Asparagine-Linked Glycans of *Cryptosporidium parvum* Contain a Single Long Arm, Are Barely Processed in the Endoplasmic Reticulum (ER) or Golgi, and Show a Strong Bias for Sites with Threonine*§

John R. Haserick‡§, Deborah R. Leon‡, John Samuelson§, and Catherine E. Costello‡¶

*Cryptosporidium parvum* causes severe diarrhea in infants in developing countries and in immunosuppressed persons, including those with AIDS. We are interested in the Asn-linked glycans (N-glycans) of *C. parvum*, because (1) the N-glycan precursor is predicted to contain five mannose and two glucose residues on a single long arm versus nine mannose and three glucose residues on the three-armed structure common in host N-glycans, (2) *C. parvum* is a rare eukaryote that lacks the machinery for N-glycan-dependent quality control of protein folding in the lumen of the Endoplasmic Reticulum (ER), and (3) ER and Golgi mannosidases, as well as glycosyltransferases that build complex N-glycans, are absent from the predicted proteome. The *C. parvum* N-glycans reported here, which were determined using a combination of collision-induced dissociation and electronic excitation dissociation, contain a single, unprocessed mannose arm ± terminal glucose on the trimannosyl chitobiase core. Upon nanoUPLC-MS/MS separation and analysis of the *C. parvum* tryptic peptides, the total ion and extracted oxonium ion chromatograms delineated 32 peptides with occupied N-glycan sites; these were derived from 16 glycoproteins. Although the number of potential N-glycan sites with Thr (NxT) is only about twice that with Ser (NxS), almost 90% of the occupied N-glycan sites contain NxT. The two most abundant *C. parvum* proteins modified with N-glycans were an immunodominant antigen on the surface of sporozoites (gp900) and the possible oocyst wall protein 1 (POWP1). Seven other glycoproteins with N-glycans were unique to *C. parvum*; five shared common ancestry with other apicomplexans; two glycoproteins shared common ancestry with many organisms. In summary, *C. parvum* N-glycans are remarkable for the absence of ER and Golgi modification and for the strong bias toward occupancy of N-glycan motifs containing Thr. *Molecular & Cellular Proteomics* 16: 10.1074/mcp.M116.066035, S42–S53, 2017.

Cryptosporidium parvum and *Cryptosporidium hominis* are coccidian parasites (walled apicomplexans) that infect humans (both) and cows (*C. parvum* only) (1–3). *C. parvum* and *C. hominis* are leading causes of diarrhea and death in children in the developing world and cause chronic diarrhea in AIDS patients (4–7). Although the massive outbreak of *C. parvum* in Milwaukee in 1993 was associated with contamination of municipal water, *Cryptosporidium* in developing countries is likely spread by poor hygiene (8, 9). Although there are mouse vaccine models and veterinary vaccines for *C. parvum*, there are no human vaccines (10–14). Furthermore, Nitazoxanide, the drug used to treat *C. parvum*, is not effective in immune suppressed persons (15).

We are interested in the N-glycans of *Cryptosporidium* for numerous reasons. Like *Toxoplasma gondii*, *Entamoeba histolytica*, and *Trichomonas vaginalis*, there is secondary loss of ALG genes, so that the predicted N-glycan precursor of *C. parvum* has a single long arm rather than the three-arm structure common in the host (supplemental Table S1) (16–21). In contrast to most other eukaryotes, *C. parvum* has a paucity of predicted mannosidases and glycosyltransferases, which could modify N-glycans in the endoplasmic reticulum (ER)1

---

1 The abbreviations used are: ER, endoplasmic reticulum; AIDS, Acquired Immune Deficiency Syndrome; ALG, Asparagine-Linked Glycosylation; ALC, average local confidence; EED, Electronic Excitation Dissociation; FA, formic acid; FDR, false discovery rate; GAP50, glycosomal-associated protein; HA, Hemagglutinin; HCD, higher-energy collision-induced dissociation; IAA, iodoacetamide; MeOH, methanol; AcOH, acetic acid; HIV, human immunodeficiency virus; IAA, iodoacetamide; OST, oligosaccharyltransferase; PNGaseF, peptide:N-glycanaseF; POWP1, possible oocyst wall protein 1; ppm, parts-per-million; UCG, unique *Cryptosporidium* glycoprotein; UPLC, Ultra Performance Liquid Chromatography.
and Golgi (22). C. parvum sporozoites label with cyanovirin-N, an anti-retroviral lectin that binds to the high mannose N-glycans of gp120 in HIV (23, 24). C. parvum is a rare eukaryote that lacks the machinery for N-glycan-dependent quality control of protein folding, and there is no positive selection for N-glycan sites in secreted proteins of C. parvum (25–27). Antigenic proteins on the surface of Cryptosporidium sporozoites (e.g., gp900 and gp40/gp15), oocyst wall proteins (COWPs) and possible oocyst wall proteins (POWPs), are glycoproteins with numerous predicted N-glycan sites (24, 28–40). Finally, Concanavalin A, which binds some N-glycans and other mannose-containing structures, recognizes numerous C. parvum antigens, Whereas release of N-glycans reduces binding of immune sera to parasite proteins on Western blots (41).

Here we used tandem mass spectrometry to identify the proteins that contain N-glycans, and determine the structures of N-glycans released with PNGaseF. We found that the N-glycans of C. parvum contain a single long arm, are barely processed in the ER or Golgi, and show an extreme bias for sequons with threonine.

**EXPERIMENTAL PROCEDURES**

**Parasites and Reagents**—C. parvum oocysts were purchased from Bunch Grass Farm (Deary, ID) and handled under BSL-2 protocols approved by the Boston University Institutional Biosafety Committee. All chemicals and reagents, including proteomics grade trypsin, were obtained from Sigma-Aldrich (St. Louis, MO), unless otherwise stated. All solvents used for LC-MS were Fisher Scientific Optima™ grade (Thermo-Fisher Scientific, Waltham, MA). PNGase F was from New England Biolabs (Ipswich, MA).

**Protein Extraction**—Two distinct methods were utilized to extract proteins from whole C. parvum oocysts. The first method used a combination of mechanical disruption and detergent extraction. Briefly, 109 oocysts were concentrated by centrifugation at 1000 g for 10 min at 4 °C. The oocysts were resuspended in phosphate buffered saline (PBS) with EDTA-free cOmplete™ protease inhibitor (Roche, Basel, Switzerland). The oocysts were broken using 0.5-mm glass beads with 4 × 5 min cycles of vigorous bead beating at 4 °C. Samples were placed in an ice bath between cycles to mitigate any heating effect. Proteins were extracted using a buffer containing 0.5-mm MeOH/0.1 M NH4OAc prior to lyophilization.

**Release and Processing of N-glycans**—N-glycans were released from total protein isolated from oocysts (100 μg) by overnight treatment at 37 °C with ten units of glycerol-free PNGase F (New England Biolabs), according to the manufacturer’s instructions, without the addition of Nonidet P-40. The product mixture was lyophilized, and the released N-glycans were separated from the proteins by addition of 0.1% trifluoroacetic acid (TFA) in LC-MS grade water. The aqueous phase was passed onto C-18 Sep-Pak cartridges (Waters Corporation, Milford, MA). The cartridges were washed with three bed volumes of 0.1% TFA/water, and the eluents were pooled and lyophilized. The N-glycan pool was reduced with 0.5 M NaBD4/2 M NH4OAc overnight at 55 °C. The reaction was quenched by dropwise addition of glacial acetic acid. The products were washed multiple times with 10% acetic acid/MeOH, dried with a gentle stream of nitrogen, and washed again multiple times with 100% MeOH.

**Permethylylation** was performed by published methods (44, 45). Briefly, a slurry of finely ground NaOH in dimethyl sulfoxide was added to the deuterio-reduced sample; the suspension was mixed and methyl iodide was added. The solution was gently mixed at RT for one hr. To assure complete derivatization, the process was repeated three times. The product was isolated by extraction with water/chloroform, and the chloroform layer was dried in the SpeedVac.

**MALDI-TOF MS**—The purified deuto-reduced sample was dissolved in 20 μl of 1:1 MeOH/water, and 0.5 μl of this solution was spotted on a stainless steel MALDI target with 2,5-dihydroxybenzoic acid as the matrix. The mass spectra were recorded with an ultrafleXtreme MALDI-TOF/TOF MS (Bruker Daltonics, Bremen, Germany) equipped with a smartbeam II Nd-YAG laser (355 nm, 3 nsec, 2 kHz). Each spectrum was acquired by summing the signals recorded after 500 shots from each of 10 locations within the sample spot.

**Electron Excitation Dissociation (EED) Fourier Transform-Ion Cyclotron Resonance (FT-ICR) MS/MS**—The released, deuto-reduced, and permethylated N-glycans were dried and re-suspended in 10 μl of 50% MeOH, 20 μM sodium acetate. The solution was loaded into a pulled glass capillary tube and directly infused into the ion source of a SolarisX 12-T hybrid Qq-FT-ICR mass spectrometer (Bruker Daltonics), using a nano-ESI source. Each [M + Na]1+ parent ion was isolated by the quadrupole and accumulated in the collision cell for 8 s. The accumulated ions were then transferred into the ICR cell.
Fragmentation by EED was achieved via 14-eV electrons generated from a cathode source heated with 1.5 A current. Electron density and energy were modulated using the following parameters: bias, 14 V; ECD lens, −13.85 V; pulse width 1.0 s. For each spectrum, 80 transients were averaged.

LC-MS/MS—The dried and desalted peptides were reconstituted in 2% ACN, 0.1% formic acid (FA) and separated on a NanoAcquity Ultra Performance Liquid Chromatography (UPLC) system (Waters), fitted with a nanoAcquity Symmetry C18 trap column (5-μm packing, 180 μm × 20 mm) and a BEH130C18 analytical column (1.7-μm packing, 150 μm × 10 cm). The mobile phase A was 99:1:0.1 HPLC grade water/ACN/FA, and mobile phase B was 99:1:0.1 ACN/HPLC grade water/FA. Each sample was loaded on the trapping column for 4 min at 4 μl/min flow rate and then separated on the analytical column using a 45 or 90 min 2–40% mobile phase B linear gradient at 0.5 μl/min flow rate. The column was washed between runs and equilibrated for 30 min.

The analytical column was coupled to a TriVersa NanoMate ion source (Advion, Ithaca, NY), and the ions were introduced into either an LTQ-Orbitrap-XL-ETD or a QE Plus mass spectrometer (both from Thermo-Fisher Scientific), which was operated in the positive-ion mode. MS spectra were obtained by scanning over the range m/z 350–2000. MS2 HCD spectra were acquired by isolating the top 5 (LTQ-Orbitrap) or top 20 (QE+) precursor ions with a 2-m/z window and fragmenting the selected precursor ions with 27, 35, or 45 V HCD energy. The MS2 HCD spectra were scanned from m/z 100 to a value that was dependent upon the parent ion.

Manual Interpretation of Glycopeptide MS/MS Spectra—Raw data files from LC-MS/MS experiments were manually interpreted using Qual Browser in the Xcalibur 2.2 software suite (Thermo-Fisher Scientific). HCD MS2 spectra containing oxonium ions were manually interpreted to determine the peptide sequence and the linear arrangement of the glycan. The y1 ion, corresponding to the residue Lys (K) or Arginine (R), was used as the starting point for most of the manually interpreted spectra. The resulting peptide tag was then searched using the online NCBI BLASTp algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the predicted C. parvum proteome, and the entire nr database (17–19). When a match was found, we determined the mass difference between the predicted tryptic generated peptide [M + H]1+ and that of the precursor, converted to [M + H]1+. The glycosidic bond fragment series, typically accounting for the most abundant peaks in the spectra, were sequenced in a similar manner, so far as each series could be followed. Missing residues were accounted for by calculating the difference between the highest member of the assigned series and the total observed molecular weight. Extracted ion chromatograms were generated to aid in the assignment of the numerous glycoconjugates.

LC-MS/MS Proteomics Database Search and Analysis—Once the possible N-glycoforms were discovered from the manual interpretation, these values could be utilized to search against the predicted C. parvum proteome as possible dynamic modifications. Database searches were performed using the PEAKS software suite version 7.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada). The following parameters were set for the search: the data refinement step corrected for the precursor m/z, for the PEAKSdenovo search stages, tryptic was specified as the enzyme, 8.0 ppm parent mass error tolerance, 0.05 Da fragment mass error tolerance, with carbamidomethyl cysteine set as a fixed modification, and possible dynamic modifications set to include methionine oxidation, HexNAc at serine/threonine; HexHexNAc2 and HexHexNAc3 on Asn. A maximum of five dynamic modifications was specified. The PEAKSDB search stage was identical to the PEAKSdenovo stage, with the exception that up to three missed tryptic cleavages were allowed, with the possibility of one nonspecific cleavage. Searches were performed against the C. parvum Iowaa-II predicted proteome release-5.0 obtained from the Cryptosporidium Genome Resource (cryptodb.org), which contained 3803 entries (18, 19). False discovery rate (FDR) estimation was enabled. For the final PEAKSPTM stage, the de novo score average local confidence (ALC) threshold was 15 and the peptide hit threshold (-10 logP) was set to 30. All possible Unimod modifications were considered for this stage. The PEAKSPTM report was exported as a mzidentML with a FDR set to 5%, ALC 50% for de novo only, and proteins with a score of (-10 logP) ≥ 20 containing unique peptides ≥ 2. Each data file was analyzed individually for all samples and replicates.

Scaffold Analysis—The mzidentML files from the PEAKSPTM searches were imported into the computer program Scaffold version 4.6 for further analysis (Proteome Software, Inc., Portland, Oregon). Three “Biosamples” and two “categories” were specified for the samples. The two categories corresponded to the method of protein extraction, either “mechanical” or “chemical.” The sample names correspond to the subsample classification, the “CHAPS” was the mechanically broken 2% CHAPS extraction buffer soluble portion, and the “phenol” and “interphase” samples correspond to the phenol and interphase layers from the chemical extraction procedure. Each sample was analyzed independently, with experiment wide grouping and protein clustering. The probability model utilized was Peptide Prophet with delta mass correction. All spectra that were assigned by the software as possible N-glycosylated peptides were manually reviewed for quality and proper assignment to compile the final lists of glycopeptides and proteins that are available in supplemental table Excel S3.

Analysis of N-glycosylation Sites—For each protein observed to be N-glycosylated, the lists of occupied and total potential N-linked sites were compared. The “occupied” data set was created from the list of peptides modified with an N-glycan, taking for each a nine-amino acid window, centered on the modified asparagine. The same window was taken for all tryptic peptides that contained a canonical N-glycosylation secon (Asn-Xxx-Ser/Thr, Xxx ≠ Pro) that could theoretically be generated from the group of observed glycoproteins. The program WebLogo v3.5.0 from the Department of Plant and Microbial Biology, University of California, Berkeley, was used to generate logos (46).

Bioinformatics—Predicted proteins of C. parvum with occupied N-glycan sites were analyzed for signal peptides and transmembrane helices using SignalP 4.0, TMHMM 2.0, and Phobius (47–51). Conserved domains were identified, and proteins were compared with those of other apicomplexans, eukaryotes, and bacteria (17–19, 50, 51). ALG enzymes, glucosidas, mannosidas, and OST peptides were predicted from whole genome sequences of C. parvum and T. gondii, using S. cerevisiae as a model (16, 20, 21, 25, 52, 53). Protein cartoon schematics were drawn using the program DOG 1.0 in combination with the software Inkscape 0.91 (54).

Analysis of Released N-glycans—To assist in the interpretation of the MS/MS spectra, we used the software GlycoWorkBench 2.1 (release 146) to generate theoretical fragmentation lists. Additional theoretical m/z values were generated using Microsoft Excel. Observed and theoretical peak lists were compared with obtain the best match. Assignments within 1-ppm error were considered to be a likely match. In the event that there were isobaric ion values, annotations were preferentially assigned to the ion that would be generated from a single fragmentation event. A single cross-ring fragment in combination with one or more glycosidic cleavages was considered only if the simple glycosidic bond fragment was also observed within the spectrum. All annotations were assigned only after a thorough manual review of the spectrum using Bruker DataAnalysis software suite version 4.05 SP5 build 283. Manual inspection helped to assign ions that didn’t fit the list of theoretical values for expected cleavages.
N-Glycans and Glycoproteins of Cryptosporidium parvum

---

**RESULTS**

*N-glycans of C. parvum Are Much Simpler Than Those of the Host and Most Other Parasites*—The predicted N-glycan precursor of *C. parvum*, based upon its Alg enzymes, is Glc2Man2GlcNAc2 ([supplemental table Excel S1](16, 20, 21, 27)). MALDI-TOF MS of N-glycans that had been deuterobased upon its Alg enzymes, is Glc2Man2GlcNAc2 ([supplemental table Excel S1](16, 20, 21, 27)). MALDI-TOF MS of N-glycans that had been deuteroreduced and permethylated after being released by PNGaseF from oocyst glycoproteins showed Hex5HexNAc2 (M + Na)1+ m/z 1800.906) to be the most abundant form, whereas Hex6HexNAc2 (M + Na)1+ m/z 1596.805) is less abundant (Fig. 1). The *C. parvum* N-glycans are much simpler than those of calf glycoproteins, suggesting oocysts, which are washed with PBS and purified on a CsCl gradient, are clean of those of calf glycoproteins, suggesting oocysts, which are washed with PBS and purified on a CsCl gradient, are clean of host tissues (56). The released N-glycans also match those present on glycopeptides that were separated by reversed host tissues (56). The released N-glycans also match those present on glycopeptides that were separated by reversed

The EED FT-ICR MS/MS was also performed on Hex5HexNAc2, which is the less abundant *C. parvum* N-glycan ([supplemental Figs. S1 and S2, supplemental table Excel S1](16, 20, 21, 27)). The same topology as described for the aforementioned Hex5HexNAc2 (minus a terminal hexose on the long arm) is indicated by the complete (Cn - 2H), Z, (Yn - 2H), and (Zn/Z3β + 2H) series where n = 3, 4, or 5 ([supplemental Figs. S1 and S2, supplemental table Excel S1](16, 20, 21, 27)). A series of cross-ring fragments similar to the Hex5HexNAc2 glycan were observed for this glycoform. The observation of the ion pairs 0,2A2 and 0,2X2 ions suggest that second and third hexoses on the long arm are either 2- or 3-linked. 0,2A2 or 0,3A2 fragments would be expected if these hexoses were linked at the 3 position; their absence suggests that the links are on the 2 position. The presence of an internal fragment 1,3A4 ions suggest that the less abundant 1,3A4 ions are isobaric to the 2,4A5 and 2,4X2 ions and these might allow assignment of the linkage to the 3 or 4 position, the presence of the 3,5A5 ion rules out linkage at the 4 position. In sum, these cross-ring fragments indicate that the long arm containing four hexoses is attached by a 1,3-linkage to the central mannos.

The observation of 1,3A4 ions suggest that second and third hexoses on the long arm are either 2- or 3-linked. 0,2A2 or 0,3A2 fragments would be expected if these hexoses were linked at the 3 position; their absence suggests that the links are on the 2 position. The presence of an internal fragment 1,3A4 ions suggest that the less abundant 1,3A4 ions are isobaric to the 2,4A5 and 2,4X2 ions and these might allow assignment of the linkage to the 3 or 4 position, the presence of the 3,5A5 ion rules out linkage at the 4 position. In sum, these cross-ring fragments indicate that the long arm containing four hexoses is attached by a 1,3-linkage to the central mannos.

EED FT-ICR MS/MS was also performed on Hex5HexNAc2, which is the less abundant *C. parvum* N-glycan ([supplemental Figs. S1 and S2, supplemental table Excel S1](16, 20, 21, 27)). The same topology as described for the aforementioned Hex5HexNAc2 (minus a terminal hexose on the long arm) is indicated by the complete (Cn - 2H), Z, (Yn - 2H), and (Zn/Z3β + 2H) series where n = 3, 4, or 5 ([supplemental Figs. S1 and S2, supplemental table Excel S1](16, 20, 21, 27)). A series of cross-ring fragments similar to the Hex5HexNAc2 glycan were observed for this glycoform. The observation of the ion pairs 0,2A2 and 0,2X2 ions suggest that second and third hexoses on the long arm are either 2- or 3-linked. 0,2A2 or 0,3A2 fragments would be expected if these hexoses were linked at the 3 position; their absence suggests that the links are on the 2 position. The presence of an internal fragment 1,3A4 ions suggest that the less abundant 1,3A4 ions are isobaric to the 2,4A5 and 2,4X2 ions and these might allow assignment of the linkage to the 3 or 4 position, the presence of the 3,5A5 ion rules out linkage at the 4 position. In sum, these cross-ring fragments indicate that the long arm containing four hexoses is attached by a 1,3-linkage to the central mannos.

---

**Fig. 1.** MALDI-TOF MS: released, deuto-reduced, permethylated *N*-glycans ([M + Na]1+). Total *N*-glycans released from *C. parvum* glycoproteins with PNGaseF, reduced with sodium borodeuteride, and permethylated. Only two glycoforms are observed: Hex5HexNAc2 and Hex6HexNAc2. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE ([74](http://www.ebi.ac.uk/pride/archive/projects/PXD005503)) partner repository with the data set identifier PXD005503 and 10.6019/PXD005503. **http://www.ebi.ac.uk/pride/archive/projects/PXD005503.**
differentiate between isobaric monosaccharides, (e.g. mannose from glucose), these methods can accurately define the topology of the glycan and the linkages connecting each monosaccharide. The structures we have defined are consistent with the N-glycan structure that has been proposed on the basis of the presence or absence of the highly conserved N-glycosylation biosynthetic pathway enzymes identified in the C. parvum genome (supplemental table Excel S1 and supplemental Fig. S3). The topology of the Hex5HexNAc2 structure defined here, presumably Man5GlcNAc2, is different from the Man9GlcNAc2 glycoform produced when host Man9GlcNAc2 is processed by ER mannosidase 1 in higher organisms, which has 1,3 and 1,6 dimannosyl branches off the first 1,6-linked Man (16, 25).

Confident Assignment of Cryptosporidium parvum N-glycosylated Peptides—To rule out the possibility of host cell contamination that could occur because C. parvum is an obligate intracellular parasite, the MS/MS spectra of glycopeptides from a whole oocyst lysate were manually interpreted. Peptides generated from a trypsin digestion of proteins isolated from oocysts were separated on a reversed phase-C18 nanoflow column interfaced to a mass spectrometer, as described in the methods section. Extracted oxonium ion chromatograms for m/z 204.08 (HexNAc) and m/z 366.13 (HexNAc-Hex) were very abundant throughout the MS/MS spectra recorded across the HPLC separation (Fig. 4A). These XIC pointed out which spectra should be manually interpreted to obtain sequence information on both glycan and peptide. A representative HCD spectrum for [M+H][\text{Na}]^+ m/z 1800.9192 eluting at 15 min, marked by the carat in Fig. 4A, is interpreted in Figs. 4B and 4C. Ions containing sequential glycosidic bond fragments dominate the spectrum and provide the linear sequence for Hex6HexNAc2 (Fig. 4B). The glycosidic fragments could be traced down to the aglycon, the peptide with [M+H] m/z 806.3992. Lower abundance peptide backbone fragments are observable in the magnified view of the spectrum.
The complete y-series is interpretable, starting from y1 \( m/z \) 175.1189, to the full-length peptide, observed at \( m/z \) 806.3992, thus revealing the peptide sequence NSTTEVR, and indicating the Hex6HexNAc2 was linked to Asn (Fig. 4C). Checking the peptide sequence against the NCBI nr and C. parvum proteome databases, utilizing the blastp algorithm, revealed that the peptide belongs to the C. parvum protein POWP1 (Table I). In summary, the linear glycan sequence, conjugation site, and complete peptide sequence, can all be determined from a single spectrum.

This method of manual interpretation was continued systematically for the remaining spectra containing one or more oxonium ion(s). Many of the most abundant peptides mapped to the same protein, POWP1. Five N-glycosylation sites were
mapped to this protein; all contain either Hex5HexNAc2 or Hex6HexNAc2 (Fig. 5 and Table I). Many of the MS/MS spectra assigned to glycopeptides contained Y1 and Y2 ions that arose via glycosidic cleavages adjacent to the HexNAc residues in the chitobiose core, with charge retention on the peptide fragment, but none of these MS/MS spectra contained (Y1/H11001)146 or (Y1/H11001)162 ions that would indicate the presence of a deoxyhexose or hexose branch on the inner HexNAc residue. No spectra indicated the presence of glycopeptides that did not originate from Cryptosporidium parvum. This result demonstrated that the preparation was clean from contaminating host material, and assured that the released N-glycans are of parasite origin, as the results had already suggested (Figs. 2, 3, supplemental Figs. S1, S2). The manual interpretations of the glycopeptide spectra are consistent with the results obtained by analysis of the released glycans and underscore the very limited repertoire N-glycans in this organism. This information could then be applied to perform semi-automated database searches to dig deeper in the spectra, allowing for faster processing of replicate samples.

Heavily Glycosylated Proteins Include an Immunodominant Protein (gp900) and a Putative Oocyst Wall Protein (POWP1)—Thirty-two tryptic peptides with occupied N-glycan sites were identified. These peptides derive from 16 glycoproteins, which include the vaccine candidate gp900 and a probable oocyst wall protein POWP1 (Fig. 5) (24, 30, 31, 35). Some proteins (e.g. gp900) but not others (e.g. POWP1) showed a higher relative abundance for Hex6HexNAc2 versus Hex5HexNAc2 (Table I), as did the released N-glycans (Fig. 1) (24). No other N-glycoforms were detected. Like gp900, two Cryptosporidium N-linked glycoproteins that are also unique (UCG1 and UCG2) contain long runs of Thr, which are likely modified by O-linked GalNAc (30, 35, 57). Five other unique glycoproteins with occupied N-glycan sites (UCG3 to UCG7) remain uncharacterized. Other observed glycoproteins have analogs elsewhere in apicomplexa: the glideosome-associated protein (GAP50), three putative adhesion proteins with a Limulus coagulation factor C lectin (LCCL) domain (CCP1, CCP2, and FNPA), and a copper amine oxi-

![Fig. 4](image-url). De novo identification of glycopeptides from a whole C. parvum oocyst lysate. A. Base peak and extracted oxonium ion chromatograms from a reversed phase C18 separation. The top trace shows the base peak chromatogram from the MS. The middle and bottom chromatograms are extracted oxonium ion chromatograms (XIC) from the 35-V HCD MS/MS spectra, corresponding to m/z 204.08 (HexNAc) and m/z 366.13 (HexNAc-Hex), respectively. The carat located at 15 min. indicates the time point for recording of the MS/MS spectrum shown in Fig. 4 B and 4 C. B. 35-V HCD MS/MS spectrum of an N-glycosylated peptide: NSTTEVR modified with Hex5HexNAc2. [M + 2H]2+ m/z 1092.9426. Prominent glycosidic bond fragmentation is observed, delineating the sequence of the glycan. C. Peptide sequence of the aglycon. Lower intensity y-ion peptide backbone fragments are observed in the same spectrum. Peptide fragment ion assignments are shown on this magnified view; these extend from y1 (Arg) at m/z 175.1189 to the complete aglycon, m/z 806.3992, defining the complete peptide sequence as NSTTEVR.

![Fig. 5](image-url). Occupied N-glycosylation sites of POWP1 (cgd2_490). The most densely N-glycosylated protein is represented as a cartoon schematic. The occupied peptides are shown; the bold and italicized asparagine residues indicate the site of attachment of the glycan.
TABLE I

Summary of identified N-glycosylated peptides

<table>
<thead>
<tr>
<th>Name</th>
<th>Ac. Number</th>
<th>Unique N-glycosylated peptides</th>
<th># Spectra</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp900</td>
<td>cgd7_4020</td>
<td>(107)-RMVDPVSLMDLFDWSTGVMDPNTNSILEGSIAIR-(141)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(989)-SGNLVHYPNTQMTSMSGLSVSPLA-(1012)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1224)-LINPTVNMTDSFAGAYK-(1242)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1388)-DPVTNTQYNSNTGNIIIPETGK-(1409)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1460)-LPWSGSADEITLVEVNTVIDTGLDLETGLPR-(1496)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1497)-DPVSVLPQLPNSTLVDPSNK-(1516)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1524)-SGF GNWSTGESHEK-(1538)</td>
<td>11</td>
</tr>
<tr>
<td>POWP1</td>
<td>cgd2_490</td>
<td>(31)-NSTRTEVR-(37)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(195)-NQTSSSGNPNVNLNR-(211)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(261)-NINPNLYETISSDGGK-(275)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(277)-YNTDASPDK-(285)</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(286)-TPEIVYYNTSNLR-(299)</td>
<td>3</td>
</tr>
<tr>
<td>UCG1</td>
<td>cgd5_1210</td>
<td>(328)-APGSSNTMQITNRLNNER-(345)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(362)-FALSPLAVGTEPLFSK-(378)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(97)-VLLGNDSLTK-(106)</td>
<td>1</td>
</tr>
<tr>
<td>UCG2</td>
<td>cgd6_710</td>
<td>(130)-GTIYNTISVDDLQNSR-(146)</td>
<td>0</td>
</tr>
<tr>
<td>UCG3</td>
<td>cgd1_640</td>
<td>(122)-EHVAFVTVGTQLGEVK-(138)</td>
<td>1</td>
</tr>
<tr>
<td>UCG4</td>
<td>cgd1_660</td>
<td>(290)-EIDDVIPHTETM-(303)</td>
<td>17</td>
</tr>
<tr>
<td>UCG5</td>
<td>cgd2_1290</td>
<td>(37)-ATNQTDWSFNLDR-(52)</td>
<td>0</td>
</tr>
<tr>
<td>UCG6</td>
<td>cgd3_660</td>
<td>(58)-ANVSTIFGDLNIS-(71)</td>
<td>0</td>
</tr>
<tr>
<td>UCG7</td>
<td>cgd8_4660</td>
<td>(175)-ISFLSEGSSITNTFTSMYR-(194)</td>
<td>0</td>
</tr>
<tr>
<td>GMCO</td>
<td>cgd2_2510</td>
<td>(209)-VYNLFWSDHGFR-(221)</td>
<td>0</td>
</tr>
<tr>
<td>GAP50</td>
<td>cgd2_640</td>
<td>(236)-MNVTSVPIVTNLTLYNKT-(255)</td>
<td>0</td>
</tr>
<tr>
<td>COA</td>
<td>cgd3_3430</td>
<td>(109)-NVTYSNDDIFPR-(122)</td>
<td>0</td>
</tr>
<tr>
<td>O-GNT4</td>
<td>cgd7_1310</td>
<td>(311)-SLNETQSGVDLOR-(324)</td>
<td>1</td>
</tr>
<tr>
<td>CDP2</td>
<td>cgd7_1730</td>
<td>(339)-VSINLSAMYQLK-(352)</td>
<td>0</td>
</tr>
<tr>
<td>CDP2</td>
<td>cgd7_300</td>
<td>(44)-SIIDTQDLGES          VTYSIK-(59)</td>
<td>0</td>
</tr>
<tr>
<td>CDP2</td>
<td>cgd7_4810</td>
<td>(144)-YASTCGSGQSVISSSR-(158)</td>
<td>0</td>
</tr>
</tbody>
</table>

**DISCUSSION**

ALG enzymes, which are required for the synthesis of N-glycan precursors, are reliable predictors of the types of N-glycans transferred to the nascent peptides, because these glycosyltransferases are constitutively expressed in the ER (16, 21). Two peculiarities present themselves with regards to the ALG enzymes of *C. parvum*. First, the ALG13 peptide of the glycosyltransferase that adds the second GlcNAc to the pyrophosphate-linked precursor can easily be identified in *Cryptosporidium muris* and in all other organisms that make N-glycans, but ALG 13 is absent from the predicted proteins of *C. parvum* and *C. hominis* (supplemental table Excel S1) (17–19, 52). Second, although the *C. parvum* N-glycan precursor is predicted to be Glc2Man_GlcNAc2, Hex2HexNAc2, Hex2HexNAc2 was absent from the N-glycans released with PNGaseF and from tryptic glycopeptides (Figs. 1 to 4 and Table I). This result suggests that ALG8, which adds the second glucose to the N-glycan precursor, is not active, or a glucose residue is rapidly removed by glucosidase 2 from Glc2Man_GlcNAc2 after it is transferred to the nascent peptide. In contrast, *T. gondii*, which has a predicted N-glycan precursor composed of Glc2Man2GlcNAc2, has been shown to have glycoproteins containing Hex2HexNAc2 and Hex2HexNAc2 (57).

Although it is well-known that the oligosaccharyltransferase (OST) that adds N-glycans to the nascent peptide prefers N-glycan sites with Thr over those with Ser, such a strong bias for Thr as that observed here for occupied N-glycan sites of *C. parvum* has not previously been de-
Parasites in the relative abundance of GlcMan5GlcNAc2 and from a precursor composed of Man5GlcNAc2 (61–63).

The canonical WebLogo 3.5.0, were compiled for all the peptides containing the served N sugars presented by the (Occupied) logo.

Mice, humans, cats, xylose (65–68). Finally, the merous complex glycans precursors with three mannose arms and make nu-

Whether the high mannose glycans on gp120 of HIV and on HA of influenza virus, remains to be determined (69–71). It has been established that many of the C. parvum proteins that elicit a strong immune response are N-linked glycoproteins (35, 41). Attempts have been made to develop vaccines from several of these glycoproteins; however, the critical details such as which amino acids are modified and with what glycan structure(s) were left unanswered (72, 73). The results we have presented here fill in the missing details regarding the N-glycosylation of the immunodominant antigen gp900, and we also expand upon the number of N-glycosylated proteins previously described in the literature. Of particular interest is the abundant and densely glycosylated protein POWP1. The function of POWP1 remains to be determined. These details may be crucial in providing a means to finally developing an effective, synthetic glycoprotein or glycopeptide-based vaccine against cryptosporidiosis.

The protein names referred to in this publication map to the following UniProtKB accession numbers: UCG3, UniProtKB = UniProt # Q5CT01; UCG4, UniProtKB = UniProt # Q5CSZ9; POWP1, UniProtKB = UniProt # Q5CU33; GAP50, UniProtKB = UniProt # Q5CU19; UCG5, UniProtKB = UniProt # Q5CTW5; GMCO, UniProtKB = UniProt # Q5CT6L; UCG6, UniProtKB = UniProt # Q5CV13; COA, UniProtKB = UniProt # Q5CUC0; UCG1, UniProtKB = UniProt # Q5CRW6; UCG2, UniProtKB = UniProt # Q5CXK6; CCP2, UniProtKB = UniProt # Q5CZ08; O-GAT4, UniProtKB = UniProt # Q5CYR4; CCP1, UniProtKB = UniProt # Q5CYM9; gp900, UniProtKB = UniProt # Q5CY21; FNPA, UniProtKB = UniProt # Q5CXV2; UCG7, UniProtKB = UniProt # Q5CV8B.

Supplementary material is freely available for downloading from the MCP website. It contains the EED MS/MS spectrum of the Hex4HexNAc2 glycan, glycosidic and cross-ring fragment assignments, a list of the C. parvum glycosyltransferases and a figure showing the predicted glycans of C. parvum. Excel files S1 and S2 contain the lists of fragment assignments, a list of the glycosyltransferases and a figure showing the predicted glycans of C. parvum. Excel files S3 contain the lists of fragment ion assignments. Excel file S3 contains the complete list of glycopeptides, proteins, and related bioinformatics data.

Acknowledgments—We thank Yi Pu and Cheng Lin for their assistance setting up the EED FT-ICR MS/MS experiment. Furthermore, Cheng Lin provided valuable feedback and questions pertaining to the interpretation of the EED spectra. Additionally, we thank Giulia Bandini for her insight on glycosyltransferases and knowledge of T. gondii.

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (http://www.ebi.ac.uk/pride) with the dataset identifier PXD005503 and 10.6019/PXD005503 (http://www.ebi.ac.uk/pride/archive/projects/PXD005503).

* Support for this study came from NIH grants R01 AI110638 and R01 GM031318 (J.S.), and P41 GM104603, S10 RR025082 and S10 OD010724 (C.E.C.), and NIH-NHLBI contract HHSN268201000031C (C.E.C.). We thank Thermo-Fisher Scientific for loan of the Q Exactive Plus system. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains supplemental material.
REFERENCES


