

Title: Changes in the *N*-glycome (glycoproteins with Asn-linked glycans) of *Giardia lamblia* with differentiation from trophozoites to cysts

Daniel M. Ratner^{1,2#}, Jike Cui¹, Martin Steffen³, Landon L. Moore³, Phillips W. Robbins¹, and John Samuelson^{1*}

Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, Boston, Massachusetts 02118¹, Section of Infectious Diseases, Boston Medical Center, Boston, MA 02118², and Department of Genetics and Genomics, Boston University School of Medicine, Boston, MA, 02118³

Running title: Changes in *Giardia* glycoproteins with differentiation

*Corresponding author. Mailing address: Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, 715 Albany Street, Evans 425, Boston, MA 02118, USA. Phone: (617) 414-1054; Fax: (617) 414-1041; E-mail: jsamuels@bu.edu.

#Present address: Department of Bioengineering, University of Washington, Seattle, WA 98195-5061

1 Abstract

2 *Giardia lamblia* are present in the intestinal lumen as binucleate, flagellated trophozoites
3 or quadrinucleate, immotile cysts. Here we used the plant lectin wheat germ agglutinin
4 (WGA) that binds to the disaccharide di-N-acetyl-chitobiose (GlcNAc₂), which is the truncated
5 Asn-linked glycan (*N*-glycan) of *Giardia*, to affinity purify the *N*-glycome (glycoproteins with *N*-
6 glycans) of trophozoites and cysts. Fluorescent WGA bound to the perinuclear membranes,
7 peripheral acidified vesicles, and plasma membrane of trophozoites. In contrast, WGA bound
8 strongly to membranes adjacent to the wall of *Giardia* cysts and less strongly to ER and
9 acidified vesicles. WGA lectin-affinity chromatography dramatically enriched secreted and
10 membrane proteins of *Giardia*, including proteases and acid phosphatases that retain their
11 activities. With mass spectroscopy, we identified 91 glycopeptides with *N*-glycans and 194
12 trophozoite secreted and membrane proteins, including 42 unique proteins. The *Giardia*
13 oligosaccharyltransferase, which contains a single catalytic subunit, preferred *N*-glycosylation
14 sites with Thr to those with Ser *in vivo* but had no preference for flanking amino acids. The
15 most abundant glycoproteins in the *N*-glycome of trophozoites were lysosomal enzymes,
16 folding-associated proteins, and unique transmembrane proteins with Cys-, Leu-, or Gly-rich
17 repeats. We identified 157 secreted and membrane proteins in the *Giardia* cysts, including 20
18 unique proteins. Compared to trophozoites, cysts were enriched in Gly-rich repeat
19 transmembrane (GRREAT) proteins, cyst wall proteins, and unique membrane proteins but
20 had relatively fewer Leu-rich repeat proteins, folding-associated proteins, and unique secreted
21 proteins. In summary, there are major changes in the *Giardia* *N*-glycome with differentiation
22 from trophozoites to cysts.

23 *Giardia lamblia* is a deeply divergent protist, which causes diarrhea (1, 32).
24 While *Giardia* is a tremendous problem in the developing world where hygiene is inadequate to
25 block its transmission by the fecal-oral route, some two million Americans are infected each
26 year with *Giardia*, which is present in streams and lakes or is transmitted in day-care centers
27 (46).

28 Binucleate trophozoites of *Giardia* are motile by means of eight flagellae and are
29 adherent by means of a ventral disc (13, 41). Secreted and membrane proteins of *Giardia* are
30 present in the ER; in peripheral vesicles, which have features in common with lysosomes; and
31 in the plasma membrane, which has variant-specific surface proteins (VSPs) that are rich in
32 Cys residues (1, 15, 27, 28, 32, 35, 44, 45).

33 The infectious and diagnostic stage of *Giardia* is the quadrinucleate cyst, the wall of
34 which contains a unique β -1,3-linked GalNAc polymer (12). Three abundant cyst wall proteins
35 (CWP1, CWP2, and CWP3), which are present in encystation-associated vesicles (ESV), have
36 a series Leu-rich repeats and a C-terminal Cys-rich domain (15, 33, 47). CWP1 is the target
37 for diagnostic monoclonal antibodies to *Giardia* in clinical specimens. There is a complex set
38 of membranes that is deep to the cyst wall of *Giardia* (8). Cys-rich proteins resembling VSPs
39 have been implicated in encystation (9), while numerous cysteine proteinases have been
40 associated with either encystation or excystation (50, 54).

41 We are interested in Asn-linked glycans (*N*-glycans) and glycoproteins with *N*-glycans
42 (the so-called *N*-glycome) of *Giardia* for many reasons. First, while the vast majority of
43 eukaryotes (animals, fungi, and plants) synthesize *N*-glycans by means of a lipid-linked
44 precursor containing 14 sugars (Glc₃Man₉GlcNAc₂-PP-dolichol), *Giardia* makes a lipid-linked

45 precursor with just two sugars (GlcNAc₂-PP-dolichol) (16, 42). This is because *Giardia* lacks
46 Alg enzymes, which add Man and Glc to Asn-linked glycans (42). Although secondary loss is
47 the most likely cause for the absence of Alg enzymes from *Giardia* and other protists, there is
48 presently no good explanation for why so many medically important protists have no N-glycans
49 (microsporidia and *Theileria*), very short N-glycans containing 1 or 2 sugars (*Giardia* and
50 *Plasmodium*), or truncated N-glycans containing 7 sugars (*Entamoeba* and *Trichomonas*) (42).

51 Second, the oligosaccharyltransferase (OST), which transfers the glycan from the
52 dolichol-linked precursor to the Asn residues on nascent peptides, is composed of 4 to 8
53 subunits in most eukaryotes but is composed of a single catalytic residue (STT3) in *Giardia*
54 and *Trypanosoma* (21). Eukaryotic sequons (sites of N-glycosylation) are specified by Asn-
55 Xaa-Thr or Asn-Xaa-Ser (NxT or NxS), where Xaa cannot be Pro, and sequons containing Thr
56 are more frequently occupied (5, 20, 22). The sequons of *Campylobacter*, which also has a
57 single catalytic peptide in the OST, have a negatively charged residue at the -2 position (Asp-
58 or Glu-Xaa-Asn-Xaa-Ser or -Thr) (24).

59 Third, *Giardia* N-glycans, which are predominantly composed of GlcNAc₂, are not
60 modified in the ER or Golgi (42). Indeed *Giardia* has a single nucleotide sugar transporter for
61 UDP-GlcNAc, which appears to be involved in synthesis of glycolipids rather than
62 glycoproteins (2). In contrast, GlcNAc, Gal, Fuc, and NANA are added from nucleotide-sugar
63 donors to make complex metazoan N-glycans (6, 26).

64 Fourth, *Giardia* is missing the machinery for N-glycan-dependent quality control of
65 protein folding and degradation, which is present in higher eukaryotes (3, 16, 51). However,
66 *Giardia* has N-glycan-independent quality control of protein folding in the ER (chaperones,

67 protein disulfide isomerases, and peptidyl-prolyl cis-trans isomerases) and degradation (Der1,
68 Ccd48, Np14, and Ufd1) (3, 23).

69 Fifth, wheat germ agglutinin (WGA), which recognizes GlcNAc polymers as well as sialic
70 acid, binds to the surface of trophozoites and cysts and blocks encystation (30, 38). Because
71 *Giardia* has none of the enzymatic machinery to make sialic acid or to transfer sialic acid from
72 host glycoproteins (32) (data not shown), WGA is likely binding to *Giardia* N-glycans
73 composed GlcNAc₂ (42). WGA-affinity columns have been used to dramatically enrich for
74 serum glycoproteins (53).

75 Sixth, while many secreted and membrane proteins are predicted by whole genome
76 sequencing of *Giardia* (32), it is not known how many of these proteins are present in
77 trophozoites or cysts. These secreted and membrane proteins of *Giardia* are likely important
78 for pathogenesis (e.g. interaction with the host epithelium or cyst wall formation) and may
79 include novel targets for anti-*Giardia* vaccines (1, 32, 37).

80 Four interrelated goals here were to: 1. localize N-glycans of *Giardia* trophozoites and
81 cysts using fluorescent WGA that binds to GlcNAc₂; 2. use WGA-affinity to enrich *Giardia*
82 secreted and membrane glycoproteins with N-glycans (the so-called N-glycome) and test the
83 activity of selective enzymes; 3. use mass spectroscopy to identify glycopeptides containing
84 occupied sequons and so determine the recipient peptide specificity of the single subunit
85 *Giardia* OST; 4. use mass spectroscopy to identify the *Giardia* N-glycome in order to compare
86 secreted and membrane proteins of trophozoites and cysts.

87 MATERIALS AND METHODS

88 **Parasites examined.** WB (genome project strain) and strain MR4 of *Giardia lamblia*
89 were grown axenically in TYI-S-33 medium supplemented with bile salts. Trophozoites were
90 encysted by standard methods (14). Briefly, *Giardia* trophozoites were cultured in pre-
91 encystation TYI-S-33 medium lacking bile salts, upon reaching ~60% confluence of adherent
92 trophozoites, the pre-encystation medium is exchanged for TYI-S-33 medium supplemented
93 with 10 mg/ml bile salts and 5 mM lactic acid hemi-calcium salt, pH 7.8. Non-adherent water-
94 resistant cysts were isolated and washed twice with deionized water. For fluorescence
95 microscopy, cysts of H3 strain of *Giardia*, which had been passaged through gerbils, were also
96 obtained from Waterborne Incorporated (New Orleans, LA). **To further characterize *in vitro* or**
97 **gerbil-derived cysts, excystation was performed using standard protocols (4).**

98 **Three-dimensional high-resolution fluorescence microscopy.** Logarithmic-phase
99 trophozoites were chilled to release adherent *Giardia*, concentrated by low speed
100 centrifugation, and washed in chilled phosphate buffered saline (PBS). For labeling the
101 surface of live trophozoites with an amine-reactive fluorescent probe, WB *Giardia* were placed
102 in 1 ml of 200 mM carbonate-bicarbonate buffer (to maximize the number of deprotonated
103 amino groups on their surface) and incubated with 20 ug/ml of Alexafluor 488 5-TFP
104 (Molecular Probes) for 60 min on ice (13). Alexafluor-labeled *Giardia* were washed 2X in PBS
105 and then fixed for 10 min at 4°C in 2% paraformaldehyde in 100 mM phosphate (no saline), pH
106 7.4. Immediately prior to rinsing, Triton X-100 was added to a final concentration of 0.1% to
107 permeabilize the organisms. Cells were gently pelleted by centrifugation and washed with
108 PBS + 2% BSA. After washing with PBS-BSA to quench free-aldehydes, *Giardia* were co-
109 labeled with 20 µg/ml TRITC-conjugated WGA in PBS-BSA for 30 min at 4°C, and then

110 washed twice in PBS. The nuclei of Alexafluor- and WGA-labeled *Giardia* were labeled with
111 0.1 µg/ml DAPI, and SlowFade anti-fade solution (Invitrogen) was added.

112 Because cysts of *Giardia* are often very poorly permeable to lectins or antibodies even
113 after treatment with non-ionic detergent, *Giardia* cysts, which had been fixed, were frozen and
114 thawed three times in order to produce an ice artifact prior to incubation with WGA and DAPI.
115 This method has been used to improve internal morphology and labeling of *Cryptosporidium*
116 sporocysts (18) and *Entamoeba* cysts (our unpublished data).

117 To label the plasma membrane with TRITC-WGA, live *Giardia* were incubated with
118 WGA in PBS at 4°C, as described, and subsequently rinsed in PBS, fixed in paraformaldehyde
119 (w/ 0.15% glutaraldehyde), and labeled with DAPI. To compare the WGA localization with that
120 of acidified vesicles, live *Giardia* were incubated for 90 min at 37°C with 100 nM
121 LysoTrackerRed DND-99 (Invitrogen), which is a fluorophore linked to a weak base that
122 accumulates in acidified vesicles. To label the cyst wall, encysting *Giardia* were incubated with
123 Waterborn Inc.'s Alexafluor 488-labeled monoclonal antibody, which is specific for CWP1 (33).

124 Slides were examined by three-dimensional multiple wavelength fluorescence
125 microscopy using an Olympus IX70 microscope equipped for DeltaVision deconvolution
126 (Applied Precision). This system employs restorative as well as deconvolution techniques to
127 provide resolutions up to four times greater than conventional light microscopes and is used to
128 study the ultrastructure of intracellular structures such as the kinetochore (7, 31). Images were
129 collected at 0.2 µm optical sections for the indicated wavelengths and were subsequently
130 deconvolved using SoftWoRx (Applied Precision). Data were examined as either optical
131 sections or as a projection of the entire stack.

132 **WGA-affinity chromatography of *Giardia* proteins.** Adherent logarithmic-phase
133 *Giardia* and water-resistant cysts were harvested on ice, washed in PBS and sonicated on an
134 ice-water slurry in the presence of 0.1% Triton X-100 and EDTA-free Complete protease
135 inhibitor cocktail (Roche). Insoluble material was removed by centrifugation ($>12,000 \times g$),
136 and soluble proteins were applied to a WGA-Sepharose column (EY Laboratories, Inc.) (53).
137 Because there was non-specific binding to this column, *Giardia* N-linked glycoproteins were
138 selectively eluted with 50 mM tetra-N-acetylchitotriose rather than with SDS. Proteins eluted
139 from the WGA column were run on SDS-PAGE containing 4 to 20% gradient of acrylamide
140 (Bio-Rad). In parallel lanes were 1) *Giardia* proteins solubilized with 0.1% Triton X-100 but not
141 applied to the WGA column, and 2) *Giardia* proteins that were treated twice with 1000 units
142 peptide:N-glycanase F (PNGaseF, New England Biolabs) for 9 hours each at 37°C in NEB G7
143 phosphate buffer. SDS-PAGE were fixed and stained with silver or Coomassie blue, or
144 proteins were transferred to nitrocellulose membranes by electroporation, incubated with HRP-
145 conjugated WGA, and developed with ECL chemiluminescent substrate (Pierce).

146 To demonstrate the ability of WGA affinity to enrich for enzymatic activity among the
147 *Giardia* N-glycome, trophozoite glycoproteins were prepared as described above with the
148 following differences. To isolate active proteases and acid phosphatases, we harvested
149 *Giardia* trophozoites at 0°C in the absence of protease inhibitor cocktail. Lectin affinity and
150 sample handling was carefully performed to ensure that the material remained on ice or at 4°C.
151 The resulting enriched glycoproteins were run on a Zymogram Gel containing 10% acrylamide
152 and gelatin (Bio-Rad). Samples were prepared using a non-reducing zymogram sample buffer
153 (Bio-Rad). The resulting gels were rinsed, incubated 1 hr in zymogram renaturation buffer,
154 and then incubated at 37°C in 50 mM acetate at pH 5.5 for 2 hrs. Proteolytic activity was

155 indicated by negative Coomassie staining on the zymogram gel. Acid phosphatase activity of
156 the enriched glycoproteins was assayed using an acid phosphatase assay kit (Sigma).

157 **Mass spectroscopy.** WGA-bound *Giardia* glycoproteins were run 0.5 cm into the
158 running gel of SDS-PAGE, excised, digested in-gel with sequencing grade trypsin (43), and
159 identified by LC/MS/MS, using methods that we have previously used to identify the cyst
160 proteins of *Entamoeba* (52). Briefly, reversed-phase chromatography was carried out using a
161 nano-HPLC pump and auto-sampler (Surveyor and MicroAS, Thermo Finnigan, San Jose, CA)
162 on a 10 cm by 100 micron ID MAGIC C18 reversed phase capillary column (Michrom, Auburn,
163 CA, USA) at the Boston University Proteomics Core Facility or at the MIT Center for Cancer
164 Research Biopolymers Laboratory. Peptides were separated using gradients of 5% to 90%
165 acetonitrile over 30-120 minutes in the presence of 0.5% acetic acid (55). Peptides were
166 analyzed using an LTQ ProteomeX ion trap mass spectrometer (Thermo Finnigan), and mass
167 spectra were compared to tryptic digests of *Giardia* proteins predicted from whole genome
168 sequencing (32) using SEQUEST and theGPM software (www.thegpm.org). All searches
169 were conducted on a reverse database to ensure that the false-positive rate for protein
170 identification was kept below 2%. Tryptic peptides with a SEQUEST XCorr score of > 1.5, 2.5
171 or 3.5 for Z = 1, 2, or 3 respectively, and a peptide score <0.05 were considered a match.
172 GPM-based searches were scored based on having a log(e) value greater than a cut-off
173 threshold of 2% false positives, typically $\log(e) < -10$. Proteins with one or more high scoring
174 fully-tryptic peptides were considered present. In addition, we used a neutral loss of 203.08
175 daltons or 406.16 daltons to identify glycopeptides with Asn-linked GlcNAc or GlcNAc₂
176 respectively. **Mass spectroscopy data has been submitted to the GiardiaDB website.**

177 **Additional analysis of the mass spectroscopic data.** Two dimensional protein gels
178 were simulated from mass spectroscopy data using theGPM, where the position of each
179 protein was determined by its predicted pI and mass, not including post-translational
180 modifications, and the size of the spot was proportional to the number of observed ions
181 corresponding to that protein. These two-dimensional gels highlighted the relative abundance
182 of secreted and plasma membrane proteins (defined by either an N-terminal ER-targeting
183 sequence or a transmembrane helix (25, 36) versus nucleocytoplasmic proteins (defined by
184 the absence of these features). **The ratio of secreted to nucleocytosolic proteins in a given
185 sample was determined by comparing the SEQUEST Area/Height results for the set of
186 secreted proteins with the SEQUEST Area/Height results for the nucleocytosolic proteins.**

187 *Giardia* proteins identified by mass spectroscopy were compared to annotated *Giardia*
188 proteins at *Giardia* DB and to proteins in the NR database at the NCBI. Secreted and
189 membrane proteins were sorted into ER, lysosomes, and plasma membrane based upon their
190 previous characterization in *Giardia* or in other eukaryotes. Proteins identified by LC/MS/MS
191 were characterized as unique *Giardia* proteins in cases where 1) there was no similarity to
192 other proteins in the NR data base and 2) the predicted open reading frame was designated
193 “hypothetical protein” in *Giardia* DB or the NR database at the NCBI.

194 **Methods for comparing the N-glycomes of the two stages.** Hierarchical clustering
195 of the protein data was performed using the Gene Expression Pattern Analysis Suite (GEPAS,
196 <http://gepas.bioinfo.cipf.es/>) to generate a heat map comparing the stage-specificity of the
197 proteins identified by mass spectrometry (17). Briefly, SEQUEST Area/Height results for all
198 samples were normalized via a Z-transformation and submitted as a single dataset for cluster

199 analysis. Complete linkage analysis was performed via GEPAS using a weighted pair-group
200 method (arithmetic averages), and distances were calculated using Spearman's Correlation
201 Coefficient Method. The resulting heat map (Supplemental Fig. 1) clearly groups samples as
202 either cyst or trophozoite, and the stage specific identification of individual proteins can be
203 identified by color coding on the map.

204 An additional stage specific comparison of identified proteins was performed by hand
205 (Supplemental Excel file) by listing proteins by their average SEQUEST Area/Height result or
206 by theGPM (X!Tandem) ion intensity for all identified peptides. Stage specificity was estimated
207 by calculating the ratio of the average abundance of proteins associated with the trophozoite
208 vs. cyst (troph/cyst ratio). Proteins with a troph/cyst ratio less than 0.5 were estimated to be
209 predominantly cyst-associated, and proteins with a troph/cyst ratio greater than 2 were
210 trophozoite-associated.

211 RESULTS

212 **WGA co-localizes to acidified vesicles of *Giardia* trophozoites and binds in a**
213 **punctate manner to the plasma membrane.** The first goal here was to compare the binding
214 of fluorescent WGA to that of Alexafluor, which binds evenly to the surface of the body and
215 flagellae of *Giardia* trophozoites (Figs. 1A and 1B) (13), or to that of lysotracker, which is taken
216 up into acidified vesicles by *Giardia* trophozoites (Figs. 2E and 2F) (11). The most important
217 observations with **three-dimensional high-resolution fluorescence microscopy** included the
218 following:

219 First, binding of WGA to *Giardia* trophozoites was remarkably similar from one parasite
220 to the next, so that all of the micrographs of individual *Giardia* in Fig. 1 are representative of
221 the group. Second, WGA labeled the plasma membrane of non-permeabilized *Giardia*
222 trophozoites in a punctate pattern, where the flagellae appeared less bright than the body of
223 *Giardia* (Fig. 1C). This punctate pattern of WGA, for which we have no explanation, was
224 distinct from the even labeling of the *Giardia* surface with Alexafluor, which included brightly
225 labeled flagellae (Figs. 1A and 1B) (13). Third, WGA labeled perinuclear membranes and
226 small vesicles, which were distributed throughout the cytosol (Figs. 1A and 1B). The
227 distribution of WGA labeling was similar to that previously shown for ER markers BiP and
228 protein disulfide isomerase (11, 28, 44). Fourth, WGA labeling was stronger to the intracellular
229 vesicles than to the plasma membrane. This result is similar to our experience with the anti-
230 retroviral lectin cyanovirin-N binding to *Entamoeba* N-glycans, which is greater on the internal
231 membranes of permeabilized trophozoites than to the surface (29). Fifth, WGA bound to the
232 same acidified vesicles in the body of *Giardia* trophozoites that accumulate lysotracker (Figs.
233 1D to 1F) (11).

234 **WGA-affinity column dramatically enriches the N-glycome (glycoproteins with N-**
235 **glycans) of *Giardia* trophozoites.** This result was shown in five ways:

- 236 1) Glycoproteins that bind WGA were dramatically enriched when solubilized proteins of
237 *Giardia* trophozoites were bound to a WGA-affinity column and eluted with chito-tetrose (Fig.
238 2B).
- 239 2) N-glycanase treatment of *Giardia* glycoproteins affinity purified with WGA entirely removed
240 the binding of WGA (Fig. 2E).

241 3) WGA-affinity dramatically increased the activity of Cys-proteases in zymograms of *Giardia*
242 proteins (Fig. 2C). Similarly, WGA-affinity increased the acid phosphatase activity of *Giardia*
243 protein lysate by 10-fold (data not shown).

244 4) WGA-affinity dramatically increased the number and coverage of secreted and membrane
245 proteins identified by mass spectroscopy. In the absence of WGA, most *Giardia* trophozoite
246 proteins identified by LC/MS/MS were nucleocytosolic (shown in blue and labeled with
247 numbers in Fig. 3A). In contrast after WGA-affinity, most proteins identified by LC/MS/MS
248 were secreted or membrane proteins (shown in red and labeled with letters in Fig. 3B). Please
249 see below for a more complete description of the *Giardia* N-glycome.

250 5) WGA-affinity chromatography and mass spectroscopy identified 91 glycopeptides containing
251 either N-linked GlcNAc₂ or GlcNAc (Fig. 4 and Fasta File in the Supplemental Data). In
252 contrast, in the absence of WGA, no glycopeptides with N-linked GlcNAc₂ or GlcNAc were
253 identified. In some relatively abundant glycoproteins, where many peptides were identified, 4
254 of 4 potential sites of N-glycosylation were occupied (GI50803_8360, a spingomyelin-like
255 phosphodiesterase), 2 of 2 were occupied (GI50803_16380, a cathepsin), or 3 of 4 sites were
256 occupied (GI50803_9780, a unique secreted trophozoite protein).

257 **The single-subunit OST of *Giardia* shows a preference for sequons with Thr *in***
258 ***vivo*.** All of the *Giardia* N-glycans identified by mass spectroscopy were present in sequons
259 (Asn-Xaa-Thr or Asn-Xaa-Ser, where Xaa can be any amino acid except Pro), as previously
260 shown for N-glycans of higher eukaryotes (22) (Fig. 4A). There was no preference by the
261 *Giardia* OST for a negatively charged amino acid at position -2, as has been shown for the
262 single subunit OST of bacteria (Fig. 4B) (24). Sixty-six percent of occupied sequons of *Giardia*

263 contained Thr, while 48% of the potential sites of *N*-linked glycosylation contained Thr (Fig.
264 4C). This result suggests moderate selection (~40% increase) *in vivo* for sequons with Thr
265 over those with Ser by the *Giardia* OST, which contains a single catalytic subunit and transfers
266 GlcNAc₂ (5, 20-22, 42).

267 We compared the data for *Giardia* with recent data for *Caenorhabditis elegans*, where
268 1191 unique occupied sequons were identified by lectin chromatography and mass
269 spectroscopy of tryptic peptides (Fig. 4C) (19). There was modest selection (~20% increase)
270 *in vivo* for sequons with Thr over those with Ser by the *Caenorhabditis* OST, which contains 8
271 subunits and transfers a 14-sugar *N*-glycan (16, 21, 42).

272 ***N*-glycome of *Giardia* trophozoites contains proteins involved in ER QC and**
273 **lysosomal enzymes.** The WGA-affinity purified glycoproteins (the so-called *N*-glycome) of
274 *Giardia* trophozoites included 169 secreted and plasma membrane proteins (Figs. 3, 5, and 6
275 and Table 1) (Fasta file and Excel File in the Supplemental Data). The 20 most abundant *N*-
276 glycome proteins of *Giardia* trophozoites were each at least 1% of the total proteins identified
277 and together comprised 45% of the proteins identified (Table 1, which shows an average of 9
278 experiments). Of these 20 glycoproteins, all but two were enriched two-fold or more in
279 trophozoites versus cysts (Fig. 6 and Table 1).

280 The most abundant category of trophozoite glycoproteins were enzymes associated
281 with lysosomes in other organisms, which include serine-, cysteine-, alanyl-peptidases, as well
282 as phosphodiesterases, phospholipases, and phosphatase (Figs. 3, 5, and 6 and Table 1).
283 Some of these enzymes have been localized to peripheral vesicles of *Giardia*, which were
284 labeled by WGA here (48, 49).

285 The second most abundant glycoproteins in the *Giardia* *N*-glycome were those involved
286 in *N*-glycan-independent quality control of protein folding in the ER lumen, which include
287 protein disulfide isomerases and chaperones (Hsp70 and Hsp90) (Figs. 3, 5, and 6 and Table
288 1) (3, 11, 23, 28, 44). Folding-associated proteins, which make up 17% of the trophozoite *N*-
289 glycome, were markedly decreased in cysts, where they make up just 4% of the *N*-glycome
290 (Figs. 5 and 6 and see below). The abundance of folding-associated proteins in *Giardia*
291 trophozoites is consistent with heavy labeling by WGA of perinuclear membranes and small
292 cytosolic vesicles (Fig. 1), where Hsp70 (also known as BiP) and protein disulfide isomerases
293 have been localized (11, 28, 44). As noted in the Introduction, *Giardia* is missing the set of
294 proteins involved in *N*-glycan-dependent quality control of protein folding (3).

295 The third most abundant glycoproteins in the *Giardia* *N*-glycome were membrane
296 proteins, which included multiple Leu-rich repeat (LRR) proteins and Cys-rich repeat (CRR)
297 proteins including VSPs (Figs. 3, 5, and 6 and Table 1) (1, 9, 32, 35). While CRR proteins
298 were also abundant in cysts, LRR proteins made up a larger percentage of the trophozoite *N*-
299 glycome (12%) than of the cyst *N*-glycome (5%) (Figs. 5 and 6).

300 ***N*-glycome of *Giardia* trophozoites contains dozens of previously hypothetical**
301 **proteins.** With WGA-affinity chromatography and mass spectroscopy, we identified peptide
302 sequences for 33 previously “hypothetical proteins,” which we now refer to as unique proteins
303 (Figs. 3, 5, and 6 and Table 1) (Fasta file and Excel File in the Supplemental Data) (32). In
304 trophozoites, the most abundant unique proteins contained an *N*-terminal signal peptide but no
305 transmembrane helices or GPI-anchors (10, 25, 36). Unique secreted proteins accounted for
306 17% of the *N*-glycome of *Giardia* trophozoites, including a glycoprotein (GI50803_9780) that is

307 6% of the N-glycome of trophozoites and relatively specific to trophozoites (Table 1 and
308 Supplemental Fasta file).

309 Unique transmembrane proteins included a novel family of Gly-rich repeat (GRREAT)
310 proteins, which contain an *N*-terminal signal-peptide and a *C*-terminal transmembrane helix
311 (Fig. 8). Because the GRREAT proteins were more abundant in the *N*-glycome of cysts than
312 in that of trophozoites, these unique transmembrane proteins are described below.

313 **WGA predominantly binds to the surface of encysted *Giardia*, which is closely**
314 **apposed to the wall of *Giardia* cysts.** The walls of encysting *Giardia* were densely labeled
315 with a monoclonal antibody to CWP1 (green in Fig. 7) (33). WGA (red in Fig. 7) bound
316 strongly to surface of encysted *Giardia*, which is closely apposed to the cyst wall, where
317 numerous membranes have been localized by electron microscopy (8). In contrast, WGA
318 bound less strongly to perinuclear membranes, peripheral vesicles, and the cyst wall itself.
319 *Giardia* cysts isolated from the intestines of infected gerbils, which readily excysted *in vitro*,
320 consistently showed four intact nuclei with DAPI staining (Figs. 7B and 7C) (data not shown)
321 (4, 54). In contrast, numerous *Giardia* cysts made *in vitro* often showed degenerative
322 changes, which included WGA-binding to structures that have pulled away from the cyst wall,
323 loss of nuclear integrity, and/or failure to excyst *in vitro* (Fig. 7A). The tendency for *Giardia*
324 cysts made *in vitro* to lose their viability after the cyst wall is formed has been noted previously
325 (14).

326 **The *N*-glycome of *Giardia* cysts varies dramatically from that of trophozoites.** The
327 WGA-affinity purified glycoproteins (the so-called *N*-glycome) of *Giardia* cysts made *in vitro*
328 included 157 secreted and plasma membrane proteins (Figs. 5 and 6 and Table 2) (Fig. 1 and

329 Excel File in the Supplemental Data). The 20 most abundant *N*-glycome proteins of *Giardia*
330 cysts were each at least 0.8% of the total proteins identified and together comprised 40% of
331 the proteins identified (Table 1, which shows an average of six experiments). The relative
332 paucity of cyst wall proteins (CWP1, CWP2, and CWP3) (15, 33, 47) likely results from 1)
333 difficulties releasing these proteins from the cyst walls with non-ionic detergent and 2) the
334 relative paucity of trypsin cleavage sites in these proteins, which produce fewer usable
335 peptides for mass spectroscopy.

336 Dramatic differences in the *N*-glycome of cysts from that of trophozoites were shown by
337 two-dimensional reconstruction of the mass spectroscopy data of a representative experiment
338 (Fig. 6) and by hierarchical clustering of numerous experiments (“heat maps” in Fig. 1 in
339 Supplemental Data). These results, which are summarized in pie charts in Fig. 5, are detailed
340 in the Excel file in the Supplemental Data. Compared to trophozoites, cysts had fewer Leu-rich
341 repeat proteins, folding proteins, and unique secreted proteins (Figs. 5 and 6 and Tables 1 and
342 2). In contrast, cysts were relatively enriched in Gly-rich repeat transmembrane (GRREAT)
343 proteins (see next paragraph), cyst wall proteins, and unique transmembrane proteins. In
344 addition, while the *N*-glycomes of trophozoites and cysts had similar overall percentages of
345 lysosomal enzymes, numerous of the proteases and phospholipases were specific for one
346 stage or the other (Figs. 5 and 6 and Table 1).

347 Among the set of unique *Giardia* cyst transmembrane proteins were GRREAT proteins,
348 which are type 1 membrane proteins that contain a large region of relatively degenerate, Gly-
349 rich repeats (Fig. 8). GIGRREAT1 (GI50803_114210), which composed ~9% of the *N*-glycome

350 of *Giardia* cysts, is 1,088 amino acid long and shares with other GRREAT proteins a C-
351 terminal Cys-rich domain adjacent to the transmembrane helix.

352 DISCUSSION

353 While the binding of WGA to *Giardia* cyst walls and inhibition of encystation by WGA
354 have been shown previously (30, 38), this is the first demonstration that WGA binds to its very
355 short *N*-glycans that contain just two GlcNAc, the first fine localization of WGA on the surface
356 and interior of trophozoites and cysts, the first use of WGA to identify large groups of
357 trophozoite and cyst glycoproteins with *N*-glycans (the *N*-glycome), and the first demonstration
358 of the abundance of GRREAT proteins in cysts.

359 This is also the first time that a large number of occupied sequons (sites of *N*-linked
360 glycans) have been identified from any of the protists. This is the first demonstration that an
361 OST with a single catalytic subunit has the same preference for sequons containing Thr over
362 those containing Ser (as is the case for OSTs of metazoa and fungi that contain 7 to 8
363 peptides) (6, 16, 20-22). While the *Giardia* OST resembles that of bacteria in that both have a
364 single catalytic peptide, the *Giardia* OST uses the eukaryotic sequon rather than the bacterial
365 sequon, which contains a negative charge in the -2 position (24).

366 Although *N*-glycans are added to sequons on many different glycoproteins and all of the
367 sequons on a given *Giardia* glycoprotein may be occupied, it is not clear what the very short *N*-
368 glycans of *Giardia* are doing, as they do not participate in *N*-glycan-dependent quality control
369 of protein folding or degradation (3). While there is positive selection for sites of *N*-linked
370 glycosylation in secreted and membrane proteins of eukaryotes with *N*-glycan-dependent

371 quality control of protein folding, there is no positive selection for sites of N-linked glycosylation
372 in secreted proteins and membrane proteins of *Giardia* and other protists (e.g. *Plasmodium*,
373 *Cryptosporidium*, and *Toxoplasma*), which lack N-glycan-dependent quality control of protein
374 folding (our unpublished data). We speculate that GlcNAc₂ on *Giardia* N-glycans stimulates
375 the innate immune system, as has recently been described for chitin (40). We are presently
376 attempting to knock down expression of *Giardia* Alg enzymes and STT3, which are crucial for
377 synthesis of N-glycans (16, 21, 42), to determine the effect on protist viability, phenotype, and
378 differentiation.

379 While WGA-affinity has been used previously to enrich serum glycoproteins containing
380 sialic acid (53), this is the first time that WGA has been used to enrich glycoproteins with very
381 short N-glycans containing GlcNAc₂. Advantages of the use of lectin affinity are 1) it
382 dramatically enriches the secreted and membrane glycoproteins of *Giardia*, many of which are
383 likely important for pathogenesis. 2) Glycoproteins may be released with chito-
384 oligosaccharides, so that enzymes are still active and proteins are still in their native form for
385 subsequent immunization. 3) Multiple repeats of the same experiment can be performed
386 quickly, so that low abundance glycoproteins may be detected, and better estimates of the
387 relative abundance of proteins can be made.

388 4) It appears that both trophozoites and cysts add N-glycans to many glycoproteins, as
389 shown by mass spectroscopy and by three-dimensional high-resolution fluorescence
390 microscopy. Further, there is no evidence in any organism that Alg genes, which encode
391 enzymes that make N-glycan precursors, show stage-specific expression (16). As well, there
392 is no subsequent modification of *Giardia* N-glycans (42), which might be affected by

393 differentiation. 5) Identification of numerous tryptic peptides makes it possible to shift
394 numerous *Giardia* proteins from the “hypothetical” status to the “unique protein” status. 6)
395 Lectin affinity also works very well to enrich the secreted and membrane proteins of
396 *Entamoeba*, *Trichomonas*, and *Cryptosporidium* (our unpublished data).

397 Disadvantages of lectin affinity are 1) it misses nucleocytosolic proteins, which include
398 many of the enzymes important for *Giardia* to adapt to its anaerobic environment and for
399 synthesizing precursors of its cyst wall polymer (12, 34). We have recently used two-
400 dimensional liquid chromatography and many mass spectroscopic runs to identify peptides
401 from >1400 total proteins of *Giardia*, which have also been deposited at the GiardiaDB
402 website. 2) The method does not work well with secreted and membrane proteins that lack
403 sequons or that have sequons that are not N-glycosylated. Conversely, glycoproteins that
404 contain many sequons or glycoproteins that are heavily glycosylated will be over-represented
405 in mass spectroscopic data. 3) WGA binds weakly or not at all to N-glycans containing a single
406 GlcNAc. This is not a problem in *Giardia* where most of the N-glycans contains GlcNAc₂ but is
407 a problem in *Plasmodium falciparum*, where many of the N-glycans contain a single GlcNAc
408 (42) (our unpublished data).

409 4) The method does not work if the secreted or membrane proteins are not solubilized
410 with non-ionic detergent, which is likely the case for cyst wall proteins (15, 33, 47). 5) The
411 method does not discriminate between plasma membrane proteins and those located in the
412 ER or lysosomes. 6) The make up of a given sample is calculated based upon either MS/MS
413 peak area/height analysis or on peptide coverage. These methods provide a useful estimate
414 of protein abundance, but cannot determine absolute values for protein composition in a

415 sample. 7) Similarly, hierarchical clustering, which compares the normalized abundance of
416 proteins from different samples (cyst vs. trophozoite), gives the same weight to both minor and
417 major protein constituents. Due to this normalization, the resulting heat maps tend to
418 dramatically overstate the differences between samples due to variability in runs on a
419 particular mass spectrometer, while at the same time identifying significant clusters of proteins
420 that are stage-specific.

421 7) There is no direct correlation between protein abundance and protein importance,
422 although it is often the case that monoclonal antibodies identify vaccine candidates that are
423 abundant (e.g. VSPs of *Giardia* or Gal/GalNAc lectins of *Entamoeba*) (35, 37, 39). 8) These
424 mass spectroscopic methods, like microarray data, generate large lists of proteins, each of
425 which calls for further investigation and better characterization.

426 The composition of the *N*-glycome of *Giardia* trophozoites was remarkable for the large
427 numbers of lysosomal enzymes and ER enzymes involved in *N*-glycan-independent quality
428 control of protein folding. In contrast, the results here, which focus on *N*-glycans that are
429 added in the ER lumen and do not appear to be subsequently modified, shed little light on
430 glycosylation in the Golgi and do not address other Golgi functions (29, 42, 45, 49).

431 The large number of unique secreted and membrane glycoproteins of *Giardia* identified
432 here likely includes novel vaccine candidates. A possible advantage of these unique *Giardia*
433 proteins is that they likely do not belong to a large gene family that is variably expressed, as is
434 the case for *Giardia* VSPs (35). It is likely that WGA-affinity methods may be used to collect
435 *Giardia* glycoproteins in their native state for immunization and for screening monoclonal
436 antibodies (37).

437 The *Giardia* cysts made *in vitro* here do not excyst nearly as well as those isolated from
438 the intestines of gerbils, so that it is likely that our mass spectroscopic data here will not
439 perfectly reflect proteins of the *in vivo Giardia* cysts. Further the conditions we used for
440 encystation did not produce well-synchronized cysts, so that we only used end-points to
441 describe the transition from trophozoites to cysts. With these caveats and those described
442 above for the WGA-affinity methods, the composition of the *N*-glycome of *Giardia* cysts was
443 remarkable for the relative paucity of ER folding proteins and unique secreted proteins and for
444 the abundance of encystation-associated proteases, unique transmembrane proteins, and
445 GRREAT family proteins. These results support previous experimental demonstrations of the
446 importance of proteases (50, 54) and Cys-rich proteins (9) in encystation. These results are
447 also consistent with the idea that secretion is less important to encysting *Giardia* than is
448 development of membranous structures (8). Identification of VSPs in *Giardia* cysts is
449 consistent with previous demonstration of encystation-associated VSP-like proteins (9). Of
450 great interest is the function of GRREAT proteins, which are very abundant in *Giardia* cysts,
451 less abundant in *Giardia* trophozoites, but have no homologs in other organisms.

452 **Acknowledgments.** This work was supported by NIH grants AI048082 (to J.S.) and
453 GM31318 (to P.W.R.). Support for D.M.R. was provided by the Training Program in Host
454 Pathogen Interactions (T32 AI052070). Thanks to Dick Cook of MIT for some mass
455 spectroscopy data.

456 **References.**

457 1. Adam, R. D. 2001. Biology of *Giardia lamblia*. Clin. Microbiol. Rev. 14:447-475.

- 458 2. **Banerjee, S., J. Cui, P. W. Robbins, and J. Samuelson.** 2008. Use of *Giardia*, which has
459 a single nucleotide-sugar transporter for UDP-GlcNAc, to identify the UDP-Glc transporter of
460 *Entamoeba*. *Mol. Biochem. Parasitol.*
- 461 3. **Banerjee, S., P. Vishwanath, J. Cui, D. J. Kelleher, R. Gilmore, P. W. Robbins, and J.**
462 **Samuelson.** 2007. Evolution of quality control of protein- folding in the ER lumen. *Proc. Natl.*
463 *Acad. Sci. USA* **104**:11676-11681.
- 464 4. **Boucher, S. E., and F. D. Gillin.** 1990. Excystation of *in vitro*-derived *Giardia lamblia* cysts.
465 *Infect Immun.* **58**:3516-3522.
- 466 5. **Breuer, W., R. A. Klein, B. Hardt, A. Bartoschek, and E. Bause.** 2001.
467 Oligosaccharyltransferase is highly specific for the hydroxy amino acid in Asn-Xaa-Thr/Ser.
468 *FEBS Lett.* **501**:106-110.
- 469 6. **Caffaro, C. E., and C. B. Hirschberg.** 2006. Nucleotide sugar transporters of the Golgi
470 apparatus: from basic science to diseases. *Acc. Chem. Res.* **39**:805-812.
- 471 7. **Carrington, W. A., R. M. Lynch, E. D. W. Moore, G. Isenberg, K. E. Fogarty, and F. S.**
472 **Fay.** 1995. Superresolution three-dimensional images of fluorescence in cells with minimal
473 light exposure. *Science* **268**:1483–1487.
- 474 8. **Chavez-Munguia, B., R. Cedillo-Rivera, and A. Martinez-Palomo.** 2004. The
475 ultrastructure of the cyst wall of *Giardia lamblia*. *J. Eukaryot. Microbiol.* **51**:220-226.

- 476 9. **Dauids, B. J., D. S. Reiner, S. R. Birkeland, S. P. Preheim, M. J. Cipriano, A. G.**
477 **McArthur, and F. D. Gillin.** 2006. A new family of giardial cysteine-rich non-VSP protein
478 genes and a novel cyst protein. *PLoS ONE* **1**:e44.
- 479 10. **Eisenhaber, B., P. Bork, and F. Eisenhaber.** 1998. Sequence properties of GPI-
480 anchored proteins near the omega-site: constraints for the polypeptide binding site of the
481 putative transamidase. *Protein Eng.* **11**:1155-1161.
- 482 11. **Gaechter, V., E. Schraner, P. Wild, and A. B. Hehl.** 2008. The single dynamin family
483 protein in the primitive protozoan *Giardia lamblia* is essential for stage conversion and
484 endocytic transport. *Traffic* **9**:57-71.
- 485 12. **Gerwig, G. J., J. A. van Kuik, B. R. Leeflang, J. P. Kamerling, J. F. Vliegenthart, C. D.**
486 **Karr, and E. L. Jarroll.** 2002. The *Giardia intestinalis* filamentous cyst wall contains a novel
487 beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study.
488 *Glycobiology* **12**:499-505.
- 489 13. **Ghosh, S. K., R. Rogers, and J. Samuelson.** 2001. How giardia swim and divide.
490 *Infect. Immun.* **69**:7866-7872.
- 491 14. **Gillin, F. D., D. S. Reiner, and S. E. Boucher.** 1988. Small-intestinal factors promote
492 encystation of *Giardia lamblia in vitro*. *Infect Immun.* **56**:705-707.
- 493 15. **Gillin, F. D., D. S. Reiner, and J. M. McCaffery.** 1996. Cell biology of the primitive
494 eukaryote *Giardia lamblia*. *Annu. Rev. Microbiol.* **50**:679-705.

- 495 16. **Helenius, A., and M. Aebi.** 2004. Roles of N-linked glycans in the endoplasmic reticulum.
496 *Annu. Rev. Biochem.* **73**:1019-1049.
- 497 17. **Herrero, J., A. Valencia, and J. Dopazo.** 2001. A hierarchical unsupervised growing
498 neural network for clustering gene expression patterns. *Bioinformatics* **17**:126-136.
- 499 18. **Jenkins, M. C., C. Murphy, J. Trout, and R. Fayer.** 2006. An improved electron
500 microscopic technique for the immunolabeling of *Cryptosporidium parvum* oocysts. *J.*
501 *Parasitol.* **92**:403-405.
- 502 19. **Kaji, H., Y. Yamauchi, N. Takahashi, and T. Isobe.** 2006. Mass spectrometric
503 identification of N-linked glycopeptides using lectin-mediated affinity capture and
504 glycosylation site-specific stable isotope tagging. *Nat. Protoc.* **1**:3019-3027.
- 505 20. **Kasturi, L., J. R. Eshleman, W. H. Wunner, and S. H. Shakin-Eshleman.** 1995. The
506 hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation
507 efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.*
508 **270**:14756-14761.
- 509 21. **Kelleher, D. J., and R. Gilmore R.** 2006. An evolving view of the eukaryotic
510 oligosaccharyltransferase. *Glycobiol.* **16**:47R-62R.
- 511 22. **Kornfeld, R., and S. Kornfeld.** 1985. Assembly of asparagine-linked oligosaccharides.
512 *Annu. Rev. Biochem.* **54**:631-664.
- 513 23. **Kostova, Z., and Wolf, D. H.** 2003. For whom the bell tolls: protein quality control of the
514 endoplasmic reticulum and the ubiquitin-proteasome connection. *EMBO J.* **22**:2309-2317.

- 515 24. **Kowarik, M., N. M. Young, S. Numa, B. L. Schulz, I. Hug, N. Callewaert, D. C. Mills,**
516 **D. C. Watson, M. Hernandez, J. F. Kelley, M. Wacker, and M. Aebi.** 2006. Definition of
517 the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**:1957-1966.
- 518 25. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting
519 transmembrane protein topology with a hidden Markov model: application to complete
520 genomes. *J. Mol. Biol.* **305**:567-580.
- 521 26. **Lowe, J. B., and J. D. Marth.** 2003. A genetic approach to mammalian glycan function.
522 *Annu. Rev. Biochem.* **72**:643-691.
- 523 27. **Luján, H. D., A. Marotta, M. R. Mowatt, N. Sciaky, J. Lippincott-Schwartz, and T. E.**
524 **Nash.** 1995. Developmental induction of Golgi structure and function in the primitive
525 eukaryote *Giardia lamblia*. *J. Biol. Chem.* **270**:4612-4618.
- 526 28. **Lujan, H. D., M. R. Mowatt, J. T. Conrad, and T. E. Nash.** 1996. Increased expression of
527 the molecular chaperone BiP/GRP78 during the differentiation of a primitive eukaryote. *Biol.*
528 *Cell* **86**:11-18.
- 529 29. **Magnelli, P., J. F. Cipollo, D. M. Ratner, J. Cui, D. Kelleher, R. Gilmore, C. E. Costello,**
530 **P. W. Robbins, and J. Samuelson J.** 2008. Unique Asn-linked oligosaccharides of the
531 human pathogen *Entamoeba histolytica*. *J. Biol. Chem.* **283**:18355-18364.
- 532 30. **Meng, T. C., M. L. Hetsko, and F. D. Gillin.** 1996. Inhibition of *Giardia lamblia* excystation
533 by antibodies against cyst walls and by wheat germ agglutinin. *Infect. Immun.* **64**:2151-2157.

- 534 **31. Moore, L. L., M. Morrison, and M. B. Roth.** 1999. HCP-1, a protein involved in
535 chromosome segregation, is localized to the centromere of mitotic chromosomes in
536 *Caenorhabditis elegans*. *J. Cell Biol.* **147**:471-480.
- 537 **32. Morrison, H. G., A. G. McArthur, F. D. Gillin, S. B. Aley, R. D. Adam, G. J. Olsen, et al.**
538 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science*
539 **317**:1921-1926.
- 540 **33. Mowatt, M. R., H. D. Lujan, D. B. Cotten, B. Bowers, J. Yee, T. E. Nash, and H. H.**
541 **Stibbs.** 1995. Developmentally regulated expression of a *Giardia lamblia* cyst wall protein
542 gene. *Mol. Microbiol.* **15**:955-963.
- 543 **34. Müller, M.** 1988. Energy metabolism of protozoa without mitochondria. *Annu. Rev.*
544 *Microbiol.* **42**:465-488.
- 545 **35. Nash, T. E.** 2002. Surface antigenic variation in *Giardia lamblia*. *Mol. Microbiol.* **45**:585-
546 590.
- 547 **36. Nielsen, H., S. Brunak, and G. von Heijne.** 1999. Machine learning approaches for the
548 prediction of signal peptides and other protein sorting signals. *Protein Eng.* **12**:3-9.
- 549 **37. Olson, M. E., H. Ceri, and D. W. Morck.** 2000. *Giardia* vaccination. *Parasitol. Today*
550 **16**:213-217.
- 551 **38. Ortega-Barria, E., H. D. Ward, J. E. Evans, and M. E. Pereira.** 1990. N-acetyl-D-
552 glucosamine is present in cysts and trophozoites of *Giardia lamblia* and serves as receptor for
553 wheatgerm agglutinin. *Mol. Biochem. Parasitol.* **43**:151-165.

- 554 **39. Petri, W. A., Jr., and R. Haque, B. J. Mann.** 2002. The bittersweet interface of parasite
555 and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba*
556 *histolytica*. *Annu. Rev. Microbiol.* **56**:39-64.
- 557 **40. Reese, T. A., H. E. Liang, A. M. Tager, A. D. Luster, N. Van Rooijen, D. Voehringer,**
558 **and R. M. Locksley.** 2007. Chitin induces accumulation in tissue of innate immune cells
559 associated with allergy. *Nature* **447**:92-96.
- 560 **41. Sagolla, M. S., S. C. Dawson, J. J. Mancuso, and W. Z. Cande.** 2006. Three-
561 dimensional analysis of mitosis and cytokinesis in the binucleate parasite *Giardia intestinalis*.
562 *J. Cell Sci.* **119**:4889-4900.
- 563 **42. Samuelson, J., S. Banerjee, P. Magnelli, J. Cui, D. J. Kelleher, R. Gilmore, and P. W.**
564 **Robbins.** 2005. The diversity of protist and fungal dolichol-linked precursors to Asn-linked
565 glycans likely results from secondary loss of sets of glycosyltransferases. *Proc. Natl. Acad.*
566 *Sci. U. S. A.* **102**:1548-1553.
- 567 **43. Shevchenko, A., M. Wilm, O. Vorm, and M. Mann.** 1996. Mass spectrometric
568 sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**:850-858.
- 569 **44. Soltys, B. J., M. Falah, and R. S. Gupta.** 1996. Identification of endoplasmic reticulum in
570 the primitive eukaryote *Giardia lamblia* using cryoelectron microscopy and antibody to Bip. *J.*
571 *Cell Sci.* **109**:1909-1917.

- 572 45. **Stefanic, S., D. Palm, S. G. Svard, and A. B. Hehl.** 2006. Organelle proteomics reveals
573 cargo maturation mechanisms associated with Golgi-like encystation vesicles in the early-
574 diverged protozoan *Giardia lamblia*. *J. Biol. Chem.* **281**:7595-7604.
- 575 46. **Steiner, T. S., N. M. Thielman, and R. L. Guerrant.** 1997. Protozoal agents: what are the
576 dangers for the public water supply? *Annu. Rev. Med.* **48**:329-340.
- 577 47. **Sun, C. H., J. M. McCaffery, D. S. Reiner, and F. D. Gillin.** 2003. Mining the *Giardia*
578 *lamblia* genome for new cyst wall proteins. *J. Biol. Chem.* 278:21701-21708.
- 579 48. **Thirion, J., R. Wattiaux, and M. Jadot.** 2003. The acid phosphatase positive organelles of
580 the *Giardia lamblia* trophozoite contain a membrane bound cathepsin C activity. *Biol. Cell*
581 **95**:99-105.
- 582 49. **Touz, M. C., L. Kulakova, and T. E. Nash.** 2004. Adaptor protein complex 1 mediates the
583 transport of lysosomal proteins from a Golgi-like organelle to peripheral vacuoles in the
584 primitive eukaryote *Giardia lamblia*. *Mol. Biol. Cell* **15**:3053-3060.
- 585 50. **Touz, M. C., M. J. Nores, I. Slavin, C. Carmona, J. T. Conrad, M. R. Mowatt, T. E.**
586 **Nash, C. E. Coronel, and H. D. Luján.** 2002. The activity of a developmentally regulated
587 cysteine proteinase is required for cyst wall formation in the primitive eukaryote *Giardia*
588 *lamblia*. *J. Biol. Chem.* **277**:8474-8481.
- 589 51. **Trombetta, E. S., and A. J. Parodi.** 2003. Quality control and protein folding in the
590 secretory pathway. *Annu. Rev. Cell Dev. Biol.* **19**:649-676.

591 52. **Van Dellen, K. L., A. Chatterjee, D. Ratner, P. E. Magnelli, J. Cipollo, M. Steffen, P. W.**
592 **Robbins, and J. Samuelson.** 2006. Unique posttranslational modifications of chitin-binding
593 lectins of *Entamoeba invadens* cyst walls. *Eukaryotic Cell* **5**:836-848.

594 53. **Wang, Y., and S.-I. Wu, and W. S. Hancock.** 2006. Approaches to the study of N-linked
595 glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled
596 to electrospray linear ion trap—Fourier transform mass spectrometry. *Glycobiology* **16**:514-
597 523.

598 54. **Ward, W., L. Alvarado, N. D. Rawlings, J. C. Engel, C. Franklin, and J. H. McKerrow.**
599 1997. A primitive enzyme for a primitive cell: the protease required for excystation of *Giardia*.
600 *Cell* **89**:437-444.

601 55. **Yates, J. R., 3rd, E. Carmack, L. Hays, A. J. Link, and J. K. Eng.** 1999. Automated
602 protein identification using microcolumn liquid chromatography-tandem mass spectrometry.
603 *Methods Mol. Biol.* **112**:553-569.

604 **Figures legends.**

605 Fig. 1. **Three-dimensional high-resolution fluorescence** microscopy of *Giardia* trophozoites
606 labeled with WGA. While Alexafluor (green in A and B) bound evenly to the surface of
607 trophozoites including the flagella (13), WGA (red in C) bound in a punctate pattern to the
608 surface of non-premeabilized trophozoites with relatively **less** labeling of the flagella. WGA
609 binding to permeabilized trophozoites (red in A and B) was greatest to membranes
610 surrounding the two nuclei, which were labeled blue with DAPI. WGA binding to
611 permeabilized *Giardia* (red in D) extensively overlapped with binding of lysotracker (green in

612 E), so that the overlay of WGA and lysotracker (F) was for the most part yellow (a mix of red
613 and green). All of the images are three-dimensional reconstructions and use pseudocolors.

614 Fig. 2. Protein gels of WGA-affinity purified proteins. A. Coomassie blue stain of *Giardia*
615 proteins before (-) and after (+) WGA-affinity. B. WGA blot of the same *Giardia* proteins
616 transferred to PVDF shows marked enrichment of lectin-binding to *Giardia* proteins after (+)
617 WGA-affinity. C. Zymogram of the same proteins run through a gel containing gelatin shows
618 marked enrichment of two protease bands after (+) WGA affinity. D. Silver stains of *Giardia*
619 proteins after WGA enrichment before (-) and after (+) treatment with PNGaseF, which
620 removes N-glycans. E. WGA blot shows that lectin-binding was removed by (+) treatment with
621 PNGaseF.

622 Fig. 3. A representative experiment showing substantial enrichment of secreted and
623 membrane proteins of *Giardia* using WGA chromatography. Computer-derived two-
624 dimensional gels of mass spectroscopy data of unfractionated trophozoites (A) and trophozoite
625 proteins after WGA affinity (B). The size of each spot is proportional to the coverage of the
626 protein by mass spectroscopy. Secreted proteins, which are enriched after WGA, are marked
627 in red. Cytosolic proteins, which are abundant in unfractionated proteins, are marked in blue.
628 By definition, there is no overlap between cytosolic and secreted proteins, but some proteins
629 may be present in both unfractionated *Giardia* and WGA-enriched fractions. C. Relatively
630 abundant secreted proteins are labeled with a lower case letter on the two dimensional gels
631 and identified by 1) an abbreviated gene number (e.g. 17121 is short for GI50803_17121) and
632 2) annotation at GiardiaDB (e.g. phosphodiesterase) or annotation here (e.g. unique secreted
633 protein) (see also Table 1). Similarly, a sample of cytosolic proteins are numbered in the two-

634 dimensional gels and identified by their abbreviated genes numbers and annotation. OCT =
635 ornithine carbamoyltransferase. PFOR = pyruvate:flavodoxin oxidoreductase.

636 Fig. 4. The single-subunit OST of *Giardia* shows a preference for sequons with Thr *in vivo*. A.
637 Mass spectroscopy of a representative *Giardia* glycopeptide containing an N-glycan composed
638 of GlcNAc₂ showed neutral loss of a single GlcNAc (resulting in a peak of 203.08) or of two
639 GlcNAc (resulting in a peak of 406.16). Other fragments enabled us to confirm the sequence
640 of the glycopeptide. B. Weblogo representation of results from analysis of 91 glycopeptides
641 showed 1) all *Giardia* N-glycans are on sequons (NxT or NxS); 2) there was no negative
642 charge at the -2 position, and 3) the majority of occupied sequons contained Thr. C. While
643 roughly 2/3rds of the occupied sequons of *Giardia* contained Thr, only 50% of the potential
644 sites of N-glycosylation contained Thr. This result showed the OST of *Giardia* prefers sequons
645 with Thr to those with Ser *in vivo*.

646 Fig. 5. The overall composition of secreted and membrane proteins of *Giardia* changes during
647 encystation. Pie diagrams showed average areas for nine sets of trophozoites (A) or 6 sets of
648 cyst (B) proteins, which were enriched by WGA affinity chromatography and identified by mass
649 spectroscopy. Compared to cysts, trophozoites were relatively enriched in Leu-rich repeat
650 proteins (LRR), proteins involved in folding in the ER, and unique secreted proteins. In
651 contrast, cysts were relatively enriched in Gly-rich repeat (GRR) proteins, cyst wall proteins
652 (CWPs), and unique membrane proteins. In addition, cysts contained a small set of
653 encystation-associated proteases (GI50803_2897, a serine protease; GI50803_113303 and
654 GI50803_114165, cysteine protease, and GI50803_14566, a dipeptidyl protease), which were
655 nearly absent in trophozoites (see also Table 2).

656 Fig. 6. A representative experiment showing marked differences in the secreted proteins of
657 trophozoites and cysts, which were affinity purified with WGA and identified by mass
658 spectroscopy. Computer-derived gel comparing secreted proteins of trophozoites (purple) with
659 those of cysts (light green), where proteins present in both trophozoites and cysts appear dark
660 green. Proteins that are relatively enriched in cysts were identified by single lower case letter
661 in the two-dimensional gel, while those that are relatively enriched in trophozoites are marked
662 with a pair of lower cases letters (e.g. aa) (see also Table 2). Stage-specific expression of
663 many different proteins over numerous experiments is shown in hierarchical clusters in Fig. 1
664 of the Supplemental Data.

665 Fig. 7. High-resolution fluorescence microscopy of *Giardia* cysts labeled with WGA (red) and
666 antibodies to the most abundant cyst wall protein (CWP1 in green). *In vitro* derived cysts (A),
667 which excysted poorly, had a well-developed cyst wall, but the nuclei (labeled blue with DAPI)
668 of the encysting organisms were often falling apart. The surfaces of these organisms, which
669 were densely labeled with WGA, were often retracted from the cyst wall. In contrast, cysts
670 isolated from gerbils (B and C), which excysted well, had well-preserved nuclei, and their
671 surfaces, which were densely labeled with WGA, were closely apposed to the cyst wall
672 (arrowheads). Relative to trophozoites, there was less WGA-labeling of perinuclear and
673 peripheral vesicles in cysts. The variability in WGA-labeling of *Giardia* cysts obtained from
674 gerbils in C is caused by the relative impermeability of the intact cyst walls. To increase the
675 penetration of WGA into cysts, *Giardia* were frozen and thawed three times prior to labeling
676 with the lectin. A and B are cross-sections, while C is a three-dimensional reconstruction.

677 Fig. 8. This *Giardia lamblia* Gly-rich-repeat transmembrane protein (GIGRREAT1)
678 (GI50803_114210), was the most abundant protein in the *N*-glycome of *Giardia* cysts.
679 GIGRREAT1 is 1,088 amino acids long and has an *N*-terminal signal sequence (pink) and *C*-
680 terminal transmembrane helix (blue). Gly residues, which are part of degenerate repeats, are
681 highlighted in red. Cys residues, which are part of conserved sequences adjacent to the
682 transmembrane helix, are highlighted in green.

ACCEPTED

Table 1. Abundant *Giardia* trophozoite proteins.

	ORF	Average Area	Troph/Cyst Ratio
Unique Secreted Protein	9780	6.09	3.94
Phosphodiesterase	8360	5.12	2.62
Leucine-rich repeat protein	5795	4.43	6.08
Peroxiredoxin 1	15383	3.34	2.87
Serine protease	10843	2.75	2.77
Protein disulfide isomerase 2	9413	2.71	22.74
Serine peptidase	15871	2.42	2.92
Lysophosphatidic acid phosphatase	11599	2.33	15.76
Leucine-rich repeat protein	11684	2.11	1.48
GRREAT	16507	1.60	0.93
Phospholipase B	93548	1.58	66.13
Alanyl dipeptidyl peptidase	6148	1.56	2.40
Heat shock protein 70	17432	1.54	5.73
Acid phosphatase	7556	1.52	9.52
Alanyl dipeptidyl peptidase	15574	1.40	2.19
Grp94/Hsp90	15247	1.31	Troph only
Unique Secreted Protein	10016	1.29	27.10
Cathepsin L precursor	14983	1.15	3.07
Unique Transmembrane Protein	16916	1.12	6.72
Cathepsin B precursor	14019	1.06	2.97

Table 2. Abundant *Giardia* cyst proteins.

	ORF	Average Area	Cyst/Troph Ratio
Glycine-rich repeat protein	114210	9.22	32.48
Serine protease	2897	6.46	932.44
Cathepsin B-like cysteine protease	113303	4.74	384.75
Absent in current <i>Giardia</i> DB	16318	2.13	17.97
Cathepsin B-like cysteine protease	114165	2.03	155.06
Phosphodiesterase	8360	1.95	0.38
Glycine-rich repeat protein	16507	1.72	1.08
Unique Secreted Protein	9780	1.55	0.25
Leucine-rich repeat protein	11684	1.42	0.68
Glycine-rich repeat protein	7188	1.30	13.28
Amino acid transporter, putative	11299	1.28	110.80
Phospholipase B	17277	1.22	3.08
Peroxiredoxin 1	15383	1.16	0.35
Cyst wall protein 3	2421	1.08	467.33
Variant-specific surface protein	12063	1.08	1.14
Dipeptidyl-peptidase I precursor	14566	1.04	84.63
Serine protease	10843	0.99	0.36
CEGP1 protein	17120	0.98	Cyst only
Unique transmembrane protein	113723	0.92	8.34
Leucine-rich repeat protein	17198	0.86	1.27

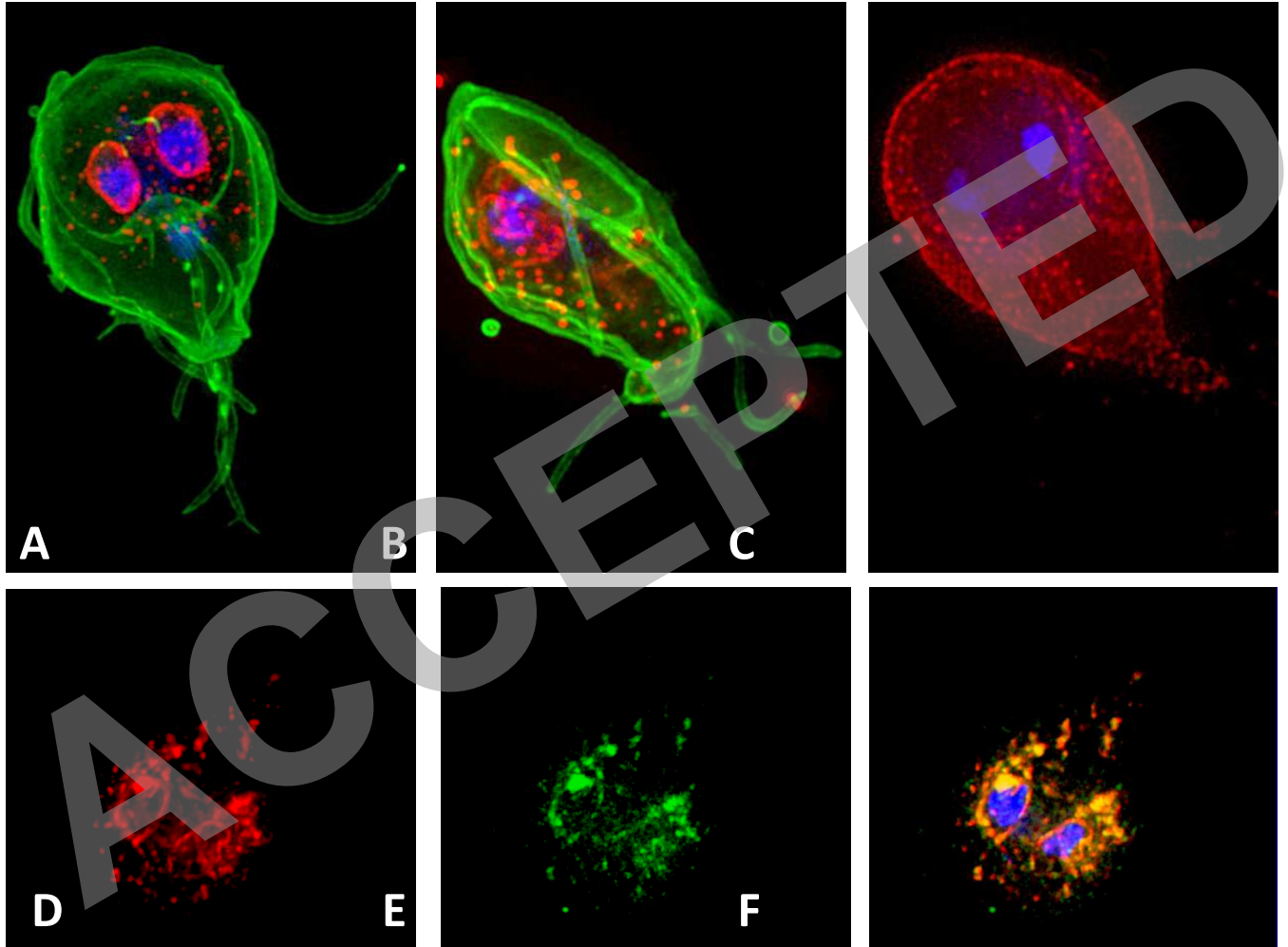


Fig. 1. Revised EC00268-08

A. Coomassie B. WGA blot C. Zymogram D. Silver stain E. WGA blot

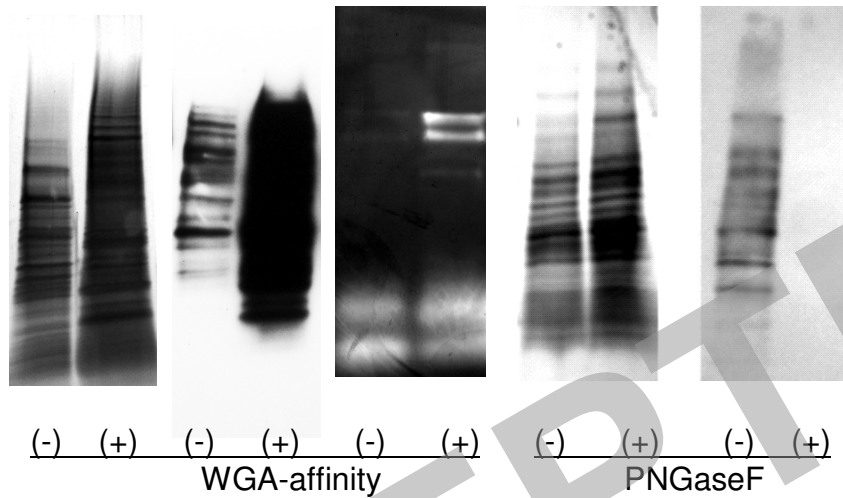


Fig. 2. Revised EC00268-08

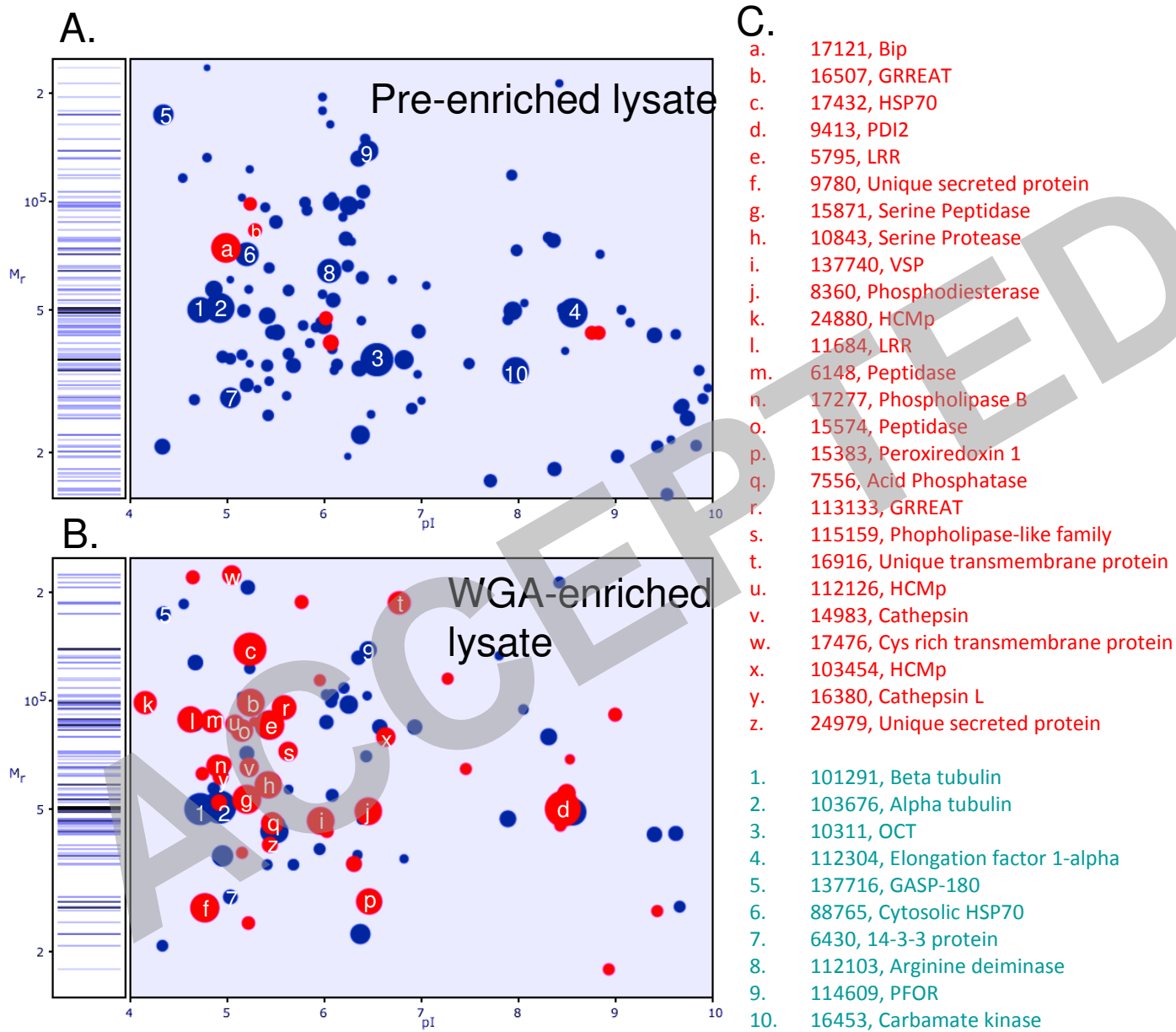


Fig. 3. Revised EC00268-08

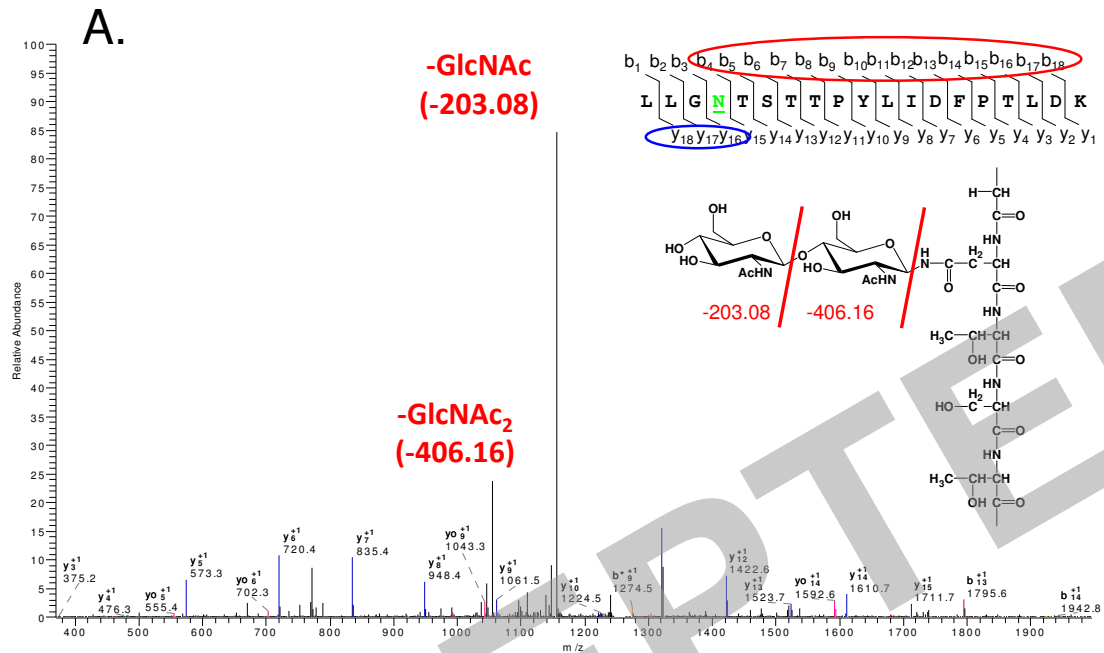


Fig. 4. Revised EC00268-08

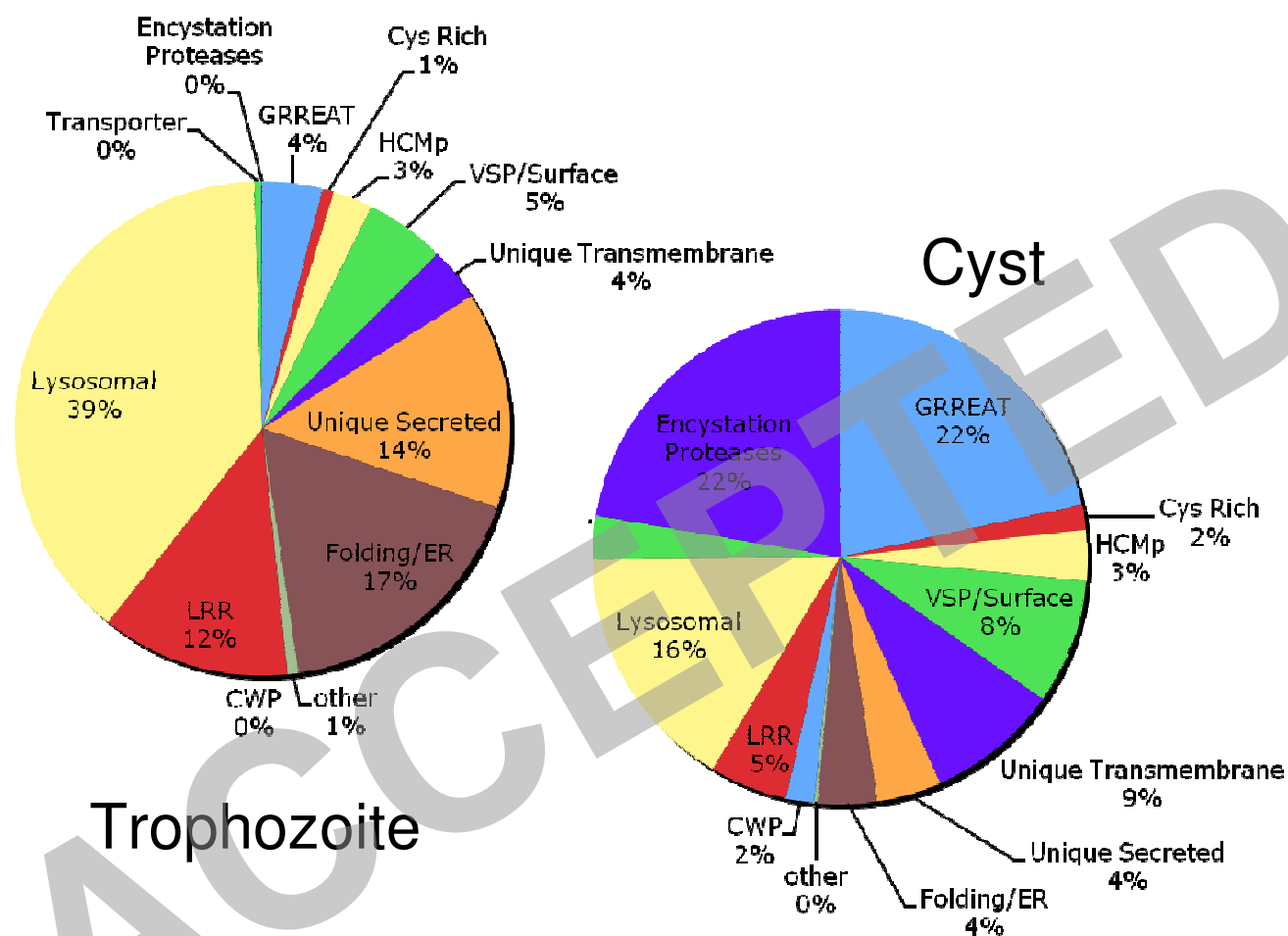
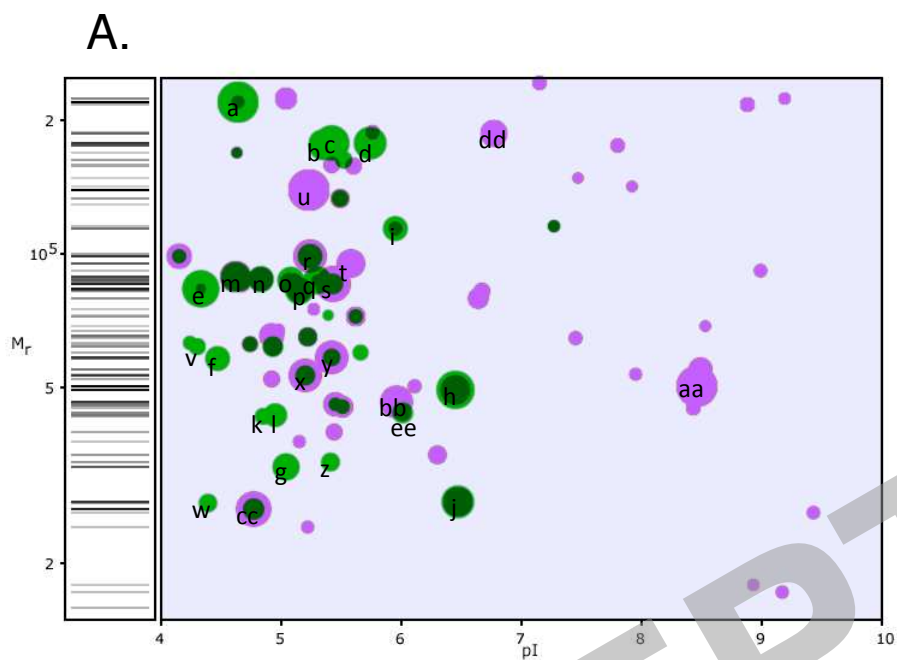


Fig. 5. Revised EC00268-08



B.

- | | |
|----------------------------------|----------------------------|
| a. 115154, VSP-like | p. 15574, Peptidase |
| b. 137727, HCMp | q. 113303, GRREAT |
| c. 114626, HCMp | r. 16507, GRREAT |
| d. 113531, HCMp | s. 5795, LRR |
| e. 2897, Serine protease | t. 113133, GRREAT |
| f. 17120, Unique | u. 17432, HSP70 |
| g. 14019, Cathepsin B | v. 95162, Cys-rich unique |
| h. 8360, Phosphodiesterase | w. 2421, CWP3 |
| i. 114210, GRREAT | x. 15871, Carboxypeptidase |
| j. 15383, Thioredoxin peroxidase | y. 10843, Serine protease |
| k. 113723, Unique | z. 16468, Cathepsin B |
| l. 16622, Unique | aa. 9413, PDI-2 |
| m. 11684, LRR | bb. 137740, VSP |
| n. 6148, Peptidase | cc. 9780, Unique |
| o. 114165, Cathepsin B | dd. 16916, Unique |
| | ee. 12063, VSP |

Fig. 6. Revised EC00268-08

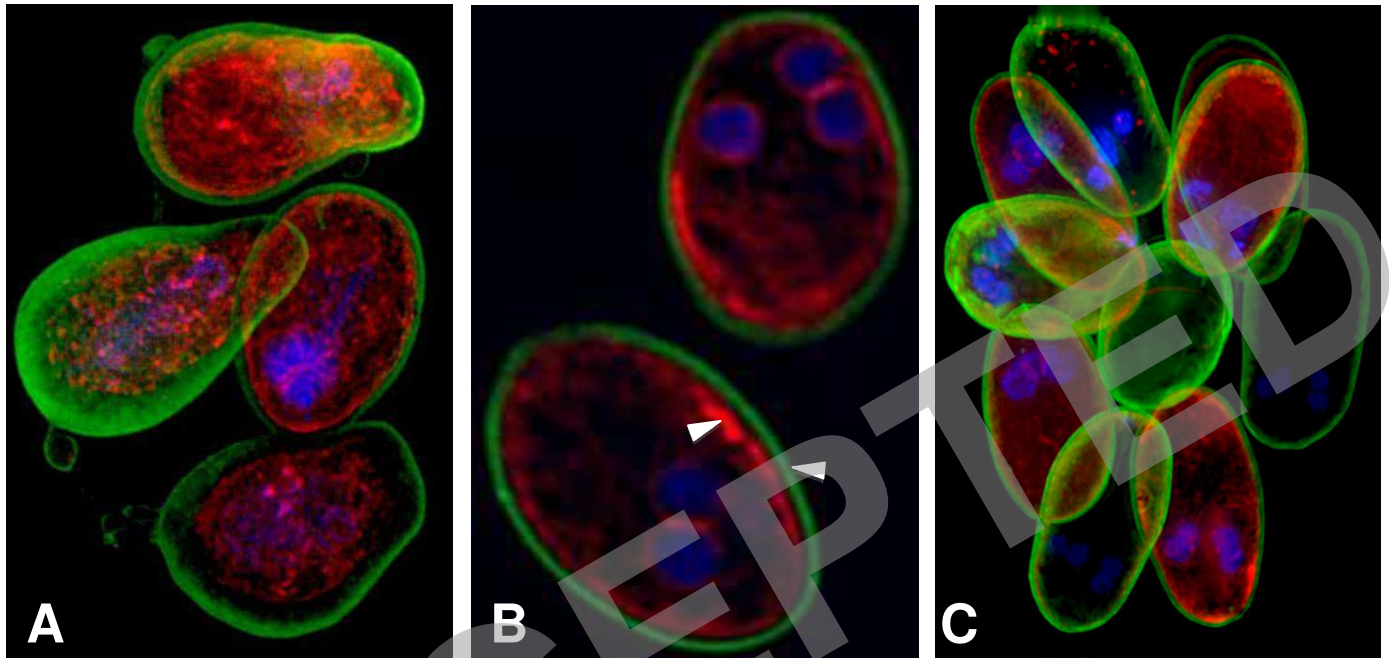


Fig. 7. Revised EC00268-08

MICLTWLLVLSLTEECQR

RDTTIGIYSESDFNKIRDPCCCDYAIGADIAFQRPIVPIPVFT
GSLYGQNHTLSNIRFDLSGTPNEYVNCGIFGFLDQGTVAAVSLVISHTMVDKILNL
GTVAGYAYKPRLYMINVTGTIEGTFSTPSSSEHPYLDKSSSI
GGIVGLENENGEPIITEMRGFVNVSLSNVNSVYYPLNI
GSFFGRFESDTPLOHFIHWSSTMQVALKSNTTAYV
GGAVGYCAADVWHATIGTNQTSGTITGKGTPTLFL
GGAFGSVGGLYKILSWNKNTTCAAEEGTCVS
GAIAGMVSTRKAVYNLTFGGPSTVSCDALSCCT
GGVIGSADAGKFDTMLFNGLSVCKARDTVTC
GGVFGHVFSSLDRIYSYAQSVVASSVASSSDMIMAV
GGLAGIIGDSLNGEMLVINSSVSNVTTVSGTGPGSIFI
GGFAGKIAMQAVRLIRIYSCAAYSNTVNATETSSSYAISNHATA
GGFVAFSKMAKYDFSYSYNSQNI SAESKLSSAAA
GGFVGDGLETHGNHMMIYNAYINGTSGGPEQASVRV
GGFGGGLEGNALVRYTGAFSKALIGKSVQAALNV
GGFAGILQGAHSYGGFCEADIIDVTATGDAGRYSV
GGFVGTNLNGVTSYVNGAWAKANIKVTNPGSAAFDSKKGMRI
GGFVGTMESSVLNSTQAFGSITLSQAQSAKA
GAMVGAHTANISACMSNVTFSSSPSQSDPFKMV
GYS DGTISMDDGSMNFQIVETE GQASETNVTYLTW DKA KKNSSYQTTS
GSYHGFNLNPNRRQYLIKENEYPLQVLPNPGSVDNPRWREGNYC
IFVEFNVNITLSSEYMDGKPFVAAHVAPDT
CPDFSVCRGHYTSPLTVDCNPGFGLRCGNCTTDSACAHGGSCDKSF
GSSYGSCKCVRGKGVDC TYETCGGYKNDQCNGLGMCTYVNDGFYI
CDCPSKEYYIKNQLCAKGCMLTTGMCVPPSNLDRNILCPPGQSYKTF
CGSQSGVNQSGYIASISIMTIIAVVAVILLVLLVALGKVSFT
CCKANGGHQRDLGDSMPLADREMVNSIHSPLDASVRRRIQSVSGGVSIL*

Fig. 8. Revised EC00268-08