

Unique Posttranslational Modifications of Chitin-Binding Lectins of *Entamoeba invadens* Cyst Walls

Katrina L. Van Dellen,¹ Anirban Chatterjee,¹ Daniel M. Ratner,^{1,2} Paula E. Magnelli,¹ John F. Cipollo,¹ Martin Steffen,³ Phillips W. Robbins,¹ and John Samuelson^{1,4*}

Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, Boston, Massachusetts 02118¹; Section of Infectious Diseases, Boston Medical Center, Boston, Massachusetts 02118²; and Departments of Genetics and Genomics³ and Microbiology,⁴ Boston University School of Medicine, Boston, Massachusetts 02188

Received 17 January 2006/Accepted 3 March 2006

Entamoeba histolytica, which causes amebic dysentery and liver abscesses, is spread via chitin-walled cysts. The most abundant protein in the cyst wall of *Entamoeba invadens*, a model for amebic encystation, is a lectin called EiJacob1. EiJacob1 has five tandemly arrayed, six-Cys chitin-binding domains separated by low-complexity Ser- and Thr-rich spacers. *E. histolytica* also has numerous predicted Jessie lectins and chitinases, which contain a single, N-terminal eight-Cys chitin-binding domain. We hypothesized that *E. invadens* cyst walls are composed entirely of proteins with six-Cys or eight-Cys chitin-binding domains and that some of these proteins contain sugars. *E. invadens* genomic sequences predicted seven Jacob lectins, five Jessie lectins, and three chitinases. Reverse transcription-PCR analysis showed that mRNAs encoding Jacobs, Jessies, and chitinases are increased during *E. invadens* encystation, while mass spectrometry showed that the cyst wall is composed of an ~30:70 mix of Jacob lectins (cross-linking proteins) and Jessie and chitinase lectins (possible enzymes). Three Jacob lectins were cleaved prior to Lys at conserved sites (e.g., TPSVDK) in the Ser- and Thr-rich spacers between chitin-binding domains. A model peptide was cleaved at the same site by papain and *E. invadens* Cys proteases, suggesting that the latter cleave Jacob lectins in vivo. Some Jacob lectins had *O*-phosphodiester-linked carbohydrates, which were one to seven hexoses long and had deoxysugars at reducing ends. We concluded that the major protein components of the *E. invadens* cyst wall all contain chitin-binding domains (chitinases, Jessie lectins, and Jacob lectins) and that the Jacob lectins are differentially modified by site-specific Cys proteases and *O*-phosphodiester-linked glycans.

The infectious form of *Entamoeba histolytica*, a protozoan parasite that causes amebic dysentery and liver abscesses, is the quadrinucleate cyst (10, 15, 25). Because *E. histolytica* does not encyst in axenic culture, cyst formation has been studied in *Entamoeba invadens*, a reptilian pathogen that also forms quadrinucleate cysts (7, 31). *E. invadens* converts to chitin-walled cysts within 2 days when subjected to osmotic shock and/or glucose deprivation (30). The *E. invadens* cyst wall resembles the *E. histolytica* cyst wall, and *E. invadens* cysts are able to excyst readily when placed in full medium. Cyst formation is blocked by the addition of inhibitors of Cys or Ser proteases, although the mechanism of inhibition is unclear (21, 23). *E. histolytica* has 20 lysosomal Cys protease genes, 8 of which are transcribed by trophozoites in culture, while *E. invadens* has multiple Cys proteases (4, 23).

The most abundant protein in the cyst wall of *E. invadens* is the Jacob lectin (EiJacob1), a secreted glycoprotein that contains five tandemly arranged chitin-binding domains (CBDs) (9). Each EiJacob1 CBD contains six conserved Cys residues and numerous conserved aromatic amino acids. EiJacob1 was identified by sequencing the largest of >20 spots on two-dimensional (2-D) gels containing proteins isolated from purified *E. invadens* cyst walls. The finding that quadrinucleate, “wall-less” cysts are formed when encystation is induced in the

presence of galactose suggested a two-lectin model of *E. invadens* cyst wall formation. A plasma membrane galactose lectin binds sugars on Jacob, and the Jacob lectin, in turn, cross-links chitin fibrils (9, 16).

Entamoeba chitinases, which may be involved in remodeling the walls of encysting parasites or in degrading the walls of excysting parasites, have a unique, N-terminal CBD that contains eight conserved Cys residues (6, 29). The same eight-Cys CBD is also present in a family of *E. histolytica* lectins named Jessie, which have an unknown function. The six-Cys and eight-Cys CBDs of *Entamoeba* are reminiscent of, but not homologous to, the Cys-rich CBDs found in insect chitinases and peritrophic membrane proteins and in plant lectins, such as wheat germ agglutinin (26, 35).

E. histolytica trophozoites have two types of glycans on their surfaces. Asn-linked glycans (*N*-glycans) are built upon a truncated lipid precursor that contains 7 sugars (Man₅GlcNAc₂) rather than the 14 sugars present in precursors from animals, plants, and fungi (Glc₃Man₅GlcNAc₂) (22). *E. histolytica* *N*-glycans include unprocessed Man₅GlcNAc₂ as well as complex glycans containing Gal and Glc (our unpublished data). *E. histolytica* trophozoites also have proteophosphoglycans, which are glycosylphosphatidylinositol-anchored peptides with *O*-phosphodiester-linked glycans (*O*-P-glycans), on their surfaces (18). The *O*-P-glycans, which are immunogenic, contain Gal at the reducing end and are extended by Glc.

The goal of the present study was to answer a number of questions concerning the *E. invadens* cyst wall, as follows. First, how many genes encoding Jacob lectins, Jessie lectins, and

* Corresponding author. Mailing address: Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, 715 Albany St., Evans 426, Boston, MA 02118. Phone: (617) 414-1054. Fax: (617) 414-1041. E-mail: jsamuels@bu.edu.

TABLE 1. Primers used for RT-PCR analysis

Target	Sense primer	Antisense primer
EiJacob1	CCATACACATTAAGCCCTTGT	CAGACGCATGGATCAGTGTC
EiJacob2	AATGGTATCGATTGCAAGGAA	TAACACAATACAACCCGTCTT
EiJacob3	ATGTTGATACTACTGTTACTG	CAACATCATCAAGACAATAAT
EiJacob4	TGTTCTTCTTGTATTCTTC	TGTTGAGTAAGCGACTTGGTT
EiJacob5	TTACAGTCACTCTCTGATG	TTGATTATTGTGGATCCATC
EiJacob6	CACTTCTCTTTGCGATAACAG	GAGTTTCTGGTTCCCTTGATT
EiJacob7	CGTTCCAAATGAGAGCTGTTT	AATCGAAAGATTCAATATAAAA
EiJessie1a	ATGTTGACTTACTTTGAAGT	TCATACCTCGTCGGAAATGTA
EiJessie1b	ATGAATAGTTTCTTCTATTG	TTAAATTAAGCGAACAATAA
EiJessie1c	ATGTTCTTGGTTGTGTTAACT	TTATAATAGAACATAAAGTAC
EiJessie3a	ATGGCAAAGAGAAAGACAAGA	TGTAGCACAGAATGTAATGG
EiJessie3b	ATGAAGAAAAGAACAACAAGA	TGTCGAGCAGAACGAATAAGG
EiChitinase1	ATGACGTACGATCTGCATGGT	CAAGTCCAACTGGCAATCAT
EiChitinase2	TCAACAAACAATACAACCGAACACA	GTGCAACGCCACATTAC
EF1 α	CCTCTGGTATCGGAACAGTCCC	AAAGTCTCAACACAGAGTGGC
Actin	CACTGACTACCTCATGAAGAT	AGTCTGTCTGTTGATGCCTGGG

chitinases are present in the *E. invadens* genome, and are they encystation specific like EiJacob1 and *E. invadens* chitinase 1? Second, what are the identities of the >20 uncharacterized spots present on two-dimensional gels containing *E. invadens* cyst wall proteins? Are they mostly Jacob lectins, which cross-link chitin, or are there other chitin-binding proteins in the cyst wall? Can other proteins be identified by mass spectrometry (MS) of tryptic fragments of cyst wall proteins? Third, what are the posttranslational modifications (e.g., glycosylation or proteolytic processing) of Jacob lectins and other proteins present in the *E. invadens* cyst wall? What is the endopeptidase(s) that cleaves between CBDs of Jacob lectins?

MATERIALS AND METHODS

Database searches and sequence assembly. The *E. invadens* genome has been extensively sequenced, so complete genes were most often present within large contigs assembled at The Institute for Genomic Research (<http://www.tigr.org/tdb/e2k1/eha1/>) (31). *E. invadens* Jacob lectins were identified in TBLASTN searches of the strain IP-1 genomic DNA sequence using the EiJacob1 lectin sequence (GenBank accession number AF175527) (1, 9). Five novel *E. invadens* Jacob lectin genes were complete (EiJacob2 to EiJacob6), while EiJacob7 was missing the 3' end. Five complete *E. invadens* Jessie lectins (EiJessie1a, -1b, and -1c and EiJessie3a and -3b) were identified with the sequences for *E. histolytica* Jessie lectins 1 to 3 (GenBank accession numbers AF401986 to AF401988) (1, 29, 31). PCR was used to link the 5' and 3' portions of the EiJessie3a gene. Two additional complete *E. invadens* chitinase genes (*E. invadens* chitinases 2 and 3) were identified with the *E. invadens* chitinase 1 sequence (GenBank accession number AAB52724) (1, 6, 31).

Signal peptides and transmembrane helices of *E. invadens* proteins were predicted using SignalP and TMHMM, respectively (14, 19). Jessie lectin and chitinase sequences were aligned using ClustalW (27).

RT-PCR analysis of Jacob expression in encysting *E. invadens*. *E. invadens* strain IP-1 was induced to encyst by transferring cells from TYI-S-33 to 47% low-glucose (LG) medium (8, 9, 30). For RNA preparations, cells were harvested immediately after the transfer (0 h) and after 22 h in 47% LG medium. Total RNA was isolated using an RNAqueous-4PCR kit (Ambion, Austin, TX) and was treated with DNase I for 1 h at 37°C. First-strand cDNA synthesis was done with RETROscript (Ambion), using oligo(dT) primers. Reverse transcriptase (RT) was omitted from negative control reactions. PCR was performed for 25 cycles of 30 s at 95°C, 30 s at 50°C, and 1 min at 72°C, using AmpliTaq Gold (Applied Biosystems, Foster City, CA). PCR primers, all of which were tested with *E. invadens* genomic DNA, are listed in Table 1.

Mass spectrometry of tryptic fragments of cyst wall proteins. *E. invadens* strain IP-1 was induced to encyst, and purified cyst walls were obtained by previously described methods (9). Previous transmission microscopy had shown that *E. invadens* cyst wall preparations are, for the most part, clean of membranes or organelles (9). *E. invadens* cyst wall proteins were released in 1% sodium

dodecyl sulfate (SDS), separated by 1-D SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and stained with Coomassie blue. The entire set of cyst wall proteins or prominent protein bands was cut out and digested in the gel with trypsin (24). Reverse-phase chromatography was carried out using a nano-HPLC pump and autosampler (Surveyor and MicroAS, respectively; Thermo Finnigan, San Jose, CA) on a 10-cm by 100- μ m ID MAGIC C₁₈ reverse-phase capillary column (Michrom, Auburn, CA) at the Boston University Proteomics Core Facility and the MIT CCR Biopolymers Laboratory. Peptides were separated using gradients of 2% to 98% acetonitrile over 30 to 200 min in the presence of 0.1% formic acid (36). Peptides were analyzed using an LTQ ProteomeX ion-trap mass spectrometer (Thermo Finnigan, San Jose, CA), and mass spectra were compared to tryptic digests of predicted *E. invadens* proteins (see below) by using SEQUEST (12). Tryptic peptides with a SEQUEST XCorr score of >1.75, 2.5, or 3.5 for a Z value of 1, 2, or 3, respectively, a peptide probability of <0.01, and a protein probability of <0.001, and proteins with two or more high-scoring fully tryptic peptides were considered present.

Because the *E. invadens* genome sequence is incomplete, it was necessary to manually curate the set of *E. invadens* proteins used to identify the tryptic peptides (31). The *E. invadens* proteins included all those present in GenBank, as well the Jacobs, Jessies, and chitinases identified here. In addition, >100 other *E. invadens* proteins which have signal peptides or transmembrane helices (Cys peptidases, Gal/GalNAc lectins, receptor kinases, etc.) were included. A set of >30 *E. invadens* cytosolic proteins included thioredoxins, actins and associated proteins, and tubulins. The mass spectrometry results for *E. invadens* cyst walls were also searched using predicted tryptic peptides from the *E. invadens* genome sequence (updated 24 January 2006), translated in all six frames (31).

Compositional analysis was performed using the BioWorks 3.2 protein area/height calculation. In short, this algorithm generates a reconstructed ion chromatogram by calculating the precursor mass from the raw data. The smoothed reconstructed ion chromatogram is used to integrate the peak areas for all peptides identified by the SEQUEST algorithm. These integrals are used to create an approximate composition of the sample, as determined by the percentage of the total integral comprising the integral of a given peptide.

Mass spectrometric identification of unmasked phosphopeptides. Isolated cyst wall proteins were treated with 40 mM trifluoroacetic acid (TFA) at 95°C for 10 min to cleave O-P-glycans. The subsequent proteins contained an O-phosphate modification at the serine or threonine that had been occupied by O-P-glycan. These proteins were digested with trypsin (as previously described), and the tryptic peptides were subjected to mass spectrometric identification. Phosphopeptides were identified with SEQUEST or X!Tandem (The Global Proteome Machine Organization) by searching for a neutral loss of 98 Da, 49 Da, or 33 Da from parent ions for a Z value of 1, 2, or 3, respectively. This neutral loss event created a dominant peak with the mass of the parent peptide less a phosphate and water (2, 12). Peptide identification was performed on the suppressed y and b ions when this characteristic neutral loss was observed.

Two-dimensional SDS-PAGE analysis of cyst wall proteins. For each gel, a 20- to 30- μ l packed volume of purified walls was boiled in 2 volumes of SDS buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 5% β -mercaptoethanol, 5% glycerol). Forty microliters of the supernatant from subsequent microcentrifugation was added to 360 μ l of CHAPS buffer (7 M urea, 2 M thiourea, 2% CHAPS, 2% ampholytes,



FIG. 1. Primary structures of seven predicted *E. invadens* Jacob lectins, with locations of tryptic peptides and N-terminal sequences marked. (A) EiJacob1 to EiJacob5 contain short Ser- and Thr-rich spacers between CBDs. (B) EiJacob6 and EiJacob7 contain a large, low-complexity region between CBDs. Stop codons are marked with asterisks, while the incomplete sequence at the C-terminal end of EiJacob7 is marked by a pound symbol (#). Italics mark signal peptides. CBDs for which Cys residues are shown enlarged and in bold are numbered from the N to the C terminus. Conserved peptides (e.g., TPPVD) prior to proven or hypothetical endoprotease sites are also shown enlarged and in bold. Periods mark gaps in the alignments of CBDs, while spaces indicate sites of identified endoprotease cleavages. Single underlines mark tryptic peptides identified by mass spectrometry, while double underlines mark N-terminal sequences obtained by Edman degradation of proteins isolated from 2-D gels (see Fig. 5B). Shaded residues mark Ser residues shown by mass spectrometry to be sites of O-phosphodiester glycosylation (see Fig. 9).

65 mM dithiothreitol, Serdolit MB-1, and 0.01% bromophenol blue) and electrophoresed in a two-dimensional gel (20). 2-D gel electrophoresis was performed using an Investigator 2-D electrophoresis system (Genomics Solutions, Ann Arbor, MI) by K. Doud at the Harvard University NIEHS Center for Environmental Health Proteomics Facility. Precast gels contained ampholytes from pH 4 to 7 in the first (isoelectric focusing) dimension (Amersham Biosciences, Piscataway, NJ) and a gradient of acrylamide from 8 to 18% in the second (SDS-PAGE) dimension. Molecular weight markers used for the second dimension were broad-range SDS-PAGE standards (Bio-Rad, Hercules, CA). SYPRO Ruby protein gel stain (Molecular Probes, Eugene, OR) was used to detect proteins in 2-D gels. Glycoproteins were detected with Pro-Q Emerald 488 glycoprotein stain (Molecular Probes) (11). The stained gels were imaged on a Molecular Imager FX Pro Plus system (Bio-Rad), using a 532-nm diode-pumped solid-state laser (SYPRO Ruby) or a 488-nm argon-ion laser (Pro-Q Emerald 488).

N-terminal sequencing of cyst wall proteins. Cyst wall proteins were separated by 2-D gel electrophoresis and transferred to an Immobilon-P membrane (Millipore, Billerica, MA) using a Panther HEP-1 semidry electroblotter (Owl Separation Systems, Portsmouth, NH) and a three-buffer system (3). The blot was stained with Coomassie blue, and spots were excised for N-terminal protein sequencing by automated Edman degradation. Sequencing was performed by D. McCourt at Midwest Analytical (St. Louis, MO), using protein sequencers from Applied Biosystems. N-terminal sequences of 2-D spots were searched against

peptide sequences predicted from the *E. invadens* strain IP-1 genomic sequence reads and scaffold assemblies.

Cleavage of a model peptide by endogenous and exogenous peptidases. A synthetic peptide (TPSVDKNEDI) labeled at the N terminus with fluorescein isothiocyanate (FITC) was made at the Harvard Biopolymers Laboratory. The peptide (molecular mass, 1,475.5 Da; 0.03 mg/ml) was digested with 20 μ M trypsin (positive control) in 1 mM CaCl₂, 50 mM Tris-HCl (pH 8), or 23 μ M papain in 2.7 mM cysteine-HCl, 0.27 mM sodium EDTA (pH 4.5). Alternatively, the peptide was digested at pH 8 with a lysate of *E. invadens* trophozoites, which was prepared by sonicating amebae in the presence of 0.1% Triton X-100 in phosphate-buffered saline on ice. All assays were performed at 37°C with a shaking Thermomixer (Eppendorf AG, Germany). Protease inhibitors (E-64 and phenylmethylsulfonyl fluoride [PMSF]) were added to a final concentration of 2 mM. Reactions were stopped by adding 3 volumes of 100% methanol, and reaction mixtures were centrifuged through Microcon spin columns with a 3,000-Da cutoff (Millipore) at 10,000 \times g for 10 min at 4°C to remove proteins that might interfere with mass spectrometric analysis. Peptide samples (<3,000 Da) were desalted using a ZipTip C₁₈ reverse-phase microtip (Millipore). Mass spectrometry of the peptides was performed in reflector mode on a Reflex IV mass spectrometer (Bruker, Germany) with a laser power of 15 to 20% in the presence of 10 mg/ml of α -cyano 4-hydroxycinnamic acid matrix and peptide standards (Bruker, Germany).

Signal peptides	
EhJessie1	<u>MTLIIFLIISLSFSL</u>
EiJessie1a	<u>MLTYFEAILFCLFVVC</u> TAAKNN
EiJessie1b	<u>MNSFLLLLVVS</u> AFSQTYE
EiJessie1c	<u>MFLVVLTFIATALS</u> SIGP
EiJessie3a	<u>MKITFIVLCLLCLALS</u> SLAFNKTAQRTN
EiJessie3b	<u>MNRAIITLLFICAALS</u> RELNRGFKNITRTATRIN
EhJessie3	<u>MKLAVLTVTIFMTLSY</u> GLNITFSQRTN
8-Cys chitin-binding domains	
EhJessie1	<u>CTNVKKG</u> FY <u>CLDRSKFVW</u> <u>CSG</u> -- <u>TNQSM</u> AIT <u>CFK</u> ETV <u>CK</u> <u>CG</u> KTKYNP <u>CV</u> FSFQEL- <u>DD</u> <u>CE</u> GLPGDI INEP
EiJessie1a	<u>CSRLEK</u> AAY <u>CLDRTRFVW</u> <u>CSG</u> -- <u>NGHG</u> IGMK <u>CDE</u> ETV <u>CK</u> <u>CG</u> KTYGYP <u>CV</u> FTFQEL- <u>PD</u> <u>CT</u> GLPGDYISDE
EiJessie1b	<u>CYNLTY</u> GIFY <u>CY</u> DKSIF <u>MH</u> <u>CR</u> GEN-AKAV <u>WLK</u> <u>CR</u> GGTV <u>CK</u> <u>CG</u> RTTYNP <u>CV</u> FVDYNDV- <u>DD</u> <u>CE</u> GKPGSYLNEK
EiJessie1c	<u>CDNLT</u> LGGY <u>CY</u> DKSIFLY <u>CP</u> GSDFVES <u>SWR</u> <u>CK</u> GGSI <u>CK</u> <u>CG</u> KTVSNP <u>CAF</u> DYSDV- <u>DF</u> <u>CN</u> GRVGTYINSS
EiJessie3a	<u>CDGLD</u> VGFY <u>CVDQ</u> TKYNW <u>CY</u> G-- <u>QSLY</u> KTTD <u>CP</u> AGLV <u>CK</u> <u>CG</u> FTIYNP <u>CA</u> WPWSDL-AI <u>CD</u> GEPGKYFGDT
EiJessie3b	<u>CNGLD</u> IGFY <u>CVDK</u> NTYNW <u>CF</u> G-- <u>QSAY</u> RSTS <u>CP</u> AGLE <u>CK</u> <u>CG</u> FTTNP <u>CA</u> WSYQDLGNM <u>CV</u> GKPGDYFVDE
EhJessie3	<u>CDGL</u> ESGFY <u>CVD</u> STYYW <u>CY</u> G-- <u>QSRF</u> KQAT <u>CP</u> NGLE <u>CK</u> <u>CG</u> KTIYNP <u>CM</u> WPFQGNTPD <u>CD</u> GKPGKYFSGT
Ser- and Thr-rich low complexity spacers	
EhJessie1	SKFYENYK*
EiJessie1a	V*
EiJessie1b	SLTPPEIIEEESNTNKPTTQSTEKKPSAAFSYLLYSVLLFALI*
EiJessie1c	IVTPPETFANDKLYSSQAQEVNESSSLGVMNDGTTIASVLFVVLVYVLL*
EiJessie3a	PVEPESSEPTPPQESSDTPIPPESSETPLPPTPESSDNPILPSSNEEP
EiJessie3b	EPEIPSEEPSKEPSSGNDDFSSENIPSEEPGSDS
EhJessie3	TPEEPSESSTVNPPESSSIPSPDSSSTFVSSSSHTPEPSESSSQKSESSTSSQSSNNSGDD
conserved domain like that of a putative <i>Clostridium</i> chitin-binding protein	
EiJessie3a	YDPDWPDVEAGVYSMTPTTHLPLVLKFEKSNWQQLKEVLSYKDYNNKDLKPTDSTYECSLHPPAKYDT
EiJessie3b	YNPEWPDVEDGVYSMKPSAHLPLVLKFDKDNWQEQIKEVVKGENYNYNEKLYTPVDSNYECALHPPKNHND
EhJessie3	YNPDWPDVEKGVYSMAFNTHLPLVLKFDKKNKWEQIKEVVSNSNYNNEALYQPATSDYACALHPPSKYDN
EiJessie3a	<u>NPTQI</u> WIGRPSETVT-FSYTPGVAIRLPDKYGLFLGYAMDAYGLNPGMLVGLGAKESFSFSRFAANDDG
EiJessie3b	<u>NPTQM</u> WIGRPSETVT-ISYTPGVAVRLPDKYGLFLGYAMDAYGLNPGMLIGLGAKESFSFTRFDATDDG
EhJessie3	<u>NPTQI</u> WIGRPSQSVT-ISYTPGVAIRLPDKYGMFLGYAMDAYGLNPGMLIGLGAKESFSFARYQATDDG
<i>Clostridium</i>	^DTQLIIPMPKKMLVPIINGVNLRIWMPKSKYLAMGIGTGTEYFGLNPNFMVGLSIKENFTCGLAPLESKY
EiJessie3a	<u>SYFIVE</u> KEDEHYDCYSNSQRGLCRDGNLDGPFQVETGGMATDVAIILPNRFYIGDETIPKDKRRIKLYMYDN
EiJessie3b	<u>SYFIVE</u> KEDEHYDCYSNSQRGLCRDGNLDGPFQVETGGMSTDVAIILPNRFYVGDNNTSKDKRRIKLYMFDN
EhJessie3	<u>SYFIV</u> ANENEHYDCYSSQRGLCRDGNLDGPFQVETGGMSTDVAIILPNRFWIGESTTPKSERKIKYMFND
<i>Clostridium</i>	TENIVTVDGQKWSWP-----IQKKHPDGPFQQEKGNFNEIKKQYFD--YLPDS--AEHENYVTLTKTGE
EiJessie3a	<u>EVLTS</u> AGFREYHDFFTLTPGRAFILSSLDHFHRHNLVMMKMKLGLGEAMAKRKTREARDSLEFATAMYTY
EiJessie3b	<u>EVLTS</u> AGFREYHDYYTLNVGRAFVLSLDFHFRHNLVMMKMKIGLGDAMKKRRTREARDSLEFATAMYTY
EhJessie3	<u>EVLTY</u> AGFRAYHDYFTLNYGRAFVLTSLDFHFRHNLVMMKMKVGLRDALSKRKTREQRDSLEFATAMYTY
<i>Clostridium</i>	PDDPSYVHAAMSSYSLTMTREFLYAIPNNDFSG--VLKDAKDPWAEFVLVDNAYNRGVYGLLQKKLTFE
EiJessie3a	<u>NRGV</u> FDTQLISMLGTCDANMDPCLDCKLDGYYGGHSTDIRVCKAVDSAPGEEVYDYDLSKEDVEYFLEIL
EiJessie3b	<u>NRGV</u> FDTQLIAMLGNCDADMDPCLDCKLDGYYGGHTDIRTVCKVVDVSDAPNEEYDYELKKEVDVYFVDIL
EhJessie3	<u>NRGV</u> FDTQLIQMLNGCNADMDPCTDCKLDGYYGGHTDIRVCKAVDSVFNKEVYDYDLSKEDVEYFISTII
<i>Clostridium</i>	HR----DKLINSF-----DINKEFNLSGFANHIENIQNVIKAMDSET-ESFYDANITWDDMENYFKEI
EiJessie3a	ESTLP---FDNVDWKTLRNDVDEAYDMLKTARGGKYISFRYDWRSLAVVRMHLPAIEYFVGDQVQKSFQN
EiJessie3b	ETTFP---FDNVDAQVVRTDVKQAYQFLSSMRKKDTISFRYDWRALLAVVRMHLPAIEYFVGEQVQKSFQN
EhJessie3	QSTYP---FNNVDWKTLSRDVDTAYDYLSQRSNSNAISFRYDWRLLAVIRMHLPPIEYFVGDQVQKTFQD
<i>Clostridium</i>	RLYYGRNIPNDADWNIMKADVKSYSYDILSKHWGGDHSILRYDFLTLRLVCEKHLPENKQPGSPGSWIEQ
EiJessie3a	<u>YWG</u> DNLNADLGPYQNVKDFPFTFCATAGGRKNLNCNAKSSFFVQHMYNTINK*
EiJessie3b	<u>YWG</u> SDLNTDLGPYQNVTPFPYSPFCSTAGGRKNLCK*
EhJessie3	<u>YWG</u> VLNDDLGPYQSVSNFPFKFCSTSGGRSNLNCSSKSHKQVHYSK*
<i>Clostridium</i>	VNSANNIH*

FIG. 2. Primary sequences of *E. histolytica* and *E. invadens* Jessie lectins aligned with each other and with that of a predicted *Clostridium botulinum* protein. Sequences are marked as described in the legend to Fig. 1. The caret indicates that the *C. botulinum* protein, which is complete in the database, has been truncated here. A putative N-terminal CBD in the *C. botulinum* protein is not shown.

highly conserved spacing of Cys residues as well as conserved aromatic residues that may participate in chitin binding (34). Each Jacob sequence predicted a signal peptide, while none of the Jacob lectins (or Jessie lectins and chitinases [described below]) had transmembrane helices (14). Four *E. invadens* Jacob lectins (EiJacob1 to EiJacob4) had five closely spaced six-Cys CBDs, while EiJacob5 had just three CBDs (Fig. 1A). Between the CBDs of these Jacob lectins were short, low-

complexity, Ser- and Thr-rich spacers. Some of these spacers contained conserved sequences (e.g., TPSVDKS), which are the sites of endoprotease cleavage (see the proteomic data below).

Two novel Jacob lectins (EiJacob6 and EiJacob7) contained much longer low-complexity sequences between CBDs (Fig. 1B). These low-complexity sequences were rich in Ser, Glu, Lys, and His and contained several types of short internal

	<u>Signal peptides</u>
Eichitinase1	MSAMAFVCFVLALANSQT
Eichitinase2	MKSVTFIVILMLCDITTFGLVLRKREKRMFELHTEKNKHTEMSEYTE
Eichitinase3	MRIILLACMLMVSMAEQKE
	<u>8-Cys chitin-binding domain</u>
Eichitinase1	CEGLDNGFYCIDDDTSYLWCYGHQTQQQITKCSSEGLVCCKCGKTAYTPCVWSWTDLPDCSLKPGDYFD
	<u>low complexity Ser- and Thr-rich sequence</u>
Eichitinase1	KSSEEPLEPESSEPEPEDESSEVKPEPPAESSESSEKPKPESESSEQPKPESESSEKPKPESESSEESSESGF
	<u>Glycohydrolase 18 domain</u>
Eichitinase1	KKVVSYYTNNWAQYRQNSIDGWACKYTPDNIDPTLVVDVINYAFVVFSSSYTVKEYEWNDDQMIKPIVAMKSKNPNL
Eichitinase2	RKVVGYTNNWSQYRTATVGGWPSKFTPERIDAKLFDVLFNFAFVTFDDTYEVKEYEWNNDVMIQPLVSLKSHNPNL
Eichitinase3	YKVIQYFSAWAQYRNGQLGKSNFQFLPKHIDPTQYDVLNYAFVGDKNQQTIFLDENDGTTIPQVINLKSQNSKL
Eichitinase1	QVLASIGGWNFNFYDSTKHLFSEMAEKQTSRAAFIKSAMSFAKYNLGDIDWEYPANKDQGGRPVDVQSFILL
Eichitinase2	KVCVSIIGGWNFNANAKTKHLFHDMAATSVSRKRFIKSAIQFARKYNMDDGIDWEYPNGELQGGHPTDFLNFNCL
Eichitinase3	KIFASIGGWDFSSSEKATRGIFSKLSKSREERAARAFIRNAVSVFCRRYDFDGDIDWEYPGDEDNEGSKEVDVNNFAAL
Eichitinase1	LKEFREAIDAEEKLSGGRSRLLLTIAAPAGPKNIENLEISKFHLYLDWINLMTYDLHGSWDDVTGSHTALYADDAL
Eichitinase2	VKEFREFEVIDSKEENKEPLLITVAAPAGIQNIQNIQNDVETVSKYIDWFLNMTYDLHGAWDNITGPHTALYSTDGL
Eichitinase3	MKEFYEYKISEAQSSGKAEIILSAATSSSKEFYGNKYLSETSKYVDFWVNVMTYDHDSDDDRVRNSHTPLKSLNSG
Eichitinase1	S--VDDAVKAYLSQGVPPAKMFVGMHYGRGWTLS-ANKHDMGSPASGASTSGKCTGENGYLSKYEIDSLIPAA
Eichitinase2	S--VNDCVSEYLSKSGVPSHKLVGLLAHYARGWTVNNNTTEHKMGEEASGPFSECGRCTGENGYLAKYEIDEIIPKN
Eichitinase3	SESIETTINDYVSGVPLEKIILGIALYGRGWSLET-IVNHNIGDKAIGPSKPGAYTQEGELMANIETIETRT--V
Eichitinase1	NIQYDAKSMTFGYSQDQWFSFDDKDTFEHKVQYVCDHKLGGAMIWSLDDLKKNYVNTKFIKKELDQC*
Eichitinase2	NIEFFDDTSKTLFGYFDNKFYTFDDKETFKMKVEYAKEKSLGGVMMWSVDLDDKDFVNTKYVRSVLASD*
Eichitinase3	NERYDLTAVCKYGWYSNQWFSFDDDEITIKLKAKYAKERQLGGVMFWAIDLDPPEMRYIKPVIDELKN*

FIG. 3. Primary sequences of *E. invadens* chitinases with Cys-rich domains and tryptic peptides, marked as described in the legend to Fig. 1. While *E. invadens* chitinase 1 has an eight-Cys CBD, this domain is missing from *E. invadens* chitinase 2 and chitinase 3.

repeats. EiJacob6 contained four predicted endoprotease sites between six tandemly arranged CBDs. The low-complexity sequences of EiJacob7, which were flanked by single CBDs, contained internal repeats that included eight predicted endoprotease sites. The long low-complexity sequences of these *E. invadens* Jacob lectins resemble those predicted for a second *E. histolytica* Jacob lectin that contains three CBDs (15, 29; unpublished data).

The *E. invadens* genome predicted five Jessie lectins, three of which had a signal peptide, an N-terminal eight-Cys lectin domain, and little else (EiJessie1a to EiJessie1c) (Fig. 2). This was similar to the case for EhJessie1, EhJessie2, and a fourth predicted *E. histolytica* Jessie lectin (29; unpublished data). The other *E. invadens* Jessie lectins (EiJessie3a and -3b) had an N-terminal eight-Cys CBD, a Ser- and Thr-rich low-complexity spacer, and a conserved C-terminal domain, which is also present in EhJessie3 (29). The EhJessie3 C-terminal domain shares 17% identity with a hypothetical *Clostridium botulinum* protein predicted from genomic sequence data (1, 13, 29). The N terminus of this *C. botulinum* protein, which does not share significant sequence homology with EhJessie3, is likely a bacterial CBD that is also present in a putative glycosyl hydrolase from *Clavibacter michiganensis* and in a chitinase and chitodextrinase from *Microbulbifer degradans* (13; data not shown). These results suggested the likelihood that the unknown domain common to *Entamoeba* Jessie3 lectins and the putative *C. botulinum* protein is an enzyme involved in modifying chitin.

The *E. invadens* genome predicted two additional chitinases (*E. invadens* chitinase 2 and *E. invadens* chitinase 3), each of which contained a signal peptide but was lacking the N-terminal eight-Cys CBD present in *E. invadens* chitinase 1 (Fig. 3) (6, 29). Multiple chitinase activities have previously been iden-

tified in encysting *E. invadens* (6). The presence of the N-terminal eight-Cys CBD in *E. invadens* chitinase 1 is likely important because *E. invadens* chitinase 1 was abundant in the *E. invadens* cyst wall, while *E. invadens* chitinase 2 and chitinase 3 were not detected (see the proteomic data below).

mRNA levels for most *E. invadens* Jacob lectins, Jessie lectins, and chitinases increased during encystation. To determine if mRNA levels for the predicted *E. invadens* Jacobs, Jessies, and chitinases increase during encystation, as previously shown for EiJacob1 and *E. invadens* chitinase 1 (6, 8), RT-PCR was performed using RNAs isolated from parasites collected 0 h and 22 h after transfer into encystation medium (Fig. 4). After 22 h in encystation medium, mRNA levels for most, but not all, of the Jacob lectins, Jessie lectins, and chitinases appeared higher than they were at 0 h, while mRNA levels for the housekeeping proteins actin and elongation factor 1 α (EF1 α) remained the same. No quantification of these changes was attempted, because Jacob lectins, Jessie lectins, and chitinases were directly identified in the cyst walls by proteomic methods. Low levels of some Jacob, Jessie, and chitinase mRNAs could be detected at 0 h, most likely because some *E. invadens* trophozoites spontaneously encyst in regular culture medium. These results suggested that Jacob, Jessie, and chitinase genes are expressed during encystation but did not determine the fate of the encoded proteins, which is addressed below.

Abundant *E. invadens* cyst wall proteins all contained chitin-binding domains. Proteins in purified cyst walls of *E. invadens* trophozoites were analyzed by mass spectrometry of tryptic peptides and by N-terminal sequencing of spots isolated on 2-D protein gels. Multiple repeats of mass spectrometry identified six of seven predicted Jacob lectins, some with many peptides (EiJacob1 to EiJacob3) (Fig. 1 and Table 2) (9).

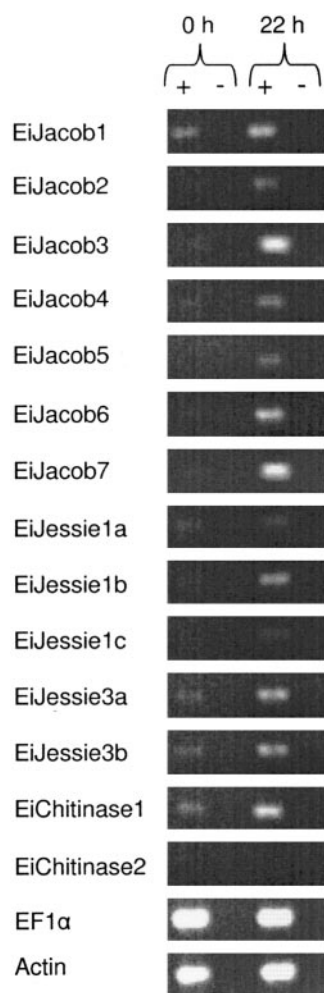


FIG. 4. RT-PCR analysis of Jacob and Jessie lectin expression in encysting *E. invadens*. Total RNA for RT-PCR analysis was isolated 0 and 22 h after transfer of cells to encystation medium. + and –, presence and absence, respectively, of reverse transcriptase in the first-strand cDNA synthesis reaction. EF1 α and actin were examined as controls. These results are qualitative rather than quantitative.

EiJacob1 to EiJacob3, each of which contains five six-Cys CBDs and short Ser- and Thr-rich spacers, also accounted for most of the spots on 2-D gels containing *E. invadens* cyst wall proteins (Fig. 5 and see below). EiJacob6 and EiJacob7, which contain long low-complexity sequences, were much less abundant (Fig. 1 and Table 2). EiJacob5 was not identified.

Mass spectrometry of *E. invadens* cyst walls identified (with many peptides) EiJessie3a and -3b, each of which has an N-terminal eight-Cys CBD and a C-terminal domain like that of a *C. botulinum* protein (Fig. 2 and Table 2) (13, 29). None of the Jessies with a single CBD (EiJessie1a to EiJessie1c) were identified. A partial explanation for this result may be the low levels mRNA expression for Jessie1a and Jessie1c (Fig. 4). However, we could not rule out the possibility that EiJessie1a to EiJessie1c (and *E. invadens* chitinase 2 and chitinase 3 [see below]) are much less abundant in the *E. invadens* cyst wall and thus were not identified by mass spectrometry.

E. invadens chitinase 1, which contains an N-terminal eight-

Cys CBD, was identified with many peptides, while *E. invadens* chitinase 2 and chitinase 3 were absent from multiple mass spectroscopic studies of *E. invadens* cyst wall proteins (Fig. 3 and Table 2) (6). It is possible that *E. invadens* chitinase 2 and chitinase 3 are absent from the cyst wall, because they each lack an N-terminal CBD.

These results for chitinases, Jacobs, and Jessies showed that lectins with predicted CBDs comprise the vast majority of *E. invadens* cyst wall proteins (Table 2). Estimates of the relative abundance of each protein in the cyst walls (using peak heights of tryptic peptides) suggested that Jacob lectins, which are likely structural proteins that cross-link chitin fibrils, comprise ~30% of the cyst wall protein (Table 2). Chitinase and Jessie3 lectins, which may be enzymes involved in remodeling chitin, comprised ~70% of the cyst wall protein.

In contrast, we failed to identify >100 other secreted and membrane proteins, including 2 chitin synthases, 4 chitin deacetylases, 7 Gal/GalNAc lectin heavy and light subunits, Hsp70, 2 calreticulins, VIP36, 15 protein disulfide isomerases, 38 Cys proteases, 2 dipeptidyl proteases, 4 amylases, and 24 receptor kinases (31). Actin and actin-binding proteins (together comprising 7 to 10% of the total protein, depending on the run) were the major cytosolic contaminants in the *E. invadens* cyst wall preparations (Table 2). Because sequencing of the *E. invadens* genome is not complete, we cannot rule out the presence of other proteins in the *E. invadens* cyst wall for which genomic sequences are not available. Similarly, we cannot rule out the possibility that some of the chitin-binding lectins identified by mass spectrometry were present in secretory vesicles prior to disruption of cysts and then bound to the chitin in cyst walls during the purification procedure. It is likely that these chitin-binding proteins, which are encystation specific, were destined for the cyst wall anyway.

Jacob lectins were cleaved at conserved sites between chitin-binding domains by an endogenous Cys protease. *E. invadens* cyst wall proteins were separated in 2-D gels with a pH range of 4 to 7 in the first dimension. This gave better resolution of the large acidic proteins, where EiJacob1 had previously been identified by mass spectrometry (Fig. 5A) (9, 20). These 2-D gels resolved >50 *E. invadens* cyst wall proteins, which fell into two groups. Acidic proteins, which appeared to be the most abundant, formed elongated spots that varied in size from ~10 to ~100 kDa. Neutral proteins, which formed circular spots, varied from ~10 to ~65 kDa. Remarkably, mass spectrometry of an excised 100-kDa acidic spot and a 65-kDa neutral spot

TABLE 2. Composition of *E. invadens* cyst walls

Protein	Composition (%) ^a
EiJessie3a	22–30
EiChitinase1	20–26
EiJessie3b	18–21
EiJacob1.....	11–14
EiJacob3.....	7–9
EiJacob2.....	6
EiJacob6.....	2–4
EiJacob4.....	0.1–1
EiJacob7.....	0–0.1
Nonwall proteins.....	7–10

^a Range for two independent experiments. Nonwall proteins, which likely are contaminants, included actin and actin-binding proteins.

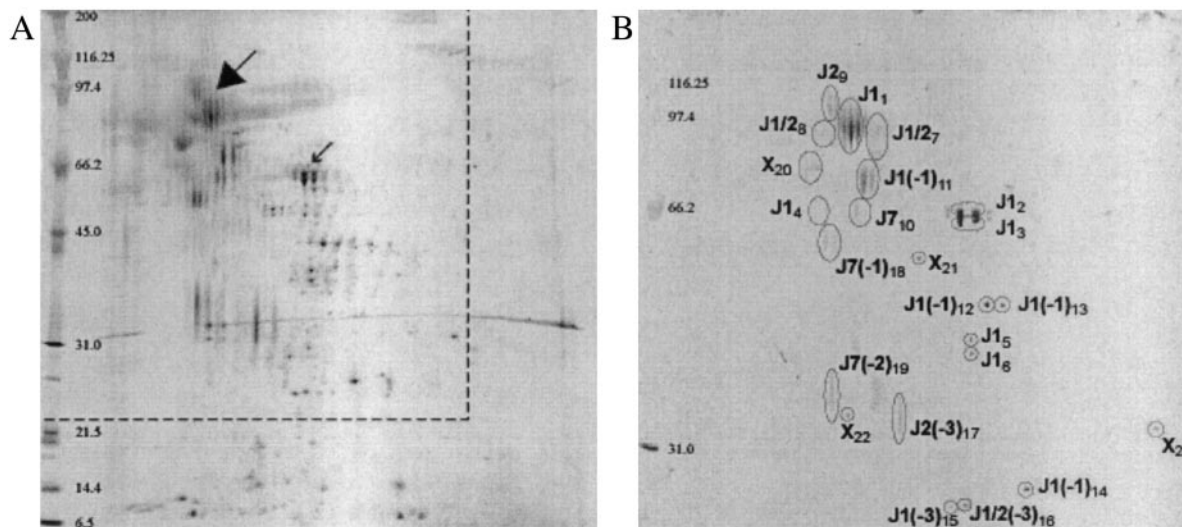


FIG. 5. 2-D gel separation of *E. invadens* cyst wall proteins for N-terminal sequencing. Two gels were run in parallel, using the same preparation of cyst wall proteins. The first gel (A) was stained with SYPRO Ruby, a fluorescent total protein stain. The second gel (B) was transferred to a polyvinylidene difluoride membrane and stained with Coomassie blue. The blotted area corresponds to the upper left quadrant of panel A (marked by the dotted line). Molecular weight markers used for SDS-PAGE appear on the left side of each gel. (B) The 23 circled spots were excised and subjected to N-terminal sequencing (see Fig. 1). Spots are labeled with their protein names (J1 = EiJacob1, J2 = EiJacob2, J3 = EiJacob3) or with an "X" if no N-terminal sequencing data could be obtained. Negative numbers in parentheses denote the numbers of CBDs missing from the N termini of proteins. Proteins which are not labeled with negative numbers were cleaved after their N-terminal signal sequences. The sample number of each spot appears as a subscript.

(large and small arrows, respectively, in Fig. 5A) showed that both were EiJacob1. We present evidence below that the differences in size and charge of the two EiJacob1s were secondary to differences in the abundances of O-P-glycans.

The N-terminal sequences of 19 *E. invadens* cyst wall proteins from 2-D gels revealed two remarkable findings (Fig. 5A and B). First, all of the N-terminal sequences obtained could be mapped to only three different Jacob lectins (Fig. 1 and 5B) (9). Eleven sequences matched EiJacob1, two matched EiJacob2, three matched EiJacob3, and three matched both EiJacob1 and EiJacob2, whose sequences resemble each other. It is not clear why we failed to identify EiJessie3 lectins and *E. invadens* chitinase 1, which were very abundant by mass spectrometry of tryptic peptides (Table 2).

Second, while *E. invadens* Jacob lectins were cleaved after the signal peptide, EiJacob1 to EiJacob3 were also cleaved at conserved sites in Ser- and Thr-rich spacers between CBDs (Fig. 1 and 5B). EiJacob1 was cleaved between its first and second CBDs (Fig. 5B, spots 11 to 14) and between its third and fourth CBDs (spot 15). EiJacob2 was cleaved between its third and fourth CBDs (spot 17). EiJacob3 was cleaved between its first and second CBDs (spot 18) and between its second and third CBDs (spot 19). The differences in the apparent molecular weights of groups of EiJacob1 spots with the same N-terminal sequence were likely due to missing C-terminal sequences in the smaller spots (e.g., compare spots 2 and 3 with spots 5 and 6 or spots 12 and 13 with spot 14).

The cleavages in the Jacob lectins occurred prior to Lys at sites resembling TPSVDK. Additional predicted sites for the endoprotease were also present in EiJacob2, EiJacob6, and EiJacob7 (Fig. 1). These results suggested that Jacob lectins, which contain as many as seven CBDs, may be cleaved at one

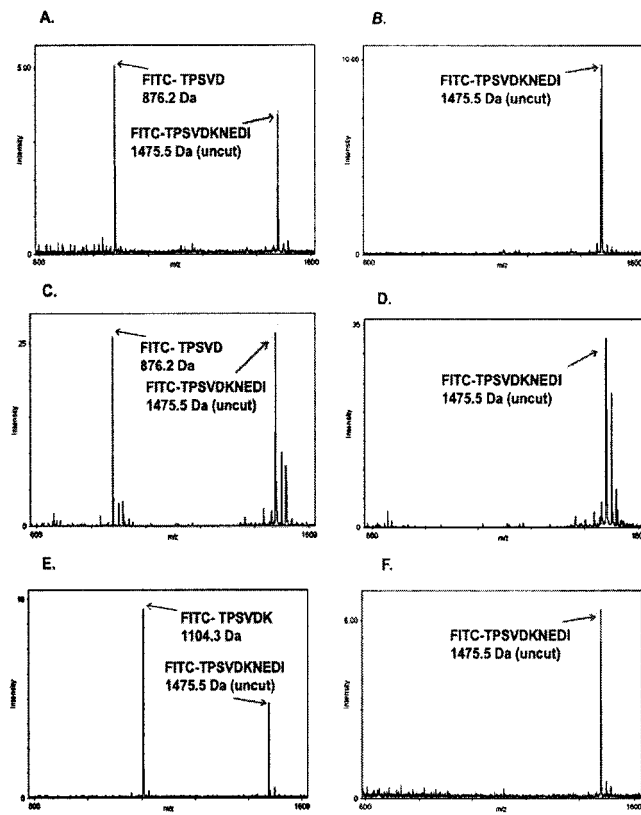


FIG. 6. Mass spectrometry of a model peptide (FITC-TPSVD KNEDI) incubated with an *E. invadens* lysate in the absence (A) or presence (B) of the Cys protease inhibitor E-64. The same peptide was incubated with papain in the absence (C) or presence (D) of E-64 or with trypsin in the absence (E) or presence (F) of the Ser protease inhibitor PMSF.

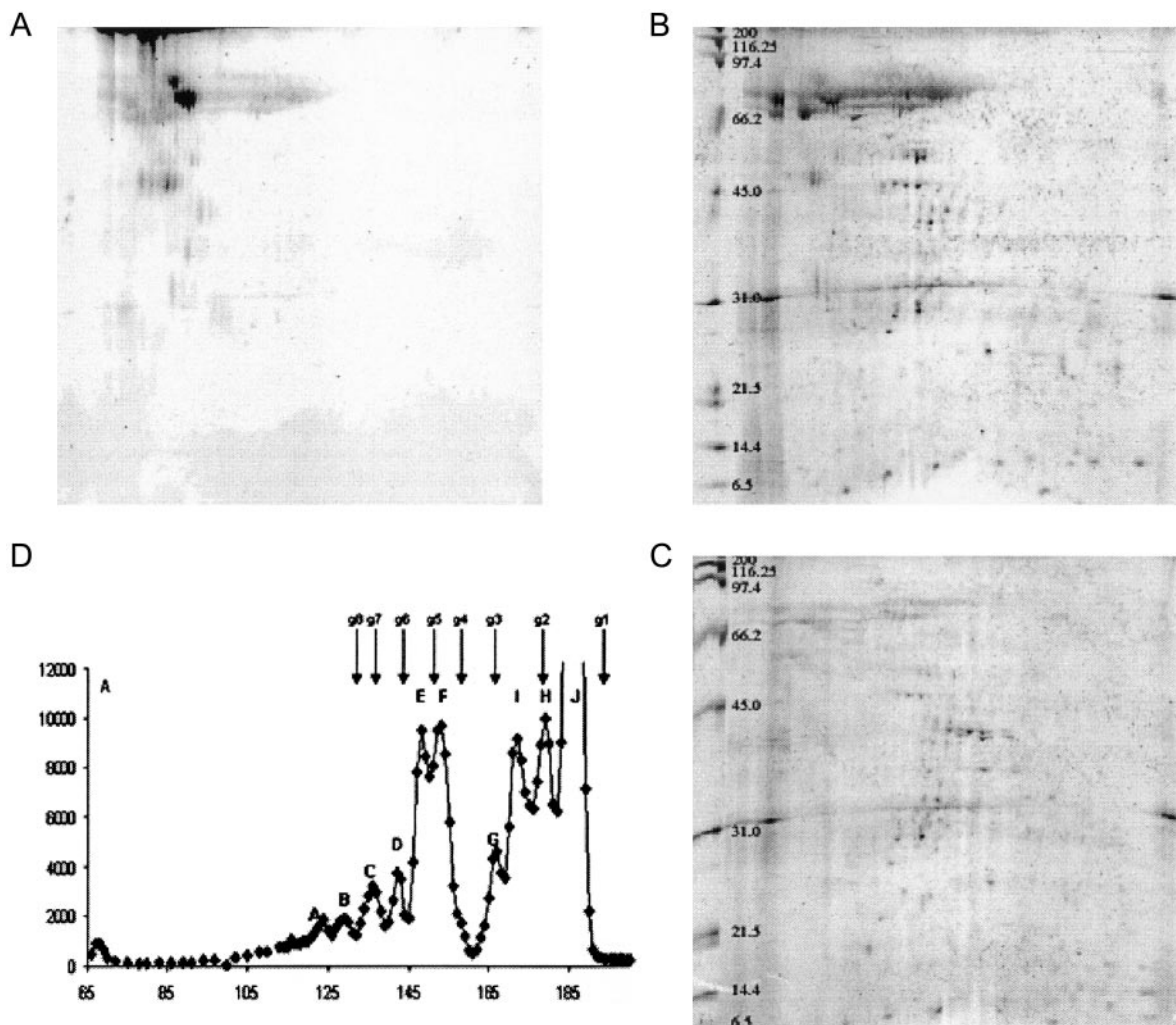


FIG. 7. Detection and analysis of glycans of *E. invadens* cyst wall proteins separated in 2-D gels. (A) Untreated cyst wall proteins visualized at 488 nm after being stained with Pro-Q Emerald 488 glycoprotein stain. (B) Same cyst wall protein gel as that shown in panel A, but visualized at 532 nm after SYPRO Ruby staining, which stains all proteins. (C) *E. invadens* cyst wall proteins were treated with 40 mM TFA for 10 min at 100°C to remove *O*-P-glycans and were visualized at 532 nm after SYPRO Ruby staining. No staining of TFA-treated cyst wall proteins was observed at 488 nm after glycoprotein staining (data not shown). (D) BioGel P4 chromatography of cyst wall *O*-P-glycans released by mild acid hydrolysis. Glycans were reductively labeled with NaB^3H_4 , so the y axis shows counts per min. The arrows indicate the elution positions of glucose oligosaccharide standards (linear β -1,6-glucosyl glucosides, one to eight units long). Peaks A to D are further characterized in Fig. 8.

or more conserved endopeptidase sites within the Ser- and Thr-rich spacers to make processed lectins with fewer CBDs. This cleavage likely did not occur during the purification of cyst walls, as the Cys protease inhibitor E-64 was added to cysts prior to sonication.

A model Jacob peptide (FITC-TPSVDKNEDI) which contained the conserved endoprotease site prior to Lys was cleaved at this site by an *E. invadens* lysate (Fig. 6A). This cleavage was inhibited by E-64 (Fig. 6B), strongly suggesting that Cys proteases are involved in site-specific cleavage (4, 23). Papain (a plant Cys protease) also cleaved the Jacob peptide at the same site as the amebic lysate and was inhibited by E-64 (Fig. 6C and D), while trypsin (a bovine Ser protease) cleaved the Jacob peptide after the Lys and was inhibited by PMSF (Fig. 6E and F). The latter result suggested that Ser proteases,

which have been implicated in *E. invadens* encystation, likely do not cleave Jacob lectins (21).

Jacob lectins were extensively modified by *O*-phosphodiester-linked glycans, which had deoxysugars at the reducing ends. Previous lectin-binding experiments showed that the acidic, high-molecular-weight EiJacob1 lectin is a glycoprotein (9). The Pro-Q Emerald 488 glycoprotein stain, which reacts with periodic acid-oxidized carbohydrate groups (11), stained the acidic proteins (including all three Jacob lectins) in 2-D gels of *E. invadens* cyst wall proteins but did not stain the more neutral EiJacob1 protein spots in the center of the gel (Fig. 7A). Subsequent staining of the gel with SYPRO Ruby showed that these previously unstained proteins were indeed present in the gel (Fig. 7B). The Pro-Q Emerald 488-stained glycans on Jacob lectins were most likely *O*-P-glycans, because the high-molec-

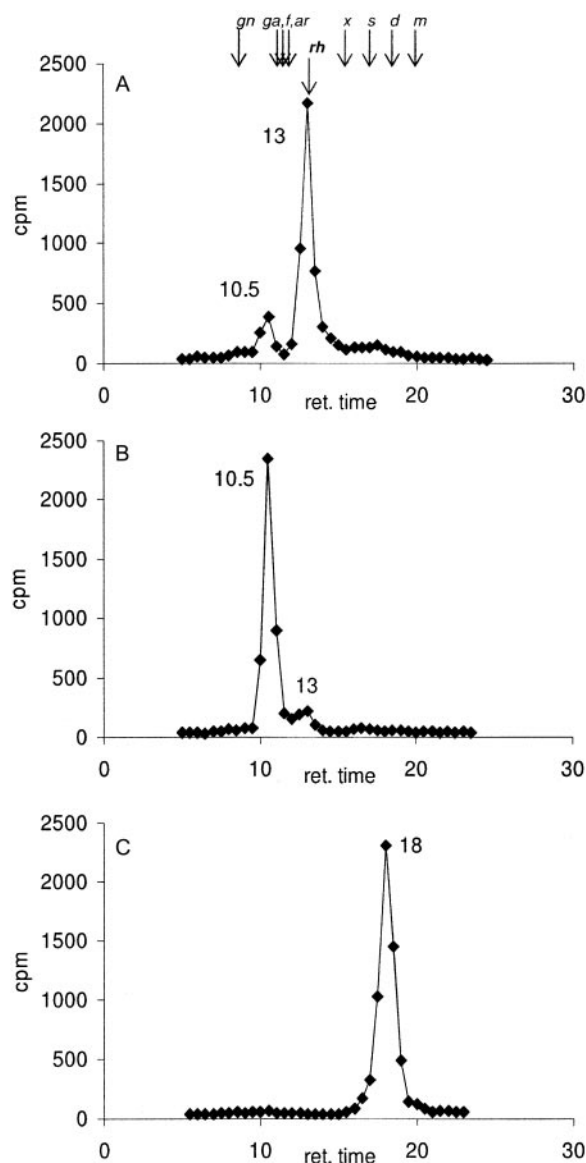


FIG. 8. Characterization of reduced sugars at ends of *O*-P-glycans of *E. invadens* cyst walls. The *O*-P-glycans shown in Fig. 7D were isolated from P4 columns, reduced with NaB^3H_4 , hydrolyzed, and separated by a CarboPac MA1 column run. Alditol standards included glucosaminitol (gn), galactosaminitol (ga), fucitol (f), arabinitol (ar), rhamnitol (rh), xylitol (x), glucitol (s), galactitol (d), and mannitol (m). (A) The reduced ends of the *O*-P-glycans in peaks A and B in Fig. 7D ran with a rhamnitol standard at 13 min. (B) The reduced ends of the *O*-P-glycans in peaks C and D did not run with a standard, but mass spectrometry of intact peak D showed a deoxysugar at the reducing end. (C) As a control, the reduced sugar from *E. histolytica* proteo-phosphoglycans ran with the galactitol standard (16).

ular-weight, acidic Jacob spots were nearly absent from 2-D gels containing *E. invadens* cyst wall proteins pretreated with 40% TFA under conditions that hydrolyze phosphodiester linkages (Fig. 7C) (17, 18). In contrast, the more neutral Jacob spots, which were not stained by Pro-Q Emerald 488, were not affected by mild acid treatment.

E. invadens cyst wall *O*-P-glycans, which were removed by TFA and labeled at their reducing ends with borotritide, were a het-

erogeneous mix (Fig. 7D). Total hydrolysis of these glycans followed by monosaccharide analysis of the alditols at the free ends showed two types of reduced sugars, eluting at 10.5 and 13 min. These alditols, which differed in their relative abundances depending on the particular peak analyzed (Fig. 8A and B), may be deoxysugars, based upon the following evidence. First, these alditols were neutral, as defined by their lack of affinity for ion-exchange resins. Second, the reduced sugar eluting at 13 min coeluted with an internal rhamnitol (6-deoxy-mannitol) standard. Third, although the reduced sugar eluting at 10.5 min did not comigrate with any of the available standards, mass spectrometry showed that peak D in Fig. 7D is composed of four hexoses and one deoxyhexose. Fourth, *Entamoeba* has at least two genes encoding putative nucleotide-sugar 4,6-dehydratases, which are involved in the synthesis of deoxysugars in the secretory pathway (data not shown). The alditols derived from the *E. invadens* cyst wall were also distinct from galactitol, which is the most abundant reduced sugar of *O*-P-glycans of the proteo-phosphoglycans of *E. histolytica* trophozoites (Fig. 8C) (18). While these results suggested that the reducing ends of cyst wall *O*-P-glycans are deoxysugars, further analysis will need to be performed to confirm their identities. Monosaccharides in peak J of Fig. 7D, which included glucitol and mannitol, were likely contaminants of *E. invadens* cyst wall purification.

The locations of a small, but likely representative, number of *O*-P-glycans on Jacob lectins were determined by mass spectrometric identification of phosphopeptides from cyst wall proteins treated with TFA to remove glycans (Fig. 1 and 9). Phosphorylated peptides were readily identified by their unique fragmentation patterns. Because the phosphorylated peptides were present in the TFA-treated samples but not in the untreated samples, each phosphate must have been present originally as an *O*-P-glycan. It is likely, however, that numerous phosphopeptides were missed because (i) there was limited sampling of peptides from the Ser- and Thr-rich spacers between CBDs and (ii) no attempts were made to enrich phosphopeptides prior to mass spectroscopy.

DISCUSSION

The results reported here suggest a model of the *E. invadens* cyst wall in which all of the abundant proteins are unique, encystation-specific lectins with either six-Cys CBDs (Jacob lectins) or an eight-Cys CBD (*E. invadens* chitinase 1, EiJessie3a, and EiJessie3b) (Fig. 10). The Jacob lectins, which show extensive gene duplication, have tandem arrays of CBDs that cross-link chitin fibrils and may protect chitin from glycohydrolases. The most abundant Jacob lectins (EiJacob1 to EiJacob3) also have Ser- and Thr-rich spacers between CBDs which undergo posttranslational modifications (discussed below). The less abundant Jacob lectins (EiJacob6 and EiJacob7) contain long, low-complexity sequences between CBDs. Insect peritrophins, which cross-link chitin fibrils in the wall around the blood meal, have tandem arrays of eight-Cys CBDs that are separated by low-complexity spacers (26). The spore coat proteins of *Dictyostelium* also have tandem arrays of cellulose-binding domains that are separated by low-complexity spacers (32). While *Saccharomyces* wall proteins have Ser- and Thr-rich domains that are extensively glycosylated, fungal wall pro-

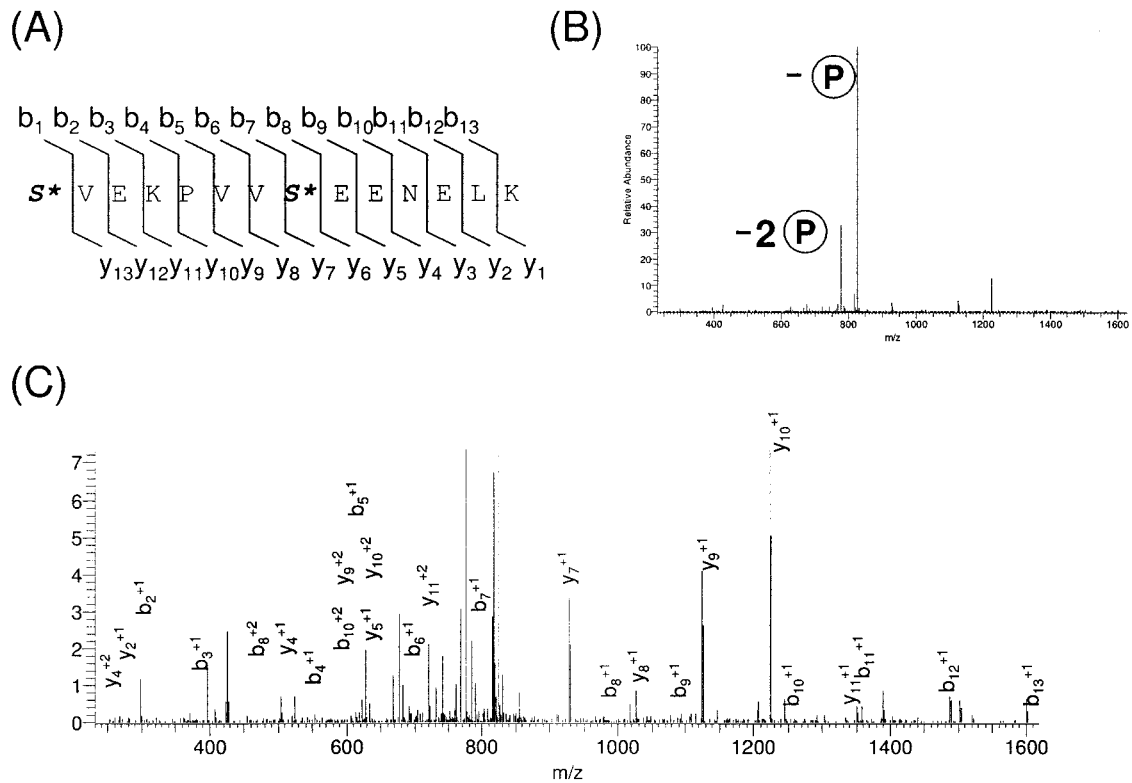


FIG. 9. Identification of phosphopeptides from TFA-treated cyst wall proteins. (A) Sample tryptic peptide from EiJacob1 lectin with corresponding MS/MS cleavage pattern revealing y- and b-ion fragments (*, sites of phosphorylation). (B) MS/MS spectrum revealing neutral losses of 98 and 196 Da, corresponding to losses of one and two phosphates, respectively. (C) Suppressed peaks revealing y- and b-ion peptide fragments for identification of the peptide sequence by SEQUEST or X!Tandem.

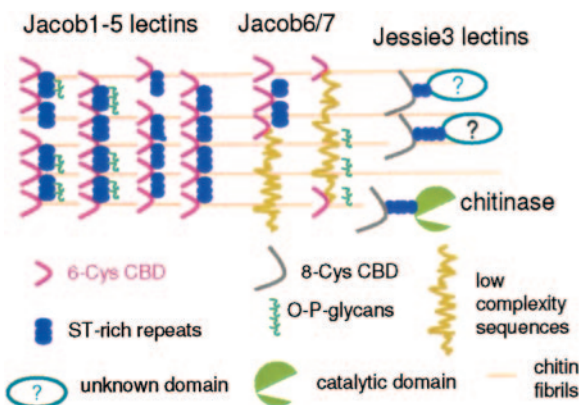


FIG. 10. Best model of *E. invadens* cyst wall. Wall proteins are restricted to those that have six-Cys CBDs (Jacob lectins) or eight-Cys CBDs (Jessie3 lectins and chitinases). Multiple Jacob lectins vary in their numbers of six-Cys CBDs, secondary to differences in the numbers of CBDs encoded and to cleavages by Cys proteases in Ser- and Thr-rich spacers. Jacob lectins also vary in the presence or absence of O-P-glycans and long, low-complexity sequences. An N-terminal CBD is present in a single chitinase and in two Jessie3 lectins, which have a C-terminal domain of unknown function. Because chitinases and Jessie3 lectins were not identified on 2-D gels, it is not known whether these proteins are cleaved by proteases.

teins have not been shown to be lectins and are covalently bound to chitin and β -1,3-glucan (36).

We were surprised by the large amounts of *E. invadens* chitinase 1, EiJessie3a, and EiJessie3b, which comprise $\sim 70\%$ of the total protein in the *E. invadens* cyst wall. These proteins likely do not cross-link chitin fibrils, because each has only a single eight-Cys CBD (29). Instead, *E. invadens* chitinase 1 may be involved in remodeling the chitin fibrils during cyst wall formation and/or degrading chitin fibrils during excystation. As discussed above, the unknown domain common to *Entamoeba* Jessie3s and the putative *C. botulinum* protein is also likely involved in modifying chitin (13). These results suggest that the walls of *E. invadens* cysts, like those of fungi, may be dynamic rather than static. *Saccharomyces* has numerous wall proteins (e.g., Bgl2 and Gas1) which are enzymes involved in remodeling β -1,3-glucans (36).

The simple model of the *E. invadens* cyst wall is made somewhat more complicated by posttranslational modifications of *E. invadens* Jacob lectins (Fig. 10). Cleavage of Jacob lectins between CBDs, which was highly reproducible, occurred at conserved sites located within otherwise low-complexity Ser- and Thr-rich sequences (Fig. 1, 5, and 6). Studies with a model synthetic peptide strongly suggested that Jacob lectins are cleaved by endogenous Cys proteases, which have previously been shown to be involved with encystation and amebic invasion into host tissues (4, 10, 23, 25). Specific cleavages between CBDs of Jacob lectins may be a mechanism of regulating the

thickness of the chitin wall during cyst construction and/or facilitate excystation. *Giardia* has a developmentally regulated Cys protease which is required for the proteolytic processing of a major cyst wall protein (28). A second developmentally regulated Cys protease is required for the excystation of *Giardia*. Future studies will determine whether there are encystation- or excystation-specific Cys proteases in *Entamoeba* and whether any of these proteases have CBDs, as shown for *E. invadens* chitinase 1 and EiJessie3s.

Jacob lectins in the *E. invadens* cyst wall had numerous O-P-glycans, which were released with TFA (Fig. 7 and 8). Why the same Jacobs appeared in heavily glycosylated forms and nonglycosylated forms is not clear, but the observation is remarkable. The O-P-glycans likely decorate Ser- and Thr-rich spacers between CBDs of Jacob lectins, in the same way that O-P-glycans are bound to Ser- and Thr-rich spacers between cellulose-binding domains of *Dictyostelium* spore coat glycoproteins (32). The *Dictyostelium* enzyme which adds O-P-glycans to Ser and Thr has not been characterized, but there are candidate enzymes that resemble bacterial enzymes that add O-P-glycans (33). Because *E. invadens* chitinases and Jessie lectins also contain Ser- and Thr-rich domains, it is likely that these cyst wall proteins also have O-P-glycans.

While the proteophosphoglycans of *E. histolytica* trophozoites have O-P-glycans terminating in glucose (18), the O-P-glycans on Jacob lectins were shorter and contained a deoxy-sugar at the reducing end rather than Gal. The O-P-glycans on Jacobs may protect cysts from degradation as they pass through the stomach and intestines. The cyst wall O-P-glycans may also be bound by the plasma membrane Gal/GalNAc lectin, as suggested by the formation of wall-less *E. invadens* cysts in the presence of excess Gal (9).

The *E. invadens* cyst wall model developed here has important implications for the pathogenesis of *E. histolytica* (Fig. 10) (10, 25). First, *E. histolytica* chitinases, Jacob lectins, and Jessie lectins, which have been shown to have chitin-binding activities in transfected *E. histolytica* (29), are likely abundant proteins in the *E. histolytica* cyst wall. Second, since *E. histolytica* chitinases, Jacobs, and Jessies also have Ser- and Thr-rich spacers between CBDs, these *E. histolytica* lectins are likely modified by O-P-glycans and may be cleaved by Cys proteases. Third, because *E. histolytica* chitinases, Jacobs, Jessies, and their attached O-P-linked glycans are unique, they appear to be good targets for new antiamebic diagnostics and/or vaccines.

ACKNOWLEDGMENTS

We thank Katie Doud of the Harvard University NIEHS Center for Environmental Health Proteomics Facility for performing two-dimensional gel electrophoresis. We also thank David McCourt of Midwest Analytical for his N-terminal sequencing work and Dick Cook of the Cancer Institute at MIT for some of the mass spectrometry data. We thank Neil Hall of TIGR for providing assemblies of the *E. invadens* genome.

This work was supported in part by National Institutes of Health grants AI44070 (to J.C.S.) and GM31318 (to P.W.R.). Training support (to D.M.R.) was provided by the Training Program in Host Pathogen Interactions (5T32 AI052070). M.S. thanks the Whitaker Foundation Leadership Award to Boston University for financial support.

REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Beausoleil, S. A., M. Jedrychowski, D. Schwartz, J. E. Elias, J. Villen, J. Li, M. A. Cohn, L. C. Cantley, and S. P. Gygi. 2004. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. USA* **101**:12130–12135.
- Bjerrum, O. J., and C. Schafer-Nielsen. 1986. Buffer systems and transfer parameters for semi-dry electroblotting with horizontal apparatus, p. 315–327. *In* J. J. Dunn (ed.), *Electrophoresis '86*. VCH, Weinheim, Germany.
- Bruchhaus, I., B. J. Loftus, N. Hall, and E. Tannich. 2003. The intestinal protozoan parasite *Entamoeba histolytica* contains 20 cysteine protease genes, of which only a small subset is expressed during in vitro cultivation. *Eukaryot. Cell* **2**:501–509.
- Ciucanu, I., and F. Kerek. 1984. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**:209–217.
- de la Vega, H., C. A. Specht, C. E. Semino, P. W. Robbins, D. Eichinger, D. Caplivski, S. Ghosh, and J. Samuelson. 1997. Cloning and expression of chitinases of *Entamoebae*. *Mol. Biochem. Parasitol.* **85**:139–147.
- Eichinger, D. 2001. Encystation in parasitic protozoa. *Curr. Opin. Microbiol.* **4**:421–426.
- Field, J., K. Van Dellen, S. K. Ghosh, and J. Samuelson. 2000. Responses of *Entamoeba invadens* to heat shock and encystation are related. *J. Eukaryot. Microbiol.* **47**:511–514.
- Frisardi, M., S. K. Ghosh, J. Field, K. Van Dellen, R. Rogers, P. Robbins, and J. Samuelson. 2000. The most abundant glycoprotein of amebic cyst walls (Jacob) is a lectin with five Cys-rich, chitin-binding domains. *Infect. Immun.* **68**:4217–4224.
- Haque, R., C. D. Huston, M. Hughes, E. Houpt, and W. A. Petri, Jr. 2003. Amebiasis. *N. Engl. J. Med.* **348**:1565–1573.
- Hart, C., B. Schulenberg, T. H. Steinberg, W. Y. Leung, and W. F. Patton. 2003. Detection of glycoproteins in polyacrylamide gels and on electroblots using Pro-Q Emerald 488 dye, a fluorescent periodate Schiff-base stain. *Electrophoresis* **24**:588–598.
- Higdon, R., N. Kolker, A. Picone, G. van Belle, and E. Kolker. 2004. LIP index for peptide classification using MS/MS and SEQUEST search via logistic regression. *OMICS* **8**:357–369.
- Howard, M. B., N. A. Ekborg, L. E. Taylor, R. M. Weiner, and S. W. Hutcheson. 2003. Genomic analysis and initial characterization of the chitinolytic system of *Microbulbifer degradans* strain 2–40. *J. Bacteriol.* **185**:3352–3360.
- Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
- Loftus, B., I. Anderson, R. Davies, U. C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, et al. 2005. The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**:865–868.
- Mann, B. J., B. E. Torian, T. S. Vedvick, and W. A. Petri, Jr. 1991. Sequence of a cysteine-rich galactose-specific lectin of *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. USA* **88**:3248–3252.
- McConville, M. J., J. E. Thomas-Oates, M. A. Ferguson, and S. W. Homans. 1990. Structure of the lipophosphoglycan from *Leishmania major*. *J. Biol. Chem.* **265**:19611–19623.
- Moody-Haupt, S., J. H. Patterson, D. Mirelman, and M. J. McConville. 2000. The major surface antigens of *Entamoeba histolytica* trophozoites are GPI-anchored proteophosphoglycans. *J. Mol. Biol.* **297**:409–420.
- Nielsen, H., S. Brunak, and G. von Heijne. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* **12**:3–9.
- O'Farrell, P. Z., H. M. Goodman, and P. H. O'Farrell. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell* **12**:1133–1141.
- Riahi, Y., and S. Ankri. 2000. Involvement of serine proteinases during encystation of *Entamoeba invadens*. *Arch. Med. Res.* **31**:S187–S189.
- Samuelson, J., S. Banerjee, P. Magnelli, J. Cui, D. J. Kelleher, R. Gilmore, and P. W. Robbins. 2005. The diversity of protist and fungal dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc. Natl. Acad. Sci. USA* **102**:1548–1553.
- Sharma, M., K. Hirata, S. Herdman, and S. Reed. 1996. *Entamoeba invadens*: characterization of cysteine proteinases. *Exp. Parasitol.* **84**:84–91.
- Shevchenko, A., M. Wilm, O. Vorm, and M. Mann. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**:850–858.
- Stanley, S. L., Jr. 2003. Amoebiasis. *Lancet* **361**:1025–1034.
- Tellam, R. L., G. Wiffels, and P. Willadsen. 1999. Peritrophic matrix proteins. *Insect Biochem. Mol. Biol.* **29**:87–101.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- Touz, M. C., M. J. Nores, I. Slavin, C. Carmona, J. T. Conrad, M. R. Mowatt, T. E. Nash, C. E. Coronel, and H. D. Lujan. 2002. The activity of a developmentally regulated cysteine proteinase is required for cyst wall formation in the primitive eukaryote *Giardia lamblia*. *J. Biol. Chem.* **277**:8474–8481.

29. **Van Dellen, K., S. K. Ghosh, P. W. Robbins, B. Loftus, and J. Samuelson.** 2002. *Entamoeba histolytica* lectins contain unique 6-Cys or 8-Cys chitin-binding domains. *Infect. Immun.* **70**:3259–3263.
30. **Vazquezdelara-Cisneros, L. G., and A. Arroyo-Begovich.** 1984. Induction of encystation of *Entamoeba invadens* by removal of glucose from the culture medium. *J. Parasitol.* **70**:629–633.
31. **Wang, Z., J. Samuelson, C. G. Clark, D. Eichinger, J. Paul, K. Van Dellen, N. Hall, I. Anderson, and B. Loftus.** 2003. Gene discovery in the *Entamoeba invadens* genome. *Mol. Biochem. Parasitol.* **129**:23–31.
32. **West, C. M.** 2003. Comparative analysis of spore coat formation, structure, and function in *Dictyostelium*. *Int. Rev. Cytol.* **222**:237–293.
33. **West, C. M., H. van der Wel, P. M. Coutinho, and B. Henrissat.** 2005. Glycosyltransferase genomics in *Dictyostelium*, p. 235–264. *In* W. F. Loomis and A. Kuspa (ed.), *Dictyostelium* genomics. Horizon Bioscience, Norfolk, United Kingdom.
34. **Wright, H. T., G. Sandrasegaram, and C. S. Wright.** 1991. Evolution of a family of *N*-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. *J. Mol. Evol.* **33**:283–294.
35. **Yates, J. R., III, E. Carmack, L. Hays, A. J. Link, and J. K. Eng.** 1999. Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry. *Methods Mol. Biol.* **112**:553–569.
36. **Yin, Q. Y., P. W. de Groot, H. L. Dekker, L. de Jong, F. M. Klis, and C. G. de Koster.** 2005. Comprehensive proteomic analysis of *Saccharomyces cerevisiae* cell walls: identification of proteins covalently attached via glycosylphosphatidylinositol remnants or mild alkali-sensitive linkages. *J. Biol. Chem.* **280**:20894–20901.