Exploring the N-glycosylation Pathway in Chlamydomonas reinhardtii Unravels Novel Complex Structures*§

Elodie Mathieu-Rivet‡§, Martin Scholz§¶, Carolina Arias§¶, Flavien Dardelle‡**, Stefan Schulze¶¶, François Le Mauff‡‡, Gavin Teo‡‡, Ana Karina Hochmal¶, Amaya Blanco-Rivero||, Corinne Loutelier-Bourhis §§, Marie-Christine Kiefer-Meyer‡, Christian Fufezan¶, Carole Burel‡, Patrice Lerouge‡, Flor Martinez¶, Muriel Bardor‡§¶¶, and Michael Hippler§¶¶¶

Chlamydomonas reinhardtii is a green unicellular eukaryotic model organism for studying relevant biological and biotechnological questions. The availability of genomic resources and the growing interest in C. reinhardtii as an emerging cell factory for the industrial production of biopharmaceuticals require an in-depth analysis of protein N-glycosylation in this organism. Accordingly, we used a comprehensive approach including genomic, glycomic, and glycoproteomic techniques to unravel the N-glycosylation pathway of C. reinhardtii. Using mass-spectrometry-based approaches, we found that both endogenous soluble and membrane-bound proteins carry predominantly oligomannosides ranging from Man-2 to Man-5. In addition, minor complex N-linked glycans were identified as being composed of partially 6-O-methylated Man-3 to Man-5 carrying one or two xylose residues. These findings were supported by results from a glycoproteomic approach that led to the identification of 86 glycoproteins. Here, a combination of in-source collision-induced dissociation (CID) for glycan fragmentation followed by mass tag-triggered CID for peptide sequencing and PNGase F treatment of glycopeptides in the presence of 18O-labeled water in conjunction with CID mass spectrometric analyses were employed. In conclusion, our data support the notion that the biosynthesis and maturation of N-linked glycans in the endoplasmic reticulum and Golgi apparatus occur via a GnT I-independent pathway yielding novel complex N-linked glycans that mature differently from their counterparts in land plants. Molecular & Cellular Proteomics 12: 10.1074/mcp.M113.028191, 3160–3183, 2013.

Chlamydomonas reinhardtii is a green alga that is used as a model organism for studying a number of biological processes such as photosynthesis, flagellar assembly and function, organelle biosynthesis, phototaxis, and circadian rhythms (1). Studies on glycosylation pathways in C. reinhardtii have been mostly focused on O-glycosylation processing, as the cell wall of this organism consists of a vast framework of O-glycosylated hydroxyproline-rich glycoproteins (2, 3). More recently, Bollig et al. even demonstrated that O-glycans from C. reinhardtii cell wall glycoproteins contain arabinose and galactose, the latter being in the furanose form (4). In contrast, the N-glycosylation pathway, although a major post-translational modification step in the maturation of secreted proteins in eukaryotes, has received very little attention so far. In N-glycan processing, a Man₉GlcNAc₂-PP-dolichol-oligosaccharide intermediate is first assembled onto a dolichol pyrophosphate on the cytosolic face of the endoplasmic reticulum (ER). ¹ After translocation of this intermediate by a flippase, the biosynthesis continues in the lumen of the ER

¹ The abbreviations used are: 2-AB, 2-aminobenzamide; ALG, asparagine-linked glycosylation; Con A, concanavalin A; DeHex, deoxyhexose; ER, endoplasmic reticulum; FASP, filter-assisted sample preparation; GC-EIMS, gas chromatography coupled to electron ionization mass spectrometry; GnT I, N-acetylgalactosaminyltransferase I; Hex, hexose; HexNAc, N-acetylgalcosamine; HPAEC-PAD, high-pH anion exchange chromatography coupled to pulse amperometric detection; IS-CID, in-source collision-induced dissociation; MALDI-TOF, matrix-assisted laser desorption ionization–time-of-flight; MeHex, methyl hexose; MSA, multistage activation; OST, oligosaccharyltransferase; Pent, pentose; PKHD1, polycystic and hepatic disease 1; PKHD1L1, human polycystic kidney and hepatic disease 1-like 1 protein; SN, supernatant; TCE, total cell extract; UGGT, UDP glucose: glycoprotein glucosyltransferase.
until a Glc₃Man₃GlcNAc₂−PP-dolichol N-glycan precursor is completed (9). This precursor is then transferred by the oligosaccharyltransferase (OST) multisubunit complex onto the asparagine residues of the consensus Asn−X−Ser/Thr sequences of a protein (5). The precursor is then deglucosylated/reglucosylated to ensure the quality control of the neosynthesized protein through the interaction with ER-resident chaperones such as calnexin and calreticulin. These ER events are crucial for the proper folding of secreted proteins (6), conserved in eukaryotes investigated so far, and involve a limited number of oligomannoside N-glycans. In contrast, the evolutionary adaptation of N-glycan processing in the Golgi apparatus gives rise to a large variety of organism-specific complex structures (7). Type I mannosidases located in this compartment first degrade the oligosaccharide precursor into oligomannoside N-glycans ranging from Man₉GlcNAc₂ (Man-9) to Man₃GlcNAc₂ (Man-5). N-acetylglucosaminyltransferase I (GnT I) then transfers a first GlcNAc residue on the α(1,3)-mannose arm of Man-5 to initiate the synthesis of polyantennary complex-type N-glycans (7).

To date, a few studies carried out in Chlorophyceae using on-blots of different or a combination of exoglycosidase digestions and two-dimensional HPLC separation have suggested that proteins secreted by these microalgae harbor mainly oligomannosides or mature N-glycans having a core xylose residue (8–10). Deeper insight into the structure of glycans N-linked to proteins secreted by two algal species, Porphyridium sp. and Phaeodactylum tricornutum, has been recently reported. A cell wall glycoprotein from the red microalgae Porphyridium sp. was found to carry Man-8 and Man-9 oligomannosides containing 6-O-methyl mannose and substituted by one or two xylose residues (11). In contrast, glycans N-linked to proteins secreted by the diatom P. tricornutum can be processed through a GnT I-dependent pathway into paucimannosidic oligosaccharides (12).

In contrast to glycomic analysis, which focuses on the structure of N-linked oligosaccharides irrespective of the carrier proteins, glycoproteomics is used to characterize and determine the cell localization of individual proteins carrying these carbohydrate post-translational modifications. Whereas mammalian N-glycoproteomes have been studied extensively down to tissue- and cell-type-specific levels (13–17), less information is available regarding the N-glycoproteomes of plants and green algae (18, 19). The use of glycoproteomic approaches could help unravel the identity of endogenous glycoproteins from C. reinhardtii. As this green alga possesses many animal-like features (20), glycoproteomic analyses will help provide information concerning similarities and differences relative to not only mammalian but also vascular plant N-glycosylation pathways and glycoprotein trafficking.

Recently, microalgae have emerged as an alternative system for the production of biopharmaceuticals, which represents a multibillion-dollar industry worldwide (21). The high expense and complicating factor of potential virus contamination encountered with commonly used expression systems have driven scientists to seek alternatives such as C. reinhardtii cells. Actually, they are cheap, easy to grow, safe, and scalable for the production of a high amount of proteins, making them ideal hosts for industrial production (22). Several studies have already demonstrated that the green alga C. reinhardtii is a convenient platform for producing recombinant proteins, including those of human origin (23). For example, a large single-chain antibody directed against glycoprotein D of the herpes simplex virus (24) and full-length IgG1 monoclonal antibodies directed against anthrax protective antigen 83 (25) have been successfully expressed in the chloroplast of transgenic C. reinhardtii cells. The production of secreted therapeutic proteins such as erythropoietin has also been evaluated (26). In contrast to the expression of proteins in the chloroplast, protein post-translational modifications such as N-glycosylation acquired by the secreted recombinant protein are a major concern for biopharmaceuticals, as more than half of the approved ones are glycosylated (27). Moreover, glycosylation is a critical quality attribute for biopharmaceuticals, because the presence and structures of the N-glycans are required for their biological activity, stability, and half-life (28, 29). However, given that unsuitable N-glycan structures can induce immune responses in humans (30–32) and generate adverse reactions, as reported for α(1,3)-Gal epitope on therapeutic drugs like cetuximab (33), it is essential to take into account the N-glycosylation capacity for an optimal expression system. Therefore, a suitable expression system should allow the production of glycomolecules harboring N-glycans and/or O-glycans compatible with human therapeutic applications and better efficacy of the therapeutic drug (34).

In this study, we used a comprehensive approach including genomic, glycomic, and glycoproteomic analyses to investigate the N-glycosylation pathway occurring in C. reinhardtii. Our results revealed that the biosynthesis and maturation of N-glycans occur in the ER and Golgi apparatus through a GnT I-independent pathway and yield novel complex structures in addition to oligomannoside N-glycans.

EXPERIMENTAL PROCEDURES

Strains and Growth Conditions—CC-503 cw92 and CC-1036 pf18 strains were obtained from the Chlamydomonas Culture Collection at Duke University (Durham, NC) and grown in batch cultures at 26 °C, illuminated with a photosynthetic photon flux density of 150 μmol m⁻² s⁻¹ supplied from cool, white fluorescent lamps (TDL 150 W, Philips, Eindhoven, The Netherlands), using minimal medium (35) and aeration with air enriched with 5% CO₂. Cells were harvested via centrifugation, frozen in liquid nitrogen, and stored at –80 °C until use. Cultivation of C. reinhardtii CC-400-cw15 under iron-deficient and iron-deficient conditions was carried out in TAP medium as described elsewhere (36).

In Silico Genome Analysis—Annotation of genes involved in the N-glycosylation pathway in the C. reinhardtii genome (Chlamydomonas reinhardtii v4.3, available online at the Phytozone website) was carried out via TBLASTN analysis using protein sequences from Homo sapiens, Mus musculus, Arabidopsis thaliana, Drosophila melanogaster, Saccharomyces cerevisiae, and Physcomitrella patens as
Protein N-glycosylation in *Chlamydomonas reinhardtii*

queries. Sequence alignments were done with ClustalW 1.8 (37) from the BioEdit 7.0.9.0 package. Signal peptides and cell localization/targeting of mature proteins were predicted using SignalP4.0 and TargetP 1.0. Searches for the presence of predicted transmembrane domain(s) and for specific Pfam domains were done, respectively, by TMHMM and Pfam (Wellcome Trust, Sanger Institute, Cambridge, UK).

**Soluble and Membrane-bound Protein Preparation**—Ten liters of cells were harvested via 5 min of centrifugation at 2,500g (SORVALL® RCSC Plus). The cells were washed with 20 mM potassium phosphate buffer (pH 7.4) and underwent 5 min of centrifugation at 2,500g after washing. The cell pellet was packed in 20 ml of 20 mM potassium phosphate buffer (pH 7.4) plus 1 ml PIC 25x (Protease inhibitor mixture, Roche, Meylan, France) and broken with a French press (SLM Amincor, SLM Instruments, Inc., Urbana, IL) at 1,300 Pa. Cell suspensions were centrifuged at 300g for 3 min to remove intact cells and debris. The supernatant was centrifuged again at 20,000g for 30 min and then ultracentrifuged at 100,000g for 1 h (Centrifok T-2070, Kontron Instruments Montigny Le Bretonneux, France) to pellet the microsomal fraction, corresponding to the membrane-bound proteins. Concentration to 1 ml of the supernatant containing the soluble proteins was done using Amicon® Ultra centrifugal filters ULTRACEL® 10K (Millipore, Billerica, MA). All the steps in the protein preparation were carried out at 4 °C.

**Isolation of N-glycans from *C. reinhardtii* Proteins**—

**In-gel Trypsin Digestion**—Five milligrams of protein extract were loaded on an SDS-PAGE gel (4–12%, Bis-Tris, Invitrogen) run at 150 V for more than 1 h in MES buffer 1X (Invitrogen). After fixation in 20% ethanol/10% acetic acid, the gel was cut in small pieces and washed for more than 1 h in MES buffer 1X (Invitrogen). After fixation in 20% ethanol/10% acetic acid, the gel was cut in small pieces and washed for more than 1 h in MES buffer 1X (Invitrogen). After fixation in 20% ethanol/10% acetic acid, the gel was cut in small pieces and washed for more than 1 h in MES buffer 1X (Invitrogen).

**Peptide N-glycosidase A Digestion**—The eluate containing the glycopeptides was then dried down and reconstituted in 100 mM sodium acetate, pH 5.5, prior to peptide N-glycosidase A digestion (Roche). 1.5 mM of enzyme was added for every 5 mg of protein and incubated at 37 °C overnight under agitation. After digestion, the amount of released N-glycans was determined using a C18 cartridge (Waters, Milford, MA). The column was conditioned with 5 ml of ethanol followed by 5 ml of water. The sample was then loaded onto the resin prior to a washing step using 5 ml of water. The interesting glycopeptides were eluted with 4 ml of 50% acetonitrile.

**Peptide N-glycosidase F Digestion**—The peptide N-glycosidase F (Roche) digestion was performed on *C. reinhardtii* proteins according to the procedure outlined in Ref. 12. The released N-glycans were then cleaned up and labeled with 2-aminobenzamide (2-AB) prior to MALDI-TOF-MS analysis with or without α-mannosidase treatment.

2-AB Labeling of N-glycans—The 2-AB labeling was done according to Ref. 38 in order to increase the sensitivity and the ionization efficiency of the MALDI-TOF analysis as previously described for plant N-glycan analysis. The excess reagent was then removed using a cartridge S from Prozyme, Hayward, CA, USA following the manufacturer’s instructions. The 2-AB-labeled N-glycans were analyzed via MALDI-TOF MS1 and MS2 prior to eventual further permethylation. The 2-AB labeling reaction was carried out to completion, and there were no leftover unlabeled oligosaccharides—this was checked carefully during the experiment through MALDI-TOF-MS analysis prior to and after the labeling procedure.

**Permethylation**—The 2-AB-labeled N-glycan preparation was permethylated using the sodium hydroxide procedure (39, 40). The permethylated N-glycans were cleaned up using Sep-Pack C18 cartridges (Waters) according to the procedure described in Ref. 41.

**MALDI-TOF-MS Analysis**—The permethylated glycans were dried down and reconstituted in 10 μl of 90% methanol–0.1% trifluoroacetic acid (TFA) and 2-AB-labeled N-glycans were reconstituted in 10 μl of water containing 0.1% TFA before 0.5 μl of sample were spotted on a MALDI target plate at a ratio of 1:1 with 2,5-dihydroxybenzoic acid matrix at 10 mg ml⁻¹ (Waters, Milford, MA) dissolved in 80% (v/v) methanol in water. The analysis was then performed on a MALDI-TOF-TOF 5800 (AB Sciex, Framingham, MA). The MS acquisition was done in reflector positive mode with the laser intensity fixed at 63% and a pulse rate of 400 Hz. The detector’s voltage was about 1.95 kV. The MS2 experiments were performed at a voltage of 2 kV combined with activation of collision-induced dissociation (CID) by argon gas at a pressure of 5 psi. 10,000 laser shots were accumulated for each spectrum of MS1 and MS2. The 4.700 calibration standard kit Cal Mix (AB Sciex) was used for external calibration. Spectra were analyzed using Data Explorer® software (AB Sciex). The parameters used to reject or exclude outliers were a signal-to-noise ratio threshold of 3%, a centroid of 50, a noise window width of 250, and a threshold (m/z) after signal-to-noise ratio recalculating of 10. Relative quantification of the different N-glycan species was based on the MALDI-TOF-MS spectra of permethylated N-glycans as previously described (42). For this purpose, the height of the interesting ions was used to calculate their relative intensity as compared with that of all the glycan structures identified. The values presented in Table 1 correspond to the mean of data obtained from five independent N-glycan preparations and MALDI-TOF analyses. The standard deviation has been calculated and indicated. The biological reliability of all measurements was validated using at least three independent experiments for each of the three biological replicates.

α-Mannosidase Treatment—1.5 μl of 2-AB-labeled N-glycans were submitted to 215 μl of α-mannosidase from proteomic-grade *C. reinhardtii* (Sigma-Aldrich, St Louis, MO) in commercial buffer diluted five times (Sigma-Aldrich) for 24 h at 37 °C under agitation. Then, the digest was directly analyzed via MALDI-TOF-MS.

**HPAEC-PAD Analysis of N-glycans—Oligosaccharides, especially Man-5 (MC0731)** from Dextra Laboratories, Reading, UK were used as standards. These were prepared by dissolving 20 μg each of standard in 1 ml of water. The Dionex ICS-5000 system with integrated amperometry was used for HPAEC analysis. A carboPac PA200 analytical column (3 × 250 mm) with a PA200 guard column (3 × 50 mm) from Dionex (Sunnyvale, CA) and three eluents were used for the separation. The eluents were 500 mM sodium acetate, 500 mM NaOH, and deionized water. The gradient program for the elution of both neutral and charged oligosaccharides began with isocratic mode with 20% NaOH and 80% water for 10 min, followed by a ramp gradient for sodium acetate to 34% while maintaining NaOH at 20% until 78 min had passed. After the elution period of 25 min with 20% NaOH and 80% water was allowed for the next run. The waveform used was E1 = +0.05 V, t1 = 400 ms; E2 = -0.20 V, t2 = 200 ms; E3 = +0.30 V, t3 = 200 ms; E4 = +0.50 V, t4 = 500 ms; E5 = +0.65 V, t5 = 500 ms; E6 = +0.70 V, t6 = 500 ms; E7 = +0.75 V, t7 = 500 ms; E8 = +0.80 V, t8 = 500 ms. The waveform was repeated for a total of 5 minutes.
+0.75 V, \( t_2 = 200 \text{ ms} \); \( E_3 = -0.15 \text{ V} \), \( t_3 = 400 \text{ ms} \). The flow rate was kept constant at 0.3 ml min\(^{-1}\), and the volume of sample/standard injected was 30 \( \mu l \).

Monosaccharide Composition of 2-AB-labeled N-glycans by GC-EIMS—The monosaccharide composition of 2-AB-labeled N-glycans was determined via GC-EIMS. Samples were hydrolyzed with 2 M TFA, reduced with NaBD\(_4\), and peracylated. The resulting alditol acetates were separated via GC (Hewlett-Packard 6890 series gas chromatographic system) on an HP-5MS capillary column (0.25 mm inner diameter \( \times 30 \text{ m} \), 0.25-\( \mu \)m film thickness; Hewlett-Packard, Palo Alto, CA) and analyzed via electron ionization using an Autospec mass spectrometer of EBE geometry (Micromass, Manchester, UK) equipped with an Opus 3.1 data system. Helium was the carrier gas, and the flow rate was 0.8 ml min\(^{-1}\). The oven temperature was as follows: 100 °C for 1 min, 100 °C to 160 °C at 10 °C/min, 160 °C to 200 °C at 2 °C/min, 200°C to 300 °C at 15 °C/min, and 300 °C for 1 min. The temperature of the injector, the interface, and the transfer lines was 250 °C. Injections of 0.5 or 1 \( \mu l \) were performed with a split ratio of 10 or in splitless mode. The mass spectra were recorded using an ionizing electron energy of 70 eV and a trap current of 200 nA.

An ionizing electron energy of 70 eV and a trap current of 200 nA was used to confirm glyco-FASP (Amicon Ultra, 0.5-ml capacity, 30-kDa molecular-weight cutoff) according to the manufacturer’s instructions.

For the solubilization of plasma membranes, TCEs, chloroplasts, and proteins and the reduction of cysteine residues, lysis buffer (2% SDS/0.1 M DTT in 0.1 M Tris-HCl, pH 7.8) was added to the samples, and the samples were then incubated at 95 °C for 3 min. Samples were centrifuged at 16,000 \( g \) for 10 min, and SNs containing solubilized proteins underwent either glycoprotein enrichment or N-glycoproteomic analysis via the filter-assisted sample-preparation method (N-glyco-FASP) (see below). The lysis step was omitted for secreted proteins; instead, a volume of SN concentrate corresponding to 300 \( \mu g \) of protein was transferred to centrifugal filters and concentrated further to a volume of 40 \( \mu l \). Then 350 \( \mu l \) of 8 M urea/0.1 M DTT in 0.1 M Tris/HCl pH 8.5 was added and denaturation/reduction was carried out at room temperature for 45 min. After that, SN samples were immediately used for glycoprotein enrichment or N-glyco-FASP.

Glycopeptide Enrichment and PNGase F Treatment—Carbamidomethylation of cysteines, trypptic digestion, glycopeptide enrichment, and PNGase F-mediated glycan hydrolysis in \( ^{18} \text{O} \)-labeled water were performed in centrifugal filter devices (Amicon Ultra, 0.5-ml capacity, 30-kDa molecular-weight cutoff) according to the N-glyco-FASP protocol (16) with the following modifications: 300 \( \mu g \) of protein were used per sample, and glycopeptide enrichment was carried out using 150 \( \mu l \) of agarose-bound Con A (50% slurry, Vector Laboratories Inc., Burlingame, CA). PNGase F (catalog no. P0704S) was obtained from New England Biolabs, Ipswich, MA. After the elution of \( ^{18} \text{O} \)-labeled peptides, samples were dried in a vacuum centrifuge and stored at –20 °C.

Glycoprotein Enrichment for In-source Collision-Induced Dissociation Analyses—300 \( \mu g \) of protein were transferred to centrifugal filters (Amicon Ultra, 0.5-ml capacity, 30-kDa molecular-weight cutoff) and concentrated via centrifugation at 14,000 \( g \) for 15 min at room temperature to a volume of 40 \( \mu l \). 100 \( \mu l \) of UA buffer (8 M urea in 10 mM HEPES, pH 6.5) was added, and samples were centrifuged as described above (this step was repeated twice). Subsequently, samples were incubated with 100 \( \mu l \) of 50 mM iodoacetamide in UA buffer for 20 min in the dark and then centrifuged as before. Filters were washed twice with 100 \( \mu l \) of UA and twice with 200 \( \mu l \) of lectin binding buffer (500 mM NaCl, 1 mM CaCl\(_2\), 1 mM MnCl\(_2\) in 20 mM Tris-HCl, pH 7.6). Afterward, protein samples were transferred to new collection tubes via the addition of 100 \( \mu l \) binding buffer and centrifugation of the inverted filter unit at 14,000 \( g \) for 5 min. The transfer step was repeated once. Then, 200 \( \mu l \) of agarose-bound Con A were washed and equilibrated three times with 300 \( \mu l \) of binding buffer in a 1.5-ml reaction tube by means of mixing, centrifugation at 10,000 \( g \) for 5 min, and finally removal of the SN. Protein samples were added to the preconditioned Con A and incubated at room temperature overnight in a thermomixer while being shaken (1,000 rpm). Unbound proteins were removed by three washes with 300 \( \mu l \) NaCl-free binding buffer, with centrifugation at 10,000 \( g \) for 5 min between washes. Glycoproteins were eluted via the addition of 150 \( \mu l \) of 0.5 M \( \alpha \)-methyl D-mannopyranoside in NaCl-free binding buffer and incubation for 20 min at room temperature with shaking (1,000 rpm). After centrifugation (10,000 \( g \) for 5 min), the elution was repeated once, and the pooled eluates were transferred to a centrifugal filter unit (30-kDa molecular-weight cutoff). The samples were centrifuged at 14,000 \( g \) for 15 min, and \( \alpha \)-methyl D-mannopyranoside was removed by three successive washes with 200 \( \mu l \) of 50 mM ammonium bicarbonate. Afterward, glycoproteins were digested by the addition of 2 \( \mu l \) of trypsin (sequencing-grade modified, Promega, Madison, WI) in 40 \( \mu l \) ammonium bicarbonate and overnight incubation at 37 °C. Peptides were otherwise not disclosed.
were eluted via centrifugation (14,000g for 10 min). Elution was repeated twice with 50 μl of ammonium bicarbonate. Finally, peptides were dried down in a vacuum centrifuge and stored at −20 °C.

LC-MS Analysis of 18O-labeled Peptides—Chromatographic separation of glycopeptides was performed on an Ultimate 3000 Nanoflow HPLC system (Dionex), which was coupled via a nanospray source to an LTQ Orbitrap XL mass spectrometer (Thermo Scientific). The mobile phases were composed of 2% (v/v) acetonitrile/0.1% (v/v) formic acid in ultrapure water (A) and 80% acetonitrile/0.1% formic acid in ultrapure water (B). The sample was dissolved in eluent A and loaded on a trap column (C18 PepMap100, 300 μm × 5 mm, 5-μm particle size, 100-Å pore size; Dionex) using eluent A at flow rate of 25 μl min−1 for 1 min. For peptide separation, the trap column was switched in line with an Acclaim PepMap 100 C18 capillary column (75 μm × 150 mm, 3-μm particle size, 100-Å pore size, Dionex) and the following gradient was applied: 0% to 35% B (100 min), 35% to 50% B (15 min), 50% to 100% B (1 min), and 100% B (5 min).

The mass spectrometer was operated in positive ion mode. MS full scans (m/z 400–1800) were acquired via Fourier transform MS in the Orbitrap at a resolution of 60,000 (full width at half-maximum). The five most abundant ions of each full scan were fragmented in the linear ion trap via CID (35% normalized collision energy) using three microscans per spectrum. Dynamic exclusion was enabled with a repeat duration of 30 s, exclusion duration of 90 s, repeat count of 1, list size of 500, and exclusion mass window of ±25 ppm. Unassigned charge states and charge states of 1 were rejected.

LC-MS Analysis of Intact Glycopeptides—Instrumentation, the composition of mobile phases, and the sample loading procedure were the same as described for the analysis of 18O-labeled peptides. HPLC gradient conditions were as follows: 0% to 30% B (70 min), 30%–100% B (5 min), and 100% B (10 min).

The mass spectrometer was operated in positive ion mode with in-source (IS)-CID enabled during MS full scans and MS2 for the partial removal of glycans. The default IS voltage was 90 V. In addition, some samples were analyzed multiple times using 60 V and 80 V. MS full scans (m/z 400–2000) were acquired via Fourier transform MS in the Orbitrap at a resolution of 60,000 (full width at half-maximum). The “mass tags” option of XCalibur was enabled for the online identification of ion pairs differing by 203.0794 and 406.1587 Da. Conflicting peptide-spectrum matches were filtered on the basis of e-values. If e-values differed by 2 orders of magnitude or more, the peptide-spectrum match with the lower score was retained. Otherwise, peptide-spectrum matches were regarded as ambiguous and all corresponding identifications were discarded.

Additional Variable Modifications Used for the Identification of 18O-Labeled Peptides—Preliminary analyses led to the identification of numerous peptides derived from Con A and PNGase F, indicating high residual trypsin activity during the glycopeptide enrichment step of N-glyco-FASP. In order to take account of trypsin-mediated incorporation of 18O into the C termini of peptides potentially resulting in false glycopeptide identifications, spectra files were analyzed four times by X!Tandem, each time using a different set of variable modifications (52): (1) deamidation of asparagine in 18O-labeled water (+2.9883 Da) and single incorporation of 18O at the peptide C terminus (+2.0043 Da); (2) deamidation of asparagine in 18O-labeled water and double incorporation of 18O at the peptide C terminus (+4.0085 Da); (3) deamidation of asparagine in 18O water (+0.9840 Da) and single incorporation of 18O at the peptide C terminus (+2.0043 Da); and (4) deamidation of asparagine in 18O-labeled water (+0.9840 Da) and double incorporation of 18O at the peptide C terminus (+4.0085 Da). All results were combined, and conflicting peptide-spectrum matches were filtered on the basis of e-values. If e-values differed by 2 orders of magnitude or more, the peptide-spectrum match with the lower score was retained. Otherwise, peptide-spectrum matches were regarded as ambiguous and all corresponding identifications were discarded.

Additional Variable Modifications Used for the Identification of Intact Glycopeptides via IS-CID—X!Tandem searches were performed twice, each time with different sets of variable modifications: (1) modification of asparagine, serine, and threonine by HexNAc (+203.0794 Da); and (2) modification of asparagine by chitobiose (+406.1587 Da). Conflicting peptide-spectrum matches were filtered on the basis of e-values as described for the identification of 18O-labeled peptides. However, peptide glycosylations considered as ambiguous were not discarded automatically and were validated through manual inspection of the fragmentation spectra.

All the raw MS data have been placed in a public database repository at PeptideAtlas.

RESULTS

To comprehensively identify the N-glycosylation pathway of C. reinhardtii, different approaches including genomic, glycomic, and glycoproteomic techniques were employed.

Soluble and Membrane N-glycoproteins from C. Reinhardtii Bear Mainly Oligomannoside N-glycans—The characterization of N-glycans from C. reinhardtii was carried out on protein extracts from three different strains: CC-503 cw92, CC-1036 p18, and CC-400 cw15 (later called CC-503, CC-1036, and CC-400, respectively) (1). CC-400 and CC-503 are cell-wall-deficient strains used as references (20), whereas CC-1036, the motility of which is completely impaired, possesses a cell wall (53). Both soluble and membrane-bound proteins were isolated and separated on SDS-PAGE, trypsinized prior to the
N-glycan release using PNGase A, or directly deglycosylated using PNGase F. The resulting N-glycans were then labeled with a fluorescent tag (2-AB) before their analysis using MALDI-TOF-MS (Fig. 1). In both CC-503 and CC-1036 strains, N-glycans released from soluble and membrane-bound proteins showed identical profiles regardless of whether PNGase A or PNGase F was used. As presented in Fig. 1, the N-glycan processing does not depend on the strains or on the final destination of the secreted proteins. Moreover, the absence of cell wall does not influence the N-glycan processing. Based on the m/z values of [M+Na]+ ions, major ions were assigned to 2-AB derivatives of hexose-5 N-acetylglucosamine2 (Hex2-5GlcNAc2) (Fig. 1; supplemental Table S1). Traces of larger oligomers up to Hex3GlcNAc2 were also detected (Table I, supplemental Table S1). The pool of 2-AB-labeled N-glycans was subjected to exo-glycosidase digestion with jack bean α-mannosidase. Consistent with the presence of α-linked mannose residues, ions corresponding to Hex3-5GlcNAc2 were converted into single species corresponding to a HexGlcNAc2 composed of a β-mannose linked to the chitobiose unit (supplemental File S2). These results are in agreement with the affinodetection with Con A, a lectin specific for oligomannoside N-glycans (supplemental File S3A). All these data allow us to assign these ions to oligomannoside N-glycans ranging from Man3GlcNAc2 to Man5GlcNAc5, as previously reported for other eukaryotes (54). For confirmation, HPAEC-PAD of the N-glycan pool showed that Man-5 from C. reinhardtii had the same elution time as a commercially available standard of Man-5 (not shown).

Complex N-glycans in C. Reinhardtii Carry Xylose Residues and Are Partially Methylated—Remaining minor ions (indicated by lowercase letters in Fig. 1 and Table I) were assigned to 2-AB-labeled complex N-glycans. Most of these ions were resistant to jack bean α-mannosidase digestion (supplemental File S2), suggesting the absence of free terminal α-mannose residues in these oligosaccharides. In order to determine the monosaccharide composition of these complex-type N-linked glycans, 2-AB-labeled N-glycans were hydrolyzed and monomers were converted into alditol acetates prior to their analysis via GC-EIMS. Mannose was identified as the main monosaccharide, along with low amounts of xylose. A 6-O-methyl hexose was also detected on the basis of its fragmentation pattern in EIMS. A search for specific monosac-
| Identified oligomannoside N-glycans | Relative proportion +/- s.d. | Identified complex N-glycans | Relative proportion +/- s.d. |
|-----------------------------------|-------------------------------|-------------------------------|-------------------------------|
| ![Oligomannoside N-glycan](image1) | 15.4% +/- 2.8%                | ![Complex N-glycan](image2)   | 3.9% +/- 1.3%                 |
|                                   | 17.2% +/- 0.6%                | ![Complex N-glycan](image3)   | 5.6% +/- 0.8%                 |
|                                   | 13.0% +/- 1.3%                | ![Complex N-glycan](image4)   | 3.1% +/- 0.3%                 |
|                                   | 22.9% +/- 0.6%                | ![Complex N-glycan](image5)   | 6.3% +/- 1.3%                 |
|                                   | 0.8% +/- 0.0%                 | ![Complex N-glycan](image6)   | traces                       |
|                                   | traces                        | ![Complex N-glycan](image7)   | traces                       |
|                                   | traces                        | ![Complex N-glycan](image8)   | traces                       |
|                                   | traces                        | ![Complex N-glycan](image9)   | traces                       |
|                                   | traces                        | ![Complex N-glycan](image10)  | traces                       |

Relative percentages are the mean of five independent N-glycan preparations and analyses. The quantification was run on permethylated 2-AB-labeled N-glycans. Oligomannoside N-glycans accounted for almost 70% of the N-glycan population, whereas the complex-type N-glycans substituted by one or two pentose residues represented 14.1% and 16.6%, respectively. The symbols used are the ones adopted by the Consortium for Functional Glycomics (101). ■, N-acetylglycosamine; ○, mannose; ★, xylose; ▲, fucose.
charides such as sialic acid was also carried out. As illustrated in supplemental File S4, neither Neu5Ac nor Neu5Gc was detected with HPAEC-PAD. Complementary Western blotting using a sialic acid specific lectin such as Sambucus nigra lectin did not reveal any sialic acid modified glycoproteins (not shown). Altogether, these results indicate that C. reinhardtii proteins do not carry any detectable sialic acid.

Among the 2-AB-labeled complex N-glycans (Figs. 1 and 2), most of the species exhibited differences of 132 or 264 Da, which could correspond to the presence of one or two pentose residues on the oligosaccharides (Fig. 2). Traces of a fucosylated N-glycan (Table I) were also detected, but due to its very low amount, this glycan could not be investigated further. The pentose residue is likely to be xylose, based on the monosaccharide composition. Moreover, some complex N-glycans exhibited mass differences of 14 Da (Figs. 2A and 2B). This shift may have resulted either from the substitution of xylose by a deoxyhexose residue or from the methylation of the N-glycan. To discriminate between these two possibilities, the pool of 2-AB-labeled N-glycans isolated from CC-503 soluble proteins was then permethylated and analyzed via MALDI-TOF-MS. The MS1 profile comparison of native and permethylated 2-AB-labeled N-glycans is shown in Fig. 2. N-glycans that had previously displayed a mass difference of 14 Da were converted into unique derivatives corresponding to permethylated 2-AB-labeled Man₅Xyl₂GlcNAc₂ and Man₆Xyl₂GlcNAc₂, which indicated that the 14-Da mass shifts in the native oligosaccharides resulted from the partial methylation of the N-glycans (Figs. 2C and 2D), with the single exception of the ion at m/z 1331, which was converted into m/z derivatives of 1653 and 1667, corresponding to the dixylosylated glycan E and to Man₅GlcNAc₂ bearing both fucose and xylose residues, respectively.

Xylose Residues Are Located on a Terminal Mannose Residue and on the Core β-mannose of the Complex N-glycans—in order to precisely determine the position of the xylose residues on C. reinhardtii complex N-glycans, tandem mass spectrometry was carried out on permethylated 2-AB-labeled N-glycans from CC-503 soluble proteins (Fig. 3). For example, MS2 fragmentation of the precursor ion at m/z 2062.0, which corresponded to the sodiated adduct of the permethylated derivative of Man₅Xyl₂GlcNAc₂, yielded product ions at m/z 462.4 and 707.5, which resulted from the fragmentation of the 2-AB-labeled chitobiose unit (Fig. 3A). No product ion was found that could support the notion that
the xylose residue was linked to the proximal GlcNAc of the chitobiose unit. In contrast, the presence of characteristic product ions at $m/z$ 735.5 and 939.6 indicated the location of one xylose on the core $\beta$-mannose (Fig. 3A). Indeed, these two ions were shown to result from the core $\beta$-mannose’s cross-ring fragmentations $^{1,5}X_2$ ($m/z$ 735.5) and $^{2,5}X_2$ ($m/z$ 939.6) by Domon and Costello (55). The presence of these two ions can be explained by the substitution of the C2 of the core $\beta$-Man with a xylose residue. Because C. reinhardtii proteins were also immunodetected with specific core $\beta$(1,2)-xylose antibodies (supplemental File S3B), we concluded that this xylose is $\beta$(1,2)-linked to the core $\beta$-Man, as has been demonstrated in land plants (56).

Moreover, the presence of the ion at $m/z$ 401.3 (B3) indicated the location of the other xylose on a terminal mannose residue (Fig. 3A). Additionally, the presence of the ion at $m/z$ 335.3 showed that the C4 of a terminal mannose was substituted. Based on this MS2 information and the resistance to the $\alpha$-mannosidase treatment, the second xylose is proposed to be located on a terminal mannose residue. The same conclusions could be drawn from the MS2 analysis of the permethylated Man$_4$Xyl$_2$GlcNAc$_2$ and Man$_3$Xyl$_2$GlcNAc$_2$.

Terminal Mannose Residues Are Methylated in Complex N-glycans—The position of the methyl groups on C. reinhardtii complex N-glycans was further investigated by MS2 fragmentation on native 2-AB-labeled N-glycans. MS2 analysis of precursor ions at $m/z$ 1683.6, corresponding to the Man$_5$Xyl$_2$GlcNAc$_2$ modified by three methyl groups (Figs. 3B and 3C), revealed product ions at $m/z$ 364.2 and 567.2, which corresponded, respectively, to one and two $\text{N}$-acetylglucosamine residues linked to the 2-AB. The presence of these product ions suggested that the methylation did not occur on the chitobiose unit. Furthermore, the presence of the specific product ions at $m/z$ 1199.5 and 1507.6 indicated the presence of two methyl groups on two different outer mannose residues. The product ion at $m/z$ 1023.4 was attributed to the triple fragmentation $Y_{3\alpha}/Y_{4\alpha}/Y_{4\beta}$ (55) of the N-glycan according to its low intensity, signifying that the two inner mannose residues were not methylated. Consequently, the three methyl groups were assigned to the three outer mannose residues that were supposed to be linked to the C6 of each residue based on the GC-EIMS data. Those results linked to the presence of the ion at $m/z$ 1551.5 confirmed that the methylation occurred on the mannose residues rather than on the xylose (Figs. 3B and 3C).

Relative quantitation based on the intensity of the ions corresponding to the 2-AB-labeled permethylated N-glycans revealed that oligomannosidic N-glycans represented almost 70% of the total N-glycan population (Table I). Although these analyses gave rise to thorough insights into N-glycan structures, no information about the proteins that carry these post-translational modifications can be inferred. In order to shed light on N-glycoprotein identities, glycoproteomic studies were conducted.

Glycoproteomic Analyses Led to the Identification of 135 Glycopeptides and Confirmed Hexose Methylation—In order to obtain information regarding the identity and cellular distribution of glycoproteins in C. reinhardtii, a glycoproteomic approach was performed using proteins from TCEs, plasma
membranes, chloroplasts, and the culture SN expressed under iron-sufficient and -deficient conditions. A detailed workflow scheme depicting protein isolation and glycoprotein/glycopeptide enrichment strategies is presented in supplemental File S1.

Two complementary methods were used to identify glycoproteins and glycosylation sites. For the first approach, mannosose-rich glycopeptides were enriched using Con A and then samples underwent PNGase F–mediated deglycosylation in the presence of H218O. N-glycosylation site occupancy was then detected by a mass increase of +2.9883 Da of precursor and fragment ions induced by the deamidation of modified asparagine residues upon glycan hydrolysis (N-glyco-FASP) (16, 57). The second approach, unbiased by the mode of core fucosylation, focused on the analysis of intact glycopeptides. Here, IS-CID was applied for the partial removal of glycan structures. IS-CID leads to the fragmentation of glycosidic bonds while preserving the integrity of the peptide backbone. Therefore, IS-CID in combination with high-resolution spectrum acquisition allows for the detection of glycopeptides on the MS1 level through the detection of ion pairs that differ in mass by a single carbohydrate residue. These ions were determined “on the fly” through activation of the mass tag option of XCalibur with a mass setting of one N-acetylglucosamine (HexNAc) unit (203.0794 Da). Subsequently, all ions potentially differing by one HexNAc residue were sequentially isolated and fragmented via MSA-CID in order to obtain peptide sequence information. Because IS-CID affects all glycans irrespective of their linkage types, measures were taken to rule out that peptides modified by O-glycans were falsely identified as N-glycosylated. Firstly, peptide identification by X!Tandem was performed using the modification of serine and threonine by HexNAc as an additional variable parameter, thereby creating a competitive environment in terms of peptide-spectrum matching and scoring. Secondly, all glycopeptide identifications and glycosylation sites were validated through manual inspection of MS2 fragmentation spectra. However, IS-CID-MS1 spectra do not provide information regarding the isomeric nature of carbohydrates, nor do they allow for the determination of linkage types. Although the identification of oligomannoside N-glycans via IS-CID is facile, ion signals of branched and multiply N-glycosylated peptides often are highly ambiguous, and caution must be exercised with respect to spectrum interpretation.

A total of 134 distinct glycopeptides corresponding to 137 glycosylation sites from 86 proteins were identified (Table II). Using the PNGase F/18O-method and IS-CID, we identified 124 and 31 glycopeptides, respectively, with an overlap of 21 glycopeptides. Through the 18O-method, six additional glycosylation sites were found that did not match the consensus motif N[XIP][S/T], suggesting spontaneous deamidation during incubation in the presence of H218O. The IS-CID method yielded a considerably lower number of glycopeptide identifications than the PNGase F/18O-approach. This might have been due to the generally low ionization efficiency of glycopeptides and the weak retention of glycopeptides on the reversed-phase trap column. Moreover, when intact glycopeptides are being analyzed, the presence of glycoforms of distinct glycopeptides exhibiting slightly shifted retention times may lead to peak spreading and ultimately to signal intensities too low for detection (58).

The number of detected glycoproteins differed considerably among the cell fractions analyzed (supplemental File S5). The majority of N-glycosylated proteins (62) were identified in the culture medium (SN). Only a few of these secreted proteins are functionally annotated in the JGI 4.3 Augustus 10.2 gene model database, but conserved domain searches indicated that most proteins are related to protein lysis, cell wall degradation, and carbohydrate binding.

Among secreted N-glycosylated proteins of C. reinhardtii cultivated under iron-deficient conditions, FEA1 and FEA2 were identified. These are two related proteins that have been proposed as members of the iron uptake pathway (59). All glycosylation sites of FEA1/2 were detected independently with both the PNGase F method and IS-CID. In the same fraction, another N-glycosylated protein (MnSOD3) that was previously proposed to play an important role in iron-deficiency responses was also identified (60).

The highest numbers of glycosylation sites were determined for the flagellar proteins PKHD1–1 and FMG1–1/2. PKHD1–1 is a close homolog of human polycystic kidney and hepatic disease 1-like 1 protein (PKHD1L1) (BLASTp E-value = 0.0, 28% identity). An alignment of the peptide sequences of algal and human proteins revealed that four glycosylation sites were highly conserved (supplemental File S6). The IS-CID-MS1 spectrum of the PKHD1–1 glycopeptide TITVANN*GTHSTATILK showed considerable clustering of glycopeptide ions, which can be attributed to extensive glycan heterogeneity (Fig. 4; supplemental File S7). Within individual clusters and as observed in the total N-glycan profile (Fig. 1), several ions exhibited a difference of 14 Da, suggesting either partial glycan methylation or the presence of xylose and fucose. In fact, the IS-CID-MS1 spectrum indicated that both possibilities applied. However, MALDI-TOF analyses of released N-glycans clearly suggested that clustering was predominantly caused by the presence of methylated hexoses. Moreover, no ions exhibiting a mass difference of 291.0954 Da corresponding to sialic acid were present in the IS-CID-MS1 spectra of PKHD1–1. This was also true for all other MS1 spectra containing glycopeptide ions with identities confirmed by means of MSA-CID. Figs. 4B and 4C show MSA-CID spectra of the peptide TITVANN*GTHSTATILK modified by one and two HexNAc residues, respectively. N-glycosylation sites of PKHD1–1 are not evenly distributed but in three clusters. Cluster 1 (342–519) contains four N-glycosylation sites, whereas cluster 2 (1944–1967) and cluster 3 (2671–2716) contain two N-glycosylation sites each.
| Identifier (JGI 4.3 Augustus 10.2) | Protein name or conserved domain | Peptide | Source | Method | m/z (#HexNAc) | z | E-value |
|-----------------------------------|---------------------------------|---------|--------|--------|---------------|---|--------|
| Cre07.g340450.11.2 {PKHD1–1}     | G8 domain (found in disease proteins PKHD1 and KIAA1199) | TSWSATWSNGSSAEYFK | SN | ^{16}O | 1012.9609 | 2 | 1.2e-14 |
|                                  |                                 | GVEYEFYVSLGVSNLWR | IS | ^{16}O | 1113.0093 (1) | 2 | 6.3e-10 |
|                                  |                                 | FYMOADDGQNMYTDQTVVR | SN | ^{16}O | 1462.1926 | 2 | 5.8e-12 |
|                                  |                                 | FTVMQFSNNTAHSMFGLR | SN | ^{16}O | 832.7168 | 3 | 1.5e-03 |
|                                  |                                 | TTVAMNGTHSATILK | SN | ^{16}O | 973.0201 (1) | 2 | 1.6e-03 |
|                                  |                                 | TSGGPSAGNNTGIGSR | SN | ^{16}O | 959.9818 (1) | 2 | 4.9e-08 |
|                                  |                                 | YLYGAGANTK | IS | ^{16}O | 818.3947 (2) | 2 | 5.5e-03 |
|                                  |                                 | TSDALNTDTPTAFTWNPNTVR | SN | ^{16}O | 1.383.6801 | 2 | 1.9e-09 |
| Cre06.g279700.11.2              | G8 domain (found in disease proteins PKHD1 and KIAA1199) | TSWSATWSNGSSAEYFK | SN | ^{16}O | 1012.9609 | 2 | 1.2e-14 |
| Cre16.g676150.11.1 (MnSOD3)    | Mn superoxide dismutase | WGNATALLSLR | SN | ^{16}O | 660.3345 | 2 | 4.5e-04 |
| Cre17.g718500.11.2 {MMP1}      | Matrix metalloproteinase, gamete lytic enzyme (G-lysin) | KNTTAGYDSGLTDFHK | IS | ^{16}O | 1021.0128 | 2 | 3.3e-15 |
| Cre17.g718488.11.1 (MMP2)    | Peptidase M11 superfamily domain (Gametolysin) | KNTTAGYDSGLTDFHK | IS | ^{16}O | 1121.0601 | 2 | 3.8e-07 |
| Cre17.g718350.11.1 (MMP11)   | Peptidase M11 superfamily domain (Gametolysin) | LLVHEVEVPATMDNLQLYR | SN | ^{16}O | 1073.5469 | 2 | 9.4e-15 |
| Cre07.g324500.11.1             | Peptidase M11 superfamily domain (Gametolysin) | TMVLYHSVYNGGTAVSYQR | SN | ^{16}O | 672.6630 | 3 | 5.9e-04 |
| Cre02.g133500.11.2             | Peptidase M11 superfamily domain (Gametolysin) | VLVHFINQGSASER | SN | ^{16}O | 733.3726 | 2 | 1.5e-17 |
| Cre13.g596600.11.1             | Peptidase M11 superfamily domain (Gametolysin) | VYFVHEFNETADNKPSQDQNPPLIMAVLVDVK | SN | ^{16}O | 1129.2257 | 3 | 9.3e-10 |
| Cre13.g596750.11.1             | Peptidase M11 superfamily domain (Gametolysin) | IFIHEFNETADNPDTDSYPPLUR | SN | ^{16}O | 941.1145 | 3 | 8.4e-03 |
| Cre14.g625850.11.2             | Peptidase M11 superfamily domain (Gametolysin) | IYHNFINATLR | SN | ^{16}O | 682.8665 | 2 | 6.0e-05 |
| Cre60.g792000.11.1             | Matrix metalloproteinase | WWVHEVNETAGLTANL | SN | ^{16}O | 1031.5098 | 2 | 1.5e-11 |
| Cre01.g011300.11.1             | Serine carboxypeptidase domain | GFTINATGIIATMDFTDR | SN | ^{16}O | 859.9225 | 2 | 8.8e-09 |
| Cre05.g242750.11.2             | Multiple peptidase S8 family domains | NPDIAAFANQNGSDALTPGGSGTPAK | SN | ^{16}O | 1444.7141 | 2 | 6.8e-11 |
| Cre12.g513400.11.2             | Multiple glycosyl hydrolase family 81 domains | NISITAEFGHSV | SN | ^{16}O | 684.3572 | 2 | 5.5e-10 |
| Cre09.g400850.11.2             | Carbohydrate binding domains (FS/8 type C, WSC, C- and H-type lectin) | LNLMSDIVSMGR | SN | ^{16}O | 1301.1613 | 2 | 1.3e-09 |
| Cre06.g309950.11.2             | Multiple C-type lectin (CTL/C-type lectin-like (CTLD) domains) | YLVTFIDNATYAR | SN | ^{16}O | 775.3936 | 2 | 5.1e-09 |
| Cre05.g245259.11.1             | Multiple C-type lectin-like domain (CTLD), GHT18 chitinase-like domains | TFTOLSPLDLAGPSFYIDTNTTTR | SN | ^{16}O | 1468.2206 | 2 | 1.0e-15 |
|                                  |                                 | QQVTVGQSTTIR | SN | ^{16}O | 748.3880 | 2 | 3.4e-05 |
|                                  |                                 | SGAVFQGSQVQPSLIPPASIGR | SN | ^{16}O | 1279.6560 | 2 | 6.8e-10 |
Table II—continued

| Identifier (JGI 4.3 Augustus.10.2) | Protein name or conserved domain | Peptide | Source | Method | mz # HexNac | z | E-value |
|-----------------------------------|----------------------------------|---------|--------|--------|-------------|---|---------|
| (a) Cre14.g631100.t1.11.2         | (a) Carbohydrate binding domains (F3/8 type C, WSC, C-type lectin), scavenger receptor Cys-rich domain, peptidase C1A domain | VGNASVTSSTDLSLYGNTLVWκ | SN | 18O | 1101.5472 | 2 | 2.1e-11 |
|                                   | (b) Peptidase C1 superfamily domain, WSC domain | LALQPSSLLFNGSGAEEWκ | SN | 18O | 949.4857 | 2 | 2.4e-03 |
| Cre04.g226500.t1.11.2             | Sulfatase superfamily domain | LNQLFLSSDEAEVNDLNLK | SN | 18O | 1139.5939 | 2 | 1.8e-13 |
| Cre04.g226600.t1.11.2             | Sulfatase superfamily domain | SDPKNFVITTDGGDLNITHPPYMALNR | SN | 18O | 902.9440 | 4 | 4.0e-05 |
| Cre10.g432600.t1.11.2             | Sulfatase superfamily domain | YTHNWWTSNPEHPGWSK | SN | 18O | 750.0142 | 3 | 1.2e-04 |
| Cre10.g431800.t1.11.2             | Sulfatase | ALPNATLWGDTFGTAAR | SN | 18O | 984.8986 | 2 | 1.4e-07 |
| Cre02.g097000.t1.11.1             | Dihydropyrimidinase domain | VIGPASQLADSEPVDSNSFTR | SN | 18O | 1281.6359 | 3 | 3.4e-14 |
| Cre01.g028850.t1.11.1             | Rhodanese homology domain (Rhod) | NITTFDLIR | SN | 18O | 491.7594 | 9 | 9.8e-05 |
|                                   |                                  | TEANFTASHAGAVNPK | IS | 18O | 615.3191 | 3 | 3.5e-07 |
| Cre02.g077750.t1.11.1 [FAP211]    | Flagellar associated protein (a) No conserved domains | APTILVSAADALLLDDKNITTFDLIR | SN | 18O | 935.1729 | 3 | 7.4e-04 |
|                                   | (b) No conserved domains | SGYVAQAVGLTRNPILRL | SN | 18O | 640.3583 | 3 | 2.7e-05 |
|                                   |                                  | GGQYHAGP/GFTPNLMTDPR | SN | 18O | 794.4139 | 3 | 3.4e-07 |
|                                   |                                  | NMTQGQALTPAPGLWYDSPNFAANNK | SN | 18O | 1447.2091 | 2 | 4.9e-13 |
| (a) Cre16.g661750.t1.1            | CaMK II association domain (a) No conserved domains | WLVIEHHSSAMPENALVMADFQWQNDALATLNASK | SN | 18O | 1028.7544 | 4 | 5.2e-03 |
| (b) Cre16.g661850.t1.2            | CaMK II association domain | NTSHTDDYAAALNVTAR | SN | 18O | 891.4289 | 3 | 1.3e-11 |
| Cre13.g569550.t1.11.2             | Polycystin cation channel protein domain, leucine-rich repeat receptor-like protein kinase domain | LVSSGQNTALDAAR | SN | 18O | 738.8793 | 2 | 5.6e-07 |
|                                   |                                  | DSWLVFGGDNDNFR | SN | 18O | 940.4707 (2) | 2 | Manκ |
| Cre01.g062570.t1.11.2             | TRP superfamily domain, galactose oxidase (central domain) | SQHTALPYENSLVFQGGSNK | SN | 18O | 765.7286 | 3 | 4.6e-05 |
| Cre45.g788400.t1.11.1             | DUF3707 (Phosphokinase) domain | ETSQAFMWLQLNR | SN | 18O | 812.4184 | 2 | 7.1e-05 |
| Cre11.g76250.t1.11.2              | C3HC4-type RING-finger domain | GAQPVNVTYSGQYYAPPR | SN | 18O | 978.9732 | 2 | 2.3e-12 |
| Cre06.g266050.t1.11.2             | DUF288 family domain | GFFPOITQNYVVVR | SN | 18O | 725.8926 | 2 | 2.2e-09 |
| Cre13.g598600.t1.11.1 [FAST]     | Multiple fasciclin domains | OWQEFALPGYNYTIK | SN | 18O | 875.4777 | 2 | 1.8e-07 |
| Cre16.g694400.t1.11.1             | No conserved domains | VWHSHESETAGAPAAGPHTLIR | SN | 18O | 835.4196 | 3 | 3.8e-07 |
| Cre09.g398900.t1.11.2             | No conserved domains | TVGSVANVTIR | SN | 18O | 560.3170 | 2 | 5.1e-04 |
| Cre06.g258800.t1.11.2             | No conserved domains | NITWTAPAGITTIR | SN | 18O | 819.4330 (2) | 2 | 5.4e-04 |
| Cre12.g487950.t1.11.2             | No conserved domains | ISDGVSTYSEPFDR | SN | 18O | 901.9235 | 2 | 1.2e-08 |
| Cre09.g401050.t1.11.2             | No conserved domains | AGQNYTFSNCMETYNGAR | SN | 18O | 947.9026 | 2 | 2.7e-04 |
| Cre02.g080150.t1.11.2             | No conserved domains | VLGELNATSWTEAGAELAPLEAVATAAR | SN | 18O | 1535.3008 | 2 | 3.8e-10 |
| Cre02.g122550.t1.11.1             | No conserved domains | FYFNWITVR | SN | 18O | 649.7910 | 2 | 7.3e-04 |
| Cre06.g308050.t1.11.1             | No conserved domains | LANGTVDGSPAYFSR | SN | 18O | 786.3850 | 2 | 4.4e-05 |
| Cre17.g070070.t1.11.2             | No conserved domains | RPPDINATAGLVDGSYHINDR | SN | 18O | 719.0178 | 3 | 1.4e-05 |
| Cre07.g333100.t1.11.2             | No conserved domains | VDLGDSYELNNFSK | SN | 18O | 819.3798 | 2 | 5.5e-09 |
| Cre17.g070850.t1.11.1             | No conserved domains | GGAIVLNOTDTSK | SN | 18O | 905.9520 | 2 | Manκ |
| Identifier (JGI 4.3 Augustus10.2) | Protein name or conserved domain | Peptide | Source | Method | \( m/z \) (# HexNAc) | E-value |
|----------------------------------|----------------------------------|---------|--------|--------|----------------------|---------|
| Cre02.g12115011.2 (CrSTT3A)     | Oligosaccharidyltransferase STT3 subunit | TVIDNTWNTSHIATVGR | TCE      |        | 1051.0349 2 | 1.5e-10 |
| Cre07.g33010011.1 (CrSTT3B)     | COG1287 (uncharacterized membrane protein, required for N-linked glycosylation) | VASWWDYGQTAMANRFT | TCE      |        | 1011.9531 2 | 7.3e-15 |
| Cre05.g23325011.2 (CrUGGC)      | UDP-glucose:glycoprotein glucosyltransferase | TVIDNTWNTSHIATVGR | TCE      |        | 1047.9688 2 | 6.7e-11 |
| Cre05.g23325011.2 (CrUGGC)      | UDP-glucose:glycoprotein glucosyltransferase | VASWWDYGQTAMANRFT | TCE      |        | 710.0291 3 | 3.3e-03 |
| Cre07.g33075011.2 | DUF3707 (Phosphorin) domain | ATIAASTFG | PM | \( ^{18}O \) | 546.8153 2 | 1.1e-03 |
| Cre10.g46330011.1 | No conserved domains | NMTISGPLDSPIK | PM | \( ^{18}O \) | 688.3521 2 | 1.4e-04 |
| Cre17.g70550011.2 | Lysosomal cystine transporter domain, PQ loop repeat domain | ISLASDGFGV/NTATYNTGTYAIL GAK | PM | \( ^{18}O \) | 1255.6277 2 | 1.0e-15 |
| Cre05.g23810011.1 | Protein tyrosine kinase | NMTISGPLDSPIK | PM | \( ^{18}O \) | 1393.7101 2 | 3.0e-09 |
| Cre02.g09005011.2 | Flagellar associated protein | LLGGSDLVTVYSQIV GLGRS | PM | \( ^{18}O \) | 818.4133 2 | 2.6e-05 |
| Cre07.g71290011.1 | No conserved domains | LEQSLG | PM | \( ^{18}O \) | 637.3177 3 | 1.8e-03 |
| Cre17.g72250011.1 | DUF3707 (Phosphorin) domain | ISLASDGFGV/NTATYNTGTYAIL GAK | PM | \( ^{18}O \) | 1246.3956 4 | 3.8e-03 |
| Cre05.g24135011.2 | Eye pigment and drug resistance transporter superfamily domain (ABCG_EPDOR), ABC-type cobalt transport system (CbiO) | LFGVPVASAYGTAVQLLAYDYYVK | PM | \( ^{18}O \) | 546.8153 2 | 1.1e-03 |
| Cre09.g40900011.2 | Fe-assimilating protein | TNGCSVEDQDLERIFAPETLSVYK | CP | \( ^{18}O \) | 953.8032 3 | 2.3e-05 |
| Cre12.g54655011.1 | Flagellar membrane glycoprotein, major form | LFGVPV/NTATYNTGTYAIL GAK | CP | \( ^{18}O \) | 1363.2107 2 | 1.1e-12 |
| Cre09.g39315011.1 | Multicopper ferredoxase | TNGCSVEDQDLERIFAPETLSVYK | PM | \( ^{18}O \) | 1463.2527 1 (1) | 1.3e-10 |
| Cre12.g54660011.1 | Fe-assimilating protein | LFGVPV/NTATYNTGTYAIL GAK | PM | \( ^{18}O \) | 1317.1770 1 (1) | 3.9e-08 |
| Cre09.g39315011.1 | Multicopper ferredoxase | LFGVPV/NTATYNTGTYAIL GAK | PM | \( ^{18}O \) | 982.9630 2 | 1.6e-12 |
| Cre12.g54660011.1 | Fe-assimilating protein | FASYTANGSVEPLHDSILAGK | TCE | \( ^{18}O \) | 1083.0070 2 | 9.5e-08 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE | \( ^{18}O \) | 643.8174 2 | 4.7e-10 |
| Cre12.g54660011.1 | Fe-assimilating protein | FASYTANGSVEPLHDSILAGK | TCE | \( ^{18}O \) | 743.8636 1 (1) | 8.8e-04 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE | \( ^{18}O \) | 881.8932 2 | 1.0e-08 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE | \( ^{18}O \) | 1265.2036 1 (1) | 2.1e-15 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE | \( ^{18}O \) | 747.3456 2 | 1.5e-07 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE | \( ^{18}O \) | 983.4730 2 | 5.0e-11 |
| Cre12.g54655011.1 | Fe-assimilating protein | FASYTANGSVEPLHDSILAGK | TCE, SN | \( ^{18}O \) | 1147.0874 2 | 5.7e-08 |
| Cre12.g54660011.1 | Fe-assimilating protein | FASYTANGSVEPLHDSILAGK | TCE, SN | \( ^{18}O \) | 1247.1337 2 | 3.9e-14 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE, SN | \( ^{18}O \) | 1147.0874 2 | 5.7e-08 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE, SN | \( ^{18}O \) | 1247.1337 2 | 3.9e-14 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE, SN | \( ^{18}O \) | 1147.0874 2 | 5.7e-08 |
| Cre09.g39315011.1 | Multicopper ferredoxase | FASYTANGSVEPLHDSILAGK | TCE, SN | \( ^{18}O \) | 1247.1337 2 | 3.9e-14 |
| Identifier (JGI 4.3 Augustus10.2) | Protein name or conserved domain | Peptide | Source | Method | $\text{m/z}$ (# HexNAc) | z | E-value |
|---------------------------------|---------------------------------|---------|-------|--------|------------------------|--|---------|
| Cre10.g439900.t1.1 (HSP70G)    | ER-located HSP110/SSE-like protein | YNTSGQISLR | TCE   | $^{18}$O | 571.2894  | 2  | 1.7e-06 |
|                                 |                                 | HILDAEAVALGAGFLAANLTSFR | TCE   | $^{18}$O | 1225.1179 | 2  | 1.0e-15 |
|                                 |                                 | FISLAVYNNASTHHLGLPPGW | TCE, SN<sup>a</sup> | $^{18}$O | 949.9884  | 2  | 1.6e-09 |
|                                 |                                 | IIEVPVENETDTATGAEGAGADADTK | TCE, SN, CP | $^{18}$O | 1224.5840 | 2  | 1.0e-15 |
|                                 |                                 | TCE, SN, CP | IS    | $^{18}$O | 1488.7189 (1) | 2  | 6.6e-09 |
|                                 |                                 | AQQGDKAEAEAEAGKEADAGAGSANNASTNA, NAKPATVK | SN<sup>a</sup> | $^{18}$O | 1002.7274 | 2  | 4.1e-08 |
|                                 |                                 |                      |       |        |            |   |         |
| (a) Cre06.g272250.t1.1          | (a), (b) DUF3707 (Pherophorin) domain | IMAFNTTPATDYSK<sup>c</sup> | TCE   | $^{18}$O | 781.8712  | 2  | 1.3e-04 |
| (b) Cre03.g155300.t1.1          |                                 | KLYNSYSPFDGPEK<sup>c</sup> | TCE   | $^{18}$O | 838.3884  | 2  | 6.1e-09 |
|                                 |                                 | LGYQLGSTVAPNGTMFNDLAK<sup>c</sup> | TCE   | $^{18}$O | 1159.0795 | 2  | 1.3e-11 |
|                                 |                                 | SVNVTGYIDDLDEALCDY<sup>c</sup> | CP    | $^{18}$O | 1107.0132 | 2  | 3.1e-11 |
|                                 |                                 |                      | TCE, SN, CP | $^{18}$O | 1051.9949 (1) | 2  | 6.6e-09 |
|                                 |                                 |                      |       |        |            |   |         |
| Cre17.g741000.t1.1 (TRAP1)     | Translocon-associated protein (TRAP), alpha subunit | YNSAMVGSLLNAGDFR | TCE, SN, CP | $^{18}$O | 951.9250  | 2  | 3.1e-11 |
|                                 |                                 |                      |       |        |            |   |         |
| Cre17.g715300.t1.1 (PKD2)      | Polycystin cation channel protein | GTTLYNPDYDDNGTK | TCE, PM, SN | $^{18}$O | 888.3962  | 2  | 6.7e-12 |
| Cre13.g592050.t1.1 (DAL1)      | Allantoicase | IV/AVSDFPVTIPSHMLNFK | TCE   | $^{18}$O | 697.6844  | 3  | 1.5e-03 |
|                                 |                                 | VEWEKIGWNGTR | TCE, SN<sup>a</sup> | $^{18}$O | 675.3238  | 3  | 3.8e-04 |
|                                 |                                 |                      | TCE, PM | $^{18}$O | 999.5161  | 2  | 4.6e-06 |
|                                 |                                 |                      | TCE, SN, CP | $^{18}$O | 1356.1448 (1) | 2  | 2.3e-04 |
|                                 |                                 |                      |       |        |            |   |         |
| Cre06.g294400.t1.2             | Nicastrin | TLHLSAGSPETPALEV<sup>a</sup> | PM    | $^{18}$O | 632.3205  | 3  | 1.4e-03 |
|                                 |                                 |                      |       |        |            |   |         |
| (a) Cre01.g067150.t1.1          | (a), (b) EGF-like domain | ITTSDLSPLDSFAPNFLTDK | TCE, SN | $^{18}$O | 699.3804  | 2  | 6.9e-10 |
| (b) Cre01.g067059.t1.1          |                                 |                      |       |        |            |   |         |
| Cre45.g788350.t1.2 (GAS31)     | Hydroxyproline-rich cell wall protein pherophorin | LSYVTNTSSLGVR | TCE, SN | $^{18}$O | 699.3804  | 2  | 6.9e-10 |
|                                  |                                 |                      |       |        |            |   |         |
| Cre06.g304500.t1.1 (ZYS3-2)    | Zygote-specific protein | LNSVQIEKHPEVQAMLADIEK | TCE, SN | $^{18}$O | 1247.1454 | 2  | 1.0e-10 |
|                                 |                                 |                      |       |        |            |   |         |
| Cre15.g635079.t1.1             | Periplasmic binding protein (PBPs) superfamily domain | NLTGNGYSGLRR | TCE, PM | $^{18}$O | 707.3672  | 2  | 1.3e-07 |
|                                  |                                 |                      |       |        |            |   |         |
| Cre05.g249490.t1.1             | Translocon-associated protein beta (TRAP beta) domain | YTDVLPEPLATVEGSLDEVFGK | TCE, CP | $^{18}$O | 1200.0963 | 2  | 1.0e-15 |
|                                  |                                 |                      |       |        |            |   |         |
| Cre01.g042550.t1.1             | DUF1620 | QTLVFDLPANGTALK | TCE, CP | $^{18}$O | 795.9345  | 2  | 1.5e-04 |
| Cre07.g321400.t1.2 (FAP113)   | Flagellar associated protein | KNVTVQPTQISNIEFK | SN    | $^{18}$O | 1024.5787 | 2  | 1.7e-07 |
|                                  |                                 |                      |       |        |            |   |         |
| Cre14.g561250.t1.2             | No conserved domains | VLDYVAYGNSDANLPAGSVSLVPLDPGA | TCE, SN<sup>a</sup> | $^{18}$O | 1501.7690 | 2  | 5.2e-10 |
|                                  |                                 |                      |       |        |            |   |         |
| Cre06.g283200.t1.1             | No conserved domains | DINGLGSSTVDLVAGK | TCE, SN<sup>a</sup> | $^{18}$O | 867.4476  | 2  | 9.7e-11 |
|                                  |                                 |                      |       |        |            |   |         |
| Cre02.g398500.t1.1             | No conserved domains | LPIATASATFDGLR | TCE, SN<sup>a</sup> | $^{18}$O | 731.8952  | 2  | 4.8e-08 |

Proteins not functionally annotated in the JGI 4.3 Augustus 10.2 assembly were analyzed using the Conserved Domains search tool provided by NCBI. Search results (E-value cutoff: $1 \times 10^{-10}$) are printed in italics. Glycosylation sites are denoted by bold letters. Members of protein groups are indicated by alphabetical prefixes. Superscript roman letters indicate the association of peptides with corresponding protein group members.

<sup>a</sup> Peptides were exclusively identified in samples derived from iron-deficient conditions.

<sup>b</sup>–<sup>h</sup> These peptides were originally identified as O-glycosylated, but manual inspection revealed N-glycosylation (original E-values: $^b 4.7e-09$, $^c 2.5e-10$, $^d 1.10e-05$, $^e 5.9e-03$, $^f 2.1e-09$, $^g 5.5e-05$, $^h 7.1e-03$).

SN, supernatant (culture medium); TCE, total cell extract; PM, plasma membrane; CP, chloroplast; $^{18}$O, PNGase F/$^{18}$O-method; IS, in-source CID; z, charge; man, manual inspection.
FIG. 4. Mass spectrometric analysis of the PKHD1–1 peptide TITVANN*GTHSTATILK. A, MS1 spectrum of doubly charged, N-glycosylated peptide ions (P) differing in glycan composition and chain length. Distinctive peak clustering suggests multiple branching of a complex-type glycan and/or the co-elution of glycoforms. The presence of $\rho$ + HexNAc = Hex and similar species indicates concomitant O-glycosylation or glycan rearrangement during fragmentation. Only major peaks are annotated. For detailed annotation of cluster 7 ranging
Six distinct glycopeptides of HSP70G corresponding to seven glycosylation sites were identified in the SN, TCE and chloroplast fractions. Of these, only one peptide (IIEVPVN*ETDTATGAEGAGADADTKAEK) was detected with IS-CID (Fig. 5). The IS-CID-MS1 spectrum (Fig. 5A) showed signals of the HSP70G peptide modified by up to two HexNAc and five hexose residues. No peak clustering was observable; thus the glycan could be unambiguously classified as oligomannoside.

Targeted fragmentation of \( m/z 1488.7174 \) via MSA-CID provided the information required for the determination of peptide from \( m/z 1389 \) to 1419, please refer to the supplemental material. Possible glycan compositions: \( X_1 \): Hex+MeHex+Pent; \( X_2 \): Hex+MeHex+Pent; \( X_3 \): Hex+MeHex+DeHex or Hex+MeHex+DeHex+Pent or MeHex+Pent; \( X_4 \): Hex+MeHex+DeHex+Pent or Hex+MeHex+DeHex+Pent or MeHex+Pent; \( X_5 \): Hex+MeHex+DeHex+Pent or Hex+MeHex+DeHex+Pent or MeHex+Pent; \( X_6 \): Hex+MeHex+DeHex+Pent or Hex+MeHex+DeHex+Pent or MeHex+Pent; \( X_7 \): Hex+MeHex+DeHex+Pent or Hex+MeHex+DeHex+Pent or MeHex+Pent; \( X_8 \): Hex+MeHex+DeHex+Pent or Hex+MeHex+DeHex+Pent or MeHex+Pent. B, multistage-activation (MSA)-CID spectrum of TITVANN*GTHSTATILK (precursor \( m/z: 1488.7174 \) (\( MH^2+ \))) modified by one HexNAc residue. Fragmentation was carried out by MSA of precursor and neutral loss ions arising from the cleavage of the HexNAc residue (\( \sim 203 \) Da). The resulting composite spectrum contains several b- and y-type ions showing modifications by HexNAc. ■, HexNAc; *, loss of water or ammonia.

Fig. 5. Mass spectrometric analysis of the HSP70G peptide IIEVPVN*ETDTATGAEGAGADADTKAEK modified by a HexNAc\textsubscript{2}Hex\textsubscript{5} glycan. A, in-source CID MS1 spectrum of several doubly and triply charged peptide ions (P) differing in glycan chain length. B, CID spectrum of IIEVPVN*ETDTATGAEGAGADADTKAEK (precursor \( m/z: 1488.7174 \) (\( MH^2+ \))) modified by one HexNAc residue. Fragmentation was carried out by MSA of precursor and neutral loss ions arising from the cleavage of the HexNAc residue (\( \sim 203 \) Da). The resulting composite spectrum contains several b- and y-type ions showing modifications by HexNAc. ■, HexNAc; *, loss of water or ammonia.

Protein N-glycosylation in *Chlamydomonas reinhardtii*
Protein N-glycosylation in *Chlamydomonas reinhardtii*

| Glycoproteins conserved in *C. reinhardtii* and humans | Best BLASTp hit (human) | Accession number | BLASTp E-value | Conserved N-glycosylation sites |
|---|---|---|---|---|
| Cre07.g340450.t1.2 | PKHD1-1 | EAW91931.1 | 0.0 | 4 out of 8 |
| Cre01.g011300.t1.1 | CPVL | AAG37991.2 | 4e-52 | 1 out of 1 |
| Cre04.g226600.t1.2 | N-acetylglucosamine-6-sulfatase precursor | NP_002067.1 | 4e-52 | - |
| Cre10.g431800.t1.1 | N-acetylglucosamine-6-sulfatase precursor | NP_002067.1 | 4e-53 | - |
| Cre02.g097000.t1.1 | DPYS | NP_001376.1 | 1e-141 | - |
| Cre08.g383400.t1.2 | Deleted in malignant brain tumor 1 protein isoform | CAB56155.1 | 3e-48 | - |
| Cre02.g121650.t1.2 | STT3A/STT3B | NP_689926.1/NP_849193.1 | 0.0/0.0 | 2 out of 2 |
| Cre08.g383600.t1.2 | DMBT1/8kb.2 protein | NP_015568.2 | 4e-52 | - |
| Cre07.g330100.t1.1 | STT3A/STT3B | NP_689926.1/NP_849193.1 | 0.0/0.0 | 2 out of 2 |
| Cre09.g397900.t1.1 | CLPT1 | NP_001285.1 | 4e-140 | - |
| Cre14.g611850.t1.1 | TM9SF3 | NP_004818.2 | 0.0 | 1 out of 1 |
| Cre05.g241350.t1.2 | ABC transporter (ABCG2) | AAG52982.1 | 4e-41 | - |
| Cre09.g393150.t1.1 | FOX1 | NP_001124332.1 | 2e-147 | - |
| Cre10.g439900.t1.1 | HYOU1 | NP_006380.1 | 1e-111 | - |
| Cre06.g279700.t1.2 | Fibrocystin L | NP_803875.2 | 6e-38 | - |
| Cre01.g042550.t1.1 | ER membrane protein complex subunit 1 | NP_001258357.1 | 4e-69 | - |

BLASTp was performed using full-length amino acid sequences and an E-value cutoff of 1e**-30.** Peptide sequence alignments of candidates with conserved glycosylation sites are provided as supplemental material.

Among the glycoproteins of the TCE fraction, we found three candidates of the N-glycan pathway: CrSTT3A, CrSTT3B, two subunits of the OST complex, and CrUGGT, an UDP glucosyl:glycoprotein glucosyltransferase involved in the ER quality control of neosynthesized glycoproteins. As shown in supplemental File S9, the glycosylation sites of CrSTT3A and CrSTT3B are highly conserved among eukaryotic organisms. BLASTp analyses revealed that 14 N-glycosylated proteins from *C. reinhardtii* exhibited high sequence similarity to human proteins (E-value cutoff: 1e**-30.** Tables III and IV). Based on these results, peptide sequence alignments were repeated using ClustalW2, which led to the identification of five human proteins with either potential (PKHD1L1, CPVL) or repeated using ClustalW2, which led to the identification of five human proteins with either potential (PKHD1L1, CPVL) or

Taking advantage of the sequenced genome (20) and based on sequence similarity to genes encoding enzymes involved in these ER steps, we identified most of the enzymes involved in the biosynthesis of dolichol pyrophosphate-linked oligosaccharide donor Glc3Man9GlcNAc2-PP-dolichol and its transfer by the OST onto asparagine residues of proteins in the lumen of the rough ER. Then, this precursor is deglucosylated by glucosidases I and II and reglucosylated by UGGT to ensure its interaction with chaperones responsible for protein folding.

In *Silico* Analysis of the Chlamydomonas Reinhardtii Genome—In eukaryotes, the N-glycan pathway starts with the biosynthesis of the dolichol pyrophosphate-linked oligosaccharide donor Glc3Man9GlcNAc2-PP-dolichol and its transfer by the OST onto asparagine residues of proteins in the lumen of the rough ER. Then, this precursor is deglucosylated by glucosidases I and II and reglucosylated by UGGT to ensure its interaction with chaperones responsible for protein folding. Taking advantage of the sequenced *C. reinhardtii* genome (20) and based on sequence similarity to genes encoding enzymes involved in these ER steps, we identified most of the enzymes involved in the biosynthesis of dolichol pyrophosphate-linked oligosaccharide (Table IV). Some of these predicted enzymes show strong homologies with the corresponding asparagine-linked glycosylation (ALG) orthologs described in other eukaryotes (61). No candidate gene was found to correspond to ALG3, ALG9, ALG10, and ALG12. Putative transferases able to catalyze the formation of dolichol-activated mannose and glucose (CrDPM1 and CrALG5, respectively) required for the biosynthetic steps arising in the ER lumen were also predicted. In addition, genes whose translation products display high percentages of identity with the flippase involved in the translocation of the dolichol pyrophosphate-associated intermediate and with subunits of the OST complex (STT3A, STT3B, DLG1, DAD1, ribophorin I and II, and OST3) were also identified (Table IV).

A search for putative proteins involved in the quality control of the proteins in the ER led to the identification of sequences encoding a glucosidase I, as well as the α and β subunits of glucosidase II. The α subunit contains the DMNE sequence (62) and a lectin domain involved in the binding of mannose residues (63). Glucosidase II is responsible for the cleavage of two α(1,3)-linked glucose residues from the precursor N-glycan. The trimming of terminal glucose residues allows the
| Putative function                  | Protein name | Gene name (JGI v4.3) | Gene location | Best hit in A. thaliana* | SignalP prediction | TMHMN prediction | Predicted domains |
|-----------------------------------|--------------|----------------------|---------------|--------------------------|--------------------|------------------|------------------|
| **ER cytosolic enzymes**          |              |                      |               |                          |                    |                  |                  |
| N-acetylglucosamine phosphoasfrase| CrALG7       | Cre16g663100         | Chromosome_16: 1990552–1993342 | At3g57220 (43%) | Signal peptide 10 TMD | PF00953          |                  |
| β-1,4-N-acetylglucosaminyl transferase | CrALG13  | Cre13g358580         | Chromosome_13: 3255029–3257117 | At4g16710 (44%) | None | None | PF04101          |
| β-1,4-N-acetylglucosaminyl transferase | CrALG14  | Cre16g669950         | Chromosome_16: 3067735–3069662 | At4g18230 (40%) | Signal peptide 1 TMD | PF08660          |                  |
| β-1,4-mannosyltransferase         | CrALG1      | Cre12g516550         | Chromosome_12: 3825477–3826903 | At1g16570 (44%) | Signal anchor None | PF00534          |                  |
| α-1,3-mannosyltransferase         | CrALG2      | Cre11g474450         | Chromosome_11: 1150916–1152886 | At1g78800 (46%) | None | None | PF00534          |
| α-1,2-mannosyltransferase         | CrALG11     | Cre23g767350         | Scaffold_23: 188138–193311 | At2g40190 (40%) | Signal anchor 2 TMD | PF00534          |                  |
| Flippase                          | CrRFT       | Cre22g765100         | Scaffold_22: 349328–353741 | At5g07630 (22%) | Signal peptide 9 TMD | PF04506          |                  |
| **ER lumenal enzymes**            |              |                      |               |                          |                    |                  |                  |
| α-1,3-glucosyltransferase         | CrALG6      | Cre16g690150         | Chromosome_16: 5766199–5770411 | At5g28460 (34%) | Signal peptide 11 TMD | PF03155          |                  |
| α-1,3-glucosyltransferase         | CrALG8      | Cre09g414250         | Chromosome_9: 4325467–4328535 | At2g44660 (44%) | None | 10 TMD | PF03155          |
| Calnexin                          | CrCLNX      | Cre07g357900         | Chromosome_7: 6143630–6147986 | At5g61790 (35%) | Signal peptide 1 TMD | PF03155          |                  |
| Calreticulin                      | CrCLRT      | Cre01g038400         | Chromosome_1: 529313–5297125 | At1g09210 (59%) | Signal peptide None | PF03155          |                  |
| Dolichol-phosphate mannosyltransferase | CrDPM1   | Cre03g150950         | Chromosome_3: 438314–439077 | At1g20575 (52%) | None | None | PF00535          |
| Dolichol-phosphate glucosyltransferase | CrALG5    | Cre16g652850         | Chromosome_3: 679141–682477 | At2g29630 (38%) | Signal peptide 1 TMD | PF00535          |                  |
| Glucosidase I                     | CrGSI       | Cre13g579750         | Chromosome_13: 2459268–2461998 | At1g67490 (36%) | None | None | PF03200          |
| Glucosidase II, α-subunit         | CrGSIiA     | Cre17g190500         | Chromosome_17: 4955490–4962655 | At5g63840 (49%) | Signal peptide 1 TMD | PF01055          |                  |
| Glucosidase II, β-subunit         | CrGSIiB     | Cre17g725350         | Chromosome_17: 3554839–3558301 | At5g56360 (25%) | None | None | PF07915          |
| UDP-glucose:glycoprotein glucosyltransferase | CrUGGC | Cre05g232250         | Chromosome_5: 672374–689328 | At1g71220 (15%) | Signal peptide None | PF06427          |                  |
| **Oligosaccharyltransferase complex subunits** | | | | | | |
| DDPGT subunit a                    | CrDGL1      | Cre14g614100         | Chromosome_14: 987242–990915 | At5g66680 (43%) | Signal peptide 2 TMD | PF03345          |                  |
| Ribophorin I                      | CrRPN1      | Cre12g523300         | Chromosome_12: 4550885–4553389 | At1g76400 (38%) | Signal peptide 1 TMD | PF04597          |                  |
| Ribophorin II                     | CrRPN2      | Cre09g368450         | Chromosome_9: 593746–596364 | At4g21150 (10%) | Signal peptide 1 TMD | PF05817          |                  |
| DDPGT subunit b                    | CrDAD1      | Cre02g108400         | Chromosome_2: 4617959–4619999 | At1g32210 (48%) | Signal anchor 3 TMD | PF02109          |                  |
| DDPGT subunit b                    | CrSTT3A     | Cre02g121650         | Chromosome_3: 6262471–6270238 | At5g19960 (27%) | Signal peptide 3 TMD | PF02516          |                  |
| DDPGT subunit b                    | CrOST3      | Cre01g063500         | Chromosome_1: 8748501–8751480 | At1g61790 (28%) | Signal peptide 4 TMD | PF04756          |                  |
| **GolgI enzymes**                 |              |                      |               |                          |                    |                  |                  |
| Endomannosidase                    | CrEMAN      | Cre03g189050         | Chromosome_3: 4795688–4801485 | - | None | 1 TMD | PF03659          |
| α-1,2-mannosidase I                | CrMAN1      | Cre07g336600         | Chromosome_7: 304731–311367 | At1g30000 (28%) | Signal peptide None | PF01532          |                  |
| β-1,2-xyllosyltransferase          | CrXYLT      | Cre07g125700         | Chromosome_7: 683318–6837848 | At5g55550 (17%) | Signal peptide None | DUF563 (PF04557) |                  |
| α-1,3-fucosyltransferase           | CrFUT1      | Cre31g780450         | Scaffold_31: 155613–159255 | At3g19280 (20%) | Signal anchor 1 TMD | PF00852          |                  |

Annotation of genes was carried out in the *C. reinhardtii* v4.3 genome.

*Pairwise alignments were done using full-length amino-acid sequences with Clustal W1.8 and percentages of identity are given into parenthesis.

DDPGT is the abbreviation for dolichol-diphospho-oligosaccharide protein glycosyltransferase.
binding and release of monoglycosylated glycoproteins with calnexin and calreticulin, two ER-resident lectin-like chaperones that are involved in the retention of misfolded or incompletely folded proteins (64). A sequence encoding a UGGT, involved in the entry of incompletely folded proteins into cycles of calnexin/calreticulin-assisted folding (65), was also identified in the C. reinhardtii genome.

After their ER processing, the glycoproteins move to the Golgi apparatus, where the oligomannoside N-glycans are stepwise matured into complex-type N-glycans. Three types of mannosidases are predicted in the C. reinhardtii genome (Table IV). An endo-mannosidase belonging to the CAZy family GH 99, CrEMAN, has been identified exhibiting 36.5% identity with the human homolog. This mannosidase, identified in animals but not in plants, is able to release a Man-8 oligomannoside by cleaving internally the glucosylated precursors (66). Endo-mannosidases are usually located in cis Golgi and provide an alternative pathway for the processing of the ER N-glycan precursor (67). In addition to this endo-mannosidase, one putative type-I mannosidase is predicted in the genome (Table IV). Although this glycosidase, CrMANI, does not display the typical topology of Golgi enzymes, the sequence exhibits 26% to 28% identity with human and plant α-MANI, as well as the conserved aminoacids of the catalytic domain involved in Ca²⁺ and oligomannoside bindings and Cys residues involved in its folding (68, 69). In addition, a sequence encoding α-mannosidase II belonging to the CAZy family GH38 was predicted (Table IV). This putative mannosidase displays the greatest similarity to the human cytosolic type-II mannosidase C (MANIC, NP_067062.2), which has been shown to be involved in the turnover of free oligosaccharides (70, 71). However, a putative function as a Golgi mannosidase involved in the N-glycan trimming cannot be definitively ruled out.

Usually, the synthesis of complex-type N-glycans starts with the transfer of a GlcNAc residue on the α(1,3)-mannose arm of Man₉GlcNAc₂ by the action of a GnT I. However, no putative GnT I or GnT II sequence was identified in C. reinhardtii, suggesting the absence of a GnT I-dependent pathway in this green microalga. A search for a putative xylanotransferase revealed the presence of one sequence (CrXYLT) exhibiting about 16.5% identity with β(1,2)-xylosyltransferase from Arabidopsis thaliana, in which this enzyme is responsible for the transfer of a β-xylene onto the β-mannose of the core N-glycan (72). However, considering the lack of information regarding conserved peptide domains required for β(1,2)-xylosyltransferase activity on N-linked glycans, the assignment of such a sequence remains highly speculative. A putative fucosyltransferase, CrFUT1, exhibiting 20% and 21% identity with α(1,3)-fucosyltransferases from A. thaliana AtFUT11 and AtFUT12, respectively, was also predicted in the genome (Table IV). This protein sequence exhibited the expected type-II membrane protein topology and motifs required for α(1,3)-fucosyltransferase activity (73–75), as well as conserved Cys residues and a CXXC motif located at the C-terminal end that is involved in the formation of disulfide bonds in plant α(1,3)-fucosyltransferases (76).

**DISCUSSION**

Here, we developed an integrated genomic, glycomic, and glycoproteomic approach to unravel the N-glycosylation pathway of C. reinhardtii and shed light on N-glycan structures and N-glycosylated proteins. Based on sequence similarities, we identified in the genome of C. reinhardtii a set of putative sequences encoding proteins involved in the synthesis in the ER of the dolichol pyrophosphate–linked oligosaccharide donor Glc₃Man₉GlcNAc₂–PP-dolichol, its transfer by OST onto asparagine residues of proteins, and the deglycosylation/reglycosylation of the precursor N-glycan allowing its interaction with chaperones involved in the quality control of secreted proteins (Fig. 6, Table III and Table IV). Some of these proteins (STT3A/STT3B, UGGT) were identified in the proteome analysis of C. reinhardtii. In addition, the biochemical investigation of the N-glycan structures showed that both secreted and membrane-bound C. reinhardtii glycoproteins bear mainly Man₉GlcNAc₃ to Man₉GlcNAc₅ structures representing almost 70% of the total N-glycan population. Although some ALGs were not clearly identified in the C. reinhardtii genome, the identification of large oligomannosides up to Man₉GlcNAc₃ (Man-9) (Table I) suggested that the biosynthesis of C. reinhardtii N-glycan in the ER is similar to that described in other eukaryotes. In addition, the structure of Man₉GlcNAc₅ (Man-5) detected in C. reinhardtii N-glycan pools is identical to the one usually observed on eukaryote N-linked proteins. Complex N-glycans were also identified on secreted and membrane-bound proteins isolated from C. reinhardtii. These N-glycans are partially O-methylated Man₉GlcNAc₂ to Man₉GlcNAc₂ bearing one or two xylose residues. Based on MS2 fragmentation and immunoblotting data, we demonstrated that one of these xylose residues is linked in β(1,2) to the core β-Man as previously reported in plants (56), whereas the second one is linked in C4 on one outer terminal mannose. Although a putative fucosyltransferase, CrFUT1, is predicted in the C. reinhardtii genome, only traces of fucosylated glycans were detected in the N-glycan profiles. These complex N-linked glycans were also observed on an individual peptide, the PKHD1L1 peptide TITVANN*GTHSTATLK, exhibiting extensive glycan heterogeneity (Fig. 4, supplemental File S7). These results contrast with those obtained in Porphyridium sp., in which a cell wall glycoprotein was found to carry Man₉GlcNAc₂ and Man₉GlcNAc₂ containing 6-O-methyl mannose and substituted by one or two xylose residues, with one xylose located on the chitobiase unit (11).

Based on both in silico and biochemical analyses, we postulate, as illustrated in Fig. 6, that after their synthesis in the ER, oligomannoside N-glycans are processed into Man₉GlcNAc₂ in the Golgi apparatus by Golgi-residing mannosidases such as
the putative type I-mannosidase CrMANI (Table IV). The formation of complex-type N-glycans then occurs via additional maturation steps such as xylosylation and methylation of mannoses (Fig. 6). Although functional characterization of Golgi putative transferases is required in order to definitively establish the precise order of Golgi events, the absence in N-glycan profiles of methylated Man$_2$GlcNAc$_2$ and Man$_1$GlcNAc$_2$ suggests that O-methylation of mannose residues likely occurs after the xylosylation of oligomannosides.

In most eukaryotic organisms, GnT I transfers a GlcNAc residue on the $\alpha$(1,3)-mannose arm of Man$_5$GlcNAc$_2$ to initiate the synthesis of complex-type N-glycans. However, because no gene encoding a putative GnT I could be identified in the C. reinhardtii genome and neither MALDI-TOF analyses of N-glycans nor IS-CID experiments indicated any GnT I–dependent activities, we conclude that the maturation of complex N-glycans occurs through a GnT I–independent pathway. N-glycan processing in a GnT I–independent pathway has already been demonstrated to occur in GnT I mutants (77, 78) or in organisms devoid of GnT I activity such as mushrooms (79). In contrast, N-glycans are processed in a GnT I–dependent manner in the diatom *Phaeodactylum tricornutum* (12), implying the existence of distinct N-glycosylation pathways in microalgae depending on the phyla they belong to.

From the proteomic data, it is clear that *C. reinhardtii* possesses numerous functionally interesting N-glycosylated proteins. The highest number of distinct N-glycoproteins was detected in the culture medium of *C. reinhardtii*, which is not surprising, because glycosylation is a common characteristic of secreted proteins (19, 80). Moreover, glycoproteomic analyses were carried out on the *C. reinhardtii* strain CC-400, which easily releases periplasmic proteins into the growth medium because of its cell wall deficiency (81). To compensate for the “loss” of extracellular proteins, the expression of secreted proteins may be up-regulated in this strain. Most of the identified proteins lack functional annotation, yet many of them feature conserved domains that suggest proteolytic and/or carbohydrate-binding activities. Correspondingly, they may be involved in processes such as nutrient acquisition, cell−cell recognition, or cell wall degradation. The latter function was confirmed for the matrix metalloprotease MMP1 (G-lysin), which is induced during gametogenesis (82–85). However, it remains unknown whether the seven uncharacterized glycoproteins containing gametolysin domains serve a similar function.

Through BLASTp searches, 14 human proteins were identified that showed high sequence similarity to glycoproteins from *C. reinhardtii* (Table III). Among these, five proteins showed sequence conservation even with respect to the localization of the NXT/NXS motif. For example, two N-glycosylation sites were identified in each of the predicted oligosaccharyltransferases CrSTT3A and CrSTT3B. The peptide sequence alignments showed that these sites are located within a region that is highly conserved in eukaryotic organisms and is proposed to harbor the catalytic site (86).

![Proposed N-glycosylation biosynthesis pathway in Chlamydomonas reinhardtii.](image)

**Fig. 6.** Proposed N-glycosylation biosynthesis pathway in *C. reinhardtii*. The proposed pathway is based on the major N-glycan structures found according to the *in silico* analysis. N-glycan structures have been drawn using the symbolic nomenclature adopted by the Consortium for Functional Glycomics (101). ■, N-acetylglucosamine; ●, mannose; ○, xylose.
made for the corresponding residue (N535) of yeast STT3. Moreover, mutational studies led to the conclusion that non-glycosylated N535 is essential for proper enzyme function (86). In humans, however, this residue is indeed N-glycosylated, in both STT3A and STT3B (13, 89). The example of STT3 proteins demonstrates that N-glycosylation sites are highly conserved across distantly related organisms when N-glycans are essential for enzyme activity. The subtle differences in the glycosylation patterns of human, yeast, and C. reinhardtii STT3 proteins may provide fundamental information regarding the principles of N-glycosylation in eukaryotes.

Polypeptide sequence alignments showed that four out of eight N-glycosylation sites determined for PKHD1–1 (fibrocystin-like protein) aligned perfectly with NXT/NXS motifs of the human homolog PKHD1L1 (synonym: PKHDL1; supplemental Files S6 and S7). No glycopeptidomic data are available indicating N-glycosylation of PKHD1L1. However, its paralog fibrocystin (polycystic and hepatic disease 1 (PKHD1)) was known to be highly N-glycosylated (90). Fibrocystin-like proteins are proposed to be evolutionary ancestors of fibrocystin and thus exhibit many structural similarities (91–93). In fact, although it is more similar to PKHD1L1 on the peptide-sequence level, PKHD1–1 is actually suggested to be the functional homolog of fibrocystin, because PKHD1–1 and PKHD1 are localized to flagella and primary cilia/basal bodies, respectively (93, 94). In contrast, PKHD1L1 might play a role in cellular immunity, as was concluded from the widespread PKHD1L1 expression in human and murine blood-derived cell lines and its up-regulation in T lymphocytes (91). The function of fibrocystin remains unknown, but mutations in the PKHD1 gene are linked to autosomal-recessive polycystic kidney disease (92, 95). The structural similarity of fibrocystin and fibrocystin-like proteins in addition to the presence of conserved (potential) N-glycosylation sites underlines that C. reinhardtii may be a suitable model system for studying human ciliary dysfunctions.

Glycopeptides of the heat shock protein HSP70G were identified in several cell fractions. Whether the widespread distribution was caused by cross-contaminations or HSP70G is indeed localized to several cellular compartments is currently being elucidated. HSP70G was initially presumed to be an ER resident protein as predicted by TargetP and inferred from its sequence similarity to the human ER-localized HYOU1 protein (96). However, recent studies have shown that HSP70G is also chloroplast localized in C. reinhardtii (97). Interestingly, HYOU1 localization was found not to be restricted to the ER. In rats, glycosylated HYOU1 was detected in mitochondria, and an N-terminally truncated form was found in the cytoplasm (98, 99). Accordingly, we assume that HSP70G and its homologs may be localized to several cellular compartments as part of a multiple targeting strategy. The same conclusion may be drawn for MnSOD3, as it was found recently to be a chloroplast-located superoxide dismutase (60). Because MnSOD3 was found to be N-glycosylated and was identified in the culture medium, it is evident that the protein must take a route through the secretory pathway with subsequent distribution to multiple destinations. A corresponding targeting mechanism of proteins to the chloroplast via the ER and Golgi apparatus was already described in Arabidopsis for N-glycosylated α-carbinic anhydrase (100).

The exploration of the C. reinhardtii N-glycan pathway as done in the present study represents an important first step toward the design of genetically engineered driven remodeling of the alga to produce Chlamydomonas-derived biopharmaceuticals carrying N-linked glycans compatible with human therapeutical applications. Notably, our comprehensive analyses also revealed N-glycosylated proteins in the chloroplast, as well as in the extracellular space, thereby providing information for future targeting experiments for the expression of glycoproteins of biotechnological interest.

Acknowledgments—We would like to dedicate this paper to Dr. Arsenio Villarejo, a friend and co-worker who was taken from us too early on May 2, 2011, and who initiated this research with us.

* Collaborative research conducted by the universities of Rouen (France), Münster (Germany), and Autonoma of Madrid (Spain) was supported by the ANR, PLANT-KBBE 2008 program in the frame of the ALGALGLYCO project. The experiments done at Bioprocessing Technology Institute (BTI, Singapore) were supported by the Biomedical Research Council of A*STAR (Agency for Science, Technology and Research). The collaboration between the University of Rouen (France) and BTI (Singapore) is supported by the MERLION 2011 initiative, project Glyco-TOOLS. This work has also been supported by INSIA Rouen, CNRS, Region Haute-Normandie, EFRD (No. 31708), and Labex SynOrg (ANR-11-LABX-0029).

This article contains supplemental material.

To whom correspondence should be addressed: Dr. Muriel Bardor, Laboratoire Glyco-MEV, EA 4358, Institut de Recherche et d’Innovation Biomédicale (IRIB), Faculté des Sciences et Techniques, 76821 Mont-Saint-Aignan Cedex, France, Tel.: 33–2–35–14–67–58, Fax: 33–2–35–14–66–15, E-mail: muriel.bardor@univ-rouen.fr; Prof. Michael Hippler, Institute of Plant Biology and Biotechnology, Schlossplatz 8, University of Münster, D-48143, Germany, Tel.: 49–251–83–24790, Fax: 49–251–83–28371, E-mail: mhippler@uni-muenster.de.

These authors contributed to this work equally.

The experiments done at Bioprocessing Technology Institute (BTI, Singapore) is supported by the MERLION 2011 initiative, project Glyco-TOOLS. This work has also been supported by INSIA Rouen, CNRS, Region Haute-Normandie, EFRD (No. 31708), and Labex SynOrg (ANR-11-LABX-0029).

These authors contributed to this work equally.

Author contributions: E.M.-R., C.A., F.D., F.L.M., G.T., M.S., S.S., A.K.H., A.B.R., C.L.-B., M.-C.K.-M., and C.B. performed sample preparations, glycan and glycoproteome analysis, analysis of monosaccharide composition, and bioinformatic analyses and generated the data. C.F. helped in bioinformatic analyses of the proteomic data. E.M.-R., F.D., F.L.M., P.L., M.B., M.S., and M.H. wrote the paper. P.L., M.B., F.M., and M.H. coordinated research efforts among authors. All authors read and approved the manuscript.

REFERENCES
1. Harris, E. H. (2008) The Chlamydomonas Sourcebook: Introduction to Chlamydomonas and Its Laboratory Use, Vol. 1, 2nd ed., Elsevier, Oxford, UK
2. Imam, S. H., Buchanan, M. J., Shin, H. C., and Snell, W. J. (1985) The Chlamydomonas cell wall: characterization of the wall framework. J. Cell Biol. 101, 1599–1607
3. Ferris, P. J., Woessner, J. P., Waffenschmidt, S., Kitz, S., Drees, J., and Goodenough, U. W. (2001) Glycosylated polyproline II rods with kinks as
a structural motif in plant hydroxyproline-rich glycoproteins. Biochemistry 40, 2978–2987.

4. Bollig, K., Lamshöft, M., Schweimer, K., Marner, F. J., Budzikiewicz, H., and Waffenschmidt, S. (2007) Structural analysis of linear hydroxyproline-bound O-glycans of Chlamydomonas reinhardtii—conservation of the inner core in Chlamydomonas and land plants. Carbohydr. Res. 342, 2057–2065.

5. Burda, P., and Aebi, M. (1999) The dolichol pathway of N-linked glycosylation. Biochim. Biophys. Acta 1426, 239–257.

6. Helenius, A., and Aebi, M. (2001) Intracellular functions of N-linked glycoproteins. Science 291, 2364–2369.

7. Varki, A. (2011) Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. Cold Spring Harb. Perspect. Biol. 11, 1–14.

8. Grunow, A., Becker, B., and Melkonian, M. (1993) Isolation and characterization of the Golgi apparatus of a flagellate scaly green alga. Eur. J. Cell Biol. 61, 10–20.

9. Becker, B., Perasso, L., and Melkonian, M. (1996) Scale-associated glycoproteins of Scherffelia dubia (Chlorophyta) form high-molecular-weight complexes between the scale layers and the flagellar membrane. Planta 199, 503–510.

10. Balasubramanian, G., and Jaenicke, L. (1999) The oligosaccharides of the glycoprotein pheromone of Volvox carteri f. nagariensis iyengar (Chlorophyceae). Eur. J. Biochem. 192, 231–237.

11. Levy-Ontman, O., Arad, S., Harvey, J. D., Parsons, T. B., Fairbanks, A., and Tekoah, Y. (2011) Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalgae Porphyridium sp. J. Biol. Chem. 286, 21340–21352.

12. Bailleul, B., Burel, X., 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-glycans of Phaeodactylum tricornutum diatom and functional characterization of its N-acetylgalactosaminyltransferase I enzyme. J. Biol. Chem. 286, 6152–6164.

13. Chen, R., Jiang, X., Sun, D., Han, G., Wang, F., Ye, M., Wang, L., and Zou, H. (2009) Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide. J. Proteome Res. 8, 651–661.

14. Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Kiefer-Meyer, M.-C., and Mayfield, S. P. (2010) Production of recombinant proteins in algae, analysis of expression of seven human proteins in the chloroplast of Chlamydomonas reinhardtii. Plant Biotechnol. J. 8, 719–733.

15. Mayfield, S. P., Franklin, S. E., and Lerner, R. A. (2003) Expression and assembly of a full active antibody in algae. Nat. Protoc. 8, 438–442.

16. Tran, M., Zhou, B., Pettersson, P. L., Gonzalez, M. J., and Mayfield, S. P. (2009) Synthesis and expression of a full-length monoclonal antibody in algae chloroplasts. Biotechnol. Bioeng. 104, 663–673.

17. Eichler-Stahlberg, A., Weisheit, W., Ruecker, O., and Heitzer, M. (2009) Strategies to facilitate transgene expression in Chlamydomonas reinhardtii. Planta 229, 873–883.

18. Walsh, G., and Jeffers, R. (2006) Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. 24, 1241–1252.

19. Jiang, X., Song, A., Bergelson, S., Arroll, T., Parekh, B., May, K., Chung, S., Strouse, R., Mire-Sluis, A., and Schennerman, M. (2011) Advances in the assessment and control of the effecter functions of therapeutic antibodies. Nat. Rev. Drug Discov. 10, 101–111.

20. Ling, N., Zhang, P., Song, Z., and Bardor, M. (2012) The sweet tooth of biopharmaceuticals: importance and analysis of recombinant protein glycosylation. Biotechnol. J. 7, 1462–1472.

21. Bardor, M., Faveeuw, C., Flitsch, A. C., Gilbert, D., Carlier, A. M., and Marin-Navarro, J. (2007) Production of therapeutic proteins in algae, analysis of the inner core in Chlamydomonas and land plants. Biotechnol. Bioeng. 104, 663–673.

22. Galili, U. (2004) Immune response, accommodation, and tolerance to transplantation carbohydrate antigens. Transplantation 78, 1093–1098.

23. Padler-Karavani, V., Yu, H., Cao, H., Chokhawala, H., Karp, F., Varki, N., Chen, X., and Varki, A. (2008) Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: potential implications for disease. Glycoconjugate J. 23, 447–434.

24. Lu, Q. (2004) Immunoreactivity, accommodation, and tolerance to transplantation carbohydrate antigens. Transplantation 78, 1093–1098.

25. Moseley, J. L., Allinger, T., Herzog, S., Horeth, P., Wehinger, E., Merchant, S., and Hippler, M. (2002) Adaptation to Fe-deficiency requires remodeling of the photosynthetic apparatus. EMBO J. 21, 6709–6720.

26. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

27. Sestim, N., Cabrera, G., Triguero, A., Burel, C., Leprince, J., Rihouey, C., Vézina, L.-P., D'oust, M.-A., Ruff, P., Roy, L., Dwek, R., Harvey, J., Leriche, P., Cremata, J., and Bardor, M. (2008) Plant N-glycan profiling...
fucosyltransferase cDNA from Drosophila melanogaster. Potential basis of the neural anti-horse radish peroxidase epitope. J. Biol. Chem. 276, 28058–28067
76. Holmes, E. H., Yen, T. Y., Thomas, S., Joshi, R., Nguyen, A., Long, T., Gallet, F., Mattaf, A., Julien, R., and Macher, B. A. (2000) Human alpha 1,3/4 fucosyltransferases. Characterization of highly conserved cysteine residues and N-linked glycosylation sites. J. Biol. Chem. 275, 24237–24245
77. Zhu, S., Hanneman, A., Reinhold, V. N., Spence, A. M., and Schachter, H. (2004) Caenorhabditis elegans triple null mutant lacking UDP-N-acetyl-D-glucosamine:alpha-3-D-mannoside beta-1,2-N-acetylglucosaminyl-transferase I. Biochem. J. 382, 995–1001
78. Crispin, M., Harvey, D. J., Chang, V. T., Yu, C., Aricescu, A. R., Jones, E. Y., Davis, S. J., Dwek, R. A., and Rudd, P. M. (2006) Inhibition of hybrid- and complex-type glycosylation reveals the presence of the GlcNAC transferase I-independent fucosylation pathway. Glycobiology 16, 748–756
79. Grass, J., Pabst, M., Kolarich, D., Pöltl, G., Léonard, R., Brecker, L., and Altmann, F. (2011) Discovery and structural characterization of fucosylated oligomannosidic N-glycans in mushrooms. J. Biol. Chem. 286, 5977–5984
80. Agrawal, G. K., Jwa, N.-S., Lebrun, M.-H., Job, D., and Rakwal, R. (2010) Kinoshita, T., Fukuzawa, H., Shimada, T., Saito, T., and Matsuda, Y. (1992) Studies on the function of oligosaccharyltransferase subunits. Stt3p is directly involved in the glycosylation process. J. Biol. Chem. 277, 47692–47700
81. Kubo, T., Saito, T., Fukuzawa, H., and Matsuda, Y. (1992) Primary structure and expression of a gamete lytic enzyme in Chlamydomonas reinhardtii: similarity of functional domains to matrix metalloproteases. Proc. Natl. Acad. Sci. U.S.A. 89, 4693–4697
82. Matsuda, Y., Saito, T., Yamaguchi, T., and Kawase, H. (1985) Cell wall lytic enzyme released by mating gametes of Chlamydomonas reinhardtii and its application as a high-CO2-responsive marker for the study of the high-CO2-sensing mechanism in Chlamydomonas reinhardtii. Plant Cell Physiol. 26, 299–309
83. Kinoshita, T., Fukuzawa, H., Shimada, T., Saito, T., and Matsuda, Y. (1992) Cystin is expressed on primary cilia. Hum. Mol. Genet. 1, 572–583
84. Malerod, H., Graham, R. L. J., Sweredoski, M. J., and Hess, S. (2013) Analysis of a eukaryotic cillum. J. Cell Biol. 170, 103–113
85. Onuchic, L. F., Furu, L., Nagasawa, Y., Hou, X., Eggemann, T., Ren, Z., Berghman, C., Senderek, J., Esquivel, E., Zeltner, R., Rudnik-Schöneborn, S., Trujillo, D., Sweeney, W., Avner, E. D., Zerres, K., Guay-Woodford, L. M., Somlo, S., and Gemino, G. G. (2002) PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix 1 repeats. Am. J. Hum. Genet. 70, 1305–1317
86. Schroda, M., and Vallon, O. (2008) Chaperones and proteases. In The Chlamydomonas Sourcebook (Stem, D. B., ed.) 2nd ed., Vol. 2, pp. 671–729, Elsevier, Oxford, UK
87. Terashima, M., Specht, M., Naumann, B., and Hippler, M. (2010) Characterizing the anaerobic response of Chlamydomonas reinhardtii by quantitative proteomics. Mol. Cell. Proteomics 9, 1514–1532
88. Arriaga, D. D., and Schnell, R. G. (2008) Targeting of the molecular chaperone oxygen-regulated protein 150 (ORP150) to mitochondria and its induction by cellular stress. Am. J. Physiol. Cell Physiol. 294, C641–C650
89. Yu, L.-G., Andrews, N., Weldon, M., Gerassimenko, O. V., Campbell, B. J., Singh, R., Grierson, I., Petersen, O. H., and Rhodes, J. M. (2002) An N-terminal truncated form of Orp150 is a cytoplasmic ligand for the anti-proliferative mushroom Agaricus bisporus lectin and is required for nuclear localization sequence-dependent nuclear protein import. J. Biol. Chem. 277, 24538–24545
90. Villarejo, A., Burián, S., Larsson, S., Dénardin, A., Monné, M., Rudhe, C., Karlsson, J., Janson, S., Lerouge, P., Rolland, N., Von Heijne, G., Grebe, M., Bakó, L., and Samuelsson, G. (2005) Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat. Cell Biol. 7, 1224–1231
91. Varri, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth, J. D., Bortzoi, C. R., Hart, G. W., and Etzler, M. E. (2009) Symbol nomenclature for glycan representation. Proteomics 9, 5398–5399
92. Deleted in proof
93. Deleted in proof
94. Deleted in proof
95. Deleted in proof
96. Deleted in proof
97. Deleted in proof
98. Deleted in proof
99. Deleted in proof
100. Deleted in proof
101. Deleted in proof
102. Deleted in proof
103. Deleted in proof