



HAL
open science

Exploring protein N-glycosylation in the green microalga *Dunaliella salina*

Jaber Dehghani, Juliette Balieu, Olivier Perruchon, Elodie Mathieu-Rivet,
Narimane Mati-Baouche, Patrice Lerouge, Muriel Bardor

► To cite this version:

Jaber Dehghani, Juliette Balieu, Olivier Perruchon, Elodie Mathieu-Rivet, Narimane Mati-Baouche, et al.. Exploring protein N-glycosylation in the green microalga *Dunaliella salina*. *Algal Research - Biomass, Biofuels and Bioproducts*, 2024, 10.1016/j.algal.2024.103711 . hal-04700205

HAL Id: hal-04700205

<https://hal.science/hal-04700205v1>

Submitted on 17 Sep 2024

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.

Journal Pre-proof

Exploring protein N-glycosylation in the green microalga
Dunaliella salina

Jaber Dehghani, Juliette Balieu, Olivier Perruchon, Elodie Mathieu-Rivet, Narimane Mati-Baouche, Patrice Lerouge, Muriel Bardor



PII: S2211-9264(24)00323-0

DOI: <https://doi.org/10.1016/j.algal.2024.103711>

Reference: ALGAL 103711

To appear in: *Algal Research*

Received date: 17 June 2024

Revised date: 4 September 2024

Accepted date: 10 September 2024

Please cite this article as: J. Dehghani, J. Balieu, O. Perruchon, et al., Exploring protein N-glycosylation in the green microalga *Dunaliella salina*, *Algal Research* (2024), <https://doi.org/10.1016/j.algal.2024.103711>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier B.V.

Exploring protein *N*-glycosylation in the green microalga *Dunaliella salina*

Jaber Dehghani, Juliette Balieu, Olivier Perruchon, Elodie Mathieu-Rivet, Narimane Mati-Baouche, Patrice Lerouge[†] and Muriel Bardor^{†*}

Université de Rouen Normandie, Laboratoire GlycoMEV UR 4358, SFR Normandie Végétal FED 4277, Innovation Chimie Carnot, GDR CNRS Chemobiologie, RMT BESTIM, F-76000 Rouen, France

*Correspondence author; muriel.bardor@univ-rouen.fr (ORCID: 0000-0002-0966-903X)

[†] These authors contributed equally to this work.

Abstract

N-glycosylation is a major post-translational modification of proteins that has a crucial influence on cell targeting, activity, and half-life. This process starts in the endoplasmic reticulum where an oligosaccharide precursor is added to the newly synthesized protein and continues in the Golgi apparatus where the *N*-linked carbohydrate sequences are processed. Importantly, the most approved recombinant pharmaceutical proteins (so-called biologics) are glycoproteins mainly currently produced in mammalian cells which is a lengthy, costly, and complex process. Today, several microalgae such as the diatom *Phaeodactylum tricornutum*, and the green microalgae *Chlamydomonas reinhardtii*, *Chlorella vulgaris*, and *Dunaliella salina* are considered as efficient and eco-friendly alternative platforms for the production of biologics. However, unlike for *C. reinhardtii*, *C. vulgaris*, and *P. tricornutum*, there is to date no data reported regarding the protein *N*-glycosylation pathway in *D. salina*. Here, we first investigated the protein *N*-glycosylation in this green microalga by MALDI-TOF mass spectrometry. These analyses showed that proteins from *D. salina* are *N*-glycosylated with Man₅GlcNAc₂ oligomannoside. Using genome mining approaches, we then identified genes encoding proteins involved in the *N*-glycosylation pathways in *D. salina*. Genetic similarities and phylogenetic relationships of the putative sequences with homologues from *C. reinhardtii*, *P. tricornutum*, and humans were investigated. These data revealed that in *D. salina* the biosynthesis of nucleotide sugars and *N*-glycan biosynthesis share mainly similarities with the GnT I-independent pathway of *C. reinhardtii* that gives rise to the synthesis of a non-canonical oligomannoside Man₅GlcNAc₂. Although an $\alpha(1,3)$ -fucosyltransferase is identified in the *D. salina* genome, impairment of the cytosolic GDP-Fuc biosynthesis prevents the Golgi fucosylation of *N*-glycans. Taken together, these data demonstrated that proteins from *D. salina* are homogeneously *N*-glycosylated with a non-canonical Man₅GlcNAc₂.

Keywords: *Dunaliella salina*; *N*-glycosylation; microalgae; fucosylation; mannosidase; GnT I

1. Introduction

Protein *O*- and *N*-glycosylation are two of the most prevalent post-translational modifications (PTMs) in eukaryotic organisms [1, 2]. Indeed, protein glycosylation is involved in different protein functions including cell-cell interactions, targeting, cell differentiation, and localization [1]. Moreover, *N*-linked glycosylation has a key impact on the dynamics, proper folding, and half-life of glycoproteins [3]. Among the most approved recombinant pharmaceutical proteins, named biologics, seventy percent are glycosylated proteins [4, 5]. Today, biologics are mainly produced in mammalian cell lines because of the similarity of PTMs compared to those of human cells [6]. However, high sensitivity to stresses and pathogens, complex culture requirements, and costly production and purification processes have encouraged many researchers to explore alternative cost-effective platforms for the production of biologics [4]. In this context, microalgae have recently been investigated for their capability to produce recombinant glycoproteins. Until now, several microalgae, especially *Chlamydomonas reinhardtii*, *Phaeodactylum tricorutum*, *Chlorella vulgaris*, and *Dunaliella* species, have attracted much attention as innovative expression systems of biologics [7-11]. These microorganisms present several important advantages as a production platform, including having no common pathogens with humans, being approved as human food, exhibiting a fast-growing rate in inexpensive culture media, and having the capability to express recombinant proteins [4, 12]. Moreover, microalgae are able to properly perform PTMs of proteins, including *N*-glycosylation [13]. Currently, different glycoproteins, such as monoclonal antibodies, cytokines, enzymes, hormones, and antigens have been successfully produced in microalgae [4, 14, 15]. Importantly, it is reported that glycoproteins produced by microalgae provide acceptable functionality and stability [10, 16]. *Dunaliella salina* is a cell wall-less green microalga that synthesizes high-value metabolites such as β -carotene, unsaturated lipids, and essential amino acids [13, 17]. Moreover, *Dunaliella salina* has been recently reported as a suitable platform for the production of biologics [13, 18]. However, in contrast to *C. reinhardtii*, *C. vulgaris* and *P. tricorutum* [19], data on protein *N*-glycosylation in *Dunaliella* is missing to date.

The human *N*-glycosylation process starts in the endoplasmic reticulum (ER) through the successive action of asparagine-linked glycosylation enzymes (ALG) which transfer activated monosaccharides to a dolichol pyrophosphate (PP-Dol) oligosaccharide to perform the biosynthesis of a membrane-anchored $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$ [5] (Fig. 1). This precursor is then transferred *en bloc* by the oligosaccharyltransferase complex (OST) on Asn residues of

consensus *N*-glycosylation sites of newly synthesized proteins [20]. It is usually assumed that *N*-glycosylation events occurring in the ER are highly conserved in eukaryotic organisms because they give rise to *N*-linked glycans that ensure the quality control of secreted proteins through the interaction with lectin-chaperones such as calnexin and calreticulin [21]. The two glucoses of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ are first cleaved by glucosidases I and II (GCS I and GCS II). Then, the resulting $\text{GlcMan}_9\text{GlcNAc}_2$ is de- and re-glycosylated by GCS II and the glucosyltransferase UGGT, respectively. Interaction of this mono-glucosylated *N*-glycan with the lectins, calnexin, and calreticulin, ensures that only properly folded glycoproteins are transported to the Golgi apparatus for further processing. With regards to misfolded glycoproteins, their mannose trimming by α -mannosidases (EDEM1/2) results in specific *N*-glycan sequences that are recognized by the ER-associated degradation (ERAD) system [22].

Differences in the *N*-glycan processing between eukaryotes then occur in the Golgi apparatus depending on their specific enzyme repertoires. In human cells, $\text{Man}_9\text{GlcNAc}_2$ arising from the removal of Glc residues from the precursor during the protein quality control process is converted into poly antennary *N*-glycans composed of a core $\text{Man}_3\text{GlcNAc}_2$ bearing an $\alpha(1,6)$ -fucose on the proximal GlcNAc residue and sialyllactosamine extensions on mannose residues (Fig. 1). These complex *N*-glycans result from the stepwise actions of α -mannosidase I (MAN I), *N*-acetylglucosaminyltransferase I (GnT I), α -mannosidase II (MAN II), *N*-acetylglucosaminyltransferase II (GnT II), $\alpha(1,6)$ -fucosyltransferase ($\alpha(1,6)$ -FucT), $\beta(1,4)$ -galactosyltransferase ($\beta(1,4)$ -GalT) and finally sialyltransferases (Fig. 1) [1]. The *N*-glycan biosynthetic pathway in the diatom *P. tricornutum* also occurs through a GnT I-dependent process leading to oligomannosides and a fucosylated $\text{Man}_3\text{GlcNAc}_2$ [23]. The biosynthesis of protein *N*-linked glycans in this diatom starts, as in the human pathway, with a $\text{Man}_9\text{GlcNAc}_2$ that is successively processed in the GA by MAN I, GnT I, MAN II, $\alpha(1,3)$ -FucT and hexosaminidases (Fig. 1) [19]. In contrast, the protein *N*-glycosylation pathway in *C. reinhardtii* is independent of a GnT I glycosyltransferase. Mature *N*-glycans are mainly composed of $\text{Man}_5\text{GlcNAc}_2$ with substitution by one or two xyloses, one fucose and also 6-*O*-methylation of mannose residues (Fig. 1) [19, 24]. However, $\text{Man}_5\text{GlcNAc}_2$ *N*-linked to *C. reinhardtii* proteins exhibits an arrangement of mannose residues different from the canonical oligosaccharide described in *P. tricornutum* or human (Figs. 1 and S1) [25]. This is due to the absence of ER-resident mannosyltransferases ALG3, 9, and 12 in *C. reinhardtii* and, as a consequence, of the

synthesis of a truncated $\text{Glc}_3\text{Man}_5\text{GlcNAc}_2\text{-PP-Dol}$ precursor (Fig. 1) [26]. Mature *N*-glycans then result from the action of a fucosyltransferase, xylosyltransferases, and *O*-methyltransferases on this non-canonical $\text{Man}_5\text{GlcNAc}_2$ (Fig. 1) [27]. With regards to *C. vulgaris*, mainly highly *O*-methylated oligomannosides have been found *N*-linked to proteins of this green microalgae [27, 28]. It should be highlighted that both experimental and *in silico* approaches confirmed that galactose and sialic acid are absent in the *N*-glycans of proteins from the microalgae *C. vulgaris*, *C. reinhardtii*, and *P. tricornutum* [23, 29, 30]. Sialic acids linked to mammalian glycoconjugates are involved in many vital events such as cell-cell and carbohydrate-protein interactions, immune reactions, targeting, and proper stability [31].

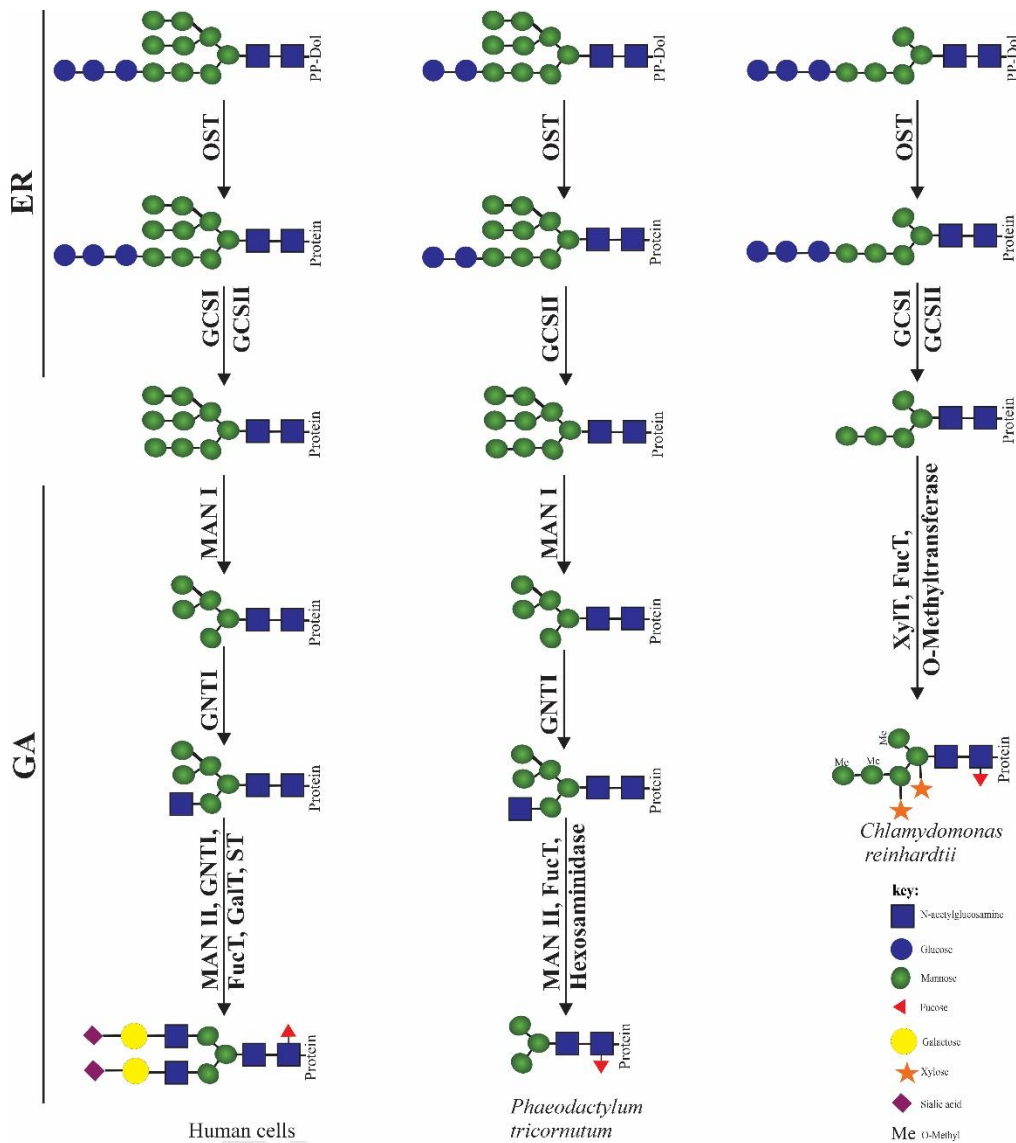


Fig. 1. N-glycan processing in human cells and in the microalgae *P. tricornutum* and *C. reinhardtii*.

Therefore, attributable to the lack of reported data on the *D. salina* glycosylation pathway and considering that this microalgae species is intended to be used as a cell bio-factory for the production of biologics, we here investigate the N-glycan profile of endogenous proteins from this microalga and search in its genome for genes involved in nucleotide sugar and N-glycan biosynthesis through genome mining approaches. Candidates were identified in the genome of *D. salina* by a search for homologues of genes encoding proteins of the glycosylation pathways in humans, *C. reinhardtii*, and *P. tricornutum*. The two later microalgal organisms were chosen because they are representative of GnT I-dependent and GnT I-independent N-glycan pathways

(Fig. 1). Identification of Pfam-specific domains and phylogenetic analyses showed that the nucleotide sugar biosynthesis and the *N*-glycosylation process in *D. salina* are similar to those of *C. reinhardtii*, except for the absence of putative xylosyltransferases, and give rise to the synthesis of a main *N*-glycan being a non-canonical Man₅GlcNAc₂ (Fig. S1).

2. Materials and methods

2.1. Microalga growth conditions

The *D. salina* strain CCAP 19/18 was purchased from the Culture Collection of Algae and Protozoa (Dunbeg, Scotland, United Kingdom). The cells were cultured in the Modified Johnson's medium (1.5 g MgCl₂·6 H₂O, 0.5 g MgSO₄·7 H₂O, 0.2 g KCl, 0.2 g CaCl₂·2 H₂O, 1 g KNO₃, 0.043 g NaHCO₃, 0.035 g KH₂PO₄, 10 mL Fe solution and 10 mL of trace element solution) with 0.5M NaCl salinity at 19°C without agitation at a light density of 30 μmol photon m⁻²s⁻¹ irradiance with a light-dark photoperiod of 16h/8h [32].

2.2. Protein extraction

The cells were grown in 500 mL Erlenmeyer until reaching the optical density of 25×10^7 cells.ml⁻¹ and then centrifuged at 5,000 g for 10 min (Beckman Coulter Allegra® X-15 R, USA). The total proteins were extracted from the cell pellet with 100 mM Tris buffer pH 7.5. The cell pellets were resuspended with 1.5 mL of extraction buffer into a Lysing Matrix E tube (MP biomedical, Illkirch-Graffenstaden, France). Furthermore, the cells were homogenized by a high-speed homogenizer (Fast Prep-24 5GTM High-Speed Homogenizer, MP biomedical, Illkirch-Graffenstaden, France) through three cycles of 30 s at a speed of 6.5 m.s⁻¹. Afterward, the tubes were centrifuged at 10,000 g for 10 min at 4 °C and the supernatants were collected and stored at -20 °C until further treatments.

2.3. Monosaccharide composition

The monosaccharide composition of *D. salina* crude extracts was determined by gas chromatography (GC). 2-5 mg of samples were first submitted to hydrolysis with 200 μL of 2M TFA at 110°C for 2h. After evaporation of the acidic solution, the residue was submitted to methanolysis by heating in 1 mL of 1 M HCl in methanol at 80°C overnight. After evaporation of the methanol-HCl solution, the samples were then trimethylsilylated by heating for 20 min at

110°C in hexamethyldisilazane: trimethylchlorosilane: pyridine (2:1:10) to yield trimethylsilyl O-methyl glycosides. After evaporation of the reagent, the samples were dissolved in cyclohexane before being analyzed by GC which was performed on an Agilent 8860 GC instrument (Agilent Technologies, Palo Alto, CA, USA). Separations were carried out on a CP-Sil 5CB capillary column (Agilent Technologies) with a film thickness of 250 μm . The carrier gas was 99.9% helium at a flow rate of 1.3 $\text{mL}\cdot\text{min}^{-1}$. The injector and ion source temperatures were set to 280 and 230°C, respectively. Samples were injected with 1:5 split mode. The temperature of the GC oven was first maintained at 40°C for 3 min and then increased up to 160°C at a rate of 15°C $\cdot\text{min}^{-1}$ then up to 280°C at a rate of 1.5°C $\cdot\text{min}^{-1}$. Separated molecules were detected *via* a flame ionization detector (FID). Quantification of monosaccharides was performed by comparison of peak surfaces with those of standard monosaccharides (Ara, Fuc, Rha, Xyl, Man, Gal, Glc, GlcUA GalUA), an internal standard (inositol) and taking into account the specific response factor of each monomer. Analyses were carried out in triplicate.

2.4. N-glycan profiling by MALDI-TOF mass spectrometry

About 5 mg of total proteins were separated on 12 % SDS-PAGE gel and stained by Coomassie Brilliant blue R250. Protein bands were cut into small pieces and washed 3 times with NH_4HCO_3 0.1M pH 8 / acetonitrile (v/v). Next, proteins were reduced with 0.1M dithiothreitol in 0.1M NH_4HCO_3 pH 8 for 45 min at 56°C. Afterward, the cysteine residues were alkylated with 55mM iodoacetamide in 0.1M NH_4HCO_3 pH 8 for 30 min at room temperature in the dark and then the supernatant was discarded. Gel pieces corresponding to the entire gel length were first incubated at 37°C overnight by trypsin (Sigma-Aldrich and Promega, USA; in a ratio 1: 50) in 0.1M NH_4HCO_3 pH 8. The sample was then heated at 100 °C for 10 minutes to inactivate the trypsin. Gel pieces were further incubated with chymotrypsin (Sigma-Aldrich, USA; in a ratio 1: 50) for 3 h at 37 °C in 0.1M NH_4HCO_3 pH 8 before heating at 100°C for 10 minutes. Peptides and glycopeptides were recovered from the gel pieces by successive washes with 50% acetonitrile (v/v), 5% formic acid (v/v), 0.1M NH_4HCO_3 pH 8, 100 % acetonitrile (v/v) and 5% formic acid (v/v). The solution containing the peptides and glycopeptides was dried down using the SpeedVac™ SPD120 (Thermo Fisher Scientific, USA), resuspended in 50 mM of sodium acetate pH 5, and then incubated with PNGase A (Agilent, USA; in a ratio 0.5 mU for 2.5 mg of proteins) at 37 °C overnight with vigorous agitation. The released N-glycans were isolated using

a C18 cartridge (Thermo Fisher Scientific, USA) and purified by Carbograph cartridges (Thermo Fisher Scientific, USA) as previously described [29]. For 2-amidobenzamide (2AB) labelling, the *N*-glycans were incubated in 2AB 0.35M dissolved in DMSO/acetic acid containing 1M sodium cyanoborohydride for 2 h at 60 °C. 2AB labelled *N*-glycans were then purified by a cartridge S (Agilent Technologies, USA) and eluted in water. 2AB-labelled *N*-glycans were analysed by MALDI-TOF mass spectrometry as previously described [23, 33]. For MALDI-TOF MS/MS analysis, the 2AB-labelled *N*-glycans were permethylated. To proceed, 0.9 g of dried sodium hydroxide was ground in 3 mL of anhydrous dimethyl sulfoxide. Half mL of this preparation and 0.5 mL of iodomethane were added to 2AB-labelled *N*-glycans and then incubated overnight at room temperature under agitation. Then, 1 mL of water was added drop by drop in order to stop the reaction. 1 mL of chloroform and 3 mL of water were then added to the sample and mixed. The upper aqueous phase was discarded and 3 mL of water was added. This step was repeated until pH was neutralized. The chloroform phase containing permethylated 2AB-labelled *N*-glycans was dried down. Permethylated 2AB-labelled *N*-glycans were re-suspended in 80% methanol (v/v) and purified on a C18 cartridge (Thermo Fisher Scientific, USA) by successive elution with 1.8 mL of 15%, 35%, 50% and 75% of acetonitrile in water. Permethylated 2AB-labelled *N*-glycans were mainly recovered in the 50% acetonitrile fraction. Major ions from MALDI-TOF-MS permethylated 2AB-labelled *N*-glycans analyses were selected for MS/MS analyses (collision-induced dissociation) with a window range of 10 Da. MS/MS scans were recorded in positive ionization mode by accumulating 10,000 laser shots.

2.5. Datasets, search methodology, and analyses

Information on genes encoding proteins involved in the human *N*-glycosylation biosynthesis was obtained from literature and related pathways annotated in the UniProt database (www.uniprot.org). The presence of the genes or their homologue counterparts in *D. salina*, *C. reinhardtii*, and *P. triconutum* were investigated through the Phytozome, UniProt, and Ensembl Protists databases. The human sequences were used as queries against the *D. salina* genome by BLASTX and BLASTP programs to find the best match for the corresponding genes [34]. Moreover, annotation of the identified sequences was retrieved via TBLASTN analysis using protein sequences of *Homo sapiens* counterparts as a query using NCBI and/or UniProt databases. Sequence alignments and subsequent phylogenetic analyses were performed with the

Clustal W algorithm included in the Mega software version 11 by default parameters. Additionally, the maximum likelihood (ML) approach was carried out by Mega software version 11 using the Whelan and Goldman model+Gama Distributed (WAG+G) model which is predicted as the best model based on BIC and AICc values [32]. The predicted *N*-glycosylation pathways were also drawn using CorelDRAW software version 2019. Moreover, searches for the presence of specific Pfam domains were investigated using ProtCID (<http://dunbrack2.fccc.edu/ProtCiD/Default.aspx>), Pfam (<http://pfam.xfam.org/>) and InterProScan (<https://www.ebi.ac.uk/interpro/search/sequence/>) datasets.

3. Results

3.1. Monosaccharide composition of *D. salina* crude extract

The sugar composition of *D. salina* crude extracts was determined to investigate the monosaccharide contents of glycoconjugates and polysaccharides in this green microalga. Crude extracts of *D. salina* were hydrolyzed and the resulting monosaccharides were converted into trimethylsilyl methyl glycoside derivatives and analyzed by gas chromatography. Table 1 shows the sugar composition which indicates that Glc and Gal are the main monomers. In contrast, fucose was not detected in the *D. salina* crude extracts.

Table 1. Sugar composition (pentoses, deoxyhexose, uronic acids, and hexoses) of crude extracts of *D. salina*

| | Ara | Rha | Xyl | Man | GlcA | GalA | Gal | Glc |
|--------------------------------------|-------------|-------------|-------------|-------------|-------------|--------------|--------------|------------|
| Relative % +/- SD^a | 2.3 +/- 0.1 | 3.9 +/- 0.1 | 2.5 +/- 0.1 | 2.7 +/- 0.2 | 2.8 +/- 0.2 | 10.1 +/- 0.6 | 26.2 +/- 0.8 | 49.6 +/- 2 |

^a SD: Standard deviation of three technical replicates.

3.2. *N*-glycan profiling of *D. salina* proteins

For the determination of the *N*-glycosylation profile of proteins isolated from *D. salina*, *N*-glycans were released from total proteins by PNGase A which is a peptide *N*-glycosidase able to release *N*-glycans from Asn residues including those carrying a fucose $\alpha(1,3)$ -linked to the proximal GlcNAc residue of the chitobiose motif [35]. *N*-glycans were then labelled with 2-aminobenzamide (2AB) to improve their detection in mass spectrometry. MALDI-TOF mass spectrum of 2AB-labelled *N*-glycans released from total proteins showed a major ion $[M+Na]^+$ at m/z 1377.512 that was assigned to Man₅GlcNAc₂-2AB (molecular formula: C₅₄H₈₈O₃₆N₃Na;

calculated m/z 1377.505) (Fig. 2A). A minor *N*-glycan with 6 hexoses was also detected and could be attributed to 2AB-labelled $\text{GlcMan}_5\text{GlcNAc}_2$ in which an additional Glc residue is attached to $\text{Man}_5\text{GlcNAc}_2$. Such a monoglucosylated *N*-glycan likely results from an uncompleted removal of terminal Glc residues by GCS II as previously reported in *C. reinhardtii* [29]. For confirmation of the structure of the major glycan, the sample was permethylated and the fragmentation pattern of the resulting permethylated $\text{Man}_5\text{GlcNAc}_2$ -2AB at m/z 1741.8 was investigated in MALDI-TOF-MS/MS. As depicted in Fig. 2B, the cleavage of the chitobiose motif gives rise to two major fragment ions at m/z 1279 and 462 as previously observed for permethylated $\text{Man}_5\text{GlcNAc}_2$ *N*-glycans [22] (Fig. 2B). We thus concluded that *D. salina* proteins mainly carry $\text{Man}_5\text{GlcNAc}_2$. It should be noted that the MS/MS fragmentation did not allow to discriminate between canonical and non-canonical $\text{Man}_5\text{GlcNAc}_2$ isomers (Fig. S1).

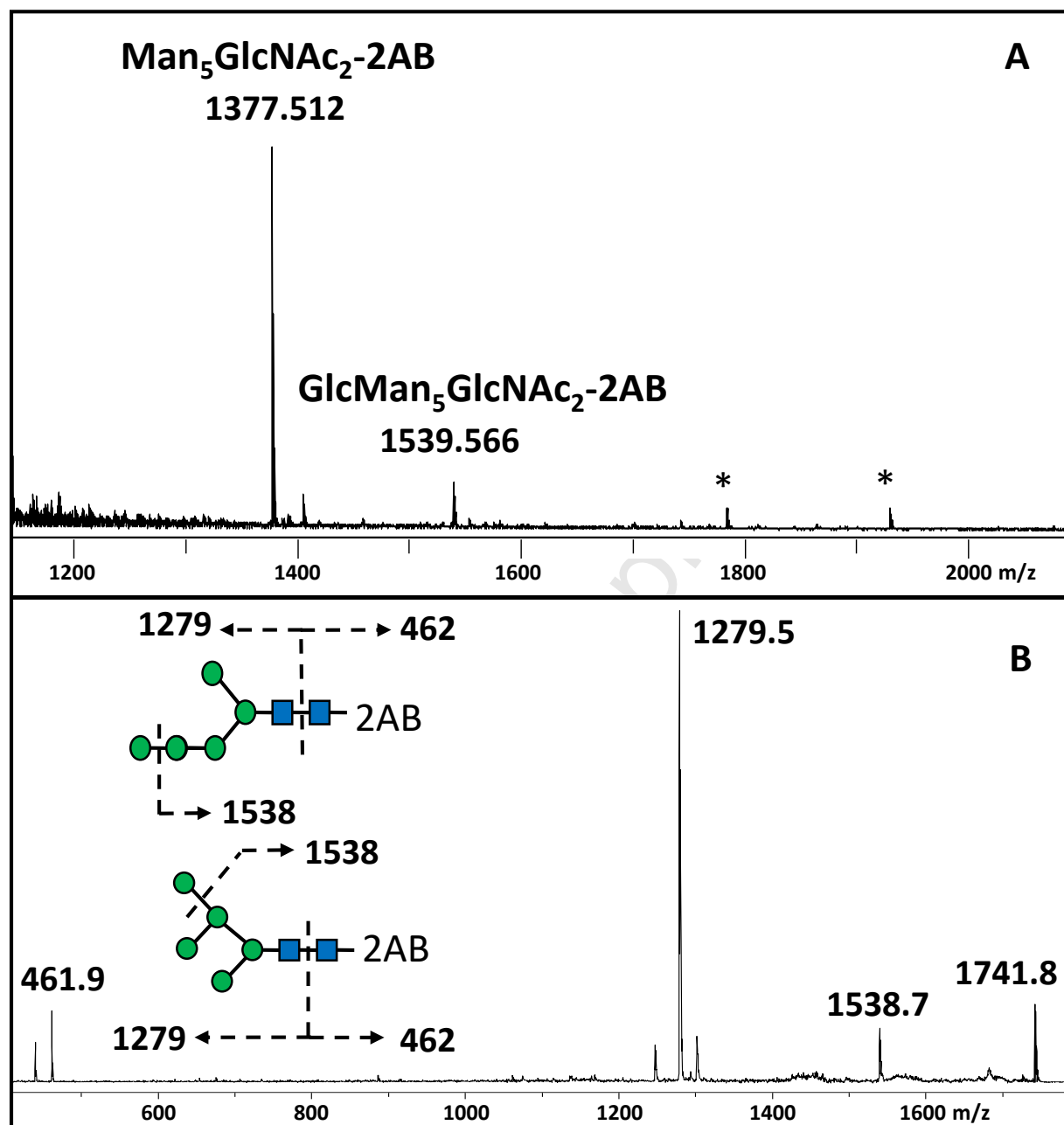


Fig. 2. (A) MALDI-TOF mass spectrum of $[M+Na]^+$ ions assigned to 2AB-labelled *N*-glycans released from total endogenous proteins of *D. salina*. *: contaminants. (B) MALDI-TOF-MS/MS spectrum of $[M+Na]^+$ ion of permethylated 2AB-labelled $Man_5GlcNAc_2$ at m/z 1741.8. Green circle: Mannose; blue square: *N*-acetylglucosamine.

3.3. Prediction of the biosynthesis pathway of nucleotide sugars in *D. salina*

In eukaryotes, nucleotide sugars are produced in the cytosol through nucleotide sugar interconversion pathways involving the action of specific epimerases, reductases, oxidases, and transaminases. The search in the *D. salina* genome using protein sequences from *C. reinhardtii*, *P. tricornutum*, and humans as queries allowed the identification of 14 sequences exhibited Pfam domains that characterize enzymes involved in nucleotide sugar interconversion pathways (Fig. 3, Table 2). With regard to monosaccharides identified in the sugar composition of *D. salina* extracts (Table 1), genes encoding homologues of enzymes involved in the biosynthesis of UDP-Xyl, UDP-Rha, UDP-uronic acids (UDP-GalA and UDP-GlcA) and hexose nucleotides (GDP-Man and UDP-Gal) are identified in the *D. salina* genome. Homologues of enzymes involved in the biosynthesis of UDP-GlcNAc from Fru-6P that is required for *N*-glycan biosynthesis were also identified. However, although Ara was detected in the sugar composition (Table 2), we failed to identify in the genome of *D. salina* a gene encoding a putative UDP-xylose 4-epimerase (UXE) which is responsible for converting UDP-xylose to UDP-arabinose while it is identified in *C. reinhardtii* (Gene ID; Cre09.g401022).. With regards to the synthesis of GDP-Fuc, a GDP-L-fucose synthase (TSTA3) was identified in the *D. salina* genome that shared the specific Pfam domain (PF01370) (Table 2). Despite this prediction in *D. salina*, we were unable to identify a homologue of GDP-mannose 4, 6-dehydratase (GMDS) that provides the substrate for TSTA3 in the GDP-fucose biosynthesis. The lack of GMDS activity may explain the absence of fucose in the sugar composition of *D. salina* crude extracts (Table 1).

It is worth noting that the phosphoglucomutase 1 (PGM1 for the isomerization of Glc-6P to Glc-1P) identified in *D. salina* presented lower similarities to their *P. tricornutum* (20.32%) and human (18.95%) homologues while it showed more similarity to that found in *C. reinhardtii* (49%). We thus concluded that PGM1 from *D. salina* and *C. reinhardtii* are evolutionary far from the human sequence although they share the conserved phosphoglucomutase domain I (PF02878) (Fig. S2 and Table 2). Genes encoding enzymes involved in the CMP-sialic acid biosynthesis (*i.e.*, UDP-GlcNAc 2-epimerase, Neu5Ac synthase, Neu5Ac-9P phosphatase, and CMP-Neu5Ac synthase) were not identified in *D. salina* genome suggesting the absence of a sialylation process as previously reported in microalgae and plants [36].

Table 2. List of the identified genes encoding putative enzymes involved in the nucleotide sugar biosynthesis in *D. salina*

| Enzyme name | Gene ID | Pfam | Functionality | URL in Phytozome database |
|-------------|------------------|---------|---|---------------------------|
| GAE | Dusal.2930s00002 | PF04321 | UDP-glucuronate 4-epimerase | Dusal.2930s00002 |
| GALE | Dusal.0081s00005 | PF16363 | UDP-glucose 4-epimerase | Dusal.0081s00005 |
| GCK | Dusal.0652s00009 | PF00349 | Glucokinase | Dusal.0652s00009 |
| GFPT | Dusal.0953s00002 | PF01380 | Glutamine-fructose-6-phosphate transaminase | Dusal.0953s00002 |
| GMPPB | Dusal.1025s00004 | PF00483 | Mannose-1-phosphate guanylyltransferase | Dusal.1025s00004 |
| GNPNAT1 | Dusal.0100s00013 | PF00583 | N-acetyltransferase | Dusal.0100s00013 |
| GPI | Dusal.0233s00003 | PF04188 | Phosphoglucose isomerase | Dusal.0233s00003 |
| MPI | Dusal.0799s00005 | PF01238 | Phosphomannose isomerase | Dusal.0799s00005 |
| PGM1 | Dusal.0358s00001 | PF02878 | Phosphoglucomutase 1 | Dusal.0358s00001 |
| PGM3 | Dusal.0057s00021 | PF02880 | N-acetylglucosamine phosphomutase 3 | Dusal.0057s00021 |
| PMM1 | Dusal.0091s00023 | PF03332 | Phosphomannomutase | Dusal.0091s00023 |
| RHM | Dusal.1480s00001 | PF01370 | Rhamnose biosynthetic enzyme 1-related | Dusal.1480s00001 |
| TSTA3 | Dusal.0366s00015 | PF01370 | GDP-L-fucose synthase | Dusal.0366s00015 |
| UAP1 | Dusal.0347s00016 | PF01704 | UTP-N-acetylglucosamine-1-phosphate uridylyltransferase | Dusal.0347s00016 |
| UER1 | Dusal.1013s00001 | PF04321 | UDP-4-keto-rhamnose-4-keto-reductase activity | Dusal.1013s00001 |
| UGDH | Dusal.0681s00003 | PF03721 | UDP-glucose dehydrogenase | Dusal.0681s00003 |
| UGP2 | Dusal.0096s00007 | PF01704 | UTP-glucose-1-phosphate uridylyltransferase | Dusal.0096s00007 |
| UXS1 | Dusal.0032s00001 | PF11803 | UDP-glucuronate decarboxylase | Dusal.0032s00001 |

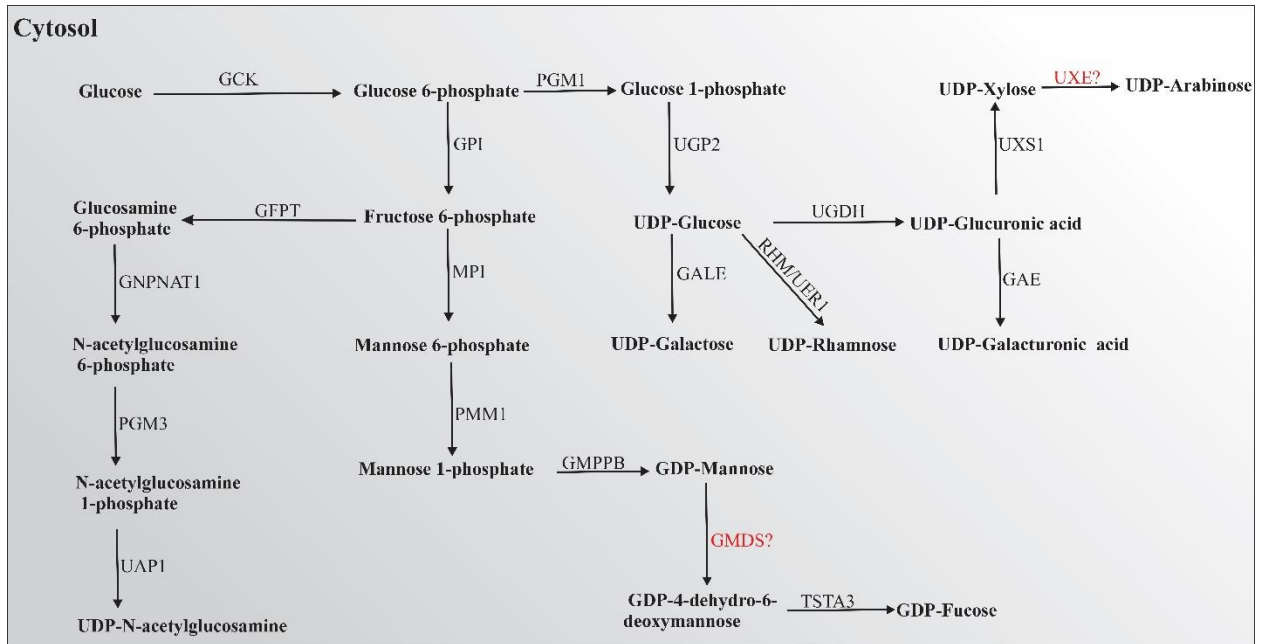


Fig. 3. Predicted scheme for the biosynthesis of nucleotide sugars in the *D. salina*. See Table 2 for the functionality and full names of enzymes. We could not identify the GMDS and UXE enzymes in *D. salina* (pointed out in red) through BLASTX and BLASTP programs.

3.4. N-glycan biosynthesis and glycoprotein quality control in the ER of *D. salina*

Among ALG genes, 9 homologues were identified in the *D. salina* genome (Fig S3). However, candidate genes for ALG3, ALG9, ALG11, and ALG12 were not found. The *D. salina* ALGs exhibit Pfam-specific domains (Table 3). Genes encoding the subunits of the dolichol-phosphate mannosyltransferase (DPM), as well as the dolichol-phosphate glucosyltransferase ALG5, were also identified in the *D. salina* genome. Both DPM and ALG5 are hexose donors for ALG on the luminal side of the ER. RFT1, a homologue of the human flippase, that allows the flipping of the lipid-linked intermediate on the ER luminal side, was also identified in the *D. salina* genome. The Pfam and phylogenetic data suggested that the *D. salina* RFT1 is closely related to that found in the microalgae and human cells (Table 3 and Fig. S3).

The absence of ALG3, 9, and 12 that encode ER luminal mannosyltransferases has already been reported in *C. reinhardtii* and gives rise to the biosynthesis of a truncated dolichol pyrophosphate precursor in the ER of this green microalga [29]. Based on the bioinformatic

prediction highlighting in *D. salina* the absence of ALG3, 9 and 12, we thus postulated that this microalga also accumulates a truncated Glc₃Man₅GlcNAc₂-PP-Dol precursor as depicted for *C. reinhardtii* in Fig. 1. Two genes encoding putative $\alpha(1,2)$ -mannosyltransferases were previously reported in the literature [37] and categorized as ALG11 in the KEGG database (EC 2.4.1.131), that is the enzyme responsible for the transfer of two Man residues on Man₃GlcNAc₂-PP-Dol. However, these transferases exhibit a Pfam domain of PF01793 characteristic for glycolipid 2- α -mannosyltransferases whereas ALG11 enzymes involved in the *N*-glycosylation process share a Pfam domain of PF15924 [29, 37] (Table 3). Notably, these sequences share low similarities with ALG11 from *C. reinhardtii*, *P. tricornutum*, and human homologues (22.14%, 26.04%, and 26.42%, respectively). Moreover, the phylogenetic tree showed these putative ALG11 sequences were not clustered with their counterparts from microalgae and humans (Fig. S3).

Table 3. List of the identified genes encoding putative enzymes involved in the different steps of *N*-glycosylation processing in *D. salina*.

| Name | Gene ID | Pfam | Functionality | URL in Phytozome database |
|---|------------------|---------|---|---------------------------|
| Precursor biosynthesis | | | | |
| ALG1 | Dusal.1301s00001 | PF13439 | Mannosyltransferase CAZy GT4 | Dusal.1301s00001 |
| ALG2 | Dusal.0116s00018 | PF13439 | Mannosyltransferase CAZy GT4 | Dusal.0116s00018 |
| ALG5 | Dusal.0262s00013 | PF00535 | Dolichol-phosphate glucosyltransferase | Dusal.0262s00013 |
| ALG6 | Dusal.0804s00001 | PF03155 | Glucosyltransferase CAZy GT57 | Dusal.0804s00001 |
| ALG7 | Dusal.0267s00017 | PF00953 | <i>N</i> -acetylglucosaminyltransferase CAZy GT4 | Dusal.0267s00017 |
| ALG8 | Dusal.0782s00002 | PF03155 | Glucosyltransferase CAZy GT57 | Dusal.0782s00002 |
| ALG10 | Dusal.0425s00005 | PF04922 | Glucosyltransferase CAZy GT59 | Dusal.0425s00005 |
| ALG11 | Dusal.0218s00001 | PF01793 | Mannosyltransferase CAZy GT15 | Dusal.0218s00001 |
| ALG11 | Dusal.0218s00005 | PF01793 | Mannosyltransferase CAZy GT15 | Dusal.0218s00005 |
| ALG13 | Dusal.0104s00004 | PF04101 | <i>N</i> -acetylglucosaminyltransferase CAZy GT28 | Dusal.0104s00004 |
| ALG14 | Dusal.0138s00012 | PF08660 | <i>N</i> -acetylglucosaminyltransferase CAZy GT1 | Dusal.0138s00012 |
| DPM1 | Dusal.0457s00012 | PF00535 | Dolichol-phosphate mannosyltransferase subunit 1 | Dusal.0457s00012 |
| DPM2 | Dusal.1011s00006 | PF07297 | Dolichol-phosphate mannosyltransferase subunit 2 | Dusal.1011s00006 |
| DPM3 | Dusal.0686s00001 | PF08285 | Dolichol-phosphate mannosyltransferase subunit 3 | Dusal.0686s00001 |
| RFT1 | Dusal.0238s00005 | PF04506 | Flippase | Dusal.0238s00005 |
| Oligosaccharyltransferase (OST) subunits | | | | |
| DAD1 | Dusal.0122s00021 | PF02109 | DAD family | Dusal.0122s00021 |
| DDOST | Dusal.1270s00002 | PF03345 | Oligosaccharyltransferase, subunit beta | Dusal.1270s00002 |
| OST3/ OST6 | Dusal.0600s00012 | PF04756 | OST3 / OST6 family | Dusal.0600s00012 |
| OST4 | Dusal.0348s00012 | PF10215 | OST4 family | Dusal.0348s00012 |
| RPN1 | Dusal.1458s00001 | PF04597 | Ribophorin I | Dusal.1458s00001 |

| | | | | |
|---------------------------|------------------|--------------------|--|------------------|
| RPN2 | Dusal.0052s00036 | PF05817 | Ribophorin II | Dusal.0052s00036 |
| STT3B | Dusal.0132s00020 | PF02516 | STT3 subunit | Dusal.0132s00020 |
| ER quality control | | | | |
| CALR | Dusal.0454s00001 | PF00262 | Calreticulin family | Dusal.0454s00001 |
| HRD1 | Dusal.0127s00002 | PF13639 | E3 ubiquitin-protein ligase | Dusal.0127s00002 |
| EDEM2 | Dusal.0172s00002 | PF01532 | Mannosylhydrolase CAZy GH47 | Dusal.0172s00002 |
| GCS I | Dusal.1215s00001 | PF16923 | Glucosylhydrolase CAZy GH63 | Dusal.1215s00001 |
| GCS II | Dusal.0307s00020 | PF01055 | Glucosylhydrolase CAZy GH31 | Dusal.0307s00020 |
| UGGT | Dusal.0003s00028 | PF06427 | UDP-glucose:glycoprotein glucosyltransferase | Dusal.0003s00028 |
| Golgi apparatus | | | | |
| MAN I | Dusal.0847s00002 | PF01532 | Mannosylhydrolase CAZy GH47 | Dusal.0847s00002 |
| FUT 11 | Dusal.1386s00002 | PF00852 | Fucosyltransferase CAZy GT10 | Dusal.1386s00002 |
| Unknown functions | | | | |
| B4GALT4 | Dusal.1441s00001 | PF02709 PF13733 | C-terminal domain of GalT N-terminal domain of GalT | Dusal.1441s00001 |

In eukaryotes, the *N*-glycan precursor is transferred on newly synthesized proteins in the lumen of the ER by a multimeric oligosaccharyltransferase (OST). Among the 12 subunits of the human OST, 7 homologues were identified in the *D. salina* genome (Table 3). The phylogenetic data revealed that OST subunits identified in *D. salina* are closer to their counterparts predicted in *C. reinhardtii*, rather than *P. tricornutum*, and human homologues (Fig. S4). It should be noted that the *D. salina* OST4 subunit is not identified in *P. tricornutum* and *C. reinhardtii*. This subunit is a small membrane protein that interacts with STT3 and RPN subunits to stabilize the OST complex [38].

ER quality control of glycoproteins involves glucosidases GCS I and II, the glucosyltransferase UGGT, chaperones calreticulin (CALR) and calnexin, as well as mannosidases EDEM1/2. Sequences encoding putative GCSI, GCSII, EDEM2, UGGT, and CALR (but not calnexin) were identified in the *D. salina* genome (Table 3, Fig. S3). However, EDEM1 was not found neither in *D. salina*, nor in *C. reinhardtii* or *P. tricornutum*. Moreover, the E3-ubiquitin ligase Hydroxymethyl glutaryl-coenzyme A reductase degradation protein 1 (HRD1), the core component of ERAD, was identified in *D. salina* (Table 3). Based on these data, we assumed that the biosynthesis and the protein quality control in the ER of *D. salina* occur as depicted in Fig. 3. After the biosynthesis of a $\text{Glc}_3\text{Man}_5\text{GlcNAc}_2$ truncated precursor and its transfer to the Asn residues of the *N*-glycosylation consensus sites, the glycoprotein is submitted to the quality control and then transported to the secretory pathway. The unfolded or

misfolded *N*-glycans are recognized by EDEM2 and HRD1 to address them to the cytosol to be degraded by the proteasome (Fig. 4).

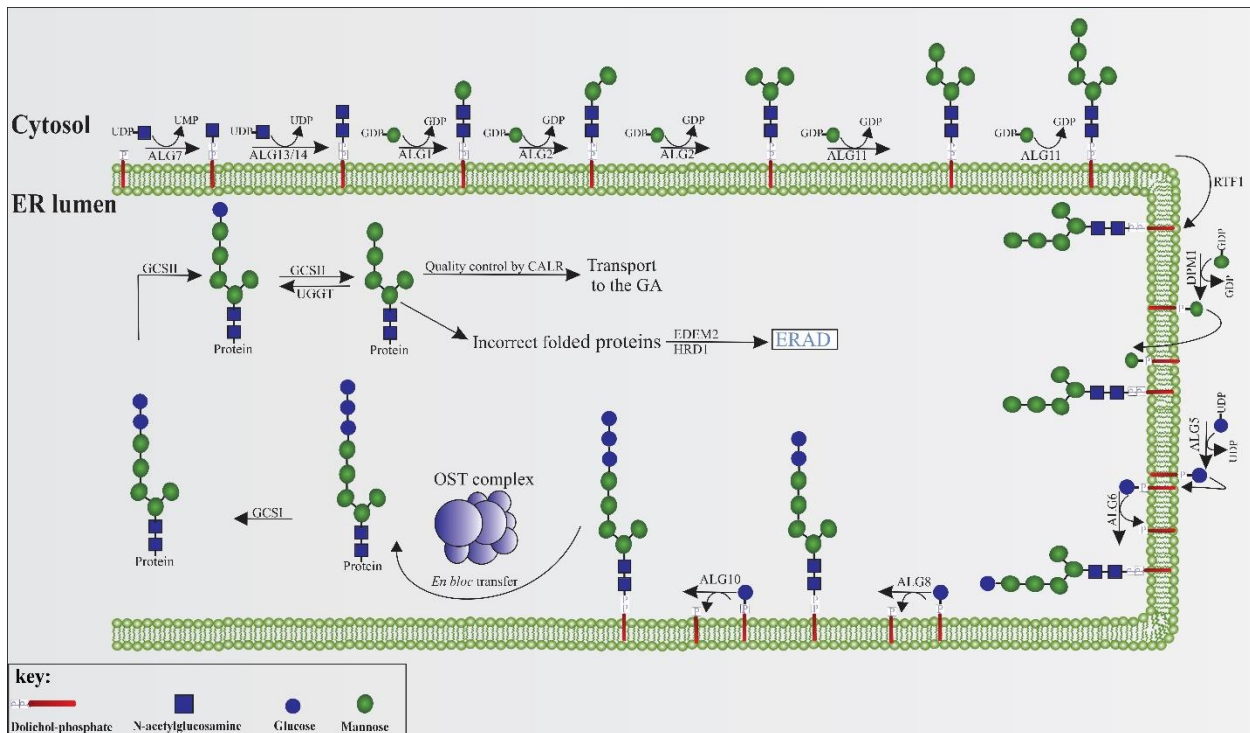


Fig. 4. Predicted scheme for the *N*-glycan biosynthesis and processing in ER of *D. salina*. It appears that this pathway is similar to that of the microalga *C. reinhardtii*, as no sequence encoding the luminal mannosyltransferases ALG3, ALG9, and ALG12 was identified in *D. salina*.

3.5. *N*-glycan processing in the Golgi apparatus of *D. salina*

In eukaryotes, *N*-glycans leaving the ER are processed by GnT I-dependent or independent pathways involving a set of Golgi glycosidases and glycosyltransferases of different specificities (Fig. 1). Gene encoding GnT I homologue was not identified in the *D. salina* genome indicating that the processing of *N*-linked glycans in proteins of *D. salina* occurs in a GnT I-independent manner as previously reported for *C. reinhardtii*. GnT II and Man II, two key enzymes of GnT I-dependent pathways, are also missing in *D. salina*. In its genome, we could only identify homologue genes encoding α -mannosidase I (MAN I) and α (1,3)-fucosyltransferase (FUT11) exhibiting specific Pfam domains named PF01532 and PF00852 (Table 3). The putative FUT11 enzyme identified in *D. salina* shared 36% and 34% of similarity with their *C. reinhardtii* and human homologues respectively. Regarding the predicted MAN I, highest similarities (18% to

20%) are found with their counterparts found in Arabidopsis, human and *P. tricornutum*. It should be mentioned that we could not find any gene coding $\beta(1,2)$ -xylosyltransferase in *D. salina*.

In addition to MAN I and FUT11, we could identify a putative protein sequence in *D. salina* that shares 33.18% similarity with human $\beta(1,4)$ -galactosyltransferase (B4GALT4) (Table 3). The putative GalT sequence exhibits the two Pfam domains PF02709 and PF13733 that are respectively found in C- and N-terminal ends of $\beta(1,4)$ -galactosyltransferases. It also contains the DVD metal binding motif and the WGWGGEDDD sequence of the short flexible loop that is conserved in the catalytic domain of these enzymes [39, 40] (Table 3). Interestingly, we could also identify a homologue of this sequence in *P. tricornutum* but not in *C. reinhardtii*. The precise function of this protein needs more investigation because B4GALT4 is able to transfer galactose on terminal GlcNAc in the *N*-glycan biosynthesis. However, terminal GlcNAc residues are absent in *D. salina* *N*-glycans because of the lack of GnT I activity in this green microalga. As a consequence, we postulate that this galactosyltransferase candidate may be involved in the biosynthesis of other galactose-containing glycoconjugates in *D. salina*.

4. Discussion

Glycosylation is considered as one of the most fundamental PTMs in eukaryotic cells that could affect the structural, physicochemical, and functional properties of proteins. This process, especially the *N*-glycosylation pathways, has been deeply explored in mammals, yeasts, insects, filamentous fungi, and plants [1, 41]. In contrast, *N*-glycosylation processing is still poorly understood in microalgae and so far, data are mainly available for *C. reinhardtii*, *P. tricornutum*, *C. vulgaris*, and *Porphyridium* [5]. Importantly, there is no reported data about *N*-glycosylation pathways in *D. salina* although the expression of biologics has been successfully performed in it [5]. For example, the VP28 glycoprotein that can be used as a vaccine against white spot syndrome virus (WSSV) in shrimps was produced in *D. salina*. Interestingly, oral delivery of transgenic *Dunaliella* cells could induce protection against WSSV in crayfish [42].

In silico analysis of the *N*-glycosylation pathway in *D. salina* suggests that the cytosolic biosynthesis of monosaccharides is closely related to those of other eukaryotes. The only exception is arabinose because UDP-xylose 4-epimerase was not identified in the genome of this

green microalga [43]. Since arabinose was detected in the sugar composition of *D. salina* (Table 1), we postulate that this monosaccharide may arise from an alternative pathway such as a bifunctional nucleotide sugar 4-epimerase as reported in plants [44, 45]. With regards to fucose, the biosynthesis pathway of this deoxyhexose lacks a key enzyme (GMDS) likely causing the absence of fucose in the sugar composition of *D. salina*. This suggests that the availability of this monosaccharide is not crucial for the development of this cell wall-less green microalga in contrast to data reported in other eukaryotes [44, 45]. Enzymes involved in the CMP-sialic acid biosynthetic pathway are also not predicted in *D. salina* as previously reported in other microalgae [19].

The biosynthesis of the lipid-linked ER precursor, its transfer to Asn residues of protein by the OST complex, and the quality control of newly synthesized glycoproteins likely occur in the ER as previously reported in other eukaryotes. Among all ALG sequences reported in Table 3, the identification of a gene encoding a putative ALG11 remains questionable. However, based on the biochemical data, we assume that the absence of ALG3, 9, and 12 in *D. salina* results in the biosynthesis of a truncated Dol-PP-linked $\text{Glc}_3\text{Man}_5\text{GlcNAc}_2$ precursor in the ER as already reported for *C. reinhardtii* [29]. The absence of GnT I in *D. salina* also supports this hypothesis because this Golgi enzyme only transfers a GlcNAc residue onto the terminal mannose $\alpha(1,3)$ -linked to the β -mannose of the core of canonical $\text{Man}_5\text{GlcNAc}_2$ (Fig. S1). With regards to the OST, *D. salina* contains the OST4 subunit which was not identified in *C. reinhardtii*. This subunit is a very small membrane protein that interacts with STT3 and RPN subunits resulting in the maintenance of the stability of the OST complex [38].

We found very little data about the processing that occurs in the GA in *D. salina*. Sequences encoding putative fucosyltransferase (FUT11) and α -mannosidase (MAN I) are identified in the genome. These candidates share homology with homologues from land plants α -mannosidase I and $\alpha(1,3)$ -fucosyltransferase that is involved in the Golgi trimming of oligomannosides and the $\alpha(1,3)$ -fucosylation of the proximal GlcNAc residue. In contrast, gene sequences encoding GnT I, GnT II, and MAN II which are key enzymes in the GnT I-dependent *N*-glycosylation pathway, as well as xylosyltransferases, were not identified in the *D. salina* genome. *N*-glycan profiling by MALDI-TOF mass spectrometry and MS/MS fragmentation shows that proteins from *D. salina* are homogeneously *N*-glycosylated by $\text{Man}_5\text{GlcNAc}_2$,

although the presence of low-abundant shorter oligomannosides cannot be ruled out. Taken together, *in silico* and analytical approaches suggest that *D. salina* proteins are *N*-glycosylated with a non-canonical $\text{Man}_5\text{GlcNAc}_2$ resulting from the biosynthesis of a truncated $\text{Glc}_3\text{Man}_5\text{GlcNAc}_2$ precursor in the ER and then removal of terminal Glc residues by GCS I and GCS II. This is an unexpected result because the Golgi processing of protein *N*-glycans in eukaryotes usually gives rise to a wide diversity of oligosaccharides that depend on the Golgi enzyme repertoires [46]. Moreover, to our knowledge, this is the first example of a protein *N*-glycan pathway where protein *N*-glycan arising from the ER biosynthesis is not further processed in the GA. The impairment of fucose biosynthesis likely contributes to this observation because the identified FUT11 is not supplied with activated GDP-Fuc nucleotide sugar in the GA.

From a biotech perspective, this homogeneous oligomannoside glycosylation offers many advantages because immunogenic glycoepitopes such as $\alpha(1,3)$ -fucose and $\beta(1,2)$ -xylose and also *O*-methylation are missing in *D. salina*, while these allergic motifs have been reported in *C. reinhardtii* [5] and *C. vulgaris* [46]. Moreover, *N*-glycan humanization by glycoengineering to get human-compatible biologics is more achievable because it concerns a single *N*-glycan rather than a mixture of glycan structures. For instance, this can be achieved by expression of an $\alpha(1,2)$ -mannosidase in *D. salina* to remove the two $\alpha(1,2)$ -mannoses of the non-canonical $\text{Man}_5\text{GlcNAc}_2$ and then complementation with GnT I and GalT to get the core *N*-glycan found in mammals.

CRedit authorship contribution statement

Jaber Dehghani: Investigation, Methodology, Validation, Writing – original draft, Funding acquisition. **Juliette Balieu:** Investigation, Methodology, Validation. **Olivier Perruchon:** Methodology, Formal analysis. **Elodie Mathieu-Rivet:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Funding acquisition. **Narimane Mati-Baouche:** Validation, Formal analysis, Writing – review & editing, Funding acquisition. **Patrice Lerouge:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Funding acquisition, supervision. **Muriel Bardor:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Funding acquisition, supervision.

Acknowledgments

We wish to thank the Normandy Region (France), the Université de Rouen Normandie, the Laboratoire GlycoMEV UR 4358, and the European Union (Horizon 2020 Research and Innovation Programme) for their financial support. The authors would like to thank Carole Burel, assistant engineer, for technical help and advice in this work.

Declaration of competing interest

All authors declare that there are no competing or financial conflicts of interest.

Funding

J.D. is a recipient of the WINNING Normandy Program supported by the Normandy Region, France. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101034329. This work was also financially supported by the French government through the ANR agency under the ANR PRCE DAGENTA project (ANR-21-CE20-0038-001) and the program «Grand défi Biomédicament: améliorer les rendements et maîtriser les coûts de production: Nouveaux Systèmes d'Expression—2020» (PHAEOMABS project—ANR-21-F2II-0005).

References

- [1] K.T. Schjoldager, Y. Narimatsu, H.J. Joshi, H. Clausen, Global view of human protein glycosylation pathways and functions, *Nat Rev Mol Cell Biol*, 21 (2020) 729-749, DOI: 10.1038/s41580-020-00294-x.
- [2] S. Ramazi, J. Zahiri, Posttranslational modifications in proteins: resources, tools and prediction methods, *Database (Oxford)*, 2021 (2021), DOI: 10.1093/database/baab012.
- [3] H.S. Lee, Y. Qi, W. Im, Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study, *Sci Rep*, 5 (2015) 8926, DOI: 10.1038/srep08926.
- [4] J. Dehghani, K. Adibkia, A. Movafeghi, H. Maleki-Kakelar, N. Saeedi, Y. Omid, Towards a new avenue for producing therapeutic proteins: Microalgae as a tempting green biofactory, *Biotechnol Adv*, 40 (2020) 107499, DOI: 10.1016/j.biotechadv.2019.107499.
- [5] E. Mathieu-Rivet, M.C. Kiefer-Meyer, G. Vanier, C. Ovide, C. Burel, P. Lerouge, M. Bardor, Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals, *Front Plant Sci*, 5 (2014) 359, DOI: 10.3389/fpls.2014.00359.
- [6] F.J. Krambeck, S.V. Bennun, M.R. Andersen, M.J. Betenbaugh, Model-based analysis of N-glycosylation in Chinese hamster ovary cells, *PLoS One*, 12 (2017) e0175376, DOI: 10.1371/journal.pone.0175376.
- [7] A.J. Berndt, T.N. Smalley, B. Ren, R. Simkovsky, A. Badary, A.E. Sproles, F.J. Fields, Y. Torres-Tiji, V. Heredia, S.P. Mayfield, Recombinant production of a functional SARS-CoV-2 spike receptor binding domain in the green algae *Chlamydomonas reinhardtii*, *PLoS One*, 16 (2021) e0257089, DOI: 10.1371/journal.pone.0257089.

- [8] F. Hempel, J. Lau, A. Klingl, U.G. Maier, Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*, *PLoS One*, 6 (2011) e28424, DOI: 10.1371/journal.pone.0028424.
- [9] B. Yang, J. Liu, Y. Jiang, F. Chen, *Chlorella* species as hosts for genetic engineering and expression of heterologous proteins: Progress, challenge and perspective, *Biotechnol J*, 11 (2016) 1244-1261, DOI: 10.1002/biot.201500617.
- [10] J. Dehghani, K. Adibkia, A. Movafeghi, M.M. Pourseif, Y. Omid, Designing a new generation of expression toolkits for engineering of green microalgae; robust production of human interleukin-2, *BioImpacts* : BI, 10 (2020) 259-268, DOI: 10.34172/bi.2020.33.
- [11] A. Banerjee, V. Ward, Production of recombinant and therapeutic proteins in microalgae, *Curr Opin Biotechnol*, 78 (2022) 102784, DOI: 10.1016/j.copbio.2022.102784.
- [12] J. Dehghani, K. Adibkia, A. Movafeghi, A. Barzegari, M.M. Pourseif, H. Maleki Kakelar, A. Golchin, Y. Omid, Stable transformation of *Spirulina (Arthrospira) platensis*: a promising microalga for production of edible vaccines, *Appl Microbiol Biotechnol*, 102 (2018) 9267-9278, DOI: 10.1007/s00253-018-9296-7.
- [13] J. Dehghani, A. Movafeghi, A. Barzegari, J. Barar, Efficient and stable transformation of *Dunaliella pseudosalina* by 3 strains of *Agrobacterium tumefaciens*, *Bioimpacts*, 7 (2017) 247-254, DOI: 10.15171/bi.2017.29.
- [14] B.A. Rasala, P.A. Lee, Z. Shen, S.P. Briggs, M. Mendez, S.P. Mayfield, Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide, *PLoS One*, 7 (2012) e43349, DOI: 10.1371/journal.pone.0043349.
- [15] G. Vanier, F. Hempel, P. Chan, M. Rodamer, D. Vaudry, U.G. Maier, P. Lerouge, M. Bardor, Biochemical Characterization of Human Anti-Hepatitis B Monoclonal Antibody Produced in the Microalgae *Phaeodactylum tricornutum*, *PLoS One*, 10 (2015) e0139282, DOI: 10.1371/journal.pone.0139282.
- [16] G. Vanier, S. Stelter, J. Vanier, F. Hempel, U.G. Maier, P. Lerouge, J. Ma, M. Bardor, Alga-Made Anti-Hepatitis B Antibody Binds to Human Fcγ Receptors, *Biotechnol J*, 13 (2018) e1700496, DOI: 10.1002/biot.201700496.
- [17] Y. Xu, I.M. Ibrahim, C.I. Wosu, A. Ben-Amotz, P.J. Harvey, Potential of New Isolates of *Dunaliella Salina* for Natural β-Carotene Production, *Biology (Basel)*, 7 (2018), DOI: 10.3390/biology7010014.
- [18] S. Feng, X. Li, Z. Xu, J. Qi, *Dunaliella salina* as a novel host for the production of recombinant proteins, *Appl Microbiol Biotechnol*, 98 (2014) 4293-4300, DOI: 10.1007/s00253-014-5636-4.
- [19] E. Mathieu-Rivet, P. Lerouge, M. Bardor, *Chlamydomonas reinhardtii*: Protein Glycosylation and Production of Biopharmaceuticals, in: M. Hippler (Ed.) *Chlamydomonas: Biotechnology and Biomedicine*, Springer International Publishing, Cham, 2017, pp. 45-72.
- [20] P. Burda, M. Aebi, The dolichol pathway of N-linked glycosylation, *Biochim Biophys Acta*, 1426 (1999) 239-257, DOI: 10.1016/s0304-4165(98)00127-5.
- [21] L.A. Rutkevich, D.B. Williams, Participation of lectin chaperones and thiol oxidoreductases in protein folding within the endoplasmic reticulum, *Curr Opin Cell Biol*, 23 (2011) 157-166, DOI: 10.1016/j.ceb.2010.10.011.
- [22] G.Z. Lederkremer, Glycoprotein folding, quality control and ER-associated degradation, *Curr Opin Struct Biol*, 19 (2009) 515-523, DOI: 10.1016/j.sbi.2009.06.004.
- [23] B. Baïet, C. Burel, B. Saint-Jean, R. Louvet, L. Menu-Bouaouiche, M.C. Kiefer-Meyer, E. Mathieu-Rivet, T. Lefebvre, H. Castel, A. Carlier, J.P. Cadoret, P. Lerouge, M. Bardor, N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme, *J Biol Chem*, 286 (2011) 6152-6164, DOI: 10.1074/jbc.M110.175711.
- [24] A. Oltmanns, L. Hoepfner, M. Scholz, K. Zinzius, S. Schulze, M. Hippler, Novel Insights Into N-Glycan Fucosylation and Core Xylosylation in *C. reinhardtii*, *Front Plant Sci*, 10 (2019) 1686, DOI: 10.3389/fpls.2019.01686.

- [25] G. Vanier, P.-L. Lucas, C. Loutelier-Bourhis, J. Vanier, C. Plasson, M.-L. Walet-Balieu, P.C. Tchi-Song, I. Remy-Jouet, V. Richard, S. Bernard, A. Driouich, C. Afonso, P. Lerouge, E. Mathieu-Rivet, M. Bardor, Heterologous expression of the N-acetylglucosaminyltransferase I dictates a reinvestigation of the N-glycosylation pathway in *Chlamydomonas reinhardtii*, *Scientific reports*, 7 (2017) 10156-10156, DOI: 10.1038/s41598-017-10698-z.
- [26] P.L. Lucas, E. Mathieu-Rivet, P.C.T. Song, A. Oltmanns, C. Loutelier-Bourhis, C. Plasson, C. Afonso, M. Hippler, P. Lerouge, N. Mati-Baouche, M. Bardor, Multiple xylosyltransferases heterogeneously xylosylate protein N-linked glycans in *Chlamydomonas reinhardtii*, *Plant J*, 102 (2020) 230-245, DOI: 10.1111/tpj.14620.
- [27] P.L. Lucas, R. Dumontier, C. Loutelier-Bourhis, A. Mareck, C. Afonso, P. Lerouge, N. Mati-Baouche, M. Bardor, User-friendly extraction and multistage tandem mass spectrometry based analysis of lipid-linked oligosaccharides in microalgae, *Plant Methods*, 14 (2018) 107, DOI: 10.1186/s13007-018-0374-8.
- [28] R. Mócsai, R. Figl, C. Troschl, R. Strasser, E. Svehla, M. Windwarder, A. Thader, F. Altmann, N-glycans of the microalga *Chlorella vulgaris* are of the oligomannosidic type but highly methylated, *Sci Rep*, 9 (2019) 331, DOI: 10.1038/s41598-018-36884-1.
- [29] E. Mathieu-Rivet, M. Scholz, C. Arias, F. Dardelle, S. Schulze, F. Le Mauff, G. Teo, A.K. Hochmal, A. Blanco-Rivero, C. Loutelier-Bourhis, M.C. Kiefer-Meyer, C. Fufezan, C. Burel, P. Lerouge, F. Martinez, M. Bardor, M. Hippler, Exploring the N-glycosylation pathway in *Chlamydomonas reinhardtii* unravels novel complex structures, *Mol Cell Proteomics*, 12 (2013) 3160-3183, DOI: 10.1074/mcp.M113.028191.
- [30] X. Xie, H. Du, J. Chen, M. Aslam, W. Wang, W. Chen, P. Li, H. Du, X. Liu, Global Profiling of N-Glycoproteins and N-Glycans in the Diatom *Phaeodactylum tricornutum*, *Front Plant Sci*, 12 (2021) 779307, DOI: 10.3389/fpls.2021.779307.
- [31] S. Ghosh, Sialic acid and biology of life: An introduction, *Sialic Acids and Sialoglycoconjugates in the Biology of Life, Health and Disease*, (2020) 1-61, DOI: 10.1016/B978-0-12-816126-5.00001-9.
- [32] J. Dehghani, E. Atazadeh, Y. Omid, A. Movafeghi, The use of 18S ribosomal DNA, ITS and molecular markers to study the genus (*Dunaliellaceae*) in Iranian samples: A phylogenetic approach, *Oceanological and Hydrobiological Studies*, 49 (2020) 88-98,
- [33] J. Balieu, J.W. Jung, P. Chan, G.P. Lomonosoff, P. Lerouge, M. Bardor, Investigation of the N-Glycosylation of the SARS-CoV-2 S Protein Contained in VLPs Produced in *Nicotiana benthamiana*, *Molecules*, 27 (2022), DOI: 10.3390/molecules27165119.
- [34] P. McVeigh, K. Cwiklinski, A. Garcia-Campos, G. Mulcahy, S.M. O'Neill, A.G. Maule, J.P. Dalton, In silico analyses of protein glycosylating genes in the helminth *Fasciola hepatica* (liver fluke) predict protein-linked glycan simplicity and reveal temporally-dynamic expression profiles, *Sci Rep*, 8 (2018) 11700, DOI: 10.1038/s41598-018-29673-3.
- [35] V. Tretter, F. Altmann, L. März, 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 (1991) 647-652, DOI: 10.1111/j.1432-1033.1991.tb16166.x.
- [36] M. Séveno, M. Bardor, T. Paccalet, V. Gomord, P. Lerouge, L. Faye, Glycoprotein sialylation in plants?, *Nat Biotechnol*, 22 (2004) 1351-1352; author reply 1352-1353, DOI: 10.1038/nbt1104-1351.
- [37] J.E.W. Polle, S. Calhoun, Z. McKie-Krisberg, S. Prochnik, P. Neofotis, W.C. Yim, L.T. Hathwaik, J. Jenkins, H. Molina, J. Bunkenborg, I.V. Grigoriev, K. Barry, J. Schmutz, E. Jin, J.C. Cushman, J.K. Magnusson, Genomic adaptations of the green alga *Dunaliella salina* to life under high salinity, *Algal Research*, 50 (2020) 101990, DOI: <https://doi.org/10.1016/j.algal.2020.101990>.
- [38] A. Dumax-Vorzet, P. Roboti, S. High, OST4 is a subunit of the mammalian oligosaccharyltransferase required for efficient N-glycosylation, *Journal of cell science*, 126 (2013) 2595-2606, DOI: 10.1242/jcs.115410.
- [39] R. Strasser, Plant protein glycosylation, *Glycobiology*, 26 (2016) 926-939, DOI: 10.1093/glycob/cww023.

- [40] B. Ramakrishnan, P.K. Qasba, Structure-based evolutionary relationship of glycosyltransferases: a case study of vertebrate β 1,4-galactosyltransferase, invertebrate β 1,4-N-acetylgalactosaminyltransferase and α -polypeptidyl-N-acetylgalactosaminyltransferase, *Curr Opin Struct Biol*, 20 (2010) 536-542, DOI: 10.1016/j.sbi.2010.07.004.
- [41] V. Gomord, A.C. Fitchette, L. Menu-Bouaouiche, C. Saint-Jore-Dupas, C. Plasson, D. Michaud, L. Faye, Plant-specific glycosylation patterns in the context of therapeutic protein production, *Plant Biotechnol J*, 8 (2010) 564-587, DOI: 10.1111/j.1467-7652.2009.00497.x.
- [42] S. Feng, W. Feng, L. Zhao, H. Gu, Q. Li, K. Shi, S. Guo, N. Zhang, Preparation of transgenic *Dunaliella salina* for immunization against white spot syndrome virus in crayfish, *Arch Virol*, 159 (2014) 519-525, DOI: 10.1007/s00705-013-1856-7.
- [43] T. Kotake, R. Takata, R. Verma, M. Takaba, D. Yamaguchi, T. Orita, S. Kaneko, K. Matsuoka, T. Koyama, W.D. Reiter, Y. Tsumuraya, Bifunctional cytosolic UDP-glucose 4-epimerases catalyse the interconversion between UDP-D-xylose and UDP-L-arabinose in plants, *Biochem J*, 424 (2009) 169-177, DOI: 10.1042/bj20091025.
- [44] D.J. Becker, J.B. Lowe, Fucose: biosynthesis and biological function in mammals, *Glycobiology*, 13 (2003) 41r-53r, DOI: 10.1093/glycob/cwg054.
- [45] L. Zhang, B.C. Paasch, J. Chen, B. Day, S.Y. He, An important role of l-fucose biosynthesis and protein fucosylation genes in *Arabidopsis* immunity, *New Phytol*, 222 (2019) 981-994, DOI: 10.1111/nph.15639.
- [46] C. Toustou, M.L. Walet-Balieu, M.C. Kiefer-Meyer, M. Houdou, P. Lerouge, F. Foulquier, M. Bardor, Towards understanding the extensive diversity of protein N-glycan structures in eukaryotes, *Biol Rev Camb Philos Soc*, 97 (2022) 732-748, DOI: 10.1111/brv.12820.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Highlights

- * The *N*-glycosylation profile of *Dunaleilla salina* proteins was reported for the first time by biochemical analysis and *in silico* approaches.
- * The *N*-glycosylation pathway in *D. salina* is independent of GnT I enzyme as reported previously for the microalga *Chlamydomonas reinhardtii*.
- * No fucose and xylose residues were identified on the glycoproteins of *D. salina*.
- * The lack of allergic or immunogenic residues in *D. salina* glycan made this microalga an attractive platform for the production of recombinant pharmaceutical glycoproteins.

Journal Pre-proof