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

Seema Bhalchandra,^a Jacob Ludington,^{a,b,c} Isabelle Coppens,^d Honorine D. Ward^{a,b,c}

Division of Geographic Medicine and Infectious Diseases, Tufts Medical Center,^a Tufts University School of Medicine,^b and Tufts Sackler School of Graduate Biomedical Sciences,^c Boston, Massachusetts, USA; Department of Molecular Microbiology and Immunology and Malaria Research Institute, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA^d

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 (CTLTD) and has orthologs in *C. hominis* and *C. muris*. CTLTD-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 CTLTD-containing protein in protozoa and in *Apicomplexa*. CpClec is predicted to be a type 1 membrane protein, with a CTLTD, an O-glycosylated mucin-like domain, a transmembrane domain, and a cytoplasmic tail containing a YXX Φ sorting motif. The predicted structure of CpClec displays several characteristics of canonical CTLTD-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 CTLTD-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 vitro* 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 (CTLTD).

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 CTLTDs were originally identified as calcium-dependent, carbohydrate recognition domains, many CTLTDs 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).

CTLTD-containing proteins are present in all metazoans and most nonmetazoans (10). However, CTLTD-containing proteins have not previously been reported in protozoans. Here, we report the identification and characterization of the *C. parvum* Clec (CpClec), a novel CTLTD-containing, mucin-like glycoprotein with orthologs in *C. hominis* and *C. muris*. In addition, we report identification of putative CTLTD-containing proteins in three other apicomplexan parasites. Understanding the structure and function of CpClec as well as other apicomplexan CTLTD-contain-

Received 6 April 2013 Returned for modification 30 April 2013

Accepted 17 June 2013

Published ahead of print 1 June 2013

Editor: J. F. Urban, Jr.

Address correspondence to Honorine D. Ward, hward@tuftsmedicalcenter.org.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/IAI.00436-13>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/IAI.00436-13

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 $10,000 \times g$ 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) benzenesulfonyl fluoride hydrochloride, 80 μ M aprotinin, 4 mM bestatin, 1.4 mM E-64, 2 mM leupeptin, 1.5 mM pepstatin A; Sigma-Aldrich, St. Louis, MO] and excysted at 37°C for 1 h. After washing with PBS, the resultant mixture of sporozoites and unexcysted oocysts was centrifuged at $4,500 \times g$ 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 $10,000 \times g$ 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- μ m 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 μ g of streptomycin per ml at 37°C in 5% CO₂.

Identification and *in silico* analysis of *Cryptosporidium* species and other apicomplexan Clects. The CpClec sequence (Cgd3_440) was identified from the *C. parvum* genome sequence in the CryptoDB database (www.cryptodb.org) (13) by using a text search with the word mucin. The *C. hominis* and *C. muris* Clects were identified from their genome sequences by BLAST comparison to the CpClec sequence. Both the *C. parvum* and *C. hominis* Clect 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 *Plasmodium*, *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/scanprosite>). 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 CpClec 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 CpClec 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 CpClec gene expression. Amplification of CpClec and Cp18S rRNA cDNA 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 $2^{-\Delta\Delta C_T}$ method, where the amount of target, normalized to the endogenous reference 18S rRNA gene and relative to expression in uninfected cells, is given by $2^{-\Delta\Delta C_T}$, and C_T 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 Pfx 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 Qiaquick 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 Qiaquick kit (Qiagen) and cloned into the TOPO pCRII vector (Invitrogen), and plasmids were isolated 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 injection.

<i>C. parvum</i>	1	MRLINELKDCILIVLLSLIVSVGVFGREVS KLGDTHAN	100
<i>C. hominis</i>	1	MRLINELKDFILIVLLSLIVSVGVFGREVS KLGDTHAN	100
<i>C. muris</i>	1	MMINGFIDEAIVLLIWLTLWF---HGILSSNTAGRAI	96
		* * * * *	
<i>C. parvum</i>	101	YKLSDEEAIRRTLVEQGDHKTANGCFVGLRKYYAFFRYPGGWYIDSGSPVSRDSEIMWTSSYNASWLLRFKFPFEKCGFFDGGGLRRPCWLIDSELS	200
<i>C. hominis</i>	101	YKLSDEEAIRRTLVEQGDHKTANGCFVGLRKYYAFFRYPGGWYIDSGSPVSRDSEIMWTSSYNASWLLRFKFPFEKCGFFDGGGLRRPCWLIDSELS	200
<i>C. muris</i>	97	HKKTDAVKRVLNFPNPTKYKTANGCFVGLRKYYLFRYPGGWYIDEEHPVSIENASWLLRFKFPFEKCGFFDEGGLRRPCWLIDSELS	196
		* . * . * * *	
<i>C. parvum</i>	201	CAVCGTVAYDSGGLNPNKYSELIKQTDEIISKEPKLDLPGYNNLCTDAWENNLSVGLCQVNSNILDASWSSYQSSLNFPISFNWFKEDSTNEGGDQVYH	300
<i>C. hominis</i>	201	CAVCGTVAYDSGGLNPNKYSELIKQTDEIISKEPKLDFPGYNNLCTDAWENNLSVGLCQVNSNILDASWSSYQSSLNFPISFNWFKEDSTNEGGDQVYH	300
<i>C. muris</i>	197	CAICGTVAKDSGGGGPNKYSELIQVDEIISKEPKFDIPQSSILMCSLDNNSNATLNDVNSIINNAWLNTHSS-TPNINFWEKRSYEGGGLNLFN	295
		* * . * * * * *	
<i>C. parvum</i>	301	NSYLDLPRYKQKKT FYWQDQPGTIPALSDMKLIGLSAQPTYHPLDRSSSGSFESDSTESGRANEERNDTLEEMYKQSSSDIEDILVPVTKPIFSDYSS	400
<i>C. hominis</i>	301	NSYLDLPRYKQKKT FYWQDQPGTIPALSDMKLIGLSAQPTYHPLARSSSGSFESDSTESGRVNEERNDTLEEMYKQSSSDIEDILVPITKPIFSDYSS	400
<i>C. muris</i>	296	NTNPHDIRLKQKGYSSSSE-NISEMSDSMLVLPVTVTSQKYGL-----DIEN-----NKYSH	349
		* . * * * * *	
<i>C. parvum</i>	401	IGRNSGVEEVYSSNVSTISSKGVVPFNTTSSDNDFTKNNPANSSKH	500
<i>C. hominis</i>	401	IGRNSGVEEVYSSNVSTISSKGVVPFNTTASSDNTKNNSSNSKH	497
<i>C. muris</i>	350	MSRTS-----TNTVNSEFTSRTTITTVATSTRTAVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-----TTTTTTTTTSSNI	422
		* * * * *	
<i>C. parvum</i>	498	TTTLRPRFYTTGKSLYTPKTENVRTSAYVKKNSNSLTGNVVKRDSGIIIMNSDKEAENSLLGFAGAGEY-VNSESDDPQETQELS-YLNEAG-QVKMEKS	594
<i>C. hominis</i>	501	TTTLRPRFYTTGKSLYTPKTENVRTSAYVKKNSNSLTGNVVKRDSGIIIMNSDKEAENSLLGFAGAGEY-VNSESDDPQETQELS-YLNEAG-QVKMEKS	597
<i>C. muris</i>	423	VTTIE---YTS-----VPIVNTLNYIEK---QLIIGSSEPNYES-----KMDLDDLMLMGFGEENEIMSQSFYPYHSQEILGYQDDINTVQVDLN	503
		* * . * * * *	
<i>C. parvum</i>	595	IEGHFDVIERTVNKKLIIGVIGILLILILMFCCVFFLRRSSDDTESAQNVDSNKEIIRKNL-SQLASIKPILARKGSKIIVHMSQKSIYSYDSNIAP	693
<i>C. hominis</i>	598	IEGHFDVIERTVNKKLIIGVIGILLILILMFCCVFFLRRSSDDTESAQNVDSNKEIIRKNL-SQLASIKPILARKGSKIIVHMSQKSIYSYDSNIAP	696
<i>C. muris</i>	504	NESLQKLSDRPANSAIIGSSVGGAVALGSIIILCCVACRNKNDNGD-FNSENQGNIQISNISMDSSIRENSLEPACS-DSPAFMIIDKSVNHNVLMTQ	601
		* * * * *	
<i>C. parvum</i>	694	PEKKSIVSVEDTTFSPYMNMSGEGGAETIPYVAQSQANTLQLHSGRNINLGYDTRPQTQKQAVLKS PKYHDFFYFE-----SSR 775	
<i>C. hominis</i>	697	PEKKSIVSVEDTTFSPYMNMSGEGGAETIPYVAQSQANTLQLHSGRNINLGYDTRPQTQKQAVLKS PKYHDFFYFE-----PSR 778	
<i>C. muris</i>	602	SDQ-VSISIEDTTFSPYITQNGTPETGKQP-FCVAQSQAALNLHNSNNIAIHYDTRQGNFKNALKS AKYHEEYFYSQGGDSNSSK 686	
		* . * * * * *	

FIG 1 Alignment of *Cryptosporidium* CTLD-containing proteins (Clecs). Sequences were identified in CryptoDB and aligned by using MacVector v. 12. The 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, the mucin-like domain is in light gray, and the transmembrane domain is underlined. The YXX ϕ sorting motif (YHD/EF) in the cytoplasmic tail is shown in bold and underlined. Predicted O-glycosylation sites are in bold, and N-glycosylation sites are italicized and underlined.

tion. For subsequent boosts, mice were immunized with the same antigen emulsified with incomplete Freund's adjuvant at 3-week intervals for a total of 3 boosts. Mice were bled from the tail vein, and antibody (anti-rCpClec) production was monitored by SDS-PAGE and immunoblotting. Rabbit polyclonal antibody to the recombinant *C. parvum* surface protein gp15 (anti-rgp15) has been described previously (20).

SDS-PAGE and immunoblotting. *C. parvum* lysates were heated with reducing or nonreducing sample buffer for 5 min at 95°C, separated by SDS-PAGE, transferred to nitrocellulose, and probed with anti-rCpClec. Horseradish peroxidase (HRP)-conjugated anti-rat and anti-mouse IgGs (Southern Biotech) were used as secondary antibodies. Immunoblots were developed with the SuperSignal West Pico chemiluminescent substrate (Thermo Scientific, Waltham, MA).

Immunofluorescence assay (IFA). Hypochlorite-treated oocysts were resuspended in PBS and excysted on poly-L-lysine-coated coverslips at 37°C for 1 h, then fixed with ice-cold methanol for 20 min at room temperature (RT) or freshly prepared 4% paraformaldehyde in PBS for 30 min at RT. Coverslips were placed in 6-well plates, and nonspecific binding was blocked with 1% bovine serum albumin in PBS at RT and then probed with anti-rCpClec and anti-rgp15 antibodies. Reactivity was detected with Alexa Fluor 594-conjugated goat anti-rabbit IgG and Alexa Fluor 488-conjugated goat anti-mouse IgG, respectively (Invitrogen). For localization of CpClec in intracellular stages of *C. parvum*, Caco-2 cells were grown to confluence on coverslips in 6-well plates, infected with *C. parvum* oocysts for 24 h at 5% CO₂ and 37°C, and fixed in methanol at RT for 10 min, followed by incubation in PBS for 10 min. Coverslips were incubated with mouse anti-rCpClec and anti-rgp15 for 1 h at RT, followed by incubation with secondary antibodies (Alexa 488-labeled anti-mouse IgG and Alexa 594-labeled anti-rabbit IgG; Invitrogen) for 1 h at RT. After washing, the slides were mounted with Vectashield mounting medium containing 4',6'-diamidino-2-phenylindole (DAPI; Vector Laboratories, Burlingame, CA) and examined by differential interference contrast (DIC) and fluorescence microscopy using a Zeiss Axioimager Z.1 fluores-

cence microscope (Carl Zeiss Microscopy, Jena, Germany). Images were deconvolved and captured with an IEEE1394 digital charge-coupled-device camera (Hamamatsu, Hamamatsu City, Japan), and colocalization of the fluorescent labels was achieved by using Volocity software (Improvion Inc., Lexington, MA).

Immunoelectron microscopy (IEM). Monolayers of Caco-2 cells were infected with *C. parvum* sporozoites at a multiplicity of infection of 2 for 36 h before fixation, first in 4% paraformaldehyde (Electron Microscopy Sciences, PA) in 0.25 M HEPES (pH 7.4) for 1 h at RT and then in 8% paraformaldehyde in the same buffer overnight at 4°C. Samples were infiltrated, frozen, and sectioned as previously described (21). The sections were immunolabeled with anti-rCpClec at a 1:50 dilution in PBS containing 1% fish skin gelatin and then with anti-mouse IgG, followed directly by 10-nm protein A-gold particles, before examination with a CM120 electron microscope (Philips, Eindhoven, Netherlands) at 80 kV.

RESULTS

Identification and *in silico* analysis of *Cryptosporidium* species ClecS. While mining the *C. parvum* genome for genes encoding mucin-like glycoproteins, we identified a novel gene that encoded a hypothetical protein that was predicted to contain a mucin-like domain and a CTLD and that had orthologs in *C. hominis* and *C. muris*. The sequence of the *in silico*-identified *C. parvum* gene was identical to that of the cloned cDNA from infected Caco-2 cells. *CpClec* has a two-exon gene structure with a short 78-bp intron within the 5' end. The first exon encodes the signal peptide, and the second exon encodes the rest of the protein. *CpClec* is predicted to be a type 1 membrane protein with a signal peptide, an extracellular region containing the CTLD, and a mucin-like domain, a transmembrane domain, and a short cytoplasmic tail (Fig. 1). The mucin-like domain is predicted to display mucin-

TABLE 1 Properties of *Cryptosporidium* CTLD-containing proteins

Property	<i>C. parvum</i>	<i>C. hominis</i>	<i>C. muris</i>
Gene ID	cgd3_440	chro.30061	cmu_038310
GenBank accession no.	JX417972	JX417971	EEA04765.1
No. of amino acids ^a	775	778	686
Mass (kDa) ^b	86.4	86.6	76.1
Isoelectric point ^b	5.87	5.96	4.99
Amino acid content (%) ^c			
Proline	4	4	3.8
Serine	11.5	11.2	10.8
Threonine	10.7	11.2	12.8
O-glycosylation sites ^d	40	45	49
N-glycosylation sites ^e	6	5	7
% nucleotide identity ^f		97.2	54.3
% amino acid identity ^f		97.7	37.7
No. of ESTs ^a	3	N/A	N/A

^a Obtained from CryptoDB (18).^b Determined using Compute pI/MW.^c Determined using ProtParam.^d Identified using NetOGlyc 3.^e Identified using NetNGlyc1.^f Nucleotide/deduced amino acid sequence compared to that of *C. parvum* Clec.

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 YXXΦ 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.

Comparison of the CpClec CTLD with those of other apicomplexan CTLD-containing proteins. In addition to *Cryptosporidium*, we also identified for the first time genes that encode CTLD-containing proteins from three other apicomplexan parasites whose genomes have been sequenced, namely, *Toxoplasma gondii*, *Neospora caninum*, and *Eimeria tenella*, which we named TgClec, NcClec, and EtClec, respectively (Fig. 2B). The genomes of 3 strains of *T. gondii* (ME49, VEG, and GT1) contained genes that encode TgClec, all of which were identical. We identified an identical TgClec in the type I RH strain by PCR amplification, cloning, and sequencing of a genomic DNA clone from this strain (Fig. 2B, CTLD sequence).

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 apicomplexan 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* macrophage 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,

FIG 2 Alignment of the *CpClec* CTLD with those of other CTLD-containing proteins. The deduced amino acid sequence of the *CpClec* CTLD and those of representative known CTLD-containing proteins (A) or apicomplexan CTLDs (B) were aligned by using MacVector v.12. Alignments were manually adjusted for structurally equivalent positions by introducing gaps. Accession numbers are as follows: *CpClec* (JX417972), *ChClec* (JX417971.1), *CmClec* (EEA04765.1), *S. mansoni* M23B CTLD 4 (CCD81213.1), *C. elegans* Clec55 (NP_506591.1), *M. musculus* Dectin (NP_064392), *R. norvegicus* MBPA (NP_036731.2), *Homo sapiens* MMR CTLD 4 (NP_002429.1), *Homo sapiens* CD69 (CAA80298.1), *TgClec* (KC812391), *EtClec* (XM_001238736), *NcClec* (XP_003885782.1). The consensus sequence shows identities as asterisks and similarities as dots (periods). Conserved residues important for the structure of the CTLD are shaded dark gray (identical) or light gray (similar). Conserved cysteine residues are in bold and italics and labeled C to C4. The conserved E69 residue and the WIGL motif are in bold and labeled. Conserved residues EPN and WND, implicated in calcium-dependent carbohydrate binding, are in bold, and the E139 residue that is part of calcium binding site 2 is in bold and labeled. Other conserved residues that are shaded are important for formation of the hydrophobic cores.

and *N*-acetylgalactosamine 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 M_r 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-

FIG 3 Predicted tertiary structure of *CpClec*. (A) The tertiary structure of *CpClec* was predicted using Phyre². The ribbon model is colored based on a rainbow from the N to C terminus, with the predicted signal peptide (SP), CTLD, mucin-like domain (MLD), and transmembrane domain (TMD). (B) The 3DLigandSite server was used to identify predicted ligand-binding sites in the *CpClec* CTLD for ligands that include Ca^{2+} , mannose (Man), galactose (Gal), and N-acetylgalactosamine (GalNAc). Amino acid residues are numbered according to the CTLD sequence, and residues predicted to be involved in ligand binding are indicated in blue. Also shown are the 6 conserved cysteine residues (C0, C0', C1, C2, C3, and C4), the highly conserved E69 residue, and the E139 residue known to be involved in calcium-dependent carbohydrate binding. Also indicated is the long loop region (LLRHC).

FIG 4 Expression of *CpClec* in *C. parvum*-infected Caco-2 cells. (A) 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. (B) 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 .

ized with anti-rgp15 against the surface membrane-associated gp15 antigen (Fig. 6). 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).

Ultrastructural localization of *CpClec* in intracellular stages by immunoelectron microscopy. Immunogold staining analyses with anti-r*CpClec* 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

FIG 5 Purification of r*CpClec* and identification of native *CpClec*. (A) The r*Clec* 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. (B) Immunoblot analysis of *C. parvum* lysate probed with preimmune sera (lane 1) or anti-r*CpClec* (lane 2). The arrow indicates the mass of native *CpClec*.

FIG 6 Localization of *CpClec* in *C. parvum* by immunofluorescence. Sporozoites (a to h) and merozoites within meronts in infected Caco-2 cells (i to p) were stained with anti-r*CpClec*_{633–775} (α-r*CpClec*_{633–775}) (a to d and i to l) or preimmune sera (e to h and m to p) and anti-Cprgp15, followed by Alexa 488-labeled anti-mouse IgG and Alexa 594-labeled anti-rabbit IgG. Staining was localized to the surface and apical region of sporozoites (b and c) and merozoites within intracellular meronts (j and k). Nuclei were stained with DAPI.

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

CTLCD-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 CTLCD 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 CTLCD-containing proteins in protozoans and in apicomplexans, and *Cryptosporidium* species in particular. In this study, we identified, characterized, and expressed *CpClec*, a novel CTLCD-containing, mucin-like glycoprotein with orthologs in all three *Cryptosporidium* species whose genomes have been sequenced.

CTLCDs can be structurally divided into two groups: canonical and compact. Canonical CTLCDs contain the long loop region hydrophobic core, which is involved in calcium-dependent carbohydrate binding, and the highly conserved WIGL motif, while compact CTLCDs lack the long loop region hydrophobic core (11). Our analyses revealed that *CpClec* has a canonical CTLCD, 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* CTLCD, a key glutamic acid

FIG 7 Ultrastructural localization of *CpClec* in *C. parvum* by immunoelectron microscopy. Immunogold staining of *C. parvum* infecting Caco-2 cells for 36 h with anti-r*CpClec*_{633–775} (α-r*CpClec*_{633–775}) antibody was revealed with protein A-gold particles. (A) PV with 6 merozoites decorated with gold particles on the plasma membrane (PM) at the apical pole, rich in micronemes (left inset), and in the median part, where the nucleus (nu) is located. (B) A parasite exhibits intracellular staining on dense granules (DG). No labeling of other secretory organelles, such as micronemes (mi) or rhoptries, was observed. (C) An extracellular sporozoite labeled similarly to the intravacuolar merozoites. In panels A (inset at bottom) and D, staining on the feeder organelle (FO; arrows) as well as the dense band (DB; arrowheads) is shown at the interface between the PV and the host cell (hc). Bars, 200 nm.

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* CTLCD. However, it remains to be experimentally determined whether *CpClec* binds glycans and if so, whether binding is calcium dependent or independent.

CTLCD-containing proteins are currently classified into 17 groups (10, 11). The predicted tertiary structure of the *CpClec* CTLCD 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)-D-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 Clects from all three *Cryptosporidium* species contains the evolutionarily conserved YXX Φ 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 YXX Φ motif is essential for targeting membrane proteins to secretory organelles of the apical complex. The YXX Φ motifs in the cytoplasmic tails of *T. gondii* rhoptry-associated protein 2 (ROP-2) (32) and *Plasmodium* thrombospondin-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 YXX Φ 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 reinvasion of epithelial cells by merozoites (37). A recent study of the transcriptome of the intra-

cellular 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 CTLDcps. 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* Clects. 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 EtClec 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 Clects 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 CTLs (45). Invasin 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, Tc-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 *CpClec* 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 *CpClec*. However, taken together, the predicted structural features and localization of *CpClec* along with the known functions of *C. parvum* mucin-like glycoproteins and of CTLD-containing proteins strongly implicate *CpClec* 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; Kami 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 Sackler School of Graduate Biomedical Sciences Program in Immunology and T32 GM008448 to Tufts University School of Medicine Medical Scientist Training Program.

REFERENCES

- Leav BA, Mackay M, Ward HD. 2003. *Cryptosporidium* species: new insights and old challenges. *Clin. Infect. Dis.* 36:903–908.
- Collinet-Adler S, Ward HD. 2010. *Cryptosporidiosis*: environmental, therapeutic, and preventive challenges. *Eur. J. Clin. Microbiol. Infect. Dis.* 29:927–935.
- Tzipori S, Ward H. 2002. *Cryptosporidiosis*: biology, pathogenesis and disease. *Microbes Infect.* 4:1047–1058.
- Shirley DA, Moonah SN, Kotloff KL. 2012. Burden of disease from cryptosporidiosis. *Curr. Opin. Infect. Dis.* 25:555–563.
- O'Connor RM, Shaffie R, Kang G, Ward HD. 2011. *Cryptosporidiosis* in patients with HIV/AIDS. *AIDS* 25:549–560.
- Cabada MM, White AC, Jr. 2010. Treatment of cryptosporidiosis: do we know what we think we know? *Curr. Opin. Infect. Dis.* 23:494–499.
- Abubakar I, Aliyu SH, Arumugam C, Hunter PR, Usman NK. 2007. Prevention and treatment of cryptosporidiosis in immunocompromised patients. *Cochrane Database Syst. Rev.* 1:CD004932.
- Wanyiri J, Ward H. 2006. Molecular basis of *Cryptosporidium*-host cell interactions: recent advances and future prospects. *Future Microbiol.* 1:201–208.
- Bouazid M, Hunter PR, Chalmers RM, Tyler KM. 2013. *Cryptosporidium* pathogenicity and virulence. *Clin. Microbiol. Rev.* 26:115–134.
- Cummings RD, McEver RP. 2009. C-type lectins, p 439–457. *In* Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (ed), *Essentials of glycobiology*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Zelensky AN, Gready JE. 2005. The C-type lectin-like domain superfamily. *FEBS J.* 272:6179–6217.
- Drickamer K. 1999. C-type lectin-like domains. *Curr. Opin. Struct. Biol.* 9:585–590.
- Heiges M, Wang H, Robinson E, Aurrecochea C, Gao X, Kaluskar N, Rhodes P, Wang S, He CZ, Su Y, Miller J, Kraemer E, Kissinger JC. 2006. *CryptoDB*: a *Cryptosporidium* bioinformatics resource update. *Nucleic Acids Res.* 34:D419–D422.
- Aurrecochea C, Brestelli J, Brunk BP, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer ET, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Srinivasamoorthy G, Stoeckert CJ, Jr, Thibodeau R, Treatman C, Wang H. 2010. *EuPathDB*: a portal to eukaryotic pathogen databases. *Nucleic Acids Res.* 38:D415–D419.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA 5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731–2739.
- Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4:363–371.
- Wass MN, Kelley LA, Sternberg MJ. 2010. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* 38:W469–W473.
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55:611–622.
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C_T)$ method. *Methods* 25:402–408.
- O'Connor RM, Wanyiri JW, Cevallos AM, Priest JW, Ward HD. 2007. *Cryptosporidium parvum* glycoprotein gp40 localizes to the sporozoite surface by association with gp15. *Mol. Biochem. Parasitol.* 156:80–83.
- Folsch H, Pypaert M, Schu P, Mellman I. 2001. Distribution and function of AP-1 clathrin adaptor complexes in polarized epithelial cells. *J. Cell Biol.* 152:595–606.
- Bonifacino JS, Traub LM. 2003. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* 72:395–447.
- Zelensky AN, Gready JE. 2003. Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). *Proteins* 52:466–477.
- Llera AS, Viedma F, Sanchez-Madrid F, Tormo J. 2001. Crystal structure of the C-type lectin-like domain from the human hematopoietic cell receptor CD69. *J. Biol. Chem.* 276:7312–7319.
- Brown J, O'Callaghan CA, Marshall AS, Gilbert RJ, Siebold C, Gordon S, Brown GD, Jones EY. 2007. Structure of the fungal beta-glucan-binding immune receptor dectin-1: implications for function. *Protein Sci.* 16:1042–1052.
- Sancho D, Gomez M, Sanchez-Madrid F. 2005. CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol.* 26:136–140.
- Kavan D, Kubickova M, Bily J, Vanek O, Hofbauerova K, Mrazek H, Rozbesky D, Bojarova P, Kren V, Zidek L, Sklenar V, Bezouska K. 2010. Cooperation between subunits is essential for high-affinity binding of N-acetyl-D-hexosamines to dimeric soluble and dimeric cellular forms of human CD69. *Biochemistry* 49:4060–4067.
- Pavlicek J, Sopko B, Ettrich R, Kopecky V, Jr, Baumruk V, Man P, Havlicek V, Vrbacky M, Martinkova L, Kren V, Pospisil M, Bezouska K. 2003. Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry* 42:9295–9306.
- Xie J. 2012. The C-type lectin-like receptors of Dectin-1 cluster in natural killer gene complex. *Glycoconj. J.* 29:273–284.
- Pandey KN. 2009. Functional roles of short sequence motifs in the endocytosis of membrane receptors. *Front. Biosci.* 14:5339–5360.
- Sancho D, Reis e Sousa C. 2012. Signaling by myeloid C-type lectin receptors in immunity and homeostasis. *Annu. Rev. Immunol.* 30:491–529.
- Hoppe HC, Ngo HM, Yang M, Joiner KA. 2000. Targeting to rhoptry organelles of *Toxoplasma gondii* involves evolutionarily conserved mechanisms. *Nat. Cell Biol.* 2:449–456.
- Bhanot P, Frevert U, Nussenzweig V, Persson C. 2003. Defective sorting of the thrombospondin-related anonymous protein (TRAP) inhibits *Plasmodium* infectivity. *Mol. Biochem. Parasitol.* 126:263–273.
- Bonnin A, Gut J, Dubremetz JF, Nelson RG, Camerlynck P. 1995. Monoclonal antibodies identify a subset of dense granules in *Cryptosporidium parvum* zoites and gamonts. *J. Eukaryot. Microbiol.* 42:395–401.
- Petry F, Harris JR. 1999. Ultrastructure, fractionation and biochemical analysis of *Cryptosporidium parvum* sporozoites. *Int. J. Parasitol.* 29:1249–1260.
- Blackman MJ, Bannister LH. 2001. Apical organelles of Apicomplexa: biology and isolation by subcellular fractionation. *Mol. Biochem. Parasitol.* 117:11–25.
- Mauzy MJ, Enomoto S, Lancto CA, Abrahamsen MS, Rutherford MS. 2012. The *Cryptosporidium parvum* transcriptome during in vitro development. *PLoS One* 7(3):e31715. doi:10.1371/journal.pone.0031715.
- Belli SI, Smith NC, Ferguson DJ. 2006. The coccidian oocyst: a tough nut to crack! *Trends Parasitol.* 22:416–423.

39. Lau AO. 2009. An overview of the Babesia, Plasmodium and Theileria genomes: a comparative perspective. *Mol. Biochem. Parasitol.* **164**:1–8.
40. Hanisch FG. 2001. O-glycosylation of the mucin type. *Biol. Chem.* **382**:143–149.
41. Barnes DA, Bonnin A, Huang JX, Gousset L, Wu J, Gut J, Doyle P, Dubremetz JF, Ward H, Petersen C. 1998. A novel multi-domain mucin-like glycoprotein of *Cryptosporidium parvum* mediates invasion. *Mol. Biochem. Parasitol.* **96**:93–110.
42. Cevallos AM, Bhat N, Verdon R, Hamer DH, Stein B, Tzipori S, Pereira ME, Keusch GT, Ward HD. 2000. Mediation of *Cryptosporidium parvum* infection in vitro by mucin-like glycoproteins defined by a neutralizing monoclonal antibody. *Infect. Immun.* **68**:5167–5175.
43. Cevallos AM, Zhang X, Waldor MK, Jaison S, Zhou X, Tzipori S, Neutra MR, Ward HD. 2000. Molecular cloning and expression of a gene encoding *Cryptosporidium parvum* glycoproteins gp40 and gp15. *Infect. Immun.* **68**:4108–4116.
44. O'Connor RM, Burns PB, Ha-Ngoc T, Scarpato K, Khan W, Kang G, Ward H. 2009. Polymorphic mucin antigens CpMuc4 and CpMuc5 are integral to *Cryptosporidium parvum* infection in vitro. *Eukaryot. Cell* **8**:461–469.
45. Shaw PL, Kirschner AN, Jardetzky TS, Longnecker R. 2010. Characteristics of Epstein-Barr virus envelope protein gp42. *Virus Genes* **40**:307–319.
46. Dersch P, Isberg RR. 2000. An immunoglobulin superfamily-like domain unique to the *Yersinia pseudotuberculosis* invasin protein is required for stimulation of bacterial uptake via integrin receptors. *Infect. Immun.* **68**:2930–2938.
47. Batchelor M, Prasannan S, Daniell S, Reece S, Connerton I, Bloomberg G, Dougan G, Frankel G, Matthews S. 2000. Structural basis for recognition of the translocated intimin receptor (Tir) by intimin from enteropathogenic *Escherichia coli*. *EMBO J.* **19**:2452–2464.
48. Loukas A, Maizels RM. 2000. Helminth C-type lectins and host-parasite interactions. *Parasitol. Today* **16**:333–339.
49. Loukas A, Doedens A, Hintz M, Maizels RM. 2000. Identification of a new C-type lectin, TES-70, secreted by infective larvae of *Toxocara canis*, which binds to host ligands. *Parasitology* **121**:545–554.