This paper is available on line at http://www.mcponline.org

Exploring the *N*-glycosylation Pathway in *Chlamydomonas reinhardtii* Unravels Novel Complex Structures*[S]

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 dissodiation (CID) for glycan fragmentation followed by mass tag-triggered CID for peptide sequencing and PNGase F treatment of glycopeptides in the presence of ^{18}O-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 maturate 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 $Man_5GlcNAc_2$-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

From the ‡Université de Rouen, Laboratoire Glyco-MEV, EA 4358, Institut de Recherche et d'Innovation Biomédicale (IRIB), 76821 Mont-Saint-Aignan Cedex, France; ¶Institute of Plant Biology and Biotechnology, Schlossplatz 8, University of Münster, D-48143, Germany; ‖Comisión Docente de Fisiología Vegetal, Departamento de Biología, Edificio de Biología Universidad Autónoma de Madrid, 28049 Madrid, Spain; ‡‡Bioprocessing Technology Institute, Agency for Science Technology and Research (A*STAR), 20 Biopolis Way, #06-01, Centros, Singapore, 138668; §§Université de Rouen, Laboratoire COBRA UMR 6014 & FR 3038, INSA de Rouen, 1 Rue Tesnière, 76821 Mont St Aignan Cedex, France

Received February 6, 2013, and in revised form, August 1, 2013

Published, MCP Papers in Press, August 2, 2013, DOI 10.1074/mcp.M113.028191

[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*-acetylglucosaminyltransferase I; Hex, hexose; HexNAc, *N*-acetylglucosamine; 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 $Glc_3Man_9GlcNAc_2$-PP-dolichol *N*-glycan precursor is completed (5). 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 $Man_9GlcNAc_2$ (Man-9) to $Man_5GlcNAc_2$ (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 *Chlorophycaea* using on-blot affinodetection 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 therapeutical 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% CO_2. 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-sufficient 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 Phytozome 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

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 2,500*g* (SORVALL® RC5C Plus). The cells were washed with 20 mM potassium phosphate buffer (pH 7.4) and underwent 5 min of centrifugation at 2,500*g* 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 300*g* for 3 min to remove intact cells and debris. The supernatant was centrifuged again at 20,000*g* for 30 min and then ultracentrifuged at 100,000*g* for 1 h (Centrikon 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 with an ammonium bicarbonate 50 mM acetonitrile (1/1 V:V) solution for 15 min. This washing step was repeated once. Reduction and alkylation were then performed by means of incubation with, respectively, 25 mM dithiothreitol (DTT) for 45 min at 56 °C and 55 mM iodoacetamide at room temperature in the dark for 20 min; both DTT and iodoacetamide were dissolved in 50 mM ammonium bicarbonate buffer. Then, alkylated proteins were in-gel digested overnight by trypsin treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone TPCK needs to be add in the brackets (Sigma) at a ratio of 20:1 (protein:trypsin) at 37 °C under agitation. The resulting peptides and glycopeptides were extracted from the gel via a succession of washes with 100% acetonitrile, 100 mM ammonium bicarbonate, 100% acetonitrile, and 5% formic acid. The peptide and glycopeptide mixtures were separated 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 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 mU of enzyme was added for every 5 mg of protein and incubated at 37 °C overnight under agitation. After the digestion, isolation of the released *N*-glycans was performed using a C18 cartridge (Waters) conditioned as previously described. After the sample had been loaded, the free *N*-glycans were recovered by 5 ml of water. This fraction was dried down prior to purification by a Hypersep Hypercarb cartridge following the manufacturer's instructions (Thermo Scientific).

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 sensibility 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 recalculation 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 I 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 mU of α-mannosidase from proteomic-grade *Canavalia ensiformis* (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 of each 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 $\times$ 250 mm) with a PA200 guard column (3 $\times$ 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 that, a re-equilibration 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; $E2 =$

+0.75 V, *t*2 = 200 ms; *E*3 = −0.15 V, *t*3 = 400 ms. The flow rate was kept constant at 0.3 ml min^{-1}, and the volume of sample/standard injected was 30 μl.

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 peracetylated. 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 × 30 m, 0.25-μm 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 μl 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 μA, and the pressure and temperature of the ion source were 2.10^{-6} mbar and 250 °C, respectively. The acceleration voltage was 8 kV, the resolution was 1,000 (10% valley definition), and the magnet scan rate was 1 s/decade over *m/z* range 600–38. The assignment to monomers was carried using standards of monosaccharides, as well as on the basis of their electron ionization fragmentation patterns.

Sialic Acid Release and HPAEC-PAD Analysis—Sialic acids bound to soluble and membrane proteins were released from 6 to 15 mg of CC503, CC1036 protein extracts by means of acetic acid hydrolysis and recovered through a 5-kDa vivaspin filter (Sartorius Stedim Biotech, Aubagne, France) following the procedure described in Ref. 43. Then the released sialic acids were analyzed via HPAEC-PAD. The experiment was run on a Dionex ICS-5000 system (Dionex, Sunnyvale, CA) equipped with an electrochemical detector. A carboPac PA 20 column (3 × 150 mm, Dionex) with a guard (3 × 30 mm, Dionex) was used for the analysis using a flow rate of 0.5 ml min^{-1} and the gradient conditions described in Ref. 44.

Affino- and Immunoblotting Analyses—*C. reinhardtii* total cell protein extracts were separated via SDS-PAGE and electrotransferred onto nitrocellulose membrane (membrane blotting, Pall Corporation, Port Washington, NY) for immunoblot or affinoblot analysis. Affinodetections with concanavalin A (Con A) (Sigma-Aldrich, St. Louis, MO) and biotinylated *Sambucus nigra* lectin (Vectorlabs, Burlingame, CA), and immunodetection with β(1,2)-xylose specific antibodies (Agrisera, Vännäs, Sweden) were performed according to the procedure described in Ref. 45. Total protein extracts from different organisms were used as controls: *Drosophila melanogaster*, which does not contain any core β(1,2)-xylose; *Arabidopsis thaliana* wild type; *A. thaliana* double mutant fut11/fut12, which does not contain any core α(1,3)-fucose in complex-type *N*-glycans; and the GnT I mutant of *A. thaliana* (cgl1), which is unable to synthesize complex-type *N*-glycans because of the lack of GnT I activity in this organism and produces only oligomannoside-type *N*-glycans.

Identification of Glycopeptides Using a Proteomic Approach Combined with Liquid Chromatography–Electrospray Ionization–Fourier Transform Mass Spectrometry—

Culture Conditions and Protein Isolation—A schematic overview of the various protein isolation and cell fractionation procedures is given in supplemental File S1. Cells required for the preparation of total cell extracts (TCEs) were harvested via centrifugation (3 min at 2,000*g*), washed in a small volume of fresh culture medium, and centrifuged again (10 min at 10,000*g*). The supernatants (SNs) containing secreted glycoproteins were combined and concentrated using 15-ml centrifugal filter devices (Amicon Ultra, 30-kDa molecular-weight cutoff). Both cell pellets and SNs were stored at −80 °C until further use. Isolation of chloroplasts and plasma membranes was performed according to established protocols (46, 47). Protein concentrations of all samples were determined using the Pierce BCA assay kit (Thermo Fisher Scientific) 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 7.6) 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 μg of protein was transferred to centrifugal filters and concentrated further to a volume of 40 μl. Then 350 μl 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, tryptic digestion, glycopeptide enrichment, and PNGase F–mediated glycan hydrolysis in ^{18}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 μg of protein were used per sample, and glycopeptide enrichment was carried out using 150 μl 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}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 μg 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 μl. 100 μl 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 μl of 50 mM iodoacetamide in UA buffer for 20 min in the dark and then centrifuged as before. Filters were washed twice with 100 μl of UA and twice with 200 μl 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 μl binding buffer and centrifugation of the inverted filter unit at 14,000*g* for 5 min. The transfer step was repeated once. Then, 200 μl of agarose-bound Con A were washed and equilibrated three times with 300 μl 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 μl NaCl-free binding buffer, with centrifugation at 10,000*g* for 5 min between washes. Glycoproteins were eluted via the addition of 150 μl of 0.5 M α-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 α-methyl D-mannopyranoside was removed by three successive washes with 200 μl of 50 mM ammonium bicarbonate. Afterward, glycoproteins were digested by the addition of 2 μg of trypsin (sequencing-grade modified, Promega, Madison, WI) in 40 μl ammonium bicarbonate and overnight incubation at 37 °C. Peptides

were eluted via centrifugation (14,000*g* 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 ^{18}O-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 width 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 ^{18}O-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, corresponding to the neutral loss of one and two HexNAc residues, respectively (48). These ion pairs potentially represented peptides whose glycans had been trimmed down to the chitobiose core by IS-CID, with the high-mass partner bearing one HexNac residue more than the low-mass partner. The three most intense ion pairs of each full scan were fragmented in the linear ion trap via CID (35% normalized collision energy) with multistage activation (MSA) of the neutral loss of HexNAc (−203.1 Da, −101.5 Da, and −67.7 Da for $z = 1$, 2, and 3). Each ion pair partner was isolated and fragmented individually. Consequently, a maximum of six fragmentation events were triggered per MS1 scan. Dynamic exclusion was enabled with a repeat and exclusion duration of 10 s, a repeat count of 1, a list size of 500, and an exclusion mass width of ±10 ppm. Unassigned charge states and charge states of 1 were rejected.

Glycopeptide Identification—For peptide identification, X!Tandem CYCLONE (49) incorporated into the proteomics data-processing pipeline Proteomatic (50) was used. MSA-CID spectra were matched against a target-decoy database composed of JGI4.3 Augustus 10.2 gene models merged with mitochondrial and chloroplast protein sequences from NCBI databases BK000554.2 and NC_001638.1, respectively. This database was supplemented with the protein sequences of jack bean Con A (UniProtKB: CVJB), *Flavobacterium meningosepticum* PNGase F (UniProtKB: P21163), and sequences of contaminant proteins from the Common Repository of Adventitious Proteins (version 1.0, released January 1, 2012). The total number of protein entries in the composite database was 16,940. Decoy protein sequences were generated by randomly shuffling tryptic peptides while retaining the redundancy of non-proteotypic peptides. The maximum number of missed cleavages allowed was two. The mass accuracy was set at 5 ppm for MS1 precursor ions and 0.5 Da for product ions. X!Tandem analyses were performed several times on spectra files, each time with a slightly modified set of glycosylation-related variable modifications (see below). The following modifications were used for all X!Tandem analyses: carbamidomethylation of cysteine (static), oxidation of methionine (variable), and deamidation of asparagine (variable). Peptide identifications were statistically validated using Qvality (version 2.02 (51)), with a *q*-value threshold of 0.01.

Additional Variable Modifications Used for the Identification of ^{18}O-labeled Peptides—Preliminary analyses led to the identification of numerous peptides derived from Con A and PNGase F, indicating high residual tryptic activity during the glycopeptide enrichment step of *N*-glyco-FASP. In order to take account of trypsin-mediated incorporation of ^{18}O 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 ^{18}O-labeled water (+2.9883 Da) and single incorporation of ^{18}O at the peptide C terminus (+2.0043 Da); (2) deamidation of asparagine in ^{18}O-labeled water and double incorporation of ^{18}O at the peptide C terminus (+4.0085 Da); (3) deamidation of asparagine in ^{16}O water (+0.9840 Da) and single incorporation of ^{18}O at the peptide C terminus (+2.0043 Da); and (4) deamidation of asparagine in ^{16}O-labeled water (+0.9840 Da) and double incorporation of ^{18}O 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 ^{18}O-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 pf18, 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

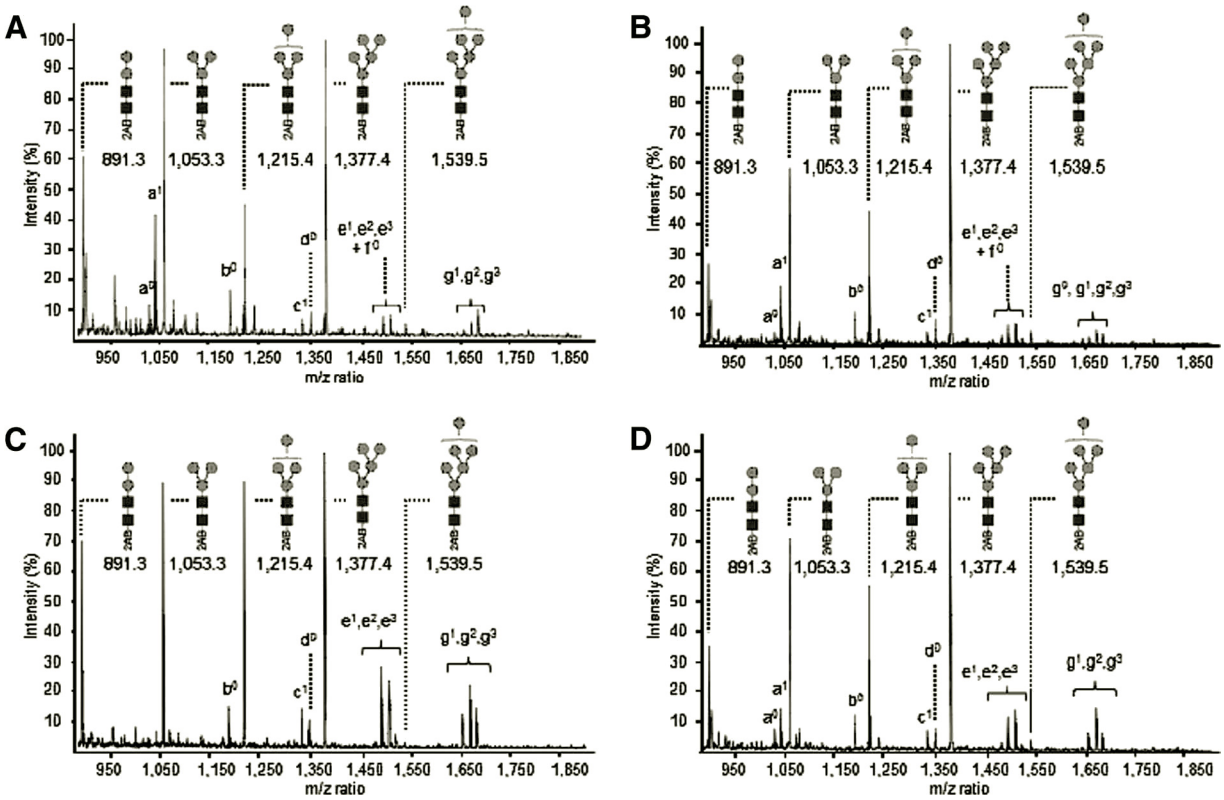


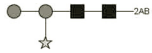
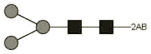
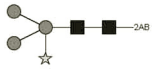
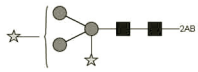
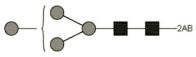
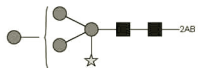
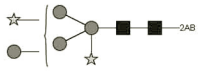
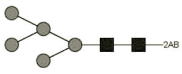
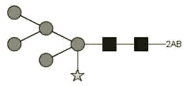
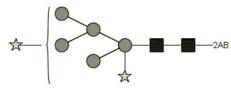
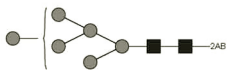
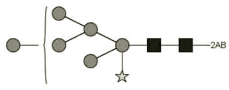
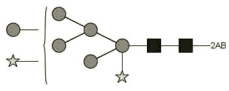
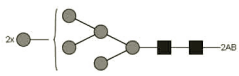
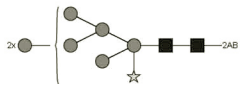
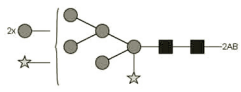
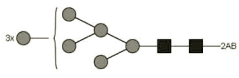
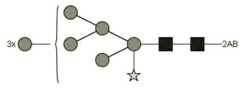
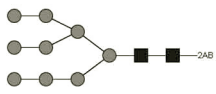
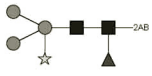
FIG. 1. **MALDI-TOF mass spectra of *N*-glycans borne by soluble and membrane proteins from two different strains of *C. reinhardtii*, CC-503 and CC-1036.** The *N*-glycans were released using PNGase A and labeled with the 2-AB fluorescent tag prior to MALDI-TOF-MS analysis. Mass spectra of the 2-AB-labeled *N*-glycans released from CC-503 soluble (*A*) and membrane-bound (*B*) proteins and from CC-1036 soluble (*C*) and membrane-bound (*D*) proteins are shown. Ions observed are sodium adducts. Oligomannoside *N*-glycans are annotated with their proposed carbohydrate structures according to the symbolic nomenclature adopted by the Consortium for Functional Glycomics (101). ■, *N*-acetylglucosamine; ●, mannose. The alphanumeric code indicates complex-type *N*-glycan structures (letters) and the number of methyl groups present on the structure (digits). The asterisks indicate ions that have been identified but not annotated in the spectra.

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 $hexose_{2-5}$ N-acetylglucosamine$_2$ ($Hex_{2-5}GlcNAc_2$) (Fig. 1; supplemental Table S1). Traces of larger oligomers up to $Hex_9GlcNAc_2$ 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 $Hex_{2-5}GlcNAc_2$ were converted into single species corresponding to a $HexGlcNAc_2$ 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 S3*A*). All these data allow us to assign these ions to oligomannoside *N*-glycans ranging from $Man_2GlcNAc_2$ to $Man_5GlcNAc_2$, 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-

TABLE I
Relative quantification of the N-glycans found on CC-503 soluble proteins

Identified oligomannoside N-glycans	Relative proportion +/- s.d.	Identified complex N-glycans	Relative proportion +/- s.d.
	15.4% +/- 2.8%	 a and A	3.9% +/- 1.3%
	17.2% +/- 0.6%	 b and B	5.6% +/- 0.8%
		 c and C	2.5% +/- 0.3%
	13.0% +/- 1.3%	 d and D	3.1% +/- 0.3%
		 e and E	6.3% +/- 1.3%
	22.9% +/- 0.6%	 f and F	1.6% +/- 0.3%
		 g and G	7.9% +/- 1.3%
	0.8% +/- 0.0%		traces
			traces
	traces		traces
			traces
	traces		traces
	traces	 h and H	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*-acetylglucosamine; ●, mannose; ☆, xylose; ▲, fucose.

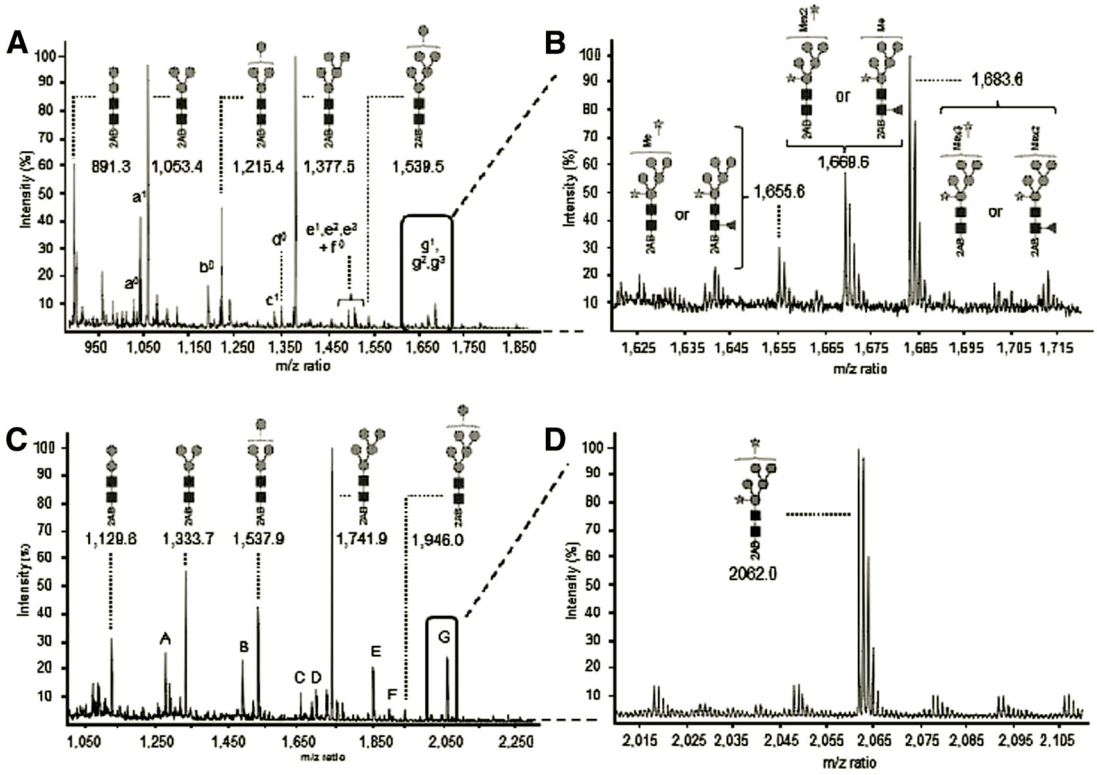


FIG. 2. **MALDI-TOF mass spectra of 2-AB-labeled *N*-glycans released from CC-503 soluble proteins using PNGase A.** *A*, mass spectrum of the 2-AB-labeled *N*-glycans released from CC-503 soluble proteins. *B*, zoomed view of the outlined area in panel *A*. *C*, mass spectrum of the 2-AB-labeled *N*-glycans released from CC-503 soluble proteins after permethylation. *D*, zoomed view of the outlined area in panel *C*. Ions observed are sodium adducts and are annotated with their proposed carbohydrate structures according to the symbolic nomenclature from the Consortium for Functional Glycomics (101). ■, *N*-acetylglucosamine; ●, mannose; ☆, xylose; ▲, fucose; Me, methyl group. The alphanumeric code indicates complex *N*-glycan structures (letters) and the number of methyl groups present on the structure (digits). The asterisks indicate ions that have been identified but not annotated in the spectra.

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. 2*A* and 2*B*). 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 $Man_4Xyl_2GlcNAc_2$ and $Man_5Xyl_2GlcNAc_2$, which indicated that the 14-Da mass shifts in the native oligosaccharides resulted from the partial methylation of the *N*-glycans (Figs. 2*C* and 2*D*), 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 $Man_3GlcNAc_2$ 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 $Man_5Xyl_2GlcNAc_2$, yielded product ions at *m/z* 462.4 and 707.5, which resulted from the fragmentation of the 2-AB-labeled chitobiose unit (Fig. 3*A*). No product ion was found that could support the notion that

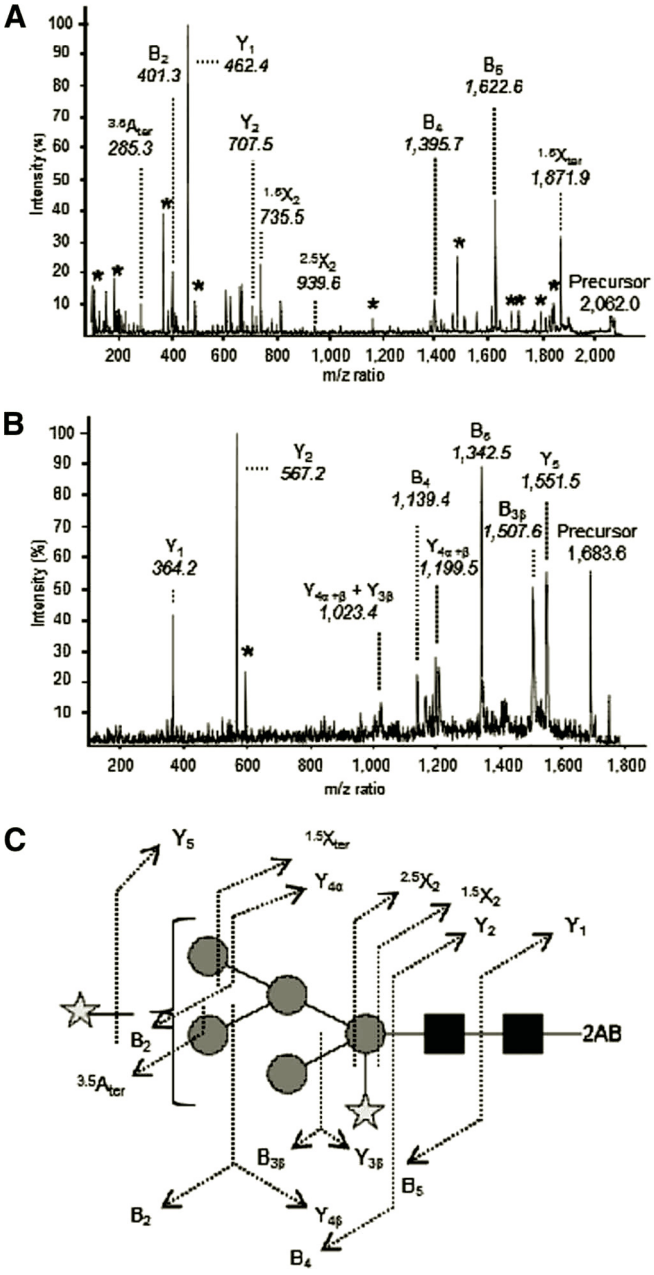


FIG. 3. **Fragmentation by MALDI-TOF-TOF-MS of the $Man_5Pen_2GlcNAc_2$ *N*-glycan revealing the linkage information for the pentose residues.** *A*, MS2 spectrum of permethylated 2-AB-labeled $Man_5Pen_2GlcNAc_2$. *B*, MS2 spectrum of native 2-AB-labeled $Man_5Pen_2GlcNAc_2$ indicating the positions of the methyl groups. *C*, assignment of the product ions observed in MS2 for the indicated *N*-glycan structure. ■, *N*-acetylglucosamine; ●, mannose; ☆, xylose; Me, methyl group. The asterisks indicate fragment ions that have been identified but not annotated on the structure in panel *C*.

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 α-mannose (Fig. 3*A*). Indeed, these two ions were shown to result from the core β-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 β-Man with a xylose residue. Because *C. reinhardtii* proteins were also immunodetected with specific core β(1,2)-xylose antibodies (supplemental File S3*B*), we concluded that this xylose is β(1,2)-linked to the core β-Man, as has been demonstrated in land plants (56).

Moreover, the presence of the ion at *m/z* 401.3 (B_2) indicated the location of the other xylose on a terminal mannose residue (Fig. 3*A*). Additionally, the presence of the ion at *m/z* 285.3 showed that the C4 of a terminal mannose was substituted. Based on this MS2 information and the resistance to the α-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 $Man_4Xyl_2GlcNAc_2$ and $Man_3Xyl_2GlcNAc_2$.

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 $Man_5Xyl_2GlcNAc_2$ modified by three methyl groups (Figs. 3*B* and 3*C*), revealed product ions at *m/z* 364.2 and 567.2, which corresponded, respectively, to one and two *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 $Y_{3\beta}/Y_{4\alpha}/Y_{4\beta}$ (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. 3*B* and 3*C*).

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, mannose-rich glycopeptides were enriched using Con A and then samples underwent PNGase F–mediated deglycosylation in the presence of $H_2{}^{18}O$. *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/^{18}O-method and IS-CID, we identified 124 and 31 glycopeptides, respectively, with an overlap of 21 glycopeptides. Through the ^{18}O-method, six additional glycosylation sites were found that did not match the consensus motif N[X!P][S/T], suggesting spontaneous deamidation during incubation in the presence of $H_2{}^{18}O$. The IS-CID method yielded a considerably lower number of glycopeptide identifications than the PNGase F/^{18}O-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/FMG1–2. 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. 4*B* and 4*C* 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.

TABLE II
Glycoproteins and glycopeptides identified in C. reinhardtii

Identifier (JGI 4.3 Augustus10.2)	Protein name or conserved domain	Peptide	Source	Method	*m/z* {# HexNAc}	*z*	E-value
Cre07.g340450.t1.2 {PKHD1–1}	*G8 domain (found in disease proteins PKHD1 and KIAA1199)*	TSWSATWS**N**GSSAEYFIK	SN	^{18}O	1012.9609	2	1.2e-14
				IS	1113.0093 {1}	2	Man[b]
		GVEYEFY**N**VSLSGVSNLWR	SN	^{18}O	1111.5408	2	6.3e-10
		FYMQADDIGQL**N**ITYTDTNNQVVTR	SN	^{18}O	1462.1926	2	5.8e-12
		FTQMVQFS**N**NTAHSNMFYGLR	SN	^{18}O	832.7168	3	1.5e-03
		TITVAN**N**GTHSTATILK	SN	IS	973.0201 {1}	2	1.6e-03
		TSGGPSGIAG**N**NTVIGSAR	SN	IS	959.9818 {1}	2	4.9e-08
		YLYGAGA**N**TTAK	SN	IS	818.3947 {2}	2	5.5e-03
		TSDALLNTDTTPATFWITNP**N**NTVR	SN[a]	^{18}O	1383.6801	2	1.9e-09
				IS	1483.7289 {1}	2	2.1e-07
Cre06.g279700.t1.2	*G8 domain (found in disease proteins PKHD1 and KIAA1199)*	STFDPTDPA**N**SSLPVK	SN	^{18}O	839.9061	2	1.6e-11
Cre16.g676150.t1.1 {MSD3/MnSOD3}	Mn superoxide dismutase	WG**N**ATALLDSLR	SN[a]	^{18}O	660.3435	2	4.5e-04
Cre17.g718500.t1.2 {MMP1}	Matrix metalloproteinase, gamete lytic enzyme (G-lysin)	IK**N**TTAGGYDSGLTLDFHK	SN	^{18}O	1021.0128	2	3.3e-15
				IS	1121.0601 {1}	2	3.8e-07
		NTTAGGYDSGLTLDFHK	SN	^{18}O	900.4224	2	2.6e-14
				IS	1000.47 {1}	2	1.6e-08
		R**N**DTYDDWWDLSK	SN	^{18}O	858.8741	2	5.4e-08
				IS	958.9223 {1}	2	1.9e-07
Cre17.g718468.t1.1 {MMP2}	*Peptidase M11 superfamily domain (Gametolysin)*	IK**N**TTAGGYDSGLTTDFHK	SN	^{18}O	676.9949	3	1.0e-05
		NTTAGGYDSGLTTDFHK	SN	^{18}O	894.4036	2	9.3e-10
Cre09.g388350.t1.2 {MMP11}	*Peptidase M11 superfamily domain (Gametolysin)*	LLVHEV**N**ATMDNNLQLYR	SN[a]	^{18}O	1073.5469	2	9.4e-15
Cre07.g324500.t1.1	*Peptidase M11 superfamily domain (Gametolysin)*	TMVLVHSY**N**GTAVSSYQR	SN	^{18}O	672.6630	3	5.9e-04
Cre02.g133500.t1.2	*Peptidase M11 superfamily domain (Gametolysin)*	VLVHFF**N**GSASER	SN	^{18}O	733.3726	2	1.5e-07
Cre13.g596600.t1.1	*Peptidase M11 superfamily domain (Gametolysin)*	VFVHEF**N**ETADNKPSDQDNPPLIMAVLDVK	SN	^{18}O	1129.2257	3	9.3e-10
Cre13.g596550.t1.2	*Peptidase M11 superfamily domain (Gametolysin)*	IFIHEF**N**ETADNNPTDDSYPPLIR	SN	^{18}O	941.1145	3	8.4e-03
Cre14.g625850.t1.2	*Peptidase M11 superfamily domain (Gametolysin)*	IYIHNF**N**ATLR	SN	^{18}O	682.8665	2	6.0e-05
Cre60.g792000.t1.1	Matrix metalloproteinase	VWVHEY**N**ETANGLTANLK	SN	^{18}O	1031.5098	2	1.5e-11
				IS	1131.5579 {1}	2	3.0e-08
Cre01.g011300.t1.1	*Serine carboxypeptidase domain*	GFIT**N**ATGIATMFDTR	SN[a]	^{18}O	859.9225	2	8.8e-09
Cre05.g242750.t1.2	*Multiple peptidase S8 family domains*	NPDSIAFIAAGN**N**GSDALTPGGSIGTPATAK	SN[a]	^{18}O	1444.7141	2	6.8e-11
Cre12.g513400.t1.2	*Multiple glycosyl hydrolase family 81 domains*	**N**ISITAAEGFVSR	SN	^{18}O	684.3572	2	5.5e-10
		LNAAGTGN**N**ASLVYDTTWGGLIVYK	SN	^{18}O	1301.1613	2	1.3e-09
Cre09.g400850.t1.2	*Carbohydrate binding domains (F5/8 type C, WSC, C- and H-type lectin)*	LTL**N**MSDIVGMR	SN	^{18}O	676.8456	2	2.8e-06
Cre06.g309950.t1.2	*Multiple C-type lectin (CTL)/C-type lectin-like (CTLD) domains*	YLVTIFD**N**ATYAR	SN	^{18}O	775.3936	2	5.1e-09
Cre05.g245259.t1.1	*Multiple C-type lectin-like domain (CTLD), GH18 chitinase-like domains*	TFTQLSPWLDLAGSPFYIDTS**N**TTTR	SN	^{18}O	1468.2206	2	1.0e-15
		QVFVTVGSG**N**DTIR	SN	^{18}O	748.3880	2	3.4e-05
		SGAVFGGDSIPV**N**DTSLIQPPASIGR	SN	^{18}O	1279.6560	2	6.8e-10

Table II—*continued*

Identifier (JGI 4.3 Augustus10.2)	Protein name or conserved domain	Peptide	Source	Method	*m/z* {# HexNAc}	*z*	E-value
(a) Cre14.g631100.t1.2 (b) Cre14.g631150.t1.2	(a) *Carbohydrate binding domains (F5/8 type C, WSC, C-type lectin), scavenger receptor Cys-rich domain, peptidase C1A domain* (b) *Peptidase C1 superfamily domain, WSC domain*	VG**N**ASVTSTSDSLYGNTLVWK[a] LALQPSSLFF**N**GSAEWK[a,b]	SN SN	^{18}O ^{18}O	1101.5472 949.4857	2 2	2.1e-11 2.4e-03
Cre04.g226050.t1.2	*Sulfatase superfamily domain*	LNQLF**N**LSSDEAEVNDLLLK	SN	^{18}O	1139.5939	2	1.8e-13
Cre04.g226600.t1.2	*Sulfatase superfamily domain*	SDKPNFIVIITDDQDDIL**N**STHPYYMPALNR	SN	^{18}O	902.9440	4	4.0e-05
Cre10.g432600.t1.2	*Sulfatase superfamily domain*	YTHNN**N**VTSNIEPHGSFWK	SN	^{18}O	750.0142	3	1.2e-04
		ALP**N**ATLGWGDTFFGTAAR	SN[a]	^{18}O	984.9886	2	1.4e-07
Cre10.g431800.t1.1.	*Sulfatase*	LTHNH**N**VTSNQAPQGGWK	SN	^{18}O	664.6580	3	9.7e-06
Cre02.g097000.t1.1	*Dihydropyrimidinase domain*	VIGEPVASGLALDESPVWDS**N**FTR	SN[a]	^{18}O	1281.6359	2	3.4e-14
		ALASGVLQLVATDHAVF**N**SSQK	SN[a]	^{18}O	753.7356	3	1.7e-06
		LLAA**N**VTGPEGHPLSRPPALEGEATGR	SN[a]	^{18}O	905.1431	3	5.1e-07
Cre01.g028850.t1.1	*Rhodanese homology domain (RHOD)*	**N**TTFLDIR	SN	^{18}O	491.7594	2	9.8e-05
		TEA**N**FTASHIAGAVNIPK	SN	^{18}O	615.3191	3	3.5e-07
				IS	1022.5206 {1}	2	3.3e-06
		APDVLTVSAADALALLDGK**N**TTFLDIR	SN	^{18}O	935.1729	3	7.4e-04
		SGVYAGAVQLTRP**N**ITLR	SN	^{18}O	640.3583	3	2.7e-05
		GGQVHAGPVGIFTP**N**LTLMTDPR	SN	^{18}O	794.4139	3	3.4e-07
Cre02.g077750.t1.1 {FAP211}	Flagellar associated protein	**N**MTGQGTLLPAGPLIWYDSPNFAANNK	SN	^{18}O	1447.2091	2	4.9e-13
(a) Cre02.g077850.t1.1 {FAP212} (b) Cre02.g077800.t1.1	(a) *No conserved domains* (b) *No conserved domains*	**N**QTAINSLVDDIQNTYAK[a,b] VVEDFVVTYQ**N**QTIGDQDPADLQK[a,b]	SN SN	IS IS	1106.0414 {1} 1463.2085 {1}	2 2	9.7e-14 Man[c]
		ALGV**N**ATAIVVR[a,b]	SN	^{18}O	593.8557	2	1.6e-08
		A**N**DSMVTVPLFFK[b]	SN	^{18}O	736.3755	2	4.5e-05
				IS	836.4227 {1}	2	1.8e-04
(a) Cre16.g661750.t1.1 (b) Cre16.g661850.t1.2	(a) *CaMK II association domain* (b) *CaMK II association domain*	WLIVEHHSSAMPENEAALVMDAFVQWNDALATL**N**ASK	SN	^{18}O	1028.7544	4	5.2e-03
Cre13.g569550.t1.2	*Polycystin cation channel protein domain, leucine-rich repeat receptor-like protein kinase domain*	**N**TSHTDYAAALA**N**VTAR	SN	^{18}O	891.4289	2	1.3e-11
(a) Cre08.g383400.t1.2 (b) Cre08.g383600.t1.2	*Multiple scavenger receptor cysteine-rich domains*	LVSSG**N**QTALDAAAR	SN	^{18}OIS	738.8793 940.4707 {2}	2 2	5.6e-07 Man[d]
Cre17.g706700.t1.2	*TRP superfamily domain, galactose oxidase (central domain)*	DSLWVFGGLD**N**FTR	SN	^{18}O	815.3929	2	5.9e-09
Cre01.g052750.t1.2	*TRP superfamily domain*	SGHTALPY**N**ESVLVFGGSVI**N**K.T	SN	^{18}O	765.7286	3	4.6e-05
Cre45.g788400.t1.1	*DUF3707 (Pherophorin) domain*	ETSQAF**N**VTLQLNR	SN	^{18}O	812.4184	2	7.1e-05
Cre11.g476250.t1.1	*C3HC4-type RING-finger domain*	GAQP**N**VTYSGGGVYYAPPR	SN[a]	^{18}O	978.9732	2	2.3e-12
Cre06.g260650.t1.2	*DUF288 family domain*	GFPLTQIT**N**VTTR	SN	^{18}O	725.8926	2	2.2e-09
Cre13.g596800.t1.1 {FAS7}	*Multiple fasciclin domains*	QVVETALPGY**N**LTITK	SN	^{18}O	875.4777	2	1.8e-07
Cre16.g694200.t1.2	*No conserved domains*	VWVHES**N**ETATGAPAAPGSHTLIR	SN	^{18}O	835.4196	3	3.8e-07
Cre09.g398900.t1.2 {GP1}	*No conserved domains*	TVGSVA**N**VTIR	SN	^{18}O	560.3170	2	5.1e-04
Cre06.g258800.t1.2 {GP2}	*No conserved domains*	**N**TTWTAPAGTTVR	SN	IS	891.4330 {2}	2	5.4e-04
Cre12.g487950.t1.2	*No conserved domains*	ISDGVST**N**YSEPFDIR	SN	^{18}O	901.9235	2	1.2e-08
Cre09.g401050.t1.2	*No conserved domains*	AG**N**VTFSNCEMYNTGAR	SN	^{18}O	947.9026	2	2.7e-04
Cre02.g080150.t1.2	*No conserved domains*	VLGEL**N**ATSWTEAAGAELAALPLEAVATAAR	SN[a]	^{18}O	1535.3008	2	3.8e-10
Cre02.g122550.t1.1	*No conserved domains*	FYFN**N**GTNYR	SN	^{18}O	649.7910	2	7.3e-04
Cre06.g308050.t1.1	*No conserved domains*	LA**N**GTTVDGPAYFSR	SN	^{18}O	786.3850	2	4.4e-05
Cre17.g700700.t1.2	*No conserved domains*	RPPD**N**ATAGLAVDGLYHDNR	SN	^{18}O	719.0178	3	1.4e-05
Cre07.g333100.t1.2	*No conserved domains*	VDLGDSYYL**N**NSFK	SN[a]	^{18}O	819.3798	2	5.5e-09
Cre17.g708750.t1.1	*No conserved domains*	GGAIVL**N**QTTDSTK	SN	IS	905.9520 {2}	2	Man[e]

Table II—*continued*

Identifier (JGI 4.3 Augustus10.2)	Protein name or conserved domain	Peptide	Source	Method	*m/z* {# HexNAc}	*z*	E-value
Cre02.g121650.t1.2 {CrSTT3A}	*Oligosaccharyltransferase STT3 subunit*	TVIVDNNTW**N**TSHIATVGR	TCE	^{18}O	1051.0349	2	1.5e-10
		VASWWDYGYQTTAMA**N**R.T	TCE	^{18}O	1011.9531	2	7.3e-15
Cre07.g330100.t1.1 {CrSTT3B}	*COG1287 (uncharacterized membrane protein, required for N-linked glycosylation)*	IMSWWDYGYQITAMG**N**R.T	TCE	^{18}O	1047.9698	2	6.7e-11
		TVIVDNNTW**N**NTHIATVGR	TCE	^{18}O	710.0291	3	3.3e-03
Cre05.g233250.t1.2 {CrUGGC}	*UDP-glucose:glycoprotein glucosyltransferase*	F**N**ATSYLLEALEFLAEEEPALVWK	TCE	^{18}O	1393.7101	2	3.0e-09
Cre07.g330750.t1.2	*DUF3707 (Pherophorin) domain*	ATIAASTFG**N**VSK	TCE	^{18}O	635.3322	2	4.2e-07
Cre17.g705500.t1.2	*DUF3707 (Pherophorin) domain*	LYVLPEIANSAAITSVMFN**N**K.T	TCE	^{18}O	1149.6047	2	5.3e-13
Cre07.g326600.t1.1	*Thioredoxin (TRX)-like domain*	QLSG**N**VSAELAALDAR	TCE	^{18}O	809.4216	2	1.2e-07
Cre17.g722500.t1.1	*Lysosomal cystine transporter domain, PQ loop repeat domain*	Acetyl-ADLL**N**TTSVVLK	TCE	^{18}O	659.8724	2	1.2e-04
Cre14.g611850.t1.1	*Endomembraneprotein 70 (EMP70)*	IIQV**N**LTTADPVPVAPGAK	TCE	^{18}O	954.0396	2	1.6e-08
Cre10.g463300.t1.1	*No conserved domains*	GGV**N**SSHVVAQEAGYLYR	TCE	^{18}O	637.3177	3	1.8e-03
Cre16.g656050.t1.1	*No conserved domains*	AV**N**TTATAPPPAPSVRPQAPAPDVTGG… LEGLEEAATTVAAAASAAASVVDAAAK	TCE	^{18}O	1246.3956	4	3.8e-03
Cre49.g789700.t1.1	*No conserved domains*	GPY**N**VTVVLK	TCE	^{18}O	546.8153	2	1.1e-03
(a) Cre12.g516600.t1.2 (b) Cre12.g517000.t1.2 {MAPKKK7} (c) Cre12.g516650.t1.2 {STPK6}	(a) *Protein kinases (PKs), catalytic domain* (b) *Protein kinases (PKs), catalytic domain* (c) Serine/threonine protein kinase	LE**N**CTLVVSAEELR	PM	^{18}O	818.4133	2	2.6e-05
Cre05.g238100.t1.1	Protein tyrosine kinase	**N**MTISGPLDSPIK	PM	^{18}O	688.3521	2	1.4e-04
Cre02.g090050.t1.2	Flagellar associated protein	ISLASDGGFV**N**ATYT**N**GTAYILGAK	PM	^{18}O	1255.6277	2	1.0e-15
Cre17.g712900.t1.1	*No conserved domains*	**N**VTAAQLGGNDFDINPTAV**N**R.T	PM	^{18}O	1097.0371	2	3.7e-11
Cre03.g152250.t1.2	*No conserved domains*	IYP**N**YSDPSIYLQLR	PM	^{18}O	922.9720	2	3.6e-05
Cre05.g241350.t1.2	*Eye pigment and drug resistance transporter subfamily domain (ABCG_EPDR), ABC-type cobalt transport system (CbiO)*	L**N**SSAPADALPR	CP	^{18}O	607.8179	2	1.1e-04
Cre09.g409900.t1.2	*DUF3707 (Pherophorin) domain*	T**N**CSVSEVDQLELFIAPETLNSVYK	CP	^{18}O	953.8032	3	2.3e-05
(a) Cre29.g778950.t1.1 {FMG1-1} (b) Cre31.g780700.t1.2 {FMG1-2}	(a) Flagella membrane glycoprotein, major form (b) Flagella membrane glycoprotein, minor form	LFGVPV**N**ASAYGTAVQLLAYDYYVK[a]	PM	^{18}O	1363.2107	2	1.1e-12
				IS	1463.2527 {1}	2	1.3e-10
		LLSAG**N**FSAGDTVNIKPEQAELR[a]	PM, SN	IS	1317.1770 {1}	2	3.9e-08
		LLG**N**NSDVYTGGDTFNFK[a]	PM, SN	^{18}O	982.9630	2	1.6e-12
				IS	1083.0070 {1}	2	9.5e-08
		FFDGL**N**TSVAGR[a]	PM	^{18}O	643.8174	2	4.7e-10
				IS	743.8636 {1}	2	8.8e-04
		ADCDAVFVFSGAG**N**TTK[a]	PM	^{18}O	881.8932	2	1.0e-08
		LAAQVVNPALFANA**N**LTSATAITVR[a,b]	TCE, PM, SN	^{18}O	1265.2036	2	1.0e-15
		NAFSYFDLN**N**GTK[b]	TCE	^{18}O	747.3456	2	1.5e-07
		LLG**N**TTDVYASGDTFNFK[b]	TCE, SN[a]	^{18}O	983.4730	2	5.0e-11
Cre12.g546550.t1.1 {FEA1}	Fe-assimilating protein	FASYITA**N**GSVEPLHDSILAGK	TCE, SN	^{18}O	1147.0874	2	5.7e-08
				IS	1247.1337 {1}	2	Man[f]
		AAMAAG**N**FTEALSIYSTGK	SN	^{18}O	953.4623	2	1.0e-15
				IS	1053.5095 {1}	2	7.7e-14
Cre12.g546600.t1.2 {FEA2}	Fe-assimilating protein	D**N**GTLSSAVY**N**ASR	TCE	^{18}O	730.8359	2	1.4e-09
		K_D**N**GTLSSAVY**N**ASR	SN	^{18}O	794.8843	2	1.5e-08
				IS	994.9776 {2}	2	1.1e-04
Cre09.g393150.t1.1 {FOX1}	Multicopper ferroxidase	GVDLMVPLYWQVVDENSSPFLDLNVEAAQL**N**VTK	PM	^{18}O	1269.6463	3	5.8e-09
		LGDGGALAAQLAA**N**ATEMTALVTDPVFMEHMLK	TCE	^{18}O	1130.5664	3	7.1e-06

Table II—*continued*

Identifier (JGI 4.3 Augustus10.2)	Protein name or conserved domain	Peptide	Source	Method	*m/z* {# HexNAc}	*z*	E-value
Cre10.g439900.t1.1 {HSP70G}	ER-located HSP110/SSE-like protein	Y**N**TSGQISLR	TCE	^{18}O	571.2894	2	1.7e-06
		HLDADEAVALGAGLFAA**N**LSTSFR	TCE	^{18}O	1225.1179	2	1.0e-15
		FSLAY**N**ASTHHGLPPGVK	TCE, SN[a]	^{18}O	949.9884	2	1.6e-09
		IIEVPV**N**ETDTATGAEGAGADADTK	TCE, SN, CP	^{18}O	1224.5840	2	1.0e-15
		IIEVPV**N**ETDTATGAEGAGADADTKAEK	TCE, SN,[a] CP	IS	1488.7189 {1}	2	6.6e-09
		AGGDKAEEAEGEAKEGADAGAGSA**N**AT**N**AS**N**STA… NAKPATVIK	SN[a]	^{18}O	1002.7274	2	4.1e-08
(a) Cre06.g272250.t1.1 (b) Cre03.g155300.t1.1	(a), (b) *DUF3707 (Pherophorin) domain*	IMAF**N**TTPATDYSK[a]	TCE	^{18}O	781.8712	2	1.3e-04
		KL**N**YSYPDFDGPEK[a,b]	TCE	^{18}O	838.3884	2	6.1e-09
		LGYQLSGVAP**N**QTMFNFDLAK[b]	TCE	^{18}O	1159.0795	2	1.3e-11
		VV**N**VTGIDYDLEEEALCDR[a,b]	CP	^{18}O	1107.0132	2	3.1e-11
Cre17.g741000.t1.1 {TRAPA1}	*Translocon-associated protein (TRAP), alpha subunit*	Y**N**ISAVMGSLNNAQDFR	TCE, SN,[a] CP	^{18}OIS	951.9520 1051.9949 {1}	2 2	3.1e-11 man[g]
Cre17.g715300.t1.1 {PKD2}	*Polycystin cation channel protein*	GTTLYNPDYDDV**N**GTK	TCE, PM, SN	^{18}O	888.3962	2	6.7e-12
Cre13.g592050.t1.1 {DAL1}	Allantoinase	IVAVSD**N**TTVDSPHMLNFK	TCE	^{18}O	697.6844	3	1.5e-03
		VEWEGIW**N**GTR	TCE, SN[a]	^{18}O	675.3238	2	3.8e-04
Cre06.g294400.t1.2	Nicastrin	TLHSLAGSPETPALEV**N**R.T	PM	^{18}O	632.3325	3	1.4e-03
		FVYNYLG**N**LTAAPLPADR	TCE, PM	^{18}O	999.5161	2	4.6e-06
(a) Cre01.g067150.t1.1 (b) Cre01.g067059.t1.1	(a), (b) *EGF-like domain*	**N**TTEWSLDPLDSFPAPNFLTDK	TCE, SN	^{18}OIS	1256.0959 1356.1448 {1}	2 2	8.6e-13 2.3e-04
Cre45.g788350.t1.2 {GAS31}	Hydroxyproline-rich cell wall protein pherophorin	LSYV**N**TSSLVGVR	TCE, SN	^{18}O	699.3804	2	6.9e-10
Cre06.g304500.t1.1 {ZYS3–2}	Zygote-specific protein	L**N**VSQIEKPHEVPEAMLADIEK	TCE, SN	^{18}O	1247.1454	2	1.0e-10
				IS	1347.1940 {1}	2	6.1e-08
Cre15.g635079.t1.1	*Periplasmic binding protein (PBP) superfamily domain*	**N**LTGYGYSGPLLR	TCE, PM	^{18}O	707.3672	2	1.3e-07
Cre05.g244950.t1.1	*Translocon-associated protein beta (TRAPB) domain*	YTDVLPE**N**ATLVEGSLEVDFGK	TCE, CP	^{18}O	1200.0963	2	1.0e-15
Cre01.g042550.t1.1	*DUF1620*	QTLFVDLPA**N**GTALK	TCE, CP	^{18}O	795.9345	2	1.5e-04
Cre07.g321400.t1.2 {FAP113}	Flagellar associated protein	KNVITVPTQIS**N**ITIEFK	SN	^{18}O	1024.5787	2	1.7e-07
				IS	1124.6289 {1}	2	1.9e-12
		ITYATTAAAVTNA**N**LSSYK	SN, PM	IS	1082.0464 {1}	2	2.2e-10
Cre09.g394200.t1.2 {FAP102}	Flagellar associated protein	SSNLA**N**ATFWVASK	TCE	^{18}O	749.8744	2	1.8e-10
		MTS**N**ITIADIPVQR	TCE, SN	^{18}O	781.4117	2	1.5e-03
				IS	982.9952 {2}	2	Man[h]
Cre14.g612650.t1.2	*No conserved domains*	VLDYVAYG**N**DSANPLPAGSVSLVPLDGPAK	TCE, SN[a]	^{18}O	1501.7690	2	5.2e-10
				IS	1601.8141	2	2.0e-05
Cre06.g283200.t1.1	*No conserved domains*	DINQLG**N**SSTVDLVAGK	TCE, SN[a]	^{18}O	867.4476	2	9.7e-11
Cre02.g138500.t1.1	*No conserved domains*	LPIA**N**ATAFTDGLR	TCE, SN[a]	^{18}O	731.8951	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: $1 \cdot 10^{-2}$) 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.

[a] Peptides were exclusively identified in samples derived from iron-deficient conditions.

[b–h] 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.

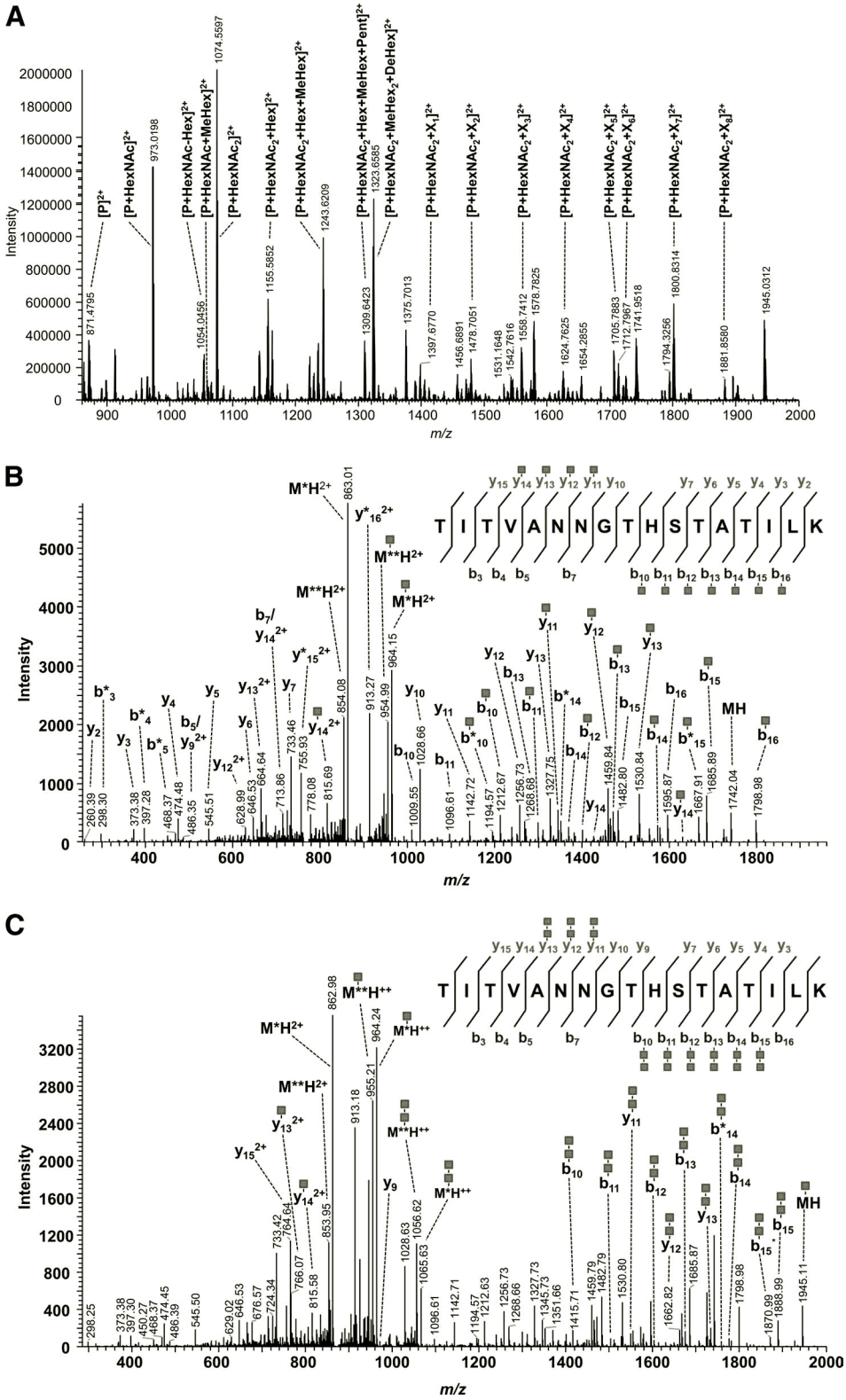


FIG. 4. **Mass spectrometric analysis of the PKHD1–1 peptide TITVANN*GTHSTATILK.** *A*, MS1 spectrum of doubly charged, *N*-glycosylated peptide ions (P) differing in glycan composition and chain length. Distinctive peak clustering suggests multiple branching of a complex-type glycan and/or the co-elution of glycoforms. The presence of *p* + 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

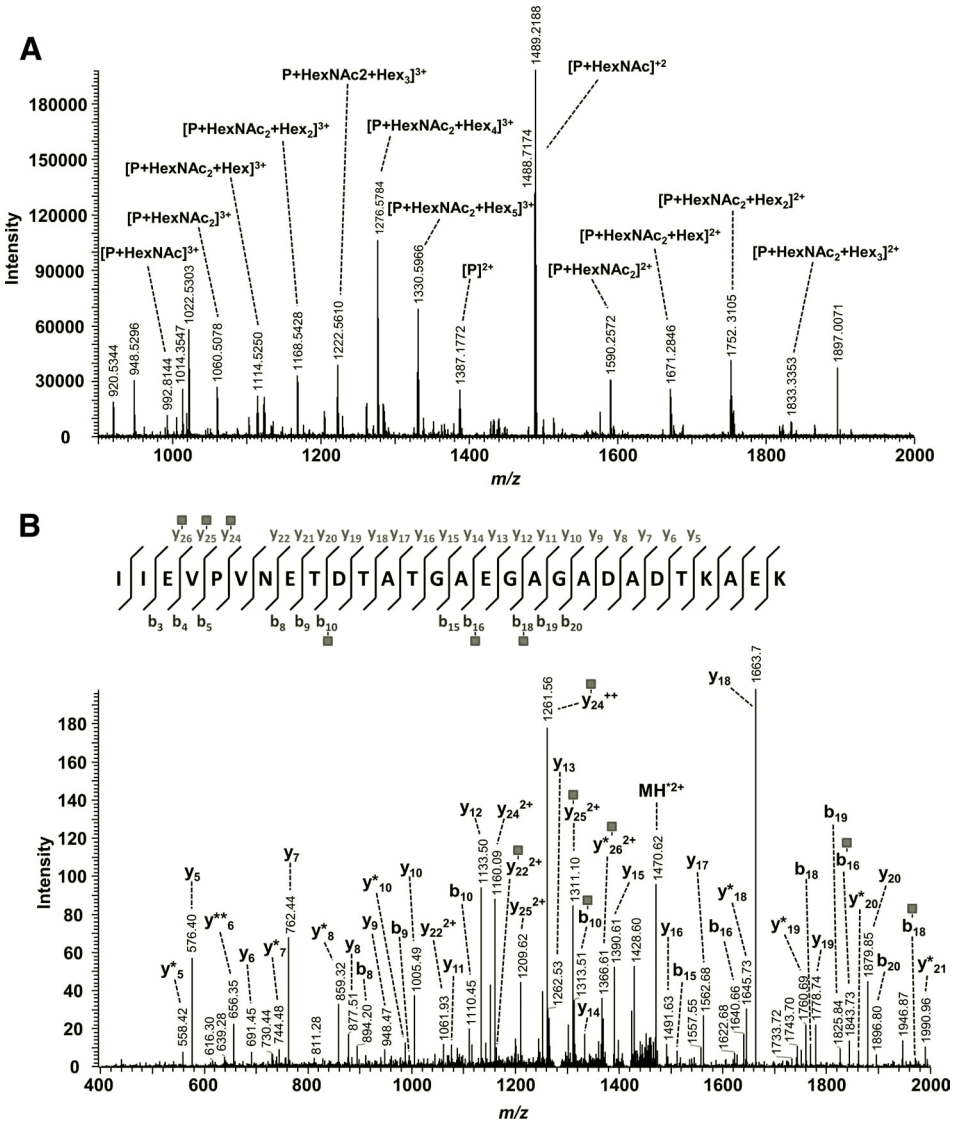


FIG. 5. **Mass spectrometric analysis of the HSP70G peptide IIEVPVN*ETDTATGAEGAGADADTKAEK modified by a $HexNAc_2Hex_5$ glycan.** *A*, in-source CID MS1 spectrum of several doubly and triply charged peptide ions (P) differing in glycan chain length. *B*, CID spectrum of IIEVPVN*ETDTATGAEGAGADADTKAEK (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 (−203 Da). The resulting composite spectrum contains several b- and y-type ions showing modifications by HexNAc. ■, HexNAc; *, loss of water or ammonia.

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. 5*A*) 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 pep-

from *m/z* 1389 to 1419, please refer to the supplemental material. Possible glycan compositions: X_1: Hex+$MeHex_2$+Pent; X_2: Hex_2+$MeHex_2$+Pent; X_3: Hex_2+$MeHex_2$+$DeHex_2$ or Hex+$MeHex_3$+DeHex+Pent or $MeHex_4$+$Pent_2$; X_4: Hex_2+$MeHex_2$+$DeHex_2$+Pent or Hex+$MeHex_3$+DeHex+$Pent_2$ or $MeHex_4$+$Pent_3$; X_5: Hex_3+$MeHex_2$+$DeHex_2$+Pent or Hex_2+$MeHex_3$+DeHex+$Pent_2$ or Hex+$MeHex_4$+$Pent_3$; X_6: Hex_2+$MeHex_3$+$DeHex_2$+Pent or Hex+$MeHex_4$+DeHex+$Pent_2$ or $MeHex_5$+$Pent_3$; X_7: Hex_2+$MeHex_4$+$DeHex_2$+Pent or Hex+$MeHex_5$+DeHex+$Pent_2$ or $MeHex_6$+$Pent_3$; X_8: Hex_3+$MeHex_4$+$DeHex_2$+Pent or Hex_2+$MeHex_5$+DeHex+$Pent_2$ or Hex+$MeHex_6$+$Pent_3$. *B*, multistage-activation (MSA)-CID spectrum of TITVANN*GTHSTATILK (precursor *m/z* 973.0197(MH^{2+})) modified by one HexNAc residue. *C*, MSA-CID spectrum of the same peptide as in *B* differing by one additional core HexNAc residue (precursor *m/z* 1074.5597 (MH^{2+})). The majority of ions consistent with those in *B* are not annotated. ■, HexNAc; *, loss of water/ammonia.

TABLE III
Glycoproteins conserved in C. reinhardtii and humans

Identifier (JGI 4.3 Augustus 10.2)	Best BLASTp hit (human)	Accession number	BLASTp E-value	Conserved glycosylation sites
Cre07.g340450.t1.2 {PKHD1-1}	Polycystic kidney and hepatic disease 1 (autosomal recessive)-like 1	EAW91931.1	0.0	4 out of 8
Cre01.g011300.t1.1	Vitellogenic carboxypeptidase-like protein (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	Dihydropyrimidinase (DPYS)	NP_001376.1	1e-141	-
(a) Cre08.g383400.t1.2 (b) Cre08.g383600.t1.2	(a) Deleted in malignant brain tumor 1 protein isoform (b) DMBT1/8kb.2 protein	(a) NP_015568.2 (b) CAB56155.1	(a) 6e-47 (b) 3e-48	-
Cre02.g121650.t1.2 {CrSTT3A}	STT3A/STT3B	NP_689926.1/NP_849193.1	0.0/0.0	2 out of 2
Cre07.g330100.t1.1 {CrSTT3B}	STT3A/STT3B	NP_689926.1/NP_849193.1	0.0/0.0	2 out of 2
Cre09.g397900.t1.1	Cleft lip and palate 1 (CLPT1)	NP_001285.1	4e-140	-
Cre14.g611850.t1.1	SM-11044 binding protein/transmembrane 9 superfamily member 3 (TM9SF3)	NP_004818.2	0.0	1 out of 1
Cre05.g241350.t1.2	ABC transporter (ABCG2)	AAG52982.1	7e-41	-
Cre09.g393150.t1.1 {FOX1}	Hephaestin	NP_001124332.1	2e-147	-
Cre10.g439900.t1.1 {HSP70G}	Hypoxia upregulated 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.

tide sequence and glycosylation site (Fig. 5*B*). BLASTp revealed a high similarity of HSP70G to human HYOU1 (E-value: $1e^{-111}$; Table III). However, none of the *N*-glycosylation sites of HSP70G aligned with those determined for HYOU1 (supplemental File S8).

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 glucose: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 confirmed (STT3A, STT3B, TM9SF3) *N*-glycosylation sites matching those determined for *C. reinhardtii* glycoproteins.

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 $Glc_3Man_9GlcNAc_2$-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

TABLE IV
Sequences encoding putative proteins involved in the N-glycan pathway in C. reinhardtii

Putative function	Protein name	Gene name (JGI v4.3)	Gene location	Best hit in *A. thaliana*[a]	SignalP prediction	TMHMN prediction	Predicted domains
ER cytosolic enzymes							
N-acetylglucosamine phosphostransferase	CrALG7	Cre16g663100	Chromosome_16: 1990552–1993342	At3g57220 (43%)	Signal peptide	10 TMD	PF00953
β-1,4-*N*-acetylglucosaminyl transferase	CrALG13	Cre13g585850	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-mannosyl transferase	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: 4322467–4325535	At2g44660 (44%)	None	10 TMD	PF03155
Calnexin	CrCLNX	Cre07g357900	Chromosome_7: 6143630–6147986	At5g61790 (35%)	Signal peptide	1 TMD	PF00262
Calreticulin	CrCLRT	Cre01g038400	Chromosome_1: 5293813–5297125	At1g09210 (59%)	Signal peptide	None	PF00262
Dolichol-phosphate mannosyltransferase	CrDPM1	Cre03g150950	Chromosome_3: 438314–439077	At1g20575 (52%)	None	None	PF00535
Dolichol-phosphate glucosyltransferase	CrALG5	Cr16g652850	Chromosome_16: 679141–682477	At2g29630 (38%)	Signal peptide	1 TMD	PF00535
Glucosidase I	CrGSI	Cre13g579750	Chromosome_13: 2459268–2461989	At1g67490 (36%)	None	None	PF03200
Glucosidase II, α-subunit	CrGSIIA	Cre03g190500	Chromosome_3: 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	Cre05g233250	Chromosome_5: 672374–689328	At1g71220 (15%)	Signal peptide	None	PF06427
Oligosaccharyltransferase complex subunits							
DDPGT subunit[b]	CrDGL1	Cre14g614100	Chromosome_14: 987242–990915	At5g66680 (43%)	Signal peptide	2 TMD	PF03345
Ribophorin I	CrRPN1	Cre12g523300	Chromosome_12: 4550885–4555389	At1g76400 (38%)	Signal peptide	1 TMD	PF04597
Ribophorin II	CrRPN2	Cre08g368450	Chromosome_8: 1593746–1596364	At4g21150 (10%)	Signal peptide	1 TMD	PF05817
DDPGT subunit[b]	CrDAD1	Cre02g108400	Chromosome_2: 4617959–4619399	At1g32210 (48%)	Signal anchor	3 TMD	PF02109
DDPGT subunit[b]	CrSTT3B	Cre07g330100	Chromosome_7: 2282859–2291158	At1g34130 (57%)	None	11 TMD	PF02516
DDPGT subunit[b]	CrSTT3A	Cre02g121650	Chromosome_2: 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	CrMANI	Cre07g336600	Chromosome_7: 3304731–3311367	At1g30000 (28%)	Signal peptide	None	PF01532
β-1,2-xylosyltransferase	CrXYLT	Cre02g126700	Chromosome_2: 6833318–6837848	At5g55500 (17%)	Signal peptide	None	DUF563 (PF04577)
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.
[a] Pairwise alignments were done using full-length amino-acid sequences with Clustal W1.8 and percentages of identity are given into parenthesis.
[b] DDPGT is the abbreviation for dolichol-diphospho-oligosaccharide protein glycosyltransferase.

binding and release of monoglucosylated glycoproteins with calnexin and calreticulin, two ER-resident lectin-like chaperons 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 maturated 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 Ca^{2+} 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 (MANIIC, NP_06706.2), 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 $Man_5GlcNAc_2$ 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 xylosyltransferase 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-xylose 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 α1,3-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 $Glc_3Man_9GlcNAc_2$-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 $Man_2GlcNAc_2$ to $Man_5GlcNAc_2$ structures representing almost 70% of the total *N*-glycan population. Although some ALGs were not clearly identified in the *C. rheinardtii* genome, the identification of large oligomannosides up to $Man_9GlcNAc_2$ (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 $Man_5GlcNAc_2$ (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 $Man_3GlcNAc_2$ to $Man_5GlcNAc_2$ 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*GTHSTATILK, 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 $Man_8GlcNAc_2$ and $Man_9GlcNAc_2$ containing 6-O-methyl mannose and substituted by one or two xylose residues, with one xylose located on the chitobiose 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 $Man_5GlcNAc_2$ in the Golgi apparatus by Golgi-residing mannosidases such as

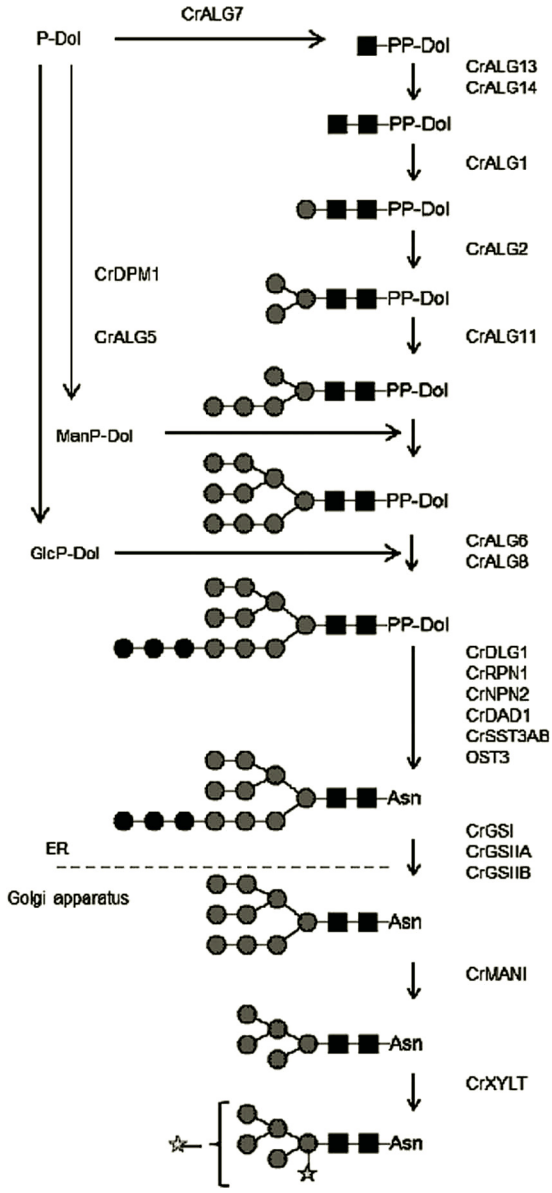


FIG. 6. **Proposed *N*-glycosylation biosynthesis pathway in *C. reinhardtii*.** The proposed pathway is based on the major *N*-glycan structures found according to the *in silico* analysis. *N*-glycan structures have been drawn using the symbolic nomenclature adopted by the Consortium for Functional Glycomics (101). ■, *N*-acetylglucosamine; ●, mannose; ☆, xylose.

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 $Man_2GlcNAc_2$ and $Man_1GlcNAc_2$ 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 α(1,3)-mannose arm of $Man_5GlcNAc_2$ 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). *N*-glycosylation sites corresponding to those of *C. reinhardtii* have already been reported for STT3 from *Saccharomyces cerevisiae* and human STT3A/STT3B (13, 86–89) (supplemental File S9). In yeast, glycosylation of N539 (corresponding to *N*-glycosylated N595/N986 in CrSTT3A/CrSTT3B) was shown to be essential for the enzymatic function of STT3. N591 and N582 of CrSTT3A and CrSTT3B, respectively, were not found to be glycosylated, although they were located within consensus motifs of *N*-glycosylation. The same observation was

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 glycoproteomic 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-carbonic 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 INSA Rouen, CNRS, Region Haute-Normandie, EFRD (No. 31708), and Labex SynOrg (ANR-11-LABX-0029).

S This article contains supplemental material.

¶¶ To whom correspondence should be addressed: Dr. Muriel Bardor, Université de Rouen, 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.

** 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 analysis 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., Kilz, 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,** 2557–2566
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 glycans. *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.* **2011,** 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. Balshüsemann, D., and Jaenicke, L. (1990) 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, D. J., Parsons, T. B., Fairbanks, A., and Tekoah, Y. (2011) Unique *N*-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga Porphyridium sp. *J. Biol. Chem.* **286,** 21340–21352
12. Baiet, B., Burel, C., Saint-Jean, B., Louvet, R., Menu-Bouaouiche, L., Kiefer-Meyer, M.-C., Rivet, E., Castel, H., Lefebvre, T., Carlier, A., Cadoret, J.-P., Lerouge, P., and Bardor, M. (2011) *N*-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase 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., Schiess, R., Aebersold, R., and Watts, J. D. (2009) Mass-spectrometric identification and relative quantification of *N*-linked cell surface glycoproteins. *Nat. Biotechnol.* **27,** 378–386
15. Whitmore, T. E., Peterson, A., Holzman, T., Eastham, A., Amon, L., McIntosh, M., Ozinsky, A., Nelson, P. S., and Martin, D. B. (2012) Integrative analysis of *N*-linked human glycoproteomic data sets reveals PTPRF ectodomain as a novel plasma biomarker candidate for prostate cancer. *J. Proteome Res*. **11,** 2653–2665
16. Zielinska, D. F., Gnad, F., Wiœniewski, J. R., and Mann, M. (2010) Precision mapping of an in vivo *N*-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141,** 897–907
17. Yin, X., Bern, M., Xing, Q., Ho, J., Viner, R., and Mayr, M. (2013) Glycoproteomic analysis of the secretome of human endothelial cells. *Mol. Cell. Proteomics* **12,** 956–978
18. Zhang, Y., Giboulot, A., Zivy, M., Valot, B., Jamet, E., and Albenne, C. (2011) Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry* **72,** 1109–1123
19. Ruiz-May, E., Kim, S.-J., Brandizzi, F., and Rose, J. K. C. (2012) The secreted plant N-glycoproteome and associated secretory pathways. *Front. Plant Sci.* **3,** 117
20. Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., Marshall, W. F., Qu, L. H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P, Lindquist, E., Shapiro, H., Lucas, S. M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C. L., Cognat, V., Croft, M. T., Dent, R., Dutcher, S., Fernández, E., Fukuzawa, H., González-Ballester, D., González-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P. A., Lemaire, S. D., Lobanov, A. V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J. V., Moseley, J., Napoli, C., Nedelcu, A. M., Niyogi, K., Novoselov, S. V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J. P., Riaño-Pachón, D. M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S. L., Allmer, J., Balk, J., Bisova, K., Chen, C. J., Elias, M., Gendler, K., Hauser, C., Lamb, M. R., Ledford, H., Long, J. C., Minagawa, J., Page, M. D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A. M., Yang, P., Ball, S., Bowler, C., Dieckmann, C. L., Gladyshev, V. N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R. T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y. W., Jhaveri, J., Luo, Y., Martínez, D., Ngau, W. C., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I. V., Rokhsar, D. S., and Grossman, A. R. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318,** 245–250
21. Walsh, G. (2010) Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* **28,** 917–924
22. Mayfield, S. P., Manuell, A. L., Chen, S., Wu, J., Tran, M., Siefker, D., Muto, M., and Marin-Navarro, J. (2007) *Chlamydomonas reinhardtii* chloroplasts as protein factories. *Curr. Opin. Biotechnol.* **18,** 126–133
23. Rasala, B. A., Muto, M., Lee, P. A., Jager, M., Cardoso, R. M., Behnke, C. A., Kirk, P., Hokanson, C. A., Crea, R., Mendez, M., and Mayfield, S. P. (2010) Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* **8,** 719–733
24. Mayfield, S. P., Franklin, S. E., and Lerner, R. A. (2003) Expression and assembly of a fully active antibody in algae. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 438–442
25. Tran, M., Zhou, B., Pettersson, P. L., Gonzalez, M. J., and Mayfield, S. P. (2009) Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol. Bioeng.* **104,** 663–673
26. Eichler-Stahlberg, A., Weisheit, W., Ruecker, O., and Heitzer, M. (2009) Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta* **229,** 873–883
27. Walsh, G., and Jefferis, R. (2006) Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.* **24,** 1241–1252
28. Jiang, X. R., Song, A., Bergelson, S., Arroll, T., Parekh, B., May, K., Chung, S., Strouse, R., Mire-Sluis, A., and Schenerman, M. (2011) Advances in the assessment and control of the effector functions of therapeutic antibodies. *Nat. Rev. Drug Discov.* **10,** 101–111
29. Lingg, 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
30. Bardor, M., Faveeuw, C., Fitchette, A. C., Gilbert, D., Galas, L., Trottein, F., Faye, L., and Lerouge, P. (2003) Immunoreactivity in mammals of two typical plant glyco-epitopes, core α(1,3)-fucose and core-xylose. *Glycobiology* **13,** 427–434
31. Galili, U. (2004) Immune response, accommodation, and tolerance to transplantation carbohydrate antigens. *Transplantation* **78,** 1093–1098
32. 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. *Glycobiology* **18,** 818–830
33. Chung, C. H., Mirakhur, B., Chan, E., Le, Q. T., Berlin, J., Morse, M., Murphy, B. A., Satinover, S. M., Hosen, J., Mauro, D., Slebos, R. J., Zhou, Q., Gold, D., Hatley, T., Hicklin, D. J., and Platts-Mills, T. A. (2008) Cetuximab-induced anaphylaxis and IgE specific for galactose-α-1,3-galactose. *N. Engl. J. Med.* **358,** 1109–1117
34. Sola, R. J., and Griebenow, K. (2010) Glycosylation of therapeutic proteins: an effective strategy to optimize efficacy. *BioDrugs* **24,** 9–21
35. Sueoka, N. (1960) Mitotic replication of deoxyribonucleic acid in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U.S.A.* **46,** 83–91
36. Moseley, J. L., Allinger, T., Herzog, S., Hoerth, P., Wehinger, E., Merchant, S., and Hippler, M. (2002) Adaptation to Fe-deficiency requires remodeling of the photosynthetic apparatus. *EMBO J.* **21,** 6709–6720
37. 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
38. Séveno, M., Cabrera, G., Triguero, A., Burel, C., Leprince, J., Rihouey, C., Vézina, L.-P., D'Aoust, M.-A., Rudd, P., Royle, L., Dwek, R., Harvey, J., Lerouge, P., Cremata, J., and Bardor, M. (2008) Plant N-glycan profiling

of minute amounts of material. *Anal. Biochem.* **379,** 66–72
39. Ciucanu, I., and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131,** 209–217
40. Dell, A., Reason, A. J., Khoo, K. H., Panico, M., McDowell, R. A., and Morris, H. R. (1994) Mass spectrometry of carbohydrate-containing biopolymers. *Methods Enzymol.* **230,** 108–132
41. North, S. J., Huang, H. H., Sundaram, S., Jang-Lee, J., Etienne, A. T., Trollope, A., Chalabi, S., Dell, A., Stanley, P., and Haslam, S. M. (2010) Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals *N*-glycans of a novel size and complexity. *J. Biol. Chem.* **285,** 5759–5775
42. Wada, Y., Azadi, P., Costello, C. E., Dell, A., Dwek, R. A., Geyer, H., Geyer, R., Kakehi, K., Karlsson, N. G., Kato, K., Kawasaki, N., Khoo, K. H., Kim, S., Kondo, A., Lattova, E., Mechref, Y., Miyoshi, E., Nakamura, K., Novotny, M. V., Packer, N. H., Perreault, H., Peter-Katalinic, J., Pohlentz, G., Reinhold, V. N., Rudd, P. M., Suzuki, A., and Taniguchi, N. (2007) Comparison of the methods for profiling glycoprotein glycans—HUPO human disease glycomics/proteome initiative multi-institutional study. *Glycobiology* **17,** 411–422
43. Bardor, M., Nguyen, D. H., Diaz, S., and Varki, A. (2005) Mechanism of uptake and incorporation of the non-human sialic acid *N*-glycolylneuraminic acid into human cells. *J. Biol. Chem.* **280,** 4228–4237
44. Rohrer, J. S., Thayer, J., Weitzhandler, M., and Avdalovic, N. (1998) Analysis of the *N*-acetylneuraminc acid and *N*-glycolylneuraminic acid contents of glycoproteins by high-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). *Glycobiology* **8,** 35–43
45. Fitchette, A. C., Tranh Dinh, O., Faye, L., and Bardor, M. (2007) Plant proteomics and glycosylation, in *Plant Proteomics: Methods and Protocols* (Thiellement, H., Mehin, V., Damerval, C., and Zivy, M., eds.), Methods in Molecular Biology Series, Vol. 355, pp. 317–342, Humana Press, Totowa, NJ
46. Naumann, B., Stauber, E. J., Busch, A., Sommer, F., and Hippler, M. (2005) N-terminal processing of Lhca3 is a key step in remodeling of the photosystem I-light-harvesting complex under iron deficiency in *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **280,** 20431–20441
47. Herbik, A., Haebel, S., and Buckhout, T. J. (2002) Is a ferroxidase involved in the high-affinity iron uptake in *Chlamydomonas reinhardtii*? *Plant Soil* **241,** 1–10
48. Hsiao, H. H., and Urlaub, H. (2010) Pseudo-neutral-loss scan for selective detection of phosphopeptides and N-glycopeptides using liquid chromatography coupled with a hybrid linear ion-trap/orbitrap mass spectrometer. *Proteomics* **10,** 3916–3921
49. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467
50. Specht, M., Kuhlgert, S., Fufezan, C., and Hippler, M. Proteomics to go: Proteomatic enables the user-friendly creation of versatile MS/MS data evaluation workflows. *Bioinformatics* **27,** 1183–1184
51. Käll, L., Storey, J. D., and Noble, W. S. (2008) Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics* **24,** i42-i48
52. Angel, P. M., Lim, J., Wells, L., Bergmann, C., and Orlando, R. (2007) A potential pitfall in site mapping O-based N-linked glycosylation. *Rapid Commun. Mass Spectrom.* **21,** 674–682
53. Witman, G. B., Carlson, K., Berliner, J., and Rosenbaum, J. L. (1972) *Chlamydomonas* flagella: isolation and electrophoretic analysis of microtubules, matrix, membranes, and mastigonemes. *J. Cell Biol.* **54,** 507–539
54. Brooks, S. (2011) Species differences in protein glycosylation and their implication for biotechnology, in *Glycosylation in Diverse Cell Systems: Challenges and New Frontiers in Experimental Glycobiology* (Brooks, S., Rudd, P., and Apelmelk, B., eds.), Essential Reviews in Experimental Biology Series, Vol. 4, pp. 1–24, The Society for Experimental Biology, London, UK
55. Domon, B., and Costello, C. (1988) A systematic nomenclature for carbohydrate fragmentation in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5,** 253–257
56. Bardor, M., Burel, C., Villarejo, A., Cadoret, J. P., Carlier, A., and Lerouge, P. (2011) Plant *N*-glycosylation: an engineered pathway for the production of therapeutical plant-derived glycoproteins, in *Glycosylation in Diverse Cell Systems: Challenges and New Frontiers in Experimental Glycobiology* (Brooks, S., Rudd, P., and Apelmelk, B., eds.), Essential Reviews in Experimental Biology Series, Vol. 4, pp. 93–118, The Society for Experimental Biology, London, United Kingdom
57. Küster, B., and Mann, M. (1999) ^{18}O-labeling of *N*-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal. Chem.* **71,** 1431–1440
58. Balaguer, E., and Neusüss, C. (2006) Protein and glycan analysis by capillary spectrometry. *Anal. Chem.* **78,** 5384–5393
59. Allen, M. D., Kropat, J., Tottey, S., Del Campo, J. A., and Merchant, S. S. (2007) Manganese deficiency in *Chlamydomonas* results in loss of photosystem II and MnSOD function, sensitivity to peroxides, and secondary phosphorus and iron deficiency. *Plant Physiol.* **143,** 263–277
60. Page, M. D., Allen, M. D., Kropat, J., Urzica, E. I., Karpowicz, S. J., Hsieh, S. I., Loo, J. A., and Merchant, S. S. (2012) Fe sparing and Fe recycling contribute to increased superoxide dismutase capacity in iron-starved *Chlamydomonas reinhardtii*. *Plant Cell* **24,** 2649–2665
61. Weerapana, E., and Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16,** 91R-101R
62. Feng, J., Romaniouk, A. V., Samal, S. K., and Vijay, I. K. (2004) Processing enzyme glucosidase II: proposed catalytic residues and developmental regulation during the ontogeny of the mouse mammary gland. *Glycobiology* **14,** 909–921
63. Wilkinson, B. M., Purswani, J., and Stirling, C. J. (2006) Yeast GTB1 encodes a subunit of glucosidase II required for glycoprotein processing in the endoplasmic reticulum. *J. Biol. Chem.* **281,** 6325–6333
64. Kwiatkowski, B. A., Zielinska-Kwiatkowska, A. G., Migdalski, A., Kleczkowski, L. A., and Wasilewska, L. D. (1995) Cloning of two cDNAs encoding calnexin-like and calreticulin-like proteins from maize (Zea mays) leaves: identification of potential calcium-binding domains. *Gene* **165,** 219–222
65. Guerin, M., and Parodi, A. J. (2003) The UDP-glucose:glycoprotein glucosyltransferase is organized in at least two tightly bound domains from yeast to mammals. *J. Biol. Chem.* **278,** 20540–20546
66. Lubas, W. A., and Spiro, R. G. (1987) Golgi endo-α-D-mannosidase from rat liver, a novel N-linked carbohydrate unit processing enzyme. *J. Biol. Chem.* **262,** 3775-3781
67. Zuber, C., Spiro, M. J., Guhl, B., Spiro, R. G., and Roth, J. (2000) Golgi apparatus immunolocalization of endomannosidase suggests postendoplasmic reticulum glucose trimming: implications for quality control. *Mol. Biol. Cell* **11,** 4227–4240
68. Lobsanov, Y. D., Vallee, F., Imberty, A., Yoshida, T., Yip, P., Herscovics, A., and Howell, P. L. (2002) Structure of *Penicillium citrinum* alpha 1,2-mannosidase reveals the basis for differences in specificity of the endoplasmic reticulum and Golgi class I enzymes. *J. Biol. Chem.* **277,** 5620–5630
69. Tempel, W., Karaveg, K., Liu, Z. J., Rose, J., Wang, B. C., and Moremen, K. W. (2004) Structure of mouse Golgi alpha-mannosidase IA reveals the molecular basis for substrate specificity among class 1 (family 47 glycosylhydrolase) alpha1,2-mannosidases. *J. Biol. Chem.* **279,** 29774–29778
70. Kuokkanen, E., Smith, W., Mäkinen, M., Tuominen, H., Puhka, M., Jokitalo, E., Duvet, S., Berg, T., and Heikinheimo, P. (2007) Characterization and subcellular localization of human neutral class II α-mannosidase cytosolic enzymes/free oligosaccharides/glycoside hydrolase family 38/M2C1/*N*-glycosylation. *Glycobiology* **17,** 1084–1093
71. Chantret, I., and Moore, S. H. E. (2008) Free oligosaccharide regulation during mammalian protein N-glycosylation. *Glycobiology* **18,** 210–224
72. Strasser, R., Mucha, J., Mach, L., Altmann, F., Wilson, I., Glössl, J., and Steinkellner, H. (2000) Molecular cloning and functional expression of beta1,2-xylosyltransferase cDNA from *Arabidopsis thaliana*. *FEBS Lett.* **472,** 105–108
73. Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L., and Breton, C. (1999) Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology* **9,** 323–334
74. Both, P., Sobczak, L., Breton, C., Hann, S., Nöbauer, K., Paschinger, K., Kozmon, S., Mucha, I., and Wilson, I. B. H. (2011) Distantly related plant and nematode core α1,3-fucosyltransferases display similar trends in structure-function relationships. *Glycobiology* **21,** 1401–1415
75. Fabini, G., Freilinger, A., Altmann, F., and Wilson, I. B. (2001) Identification of core alpha 1,3-fucosylated glycans and cloning of the requisite

fucosyltransferase cDNA from Drosophila melanogaster. Potential basis of the neural anti-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., Maftah, 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-acetylglucosaminyltransferase 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) Plant secretome: unlocking secrets of the secreted proteins. *Proteomics* **10,** 799–827
81. Hanawa, Y., Watanabe, M., Karatsu, Y., Fukuzawa, H., and Shiraiwa, Y. (2008) Induction of a high-CO_2-inducible, periplasmic protein, H43, and its application as a high-CO_2-responsive marker for study of the high-CO_2-sensing mechanism in *Chlamydomonas reinhardtii*. *Plant Cell Physiol.* **48,** 299–309
82. Kinoshita, T., Fukuzawa, H., Shimada, T., Saito, T., 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
83. Matsuda, Y., Saito, T., Yamaguchi, T., and Kawase, H. (1985) Cell wall lytic enzyme released by mating gametes of Chlamydomonas reinhardtii is a metalloprotease and digests the sodium perchlorate-insoluble component of cell wall. *J. Biol. Chem.* **260,** 6373–6377
84. Kubo, T., Saito, T., Fukuzawa, H., and Matsuda, Y. (2001) Two tandemly-located matrix metalloprotease genes with different expression patterns in the *Chlamydomonas* sexual cell cycle. *Curr. Genet.* **40,** 136–143
85. Kubo, T., Kaida, S., Abe, J., Saito, T., and Fukuzawa, H. (2009) The Chlamydomonas hatching enzyme, sporangin, is expressed in specific phases of the cell cycle and is localized to the flagella of daughter cells within the sporangial cell wall. *Plant Cell Physiol.* **50,** 572–583
86. Li, G., Yan, Q., Nita-Lazar, A., Haltiwanger, R. S., and Lennarz, W. J. (2005) Studies on the *N*-glycosylation of the subunits of oligosaccharyl transferase in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280,** 1864–1871
87. Yan, Q., and Lennarz, W. J. (2002) Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* **277,** 47692–47700
88. Kelleher, D. J., Karaoglu, D., Mandon, E. C., and Gilmore, R. (2003) Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol. Cell* **12,** 101–111
89. Malerod, H., Graham, R. L. J., Sweredoski, M. J., and Hess, S. (2013) Comprehensive profiling of N-linked glycosylation sites in HeLa cells using hydrazide enrichment. *J. Proteome Res.* **12,** 248–259
90. Bakeberg, J. L., Tammachote, R., Woollard, J. R., Hogan, M. C., Tuan, H.-F., Li, M., Van Deursen, J. M., Wu, Y., Huang, B. Q., Torres, V. E., Harris, P. C., and Ward, C. J. (2011) Epitope-tagged Pkhd1 tracks the processing, secretion, and localization of fibrocystin. *J. Am. Soc. Nephrol.* **22,** 2266–2277
91. Hogan, M. C., Griffin, M. D., Rossetti, S., Torres, V. E., Ward, C. J., and Harris, P. C. (2003) PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. *Hum. Mol. Genet.* **12,** 685–698
92. Ward, C. J., Yuan, D., Masyuk, T. V., Wang, X., Punyashthiti, R., Whelan, S., Bacallao, R., Torra, R., LaRusso, N. F., Torres, V. E., and Harris, P. C. (2003) Cellular and subcellular localization of the ARPKD protein; fibrocystin is expressed on primary cilia. *Hum. Mol. Genet.* **12,** 2703–2710
93. Pazour, G. J., Agrin, N., Leszyk, J., and Witman, G. B. (2005) Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **170,** 103–113
94. Zhang, M.-Z., Mai, W., Li, C., Cho, S., Hao, C., Moeckel, G., Zhao, R., Kim, I., Wang, J., Xiong, H., Wang, H., Sato, Y., Wu, Y., Nakanuma, Y., Lilova, M., Pei, Y., Harris, R. C., Li, S., Coffey, R. J., Sun, L., Wu, D., Chen, X.-Z., Breyer, M. D., Zhao, Z. J., McKanna, J. A., and Wu, G. (2004) PKHD1 protein encoded by the gene for autosomal recessive polycystic kidney disease associates with basal bodies and primary cilia in renal epithelial cells. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 2311–2316
95. Onuchic, L. F., Furu, L., Nagasawa, Y., Hou, X., Eggermann, T., Ren, Z., Bergmann, C., Senderek, J., Esquivel, E., Zeltner, R., Rudnik-Schöneborn, S., Mrug, M., Sweeney, W., Avner, E. D., Zerres, K., Guay-Woodford, L. M., Somlo, S., and Germino, 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
96. Schroda, M., and Vallon, O. (2008) Chaperones and proteases, in *The Chlamydomonas Sourcebook* (Stern, D. B., ed.), 2nd ed., Vol. 2, pp. 671–729, Elsevier, Oxford, UK
97. 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
98. Arrington, D. D., and Schnellmann, 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
99. Yu, L.-G., Andrews, N., Weldon, M., Gerasimenko, 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
100. Villarejo, A., Burén, S., Larsson, S., Déjardin, A., Monné, M., Rudhe, C., Karlsson, J., Jansson, S., Lerouge, P., Rolland, N., Von Heijne, G., Grebe, M., Bako, 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
101. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Marth, J. D., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. (2009) Symbol nomenclature for glycan representation. *Proteomics* **9,** 5398–5399
102. Deleted in proof
103. Deleted in proof