The Phenotypically Variable Surface Protein of Trichomonas vaginalis Has a Single, Tandemly Repeated Immunodominant Epitope

DON C. DAILEY AND J. F. ALDERETE*

Department of Microbiology, The University of Texas Health Science Center, San Antonio, Texas 78284-7758

Received 10 January 1991/Accepted 3 April 1991

Trichomonas vaginalis is a sexually transmitted protozoan parasite that undergoes phenotypic variation for numerous surface proteins. A monoclonal antibody (MAb) was used to isolate an ~400-bp cDNA encoding a fragment of an important phenotypically varying immunogen of *T. vaginalis* ($M_r = 270$ kDa; P270). The MAb completely inhibited the binding of P270 by antibody in sera of patients and by antibody in monospecific antiserum obtained toward purified P270, indicating that P270 contained only one immunodominant epitope. Hydrophilicity plot analysis of the deduced amino acid sequence of the recombinant protein predicted the hexapeptide sequence DREGRD as the antigenic determinant of P270. Synthetic peptides synthesized to this region demonstrated that the amino acid sequence DREGRD is important for antibody binding. Seven adjacent amino acids also contributed substantially to maximal recognition of the epitope by the MAb. A single transcript of ~9.5 kb, a size compatible with the reported M_r of the immunogen, hybridized to the cDNA in Northern blots of total RNA from *T. vaginalis*. DNA sequence and Southern blot analysis determined the epitope to be encoded by a 339-bp unit, which was found to be tandemly repeated at least 12 times within the single-copy gene. This 12-mer unit would only constitute ~50% of the protein, yet it is responsible for all of the serum antibody to the immunogen produced by animals and humans. The epitope sequence was found in all fresh and long-term-grown organisms examined to date, demonstrating the stability and conservation of this gene.

Trichomonas vaginalis is a flagellated protozoan responsible for the world's most common sexually transmitted parasitic disease. Trichomoniasis of the vagina is non-selflimiting and imparts a severe emotional and economic burden on women of all world societies. A major problem with diagnosis and control of this disease is the dramatic variation in host symptomatology, ranging from an asymptomatic carrier state to a severe discomfort state. Host or parasite factors responsible for this variation in pathogenesis remain undefined.

Independent investigations have indicated a dramatic variation in the antigenic structure of T. vaginalis isolated from patients (12, 19, 20). That the surface disposition of a repertoire of high-molecular-weight, prominent immunogens may be responsible for some of this antigenic diversity is now appreciated (1, 5, 6). A monoclonal antibody (MAb) reactive with one immunogen having a M_r of ~270 kDa (P270) demonstrated the existence of two isolate types; some isolates consist of a homogeneous population of parasites that synthesize but lack P270 on their surface (type I). Other isolates (type II) comprise a heterogeneous population of organisms with regard to surface expression of the P270 immunogen (5), and fluorescence-activated cell sorting showed for the first time the ability of these trichomonads to undergo phenotypic variation for surface expression of immunogens (3). More recent data indicate that the phenotypically varying isolates predominate in vivo, a finding that suggests that the property of phenotypic variation for P270 is an excellent molecular marker for the possibly more virulent isolates. The P270 immunogen elicits the highest levels of antibody in experimentally infected animals (5), and most, if not all, patients make antibody to this molecule (1). The Of special significance to this investigation was a recent finding that only a single epitope on P270 was accessible to antibody binding on live parasites (unpublished observations). These data initiated our interest in the molecular biology of the immunodominant epitope of this important immunogen of T. vaginalis.

MATERIALS AND METHODS

Microorganisms. T. vaginalis NYH 286 undergoes phenotypic variation for surface expression of immunogens (1, 3). Trichomonads were cultured in Trypticase-yeast extractmaltose medium supplemented with 10% heat-inactivated horse serum (8). Nucleic acid and antigen analyses were performed on organisms grown for 20 h (17).

Escherichia coli strains TB1, HB101, and Y1090 were cultured in Luria-Bertani broth with ampicillin (100 μ g/ml) added as needed (14).

Nucleic acid isolation from T. vaginalis. DNA from trichomonads lysed with 4 M guanidinium thiocyanate was banded on CsCl gradients (21). DNA fractions were collected, diluted twofold with TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]), extracted with chloroform, and precipitated by ethanol (14). For RNA, trichomonads were lysed in 4 M guanidinium thiocyanate, and the RNA was collected by ultracentrifugation through a 5.7 M CsCl cushion prepared in 0.1 M EDTA (7). The poly(A)⁺ RNA (mRNA) was purified from total cellular RNA by chromatography through poly(U) Sephadex G-10 (GIBCO BRL, Gaithersburg, Md.) as described by the manufacturer. This purified mRNA contained no rRNA by electrophoresis analysis on agarose gels (14).

significance of the highly immunogenic and phenotypically varying nature of P270 and its role in the host-parasite biology are unknown and require clarification for any future hope of controlling this major pathogen.

^{*} Corresponding author.

Electrophoresis and hybridizations. Southern blot analysis was performed on samples electrophoresed on 1% agarose gels in 40 mM Tris-acetate buffer (14), and the DNA was transferred to a Zeta-probe membrane (Bio-Rad Laboratories, Richmond, Calif.). Hybridizations with nick-translated cDNA (14) were performed at 42°C in 50% formamide containing $5 \times$ SSC (1 \times SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 2.5 \times Denhardt's solution, 0.1% sodium dodecyl sulfate (SDS), and 1 mM EDTA (14). The filters were then washed at 60°C with 2 \times SSC-0.1% SDS and then washed in 0.1 \times SSC-0.1% SDS at 60°C.

RNA was electrophoresed on 1% agarose-2.2 M formaldehyde gels (14) and transferred to a Zeta-probe membrane in 40 mM NaOH as recommended by the manufacturer (Bio-Rad). Hybridization of the filter was as described above.

Synthesis and cloning of a cDNA encoding the P270 epitope. Double-stranded cDNA was prepared from 2 μ g of trichomonad mRNA by the RNase H method (cDNA Synthesis Kit; Boehringer-Mannheim, Indianapolis, Ind.) by the manufacturer's instructions. After the fill-in reaction, *Eco*RI linkers were added for λ gt11-cloning. The cDNA was then ligated into the bacteriophage vector λ gt11, packaged by using the Gigapack Gold packaging extracts (Stratagene, San Diego, Calif.), and plated on *E. coli* Y1090 cells (14).

By using standard procedures, a recombinant plaque was isolated following immunodetection with MAb C20A3, which is specific for the phenotypically varying P270 immunogen (1, 3). The cDNA insert was purified from an *Eco*RI digest of the bacteriophage DNA and subcloned into the *Eco*RI site of pUC19. Recombinant plasmids were recovered following transformation of *E. coli* TB1 (14).

Expression of the recombinant cDNA molecule in *E. coli.* To facilitate an accurate determination of the size of the protein encoded by the cDNA, the insert was subcloned into the plasmid expression vector pINI-C2 (15). Recombinant plasmids were transformed into *E. coli* strain HB101, and colonies were screened with MAb C20A3 for expression of the recombinant protein. An immunoreactive clone was isolated and inoculated into 5 ml of Luria-Bertani broth containing 100 μ g of ampicillin per ml. Total protein lysates were prepared by resuspending the cells harvested from 1 ml of the broth culture in 100 μ l of SDS-polyacrylamide gel electrophoresis (PAGE) sample buffer (5). Aliquots were then subjected to electrophoresis and immunoblot analysis with MAb C20A3.

Purification of antibody from recombinant protein. Antibody bound to recombinant bacteriophage plaques was eluted with a solution of 0.2 M glycine (pH 2.8)-0.2 M NaCl-0.1% bovine serum albumin. The eluted antibody solution was neutralized with 1 M Tris-HCl and dialyzed against phosphate-buffered saline.

SDS-PAGE and immunoblot analysis. Procedures for SDS-PAGE and immunoblot of T. vaginalis proteins have been described previously (5, 6).

Sequence analysis of the cDNA. The cDNA insert in pUC19 was sequenced by the double-stranded DNA sequencing technique (T7 Sequencing Kit; Pharmacia, Piscataway, N.J.) with the sequencing and reverse sequencing primers (Boehringer-Mannheim). The λ gt11 sequencing primers (Boehringer-Mannheim) were used to confirm the orientation and open reading frame of the DNA insert in the recombinant bacteriophage. The reaction products were visualized by electrophoresis on 6% polyacrylamide sequencing gels and autoradiography (14). The sequence was analyzed with the PC/GENE analysis software (IntelliGenetics, Inc., Moun-

INFECT. IMMUN.



FIG. 1. Identification of the cDNA insert and the recombinant protein encoding the C20A3-reactive epitope. (A) Bacteriophage DNA purified from an immunoreactive plaque was digested completely with EcoRI, and a single insert of ~400 bp was observed after electrophoresis on a 0.8% agarose gel (lane 1, arrow). Identically treated DNA from a duplicate preparation of the same amount of the original bacteriophage vector did not contain the fragment (lane 2). The molecular-weight markers are HindIII-digested lambda DNA. (B) A recombinant protein containing the C20A3-reactive epitope was expressed from the cDNA insert subcloned into the plasmid pINI-C2. Blots of total cell lysates of the recombinant E. coli separated on a 12.5% SDS-polyacrylamide gel identified a protein of ~14.3 kDa when probed with MAb C20A3 (lane 3). There was no recognition of a protein of this size in control cell lysates of E. coli HB101 harboring the plasmid without an insert (lane 2). Lane 1 contains trichloroacetic acid-precipitated proteins of T. vaginalis NYH 286, demonstrating the reactivity of the MAb with P270.

tain View, Calif.). Both strands of the cDNA molecule were completely sequenced at least three times.

Nucleotide sequence accession numbers. The nucleotide sequence data have been assigned GenBank accession numbers M62853 (Fig. 3) and M62854 (Fig. 6).

RESULTS

Isolation of a cDNA encoding the C20A3 MAb-reactive epitope of P270. A high-titered λ gt11 cDNA library (~10⁶ PFU) was screened with MAb to P270 (5). An immunoreactive plaque was isolated, and insert analysis of the bacteriophage DNA revealed the presence of a single insert of ~400 bp (Fig. 1A).

The coding capacity of the cDNA insert was determined by subcloning the DNA into the plasmid expression vector pINI-C2 (15). An immunoreactive colony was detected by a colony-probing enzyme immunoassay, and the recombinant protein was identified by immunoblot analysis following separation on SDS-polyacrylamide gels (Fig. 1B). *E. coli* HB101 harboring the recombinant plasmid expressed a unique MAb-reactive protein of ~14.3 kDa (lane 3), which was absent from identically prepared samples of *E. coli* HB101 transformed with the vector alone as a control (lane 2).

Demonstration that the C20A3 MAb is to a single epitope on the P270 immunogen. To further confirm that the cDNA encoded a key epitope of the P270 immunogen, preparative plaques of the bacteriophage clone expressing the recombinant protein were first treated with monospecific anti-P270 serum. Antibody was eluted from the preparative plaques,

EcoRI



FIG. 2. Demonstration that the recombinant bacteriophage clone encodes the single immunodominant epitope of P270. (A) Confirmation that the recombinant protein encoded an epitope of P270 was obtained by immunoblot analysis of the autodegraded immunogen as described previously (4). Identical patterns of reactivity were noted on blots treated with MAb C20A3 (lane 1) and monospecific, polyclonal rabbit antibody eluted from preparative phage plaques (lane 3). Preincubation of the blot with MAb totally competed with the binding of eluted rabbit antibody (lane 2). Blots were developed with alkaline phosphatase-conjugated goat anti-mouse (lane 1) or anti-rabbit (lanes 2 and 3) antibody. (B) The binding of antibody in pooled sera from patients (5) to the autodigested P270 (4) is shown in lane 1. Treatment of blots with either MAb (lane 2) or eluted rabbit antibody (lane 3) prior to incubation with the pooled patient sera as a source of anti-P270 antibodies totally inhibited binding of human serum antibody to autodegraded P270. An alkaline phosphatase-conjugated anti-human antibody was used as the probe to measure human antibody recognition of the blotted immunogen.

and this affinity-purified antibody was used in immunoblot analysis of trichomonad proteins.

A recent publication demonstrated that P270 was autodegraded by cysteine proteinases present in a trichomonal detergent extract (4). The autodegradation of the parent P270 molecule resulted in a ladder pattern with a periodicity of ~11 kDa on immunoblots as detected by the C20A3 MAb, which was distinctive for this protein (4). The rabbit antibody purified from preparative phage plaques gave patterns of reactivity on immunoblots of the autodegraded immunogen identical to those of the MAb (Fig. 2A, lanes 1 and 3) (4). Control antibody prepared from similarly treated plaques of the bacteriophage vector lacking a cDNA insert failed to recognize P270, reaffirming that the cDNA encoded a portion of the P270 immunogen.

Much to our surprise, however, the purified rabbit antibody derived from the monospecific serum only recognized a single epitope of P270, as demonstrated by a competition immunoblot experiment. Preincubation of the immunoblot with only the C20A3 MAb totally inhibited the binding by the eluted rabbit antibody to P270 (Fig. 2A, lane 2). This abolishment of recognition by eluted rabbit antibody of the parent P270 molecule by MAb suggested that the recombinant protein, $M_r \sim 14$ kDa, encoded the single epitope of P270.

Similar blocking was also observed by the MAb and the rabbit antibody from the phage plaques of recognition of P270 by anti-P270 antibody present in sera of patients with trichomoniasis (Fig. 2B). Lane 1 shows the characteristic multiple-band pattern obtained on immunoblots of the autodegraded immunogen with the patient antibody. Treatment

1 1	gaa	ttc	GG R	GAT	AAC N	GTT V	AGA R	AGT S	AAA K	GGA G	CTC V	ACA T	TTA L	CGC R	GCA A	GCA A	Q Q	Q Q
55	<u>दि</u> व	CCA	CCA	тсс	ATT	AGC	GAT	TTT	ACA	ATA	GAA	GGT	GGC	ACA	GAA	CTG	ACA	ATT
17		P	P	s	I	S	D	F	T	I	E	G	G	T	E	L	T	I
109	GGT	AAT	ACA	TAT	CCA	ATC	ACT	ATC	ACA	CTT	TCG	CCA	TCA	TCA	GAT	TTA	GCA	GAT
35	G	N	T	Y	P	I	T	I	T	L	S	P	S	S	D	L	A	D
163	TGC	TTT	TAT	GCT	TTC	GAC	ACA	GAA	ACT	CAG	CAT	ACA	TTC	CCA	GGT	GAT	GCT	GCC
53	C	F	Y	A	F	D	T	E	T	Q	H	T	F	P	G	D	A	A
217	AGT	AAA	AGC	CAG	TGC	ACA	GAA	CTA	TTG	GGT	AAC	тст	GAT	AAA	ACA	GAG	TAT	ACA
71	S	K	S	Q	C	T	E	L	L	G	N	s	D	K	T	E	Y	T
271 89	GCC A	AAA K	TTA L	<u>ні</u> С <u>АА</u> Q	ndII GCT A	I _JCC S	GGT G	тст s	GCA A	GGC G	AGT S	TTC F	AAT N	CTT L	TTC F	ATC I	CAA Q	GTT V
325 107	GTT V	GAT D	AGA R	GAA E	GCT G	AGG	GAT	AAC N	GTT V	AGA R	AGT S	AAA K	GGA G	GTC V	ACA T	TTA L	CGC R	GCA A
379 125	GCA A	CAA Q	CAA Q	3	<u>Ec</u> gaa	<u>oRI</u> ttc												

FIG. 3. Sequence and open reading frame of the cDNA encoding the immunodominant epitope of the P270 immunogen. Direct repeats of 49 nucleotides (bracketed region) encoding 17 amino acids flanked this cDNA sequence. Single letters representing the amino acids are listed below each codon. Nucleotides and amino acids are numbered at the left, with the first amino acid being the arginine (R) residue immediately following the *Eco*RI linker. A *Hind*III restriction endonuclease site is present at nucleotide 281 (single underline). The amino acid sequence DREGRD (double underline) was predicted as having the greatest probability of forming an antigenic determinant (10).

of blots with either MAb (lane 2) or eluted rabbit antibody (lane 3) resulted in abolishment of patient antibody binding to P270. Identical inhibition results of binding of monospecific antiserum to the parental P270 molecule were obtained with MAb or eluted antibody (data not shown). This shows that antibody in sera of patients is directed toward the single epitope of the recombinant protein and further reinforces the idea that the P270 molecule itself has only one immunodominant epitope. Control serum or antibody preparations from the bacteriophage vector or MAb to an unrelated surface immunogen of *T. vaginalis* (but of the same isotype) (1) did not interfere with patient antibody recognition of P270.

Identification of a repeated sequence element in the cDNA. The cDNA insert was sequenced, and the open reading frame encoding the recombinant protein was determined (Fig. 3). The correct open reading frame was confirmed by sequencing the cDNA insert in the recombinant bacteriophage clone. The cDNA codes for a protein of 14,218 molecular weight, consistent with that obtained from SDS-PAGE of the recombinant E. coli lysate (Fig. 1B). Analysis of the DNA sequence revealed the presence of a 49-bp direct repeat within the cDNA molecule (bases 8 through 56 on the 5' side and 341 through 389 on the 3' side), resulting in the reiteration of 17 amino acids in the recombinant protein. It is important that the first and last 7 bp of the cDNA sequence are derived from the *Eco*RI linkers used in the construction of the library. It was fortuitous that the juncture of the linker DNA with the cDNA resulted in the retention of the correct amino acids at positions 1 and 128 (Fig. 3).

Demonstration of the epitope sequence recognized by MAb C20A3. The algorithm of Hopp and Woods (10) identified the six-amino acid sequence DREGRD (Fig. 3, double underline) as the MAb-binding site on the basis of its hydrophilic-



FIG. 4. Demonstration of the antigenic determinant recognized by the MAb. The hydrophilicity plot of the carboxy-terminal region of the recombinant protein is shown with the corresponding single letter representing the amino acid. The peptide containing the putative epitope sequence is designated as 1, while other synthetic peptides with altered 5' and 3' ends of the peptide 1 sequence are designated as peptides 1-1 and 1-2, respectively. Synthetic peptides corresponding to the indicated sequences were purchased (Multiple Peptide Systems, San Diego, Calif.). A Bio-Dot apparatus (Bio-Rad Laboratories) was used to apply 400 μ g of each peptide onto nitrocellulose, and the blot was treated with MAb by standard immunoblot procedures. The reactivity of the MAb with the peptides is displayed above the corresponding hydrophilic region on the graph.

ity and probability of forming an antigenic determinant (Fig. 4). The seven amino acids immediately downstream of the DREGRD sequence also contributed significantly to the hydrophilicity of this region. Therefore, a 14-amino acid synthetic peptide containing the DREGRD core sequence plus one amino acid upstream and the seven amino acids downstream was synthesized. This peptide was shown by dot blot analysis to strongly bind the C20A3 MAb (Fig. 4, peptide labeled 1). It is noteworthy that a control peptide synthesized to an unrelated hydrophilic region of the recombinant protein (amino acids 80 through 93, peptide 2) failed to bind the MAb under the same conditions. Importantly, and as expected, a peptide specifically altered within the DREGRD sequence (peptide 1-1) and another containing DREGRD but with the seven adjacent amino acids scrambled (peptide 1-2) demonstrated an almost total absence of MAb binding, indicating that although the DREGRD sequence may be the most important, all of the amino acids constituting this entire hydrophilic region of the protein are needed for optimal antibody recognition. Site-directed mutagenesis will be employed in future experiments to definitively map the specific amino acids involved in antibody binding.

Genomic organization and expression of the gene encoding P270. Northern (RNA) and Southern blot analysis were used to investigate the genetic structure of the P270 gene. A single transcript of ~9.5 kb, consistent with the size needed to code for P270, was detected by the cDNA in RNA from the phenotypically varying isolate NYH 286 (Fig. 5A). Southern blots of trichomonad DNA digested with several restriction endonucleases that do not cleave the cDNA sequence, such as EcoRI and SacI (Fig. 3), resulted in the hybridization of a



FIG. 5. Characterization of the gene encoding the P270 immunogen. (A) A single transcript complementary to the cDNA probe was detected by Northern blot analysis of 10 μ g of total RNA from T. vaginalis. (B) Southern blot analysis of restriction endonuclease digests of the trichomonal DNA suggest that the gene is single copy. Approximately 5 µg of DNA was digested with EcoRI (E), HindIII (H), or SacI (S), electrophoresed on 1% agarose gels, and blotted to Zeta-probe membranes. Hybridization was performed with the nick-translated cDNA probe. (C) The tandemly repeated nature of the epitope sequence was demonstrated following limited digestion of trichomonad DNA with HindIII prior to hybridization with the cDNA probe. The DNA (5 μ g) was restricted with 2 U of HindIII for increasing periods of time, electrophoresed on 1% agarose, and analyzed by Southern blotting. HindIII was selected because the sequence recognized by this endonuclease is located in between the repeated segments of the cDNA (Fig. 3), and only a single band was hybridized by the cDNA in Southern blots of complete digests of the trichomonad DNA (panel B).

single band, suggesting the presence of only a single copy of the gene in the parasite genome (Fig. 5B).

Finally, hybridization of HindIII digests of genomic trichomonad DNA probed with the cDNA gave a single band of \sim 330 bp. This was surprising, since two bands were expected because of the HindIII site at position 281 in the cDNA (Fig. 3). Because the cDNA contains directly repeated sequences at the 5' and 3' termini (Fig. 3), we hypothesized that the genomic DNA was tandemly repeated. Indeed, a ladder pattern of recognition of DNA bands differing by ~330 bp was observed in a Southern blot analysis of partially HindIII-digested trichomonad DNA. demonstrating that the DNA sequence encoding the epitope was tandemly repeated at least 12 times (Fig. 5C). Therefore, a HindIII monomer of the repeated sequence was cloned, and the protein encoded by this region was derived from the DNA sequence (Fig. 6, uppercase letters). The sequence data matched those obtained earlier (Fig. 3), and the repeat element was 339 bp, encoding 112 amino acids for a protein of 11,700 molecular weight. To more clearly demonstrate the repetitive nature of this element, we have included upstream and downstream sequences of the HindIII monomer (Fig. 6, lowercase letters). Sequence data of a recently obtained

HindIII cAA GCT TCC GGT TCT GCA GGC AGT TTC AAT CTT TTC ATC CAA GTT GTT GAT 1 q A S G T TCC GGT TCT GCA GGC AGT TTC AAT CTT TTC ATC CAA GTT GTT GAT 54 GAA GCT AGC GAT AAC GTT AGA AGT AAA GCA GTC ACA TTA CCC GCA GCA CAA 18 E G R D N V R S K G V T L R A Q 54 GAA GCT AGC GAT AAC GTT AGA AGT AAA GCA GTC ACA TTA CCC GCA GAA CAA Q N V R S K G V T L R A Q 108 GCA CCA CCA TCC ATT AGC GAT TTT ACA ATA GAA GGT GCG ACA GAA CTA AA Q F T I E G G T E L T 162 GCT AAT ACA TAT CCA ATC ACT ATC ACT ATC ACA CTT TCC CCA TCA T		cag q	tgc c	aca t	gaa e	cta 1	ttg 1	ggt g	aac n	tct s	gat d	aaa k	aca t	gag e	tat y	aca t	gcc a	aaa k	tta 1	
1 CAA GCT TCC GCT TCT GCA GGC AGT TTC AAT CTT TTC ATC CAA GTT GTT GAT 1 q A S G S A G S F N L F I Q V V <u>D</u> 54 GAA GCT A <u>G</u> G GAT AAC CTT AGA AGT AAA GGA GTC ACA TTA CGC GCA CCA CAA 18 <u>E G R D</u> N V R S K G V T L R A A Q 108 <u>GG</u> A CCA CCA TCC ATT AGC GAT TTT ACA ATA GAA GGT GGC ACA GAA CTG ACA 36 <u>G</u> P P S I S D F T I E G G T E L T 162 GGT AAT ACA TAT CCA ATC ACT ATC ACA CTT TGC CCA TCA TCA GAT TTA GCA 54 G N T Y P I T I T L S P S S D L A 216 TCC TTT TAT GCT TTC GAC ACA GAA ACT CAG CAT ACA TTC CCA GGT GCT GCT 72 C F Y A F D T E T Q H T F P G D A 270 AGT AAA AGC CAG TGC ACA GAA CTA TTG GGT AAC TCT GAT AAA ACA GAC TAT 89 S K S Q C T E L L G N S D K T E Y 324 GCC AAA TTA CAA GT TCC <u>GAT</u> CAG <u>GT</u> ACA <u>TC</u> <u>C</u> C	_	Hir	ndIII																	
54 GAA GGT AGG GAT AAC GTT AGA AGT AAA GGA GTC ACA TTA CGC GCA GCA CAA 18 $E - G - R - D = N = V = R = S = K = G = V = T = L = R = A = Q$ 108 GCA CCA CCA TCC ATT AGC GAT TTT ACA ATA GAA GGT GGC ACA GAA CTG ACA 36 $G = P = P = S = I = S = D = F = T = I = E = L = T$ 162 GGT AAT ACA TAT CCA ATC ACT ATC ACA CTT TCG CCA TCA TCA GAT TTA GCA 37 $G = P = S = I = S = D = F = T = I = E = L = T$ 162 GGT AAT ACA TAT CCA ATC ACT ATC ACA CTT TCG CCA TCA TCA GAT TTA GCA 37 $G = N = T = P = I = T = I = T = L = T = C = $	1	cAA q	GCT A	S	GGT G	TCT S	GCA A	GGC G	AGT S	F	AAT N	L	TTC F	ATC I	Q	GTT V	GTT V	GAT D	AGA	
18 <u>E. G. R. D.</u> N. V. R. S. K. G. V. T. L. R. A. A. Q. 108 <u>GGA</u> CCA CCA TCC ATT AGC GAT TTT ACA ATA GAA GGT GGC ACA GAA CTG ACA 36 <u>G</u> P. P. S. I. S. D. F. T. I. E. G. G. T. E. L. T. 162 GGT AAT ACA TAT CCA ATC ACT ATC ACA CTT TCG CCA TCA TCA CAT TTA GCA 54 G. N. T. Y. P. I. T. I. T. L. S. P. S. S. D. L. A. 216 TGC TTT TAT GCT TTC GAC ACA GAA ACT CAG CAT ACA TTC CCA GGT GAT GCT 72 C. F. Y. A. F. D. T. E. T. Q. H. T. F. P. G. D. A. 270 AGT AAA ACC CAG TGC ACA GAA CTA TTG GCT AAC TCT CAT AAA ACA GAG TAT 89 S. K. S. Q. C. T. E. L. L. G. N. S. D. K. T. E. Y. 324 GCC AAA TTA CAA GCT TCC ggt tct gca ggc agt tcc aat ctt tcc atc caa 64 C. C. AAA TTA CAA GCT TCC ggt tct gca ggc agt tcc aat ctt tcc atc caa 65 C. C. AAA TTA CAA GCT TCC ggt tct gca ggc agt tcc aat ctt tcc atc caa 66 A. K. L. Q. A. S.	54	GAA	GGT	AGG	GAT	AAC	GTT	AGA	AGT	-	GGA	GTC	ACA	TTA	CGC	GCA	GCA	CAA	CAA	
108 \overrightarrow{GA} CCA CCA TCC ATT AGC GAT TTT ACA ATA GAA GGT GGC ACA GAA CTG ACA 109 \overrightarrow{G} P P S I S D F T I E G G C T E L T 1102 \overrightarrow{GT} AAT ACA TAT CCA ATC ACT ATC ACA CTT TGG CCA TCA TCA GAT TTA GCA 1103 \overrightarrow{G} N T Y P I T I T L S P S S D L A 1104 \overrightarrow{G} N T Y P I T C A ACT ATC ACA CTT TCG CCA TCA TCA GAT TTA GCA 1105 \overrightarrow{G} TTT TAT GCT TTC GAC ACA GAA ACT CAC CAT ACA TTC CCA GGT GAT GCT 1105 \overrightarrow{G} TTT TAT GCT TTC GAC ACA GAA ACT CAC CAT ACA TTC CCA GGT GAT GCT 1106 \overrightarrow{G} ACA AAA ACC CAC TGC ACA GAA ACT TC GGT AAC TCT CAT AAA ACA CAC TAT 1107 \overrightarrow{G} ACT AAA ACC CAC TGC ACA GAA CTA TTC GGT AAC TCT GAT AAA ACA CAC TAT 1108 \overrightarrow{G} ACC AAA TTA CAA GCT TCC ggt tct gca ggc agt tct aat ctt tct atc cata 1103 \overrightarrow{G} ACC AAA TTA CAA GCT TCC Ggt AC S S D L I S D	18	<u>E</u>	G	B	<u>_</u> D	N	v	R	S	K	G	v	Т	L	R	A	A	Q	Q	
36 <u>G</u> P P S I S D F T I E G G T E L T 162 GGT AAT ACA TAT CCA ATC ACT ATC ACA CTT TCG CCA TCA TCA GAT TTA GCA 54 G N T Y P I T I T L S P S S D L A 216 TGC TTT TAT GCT TTC GAC ACA GAA ACT CAG CAT ACA TTC CCA GGT GAT GCT 72 C F Y A F D T E T Q H T F P G D A 270 AGT AAA AGC CAG TGC ACA GAA CTA TTG GGT AAC TCT GAT AAA ACA GAG TAT 89 S K S Q C T E L L G N S D K T E Y 324 GCC AAA TTA CAA GCT TCC ggt tct gca ggc agt tcc aat ctt tcc atc caa 66 A K L Q A S E S A S F N I I I I I	108	GGA	CCA	CCA	TCC	ATT	AGC	GAT	TTT	ACA	ATA	GAA	GGT	GGC	ACA	GAA	CTG	ACA	ATT	
162 GGT AAT ACA TAT CCA ATC ACT ATC ACA GTT TGG CCA TCA TCA GAT TTA GCA 54 G N T Y P I T I T L S P S S D L A 216 TGC TTT TAT GCT TTC GAG ACA GAA ACT CAG CAT ACA TTC CCA GGT GAT GCT 72 C F Y A F D T E T Q H T F P G D A 270 AGT AAA ACC CAG TGC ACA GAA CTA TTG GGT AAC TCT GAT AAA ACA GAG TAT 89 S K S Q C T E L L G N S D K T E Y HindIII 324 GCC AAA TTA CAA GCT TCC ggt tct gca ggc agt tct aat ctt tct atc caat 80 A K L Q A S S S S S D L A	36	_9	P	P	S	I	S	D	F	Т	I	E	G	G	T	E	L	Т	I	
216 TGC TTT TAT GCT TTC GAC ACA GAA ACT CAG CAT ACA TTC CCA GGT GAT GCT 72 C F Y A F D T E T Q H T F P G D A 270 AGT AAA AGC CAG TGC ACA GAA CTA TTG GGT AAC TCT GAT AAA ACA GAG TAT 89 S K S Q C T E L L G N S D K T E Y <u>HindIII</u> 324 GCC AAA TTA CAA GCT Tcc ggt tct gca ggc agt tcc aat ctt tcc atc caat 89 A L Q A S S S S S D L A	162	GCT	AAT	ACA	TAT	CCA	ATC	ACT	ATC	ACA	стт	TCG	CCA	TCA	TCA	GAT	TTA	GCA	GAT	
216 TGC TTT TAT GCT TTC GAG AGA GAA AGT GAG GAT AGA TTC GGA GGT GAT GGT 72 C F Y A F D T E T Q H T F P G D A 270 AGT AAA AGC CAG TGC AGA GAA CTA TTG GGT AAC TCT GAT AAA AGA GAG TAT 89 S K S Q C T E L L G N S D K T E Y HindILI 324 GCC AAA TTA GAA GGT TCC ggt tct gca ggc agt tcc aat ctt ttc atc caat 89 A L Q A S E S E S E T L I G N L I I I I I I I I I I I I I I I I I I	54	6	N	1	1	r	1	1	1	1	L	3	P	5	5	D	L	A	D	
72 C F T A F D T E T Q H T F P G D A 270 AGT AAA AGC CAG TGC ACA GAA CTA TTG GGT AAC TCT GAT AAA ACA GAG TAT 89 S K S Q C T E L L G N S D K T E Y <u>HindIII</u> 324 GCC AAA TTA CAA GGT Tcc ggt tct gca ggc agt ttc aat ctt ttc atc caa 08 A K L Q A s g s a g s f n l f i g	216	TGC	TTT	TAT	GCT	ттс	GAC	ACA	GAA	ACT	CAG	CAT	ACA	ттс	CCA	GGT	GAT	GCT	GCC	
270 AGT AAA AGC CAG TGC AGA GAA CTA TTG GGT AAC TCT GAT AAA AGA GAG TAT 89 S K S Q C T E L L G N S D K T E Y <u>HindIII</u> 324 GCC AAA TTA CAA GGT Tcc ggt tct gca ggc agt ttc aat ctt ttc atc caa 108 A K L Q A s g s a g s f n l f i n	72	C	r	Ŷ	•	F	U	т	E	т	Q	н	т	F	P	G	D	A	•	
89 S K S Q C T E L L G N S D K T E Y <u>HindIII</u> 324 GCC AAA TTA CAA GCT Tcc ggt tct gca ggc agt ttc aat ctt ttc atc caa 108 A K L Q A s g s a g s f n l f i o	270	AGT	***	AGC	CAG	TGC	ACA	GAA	CTA	TTG	GGT	AAC	тст	GAT	AAA	ACA	GAG	TAT	ACA	
HindIII 324 GCC AAA TTA CAA GCT Tec ggt tet gea gge agt tte aat ett tte ate caa 108 A K L O A s g s a g s f n l f i g	89	5	ĸ	S	Q	C	т	E	L	L	G	N	S	D	ĸ	т	E	Y	т	
108 A K L Q A s g s a g s f n l f i o		HindIII																		
	108	GCC	AAA V	T		GCT	100	ggt	tet	gca	ggc	agt	ttc	aat	ctt	ttc	atc	caa	gtt	
	100	ñ	ĸ	L	v	^	5	B	5	a	g	5	ſ	n	T	r	1	q	v	
gtt gat aga gaa ggt agg gat aac gtt		gtt	gat	aga	gaa	ggt	agg	gat	aac	gtt										

FIG. 6. Sequence of a *Hind*III monomer of genomic DNA reactive with the cDNA probe (uppercase letters). The open reading frame of this fragment was deduced from the sequence data in Fig. 2. The single letters representing the amino acids are listed below each codon. The epitope sequence DREGRD is doubly underlined, and the brackets show the 49 nucleotides that were repeated in the sequence data of Fig. 3. Lowercase letters denote sequences immediately upstream and downstream of this gene segment. Confirmation that this 339-bp region is tandemly repeated in the gene for P270 was obtained by sequencing a larger (1.1 kb) cDNA encoding the MAb-binding epitope (data not shown), some of which is shown in lowercase letters.

cDNA insert of ~ 1.1 kb, containing tandemly repeated elements (data not shown), reaffirm the data presented in Fig. 6.

DISCUSSION

Our data demonstrate the epitope immunodominance of the amino acid sequence DREGRD and of neighboring amino acids recognized by a MAb. These results confirm earlier protein autodegradation data (4), which suggested a tandemly repeated determinant on the protein. The highly immunogenic nature of this protein could serve to aid the parasite in diverting and evading the immune system. It has been shown that surface proteins containing repetitive epitopes (11) may influence the host antibody response, resulting in interference with or subversion of the immune response to other important molecules. Since all patients appear to produce antibody to P270 (1), the property of phenotypic variation for surface disposition of P270 (1, 3) ensures parasite survival, even in the presence of a cytolytic antibody response (2).

Previous work suggested a repeated structure in P270 with an equidistant cleavage pattern that had a periodicity of ~ 11 kDa (4), which was consistent with the 11,700 molecular weight of the protein encoded by the repeat element (Fig. 6). The similarity in the periodicity of the proteinase cleavage and the size of the repeat element indicated that the autodegradation pattern of P270 resulted from proteinase digestion within the repeated sequence. Previously published data have already documented that the trichomonad cysteine proteinases are responsible for the autodegradation of P270 (4). Very little is known about the substrate specificity of cysteine proteinases other than a preference for arginine in the P1 or P2 position (16). The repeated amino acid sequence contains four arginine residues clustered around the epitope sequence. It is intriguing and noteworthy that partial digestion of P270 at these residues would result in protein fragments differing by 11,700 Da (Fig. 2), consistent with the ladder pattern periodicity reported earlier (4). Furthermore, it is significant that complete digestion of P270, if cleavage occurs at the arginine residues, would degrade all DREGRD epitopes, finally explaining the lack of accumulation of any small immunoreactive peptides in autodegradation experiments of the earlier report (4). These investigations of the repeat element and its amino acid sequence might ultimately define substrate specificity of the poorly understood and uncharacterized trichomonad cysteine proteinases (16).

We now describe the highly immunogenic nature of the single repetitive DREGRD epitope of the P270 immunogen of T. vaginalis. The immunodominance of the repeat sequence is similar to that described for the tandemly repeated peptide of the circumsporozoite protein of Plasmodium knowlesi, which is responsible for ~95% of the antibody response (22). However, an interesting feature of the repeat element of P270 of T. vaginalis was the identical number and size of this repeat among numerous trichomonal isolates (1, 3, 5, 6). The stability of this gene and the conservation of the DREGRD epitope sequence were further illustrated by analysis of isolates cultured in vitro for >10 years (3, 8, 17). This lack of molecular divergence may be indicative of some important biological function for this molecule, apart from the suggested immunological function. Proteins containing tandemly repeated segments have been widely described for the parasitic protozoa, including the insect stages of trypanosomes (9, 13, 18, 22). Since the immune status of an insect is dramatically different from that of a human, these tandemly repeated regions must serve functions in addition to contributing to the overall immune evasion by the parasite. A recent report identified tandemly repeated sequences in a microtuble-associated protein of Trypanosoma brucei (18), indicating that the P270 proteins may also mediate important biofunctional roles in trichomonads. Further experimentation into the molecular biochemistry of the P270 immunogen and its tandemly repeated epitope will likely provide answers to important questions involving not only its specific role in the biology of T. vaginalis but also the regulation of phenotypic variation for this and other members of the immunogen repertoire (6).

ACKNOWLEDGMENTS

We thank Diana Hinojosa for the excellent secretarial assistance in the preparation of the manuscript. The technical assistance of Ricky Ochoa and Melissa White, supported by summer internship fellowships of this department and the American Society for Microbiology, respectively, is gratefully acknowledged. The help of Alan Barbour in reviewing the manuscript was greatly appreciated.

This investigation was supported by Public Health Service grant AI-18768 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

- Alderete, J. F., P. Demeš, A. Gombošova, M. Valent, A. Yánoška, H. Fabušova, L. Kasmala, G. E. Garza, and E. C. Metcalfe. 1987. Phenotype and protein-epitope phenotypic variation among fresh isolates of *Trichomonas vaginalis*. Infect. Immun. 55:1037-1041.
- 2. Alderete, J. F., and L. Kasmala. 1986. Monoclonal antibody to a major glycoprotein immunogen mediates differential comple-

INFECT. IMMUN.

ment-independent lysis of *Trichomonas vaginalis*. Infect. Immun. 53:697-699.

- 3. Alderete, J. F., L. Kasmala, E. C. Metcalfe, and G. E. Garza. 1986. Phenotypic variation and diversity among *Trichomonas vaginalis* and correlation of phenotype with contact-dependent host cell cytotoxicity. Infect. Immun. 53:285–293.
- 4. Alderete, J. F., and K. A. Neale. 1989. Relatedness of structures of a major immunogen in *Trichomonas vaginalis* isolates. Infect. Immun. 57:1849–1853.
- Alderete, J. F., L. Suprun-Brown, and L. Kasmala. 1986. Monoclonal antibody to a major surface immunogen differentiates isolates and subpopulations of *Trichomonas vaginalis*. Infect. Immun. 52:70–75.
- Alderete, J. F., L. Suprun-Brown, L. Kasmala, J. Smith, and M. Spence. 1985. Heterogeneity of *Trichomonas vaginalis* and discrimination among trichomonal isolates and subpopulations by sera of patients and experimentally infected mice. Infect. Immun. 49:463–468.
- Chirgwin, J. M., A. E. Przybyla, R. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294– 5299.
- Diamond, L. S. 1957. The establishment of various trichomonads of animals and man in axenic cultures. J. Parasitol. 43:488–490.
- Hoft, D. F., K. S. Kim, K. Otsu, D. R. Moser, W. J. Yost, J. H. Blumin, J. E. Donelson, and L. V. Kirchhoff. 1989. *Trypano*soma cruzi expresses diverse repetitive protein antigens. Infect. Immun. 57:1959–1967.
- Hopp, T. P., and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequence. Proc. Natl. Acad. Sci. USA 78:3824–3828.
- 11. Kemp, D. J., A. F. Couman, and D. Walliker. 1990. Genetic diversity in *Plasmodium falciparum*. Adv. Parasitol. 29:75-128.
- 12. Krieger, J. N., K. K. Holmes, M. R. Spence, M. F. Rein, W. M.

McCormack, and M. R. Tam. 1985. Geographic variation among isolates of *Trichomonas vaginalis*: demonstration of antigenic heterogeneity by using monoclonal antibodies and the indirect immunofluorescence technique. J. Infect. Dis. 152:979–984.

- Lafaille, J. J., J. Linss, M. A. Krieger, T. Souto-Padrón, W. de Souza, and S. Goldenberg. 1989. Structure and expression of two *Trypanosoma cruzi* genes encoding antigenic proteins bearing repetitive epitopes. Mol. Biochem. Parasitol. 35:127-136.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 15. Masui, Y., J. Coleman, and M. Inouye. 1983. Multipurpose expression cloning vehicles in *Escherichia coli*, p. 15–32. *In* M. Inouye (ed.), Experimental manipulation of gene expression. Academic Press, Inc., New York.
- North, M. J., J. C. Mottram, and G. H. Coombs. 1990. Cysteine proteinases of parasitic protozoa. Parasitol. Today 6:270–275.
- 17. Peterson, K. M., and J. F. Alderete. 1982. Host plasma proteins on the surface of pathogenic *Trichomonas vaginalis*. Infect. Immun. 37:755-762.
- Schneider, A., A. Hemphill, T. Wyler, and T. Seebeck. 1988. Large microtubule-associated protein of *T. brucei* has tandemly repeated, near-identical sequences. Science 241:459–462.
- 19. Su-Lin, K. E., and B. M. Honigberg. 1983. Antigenic analysis of *Trichomonas vaginalis* strains by quantitative fluorescent antibody methods. Z. Parasitenkd. 69:161–181.
- Teras, J. K. 1966. Difference in the antigenic properties within strains of *Trichomonas vaginalis*. Wiad. Parazytol. 12:357–363.
- Wang, A. A., and C. C. Wang. 1985. Isolation and characterization of DNA from *Trichomonas vaginalis* and *Tritrichomonas foetus*. Mol. Biochem. Parasitol. 14:323–335.
- Zavala, F., A. H. Cochrane, E. H. Nardin, R. S. Nussenzweig, and V. Nussenzweig. 1983. Circumsporozoite proteins of malaria parasites contain a single immunodominant region with two or more identical epitopes. J. Exp. Med. 157:1947–1957.