Etude de la voie de N-Glycosylation et caractérisation de la N-acétylglucosamiltransférase I chez la diatomée Phaeodactylum tricornutum

Bérengère Baïet

▶ **To cite this version:**

Bérengère Baïet. Etude de la voie de N-Glycosylation et caractérisation de la N-acétylglucosamiltransférase I chez la diatomée Phaeodactylum tricornutum. Biochimie [q-bio.BM]. Université de Rouen, 2013. Français. NNT : . tel-02053590

HAL Id: tel-02053590

https://normandie-univ.hal.science/tel-02053590

Submitted on 14 Mar 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.

UNIVERSITE DE ROUEN
UFR DES SCIENCES ET TECHNIQUES

THESE DE DOCTORAT
Ecole Doctorale Normande de Biologie Intégrative, Santé et Environnement
Spécialité : Biochimie

Soutenue le 16 décembre 2013 par :

Bérengère BAÏET HAMONIC

Etude de la voie de N-glycosylation et caractérisation de la N-acétylglucosaminyltransférase I chez la diatomée *Phaeodactylum tricornutum*

Laboratoire d'accueil :
Laboratoire de Glycobiologie et Matrice Extracellulaire Végétale
(EA 4358 Glyco-MEV)

Directeur de thèse : Mme Muriel BARDOR, MCF-HDR

Membres du jury :

Mr François-Yves Bouget	Directeur recherche, Laboratoire Arago, Banyuls	Rapporteur
Mme Christelle Breton	Professeur, CERMAV, Université J. Fourier, Grenoble	Rapporteur
Mr Jean-Paul Cadoret	Directeur recherche, IFREMER de Nantes	Examinateur
Mr Azeddine Driouich	Professeur, Université de Rouen	Président
Mr Ian Probert	Ingénieur recherche, Station Biologique de Roscoff	Examinateur

Remerciement

Je remercie tout d'abord 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 pour effectuer ma thèse.

Je remercie Madame Christelle Breton, Professeur à l'Université Joseph Fourier de Grenoble, et Monsieur François-Yves Bouget, Directeur de Recherche au sein du Laboratoire Arago de Banyuls-sur-mer, de m'avoir fait l'honneur de rapporter sur ce travail.

Je remercie Monsieur Ian Probert, Ingénieur de Recherche à la Station Biologique de Roscoff, d'avoir accepté de participer au jury de cette thèse, ainsi que Monsieur Jean-Paul Cadoret, Directeur de Recherche à l'IFREMER de Nantes, qui m'a également gentiment accueilli au sein de son laboratoire pour les expériences de RT-qPCR.

Je remercie le Professeur Azeddine Driouich, co-Directeur du Laboratoire Glyco-MEV de l'Université de Rouen, d'avoir accepté de présider mon jury de thèse.

Je remercie tout particulièrement ma directrice de thèse, le Docteur Muriel Bardor, Maître de Conférences à l'Université de Rouen, pour sa confiance, ses encouragements, sa disponibilité malgré la distance géographique qui nous séparait, ses conseils et pour le temps consacré à la correction de ce manuscrit. Cela a été un grand plaisir de travailler avec toi et une belle rencontre sur le plan personnel.

Mes remerciements vont également au Docteur Hélène Castel, au Docteur Thomas Lefebvre, à Marie - Thérèse Schouft et à Laurence Desrues du laboratoire DC2N de l'Université de Rouen pour leur aide et leurs conseils dans les expériences de transfection des cellules de mammifères.

Je remercie également le Docteur Bruno Saint-Jean de l'IFREMER de Nantes avec qui j'ai réalisé les expériences de RT-qPCR et qui m'a apporté de façon générale de précieux conseils en biologie moléculaire.

Je remercie Carole Burel ainsi que les Docteurs Marie-Christine Kiefer-Meyer, Laurence Menu-Bouaouiche et Elodie Mathieu-Rivet du laboratoire Glyco-MEV pour leur collaboration à ce travail et pour toute l'aide apportée.

Je remercie mon jeune padawan Clément Ovide que j'ai eu l'honneur de co-encadrer durant ses stages de Master 1 et 2. Ca a été un bonheur de venir travailler avec toi tous les matins. Merci de t'être prêté à mes blind-tests pendant nos longues heures de repiquage, et d'avoir survécu à mes chants mélodieux que je te faisais subir toute la journée. Je te souhaite bon courage pour la suite.

Je souhaite remercier tous les membres du Laboratoire Glyco-MEV, notamment pour la bonne ambiance lors des repas du midi. J'espère que vous continuez à fêter les anniversaires, Jean-Claude et Philippe je compte sur vous pour tenir à jour le calendrier !

Ces années n'auraient pas été les mêmes si elles n'avaient pas été partagées avec d'autres thésards. Merci tout particulièrement à Marco, Flavien, Romain et Sophie pour tous nos fous rires, nos blagues de potaches et l'ambiance certes studieuse mais aussi délirante que nous avons su générer. Bien que nous soyons tous partis du laboratoire, à part l'irréductible Sophie, et que nous soyons tous éparpillés géographiquement, j'espère que nous continuerons à rester en contact.

Pour leur soutien et leur présence, je remercie chaleureusement toute ma famille et mes amis. Un grand merci en particulier à mes nourrices improvisées de ces derrières semaines, Benji, TiClo, poupoule et mes parents, sans qui l'achèvement de ce manuscrit aurait été difficile.

A mon homme qui a eu la folle idée d'accepter de passer sa vie avec moi. Il serait trop long d'expliquer toute l'importance que tu as dans ma vie. Je te remercie pour ton aide, ton soutien, ta joie de vivre et tout ton amour, mais plus encore pour le plus beau cadeau que tu m'aies donné il y a un peu plus d'un an maintenant, notre petite Lola ! Ma belle, je finirai par toi, tu es la plus belle chose qui me soit arrivée dans la vie. Merci d'avoir été patiente et d'avoir bien voulu me partager avec la rédaction de cette thèse. J'espère que tu seras fière de ta maman. Je t'aime très fort !

M^me^ Baïet Hamonic Bérengère

berengere.hamonic@gmail.com
Immeuble A escalier 2
144 bis rue du renard
76000 Rouen

Age 29 ans
Née le 02/04/1984
A Dieppe
Mariée, 1 enfant
Nationalité française

Présentation de la candidate

Parcours Universitaire

2009-2013 : **Doctorat Biologie cellulaire** de l'Université de Rouen, spécialité biochimie et biologie moléculaire, allocation ministérielle du MESR
2008-2009 : **Master 2 biochimie et biologie moléculaire** de l'Université de Rouen**,** mention très bien, majeur de promotion.
2007-2008 : **Master 1 biochimie et biologie moléculaire** de l'Université de Rouen**,** mention bien, majeur de promotion.
2006-2007 : **Licence 3 biochimie et biologie moléculaire** de l'Université de Rouen**,** mention bien, majeur de promotion.
2005-2006 : **Licence 2 biologie cellulaire et physiologie** de l'Université de Rouen**,** mention assez bien.

Activité dans le cadre d'une mission enseignement

2009-2012 : Enseignement à l'Université de Rouen de **cours**, **TD** et **TP** pour des **1^ère^**, **2^ème^** et **3^ème^ années de Licence SVTE** (**192 h sur 3 ans**). Matières enseignées : biochimie, biologie moléculaire et biologie cellulaire.

Co-Encadrement de stage

2011-2012 : Co-encadrement du projet de recherche d'un étudiant de **Master 2** biochimie et biologie cellulaire.
2010-2011 : Co-encadrement du projet de recherche d'un étudiant de **Master 1** biochimie et biologie cellulaire.
2009-2010 : Co-encadrement du projet de recherche d'une étudiante de **3^ème^ année de Licence** SVTE parcours Ecologie et Biologie des Organismes.

Communications

Le nom du présentateur est souligné.

Communications orales :

Congrès nationaux :

- Mathieu-Rivet E., Ovide C., **Baïet B.**, Kiefer-Meyer M.-C., Burel C., Plasson C., Lerouge P. and Bardor M. (**2013**) 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.

- **Baïet B.***, Rivet-Mathieu E.*, Dardelle F., Burel C., Saint-Jean B., Le Mauff F., Teo G., Kiefer-Meyer M.-C., Cadoret J.-P., Hippler M., Martínez F., Lerouge P. and Bardor M. (**2012**) Towards unraveling the N-glycosylation pathway in microalgae, 24ème Journée du Groupe Français des Glycosciences GFG 2012, Domaine du Val Joly, France, 21-25 mai. * Participation équivalente des deux premiers co-auteurs dans le travail.
- **Baïet B.**, Burel C., Saint-Jean B., Louvet R., Menu-Bouaouiche L., Kiefer-Meyer M. C., Mathieu-Rivet E., Lefebvre T., Castel H., Carlier A., Cadoret J. P., Lerouge P. and Bardor M. (**2011**) Characterization of N-glycosylation pathway in a diatom *Phaeodactylum tricornutum.* XVIIème Journée de l'IFRMP 23, Mont-Saint-Aignan. 17 juin.

Congrès internationaux :

- 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. **(2013)** Deciphering the *N*-glycosylation pathway in microalgae using the diatom *Phaeodactylum tricornutum* as a model. EMBO Workshop – The molecular Life of diatoms, Paris, France, 25-28 juin. *Participation équivalente des deux premiers co-auteurs dans le travail.
- Lerouge P., **Baïet B.**, Burel C., Saint-Jean B., Cauchard A., Louvet R., Carlier A., Cadoret J-P., and Bardor M. (**2009**) Exploring microalgae as cell factory for the production of therapeutic proteins. Biotechnologia Habana 2009, Cuba. 2-5 novembre.

Communications par affiche :

Congrès nationaux :

- Rivet-Mathieu E.*, **Baiet B.***, Burel C., Saint-Jean B., Le Mauff F., Teo G., Kiefer-Meyer M.-C., Dardelle F., Follet-Gueye M.-L., Cadoret J.-P., Hippler M., Martínez F., Lerouge P. and Bardor M. (**2011**) Microalgae: a suitable system for the production of glycosylated biopharmaceuticals. 5ème édition des Rencontres Normandes en Chimie-Biologie-Santé, Rouen, 9 décembre. *Participation équivalente des deux premiers co-auteurs dans le travail.
- Galas L., Bénard M., Schapman D., Bernard S., Lebon A., Banquet S., Edwards-Lévy F., Richard V., Brakenhielm E., Percoco, G., Ramdani, Y., Lati E., Yvergnaux L., Lefeuvre L., Ridziniak G., Follet-Gueye M.L., Driouich A., Lerouge, P., **Baïet B.,** Burel C., Bardor M., Young C., Gorecki D., Denorme, M., C. Roux, C, Gonzalez B.J., Anouar, Y., Yon, L., Dubessy, C., Brifault C., Gras M., Hilber P., Wurtz O., Vaudry H. and Vaudry D. (**2011**) Complementary cell imaging technologies for biotherapy strategies. 5ème édition des Rencontres Normandes en Chimie-Biologie-Santé. Rouen. 9 décembre.
- Mathieu-Rivet E., Arias C., Loutelier-Bourhis C., Kiefer-Meyer M.-C., Le Mauff F., Burel C., **Baïet B.**, Hippler M., Martinez F., Lerouge P., Villarejo A. and Bardor M. (**2011**) Characterization of N-glycosylation pathway in the green microalga, *Chlamydomonas reinhardtii.* XVIIème Journée de l'IFRMP 23, Mont-Saint-Aignan. 17 juin.
- **Baïet B.**, Burel C., Saint-Jean B., Louvet R., Menu-Bouaouiche L., Kiefer-Meyer M-C., Rivet E., Castel H., Lefebvre T., Carlier A., Cadoret J-P., Lerouge P. and Bardor M. (**2011**) N-glycosylation pathway in the *Phaeodactylum tricornutum.* XIVième Journées de l'EdNBISE, Le Havre, 6-7 juin.

- Rivet E., **Baïet B.,** Burel C., Saint-Jean B., Kiefer-Meyer M-C., Louvet R., Lefebvre T., Cadoret J.-P., Lerouge P. and Bardor M. (**2010**) Analysis of protein N-glycosylation in microalgae. XVI^{ième} Journée de l'IFRMP23, Rouen, 4 juin.
- **Baïet B.**, Burel C., Saint-Jean B., Cauchard A., Louvet R., Lefebvre T., Carlier A., Castel H., Cadoret J.-P., Lerouge P. and Bardor M. **(2009)** Characterization of the microalgae N-glycosylation pathway: first evidences in the diatom, *Phaeodactylum tricornutum.* Glucidoc2009, Autrans. 13-16 octobre.
- **Baïet B.**, Burel C., Saint-Jean B., Louvet R., Lefebvre T., Carlier A., Castel H., Cadoret J-P., Lerouge P. and Bardor M. (**2009**) Protein N-linked glycosylation in *Phaeodactylum tricornutum* : a regulated process ? XV^{ème} Journée de l'IFRMP 23, Mont-Saint-Aignan. 5 juin

Congrès internationaux :

- Ovide C.*, **Baiet B.***, Mathieu-Rivet E., Dardelle F., Burel C., Saint-Jean B., Kiefer-Meyer M.-C., Cadoret J.-P., Lerouge P. and Bardor M. **(2013)** Deciphering the *N*-glycosylation pathway in microalgae using the diatom *Phaeodactylum tricornutum* as a model. EMBO Workshop – The molecular Life of diatoms, Paris, France, 25-28 juin. *Participation équivalente des deux premiers co-auteurs dans le travail.
- Rivet E., Burel C., Arias C., **Baïet B.,** Louvet R., Kiefer-Meyer M-C., Hippler M., Lerouge P., Villajero A. and Bardor M. (**2010**) From plant to microalgae glycobiology: *Chlamydomonas reinhardtii* as a model to understand N-glycosylation pathway. Cell and molecular biology of *Chlamydomonas*, Wheaton College, Norton, USA, 6-10 juin.
- **Baïet B.**, Burel C., Saint-Jean B., Cauchard A., Louvet R., Carlier A., Cadoret J-P., Michel R., Lerouge P. and Bardor M. (**2009**) Evidences for N-glycosylation post-translational modifications in microalgae: *Phaeodactylum tricornutum* as a model. Post-translational modifications 2009, Pragues. 10-11 septembre.

Publication-Brevet

- **N-GLYCANS OF *PHAEODACTYLUM TRICORNUTUM* DIATOM AND FUNCTIONAL CHARACTERIZATION OF ITS N-ACETYLGLUCOSAMINYLTRANSFERASE I ENZYME.** Auteurs : **Baïet B.**, Burel C., Saint-Jean B., Louvet R., Menu-Bouaouiche L., Kiefer-Meyer M.C., Mathieu-Rivet E., Lefebvre T., Castel H., Carlier A., Cadoret J.-P., Lerouge P. and Bardor M. Journal de parution : Journal of Biological Chemistry. Année de publication : **2011**. Références complètes : J Biol Chem. 2011.

- **N-GLYCOSYLATION IN TRANSFORMED *PHAEODACTYLUM TRICORNUTUM.*** Bardor M., Louvet R., Saint-Jean B., Burel C., **Baïet B.,** Carlier A., Cadoret J-P. and Lerouge P. (**2011**) Brevet PCT/EP2011/003756, Université de Rouen/ IFREMER/ Algenics. Licence d'exploitation avec la société Algenics, Nantes en cours de négociation.
- **N-GLYCOSYLATION IN TRANSFORMED *PHAEODACTYLUM TRICORNUTUM.*** Bardor M., Louvet R., Saint-Jean B., Burel C., **Baïet B.**, Carlier A., Cadoret J-P. and Lerouge P. (**2010**). Brevet européen n°10007813.8 Université de Rouen/IFREMER.

Séquences déposées dans une base de données

1. Baïet B., Louvet R., Kiefer-Meyer M-C. and Bardor M. (**2010**) *P. tricornutum* GnT I. GenBank™ BankIt 1370414.

2. Baïet B., Louvet R., Kiefer-Meyer M-C. and Bardor M. (**2010**) Modified *P. tricornutum* GnT I expressed in CHO GenBank™ BankIt 1370344.

Stages complémentaires de formation

- Janvier–février 2011 : Analyse des nucléotides-sucres et des N-glycannes dans les différents morphotypes de *Phaeodactylum tricornutum*. **Institut Bioprocessing Technology Institut (BTI), A*STAR, Singapour.**
- Avril 2010 : Formation d'utilisation du spectromètre de masse MALDI-TOF, plate-forme PRIMACEN, Université de Rouen.
- 18-28 janvier et 15-26 mars 2010 : Analyse par RT-qPCR de l'expression du gène de la GnT I de *Phaeodactylum tricornutum* en fonction du taux de nitrate dans la culture. **Laboratoire de Physiologie et Biotechnologie des Algues (PBA) de l'IFREMER de Nantes.**

Formations suivies dans le cadre de l'Ecole Doctorale et de la mission enseignement

- Niveau 1 de secourisme (Université de Rouen).
- Conduite de réunion (Université de Rouen).
- Handicaps et adaptations pédagogiques (IUFM de Rouen).
- Veille et réseautage social (IUFM de Rouen).
- Le schéma heuristique, un outil pour la formation (IUFM de Rouen).
- Ethique de l'enseignement (IUFM de Rouen).
- Enseigner et former à l'université (IUFM de Rouen).
- Eléments de didactique des sciences pour enseigner dans le supérieur (biologie) (IUFM de Rouen).
- Groupes d'analyses de situations professionnelles (IUFM de Rouen).
- Gérer son corps, sa voix, sa communication pour mieux enseigner (IUFM de Rouen).
- Production d'outils numériques à destination des étudiants (IUFM de Rouen).

Abstract

Study of the N-glycosylation pathway and characterization of N-acetylglucosaminyltransferase I in *Phaeodactylum tricornutum.*

N-glycosylation is a major co- and post-translational event in the protein synthesis. Beside a lot of information in higher eukaryotes, very little is known regarding this pathway in microalgae. In this work, we describe the N-glycosylation pathway of a model diatom, *Phaeodactylum tricornutum*. First, bio-informatic analysis of its genome revealed the presence of sequences potentially encoding for transferases responsible for the synthesis of the precursor oligosaccharide, some subunits of the oligosaccharyltransferase complex, ER glucosidases and chaperones required for protein quality control, and the Golgi α-mannosidase I involved in the trimming of the N-glycan into high mannose-type N-glycan. Moreover, structural analyses of N-linked glycans demonstrated that *P. tricornutum* endogenous proteins carry mainly high mannose-type N-glycans ranging from Man-5 to Man-9. However, a minor population of more mature N-glycans were also detected, in particular $Man_3FucGlcNAc_2$. Consistently, one gene potentially encoding for a N-acetylglucosaminyltransferase I (GnT I), a Golgi glycosyltransferase that initiates the formation of complex-type N-glycans, was predicted in the diatom genome.We demonstrated that the GnT I gene encodes for an active protein which is able to restore complex-type N-glycan maturation in the Chinese Hamster Ovary Lec1 mutant, defective in its endogeneous GnT I activity. Furthermore, study of the GnT I expression in *P. tricornutum* according to culture conditions, morphotype and strain revealed the presence of GnT I transcripts. Taken together, all these data suggest the occurrence of a GnT I dependent N-glycosylation pathway in *P. tricornutum*.

Keywords: Carbohydrate Biosynthesis, Carbohydrate Structure, Glycosylation, Golgi apparatus, *N*-Acetylglucosaminyltransferase I, *Phaeodactylum tricornutum*, Diatom, Glycosyltransferase, Microalgae

Sommaire

REMERCIEMENT **3**

PRESENTATION DE LA CANDIDATE **7**

ABSTRACT **13**

SOMMAIRE **15**

LISTE DES FIGURES **23**

LISTE DES TABLEAUX **29**

LISTE DES ABREVIATIONS **31**

INTRODUCTION GENERALE **37**

CHAPITRE I : SYNTHESE BIBLIOGRAPHIQUE **41**

I- Modèle d'étude : les diatomées. **43**

I-1- Microalgues. **43**

I-1-1- Position phylogénétique. ***43***

I-1-2- Origine endosymbiotique. ***45***

I-2- Diatomées. **46**

I-2-1- Biologie des diatomées. ***46***

I-2-1-1- Morphologie et cytologie. 46

I-2-1-2- Reproduction. 48

I-2-2- Classification et histoire des diatomées. ***50***

I-2-3- Les diatomées dans la biosphère. ***52***

I-2-3-1- Mode de vie. 52

I-2-3-2- Rôle dans la biosphère. 54
I-2-4- Génome des diatomées. ***56***
I-2-5- Utilisations des diatomées. ***60***
I-2-5-1- Diatomite 61
I-2-5-2- Alimentation. 62
I-2-5-3- Santé, médecine. 62
I-2-5-4- Chimie fine 63
I-2-5-5- Energie 64
I-2-5-6- Environnement 64
I-2-5-7- Nanotechnologie 65
I-2-5-8- Usine cellulaire 65
I-3- La diatomée modèle *Phaeodactylum tricornutum*. 66

II- N-glycosylation. 69
II-1- Réticulum endoplasmique, siège de la N-glycosylation des protéines. 69
II-1-1- Synthèse du précurseur oligosaccharidique. ***70***
II-1-1-1- Etapes cytosoliques. 70
II-1-1-2- Etapes luminales. 71
II-1-2- Transfert du précurseur oligosaccharidique sur une protéine par l'OST. ***72***
II-1-2-1- Structure et organisation de l'OST. 73
II-1-2-2- Spécificité de substrat de l'OST. 75
II-1-2-2-1- Spécificité pour le substrat donneur. *75*
II-1-2-2-2- Spécificité pour le substrat accepteur. *75*
II-1-3- Repliement et contrôle qualité des glycoprotéines. ***76***
II-1-3-1- Déglucosylation de l'oligosaccharide $Glc_3Man_9GlcNAc_2$ en $Glc_1Man_9GlcNAc_2$ par l'action successive des α-glucosidases I et II. 77
II-1-3-2- Cycle de contrôle qualité des glycoprotéines impliquant les chaperonnes calnexine et calréticuline. 78
II-2- Appareil de Golgi : siège de la diversité glycannique....81
II-2-1- Maturation des N-glycannes chez les levures....82
II-2-2- Voie de biosynthèse des N-glycannes complexes chez les mammifères, les plantes et les insectes : un processus initié par la GnT I....83
II-2-2-1- N-glycannes matures des plantes. 87
II-2-2-2- N-glycannes matures des mammifères. 89

II-2-2-3- N-glycannes matures des insectes. 90
II-2-3- Structure et caractéristique de la GnT I. *91*
II-2-4- Localisation subcellulaire des glycosidases et glycosyltransférases golgiennes. *92*
II-2-5- Rôle des N-glycannes complexes. *94*

III- Objectifs de la thèse. **95**

CHAPITRE II : MATERIEL ET METHODES **97**

I- MATERIEL **99**
I-1- Matériel végétal : *Phaeodactylum tricornutum* **99**
I-2- Matériel animal : Cellules d'ovaire de hamster chinois (CHO). **99**
I-3- Plasmides et vecteurs utilisés. **100**
I-3-1- pUC57+FuT46109 et pUC57+FuT54599 ....*100*
I-3-2- pcDNA3.1/V5-His-TOPO® *101*
I-4- Amorces PCR. **102**

II- METHODES **103**
II-1- Mode de culture **103**
II-1-1- Culture en Batch *103*
II-1-2- Culture continue *104*
II-2- Analyses bioinformatiques **105**
II-2-1- Alignements multiples. *105*
II-2-2- Analyses des séquences nucléotidiques. *106*
II-2-3- Analyses des séquences protéiques. *106*
II-2-4- Recherche d'homologies. *107*
II-2-4-1- Par comparaison de séquences. 107
II-2-4-2- Par recherche de motifs. 107
II-2-5- Vérification des séquences protéiques exprimées dans les cellules de mammifères. *108*
II-2-6- Analyse du séquençage. *108*
II-3- Techniques de biologie moléculaire. **109**
II-3-1- Analyses des acides nucléiques. *109*
II-3-1-1- Analyse quantitative : dosage par spectrophotométrie. 109

II-3-1-2- Analyse qualitative : électrophorèse en gel d'agarose.....109
II-3-2- Clonage et expression des gènes d'intérêt. 110
II-3-2-1- Extraction de l'ADN génomique.....110
II-3-2-2- Amplification par Polymerase Chain Reaction (PCR). 110
II-3-2-3- Purification du produit PCR.....112
II-3-2-4- Ligation et transformation bactérienne. 112
II-3-2-5- Extraction d'ADN plasmidique.....112
II-3-2-6- Contrôle des clones sélectionnés. 113
II-3-2-6-1- Sélection des clones positifs par digestion enzymatique. 113
II-3-2-6-2- Séquençage.....114
II-3-2-7- Linéarisation du vecteur recombinant. 114
II-3-2-8- Transfection stable des cellules CHO. 114
II-3-2-8-1- Par électroporation. 115
II-3-2-8-2- Par lipofection.....116
II-3-2-8-3- Sélection monoclonale. 116
II-3-3- Etude de l'expression du gène de la GnT I.....117
II-3-3-1- Extraction des ARN totaux. 117
II-3-3-2- Purification des ARN.....117
II-3-3-3- Reverse transcription PCR (RT-PCR). 117
II-3-3-3-1- Transcription inverse.....117
II-3-3-3-2- Amplification par PCR. 118
II-3-3-4- PCR quantitative en temps réel. 119
II-3-3-4-1- Transcription inverse.....119
II-3-3-4-2- Amplification par PCR. 119
II-3-4- Analyse des données de Q-PCR.....120
II-4- Techniques de biochimie.....120
II-4-1- Préparation des extraits protéiques. 120
II-4-1-1- Extraction des protéines de *P. tricornutum*.....120
II-4-1-2- Extraction des protéines des cellules CHO. 121
II-4-1-3- Dosage protéique. 121
II-4-2- Déglycosylation. 122
II-4-2-1- Digestion par l'Endoglycosidase H. 122
II-4-2-2- Digestion par la peptide N-glycosidase F. 123
II-4-2-3- Digestion par la peptide N-glycosidase A. 124

II-4-3- Immuno- et affino-blot. *125*
II-4-3-1- Séparation des protéines par gel d'électrophorèse SDS-PAGE. 125
II-4-3-2- Transfert électrophorétique de protéines sur membrane de nitrocellulose. ..126
II-4-3-3- Immuno-marquage sur empreinte. 127
II-4-3-3-1- Anticorps anti-épitope V5. ..... *127*
II-4-3-3-2- Anticorps anti-xylose et anti-fucose. *128*
II-4-3-4- Affino-marquage sur empreinte. 128
II-4-3-5- Révélation. 129
II-4-3-5-1- Electrochimioluminescence. *129*
II-4-3-5-2- 4-chloro-1-naphtol. *129*
II-4-3-5-3- Réactif BCIP/NBT *129*
II-4-4- Immunodétection sur cellules de mammifères. ***130***
II-4-4-1- Fixation et perméabilisation des cellules de mammifère. 130
II-4-4-2- Immunomarquage sur cellules avec l'anticorps anti-V5. 131
II-4-5- Purification des N-glycannes. ***131***
II-4-5-1- Chromatographie sur colonne hydrophobe C18. 131
II-4-5-2- Chromatographie sur colonne de charbon actif (Carbograph). 132
II-4-6- Purification des N-glycannes. ***132***
II-4-6-1- Marquage des N-glycannes. 132
II-4-6-2- Digestion par exoglycosidase des N-glycannes couplés au 2-AB. 132
II-4-7- Analyse des N-glycannes par spectrométrie de masse MALDI TOF. ***133***

CHAPITRE 3 : RESULTATS **135**

I- Processus de N-glycosylation et caractérisation fonctionnelle de la GnT I de *Phaeodactylum tricornutum*. **137**

II- Résultats complémentaires. **152**
II-1- Etude de l'expression du gène de la GnT I en fonction des conditions de culture, du morphotype et de la souche. **152**
II-1-1- Analyse transcriptionnelle du gène de la GnT I de P. tricornutum en fonction de la souche et du morphotype par RT-PCR. ***153***
II-1-2- Analyse transcriptionnelle du gène de la GnT I en fonction de la quantité de nitrate dans le milieu de culture par RT-qPCR. ***154***

II-2- Initiation de la caractérisation fonctionnelle des 3 gènes de *P. tricornutum* codant potentiellement pour des α-1,3-Fucosyltransférases. 157

II-2-1- Analyse bioinformatique des gènes 46109, 46110 et 54599 de P. tricornutum. 157

II-2-2- Clonage des trois gènes 54599, 46109 et 46110 de P. tricornutum codant potentiellement pour des α-1,3-fucosyltransférases. 162

II-2-3- Transfection des cellules CHO FUT8-/- avec les plasmides pcDNA3.1/V5-His-TOPO + 54599, pcDNA3.1/V5-His-TOPO + 46109 et pcDNA3.1/V5-His-TOPO + 46110 et contrôle de l'expression des protéines recombinantes. 164

II-2-4- Test préliminaire de l'activité de FuT54599 par immunomarquage sur empreinte avec l'anticorps anti-α-1,3-fucose du core. 166

CHAPITRE 4 : DISCUSSION-PERSPECTIVES.....169

I- La voie de N-glycosylation de *Phaeodactylum tricornutum : un processus GnT I-dépendant.* 172

II- Fucosylation des N-glycannes de *Phaeodactylum tricornutum.* 177

III- Rôle des N-glycannes complexes chez *Phaeodactylum tricornutum : une possible fonction dans l'adaptation à un stress.* 179

IV- *Phaeodactylum tricornutum* : nouvelle usine cellulaire pour la production de glycoprotéines à intérêt thérapeutique. 182

CONCLUSION GENERALE 187

ANNEXES 191

Annexe 1.....193

Annexe 2.....201

REFERENCES BIBLIOGRAPHIQUES 255

Liste des figures

Figure 1 : Arbre phylogénétique des lignées eucaryotes d'après Keeling, 2013. 42
Figure 2 : Origine endosymbiotique des microalgues d'après Keeling, 2013. 44
Figure 3 : Structure et diversité du frustule des diatomées. 46
Figure 4 : Ultrastructure d'une diatomée. 47
Figure 5 : Cycle de vie d'une diatomée : multiplication végétative et reproduction sexuée (schéma de Russell Habermann extrait du site http://www.seagrant.umn.edu/newsletter/2013/01/what_good_is_a_diatom.htm).....49
Figure 6 : Photographies prises en microscopie électronique à balayage des quatre classes de diatomées : centrique radiale, centrique polaire, pennée raphidée et pennée araphidée (Kroger et Poulsen, 2008). 50
Figure 7 : Schéma présentant la chronologie des divergences entre les quatre lignées majeures de diatomées et les événements géologiques qui coïncident (d'après Armbrust, 2009). 51
Figure 8 : Diversité des formes coloniales des diatomées (photos au microscope optique de J. Oyadomari, http://www.keweenawALGae. mtu.edu/gallery_pages/diatoms2.htm). 53
Figure 9 : Cycle biogéochimique de l'azote en milieu marin. 55
Figure 10 : Schéma présentant l'origine du génome des diatomées par endosymbiose primaire et secondaire et par transfert horizontal de gènes bactériens (Finazzi et al., 2010). 58
Figure 11 : Les différents domaines d'application des diatomées (Lopez et al., 2005). 60
Figure 12 : Schéma présentant les changements morphotypiques de *P. tricornutum* en réponse à des signaux environnementaux (De Martino et al., 2011). 67
Figure 13 : Voie de biosynthèse du précurseur oligosaccharidique $Glc_3Man_9GlcNAc_2$ d'après (Aebi, 2013). 71
Figure 14 : Transfert de l'oligosaccharide précurseur $Glc_3Man_9GlcNAc_2$ sur un polypeptide en cours d'élongation par l'oligosaccharyltransférase (OST) d'après Schwarz et Aebi, 2011. 73
Figure 15 : Différentes structures générales de l'oligosaccharyltransférase (OST) chez les eucaryotes (Schwarz et Aebi, 2011). 74
Figure 16 : Repliement et cycle de contrôle qualité des glycoprotéines N-liées dans le réticulum endoplasmique (RE ; d'après Williams, 2006).. 79
Figure 17 : Caractéristiques communes et points de divergence dans la biosynthèse des N-glycannes matures de mammifère, d'insecte, de plante et de levure.....80

Figure 18 : Structure d'une protéine membranaire de type II (d'après Wildt et Gerngross, 2005)..... 81
Figure 19 : Voie de N-glycosylation de la levure *Saccharomyces cerevisiae* (d'après De Pourcq et al., 2010)..... 83
Figure 20 : Biosynthèse et maturation des N-glycannes de plante, de mammifère et d'insecte dans le réticulum endoplasmique et l'appareil de Golgi. 84
Figure 21 : Biosynthèse et maturation des N-glycannes de plante dans le réticulum endoplasmique, l'appareil de Golgi et les lieux de stockage comme la vacuole et la matrice extracellulaire (MEC). 86
Figure 22 : Biosynthèse et maturation des N-glycannes de mammifère dans l'appareil de Golgi (d'après Nettleship, 2012). 88
Figure 23 : Structures type de paucimannoses et de N-glycannes complexes d'insecte. 90
Figure 24 : Diagramme en ruban de la partie catalytique de la GnT I de lapin (Unligil et al., 2000). 92
Figure 25 : Localisation sub-cellulaire des glycosyltransférases et glycosidases golgiennes impliquées dans la maturation des N-glycannes complexes (d'après Wildt et Gerngross, 2005 ; Schoberer et al., 2013). 93
Figure 26 : Cellules d'ovaire de hamster chinois (CHO) sauvages et CHO Lec1, lignée mutante déficiente en activité N-acétylglucosaminyltransférase I endogène. 99
Figure 27 : Cellules d'ovaire de hamster chinois sauvages (CHO K1) et CHO FuT8-/-, lignée mutée sur le gène de la fucosyltransférase 8 (FuT8) endogène.. 100
Figure 28 : Cartes des vecteurs pUC57+FuT46109 et pUC57+FuT54599 réalisées avec le logiciel pDRAW32 version 1.1.114. 100
Figure 29 : Carte du vecteur pcDNA3.1/V5-His-TOPO® (Invitrogen, K4800-01)..... 101
Figure 30 : Suivi de la culture continue de la souche Pt1.8.6 de *Phaeodactylum tricornutum* en fonction de la concentration en nitrate. 105
Figure 31 : Principe de la transfection stable par lipofectamine (A) ou électroporation (B). 115
Figure 32 : Représentation des sites de coupure des glycoprotéines par l'enzyme de dé-N-glycosylation Endoglycosidase H. 122
Figure 33 : Représentation des sites de coupure des glycoprotéines par l'enzyme de dé-N-glycosylation peptide N-glycosidase F. 123
Figure 34 : Représentation des sites de coupure des glycopeptides par l'enzyme de dé-N-glycosylation peptide N-glycosidase A. 124
Figure 35 : Marqueurs de taille protéiques. 126

Figure 36 : Représentation schématique de la cassette de transfert des gels de polyacrylamide sur membrane de nitrocellulose.....127
Figure 37 : Représentation des épitopes reconnus par les anticorps anti-α-1,3-fucose (Anti-Fuc) et anti-β-1,2-xylose (Anti-Xyl). 128
Figure 38 : Principe de l'immunodétection sur cellules par l'anticorps anti-V5. 130
Figure 39 : Profil transcriptomique du gène de la N-acétylglucosaminyltransférase I (GnT I) de *P. tricornutum* selon 16 conditions étudiées (http://genome.jgi-psf.org/cgi-bin/browser Load/?db=Phatr2&position=chr_16:89066-90505).. 152
Figure 40 : Electrophorèse en gel d'agarose révélant le produit d'amplification PCR obtenu à partir des ADNc de *P. tricornutum* grâce à l'utilisation d'amorces spécifiques du gène de la GnT I et du gène de l'actine 1 (gène de référence) en fonction de la souche et du morphotype. 153
Figure 41 : Expression du gène de la GnT I de *P. tricornutum* par PCR quantitative en temps réel en fonction de la concentration en nitrate (NO_3^-) dans le milieu de culture. 156
Figure 42 : Alignement multiple réalisé avec le logiciel MultAlin des séquences protéiques prédites FuT46109, FuT46110 et FuT54599 de *P. tricornutum*, avec celles de la FucT A et de la FucT B d' *A. thaliana.* 158
Figure 43: Schéma basé sur les prédictions bioinformatiques des gènes de la FuT46109, FuT46110 et FuT54599 de *P. tricornutum*, ainsi que leur protéine putative. 159
Figure 44 : Alignement réalisé avec le logiciel MultAlin de la séquence théorique FuT46110 de *P. tricornutum*, avec celle générée chez les vertébrés par le logiciel GENSCAN. 160
Figure 45 : Résultat des prédictions du peptide signal et du domaine transmembranaire pour la séquence protéique FuT46110 prédite chez les vertébrés par le logiciel GENSCAN à partir du gène 46110 de *P. tricornutum*. 161
Figure 46 : Construction génétique des plasmides recombinants pcDNA3.1/V5-His-TOPO+46109, pcDNA3.1/V5-His-TOPO+46110 et pcDNA3.1/V5-His-TOPO+54599. 163
Figure 47 : Immunodétection sur empreinte de l'expression de la protéine recombinante FuT54599-V5/His dans les cellules CHO FuT8-/- transfectées de façon stable. 164
Figure 48 : Immunodétection sur cellules de l'expression de la protéine recombinante FuT54599-V5/His dans les clones CHO FuT8-/- transfectés de façon stable avec le gène 54599 de *P. tricornutum* par l'anticorps anti-épitope V5. 165
Figure 49 : Etude de l'activité de la protéine recombinante FuT54599-V5/His dans les cellules CHO FuT8-/- transfectées de façon stable par immunodétection sur empreinte avec l'anticorps anti-α1,3-fucose. 166

Figure 50 : Schéma récapitulatif de la voie de N-glycosylation de *Phaeodactylum tricornutum* établie à partir des données bioinformatiques et biochimiques obtenus au cours de cette thèse. 188

Liste des tableaux

Tableau 1 : Liste complète des amorces PCR utilisées dans ce projet. 102

Tableau 2 : Conditions de PCR utilisées pour l'amplification des gènes codant la GnT I et les FuT46109, FuT46110 et FuT54599 avec la polymérase Extensor HI-fidelity PCR Enzyme Mix (Thermo Scientific). 111

Tableau 3 : Composition du mélange réactionnel utilisé pour l'amplification PCR des gènes codant la GnT I et les FuT46109, FuT46110 et FuT54599 avec la polymérase Extensor HI-fidelity PCR Enzyme Mix (Thermo Scientific). 111

Tableau 4 : Composition du mélange réactionnel utilisé pour les digestions avec les enzymes de restriction Xho I, Hind III et Kpn I (Promega). 113

Tableau 5 : Composition du mélange réactionnel utilisé pour la transcription inverse avec la M-MLV RT de Promega. 118

Tableau 6 : Conditions de PCR utilisées pour l'amplification des ADNc avec la GoTaq polymérase (Promega). 119

Tableau 7 : Composition du mélange réactionnel utilisé pour l'amplification PCR des ADNc et des ARN avec la GoTaq polymérase (Promega). 119

Tableau 8 : Composition des gels de concentration et de séparation SDS-PAGE 12% 125

Liste des abréviations

2-AB : 2-aminobenzamide
α-Man : α-mannosidase
α-MEM : Minimum Essential Media
A : adénosine
Act : actine
ADN : Acide DésoxyriboNucléique
ADNc : Acide DésoxyriboNucléique complémentaire
ALG : Asparagine Linked Glycosylation
Anti-Fuc : Anticorps anti-α(1,3)-fucose
Anti-Xyl : Anticorps anti-β(1,2)-xylose
ARN = RNA : Acide RiboNucléique
ARNm : Acide RiboNucléique messager
Asn : Asparagine
BCIP : 5-bromo-4-chloro-3-indolyl-phosphate
BET : Bromure d'éthidium
BLAST : Basic Local Alignment Search Tool
b_R : Pentes des courbes standard du gène de référence.
BSA : Bovine serum albumin
b_X : Pentes des courbes standard du gène cible
C : Cytidine
Ca^{2+} : Ion calcium
CDG : Anomalie congénitale de la glycosylation
CHO : Chinese Hamster Ovary
CHO Lec1 : Cellules d'ovaires de hamster chinois mutées sur le gène GnT I.
CO_2 : Dioxyde de carbone
Con A : Concanavaline A
C_T : Cycles seuil
CTP : Cytidine triphosphate
$C_{T,R}$: C_T du gène de référence
CTS : Cytoplasmic-transmembrane-stem
$C_{T,X}$: C_T du gène cible
Cys : Cystéine
DHB : Acide 2,5-dihydroxybenzoïque
DMB : 1,2-diamino-4,5-methylene dioxybenzene
DNase : Désoxyribonucléase
DO : Densité optique
Dol-P : Dolichol phosphate
Dol-PP : Dolichol pyrophosphate
DPBS : Dulbecco`s Phosphate Buffered Saline
DTT : Dithiothreitol

ECA : Erythrina Crista Galli Agglutinine
EDEM : ER Degradation-Enhancing Mannosidases
Endo H : Endoglycosidase H
ENGases : endo-N-acetyl-β-D-glucosaminidases
EPA : Acide eicosapentaénoïque
EPS : Extra Polymeric Substances (substances extracellulaires polymériques)
ERAD : ER associated protein degradation
EST : Expressed Sequence Tag
ET : Eléments transposables
Fuc : Fucose
FuT (= FucT) : fucosyltransférase
GalNAc : N-acétylgalactosamine
G : Guanosine
GDP : Guanosine diphosphate
GH : Glycosylhydrolases
Glc : Glucose
GlcNAc : N-acétylglucosamine
GnT : N-acétylglucosaminyltransférase
GT : Glycosyltransférases
H4 : Histone 4
HEK : Human Embryonic Kidney
Hex : Hexosaminidases
HRP : Peroxydase de raifort
kb : kilobase
kDa : kiloDalton
M-Pol : mannane polymérase
Ma : Millions d'années
MALDI-TOF : Matrix Assisted Laser Desorption Ionisation-Time Of Flight
Man : Mannose
Man-8 : Man8GlcNAc2
Man-9 : Man9GlcNAc2
MEC : Matrice extracellulaire
miRNA : micro ARN
mL : Millilitre
mM : Millimolaire
Mn2+ : Ion manganèse
Mpb : Méga paires de bases
MR : Poid moléculaire des amplicons du gène de référence (R0).
MRH : Mannose 6-phosphate receptor homolog
MX : Poids moléculaires des amplicons du gène cible (X0)
NaCl : Chlorure de sodium
NaOH : Hydroxyde de sodium
NBT : Nitro blue tetrazolium
nm : nanomètre

NO2- : Ion nitrite
NO3- : Ion nitrate
NH4+ : Ion ammonium
O2 : Dioxygène
ORF : Cadre ouvert de lecture
OST : OligoSaccharylTransférase
P. tricornutum = Pt : Phaeodactylum tricornutum
pb : paires de bases
PBS : Phosphate Buffered Saline
PCR : Polymerase Chain Reaction
PHA-E et L : Phytohémagglutinine E et L
PHB : Poly-3-hydroxybutyrate
PLA2 : Phospholipase A2
PNA : Peanut agglutinin ou Agglutinine de cacahuète
PNGase : Peptide N-Glycosidase
PP-Dol : Dolichol PyroPhosphate
PPN : Production primaire nette
(R) : Gène de référence
RCA : Ricinus Communis Agglutinine
RE = ER : Réticulum endoplasmique
RNase : Ribonucléase
RNA-seq : RNA sequencing
RPS : Small subunit ribosomal protein
RT : Reverse Transcription ou transcription inverse
RT-PCR : Reverse transcription PCR
RT-qPCR = Q-PCR : PCR quantitative en temps réel
SAR : Straménopiles-Alvéolates-Rhizaria
SDS : Dodécylsulfate de sodium
SDS-PAGE : Sodium dodécyl sulfate poly acrylamide gel electrophoresis
Ser : Sérine
SialT=ST : Sialyltransférase
siRNA : Petit ARN interférent
SVF : Sérum de veau fœtal
T : Thymidine
TBE : Tris-Borate-EDTA
TBS : Tris-Buffered Saline
TBST : TBS + Tween 20
TE : Tris-EDTA
TEMED : Tetramethylethylenediamine
TFA : acide trifluoroacétique
TGN : Réseau trans-golgien
THG : Transferts horizontaux de gènes
Thr : Thréonine
TPBS : PBS + Tween

Tris-HCL : Tris(hydroxymethyl)aminomethane hydrochloride
UDP : Uracile DiPhosphate
UGGT : UDP-glucose : glycoprotéine glucosyltransférase
VIH : Virus de l'Immunodéficience Humaine
VWR : Membrane de nitrocellulose
(X) : Gène cible
Xyl : Xylose
XylT : Xylosyltransférase

Introduction générale

La N-glycosylation est une importante modification co- et post-traductionnelle des protéines chez les eucaryotes. Elle permet l'ajout et la maturation de glycannes N-liés sur une protéine empruntant le système endomembranaire de sécrétion. Cette modification est essentielle lors de l'acquisition de la structure d'une glycoprotéine. Bien que les premières étapes de ce processus soient universellement retrouvées chez les eucaryotes, les N-glycannes matures résultants se sont diversifiés au cours de l'évolution donnant lieu à une grande variété de structures organisme-spécifiques. En effet, au niveau du réticulum endoplasmique, la synthèse et le transfert du précurseur oligosaccharidique sur la protéine, ainsi que sa maturation en N-glycannes oligomannosidiques, semblent conservés chez tous les eucaryotes étudiés à ce jour. En revanche, la suite de la maturation des N-glycannes, ayant essentiellement lieu dans l'appareil de golgi, divergent en fonction du type d'organisme. Ainsi, il a été montré que les levures présentent des structures glycanniques polymannosylées, contrairement aux eucaryotes supérieurs tels que les mammifères, les plantes et les invertébrés (insectes et vers) qui possèdent des N-glycannes matures appelés complexes. La première étape de maturation des N-glycannes oligomanosidiques en N-glycannes complexes a lieu dans l'appareil de Golgi sous le contrôle de la N-acétylglucosaminyltransférase I (GnT I). Cette enzyme, permettant l'ajout d'un résidu N-acétylglucosamine sur un glycanne oligomannosidique contenant 5 mannoses, est l'enzyme clé de la voie des N-glycannes complexes. Ces derniers sont très importants car ils permettent à la protéine d'acquérir des fonctions supplémentaires et seraient en partie responsable de la survie et de la longévité des vertébrés et invertébrés au cours de l'évolution. Suite à la GnT I, diverses glycosyltransférases et glycosidases peuvent agir aboutissant à la formation d'une grande variété de N-glycannes complexes matures. Or c'est l'action d'un répertoire bien particulier de ces enzymes qui va conférer aux organismes leurs N-glycannes complexes spécifiques. La fucosylation des N-glycannes complexes est l'une des principales modifications à l'origine de cette diversité.

La N-glycosylation fait depuis de nombreuses années l'objet de recherches afin de mieux comprendre ce processus non seulement d'un point de vue fondamental mais également dans un but de valorisation économique. En effet, la grande majorité des protéines présentant un intérêt thérapeutique sont en réalité des protéines N-glycosylées. Or, la N-glycosylation, indispensable pour l'obtention d'une conformation fonctionnelle chez ces protèines, rend ces molécules bien trop complexes pour être produites par simple synthèse

chimique. Dès lors, seule une synthèse *in vivo* de ces protéines est possible et envisageable pour le moment. C'est pourquoi, depuis une 30^{aine} d'années, une course internationale est menée afin de trouver le meilleur système vivant pouvant être utilisé comme usine cellulaire pour la production de glycoprotéines à intérêt thérapeutique.

C'est dans ce contexte et fort de son expérience sur l'étude de la N-glycosylation chez les plantes supérieures, que le laboratoire Glyco-MEV a initié, depuis quelques années, une thématique de recherche dirigée par le Dr. Muriel Bardor, visant à élucider les mécanismes de N-glycosylation chez les microalgues.

Les microalgues sont des organismes microscopiques et photosynthétiques, faisant partie des premiers êtres vivants apparus sur la Terre. Bien qu'elles représentent moins de 1 % de la biomasse photosynthétique totale, elles sont responsables de plus de 45% de la production primaire nette (PPN) globale et produisent la moitié de l'oxygène de notre planète. Les microalgues constituent un vaste ensemble d'organismes polyphylétiques parmi lesquels les diatomées représentent l'un des groupes les plus diversifiés à l'importance écologique majeure. En effet, ces microalgues brunes, réparties sur l'ensemble de la planète, constituent la base du phytoplancton et donc de la chaine alimentaire marine. De plus, leurs aptitudes biologiques en font des acteurs principaux dans bon nombre d'évènements biogéochimiques tels que la fixation du CO_2 et la production de l'O_2.

Certaines de leurs particularités physiologiques et métaboliques suscitent également un intérêt pour l'industrie. Ainsi, les diatomées sont déjà utilisées pour la production de diverses molécules à haute valeur ajoutée, comme des antioxydants, des pigments et des polysaccharides. Leur importante teneur en acide gras permet également de les envisager comme potentielle source de biocarburants. De même, leur facilité à être cultivée et leur possible transformation génétique en font des candidats sérieux pour la production de protéines à intérêt thérapeutique.

Parallèlement, la position phylogénétique des diatomées, bien distincte des plantes et des animaux, rendent ces organismes très intéressants pour comprendre l'évolution de certaines voies métaboliques, telle que la N-glycosylation. Ces études sont aujourd'hui grandement facilitées par l'accessibilité à la séquence du génome de certaines diatomées comme *Thalassiosira pseudonana* et *Phaeodactylum tricornutum* ainsi que par le développement d'outils de biologie moléculaire.

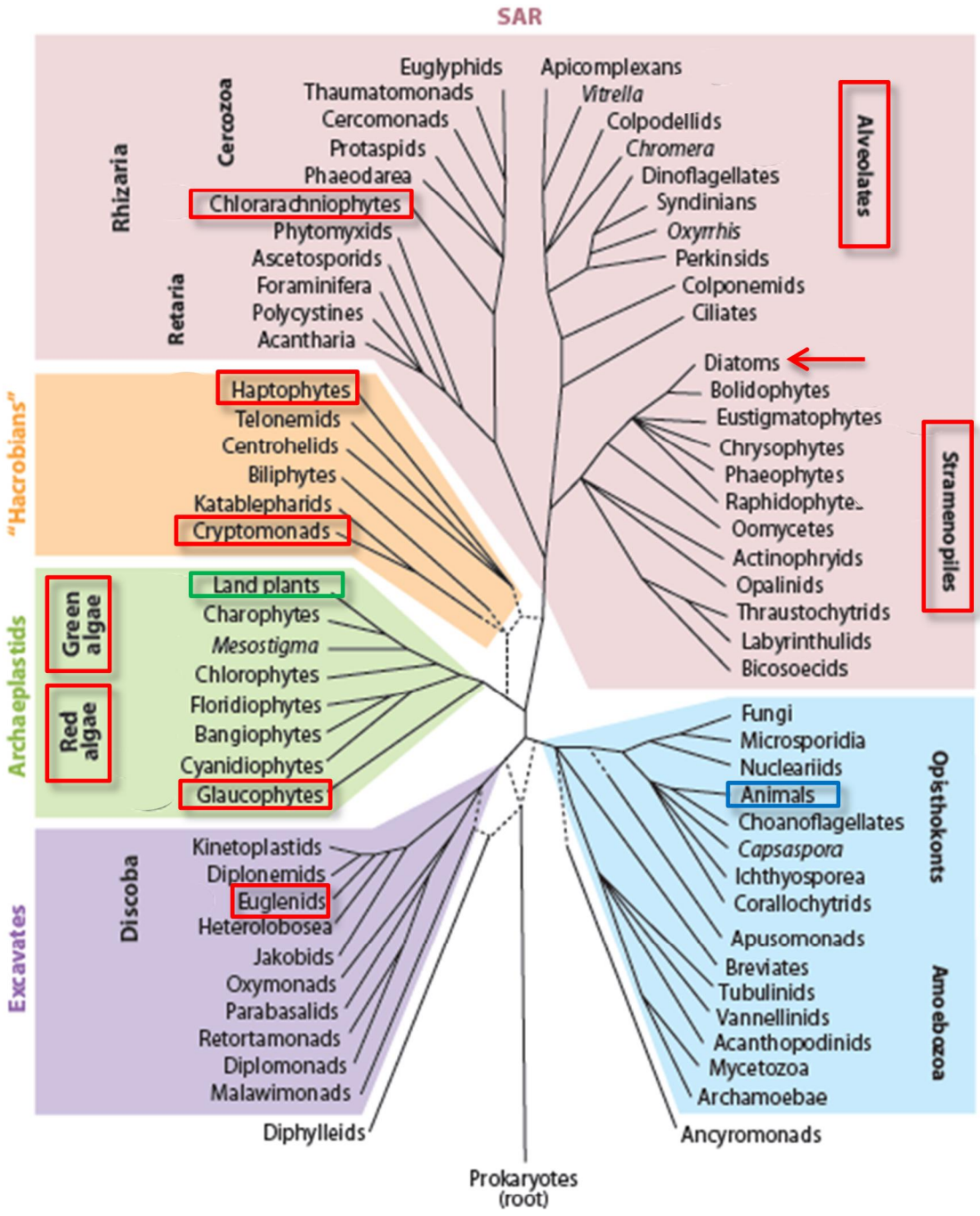
Ce manuscrit présente le travail qui a été accompli au cours de ma thèse, visant à étudier la voie de N-glycosylation de *Phaeodactylum tricornutum* et à réaliser la caractérisation fonctionnelle de la GnT I de cette diatomée.

Le premier chapitre consiste en une synthèse bibliographique divisé en deux grandes parties. La première partie présente les diatomées, leur origine, leur mode de vie, leur structure, leurs applications, ainsi que le cas particulier du modèle d'étude choisi par le laboratoire, la microalgue *Phaeodactylum tricornutum*. La seconde partie concerne la voie de N-glycosylation, les différentes étapes de ce processus, les variations organisme-spécifiques, les fonctions attribuées aux N-glycannes et une description plus détaillée de la GnT I.

Après un second chapitre consacré aux matériels et méthodes utilisés, les résultats obtenus seront exposés dans un troisième chapitre divisé en 3 parties. La première partie sera présentée sous forme d'un article scientifique décrivant la voie de N-glycosylation chez *Phaeodactylum tricornutum* établie à partir d'analyses bioinformatiques et biochimiques, ainsi que la caractérisation fonctionnelle de sa Gn T I réalisée par complémentation dans des cellules d'ovaire de hamster chinois mutantes présentant une déficience en activité GnT I endogène. La seconde partie sera consacrée aux résultats complémentaires obtenus au cours de cette thèse, avec dans un premier temps, l'étude de l'expression du gène de la GnT I en fonction du morphotype, de la souche et des conditions en nitrate dans le milieu de culture. Dans un second temps, les résultats préliminaires de la caractérisation des 3 fucosyltransférases de *Phaeodactylum tricornutum* seront présentés.

Un quatrième chapitre sera consacré à une discussion générale, et suivi d'une conclusion et des perspectives.

Chapitre I : Synthèse Bibliographique



<u>Figure 1</u> : Arbre phylogénétique des lignées eucaryotes d'après Keeling, 2013.
Les relations entre grandes lignées eucaryotes ont été réalisées à partir d'études basées sur des analyses phylogénomiques ou sur des caractéristiques génomiques rares. Les lignes pointillées indiquent des relations qui font l'objet de litiges dans la communauté scientifique. Dans certains cas, deux possibilités contradictoires sont affichées. Les principaux groupes de microalgues sont encadrés en rouge et répartis dans quatre supergroupes différents : le supergroupe des Straménopiles-Alvéolates-Rhizaria (SAR), les Hacrobians (supergoupe encore contesté contenant les haptophytes et les cryptomonades, qui formaient initialement les Chromalvéolates avec les Straménopiles et les Alvéolates), les Excavates (contenant les euglènes) et les Archaeplastides ou Plantae (contenant les algues rouges, les algues vertes et les glaucophytes). Les plantes terrestres et les animaux sont encadrés respectivement en vert et bleu. La flèche rouge, présente dans le phylum des Straménopiles, indique la lignée des diatomées au sein de laquelle se classe *Phaeodactylum tricornutum*.

I- Modèle d'étude : les diatomées.

I-1- Microalgues.

I-1-1- Position phylogénétique.

C'est dans une atmosphère terrestre primitive, majoritairement composée de gaz carboniques, que sont apparus les premiers organismes photosynthétiques appelés microalgues. Leur réussite adaptative et leur prolifération ont permis l'accumulation de l'oxygène atmosphérique, indispensable au développement des espèces animales.

Le terme « microalgue » définit un groupe polyphylétique d'organismes microscopiques et photosynthétiques très divers allant des cyanobactéries aux microalgues eucaryotes. Apparues il y a environ 3,5 milliards d'années, les cyanobactéries sont des organismes procaryotes unicellulaires, également appelées microalgues bleues en raison de leur coloration bleu-vert. Les microalgues eucaryotes regroupent, quant à elles, des organismes très variés aussi bien uni- que pluricellulaires répartis dans des phylums différents (**figure 1** ; Keeling, 2013). Elles rassemblent :

- les microalgues rouges, les microalgues vertes et les glaucophytes qui appartiennent aux Plantae (ou Archaeplastides), supergroupe contenant également les plantes terrestres.
- les Straménopiles, les Alvéolates et les Chlorarachniophytes (appartenant aux Rhizaria) qui forment le supergroupe SAR (Burki et al., 2007 ; Hackett et al., 2007 ; Baurain et al., 2010).
- ainsi que les Haptophytes et les Cryptomonades qui appartiennent aux Hacrobians (Patron et al., 2007 ; Okamoto et al., 2009). L'existence de ce supergroupe fait l'objet d'un profond désaccord dans la communauté scientifique (Burki et al., 2012). Initialement, ces organismes étaient placés avec les Alvéolates et les Straménopiles formant le supergroupe des Chromalvéolates (Cavalier-Smith, 1999).

Les microalgues sont donc réparties dans les différents embranchements de l'arbre phylogénétique des eucaryotes à l'exception des Unikonts, supergroupe auquel appartiennent les animaux (Keeling, 2013).

Les diatomées ou Bacillariophycées, appartenant aux Straménopiles (**figure 1**), représentent l'un des groupes phytoplanctoniques les plus diversifiés avec plus de 200 genres et 100 000 à 200 000 espèces estimées (Round et al., 1990 ; Mann, 1999). Elles sont généralement divisées en deux groupes que sont les diatomées centriques et les diatomées pennées. Leur position phylogénétique, bien distincte des plantes terrestres et des animaux, en fait des organismes très intéressants pour l'étude de l'évolution des différentes voies métaboliques (**figure 1**).

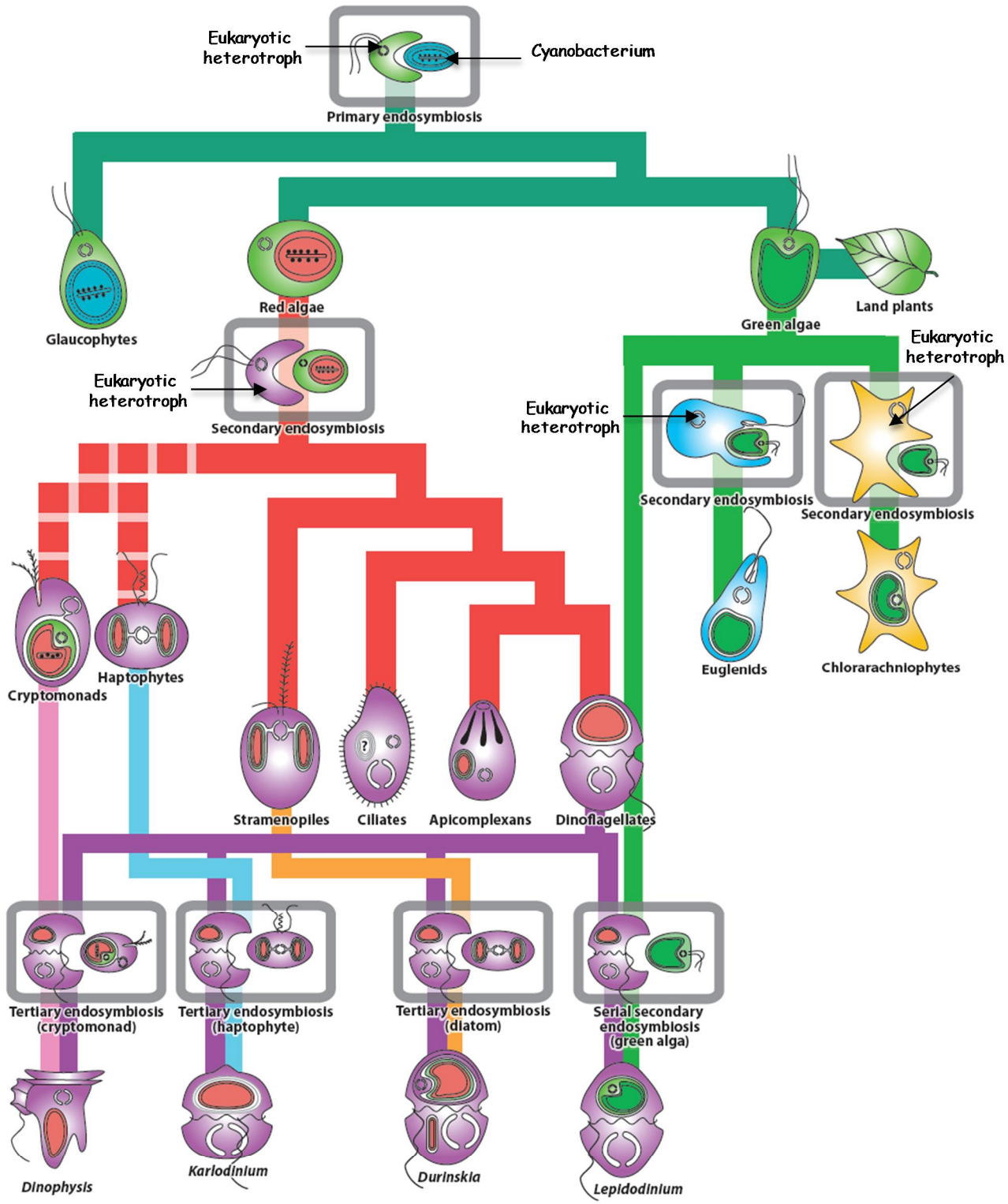


Figure 2 : Origine endosymbiotique des microalgues d'après Keeling, 2013.
Les encadrés gris représentent les évènements endosymbiotiques primaires, secondaires et tertiaires. L'incertitude de l'origine commune des haptophytes et des cryptophytes avec les straménopiles est représentée par une ligne d'embranchement en pointillé rouge.

I-1-2- Origine endosymbiotique.

Bien que les microalgues aient une histoire évolutive différente, elles semblent présenter une origine endosymbiotique commune. Actuellement, plusieurs théories évolutives, s'appuyant sur des données génomiques, biochimiques ou structurales, coexistent. Les principaux désaccords portent essentiellement sur la datation de certains évènements et sur l'embranchement de quelques espèces. Cependant, la grande majorité des scientifiques s'accorde sur l'hypothèse de l'origine procaryotique et symbiotique des plastes qui aurait aboutie à la formation des microalgues (Gibbs, 1981 ; Margulis, 1992).

La photosynthèse oxygénique, apparue chez les cyanobactéries il y a plus de 2,8 milliards d'années (Brocks et al., 1999), aurait été transférée chez les eucaryotes par une succession d'évènements endosymbiotiques. En effet, les organismes appartenant au supergroupe des Plantae, tels que les algues vertes (ou Chlorophytes) et rouges (ou Rhodophytes), seraient issues de l'absorption et de la rétention d'une cyanobactérie par un organisme eucaryote hétérotrophe, il y a plus de 1,5 milliards d'années (**figure 2** ; McFadden, 2001 ; Yoon et al., 2004 ; Armbrust, 2009). Cette endosymbiose initiale a été l'un des événements les plus importants dans l'évolution de la vie.

Peu de temps après, les algues rouges et vertes ont divergées. Une endosymbiose secondaire serait alors rapidement survenue au cours de laquelle un autre organisme eucaryote hétérotrophe aurait intégré une algue rouge. Cet événement endosymbiotique secondaire aurait eu lieu il y a environ 1,3 milliards d'années. Il serait à l'origine de l'ancêtre commun des Chromistes dont font partie les cryptophytes, les haptophytes et les hétérokontes (ou Straménopiles) auxquels appartiennent les diatomées (**figure 2** ; Yoon et al., 2004 ; Keeling, 2009 ; Baurain et al., 2010).

De même, les euglènes et les chlorarachniophytes seraient issus d'endosymbioses secondaires similaires ayant eu lieu entre un organisme eucaryote hétérotrophe et une algue verte (**figure 2**). Enfin, des évènements d'endosymbiose tertiaire seraient à l'origine des dinophytes (**figure 2**).

Cependant, des incertitudes persistent quant à l'ordre d'apparition et l'indépendance de certains évènements endosymbiotiques. En particulier, les analyses actuelles tendent à montrer que les plastes des haptophytes, des cryptophytes et des straménopiles dériveraient d'endosymbiotes secondaires indépendants (Armbrust, 2009 ; Keeling, 2013).

I-2- Diatomées.

I-2-1- Biologie des diatomées.

I-2-1-1- Morphologie et cytologie.

Les diatomées sont des algues eucaryotes microscopiques allant de quelques micromètres à un demi-millimètre pour les plus grandes. Elles présentent la particularité de posséder un squelette externe rigide et transparent appelé frustule. Cette enveloppe, composée de silice faiblement cristallisée associée à des composants organiques, confère aux diatomées une grande résistance face aux contraintes mécaniques et chimiques pouvant être exercées par leur environnement (Kroger et Poulsen, 2008). Le frustule est constitué de deux valves que sont l'épivalve et l'hypovalve réunies par un ensemble de ceintures connectives (**figure 3A** ; Falciatore et Bowler, 2002). L'hypovalve, qui possède une taille légèrement plus petite, est imbriquée dans l'épivalve à la façon d'une boite de pétri. Chaque valve est finement ornée de pores ou aéroles, d'épines, de stries, de côtes, ou tout autre ornement formant des motifs nanométriques pouvant être assimilés à de la dentelle (**figure 3B** ; Round et al., 1990 ; Losic et al., 2006). Certaines diatomées présentent également une fente longitudinale au niveau de chacune de leurs valves, appelée raphé (**figure 3A**). L'organisation de ces motifs est très variée et est utilisée comme critère de détermination des espèces.

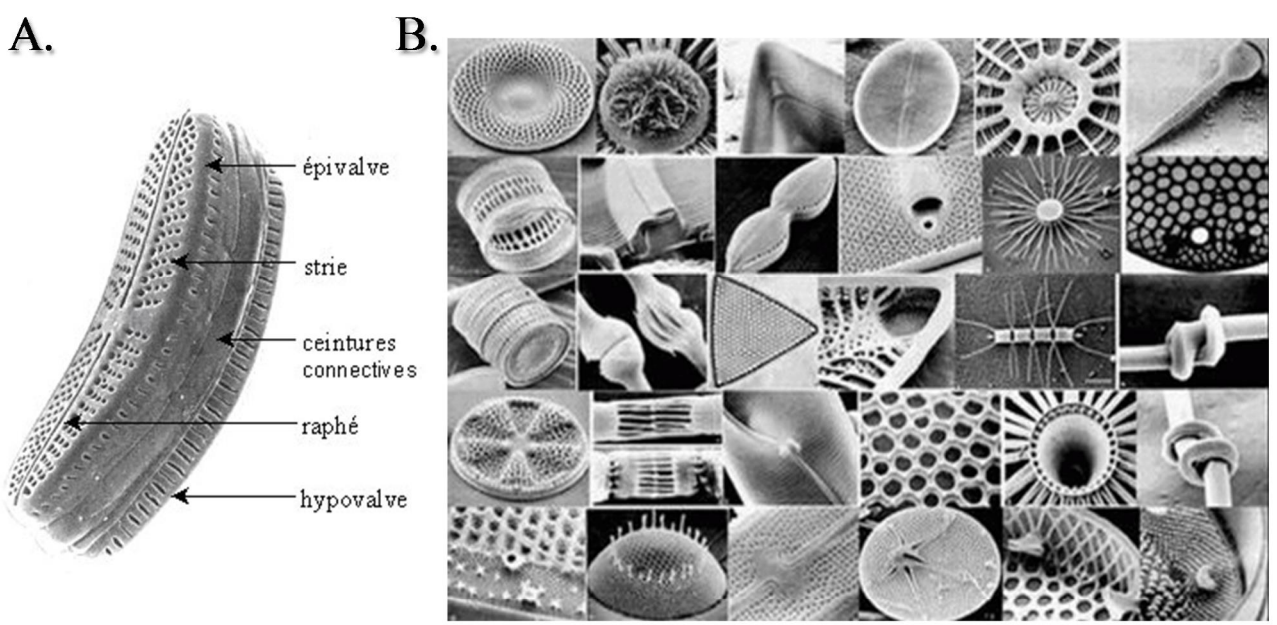


Figure 3 : Structure et diversité du frustule des diatomées.
A. = Détail de la structure du frustule d'une diatomée (*Achnantes*). Photographie en microscopie électronique à balayage avec annotations, extraite du site www.pays-de-la-loire.developpement-durable.gouv.fr.
B. = Diversité ornementale du frustule de différentes diatomées. Photographies en microscopie électronique à balayage issues de (Round et al., 1990).

La cohérence de cette enveloppe est renforcée par une matrice organique extracellulaire. En effet, les diatomées excrètent des substances extracellulaires polymériques (EPS) au travers des perforations présentes sur toute la surface du frustule. Les EPS peuvent s'organiser sous différentes formes comme des tubes, des pédoncules, des fibrilles ou bien encore former une enveloppe autour du frustule. L'excrétion de telles substances mucilagineuses joue un rôle important dans leur mobilité, leur adhésion au substrat, leur reproduction et dans la constitution et la protection des colonies (Davidovich et al., 2012 ; Wang et al., 2013).

La paroi siliceuse des diatomées protège un protoplasme qui, comme dans tous les organismes eucaryotes, présente des organites différenciés et structurés : noyau, réticulum endoplasmique, appareil de Golgi, mitochondries, peroxysomes, vacuoles, plastes…(**figure 4** ; Borowitzka et Volcani, 1978 ; Hoops et Floyd, 1979 ; De Martino et al., 2011).

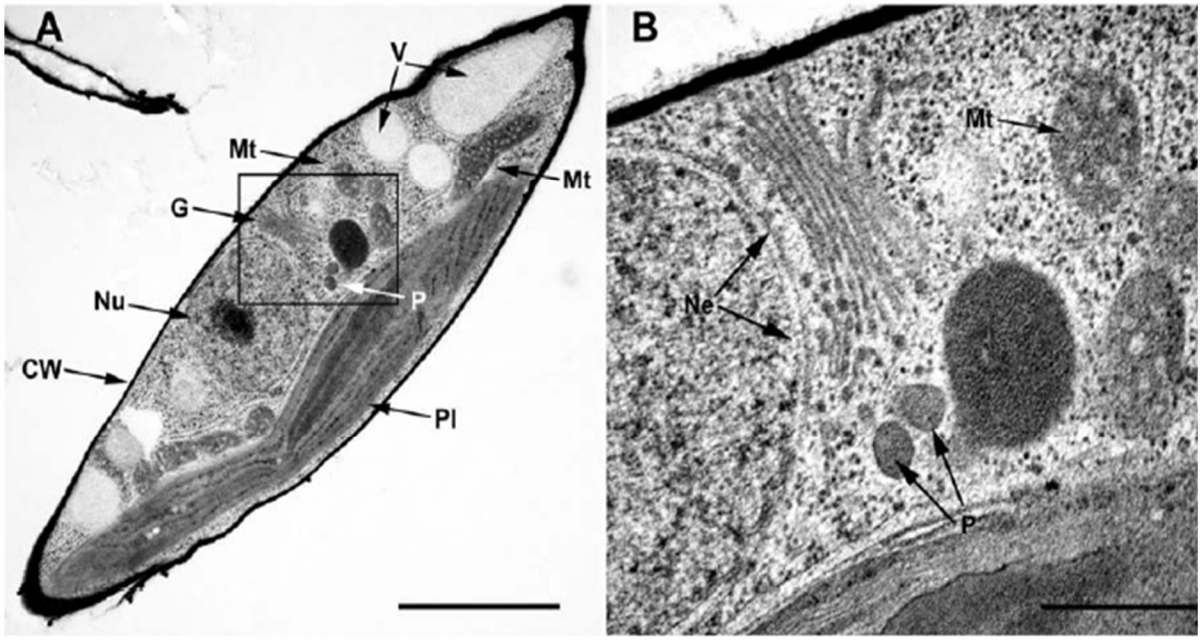


Figure 4 : Ultrastructure d'une diatomée.
Photographies prises en microscopie électronique à transmission présentant l'ultrastructure de la diatomée *P. tricornutum* (Gonzalez et al., 2011). La zone encadrée correspond à un grossissement supérieur du contenu cytoplasmique présenté en (B). CW : paroi ; G : appareil de Golgi ; Mt : mitochondrie ; Ne : enveloppe nucléaire ; Nu : noyau ; P : peroxysome ; Pl : plaste ; V : vacuole. La barre d'échelle noire représente 2 µm pour (A) et 500 nm pour (B).

Les chloroplastes des diatomées renferment différents pigments tels que la chlorophylle a et c et une importante quantité de caroténoïdes (β-carotène, fucoxanthine, diatoxanthine, diadinoxanthine) ce qui leur confère une couleur pouvant varier du jaune très pâle au brun et leur assure un large spectre d'absorption de la lumière. L'organisation de ces plastes résulte de leur origine endosymbiotique (Armbrust, 2009). En effet, les diatomées possèdent des plastes secondaires comportant quatre membranes (McFadden, 1999 ; Keeling, 2013). A la vue des données sur la topologie et la composition biochimique des membranes chloroplastiques, il semblerait que les deux membranes les plus internes correspondent à celles de la cyanobactérie phagocytée au cours de la première endosymbiose (Cavalier-Smith, 1982). La membrane du premier hôte eucaryote aurait donc été perdue faisant de la cyanobactérie un organite directement immergé dans le cytoplasme. La troisième membrane serait, quant à elle, la membrane de l'algue rouge résultant de cet évènement endosymbiotique primaire, suite à son intégration dans un deuxième organisme eucaryote lors de l'endosymbiose secondaire. Enfin, la membrane la plus externe correspondrait à celle de ce dernier organisme hôte (McFadden, 1999). Ainsi, contrairement, au premier évènement endosymbiotique, la membrane du second hôte eucaryote n'a pas été perdue.

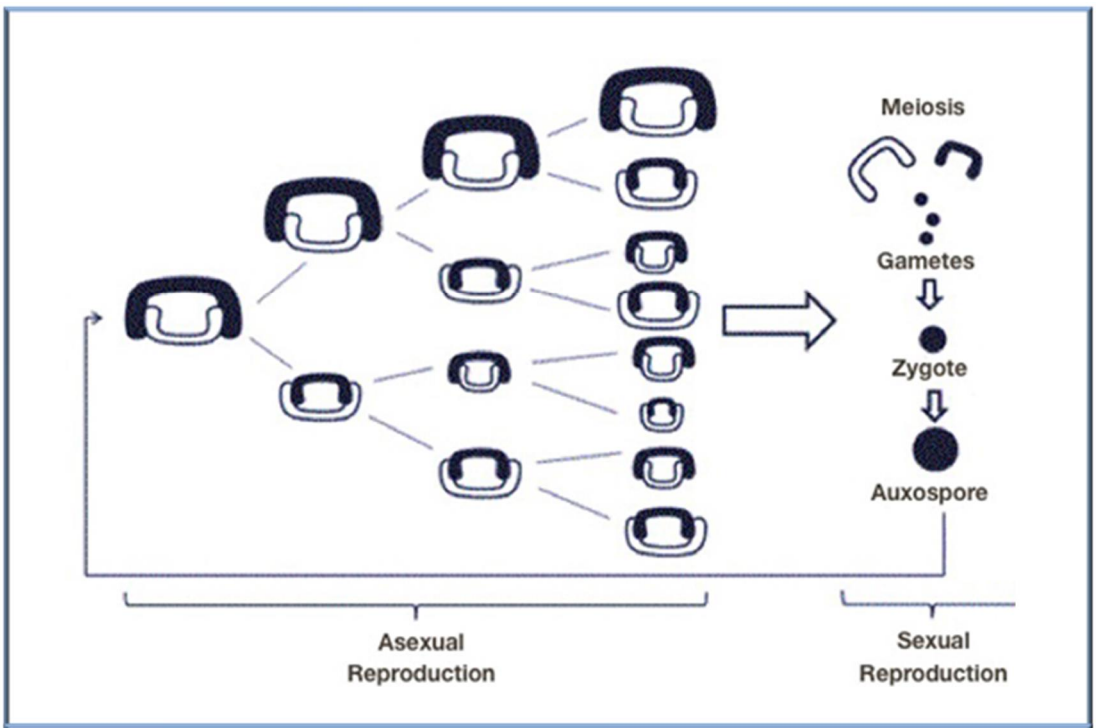
I-2-1-2- Reproduction.

Le cycle de développement des diatomées est essentiellement diplophasique et relativement rapide, allant de quelques heures à quelques jours. Les diatomées se reproduisent principalement par multiplication végétative au cours de laquelle chaque cellule donne naissance à deux cellules filles (Round et al., 1990). Ainsi, lorsque les conditions sont favorables à leur prolifération, les diatomées se multiplient par mitose formant des populations monoclonales où tous les individus possèdent le même génotype (Kroger et Poulsen, 2008). Lors de cette mitose, chaque cellule fille conserve une valve de la cellule mère et en synthétise une nouvelle plus petite que la première (**figure 5**) (Falciatore et Bowler, 2002). Une fois le frustule achevé, les deux cellules filles se séparent ou restent associées dans le cas des espèces coloniales. Au fil des générations, une réduction de la taille des cellules est donc observée. Cependant, cette diminution n'est pas infinie. En effet, ces évènements mitotiques, se déroulant pendant plusieurs mois voire plusieurs années, ont lieu jusqu'à ce que les cellules atteignent une taille minimale correspondant à 30 % de leur taille initiale (**figure 5**).

Au-dessous de cette taille minimale, appelée taille critique, démarre une phase de reproduction sexuée au cours de laquelle les cellules entrent en méiose et produisent des gamètes (**figure 5** ; Mann, 1993). La reproduction sexuée, qui n'est pas uniquement initiée par la taille des diatomées mais également par des facteurs environnementaux (lumière, température, nature du substrat,…), permet l'obtention d'un nouvel individu souche de taille « normale » (Armbrust et al., 1990). Le fonctionnement de cette reproduction est complexe et diffère entre les diatomées centriques et pennées (Round et al., 1990).

Chez les diatomées centriques, les cellules se transforment soit en gamètes femelles soit en gamètes mâles. Ces derniers possèdent un flagelle avec lequel ils migrent vers les gamètes femelles afin de les féconder (Manton et al., 1970 ; Jensen et al., 2003). A l'inverse, les gamètes mâles des diatomées pennées ne possèdent pas de flagelle. Dans ce cas, la rencontre et l'adhésion cellulaire entre deux cellules sont alors facilitées par la production de mucilage (Davidovich et al., 2012). Lorsqu'elles sont ainsi en contact, les deux cellules évoluent alors chacune en un ou deux gamètes actifs (mâles), ou passifs (femelles) qui fusionnent ensuite par divers moyens (Chepurnov et Mann, 2004).

Toutefois, que ce soit chez les diatomées centriques ou pennées, la fusion de deux gamètes engendre la formation d'un zygote appelé auxospore (**figure 5** ; Davidovich et Bates, 1998). Ce zygote croît jusqu'à atteindre une taille maximale qui est propre à chaque espèce, avant de former un nouveau frustule et repartir dans un nouveau cycle de multiplication végétative (**figure 5**).



<u>Figure 5</u> : Cycle de vie d'une diatomée : multiplication végétative et reproduction sexuée (schéma de Russell Habermann extrait du site http://www.seagrant.umn.edu/newsletter/2013/01/what_good_is_a_diatom.htm).

I-2-2- Classification et histoire des diatomées.

Comme évoqué dans les paragraphes précédents, les diatomées sont traditionnellement divisées en deux catégories : les centriques et les pennées (Round et al., 1990). Elles se distinguent essentiellement par les caractéristiques de leur frustule comme la forme, l'agencement des ornements et en particulier la symétrie de la cellule (radiale pour les centriques et latérales pour les pennées) ainsi que par leur mode de reproduction sexuée lié principalement à la forme de leurs gamètes (petits gamètes flagellés chez les centriques et larges gamètes non flagellés chez les pennées (Round et al., 1990 ; Davidovich et al., 2012)). Ces deux catégories sont elles-mêmes subdivisées en deux groupes que sont les radiales et bi-/multi-polaires pour les centriques, et les raphidées et araphidées (qui ne présentent pas de raphé) pour les pennées (**figure 6** ; Kroger et Poulsen, 2008).

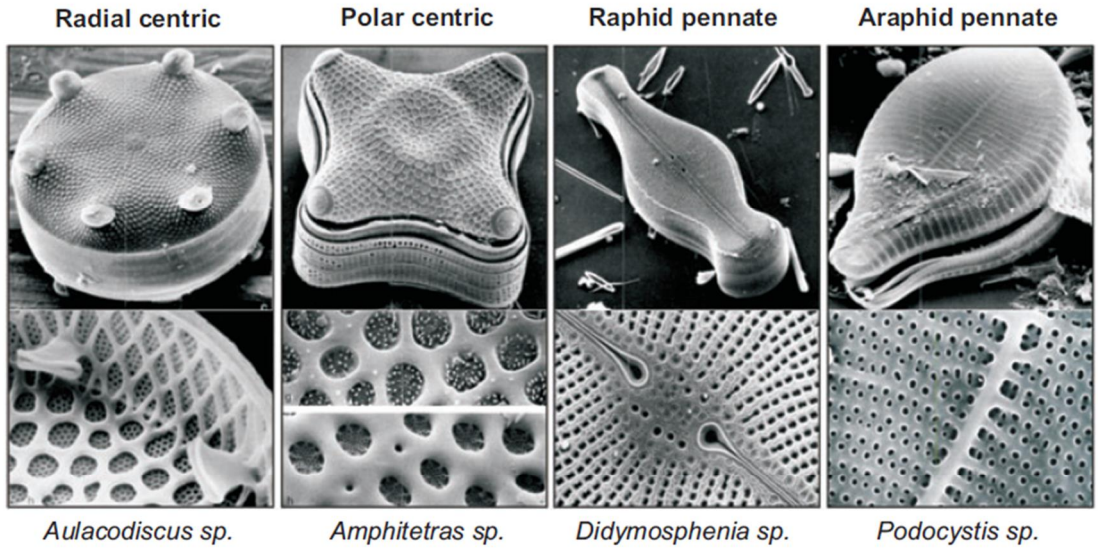


Figure 6 : Photographies prises en microscopie électronique à balayage des quatre classes de diatomées : centrique radiale, centrique polaire, pennée raphidée et pennée araphidée (Kroger et Poulsen, 2008).

Ces quatre classes de diatomées ne sont pas apparues simultanément au cours de l'évolution. Des estimations provenant à la fois d'analyses de marqueurs moléculaires et d'observations stratigraphiques indiquent que les diatomées auraient émergées au début du Jurassique, il y a environ 200 millions d'années (Ma ; Medlin, 2011 ; Medlin et al., 1997). En effet, les plus anciens fossiles retrouvés à ce jour sont ceux de diatomées centriques radiales et datent de 185 Ma (Sims et al., 2006). Cependant, les analyses basées sur l'horloge moléculaire suggèrent que des formes plus ancestrales de ces diatomées auraient été présentes

dès le Triasique (Medlin, 2011). L'origine des diatomées seraient donc située entre 185 et 250 Ma (**figure 7** ; Sorhannus, 2007). Suite à une extinction massive de nombreuses espèces liée à la crise du Triasique, les diatomées centriques radiales modernes se sont fortement diversifiées (< 183 Ma). C'est également à cette époque, entre le Jurassique et le Crétacé inférieur, que les diatomées centriques bi- et multipolaires ont émergé (125 à 150 Ma) (**figure 7** ; Sorhannus, 2007). En revanche, l'ancêtre commun des diatomées pennées n'est apparu qu'à la fin du Crétacé supérieur (> 96 Ma) (Sorhannus, 2007). La lignée araphidée, datant de moins de 90 Ma, est la première lignée pennée à avoir divergé (**figure 7** ; Kooistra et Medlin, 1996 ; Armbrust, 2009). La seconde extinction massive, liée à la crise de la fin du Crétacé supérieur, a laissé de nombreuses niches écologiques vides ce qui a favorisé l'expansion des diatomées et l'apparition de la lignée pennée raphidée, il y a moins de 50 Ma (**figure 7** ; Kooistra et Medlin, 1996 ; Falkowski et al., 2004).

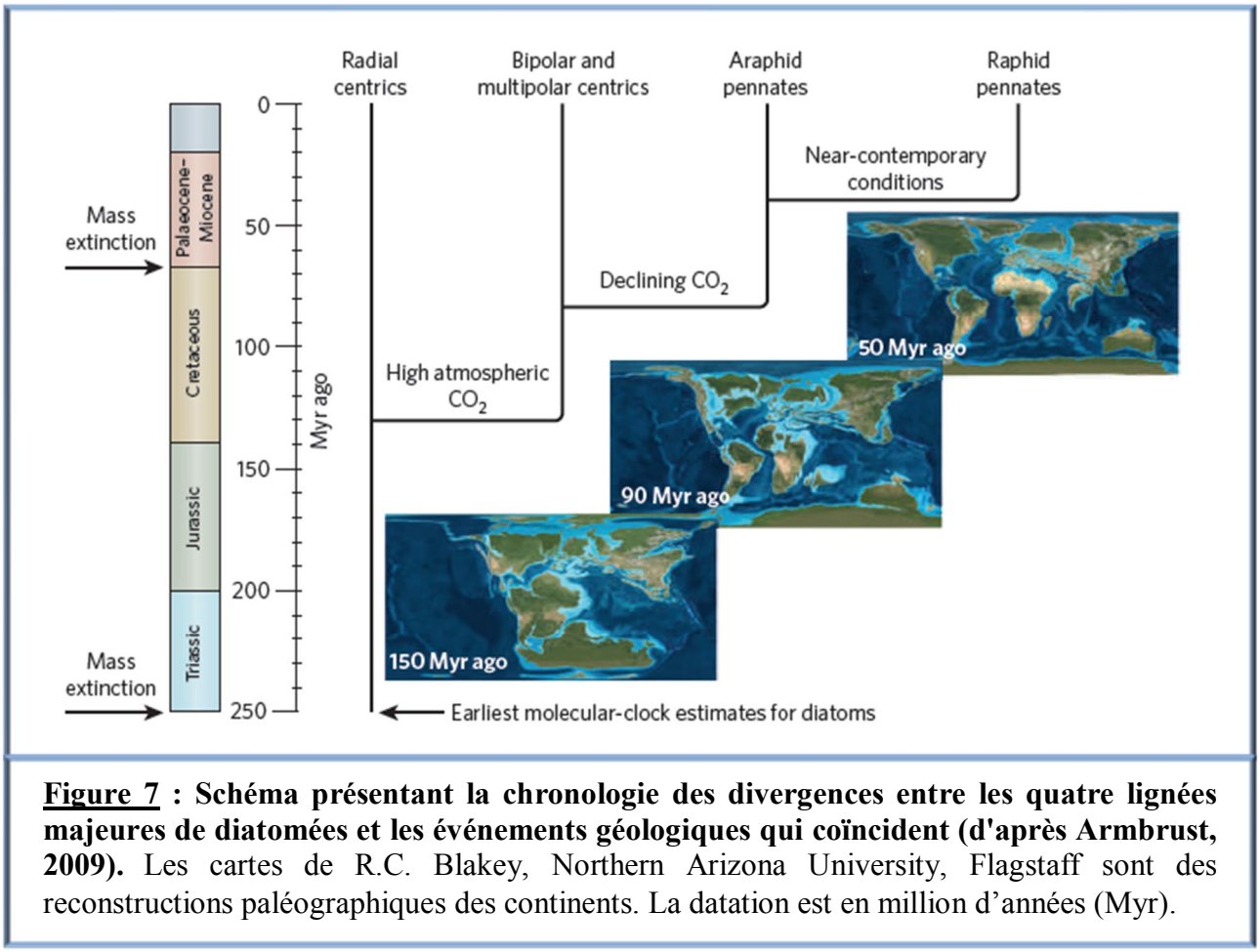


Figure 7 : Schéma présentant la chronologie des divergences entre les quatre lignées majeures de diatomées et les événements géologiques qui coïncident (d'après Armbrust, 2009). Les cartes de R.C. Blakey, Northern Arizona University, Flagstaff sont des reconstructions paléographiques des continents. La datation est en million d'années (Myr).

I-2-3- Les diatomées dans la biosphère.

I-2-3-1- Mode de vie.

Les diatomées sont des microalgues brunes cosmopolites et généralement aquatiques bien que certaines espèces soient terrestres ou aériennes. Leur développement est étroitement lié à différents facteurs environnementaux comme la température, le pH, l'intensité lumineuse, les caractéristiques hydrauliques et physicochimiques du milieu telles que la salinité, la teneur en oxygène, en nitrate, en phosphate, en silice et en fer (Yongmanitchai et Ward, 1991 ; McKay et al., 1997 ; Li et al., 2012 ; Schneider et al., 2013). Par conséquence, la prolifération des diatomées est sensible aux variations saisonnières et annuelles de leur environnement.

Cependant, leur grande capacité d'adaptation face aux variations environnementales leur ont permis de peupler un grand nombre de niches écologiques différentes. En effet, les diatomées ont colonisé des milieux aussi variés que les lacs, les rivières, les mers et les océans que ce soit dans les eaux douces, saumâtres ou salées, courantes ou stagnantes, froides ou tropicales, mais également dans des milieux plus inhospitaliers comme les glaciers, les mares temporaires, les zones fortement polluées ou les sols (paroi de grotte, boues, sol humide…) (Tolun et al., 2001 , Smol et Stoermer, 2010).

Par ailleurs, les diatomées peuvent également être impliquées dans diverses relations avec d'autres organismes. Ainsi, certaines espèces sont attachées aux plumes d'oiseaux aquatiques (Croll et Holmes, 1982), à l'épiderme de cétacés (Bennett, 1920) ou aux écailles de tortues (Wetzel et al., 2010). D'autres encore vivent comme endosymbiotes de foraminifères (Lee et al., 1995), dans des thalles de lichens (Lakatos et al., 2004), à l'intérieur d'éponges (Bavestrello et al., 2000) ou sont au contraire les hôtes de cyanobactéries (Stewart et al., 1983).

Dans le milieu aquatique, il existe deux types de diatomées que sont les espèces planctoniques et les espèces benthiques. Ces deux catégories se distinguent par leur position dans la colonne d'eau. Ainsi, les espèces dites planctoniques, concernant principalement des diatomées centriques ou pennées araphidées, sont en suspension dans le milieu et constituent la majeure partie du phytoplancton lacustre et marin. En revanche, les espèces benthiques, qui correspondent essentiellement à des diatomées pennées raphidées, évoluent à proximité ou fixées sur les fonds ou à un substrat. Cette fixation est rendue possible par la sécrétion de substances mucilagineuses au travers du raphé. Les espèces benthiques sont elles-mêmes différenciées en fonction du type de substrat auquel elles adhèrent. A titre d'exemple, les diatomées formant l'épiphyton se fixent sur des plantes aquatiques, celles de l'épilithon sur

des substrats durs comme la pierre ou les galets, celles de l'épipélon et de l'endopélon vivent libres dans les sédiments ou à leur surface.

Les diatomées sont en grande majorité autotrophes, bien que quelques rares espèces aient développé des stratégies leur permettant de s'affranchir de l'énergie lumineuse (Hellebust, 1977). Ces microorganismes unicellulaires vivent généralement de façon isolée. Cependant, certaines espèces constituent des colonies aux formes très variables : rosette, arborescence, éventail, chaînette, ruban, étoile ou dans un tube muqueux au sein duquel les diatomées sont mobiles (**figure 8**). Les diatomées libres peuvent également être douées de mouvements autonomes en particulier les pennées raphidées *(Navícula, Nitzschia).* Ces mouvements sont réalisés par sécrétion d'une substance mucilagineuse qui facilite ainsi la colonisation de nouveaux substrats et milieux (Iwasa et Shimizu, 1972 ; Gupta et Agrawal, 2007 ; Wang et al., 2013). Les diatomées les plus rapides peuvent atteindre des vitesses de 20 $\mu m.s^{-1}$ (Umemura et al., 2013).

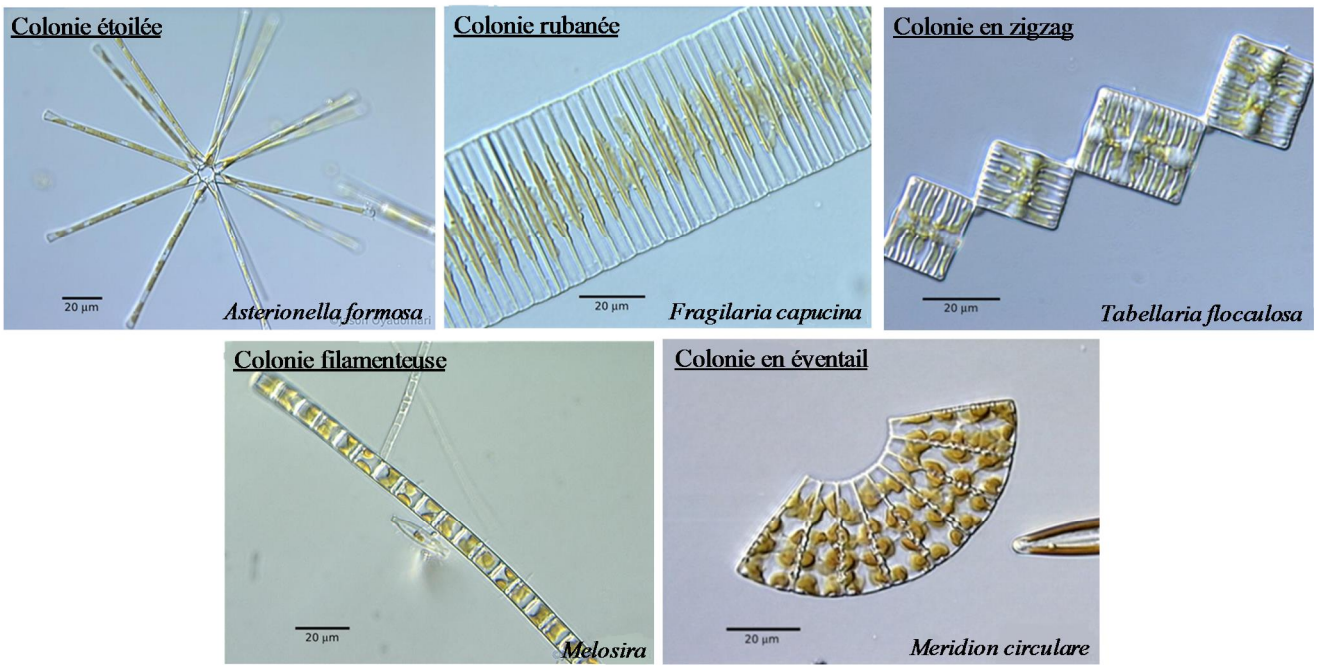


Figure 8 : Diversité des formes coloniales des diatomées (photos au microscope optique de J. Oyadomari, http://www.keweenawALGae. mtu.edu/gallery_pages/diatoms2.htm).

I-2-3-2- Rôle dans la biosphère.

Les diatomées ont une importance écologique considérable. Elles forment une part importante du phytoplancton et constituent la base de la chaîne alimentaire en milieu aquatique. Alors qu'elles représentent moins de 1% de la biomasse photosynthétique totale de la planète, les diatomées sont responsables de près de 40% de la production primaire nette (PPN) marine et de plus de 20% de la PPN mondiale (Falkowski et al., 1998 ; Field et al., 1998 ; Goldman, 1993).

Les diatomées jouent également un rôle essentiel dans de nombreux cycles biogéochimiques mondiaux. En particulier, leur capacité photosynthétique en fait des acteurs principaux dans le cycle du carbone et de l'oxygène (Kemp et al., 2006 ; Smetacek et al., 2012). Après avoir été en grande partie responsable de la formation de l'atmosphère moderne, les diatomées sont encore aujourd'hui des organismes clé dans le maintien de sa composition. En effet, elles participent activement à la formation de l'O_2 et au stockage du CO_2 atmosphérique. Les diatomées sont ainsi responsables de la production de près de 25% de l'O_2 atmosphérique, de la fixation annuelle d'un quart du carbone inorganique des océans (Falkowski et Raven JA, 1997 ; Granum et al., 2005), et de 20% du carbone fixé par voie photosynthétique à l'échelle de toute la biosphère (Field et al., 1998 ; Smetacek, 1999). Pour une même valeur de biomasse, les diatomées possèdent un pouvoir de fixation du carbone très largement supérieur aux plantes terrestres. Par exemple, dans certaines régions océaniques, les diatomées peuvent fixer la même quantité de carbone journalier qu'une forêt de végétaux terrestres, alors qu'elles ne représentent qu'une infime partie de la biomasse photosynthétique (Smetacek, 2001).

De par la composition de leur frustule, les diatomées constituent aussi un élément crucial dans le cycle de la silice (Nelson et al., 1995) (Hutchins et Bruland, 1998 ; Kemp et al., 2006). Leur besoin en silice est tel qu'elles représentent l'un des plus importants consommateurs d'acide silicique dissous dans l'océan, le faisant passer de l'hydrosphère à la biosphère (Tréguer et al., 1995). Après leur mort, une partie de la silice présente dans leur frustule est directement dissoute dans l'eau de mer et donc à nouveau disponible, alors que le reste du squelette tombe dans les fonds marin, participant ainsi à la formation de sédiments appelés diatomites (Tréguer et al., 1995).

Les diatomées participent également au cycle du fer (Sutak et al., 2012 ; Hutchins et Bruland, 1998 ; Johnson et al., 2001). Ces microalgues brunes, comme tout organisme photosynthétique, ont besoin de fer dans bon nombre de leurs voies métaboliques. Or, la concentration de cet élément est fortement limitante dans les océans (Moore et al., 2002). Par

conséquence, les diatomées se sont adaptées en développant de nouveaux mécanismes permettant de réduire leurs besoins en fer dans des conditions de carence (Allen et al., 2008), mais également en étant capable de le stocker (Raven, 2013). En effet, des études génomiques ont montré récemment que les diatomées pennées présentent la particularité de posséder une protéine de stockage du fer appelée ferritine (Marchetti et al., 2009 ; Pfaffen et al., 2013).

Il en est de même avec le cycle de l'azote, élément indispensable chez les êtres vivants puisqu'il entre dans la composition des acides nucléiques et des acides aminés (Collos, 1983 ; Weber et Deutsch, 2012). Les diatomées sont capable d'incorporer des formes inorganiques (NO_3^-, NO_2^-, NH_4^+) et organiques (urée, acides aminées) d'azote et permettent la migration verticale ascendante du nitrate dans les océans (Waser et al., 1998 ; Waser et al., 1998 ; Villareal et al., 1999 ; **figure 9**). Elles sont généralement très répandues dans les eaux riches en nitrate où elles tendent à dominer les autres populations de microorganismes, alors qu'elles sont assez rares dans les milieux qui en sont dépourvus. Cependant, dans les mers chaudes oligotrophes, il est possible de trouver des symbioses entre des cyanobactéries fixatrices d'azote et des diatomées (Villareal, 1989 ; Foster et Zehr, 2006).

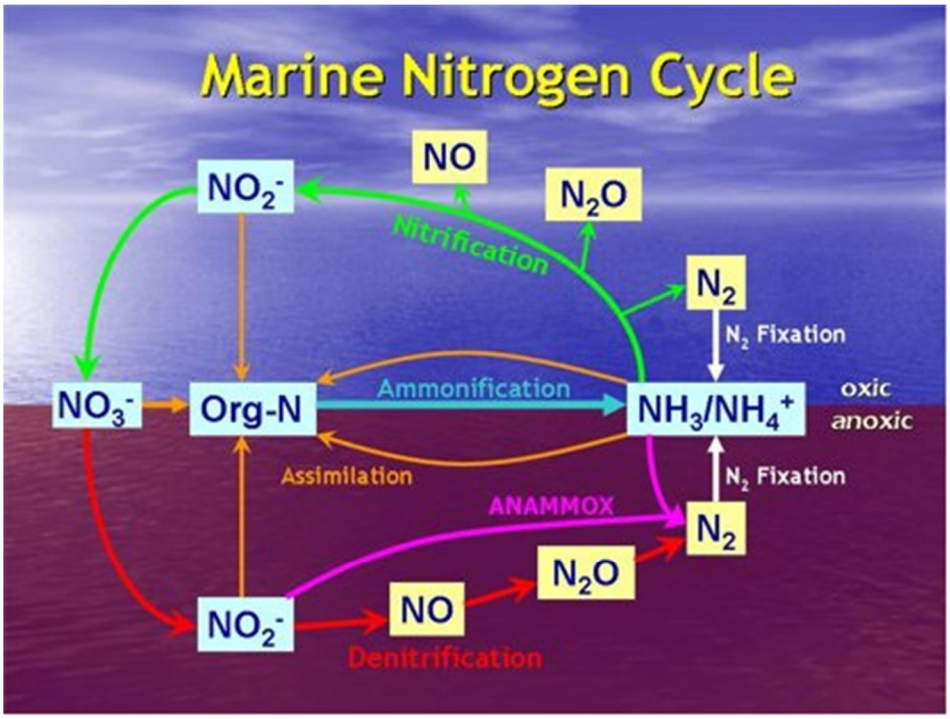


Figure 9 : Cycle biogéochimique de l'azote en milieu marin.
Schéma extrait du site www.mpi-bremen.de/Binaries/Binary5370/nitrogen%20cycle.jpg.

I-2-4- Génome des diatomées.

La maîtrise des techniques de biologie moléculaire et le développement de la bioinformatique ont favorisé l'isolement, le séquençage et l'annotation fonctionnelle des génomes. En raison de leur importance écologique et de leur intérêt technologique, les génomes de plusieurs espèces de diatomées ont été séquencés au cours de la dernière décennie (Tirichine et Bowler, 2011). Ainsi, les séquences des génomes complets de deux espèces modèles de diatomées que sont *Thalassiosira pseudonana* (centrique) et *Phaeodactylum tricornutum* (pennée) ont été publiées et sont maintenant accessibles au public (Armbrust et al., 2004 ; Bowler et al., 2008). Plus récemment, la diatomée pennée psychrophile (adaptée au froid) *Fragilariopsis cylindrus* a également été séquencée et son génome est désormais disponible sur le site http://genome.jgi-psf.org/Fracy1/Fracy1.home.html. D'autres espèces de diatomées pennées, *Pseudo-nitzschia multiseries* (diatomée produisant des produits toxiques), *Fistulifera sp*. et *Seminavis variabilis,* sont elles aussi en cours de séquençage ce qui constituera une source d'informations supplémentaires permettant d'approfondir les travaux de génomique comparative.

L'analyse des génomes de *T. pseudonana* et de *P. tricornutum* a révélé une mine d'informations sur les gènes des diatomées ainsi que sur la biologie et l'origine évolutive de ces organismes. Ces deux génomes ont une petite taille de 32,4 Mpb pour *T. pseudonana* et 27,4 Mpb pour *P. tricornutum,* contiennent respectivement 11 776 et 10 402 gènes et possèdent un pourcentage de GC de l'ordre de 48% (Bowler et al., 2008 ; Armbrust et al., 2004 ; Scala et al., 2002). Les gènes de *T. pseudonana* présentent plus d'introns que ceux de *P. tricornutum* avec en moyenne 1,52 introns par gène pour la première et 0,79 pour la seconde, ce qui suggère une récente acquisition d'introns dans le génome des diatomées centriques (Bowler et al., 2008). De plus, *P. tricornutum* partage 57% de ses gènes en commun avec *T. pseudonana* (Bowler et al., 2008).

La comparaison des séquences génomiques de ces deux diatomées avec celles d'autres organismes eucaryotes comme des plantes, des algues vertes et rouges, des métazoaires, des champignons et d'autres chromalvéolates, a également permis de mettre en évidence la présence de caractéristiques particulières qui sont le reflet de leur histoire évolutive. En effet, comme évoqué précédemment, les diatomées seraient issues d'un évènement endosymbiotique secondaire entre une algue rouge et un eucaryote hétérotrophe (Yoon et al., 2004 ; Keeling, 2009 ; Li et al., 2006). Au cours de ce processus, un transfert de gènes se serait produit entre les génomes nucléaires, chloroplastiques et mitochondriaux de l'hôte et de l'endosymbiote (Finazzi et al., 2010 ; Keeling, 2013 ; **figure 10**). Ces analyses génomiques comparatives ont

effectivement révélées la présence de gènes issus des algues rouges dans les génomes chloroplastiques et nucléaires de *T. pseudonana* et de *P. tricornutum* (Bowler et al., 2008). Ces gènes codent essentiellement pour des protéines impliquées dans la photosynthèse comme la chlorophylle c (Cavalier-Smith, 1999 ; Green, 2011).

En outre, la présence de gènes similaires à ceux de la lignée verte a fait naître l'hypothèse d'un événement endosymbiotique indépendant où l'ancêtre des diatomées aurait phagocyté une algue verte (Moustafa et al., 2009 ; Deschamps et Moreira, 2012). Cette dernière aurait été intégrée puis totalement réduite après avoir transférée certains de ses gènes au noyau de la cellule hôte (**figure 10**). Cette endosymbiose secondaire serait antérieure à celle de l'algue rouge.

De même, les génomes de *T. pseudonana* et de *P. tricornutum* présentent de nombreux gènes très proches de ceux des mammifères. En effet, il a été montré que 806 protéines de *T. pseudonana* s'alignent avec le protéome de souris (*Mus. Musculus*) mais pas avec celui des plantes vertes ni celui des algues rouges, ce qui suggère que les gènes codant ces protéines seraient issus de l'hôte secondaire hétérotrophe (Armbrust et al., 2004 ; **figure 10**). Par ailleurs, *T. pseudonana* et *P. tricornutum* présentent également de nombreux gènes communs aux eucaryotes qui possèdent plus de similarités de séquence avec leurs homologues animaux qu'avec ceux des végétaux (Scala et al., 2002 ; Montsant et al., 2005 a).

Enfin, ces analyses génomiques comparatives ont montré que les génomes de *T. pseudonana* et *P. tricornutum* présentent environ 10 % de gènes originaux, dits « diatomée-spécifique », ne possédant pas d'homologue chez les autres organismes eucaryotes auxquels ils ont été comparés (Armbrust et al., 2004 ; Montsant et al., 2005 a ; Bowler et al., 2008). Néanmoins, la majorité de ces gènes serait d'origine procaryotique, issus de transferts horizontaux de gènes (THG) avec des bactéries (Chan et al., 2012 ; Keeling et Palmer, 2008 ; **figure 10**). En effet, environ 7% des gènes répertoriés dans le génome des diatomées seraient issus de THG indépendants avec des procaryotes ce qui est plus élevé que chez les autres eucaryotes et est comparable à ce qui est retrouvé entre bactéries (Armbrust et al., 2004 ; Bowler et al., 2008). Ces THG seraient à l'origine de l'acquisition de nombreuses voies métaboliques particulières, responsables de la grande capacité d'adaptation des diatomées face à des conditions environnementales hostiles (Montsant et al., 2005 a). Ainsi, ils seraient à l'origine, par exemple, de l'acquisition des mécanismes d'utilisation et de stockage du fer dans des conditions de fer limitant (Armbrust, 2009), de la fixation du carbone (photosynthèse en C4 (Reinfelder et al., 2000)), et surtout de la particularité de posséder un cycle complet de l'urée, ce qui n'a encore jamais été observé chez d'autres eucaryotes autotrophes (Armbrust et

al., 2004 ; Allen et al., 2011). Cependant, contrairement aux métazoaires, qui excrètent l'urée dans leur environnement afin d'évacuer l'excès d'azote, les diatomées possèdent une uréase active capable de le métaboliser. Ainsi, cette protéine confère aux diatomées la capacité d'évoluer dans des milieux contenant de l'urée comme seule source d'azote.

Le génome des diatomées est donc une remarquable mosaïque de séquences génomiques provenant d'au moins quatre sources distinctes, illustrée par la coexistence de métabolismes spécifiques de plantes, d'animaux et de bactéries, pouvant certainement expliquer leur réussite écologique (Armbrust et al., 2004 ; Bowler et al., 2010 ; Chan et al., 2012 ; **figure 10**).

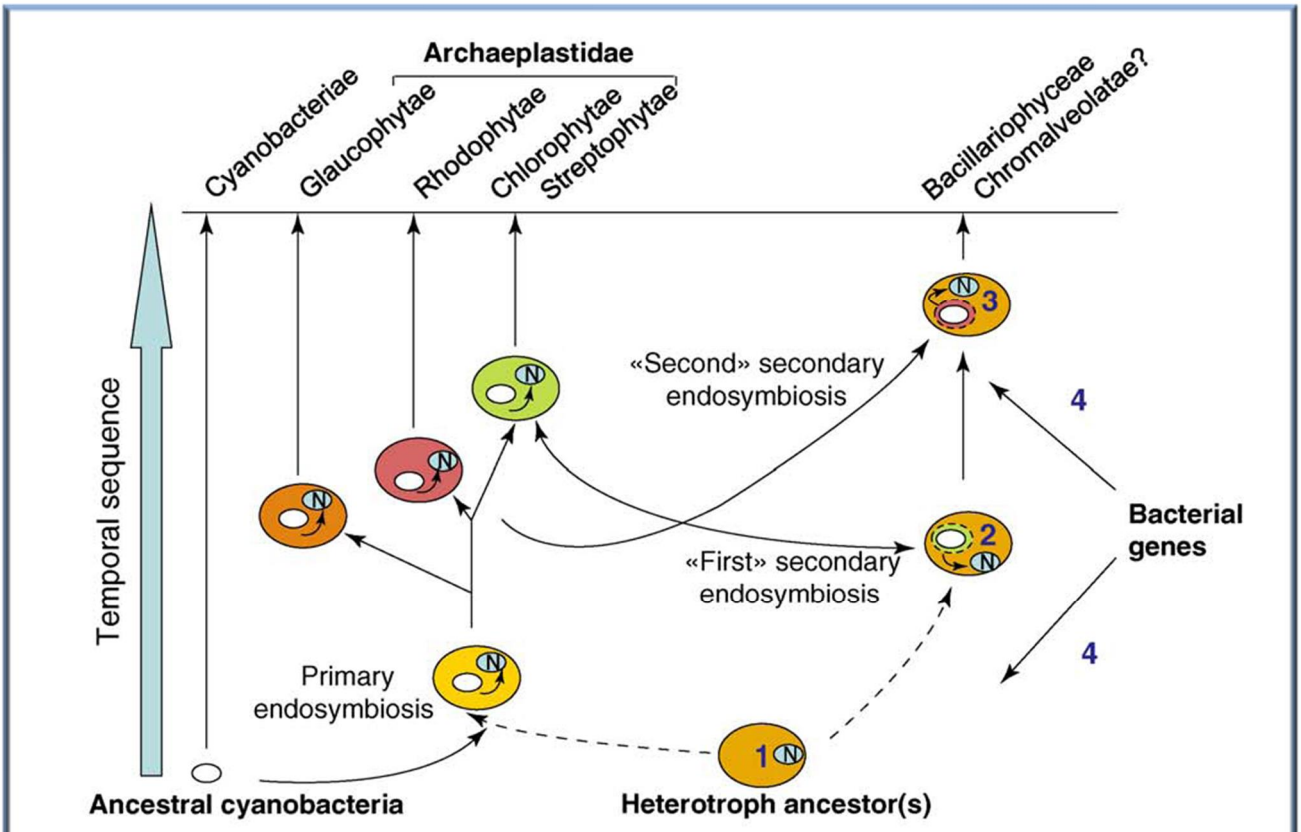


Figure 10 : Schéma présentant l'origine du génome des diatomées par endosymbiose primaire et secondaire et par transfert horizontal de gènes bactériens (Finazzi et al., 2010). Le génome des diatomées (Bacillariophyceae) est une chimère de gènes dérivés d'au moins quatre sources différentes comprenant : (1) l'ancêtre hétérotrophe, (2) une algue verte (« première » endosymbiote secondaire), (3) une algue rouge (« seconde » endosymbiote secondaire) et (4) une gamme de bactéries via de multiples événements de transfert horizontal de gènes. N: noyau. Les petites flèches à l'intérieur des cellules représentent des événements de transfert de gènes des endosymbiotes vers les noyaux des cellules hôtes.

Néanmoins, l'acquisition de nouveaux gènes par transfert horizontal ou endosymbiose ne peut pas expliquer à elle seule la grande capacité d'adaptation des diatomées face aux changements environnementaux. Des mécanismes plus flexibles basés sur des processus épigénétiques pourraient fournir une explication complémentaire. Ces processus correspondent, par exemple, à la méthylation des histones et de l'ADN qui joue un rôle crucial dans la régulation de la structure du génome et de la transcription (Lin et al., 2012). Le méthylome du génome entier de *P. tricornutum* a récemment été étudié et montre que cette modification concerne environ 6% du génome et plus de 350 gènes (Veluchamy et al., 2013).

Par ailleurs, il a été mis en évidence que le génome des diatomées possède des éléments mobiles appelés éléments transposables (ET) (Armbrust et al., 2004 ; Bowler et al., 2008 ; Maumus et al., 2009). *P. tricornutum* possède 6,4 % d'ET dans son génome, soit trois fois plus que *T. pseudonana* qui n'en contient que 2 % (Armbrust et al., 2004 ; Maumus et al., 2009). Certains de ces ET sont exprimés et régulés de façon différente selon les conditions de culture, ce qui suggère que l'expression des ET pourrait avoir un impact sur la capacité d'adaptation des diatomées. De plus, des séquences correspondant à des microARN ont également été identifiées dans le génome de *P. tricornutum* (Huang et al., 2011) et de *T. pseudonana* (Norden-Krichmar et al., 2011). Leurs cibles potentielles sont impliquées dans divers processus, tels que la transduction du signal, la régulation de la transcription, la phosphorylation des acides aminés et le processus de biosynthèse des acides gras (Huang et al., 2011). L'ensemble de ces mécanismes de régulation fournissent une réponse supplémentaire quant à la réussite écologique de ces microalgues.

I-2-5- Utilisations des diatomées.

Pendant de nombreuses années, les diatomées ont presque exclusivement été étudiées dans le cadre de recherches fondamentales universitaires avec peu de considération pour leurs potentielles applications industrielles. Néanmoins, de par leur biodiversité, leurs propriétés biochimiques, leur répartition géographique, leur structure et leurs aptitudes biologiques particulières, les diatomées ont reçu ces dernières années une plus grande attention en recherche appliquée. En particulier, les propriétés physico-chimiques et la construction géométrique de leur frustule ainsi que leur capacité à produire naturellement des molécules à haute valeur ajoutée, leur confère un attrait considérable dans un grand nombre de secteurs d'activité. Par ailleurs, les diatomées bénéficient d'un temps de génération très court et leur culture en grands volumes est bien maîtrisée, rendant leur utilisation tout à fait adaptée à l'échelle industrielle. Les diatomées suscitent donc aujourd'hui un intérêt aussi bien économique que purement scientifique et se révèlent très prometteuses dans des domaines d'application très variés (**figure 11** ; Vizinet et De reviers, 1995 ; Lopez et al., 2005 ; Bozarth et al., 2009).

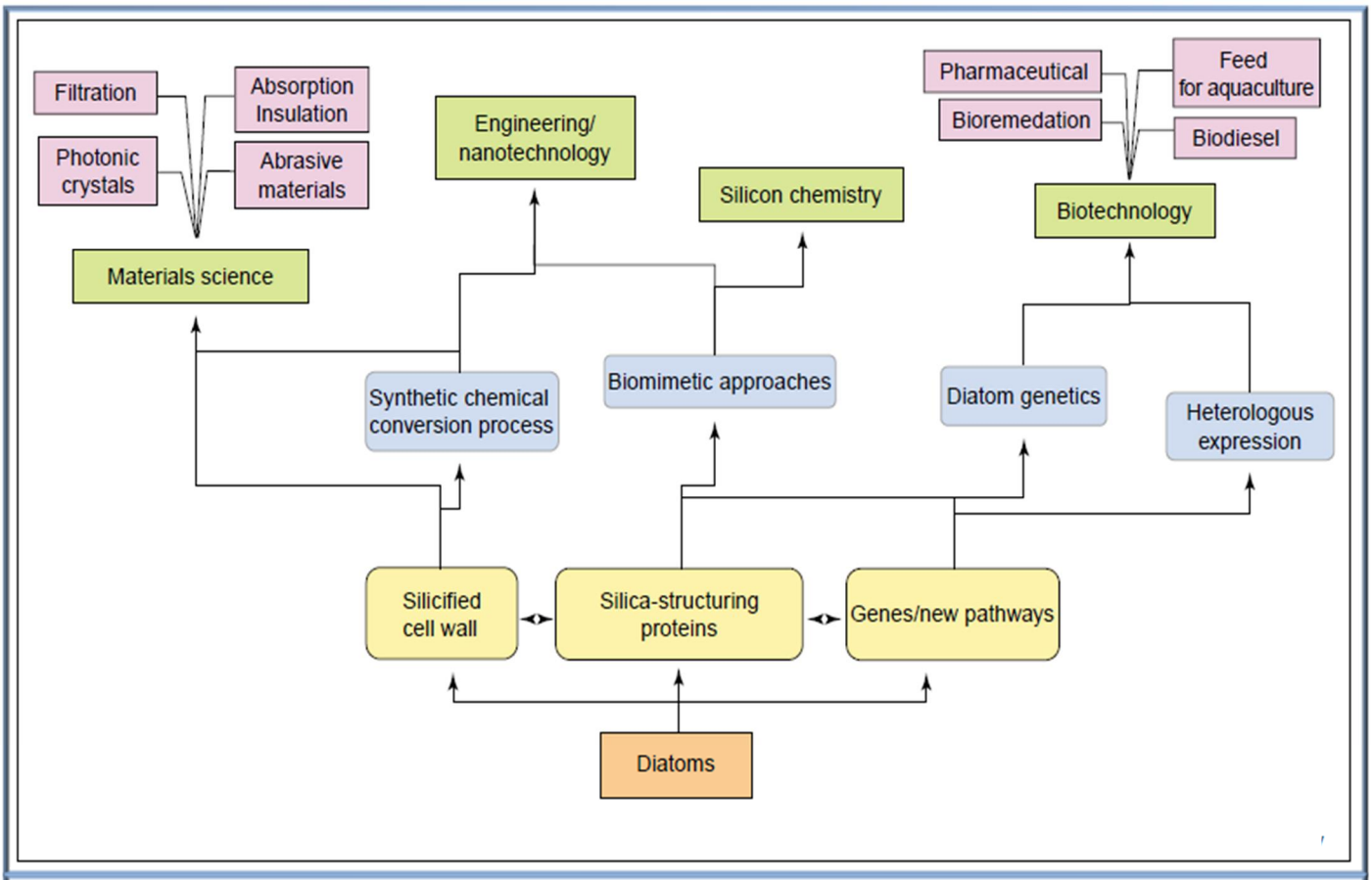


Figure 11 : Les différents domaines d'application des diatomées (Lopez et al., 2005).

I-2-5-1- Diatomite

La diatomite ou « terre de diatomée » est une roche sédimentaire siliceuse, poreuse et friable constituée entièrement ou majoritairement de squelettes (frustules) de diatomées fossilisées. Cette roche est exploitée industriellement pour ses diverses propriétés, notamment pour son caractère abrasif (Vizinet et De reviers, 1995 ; Lopez et al., 2005). Cet aspect est employé dans différents secteurs comme la cosmétologie, où la terre de diatomée entre, par exemple, dans la composition de dentifrices ou de masques exfoliants. Elle possède également un pouvoir absorbant utilisé notamment pour la fabrication de litière pour animaux domestiques.

Ces deux propriétés, absorbante et abrasive, ont conduit certaines entreprises à proposer la diatomite comme insecticide naturel (Korunic , 1997 ; Kljajic et al., 2010). En effet, cette roche, sous sa forme broyée, agit sur les insectes comme de petits cristaux de verres concassés. Après avoir adhéré au corps de l'insecte, la diatomite perfore les différents tissus tels que les couches de cire protectrice, favorisant ainsi une perte de liquide entrainant la déshydratation puis la mort de l'insecte. Cet effet est renforcé par le caractère absorbant de cette roche. La terre de diatomée a cependant le désavantage de ne pas être sélective, car elle génère la mort de toutes espèces d'insectes, qu'elles soient bénéfiques ou nuisibles.

La diatomite est également utilisée comme système de filtration appelé filtre à diatomées. En effet, de par la microporosité de cette roche, la filtration à diatomées élimine très efficacement les impuretés et les bactéries (Lange et al., 1986). La propriété filtrante de la diatomite, due à ses interstices de l'ordre de 0,1 micron, permet de retenir la plupart des organismes uni- et pluricellulaires (Schuler et al., 1991). Ainsi, la filtration à diatomées est un procédé approprié et efficace pour traiter les eaux usées. Ce système est notamment utilisé dans la fabrication de filtres de piscine.

De même, la faible conductivité thermique de cette roche lui confère des propriétés isolantes utilisées dans certains matériaux de construction (Lopez et al., 2005). Enfin, la diatomite est utilisée par l'industrie dans la fabrication de nombreux produits du quotidien comme la peinture, le papier, le béton, le vernis, les pneus, les allumettes... pour ses propriétés filtrantes, d'agent abrasif, d'isolant ou encore comme stabilisant.

I-2-5-2- Alimentation.

Les diatomées peuvent constituer un réel apport nutritif. Elles sont couramment utilisées dans l'alimentation animale, notamment en aquaculture. En effet, les diatomées représentent une composante essentielle de l'alimentation des organismes marins d'élevage tels que les poissons et les mollusques. Par exemple, *Haslea ostrearia,* qui est responsable du verdissement des huitres, est une diatomée couramment utilisée dans le nourrissage en ostréiculture.

Les diatomées sont également utilisées dans l'alimentation humaine. En effet, de par leur teneur en vitamines de haute qualité, en acides gras insaturés essentiels et en acides aminés, les diatomées sont naturellement devenues des sources intéressantes de compléments alimentaires comparables aux végétaux (Becker, 2007). Elles sont notamment capables de synthétiser des acides gras poly-insaturés, précurseurs d'oméga 3, tels que l'acide eicosapentaénoïque (EPA ; Lebeau et Robert, 2003 ; Meiser et al., 2004), l'acide arachidonique (Lebeau et Robert, 2003) et l'acide docosahexaénoïque (Kroth, 2007). Certaines espèces de diatomées peuvent produire ces biomolécules en quantité aussi élevée que chez les poissons, l'autre source naturelle riche en oméga 3. En particulier, l'EPA est un acide gras polyinsaturé très abondant chez les diatomées qui constitue jusqu'à 30% de tous les acides gras. Chez *P. tricornutum*, il représente entre 3,9 et 5 % du poids sec. La purification de cet acide gras chez les diatomées est peu coûteuse contrairement à l'huile de foie de poisson, et ne présente pas l'inconvénient de posséder un goût désagréable (Lebeau et Robert, 2003 ; Bozarth et al., 2009).

I-2-5-3- Santé, médecine.

De nombreuses utilisations possibles des diatomées ont été montrées à ce jour dans le domaine de la médecine. En effet, les diatomées représentent une mine de biomolécules pouvant avoir des effets très variés sur la santé (Lincoln et al., 1990). Tout d'abord, comme expliqué dans la partie précédente, les diatomées sont des organismes riches en acides gras insaturés. Ces composés permettent de réduire les maladies cardiovasculaires (Lebeau et Robert, 2003) et sont des précurseurs d'hormones tissulaires importantes. Ils sont présents dans 20% de la matière grise du cerveau (Kroth, 2007) et sont généralement considérés comme anticancéreux. Un grand nombre d'autres molécules produites par diverses diatomées suscite un intérêt dans le domaine de la santé. En particulier, certaines diatomées synthétisent naturellement des composés biologiquement actifs pouvant avoir un effet antibiotique

(exemple chez *Skeletonema Costatum et Cheatoceros Lauderi* (Pesando et al., 1979 ; Lincoln et al., 1990 ; Lebeau et Robert, 2003), d'inhibiteur d'enzyme, de dépresseur du système nerveux central et de relaxant musculaire potentiel (exemple chez *P. tricornutum* (Laguna et al., 1990 ; Villar et al., 1992 ; Villar et al., 1994), d'antifongique (exemple de polysaccharidiques chez *Chaetoceros lauderi* (Gueho et al., 1977 ; Pesando et al., 1979), d'anti-inflammatoire (exemple de polysaccharides chez *P. tricornutum* (Guzmán et al., 2003)) et d'analgésique (Guzmán et al., 2001). Des composés organiques provenant de *S. costatum* (Bergé et al., 1997) et *Haslea ostrearia* (Rowland et al., 2001) ont également montré une activité antitumorale *in vitro* contre le cancer du poumon chez l'homme et des effets anti-VIH. De même, la fucoxanthine qui est un caroténoïde hautement présent chez certaines diatomées comme *P. tricornutum,* possède des propriétés biologiques remarquables avec des applications potentielles importantes et prometteuses en matière de santé humaine (Kim et al., 2012). Cette molécule possède des propriétés anti-inflammatoires, antioxydantes, anticancéreuses, anti-obésité, anti-diabétiques, anti-angiogéniques et antipaludiques, et des effets protecteurs sur le foie, les poumons, les vaisseaux sanguins du cerveau, les os, la peau et les yeux (Moreau et al., 2006 ; Peng et al., 2011).

Le frustule des diatomées présente également un intérêt en médecine comme vecteur de médicaments. Il possède une répartition uniforme des pores sur sa structure et une neutralité chimique qui le rende biocompatible. Ainsi, le frustule est rempli avec un médicament puis injecté dans le circuit sanguin ou administré par voie orale. Très récemment, des essais ont ainsi été réalisés avec deux médicaments, que sont la prednisone et la mésalamine, et ont montré une libération prolongée et une amélioration de la perméabilité de ces composés après administration orale (Zhang et al., 2013).

En médecine légale, l'étude des diatomées retrouvées dans les tissus d'une victime représente un complément d'information important de l'enquête essentiellement dans les affaires d'homicide par noyade (Pollanen, 1998).

I-2-5-4- Chimie fine

Hormis les nombreuses applications liées à la diatomite, présentée précédemment, les diatomées sont également la source d'autres molécules très convoitées dans l'industrie telles que les antioxydants et les pigments (Pasquet et al., 2011). En particulier, les caroténoïdes, comme le 3-carotène qui est un pigment de couleur orange, sont très utilisés comme additif, antioxydant et agent colorant dans diverses industries (agro-alimentaire, pharmaceutique, cosmétologique, textile…).

I-2-5-5- Energie

Dans le domaine de l'énergie, les diatomées présentent également un intérêt puisqu'elles contiennent les trois éléments principaux permettant la production de différents types de carburant : des glucides pour la production d'éthanol par fermentation, des protéines pour la production de méthane par gazéification anaérobie et des lipides pour la production de biodiesel (McGinnis et al., 1997 ; Cadoret et Bernard, 2008 ; Song et al., 2013). La fraction lipidique de nombreuses algues est très élevée pouvant représenter jusqu'à plus de 70% de la masse sèche (Chisti, 2007). De plus, les caractéristiques des huiles naturelles provenant des microalgues sont similaires à celles des poissons et des huiles végétales, et sont donc fortement envisagées comme substituts potentiels aux huiles fossiles. Pour cela, de nombreuses études sont consacrées à la recherche d'espèces de diatomées présentant le plus fort taux d'acides gras et le meilleur rendement.

I-2-5-6- Environnement

Outre leur importance environnemental dans un grand nombre de cycles biogéochimiques, les diatomées possèdent de nombreuses qualités pour être utilisées en tant que bioindicateurs pour le contrôle de la qualité des eaux (Scherer, 2002 ; Herlory et al., 2013). En effet, les diatomées colonisent tous les milieux aquatiques et regroupent un grand nombre d'espèces avec des préférences écologiques variées vis- à -vis des facteurs environnementaux ce qui constitue une grande source d'informations sur les caractéristiques de l'habitat dans lequel elles se développent. De plus, leur cycle de développement très court leur confère un temps de réponse rapide face aux variations du milieu.

Par ailleurs, les diatomées suscitent un intérêt pour une utilisation industrielle dans des programmes de phytoremédiation (Lopez et al., 2005). Leurs propriétés métaboliques leur permettent de servir de filtres naturels à un grand nombre de particules toxiques. Ainsi, les diatomées sont envisagées comme solution pour le traitement des eaux contaminées, notamment dans l'épuration du $C0_2$, du phosphate, de l'azote et des métaux lourds (Craggs et al., 1997 . Lebeau et Robert, 2003). La résistance à une forte concentration en métaux lourds et la capacité d'en absorber des quantités non négligeables ont été montrées chez plusieurs diatomées telles que *Cylindrotheca fusiformis*, *C. cryptica*, *P. tricornutum*, *S. costatum* et *T. pseudonana* (Bræk et al., 1980 ; Morelli et Pratesi, 1997 ; Pistocchi et al., 2000). Certaines diatomées comme *Chaetoceros calcitrans* sont également capables de dégrader des hydrocarbures.

Enfin, les diatomées sont très utilisées pour les reconstitutions paléo-environnementales et paléo-océanographiques. Certaines espèces sont des fossiles stratigraphiques intéressants. En effet, leur abondance et leur large répartition géographique permet d'effectuer des corrélations stratigraphiques entre des régions éloignées.

I-2-5-7- Nanotechnologie

Les diatomées sont également utilisées comme modèle d'étude dans les secteurs de recherche et développement en nanotechnologie et sciences de l'ingénieur (Parkinson et Gordon, 1999 ; Jamali et al., 2012). En effet, l'architecture géométrique et nanométrique du frustule des diatomées intéresse fortement les chercheurs. Il est possible, par exemple, de se servir de cette structure naturelle comme matrice pour la fabrication d'une réplique dans le matériau de son choix (Bozarth et al., 2009). Pour cela, il s'agit d'enrober le frustule ou de remplacer la silice par d'autres composés. Des expériences ont déjà été menées consistant à recouvrir l'enveloppe des diatomées avec des matériaux tels que le dioxyde de titane, l'oxyde de magnésium, le silicium ou le carbone (Drum et Gordon, 2003). Des procédés similaires ont été réalisés avec des métaux comme le cuivre, l'argent, l'or, le platine et le nickel, après quoi la silice sous-jacente a été dissoute permettant d'obtenir une réplique métallique 3D d'un frustule. Ces nanocomposés possèdent de nombreuses applications possibles comme la fabrication de filtres optiques ou de nanopuces.

I-2-5-8- Usine cellulaire

Une compétition internationale visant à élaborer le meilleur système d'expression pour la production de molécules à haute valeur ajoutée, comme des protéines à intérêt thérapeutique, a été initiée il y a plusieurs décennies. Les recherches ont tout d'abord été développées chez les bactéries, les levures, les plantes terrestres et les animaux. Cependant, le développement d'outils de biologie moléculaire et l'accessibilité au génome, ainsi que leur facilité à être cultivé et leur vitesse de croissance élevée, ont fait des diatomées de sérieux candidats pour la production de molécules biologiques à fort intérêt économique comme les anticorps, les vaccins, les hormones et les enzymes (Cadoret et al., 2008 ; Bozarth et al., 2009 ; Mimouni et al., 2012).

I-3- La diatomée modèle *Phaeodactylum tricornutum*.

P. tricornutum est une diatomée pennée identifié par K. Bohlin en 1897. Pendant de nombreuses décennies, elle a servi essentiellement de modèle dans des études portant sur la physiologie des diatomées. Toutefois, avec le développement des outils de biologie moléculaire, elle a également été choisie comme espèce représentante des diatomées lors d'analyses génomiques et post-génomiques. En effet, la petite taille de son génome, sa capacité à être transformée génétiquement ainsi que sa facilité à être cultivée ont fait de *P. tricornutum* un candidat évident pour les études de génomique fonctionnelle réalisées chez les diatomées.

Contrairement à *T. pseudonana*, *P. tricornutum* ne montre pas une importance écologique majeure. Néanmoins, cette diatomée marine semble être cosmopolite car elle a été retrouvée à différents endroits de la planète, aussi bien dans des eaux froides que tropicales. Elle est principalement localisée au niveau des zones côtières, dans des environnements instables tels que les estuaires (De Martino et al., 2007). Dans ces milieux, la salinité et la température peuvent changer rapidement sous l'effet des marées, de l'irradiation solaire et des activités humaines. La capacité d'adaptation à une telle fluctuation des conditions environnementales pourrait être liée au caractère polymorphique unique des cellules de *P. tricornutum*. En effet, cette diatomée présente la particularité d'exister à l'état naturel sous trois morphotypes différents : fusiforme, ovale et triradié (**figure 11** ; Lewin et al., 1958 ; Borowitzka et Volcani, 1978 ; De Martino et al., 2007). Ce dernier est responsable de l'origine étymologique de son nom tricornutum avec « tri » pour trois et « cornutum » pour cornes. La plasticité morphologique unique de cette diatomée est possible car la silicification de sa paroi est facultative (Borowitzka and Volcani, 1978 ; De Martino et al., 2007). En effet, seul le morphotype ovale possède un réel frustule de silice. Le passage d'un morphotype à un autre peut être induit par modification des conditions environnementales (Lewin et al., 1958 ; Borowitzka et Volcani, 1978 ; De Martino et al., 2007). Les paramètres de culture permettant ce changement morphologique *in vitro* commencent à être mieux maitrisés (**figure 11 ;** Gutenbrunner et al., 1994 ; De Martino et al., 2007). Par exemple, il a été montré qu'une variation de la température et de la salinité était impliquée dans ce phénomène (De Martino et al., 2007 ; De Martino et al., 2011). Néanmoins, les processus biologiques qui régissent ce changement morphotypique ne sont pas encore bien connus.

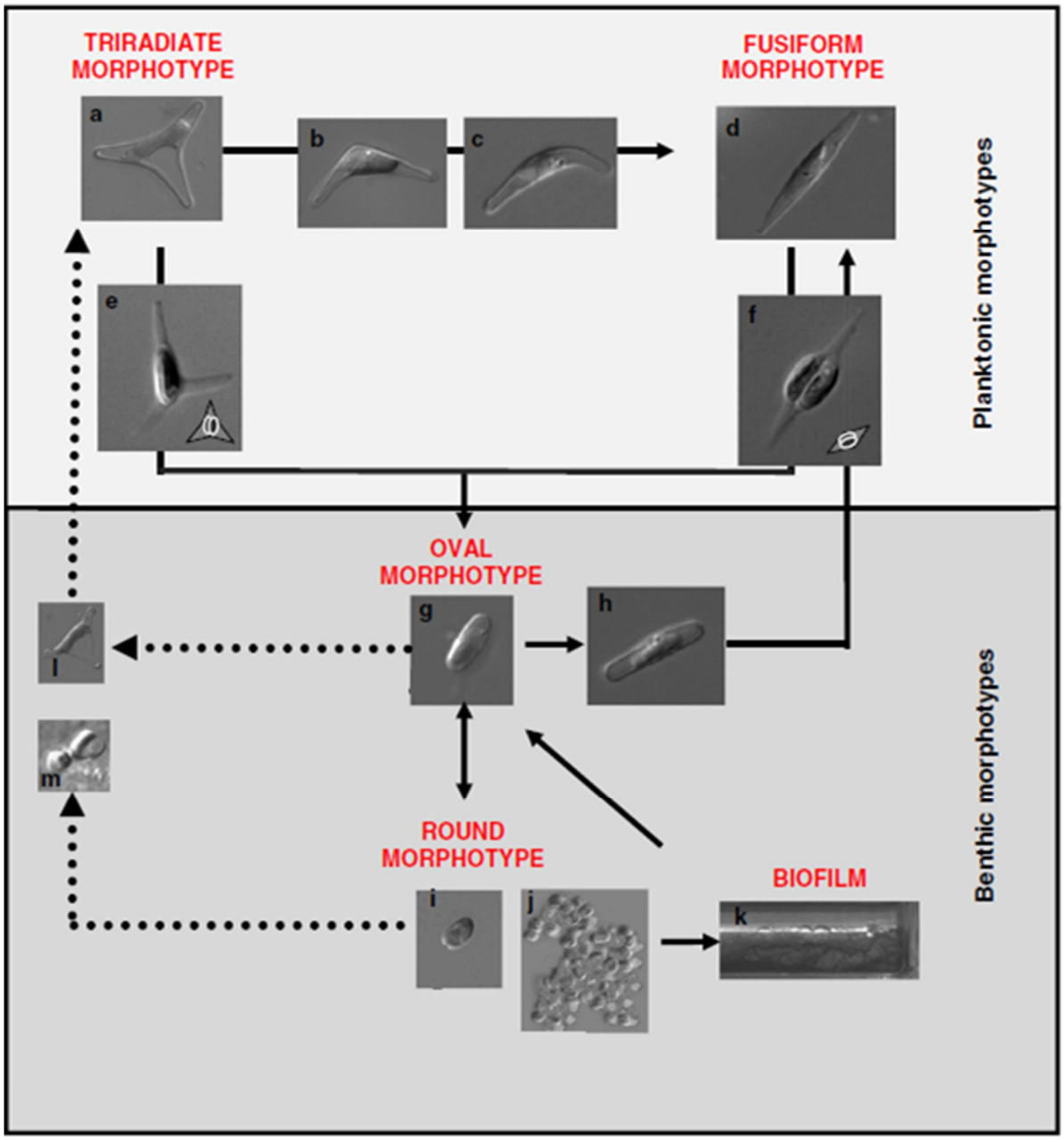


Figure 12 : Schéma présentant les changements morphotypiques de *P. tricornutum* en réponse à des signaux environnementaux (De Martino et al., 2011).

P. tricornutum présente trois morphotypes principaux : ovale (g), fusiforme (d) et triradié (a). Ces deux derniers sont stables en phase exponentielle dans des conditions standards de culture c'est-à-dire dans l'eau de mer, sous agitation et à une température avoisinant les 19°C. Le morphotype triradié change en fusiforme en passant par des formes intermédiaires (boomerang) (b, c). Les morphotypes fusiforme et triradié se convertissent en ovale par division cellulaire (e, f)) en réponse à une hyposalinité et des températures froides (15◦C). Le processus inverse s'effectue par un allongement de la cellule (h). Dans des conditions de stress prolongées, les cellules ovales s'arrondissent (i) et forment des agrégats (j) qui peuvent générer des biofilms (k) Toutes les conversions sont réversibles à l'exception de fusiforme vers triradié. Des petites cellules triradiées ont été observées occasionnellement (l) pouvant correspondre à un état intermédiaire entre ronde et triradiée (l, m).

Différents écotypes de *P. tricornutum* ont été échantillonnés à travers le monde. Parmi les dix écotypes les plus étudiés, les souches Pt1, Pt3 et Pt8 sont respectivement et majoritairement présentes sous les morphotypes fusiforme, ovale et triradié en conditions standards de culture (température : 19°C, salinité : 100% eau de mer). Elles se sont donc imposées comme souches références pour l'étude de leur morphotype majoritaire respectif. (De Martino et al., 2007). En outre, la souche Pt1 présente un intérêt tout particulier car elle est à l'origine du clone Pt1.8.6 qui a servi au séquençage du génome de *P. tricornutum* (Bowler et al., 2008*).*

Afin d'approfondir les connaissances sur la biologie des diatomées, plusieurs outils de biologie moléculaire ont été développés chez cette microalgue (Montsant et al., 2005 b). Tout d'abord, *P. tricornutum* est transformable génétiquement par biolistique (Apt et al., 1996 ; Falciatore et al., 1999 ; Siaut et al., 2007) et plus récemment par électroporation (Niu et al., 2012 ; Miyahara et al., 2013). De même, des vecteurs de sur-expression et répression des gènes par RNAi (De Riso et al., 2009) ainsi que plusieurs antibiotiques de sélection (Apt et al., 1996 ; Falciatore et al., 1999) ont été développés chez cette diatomée. Une banque de small RNAs (240 000) (Huang et al., 2011) et des outils pour des analyses transcriptomiques (whole transcriptome (Sapriel et al., 2009) et microarrays (Hook and Osborn, 2012)) ont aussi été élaborés.

De plus, une base de données numérique de l'expression des gènes de *P. tricornutum* est également disponible (http://www.diatomics.biologie.ens.fr/EST3/index.php). Cette banque regroupe des étiquettes de séquences transcrites (Expressed Sequence Tag : EST) qui sont de courts fragments d'ADNc sélectionnés aléatoirement puis séquencés partiellement, permettant la caractérisation des gènes sans ambiguïté. Les EST représentent les gènes exprimés dans une cellule à un temps donné. *P. tricornutum* possède une banque de plus de 130 000 EST obtenues à partir de cellules cultivées dans 16 conditions différentes (Maheswari et al., 2009 ; Maheswari et al., 2010).

L'accessibilité récente au génome et le développement d'outils de biologie moléculaire facilitent désormais l'étude des différentes voies métaboliques de cette diatomée (Montsant *et al.*, 2005b), telle que la N-glycosylation.

II- N-glycosylation.

La N-glycosylation est une modification co- et post-traductionnelle majeure des protéines conduisant à l'ajout et à la maturation de structures glucidiques sur le squelette protéique. Dans les organismes eucaryotes, ce processus constitue une étape essentielle pour l'acquisition de la conformation fonctionnelle des glycoprotéines destinées à être sécrétées ou intégrées dans les membranes. La N-glycosylation se déroule plus spécifiquement dans deux compartiments subcellulaires, appartenant au système endomembranaire de sécrétion, que sont le réticulum endoplasmique (RE) et l'appareil de Golgi. Les protéines pénètrent tout d'abord à l'intérieur du RE où elles peuvent subir les premières étapes de la N-glycosylation, puis transitent dans l'appareil de Golgi dans lequel la maturation de leurs N-glycannes peut éventuellement se poursuivre. Cette voie métabolique est caractérisée par l'action de deux enzymes clés. Il s'agit en premier lieu de l'oligosaccharyltransférase (OST) localisée dans le RE et dont le rôle est de catalyser le transfert de l'oligosaccharide précurseur sur la chaîne polypeptidique en cours de synthèse, élément indispensable au repliement des glycoprotéines. La seconde enzyme est la N-acétylglucosaminyltransférase I (GnT I) qui intervient dans l'appareil de Golgi où elle initie la maturation des N-glycannes oligomannosidiques en N-glycannes complexes. Ces derniers jouent un rôle physiologique essentiel en permettant aux protéines d'acquérir un grand nombre de fonctions supplémentaires. L'importance biologique de ces N-glycannes est de mieux en mieux renseignée notamment aux travers de nombreuses études portant sur des maladies congénitales ou des mutants présentant une défaillance dans leur voie de N-glycosylation.

II-1- Réticulum endoplasmique, siège de la N-glycosylation des protéines.

La N-glycosylation des protéines dans le RE est un processus très conservé chez les eucaryotes qui se décompose en trois phases (Breitling et Aebi, 2013). Elle débute par la synthèse d'un précurseur oligosaccharidique sur une ancre lipidique, le dolichol pyrophosphate (Aebi, 2013). Ce précurseur est ensuite transféré sur un polypeptide en cours de synthèse par l'OST, étape essentielle au bon repliement des protéines leur permettant ainsi d'acquérir leur conformation fonctionnelle (Helenius, 1994 ; Helenius et Aebi, 2004). Afin de vérifier le repliement de la glycoprotéine, cette dernière entre dans un cycle de contrôle qualité à l'issue duquel elle sera soit dégradée soit acheminée le long du système sécrétoire pour effectuer la suite de sa maturation (Aebi et al., 2010). Les N-glycannes obtenus après les différentes étapes opérées dans le RE sont des N-glycannes riches en résidus mannose appelés N-glycannes oligomannosidiques.

II-1-1- Synthèse du précurseur oligosaccharidique.

L'oligosaccharide précurseur est assemblé par l'ajout successif de trois monosaccharides différents que sont la N-acétylglucosamine (GlcNAc), le mannose (Man) et le glucose (Glc), sur une ancre lipidique appelée dolichol phosphate (Dol-P). Ces réactions d'addition sont catalysées par un ensemble de glycosyltransférases spécifiques nommées ALG (Asparagine Linked Glycosylation) et localisées au niveau de la membrane du RE (Weerapana et Imperiali, 2006 ; Breitling et Aebi, 2013). Avant tout ajout glucidique, l'ancre lipidique est préalablement phosphorylée par l'ajout d'un résidu phosphate sur le dolichol par la protéine SEC59 à partir du nucléotide CTP (**figure 13**).

II-1-1-1- Etapes cytosoliques.

La synthèse du précurseur oligosaccharidique débute sur la face cytosolique de la membrane du RE. Les glycosyltransférases intervenant à ce niveau utilisent des nucléotides sucres activés comme substrat (Aebi et al., 2010). En effet, l'ajout cytosolique de N-acétylglucosamine et de mannose sur l'ancre lipidique est effectué respectivement à partir d'UDP-GlcNAc et de GDP-Man.

La voie de biosynthèse du précurseur oligosaccharidique est initiée par la N-acétylglucosamine-phosphate-transférase (ALG7) qui catalyse l'ajout d'un résidu GlcNAc-phosphate sur un Dol-P, formant le dolichol pyrophosphate-GlcNAc (Dol-PP-GlcNAc) (**figure 13** ; Aebi, 2013). Un complexe enzymatique, formé par les protéines ALG13 et ALG14, transfère ensuite le second résidu GlcNAc (**figure 13** ; Bickel et al., 2005 ; Gao et al., 2008). Ces trois enzymes, utilisant l'UDP-GlcNAc comme substrat, constituent un complexe polymérique dont la formation est organisée par la sous-unité ALG14 (**figure 13** ; Noffz et al., 2009 ; Lu et al., 2012).

Interviennent alors trois enzymes responsables de l'addition de 5 résidus mannoses sur l'oligosaccharide en cours de synthèse. Ces trois mannosyltransférases utilisant du GDP-Man comme substrat, forment également un complexe protéique (**figure 13** ; Gao et al., 2004). L'ajout du premier résidu mannose sur le $GlcNAc_2$-PP-Dol est catalysé par la β-1,4-mannosyltransférase ALG1 (**figure 13** ; Couto et al., 1984). Deux mannoses ramifiés sont ensuite ajoutés en α-1,3 puis en α-1,6 sur le premier mannose par une enzyme unique, l'ALG2 (**figure 13** ; O'Reilly et al., 2006 ; Kämpf et al., 2009). Enfin, la mannosyltransférase ALG11 transfère successivement deux autres résidus mannose ce qui entraîne la formation du composé $Man_5GlcNAc_2$-PP-Dol, produit final de la biosynthèse du précurseur oligosaccharidique au niveau cytosolique (**figure 13** ; Cipollo et al., 2001).

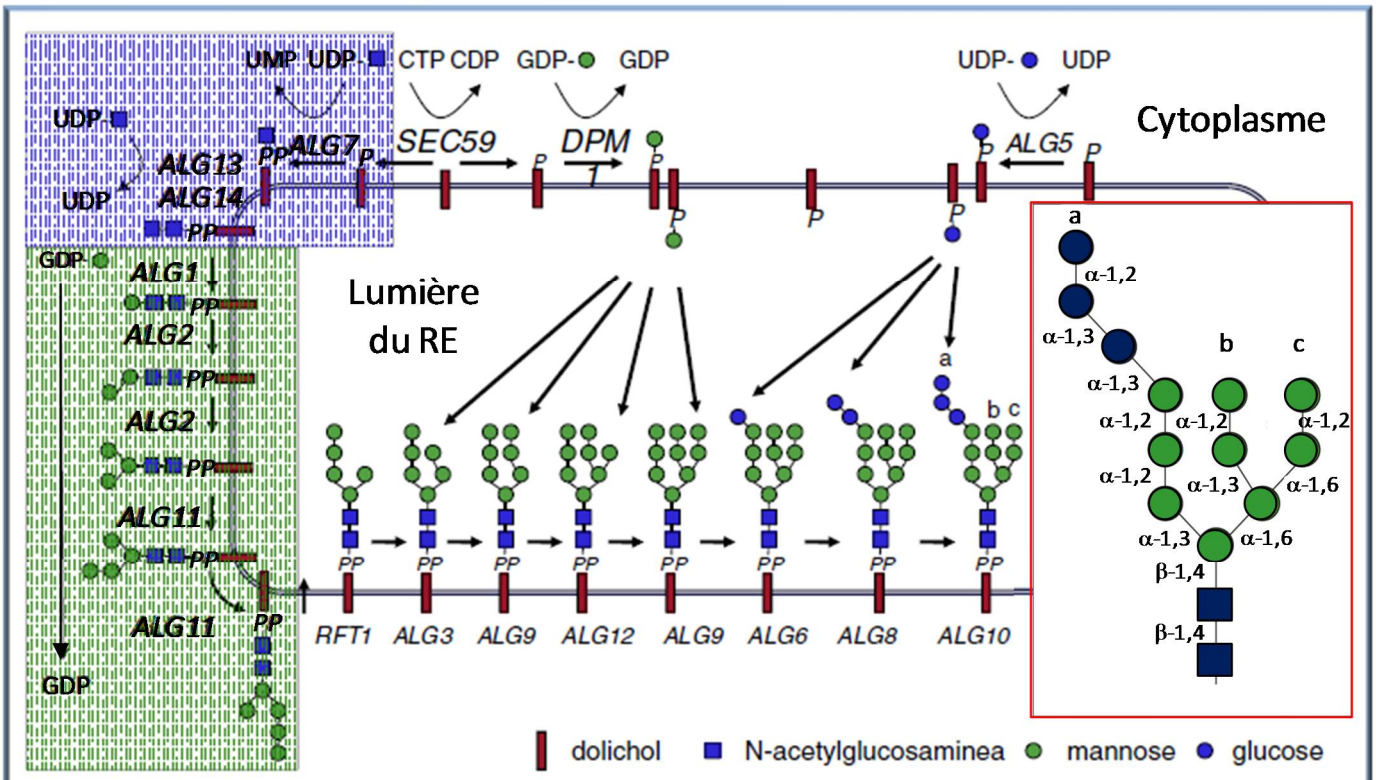


Figure 13 : Voie de biosynthèse du précurseur oligosaccharidique $Glc_3Man_9GlcNAc_2$ d'après (Aebi, 2013). Ce processus est catalysé par différentes glycosyltransférases spécifiques nommées ALG (Asparagine Linked Glycosylation) indiquées sur le schéma. Les glycosyltransférases actives sur la face cytosolique de la membrane du réticulum endoplasmique (RE) et qui utilisent l'UDP-GlcNAc et le GDP-Man forment respectivement deux complexes enzymatiques représentés par un carré bleu et vert. Les lettres a, b et c désignent les trois antennes du précurseur oligosaccharidique final $Glc_3Man_9GlcNAc_2$. Un agrandissement indiquant les différentes liaisons de ce composé est représenté dans le cadre rouge.

II-1-1-2- Etapes luminales.

Les étapes suivantes de la biosynthèse du précurseur oligosaccharidique se produisent sur la face luminale de la membrane du RE. Pour cela, l'oligosaccharide en cours de synthèse est transloqué à travers la membrane par la flippase Rft1 (**figure 13** ; Snider et Rogers, 1984 ; Helenius et al., 2002 a et b). Les glycosyltransférases, agissant au niveau de la face luminale du RE, ont pour substrat le Dol-P-Man et le Dol-P-Glc. Ces deux composés sont synthétisés sur la face cytosolique de la membrane du RE. Le Dol-P-Man est 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 (**figure 13** ; Orlean et al., 1988). De même, la synthèse du Dol-P-Glc consiste en l'ajout d'un résidu glucose provenant d'un UDP-Glc sur un Dol-P par la Dol-P-Glc synthase ALG5 (**figure 13** ; Heesen et al., 1994). Ces deux composés sont ensuite transloqués à travers la membrane afin d'être utilisés comme substrat dans la lumière du RE (**figure 13** ; Rush et Waechter, 1995 ; Rush et al., 1998 ; Sanyal et Menon, 2010).

La biosynthèse luminale du précurseur oligosaccharidique consiste en l'ajout de quatre résidus mannose supplémentaires et de trois résidus glucose. Les mannosyl- et glucosyltransférases responsables de ces réactions sont des enzymes hautement conservées caractérisées par la présence de multiples segments transmembranaires et appartenant à la classe GT- C des glycosyltransférases (Lairson et al., 2008).

La première étape est catalysée par l'α-1,3-mannosyltransférase ALG3 qui ajoute un résidu mannose lié en α-1,3 sur le précurseur oligosaccharidique $Man_5GlcNAc_2$-PP-Dol (**figure 13** ; Aebi et al., 1996). Cette étape est suivie de l'addition d'un résidu mannose lié en α-1,2 par la protéine ALG9, formant ainsi l'antenne « b » de l'oligosaccharide (**figure 13** ; Frank et Aebi, 2005). La synthèse de l'antenne « c » débute, quant à elle, par l'action de l'α-1,6-mannosyltransférase ALG12. L'ajout d'un résidu mannose lié en α-1,2 est ensuite catalysé par la protéine ALG9, aboutissant à la formation de l'oligosaccharide $Man_9GlcNAc_2$-PP-Dol (**figure 13** ; Burda et al., 1999 ; Frank et Aebi, 2005). Cependant, la synthèse de ce composé n'est pas encore achevée lorsque la protéine ALG6 initie la glucosylation de l'antenne « a » du précurseur (**figure 13** ; Reiss et al., 1996). Un second résidu glucose est ensuite ajouté sur le premier en α-1,3 par la glucosyltransférase ALG8 (**figure 13** ; Stagljar et al., 1994). Enfin, la protéine ALG10 catalyse l'addition du troisième résidu glucose en α-1,2, constituant l'étape finale de la synthèse du précurseur oligosaccharidique $Glc_3Man_9GlcNAc_2$-PP-Dol (**figure 13** ; Burda et Aebi, 1998). Ce dernier glucose est essentiel pour la reconnaissance du substrat par l'OST lors du transfert du précurseur oligosaccharidique sur la protéine.

II-1-2- Transfert du précurseur oligosaccharidique sur une protéine par l'OST.

Après avoir été synthétisé, le précurseur oligosaccharidique est transféré en bloc par l'OST sur la fonction amide de la chaîne latérale d'une asparagine appartenant à une séquence consensus Asn-X-Ser/Thr (où X peut-être un acide aminé quelconque à l'exception de la proline) d'un polypeptide en cours d'élongation (**figure 14** ; Imperiali et Hendrickson, 1995). Cette modification est réalisée dans la lumière du RE et peut s'effectuer de façon co- ou post-traductionnelle. L'OST utilise le précurseur $Glc_3Man_9GlcNAc_2$-PP-Dol en tant que substrat donneur et les protéines possédant un site consensus de N-glycosylation comme substrat accepteur (Bause, 1983 ; Breton et Imberty, 1999 ; Helenius et Aebi, 2004). Le transfert du précurseur sur le polypeptide en cours de synthèse aboutit à la formation d'une liaison covalente N-osidique. Cette addition glucidique modifie les propriétés biophysiques du polypeptide et affecte par conséquence son repliement de façon directe (Shental-Bechoret Levy, 2008 ; Hanson et al., 2009 ; Culyba et al., 2011) ou indirecte (Helenius et Aebi, 2004).

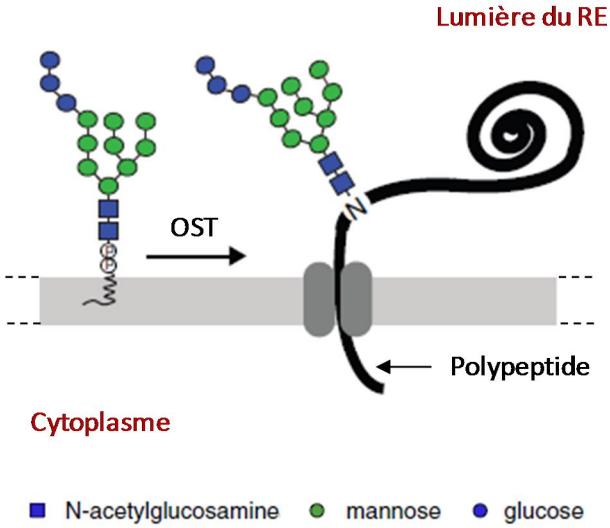


Figure 14 : Transfert de l'oligosaccharide précurseur $Glc_3Man_9GlcNAc_2$ sur un polypeptide en cours d'élongation par l'oligosaccharyltransférase (OST) d'après Schwarz et Aebi, 2011.

II-1-2-1- Structure et organisation de l'OST.

L'OST forme un complexe multimérique composé d'enzymes transmembranaires situées au niveau du RE (Yan et Lennarz, 1999 ; Mohorko et al., 2011). Il a été clairement établi que la sous-unité STT3 était responsable de l'activité catalytique de l'OST (Yan et Lennarz, 2002 ; Kelleher et al., 2003). Cette sous-unité est une protéine transmembranaire hautement conservée qui est retrouvée chez tous les eucaryotes étudiés jusqu'à ce jour. Dans certains organismes comme les protistes, elle porte à elle seule l'activité oligosaccharyltransférase (Hese et al., 2009). Par exemple, chez *Trypanosoma brucei* les protéines STT3 qui sont codées par trois paralogues différents, sont chacune autosuffisantes pour le transfert des oligosaccharides précurseurs (**figure 15** ; Nasab et al., 2008 ; Izquierdo et al., 2009).

A l'inverse, chez les eucaryotes supérieurs tels que les champignons, les animaux et les plantes, l'OST fonctionne comme un complexe multiprotéique constitué d'un maximum de huit polypeptides, construit comme une extension du core STT3 (**figure 15** ; Kelleher et Gilmore, 2006 ; Mohorko et al., 2011). Dans ces OST multimériques, certaines des sous-unités supplémentaires permettent la structuration du complexe alors que d'autres facilitent le processus de N-glycosylation. Ainsi, les petites sous-unités comme les Ost4 et Ost5 seraient impliquées dans l'assemblage du complexe OST (Spirig et al., 2005 ; Dumax-Vorzet et al., 2013). En revanche, la sous-unité Ost1 agit comme une protéine chaperonne favorisant la N-glycosylation des polypeptides (Wilson et al., 2005, 2007 et 2008). De même, les domaines

exposés à la lumière du RE de certaines grandes sous-unités peuvent interagir directement avec les substrats polypeptidiques empêchant ainsi leur repliement avant que la glycosylation n'ait lieu. C'est le cas, notamment, des sous-unités Ost3 et 6 qui possèdent une activité oxydoréductase et se lient aux polypeptides cibles par l'intermédiaire d'interactions non covalentes et de ponts disulfures transitoires (Schulz et al., 2009). Il semblerait que ces sous-unités permettent le ralentissement des premières étapes de repliement des protéines, en maintenant la chaîne polypeptidique dans une conformation développée plus propice à l'ajout du précurseur oligosaccharidique. Ainsi, le repliement des protéines peut être considéré comme un événement en compétition avec la N-glycosylation, en particulier pour les régions protéiques destinées à être localisées dans un domaine fortement structuré de la protéine finale. D'autre part, chez de nombreux organismes eucaryotes, l'OST est associée aux ribosomes et aux polypeptides en cours de synthèse (Whitley et al., 1996 ; Shibatani et al., 2005 ; Harada et al., 2009). Cette proximité augmente la concentration locale de substrat accepteur et favorise le processus de glycosylation avant le repliement de la protéine.

Le processus de N-glycosylation peut également avoir lieu de façon post-traductionnelle après un précédent ajout co-traductionnel de glycanne. Pour cela, les différentes isoformes de l'OST coopèrent et agissent de manière séquentielle. Par exemple, dans les cellules de mammifères, les isoformes STT3A et STT3B catalysent le transfert de l'oligosaccharide précurseur de façon co-traductionnelle. Néanmoins l'isoforme STT3B, contrairement à STT3A, est également capable de glycosyler la région C-terminale d'une protéine dépliée après la fin de la traduction (Ruiz-Canada et al., 2009).

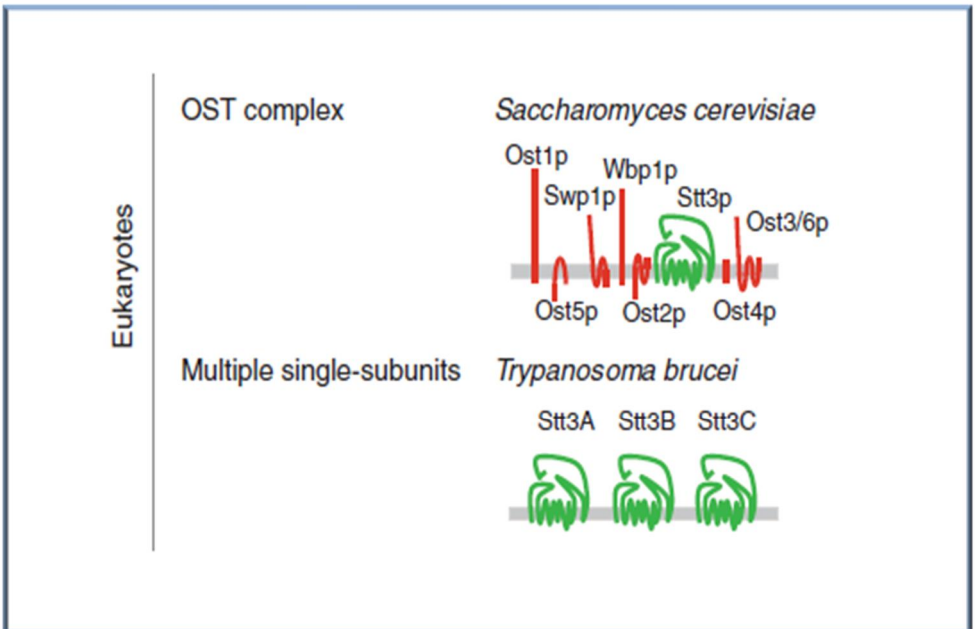


Figure 15 : Différentes structures générales de l'oligosaccharyltransférase (OST) chez les eucaryotes (Schwarz et Aebi, 2011). L'OST peut s'organiser sous forme de complexe oligomérique (ex : *Saccharomyces cerevisiae*) ou de multiple sous-unités indépendantes (ex : *Trypanosoma brucei*).

II-1-2-2- Spécificité de substrat de l'OST.

II-1-2-2-1- Spécificité pour le substrat donneur.

L'OST possède une spécificité de substrat très élevée pour le précurseur oligosaccharidique (Shrimal et al., 2013). L'oligosaccharide $Glc_3Man_9GlcNAc_2$-PP-Dol est très conservé dans les espèces animales, végétales et fongiques. Bien que des études *in vitro* aient révélé que le $GlcNAc_2$-PP-Dol était la structure minimale qui pouvait être acceptée par l'OST (Tai et Imperiali, 2001), les études *in vivo* indiquent au contraire une nette préférence pour l'oligosaccharide entièrement assemblé (Burda et al., 1999 ; Aebi, 2013). En effet, il a été montré que le résidu α-1,2-glucose terminal de la branche « a » est le facteur déterminant pour la reconnaissance du substrat par l'OST, alors que les résidus mannose des branches « b » et « c » sont d'une importance bien plus limitée (Burda et al., 1999 ; Aebi, 2013). Toutefois, les détails moléculaires du mécanisme de reconnaissance de l'oligosaccharide précurseur par l'OST restent encore inconnus.

A l'inverse, dans certaines espèces protozoaires, les précurseurs oligosaccharidiques présentent une structure moins complexe. Par exemple, chez le protozoaire *Trypanosoma brucei,* il a été mis en évidence que la protéine STT3A transfère l'oligosaccharide $Man_5GlcNAc_2$ sur des sites bien spécifiques de certaines protéines, tandis que la sous-unité STT3B utilise l'oligosaccharide $Man_9GlcNAc_2$ comme substrat pour tous les autres polypeptides accepteurs (Izquierdo et al., 2009). Ces données montrent qu'une sous-unité catalytique de l'OST peut distinguer seule des substrats oligosaccharidiques qui présentent une structure différente. Par ailleurs, la préférence des isoformes STT3A et STT3B pour leur substrat donneur serait due à la branche « c » de l'oligosaccharide $Man_9GlcNAc_2$-PP-Dol, de telle sorte que la présence de cette branche empêcherait la reconnaissance et / ou le transfert par STT3A, alors qu'elle améliorerait la reconnaissance et / ou le transfert par STT3B (Izquierdo et al., 2012).

II-1-2-2-2- Spécificité pour le substrat accepteur.

Contrairement au substrat oligosaccharidique, les polypeptides accepteurs de l'OST sont très divers. En effet, un grand nombre de protéines aux propriétés physico-chimiques variées sont N-glycosylées dans la voie de sécrétion. Elles possèdent toutes un peptide signal ubiquitaire leur permettant d'entrer dans le RE, ainsi qu'un ou plusieurs sites potentiels de N-glycosylation définis par la séquence Asn-X-Ser/Thr et dans de rares cas par Asn-X-Cys.

Malgré la grande diversité des polypeptides accepteurs, l'OST possède une grande spécificité pour la séquence Asn-X-Ser/Thr (Breuer et al., 2001). Ce tripeptide hautement conservé chez les eucaryotes est fréquemment retrouvé dans la séquence primaire des protéines (Marshall, 1972). Cependant, la présence d'un tel motif dans la séquence d'un polypeptide n'est pas suffisante pour définir un site réel de N-glycosylation. En effet, plusieurs études suggèrent que la séquence Asn-X-Ser/Thr ne peut être modifiée que si elle se situe dans un domaine souple du polypeptide (Whitley et al., 1996 ; Kowarik et al., 2002 ; Lizak et al., 2011). Ainsi, il est important de noter que tous les sites de N-glycosylation potentiels d'un polypeptide ne sont pas nécessairement glycosylés *in vivo*.

Les mécanismes moléculaires expliquant la spécificité de substrat de l'OST pour ces sites de N-glycosylation sont désormais connus. Le peptide accepteur forme une boucle de 180° et l'acide aminé hydroxylé à la position +2 de l'asparagine (sérine ou thréonine) est reconnu par un site de liaison de l'OST formé par un motif WWDYG très conservé dans la famille des protéines STT3 (Yan et Lennarz, 2002 ; Lizak et al., 2011). Par ailleurs, une préférence de l'OST pour les séquences Asn-X-Thr par rapport à Asn-X-Ser a été montrée dans toutes les espèces étudiées à ce jour (Miletich et Broze, 1990 ; Chen et al., 2007).

II-1-3- Repliement et contrôle qualité des glycoprotéines.

Une fois transféré sur le polypeptide, l'oligosaccharide agit comme une molécule de signalisation jouant un rôle essentiel dans les processus de repliement et de contrôle qualité de la protéine finale (Helenius, 1994 ; Parodi, 2000 ; Gidalevitz et al., 2013). Tout d'abord, sa structure hydrophile affecte la solubilité et le repliement des protéines (Wormald et Dwek, 1999 ; Helenius et Aebi, 2004). De plus, ce précurseur oligosaccharidique est partiellement déglucosylé par des α-glucosidases I et II permettant à la glycoprotéine d'entrer dans un cycle de repliement et de contrôle qualité impliquant différentes protéines dont deux lectines chaperonnes du RE que sont la calnexine et la calréticuline (Trombetta et Helenius, 1998 ; Williams, 2006). A l'issue de ce contrôle, les protéines mal repliées sont dégradées alors que les autres sont acheminées vers l'appareil de Golgi. Ces différentes étapes de contrôle, qui ont lieu dans le RE, sont cruciales pour le repliement correct des protéines sécrétées et sont hautement conservées chez les eucaryotes (Helenius et Aebi, 2001 ; Aebi et al., 2010).

II-1-3-1- Déglucosylation de l'oligosaccharide $Glc_3Man_9GlcNAc_2$ en $Glc_1Man_9GlcNAc_2$ par l'action successive des α-glucosidases I et II.

Immédiatement après son transfert sur le polypeptide, l'oligosaccharide $Glc_3Man_9GlcNAc_2$ peut subir l'élimination de son glucose distal catalysée par l'α-glucosidase I (**figure 16** ; Hettkamp et al., 1984; Shailubhai et al., 1991 ; Barker et Rose, 2013). Cette α-1,2-exoglucosidase est une protéine membranaire de type II appartenant à la famille des glycosylhydrolases 63 (GH 63 ; Calì et al., 2008). En outre, elle est associée au complexe de translocation à proximité de l'OST (Deprez et al., 2005). Tout comme la forme triglucosylée du précurseur, l'oligosaccharide $Glc_2Man_9GlcNAc_2$ lié à une protéine possède une demi-vie très courte et ne présentait jusqu'à présent, que peu d'intérêt. En effet, ce glycanne était uniquement considéré comme un intermédiaire transitoire rapidement hydrolysé par une α-glucosidase II. Toutefois, la caractérisation récente de la malectine montre que cette protéine membranaire du RE possède une activité putative de lectine capable de reconnaître les structures Glc_2-polymannosylées, ce qui suggère une possible fonction de l'intermédiaire $Glc_2Man_9GlcNAc_2$ dans le repliement et le contrôle qualité des protéines (Hebert, 2012).

L'α-glucosidase II commence tout d'abord par éliminer le second résidu glucose de l'oligosaccharide $Glc_2Man_9GlcNAc_2$ (**figure 16** ; Brada et Dubach, 1984 ; Kaewmuangmoon et al., 2013). Cette α-1,3-exoglucosidase appartient à la famille GH 31 (Henrissat et Romeu, 1995 ; Arendt et Ostergaard, 1997 ; Feng et al., 2004). Elle forme un hétérodimère composé d'une sous-unité α qui présente l'activité catalytique et d'une sous-unité β régulatrice (Trombetta et al., 1996 ; Calì et al., 2008 ; Watanabe et al., 2013). La sous-unité α est une protéine soluble qui possède un domaine lectine de type C permettant de se lier au résidu mannose et un motif protéique DMNE responsable de son activité catalytique (Trombetta et al., 2001 ; Feng et al., 2004). La sous-unité β contient, quant à elle, un domaine MRH (mannose 6-phosphate receptor homolog ; Munro, 2001 ; Olson et al., 2013) homologue à celui du récepteur au mannose 6-phosphate et un signal de rétention C-terminal XDEL lui permettant de se maintenir dans la lumière du RE (Trombetta et al., 1996 ; Trombetta et al ., 2001). Bien que la sous-unité α possède à elle seule l'activité hydrolytique, la sous-unité β est également nécessaire pour la déglycosylation de l'oligosaccharide $Glc_2Man_9GlcNAc_2$ (Wilkinson et al., 2006 ; Watanabe et al., 2009).

II-1-3-2- Cycle de contrôle qualité des glycoprotéines impliquant les chaperonnes calnexine et calréticuline.

Le glycanne monoglucosylé $Glc_1Man_9GlcNAc_2$ résultant de l'action de l'α-glucosidase II est reconnu par les lectines chaperonnes calnexine et calréticuline localisées dans le RE (**figure 16** ; Hahn et al., 1998 ; Molinari et al., 2004). Ces deux lectines associées à la co-chaperonne ERp57 favorisent le repliement des glycoprotéines dans le RE (High et al., 2000). La calnexine est une protéine transmembranaire de type I qui interagit préférentiellement avec les glycannes monoglucosylés liés aux polypeptides proches de la membrane, alors que la calréticuline qui est son homologue soluble s'associe plutôt avec les glycannes plus périphériques (Wada et al., 1995 ; Pieren et al., 2005 ; Olsen et al., 2013). Ces deux protéines sont composées de trois domaines : un domaine globulaire N-terminal avec une activité lectine permettant la liaison avec les oligosaccharides (Schrag et al., 2001), un domaine C-terminal responsable en grande partie de la liaison avec du Ca^{2+} et qui contient dans le cas de la calréticuline un signal de rétention dans le RE défini par le motif XDEL, et enfin un domaine P caractérisé par la répétition de courtes séquences riches en proline permettant leur association avec l'oxydoréductase ERp57 (Frickel et al., 2002 ; Paquet et al., 2005). Cette dernière catalyse la formation de ponts disulfures dans des polypeptides en cours de synthèse (**figure 16** ; Frickel et al., 2002 ; Kozlov et al., 2006).

L'α-glucosidase II clive ensuite le dernier résidu glucose générant de ce fait la structure $Man_9GlcNAc_2$ ce qui met fin à l'interaction entre le polypeptide et la calnexine ou calréticuline (**figure 16** ; Ellgaard and Helenius, 2001). Les protéines parfaitement repliées peuvent alors subir l'élimination d'un résidu α-1,2-mannose, préférentiellement sur la branche « b » du précurseur, par la mannosidase I du RE (Mast et Moremen, 2006 ; Wilson, 2012). Les glycoprotéines sont ensuite transportées vers l'appareil de Golgi. A l'inverse, les glycoprotéines mal repliées ou repliées de façon incomplète sont reconnues par l'UDP-glucose : glycoprotéine glucosyltransférase (UGGT) (Molinari et al., 2005). Cette protéine localisée dans le RE est un membre de la famille des glycosyltransférases 24 (Calì et al., 2008). Elle est composée d'un grand domaine N-terminal qui lui permet de se lier aux protéines non correctement repliées et d'un domaine C-terminal possédant une activité glucosyltransférase (Guerin et Parodi, 2003 ; Aebi et al., 2010). Cette bi-fonctionnalité est utilisée pour re-glucosyler spécifiquement et exclusivement les polypeptides présentant un défaut de conformation (**figure 16**). Les glycoprotéines mal repliées sont alors réintégrées dans le cycle calnexine / calréticuline, jusqu'à ce qu'elles soient à nouveau déglucosylées par

l'α-glucosidase II (Caramelo et Parodi, 2008). Si la conformation native n'est pas atteinte à l'issue de ce second cycle, la mannosidase I du RE clive un résidu α-1,2-mannose (**figure 16**, Aikawa et al., 2012). La glycoprotéine mal repliée est alors reconnue, puis transloquée à travers la membrane de RE par les protéines EDEM (ER Degradation-Enhancing Mannosidases), afin d'être dégradée par le système ERAD (ER associated protein degradation) couplé aux protéasomes dans le cytoplasme (**figure 16** ; Parodi, 2000 ; Meusser et al. , 2005 ; Olivari et al., 2006 ; Määttänen et al., 2010).

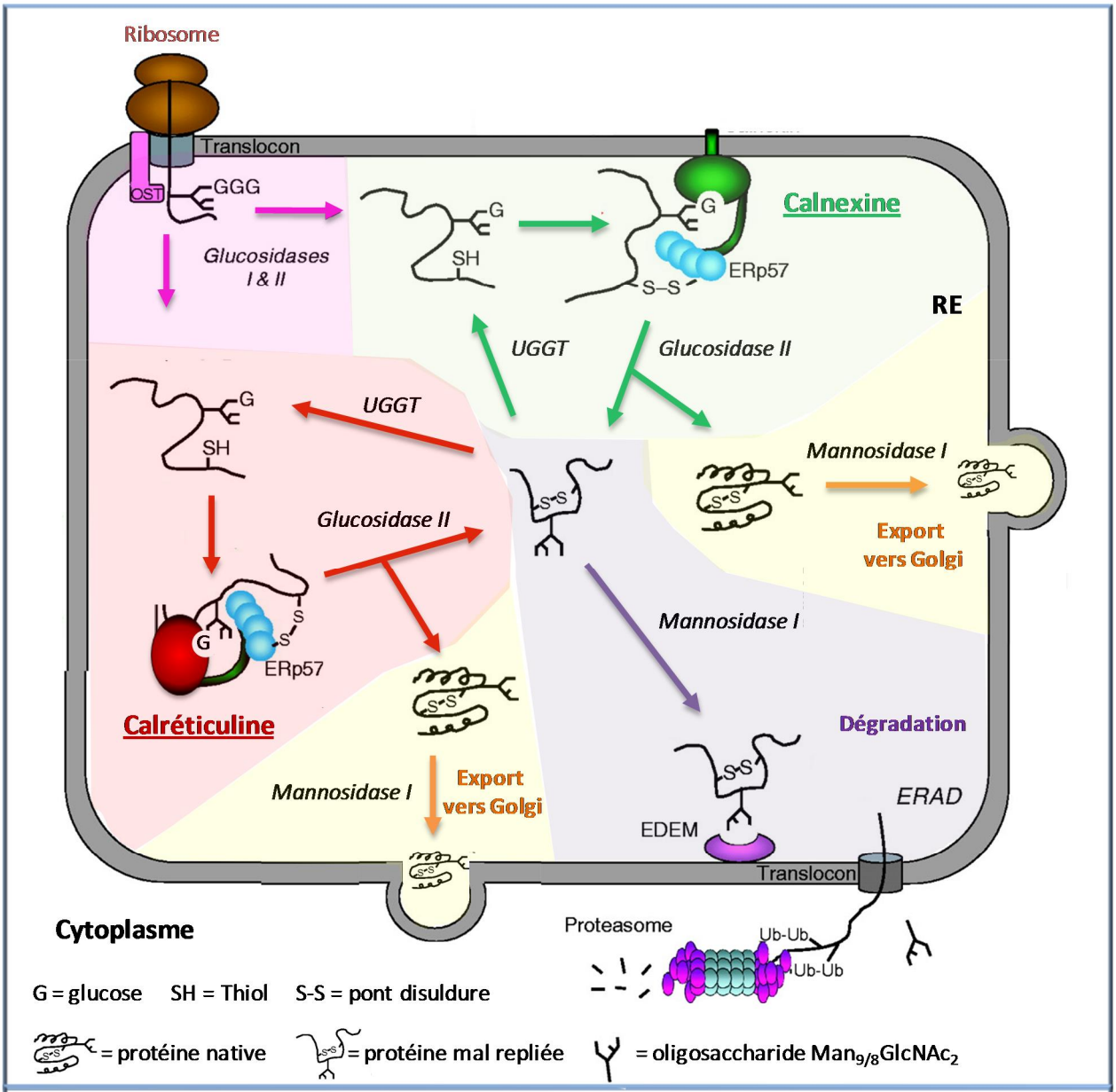


Figure 16 : Repliement et cycle de contrôle qualité des glycoprotéines N-liées dans le réticulum endoplasmique (RE ; d'après Williams, 2006). Ces processus impliquent deux lectines chaperonnes que sont la calnexine et la calréticuline ainsi que la co-chaperonne ERp57. La zone rose illustre la déglucosylation initiatrice de la glycoprotéine par la glucosidase I et II. Les zones verte et rouge correspondent respectivement au 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). La zone bleue/violette représente la voie de dégradation des glycoprotéines mal repliées par l'intermédiaire du système EDEM/ERAD/protéasome. Les zones jaunes illustrent la voie de transit des protéines natives vers l'appareil de Golgi faisant intervenir la mannosidase I du RE.

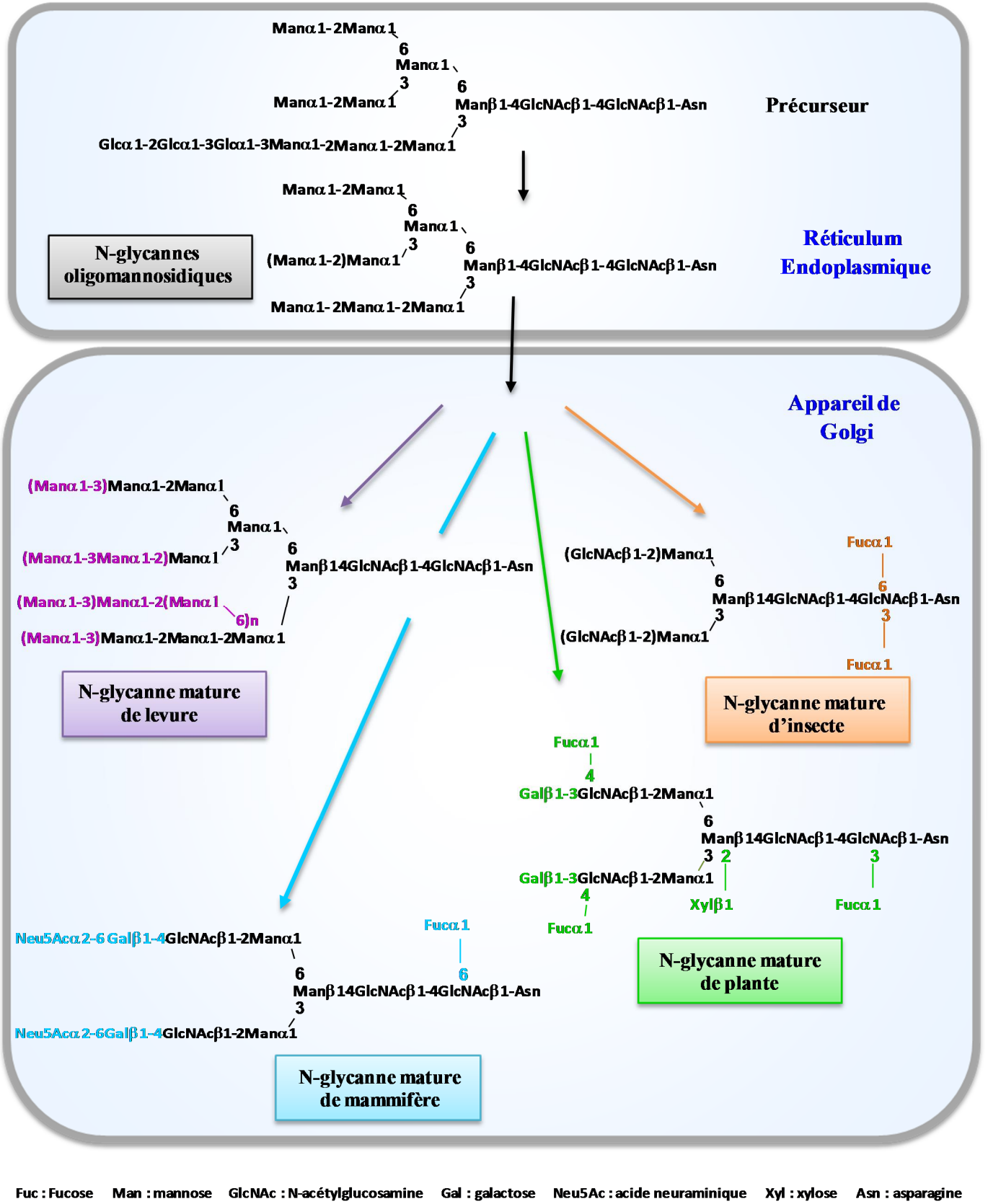


Figure 17 : Caractéristiques communes et points de divergence dans la biosynthèse des N-glycannes matures de mammifère, d'insecte, de plante et de levure. Chez les plantes, les mammifères et les insectes, les monosaccharides en noir représentent la structure bi-antennée commune. Les monosaccharides écrits en bleu, vert et orange correspondent respectivement aux motifs glucidiques spécifiques des mammifères, des plantes et des insectes. Chez les levures, les monosaccharides écrits en noir représentent la structure commune avec le précurseur oligomanosidique et ceux en violet correspondent aux motifs glucidiques spécifiquement ajoutés par cet organisme.

II-2- Appareil de Golgi : siège de la diversité glycannique.

La maturation des N-glycannes se poursuit lors du transport des glycoprotéines à travers l'appareil de Golgi, à condition que les chaînes oligosaccharidiques soient accessibles aux enzymes de maturations golgiennes (Faye et al., 1986 ; Lu et Wold, 1991 ; Czlapinski et Bertozzi, 2006). Contrairement à la glycosylation des protéines dans le RE qui est très conservée chez les eucaryotes, la maturation des N-glycannes dans l'appareil de Golgi a divergé au cours de l'évolution, aboutissant à la formation de structures glycanniques variées pouvant être organisme- voire tissu-spécifiques. C'est la présence et l'expression d'un répertoire bien particulier de glycosyltransférases et de glycosidases propres à chaque système qui est à l'origine de cette diversité glycannique. Ainsi, il a été montré que les glycoprotéines matures des levures portent des N-glycannes polymannosylés, alors que celles des eucaryotes supérieurs tels que les mammifères, les plantes et les insectes possèdent des N-glycannes complexes présentant une structure commune bi-antennée appelée core glycannique décorée de motifs glucidiques spécifiques (**figure 17**).

Les glycosyltransférases intervenant dans la maturation golgienne agissent de façon séquentielle et spécifique en utilisant des nucléotides sucres comme substrat. Ces enzymes sont des protéines membranaires de type II possédant, par conséquence, une région N-terminale comprenant une courte queue cytoplasmique, un domaine transmembranaire et une tige luminale. L'ensemble forme la région CTS (cytoplasmic-transmembrane-stem). Le domaine catalytique est quant à lui situé dans une région C-terminale localisée dans la lumière de l'appareil de Golgi (**figure 18** ; Breton et Imberty, 1999 ; Breton et al., 2006 ; Schoberer et al., 2013).

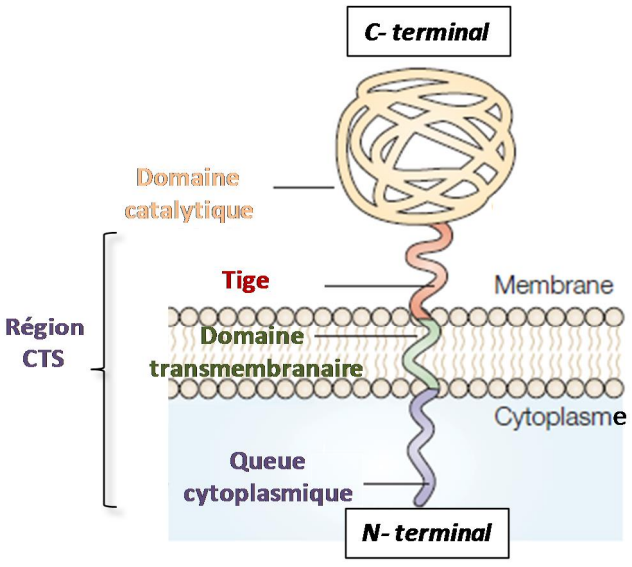


Figure 18 : Structure d'une protéine membranaire de type II (d'après Wildt et Gerngross, 2005). CTS = cytoplasmic-transmembrane-stem.

II-2-1- Maturation des N-glycannes chez les levures

La voie de N-glycosylation des levures a principalement été étudiée chez *Saccharomyces cerevisiae*. Comme dans tous les organismes eucaryotes, ce processus est initié dans le RE aboutissant à la formation d'un N-glycanne oligomannosidique de type $Man_8GlcNAc_2$ (Man-8) à l'exception de *Schizosaccharomyces pombe* et *Kluyveromyces lactis* dont le N-glycanne final du RE est l'oligosaccharide $Man_9GlcNAc_2$ (Man-9) (De Pourcq et al., 2010). Les glycoprotéines sont ensuite transportées vers l'appareil de Golgi, où la maturation des oligosaccharides est très différente de ce qui est retrouvé dans les cellules de mammifères, de plantes et d'invertébrés et peut présenter quelques variations suivant les espèces de levure.

Contrairement aux eucaryotes supérieurs, les levures ne sont généralement pas capables de démannosyler le glycanne $Man_8GlcNAc_2$. La maturation golgienne de leurs N-glycannes s'effectue donc par extension de la structure glycannique suite à l'ajout successif de résidus mannose catalysé par plusieurs mannosyltransférases (Tanner et Lehle, 1987). Tout d'abord, l'oligosaccharide $Man_{8/9}GlcNAc_2$ de la glycoprotéine nouvellement importée est modifié par l'action d'une α-1,6-mannosyltransférase Och1 qui catalyse l'addition d'un résidu mannose en α-1,6 sur l'α-1,3-mannose du core glycannique (**figure 19** ; Nakayama et al., 1992). Les levures sont généralement capables de former deux catégories de N-glycannes : les N-glycannes de type core qui possèdent moins de 15 hexoses et les N-glycannes hyperglycosylés contenant plus de 15 hexoses et dont l'extension s'effectue sur une chaine extérieure (Gemmill et Trimble, 1999). Pour former la structure des N-glycannes hyperglycosylés, la chaîne extérieure est allongée par l'ajout successif de nombreux résidus α-1,6-mannose catalysé par deux complexes protéiques que sont la mannane polymérase I (M-Pol-I) composée des protéines Mnn9 et Van1, et la mannane polymérase II (M-Pol-II) constituée des protéines Mnn9, Anp1, Mnn10, Mnn11 et Hoc1 (**figure 19** ; De Pourcq et al., 2010). Ce squelette d'α-1,6-mannoses est ensuite modifié par l'addition de résidus α-1,2-mannose par les α-1,2-mannosyltransférases Mnn2 et Mnn5, puis de résidus mannosylphosphate par les protéines Mnn4 et Mnn6, et enfin de résidus α-1,3-mannose par Mnn1 (**figure 19** ; Gemmill et Trimble, 1999 ; De Pourcq et al., 2010 ; Parsaie Nasab et al., 2013). Ces deux dernières étapes ne sont pas toujours présentes chez les levures. Par exemple, l'addition des résidus α-1,3-mannose n'existe pas chez *Pichia pastoris*. Les oligosaccharides hypermannosylés sont généralement associés à des glycoprotéines extracellulaires. A l'inverse les N-glycannes de type core sont, quant à eux, majoritairement retrouvés dans les organites cellulaires (Fukazawa et al., 1995). Ces glycannes ne présentent pas d'élongation de

la chaine extérieure par des résidus α-1,6-mannose, mais subissent directement l'ajout d'un résidu α-1,2-mannose, puis d'un mannosylphosphate, et enfin de deux α-1,3-mannoses (**figure 19** ; De Pourcq et al., 2010 ; Parsaie Nasab et al., 2013).

En plus des résidus mannoses et du phosphate, certaines espèces de levures sont également capables d'ajouter à la structure de leurs N-glycannes, du galactose, de l'acide glucuronique, du pyruvate et plus rarement du xylose (Fukazawa et al., 1995 ; Gemmill et Trimble, 1996). La composition et le nombre de chaque résidu est à l'origine de la grande diversité des N-glycannes chez les levures.

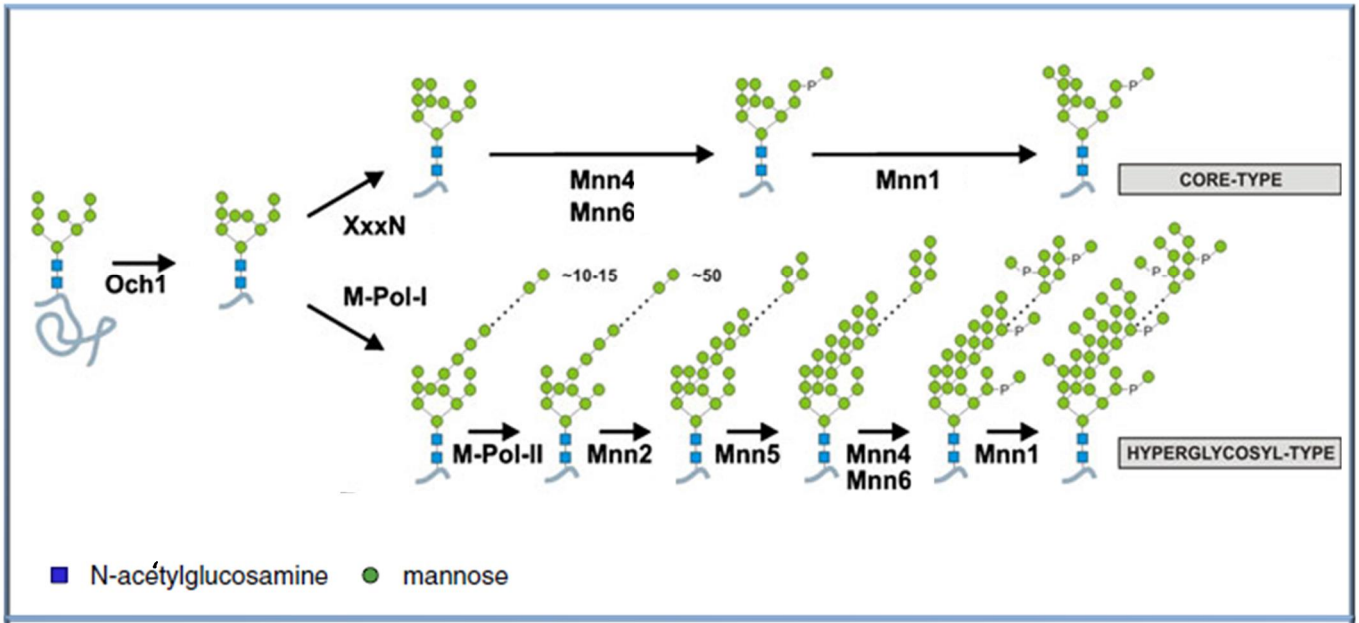


Figure 19 : Voie de N-glycosylation de la levure *Saccharomyces cerevisiae* (d'après De Pourcq et al., 2010). Ce schéma présente la synthèse des N-glycannes type core et des N-glycannes hyperglycosylés par l'ajout successif de résidus mannose et mannose-phosphate catalysé par un ensemble de mannosyltransférases dont les deux complexes protéiques mannane polymérase I (M-Pol-I) et II (M-Pol-II).

II-2-2- Voie de biosynthèse des N-glycannes complexes chez les mammifères, les plantes et les insectes : un processus initié par la GnT I.

Chez les eucaryotes supérieurs tels que les mammifères, les plantes et les invertébrés, la maturation golgienne des N-glycannes est initiée par l'α-mannosidase I (**figure 20** ; Lal et al., 1994 ; Kajiura et al., 2010). Cette enzyme catalyse l'hydrolyse d'un à 4 résidus mannoses de l'oligosaccharide Man-8/9 aboutissant à la formation de N-glycannes oligomannosidiques possédant de 8 à 5 résidus mannose (Man-8 à 5 ; Sturm *et al.*, 1987a ; Szumilo *et al.*, 1986b). Tout comme la mannosidase I intervenant dans le RE, elle appartient à la famille des glycosylhydrolases 47 (GH47) qui sont hautement spécifiques pour les résidus α-1,2-mannose

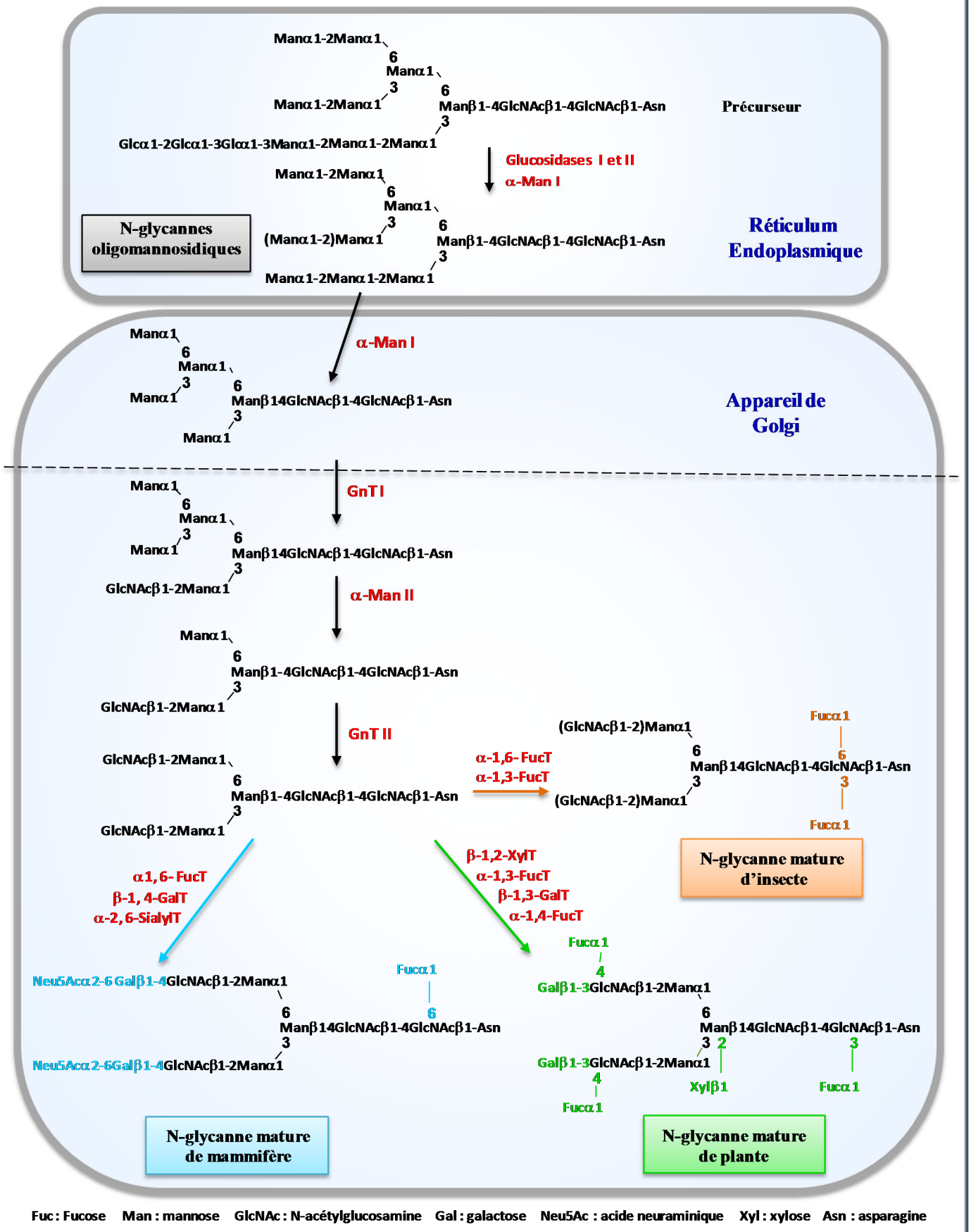


Figure 20 : Biosynthèse et maturation des N-glycannes de plante, de mammifère et d'insecte dans le réticulum endoplasmique et l'appareil de Golgi. Les motifs glucidiques spécifiques des plantes, des mammifères et des insectes sont respectivement notés en vert, bleu et orange. La ligne en pointillé sépare les N-glycannes oligomannosidiques des N-glycannes complexes. α-Man I : α-mannosidase I, GnT I : N-acétylglucosaminyltransférase I, α-Man II : α-mannosidase II, GnT II : N-acétylglucosaminyltransférase II, β-1,2-XylT : β-1,2-xylosyltransférase, α-1,3-FucT : α-1,3-fucosyltransférase, β-1,3-GalT : β-1,3-galactosyltransférase et α-1,4-FucT : α-1,4-fucosyltransférase.

et requiert du Ca^{2+} pour son activité catalytique (Lobsanov et al., 2002 ; Mast et Moremen, 2006). Jusqu'à cette étape, les glycoprotéines portent uniquement des N-glycannes oligomannosidiques dont le rôle est essentiellement de participer à la structure de la protéine.

La différenciation des N-glycannes oligomannosidiques en N-glycannes complexes est réalisée par la GnT I qui catalyse le transfert d'un résidu GlcNAc en β-1,2 sur le mannose α-1,3 du core glycannique du Man-5 (**figure 20** ; Vella et al., 1984 ; Tezuka et *a*l., 1992). L'addition de ce premier monosaccharide au niveau de l'appareil de Golgi initie donc la voie de biosynthèse des N-glycannes complexes. Ces derniers sont très importants d'un point de vue physiologique car ils confèrent aux protéines associées des fonctions biologiques supplémentaires. En effet, dans de nombreux cas, ces oligosaccharides servent de ligands dans des processus faisant intervenir des interaction protéine / sucre, comme le ciblage, l'adhésion cellulaire et la signalisation (Varki, 1993).

Suite à l'action de la GnT I, l'α-mannosidase II catalyse l'hydrolyse successive de deux résidus mannose liés à l'α-1,6-mannose du core glycannique du Man-5 (**figure 20** ; Kaushal et al., 1990 ; Altmann et März, 1995 ; Strasser et al., 2006). Un deuxième résidu GlcNAc est ensuite ajouté sur le mannose α-1,6 du core pentasaccharidique par la N-acétylglucosaminyltransférase II (GnT II), aboutissant à la formation du glycanne $GlcNAc_2Man_3GlcNAc_2$ (**figure 20** ; Johnson et Chrispeels, 1987 ; Strasser *et al.,* 1999b). Cet oligosaccharide constitue la dernière structure glycannique commune présente chez les eucaryotes supérieurs comme les mammifères, les plantes et les invertébrés.

Par la suite, un ensemble de glycosidases et de glycosyltransférases peuvent agir, aboutissant à la formation de N-glycannes complexes matures. Au cours de l'évolution, ces étapes de maturation golgienne ont divergé entre les différents systèmes eucaryotes ce qui a conduit à la formation de N-glycannes complexes matures variés qui peuvent être organisme-voire tissu-spécifiques. Les principales étapes de maturation à l'origine de cette diversité sont sous le contrôle de glycosyltransférases caractéristiques telles que les fucosyltranférases, les galactosyltransférases, les xylosyltransférase ou les sialyltransférases (**figure 20**).

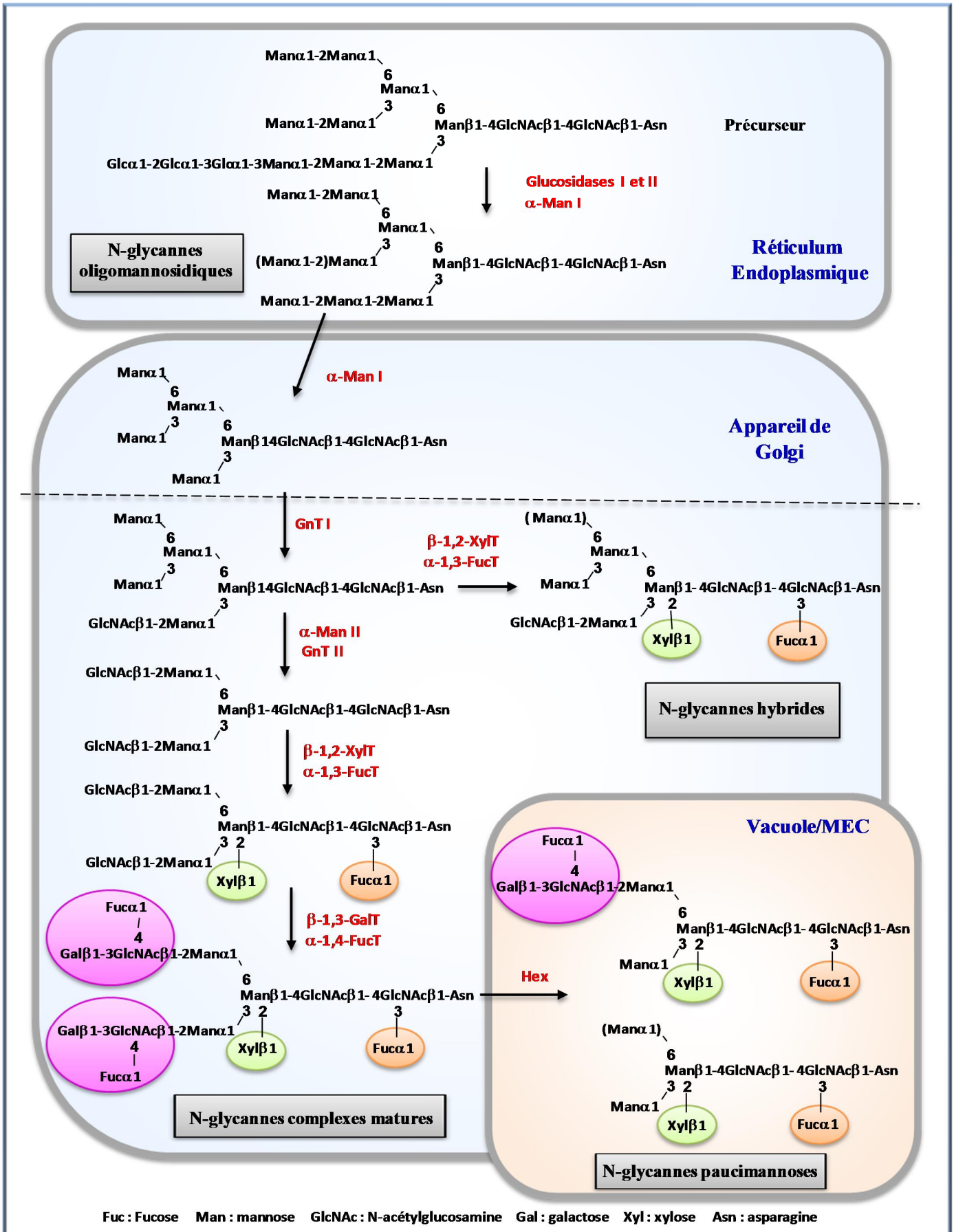


Figure 21 : Biosynthèse et maturation des N-glycannes de plante dans le réticulum endoplasmique, l'appareil de Golgi et les lieux de stockage comme la vacuole et la matrice extracellulaire (MEC). Les motifs glucidiques spécifiques de plante β-1,2-xylose, α-1,3-fucose et l'épitope Lewis[a] sont respectivement entourés sur le schéma en vert, orange et rose. α-Man I et II : α-mannosidase I et II, GnT I et II : N-acétylglucosaminyltransférase I et II, β-1,2-XylT : β-1,2-xylosyltransférase, α-1,3-FucT : α-1,3-fucosyltransférase, β-1,3-GalT : β-1,3-galactosyltransférase, α-1,4-FucT : α-1,4-fucosyltransférase, Hex : hexosaminidase et MEC : matrice extracellulaire.

II-2-2-1- N-glycannes matures des plantes.

Comme tous les autres eucaryotes supérieurs étudiés à ce jour, les glycoprotéines de plantes portent des N-glycannes complexes matures possédant un core glycannique commun $GlcNAc_2Man_3GlcNAc_2$. Ce core est substitué spécifiquement par des résidus β-1,2-xylose (β-1,2-Xyl), α-1,3-fucose (α-1,3-Fuc) et dans certains cas, par des épitopes Lewis^a terminaux ce qui aboutit à la formation de structures complexes caractéristiques de plante (**figure 21**). Les motifs spécifiques β-1,2-Xyl et α-1,3-Fuc sont reconnus comme épitopes antigéniques chez les mammifères possédant de ce fait un pouvoir allergisant et immunogène (Wilson et Altmann, 1998 ; van Ree et al., 2000 ; Bardor et al., 2003).

La maturation du core glycannique débute par l'intervention de l'α-1,3-fucosyltransférase (α-1,3-FucT) et de la β-1,2-xylosyltransférase (β-1,2-XylT) qui catalysent respectivement l'ajout d'un résidu fucose en α-1,3 sur la première GlcNAc proximale et d'un résidu xylose en β-1,2 sur le premier mannose du core (**figure 21** ; Johnson et Chrispeels, 1987 ;Breton et al., 1998 ; Lerouge et al., 1998 ; Wilson, 2002). Le transfert de ces deux résidus nécessite la présence d'au moins un résidu GlcNAc terminal sur le glycanne accepteur (Johnson et Chrispeels, 1987 ; Staudacher et al., 1995). Ces deux enzymes peuvent également agir directement sur le produit de la GnT I conduisant ainsi à la formation des structures hybrides (**figure 21** ; Johnson et Chrispeels, 1987 ; Oxley and Bacic, 1995 ; Gray *et al.*, 1996 ; Oxley *et al.*, 1996 ; Yang *et al.*, 1996). Plusieurs α-1,3-FucT de plantes ont déjà été caractérisées comme celle de *Vigna radiata* (Leiter et al., 1999) et celles d'*A. thaliana* (Wilson et al., 2001), appartenant toutes à la famille GT 10 (Breton et al., 1998). Il en est de même pour la β1,2-XylT déjà clonée chez *A. thaliana* (Strasser et al., 2000 ; Bencúr et al., 2005). Les maturations tardives des N-glycannes complexes de plante correspondent à la formation des épitopes Lewis^a qui résultent de l'action d'une β-1,3-galactosyltransférase (β-1,3-GalT) et d'une α-1,4-fucosyltransférase (α-1,4-FucT) sur une GlcNAc terminale (**figure 21** ; Melo et al., 1997 ; Fitchette-Lainé et al., 1997 ; Breton et al., 1998 ; Strasser et al., 2007).

Après leur maturation dans l'appareil de Golgi, les glycoprotéines sont dirigées soit vers la matrice extracellulaire (MEC), soit vers un organite de stockage (ex : vacuole). Au cours de ce transport ou dans le compartiment de stockage lui-même, les N-glycannes complexes peuvent subir une hydrolyse partielle catalysée par des hexosaminidases (Hex) aboutissant à la formation de glycannes appelés paucimannoses (**figure 21** ; Rayon et al., 1998 ; Gutternigg et al., 2007). C'est le cas notamment d'un grand nombre de glycoprotéines vacuolaires qui sont glycosylées par des paucimannoses de type $Man_{3/2}XylFucGlcNAc_2$.

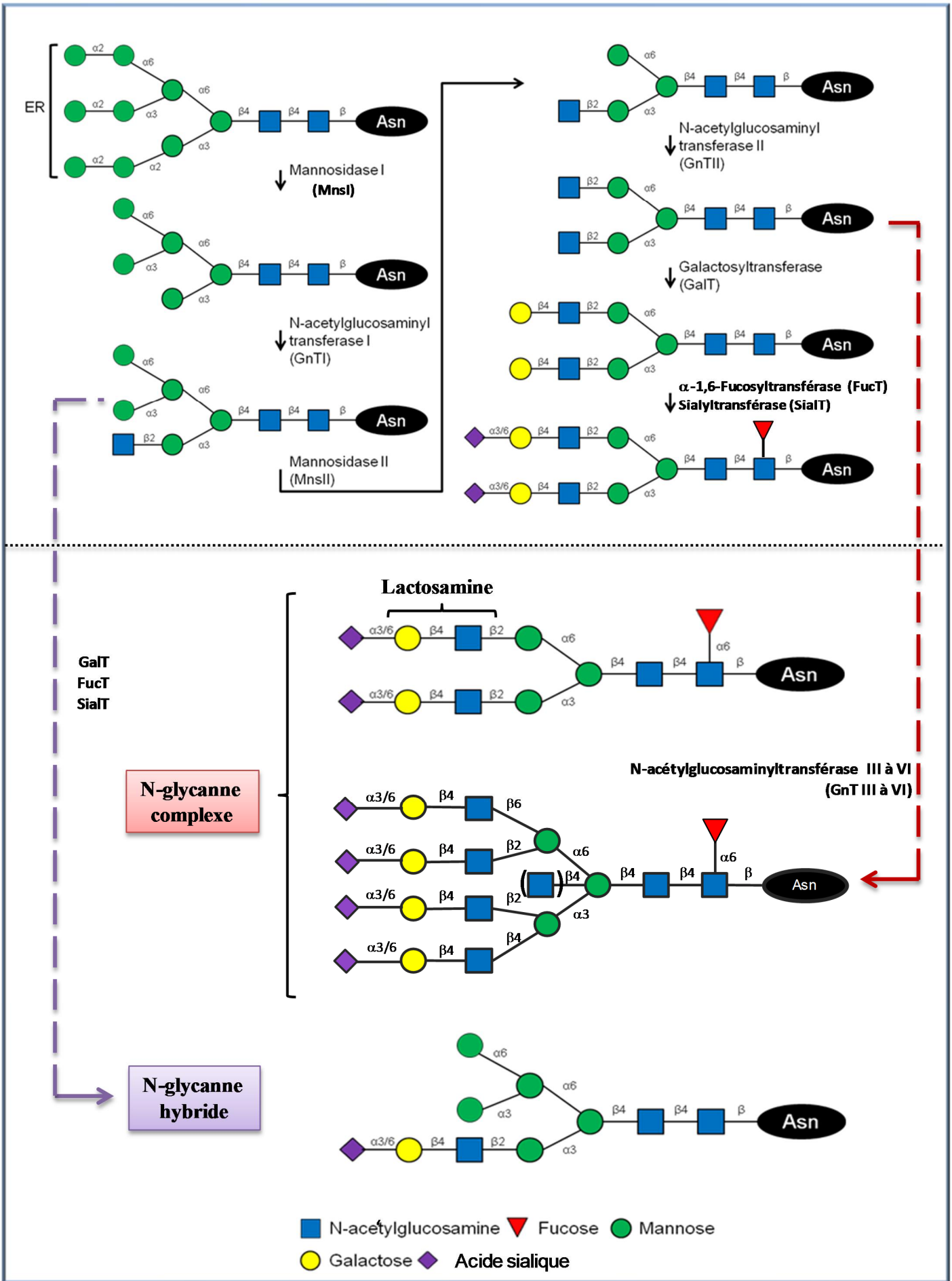


Figure 22 : Biosynthèse et maturation des N-glycannes de mammifère dans l'appareil de Golgi (d'après Nettleship, 2012). Ce schéma présente certaines voies de biosynthèse et les structures types de N-glycannes oligomannosidiques, hybrides et complexes bi- et tétra-antennés retrouvées sur les N-glycannes de mammifère. ER : Réticulum endoplasmique ; Asn : asparagine.

II-2-2-2- N-glycannes matures des mammifères.

La N-glycosylation des protéines de mammifère a fait l'objet de nombreux travaux qui ont abouti à une connaissance précise de leurs voies de biosynthèse et à la mise en évidence de la complexité du processus de maturation des N-glycannes dans ces organismes. En effet, les N-glycannes matures de mammifère présentent des structures variées possédant de deux à quatre antennes et/ou une GlcNAc bissectrice liée en β-1,4 au premier mannose du core (**figure 22** ; Takahashi et al., 2009). La biosynthèse de ces différentes structures est due à l'action des N-acétylglucosaminyltransférases III à VI sur l'oligosaccharide $GlcNAc_2Man_3GlcNAc_2$ agissant de concert voire en compétition (Brockhausen et Schachter, 1997).

Sur ces différentes antennes viennent se greffer divers résidus ou séquences oligosaccharides. Par exemple, un résidu galactose peut être ajouté en β-1,4 aux résidus GlcNAc distales sous l'action de la β-1,4-galactosyltransférase, ce qui diffère des plantes où le galactose est lié en β-1,3 (Breton et al., 1998 ; Imberty et al., 1999). L'association de ces deux sucres forme un disaccharide appelé lactosamine (**figure 22**). Ce composé peut être assemblé plusieurs fois à la suite, aboutissant à la formation de structures glycanniques de grande taille. Les résidus galactose peuvent également être modifiés par l'ajout d'acides sialiques en α-2,3- ou α-2,6 catalysé par des sialyltransférases (**figure 22** ; Jeanneau et al., 2004 ; Audry et al., 2011 ; Choi et al., 2013). De plus, contrairement aux plantes qui présentent un fucose lié en α-1,3 sur la première GlcNAc proximale, la fucosylation du core glycannique chez les mammifères est catalysée par une α-1,6-fucosyltransférase appartenant à la famille GT 23, qui permet l'ajout d'un résidu fucose lié en α-1,6 (**figure 22** ; Wilson et al., 1976 ; Breton et al., 1998 ; Voynow et al., 1991). De même, une fucosylation des résidus GlcNAc distale peut avoir lieu en α-1,3 ce qui aboutit à la formation d'épitope $Lewis^x$ correspondant au trisaccharide β-1,4-galactose/α-1,3-fucose/ GlcNAc distale.

Parallèlement à cette voie de maturation des N-glycannes complexes, les mammifères sont également capables d'ajouter ces différents résidus glucidiques directement sur le produit de la GnT I aboutissant à la formation de N-glycannes hybrides (**figure 22**). Les N-glycannes bi- à tétra-antennés représentés sur la **figure 22** sont donnés à titre d'exemple.

II-2-2-3- N-glycannes matures des insectes.

Les études structurales des N-glycannes d'insecte ont été menées sur des glycoprotéines, telles que la phospholipase A_2 (PLA_2) de venin d'abeille (Kubelka et al., 1993) ou encore des glycoprotéines isolées à partir de cellules d'insecte (Kubelka et al., 1994). Elles ont permis de montrer que les N-glycannes matures d'insecte présentent la particularité de pouvoir être bi-fucosylée sur la GlcNAc proximale. En effet, le core glycannique peut subir l'ajout de résidus fucose liés en α-1,3 et/ou en α-1,6 (**figure 23** ; Williams et al., 1991 ; Fabini et al., 2001 ; Wilson, 2002). Ces deux résidus sont ajoutés de façon ordonnée. Ainsi, il a été mis en évidence que l'α-1,6-fucosyltransférase est la première à agir en ajoutant un résidu fucose lié en α-1,6 sur la GlcNAc proximale de l'oligosaccharide $GlcNAc_2Man_3GlcNAc_2$. L'α-1,3-fucosyltransférase peut ensuite intervenir en catalysant l'ajout d'un fucose en α-1,3 sur l'oligosaccharide précédent ce qui aboutit à la formation d'un N-glycanne bi-fucosylé $GlcNAc_2Man_3Fuc_2GlcNAc_2$ (Staudacher et März, 1998 ; Paschinger et al., 2005). Certains N-glycannes d'insecte ont été décrits avec des GlcNAc terminales sur le bras α-1,3 ou sur le bras α-1,6 du core pentasaccharidique substituées par des résidus α-1,3 fucose et N-acétylgalactosamine (GalNAc) conduisant à la formation d'un motif glucidique Lewis-type (**figure 23** ; Kubelka et al., 1995 ; Vadaie et al., 2002 ; Takemae et al., 2003 ; Rendić et al., 2007). Par ailleurs, une dégradation des N-glycannes d'insecte est souvent observée par l'action de la β-N-acétylglucosaminidase aboutissant à la formation de paucimannoses tels que $Man_3Fuc_2GlcNAc_2$ (**figure 23** ; Altmann et al., 1995 ; Gutternigg et al., 2007 ; Schachter, 2009). *In fine*, ces composés sont les N-glycannes majoritairement retrouvés chez les insectes.

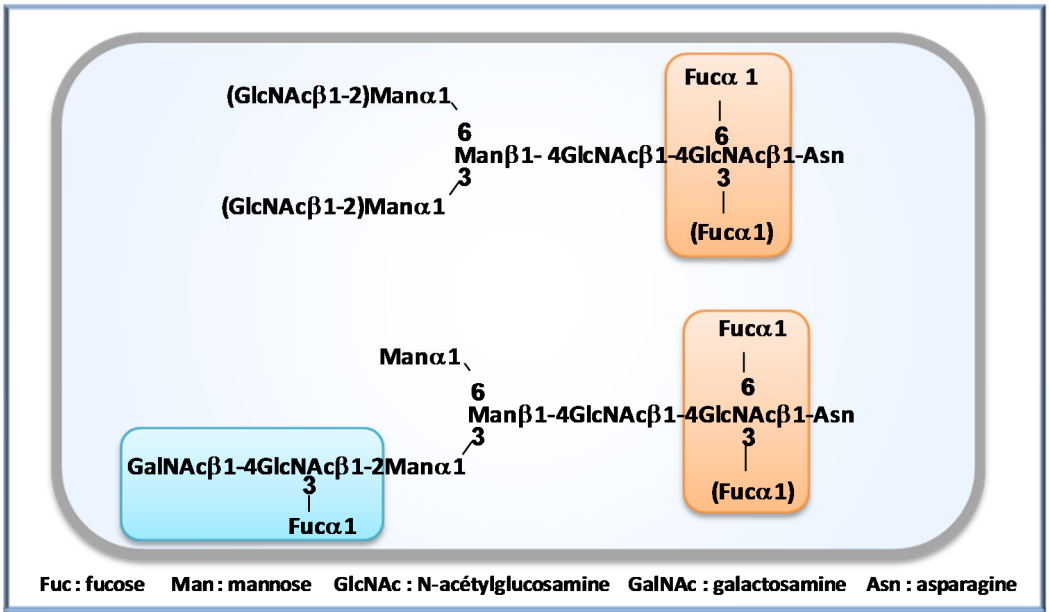


Figure 23 : Structures type de paucimannoses et de N-glycannes complexes d'insecte. Les motifs glucidiques spécifiques d'insecte que sont la bi-fucosylation α-1,6/α-1,3 et l'épitope Lewis-type sont respectivement entourés sur le schéma en orange et bleu.

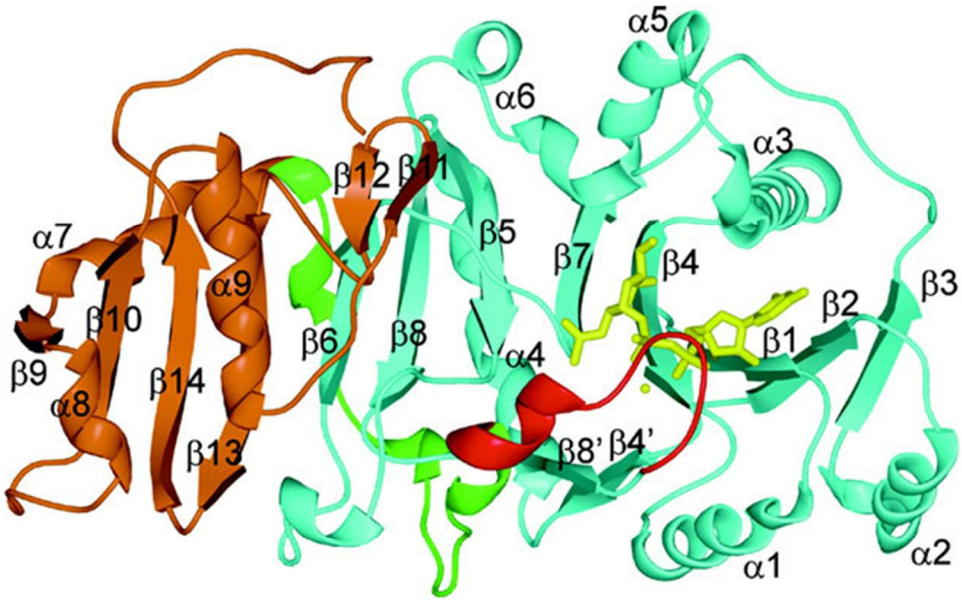
II-2-3- Structure et caractéristique de la GnT I.

Comme évoqué précédemment, la GnT I est une enzyme golgienne essentielle pour la maturation des N-glycannes oligomannosidiques en N-glycannes complexes voire hybrides (Kornfeld et Kornfeld, 1985). Elle catalyse le transfert d'un résidu GlcNAc à partir UDP-GlcNAc pour le lier en β-1,2 sur le substrat accepteur $Man_5GlcNAc_2$, ce qui génère la formation d'une structure glycannique de type $GlcNAcMan_5GlcNAc_2$. La GnT I est une glycosyltransférase de la famille GT 13 dont l'activité catalytique nécessite l'utilisation de l'ion Mn^{2+}. L'enzyme se lie d'abord au Mn^{2+} / UDP-GlcNAc puis à l'accepteur $Man_5GlcNAc_2$. Après le transfert, l'oligosaccharide $GlcNAcMan_5GlcNAc_2$ produit est ensuite relâché, suivi par l'UDP.

Chez les eucaryotes supérieurs, l'ajout de ce résidu GlcNAc est pré-requis pour l'action ultérieure de toutes les autres enzymes golgiennes intervenant dans la maturation des N-glycannes complexes (Johnson et Chrispeels, 1987). En effet, en l'absence de GnT I, les glycoprotéines portent uniquement des N-glycannes de type oligomannosidique (Strasser et al., 2005). Cependant dans de rares cas, généralement associée à une défaillance de la GnT I, une voie GnT I indépendante peut se mettre en place. Ce phénomène a notamment été mis en évidence dans des lignées cellulaires de mammifères (CHO et HEK) dont l'activité GnT I a été supprimée par mutation chimique de leur gène. Les N-glycannes extraits de ces cellules présentent majoritairement des structures oligomannosidiques mais également une petite population de N-glycannes oligomannosidiques substitués par un résidu fucose lié en α-1,6 sur la GlcNAc proximale (Lin et al., 1994 ; Crispin et al., 2006).

La GnT I a été largement étudiée dans un grand nombre d'organisme. En effet, la caractérisation de cette enzyme a déjà été menée chez plusieurs plantes comme le tabac (Strasser et al., 1999), *Arabidopsis thaliana* (Bakker *et al.*, 1999) et la pomme de terre (Wenderoth and von Schaewen, 2000), chez les invertébrés comme *Drosophila melanogaster* (Sarkar et Schachter, 2001) et *Caenorhabditis elegans* (Chen et al., 1999) mais également chez les mammifères comme le lapin (Unligil et al., 2000 ; Gordon et al., 2006). Ces études ont permis de mettre en évidence la présence de motifs très conservés responsables de son activité. Comme toutes les glycosyltransférases intervenant dans la maturation des N-glycannes au niveau de l'appareil de Golgi, la GnT I est une protéine membranaire de type II avec une partie globulaire C-terminale contenant l'activité catalytique. Cette région orientée dans la lumière de l'appareil de Golgi, est séparée en deux domaines : le domaine 1 en position N-terminale qui présente l'activité catalylique et le domaine 2 en position C-terminale (**figure 24** ; Unligil et al., 2000 ; Breton et al., 2006). C'est au sein du domaine 1

que tous les motifs conservés caractéristiques de cette enzyme et qui sont impliqués dans son activité, ont été retrouvés. En effet, ce domaine comprend un motif de type DxD (EDD) qui serait impliqué dans la liaison avec l'ion Mn^{2+} et dans l'activité catalytique (Breton et al., 1998), une boucle d'une vingtaine d'acides aminés permettant la fixation du nucléotide sucre associé au Mn^{2+} ainsi qu'un motif SQD qui interagit avec le cycle uracile du substrat donneur (**figure 24**).



<u>Figure 24</u> : Diagramme en ruban de la partie catalytique de la GnT I de lapin (Unligil et al., 2000). Le domaine 1 est représenté en bleu ciel, la boucle responsable de la liaison avec l'UDP-GlcNAc en rouge, la séquence reliant les domaines 1 et 2 en vert, le domaine 2 en brun et l'UDP-GlcNAc et l'ion Mn2 + en jaune.

II-2-4- Localisation subcellulaire des glycosidases et glycosyltransférases golgiennes.

Bien que la spécificité de substrat des glycosyltransférases et glycosidases golgiennes puisse expliquer à elle seule l'action séquentielle de ces enzymes dans la maturation des N-glycannes, il a été également montré que ces protéines étaient réparties tout le long de l'appareil de Golgi dans des compartiments bien distincts (Schoberer et al., 2013). En effet, les études de localisation sub-cellulaire, principalement réalisées avec des fusions de protéines fluorescentes, suggèrent que les enzymes agissant au début de la voie de maturation des N-glycannes comme la mannosidase I golgienne, la GnT I, la mannosidase II, la GnT II, les FucT du core glycannique mais également la XylT chez les plantes et les GnT III à V chez les

mammifères, ont tendance à résider dans les citernes cis et médian de l'appareil de Golgi (**figure 25** ; Pagny et al., 2003 ; Reichardt et al., 2007 ; Schoberer et Strasser, 2011). A l'inverse, la sialylation des N-glycannes de mammifère par les sialyltransférases (ST) et la formation de l'épitope Lewis^a par la β-1,3-galactosyltransférase et la α-1,4-fucosyltransférase chez les plantes ont lieu plus tardivement et semblent être localisé dans les citernes trans-golgiennes voire dans le réseau trans-Golgi (TGN). (**figure 25** ; Fitchette et al., 1999 ; Wildt et Gerngross, 2005 ; Schoberer et Strasser, 2011).

Cette répartition différentielle des glycosyltransférases et glycosidases serait en partie due à la région CTS de ces protéines membranaires. En effet, cette région ne servirait pas uniquement de signal d'adressage à l'appareil de Golgi, mais serait également impliquée dans la localisation des protéines dans les différentes citernes golgiennes (Saint-Jore-Dupas et al., 2006 ; Schoberer et al., 2009 ; Schoberer et al., 2013).

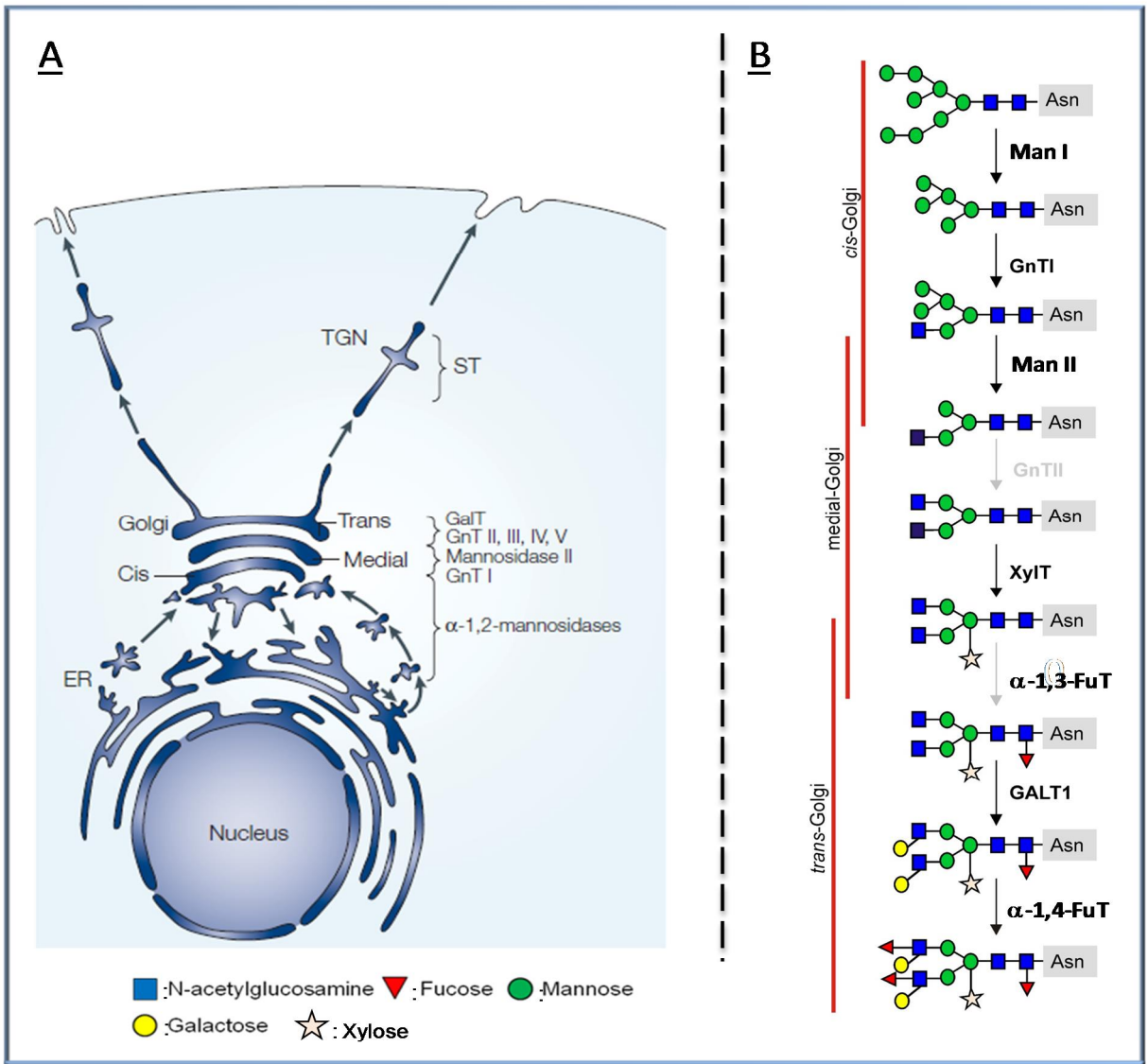


Figure 25 : Localisation sub-cellulaire des glycosyltransférases et glycosidases golgiennes impliquées dans la maturation des N-glycannes complexes (d'après Wildt et Gerngross, 2005 ; Schoberer et al., 2013). A : Répartition des enzymes golgiennes au sein du système sécrétoire de mammifère. B : Répartition des enzymes golgiennes de plante dans les différents compartiments de l'appareil de Golgi. ER : Réticulum endoplasmique ; Nucleus : noyau ; TGN : réseau trans-Golgi

II-2-5- Rôle des N-glycannes complexes.

Alors que les structures oligomannosidiques sont présentes chez tous les eucaryotes, les N-glycannes complexes sont généralement caractéristiques des organismes pluricellulaires. Ils sont impliqués dans de nombreux processus biologiques d'interaction et d'adhésion cellulaire comme l'endocytose, la reconnaissance immunitaire, l'inflammation, la pathogénicité, la signalisation et la motricité cellulaire (Varki, 1993 ; Parodi, 2000 ; Sun et al., 2005; Freeze et al., 2012). Chez les mammifères y compris chez l'homme, ces N-glycannes sont essentiels pour la survie et le développement normal de l'individu. En effet, la présence de mutations dans les gènes de la voie de N-glycosylation de ces organismes est à l'origine de sévères maladies congénitales et peut même s'avérer être embryo-létale (Aebi et Hennet, 2001 ; Grunewald et al., 2002 ; Eklund et al., 2005 ; Hennet, 2012). Ainsi, il a été montré que les embryons de souris mutés sur le gène de la GnT I, et donc dépourvus de N-glycannes complexes, mouraient environ 10 jours après fertilisation et présentaient plusieurs anomalies morphologiques et développementales (Ioffe et Stanley, 1994 ; Metzler et al., 1994). De même, l'inactivation de la GnT I du ver et de la mouche réduit fortement leur viabilité et perturbe leur activité locomotrice (Sarkar et al., 2006 ; Schachter, 2009). Au contraire, les cultures cellulaires animales présentant une mutation du gène de la GnT I croissent normalement (Stanley et al., 1975 ; Lin et al., 1994). Une déficience dans la fucosylation du core serait également impliquée dans l'apparition de maladies chez les mammifères. Il a ainsi été mis en évidence chez la souris qu'une déficience en activité α-1,6-fucosyltransférase générait un phénotype schizophrène ainsi qu'un retard de croissance (Fukuda et al., 2011).

A l'inverse chez les plantes, il est généralement admis que les N-glycannes complexes ne sont pas essentiels pour le développement normal de l'individu (Strasser et al., 2004). Pour exemple, le mutant *cgl I* d'*Arabidopsis thaliana*, dépourvu d'activité GnTI, est viable, fertile et ne montre aucun phénotype évident dans des conditions normales de culture (von Schaewen et al., 1993). Cependant, lorsque ces plantes sont cultivées en conditions de stress salin, elles présentent un phénotype fort, suggérant que les N-glycannes complexes de plante jouent un rôle dans des processus physiologiques spécifiques (von Schaewen et al., 2008 ; Kang et al., 2008). De même, il a été récemment démontré que l'inhibition de la GnT I chez le riz entraînait un défaut de croissance des plantules et de leur racine, un problème dans la synthèse de la paroi chez ce mutant et enfin, une perturbation de la voie de signalisation par les cytokinines (Fanata et al., 2013). De façon générale, les N-glycannes complexes ont semble-t-il fortement contribué à l'adaptabilité et à la réussite des espèces au cours de l'évolution.

III- Objectifs de la thèse.

Les données bibliographiques présentées dans cette introduction ont montré que les étapes de la voie de N-glycosylation dans le RE sont très conservées chez les eucaryotes, alors que la maturation des N-glycannes dans l'appareil de Golgi s'est diversifiée au cours de l'évolution, aboutissant à la formation de structures glycanniques pouvant être organisme- voire tissu-spécifiques. Ce processus à l'importance physiologique majeure a très largement été étudié chez les levures, les plantes, les mammifères et les invertébrés et continue à l'être encore aujourd'hui en raison des intérêts économiques qu'il suscite notamment dans le cadre de la production de protéines thérapeutiques dans des systèmes d'expression hétérologue. Néanmoins, à notre connaissance, cette voie de biosynthèse n'avait encore jamais été étudiée chez les microalgues avant le début des travaux présentés dans cette thèse. Par ailleurs, la recherche sur les diatomées constitue un intérêt fondamental dans la compréhension des différents processus biologiques comme la N-glycosylation, du fait de la position phylogénétique et de l'origine évolutive de ces organismes. Plusieurs diatomées modèles ont été développées ces dernières années, en particulier *P. tricornutum* représentante des diatomées pennées qui a été choisie en fonction de ses avantages scientifiques (particularités biologiques) et techniques (transformable, cycle de vie contrôlé en laboratoire, approches génétiques, génome séquencé), ainsi que pour ses applications biotechnologiques potentielles. C'est dans ce contexte scientifique, que s'est inscrit mon projet de thèse dont le but était d'initier l'étude de la voie de N-glycosylation de la diatomée modèle *P. tricornutum*. Pour cela, il s'agissait dans un premier temps de déterminer par analyse bioinformatique du génome de cette diatomée, la présence de gènes potentiellement impliqués dans la biosynthèse des N-glycannes, puis de mener des études biochimiques afin de caractériser les structures glycanniques portées par les protéines endogènes de *P. tricornutum*.

Le second objectif de ma thèse était de réaliser la caractérisation du gène identifié dans le génome de cette diatomée comme codant potentiellement pour une GnT I, par complémentation fonctionnelle dans des cellules de mammifère déficientes en activité GnT I endogène. Cette étude a été complétée par l'analyse du niveau d'expression de ce gène par PCR quantitative. L'ensemble de ces travaux est présenté sous la forme d'un article scientifique dans la première partie des résultats. La seconde partie consiste à présenter des résultats complémentaires sur l'expression du gène de la GnT I, ainsi que sur des analyses préliminaires initiées au cours de cette thèse sur les gènes identifiés dans le génome de *P. tricornutum* comme codant potentiellement pour des fucosyltransférases. Ces résultats complémentaires permettront d'apporter des éléments supplémentaires pour la discussion.

Chapitre II : Matériel et méthodes

I- MATERIEL

I-1- Matériel végétal : *Phaeodactylum tricornutum.*

Phaeodactylum tricornutum est une diatomée pennée identifiée en 1897 par K. Bohlin. Les travaux décrits dans ce manuscrit ont été réalisés essentiellement sur des cellules issues de la souche séquencée Pt1.8.6 (CCAP1055/1). La souche Pt3 a également été utilisée lors de l'étude de l'expression du gène de la GnT I en fonction du morphotype. Ces microalgues ont généreusement été fournies, à l'initiation de ce travail, 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.

I-2- Matériel animal : Cellules d'ovaire de hamster chinois (CHO).

Les cellules CHO Lec1, mutées sur le gène de la GnT I endogène, ont été achetées chez « American Type Culture Collection » (ATCC, CRL-1735 ; **figure 26**). La lignée sauvage (CHO) correspondante a été aimablement fournie par le Dr H. Castel du laboratoire DC2N, INERM U982 de l'université de Rouen. Ces deux lignées CHO ont été cultivées dans du milieu α-MEM (Invitrogen) complémenté avec 10% de sérum de veau fœtal (SVF, Fisher), puis placées dans un incubateur à 37°C contenant 5% de CO_2. Ces cellules ont été utilisées pour la mise en évidence *in vivo* de l'activité de la GnT I de *P. tricornutum*.

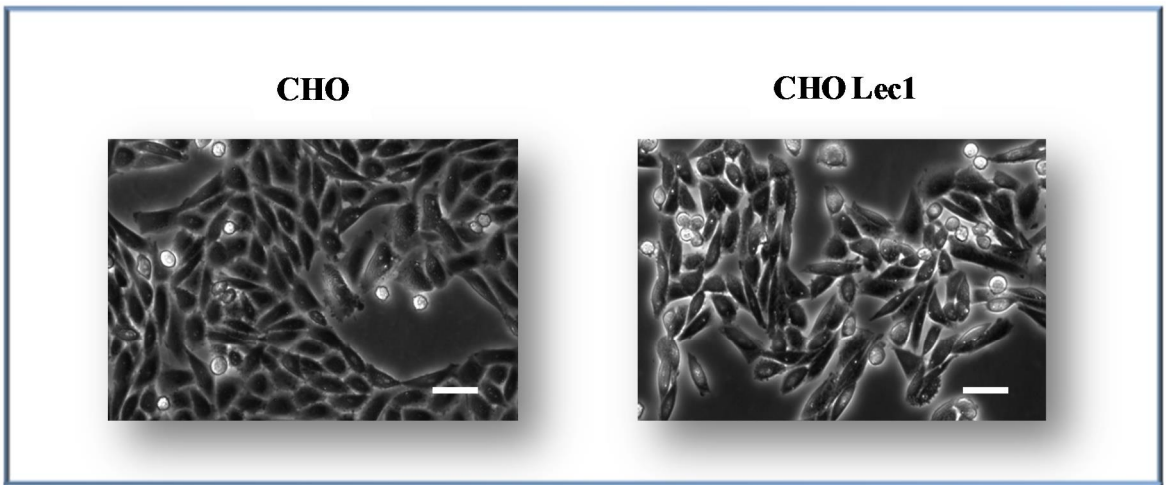


Figure 26 : Cellules d'ovaire de hamster chinois (CHO) sauvages et CHO Lec1, lignée mutante déficiente en activité N-acétylglucosaminyltransférase I endogène. Photos prises par le Dr Muriel Bardor au microscope optique avec un grossissement x400. La barre d'échelle indiquée par un trait blanc correspond à 50 µm.

Les cellules CHO FuT8-/-, mutées sur le gène codant pour la fucosyltransférase 8, sont issues de la lignée sauvage CHO K1 (**Figure 27**). Ces deux lignées ont été généreusement fournies par le Dr Gregory J. Cost de Sangamo Biosciences en Californie, USA. Ces cellules ont été cultivées dans du milieu F12 (Invitrogen) complémenté avec 10% de SVF, puis placées dans un incubateur à 37°C contenant 5% de CO_2. Elles ont été utilisées dans l'étude de l'activité *in vivo* des trois fucosyltransférases de *P. tricornutum*.

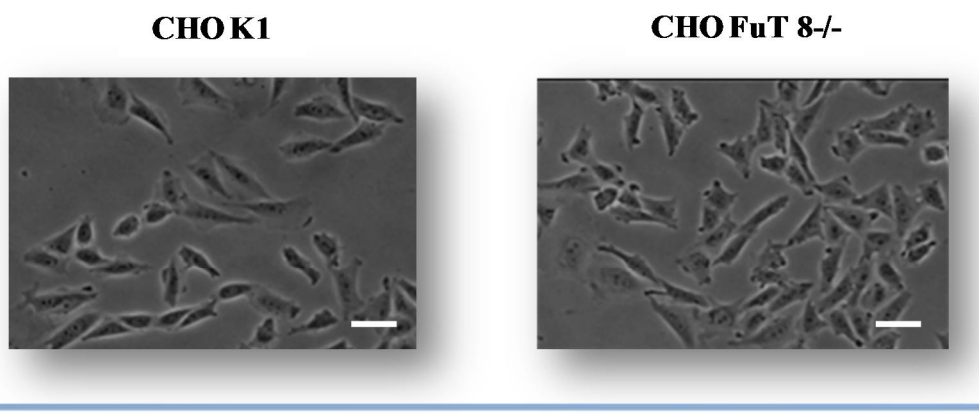


Figure 27 : Cellules d'ovaire de hamster chinois sauvages (CHO K1) et CHO FuT8-/-, lignée mutée sur le gène de la fucosyltransférase 8 (FuT8) endogène. Photos prises par Clément Ovide au microscope optique avec un grossissement x400. La barre d'échelle indiquée par un trait blanc correspond à 50 µm.

I-3- Plasmides et vecteurs utilisés.

I-3-1- pUC57+FuT46109 et pUC57+FuT54599

Les gènes codant potentiellement pour les fucosyltransférases 46109 et 54599 ont été synthétisés par la société Genecust (www.genecust.com) à l'intérieur du vecteur pUC57. Les cartes respectives des plasmides pUC57+FuT46109 et pUC57+FuT54599 sont présentées sur la **figure 28**.

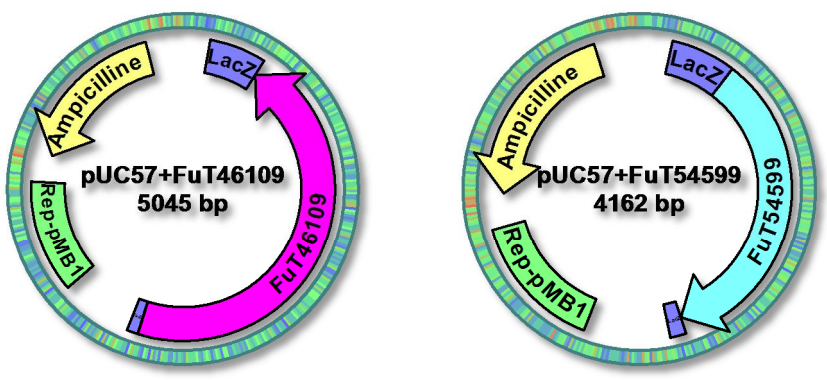


Figure 28 : Cartes des vecteurs pUC57+FuT46109 et pUC57+FuT54599 réalisées avec le logiciel pDRAW32 version 1.1.114.

I-3-2- 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 de mammifère. 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. La cassette de polyclonage est encadrée par les séquences T7 et BGH utilisées pour le séquençage du plasmide recombinant. Ce vecteur permet l'ajout de deux étiquettes en 3' du produit PCR que sont les séquences nucléotidiques codant pour un épitope V5 et une queue hexahistidine. Il possède également deux gènes de sélection : le gène de résistance à l'ampicilline qui est utilisé lors du clonage dans les bactéries et celui de la néomycine permettant la sélection des cellules de mammifère possédant le transgène. La carte du vecteur est détaillée sur la **figure 29**.

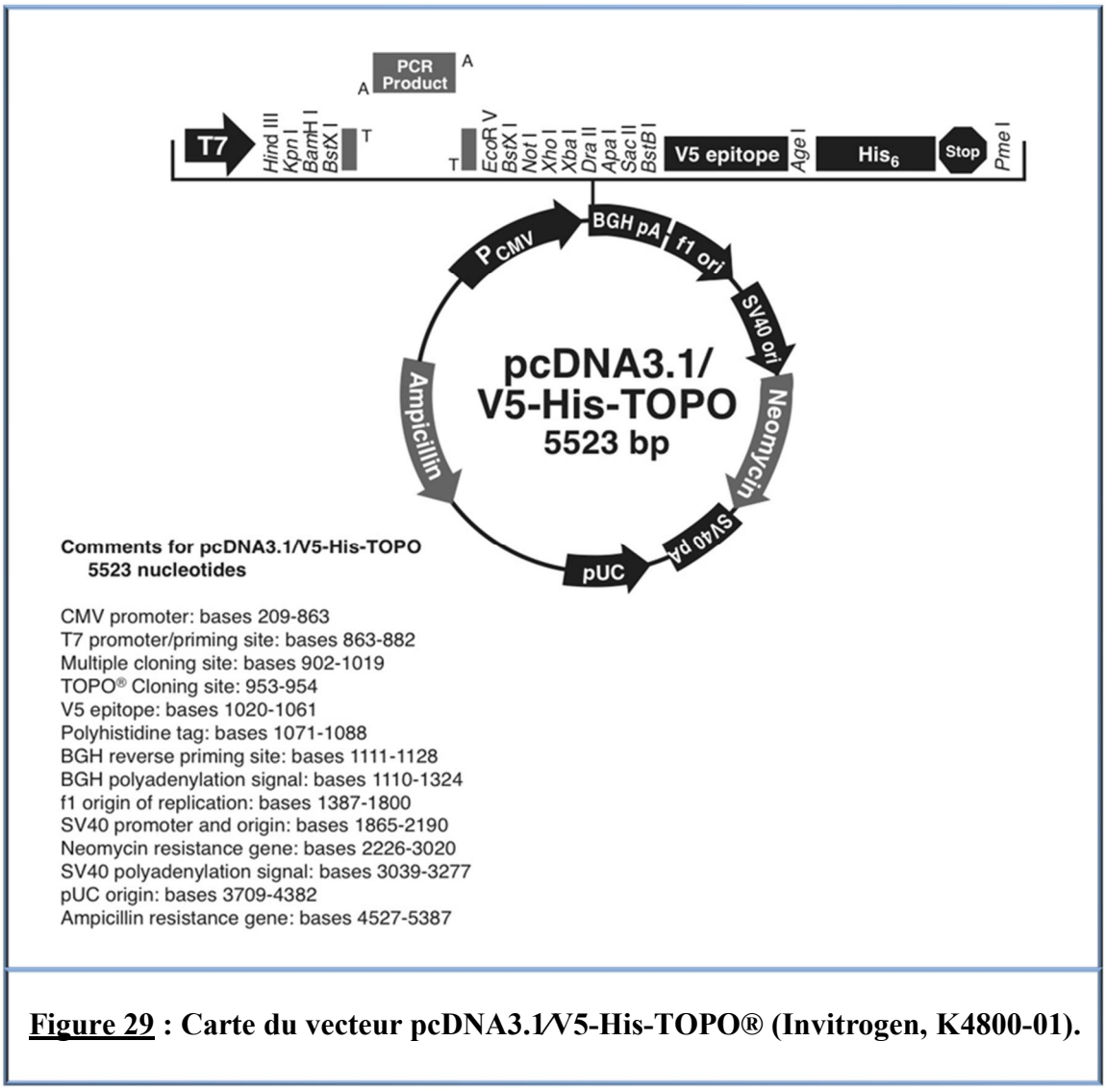


Figure 29 : Carte du vecteur pcDNA3.1/V5-His-TOPO® (Invitrogen, K4800-01).

I-4- Amorces PCR.

Les séquences des amorces utilisées pour les expériences de clonage, séquençage et analyses par RT PCR (*Reverse Transcription* PCR) ont été définies à l'aide du programme *Oligonucleotide Properties Calculator* (http://www.basic.northwestern.edu/biotools/oligocalc.html), puis commandées chez Eurogentec (http://www.eurogentec.com). Les amorces utilisées pour la PCR quantitative en temps réel (Q-PCR) ont été élaborées à l'aide du programme *Primer3Plus* (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Dans les deux cas, les séquences ont été définies en veillant à ce que le pourcentage de GC soit compris entre 45 et 55. La liste des amorces utilisées est présentée dans le **tableau 1**.

Gène cible	Nom de l'amorce	Séquence nucléotidique	Fonction(s)
GnT1	GNT1-5F	5'-ATGCGGTTGTGGAAACGTAC-3'	Clonage, séquençage et RT-PCR
	GNT1-3R	5'-TCTTTTCGGTGACGGAATG-3'	Clonage et séquençage
	GNT1-Seq1F	5'-TATTGCGGTCCGACTTTTTC-3'	Séquençage
	GNT1-Seq2R	5'-TCTGGGATGGTGTACGTCAA-3'	
	GNT1-Seq4F	5'-ACGCACCCGGGGCGTAC-3'	
	GNT1-Seq5F	5'-GACAAACGCTGTCCGGGAG-3'	
	GNT I AS12	5'-CTCGCGTACCACACGTTGG-3'	RT-PCR
	GNT1-Q-Fwd	5'-CGTACGAATCGCCCTTACTC-3'	Q-PCR
	GNT1-Q-Rev	5'-TTGCCGTCTTGTGAAATTACC-3'	
FuT46109	FuT46109 5F	5'-ATGCAGAGTTTGAATTTGACC-3'	Clonage
	FuT46109 3R	5'-TGACGATACGTAAAATAACTC-3'	
FuT46110	FuT46110 5F	5'-ATGATTCGCATACCATACGCG-3'	Clonage
	FuT46110 3R	5'-CACGAAGGTTTCCACAGGATT-3'	
FuT54599	FuT54599 5F	5'-ATGTCACTTCGCAAGCGTGGTG-3'	Clonage
	FuT54599 3R	5'-CGGATCGAACTTCCACGAGG-3'	
Actine 1	Act1 F	5'-GATTGTTGCTCCTCCGGAACG-3'	RT-PCR
	Act1 R	5'-GCTCAGTAGCTTAGAAGCACTTGCG-3'	
Histone H4	Q-H4-F	5'-AGGTCCTTCGCGACAATATC-3'	Q-PCR
	Q-H4-R	5'-ACGGAATCACGAATGACGTT-3'	
RPS (Ribosomal Protein Small subunit 30 S)	Q-rps-fw	5'-CGAAGTCAACCAGGAAACCAA-3'	Q-PCR
	Q-rps-rv	5'-GTGCAAGAGACCGGACATACC-3'	
Vecteur pcDNA3.1/V5-His-TOPO®	T7	5'-TAATACGACTCACTATAGGG-3'	Clonage
	BGH	5'-TAGAAGGCACAGTCGAGG-3'	

Tableau 1 : Liste complète des amorces PCR utilisées dans ce projet.

Les séquences des amorces de l'histone H4 et de la RPS (Ribosomal Protein Small subunit 30 S), présentées dans le **tableau 1**, ont été définies par (Siaut et al., 2007). Les amorces T7 et BGH ont, quant à elles, été fournies dans le kit du vecteur pcDNA™3.1/V5-His TOPO® TA Expression (Invitrogen).

II- METHODES

II-1- Mode de culture

II-1-1- Culture en Batch

Deux méthodes de culture en batch ont été utilisées :

La souche Pt1.8.6 de *P. tricornutum* a été cultivée par le laboratoire PBA, dans des bouteilles de deux litres contenant de l'eau de mer naturelle filtrée à 0.22 µm et complémentée avec 40 $mg.L^{-1}$ de sodium métasilicate et 0.1% (v/v) de milieu de Conway. Ce dernier est un mélange de trois solutions dont la composition est indiquée ci-dessous :

Solution principale :	
-Na_2-EDTA	45 g
-$NaNO_3$	100 g
-H_3BO_3	33,6 g
-NaH_2PO_4, 2 H_2O	26 g
-$MnCl_2$, 4 H_2O	0,36 g
-$FeCl_3$, $6H_2O$	1,28 g
-Eau distillée	qsp 1 L

(A dissoudre à chaud puis autoclaver)

Solution traces de métaux :	
-$ZnCl_2$	21 mg
-$CoCl_2$, $6H_2O$	20 mg
-$(NH_4)6MO_7O_{24}$, $4H_2O$	9 mg
-$CuSO_4$, $5H_2O$	20 mg
-Eau distillée	qsp 1 mL

Solution vitaminique :	
-Vitamine B1 thiamine hydrochloride	0,1 g
-Vitamine B12 cyanocobalamine	5 mg
-Eau distillée	qsp 50 mL

Les ballons ont été placés dans une chambre de culture à 20°C sous une illumination continue de 280-350 µmol photons.m^{-2}.s^{-1}. Les cultures, ayant atteint une concentration de 20.10^6 cellules.mL^{-1}, ont été centrifugées à 5000 g pendant 20 minutes à 4°C. Les culots cellulaires ainsi obtenus ont été lyophilisés puis envoyés au laboratoire GlycoMev où ils ont été stockés à -20°C jusqu'à leur utilisation dans les analyses biochimiques.

P. tricornutum a également été cultivée, au sein du laboratoire Glyco-MEV, dans des erlenmeyers de 500 mL, par repiquage de 25 mL de la culture mère dans 200 mL de milieu frais. Ce milieu est composé d'eau de mer artificielle contenant 33,33 g de sel marin par litre (Aquarium Systems Instant Ocean), filtrée à 0.45µm puis autoclavée pendant 20 minutes à 1 bar. Il a ensuite été complémenté avec 40 $mg.L^{-1}$ de sodium métasilicate (VWR, 28092.290) et 0.1% (v/v) de milieu de Conway fourni par le laboratoire PBA. Les microalgues ont alors été placées dans une armoire de culture, à 19°C, pendant une semaine, sous une irradiation de 50 $\mu Einstein.m^{-2}.s^{-1}$ avec une alternance jour/nuit : 16 heures/8 heures. Ces conditions de culture ont été utilisées pour l'entretien de la souche Pt1.8.6 employée pour le clonage des gènes codant la GnT I et la FuT 46110 de *P. tricornutum*. Il en est de même pour les souches Pt1.8.6 et Pt3 cultivées en morphotype fusiforme et Pt3 triradié utilisées lors des analyses par RT-PCR de l'expression de la GnT I en fonction du morphotype. La souche Pt3 cultivée en morphotype ovale a quant à elle été obtenue dans un milieu similaire mais ne contenant que 10% d'eau de mer (conditions de stress salin).

II-1-2- Culture continue

Une culture continue de la souche Pt1.8.6 de *P. tricornutum* a été réalisée par le laboratoire PBA dans le même milieu d'eau de mer naturelle que celui décrit précédemment et en suivant le protocole indiqué dans (Bougaran et al., 2003). Cette culture a été utilisée lors des expériences de PCR quantitative en temps réelle visant à déterminer le niveau d'expression du gène de la GnT I en fonction du taux de nitrate présent dans le milieu de culture. Pour cela la concentration en nitrate, contenue dans la culture, a été modulée dans le temps, générant ainsi trois phases : condition standard (phase 1 ; $[NO_3]$=1mM), condition nitrate déplété (phase 2 ; $[NO_3]$=0,05mM) et condition nitrate replété (phase 3 ; $[NO_3]$=1mM). Quatorze prélèvements ont été effectués au cours de cette étude : 5 en phase 1, 6 en phase 2, et 3 en phase 3 (voir **tableau 1** paragraphe I-4). Un dénombrement de la culture et une mesure de la densité optique (DO) à 680 nm et 800 nm ont été réalisés sur chaque prélèvement avant d'être réparti en trois échantillons de 30 mL et centrifugés à 4000 g pendant 20 minutes à 4°C. Après avoir éliminé les surnageants, les culots cellulaires ont été repris dans 1 mL de TRIzol puis immédiatement congelés et stockés à -80°C jusqu'à l'extraction des ARN. Le suivi de la culture est présenté sur le graphique de la **figure 30**.

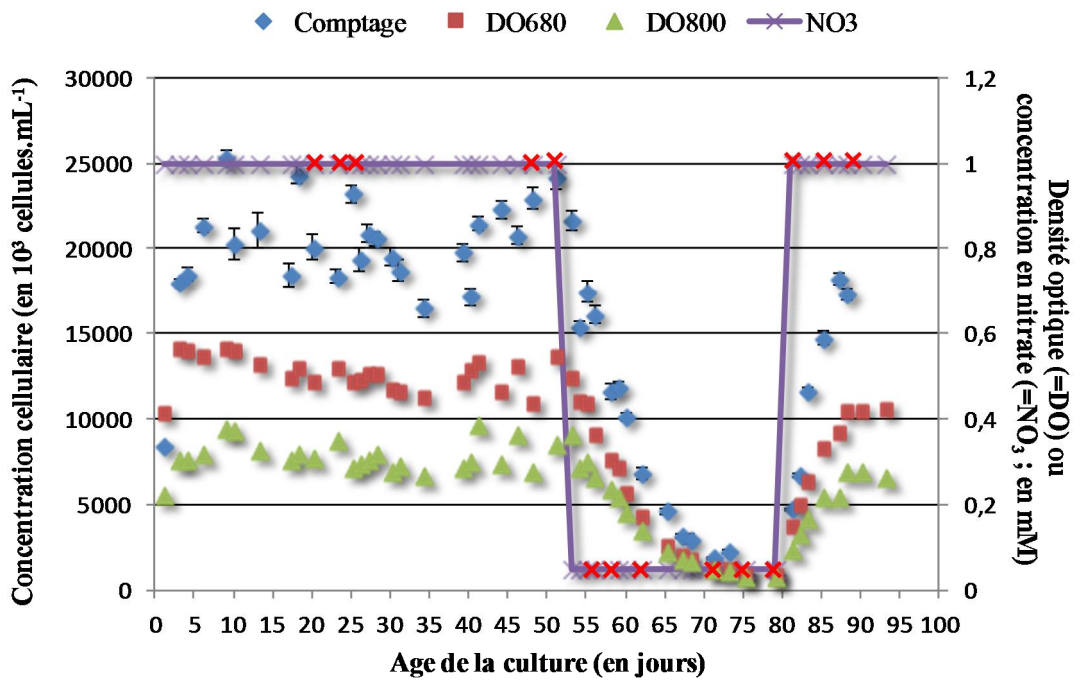


Figure 30 : Suivi de la culture continue de la souche Pt1.8.6 de *Phaeodactylum tricornutum* en fonction de la concentration en nitrate. La concentration cellulaire ainsi que la densité optique (DO) à 680 et 800 nm ont été mesurées au cours du temps pour une concentration en nitrate fixée. Les croix rouges indiquent les prélèvements effectués pour réaliser l'étude de l'expression du gène de la N-acétylglucosaminyltransférase I par PCR quantitative en temps réel.

II-2- Analyses bioinformatiques

II-2-1- Alignements multiples.

Les alignements multiples de séquences nucléotidiques et protéiques ont été réalisés sur l'intégralité des séquences en utilisant les programmes ClustalW du serveur EMBL-EBI (http://www.ebi.ac.uk/Tools/msa/clustalw2/ ; (Thompson *et al.*, 1994)) et MultAlin/ESPript version 2.2 (http://multalin.toulouse.inra.fr/multalin/ (Corpet, 1988) et http://espript.ibcp.fr/ESPript/ESPript/ ; (Gouet et al., 1999)). ESPript permet l'ajout de la structure secondaire de la protéine au-dessus de l'alignement protéique.

Le programme ClustalW calcule les scores de similitude et construit une matrice de distances pour l'ensemble des séquences alignées deux à deux. Le programme génère ensuite une séquence consensus à partir de l'alignement multiple des deux séquences les plus proches. Ce consensus est aligné avec la troisième séquence la plus proche. L'alignement se construit progressivement par l'établissement de consensus successifs jusqu'à sa forme multiple finale.

II-2-2- Analyses des séquences nucléotidiques.

La recherche de séquences nucléotidiques complémentaires et la traduction en séquences protéiques ont été réalisées avec DNA Tools (http://biology.semo.edu/cgi-bin/dnatools.pl). Les cartes de restriction des inserts et des plasmides ont été effectuées avec le programme pDRAW32 version 1.1.114.

II-2-3- Analyses des séquences protéiques.

La prédiction de peptide signal et de localisation subcellulaire des protéines a été obtenue en utilisant les programmes Signal-BLAST (Frank and Sippl, 2008), Target P (Emanuelsson et al., 2007) et Signal P3.0 (Bendtsen et al., 2004). Selon le type de données utilisées comme références (eucaryotes, bactéries Gram et Gram+), le serveur annonce, après un calcul algorithmique de type «Markov caché », les peptides signaux éventuels et les sites de clivages probables. Les peptides signaux sont de courts segments peptidiques positionnés à l'extrémité N-terminale des protéines destinées à être sécrétées ou exportées dans un compartiment intracellulaire. Ces signaux sont responsables de l'adressage des protéines aux translocons.

Les domaines transmembranaires ont été prédits avec les programmes TMHMM (Emanuelsson et al., 2007), HMMTOP (Tusnády and Simon, 2001) et TopPred (Claros and von Heijne, 1994). Le fonctionnement de ce dernier est décrit à titre d'exemple. Il utilise le fait qu'un indice d'hydrophobicité est attribué à chaque acide aminé d'après une échelle conventionnelle élaborée par (Kyte and Doolittle, 1982). Un profil d'hydrophobicité de la protéine est alors établit permettant de déterminer des zones apolaires définies par la succession d'acides aminés hydrophobes. Pour tracer le profil d'hydrophobicité d'une protéine, le score de chaque résidu est pris en compte. Le profil est ensuite lissé en calculant le score moyen sur une fenêtre de résidus de taille variable, glissant le long de la séquence. Une fenêtre de 19 à 21 résidus est préconisée pour la recherche de domaines transmembranaires. Si le score moyen dépasse 1,8 dans une zone, alors celle-ci est transmembranaire. Une hélice transmembranaire correspond à une zone transmembranaire de plus de 5 résidus consécutifs.

Les domaines Pfam ont été identifiés avec le programme Pfam sanger (http://pfam.sanger.ac.uk/ ; (Finn et al., 2010)). La base de données Pfam est une grande collection de familles de protéines, chacune étant représentée par des alignements multiples de séquences, des motifs particuliers et basée sur des modèles de Markov cachés.

II-2-4- Recherche d'homologies.

Afin d'identifier les gènes correspondant à une fonction donnée, plusieurs processus d'analyses bioinformatiques ont été menées. La première méthode utilisée consiste à comparer la séquence de gènes dont la fonction a été caractérisée avec le génome de l'organisme d'étude. La seconde méthode réside dans la recherche de motifs caractéristiques permettant ainsi d'identifier une fonction putative ou une appartenance à une famille de protéines.

II-2-4-1- Par comparaison de séquences.

La recherche de similitudes entre séquences dans les banques de données a été réalisée à l'aide d'un BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990). La valeur E calculée par le programme pour chaque alignement permet d'estimer si l'alignement observé a une réelle signification biologique. Plus la valeur E est proche de 0 et plus l'alignement et l'identification du gène sont statistiquement significatifs et fiables.

Les gènes de *P. tricornutum*, codant potentiellement pour des enzymes impliquées dans la voie de N-glycosylation ont été identifiés par des analyses de BLASTP et TBLASTN réalisée sur le site du génome (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) en utilisant comme référence des séquences génomiques et protéiques provenant de diverses espèces modèles telles que *Homo sapiens*, *Mus musculus, Arabidopsis thaliana, Drosophila melanogaster, Saccharomyces cerevisiae, Physcomitrella patens, Zea mays* et *Oryza sativa*. L'ensemble des séquences qui ont été utilisées sont recensées dans les **annexes 1** et **2**.

II-2-4-2- Par recherche de motifs.

La recherche de motifs conservés et de résidus particuliers nécessaires à la structure/fonction des protéines a également permis l'identification de ces gènes. Ces analyses ont été effectuées à partir d'alignements multiples de séquences protéiques en recherchant la présence de motifs ou résidus caractéristiques d'une famille de protéines décrits dans la littérature.

II-2-5- Vérification des séquences protéiques exprimées dans les cellules de mammifères.

Dans le but de vérifier que les séquences protéiques qui seront exprimées dans les cellules de mammifères correspondent bien à celles annotées sur le génome, une analyse bioinformatique a été réalisée. Pour cela, les séquences nucléotidiques des gènes putatifs obtenues sur le site du Joint Genome Institute ont été analysées avec le logiciel GENSCAN. Ce programme permet de réaliser des simulations d'épissage des ARNm chez les vertébrés et les plantes, puis de prédire la protéine correspondante. Les séquences protéiques générées ont ensuite été alignées avec les séquences annotées sur le génome à l'aide du programme MultAlin.

II-2-6- Analyse du séquençage.

Les séquences nucléotidiques obtenues à l'issue du séquençage ont été alignées à l'aide du programme ClustalW avec la séquence théorique, afin de vérifier d'éventuelles erreurs lors de l'amplification. La séquence nucléotidique consensus ainsi obtenue a ensuite été traduite en séquence protéique, puis alignée avec la séquence protéique théorique à l'aide des programmes MultAlin et Espript.

Le meilleur clone obtenu pour le plasmide contenant le gène de la GnT I de *P. tricornutum* contient des mutations au niveau protéique par substitutions de six acides aminés. L'influence de ces mutations sur la structure de la protéine a été analysée par le Dr Laurence Menu-Bouaouiche avec le logiciel Swiss-PdbViewer (Guex N. and Peitsch M.C., 1997). Il en résulte que les mutations observées ne semblent pas affecter la conformation de la protéine. Ce clone a donc été utilisé lors de la caractérisation fonctionnelle de la GnT I dans les cellules CHO Lec1.

II-3- Techniques de biologie moléculaire.

II-3-1- Analyses des acides nucléiques.

II-3-1-1- Analyse quantitative : dosage par spectrophotométrie.

La quantification des acides nucléiques a été réalisée par spectrophotométrie à la longueur d'onde de 260 nm. En effet, les bases puriques et pyrimidiques des acides nucléiques absorbent fortement dans l'ultraviolet à cette longueur d'onde. Il est admis qu'une unité d'absorbance à 260 nm (A260) correspond à une solution d'ARN à 40 $\mu g.mL^{-1}$ ou à une solution d'ADN double brin à 50 $\mu g.mL^{-1}$. Un spectre d'absorption a été réalisé entre 210 et 320 nm avec le NanoDrop 1000 (Thermo scientific) pour déceler la présence d'éventuels composés phénoliques et de protéines qui absorbent respectivement à 230 et 280 nm. L'estimation du degré de pureté de l'échantillon a été évalué par le calcul des rapports A260/A280 et A230/A280 dont les valeurs doivent être comprises entre 1,8 et 2,0 pour une qualité optimale.

II-3-1-2- Analyse qualitative : électrophorèse en gel d'agarose.

Une électrophorèse en gel d'agarose permet de séparer et d'identifier des fragments d'acides nucléiques selon leur taille. Cette technique peut également être utilisée pour évaluer la pureté et la qualité des échantillons. Les acides nucléiques, étant chargés négativement, vont migrer vers l'anode de la cuve d'électrophorèse avec une vitesse dépendant de leur taille et de la concentration en agarose présente dans le gel. Afin d'identifier des fragments de 0,2 kb à 5 kb, les gels utilisés sont composés de 1% d'agarose. Pour l'analyse des fragments de plus de 5 kb la concentration d'agarose est de 0,8%.

Les gels ont été préparés à partir de tampon Tris-Borate-EDTA 1X (TBE ; 80 mM Tris Base, 89 mM d'acide borique, 2 mM EDTA, pH 8,3) auquel a été ajouté de l'agarose à la concentration souhaitée ainsi que 0,1 $\mu g.mL^{-1}$ de bromure d'éthidium (BET). Ce dernier est un agent fluorescent qui s'intercale entre les bases azotées des acides nucléiques. Il fluoresce lorsqu'il est excité sous l'action d'un rayon ultraviolet. Une fois préparés, les gels ont été transférés dans une cuve d'électrophorèse (Mupid-ex, Advance) remplie de tampon de migration TBE 0,5X. Les échantillons d'acides nucléiques ont été additionnés de tampon de charge (composé de quelques gouttes de bleu de bromophénol 0,01 %, de 600 µL de glycérol 50 % et de 400 µL de TE 1X (10 mM Tris-Cl, pH 7,5 ; 1 mM EDTA)), puis déposés dans les puits du gel. En parallèle, 6 µL de marqueur de taille (Track it (Invitrogen), Smart Ladder

(Biorad) ou Smart ladder 0,2 kb-10 kb (Eurogentec)) ont également été chargés dans le gel, afin d'estimer la taille des fragments d'intérêt. La migration a ensuite été réalisée dans du TBE 0.5X sous une tension constante de 100 V pendant 30 minutes et la visualisation des acides nucléiques a été effectuée sous ultraviolets (à 302 nm) à l'aide d'un appareil Imager (Bio-capt, Vilber Lourmat).

II-3-2- Clonage et expression des gènes d'intérêt.

II-3-2-1- Extraction de l'ADN génomique.

Afin d'amplifier le gène de la FuT46110 putative, 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, les algues cultivées en erlermeyer ont été récoltées par centrifugation à 4000 g pendant 3 minutes. Le culot cellulaire a alors été repris dans 400 µL de solution de lyse puis déposé dans un « lysing matrix tube E » (MP Biomedicals) contenant des billes en céramique de 1,4 mm, des billes en silice de 0,1 mm et une bille en verre de 4 mm. 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 $m.s^{-1}$. L'échantillon a ensuite été centrifugé à 1200 g pendant 15 minutes pour culotter les billes et les débris cellulaires. Le surnageant a été récupéré et la suite de l'extraction a été effectuée avec les colonnes du kit en suivant les instructions du fournisseur Macherey-Nagel.

L'ADN génomique utilisé pour l'amplification du gène de la GnT I a été extrait en utilisant le même protocole que celui de (Falciatore et al., 1999).

II-3-2-2- Amplification par Polymerase Chain Reaction (PCR).

La PCR, mise au point par (Mullis et al., 1986), est une technique qui permet de synthétiser et d'amplifier de l'ADN *in vitro*. Elle est basée sur des cycles d'amplification, composés chacun de trois étapes :

- une étape de dénaturation qui permet de séparer les doubles brins d'ADN matrice et de supprimer leur structure secondaire.

- 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 enfin une 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.

L'ADN polymérase utilisée pour réaliser les constructions est la « Extensor HI-fidelity PCR Enzyme Mix » (Thermo Scientific, AB-0720). Cette enzyme est une polymérase de haute fidélité qui limite l'apparition de mutations au cours de la réaction. Elle permet également l'ajout d'un résidu adénosine à chaque extrémité 3' du produit PCR aboutissant à la formation d'extrémités dépassantes nécessaires au clonage. Les couples d'amorces spécifiques GnT I-5F / GnT I-3R, FuT46109-5F / FuT46109-3R, FuT46110-5F / FuT46110-3R et FuT54599-5F / FuT54599-3R ont respectivement été utilisés pour réaliser l'amplification des gènes de la GnT I, FuT46109, FuT46110 et FuT54599 (voir **tableau 1** paragraphe I-4). Le mélange réactionnel de PCR a été préparé comme indiqué dans le **tableau 2** dans des microtubes de 0,2 mL puis homogénéisé délicatement avant d'être placé dans un thermocycleur (Labcycler (Sensoquest) ou Mastercycleur personal (Eppendorf)). Les conditions de PCR utilisées sont résumées dans le **tableau 3**. Pour le témoin négatif, le volume d'ADN a été remplacé par un volume équivalent d'eau stérile.

Solutions	Volume
Tampon 1 (10X)	5 µL
Hi-fidelity PCR Enzyme mix (5 $U.\mu L^{-1}$)	0,5 µL
dNTPs (20 mM , soit 5 mM de chaque dNTP)	5 µL
ADN	1 µL
Amorce 1 (10µM)	1 µL
Amorce 2 (10µM)	1 µL
Eau stérile pour un volume final de 50 µL	

Tableau 3 : Composition du mélange réactionnel utilisé pour l'amplification PCR des gènes codant la GnT I et les FuT46109, FuT46110 et FuT54599 avec la polymérase Extensor HI-fidelity PCR Enzyme Mix (Thermo Scientific).

Etape	Température	Durée	Nombre de cycle
Dénaturation initiale	94°C	2 min	1
Dénaturation Hybridation Extension	94°C 60°C / 50°C / 56°C 68°C	30 s 30 s 90 s / 2 min / 90 s	30 / 35 / 30
Extension finale	68°C	7 min	1

Tableau 2 : Conditions de PCR utilisées pour l'amplification des gènes codant la GnT I et les FuT46109, FuT46110 et FuT54599 avec la polymérase Extensor HI-fidelity PCR Enzyme Mix (Thermo Scientific).

II-3-2-3- Purification du produit PCR.

La purification des produits PCR a été réalisée directement en solution avec le kit « QIAquick PCR Purification kit » (QIAGEN, 28104) ou à partir d'un gel d'agarose en utilisant le kit « Wizard SV Gel and PCR Clean-Up System » (Promega, A9282). Dans les deux cas, l'utilisation de ces kits a été effectuée en suivant les instructions des fabricants respectifs.

II-3-2-4- Ligation et transformation bactérienne.

La ligation du produit PCR dans le vecteur linéaire pcDNA 3.1/V5-His-TOPO® et sa transformation dans des cellules thermocompétentes « one shot top 10 chemically competent *E-coli* » ont été réalisées avec le kit « pcDNA 3.1/V5-His-TOPO TA expression kit » (Invitrogen, K4800-01) en suivant le protocole du fournisseur. Le rapport de ligation utilisé est de 4 µL de produit PCR pour 1 µL de vecteur. Après transformation, 150 µL de solution bactérienne ont été ensemencés sur une boîte de pétri contenant du milieu LB gélosé (Sigma) auquel ont été ajoutés 50 $\mu g.mL^{-1}$ d'ampicilline (Duchefa). La boîte a ensuite été incubée à 37°C pendant une nuit.

II-3-2-5- Extraction d'ADN plasmidique.

Plusieurs colonies isolées obtenues suite à la transformation bactérienne ont été sélectionnées, puis repiquées dans 5 mL de milieu de culture LB liquide (Sigma) contenant 50 $\mu g.mL^{-1}$ d'ampicilline et placées à 37°C pendant la nuit sous agitation. L'extraction et la purification d'ADN plasmidique a été réalisée à partir de 3 mL de préculture avec les kits « PureYield™ Plasmid Midiprep System » (Promega, A1222) et QIAGEN® Plasmid Mini kit (QIAGEN, 12125), selon les indications fournies par les fabricants. Le principe de ces kits repose sur la méthode de lyse alcaline des cellules initialement proposée par (Birnboim and Doly, 1979) suivie d'une purification de l'ADN plasmidique sur colonne de silice. L'ADN retenu sur la colonne a ensuite été élué avec de l'eau stérile.

II-3-2-6- Contrôle des clones sélectionnés.

II-3-2-6-1- Sélection des clones positifs par digestion enzymatique.

Une digestion enzymatique a été réalisée sur les mini préparations d'ADN plasmidique, afin de sélectionner les clones positifs contenant le gène d'intérêt orienté dans le bon sens. Pour cela, les plasmides pcDNA3.1/V5-His + GnT I et pcDNA3.1/V5-His + FuT54599 ont été digérés respectivement avec les enzymes de restriction Xho I (Promega, R6161) et Kpn I (Promega, R6341). La digestion contrôle des plasmides pcDNA3.1/V5-His + FuT46109 et pcDNA3.1/V5-His + FuT46110 a quant à elle été effectuée avec l'enzyme Hind III (Promega, R6041). La composition du mélange réactionnel utilisé pour ces digestions est indiquée dans le **tableau 4**. Une fois le mélange réalisé, les échantillons ont été délicatement homogénéisés, puis incubés au bain-marie pendant 4 heures à 37°C. Le résultat des digestions enzymatiques a ensuite été visualisé par migration électrophorétique en gel d'agarose (cf paragraphe II-3-1-2).

Solutions	Volume
ADN plasmidique	0,2 à 1,5 µg
Tampon D, E ou J (10X)	2 µL
BSA acétylée (10 $\mu g.\mu L^{-1}$)	0,2 µL
Enzyme de restriction XhoI, Hind III ou Kpn I (10 $U.\mu L^{-1}$)	0,5 µL
Eau stérile pour un volume final de 20 µL	

Tableau 4 : Composition du mélange réactionnel utilisé pour les digestions avec les enzymes de restriction Xho I, Hind III et Kpn I (Promega).

II-3-2-6-2- Séquençage.

La préparation des échantillons pour le séquençage du plasmide pcDNA3.1/V5-His + GnT I a été réalisée avec le kit « Dye Terminator Cycle Sequencing (DTCS) Quick Start Kit » (BECKMAN COULTER), en suivant les instructions du fabriquant. Cette préparation comporte une étape d'amplification par PCR de la partie à séquencer nécessitant l'utilisation d'amorces spécifiques (voir **tableau 1** paragraphe I-4), afin de générer des fragments chevauchants. L'appareil utilisé pour cette amplification est le thermocycleur Mastercycleur gradient d'Eppendorf et les conditions de PCR sont celles indiquées par le fabriquant du kit. Une fois les échantillons préparés, ils ont été déposés sur une plaque contenant 96 puits, puis recouverts d'une goutte d'huile minérale, de façon à limiter l'évaporation pendant la réaction. La plaque a ensuite été positionnée dans le séquenceur « Ceq 8000 Genetic analysis system » (Beckman Coulter). Le logiciel de séquençage utilisé est « Ceq 8000 System ».

Les clones sélectionnés des vecteurs recombinants pcDNA3.1/V5-His+FuT46109, pcDNA3.1/V5-His+FuT46110 et pcDNA3.1/V5-His+FuT54599 ont été envoyés à séquencer à l'entreprise GATC Biotech (www.gatc-biotech.com) afin de vérifier l'absence de mutation. Seule la séquence nucléotidique amplifiée par les amorces T7 et BGH a été séquencée.

II-3-2-7- 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, 30 µg de vecteur recombinant purifié ont été digérés par 7,5 µL d'enzyme de restriction Sca I (10 $U.\mu L^{-1}$, Promega, R6211) en présence de 0,75 µL de BSA acétylée (10 $\mu g.\mu L^{-1}$), 7,5 µL de tampon K 10X, et dans un volume réactionnel de 75 µL, à 37°C pendant 4 heures. La linéarisation des vecteurs a été vérifiée par migration du produit de digestion sur gel d'agarose. Les vecteurs linéarisés ont ensuite été purifiés sur gel avec le kit « Wizard SV Gel and PCR Clean-Up System » (Promega, A9282) en suivant les consignes du fabriquant.

II-3-2-8- Transfection stable des cellules CHO.

La transfection stable des cellules CHO Lec1 et CHO FuT8-/- a respectivement été réalisée par électroporation et par lipofection selon le principe présenté dans la **figure 31**.

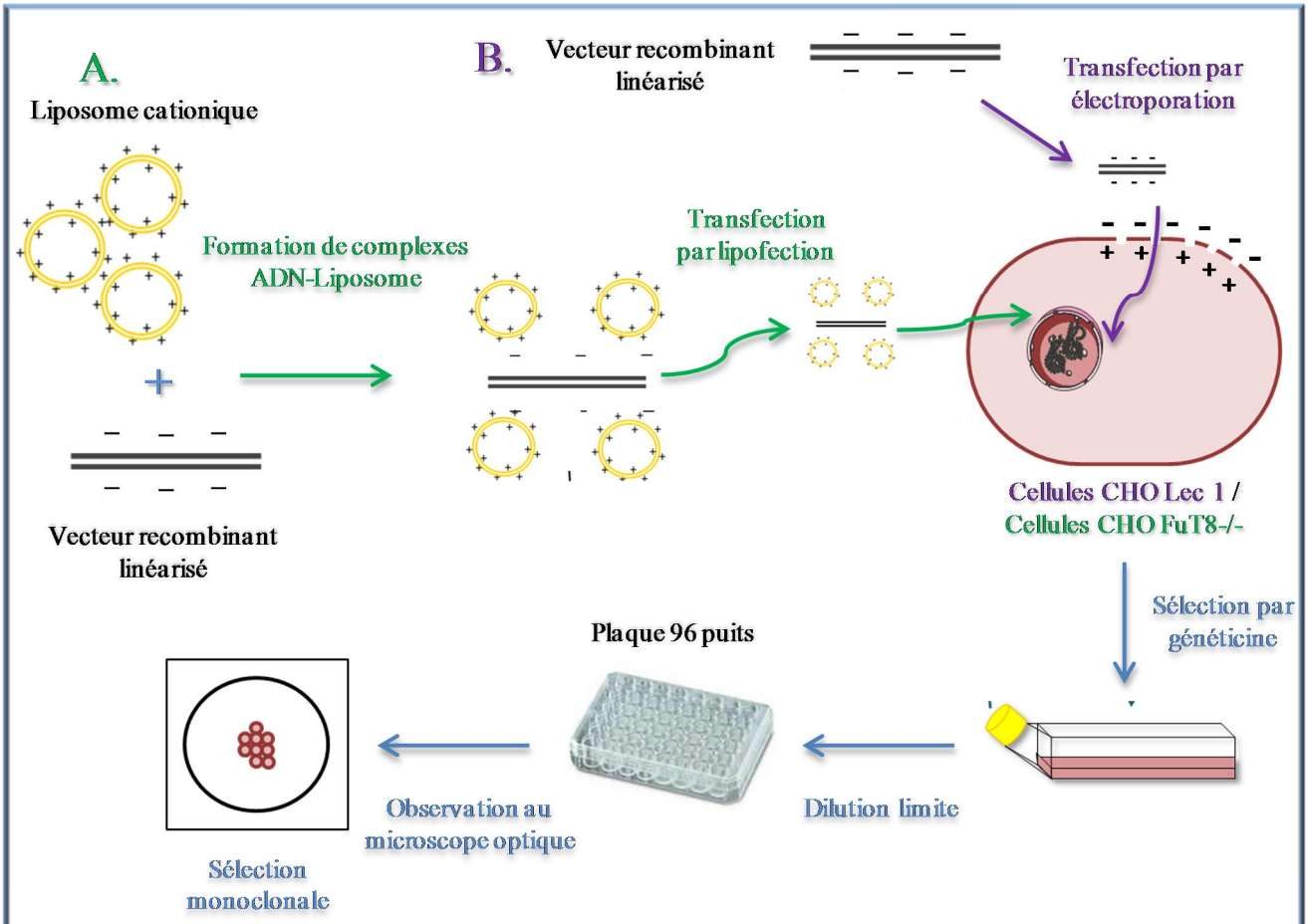


Figure 31 : Principe de la transfection stable par lipofectamine (A) ou électroporation (B). Les cellules CHO mutantes sont transfectées avec le vecteur recombinant soit par électroporation soit par lipofection sous forme de complexes avec des liposomes cationiques. Les transformants ont été sélectionnés avec la généticine. Une sélection monoclonale a ensuite été réalisée en utilisant la méthode de dilution limite associée à une observation des clones au microscope optique en lumière transmise.

II-3-2-8-1- Par électroporation.

Cette expérience a été réalisée avec le kit T (VCA-1002) sur approximativement 2.10^6 cellules CHO Lec1 récupérées par digestion trypsique (0.05%), puis culotées par centrifugation à 100 g pendant 5 minutes. Les cellules ont ensuite été reprises dans 100 µL de tampon de transfection fourni dans le kit. Cinq µg d'ADN plasmidique pcDNA3.1/V5-His + GnT I linéarisé ont alors été ajoutés ainsi que 500 µg d'ADN de sperme de saumon stérile. Les cellules CHO Lec1 ont alors été transfectées par électroporation en utilisant l'appareil Nucleofector d'Amaxa, puis déposées dans des flasques T75 contenant 12 mL de milieu frais d'αMEM (Invitrogen) et 10% de SVF (Invitrogen) et placées dans un incubateur à 37°C contenant 5% de CO_2. Au bout de 2 jours, le milieu de culture a été substitué par 12 mL de milieu frais contenant 400 $\mu g.mL^{-1}$ de généticine (Invitrogen).

II-3-2-8-2- Par lipofection.

La lipofection est une technique qui permet de former un complexe cationique autour de l'ADN lui permettant ainsi d'entrer dans les cellules. Afin de réaliser la transfection des cellules CHO Fut8-/- avec les plasmides pcDNA3.1/V5-His + FuT46109, pcDNA3.1/V5-His + FuT46110 et pcDNA3.1/V5-His + FuT54599, une première solution contenant 8 µg d'ADN plasmidique et 300 µL de milieu Opti-MEM (Invitrogen) a été préparée pour chaque vecteur recombinant. Une seconde solution contenant 32 µL de lipofectamine 2000 (Invitrogen) et 300 µL d'Opti-MEM a également été réalisée. Les deux solutions ont été mélangées et incubées pendant 30 minutes à température ambiante. Le mélange ADN/agent transfectant a alors été déposé dans une flasque T25 contenant les cellules CHO FuT8-/- cultivées depuis 24 heures et ayant atteint environ 50 % de confluence. Les flasques ont ensuite été placées dans un incubateur à 37°C contenant 5 % de CO_2. Après 24 heures d'incubation, le milieu de culture a été remplacé par du milieu frais contenant 400 $\mu g.mL^{-1}$ de généticine (Invitrogen).

II-3-2-8-3- Sélection monoclonale.

Afin d'obtenir une lignée transfectée stable pour chaque construction, une sélection monoclonale a été réalisée dans des plaques de 96 puits en utilisant la technique de dilution limite. Pour cela, les cultures ont été diluées dans du milieu frais contenant 400 $\mu g.mL^{-1}$ de généticine, jusqu'à l'obtention d'une seule cellule par puits. Les plaques ont ensuite été observées quotidiennement au microscope optique en lumière transmise afin de sélectionner les puits contenant des cellules issues d'une lignée monoclonale.

II-3-3- Etude de l'expression du gène de la GnT I.

II-3-3-1- Extraction des ARN totaux.

Afin de réaliser l'extraction des ARN totaux, les microalgues reprises dans du TRIzol (Invitrogen) ont été soniquées 8 fois 40 secondes dans la glace avec une puissance de 5 Watts à l'aide d'un sonicateur microtige. L'efficacité de la lyse cellulaire a été contrôlée en visualisant les cellules au microscope optique en lumière transmise à un grossissement x400. Deux cents microlitres de chloroforme ont alors été ajoutés au lysat, puis l'échantillon a été agité par retournement et incubé 5 minutes à température ambiante. Les échantillons ont alors été centrifugés à 12 000 g pendant 15 minutes à 4°C et l'épiphase aqueuse (~ 450 µL) contenant les ARN totaux a été récupérée.

II-3-3-2- Purification des ARN.

La purification des ARN est nécessaire pour éliminer notamment les traces résiduelles d'ADN génomique, de protéines et de composés phénoliques. Cette purification a été réalisée en trois étapes. Les extraits ont tout d'abord été déposés et purifiés sur une colonne en silice *RNeasy spin* du kit « RNeasy Mini Kit » (QIAGEN, 74104) en suivant les instructions du fournisseur. Afin d'éliminer l'ADN génomique restant, les échantillons ont ensuite été digérés avec la DNase RQ1 (Promega, M6101) dans un volume réactionnel de 100 µL contenant 30 µL d'éluat, 10 µL de tampon RQ1 RNase-free 10X et 1 $U.\mu g^{-1}$ d'ARN de DNase RQ1. La digestion a été réalisée à 37°C au bain marie pendant 40 minutes, puis arrêtée par ajout de 10 µL de solution RQ1 DNase stop suivi d'une incubation à 65°C pendant 10 minutes. Les ARN ont ensuite été purifiés par un second passage sur colonne *RNeasy spin*, puis finalement élués avec de l'eau stérile RNase-free.

II-3-3-3- Reverse transcription PCR (RT-PCR).

II-3-3-3-1- Transcription inverse.

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

Dans ces travaux, les ARNm ont été rétro-transcrits en ADNc en utilisant la transcriptase inverse M-MLV RT (Promega, M170A). Pour cela, 1 µg d'ARN est dénaturé durant 5 minutes à 70°C en présence d'1 µg de random primer (Promega), puis immédiatement placé sur de la glace. Le mélange réactionnel présenté dans le **tableau 5** a alors été ajouté aux 10 µL précédents et les tubes ont été placés dans un thermocycleur (Mastercycleur personal, Eppendorf) pendant 1 heure à 37°C. Les échantillons ont été conservés à –20 °C.

Solutions	Volume
Tampon M-MLV 5X	5 µl
dATP, 10 mM	1,25 µl
dCTP, 10 mM	1,25 µl
dGTP, 10 mM	1,25 µl
dTTP, 10 mM	1,25 µl
M-MLV Reverse Transcriptase	1 µl
Eau RNase-Free pour un volume final de 25 µl	

Tableau 5 : Composition du mélange réactionnel utilisé pour la transcription inverse avec la M-MLV RT de Promega.

II-3-3-3-2- Amplification par PCR.

L'amplification par PCR des ADNc de la GnT I a été réalisée en utilisant l'ADN polymérase GoTaq (Promega) ainsi que le couple d'amorces spécifiques GnT I 5F / GnT I AS12. Le mélange réactionnel a été préparé comme indiqué dans le **tableau 6**, dans des microtubes, puis homogénéisé délicatement avant d'être placé dans un Thermocycleur (Mastercycleur personal, Eppendorf). Les conditions de PCR utilisées sont indiquées dans le **tableau 7**. Différents témoins ont été utilisés pour cette étude. En effet, le gène de référence de l'actine 1 a été amplifié en parallèle, avec le couple d'amorces spécifiques Act F / Act R, afin de contrôler la présence d'acides nucléiques dans les échantillons. De même, la réaction a

été directement réalisée sur les ARN, afin de vérifier l'absence de contamination par de l'ADN génomique résiduel.

Solutions	ADNc	ARN
Tampon 5X green	10 µl	10 µl
dNTP 10mM chacun	1 µl	1 µl
Amorce 1 (10µM)	5 µl	5 µl
Amorce 2 (10µM)	5 µl	5 µl
GoTaq	0,25 µl	0,25 µl
ADNc ou ARN (1 µg / 25 µl)	5 µl	5 µl
Eau	23,75 µl	23,75 µl
Volume total	50 µl	50 µl

Tableau 7 : Composition du mélange réactionnel utilisé pour l'amplification PCR des ADNc et des ARN avec la GoTaq polymérase (Promega).

Etape	Température	Durée	
Dénaturation initiale	94°C	3 min	
Dénaturation Hybridation Extension	94°C 58°C 72°C	1 min 45 sec 1 min	x 39
Extension finale	72°C	7 min	

Tableau 6 : Conditions de PCR utilisées pour l'amplification des ADNc avec la GoTaq polymérase (Promega).

II-3-3-4- PCR quantitative en temps réel.

II-3-3-4-1- Transcription inverse.

Les ADN complémentaires ont été synthétisés à partir de 350 ng d'ARN totaux avec le kit « High Capacity cDNA Reverse Transcription Kit » (Applied Biosystems) en utilisant le protocole conseillé par le fournisseur.

II-3-3-4-2- Amplification par PCR.

La PCR quantitative a été réalisée avec le réactif « *Power* SYBR® Green I PCR Master Mix » (Applied Biosystems) en suivant les instructions du fabricant. Les réactions ont été effectuées dans un volume réactionnel total de 25 µL contenant 5 µL d'ADNc dilué au 10ème et 0,1 µM d'amorces spécifiques. L'amplification du gène de la GnT I de *P. tricornutum* a été réalisée avec les amorces GNT1-Q-Fwd et GNT1-Q-Rev spécifiques de la séquence codant le domaine catalytique (voir **tableau 1** paragraphe I-4). Les mesures ont été effectuées en

duplicata avec l'appareil « Stratagene Mx3000PÊ Q-PCR system ». La PCR a été réalisée en utilisant le programme suivant : un cycle à 95°C pendant 10 minutes suivi de 40 cycles contenant une étape de 30 secondes à 95°C et une étape d'1 minute à 60°C. Les gènes de référence utilisés dans cette étude sont les gènes codant pour la Ribosomal Protein Small subunit 30 S (RPS) et pour l'histone H4 en accord avec les données de la littérature (Siaut et al., 2007).

II-3-4- Analyse des données de Q-PCR.

L'analyse de l'expression relative du gène de la GnT I a été réalisée en utilisant la méthode décrite par (Kang et al., 2009) et (Chung et al., 2005). Dans cette approche, la méthode comparative des cycles seuil C_T (Livak and Schmittgen, 2001) et la méthode de courbe standard ont été combinées pour calculer le ratio molaire des ARN entre le gène cible (X) et le gène de référence (R). La formule utilisée est présentée ci-dessous :

$$\log\left(\frac{X_0}{R_0}\right) = \log\left(\frac{M_R}{M_X}\right) + \frac{C_{T,X}}{b_X} - \frac{C_{T,R}}{b_R}$$

Où : - M_X et M_R sont respectivement les poids moléculaires des amplicons du gène cible (X_0) et du gène de référence (R_0).

- $C_{T,X}$ et $C_{T,R}$ sont respectivement les C_T du gène cible et du gène de référence.
- b_X et b_R sont respectivement les pentes des courbes standard du gène cible et du gène de référence.

Afin de déterminer si l'expression de la GnT I varie en fonction des conditions en nitrate, des tests statistiques par analyse de la variance à un facteur (ANOVA) ont été réalisés sur les ratios molaires avec le logiciel Excel.

II-4- Techniques de biochimie.

II-4-1- Préparation des extraits protéiques.

II-4-1-1- Extraction des protéines de *P. tricornutum.*

Deux grammes de cellules lyophilisées de *P. tricornutum* ont été repris dans un tampon d'extraction pH 8 contenant 750 mM de Tris-HCl, 15% (m/v) de sucrose, 2% (v/v) de β-mercaptoéthanol et 1 mM de phénylméthylsulfonylfluoride. L'échantillon a été transféré dans un tube « Lysing Matrix D » de 50 mL (MP Biomedicals ; 116953010) contenant des

billes en céramique de 1,4 mm de diamètre. Le tube a ensuite été placé dans un agitateur Fastprep 24 (MP Biomedicals ; 116004500) où les cellules ont été lysées par 15 cycles de 30 secondes avec une vitesse de rotation de 6,5 $m.s^{-1}$. Le lysat cellulaire a été centrifugé à 4 000 g pendant 5 minutes à 4°C. Le culot a ensuite été lavé une fois avec 10 mL de tampon d'extraction et centrifugé à nouveau à 4 000 g pendant 5 minutes à 4°C. Les surnageants résultants ont été regroupés puis centrifugés à 15 000 g pendant 30 minutes à 4°C. Le surnageant obtenu à l'issu de cette étape a ensuite été dialysé dans l'eau pendant 48 heures à 4°C avant d'être lyophilisé.

Un extrait de protéines d'oignon vert a été préparé en parallèle des précédents échantillons et utilisé comme témoin positif lors des analyses d'affino- et immuno-détection.

II-4-1-2- Extraction des protéines des cellules CHO.

L'extraction protéique des différentes lignées cellulaires de CHO a été réalisée à partir de deux flasques T150 (Dutscher, 190151). Pour cela, les cellules ont été récoltées à l'aide d'un râteau stérile dans 5 ml de solution DPBS 1X (*Dulbecco*`s *Phosphate Buffered Saline* ; Fisher, E3468W), puis centrifugées à 100 g pendant 3 minutes à 4°C. Les culots cellulaires ont été repris dans 1 mL de tampon d'extraction (Tris-HCl 100 mM, SDS 0,1%, pH7,5), et soniqués 4 fois 20 secondes dans la glace à une amplitude de 7 Watts avec un sonicateur microtige (Microson, ultrasonic cell disruptor XL). Les lysats cellulaires ont ensuite été centrifugés à 100 g pendant 10 minutes à 4°C et les surnageants contenant les protéines ont été stockées à -20°C avant analyse.

II-4-1-3- Dosage protéique.

La quantification des protéines a été réalisée avec le kit de dosage protéique « BCA™ Protein Assay Kit » commercialisé par la société Pierce (Thermo Scientific, 23227), en suivant le protocole indiqué par le fabriquant. La protéine BSA (bovine serum albumin) a été utilisée comme protéine standard pour l'établissement de la gamme étalon. Les échantillons ont été dosés au 10ème. Deux millilitres de réactif (mélange des solutions A et B dans un ratio 50 : 1) ont été ajoutés à 100 µL de chaque solution à doser. Les échantillons ont ensuite été placés 30 minutes à 37°C avant de lire leur absorbance à 562 nm au spectrophotomètre.

II-4-2- Déglycosylation.

Dès le début des années 70, des enzymes capables d'éliminer entièrement ou partiellement la partie glycannique des glycoprotéines ont été mises en évidence et caractérisées à partir de bactéries, de champignons, de plantes et d'animaux (Berger et al., 1995). Ces enzymes capables de dé-N-glycosylation sont les peptide-N^4-(*N*-acétyl-β-D-glucosaminyl) asparagine amidases (PNGases) et les endo-*N*-acetyl-β-D-glucosaminidases (ENGases). Les PNGases clivent la liaison glycosylamine entre la GlcNAc proximale et l'Asn de la chaîne protéique. Les ENGases agissent au niveau de la liaison glycosidique entre les deux premières GlcNAc du core glycannique.

II-4-2-1- Digestion par l'Endoglycosidase H.

L'Endoglycosidase H (Endo H) est une ENGase isolée à partir de *Streptomyces griseus* ou *Streptomyces plicatus* (Tarentino and Maley, 1974 ; Robbins et al., 1984) clivant exclusivement des structures oligomannosidiques (**figure 32**). Afin de réaliser cette digestion, 0,5 mg de protéines extraites ont été repris dans une solution de SDS 1 % et dénaturées à 100°C pendant 5 minutes. L'échantillon a alors été dilué cinq fois dans 500 µL de tampon acétate de sodium 150 mM, pH 5,7 et incubé toute la nuit à 37°C avec 10 mU d'Endo H (Roche, 11088726001). Les protéines ainsi digérées ont été précipitées dans 4 volumes d'éthanol sur la nuit à -20°C, afin d'être analysées par affino-détection.

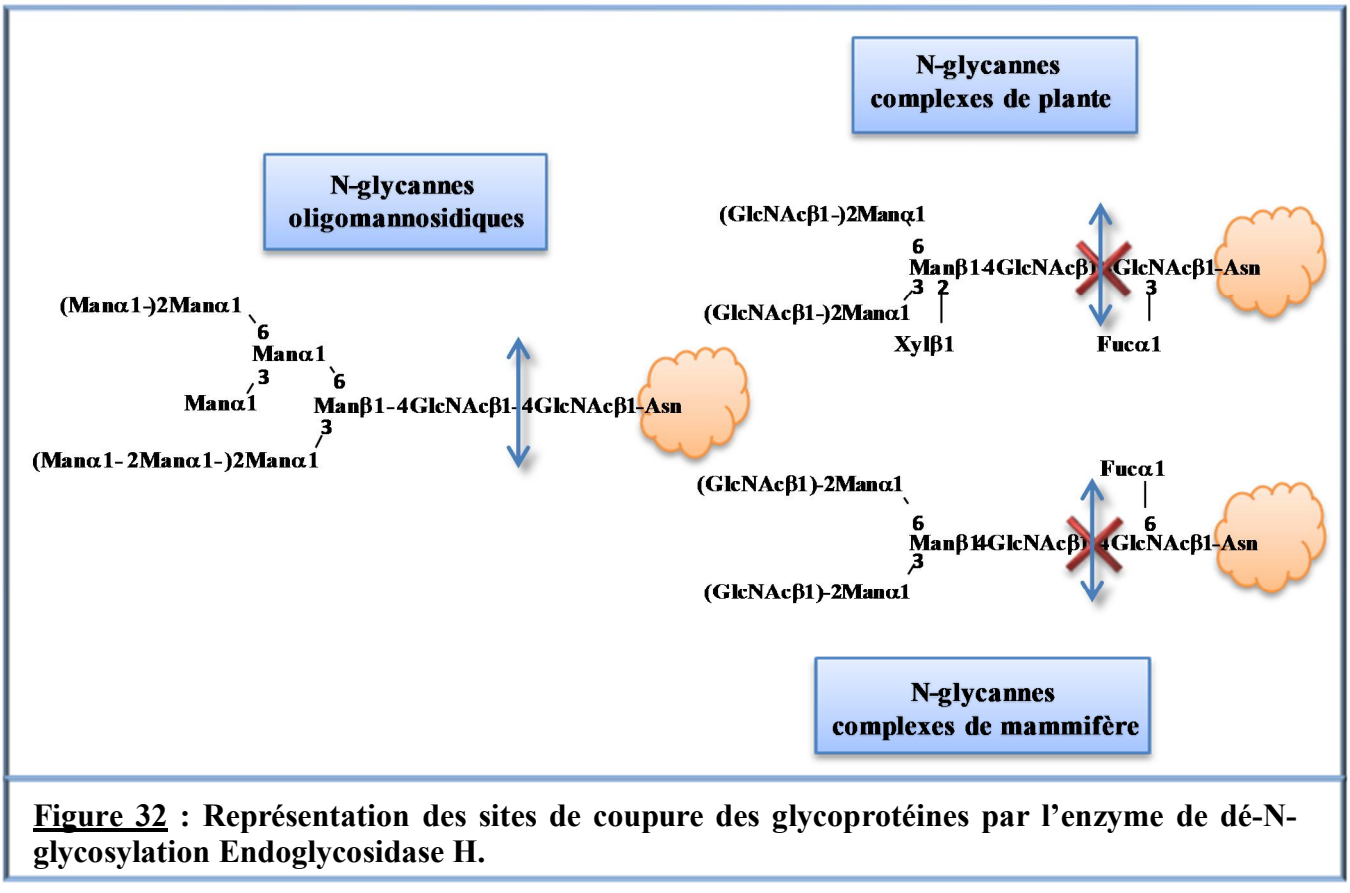


Figure 32 : Représentation des sites de coupure des glycoprotéines par l'enzyme de dé-N-glycosylation Endoglycosidase H.

II-4-2-2- Digestion par la peptide N-glycosidase F.

La peptide N-glycosidase F (PNGase F) a été isolée à partir de la souche bactérienne *Flavobacterium meningosepticum* (Plummer et al., 1984). Cette enzyme, couramment utilisée pour déglycosyler les glycoprotéines de mammifères (Tarentino et al., 1985), permet de cliver tout types de N-glycannes, à l'exception des N-glycannes complexes portant un résidu fucose lié en α(1,3) sur la GlcNAc proximale (Tretter et al., 1991 ; **figure 33**).

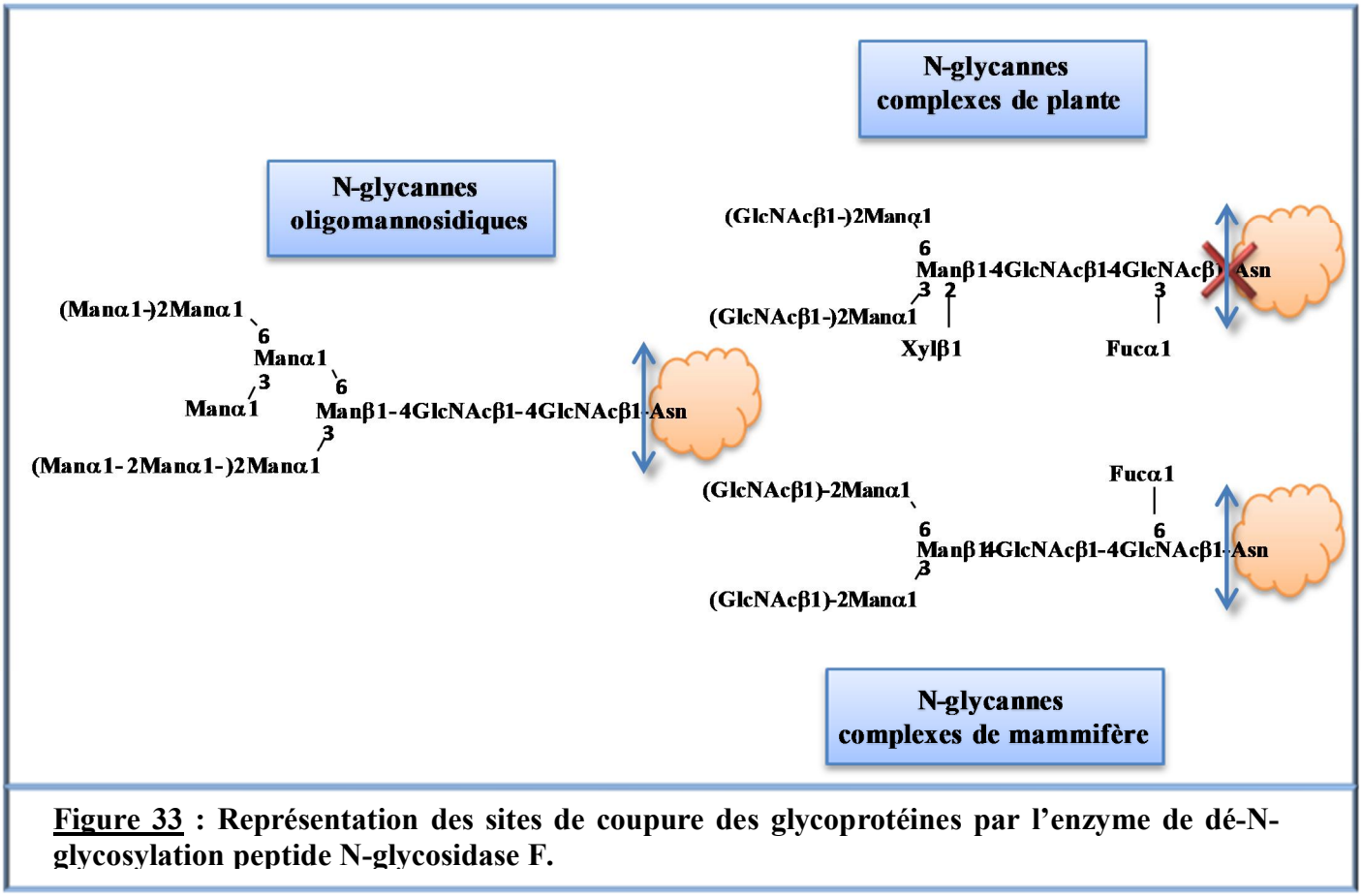


Figure 33 : Représentation des sites de coupure des glycoprotéines par l'enzyme de dé-N-glycosylation peptide N-glycosidase F.

Bien que la PNGase F soit une enzyme active sur les glycoprotéines natives, son action est favorisée par une dénaturation préalable des protéines (D'Andrea et al., 1993). Pour cela, les protéines reprises dans un tampon Tris-HCl 100 mM pH 7,5 contenant 0,1% de SDS ont été dénaturées à 100°C pendant 10 minutes. Après refroidissement, un volume équivalent de tampon Tris-HCl 100 mM pH 7,5 contenant 0,5% de Nonidet P40 (Igepal CA-630, SIGMA) a été ajouté, ainsi que 10 U de PNGase F (Roche, 11365185001). Les échantillons ont ensuite été placés dans un incubateur agitant à 37°C pendant 24 heures. Les protéines déglycosylées de *P. tricornutum*, utilisées pour les analyses d'affinodétection, ont été précipitées dans 4 volumes d'éthanol sur la nuit à -20°C.

II-4-2-3- Digestion par la peptide N-glycosidase A.

La peptide N-glycosidase A (PNGase A), isolée à partir d'amande douce (Takahashi et al., 1982), permet de cliver tout type de N-glycannes, y compris les structures complexes de plante comportant un fucose lié en α(1,3) à la GlcNAc proximale (Tretter et al., 1991). L'action de cette enzyme est plus efficace sur des glycopeptides, nécessitant ainsi une digestion préalable des protéines par une endoprotéase (**figure 34**).

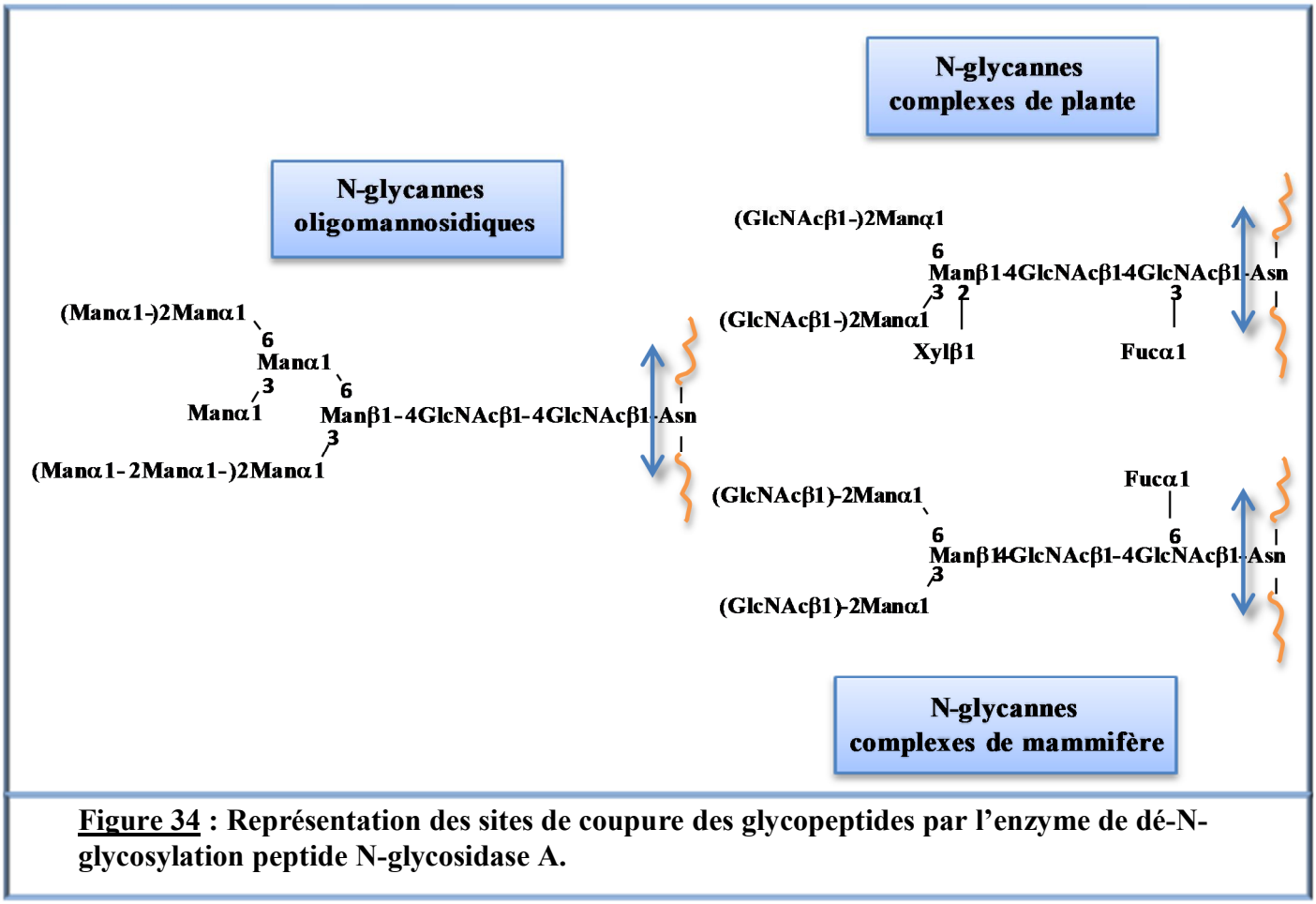


Figure 34 : Représentation des sites de coupure des glycopeptides par l'enzyme de dé-N-glycosylation peptide N-glycosidase A.

La déglycosylation par la PNGase A a été réalisée sur 5 mg de protéines lyophilisées reprises dans 3 mL d'une solution 4 M Tris-Guanidine HCl, pH 8,5, avant d'être dénaturées avec 2 $mg.mL^{-1}$ de dithiothreitol (DTT). Après une courte sonication de 30 secondes, l'échantillon a été incubé à 50°C pendant 2 heures. Un millilitre et demi d'une solution de 12 $mg.mL^{-1}$ d'iodoacétamide préparée dans un tampon Tris 0,6 M, pH 8,5, a été ajouté à l'échantillon, puis incubé dans le noir à température ambiante pendant 2 heures. L'échantillon a alors été dialysé pendant 72 heures contre de l'eau et les protéines ont été digérées à 37°C pendant 48 heures avec 10 mg de pepsine dissoute dans 2 mL d'une solution 10 mM HCl, pH

2,2. Après neutralisation de l'échantillon avec une solution 1 M d'hydroxyde d'ammonium, les peptides ont été dénaturés à 100°C pendant 5 minutes, puis lyophilisés. Les glycopeptides ont alors été repris dans un tampon 50 mM d'acétate de sodium, pH 5,5, et déglycosylés sur la nuit à 37°C sous agitation avec 1,5 mU de PNGase A (Roche, 11642995001).

II-4-3- Immuno- et affino-blot.

II-4-3-1- Séparation des protéines par gel d'électrophorèse SDS-PAGE.

La séparation électrophorétique des protéines a été effectuée selon la méthode de (Laemmli, 1970). Cette technique permet de séparer les protéines, ou leurs sous-unités constitutives, 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 8**.

Solutions	Gel de concentration	Gel de séparation
H_2O déionisée	0,73 mL	2,15 mL
Acrylamide/Bis-acrylamide (29 : 1) 40%	0,17 mL	1,5 mL
Tris-HCl 1,5 M pH8.8	-	1,25 mL
Tris-HCl 1,5 M pH6,8	0,08 mL	-
Dodécylsulfate de sodium (SDS ; 10 %)	10 µL	50 µL
Persulfate d'ammonium (10 %)	10 µL	50 µL
TEMED	1 µL	2 µL

Tableau 8 : Composition des gels de concentration et de séparation SDS-PAGE 12%.

Pour chaque échantillon à analyser, 50 µg de protéines ont été dénaturés pendant 5 minutes à 100°C dans le tampon de charge (Tris-HCl 62,5 mM, pH 6,8 ; SDS 2 %, glycérol 10 %, β-mercaptoéthanol 2,5 %, bleu de bromophénol 0,005 %), avant d'être déposés dans les puits du gel de polyacrylamide. Cinq microlitres de marqueur de taille « BioReagents EZ-Run Prestained *Rec* Protein Ladder » (Fisher, BP3603 Series) ou « Precision Plus Protein All Blue

Standards » (Biorad, 161-0373) ont été déposés en parallèle (**figure 35**). La migration a été réalisée à température ambiante sous un courant de 150 V dans un tampon contenant 25 mM de Tris-HCl, 1 % de SDS et 192 mM de glycine.

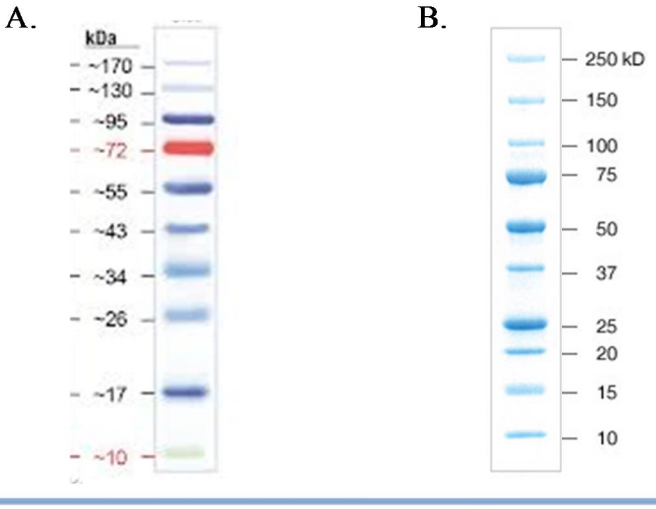


Figure 35 : Marqueurs de taille protéiques.
A. = BioReagents EZ-Run Prestained *Rec* Protein Ladder (Fisher, BP3603)
B. = Precision Plus Protein All Blue Standards » (Biorad, 161-0373)

II-4-3-2- Transfert électrophorétique de protéines sur membrane de nitrocellulose.

Les protéines du gel d'électrophorèse ont été transférées sur une membrane de nitrocellulose (VWR) selon une technique très proche de celle décrite par (Towbin et al., 1979). La membrane a préalablement été hydratée pendant 1 minute dans de l'eau distillée. A la fin de l'électrophorèse, la cassette de transfert a été réalisée comme illustrée sur la **figure 36**, puis disposée dans une cuve de transfert contenant 600 mL de tampon de transfert (200 mM glycine, 20% éthanol, 25 mM Tris et 0,1% SDS). Le transfert des protéines du gel d'électrophorèse sur la membrane de nitrocellulose a été obtenu en 2 heures sous une intensité électrique de 0,6 A. L'efficacité du transfert a ensuite été contrôlée par une coloration réversible des protéines au rouge Ponceau S (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 (10 mM Tris-HCl, 500 mM 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 ont été analysées par affino- ou immunodétection.

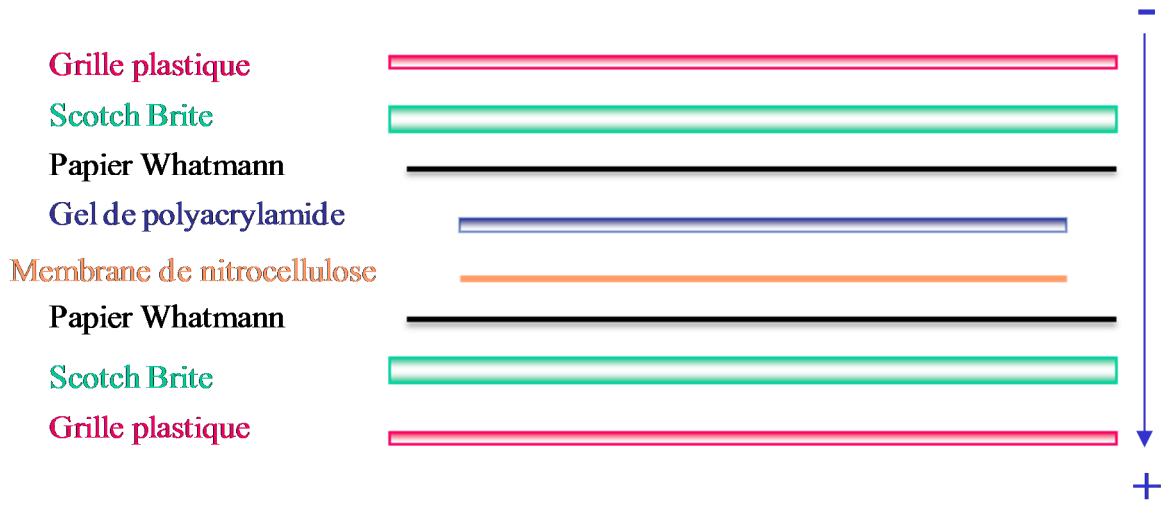


Figure 36 : Représentation schématique de la cassette de transfert des gels de polyacrylamide sur membrane de nitrocellulose.

II-4-3-3- Immuno-marquage sur empreinte.

II-4-3-3-1- Anticorps anti-épitope V5.

La membrane de nitrocellulose a été saturée sur la nuit sous agitation à température ambiante, dans une solution tampon de TBST (tampon TBS + 0,1 % Tween 20) contenant 5 % de lait écrémé, puis incubée pendant 2 heures dans 4 mL d'une solution d'anticorps primaire anti-épitope V5 (Invitrogen) dilué au 1/5000ème dans du tampon TBST contenant 1 % de lait écrémé. Cet anticorps monoclonal produit chez la souris reconnaît l'épitope V5 placé à l'extrémité C-terminale des protéines recombinantes. A l'issu 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 2 heures avec un anticorps secondaire anti-souris couplé soit à la phosphatase alcaline (dilution au 1/30000ème), soit à la peroxydase de raifort (HRP, dilution au 1/3000ème) dans une solution de tampon TBST contenant 1 % de lait écrémé. Suite à ce deuxième marquage, une série de 6 lavages de 5 minutes dans le tampon TBST suivie de 3 lavages de 5 minutes dans le tampon TBS a été réalisée. Toutes ces étapes de traitement de la membrane ont été effectuées sous agitation à température ambiante.

II-4-3-3-2- Anticorps anti-xylose et anti-fucose.

Les épitopes α-1,3-fucose et β-1,2-xylose portés sur le core glycannique sont spécifiques des glycoprotéines végétales et sont fortement immunogènes chez les mammifères (Bardor et al., 2003). Des anticorps dirigés contre ces épitopes ont été élaborés et purifiés à partir de sera de lapins (Faye et al., 1993 ; **figure 37**).

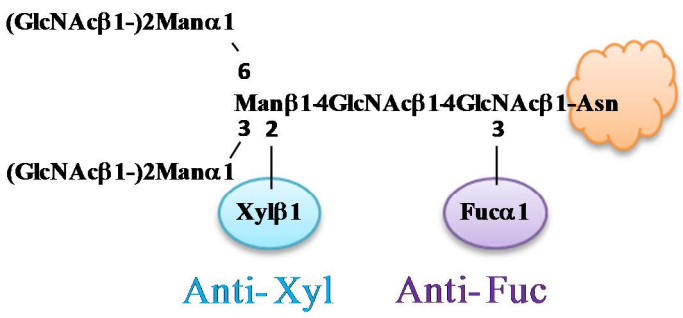


Figure 37 : Représentation des épitopes reconnus par les anticorps anti-α-1,3-fucose (Anti-Fuc) et anti-β-1,2-xylose (Anti-Xyl).

L'immuno-marquage avec les anticorps anti-α-1,3-fucose et anti-β-1,2-xylose ont été réalisés en suivant le même protocole que celui de (Faye et al., 1993). Ce protocole est très proche de celui décrit dans le paragraphe précédent. Cependant la saturation a été réalisée dans du tampon TBST et le second marquage a été réalisé avec des anticorps anti-lapin couplé à la HRP dilués dans une solution de tampon TBST.

II-4-3-4- Affino-marquage sur empreinte.

Les affino-détections ont été réalisées avec des lectines spécifiques des N-glycannes telles que la concanavaline A (Con A ; GE Healthcare) ou des lectines biotinylées comme la Phytohémagglutinine E et L (PHA-E et L), l'*Erythrina Crista Galli* Agglutinine (ECA), la *Ricinus Communis* Agglutinine 120 (RCA 120) et l'Agglutinine de cacahuète (PNA) (Vector Laboratories Inc.). Les marquages ont été effectués comme précédemment décrit par (Fitchette et al., 2007). Les lectines biotinylées ont été détectées en utilisant la streptavidine couplée avec la HRP (GE Healthcare). La Con A, quant à elle, a directement été détectée par la HRP.

II-4-3-5- Révélation.

II-4-3-5-1- Electrochimioluminescence.

Les membranes de nitrocellulose marquées avec un anticorps couplés à la HRP ont été révélées par électrochimioluminescence en utilisant les réactifs « ECL™ Western Blotting Analysis System » commercialisés par la société GE HealthCare Amersham™. Afin de réaliser cette révélation, la membrane a été égouttée pour éliminer le tampon TBS, puis placée sur un film plastique. Un millilitre de la solution 1 du kit « ECL™ Western Blotting Analysis System » a été ajouté à 1 ml de la solution 2, puis le mélange a été déposé pendant 1 minute sur la membrane. 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 10 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 1/4 dans de l'eau, puis rincé à l'eau, avant d'être déposé dans une solution de fixation (Kodak) diluée au 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.

II-4-3-5-2- 4-chloro-1-naphtol.

Le 4-chloro-1-naphtol est un réactif utilisé lors de la révélation de membrane marquée avec un anticorps couplé à la HRP. En effet, ce composé est reconnu comme substrat par la HRP générant ainsi un produit coloré. Afin de réaliser cette révélation, une solution contenant 30 mg de 4-chloro-1-naphtol dilués dans 10 mL de méthanol a été mélangée extemporanément sur la membrane avec une solution contenant 50 mL de TBS et 500 µL d'eau oxygénée 30%. La réaction a été effectuée sous agitation à température ambiante.

II-4-3-5-3- Réactif BCIP/NBT

Le réactif BCIP/NBT (5-bromo-4-chloro-3-indolyl-phosphate/nitro blue tetrazolium ; Western blue stabilized substrate for alkaline phosphatase, Promega) a été utilisé pour révéler les membranes marquées avec des anticorps couplés à la phosphatase alkaline. La solution a directement été appliquée sur la membrane et la révélation a été réalisée à température ambiante.

II-4-4- Immunodétection sur cellules de mammifères.

L'immunodétection des protéines sur cellules a été réalisée selon le principe présenté dans la **figure 38**.

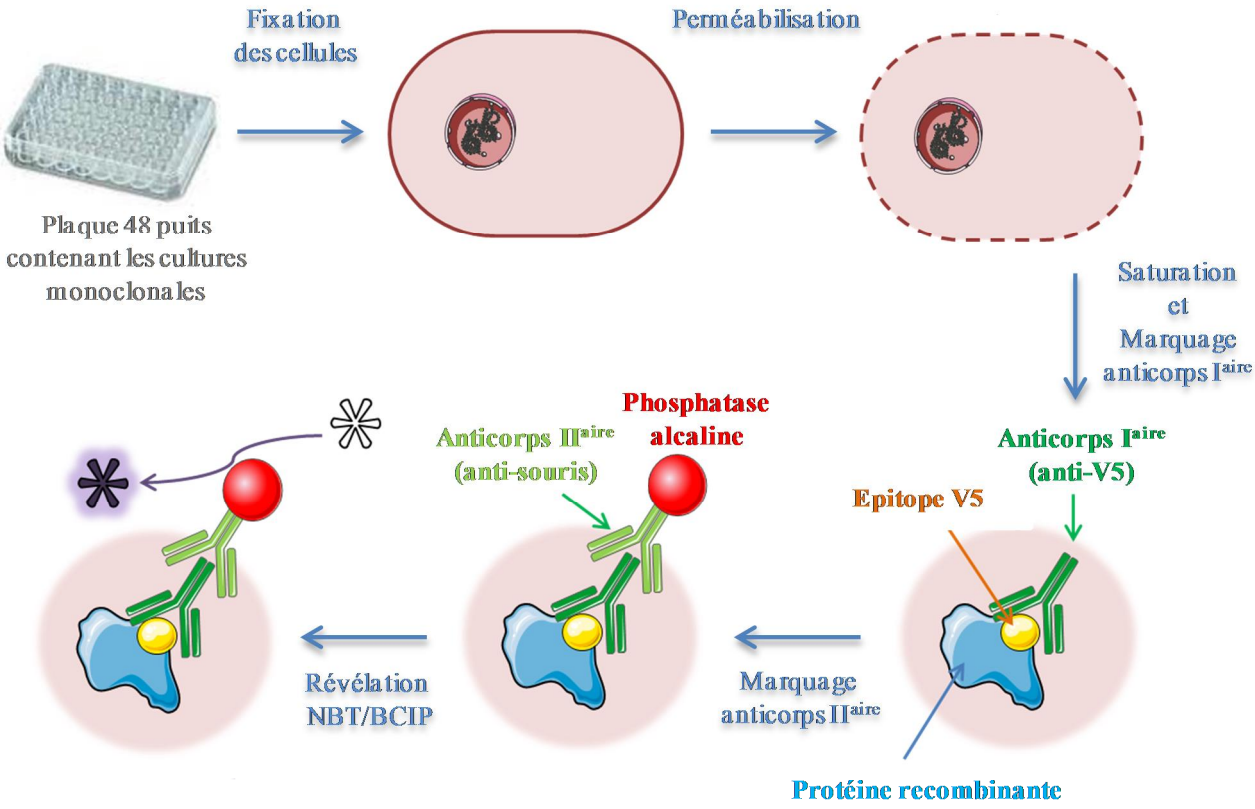


Figure 38 : Principe de l'immunodétection sur cellules par l'anticorps anti-V5.
L'épitope V5 des protéines recombinantes est reconnu par un anticorps primaire anti-V5. Cet anticorps, produit chez la souris, est lui-même reconnu par un anticorps secondaire anti-souris couplé à la phosphatase alcaline. La révélation est réalisée par oxydation du réactif NBT/BCIP.

II-4-4-1- Fixation et perméabilisation des cellules de mammifère.

L'immunodétection sur les cellules de mammifères a été effectuée sur une plaque 48 puits contenant des cultures monoclonales de cellules CHO FuT8-/- transfectées avec le gène de la FuT54599. Afin de réaliser cette immunodétection, les cellules ont été fixées et perméabilisées. Pour cela, le milieu de culture de chaque puits a été retiré et les cellules ont été fixées pendant 1 heure 30 minutes à température ambiante avec une solution contenant 4 % de paraformaldéhyde. Les cellules ont ensuite été lavées 3 fois pendant 5 minutes avec du Dulbecco's Phosphate-Buffered Saline (DPBS ; Gibco), puis perméabilisées avec du DPBS contenant 0,1 % de Triton X-100 (Sigma) pendant 10 minutes, avant d'être lavées à nouveau 2 fois pendant 5 minutes avec du DPBS.

II-4-4-2- Immunomarquage sur cellules avec l'anticorps anti-V5.

Le criblage des lignées monoclonales CHO FuT8-/- transfectées avec la FucT54599 a été réalisé par l'étude de l'expression de la protéine recombinante révélée par immunodétection sur cellules avec l'anticorps anti-V5 (Invitrogen). Pour éviter le bruit de fond, les puits ont été saturés pendant 30 minutes avec du PBS contenant 0,1 % Tween (TPBS) et 5 % de lait écrémé. La plaque a ensuite été incubée pendant 2 heures à température ambiante, sous agitation, avec 125 µL d'une solution d'anticorps anti-V5 dilué au 1/200ème dans du TPBS contenant 1 % de lait écrémé. Six lavages de 5 minutes ont alors été réalisés avec du TPBS, suivi d'un marquage avec l'anticorps secondaire. Pour cela, 125 µL d'une solution d'anticorps anti-souris couplés à une phosphatase alcaline (Sigma) dilués au 1/50ème dans du TPBS contenant 1 % de lait ont été déposés dans chaque puits et mis à incuber pendant 1 heure 30 minutes à température ambiante sous agitation. Les puits ont été lavés de nouveau 6 fois pendant 5 minutes avec du TPBS, puis 5 minutes avec du TBS. La révélation a été effectuée par ajout de 100 µL de réactif BCIP/NBT (5-bromo-4-chloro-3-indolyl-phosphate/nitro blue tetrazolium ; Western blue stabilized substrate for alkaline phosphatase ; Promega) dans chacun des puits.

II-4-5- Purification des N-glycannes.

Les N-glycannes isolés par digestion avec la PNGase A et F dans le but d'être analysés par spectrométrie de masse ont été purifiés par élutions successives à travers une colonne hydrophobe C18 et une colonne de charbon actif.

II-4-5-1- Chromatographie sur colonne hydrophobe C18.

Les colonnes C18 (Varian) sont des colonnes apolaires retenant les composés hydrophobes, tels que les protéines, et laissant passer les composés plutôt hydrophiles tels que les sucres. Les extraits protéiques digérés par la PNGase F ou A ont été purifiés sur une colonne hydrophobe C18, afin de séparer les N-glycannes des peptides. La colonne de phase inverse a été conditionnée avec 5 mL d'éthanol, puis 5 mL d'eau déionisée. L'échantillon a alors été déposé et les N-glycannes libres ont été obtenus dans la fraction non retenue et par lavage de la colonne avec 5 mL d'eau déionisée. Les N-glycannes ont ensuite été congelés, puis lyophilisés.

II-4-5-2- Chromatographie sur colonne de charbon actif (Carbograph).

Les N-glycannes libres, récupérés après purification par chromatographie sur colonne hydrophobe C18, ont été dessalés sur une colonne de carbone graphité (Alltech) selon Packer *et al.,* 1998. Cette colonne a préalablement été conditionnée par l'ajout successif de 2 mL de méthanol, 2 mL de NaOH 1M, 4 mL d'eau distillée, 4 mL d'acide acétique 30%, 4 mL d'eau distillée, 2 mL d'une solution contenant 50% d'acétonitrile et 0,1% d'acide trifluoroacétique (TFA) et 4 mL d'une solution contenant 5% d'acétonitrile et 0,1% de TFA. Les N-glycannes lyophilisés ont été repris dans 1 mL d'eau distillée, puis déposés sur la colonne. Celle-ci a ensuite été lavée avec 4 mL d'eau distillée, puis 2 mL d'une solution d'acétonitrile 5% contenant 0,1% de TFA. Les sucres, contenus dans la fraction retenue, ont été élués avec 2 mL d'une solution contenant 50% d'acétonitrile et 0,1% de TFA. Cette fraction a ensuite été placée sous une rampe à air, afin d'évaporer l'acétonitrile présent dans l'échantillon, puis congelée et lyophilisée.

II-4-6- Purification des N-glycannes.

II-4-6-1- Marquage des N-glycannes.

Les N-glycannes ont été marqués avec du 2-aminobenzamide (2-AB : Sigma-Aldrich) en utilisant le protocol décrit par (Bigge et al., 1995). Après incubation à 60°C pendant 2 heures, les N-glycannes marqués au 2-AB ont été purifiés par chromatographie sur papier comme décrit par (Séveno et al., 2008) et analysés par spectrométrie de masse MALDI-TOF

II-4-6-2- Digestion par exoglycosidase des N-glycannes couplés au 2-AB.

Les N-glycannes marqués au 2-AB ont été digérés avec l'α-mannosidase ou l'α-L-fucosidase en suivant le principe décrit dans (Bakker et al., 2001). Pour la digestion avec l'α-mannosidase de haricot (Sigma-Aldrich), 2,5 µL de N-glycannes marqués au 2-AB ont été incubés avec 1 µL d'eau et 214 mU d'enzyme pendant 48 heures à 37°C. Pour la digestion avec l'α-L-fucosidase de rein bovin (Sigma-Aldrich), l'enzyme a été dessalée avant utilisation et reprise dans une solution d'acétate de sodium 40 µM, pH 5,5. Deux microlitres et demi de N-glycannes marqués au 2-AB ont été incubés avec 80 mU d'enzyme à 37°C pendant 48 heures. Dans les deux cas, les échantillons ont été lyophilisés après digestion et repris dans 10 µL d'eau contenant 0,1% de TFA avant d'être analysés par spectrométrie de masse.

II-4-7- Analyse des N-glycannes par spectrométrie de masse MALDI TOF.

Les N-glycannes couplés au 2-AB ont été analysés avec un spectromètre de masse à temps de vol Voyager DE-Pro (Applied Biosystems) équipé d'un laser à azote de 337 nm. La préparation de la cible a été réalisée par dépôt de 1 µL de matrice DHB sur la plaque. Cette matrice est composée de 5 $mg.mL^{-1}$ d'acide 2,5-dihydroxybenzoïque (DHB, Gentisic acid, SIGMA) dissout dans un mélange acétonitrile / TFA 0,1% avec un ratio de 30 : 70 (v/v). Parallèlement, 1 µL de N-glycannes purifiés repris dans 10 µL d'eau déionisée contenant 0,1% de TFA, a été ajouté sur la plaque et mélangé à la matrice. Les spectres ont été enregistrés en mode réflectron positif, sous une tension d'accélération de 20 kV et un temps de délai de 100 ns. Le paramétrage du spectromètre de masse et le traitement des données ont été respectivement réalisés avec les logiciels Data Voyager™ et Data Explorer™ 4.0 Software. Chaque acquisition est une moyenne d'environ 5 fois 1000 tirs lasers. Les spectres ont été calibrés de façon externe avec des calibrants commerciaux (Applied Biosystems) contenant les peptides suivants : des-Arg^1-Bradykinin (904, 4681 Da), Angiotensine I (1296, 6853 Da), Glu^1-Fibrinopeptide B (1570,6774 Da) et $ACTH_{18-39}$ (2,465.1989 Da).

Chapitre 3 : Résultats

I- Processus de N-glycosylation et caractérisation fonctionnelle de la GnT I de *Phaeodactylum tricornutum*.

Dans cette première partie, les résultats sont présentés sous la forme d'un article scientifique publié dans « *The Journal of Biological Chemistry* » et intitulé « N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its *N*-acetylglucosaminyltransferase I enzyme ». Ces travaux démontrent, dans un premier temps, l'existence d'une voie de N-glycosylation chez *P. tricornutum*. En effet, des analyses bioinformatiques réalisées sur le génome de cette diatomée ont permis d'identifier un jeu complet de séquences codant potentiellement pour des protéines impliquées dans la synthèse du précurseur oligosaccharidique $Glc_3Man_9GlcNAc_2$-PP-dolichol, des sous-unités catalytiques de l'OST, des glucosidases et des chaperonnes du RE, ainsi que pour une protéine EDEM. Cela suggère que *P. tricornutum* est capable de synthétiser un précurseur oligosaccharidique puis de le transférer sur ces protéines et qu'elle possède la machinerie du RE requise pour la mise en place du cycle de contrôle qualité et le bon repliement des protéines. Ces analyses génomiques ont également mis en évidence la présence d'un gène codant potentiellement pour une α-mannosidase I golgienne suggérant l'existence, chez cette diatomée, de N-glycannes oligomannosidiques allant de Man-9/8 à Man-5.

En accord avec ces prédictions bioinformatiques, des analyses biochimiques ont montré que les protéines de la souche Pt 1.8.6 fusiforme de *P. tricornutum* sont principalement glycosylées par des structures oligomannosidiques. En effet, les protéines provenant de cette souche sont affinodetectées par la concanavaline A, une lectine spécifique des N-glycannes mannosylés. De plus, des analyses structurales détaillées menées par spectrométrie de masse MALDI-TOF sur les N-glycannes isolés et purifiés à partir des glycoprotéines de cette même souche ont permis la détection d'oligosaccharides majoritaires correspondant à des N-glycannes oligomannosidiques allant de Man-9 à Man-5. Cependant, d'autres espèces glycanniques minoritaires du type paucimannose telles que Man-3, Man-4 et $Man_3FucGlcNAc_2$ ont également été détectées. La présence de ce N-glycanne fucosylé corrobore le faible signal obtenu par immunodétection sur empreinte des protéines avec un anticorps spécifique de l'épitope α-1,3-fucose présent sur le core glycannique et avec la présence de 3 gènes codant potentiellement pour des α-1,3-fucosyltransférases prédits dans le génome de *P. tricornutum*. Par ailleurs, des gènes codant potentiellement pour une GnT I et une α-mannosidase II, enzymes responsables des premières étapes de maturation des N-glycannes complexes, ont également été prédits dans le génome de cette diatomée. Bien

qu'aucun glycanne présentant un résidu GlcNAc terminal, produit de la GnT I, n'ait été détecté sur les glycoprotéines par spectrométrie de masse, l'identification d'oligosaccharides comme Man-3, Man-4 et $Man_3FucGlcNAc_2$ suggère l'existence d'une voie de N-glycosylation GnT I dépendante chez *P. tricornutum*. En effet, chez les invertébrés et les plantes, ces structures paucimannoses proviennent de la dégradation, par des N-acétylglucosaminidases, de N-glycannes complexes possédant une GlcNAc terminale (Vitale et Chrispeels, 1984 ; Altmann et al., 1995 ; Zhang et al., 2003). Or, deux gènes codant potentiellement pour des N-acétylglucosaminidases ont été identifiés dans le génome de cette microalgue.

Afin d'étayer l'hypothèse de la présence d'une voie de N-glycosylation GnT I dépendante chez *P. tricornutum*, la seconde partie de ces travaux a consisté en la caractérisation fonctionnelle de la GnT I de cette diatomée. Pour cela, le gène codant pour la GnT I de *P. tricornutum* a été cloné puis exprimé dans des cellules mutantes d'ovaires de hamster chinois nommées CHO Lec1, déficientes en activité GnT I endogène et donc incapables de réaliser la synthèse de leurs N-glycannes complexes (Stanley et al., 1975 ; Ripka et al., 1990). Cette approche par complémentation dans un système hétérologue a permis de démontrer que le gène prédit chez *P. tricornutum* code pour une GnT I fonctionnelle capable de restaurer la maturation de N-glycannes oligomannosidiques en N-glycannes complexes dans les cellules CHO Lec1.

Ces résultats constituent la première description de la voie de N-glycosylation d'une microalgue. De plus, la GnT I de *P. tricornutum* représente la première glycosyltransférase de microalgue à avoir été clonée et caractérisée d'un point de vue fonctionnel. Une partie des résultats présentés dans cet article a également été valorisée au travers d'un brevet déposé le 27 juillet 2010 qui a été étendu en PCT européen un an plus tard (**annexe 2**).

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 8, pp. 6152–6164, February 25, 2011
© 2011 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

N-Glycans of *Phaeodactylum tricornutum* Diatom and Functional Characterization of Its *N*-Acetylglucosaminyltransferase I Enzyme*

Received for publication, August 16, 2010, and in revised form, December 8, 2010 Published, JBC Papers in Press, December 17, 2010, DOI 10.1074/jbc.M110.175711

Bérengère Baïet[‡1], Carole Burel[‡], Bruno Saint-Jean[§], Romain Louvet[‡], Laurence Menu-Bouaouiche[‡], Marie-Christine Kiefer-Meyer[‡], Elodie Mathieu-Rivet[‡], Thomas Lefebvre[¶], Hélène Castel[¶], Aude Carlier[§2], Jean-Paul Cadoret[§], Patrice Lerouge[‡], and Muriel Bardor[‡3]

From the [‡]Université de Rouen, Laboratoire Glycobiologie et Matrice Extracellulaire Végétale, Faculté des Sciences, 76821 Mont-Saint-Aignan Cédex, France, the [§]Laboratoire Physiologie et Biotechnologies des Algues, Institut Français de Recherche pour l'Exploitation de la Mer, Centre de Nantes, 44311 Nantes Cedex 03, France, and the [¶]Université de Rouen, Laboratoire de Différenciation et Communication Neuronale et Neuroendocrine, INSERM U982, IFRMP23, Faculté des Sciences, 76821 Mont-Saint-Aignan Cedex, France

N-Glycosylation, a major co- and post-translational event in the synthesis of proteins in eukaryotes, is unknown in aquatic photosynthetic microalgae. In this paper, we describe the *N*-glycosylation pathway in the diatom *Phaeodactylum tricornutum*. Bio-informatic analysis of its genome revealed the presence of a complete set of sequences potentially encoding for proteins involved in the synthesis of the lipid-linked $Glc_3Man_9GlcNAc_2$-PP-dolichol *N*-glycan, some subunits of the oligosaccharyltransferase complex, as well as endoplasmic reticulum glucosidases and chaperones required for protein quality control and, finally, the α-mannosidase I involved in the trimming of the *N*-glycan precursor into Man-5 *N*-glycan. Moreover, one *N*-acetylglucosaminyltransferase I, a Golgi glycosyltransferase that initiates the synthesis of complex type *N*-glycans, was predicted in the *P. tricornutum* genome. We demonstrated that this gene encodes for an active *N*-acetylglucosaminyltransferase I, which is able to restore complex type *N*-glycans maturation in the Chinese hamster ovary Lec1 mutant, defective in its endogeneous *N*-acetylglucosaminyltransferase I. Consistent with these data, the structural analyses of *N*-linked glycans demonstrated that *P. tricornutum* proteins carry mainly high mannose type *N*-glycans ranging from Man-5 to Man-9. Although representing a minor glycan population, paucimannose *N*-glycans were also detected, suggesting the occurrence of an *N*-acetylglucosaminyltransferase I-dependent maturation of *N*-glycans in this diatom.

Microalgae are a group of aquatic photosynthetic microorganisms that are so diverse that they have been gathered in a "paraphylum." Among these microalgae, diatoms belong to the heterokont group and are responsible for approximately 40% of marine primary productivity (1, 2). Despite their physiological relevance in the marine ecosystem, molecular and cellular processes in diatom remain widely unknown. For example, so far, little is known about secretion, post-translational modifications, and intracellular trafficking of proteins in diatoms. Diatom species are usually classified into two major groups, the bi/multipolar centrics and the pennates. Recently, the genome of a pennate diatom, *Phaeodactylum tricornutum*, became available (3) revealing a wealth of information about diatom biology. Access to this data (Joint Genome Institute, Walnut Creek, CA), together with the fact that *P. tricornutum* is easy to culture *in vitro* and can be genetically transformed (4, 5), provides the opportunity to perform comparative genomic studies and to dissect biosynthetic pathways.

N-Glycosylation is a major co- and post-translational modification in the synthesis of proteins in eukaryotes. *N*-Glycan processing occurs in the secretory pathway and is essential for glycoproteins destined to be secreted or integrated in the membranes. In this process, a $Man_5GlcNAc_2$-PP-dolichol oligosaccharide intermediate is assembled by the stepwise addition of monosaccharides to dolichol pyrophosphate on the cytosolic face of the endoplasmic reticulum (ER).[4] This intermediate is then extended in the lumen of the ER until a $Glc_3Man_9GlcNAc_2$-PP-dolichol *N*-glycan precursor is completed (6). This precursor is transferred by the oligosaccharyltransferase (OST) multisubunit complex onto the asparagine residue of the consensus Asn-Xaa-Ser/Thr sequences of a target nascent protein (6). The precursor is then deglucosylated/reglucosylated to ensure the quality control of the neosynthetized protein through the interaction with ER-resident chaperones such as calnexin and calreticulin. These ER events are crucial for the proper folding of the secreted proteins and are highly conserved in the eukaryotes investigated so far (7).

* This work was supported by funds from the University of Rouen (to P. L. and M. B.).

[1] Fellow of the French Ministry for Research and Higher Education.

[2] Present address: Algenics S. A. S., Pôle Bio Ouest, Rue du Moulin de la Rousselière, 44800 Saint Herblain, France.

[3] To whom correspondence should be addressed. Present address: Bioprocessing Technology Institute, Agency for Science Technology and Research (A*STAR), 20 Biopolis Way, #06-01, Centros, Singapore 138668, Singapore.

[4] The abbreviations used are: ER, endoplasmic reticulum; α-Man, α-mannosidase; GnT, *N*-acetylglucosaminyltransferase; DMB, 1,2-diamino-4,5-methylene dioxybenzene; 2-AB, 2-aminobenzamide; RCA 120, *R. communis* agglutinin 120; PNGase F, peptide *N*-glycosidase F; PNGase A, peptide *N*-glycosidase A; Endo H, endoglycosidase H; ALG, asparagine-linked glycosylation; GlcNAc, *N*-acetylglucosamine; OST, oligosaccharyltransferase; FucT, fucosyltransferase.

In contrast, evolutionary adaptation of *N*-glycan processing in the Golgi apparatus has given rise to a large variety of organism-specific complex structures that allow the protein to carry out diverse glycan-mediated biological functions. α-Mannosidase I (α-Man I) located in the early compartment of the Golgi apparatus (Cis cisternae) first degrades the oligosaccharide precursor into high mannose type *N*-glycans ranging from $Man_9GlcNAc_2$ (Man-9) to $Man_5GlcNAc_2$ (Man-5). *N*-Acetylglucosaminyltransferase I (GnT I) then transfers the first *N*-acetylglucosamine (GlcNAc) residue on the α(1,3)-mannose arm of $Man_5GlcNAc_2$, enabling the initiation of the synthesis of multiple structurally different complex type *N*-glycans. Following GnT I action, α-mannosidase II (α-Man II) and *N*-acetylglucosaminyltransferase II (GnT II) give rise to the synthesis of the core $GlcNAc_2Man_3GlcNAc_2$, which is finally matured into organism-specific complex *N*-glycans by transfer of various monomers by characteristic glycosyltransferases. GnT I and thus the GnT I-dependent maturation of *N*-glycans appeared during evolution at the same period as metazoans (8). Complex type *N*-glycans were demonstrated to be engaged in crucial steps of the development of pluricellular organisms (8–11). For instance, GnT I-null embryos of mice die at ~10 days after fertilization, indicating that complex *N*-glycans are required for morphogenesis in mammals (9, 10). Similarly, inactivation of the GnT I in worm and fly reduces their viabilities (8, 11). In plants, *Arabidopsis cgl* mutant, defective in GnT I activity, was demonstrated to grow normally in standard culture conditions (12). However, these plants exhibited a strong phenotype in salt-induced stress conditions, for example, suggesting a role for mature plant *N*-glycans in specific physiological processes (13–15). In contrast, animal cultured cells having GnT I null mutations usually grow normally (16).

Although major data regarding protein *N*-glycosylation have been established in yeast and higher eukaryotes, nothing is known on *N*-glycan biosynthesis and structures in microalgae. In this paper, we describe the *N*-glycosylation pathway in the diatom *P. tricornutum.* We also demonstrate that the predicted GnT I from *P. tricornutum* is able to complement the biosynthesis of complex type *N*-glycans in the CHO Lec1 mutant, which is defective in its endogenous GnT I. To the best of our knowledge, this is the first functional characterization of a *N*-glycan glycosyltransferase from microalgae.

EXPERIMENTAL PROCEDURES

Materials—TRIzol reagent, anti-V5 antibodies, α-minimum essential medium, FBS, geneticin, and pcDNA3.1/V5-His-TOPO TA kit were purchased from Invitrogen. Silica membrane spin column and RNeasy mini kit were obtained from Qiagen. RQ1 DNase was from Promega (Charbonnières-les-bains, France). A high capacity cDNA reverse transcription kit and 2× *Power* SYBR® Green I PCR Master Mix, Sequazyme™ peptide mass standards kit, as well as the Voyager DE-Pro MALDI-TOF instrument, were from Applied Biosystems (Foster City, CA). CHO K1 (wild type) and CHO Lec1 cell lines were obtained from American Type Culture Collection (Manassas, VA). Nucleofactor device was from Amaxa (Köln, Germany). Pierce BCA protein assay kit and Extensor Hi-Fidelity PCR enzyme mix were purchased from Thermo Scientific (Rockford, IL). Pepsin, jack bean α-mannosidase proteomic grade, α-L-fucosidase from bovine kidney, 2,5-dihydroxybenzoic acid, 1,2-diamino-4,5-methylene dioxybenzene (DMB), UDP-diphosphogalactose, 2-aminobenzamide (2-AB), β(1,4)-galactosyltransferase from bovine milk, and Nonidet P40 (Igepal CA-630) were obtained from Sigma-Aldrich. Biotinylated lectins such as phytohemagglutinin E and L, *Ricinus communis* agglutinin 120 (RCA 120), *Erythrina crista galli* agglutinin, and peanut agglutinin were from Vector Laboratories Inc. (Burlingame, CA). Concanavalin A and streptavidin-horseradish peroxidase conjugate were from GE Healthcare. An ECL revelation kit was purchased from Amersham Biosciences. Antibodies directed against core anti-β(1,2)-xylose and core α(1,3)-fucose were from AgriSera (Vännäs, Sweden). Peptide *N*-glycosidase F (PNGase F), peptide *N*-glycosidase A (PNGase A), and Endoglycosidase H (Endo H) were from Roche Applied Science. C18 column (C18 monomeric S/N E000930-10-2) and C18 cartridge were from Varian Inc. (Palo Alto, CA). Carbograph cartridges were from Alltech/Grace Davison Discovery Sciences (Templemars, France). NanoDrop® ND-1000 was from Thermo Fisher (Illkirch, France). Mx3000PTM Q-PCR system was from Stratagene (La Jolla, CA). *Fastprep* 24 device and 50-ml BigD Lysing tubes were purchased from MP Biomedicals (Illkirch, France).

In Silico Genome Analysis—In the *P. tricornutum* genome, annotation of genes involved in the *N*-glycan pathway was carried out by BLASTP or TBLASTN analyses with genomic sequences from *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Physcomitrella patens*, *Medicago truncatula*, *Zea mays*, *Nicotiana plumbaginifolia*, and *Oryza sativa*. Searches for signal peptides and cell localization/targeting of mature proteins were done using Signal P, Signal-BLAST, and Target P. Transmembrane domains were predicted using TMHMM, TOPPRED, and HMMTOP. Pfam domains were identified using Pfam (Wellcome Trust Sanger Institute, Cambridge, UK). The phylogenetic tree was drawn using the *Phylogeny.fr* platform (17) and following three steps: (i) complete sequences were aligned with ClustalW (v2.0.3) (18); (ii) after alignment, ambiguous regions (*i.e.* containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) (19); and (iii) the phylogenetic tree was built using the maximum likelihood method implemented in the PhyML program (v3.0 aLRT) (20, 21). Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3) (22). Thirty-one sequences were selected from the CAZy GT13 glycosyltransferase family (23) or retrieved from the JGI database. Origin and accession numbers for the different GnT I protein sequences are as follows: *A. thaliana*, MGAT1_ARATH; *Bos taurus*, Q5E9I4_BOVIN; *Caenorhabditis elegans*, MGAT1_CAEEL; *Cricetulus griseus*, P70680_CRIGR; *Danio rerio* (1), A0JMK2_DANRE; *D. rerio* (2), Q6PE41_DANRE; *D. melanogaster*, Q60GL7_DROME; *Glycine max*, C6T9Z3_SOYBN; *H. sapiens*, MGAT1_HUMAN; *Mesaucricetus auratus*, Q9QWX4_MESAU; *M. musculus*, MGAT1_

MOUSE; *Nicotiana tabacum*, Q8VXD3_TOBAC; *Oryctolagus cuniculus*, MGAT1_RABIT; *O. sativa*, Q8RW24_ORYSA; *Populus trichocarpa*, B9HPX2_POPTR; *P. patens*, Q8L5D3_PHYPA; *Pongo abelii*, Q5RB13_PONPY; *Rattus norvegicus*, MGAT1_RAT; *Salmo salar*, B5X2V8_SALSA; *Schistosoma mansoni*, C4Q4I7_SCHMA; *Solanum tuberosum*, Q8VX56_SOLTU; *Sorghum bicolor*, C5YS64_SORBI; *Sus scrofa*, Q0KKC2_PIG; *Xenopus laevis*, Q90W56_XENLA; and *Z. mays*, B6TLH4_MAIZE. The GenBank™ accession number for *Triticum aestivum* was ABT33168. The JGI accession numbers were: *Micromonas pusilla* CCMP1545 (jgi|MicpuC2|32201|estExt_Genewise1Plus.C_30312); *Emiliania huxleyi* (jgi|Emihu1|460566|estExtDG_Genemark1.C_13800003); *Fragilariopsis cylindrus* (jgi|Fracy1|189180|e_gw1.9.153.1); *P. tricornutum* (GenBank™ TPA BK007891); and *Thalassosira pseudonana* (jgi|Thaps3|1600|fgenesh1_pg.C_chr_1000734).

Microalgal Strain and Culture Conditions—The strain of *P. tricornutum* P.t1.8.6 (CCAP1055/1) was grown in batch culture method using 2-liter flat-bottomed flasks. The nutritive medium used for this experiment consisted of natural seawater, sterilized by filtration through a 0.22-μm filter, enriched in Conway medium (24) and containing 40 mg·liter^{-1} sodium metasilicate. Diatom cells were maintained at 20 °C under continuous illumination (280–350 μmol photons m^{-2}·s^{-1}). The cells (20 × 10^6 cells·ml^{-1}) were then centrifuged at 5,000 × *g* for 20 min at 4 °C, and the resulting pellet was freeze-dried before biochemical analyses. *P. tricornutum* were grown in continuous culture conditions as described previously (25) for real time quantitative PCR experiments. At the steady state (15–20 × 10^6 cells·ml^{-1}), five samples (30 ml of each in triplicate) were harvested at 4,000 × *g* for 20 min at 4 °C. The supernatant was removed rapidly, and the cell pellets were resuspended in 1 ml of TRIzol, immediately frozen, and stored at −80 °C until RNA extraction.

Real Time Quantitative PCR Experiments—Total RNA was extracted from cells using the TRIzol method and then treated with RQ1 DNase to avoid DNA contamination and finally purified using the RNeasy mini kit. cDNA templates for PCR amplification were synthesized from 350 ng of total RNA using the High Capacity cDNA reverse transcription kit. Quantitative PCR was performed using *Power* SYBR® Green I PCR master mix in a final volume reaction of 25 μl. All of the reactions were performed following the instructions of the manufacturer with 5 μl of diluted cDNA (1/10) and 0.1 μM of specific primers. Quantitative measurements were performed in duplicate with a Stratagene Mx3000P™ Q-PCR system. The cycling parameters were one cycle of 10 min at 95 °C, followed by 40 cycles of 30 s at 95 °C and 60 s at 60 °C. The results were represented as the relative gene expression normalized to reference genes encoding for ribosomal protein small subunit 30 S and histone H4 (26). Specific primers for the catalytic domain of the *P. tricornutum* GnT I gene (GNT1-Q-Fwd, 5′-CGTACGAATCGCCCTTACTC-3′; and GNT1-Q-Rev, 5′-TTGCCGTCTTGTGAAATTACC-3′) were designed using the Primer3Plus program. The relative GnT I gene expression analysis was performed using the method already described (27, 28) where the comparative C_T method (29) and standard curve method were combined to calculate RNA molar ratio between the target and housekeeping genes (28).

Expression of P. tricornutum GnT I in CHO Cells—Genomic DNA was extracted from *P. tricornutum* cell pellets as described in Ref. 30. The GnT I gene was amplified from *P. tricornutum* genomic DNA using primers 5′-ATGCGGTTGTGGAAACGTAC-3′ and 5′-TCTTTTCGGTGACGGAATG-3′ and Extensor Hi-Fidelity PCR enzyme mix. Then this GnT I gene was cloned according to the supplier's instructions in the pcDNA3.1/V5-His-TOPO vector, leading to the expression of the *P. tricornutum* GnT I fused with the V5 epitope under the control of the T7 promoter. This construction pcDNA3.1/V5-His-TOPO-GnT I was sequenced and presented the following mutations: Gln-130 → Arg; Val-148 → Ile; Gly-159 → Ser; Lys-196 → Gln; Val-337 → Ala; and Asn-422 → Tyr (GenBank™ BankIt 1370344). CHO Lec1 mutant cells were transfected by electroporation with this construct. The cells were trypsinized (0.05%), triturated in α-minimum essential medium containing 10% FBS, pelleted by centrifugation, and resuspended in 100 μl of solution V for nucleofection by an Amaxa Nucleofector device set to program U-016 with the linearized pcDNA3.1/V5-His-TOPO-GnT I and 500 μg of sterile sonicated salmon sperm DNA. Then the transfection was followed by repetitive rounds of limiting dilution of cells in 400 μg·ml^{-1} of geneticin for selection. CHO wild type and CHO Lec1 mutant were grown in α-minimum essential medium supplemented with 10% FBS at 37 °C in a humidified incubator with an atmosphere of 5% CO_2. CHO Lec1 mutant complemented with *P. tricornutum* GnT I was grown in the same conditions with 600 μg·ml^{-1} of geneticin.

Extraction of Proteins from P. tricornutum—Two g of lyophilized *P. tricornutum* cells were lysed in 750 mM Tris-HCl pH 8 buffer containing 15% (w/v) of sucrose, 2% (v/v) of β-mercaptoethanol, and 1 mM phenylmethylsulfonyl fluoride (extraction buffer). The cell lysis was done in a 50-ml BigD Lysing tube and assisted by the *Fastprep* 24 (15 times 30 s at 6 M·s^{-1}). The mixture was then centrifuged at 4 °C for 5 min at 4,000 × *g*. The pellet was washed once with 10 ml of extraction buffer and spun again at 4 °C for 5 min at 4,000 × *g*. The resulting supernatants were pooled prior to centrifugation at 15,000 × *g* for 30 min at 4 °C. The supernatant was then dialyzed against water for 48 h at 4 °C prior to lyophilization. Protein quantification was then performed using the Pierce BCA protein assay kit and bovine serum albumin as protein standard. Proteins from green onion were prepared in parallel and used as a positive control for affino- and immunodetection analyses.

Immunoblotting and Affinoblotting Analysis—Fifty μg of total protein extract from *P. tricornutum* were separated by SDS-PAGE using a 12% polyacrylamide gel. The proteins were transferred onto nitrocellulose membrane and stained with Ponceau Red to control the transfer efficiency. Affinodetections were carried out as described previously (31) using concanavalin A or biotinylated lectins such as phytohemagglutinin E and L, *E. crista galli* agglutinin, RCA 120, and peanut agglutinin. Biotinylated lectins were detected using

streptavidin coupled with horseradish peroxidase, and concanavalin A was directly detected by horseradish peroxidase. Final developments of the blots were obtained using 4-chloro-1-naphtol or ECL as substrate. Immunodetections were performed using specific core β(1,2)-xylose and core α(1,3)-fucose antibodies as reported previously (32). Oxidation of the glycan moiety of glycoproteins was carried out on the blots using sodium periodate according to (33). Immunodetection with anti-V5 antibodies was performed following the instructions of the supplier for dilution of the antibody and revelation (ECL kit).

Deglycosylation by PNGase F or Endo H—For deglycosylation with PNGase F, 0.5 mg of proteins was dissolved in 2 ml of a 0.1 M Tris-HCl buffer, pH 7.5, containing 0.1% SDS. The sample was then heated for 5 min at 100 °C for protein denaturation. After cooling down, 2 ml of 0.1 M Tris-HCl buffer, pH 7.5, containing 0.5% Nonidet P-40 were added to the sample. Digestion was performed with 10 units of PNGase F for 24 h at 37 °C. For deglycosylation by Endo H, 0.5 mg of protein extract was dissolved in 1% SDS and denaturated by heating for 5 min at 100 °C. The sample was then diluted five times in 500 μl of 150 mM sodium acetate buffer, pH 5.7, and incubated overnight at 37 °C with 10 milliunits of Endo H. Finally, proteins digested from either PNGase F or Endo H were precipitated by the addition of 4 volumes of ethanol overnight at −20 °C, separated by SDS-PAGE, and affinodetected with concanavalin A as reported previously (31).

In Vitro Galactosylation—The *in vitro* galactosylation was performed by treating 50 μg of protein at 37 °C for 24 h with 50 milliunits of β(1,4)-galactosyltransferase from bovine milk in 1 ml of 100 mM sodium cacodylate buffer, pH 6.4, supplemented with 5 μmol of UDP-galactose and 5 μmol of $MnCl_2$ (34). The sample was then freeze-dried. Proteins and glycoproteins were separated by SDS-PAGE and electroblotted onto nitrocellulose membrane. Glycoproteins were then affinodetected using biotinylated RCA 120 lectin (34).

Sialic Acid Analysis—Two mg of proteins were treated as described (35). Then the sample was submitted to DMB derivatization according to Ref. 36. DMB derivatives were separated by high performance liquid chromatography using a C18 column and detected by fluorescence using excitation and emission wavelengths of 373 and 448 nm, respectively (37). Neu5Ac was also coupled to DMB and used as a standard.

Isolation of N-linked Glycans from P. tricornutum—For *N*-glycan profiling, both PNGase A and PNGase F were used. In contrast to PNGase F, PNGase A is able to release *N*-linked oligosaccharides carrying a fucose α(1,3)-linked to the proximal glucosamine residue (38). Proteins (above 13 mg) were deglycosylated by PNGase F as described for CHO *N*-glycan profiling. For PNGase A, 5 mg of freeze-dried proteins were resuspended in 3 ml of 4 M Tris-guanidine HCl, pH 8.5, prior to denaturation with a 2 $mg \cdot ml^{-1}$ DTT solution. After a short 30 s sonication, the sample was incubated at 50 °C for 2 h. 1.5 ml of iodoacetamide, prepared in 0.6 M Tris buffer, pH 8.5, at 12 $mg \cdot ml^{-1}$, was then added to the sample, which was incubated in the dark for 2 h at room temperature. The sample was dialyzed for 72 h against water. The proteins were digested at 37 °C for 48 h with 10 mg of pepsin dissolved in 2 ml of 10 mM HCl, pH 2.2. After neutralization with 1 M ammonium hydroxide, the solution was heated for 5 min at 100 °C and lyophilized. Glycopeptides were then deglycosylated overnight at 37 °C with 1.5 milliunits of PNGase A in a 50 mM sodium acetate buffer, pH 5.5. *N*-Glycans released by either PNGase A or F were purified by successive elutions through a C18 and a Carbograph cartridges according to Ref. 39.

Isolation of N-linked Glycans from CHO Cells—Cells from wild type CHO and CHO Lec1 mutant and CHO Lec1 GnT I complemented cells were lysed by sonication (four times for 20 s) in 1 ml of 100 mM Tris-HCl buffer, pH 7.5, SDS 0.1%. After centrifugation at 100 × *g*, proteins from the supernatant were deglycosylated by PNGase F as described above, and then *N*-glycans were purified as already mentioned. Then *N*-glycan samples were concentrated and finally analyzed by MALDI-TOF mass spectrometer.

Preparation and Exoglycosidase Digestion of 2-AB Oligosaccharides—*N*-Glycans were labeled with 2-AB using the protocol described in Ref. 40. After incubation at 60 °C for 2 h, 2-AB-labeled *N*-glycans were purified by paper chromatography according to (41). The 2-AB-labeled *N*-glycans were finally analyzed by MALDI-TOF mass spectrometry before and after jack bean α-mannosidase or α-L-fucosidase treatments following the principle described in Ref. 39. For the α-mannosidase digestion, 2.5 μl of 2-AB labeled *N*-glycans were incubated with 1 μl of water and 214 milliunits of enzyme for 48 h at 37 °C. The α-L-fucosidase from bovine kidney was desalted prior to use and resuspended in 40 μM of sodium acetate, pH 5.5. Then 80 milliunits of enzyme was incubated with 2.5 μl of 2-AB labeled glycans at 37 °C for 48 h. Both digested samples were freeze-dried and resuspended in 10 μl of water, 0.1% TFA prior to mass spectrometry analysis.

MALDI-TOF Mass Spectrometry Analysis—Mass spectra were acquired on a Voyager DE-Pro MALDI-TOF instrument equipped with a 337-nm nitrogen laser. Mass spectra were performed in the reflector delayed extraction mode using 2,5-dihydroxybenzoic acid. This matrix, freshly dissolved at 5 $mg \cdot ml^{-1}$ in 70:30 acetonitrile, 0.1% TFA, was mixed with the water solubilized oligosaccharides in a ratio 1:1 (v/v). These spectra were recorded in a positive mode, using an acceleration voltage of 20,000 V with a delay time of 100 ns. They were smoothed once and externally calibrated using commercially available mixtures of peptides and proteins. In this study, the spectra were externally calibrated using des-Arg^1-bradykinin (904.4681 Da), angiotensin I (1,296.6853 Da), Glu^1-fibrinopeptide B (1,570.6774 Da), and $ACTH_{18-39}$ (2,465.1989 Da). Laser shots were accumulated for each spectrum, at least 5000 laser shots. The mass accuracy obtained is 0.011% on average, which is in agreement with the specifications of the instrument used in this study.

RESULTS

In Silico Analysis of the P. tricornutum Genome Revealed a Set of Genes Encoding for Proteins Involved in the N-Glycosylation Pathway—In eukaryotes, the *N*-glycan biosynthetic pathway can be divided into three steps: (i) the synthesis of

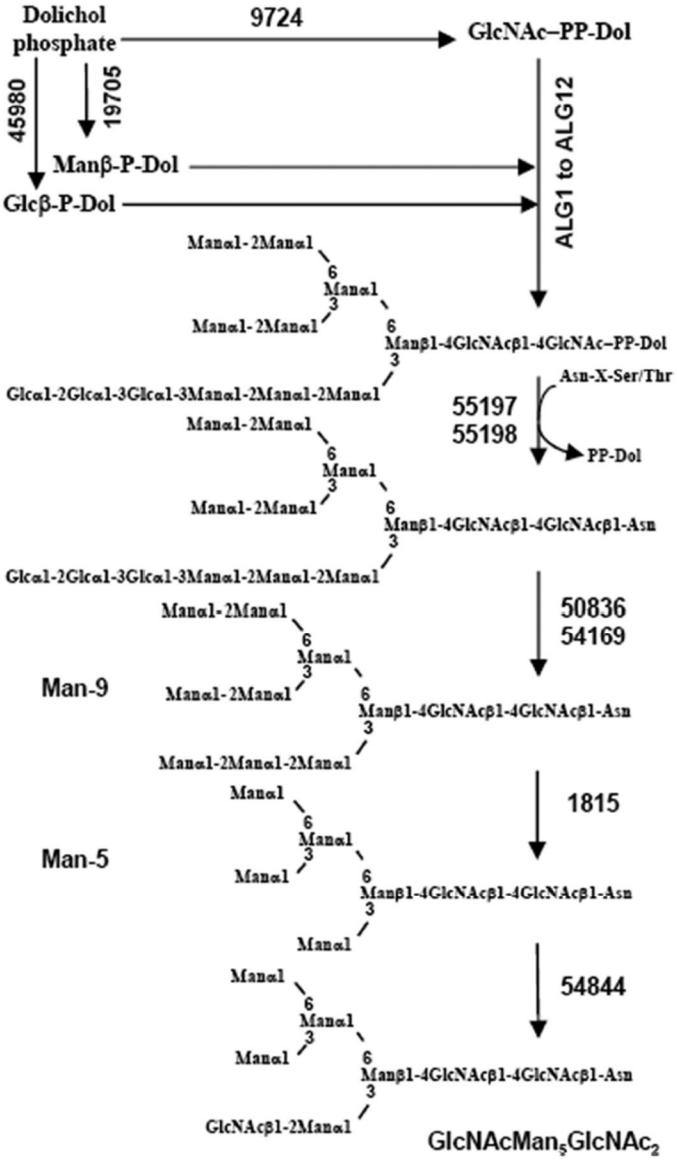


FIGURE 1. ***N*-Glycosylation pathway in *P. tricornutum* based on bioinformatic analysis of the genome database.** Sequences of the *N*-glycosylation pathway identified in the *P. tricornutum* genome are numbered in *bold type*. *Glc*, glucose; *PP-Dol*, dolicholpyrophosphate; *Man*, mannose; *GlcNAc*, *N*-acetylglucosamine; *ALG*, asparagine linked glycosylation.

the $Glc_3Man_9GlcNAc_2$-PP-dolichol precursor and its transfer by the OST onto asparagine residues of nascent polypeptides entering the lumen of the rough ER; (ii) deglucosylation/reglucosylation of the precursor *N*-glycan in the ER, allowing the interaction with chaperones responsible for proper folding and oligomerization; and finally (iii) maturation in the Golgi apparatus of the high mannose type *N*-linked oligosaccharides into complex type *N*-glycans. Based on sequence homologies, we identified in the genome of *P. tricornutum* a set of putative sequences that are likely involved in the different steps of the *N*-glycan biosynthesis and maturation (Fig. 1 and Table 1). Most of these identified genes have expressed sequence tag support (Fig. 1 and Table 1).

All of the genes encoding for enzymes involved in the biosynthesis of dolichol pyrophosphate-linked oligosaccharide on the cytosolic face and in the lumen of the ER were identified in the genome of *P. tricornutum* (Fig. 1 and Table 1). The sequences and topologies of the predicted proteins are highly similar to the corresponding asparagine-linked glycosylation (ALG) orthologs described in other eukaryotes (42), except for ALG 10, for which a *P. tricornutum* candidate sequence was not clearly identified. Putative transferases, which enabled catalyzation of the formation of dolichol-activated mannose and glucose, were also found. Those two activated sugars are required for the elongation steps arising in the ER lumen. In addition to sequences involved in the biosynthesis of the dolichol pyrophosphate-linked oligosaccharide, two putative genes encoding for orthologs of the STT3 catalytic subunit of OST multisubunit complex were identified (Table 1). These multi-spanned sequences, sharing 34 and 37%, respectively, of identities with *A. thaliana* and *H. sapiens* STT3 subunits, contain the conserved WWDYG domain required for the STT3 transferase activity (43).

Genes encoding for polypeptides involved in the quality control of proteins in the ER were also found in the *P. tricornutum* genome. Indeed, α and β subunits of α-glucosidase II were identified. The α subunit contains the characteristic DMNE sequence (44) and a C-type lectin domain involved in mannose binding (45). A putative UDP-glucose: glycoprotein glucosyltransferase and a calreticulin, two molecules ensuring the quality control of the glycoproteins in the ER, are also predicted. Calreticulin is a soluble Ca^{2+}-binding protein of the ER lumen involved in the retention of incorrectly or incompletely folded proteins. Putative *P. tricornutum* calreticulin exhibits more than 50% of identity with orthologs from *N. plumbaginifolia* (56%) and *A. thaliana* (53%). Structurally, the *P. tricornutum* calreticulin contains the three specific domains required for its biological function: a N-terminal domain of ~180 amino acids, a central domain of ~70 residues containing three repeats of an acidic 17-amino acid motif, and a C-terminal domain rich in acidic and lysine residues, both responsible for Ca^{2+} binding (46). *P. tricornutum* calreticulin also exhibited a predicted signal peptide and a C-terminal YDEF tetrapeptide that may ensure its retention in the ER as HDEL, KDEL, or YDEL signals that are known to play this function in higher eukaryotes (47–49).

In regard to Golgi enzymes involved in *N*-glycan biosynthesis, *P. tricornutum* genome contains two sequences encoding for proteins that belong to the glycosylhydrolase family GH47 that catalyze the hydrolysis of the terminal α(1,2)-mannose residues of high mannose type *N*-glycans. The first predicted sequence (sequence 52346) encodes for a protein sharing 32 and 30%, respectively, of identity with MNS4 and MNS5 from *A. thaliana*, which were characterized as being ER degradation-enhancing mannosidases (50). The second sequence (Fig. 1 and Table 1) encodes for a protein sharing 35 and 34%, respectively, of identity with MNS1 and MNS2, two *A. thaliana* α-Man I located in the Golgi apparatus and able to perform the trimming of Man-9 into Man-5 (50, 51). Furthermore, this putative *P. tricornutum* mannosidase exhibits the three conserved catalytic motifs of α-Man I, the threonine residue of the motif III, and the two cysteine residues (Cys-301 and Cys-333) essential for the

TABLE 1

References of predicted proteins involved in the *N*-glycan biosynthesis and quality control of secreted proteins in *P. tricornutum*

SP/SA, signal peptide or signal anchor; EST, expressed sequence tag; TMD, transmembrane domain; Y, yes; N, no; GlcNAcT, *N*-acetylglucosaminyltransferase; ManT, mannosyltransferase; GlcT, glucosyltransferase; P-Dol GlcT, dolichol-phosphate β-glucosyltransferase; P-Dol ManT, dolichol-phosphate mannosyltransferase; α-Glc II, α-glucosidase II; α-Man I, α-mannosidase I; GnT I, *N*-acetylglucosaminyltransferase I; α-Man II, α-mannosidase II.

Protein number	Gene location	Predicted protein function	Length	EST	SP/SA	TMD	Pfam
9724	chr_2:964854–966112	GlcNAc-P-transferase ALG 7	440	N	Y	9	PF00953
9427	chr_1:864467–864835	β(1,4)-GlcNAcT ALG 13	170	N	Y	0	PF04101
14444	chr_15:95429–95971	β(1,4)-GlcNAcT ALG 14	180	Y	N	2	PF08660
14002	chr_13:187936–189201	β(1,4)-ManT ALG 1	448	Y	N	2	PF00534
22554	chr_18:299673–301001	α(1,3)-ManT ALG 2	503	Y	N	1	PF00534
54621	chr_11:102045–103589	α(1,2)-ManT ALG 11	433	Y	Y	4	PF00534
10976	chr_4:589265–590525	α(1,3)-ManT ALG 3	414	Y	Y	9	PF05208
44574	chr_4:944634–947670	α(1,2)-ManT ALG 9	556	Y	N	10	PF03901
44425	chr_4:509561–511432	α(1,2)-ManT ALG 12	581	Y	Y	6	PF03901
44117	chr_3:1027730–1029328	α(1,3)-GlcT ALG 6	532	Y	Y	11	PF03155
44905	chr_5:604679–606258	α(1,3)-GlcT ALG 8	436	Y	Y	9	PF03155
45980	chr_8:845268–846314	P-Dol GlcT ALG 5	348	Y	Y	1	PF00535
19705	chr_6:554632–555416	P-Dol ManT (DPM1)	236	Y	N	0	PF00535
55197	chr_30:46605–50076	OST (STT3 subunit)	911	Y	Y	10	PF02516
55198	chr_30:50287–53582	OST (STT3 subunit)	894	Y	Y	10	PF02516
50836	chr_5:359528–361787	α-Glc II, subunit α	712	Y	N	0	PF01055
54169	chr_3:489007–491615	α-Glc II, subunit β	803	N	Y	0	PF07915
41172	chr_28:129840–131182	Calreticulin	421	Y	Y	0	PF00262
54787	chr_14:637216–637863	UDP-glucose: glycoprotein glucosyltransferase	499	Y	N	0	PF06427
1815	bd_32 × 35: 113900–1161195	α-Man I	666	Y	Y	1	PF01532
54844	chr_16:89066–90505	GnT I	444	Y	Y	1	PF03071
52248	chr_14:60509–61027	α-Man II	1498	Y	N	1	PF01074 PF09261 PF07748
54599	chr_10:692088–693836	α(1,3)-FucT	481	Y	Y	1	PF00852
46109	chr_9:288291–290659	α(1,3)-FucT	770	Y	Y	1	PF00852
46110	chr_9:291303–293777	α(1,3)-FucT	718	Y	Y	1	PF00852

mannosidase activity (51, 52). A signal anchor is also predicted in the N-terminal part of the protein as required for a type II transmembrane protein (Table 1). Moreover, there is some expressed sequence tag supporting an expression for this enzyme in *P. triconutum* (Table 1).

In addition, one putative GnT I and one putative α-Man II were also identified in *P. tricornutum* genome (Table 1). These enzymes are involved in the *N*-glycan maturation into complex oligosaccharides by transferring a terminal GlcNAc on the α(1,3)-mannose arm of Man-5 and then removing the two mannose residues located on the α(1,6)-mannose arm. The putative GnT I sequence is predicted to be a typical type II membrane protein. Its cytoplasmic tail contains three basic amino acids that could promote ER exit as demonstrated for *N. tabacum* GnT I (53). This sequence also possesses a luminal part sharing 37% of identity with the rabbit GnT I (Fig. 2). From the crystal structure of this mammalian transferase, 22 amino acid residues in the catalytic domain were shown to form direct or water-mediated interactions with the UDP-GlcNAc nucleotide sugar and the Mn^{2+} ion (54, 55). Fourteen of these residues are strictly identical in the *P. tricornutum* GnT I, whereas the other residues are closely conserved (Fig. 2). Moreover, the SQD motif, which has been demonstrated to be important because this motif interacts with the uracil ring of the donor substrate, is also present in the *P. tricornutum* GnT I sequence (54). This conserved motif is present in all GnT I characterized so far (58). The EDD motif that is a variation of the canonical acidic metal-binding D*X*D motif is conserved in the *P. tricornutum* sequence. This motif has been demonstrated to have critical interaction with UDP-GlcNAc and the metal ion in rabbit GnT I (55). The predicted α-Man II consisted of a large protein containing the three Pfam domains of CAZy GH38 glycosylhydrolases and the conserved residues involved in Zn^{2+} binding in the catalytic site of *D. melanogaster* α-Man II (59, 60).

In eukaryotes, α(1,3)- and α(1,6)-fucosyltransferases transfer fucose residues onto the proximal GlcNAc unit of the *N*-linked glycan core. *In silico* analysis of the *P. tricornutum* genome revealed the presence of three genes encoding for putative fucosyltransferases (FucT). These candidates exhibit the appropriate type II membrane protein topology (Table 1) and 23% (sequence 46109), 28% (sequence 46110), and 25% (sequence 54559), respectively, of identity with *A. thaliana* FucTA. These FucT candidates exhibited the motifs I and II of α(1,3)-FucT (61, 62), the SNC(G/A)A(R/H)N sequence, specific for plant and *D. melanogaster* α(1,3)-FucT (63–65), as well as the C*XX*C motif located at the C-terminal sequence involved in the formation of disulfide bonds (66). A putative type II xylosyltransferase is also predicted in the genome. This sequence shares 24% of identity with the luminal part of *A. thaliana* β(1,2)-xylosyltransferase involved in the transfer of a β(1,2)-xylose residue onto the β-Man of the *N*-glycan core (67, 68). Nevertheless, in the absence of reported motifs specific for β(1,2)-xylosyltransferase activity, the involvement of this putative transferase in *P. tricornutum N*-glycan pathway remains highly hypothetical. Searches for sequences encoding other *N*-glycan-maturating transferases, such as *N*-acetylglucosaminyltransferases ranging from GnT II to GnT VI that allow the formation of polyantennary *N*-glycans or sialyltransferases, did not reveal any ortholog in the *P. tricornutum* genome.

P. tricornutum Proteins Mainly Carry High Mannose Type N-Glycans—Analysis of glycans *N*-linked to *P. tricornutum* proteins was first investigated by Western blot on a total pro-

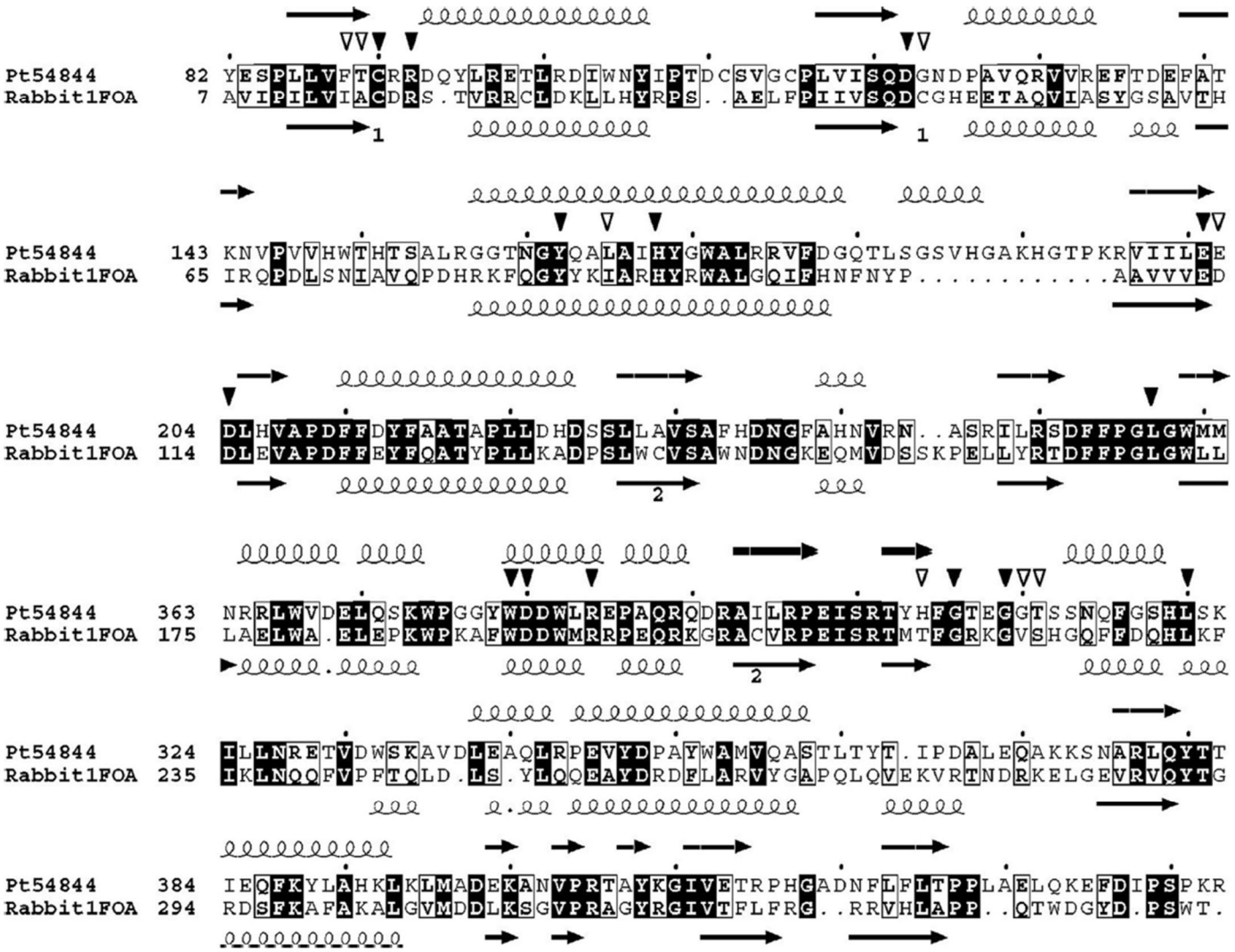


FIGURE 2. **Catalytic amino acids are very conserved in the putative GnT I protein from *P. tricornutum*.** Protein sequences alignment between rabbit (1FOA) and *P. tricornutum*, as proposed by the Swiss-Pdb viewer program (56). Secondary structural elements are represented above the alignment for the *P. tricornutum* GnT I and below the alignment for the rabbit GnT I with a *bold right arrow* as the β strand and a *looped line* as the α helix. Essential residues for the binding of the donor substrate (UDP-GlcNAc) are indicated by *arrowheads* above the alignment: in *black* when identical and in *white* when not. Rabbit GnT I disulfide bridges are also numbered. The figure was created with the Espript program (57).

tein extract using probes specific for glycan epitopes. As illustrated in Fig. 3*A*, some *P. tricornutum* proteins were affinodetected by concanavalin A, a lectin specific for high mannose sequences (69). This affinodetection was largely suppressed upon treatment with Endo H or PNGase F (Fig. 3*B*), two enzymes able to cleave *N*-glycans. In contrast, affinodetections with *E. crista galli* agglutinin, peanut agglutinin, or phytohemagglutinin E and L, four lectins specific for the lactosamine motif, the galβ1,3-GalNAc epitope, the bisected di-triantennary complex type *N*-glycans, and tri- and tetraantennary complex type *N*-glycans, respectively, did not show anything (data not shown). Immunodetections with antibodies raised against plant glycoepitopes (32) were then carried out (Fig. 3*A*). Antibodies specific for xylose epitope β(1,2)-linked to the core $Man_3GlcNAc_2$ did not detect proteins from *P. tricornutum*, whereas proteins from green onion used as positive control were labeled (Fig. 3*A*). In contrast, total proteins were weakly immunodetected using core-α(1,3)-fucose antibodies (Fig. 3*A*). These signals disappeared after oxidation of the proteins with sodium periodate, revealing that antibody recognition involved a glycan sequence (data not shown).

To investigate the presence in *P. tricornutum* proteins of complex glycans carrying terminal GlcNAc, we treated the protein extract with a β(1,4)-galactosyltransferase, an enzyme able to transfer a galactose residue onto terminal GlcNAc residues, and then analyzed by affinoblotting the resulting protein preparation with RCA 120, a lectin that binds specifically to Galβ1–4GlcNAc sequences (34). In contrast to a plant-derived IgG used as a positive control of galactose transfer, no signal was detected in *P. tricornutum* sample after this treatment (Fig. 3*C*), thus indicating that this diatom does not exhibit terminal GlcNAc onto its proteins at detectable level (0.5 μg). Moreover, the presence of *N*-acetylneuraminic acid (Neu5Ac), the main sialic acid found in mammals, was investigated by coupling to DMB (37) and analysis of the resulting DMB derivatives by liquid chromatography. Although low intensity peaks were detected by fluorescence, none of them co-migrated with a standard of DMB-Neu5Ac (data not shown).

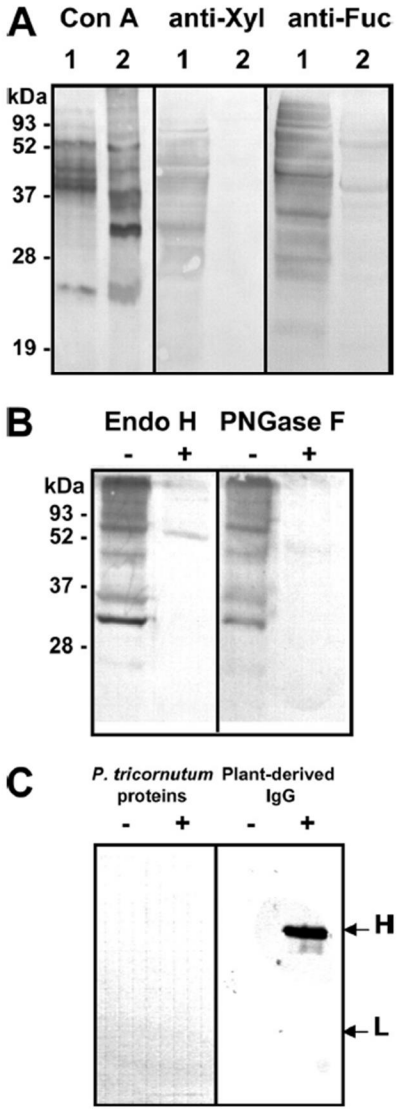


FIGURE 3. ***P. tricornutum* glycoproteins harbor *N*-linked oligosaccharides.** *A*, affinodetection using concanavalin A (*Con A*) and immunodetection using antibodies raised against the core β(1,2)-xylose (anti-Xyl) and core α(1,3)-fucose (anti-Fuc) epitopes of proteins isolated from green onion used as a positive control (*lanes 1*) and from *P. tricornutum* (*lanes 2*). *B*, affinodetection by concanavalin A of proteins extracted from *P. tricornutum* treated (+) or not (−) with Endo H and PNGase F. *C*, affinodetection with RCA 120 of *P. tricornutum* proteins treated (+) or not (−) with bovine β(1,4)-galactosyltransferase. Plant-derived IgG was used as a positive control of the galactose transfer efficiency (34). The *arrows* indicate the migration of heavy (*H*) and light (*L*) chains.

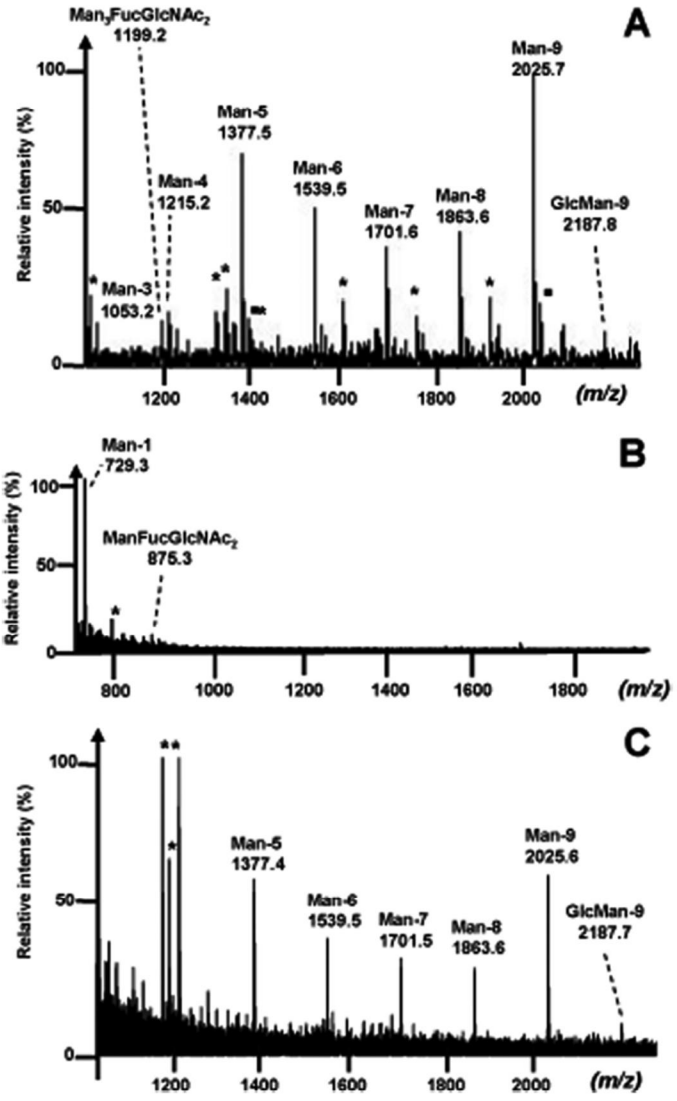


FIGURE 4. **High mannose type *N*-glycans are the main oligosaccharides *N*-linked to *P. tricornutum* proteins.** *A*, MALDI-TOF mass spectrum of *N*-linked glycans released by PNGase A from glycoproteins of *P. tricornutum* and labeled with 2-AB. *B*, MALDI-TOF mass spectrum of the pool of *N*-glycans after treatment with jack bean α-mannosidase. *C*, MALDI-TOF mass spectrum of 2-AB-labeled *N*-linked glycans released by PNGase F from glycoproteins of *P. tricornutum*. Man-3 to Man-9 are the paucimannose and high mannose type *N*-glycans $Man_3GlcNAc_2$ to $Man_9GlcNAc_2$. *, contaminants; ■, potassium adducts.

To investigate their detailed *N*-glycan structures, *N*-glycans were released from proteins by PNGase A treatment (31). The resulting *N*-glycans were then coupled to 2-AB (40, 41) to facilitate their detection and analysis by MALDI-TOF mass spectrometry. As illustrated in Fig. 4*A*, major ions correspond to $(M + Na)^+$ adducts of 2-AB derivatives of $Hexose_{5-9}GlcNAc_2$. Other minor ions were also detected in the mass spectrum profile with *m*/*z* values corresponding to $Hexose_{3-4}GlcNAc_2$ oligosaccharides. Moreover, the ion at *m*/*z* 1199 was assigned to $Hexose_3DeoxyhexoseGlcNAc_2$ *N*-linked glycan. The pool of glycans was then submitted to exoglycosidase digestions. The oligosaccharide mixture was converted to $HexoseGlcNAc_2$ and $HexoseDeoxyHexoseGlcNAc_2$ upon a treatment with jack bean α-mannosidase (Fig. 4*B*), demonstrating the presence of α-linked mannose residues in PNGase A-released *N*-glycans. Furthermore, a treatment of the sample with α-L-fucosidase resulted in the suppression of ion at *m*/*z* 1199 (not shown). As a consequence, main ions detected in MALDI-TOF mass spectrum (Fig. 4*A*) were assigned to high mannose *N*-glycans ranging from Man-5 to Man-9, and minor ions were assigned to Man-3, Man-4, and $Man_3FucGlcNAc_2$. To investigate the location of the fucose residue onto the core *N*-glycan in this later *N*-glycan, proteins were submitted to a deglycosylation experiment by PNGase F, a deglycosylating enzyme that is not able to cleave *N*-linked oligosaccharides harboring a fucose α(1,3)-linked to the proximal glucosamine residue (38). The ion assigned to $Man_3FucGlcNAc_2$ was not observed in the mass spectrum,

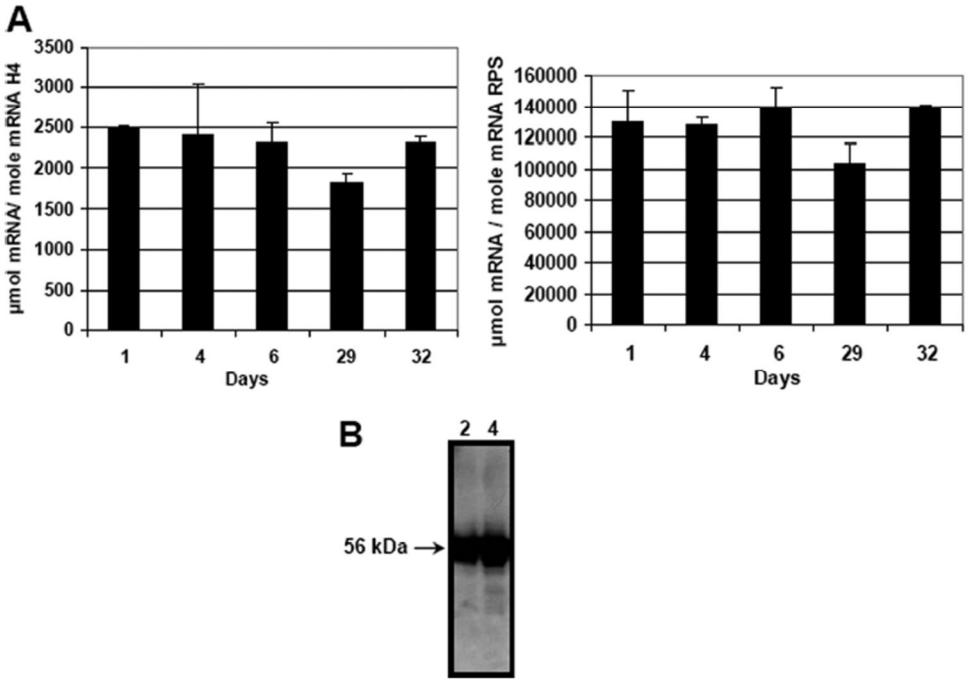


FIGURE 5. **Expression of the GnT I gene in *P. tricornutum* and in CHO Lec1 mutant.** *A*, expression of the GnT I gene of *P. tricornutum* by real time quantitative PCR over a 1-month period. The relative gene expression was normalized to two reference genes encoding for histone H4 and ribosomal protein small subunit 30 S. *B*, Western blot analysis using anti-V5 antibodies of proteins isolated from two transformants (*lanes 2* and 4) of CHO Lec1 expressing the *P. tricornutum* GnT I fused to a V5 tag. GnT I was detected at the expected molecular mass of 56 kDa.

indicating that this glycan carries a core α(1,3)-fucose residue (Fig. 4*C*).

Putative P. tricornutum GnT I Gene Is Able to Complement the Deficiency in N-Glycan Maturation in CHO Lec1 Mutant—Bio-informatic analysis of *P. tricornutum* genome revealed a gene potentially encoding for a GnT I glycosyltransferase. The expression of this gene was monitored in a continuous culture over a 1-month period by real time quantitative PCR. The relative gene expression was normalized to two reference genes encoding for ribosomal protein small subunit 30 S and histone H4 that have been recently described as appropriate housekeeping genes for real time quantitative PCR in *P. tricornutum* (26). The results show that *P. tricornutum* GnT I is steadily expressed in standard culture conditions over 32 days (Fig. 5*A*).

To demonstrate that this putative transferase is able to transfer *in vivo* a GlcNAc residue onto the proteins carrying Man-5 *N*-linked glycans, the *P. tricornutum* GnT I (GenBank™ accession number 1370344) was expressed in fusion with a V5 tag into CHO Lec1 mutant deficient in its endogenous GnT I (16). On the basis of the immunodetection of the V5 epitope, CHO transformants were found to efficiently express the recombinant GnT I. Two transformants (2 and 4) were selected for *N*-linked glycan analysis (Fig. 5*B*). Proteins from these two clones, from the wild type CHO and from the CHO Lec1 mutant, were isolated, and their *N*-linked glycans were released by treatment with PNGase F. As illustrated in Fig. 6*A*, MALDI-TOF mass spectrum of *N*-glycans released from the CHO Lec1 mutant showed that it accumulates high mannose type *N*-glycans, contrasting with the wild type CHO cell exhibiting both high mannose and complex type *N*-glycans (Fig. 6*B*). In Fig. 6*C*, glycoproteins from the complemented line (clone 4) displayed both high mannose oligosaccharides and a complete set of complex type *N*-glycans identical to the one observed in wild type CHO cells (Fig. 6*B*). The same *N*-glycan profile was obtained with clone 2 (data not shown). This demonstrates that the expression of the *P. tricornutum* GnT I was able to restore the biosynthesis of complex *N*-glycans in the mammalian cell mutant.

DISCUSSION

Bio-informatic analysis of the *P. tricornutum* genome revealed the presence of a completed set of sequences potentially encoding for proteins involved in the synthesis of the lipid-linked $Glc_3Man_9GlcNAc_2$-PP-dolichol *N*-glycan, the subunits of the OST complex that catalyze its transfer onto the asparagine residues of target proteins, as well as ER glucosidases and chaperones (6, 7). This suggests that this diatom possesses the ER machinery required for glycoprotein quality control previously characterized for other eukaryotes (71). The genome analysis also revealed the presence of one Golgi mannosidase I involved in the trimming of the *N*-glycan precursor into Man-5 high mannose type *N*-glycan (50, 51). Consistent with these sequence predictions, two lines of biochemical evidence strongly suggest that proteins from the *P. tricornutum* Pt 1.8.6 strain are mainly *N*-glycosylated by high mannose type oligosaccharides. First, proteins from this strain are affinodetected by concanavalin A, a lectin specific for high mannose sequences. This detection is largely suppressed upon treatment with Endo H or PNGase F, two deglycosylating enzymes specific for *N*-linked glycans. Furthermore, affino- and immunodetection with other glycan-specific probes, as

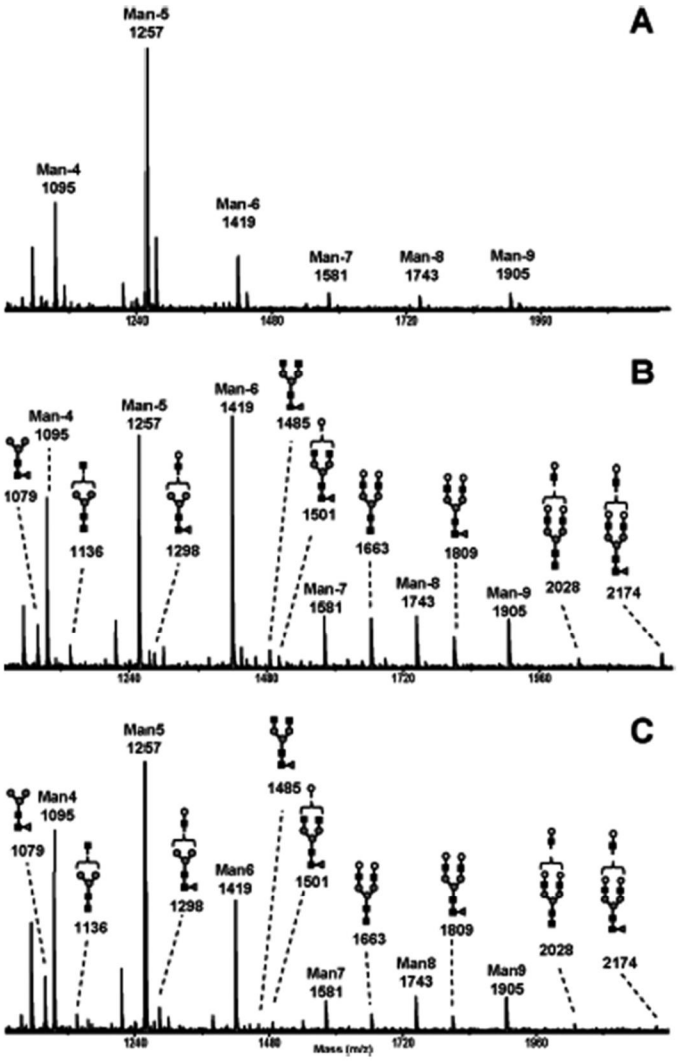


FIGURE 6. ***P. tricornutum* GnT I complements *N*-glycan maturation deficiency in CHO Lec1 mutant.** MALDI-TOF mass spectra of glycans *N*-linked to proteins extracted from CHO cells. *A*, CHO Lec1 mutant; *B*, CHO wild type; and *C*, transformant 4 of CHO Lec1 mutant complemented with *P. tricornutum* GnT I gene. Man-4 to Man-9 are the high mannose type *N*-glycans $Man_4GlcNAc_2$ to $Man_9GlcNAc_2$ 70). *Black square*, GlcNAc; *gray circle*, Man; *white circle*, Gal; *gray triangle*, fucose.

well as a search for sialic acids, were unsuccessful. Second, MALDI-TOF mass spectrometry of the *N*-glycan population released by PNGase A allowed the detection of $Hexose_{5-9}GlcNAc_2$ oligosaccharides, sensitive to an α-mannosidase treatment, as major oligosaccharide species. We concluded that proteins from the *P. tricornutum* mainly carry Man-5 to Man-9 high mannose type *N*-glycans. Other minor glycan species *i.e.* Man-3, Man-4, and $Man_3FucGlcNAc_2$ carrying a fucose α(1,3)-linked to the proximal glucosamine residue were also detected. The presence of this later sequence corroborates the weak detection of core α(1,3)-fucose epitopes on Western blot, potential product of the putative FucT sequences predicted in the *P. tricornutum* genome.

In eukaryotes, the first steps of the *N*-glycan processing into complex *N*-glycans are controlled by GnT I, α-Man II, and GnT II. The resulting core *N*-glycan is modified by the action of a wide variety of glycosyltransferases giving rise to mature *N*-linked glycans involved in various biological processes (8–11). Putative GnT I and α-Man II are predicted in the *P. tricornutum* genome. We mainly focused on GnT I characterization because this transferase is the first enzyme initiating the complex type maturation of oligosaccharides *N*-linked to secreted proteins. In standard culture conditions, we demonstrated that the gene encoding this putative transferase is expressed. However, no glycan carrying terminal GlcNAc residues has been detected on *P. tricornutum* proteins by either a galactosyltransferase assay or by MALDI-TOF mass spectrometry analysis of the PNGase A-released oligosaccharides. Therefore, the *in vivo* activity of this putative GnT I was investigated by expressing the full-length protein in CHO Lec1 mutant lacking its endogenous GnT I activity (16). Wild type-like *N*-glycosylation profiles were detected in transformed cell lines, thus demonstrating that the putative GnT I from this diatom was able to restore the biosynthesis of complex type *N*-glycans in GnT I-null CHO cells. This shows that the *P. tricornutum* gene encodes for a protein able to perform *in vivo* the processing of oligomannosides into complex type *N*-glycans, thus corresponding to a GnT I activity. To our knowledge, this work is the first functional characterization of a microalgal *N*-glycan glycosyltransferase. These data also suggest that both the targeting and the retention mechanism of Golgi enzymes are conserved between mammals and diatoms.

A search for putative GnT I was carried out in genomes of species belonging to the three main microalgae lineages, *i.e. Viridiplantae*, *Heteroconta*, and *Haptophyta* of the phylogenetic tree. Genes encoding for putative GnT I were identified in two other heterokonts, *i.e. F. cylindrus* and *T. pseudonana*, as well as in the haptophyte *E. huxleyi.* Search for GnT I in *Viridiplantae* only revealed one sequence in *M. pusilla* and not in other species (Fig. 7). The microalgal GnT I complete sequences share ~25% of identity with plant and mammal GnT I. *Heteroconta* and *Haptophyta* GnT I are gathered in a distinct lineage that could be clearly separated from other GnT I as seen in the phylogenetic tree (Fig. 7). So far, the processing of *N*-linked glycans into complex oligosaccharide has been mainly described in multicellular higher eukaryotes such as animals and land plants and has been demonstrated to be required for normal morphogenesis in animals (9, 10). Our results show that this key Golgi transferase is also involved in the processing of *N*-linked glycans in unicellular microalgae species.

Mainly high mannose type *N*-glycans were detected on *P. tricornutum* proteins, which suggests that GnT I possesses a limited *in vivo* impact on glycans *N*-linked to secreted proteins of this diatom. No glycan carrying terminal GlcNAc residue has been detected onto *P. tricornutum* proteins. However, although representing a minor glycan population, small size *N*-glycans Man-3 and Man-4, as well as $Man_3FucGlcNAc_2$, were detected in the *P. tricornutum* *N*-glycan mass profile. These glycans, named paucimannose structures, have been previously found in invertebrates and plants and result from the degradation of GlcNAc-terminated complex glycans ($GlcNAcMan_{3-4}GlcNAc_2$) by *N*-acetylglucosaminidases after their biosynthesis in the Golgi apparatus. Indeed, in

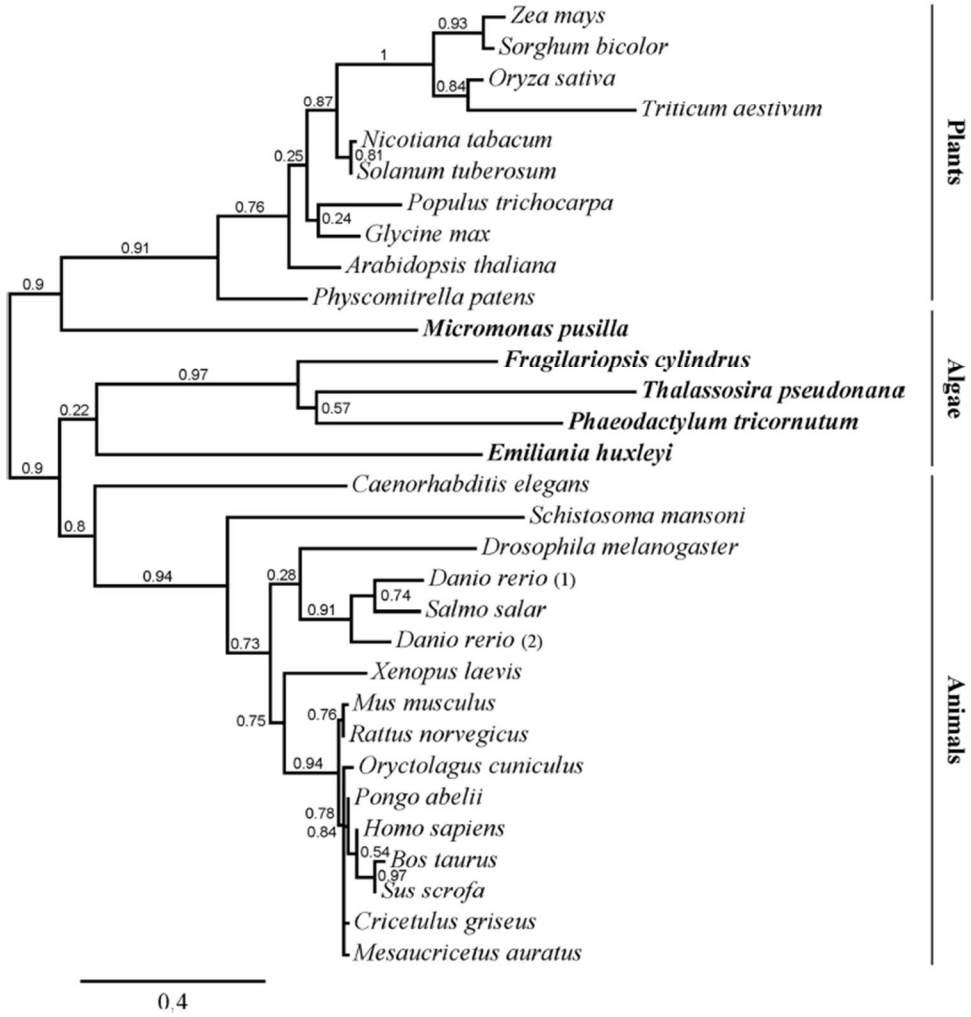


FIGURE 7. **GnT I are predicted in other microalgae.** Phylogenetic tree of GnT I from algae, plants, and animals based on the maximum likelihood method. The scale bar (0.4) represents the number of amino acid residue substitutions per site. Microalgal GnT I sequences are indicated in bold type. Accession numbers for the different reference sequences are indicated under "Experimental Procedures."

the GnT I-dependent pathway, $GlcNAcMan_5GlcNAc_2$, the product of GnT I (Fig. 1), is successively converted in the Golgi apparatus into $GlcNAcMan_4GlcNAc_2$ and then into $GlcNAcMan_3GlcNAc_2$ by action of the α-Man II and then into $GlcNAcMan_3FucGlcNAc_2$ by α-FucT. Elimination of terminal GlcNAc by β-*N*-acetylglucosaminidases in the secretory system or in compartments where proteins accumulate can then degrade these oligosaccharides into Man-3, Man-4, and $Man_3FucGlcNAc_2$. Such a processing was demonstrated in insect cells (72), *C. elegans* (73), and plants (74). Two putative processing β-*N*-acetylglucosaminidases belonging to the CAZy glycosylhydrolase GH20 family are predicted in the *P. tricornutum* genome (sequences 49563 and 45073). These glycosidases share 43 and 36%, respectively, of identity with DmFDL, a β-*N*-acetylglucosaminidase from *Drosophila* that is able to specifically hydrolyze a GlcNAc residue located onto the α(1,3)-antenna of *N*-glycans giving rise to paucimannose oligosaccharides (75, 76). Taken together, these data suggest that such processing may also occur in diatoms.

The biochemical characterization of the core-modifying FucTs and of the processing β-*N*-acetylglucosaminidases are currently studied. The *N*-glycosylation patterns of proteins from *P. tricornutum* grown in different conditions are currently under investigation to study this major post-translational modification in relation to pleomorphism and/or stress environmental conditions and the function of complex *N*-glycans in the diatom physiology.

Acknowledgments—We thank C. Bowler (ENS, Paris) and the Joint Genome Institute for giving priority access to the P. tricornutum genome. Authors also thank S. Calbo (INSERM U905, University of Rouen), C. Loutelier-Bourhis (CNRS UMR614, University of Rouen), C. Rihouey (CNRS UMR 6522, University of Rouen), and R. Kaas (Laboratoire Physiologie et Biotechnologies des Algues, Nantes, Institut Français de Recherche pour l'Exploitation de la Mer) for technical help and advice as well as Margot Lerouge for help with producing figures. We thank J.-C. Mollet, A. Driouich, and S. Calbo (University of Rouen) for critical readings of our manuscript.

REFERENCES

1. Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A., and Queguiner, B., (1995) *Globalal. Biogeochem. Cycles* **9**, 359–372
2. Raven, J. A., and Waite, A. M. (2004) *New Phytologist* **162**, 45–61

3. Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J. P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kröger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Parker, M. S., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van de Peer, Y., and Grigoriev, I. V. (2008) *Nature* **456,** 239–244
4. Falciatore, A., and Bowler, C. (2002) *Annu. Rev. Plant Biol.* **53,** 109–130
5. Scala, S., Carels, N., Falciatore, A., Chiusano, M. L., and Bowler, C. (2002) *Plant Physiol.* **129,** 993–1002
6. Burda, P., and Aebi, M. (1999) *Biochim. Biophys. Acta* **1426,** 239–257
7. Helenius, A., and Aebi, M. (2001) *Science* **291,** 2364–2369
8. Schachter, H. (2009) *Carbohydr. Res.* **344,** 1391–1396
9. Ioffe, E., and Stanley, P. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91,** 728–732
10. Metzler, M., Gertz, A., Sarkar, M., Schachter, H., Schrader, J. W., and Marth, J. D. (1994) *EMBO J.* **13,** 2056–2065
11. Sarkar, M., Leventis, P. A., Silvescu, C. I., Reinhold, V. N., Schachter, H., and Boulianne, G. L. (2006) *J. Biol. Chem.* **281,** 12776–12785
12. von Schaewen, A., Sturm, A., O'Neill, J., and Chrispeels, M. J. (1993) *Plant Physiol.* **102,** 1109–1118
13. Kang, J. S., Frank, J., Kang, C. H., Kajiura, H., Vikram, M., Ueda, A., Kim, S., Bahk, J. D., Triplett, B., Fujiyama, K., Lee, S. Y., von Schaewen, A., and Koiwa, H. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 5933–5938
14. von Schaewen, A., Frank, J., and Koiwa, H. (2008) *Plant Signal Behav.* **3,** 871–873
15. Frank, J., Kaulfürst-Soboll, H., Rips, S., Koiwa, H., and von Schaewen, A. (2008) *Plant Physiol.* **148,** 1354–1367
16. Stanley, P., Narasimhan, S., Siminovitch, L., and Schachter, H. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72,** 3323–3327
17. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008) *Nucleic Acids Res.* **36,** W465–W469
18. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
19. Castresana, J. (2000) *Mol. Biol. Evol.* **17,** 540–552
20. Guindon, S., and Gascuel, O. (2003) *Syst. Biol.* **52,** 696–704
21. Anisimova, M., and Gascuel, O. (2006) *Syst. Biol.* **55,** 539–552
22. Chevenet, F., Brun, C., Bañuls, A. L., Jacq, B., and Christen, R. (2006) *BMC Bioinformatics* **7,** 439
23. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucleic Acids Res.* **37,** D233–D238
24. Walne, P. R. (1966) *Fisheries Invest.* **2,** 1–52
25. Bougaran, G., Le Dean, L., Lukomska, E., Kaas, R., and Baron, R. (2003) *Aquat. Living Resour.* **16,** 389–394
26. Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., Manfredonia, A., Falciatore, A., and Bowler, C. (2007) *Gene* **406,** 23–35
27. Chung, C. C., Hwang, S. P., and Chang, J. (2005) *Appl. Environ Microbiol.* **71,** 8744–8751
28. Kang, L. K., Hwang, S. P. L., Lin, H. J., Chen, P. C., and Chang, J. (2009) *J. Phycol.* **45,** 864–872
29. Livak, K. J., and Schmittgen, T. D. (2001) *Methods* **25,** 402–408
30. Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C., and Bowler, C. (1999) *Mar. Biotechnol.* **1,** 239–251
31. Fitchette, A. C., Tran Dinh, O., Faye, L., and Bardor, M. (2007) *Methods Mol. Biol.* **355,** 317–342
32. Faye, L., Gomord, V., Fitchette-Lainé, A. C., and Chrispeels, M. J. (1993) *Anal. Biochem.* **209,** 104–108
33. Fitchette-Lainé, A. C., Denmat, L. A., Lerouge, P., and Faye, L. (1998) *Methods Biotechnol.* **3,** 271–290
34. Bardor, M., Loutelier-Bourhis, C., Paccalet, T., Cosette, P., Fitchette, A. C., Vézina, L. P., Trépanier, S., Dargis, M., Lemieux, R., Lange, C., Faye, L., and Lerouge, P. (2003) *Plant Biotechnol. J.* **1,** 451–462
35. Bardor, M., Nguyen, D. H., Diaz, S., and Varki, A. (2005) *J. Biol. Chem.* **280,** 4228–4237
36. Hara, S., Yamaguchi, M., Takemori, Y., Furuhata, K., Ogura, H., and Nakamura, M. (1989) *Anal. Biochem.* **179,** 162–166
37. Klein, A., Diaz, S., Ferreira, I., Lamblin, G., Roussel, P., and Manzi, A. E. (1997) *Glycobiology* **7,** 421–432
38. Tretter, V., Altmann, F., and März, L. (1991) *Eur. J. Biochem.* **199,** 647–652
39. Bakker, H., Bardor, M., Molthoff, J. W., Gomord, V., Elbers, I., Stevens, L. H., Jordi, W., Lommen, A., Faye, L., Lerouge, P., and Bosch, D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98,** 2899–2904
40. Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M., and Parekh, R. B. (1995) *Anal. Biochem.* **230,** 229–238
41. Séveno, M., Cabrera, G., Triguero, A., Burel, C., Leprince, J., Rihouey, C., Vézina, L. P., D'Aoust, M. A., Rudd, P. M., Royle, L., Dwek, R. A., Harvey, D. J., Lerouge, P., Cremata, J. A., and Bardor, M. (2008) *Anal. Biochem.* **379,** 66–72
42. Weerapana, E., and Imperiali, B. (2006) *Glycobiology* **16,** 91R–101R
43. Yan, Q., and Lennarz, W. J. (2002) *J. Biol. Chem.* **277,** 47692–47700
44. Feng, J., Romaniouk, A. V., Samal, S. K., and Vijay, I. K. (2004) *Glycobiology* **14,** 909–921
45. Wilkinson, B. M., Purswani, J., and Stirling, C. J. (2006) *J. Biol. Chem.* **281,** 6325–6333
46. Kwiatkowski, B. A., Zielińska-Kwiatkowska, A. G., Migdalski, A., Kleczkowski, L. A., and Wasilewska, L. D. (1995) *Gene* **165,** 219–222
47. Matsushima, R., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2003) *Plant J.* **33,** 493–502
48. Pagny, S., Denmat-Ouisse, L. A., Gomord, V., and Faye, L. (2003) *Plant Cell Physiol.* **44,** 173–182
49. Gnanasambandam, A., Polkinghorne, I. G., and Birch, R. G. (2007) *Plant Biotechnol. J.* **5,** 290–296
50. Liebminger, E., Hüttner, S., Vavra, U., Fischl, R., Schoberer, J., Grass, J., Blaukopf, C., Seifert, G. J., Altmann, F., Mach, L., and Strasser, R. (2009) *Plant Cell* **21,** 3850–3867
51. Kajiura, H., Koiwa, H., Nakazawa, Y., Okazawa, A., Kobayashi, A., Seki, T., and Fujiyama, K. (2010) *Glycobiology* **20,** 235–247
52. Tempel, W., Karaveg, K., Liu, Z. J., Rose, J., Wang, B. C., and Moremen, K. W. (2004) *J. Biol. Chem.* **279,** 29774–29786
53. Schoberer, J., Vavra, U., Stadlmann, J., Hawes, C., Mach, L., Steinkellner, H., and Strasser, R. (2009) *Traffic* **10,** 101–115
54. Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J. M. (2000) *EMBO J.* **19,** 5269–5280
55. Gordon, R. D., Sivarajah, P., Satkunarajah, M., Ma, D., Tarling, C. A., Vizitiu, D., Withers, S. G., and Rini, J. M. (2006) *J. Mol. Biol.* **360,** 67–79
56. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18,** 2714–2723
57. Gouet, P., Courcelle, E., Stuart, D. I., and Métoz, F. (1999) *Bioinformatics* **15,** 305–308
58. Strasser, R., Stadlmann, J., Svoboda, B., Altmann, F., Glössl, J., and Mach, L. (2005) *Biochem. J.* **387,** 385–391
59. Strasser, R., Schoberer, J., Jin, C., Glössl, J., Mach, L., and Steinkellner, H. (2006) *Plant J.* **45,** 789–803
60. van den Elsen, J. M., Kuntz, D. A., and Rose, D. R. (2001) *EMBO J.* **20,** 3008–3017
61. Breton, C., and Imberty, A. (1999) *Curr. Opin. Struct. Biol.* **9,** 563–571
62. Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L., and Breton, C. (1999) *Glycobiology* **9,** 323–334
63. Bakker, H., Schijlen, E., de Vries, T., Schiphorst, W. E., Jordi, W., Lommen, A., Bosch, D., and van Die, I. (2001) *FEBS Lett.* **507,** 307–312
64. Wilson, I. B., Rendić, D., Freilinger, A., Dumić, J., Altmann, F., Mucha, J., Müller, S., and Hauser, M. T. (2001) *Biochim. Biophys. Acta* **1527,** 88–96
65. Fabini, G., Freilinger, A., Altmann, F., and Wilson, I. B. (2001) *J. Biol. Chem.* **276,** 28058–28067
66. 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) *J. Biol. Chem.* **275,** 24237–24245
67. Strasser, R., Mucha, J., Mach, L., Altmann, F., Wilson, I. B., Glössl, J., and Steinkellner, H. (2000) *FEBS Lett.* **472,** 105–108
68. Pagny, S., Bouissonnie, F., Sarkar, M., Follet-Gueye, M. L., Driouich, A., Schachter, H., Faye, L., and Gomord, V. (2003) *Plant J.* **33,** 189–203
69. Faye, L., and Chrispeels, M. J. (1985) *Anal. Biochem.* **149,** 218–224
70. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth, J. D., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. (2009) *Proteomics* **9,** 5398–5399
71. Banerjee, S., Vishwanath, P., Cui, J., Kelleher, D. J., Gilmore, R., Robbins, P. W., and Samuelson, J. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 11676–11681
72. Altmann, F., Schwihla, H., Staudacher, E., Glössl, J., and März, L. (1995) *J. Biol. Chem.* **270,** 17344–17349
73. Zhang, W., Cao, P., Chen, S., Spence, A. M., Zhu, S., Staudacher, E., and Schachter, H. (2003) *Biochem. J.* **372,** 53–64
74. Vitale, A., and Chrispeels, M. J. (1984) *J. Cell Biol.* **99,** 133–140
75. Léonard, R., Rendic, D., Rabouille, C., Wilson, I. B., Préat, T., and Altmann, F. (2006) *J. Biol. Chem.* **281,** 4867–4875
76. Geisler, C., Aumiller, J. J., and Jarvis, D. L. (2008) *J. Biol. Chem.* **283,** 11330–11339

II- Résultats complémentaires.

II-1- Etude de l'expression du gène de la GnT I en fonction des conditions de culture, du morphotype et de la souche.

L'étude des données transcriptomiques disponibles pour *P. tricornutum* suggèrent l'existence d'une expression différentielle du gène codant pour la GnT I, selon les conditions de culture, le morphotype et la souche de *P. tricornutum* (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html, Maheswari et al., 2005). En effet, parmi 16 conditions testées, des EST (Expressed Sequence Tag) correspondant à ce gène ont été uniquement identifiées dans la souche Pt3 en morphotype ovale et dans la souche Pt1.8.6 cultivée dans un milieu possédant une faible teneur en ammonium ou en condition appauvrie en nitrate (**figure 39**). Ainsi, aucune EST correspondant spécifiquement au gène de la GnT I n'a été détectée à ce jour dans la souche Pt1.8.6 cultivée en conditions standards à partir de laquelle l'ensemble des résultats de N-glycosylation précédemment exposés ont été obtenus. Cependant, l'étude du transcriptome par création et analyse d'une banque d'EST est une approche non exhaustive, nécessitant de ce fait de vérifier les informations obtenues par amplification spécifique des ADNc générés à partir des ARNm du gène d'intérêt. Par conséquence, une analyse transcriptionnelle du gène de la GnT I de *P. tricornutum* a été entreprise par RT-PCR (*Reverse Transcription PCR*) et RT-qPCR (*Real Time quantitative* PCR) en fonction des conditions de culture, de la souche et du morphotype de *P. tricornutum*.

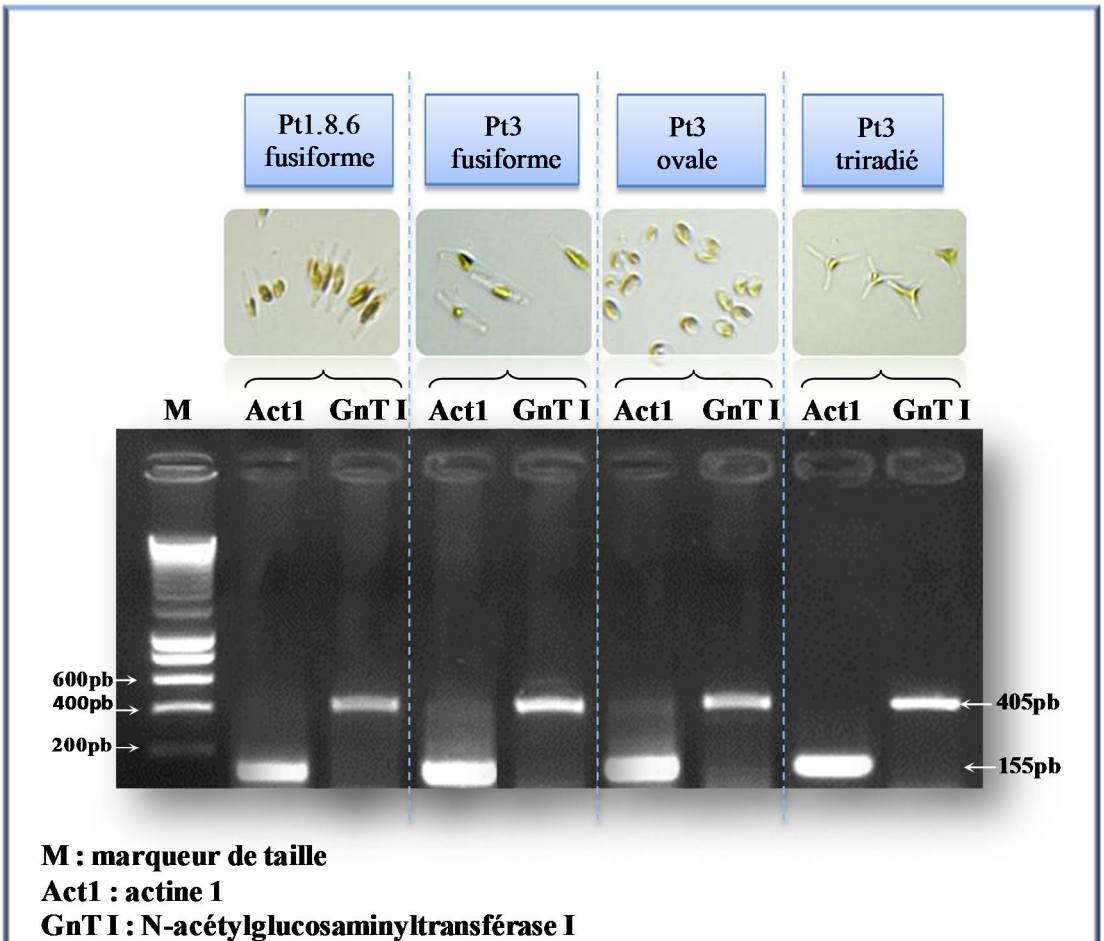
Figure 39 : Profil transcriptomique du gène de la N-acétylglucosaminyltransférase I (GnT I) de *P. tricornutum* selon 16 conditions étudiées (http://genome.jgi-psf.org/cgi-bin/browserLoad/?db=Phatr2&position=chr_16:89066-90505). Le gène de la GnT I (rectangle bleu) est aligné avec les EST (Expressed Sequence Tag) correspondantes (rectangles verts).

II-1-1- Analyse transcriptionnelle du gène de la GnT I de P. tricornutum en fonction de la souche et du morphotype par RT-PCR.

L'expression du gène de la GnT I de *P. tricornutum* a tout d'abord été étudiée en fonction de la souche et du morphotype, par RT-PCR, sur quatre cultures différentes en phase exponentielle de croissance :

- la souche Pt1.8.6 cultivée dans des conditions standards sous le morphotype fusiforme.
- la souche Pt3 cultivée dans des conditions standards sous le morphotype fusiforme.
- la souche Pt3 cultivée dans des conditions standards sous le morphotype triradié.
- la souche Pt3 cultivée en conditions de stress salin sous le morphotype ovale.

Afin de réaliser cette étude, des ADNc ont été générés à partir des ARNm totaux extraits de ces quatre cultures, puis amplifiés par PCR avec des couples d'amorces spécifiques. Le premier couple d'amorces utilisé est spécifique du gène de l'actine 1 servant ici de gène de référence. Ce couple conduit à l'amplification d'un fragment de 155 pb. Le second couple d'amorce est, quant à lui, spécifique du gène de la GnT I de *P. tricornutum* conduisant après amplification à la formation d'un fragment de 405 pb. Les résultats de cette expérience ont été visualisés par un gel d'électrophorèse et sont présentés sur la **figure 40**.



M : marqueur de taille
Act1 : actine 1
GnT I : N-acétylglucosaminyltransférase I

Figure 40 : Electrophorèse en gel d'agarose révélant le produit d'amplification PCR obtenu à partir des ADNc de *P. tricornutum* grâce à l'utilisation d'amorces spécifiques du gène de la GnT I et du gène de l'actine 1 (gène de référence) en fonction de la souche et du morphotype.

Les résultats obtenus montrent que le profil d'expression est le même pour chacune des quatre cultures étudiées (**figure 40**). En effet, les souches Pt1.8.6 et Pt3 dans le morphotype fusiforme ainsi que la souche Pt3 dans les morphotypes ovale et triradié, présentent une bande unique à la taille attendue avec les amorces spécifiques des gènes de l'actine 1 et de la GnT I soit respectivement 155 pb et 405 pb (**figure 40**). Ainsi, cette étude transcriptionnelle démontre que le gène de la GnT I est exprimé dans la souche Pt1.8.6 fusiforme de *P. tricornutum* en conditions standards de culture, contrairement à ce que pouvaient laisser penser les données d'EST. De même, il a été mis en évidence dans la souche Pt3 que ce gène est exprimé quel que soit le morphotype représenté (fusiforme, ovale et triradié). Ces résultats suggèrent que les deux souches Pt1.8.6 et Pt3 et les trois morphotypes de *P. tricornutum* sont potentiellement capables de synthétiser des N-glycannes complexes.

II-1-2- Analyse transcriptionnelle du gène de la GnT I en fonction de la quantité de nitrate dans le milieu de culture par RT-qPCR.

L'expression du gène de la GnT I de *P. tricornutum* a été étudiée sur une culture continue de la souche Pt1.8.6 par PCR quantitative en temps réel en fonction de la quantité de nitrate présente dans le milieu. Pour cela, la concentration en nitrate a été modulée au cours du temps générant ainsi trois phases : phase 1 : conditions standards ($[NO_3^-]$=1 mM), phase 2 : carence en nitrate ($[NO_3^-]$=0,05 mM) et phase 3 : enrichissement en nitrate ($[NO_3^-]$=1 mM). A l'issue de cette expérience, l'expression du gène de la GnT I a été normalisée par rapport à deux gènes de référence choisis en accord avec la littérature (Siaut et al., 2007) et codant respectivement pour l'histone H4 (H4 ; **figure 41A**) et la RPS (ribosomal protein small subunit 30 S ; **figure 41B**).

Le profil d'expression du gène de la GnT I obtenu suite à la normalisation avec la RPS est similaire à celui obtenu après normalisation avec l'histone H4 pour les phases 1 et 2, mais diffère pour la phase 3 (**figure 41**). Il sera donc difficile de conclure sur le comportement de l'expression du gène de la GnT I de *P. tricornutum* en condition « ré-enrichie » en nitrate (phase 3). Néanmoins, ces données permettent de confirmer la présence de transcrits GnT I dans ces conditions de culture, en accord avec les données de transcriptomique précédemment décrites.

De plus, ces résultats permettent de démontrer que le gène de la GnT I est exprimé de façon stable dans les conditions standards de culture (Phase 1 ; **figure 41**). En effet, le niveau d'expression de ce gène dans les différents prélèvements effectués au cours de cette phase ne

présente pas de variation significative. Au contraire, l'expression du gène de la GnT I de *P. tricornutum* semble rapidement augmenter suite à la déplétion en nitrate, puis décroit tout aussi rapidement et remonte légèrement jusqu'à atteindre un niveau d'expression à nouveau stable (**figure 41**). Il est important de noter que le niveau d'expression du gène de la GnT I dans cet état stable est du même ordre de grandeur que celui observé dans les conditions standards de culture.

En conclusion, ces analyses transcriptomiques semblent montrer que le gène de la GnT I de *P. tricornutum* est rapidement et brièvement surexprimé en réponse à une carence en nitrate dans le milieu de culture. Ces résultats suggèrent une possible implication des glycoprotéines portant des N-glycannes complexes dans la voie de signalisation et/ou dans la mise en place d'un système d'adaptation vis-à-vis de ces conditions de stress chez *P. tricornutum*.

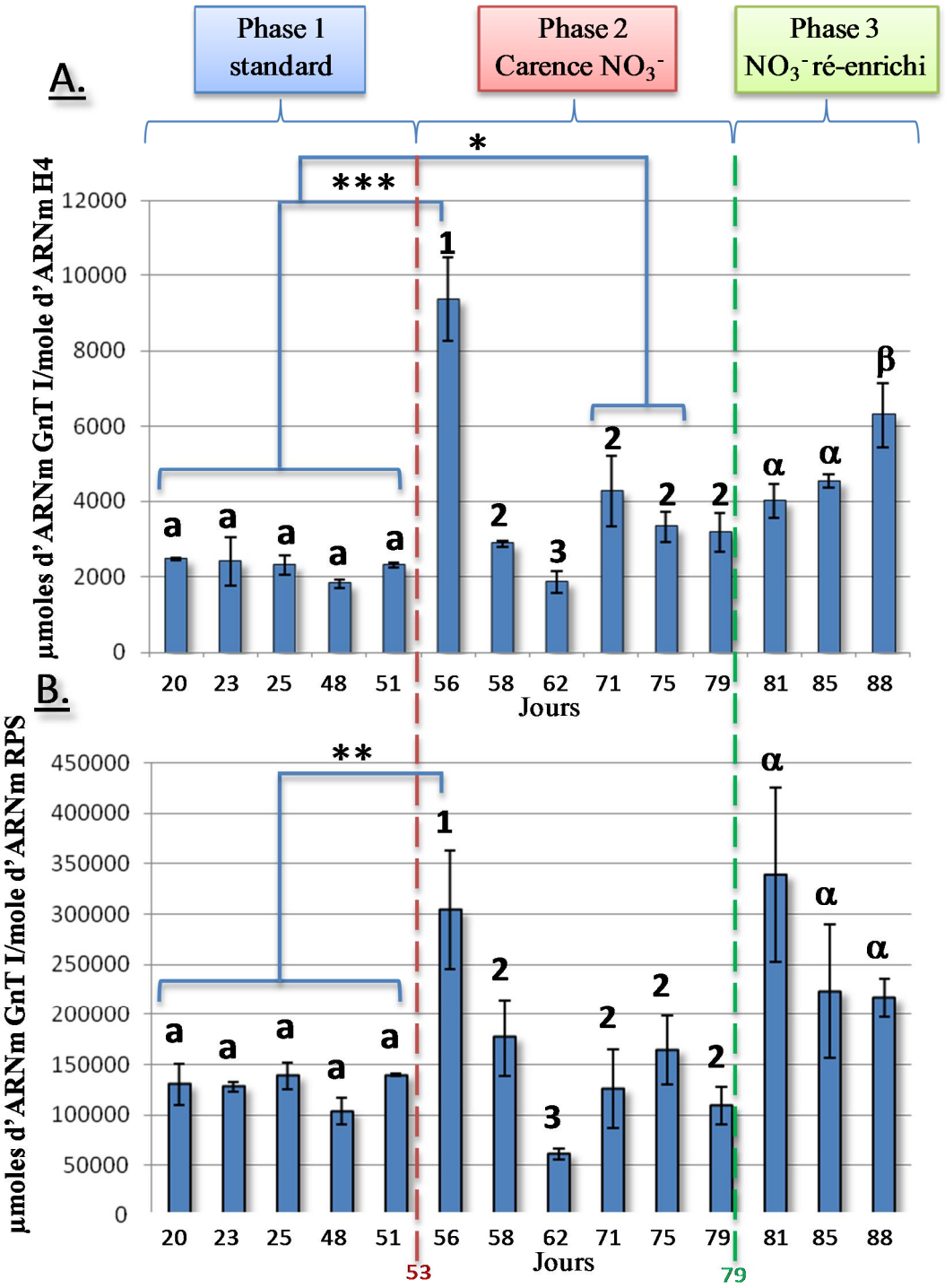


Figure 41 : Expression du gène de la GnT I de *P. tricornutum* par PCR quantitative en temps réel en fonction de la concentration en nitrate (NO_3^-) dans le milieu de culture. Cette expérience a été réalisée sur une culture Pt1.8.6 continue dont le taux de NO_3^- a été régulé générant ainsi trois phases : 1 : standard ([NO_3^-]=1 mM), 2 : appauvrie en NO_3^- ([NO_3^-]=0,05 mM) et 3 : réenrichie en NO_3^- ([NO_3^-]=1 mM). L'expression du gène de la GnT I a été normalisée par rapport à deux gènes de référence codant pour A. : l'histone H4 (H4) et B. : la RPS (ribosomal protein small subunit 30 S). Des tests statistiques ANOVA ont été réalisés au sein de chaque phase. Ainsi, chaque phase possède son propre système de notation des données statistiques qui sont indiquées par des lettres minuscules pour la phase 1, des chiffres pour la phase 2 et des lettres grecques pour la phase 3. Au sein de la même phase, deux échantillons affublés d'une lettre ou d'un chiffre différents présentent une variation significative de leur taux d'expression du gène de la GnT I. Par ailleurs, des tests ANOVA ont également été réalisés entre les prélèvements de phases différentes. Les différences significatives de l'expression du gène de la GnT I entre les prélèvements de la phase 1 et 2 sont indiquées sur le schéma avec des astérisques en suivant la nomenclature officielle qui tient compte du niveau de significativité.

II-2- Initiation de la caractérisation fonctionnelle des 3 gènes de *P. tricornutum* codant potentiellement pour des α-1,3-Fucosyltransférases.

La maturation des N-glycannes complexes chez les eucaryotes, se déroule exclusivement dans l'appareil de Golgi grâce à l'action de glycosyltransférases spécifiques. Parmi celles-ci, les fucosyltransférases permettent l'ajout de résidus fucose sur les GlcNAc distales ou proximale des N-glycannes complexes (voire sur un résidu galactose comme dans l'épitope Lewis^b présent chez les mammifères). Or, les résultats présentés dans la première partie de ce chapitre, ont permis de démontrer, au travers d'analyses biochimiques, que les glycoprotéines extraites de la souche Pt1.8.6 de *P. tricornutum* cultivée dans des conditions standards, portent majoritairement des N-glycannes oligomannosidiques, mais également une population mineure de N-glycannes complexes fucosylés. De plus, les analyses bioinformatiques du génome de *P. tricornutum* ont permis l'identification de trois gènes nommés respectivement 46109, 46110 et 54599, codant potentiellement pour des fucosyltransférases impliquées dans la maturation des N-glycannes complexes. Par conséquence, nos travaux se sont ensuite focalisés sur l'étude de la fucosylation des N-glycannes chez *P. tricornutum* à travers l'initiation de la caractérisation fonctionnelle de ces trois fucosyltransférases putatives. Pour cela, une approche de complémentation fonctionnelle par expression hétérologue, similaire à celle utilisée pour la caractérisation de la GnT I, a été envisagée dans des cellules mutantes CHO FUT8-/- déficientes en activité α-1,6-fucosyltransférase, enzyme responsable de la fucosylation du core glycannique chez les mammifères.

II-2-1- Analyse bioinformatique des gènes 46109, 46110 et 54599 de P. tricornutum.

La recherche effectuée par homologie de séquence sur le génome de *P. tricornutum* a permis d'identifier trois gènes codant potentiellement pour des fucosyltransférases impliquées dans la voie de biosynthèse des N-glycannes (54599, 46109 et 46110). Ces trois gènes codent pour des protéines membranaires de type II possédant tous les motifs structuraux caractéristiques des fucosyltransférases appartenant à la famille GT 10 (voir les résultats de la partie I présentés sous forme d'article). En effet, ces trois protéines FuT46109, FuT46110 et FuT54599 contiennent les motifs I et II des α-1,3-fucosyltransférases, la séquence SNC(G/A)A(R/H)N spécifique des α-1,3-fucosyltransférases de plante et de D. melanogaster ainsi que le motif CXXC situé à l'extrémité C-terminale qui est impliqué dans la formation de pont disulfure (**figure 42**).

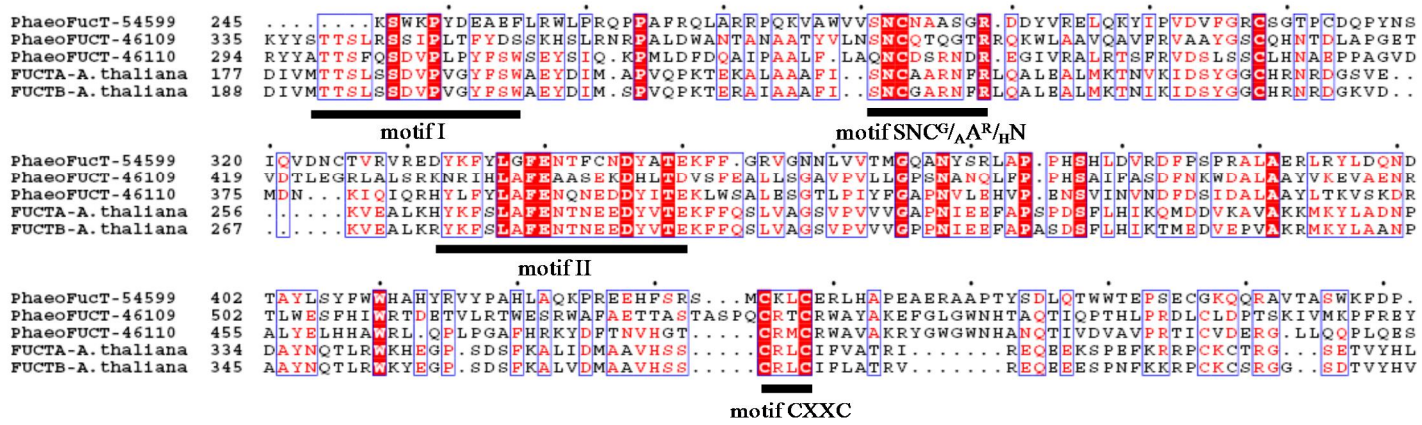


Figure 42 : Alignement multiple réalisé avec le logiciel MultAlin des séquences protéiques prédites FuT46109, FuT46110 et FuT54599 de *P. tricornutum*, avec celles de la FucT A et de la FucT B d' *A. thaliana*. Les motifs structuraux caractéristiques des α-1,3-fucosyltransférases appartenant à la famille GT 10 sont présentés sous cet alignement.

La famille GT 10 regroupe aussi bien les α-1,3-fucosyltransférases responsables de la fucosylation du core chez les plantes et les invertébrés, que les α-1,3/4-fucosyltransférases permettant l'ajout de résidus fucose en α-1,3 ou en α-1,4 sur les GlcNAc distales impliquées dans la formation des épitopes de type Lewis. Les analyses par similitudes de séquences réalisées par BLASTP sur NCBI contre tous les génomes montrent que les protéines codées par les gènes 46109 et 46110 de *P. tricornutum* présentent plus de similarité avec les α-1,3-fucosyltransférases du core alors que la protéine codée par le gène 54599 serait plus proche des α-1,3-fucosyltransférases d'insectes permettant la fucosylation des GlcNAc distales.

Les prédictions bioinformatiques indiquent que les gènes 54599 et 46109 ne semblent pas présenter d'intron et codent pour des protéines comportant respectivement 481 et 770 acides aminés (**figure 43**). Le gène 46110 possèderait, quant à lui, deux séquences introniques et coderait pour une protéine contenant 718 acides aminés (**figure 43**).

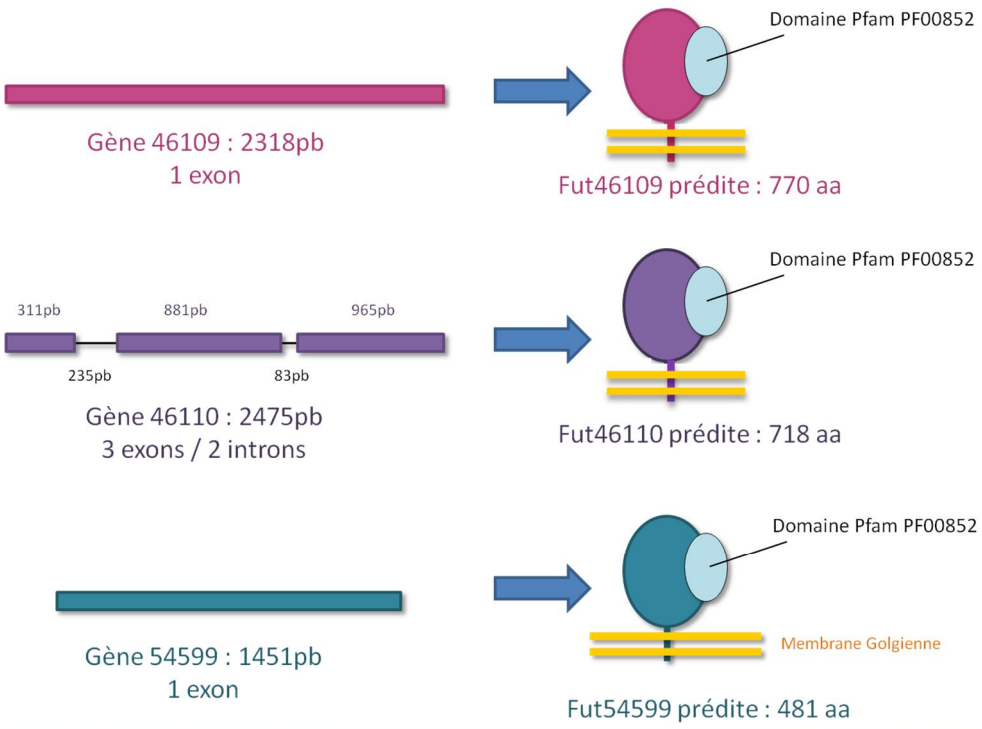


Figure 43: Schéma basé sur les prédictions bioinformatiques des gènes de la FuT46109, FuT46110 et FuT54599 de *P. tricornutum*, ainsi que leur protéine putative.

L'objectif de cette partie du projet de thèse étant de réaliser une complémentation fonctionnelle par insertion des gènes 54599, 46109 et 46110 de *P. tricornutum* dans un système hétérologue CHO, il est nécessaire de vérifier au préalable que ces gènes seront correctement traduits dans les cellules de mammifère. En effet, les sites d'épissage des ARN messagers (ARNm), et par conséquence le cadre de lecture, peuvent varier en fonction des organismes générant de ce fait un ARNm mature qui peut être différent en fonction du système dans lequel est exprimé le gène. Par conséquent, une protéine exprimée dans un système hétérologue peut différer de celle prédite dans l'organisme d'origine. C'est pourquoi, afin de vérifier que les séquences protéiques de *P. tricornutum* que nous souhaitons caractériser soient identiques à celles qui seront exprimées dans les cellules de mammifères, une analyse bioinformatique a été menée sur les trois gènes 46109, 46110 et 54559 avec le logiciel GENSCAN. Ce logiciel permet de simuler l'épissage des ARNm d'un gène en fonction du système d'expression, et de prédire les protéines qui seront traduites. Les séquences protéiques FuT46109 et FuT54599 obtenues suite à la prédiction chez les vertébrés sont identiques à celles annotées sur le site du génome de *P. tricornutum*. En revanche, l'alignement de la séquence protéique théorique FuT46110, avec celle générée chez les vertébrés, a montré une différence pour les 104 premiers acides aminés correspondant à la

région N-terminale comprenant le domaine transmembranaire et le signal d'ancrage (**figure 44**). Le domaine catalytique, quant à lui, est conservé dans son intégralité. Cette différence est liée au fait que l'intron 1 du gène 46110 n'est pas reconnu comme une partie intronique par le logiciel GENSCAN, mais comme une séquence codante. La protéine FuT46110 ainsi prédite possède une séquence contenant 707 acides aminés.

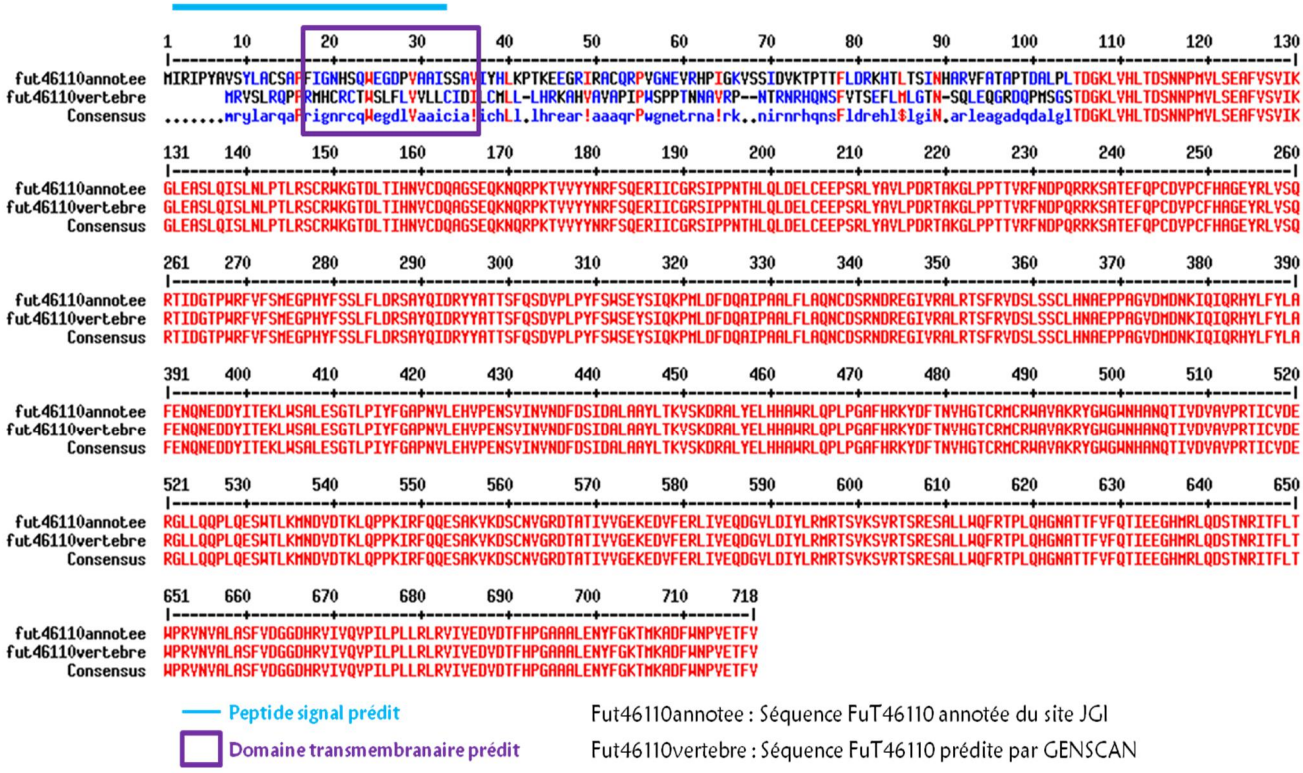
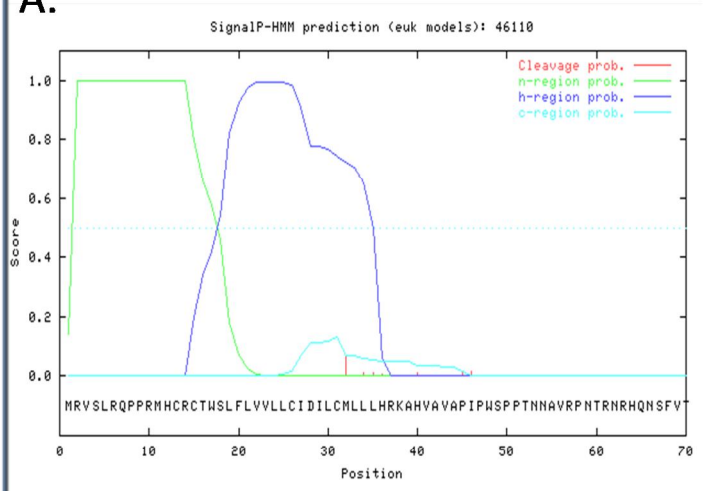
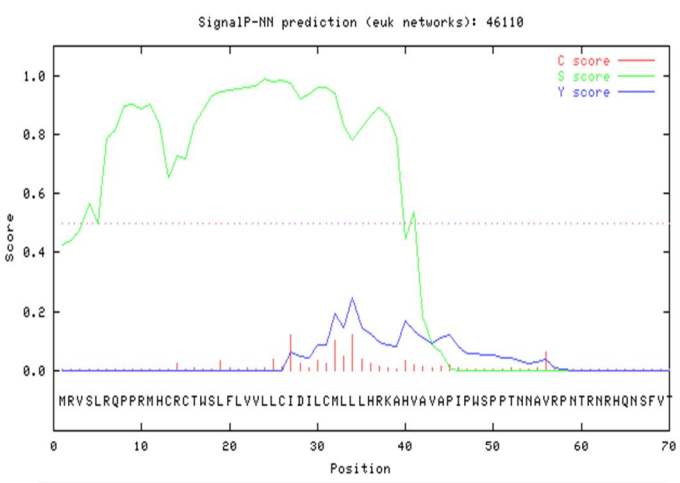


Figure 44 : Alignement réalisé avec le logiciel MultAlin de la séquence théorique FuT46110 de *P. tricornutum*, avec celle générée chez les vertébrés par le logiciel GENSCAN.

Une analyse bioinformatique avec les logiciels SignalP et TMHMM de cette séquence protéique FuT46110 prédite par GENSCAN a permis de mettre en évidence la présence d'un peptide d'ancrage et d'un domaine transmembranaire (**figure 45**). D'après l'ensemble de ces résultats, le gène 46110 de *P. tricornutum* exprimé chez les mammifères coderait pour une protéine membranaire de type II de 707 acides aminés. Les mêmes résultats ont été obtenus lorsque la simulation d'épissage de ce gène a été réalisée chez la plante. Ainsi, la différence de séquence protéique observée n'est pas spécifique à l'expression du gène 46110 chez les vertébrés. Ces résultats suggèrent donc un problème d'annotation de ce gène dans le génome de *P. tricornutum*. En effet, le gène 46110 ne comporterait qu'un seul intron (intron 2) car la séquence nucléotidique annotée sur le génome comme l'intron 1 serait en réalité une séquence codante.

A.

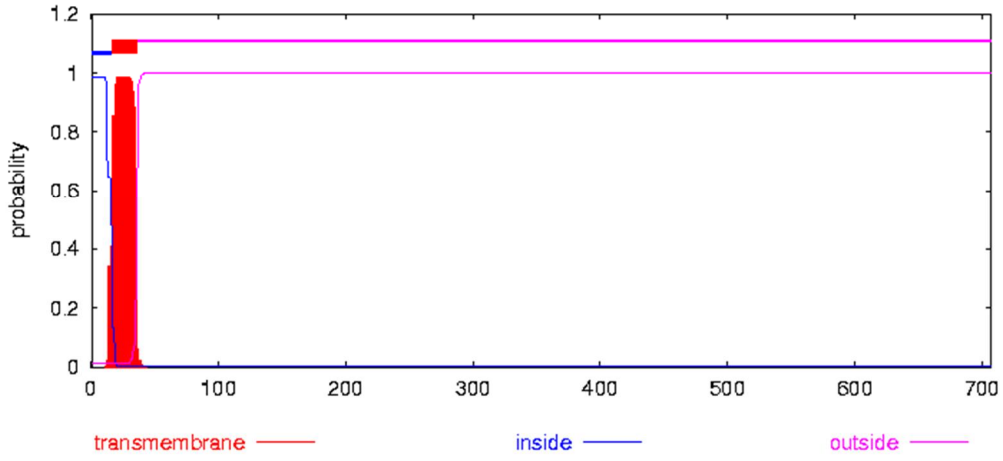





```
>46110
Prediction: Signal anchor
Signal peptide probability: 0.144
Signal anchor probability: 0.856
Max cleavage site probability: 0.065 between pos. 31 and 32
```

```
>46110                  length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C     34       0.124   0.32   NO
  max. Y     34       0.248   0.33   NO
  max. S     24       0.987   0.87   YES
  mean S      1-33    0.831   0.48   YES
       D      1-33    0.540   0.43   YES
# Most likely cleavage site between pos. 33 and 34: CML-LL
```

B.



<u>Figure 45</u> : Résultat des prédictions du peptide signal et du domaine transmembranaire pour la séquence protéique FuT46110 prédite chez les vertébrés par le logiciel GENSCAN à partir du gène 46110 de *P. tricornutum*. La prédiction d'un peptide signal et d'un signal d'ancrage a été réalisée par le logiciel SignalP (A.). La prédiction du domaine transmembranaire a été réalisée par le logiciel TMHMM (B.).

II-2-2- Clonage des trois gènes 54599, 46109 et 46110 de P. tricornutum codant potentiellement pour des α-1,3-fucosyltransférases.

La première étape dans la caractérisation fonctionnelle des trois fucosyltransférases de *P. tricornutum* a été de cloner les gènes 54599, 46109 et 46110 dans un vecteur d'expression de mammifère. Le clonage de ces gènes n'a délibérément pas été réalisé à partir de la séquence d'ADNc correspondante mais à partir de la séquence génomique. En effet, pour les gènes 46109 et 54599 qui ne présentent pas d'intron la séquence génomique et ADNc est la même. Cependant, pour le gène 46110, des doutes concernant sa partie intronique n°1 ont été émis à l'issue des analyses bioinformatiques exposées précédemment. Ainsi, il a été choisi de conserver ces potentiels introns dans la construction, c'est pourquoi là encore le clonage a été réalisé à partir de la séquence génomique du gène 46110 et non de l'ADNc. Pour effectuer ces clonages, les gènes d'intérêt ont été amplifiés par PCR par une ADN polymérase de haute fidélité permettant l'ajout d'extrémités adénosines en 3' du produit PCR, nécessaires à sa ligation dans le vecteur d'expression pcDNA 3.1/V5-His-TOPO. En effet, ce dernier est un vecteur linéaire de mammifère possédant des extrémités T (Thymidine) dépassantes qui vont s'hybrider avec les extrémités A du produit PCR. De ce fait, le produit PCR pourra s'insérer dans les deux sens avec la même probabilité.

Suite à la ligation et afin d'identifier des clones présentant le gène d'intérêt inséré dans le bon sens, une digestion des plasmides recombinants a été réalisée avec une enzyme de restriction. L'enzyme utilisée présente un premier site de restriction situé dans le vecteur, et un second localisé au niveau de l'extrémité 5' ou 3' de l'insert. A l'issue de ce contrôle, deux clones positifs obtenus pour les constructions pcDNA3.1/V5-His-TOPO + 54599 et pcDNA3.1/V5-His-TOPO + 46109 et trois clones positifs obtenus pour le plasmide recombinant pcDNA3.1/V5-His-TOPO + 46110 ont été séquencés afin de s'assurer qu'aucune mutation ne soit apparue au cours des étapes de clonage et d'amplification. Suite à ce séquençage, un clone a été sélectionné pour chaque construction afin de réaliser la transfection des cellules de mammifères mutantes CHO FUT8-/-.

Le vecteur pcDNA 3.1/V5-His-TOPO a été choisi pour réaliser cette étude car il permet l'ajout d'une séquence codant pour deux étiquettes courtes que sont l'épitope V5 et une queue hexahistidine, en 3' du gène d'intérêt permettant d'identifier et/ou de purifier la protéine recombinante. Les cartes des plasmides recombinants pcDNA3.1/V5-His-TOPO + 54599, pcDNA3.1/V5-His-TOPO + 46109 et pcDNA3.1/V5-His-TOPO + 46110 sont présentées sur la **figure 46**.

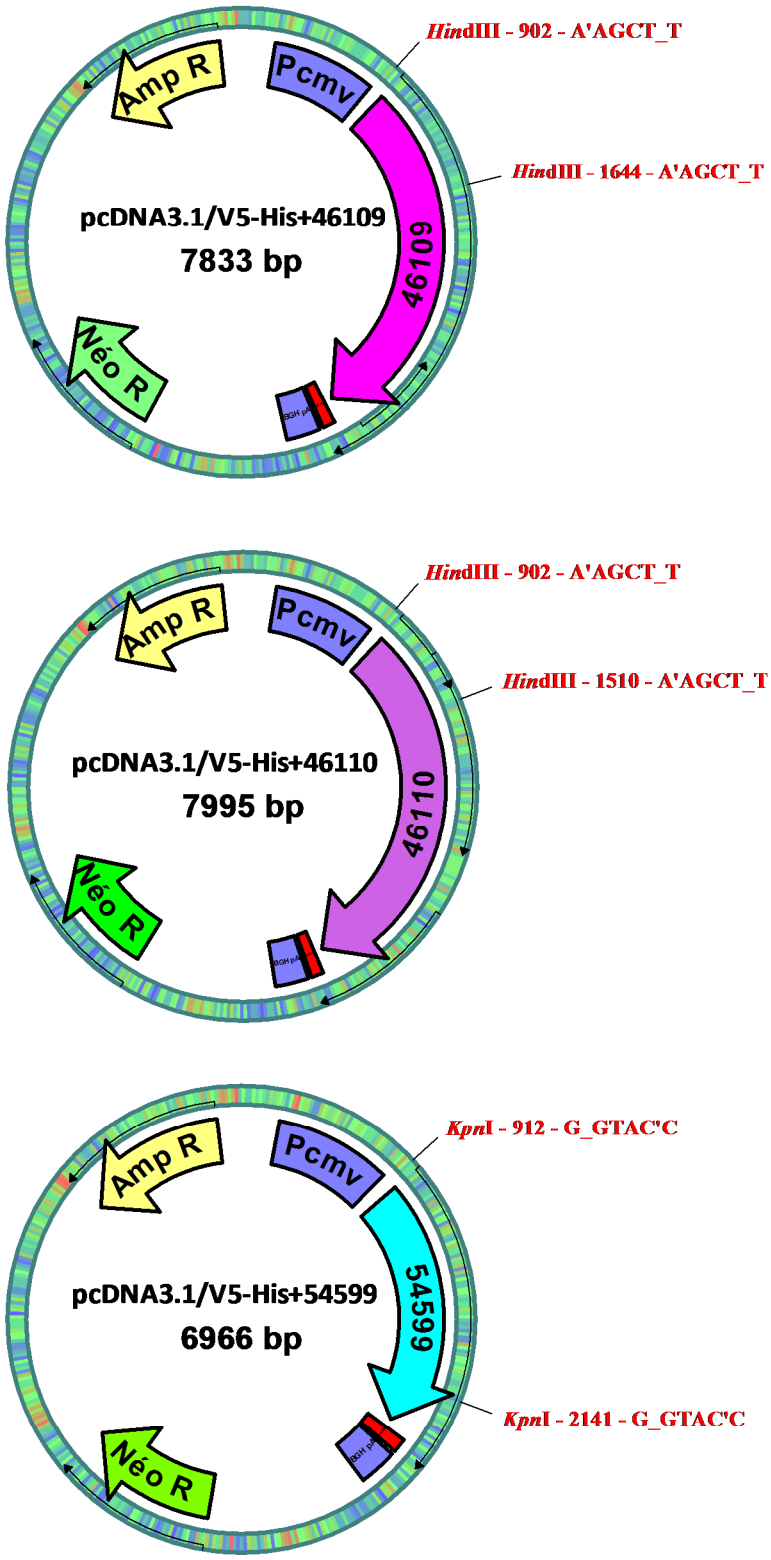
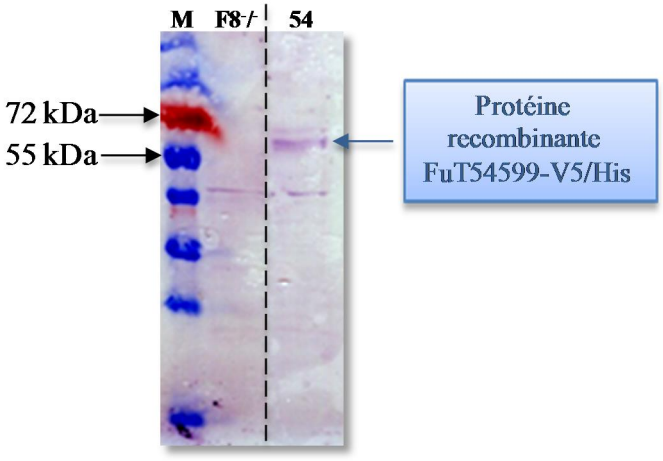


Figure 46 : Construction génétique des plasmides recombinants pcDNA3.1/V5-His-TOPO+46109, pcDNA3.1/V5-His-TOPO+46110 et pcDNA3.1/V5-His-TOPO+54599. Les cartes ont été réalisées et éditées avec le logiciel pDRAW32. Le pourcentage de GC des séquences est signalé par un code de couleur. Les flèches indiquent la présence et le sens des ORF. Le promoteur P_{CMV} et le terminateur BGH polyadénylation sont indiqués sur fond bleu foncé. Les inserts 46109, 46110 et 54599 sont respectivement sur fond rose, violet et bleu ciel. La séquence codant pour les étiquettes épitopeV5/Hexahistidine est représenté sur fond rouge. Le gène de résistance à l'ampicilline est en jaune et celui de la Néomycine en vert. Les sites de restrictions utilisés pour la sélection de clones positifs sont indiqués en rouge (HindIII pour pcDNA3.1/V5-His-TOPO+46109 et pcDNA3.1/V5-His-TOPO+46110 et Kpn I pour pcDNA3.1/V5-His-TOPO+54599).

II-2-3- Transfection des cellules CHO FUT8-/- avec les plasmides pcDNA3.1/V5-His-TOPO + 54599, pcDNA3.1/V5-His-TOPO + 46109 et pcDNA3.1/V5-His-TOPO + 46110 et contrôle de l'expression des protéines recombinantes.

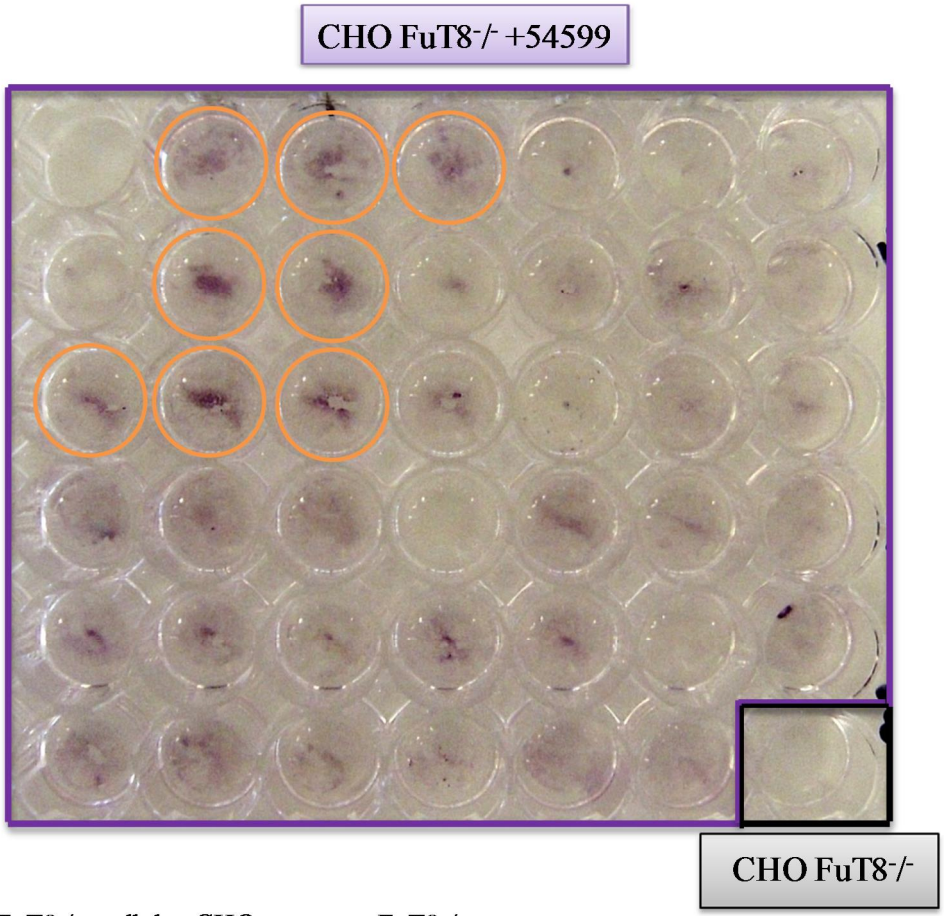
Les vecteurs pcDNA3.1/V5-His-TOPO + 54599, pcDNA3.1/V5-His-TOPO + 46109 et pcDNA3.1/V5-His-TOPO + 46110 ont été utilisés pour transfecter de façon stable des cellules mutantes CHO FuT8-/- par lipofection. Pour les cellules CHO FUT8-/- transfectées avec les plasmides recombinants pcDNA3.1/V5-His-TOPO + 46109 et pcDNA3.1/V5-His-TOPO + 46110, une sélection monoclonale dans des plaques 96 puits a été initiée et est actuellement en cours de réalisation. Pour les cellules CHO FUT8-/- transfectées avec le plasmide pcDNA3.1/V5-His-TOPO + 54599, l'expression de la protéine recombinante FucT54599-V5/His a été mise en évidence sur un pool total de cellules transfectées par immunodétection sur empreinte avec un anticorps spécifiquement dirigé contre l'épitope V5. L'anticorps secondaire utilisé est couplé à la phosphatase alcaline et révélé par colorimétrie. Des cellules CHO FuT8-/- non transformées ont servi de témoin négatif dans cette analyse. Les résultats obtenus, présentés sur la **figure 47**, montrent le marquage spécifique d'une bande apparaissant à 61 kDa. Cette masse moléculaire correspond à celle attendue pour la protéine recombinante FuT54599-V5/His. De plus, ces résultats indiquent que le taux d'expression de la protéine recombinante FuT54599-V5/His semble tout à fait satisfaisant.



M : marqueur de taille protéique.
F8-/- : Extrait protéique des cellules CHO mutantes FuT8-/-.
54 : Extrait protéique des cellules CHO mutantes FuT8-/- transfectées avec le gène 54599 de *P. tricornutum*.

Figure 47 : Immunodétection sur empreinte de l'expression de la protéine recombinante FuT54599-V5/His dans les cellules CHO FuT8-/- transfectées de façon stable. L'analyse de l'expression protéique a été réalisée grâce à l'anticorps anti-V5 reconnaissant l'étiquette spécifiquement ajoutée sur les protéines recombinantes. L'anticorps secondaire utilisé est couplé à la phosphatase alcaline, ce qui a permis une révélation colorimétrique de la membrane par ajout du réactif NBT/BCIP. Un contrôle négatif a été réalisé avec un extrait protéique total des cellules CHO FuT8-/- non transformées.

La sélection monoclonale de cette lignée cellulaire CHO FUT8-/- transfectée avec le plasmide pcDNA3.1/V5-His-TOPO + 54599 a été achevée. Les clones ont ensuite été criblés puis sélectionnés par analyse de l'expression de la protéine recombinante FuT54599-V5/His. Pour cela, une immunodétection sur cellules dans des plaques 48 puits a été réalisée avec l'anticorps anti-épitope V5 (**figure 48**). A l'issue de ce contrôle, huit clones positifs ont été sélectionnés pour la suite des expériences.

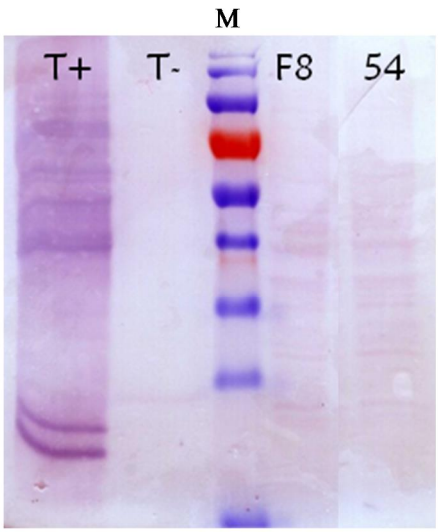


CHO FuT8$^{-/-}$: cellules CHO mutantes FuT8-/-.
CHO FuT8$^{-/-}$ +54599 : cellules CHO mutantes FuT8-/- transfectées avec le gène 54599 de *P. tricornutum*.

Figure 48 : Immunodétection sur cellules de l'expression de la protéine recombinante FuT54599-V5/His dans les clones CHO FuT8-/- transfectés de façon stable avec le gène 54599 de *P. tricornutum* par l'anticorps anti-épitope V5. L'anticorps secondaire utilisé est couplé à la phosphatase alcaline, ce qui a permis une révélation colorimétrique des cellules par ajout du réactif NBT/BCIP. Un contrôle négatif a été réalisé avec des cellules CHO FuT8-/- non transformées. Les clones positifs sélectionnés pour la suite de la caractérisation fonctionnelle sont entourés en orange.

II-2-4- Test préliminaire de l'activité de FuT54599 par immunomarquage sur empreinte avec l'anticorps anti-α-1,3-fucose du core.

Une étude préliminaire de l'activité de la FuT54599 recombinante a ensuite été initiée par immunomarquage sur empreinte des protéines avec des anticorps anti-α-1,3-fucose spécifiques de la fucosylation du core des N-glycannes. Cette analyse a été réalisée sur des protéines extraites à partir d'un pool total de cellules CHO FuT8-/- transfectées avec le plasmide pcDNA3.1/V5-His-TOPO + 54599 et sur des extraits protéiques totaux de feuilles d'*Arabidopsis thaliana* utilisées comme témoins. En effet, les protéines extraites des feuilles de la lignée sauvage columbia-0 ont été utilisées comme témoin positif, et celles des feuilles du double mutant FuTA/FuTB (mutant présentant une déficience de la fucosylation du core en α-1,3) comme témoin négatif. Les résultats obtenus ne montrent pas de marquage spécifique α-1,3-fucose sur les glycoprotéines de la lignée cellulaire CHO FUT8-/- transfectée avec le plasmide pcDNA3.1/V5-His-TOPO + 54599 (**Figure 49**). Ces analyses préliminaires semblent indiquer que cette protéine recombinante FucT54599 ne présente pas d'activité α-1,3-fucosyltransférase responsable de la fucosylation du core des N-glycannes, en accord avec les prédictions bioinformatiques qui prédisent que cette protéine serait plutôt une fucosyltransférase permettant l'ajout de fucose en α-1,3 sur les résidus GlcNAc distales.



M : marqueur de taille protéique.
T+ : Extrait protéique des feuilles de la lignée sauvage columbia-0 d'*Arabidopsis thaliana*.
T- : Extrait protéique des feuilles du double mutant FuTA/FuTB d'*Arabidopsis thaliana*.
F8 : Extrait protéique des cellules CHO mutantes FuT8$^{-/-}$.
54 : Extrait protéique des cellules CHO mutantes FuT8$^{-/-}$ transfectées avec le gène 54599 de *P. tricornutum*.

Figure 49 : Etude de l'activité de la protéine recombinante FuT54599-V5/His dans les cellules CHO FuT8-/- transfectées de façon stable par immunodétection sur empreinte avec l'anticorps anti-α1,3-fucose. L'anticorps secondaire utilisé est couplé à une phosphatase alcaline, ce qui a permis une révélation colorimétrique par ajout du réactif NBT/BCIP (Promega).

Les travaux concernant la caractérisation fonctionnelle des fucosyltransférases de *P. tricornutum* n'ont pu être terminés dans le temps imparti de mon contrat doctoral (3 ans). En conséquence, ce travail a été repris et est actuellement poursuivi dans le cadre de la thèse de Clément Ovide.

Chapitre 4 : Discussion-Perspectives

La N-glycosylation est une modification essentielle dans le repliement des protéines empruntant le système endomembranaire de sécrétion. Les premières étapes de ce processus que sont la synthèse du précurseur oligosaccharidique, son transfert sur le polypeptide en cours de synthèse et le cycle de contrôle qualité du repliement des glycoprotéines sont localisées dans le RE et sont des étapes universelles rencontrées chez tous les eucaryotes (Burda and Aebi, 1999 ; Aebi, 2013). A l'inverse, la maturation des N-glycannes qui a lieu dans l'appareil de Golgi a divergé au cours de l'évolution, aboutissant à la formation de N-glycannes matures complexes organisme- voire tissu-spécifiques. Ce processus a largement été décrit chez les levures et chez les eucaryotes supérieurs comme les mammifères, les plantes et les invertébrés (insectes et vers)(Kubelka et al., 1993 ; Lerouge et al., 1998 ; Gemmill et Trimble, 1999 ;). Cependant, lorsque ce travail a été initié, peu d'informations étaient disponibles sur la N-glycosylation des microalgues et aucune voie de biosynthèse des N-glycannes n'avait été caractérisée chez ces organismes. Comme nous l'avons vu dans la première partie de la synthèse bibliographique, les microalgues regroupent des organismes très divers appartenant à des phyla différents (Keeling, 2013). Parmi celles-ci, les diatomées présentent un intérêt tout particulier en raison de leur position phylogénétique. En effet, les diatomées appartiennent au phylum des Straménopiles qui est bien distinct de celui des plantes et des animaux et qui regroupe un ensemble d'organismes très variés dont les voies métaboliques sont souvent assez mal connues, comme la voie de N-glycosylation. Les quelques organismes de ce phylum pour lesquels des informations sur le processus de N-glycosylation sont disponibles, présentent généralement des N-glycannes matures similaires à ceux décrits chez les levures (Gemmill et Trimble, 1999). Ainsi, le manque de connaissance sur la voie de N-glycosylation de cette branche de l'arbre phylogénétique des eucaryotes, associé à la possibilité d'accéder au génome de plusieurs espèces, aux outils de biologie moléculaire et au matériel biologique, a fortement contribué à choisir les diatomées comme microalgue modèle dans notre étude. Dans ce contexte, le projet de thèse qui m'a été confié a consisté à étudier la voie de N-glycosylation de la diatomée modèle *P. tricornutum*. Ces travaux se sont inscrits dans une thématique de recherche plus globale du laboratoire visant à mieux comprendre ce processus métabolique d'un point de vue évolutif. En effet, l'étude de la voie de N-glycosylation des organismes répartis dans les différents phyla de l'arbre du vivant permettrait, par comparaison des séquences de glycoenzymes et des structures glycanniques détaillées, de déterminer l'ordre d'apparition, au cours de l'évolution, des glycosyltransférases spécifiques intervenant dans la maturation des N-glycannes complexes (Breton et al., 2012). Cela permettrait de comprendre pourquoi certaines enzymes impliquées dans la voie de N-

glycosylation sont présentes dans un organisme et pas dans un autre. Un parallèle pourrait alors être fait avec les différents évènements géologiques et les conditions environnementales dans lesquelles s'est développé chaque organisme au cours du temps, ce qui permettrait d'approfondir les connaissances sur le rôle des différents motifs glucidiques des N-glycannes complexes dans l'adaptation des organismes au cours de l'évolution. Pour atteindre cet objectif, le laboratoire a donc récemment choisi des microalgues représentatives des différents phyla afin non seulement de mener des études comparatives des séquences génomiques et protéiques codant pour leurs glycosidases et leurs glycosyltransférases putatives, mais également de réaliser l'analyse structurale des N-glycannes portés par leurs protéines endogènes.

I- La voie de N-glycosylation de *Phaeodactylum tricornutum* : un processus GnT I-dépendant.

Les recherches bio-informatiques par similitudes de séquences menées sur le génome de la souche Pt1.8.6 de *P. tricornutum* ont révélé la présence d'un jeu complet de séquences codant potentiellement pour des protéines impliquées dans la synthèse du précurseur oligosaccharidique $Glc_3Man_9GlcNAc_2$-PP-Dol, des sous-unités du complexe OST, ainsi que des glucosidases et des chaperonnes du RE responsables du contrôle qualité des protéines (Helenius, 1994 ; Gidalevitz et al., 2013). Cela suggère que *P. tricornutum* possède toute la machinerie nécessaire aux différentes étapes de la N-glycosylation ayant lieu dans le RE, à l'image de tous les eucaryotes étudiés jusqu'à ce jour (Chevet et al., 2001 ; Lederkremer, 2009). L'analyse du génome a également permis d'identifier un gène codant potentiellement pour une α-mannosidase I golgienne, glycosidase impliquée dans l'hydrolyse de l'oligosaccharide Man-9/8 en N-glycannes oligomannosidiques contenant de 5 à 8 résidus mannose (Liebminger et al., 2009 ; Kajiura et al., 2010).

En accord avec ces prédictions bioinformatiques, les analyses biochimiques ont révélé que les protéines de la souche Pt1.8.6 de *P. tricornutum* cultivée en conditions standards sont principalement glycosylées par des structures oligomannosidiques allant de Man-9 à Man-5. Cependant, d'autres espèces glycanniques minoritaires de type paucimannose telles que Man-3, Man-4 et $Man_3FucGlcNAc_2$ ont également été détectées par spectrométrie de masse MALDI-TOF dans des extraits glycanniques réalisés avec la PNGase A. L'absence de l'oligosaccharide fucosylé $Man_3FucGlcNAc_2$ dans les échantillons glycanniques obtenus après digestion par la PNGase F, suggère que le résidu fucose est probablement lié en α-1,3 sur le

core de ce N-glycanne. En effet, la PNGase F est une enzyme incapable de catalyser la déglycosylation des protéines lorsqu'un résidu fucose est lié en α-1,3 sur la GlcNAc proximale impliquée dans la liaison N-osidique formée entre l'oligosaccharide et la protéine (Tretter et al., 1991). La présence d'un tel glycanne fucosylé a été corroborée par l'obtention d'un faible signal par immunodétection sur empreinte des glycoprotéines avec un anticorps spécifique des résidus fucose liés en α-1,3 sur le core glycannique, et par la présence de 3 gènes codant potentiellement pour des α-1,3-fucosyltransférases prédits dans le génome de *P. tricornutum.*

Les structures paucimannosidiques Man-3, Man-4 et $Man_3FucGlcNAc_2$ qui ont été détectées sur les glycoprotéines de *P. tricornutum,* sont également retrouvées chez les plantes et les invertébrés. Dans ces organismes, les paucimannoses proviennent de la dégradation de N-glycannes complexes possédant une GlcNAc terminale par des N-acétylglucosaminidases (Vitale et Chrispeels, 1984 ; Altmann et al., 1995 ; Zhang et al., 2003). Ainsi, la présence de telles structures paucimannosidiques chez *P. tricornutum* suggère que cette microalgue est capable de synthétiser des N-glycannes complexes qui seraient dégradés par la suite par des N-acétylglucosaminidases.

Chez les eucaryotes supérieurs, la synthèse des N-glycannes complexes est initiée par la GnT I qui est l'enzyme responsable de la maturation des N-glycannes oligomannosidiques en N-glycannes complexes, par ajout d'un résidu GlcNAc terminal sur un oligosaccharide de type Man-5 (Chen et al., 1999 ; Strasser et al., 1999 ; Unligil et al., 2000). Cette modification est alors suivie par l'action d'une α-mannosidase II et d'une GnT II, aboutissant à la formation d'un core glycannique commun aux eucaryotes supérieurs, $GlcNAc_2Man_3GlcNAc_2$ (Kaushal et al., 1990 ; Altmann et März, 1995 ; Strasser et al., 2006). Or, des gènes codant potentiellement pour une GnT I et une α-mannosidase II ont été identifiés dans le génome de *P. tricornutum*, mais aucun n'a été prédit pour une GnT II. Cependant, l'analyse bioinformatique de la séquence protéique codée par le gène putatif de l'α-mannosidase II a révélé la présence d'un domaine Pfam05060 caractéristique des GnT II. Ces données suggèrent que la même protéine pourrait être dotée de ces deux activités catalytiques. Néanmoins, la séquence protéique de cette α-mannosidase II codée par le gène 52248 reste encore incertaine. En effet, le gène annoté sur le génome de *P. tricornutum* code pour une protéine de 173 acides aminés qui ne possède pas de codon stop ni de domaine transmembranaire et de peptide signal, caractéristiques des enzymes golgiennes impliquées

dans la maturation des N-glycannes. Suite à nos analyses bioinformatiques, la séquence a été allongée en C-terminal jusqu'au premier codon stop apparaissant dans le même cadre de lecture. C'est sur cette portion supplémentaire qu'est apparue une séquence correspondant au domaine catalytique spécifique des GnT II. De plus, des EST recouvrant la séquence nucléotidique codant potentiellement pour le domaine Pfam05060 ont été identifiées dans plusieurs conditions de stress testées telles que la carence en silice et en fer (Maheswari et al., 2009). Par ailleurs, la séquence protéique a également été allongée en N-terminal, ce qui a fait apparaitre la présence d'une séquence correspondant à un domaine transmembranaire mais pas de peptide signal. La séquence protéique de l'α-mannosidase II putative ainsi obtenue contient 1498 acides aminés. Bien que la simulation de cette séquence dans GENSCAN n'ait pas révélé la présence d'intron et ait indiqué que cette séquence représente une seule et même ORF, la réelle séquence de cette protéine reste encore incertaine. C'est pourquoi, une étude visant à cloner la totalité de cette séquence est envisagée afin de l'exprimer dans une lignée mutante HEK déficiente en activité α-mannosidase II. Ces travaux permettront d'une part de réaliser la caractérisation fonctionnelle de cette α-mannosidase II, mais également de purifier la protéine afin de la séquencer et de déterminer sa séquence réelle. De la même manière, il pourrait être intéressant de réaliser une expérience similaire dans des cellules mutantes déficientes en activité GnT II. En parallèle, il est envisagé d'étudier l'expression de ce gène, ce qui permettrait de connaitre la véritable séquence de ces transcrits.

Bien qu'aucun N-glycanne reflétant directement l'action de la GnT I n'ait pu être observé sur les glycoprotéines de la souche Pt 1.8.6 cultivée en conditions standards, l'expression de son gène a néanmoins été montrée dans cette même culture par PCR quantitative en temps réel. De plus, la caractérisation fonctionnelle de la protéine codée par ce gène a également été réalisée par complémentation de cellules mutantes d'ovaires de hamster chinois nommées CHO Lec1, déficientes en activité GnT I endogène et donc incapables de réaliser la synthèse de N-glycannes complexes (Stanley et al., 1975 ; Ripka et al., 1990). Cette approche par complémentation dans un système hétérologue a permis de démontrer que le gène prédit chez *P. tricornutum* code pour une GnT I fonctionnelle capable de restaurer la maturation de N-glycannes oligomannosidiques en N-glycannes complexes dans les cellules CHO Lec1. A notre connaissance, ce travail constitue la première caractérisation fonctionnelle d'une glycosyltransférase de microalgue. L'ensemble de ces données suggèrent que *P. tricornutum* est capable de synthétiser des N-glycannes complexes, et que la localisation de la GnT I et les mécanismes de rétention des enzymes golgiennes sont

conservés entre les mammifères et les diatomées. Par ailleurs, deux gènes codant potentiellement pour des N-acétylglucosaminidases appartenant à la famille GT 20 ont été révélés dans le génome de *P. tricornutum,* identifiés par les séquences génomiques 49563 et 45073.

L'ensemble de ces résultats indique que les structures paucimannosidiques détectées dans la souche Pt1.8.6 cultivée en conditions standards, seraient synthétisées de la même manière que chez les plantes et les invertébrés (Zhang et al., 2003 ; Schachter, 2009). Ainsi, l'oligosaccharide $GlcNAcMan_5GlcNAc_2$, produit de la GnT I, serait dégradé successivement dans l'appareil de Golgi en $GlcNAcMan_4GlcNAc_2$ et $GlcNAcMan_3GlcNAc_2$ par l'α-mannosidase II, puis subirait potentiellement le transfert d'une seconde GlcNAc terminale, et enfin l'ajout d'un résidu fucose en α-1,3 sur la GlcNAc du core glycannique par une α-1,3-fucosyltransférase. Le ou les résidus GlcNAc terminale(s) serai(en)t alors éliminé(s) par des N-acétylglucosaminidases officiant certainement soit dans le système sécrétoire soit dans le compartiment de stockage, par analogie avec ce qui a lieu chez les plantes et les invertébrés. L'ensemble de ces travaux démontre donc l'existence d'une voie de N-glycosylation GnT I–dépendante chez *P. tricornutum.*

Ce processus de maturation des N-glycannes diverge fortement de celui observé chez la majorité des eucaryotes inférieurs étudiés à ce jour, comme les levures ou l'oomycète *Phytophthora infestans,* appartenant au même phylum que les diatomées. Ces organismes ne possèdent pas de GnT I et ne sont donc pas capables de synthétiser de N-glycannes complexes. La maturation de leurs N-glycannes aboutit généralement à la formation de structures poly- voire hypermannosylées (Herscovics et Orlean, 1993 ; Gemmill et Trimble, 1999).

En revanche, une recherche de gènes codant potentiellement pour une GnT I a été réalisée dans les génomes de plusieurs autres espèces de microalgues appartenant aux trois lignées principales que sont les viridiplantae, les diatomées (ou hétérocontes) et les haptophytes. Des gènes codant potentiellement pour une GnT I ont été identifiés dans deux autres diatomées que *P. tricornutum*, à savoir *F. cylindrus* et *T. pseudonana,* ainsi que dans l'haptophyte *E. huxleyi*. Les analyses menées chez les viridiplantae (auxquelles appartiennent les microalgues vertes) ont uniquement révélé la présence d'une séquence GnT I dans le génome de *C. vulgaris* et *M. pusilla*. Il a été montré que les séquences complètes des GnT I de microalgues partagent environ 25 % d'identité avec celles des plantes et des mammifères. De

plus, l'analyse des séquences montre que les GnT I des diatomées et des haptophytes sont rassemblées dans un même groupe et semblent être clairement séparées des autres GnT I observées dans l'arbre phylogénétique (cf figure 7 de l'article publié dans *Journal of Biological Chemistry*). Ainsi, ces résultats indiquent que la GnT I et donc la maturation GnT I-dépendante des N-glycannes semble être conservée chez les diatomées et qu'elle est potentiellement présente dans quelques autres espèces de microalgues.

Contrairement à *P. tricornutum,* les N-glycannes de microalgues unicellulaires qui ont été caractérisés à ce jour, comme ceux de la microalgue rouge *Porphyridium* sp ou de la microalgue verte *C. reinhardtii*, ne présentent pas de voie de N-glycosylation faisant intervenir la GnT I. En effet, depuis notre étude, la structure des N-glycannes portés par une glycoprotéine pariétale de 66kDa de *Porphyridium sp* a été publiée (Levy-Ontman et al., 2011). Cette microalgue présente des structures glycanniques composées de 7 à 9 résidus mannose dont trois d'entre eux sont substitués par un groupement O-méthyl. Un résidu xylose a également été mis en évidence sur le mannose distal de la branche « c ». De même, un second résidu xylose semble pouvoir être ajouté sur le premier mannose du core glycannique de façon similaire à ce qui est présent chez les plantes. Par ailleurs, la caractérisation de la voie de N-glycosylation a été réalisée chez *C. reinhardtii* et publiée récemment par notre laboratoire (Mathieu-Rivet et al., 2013). Il a été montré que les protéines endogènes solubles et membranaires de cette microalgue verte possèdent principalement des N-glycannes oligomannosidiques allant de Man-2 à Man-5 (environ 70% des N-glycannes détectés). En outre, 30% des autres N-glycannes identifiés correspondent à des structures Man-3 à Man-5 modifiées par des résidus 6-O-méthyl et par un ou deux xyloses. Ces résidus xylose sont liés respectivement en β-1, 2 sur le core glycannique et en β-1, 4 sur un résidu mannose terminal d'une des branches glycanniques. Ainsi, contrairement à *P. tricornutum,* ces microalgues possèdent une voie de maturation des N-glycannes GnT I-indépendante. L'existence de structures comportant des motifs glucidiques spécifiques en l'absence d'activité GnT I avait déjà été mise en évidence dans des lignées cellulaires mutantes de mammifères (ex : CHO Lec1)(Crispin et al., 2006).

De par leur histoire évolutive, le génome des diatomées est une mosaïque de séquences génomiques provenant d'au moins quatre sources distinctes, ce qui est illustré par la coexistence chez ces microalgues de métabolismes spécifiques de plantes, d'animaux et de bactéries (Armbrust et al., 2004 ; Bowler et al., 2010 ; Chan et al., 2012). Ainsi, se pose la question de l'origine du gène de la GnT I chez les diatomées. Or, les processus de maturation GnT I-dépendante des N-glycannes sont généralement identifiés chez des eucaryotes

supérieurs. De plus, les résultats précédents indiquent que la maturation des N-glycannes complexes par une voie GnT I-dépendante n'est pas une caractéristique générale des microalgues. Enfin, la comparaison des séquences GnT I présentes chez les microalgues avec celles des animaux et des végétaux montre qu'elles possèdent un pourcentage d'identité similaire. Par conséquent, il est probable que le gène de la GnT I des diatomées provienne d'un ancêtre eucaryote hétérotrophe commun avec les animaux et les plantes. Cependant, la divergence avec ces organismes a dû avoir lieu très tôt car les GnT I des diatomées et des haptophytes forment un groupe bien distinct de celui formé par les GnT I des animaux et des plantes.

II- Fucosylation des N-glycannes de *Phaeodactylum tricornutum.*

La fucosylation est l'une des principales étapes de la maturation des N-glycannes qui diverge entre les organismes. Elle consiste en l'ajout d'un résidu fucose, à partir du GDP-fucose, sur une GlcNAc du N-glycanne par une fucosyltransférase. Chez les eucaryotes supérieurs, cette modification peut avoir lieu sur deux résidus différents du N-glycanne. D'une part, l'ajout du résidu fucose peut être effectué sur les GlcNAc terminales, participant à la formation de motifs glycanniques de type Lewis (Breton et al., 1998). Dans ce cas, il est transféré en α-1,3 chez les mammifères (de Vries et al., 2001) et chez les insectes (Kubelka et al., 1993), et en α-1,4 chez les plantes (Lerouge et al., 1998). D'autre part, la fucosylation peut également être réalisée au niveau de la GlcNAc proximale du core glycannique (Breton et al., 1998). Ce fucose est lié en α-1,3 chez les plantes (Johnson et Chrispeels, 1987), en α-1,6 chez les animaux (Voynow et al., 1991), et en α-1,3 et α-1,6 chez les invertébrés, présentant ainsi une bifucosylation de leurs N-glycannes (Staudacher et März, 1998). Les fucosyltransférases responsables de ces différentes réactions appartiennent à deux familles de glycosyltransférases bien distinctes. En effet, l'α-1,3-fucosyltransférase responsable de la fucosylation du core en α-1,3, et les α-1,3/4-fucosyltransférases catalysant l'ajout du fucose sur les branches terminales du N-glycanne, appartiennent à la même famille des glycosyltransférases GT10 (Breton et al., 1998). En revanche, le transfert du fucose en α-1,6 est, quant à lui, catalysé par une α-1,6-fucosyltransférase appartenant à la famille GT23.

Les recherches d'homologie de séquences menées sur le génome de *P. tricornutum* ont révélé la présence de trois gènes codant potentiellement pour des α-1,3-fucosyltransférases, mais à priori aucun correspondant à une α-1,6-fucosyltransférase. Les analyses bioinformatiques de leurs séquences indiquent que ces gènes coderaient pour des fucosyltransférases membranaires de type II appartenant à de la famille GT10. De plus, les

résultats de ces recherches montrent que les gènes 46109 et 46110 de *P. tricornutum* présentent plus de similarité avec les α-1,3-fucosyltransférases du core, et seraient donc potentiellement responsables de l'ajout du fucose détecté sur les N-glycannes de cette diatomée lors des analyses biochimiques réalisées au cours de cette thèse. Cependant, l'annotation du gène 46110 dans le génome de *P. tricornutum* semble erronée. En effet, l'intron n°1 prédit dans ce gène semble en réalité correspondre à une séquence codante. Ces résultats ont récemment été confirmés au laboratoire par l'amplification PCR des ADNc générés à partir des ARNm matures de ce gène.

A l'inverse, les analyses bioinformatiques indiquent que la protéine codée par le gène 54599 serait plus proche des α-1,3-fucosyltransférases d'insectes permettant la fucosylation des GlcNAc distales. Or, aucun N-glycanne $FucGlcNAcMan_3GlcNAc_2$ ou $FucGlcNAcMan_3FucGlcNAc_2$ correspondant au produit de cette enzyme n'a été observé dans nos analyses par spectrométrie de masse MALDI TOF réalisées sur des extraits glycanniques de la souche Pt1.8.6 de *P. tricornutum* cultivée en conditions standards. L'absence de ces épitopes glycanniques particuliers peut être due à une régulation de ce gène en fonction des conditions environnementales ou bien encore à l'action de N-acétylglucosaminidases comme exposé précédemment. En effet, les GlcNAc terminales sur lesquelles viennent potentiellement s'ajouter ce résidu fucose sont également la cible de ces glycosidases. Par conséquence, il est possible qu'il y ait une compétition entre ces deux enzymes possédant le même substrat, ce qui pourrait expliquer qu'aucun produit de la FuT54599 putative ne soit observé. Par ailleurs, l'intensité des ions paucimannosidiques observés sur les spectres de masse n'étant pas très élevée, il n'est pas exclu que de tels glycannes $FucGlcNAcMan_3GlcNAc_2$ ou $FucGlcNAcMan_3FucGlcNAc_2$ soient présents en très faible proportion et qu'ils n'aient pas pu être détectés au cours de nos analyses. Ainsi, il semble indispensable dans la poursuite de cette étude de perfectionner les étapes de préparation des N-glycannes de *P. tricornutum*, afin d'obtenir des spectres de masse présentant une meilleure sensibilité, ce qui pourrait éventuellement permettre l'observation d'une population supplémentaire de N-glycannes complexes tels que les produits de la GnT I et de la FuT54599. L'obtention de spectres de masse avec une meilleure sensibilité permettrait également de réaliser une fragmentation des structures complexes afin de les analyser de façon plus approfondie par MS^2, permettant ainsi de localiser précisément le résidu fucose de l'ion $Man_3FucGlcNAc_2$ sur le squelette glycannique et de déterminer définitivement le type de liaison dans laquelle il est impliqué.

Afin d'identifier la réelle fonction de ces gènes, la caractérisation fonctionnelle des trois fucosyltransférases putatives de *P. tricornutum* a été initiée au cours de cette thèse. Cette étude constitue désormais un objectif important de la thématique de recherche du laboratoire qui vise à compléter la compréhension des mécanismes golgiens de maturation des N-glycannes chez cette diatomée. En complément, des études de localisation et de cartographie golgiennes par microscopie électronique à transmission sur la GnT I et les FuT seront prochainement initiées, afin de déterminer si l'organisation spatiale des glycosyltransférases dans l'appareil de Golgi chez *P. tricornutum* est similaire à celle observée chez les plantes et les mammifères (Chevalier et al., 2010 ; Driouich et al., 2012).

Pour conclure, l'ensemble des résultats présentés dans ces travaux de thèse démontre l'existence d'une fucosylation des N-glycannes de diatomée qui diverge de celle observée chez les mammifères et qui serait en partie similaire à celle présente chez les invertébrés et les plantes.

III- Rôle des N-glycannes complexes chez *Phaeodactylum tricornutum : une possible fonction dans l'adaptation à un stress.*

Les N-glycannes complexes sont généralement caractéristiques des eucaryotes supérieurs et jouent un rôle essentiel dans le développement de ces organismes pluricellulaires (Sarkar et al., 2006 ; Schachter, 2009). Les fonctions de ces N-glycannes ont surtout été mises en évidence chez les mammifères notamment au travers de mutants de glycosylation ou de maladies liées à un déficit de glycosylation (CDG). En effet, chez les mammifères, y compris chez l'homme, les N-glycannes complexes sont physiologiquement indispensables car ils interviennent dans de nombreux processus biologiques faisant intervenir des phénomènes de reconnaissance et d'adhésion cellule/cellule, protéine/cellule ou ligand/récepteur par l'interaction de ces N-glycannes avec des lectines spécifiques (Varki, 1993 ; Freeze et al., 2012). En l'absence de telles structures, les mammifères sont sujets à de sévères pathologies comme dans les CDG, voire à de l'embryolétalité (Sun et al., 2005; Hennet, 2012 ; Freeze et al., 2012). Ainsi, il a été démontré que les souris mutées sur le gène de la GnT I et dépourvues de N-glycannes complexes, mouraient environ 10 jours après fertilisation et présentaient plusieurs anomalies morphologiques (Ioffe et Stanley, 1994 ; Metzler et al., 1994). Un défaut de fucosylation serait également impliqué dans l'apparition de pathologies chez les mammifères, et en particulier chez l'homme, donnant de sévères symptômes comme des retards mentaux et de croissance, des malformations au niveau des pieds, des phénomène de

schizophrénie et enfin des infections récurrentes entraînant généralement la mort de l'enfant dès son plus jeune âge (4-5 ans) (Fukuda et al., 2011).

Chez les invertébrés comme les insectes et les vers, il a également été mis en évidence que les N-glycannes complexes avaient un rôle essentiel dans leur développement. En effet, l'inactivation de la GnT I chez le ver et la mouche a montré une forte réduction de leur viabilité et des perturbation de leur activité locomotrice (Sarkar et al., 2006 ; Schachter, 2009).

A l'inverse, peu de données sont connues quant à la fonction de telles structures complexes chez les plantes. Aucun phénotype particulier n'a été initialement observé chez les mutants de plante déficients en activité GnT I comme le mutant *cgl I* d'*A. thaliana* quand ils sont cultivés en conditions normales en laboratoire (von Schaewen et al., 1993 ; Strasser et al., 2004). Néanmoins, il a été mis en évidence que ces mutants *cgl I* présentent un phénotype fort lorsqu'ils sont cultivées en conditions de stress salin, suggérant que les N-glycannes complexes de plante jouent un rôle dans des processus physiologiques spécifiques notamment en réponse à des stress environnementaux (von Schaewen et al., 2008 ; Kang et al., 2008). De même, il a été récemment démontré que l'inhibition de la GnT I chez le riz entraînait chez ce mutant un défaut de croissance des plantules et de leur racine, un problème dans la synthèse de la paroi et enfin une perturbation de la voie de signalisation par les cytokinines (Fanata et al., 2013).

Contrairement à ce qui a été montré chez les eucaryote supérieurs, l'absence de N-glycannes complexes dans des lignées cellulaires, telles que les lignées de cellules de mammifères CHO Lec 1 et CHO FuT-/-8, ne présente pas d'effet visible. Ainsi, de façon générale, les N-glycannes complexes semblent avoir fortement contribué à l'adaptabilité et à la réussite des organismes pluricellulaires au cours de l'évolution.

De la même manière, il serait intéressant de comprendre le rôle des structures glycanniques complexes chez les diatomées qui sont des organismes unicellulaires, soit par la création de mutants de glycosylation soit par des méthodes d'inactivation de gènes avec des ARN interférents (iRNA, miRNA, siRNA), techniques déjà développées et utilisées chez *P. tricornutum* (De Riso et al., 2009). Etant donné que les glycoprotéines extraites de la souche Pt1.8.6 cultivée en conditions standards sont principalement glycosylées par des N-glycannes oligomannosidiques, il semblerait que l'action *in vivo* de la GnT I dans cette culture soit limitée. Par conséquence, il serait intéressant à l'avenir, après avoir caractérisé l'activité enzymatique des N-acétylglucosaminidases potentielles, d'inactiver spécifiquement leurs gènes chez *P. tricornutum*. Cette approche permettrait de constater si des N-glycannes

complexes portant des résidus GlcNAc voire des motifs GlcNAc-Fuc en position terminale, commenceraient à s'accumuler, dans quel cas il serait intéressant d'évaluer l'impact de ces structures sur la physiologie des diatomées.

A l'inverse, une inactivation du gène de la GnT I de *P. tricornutum* permettrait d'évaluer l'impact d'une déficience en N-glycannes complexes observés dans la souche Pt1.8.6 sur la biologie des diatomées que ce soit en conditions standards mais également en conditions de stress. En effet, les diatomées sont des organismes connus pour réagir rapidement aux variations de leurs paramètres environnementaux par mise en place de voies métaboliques leur permettant de s'adapter à ces conditions de stress (Allen et al., 2008 ; Allen et al., 2011). Or, nos travaux ont mis en évidence la présence de transcrits de la GnT I dans la souche Pt1.8.6 de *P. tricornutum* cultivée en conditions standards, ainsi qu'une surexpression rapide et brève de ce gène en réponse à une carence en nitrate. Ceci suggère une possible implication des glycoprotéines portant des N-glycannes complexes dans la voie de signalisation et/ou dans la mise en place d'une réponse adaptative vis-à-vis de ces conditions de stress chez cette diatomée. Un impact de la variation du taux de nitrate dans le milieu de culture a déjà été montré sur plusieurs voies métaboliques chez *P. tricornutum* notamment au travers des analyses d'EST et du méthylome (Montsant et al., 2005 a ; Maheswari et al., 2009; Veluchamy et al., 2013). Nos résultats ont également permis de montrer, par une approche qualitative de RT-PCR, que le gène de la GnT I était exprimé dans les deux souches Pt1.8.6 et Pt3 de *P. tricornutum* ainsi que dans les trois morphotypes (fusiforme, ovale et triradié) de cette dernière souche. Cela suggère que cette diatomée est capable de synthétiser des N-glycannes complexes quel que soit le morphotype et peut être même la souche. Cependant, bien que les mécanismes moléculaires régissant le changement morphotypique de *P. tricornutum* soient encore inconnus, il a été montré que le passage d'un morphotype à un autre était généralement obtenu en réponse à des modifications environnementales comme les conditions de salinité. En effet, il a été mis en évidence *in vitro* qu'un stress salin pouvait engendrer le changement du morphotype fusiforme en ovale (De Martino et al., 2011). Le pléomorphisme de cette diatomée serait donc potentiellement un moyen pour elle de s'adapter rapidement aux variations qui l'entourent. Ainsi, il serait intéressant d'étudier dans les trois morphotypes non seulement le profil des N-glycannes, mais également l'expression du gène de la GnT I par une approche quantitative. Cela permettrait de mettre en évidence une potentielle différence dans la voie de maturation des N-glycannes en fonction du morphotype, suggérant alors l'importance de glycoprotéines portant des N-glycannes complexes dans le changement morphotypique, et donc par extension, le rôle potentiel des N-glycannes dans ce

procesus d'adaptation. Le laboratoire Glyco-MEV, conscient de sa particularité à posséder les trois morphotypes provenant de la même souche (Pt3), a d'ores et déjà entamé un projet transcriptomique en utilisant une approche de RNA sequencing (RNA-Seq) sur ces trois morphotypes. L'objectif de cette étude est de comprendre les mécanismes moléculaires et les voies métaboliques impliquées dans ces changements morphotypiques, d'identifier la présence de transcrits faiblement exprimés et enfin de mettre en évidence des mécanismes de régulation de l'expression de certains gènes. En particulier, ces travaux permettront de déterminer si les glycosidases et les glycosyltransférases de la voie de N-glycosylation sont exprimées de façon différentielle entre ces trois morphotypes, ce qui indiquerait la présence d'un système de régulation de ces protéines et soulignerait l'importance de la glycosylation dans la morphogénèse.

En conclusion, il est possible que, tout comme chez les eucaryotes supérieurs, les N-glycannes complexes et donc l'existence d'une voie de N-glycosylation GnT I-dépendante aient joué un rôle dans l'adaptation de ces organismes aux changements environnementaux sévères qui ont eu lieu au cours de l'évolution permettant ainsi la survie et la grande réussite écologique des diatomées.

IV- *Phaeodactylum tricornutum* : nouvelle usine cellulaire pour la production de glycoprotéines à intérêt thérapeutique.

La majorité des protéines à intérêt thérapeutique (95%), y compris les anticorps, sont des glycoprotéines. Ceci sous-entend que le système d'expression hétérologue, utilisé pour la production de ces molécules, doit être capable non seulement de synthétiser une partie peptidique identique à la protéine naturelle, mais également d'additionner des glycannes sur les sites de N-glycosylation de cette protéine. Ces N-glycannes sont d'autant plus importants qu'ils peuvent influencer l'activité et la fonctionnalité de la protéine à intérêt thérapeutique (Lingg et al., 2012) mais également poser problème lors de leur utilisation en thérapie(van Beers et Bardor, 2012), car les processus de N-glycosylation sont spécifiques des systèmes d'expression hétérologues utilisés. En effet, les principaux modèles utilisés ces dernières années pour la production de ces glycoprotéines, que sont les levures, les insectes, les mammifères et les plantes, sont capables de N-glycosyler une protéine, mais les structures des oligosaccharides matures diffèrent d'un système à l'autre (Jenkins et Curling, 1994 ; De Pourcq et al., 2010). Chez les plantes, il a été montré que leurs résidus spécifiques α-1,3-fucose et β-1,2-xylose présents sur le core glycannique étaient fortement immunogènes et

peuvent entrainer de fortes allergies voire un choc anaphylactique. Néanmoins, depuis une dizaine d'années, l'ingénierie des voies de N-glycosylation dans ce système s'est fortement développée. Il est possible aujourd'hui de produire chez les plantes des glycoprotéines à intérêt thérapeutique possédant des N-glycannes humanisés dépourvus d'épitopes immunogènes comme l'α-1,3-fucose et le β-1,2-xylose et présentant des caractéristiques de mammifères comme des acides sialiques (Castilho et al., 2013 ; Schneider et al., 2013 ; Jez et al., 2013).

Le développement de nombreux outils de biologie moléculaire et l'accessibilité au génome ainsi que leur facilité à être cultivée et leur rapidité de reproduction ont fait des diatomées de sérieux candidats pour la production de molécules biologiques à fort intérêt économique (Hempel et Maier, 2012 ; Niu et al., 2012 ; Miyahara et al., 2013). Par exemple, un anticorps monoclonal humain fonctionnel, spécifique d'une protéine de surface de l'hépatite B a récemment été produit chez *P. tricornutum* (Hempel et al., 2011a ; Hempel et Maier, 2012). Les tests d'activités montrent qu'il reconnait efficacement son antigène et les analyses biochimiques par affinodétection sur empreinte réalisées avec la concanavaline A indiquent que cet anticorps produit chez *P. tricornutum* est bien glycosylé. Le laboratoire Glyco-MEV a initié il y a un an une collaboration avec le Dr Hempel et travaille actuellement sur l'analyse fine des structures glycanniques associées à cet anticorps recombinant produit et sécrété par *P. tricornutum*.

Dans ce contexte, une stratégie d'ingénierie de la voie de N-glycosylation est envisageable chez *P. tricornutum*. En effet, la structure des N-glycannes semble relativement dépourvue de résidus spécifiques allergisants à l'exception du résidu α-1,3-fucose présent sur le core glycannique de certains glycannes mineurs. Une fois que l'identification du ou des gènes responsables de cette fucosylation aura été terminée, à l'issue de nos travaux de caractérisation fonctionnelle, ils pourront alors être inactivés par le biais d'ARN interférents (De Riso et al., 2009). De même, l'expression des N-acétylglycosaminyltransférases putatives identifiées chez *P. tricornutum* pourrait être modulée par un procédé similaire. Par ailleurs, des gènes codant pour des glycosyltransférases responsables de l'ajout de motifs spécifiques de mammifères, comme des sialyltransférases, des galactosyltransférases, ou des GnT III à V, pourraient également être insérés dans le génome de cette diatomée. Une telle ingénierie de voie métabolique a déjà été réalisée chez *P. tricornutum* afin de produire le polymère plastique Poly-3-hydroxybutyrate (PHB), par l'insertion de trois gènes codant pour des enzymes responsables de la synthèse de ce composé (Hempel et al., 2011b).

Une compétition internationale visant à élaborer le meilleur système modèle pour la production de molécules à haute valeur ajoutée comme les glycoprotéines à intérêt thérapeutique a été initiée il y a plusieurs décennies. Or, chaque système d'expression possède des avantages et des inconvénients. Le meilleur système pour une telle production, si tenté qu'il puisse y en avoir un, sera celui qui présente le moins d'inconvénients. Les recherches ont tout d'abord été développées chez les bactéries, les levures, les plantes terrestres et les animaux. Cependant, ces systèmes ne permettent pas de répondre complètement de façon satisfaisante aux différentes contraintes générées par une production à l'échelle industrielle de ces molécules d'intérêt.

Contrairement au système plante, les diatomées peuvent être cultivées dans des structures confinées (photobioréacteurs) et cela dans des espaces désertiques non cultivables ne rentrant donc pas en compétition avec la culture de plantes alimentaires (Nigh et al., 2000 ; Wolfenbarger et Phifer, 2000). De plus, leur temps de génération est court, leurs récoltes sont continues sur l'année (permaculture) et leurs cultures ne nécessitent pas l'apport de produits phytosanitaires ni d'engrais comme chez les plantes, reconnus comme étant onéreux et particulièrement nocifs pour l'environnement. De même, à l'instar des bactéries, les diatomées sont des organismes possédant une machinerie moléculaire complexe et sont donc capables de réaliser des modifications post-traductionnelles. Par ailleurs, contrairement aux animaux, les diatomées présentent l'avantage de ne pas transmettre de pathogènes à l'homme et possèdent un coût d'entretien largement inférieur. En effet, les diatomées ont uniquement besoin de dioxyde de carbone, d'eau, de sels inorganiques et de lumière pour se développer (Lopez et al., 2005). Pour finir, la production en masse est envisageable à faible coût ce qui n'est pas le cas des systèmes plantes et animaux. Néanmoins, la production à l'échelle industrielle de ce type de molécule chez *P. tricornutum* nécessite l'optimisation des systèmes de production notamment au niveau des étapes d'extraction et de purification, afin d'abaisser encore le coût de production et de devenir réellement compétitifs face aux modèles déjà existants. Bien que les outils de biologie moléculaire de cette diatomée se soient fortement développés ces dernières années, alors que ce modèle est récent, ils restent encore limités en comparaison de ceux disponibles dans les autres systèmes. De même, les processus physiologiques et cellulaires tels que la sécrétion sont encore mal connus.

En conclusion, *P. tricornutum* est un modèle qui a déjà été utilisé pour la production de protéines à intérêt thérapeutique dans le cadre d'études expérimentales servant pour le moment de preuves de concept. Cette diatomée présente des caractéristiques très intéressantes

qui permettent de l'envisager comme une usine cellulaire compétitive à l'échelle industrielle qui n'a rien à envier aux autres systèmes d'expression déjà utilisés.

Conclusion générale

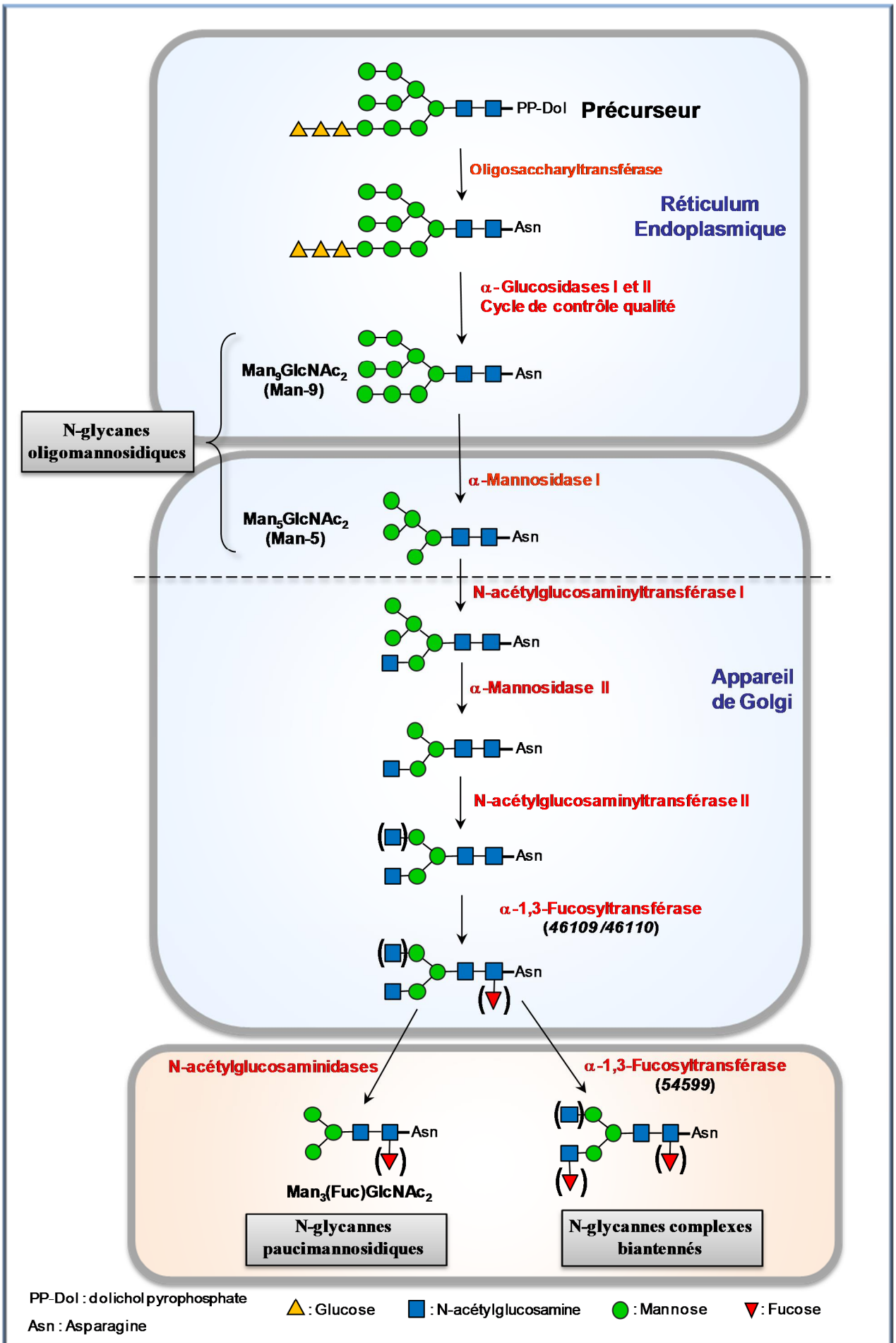


Figure 50 : Schéma récapitulatif de la voie de N-glycosylation de *Phaeodactylum tricornutum* établie à partir des données bioinformatiques et biochimiques obtenus au cours de cette thèse. La ligne en pointillé délimite les N-glycannes oligomannosidiques des N-glycannes complexes. Le numéro des gènes codant pour les α-1,3-fucosyltransférases est indiqué entre parenthèse en italique noir. Les N-glycannes observés sont écrits en gras noir.

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, l'objectif plus spécifique de ma thèse était d'étudier la voie de N-glycosylation de la diatomée pennée polymorphique *P. tricornutum.*

Les données bioinformatiques du génome de *P. tricornutum* et les analyses biochimiques menées sur les glycoprotéines de la souche Pt1.8.6 cultivée en conditions standards indiquent que les étapes de la voie de N-glycosylation localisées au niveau du RE sont également conservées chez cette diatomée et que la maturation des N-glycannes de cette microalgue aboutit principalement à la formation de N-glycannes oligomannosidiques allant de Man-9 à Man-5 (**figure 50**). Il a également été démontré que cette diatomée possède une voie de N-glycosylation GnT I-dépendante permettant la formation de N-glycannes paucimannosidiques tels que Man-3, Man-4 et $Man_3FucGlcNAc_2$. La formation de ces paucimannoses serait similaire à celle retrouvée chez les plantes et les invertébrés et ferait intervenir une GnT I, qui a été caractérisée d'un point de vue fonctionnel au cours de cette thèse, une α-mannosidase II, potentiellement une GnT II, des α-1,3-fucosyltransférases (FuT46109 et FuT46110) putatives capable de fucosyler le core glycannique en α-1,3 et enfin des N-acétylglucosaminidases (**figure 50**). Une fucosylation des GlcNAc terminales, comme dans le motif de type Lewis, semble également envisageable par une α-1,3-fucosyltransférase putative (FuT54599), bien que le produit de ces enzymes n'ait pas encore été identifié sur les glycoprotéines endogènes de la souche Pt1.8.6 de *P. tricornutum* cultivée en conditions standards.

Par ailleurs, l'étude de l'expression du gène de la GnT I a montré que ce gène était transcrit dans les trois morphotypes de *P. tricornutum*, ainsi que dans deux de ces souches P.t1.8.6 et Pt3. Ces analyses ont également permis de montrer, par PCR quantitative en temps réel, que ce gène semblait être surexprimé en réponse à une carence en nitrate dans le milieu de culture, suggérant une possible implication des N-glycannes complexes dans l'adaptation des diatomées face à des stress exercés dans leur environnement.

Annexes

Annexe 1

Les tableaux suivants contiennent les numéros des gènes correspondant aux séquences utilisées pour les analyses de recherche d'homologie et pour la recherche de motifs spécifiques des différentes protéines impliquées dans la voie de N-glycosylation. Les organismes de référence utilisés sont *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Physcomitrella patens*, *Zea mays* et *Oryza sativa*.

Organismes	Alg 1	Alg 2	Alg 3
Homo sapiens	56052	85365	10195
Mus musculus	208211	56737	208624
Drosophila melanogaster	Dmel_CG18012	Dmel_CG1291	Dmel_CG4084
Caenorhabditis elegans	T26A5.4	F09E5.2	K09E4.2
Saccharomyces cerevisiae	YBR110W	YGL065C	YBL082C
Physcomitrella patens	PHYPADRAFT_50684	PHYPADRAFT_222991	PHYPADRAFT_13871
Arabidopsis thaliana	AT1G16570	AT1G78800	AT2G47760
Zea mays	100282626	100279297	100276684
Oryza sativa	4331831	4336813	4327643

Organismes	Alg 5	Alg 6	Alg 7
Homo sapiens	29880	29929	1798
Mus musculus	66248	320438	13478
Drosophila melanogaster	Dmel_CG7870	Dmel_CG5091	Dmel_CG5287
Caenorhabditis elegans	H43I07.3	C08B11.8	Y60A3A.14
Saccharomyces cerevisiae	YPL227C	YOR002W	YBR243C
Physcomitrella patens	PHYPADRAFT_115941 / PHYPADRAFT_132439	PHYPADRAFT_149102	PHYPADRAFT_171226
Arabidopsis thaliana	AT2G39630	AT5G38460	AT2G41490 / AT3G57220
Zea mays	100273013	100278369	100274497
Oryza sativa	4334604	4342475	4344180

Organismes	Alg 8	Alg 9	Alg 10
Homo sapiens	79053	79796	144245 / 84920
Mus musculus	381903	102580	380959
Drosophila melanogaster	Dmel_CG4542	Dmel_CG11851	Dmel_CG32076
Caenorhabditis elegans	C08H9.3	C14A4.3	T24D1.4
Saccharomyces cerevisiae	YOR067C	YNL219C	YGR227W
Physcomitrella patens	PHYPADRAFT_179078	PHYPADRAFT_173520	PHYPADRAFT_148859
Arabidopsis thaliana	AT2G44660	AT1G16900	AT5G02410
Zea mays		100282615	
Oryza sativa	4330363	4325533	4336325

Organismes	Alg 11	Alg 12	Alg 13
Homo sapiens	440138	79087	79868
Mus musculus	207958	223774	67574
Drosophila melanogaster	Dmel_CG11306	Dmel_CG8412	Dmel_CG14512
Caenorhabditis elegans	B0361.8	ZC513.5	R10D12.12
Saccharomyces cerevisiae	YNL048W	YNR030W	YGL047W
Physcomitrella patens	PHYPADRAFT_196599	PHYPADRAFT_206317	PHYPADRAFT_145686
Arabidopsis thaliana	AT2G40190	AT1G02145	AT4G16710
Zea mays	100274516	100282442	100217271 / 100280032
Oryza sativa	4352649	4336560	4329330

Organismes	DPM 1	DPM 2	DPM 3
Homo sapiens	8813	8818	54344
Mus musculus	13480	13481	68563
Drosophila melanogaster	Dmel_CG10166		Dmel_CG33977
Caenorhabditis elegans	Y66H1A.2		F28D1.11
Saccharomyces cerevisiae	YPR183W	YIL102C-A	
Physcomitrella patens	PHYPADRAFT_89385	PHYPADRAFT_147780	
Arabidopsis thaliana	AT1G20575	AT1G74340	AT1G48140
Zea mays			
Oryza sativa	4334626	9269261	9272161

Organismes	Alg 14	RFT I
Homo sapiens	199857	91869
Mus musculus	66789	328370
Drosophila melanogaster	Dmel_CG6308	Dmel_CG3149
Caenorhabditis elegans	M02B7.4	ZK180.3
Saccharomyces cerevisiae	YBR070C	YBL020W
Physcomitrella patens	PHYPADRAFT_142865	PHYPADRAFT_117127
Arabidopsis thaliana	AT4G18230	AT5G07630
Zea mays	100285424	100382966
Oryza sativa	4333143	4343254

Organismes	OST 1	OST 2	OST 3
Homo sapiens	6184	1603	7991
Mus musculus	103963	13135	80286
Drosophila melanogaster	Dmel_CG33303	Dmel_CG13393	Dmel_CG7830
Caenorhabditis elegans	T22D1.4	F57B10.10	ZK686.3
Saccharomyces cerevisiae	YJL002C	YOR103C	YML019W / YOR085W
Physcomitrella patens	PHYPADRAFT_123403	PHYPADRAFT_227891 / PHYPADRAFT_74146	PHYPADRAFT_118668 / PHYPADRAFT_60432
Arabidopsis thaliana	AT1G76400	AT1G32210	AT1G11560 / AT1G61790
Zea mays	100285254		100282286
Oryza sativa	4338348	4335701	4333265

Organismes	OST 4	OST 5	STT 3
Homo sapiens			201595 / 3703
Mus musculus			16430 / 68292
Drosophila melanogaster			Dmel_CG1518
Caenorhabditis elegans			T12A2.2
Saccharomyces cerevisiae	YDL232W	YGL226C-A	YGL022W
Physcomitrella patens			PHYPADRAFT_186785 / PHYPADRAFT_200115
Arabidopsis thaliana			AT1G34130 / AT5G19690
Zea mays			100192699 / 100279881
Oryza sativa			4337389

Organismes	SWP 1	WBP 1	Glucosidase I
Homo sapiens	6185	1650	7841
Mus musculus	20014	13200	57377
Drosophila melanogaster	Dmel_CG6370	Dmel_CG9022	Dmel_CG1597
Caenorhabditis elegans		T09A5.11	F13H10.4
Saccharomyces cerevisiae	YMR149W	YEL002C	YGL027C
Physcomitrella patens	PHYPADRAFT_6306	PHYPADRAFT_234790	PHYPADRAFT_159281
Arabidopsis thaliana	AT4G21150	AT5G66680	AT1G24320
Zea mays	100273881 / 100279928	100191802 / 100274092	
Oryza sativa	4324876	4342702	4327748

Organismes	Glucosidase II	Calnexine	Calréticuline
Homo sapiens	23193	821	125972 / 811
Mus musculus	14376	12330	108802 / 73316 / 12317
Drosophila melanogaster	Dmel_CG14476	Dmel_CG9906 / Dmel_CG11958	Dmel_CG9429
Caenorhabditis elegans	CELE_F40F9.6 / CELE_F52D1.1	ZK632.6	Y38A10A.5
Saccharomyces cerevisiae	YBR229C	YAL058W	
Physcomitrella patens	PHYPADRAFT_187728	PHYPADRAFT_118597 / PHYPADRAFT_59289 / PHYPADRAFT_191189	PHYPADRAFT_164102
Arabidopsis thaliana	AT5G63840	AT5G61790 / AT5G07340	AT1G09210 / AT1G56340 / AT1G08450
Zea mays	100383593	100283947	542732
Oryza sativa	4332068	4335732	4334509 / 4334675

Organismes	Man I	GnT I
Homo sapiens	10905/11253/4121/57134	4245
Mus musculus	17155 / 17156 / 227619 / 230815	17308
Drosophila melanogaster	Dmel_CG11874 / Dmel_CG31202 / Dmel_CG42275	Dmel_CG13431
Caenorhabditis elegans	CELE_D2030.1 / CELE_T03G11.4 / CELE_ZC410.3	CELE_B0416.6 / CELE_F48E3.1 / CELE_M01F1.1
Saccharomyces cerevisiae	YJR131W / YLR057W	
Physcomitrella patens	PAS_chr2-1_0753	PHYPADRAFT_214349
Arabidopsis thaliana	AT1G30000 / AT1G51590 / AT3G21160	AT4G38240
Zea mays	100282405 / 100502479	100274447
Oryza sativa	4336917 / 4338088	4331263

Organismes	Man II	GnT II	GnT III à VI
Homo sapiens	4122 / 4124	4247	**Voir annexe 2**
Mus musculus	140481 / 17158	217664	
Drosophila melanogaster	Dmel_CG18802	Dmel_CG7921	
Caenorhabditis elegans	CELE_F58H1.1	CELE_C03E10.4	
Saccharomyces cerevisiae			
Physcomitrella patens	PHYPADRAFT_159159	PHYPADRAFT_113818/PHYPADRAFT_226628	
Arabidopsis thaliana	AT5G14950	AT2G05320	
Zea mays	100502340	100381400	
Oryza sativa	4340636	4331021	

Organismes	β-1,3-GalT	β-1,4-GalT	β-1,3-xylt
Homo sapiens		2683/8703/8704	
Mus musculus		14595/53418/57370	
Drosophila melanogaster			
Caenorhabditis elegans			
Saccharomyces cerevisiae			
Physcomitrella patens			
Arabidopsis thaliana	839224		835643
Zea mays			732746
Oryza sativa			4345969

Organismes	α-1,3-FuT core	α-1,6-FuT core	α-1,3/4-FuT Lewis
Homo sapiens		2530	2528/2526/2525/10690/2527
Mus musculus		53618	14345/14348
Drosophila melanogaster	39653	32122	3355171
Caenorhabditis elegans	174400	178985	
Saccharomyces cerevisiae			
Physcomitrella patens			
Arabidopsis thaliana	821462/841394		843530
Zea mays	778434		
Oryza sativa			

Colonne1	SialT	N-acétylglucosaminidase
Homo sapiens	**Voir annexe 2**	**Voir annexe 2**
Mus musculus		
Drosophila melanogaster		
Caenorhabditis elegans		
Saccharomyces cerevisiae		
Physcomitrella patens		
Arabidopsis thaliana		
Zea mays		
Oryza sativa		

Annexe 2

Brevet PCT/EP2011/003756, Université de Rouen/ IFREMER/ Algenics. Licence d'exploitation avec la société Algenics, Nantes en cours de négociation. (**2011**) :

N-GLYCOSYLATION IN TRANSFORMED *PHAEODACTYLUM TRICORNUTUM.*

Bardor M., Louvet R., Saint-Jean B., Burel C., Baïet B., Carlier A., Cadoret J-P. and Lerouge P.

1

N-GLYCOSYLATION IN TRANSFORMED PHAEODACTYLUM TRICORNUTUM

FIELD OF THE INVENTION

The present invention is directed to methods for improving the N-glycosylation pattern of proteins in *Phaeodactylum tricornutum* in order to produce in these microalgae glycoproteins having N-glycan profiles similar to those of glycoproteins produced by animal cells, especially human cells, which can be used as human or animal therapeutic agents.

BACKGROUND OF THE INVENTION

Pharmaceutical proteins are produced as recombinant proteins by expression in eukaryotic expression systems. After the synthesis of the protein backbone, the recombinant protein is submitted to further post-translational processing including the attachment of sugar residues, a process known as glycosylation. However, eukaryotic organisms exhibit different glycosylation processing involving specific enzymes (glycosyltransferases and glycosidases), and so that the glycosylation patterns, even of the same protein, will be different depending on the eukaryotic cell in which the particular protein is being produced. Thus, the glycosylation pattern of pharmaceutical proteins expressed in eukaryotic host cells differs substantially from the glycosylation pattern of the natural proteins produced in humans and other mammals.

N-glycosylation: a major post-translational modification of secreted proteins

N-glycosylation is a major post-translational modification step in the synthesis of proteins in eukaryotes. N-glycan processing in the secretory pathway is essential for proteins intended to be secreted or integrated into membranes. N-glycosylation starts when the protein is translated and translocated from the ribosome into the lumen of the endoplasmic reticulum (ER). In this processing, a dolicholphosphate oligosaccharide precursor ($Glc_3Man_9GlcNAc_2$-PP-dolichol) is initially assembled at the cytoplasmic face and finished in the luminal face of the ER membrane (BURDA AND AEBI, *Biochimica et Biophysica Acta*, vol.1426, p: 239-257, 1999). This precursor is used by the oligosaccharyltransferase (OST) multisubunit complex that catalyses its transfer onto the asparagine residues of the consensus sequences Asn-X-Ser/Thr, when X is different than proline and aspartic acid, of a target protein (BURDA AND AEBI, above mentioned, 1999). The precursor is then deglucosylated/reglucosylated to ensure the quality control of the neosynthesised protein through the interaction with ER-resident chaperones calreticulin and calnexin. These ER

events are crucial for proper folding and oligomerization of secreted proteins (HELENIUS AND AEBI, *Science*, vol. 291, p: 2364-2369, 2001), highly conserved in eukaryotes investigated so far. These steps lead to the formation of a limited set of high-mannose-type N-glycans (Fig. 1). In contrast, evolutionary adaptation of N-glycan processing in the Golgi apparatus has given rise to a large variety of organism-specific complex structures. Mannosidases located in this compartment first degrade the oligosaccharide precursor into high-mannose-type N-glycans ranging from $Man_9GlcNAc_2$ (Man-9) to $Man_5GlcNAc_2$ (Man-5). N-acetylglucosaminyltransferase I (GnT I) then transfers a first GlcNAc residue on the β(1,3)-mannose arm of $Man_5GlcNAc_2$ and opens the door to the synthesis of multiple structurally different complex-type N-glycans (Fig. 1). Then, the actions of α-mannosidase II and GnT II allow the synthesis of the core $GlcNAc_2Man_3GlcNAc_2$. The complex-type oligosaccharides arise from the transfer in the Golgi apparatus of monosaccharide residues onto the core $GlcNAc_2Man_3GlcNAc_2$ under the action of organism-specific glycosyltransferases. As a consequence, mature proteins leaving the secretory pathway harbour multiple organism-specific complex N-glycans allowing the protein to acquire a set of glycan-mediated biological functions. As illustration, N-glycans in mammals are maturated into poly-antennary, poly-sialylated structures harbouring an α(1,6)-linked fucose residue on the proximal N-acetylglucosamine of the core (core-α(1,6)-fucose) (Fig. 1).

Remodelling into human-like N-glycans by knock-in strategies

Since glycosylation profiles differs between mammals and eukaryotic host cells, strategies have been developed for the *in vivo* remodelling of the protein N-linked glycan structures. These strategies include the knock-out of endogenous genes that are involved in the transfer of some specific monomers, and knock-in methodologies based on the expression in the host cells of mammalian enzymes. The knock-in approach results, by complementing the enzyme repertoire of the host cell, in the synthesis in the recombinant expression system of N-linked glycans similar to those found in mammalian cells. As illustration, the remodelling of plant N-glycans into mammalian-like N-glycans has been achieved by expressing a human β(1,4)-galactosyltransferase in plant cells. Targeted insertion of the human β(1,4)-galactosyltransferase in *Physcomitrella patens* has also been carried out leading to the addition of terminal β(1,4)-galactose to endogenous N-glycans. Human N-acetylglucosaminyltransferase III (GnT III) has also been successfully expressed in plants in order to *in planta* engineer endogenous N-glycans. This transferase is able to introduce

β(1,4)-GlcNAc residue on the β-mannose of the core mammalian N-glycans (bisecting GlcNAc).

With the exception of IgG, human serum proteins require sialic acid on terminal positions of their N-glycans (Fig. 1). Most non-mammalian eukaryotic cells, such as plants, do not synthesize N-acetyl- and N-glycolylneuraminic acids (Neu5Ac and Neu5Gc), the two main mammalian sialic acids. As consequence, genetic manipulation has been developed for the *in planta* synthesis of sialylated proteins by expressing enzymes able to synthesise CMP-Neu5Ac, its Golgi transporter and the appropriate sialyltransferases (Paccalet et al., 2007, *Plant Biotechnology Journal*, vol.5, p12-25; Castilho *et al.,* 2008, Plant Physiol., vol. 147 (11), p331-339; Castilho *et al.,* 2010, The journal of Biological Chemistry, vol.285 (21), p15923-30.

BRIEF DESCRIPTION OF DRAWINGS

The figure 1 shows the biosynthesis of N-linked glycans in mammals.

The figure 2 shows the affinodetection using concanavalin A (Con A) and immunodetection using antibodies raised against the core β(1,2)-xylose (anti-Xyl) and core α(1,3)-fucose (anti-Fuc) epitopes of proteins isolated from green onion used as a positive control (Lane 1) and from *Phaeodactylum tricornutum* (Lane 2).

The Figure 3 shows the transformation vector BSJ-25 (SEQ ID N°79). The expression cassette is made up of a double cauliflower mosaic virus (CaMV) 35S promoter (SEQ ID N°80), a tobacco mosaic virus-Ω sequence as translation enhancer and an Arabidopsis chitinase signal peptide (SP) (SEQ ID N°81) which are placed upstream of the multi-clonal site (MCS) and the sequence of *FcpA* terminator of *Phaeodactylum tricornutum* at the end. Selection marker, the bleomycin-resistant gene, *ShBle* is driven by the *FcpB* promoter. The construct also contains the ampicillin-resistant gene, *Amp* and the *Escherichia coli* replication origin.

The figure 4 shows the expression cassette of *Phaeodactylum tricornutum* GnT I. The first cassette comprised the sequence coding for the endogenous GnT I placed under the control of endogenous regulatory sequences. In the second construction, the GnT I was fused to the Green Fluorescent Protein (GFP).

The figure 5 shows the screening of transformed microalgae with the sequence coding for the *Phaeodactylum tricornutum* GnT I (*Pt*-GnT I). *Pt*-GnT I : potentially transformed

microalgae; T- : water amplification negative control; T+ : PCR amplification realised on the transformation vector.

The figure 6 shows PCR amplifications of the H4 housekeeping gene realised on the cDNA obtained from the *P. tricornutum* lineages which where transformed with the sequence coding for the GnT I protein (*Pt*-GnT I). *Pt :* wild strain of *P. tricornutum*; T- : water amplification negative control; RT+: PCR amplification realised on the cDNA; RT- : Reverse Transcription negative control.

The figure 7 shows PCR amplification of GnT I realised on the cDNA obtained from *P. tricornutum* which were transformed with the sequence coding for the endogenous GnT I (*Pt*-GnT I). *Pt :* wild strain of *P. tricornutum*; T- : water amplification negative control; RT+: PCR amplification realised on the cDNA; RT- : Reverse Transcription negative control.

The figure 8 shows Q-PCR realised on a wild type *P. tricornutum* lineage and on a lineage expressing the GnT I transgene. A: Fusion curve of the amplicons; B: amplification profiles of the wild-type and transformed lineages of *P. tricornutum*. *Pt*-GnT I : amplification curves obtained for 3 dilutions of cDNA from the transformed lineage with the GnT I gene. *Pt*-wt : amplification curves obtained for 3 dilutions of cDNA from the wild-type lineage.

The figure 9 shows confocal microscopy observation of *P. tricornutum* microalgae which were transformed with the genes coding for a cytosolic Green Fluorescent Protein (*Pt*-eGFP) and a GnT I-eGFP fusion protein (*Pt*-GnT I-eGFP).

The figure 10 shows the fluorescence observed in the Chloroplast and Endoplasmic Reticulum of *P. tricornutum* microalgae which express the eGFP protein and the fluorescence observed in the Golgi apparatus of microalgae expressing GnT I-eGFP, after a culture of said microalgae in brefeldin A (an agent capable of dismantling the structure of the Golgi apparatus by blocking the transport from the Endoplasmic Reticulum to the Golgi apparatus).

The figure 11 shows the detection of the GnT I-eGFP fusion protein by Western Blot. The *Pt*-eGFP sample corresponds to a protein extract from *P. tricornutum* which produce cytosolic eGFP protein. The negative sample (*Pt*) corresponds to protein extract from a non-transformed *P. tricornutum* microalga. Finally, the *Pt*-GnT I-eGFP correspond to protein extracts from two *P. tricornutum* clones which were transformed with a sequence coding for the GnT I-eGFP fusion protein and presenting fluorescence at the Golgi apparatus.

The figure 12 shows the constructs used for the expression of GnT I in *P. tricornutum*.

The figure 13 shows High mannose-type N-glycans are the main oligosaccharides N-linked to *P. tricornutum* proteins. (A) MALDI-TOF mass spectrum of N-linked glycans released by PNGase A from glycoproteins of *P. tricornutum* and labelled with 2-aminobenzamide (2-AB). (B) MALDI-TOF mass spectrum of the pool of N-glycans after treatment with Jack bean a-mannosidase. (C) MALDI-TOF mass spectrum of 2-AB-labelled N-linked glycans released by PNGase F from glycoproteins of *P. tricornutum*. Man-3 to Man-9: paucimannose and high mannose-type N-glycans $Man_3GlcNAc_2$ to $Man_9GlcNAc_2$. *: contaminants. ■: potassium adducts.

The figure 14 shows *P. tricornutum* GnT I complements N-glycan maturation deficiency in CHO Lec1 mutant. MALDI-TOF mass spectra of glycans N-linked to proteins extracted from CHO cells. (A) wild-type, (B) CHO Lec1 mutant and (C) transformant 4 of CHO Lec1 mutant complemented with *P. tricornutum* GnT I gene. Man-4 to Man-9: high mannose-type N-glycans $Man_4GlcNAc_2$ to $Man_9GlcNAc_2$. Symbols (70): black square with black outline: GlcNAc, grey circle with black outline: Man, white circle with black outline: Gal, grey triangle with black outline: Fucose.

SUMMARY OF THE INVENTION

A first aspect of the invention concerns a transformed *Phaeodactylum tricornutum* whose N-glycosylation pathway has been modified by the inactivation of at least one β-N-acetylglucosaminidase and/or the expression of at least one glycosylation enzyme encoded by a nucleic acid sequence operatively linked to a promoter, wherein

(i) said nucleic acid sequence encodes an N-acetylglucosaminyltransferase I having the amino acid sequence SEQ ID N°1, a fragment or a derivative thereof and said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°3 or a fragment thereof, and/or

(ii) said nucleic acid sequence encodes an α-Mannosidase II having the amino acid sequence SEQ ID N°5, a fragment or a derivative thereof and said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°7 or a fragment thereof.

In a preferred embodiment, said *P. Tricornutum* further comprises another nucleic acid sequence operatively linked to a promoter, said other nucleic acid sequence encoding a polypeptide that is expressed and glycosylated in the transformed *P. tricornutum*.

A second aspect of the invention concerns a method for producing a glycosylated polypeptide, said method comprising the steps of

(i) Culturing a transformed *P. tricornutum* as disclosed previously,

(ii) Purifying said polypeptide that is expressed and glycosylated in said transformed *P. tricornutum*.

In a preferred embodiment, said method comprises a further step (iii) of determining the glycosylation pattern of said polypeptide.

A third aspect of the invention concerns a use of a transformed *P. tricornutum* as disclosed previously for producing a glycosylated polypeptide.

DETAILED DESCRIPTION OF THE INVENTION

A gene encoding a putative N-acetylglucosaminyltransferase I (GnT I) has been predicted in the *Phaeodactylum tricornutum (P. tricornutum)* genome, but the inventors established that this putative GnT I does not exhibit any significant activity in *Phaeodactylum tricornutum* under standard culture conditions, as proteins extracted from *Phaeodactylum tricornutum* did not exhibit any GlcNAcMan5GlcNAc2 glycosylation pattern and carry about 95-97% of high-mannose-type N-glycans ranging from $Man_9GlcNAc_2$ (Man-9) to $Man_5GlcNAc_2$ (Man-5).

However, the inventors surprisingly found in further experiments, presented in the following examples, that this putative GnT I was able to restore the maturation of N-linked glycans into complex-type N-glycans in CHO *Lec1* mutants, defective in their endogenous GnT I.

Consequently, the inventors show that this putative GnT I has an enzymatic activity, which enzymatic activity can restore defective mammalian GnT I activity.

Moreover, The N-acetylglucosaminyltransferase I is not the only enzyme identified in the genome of *Phaeodactylum tricornutum* by the inventors. A further gene encoding an alpha-Mannosidase II (α-Man II) has been identified in the genome of *Phaeodactylum tricornutum*, whereas there was no detectable of α-Man II activity in *P. tricornutum* under standard culture conditions.

$GlcNAcMan_5GlcNAc_2$, the product of GnT I, is successively converted in the Golgi apparatus into $GlcNAcMan_4GlcNAc_2$ and then $GlcNAcMan_3GlcNAc_2$ by the action of the α-Man II, followed by the production of $GlcNAc_2Man_3GlcNAc_2$ under the action of GnT II.

Nevertheless, some organisms express β-N-acetylglucosaminidases, which are enzymes responsible for the degradation of GlcNAc-terminated N-glycans after their biosynthesis in the Golgi apparatus with the action of GnT I and Man II. Elimination of terminal GlcNAc by β-N-acetylglucosaminidases in the secretory system or in compartments where proteins accumulate can then convert these oligosaccharides into $Man_4GlcNAc_2$ and $Man_3GlcNAc_2$, thus annealing the Glycosylation pathway.

The inventors identified genes encoding putative β-N-acetylglucosaminidases in the genome *of Phaeodactylum tricornutum*: a first β-N-acetylglucosaminidase of amino acid sequence SEQ ID N°9 is encoded by the nucleic acid sequence SEQ ID N°10 (Accession number 45073), whereas a second β-N-acetylglucosaminidase of amino acid sequence SEQ ID N°11 is encoded by the nucleic acid sequence SEQ ID N°12 (Accession number 49563).

These putative genes may explain the absence of the detectable amount of $GlcNAcMan_5GlcNAc_2$, the product of GnT I, or $GlcNAcMan_4GlcNAc_2$ and $GlcNAcMan_3GlcNAc_2$, products of GnT I and α-Man II, on *P. tricornutum* proteins.

Therefore, a first object of the invention is a transformed *Phaeodactylum tricornutum* whose N-glycosylation pathway has been modified by the inactivation of at least one β-N-acetylglucosaminidase and/or the expression of at least one glycosylation enzyme encoded by a nucleic acid sequence operatively linked to a promoter, wherein

(i) said nucleic acid sequence encodes an N-acetylglucosaminyltransferase I having the amino acid sequence SEQ ID N°1, a fragment or a derivative thereof and said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°3 or a fragment thereof, and/or

(ii) said nucleic acid sequence encodes an α-Mannosidase II having the amino acid sequence SEQ ID N°5, a fragment or a derivative thereof and said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°7 or a fragment thereof.

In a first embodiment, said transformed *Phaeodactylum tricornutum* comprises a nucleic acid sequence operatively linked to a promoter, wherein:

(i) said nucleic acid sequence encodes an N-acetylglucosaminyltransferase I having the amino acid sequence SEQ ID n°1, a fragment or a derivative thereof; and

(ii) said operatively linked promoter has a sequence identity of less 50% with SEQ ID n°3 or a fragment thereof.

Phaeodactylum tricornutum is a microalga which belongs to the Bacillariophyceae class, to the Naviculales order, to the Phaeodactylaceae family and to the Phaeodactylum genus.

The term "nucleic acid sequence" used herein refers to DNA sequences (e.g., cDNA or genomic or synthetic DNA), as well as analogs of DNA containing non-natural nucleotide analogs, non-native internucleoside bonds, or both. Preferably, said nucleic acid sequence is a DNA sequence. This nucleic acid sequence can be in any topological conformation, like linear or circular.

The expression "Operatively linked" promoter refers to a linkage in which the promoter is contiguous with the gene of interest to control the expression of said gene.

The expression “fragment” with reference to SEQ ID N°3 refers to a nucleic acid sequence of at least 100 nucleic acids of said SEQ ID N°3, preferably of at least 150 nucleic acids of SEQ ID N°3, most preferably of at least 200 nucleic acids of SEQ ID N°3.

The term "transformed *Phaeodactylum tricornutum*" refers to a *P. tricornutum* wherein the nucleic acid sequence operatively linked to a promoter has been introduced in said microalgae by conventional methods of transformation, as described below, so as to express said nucleic acid molecule in the nucleus of said *P. tricornutum*.

Transformation of *P. tricornutum* can be carried out by conventional methods such as microparticles bombardment, electroporation, glass beads, polyethylene glycol (PEG), silicon carbide whiskers, or use of viruses or agrobacterium. Such a protocol is disclosed in the examples. The nucleic acid sequence may be introduced into *Phaeodactylum tricornutum via* a plasmid, virus sequences, double or simple strand DNA, circular or linear DNA. It is generally desirable to include into each nucleic acid sequence or vector at least one selectable marker to allow selection of *Phaeodactylum tricornutum* that have been stably transformed. Examples of such markers are antibiotic resistant genes such as *sh ble* gene enabling resistance to zeocin, *nat* or *sat-1* genes enabling resistance to nourseothricin, *bar* gene enabling resistance to glufosinate.

N-acetylglucosaminyltransferase I, also known as GnT I or mannoside acetylglucosaminyltransferase I (MGAT I) is an enzyme from the N-glycosylation pathway,

which is capable of adding an N-acetylglucosamine (GlcNAc) residue to $Man_5GlcNAc_2$ to produce a $GlcNAcMan_5GlcNAc_2$.

The N-acetylglucosaminyltransferase I having the amino acid sequence SEQ ID N°1 corresponds to the "endogenous N-acetylglucosaminyltransferase" encoded by the nucleic acid sequence SEQ ID N°2 present in the genome of wild-type *Phaeodactylum tricornutum*.

The expression "fragment" with reference to SEQ ID N°1 refers to an amino acid sequence comprising at least SEQ ID N°4 corresponding to the amino acid sequence of the catalytic site of N-acetylglucosaminyltransferase I having said SEQ ID N°1.

The expression "catalytic site" refers to the amino acid sequence of an enzyme, said amino acid sequence being responsible for the enzymatic activity. Therefore, the catalytic site of the N-acetylglucosaminyltransferase I according to the invention corresponds to the amino acid sequence responsible for the addition of an N-acetylglucosamine (GlcNAc) residue to $Man_5GlcNAc_2$ to produce a $GlcNAcMan_5GlcNAc_2$.

As used herein the term "N-acetylglucosaminyltransferase I derivative" refers to an amino acid sequence capable of adding an N-acetylglucosamine (GlcNAc) residue to $Man_5GlcNAc_2$ to produce a $GlcNAcMan_5GlcNAc_2$ and having more than 85% of identity with amino acid sequence SEQ ID N°1 or a fragment thereof, preferably more than 90% of identity with amino acid sequence SEQ ID N°1 or a fragment thereof, and more preferably more than 95% of identity with amino acid sequence SEQ ID N°1 or a fragment thereof.

As used herein, "percentage of identity" between two amino acids sequences, means the percentage of identical amino-acids, between the two sequences to be compared, obtained with the best alignment of said sequences, this percentage being purely statistical and the differences between these two sequences being randomly spread over the amino acids sequences. As used herein, "best alignment" or "optimal alignment", means the alignment for which the determined percentage of identity (see below) is the highest. Sequences comparison between two amino acids sequences are usually realized by comparing these sequences that have been previously align according to the best alignment; this comparison is realized on segments of comparison in order to identify and compared the local regions of similarity. The best sequences alignment to perform comparison can be realized by using computer softwares using such algorithms (GAP, BESTFIT, BLAST P, BLAST N, FASTA, TFASTA in the Wisconsin Genetics software Package). To get the best local alignment, one can preferably used BLAST software, with the BLOSUM 62 matrix, or the PAM 30 matrix. The identity

percentage between two sequences of amino acids is determined by comparing these two sequences optimally aligned, the amino acids sequences being able to comprise additions or deletions in respect to the reference sequence in order to get the optimal alignment between these two sequences. The percentage of identity is calculated by determining the number of identical position between these two sequences, and dividing this number by the total number of compared positions, and by multiplying the result obtained by 100 to get the percentage of identity between these two sequences.

Many different promoters allowing the expression of a nucleic acid sequence in *Phaeodactylum tricornutum* are known from the skilled person. As an example of such promoters, one can cite the nuclear promoters such fcpA and fcpB from *Phaeodactylum tricornutum* disclosed in ZAVLASKAÏA *et al.* (*J. Phycol.*, vol.36, p: 379-386, 2000). Nevertheless, this promoter has a sequence identity of less 50% with SEQ ID n°3 corresponding to the sequence of 1047 pb nucleic acid sequence upstream of the ATG of nucleic acid sequence SEQ ID N°2 present in the genome of wild-type *Phaeodactylum tricornutum*, preferably of less than 25% and most preferably of less than 10%.

In another preferred embodiment, said transformed *Phaeodactylum tricornutum* further comprises a nucleic acid sequence operatively linked to a promoter, wherein:

(i) said nucleic acid sequence encodes an alpha-Mannosidase II having the amino acid sequence SEQ ID n°5, a fragment or a derivative thereof; and

(ii) said operatively linked promoter has a sequence identity of less 50% with SEQ ID n° 7 or a fragment thereof.

In a second embodiment, said transformed *Phaeodactylum tricornutum* comprises a nucleic acid sequence operatively linked to a promoter, wherein:

(i) said nucleic acid sequence encodes an alpha-Mannosidase II having the amino acid sequence SEQ ID n°5, a fragment or a derivative thereof; and

(ii) said operatively linked promoter has a sequence identity of less 50% with SEQ ID n° 7 or a fragment thereof.

The expression "fragment" with reference to SEQ ID N°7 refers to a nucleic acid sequence of at least 100 nucleic acids of said SEQ ID N°7, preferably of at least 150 nucleic acids of SEQ ID N°7, most preferably of at least 200 nucleic acids of SEQ ID N°7.

α-Mannosidase II, also known as α-Man II, is an enzyme which catalyzes the first committed step in the biosynthesis of complex N-glycans. α-Man II is capable of hydrolysing the terminal (1->3)- and (1->6)-linked alpha-D-mannose residues in the mannosyl-oligosaccharide $GlcNAcMan_5GlcNAc_2$. $GlcNAcMan_5GlcNAc_2$, the product of GnT I, is successively converted in the Golgi apparatus into $GlcNAcMan_4GlcNAc_2$ and then $GlcNAcMan_3GlcNAc_2$ by action of the α-Man II.

The α-Mannosidase II having the amino acid sequence SEQ ID N°5 corresponds to the "endogenous α-Mannosidase II" encoded by the nucleic acid sequence SEQ ID N°6 present in the genome of wild-type *Phaeodactylum tricornutum*.

The expression "fragment" with reference to SEQ ID N°5 refers to an amino acid sequence of at least SEQ ID N°8 corresponding to the amino acid sequence of the luminal part of α-Man II, which comprises the catalytic site of α-Man II.

The expression "catalytic site" refers to the amino acid sequence of an enzyme, said amino acid sequence being responsible for the enzymatic activity. Therefore, the catalytic site of the α-Mannosidase II according to the invention corresponds to the amino acid sequence responsible for the conversion of $GlcNAcMan_5GlcNAc_2$, the product of GnT I, in the Golgi apparatus into $GlcNAcMan_4GlcNAc_2$ and then $GlcNAcMan_3GlcNAc_2$.

As used herein the term "α-Man II derivative" refers to an amino acid sequence capable of converting $GlcNAcMan_5GlcNAc_2$, the product of GnT I, in the Golgi apparatus into $GlcNAcMan_4GlcNAc_2$ and then $GlcNAcMan_3GlcNAc_2$ and having more than 85% of identity with amino acid sequence SEQ ID N°5 or a fragment thereof, preferably more than 90% of identity with amino acid sequence SEQ ID N°5 or a fragment thereof, and more preferably more than 95% of identity with amino acid sequence SEQ ID N°5 or a fragment thereof.

Many different promoters allowing the expression of a nucleic acid sequence in *Phaeodactylum tricornutum* are known from the skilled person. As an example of such promoters, one can cite the nuclear promoters such fcpA and fcpB from *Phaeodactylum tricornutum* disclosed in ZAVLASKAÏA *et al.* (*J. Phycol.*, vol.36, p: 379-386, 2000). Nevertheless, this promoter has a sequence identity of less 50% with SEQ ID n°7 corresponding to the sequence of 1000 pb nucleic acid sequence upstream of the ATG of nucleic acid sequence SEQ ID N°6 present in the genome of wild-type *Phaeodactylum tricornutum*, preferably of less than 25% and most preferably of less than 10%.

In a third embodiment, said transformed *Phaeodactylum tricornutum* has at least one β-N-acetylglucosaminidase which has been inactivated, preferably said at least one β-N-acetylglucosaminidase being a β-N-acetylglucosaminidase of amino acid sequence SEQ ID N°9 or a β-N-acetylglucosaminidase of amino acid sequence of SEQ ID N°11.

In a preferred embodiment, both β-N-acetylglucosaminidases of amino acid sequences SEQ ID N°9 and SEQ ID N°11 have been inactivated.

The term "inactivated" with reference to β-N-acetylglucosaminidases refers to β-N-acetylglucosaminidases which do not present any enzymatic activity.

This type of enzyme is able to hydrolyze exclusively the GlcNAc residue attached to the α-1,3/1,6-linked mannose of the core pentasaccharide of *N*-glycans. It is noteworthy that this enzyme could not hydrolyze $Man_5GlcNAc$, but acted only further "downstream" on $GlcNAcMan_5GlcNAc_2$, $GlcNAcMan_4GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$ or $GlcNAcMan_3GlcNAc_2$ substituted by fucose residues linked to the proximal GlcNAc.

By β-N-acetylglucosaminidase enzymatic activities, we mean the removal of the terminal GlcNAc residue linked to the mannoses of the core of complex-type glycans N-linked to proteins.

Such inactivation can be obtained by several ways known from the person skilled in the art. For example, methods of inactivation comprise gene silencing with RNA interference (miRNA, siRNA) which have been used in microalgae (Zhao T. *et al* 2009, The Plant Journal, vol. 58, p:157–164; Molnar A. *and al*., 2009, The Plant Journal vol. 58, p:165–174), as in *P. tricornutum* (De Riso V. *et al.* , 2009, Nucleic Acids Research, p:1–12). The inactivation of said β-N-acetylglucosaminidases can also be obtained by the knock out of the corresponding genes but also by the use of a method of inhibition of the enzymatic activity, for example by using antibody directed to the catalytic site of said β-N-acetylglucosaminidases.

The inactivation of β-N-acetylglucosaminidases can be confirmed by testing their enzymatic activity, which can be measured by techniques which are known from the skilled person in the art, for example by using one of the two different following tests of activity described in Léonard R. *et al.,* 2006, The Journal of Biological Chemistry, Vol. 281, p:4867-4875.

According to a first test, *N*-acetylglucosaminidases are incubated with the different substrates at 37 °C for 1-20 h. For experiments with *p*-nitrophenyl-GlcNAc, the substrate concentration is 5 mM in a total volume of 0.04 ml of 0.1 M citrate/phosphate buffer at pH 3-

8. The reactions are terminated by the addition of 0.26 ml of 0.4 M glycine/NaOH buffer at pH 10.4, and absorbance at 405 nm is measured with a microtiter plate reader.

According to a second test, Pyridylaminated oligosaccharides are used at a final concentration of 0.1 mM in a total volume of 0.02 ml of 0.1 M citrate/phosphate buffer at pH 3-8. Incubation is terminated by the addition of 0.18 ml of 20 mM ice-cold sodium borate. Aliquots of 0.05 ml are analyzed by reverse-phase HPLC as described previously.

As used herein, a "transformed" *Phaeodactylum tricornutum* may also correspond to a *P. tricornutum*, wherein at least one β-N-acetylglucosaminidase has been inactivated.

In a preferred embodiment, said β-N-acetylglucosaminidases have been inactivated with the technique of RNA interference.

In a preferred embodiment, said transformed *Phaeodactylum tricornutum* having at least one β-N-acetylglucosaminidase which has been inactivated further comprises a nucleic acid sequence operatively linked to a promoter, wherein

(i) said nucleic acid sequence encodes an N-acetylglucosaminyltransferase I having the amino acid sequence SEQ ID N°1, a fragment or a derivative thereof, and

(ii) said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°3 or a fragment thereof.

In another preferred embodiment, said transformed *Phaeodactylum tricornutum* having at least one β-N-acetylglucosaminidase which has been inactivated as disclosed previously further comprises a nucleic acid sequence operatively linked to a promoter, wherein

(iii) said nucleic acid sequence encodes an α-Mannosidase II having the amino acid sequence SEQ ID N°5, a fragment or a derivative thereof, and

(iv) said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°7 or a fragment thereof.

Preferably, all the transformed *P. tricornutum* as described previously further comprises a further nucleic acid sequence operatively linked to a promoter, said further nucleic acid sequence encoding a polypeptide that is expressed and glycosylated in the transformed *P. tricornutum*.

The term "polypeptide" as used herein refers to an amino acid sequence comprising more than 50 amino acids which are linked by peptide bonds.

After transformation of *P. tricornutum*, transformants producing the desired polypeptide are selected. Selection can be carried out by one or more conventional methods comprising: enzyme-linked immunosorbent *assay* (ELISA), mass spectroscopy such as MALDI-TOF-MS, ESI-MS chromatography, characterization of cells using fluorescence activated cell sorter, spectrophotometer, fluorimeter, immunocytochemistry by exposing cells to an antibody having a specific affinity for the desired protein. Such methods are detailed in examples below.

The glycosylated polypeptides have at least one $GlcNAcMan_5GlcNAc_2$ structure. Preferably, said glycosylated polypeptides have at least one $GlcNAcMan_4GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$ or $GlcNAc_2Man_3GlcNAc_2$.

Advantageously, the polypeptide expressed and glycosylated by the transformed *Phaeodactylum tricornutum* is a polypeptide of animal origin, preferably of mammalian origin, and most preferably of human origin.

In a still preferred embodiment, the polypeptide expressed and glycosylated by the transformed *P. tricornutum* of the invention is a polypeptide having a therapeutic interest. Preferably, said polypeptide is selected in the group comprising erythropoietin, cytokines such as interferons, antibodies and their fragments, coagulation factors, hormones, beta-glucocerebrosidase, pentraxin-3, anti-TNFs, acid α-glucosidase, α-L-iduronidase and derivatives thereof.

An antibody is an immunoglobulin molecule corresponding to a tetramer comprising four polypeptide chains, two identical heavy (H) chains (about 50-70 kDa when full length) and two identical light (L) chains (about 25 kDa when full length) inter-connected by disulfide bonds. Light chains are classified as kappa and lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, and define the antibody's isotype as IgG, IgM, IgA, IgD, and IgE, respectively. Each heavy chain is comprised of a N-term heavy chain variable region (abbreviated herein as HCVR) and a heavy chain constant region. The heavy chain constant region is comprised of three domains (CH1, CH2, and CH3) for IgG, IgD, and IgA; and 4 domains (CH1, CH2, CH3, and CH4) for IgM and IgE. Each light chain is comprised of a N-term light chain variable region (abbreviated herein as LCVR) and a light chain constant region. The light chain constant region is comprised of one domain, CL. The HCVR and

LCVR regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDRs), interspersed with regions that are more conserved, termed framework regions (FR). Each HCVR and LCVR is composed of three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4. The assignment of amino acids to each domain is in accordance with well-known conventions. The functional ability of the antibody to bind a particular antigen depends on the variable regions of each light/heavy chain pair, and is largely determined by the CDRs.

The term "antibody", as used herein, refers to a monoclonal antibody *per se*. A monoclonal antibody can be a human antibody, chimeric antibody and/or humanized antibody.

The term “antibody fragments” as used herein refers to antibody fragments that bind to the particular antigens of said antibody. For example, antibody fragments capable of binding to particular antigens include Fab (e.g., by papain digestion), Fab' (e.g., by pepsin digestion and partial reduction) and F(ab')2 (e.g., by pepsin digestion), facb (e.g., by plasmin digestion), pFc’ (e.g., by pepsin or plasmin digestion), Fd (e.g., by pepsin digestion, partial reduction and reaggregation), Fv or ScFv (e.g., by molecular biology techniques) fragments, are encompassed by the invention.

Such fragments can be produced by enzymatic cleavage, synthetic or recombinant techniques, as known in the art and/or as described herein. Antibodies can also be produced in a variety of truncated forms using antibody genes in which one or more stop codons have been introduced upstream of the natural stop site. For example, a combination gene encoding a $F(ab')_2$ heavy chain portion can be designed to include DNA sequences encoding the CH_1 domain and/or hinge region of the heavy chain. The various portions of antibodies can be joined together chemically by conventional techniques, or can be prepared as a contiguous protein using genetic engineering techniques.

The term “Cytokines” refers to signalling proteins which are released by specific cells of the immune system to carry a signal to other cells in order to alter their function. Cytokines are immunomodulating agents and are extensively used in cellular communication. The term cytokines encompasses a wide range of polypeptide regulators, such as interferons, interleukins, chemokins or Tumor Necrosis Factor.

The term "Coagulation factors" refers to the plasma proteins which interact with platelets in a complex cascade of enzyme-catalyzed reactions, leading to the formation of fibrin for the initiation of a blood clot in the blood coagulation process. Coagulation factors, at the number of 13, are generally serine proteases, but also comprise glycoproteins (Factors VIII and V) or others types of enzyme, such as transglutaminase (Factor XIII).

The term "Hormones" refers to chemical messengers secreted by specific cells in the plasma or the lymph to produce their effects on other cells of the organism at a distance from their production sites. Most hormones initiate a cellular response by initially combining with either a specific intracellular or cell membrane associated receptor protein. Common known hormones are, for example, insulin for the regulation of energy and glucose in the organism, or the Growth Hormone which stimulates growth and cell reproduction and regeneration.

In another preferred embodiment, the invention relates to a transformed *P. tricornutum* as described above, further comprising another nucleic acid sequence operatively linked to a promoter, wherein said nucleic acid sequence encodes an N-acetylglucosaminyltransferase II, a fragment or a derivative thereof.

N-acetylglucosaminyltransferase II, also known as GnT II or mannoside acetylglucosaminyltransferase II (MGAT II) is an enzyme from the N-glycosylation pathway, which is capable of adding an N-acetylglucosamine (GlcNAc) residue to $GlcNAcMan_3GlcNAc_2$, product of α-Man II, to produce a $GlcNAc_2Man_3GlcNAc_2$

Examples of GnT II, also known as mannosyl (alpha- 1 ,6-)-glycoprotein beta- 1,2 -N-acetylglucosaminyltransferase (MGAT 2), include GnT II from Mus musculus (SEQ ID N°13, Accession number NP_666147), from Homo sapiens (SEQ ID N°14, Accession number NP_002399) or from *Phaeodactylum tricornutum* (SEQ ID N°5).

Preferably, N-acetylglucosaminyltransferase II has an amino acid sequence of SEQ ID N°5, a fragment or a derivative thereof and said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°7 or a fragment thereof.

In still another preferred embodiment, said N-acetylglucosaminyltransferase II comprises the amino acid sequence SEQ ID N°15 and said operatively linked promoter has a sequence identity of less 50% with SEQ ID n° 7 or a fragment thereof.

Inventors have noticed that α-Mannosidase II have an N-acetylglucosaminyltransferase II domain of amino acid sequence SEQ ID N°15.

The expression "fragment" with reference to SEQ ID N°5 and GnT II refers to an amino acid sequence of at least SEQ ID N°15 corresponding to the amino acid sequence of the GnT II domain of α-Mannosidase II.

The expression "catalytic site" refers to the amino acid sequence of an enzyme, said amino acid sequence being responsible for the enzymatic activity. Therefore, the catalytic site of the GnT II according to the invention corresponds to the amino acid sequence responsible for the addition of an N-acetylglucosamine (GlcNAc) residue to $GlcNAcMan_3GlcNAc_2$ to produce a $GlcNAc_2Man_3GlcNAc_2$.

As used herein the term "GnT II derivative" refers to an amino acid sequence capable of adding an N-acetylglucosamine (GlcNAc) residue to $GlcNAcMan_3GlcNAc_2$, product of α-Man II, to produce a $GlcNAc_2Man_3GlcNAc_2$ and having more than 85% of identity with the amino acid sequence SEQ ID N°5 or a fragment thereof, preferably more than 90% of identity with amino acid sequence SEQ ID N°5 or a fragment thereof, and more preferably more than 95% of identity with amino acid sequence SEQ ID N°5 or a fragment thereof.

In another preferred embodiment, the invention relates to a transformed *P. tricornutum* as described above, further comprising another nucleic acid sequence operatively linked to a promoter, said nucleic acid sequence encoding at least an enzyme of the human N-glycosylation pathway such as N-acetylglucosaminyltransferases III, IV, V, VI, and glycosyltransferases such as galactosyltransferases, fucosyltransferases or sialyltransferases. Said enzymes are expressed in said transformed *P. tricornutum* and enable the N-glycosylation of a polypeptide.

In another preferred embodiment, said transformed *P. tricornutum* as described above, further comprising another nucleic acid sequence operatively linked to a promoter comprises a nucleic acid sequence encoding N-acetylglucosaminyltransferases III, IV, V and VI.

In still another preferred embodiment said transformed *P. tricornutum* as described above, further comprising another nucleic acid sequence operatively linked to a promoter comprises a nucleic acid sequence encoding glycosyltransferases comprising galactosyltransferases, fucosyltransferases and sialyltransferases.

GnT III, GnT IV, GnT V, GnT VI, fucosyltransferase, galactosyltransferase (GaIT) and sialyltransferases (ST) are well known from one of skilled in the art.

Examples of GnT III, also known as mannosyl (beta- 1 ,4-)-glycoprotein beta-l,4-N-acetylglucosaminyltransferase (MGAT 3), include GnT III from Mus musculus (SEQ ID

N°17, Accession number NP_034925) or from Homo sapiens (SEQ ID N°18, Accession number NP_002400). Preferably, said N-acetylglucosaminyltransferase III (GnT III) corresponds to SEQ ID N° 18 (Accession number NP_02400).

Examples of GnT IV, also known as mannosyl (alpha- 1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (MGAT4), include GnT IV isozyme A from Mus musculus (SEQ ID N°19, Accession number NP_776295), isozyme B from Mus musculus (SEQ ID N°20, Accession number NP_666038), isozyme C from Mus musculus (SEQ ID N°21, Accession number NP_080519), GnT IV isozyme A from Homo sapiens (SEQ ID N°22, Accession number NP_036346), GnT IV isozyme B from Homo sapiens (isoform 1, SEQ ID 23, Accession number NP_055090 or isoform 2, SEQ ID N°24, Accession number NP_463459) or GnT IV isozyme C from Homo sapiens (SEQ ID N°25, Accession number NP_037376).

Examples of GnT V, include GnT V from Mus musculus (SEQ ID N°26, Accession number NP_660110), GnT V isozyme B from Mus musculus (SEQ ID N°27, Accession number NP_766536), GnT V from Homo sapiens (SEQ ID N°28, Accession number NP 002401), GnT V isozyme B from Homo sapiens (isoform 1, SEQ ID N°29, Accession number NP 653278 or isoform 2, SEQ ID N°30, Accession number NP_945193).

Example of GnT VI includes GnT VI from Gallus gallus (SEQ ID N°31, Accession number NP_990012).

Fucosyltransferases are well known from the skilled person and include, as an example alpha (1,6) fucosyltransferase (fucosyltransferase 8 (FUT8)), like FUT8 from Mus musculus (SEQ ID N°32, Accession number NP_058589) or FUT8 from Homo sapiens (SEQ ID N°33, Accession number Q9BYC5). Preferably, said fucosyltransferase corresponds to SEQ ID N°33 (Accession number Q9BYC5).

Galactosyltransferase are well known from the skilled person and include, as an example, one beta- (1,4) - galactosyltransferase (B4GALT1), like B4GALT1 from Homo sapiens (SEQ ID N°34, Accession number NP_001488), or B4GALT1 from Mus musculus (SEQ ID N°35, Accession number CAM14782). Preferably, said galactosyltransferase corresponds to SEQ ID N° 34 (Accession number NP_001488).

Sialyltransferase are well known from the skilled person and include, as an example Alpha 2, 6 Sialyltransferase (ST6 beta-galactosamide alpha-2,6-sialyltranferase 1 (ST6GAL1) or beta galactoside alpha 2,6 sialyltransferase 2 (ST6GAL2)), like ST6GAL2 from Mus

musculus (SEQ ID N°36, Accession number NP_766417) or ST6GAL1 from Homo sapiens (isoform a, SEQ ID 37, Accession number NP_775323 or isoform b, SEQ ID N°38, Accession number NP_775324), or Alpha 2, 3 Sialyltransferase (ST3 beta- galactoside alpha-2,3-sialyltransferase 6 (ST3GAL6), ST3 beta-galactoside alpha-2,3- sialyltransferase 1 (ST3GAL1), ST3 beta-galactoside alpha-2,3 -sialyltransferase 2 (ST3GAL2), ST3 beta-galactoside alpha-2,3 - sialyltransferase 3 (ST3GAL3), like ST3GAL1 from Mus musculus (SEQ ID N°39, Accession number NP_033203) or from Homo sapiens (SEQ ID N°40, Accession number NP_003024), ST3GAL2 from Homo sapiens (SEQ ID N°41Accession number NP_008858), ST3GAL3 from Homo sapiens (isoform a, SEQ ID N°42, Accession number NP_777623, isoform b, SEQ ID N°43, Accession number NP_777624, isoform c, SEQ ID N°44, Accession number NP_777625, isoform f, SEQ ID N°45, Accession number NP_777628, isoform j, SEQ ID N°46, Accession number NP_006270, isoform d, SEQ ID N°47, Accession number NP_777626, isoform e, SEQ ID N°48, Accession number NP 777627, isoform i, SEQ ID N°49, Accession number NP_777631, isoform g, SEQ ID N°50, Accession number NP_777629, isoform h, SEQ ID N°51, Accession number NP_777630), or ST3GAL6 from Homo sapiens (SEQ ID N°52, Accession number NP_006091).

For a glycosyltransferase to function satisfactorily in the Golgi apparatus, it is necessary for the enzyme to be provided with sufficient concentrations of an appropriate nucleotide sugar, which is the high-energy donor of the sugar moiety added to a nascent glycoprotein. In humans, the full range of nucleotide sugar precursors are generally synthesized in the cytosol and transported into the Golgi apparatus, where they are attached to the core oligosaccharide by glycosyltransferases. The Applicant observed in microalgae a sufficient concentration of GlcNAc, mannose, fucose and galactose but not of sialic acid.

Therefore, for a sialyltransferase to function satisfactorily in the Golgi apparatus, it is necessary to express in the microalgae one or more enzymes needed for sialic acid synthesis, its activation and its transport within the Golgi apparatus among UDP-GlcNAc 2-epimerase, GlcNAc 2-epimerase, GlcNAc-6P 2-epimerase, NeuAc synthase. NeuAc-9P synthase, CMP-NeuAc synthase and CMP-sialic acid transporter (see for example works done in plants: Misaki R et al. Biochem Biophys Res Commun. 2006 Jan 27; 339 (4): 1184- 9; Paccalet T et al. Plant Biotechnol J. 2007 Jan;5(l): 16-25), Castihlo et al., 2008 and Castilho et al., 2010 as described above).

UDP-GlcNAc 2-epimerase, which is also known as glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase (GNE), is well known from the skilled person and

include, as an example GNE from Mus musculus (SEQ ID N°53, Accession number NP_056643) or GNE from Homo sapiens (SEQ ID N°54, Accession number NP_005467). Preferably, said GNE corresponds to SEQ ID N° 54 (Accession number NP_005467).

GlcNAc 2-epimerase is well known from the skilled person and includes, as an example, the renin binding protein (RENBP) from Homo sapiens (SEQ ID N°55, Accession number NP_002901).

NeuAc-9-P synthase, also called N-acetylneuraminic acid synthase (NANS), is well known from the skilled person and include, as an example, NANS from Homo sapiens (SEQ ID N°56, Accession number NP_061819).

CMP-NeuAc synthase, which is also known as cytidine monophospho-N-acetylneuraminic acid synthetase (CMAS), is well known from the skilled person and include, as an example CMAS from Mus musculus (SEQ ID N°57, Accession number NP_034038) or from Homo sapiens (SEQ ID N°58, Accession number NP_061156). Preferably, said CMAS corresponds to SEQ ID N° 58 (Accession number NP_061156).

CMP-sialic acid transporters are also well known from the skilled person and include, as an example, solute carrier family 35 (CMP-sialic acid transporter), member Al (SLC35A1) from Mus musculus (SEQ ID N°59, Accession number NP_036025) or from Homo sapiens (SEQ ID N°60, Accession number NP_006407). Preferably, said CMP- sialic acid transporter corresponds to SLC35A1 from Homo sapiens (SEQ ID N°60, Accession number NP_006407).

The added transporter protein conveys a nucleotide sugar from the cytosol into the Golgi apparatus, where the nucleotide sugar may be reacted by the glycosyltransferase, e.g. to elongate an N-glycan. The reaction liberates a nucleoside diphosphate or monophosphate, e.g. UDP, GDP, or CMP. As accumulation of a nucleoside diphosphate inhibits the further activity of a glycosyltransferase, it is frequently also desirable to provide an expressed copy of a gene encoding a nucleotide diphosphatase. The diphosphatase (specific for UDP or GDP as appropriate) hydrolyzes the diphosphonucleoside to yield a nucleoside monosphosphate and inorganic phosphate. The nucleoside monophosphate does not inhibit the glycosyltransferase and in any case is exported from the Golgi by an endogenous cellular system.

Another object of the invention is a method for producing a glycosylated polypeptide, said method comprising the steps of:

(i) Culturing a transformed *P. tricornutum* as described here above;

(ii) Purifying said polypeptide expressed and glycosylated in said transformed *P. tricornutum*.

In a preferred embodiment, said method for producing a glycosylated polypeptide comprises a former step of transforming a *Phaeodactylum tricornutum* so as to obtain a *P. tricornutum* as defined previously.

Methods which can be employed for the transformation of *P. tricornutum* are described here above. Such transformation, culture of *P. tricornutum* and purification of glycosylated polypeptides are also exemplified below.

Advantageously, the method of the invention further comprises a step (iii) of determining the glycosylation pattern of said polypeptide.

This glycosylation pattern can be determined by method well known from the skilled person. As an example, preliminary informations about N-glycosylation of the recombinant glycoprotein can be obtained by affino- and immunoblotting analysis using specific probes such as lectins (CON A; ECA; SNA; MAA...) and specific N-glycans antibodies (anti- βl,2-xylose; anti-α-1,3-fucose; anti-Neu5Gc, anti-Lewis ...). This is made according to FITCHETTE *et al.*, (Methods Mol. Biol., vol.355, p: 317-342, 2007) and could be completed by deglycosylation assays.

To investigate the detailed N-glycan profile of recombinant protein, N-linked oligosaccharides is then released from the protein in a non specific manner using enzymatic digestion or chemical treatment (FITCHETTE *et al.*, above mentioned, 2007; SEVENO *et al., Anal. Biochem.*, vol.379(1), p:66-72, 2008). The resulting mixture of reducing oligosaccharides can be profiled by HPLC and/or mass spectrometry approaches (ESI-MS-MS and MALDI-TOF essentially) (BARDOR *et al., Curr Opin Struct Biol.*, vol.16 (5), p:576-583, 2006; SEVENO *et al.,* above mentioned, 2008). These strategies, coupled to exoglycosidase digestion, enable N-glycan identification and quantification (SEVENO *et al.,* above mentioned, 2008).

Another alternative to study N-glycosylation profile of recombinant protein is to work directly on its glycopeptides after protease digestion of the protein, purification and mass spectrometry analysis of the glycopeptides as disclosed in BARDOR *et al.* (*Plant Biotechnol. J.*, vol.1 (6), p: 451-462, 2003).

Another object of the invention relates to the use of a transformed *P. tricornutum* as defined previously for producing a glycosylated polypeptide.

In the following, the invention is described in more detail with reference to methods. Yet, no limitation of the invention is intended by the details of the examples. Rather, the invention pertains to any embodiment which comprises details which are not explicitly mentioned in the examples herein, but which the skilled person finds without undue effort.

EXAMPLES

Example 1 No significant GnT-I activity was detected in *Phaeodactylum tricornutum*

A gene encoding a putative GnT-I was predicted in the *P. tricornutum* genome (Pt54844; http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) (SEQ ID N°2). In eukaryotes, this enzyme is involved in the N-glycan maturation into complex-type N-glycan by transfer of a terminal GlcNAc onto Man-5 (Fig. 1). This putative GnT I sequence SEQ ID N°2 is predicted to be a type II protein with a luminal part sharing high structural homologies with the catalytic domain of the rabbit GnT I.

In order to determinate if this gene encodes for an active GnT I, the glycosylation pattern of *P. tricornutum* proteins was analyzed.

Phaeodactylum tricornutum was cultivated using a standard batch culture method using a scale-up from 2 to 10 L glass carboys in sterilized Conway media (WALNE, L. Fish Invest Serie II, vol.25(4), p:1–53, 1966) with seawater (salinity = 3.3–3.4 %), 1 µm-filtered and aerated with a 2% CO_2/air mixture to maintain the pH in a range of 7.5–8.1. Sodium metasilicate was added to the media to the 40 mg/L final concentration.

Phaeodactylum tricornutum were grown at 22–23°C under continuous illumination (280–350 µmol photons m^2s^{-1}). The concentrated culture (about 20.10^6 cells / mL) is first centrifuged at 5,000 g for 20 min at 4°C and the pellet was then lyophilised.

Two grams of lyophilised microalgae were grind in presence of sand in a mortar using a 750 mM Tris-HCl pH 8 buffer containing 15% (w/v) of sucrose, 2% (v/v) of β-mercaptoethanol and 1 mM phenylmethylsulfonylfluoride and then centrifuged at 4°C for 30 min at 11,500 g. Proteins from the supernatant were then precipitated with 90% ammonium sulfate during 2 hours at room temperature. The pellet was solubilized in water and then, dialysed against water overnight at 4°C. Finally, the total protein extract was ultra-centrifuged at 100 000 g for 1 hour at 4°C and resuspended in the smallest volume of water, prior to protein quantification and further analyses. Protein quantification was performed on the total protein extracts from *Phaeodactylum tricornutum* using the BCA protein assay kit from PIERCE according to the manufacturer's instructions.

Structural analysis of glycans N-linked to *P. tricornutum* proteins was then investigated by western-blot analysis on a total protein extract using probes specific for glycan epitopes. For this analysis, 50 µg of total proteins were separated by SDS-PAGE. Onion proteins were used as a control. The separated proteins were transferred onto nitrocellulose membrane and stained with Ponceau Red in order to control transfer efficiency. Affinodetection using concanavalin A was performed by incubation with the lectin at 25 $\mu g.mL^{-1}$ during 2 h at RT in TBS-T, complemented with 1 mM $CaCl_2$ and 1 mM $MgCl_2$. After washing with TBS-T complemented with $CaCl_2$ and $MgCl_2$ (6 times, 5 minutes), binding of this lectin was detected using horseradish peroxidase diluted at 50 $\mu g.mL^{-1}$, 1 h at RT in TBS-T complemented with 1 mM $CaCl_2$ and 1 mM $MgCl_2$. After washing with the same TBS-T and then TBS, final development of the blots was performed by using 4-chloro-1-naphtol as previously described (FITCHETTE *et al.*, *Methods in Molecular Biology published by Humana Press, USA (Totowa, New Jersey)*, p:317-342, 2006).

Immunodetection using home-made specific core-β(1, 2)-xylose and core-α(1, 3)-fucose antibodies (1:1,000 in TBS containing 1% of gelatin, 2h, RT) was also performed. After washing with TBS-T (6 times, 5 minutes), binding of antibodies was detected using a secondary horseradish peroxidase-conjugated goat anti-rabbit IgG antibody diluted at 1:3,000 in TBS containing 1% gelatin for 90 min at RT (Bio-Rad). Final development of the blots was performed by using 4-chloro 1-naphtol as previously described (FITCHETTE *et al.*, above mentioned, 2006).

The results are presented in Figure 2 showing affinodetection using concanavalin A (Con A) and immunodetection using antibodies raised against the core β(1,2)-xylose (anti-Xyl) and core α(1,3)-fucose (anti-Fuc) epitopes of proteins isolated from green onion as a positive control (Lane 1) and from *Phaeodactylum tricornutum* grown in standard conditions (Lane 2)

The results show that the proteins expressed in *Phaeodactylum tricornutum* do not exhibit any β(1,2)-xylose (anti-Xyl) and core α(1,3)-fucose (anti-Fuc) epitopes. Nevertheless, *Phaeodactylum tricornutum* proteins exhibit high-mannose sequences as revealed by Con A binding, a lectin specific for high-mannose sequences. The presence of high-mannose sequences in *Phaeodactylum tricornutum* proteins was also confirmed by deglycosylation assays using peptide N-glycosidase F (PNGase F) and Endoglycosidase H (Endo H) (data not shown).

In order to determine the specific glycosylation pattern of *Phaeodactylum tricornutum* proteins, said proteins were digested prior to mass spectrometry analysis.

Total proteins were digested by successive treatments with pepsin and PNGase A as previously described in FITCHETTE *et al.* (above mentioned, 2006). 4 mg of proteins were digested with 6 mg of pepsin in 2 mL of 10 mM HCl, pH 2.2, at 37°C for 48 h. After neutralization with 1 M ammonium hydroxide, the solution was heated for 5 min at 100°C and lyophilized. Glycopeptides were then deglycosylated overnight at 37°C with PNGase A (10 mU, BOEHRINGER MANNHEIM) in a 100 mM sodium acetate buffer, pH 5.0. N-Glycans were purified by successive elution through an AG 50W-X2 column (BIO-RAD) and a C18 cartridge (VARIAN) according to FITCHETTE *et al.* (above mentioned, 2006). The purified N-glycans were then labelled by 2-aminobenzamide (2-AB) using the optimized protocol described in BIGGE *et al.* (*Anal Biochem,* vol.230, p: 229-238, 1995). Then, labelled N-glycans were detected with an UV light and eluted using water. The eluted labelled N-glycans were finally lyophilised prior to exoglycosidase digestion and MALDI-TOF mass spectrometry analysis. For exoglycosidase digestion, 200 milliunits of Jack bean α-mannosidase (SIGMA-ALDRICH) were desalted by ultrafiltration using a Centricon and incubated overnight at 37°C with approximately 50 pmoles of 2-AB labelled N-glycan mixture. Then, the 2-AB labelled N-glycans was directly analysed by matrix assisted laser desorption ionisation-time of flight (MALDI-TOF) mass spectrometry on a Voyager DE-Pro MALDI-TOF instrument (APPLIED BIOSYSTEMS) equipped with a 337 nm nitrogen laser. Mass spectra were performed in the reflector delayed extraction mode using 2,5-dihydroxybenzoic acid (SIGMA-ALDRICH) as matrix. The matrix, freshly dissolved at 5 mg/mL in a 70:30 acetonitrile/ 0.1% TFA, was mixed with the water solubilized oligosaccharides in a ratio 1:1 (v/v). These spectra were recorded in a positive mode, using an acceleration voltage of 20,000 V with a delay time of 100 ns. They were smoothed once and externally calibrated using commercially available mixtures of peptides and proteins (APPLIED BIOSYSTEMS). In this study, the spectra have been externally calibrated using des-Arg^1-bradykinin (904.4681 Da), angiotensin I (1296.6853), Glu^1-fibrinopeptide B (1570.6774 Da), $ACTH_{18-39}$ (2465.1989 Da) and bovine insulin (5730.6087 Da). 1000 Laser shots were accumulated for each spectrum and several spectra were accumulated (between 5 to 10 spectra) in order to obtain a good signal to noise ratio.

The results in figure 13 have shown that the major ions (95-97%) for the resulting pool of labeled N-glycans correspond to $(M+Na)^+$ ions of 2-AB derivatives of $Hexose_{5-9}GlcNAc_2$.

Nevertheless, no derivatives GlcNAc-Hexose$_5$GlcNAc$_2$ was detected.

Consequently, we show that *P. tricornutum* does not exhibit a significant GnT I activity.

Example 2 A putative GnT I detected in *P. tricornutum* GnT I complement CHO$^{\Delta GnT\text{-}I}$ cells

Even if no detectable GnT I activity was identify in *Phaeodactylum tricornutum*, we try to express the nucleic acid sequence corresponding to the *P. tricornutum* putative GnT I full-length sequence (Pt54844) (SEQ ID N°2) in CHO *Lec1* mutant, which is mutated on its endogenous GnT I. Said *P. tricornutum* sequence was expressed in the CHO mutant in fusion with a V5 epitope to monitor the expression of the fusion protein in transformants.

On the basis of the detection of this epitope with anti-V5 antibodies, we show that most of the CHO transformants were found to express the V5 fusion protein (data not shown). Two cell lines expressing the V5 fusion protein were selected for N-linked glycan analysis.

Proteins from these lines as well as from wild-type and *Lec1* CHO cells were isolated and their N-linked glycans were released by treatment with PNGase F followed by MALDI-TOF mass spectrometry analysis.

The results shown that the CHO *Lec1* mutant accumulates high-mannose-type N-glycans, in contrast to wild-type CHO cells which exhibited both high-mannose-type and complex-type N-glycans. Surprisingly, the proteins from the CHO *Lec1* mutant expressing V5 fusion protein carry both high-mannose-type and a complete set of complex N-glycans identical to the one observed in wild-type CHO cells (figure 14).

Finally, these results show that even if no GnT I activity was significantly detectable in *P. tricornutum*, the expression of the Pt54844 gene was able to restore the biosynthesis of complex N-glycans in mammalian cells.

In conclusion, Pt54844 gene encodes for a functional transferase designed as Pt GnT I.

Example 3 Expression of GnT I in *Phaeodactylum tricornutum* and validation by RT-PCR

The Pt54844 gene disclosed previously (SEQ ID N°2) was cloned under of an enhanced promoter of Cauliflower Mosaic Virus in a pPha-T1 based vector called BSJ-25 vector. As a control, vectors comprising GnT I from *Arabidopsis thaliana* (AtGnT I, At4g38240) and GnT I from human cell (hGnT I, MGAT1) were also constructed. BSJ-25

(SEQ ID N°79) was derived from pPha-T1 (Zaslavskaia and Lippmeier, 2000, J. Phycol. vol.36(2), p: 379-386.) vector by replacing the *FcpA* promoter of the expression cassette by the double enhanced Cauliflower Mosaic Virus (CaMV35S) (SEQ ID N°80) fused to the plant signal peptide. For this, pPha-T1 containing bleomycin-resistance cassette driven by *FcpB* promoter was digested with Nde1 and EcoR1 to remove the FcpA promoter. This construct was designated as "pPha-T1-PfcpA deleted". The double enhanced Cauliflower Mosaic Virus 35S promoter of the expression cassette was amplified by PCR with forward primer, CaMV35Sfwd (5'-GAACATATGGTGGATTGATGTGATCTACTCC-3') (SEQ ID N°61) and reverse primer, CaMV35Srev (5'-AATTCTCGAGGAATTCGGCCGAGG -3') (SEQ ID N°62) on the PS1 construct (Kotzer *et al.,* 2004, J Cell Sci, vol.117(Pt 26), p: 6377-89). PS1 construct contains a double Cauliflower Mosaic Virus 35S promoter (SEQ ID N°80), a tobacco Mosaic Virus - Ω sequence as translation enhancer fused to the tobacco chitinase signal peptide of SEQ ID N°81 (Haseloff *et al.,* 1997, Proc Natl Acad Sci U S A, vol.94(6), p: 2122-7.; Batoko *et al.,* 2000, Plant Cell, vol.12(11), p: 2201-18). The PCR product was digested by Nde1 and EcoR1 and then was cloned into "pPha-T1-PfcpA deleted" to generate BSJ-25 vector.

To resume, the expression cassette of BSJ-25 vector contain a double CaMV promoter, a translational enhancer, a signal peptide, a multi-cloning site and the *FcpA* terminator (Fig 3). The GnT I gene from various origin has been introduced in this vector using enzymatic digestion from the multi-cloning site and the inserts had been cloned between the signal peptide and the terminator sequence.

Specific vectors expressing the abovementioned genes in fusion to Green Fluorescent Protein (GFP) were also constructed to investigate the cellular localisation in the Golgi apparatus of the corresponding fusion proteins.

The constructs can be seen in Figure 12.

Said vectors were used to transform *Phaeodactylum tricornutum*.

Transformation of *Phaeodactylum tricornutum* was carried out as described by Zaslavskaia *et al.* (2000, J. Phycol. Vol.36(2), p: 379-386). *P. tricornutum* 1.8.6 are cultivated (flask or agar plate 10%) in sterile sea water (0.22µm filtered) enriched with 0.1% (v/v) of the nutritive medium (Conway) and 0.1% (v/v) of a silica solution (0.4 g/ml). All the cultures are maintained at 20°C under continuous lighting in sterile conditions. For genetic transformation, cultures of microalgae in exponential phase of growth are counted and

concentrated by centrifugation, diluted in sterile sea water and approximately 10^8 cells were inoculated as a plaque of 2.5 cm of diameter on the surface of the medium agar plate. Five hundred micrograms or gold microcarrier (0.6 µm of particle size) was coated with 1µg of vector DNA in the presence of 1.25M of $CaCl_2$ and 20µM of spermidin. The transformation was carried out with the BIORAD PDS-1000/He biolistic particle delivery system for particles bombardment. Experiments are performed under a hood. The bombardment was performed at 900 and 1100 psi under a negative pressure of 27 Hg with different target distance (6 – 8 cm). After bombardment, cells transformed were suspended in 600 µL of nutritive sea water and were cultivated 1 day under illumination at 20°C. Transformed cells were spread on agar plate medium containing 100 $\mu g.mL^{-1}$ of Zeocin and incubated to grown 3 or 4 weeks under continuous light.

The presence of transcripts for the recombinant GnT I was then monitored by reverse transcription PCR (RT-PCR). Microalgae pellet were resuspended in 1.2 mL of Trizol (INVITROGEN) and homogenized in 2 mL Lysing Matrix E by 10 Fastprep-24 run (6.5m / s, 60 sec, MP BIOMEDICALS). 300 µL of chloroform were added to the supernatant, vigorously homogenized and incubated 3 min at room temperature. After a 15 min centrifugation at 4°C, the aqueous was mixed with 1 volume of absolute ethanol and the RNA were purified by RNEASY MINI kit (QIAGEN). The purified RNA were eluted in 50 µL of RNase-free water, dosed with NanoDrop (THERMO SCIENTIFIC) and were digested by RQ1 DNase (PROMEGA). After second purification by RNEASY MINI kit and Nanodrop quantification, RNA were conserved at -80°C.

Reverse Transcription (RT) was performed on 1 µg of purified RNA. The first cDNA strand was synthesised by 200 units of M-MLV RT RNase H minus (PROMEGA) with oligodT primers. Two µL of cDNA were used for PCR with GoTaq polymerase (PROMEGA) in 50 µL. The annealing was performed with specific primers of GnT I and actin sequence.

The results confirmed the expression of transcripts for the recombinant GnT in the selected recombinants (data not shown).

The glycosylation pattern of proteins expressed in said transformants is then performed as disclosed previously. Simultaneously, the cellular localisation of the fusion proteins is investigated by observing the GFP fluorescence of GnT I-GFP fusion proteins by confocal microscopy.

Another set of tests were run in order to demonstrate the expression and localization of endogenous GnT I in transformed *Phaeodactylum tricornutum*.

Example 4 Expression of endogenous GnT I and validation by RT-PCR and Q-PCR.

1) Constructions comprising the sequence of GnT I

Different constructs comprising the sequence coding for the endogenous GnT I from *Phaeodactylum tricornutum* were realised with the cloning vector pPHA-T1 built by Zavlaskaïa et al. (2000) for the genetic transformation of *Phaeodactylum tricornutum*, said vector including sequences of *P. tricornutum* promoters fcpA and fcpB (fucoxanthin-chlorophyll a/c-binding proteins A and B) and the terminator of fcpA. It contains a selection cassette with the gene she ble and a MCS flanking the fcpA promoter.

Said constructs are schematised in figure 4.

The first cassette comprised the sequence coding for the endogenous GnT I placed under the control of endogenous regulatory sequences. In the second construction, the GnT I was fused to the Green Fluorescent Protein (GFP).

The endogenous GnT I sequence of *Phaeodactylum tricornutum* was cloned in an expression vector with regulatory sequences from said microalgae. The GnT I was cloned alone or fused to the Green Fluorescent Protein (eGFP). The expression of said fusion protein enabled to visualise the expression and localization of the GnT I in the microalgae.

The vectors used for the transformation of *Phaeodactylum tricornutum* also comprised a selection cassette comprising a zeocin resistance gene. They enabled the genetic transformation of the Pt186 *Phaeodactylum tricornutum* strain.

The obtained clones were isolated and cultured in order to be analysed.

The transformation of *Phaeodactylum tricornutum* was realized as described previously.

2) Screening of transformed microalgae

The clones obtained after the transformation of the Pt186 *Phaeodactylum tricornutum* strain were isolated on fresh culture medium. The insertion of the construction comprising a gene coding for GnT I in the genome of the microalga was verified by PCR amplification with a set of specific primers for the transgene (SEQ ID N°69 for the vector and SEQ ID N°70 for the GnT I).

The PCR reaction was carried out in a final volume of 50 µl consisting of 1X PCR buffer, 0.2 mM of each dNTP, 5 µM of each primer, 20 ng of template DNA and 1.25 U of *Taq* DNA polymerase (Taq DNA polymerase, ROCHE). Thirty cycles were performed for the amplification of template DNA. Initial denaturation was performed at 94°C for 3 min. Each subsequent cycle consisted of a 94°C (1 min) melting step, a 55°C (1 min) annealing step, and a 72°C (1 min) extension step. Samples obtained after the PCR reaction were run on agarose gel (1%) stained with ethidium bromide.

The results are shown on figure 5 and present the screening obtained for 5 different lineages of potentially transformed *Phaeodactylum tricornutum*. The amplification of a specific sequence of 120 bp from the transgene obtained for the positive control (plasmidic DNA) was also found in the 5 analysed clones. The clones of the microalga therefore did integrate the genetic construction comprising the gene coding for the endogenous GnT I.

3) Screening of microalgae expressing the GnT I

a) Analysis of the expression of GnT I by RT-PCR

A transcriptomic analysis was realised on transformed clones with the sequence coding for the *P. tricornutum* GnT I. The total RNAs from the different clones were purified. The corresponding cDNA were synthesized then analysed by Reverse Transcription PCR amplification according to the instructions as disclosed in Example 3 with specific primers for the transgene (SEQ ID N°71 and SEQ ID N°72) and with the H4 housekeeping gene (SEQ ID N°73 and SEQ ID N°74).

The results of the amplifications realised with specific primers from the H4 housekeeping gene are presented in figure 6. The absence of amplification for the RT- samples enables to validate that the samples did not contain genomic DNA.

The amplifications realised on the RT+ samples enabled to obtain amplicons at 150 bp which were homogenous between all the *Pt*-GnT I clones. This study therefore enabled to validate the experimental conditions.

The *Pt*-GnT I clones were then analysed with primers enabling the highlight of the expression of GnT I (see figure 7).

The figure 7 represents the RT-PCR results obtained for the expression of GnT I in the *P. tricornutum* transformed clones. The absence of amplification in the RT- samples validates the quality of the samples. Furthermore, the amplification fragment of 212 bp with a higher

intensity than the one of the wild type strain is obtained for the clones 1 and 3. This difference of signal corresponds to a surexpression of GnT I. Therefore, the inventors were able to identify microalgae clones which surexpressed the GnT I. A thinner analysis of the expression of GnT I was proceeded by quantitative PCR (Q-PCR).

b) Analysis of the expression of GnT I by Q-PCR

A lineage from transformed *P. tricornutum* with a construction comprising the GnT I gene was selected by RT-PCR. The expression of GnT I was then analysed by Q-PCR as disclosed in Siaut *et al.*, 2007, Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*, Gene 406 (1-2): 23-35, by using primers directed to the transgene (SEQ ID N°75 and SEQ ID N°76) and the H4 housekeeing gene (SEQ ID N° 77 and SEQ ID N°78).

The figure 8A corresponds to a fusion curve. It enables to verify that only one PCR product has been amplified. The presence of one peak shows that only one amplification fragment was obtained. The dissociation temperature of 81,2°C shows a strong specificity of the amplicon-primers association.

The figure 8B presents the detection profile of GnT I in the transformed cells (*Pt*-GnT I) and the wild-type cells (*Pt*-wt). For each strain, 3 dilutions of cDNA were analysed. The results were expressed in number of cycles compared to the relative values of detection of the GnT I amplicon.

The 3 sigmoids corresponding to negative controls (cDNA- and water) show a late detection, which is indicative of a non-significative amplification of GnT I.

The analysis of the transformed lineage with GnT I showed a detection which was significantively earlier than the wild-type cells. The data obtained were normalized with a housekeeping gene H4. The normalisation compared to the housekeeping gene by using the comparative Ct method also called as $2^{-[delta][delta]Ct}$ (△△CT method) revealed a clear difference of expression profile between the strain expressing the GnT I transgene and the wild-type cells (△△CT = 7).

Example 5 Display of the expression and localisation of the GnT I-eGFP protein

1) By Confocal microscopy

The clones of the *P. tricornutum* microalgae which were transformed with a sequence coding for a GnT I-eGFP fusion protein were analyzed by confocal microscopy with standard parameters. The figure 9 presents the results of said analysis.

The observations realized clearly show a difference of fluorescence localization between the microalgae expressing a cytosolic eGFP and the GnT I-eGFP fusion protein. In that case, the fluorescence is much more localised and is presented under the form of a bowl or a bean. This marking is typical of the Golgi apparatus in *P. tricornutum* microalgae. The inventors therefore demonstrated that the endogenous GnT I was expressed and localized in the suitable cell compartment.

2) By analysis after treatment of the culture with brefeldin A and epifluorescence microscopy

In order to validate those results, the inventors analyzed the clones expressing a GnT I-eGFP fusion protein after a treatment of the culture of the transformed microalgae with brefeldin A compared to microalgal strain of *P. tricornutum* expressing eGFP in the chloroplast, in the endoplasmic reticulum or in the Golgi apparatus after a culture with brefeldin A (Figure 10).

The brefeldin A is a molecule presenting a particularity of dismantling the structure of the Golgi apparatus in the cells. Therefore, it is possible to confirm the localization of the fusion protein which is produced in *P. tricornutum* microalgae. The cells were incubated with 50µM of brefeldin A and the observations were conducted with an epifluorescence microscopy on a period of 48 hours.

The figure 10 presents the results which were obtained. No fluorescence variation was observed in the *P. tricornutum* microalgae expressing eGFP in the chloroplast or in the Endoplasmic Reticulum after a culture with brefeldin A. In the microalgae expressing the GnT I-eGFP fusion protein, small fluorescent vesicles rapidly appeared (after 5 hours on the picture), which is indicative of the effect of brefeldin A on the Golgi apparatus. Said effect became more and more visible after 24 and 48 hours of culture with a marking spreading to the whole cell. The Golgi marking which was very targeted almost disappeared.

This analysis enabled to validate that the GnT I-eGFP fusion protein is expressed and localized in the Golgi apparatus of transformed *P. tricornutum* microalgae.

3) Immunoblotting analysis by Western Blot

Aliquotes of wild-type and transformed cells of *P. tricornutum* culture at exponential phase of growth are collected and cells are separated from the culture medium by centrifugation (10 minutes, 2150 g, 20°C). Cell pellets are resuspended in Tris-HCl 0.15 M pH 8, saccharose 15%, SDS 0.5%, PMSF 1 mM, protease inhibitor cocktail 1% (SIGMA) and sonicated for 30 min. Cell suspensions obtained are centrifuged (60 minutes, 15000 g, 4°C) to remove cell debris and supernatants correspond to the intracellular fraction.

Ten µL of intracellular fractions from *Pt*-eGFP, *Pt*-GntI-eGFP transformed and wild-type cells are separated by SDS-PAGE using a 12% polyacrylamide gel. The separated proteins are transferred onto nitrocellulose membrane and stained with Ponceau Red in order to control transfer efficiency. The nitrocellulose membrane is blocked overnight in milk 5% dissolved in TBS for immunodetection. Immunodetection is then performed using horseradish peroxidase-conjugated anti-GFP (Santa Cruz, sc-9996-HRP) (1:2000 in TBS-T containing milk 1% for 2h at room temperature). Membranes are then washed with TBS-T (6 times, 5 minutes, room temperature) followed by a final wash with TBS (5 minutes, room temperature). Final development of the blots is performed by chemiluminescence method.

The total protein extracts from wild-type *P. tricornutum* microalgae, from *P. tricornutum* microalgae expressing the eGFP protein (*Pt*-eGFP) and from *P. tricornutum* microalgae expressing the GnT I-eGFP fusion protein (*Pt*-GnT I-eGFP) were analyzed by Western Blot with anti-GFP antibodies which were coupled to peroxidase.

The figure 11 presents the results obtained for the detection of the GnT I-eGFP fusion protein by Western Blot. The figure shows a detection at around 28 kDa for the cytosolic eGFP and around 75 kDa for the clones which were transformed with a sequence coding for the GnT I-eGFP fusion protein, which theoretical size corresponds to 77 kDa.

The presence of the fusion protein in the protein samples from transformed microalgae is therefore demonstrated. The absence of degradation is a proof that the fusion protein is stable.

This analysis therefore demonstrates that it is possible to express the *P. tricornutum* endogenous GnT I sequence and that said protein is addressed to the suitable cell compartment, i.e. the Golgi apparatus.

Example 6 Glycosylation pattern of proteins expressed in transformed *P. tricornutum* with endogenous GnT I

In order to investigate the presence in *P. tricornutum* proteins of complex-type N-glycans carrying terminal GlcNAc, we treat the protein extract with a β1,4-galactosyltransferase, an enzyme able to transfer a galactose residue onto terminal GlcNAc residues, and then we analyze by affinoblotting the resulting protein preparation either with RCA 120 or ECA, which are lectins that bind specifically to Galβ1-4GlcNAc sequences according to a strategy previously reported (Bardor *et al.*, 2003, *Plant Biotech* J, vol.**1**, 451-462). Alternatively, the glycosylation pattern of proteins expressed in transformed *P. tricornutum* with endogenous GnT I is study by approaches which has been mentioned above.

Example 7 Expression of α-Man II in *Phaeodactylum tricornutum*

A gene encoding a putative α-Man II was predicted in the *P. tricornutum* genome (Pt52108).

The Pt52108 gene (SEQ ID N°6) is cloned under of an enhanced promoter of Cauliflower Mosaic Virus in a pPha-T1 based vector called BSJ-25 vector. As a control, vectors comprising α-Man II from Arabidopsis thaliana (NP_196999) and from human cell (Q16706 or AAC50302) are also constructed. BSJ-25 is derived from pPha-T1 vector (Zaslavskaia and Lippmeier, 2000, J. Phycol. vol.36(2), p: 379-386) by replacing the *FcpA* promoter of the expression cassette by the double enhanced Cauliflower Mosaic Virus (CaMV35S) fused to the plant signal peptide. For this, pPha-T1 containing bleomycin-resistance cassette driven by *FcpB* promoter is digested with Nde1 and EcoR1 to remove the FcpA promoter. This construct was designated as "pPha-T1-PfcpA deleted". The double enhanced Cauliflower Mosaic Virus 35S promoter of the expression cassette is amplified by PCR with forward primer, CaMV35Sfwd (5'-GAACATATGGTGGATTGATGTGATCTACTCC-3') (SEQ ID N°61) and reverse primer, CaMV35Srev (5'-AATTCTCGAGGAATTCGGCCGAGG -3') (SEQ ID N°62) on the PS1 construct (Kotzer *et al.,* 2004, J Cell Sci, vol.117(Pt 26), p: 6377-89). PS1 construct contains a double Cauliflower Mosaic Virus 35S promoter, a tobacco Mosaic Virus - Ω sequence as translation enhancer fused to the tobacco chitinase signal peptide (Haseloff *et al.,* 1997, Proc Natl Acad Sci U S A, vol.94(6), p: 2122-7.; Batoko *et al.,* 2000, Plant Cell, vol.12(11), p: 2201-18). The PCR product is digested by Nde1 and EcoR1 and then is cloned into "pPha-T1-PfcpA deleted" to generate BSJ-25 vector.

To resume, the expression cassette of BSJ-25 vector contain a double CaMV promoter, a translational enhance, a signal peptide, a multi-cloning site and the *FcpA*

terminator (Fig 3).

Specific vectors expressing the abovementioned genes in fusion to Green Fluorescent Protein (GFP) are also constructed to investigate the cellular localisation in the Golgi apparatus of the corresponding fusion proteins.

Said vectors are used to transform *Phaeodactylum tricornutum*.

Transformation of *Phaeodactylum tricornutum* is carried out as disclosed previously in the example of expression of GnT I in *Phaeodactylum tricornutum*.

The presence of transcripts for the recombinant α-Man II is then monitored by reverse transcription PCR (RT-PCR). Microalgae pellet are resuspended in 1.2 mL of Trizol (INVITROGEN) and homogenized in 2 mL Lysing Matrix E by 10 Fastprep-24 run (6.5m / s, 60 sec, MP BIOMEDICALS). 300 µL of chloroform are added to the supernatant, vigorously homogenized and incubated 3 min at room temperature. After a 15 min centrifugation at 4°C, the aqueous is mixed with 1 volume of absolute ethanol and the RNAs were purified by RNEASY MINI kit (QIAGEN). The purified RNAs are eluted in 50 µL of RNase-free water, dosed with NanoDrop (Thermo SCIENTIFIC) and are digested by RQ1 DNase (PROMEGA). After second purification by RNEASY MINI kit and Nanodrop quantification, RNAs are conserved at -80°C.

Reverse Transcription (RT) is performed on 1 µg of purified RNA. The first cDNA strand is synthesised by 200 units of M-MLV RT RNase H minus (PROMEGA) with oligodT primers. Two µL of cDNA are used for PCR with GoTaq polymerase (PROMEGA) in 50 µL. The annealing is performed with specific primers of GnT I and actin sequence.

The results confirm the expression of transcripts for the recombinant α-Man II in the selected recombinants (data not shown).

The glycosylation pattern of proteins expressed in said transformants is then performed as disclosed previously. Simultaneously, the cellular localisation of the fusion proteins is investigated by observing the GFP fluorescence of GnT I-GFP fusion proteins by confocal microscopy.

Example 8 Glycosylation pattern of proteins expressed in transformed *P. tricornutum* with endogenous alpha Man II

The glycosylation pattern of proteins expressed in transformed *P. tricornutum* with endogenous alpha Man II is studying by HPLC and Mass spectrometry approaches as mentioned above.

Example 9 Inactivation of N-acetylglucosaminidases in *Phaeodactylum tricornutum* with RNA interference

a) by the use of vectors

This example is based on the use of constructs containing antisense RNA or inverted-repeat RNA for the inactivation of β-N-acetylglucosaminidase expression as described in De Riso V. *et al.* , 2009, Nucleic Acids Research, p:1–12.

For the generation of a β -N-acetylglucosaminidase vector, two fragments from the β -N-acetylglucosaminidase cDNA are amplified with specific primers, wherein one of the fragments is longer than the other fragment and also contains the total sequence of the shorter fragment. The two obtained fragments are then digested with restriction enzymes which can also be used for the linearization of the vector.

For the antisense construct, one of the two fragments obtained is inserted into the vector in the antisense orientation.

For the inverted-repeat construct, fragments are ligated in sense and antisense orientations in the integration site of the vector.

Said vectors are then used together or separately by choosing one vector to transform *Phaeodactylum tricornutum*. Transformation of *Phaeodactylum tricornutum* is carried out as described in example 3.

b) By the use of siRNAs

The gene silencing of β -N-acetylglucosaminidases can also be obtained with the transformation of *Phaeodactylum tricornutum* with the use of siRNAs specific of at least one gene encoding a β-N-acetylglucosaminidase in said microalga (1 µg for each siRNA). Transformation of *Phaeodactylum tricornutum* is carried out as described in example 3.

siRNAs used for the transformation of *Phaeodactylum tricornutum* have the following sequences :

- siRNAs specific of a first β-N-acetylglucosaminidase of amino acid sequence SEQ ID N°9 which is encoded by the nucleic acid sequence SEQ ID N°10

SEQ ID N°63	siRNA1*(β-N-acetylglucosaminidase SEQ ID N°10)*	GGCCAUUCGUUACUAGCAA
SEQ ID N°64	siRNA2*(β-N-acetylglucosaminidase SEQ ID N°10)*	GUGGUUCGUUGGGAAAUGA
SEQ ID N°65	siRNA3*(β-N-acetylglucosaminidase SEQ ID N°10)*	CCGUCUGUGUGAAAUUGGU

- siRNAs specific of a second β-N-acetylglucosaminidase of amino acid sequence SEQ ID N°11 which is encoded by the nucleic acid sequence SEQ ID N°12

SEQ ID N°66	siRNA1*(β-N-acetylglucosaminidase SEQ ID N°12)*	CGGUAGUAGUGCUUGUUGU
SEQ ID N°67	siRNA2*(β-N-acetylglucosaminidase SEQ ID N°12)*	CUGCCAUGGAUAUUGUCAA
SEQ ID N°68	siRNA3*(β-N-acetylglucosaminidase SEQ ID N°12)*	GCCUUUGGUCCUGAAGAAA

Example 10 Expression and Glycosylation of a polypeptide in transformed *P. tricornutum*

As explained previously, polypeptides of interest to be expressed and glycosylated in transformed *P. tricornutum* according to the invention are proteins of therapeutic interest. We choose the erythropoietin to exemplify the expression and glycosylation of such proteins in transformed *P. tricornutum* according to the invention. However, the invention is not limited to said erythropoietin and could be applied to any protein of therapeutic interest which needs to be glycosylated to present the adequate glycosylation pattern.

a) Expression construct for Erythropoietin

The vector used for the expression of Erythropoietin comprises a nucleic acid sequence operably linked to a promoter, said nucleic acid sequence encoding Erythropoietin. Said vector preferably contains a selectable marker distinct from the selectable markers present on the BSJ-25 vector according to the example 3.

b) Transformation of *P. tricornutum*

Transformed *P. tricornutum* of example 3 are also transformed with a vector according to step a).

c) Glycosylation pattern analysis of expressed and glycosylated Erythropoietin

The glycosylation pattern of Erythropoietin that is expressed and glycosylated in the transformed *P. tricornutum* according to the invention is described in the example 1.

CLAIMS

1. A transformed *Phaeodactylum tricornutum* whose N-glycosylation pathway has been modified by the inactivation of at least one β-N-acetylglucosaminidase and/or the expression of at least one glycosylation enzyme encoded by a nucleic acid sequence operatively linked to a promoter, wherein

   (i) said nucleic acid sequence encodes an N-acetylglucosaminyltransferase I having the amino acid sequence SEQ ID N°1, a fragment or a derivative thereof and said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°3 or a fragment thereof, and/or

   (ii) said nucleic acid sequence encodes an α-Mannosidase II having the amino acid sequence SEQ ID N°5, a fragment or a derivative thereof and said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°7 or a fragment thereof.

2. A transformed *Phaeodactylum tricornutum* according to claim 1, wherein the N-glycosylation pathway of said transformed *Phaeodactylum tricornutum* has been modified by the inactivation of at least one β-N-acetylglucosaminidase.

3. A transformed *Phaeodactylum tricornutum* according to claim 1, wherein said transformed *Phaeodactylum tricornutum* whose N-glycosylation pathway has been modified comprises a nucleic acid sequence operatively linked to a promoter,

   (i) said nucleic acid sequence encodes an N-acetylglucosaminyltransferase I having the amino acid sequence SEQ ID N°1, a fragment or a derivative thereof, and

   (ii) said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°3 or a fragment thereof.

4. The transformed *P. tricornutum* according to claim 1, wherein said transformed *P. tricornutum* whose N-glycosylation pathway has been modified comprises a nucleic acid sequence operatively linked to a promoter,

   (i) said nucleic acid sequence encodes an α-Mannosidase II having the amino acid sequence SEQ ID N°5, a fragment or a derivative thereof, and

   (ii) said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°7 or a fragment thereof.

5. A transformed *Phaeodactylum tricornutum* according to claims 2 and 3, or according to claims 3 and 4, or according to claims 2 and 4, or according to claims 2 and 3 and 4.

6. The transformed *P. tricornutum* according to any one of claims 1 to 5, further comprising a nucleic acid sequence operatively linked to a promoter, wherein said nucleic acid sequence encodes an N-acetylglucosaminyltransferase II, a fragment or a derivative thereof.

7. The transformed *P. tricornutum* according to claim 6, wherein

   (i) said N-acetylglucosaminyltransferase II comprises the amino acid sequence SEQ ID N°15, and

   (ii) said operatively linked promoter has a sequence identity of less 50% with SEQ ID N°7 or a fragment thereof.

8. The transformed *P. tricornutum* according to any one of claims 1 to 7, wherein said *P. Tricornutum* further comprises another nucleic acid sequence operatively linked to a promoter, said other nucleic acid sequence encoding a polypeptide that is expressed and glycosylated in the transformed *P. tricornutum*.

9. The transformed *P. tricornutum* according to claim 8, wherein said glycosylated polypeptide has at least one $GlcNAcMan_5GlcNAc_2$ structure, preferably one $GlcNAcMan_4GlcNAc_2$,$GlcNAcMan_3GlcNAc_2$ and/or $GlcNac_2Man_3GlcNAc_2$ structure.

10. The transformed *P. tricornutum* according to any of claims 8 or 9, wherein said polypeptide is selected from the group comprising erythropoietin, cytokines, antibodies and their fragments, coagulation factors, hormones, beta-glucocerebrosidase, pentraxin-3, anti-TNFs, acid α-glucosidase, α-L-iduronidase and derivatives thereof .

11. The transformed *P. tricornutum* according any one of claims 1 to 10 wherein said *P. tricornutum* further comprises still another nucleic acid sequence operatively linked to a promoter, said nucleic acid sequence encoding at least one enzyme selected among N-acetylglucosaminyltransferase III, IV, V and VI.

12. The transformed *P. tricornutum* according to claim 11, wherein said *P. tricornutum* further comprises another nucleic acid sequence operatively linked to a promoter, said

nucleic acid sequence encoding at least one glycosyltransferase enzyme selected in the group comprising galactosyltransferases, fucosyltransferases and sialyltransferases.

13. A method for producing a glycosylated polypeptide, said method comprising the steps of

(i) Culturing a transformed *P. tricornutum* as defined in any one of claims 8 to 12

(ii) Purifying said polypeptide that is expressed and glycosylated in said transformed *P. tricornutum*.

14. The method according to claim 13 wherein said method comprises a step (iii) of determining the glycosylation pattern of said polypeptide.

15. A use of a transformed *P. tricornutum* as defined in any one of claims 8 to 12 for producing a glycosylated polypeptide.

ABSTRACT

The present invention concerns transformed *Phaeodactylum tricornutum* comprising a nucleic acid sequence operatively linked to a promoter, wherein said nucleic acid sequence encodes an N-acetylglucosaminyltransferase I and/or an α-Mannosidase II and wherein at least one β-N-acetylglucosaminidase of said transformed *Phaeodactylum tricornutum* has been inactivated; a method for producing a glycosylated polypeptide, said method comprising the steps of (i) culturing a transformed *P. tricornutum* as defined previously and (ii) purifying said polypeptide that is expressed and glycosylated in said transformed *P. tricornutum*; and the use of such a transformed *P. tricornutum* for producing a glycosylated polypeptide.

1/10

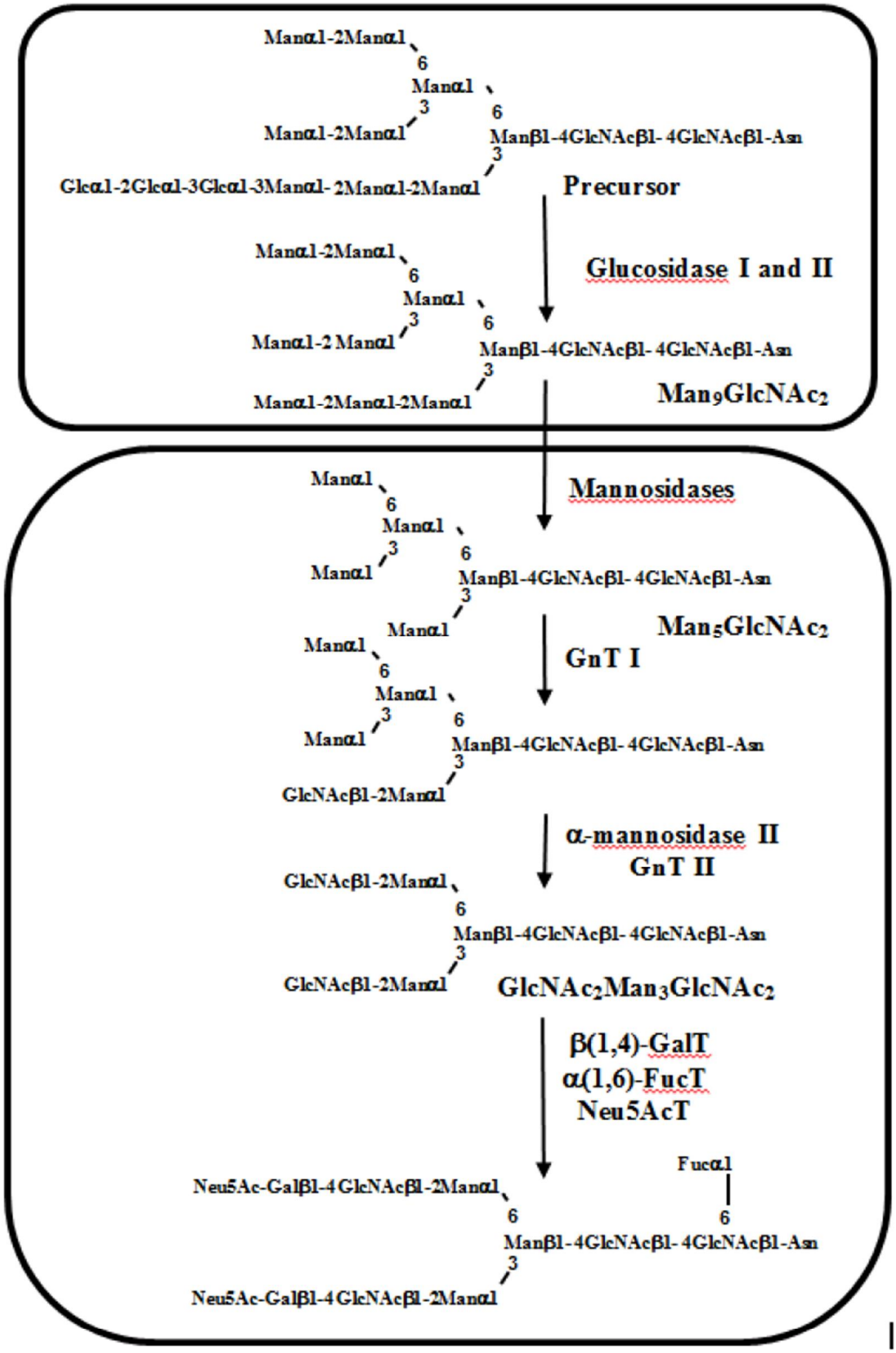

Manα1-2Manα1
6
Manα1
3
6
Manα1-2Manα1
Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn
3
Glcα1-2Glcα1-3Glcα1-3Manα1-2Manα1-2Manα1
Precursor
Glucosidase I and II
Man9GlcNAc2
Mannosidases
Man5GlcNAc2
GnT I
α-mannosidase II
GnT II
GlcNAc2Man3GlcNAc2
β(1,4)-GalT
α(1,6)-FucT
Neu5AcT
Fucα1
Neu5Ac-Galβ1-4GlcNAcβ1-2Manα1

Figure 1

2/10

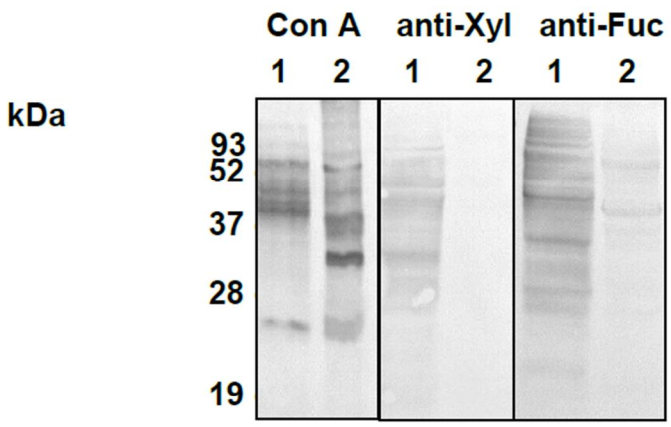


Figure 2

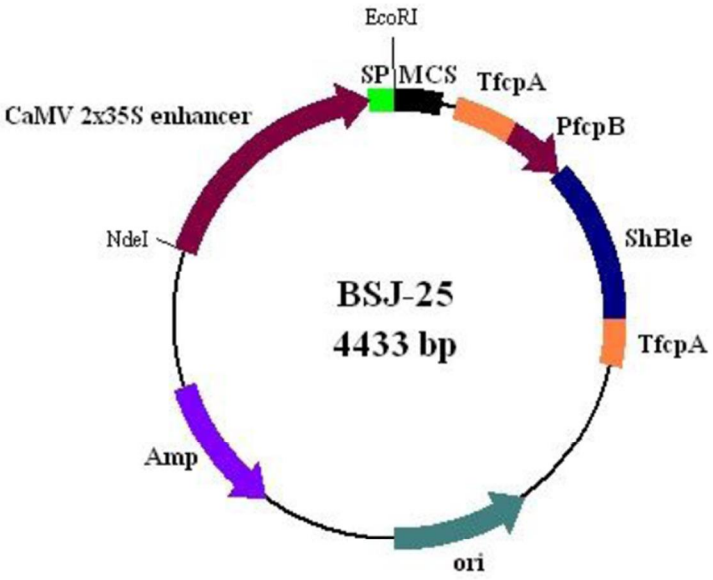


Figure 3

3/10

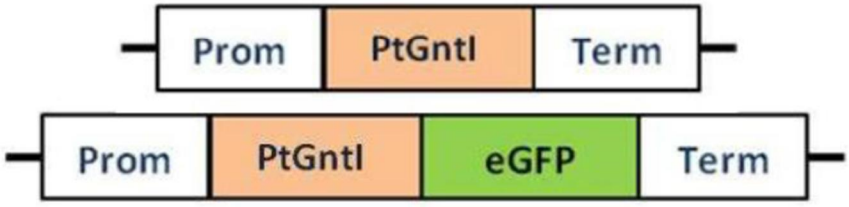


Figure 4

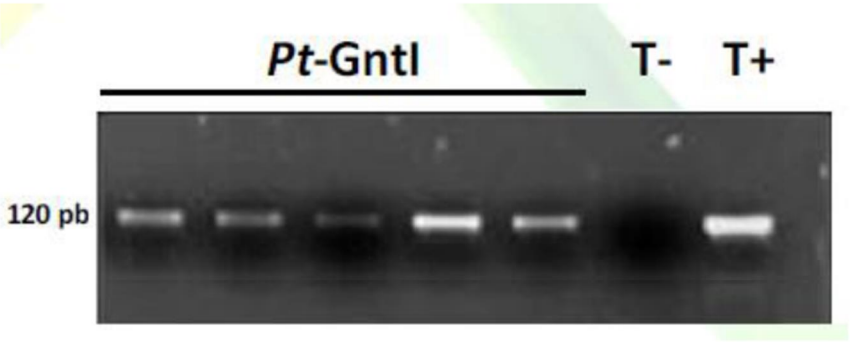


Figure 5

4/10

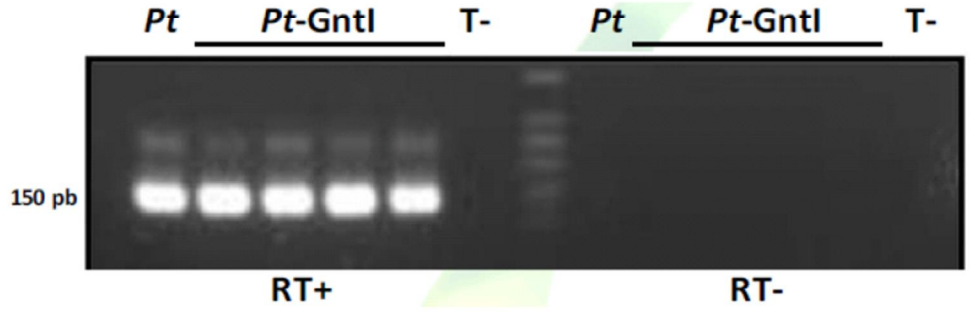


Figure 6

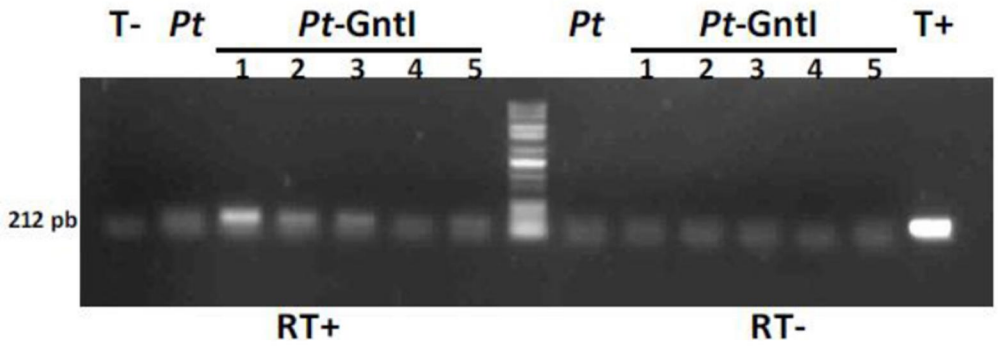


Figure 7

5/10

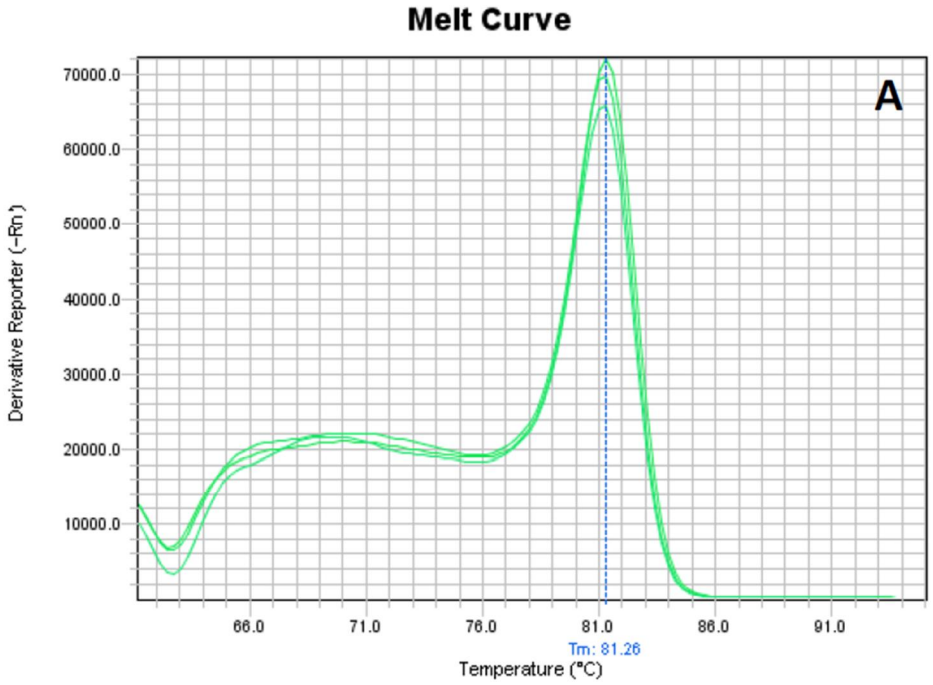


Figure 8A

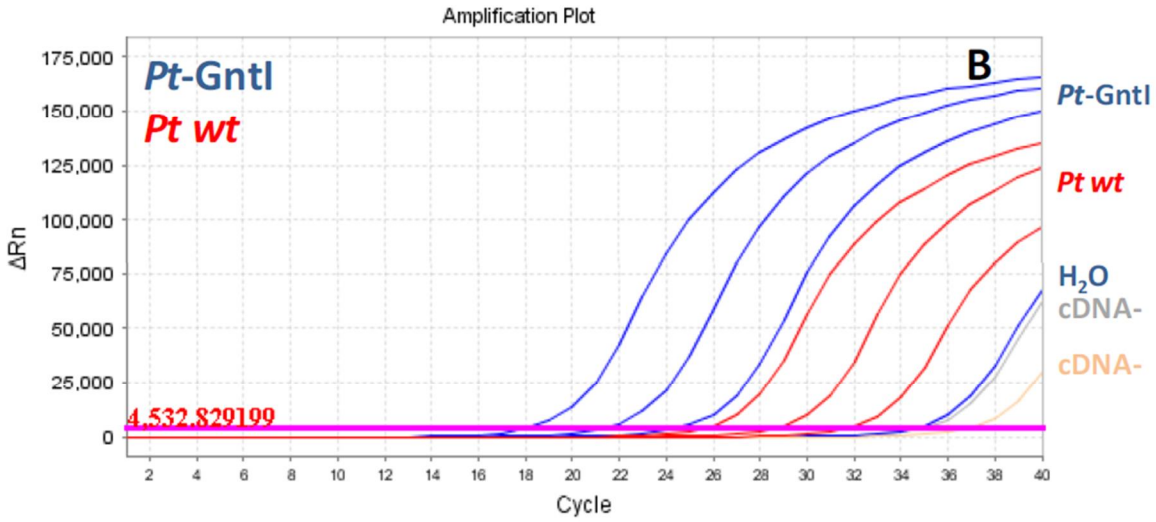


Figure 8B

6/10

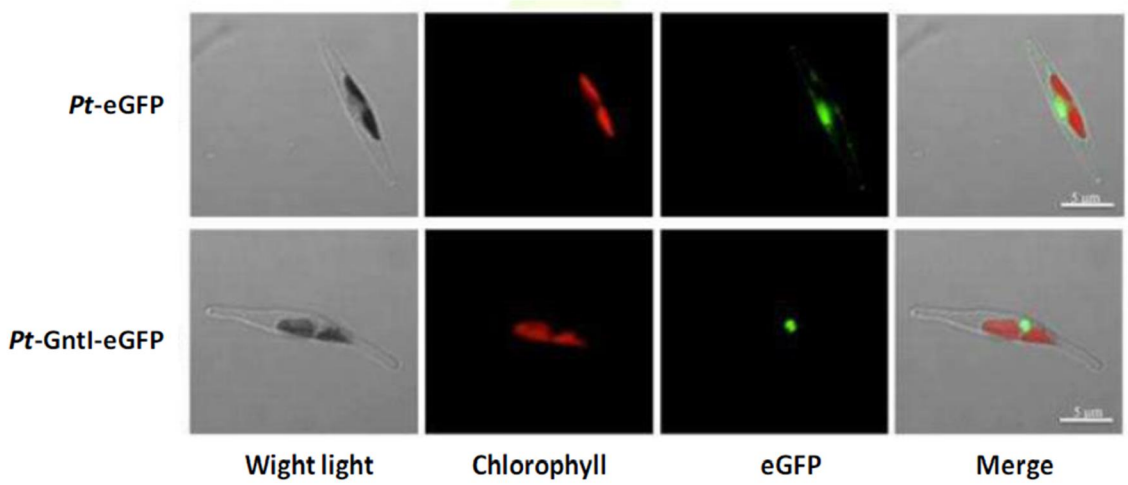


Figure 9

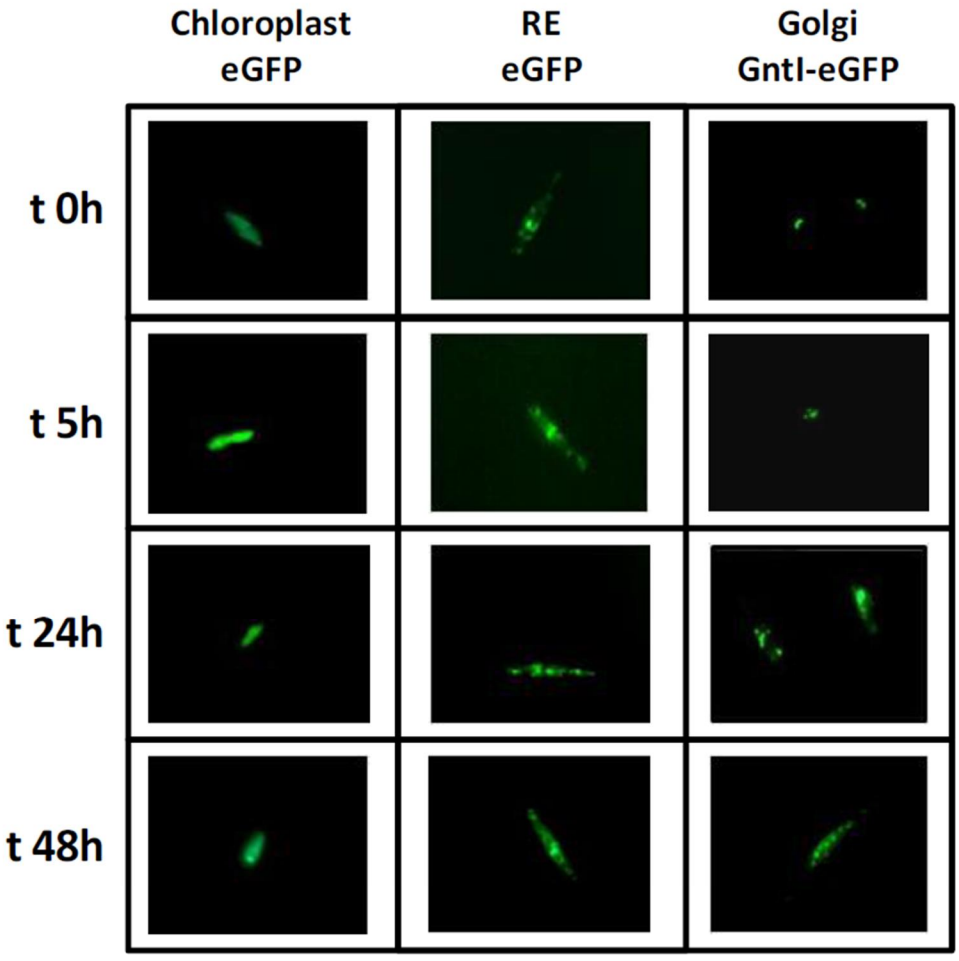


Figure 10

7/10

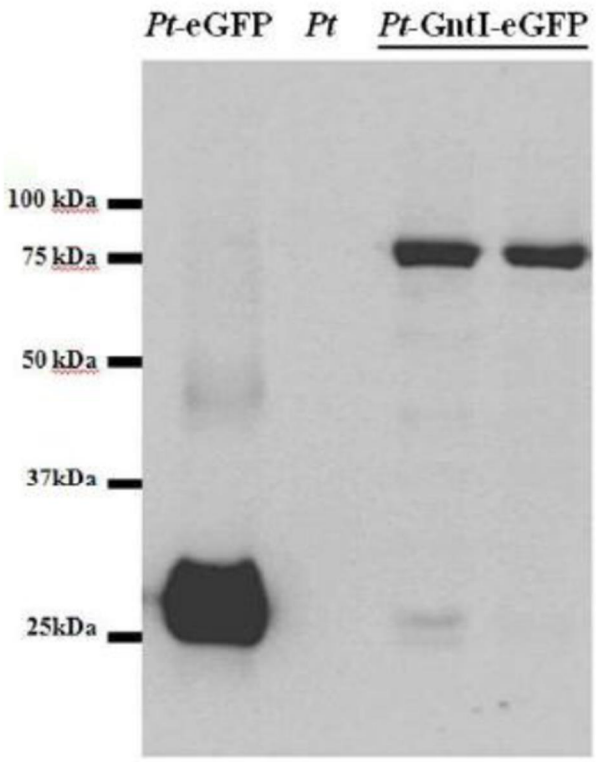


Figure 11

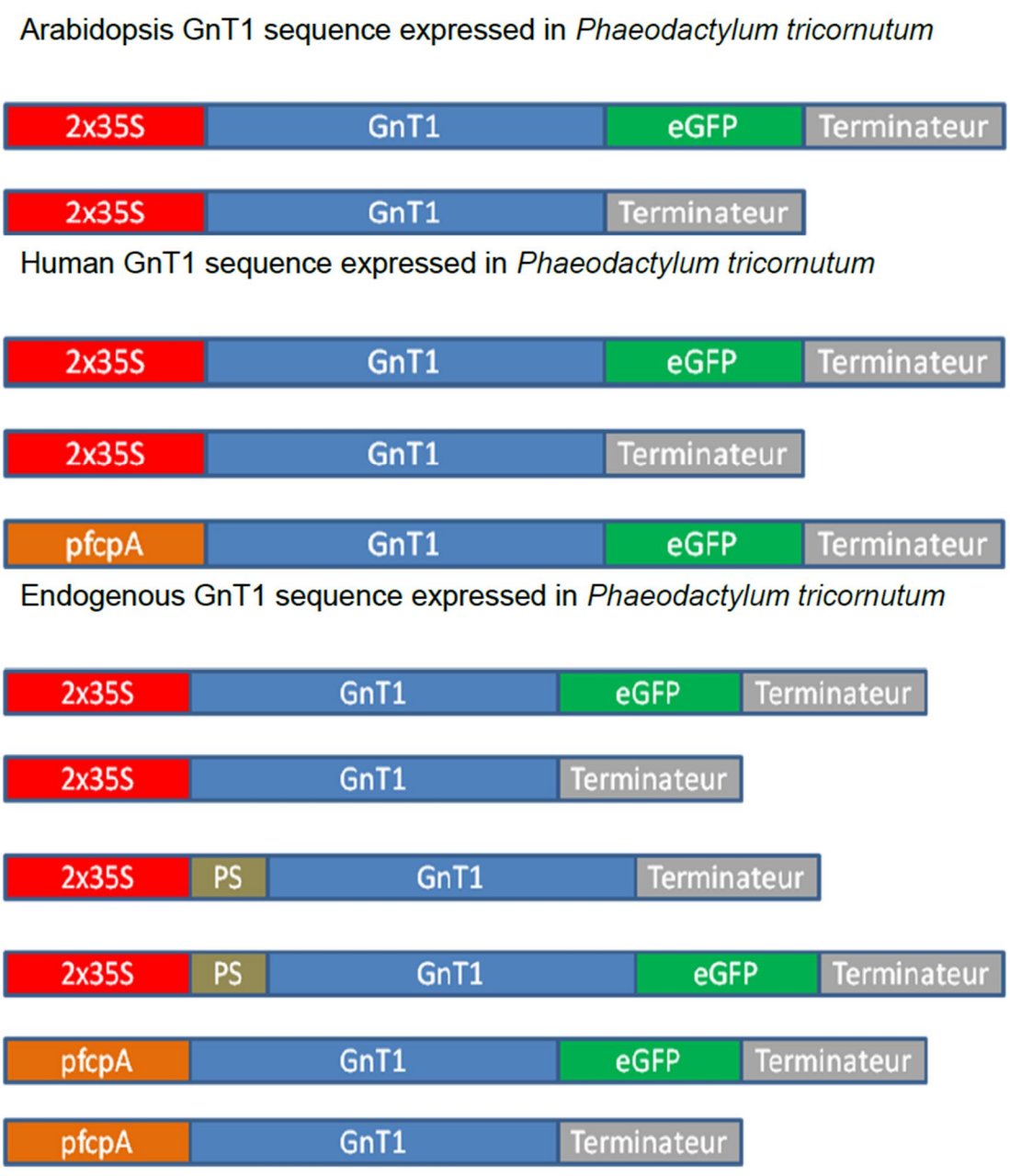

Arabidopsis GnT1 sequence expressed in *Phaeodactylum tricornutum*
2x35S
GnT1
eGFP
Terminateur
2x35S
GnT1
Terminateur
Human GnT1 sequence expressed in *Phaeodactylum tricornutum*
2x35S
GnT1
eGFP
Terminateur
2x35S
GnT1
Terminateur
pfcpA
GnT1
eGFP
Terminateur
Endogenous GnT1 sequence expressed in *Phaeodactylum tricornutum*
2x35S
GnT1
eGFP
Terminateur
2x35S
GnT1
Terminateur
2x35S
PS
GnT1
Terminateur
2x35S
PS
GnT1
eGFP
Terminateur
pfcpA
GnT1
eGFP
Terminateur
pfcpA
GnT1
Terminateur

Figure 12

9/10

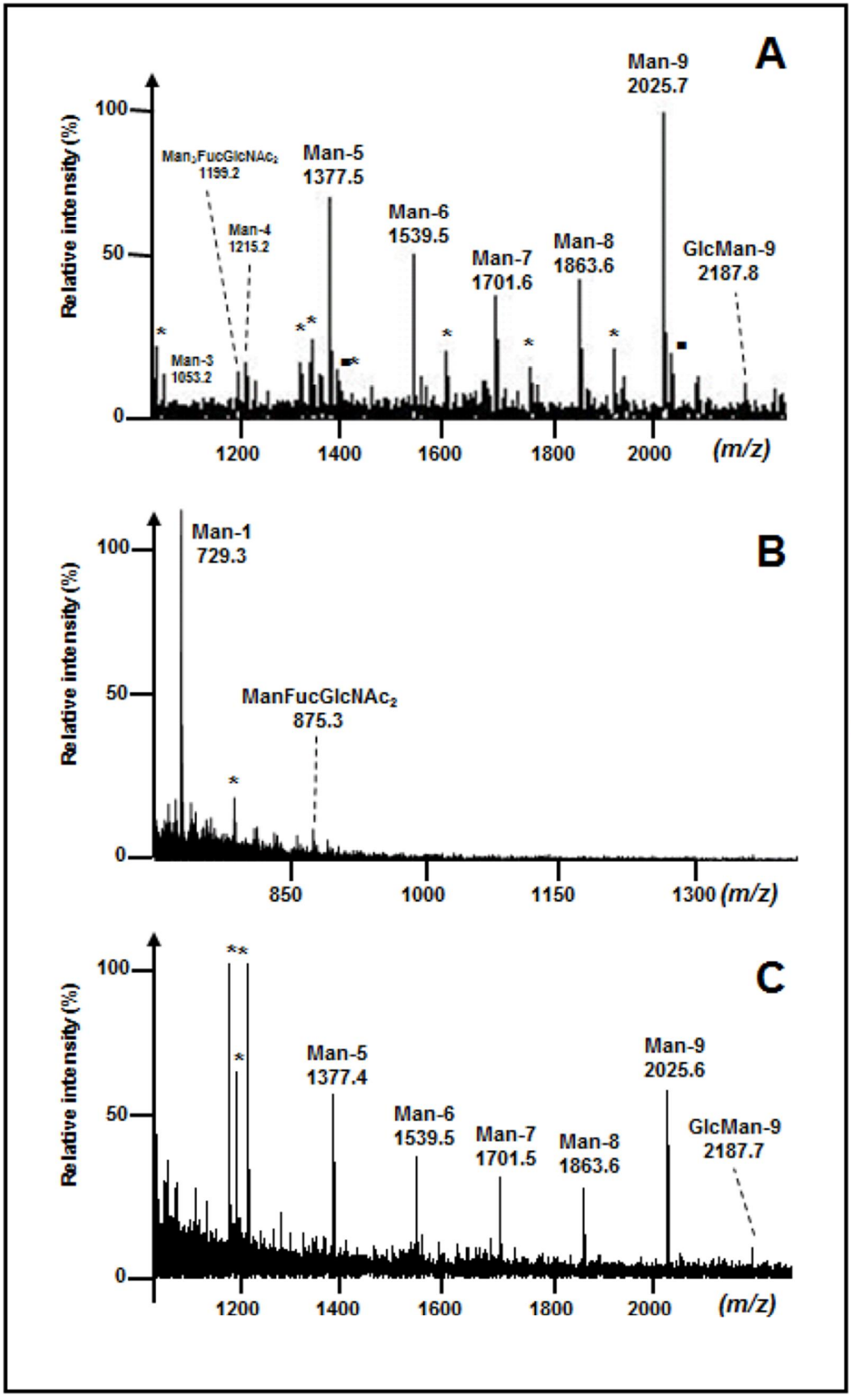


Figure 13

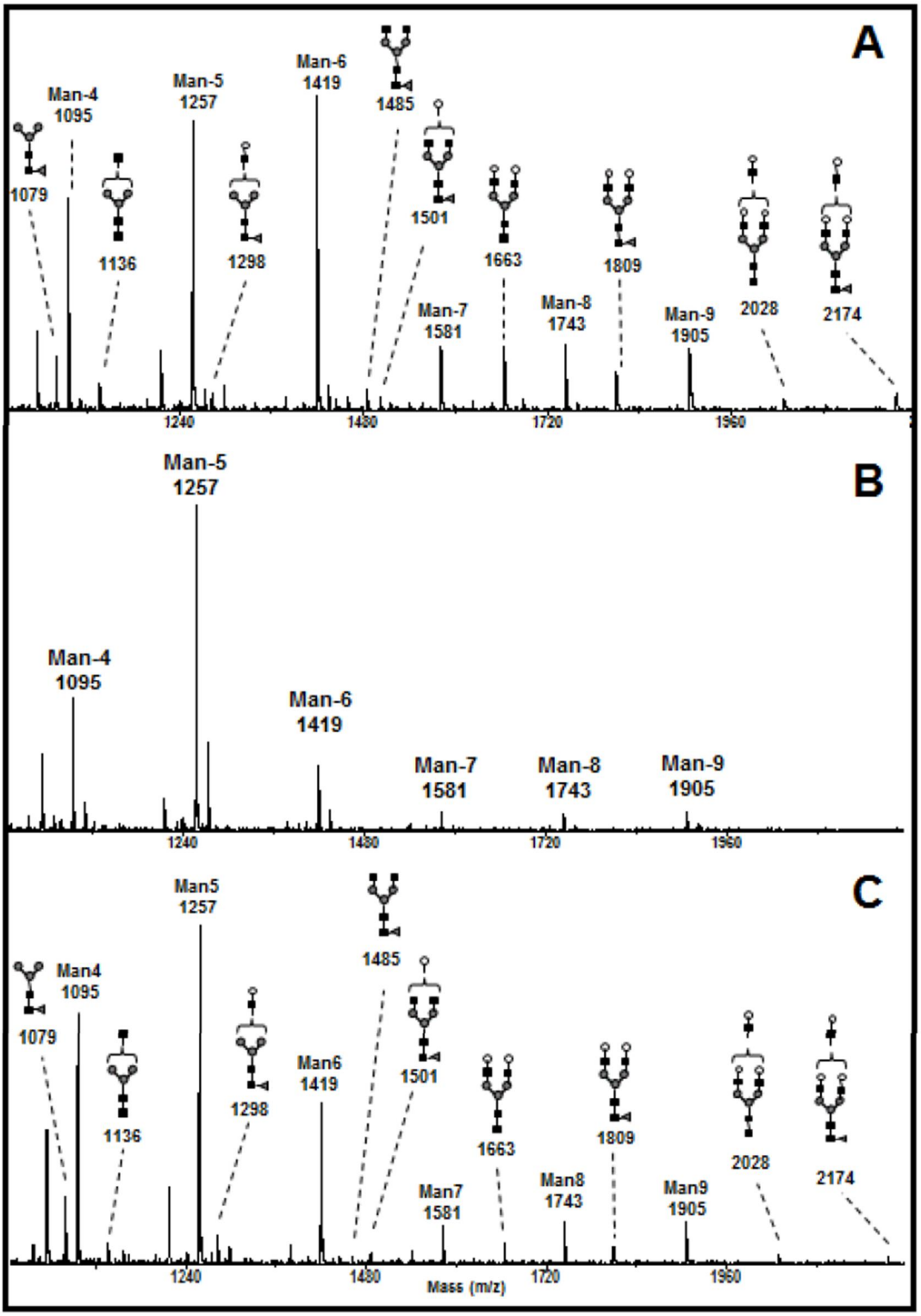

A
Man-4
1095
Man-5
1257
Man-6
1419
1485
1079
1136
1298
1501
1663
1809
2028
2174
Man-7
1581
Man-8
1743
Man-9
1905
1240
1480
1720
1960
B
Man-5
1257
Man-4
1095
Man-6
1419
Man-7
1581
Man-8
1743
Man-9
1905
1240
1480
1720
1960
C
Man5
1257
1485
Man4
1095
1079
1136
1298
Man6
1419
1501
1663
1809
2028
2174
Man7
1581
Man8
1743
Man9
1905
1240
1480
1720
1960
Mass (m/z)

Figure 14

Références bibliographiques

References

Aebi, M. (**2013**). N-linked protein glycosylation in the ER. Biochim. Biophys. Acta *1833*, 2430-2437.

Aebi, M. and Hennet, T. (**2001**). Congenital disorders of glycosylation: genetic model systems lead the way. Trends Cell Biol. *11*, 136-141.

Aebi, M., Bernasconi, R., Clerc, S. et al. (**2010**). N-glycan structures: recognition and processing in the ER. Trends Biochem. Sci. *35*, 74-82.

Aebi, M., Gassenhuber, J., Domdey, H. et al. (**1996**). Cloning and characterization of the ALG3 gene of Saccharomyces cerevisiae. Glycobiology *6*, 439-444.

Aikawa, J., Matsuo, I. and Ito, Y. (**2012**). In vitro mannose trimming property of human ER α-1,2 mannosidase I. Glycoconj. J. *29*, 35-45.

Allen, AE., Dupont, CL., Oborník, M. et al. (**2011**). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature *473*, 203-207.

Allen, AE., Laroche, J., Maheswari, U. et al. (**2008**). Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. Proc. Natl. Acad. Sci. U.S.A. *105*, 10438-10443.

Altmann, F. and März, L. (**1995**). Processing of asparagine-linked oligosaccharides in insect cells: evidence for alpha-mannosidase II. Glycoconj. J. *12*, 150-155.

Altmann, F. and März, L. (**1995**). Processing of asparagine-linked oligosaccharides in insect cells: evidence for alpha-mannosidase II. Glycoconj. J. *12*, 150-155.

Altmann, F., Schwihla, H., Staudacher, E. et al. (**1995**). Insect cells contain an unusual, membrane-bound beta-N-acetylglucosaminidase probably involved in the processing of protein N-glycans. J. Biol. Chem. *270*, 17344-17349.

Altschul, SF., Gish, W., Miller, W. et al. (**1990**). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

Apt, KE., Kroth-Pancic, PG. and Grossman, AR. (**1996**). Stable nuclear transformation of the diatom Phaeodactylum tricornutum. Mol. Gen. Genet. *252*, 572-579.

Arendt, CW. and Ostergaard, HL. (**1997**). Identification of the CD45-associated 116-kDa and 80-kDa proteins as the alpha- and beta-subunits of alpha-glucosidase II. J. Biol. Chem. *272*, 13117-13125.

Armbrust, EV. (**2009**). The life of diatoms in the world's oceans. Nature *459*, 185-192.

Armbrust, EV., Berges, JA., Bowler, C. et al. (**2004**). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science *306*, 79-86.

Armbrust, EV., Chisholm, SW. and Olson, RJ. (**1990**). ROLE OF LIGHT AND THE CELL CYCLE ON THE INDUCTION OF SPERMATOGENESIS IN A CENTRIC DIATOM1. J. Phycol. *26*, 470-478.

Audry, M., Jeanneau, C., Imberty, A. et al. (**2011**). Current trends in the structure-activity relationships of sialyltransferases. Glycobiology *21*, 716-726.

Bakker, H., Bardor, M., Molthoff, JW. et al. (**2001**). Galactose-extended glycans of antibodies produced by transgenic plants. Proc. Natl. Acad. Sci. U.S.A. *98*, 2899-2904.

Bardor, M., Faveeuw, C., Fitchette, A. et al. (**2003**). Immunoreactivity in mammals of two typical plant glyco-epitopes, core alpha(1,3)-fucose and core xylose. Glycobiology *13*, 427-434.

Barker, MK. and Rose, DR. (**2013**). Specificity of Processing α-glucosidase I is guided by the substrate conformation: crystallographic and in silico studies. J. Biol. Chem. *288*, 13563-13574.

Baurain, D., Brinkmann, H., Petersen, J. et al. (**2010**). Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol. Biol. Evol. *27*, 1698-1709.

Bause, E. (**1983**). Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem. J. *209*, 331-336.

Bavestrello, G., Arillo, A., Calcinai, B. et al. (**2000**). Parasitic diatoms inside antarctic sponges. Biol. Bull. *198*, 29-33.

Becker, E. (**2007**). Micro-algae as a source of protein. Biotechnol. Adv. *25*, 207 - 210.

Bencúr, P., Steinkellner, H., Svoboda, B. et al. (**2005**). Arabidopsis thaliana beta1,2-xylosyltransferase: an unusual glycosyltransferase with the potential to act at multiple stages of the plant N-glycosylation pathway. Biochem. J. *388*, 515-525.

Bendtsen, JD., Nielsen, H., von Heijne, G. et al. (**2004**). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783-795.

Bennett, AG. (**1920**). On the Occurrence of Diatoms on the Skin of Whales. Proceedings of the Royal Society of London. *91 (641)*, 352-357.

Berger, S., Menudier, A., Julien, R. et al. (**1995**). Endo-N-acetyl-beta-D-glucosaminidase and peptide-N4-(N-acetyl-glucosaminyl) asparagine amidase activities during germination of Raphanus sativus. Phytochemistry *39*, 481-487.

Bergé, JP., Bourgougnon, N., Carbonnelle, D. et al. (**1997**). Antiproliferative effects of an organic extract from the marine diatom Skeletonema costatum
(Grev.) Cleve. against a non-small-cell bronchopulmonary

carcinoma line (NSCLC-N6).. Anticancer Res. *17*, 2115–2120.

Bickel, T., Lehle, L., Schwarz, M. et al. (**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.

Bigge, JC., Patel, TP., Bruce, JA. et al. (**1995**). Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. Anal. Biochem. *230*, 229-238.

Birnboim, HC. and Doly, J. (**1979**). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. *7*, 1513-1523.

Borowitzka, MA. and Volcani, BE. (**1978**). THE POLYMORPHIC DIATOM PHAEODACTYLUM TRICORNUTUM: ULTRASTRUCTURE OF ITS MORPHOTYPES1,2. J. Phycol. *14*, 10-21.

Bougaran, G., DÃ©an, LL., Lukomska, E. et al. (**2003**). Transient initial phase in continuous culture of Isochrysis galbana affinis Tahiti. Aquat. Living Resour. *16*, 389 - 394.

Bowler, C., Allen, AE., Badger, JH. et al. (**2008**). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature *456*, 239-244.

Bowler, C., Vardi, A. and Allen, AE. (**2010**). Oceanographic and biogeochemical insights from diatom genomes. Ann Rev Mar Sci *2*, 333-365.

Bozarth, A., Maier, U. and Zauner, S. (**2009**). Diatoms in biotechnology: modern tools and applications. Appl. Microbiol. Biotechnol. *82*, 195-201.

Brada, D. and Dubach, UC. (**1984**). Isolation of a homogeneous glucosidase II from pig kidney microsomes. Eur. J. Biochem. *141*, 149-156.

Breitling, J. and Aebi, M. (**2013**). N-linked protein glycosylation in the endoplasmic reticulum. Cold Spring Harb Perspect Biol *5*, a013359.

Breton, C. and Imberty, A. (**1999**). Structure/function studies of glycosyltransferases. Curr. Opin. Struct. Biol. *9*, 563-571.

Breton, C., Bettler, E., Joziasse, DH. et al. (**1998**). Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. J Biochem *123*, 1000-1009.

Breton, C., Fournel-Gigleux, S. and Palcic, MM. (**2012**). Recent structures, evolution and mechanisms of glycosyltransferases. Curr. Opin. Struct. Biol. *22*, 540-549.

Breton, C., Oriol, R. and Imberty, A. (**1998**). Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. Glycobiology *8*, 87-94.

Breton, C., Snajdrová, L., Jeanneau, C. et al. (**2006**). Structures and mechanisms of glycosyltransferases. Glycobiology *16*, 29R-37R.

Breuer, W., Klein, RA., Hardt, B. et al. (**2001**). Oligosaccharyltransferase is highly specific for the hydroxy amino acid in Asn-Xaa-Thr/Ser. FEBS Lett. *501*, 106-110.

[Brockhausen1997] Brockhausen, I. and Schachter, H. Glycosyltransferases involved in N- and O-glycan biosynthesis. In *Glycosciences*. . 1997. 79-113.

Brocks, JJ., Logan, GA., Buick, R. et al. (**1999**). Archean molecular fossils and the early rise of eukaryotes. Science *285*, 1033-1036.

Bræk, GS., Malnes, D. and Jensen, A. (**1980**). Heavy metal tolerance of marine phytoplankton. IV. Combined effect of zinc and cadmium on growth and uptake in some marine diatoms. J. Exp. Mar. Biol. Ecol. *42*, 39 - 54.

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. and Aebi, M. (**1999**). The dolichol pathway of N-linked glycosylation. Biochimica et Biophysica Acta (BBA) - General Subjects *1426*, 239-257.

Burda, P., Jakob, CA., Beinhauer, J. et al. (**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.

Burki, F., Okamoto, N., Pombert, J. et al. (**2012**). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc. Biol. Sci. *279*, 2246-2254.

Burki, F., Shalchian-Tabrizi, K., Minge, M. et al. (**2007**). Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE *2*, e790.

Cadoret, J. and Bernard, O. (**2008**). [Lipid biofuel production with microalgae: potential and challenges]. J. Soc. Biol. *202*, 201-211.

Cadoret, J., Bardor, M., Lerouge, P. et al. (**2008**). [Microalgae as cell factories producing recombinant commercial proteins]. Med Sci (Paris) *24*, 375-382.

Calì, T., Vanoni, O. and Molinari, M. (**2008**). The endoplasmic reticulum crossroads for newly synthesized polypeptide chains. Prog Mol Biol Transl Sci *83*, 135-179.

Caramelo, JJ. and Parodi, AJ. (**2008**). Getting in and out from calnexin/calreticulin cycles. J. Biol. Chem. *283*, 10221-10225.

Castilho, A., Neumann, L., Gattinger, P. et al. (**2013**). Generation of biologically active multi-sialylated recombinant human EPOFc in plants. PLoS ONE *8*, e54836.

Cavalier-Smith, T. (**1982**). The origins of plastids. Biological Journal of the Linnean Society *17*, 289-306.

Cavalier-Smith, T. (**1999**). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J. Eukaryot. Microbiol. *46*, 347-366.

Chan, CX., Bhattacharya, D. and Reyes-Prieto, A. (**2012**). Endosymbiotic and horizontal gene transfer in microbial eukaryotes: Impacts on cell evolution and the tree of life. Mob Genet Elements *2*, 101-105.

Chen, MM., Glover, KJ. and Imperiali, B. (**2007**). From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. Biochemistry *46*, 5579-5585.

Chen, S., Zhou, S., Sarkar, M. et al. (**1999**). Expression of three Caenorhabditis elegans N-acetylglucosaminyltransferase I genes during development. J. Biol. Chem. *274*, 288-297.

Chepurnov, VA. and Mann, DG. (**2004**). Auxosporulation of Licmophora communis (Bacillariophyta) and a review of mating systems and sexual reproduction in araphid pennate diatoms. Phycological Res. *52*, 1-12.

Chevalier, L., Bernard, S., Ramdani, Y. et al. (**2010**). Subcompartment localization of the side chain xyloglucan-synthesizing enzymes within Golgi stacks of tobacco suspension-cultured cells. Plant J. *64*, 977-989.

Chevet, E., Cameron, PH., Pelletier, MF. et al. (**2001**). The endoplasmic reticulum: integration of protein folding, quality control, signaling and degradation. Curr. Opin. Struct. Biol. *11*, 120-124.

Chisti, Y. (**2007**). Biodiesel from microalgae. Biotechnol. Adv. *25*, 294-306.

Choi, YH., Kim, JH., Park, JH. et al. (**2013**). Protein engineering of α2,3/2,6-sialyltransferase to improve the yield and productivity of in vitro sialyllactose synthesis. Glycobiology , .

Cipollo, JF., Trimble, RB., Chi, JH. et al. (**2001**). The yeast ALG11 gene specifies addition of the terminal alpha 1,2-Man to the Man5GlcNAc2-PP-dolichol N-glycosylation intermediate formed on the cytosolic side of the endoplasmic reticulum. J. Biol. Chem. *276*, 21828-21840.

Collos, Y. (**1983**). Transient situations in nitrate assimilation by marine diatoms. 4. Non-linear phenomena and the estimation of the maximum uptake rate. J. Plankton Res. *5(5)*, 677-691.

Corpet, F. (**1988**). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. *16*, 10881-10890.

Couto, JR., Huffaker, TC. and Robbins, PW. (**1984**). Cloning and expression in Escherichia coli of a yeast mannosyltransferase from the asparagine-linked glycosylation pathway. J. Biol. Chem. *259*, 378-382.

Craggs, RJ., McAuley, PJ. and Smith, VJ. (**1997**). Wastewater nutrient removal by marine microalgae grown on a corrugated raceway. Water Res. *31*, 1701 - 1707.

Crispin, M., Harvey, DJ., Chang, VT. 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.

Croll, DA. and Holmes, RW. (**1982**). A Note on the Occurrence of Diatoms on the Feathers of Diving Seabirds. The Auk *99 (4)*, 765-766.

Culyba, EK., Price, JL., Hanson, SR. et al. (**2011**). Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. Science *331*, 571-575.

Czlapinski, JL. and Bertozzi, CR. (**2006**). Synthetic glycobiology: exploits in the Golgi compartment. Current Opinion in Chemical Biology *10*, 645 - 651.

D'Andrea, G., Salucci, ML., Pitari, G. et al. (**1993**). Exhaustive removal of N-glycans from ascorbate oxidase: effect on the enzymatic activity and immunoreactivity. Glycobiology *3*, 563-565.

Davidovich, NA. and Bates, SS. (**1998**). SEXUAL REPRODUCTION IN THE PENNATE DIATOMS PSEUDO-NITZSCHIA MULTISERIES AND P. PSEUDODELICATISSIMA (BACILLARIOPHYCEAE). J. Phycol. *34*, 126-137.

Davidovich, NA., Kaczmarska, I., Karpov, SA. et al. (**2012**). Mechanism of male gamete motility in araphid pennate diatoms from the genus Tabularia (Bacillariophyta). Protist *163*, 480-494.

De Martino, A., Bartual, A., Willis, A. et al. (**2011**). Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom Phaeodactylum tricornutum. Protist *162*, 462-481.

De Martino, A., Meichenin, A., Shi, J. et al. (**2007**). Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions1. J. Phycol. *43*, 992-1009.

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. et al. (**2009**). Gene silencing in the marine diatom Phaeodactylum tricornutum. Nucleic Acids Res. *37*, e96.

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.

Deschamps, P. and Moreira, D. (**2012**). Reevaluating the green contribution to diatom genomes. Genome Biol Evol *4*, 683-688.

Driouich, A., Follet-Gueye, M., Bernard, S. et al. (**2012**). Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. Front Plant Sci *3*, 79.

Drum, RW. and Gordon, R. (**2003**). Star Trek replicators and diatom nanotechnology. Trends Biotechnol. *21*, 325-328.

Dumax-Vorzet, A., Roboti, P. and High, S. (**2013**). OST4 is a subunit of the mammalian oligosaccharyltransferase required for efficient N-glycosylation. J. Cell. Sci. *126*, 2595-2606.

Eklund, EA., Sun, L., Westphal, V. et al. (**2005**). Congenital disorder of glycosylation (CDG)-Ih patient with a severe hepato-intestinal phenotype and evolving central nervous system pathology. J. Pediatr. *147*, 847-850.

Emanuelsson, O., Brunak, S., von Heijne, G. et al. (**2007**). Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc *2*, 953-971.

Fabini, G., Freilinger, A., Altmann, F. et al. (**2001**). Identification of core alpha 1,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from Drosophila melanogaster. Potential basis of the neural anti-horseadish peroxidase epitope. J. Biol. Chem. *276*, 28058-28067.

Falciatore, A. and Bowler, C. (**2002**). Revealing the molecular secrets of marine diatoms. Annu Rev Plant Biol *53*, 109-130.

Falciatore, A., Casotti, R., Leblanc, C. et al. (**1999**). Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar. Biotechnol. *1*, 239-251.

Falkowski, P. and Raven JA (**1997**). . Aquatic photosynthesis. Malden, MA, USA: Blackwell Science , .

Falkowski, P., Barber, R. and Smetacek, V. (**1998**). Biogeochemical Controls and Feedbacks on Ocean Primary Production. Science *281*, 200-207.

Falkowski, PG., Katz, ME., Knoll, AH. et al. (**2004**). The evolution of modern eukaryotic phytoplankton. Science *305*, 354-360.

Fanata, WID., Lee, KH., Son, BH. et al. (**2013**). N-glycan maturation is crucial for cytokinin-mediated development and cellulose synthesis in Oryza sativa. Plant J. *73*, 966-979.

Faye, L., Gomord, V., Fitchette-Lainé, AC. et al. (**1993**). Affinity purification of antibodies specific for Asn-linked glycans containing alpha 1-->3 fucose or beta 1-->2 xylose. Anal. Biochem. *209*, 104-108.

Faye, L., Sturm, A., Bollini, R. et al. (**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.

Feng, J., Romaniouk, AV., Samal, SK. et al. (**2004**). Processing enzyme glucosidase II: proposed catalytic residues and developmental regulation during the ontogeny of the mouse mammary gland. Glycobiology *14*, 909-921.

Field, C., Behrenfeld, M., Randerson, J. et al. (**1998**). Primary production of the biosphere: integrating terrestrial and oceanic components. Science *281*, 237-240.

Finazzi, G., Moreau, H. and Bowler, C. (**2010**). Genomic insights into photosynthesis in eukaryotic phytoplankton. Trends Plant Sci. *15*, 565-572.

Finn, RD., Mistry, J., Tate, J. et al. (**2010**). The Pfam protein families database. Nucleic Acids Res. *38*, D211-22.

Fitchette, A., Dinh, OT., Faye, L. et al. (**2007**). Plant proteomics and glycosylation. Methods Mol. Biol. *355*, 317-342.

Fitchette, AC., Cabanes-Macheteau, M., Marvin, L. et al. (**1999**). Biosynthesis and immunolocalization of Lewis a-containing N-glycans in the plant cell. Plant Physiol. *121*, 333-344.

Fitchette-Lainé, AC., Gomord, V., Cabanes, M. et al. (**1997**). N-glycans harboring the Lewis a epitope are expressed at the surface of plant cells. Plant J. *12*, 1411-1417.

Foster, RA. and Zehr, JP. (**2006**). Characterization of diatom-cyanobacteria symbioses on the basis of nifH, hetR and 16S rRNA sequences. Environ. Microbiol. *8*, 1913-1925.

Frank, CG. and Aebi, M. (**2005**). ALG9 mannosyltransferase is involved in two different steps of lipid-linked oligosaccharide biosynthesis. Glycobiology *15*, 1156-1163.

Frank, K. and Sippl, MJ. (**2008**). High-performance signal peptide prediction based on sequence alignment techniques. Bioinformatics *24*, 2172-2176.

Freeze, HH., Eklund, EA., Ng, BG. et al. (**2012**). Neurology of inherited glycosylation disorders. The Lancet Neurology *11*, 453-466.

Frickel, E., Riek, R., Jelesarov, I. et al. (**2002**). TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. Proc. Natl. Acad. Sci. U.S.A. *99*, 1954-1959.

Fukazawa, Y., Kagaya, K. and Shinoda, T. (**1995**). Cell wall polysaccharides of pathogenic yeasts. Curr Top Med Mycol *6*, 189-219.

Fukuda, T., Hashimoto, H., Okayasu, N. et al. (**2011**). Alpha1,6-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.

Gao, X., Moriyama, S., Miura, N. et al. (**2008**). Interaction between the C termini of Alg13 and Alg14 mediates formation of the active UDP-N-acetylglucosamine transferase complex. J. Biol. Chem. *283*, 32534-32541.

Gao, X., Nishikawa, A. and Dean, N. (**2004**). Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum. Glycobiology *14*, 559-570.

Gemmill, TR. and Trimble, RB. (**1996**). Schizosaccharomyces pombe produces novel pyruvate-containing N-linked oligosaccharides. J. Biol. Chem. *271*, 25945-25949.

Gemmill, TR. and Trimble, RB. (**1999**). Overview of N- and O-linked oligosaccharide structures found in various yeast species. Biochim. Biophys. Acta *1426*, 227-237.

Gibbs, SP. (**1981**). The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. Ann. N. Y. Acad. Sci. *361*, 193-208.

Gidalevitz, T., Stevens, F. and Argon, Y. (**2013**). Orchestration of secretory protein folding by ER chaperones. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1833*, 2410-2424.

Goldman, J. (**1993**). Potential role of large oceanic diatoms in new primary production. Deep-sea Research Part I-oceanographic Research Papers *40*, 159-168.

Gordon, RD., Sivarajah, P., Satkunarajah, M. et al. (**2006**). X-ray crystal structures of rabbit N-acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. J. Mol. Biol. *360*, 67-79.

Gouet, P., Courcelle, E., Stuart, DI. et al. (**1999**). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305-308.

Granum, E., Raven, JA. and Leegood, RC. (**2005**). How do marine diatoms fix 10 billion tonnes of inorganic carbon per year?. Can. J. Bot *83*, 898-908.

Green, BR. (**2011**). After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl c. Photosyn. Res. *107*, 103-115.

Grunewald, S., Matthijs, G. and Jaeken, J. (**2002**). Congenital disorders of glycosylation: a review. Pediatr. Res. *52*, 618-624.

Gueho, E., Pesando, D. and Barelli, M. (**1977**). Proprietes antifongiques d'une diatomee Chaetoceros lauderi ralfs C C. Mycopathologia *60*, 105-107.

Guerin, M. and Parodi, AJ. (**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.

Gupta, S. and Agrawal, SC. (**2007**). Survival and motility of diatoms Navicula grimmei and Nitzschia palea affected by some physical and chemical factors. Folia Microbiol. (Praha) *52*, 127-134.

Gutenbrunner, SA., Thalhamer, J. and Schmid, AM. (**1994**). PROTEINACEOUS AND IMMUNOCHEMICAL DISTINCTIONS BETWEEN THE OVAL AND FUSIFORM MORPHOTYPES OF PHAEODACTYLUM TRICORNUTUM (BACILLARIOPHYCEAE)1. J. Phycol. *30*, 129-136.

Gutternigg, M., Kretschmer-Lubich, D., Paschinger, K. et al. (**2007**). Biosynthesis of truncated N-linked oligosaccharides results from non-orthologous hexosaminidase-mediated mechanisms in nematodes, plants, and insects. J. Biol. Chem. *282*, 27825-27840.

Guzmán, S., Gato, A. and Calleja, JM. (**2001**). Antiinflammatory, analgesic and free radical scavenging activities of the marine microalgae Chlorella stigmatophora and Phaeodactylum tricornutum. Phytother Res *15*, 224-230.

Guzmán, S., Gato, A., Lamela, M. et al. (**2003**). Anti-inflammatory and immunomodulatory activities of polysaccharide from Chlorella stigmatophora and Phaeodactylum tricornutum. Phytother Res *17*, 665-670.

Hackett, JD., Yoon, HS., Li, S. et al. (**2007**). Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. Mol. Biol. Evol. *24*, 1702-1713.

Hahn, M., Borisova, S., Schrag, JD. et al. (**1998**). Identification and Crystallization of a Protease-Resistant Core of Calnexin That Retains Biological Activity. J. Struct. Biol. *123*, 260-264.

Hanson, SR., Culyba, EK., Hsu, T. et al. (**2009**). The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. Proc. Natl. Acad. Sci. U.S.A. *106*, 3131-3136.

Harada, Y., Li, H., Li, H. et al. (**2009**). Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. Proc. Natl. Acad. Sci. U.S.A. *106*, 6945-6949.

Hebert, DN. (**2012**). An MBoC favorite: Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. Mol. Biol. Cell *23*, 2236.

Heesen, S., Lehle, L., Weissmann, A. et al. (**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*, 253-265.

Helenius, A. and Aebi, M. (**2001**). Intracellular functions of N-linked glycans. Science *291*, 2364-2369.

Helenius, A. and Aebi, M. (**2004**). Roles of N-linked glycans in the endoplasmic reticulum. Annu. Rev. Biochem. *73*, 1019-1049.

Helenius, J., Ng, DTW., Marolda, CL. et al. (**2002**). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. Nature *415*, 447-450.

Hempel, F., Bozarth, AS., Lindenkamp, N. et al. (**2011**). Microalgae as bioreactors for bioplastic production. Microb. Cell Fact. *10*, 81.

Hempel, F., Lau, J., Klingl, A. et al. (**2011**). Algae as Protein Factories: Expression of a Human Antibody and the Respective Antigen in the Diatom Phaeodactylum tricornutum. PLoS ONE *6*, e28424.

Hennet, T. (**2012**). Diseases of glycosylation beyond classical congenital disorders of glycosylation. Biochimica et Biophysica Acta (BBA) - General Subjects *1820*, 1306-1317.

Henquet, M., Lehle, L., Schreuder, M. et al. (**2008**). Identification of the gene encoding the alpha1,3-mannosyltransferase (ALG3) in Arabidopsis and characterization of downstream n-glycan processing. Plant Cell *20*, 1652-1664.

Henrissat, B. and Romeu, A. (**1995**). Families, superfamilies and subfamilies of glycosyl hydrolases. Biochem. J. *311 (Pt 1)*, 350-351.

Herlory, O., Bonzom, J., Gilbin, R. et al. (**2013**). Use of diatom assemblages as biomonitor of the impact of treated uranium mining effluent discharge on a stream: case study of the Ritord watershed (Center-West France). Ecotoxicology *22*, 1186-1199.

Herscovics, A. and Orlean, P. (**1993**). Glycoprotein biosynthesis in yeast. FASEB J. *7*, 540-550.

Hese, K., Otto, C., Routier, FH. et al. (**2009**). The yeast oligosaccharyltransferase complex can be replaced by STT3 from Leishmania major. Glycobiology *19*, 160-171.

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.

High, S., Lecomte, FJ., Russell, SJ. et al. (**2000**). Glycoprotein folding in the endoplasmic reticulum: a tale of three chaperones?. FEBS Lett. *476*, 38-41.

Hook, SE. and Osborn, HL. (**2012**). Comparison of toxicity and transcriptomic profiles in a diatom exposed to oil, dispersants, dispersed oil. Aquat. Toxicol. *124-125*, 139-151.

Hoops, HJ. and Floyd, GL. (**1979**). Ultrastructure of the centric diatom, Cyclotella meneghiniana: vegetative cell and auxospore development*. Phycologia *18*, 424-435.

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*, 337.

Hutchins, DA. and Bruland, KW. (**1998**). Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. Nature *393*, 561-564.

Imberty, A., Monier, C., Bettler, E. et al. (**1999**). Fold recognition study of alpha3-galactosyltransferase and molecular modeling of the nucleotide sugar-binding domain. Glycobiology *9*, 713-722.

Imperiali, B. and Hendrickson, TL. (**1995**). Asparagine-linked glycosylation: Specificity and function of oligosaccharyl transferase. Bioorg. Med. Chem. *3*, 1565-1578.

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.

Iwasa, K. and Shimizu, A. (**1972**). Motility of the diatom, Phaeodactylum tricornutum. Exp. Cell Res. *74*, 552-558.

Izquierdo, L., Mehlert, A. and Ferguson, MAJ. (**2012**). The lipid-linked oligosaccharide donor specificities of Trypanosoma brucei oligosaccharyltransferases. Glycobiology *22*, 696-703.

Izquierdo, L., Schulz, BL., Rodrigues, JA. et al. (**2009**). Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. EMBO J. *28*, 2650-2661.

Jamali, AA., Akbari, F., Ghorakhlu, MM. et al. (**2012**). Applications of diatoms as potential microalgae in nanobiotechnology. Bioimpacts *2*, 83-89.

Jeanneau, C., Chazalet, V., Augé, C. et al. (**2004**). Structure-function analysis of the human sialyltransferase ST3Gal I: role of n-glycosylation and a novel conserved sialylmotif. J. Biol. Chem. *279*, 13461-13468.

Jenkins, N. and Curling, EM. (**1994**). Glycosylation of recombinant proteins: problems and prospects. Enzyme Microb. Technol. *16*, 354-364.

Jensen, KG., Moestrup, Ã. and Schmid, AM. (**2003**). Ultrastructure of the male gametes from two centric diatoms, Chaetoceros laciniosus and Coscinodiscus wailesii (Bacillariophyceae). Phycologia *42*, 98-105.

Jez, J., Castilho, A., Grass, J. et al. (**2013**). Expression of functionally active sialylated human erythropoietin in plants. Biotechnol J *8*, 371-382.

Johnson, KD. and Chrispeels, MJ. (**1987**). Substrate Specificities of N-Acetylglucosaminyl-, Fucosyl-, and Xylosyltransferases that Modify Glycoproteins in the Golgi Apparatus of Bean Cotyledons. Plant Physiol. *84*, 1301-1308.

Johnson, KS., Chavez, FP., Elrod, VA. et al. (**2001**). The annual cycle of iron and the biological response in central California coastal waters. Geophysical Research Letters *28*, 1247-1250.

Kaewmuangmoon, J., Kilaso, M., Leartsakulpanich, U. et al. (**2013**). Expression of a secretory α-glucosidase II from Apis cerana indica in Pichia pastoris and its characterization. BMC Biotechnol. *13*, 16.

Kajiura, H., Koiwa, H., Nakazawa, Y. et al. (**2010**). Two Arabidopsis thaliana Golgi alpha-mannosidase I enzymes are responsible for plant N-glycan maturation. Glycobiology *20*, 235-247.

Kang, JS., Frank, J., Kang, CH. et al. (**2008**). Salt tolerance of Arabidopsis thaliana requires maturation of N-glycosylated proteins in the Golgi apparatus. Proc. Natl. Acad. Sci. U.S.A. *105*, 5933-5938.

Kaushal, GP., Szumilo, T., Pastuszak, I. et al. (**1990**). Purification to homogeneity and properties of mannosidase II from mung bean seedlings. Biochemistry *29*, 2168-2176.

Keeling, PJ. (**2009**). Chromalveolates and the evolution of plastids by secondary endosymbiosis. J. Eukaryot. Microbiol. *56*, 1-8.

Keeling, PJ. (**2013**). The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. Annu Rev Plant Biol , .

Keeling, PJ. and Palmer, JD. (**2008**). Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. *9*, 605-618.

Kelleher, DJ. and Gilmore, R. (**2006**). An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology *16*, 47R-62R.

Kelleher, DJ., Karaoglu, D., Mandon, EC. et al. (**2003**). Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. Mol. Cell *12*, 101-111.

Kemp, AES., Pearce, RB., Grigorov, I. et al. (**2006**). Production of giant marine diatoms and their export at oceanic frontal zones: Implications for Si and C flux from stratified oceans. Global Biogeochemical Cycles *20*, n/a-n/a.

Kim, SM., Jung, Y., Kwon, O. et al. (**2012**). A potential commercial source of fucoxanthin extracted from the microalga Phaeodactylum tricornutum. Appl. Biochem. Biotechnol. *166*, 1843-1855.

KljajiÄ‡, P., AndriÄ‡, G., AdamoviÄ‡, M. et al. (**2010**). Laboratory assessment of insecticidal effectiveness of natural zeolite and diatomaceous earth formulations against three stored-product beetle pests. Journal of Stored Products Research *46*, 1 - 6.

Kooistra, WH. and Medlin, LK. (**1996**). Evolution of the Diatoms (Bacillariophyta): IV. A Reconstruction of Their Age from Small Subunit rRNA Coding Regions and the Fossil Record. Mol. Phylogenet. Evol. *6*, 391-407.

Kornfeld, R. and Kornfeld, S. (**1985**). Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. *54*, 631-664.

KoruniÄ†, Z. (**1997**). Rapid assessment of the insecticidal value of diatomaceous earths without conducting bioassays. Journal of Stored Products Research *33*, 219 - 229.

Kowarik, M., Küng, S., Martoglio, B. et al. (**2002**). Protein folding during cotranslational translocation in the endoplasmic reticulum. Mol. Cell *10*, 769-778.

Kozlov, G., Maattanen, P., Schrag, JD. et al. (**2006**). Crystal structure of the bb' domains of the protein disulfide isomerase ERp57. Structure *14*, 1331-1339.

Kroger, N. and Poulsen, N. (**2008**). Diatoms- ”From Cell Wall Biogenesis to Nanotechnology. Annu. Rev. Genet *42*, 83-107.

Kroth, P. (**2007**). Molecular biology and the biotechnological potential of diatoms. Adv. Exp. Med. Biol. *616*, 23-33.

Kubelka, V., Altmann, F. and März, L. (**1995**). The asparagine-linked carbohydrate of honeybee venom hyaluronidase. Glycoconj. J. *12*, 77-83.

Kubelka, V., Altmann, F., Kornfeld, G. et al. (**1994**). Structures of the N-linked oligosaccharides of the membrane glycoproteins from three lepidopteran cell lines (Sf-21, IZD-Mb-0503, Bm-N). Arch. Biochem. Biophys. *308*, 148-157.

Kubelka, V., Altmann, F., Staudacher, E. et al. (**1993**). Primary structures of the N-linked carbohydrate chains from honeybee venom phospholipase A2. Eur. J. Biochem. *213*, 1193-1204.

Kämpf, M., Absmanner, B., Schwarz, M. et al. (**2009**). Biochemical characterization and membrane topology of Alg2 from Saccharomyces cerevisiae as a bifunctional alpha1,3- and 1,6-mannosyltransferase involved in lipid-linked oligosaccharide biosynthesis. J. Biol. Chem. *284*, 11900-11912.

LEWIN, JC., LEWIN, RA. and PHILPOTT, DE. (**1958**). Observations on Phaeodactylum tricornutum. J. Gen. Microbiol. *18*, 418-426.

Laemmli, UK. (**1970**). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Laguna, MR., Villar, RM., Cadavid, I. et al. (**1990**). Effects of Phaeodactylum tricornutum and Dunaliella tertiolecta extracts on the central nervous system. Planta Med. *56*, 152-157.

Lairson, LL., Henrissat, B., Davies, GJ. et al. (**2008**). Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. *77*, 521-555.

Lakatos, M., Lange-Bertalot, H. and BÃ¼del, B. (**2004**). DIATOMS LIVING INSIDE THE THALLUS OF THE GREEN ALGAL LICHEN COENOGONIUM LINKII IN NEOTROPICAL LOWLAND RAIN FORESTS1. J. Phycol. *40*, 70-73.

Lal, A., Schutzbach, JS., Forsee, WT. et al. (**1994**). Isolation and expression of murine and rabbit cDNAs encoding an alpha 1,2-mannosidase involved in the processing of asparagine-linked oligosaccharides. J. Biol. Chem. *269*, 9872-9881.

Lange, KP., Bellamy, WD. and Hendricks, DW. (**1986**). Diatomaceous earth filtration of Giardia cysts and other substances.. Journal of the American Water Works Association *78(1)*, 76-84.

Lebeau, T. and Robert, JM. (**2003**). Diatom cultivation and biotechnologically relevant products. Part II: Current and putative products. Appl. Microbiol. Biotechnol. *60*, 624-632.

Lederkremer, GZ. (**2009**). Glycoprotein folding, quality control and ER-associated degradation. Curr. Opin. Struct. Biol. *19*, 515-523.

Lee, JJ., Morales, J., Symons, A. et al. (**1995**). Diatom symbionts in larger foraminifera from Caribbean hosts. Mar. Micropaleontol. *26*, 99-105.

Leiter, H., Mucha, J., Staudacher, E. et al. (**1999**). Purification, cDNA cloning, and expression of GDP-L-Fuc:Asn-linked GlcNAc alpha1,3-fucosyltransferase from mung beans. J. Biol. Chem. *274*, 21830-21839.

Lerouge, P., Cabanes-Macheteau, M., Rayon, C. et al. (**1998**). N-glycoprotein biosynthesis in plants: recent developments and future trends. Plant Mol. Biol. *38*, 31-48.

Levy-Ontman, O., Arad, SM., Harvey, DJ. et al. (**2011**). Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga Porphyridium sp. J. Biol. Chem. *286*, 21340-21352.

Li, S., Nosenko, T., Hackett, JD. et al. (**2006**). Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol. Biol. Evol. *23*, 663-674.

Li, W., Gao, K. and Beardall, J. (**2012**). Interactive Effects of Ocean Acidification and Nitrogen-Limitation on the Diatom Phaeodactylum tricornutum. PLoS ONE *7*, e51590.

Liebminger, E., Hüttner, S., Vavra, U. et al. (**2009**). Class I alpha-mannosidases are required for N-glycan processing and root development in Arabidopsis thaliana. Plant Cell *21*, 3850-3867.

Lin, AI., Philipsberg, GA. and Haltiwanger, RS. (**1994**). Core fucosylation of high-mannose-type oligosaccharides in GlcNAc transferase I-deficient (Lec1) CHO cells. Glycobiology *4*, 895-901.

Lin, X., Tirichine, L. and Bowler, C. (**2012**). Protocol: Chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species. Plant Methods *8*, 48.

Lincoln, RA., Strupinski, K. and Walker, JM. (**1990**). BIOLOGICALLY ACTIVE COMPOUNDS FROM DIATOMS. Diatom Research *5*, 337-349.

Lingg, N., Zhang, P., Song, Z. et al. (**2012**). The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. Biotechnol J *7*, 1462-1472.

Lizak, C., Gerber, S., Numao, S. et al. (**2011**). X-ray structure of a bacterial oligosaccharyltransferase. Nature *474*, 350-355.

Lobsanov, YD., Vallée, F., Imberty, A. et al. (**2002**). Structure of Penicillium citrinum alpha 1,2-mannosidase reveals the basis for differences in specificity of the endoplasmic reticulum and Golgi class I enzymes. J. Biol. Chem. *277*, 5620-5630.

Lopez, PJ., Desclés, J., Allen, AE. et al. (**2005**). Prospects in diatom research. Curr. Opin. Biotechnol. *16*, 180-186.

Losic, D., Rosengarten, G., Mitchell, JG. et al. (**2006**). Pore architecture of diatom frustules: potential nanostructured membranes for molecular and particle separations. J Nanosci Nanotechnol *6*, 982-989.

Lu, J., Takahashi, T., Ohoka, A. et al. (**2012**). Alg14 organizes the formation of a multiglycosyltransferase complex involved in initiation of lipid-linked oligosaccharide biosynthesis. Glycobiology *22*, 504-516.

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.

Maheswari, U., Jabbari, K., Petit, J. et al. (**2010**). Digital expression profiling of novel diatom transcripts provides insight into their biological functions. Genome Biol. *11*, R85.

Maheswari, U., Mock, T., Armbrust, EV. et al. (**2009**). Update of the Diatom EST Database: a new tool for digital transcriptomics. Nucleic Acids Res. *37*, D1001-5.

Mann, D. (**1993**). Patterns of sexual reproduction in diatoms.. Hydrobiologia *269/270*, 11-20.

Mann, DG. (**1999**). The species concept in diatoms. Phycologia *38*, 437-495.

Manton, I., Kowallik, K. and von Stosch, HA. (**1970**). Observations on the fine structure and development of the spindle at mitosis and meiosis in a marine centric diatom (Lithodesmium undulatum). IV. The second meiotic division and conclusion. J. Cell. Sci. *7*, 407-443.

Marchetti, A., Parker, MS., Moccia, LP. et al. (**2009**). Ferritin is used for iron storage in bloom-forming marine pennate diatoms. Nature *457*, 467-470.

Margulis, L. (**1992**). Biodiversity: molecular biological domains, symbiosis and kingdom origins. BioSystems *27*, 39 - 51.

Marshall, RD. (**1972**). Glycoproteins. Annu. Rev. Biochem. *41*, 673-702.

Mast, SW. and Moremen, KW. (**2006**). Family 47 alpha-mannosidases in N-glycan processing. Meth. Enzymol. *415*, 31-46.

Mathieu-Rivet, E., Scholz, M., Arias, C. et al. (**2013**). Exploring the N-glycosylation Pathway in Chlamydomonas reinhardtii Unravels Novel Complex Structures. Mol. Cell Proteomics *12*, 3160-3183.

Maumus, F., Allen, AE., Mhiri, C. et al. (**2009**). Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. BMC Genomics *10*, 624.

McFadden, GI. (**1999**). Plastids and protein targeting. J. Eukaryot. Microbiol. *46*, 339-346.

McFadden, GI. (**2001**). PRIMARY AND SECONDARY ENDOSYMBIOSIS AND THE ORIGIN OF PLASTIDS. J. Phycol. *37*, 951-959.

McGinnis, K., Dempster, T. and Sommerfeld, M. (**1997**). Characterization of the growth and lipid content of the diatom Chaetoceros muelleri. Journal of Applied Phycology *9*, 19-24.

McKay, RML., Geider, RJ. and LaRoche, J. (**1997**). Physiological and Biochemical Response of the Photosynthetic Apparatus of Two Marine Diatoms to Fe Stress. Plant Physiol. *114*, 615-622.

Medlin, K., Kooistra, W., Gersonde, R. et al. (**1997**). Is the origin of the diatoms related to the end-permian mass extinction ?. Nova Hedwigia *65 (1-4)*, 1-11.

Medlin, LK. (**2011**). The Permianâ "Triassic mass extinction forces the radiation of the modern marine phytoplankton. Phycologia *50*, 684-693.

Meiser, A., Schmid-Staiger, U. and TrÃ¶sch, W. (**2004**). Optimization of eicosapentaenoic acid production byPhaeodactylum tricornutumin the flat panel airlift (FPA) reactor. Journal of Applied Phycology *16*, 215-225.

Melo, NS., Nimtz, M., Conradt, HS. et al. (**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. et al. (**1994**). Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. EMBO J. *13*, 2056-2065.

Miletich, JP. and Broze, GJJ. (**1990**). Beta protein C is not glycosylated at asparagine 329. The rate of translation may influence the frequency of usage at asparagine-X-cysteine sites. J. Biol. Chem. *265*, 11397-11404.

Mimouni, V., Ulmann, L., Pasquet, V. et al. (**2012**). The potential of microalgae for the production of bioactive molecules of pharmaceutical interest. Curr Pharm Biotechnol *13*, 2733-2750.

Miyahara, M., Aoi, M., Inoue-Kashino, N. et al. (**2013**). Highly Efficient Transformation of the Diatom Phaeodactylum tricornutum by Multi-Pulse Electroporation. Biosci. Biotechnol. Biochem. *77*, 874-876.

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, KK., Calanca, V. et al. (**2004**). Contrasting Functions of Calreticulin and Calnexin in Glycoprotein Folding and ER Quality Control. Mol. Cell *13*, 125-135.

Molinari, M., Galli, C., Vanoni, O. et al. (**2005**). Persistent Glycoprotein Misfolding Activates the Glucosidase II/UGT1-Driven Calnexin Cycle to Delay Aggregation and Loss of Folding Competence. Mol. Cell *20*, 503-512.

Montsant, A., Jabbari, K., Maheswari, U. et al. (**2005 a**). Comparative genomics of the pennate diatom Phaeodactylum tricornutum. Plant Physiol. *137*, 500-513.

Montsant, A., Maheswari, U., Bowler, C. et al. (**2005 b**). Diatomics: toward diatom functional genomics. J Nanosci Nanotechnol *5*, 5-14.

Moore, J., Doney, S., Glover, D. et al. (**2002**). Iron cycling and nutrient limitation patterns in surface waters of the world ocean. Deep-Sea Res. II *49*, 463-508.

Moreau, D., Tomasoni, C., Jacquot, C. et al. (**2006**). Cultivated microalgae and the carotenoid fucoxanthin from Odontella aurita as potent anti-proliferative agents in bronchopulmonary and epithelial cell lines. Environ. Toxicol. Pharmacol. *22*, 97-103.

Morelli, E. and Pratesi, E. (**1997**). Production of Phytochelatins in the Marine Diatom Phaeodactylum tricornutum in Response to Copper and Cadmium Exposure. Bulletin of Environmental Contamination and Toxicology *59*, 657-664.

Moustafa, A., Beszteri, B., Maier, UG. et al. (**2009**). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science *324*, 1724-1726.

Mullis, K., Faloona, F., Scharf, S. et al. (**1986**). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. *51 Pt 1*, 263-273.

Munro, S. (**2001**). The MRH domain suggests a shared ancestry for the mannose 6-phosphate receptors and other N-glycan-recognising proteins. Curr. Biol. *11*, R499-501.

MÃ¤Ã¤ttÃ¤nen, P., Gehring, K., Bergeron, JJ. et al. (**2010**). Protein quality control in the ER: The recognition of misfolded proteins. Semin. Cell Dev. Biol. *21*, 500-511.

Määttänen, P., Gehring, K., Bergeron, JJ. et al. (**2010**). Protein quality control in the ER: The recognition of misfolded proteins. Semin. Cell Dev. Biol. *21*, 500-511.

Nakayama, K., Nagasu, T., Shimma, Y. et al. (**1992**). OCH1 encodes a novel membrane bound mannosyltransferase: outer chain elongation of asparagine-linked oligosaccharides. EMBO J. *11*, 2511-2519.

Nasab, FP., Schulz, BL., Gamarro, F. et al. (**2008**). All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. Mol. Biol. Cell *19*, 3758-3768.

Nelson, DM., TrÃ©guer, P., Brzezinski, MA. et al. (**1995**). Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Global Biogeochemical Cycles *9*, 359-372.

Nigh, R., Benbrook, C., Brush, S. et al. (**2000**). Transgenic crops: a cautionary tale. Science *287*, 1927.

Niu, Y., Yang, Z., Zhang, M. et al. (**2012**). Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. BioTechniques *0*, 1-3.

Noffz, C., Keppler-Ross, S. and Dean, N. (**2009**). Hetero-oligomeric interactions between early glycosyltransferases of the dolichol cycle. Glycobiology *19*, 472-478.

Norden-Krichmar, TM., Allen, AE., Gaasterland, T. et al. (**2011**). Characterization of the small RNA transcriptome of the diatom, Thalassiosira pseudonana. PLoS ONE *6*, e22870.

O'Reilly, MK., 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 *45*, 9593-9603.

Okamoto, N., Chantangsi, C., Horák, A. et al. (**2009**). Molecular phylogeny and description of the novel katablepharid Roombia truncata gen. et sp. nov., and establishment of the Hacrobia taxon nov. PLoS ONE *4*, e7080.

Olivari, S., Cali, T., Salo, KE. et al. (**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.

Olsen, DT., Peng, L., Træholt, SD. et al. (**2013**). Purification and characterization of a soluble calnexin from human placenta. Protein Expr. Purif. *92*, 105-111.

Olson, LJ., Orsi, R., Alculumbre, SG. et al. (**2013**). Structure of the lectin mannose 6-phosphate receptor homology (MRH) domain of glucosidase II, an enzyme that regulates glycoprotein folding quality control in the endoplasmic reticulum. J. Biol. Chem. *288*, 16460-16475.

Orlean, P., Albright, C. and Robbins, PW. (**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. et al. (**2003**). Structural requirements for Arabidopsis beta1,2-xylosyltransferase activity and targeting to the Golgi. Plant J. *33*, 189-203.

Paquet, M., Leach, MR. and Williams, DB. (**2005**). In vitro and in vivo assays to assess the functions of calnexin and calreticulin in ER protein folding and quality control. Methods *35*, 338-347.

Parkinson, J. and Gordon, R. (**1999**). Beyond micromachining: the potential of diatoms. Trends Biotechnol. *17*, 190 - 196.

Parodi, AJ. (**2000**). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. Biochem. J. *348 Pt 1*, 1-13.

Parsaie Nasab, F., Aebi, M., Bernhard, G. et al. (**2013**). A combined system for engineering glycosylation efficiency and glycan structure in Saccharomyces cerevisiae. Appl. Environ. Microbiol. *79*, 997-1007.

Paschinger, K., Staudacher, E., Stemmer, U. et al. (**2005**). Fucosyltransferase substrate specificity and the order of fucosylation in invertebrates. Glycobiology *15*, 463-474.

Pasquet, V., ChÃ©rouvrier, J., Farhat, F. et al. (**2011**). Study on the microalgal pigments extraction process: Performance of microwave assisted extraction. Process Biochemistry *46*, 59-67.

Patron, NJ., Inagaki, Y. and Keeling, PJ. (**2007**). Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr. Biol. *17*, 887-891.

Peng, J., Yuan, J., Wu, C. et al. (**2011**). Fucoxanthin, a marine carotenoid present in brown seaweeds and diatoms: metabolism and bioactivities relevant to human health. Mar Drugs *9*, 1806-1828.

Pesando, D., Gnassia-Barelli, M., Gueho, E. et al. (**1979**). Isolement, étude structurale et propriétés antibiotiques et antifongiques d'un composant polysaccharidique de la diatomée marine Cheatoceros Lauderi (Ralfs CC). Océanis , 561-568.

Pfaffen, S., Abdulqadir, R., Le Brun, NE. et al. (**2013**). Mechanism of Ferrous Iron Binding and Oxidation by Ferritin from a Pennate Diatom. J. Biol. Chem. , .

Pieren, M., Galli, C., Denzel, A. et al. (**2005**). The use of calnexin and calreticulin by cellular and viral glycoproteins. J. Biol. Chem. *280*, 28265-28271.

Pistocchi, R., Mormile, M., Guerrini, F. et al. (**2000**). Increased production of extra- and intracellular metal-ligands in phytoplankton exposed to copper and cadmium. Journal of Applied Phycology *12*, 469-477.

Plummer, THJ., Elder, JH., Alexander, S. et al. (**1984**). Demonstration of peptide:N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations. J. Biol. Chem. *259*, 10700-10704.

Pollanen, MS. (**1998**). Diatoms and homicide. Forensic Sci. Int. *91*, 29-34.

Raven, JA. (**2013**). Iron acquisition and allocation in stramenopile algae. J. Exp. Bot. , .

Rayon, C., Lerouge, P. and Faye, L. (**1998**). The protein N-glycosylation in plants . J. Exp. Bot. *49(326)*, 1463-1472.

Reichardt, I., Stierhof, Y., Mayer, U. et al. (**2007**). Plant cytokinesis requires de novo secretory trafficking but not endocytosis. Curr. Biol. *17*, 2047-2053.

Reinfelder, JR., Kraepiel, AM. and Morel, FM. (**2000**). Unicellular C4 photosynthesis in a marine diatom. Nature *407*, 996-999.

Reiss, G., te Heesen, S., Zimmerman, J. et al. (**1996**). Isolation of the ALG6 locus of Saccharomyces cerevisiae required for glucosylation in the N-linked glycosylation pathway. Glycobiology *6*, 493-498.

Rendić, D., Klaudiny, J., Stemmer, U. et al. (**2007**). Towards abolition of immunogenic structures in insect cells: characterization of a honey-bee (Apis mellifera) multi-gene family reveals both an allergy-related core alpha1,3-fucosyltransferase and the first insect Lewis-histo-blood-group-related antigen-synthesizing enzyme. Biochem. J. *402*, 105-115.

Ripka, J., Pierce, M. and Fregien, N. (**1990**). Co-transformation of Lec 1 CHO cells with N-acetylglucosaminyltransferase 1 activity and a selectable marker. J. Cell. Biochem. *42*, 117-122.

Robbins, PW., Trimble, RB., Wirth, DF. et al. (**1984**). Primary structure of the Streptomyces enzyme endo-beta-N-acetylglucosaminidase H. J. Biol. Chem. *259*, 7577-7583.

Round, F., Crawford, R. and Mann, D. (**1990**). The Diatoms. Biology & morphology of the genera.. Cambridge Univ.Press Ed. , .

Rowland, S., Belt, S., Wraige, E. et al. (**2001**). Effects of temperature on polyunsaturation in cytostatic lipids of Haslea ostrearia. Phytochemistry *56*, 597 - 602.

Ruiz-Canada, C., Kelleher, DJ. and Gilmore, R. (**2009**). Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. Cell *136*, 272-283.

Rush, JS. and Waechter, CJ. (**1995**). Transmembrane movement of a water-soluble analogue of mannosylphosphoryldolichol is mediated by an endoplasmic reticulum protein. J. Cell Biol. *130*, 529-536.

Rush, JS., van Leyen, K., Ouerfelli, O. et al. (**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.

Saint-Jore-Dupas, C., Nebenführ, A., Boulaflous, A. et al. (**2006**). Plant N-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. Plant Cell *18*, 3182-3200.

Sanyal, S. and Menon, AK. (**2010**). Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase. Proc. Natl. Acad. Sci. U.S.A. *107*, 11289-11294.

Sapriel, G., Quinet, M., Heijde, M. et al. (**2009**). Genome-wide transcriptome analyses of silicon metabolism in Phaeodactylum tricornutum reveal the multilevel regulation of silicic acid transporters. PLoS ONE *4*, e7458.

Sarkar, M. and Schachter, H. (**2001**). Cloning and expression of Drosophila melanogaster UDP-GlcNAc:alpha-3-D-mannoside beta1,2-N-acetylglucosaminyltransferase I. Biol. Chem. *382*, 209-217.

Sarkar, M., Leventis, PA., Silvescu, CI. et al. (**2006**). Null mutations in Drosophila N-acetylglucosaminyltransferase I produce defects in locomotion and a reduced life span. J. Biol. Chem. *281*, 12776-12785.

Scala, S., Carels, N., Falciatore, A. et al. (**2002**). Genome properties of the diatom Phaeodactylum tricornutum. Plant Physiol. *129*, 993-1002.

Schachter, H. (**2009**). Paucimannose N-glycans in Caenorhabditis elegans and Drosophila melanogaster. Carbohydr. Res. *344*, 1391-1396.

Scherer, R. (**2002**). The diatoms: applications for the environmental and earth sciences E. F. Stoermer and J. P. Smol Cambridge University Press, Cambridge 1999 (469 pp) Â£70/US$115 ISBN 0 521 58281 4. Journal of Quaternary Science *17*, 283-284.

Schneider, JD., Castilho, A., Neumann, L. et al. (**2013**). Expression of human butyrylcholinesterase with an engineered glycosylation profile resembling the plasma-derived orthologue. Biotechnol J , .

Schneider, SC., Kahlert, M. and Kelly, MG. (**2013**). Interactions between pH and nutrients on benthic algae in streams and consequences for ecological status assessment and species richness patterns. Sci. Total Environ. *444*, 73-84.

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., Liebminger, E., Botchway, SW. et al. (**2013**). Time-resolved fluorescence imaging reveals differential interactions of N-glycan processing enzymes across the Golgi stack in planta. Plant Physiol. *161*, 1737-1754.

Schoberer, J., Vavra, U., Stadlmann, J. et al. (**2009**). Arginine/lysine residues in the cytoplasmic tail promote ER export of plant glycosylation enzymes. Traffic *10*, 101-115.

Schollen, E., Grünewald, S., Keldermans, L. et al. (**2005**). CDG-Id caused by homozygosity for an ALG3 mutation due to segmental maternal isodisomy UPD3(q21.3-qter). Eur J Med Genet *48*, 153-158.

Schrag, JD., Bergeron, JJ., Li, Y. et al. (**2001**). The Structure of Calnexin, an ER Chaperone Involved in Quality Control of Protein Folding. Mol. Cell *8*, 633-644.

Schuler, PF., Ghosh, MM. and Gopalan, P. (**1991**). Slow sand and diatomaceous earth filtration of cysts and other particulates. Water Res. *25*, 995 - 1005.

Schulz, BL., Stirnimann, CU., Grimshaw, JPA. et al. (**2009**). Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. Proc. Natl. Acad. Sci. U.S.A. *106*, 11061-11066.

Shailubhai, K., Pukazhenthi, BS., Saxena, ES. et al. (**1991**). Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. J. Biol. Chem. *266*, 16587-16593.

Shental-Bechor, D. and Levy, Y. (**2008**). Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. Proc. Natl. Acad. Sci. U.S.A. *105*, 8256-8261.

Shibatani, T., David, LL., McCormack, AL. et al. (**2005**). Proteomic analysis of mammalian oligosaccharyltransferase reveals multiple subcomplexes that contain Sec61, TRAP, and two potential new subunits. Biochemistry *44*, 5982-5992.

Shrimal, S., Ng, BG., Losfeld, M. et al. (**2013**). Mutations in STT3A and STT3B cause two congenital disorders of glycosylation. Hum. Mol. Genet. , .

Siaut, M., Heijde, M., Mangogna, M. et al. (**2007**). Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. Gene *406*, 23-35.

Sims, PA., Mann, DG. and Medlin, LK. (**2006**). Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia *45*, 361-402.

Smetacek, V. (**1999**). Diatoms and the Ocean Carbon Cycle. Protist *150*, 25-32.

Smetacek, V. (**2001**). A watery arms race. Nature *411*, 745-745.

Smetacek, V., Klaas, C., Strass, VH. et al. (**2012**). Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. Nature *487*, 313-319.

Smol, JP. and Stoermer, EF. (**2010**). The diatoms: applications for the environmental and earth sciences. Seconde édition.. , 1-686.

Snider, MD. and Rogers, OC. (**1984**). Transmembrane movement of oligosaccharide-lipids during glycoprotein synthesis. Cell *36*, 753-761.

Song, M., Pei, H., Hu, W. et al. (**2013**). Evaluation of the potential of 10 microalgal strains for biodiesel production. Bioresour. Technol. , .

Song, W., Mentink, RA., Henquet, MGL. et al. (**2013**). N-glycan occupancy of Arabidopsis N-glycoproteins. J Proteomics , .

Sorhannus, U. (**2007**). A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. Mar. Micropaleontol. *65*, 1-12.

Spirig, U., Bodmer, D., Wacker, M. et al. (**2005**). The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. Glycobiology *15*, 1396-1406.

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. U.S.A. *91*, 5977-5981.

Stanley, P., Narasimhan, S., Siminovitch, L. et al. (**1975**). Chinese hamster ovary cells selected for resistance to the cytotoxicity of phytohemagglutinin are deficient in a UDP-N-acetylglucosamine--glycoprotein N-acetylglucosaminyltransferase activity. Proc. Natl. Acad. Sci. U.S.A. *72*, 3323-3327.

Staudacher, E. and März, L. (**1998**). Strict order of (Fuc to Asn-linked GlcNAc) fucosyltransferases forming core-difucosylated structures. Glycoconj. J. *15*, 355-360.

Staudacher, E., Dalik, T., Wawra, P. et al. (**1995**). Functional purification and characterization of a GDP-fucose: beta-N-acetylglucosamine (Fuc to Asn linked GlcNAc) alpha 1,3-fucosyltransferase from mung beans. Glycoconj. J. *12*, 780-786.

Stewart, WD., Rowell, P. and Rai, AN. (**1983**). Cyanobacteria-eukaryotic plant symbioses. Ann. Microbiol. (Paris) *134B*, 205-228.

Strasser, R., Altmann, F., Mach, L. et al. (**2004**). Generation of Arabidopsis thaliana plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. FEBS Lett. *561*, 132-136.

Strasser, R., Bondili, JS., Vavra, U. et al. (**2007**). A unique beta1,3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis a structures in Arabidopsis thaliana. Plant Cell *19*, 2278-2292.

Strasser, R., Mucha, J., Mach, L. et al. (**2000**). Molecular cloning and functional expression of beta1, 2-xylosyltransferase cDNA from Arabidopsis thaliana. FEBS Lett. *472*, 105-108.

Strasser, R., Mucha, J., Schwihla, H. et al. (**1999**). Molecular cloning and characterization of cDNA coding for beta1, 2N-acetylglucosaminyltransferase I (GlcNAc-TI) from Nicotiana tabacum. Glycobiology *9*, 779-785.

Strasser, R., Schoberer, J., Jin, C. et al. (**2006**). Molecular cloning and characterization of Arabidopsis thaliana Golgi alpha-mannosidase II, a key enzyme in the formation of complex N-glycans in plants. Plant J. *45*, 789-803.

Strasser, R., Stadlmann, J., Svoboda, B. et al. (**2005**). Molecular basis of N-acetylglucosaminyltransferase I deficiency in Arabidopsis thaliana plants lacking complex N-glycans. Biochem. J. *387*, 385-391.

Sun, L., Eklund, EA., Chung, WK. et al. (**2005**). Congenital disorder of glycosylation id presenting with hyperinsulinemic hypoglycemia and islet cell hyperplasia. J. Clin. Endocrinol. Metab. *90*, 4371-4375.

Sutak, R., Botebol, H., Blaiseau, P. et al. (**2012**). A comparative study of iron uptake mechanisms in marine microalgae: iron binding at the cell surface is a critical step. Plant Physiol. *160*, 2271-2284.

Séveno, M., Cabrera, G., Triguero, A. et al. (**2008**). Plant N-glycan profiling of minute amounts of material. Anal. Biochem. *379*, 66-72.

Tai, VW. and Imperiali, B. (**2001**). Substrate specificity of the glycosyl donor for oligosaccharyl transferase. J. Org. Chem. *66*, 6217-6228.

Takahashi, M., Kuroki, Y., Ohtsubo, K. et al. (**2009**). Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. Carbohydr. Res. *344*, 1387-1390.

Takahashi, N., Toda, H., Nishibe, H. et al. (**1982**). Isolation and characterization of Taka-amylase A apoprotein deglycosylated by digestion with almond glycopeptidase immobilized on Sepharose. Biochim. Biophys. Acta *707*, 236-242.

Takemae, H., Ueda, R., Okubo, R. et al. (**2003**). Proteoglycan UDP-galactose:beta-xylose beta 1,4-galactosyltransferase I is essential for viability in Drosophila melanogaster. J. Biol. Chem. *278*, 15571-15578.

Tanner, W. and Lehle, L. (**1987**). Protein glycosylation in yeast. Biochim. Biophys. Acta *906*, 81-99.

Tarentino, AL. and Maley, F. (**1974**). Purification and properties of an endo-beta-N-acetylglucosaminidase from Streptomyces griseus. J. Biol. Chem. *249*, 811-817.

Tarentino, AL., Gómez, CM. and Plummer, THJ. (**1985**). Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. Biochemistry *24*, 4665-4671.

Tirichine, L. and Bowler, C. (**2011**). Decoding algal genomes: tracing back the history of photosynthetic life on Earth. Plant J. *66*, 45-57.

Tolun, LG., Okay, OS., Gaines, AF. et al. (**2001**). The pollution status and the toxicity of surface sediments in Izmit Bay (Marmara Sea), Turkey. Environ Int *26*, 163-168.

Towbin, H., Staehelin, T. and Gordon, J. (**1979**). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. U.S.A. *76*, 4350-4354.

Tretter, V., Altmann, F. and März, L. (**1991**). Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. Eur. J. Biochem. *199*, 647-652.

Trombetta, ES. and Helenius, A. (**1998**). Lectins as chaperones in glycoprotein folding. Curr. Opin. Struct. Biol. *8*, 587-592.

Trombetta, ES., Fleming, KG. and Helenius, A. (**2001**). Quaternary and domain structure of glycoprotein processing glucosidase II. Biochemistry *40*, 10717-10722.

Trombetta, ES., Simons, JF. 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.

Tréguer, P., Nelson, DM., Van Bennekom, AJ. et al. (**1995**). The silica balance in the world ocean: a reestimate. Science *268*, 375-379.

Umemura, K., Haneda, T., Tanabe, M. et al. (**2013**). Semi-circular microgrooves to observe active movements of individual Navicula pavillardii cells. J. Microbiol. Methods *92*, 349-354.

Unligil, UM., Zhou, S., Yuwaraj, S. et al. (**2000**). X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. EMBO J. *19*, 5269-5280.

Vadaie, N., Hulinsky, RS. and Jarvis, DL. (**2002**). Identification and characterization of a Drosophila melanogaster ortholog of human beta1,4-galactosyltransferase VII. Glycobiology *12*, 589-597.

Varki, A. (**1993**). Biological roles of oligosaccharides: all of the theories are correct. Glycobiology *3*, 97-130.

Vella, GJ., Paulsen, H. and Schachter, H. (**1984**). Control of glycoprotein synthesis. IX. A terminal Man alpha l-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. *62*, 409-417.

Veluchamy, A., Lin, X., Maumus, F. et al. (**2013**). Insights into the role of DNA methylation in diatoms by genome-wide profiling in Phaeodactylum tricornutum. Nat Commun *4*, 2091.

Villar, R., Laguna, MR., Cadavid, I. et al. (**1994**). Effects of aqueous extracts of six marine microalgae on smooth muscle contraction in rat duodenum and vas deferens. Planta Med. *60*, 521-526.

Villar, R., Laguna, MR., Calleja, JM. et al. (**1992**). Effects of Phaeodactylum tricornutum and Dunaliella tertiolecta extracts on the central nervous system. Planta Med. *58*, 405-409.

Villareal, TA. (**1989**). Division cycles in the nitrogen-fixing Rhizosolenia (Bacillariophyceae)-Richelia (Nostocaceae) symbiosis. British Phycological Journal *24*, 357-365.

Villareal, TA., Pilskaln, C., Brzezinski, M. et al. (**1999**). Upward transport of oceanic nitrate by migrating diatom mats. Nature *397*, 423-425.

Vitale, A. and Chrispeels, MJ. (**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.

Vizinet, J. and De reviers, B. (**1995**). Les utilisations des diatomées.. Vie et milieu *45*, 165.

Voynow, JA., Kaiser, RS., Scanlin, TF. et al. (**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.

Wada, I., Imai, S., Kai, M. et al. (**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, J., Cao, S., Du, C. et al. (**2013**). Underwater locomotion strategy by a benthic pennate diatom Navicula sp. Protoplasma , .

Waser, NA., Yin, K., Yu, Z. et al. (**1998**). Nitrogen isotope fractionation during nitrate, ammonium and urea uptake by marine diatoms and coccolithophores under various conditions of N availability. Mar. Ecol. Prog. Ser. *169*, 29-41.

Waser, NAD., Harrison, PJ., Nielsen, B. et al. (**1998**). Nitrogen isotope fractionation during the uptake and assimilation of nitrate, nitrite, ammonium, and urea by a marine diatom. Limnology & Oceanography *43(2)*, 215-224.

Watanabe, S., Kakudo, A., Ohta, M. et al. (**2013**). Molecular cloning and characterization of the α-glucosidase II from Bombyx mori and Spodoptera frugiperda. Insect Biochem. Mol. Biol. *43*, 319-327.

Watanabe, T., Totani, K., Matsuo, I. et al. (**2009**). Genetic analysis of glucosidase II beta-subunit in trimming of high-mannose-type glycans. Glycobiology *19*, 834-840.

Weber, T. and Deutsch, C. (**2012**). Oceanic nitrogen reservoir regulated by plankton diversity and ocean circulation. Nature *489*, 419-422.

Weerapana, E. and Imperiali, B. (**2006**). Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. Glycobiology *16*, 91R-101R.

Wetzel, CE., Van de vijver, B. and Ector, L. (**2010**). Luticola deniseae sp. nov. A new epizoic diatom from the Rio Negro (Amazon hydrographic basin). Vie et Milieu *60 (3)*, 177-184.

Whitley, P., Nilsson, IM. and von Heijne, G. (**1996**). A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. J. Biol. Chem. *271*, 6241-6244.

Wildt, S. and Gerngross, TU. (**2005**). The humanization of N-glycosylation pathways in yeast. Nat. Rev. Microbiol. *3*, 119-128.

Wilkinson, BM., Purswani, J. and Stirling, CJ. (**2006**). Yeast GTB1 encodes a subunit of glucosidase II required for glycoprotein processing in the endoplasmic reticulum. J. Biol. Chem. *281*, 6325-6333.

Williams, DB. (**2006**). Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. J. Cell. Sci. *119*, 615-623.

Williams, PJ., Wormald, MR., Dwek, RA. et al. (**1991**). Characterisation of oligosaccharides from Drosophila melanogaster glycoproteins. Biochim. Biophys. Acta *1075*, 146-153.

Wilson, CM., Kraft, C., Duggan, C. et al. (**2005**). Ribophorin I associates with a subset of membrane proteins after their integration at the sec61 translocon. J. Biol. Chem. *280*, 4195-4206.

Wilson, IB. (**2002**). Glycosylation of proteins in plants and invertebrates. Curr. Opin. Struct. Biol. *12*, 569-577.

Wilson, IB. and Altmann, F. (**1998**). Structural analysis of N-glycans from allergenic grass, ragweed and tree pollens: core alpha1,3-linked fucose and xylose present in all pollens examined. Glycoconj. J. *15*, 1055-1070.

Wilson, IB., Rendić, D., Freilinger, A. et al. (**2001**). Cloning and expression of cDNAs encoding alpha1,3-fucosyltransferase homologues from Arabidopsis thaliana. Biochim. Biophys. Acta *1527*, 88-96.

Wilson, IBH. (**2012**). The class I α1,2-mannosidases of Caenorhabditis elegans. Glycoconj. J. *29*, 173-179.

Wilson, JR., Williams, D. and Schachter, H. (**1976**). The control of glycoprotein synthesis: N-acetylgluocosamine linkage to a mannose residue as a signal for the attachment of L-fucose to the asparagine-linked N-acetylglucosamine residue of glycopeptide from Î±1-acid glycoprotein. Biochem. Biophys. Res. Commun. *72*, 909 - 916.

Wolfenbarger, LL. and Phifer, PR. (**2000**). The ecological risks and benefits of genetically engineered plants. Science *290*, 2088-2093.

Wormald, MR. and Dwek, RA. (**1999**). Glycoproteins: glycan presentation and protein-fold stability. Structure *7*, R155-60.

Yan, Q. and Lennarz, WJ. (**1999**). Oligosaccharyltransferase: A Complex Multisubunit Enzyme of the Endoplasmic Reticulum. Biochem. Biophys. Res. Commun. *266*, 684-689.

Yan, Q. and Lennarz, WJ. (**2002**). Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. J. Biol. Chem. *277*, 47692-47700.

Yongmanitchai, W. and Ward, OP. (**1991**). Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. Appl. Environ. Microbiol. *57*, 419-425.

Yoon, HS., Hackett, JD., Ciniglia, C. et al. (**2004**). A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. *21*, 809-818.

Zhang, H., Shahbazi, M., Mäkilä, EM. et al. (**2013**). Diatom silica microparticles for sustained release and permeation enhancement following oral delivery of prednisone and mesalamine. Biomaterials *34*, 9210-9219.

Zhang, W., Cao, P., Chen, S. et al. (**2003**). Synthesis of paucimannose N-glycans by Caenorhabditis elegans requires prior actions of UDP-N-acetyl-D-glucosamine:alpha-3-D-mannoside beta1,2-N-acetylglucosaminyltransferase I, alpha3,6-mannosidase II and a specific membrane-bound beta-N-acetylglucosaminidase. Biochem. J. *372*, 53-64.

de Vries, T., Knegtel, RM., Holmes, EH. et al. (**2001**). Fucosyltransferases: structure/function studies. Glycobiology *11*, 119R-128R.

van Beers, MMC. and Bardor, M. (**2012**). Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. Biotechnol J *7*, 1473-1484.

van Ree, R., Cabanes-Macheteau, M., Akkerdaas, J. et al. (**2000**). Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. J. Biol. Chem. *275*, 11451-11458.

von Schaewen, A., Frank, J. and Koiwa, H. (**2008**). Role of complex N-glycans in plant stress tolerance. Plant Signal Behav *3*, 871-873.

von Schaewen, A., Sturm, A., O'Neill, J. et al. (**1993**). Isolation of a mutant Arabidopsis plant that lacks N-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans. Plant Physiol. *102*, 1109-1118.