Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 3248–3252, April 1991
Medical Sciences

Sequence of a cysteine-rich galactose-specific lectin of *Entamoeba histolytica*

(vaccine/amebiasis/PCR)

BARBARA J. MANN*†, BRUCE E. TORIAN‡, THOMAS S. VEDVICK§, AND WILLIAM A. PETRI, JR.*

*Departments of Medicine and Microbiology, University of Virginia, Charlottesville, VA 22908; ‡Department of Microbiology, Immunology, and Parasitology, Louisiana State University Medical Center, New Orleans, LA 70112; and §Sibia, Inc., La Jolla, CA 92037

Communicated by Louis H. Miller, January 18, 1991

ABSTRACT **_Entamoeba histolytica_ trophozoites adhere to human colonic mucins and epithelial cells by a cell surface galactose-specific lectin. This lectin, which is composed of two subunits linked by disulfide bonds, has been shown to be a protective antigen in an animal model of amebiasis. We have determined the sequence of the mature form of the 170-kDa heavy subunit from cDNA clones and PCR-amplified fragments. The heavy subunit sequence consisted of a putative extracellular domain containing 1209 amino acids with 16 potential sites for N-linked glycosylation, a 26-amino acid hydrophobic region, and a 41-amino acid cytoplasmic tail. The presence of N-linked oligosaccharides was confirmed by culturing amebae with tunicamycin, which resulted in a decrease in the heavy subunit molecular mass to 160 kDa and a loss of lectin activity. The extracellular domain was remarkable for an extensive cysteine-rich domain that shared identity with similar regions of several other cell surface proteins and appeared to confer protease resistance to the subunit.**

Entamoeba histolytica is a cytopathic enteric protozoan that infects 10% of the world's population (1). Despite the existence of effective anti-amebic chemotherapy, invasive amebiasis is surpassed only by malaria and schistosomiasis as the leading parasitic cause of death. Primates and humans are the only epidemiologically significant reservoirs of *E. histolytica* infection, and there are no known arthropod vectors. If an anti-amebic vaccine were developed that could prevent large bowel colonization by the parasite, it would be theoretically possible to eliminate *E. histolytica* as a cause of human disease.

A galactose-specific lectin of *E. histolytica* has been identified and shown to mediate *in vitro* adherence to physiologically relevant receptors including human colonic mucin glycoproteins and colonic epithelium (2–4). Blockade of this lectin with galactose prevents amebic contact-dependent lysis of mammalian cells (2). The purified lectin is a heterodimer of heavy (170 kDa) and light (35 kDa) subunits linked in a 1:1 molar ratio by disulfide bonds (4). Monoclonal antibodies directed to six epitopes on the heavy subunit have been shown to enhance or inhibit the galactose-binding activity of the lectin, providing evidence that the heavy subunit contains the carbohydrate-binding domain (5).

The lectin heavy subunit is an immunodominant antigen, recognized by the immune sera of virtually all patients with amebic liver abscess (6). The lectin is also antigenically stable, as judged by the conservation of the six murine monoclonal antibody-defined epitopes in all 16 pathogenic strains of *E. histolytica* examined to date (7). Immunization of gerbils with the purified lectin was demonstrated to provide complete protection against amebic liver abscess in 67% of animals challenged intrahepatically with *E. histolytica* (8). To further advance our understanding of the lectin's structure and function, we have cloned and sequenced the gene for the heavy subunit.¶

MATERIALS AND METHODS

***E. histolytica* Culture.** Trophozoites (strain HM-1:IMSS) were grown in TYI-S-33 (2) with penicillin (100 μg/ml) and streptomycin sulfate (Pfizer; 100 μg/ml) at 37°C.

λgt11 Library Screening with Antibodies. Construction of an *E. histolytica* cDNA λgt11 library from strain H-302:NIH has been described elsewhere (9). Recombinant bacteriophage was grown on Petri plates and overlaid with a nitrocellulose filter that had been soaked in isopropyl β-D-thiogalactopyranoside as described in the ProtoBlot protocol (Promega) and in ref. 10. The primary antibody used for screening was a rabbit anti-lectin antiserum that had been preabsorbed with *Escherichia coli* lysate coupled to CNBr-activated Sepharose 4B beads (Pharmacia). Filters were blocked in 3% (wt/vol) bovine serum albumin and incubated with primary antibody at a 1:100 dilution. The secondary antibody was an alkaline phosphatase-conjugated anti-rabbit IgG, used at a 1:5000 dilution. Positive plaques producing a lectin–LacZ fusion protein were visualized with alkaline phosphatase color development substrates. Bacteriophage DNA was isolated from purified plaques using λSorb (Promega).

Northern Blot Analysis. Total *E. histolytica* RNA was isolated using guanidinium isothiocyanate (11). Electrophoresis of RNA on formaldehyde gels and RNA transfer to solid support were performed according to Sambrook (12). After electrophoresis total RNA was transferred to a nylon membrane with 25 mM sodium phosphate (pH 6.5). Hybridization conditions were 6× standard saline/citrate/5× Denhardt's solution/0.1% SDS/sonicated salmon sperm DNA (100 μg/ml)/50 mM sodium phosphate, pH 6.5, at 37°C. Washes were 0.2× SSC/0.1% SDS or 3 M tetramethylammonium chloride (13), as indicated. Oligonucleotides were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Double-stranded DNA was labeled with random primers, the Klenow fragment of DNA polymerase I, and [α-^{32}P]dCTP. Double-stranded probes were boiled before use.

DNA Sequencing. The cDNA clone was subcloned into a plasmid vector and sequenced using the dideoxynucleotide chain-termination method of Sanger *et al.* (14) and a Sequenase kit (United States Biochemical). Deletion plasmids for sequencing were constructed with *Exo* III and S1 nuclease (15). PCR products were prepared for sequencing by polyethylene glycol (PEG) precipitation (16). Briefly, 50-μl PCR mixtures were extracted once with phenol/chloroform and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: TFMS, trifluoromethanesulfonic acid.

†To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M59850).

once with chloroform. One-sixth volume of 20% (wt/vol) PEG 6000 in 2.5 M NaCl was added to the reaction mixture and then incubated at 37°C for 15 min. The mixture was centrifuged for 15 min, washed once with 80% ethanol, and dried. The pellet was dissolved in water. PCR products were sequenced with ^{32}P-end-labeled oligonucleotides and *Taq* DNA polymerase.

PCRs. PCR experiments were carried out using the Amplitaq kit (United States Biochemical). Standard reaction incubation times were 1 min at 93°C, 1 min at 42°C, and 3 min at 72°C, repeated 30 times. Each reaction was preceded with an incubation at 93°C for 5 min and terminated with an incubation at 72°C for 15 min. In some cases, the 42°C annealing temperature was lowered to 37°C. Primers were synthesized at the University of Virginia Sequencing Center or Oligos Etc. (Guilford, CT).

Computer Analysis. Comparisons of the derived lectin protein sequence with the National Biomedical Research Foundation, Swiss-Prot, and GenBank data bases were done using the FASTA program (17).

Deglycosylation of the Galactose-Specific Lectin. The lectin was chemically deglycosylated by trifluoromethanesulfonic acid (TFMS) (18). N-linked glycosylation was blocked by growth of amebic trophozoites in tunicamycin (Sigma)/TYI broth for 24 hr. Ninety minutes after tunicamycin addition, Tran^{35}S-label methionine/cysteine (20 μCi/ml) was added to label newly synthesized protein. After incubation, lectin was immunoprecipitated with mouse polyclonal anti-lectin antisera (19). Adherence of the tunicamycin-treated trophozoites was measured in a rosetting assay (2).

Protease Digestion. Reduction and alkylation of the 170-kDa subunit of the monoclonal-antibody-affinity-purified lectin, electroeluted from an SDS/PAGE gel, were accomplished by suspending the lectin in 2 mM dithiothreitol for 2 hr at 37°C, then adding iodoacetic acid to 5 mM, and incubating in the dark for 2 hr at room temperature. The treated lectin was dialyzed against 6 liters of water and then subjected to digestion by endoproteinase AspN (Boehringer Mannheim) in 0.1 M Tris·HCl, pH 8.0/0.01% SDS. Endoproteinase AspN was added at a 1:50 ratio by weight (enzyme/lectin) and incubated at room temperature for 17 hr. Digestion products were applied to a 15% SDS/PAGE gel and visualized by staining with Coomassie blue.

Sequencing of CNBr Peptide Fragments. The 170-kDa subunit was electroeluted from an SDS/PAGE gel of the monoclonal-antibody-affinity-purified lectin and digested with excess CNBr. The resulting peptides were separated on a 10–20% gradient gel by SDS/PAGE, transferred on to Immobilon membranes, and visualized on the membrane by Coomassie blue staining. Individual peptides were sequenced off the membranes by the method of Matsudaira (20).

RESULTS

Cloning and Sequencing the Heavy Subunit. A λgt11 cDNA library was screened with anti-lectin antisera. A single plaque was isolated that contained two *Eco*RI inserts of 1.97 and 0.46 kilobases (kb). These fragments were subcloned into the plasmid vector pSKII for further analysis. The 1.97-kb fragment was used as a probe to rescreen the library but no new clones were identified. The sequence of the cDNA fragments was determined by the dideoxynucleotide chain-termination method of Sanger *et al.* (14). Both DNA strands were sequenced for the entire gene. The two *Eco*RI fragments were shown to be contiguous in the genome by PCR amplification and sequencing of a fragment that spanned the *Eco*RI site. A single continuous open reading frame was identified and confirmed by the identification of the protein sequences of two CNBr peptides of the 170-kDa subunit (Fig. 1, underlined). The cDNA clone was incomplete, lacking the 5′ end of the coding region. The DNA sequence determined from the cDNA clone begins at residue 483.

```
   1: GRLDEFSADNDYYDGGIMSRGKNAGSWYHSYTHQYDVFYYLAMQPWRHFVWTTCDKNDNT
  61: ECYKYTINEDHNVKVEDINKTNIKQDFCQKEYAYPIEKYEVDWDNVPVDEQRIESVDING
 121: KTCFKYAAKRPLAYVYLNTKMTYATKTEAYDVCRMDFIGGRSITFRSFNTENKAFIDQYN
 181: TNTTSKCLLNVYDNNVNTHLAIIFGITDSTVIKSLQENLSLLSQLKTVKGVTLYYLKDDT
 241: YFTVNITLDQLKYDTLVKYTAGTGQVDPLINIAKNDLATKVADKSKDKNANDKIKRGTMI
 301: VLMDTALGSEFNAETEFDRKNISVHTVVLNRNKDPKITRSALRLVSLGPHYHEFTGNDEV
 361: NATITALFKGIRANLTERCDRDKCSGFCDAMNRCTCPMCCENDCFYTSCDVETGSCIPWP
 421: KAKPKAKKECPATCVGSYECRDLEGCVVTKYNDTCQPKVKCMVPYCDNDKNLTEVCKQKA
 481: NCEADQKPSSDGYCWSYTCDQTTGFCKKDKRGKEMCTGKTNNCQEYVCDSEQRCSVRDKV
 541: CVKTSPYIEMSCYVAKCNLNTGMCENRLSCDTYSSCGGDSTGSVCKCDSTTGNKCQCNKV
 601: KNGNYCNSKNHEICDYTGTTPQCKVSNCTEDLVRDGCLIKRCNETSKTTYWENVDCSNTK
 661: IEFAKDDKSETMCKQYYSTTCLNGKCVVQAVGDVSNVGCGYCSMGTDNIITYHDDCNSRK
 721: SQCGNFNGKCIKGSDNSYSCVFEKDKTSSKSDNDICAECSSLTCPADTTYRTYTYDSKTG
 781: TCKATVQPTPACSVCESGKFVEKCKDQKLERKVTLENGKEYKYTIPKDCVNEQCIPRTYI
 841: DCLGNDDNFKSIYNFYLPCQAYVTATYHYSSLFNLTSYKLHLPQSEEFMKEADKEAYCTY
 901: EITTRECKTCSLIETREKVQEVDLCAEETKNGGVPFKCKNNNCIIDPNFDCQPIECKIQE
 961: IVITEKDGIKTTTCKNTTKATCDTNNKRIEDARKAFIEGKEGIEQVECASTVCQNDNSCP
1021: IITDVEKCNQNTEVDYGCKAMTGECDGTTYLCKFVQLTDDPSLDSEHFRTKSGVELNNAC
1081: LKYKCVESKGSDGKITHKWEIDTERSNANPKPRNPCETATCNQTTGETIYTKKTCTVSEF
1141: PTITPNQGRCFYCQCSYLDGSSVLTMYGETDKEYYDLDACGNCRVWNQTDRTQQLNNHTE
1201: CILAGEINNVGAIAAATTVAAVIVAVVVALIVVSIGLFKTYQLVSSAMKNAITITNENAE
1261: YVGADNEATNAATFNG
```

FIG. 1. Derived protein sequence of the 170-kDa subunit. Underlined residues indicate the amino-terminal and CNBr sequences determined by Edman degradation. Overlined residues indicate putative transmembrane domain. Asterisk residues are potential glycosylation sites. Cysteines are highlighted with a solid circle. Residue 483 is the 5′ end of the cDNA clone. The 3′ end of the PCR-amplified fragment was at amino acid 596. The amino acid sequences were determined by Edman degradation for the amino terminus (GKLNEFSADIDYYDL) and CNBr peptides (MDFIGGHSITFESFNTEHKNFIDQ, MATDNIITYHN, and MKEADKDAYCTYEITT). (Underlined residues indicate differences from the derived sequence.)

The remainder of the gene sequence of the mature protein was obtained using a PCR approach. Two oligonucleotide primers were synthesized for PCR amplification of the missing 5′ end of the gene from genomic DNA. One was an antisense oligonucleotide (17-mer) complementary to the 5′ end of the cDNA sequence (Fig. 1, amino acids 591–596). The sequence of the sense primer was generated from the amino-terminal amino acid sequence of the 170-kDa subunit. This primer was 30 nucleotides long containing 17 nucleotides of lectin sequence (Fig. 1, amino acids 1–6) and a 13-base *Not* I restriction site and was 256-fold degenerate. Codon selection was based on the codon bias exhibited by the actin gene of *E. histolytica* (21, 22). The amplified fragment produced by these primers was 1840 bases long.

The 1840-base-pair segment was not subcloned prior to sequencing. Direct sequencing of the PCR product was undertaken to circumvent *Taq* DNA polymerase-introduced errors in the sequence (23). A single-base discrepancy was observed in >1800 nucleotides sequenced from a PCR template. The only ambiguity occurred at amino acid 441 where a guanine or adenine residue was observed approximately 50% of the time. This resulted in an amino acid change of arginine to lysine. The amino acid composition of the derived protein sequence was not significantly different from the composition determined from the purified protein (4). The amino-terminal sequence and an additional CNBr peptide were identified within the protein sequence of this fragment,

further corroborating the reading frame. The amino-terminal sequence differed from the published sequence (4) at 4 of 15 positions. The CNBr peptide sequences, determined by Edman degradation, differed by 1–4 residues from the derived sequence (Fig. 1). These differences may represent ambiguities in interpreting the Edman degradation or microheterogeneity in the lectin sequence if the lectin is a multicopy gene. Evidence for protein sequence heterogeneity has been observed for the light subunit as well (4).

Fig. 1 represents a composite between the cDNA sequence of strain H-302:NIH and the genomic sequence of HM-1:IMSS, which may not exist in reality in the organism. However, no nucleotide differences were found in 200 bases of overlap that were sequenced in both the genomic and cDNA templates.

The open reading frame identified in the PCR genomic fragment was uninterrupted by introns. To identify introns in the cDNA portion of the gene, several sense–antisense primer pairs were synthesized that encompassed the entire cDNA region and were used in PCR experiments. The sizes of the amplified fragments were identical to the sizes predicted from the cDNA sequence (data not shown). These results indicated that the lectin genomic sequence is not interrupted by introns.

Characteristics of the Lectin Sequence. Hydropathy plots (24) suggested the existence of at least four structural domains in the derived amino acid sequence of the heavy subunit. The initiation methionine and signal sequence were lacking from the derived sequence, which begins with the amino-terminal and presumably extracellularly oriented glycine of the mature protein. The region from amino acids 1 to 187 is cysteine- and tryptophan-rich (3.2% cysteine and 2.1% tryptophan) and hydrophilic. Residues 188–378 contain alternating hydrophobic and hydrophilic stretches of amino acids and are cysteine-free. No significant protein sequence identities could be found to any other proteins in the National Biomedical Research Foundation data bank by FastA searches of the amino-terminal 378 amino acids.

An extremely cysteine-rich (10.8% cysteine) and hydrophilic region followed from residues 379 to 1209. This region was shown to contain repetitive elements by FastA plot of the lectin sequence against itself (Fig. 2*A*). This region also shares some similarity to a number of surface proteins, including human complement receptor CR1, β_1 integrins, differentiation-inducing factor-induced prestalk protein of *Dictyostelium*, and Glp1 of *Caenorhabditis*. Fig. 2*B* also shows a FastA plot of the lectin and the Glp1 protein. The areas of similarities are again confined to the cysteine-rich regions of each protein.

The next putative domain is a hydrophobic stretch of 26 amino acids (residues 1210–1235). This is likely a transmembrane region of the protein and suggests that the lectin is an integral membrane protein. The carboxyl-terminal 41 amino acids presumably form the cytoplasmic tail of the heavy subunit. Contained within the carboxyl terminus is a region surrounding tyrosine-1261 that shares identity with the tyrosine autophosphorylation site of the epidermal growth factor receptor (T**N**E**N**AEY**V**G**A**DN; identical residues are bold; conservative substitutions are underlined).

Northern Blot Analysis. An mRNA of 3.8 kb encoding the 170-kDa subunit was identified by hybridization of total RNA with an oligonucleotide probe to the amino terminus of the 170-kDa subunit or with the cDNA clone (Fig. 3). An oligonucleotide probe generated from the amino-terminal sequence of the 35-kDa subunit strongly hybridized to an RNA of 1.2 kb and weakly hybridized to a band of 1.4 kb. The light subunit oligonucleotide probe was a 48-mer, corresponding to amino acids 6–21 of the amino terminus (4), and was 4096-fold degenerate. Distinct mRNAs hybridizing to the 170-kDa and 35-kDa probes suggest that the subunits are encoded by separate genes, which is consistent with previous observations that the subunits have unique amino-terminal amino acid sequences and are antigenically distinct (4).

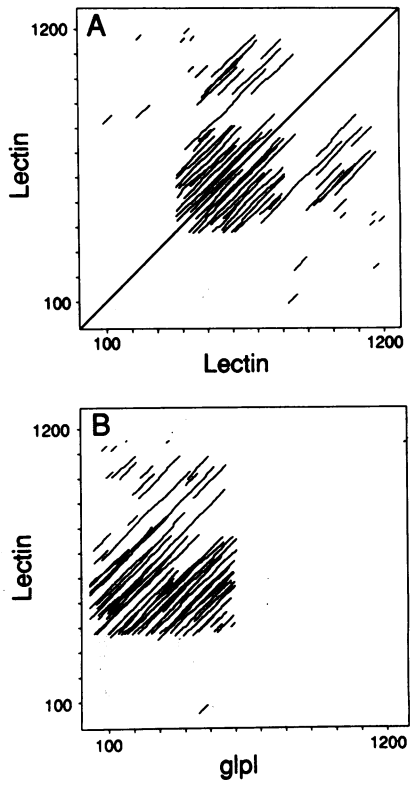

FIG. 2. FastA plots (17) of the 170-kDa subunit compared with itself and the Glp1 protein of *Caenorhabditis elegans*. Sequences were compared one nucleotide at a time (*k*-tup = 1) using the PAM250 matrix. All alignments with scores greater than a cut-off value, calculated based on sequence length, are displayed. (*A*) The 170-kDa subunit compared with itself. (*B*) The 170-kDa subunit (*y* axis) compared to Glp1 (*x* axis).

Extent and Function of Glycosylation. The 170-kDa subunit is a glycoprotein (4). The derived amino acid sequence has a calculated molecular mass of 143 kDa and there are 16 potential N-linked glycosylation sites within this sequence. An estimation of the extent of actual glycosylation was determined by two methods. Purified lectin was chemically deglycosylated with TFMS or the lectin was immunoprecipitated from trophozoites grown in the presence of tunicamycin to block N-linked glycosylation (Fig. 4). Both TFMS and tunicamycin treatment shifted the electrophoretic mobility of

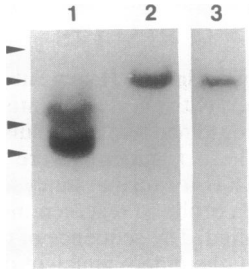

FIG. 3. Northern blots of *E. histolytica* RNA with heavy and light subunit probes. Total RNA was hybridized with an oligonucleotide to the 35-kDa subunit (lane 1), an oligonucleotide to the 170-kDa subunit amino terminus (lane 2), and the 1.97-kb cDNA fragment of the 170-kDa subunit (lane 3). Lanes 1 and 2 were washed in 3 M tetramethylammonium chloride at 50°C, lane 3 was washed in 0.2× SSC/0.1% SDS at 65°C. Molecular size standards are 7.46, 4.40, 2.37, and 1.37 kb (indicated by arrowheads).

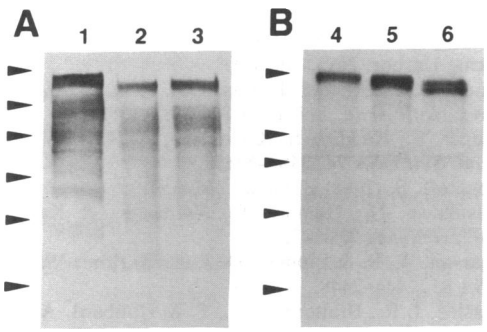

FIG. 4. Apparent molecular masses of the 170-kDa subunit isolated from tunicamycin-treated amebae or deglycosylated with TFMS. (*A*) Purified native lectin was analyzed by SDS/PAGE and Western blots with anti-lectin antisera before (lane 1) or after treatment for 2 hr (lane 2) or 6 hr (lane 3) with TFMS. Molecular mass standards (indicated by arrowheads) are 200, 92.5, 69, 45, 30, and 21.5 kDa. (*B*) Lectin immunoprecipitated from [^{35}S]Met/Cys-labeled trophozoites for 24 hr grown in absence of tunicamycin (lane 4) or in presence of tunicamycin at 0.5 μg/ml (lane 5) or 5 μg/ml (lane 6). Molecular mass standards are 200, 116, 97, 66, and 43 kDa.

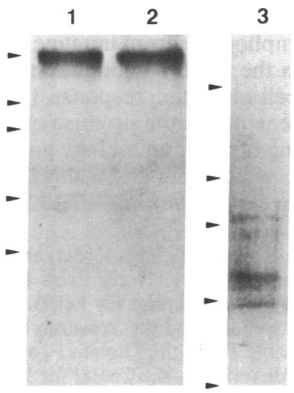

FIG. 5. Effect of reduction and alkylation of 170-kDa subunit on protease sensitivity. The affinity-purified electroeluted 170-kDa subunit (lane 1) was digested with endoproteinase AspN (protease/protein weight ratio, 1:50) for 17 hr at room temperature before (lane 2) or after (lane 3) reduction and alkylation. Molecular mass standards are 200, 92.5, 69, 45, and 30 kDa (indicated by arrowheads).

the 170-kDa subunit to approximately 160 kDa. These results suggest that the major type of glycosylation is N-linked and represents about 6% of the apparent molecular mass. The decrease in size due to deglycosylation did not entirely account for the difference in the apparent and actual molecular mass.

Amebae grown in tunicamycin (5 μg/ml) for 24 hr had dramatically reduced adherence. As measured by a rosetting assay of Chinese hamster ovary cells, adherence at 120 min decreased from control levels of 39 ± 15% to 1.6 ± 0.8% in the tunicamycin-treated amebae (mean ± SEM, n = 5).

Protease Treatment. One postulated function of cysteine-rich regions of cell surface proteins is to confer resistance to protease digestion. The purified lectin heavy subunit was in fact resistant to digestion with trypsin, chymotrypsin, and endoproteinase AspN. However, upon alkylation and reduction the protein became sensitive to proteolytic attack, supporting a contribution of the cysteine-rich areas to the protease resistance of the native lectin (Fig. 5).

DISCUSSION

We have isolated and sequenced clones and PCR-amplified fragments coding for the entire sequence of the mature form of the heavy subunit of the galactose lectin. The heavy subunit consists of a 1209-amino acid putative extracellular region, a 26-amino acid hydrophobic region, and a 41-amino acid carboxyl-terminal putative cytoplasmic region. The sequence encodes an unglycosylated protein with a predicted molecular mass of 143 kDa, which is smaller than the 160-kDa estimated molecular mass of the deglycosylated lectin. This suggests that either the SDS/PAGE estimate is inaccurate or that post-translational modifications other than glycosylation explain the differences in observed and calculated molecular masses.

The heavy subunit sequence does not share significant amino acid identities with the conserved regions of the carbohydrate-binding domains of the C- or S-type animal lectins, the *Escherichia coli* α1 → 4-galactose-specific pilus or plant lectins that have been sequenced to date (25, 26). The evidence that the carbohydrate-binding region is in the heavy subunit is indirect since the reduction in disulfide bonds required to separate the subunits results in loss of the lectin activity (W.A.P., unpublished data). Potentially the galactose-binding domain could reside in the light subunit or be formed by the two subunits together. However, the abilities of anti-heavy subunit monoclonal antibodies to dramatically enhance or inhibit adherence (5) make it more likely that the heavy subunit contains a carbohydrate-binding region whose sequence is not similar to previously described classes of eukaryotic lectins. Sequencing the light subunit and resolving the issue of the contributions of the two subunits to adherence are active areas of investigation.

The galactose lectin is the second cysteine-rich surface antigen identified in *E. histolytica* (9). *Giardia lamblia*, another enteric protozoan parasite, also has several cysteine-rich surface proteins (27, 28). Gillin *et al.* (28) have postulated that cysteine-rich surface proteins may be important for survival in the harsh environment of the gut. The protease resistance of the nonreduced heavy subunit supports such a role for this type amino acid composition in *E. histolytica*.

The evidence from the sequence that the lectin is an integral membrane protein is consistent with the ability of the lectin to cap on the surface of the trophozoites in the presence of anti-lectin antibody and the involvement of the lectin in the endocytosis of bacteria and erythrocytes (29, 30).

The presence of an area with sequence identity to the autophosphorylation site of the epidermal growth factor receptor is intriguing in light of evidence for conformational control of the lectin's activity (5) and data that cytoplasmic tyrosine phosphorylation is involved in the control of integrin cell adhesive activity (31). However, this is speculative in the absence of evidence that tyrosine-1261 is phosphorylated *in vivo*.

The putative extracellular region of the lectin heavy subunit contained 16 potential sites for N-linked glycosylation, and TFMS deglycosylation and tunicamycin treatment indicated that approximately 10 kDa of the lectin consists of N-linked carbohydrate. Inhibition of N-linked oligosaccharide addition to the lectin by growth of the amebae in tunicamycin (5 μg/ml) did not affect amebic growth, viability, or [^{35}S]methionine/cysteine incorporation into the lectin, yet profoundly inhibited the adherence of trophozoites. It has been clearly shown for the glycoproteins of vesicular stomatitis virus, influenza virus, and paramyxoviruses that oligosaccharide chains play an important role in promoting the appropriate folding and disulfide bond formation that results in a fully functional glycoprotein (32–34). The loss of lectin activity at tunicamycin concentrations that did not inhibit protein synthesis is evidence for the importance of N-linked glycosylation in *E. histolytica* as well.

The galactose-specific lectin is a protective antigen in an animal model of amebic liver abscess (8). In humans, antibodies against the lectin's heavy subunit have been detected

in 95% of patients convalescing from amebic liver abscess (6). Incubation of lymphocytes from immune individuals with the lectin resulted in the *in vitro* production of γ-interferon and an amebicidal cell-mediated response (35). Cloning of the heavy subunit gene offers the promise of future advances in our understanding of the immunology and molecular biology of this parasite, which remains a devastating problem for people of the developing world.

We thank Erika Shugart, Kendra Kress, and Terri Snodgrass for excellent technical assistance, Doris Estes for preparation of the manuscript, and William Pearson for helpful advice. W.A.P. is a Lucille P. Markey Scholar and this research has been supported by the Markey Charitable Trust, the Thomas F. and Kate Miller Jeffress Memorial Trust, Alexon Inc., the Center for Innovative Technology of the Commonwealth of Virginia, and U.S. Public Health Service Grant AI-26649. B.E.T. has been supported by U.S. Public Health Service Grant AI-28188.

1. Walsh, J. A. (1986) *Rev. Infect. Dis.* **8**, 228–238.
2. Ravdin, J. I. & Guerrant, R. L. (1981) *J. Clin. Invest.* **68**, 1305–1313.
3. Chadee, K., Petri, W. A., Jr., Innes, D. J. & Ravdin, J. I. (1987) *J. Clin. Invest.* **80**, 1245–1254.
4. Petri, W. A., Jr., Chapman, M. D., Snodgrass, T., Mann, B. J., Broman, J. & Ravdin, J. I. (1989) *J. Biol. Chem.* **264**, 3007–3012.
5. Petri, W. A., Jr., Snodgrass, T. L., Jackson, T. F. H. G., Gathiram, V., Simjee, A. E., Chadee, K. & Chapman, M. D. (1990) *J. Immunol.* **144**, 4803–4809.
6. Ravdin, J. I., Jackson, T. F. H. G., Petri, W. A., Jr., Murphy, C. F., Ungar, B. L. P., Gathiram, V., Skilogiannis, J. & Simjee, A. E. (1990) *J. Infect. Dis.* **162**, 768–772.
7. Petri, W. A., Jr., Jackson, T. F. H. G., Gathiram, V., Kress, K., Saffer, L., Snodgrass, T. L., Chapman, M. D., Keren, Z. & Mirelman, D. (1990) *Infect. Immun.* **58**, 1802–1806.
8. Petri, W. A., Jr., & Ravdin, J. I. (1991) *Infect. Immun.* **59**, 97–101.
9. Torian, B. E., Flores, B. M., Stroeher, V. L. & Hagen, F. S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6358–6362.
10. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49–78.
11. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
12. Sambrook, J., Fritsch, E. F. & Maniatas, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
13. Wood, W. I., Gitshier, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1585–1588.
14. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Henikoff, S. (1984) *Gene* **28**, 351–359.
16. Kusukawa, N., Uemori, T., Asada, K. & Kato, I. (1990) *BioTechniques* **9**, 66–72.
17. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
18. Bartles, J. R., Braiterman, L. T. & Hubbard, A. L. (1985) *J. Biol. Chem.* **260**, 12792–12802.
19. Petri, W. A., Jr., Broman, J., Healy, G., Quinn, T. & Ravdin, J. I. (1989) *Am. J. Med. Sci.* **297**, 163–165.
20. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038.
21. Edman, U., Meza, I. & Agabian, N. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3024–3028.
22. Huber, M., Garfinkel, L., Gitler, C., Mirelman, D., Revel, M. & Rozenblatt, S. (1987) *Mol. Biochem. Parasitol.* **24**, 227–235.
23. Innis, M. A., Myambo, K. B., Gelfand, D. H. & Brow, M. D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9436–9440.
24. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
25. Drickamer, K. (1988) *J. Biol. Chem.* **263**, 9557–9560.
26. Lund, B., Lindberg, F., Marklund, B. T. & Normark, S. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5898–5902.
27. Adam, R. D., Aggarwal, A., Lal, A. A., De La Cruz, V. F., McCutchan, T. & Nash, T. E. (1988) *J. Exp. Med.* **167**, 109–118.
28. Gillin, F. D., Hagbloom, P., Harwood, J., Aley, S. B., Reiner, D. S., McCaffery, M., Magdalene, S. & Guiney, D. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4463–4467.
29. Bailey, G. B., Nudelman, E. D., Day, D. B., Harper, C. C. & Gilmour, J. R. (1990) *Infect. Immun.* **55**, 43–47.
30. Bracha, R. & Mirelman, D. (1983) *Infect. Immun.* **40**, 882–887.
31. Albelda, S. M. & Buck, C. A. (1990) *FASEB J.* **4**, 2868–2880.
32. Machamer, C. E. & Rose, J. K. (1988) *J. Biol. Chem.* **263**, 5955–5960.
33. Copleland, C. S., Zimmer, K. P., Wagner, K. R., Healey, G. A., Mellman, I. & Helenius, A. (1988) *Cell* **53**, 197–209.
34. Vidal, S., Mottet, G., Kolakofsky, D. & Roux, L. (1989) *J. Virol.* **63**, 892–900.
35. Schain, D. C., Salata, R. A., Murphy, C. F. & Ravdin, J. I. (1990) *Clin. Res.* **38**, 352A (abstr.).