Identification and Characterization of Cryptosporidium parvum Clec, a Novel C-Type Lectin Domain-Containing Mucin-Like Glycoprotein

Seema Bhalchandra, Jacob Ludington, Isabelle Coppens, Honorine D. Ward

Cryptosporidium species are waterborne apicomplexan parasites that cause diarrheal disease worldwide. Although the mechanisms underlying Cryptosporidium-host cell interactions are not well understood, mucin-like glycoproteins of the parasite are known to mediate attachment and invasion in vitro. We identified C. parvum Clec (CpClec), a novel mucin-like glycoprotein that contains a C-type lectin domain (CTLD) and has orthologs in C. hominis and C. muris. CTLD-containing proteins are ligand-binding proteins that function in adhesion and signaling and are present in a wide range of organisms, from humans to viruses. However, this is the first report of a CTLD-containing protein in protozoa and in Apicomplexa. CpClec is predicted to be a type 1 membrane protein, with a CTLD, an O-glycosylated mucin-like domain, a transmembrane domain, and a cytoplasmic tail containing a YXXH sorting motif. The predicted structure of CpClec displays several characteristics of canonical CTLD-containing proteins, including a long loop region hydrophobic core associated with calcium-dependent glycan binding as well as predicted calcium- and glycan-binding sites. CpClec expression during C. parvum infection in vitro is maximal at 48 h postinfection, suggesting that it is developmentally regulated. The 120-kDa mass of native CpClec is greater than predicted, most likely due to O-glycosylation. CpClec is localized to the surface of the apical region and to dense granules of sporozoites and merozoites. Taken together, these findings, along with the known functions of C. parvum mucin-like glycoproteins and of CTLD-containing proteins, strongly implicate a significant role for CpClec in Cryptosporidium-host cell interactions.

Cryptosporidium species are protozoan parasites of the phylum Apicomplexa that cause diarrheal disease worldwide (1–4). Infection with these parasites is often asymptomatic or associated with a self-limiting diarrheal illness in immunocompetent hosts. However, cryptosporidiosis can be particularly devastating in immunocompromised individuals, such as untreated AIDS patients and malnourished children in resource-constrained countries (2, 5). Despite the worldwide impact of Cryptosporidium species (4), treatment options for cryptosporidiosis, particularly for the immunocompromised, are limited (6). Nitazoxanide, the only drug approved for cryptosporidiosis by the U.S. Food and Drug Administration, is ineffective in AIDS patients (7), and there is no vaccine available. Thus, development of effective therapeutic and preventive strategies for cryptosporidiosis in vulnerable populations is urgently needed.

The molecular mechanisms by which C. parvum infects host cells and the parasite proteins involved are not completely known. Studies to investigate these mechanisms and the proteins that mediate them have been impeded by the inability to propagate the parasite in cell culture and to genetically manipulate it (8, 9). Therefore, a number of studies have focused on indirect approaches to identify and characterize parasite proteins important in initiation of infection. These studies have revealed that C. parvum employs mucin-like glycoproteins for attachment to and invasion of host cells (reviewed in references 8 and 9). This has prompted a search for other mucin-like glycoproteins that may be important in pathogenesis. While mining the C. parvum genome for additional mucin-like glycoproteins, we identified a gene that encodes a novel mucin-like glycoprotein with a C-type lectin domain (CTLD).

C-type lectins (CTLs) are calcium-dependent, glycan-binding proteins that function as adhesion or signaling receptors in the immune system (10). The C-type lectin domain or fold is a structural motif characterized by a two-loop structure stabilized by 2 highly conserved disulfide bonds with at least 4 conserved cysteine residues, 2 to 3 hydrophobic cores, and up to 4 sites for binding calcium (10–12). Although CTLDs were originally identified as calcium-dependent, carbohydrate recognition domains, many CTLDs do not bind carbohydrates in a calcium-dependent manner. These are referred to as C-type lectin-like domains and represent a ligand-binding motif that may bind carbohydrate, lipid, protein, or inorganic ligands (11).

CTLD-containing proteins are present in all metazoans and most nonmetazoans (10). However, CTLD-containing proteins have not previously been reported in protozoans. Here, we report the identification and characterization of the C. parvum Clec (CpClec), a novel CTLD-containing, mucin-like glycoprotein with orthologs in C. hominis and C. muris. In addition, we report identification of putative CTLD-containing proteins in three other apicomplexan parasites. Understanding the structure and function of CpClec as well as other apicomplexan CTLD-contain-
ing proteins may reveal additional molecular mechanisms by which these parasites interact with host cells.

MATERIALS AND METHODS

Parasites. C. parvum oocysts (Iowa isolate) were purchased from Bunch Grass Farms, ID, or the University of Arizona Sterling Parasitology Lab, Tucson, AZ, and stored in 150 mM sodium chloride, 20 mM sodium phosphate buffer, pH 7.2 (phosphate-buffered saline [PBS]) at 4°C. Prior to use, oocysts were surface sterilized by suspension in 10% (vol/vol) commercial bleach solution (sodium hypochlorite) for 10 min on ice and then washed three times with PBS at 10,000 × g for 2 min at 4°C.

Hypochlorite-treated C. parvum oocysts were resuspended in PBS containing a protease inhibitor cocktail (104 mM 4-(2-aminoethyl) benzenesulfonfonyl Fluoride hydrochloride, 80 μM aprotinin, 4 mM bestatin, 1.4 mM E-64, 2 mM leupeptin, 1.5 mM pepstatin A Sigma-Aldrich, St. Louis, MO) and exsyrated at 37°C for 1 h. After washing with PBS, the resultant mixture of sporozoites and unexcysted oocysts was centrifuged at 4,500 × g for 15 min, and the pellet was resuspended in 20 mM Tris buffer (pH 7.2), 135 mM sodium chloride, 10 mM manganese chloride containing 1% Triton X-100 and protease inhibitor cocktail and then incubated on ice for 2 h. The clarified supernatant was collected by centrifugation at 10,000 × g for 25 min at 4°C and was designated the C. parvum lysate.

Hypochlorite-treated oocysts were excysted in 0.75% sodium taurocholate in PBS for 1 h at 37°C, and released sporozoites were separated from unexcysted oocysts and oocyst shells by filtration through a 3-μm filter (Millipore, Billerica, MA).

Toxoplasma gondii RH strain parasites were obtained from Kami Kim, Albert Einstein College of Medicine, and maintained in human foreskin fibroblast (HFF) cells.

Cell culture. The human intestinal epithelial (ileocecal adenocarcinoma) cell line Caco-2 and human foreskin fibroblasts cells were grown in Dulbecco’s modified Eagle medium (DMEM; Life Technologies, Grand Island, NY) containing 15% fetal calf serum, 25 mM HEPES, 100 U of penicillin, and 100 μg of streptomycin per ml at 37°C in 5% CO2.

Identification and in silico analysis of Cryptosporidium species and other apicomplexan Clec. The Clec sequence (Cgd3_440) was identified from the C. hominis genome sequence in the Cryptodb database (http://www.cryptodb.org) (13) by using a text search with the word mucin. The C. hominis and C. muris Clec sequences were identified from their genome sequences by BLAST comparison to the C. hominis and C. muris Clec sequences. Both the C. parvum and C. hominis Clec sequences were incomplete at the 5’ end and were extended in silico from the respective contigs in Cryptodb, using the complete C. muris sequence as a template. The genomes of other apicomplexan parasites, including Plasmmodium, Toxoplasma, Neospora, Babesia, and Theileria species, were mined for putative CTLD-containing proteins by searching their respective genomes within the Eupathdb database (http://eupathdb.org) (14) with the “identify genes by similarity/pattern Interpro domain” option for sequences containing a C-type lectin fold. The presence of the CTLD in each sequence was confirmed by using the ExpASy Prosite server (http://prosite.expasy.org/sanprosite). Sequences were aligned using MacVector v. 12 (MacVector, Inc., Cary, NC), and analysis was performed using the ExpASy Molecular Biology Server (http://us.expasy.org). Phylogenetic analysis of representative CTLD-containing proteins was performed using MEGA 5 (15) and the maximum likelihood maximum method, based on a Poisson distribution model. A bootstrap consensus tree was created from 1,000 replicates.

Predictions of tertiary structure and ligand-binding sites. The tertiary structure of Cplec was predicted using the Protein Homology/AnalogY Recognition Engine Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/), which uses homology of the input sequences with known structures to generate the output model (16). The 3DLigandsite server (http://www.sbg.bio.ic.ac.uk/3dligandsite) (17) was used to identify predicted ligand-binding sites in the CTLD of CpClec. A threshold of 95% confidence was used to evaluate the output models.

Time course of expression of Cplec RNA in infected Caco-2 cells. Confluent Caco-2 cell monolayers were infected with C. parvum oocysts, and RNA was isolated from infected and uninfected Caco-2 cells at 0, 6, 12, 24, 48, and 72 h postinfection by using an RNeasy kit (Qiagen, Valencia, CA). cDNA was synthesized using a QuantiTect reverse transcription kit (Qiagen). cDNA from uninfected Caco-2 cells was used as a negative control, and reactions without reverse transcriptase were performed in parallel to control for amplification of contaminating genomic DNA. Amplification of a portion of the C. parvum 18S rRNA gene was used as an internal control for normalization of Cplec gene expression. Amplification of Cplec and C. parvum 18S RNA was visualized following conventional PCR and 1.5% agarose gel electrophoresis, with quantification by real-time PCR following MIQE guidelines (18). SYBR green supermix, (Qiagen) and specific primers (see Table S1 in the supplemental material) were used, and reactions were run on a Stratagene MX3000P thermocycler, with cycling conditions of 95°C for 15 min and 40 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 30 s. All primers were designed using Primer 3 (http://frodo.wi.mit.edu/). Gene expression in infected Caco-2 cells was normalized to that in uninfected cells by using the 2−ΔΔCt method, where the amount of target, normalized to the endogenous reference 18S rRNA gene and relative to expression in uninfected cells, is given by 2−ΔΔCt, and Ct is the cycle number of the detection threshold (19). Assays were performed in duplicate, and the experiment was performed twice. Statistical comparisons were made using the Kruskal-Wallis test and Prism software v 6.0 (GraphPad Inc., La Jolla, CA).

Cloning, expression, and purification of recombinant CpClec. Caco-2 cells were infected with C. parvum oocysts for 24 h as described above, and total RNA was extracted with an RNeasy kit (Qiagen). Contaminating genomic DNA was removed using Turbo DNase (Ambion, Austin, TX) treatment. The full-length CpClec transcript was reverse transcribed and amplified by PCR using SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA), Platinum Pf DNA polymerase (Invitrogen), and primers specific for the putative coding sequence (see Table S1 in the supplemental material) under the following conditions: 95°C for 5 min, followed by 35 cycles at 94°C for 1 min, 60°C for 30 s, and 72°C for 2 min, with a final 72°C for 10 min. The PCR product was purified using a Qiagen Quick PCR purification kit (Qiagen) and cloned into the TOPO pcRII vector (Invitrogen), and plasmids were isolated by using a Qiaprep miniprep kit (Qiagen). DNA sequencing was performed by the dye terminator method, on a Perkin-Elmer ABI 377 sequencer at the Tufts University Core Facility.

A 429-bp cDNA fragment encoding CpClec amino acids 633 to 775 was PCR amplified from the CpClecFL cDNA clone with specific primers (see Table S1), cloned into the pGEX4T-1 expression vector (GE Healthcare, Piscataway, NJ), and overexpressed as a glutathione S-transferase (GST) fusion protein in Escherichia coli BL21(DE3)pLYsS cells following induction with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The recombinant GST fusion protein, rCpClec633–775, was purified by affinity chromatography using GST-Sepharose (Qiagen).

Cloning of T. gondii RH strain Clec. RH strain tachyzoites were harvested from infected HFF cells, and DNA was extracted using a DNeasy blood and tissue kit (Qiagen). The T. gondii RH strain Clec (TgClec) sequence was amplified from genomic DNA with specific primers (see Table S1 in the supplemental material) under the following cycling conditions: 95°C for 10 min, 35 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 30 s, and 72°C for 10 min. The amplified product was purified using a Qiagen kit (Qiagen) and cloned into the TOPO pcRII vector (Invitrogen), and plasmids were isolated by using a Qiaprep miniprep kit (Qiagen). The resulting plasmids were sequenced at the Tufts University Core Facility.

Antibodies. Purified rCpClec was run on a preparative SDS-PAGE gel and stained with 10% Coomassie blue R-250 in water. The band for the expected size (42 kDa) was excised from the gel, lyophilized, resuspended in PBS, and emulsified with complete Freund’s adjuvant. Six-week-old BALB/c mice were immunized with the antigen by intraperitoneal injec-

Downloaded from iai.asm.org on December 22, 2017 by guest
SDS-PAGE, transferred to nitrocellulose, and probed with anti-reducing or nonreducing sample buffer for 5 min at 95°C, separated by and underlined. Predicted O/H9278 the mucin-like domain is in light gray, and the transmembrane domain is underlined. The YXX\_consensus sequence shows identities with an asterisk and similarities by a period. The predicted signal sequence is double underlined, the CTLD is in dark gray,
type O-glycosylation on 38 contiguous threonine and 2 serine residues (Fig. 1; Table 1). Six predicted N-glycosylation sites were also identified. In accordance with commonly used nomenclature for CTLD-containing proteins, we named the protein CpClec.

Mining of the *C. hominis* and *C. muris* genomes revealed orthologs in both species, which we named *ChClec* and *CmClec*, respectively (Fig. 1). CpClec and *ChClec* were nearly identical at the nucleotide (97.2%) and deduced amino acid (97.2%) levels, whereas *CmClec* was more divergent (54.3% and 37.7%, respectively) (Table 1). *CpClec* and *ChClec* share identical signal peptides, CTLDs, and transmembrane domains (Fig. 1). The mucin-like domains are identical, except for the number of predicted O-glycosylation sites, which vary between both species. The predicted N-glycosylation sites are identical between the two species. However, the *CmClec* signal peptide, CTLD, mucin-like domain, and transmembrane domain are more divergent, and the protein is predicted to be more extensively O- and N-glycosylated. The cytoplasmic tails of Clec from all three species contain 4 conserved tyrosine residues, one of which is present in a conserved YHD/EF sequence and conforms to the YXXδ sorting motif in metazoans and other apicomplexan parasites (22).

Comparison with representative known CTLDs revealed that the CpClec CTLD (Fig. 2A) displays several highly conserved features critical for the fundamental structure of CTLDs (23). These include six conserved cysteine residues, four of which are predicted to form disulfide bridges that result in the signature double-looped structure characteristic of CTLDs. The second loop is the long loop region hydrophobic core (LLHRC), which is a characteristic feature of canonical CTLDs and is associated with calcium-dependent carbohydrate binding. A highly conserved glutamate residue (E69) that is known to form an ionic and/or hydrogen bond and that has an important role in stabilizing the tertiary protein structure is also conserved in the CpClec CTLD. The characteristic WIGL motif, with an invariant glycine residue that is highly conserved in canonical CTLD sequences and is implicated in stabilizing three hydrophobic core regions, is also present. The EPN, QPD, and WND motifs, which are involved in calcium-dependent carbohydrate binding in classical CTLs, are not present in the CpClec CTLD. However, a glutamate residue (E139) that is one of two amino acids from those that form the calcium-binding site 2 in canonical CTLDs is present in CpClec.

### Table 1: Properties of *Cryptosporidium* CTLD-containing proteins

<table>
<thead>
<tr>
<th>Property</th>
<th><em>C. parvum</em></th>
<th><em>C. hominis</em></th>
<th><em>C. muris</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene ID</strong></td>
<td>cgD3_440</td>
<td>cho.30061</td>
<td>cmu_038310</td>
</tr>
<tr>
<td><strong>GenBank accession no.</strong></td>
<td>JX417972</td>
<td>JX417971</td>
<td>EEA047651.1</td>
</tr>
<tr>
<td><strong>No. of amino acids</strong></td>
<td>775</td>
<td>778</td>
<td>686</td>
</tr>
<tr>
<td><strong>Mass (kDa)</strong></td>
<td>86.4</td>
<td>86.6</td>
<td>76.1</td>
</tr>
<tr>
<td><strong>Isoelectric point</strong></td>
<td>5.87</td>
<td>5.96</td>
<td>4.99</td>
</tr>
</tbody>
</table>

Amino acid content (%)

<table>
<thead>
<tr>
<th>Amino acid</th>
<th><em>C. parvum</em></th>
<th><em>C. hominis</em></th>
<th><em>C. muris</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Proline</td>
<td>4</td>
<td>4</td>
<td>3.8</td>
</tr>
<tr>
<td>Serine</td>
<td>11.5</td>
<td>11.2</td>
<td>10.8</td>
</tr>
<tr>
<td>Thrreonine</td>
<td>10.7</td>
<td>11.2</td>
<td>12.8</td>
</tr>
</tbody>
</table>

O-glycosylation sites

<table>
<thead>
<tr>
<th>O-glycosylation sites</th>
<th><em>C. parvum</em></th>
<th><em>C. hominis</em></th>
<th><em>C. muris</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>45</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>

N-glycosylation sites

<table>
<thead>
<tr>
<th>N-glycosylation sites</th>
<th><em>C. parvum</em></th>
<th><em>C. hominis</em></th>
<th><em>C. muris</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>5</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

% nucleotide identity

<table>
<thead>
<tr>
<th>% nucleotide identity</th>
<th><em>C. parvum</em></th>
<th><em>C. hominis</em></th>
<th><em>C. muris</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>97.2</td>
<td>54.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

% amino acid identity

<table>
<thead>
<tr>
<th>% amino acid identity</th>
<th><em>C. parvum</em></th>
<th><em>C. hominis</em></th>
<th><em>C. muris</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>97.7</td>
<td>37.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No. of ESTs

<table>
<thead>
<tr>
<th>No. of ESTs</th>
<th><em>C. parvum</em></th>
<th><em>C. hominis</em></th>
<th><em>C. muris</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

![Nucleotide/deduced amino acid sequence compared to that of *C. parvum* Clec.](http://iai.asm.org/)

Alignment of the apicomplexan CTLDs showed both highly conserved and variable regions (Fig. 1C). Three of the cysteine residues are conserved. In addition, the WIGL motif is present in the CTLDs of all three *Cryptosporidium* species, as well as in *T. gondii* and *N. caninum*. Other important residues conserved among most of the apicomplexans CTLDs include those involved in formation of the hydrophobic cores, although the complete LLRHC motif is only present in *Cryptosporidium* species. Motifs implicated in calcium-dependent, carbohydrate binding (EPN, WND, and QPD) are not present in any of the apicomplexan CTLDs.

**Phylogenetic analysis of *Cryptosporidium* species CTLDs.** A phylogenetic analysis of *Cryptosporidium* species CTLDs and representative known CTLDs, as well as apicomplexan CTLDs identified in this study, showed that *Cryptosporidium* species CTLDs are most closely related to each other, followed by those of the other apicomplexans, *Caenorhabditis elegans*, and lastly the mammalian CTLDs (see Fig. S1 in the supplemental material). An exception is the CTLD of *E. tenella*, which is most closely related to the CTLD of the *Homo sapiens* macropage mannose receptor. The other exception is the *Schistosoma mansoni* M23B CTLD, which is more closely related to vertebrate than invertebrate CTLDs.

**Predicted tertiary structure and ligands of CpClec CTLD.** The predicted tertiary structure of CpClec most closely resembles that of human CD69 (24) and murine Dectin-1 (25). Given its highest-confidence match (96.5%) to the CpClec CTLD, the crystal structure of CD69 was the best structural template for the construction of a three-dimensional molecular model (Fig. 3A). Although the positions and orientations of the secondary structural elements are inherited from the predicted structure, the model correlates well with data from the primary sequence (Fig. 1). The ribbon model for the predicted structure of CpClec CTLD (Fig. 3B) consists of two antiparallel β-sheets and two α-helices, with the C and N termini coming close together at the bottom of the structure and stabilized by highly conserved disulfide bridges, which is a characteristic feature of the CTLD fold (11). The upper lobe, consisting of the long loop hydrophobic core region and the highly conserved glutamate residue in the predicted calcium-binding site, was also identified (Fig. 3B). Although the characteristic carbohydrate-binding EPN, WND, and QPD motifs are not present in the CpClec CTLD, binding sites for mannose, galactose,
and N-acetylglactosamine are predicted in the tertiary structure (Fig. 3B).

**Time course for expression of CpClec.** The presence of 3 expressed sequence tags (ESTs) corresponding to the CpClec cDNA sequence (Table 1) suggested that the CpClec gene is expressed. To confirm this and to determine the transcriptional profile of the CpClec gene during intracellular development, we analyzed expression of this gene by conventional and real-time reverse transcription-PCR over a 72-h time course in C. parvum-infected Caco-2 cells. This showed that expression of the CpClec gene occurred at low levels throughout intracellular development but was maximal at 48 h postinfection, suggesting that this gene is developmentally regulated (Fig. 4).

**Identification of native CpClec.** In order to characterize native CpClec, we generated antibodies in mice to an antigenic region corresponding to amino acids 633 to 775 in the cytoplasmic tail of CpClec expressed as a GST fusion protein in E. coli (Fig. 5A). Under reducing conditions, anti-rCpClec reacted exclusively with an ∼120-kDa band in C. parvum lysates in an immunoblotting assay (Fig. 5B). There was no reactivity with preimmune sera, confirming the specificity of the antibody for CpClec. The ∼120-kDa band indicated a protein significantly larger than the predicted Mr of 86, which was likely due to the presence of O-glycosylation in the mucin-like domain of the native protein. Since CpClec contains several cysteine residues that are predicted to form disulfide bonds, we analyzed the native protein by SDS-PAGE under reducing and nonreducing conditions. Under nonreducing conditions, there was no reactivity with the antibody (data not shown), suggesting that the epitopes it recognizes are exposed only under reducing conditions.

**Localization of CpClec in sporozoites and intracellular stages by immunofluorescence.** We used anti-rCpClec to localize CpClec in C. parvum sporozoites and intracellular stages by immunofluorescence. In sporozoites, anti-rCpClec reacted predominantly with the surface of the apical region (identified as being anterior to the posteriorly located nucleus) and partially colocal-
I. Anti-rgp15 Antibody Localization

At 48 h postinfection of Caco-2 cells, we observed a similar pattern of localization at the apical region of merozoites within meronts and colocalization with anti-rgp15, indicating surface localization (Fig. 6).

II. Ultrastructural Localization of CpClec

By immunoelectron microscopy, immunogold staining analyses with anti-rCpClec antibody at the ultrastructural level confirmed the localization of CpClec on the plasma membrane of merozoites, especially at the apical pole, which contains micronemes (Fig. 7A and B). Within the parasites, CpClec was associated with dense cytoplasmic material and the plasma membrane.

III. Expression Analysis of CpClec

Quantitative reverse transcription-PCR was performed on total RNA extracted from C. parvum-infected Caco-2 cells at various times postinfection (pi), from 0.5 to 72 h, using primers specific for CpClec and the Cp18S rRNA gene as a reference control. Conventional PCR was done in parallel, utilizing the same primer sets to amplify specific regions of CpClec (A) or Cp18S rRNA (B). Genomic DNA (gDNA) was included as a positive control, and uninfected Caco-2 cells (-) were used as a negative control. *P value of <0.05; **P value of <0.01.

IV. Purification of rCpClec

The rClec fusion protein was purified by GST affinity chromatography and analyzed by 10% SDS-PAGE. Lane 1, starting lysate; lane 2, eluate. The arrow indicates the 50-kDa rGST-Clec fusion protein. Immunoblot analysis of C. parvum lysate probed with preimmune sera (lane 1) or anti-rCpClec (lane 2). The arrow indicates the mass of native CpClec.
granules prior to secretion (Fig. 7A and C). Interestingly, CpClec staining was also observed beyond the parasitophorous vacuolar membrane, on the membranous folds located at the interface of the parasitophorus vacuole (PV) and the host cell (Fig. 7A and D). More selectively, gold particles were visible on a structure resembling the feeder organelle as well as aligned along the desmosome-like band that separates the parasitophorus vacuole from the host cytoplasm (Fig. 7D).

**DISCUSSION**

CTLD-containing proteins are a large group of proteins with diverse functions, the most well known being their role in cell-cell recognition, adhesion, and signaling via calcium-dependent glycan binding (10, 11). Although the CTLD superfamily contains over 1,000 proteins found in a variety of organisms ranging from humans to viruses (10, 11), this is the first report of CTLD-containing proteins in protozoans and in apicomplexans, and Cryptosporidium species in particular. In this study, we identified, characterized, and expressed CpClec, a novel CTLD-containing, mucin-like glycoprotein with orthologs in all three Cryptosporidium species whose genomes have been sequenced.

CTLDs can be structurally divided into two groups: canonical and compact. Canonical CTLDs contain the long loop region hydrophobic core, which is involved in calcium-dependent carbohydrate binding, and the highly conserved WIGL motif, while compact CTLDs lack the long loop region hydrophobic core (11). Our analyses revealed that CpClec has a canonical CTLD, including the signature long loop region hydrophobic core and the characteristic WIGL motif. Although motifs such as EPN, QPD, and WND, which are associated with carbohydrate binding in classical CTLS, were not identified in the CpClec CTLD, a key glutamic acid residue, E69, that is known to be a part of calcium-binding site 2 was conserved, indicating that it may have a novel Ca$^{2+}$-dependent glycan-binding motif. The predicted tertiary structure also includes putative calcium-binding and glycan-binding (mannose, N-acetylgalactosamine, and galactose) sites in the CpClec CTLD. However, it remains to be experimentally determined whether CpClec binds glycans and if so, whether binding is calcium dependent or independent.

CTLD-containing proteins are currently classified into 17 groups (10, 11). The predicted tertiary structure of the CpClec CTLD is most similar to those of CD69, Dectin-1, and other members of the NK cell receptor group V that are nonclassical C-type lectins encoded by the natural killer gene complex (10). Group V is evolutionarily the most dynamic group, and its members, like CpClec, are characterized by the absence of canonical calcium and glycan-binding residues. Human CD69 is an immunoregulatory,
early leukocyte activation molecule (26) that associates with calcium via a single calcium-binding site formed by conserved amino acids in its CTLD, which leads to the formation of high-affinity binding sites for N-acetyl-d-glucosamine and N-acetylgalactosamine (27, 28). Dectin-1 is a myeloid cell surface receptor that binds β(1-3)-β-glucans on the cell walls of fungi in a calcium-independent manner by using a unique carbohydrate recognition domain (reviewed in reference 29). It has been suggested that the CTLDs in NK cell receptors may have evolved to recognize glycans by novel mechanisms (10). Although it is not yet known whether CpClec binds glycans, it is possible that this is the case for the CpClec CTLD as well. The cytoplasmic tail of Clecs from all three Cryptosporidium species contains the evolutionarily conserved YXXΦ sorting motif (22). This is the first report of a sorting motif in Cryptosporidium. In mammalian membrane proteins, this motif is associated with clathrin-mediated endocytosis, sorting to lysosomes, and the basolateral membrane of polarized epithelial cells through interaction with adaptor protein complexes (22, 30). This motif is present in the cytoplasmic tails of other CTLD-containing proteins, including human and murine macrophage galactose lectin, which is present on dendritic cells and macrophages (31). In other apicomplexans, the YXXΦ motif is essential for targeting membrane proteins to secretory organelles of the apical complex. The YXXΦ motifs in the cytoplasmic tails of T. gondii rhoptry-associated protein 2 (ROP-2) (32) and Plasmodium thrombospordin-related anonymous protein (TRAP) are responsible for targeting to rhoptries and micronemes and the cell surface, respectively (33). Our data indicate that CpClec is localized to the apical cell surface, dense granules, and membranous structures at the host interface of invasive and intracellular stages. We identified genes that encode homologs of mammalian adaptor proteins in the C. parvum genome (data not shown), which suggests that CpClec may interact with them, as is the case for mammalian and other apicomplexan membrane proteins. However, it remains to be determined whether the YXXΦ motif in CpClec is involved in targeting the protein to any of these locations.

CpClec is the first known Cryptosporidium protein shown to localize to dense granules. A previous study reported reactivity of a monoclonal antibody with a 110-kDa glycoprotein in C. parvum dense granules (34). However, the identity of this protein is not known. Another study identified five major proteins in the 120- to 180-kDa range in dense granules of C. parvum sporozoites (35). Again, the identity of these proteins is not known, but the 120-kDa mass of one of them is consistent with that of CpClec. In other apicomplexan parasites, dense granule proteins have been reported to participate in modifying the host cell after invasion (36). However, their role in Cryptosporidium has not been verified.

In addition to dense granules and the apical surface membrane of invasive zoites, our study showed localization of CpClec to the feeder organelle and the desmosome-like band that separates the PV from the host cytoplasm. These findings suggest that CpClec is secreted from dense granules and functions at the parasite surface and on membranous structures in which the PV is embedded.

CpClec is expressed at low levels throughout intracellular development but is maximal at 48 h postinfection, suggesting that it is developmentally regulated. Although intracellular development is asynchronous, at 48 h postinfection there is increased morphological complexity marked by rearrangement of epithelial cells by merozoites (37). A recent study of the transcriptome of the intracellular stages of C. parvum reported a spike in expression of 7 Cryptosporidium-specific genes, including 5 genes encoding mucin-like glycoproteins, at 48 h postinfection (37).

By mining the recently sequenced genomes of six other apicomplexans, we found that T. gondii, E. tenella, and N. caninum contained genes encoding CTLDcys. Interestingly, all three of these as well as Cryptosporidium are oocyst-forming apicomplexans that primarily infect the gastrointestinal tract (38). This raises the possibility that the CTLDs of these parasites may play a role in interaction with intestinal epithelial cells. Interestingly, CTLD-containing proteins are not present in the genomes of Plasmodium, Babesia, or Theileria species, all of which are vector-borne apicomplexan hemoparasites (39).

Although CTLDs are present in T. gondii, E. tenella, and N. caninum, the proteins containing them are very different from the Cryptosporidium Clecs. TgClec is predicted to be a 20-kDa protein that lacks a signal peptide and transmembrane domain and whose CTLD is missing a large portion of the long loop region hydrophobic core. NcClec is predicted to be a 110-kDa protein that contains a signal peptide but lacks a transmembrane domain, and ErClec is predicted to be a much larger (713-kDa) cysteine repeat modular protein that contains a tyrosine kinase growth factor receptor domain. None of them contains mucin-like domains. Thus, the Cryptosporidium species Clecs appear to be unique among apicomplexans, with a canonical CTLD that is more similar to that of higher eukaryotes.

CpClec is also predicted to be a mucin-like glycoprotein. Mucin-like glycoproteins are secreted or surface-associated glycoproteins that contain serine- and/or threonine-rich domains with extensive mucin-type O-glycosylation (40). Cryptosporidium mucin-like glycoproteins that mediate attachment and/or invasion include gp900 (41), gp40 (42, 43), and Muc4 (44), all of which localize to the surface and/or apical region of sporozoites and bind to intestinal epithelial cells. gp900 and gp40 contain heavily O-glycosylated polythreonine and polyserine mucin-like domains, respectively, which display N-acetylgalactosamine α-linked to serine or threonine residues (42). Antibodies to gp900 and gp40 as well as antibodies or lectins specific for α-N-acetylgalactosamine on these proteins block attachment and invasion in vitro, indicating the importance of these glycoproteins and their O-glycans in the initiation of infection (41–43). Like gp900, CpClec contains a polythreonine domain that is predicted to display extensive mucin-type O-glycosylation. However, the presence and type of O-glycosylation in this domain needs to be experimentally confirmed.

Viral, bacterial, and helminthic pathogens are also known to contain CTLDs that function in interactions with host cells. The Epstein-Barr virus (EBV) surface glycoprotein gp42 contains a CTLD that binds major histocompatibility complex class II molecules in a calcium- and carbohydrate-independent manner and, like CpClec, has a structure similar to NK cell receptor-like CTLDs (45). Invasive from Yersinia pseudotuberculosis contains a CTLD that mediates calcium- and glycan-independent adhesion to integrins on host cells (46). Intimin, an adhesin from enterohemorrhagic and enteropathogenic Escherichia coli, also contains a CTLD that functions in (calcium- and glycan-independent) binding to the translocated intimin receptor (Tir) on host cells (47). A number of helminths are known to express CTLD-containing proteins (48). For example, Te-CTL-4, a C-type lectin, from Toxocara canis, binds glycans on host endothelial cells in a calcium-
dependent manner (49). In addition to their role in adhesion, CTLDs of pathogens, particularly helminths, have been postulated to play a role in immune evasion because of their structural similarity to host CTLDs (48).

Possible roles for CpleC in attachment to and subsequent host cell invasion as well as in immune protection or evasion remain to be experimentally determined. Since Cryptosporidium cannot be genetically manipulated, indirect approaches are necessary to confirm the function of CpleC. However, taken together, the predicted structural features and localization of CpleC along with the known functions of C. parvum mucin-like glycoproteins and of CTLD-containing proteins strongly implicate CpleC as playing a significant role in Cryptosporidium-host cell interactions.

ACKNOWLEDGMENTS

We thank Kimberley Zichichi from the Electron Microscopy Facility at Yale University for technical assistance with immunoelectron microscopy; Anne Kane, Tufts Medical Center, for preparation of reagents and for helpful discussions; Kamil Kim, Albert Einstein College of Medicine, for the T. gondii RH strain and reagents; Lloyd Kasper, Dartmouth University for the SAG-1 antibody; and Richard Cummings, Emory University, for helpful discussions.

S.B. is supported by NIH T32 AI07389 to Tufts Medical Center, J.L. is supported by NIH T32 AI07077 to Tufts University School of Medicine, and Graduate Biomedical Sciences Program in Immunology and T32 GM008448 to Tufts University School of Medicine Medical Scientist Training Program.

REFERENCES


