



ORIGINAL ARTICLE

Zoonotic *Cryptosporidium* Parasites Possess a Unique Carbohydrate-binding Protein (Malectin) that is Absent in other Apicomplexan Lineages

Mingbo Wu[#], Bo Yang[#], Dongqiang Wang, Ying Zhang, Xiaohui Li, Yubo Zhi, Xinhui Zhao, Jigang Yin and Guan Zhu^{*}

Abstract

Objective. Malectin is a carbohydrate-binding protein that binds Glc(2)-N-glycan and is present in animals and some alveolates. This study aimed to characterize the general molecular and biochemical features of *Cryptosporidium parvum* malectin (CpMal).

Methods. Polyclonal antibodies were raised for detecting native CpMal by western blotting and immunofluorescence assays. Recombinant CpMal and human malectin (HsMal) were produced, and their binding activities to amylose and the host cell surface were compared. Far-western blotting and far-immunofluorescence assays were used to detect potential binding partners of CpMal in the parasite.

Results. Native CpMal appeared to exist in dimeric form in the parasite and was distributed in a diffuse pattern over sporozoites but was highly concentrated on the anterior and posterior sides near the nuclei. CpMal, compared with HsMal, had significantly lower affinity for binding amylose but substantially higher activity for binding host cells. Recombinant CpMal recognized three high molecular weight protein bands and labeled the sporozoite posterior end corresponding to the crystalloid body, thus suggesting the presence of its potential ligands in the parasite. Two proteins identified by proteomics should be prioritized for future validation of CpMal-binding.

Conclusion. CpMal notably differs from HsMal in molecular and biochemical properties; thus, further investigation of its biochemical and biological roles is warranted.

Key words: *Cryptosporidium parvum*, carbohydrate-binding protein, malectin, N-glycosylation, endoplasmic reticulum (ER)

Edited by:

Wei Wang, Jiangsu Institute of Parasitic Diseases, China

Reviewed by:

Jianhai Yin, National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research), Key Laboratory of Parasite and Vector Biology

The other three reviewers chose to be anonymous.

[#]These authors contributed equally to this work.

*Corresponding author:

E-mail: cryptosporida@gmail.com (GZ), wumb9914@mails.jlu.edu.cn (MW), yangbo2015hsy@163.com (BY), wdq19@mails.jlu.edu.cn (DW), ying_zhang21@mails.jlu.edu.cn (YZ), xhli19@mails.jlu.edu.cn (XL), zhiyb9919@mails.jlu.edu.cn (YBZ), zhaoxh9919@mails.jlu.edu.cn (XZ), yinjj@jlu.edu.cn (JY)

Key Laboratory of Zoonosis Research of the Ministry of Education, the Institute of Zoonosis, and the College of Veterinary Medicine, Jilin University, Changchun 130062, China

Received: March 29 2022

Revised: April 18 2022

Accepted: April 20 2022

Published Online: May 10 2022

INTRODUCTION

Cryptosporidium parvum is an important zoonotic protozoan parasite infecting humans and several mammals. The *Cryptosporidium* genus belongs to the phylum Apicomplexa, which contains many medically and veterinarily important

pathogens (e.g., *Plasmodium* spp., *Toxoplasma gondii*, *Eimeria* spp. and *Babesia* spp.). Although the life cycle and morphology of *Cryptosporidium* resemble those of intestinal coccidia, *Cryptosporidium* differs from the coccidia in many ways. Evolutionarily, the *Cryptosporidium* clade forms an early branch at the base of the

apicomplexans away from the coccidian clade [1,2]. In the parasitic lifestyle, *Cryptosporidium* is an intracellular but extra-cytoplasmic parasite (epicellular), rather than residing in the host cell cytosol [3,4]. Metabolically, *Cryptosporidium* lacks chloroplast-derived apicoplasts, typical mitochondria, and their organellar genomes and associated metabolic pathways that are present in other apicomplexans (e.g., cytochrome-based respiration and type II fatty acid synthesis) [1,5,6].

Here, we report that *Cryptosporidium* also differs from other apicomplexans by possessing a malectin, a type of carbohydrate-binding protein that is absent in other apicomplexans. Malectins are small type-I transmembrane proteins (~300 aa) first reported in the frog *Xenopus laevis* for their role in the early processing of *N*-glycosylation (high mannose-type) in the endoplasmic reticulum (ER) lumen [7]. *N*-glycosylation is a protein post-translational modification important in the biological functions of glycoproteins in eukaryotes [8,9]. During early stages of protein *N*-glycosylation in the ER, malectin recognizes and binds $\text{Glc}_2\text{Man}_9\text{GlcNAc}_2$ (or Glc_2 -*N*-glycan for short), an intermediate produced after the removal of the outermost glucose (Glc) from the precursor Glc_3 -*N*-glycan (i.e., $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ attached to an Asn residue). The remaining two glucoses ($\text{Glc}\alpha 1$ -3Glc) are removed in subsequent steps in the ER before the glycoprotein is transported to the Golgi for further processing into various types of high-mannose *N*-glycans. The subcellular location of malectin in the ER has also been confirmed by immunostaining [7,10]. Malectin binds the disaccharide $\text{Glc}\alpha 1$ -3Glc (nigerose; the Glc_2 part of the Glc_2 -*N*-glycan) at a lower binding affinity (~20% activity vs. binding Glc_2 -*N*-glycan), as well as to $\text{Glc}\alpha 1$ -4Glc (maltose) [7].

The binding of Glc_2 -*N*-glycan has been thought to aid in recruitment of glucosidase II (GII; responsible for the removal of the remaining two glucoses) to the Glc_2 -*N*-glycan on the nascent polypeptide [7]. Animal malectins have also been found to participate in the quality control of glycoproteins in the ER [10-14]. Evidence indicates that malectin interacts with ribophorin I (part of the oligosaccharyltransferase [OST] complex) by forming a complex for enhanced association with misfolded glycoproteins [15].

The small malectins are present in only metazoans and some alveolates, whereas malectin domain-containing proteins are present in plants, eubacteria and archaea [16]. In the apicomplexans, we have found that only the *Cryptosporidium* lineage has a malectin, whereas all other apicomplexan lineages lack either malectin or malectin domain-containing proteins. In this study, we report the first characterization of the primary molecular and biochemical features of the malectin from *C. parvum* (CpMal), including its phylogenetic relationship, cellular localization and binding activity towards amylose and host cell surface. This is the first such report for a protozoon. We also confirmed the presence of binding partners of CpMal in the parasite, thereby paving the way for subsequent identification of ligands and investigation of the biological role of CpMal.

MATERIALS AND METHODS

Parasite materials and in vitro culture of *C. parvum*

A strain of *C. parvum* (subtype IIaA17G2R1 at the gp60 locus) was propagated in-house in calves. Oocysts were purified from calf feces with a standard sucrose/cesium chloride gradient centrifugation protocol [17] and stored in PBS containing penicillin (10^4 unit/mL) and streptomycin (10^4 $\mu\text{g}/\text{mL}$) at 4°C. Before experiments, oocysts were treated with 4% sodium hypochlorite for 5 min on ice, then subjected to five or more washes in water by centrifugation. Free sporozoites of *C. parvum* were prepared by excystation in RPMI-1640 medium containing 0.75% bile salt, fixed in 4% paraformaldehyde in PBS for 30 min and washed with PBS by centrifugation.

The in vitro culture of *C. parvum* used ileocecal colorectal adenocarcinoma HCT-8 host cells (ATCC # CCL-244) propagated in RPMI 1640 medium containing 10% fetal bovine serum as previously described [18]. All in vitro experiments were performed at 37°C in a cell culture incubator under 5% CO_2 . Before infection, HCT-8 cells were cultured in 48-well plates until reaching >70% confluence. For immunostaining experiments, plates contained poly *L*-lysine-treated glass coverslips to support the growth of host cell monolayers. Infection started with the addition of chlorine-treated parasite oocysts into the plates (5×10^5 oocysts per well; in vitro excystation rate >80%), followed by incubation at 37°C for 3 h for excystation and invasion, removal of free parasites and oocyst walls, and continual culture of infected cells for various times before collection of specimens, as specified below.

Molecular and phylogenetic analyses

The gene encoding a malectin was identified from the *C. parvum* genome at the locus *cgd6_110* (GenBank: XM_625351). It is described as a “conserved protein with signal peptide and transmembrane domain or GPI anchor signal near C-terminus” in GenBank and as “malectin” in the CryptoDB (<https://www.cryptodb.org/>). In this study, we designate the gene and product as *CpMal* and CpMal, respectively. *CpMal* was defined by an 870 bp open reading frame containing no introns and yielding a 289 aa product. The protein sequence was further analyzed with the InterProScan server for domains and features (<https://www.ebi.ac.uk/interpro/>), thus confirming that CpMal was an authentic malectin ortholog.

For phylogenetic reconstructions, the CpMal sequence was used as the query to search CryptoDB and NCBI's reference protein databases across all major taxonomic groups. Malectin orthologs were identified from chromerids and ciliates but not from dinoflagellates. Among other major taxonomic groups, malectins were identified only from animals that shared reasonably high sequence identities for reliable phylogenetic reconstructions. Finally, a dataset containing malectin orthologs from all available alveolates (*Cryptosporidium*, chromerids and ciliates) and

representative animal species (invertebrates and vertebrates) was built and subjected to multiple sequence alignments in the MUSCLE program (v3.8.31) (<http://www.drive5.com/muscle/>). During the process, identical sequences in the trimmed alignment were deleted (mainly isoforms from the same species). The final dataset contained 17 taxa and 195 amino acid positions.

The Bayesian inference (BI) method was used to construct phylogeny in the MrBayes program (v3.2.6) as described [19]. Model selection for amino acid substitutions was set to “mixed” to allow sampling across all rate matrices. Rate heterogeneity considered the proportion of invariable sites and 4-rate gamma distribution. One million generations of tree searches were performed with two independent searches running with four chains. Trees were sampled in every 1000 generations of the run. A consensus tree was summarized with posterior probabilities from the bottom 75% of the sampled trees, as displayed with the FigTree program (v1.4.4), and annotated with Adobe Illustrator (v25.3).

qRT-PCR of *CpMal* gene transcripts

Total RNA samples were isolated from *C. parvum* oocysts, excysted sporozoites and intracellular parasites cultured in HCT-8 cells for 3 to 72 h with iScript qRT-PCR sample preparation reagent (Bio-Rad Laboratories, Hercules, CA). The detection of *CpMal* transcripts was performed with SYBR Green-based qRT-PCR with a HiScript II One-Step qRT-PCR SYBR Green Kit (Vazyme Biotech Co., Nanjing, China). The levels of *C. parvum* 18S rRNA (*Cp18S*) were detected for normalization [20,21]. The transcripts of the lactate dehydrogenase (*CpLDH*) and elongation factor 1 α (*CpEF1 α*) genes [21–24] were detected in parallel for comparison and quality control. The following primer pairs were used in qRT-PCR: 5′-GCA GTA AAG ACG AAA TCC CAA A-3′ and 5′-ATA TCA GGC TTC TGA CCC TCC T-3′ for *CpMal*; 5′-TAG AGA TTG GAG GTT GTT CCT-3′ and 5′-CTC CAC CAA CTA AGA ACG GCC-3′ for *Cp18S* rRNA; 5′-AAG CAA GGT CTT ATC ACC CAG-3 and 5′-GCA AAG TAG GCA GTT CCT GTC-3′ for *CpLDH* (cgd7_480); and 5′-CTT CCG AAA TGG GTA AGG G-3′ and 5′-TGG GGC ATC AAT GAC AGT G-3′ for *CpEF1 α* (cgd6_3990) [18,23].

Reactions were performed in 20 μ L final volume, containing 0.2 μ M of each primer, 1.0 μ L One Step SYBR enzyme mix, 10 μ L SYBR Green mix, 0.4 μ L ROX reference dye 1 (50 \times), 0.2 ng of total RNA isolated from oocysts/sporozoites or 15 ng total RNA isolated from intracellular parasites (Vazyme Biotech). Thermal cycling started at 50°C for 3 min to synthesize cDNA, followed by incubation at 95°C for 30 s to inactivate the reverse transcriptase and 40 cycles at 95°C for 10 s and 60°C for 30 s to produce amplicons. At least two technical replicate qRT-PCR reactions were performed for each sample. Relative transcript levels were calculated with an empirical $2^{(-\Delta\Delta CT)}$ formula with *Cp18S* transcript used for normalization as previously described [20].

Anti-CpMal antibody production and purification

A short peptide (DEIPKIQRPKPK-C; positions 202–213) unique to CpMal was synthesized by ChinaPeptides Company (Shanghai, China). This peptide is associated with keyhole limpet hemocyanin (KLH) via maleimidobenzoyl-*N*-hydroxysuccinimide ester [25], and it was used to immunize two specific-pathogen-free rabbits with a standard antibody production protocol [26]. Rabbits were subcutaneously administered KLH-linked peptide emulsified Freund’s complete adjuvant for the first injection (300 μ g) or incomplete adjuvants for the three subsequent injections (150 μ g each) in a 2-week interval. Pre-immune sera and antisera were collected before the first injection and 2 weeks after the last injection, respectively. The animal use protocol was reviewed and approved by the Institute Committee for Biosafety and Ethics for Animal Use, Jilin University Institute of Zoonosis (AUP # IZ-2019-084).

Rabbit polyclonal antibody was affinity-purified by a nitrocellulose membrane-based protocol with slight modifications [19,27]. Briefly, 100 μ g peptide dissolved in 300 μ L ddH₂O was immobilized to the membrane (~1.0 cm²); this was followed by blocking with 5% skim milk-TBST buffer (10 mM Tris-HCl 150 mM NaCl and 0.05% Tween-20; pH 8.0) and three washes in TBST, incubation with 4 mL of antisera (1:20 dilution) for 1 h at room temperature and overnight at 4°C, five washes with TBST and elution with 1.0 mL elution buffer containing 0.2 M glycine, 0.15 M NaCl and 0.05% Tween-20 (pH 2.7). Eluted antibody was immediately neutralized with 50 μ L of 1.0 M Tris-HCl buffer (pH 8.0) and dialyzed against PBS as previously described [27]. Affinity-purified antibody was used immediately or stored at –20°C until use. The secondary antibody was goat anti-rabbit IgG conjugated with horseradish peroxidase (Immunoway, Plano, TX, USA) for western blot analysis or goat anti-rabbit IgG conjugated with Alexa Fluor 488 (Invitrogen, Waltham, MA, USA) for immunofluorescence assays (IFAs).

Western blot analysis of native CpMal protein

Native CpMal protein in parasite sporozoites was detected by western blot analysis as previously described [19,28]. Free sporozoites were prepared as described above and suspended in RIPA lysis buffer (Thermo Fisher Scientific, Carlsbad, CA, USA) (10⁷ oocysts in 20 μ L) containing a protease inhibitor cocktail, disrupted by ten freeze/thaw cycles and centrifuged at 15,000 g for 15 min. The supernatants were mixed with loading buffer, heated at 95°C for 5 min and electrophoresed on 10% SDS-PAGE (10⁷ sporozoites per lane). After electrophoresis, proteins were transferred onto nitrocellulose membranes in a semi-dry transfer apparatus (Bio-Rad Laboratories), and this was followed by blocking for 1 h in TBST buffer containing 5% skim milk, incubation with affinity-purified rabbit anti-CpMal antibody in TBST (1:50 dilution) for 1 h and incubation with HRP-conjugated goat anti-rabbit IgG antibody (Immunoway, 1:10,000 dilution) for 1 h. Three or more washes with PBST were performed after each incubation step, and all

procedures were conducted at room temperature or as specified. The blots were developed with an enhanced chemiluminescence reagent and visualized with UVP Chemstudio analyzer (Analytik Jena, Upland, CA, USA),

IFA detection of CpMal in the parasite

Excysted sporozoites of *C. parvum* were prepared as described above and fixed in 4% paraformaldehyde for 20 min. Intact oocysts were suspended in 4% paraformaldehyde and disrupted by three freeze/thaw cycles to allow antibody access to the internal sporozoites and structures. Fixed oocysts and sporozoites were washed three times with PBS by centrifugation and applied on poly *L*-lysine-treated microscopic slides. Intracellular parasites grown with HCT-8 cells for 24 to 48 h on coverslips were prepared as described above and fixed in 4% paraformaldehyde. All samples were washed three times in PBS, permeabilized with 0.1% Triton X-100 in PBS for 5 min, blocked with 3% BSA/PBS for 50 min at room temperature, incubated with primary antibodies (i.e., purified anti-CpMal antibody at 1:10 dilution) in 3% BSA/PBS at 4°C overnight, labeled with goat anti-rabbit IgG conjugated with Alexa Fluor 488 (1:2000) (Invitrogen, Waltham, MA, USA) at 37°C for 1 h, counterstained with 4',6-diamidino-2-phenylindole (1.0 µg/mL) for 5 min and mounted with antifade mounting medium (Beyotime Biotechnology, Shanghai, China). Specimens were examined under a BX53 research microscope (Olympus, Tokyo, Japan).

Heterologous expression of recombinant CpMal and human malectin

For expression of recombinant CpMal, a DNA fragment encoding the non-cytoplasmic region of CpMal was amplified by PCR from the genomic DNA isolated from *C. parvum* oocysts, amplified by PCR (length = 224 aa; amino acid positions from 27 to 250) (Fig 1A). For expression of recombinant human malectin (HsMal), a fragment encoding the non-cytoplasmic region was amplified by PCR from a cDNA reverse-transcribed from total RNA isolated from HCT-8 cells with a PrimeScript reagent kit (length = 187 aa; amino acid positions from 42 to 228 based on GenBank # BC016297) (Takara Bio Inc, Kusatsu, Japan). The following primers were used: 5'-GAT CTG GTT CCG CGT GGA TCC GAA GTC ATT TAC GCC GTG AA-3' and 5'-CTC GAG TCG ACC CGG GAA TTC TAA TTC TTT AAC AGT GAA AAG AGG T-3' (*Bam*H I or *Eco*R I; restriction sites are underlined), and 5'-AAT CGG ATC TGG TTC CGC GTG GAT CCG CAG GCCTGC CGG AA-3' and 5'-AGT CAG TCA CGA TGC GGC CGC TCC AGT CAT TCC AGG CCC GGA TGC-3' (*Bam*H I or *Xho* I; restriction sites are underlined). Thermal cycling used the following conditions: denaturation of templates at 95°C for 5 min; 35 cycles at 94°C for 30 s, 50°C or 55°C for 45 s (for *CpMal* or *HsMal*, respectively) and 72°C for 90 s; and a final extension at 72°C for 10 min.

The PCR products were purified with a Gel/PCR Extraction Kit (Solarbio, Beijing, China) and cloned into

pGEX-4T-1 vector (Invitrogen) with a ClonExpress MultiS One Step Cloning Kit (Vazyme, Nanjing, China) for expression as a glutathione-S-transferase (GST)-fusion protein. Recombinant proteins were expressed in the BL21(DE3) strain of *Escherichia coli* (Tiangen Biotech Co., Beijing, China) according to standard protocols. Recombinant proteins were purified with glutathione-Sepharose-based affinity chromatography with glutathione-Sepharose 4B, according to the manufacturer's instructions (GE Healthcare, Stockholm, Sweden). The purity and molecular weight were evaluated with SDS-PAGE gels stained with Coomassie blue.

Evaluation of the binding activity of CpMal and HsMal to amylose

GST-fused CpMal or HsMal protein (designated as GST-CpMal or GST-HsMal; 0.5 µM) was mixed with 40 µL of amylose resin (New England Biolabs, Ipswich, MA, USA) in 400 µL of PBS and incubated for 30 min at room temperature. After centrifugation at 800 g for 4 min at 4°C, the resin pellets were washed three times with PBS by centrifugation and resuspended in 40 µL PBS. After addition of 10 µL of 5× loading buffer, samples were subjected to SDS-PAGE fractionation and transfer to nitrocellulose membranes. The detection of proteins on the blots followed the same procedures as those for western blot analysis described above. The relative intensities of the protein bands were analyzed in ImageJ software (<https://imagej.nih.gov/ij/>).

Evaluation of the binding activity of CpMal and HsMal to host cells

The binding activity of GST-CpMal and GST-HsMal to host cells was assessed with a protocol similar to ELISA [28,29]. HCT-8 cells were cultured to 100% confluence in 96-well plates, washed three times with PBST and fixed with 1% glutaraldehyde in PBS for 30 min. After three washes with PBST, plates were blocked with 5% skim milk in PBST for 1 h and incubated with GST-CpMal or GST-HsMal proteins (0 to 20 µM) in PBS containing 1 mM CaCl₂ and 0.5 mM MgCl₂ for 1 h in the 37°C. GST-tag at the same molar concentrations was used as a negative control and for background subtraction. Recombinant proteins bound to the host cell surface were detected by incubation with a monoclonal anti-GST antibody (ABclonal Technology Co., Wuhan China) at 37°C for 1 h, rinsed with PBST three times and incubated with alkaline phosphatase-conjugated goat anti-mouse IgG at 37°C for 1 h. After three washes with PBST, specimens were developed with the substrate p-nitrophenyl-phosphate, and the optical density at 405 nm (OD₄₀₅) was measured.

Detection of potential binding partners of CpMal in sporozoites

Two approaches were used to detect potential binding partners of CpMal in the parasite. The first used a far-western blot assays [30], in which protein extracts from excysted sporozoites (10⁷ per lane) were electrophoresed by SDS-PAGE,

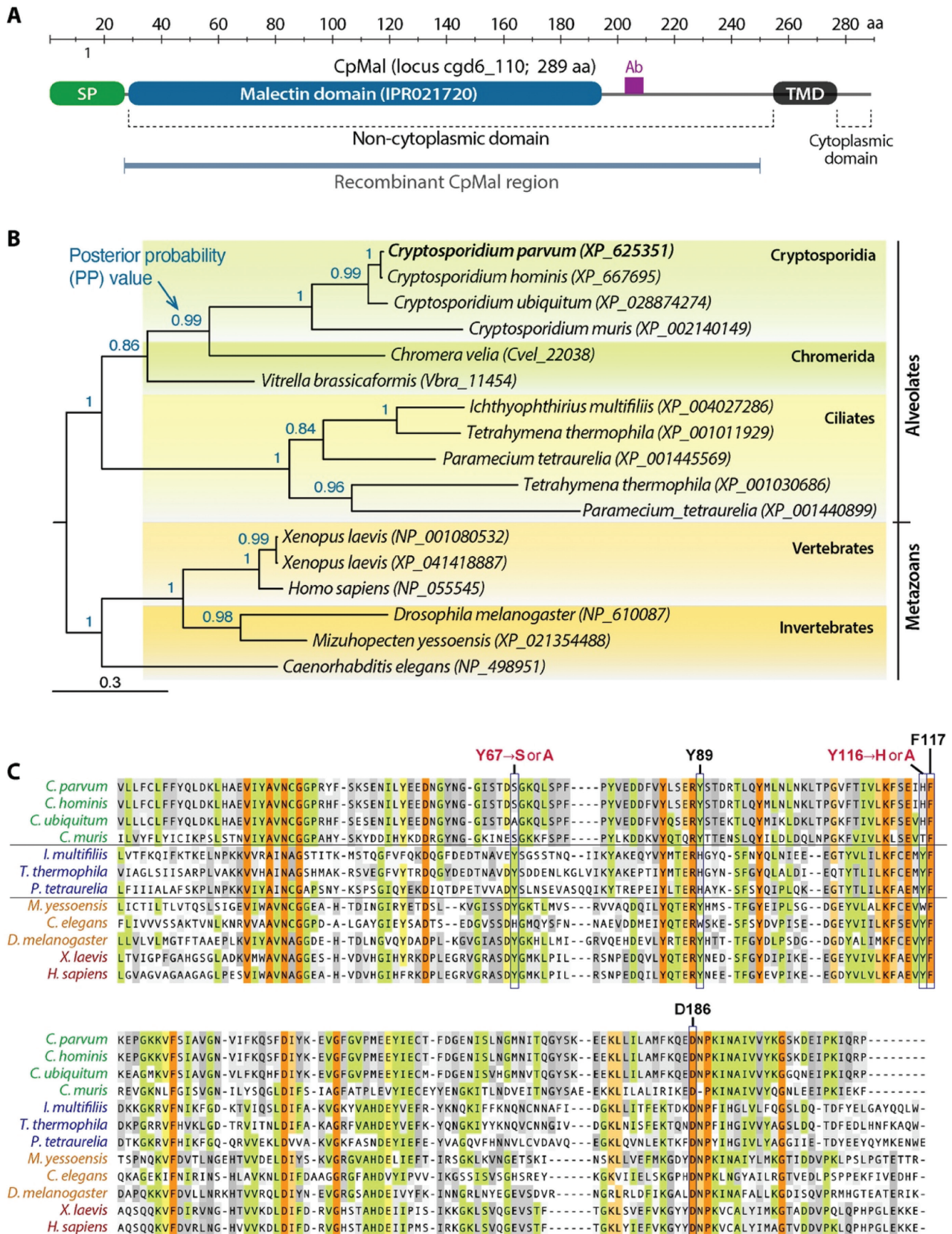


FIGURE 1 | Molecular features of *Cryptosporidium parvum* malectin (CpMal). A) Domain organization of CpMal (289 aa) containing an N-terminal signal peptide (SP; 1–26 aa), a malectin domain (27–194 aa), a transmembrane domain (TMD) near the C-terminus (259–282 aa) and a C-terminal domain (283–289 aa). In this single-pass type I transmembrane protein, the TMD separates the N-terminal non-cytoplasmic and C-terminal cytoplasmic domains. Site recognized by the rabbit polyclonal antibody (Ab) are marked (202–213 aa). B) Phylogenetic tree reconstructed by Bayesian inference (BI) analysis based on protein sequences available from the alveolate genomes and representative animal

species (17 taxa and 195 characters). Malectin domain-containing sequences in plants and other taxonomic groups were found to be highly divergent from those in animals and alveolates, and were excluded from the analysis. Numbers at the nodes are posterior probability (PP) values summarized from the bottom 75% trees derived from 10^6 generations of tree searches. Amino acid substitutions used the WAG model with the considerations of fraction of invariance and 4-rate gamma distribution. The tree is arbitrarily rooted by using animal sequences as the outgroup. The scale bar indicates the amino acid substitution rate. C) Multiple alignment of malectin protein sequences from representative *Cryptosporidium* species, ciliates and animals at the malectin domain. Different color shades indicate the degree of amino acid identity at the positions. Positions of the five residues known for mediating carbohydrate-binding are boxed, and conserved and mutated residues between *Cryptosporidium* and animal malectins are marked on top in black and red, respectively. Amino acid positions refer to the *X. laevis* sequence.

transferred onto a polyvinylidene fluoride membrane and blocked as described above. Blots were then incubated with GST-CpMal (50 μ g in 4 mL PBS) for 1 h and washed five times in TBST. The subsequent procedures for detecting the protein bands followed the same steps for western blot analysis as described above. To identify the putative CpMal partners/ligands, we excised areas in the blot corresponding to the three bands recognized by GST-CpMal for proteomic analysis (Beijing Protein Innovation Co., Beijing, China). Briefly, samples were digested with trypsin overnight and subjected to liquid chromatography with tandem mass-spectrum (LC-MS/MS) analysis according to standard protocols. The mass spectral data were managed with the Mascot platform (v2.3.01; Matrix Science, UK) for the identification of peptides by searching the NIST peptide spectral libraries with the MS PepSearch engine. Identified peptides were further mapped to specific proteins by searching of the UniProt and CryptoDB protein databases.

The second assay used a procedure similar to IFA, in which GST-CpMal protein was first incubated with excysted sporozoites to label potential binding partners and then detected by IFA. For clarity, following the terminology of far-western blotting, we named this assay “far-IFA.” In this far-IFA assay, excysted sporozoites were prepared, fixed in paraformaldehyde, applied onto microscopic slides, permeabilized and blocked as described above. Specimens were incubated with GST-CpMal (15 μ M in 30 μ L solution) for 1 h and washed three times with PBS. The subsequent procedures including incubation with mouse anti-GST monoclonal antibody and Alexa Fluor 488-conjugated goat anti-mouse IgG antibody followed the same steps as those for IFA described above.

RESULTS

Malectin or malectin domain-containing proteins are present in only select members of the SAR supergroup

Although protein glycosylation is widely present in eukaryotes, and malectin was discovered in vertebrates for its function in *N*-linked glycosylation, malectin orthologs or malectin domain-containing proteins are present in limited taxonomic groups [16]. In the phylum Apicomplexa, genes encoding malectin were found in only *Cryptosporidium* (Fig 1). At a higher taxonomic level (the SAR supergroup), genes encoding malectin or malectin domain-containing proteins were identified in the genomes of some chromerids and ciliates, but not in stramenopiles and Rhizaria. With the

CpMal protein sequence as the query, we searched NCBI's protein databases with the exclusion of *Cryptosporidium* sequences. The top hits were malectin orthologs from invertebrates rather than ciliates (e.g., E-values = $2e-21$ with 38.17% identity to the ortholog from the marine pennis worm *Priapulius caudatus* [XP_014674267] vs. E-values = $1e-14$ with 29.12% identity to the ortholog from the ciliate *Ichthyophthirius multifiliis* [XP_004027286]).

In contrast, in our BI-based phylogenetic reconstructions on malectin orthologs from all available alveolate sequences and representative animal sequences, *Cryptosporidium* sequences clustered with chromerids, rather than with animal or ciliate sequences (Fig 1B). The phylogenetic affiliation between cryptosporidium and chromerid sequences was strongly supported by the posterior probability (PP; value = 0.86). In this BI tree, ciliates and animals formed two separate clades that were robustly supported by posterior analysis (PP = 1.0 for both clades). The same topology was also obtained through maximum likelihood-based phylogenetic reconstructions (data not shown). The data suggested that *Cryptosporidium* and chromerid malectins are likely to share a common evolutionary origin, and weakly implied that the ancestral apicomplexans might contain malectins, but malectins have been lost in most the apicomplexan lineages. Our phylogenetic analysis was unable to indicate whether the *Cryptosporidium* and ciliate malectins shared a common ancestor, because of the lack of orthologs within alveolates and among the three major clusters. Although the tree displayed in Fig 1B was arbitrarily rooted with animal malectins as an outgroup, the cryptosporidia/chromerida clade could be placed closer to the animal clade than the ciliate clade by mid-point rooting; however, this finding might have simply been an artifact of long-branch attraction between highly divergent sequences.

Apicomplexan lineages vary in synthesis and processing of *N*-glycans: implications for the function of *Cryptosporidium* malectin

In animals, the endogenous ligand of malectin is the high-mannose Glc₂-*N*-glycan, which is an intermediate after the removal of the outermost glucose from the precursor Glc₃-*N*-glycan by glucosidases I (GI) (Fig 2A). Bound malectin facilitates the recruitment of glucosidases II (GII), which are responsible for the removal of the two glucoses in Glc₂-*N*-glycans. In apicomplexans, the synthesis of *N*-glycan precursors is highly divergent among lineages. Datamining

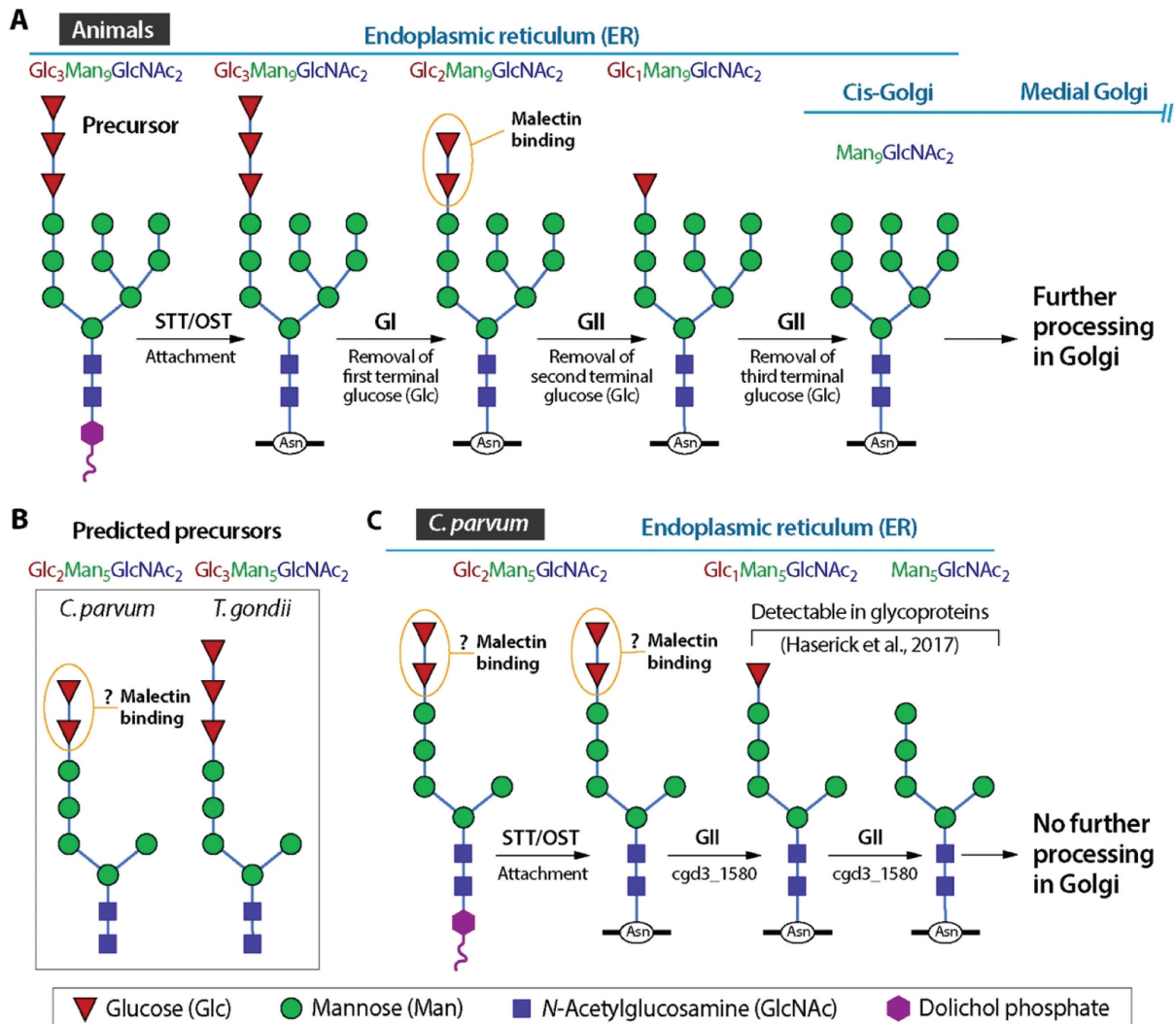


FIGURE 2 | Comparison of the *N*-glycan precursors and trimming of terminal glucoses from the precursors between animals and *Cryptosporidium*. A) Illustration of the *N*-glycan precursor ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) and trimming of the terminal glucoses by glucosidases I and II (GI and GII) in the endoplasmic reticulum (ER). B) Predicted *N*-glycan precursor in *C. parvum* ($\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$) in comparison with that from the cyst-forming apicomplexan *Toxoplasma gondii* ($\text{Glc}_3\text{Man}_5\text{GlcNAc}_2$). C) Predicted trimming steps of *N*-glycan precursors in the ER of *C. parvum*. In panels B and C, the prediction was based on enzymes identifiable from genome sequences together with available experimental evidence (details in main text). Question marks indicate that the binding of CpMal to $\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$ is not fully confirmed and requires further experimental validation. STT/OST, asparagine *N*-glycosyltransferase/oligosaccharyltransferase complex.

the apicomplexan genomes identified enzymes for synthesizing precursors from the coccidia (e.g., *Toxoplasma*) and cryptosporidia, but not the hematozoa, including *Plasmodium*, *Babesia* and *Theileria*. However, *Cryptosporidium* and *Toxoplasma* can synthesize only “simplified” *N*-glycan precursors because they lack alpha-1,6-mannosyltransferase (ALG12) and alpha-1,3-mannosyltransferase (ALG3). Additionally, *Cryptosporidium* further differs from *Toxoplasma* by lacking an alpha-1,2-glucosyltransferase (ALG10); therefore, their *N*-glycan precursors are predicted to be single long branched $\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$ and $\text{Glc}_3\text{Man}_5\text{GlcNAc}_2$, respectively (Fig 2B). This prediction has been supported by mass spectrum-based analyses of *N*-glycans in *C. parvum* and *T. gondii* [31,32]. Therefore, apicomplexans differ not only from animals but also between lineages in synthesizing the *N*-glycan precursors.

In trimming glucoses from the Glc_2 - and Glc_3 -*N*-glycan precursors, *T. gondii* has glucosidases I (GI) and II (GII), whereas *C. parvum* has only GII (Fig 2C). Therefore, the trimming of Glc_3 -*N*-glycan precursor in *Toxoplasma* resembles that in metazoans. However, *T. gondii* lacks malectin, thus indicating that the binding of malectin to Glc_2 -*N*-glycans is inessential for the *N*-glycosylation in the coccidia. In contrast, *Cryptosporidium* has a malectin but synthesizes Glc_2 -*N*-glycan as the precursor. If the malectin in *Cryptosporidium* also specifically binds Glc_2 -*N*-glycan ($\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$), the binding would start as soon as the precursor Glc_2 -*N*-glycan is synthesized (synthesis of the precursor), and continue to the attachment of Glc_2 -*N*-glycan to a protein and trimming of the two terminal glucoses (early processing of the precursor).

Cryptosporidium malectins might be predicted to have different biochemical and biological properties from the

malectins in the human and animal hosts. This notion is also suggested by sequence divergence of malectins between *Cryptosporidium* and hosts, and further supported by the binding assays as described below. In animal malectins, five amino acids were identified to mediate the carbohydrate-binding: the four aromatic residues Y67, Y89, Y116 and F117, and the aspartate D186 (positions based on *X. laevis* sequence NM_001091743), which were highly conserved in animals. In *Cryptosporidium*, malectins were relatively divergent between the intestinal species (e.g., *C. parvum* and *C. ubiquitum*) and gastric species (e.g., *C. muris*) groups, but highly conserved within each group (Fig 1B, C). Among the five binding site residues in *Cryptosporidium* malectins, three residues were identical to those in animals (i.e., Y89, F117 and D186), whereas the other two differed (i.e., the aromatic Y67 and Y116 were replaced by non-aromatic residues S/A and H/A, respectively) (Fig 1C).

The *CpMal* gene is expressed, and CpMal protein is present, in the parasite extracellular and intracellular developmental stages

CpMal is a typical malectin, which is small (289 aa), and contains an N-terminal signal peptide for targeting the protein to the ER, a malectin domain and a transmembrane domain (TMD) separating the long N-terminal non-cytoplasmic and short C-terminal cytoplasmic domains (Fig 1A). *CpMal* gene transcripts were detected by qRT-PCR in all developmental stages, thereby indicating that the gene was continually expressed (Fig 3A). The highest levels of *CpMal* transcript (normalized to those of *Cp18S* transcript) were detected in sporozoites and intracellular parasites at 72 h post-infection (hpi), followed by oocysts and parasites at 48 hpi. The levels of *CpMal* transcript were lowest in intracellular parasites between 3 and 24 hpi. The reliability of the qRT-PCR data were validated by parallel detection

of the transcripts of the previously reported *CpLDH* and *CpEF1 α* genes, which showed the expected expression patterns [21,22,24]. In summary, the levels of *CpMal* transcripts were relatively high in the extracellular stages (i.e., oocysts and sporozoites) and later intracellular stages corresponding to more advanced sexual development (i.e., 48 and 72 hpi). However, the biological importance of the varied expression levels requires further investigation.

To detect the native CpMal protein in the parasite, rabbit polyclonal antibodies were raised against an epitope in the non-cytoplasmic region (position marked in Fig 1A). In western blot analysis, affinity-purified anti-CpMal antibody recognized a single band from the sporozoite crude extract (Fig 3B), thus supporting the specificity of the antibody. However, the detected band was at ~70 kDa, nearly two times the predicted molecular weight (33 kDa). This phenomenon was persistent despite multiple attempts to change the experimental conditions. We hence concluded that the native CpMal protein was present in the parasite cells in a stable dimeric form. This notion was partly supported by the western blot detection of dimeric human malectin with a rabbit polyclonal antibody by Abcam PLC (<https://www.abcam.com/malectin-antibody-ab97616.html>; product # ab97616).

In IFAs, anti-CpMal antibody produced strong signals in the sporozoites within the oocysts (Fig 4A). The subcellular locations of the signals could not be resolved, owing to the crowding of the four sporozoites in the oocysts, in which the oocyst walls were ruptured by repeated freeze/thaw to allow access of antibodies. These results confirmed that malectin was present in sporozoites but not in any other oocyst structures, such as the lumens and walls of oocysts. In excysted sporozoites, CpMal showed a relatively diffuse pattern of distribution, but two spots on the anterior and posterior sides near the nuclei showed much stronger signals (Fig 4B).

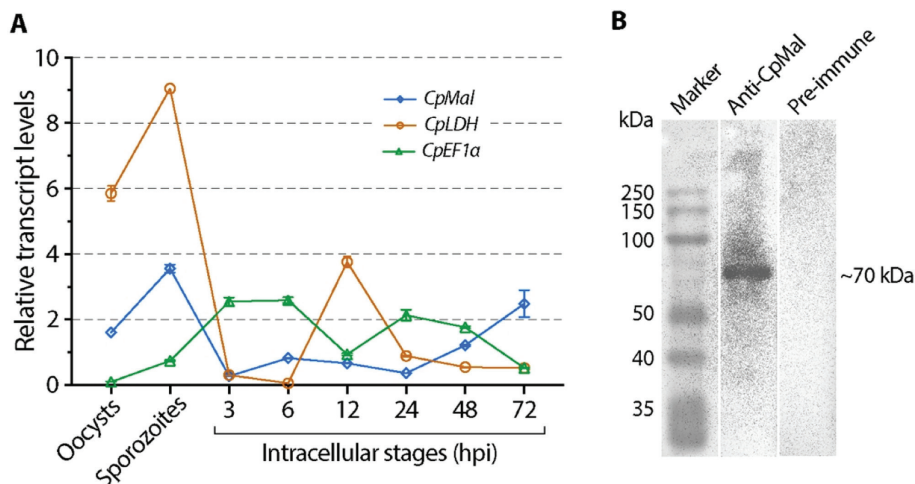


FIGURE 3 | Detection of *CpMal* transcripts by qRT-PCR and native CpMal protein by western blot analysis. A) Relative levels of *CpMal* transcripts in oocysts, excysted sporozoites and parasite intracellular stages, as determined by qRT-PCR and normalized to 18S rRNA (*Cp18S*). Transcripts of *CpLDH* and *CpEF1 α* genes were detected in parallel as references and for quality control. B) Western blot detection of native CpMal protein from the sporozoite lysates with an affinity-purified rabbit polyclonal anti-CpMal antibody. Pre-immune serum was subjected to the same affinity-purification procedures and used to probe the blot under the same experimental conditions.

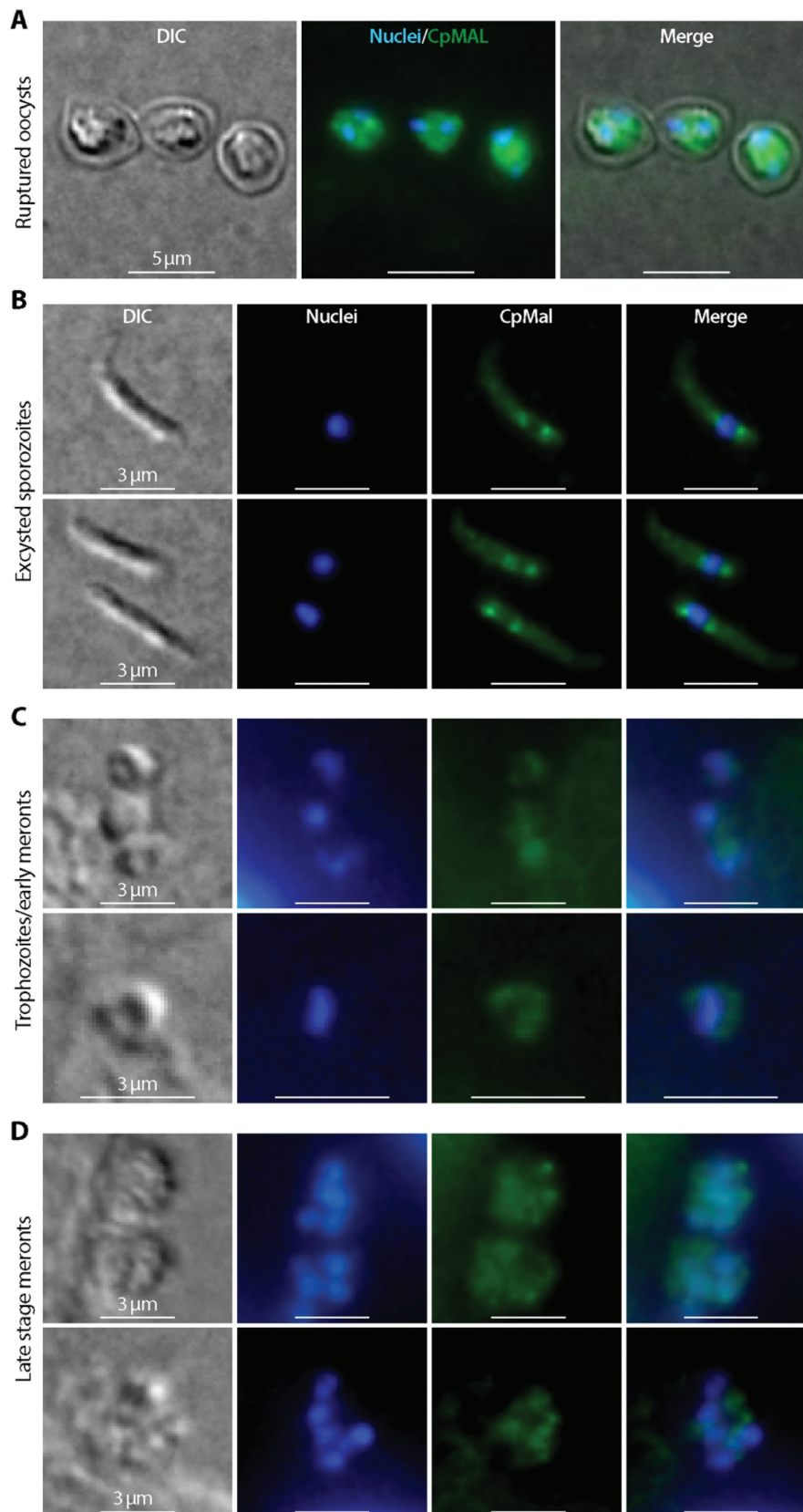


FIGURE 4 | Immunofluorescence assays (IFAs) using an affinity-purified anti-CpMal antibody to detect CpMal in the *C. parvum* oocysts ruptured by freeze/thaw cycles (A), excysted sporozoites (B) and intracellular developmental stages (C and D).

Although the structure of the ER network in *C. parvum* has not been fully defined, the IFA signals were expected for an ER network (i.e., usually all over the cytosol but more concentrated near the nuclei).

In the intracellular meronts, immunostaining produced signals that were generally weak but slightly stronger than the background (Fig 4C, D). The subcellular location of the signals was not well resolved, owing to the limited resolution of fluorescence microscopy. However, the results were sufficient to confirm that CpMal was present in the meronts contained within but not on the parasitophorous vacuole membrane, as seen for CpLDH and some other *C. parvum* proteins (e.g., [22,33]).

CpMal and HsMal differ in their binding affinity to amylose and the host cell surface

To gain a basic understanding of the carbohydrate-binding properties between the parasite and host malectins, we compared the binding affinity of CpMal and HsMal to amylose, which could be considered a polymer of maltose, and to the surfaces of fixed HCT-8 cells with various extracellular glycoproteins. We observed significant differences in binding affinities between GST-CpMal and GST-HsMal proteins. In amylose-binding assays, GST-CpMal displayed significantly weaker binding activity than GST-HsMal (i.e., $53.5 \pm 0.41\%$ vs. $100 \pm 0.72\%$) (Fig 5A, B). In contrast, CpMal showed much stronger binding activity to the host cell surface than HsMal (Fig 5C). The cell surface-binding activity of HsMal was extremely weak, only slightly above the GST background. Among the tested concentrations, CpMal displayed 6.1-fold and 6.3-fold higher binding activity than HsMal at 10 and 20 μM , respectively. Because Glc_2 -*N*-glycan was an intermediate form not expected to be present on the host cell surface, the observed low binding activity of HsMal was probably attributable to its weak affinity toward other polysaccharides on the cell surface. The relatively strong binding activity of CpMal might have resulted from its interaction

with unknown protein domains rather than carbohydrates. This possibility is partly supported by the ineffectiveness of maltose (10 mM) in the cell-binding of CpMal (Fig 5D). We were unable to compare substrate preferences between CpMal and HsMal because of the unavailability of reagents (e.g., disaccharide arrays and $\text{Glc}_{(1 \text{ to } 3)}\text{Man}_{(5 \text{ or } 9)}\text{GlcNAc}_2$ -*N*-glycans) and current technical obstacles in preparing these reagents. However, the amylose- and cell-binding results supported the conclusion that CpMal significantly differed from HsMal in binding properties.

Cryptosporidium parvum contains potential binding partners/ligands for CpMal

In addition to binding Glc_2 -*N*-glycan, malectin was found to interact with other proteins by forming a complex with ribophorin I for enhanced association between ribophorin I and misfolded glycoproteins [12,15]. In this study, we attempted to detect potential binding partners of CpMal in the parasite. In far-western blot analysis with GST-CpMal to probe the fractionated sporozoite lysates, followed by western blotting to detect the GST-tag, CpMal recognized three protein bands with sizes ranging from ~140 to 250 kDa (Fig 6A). The observed binding of CpMal was specific, because no bands were detected with GST-tag as the probe.

The three bands were excised for proteomic analysis, in which a total of 15 proteins were identified with calculated molecular weights (MWs) ranging from 11.3 to 279.5 kDa (Table 1). The ten proteins with lower than expected MWs (i.e., 112.4 kDa or lower) were likely contaminants, because they were primarily proteins known for their high abundance in cells (e.g., Hsp70, elongation factor 1 α and histones) and/or mostly showed low numbers of MS/MS spectrum matches and low scores. Among the five high MW proteins (176.2 kDa or higher), two showed both high scores and high spectral matches and therefore should be prioritized for further investigation: 1) a 1,769 aa protein annotated as “amine oxidase” in the CryptoDB (gene ID:

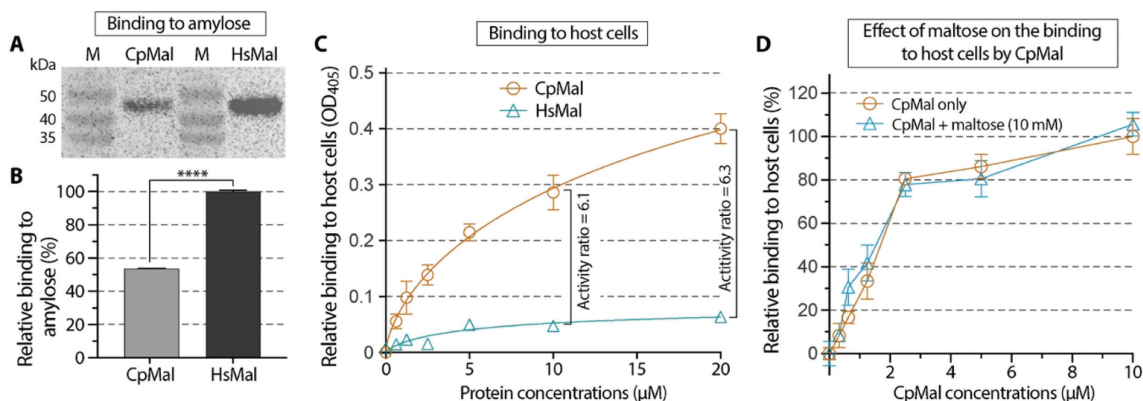


FIGURE 5 | Detection of the binding activities of glutathione-S-transferase (GST)-fused recombinant *C. parvum* and human malectins (CpMal and HsMal). A) Western blot analysis of GST-CpMal and GST-HsMal bound to amylose-resin, by using an anti-GST antibody. B) Relative levels of GST-CpMal and GST-HsMal proteins bound to the amylose-resin measured from the blots ($n = 3$). **** $p < 0.0001$ by two-tailed Student's *t*-test. C) Binding kinetics of GST-CpMal and GST-HsMal proteins to the fixed HCT-8 cells. Bound proteins were detected with anti-GST antibodies. GST-tag only was used in the assay and used for background subtraction. D) Effect of maltose (10 mM) on the binding of GST-CpMal to the host cell surface.

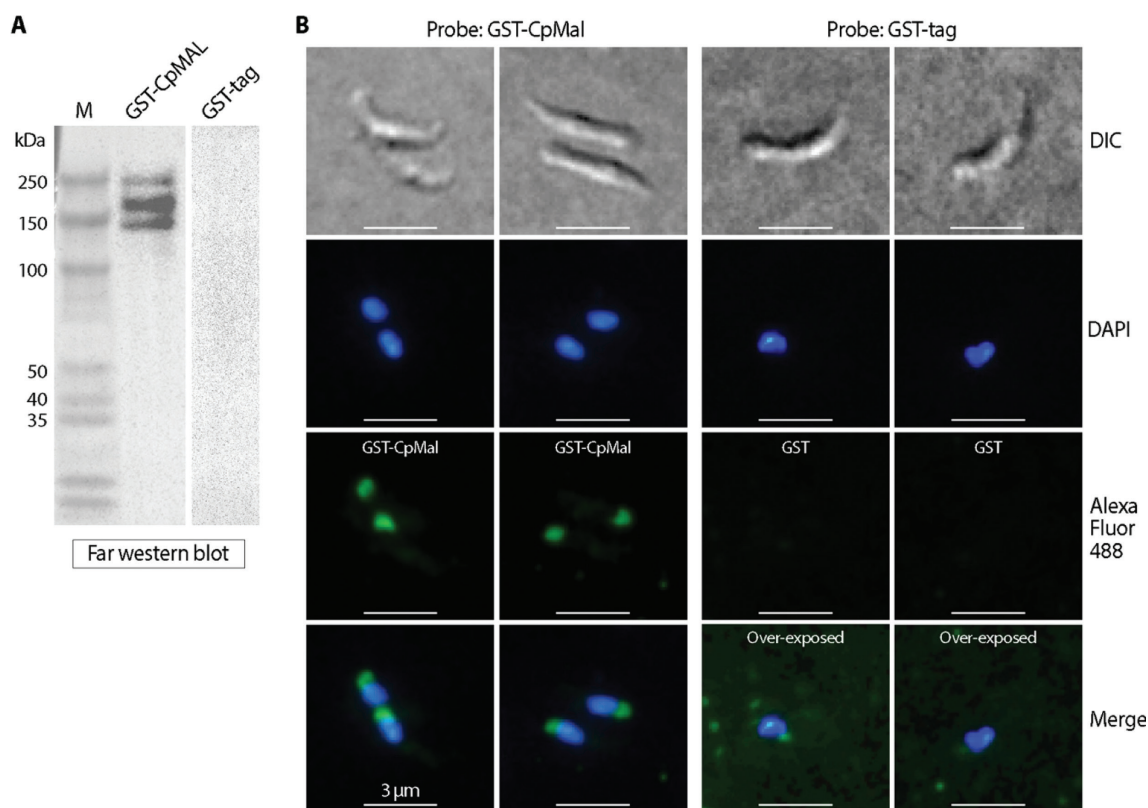


FIGURE 6 | Detection of potential binding partners/ligands of CpMal in the *C. parvum* sporozoites. A) Far-western blot detection of CpMal-binding proteins in the sporozoite lysates. The blots were first probed with GST-CpMal or GST-tag and subsequently detected by western blotting with anti-GST antibodies. B) Immunostaining of CpMal-binding proteins in excysted sporozoites. Fixed and permeabilized sporozoites were first probed with GST-CpMal or GST-tag and subsequently detected with regular immunofluorescence assay procedures with anti-GST antibodies. DIC, differential interference microscopy; DAPI, 4',6-diamidino-2-phenylindole for counterstaining nuclei; Alexa Fluor 488, secondary antibody conjugated with Alexa Fluor 488 for detecting anti-GST antibody. In the GST-tag control, merged images were over-exposed to show background signals.

TABLE 1 | List of proteins identified by proteomic analysis from the three bands recognized by CpMal in far-western blot analysis*.

No.	Gene ID	Description	MW	Scores	Matches	Sequences	emPAI	Coverage
1	cgd6_4460	Uncharacterized protein with Armadillo-like helical	279,469	36	1(1)	1(1)	0.01	0%
2	cgd4_650	Bromodomain/Zinc finger, CCHC-type	218,743	23	1(1)	1(1)	0.01	0%
3	cgd3_3430	Amine oxidase	200,926	131	13(7)	8(7)	0.12	4%
4	cgd1_2300	Uncharacterized transmembrane Protein	184,812	15	2(1)	1(1)	0.02	0%
5	cgd4_3530	Uncharacterized protein	176,189	106	4(3)	4(3)	0.06	3%
6	cgd2_3250	Tetratricopeptide repeat	112,410	20	1(1)	1(1)	0.03	0%
7	cgd2_3330	Hsp70 protein	103,487	53	2(1)	2(1)	0.03	2%
8	cgd3_3440	Heat shock protein HSP70	74,825	25	1(1)	1(1)	0.04	1%
9	cgd7_1320	Casein kinase II, alpha subunit, putative	60,381	34	2(2)	1(1)	0.11	1%
10	cgd6_3990	Elongation factor 1-alpha	48,416	103	3(3)	1(1)	0.22	3%
11	cgd4_3240	Unspecified product	45,861	25	4(1)	1(1)	0.07	1%
12	cgd2_550	Mannose-P-dolichol utilization defect 1-like protein with PQ-loop repeat	26,624	17	3(1)	1(1)	0.13	3%
13	cgd6_4153	EF-hand domain containing protein	15,819	16	13(3)	1(1)	0.21	5%
14	cgd5_940	Histone H2A	15,730	52	2(2)	2(2)	0.48	10%
15	cgd8_5230	Histone H4	11,303	51	2(2)	1(1)	0.70	11%

*Bold format indicates the two proteins' prioritized candidate ligands of CpMal. MW = molecular weight. Scores = ion score for the match of the observed MS/MS spectrum to the indicated peptides. Matches = spectral queries matched. Sequences = unique peptides matched. Numbers in parenthesis show queries with E-values <0.05. emPAI = exponentially modified protein abundance index. Coverage = percentage coverage of peptides in the indicated protein.

cgd3_3430) or “extracellular protein with a signal peptide sequence, MAM domain and a Cu amine oxidase domain” in GenBank (XP_626894) and 2) a 1,578 aa protein annotated as “uncharacterized protein” in CryptoDB (cgd4_3530) or “hypothetical protein” in GenBank (XP_625929).

In far-IFA assays with GST-CpMal as the probe, followed by IFA procedures to detect the GST-tag, CpMal specifically labeled the posterior region behind the nuclei of the sporozoites (Fig 6B). GST-tag as the probe produced no signals, thus confirming that the labeling of CpMal in the sporozoites was specific. Unexpectedly, the distribution of putative ligands for CpMal (i.e., in the posterior end of the sporozoites) was entirely different from that of native CpMal (i.e., in the sporozoite cytosol with two concentrated spots near the nuclei). In theory, GST-CpMal would recognize and label Glc₂-*N*-glycans present in the ER. A plausible explanation might be that Glc₂-*N*-glycans in the parasite were already masked by native CpMal and thus could not be accessed in GST-CpMal-binding assays, unless Glc₂-*N*-glycan was actually not a ligand for CpMal. For the same reason, we speculated that the observed binding of GST-CpMal to ligands was mediated by direct protein-protein interactions rather than interaction with the Glc₂-*N*-glycan moiety of the ligands.

DISCUSSION

Protein *N*- and *O*-glycosylation are post-translational modifications found in all three domains of life (i.e., Eukarya, Bacteria and Archaea), and glycosylated proteins have diverse biological roles [34,35]. In cryptosporidia, both *N*- and *O*-glycosylations are present in several proteins, such as the mucin-like GP900 [31,36]. Our datamining of the genomes and the structural clarification of polysaccharides released from cryptosporidial proteins, determined by other investigators, suggest that *Cryptosporidium* parasites differ from animals and other apicomplexans in *N*-glycosylation (Fig 2) [31]. However, the biological process of *N*-glycosylation in *Cryptosporidium* remains poorly understood. In studying *Cryptosporidium* biology, an apparent obstacle is the lack of availability of marker reagents. In fact, well characterized markers for the ER—a common organelle in eukaryotes—remain lacking. The morphology and function of the ER in *Cryptosporidium* are also poorly studied. Because malectins are known to participate in early processing of *N*-glycan in the ER, and because malectin orthologs were present in *Cryptosporidium* only within the Apicomplexa, we decided to characterize the unique CpMal to study the *N*-glycosylation and potentially develop an ER marker in the parasite.

Because of the unavailability and current technical difficulties in synthesizing malectin's endogenous substrates (i.e., the putative Glc₂Man₅GlcNAc₂-*N*-glycan in *Cryptosporidium* and the known Glc₂Man₉GlcNAc₂-*N*-glycan in mammalian hosts), we were unable to fully characterize the biochemical features of CpMal in comparison to HsMal. However, our current data are sufficient to show that CpMal differs substantially from HsMal at

both the sequence and biochemical levels. The substantial differences between CpMal and HsMal also allowed us to hypothesize that selective inhibitors of CpMal might be developed to interfere with the essential *N*-glycosylation in *Cryptosporidium*, thus killing the parasite.

The distribution pattern of native CpMal in sporozoites is consistent with that of an ER network in cells, i.e., present in most regions of cells but more concentrated near the nuclei (Fig 4). However, whether CpMal might serve as a standard ER marker must be further validated by immunoelectron microscopy and the development of additional ER markers, such as ER membrane-anchored enzymes involved in synthesizing the glycan precursor and processing of the signal peptide.

Another open question is the identity of potential CpMal-binding partners observed by far-western blotting and far-IFA (Fig 6). Proteomic analysis identified two *C. parvum* proteins from the areas corresponding to the three bands recognized by CpMal in the far-western blot, which could be considered potential candidate binding partners for further investigation. Both proteins contain an N-terminal signal peptide but lack any transmembrane domains, thus suggesting that they are secretory proteins. The first protein (cgd3_3430) contains a meprin, A-5 protein, and receptor protein-tyrosine phosphatase Mu (MAM) domain close to the N-terminus at amino acid positions 283 to 476 (InterPro domain IPR000998), and a copper amine oxidase domain close to the C-terminus at positions 1,251 to 1,743 (InterPro family IPR000269). The MAM domain is present in several cell surface proteins and is likely to have an adhesive function [37], whereas copper amine oxidase catalyzes the oxidation of primary amines to aldehydes with the release of ammonia and hydrogen peroxide [38]. The other (cgd4_3530) is highly enigmatic and lacks homologs to any known domains despite its massive size and its high redundancy in oocysts and sporozoites (as indicated by the current available proteomic data at the CryptoDB). Although the biological roles in the parasite remain unknown and must be elucidated, validating whether one (or both) of them is truly a binding partner/ligand for CpMal should prove interesting. The same proteomic analysis will also be repeated to produce more reliable and comparable data for identifying potential candidate binding ligands in the parasite.

CONCLUSIONS

We characterized the primary molecular and biochemical features of a malectin from the zoonotic apicomplexan *C. parvum* (CpMal). Within the phylum Apicomplexa, *Cryptosporidium* is the only lineage possessing a malectin that shares low sequence identity with orthologs from animals. CpMal is distributed in a diffuse pattern in the sporozoites but is highly concentrated in two areas on the anterior and posterior sides near the nuclei, thus implying higher *N*-glycan processing activity in the ER near the nuclei. Native CpMal is likely to be present in the parasite cells in stable dimeric form. CpMal also differs from

HsMal in its binding activity to amylose and to the surfaces of HCT-8 cells. Additionally, we confirmed the presence of binding partners of CpMal by far-western blot analysis and immunostaining-based assays. This study provides a basis for future investigation of the biological role of the unique *Cryptosporidium* malectin.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Key Research and Development Program of China (2017YFC1601200). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

REFERENCES

- Templeton TJ, Enomoto S, Chen WJ, Huang CG, Lancto CA, Abrahamsen MS, et al. A genome-sequence survey for *Ascogregarina taiwanensis* supports evolutionary affiliation but metabolic diversity between a Gregarine and *Cryptosporidium*. *Mol Biol Evol*. 2010;27(2):235-248.
- Zhu G, Keithly JS, Philippe H. What is the phylogenetic position of *Cryptosporidium*? *Int J Syst Evol Microbiol*. 2000;50(Pt 4):1673-1681.
- Zhu G, Yin J, Cuny GD. Current status and challenges in drug discovery against the globally important zoonotic cryptosporidiosis. *Animal Diseases*. 2021;1(1):3.
- Tzipori S, Widmer G. The biology of *Cryptosporidium*. *Contrib Microbiol*. 2000;6:1-32.
- Rider Jr SD, Zhu G. *Cryptosporidium*: genomic and biochemical features. *Exp Parasitol*. 2010;124(1):2-9.
- Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, et al. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*. 2004;304(5669):441-445.
- Schallus T, Jaeckh C, Feher K, Palma AS, Liu Y, Simpson JC, et al. Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein *N*-glycosylation. *Mol Biol Cell*. 2008;19(8):3404-3414.
- Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*. 2012;13(7):448-462.
- Wiederschain GY. Essentials of glycobiology. *Biochemistry*. 2009;74(9):1056-1056.
- Galli C, Bernasconi R, Solda T, Calanca V, Molinari M. Malectin participates in a backup glycoprotein quality control pathway in the mammalian ER. *PLoS One*. 2011;6(1):e16304.
- Tannous A, Pisoni GB, Hebert DN, Molinari M. *N*-linked sugar-regulated protein folding and quality control in the ER. *Semin Cell Dev Biol*. 2015;41:79-89.
- Takeda K, Qin SY, Matsumoto N, Yamamoto K. Association of malectin with ribophorin I is crucial for attenuation of misfolded glycoprotein secretion. *Biochem Biophys Res Commun*. 2014;454(3):436-440.
- Chen Y, Hu D, Yabe R, Tateno H, Qin SY, Matsumoto N, et al. Role of malectin in Glc(2)Man(9)GlcNAc(2)-dependent quality control of alpha1-antitrypsin. *Mol Biol Cell*. 2011;22(19):3559-3570.
- Takeda Y, Totani K, Matsuo I, Ito Y. Chemical approaches toward understanding glycan-mediated protein quality control. *Curr Opin Chem Biol*. 2009;13(5-6):582-591.
- Qin SY, Hu D, Matsumoto K, Takeda K, Matsumoto N, Yamaguchi Y, et al. Malectin forms a complex with ribophorin I for enhanced association with misfolded glycoproteins. *J Biol Chem*. 2012;287(45):38080-38089.
- Yang H, Wang D, Guo L, Pan H, Yvon R, Garman S, et al. Malectin/Malectin-like domain-containing proteins: a repertoire of cell surface molecules with broad functional potential. *Cell Surf*. 2021;7:100056.
- Truong Q, Ferrari BC. Quantitative and qualitative comparisons of *Cryptosporidium* faecal purification procedures for the isolation of oocysts suitable for proteomic analysis. *Int J Parasitol*. 2006;36(7):811-819.
- Zhang H, Zhu G. Quantitative RT-PCR assay for high-throughput screening (HTS) of drugs against the growth of *Cryptosporidium parvum* in vitro. *Front Microbiol*. 2015;6:991.
- Wang C, Wang D, Nie J, Gao X, Yin J, Zhu G. Unique tubulin-based structures in the zoonotic apicomplexan parasite *Cryptosporidium parvum*. *Microorganisms*. 2021;9(9):1921.
- Guo F, Zhang H, Payne HR, Zhu G. Differential gene expression and protein localization of *Cryptosporidium parvum* fatty Acyl-CoA synthetase isoforms. *J Eukaryot Microbiol*. 2016;63(2):233-246.
- Zhang H, Guo F, Zhou H, Zhu G. Transcriptome analysis reveals unique metabolic features in the *Cryptosporidium parvum* Oocysts associated with environmental survival and stresses. *BMC Genomics*. 2012;13:647.
- Zhang H, Guo F, Zhu G. *Cryptosporidium* lactate dehydrogenase is associated with the parasitophorous vacuole membrane and is a potential target for developing therapeutics. *PLoS Pathog*. 2015;11(11):e1005250.
- Mauzy MJ, Enomoto S, Lancto CA, Abrahamsen MS, Rutherford MS. The *Cryptosporidium parvum* transcriptome during in vitro development. *PLoS One*. 2012;7(3):e31715.
- Yu X, Guo F, Mouneimne RB, Zhu G. *Cryptosporidium parvum* elongation factor 1alpha participates in the formation of base structure at the infection site during invasion. *J Infect Dis*. 2020;221(11):1816-1825.
- Lateef SS, Gupta S, Jayathilaka LP, Krishnanchettiar S, Huang JS, Lee BS. An improved protocol for coupling synthetic peptides to carrier proteins for antibody production using DMF to solubilize peptides. *J Biomol Tech*. 2007;18(3):173-176.
- Greenfield EA. Standard immunization of rabbits. *Cold Spring Harb Protoc*. 2020;2020(9):100305.
- Kurien BT. Affinity purification of autoantibodies from an antigen strip excised from a nitrocellulose protein blot. *Methods Mol Biol*. 2009;536:201-211.
- Zhang T, Gao X, Wang D, Zhao J, Zhang N, Li Q, et al. A single-pass type I membrane protein from the apicomplexan parasite *Cryptosporidium parvum* with nanomolar binding affinity to host cell surface. *Microorganisms*. 2021;9(5):1015.
- Ludington JG, Ward HD. The *Cryptosporidium parvum* C-type lectin CpClec mediates infection of intestinal epithelial cells via interactions with sulfated proteoglycans. *Infect Immun*. 2016;84(5):1593-1602.
- Wu Y, Li Q, Chen XZ. Detecting protein-protein interactions by Far western blotting. *Nat Protoc*. 2007;2(12):3278-3284.
- Haserick JR, Leon DR, Samuelson J, Costello CE. 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. *Mol Cell Proteomics*. 2017;16(4 suppl 1):S42-S53.
- Fauquenoy S, Hovasse A, Sloves PJ, Morelle W, Dilezitoko Alayi T, Slomianny C, et al. Unusual N-glycan structures required for trafficking *Toxoplasma gondii* GAP50 to the inner membrane complex regulate host cell entry through parasite motility. *Mol Cell Proteomics*. 2011;10(9):M111008953.
- Zeng B, Cai X, Zhu G. Functional characterization of a fatty acyl-CoA-binding protein (ACBP) from the apicomplexan *Cryptosporidium parvum*. *Microbiology*. 2006;152(Pt 8):2355-2363.

-
34. Abu-Qarn M, Eichler J, Sharon N. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr Opin Struct Biol.* 2008;18(5):544-550.
 35. Schaffer C, Messner P. Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev.* 2017;41(1):49-91.
 36. Haserick JR, Klein JA, Costello CE, Samuelson J. *Cryptosporidium parvum* vaccine candidates are incompletely modified with O-linked-N-acetylgalactosamine or contain N-terminal N-myristate and S-palmitate. *PLoS One.* 2017;12(8):e0182395.
 37. Beckmann G, Bork P. An adhesive domain detected in functionally diverse receptors. *Trends Biochem Sci.* 1993;18(2):40-41.
 38. Parsons MR, Convery MA, Wilmot CM, Yadav KD, Blakeley V, Corner AS, et al. Crystal structure of a quinoenzyme: copper amine oxidase of *Escherichia coli* at 2 Å resolution. *Structure.* 1995;3(11):1171-1184.