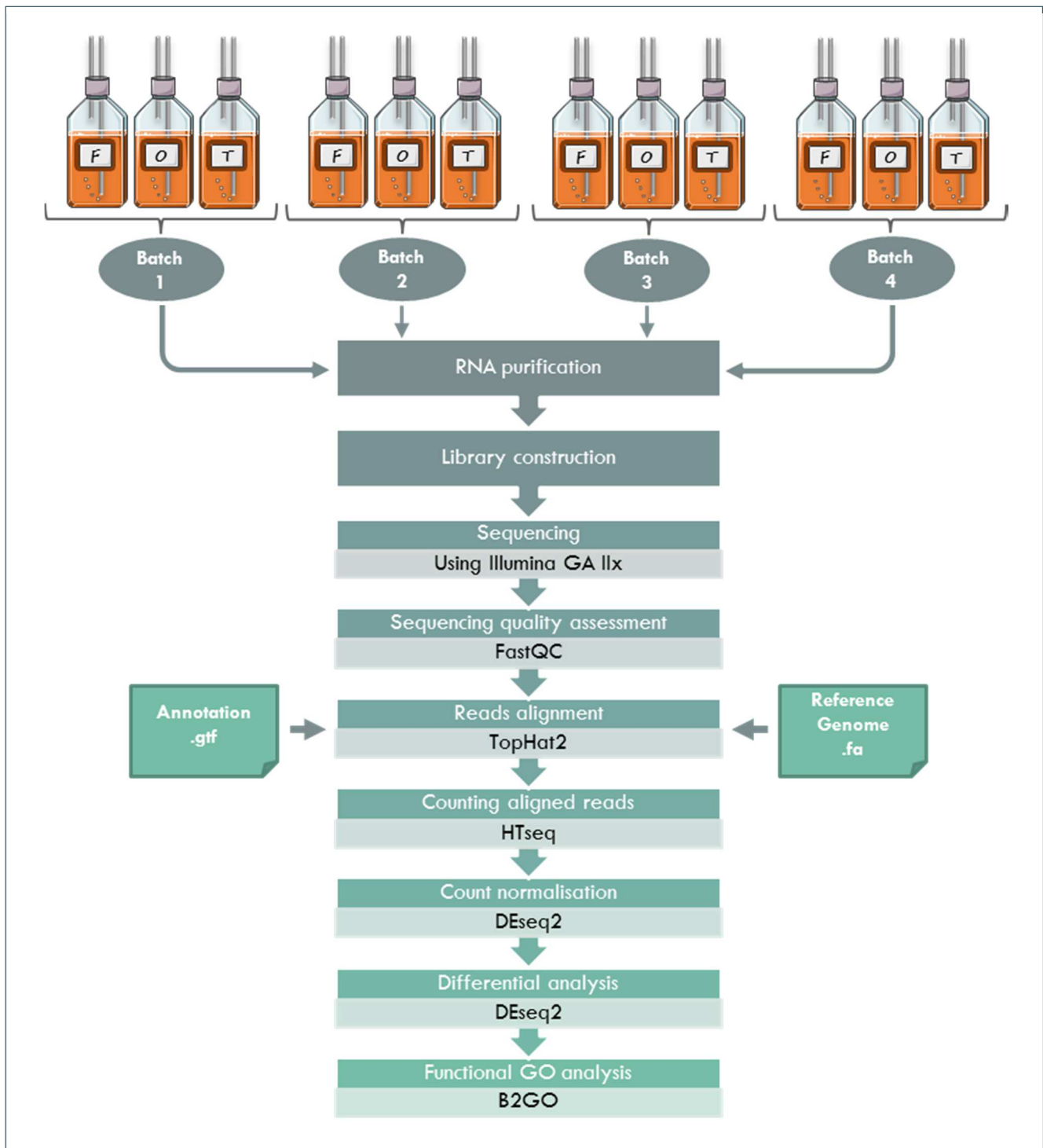


Figure 49: Summary of the experimental design and analytical pipeline used in this study.

Four biological replicates of cultures enriched for each specific morphotype were performed in parallel. The RNA were independently purified and used as a matrix to build libraries for RNA sequencing. Quality of the raw data generated have been checked with FastQC. Then, the short reads were aligned on the annotated genome with TopHat2. Aligned reads have been counted with HTseq. DESeq2 have been used to normalize counts and to perform the differential analysis. Finally, GO enrichment have been done with B2GO for functional analysis.

Briefly, 4 biological replicates of each morphotype were cultured. No obvious differences were observed in the growth curve at the sampling time of 8 days (**Figure 48**). Therefore, the cells were harvested in the middle of the exponential growth phase at day 8 and then further treated for RNA purification and cDNA library preparation in order to perform RNA-Seq analysis. After the sequencing, 38 millions quality reads were generated in average per sample. After quality assessment, RNA-Seq data were then analyzed using the pipeline described in the **Figure 49**. The high quality reads were mapped on the 27.56 Mpb genome of *P. tricornutum*'s P.t.1.8.6 strain (Bowler et al., 2008). Results indicated that around 83% of the RNA-Seq reads were aligned with in average 5% of the reads corresponding to intron/exon junctions for the fusiform morphotype. Similar results were obtained for each morphotype (**Table 23 and Figure 50**).

Conditions	F1	F2	F3	F4	T1	T2	T3	T4	O1	O2	O3	O4
Reads in right	23117449	15298228	17634774	17925360	22445606	14314071	18289429	23393382	18073097	17996250	16397447	23285568
Reads in left	23117449	15298228	17634774	17925360	22445606	14314071	18289429	23393382	18073097	17996250	16397447	23285568
Reads out right	23089050	15282358	17617406	17906233	22422393	14300476	18267749	23368277	18049817	17976769	16378341	23262680
Reads out left	23114480	15298040	17634269	17925227	22444085	14313597	18289145	23393065	18069823	17996066	16397134	23285201
Reads aligned right	19475078	12617512	14636505	15246361	18451385	11756755	15964896	20236651	14837746	15011917	13586822	19518782
% Reads aligned right	84,35	82,56	83,08	85,15	82,29	82,21	87,39	86,6	82,2	83,51	82,96	83,91
Reads aligned left	19404080	12615170	14669491	15215790	18431242	11790578	15963641	20281578	14782468	14967481	13600585	19409565
% Reads aligned left	83,95	82,46	83,19	84,88	82,12	82,37	87,28	86,7	81,81	83,17	82,94	83,36
Junctions	2219007	1333824	1363321	1557767	2177984	1333563	1680197	2160152	1610586	1796936	1579857	3589493
% Reads on junctions	5,510406	5,10881	4,535788	4,947147	5,678265	5,461346	5,092778	5,145548	5,246137	5,748145	5,616213	8,571894
nb lignes sam	40269393	26108313	30056981	31488186	38356503	24418210	32991759	41980990	30700419	31261144	28130291	41875144
% reads sam / out	87,15653	85,37598	85,26398	87,8786	85,49034	85,33637	90,24771	89,77713	84,99647	86,90209	85,82726	89,96144

Table 23: Metric of mapping from TopHat for the 4 biological replicates of the cultures enriched in the fusiform, triradiate and oval morphotype, respectively.

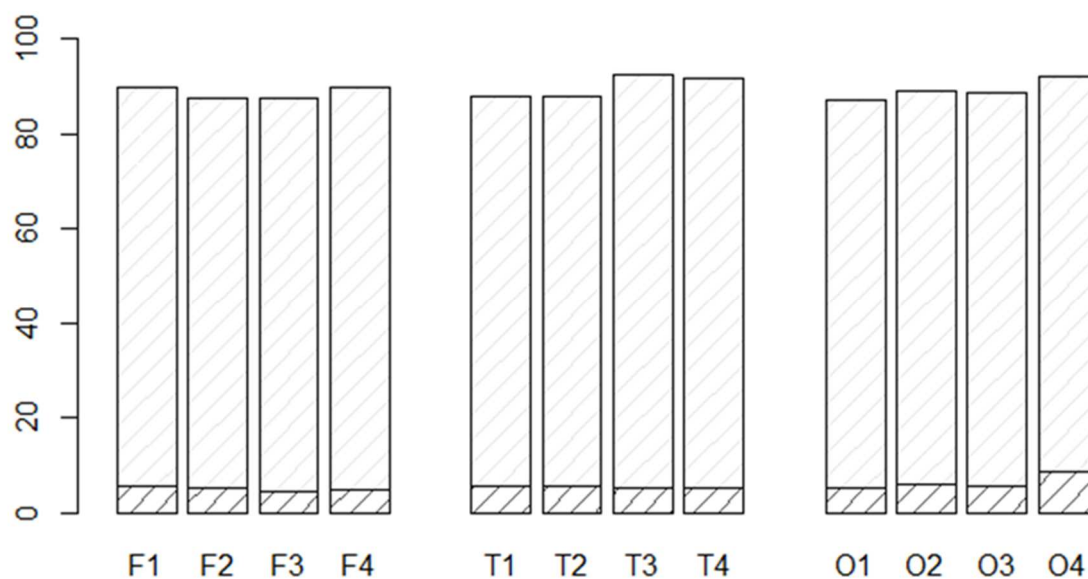


Figure 50: Histogram representing the percentages of read alignment with TopHat for the 4 biological replicates of the cultures enriched in the fusiform, triradiate and oval morphotype, respectively.

Proportion of reads aligned to intron-exon junction are also indicated at the bottom of the bars.

Raw reads were counted using HTSeq-count (Anders et al., 2015). The reads counts were therewith successfully normalized (Figure 51).

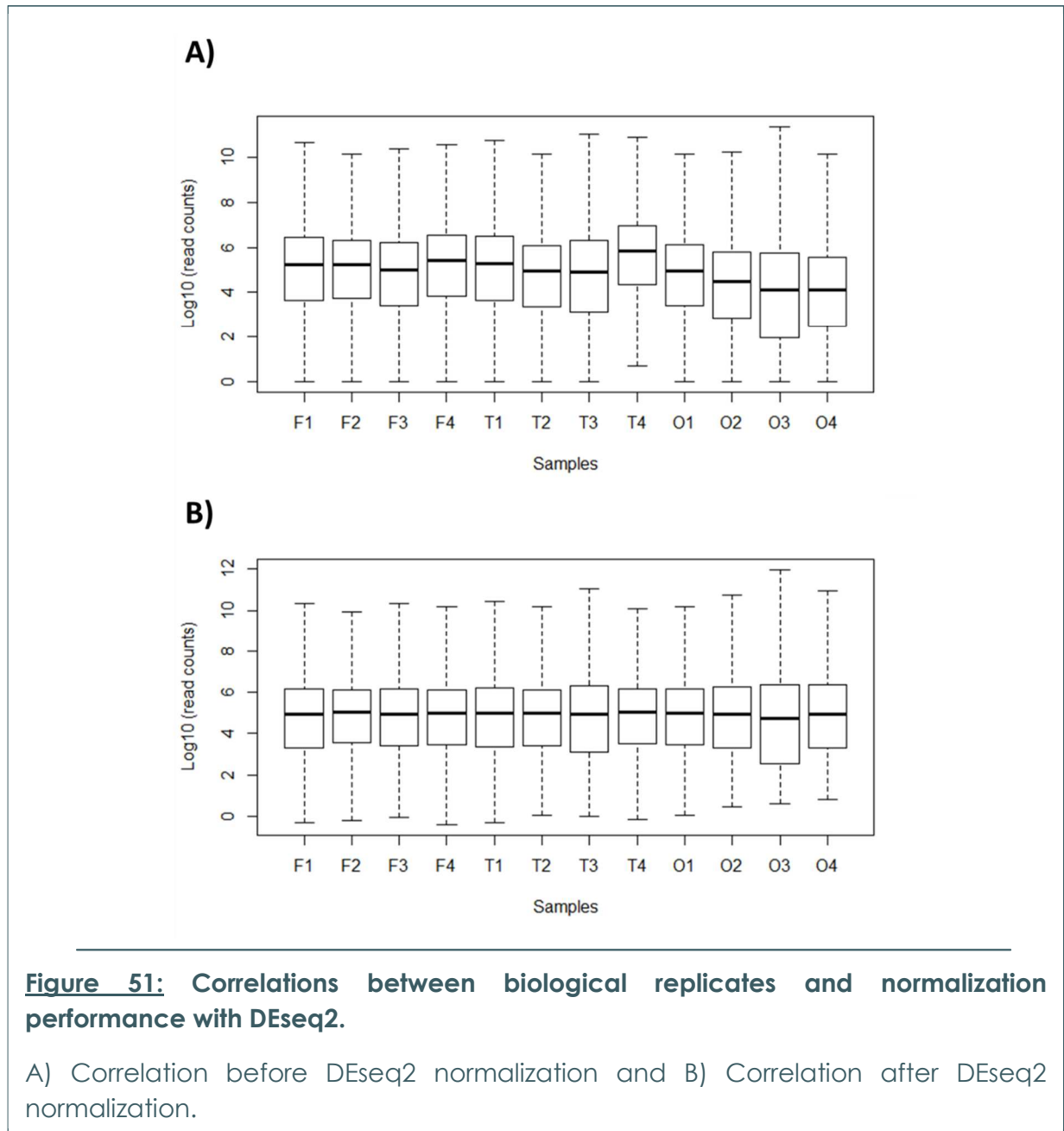


Figure 51: Correlations between biological replicates and normalization performance with DESeq2.

A) Correlation before DESeq2 normalization and B) Correlation after DESeq2 normalization.

As illustrated in **Figure 52 and Table 24**, biological replicates of the same conditions correlated together by using a distance matrix computation, thus demonstrated the robustness of our experimental design. As a consequence, the biological replicates were averaged for each specific morphotype (n=4) for the rest of the study. Finally, lists of differentially expressed genes (DEG) were generated using the DESeq2 tool (Love et al., 2014). Only the significantly DEG at an adjusted p-value below 0.01 were conserved.

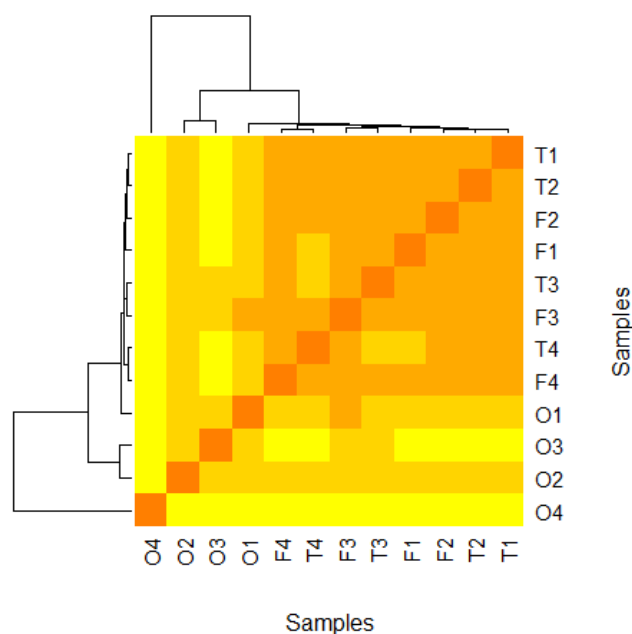


Figure 52: Heatmap illustrating the correlation between the 4 biological replicates for each culture specifically enriched in either the fusiform (F), triradiate (T) or oval (O) morphotype. The heatmap has been built using the distance matrix computation (Additional file 6).

	O1	O2	O3	O4	F1	F2	F3	F4	T1	T2	T3	T4
O1	0.0	1532481.6	1843759.0	5131383	520288.0	578130.4	495129.9	582376.2	556876.4	612720.5	627474.5	661492.1
O2	1532481.6	0.0	878881.8	4197895	1759598.7	1898608.2	1683391.5	1880525.0	1862167.4	1890419.9	1676816.6	1983021.7
O3	1843759.0	878881.8	0.0	4704847	2023886.2	2155174.0	1939502.1	2128028.1	2124432.6	2151191.3	1930803.1	2224048.4
O4	5131382.7	4197895.2	4704847.5	0	5264819.8	5391298.0	5314177.7	5430925.1	5347988.2	5372851.3	5259750.3	5497955.6
F1	520288.0	1759598.7	2023886.2	5264820	0.0	354992.3	412879.2	440533.0	286090.7	326604.7	473018.4	537771.9
F2	578130.4	1898608.2	2155174.0	5391298	354992.3	0.0	346686.7	211333.7	280036.4	220757.6	451288.3	258377.5
F3	495129.9	1683391.5	1939502.1	5314178	412879.2	346686.7	0.0	289535.2	393344.0	399601.6	366331.5	404873.4
F4	582376.2	1880525.0	2128028.1	5430925	440533.0	211333.7	289535.2	0.0	385623.2	345832.6	462094.0	181747.4
T1	556876.4	1862167.4	2124432.6	5347988	286090.7	280036.4	393344.0	385623.2	0.0	224377.6	478698.9	419906.4
T2	612720.5	1890419.9	2151191.3	5372851	326604.7	220757.6	399601.6	345832.6	224377.6	0.0	383333.3	383674.2
T3	627474.5	1676816.6	1930803.1	5259750	473018.4	451288.3	366331.5	462094.0	478698.9	383333.3	0.0	540446.1
T4	661492.1	1983021.7	2224048.4	5497956	537771.9	258377.5	404873.4	181747.4	419906.4	383674.2	540446.1	0.0

Table 24: Results of the matrix distance computation performed after the normalization with DEseq2.

Validation of the RNA-Seq results was performed by qRT-PCR analysing the expression of 22 genes selected randomly from the DGE lists (Table 25). The expression of these genes was normalized to 2 reference genes validated as described in chapter II. When compiling the results and plotting the log₂ ratio of the qRT-PCR in relation to the log₂ fold change from the RNA-Seq, a correlation value $R^2=0.799$ was obtained, thus attesting of the high quality of our transcriptomic

data (**Figure 53**). For example, the Phatr3_J35771 gene was up-regulated 6.4 and 7.5 log₂ fold change according to the RNA-Seq and qRT-PCR analyses respectively in the OF comparison. Similar results was observed for the Phatr3_J39760 gene which is down-regulated in the TF pairwise comparison with a -3.5 and 2.3 log₂ fold change in RNA-Seq and qRT-PCR analyses respectively. In contrast, this Phatr3_J39760 gene is up-regulated with a 1.7 log₂ fold change by RT-qPCR and a 1.2 log₂ fold change according to the RNA-Seq data when O is compared to F and a 4 log₂ fold change both for qRT-PCR and RNA-Seq when the OT comparison is concerned (**Table 25**).

	T/F		O/F		O/T	
	RNA-seq	qRT-PCR	RNA-seq	qRT-PCR	RNA-seq	qRT-PCR
	log ₂ fold change	log ₂ ratio	log ₂ fold change	log ₂ ratio	log ₂ fold change	log ₂ ratio
Phatr3_J38708	-2,824	-0,379090256	5,038	5,90125084	7,862	6,280341096
Phatr3_J47653	0,383	1,75578349	-5,538	-4,964945406	-5,921	-6,720728896
Phatr3_EG00182	0,375	-0,391623988	-6,431	-6,658256361	-6,806	-6,266632373
Phatr3_J54574	-1,769	-1,265992662	5,335	6,07422989	7,103	7,340222552
Phatr3_J12405	-3,129	-0,791861313	0,477	2,37514324	3,606	3,167004553
Phatr3_J44251	-2,312	-1,597432819	0,808	3,808133139	3,121	5,405565958
Phatr3_EG02507	-1,318	-0,811249332	5,494	7,30823081	6,812	8,119480141
Phatr3_EG01990	-1,847	-1,606072643	0,808	2,12231552	2,656	3,728388163
Phatr3_Jdraft1583	1,041	0,527787381	5,285	4,338316202	4,244	3,810528821
Phatr3_J48834	-2,096	-2,874953993	5,611	7,163920882	7,707	10,03887487
Phatr3_J49336	-2,684	-3,377307746	0,590	1,068683838	3,274	4,445991585
Phatr3_J35771	0,330	1,000604251	6,439	7,502492421	6,109	6,50188817
Phatr3_J33783	-1,314	1,589473355	6,083	8,299373531	7,398	6,709900177
Phatr3_J48356	1,175	0,703060539	8,627	4,958996039	7,452	4,2559355
Phatr3_J45621	-3,108	-2,830714169	-0,539	-0,971009778	2,568	1,859704392
Phatr3_EG00085	-3,017	0,987287048	3,267	4,23256625	6,284	3,245279202
Phatr3_J41599	-0,978	-1,280175063	6,286	8,338235442	7,265	9,618410505
Phatr3_J32459	-3,202	-4,538447541	0,768	-0,025093948	3,969	4,513353593
Phatr3_J33840	-2,483	-0,829839263	5,122	6,182594121	7,605	7,012433384
Phatr3_J39760	-3,448	-2,320517429	1,213	1,785186424	4,660	4,105703853
Phatr3_EG02393	-0,499	1,919298685	-0,166	8,67653821	0,333	6,757239525
Phatr3_J50304	2,323	3,919462295	-0,550	0,149828227	-2,873	-3,769634068

Table 25: Comparative table highlighting the RNA-Seq log₂ fold change and the qPCR log₂ ratio for the 22 genes randomly selected among the DGE list obtained from the TF; OF; OT pairwise comparison of the RNA-Seq analysis.

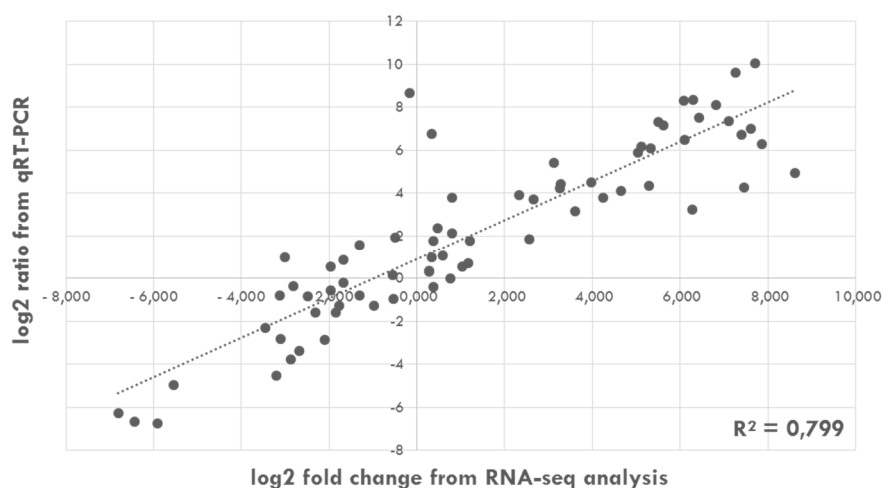


Figure 53: Validation of the RNA-Seq data using qRT-PCR analysis.

Correlation between the log₂ fold change extracted from the RNA-Seq data and the log₂ ratio obtained by qRT-PCR analysis for 22 genes randomly selected from the DEG lists.

4.3.2 Gene differential expression analysis

Pairwise comparisons of the morphotype's gene expression data were performed by considering that the fusiform morphotype (F) was the reference because it is the most common morphotype found in culture (De Martino et al., 2007; Zhao et al., 2014). The gene expression profile from the triradiate morphotype (T) was compared first to the one of the fusiform (T *versus* F, noted TF) as these two morphotypes have been previously considered to be the two basic morphotypes in agitated liquid media (Bartual et al., 2008). Among 12 320 genes described in *P. tricornutum*'s genome including 12 178 coding sequences and 142 non-coding, only 120 genes were significantly differentially expressed when comparing TF representing less than 1% of the transcriptome (**Table 26**).

Please follow the link below for downloading the table

<https://ent.normandie-univ.fr/filex/aet?k=eYzHKnfzFz5BdmGrdKs>

Table 26: List of the genes significantly differentially expressed (adjusted p-value below 0.01) in the OF, OT and TF pairwise comparisons.

As a consequence of this result, considering the small difference observed between T and F, the oval morphotype (O) were thereafter compared to F (OF) and T (OT) respectively. The pairwise comparison of OF highlighted that 2,692 genes were significantly differentially expressed between the two morphotypes, thus represented almost 22% of the transcriptome (**Table 26**). For the OT comparison, 3,553 genes were described to be significantly differentially expressed representing up to 29% (**Table 26**). Such results suggest that wide reprogramming of genes is occurring in the oval morphotype as compared to the others. Volcano plot for TF shown that 100 genes are down-regulated in the triradiate morphotype with the majority inside the interval of the log₂ fold change values comprised between 2 to 5 (**Figures 54 and 55**).

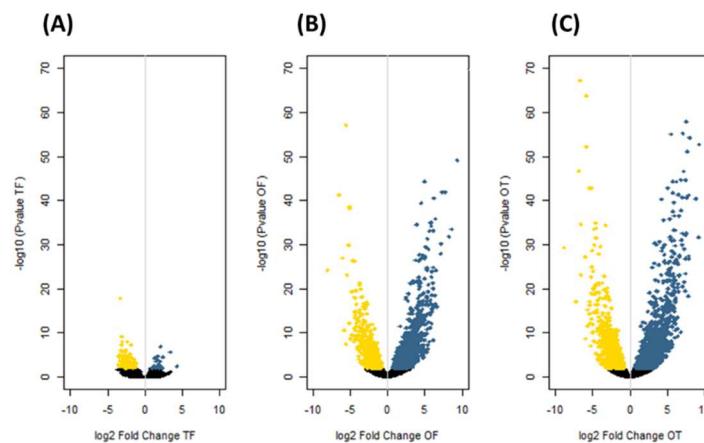


Figure 54: Volcano Plots of pairwise comparisons of (A) T versus F (TF); (B) O versus F (OF); (C) O versus T (OT).

The yellow dots represent down-regulated genes. In contrast, cyan dots represent up-regulated genes. Dark dots represent the non-significantly DE genes.

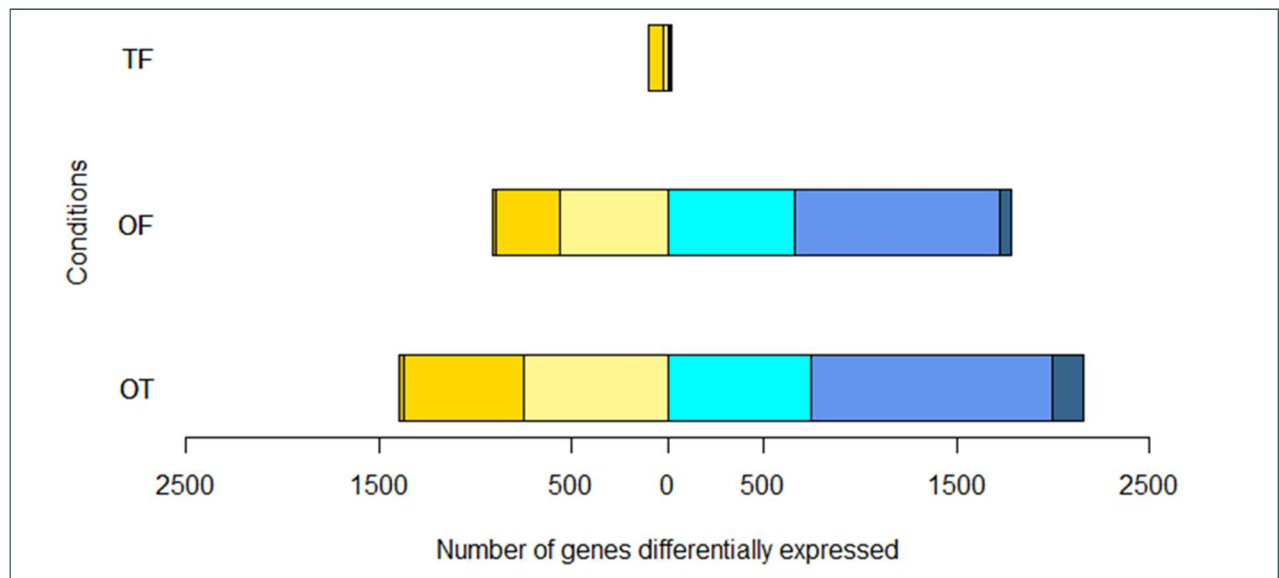


Figure 55: Graph representing the distribution of the significantly differentially expressed genes (DEG) in relation to their log₂ fold change values.

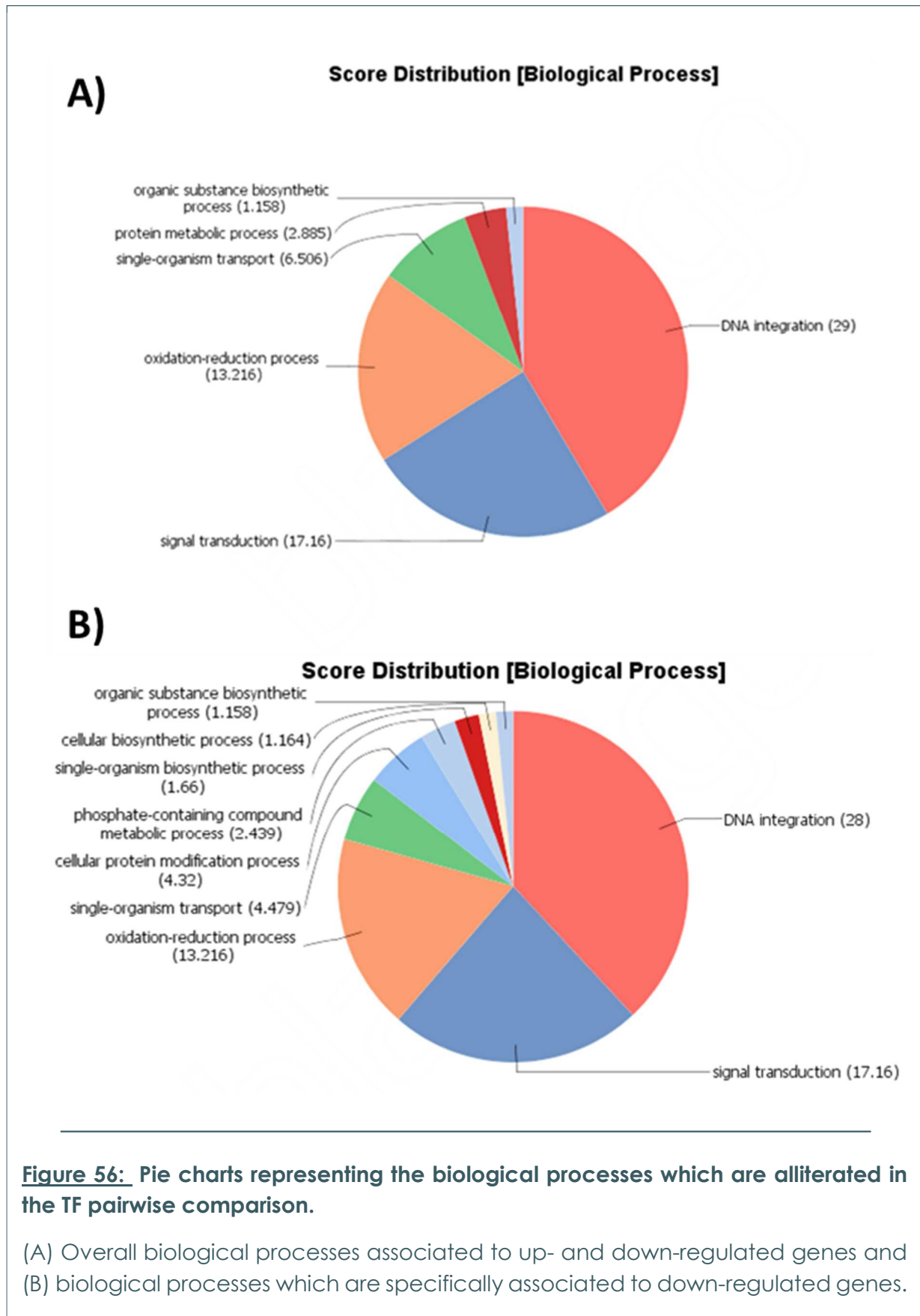
Blue squares represent up-regulated genes with a log₂fold change value comprised between 0 to 2 for the light blue, 2 to 4 for the median blue and superior to 5 for the dark blue. Yellow squares represent the down-regulated genes with different color intensities which are corresponding to the value of log₂fold change, log₂fold change value comprised between 0 to 2 for the light yellow, 2 to 4 for the median yellow and superior to 5 for the dark yellow.

When focusing on the OF comparison, 906 genes are down-regulated and 1 786 genes are up-regulated in the oval morphotype. As far as the up-regulated genes are concerned in the OF DEG list, 662 genes are differentially expressed at less than a value of 2 for the log₂fold change and 1 059 being differentially expressed at a log₂ fold change comprise between 2 to 5. Only few genes are up-regulated at a value greater than 5. The down-regulated genes of OF are mostly higher than the log₂fold change value -5, with only 11 genes being differentially expressed at a log₂fold change lower than -5, 332 genes comprised between 2 to 5 and 563 genes having a log₂fold change comprise between 0 to 2. The OT comparison reveals 3 553 DEG with 1 396 being down-regulated and 2 157 being up-regulated. When looking at the log₂ fold change value, 744 of the up-regulated genes are in the 0-2 range for their log₂ fold change value, 1 255 genes are between 2 to 5 and 158 genes are higher than 5. The down-regulated genes possess 746 genes comprised between 0 to 2 and 624 genes comprised in the interval 2 to 5. Only 26 genes are down-regulated at a log₂ fold change value greater than 5 for the OT comparison (**Figures 54 and 55**).

4.3.3 Functional profiling analysis

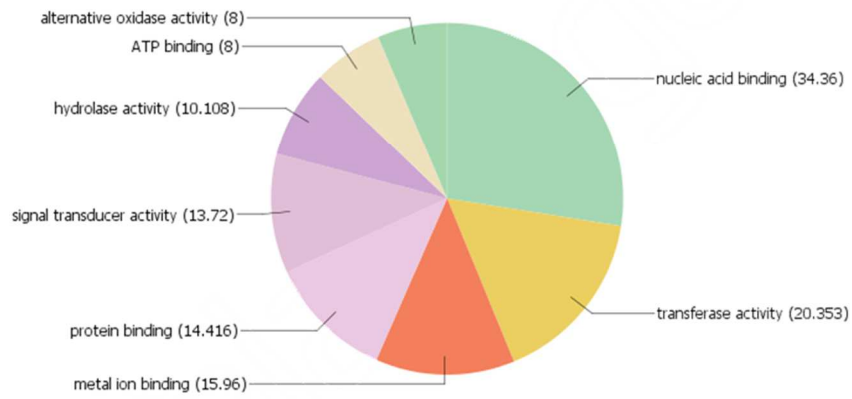
In order to characterize the biological process (BP) and molecular function in which DEG are involved, a gene ontology (GO) term analysis was performed using Blast2GO (Conesa et al., 2005; see experimental procedure for details). Above 65 to 70 % of the transcripts from RNA-Seq data have been annotated with GO terms. The pie chart of the TF comparison shown that some biological process are regulated (**Figure 56**). Among them, DNA integration represent about 40% of the DEG genes which are almost all down-regulated. Signal transduction BP is enriched to 25%, being also down-regulated. The oxidation-reduction process is also appearing as 25% of the overall down-regulated genes. Genes dedicated to single-organism transport accounted for 12.5% of the DEG and are also mainly down-regulated. Finally, protein metabolic process and organic substance biosynthetic process are completed the overall and down-regulated pie charts (**Figure 56**). Similar results were obtained when considering the score distribution at the molecular function level (**Figure 57**). Indeed, for example, above 40% of the down-regulated genes are corresponding to nucleic acid binding activity, 20% coincide with signal transducer activity and about 15-20% of the down-regulated genes are representing by alternative oxidase activity. Altogether, these results reinforce the ones achieved for the biological process.

As already mentioned, 2 692 genes are differentially expressed in the OF pairwise comparison with 906 genes down-regulated and 1 786 genes up-regulated. Among these genes, above 25% are corresponding to single-organism cellular process with the majority being up-regulated, 12.5% belong to the gene expression and 12.5% to the single-organism transport (**Figure 58**). As far as gene expression BP is concerned, similar percentage of genes can be found in the up- or down-regulated categories. In contrast, RNA metabolic process seems to be more down-regulated. Genes involved in the regulation of cellular process are equally distributed between the up- and down-regulated genes. Similar repartition is observed for genes involved in the cellular protein modification and oxidation-reduction process. In contrast, genes implicated in the cellular macromolecules biosynthetic process and organonitrogen compound metabolic process are all up-regulated whereas cellular nitrogen compound biosynthetic process are down-regulated (**Figure 58**). The analysis at the molecular function level shown that transferase activity, hydrolase activity, nucleic acid binding, metal ion binding and oxidoreductase activity are equally classified in the up- and down-regulated sub-categories (**Figure 59**). In contrast, genes dedicated to the purine binding (purine ribonucleotide binding, purine ribonucleoside triphosphate binding and purine ribonucleoside binding) are all down-regulated (**Figure 59**).

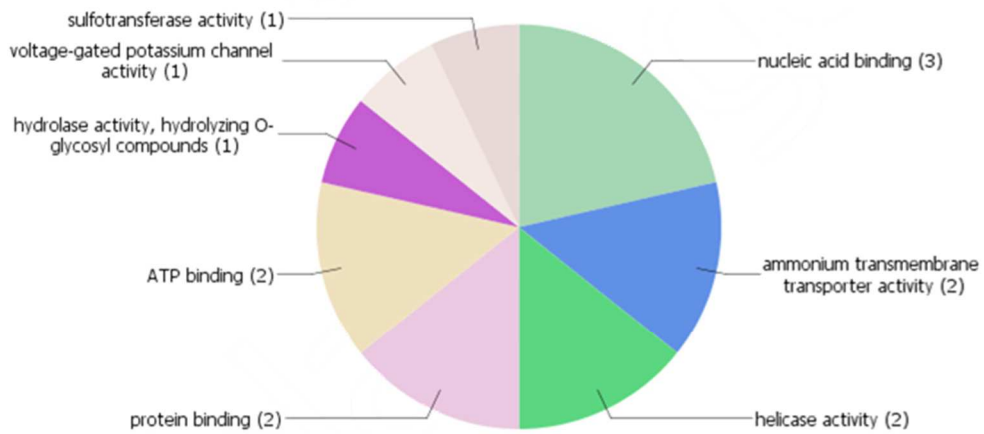


A)

Score Distribution [Molecular Function]



B)



C)

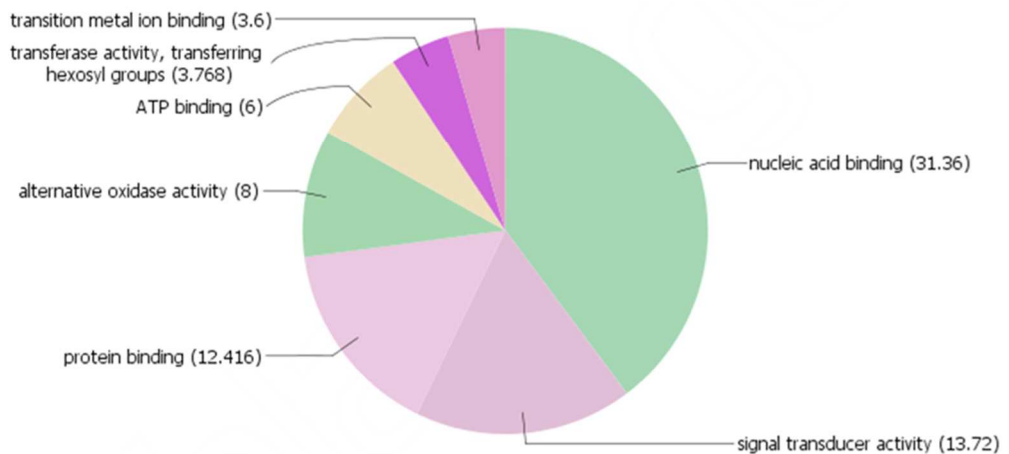


Figure 57: Pie charts representing the molecular functions which are alliterated in the TF pairwise comparison.

(A) Overall of the molecular functions associated to up- and down-regulated genes; (B) molecular functions associated to up-regulated genes and (C) molecular functions associated to down-regulated genes.

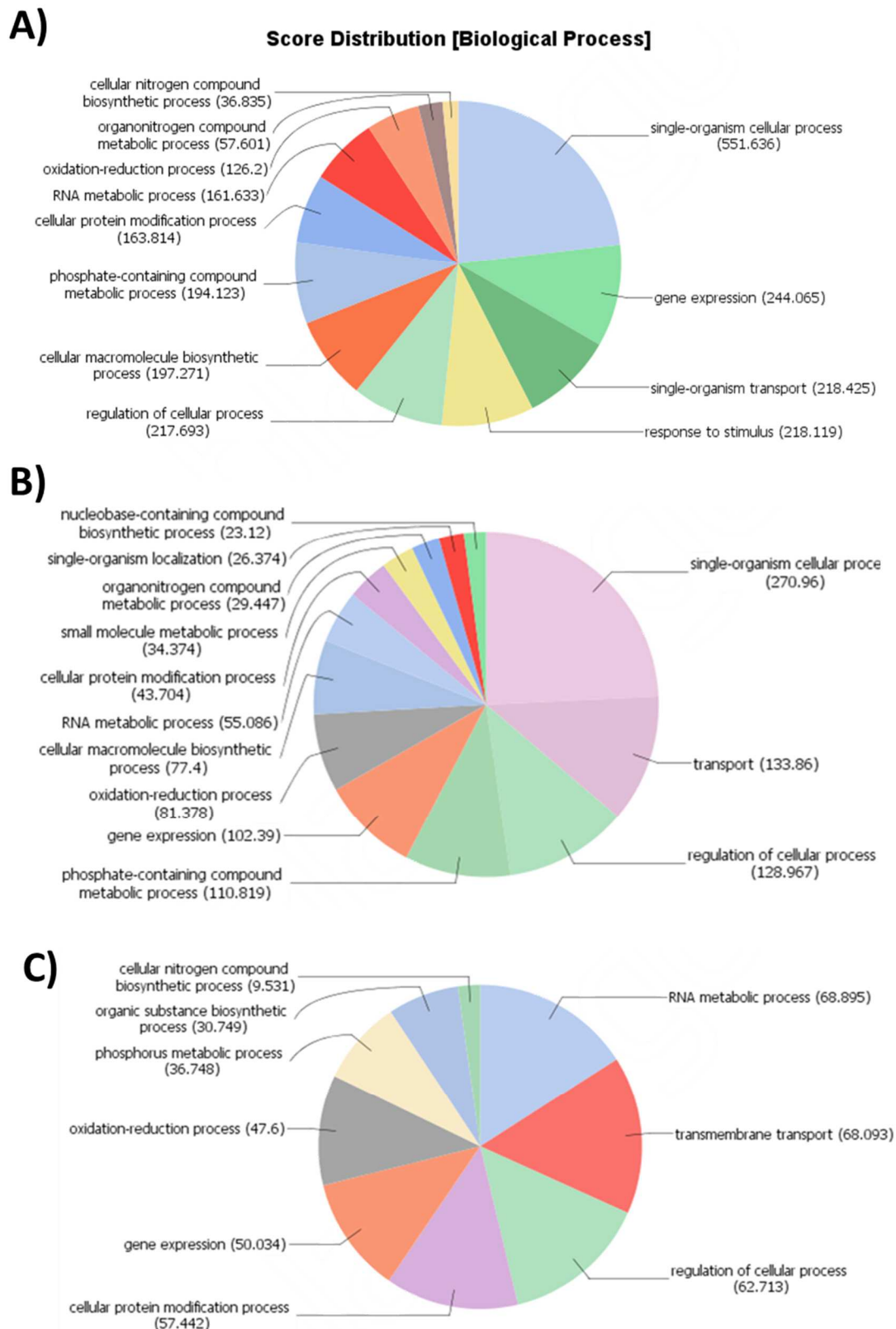


Figure 58: Pie charts representing the biological processes which are alliterated in the OF pairwise comparison.

(A) Overall biological processes associated to up- and down-regulated genes and (B) biological processes which are specifically associated to up-regulated genes, (C) biological processes associated to down-regulated genes.

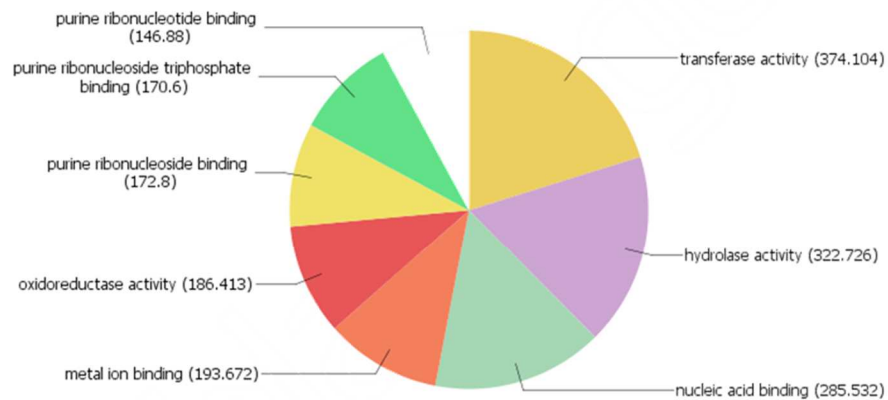
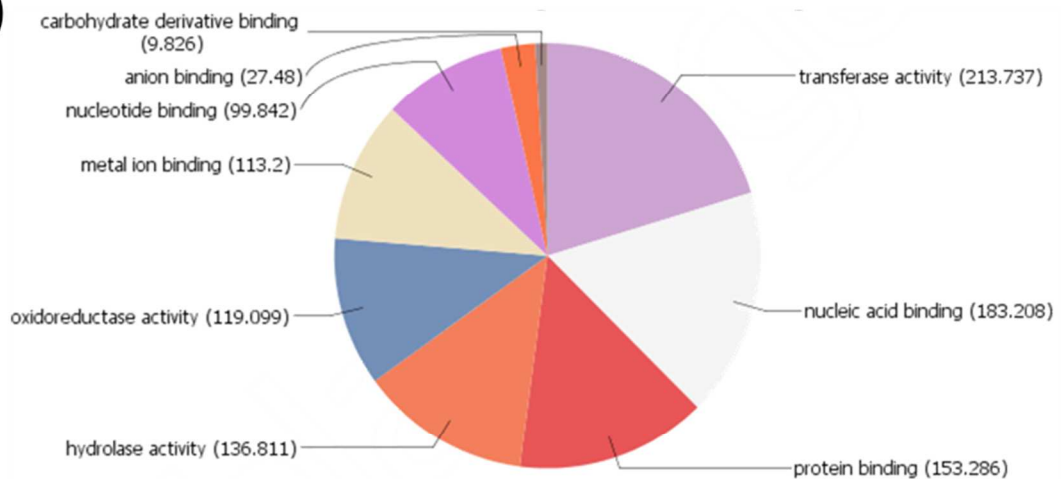
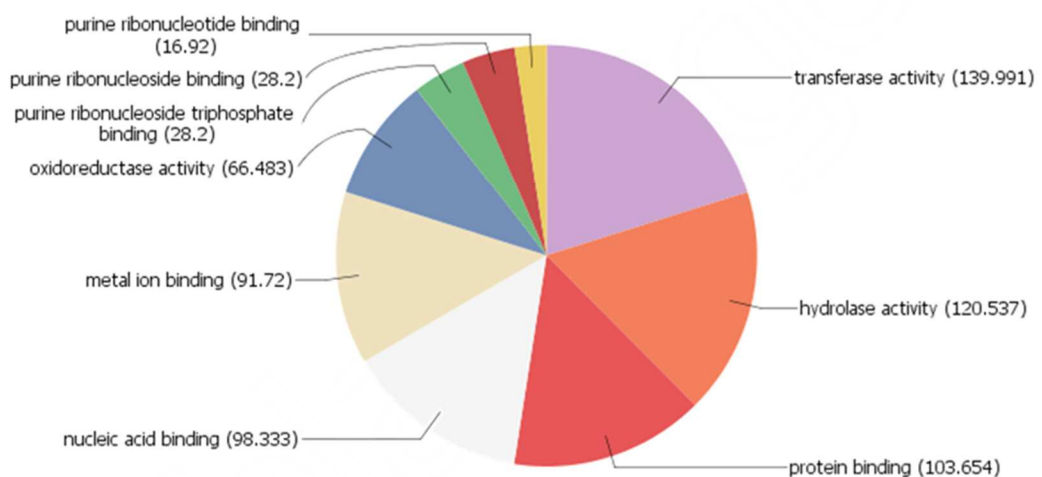
A)**Score Distribution [Molecular Function]****B)****C)**

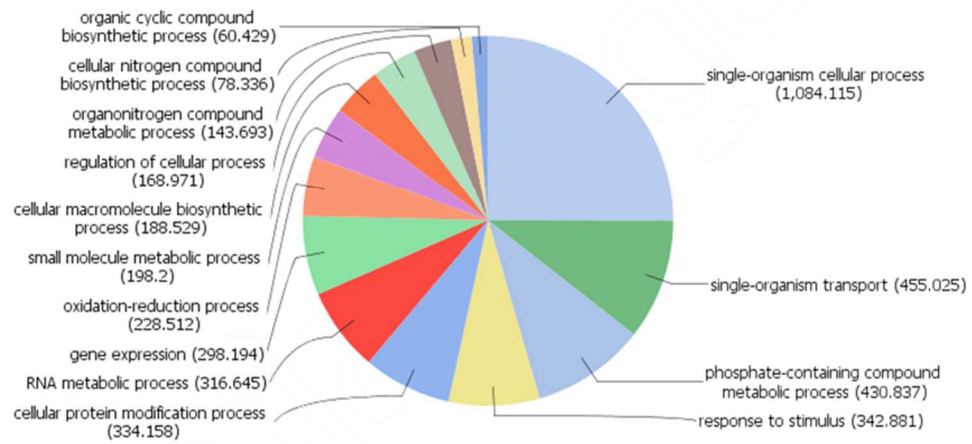
Figure 59: Pie charts representing the molecular functions which are alliterated in the OF pairwise comparison.

(A) Overall of the molecular functions associated to up- and down-regulated genes; (B) molecular functions associated to up-regulated genes and (C) molecular functions associated to down-regulated genes.

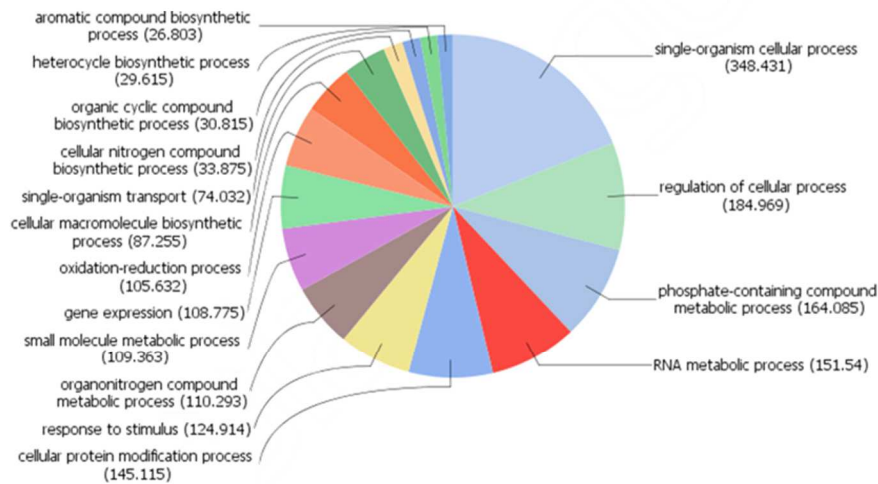
Among the 29% of the transcriptome thus representing 3 553 genes which are significantly differentially expressed in the OT pairwise comparison, 25% correlated with single-organism cellular process. Such BP seems to be mainly up-regulated in the OT condition (**Figure 60**). Some of the DEG are implicated in single-organism transport being mostly up-regulated as it is for the phosphate-containing compound metabolic process, response to stimulus, cellular protein modification process, cellular macromolecule biosynthetic process, cellular nitrogen compound biosynthetic process (**Figure 60**). By opposition, oxidation-reduction process is equally distributed between the up- and down-regulated genes and the gene expression BP is exclusively down-regulated. When looking at the molecular function analysis, it appears that the hydrolase activity, the nucleic acid binding and the oxidoreductase activity are equally represented in both the up- and down-regulated genes lists (**Figure 61**). The transferase activity is only present in the up-regulated genes category. An opposite situation appears for the ATP binding which are only down-regulated. The metal ion binding function are present in both case with a majority of genes being up-regulated (**Figure 61**).

A)

Score Distribution [Biological Process]



B)



C)

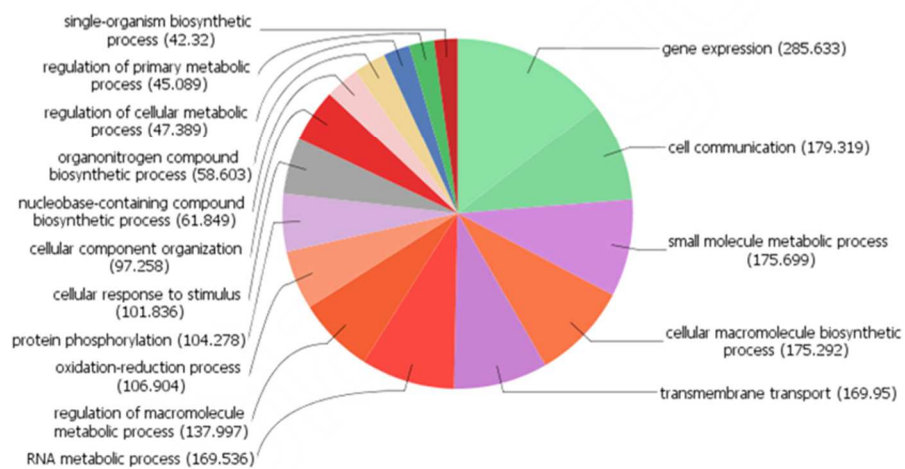
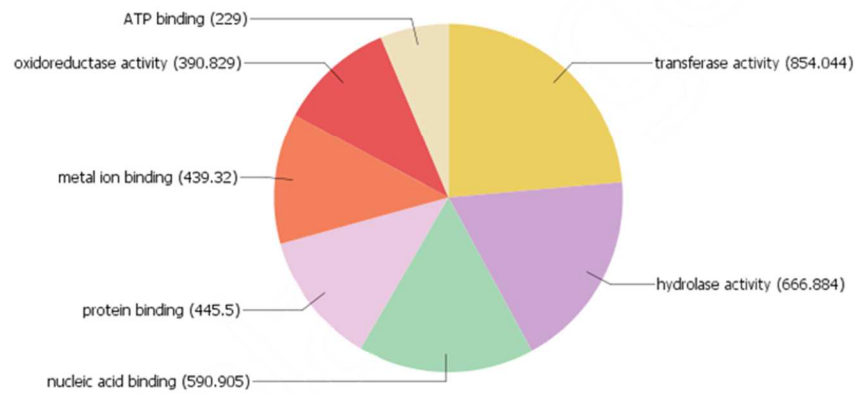


Figure 60: Pie charts representing the biological processes which are alliterated in the OT pairwise comparison.

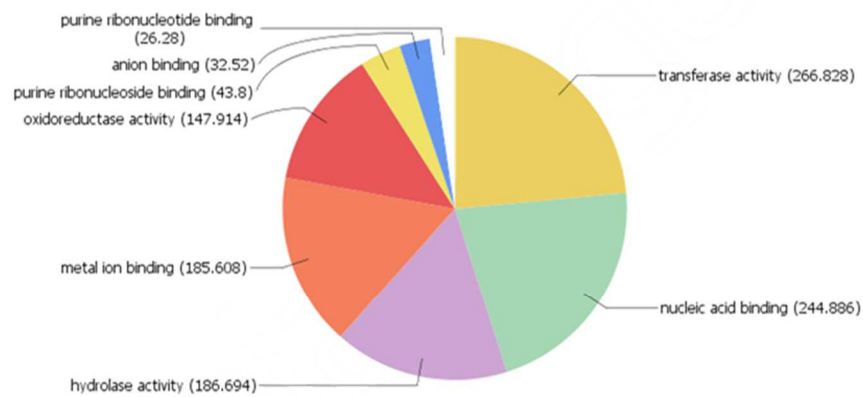
(A) Overall biological processes associated to up- and down-regulated genes and (B) biological processes which are specifically associated to up-regulated genes, (C) biological processes associated to down-regulated genes.

A)

Score Distribution [Molecular Function]



B)



C)

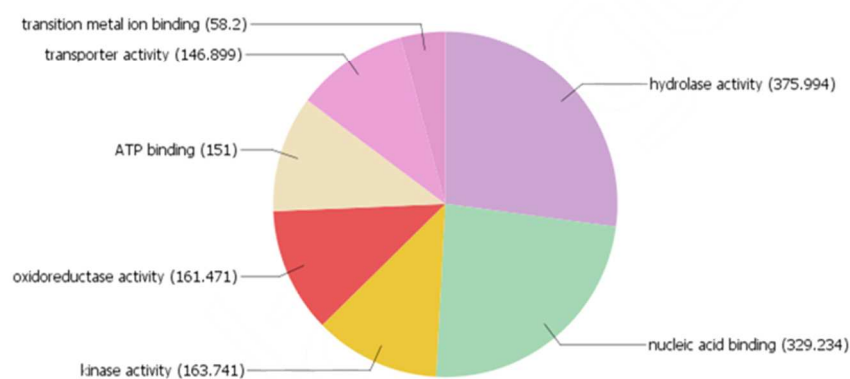
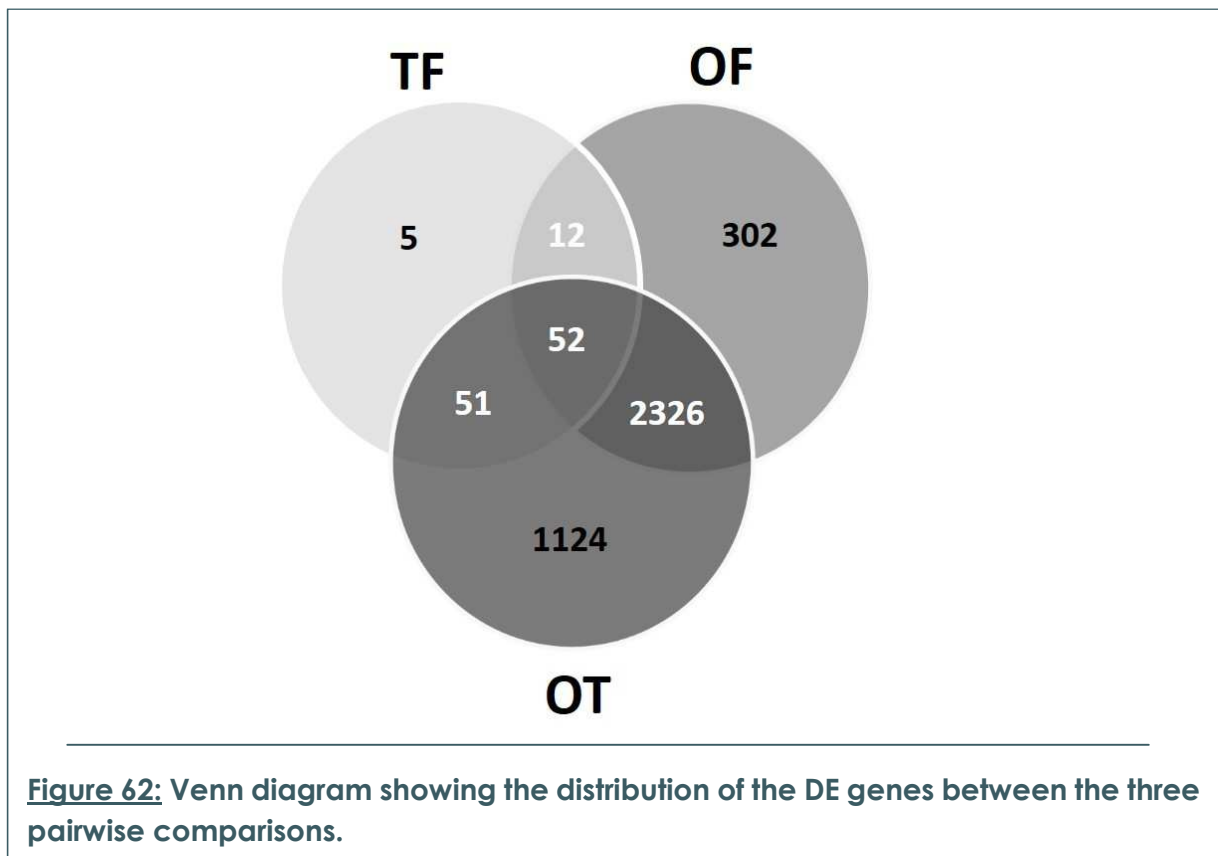


Figure 61: Pie charts representing the molecular functions which are alliterated in the OT pairwise comparison.

(A) Overall of the molecular functions associated to up- and down-regulated genes; (B) molecular functions associated to up-regulated genes and (C) molecular functions associated to down-regulated genes.

4.3.4 Intersection analysis reveals specific pathways which are up-regulated in the oval morphotype

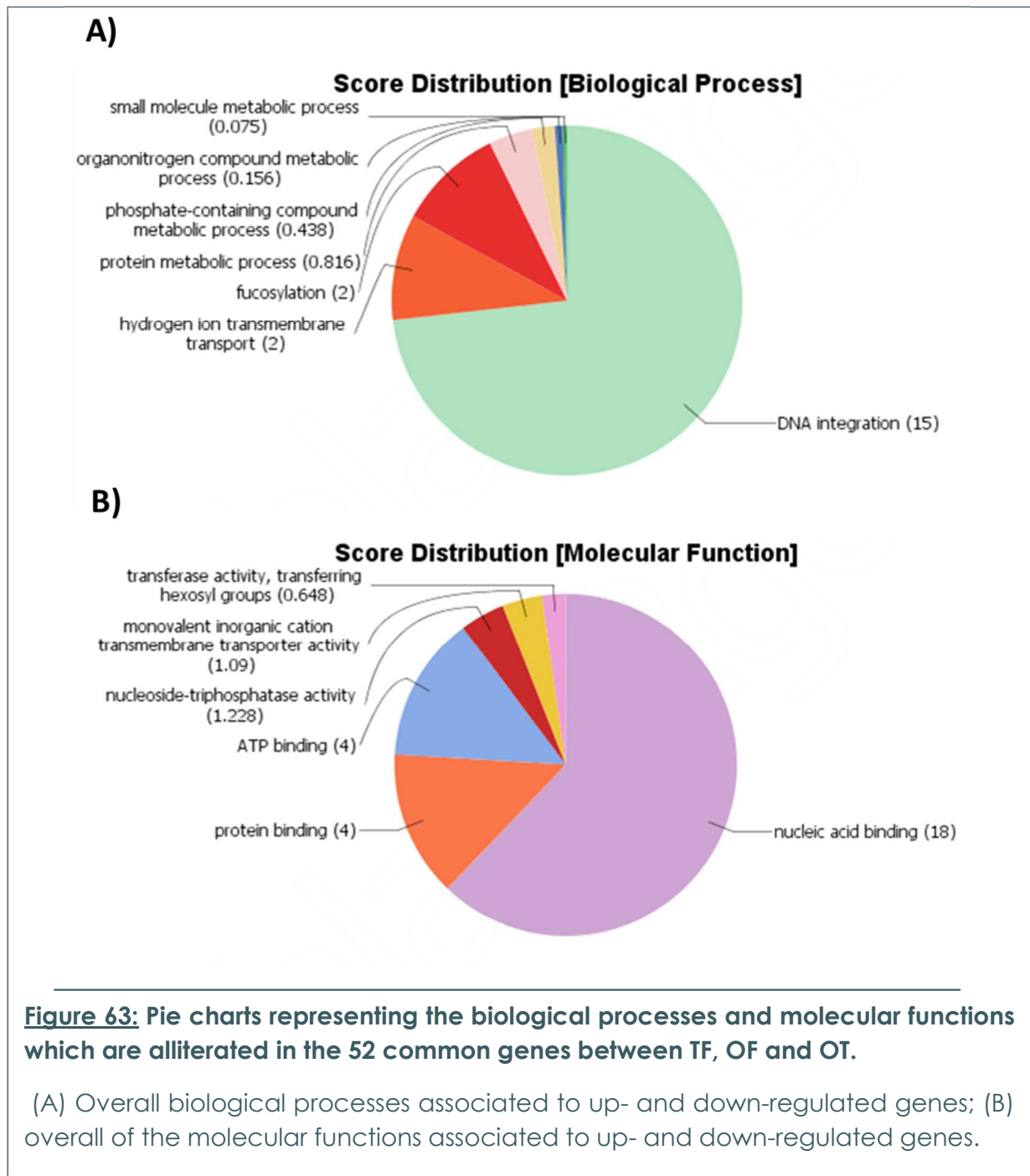
Intersection analysis was performed in order to identify specific genes which can be attributed to one morphotype. The resulting Venn diagram is presented in the **Figure 62**. It demonstrated that 52 genes are common to all pairwise comparisons. These genes are differentially expressed in a common manner for the three morphotypes. However, when looking precisely at the log₂Fold Change, this one is reflecting a specific gene expression level for each morphotype (**Table 27**). Among the list, 17 genes were not annotated with any GO term. 18 genes out of the 52 are encoding for proteins related to nucleic acid binding functions with a very high percentage (almost 70%) being DNA polymerases (Pol proteins). This fact is well illustrated in the score distribution pie charts which came out from the analysis of the BP and molecular function distribution through Blast2GO analysis. Indeed, DNA integration and nucleic acid binding occupy the biggest part of the pie charts (**Figure 63**).



Please follow the link below for downloading the table

<https://ent.normandie-univ.fr/filex/aet?k=Rk4ufUGpfCOA3aHQxWs>

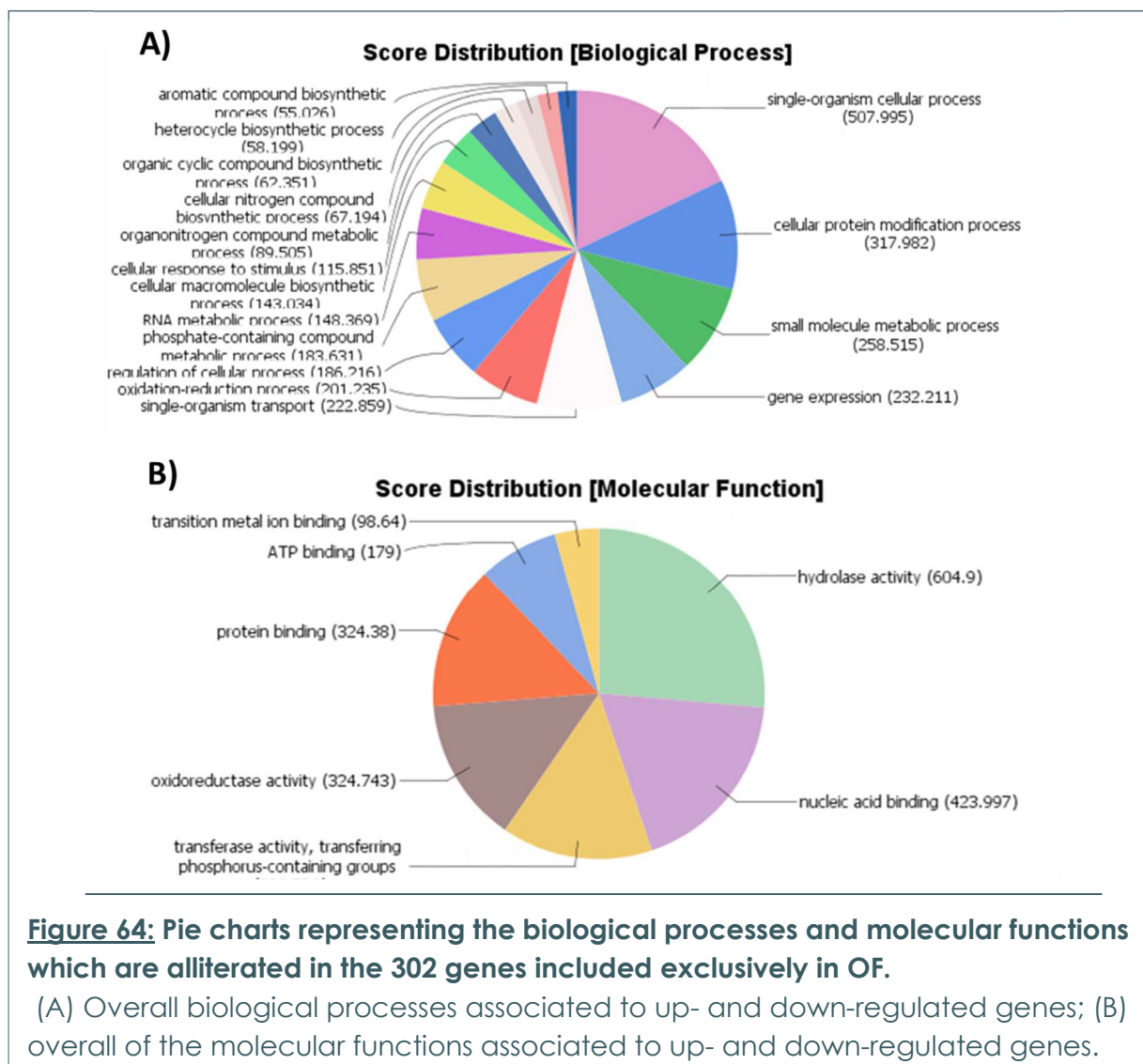
Table 27: Lists of the significantly differentially expressed genes associated with GO terms which have been obtained from the Venn diagram as a result of the intersection analysis



Coming back to the Venn diagram (**Figure 62**), it is noticeable that only 5 genes came out to be specific of the TF pairwise comparison. Among the GO terms associated, transferases activity, carbohydrate and nucleic acid binding proteins can be found. The TF and OF comparisons overlap and present 12 common genes which are either associated to any GO term (5 genes) or relative to signal transduction and DNA integration process (5 genes). Similarly, 51 genes show up to overlap between TF and OT (**Table 27**), thus being specific from the triradiate morphotype. 18 out of the 51 genes are encoding for putative proteins which currently do not have any associated GO terms.

Among the rest, signal transduction and DNA integration represent almost 45% of both the BP and molecular function distributions. Oxidation reduction process, transmembrane transport and protein phosphorylation are also enriched (**Table 27**).

In order to identify genes which are specifically down- or up-regulated and pathway affected in the oval morphotype, a particular attention has been paid to the analysis of the 2 336 DEG which are at the intersection of the OF and OT pairwise comparisons (**Table 27**). Among this DGE list, 68% of the genes were up-regulated. For this dataset, a GO and a pathway analysis were ran in parallel independently. Results of the GO analysis and enrichment for both the biological process and molecular functions are summarized in the **Figure 64**. The pie charts shown that the single-organism cellular process, the cellular protein modification process, the small molecule metabolic process and gene expression represent major biological processes which are modified within the oval morphotype from *P. tricornutum*. The analysis at the molecular function level highlight changes in the hydrolase activity, nucleic acid binding, transferase activity, oxidoreductase activity, protein and ATP binding as well as transition metal ion binding.



In parallel, genes encoding for known enzymes in the *P. tricornutum*'s genome were mapped in KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways using both B2GO and pathview (Luo and Brouwer, 2013). Among the 2 326 genes from the DGE list, 1 862 are possessing a Uniprot annotation and only 1 483 sequences are having an attributed EC numbers. The pathway analysis supports a putative activation of key primary metabolism pathways such as the glycolysis, the pentose phosphate pathway and the biosynthesis of nucleotides and triglycerides in the oval morphotype of *P. tricornutum* (**Table 27**). Indeed, the depolymerisation of chrysolaminarin, the diatom storage $\beta(1,3)$ -glucan, by a set of glucanases into free glucose and then its subsequent phosphorylation by hexokinase into glucose 6-phosphate allows to supply the glycolysis and the pentose phosphate pathway in carbon sources. Then, overexpression of the glyceraldehyde-3-phosphate dehydrogenase, that catalyzes the synthesis of glycerate 1,3 bis-phosphate in the glycolysis, as well as the overexpression of sub-units of the pyruvate dehydrogenase allow the conversion of glucose 6-phosphate into acetyl CoA for the carbon supplementation of the citrate cycle. In addition, nucleotide biosynthesis is likely triggered in the oval morphotype of *P. tricornutum*. For instance, overexpressions of the glucose-6-phosphate dehydrogenase, 6-phosphogluconolactonase and 6-phosphogluconate dehydrogenase of the cytosolic pentose phosphate pathway reflect the activation of the ribose phosphate biosynthesis that is required for the synthesis of both nucleotides and deoxynucleotides (**Table 27**). Genes encoding key enzymes involved in the biosynthesis of purines and pyrimidines are also overexpressed in the oval morphotype.

Analysis of the expression of genes encoding enzymes of the lipid pathways mainly revealed a down-regulation in the oval morphotype of genes encoding enzymes specific from the fatty degradation together with the overexpression of the diatom phosphatidic acid phosphatase as well as monoacyl and diacylglycerol O-acyltransferase. Thus, this suggests that triglyceride biosynthesis is activated in this morphotype. Among other gene family exhibiting differential expression between *P. tricornutum* morphotypes, it is worth noting that a mannose kinase is overexpressed in the oval diatom. This enzyme is able to supply the cytosol in mannose 6-phosphate that is required for the biosynthesis of GDP-Man nucleotide sugar. This may support an active biosynthesis of the glucuronomannans in this morphotype. Indeed, glucuronomannans have been previously described to be the main polysaccharide component of *P. tricornutum*'s cell wall (Abdullahi et al., 2006; Chiovitti et al., 2005; Willis et al., 2013). In addition, the gene encoding for the glucosamine fructose 6 phosphate aminotransferase (EC 2.6.1.16) is down-regulated, therefore limiting the synthesis of L-glutamate and D-glucosamine-6-phosphate. Meanwhile, the amidophosphoribosyltransferase (EC 2.4.2.14) is up-regulated meaning that the L-glutamine is used to synthesize 5-phospho

ribosylamine which enter into the purine metabolism. This example illustrate again that the oval morphotype promotes purine synthesis pathway at the expense of the aminosugars metabolism.

The metabolism of some of the amino acids is also modified within the oval morphotype (**Table 27**). For example, the Serine biosynthesis is up-regulated whereas its transformation into glycine is down-regulated leading overall to an increase of Serine. The global pool of Cysteine is also boost. Indeed, the Serine is transformed in L-Cysteine through the overexpression of the gene encoding for the cysteine synthase A (EC 2.5.1.47) and the formation of the acetyl-L-Serine intermediate. The synthesis of the D-cysteine is intensify through the up-regulation of the gene encoding for an amino acid racemase (EC 5.1.1.10). Beside the regulation of the Serine and Cysteine metabolism, some of the steps of the Arginine biosynthesis are also down regulated (EC 2.6.1.1; EC 1.4.1.1; EC 2.1.3.3) leading to a decrease of glutamate and citruline which are both precursors of Arginine and Aspartate. The gene encoding for the acetalactate synthase I/II/III large subunit (EC 2.2.1.6) is also up-regulated. This enzyme transforms pyruvate in S-2-aceto 2 hydroxybutanoate, an intermediate of the L-isoleucine biosynthesis. The tyrosine aminotransferase (EC 2.6.1.5), the aromatic amino acid transaminase (EC 2.6.1.57), the aspartate transferase (EC 2.6.1.1) and the histidinol-phosphate aminotransferase (EC 2.6.1.9) are down regulated. All together, these steps induce a reduction of the accumulation of L-glutamate, thus favoring the synthesis of Phenylalanine and Tyrosine. In agreement to this and the impact observed on Isoleucine metabolism, the overexpression of genes encoding for the phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20) and the isoleucyl-tRNA synthetase (EC 6.1.1.5) can be noticed. In complement, an overexpression genes encoding respectively for the alanyl-tRNA synthetase (EC 6.1.1.7), the lysyl-tRNA synthetase class I (EC 6.1.1.6) and the methionyl-tRNA synthetase (EC 6.1.1.10), even if the Lysine and Methionine metabolism are not affected. In constrast, genes encoding for the aspartyl-tRNA synthetase (EC 6.1.1.12) and the asparaginylyl-tRNA synthetase (EC 6.1.1.22) appears to be significantly down-regulated. As the aminoacids metabolism and aminoacyl tRNA biosynthesis was perturbed, we decided, then, to focus our analysis on the protein metabolism. The path view analysis demonstrated that the cellular protein modification process are up-regulated, thus confirming the previous BP and molecular function studies. Indeed, such analyses demonstrated that several actors involved in the protein degradation processing are up-regulated (**Figure 65 and Annexe 3**).

Among them, genes involved in protein processing in the endoplasmic reticulum (ER) are up-regulated (**Figure 65A**). This is the case for the Sec61/Sec62/Sec63 complex which are membrane protein translocators allowing the translocation of a neo-synthesized polypeptide within the ER (Conti et al., 2015; Meyer et al., 2000). In parallel, genes like Ubc6/7, Derlin, skp1 encoding for proteins belonging to the Ubiquitin ligase complex (Ciechanover et al., 2000; Glickman and Ciechanover, 2002) and genes encoding for actors of the ER-associated degradation (Derlin, NEF, sHSF, DUB) that detects misfolded proteins in the ER and targets them for destruction (Vembar and Brodsky, 2008) are also up-regulated. In addition, the ubiquitin-mediated proteolysis mechanism is also up-regulated (**Figure 65B**) as well as some genes encoding for protein subunits of the proteasome (**Figure 65C**). Thus, taken together, those results suggests that the protein degradation metabolism is much more active in the oval morphotype as compared to the others and it can be hypothesized that like in other eukaryotes, Sec61-Sec62-Sec63 complex might also, in *P. tricornutum* be responsible of the backward transport of ER proteins in order to submit them to the ubiquitin-proteasome dependent degradation pathway. We noticed that in the KEGG pathway corresponding to the protein processing in the ER, the gene encoding for SAR1, a GTPase specifically found in COP II vesicles, were up-regulated (**Figure 65A**). Therefore, we decided to focus on the vesicular transport (**Figure 66**) and observed that the gene encoding for Bet1, a Golgi vesicular membrane trafficking protein was up-regulated. This protein is a Golgi-associated membrane protein that participates in vesicular transport from the ER to the Golgi complex. In human, it has been demonstrated that this protein functions as a SNARE involved in the docking process of ER-derived vesicles with the *cis*-Golgi membrane (Morsomme et al., 2003; Parlati et al., 2000). Additionally, Sec22, a gene encoding for a SNARE involved in targeting and fusion of ER-derived transport vesicles with the Golgi complex as well as Golgi-derived retrograde transport vesicles with the ER (Lewis et al., 1997; Liu and Barlowe, 2002) is up-regulated. This particular SNARE has been recently described to be responsible for the regulation of the ER morphology in *Drosophila* (Zhao et al., 2015). Therefore, once can hypothesize that the ER morphology of *P. tricornutum*'s oval morphotype needs to be maintain to face the overactivity of the cell machinery. In our experiment, the gene encoding for the Syntaxin 6 (STX6) was also up-regulated (**Figure 66**). STX6 is a Q_{bc}-SNARE (Wendler and Tooze, 2001). It displays an important role in protein trafficking between the *trans*-Golgi network (TGN) and the endosomal system (Chao et al., 1999; Kuliawat et al., 2004; Mallard et al., 2002). Its up-regulation in the oval morphotype suggest that the *P. tricornutum* cells are secreting more materiel for degradation through the endosomal system as compared to the other morphotypes.

As a validation experiment, the gene expression data for the oval morphotype were compared to the ones of T+F combined together. This comparison came out with a similar DGE list as the one containing the 2 336 differentially expressed genes generated from the Venn diagram (data not shown).

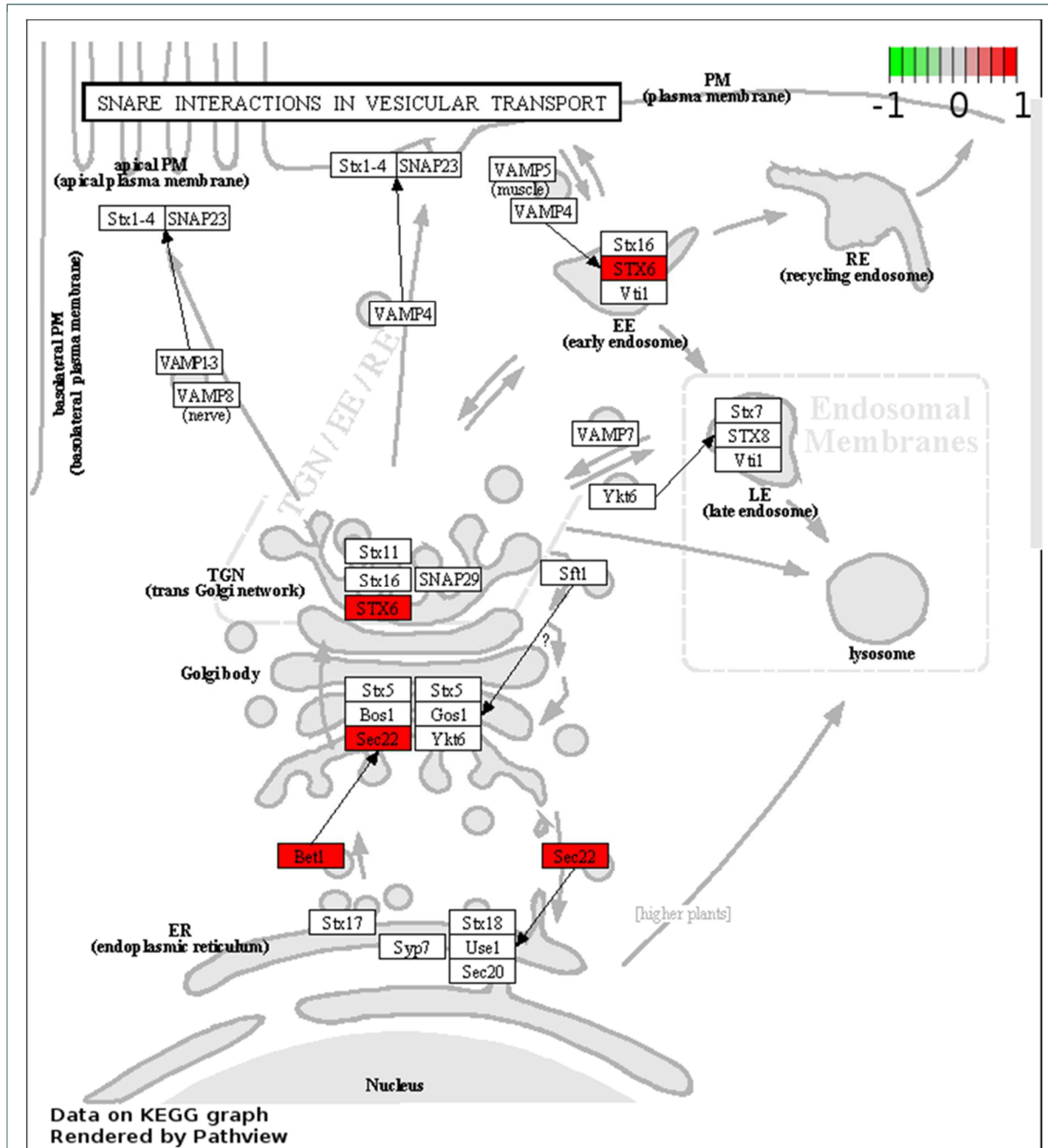


Figure 66: Mapping of the DE genes exclusive to the oval morphotype on the SNARE interactions in vesicular transport pathway.

4.4 DISCUSSION

This work is aiming at studying and comparing the transcriptomes of the three main morphotypes from *P. tricornutum*, namely the fusiform, ovale and triradiate one (Borowitzka and Volcani, 1978; He et al., 2014; Lewin et al., 1958; Tesson et al., 2009). In order to get the most accurate transcriptomic comparisons, Pt3 strain was adapted to generate enriched cultures in each specific morphotype. Such enrichment reached more than 80% for the triradiate morphotype culture and more than 90% for the fusiform and ovale cultures. Even if it was not possible to reach 100% for each morphotype, the above mentioned percentage seems to be high as compared to previous articles. Indeed, for example, Maheswari and coworkers in 2010 generated an ESTs database on the three morphotypes of *P. tricornutum* issued either from the Pt3 strain for a culture enriched up to 80% in oval; from the Pt8 strain for a culture enriched up to 70% in triradiate and finally Pt1.8.6 for a culture containing more than 95% of the fusiform cells. One year later, De Martino and collaborators reported evidence that environmental changes elicit morphological interconversion in *P. tricornutum*. This study was performed using cultures of Pt1.8.6 containing 80% of the fusiform morphotype and cultures of Pt3 containing 60% of oval cells (De Martino et al., 2011). The temperature used for these cultures was 19°C similarly to what was used for the Pt3 cultures in the current work. The mentioned previous studies were performed by using different strains as representative culture of one specific morphotype. In contrast, in the present work, we have chosen to generate morphotype enriched culture all issued from the Pt3 strain as it was clearly demonstrated that different accessions of *P. tricornutum* were significantly different genetically and morphologically (De Martino et al., 2007), thus introducing potential biases within the results. In addition, the enriched cultures for one specific morphotype generated for this study were stable over several months and years as previously described (Bartual et al., 2011; Borowitzka and Volcani, 1978; De Martino et al., 2007; He et al., 2014; Lewin et al., 1958). The integrity of the cellular organization was confirmed for the three morphotypes by TEM and was consistent with data previously described in the literature (Borowitzka and Volcani, 1978; Gügi et al., 2015; Tesson et al., 2009). Our experimental plan (Figure 2) was designed to combine recommendations for the RNA-Seq data analysis best practices as recently reviewed by Conesa and collaborators. To demonstrate the robustness of the RNA-Seq data obtained in this study, 22 DEG were randomly selected and their respective expression level was analyzed by qRT-PCR. The correlation between the qRT-PCR results and the RNA-Seq measurement came out with a R^2 value equal to 0.799, thus supporting the high quality and consistency of the dataset.

Pairwise comparisons of the morphotype's gene expression data were performed by considering that the fusiform morphotype (F) was the reference because it is the most common morphotype found in culture (De Martino et al., 2007; Zhao et al., 2014). Therefore, the comparison TF, OF and OT were done respectively. Only small differences representing less than 1% of the transcriptome were observed in the TF comparison as only 120 genes were significantly differentially expressed between these two morphotypes. This is in agreement with the report of Bartual and coworkers in 2008 which considered the fusiform and triradiate morphotypes from *P. tricornutum* as two basic morphotypes. More differences were observed in the OF and OT pairwise comparisons. Indeed, 2 692 and 3 553 genes were significantly differentially expressed in these conditions. Such results represent 22 and 29% of the transcriptomes respectively. About 83.3 % of the genes are down-regulated in the TF condition whereas 67% of the genes are up-regulated in OF and 62% are up-regulated in OT, suggesting that the oval morphotype might be more metabolically active. Amid these genes, 30 to 35% does not have any associated GO term. This could probably be due to the fact that the genome of *P. tricornutum* contains specific genes related to diatom metabolism. Indeed, previous reports indicate that at least 10 % of the diatom genomes correspond to diatom specific genes (Armbrust et al., 2004; Bowler et al., 2008; Montsant et al., 2005). Even if not completed, the GO annotation obtained in this work seems to be better than previous studies which are reported that only half of the diatom genes can be assigned to a putative function using homology-based methods (Maheswari et al., 2009). Similarly, Rogato et al., 2014 suggested recently that half of the diatom genes encode for protein of unknown functions. Hereafter, more functional characterization needs to be done in order to characterize the DGE which still do not have any associated GO term. Surprisingly, lots of these genes are presenting high values of log₂ fold change being in the extreme parts of the volcano plots (**Figure 54**).

The GO annotation and intersection analyses revealed global changes in major categories of genes involved in various pathways when comparing two by two the morphotypes from *P. tricornutum*. A particular focus has been done on the analysis of the 2 336 DEG which are present at the intersection of the OF and OT pairwise comparisons, thus being characteristic of the oval morphotype. Within the 2 336 DEG, 68% were found to be up-regulated, thus suggesting that the metabolism must be more active in the oval morphotype which is known to be resistant to stresses (De Martino et al., 2007). In our hand, the oval morphotype pathway analysis supports an activation of key primary metabolism pathways such as the glycolysis, the pentose phosphate pathway and the biosynthesis of nucleotides and triglycerides similarly to what has been described in the fusiform morphotype when studying the lipid accumulation mechanisms following nitrogen deprivation (Yang et al., 2013). Altogether, these results suggested that the triglyceride biosynthesis is activated

in the oval morphotype, that the glucuronomannan biosynthesis is more active and that the oval morphotype promotes nucleotide synthesis pathway. Such increase of the triacylglyceride biosynthesis pathway has been previously reported when Pt1 strain from *P. tricornutum* was cultured in media depleted in nitrate and exogenous phosphate (Valenzuela et al., 2012). As far as the glucuronomannan biosynthesis is concerned, one can hypothesize that the oval morphotype which is resistant to stresses need to protect itself with a bigger layer of cell wall. Indeed, glucuronomannan has previously been described to represent the major polysaccharide component of the cell wall from *P. tricornutum* (Gügi et al., 2015). *P. tricornutum* cells, like all other eukaryotic cells, contain a conventional endomembrane system composed of the nuclear envelope, the ER, the Golgi apparatus and the *trans* Golgi network which are connected to each other by vesicular shuttles (Brighthouse et al., 2010). When looking at the specific activity of the endomembrane system, we notice that the protein processing in the ER was overexpressed jointly with some actors from the Ubiquitin ligase complex as well as genes involved in the ER-associated degradation process which encode for proteins involved in the detection of misfolded proteins and their targeting for destruction. In the same guideline, the ubiquitin-mediated proteolysis mechanism, the proteasome and vesicular transport are also up-regulated. Such results demonstrate that, in the oval morphotype, more proteins are secreted and sent for degradation through the endosomal system as compared to the other morphotypes.

In the future, the RNA-Seq data generated in this work will be used to identify and discover novel transcripts, alternative splicing analysis and polymorphism in *P. tricornutum*. This will help unravel rare and novel transcripts and detect eventual layer of regulation never studied so far in *P. tricornutum*. Additionally, it should allow *de novo* transcript reconstruction which could help refining the quality of the genome annotation. It would also be interesting to analyze and compare the small RNAs transcriptome and the DNA methylation status of the three morphotypes, and evaluate whether the ovale morphotype is presenting some differences when compared to the other morphotypes. Indeed, Rogato and co-workers demonstrated in 2014 that small RNAs, including ncRNAs, could play important functional roles when *P. tricornutum* fusiform morphotype was grown at 18°C under different conditions of light and iron starvation. These results suggest that small ncRNAs may be fundamental for the biology of *P. tricornutum* especially when responses to stresses are concerned (Rogato et al., 2014). Therefore, it can be tempting to think that the oval morphotype may have a different profile of small ncRNA compared to the other morphotypes in *P. tricornutum*. This hypothesis is coherent with the fact that in our findings some actors from the spliceosome (Sm; SF3b; Prp38; DIB1; SPF27; PPIL1; data not shown) are up-regulated in the oval morphotype. Finally, proteomic study comparing the three morphotypes from *P. tricornutum* would

also be interesting to study and should allow to confirm and complement the present transcriptomic finding and oval specific features. Previous results from (Gutenbrunner et al., 1994) and co-workers suggested in 1994 from comparative SDS-PAGE experiments that the fusiform and oval protein patterns are slightly different. Based on the DEG lists that were obtained in the current RNA-Seq analysis, we can expect much more differences at the protein expression level when comparing the oval proteome to the one from the fusiform and triradiate respectively. The quality and power of such proteomic approach should be increased if the proteome used as a reference is based and enriched in reconstructed transcripts. In turn, proteogenomics which intend to identify novel peptides and proteins (not present in the reference protein sequence databases) from mass spectrometry-based proteomic by using a customized protein sequence databases generated from genomic and transcriptomic information (Nesvizhskii, 2014). In complement, proteomic data from *P. tricornutum* could be used to refine gene models based on protein-level evidence of gene expression.

4.5 MATERIELS AND METHODS

4.5.1 Culture and growth conditions

Phaeodactylum tricornutum strain Pt3 was used for this study (De Martino et al., 2007). Cultures enriched in fusiform, ovale or triradite morphotypes were adapted in the laboratory. Diatom cells were grown at 19°C in 1L bioreactors on a 16h/8h light/night cycle. The intensity of the light used was 68 $\mu\text{mol.m}^{-2}.\text{s}^{-1}$. The nutritive medium used was composed of natural seawater 100% for the fusiform and triradial morphotypes and 10% for the oval morphotype. It was sterilized by filtration through a 0.22 μm filter, autoclaved, and then enriched in Conway containing 40 mg.L^{-1} of sodium metasilicate (Na_2SiO_3) as described in (Baiet et al., 2011). The diatom cells were cultured under ambient air at a regulated flow rate of 0.83 L.min^{-1} . CO_2 from the air was the only available source of carbon for the diatoms. Bioreactors were inoculated with a final microalgae concentration of 1×10^6 cells.mL^{-1} . A cell count was performed daily on a Nageotte chamber. This cell count was done in triplicate for each culture during 15 days to follow the cell growth (**Figure 48**). According to this protocol, cultures containing a dominant morphotype were done in parallel. Four biological replicates for each morphotype's culture were performed independently (**Figure 49**).

4.5.2 RNA extraction

In order to isolate total RNA, cells were collected at day 8, in the middle of the exponential growth phase (**Figure 48**). For each culture, 100 mL (8.10^8 cells) were recovered and centrifugated during 3 min at 10 000 g at 19°C. Cell pellets were immediately resuspended in 1mL of TRIzol® Reagent (Ambion by Life Technologies) and the suspension transferred in a 2 mL tube of lysing matrix E (MP Biomedicals) and immediately flash frozen in liquid nitrogen. The samples were then stored at -80°C until RNA extraction. For RNA extraction, cells were lysed by 4 runs of 30 sec at 6.5 m.s^{-1} with the FastPrep®-24 high-speed benchtop homogenizer (MP Biomedicals). Samples were stored on ice between each run to avoid RNA degradation. Cell lysates were then incubated at room temperature during 5 min prior to a 5 min centrifugation at 14 000 g. Then, RNA was isolated using the TRIzol® reagent by using the standard protocol given by the supplier (Ambion®, Life Technologies) followed by a purification on Nucleospin RNA II column (Macherey-Nagel) including an on-column DNase I treatment and finally eluted with 60 μL of RNase-free water and stored at -80°C until library preparation. For each preparation, the RNA concentration was determined in duplicate using a Nanodrop spectrophotometer (Thermo Scientific). In addition, the RNA quality was controlled using an Agilent 2100 Bioanalyser (Agilent Technologies).

4.5.3 Library preparation and RNA-sequencing

Libraries were prepared from 4µg of total RNA using the Illumina Truseq® Stranded mRNA Sample Preparation kit according to the manufacturer's protocol (Illumina). Briefly, mRNAs were purified and fragmented before first strand synthesis, second strand synthesis was performed and adaptors were ligated to the products before PCR amplification of the samples as previously recommended to minimize bias (van Dijk et al., 2014). Indexed Libraries were normalized and pooled two by two for random multiplexing. Libraries were subjected to a 75 bp paired-end sequencing on an Illumina GAIIX, according to the manufacturer's protocol. Two samples were pooled on a flowcell lane. Raw image files were processed using the Illumina Real Time Analysis (RTA 1.9) software (Illumina). Demultiplexing was performed using CASAVA 1.8.2 (Illumina). On average, 38 millions of reads were generated per sample corresponding to 2.9 Gb per sample. 94% of bases were read with a Qscore above 30.

4.5.4 Bioinformatic analysis

4.5.4.1 Quality assessment

Quality assessment of the short reads was realized with FastQC v0.10.1 (S. Andrews, FastQC a quality-control tool for high-throughput sequence data, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/2010>) (**Figure 67**). Over represented sequences were found to correspond to adapters. In a first attempt, various trimming assays were tested using either the FASTX-Toolkit (FASTQ/A short-reads pre-processing tools, http://hannonlab.cshl.edu/fastx_toolkit/) or a home procedure with perl scripting. Such treatment of the data had no impact on the results of the mapping. Therefore, the trimming step was not necessary and we decided to keep all the reads without trimming for further steps.

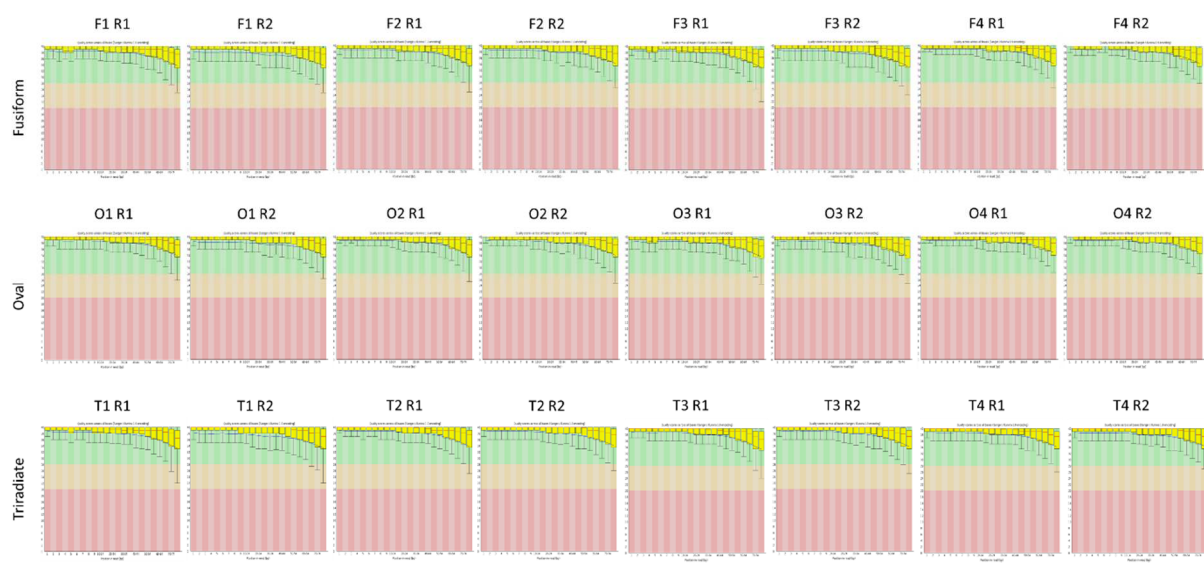


Figure 67: FastQC report depicting per base sequence quality for the 4 biological replicates of the respective enriched cultures for either the fusiform, triradiate or oval morphotype.

4.5.4.2 Read alignment

Raw reads were aligned to the *P. tricornutum* reference genome sequence (.FA), and the reference annotations (.GTF), both obtained from http://protists.ensembl.org/Phaeodactylum_tricornutum (release 25, dec 2014). The total genome sequence (JGI-Phatr2, Assembly ASM15095v2 Feb 2010; INSDC Assembly GCA_000150955.2) contains 27 568 093 bp including 33 finished nuclear chromosomes (26 138 756 bp), the chloroplast chromosome (117 369 bp) and 55 unmapped sequences (Phatr2_bd, 1,311,968 bp). All were included in this study for comprehensive analysis. Although transcripts from the chloroplast lacking polyA tail are supposed to be automatically removed due to the purification method of the mRNA by polyA selection, we nonetheless decided to include the chloroplastic genome. EnsemblProtists annotations (.GTF file) used for alignment include 12 178 protein coding genes (12 392 gene transcripts), 177 non-coding genes and 37 pseudogenes. RNA short reads were aligned to the reference genome using the splice junction mapper TopHat 2 v2.0.6 (ccb.jhu.edu/software/tophat/) (Trapnell et al., 2009) with "the option "-library-type fr-firststrand", and Bowtie v2.0.2 (Langmead and Salzberg, 2012; Langmead et al., 2009) with default parameters with the exception of Max-multihits = 1.

4.5.4.3 Read count and normalization

For gene expression quantification, the resulting alignment files for each condition, and the reference annotation (.GTF file) guiding the transcript count values, were provided to HTSeq-Count v0.5.4p1 (Anders et al., 2015) used with the intersection-nonempty mode. For each gene defined in the Phatr2 database, counts were normalized with DESeq2 v1.14.0 (Love et al., 2014; <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>). Normalized counts allowed us to estimate the fold-change in transcript levels as well as the log₂ Fold Change, for the three pairwise conditions (oval versus fusiform, oval versus triradiate and triradiate versus fusiform).

4.5.4.4 Differential analysis

In order to carry out differential analysis, DESeq2 v1.14.0 (Love et al., 2014; <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>) was used to perform pairwise comparisons between samples to find significantly differentially expressed transcripts.

RNA-Seq quantification based on normalized read counts was performed. To do so, we used discrete probability distributions by applying the Negative Binomial (Robinson and Smyth, 2007). The adjusted p-value was obtained by the Holm-Bonferroni method which is known to counteract the problem of multiple comparisons and to control the familywise error rate.

4.5.4.5 Functional analysis and enrichment

Prior to perform such analysis, EPr sequences corresponding to ncRNA (tRNA, rRNA, snoRNA) have been removed manually from the dataset. For each significantly differentially expressed transcripts (P adjusted value <0.01), GO annotation and enrichment was performed using Blast2GO v3.3.5 (Conesa et al., 2005; <https://www.blast2go.com>) with default parameters. Intersection analysis was performed with Venny v2.1 (<http://bioinfogp.cnb.csic.es/tools/venny/>). Pie charts were done with Multi-Level option in Blast2GO considering the biological process and molecular function. In order to perform mapping on metabolic pathways, significantly DGE were mapped with the help of Pathview (Luo and Brouwer, 2013; <http://pathview.uncc.edu>), and the mapping tool of Blast2GO powered with KEGG database.

4.5.5 qRT-PCR Validation

Twenty-two gene of interest (GOI) were randomly selected from the RNA-seq data analysis, picked from the 3 DEG lists resulting from the pairwise comparisons of *P. tricornutum*'s morphotypes. Primer pairs specific to each GOI were designed by using the the Primer-BLAST tool at NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) as described in chapter II. Validation steps of each primer pair in order to control its specificity and evaluate its amplification efficiency were performed as previously described (chapter II). qRT-PCR experiments were achieved as described in our previous work (chapter II).

4.5.6 TEM microscopy for *P. tricornutum* morphotype characterization

4.5.6.1 High pressure freezing (HPF) and freeze substitution (FS)

HPF was performed with the freezer HPF-EM PACT I Leica-microsystems (Studer et al., 2001). Prior to freezing, *P. tricornutum* suspension-cultured cells were treated 2 hours with 100 mM mannitol as cryoprotectant diluted in culture medium at room temperature. Pre-treated *P. tricornutum* cells were then transferred into the cavity of a copper ring used for cryofracture, which is 100 μm in depth and 1.2 mm in diameter. Excess medium was lightly absorbed by filter paper. Using a horizontal loading station, the specimen carriers were tightened securely to the pod of specimen holder. After fixation on the loading device, specimens were frozen according to a maximum cooling rate of $10\,000^\circ\text{C}\cdot\text{sec}^{-1}$, incoming pressure of 7.5 bars and working pressure of 4.8 bars. Rings containing frozen samples were stored in liquid nitrogen until the freeze substitution procedure was initiated.

4.5.6.2 Freeze substitution

After high-pressure freezing, samples were transferred to a freeze substitution automate (AFS, Leica, <http://www.leica-microsystems.com>) pre-cooled to -140°C . Freeze substitution conditions followed a modified procedure from Studer et al. (2001). Samples were substituted in anhydrous acetone with 0.5% uranyl acetate at -90°C for 96 h. The temperature was gradually raised to -60°C using a gradient of $+2^{\circ}\text{C}\cdot\text{h}^{-1}$ and stabilized during 12 h, then gradually raised to -30°C by using the same gradient during 12 h and gradually raised again to -15°C for 8 h. Finally, samples were rinsed twice with anhydrous ethanol.

Infiltration was then processed at -15°C in a solution containing ethanol: London Resin White (LRW) in a ratio 2:1 overday; 1:1 overnight and 1:2 overday and then 100% of a pure resin solution overnight. This solution was changed twice and incubated for 2x24 h. The LRW was finally polymerised into the AFS apparatus at -15°C under ultra violet light during 48 h.

Ultrathin sections (80 nm; ultracut UCT, Leica) of *P. tricornutum* cells were collected onto carbon-formvar-coated nickel grids. A classical staining using uranyl acetate and lead citrate was done before sections were observed in a Philips, FEI Tecnai 12 Biotwin transmission electron microscope operating at 80 kV, with ES500W Erlangshen CCD camera (Gatan).

Availability of data and materiel

The analyzed data are presented within the different tables and figures of this paper.

Abbreviations: BP: biological process; DEG: differentially expressed gene; ER: endoplasmic reticulum; F: *P. tricornutum* fusiform morphotype; GO: gene ontology; GOI: gene of interest (GOI); O: *P. tricornutum* oval morphotype; Pt: *Phaeodactylum tricornutum*; RNA-Seq; Ribonucleic acid sequencing; T: *P. tricornutum* triradiate morphotype; KEGG: Kyoto Encyclopedia of Genes and Genomes; EST: Expressed Sequence Tag.

Competing interests:

The authors declare that they have no competing interests.

Funding:

The authors are grateful to the GRR IRIB of the Haute-Normandie region, the University of Rouen Normandie as well as the IUF (Institut Universitaire de France) for their financial support. The

funding bodies were not involved in the design of the study, collection and interpretation of data nor in the writing of the manuscript.

Authors' contributions:

MCKM, PL and MB conceived the idea and shape the structure of the work. CO, MCKM, CP, CB, SB, IT performed the experiments; CO, MCKM, CB, NV, TL, HD, PL, MB analyzed the results; CO, MCKM, CB, HD and MB compiled the results together and wrote the paper. All the authors read the paper and agreed on its final version prior to the submission.

Acknowledgements:

The authors are thankful to Pr Chris Bowler, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR8197, Inserm U1024 for sharing its *P. tricornutum* Pt3 strain.

Chapitre IV:

Identification of reference genes for quantitative real-time PCR analysis in *Phaeodactylum tricornutum* morphotypes

5 Identification of reference genes for quantitative real-time PCR analysis in *Phaeodactylum tricornutum* morphotypes

Clément Ovide^{1*}, Carole Plasson^{1*}, Muriel Bardor^{1,2} and Marie-Christine Kiefer-Meyer^{1#}

*** These authors equally contributed to the work.**

¹ Normandie Univ, UNIROUEN, laboratoire Glyco-MEV 76000 Rouen, France.

² Institut Universitaire de France (IUF), Paris, France.

#correspondence to :

Dr Marie-Christine Kiefer-Meyer,

Laboratoire Glyco-MEV EA 4358, Université de Rouen, UFR des Sciences et Techniques, Bâtiment Henri Gadeau de Kerville, 1 rue Tesnière, 76821 Mont-Saint-Aignan Cedex, France.

Tel: +33 2 35 14 66 90; e-mail: marie-christine.kiefer-meyer@univ-rouen.fr

Short title: Identification of new reference genes in *Phaeodactylum tricornutum*

Key words: reference genes, qRT-PCR, morphotype, diatom, *Phaeodactylum tricornutum*

5.1 Abstract

This work aimed at identifying a set of genes which can be used as reference genes in qRT-PCR analysis in the diatom *P. tricornutum*. Stability of the gene expression of different potential reference genes selected from commonly used housekeeping genes and already evaluated for *P. tricornutum* (Siaut et al., 2007) or genes showing a stable expression in a RNA-seq transcriptome analysis of *P. tricornutum* have been studied. The results allow us to find most appropriate reference genes for normalization of qRT-PCR analysis of the three main morphotypes from *P. tricornutum*, ovale (O), fusiform (F) and triradiate (T). In a second step, the reference genes were used to normalize the expression of genes involved in the *N*-glycosylation pathway. Differential expression of these genes in the three morphotypes from *P. tricornutum* was determined.

5.2 Introduction/Background

5.2.1 qRT-PCR

Reverse transcription (RT) of RNAs followed by a real-time polymerase chain reaction (PCR) has become a routine tool in molecular biology for the quantification of RNA (Bustin, 2000; Gachon et al., 2004; Kubista et al., 2006; Nolan et al., 2006). qRT-PCR is a simple and very sensitive method to quantify gene expression. It is easy to carry out at high throughput and is now widely used to detect and/or measure differences in RNA levels between samples. However, to provide reliable results, the raw qRT-PCR data need to be normalized carefully with selected reference genes (Bennett et al., 2015; Guénin et al., 2009; Vandesompele et al., 2002).

5.2.2 Validation of reference genes for qRT-PCR data normalization

Traditionally, reference genes are genes encoding for proteins involved in basic cellular process (cell structure maintenance or primary metabolism), supposed to be expressed at constant levels in all conditions and known as housekeeping genes. In the literature, some of the most frequently mentioned genes for normalization are for example, actin, glyceraldehyde-3-phosphate dehydrogenase, tubulin or ubiquitin; genes that were in the past, usually used as control in Northern-blot experiments (Huggett et al., 2005). However, unlike Northern blotting which gives rather a qualitative information on gene expression, qRT-PCR analysis is intended to monitor gene expression profiles in a very accurate quantitative manner. Therefore, it is essential to carefully choose the appropriate reference genes in the qRT-PCR experimental design (Chapman and Waldenström, 2015; Gutierrez et al., 2008; Hruz et al., 2011; Kozera and Rapacz, 2013; Thellin et al., 2009). In addition, reference genes have to be validated by comparison to target genes in a well-defined biological context (Bennett et al., 2015; Bustin, 2010). In the last years, various studies carried out in different species were aiming at identifying new reference genes suitable for the normalization of gene expression analysis by qRT-PCR (Czechowski et al., 2005; Hruz et al., 2011; Li et al., 2016; Paolacci et al., 2009; Xiao et al., 2015). Indeed, it is now well-established that no gene is universally stable. In these studies, a set of candidate reference genes are tested to evaluate their stability in defined experimental conditions in comparison to the commonly used “housekeeping” genes. Different statistical programs: NormFinder (Andersen et al., 2004), BestKeeper (Pfaffl et al., 2004) geNorm (Vandesompele et al., 2002) for example, are available and used routinely to determine the stability of the potential new reference genes.

5.2.3 Identification of reference genes from transcriptomic datasets (microarray data and RNA-sequencing)

With the development of high-throughput transcriptome analysis, first by microarray profiling and very recently, by RNA-sequencing, systematic analysis of the resulting gene expression data offers the possibility to identify novel candidate genes suitable for normalization of qRT-PCR analysis. In the plant model *Arabidopsis thaliana*, Czechowski et al. (2005) have identified hundreds of very stably expressed genes based on a large set of microarrays data. Among them, 20 were tested in addition to 5 traditional reference genes in 20 different physiological and environmental conditions by qRT-PCR analysis. Most of the novel genes showed a higher stability in gene expression levels in these various conditions. Accordingly to these results, they were recommended as reference genes to normalize transcripts levels in *Arabidopsis* gene expression studies and have later on been used in qRT-PCR analysis. This is the case for example of the *PP2A gene* (At1g13320, Huang et al., 2016) and the *Tip41-like* one (AT4G34270, Xu et al., 2014). These new reference genes have also been used in *A. thaliana* as a basis to search candidate reference genes in other organisms by sequence similarity, as for example in grapevine (Tashiro et al., 2016), leading to the identification of new stably expressed genes in this specie and suggesting the possibility that these genes could also be selected in others plant species.

Recently, the RNA-Seq technology appears as an efficient tool to identify new candidate reference genes due to its ability to detect theoretically all transcripts and to access to their expression profiles. By analyzing RNA-Seq data obtained from different normal human tissue types, Eisenberg and Levanon (2013) were able to provide a list of 3 804 genes that show high expression variability across the 16 different tissues. Most of the reference genes previously identified by a microarray analysis (de Jonge et al., 2007) were present in this list. However, the 2 genes often cited as housekeeping genes in the literature, GAPDH (glyceraldehyde-3-Phosphate dehydrogenase) and ACTB (Actin B), were missing. Lin and coworkers (2014) have combined RNA-Seq and microarrays analyses for reference genes mining in maize. From a pool of 1 661 candidate genes, they finally validated 9 genes, stably expressed across various tissues and developmental stages, under normal or stress conditions, as reference genes for gene expression analysis in maize. In a recent study (Kudo et al. 2016), the datasets available from RNA-Seq experiments were analyzed with the aim to find new potential reference genes in *A. thaliana* and model crop plants such as *Glycine max* (soybean), *Solanum lycopersicum* (tomato) and *Oryza sativa* (rice). A set of novel reference genes were identified for these

species with a higher stability than commonly used reference genes. Altogether, these studies demonstrated that RNA-Seq data analysis could definitively help to identify efficient reference genes.

5.2.4 Housekeeping genes of *P. tricornutum*

In 2007, Siaux *et al.* selected 12 putative housekeeping genes of *P. tricornutum* which were chosen according to the ones used conventionally in plants, mammals and yeast. For each gene, a couple of primers was designed and tested. Among the 12 genes tested, an analysis of the qRT-PCR data with the geNorm program (Vandesompele *et al.*, 2002) demonstrated that the most stably expressed housekeeping genes were the *RPS* (ribosomal protein small subunit 30S) and *TBP* (TATA box binding protein) genes in the experimental conditions defined for the study, i.e. *P. tricornutum* Pt1 8.6 and Pt8 cells grown at 18°C with a 12 h photoperiod (175 $\mu\text{mol.m}^{-2}.\text{s}^{-1}$). The *RPS* and also the histone *H4* reference genes were then used to normalize the *FcpB* (Fucoxanthin, Chlorophyll *a/c*- binding Protein B) gene expression values. In subsequent gene expression studies of *P. tricornutum* involving qRT-PCR analysis, the best reference genes identified by Siaux and coworker in 2007, in particular *RPS* and *H4*, were generally used (Coesel *et al.*, 2008; De Riso *et al.*, 2009; Huysman *et al.*, 2010; Mus *et al.*, 2013; Zhao *et al.*, 2014). Even when studies are performed in various physiological conditions, scientists are using the same reference genes without specifying how these genes behave in the experimental conditions used. Generally, only one reference gene is mentioned. An alternative reference gene, encoding for a potential hiv-1 rev binding protein (now Phatr3_J42776), was chosen according to microarray data in an analysis which studied the mechanisms of high light acclimation in *P. tricornutum* (Nymark *et al.*, 2009).

5.3 Main objectives

In this study, our main objective was to benefit from the RNA-Seq dataset that was generated from a comparative study of the transcriptomes of the three morphotypes (O, F and T) from *P. tricornutum* to identify novel reference genes suitable for gene expression analysis in qRT-PCR by selecting the genes that appear stably expressed in the RNA-Seq data analysis. Primer pairs specific to these new potential reference genes were then validated, and used in a qRT-PCR experiment aimed at analyzing the expression levels of genes encoding the enzymes involved in the *N*-glycosylation pathways.

5.4 Material and Methods

5.4.1 Diatom cultures

P. tricornutum strain Pt3 (De Martino et al., 2007) was used in this study to purify total RNAs. Cultures specifically enriched for the three morphotypes were performed in parallel. Cells were grown at 19°C in 1L bioreactors on a 16h-8h light cycle at 68 $\mu\text{mol.m}^{-2}.\text{s}^{-1}$. The nutritive medium used consisted of natural seawater 100% for the fusiform and triradiate morphotypes and 10% for the oval morphotype. The media composition was as previously described in (Baïet et al., 2011) containing some Conway and 40 mg.L^{-1} sodium metasilicate (Na_2SiO_3). The diatom cells were cultured under ambient air at a regulated flow rate of 0.83 L.min^{-1} . CO_2 from the air was the only available source of carbon for the diatoms. Bioreactors were inoculated with a final microalgae concentration of 1×10^6 cells.mL^{-1} . Finally, cells were collected at day 8 in the middle of the exponential growth phase.

Genomic DNA extracted from *P. tricornutum* Pt 1.8.6 strain was used as a control for primers specificity verification. These *P. tricornutum* cells were grown in the same conditions as the ones described for the Pt3 strain.

5.4.1.1 DNA isolation

The NucleoSpin® Plant II kit (reference number 740770) from Macherey-Nagel was used in order to prepare genomic DNA from *P. tricornutum* cells. First, 100 mL of cell culture was centrifuged during 5 minutes at 4 000g and the cell pellet was subsequently resuspended in 800 μL of lysis buffer PL1 and transferred in a 2 mL « lysing matrix tube E » (reference number 116914050, MP Biomedicals) in order to lyse the sample in a FastPrep-24 Homogenizer (MP Biomedicals). For this purpose, the instrument was used at a speed setting of 6.5m/s for 8 rounds of 1 minute. The sample was then centrifuged during 15 minutes at 12 000g and the supernatant was taken to purify the DNA by using a mini spin column according to the protocol given by the supplier, including a 40 minutes incubation at 65°C with RNase A. DNA was finally eluted with 100 μL water, at 65 °C, in 2 steps.

5.4.1.2 RNA purification

All consumables and solutions used to work with RNAs were certified RNase-free and the experiments were carried out in a RNase-free environment. For each morphotype, four

biological replicates (4 independent cultures for each morphotype) were done and subsequently independent RNA purifications were carried out for each replicate.

In order to purify total RNA, *P. tricornutum* cells in exponential growth phase when harvested and 100 mL of each cell morphotype culture were taken and centrifugated during 3 min at 10 000 g at room temperature to harvest the cells. The cell pellets were then resuspended in 1mL of TRIzol (Life Technologies) per 100mL, transferred in a 2 mL « lysing matrix tube E » (reference number 116914050, MP Biomedicals) and immediately flash frozen in liquid nitrogen. The samples were stored at -80°C until RNA extraction. Lysis of the cells was carried out by 4 runs of 30 sec at 6.5 m.s⁻¹ in a FastPrep-24 Homogenizer (MP Biomedicals) instrument. Samples were stored on ice between each run to avoid any degradation. Lysed samples were then incubated at room temperature during 5 min and subsequently centrifugated during 5 min at 14 000g. The supernatant was recovered and transferred in a new microtube, and after addition of 200µL of chloroform, the sample was gently mixed and then incubated at room temperature during 3 min. After a 15 min centrifugation at 12 000 g at 4°C, the upper aqueous phase was transferred in a new microtube and 1 volume of 70% ethanol was added. Total RNAs were then purified on a Nucleospin column using the NucleoSpin RNA II kit (Macherey-Nagel) according to the instructions provided by the supplier. RNA were finally eluted with 60µL of RNase-free water and stored at -80°C until further use. For each sample, two aliquots were removed to evaluate the RNA concentration with a Nanodrop 2000c spectrophotometer (Thermo Scientific) and its quality with a Bioanalyzer 2100 instrument (Agilent).

5.4.2 RNA sequencing

Libraries were prepared from 4µg of total RNA using the Illumina Truseq® Stranded mRNA Sample Preparation kit according to the manufacturer's protocol. The detailed method is described elsewhere (chapter III) but briefly, mRNAs were purified and fragmented before first strand synthesis, second strand synthesis was performed and adaptors were ligated to the products before PCR amplification of the samples. Indexed libraries were normalized and pooled two by two. Libraries were subjected to a 75 bp paired-end sequencing on an Illumina GA IIX, according to the manufacturer's protocol. Two samples were pooled on a flow-cell lane. Raw image files were processed using the Illumina Real Time Analysis (RTA 1.9) software (Illumina). Demultiplexing was performed using CASAVA 1.8.2 (Illumina). The genes which are differentially expressed between 2 morphotypes were identified and their difference in expression was given in log₂Fold Change using DE-Seq2 tool (Love et al., 2014). This analysis

was done to compare O *versus* T, O *versus* F and F *versus* T leading to 3 lists of differentially expressed genes.

5.4.3 qRT-PCR analysis

5.4.3.1 cDNA synthesis

The High capacity cDNA reverse transcription kit from Applied Biosystems was used to carry out the cDNA synthesis with total RNA samples from the four biological replicates of the three different morphotypes following the manufacturer's instructions. One μg of total RNA purified as described above was used per reaction in a final volume of $20\mu\text{L}$, using random primers as cDNA priming strategy. The cDNAs were then stored at -20°C until use.

5.4.3.2 qRT-PCR primer design

The Primer-BLAST tool at NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) was used to design primer pairs specific of each selected gene. The default parameters were used for the primer searches with the exception of the "PCR product size" which was set to 70-150 bp and in the "Primer Pair Specificity Checking Parameters", *Phaeodactylum tricornutum* (taxid: 2850) was given for the "Organism". The optimal primer melting temperature was set at 60°C .

5.4.3.3 PCR amplification, DNA electrophoresis and Sanger sequencing analysis

In order to control the amplification specificity of each selected primers pair, a PCR amplification test was done using the GoTaq DNA polymerase (Promega) in standard conditions with a Veriti 96-w Thermal Cycler, 0.2 ml wells (Applied Biosystems). One microliter of DNA or cDNA templates prepared as described above were used in a $50\mu\text{L}$ volume of reaction with the following amplification parameters: an initial denaturation step of 3 min at 95°C then 30 cycles of 30 s at 95°C , 30 s at 60°C and 30 s at 72°C with a final extension step of 5 min at 72°C . Ten microliters out of the $50\mu\text{L}$ of PCR reaction were subsequently analyzed by a 2% agarose gel electrophoresis prior to check the specificity of amplified products by sequencing. The Sanger sequencing was carried out by GATC Biotech (<https://www.gatc-biotech.com>).

5.4.3.4 Real-time quantitative PCR experiments

The quantitative real-time PCR was carried out with the QuantStudioTM 12K Flex Real-Time PCR System (Applied Biosystems) and the pipetting steps were done with the Bravo Automated Liquid Handling Platform (Agilent Technologies). The reaction mix were prepared

in a 96-well reaction plate using the Fast SYBR Green Master Mix (Applied Biosystems) in a total volume of 13 μ L. The final primer concentration in the reaction mix is 100 nM each and the cDNA volume added is 3 μ L. The following amplification protocol was used: 20 s at 95°C then 40 cycles of 3 s at 95°C, 20 s at 60°C followed by the melt-curve analysis: 15 s at 95°C, 1 min at 60°C and 15 s at 95°C. In order to validate the primer pairs, a dilution set of cDNA was prepared as follow: first, a 1/5 dilution of the cDNA stock is prepared and 4 consecutive 1/3 dilutions were made from it and the 5 cDNA dilutions were used in the qRT-PCR experiment to test each primers pair and to determine the optimal cDNA dilution to use.

The qRT-PCR analysis was done with the optimal cDNA dilution determined in the validation assay, namely a 1/5 dilution was used in our experiments. The efficiency of each primers pair was calculated from this validation. The specificity of the primer annealing was also confirmed and a melt-curve analysis performed at the end of the PCR cycles.

5.4.3.5 qRT-PCR data and gene expression stability analysis

5.4.3.5.1 Selection of the best reference genes

In order to select the best reference genes, a verification of the gene expression stability of the candidate reference genes in the defined experimental conditions was done by different methods listed below.

5.4.3.5.1.1 Normfinder analysis

A “Normfinder” analysis (Andersen et al., 2004) was carried out to select the best reference genes among a set of candidates. The “Normfinder” algorithm calculates a stability value for each candidate normalization gene and ranks them according to this value.

5.4.3.5.1.2 RefFinder analysis

The “RefFinder” tool (Xie et al., 2012) was recently made available (<http://fulxie.0fees.us/?type=reference&ckattempt=1>). It allows an easily, web-based analysis of qPCR data with the different statistical programs currently used (geNorm, Normfinder, BestKeeper, and the comparative $\Delta\Delta$ Ct method) in order to analyze the gene expression stability of candidate reference. The candidate gene are ranked according to their stability determined by each approach and the geometric mean of the ranking values is then calculated for each gene in order to finally sort out the genes, those with the lowest geometric mean are considered as the most stable.

5.4.3.5.2 qRT-PCR data analysis

The qRT-PCR data were analyzed accordingly by using a combination of the methods described by Livak and Schmittgen (2001) and Vandesompele et al. (2002). In the first one, a derivation of the $2^{-\Delta\Delta Ct}$ is used to estimate the expression level of a target gene taking into account the amplification efficiency of the primer pair used specifically for this gene. According to the second method, this gene expression value is then normalized by using the geometric mean of reference genes (2 in this study) as normalization factor. The different calculation steps are detailed below:

1. First, the exponential amplification of PCR is described by the following equation:

$$N_{ct} = N_0 \times (1 + Ec)^{\overline{Ct}}$$

Where N_0 is the the initial number of target molecules (reflecting the expression level of the target gene), N_{ct} is the number of target molecules at the threshold cycle (Ct), Ec is the amplification efficiency of the target gene and \overline{Ct} is the mean of the C_t values for the technical replicates of this gene.

N_0 can be expressed for each gene (target and reference) and each biological replicate (k , $k=4$ in this analysis) of each morphotype (m): $N_{0m,k}$ as follow:

$$N_{0Tm,k} = N_{ctT} \times (1 + E_T)^{-\overline{Ct_T}}$$

$$N_{0Rm,k} = N_{ctR} \times (1 + E_R)^{-\overline{Ct_R}}$$

where E_T is the amplification efficiency of the target gene, E_R the amplification efficiency of the reference gene, $\overline{Ct_T}$ the mean of the C_t values for the technical replicates of the target gene and $\overline{Ct_R}$ the mean of the C_t values for the technical replicates of the reference gene.

2. In a second step, a normalization factor (NF) obtained by geometric averaging of the N_0 of the 2 reference genes is determined for each biological replicate:

$$NF = (N_{0R1m,k} \times N_{0R2m,k})^{1/2}$$

3. The normalized relative quantity (NRQ) for each target gene and each biological replicate is given by the formula:

$$NRQ_{m,k} = \frac{N_{0Tm,k}}{NF_{m,k}} = \frac{(1 + E_T)^{-\overline{Ct_T}}}{[(1 + E_{R1})^{-\overline{Ct_{R1}}} \times (1 + E_{R2})^{-\overline{Ct_{R2}}}]^{1/2}}$$

4. The average of the normalized $NRQ_{m,k}$ values of the 4 biological replicates is then calculated for each morphotype:

$$\overline{NRQ_{m,k}} = \frac{\sum_{k=1}^4 NRQ_{m,k}}{4}$$

5. The comparison of the expression levels of a target gene between 2 morphotypes is done by calculating the ratio of the mean of their normalized gene expression values:

$$\frac{\overline{NRQ_{mx}}}{\overline{NRQ_{my}}}$$

where mx and my represent the 2 considered morphotypes.

5.5 Results

5.5.1 Selection of candidate reference genes

5.5.1.1 Genes selected from the list of reference genes tested by Siaux *et al.*

Firstly, we decided to take the genes identified previously by Siaux *et al.* (2007) as potential reference genes to be used in qRT-PCR experiments. In this work, a list of twelve genes is provided. In this study, 11 out of these 12 genes were selected and a search for primer pairs was carried out with the Primer-BLAST software as described in the Materiel and Methods section. The Ubiquitin gene corresponding to the Protein ID 29166 was not found any longer in the databases (JGI Genome portal, <http://genome.jgi.doe.gov> and EnsemblProtists, protists.ensembl.org). Most of the genes contain at least 1 intron and, if possible, the 2 primers of a pair are separated by an intron or 1 of the primers span an exon-exon junction. For almost all of the primer pairs, one of these conditions was observed (see **Table 28** for detailed information regarding each gene). In addition, for 1 of these genes (TBP, tata-box-binding protein, Phatr3_J10199), identical primer pair as the one used by Siaux *et al.* 2007 was also tested.

At the end, 12 primer pairs corresponding to 11 different genes were examined to be suitable for RT-qPCR analysis but only 3 of them (Phatr3_J21122, Phatr3_J10847 and Phatr3_J29812) were validated in our experimental design and further used to verify the gene expression stability of these traditional “housekeeping” genes.

Sequence name Ensembl Protists release 25 - January 2015 © EBI	Sequence description	Primer name (Forward/Reverse)	Sequences (5'-3') Forward Reverse	Amplicon size (bp)
Phatr3_J10199	tata-box-binding protein	TBPF1/TBPR1	GTC-GCT-TTG-GTG-CCG-TAA-TC TAG-GCA-AAC-TTT-TTC-GCC-GC	133
		TBPfw/TBPrev (Siaut et al., 2007)	ACC-GGA-GTC-AAG-AGC-ACA-CAC CGG-AAT-GCG-CGT-ATA-CCA-GT	175
Phatr3_J10847	30s ribosomal protein s1	RPSF1/RPSR1	ACG-CCC-ATG-GAG-TTC-CAA-AT ACA-GCG-ATA-ACC-TTG-GCC-TC	149
Phatr3_J18475	elongation factor 1-alpha	EF1AF1/EF1AR1	TCG-CGT-TGT-TCA-CTC-TCT-CG ACG-CTT-ATC-GAT-ACC-TCC-ACA-C	193
Phatr3_J20262	cmgc cdk protein kinase	CDKA1F1/ CDKA1R1	TGG-AGC-TCC-CGA-TTC-TCA-AG TGA-TGA-GCA-AGT-TAG-GGG-GC	111
Phatr3_J21122	Tubulin beta-2A chain (Tub B)	TubBF1/TubBR1	CCG-TTC-GCT-CGT-TTG-ATT-CC GGC-AGA-CAT-GAC-CTC-CCA-AA	125
Phatr3_J29136	actin2	Actin2F2/Actin2R2	GCG-CAG-TTC-GTT-CTT-ATA-ATC-CA CAA-ATC-CGG-CCT-TGC-ACA-TT	157
Phatr3_J29812	actin1	Actin1F1/Actin1R1	GCC-TTA-CAC-CGC-CGT-ATT-CT TCA-ATG-ACA-AGG-GCA-GCG-AT	79
Phatr3_J34971	histone h4	H4-1bF1/H4-1bR1	TCG-TGG-TGG-TGT-GAA-ACG-TA CAA-GCG-CAA-GCT-TAA-CCT-CC	197
Phatr3_J51128	GAPDH, isoform 1	GapC2aF1/ GapC2aR1	CCG-AGG-ATG-TCT-CGT-TTG-AAG-A GAC-AAC-CTT-TTC-TGC-ACC-GC	191
Phatr3_J54534	Tubulin alpha-1C chain (Tub A)	TubAF1/TubAR1	GAG-GAT-GCC-GCC-AAC-AAC-TA GTC-GCG-TGA-AAG-ACG-AGG-AA	134
gi 45356670	18S ribosomal RNA	18SrRNAF1/18SrRNAR1	GGC-TTT-AAC-GGG-TAA-CGG-GA GGT-AAT-TTA-CGC-GCC-TGG-TG	98

Table 28: List of the candidate housekeeping genes, with the primer sequences used in qRT-PCR analysis and the amplicon size.

The sequence name is given according to the EnsemblProtist database. Genes successfully validated in our experimental design are colored in green, and non validated genes are colored in red.

5.5.1.2 Identification of reference genes based on RNA-Seq data sets from the three *P. tricornutum* morphotypes

In parallel, another strategy was carried out to identify new potential reference genes on the basis of the results obtained from a transcriptome analysis by RNA-Seq. After a differential analysis of the data by DESeq (Anders *et al.*, 2010) and the establishment of 3 lists of differentially expressed genes by pairwise comparison of the morphotypes, a search for genes which are not statistically differentially expressed was carried out by sorting the genes according to their log₂FoldChange. The genes which were not differentially expressed between 2 morphotypes showing a log₂FoldChange equal to 0 and an adjusted P_{value} near to 1 were selected.

The 3 pairwise comparison lists of expressed genes were compared and a set of 10 genes stably expressed in the 3 morphotypes was selected from these data. A primer pair was designed for each of these genes and tested in validation experiments (**Table 29**). Four out of the 10 pairs were validated and subsequently used in qRT-PCR analysis to evaluate the gene expression stability of the corresponding genes and the possibility to use them as reference genes.

Sequence name Ensembl Protists release 25 - January 2015 © EBI	Sequence description Blast2GO 3.0.8	Primer name Forward/Reverse	Sequences (5'-3') Forward Reverse	Amplicon size (pb)
Phatr3_EG02331	Predicted protein	Phatr47207_F1/ Phatr47207_R1	CGG-ACA-CGG-TCC-TGA-GAA-AA TAT-CAC-GAG-CAT-ACG-CCAQ-CC	113
Phatr3_J48731	probable cytosolic oligopeptidase a	Phatr48731_F1/ Phatr48731_R1	GGC-GAA-CAC-TCT-GGA-GGA-AA GCA-AGC-TTT-GTA-cGG-CTC-TG	76
Phatr3_EG02376	dead deah box rna	Phatr33072_F1/ Phatr33072_R1	CTT-GGT-GCG-ATC-CTT-CTG-GT CAT-TGA-GGC-CCC-CTT-TTT-GC	144
Phatr3_J45764	Histone acetyltransferase HAC1	Phatr45764_F1/ Phatr45764_R1	GTC-AGT-TCC-CGG-TAC-GTG-TT AAT-GGC-TTC-CCT-TAC-TGG-GC	84
Phatr3_J43542	predicted protein	Phatr43542_F1/ Phatr43542_R1	CGT-CGA-CCG-AAG-TTG-TTA-CG GTT-GTG-GTC-GAA-CTC-GGA-GG	131
Phatr3_J50885	glycyl-trna synthetase subunit beta	Phatr50885_F1/ Phatr50885_R1	GAT-CCG-GCA-AGA-CCT-GCT-C CGG-TTC-AAT-GCA-ATG-CGG-AA	146
Phatr3_J27639	agc grk bark protein kinase	Phatr27639_F1/ Phatr27639_R1	GCG-TCA-TGC-CTG-AAG-ACT-TT TTA-CAG-GCT-GTC-ACG-AGT-CC	73
Phatr3_J47374	mfs general substrate transporter	Phatr47374_F1/ Phatr47374_R1	AGG-GCG-TGT-AGA-AAC-GAC-AG CCC-CGT-TCG-AAT-CAA-GTA-AAC-C	142
Phatr3_J19761	glycine-trna ligase	Phatr19761_F1/ Phatr19761_R1	CCC-GAA-TTC-AAC-CGA-AAG-GC ATG-GCA-GCC-TTG-AGA-GAG-C	134
Phatr3_J44001	LPS A	Phatr44001_F1/ Phatr44001_R1	AAC-TGG-TAT-GCA-CCA-CCC-TG CGA-AGC-AGC-GAT-CGA-ACT-TG	123

Table 29: List of the candidate new reference genes selected from a RNA-Seq data analysis, with the primer sequences used in qRT-PCR analysis and the amplicon size.

The sequence name is given according to the EnsemblProtist database.

Genes successfully validated in our experimental design (efficiency above 95%) are colored in green, and non-validated genes (efficiency below 90%) are colored in red. Genes having an efficiency comprise between 90% and 95% are represented in yellow.

5.5.2 qRT-PCR primer set validation

5.5.2.1 Amplification specificity

With the aim to verify that the primer pairs are specific to the intended PCR template, each primer pair was used for PCR amplifications of cDNA of the three different morphotypes, followed by an agarose gel electrophoresis to check that only a single PCR product of the expected size was obtained. In this case, the amplified fragment was purified and a direct sequencing of the PCR product was carried out. Some primer pairs were dismissed at this PCR amplification control step as no or very low amplification of their corresponding genes was obtained in the standard PCR conditions. That was the case for the Phatr3_J50885 (glycyl-tRNA synthetase subunit beta) and the Phatr3_J27639 (agc grk bark protein kinase), 2 genes identified from our RNA-Seq data analysis. Another primer pair was not submitted to the subsequent selection steps because 2 bands were detected on the control gel (Phatr3_J43542,

unknown protein, data not shown). The specificity of the PCR reaction was also verified by a melt-curve analysis carried out at the end of the PCR cycles in the qPCR assays and no problem was detected with the primer pairs tested.

5.5.2.2 Efficiency of amplification

The efficiency of the qPCR reaction was determined for each primer pair with cDNA from each of the 3 morphotypes by establishing standard curves from dilution series (Pfaffl, 2001). The primer pairs showing efficiencies in the 92-105 % range were selected to be tested as reference genes in qRT-PCR experiments. Most of the primer pairs were excluded at this step and finally primer pairs designed for 3 common reference genes: TubB (Phatr3_J21122), Actin1 (Phatr3_J29812) and Actin2 (Phatr3_J29136), already tested by Siaux *et al.* (2007) and also for 4 new potential reference genes (**Table 30**) were retained for further analysis in qRT-PCR assays to find among them the most stably expressed genes in the experimental conditions described here.

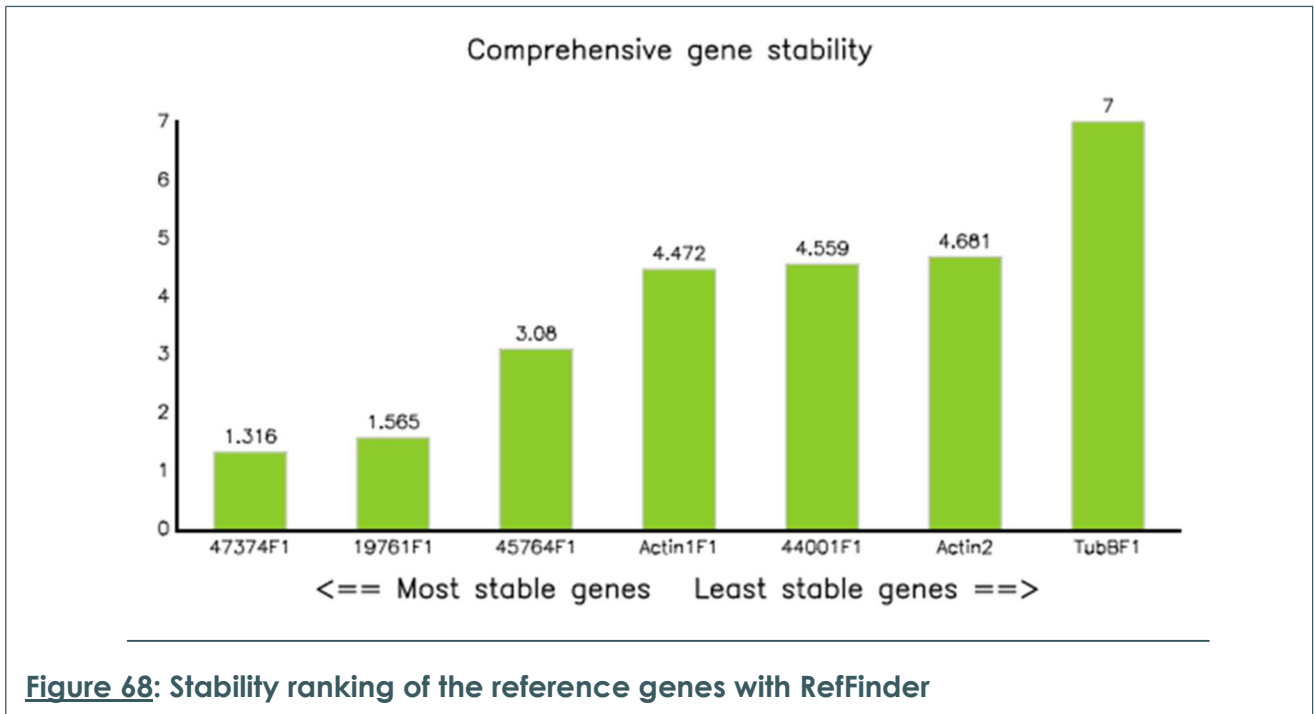
Class	Sequence name Ensembl Protists release 25 - January 2015 © EBI	Sequence description	Primer name (Forward/Reverse)	Sequences (5'-3') Forward Reverse	Amplicon size (bp)	Amplification efficiency (%)			Ct Mean		
						F	O	T	F	O	T
Literature-based	Phatr3_J21122	Tubulin beta-2A chain (Tub B)	TubBF1/TubBR1	CCG-TTC-GCT-CGT-TTG-ATT-CC GGC-AGA-CAT-GAC-CTC-CCA-AA	125	94.6	97.7	98.3	24,641	27,652	23,778
	Phatr3_J29136	actin2	Actin2F2/Actin2R2	GCG-CAG-TTC-GTT-CTT-ATA-ATC-CA CAA-ATC-CGG-CCT-TGC-ACA-TT	157	96.0	95.9	83.9	29,391	31,870	29,581
	Phatr3_J29812	actin1	Actin1F1/Actin1R1	GCC-TTA-CAC-CGC-CGT-ATT-CT TCA-ATG-ACA-AGG-GCA-GCG-AT	79	97.5	98.4	98.1	21,068	22,304	21,335
From RNA-seq data	Phatr3_J19761	glycine-trna ligase	Phatr19761_F1/ Phatr19761_R1	CCC-GAA-TTC-AAC-CGA-AAG-GC ATG-GCA-GCC-TTG-AGA-GAG-C	134	91.4	96.5	93.3	23,521	25,043	23,593
	Phatr3_J44001	LPS A	Phatr44001_F1/ Phatr44001_R1	AAC-TGG-TAT-GCA-CCA-CCC-TG CGA-AGC-AGC-GAT-CGA-ACT-TG	123	96.1	94.3	93.9	26,621	27,894	26,791
	Phatr3_J45764	Histone acetyltransferase HAC1	Phatr45764_F1/ Phatr45764_R1	GTC-AGT-TCC-CGG-TAC-GTG-TT AAT-GGC-TTC-CCT-TAC-TGG-GC	84	93.3	101.3	94.2	22,131	24,349	22,220
	Phatr3_J47374	mfs general substrate transporter	Phatr47374_F1/ Phatr47374_R1	AGG-GCG-TGT-AGA-AAC-GAC-AG CCC-CGT-TCG-AAT-CAA-GTA-AAC-C	142	96.8	96.7	96.3	25,192	26,496	25,681

Table 30: List of the candidate reference genes tested for the normalization of qRT-PCR expression studies in the 3 morphotypes of *Phaeodactylum tricornutum*. The sequences of the primers are given with the amplicon size. The amplification efficiency is given for each primer pair, in each morphotype.

5.5.3 Expression stability of the tested reference genes and validation of reference genes

After the preliminary steps of validation of a set of 7 primer pairs specific to candidate reference genes, the expression of these genes was analyzed in qRT-PCR assays in order to evaluate their expression stability in the three morphotypes in the experimental conditions that were defined. The qRT-PCR data were obtained from an experiment in which the target genes were amplified from 4 biological replicates of each morphotype (O, F and T) and 2 technical replicates were carried out for each biological replicate. In order to select the best reference genes, the data were analyzed by different dedicated algorithms leading to the following results. Using the Normfinder program (Andersen et al., 2004), a stability value is calculated for each potential reference gene, the best candidate having the lowest stability value. We performed a NormFinder analysis as recommended in the user manual and determined that in our experiment the best gene was Phatr3_J19761 (glycyl-tRNA ligase) followed by the Phatr3_J47374 gene (mfs general substrate transporter). The best combination of 2 genes for a 2 genes normalization factor are Phatr3_J21122 (TubB) and Phatr3_J47374. The stability ranking order (from the most stable to the least stable gene) was the following: Phatr3_J19761> Phatr3_J47374> Phatr3_J45764> Actin1> TubB> Phatr3_J44001> Actin2.

The RefFinder tool (Xie et al., 2012) which integrates four different programs was used to determine the comprehensive rankings of the candidate reference genes. Phatr3_J47374 and Phatr3_J19761 were identified as the 2 most stable genes in all the samples (**Figure 68**). This is in agreement with the results obtained with the NormFinder analysis and described above. Looking at the RefFinder outputs shows that Phatr3_J 47374 and Phatr3_J19761 are the 2 genes ranked at the top whatever the algorithm used to determine the most stable reference genes. Both genes were then used in a qRT-PCR analysis studying the expression of genes involved in the *N*-glycosylation processing.



5.5.3.1 Analysis of the expression patterns of a set of *N*-glycosylation genes by qRT-PCR by using the newly selected reference genes

The *N*-glycosylation is a major co- and post-translational event in the synthesis of proteins in eukaryotes. This modification is occurring within the endoplasmic reticulum (ER) and Golgi apparatus. Recently, *N*-glycan structures of endogenous proteins from *P. tricornutum* have been analyzed and found to be high mannose type *N*-glycans with tiny amount of complex fucosylated *N*-glycans (Baïet et al., 2011). Bio-informatic analysis of *P. tricornutum* genome allow the identification of putative enzymes involved in the different steps of the *N*-glycosylation processing. For example, a complete set of sequences potentially encoding for proteins involved in the synthesis of the lipid-linked precursor have been found (Baïet et al., 2011). Subunits of the oligosaccharyltransferase which is responsible of the transfer of the precursor onto the nascent protein to give birth to glycoprotein have also being identify. Enzymes and chaperones involved in the ER modifications of the *N*-glycans, glycosidases and glycosyltransferases acting in the Golgi maturation as well as nucleotide-sugar transporters were also predicted in the genome (Baïet et al., 2011; Mathieu-Rivet et al., 2014). However, no information regarding the expression level of the genes encoding for these glycoenzymes is currently available in *P. tricornutum*. Therefore, in this study a set of genes encoding proteins that are belonging to the *N*-glycosylation pathway of *P. tricornutum* were selected to evaluate their gene expression levels in the 3 morphotypes by qRT-PCR analysis using the 2 newly selected reference genes. Primer pairs specific to 9 of these *N*-glycosylation genes were

validated and used in a qRT-PCR experiment. Among these genes, 2 are encoding for putative enzymes located in the ER: Phatr3_EGO2351 (catalytic subunit of the α -Glucosidase II) and Phatr3_EGO2374 (α -mannosidase I). In contrast, Phatr3_J54844 (β 1,2-*N*-acetylglucosaminyltransferase, GnTI), Phatr3_J52248 (α -mannosidase II), Phatr3_J46109, Phatr3_J46110 and Phatr3_J54599 (putative fucosyltransferases); Phatr3_J43174 (GDP-Fucose transporter) are genes encoding proteins located in the Golgi apparatus. In addition, the Phatr3_J45073 encoding for a putative β -*N*-acetylhexosaminidase involved in the degradation of the complex-type *N*-glycans, has also been included in the analysis. The qRT-PCR data were analyzed as described in the Materiel and Methods section using the 2 most stable reference genes identified (Phatr3_J19761 and Phatr3_J 47374) in combination to normalize the data. The results are presented in **Figure 69**.

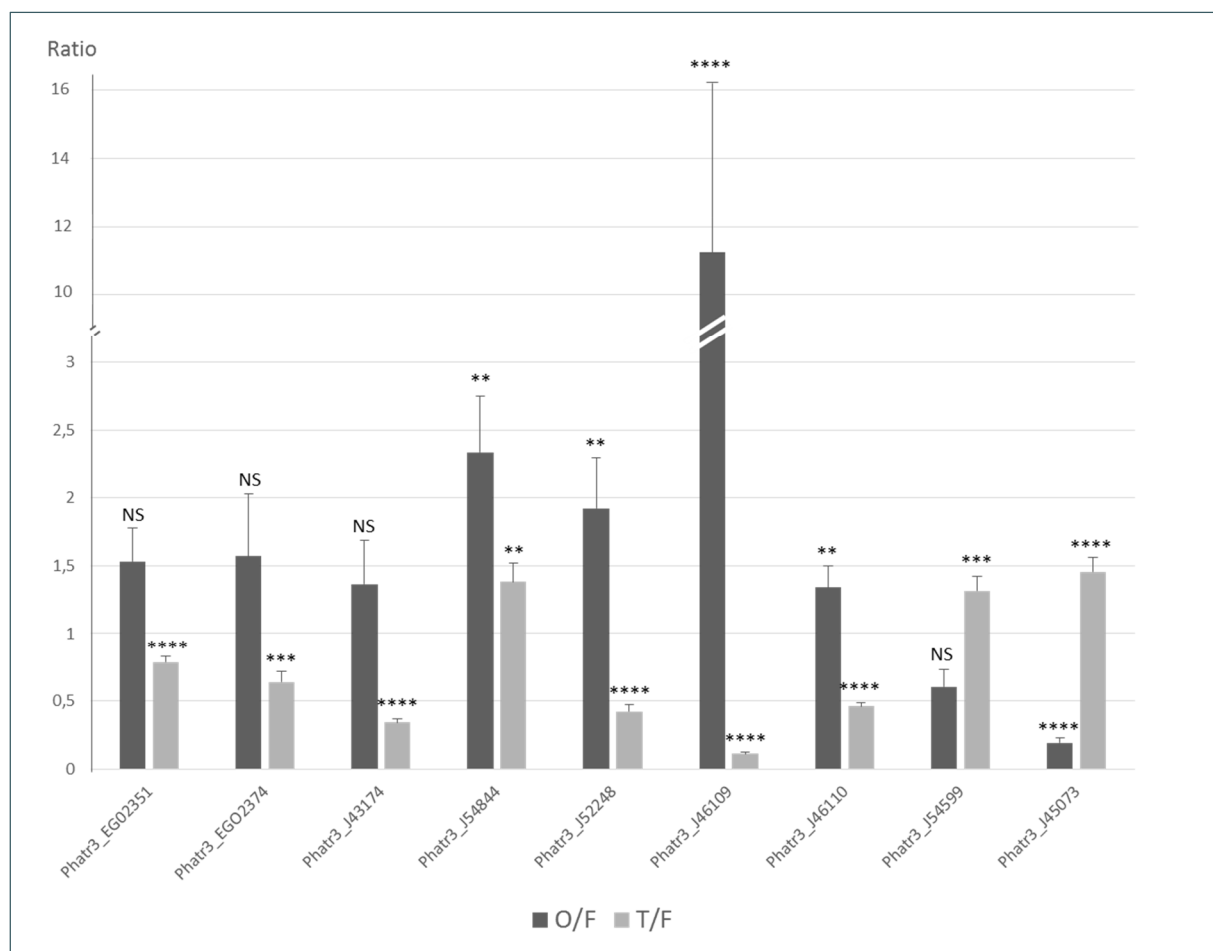


Figure 69: Histogram representing the ratio of normalized gene expression for genes belonging to the *N*-glycosylation pathway comparing the oval and fusiform morphotypes and the triradiate and fusiform morphotypes respectively.

Ratios have been performed with normalized gene expression between each one of the 4 biological replicates of each morphotype compared, leading to 16 ratio values per morphotype comparison. These 16 values were used to perform a Mann-Whitney test against the value 1, which represent theoretically a non-differentially expressed ratio. The 16 ratios were also used to calculate the mean and the SEM which are reported in the figure.

The relative expression of the different genes is showed as pairwise comparisons of the O and T morphotypes against the F morphotype, considering that the latest is the basic morphotype which is the most common one found in *P. tricornutum* cultures (De Martino et al., 2007; Zhao et al., 2014) and which is corresponding to the major morphotype of the Pt1.8.6 strain for which the genome has been sequenced (Bowler et al., 2008; De Martino et al., 2007). Various expression levels were observed for the *N*-glycosylation genes and no relationship between the expression level and the compartment where the enzymes are supposed to act can be established. With the exception of the Phatr3_J45073 gene encoding for a putative β -*N*-acetylhexosaminidase and the Phatr3_J54599 encoding for a putative fucosyltransferase which are down-regulated in the oval morphotype, all the others show a highest expression in the oval morphotype with a particularly high level of transcripts for the Phatr3_J46109 gene. Altogether, these results suggest that the *N*-glycosylation processing can be up-regulated in the oval morphotype which could lead to an accumulation of complex-type *N*-glycans. Indeed, over-expression of the Phatr3_J54844 gene encoding for the β 1,2-*N*-acetylglucosaminyltransferase, the Phatr3_J52248 gene encoding for the alpha-mannosidase II, the Phatr3_J46109 and Phatr3_J46110 encoding for the putative fucosyltransferases together with the increase in the expression of the Phatr3_J43174 gene (GDP-Fucose transporter) and the decrease in the expression of the Phatr3_J45073 gene encoding for a putative β -*N*-acetylhexosaminidase suggest that more complex-type *N*-glycans are synthesized within the Golgi apparatus and accumulated within the oval morphotype of *P. tricornutum*. Similar trend has been observed in the RNA-Seq data (Chapter III).

5.6 Discussion

The qRT-PCR method is now widely used to quantify RNA levels in gene expression studies. In order to produce reliable results, the data need to be normalized with stable reference genes. For a long time, the same set of housekeeping genes were used in all the organisms and whatever the biological conditions used but in the last years, an effort was made to identify new reference genes. The availability of growing transcriptomic data from microarrays analysis or more recently from RNA-Seq experiments provides the means to search novel reference genes, generally more stable than the commonly used.

In the diatom *P. tricornutum*, Siaut et al. (2007) have tested a set of 12 common housekeeping genes and have identified genes suitable for use to normalize qPCR data. These genes have been generally used in the few gene expression analysis carried out since then. We also tested them to determine if they would be suitable as reference genes in our experimental

conditions with the aim to compare the gene expression of targets genes among the 3 morphotypes. But in our work, only three of the genes were selected for further validation.

As a consequence, we took advantages of our RNA-Seq datasets that we generated from the three morphotypes of *P. tricornutum* to identify stably expressed genes which could be used as good references genes for qRT-PCR analysis. Among the different genes tested, only two were validated and had been chosen as efficient references genes. Therefore, we proposed to use Phatr3_J19761 and Phatr3_J47374 as reference genes suitable for qRT-PCR in the three morphotypes of *P. tricornutum*. These references genes were then used in combination to normalize the expression of 9 genes involved in the *N*-glycosylation pathway of *P. tricornutum*. The results represent the first attempt to analyze and compare *N*-glycosylation gene expression over the three morphotypes from *P. tricornutum*. Altogether, the qRT-PCR data suggest that suggests that more complex-type *N*-glycans are synthesized within the Golgi apparatus and accumulated within the oval morphotype of *P. tricornutum*. Therefore, in the future, comparative glycomic analysis performed on glycoproteins extracted from the 3 morphotypes from *P. tricornutum* would be necessary to confirm the accumulation of complex-type *N*-glycans within the oval morphotype. However, the qRT-PCR result is already coherent with the fact that the oval morphotype from *P. tricornutum* is described to be tolerant to stresses (De Martino et al., 2007) and previous reports from the literature which are describing roles for *N*-/*O*-glycoproteins in the control of growth, morphogenesis and adaptation to biotic and abiotic stresses in plants (Nguema-Ona et al., 2014).

Chapitre V :
Caractérisation des
fucosyltransférases putatives de
P. tricornutum

6 Caractérisation des fucosyltransférases de *P. tricornutum*

La voie de maturation des *N*-glycannes complexes chez les eucaryotes fait intervenir différentes glycosidases et glycosyltransférases propres à chaque organisme. Ces enzymes vont intervenir au sein de l'appareil de Golgi pour aboutir à la formation de *N*-glycannes complexes possédant des motifs glucidiques spécifiques à chaque espèce. Parmi ces glycosyltransférases, les fucosyltransférases catalysent l'ajout de résidus fucose sur les GlcNAc proximales ou distales de la structure glycanique (Johnson and Chrispeels, 1987; Kubelka et al., 1993). L'existence de fucosyltransférases putatives a été mise en évidence chez *P. tricornutum* lors d'une étude précédente menée dans le cadre de la thèse de Bérengère BAIET au sein du laboratoire Glyco-MEV. Ces travaux ont démontré la présence d'une population majoritaire de *N*-glycannes oligomannosidiques sur les protéines endogènes de la microalgue, mais aussi la présence d'une population mineure de *N*-glycannes complexes pouvant être fucosylés (Baïet et al., 2011). Parallèlement, l'analyse *in silico* du génome de *P. tricornutum* a permis de mettre en évidence par homologie de séquence avec les α 1,3-FuT A et B d'*A.thaliana* (référence Uniprot : 9LJK1 et référence Uniprot Q9FX97, respectivement), et avec l' α 1,3-FuT A de *D. melanogaster* (référence Uniprot : Q9VUL9), la présence de 3 gènes (PHATRDRAFT_46109, PHATRDRAFT_461010 et PHATRDRAFT_54599) codant potentiellement pour des α 1,3-fucosyltransférases, appelées d'après leurs identifiants protéiques (I.D.): 46109, 46110 et 54599 (Baïet et al., 2011). Ces trois gènes coderaient pour des protéines possédant tous les motifs structuraux caractéristiques des fucosyltransférases appartenant à la famille des GT10 (Baïet et al., 2011). Il est ainsi possible de retrouver sur la séquence de ces protéines putatives, la présence des motifs I, II et la séquence SNC(G/A)A(R/H)N potentiellement spécifiques des α 1,3-fucosyltransférases de plante (Wilson et al., 2001) et de *D. melanogaster* (Fabini et al., 2001), ainsi que le motif CXXC situé à l'extrémité C-terminale. Ce dernier motif est notamment impliqué dans le repliement de la protéine par la formation de ponts disulfure avec les résidus cystéine présents à l'extrémité N-terminale (Holmes et al., 2000).

L'objectif de ce chapitre de résultats est de poursuivre la caractérisation des fucosyltransférases de *P. tricornutum* en investiguant de nouveau leurs séquences protéiques, en réalisant une étude de la localisation subcellulaire de ces enzymes chez *P. tricornutum* et d'initier leur caractérisation fonctionnelle par complémentation hétérologue.

6.1 Analyse bio-informatique des gènes PHATRDRRAFT_46109, PHATRDRRAFT_46110 et PHATRDRRAFT_54599 de *P. tricornutum*

6.1.1 Analyse du gène PHATRDRRAFT_54599

Les analyses bio-informatiques concernant le gène PHATRDRRAFT_54599 de la fucosyltransférase 54599 (FuT54599) prédisent un transcrit mono-exonique codant pour une protéine constituée de 481 acides aminés. La taille prédite pour cette protéine est similaire à celles d'autres α 1,3-fucosyltransférases spécifiques du core glycanne qui ont été préalablement caractérisées chez les plantes et les invertébrés (Paschinger et al., 2004; Wilson et al., 2001). En effet, ces enzymes sont constituées d'environ 500 acides aminés. La FuT54599 posséderait une conformation typique de glycosyltransférase (**figure 28** de la partie synthèse bibliographique), avec une structure de protéine membranaire de type II comportant une courte queue cytosolique en *N*-terminal, un domaine transmembranaire (**figure 70**), ainsi qu'une grande séquence luminale contenant le domaine catalytique.

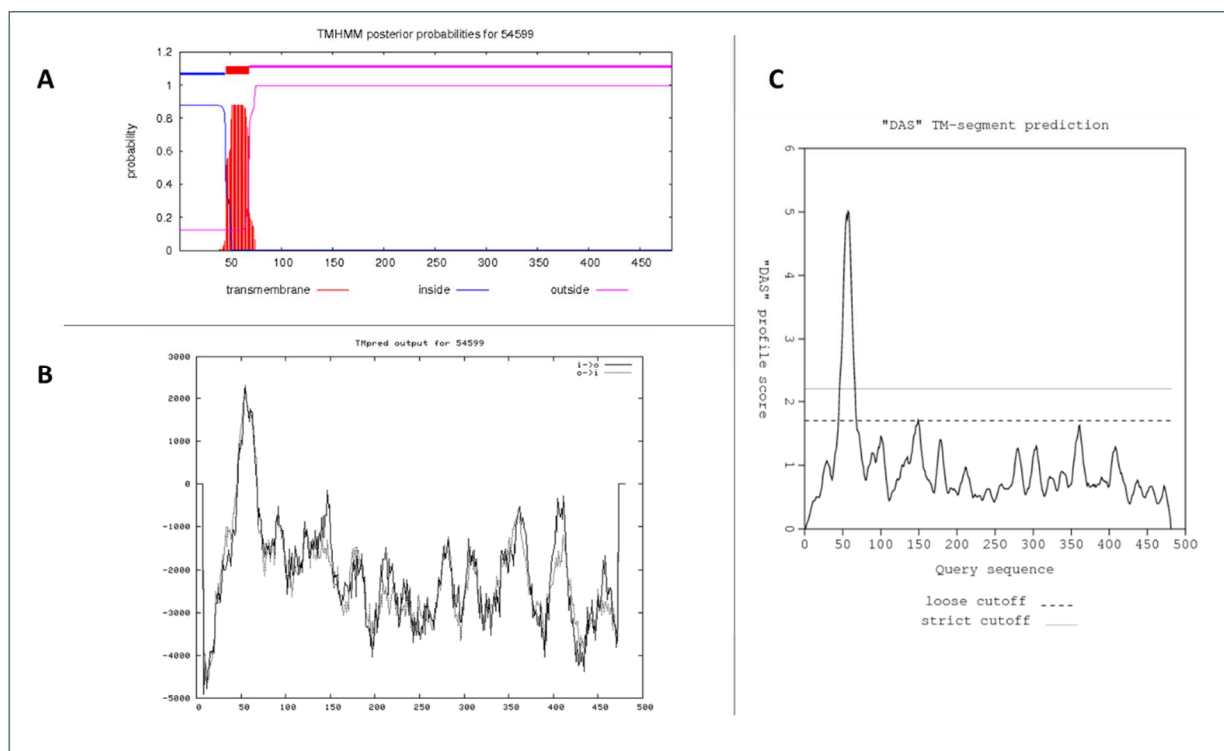


Figure 70 : Prédictions de domaines transmembranaires sur la séquence protéique FuT54599 par les programmes TMHMM (A), TMpred (B) et DAS (C).

A : Prédiction d'une protéine possédant un domaine transmembranaire situé entre les acides aminés 46 à 68, avec une extrémité *N*-terminale cytosolique et une extrémité *C*-terminale luminale

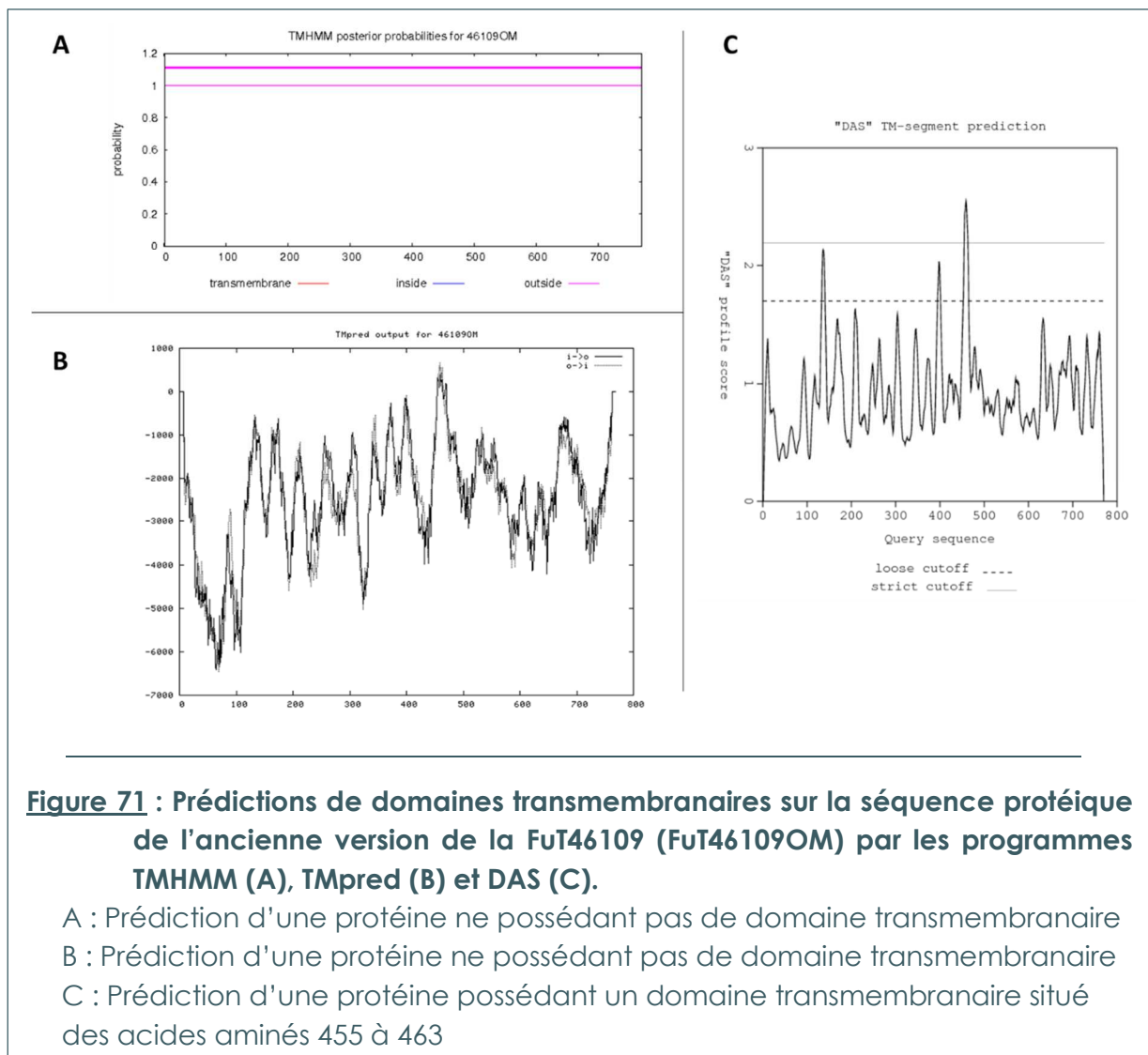
B : Prédiction d'une protéine possédant un domaine transmembranaire situé entre les acides aminés 46 à 66

C : Prédiction d'une protéine possédant un domaine transmembranaire situé entre les acides aminés 46 à 66

6.1.2 Analyse du gène PHATRRAFT_46109

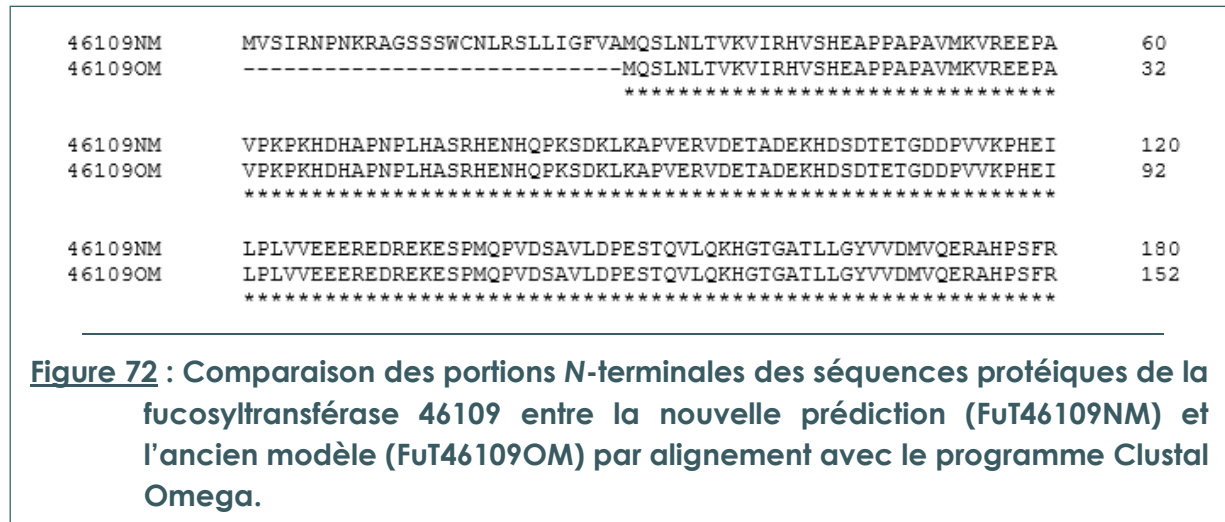
6.1.2.1 Version originale

Les prédictions bio-informatiques correspondant à l'annotation PHATRRAFT_46109 du gène 46109 indiquent la présence d'un transcrit mono-exonique codant potentiellement pour une protéine de 770 acides aminés (FuT46109OM). La taille de cette protéine est donc étonnamment grande puisqu'elle possède 237 acides aminés de plus que l' α 1,3-FuT de *P. patens* qui est une des plus longues α 1,3-FuT du core (533 acides aminés (Koprivova et al., 2003)). D'après les outils de prédiction, la FuT46109OM ne possède pas de domaine transmembranaire (TMD) évident en *N*-terminal. Cette absence de TMD prédit pour cette protéine (**figure 71**), qui en tant que fucosyltransférase devrait plutôt présenter une structure protéique membranaire de type II, nous a amené à remettre en question la prédiction existante et à proposer une nouvelle version de la séquence codante du gène PHATRRAFT_46109.

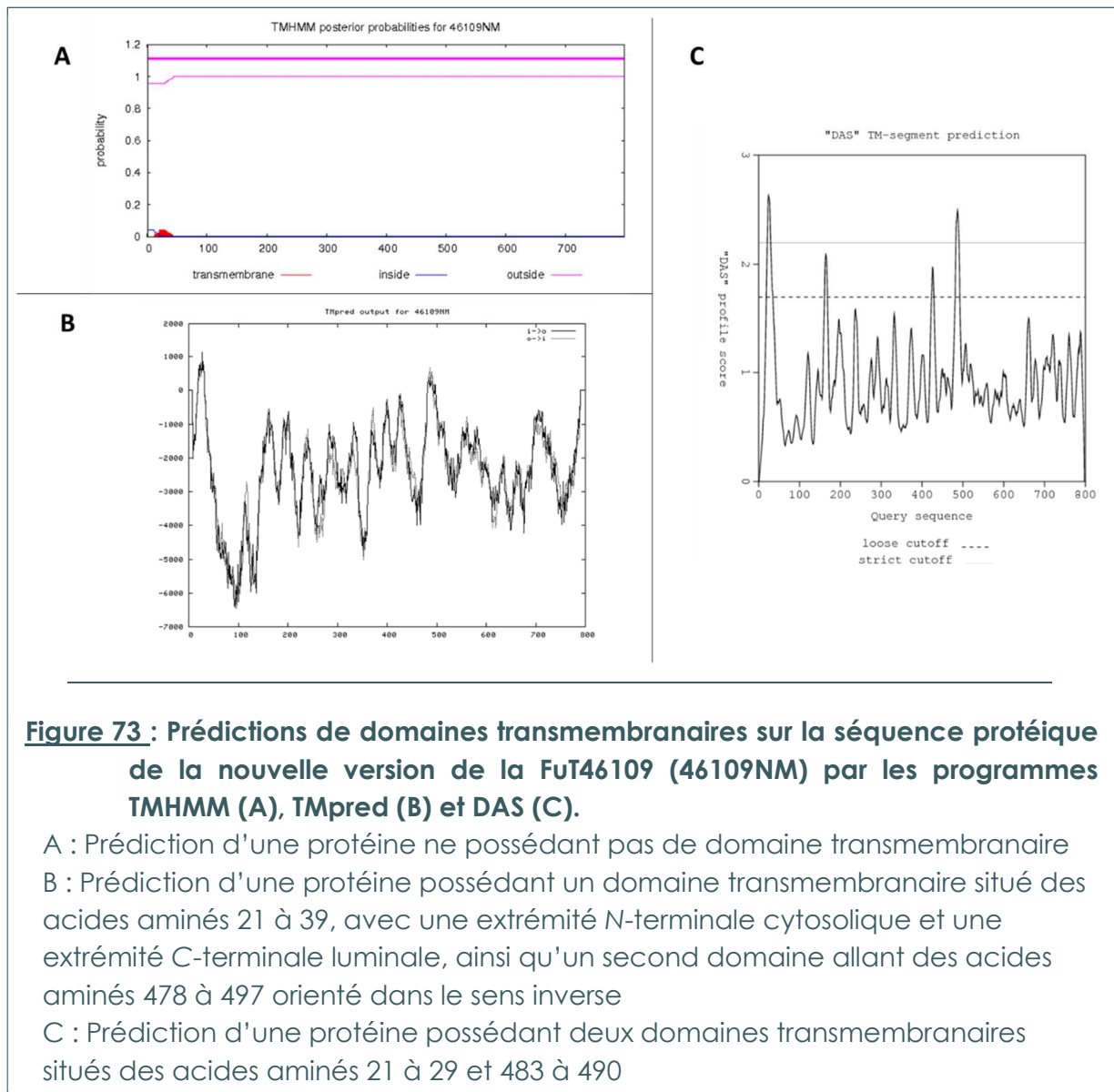


6.1.2.2 Nouvelle version

Une nouvelle version de la séquence codante du gène PHATRDRAFT_46109, que nous appellerons 46109NM, a été établie en prenant comme codon initiateur le codon ATG qui précède en 5' la précédente annotation du gène. L'ajout de cette nouvelle portion de 84 nucléotides en 5' ne change pas le cadre de lecture et n'introduit pas de codon stop (**figure 72**).



Cette nouvelle séquence codante permettrait ainsi la traduction d'une protéine de 798 acides aminés (FuT46109NM). Cependant, bien que les scores obtenus lors des analyses bio-informatiques soient plus élevés que pour la précédente annotation (**figure 73**), ils ne nous permettent toujours pas d'affirmer la présence d'un domaine transmembranaire sur la nouvelle protéine prédite. Ainsi, même si la nouvelle prédiction semble de meilleure qualité, l'étude de la localisation subcellulaire de la FuT46109 sera réalisée avec les deux séquences 46109OM et 46109NM afin de comparer l'implication de la nouvelle portion N-terminale et d'observer une éventuelle différence de localisation entre les deux modèles de FuT46109. Au contraire, l'étude fonctionnelle sera effectuée uniquement avec la 46109NM. En effet, la séquence correspondante au site catalytique de l'enzyme n'étant pas modifiée par l'ajout de la nouvelle séquence en N-terminal, aucune différence ne devrait être observée entre l'activité de la FuT46109OM et la FuT46109NM *in vitro*. Cependant, *in vivo*, le bon adressage de la glycosyltransférase à l'appareil de Golgi est nécessaire pour pouvoir observer une éventuelle modification des N-glycannes. Or chez les glycosyltransférases, l'adressage de la protéine semble conditionné par la séquence en N-terminal composant la queue cytoplasmique et le domaine transmembranaire (Saint-Jore-Dupas et al., 2006). Ainsi, la FuT46109NM qui a plus de chance de posséder ces deux domaines a été sélectionnée pour l'analyse fonctionnelle. L'ensemble de ces résultats pourrait conforter la nouvelle prédiction du gène PHATRDRAFT_46109 par une approche *in vivo*.



6.1.3 Analyse du gène PHATRDRRAFT_46110

6.1.3.1 Version originale

Contrairement aux deux autres gènes PHATRDRRAFT_54599 et PHATRDRRAFT_46109, celui codant pour la fucosyltransférase 46110 posséderait deux introns selon l'annotation PHATRDRRAFT_46110. Cette séquence coderait pour une protéine prédite de 718 acides aminés (FuT46110OM). Les analyses bio-informatiques de sa séquence ne permettent pas de détecter de domaine transmembranaire (**figure 74**).

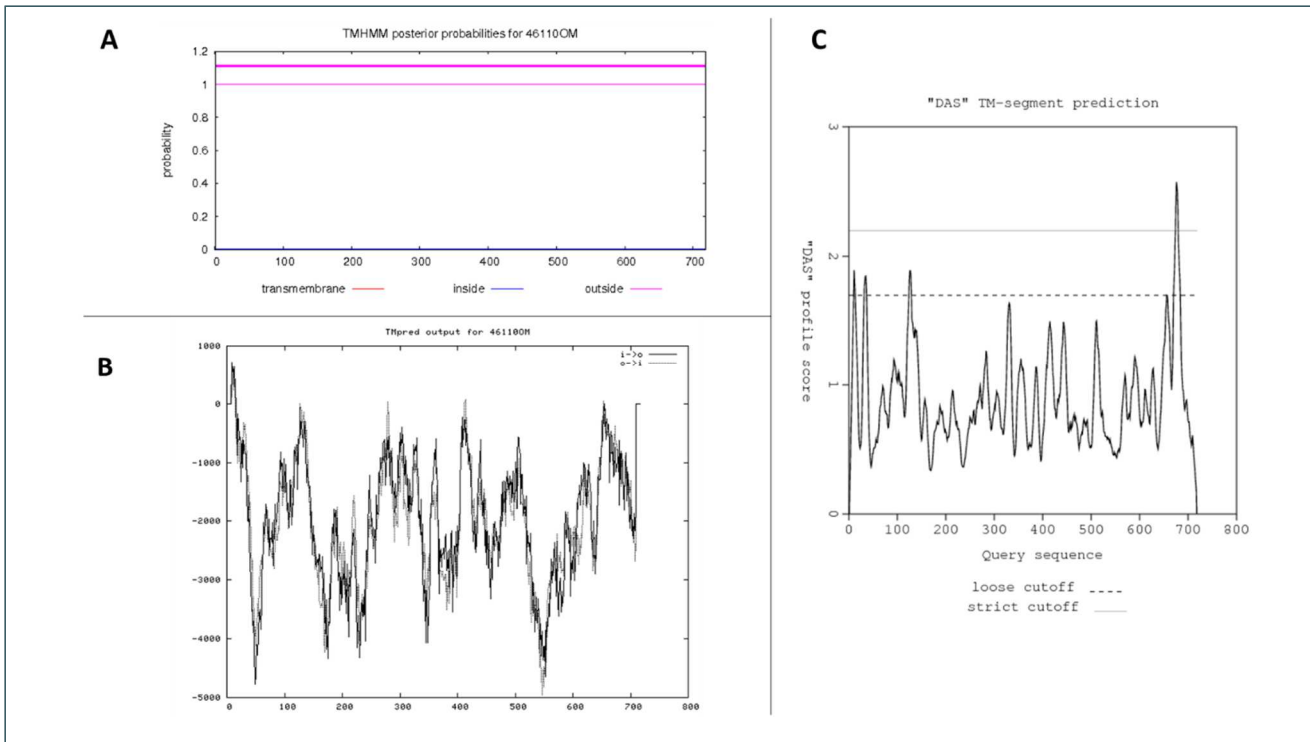


Figure 74 : Prédications de domaines transmembranaires sur la séquence protéique de l'ancienne version de la FuT46110 (FuT46110OM) par les programmes TMHMM (A), TMpred (B) et DAS (C).

A : Prédiction d'une protéine ne possédant pas de domaine transmembranaire
 B : Prédiction d'une protéine ne possédant pas de domaine transmembranaire
 C : Prédiction d'une protéine possédant un domaine transmembranaire situé des acides aminés 675 à 679

Ainsi, tout comme pour l'annotation PHATRDRRAFT_46109, l'absence de cette structure attendue sur la protéine prédite nous a conduits à repenser l'annotation du gène 46110. De plus, des analyses menées par RT-PCR nous ont permises de remettre en question l'existence du premier intron de la séquence annotée. En effet, des amplifications PCR ont été réalisées sur l'ADNc du gène 46110 en utilisant des amorces encadrant les deux introns prédits (**figure 75 A**). Ainsi, en ce qui concerne le premier intron, un amplicon de 377pb est attendu dans le cas où un événement d'épissage aurait lieu, or nous observons par migration électrophorétique que l'amplicon obtenu possède une taille de 612pb, ce qui correspond à la séquence amplifiée sans épissage de l'intron (**figure 75 B**). Concernant le second intron, un amplicon de 478 pb est attendu dans le cas d'un épissage, et un fragment de 562pb si la séquence n'est pas épissée. Ici, nous observons la présence d'un amplicon à 478pb correspondant à l'épissage du second intron (**figure 75 B**). De plus, par comparaison avec une amplification réalisée sur l'ADN génomique, les résultats observés confirment bien que la séquence nucléique située sur le premier intron de l'annotation PHATRDRRAFT_46110 n'est pas épissée contrairement au second intron (**figure 75 B**). Ce constat a pu être établi quel que soit le

morphotype de *P. tricornutum* étudié. Dès lors, ces résultats nous permettent de proposer une nouvelle prédiction pour le gène PHATRRAFT_46110 en considérant le 1^{er} intron comme une séquence codante.

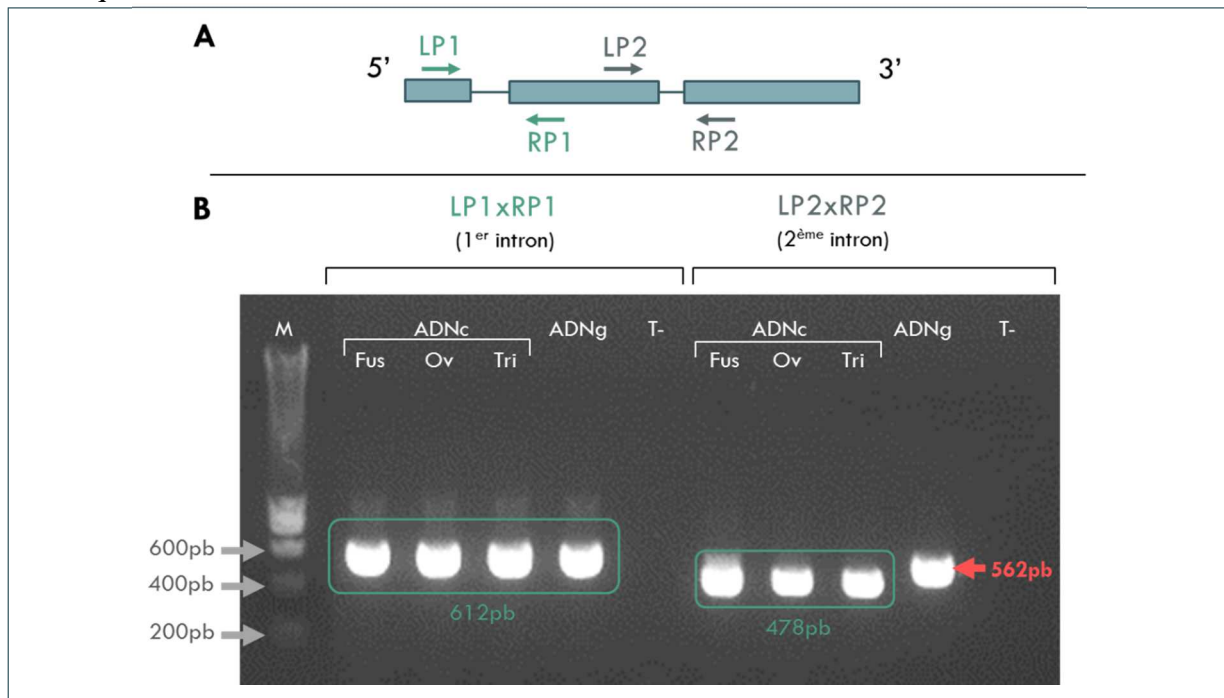


Figure 75 : Contrôle de l'épissage des introns de la séquence 46110OM par PCR.

A : Représentation schématique du positionnement des couples d'amorces utilisées pour la PCR sur la séquence 46110OM.

B : Gel de migration des fragments obtenus lors de la PCR. Fus = morphotype fusiforme ; Ov = morphotype ovale ; Tri = morphotype triradié

6.1.3.2 Nouvelle version

La nouvelle séquence établie pour le gène PHATRRAFT_46110, que nous appellerons 46110NM, coderait donc pour une protéine de 707 acides aminés (FuT46110NM), modifiant uniquement la séquence prédite sur les 104 acides aminés de l'extrémité *N*-terminale (**figure 76**). En effet, cette nouvelle version a été établie de manière à conserver le cadre de lecture qui permet ainsi la préservation du domaine catalytique.

46110OM	MIRIPYAVSYLACSAPFIGNHSQWEGDPVAAISSAV----IYHLKPTKEEGRIRACQRPV	56
46110NM	-MRVS-----LRQPPRMHCRCTWSLFLVLLCIDILCMLLLHRKAHVAVAPI-PWSPPT	52
	:* : . * : . * . * . : : * * . * . *	
46110OM	GNEVRHPIGKVSSIDVKPTPTFLDRKHTLTSINHARVFATAPTDALPLTDGKLVHLTDSN	116
46110NM	NNAVRPNTR-----NRHQNSFVTSEFLMLGTN-SQLEQGRDQPMMSGTDGKLVHLTDSN	105
	* ** : .*: . : . * : : *****	
46110OM	NPMVLSEAFVSVIKGLEASLQISLNLPTLRSCRWKGTDLTIHNVCDQAGSEQKNQRPKTV	176
46110NM	NPMVLSEAFVSVIKGLEASLQISLNLPTLRSCRWKGTDLTIHNVCDQAGSEQKNQRPKTV	165

46110OM	VYYNRFSEQERIICGRSIPPNTHLQLDELCEEPSRLYAVLPDRITAKGLPPTTVRFNDPQRR	236
46110NM	VYYNRFSEQERIICGRSIPPNTHLQLDELCEEPSRLYAVLPDRITAKGLPPTTVRFNDPQRR	225

Figure 76 : Comparaison des portions *N*-terminales des séquences protéiques de la fucosyltransférase 46110 entre la nouvelle prédiction (FuT46110NM) et l'ancien modèle (FuT46110OM) par alignement avec le programme Clustal Omega

La protéine ainsi modifiée comporterait un domaine transmembranaire en *N*-terminal (**figure 77**). Pour les mêmes raisons qu'exposés précédemment pour le gène PHATRDRAFT_46109, cette nouvelle séquence codante proposée pour le gène PHATRDRAFT_46110 sera utilisée pour l'étude fonctionnelle, et a été utilisée en parallèle de l'ancienne version pour l'étude de la localisation subcellulaire de la protéine réalisé dans le cadre de ce travail de thèse.

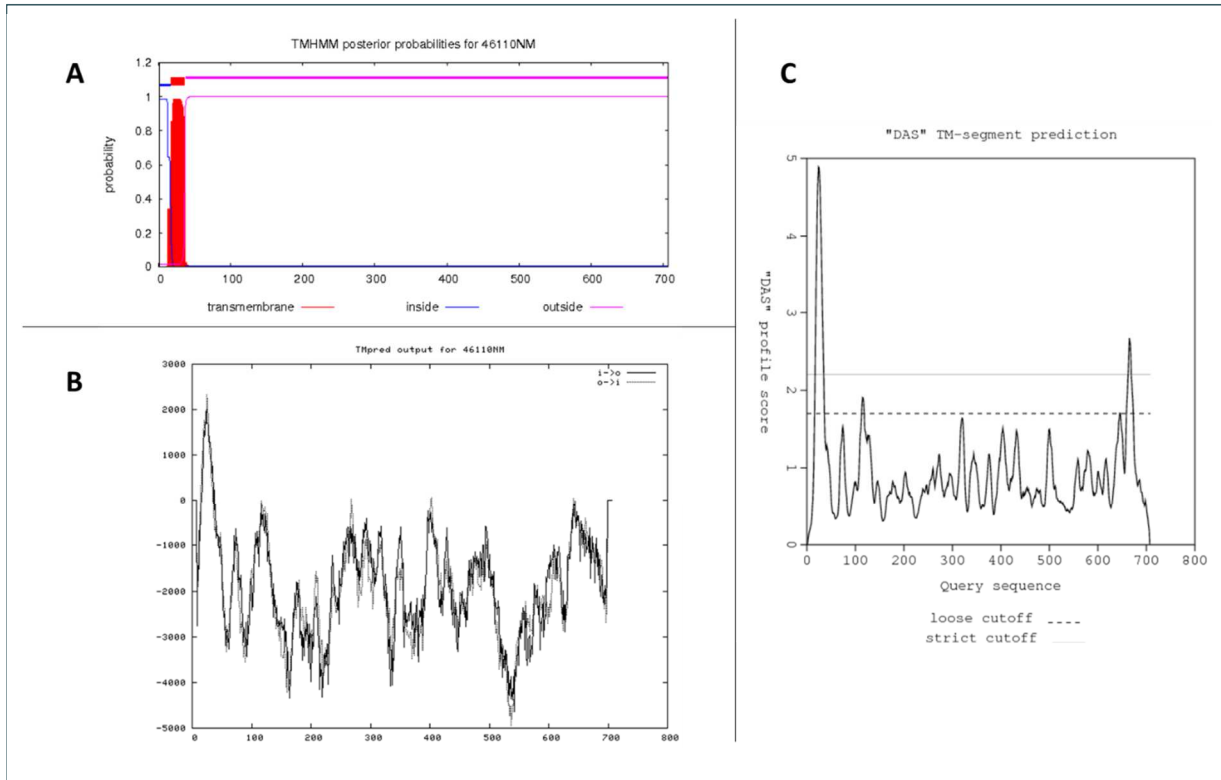


Figure 77 : Prédictions de domaines transmembranaires sur la séquence protéique de la nouvelle version de la FuT46110 (FuT46110NM) par les programmes TMHMM (A), TMpred (B) et DAS (C).

A : Prédiction d'une protéine possédant un domaine transmembranaire situé des acides aminés 17 à 36, avec une extrémité *N*-terminale cytosolique et une extrémité *C*-terminale luminale

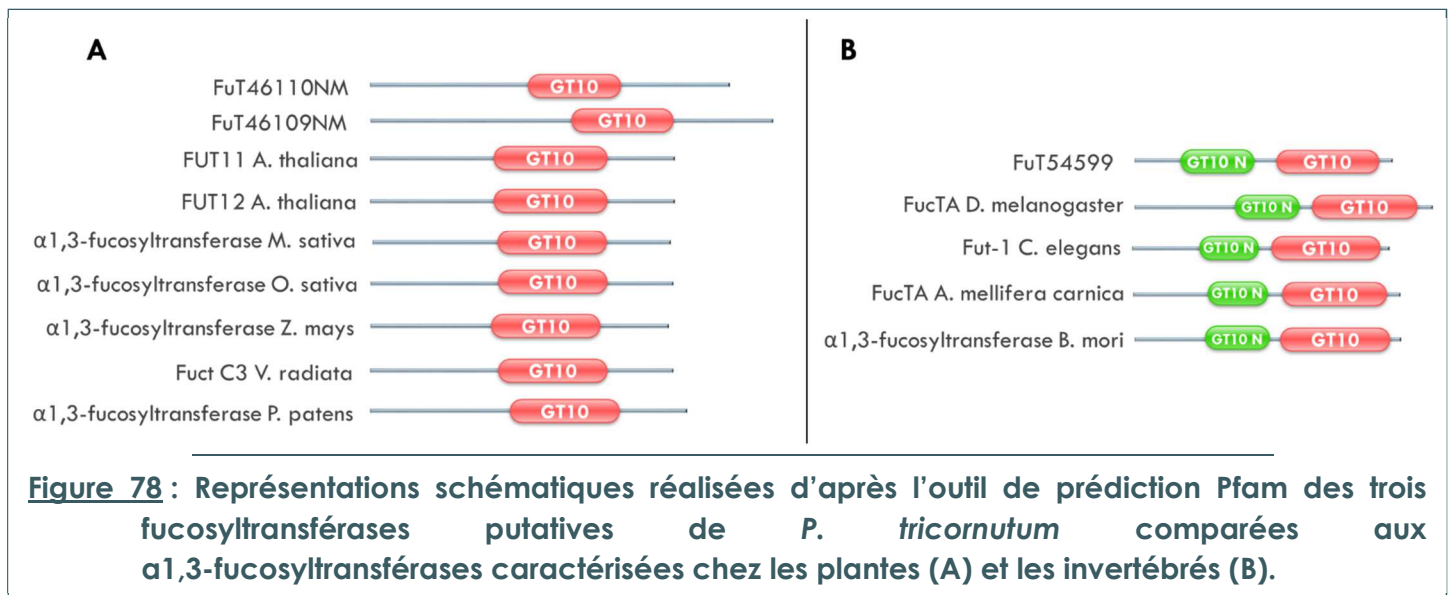
B : Prédiction d'une protéine possédant un domaine transmembranaire situé des acides aminés 15 à 34

C : Prédiction d'une protéine possédant deux domaines transmembranaires situés des acides aminés 17 à 34 et 663 à 668

6.2 Comparaison des protéines putatives FuT46109NM, FuT46110NM et FuT54599

Suite à une analyse réalisée avec le programme Pfam, les trois séquences protéiques putatives FuT46109NM, FuT46110NM et FuT54599 possèderaient des domaines Pfam GT10 à leur extrémité *C*-terminale, comme chez les α 1,3-fucosyltransférases caractérisées chez les

plantes (**figure 78-A**; Both et al., 2011). Cependant, la FuT54599 posséderait un second domaine Pfam GT10 N situé en *N*-terminal du précédent domaine, ce qui semble la rapprocher plutôt des α 1,3-fucosyltransférases caractérisées chez les invertébrés (**figure 78-B**).



En réalisant un alignement des séquences protéiques correspondantes au domaine catalytique GT10 des α 1,3-fucosyltransférases du core caractérisées chez les plantes et les invertébrés avec celles des fucosyltransférases putatives FuT46109NM, FuT46110NM et FuT54599 de *P. tricornutum*, on peut observer la préservation de motifs communs aux fucosyltransférases de la famille GT10. On peut ainsi retrouver les domaines conservés appelés « 1st cluster », « motif α 1,3-FuT » et « CXXC » (Both et al., 2011). Ces deux premiers motifs sont notamment impliqués dans la liaison au substrat donneur GDP-fucose. Par extrapolation des travaux présentés par (Both et al., 2011; Sun et al., 2007) utilisant des techniques de mutagenèse dirigée et de cristallographie, les acides aminés présentés en rouge sur la **figure 79** semblent déterminants pour la liaison au substrat accepteur, et sont effectivement bien conservés sur les différentes séquences de fucosyltransférases. Les rares acides aminés divergents parmi les séquences ont des propriétés physico-chimiques proches des acides aminés consensus. En ce qui concerne le motif CXXC, ce dernier est parfaitement conservé pour toutes les séquences. En effet, il a été démontré par mutagenèse dirigée que ce motif est nécessaire à l'activité enzymatique en permettant à la protéine d'adopter sa conformation fonctionnelle par la formation d'une boucle sur la protéine via la formation de ponts disulfures entre des cystéines de l'extrémité *N*-terminale et des cystéines du motif CXXC situé en *C*-terminal (Holmes et al., 2000).

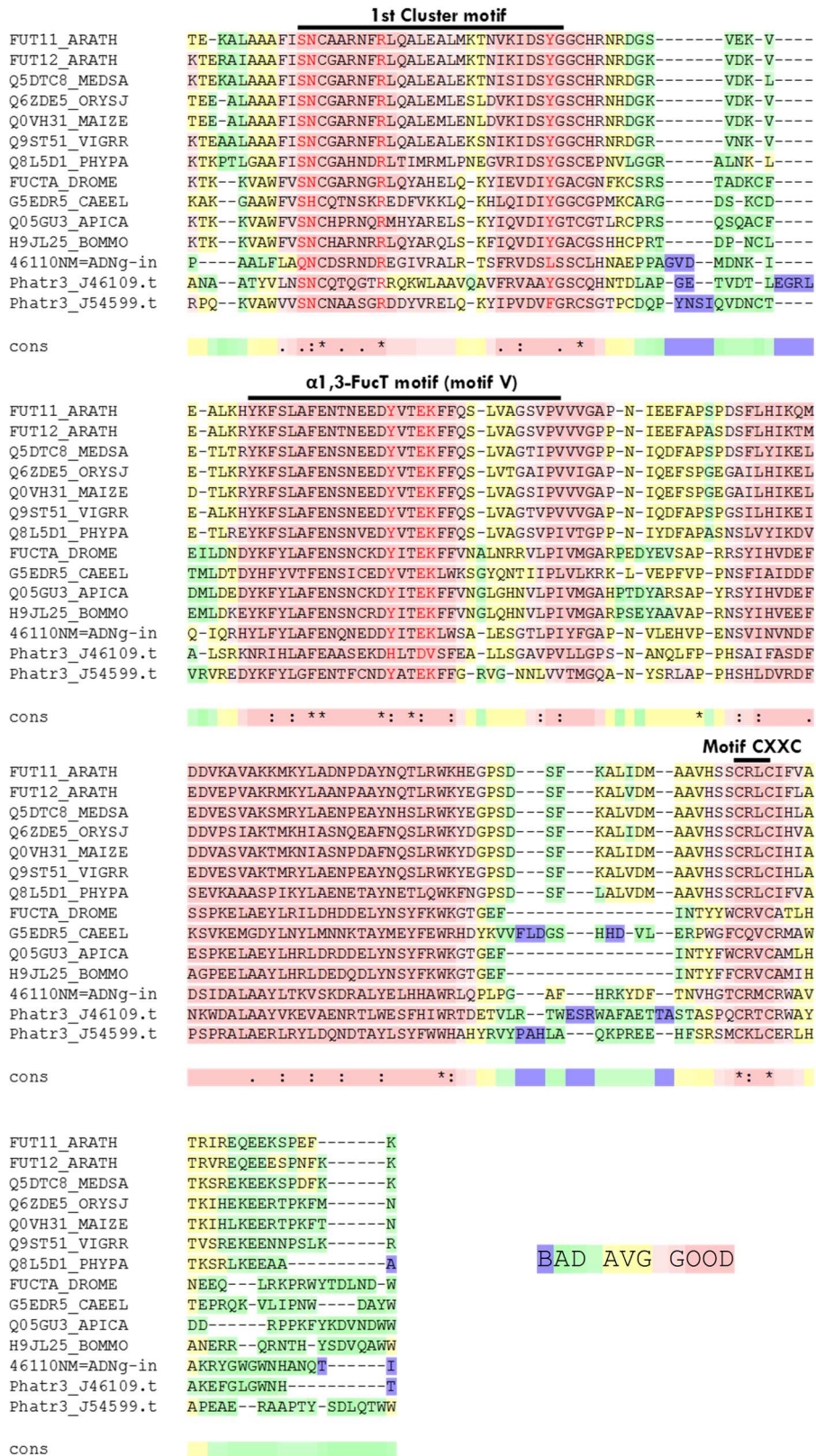


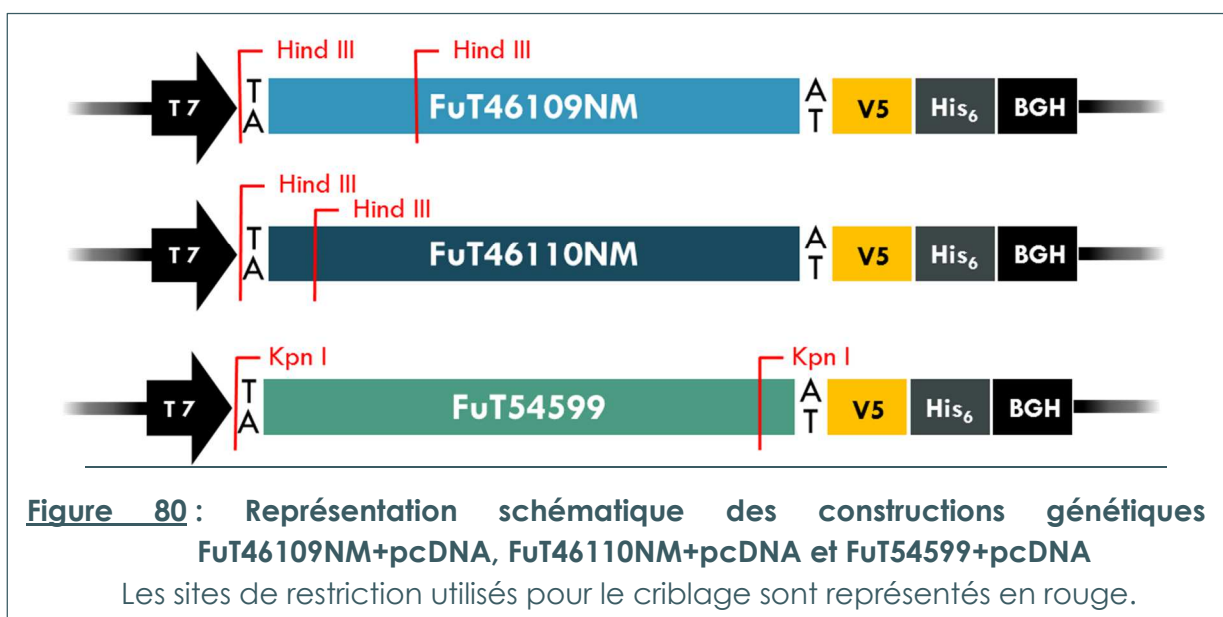
Figure 79 : Aligement des séquences protéiques correspondant aux domaines Pfam GT10 d'α1,3-fucosyltransférases caractérisées, ainsi que des séquences Fut46109NM, Fut46110NM et Fut54599 réalisé avec T-COFFEE v11 montrant la conservation des domaines « 1st cluster », « motif α1,3-FucT » et « CXXC ».

FUT 11_ARATH = FUT11 *A. thaliana* (Code Uniprot : Q9LJK1) ; FUT 12_ARATH = FUT12 *A. thaliana* (Code Uniprot : Q9FX97) ; Q5DTC8_MEDSA = α1,3-FuT *M. sativa* (Code Uniprot : Q5DTC8) ; Q6ZDE5_ORYSJ = α1,3-FuT *O. sativa subsp. japonica* (Code Uniprot : Q6ZDE5) ; Q0VH31_MAIZE = α1,3-FuT *Z. mays* (Code Uniprot : Q0VH31) ; Q9ST51_VIGRR = Fut c3 *V. radiata var. radiata* (Code Uniprot : Q9ST51) ; Q8L5D1_PHYPA = α1,3-FuT *P. patens subsp. Patens* (Code Uniprot : Q8L5D1) ; FUCTA_DROME = fuTA *D. melanogaster* (Code Uniprot : Q9VUL9) ; G5EDR5_CAEEEL = fut-1 *C. elegans* (Code Uniprot : G5EDR5) ; Q05GU3_APICA = fuTA *A. mellifera camic* (Code Uniprot : Q05GU3) ; H9JL25_BOMMO = α1,3-FuT *B. mori* (Code Uniprot : H9JL25) ; 46110NM = prédiction α1,3-FuT 46110NM *P. tricornutum* ; Phatr3_J46109 = prédiction α1,3-FuT 46109NM *P. tricornutum* ; Phatr3_J54599 = prédiction α1,3-FuT 54599 *P. tricornutum*. Les lettres rouges représentent les acides aminés impliqués dans la liaison au GDP-Fucose.

6.3 Clonage des gènes putatifs codant potentiellement pour des α 1,3-fucosyltransférases

6.3.1 Clonage des gènes dans le vecteur de mammifère pcDNA3.1/V5

Dans le but de réaliser la caractérisation fonctionnelle des trois fucosyltransférases putatives de *P. tricornutum* par complémentation chez CHO, les séquences codantes pour les protéines 46109NM, 46110NM et 54599 ont été clonées dans un vecteur d'expression spécifique des mammifères (**figure 80**). Le vecteur pcDNA 3.1/V5-His-TOPO a été choisi pour réaliser cette étude car il permet d'ajouter à l'extrémité C-terminale de la protéine recombinante une étiquette V5 et une séquence hexa-histidine, permettant ainsi de marquer, localiser et/ou purifier la protéine exprimée. Afin de réaliser le clonage des trois gènes putatifs, leur séquence génomique a été amplifiée par PCR en utilisant une ADN polymérase de haute-fidélité permettant d'ajouter des extrémités adénosine en 3' de chaque amplicon. Ainsi synthétisés, les inserts peuvent s'hybrider avec les extrémités thymidine dépassantes du vecteur linéaire pcDNA 3.1/V5-His-TOPO. Cette hybridation par T/A cloning va donc laisser à l'amplicon la possibilité de s'insérer dans les deux sens avec la même probabilité. Afin de sélectionner les vecteurs recombinants possédant l'insert correctement orienté, une digestion des constructions plasmidiques a été réalisée avec des enzymes de restriction. Les enzymes ont été sélectionnées afin de cliver les plasmides en deux fragments en utilisant un site de restriction située sur une extrémité de l'insert, ainsi qu'un second site de restriction situé sur le vecteur. Une fois l'orientation de l'insert vérifiée, les constructions ont été séquencées afin de s'assurer qu'aucune mutation n'était apparue au cours des étapes de clonage et d'amplification. Les plasmides recombinants ainsi contrôlés ont été utilisés pour la transfection des cellules de mammifères mutantes CHO Fut8^{-/-}.



6.3.2 Clonage des gènes dans le vecteur de diatomée pPha-NR

Afin de procéder à la localisation subcellulaire des trois α 1,3-fucosyltransférases putatives chez *P. tricornutum*, les séquences codantes pour les protéines FuT46109OM, FuT46109NM, FuT46110OM, FuT46110NM et FuT54599 ont été clonées dans le vecteur d'expression de diatomée pPha-NR. Ce vecteur possède un promoteur inductible au nitrate (**figure 81**). En plus de ces séquences, différents gènes témoins ont été clonés car il n'existe actuellement aucun marqueur commercial de l'appareil de Golgi testé chez les diatomées. Dans ce but, le gène codant pour la GnT I de *P. tricornutum*, attendue dans la zone compartimentale *median* Golgi (Reichardt et al., 2007), a été sélectionné comme témoin de localisation. Bien que la localisation subcellulaire de la GnT I chez *P. tricornutum* n'ait pas encore été caractérisée par microscopie, l'utilisation de cette protéine comme témoin de localisation de l'appareil de Golgi a été motivée par des résultats antérieurs. En effet, il a été montré que la complémentation de cellules CHO mutées sur la GnT I endogène par la GnT I de *P. tricornutum* permet de restaurer un profil *N*-glycannique de type complexe (Baiet et al., 2011). Or, comme l'adressage de la GnT I au niveau du saccule *median* de l'appareil de Golgi est nécessaire à la restauration de la *N*-glycosylation complexe, ce résultat suggère que la GnT I est correctement adressée à ce compartiment. Les autres témoins de localisation utilisés sont la β 1,4-galactosyltransferase humaine (EC 2.4.1.22) qui a été sélectionnée comme marqueur du *trans*-Golgi (Berger and Hesford, 1985), et une xylosyltransférase putative de *P. tricornutum* (Phatr3_J45496) qui est attendue dans l'appareil de Golgi. La fonctionnalité de cette dernière protéine n'étant pas encore caractérisée chez *P. tricornutum*, son utilisation comme témoin de localisation fait l'objet d'une extrapolation de résultats obtenus par microscopie électronique à transmission chez la plante (Chevalier et al., 2010; Pagny et al., 2003). Les analyses bio-informatiques par homologie de séquence indiquent que cette xylosyltransférase putative présente des similitudes avec les β 1,2-xylosyltransférases impliquées dans la maturation des *N*-glycannes chez les plantes. Cependant, de précédentes analyses biochimiques réalisées sur les *N*-glycannes du morphotype fusiforme de *P. tricornutum* par western-blot et spectrométrie de masse ont écarté tout ajout de β 1,2-xylose sur les *N*-glycannes des protéines endogènes de ce morphotype. Cette glycosyltransférase intervient ainsi probablement soit dans le processus de glycosylation chez un autre morphotype, soit dans un autre processus de synthèse polysaccharidique qui n'est pas encore déterminé. La construction génétique contenant cette xylosyltransférase a été réalisée et généreusement mise à notre disposition par l'équipe du Pr. Uwe Maier (Université de Marbourg, Allemagne) qui a préalablement observé par microscopie confocale

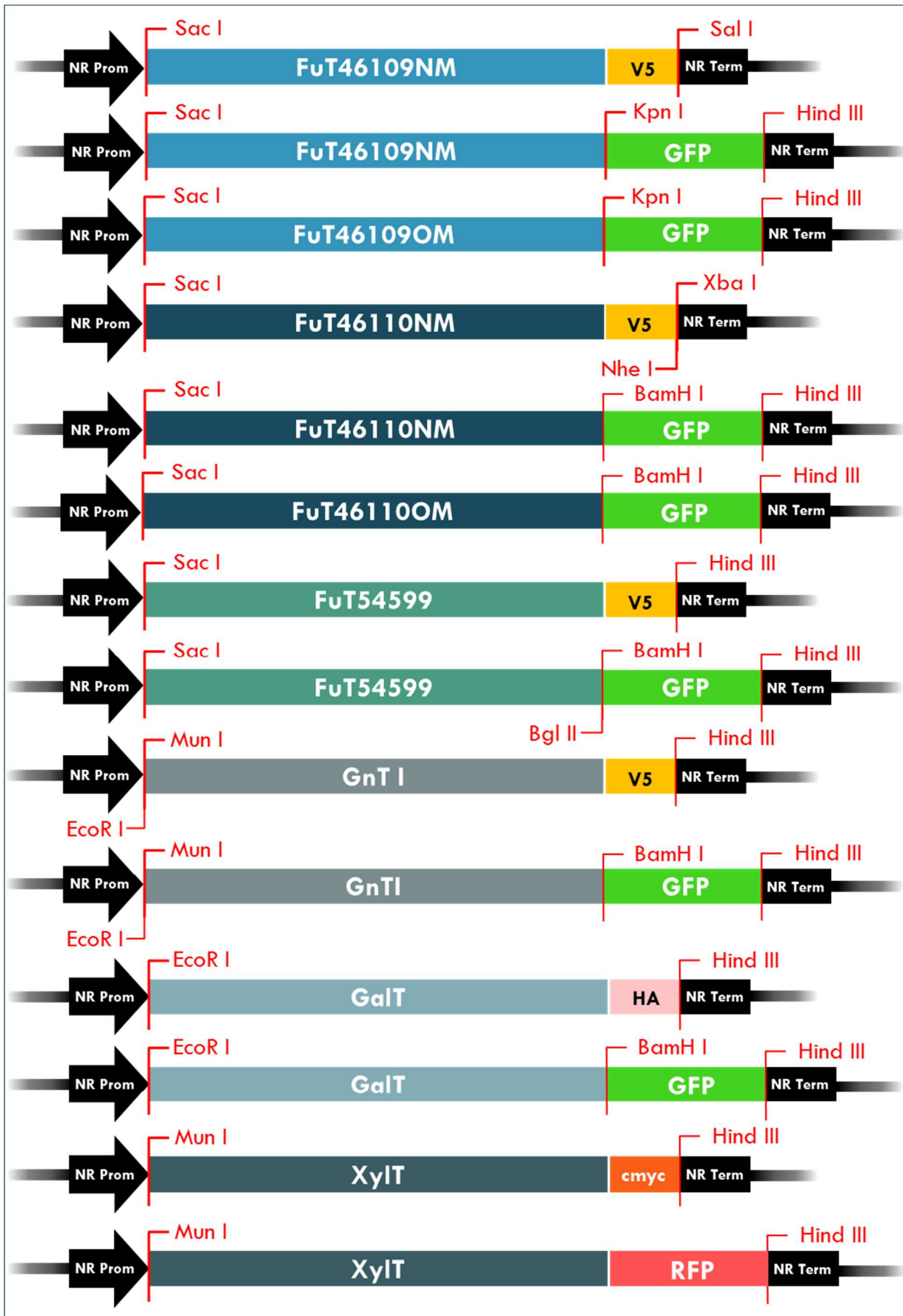


Figure 81 : Représentation schématique des constructions génétiques FuT46109NM-GFP et -V5, FuT46109OM-GFP, FuT46110NM-GFP et -V5, FuT46110OM-GFP, FuT54599-GFP et -V5, GnTI-GFP et -V5, GalT-GFP et -HA, ainsi que XylT-RFP et -cmyc dans le vecteur pPha-NR

Les sites de restriction utilisés pour le clonage sont présentés en rouge

la localisation de cette protéine comme étant située dans l'appareil de Golgi (communication personnelle).

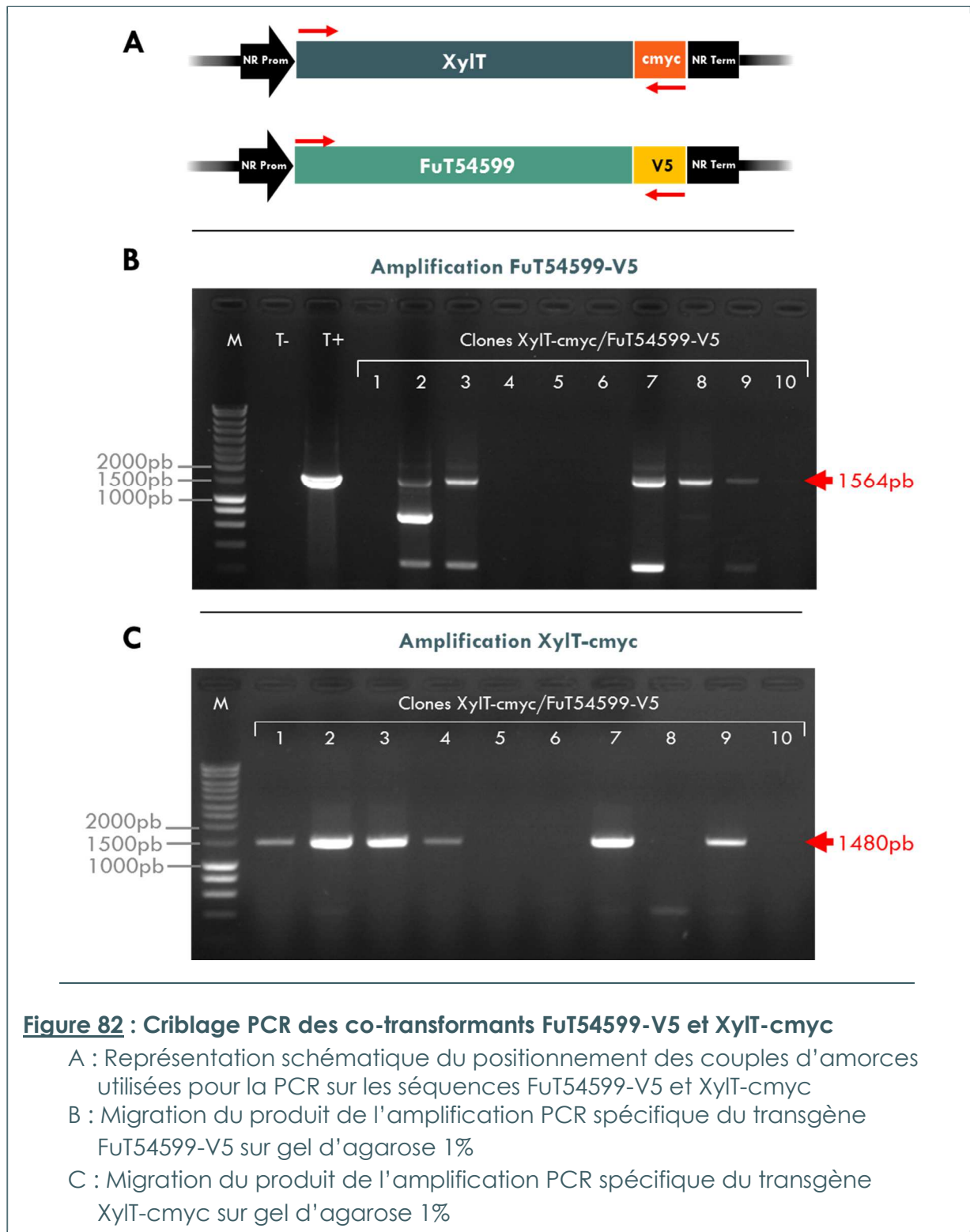
Pour réaliser l'étude des protéines codées par ces séquences, différentes fusions à des étiquettes ont été effectuées. Ainsi dans un premier temps, pour observer aisément la localisation subcellulaire de ces protéines d'intérêt par microscopie, une protéine fluorescente, GFP ou RFP, a été fusionnée à leur séquence protéique. Cependant, l'ajout de ces groupements fluorochromes de grande taille peut induire certains artefacts, comme des problèmes de localisation, ou des perturbations de l'activité enzymatique dues à la gêne stérique (Donohoe et al., 2013; Moore and Murphy, 2009). Ainsi, pour s'affranchir de ces éventuels problèmes, une 2^{ème} série de constructions possédant une étiquette peptidique plus courte, épitope V5, c-myc ou Hémagglutinine (HA), a été utilisée pour pouvoir si nécessaire caractériser biochimiquement l'activité enzymatique de ces protéines. Le choix de ces étiquettes peptidiques a été déterminé afin de pouvoir purifier éventuellement par immunoprécipitation les protéines dans le but de caractériser leur activité *in vitro*, si la surexpression de la protéine recombinante ne permet pas de modifier de manière significative le profil glycanique *in vivo* des diatomées transfectées.

Le clonage des différentes séquences d'intérêt dans le vecteur pPha-NR a été réalisé en deux étapes, comme détaillé dans la partie « matériel et méthodes » de ce manuscrit. En effet, le clonage d'une séquence dans ce vecteur de diatomée possédant un promoteur inductible au nitrate est réalisé par hybridation d'extrémités cohésives. Pour cela un premier clonage a été effectué dans le vecteur aux extrémités franches pJET1.2/Blunt afin de pouvoir par la suite digérer efficacement les extrémités de l'insert. Les constructions positives ont été digérées par les enzymes de restriction présentant des sites de coupure aux extrémités de l'insert afin d'extraire la séquence d'intérêt, tout en générant à ses bordures des extrémités cohésives. Les fragments générés ont ensuite été clonés dans le vecteur pPha-NR préalablement digéré avec des enzymes de restriction offrant des extrémités cohésives compatibles pour la ligation de l'insert. Les constructions étiquetées avec une protéine fluorescente ont été réalisées par ligation de trois fragments cohésifs en même temps (vecteur/séquence d'intérêt/étiquette). Les autres étiquettes ont été fusionnées à la séquence d'intérêt par PCR. Les cartes détaillées des différents plasmides sont présentées en annexe 1 de ce manuscrit. Les constructions génétiques ont ensuite été criblées par digestion des plasmides avec des enzymes de restriction, puis les vecteurs recombinants positifs ont été séquencés. Un clone a ainsi été sélectionné afin de procéder à la transfection des cellules de *P. tricornutum* par biolistique.

6.3.3 Analyse des transformants obtenus

6.3.3.1 Contrôle par PCR de la transgénèse des lignées transfectées

Suite à la transfection des microalgues avec les constructions génétiques possédant une étiquette peptidique (HA, V5, ou c-myc), les différentes colonies ayant poussé sur milieu sélectif ont été criblées par PCR. Cette technique permet de vérifier dans un premier temps si le transgène est bien intégré au génome de la microalgue avant de confirmer par la suite l'expression de la protéine recombinante par western-blot en utilisant des anticorps spécifiquement dirigés contre les étiquettes. Pour cette expérience de PCR, les amplifications ont été effectuées en utilisant des amorces spécifiques de chaque transgène. Un témoin positif a été réalisé avec le plasmide utilisé pour la transfection, et un témoin négatif en remplaçant l'ADN matrice par de l'eau stérile. Les résultats obtenus lors de ces amplifications nous ont permis de mettre en évidence la présence de 15 clones positifs pour les constructions pPha-NR+FuT46109-V5, 8 clones pour pPha-NR+FuT54599-V5, 8 clones pour pPha-NR+GnTI-V5 (**figure 83A**), 17 clones pour pPha-NR+GalT-HA et 6 clones pour pPha-NR+XylT-cmyc. Malheureusement, sur 16 colonies criblées, aucune colonie positive n'a pu être obtenue pour la construction pPha-NR+FuT46110NM-V5. L'élaboration de co-transformants a été initiée avec les plasmides pPha-NR+FuT54599-V5 / pPha-NR+XylT-cmyc, et pPha-NR+GnTI-V5 / pPha-NR+XylT-cmyc. Le criblage par PCR des doubles transfactions a permis de mettre en évidence la présence de co-transformants pPha-NR+FuT54599-V5 / pPha-NR+XylT-cmyc (**figure 82**), cependant pour le couple pPha-NR+GnTI-V5 / pPha-NR+XylT-cmyc, seule la transfection du gène GnT I-V5 a pu être confirmée. Les lignées ayant bien intégré le transgène ont ensuite été criblées par western-blot afin de vérifier l'expression de la protéine recombinante correspondante.



6.3.3.2 Contrôle de la transgénèse des lignées transfectées par western-blot

Un western-blot a été réalisé sur un extrait protéique provenant de la fraction microsomale des lignées cellulaires de *P. tricornutum* transfectées avec les constructions contenant une étiquette peptidique. L'expression de la protéine recombinante a été mise en évidence par immunodétection sur empreinte avec un anticorps spécifiquement dirigé contre

l'étiquette peptidique fusionnée à la protéine cible (V5 ou c-myc). L'anticorps secondaire utilisé est couplé à la peroxydase de raifort (HRP) et révélé par électrochimiluminescence. Des cellules non transfectées ont été utilisées comme témoin négatif, et un extrait protéique membranaire de CHO exprimant la GnT I de *P. tricornutum* étiquetée V5 a été utilisé comme témoin positif pour les protéines possédant cette étiquette. Les résultats de ces western-blots démontrent la présence de lignées de *P. tricornutum* exprimant les protéines recombinantes GnT I-V5 (**figure 83B**), et les protéines recombinantes XylT-cmyc / FuT54599-V5 de manière co-exprimée (**figure 84**). Au contraire, les co-transformants XylT-cmyc / GnT I-V5 n'expriment que la protéine GnT I-V5, qui est fortement exprimée par le clone 5 (**figure 85**). Aucune protéine recombinante n'a pu être observée pour les lignées FuT54599-V5 et XylT-cmyc. Les transformants 46109NM-V5 et GalT-HA ayant été acquis tardivement lors de ce travail, l'expression des protéines recombinantes chez ces lignées n'a pas pu encore être testée.

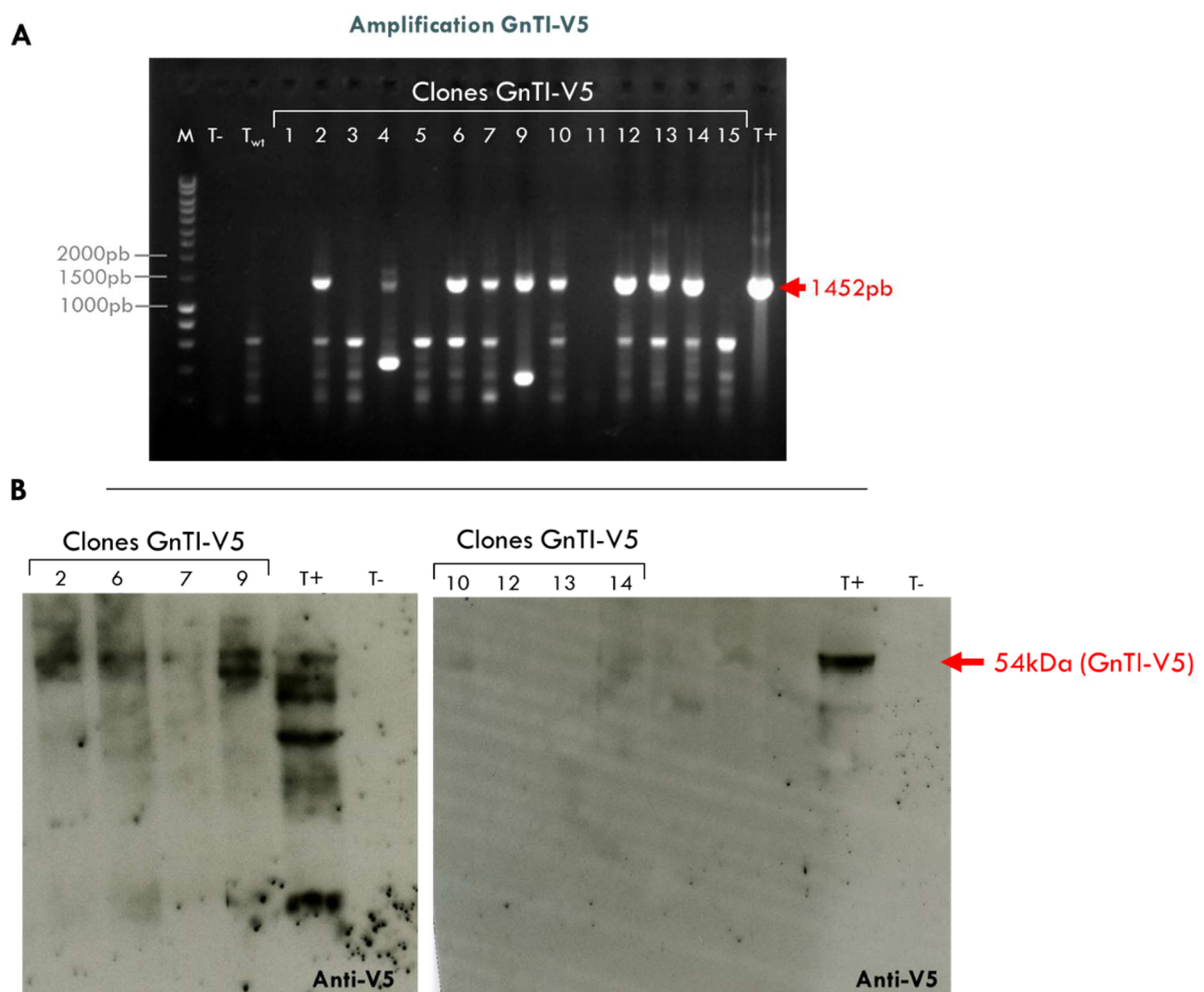


Figure 83 : Criblage des transformants GnT I-V5 par PCR et western-blot

A : Migration du produit de l'amplification PCR spécifique du transgène GnT I-V5 sur gel d'agarose 1%

B : Détection spécifique par western-blot de l'expression de la protéine GnT I-V5 avec l'anticorps anti-V5

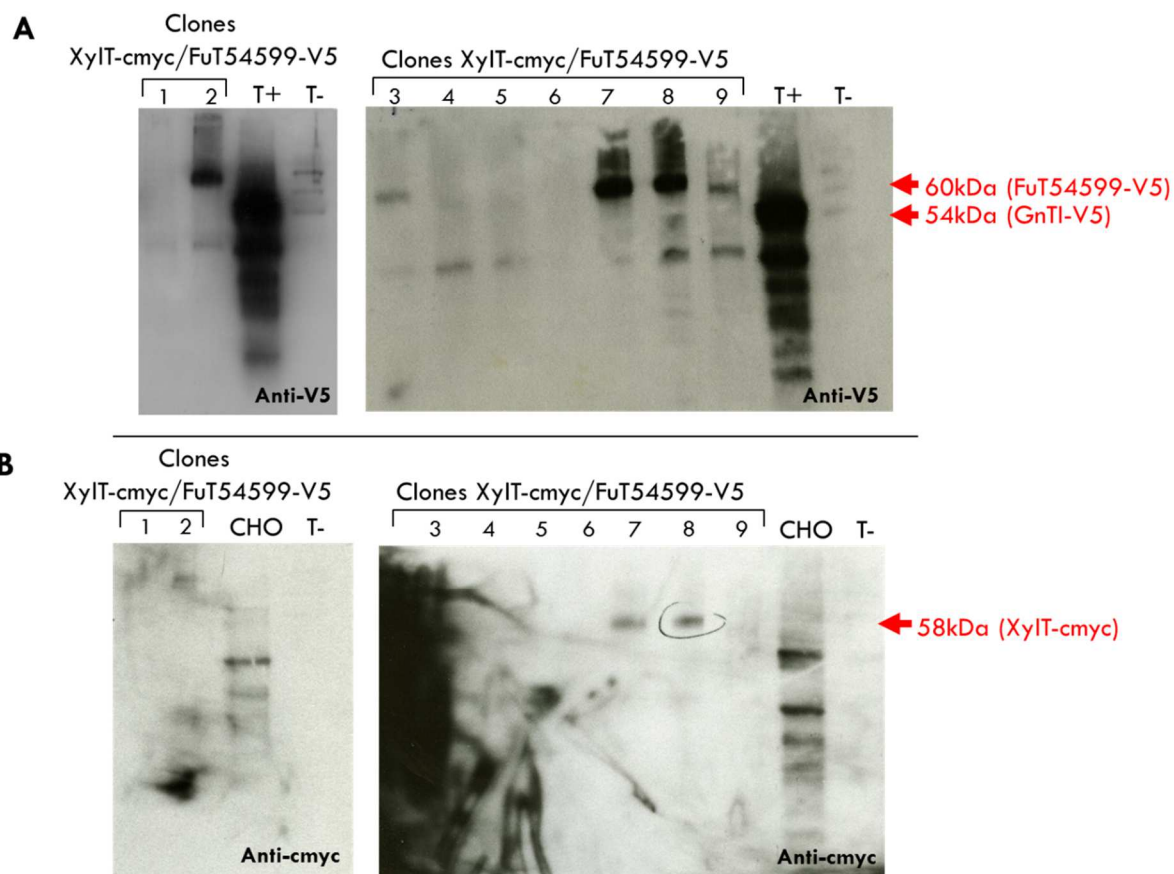


Figure 84 : Criblage par western-blot des co-transformants FuT54599-V5 et XylIT-cmyc

A : détection spécifique de l'expression de la protéine FuT54599-V5 avec l'anticorps anti-V5

B : détection spécifique de l'expression de la protéine XylIT-cmyc avec l'anticorps anti-cmyc

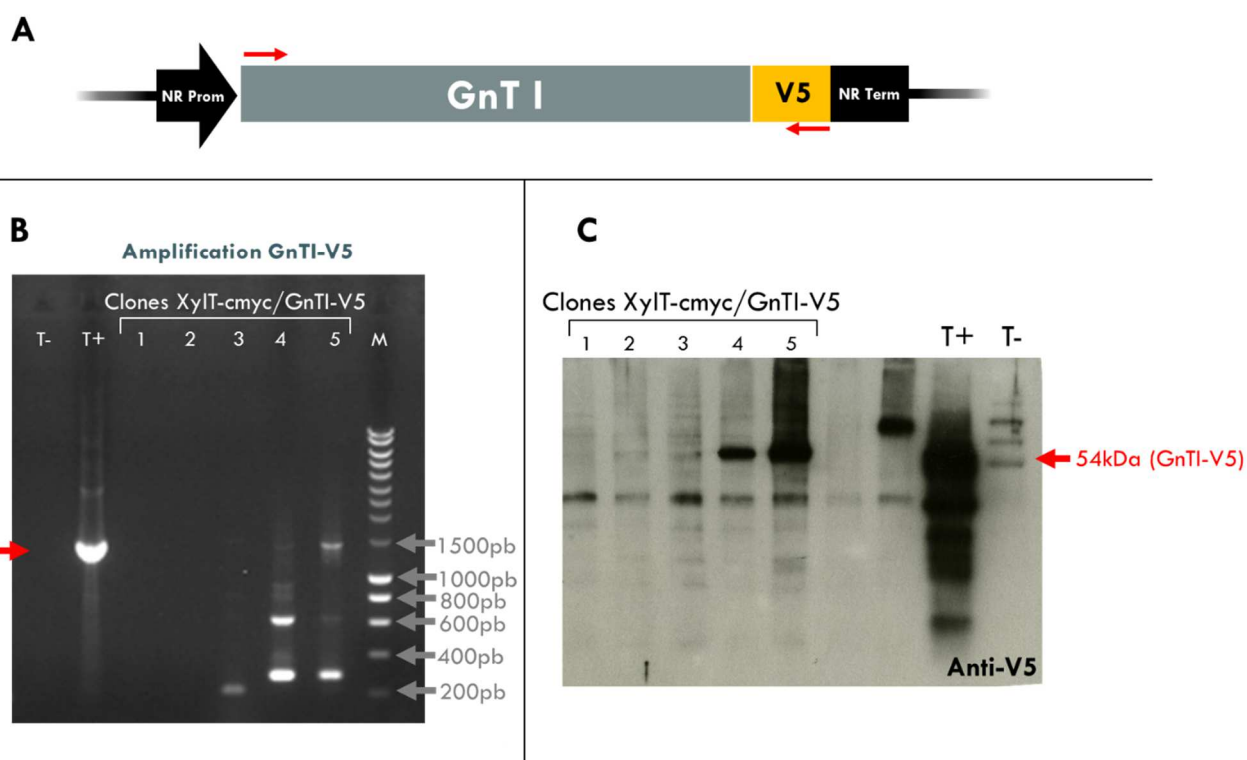


Figure 85 : Criblage des co-transformants GnT I-V5 et XylIT-cmyc par PCR et western-blot

A : Représentation schématique du positionnement des amorces utilisées pour la PCR sur la séquence GnT I-V5

B : Migration du produit de l'amplification PCR spécifique du transgène GnT I-V5 sur gel d'agarose 1%

C : détection spécifique par western-blot de l'expression de la protéine GnT I-V5 avec l'anticorps anti-V5

6.3.3.3 Contrôle de la transgénèse des lignées transfectées par microscopie confocale

Les lignées de *P. tricornutum* transfectées avec des transgènes fusionnés aux gènes codant pour une protéine fluorescente ont été directement criblées par observation au microscope confocal. Ces observations ont pu mettre en évidence l'expression d'une protéine fluorescente localisée de manière ponctiforme, voire en forme de croissant selon l'orientation de la cellule. Cette observation a pu être faite pour toutes les lignées témoin transfectées avec les gènes GnT I-GFP, GalT-GFP et XylT-RFP (**figure 86**), ce qui laisse penser que ces protéines sont correctement adressées à l'appareil de Golgi.

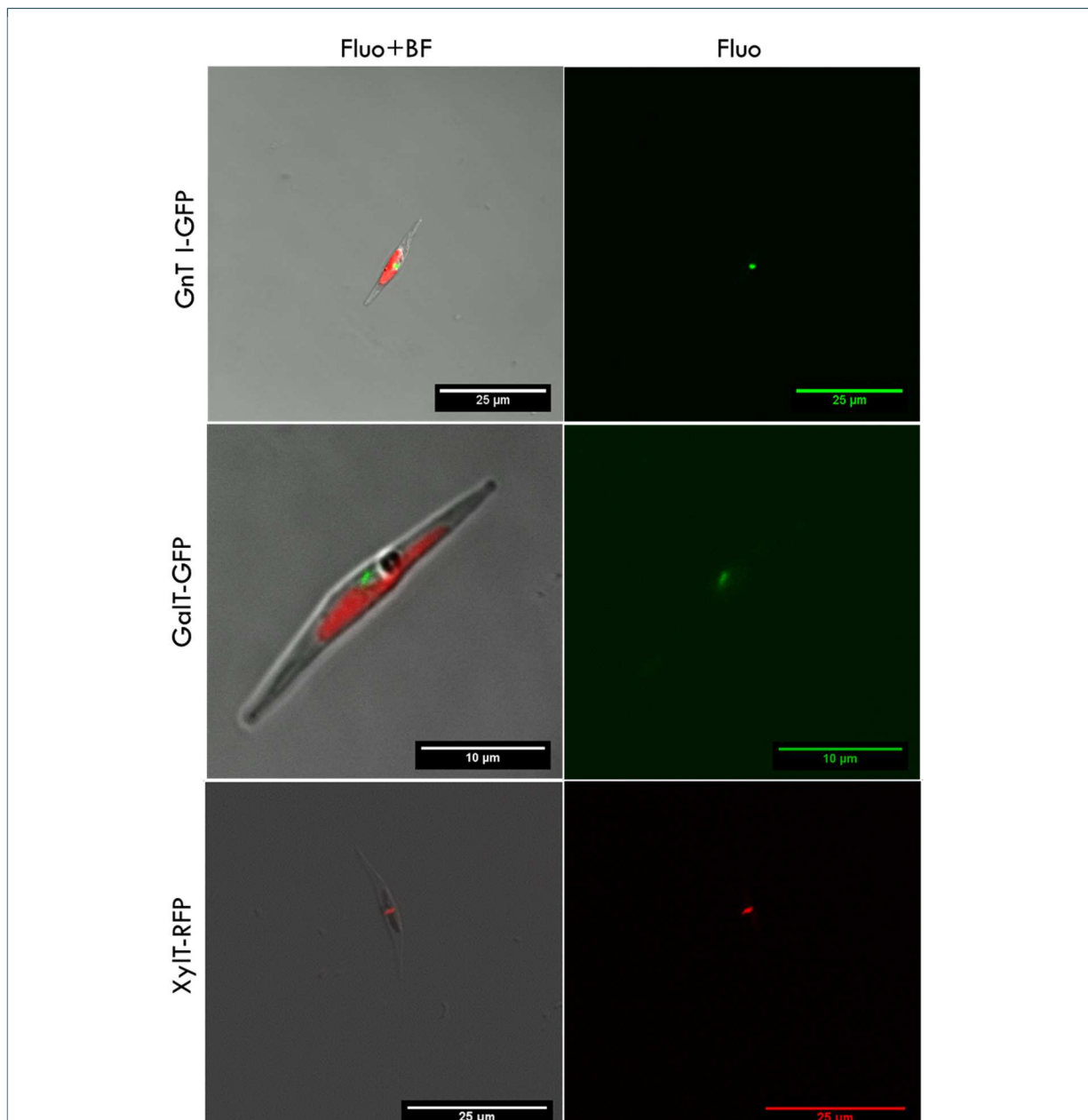
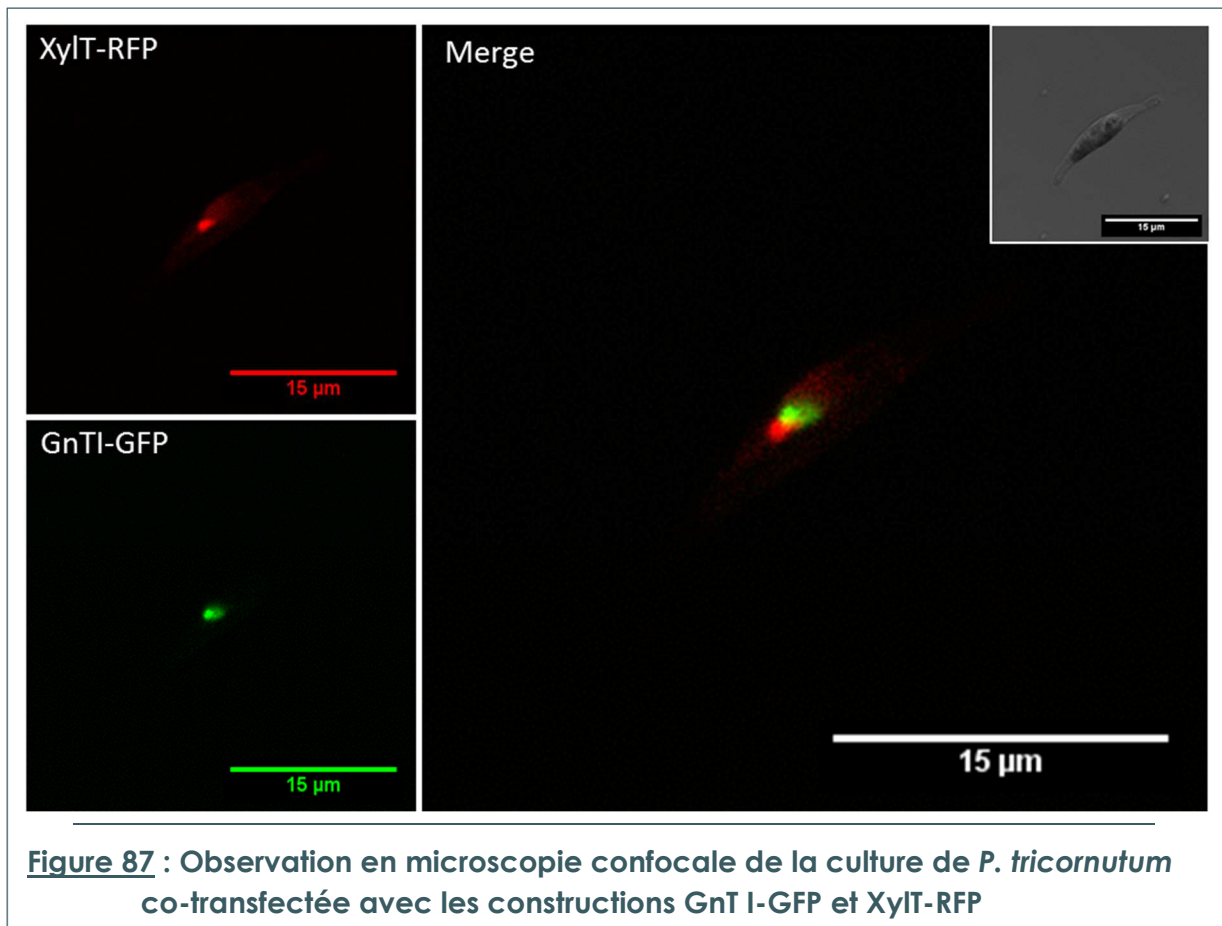


Figure 86 : Observation en microscopie confocale des cultures de *P. tricornutum* transfectées avec les constructions GnT I-GFP, GalT-GFP et XylT-RFP
Le signal coloré en rouge pour les cellules GnT I-GFP et GalT-GFP correspond à l'autofluorescence plastidiale.

Un co-transformant combinant l'expression des transgènes GnT I-GFP et XylT-RFP a aussi été analysé par microscopie confocale. Cette observation montre que les protéines recombinantes exprimées sont localisées de façon très proche mais pas complètement co-localisées puisque les signaux de fluorescence visualisés sont juxtaposés. Ainsi, ce résultat suggère que la GnT I, qui est attendue plutôt dans le compartiment *median* de l'appareil de Golgi, n'est pas localisée dans la même citerne golgienne que la XylT. Etant dans l'incapacité d'orienter l'appareil de Golgi dans cette approche, nous ne pouvons qu'émettre l'hypothèse que la XylT se situe dans les citernes *cis* ou *trans* golgiennes (**figure 87**).



Cette même observation de fluorescence ponctiforme a pu être faite pour les lignées transfectées avec les transgènes 54599-GFP, 46109NM-GFP et 46110NM-GFP, ce qui semble indiquer que les protéines recombinantes correspondantes sont bien adressées à l'appareil de Golgi. *A contrario*, la fluorescence émise par l'expression de la protéine de fusion FuT46109OM est beaucoup moins localisée et se trouve répartie dans toute la cellule (**figure 88**). Cette différence de localisation entre les protéines FuT46109OM et FuT46109NM montre que la portion protéique ajoutée en *N*-terminal du nouveau modèle est cruciale pour la localisation de la protéine contenant probablement le domaine transmembranaire indispensable pour l'adressage des glycosyltransférases.

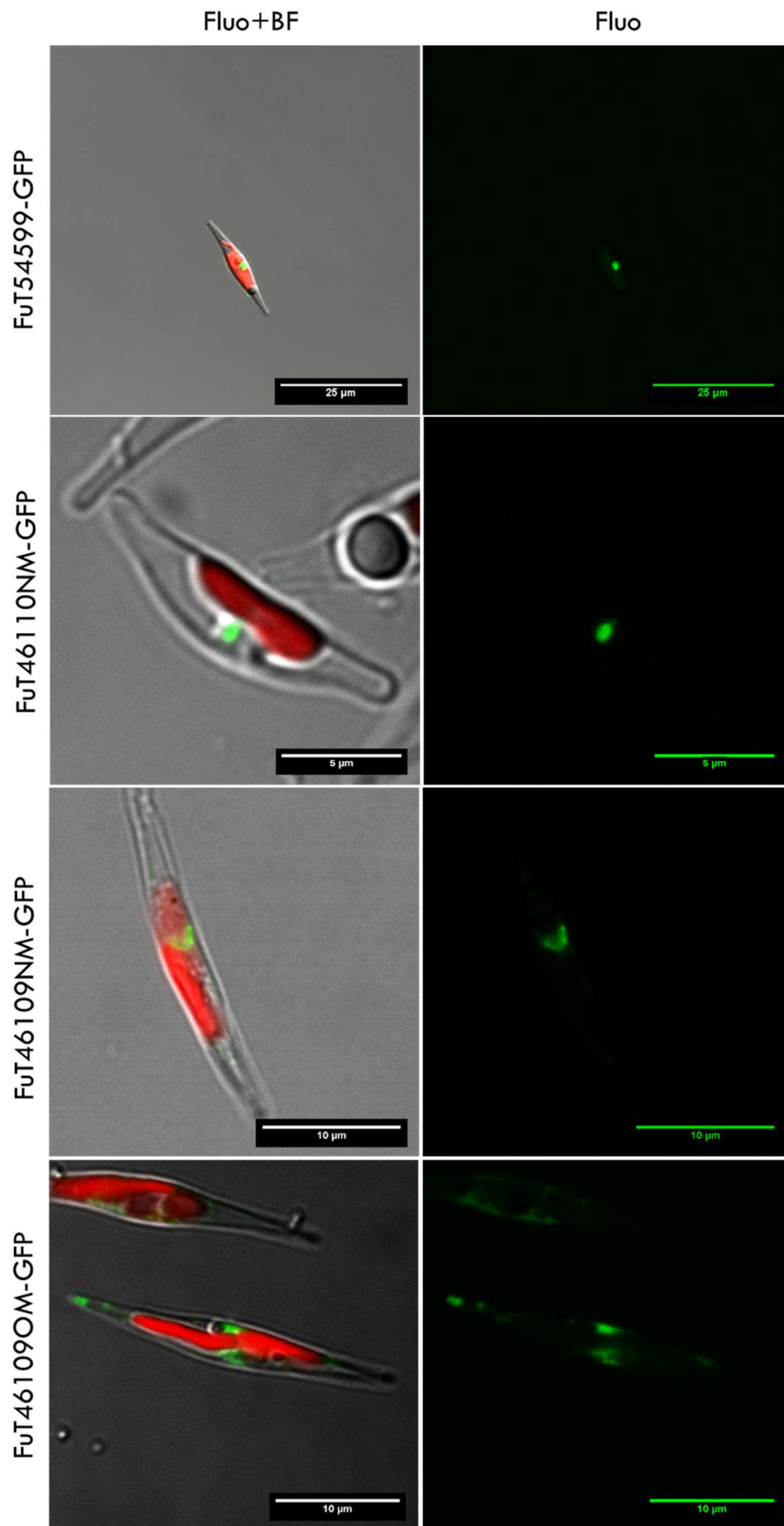
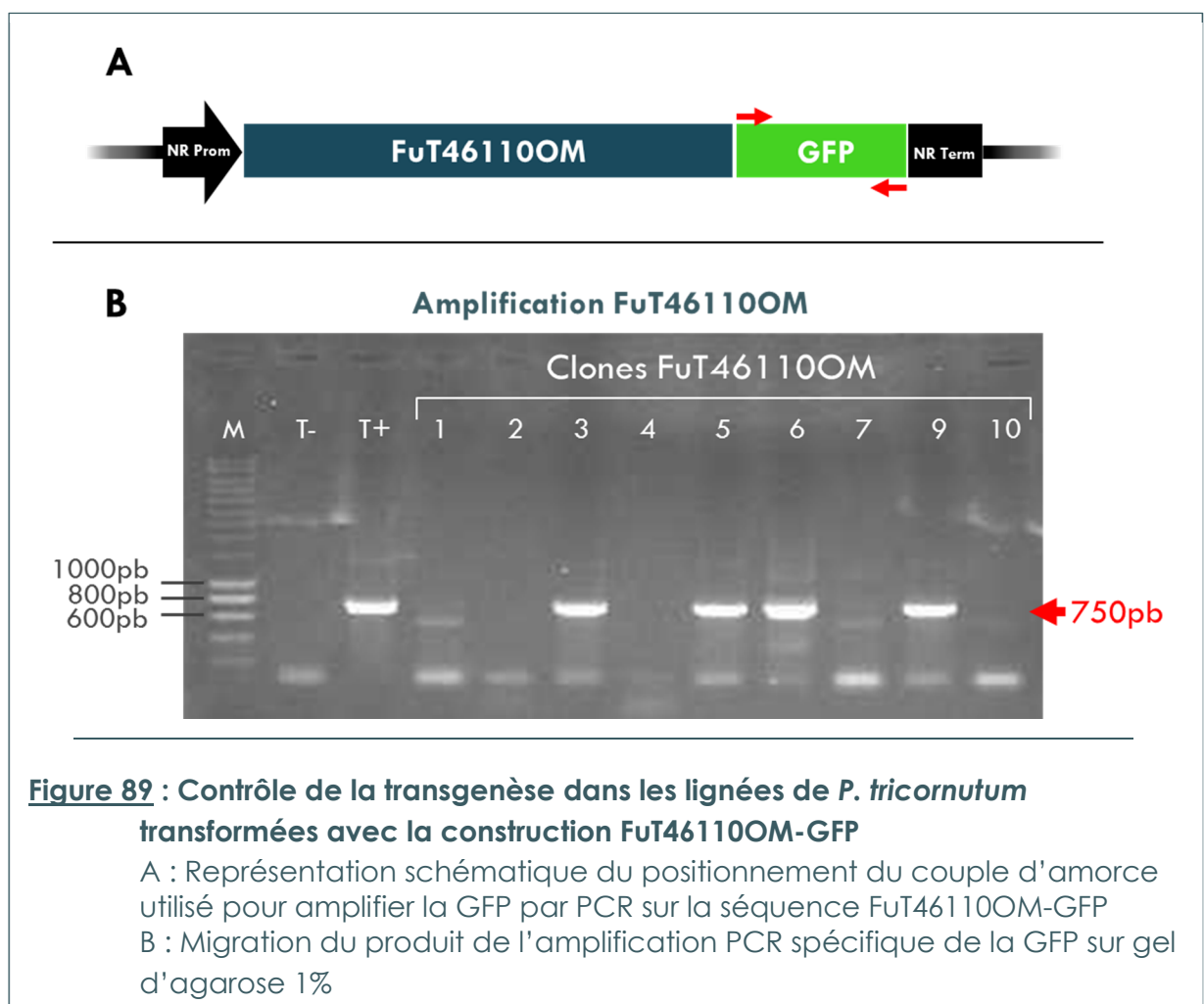


Figure 88 : Observation en microscopie confocale des cultures de *P. tricornutum* transformées avec les constructions FuT54599-GFP, FuT46110NM-GFP, FuT46109NM-GFP et FuT46109OM-GFP

Ainsi, bien que les prédictions bio-informatiques réalisées sur cette nouvelle portion ne soient pas unanimes sur la présence d'un tel domaine dans la séquence FuT46109NM, il semblerait qu'*in vivo* la FuT46109NM possède un domaine transmembranaire comme attendu pour une glycosyltransférase fonctionnelle. Cette observation valide la nouvelle annotation proposée du gène 46109.

En ce qui concerne les transformants 46110OM-GFP, aucun signal de fluorescence n'a pu être observé malgré l'intégration du transgène dans plusieurs lignées (**figure 89**). Ce résultat peut être expliqué par l'expression d'une protéine tronquée dépourvue de GFP, causé par l'absence d'épissage du premier intron prédit sur le transcrit 46110OM-GFP. En effet, la séquence nucléique présente dans le premier intron de l'annotation PHATR DRAFT_46110 comporte un codon stop. Ceci implique qu'en l'absence d'épissage de cette séquence, le transcrit produit codera pour une protéine tronquée sans étiquette fluorescente (**figure 90**). Ce résultat confirme les précédentes analyses que nous avons réalisées par PCR sur l'épissage de ce transcrit, et semble ainsi valider la nouvelle annotation proposée pour le gène 46110.



```

MIRIPYAVSYLACSAPFIGNHSQWEGDPVAAISSAVIYHLKPTKEEGRIRACQRPVG
NEVRHPIGKVSSIDVKTPPTFLDRKHTLTSINHARVFATAPTDALPLYVEPFSCRALV
HRYPLYAPAPSQSTRSRCAYSVVAAHQQCRTPQHSEPAPEFIRNV-IPHARYQFT
IGARA-SAHVWQYRWQACPPY-FE-SYGTVRSVCICDQGFSGSITTNIVELA
YVTIMPLERNGSDDTQCL-PSRIGTKPKTKDCSLLQSIFPGTHYLRPEYS
AKHTFTVG-AVRGAVSLVCGFTRSHGKRATTDHGQLRSPAPQISDGVSA
V-RALLSRG-I-ACFSTDH-RNPLAFCLFHGRTTLFFVAFLGSLRIPD-PLLR
HHIFPIRCSLAVFLVERVQYSKANARF-PSHSGGSFSGPKL-FT-RSGGYC
AGFADVLSC-FLEQLSAQRRATSRCRYGQQNSNSASLSVLFVREPKRR-
LHHRKAVERFRIRHASNLLRST-RS-ARSGKLGH-RKRF-QYRRVGCLLDK
GFQRSGFVRIAPCVAIATTTGSFPPSEIRFYECPRNVPNVSVGGG-TLRLGL
EPRESNYCGRRRPPNHMRRRTWLVATAATGILDFENE-CGYEIAAAKNSL
SAGIRKGRQLQRWQRHSNDCCWREGGCI-TINCGTRWSARHIPENENK
C-ERKDIKIRFAVAVSDATPTW-CYHLCLSNH-RGSHEIARQYKSHYVFD
VAARECCVGIIC-WGRP-SDCSGAHITVATPKSDCGGRRHLSSRSRGCPGK
LFWQNDEG-FLESCGNLRV

```

Figure 90 : Traduction prédite à partir de la séquence nucléotidique non épissée de l'ancienne version du gène 46110 (FuT46110OM)

La séquence écrite en rouge représente la séquence pouvant être traduite en protéine. Les tirets correspondent aux codons stop.

Un co-transformant a aussi été réalisé avec les transgènes 54599-GFP / XylIT-RFP. L'observation de ce co-transformant montre une nouvelle fois que la fluorescence émise par les deux protéines recombinantes est localisée de manière ponctiforme, et bien que les signaux de fluorescence soient juxtaposés, seule une faible co-localisation des deux protéines est présente (**figure 91**). Ainsi, comme observé précédemment pour le co-transformant GnT I / XylIT, les protéines recombinantes exprimées ne sont donc pas, ou peu co-localisées. Les précédents résultats de co-localisation ayant montré que la XylIT était probablement localisée dans les compartiments *cis* ou *trans*, et étant donné la proximité des signaux de fluorescence, on peut émettre l'hypothèse que la FuT54599 soit localisée dans le saccule *median* comme la GnT I. Cependant, il n'existe actuellement aucun marqueur des saccules golgiennes de *P. tricornutum* permettant de déterminer avec certitude la localisation de ces protéines recombinantes par microscopie confocale. Il a donc été entrepris d'étudier plus précisément la localisation de ces protéines recombinantes par microscopie électronique à transmission (MET).

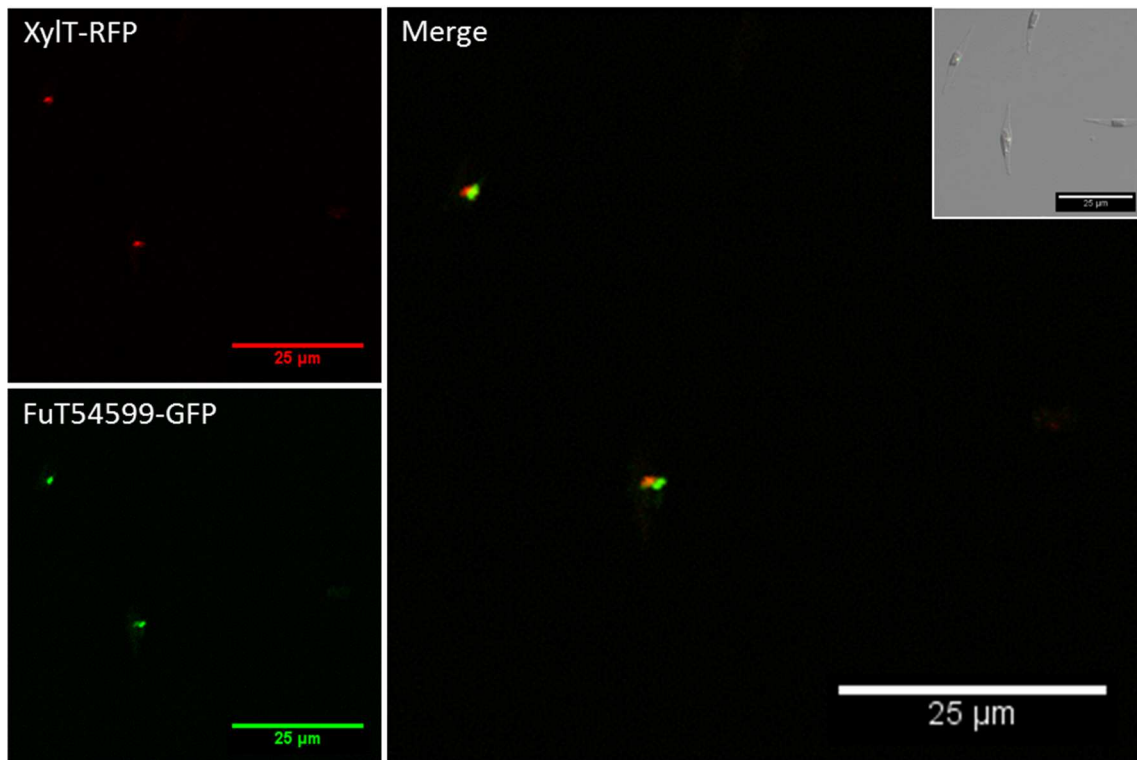


Figure 91 : Observation en microscopie confocale de la culture de *P. tricornutum* co-transfectée avec les constructions FuT54599-GFP et XylT-RFP

6.4 Localisation subcellulaire des enzymes GnT I et FuT54599 par microscopie électronique à transmission.

Dans le but de réaliser une analyse plus fine de la localisation subcellulaire de ces différentes protéines recombinantes, une observation des microalgues transformées a été entreprise par microscopie électronique à transmission (MET). Ce travail a nécessité une optimisation de la préparation des échantillons pour une bonne préservation de l'organisation cellulaire des diatomées. Les résultats obtenus par microscopie confocale sur les différents transformants étiquetés GFP étant plutôt encourageants, nous avons choisi d'utiliser ces lignées cellulaires pour réaliser les observations au microscope électronique. La détection des protéines recombinantes a été réalisée par immunomarquage en utilisant un anticorps primaire dirigé contre la GFP, puis un anticorps secondaire couplé à une bille d'or. Malheureusement, contrairement au témoin positif, aucune protéine recombinante produite par nos transformants n'a pu être observée en utilisant cette stratégie. Cette absence de détection est due probablement à une expression très faible des glycosyltransférases dans ces clones empêchant la détection par cette approche. En effet, la microscopie confocale est une technique très sensible qui détecte la protéine recombinante via la fluorescence émise par la GFP ou RFP. De plus, la finesse des

coupes en MET implique qu'une protéine faiblement exprimée aura statistiquement moins de chance d'être localisée sur la portion observée.

Une nouvelle stratégie a donc été employée en réalisant une nouvelle immunodétection sur les clones possédant les étiquettes peptidiques, qui étaient déjà disponibles et qui possèdent une bonne expression de la protéine recombinante. Ainsi, le co-transformant 54599-V5 / XylT-cmyc a été utilisé pour réaliser un double marquage anti-V5 et anti-cmyc, et le co-transformant GnT I-V5 / XylT-cmyc a été utilisé pour réaliser un simple marquage anti-V5. En effet, d'après les résultats de PCR et western blot, le clone GnT I-V5 / XylT-cmyc exprimant le plus la GnT I est dépourvu du transgène XylT-cmyc. Il a donc été considéré comme un simple transformant GnT I-V5 pour la suite de l'étude.

Les observations réalisées par MET montrent que la structure cellulaire des microalgues est parfaitement conservée suite aux différents traitements, et permet ainsi d'observer correctement l'architecture des organites (**figure 92**). Lorsque l'on se focalise sur l'appareil de Golgi, d'après (Donohoe et al., 2013) il est possible d'orienter les saccules, avec le *cis* Golgi composé de citernes épaisses à la lumière faiblement marquée, alors que les citernes *median* et *trans*, plus fines, possèdent un matériel luminal beaucoup plus marqué après contraste.

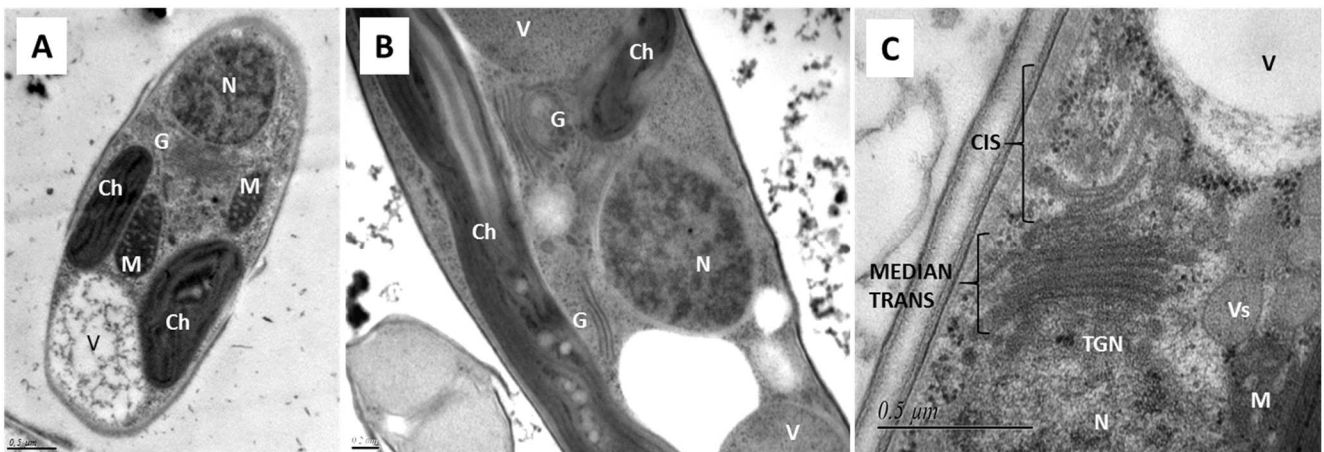


Figure 92 : Observation en microscopie électronique à transmission de l'organisation ultrastructurale de la microalgue *Phaeodactylum tricornutum* Pt1.8.6

A : vue générale de l'organisation cellulaire

B : Focalisation sur le noyau et l'appareil de Golgi

C : Structure de l'appareil de Golgi

N : noyau ; Ch : chloroplaste ; V : vacuole ; Vs : vésicule de sécrétion ; M : mitochondrie ; G : appareil de Golgi ; TGN : réseau *trans* golgien

L'observation du transformant GnT I-V5 permet de mettre en évidence que la protéine recombinante est bien située dans l'appareil de Golgi, localisée plutôt dans les saccules *median* comme attendu (**figure 93**). Par extrapolation, ce résultat permet ainsi de confirmer la localisation golgienne des protéines recombinantes observées précédemment par microscopie confocale.

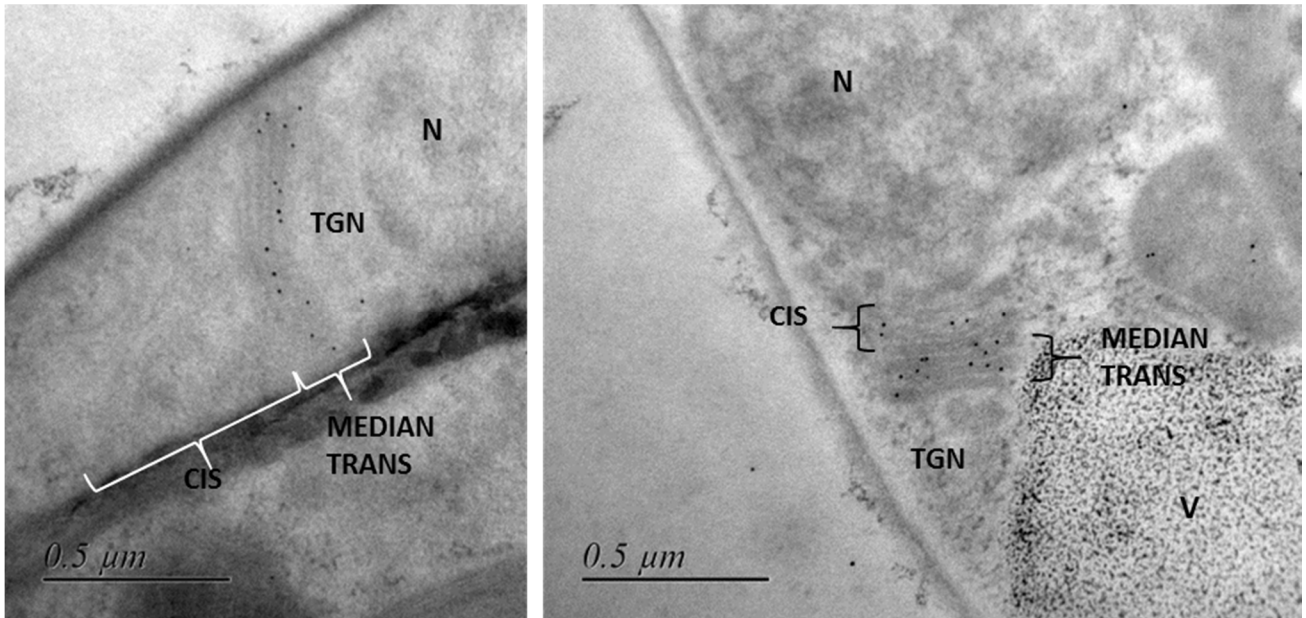


Figure 93 : Observation en microscopie électronique à transmission de la localisation golgienne de la GnT I-V5 après immunomarquage avec un anticorps primaire dirigé contre l'épitope V5, et un anticorps secondaire couplé à une particule d'or de 10nm. Les deux prises de vue proviennent de deux cellules issues du même clone GnT I-V5/XylT-cmyc n°5 ne possédant pas le transgène XylT-cmyc. N: noyau ; V : vacuole ; TGN : réseau *trans* golgien

En ce qui concerne le double transformant 54599-V5 / XylT-cmyc, seul le marquage dirigé contre la protéine FuT54599-V5 a fonctionné. Il est ainsi possible d'observer la localisation de cette protéine qui se situerait plutôt dans le compartiment *median* de l'appareil de Golgi (**figure 94**), comme conjecturé précédemment par microscopie confocale. La seule étude de localisation subcellulaire de l'activité fucosyltransférase du core (Fitchette-Lainé et al., 1994) a montré de manière indirecte que le fucose était présent majoritairement dans les saccules *trans* de l'appareil de Golgi. Ces travaux ont été réalisés par MET en utilisant une technique d'immunodétection avec des anticorps primaires dirigés contre le résidu fucose lié en alpha-1,3 des *N*-glycannes fucosylés au niveau du core. Ces travaux montrent donc que chez la plante les produits résultant de l'activité des fucosyltransférases du core sont majoritairement situés dans la saccule *trans*. Cependant, cette information n'est pas contradictoire avec notre

résultat. En effet, la fucosylation du core pourrait avoir lieu dans le saccule *médian*, puis les protéines portant des *N*-glycannes ainsi fucosylés pourraient être transportées et accumulées dans les compartiments *trans* et TGN où ils sont majoritairement retrouvés. Cette hypothèse peut être étayée par le fait qu'une quantité non négligeable (15,9 %) de *N*-glycannes fucosylés au niveau du core est aussi retrouvée dans le compartiment *médian*. Peu d'informations sont actuellement disponibles quant à la localisation des fucosyltransférases du core. Cependant, d'après des analyses biochimiques par spectrométrie de masse réalisées sur les *N*-glycannes d'*A. thaliana* (Strasser et al., 2006), la présence de structures *N*-glycanniques sous la forme $\text{Man}_5\text{XylFucGlcNAc}_2$ et $\text{GlcNAcMan}_5\text{XylFucGlcNAc}_2$ montre que les fucosyltransférases du core peuvent intervenir dans le processus de maturation des glycannes avant l'action de l' α -mannosidase II. Cette protéine étant résidente de la citerne *median* de l'appareil de Golgi (Dejgaard et al., 2007; Schoberer and Strasser, 2011; Velasco et al., 1993), il est possible d'en déduire que les fucosyltransférases du core peuvent être localisées au moins dans le même compartiment, ce que nous observons dans notre étude, voire même dans des saccules plus en amont. Ces travaux réalisés sur la localisation subcellulaire des fucosyltransférases chez *P. tricornutum* permettent ainsi d'apporter de nouvelles informations sur l'organisation des glycosyltransférases impliquées dans le processus de maturation des *N*-glycannes au sein de l'appareil de Golgi.

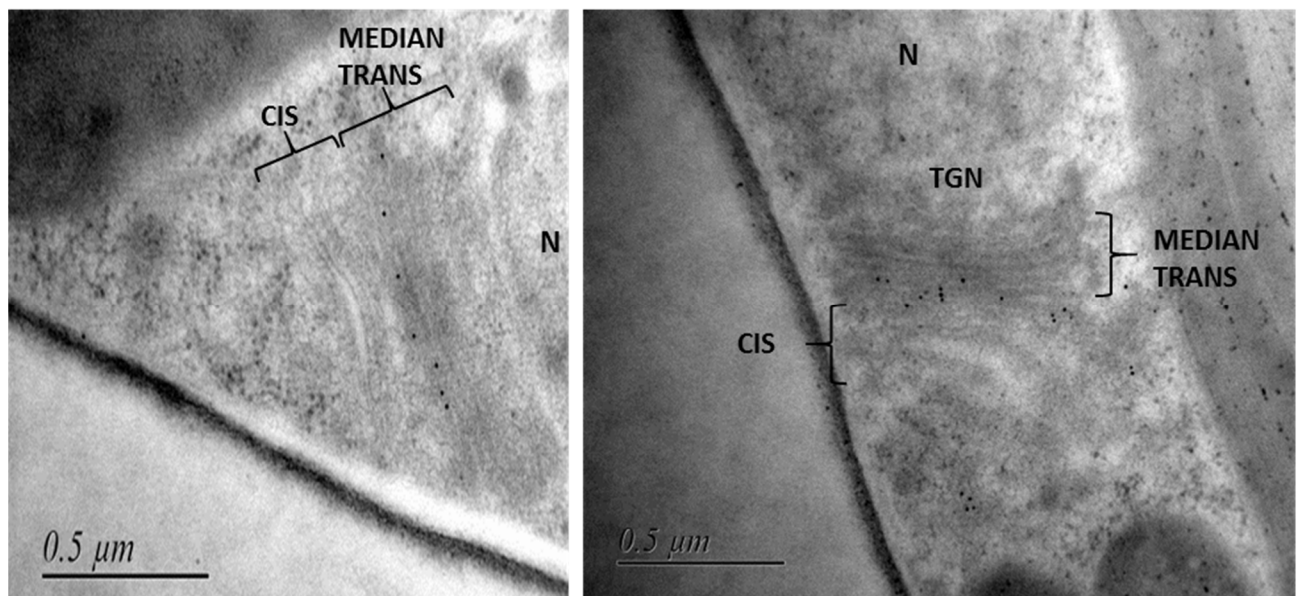


Figure 94 : Observation en microscopie électronique à transmission de la localisation golgienne de la FuT54599-V5 après immunomarquage avec un anticorps primaire dirigé contre l'épitope V5, et un anticorps secondaire couplé à une particule d'or de 10nm.

Les deux prises de vue proviennent de deux cellules issues du même clone FuT54599-V5/XylT-cmyc n°8. N: noyau ; TGN : réseau *trans* golgien

7 Discussion et perspectives de la caractérisation des fucosyltransférases de *P. tricornutum*

Les fucosyltransférases impliquées dans la synthèse des *N*-glycannes permettent d'ajouter un résidu fucose sur les GlcNAc des *N*-glycannes complexes. L'ajout de ce résidu est un des principaux points de divergence dans la voie de maturation des *N*-glycannes chez les différents organismes. L'ajout de ce résidu fucose peut ainsi être effectué sur les GlcNAc terminales, permettant la création de motifs glycaniques de type Lewis (Breton et al., 1998a). Dans cette situation, l'ajout du résidu fucose peut être effectué en α 1,3 chez les mammifères (de Vries et al., 2001) et chez certains insectes (Kubelka et al., 1993; Kurz et al., 2016), ou en α 1,4 chez les plantes (Lerouge et al., 1998). Au contraire, la fucosylation du core glycanique au niveau de la GlcNAc proximale peut être observée soit en α 1,6 chez les animaux (Voynow et al., 1991), en α 1,3 chez les plantes (Johnson and Chrispeels, 1987), et en α 1,3 et α 1,6 chez certains invertébrés, qui arboreront ainsi une bifucosylation de leurs *N*-glycannes (Staudacher and März, 1998). Les fucosyltransférases responsables de ces modifications appartiennent à deux catégories différentes, avec d'un côté les α 1,6-fucosyltransférases rattachées à la famille GT23, et d'un autre les α 1,3/4-fucosyltransférases agissant sur les GlcNAc proximales et distales qui appartiennent à la famille GT10.

Les analyses *in silico* réalisées par homologie de séquence sur le génome de *P. tricornutum* ont révélé la présence de trois gènes putatifs codant pour des fucosyltransférases potentiellement responsables de l'ajout de fucose au niveau du core glycanique. En effet, des *N*-glycannes complexes fucosylés sur le core glycanique ont été détectés en faible quantité lors de précédentes analyses biochimiques réalisées sur les *N*-glycannes de *P. tricornutum* (Baïet et al., 2011). Ces gènes possèdent des annotations sous les noms de PHATRDRAFT_46109, PHATRDRAFT_46110 et PHATRDRAFT_54599 sur la base de données du JGI. Les analyses bio-informatiques réalisées au cours de cette étude sur les gènes prédits ont permis de mettre en évidence que les prédictions PHATRDRAFT_46109 et PHATRDRAFT_46110 sont très probablement erronées. En effet, les protéines prédites par ces deux annotations semblent dépourvues de domaine transmembranaire, contrairement à ce qui est attendu pour une glycosyltransférase, protéine membranaire de type II. Pour arriver à cette conclusion, nous avons utilisé trois programmes de prédiction différents : TMHMM, TMPred et DAS, afin de croiser les résultats de différents algorithmes de prédiction. Cependant, bien que nous ayons observé que les programmes TMHMM et TMPred présentaient sensiblement les mêmes résultats de prédiction de domaines transmembranaires, il s'est avéré que le

programme DAS n'est pas assez stringent, et génère ainsi plus de faux positifs que les deux autres programmes. En effet, ceci a pu être mis en évidence par la prédiction de domaines transmembranaires inférieurs à 10, voire 5 acides aminés, alors que ces domaines sont habituellement constitués d'une vingtaine d'acides aminés (Singh and Mittal, 2016). Afin de conserver une certaine rigueur dans nos travaux, nous avons décidé de ne pas prendre en considération dans nos raisonnements les prédictions de DAS présentant des domaines transmembranaires trop éloignés des 20 acides aminés attendus. Nous avons ainsi pu établir de nouvelles prédictions qui permettent d'obtenir des protéines avec une structure en adéquation avec celle attendue pour des protéines membranaires de type II. Ainsi, la séquence nucléotidique relative à l'annotation PHATRDRAFT_46109 a été rallongée en 5', permettant ainsi d'ajouter à la protéine prédite une séquence d'acides aminés en *N*-terminal pouvant correspondre à un domaine transmembranaire. De même, nous avons mis en évidence par RT-PCR que le premier intron de la prédiction PHATRDRAFT_46110 correspondait finalement à une séquence codante. En prenant en considération cette nouvelle information, nous avons réalisé une nouvelle prédiction qui modifie ainsi la FuT46110 sur son extrémité *N*-terminale, tout en lui ajoutant un nouveau domaine transmembranaire. Ces nouvelles prédictions peuvent maintenant être confortées avec les résultats générés lors de l'analyse RNA-seq réalisée à partir des trois morphotypes de *P. tricornutum*. En effet, les profils d'expression issus de cette étude transcriptomique ont pu mettre en évidence pour le gène PHATRDRAFT_46109 l'expression d'un transcrit plus long en 5' comme prédit (**Figure 95**).

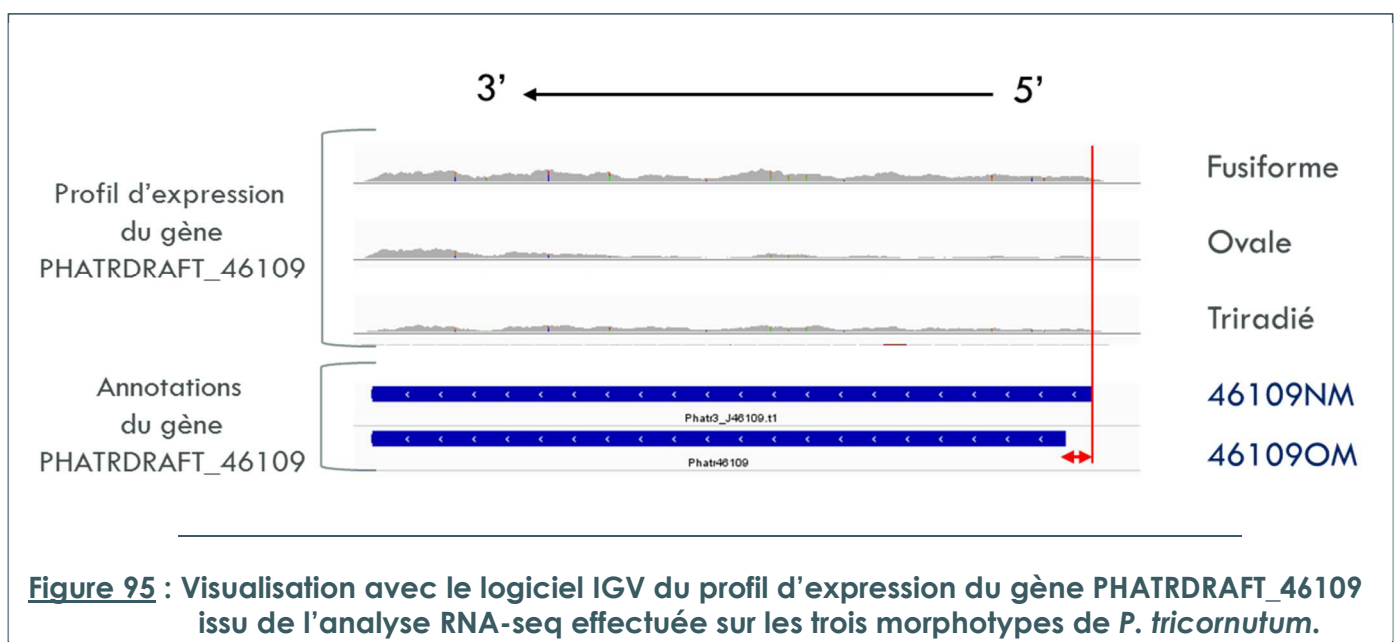
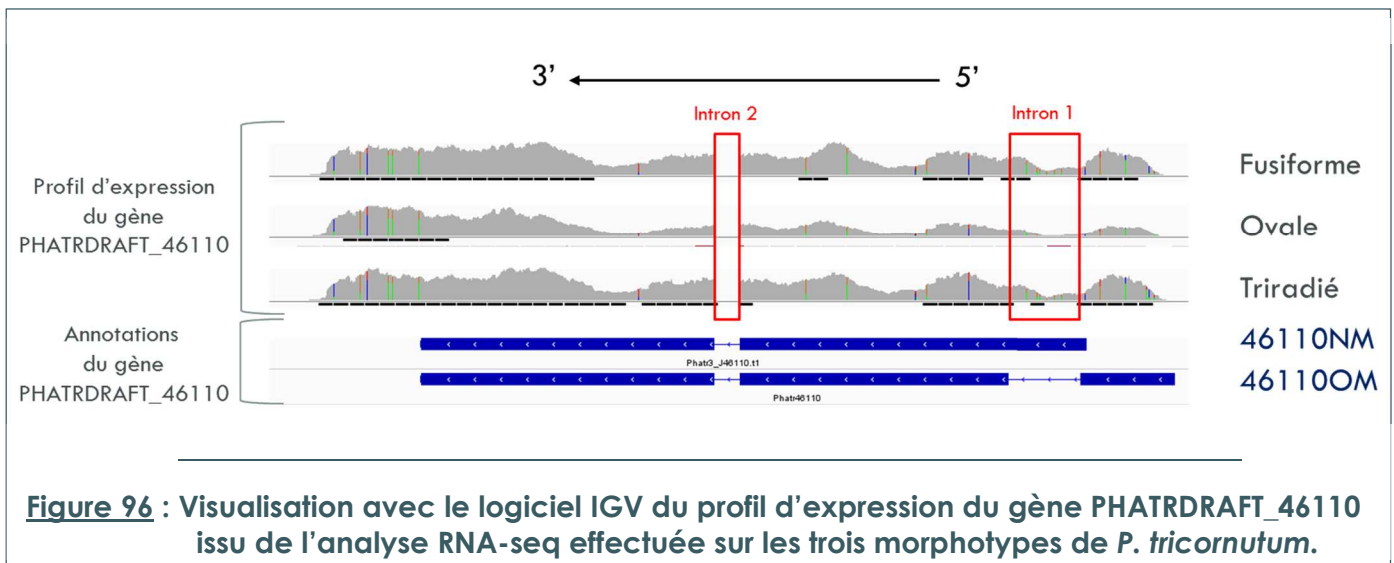


Figure 95 : Visualisation avec le logiciel IGV du profil d'expression du gène PHATRDRAFT_46109 issu de l'analyse RNA-seq effectuée sur les trois morphotypes de *P. tricornutum*.

De la même manière, pour le gène PHATRDRRAFT_46110 il est possible d'observer par l'absence de séquence que le second intron est bien épissé, contrairement au premier intron pour lequel on peut observer un profil d'expression, ce qui traduit une absence d'épissage conformément à nos précédents résultats (**Figure 96**). L'analyse transcriptomique réalisée au cours de cette étude nous a aussi permis de mettre en évidence la présence d'un polymorphisme génétique commun aux trois morphotypes, et qui est notamment observable sur les séquences de nos trois gènes PHATRDRRAFT_46109, PHATRDRRAFT_46110 et PHATRDRRAFT_54599.



De nouvelles prédictions pour les gènes PHATRDRRAFT_46109 et PHATRDRRAFT_46110 ont été mises récemment à disposition sur la base de données Ensembl Protists sous les noms respectifs Phatr3_J46109 et Phatr3_J46110. Les comparaisons de séquences entre les annotations d'Ensembl et celles que nous avons réalisées lors de ces travaux ont montré que l'annotation Phatr3_J46109 est identique à notre annotation 46109NM, ce qui appuie une nouvelle fois notre prédiction. Cependant, bien que la comparaison entre l'annotation Phatr3_J46110 et notre annotation 46110NM présente une grande similarité, notamment par l'absence du premier intron, on observe une différence à l'extrémité 5'. En effet, notre prédiction débute au codon start précédant celui de l'annotation Phatr3_J46110. Cette différence est cruciale car elle permet d'ajouter un domaine transmembranaire à la protéine prédite à partir de notre séquence 46110NM, contrairement à l'annotation Phatr3_J46110 où ce domaine est absent. Cette observation montre que notre prédiction est certainement plus fiable, et démontre ainsi que malgré une mise à jour récente des annotations du génome de *P. tricornutum*, et bien que ces dernières soient de meilleure qualité, elles demeurent perfectibles.

Les trois fucosyltransférases putatives de *P. tricornutum* possèderaient une structure membranaire de type II avec un domaine transmembranaire situé à l'extrémité *N*-terminale, comme retrouvé chez la plupart des glycosyltransférases. Elles sont aussi dotées toutes les trois d'un domaine GT10 similaire aux α 1,3/4-fucosyltransférases, ce qui coïncide avec de précédents résultats obtenus par western-blot mettant en évidence la présence de glycannes α 1,3-fucosylés (Baiet et al., 2011). Cependant, la FuT54599 présente un second domaine GT10_N, que l'on retrouve notamment chez les α 1,3-fucosyltransférases des invertébrés, et qui est absent des deux autres fucosyltransférases de *P. tricornutum*. La structure des FuT46109 et FuT46110 semble donc plus proche de celle rencontrée chez les α 1,3-fucosyltransférases de plantes. La FuT54599 diffère aussi des deux autres fucosyltransférases par sa taille. En effet, la FuT54599 possède une taille semblable à celle rencontrée pour les α 1,3-fucosyltransférases des autres organismes, contrairement aux FuT46109 et FuT46110 qui possèdent une séquence plus longue, de l'ordre de 25% à 30% de plus que la plus grande des α 1,3-fucosyltransférases caractérisées. Cela pose ainsi la question de l'origine de ces longueurs supplémentaires. Est-ce qu'elles confèrent à ces deux fucosyltransférases des fonctions supplémentaires, ou sont-ce des vestiges de l'évolution que les pressions de sélection n'auraient pas réussi à éliminer ? Nous pouvons émettre l'hypothèse que *P. tricornutum*, organisme unicellulaire particulièrement tolérant à son environnement, n'aurait pas nécessité pour son développement ou sa survie, de fucosyltransférases aussi évoluées que celles des organismes pour lesquels les fucosyltransférases ont déjà été caractérisées. En effet, l'étude des fucosyltransférases a été réalisée principalement pour des organismes pluricellulaires dont le développement est très dépendant des *N*-glycannes (Marek et al., 1999), ce qui implique une pression de sélection forte sur les glycosyltransférases, contrairement aux organismes unicellulaires comme *P. tricornutum* qui aurait pu conserver des domaines ancestraux dans ses fucosyltransférases. De manière générale, peu d'informations sont disponibles sur les fucosyltransférases GT10 d'organismes unicellulaires, et la seule à avoir déjà été caractérisée est issue d'*H. pylori* qui l'utilise pour synthétiser des lipopolysaccharides. Ces derniers sont indispensables à la survie d'*H. pylori* dans l'estomac qui est donc soumise à de grandes pressions de sélection (Sun et al., 2007).

De manière totalement opposée, on peut aussi émettre l'hypothèse que les séquences protéiques supplémentaires correspondent à des domaines fonctionnels encore non caractérisés, qui auraient été acquis évolutivement en apportant des avantages pour la survie de *P. tricornutum*.

Cependant, en suivant ces deux raisonnements, pourquoi la FuT54599, n'aurait-elle pas subi le même processus que les deux autres FuT46109 et FuT46110 ? L'activité de ces glycosyltransférases n'étant pas encore caractérisée, il est possible d'imaginer que la FuT54599 catalyse une réaction possédant un impact évolutif différent des deux autres fucosyltransférases, créant ainsi une disparité de structure stimulée par une pression de sélection différente. Cependant, l'explication la plus probable liée à la différence de taille, mais aussi surtout de structure, opposant la FuT54599 par rapport aux deux autres fucosyltransférases pourrait provenir d'un transfert horizontal du gène PHATR DRAFT_54599, à l'instar de plusieurs centaines d'autres gènes composant le génome de *P. tricornutum* (Bowler et al., 2008).

La localisation subcellulaire des trois fucosyltransférases putatives de *P. tricornutum* a pu être initiée, dans un premier temps par microscopie confocale, puis par microscopie électronique à transmission. Dans ce but, le clonage des trois fucosyltransférases putatives a été réalisé avec succès, de la même manière que les gènes codant pour la xylosyltransférase putative et la GnT I de *P. tricornutum*, ainsi que pour la β 1,4-galactosyltransférase humaine, protéines utilisées ici comme témoin de localisation dans l'appareil de Golgi. La fusion de ces gènes avec une étiquette GFP ou RFP a permis de mettre en évidence par microscopie confocale une localisation ponctiforme « *Golgi-like* » de ces protéines, à l'exception des anciennes prédictions FuT46109OM et FuT46110OM qui produisent respectivement une protéine mal localisée et une protéine tronquée non visualisable. Ces résultats nous ont permis de valider les nouvelles prédictions réalisées pour les gènes 46109 et 46110, et aussi de mettre en évidence que chez *P. tricornutum*, comme observé chez les vertébrés et les plantes (Saint-Jore-Dupas et al., 2006; Sharpe et al., 2010), l'extrémité *N*-terminale des glycosyltransférases est cruciale pour leur adressage à l'appareil de Golgi. De plus, les résultats obtenus avec les co-transformants XylT/GnT I et XylT/FuT54599 nous ont permis d'émettre les hypothèses suivantes : la XylT putative de *P. tricornutum* serait localisée dans le compartiment *cis* ou *trans* de l'appareil de Golgi, tandis que la FuT54599 serait probablement localisée dans le compartiment *median*. Ces résultats apportent de nouvelles informations concernant ces deux glycosyltransférases pour lesquelles très peu de choses sont actuellement connues. Ces observations ont ensuite pu être confirmées pour la GnT-I et la FuT54599 par microscopie électronique à transmission, montrant que ces glycosyltransférases sont bien adressées à l'appareil de Golgi. La résolution apportée par cette technique nous a permis de certifier la localisation de la GnT I dans les saccules *median* de l'appareil de Golgi, et de valider notre hypothèse quant à la localisation de la FuT54599 dans ce même compartiment. A notre connaissance, ces résultats constituent la première localisation subcellulaire de glycosyltransférases chez les microalgues, ainsi que la

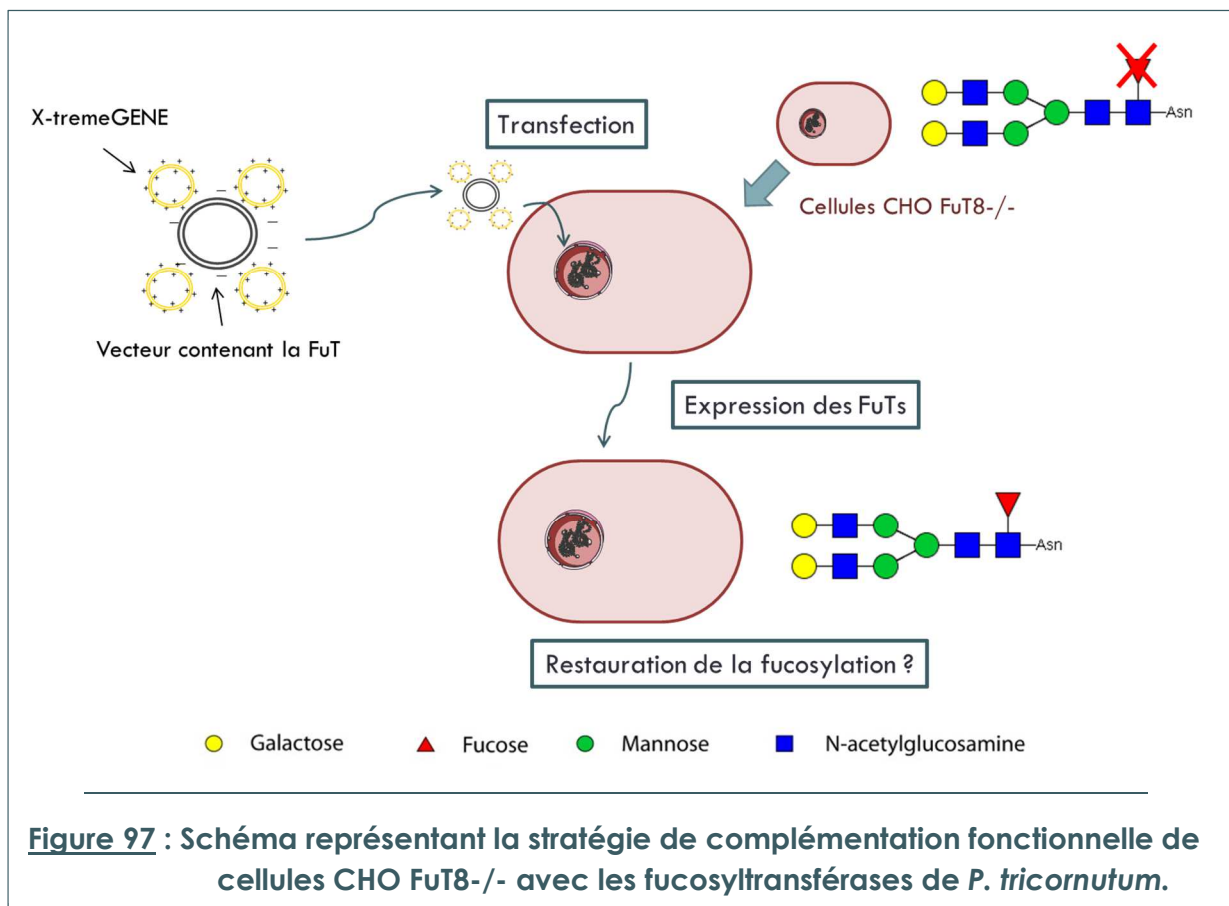
première étude par MET de la localisation d'une fucosyltransférase du core *N*-glycannique. De plus, la transfection de la GalT humaine chez *P. tricornutum* a permis l'expression de la première glycosyltransférase humaine chez cette microalgue, montrant ainsi que *P. tricornutum* est capable d'adresser correctement ces protéines à l'appareil de Golgi. Ce résultat est prometteur dans un contexte biotechnologique où l'humanisation de la voie de *N*-glycosylation de *P. tricornutum* est nécessaire pour produire des protéines recombinantes à intérêt thérapeutique (bio-médicaments). En effet, cette microalgue ne possède pas de galactose, ni d'acide sialique, à l'extrémité de ses *N*-glycannes, épitopes glucidiques qui sont importants dans l'activité et la demi-vie des bio-médicaments utilisés en thérapie chez l'Homme. Les *N*-glycannes de *P. tricornutum* étant principalement oligomannosidiques, dont majoritairement du Man-5, la stratégie d'humanisation pourrait être réalisée dans un premier temps par sur-expression de la GnT I endogène, qui produirait ainsi assez de substrat pour une Gal T recombinante. L'ajout de galactose par cette dernière permettrait ensuite l'action d'une sialyltransférase recombinante. Cependant, les recherches d'acide sialiques menées par chromatographie n'ont pas permises de mettre en évidence leur présence chez *P. tricornutum* (Baiet et al., 2011). De même, aucun gène impliqué dans sa synthèse n'a pu être retrouvé lors d'analyses bio-informatiques par homologie de séquences. Ceci implique donc que pour humaniser la voie de *N*-glycosylation chez *P. tricornutum*, en plus des glycosyltransférases citées précédemment, il sera nécessaire de transférer la microalgue avec les différentes enzymes impliquées dans la synthèse de l'acide sialique, tout comme ce qui a déjà été réalisé chez les plantes (Castilho et al., 2008, 2013; Jez et al., 2013; Kallolimath et al., 2016).

Afin de compléter l'étude de la localisation golgienne des différentes glycosyltransférases intervenant dans la voie de *N*-glycosylation, une nouvelle transfection dans *P. tricornutum* devra être réalisée pour les constructions n'ayant pas pu être intégrées au génome, ou n'ayant pas exprimé les protéines recombinantes, à savoir les constructions contenant les transgènes FuT46110NM-V5, XylT-cmyc, ainsi que le co-transformant GnT I-V5/XylT-cmyc. Les clones positifs obtenus récemment ayant intégré les transgènes FuT46109NM-V5 et GalT-HA devront être criblés par western blot afin de vérifier l'expression de la protéine recombinante, puis pourront ensuite être analysés par MET.

En parallèle de cette étude de localisation subcellulaire, une étude fonctionnelle pourra être menée avec les clones de *P. tricornutum* transformés avec ces différentes glycosyltransférases, ce qui permettra notamment de confirmer l'activité des trois fucosyltransférases putatives. Les *N*-glycannes des différents clones pourront être analysés par spectrométrie de masse afin de détecter une éventuelle modification liée à la sur-expression des

glycosyltransférases. Si la caractérisation fonctionnelle de ces enzymes n'est pas réalisable avec cette stratégie, les protéines recombinantes pourront être immunopurifiées via l'étiquette peptidique fusionnée pour ensuite être caractérisées *in vitro*.

Une caractérisation par complémentation fonctionnelle dans des cellules CHO FuT8^{-/-} a aussi été initiée pour les trois fucosyltransférases putatives de *P. tricornutum*. Les séquences de ces enzymes ont ainsi été clonées dans un vecteur permettant l'expression de protéines recombinantes dans un système mammifère. Ces constructions génétiques pourront maintenant être transfectées dans les cellules CHO mutées sur la fucosyltransférase du core. Les *N*-glycannes de ces cellules pourront ensuite être purifiés, puis analysés par spectrométrie de masse afin d'observer si les fucosyltransférases recombinantes ont permis la restauration de la fucosylation au niveau du core (**figure 97**).



Conclusion générale

Le projet de cette thèse s'est inscrit dans une thématique de recherche du laboratoire Glyco-MEV qui vise à comprendre les processus de *N*-glycosylation chez les microalgues. Dans ce contexte, les objectifs de ma thèse furent d'étudier 1- la voie de *N*-glycosylation chez la diatomée modèle *P. tricornutum* en me focalisant plus spécifiquement sur le clonage et la caractérisation de fucosyltransférases putatives et 2- les mécanismes de morphogenèse de cette microalgue polymorphique.

Lors de l'étude des 3 gènes candidats codant pour les α 1,3-fucosyltransférases putatives de *P. tricornutum*, nous avons pu remettre en cause l'annotation de deux de ces gènes, et proposer de nouvelles annotations qui ont été validées par une approche *in vivo* de localisation subcellulaire par microscopie. En effet, nous avons pu observer la bonne localisation de ces trois fucosyltransférases au sein de l'appareil de Golgi, mais également celle de la GnT I et d'une XylT putative de *P. tricornutum*, et de la β 1,4-GalT humaine, surexprimée chez *P. tricornutum* et utilisée ici comme marqueur de l'appareil de Golgi. Ce témoin a également permis de montrer les capacités de *P. tricornutum* à correctement adresser une glycosyltransférase hétérologue (humaine dans le cas présent) dans son propre appareil de Golgi. Ce résultat est particulièrement intéressant car il a été observé que les anticorps recombinants produits chez *P. tricornutum* portaient majoritairement des *N*-glycannes oligomannosidiques (Hempel et al., 2011a; Vanier et al., 2015). Or, la présence de β 1,4-galactose à l'extrémité terminale des *N*-glycannes est requise pour les fonctions effectrices de l'anticorps et pour assurer une liaison optimale de l'anticorps à toutes les classes de récepteurs (Jefferis et al., 1998). La possibilité d'exprimer correctement cette β 1,4-GalT humaine chez *P. tricornutum* constitue ainsi un résultat intéressant dans un contexte d'humanisation de sa voie de *N*-glycosylation à des fins de production de bio-médicaments. De plus, cette étude nous a permis de réaliser la première localisation subcellulaire de glycosyltransférases chez les microalgues, ainsi que, sous réserve de confirmation de l'activité fonctionnelle, la première localisation d'une fucosyltransférase du core *N*-glycannique. Enfin, des constructions génétiques ont été réalisées avec les trois α 1,3-fucosyltransférases putatives de *P. tricornutum* afin de réaliser prochainement une caractérisation fonctionnelle de ces enzymes par complémentation dans des cellules CHO avec pour objectif de démontrer leur activité biologique et, si possible, confirmer leur implication dans l'ajout de fucose sur le core *N*-glycannique.

Afin de lever les inconnues entourant la morphogenèse de *P. tricornutum*, nous avons réalisé une approche globale par transcriptomique à haut débit sur des cultures enrichies en

chacun des trois morphotypes de *P. tricornutum*. Une analyse comparative de ces trois transcriptomes nous a permis d'observer une étonnante similarité entre le transcriptome des morphotypes triradié et fusiforme, qui possèdent plus de 99% de similarité à une pvalue ajustée de 0,01%. De manière opposée, le morphotype ovale possède un transcriptome très différent des deux autres, avec 22% de ses gènes différentiellement exprimés en comparaison au morphotype fusiforme, et 29% en comparaison avec le morphotype triradié. Dans la suite de cette étude, nous nous sommes donc intéressés aux gènes spécifiquement différentiellement exprimés chez le morphotype ovale. Ceci nous a permis de mettre en évidence dans ce morphotype une activation de voies métaboliques clés comme la glycolyse, la biosynthèse des nucléotides et des triglycérides, ou encore la voie des pentoses phosphate. Nous avons aussi pu mettre en avant une régulation de la biosynthèse des purines et pyrimidines, ainsi qu'une activation de la biosynthèse des triglycérides et du GDP-Mannose. La surproduction de ce nucléotide-sucre dans le morphotype ovale comparé aux deux autres morphotypes pourrait être expliquée par l'utilisation de ce composé dans la biosynthèse de glucuronomannane qui compose majoritairement la paroi cellulaire de *P. tricornutum*, paroi qui est particulièrement épaisse dans le morphotype ovale. Les voies métaboliques associées aux acides aminés semblent aussi perturbées, allant ainsi à la faveur de la synthèse de Sérine, Cystéine, Phénylalanine, Tyrosine et Isoleucine, et vers une diminution de la synthèse de Glycine, Glutamate, Arginine et Aspartate dans le morphotype ovale. Associées à ces perturbations du métabolisme des acides aminés, des modifications de la synthèse des aminoacyl-ARNt ont pu aussi être observées, ce qui nous a menés vers l'étude du métabolisme de protéines chez le morphotype ovale. Il est possible aussi d'observer une sur-expression de nombreux gènes impliqués dans les voies de dégradation des protéines. De manière intéressante, le morphotype ovale présente une sur-expression des gènes impliqués dans le trafic des protéines et la sécrétion, ce qui semble montrer que ce morphotype possède une machinerie cellulaire plus active que les deux autres morphotypes. D'après nos connaissances, cette analyse RNA-seq est la première analyse transcriptomique à avoir été réalisée sur les trois morphotypes de *P. tricornutum* sur des cultures fortement enrichies en un morphotype (plus de 80%), et issues de la même souche Pt3. Ce travail nous permet d'élargir les connaissances sur la morphogénèse de *P. tricornutum* et de mettre à disposition de la communauté scientifique une source d'informations qualitative et quantitative. Qui plus est, ces données ont également permis de mettre en lumière de nouveaux gènes de référence stablement exprimés entre les trois morphotypes pour des analyses de qRT-PCR, gènes de références qui ont été utilisés pour valider les données de RNA-Seq et pour étudier l'expression de gènes de la voie de

N-glycosylation. A l'avenir, nous envisageons d'utiliser ces données transcriptomiques pour poursuivre l'annotation du génome de *P. tricornutum* qui nécessite encore d'être amélioré. Effectivement, la qualité des profils d'expression obtenus permet de mettre en évidence l'existence d'épissage alternatif pour de nombreux gènes et la présence de nouveaux transcrits.

Références bibliographiques

- Abdullahi, A.S., Underwood, G.J.C., and Gretz, M.R. (2006). Extracellular matrix assembly in diatoms (Bacillariophyceae). V. Environmental effects on polysaccharide synthesis in the model diatom, *Phaeodactylum tricornutum*. *J. Phycol.* 42, 363–378.
- Aebi, M. (2013). N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1833, 2430–2437.
- Aebi, M., Gassenhuber, J., Domdey, H., and te Heesen, S. (1996). Cloning and characterization of the ALG3 gene of *Saccharomyces cerevisiae*. *Glycobiology* 6, 439–444.
- Aebi, M., Bernasconi, R., Clerc, S., and Molinari, M. (2010). N-glycan structures: recognition and processing in the ER. *Trends Biochem. Sci.* 35, 74–82.
- Aikawa, J. -i., Takeda, Y., Matsuo, I., and Ito, Y. (2014). Trimming of glucosylated N-glycans by human ER 1,2-mannosidase I. *J. Biochem. (Tokyo)* 155, 375–384.
- Altmann, F., and März, L. (1995). Processing of asparagine-linked oligosaccharides in insect cells: evidence for alpha-mannosidase II. *Glycoconj. J.* 12, 150–155.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64, 5245–5250.
- Apt, K.E., Grossman, A.R., and Kroth-Pancic, P.G. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet. MGG* 252, 572–579.
- Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta BBA-Gen. Subj.* 1473, 4–8.
- Armbrust, E.V. (2009). The life of diatoms in the world's oceans. *Nature* 459, 185–192.
- Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., and Bechner, M. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306, 79–86.
- Audry, M., Jeanneau, C., Imbert, A., Harduin-Lepers, A., Delannoy, P., and Breton, C. (2011). Current trends in the structure-activity relationships of sialyltransferases. *Glycobiology* 21, 716–726.
- Baiet, B., Burel, C., Saint-Jean, B., Louvet, R., Menu-Bouaouiche, L., Kiefer-Meyer, M.-C., Mathieu-Rivet, E., Lefebvre, T., Castel, H., Carlier, A., et al. (2011). N-Glycans of *Phaeodactylum tricornutum* Diatom and Functional Characterization of Its N-Acetylglucosaminyltransferase I Enzyme. *J. Biol. Chem.* 286, 6152–6164.
- Bakker, H., Bardor, M., Molthoff, J.W., Gomord, V., Elbers, I., Stevens, L.H., Jordi, W., Lommen, A., Faye, L., and Lerouge, P. (2001). Galactose-extended glycans of antibodies produced by transgenic plants. *Proc. Natl. Acad. Sci.* 98, 2899–2904.

- Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A Kingdom-Level Phylogeny of Eukaryotes Based on Combined Protein Data. *Science* 290, 972–977.
- Bardor, M., Faveeuw, C., Fitchette, A.-C., Gilbert, D., Galas, L., Trottein, F., Faye, L., and Lerouge, P. (2003). Immunoreactivity in mammals of two typical plant glyco-epitopes, core α (1, 3)-fucose and core xylose. *Glycobiology* 13, 427–434.
- Barker, M.K., and Rose, D.R. (2013). Specificity of Processing α -Glucosidase I Is Guided by the Substrate Conformation: CRYSTALLOGRAPHIC AND IN SILICO STUDIES. *J. Biol. Chem.* 288, 13563–13574.
- Bartual, A., Gálvez, J.A., and Ojeda, F. (2008). Phenotypic response of the diatom *Phaeodactylum tricornutum* Bohlin to experimental changes in the inorganic carbon system. *Bot. Mar.* 51.
- Bartual, A., Villazán, B., and Brun, F.G. (2011). Monitoring the long-term stability of pelagic morphotypes in the model diatom *Phaeodactylum tricornutum*. *Diatom Res.* 26, 243–253.
- Bause, E. (1983). Structural requirements of *N*-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* 209, 331–336.
- Becker, E.W. (2007). Micro-algae as a source of protein. *Biotechnol. Adv.* 25, 207–210.
- van Beers, M.M.C., and Bardor, M. (2012). Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. *Biotechnol. J.* 7, 1473–1484.
- Bennett, J., Hondred, D., and Register, J.C. (2015). Keeping qRT-PCR rigorous and biologically relevant. *Plant Cell Rep.* 34, 1–3.
- Berger, E.G., and Hesford, F.J. (1985). Localization of galactosyl- and sialyltransferase by immunofluorescence: evidence for different sites. *Proc. Natl. Acad. Sci.* 82, 4736–4739.
- Bickel, T., Lehle, L., Schwarz, M., Aebi, M., and Jakob, C.A. (2005). Biosynthesis of lipid-linked oligosaccharides in *Saccharomyces cerevisiae*: Alg13p and Alg14p form a complex required for the formation of GlcNAc(2)-PP-dolichol. *J. Biol. Chem.* 280, 34500–34506.
- Blackwell, J., Parker, K.D., and Rudall, K.M. (1967). Chitin fibres of the diatoms *Thalassiosira fluviatilis* and *Cyclotella cryptica*. *J. Mol. Biol.* 28, 383–385.
- Bohlin, K.H. (1897). *Zur Morphologie und Biologie einzelliger Algen*.
- Borowitzka, M.A., and Volcani, B.E. (1978). The polymorphic diatom *Phaeodactylum tricornutum*: ultrastructure of its morphotypes. *J. Phycol.* 14, 10–21.
- Both, P., Sobczak, L., Breton, C., Hann, S., Nobauer, K., Paschinger, K., Kozmon, S., Mucha, J., and Wilson, I.B.H. (2011). Distantly related plant and nematode core 1,3-fucosyltransferases display similar trends in structure-function relationships. *Glycobiology* 21, 1401–1415.
- Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otiillar, R.P., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244.

- Bowler, C., Vardi, A., and Allen, A.E. (2010). Oceanographic and Biogeochemical Insights from Diatom Genomes. *Annu. Rev. Mar. Sci.* 2, 333–365.
- Bozarth, A., Maier, U.-G., and Zauner, S. (2009). Diatoms in biotechnology: modern tools and applications. *Appl. Microbiol. Biotechnol.* 82, 195–201.
- Brada, D., and Dubach, U.C. (1984). Isolation of a homogeneous glucosidase II from pig kidney microsomes. *Eur. J. Biochem.* 141, 149–156.
- Breton, C. (2005). Structures and mechanisms of glycosyltransferases. *Glycobiology* 16, 29R–37R.
- Breton, C., and Imberty, A. (1999). Structure/function studies of glycosyltransferases. *Curr. Opin. Struct. Biol.* 9, 563–571.
- Breton, C., Oriol, R., and Imberty, A. (1998a). Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. *Glycobiology* 8, 87–94.
- Breton, C., Bettler, E., Joziase, D.H., Geremia, R.A., and Imberty, A. (1998b). Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. *J. Biochem. (Tokyo)* 123, 1000–1009.
- Breuer, W., Klein, R.A., Hardt, B., Bartoschek, A., and Bause, E. (2001). Oligosaccharyltransferase is highly specific for the hydroxy amino acid in Asn-Xaa-Thr/Ser. *FEBS Lett.* 501, 106–110.
- Brighouse, A., Dacks, J.B., and Field, M.C. (2010). Rab protein evolution and the history of the eukaryotic endomembrane system. *Cell. Mol. Life Sci.* 67, 3449–3465.
- Brockhausen, I., and Schachter, H. (1996). Glycosyltransferases Involved in *N*- and *O*-Glycan Biosynthesis. In *Glycosciences*, H.-J. Gabius, and S. Gabius, eds. (Weinheim, Germany: Wiley-VCH Verlag GmbH), pp. 79–113.
- Burda, P., and Aebi, M. (1998). The ALG10 locus of *Saccharomyces cerevisiae* encodes the alpha-1,2 glucosyltransferase of the endoplasmic reticulum: the terminal glucose of the lipid-linked oligosaccharide is required for efficient *N*-linked glycosylation. *Glycobiology* 8, 455–462.
- Burda, P., Jakob, C.A., Beinhauer, J., Hegemann, J.H., and Aebi, M. (1999). Ordered assembly of the asymmetrically branched lipid-linked oligosaccharide in the endoplasmic reticulum is ensured by the substrate specificity of the individual glycosyltransferases. *Glycobiology* 9, 617–625.
- Bustin, S.A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 25, 169–193.
- Bustin, S.A. (2010). Why the need for qPCR publication guidelines?--The case for MIQE. *Methods San Diego Calif* 50, 217–226.
- Caramelo, J.J., Castro, O.A., Alonso, L.G., de Prat-Gay, G., and Parodi, A.J. (2003). UDP-Glc: glycoprotein glucosyltransferase recognizes structured and solvent accessible hydrophobic patches in molten globule-like folding intermediates. *Proc. Natl. Acad. Sci.* 100, 86–91.

- Castilho, A., Pabst, M., Leonard, R., Veit, C., Altmann, F., Mach, L., Glössl, J., Strasser, R., and Steinkellner, H. (2008). Construction of a functional CMP-sialic acid biosynthesis pathway in *Arabidopsis*. *Plant Physiol.* *147*, 331–339.
- Castilho, A., Neumann, L., Gattinger, P., Strasser, R., Vorauer-Uhl, K., Sterovsky, T., Altmann, F., and Steinkellner, H. (2013). Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants. *PLoS ONE* *8*, e54836.
- Castilho, A., Windwarder, M., Gattinger, P., Mach, L., Strasser, R., Altmann, F., and Steinkellner, H. (2014). Proteolytic and *N*-Glycan Processing of Human 1-Antitrypsin Expressed in *Nicotiana benthamiana*. *Plant Physiol.* *166*, 1839–1851.
- Chao, D.S., Hay, J.C., Winnick, S., Prekeris, R., Klumperman, J., and Scheller, R.H. (1999). SNARE membrane trafficking dynamics in vivo. *J. Cell Biol.* *144*, 869–881.
- Chapman, J.R., and Waldenström, J. (2015). With Reference to Reference Genes: A Systematic Review of Endogenous Controls in Gene Expression Studies. *PloS One* *10*, e0141853.
- Chen, M.M., Glover, K.J., and Imperiali, B. (2007). From Peptide to Protein: Comparative Analysis of the Substrate Specificity of *N*-Linked Glycosylation in *C. jejuni* †. *Biochemistry (Mosc.)* *46*, 5579–5585.
- Cheong, F.Y., Sharma, V., Blagoveshchenskaya, A., Oorschot, V.M.J., Brankatschk, B., Klumperman, J., Freeze, H.H., and Mayinger, P. (2010). Spatial Regulation of Golgi Phosphatidylinositol-4-Phosphate is Required for Enzyme Localization and Glycosylation Fidelity. *Traffic* *11*, 1180–1190.
- Chepurnov, V.A., Mann, D.G., Sabbe, K., and Vyverman, W. (2004). Experimental Studies on Sexual Reproduction in Diatoms. In *International Review of Cytology*, (Elsevier), pp. 91–154.
- Chevalier, L., Bernard, S., Ramdani, Y., Lamour, R., Bardor, M., Lerouge, P., Follet-Gueye, M.-L., and Driouich, A. (2010). Subcompartment localization of the side chain xyloglucan-synthesizing enzymes within Golgi stacks of tobacco suspension-cultured cells: Golgi localization of xyloglucan-synthesizing enzymes. *Plant J.* *64*, 977–989.
- Chiovitti, A., Harper, R.E., Willis, A., Bacic, A., Mulvaney, P., and Wetherbee, R. (2005). Variations in the substituted 3-linked mannans closely associated with the silicified walls of diatoms : substituted mannans of diatoms. *J. Phycol.* *41*, 1154–1161.
- Choi, Y.H., Kim, J.H., Park, J.H., Lee, N., Kim, D.-H., Jang, K.-S., Park, I.-H., and Kim, B.-G. (2014). Protein engineering of 2,3/2,6-sialyltransferase to improve the yield and productivity of in vitro sialyllactose synthesis. *Glycobiology* *24*, 159–169.
- Christensen, L.L., Jensen, U.B., Bross, P., and Orntoft, T.F. (2000). The C-terminal *N*-glycosylation sites of the human alpha1,3/4-fucosyltransferase III, -V, and -VI (hFucTIII, -V, and -VI) are necessary for the expression of full enzyme activity. *Glycobiology* *10*, 931–939.
- Ciechanover, A., Orian, A., and Schwartz, A.L. (2000). Ubiquitin-mediated proteolysis: biological regulation via destruction. *BioEssays* *22*, 442–451.
- Cipollo, J.F., Trimble, R.B., Chi, J.H., Yan, Q., and Dean, N. (2001). The Yeast ALG11 Gene Specifies Addition of the Terminal 1,2-Man to the Man5GlcNAc2-PP-dolicholN-Glycosylation

Intermediate Formed on the Cytosolic Side of the Endoplasmic Reticulum. *J. Biol. Chem.* 276, 21828–21840.

Clynes, R.A., Towers, T.L., Presta, L.G., and Ravetch, J.V. (2000). Inhibitory Fc receptors modulate in vivo cytotoxicity against tumor targets. *Nat. Med.* 6, 443–446.

Coesel, S., Oborník, M., Varela, J., Falciatore, A., and Bowler, C. (2008). Evolutionary Origins and Functions of the Carotenoid Biosynthetic Pathway in Marine Diatoms. *PLoS ONE* 3, e2896.

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

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17.

Conti, B.J., Devaraneni, P.K., Yang, Z., David, L.L., and Skach, W.R. (2015). Cotranslational Stabilization of Sec62/63 within the ER Sec61 Translocon Is Controlled by Distinct Substrate-Driven Translocation Events. *Mol. Cell* 58, 269–283.

Cooksey, K.E., and Cooksey, B. (1974). Calcium deficiency can induce the transition from oval to fusiform cells in cultures of *Phaeodactylum tricornutum* Bohlin. *J. Phycol.* 10, 89–90.

Couto, J.R., Huffaker, T.C., and Robbins, P.W. (1984). Cloning and expression in *Escherichia coli* of a yeast mannosyltransferase from the asparagine-linked glycosylation pathway. *J. Biol. Chem.* 259, 378–382.

Culyba, E.K., Price, J.L., Hanson, S.R., Dhar, A., Wong, C.-H., Gruebele, M., Powers, E.T., and Kelly, J.W. (2011). Protein Native-State Stabilization by Placing Aromatic Side Chains in *N*-Glycosylated Reverse Turns. *Science* 331, 571–575.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* 139, 5–17.

Czlapinski, J., and Bertozzi, C. (2006). Synthetic glycobiology: exploits in the Golgi compartment. *Curr. Opin. Chem. Biol.* 10, 645–651.

Darley, W.M. (1968). Deoxyribonucleic acid content of the three cell types of *Phaeodactylum tricornutum* Bohlin. *J. Phycol.* 4, 219–220.

De Martino, A., Meichenin, A., Shi, J., Pan, K., and Bowler, C. (2007). Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. *J. Phycol.* 43, 992–1009.

De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U., and Bowler, C. (2011). Physiological and Molecular Evidence that Environmental Changes Elicit Morphological Interconversion in the Model Diatom *Phaeodactylum tricornutum*. *Protist* 162, 462–481.

- De Pourcq, K., De Schutter, K., and Callewaert, N. (2010). Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl. Microbiol. Biotechnol.* *87*, 1617–1631.
- De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* *37*, e96–e96.
- Dejgaard, S.Y., Murshid, A., Dee, K.M., and Presley, J.F. (2007). Confocal microscopy-based linescan methodologies for intra-Golgi localization of proteins. *J. Histochem. Cytochem.* *55*, 709–719.
- Deprez, P., Gautschi, M., and Helenius, A. (2005). More Than One Glycan Is Needed for ER Glucosidase II to Allow Entry of Glycoproteins into the Calnexin/Calreticulin Cycle. *Mol. Cell* *19*, 183–195.
- van Dijk, E.L., Jaszczyszyn, Y., and Thermes, C. (2014). Library preparation methods for next-generation sequencing: tone down the bias. *Exp. Cell Res.* *322*, 12–20.
- Dillschneider, R., Steinweg, C., Rosello-Sastre, R., and Posten, C. (2013). Biofuels from microalgae: Photoconversion efficiency during lipid accumulation. *Bioresour. Technol.* *142*, 647–654.
- Donohoe, B.S., Kang, B.-H., Gerl, M.J., Gergely, Z.R., McMichael, C.M., Bednarek, S.Y., and Staehelin, L.A. (2013). *Cis*-Golgi Cisternal Assembly and Biosynthetic Activation Occur Sequentially in Plants and Algae: *Cis*-Golgi Cisternal Assembly and Biosynthetic Activation in Plants and Algae. *Traffic* *14*, 551–567.
- Dugdale, T.M., Willis, A., and Wetherbee, R. (2006). Adhesive Modular Proteins Occur in the Extracellular Mucilage of the Motile, Pennate Diatom *Phaeodactylum tricornutum*. *Biophys. J.* *90*, L58–L60.
- Dupuy, F., Petit, J.-M., Mollicone, R., Oriol, R., Julien, R., and Maftah, A. (1999). A Single Amino Acid in the Hypervariable Stem Domain of Vertebrate α 1, 3/1, 4-Fucosyltransferases Determines the Type 1/Type 2 Transfer CHARACTERIZATION OF ACCEPTOR SUBSTRATE SPECIFICITY OF THE LEWIS ENZYME BY SITE-DIRECTED MUTAGENESIS. *J. Biol. Chem.* *274*, 12257–12262.
- Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. *Trends Genet. TIG* *29*, 569–574.
- Ellgaard, L., and Helenius, A. (2001). ER quality control: towards an understanding at the molecular level. *Curr. Opin. Cell Biol.* *13*, 431–437.
- Ellgaard, L., and Helenius, A. (2003). Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* *4*, 181–191.
- Fabini, G., Freilinger, A., Altmann, F., and Wilson, I. (2001). Identification of Core α 1,3-Fucosylated Glycans and Cloning of the Requisite Fucosyltransferase cDNA from *Drosophila melanogaster*.
- Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C., and Bowler, C. (1999). Transformation of nonselectable reporter genes in marine diatoms. *Mar. Biotechnol.* *1*, 239–251.

- Faye, L., Sturm, A., Bollini, R., Vitale, A., and Chrispeels, M.J. (1986). The position of the oligosaccharide side-chains of phytohemagglutinin and their accessibility to glycosidases determines their subsequent processing in the Golgi. *Eur. J. Biochem.* *158*, 655–661.
- Fiedler, K., Parton, R.G., Kellner, R., Etzold, T., and Simons, K. (1994). VIP36, a novel component of glycolipid rafts and exocytic carrier vesicles in epithelial cells. *EMBO J.* *13*, 1729.
- Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* *281*, 237–240.
- Fitchette, A.C., Cabanes-Macheteau, M., Marvin, L., Martin, B., Satiat-Jeunemaitre, B., Gomord, V., Crooks, K., Lerouge, P., Faye, L., and Hawes, C. (1999). Biosynthesis and immunolocalization of Lewis a-containing *N*-glycans in the plant cell. *Plant Physiol.* *121*, 333–344.
- Fitchette-Lainé, A.-C., Gomord, V., Chekkafi, A., and Faye, L. (1994). Distribution of xylosylation and fucosylation in the plant Golgi apparatus. *Plant J.* *5*, 673–682.
- Fitchette-Lainé, A.C., Gomord, V., Cabanes, M., Michalski, J.C., Saint Macary, M., Foucher, B., Cavelier, B., Hawes, C., Lerouge, P., and Faye, L. (1997). *N*-glycans harboring the Lewis a epitope are expressed at the surface of plant cells. *Plant J. Cell Mol. Biol.* *12*, 1411–1417.
- Ford, C.W., and Percival, E. (1965). The carbohydrates of *Phaeodactylum tricornutum*. Part I. Preliminary examination of the organism, and characterisation of low molecular weight material and of a glucan. *J. Chem. Soc. Resumed* 7035–7041.
- Francius, G., Tesson, B., Dague, E., Martin-Jézéquel, V., and Dufrêne, Y.F. (2008). Nanostructure and nanomechanics of live *Phaeodactylum tricornutum* morphotypes. *Environ. Microbiol.* *10*, 1344–1356.
- Frank, C.G., and Aebi, M. (2005). ALG9 mannosyltransferase is involved in two different steps of lipid-linked oligosaccharide biosynthesis. *Glycobiology* *15*, 1156–1163.
- Frickel, E.-M., Riek, R., Jelesarov, I., Helenius, A., Wüthrich, K., and Ellgaard, L. (2002). TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proc. Natl. Acad. Sci.* *99*, 1954–1959.
- Fukuda, T., Hashimoto, H., Okayasu, N., Kameyama, A., Onogi, H., Nakagawasai, O., Nakazawa, T., Kurosawa, T., Hao, Y., Isaji, T., et al. (2011). 1,6-Fucosyltransferase-deficient Mice Exhibit Multiple Behavioral Abnormalities Associated with a Schizophrenia-like Phenotype: IMPORTANCE OF THE BALANCE BETWEEN THE DOPAMINE AND SEROTONIN SYSTEMS. *J. Biol. Chem.* *286*, 18434–18443.
- Gachon, C., Mingam, A., and Charrier, B. (2004). Real-time PCR: what relevance to plant studies. *J. Exp. Bot.* *55*, 1445–1454.
- Gao, X.-D., Nishikawa, A., and Dean, N. (2004). Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum. *Glycobiology* *14*, 559–570.

- Gavel, Y., and von Heijne, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* *3*, 433–442.
- Gillard, J., Frenkel, J., Devos, V., Sabbe, K., Paul, C., Rempt, M., Inzé, D., Pohnert, G., Vuylsteke, M., and Vyverman, W. (2013). Metabolomics Enables the Structure Elucidation of a Diatom Sex Pheromone. *Angew. Chem. Int. Ed.* *52*, 854–857.
- Glickman, M.H., and Ciechanover, A. (2002). The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* *82*, 373–428.
- Guénin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C., and Gutierrez, L. (2009). Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J. Exp. Bot.* *60*, 487–493.
- Guerin, M., and Parodi, A.J. (2003). The UDP-glucose:Glycoprotein Glucosyltransferase Is Organized in at Least Two Tightly Bound Domains from Yeast to Mammals. *J. Biol. Chem.* *278*, 20540–20546.
- Gügi, B., Costaeuec, T., Burel, C., Lerouge, P., Helbert, W., and Bardor, M. (2015). Diatom-Specific Oligosaccharide and Polysaccharide Structures Help to Unravel Biosynthetic Capabilities in Diatoms. *Mar. Drugs* *13*, 5993–6018.
- Gutenbrunner, S.A., Thalhamer, J., and Schmid, A.-M.M. (1994). Proteinaceous and immunochemical distinctions between the oval and fusiform morphotypes of *Phaeodactylum tricorutum* (Bacillariophyceae). *J. Phycol.* *30*, 129–136.
- Gutierrez, L., Mauriat, M., Pelloux, J., Bellini, C., and Van Wuytswinkel, O. (2008). Towards a systematic validation of references in real-time rt-PCR. *Plant Cell* *20*, 1734–1735.
- Hahn, M., Borisova, S., Schrag, J.D., Tessier, D.C., Zapun, A., Tom, R., Kamen, A.A., Bergeron, J.J.M., Thomas, D.Y., and Cygler, M. (1998). Identification and Crystallization of a Protease-Resistant Core of Calnexin That Retains Biological Activity. *J. Struct. Biol.* *123*, 260–264.
- Hanson, S.R., Culyba, E.K., Hsu, T.-L., Wong, C.-H., Kelly, J.W., and Powers, E.T. (2009). The core trisaccharide of an *N*-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci.* *106*, 3131–3136.
- Hasle, G.R. (1983). The Marine, Planktonic Diatoms *Thalassiosira-Oceanica Sp-Nov* and *Thalassiosira-Partheneia*. *J. Phycol.* *19*, 220–229.
- Hauri, H.-P., Kappeler, F., Andersson, H., and Appenzeller, C. (2000). ERGIC-53 and traffic in the secretory pathway. *J. Cell Sci.* *113*, 587–596.
- He, L., Han, X., and Yu, Z. (2014). A Rare *Phaeodactylum tricorutum* Cruciform Morphotype: Culture Conditions, Transformation and Unique Fatty Acid Characteristics. *PLoS ONE* *9*, e93922.
- Hecky, R.E., Mopper, K., Kilham, P., and Degens, E.T. (1973). The amino acid and sugar composition of diatom cell-walls. *Mar. Biol.* *19*, 323–331.

- Heesen, S. te, Lehle, L., Weissmann, A., and Aebi, M. (1994). Isolation of the ALG5 Locus Encoding the UDP-Glucose: Dolichyl-Phosphate Glucosyltransferase from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 224, 71–79.
- Helenius, A. (1994). How *N*-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell* 5, 253.
- Helenius, A., and Aebi, M. (2004). Roles of *N*-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* 73, 1019–1049.
- Helenius, J., Ng, D.T.W., Marolda, C.L., Walter, P., Valvano, M.A., and Aebi, M. (2002). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* 415, 447–450.
- Hemmer, W., Focke, M., Kolarich, D., Wilson, I.B.H., Altmann, F., Wöhrl, S., Götz, M., and Jarisch, R. (2001). Antibody binding to venom carbohydrates is a frequent cause for double positivity to honeybee and yellow jacket venom in patients with stinging-insect allergy. *J. Allergy Clin. Immunol.* 108, 1045–1052.
- Hempel, F., Lau, J., Klingl, A., and Maier, U.G. (2011a). Algae as Protein Factories: Expression of a Human Antibody and the Respective Antigen in the Diatom *Phaeodactylum tricornutum*. *PLoS ONE* 6, e28424.
- Hempel, F., Bozarth, A.S., Lindenkamp, N., Klingl, A., Zauner, S., Linne, U., Steinbüchel, A., and Maier, U.G. (2011b). Microalgae as bioreactors for bioplastic production. *Microb Cell Fact* 10, 81.
- Hendey, N.I., Cushing, D.H., and Ripley, G.W. (1954). Electron microscope studies of diatoms. *J. R. Microsc. Soc.* 74, 22–34.
- Hettkamp, H., Legler, G., and Bause, E. (1984). Purification by affinity chromatography of glucosidase I, an endoplasmic reticulum hydrolase involved in the processing of asparagine-linked oligosaccharides. *Eur. J. Biochem.* 142, 85–90.
- Higgins, M.J., Molino, P., Mulvaney, P., and Wetherbee, R. (2003). The structure and nanomechanical properties of the adhesive mucilage that mediates diatom-substratum adhesion and motility. *J. Phycol.* 39, 1181–1193.
- High, S., Lecomte, F.J., Russell, S.J., Abell, B.M., and Oliver, J.D. (2000). Glycoprotein folding in the endoplasmic reticulum: a tale of three chaperones? *FEBS Lett.* 476, 38–41.
- Hildebrand, M., Volcani, B.E., Gassmann, W., and Schroeder, J.I. (1997). A gene family of silicon transporters. *Nature* 385, 688–689.
- Hildebrand, M., Dahlin, K., and Volcani, B.E. (1998). Characterization of a silicon transporter gene family in *Cylindrotheca fusiformis*: sequences, expression analysis, and identification of homologs in other diatoms. *Mol. Gen. Genet. MGG* 260, 480–486.
- Hoek, C. van den, Mann, D.G., and Jahns, H.M. (1995). *Algae: an introduction to phycology* (Cambridge ; New York: Cambridge University Press).

- Holmes, E.H., Yen, T.-Y., Thomas, S., Joshi, R., Nguyen, A., Long, T., Gallet, F., Maftah, A., Julien, R., and Macher, B.A. (2000). Human α 1, 3/4 Fucosyltransferases CHARACTERIZATION OF HIGHLY CONSERVED CYSTEINE RESIDUES AND N-LINKED GLYCOSYLATION SITES. *J. Biol. Chem.* *275*, 24237–24245.
- Holst, S., Deuss, A.J.M., van Pelt, G.W., van Vliet, S.J., Garcia-Vallejo, J.J., Koeleman, C.A.M., Deelder, A.M., Mesker, W.E., Tollenaar, R.A., Rombouts, Y., et al. (2016). *N*-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homeobox 1 (CDX1)/Villin mRNA Expression. *Mol. Cell. Proteomics* *15*, 124–140.
- Hruz, T., Wyss, M., Docquier, M., Pfaffl, M.W., Masanetz, S., Borghi, L., Verbrugge, P., Kalaydjieva, L., Bleuler, S., Laule, O., et al. (2011). RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC Genomics* *12*, 156.
- Huang, A., He, L., and Wang, G. (2011). Identification and characterization of microRNAs from *Phaeodactylum tricornutum* by high-throughput sequencing and bioinformatics analysis. *BMC Genomics* *12*, 1.
- Huggett, J., Dheda, K., Bustin, S., and Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* *6*, 279–284.
- Huysman, M.J., Martens, C., Vandepoele, K., Gillard, J., Rayko, E., Heijde, M., Bowler, C., Inzé, D., Peer, Y., De Veylder, L., Vyverman, W. (2010). Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biol.* *11*, R17.
- Iida, S. (2006). Nonfucosylated Therapeutic IgG1 Antibody Can Evade the Inhibitory Effect of Serum Immunoglobulin G on Antibody-Dependent Cellular Cytotoxicity through its High Binding to Fc RIIIa. *Clin. Cancer Res.* *12*, 2879–2887.
- Imberty, A., Monier, C., Bettler, E., Morera, S., Freemont, P., Sippl, M., Flöckner, H., Rüger, W., and Breton, C. (1999). Fold recognition study of α 3-galactosyltransferase and molecular modeling of the nucleotide sugar-binding domain. *Glycobiology* *9*, 713–722.
- Imperiali, B., and Hendrickson, T.L. (1995). Asparagine-linked glycosylation: Specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.* *3*, 1565–1578.
- Imperiali, B., and O'Connor, S.E. (1999). Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* *3*, 643–649.
- Jeanneau, C. (2004). Structure-Function Analysis of the Human Sialyltransferase ST3Gal I: ROLE OF N-GLYCOSYLATION AND A NOVEL CONSERVED SIALYL MOTIF. *J. Biol. Chem.* *279*, 13461–13468.
- Jefferis, R., Lund, J., and Pound, J.D. (1998). IgG-Fc-mediated effector functions: molecular definition of interaction sites for effector ligands and the role of glycosylation. *Immunol. Rev.* *163*, 59–76.
- Jensen, K.G., Moestrup, Ø., and Schmid, A.-M.M. (2003). Ultrastructure of the male gametes from two centric diatoms, *Chaetoceros lacinosus* and *Coscinodiscus walesii* (Bacillariophyceae). *Phycologia* *42*, 98–105.

- Jez, J., Castilho, A., Grass, J., Vorauer-Uhl, K., Sterovsky, T., Altmann, F., and Steinkellner, H. (2013). Expression of functionally active sialylated human erythropoietin in plants. *Biotechnol. J.* 8, 371–382.
- Johansen, J.R. (1991). Morphological variability and cell wall composition of *Phaeodactylum tricornutum* (Bacillariophyceae). *Gt. Basin Nat.* 51, 310–315.
- Johnson, K.D., and Chrispeels, M.J. (1987). Substrate specificities of *N*-acetylglucosaminyl-, fucosyl-, and xylosyltransferases that modify glycoproteins in the Golgi apparatus of bean cotyledons. *Plant Physiol.* 84, 1301–1308.
- de Jonge, H.J.M., Fehrmann, R.S.N., de Bont, E.S.J.M., Hofstra, R.M.W., Gerbens, F., Kamps, W.A., de Vries, E.G.E., van der Zee, A.G.J., te Meerman, G.J., and ter Elst, A. (2007). Evidence based selection of housekeeping genes. *PloS One* 2, e898.
- Jost, F. (2004). Mutation of amino acids in the alpha 1,3-fucosyltransferase motif affects enzyme activity and Km for donor and acceptor substrates. *Glycobiology* 15, 165–175.
- Kaewmuangmoon, J., Kilaso, M., Leartsakulpanich, U., Kimura, K., Kimura, A., and Chanchao, C. (2013). Expression of a secretory α -glucosidase II from *Apis cerana indica* in *Pichia pastoris* and its characterization. *BMC Biotechnol.* 13, 16.
- Kallolimath, S., Castilho, A., Strasser, R., Grünwald-Gruber, C., Altmann, F., Strubl, S., Galuska, C.E., Zlatina, K., Galuska, S.P., Werner, S., et al. (2016). Engineering of complex protein sialylation in plants. *Proc. Natl. Acad. Sci.* 113, 9498–9503.
- Kampf, M., Absmanner, B., Schwarz, M., and Lehle, L. (2009). Biochemical Characterization and Membrane Topology of Alg2 from *Saccharomyces cerevisiae* as a Bifunctional 1,3- and 1,6-Mannosyltransferase Involved in Lipid-linked Oligosaccharide Biosynthesis. *J. Biol. Chem.* 284, 11900–11912.
- Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P.R., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., Jablanovic, J., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6, 6925.
- Kaushal, G.P., Szumilo, T., Pastuszak, I., and Elbein, A.D. (1990). Purification to homogeneity and properties of mannosidase II from mung bean seedlings. *Biochemistry (Mosc.)* 29, 2168–2176.
- Keeling, P.J. (2013). The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. *Annu. Rev. Plant Biol.* 64, 583–607.
- Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9, 605–618.
- Kelleher, D.J., Banerjee, S., Cura, A.J., Samuelson, J., and Gilmore, R. (2007). Dolichol-linked oligosaccharide selection by the oligosaccharyltransferase in protist and fungal organisms. *J. Cell Biol.* 177, 29–37.
- Kern, G., Kern, D., Jaenicke, R., and Seckler, R. (1993). Kinetics of folding and association of differently glycosylated variants of invertase from *Saccharomyces cerevisiae*. *Protein Sci.* 2, 1862–1868.

- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* *14*, 1.
- Koprivova, A., Altmann, F., Gorr, G., Kopriva, S., Reski, R., and Decker, E.L. (2003). *N*-Glycosylation in the Moss *Physcomitrella patens* is Organized Similarly to that in Higher Plants. *Plant Biol.* *5*, 582–591.
- Kozera, B., and Rapacz, M. (2013). Reference genes in real-time PCR. *J. Appl. Genet.* *54*, 391–406.
- Kozlov, G., Maattanen, P., Schrag, J.D., Pollock, S., Cygler, M., Nagar, B., Thomas, D.Y., and Gehring, K. (2006). Crystal Structure of the bb' Domains of the Protein Disulfide Isomerase ERp57. *Structure* *14*, 1331–1339.
- Kröger, N. (2007). Prescribing diatom morphology: toward genetic engineering of biological nanomaterials. *Curr. Opin. Chem. Biol.* *11*, 662–669.
- Kröger, N., and Poulsen, N. (2008). Diatoms—From Cell Wall Biogenesis to Nanotechnology. *Annu. Rev. Genet.* *42*, 83–107.
- Kröger, N., and Wetherbee, R. (2000). Pleuralins are Involved in Theca Differentiation in the Diatom *Cylindrotheca fusiformis*. *Protist* *151*, 263–273.
- Kröger, N., Deutzmann, R., Bergsdorf, C., and Sumper, M. (2000). Species-specific polyamines from diatoms control silica morphology. *Proc. Natl. Acad. Sci.* *97*, 14133–14138.
- Kubelka, V., Altmann, F., Staudacher, E., Tretter, V., März, L., HARD, K., Kamerling, J.P., and Vliegenthart, J.F. (1993). Primary structures of the *N*-linked carbohydrate chains from honeybee venom phospholipase A2. *Eur. J. Biochem.* *213*, 1193–1204.
- Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strömbom, L., et al. (2006). The real-time polymerase chain reaction. *Mol. Aspects Med.* *27*, 95–125.
- Kuczynska, P., Jemiola-Rzeminska, M., and Strzalka, K. (2015). Photosynthetic Pigments in Diatoms. *Mar. Drugs* *13*, 5847–5881.
- Kudo, I., Miyamoto, M., Noiri, Y., and Maita, Y. (2000). Combined effects of temperature and iron on the growth and physiology of the marine diatom *Phaeodactylum tricornutum* (Bacillariophyceae). *J. Phycol.* *36*, 1096–1102.
- Kuliawat, R., Kalinina, E., Bock, J., Fricker, L., McGraw, T.E., Kim, S.R., Zhong, J., Scheller, R., and Arvan, P. (2004). Syntaxin-6 SNARE involvement in secretory and endocytic pathways of cultured pancreatic beta-cells. *Mol. Biol. Cell* *15*, 1690–1701.
- Kurz, S., King, J.G., Dinglasan, R.R., Paschinger, K., and Wilson, I.B.H. (2016). The fucomic potential of mosquitoes: Fucosylated *N*-glycan epitopes and their cognate fucosyltransferases. *Insect Biochem. Mol. Biol.* *68*, 52–63.

- Lamriben, L., Graham, J.B., Adams, B.M., and Hebert, D.N. (2015). *N*-glycan based ER molecular chaperone and protein quality control system: the calnexin binding cycle. *Traffic* n/a-n/a.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lauc, G., KriÅ;tiÄ;†, J., and ZoldoÄ;†, V. (2014). Glycans - The third revolution in evolution. *Front. Genet.* 5.
- Lee, R.E. (1999). *Phycology* (Cambridge [England] ; New York: Cambridge University Press).
- Léonard, R., Lhernould, S., Carlué, M., Fleurat, P., Maftah, A., and Costa, G. (2005). Biochemical characterization of *Silene alba* α 4-fucosyltransferase and Lewis a products. *Glycoconj. J.* 22, 71–78.
- Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Lainé, A.-C., Gomord, V., and Faye, L. (1998). *N*-glycoprotein biosynthesis in plants: recent developments and future trends. In *Protein Trafficking in Plant Cells*, (Springer), pp. 31–48.
- Lewin, J.C. (1958). The taxonomic position of *Phaeodactylum tricornutum*. *J. Gen. Microbiol.* 18, 427–432.
- Lewin, J.C., Lewin, R.A., and Philpott, D.E. (1958). Observations on *Phaeodactylum tricornutum*. *J. Gen. Microbiol.* 18, 418–426.
- Lewis, G.D., Figari, I., Fendly, B., Wong, W.L., Carter, P., Gorman, C., and Shepard, H.M. (1993). Differential responses of human tumor cell lines to anti-p185HER2 monoclonal antibodies. *Cancer Immunol. Immunother.* 37, 255–263.
- Lewis, M.J., Rayner, J.C., and Pelham, H.R. (1997). A novel SNARE complex implicated in vesicle fusion with the endoplasmic reticulum. *EMBO J.* 16, 3017–3024.
- Li, Y., Lei, X., Zhu, H., Zhang, H., Guan, C., Chen, Z., Zheng, W., Fu, L., and Zheng, T. (2016). Chitinase producing bacteria with direct algicidal activity on marine diatoms. *Sci. Rep.* 6, 21984.
- Li, Z., Stafford, W.F., and Bouvier, M. (2001). The Metal Ion Binding Properties of Calreticulin Modulate Its Conformational Flexibility and Thermal Stability †. *Biochemistry (Mosc.)* 40, 11193–11201.
- Liebminger, E., Veit, C., Pabst, M., Batoux, M., Zipfel, C., Altmann, F., Mach, L., and Strasser, R. (2011). *N*-Acetylhexosaminidases HEXO1 and HEXO3 Are Responsible for the Formation of Paucimannosidic *N*-Glycans in *Arabidopsis thaliana*. *J. Biol. Chem.* 286, 10793–10802.
- Lin, Y., Zhang, C., Lan, H., Gao, S., Liu, H., Liu, J., Cao, M., Pan, G., Rong, T., and Zhang, S. (2014). Validation of potential reference genes for qPCR in maize across abiotic stresses, hormone treatments, and tissue types. *PLoS One* 9, e95445.

- Lingg, N., Zhang, P., Song, Z., and Bardor, M. (2012). The sweet tooth of biopharmaceuticals: Importance of recombinant protein glycosylation analysis. *Biotechnol. J.* 7, 1462–1472.
- Liu, Y., and Barlowe, C. (2002). Analysis of Sec22p in endoplasmic reticulum/Golgi transport reveals cellular redundancy in SNARE protein function. *Mol. Biol. Cell* 13, 3314–3324.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25, 402–408.
- Lizak, C., Gerber, S., Numao, S., Aebi, M., and Locher, K.P. (2011). X-ray structure of a bacterial oligosaccharyltransferase. *Nature* 474, 350–355.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15.
- Lu, Y., and Wold, F. (1991). Regulation of glycan processing by Golgi enzymes from red kidney bean (*Phaseolus vulgaris*) seedlings. *Arch. Biochem. Biophys.* 286, 147–152.
- Lu, J., Takahashi, T., Ohoka, A., Nakajima, K. -i., Hashimoto, R., Miura, N., Tachikawa, H., and Gao, X.-D. (2012). Alg14 organizes the formation of a multiglycosyltransferase complex involved in initiation of lipid-linked oligosaccharide biosynthesis. *Glycobiology* 22, 504–516.
- Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 29, 1830–1831.
- Määttänen, P., Gehring, K., Bergeron, J.J.M., and Thomas, D.Y. (2010). Protein quality control in the ER: The recognition of misfolded proteins. *Semin. Cell Dev. Biol.* 21, 500–511.
- Maheswari, U., Mock, T., Armbrust, E.V., and Bowler, C. (2009). Update of the Diatom EST Database: a new tool for digital transcriptomics. *Nucleic Acids Res.* 37, D1001–D1005.
- Maheswari, U., Jabbari, K., Petit, J.-L., Porcel, B.M., Allen, A.E., Cadoret, J.-P., De Martino, A., Heijde, M., Kaas, R., La Roche, J., et al. (2010). Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol* 11, R85.
- Mallard, F., Tang, B.L., Galli, T., Tenza, D., Saint-Pol, A., Yue, X., Antony, C., Hong, W., Goud, B., and Johannes, L. (2002). Early/recycling endosomes-to-TGN transport involves two SNARE complexes and a Rab6 isoform. *J. Cell Biol.* 156, 653–664.
- Malphettes, L., Freyvert, Y., Chang, J., Liu, P.-Q., Chan, E., Miller, J.C., Zhou, Z., Nguyen, T., Tsai, C., Snowden, A.W., et al. (2010). Highly efficient deletion of FUT8 in CHO cell lines using zinc-finger nucleases yields cells that produce completely nonfucosylated antibodies. *Biotechnol. Bioeng.* 106, 774–783.
- Mann, D.G. (1993). Patterns of sexual reproduction in diatoms. In *Twelfth International Diatom Symposium*, (Springer), pp. 11–20.
- Mann, D.G. (1999). The species concept in diatoms. *Phycologia* 38, 437–495.
- Mann, D.G., and Vanormelingen, P. (2013). An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species. *J. Eukaryot. Microbiol.* 60, 414–420.

- Manton, I. (1966). Observations on the fine structure of the male gamete of the marine centric diatom *Lithodesmium undulatum*. *J. R. Microsc. Soc.* 85, 119–134.
- Marek, K.W., Vijay, I.K., and Marth, J.D. (1999). A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality. *Glycobiology* 9, 1263–1271.
- Martin-Jezequel, V., Hildebrand, M., and Brzezinski, M.A. (2000). Silicon Metabolism in Diatoms: Implications for Growth. *J. Phycol.* 36, 821–840.
- Mast, S.W., and Moremen, K.W. (2006). Family 47 α -Mannosidases in *N*-Glycan Processing. In *Methods in Enzymology*, (Elsevier), pp. 31–46.
- Masuda, K., Kubota, T., Kaneko, E., Iida, S., Wakitani, M., Kobayashi-Natsume, Y., Kubota, A., Shitara, K., and Nakamura, K. (2007). Enhanced binding affinity for Fc γ RIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. *Mol. Immunol.* 44, 3122–3131.
- Mata, T.M., Martins, A.A., and Caetano, N.S. (2010). Microalgae for biodiesel production and other applications: A review. *Renew. Sustain. Energy Rev.* 14, 217–232.
- Mathieu-Rivet, E., Scholz, M., Arias, C., Dardelle, F., Schulze, S., Mauff, F.L., Teo, G., Hochmal, A.K., Blanco-Rivero, A., Loutelier-Bourhis, C., et al. (2013). Exploring the *N*-glycosylation Pathway in *Chlamydomonas reinhardtii* Unravels Novel Complex Structures. *Mol. Cell. Proteomics* 12, 3160–3183.
- Mathieu-Rivet, E., Kiefer-Meyer, M.-C., Vanier, G., Ovide, C., Burel, C., Lerouge, P., and Bardor, M. (2014). Protein *N*-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front. Plant Sci.* 5.
- Maumus, F., Allen, A.E., Mhiri, C., Hu, H., Jabbari, K., Vardi, A., Grandbastien, M.-A., and Bowler, C. (2009). Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* 10, 624.
- McConville, M.J., Wetherbee, R., and Bacic, A. (1999). Subcellular location and composition of the wall and secreted extracellular sulphated polysaccharides/proteoglycans of the diatom *Stauroneis amphioxys* Gregory. *Protoplasma* 206, 188–200.
- Medlin, L.K. (2000). A review of the evolution of the diatoms - a total approach using molecules, morphology and geology. Polish Academy of Sciences, Krakow, Poland.
- Melo, N.S., Nimtz, M., Conradt, H.S., Fevereiro, P.S., and Costa, J. (1997). Identification of the human Lewis(a) carbohydrate motif in a secretory peroxidase from a plant cell suspension culture (*Vaccinium myrtillus* L.). *FEBS Lett.* 415, 186–191.
- Metzler, M., Gertz, A., Sarkar, M., Schachter, H., Schrader, J.W., and Marth, J.D. (1994). Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. *EMBO J.* 13, 2056–2065.
- Meusser, B., Hirsch, C., Jarosch, E., and Sommer, T. (2005). ERAD: the long road to destruction. *Nat. Cell Biol.* 7, 766–772.

- Meyer, H.A., Grau, H., Kraft, R., Kostka, S., Prehn, S., Kalies, K.U., and Hartmann, E. (2000). Mammalian Sec61 is associated with Sec62 and Sec63. *J. Biol. Chem.* 275, 14550–14557.
- Miyahara, M., Aoi, M., Inoue-Kashino, N., Kashino, Y., and Ifuku, K. (2013). Highly Efficient Transformation of the Diatom *Phaeodactylum tricorutum* by Multi-Pulse Electroporation. *Biosci. Biotechnol. Biochem.*
- Miyoshi, E., Noda, K., Yamaguchi, Y., Inoue, S., Ikeda, Y., Wang, W., Ko, J.H., Uozumi, N., Li, W., and Taniguchi, N. (1999). The alpha1-6-fucosyltransferase gene and its biological significance. *Biochim. Biophys. Acta* 1473, 9–20.
- Miyoshi, E., Moriwaki, K., and Nakagawa, T. (2007). Biological Function of Fucosylation in Cancer Biology. *J. Biochem. (Tokyo)* 143, 725–729.
- Moeys, S., Frenkel, J., Lembke, C., Gillard, J.T.F., Devos, V., Van den Berge, K., Bouillon, B., Huysman, M.J.J., De Decker, S., Scharf, J., et al. (2016). A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom *Seminavis robusta*. *Sci. Rep.* 6, 19252.
- Mohorko, E., Glockshuber, R., and Aebi, M. (2011). Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J. Inherit. Metab. Dis.* 34, 869–878.
- Molinari, M., Eriksson, K.K., Calanca, V., Galli, C., Cresswell, P., Michalak, M., and Helenius, A. (2004). Contrasting functions of calreticulin and calnexin in glycoprotein folding and ER quality control. *Mol. Cell* 13, 125–135.
- Mollicone, R., Moore, S.E.H., Bovin, N., Garcia-Rosasco, M., Candelier, J.-J., Martinez-Duncker, I., and Oriol, R. (2009). Activity, Splice Variants, Conserved Peptide Motifs, and Phylogeny of Two New 1,3-Fucosyltransferase Families (FUT10 and FUT11). *J. Biol. Chem.* 284, 4723–4738.
- Montsant, A., Jabbari, K., Maheswari, U., and Bowler, C. (2005). Comparative Genomics of the Pennate Diatom *Phaeodactylum tricorutum*. *Plant Physiol.* 137, 500–513.
- Moore, I., and Murphy, A. (2009). Validating the Location of Fluorescent Protein Fusions in the Endomembrane System. *The Plant Cell.* 21, 1632–1636.
- Moreno, V.J., de Moreno, J.E.A., and Brenner, R.R. (1979). Biosynthesis of unsaturated fatty acids in the diatom *Phaeodactylum tricorutum*. *Lipids* 14, 15–19.
- Morsomme, P., Prescianotto-Baschong, C., and Riezman, H. (2003). The ER v-SNAREs are required for GPI-anchored protein sorting from other secretory proteins upon exit from the ER. *J. Cell Biol.* 162, 403–412.
- Moustafa, A., Beszteri, B., Maier, U.G., Bowler, C., Valentin, K., and Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324, 1724–1726.
- Murray, B.W., Wittmann, V., Burkart, M.D., Hung, S.-C., and Wong, C.-H. (1997). Mechanism of Human α -1,3-Fucosyltransferase V: Glycosidic Cleavage Occurs Prior to Nucleophilic Attack [†]. *Biochemistry (Mosc.)* 36, 823–831.
- Mus, F., Toussaint, J.-P., Cooksey, K.E., Fields, M.W., Gerlach, R., Peyton, B.M., and Carlson, R.P. (2013). Physiological and molecular analysis of carbon source supplementation and pH

stress-induced lipid accumulation in the marine diatom *Phaeodactylum tricorutum*. *Appl. Microbiol. Biotechnol.* 97, 3625–3642.

Nagai, S., and Imai, I. (1998). Enumeration of bacteria in seawater and sediment from the Seto Inland Sea of Japan that promote sperm formation in *Coscinodiscus wailesii* (Bacillariophyceae). *Phycologia* 37, 363–368.

Nesvizhskii, A.I. (2014). Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* 11, 1114–1125.

Nguema-Ona, E., Vicré-Gibouin, M., Gotté, M., Plancot, B., Lerouge, P., Bardor, M., and Driouich, A. (2014). Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Front. Plant Sci.* 5.

Nilsson, I.M., and Von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* 268, 5798–5801.

Niu, Y.-F., Yang, Z.-K., Zhang, M.-H., Zhu, C.-C., Yang, W.-D., Liu, J.-S., and Li, H.-Y. (2012). Transformation of diatom *Phaeodactylum tricorutum* by electroporation and establishment of inducible selection marker. *BioTechniques*.

Noffz, C., Keppler-Ross, S., and Dean, N. (2009). Hetero-oligomeric interactions between early glycosyltransferases of the dolichol cycle. *Glycobiology* 19, 472–478.

Nolan, T., Hands, R.E., and Bustin, S.A. (2006). Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* 1, 1559–1582.

Norden-Krichmar, T.M., Allen, A.E., Gaasterland, T., and Hildebrand, M. (2011). Characterization of the Small RNA Transcriptome of the Diatom, *Thalassiosira pseudonana*. *PLoS ONE* 6, e22870.

Nymark, M., Valle, K.C., Brembu, T., Hancke, K., Winge, P., Andresen, K., Johnsen, G., and Bones, A.M. (2009). An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricorutum*. *PloS One* 4, e7743.

Olivari, S., Cali, T., Salo, K.E., Paganetti, P., Ruddock, L.W., and Molinari, M. (2006). EDEM1 regulates ER-associated degradation by accelerating de-mannosylation of folding-defective polypeptides and by inhibiting their covalent aggregation. *Biochem. Biophys. Res. Commun.* 349, 1278–1284.

O'Reilly, M.K., Zhang, G., and Imperiali, B. (2006). In Vitro Evidence for the Dual Function of Alg2 and Alg11: Essential Mannosyltransferases in N-Linked Glycoprotein Biosynthesis †. *Biochemistry (Mosc.)* 45, 9593–9603.

Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L., and Breton, C. (1999). Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology* 9, 323–334.

Orlean, P., Albright, C., and Robbins, P.W. (1988). Cloning and sequencing of the yeast gene for dolichol phosphate mannose synthase, an essential protein. *J. Biol. Chem.* 263, 17499–17507.

- Pagny, S., Bouissonnie, F., Sarkar, M., Follet-Gueye, M.L., Driouich, A., Schachter, H., Faye, L., and Gomord, V. (2003). Structural requirements for *Arabidopsis* β 1, 2-xylosyltransferase activity and targeting to the Golgi. *Plant J.* *33*, 189–203.
- Palma, A.S., Morais, V.A., Coelho, A.V., and Costa, J. (2004). Effect of the manganese ion on human α 3/4 fucosyltransferase III activity. *Biometals* *17*, 35–43.
- Paolacci, A.R., Tanzarella, O.A., Porceddu, E., and Ciaffi, M. (2009). Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol. Biol.* *10*, 11.
- Parkinson, J., and Gordon, R. (1999). Beyond micromachining: the potential of diatoms. *Trends Biotechnol.* *17*, 190–196.
- Parlati, F., McNew, J.A., Fukuda, R., Miller, R., Söllner, T.H., and Rothman, J.E. (2000). Topological restriction of SNARE-dependent membrane fusion. *Nature* *407*, 194–198.
- Parodi, A.J. (2000). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem. J.* *348*, 1–13.
- Paschinger, K., Rendić, D., Lochnit, G., Jantsch, V., and Wilson, I.B.H. (2004). Molecular Basis of Anti-horseradish Peroxidase Staining in *Caenorhabditis elegans*. *J. Biol. Chem.* *279*, 49588–49598.
- Paulson, J.C. (1989). Glycoproteins: what are the sugar chains for ? *Trends Biochem. Sci.* *14*, 272–276.
- Petrescu, A.-J., Milac, A.-L., Petrescu, S.M., Dwek, R.A., and Wormald, M.R. (2003). Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* *14*, 103–114.
- Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* *29*, e45.
- Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* *26*, 509–515.
- Pickett-Heaps, J.D., Schmid, A.-M.M., and Tippit, D.H. (1984). Cell division in diatoms: A translation of part of Robert Lauterborn's treatise of 1896 with some modern confirmatory observations. *Protoplasma* *120*, 132–154.
- Pieren, M., Galli, C., Denzel, A., and Molinari, M. (2005). The Use of Calnexin and Calreticulin by Cellular and Viral Glycoproteins. *J. Biol. Chem.* *280*, 28265–28271.
- Poll, W.H. van de, Vrieling, E.G., and Gieskes, W.W.C. (1999). Location and expression of frustulins in the pennate diatoms *Cylindrotheca fusiformis*, *Navicula pelliculosa*, and *Navicula salinarum* (Bacillariophyceae). *J. Phycol.* *35*, 1044–1053.
- Prihoda, J., Tanaka, A., de Paula, W.B.M., Allen, J.F., Tirichine, L., and Bowler, C. (2012). Chloroplast-mitochondria cross-talk in diatoms. *J. Exp. Bot.* *63*, 1543–1557.

- Pucic, M., Knezevic, A., Vidic, J., Adamczyk, B., Novokmet, M., Polasek, O., Gornik, O., Supraha-Goreta, S., Wormald, M.R., Redzic, I., et al. (2011). High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations. *Mol. Cell. Proteomics* 10, M111.010090-M111.010090.
- Reichardt, I., Stierhof, Y.-D., Mayer, U., Richter, S., Schwarz, H., Schumacher, K., and Jürgens, G. (2007). Plant Cytokinesis Requires De Novo Secretory Trafficking but Not Endocytosis. *Curr. Biol.* 17, 2047–2053.
- Reiss, G., te Heesen, S., Zimmerman, J., Robbins, P., and Aebi, M. (1996). Isolation of the ALG6 locus of *Saccharomyces cerevisiae* required for glucosylation in the N-linked glycosylation pathway. *Glycobiology* 6, 493–498.
- Ritter, C., and Helenius, A. (2000). Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose: glycoprotein glucosyltransferase. *Nat. Struct. Mol. Biol.* 7, 278–280.
- Robinson, M.D., and Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinforma. Oxf. Engl.* 23, 2881–2887.
- Rogato, A., Richard, H., Sarazin, A., Voss, B., Navarro, S.C., Champeimont, R., Navarro, L., Carbone, A., Hess, W.R., and Falciatore, A. (2014). The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricorutum*. *BMC Genomics* 15, 1.
- Round, F.E., Crawford, R.M., and Mann, D.G. (1990). *The Diatoms: biology & morphology of the genera* (Cambridge [England] ; New York: Cambridge University Press).
- Rush, J.S., and Waechter, C.J. (1995). Transmembrane movement of a water-soluble analogue of mannosylphosphoryldolichol is mediated by an endoplasmic reticulum protein. *J. Cell Biol.* 130, 529–536.
- Rush, J.S., van Leyen, K., Ouerfelli, O., Wolucka, B., and Waechter, C.J. (1998). Transbilayer movement of Glc-P-dolichol and its function as a glucosyl donor: protein-mediated transport of a water-soluble analog into sealed ER vesicles from pig brain. *Glycobiology* 8, 1195–1205.
- Rushforth, S.R., Johansen, J.R., and Sorensen, D.L. (1988). Occurrence of *Phaeodactylum tricorutum* in the Great Salt Lake, Utah, USA. *Gt. Basin Nat.* 48, 324–326.
- Saint-Jore-Dupas, C., Nebenführ, A., Boulaflous, A., Follet-Gueye, M.-L., Plasson, C., Hawes, C., Driouich, A., Faye, L., and Gomord, V. (2006). Plant N-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. *Plant Cell Online* 18, 3182–3200.
- Sanyal, S., and Menon, A.K. (2010). Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase. *Proc. Natl. Acad. Sci.* 107, 11289–11294.
- Sato, S., Beakes, G., Idei, M., Nagumo, T., and Mann, D.G. (2011). Novel Sex Cells and Evidence for Sex Pheromones in Diatoms. *PLoS ONE* 6, e26923.

- Satoh, T., Suzuki, K., Yamaguchi, T., and Kato, K. (2014). Structural Basis for Disparate Sugar-Binding Specificities in the Homologous Cargo Receptors ERGIC-53 and VIP36. *PLoS ONE* *9*, e87963.
- Sawaguchi, A., Ide, S., Goto, Y., Kawano, J.-I., Oinuma, T., and Sugauma, T. (2001). A simple contrast enhancement by potassium permanganate oxidation for Lowicryl K4M ultrathin sections prepared by high pressure freezing/freeze substitution. *J. Microsc.* *201*, 77–83.
- Scanlan, C.N., Burton, D.R., and Dwek, R.A. (2008). Making autoantibodies safe. *Proc. Natl. Acad. Sci.* *105*, 4081–4082.
- Schaub, B.E., Berger, E.G., and Rohrer, J. (2008). Dissection of a novel molecular determinant mediating Golgi to trans-Golgi network transition. *Cell. Mol. Life Sci.* *65*, 3677–3687.
- Schiller, B., Hykollari, A., Yan, S., Paschinger, K., and Wilson, I.B.H. (2012). Complicated *N*-linked glycans in simple organisms. *Biol. Chem.* *393*.
- Schoberer, J., and Strasser, R. (2011). Sub-Compartmental Organization of Golgi-Resident *N*-Glycan Processing Enzymes in Plants. *Mol. Plant* *4*, 220–228.
- Schoberer, J., Liebinger, E., Botchway, S.W., Strasser, R., and Hawes, C. (2013). Time-Resolved Fluorescence Imaging Reveals Differential Interactions of *N*-Glycan Processing Enzymes across the Golgi Stack in Planta. *Plant Physiol.* *161*, 1737–1754.
- Schrag, J.D., Bergeron, J.J., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y., and Cygler, M. (2001). The structure of calnexin, an ER chaperone involved in quality control of protein folding. *Mol. Cell* *8*, 633–644.
- Schwarz, F., and Aebi, M. (2011). Mechanisms and principles of *N*-linked protein glycosylation. *Curr. Opin. Struct. Biol.* *21*, 576–582.
- Shailubhai, K., Pukazhenti, B.S., Saxena, E.S., Varma, G.M., and Vijay, I.K. (1991). Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. *J. Biol. Chem.* *266*, 16587–16593.
- Sharpe, H.J., Stevens, T.J., and Munro, S. (2010). A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties. *Cell* *142*, 158–169.
- Shental-Bechor, D., and Levy, Y. (2008). Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci.* *105*, 8256–8261.
- Shinkawa, T., Nakamura, K., Yamane, N., Shoji-Hosaka, E., Kanda, Y., Sakurada, M., Uchida, K., Anazawa, H., Satoh, M., Yamasaki, M., et al. (2003). The Absence of Fucose but Not the Presence of Galactose or Bisecting *N*-Acetylglucosamine of Human IgG1 Complex-type Oligosaccharides Shows the Critical Role of Enhancing Antibody-dependent Cellular Cytotoxicity. *J. Biol. Chem.* *278*, 3466–3473.
- Shrimal, S., Ng, B.G., Losfeld, M.-E., Gilmore, R., and Freeze, H.H. (2013). Mutations in STT3A and STT3B cause two congenital disorders of glycosylation. *Hum. Mol. Genet.* *22*, 4638–4645.

- Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., Manfredonia, A., Falciatore, A., and Bowler, C. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* 406, 23–35.
- Singh, S., and Mittal, A. (2016). Transmembrane Domain Lengths Serve as Signatures of Organismal Complexity and Viral Transport Mechanisms. *Sci. Rep.* 6, 22352.
- Snider, M.D., and Rogers, O.C. (1984). Transmembrane movement of oligosaccharide-lipids during glycoprotein synthesis. *Cell* 36, 753–761.
- Sousa, M., and Parodi, A.J. (1995). The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc: glycoprotein glucosyltransferase. *EMBO J.* 14, 4196.
- Stagljar, I., Te Heesen, S., and Aebi, M. (1994). New phenotype of mutations deficient in glucosylation of the lipid-linked oligosaccharide: cloning of the ALG8 locus. *Proc. Natl. Acad. Sci.* 91, 5977–5981.
- Stanley, M.S., and Callow, J.A. (2007). Whole cell adhesion strength of morphotypes and isolates of *Phaeodactylum tricornutum* (Bacillariophyceae). *Eur. J. Phycol.* 42, 191–197.
- Staudacher, E., and März, L. (1998). Strict order of (Fuc to Asn-linked GlcNAc) fucosyltransferases forming core-difucosylated structures. *Glycoconj. J.* 15, 355–360.
- Strasser, R. (2016). Plant protein glycosylation. *Glycobiology.* 26, 9, 926-939.
- Strasser, R., Mucha, J., Schwihla, H., Altmann, F., Glössl, J., and Steinkellner, H. (1999). Molecular cloning and characterization of cDNA coding for β 1, 2N-acetylglucosaminyltransferase I (GlcNAc-TI) from *Nicotiana tabacum*. *Glycobiology* 9, 779–785.
- Strasser, R., Schoberer, J., Jin, C., Glössl, J., Mach, L., and Steinkellner, H. (2006). Molecular cloning and characterization of *Arabidopsis thaliana* Golgi α -mannosidase II, a key enzyme in the formation of complex N-glycans in plants. *Plant J.* 45, 789–803.
- Strasser, R., Bondili, J.S., Schoberer, J., Svoboda, B., Liebminger, E., Glossl, J., Altmann, F., Steinkellner, H., and Mach, L. (2007). Enzymatic Properties and Subcellular Localization of *Arabidopsis* beta-N-Acetylhexosaminidases. *Plant Physiol.* 145, 5–16.
- Studer, D., Graber, W., Al-Amoudi, A., and Eggli, P. (2001). A new approach for cryofixation by high-pressure freezing. *J. Microsc.* 203, 285–294.
- Sturm, A., Johnson, K.D., Szumilo, T., Elbein, A.D., and Chrispeels, M.J. (1987). Subcellular localization of glycosidases and glycosyltransferases involved in the processing of N-linked oligosaccharides. *Plant Physiol.* 85, 741–745.
- Sumper, M., and Kröger, N. (2004). Silica formation in diatoms: the function of long-chain polyamines and silaffins. *J. Mater. Chem.* 14, 2059.
- Sun, H.-Y., Lin, S.-W., Ko, T.-P., Pan, J.-F., Liu, C.-L., Lin, C.-N., Wang, A.H.-J., and Lin, C.-H. (2007). Structure and Mechanism of *Helicobacter pylori* Fucosyltransferase: A BASIS FOR LIPOPOLYSACCHARIDE VARIATION AND INHIBITOR DESIGN. *J. Biol. Chem.* 282, 9973–9982.

- Szumilo, T., Kaushal, G.P., Hori, H., and Elbein, A.D. (1986). Purification and properties of a glycoprotein processing α -mannosidase from mung bean seedlings. *Plant Physiol.* *81*, 383–389.
- Tai, V.W.-F., and Imperiali, B. (2001). Substrate Specificity of the Glycosyl Donor for Oligosaccharyl Transferase. *J. Org. Chem.* *66*, 6217–6228.
- Takahashi, E., Ledauphin, J., Goux, D., and Orvain, F. (2009). Optimising extraction of extracellular polymeric substances (EPS) from benthic diatoms: comparison of the efficiency of six EPS extraction methods. *Mar. Freshw. Res.* *60*, 1201–1210.
- Tashiro, R.M., Philips, J.G., and Winefield, C.S. (2016). Identification of suitable grapevine reference genes for qRT-PCR derived from heterologous species. *Mol. Genet. Genomics MGG* *291*, 483–492.
- Tessier, D.C., Dignard, D., Zapun, A., Radomska-Pandya, A., Parodi, A.J., Bergeron, J.J.M., and Thomas, D.Y. (2000). Cloning and characterization of mammalian UDP-glucose glycoprotein: glucosyltransferase and the development of a specific substrate for this enzyme. *Glycobiology* *10*, 403–412.
- Tesson, B., Gaillard, C., and Martin-Jézéquel, V. (2009). Insights into the polymorphism of the diatom *Phaeodactylum tricornutum* Bohlin. *Bot. Mar.* *52*.
- Tezuka, K., Hayashi, M., Ishihara, H., Akazawa, T., and Takahashi, N. (1992). Studies on synthetic pathway of xylose-containing *N*-linked oligosaccharides deduced from substrate specificities of the processing enzymes in sycamore cells (*Acer pseudoplatanus L.*). *Eur. J. Biochem.* *203*, 401–413.
- Thellin, O., ElMoualij, B., Heinen, E., and Zorzi, W. (2009). A decade of improvements in quantification of gene expression and internal standard selection. *Biotechnol. Adv.* *27*, 323–333.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinforma. Oxf. Engl.* *25*, 1105–1111.
- Tréguer, P., Nelson, D., M., Van Bennekom, A., J., De Master, D.J., Leynaert, A., and Quéguiner, B. (1995). The silica balance in the world ocean: a reestimate. *Science* *268*, 375–379.
- Trombetta, E.S., Simons, J.F., and Helenius, A. (1996). Endoplasmic reticulum glucosidase II is composed of a catalytic subunit, conserved from yeast to mammals, and a tightly bound noncatalytic HDEL-containing subunit. *J. Biol. Chem.* *271*, 27509–27516.
- Valenzuela, J., Mazurie, A., Carlson, R.P., Gerlach, R., Cooksey, K.E., Peyton, B.M., and Fields, M.W. (2012). Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* *5*, 40.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* *3*, 1–12.

- Vanier, G., Hempel, F., Chan, P., Rodamer, M., Vaudry, D., Maier, U.G., Lerouge, P., and Bardor, M. (2015). Biochemical Characterization of Human Anti-Hepatitis B Monoclonal Antibody Produced in the Microalgae *Phaeodactylum tricornutum*. *PLOS ONE* 10, e0139282.
- Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 3, 97–130.
- Velasco, A., Hendricks, L., Moremen, K.W., Tulsiani, D.R., Touster, O., and Farquhar, M.G. (1993). Cell type-dependent variations in the subcellular distribution of alpha-mannosidase I and II. *J. Cell Biol.* 122, 39–51.
- Vella, G.J., Paulsen, H., and Schachter, H. (1984). Control of glycoprotein synthesis. IX. A terminal Man alpha 1-3Man beta 1- sequence in the substrate is the minimum requirement for UDP-N-acetyl-D-glucosamine: alpha-D-mannoside (GlcNAc to Man alpha 1-3) beta 2-N-acetylglucosaminyltransferase I. *Can. J. Biochem. Cell Biol. Rev. Can. Biochim. Biol. Cell.* 62, 409–417.
- Vembar, S.S., and Brodsky, J.L. (2008). One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.* 9, 944–957.
- Vermassen, T., Van Praet, C., Lumen, N., Decaestecker, K., Vanderschaeghe, D., Callewaert, N., Villeirs, G., Hoebeke, P., Van Belle, S., Rottey, S., et al. (2015). Urinary prostate protein glycosylation profiling as a diagnostic biomarker for prostate cancer: Prostate Protein Glycosylation in Prostate Cancer. *The Prostate* 75, 314–322.
- Volcani, B.E. (1981). Cell wall formation in diatoms: morphogenesis and biochemistry. In *Silicon and Siliceous Structures in Biological Systems*, (Springer), pp. 157–200.
- Voynow, J.A., Kaiser, R.S., Scanlin, T.F., and Glick, M.C. (1991). Purification and characterization of GDP-L-fucose-N-acetyl beta-D-glucosaminide alpha 1----6fucosyltransferase from cultured human skin fibroblasts. Requirement of a specific biantennary oligosaccharide as substrate. *J. Biol. Chem.* 266, 21572–21577.
- Vrieling, E.G., Gieskes, W.W.C., and Beelen, T.P. (1999). Silicon deposition in diatoms: control by the pH inside the silicon deposition vesicle. *J. Phycol.* 35, 548–559.
- de Vries, T., Knegt, R.M., Holmes, E.H., and Macher, B.A. (2001). Fucosyltransferases: structure/function studies. *Glycobiology* 11, 119R–128R.
- Wada, I., Imai, S. -i., Kai, M., Sakane, F., and Kanoh, H. (1995). Chaperone Function of Calreticulin When Expressed in the Endoplasmic Reticulum as the Membrane-anchored and Soluble Forms. *J. Biol. Chem.* 270, 20298–20304.
- Wang, X. (2005). Core Fucosylation Regulates Epidermal Growth Factor Receptor-mediated Intracellular Signaling. *J. Biol. Chem.* 281, 2572–2577.
- Wang, X., Weldeghiorghis, T., Zhang, G., Imperiali, B., and Prestegard, J.H. (2008). Solution Structure of Alg13: The Sugar Donor Subunit of a Yeast N-Acetylglucosamine Transferase. *Structure* 16, 965–975.
- Wang, Y., Fukuda, T., Isaji, T., Lu, J., Gu, W., Lee, H., Ohkubo, Y., Kamada, Y., Taniguchi, N., Miyoshi, E., et al. (2015). Loss of α 1,6-fucosyltransferase suppressed liver regeneration:

implication of core fucose in the regulation of growth factor receptor-mediated cellular signaling. *Sci. Rep.* 5, 8264.

Watanabe, S., Kakudo, A., Ohta, M., Mita, K., Fujiyama, K., and Inumaru, S. (2013). Molecular cloning and characterization of the α -glucosidase II from *Bombyx mori* and *Spodoptera frugiperda*. *Insect Biochem. Mol. Biol.* 43, 319–327.

Wendler, F., and Tooze, S. (2001). Syntaxin 6: the promiscuous behaviour of a SNARE protein. *Traffic Cph. Den.* 2, 606–611.

Wetherbee, R. (2002). BIOMINERALIZATION: The Diatom Glasshouse. *Science* 298, 547–547.

Whitley, P., Nilsson, I., and von Heijne, G. (1996). A Nascent Secretory Protein 5 Traverse the Ribosome/Endoplasmic Reticulum Translocase Complex as an Extended Chain. *J. Biol. Chem.* 271, 6241–6244.

Wijeyesakere, S.J., Gagnon, J.K., Arora, K., Brooks, C.L., and Raghavan, M. (2015). Regulation of calreticulin–major histocompatibility complex (MHC) class I interactions by ATP. *Proc. Natl. Acad. Sci.* 112, E5608–E5617.

Wildt, S., and Gerngross, T.U. (2005). The humanization of *N*-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* 3, 119–128.

Willis, A., Chiovitti, A., Dugdale, T.M., and Wetherbee, R. (2013). Characterization of the extracellular matrix of *Phaeodactylum tricornutum* (Bacillariophyceae): structure, composition, and adhesive characteristics. *J. Phycol.* 937–949.

Wilson, D. (1946). The triradiate and other forms of *Nitzschia closterum*.

Wilson, I.B. (2002). Glycosylation of proteins in plants and invertebrates. *Curr. Opin. Struct. Biol.* 12, 569–577.

Wilson, I.B.H. (2012). The class I α 1,2-mannosidases of *Caenorhabditis elegans*. *Glycoconj. J.* 29, 173–179.

Wilson, I.B., Harthill, J.E., Mullin, N.P., Ashford, D.A., and Altmann, F. (1998). Core α 1, 3-fucose is a key part of the epitope recognized by antibodies reacting against plant *N*-linked oligosaccharides and is present in a wide variety of plant extracts. *Glycobiology* 8, 651–661.

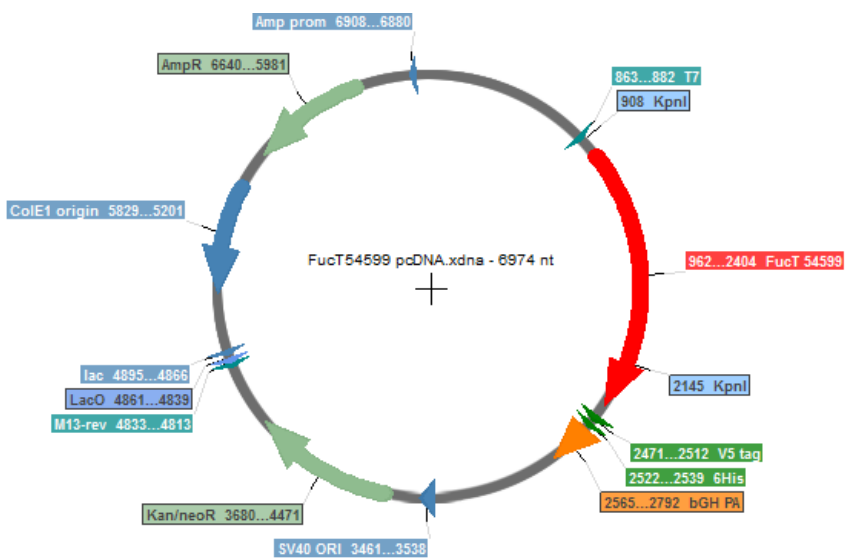
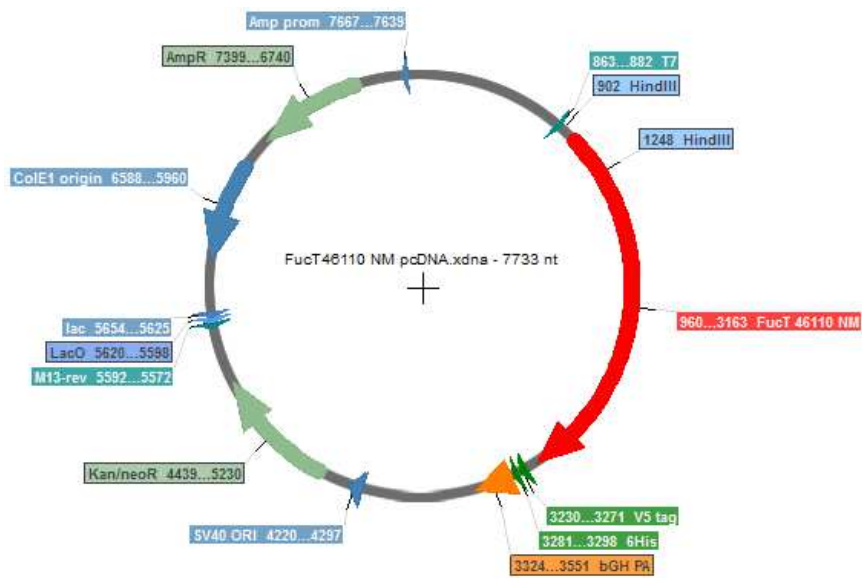
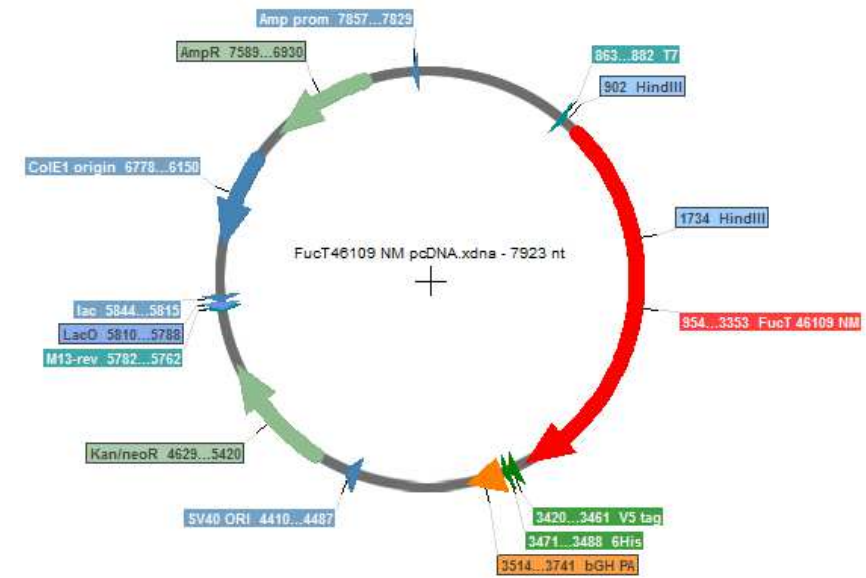
Wilson, I.B., Rendić, D., Freilinger, A., Dumić, J., Altmann, F., Mucha, J., Müller, S., and Hauser, M.-T. (2001). Cloning and expression of cDNAs encoding α 1, 3-fucosyltransferase homologues from *Arabidopsis thaliana*. *Biochim. Biophys. Acta BBA-Gen. Subj.* 1527, 88–96.

Wilson, J.R., Williams, D., and Schachter, H. (1976). The control of glycoprotein synthesis: *N*-acetylglucosamine linkage to a mannose residue as a signal for the attachment of L-fucose to the asparagine-linked *N*-acetylglucosamine residue of glycopeptide from alpha1-acid glycoprotein. *Biochem. Biophys. Res. Commun.* 72, 909–916.

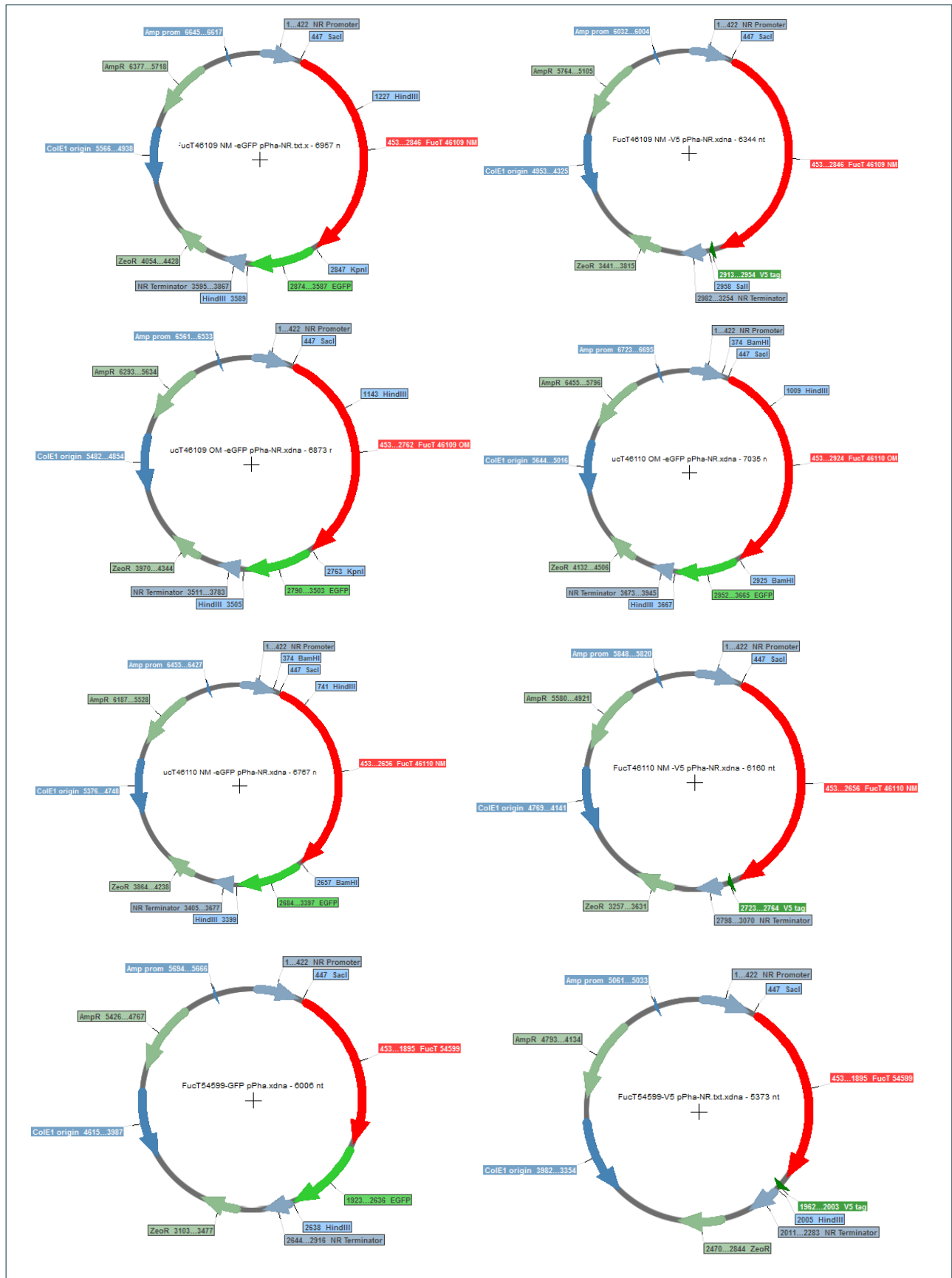
- Winckelmann, D., Bleeke, F., Thomas, B., Elle, C., and Klöck, G. (2015). Open pond cultures of indigenous algae grown on non-arable land in an arid desert using wastewater. *Int. Aquat. Res.* 7, 221–233.
- Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* 347, 1257594–1257594.
- Xiao, X., Wu, X., Ma, J., Li, P., Li, T., and Yao, Y. (2015). Systematic assessment of reference genes for RT-qPCR across plant species under salt stress and drought stress. *Acta Physiol. Plant.* 37.
- Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. (2012). miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.*
- Yan, Q., and Lennarz, W.J. (1999). Oligosaccharyltransferase: A Complex Multisubunit Enzyme of the Endoplasmic Reticulum. *Biochem. Biophys. Res. Commun.* 266, 684–689.
- Yan, Q., and Lennarz, W.J. (2002). Studies on the Function of Oligosaccharyl Transferase Subunits: Stt3p IS DIRECTLY INVOLVED IN THE GLYCOSYLATION PROCESS. *J. Biol. Chem.* 277, 47692–47700.
- Yang, Z.-K., Niu, Y.-F., Ma, Y.-H., Xue, J., Zhang, M.-H., Yang, W.-D., Liu, J.-S., Lu, S.-H., Guan, Y., and Li, H.-Y. (2013). Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnol. Biofuels* 6, 67.
- Yool, A., and Tyrrell, T. (2003). Role of diatoms in regulating the ocean's silicon cycle. *Glob. Biogeochem. Cycles* 17, n/a-n/a.
- Zhang, C., and Hu, H. (2013). High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Mar. Genomics.*
- Zhang, H., Shahbazi, M.-A., Mäkilä, E.M., da Silva, T.H., Reis, R.L., Salonen, J.J., Hirvonen, J.T., and Santos, H.A. (2013). Diatom silica microparticles for sustained release and permeation enhancement following oral delivery of prednisone and mesalamine. *Biomaterials* 34, 9210–9219.
- Zhao, P., Gu, W., Wu, S., Huang, A., He, L., Xie, X., Gao, S., Zhang, B., Niu, J., Peng Lin, A., et al. (2014). Silicon enhances the growth of *Phaeodactylum tricornutum* Bohlin under green light and low temperature. *Sci. Rep.* 4.
- Zhao, X., Yang, H., Liu, W., Duan, X., Shang, W., Xia, D., and Tong, C. (2015). Sec22 regulates endoplasmic reticulum morphology but not autophagy and is required for eye development in *Drosophila*. *J. Biol. Chem.* 290, 7943–7951.
- Zielinska, D.F., Gnad, F., Wiśniewski, J.R., and Mann, M. (2010). Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* 141, 897–907.
- Zielinska, D.F., Gnad, F., Schropp, K., Wiśniewski, J.R., and Mann, M. (2012). Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. *Mol. Cell* 46, 542–548.

Zurzolo, C., and Bowler, C. (2001). Exploring Bioinorganic Pattern Formation in Diatoms. A Story of Polarized Trafficking. *PLANT Physiol.* *127*, 1339–1345.

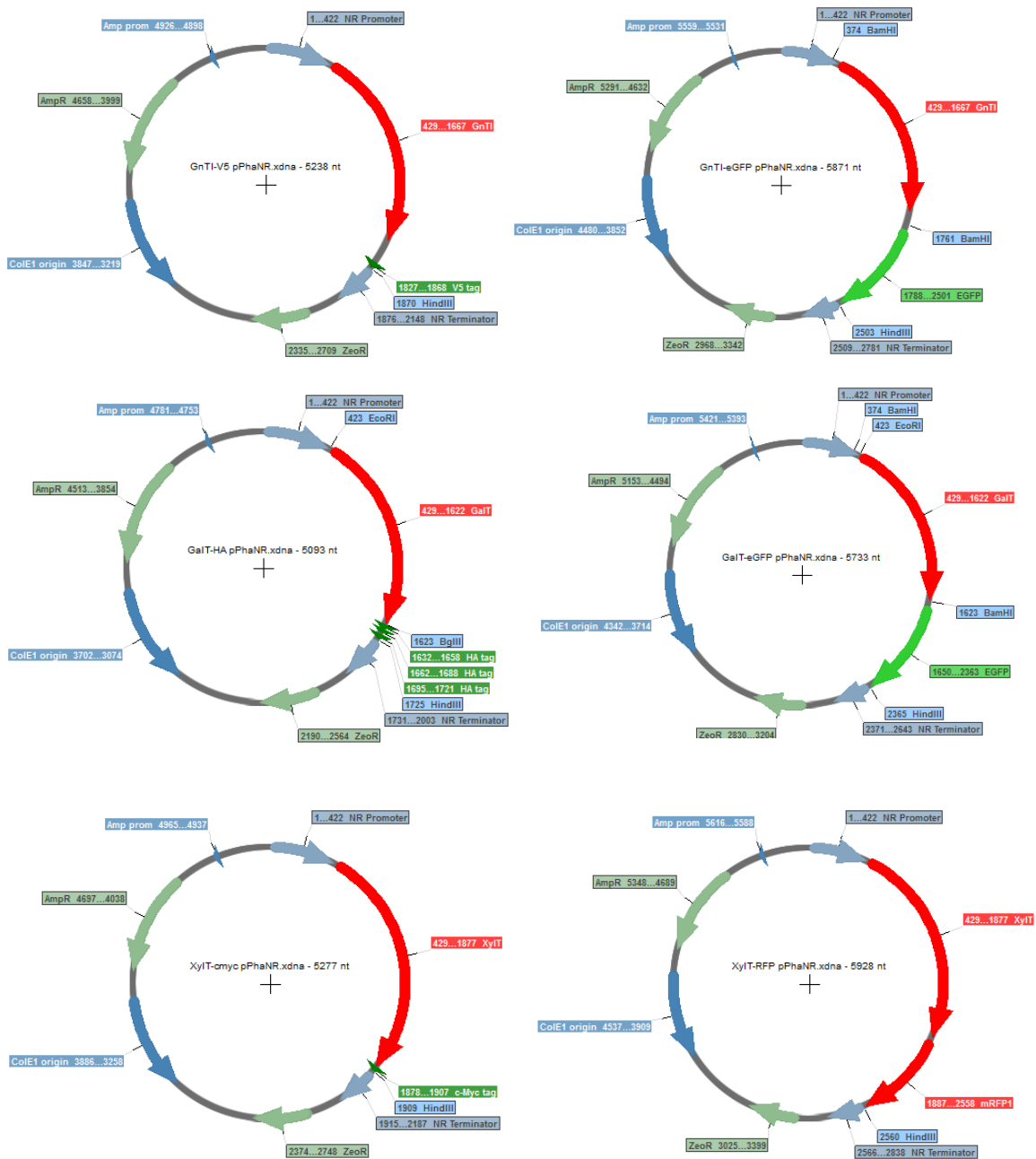
Annexe 1



Carte des constructions génétiques FucT46109NM+pcDNA, FucT46110NM+pcDNA et FucT54599+pcDNA



Carte des constructions génétiques FucT46109NM-GFP et -V5, FucT46109OM-GFP, FucT46110NM-GFP et -V5, FucT46110OM-GFP, FucT54599-GFP et -V5 dans pPha-NR



Carte des constructions génétiques GnT I-GFP et -V5, GalT-GFP et -HA, ainsi que XylIT-RFP et -cmyc dans pPha-NR

Annexe 2

Low Sample (LS) Protocol

Introduction	58
Sample Prep Workflow	60
Purify and Fragment mRNA	61
Synthesize First Strand cDNA	67
Synthesize Second Strand cDNA	70
Adenylate 3' Ends	74
Ligate Adapters	76
Enrich DNA Fragments	83
Validate Library	87
Normalize and Pool Libraries	89



Introduction

This chapter describes the TruSeq Stranded mRNA Sample Preparation low sample (LS) protocol. Illumina recommends the following kit, sample number, and protocol combinations:

Table 17 Kit and Sample Number Recommendations

Number of Samples Processed At One Time	Kit Recommended
<24	LT
24–48	LT or HT
>48	HT

Table 18 Kit and Protocol Recommendations

Kit	Number of Samples Supported	Number of Samples Processed At One Time	Protocol
LT	48	≤48	LS
		>48	HS
HT	96	≤24	LS
		>24	HS

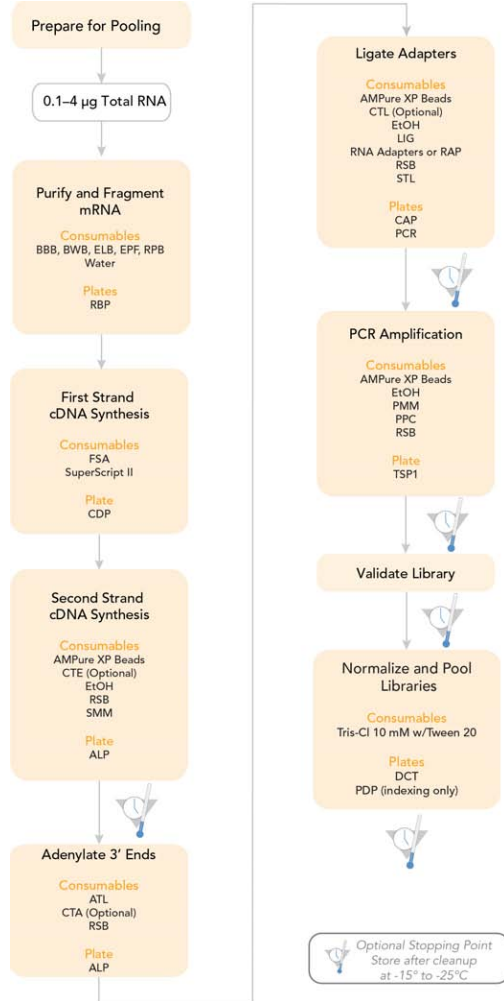
- ▶ Review *Best Practices* on page 9 before proceeding.
- ▶ Follow the protocols in the order shown, using the specified volumes and incubation parameters.
- ▶ For optimal sample tracking and quality control, fill out the Lab Tracking Form as you perform the sample preparation. For more information, see *Tracking Tools* on page 25.

- ▶ If you are pooling using adapter index tubes, record information about your samples before beginning library preparation for later use in data analysis. For more information, see *Tracking Tools* on page 25. Illumina recommends arranging samples that will be combined into a common pool in the same row. Each column should contain a common index. This will facilitate pipetting operations when dispensing indexed adapters and pooling indexed libraries later in the protocol.
- ▶ If you are pooling with the RAP, please review the planning steps in *Pooling Preparation with Adapter Plate* on page 43 before beginning library preparation.

Sample Prep Workflow

The following illustrates the processes of the TruSeq Stranded mRNA Sample Preparation LS protocol to prepare templates using 24 indexed adapter tubes or a RAP.

Figure 36 TruSeq Stranded mRNA Sample Preparation LS Workflow

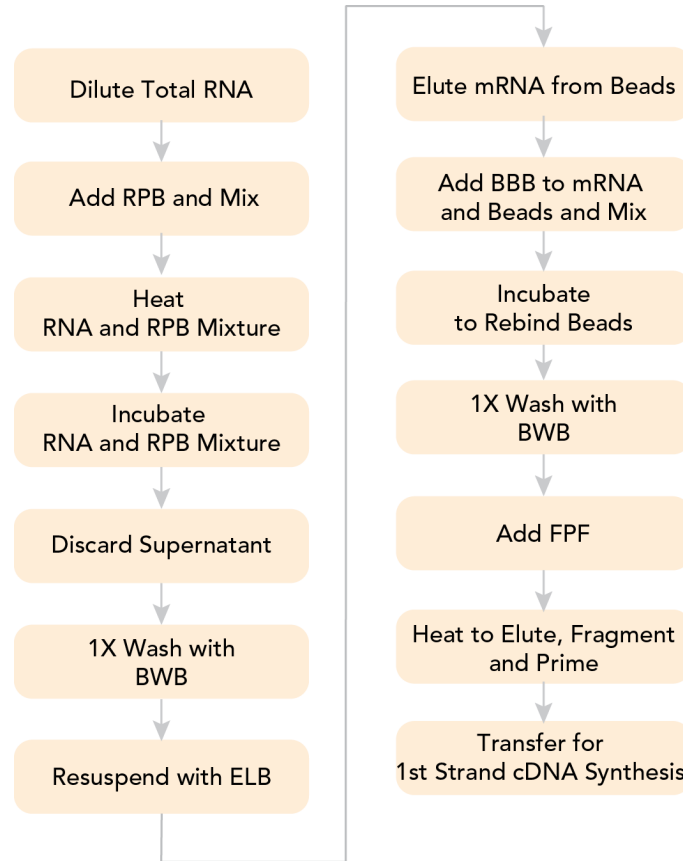


Purify and Fragment mRNA

This process purifies the poly-A containing mRNA molecules using poly-T oligo-attached magnetic beads using two rounds of purification. During the second elution of the poly-A RNA, the RNA is also fragmented and primed for cDNA synthesis.

Reference the following diagram while performing the purification procedures:

Figure 37 TruSeq Stranded mRNA Sample Prep Purification Workflow



It is important to follow this procedure exactly to be sure of reproducibility.

**NOTE**

Allow the beads to fully pellet against the magnetic stand 5 minutes. Remove the supernatant from the beads immediately while the beads are still pelleted against the magnetic stand. Do not allow the pellets to dry.

**NOTE**

Illumina recommends that you use 0.1–4 µg of total RNA and use PCR plates with a magnetic plate stand for this process. Alternatively, you can start the protocol with 10–400 ng of previously isolated mRNA. If you do so, the mRNA must be concentrated into 5 µl or less before addition to Fragment, Prime, Finish Mix. This can be done by ethanol precipitation or by concentration on a Qiagen MinElute column.

- If ethanol precipitation is used, resuspend the pellet in 18 µl Fragment, Prime, Finish Mix.
- If a Qiagen MinElute column is used, elute the mRNA with 5 µl of molecular biology grade water and add 13 µl Fragment, Prime, Finish Mix. Use of the MinElute column will result in loss of up to 50% of the mRNA due to the low elution volume.

In either case, heat the mRNA in Elute, Prime, Fragment Mix to fragment at *Incubate RFP* on page 66 in this process.

**NOTE**

For inserts larger than 120–200 bp with a median size of 150 bp, see Appendix A Alternate Fragmentation Protocols.

Consumables

Item	Quantity	Storage	Supplied By
Bead Binding Buffer (BBB)	1 tube per 48 reactions	-15° to -25°C	Illumina
Bead Washing Buffer (BWB)	1 tube per 48 reactions	-15° to -25°C	Illumina
Elution Buffer (ELB)	1 tube per 48 reactions	-15° to -25°C	Illumina
Fragment, Prime, Finish Mix (FPF)	1 tube per 48 reactions	-15° to -25°C	Illumina

Item	Quantity	Storage	Supplied By
Resuspension Buffer (RSB)	1 tube	-15° to -25°C	Illumina
RNA Purification Beads (RPB)	1 tube per 48 reactions	2° to 8°C	Illumina
RBP (RNA Bead Plate) barcode label	1 label per plate	15° to 30°C	Illumina
96-well 0.3 ml PCR plate	1	15° to 30°C	User
Microseal 'B' Adhesive Seals	3	15° to 30°C	User
RNase/DNase-free Reagent Reservoirs (if using multichannel pipettes)	6	15° to 30°C	User
RNase/DNase-free Strip Tubes and Caps (if using multichannel pipettes)	6	15° to 30°C	User

Preparation

- ▶ Remove the following from -15° to -25°C storage and thaw them at room temperature:
 - Bead Binding Buffer
 - Bead Washing Buffer
 - Elution Buffer
 - Fragment, Prime, Finish Mix
 - Resuspension Buffer



NOTE

The Resuspension Buffer can be stored at 2° to 8°C after the initial thaw.

- ▶ Remove the RNA Purification Beads tube from 2° to 8°C storage and let stand to bring to room temperature.
- ▶ Pre-program the thermal cycler with the following programs:
 - Choose the pre-heat lid option and set to 100°C
 - 65°C for 5 minutes, 4°C hold – save as **mRNA Denaturation**
 - 80°C for 2 minutes, 25°C hold – save as **mRNA Elution 1**

- 94°C for 8 minutes, 4°C hold – save as **Elution 2 - Frag - Prime**
- ▶ Set the centrifuge to 15° to 25°C, if refrigerated.
- ▶ Apply a RBP barcode label to a new 96-well 0.3 ml PCR plate.

**NOTE**

After use in this procedure, store the Bead Binding Buffer, Bead Washing Buffer, and Elution Buffer at 2° to 8°C for subsequent experiments.

Make RBP

- 1 Dilute the total RNA with nuclease-free ultra pure water to a final volume of 50 µl in the new 96-well 0.3 ml PCR plate labeled with the RBP barcode.
- 2 Vortex the room temperature RNA Purification Beads tube vigorously to completely resuspend the oligo-dT beads.
- 3 Add 50 µl of RNA Purification Beads to each well of the RBP to bind the poly-A RNA to the oligo dT magnetic beads. Gently pipette the entire volume up and down 6 times to mix thoroughly.
- 4 Seal the RBP plate with a Microseal 'B' Adhesive seal.

Incubate 1 RBP

- 1 Place the sealed RBP plate on the pre-programmed thermal cycler. Close the lid and select **mRNA Denaturation** to denature the RNA and facilitate binding of the poly-A RNA to the beads.
- 2 Remove the RBP plate from the thermal cycler when it reaches 4°C.
- 3 Place the RBP plate on the bench and incubate at room temperature for 5 minutes to allow the RNA to bind to the beads.

Wash RBP

- 1 Remove the adhesive seal from the RBP plate.
- 2 Place the RBP plate on the magnetic stand at room temperature for 5 minutes to separate the poly-A RNA bound beads from the solution.
- 3 Remove and discard all of the supernatant from each well of the RBP plate.

- 4 Remove the RBP plate from the magnetic stand.
- 5 Wash the beads by adding 200 μ l of Bead Washing Buffer in each well of the RBP plate to remove unbound RNA. Gently pipette the entire volume up and down 6 times to mix thoroughly.
- 6 Place the RBP plate on the magnetic stand at room temperature for 5 minutes.
- 7 Centrifuge the thawed Elution Buffer to 600 \times g for 5 seconds.
- 8 Remove and discard all of the supernatant from each well of the RBP plate. The supernatant contains the majority of the ribosomal and other non-messenger RNA.
- 9 Remove the RBP plate from the magnetic stand.
- 10 Add 50 μ l of Elution Buffer in each well of the RBP plate. Gently pipette the entire volume up and down 6 times to mix thoroughly.
- 11 Seal the RBP plate with a Microseal 'B' Adhesive seal.
- 12 Store the Elution Buffer tube at 4°C.

Incubate 2 RBP

- 1 Place the sealed RBP plate on the pre-programmed thermal cycler. Close the lid and select **mRNA Elution 1** to elute the mRNA from the beads. This releases both the mRNA and any contaminant rRNA that has bound the beads non-specifically.
- 2 Remove the RBP plate from the thermal cycler when it reaches 25°C.
- 3 Place the RBP plate on the bench at room temperature and remove the adhesive seal from the plate.

Make RFP

- 1 Centrifuge the thawed Bead Binding Buffer to 600 \times g for 5 seconds.
- 2 Add 50 μ l of Bead Binding Buffer to each well of the RBP plate. This allows mRNA to specifically rebind the beads, while reducing the amount of rRNA that non-specifically binds. Gently pipette the entire volume up and down 6 times to mix thoroughly.

- 3 Incubate the RBP plate at room temperature for 5 minutes and store the Bead Binding Buffer tube at 2° to 8°C.
- 4 Place the RBP plate on the magnetic stand at room temperature for 5 minutes.
- 5 Remove and discard all of the supernatant from each well of the RBP plate.
- 6 Remove the RBP plate from the magnetic stand.
- 7 Wash the beads by adding 200 µl of Bead Washing Buffer in each well of the RBP plate. Gently pipette the entire volume up and down 6 times to mix thoroughly.
- 8 Store the Bead Washing Buffer tube at 2° to 8°C.
- 9 Place the RBP plate on the magnetic stand at room temperature for 5 minutes.
- 10 Remove and discard all of the supernatant from each well of the RBP plate. The supernatant contains residual rRNA and other contaminants that were released in the first elution and did not rebind the beads.
- 11 Remove the RBP plate from the magnetic stand.
- 12 Add 19.5 µl of Fragment, Prime, Finish Mix to each well of the RBP plate. Gently pipette the entire volume up and down 6 times to mix thoroughly. The Fragment, Prime, Finish Mix contains random hexamers for RT priming and serves as the 1st strand cDNA synthesis reaction buffer.
- 13 Seal the RBP plate with a Microseal 'B' Adhesive seal.
- 14 Store the Fragment, Prime, Finish Mix tube at -15° to -25°C.

Incubate RFP

- 1 Place the sealed RBP plate on the pre-programmed thermal cycler. Close the lid and select **Elution 2 - Frag - Prime** to elute, fragment, and prime the RNA.
- 2 Remove the RBP plate from the thermal cycler when it reaches 4°C and centrifuge briefly.
- 3 Proceed immediately to *Synthesize First Strand cDNA* on page 67.

Synthesize First Strand cDNA

This process reverse transcribes the cleaved RNA fragments that were primed with random hexamers into first strand cDNA using reverse transcriptase and random primers.

Consumables

Item	Quantity	Storage	Supplied By
First Strand Synthesis Act D Mix (FSA)	1 tube per 48 reactions	-15° to -25°C	Illumina
CDP (cDNA Plate) barcode label	1 label per plate	15° to 30°C	Illumina
96-well 0.3 ml PCR plate	1	15° to 30°C	User
Microseal 'B' Adhesive Seal	1	15° to 30°C	User
RNase/DNase-free Reagent Reservoirs (if using multichannel pipettes)	1	15° to 30°C	User
RNase/DNase-free Strip Tubes and Caps (if using multichannel pipettes)	1	15° to 30°C	User
SuperScript II Reverse Transcriptase	1 tube	-15° to -25°C	User



WARNING

First Strand Synthesis Mix Act D (FSA) contains Actinomycin D, a toxin. Personal injury can occur through inhalation, ingestion, skin contact, and eye contact. Dispose of containers and any unused contents in accordance with the governmental safety standards for your region. Please refer to the product material safety data sheet (MSDS) for detailed environmental, health, and safety information. MSDSs are available for this kit on the Illumina website at <http://www.illumina.com/msds>.

Preparation

- ▶ Remove one tube of First Strand Synthesis Act D Mix from -15° to -25°C storage and thaw it at room temperature.
- ▶ Pre-program the thermal cycler with the following program and save as **Synthesize 1st Strand**:
 - Choose the pre-heat lid option and set to 100°C
 - 25°C for 10 minutes
 - 42°C for 15 minutes
 - 70°C for 15 minutes
 - Hold at 4°C
- ▶ Apply a CDP barcode label to a new 96-well 0.3 ml PCR plate.



NOTE

The First Strand Synthesis Mix Act D with SuperScript II added is stable to additional freeze-thaw cycles and can be used for subsequent experiments. If more than six freeze-thaw cycles are anticipated, divide the First Strand Synthesis Mix Act D and SuperScript II mix into smaller aliquots and store at -15° to -25°C.

Make CDP

- 1 Place the RBP plate on the magnetic stand at room temperature for 5 minutes. Do not remove the plate from the magnetic stand.
- 2 Remove the adhesive seal from the RBP plate.
- 3 Transfer 17 μ l of the supernatant (fragmented and primed mRNA) from each well of the RBP plate to the corresponding well of the new 0.3 ml PCR plate labeled with the CDP barcode.
- 4 Centrifuge the thawed First Strand Synthesis Act D Mix tube to 600 \times g for 5 seconds.
- 5 Add 50 μ l SuperScript II to the First Strand Synthesis Act D Mix tube. If you are not using the entire contents of the First Strand Synthesis Act D Mix tube, add SuperScript II at a ratio of 1 μ l SuperScript II for each 9 μ l First Strand Synthesis Act D Mix. Mix gently, but thoroughly, and centrifuge briefly.

Label the First Strand Synthesis Act D Mix tube to indicate that the SuperScript II has been added.

- 6 Add 8 μ l of First Strand Synthesis Act D Mix and SuperScript II mix to each well of the CDP plate. Gently pipette the entire volume up and down 6 times to mix thoroughly.
- 7 Seal the CDP plate with a Microseal 'B' Adhesive seal and centrifuge briefly.
- 8 Return the First Strand Synthesis Act D Mix tube to -15° to -25°C storage immediately after use.

Incubate 1 CDP

- 1 Place the sealed CDP plate on the pre-programmed thermal cycler. Close the lid and select the **Synthesize 1st Strand** program.
- 2 When the thermal cycler reaches 4°C , remove the CDP plate from the thermal cycler and proceed immediately to *Synthesize Second Strand cDNA* on page 70.


Synthesize Second Strand cDNA

This process removes the RNA template and synthesizes a replacement strand, incorporating dUTP in place of dTTP to generate ds cDNA. AMPure XP beads are used to separate the ds cDNA from the second strand reaction mix. At the end of this process, you will have blunt-ended cDNA.

Consumables

Item	Quantity	Storage	Supplied By
(Optional) End Repair Control (CTE)	1 tube per 48 reactions	-15° to -25°C	Illumina
Resuspension Buffer (RSB)	1 tube	2° to 8°C	Illumina
Second Strand Marking Master Mix (SMM)	1 tube per 48 reactions	-15° to -25°C	Illumina
ALP (Adapter Ligation Plate) barcode label	1 label per plate	15° to 30°C	Illumina
96-well 0.3 ml PCR plate	1	15° to 30°C	User
AMPure XP beads	90 µl per sample	2° to 8°C	User
Freshly Prepared 80% Ethanol (EtOH)	400 µl per sample	15° to 30°C	User
Microseal 'B' Adhesive Seals	2	15° to 30°C	User
RNase/DNase-free Reagent Reservoirs (if using multichannel pipettes)	5	15° to 30°C	User
RNase/DNase-free Strip Tubes and Caps (if using multichannel pipettes)	5	15° to 30°C	User

Preparation

- ▶ Remove the following from -15° to -25°C storage and thaw them at room temperature:
 - End Repair Control
- 

NOTE
The use of the End Repair Control is optional and it can be replaced with the same volume of Resuspension Buffer.
- Second Strand Marking Master Mix
- ▶ Remove the Resuspension Buffer from 2° to 8°C storage and bring it to room temperature.
 - ▶ Review *Handling Magnetic Beads* on page 10.
 - ▶ Remove the AMPure XP beads from storage and let stand for at least 30 minutes to bring them to room temperature.
 - ▶ Pre-heat the thermal cycler to 16°C.
 - ▶ Choose the thermal cycler pre-heat lid option and set to 30°C
 - ▶ Apply a ALP barcode label to a new 96-well 0.3 ml PCR plate.

Add SMM

- 1 Remove the adhesive seal from the CDP plate.
- 2 Do one of the following:
 - If using the in-line control reagent:
 - Centrifuge the thawed End Repair Control tube to 600 xg for 5 seconds.
 - Dilute the End Repair Control to 1/50 in Resuspension Buffer (For example, 2 µl End Repair Control + 98 µl Resuspension Buffer) before use. Discard the diluted End Repair Control after use.
 - Add 5 µl of diluted End Repair Control to each well of the CDP plate.
 - If not using the in-line control reagent, add 5 µl of Resuspension Buffer to each well of the CDP plate.
- 3 Centrifuge the thawed Second Strand Marking Master Mix to 600 xg for 5 seconds.
- 4 Add 20 µl of thawed Second Strand Marking Master Mix to each well of the CDP plate. Gently pipette the entire volume up and down 6 times to mix thoroughly.
- 5 Seal the CDP plate with a Microseal 'B' Adhesive seal.

- 6 Return the Second Strand Marking Master Mix tube to -15° to -25°C storage after use.

Incubate 2 CDP

- 1 Place the sealed CDP plate on the pre-heated thermal cycler. Close the lid and incubate at 16°C for 1 hour.
- 2 Remove the CDP plate from the thermal cycler, remove the adhesive seal, and let stand to bring the plate to room temperature.

Purify CDP



NOTE

Before performing clean up, review *Handling Magnetic Beads* on page 10 when working with AMPure XP Beads.

- 1 Vortex the AMPure XP beads until they are well dispersed, then add $90\ \mu\text{l}$ of well-mixed AMPure XP beads to each well of the CDP plate containing $50\ \mu\text{l}$ of ds cDNA. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 2 Incubate the CDP plate at room temperature for 15 minutes.
- 3 Place the CDP plate on the magnetic stand at room temperature, for 5 minutes to make sure that all of the beads are bound to the side of the wells.
- 4 Remove and discard $135\ \mu\text{l}$ of the supernatant from each well of the CDP plate.



NOTE

Leave the CDP plate on the magnetic stand while performing the following 80% EtOH wash steps (5–7).

- 5 With the CDP plate remaining on the magnetic stand, add $200\ \mu\text{l}$ of freshly prepared 80% EtOH to each well without disturbing the beads.
- 6 Incubate the CDP plate at room temperature for 30 seconds, then remove and discard all of the supernatant from each well.
- 7 Repeat steps 5 and 6 once for a total of two 80% EtOH washes.
- 8 Let the plate stand at room temperature for 15 minutes to dry and then remove the CDP plate from the magnetic stand.

- 9 Centrifuge the thawed, room temperature Resuspension Buffer to 600 \times g for 5 seconds.
- 10 Add 17.5 μ l Resuspension Buffer to each well of the CDP plate. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 11 Incubate the CDP plate at room temperature for 2 minutes.
- 12 Place the CDP plate on the magnetic stand at room temperature for 5 minutes.
- 13 Transfer 15 μ l of the supernatant (ds cDNA) from the CDP plate to the new 96-well 0.3 ml PCR plate labeled with the ALP barcode.

**SAFESTOPPING POINT**

If you do not plan to proceed to *Adenylate 3' Ends* on page 74 immediately, the protocol can be safely stopped here. If you are stopping, seal the ALP plate with a Microseal 'B' adhesive seal and store at -15° to -25°C for up to seven days.

Adenylate 3' Ends

A single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera (concatenated template) formation.

Consumables

Item	Quantity	Storage	Supplied By
(Optional) A-Tailing Control (CTA)	1 tube per 48 reactions	-15° to -25°C	Illumina
A-Tailing Mix (ATL)	1 tube per 48 reactions	-15° to -25°C	Illumina
Resuspension Buffer (RSB)	1 tube	2° to 8°C	Illumina
Microseal 'B' Adhesive Seal	1	15° to 30°C	User
RNase/DNase-free Reagent Reservoirs (if using multichannel pipettes)	3	15° to 30°C	User
RNase/DNase-free Strip Tubes and Caps (if using multichannel pipettes)	3	15° to 30°C	User

Preparation

- ▶ Remove the following from -15° to -25°C storage and thaw them at room temperature:

- A-Tailing Control



NOTE

The use of the A-Tailing Control is optional and it can be replaced with the same volume of Resuspension Buffer.

- A-Tailing Mix

- ▶ Remove the Resuspension Buffer from 2° to 8°C storage and bring it to room temperature.
- ▶ Remove the ALP plate from -15° to -25°C storage, if it was stored at the conclusion of *Purify CDP* on page 72 and let stand to thaw at room temperature.
 - Centrifuge the thawed ALP plate to 280 xg for 1 minute.
 - Remove the adhesive seal from the ALP plate.
- ▶ Pre-program the thermal cycler with the following program and save as **ATAIL70**:
 - Choose the pre-heat lid option and set to 100°C
 - 37°C for 30 minutes
 - 70°C for 5 minutes
 - Hold at 4°C

Add ATL

- 1 Do one of the following:
 - If using the in-line control reagent:
 - Centrifuge the thawed A-Tailing Control tube to 600 xg for 5 seconds.
 - Dilute the A-Tailing Control to 1/100 in Resuspension Buffer (For example, 1 µl A-Tailing Control + 99 µl Resuspension Buffer) before use. Discard the diluted A-Tailing Control after use.
 - Add 2.5 µl of diluted A-Tailing Control to each well of the ALP plate.
 - If not using the in-line control reagent, add 2.5 µl of Resuspension Buffer to each well of the ALP plate.
- 2 Add 12.5 µl of thawed A-Tailing Mix to each well of the ALP plate. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 3 Seal the ALP plate with a Microseal 'B' adhesive seal.

Incubate 1 ALP

- 1 Place the sealed ALP plate on the pre-programmed thermal cycler. Close the lid and select **ATAIL70**.
- 2 When the thermal cycler temperature is 4°C, remove the ALP plate from the thermal cycler, then proceed immediately to *Ligate Adapters* on page 76.

Ligate Adapters

This process ligates multiple indexing adapters to the ends of the ds cDNA, preparing them for hybridization onto a flow cell.

Consumables

Item	Quantity	Storage	Supplied By
(Optional) Ligation Control (CTL)	1 tube per 48 reactions	-15° to -25°C	Illumina
Choose from the following depending on the kit you are using: <ul style="list-style-type: none"> • TruSeq Stranded mRNA LT Sample Prep Kit contents: <ul style="list-style-type: none"> • RNA Adapter Indices (AR001–AR016, AR018–AR023, AR025, AR027) • TruSeq Stranded mRNA HT Sample Prep Kit contents: <ul style="list-style-type: none"> • RAP (RNA Adapter Plate) 	1 tube per column of 8 reactions, of each indices being used or 1 RAP	-15° to -25°C	Illumina
Ligation Mix (LIG)	1 tube per 48 reactions	-15° to -25°C	Illumina
Resuspension Buffer (RSB)	1 tube	2° to 8°C	Illumina
Stop Ligation Buffer (STL)	1 tube per 48 reactions	-15° to -25°C	Illumina
CAP (Clean Up ALP Plate) barcode label	1 label per plate	15° to 30°C	Illumina
PCR (Polymerase Chain Reaction) barcode label	1 label per plate	15° to 30°C	Illumina

Item	Quantity	Storage	Supplied By
RAP (RNA Adapter Plate) barcode label (if using the HT kit)	1 label per plate	15° to 30°C	Illumina
96-well 0.3 ml PCR plates	2	15° to 30°C	User
AMPure XP beads	92 µl per sample	2° to 8°C	User
Freshly Prepared 80% Ethanol (EtOH)	800 µl per sample	15° to 30°C	User
Microseal 'B' Adhesive Seals	2	15° to 30°C	User
RNase/DNase-free Reagent Reservoirs (if using multichannel pipettes)	4–28	15° to 30°C	User
RNase/DNase-free Strip Tubes and Caps (if using multichannel pipettes)	4–28	15° to 30°C	User

Preparation

- ▶ Remove the following from -15° to -25°C storage and thaw them at room temperature:
 - Appropriate RNA Adapter tubes (depending on the RNA Adapter Indices being used) or the RAP.
 - If using the RAP, review *Handling Adapter Plate* on page 44.
 - Stop Ligation Buffer



NOTE

Do not remove the Ligation Mix tube from -15° to -25°C storage until instructed to do so in the procedures.

- Ligation Control



NOTE

The use of the Ligation Control is optional and it can be replaced with the same volume of Resuspension Buffer.

- ▶ Remove the Resuspension Buffer from 2° to 8°C storage and bring it to room temperature.
- ▶ Review *Handling Magnetic Beads* on page 10.
- ▶ Remove the AMPure XP beads from storage and let stand for at least 30 minutes to bring them to room temperature.
- ▶ Pre-heat the thermal cycler to 30°C.
- ▶ Choose the thermal cycler pre-heat lid option and set to 100°C
- ▶ Apply a CAP barcode label to a new 96-well 0.3 ml PCR plate.
- ▶ Apply a PCR barcode label to a new 96-well 0.3 ml PCR plate.



NOTE

- When indexing libraries using adapter index tubes, Illumina recommends arranging samples that will be combined into a common pool in the same row. Each column should contain a common index. This will facilitate pipetting operations when dispensing indexed adapters and pooling indexed libraries later in the protocol.
- When indexing libraries with the RAP, arrange samples that will be pooled together in the same orientation as the indices in the RAP.



NOTE

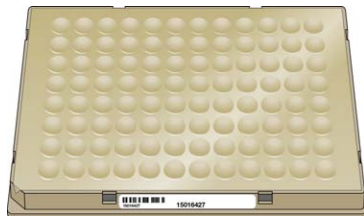
Illumina recommends that the RAP does not undergo more than 4 freeze-thaw cycles. To maximize the use of the RAP, process more than 24 samples at a time. These samples can then be pooled in any supported configuration.

Add LIG

- 1 Do one of the following:
 - If using RNA Adapter tubes, centrifuge the appropriate/desired thawed tubes to 600 xg for 5 seconds.
 - If using a RAP:
 - Thaw the plate for 10 minutes at room temperature on the benchtop. Visually inspect the wells to ensure that they all are completely thawed.
 - Remove the adapter plate tape seal.
 - Centrifuge the plate at 280 xg for 1 minute to collect all of the adapter to the bottom of the well.
 - Remove the plastic cover and save the cover if you are not processing the entire plate at once.

- If this is the first time using this RAP, apply the RAP barcode label to the plate.
- 2 Centrifuge the Ligation Control (if using Ligation Control) and Stop Ligation Buffer tubes to 600 xg for 5 seconds.
 - 3 Immediately before use, remove the Ligation Mix tube from -15° to -25°C storage.
 - 4 Remove the adhesive seal from the ALP plate.
 - 5 Do one of the following:
 - If using the in-line control reagent:
 - Dilute the Ligation Control to 1/100 in Resuspension Buffer (For example, 1 µl Ligation Control + 99 µl Resuspension Buffer) before use. Discard the diluted Ligation Control after use.
 - Add 2.5 µl of diluted Ligation Control to each well of the ALP plate.
 - If not using the in-line control reagent, add 2.5 µl of Resuspension Buffer to each well of the ALP plate.
 - 6 Add 2.5 µl of Ligation Mix to each well of the ALP plate.
 - 7 Return the Ligation Mix tube back to -15° to -25°C storage immediately after use.
 - 8 Do one of the following:
 - If using RNA Adapter tubes, add 2.5 µl of the appropriate/desired thawed RNA Adapter Index to each well of the ALP plate. Gently pipette the entire volume up and down 10 times to mix thoroughly.
 - If using a RAP:
 - Place the RAP on the benchtop so that the part number barcode on the long side of the plate is facing you and the clipped corner is located on the lower left.

Figure 38 Correct RAP Orientation



- Do one of the following to pierce the foil seal:

- If using the entire plate at once, use the bottom of a clean 96-well semi-skirted PCR plate to pierce a hole in all of the well seals simultaneously by gently but firmly pressing the clean plate over the foil seal.
 - If using only part of the plate, use the bottom of a clean eight-tube strip, with caps attached, to pierce holes in the desired columns that will be used for ligation. Repeat with a new, clean eight-tube strip, with caps attached, for each column of adapters that will be used for ligation.
 - Using an 8-tip multichannel pipette, transfer 2.5 μ l of the appropriate/desired thawed RNA Adapter from the RAP well to each well of the ALP plate. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 9 Seal the ALP plate with a Microseal 'B' adhesive seal.
 - 10 Centrifuge the ALP plate to 280 xg for 1 minute.

Incubate 2 ALP

- 1 Place the sealed ALP plate on the pre-heated thermal cycler. Close the lid and incubate at 30°C for 10 minutes.
- 2 Remove the ALP plate from the thermal cycler.

Add STL

- 1 Remove the adhesive seal from the ALP plate.
- 2 Add 5 μ l of Stop Ligation Buffer to each well of the ALP plate to inactivate the ligation. Gently pipette the entire volume up and down 10 times to mix thoroughly.

Clean Up ALP



NOTE

Before performing clean up, review *Handling Magnetic Beads* on page 10 when working with AMPure XP Beads.

- 1 Vortex the AMPure XP Beads until they are well dispersed, then add 42 μ l of mixed AMPure XP Beads to each well of the ALP plate. Gently pipette the entire volume up and down 10 times to mix thoroughly.

- 2 Incubate the ALP plate at room temperature for 15 minutes.
- 3 Place the ALP plate on the magnetic stand at room temperature for 5 minutes or until the liquid appears clear.
- 4 Remove and discard 79.5 μ l of the supernatant from each well of the ALP plate.

**NOTE**

Leave the ALP plate on the magnetic stand while performing the following 80% EtOH wash steps (5–7).

- 5 With the ALP plate remaining on the magnetic stand, add 200 μ l of freshly prepared 80% EtOH to each well without disturbing the beads.
- 6 Incubate the ALP plate at room temperature for 30 seconds, then remove and discard all of the supernatant from each well.
- 7 Repeat steps 5 and 6 once for a total of two 80% EtOH washes.
- 8 While keeping the ALP plate on the magnetic stand, let the samples air dry at room temperature for 15 minutes and then remove the plate from the magnetic stand.
- 9 Resuspend the dried pellet in each well with 52.5 μ l Resuspension Buffer. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 10 Incubate the ALP plate at room temperature for 2 minutes.
- 11 Place the ALP plate on the magnetic stand at room temperature for 5 minutes or until the liquid appears clear.
- 12 Transfer 50 μ l of the clear supernatant from each well of the ALP plate to the corresponding well of the new 0.3 ml PCR plate labeled with the CAP barcode.
- 13 Vortex the AMPure XP Beads until they are well dispersed, then add 50 μ l of mixed AMPure XP Beads to each well of the CAP plate for a second clean up. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 14 Incubate the CAP plate at room temperature for 15 minutes.
- 15 Place the CAP plate on the magnetic stand at room temperature for 5 minutes or until the liquid appears clear.
- 16 Remove and discard 95 μ l of the supernatant from each well of the CAP plate.



NOTE

Leave the CAP plate on the magnetic stand while performing the following 80% EtOH wash steps (17–19)

- 17 With the CAP plate remaining on the magnetic stand, add 200 μ l of freshly prepared 80% EtOH to each well without disturbing the beads.
- 18 Incubate the CAP plate at room temperature for 30 seconds, then remove and discard all of the supernatant from each well.
- 19 Repeat steps 17 and 18 once for a total of two 80% EtOH washes.
- 20 While keeping the CAP plate on the magnetic stand, let the samples air dry at room temperature for 15 minutes and then remove the plate from the magnetic stand.
- 21 Resuspend the dried pellet in each well with 22.5 μ l Resuspension Buffer. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 22 Incubate the CAP plate at room temperature for 2 minutes.
- 23 Place the CAP plate on the magnetic stand at room temperature for 5 minutes or until the liquid appears clear.
- 24 Transfer 20 μ l of the clear supernatant from each well of the CAP plate to the corresponding well of the new 0.3 ml PCR plate labeled with the PCR barcode.



SAFESTOPPING POINT

If you do not plan to proceed to *Enrich DNA Fragments* on page 83 immediately, the protocol can be safely stopped here. If you are stopping, seal the PCR plate with a Microseal 'B' adhesive seal and store at -15° to -25°C for up to seven days.

Enrich DNA Fragments

This process uses PCR to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. The PCR is performed with a PCR primer cocktail that anneals to the ends of the adapters. The number of PCR cycles should be minimized to avoid skewing the representation of the library.



NOTE

PCR enriches for fragments that have adapters ligated on both ends. Fragments with only one or no adapters on their ends are by-products of inefficiencies in the ligation reaction. Neither species can be used to make clusters, as fragments without any adapters cannot hybridize to surface-bound primers in the flow cell, and fragments with an adapter on only one end can hybridize to surface bound primers but cannot form clusters.

Consumables

Item	Quantity	Storage	Supplied By
PCR Master Mix (PMM)	1 tube per 48 reactions	-15° to -25°C	Illumina
PCR Primer Cocktail (PPC)	1 tube per 48 reactions	-15° to -25°C	Illumina
Resuspension Buffer (RSB)	1 tube	2° to 8°C	Illumina
TSP1 (Target Sample Plate) barcode label	1 label per plate	15° to 30°C	Illumina
96-well 0.3 ml PCR plate	1	15° to 30°C	User
AMPure XP beads	50 µl per sample	2° to 8°C	User
Freshly Prepared 80% Ethanol (EtOH)	400 µl per sample	15° to 30°C	User
Microseal 'B' Adhesive Seals	2	15° to 30°C	User

Item	Quantity	Storage	Supplied By
RNase/DNase-free Reagent Reservoirs (if using multichannel pipettes)	5	15° to 30°C	User
RNase/DNase-free Strip Tubes and Caps (if using multichannel pipettes)	5	15° to 30°C	User

Preparation

- ▶ Remove the PCR Master Mix and PCR Primer Cocktail from -15° to -25°C storage and thaw them at room temperature.
- ▶ Centrifuge the thawed PCR Master Mix and PCR Primer Cocktail tubes to 600 xg for 5 seconds.
- ▶ Remove the Resuspension Buffer from 2° to 8°C storage and bring it to room temperature.
- ▶ Review *Handling Magnetic Beads* on page 10.
- ▶ Remove the AMPure XP beads from storage and let stand for at least 30 minutes to bring them to room temperature.
- ▶ Remove the PCR plate from -15° to -25°C storage, if it was stored at the conclusion of *Clean Up ALP* on page 80 and let stand to thaw at room temperature.
 - Centrifuge the thawed PCR plate to 280 xg for 1 minute.
 - Remove the adhesive seal from the thawed PCR plate.
- ▶ Pre-program the thermal cycler with the following program and save as **PCR**:
 - Choose the pre-heat lid option and set to 100°C
 - 98°C for 30 seconds
 - 15 cycles of:
 - 98°C for 10 seconds
 - 60°C for 30 seconds
 - 72°C for 30 seconds
 - 72°C for 5 minutes
 - Hold at 4°C
- ▶ Apply a TSP1 barcode label to a new 96-well 0.3 ml PCR plate.

Make PCR

- 1 Add 5 μ l of thawed PCR Primer Cocktail to each well of the PCR plate.
- 2 Add 25 μ l of thawed PCR Master Mix to each well of the PCR plate. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 3 Seal the PCR plate with a Microseal 'B' adhesive seal.

Amp PCR

- 1 Place the sealed PCR plate on the pre-programmed thermal cycler. Close the lid and select **PCR** to amplify the plate.

Clean Up PCR



NOTE

Before performing clean up, review *Handling Magnetic Beads* on page 10 when working with AMPure XP Beads.

- 1 Remove the adhesive seal from the PCR plate.
- 2 Vortex the AMPure XP Beads until they are well dispersed, then do one of the following:
 - If using the RNA Adapter tubes, add 50 μ l of the mixed AMPure XP Beads to each well of the PCR plate containing 50 μ l of the PCR amplified library. Gently pipette the entire volume up and down 10 times to mix thoroughly.
 - If using the RAP, add 47.5 μ l of the mixed AMPure XP Beads to each well of the PCR plate containing 50 μ l of the PCR amplified library. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 3 Incubate the PCR plate at room temperature for 15 minutes.
- 4 Place the PCR plate on the magnetic stand at room temperature for 5 minutes or until the liquid appears clear.
- 5 Remove and discard 95 μ l of the supernatant from each well of the PCR plate.



NOTE

Leave the PCR plate on the magnetic stand while performing the following 80% EtOH wash steps (6–8).

- 6 With the PCR plate remaining on the magnetic stand, add 200 μl of freshly prepared 80% EtOH to each well without disturbing the beads.
- 7 Incubate the PCR plate at room temperature for 30 seconds, then remove and discard all of the supernatant from each well.
- 8 Repeat steps 6 and 7 once for a total of two 80% EtOH washes.
- 9 While keeping the PCR plate on the magnetic stand, let the samples air dry at room temperature for 15 minutes and then remove the plate from the magnetic stand.
- 10 Resuspend the dried pellet in each well with 32.5 μl Resuspension Buffer. Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 11 Incubate the PCR plate at room temperature for 2 minutes.
- 12 Place the PCR plate on the magnetic stand at room temperature for 5 minutes or until the liquid appears clear.
- 13 Transfer 30 μl of the clear supernatant from each well of the PCR plate to the corresponding well of the new 0.3 ml PCR plate labeled with the TSP1 barcode.

**SAFESTOPPING POINT**

If you do not plan to proceed to *Validate Library* on page 87 immediately, the protocol can be safely stopped here. If you are stopping, seal the TSP1 plate with a Microseal 'B' adhesive seal and store at -15° to -25°C for up to seven days.

Validate Library

Illumina recommends performing the following procedures for quality control analysis on your sample library and quantification of the DNA library templates.

Quantify Libraries

In order to achieve the highest quality of data on Illumina sequencing platforms, it is important to create optimum cluster densities across every lane of every flow cell. This requires accurate quantitation of DNA library templates. Quantify your libraries using qPCR according to the Illumina *Sequencing Library qPCR Quantification Guide*.

Quality Control

- 1 Load 1 μ l of the resuspended construct on an Agilent Technologies 2100 Bioanalyzer using a DNA specific chip such as the Agilent DNA-1000.
- 2 Check the size and purity of the sample. The final product should be a band at approximately 260 bp.

Figure 39 Example of TruSeq Stranded mRNA Sample Prep Library Size Distribution

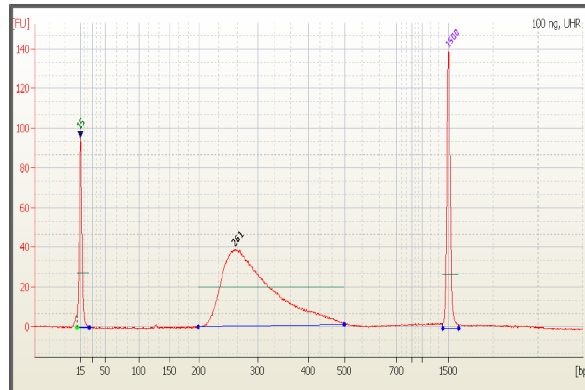
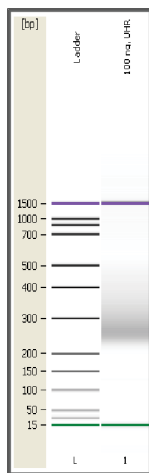


Figure 40 TruSeq Stranded mRNA Sample Prep 260 bp PCR Product



Normalize and Pool Libraries

This process describes how to prepare DNA templates that will be applied to cluster generation. Indexed DNA libraries are normalized to 10 nM in the DCT plate and then pooled in equal volumes in the PDP plate. DNA libraries not intended for indexing are normalized to 10 nM in the DCT plate without pooling.

Consumables

Item	Quantity	Storage	Supplied By
DCT (Diluted Cluster Template) barcode label	1 label per plate	15° to 30°C	Illumina
PDP (Pooled DCT Plate) barcode label (for indexing only)	1 label per plate	15° to 30°C	Illumina
96-well 0.3 ml PCR plate (for indexing only)	1	15° to 30°C	User
96-well MIDI plate	1	15° to 30°C	User
Microseal 'B' Adhesive Seals	2	15° to 30°C	User
Tris-Cl 10 mM, pH8.5 with 0.1% Tween 20	Enough to normalize the concentration of each sample library to 10 nM	15° to 30°C	User

Preparation

- ▶ Apply a DCT barcode label to a new 96-well MIDI plate.
- ▶ Apply a PDP barcode label to a new 96-well 0.3 ml PCR plate (for indexing only).
- ▶ Remove the TSP1 plate from -15° to -25°C storage, if it was stored at the conclusion of *Clean Up PCR* on page 85, and let stand to thaw at room temperature.
 - Centrifuge the thawed TSP1 plate to 280 xg for 1 minute.
 - Remove the adhesive seal from the thawed TSP1 plate.

Make DCT

- 1 Transfer 10 μ l of sample library from each well of the TSP1 plate to the corresponding well of the new MIDI plate labeled with the DCT barcode.
- 2 Normalize the concentration of sample library in each well of DCT plate to 10 nM using Tris-Cl 10 mM, pH 8.5 with 0.1% Tween 20.

NOTE
Depending on the yield quantification data of each sample library, the final volume in the DCT plate can vary from 10-400 μ l.
- 3 Gently pipette the entire normalized sample library volume up and down 10 times to mix thoroughly.
- 4 Depending on the type of library you want to generate, do one of the following:
 - For non-indexed libraries, the protocol stops here. Do one of the following:
 - Proceed to cluster generation. For more information, see the *Illumina Cluster Generation User Guide*.
 - Seal the DCT plate with a Microseal 'B' adhesive seal and store at -15° to -25°C.
 - For indexed libraries, proceed to Make PDP.

Make PDP (for indexing only)



NOTE
Do not make a PDP plate if there is no pooling.

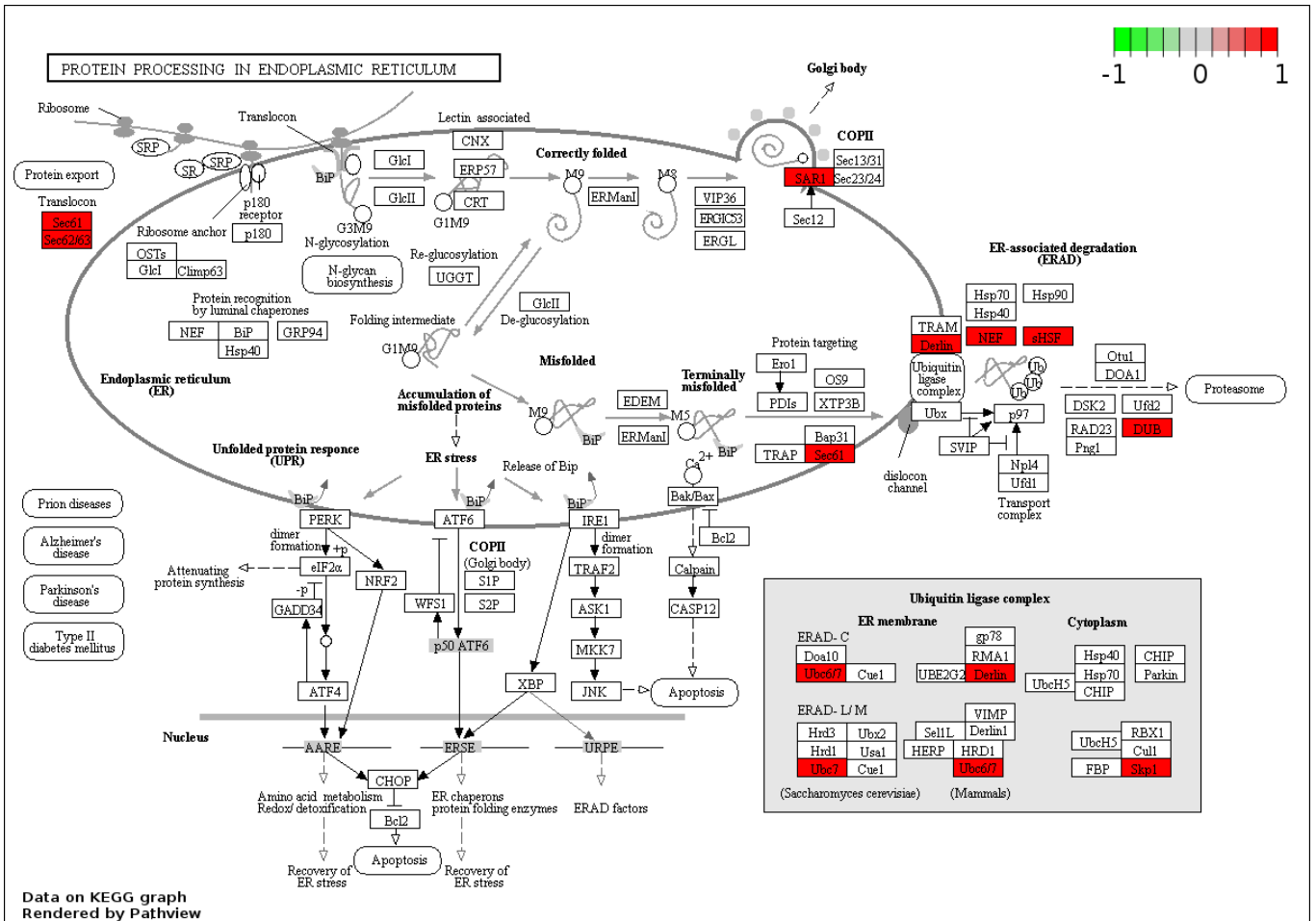
- 1 Determine the number of samples to be combined together for each pool.

NOTE
Keep track of which sample goes into which well, to avoid pooling two samples with the same index.
- 2 Do one of the following:
 - If pooling 2–24 samples:
 - Transfer 10 μ l of each normalized sample library to be pooled from the DCT plate to one well of the new 0.3 ml PCR plate labeled with the PDP barcode.

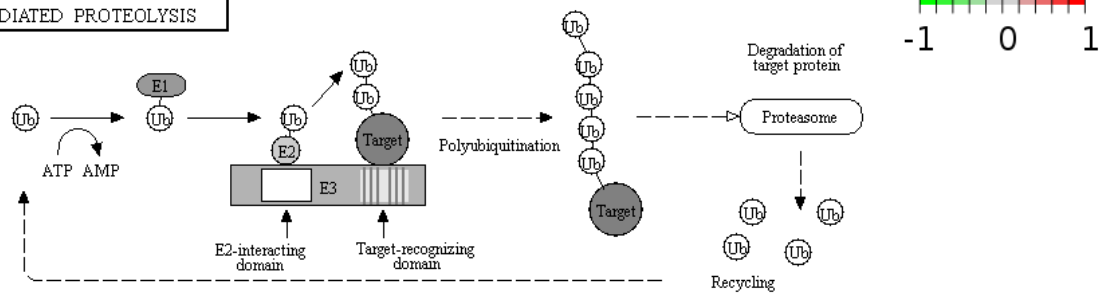
The total volume in each well of the PDP plate should be 10X the number of combined sample libraries and will be 20–240 μ l (2–24 libraries). For example, the volume for 2 samples is 20 μ l, the volume for 12 samples is 120 μ l, or the volume for 24 samples is 240 μ l.

- If pooling 25–96 samples:
 - Using a multichannel pipette, transfer 5 μ l of each normalized sample library in column 1 from the DCT plate to column 1 of the new MIDI plate labeled with the PDP barcode.
 - Transfer 5 μ l of each normalized sample library in column 2 from the DCT plate to column 1 of the PDP plate.
 - Repeat the transfer for as many times as there are remaining columns in the DCT plate. The result will be a PDP plate with pooled samples in column 1. Gently pipette the entire volume of each well of column 1 up and down 10 times to mix thoroughly.
 - Combine the contents of each well of column 1 into well A2 of the PDP plate, for the final pool.
- 3 Gently pipette the entire volume up and down 10 times to mix thoroughly.
- 4 Do one of the following:
- Proceed to cluster generation. For more information, see the *Illumina Cluster Generation User Guide*.
 - Seal the PDP plate with a Microseal ‘B’ adhesive seal and store at -15° to -25°C.

Annexe 3



UBIQUITIN MEDIATED PROTEOLYSIS



E1
(Ubiquitin-activating enzyme)

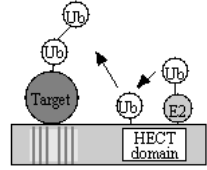
UBE1	UBLE1A	UBLE1B	UBE1C
------	--------	--------	-------

E2
(Ubiquitin-conjugating enzyme)

UBE2A	UBE2B	UBE2C	UBE2D	UBE2E	UBE2F	UBE2G1	UBE2G2	UBE2H
UBE2I	UBE2J1	UBE2J2	UBE2L3	UBE2L6	UBE2M	UBE2N	UBE2O	
UBE2Q	UBE2R	UBE2S	UBE2U	UBE2W	UBE2Z	HIP2	APC1CN	

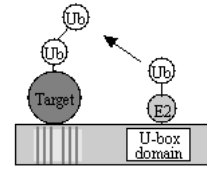
E3
(Ubiquitin ligase)

HECT type E3



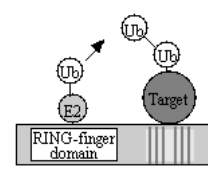
E6AP	UBE3B	UBE3C	Smurf	Itch
WWP1	WWP2	TRIP12	NEDD4	ARF-BP1
EDD1	HERC1	HERC2	HERC3	HERC4

U-box type E3



UBE4A	UBE4B	CHIP
CYC4	PRP19	UIP5

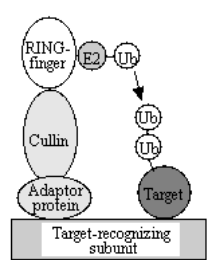
single RING-finger type E3



Mdm2	CBL	Parkin	SIAH-1	FML	TRAF6	MEKK1
COP1	PIRH2	cIAPs	PIAS	SYVN	NHLRC1	AIRE
MGRN1	BRCA1	FANCL	MID1	Trm32	Trm37	

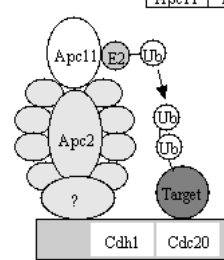
multi subunit RING-finger type E3

Cullin-Rbx E3



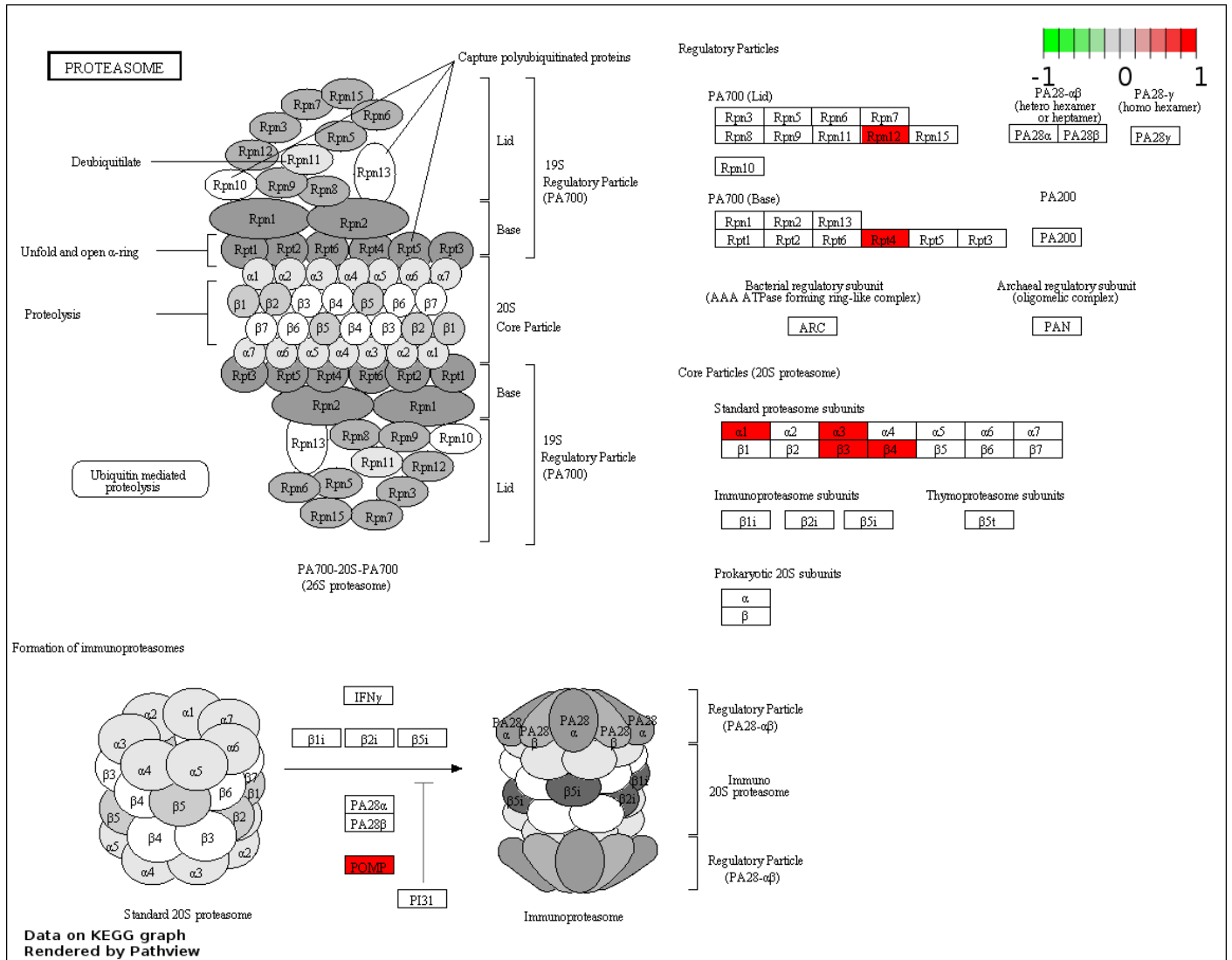
	RING finger	Cullin	Adaptor protein	Target recognizing subunit
SCF complex	RBX1	Cul1	Skp1	F-box
ECV complex	RBX1	Cul2	EloB	VHLbox
Cul3 complex	RBX1	Cul3		BTB
Cul4 complex	RBX1	Cul4	DDB1	DCAF
ECS complex	RBX2	Cul5	EloB	SOC3box
Cul7 complex	RBX1	Cul7	Skp1	Fbxw2

APC/C



RING finger	Cullin	Adaptor protein	Target recognizing subunit	Other subunits
Apc11	Apc2	?	Cdc20	Apc1, Apc3
			Cdh1	Apc4, Apc5
				Apc6, Apc7
				Apc8, Apc9
				Apc10, Apc12
				Apc13

Data on KEGG graph
Rendered by Pathview



RÉSUMÉ / ABSTRACT

La *N*-glycosylation est un événement co- et post-traductionnel majeur de la synthèse protéique chez les eucaryotes. Actuellement, peu d'informations concernant ce processus sont disponibles chez les microalgues. Dans ces travaux, nous avons poursuivi l'exploration de la voie de *N*-glycosylation de la diatomée modèle *P. tricornutum* en focalisant spécifiquement notre travail sur la caractérisation de trois α 1,3-fucosyltransférases putatives. Nos analyses ont permis de remettre en cause l'annotation pour deux des gènes codant pour ces fucosyltransférases et d'en proposer de nouvelles. Afin de réaliser l'étude de la localisation subcellulaire de ces glycosyltransférases, nous avons utilisé une stratégie de sur-expression de ces protéines fusionnées avec une étiquette. Les observations menées par des approches de microscopie confocale et de microscopie électronique à transmission ont permises de mettre en évidence la localisation subcellulaire golgienne de ces enzymes, ainsi que de trois autres glycosyltransférases utilisées comme témoin dans ce travail. Cette étude a permis de réaliser la première localisation subcellulaire de glycosyltransférases chez les microalgues.

P. tricornutum est une diatomée pléïomorphique qui possède trois morphotypes majeurs différents appelés ovale, fusiforme et triradié. Le passage d'un morphotype à un autre dépend fortement des conditions environnementales. Cependant, peu d'informations sont disponibles concernant les mécanismes et la signification physiologique de cette morphogénèse. Afin d'apporter de nouvelles réponses, nous avons réalisé une analyse comparative du transcriptome des trois morphotypes par une approche transcriptomique à haut débit appelée RNA-Seq pour *RNA-Sequencing*. Enfin, en plus d'élargir les connaissances sur la morphogénèse de *P. tricornutum*, les données générées ont également permis d'identifier de nouveaux gènes de référence utilisables pour des analyses de qRT-PCR.

N-glycosylation is a major co- and post-translational event of the protein biosynthesis in Eukaryotes. Actually, little information regarding this processing is available in microalgae. In this work, we continued the exploration of the *N*-glycosylation pathway in the model diatom *P. tricornutum* by focusing especially on the characterization of three putative α 1,3-fucosyltransferases. Our analyzes question the annotation for two of the genes encoding for these fucosyltransferases and propose new ones. In order to achieve the subcellular localization of these glycosyltransferases, we used an over-expressed fusion tag proteins strategy combined together with microscopy approaches. Observations performed by confocal microscopy and TEM allowed us to demonstrate the Golgi subcellular localization of these enzymes. This represent the first subcellular localization of glycosyltransferases in microalgae.

P. tricornutum possess three different morphotypes named oval, fusiform and triradiate. The transition from one morphotype to another strongly depends on the environmental conditions. However, little information is available regarding the mechanisms and physiological significance of this morphogenesis. In order to provide new knowledge related to the morphogenesis in *P. tricornutum*, we conducted a comparative transcriptome analysis of the three morphotypes by an RNA-Seq approach. Besides expanding our knowledge about *P. tricornutum* morphogenesis, data also helped to identify new reference genes for qRT-PCR analysis.

Mots-clés : *N*-glycosylation, *Phaeodactylum tricornutum*, appareil de Golgi, glycosyltransferase, RNA-Seq, diatomées, transcriptome, localisation subcellulaire, morphogénèse.