Microbial Pathogenesis 1998; **24**: 223–239

MICROBIAL PATHOGENESIS

Molecular analysis of the gene encoding the immunodominant phenotypically varying P270 protein of *Trichomonas vaginalis*

Oxana Musatovova & J. F. Alderete*

Department of Microbiology, The University of Texas Health Science Center, 7703 Floyd Curl Dr., San Antonio, TX 78284-7758, U.S.A.

(Received October 9, 1997, accepted in revised form December 17, 1997)

Trichomonas vaginalis is a flagellated protozoan responsible for the most common non-viral sexually transmitted disease. The immunogen P270 was previously found to be up-regulated in expression and to undergo phenotypic variation between surface versus cytoplasmic localization in trichomonads harbouring a dsRNA virus. In this report, we characterize the entire *p270* open reading frame (ORF) and the unknown flanking 5′- and 3′-unique, non-repeat coding sequences of the gene in addition to untranslated regions. Consistent with an earlier report (Dailey & Alderete, 1991, *Infect. Immun.* **59**: 2083–88), a significant portion of the gene consists of a tandemly repeated 333 bp element that contains the sequence coding for the epitope DREGRD detected by murine monoclonal antibody and antibody from the sera of patients. The non-repeat coding regions for the 5′- and 3′-ends were 69 nucleotides (23 amino acids) and 1183 nucleotides (395 amino acids), respectively. Sequencing of repeat elements showed them to be identical, affirming the highly-conserved nature of this element throughout the gene. The start codon was immediately preceded by the 12 nucleotide consensus sequence (TCATTTTTAATA) found in other trichomonad protein-coding genes. A very AT-rich, non-coding region was identified upstream of the *p270* ORF. P270 appears to contain a leader sequence at the amino-terminus and transmembrane domain at the carboxy-terminus. No significant homology was found with any reported proteins at either the nucleotide or amino acid level.

Key words: dsRNA virus, host-parasite, repeat element, *Trichomonas vaginalis*, virulence.

Introduction

Trichomonosis, the most common non-viral sexually transmitted vaginitis, is caused by *Trichomonas vaginalis.* Estimates of women infected with this protozoan range from 5 to 10 million in the U.S.A. and 250 to 300 million world-wide, showing the extent of morbidity attributable to this infection [1]. Trichomonosis is now recognized as being life-threatening for infected women. It is associated with adverse pregnancy outcomes [2], enhanced predisposition to HIV infection [3–6], and possibly cervical neoplasia [7, 8].

The high M_r immunogenic P270 protein with

* Author for correspondence.

0882–4010/98/040223+17 $25.00/0/mi970199

a tandemly repeating element (TRE) [9] is significant in the *T. vaginalis*-host relationship. The parasite has the capacity to modulate production and surface expression of P270, thereby avoiding host destruction by antibody [10–12]. Examination of parasites isolated at different time points from the same women infected with *T. vaginalis* revealed a change in the ratio of trichomonads with and without surface P270 [12], suggesting that phenotypic variation may occur *in vivo*. Furthermore, mice inoculated with organisms surface-expressing P270 showed the increase in anti-P270 antibody titre concomitant with the appearance of parasites without surface P270 [11]. Our studies have demonstrated that monoclonal antibodies to P270 were directed toward the dominant epitope of the repeated element [9] and was cytolytic for trichomonads [10]. Only sera from patients with trichomonosis, but not those without infection or with other sexually transmitted diseases, had anti-P270 antibody directed toward the same DREGRD epitope [9]. P270 appears to be highly conserved, as over 1000 isolates examined to date synthesize the immunologically reactive protein.

An association was described between *T. vaginalis* isolates harbouring a double stranded RNA virus and the ability of trichomonads to undergo phenotypic variation for P270 [13]. This finding represents, to our knowledge, the first report of an association of an important biological property with infection by a dsRNA virus. This relationship has been confirmed by examining agar-cloned virus-free progeny trichomonads derived from the virus-positive parental isolates [14]. Virus-free isolates and progeny organisms cytoplasmically express lower amounts of P270 by up to two orders of magnitude [12, 14]. These data and recent reports have provided an experimental basis for the hypothesis that virus infection directly influences *p270* gene expression and/or protein synthesis in addition to modulation of surface placement of P270 [9, 13, 14].

In our previous work, the molecular characterization of the *p270* gene was started for the laboratory isolate NYH 286 [9]. Despite repeated attempts, only a fragment originating from the TRE was recovered from a cDNA library and was found to be 333 bp long. Based on the preliminary data in this report [9], we hypothesized that P270 was encoded by a single copy gene with a TRE that probably comprised a large portion of the coding sequence and was flanked with nonrepetitive sequences at both 5′- and 3′-ends (Fig. 1). It has been shown for a wide range of pathogens that the presence of repetition in protein-coding sequences results in a repertoire of potential genetic variability. This then may be involved in response to changing environmental stimuli or a host defence mechanism. These variable genetic systems are involved in expression of surface proteins expected to function directly in pathogen–host interactions, and include numerous bacteria [15, 16] and protozoan pathogens [17–20].

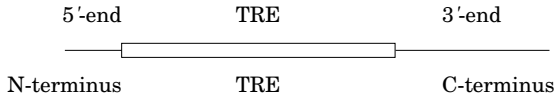


Figure 1. Model of the *p270* gene originally constructed from experiments conducted with the long-term-grown laboratory isolate *T. vaginalis* NYH 286 [9]. A large portion of the gene is represented by tandemly repeated element (TRE), and the repeat element was defined by a *Hin*dIII restriction that generated a 333 bp segment without any apparent interruptions or divergence [9]. The TRE region is flanked by putative coding, non-repeat 5′- and 3′-end sequences.

In order to understand the contribution of P270 and phenotypic variation in *T. vaginalis* pathogenesis, the next step would be to obtain the complete sequence of the *p270* gene. Information on the entire gene might help establish a functional role for this protein and begin to delineate regulation of gene expression by environmental factors or viral infection. In this report we present the molecular characterization of the entire *p270* gene from a fresh clinical isolate. We show that P270 has a leader sequence, a transmembrane domain, and a TRE encoding the immunodominant epitope [9].

Results

Characterization of P270 and the *p270* gene

Figure 2(a) shows the strong immunodetection of the P270 protein of the *T. vaginalis* clinical isolate T068-II by monoclonal antibody C20A3 [9, 11, 21]. The M_r was determined by a series of experiments using lower percentage acrylamide gels with accompanying known high-sized standard proteins. Once again, a M_r of $\sim$270 kDa

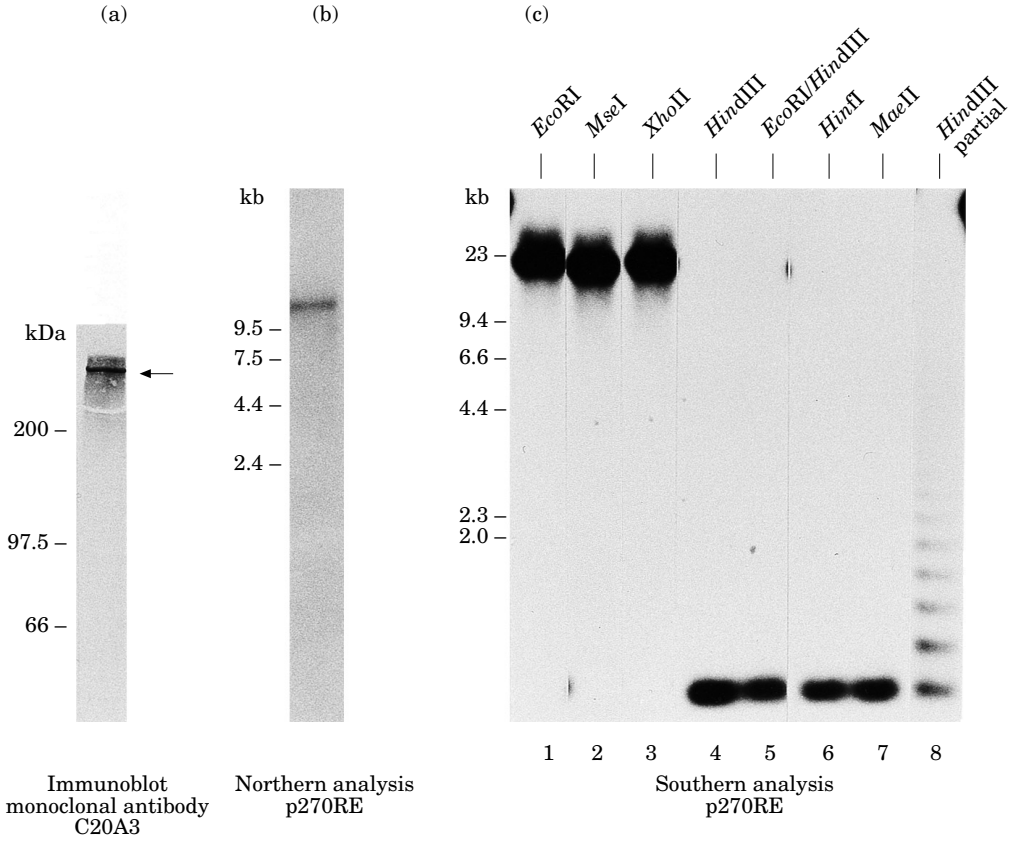


Figure 2. P270 and characterization of the *p270* gene in the T068-II isolate. (a) The immunoprecipitations of P270 for monoclonal antibody detection by immunoblot (arrow) was performed using the monoclonal antibody C20A3 [44]. The immunoprecipitated protein was electrophoresed in SDS-PAGE (7%) and transferred to nitrocellulose. Numbers on the left side represent standard size marker in kilodaltons (kDa) (BioRad). (b) Northern hybridization was performed with 10 μg of total RNA electrophoresed in 1% denaturing agarose gel, blotted onto Zeta-probe membrane, and hybridized under high stringency conditions with nick-translated p270RE. Numbers correspond to weight marker in kilobases (kb). (c) Southern analysis was performed with 5 μg of genomic DNA digested to completion, electrophoresed in 0.8% agarose gel, and transferred onto Zeta-probe membrane for hybridization with ^{32}P-labelled p270RE. Numbers represent λDNA restricted with *Hin*dIII as size markers. To determine the number of repeats, partial digestion with *Hin*dIII of trichomonad DNA was performed prior to electrophoresis in 1% agarose gel, blotting, and hybridization with ^{32}P-labelled p270RE.

was calculated, consistent with that reported previously by us for the laboratory isolate NYH 286 [9, 11, 21]. Autodegradation of P270 (not shown) detected by immunoblot also yielded a ladder pattern of equidistant protein bands, again concordant with the previously published observations [21]. Equally important, Northern analysis of total RNA probed with the cloned repeat sequence cDNA (termed p270RE) showed only one transcript of >9.5 kb (b). Southern analysis was then performed on the genomic DNA digested with twenty different enzymes that do not cut within the TRE, and large-sized bands were detected with ^{32}P-labelled p270RE probe. The smallest fragment of ~13 kb was obtained from *Mse*I restriction (c, lane 2), and ~14 kb fragments were detected after *Eco*RI (lane 1) and *Xho*II restriction (lane 3). The relative size of these large migrating bands was determined by running a series of different percentage agarose gels with both *Hin*dIII-cut λDNA and 1 kb DNA ladder reference markers. Restriction with *Hin*dIII, *Hin*fI and *Mae*II (lanes 4 to 7) that digest within the TRE yielded a single fragment of 333 bp, as determined by migration against 100-bp sized markers (not shown) and as previously reported [9], confirming the tandem organization without interruptions or divergence

of the TRE sequence. The possible absence in autoradiograms of additional bands that would represent 5′- and 3′-end unique sequence attached to the TRE could be explained by the consumption of the ^{32}P-labelled probe by the large number of repeats. However, as shown later by sequence analysis of these 5′- and 3′-end fragments (Figs 7 and 8), the small size of the 5′ coding sequence, which migrates in front of the 333 bp repeat, and the incomplete portion of the repeat on the 3′ end likely explain the absence of detectable additional bands. Finally, evaluation of X-ray film exposed for different times to hybridization blots of the *Hin*dIII partially-digested DNA allowed us to distinguish up to 18 laddered equidistant bands for the *p270* gene in this isolate. The results present in lane 8 show the ladder pattern at the bottom of the gel for X-ray film exposed for a short period of time. Collectively, these data suggest that P270 in *T. vaginalis* T068-II is also encoded by a single copy gene with many tandemly repeated 333 bp elements, reinforcing the original model (Fig. 1) and our earlier report [9]. Further, the non-repeat coding sequence of the *p270* gene appeared to be AT-rich or, alternatively, the gene was flanked by AT-rich sequences, based on restriction with AT-cutting enzymes.

The 31B4 genomic clone

A genomic library was constructed in the λ replacement vector EMBL4. A recombinant phage termed 31B4 harbouring a large insert was identified by probing with the ^{32}P-labelled p270RE. Restriction of recombinant DNA with *Eco*RI, which cleaves the insert from the EMBL4 arms, yielded an insert of a size sufficient to possibly encode for the *p270* gene [Fig. 3(a), lane 1]. Digestion of 31B4 with *Mse*I (lane 2) and *Xho*II (lane 3) yielded single large fragments, as seen above after digestion of the genomic DNA. These results suggested strongly that our genomic clone possessed the entire TRE region with accompanying 5′- and 3′-end sequences.

Visualization of stained gels after electrophoresis of the recombinant DNA digested with *Hin*dIII is shown in Fig. 3(b). Surprisingly, *Hin*dIII restriction produced only the expected 333 bp band plus the larger fragments of the vector. The only indication of additional, non-TRE sequence was the ∼6 kb band, which was larger than the expected 5 kb size of the vector right arm fragment (lane 1). The *Hin*dIII/*Eco*RI double digestion (lane 2) produced a readily visible band of ∼1.2 kb (arrow) that presumably represented non-repeat sequence. Restriction analysis suggested that the cloned trichomonad DNA fragment contained the TRE and, likely, relatively short flanking unique sequences.

We then reasoned that Southerns would detect the bands distinct from the 333 bp repeat element band. Lanes 4, 6 and 7 [Fig. 3(a)] show the weak hybridization of DNA fragments carrying unique sequence accompanied by some portion of the repeat sequence after *Hin*dIII, *Hin*fI and *Mae*II restriction (asterisks). The diminished intensity of these bands, in contrast to the intense 333 bp hybridizing band, was expected given the presence of only a part of one RE. The sizes of the hybridizing fragments reaffirmed that 31B4 harboured a short non-repeat sequence. Interestingly, the 20 kb fragment that corresponded to the left vector arm was one of the hybridizing fragments (lane 4). Based on the known repeat element sequence and that of the vector, this indicated clearly that the tandem repeats were localized very close to this arm of the vector. As expected, the 1.2 kb fragment readily visible in agarose gels after *Eco*RI/*Hin*dIII double digestion [Fig. 3(b), lane 2] did not hybridize with the repeat probe [Fig. 3(a), lane 5] confirming once again that this sequence consisted of non-repeat DNA.

Subcloning and sequencing of the 3′-end region of the *p270* gene

The non-repeat sequence close to the vector arm was derived by PCR using primers to the left arm of the vector (primer M2) (see Materials and methods) and the TRE. Since the orientation of the insert in the genomic clone had not yet been established, two primers were designed based on the p270RE sequence. Only one of the primers (primer P2) yielded PCR product, strongly suggesting that the amplified region represented the 3′-end sequence downstream of the TRE. Not unexpectedly, the P2 primer bound at multiple sites along the TRE, and the amplification resulted in up to seven bands distinguishable in agarose gels [Fig. 4(a), lane 1]. This was consistent with the map of the *p270* genomic clone (Fig. 5). The first 3 products [Fig. 4(a), lanes 2 to 4], designated p3′31B4.b, p3′31B4.b1 and p3′31B4.b2 (Fig. 5 and Table 1), were subcloned. Digestion of the PCR-derived clones with *Hin*dIII (lanes 5 to 7) confirmed that

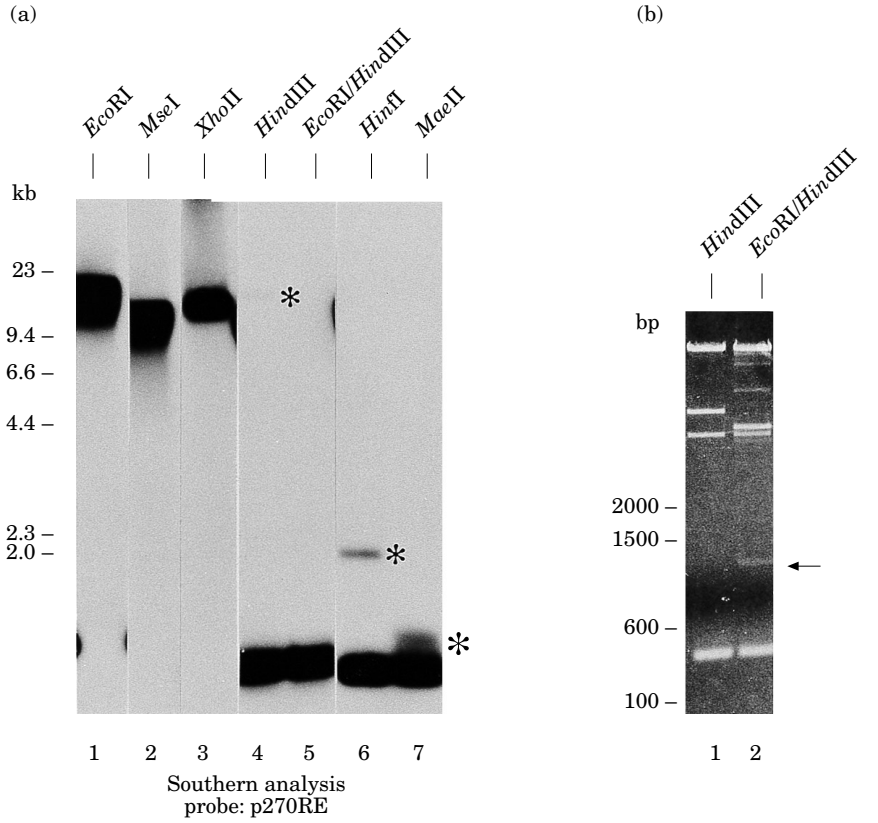


Figure 3. Characterization of the 31B4 clone. (a) Southern analysis was performed on 2.5 μg of purified recombinant DNA restricted to completion prior to electrophoresis in 1% agarose and blotting onto Zeta-probe membranes. Hybridization under high stringency was performed using ^{32}P-labelled p270RE as probe. Asterisks mark the bands carrying only a portion of the repeat unit. The large-sized hybridizing band for the *Hin*dIII-digested DNA is barely visible under these conditions due to the incomplete portion of the repeat accompanying the 3′ end fragment (Fig. 6). Numbers correspond to λDNA restricted with *Hin*dIII as size markers. (b) Restriction analysis was carried out on 5 μg of the purified recombinant DNA digested to completion. Treated DNA was then separated in 1% agarose gel and evaluated after ethidium bromide staining. The arrow indicates the fragment of ∼1.2 kb that was subcloned (p270.C3). Size markers correspond to the 100-bp DNA ladder, of which only some of the markers are shown for illustration purposes (GibcoBRL).

each had the same length of unique sequence. The increased length was due to multiples of the TRE in subclones, as evidenced by increased density of the 333 bp band (arrow). Finally, these inserts were sequenced [strategy shown in Fig. 5(d)] and confirmed the different lengths of repeat sequence (146 bp in p3′31B4.b, 479 bp in p3′31B4.b1 and 812 bp in p3′31B4.b2), each followed by 65 bp of non-repeat 3′-end sequence. The sequence of each of the three repeat elements was identical. Despite repeated attempts to transform amplified sequences into the *E. coli* host STBL2, designed to stabilize plasmids with multiple repeats, no additional larger clones were recovered, as before [9].

Sequencing the 65 bp unique DNA revealed no stop codon, suggesting that the 3′-end of the *p270* gene was missing. A cDNA library was then constructed from purified mRNA and the remainder of the 3′-end was derived by 3′-RACE (see Materials and methods). A gene-specific sense primer (P3) was designed that had 14 nucleotides of the repeat element and 13 nucleotides of the 65 bp non-repeat sequence (Fig. 6, underlined sequence *d*). Amplification yielded a PCR product of ∼1.2 kb [Fig. 4(b)] that was further subcloned and sequenced. Sequencing of 5 independent recombinant clones revealed an insert of 1267 nucleotides (Fig. 6, beginning with *d* sequence). Importantly, the first 79 nucleotides were identical with the 3′-sequence of the p3′31B4.b-clones. Fig. 6 shows the stop codon localized 1186 nucleotides downstream of the TRE. The 3′-UTR consisted of 38 nucleotides

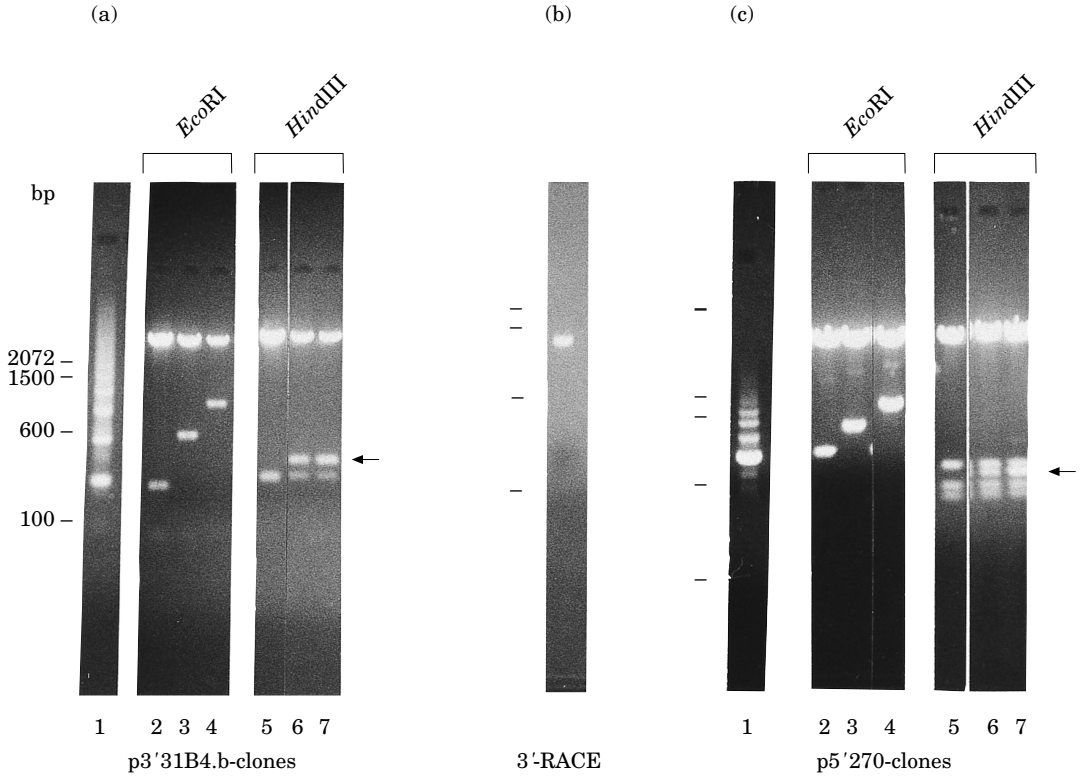


Figure 4. Subclones of the *p270* gene. (a) Amplification of the 3′-end of the 31B4 genomic insert with primer P2 yielded products with the increasing sizes (lane 1). Purified DNA of recombinant clones p3′31B4.b (lanes 2 and 5), p3′31B4.b1 (lanes 3 and 6) and p3′31B4.b2 (lanes 4 and 7) was digested to completion for electrophoresis in 1.2% agarose gel prior to ethidium bromide staining. The arrow points to the 333 bp repeat element band. (b) Amplification by 3′-RACE of the missing portion of the 3′-end. A cDNA library was generated using 1 μg of the purified mRNA used as template for amplification of the missing portion of the 3′-end with gene-specific sense primer P3. A 5.0 μl volume of the PCR reaction was eletrophoresed in 1.5% agarose gel and stained with ethidium bromide. (c) Amplification of the *p270* 5′-end (lane 1) with primer M1 again resulted in products with increasing sizes. A 5.0 μg amount of plasmid DNA purified from the recombinant clones p5′270.6 (lanes 2 and 5), p5′270.5 (lanes 3 and 6) and p5′270.19 (lanes 4 and 7) were digested to completion. After electrophoresis and staining, the 333 bp repeat element was evident (arrow). The 100-bp DNA ladder was used as size markers.

prior to the short poly-A tail, had a putative destabilizing element ATTTA [22] and no known polyadenylation signal. Importantly, primers were designed for amplification of the 3′-end portion of the gene from genomic DNA. Identical sequences were obtained for the cDNA and the numerous PCR products of genomic DNA that were subcloned and sequenced.

Subcloning and sequencing of the 5′-end region of the *p270* gene

The 1.2 kb fragment derived from *Eco*RI/*Hin*dIII double digestion of the 31B4 clone [Fig. 3(b)] was eluted from the gel and subcloned. Sequencing of the subclone, termed p270.C3 [Fig. 5(c) and Table 1] revealed the AT-rich 1107 bp sequence with no apparent ORF. Southern analysis of the restricted 31B4 was done using both the p270RE and p270.C3 as probes, and the banding patterns indicated the existence of additional ~400 bp DNA. Primers from p270.C3 sequence (P1) and the repeat element (M1) (see Materials and methods) were then used for PCR amplification of the missing portion of the genomic clone. Fig. 4(c) (lane 1) again shows a typical PCR reaction with the generation of an equidistant ladder pattern of up to nine bands with a prominent band at ~900 bp. Lanes 2 to 4 further compare sizes of the inserts after *Eco*RI restriction of the recovered clones p5′270.6, p5′270.5 and p5′270.19. Digestion with *Hin*dIII confirmed the presence of larger TRE, as evidenced by increased band density

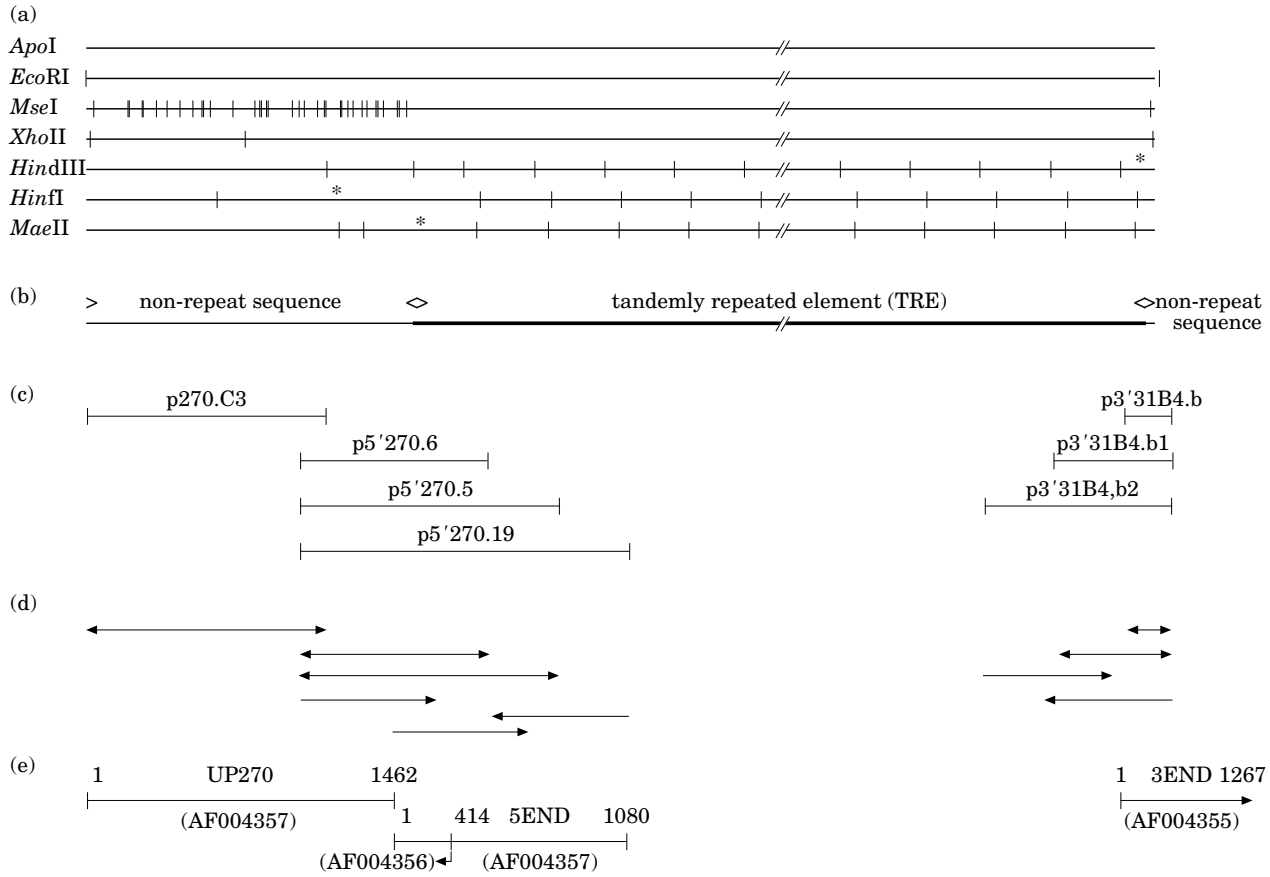


Figure 5. The genomic clone 31B4 containing a portion of *p270.* (a) The restriction maps were generated using 31B4 DNA restricted with endonucleases that either do not cut (*Apo*I, *Eco*RI, *Mse*I and *Xho*II) or cut (*Hin*dIII, *Hin*fI and *Mae*II) within the TRE. Asterisks designate the additional non-repeat fragments that hybridized with p270RE [Fig. 3(a)]. (b) Deduced model of 31B4 showing the estimated lengths of the TRE and non-repetitive sequence. (c) Subcloning strategy for the 5′- and 3′-end unique sequence. The 5′-end 1.2 kb fragment resulting from *Eco*RI/*Hin*dIII double digestion was subcloned and termed p270.C3. Primers to this DNA and to the repeat were then used for PCR where recombinant clones (p5′270.6, p5′270.5 and p5′270.19) carrying different numbers of repeat elements were recovered (Fig. 4). The 3′-end of the 31B4 genomic clone insert was amplified using primers to the EMBL4 arm and to the repeat. Clones (p3′31B4.b, p3′31B4.b1 and p3′31B4.b2) of different lengths were recovered. (d) The recombinant 5′- and 3′-end clones were sequenced using this strategy. (e) Accession numbers of the important non-repeat sequences are given.

(arrow) accompanied by additional sequence of the same size in all subclones. All inserts were further sequenced using the strategy in Fig. 5(d). The smallest clone termed p5′270.6 consisted of 888 nucleotides; the first 123 nucleotides represented overlap with p270.C3 sequence. The 3′-end of the insert contained the 345 nucleotides of the TRE sequence. This yielded a predicted 415 nucleotides of non-repeat sequence. As expected, the other two clones labelled p5′270.5 and p5′270.6 [Fig. 5(c), Table 1] contained the first 888 nucleotides identical to that of the smallest clone (p5′270.6) that were immediately followed by 333 and 666 bp of the repeat sequence, respectively.

Figure 8 shows the reading frame based on that established before for the repeat [9]. Of two possible start codons, the first ATG was preceded by 12 nucleotides (TCATTTTTAATA) that exhibited high homology (10 of 12 nucleotides) to a consensus sequence identified in other trichomonad protein-coding genes [23–26]. This putative start codon sequence was followed by 69 bp of non-repeat sequence prior to the start of the TRE region. It is noteworthy that, as found for the 3′-end, all three repeat elements of the 5′-end were identical in sequence.

Analysis of the non-coding region of 1462 nucleotides (p270.C3 and non-coding portion of

Table 1. List and designation of clones used to complete the sequencing of the *p270* gene and surrounding flanking regions

Clone designation[a]	Length of clone (bp)	Source of DNA	Vector used for clone	Clone characteristic
p270RE[b]	333	cDNA	pUC18	repeat element
31B4	~14 000	genomic	EMBL4	
p270.C3	11 107	genomic	pB SK(+)[c]	noncoding
p5′270.6	888	PCR	pCR2.1	unique 5′-end of the gene + 1 repeat
p5′270.5	1 221	PCR	pCR2.1	unique 5′-end of the gene + 2 repeats
p5′270.19	1 554	PCR	pCR2.1	unique 5′-end of the gene + 3 repeats
p3′31B4.b	211	PCR	pCR2.1	0.45 repeat + 65 bp unique 3′-end
p3′31B4.b1	544	PCR	pCR2.1	1.45 repeat + 65 bp unique 3′-end
pB3′31B4.b2	877	PCR	pCR2.1	2.45 repeat + 65 bp unique 3′-end
p3′270.7	1 167	cDNA	pCR2.1	3′-end unique sequence of the gene

[a] The clone designations are also represented in Fig. 5.
[b] The p270RE sequence represents the 333 bp repeat element reported previously [9].
[c] pB SK(+) refers to pBluescript SK(+) plasmid.

p5′270.6) immediately upstream of the *p270* gene ATG codon showed the sequence to have a high AT/GC ratio of 3.35. The AT/CG composition of the sequence resembled other non-coding sequences reported to be specific for *T. vaginalis*, that are present in numerous copies in the genome [27]. A homology search was carried out to compare this region with the available sequences in PCGENE data bank. No match was found with regulatory or promoter elements or protein-coding sequences.

Southern analysis using newly generated clones

All new subclones were then used for molecular analysis of the *p270* gene. Figure 7 shows a single band in autoradiograms when p5′270.6, a representative of the 5′-end coding region, was used as probe (lane 1, arrow). This band was also detected with ^{32}P-labelled p270RE as probe. Surprisingly, more than one fragment cross-hybridized under high stringency with the 3′-end region of the *p270* gene (lane 3), providing evidence for the existence of the similar sequence at other genomic sites. Not unexpectedly, the multiple banding patterns visualized when the AT-rich sequence was the probe confirmed the existence of AT-rich regions throughout the genome [27].

Coding region analysis

Several features make the *p270* gene different from other *T. vaginalis* genes. This sequence has a high AT content of 65.1%, compared to other trichomonad coding genes that have an average of 52% based on comparisons with 47 known coding sequences. The codon usage of the *p270* gene follows the *T. vaginalis* coding bias [28]. Three out of four arginines of repeat element are coded by the AGA codon, a rare tRNA in *E. coli* [29]. Expression of the heterologous proteins with this codon used for arginine could result in the inhibition of both the protein synthesis and the cell growth and could affect viability of the host cells [30]. These data may be relevant to the difficulties in expressing of the larger clones of P270 in *E. coli*, as indicated above.

Predicted properties of the protein P270

Finally, both 5′- and 3′-end coding regions of the protein P270 were analysed. Figures 6 and 8 present the predicted amino acid sequence of the protein. The amino-terminal non-repeat domain of the protein (Fig. 8) consists of 23 amino acids of which the first 11 have a similarity to the putative trichomonad signal sequence [23–25,31]. This 23 amino-terminal unique peptide is the most basic part of the protein, which as a discrete unit has a predicted pI of 10.01. All 5′- and 3′-end repeat elements were identical, showing the repeat to be 111 amino acids with a pI of 4.18 and an M_r of ~11.8 kDa, corresponding to the observed size of the fragment in denaturing SDS-PAGE gels [9, 21]. The last repeat peptide element in the TRE region was incomplete (318 instead of 333 nucleotides),

thereby missing five amino acids. Each repeat peptide element possessed the epitope DREGRD sequence, as previously predicted [9]. Although the exact number of repeats remains unknown, the combined M_r of the protein based on the calculated sizes of the non-repeat regions and that of repeat element suggests at least 18 repeats, which would result in an observed protein size of 270 kDa.

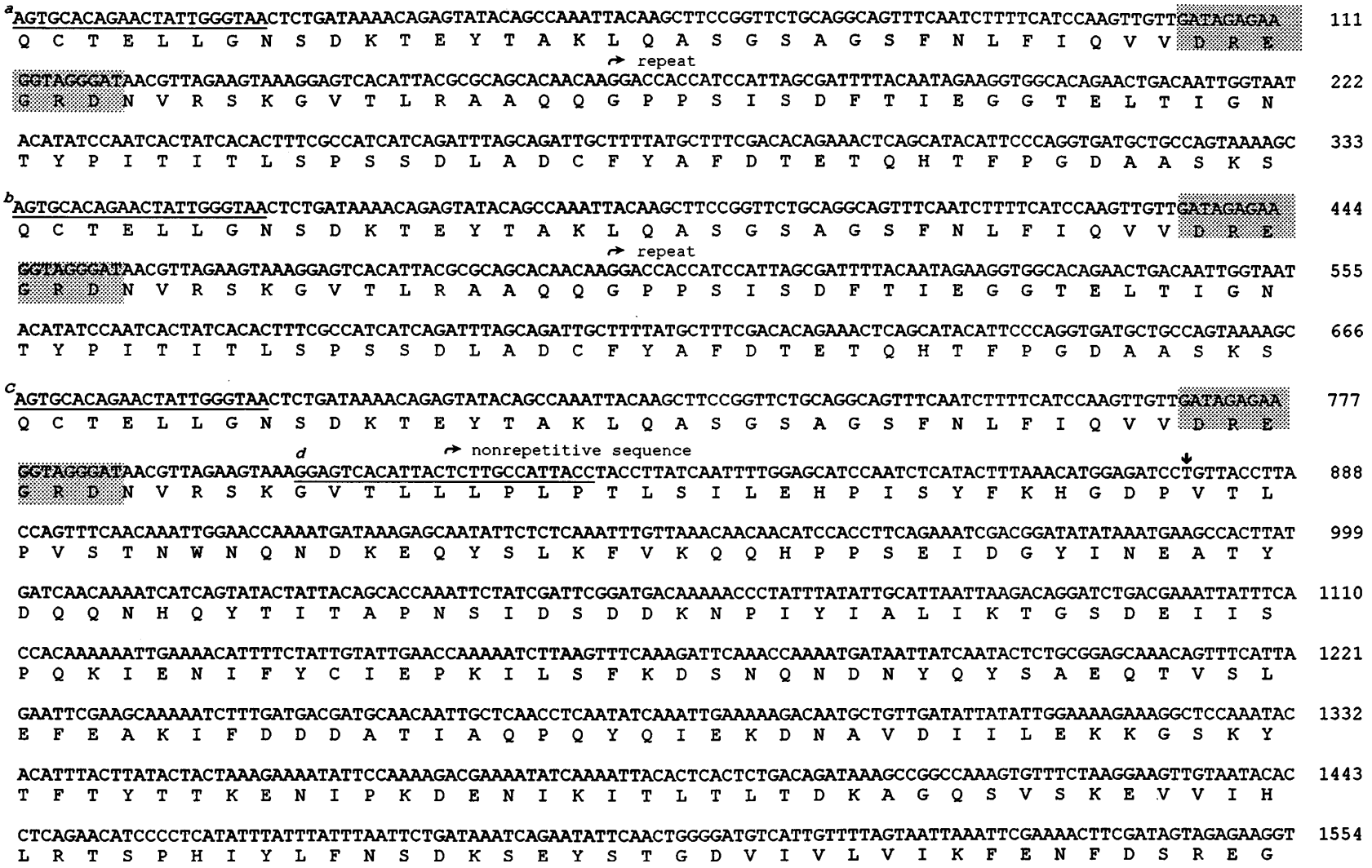
Finally, the carboxy-terminal non-repeat domain of the protein was 395 amino acids with a pI of 4.47. This domain has a putative transmembrane sequence VLLGIVIIVLIVYFVFC (Fig. 6), which may be involved in the membrane anchoring of P270. In this case, the remaining 16 amino acids downstream represent the cytoplasmic domain of the protein, and most of the protein would be exposed on the surface, a suggestion consistent with experimental evidence [10–12, 21]. Evaluation of the complete amino acid sequence allowed for us to analyse and compare in greater detail the repeat peptide elements. In all of three possible peptide elements, the 5′-end, 3′-end and internal repeats, the hydrophilicity plots [32] confirmed the existence of the same immunogenic epitope. No difference was found in accessibility of the epitope, chain flexibility and hydropathy index (data not shown) [33–35].

Discussion

The capability of invading microbial pathogens to alternate between surface versus non-surface placement of immunogens represents one of the immune evasion strategies and is considered a virulence determinant. In *T. vaginalis*, this ability clearly plays an important role in the parasite–host relationship. Parasites without surface P270 avoid the complement-independent killing by monoclonal and polyclonal antibody to P270 [10]. While the function of P270 remains unknown, the presence of P270 in all isolates of *T. vaginalis* from throughout the world, remaining unchanged in trichomonads grown *in vitro* for up to twelve years, strongly suggests its important role within the parasite. Importantly, although 50% of isolates are capable of phenotypic variation, only a small number of trichomonads examined directly from patients express P270 on the surface, suggesting either that *in vivo* environmental factors results in preferable non-surface placement of the P270 [12] or that surface-expressing parasites are eliminated by the host antibody. Of particular interest was the observation a decade ago that only isolates harbouring a dsRNA virus undergo phenotypic variation [13]. About 60% of the virus-infected isolates lose the virus soon after axenization and daily passage [13, 14, 36], suggesting that, *in vivo*, conditions are favourable for maintenance of virus infection. Nonetheless, loss of virus results in an inability for parasites to undergo phenotypic variation for P270 [9, 13, 14], showing the possible involvement of viral infection in this property. Based on these observations, we continued the study of the *p270* gene in a fresh clinical isolate that harbours a dsRNA trichomonad virus. We showed and confirmed all previously known facts on P270, including the gene organization among fresh isolates.

In this report we show the strategy for cloning and sequencing of the non-repeat coding regions of the gene and of the non-coding upstream region. We determined that the *p270* gene is preceded by an AT-rich region without an apparent reading frame that displayed homology with multiple regions of the genome as previously reported, and which are believed to represent intergenic non-coding regions of the trichomonad DNA [27]. No known regulatory or promoter elements were found within this 1462 bp upstream sequence. A 12 nucleotide region immediately upstream of the start codon displayed homology with the conserved consensus sequence found in other protein-coding genes of *T. vaginalis* [23–26]. This sequence has been recently referred to as a transcriptional initiator (Inr), which is a second type of core promoter-like element [26]. Importantly, it is also similar to an internal transcribed spacer (ITS), with such sequences being relatively short and diverse among amitochondrial protozoa [39]. The ITS includes one or two predominant nucleotides and has minimal potential secondary structure, which may form the basis for preferential processing of ITS sequences. The motif TCA or TTCA present within the ITS is complementary to the TGAA sequence found in the loop of the terminal hairpin of rRNA sequences [37] and is recently reported to be a ribosome-binding domain, perhaps suggesting an alternative or dual function for this upstream sequence.

The N-terminal non-repeat portion of P270 resembled the leader sequences previously described for other trichomonad proteins, in-


a
AGTGCACAGAACTATTGGGTAACTCTGATAAAACAGAGTATACAGCCAAATTACAAGCTTCCGGTTCTGCAGGCAGTTTCAATCTTTTCATCCAAGTTGTTGATAGAGAA 111
Q C T E L L G N S D K T E Y T A K L Q A S G S A G S F N L F I Q V V D R E
repeat
GGTAGGGATAACGTTAGAAGTAAAGGAGTCACATTACGCGCAGCACAACAAGGACCACCATCCATTAGCGATTTTACAATAGAAGGTGGCACAGAACTGACAATTGGTAAT 222
G R D N V R S K G V T L R A A Q Q G P P S I S D F T I E G G T E L T I G N
ACATATCCAATCACTATCACACTTTCGCCATCATCAGATTTAGCAGATTGCTTTTATGCTTTCGACACAGAAACTCAGCATACATTCCCAGGTGATGCTGCCAGTAAAAGC 333
T Y P I T I T L S P S S D L A D C F Y A F D T E T Q H T F P G D A A S K S
b
AGTGCACAGAACTATTGGGTAACTCTGATAAAACAGAGTATACAGCCAAATTACAAGCTTCCGGTTCTGCAGGCAGTTTCAATCTTTTCATCCAAGTTGTTGATAGAGAA 444
Q C T E L L G N S D K T E Y T A K L Q A S G S A G S F N L F I Q V V D R E
repeat
GGTAGGGATAACGTTAGAAGTAAAGGAGTCACATTACGCGCAGCACAACAAGGACCACCATCCATTAGCGATTTTACAATAGAAGGTGGCACAGAACTGACAATTGGTAAT 555
G R D N V R S K G V T L R A A Q Q G P P S I S D F T I E G G T E L T I G N
ACATATCCAATCACTATCACACTTTCGCCATCATCAGATTTAGCAGATTGCTTTTATGCTTTCGACACAGAAACTCAGCATACATTCCCAGGTGATGCTGCCAGTAAAAGC 666
T Y P I T I T L S P S S D L A D C F Y A F D T E T Q H T F P G D A A S K S
c
AGTGCACAGAACTATTGGGTAACTCTGATAAAACAGAGTATACAGCCAAATTACAAGCTTCCGGTTCTGCAGGCAGTTTCAATCTTTTCATCCAAGTTGTTGATAGAGAA 777
Q C T E L L G N S D K T E Y T A K L Q A S G S A G S F N L F I Q V V D R E
d
nonrepetitive sequence
GGTAGGGATAACGTTAGAAGTAAAGGAGTCACATTACTCTTGCCATTACCTACCTTATCAATTTTGGAGCATCCAATCTCATACTTTAAACATGGAGATCCTGTTACCTTA 888
G R D N V R S K G V T L L L P L P T L S I L E H P I S Y F K H G D P V T L
CCAGTTTCAACAAATTGGAACCAAAATGATAAAGAGCAATATTCTCTCAAATTTGTTAAACAACAACATCCACCTTCAGAAATCGACGGATATATAAATGAAGCCACTTAT 999
P V S T N W N Q N D K E Q Y S L K F V K Q Q H P P S E I D G Y I N E A T Y
GATCAACAAAATCATCAGTATACTATTACAGCACCAAATTCTATCGATTCGGATGACAAAAACCCTATTTATATTGCATTAATTAAGACAGGATCTGACGAAATTATTTCA 1110
D Q Q N H Q Y T I T A P N S I D S D D K N P I Y Y I A L I K T G S D E I I S
CCACAAAAAATTGAAAACATTTTCTATTGTATTGAACCAAAAATCTTAAGTTTCAAAGATTCAAACCAAAATGATAATTATCAATACTCTGCGGAGCAAACAGTTTCATTA 1221
P Q K I E N I F Y C I E P K I L S F K D S N Q N D N Y Q Y S A E Q T V S L
GAATTCGAAGCAAAAATCTTTGATGACGATGCAACAATTGCTCAACCTCAATATCAAATTGAAAAAGACAATGCTGTTGATATTATATTGGAAAAGAAAGGCTCCAAATAC 1332
E F E A K I F D D D A T I A Q P Q Y Q I E K D N A V D I I L E K K G S K Y
ACATTTACTTATACTACTAAAGAAAATATTCCAAAAGACGAAAATATCAAAATTACACTCACTCTGACAGATAAAGCCGGCCAAAGTGTTTCTAAGGAAGTTGTAATACAC 1443
T F T Y T T K E N I P K D E N I K I T L T L T D K A G Q S V S K E V V I H
CTCAGAACATCCCCTCATATTTATTTATTTAATTCTGATAAATCAGAATATTCAACTGGGGATGTCATTGTTTTAGTAATTAAATTCGAAAACTTCGATAGTAGAGAAGGT 1554
L R T S P H I Y L F N S D K S E Y S T G D V I V L V I K F E N F D S R E G

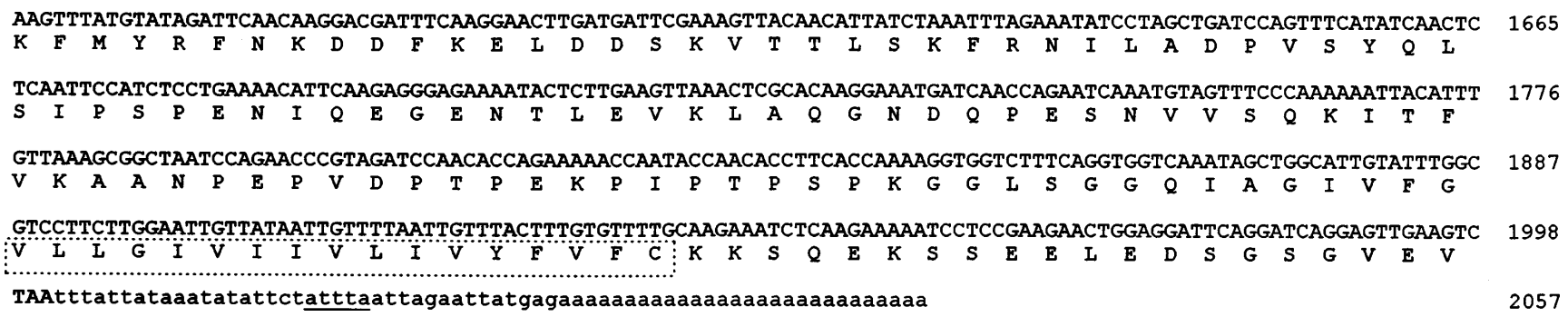


Figure 6. The DNA sequence of the 3′-end of the *p270* gene. The underlined sequences [a,b] and [c] represent the sequence of the primer P2 determining the 5′-end of the recovered clones p3′31B4.b, p3′31B4.b1 and p3′31B4.b2, respectively (Fig. 5). The arrow (↓) points to the end of the 65 bp non-repetitive sequence contained within these clones. Underlined sequence [d] represents the primer P3 used in 3′-RACE. The deduced amino acid sequence of the corresponding portion of the P270 protein is presented under the DNA sequence. The shaded sequences represent the nucleotides coding the immunodominant epitope found in all repeats. The C-terminal domain of the protein contains a putative transmembrane region (under the 'dotted box'). The motif ATTTA (underlined) is an mRNA destabilizing element [22].

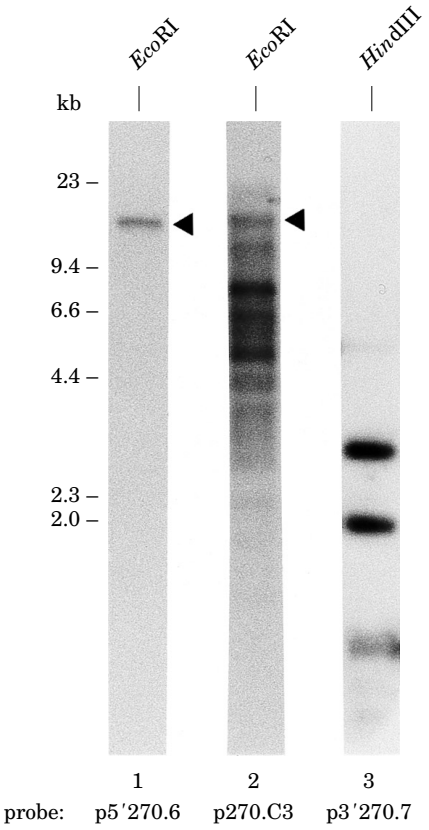


Figure 7. Southern analysis using non-repeat regions of the *p270* gene. A 5.0 μg amount of purified trichomonad DNA was digested to completion with designed enzymes and separated in 0.7% agarose. After blotting onto Zeta-probe membrane, hybridization was performed using as probes nick-translated p5′270.6 (5′-end unique DNA) (lane 1), p270.C3 (the 5′-end AT-rich sequence) (lane 2) and p3′270.7 (3′-end unique DNA) (lane 3). Arrow points to the fragment that previously hybridized with p270RE. Numbers correspond to size markers of *Hin*dIII-restricted λDNA. Not shown, are the size markers of the 100 bp ladder that was run simultaneously for more accurate size determinations of the smaller bands.

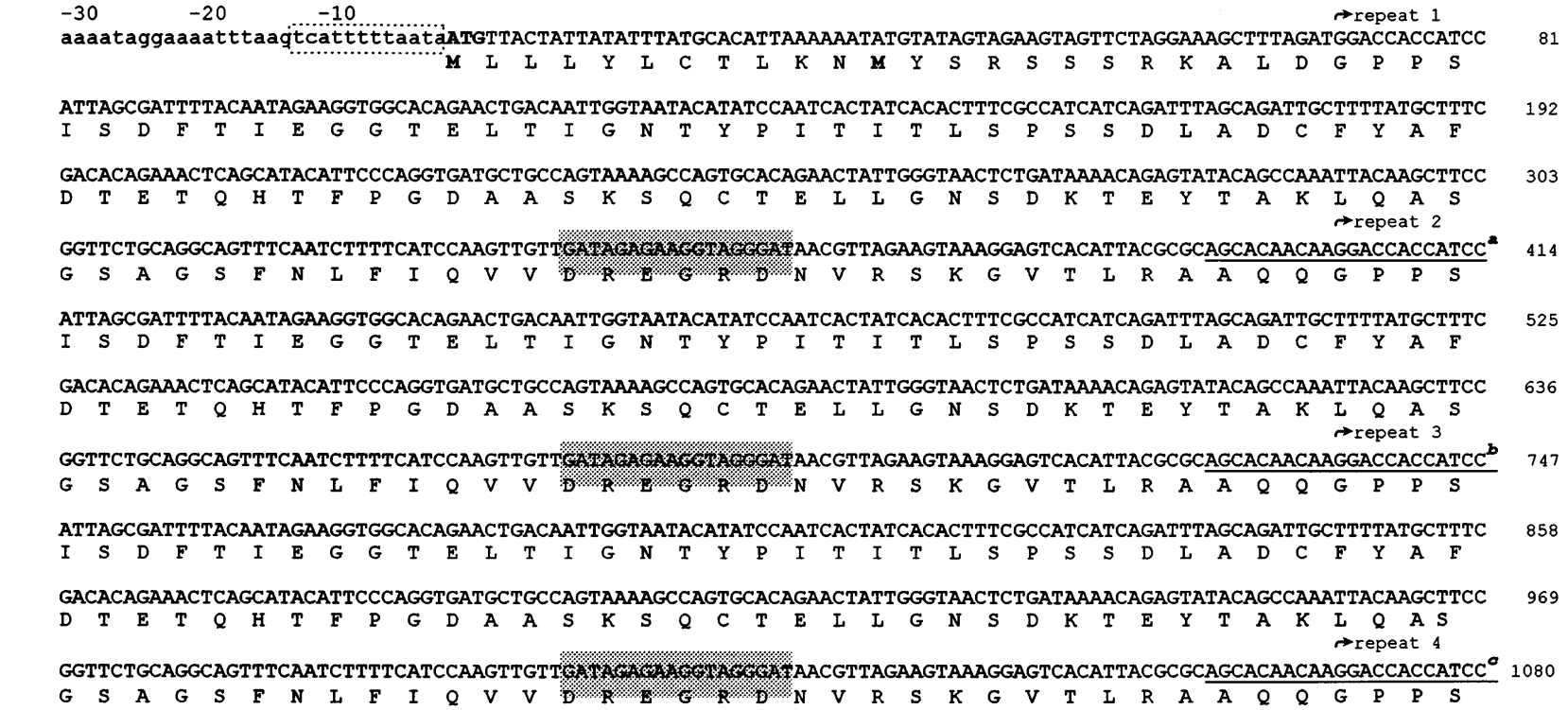
cluding the adhesins [23–25, 28] and hydrogenosomal enzymes [26, 31]. This observation now shows the existence of similar or identical leader sequences for proteins targeted to both vacuoles and surfaces. Of interest was the inability to express in *E. coli* the short 23 amino acid amino-terminal portion of P270 with or without a repeat element, despite the fact that the fragment of the TRE (382 bp) has been successfully expressed in *E. coli* [9]. It has been shown that the cloned mitochondrial leader sequence of eukaryotic proteins decreases the expression of the recombinant proteins in *E. coli*, apparently due to the secondary structure of the mRNA [38]. Lack of expression of this unique region and our past unsuccessful efforts at cloning the *p270* gene may be indicative of toxicity of this peptide to the bacterium. This notion is reinforced by the fact that all subclones of the 5′-end (~200) that were recovered had the DNA insert in the opposite orientation.

The non-repeat region of the carboxy-terminus is 395 amino acids in length, and it is in this region where a transmembrane domain is predicted. Protease treatment of parasites with surface-bound P270 results in release of almost the entire protein, affirming that the majority of the protein is extracellular. This is also consistent with the binding by monoclonal antibodies to exposed epitopes of the TRE portion of P270 [9, 10]. Further, binding of P270 by monoclonal antibodies on the parasite surface inhibited P270 autodegradation by the trichomonad cysteine proteinases, suggesting the extracellular localization of the TRE domain of the protein. These data reinforce the idea that P270 may be positioned on the cell surface through this membrane-spanning domain.

Southern analysis performed with the clone representing the 5′-end with one repeat element gave a single band, confirming that the repeat element sequence is unique within the trichomonad genome and that the *p270* gene is single copy. We found that the 3′-end non-repeat region of the gene shares homology with some other regions, a finding confirmed in additional isolates tested. Numerous clones representing both cDNA and genomic PCR products were sequenced and found to be identical. This is as expected, because the gene-specific primers were used for 3′-RACE and genomic amplifications. While we did not succeed in determining the exact number of repeats in the TRE, we showed that there are at least 18 elements through Southern analysis using limited-digestion of DNA. Nonetheless, the organization of the *p270* gene that is presented in this report (Fig. 1) and which was generated from a fresh clinical isolate of *T. vaginalis* is consistent with that proposed earlier [9].

Finally, the entire P270 protein was compared to existing amino acid sequences (PCGENE data bank). No significant homology was found with known proteins; however, among the 20 proteins with the best similarity scores, nine proteins were receptors and one protein belonged to the immunoglobulin superfamily. Representatives

(a)



(b)

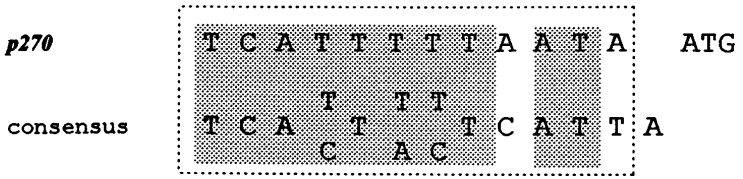


Figure 8. The DNA sequence of the 5′-end of the *p270* gene. Underlined sequences [a,b] and [c] correspond to the sequence of the primer M1 that was used for PCR amplification which yielded clones p5′270.6, p5′270.5 and p5′270.19, respectively (Fig. 5). The deduced amino acids corresponding to the nucleotide sequence are shown, and the shaded areas illustrate the DREGRD epitope conserved within each repeat. (b). The 12 nucleotides immediately preceding the ATG start codon of the *p270* gene [boxed sequence in (a)] are compared with the consensus DNA sequence reported for other trichomonad protein-coding genes [23–26,31].

of these two groups of proteins include the biliary glucoprotein 1, glutamate receptor 1, thyrotropin receptor, glutamate receptor 2 and the T-cell surface glycoprotein CD11. Interestingly, the biliary glycoprotein 1 of mice and the other proteins have been shown to serve as virus receptors [39, 40]. Further comparisons revealed that all proteins are type I surface proteins with a cytoplasmic C-terminal domain that might have enzymatic properties.

We believe that this work now provides the foundation for answering important questions. It will now be possible to establish whether the molecular size polymorphism of P270 is due to the number of repeats, especially since data collected to date suggest high conservation of both 5′- and 3′-unique regions among isolates. Second, 3′-end coding non-repeat portion of the *p270* gene (p3′270.7) crosshybridized with other regions of the trichomonad genome showing that similar, if not identical, sequences are present. Since all amplifications were carried out with gene-specific primers, it is not clear whether these sequences are transcribed and what is the level of identity. It would not be surprising if this domain of the protein was found essential to phenotypic variation, as P270 is one of a repertoire of proteins with this property [11]. Third, the 5′- and 3′-UTR sequences reveal similarities with other known trichomonad gene sequences, suggesting further involvement of these sequences as regulatory elements acting at the level of transcription or mRNA stability. In fact, the presence of the same Inr sequence on *p270* as found in iron-regulated adhesin genes makes it essential that environmental regulation of *p270* be examined. Lastly, of interest will be whether we can ultimately get expression of the amino- and carboxy-terminal domains of the proteins in *E. coli*, as this would provide the necessary reagents to prove conclusively that all antibody in patient sera toward P270 is directed solely to the DREGRD epitope.

Materials and methods

Microorganisms

T. vaginalis isolate T068-II is a fresh clinical isolate that harbours the dsRNA virus and has been used previously by our laboratory [14]. Trichomonads were grown in Trypticase-yeast extract-maltose (TYM) medium supplemented with 10% heat-inactivated horse serum (HIHS) [42] to the late-logarithmic phase of growth before harvesting. Recombinant *Escherichia coli* strains were cultured in Luria-Betani (LB) broth or LB agar with 60 μg of ampicillin/ml [42].

Nucleic acid isolation from *T. vaginalis*

Total genomic trichomonad DNA was isolated by standard procedures as previously detailed [23]. Briefly, 10^8 cells were lysed with 0.2% SDS in buffer consisting of 100 mM NaCl, 10 mM EDTA and 10 mM Tris-HCl, pH 8.0 before extracting four times with phenol-chloroform/isoamyl alcohol mixture. After ethanol precipitation, the DNA was resuspended in TE buffer (1 mM EDTA and 10 mM Tris-HCl, pH 8.0) containing 200 μg of RNase A/ml for 10 min at 37°C. Then, proteinase K (100 μg/ml final concentration) was added prior to incubation for 45 min at 37°C. DNA was phenol-chloroform extracted and precipitated again by ethanol and dissolved in TE buffer for storage at 4°C.

For total RNA isolation, 10^8 parasites were washed and immediately lysed by addition of a mixture of 4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% sarcosyl and 100 mM 2-mercaptoethanol (solution D). This was followed by the sequential addition of 2 M sodium acetate, pH 4.0, acid phenol and a chloroform–isoamyl alcohol mixture [14, 43]. The lysate was mixed by inversion after each addition followed by shaking vigorously for 10 sec and cooling on ice for 15 min. Samples were centrifuged at $10\,000 \times g$ for 10 min at 4°C, and the RNA in the aqueous phase was precipitated by ethanol. The RNA pellet was dissolved in solution D, reprecipitated and rinsed before dissolving in 0.5% SDS at 65°C for 10 min. The mRNA of isolate T068-II was isolated from 10^8 cells using the FastTrack® 2.0 Kit (Invitrogen, Carlsbad, CA) as recommended.

Recombinant DNA purification

Recombinant plasmid DNA of all clones and subclones was purified using QIAprep Spin Plasmid Kit (Qiagen, Santa Clarita, CA). For Southern and Northern hybridization analysis, inserts of plasmids cleaved with appropriate enzymes were eluted from the agarose gels by QIAEX II Gel Extraction Kit (Qiagen).

Generation of genomic library

A genomic library was constructed in the λ replacement vector EMBL4, as described by the manufacturer (Promega, Madison, WI). Trichomonad DNA was partially digested with *Sau3*A and separated on a 0.55% agarose gel. Fragments in the range of 9–18 kb were eluted for coprecipitation and ligation with vector. After packaging, the recombinant phage was grown in *E. coli* NM539, which is used for *spi* selection. The library was screened by standard hybridization procedures [14, 23, 42], and a recombinant phage, termed 31B4, hybridized to the repeat element of the *p270* gene [9]. This recombinant phage was purified (QIAGEN Lambda Starter Kit, Qiagen) for further mapping and subcloning.

Subcloning of genomic DNA fragments

Purified DNA of the 31B4 clone was completely digested with *Eco*RI and *Hin*dIII and electrophoresed in 1% agarose. A readily visible band of ~1.2 kb was eluted, ligated into pBluescript SK^+ and transformed into *E. coli* XL1-Blue as recommended by manufacturer (Strategene, La Jolla, CA).

DNA sequencing and analysis

DNA sequencing was performed by the dideoxy chain termination method using the Sequenase 2.0 DNA Sequencing kit (United States Biochemical, Cleveland, OH) as recommended by the manufacturer. Computer analysis of both DNA and amino acid sequences was carried out using PC/GENE (release 6.8) (IntelliGenetics, Inc., Mountain View, CA).

PCR amplification

For all amplifications, the hot start (94°C for 1 min) PCR procedure was performed using TaqStart Antibody (CLONTECH, Palo Alto, CA). All amplified products were then ligated and transformed to *E. coli* INVαF′ cells using the Original TA Cloning® Kit (Invitrogen Cor., Carlsbad, CA).

For isolation of the non-repeat coding and non-coding DNA at the 5′-end of the *p270* gene, the primers P1 (5′-GAATTACTCGGAAGCAATTTAACCC) and M1 (5′-GGATGGTGGTCCTTGTTGTGC) were used. P1 was designed based on the sequence of the p270.C3 subclone (Table 1) and antisense primer M1 was synthethized to bind repeat sequence [9]. Amplification was performed in 20 cycles of 94°C for 30 s, 53°C for 30 s and 68°C for 5 min. The 3′-end non-repeat coding sequence contained within genomic clone 31B4 was derived by PCR amplification using primer P2 (5′-CAGTGCACAGAACTATTGGG) corresponding to the repeat element and primer M2 designed to bind at the left arm of EMBL4 (5′-GCAACTCGTGAAAGGTAGGC). This target sequence was amplified in 25 cycles with 94°C for 30 s, 52°C for 30 s and 68°C for 5 min.

Generation and isolation of cDNA

The 3′-ampliFINDER RACE kit (CLONTECH Laboratories, Inc., Palo Alto, CA) was used for cDNA synthesis using 1 μg of mRNA as described by the manufacturer. The PCR was performed using the AP1 primer (Marathon cDNA Amplification Kit) and the gene-specific sense primer P3 (5′-AGGAGTCACATTACTCTTGCCATTACC). Hot start (TaqStart Antibody, 94°C for 1 min) PCR was done with 25 cycles of 94°C for 30 s, 53°C for 30 s and 68°C for 5 min. The amplified PCR product was ligated and transformed into *E. coli* INVαF′ (Invitrogen).

Southern and Northern hybridization

For Southern analysis, trichomonad genomic DNA or recombinant DNA was digested with specific restriction enzymes and transferred to a Zeta-probe membrane (Bio-Rad Laboratories, Richmond, CA) as recommended by the manufacturer. Hybridizations were carried out with nick-translated probes (Promega). Blots were prehybridized (50% formamide, 120 mM Na_2HPO_4, 250 mM NaCl, 7% SDS and 1 mM EDTA) for 4 h at 42°C. Hybridization was performed at 42°C for 18 h. Membranes were then sequentially washed with 2X SSC (1X SSC is 250 mM NaCl and 15 mM Na_3-citrate, pH 7.0)-0.1% SDS, 0.5X SSC-0.1% SDS and 0.1X SSC-0.1% SDS at RT for 15, 30 and 30 min, respectively. For Northern analysis, the purified total RNA was electrophoresed on 1% agarose gel before transferring to a Zeta-probe membrane for probing

with nick-translated, ^{32}P-labelled probe under the same conditions as described above. Where size estimations were needed, different percentage agarose gels were always run with size markers for either large fragments (*Hin*dIII-digested λDNA and 1 kb DNA ladder) or small fragments (100 bp DNA ladder).

SDS-PAGE and immunoblotting

P270 of isolate T068-II was purified by the immunoprecipitation assay [44] using monoclonal antibody C20A3 [9–11]. Immune complexes recovered by protein A-Sepharose were then released by boiling in electrophoresis-dissolving buffer and proteins electrophoresed in 7% polyacrylamide gel using conditions previously established. After blotting onto nitrocellulose, P270 was then detected by incubation with the C20A3 monoclonal antibody [9, 14, 21].

Nucleotide sequence accession numbers

The obtained nucleotide sequences of the 5′- and 3′-coding regions of the *p270* gene and the upstream non-coding sequence have been submitted to the GenBank data base and been assigned GenBank accession numbers AF004356, AF004355 and AF004357, respectively [Fig. 5 (e)].

Acknowledgements

This study was supported by Public Health Service grant AI-39803 from the National Institute of Health. We are indebted to Michael W. Lehker, Ph.D. for his assistance in establishing the genomic library of *T. vaginalis* and to Jean Engbring for her assistance and spirited discussions on the molecular biology of P270.

References

1 World Health Organization. An overview of selected curable sexually transmitted diseases. WHO Global Programme on AIDS Report 1995.
2 Cotch NMF, Patorek II JG, Nugent RP, *et al*. *Trichomonas vaginalis* associated with low birth weight and preterm delivery. *Sex Trans Dis* 1997; **24**: 353–60.
3 Laga M, Nzila N, Goeman J. The interrelationship of sexually transmitted diseases and HIV infection: implications for the control of both epidemics in Africa. *AIDS* 1991; **(5 Suppl 1)**: 555–63.
4 Laga M, Manoka A, Kivuvu M, *et al*. Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: results from a cohort study. *AIDS* 1993; **7**: 95–102.
5 Nzila N, Laga M, Thiam MA, *et al*. HIV and other sexually transmitted diseases among female prostitutes in Kinsasha. *AIDS* 1991; **5**: 715–21.
6 Wasserheit JN. Interrelationship between human immunodeficiency virus infection and other sexually transmitted diseases. *Sex Trans Dis* 1992; **19**: 61–77.
7 Zhang Z, Begg CB. Is *Trichomonas vaginalis* a cause of cervical neoplasma? Results from a combined analysis of 24 studies. *Int J Epidemiol* 1994; **23**: 682–90.
8 Bertini B, Hornstein M. The epidemiology of trichomonosis and the role of this infection in the development of carcinoma of the cervix. *Acta Cytol* 1970; **14**: 325–32.
9 Dailey DC, Alderete JF. The phenotypically variable surface protein of *Trichomonas vaginalis* has a single, tandemly repeated immunodominant epitope. *Infect Immun* 1991; **59**: 2083–8.
10 Alderete JF, Kasmala L. Monoclonal antibody to a major glycoprotein immunogen mediates differential complement-independent lysis of *Trichomonas vaginalis. Infect Immun* 1986; **53**: 697–9.
11 Alderete JF. *Trichomonas vaginalis* NYH 286 phenotypic variation may be coordinated for repertoire of trichomonad surface immunogens. *Infect Immun* 1987; **55**: 1957–62.
12 Alderete JF, Demeš P, Gombošova A, *et al*. Phenotype and protein/epitope phenotypic variation among fresh isolates of *Trichomonas vaginalis. Infect Immun* 1987; **55**: 1037–41.
13 Wang A, Wang CC, Alderete JF. *Trichomonas vaginalis* phenotypic variation occurs only among trichomonas infected with the double-stranded RNA virus. *J Exp Med* 1987; **166**: 142–50.
14 Khoshnan A, Alderete JF. *Trichomonas vaginalis* with a double-stranded RNA virus has upregulated levels of phenotypically variable immunogen mRNA. *J Virol* 1994; **68**: 4035–8.
15 Gravekamp C, Horensky DS, Michel JL, Madoff LC. Variation in repeat number within the alpha C protein of group B streptococci alters antigenicity and protective epitopes. *Infect Immun* 1996; **64**: 3576–83.
16 Anderson BE, McDonald GA, Jones DC, Regnery RL. A protective protein antigen of *Richettsia rickettsii* has tandemly repeated, near-identical sequences. *Infect Immun* 1990; **58**: 2760–9.
17 Ibanez DF, Affranchino JL, Macrina RA, *et al*. Multiple *Trypanosoma cruzi* antigens containing tandemly repeated amino acid sequence motifs. *Mol Biochem Parasitol* 1988; **30**: 27–33.
18 Roditi L, Carrington M, Turner M. Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei. Nature* 1987; **325**: 272–4.
19 Wallis AE, McMaster WR. Identification of *Leishmania* genes encoding proteins containing tandemly repeating peptides. *J Exp Med* 1987; **166**: 1814–24.
20 Schneider A, Hemphill A, Wyler T, Seebeck T. Large microtubule-associated protein of *T. brucei* has tandemly repeated, near-identical sequences. *Science* 1988; **241**: 459–62.

21 Alderete JF, Neale KA. Relatedness of structures of a major immunogen in *Trichomonas vaginalis* isolates. *Infect Immun* 1989; **57**: 1849–53.
22 Zubiaga AM, Belasco JG, Greenberg ME. The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Mol Cell Biol* 1995; **15**: 2219–30.
23 Alderete JF, O'Brien JL, Arroyo R, *et al*. Cloning and molecular characterization of two genes encoding adhesion proteins involved in *Trichomonas vaginalis* cytoadherence. *Mol Microbiol* 1995; **17**: 69–83.
24 Alderete JF, Engbring JA, Lauriano CM, O'Brien JL. Only two of the *Trichomonas vaginalis* triplet AP51 adhesins are regulated by iron. *Microb Pathog* 1997; (in press).
25 O'Brien JL, Lauriano CM, Alderete JF. Molecular characterization of a third malic enzyme-like AP65 adhesin gene of *Trichomonas vaginalis. Microb Pathog* 1996; **20**: 335–49.
26 Quon DVK, Delgadillo MG, Khachi A, *et al*. Similarity between a ubiquitous promoter element in an ancient eukaryote and mammalian initiator elements. *Proc Natl Acad Sci USA* 1994; **91**: 4579–83.
27 Paces J, Urbankova P, Urbanek P. Cloning and characterization of a repetitive DNA sequence specific for *Trichomonas vaginalis. Mol Biochem Parasitol* 1992; **54**: 247–56.
28 Engbring JA, Alderete JF. Three genes encode distinct AP33 proteins involved in *Trichomonas vaginalis* cytoadherence. *Mol Microbiol* (in press).
29 Buck D, Spencer ME, Guest JR. Primary structure of the succinyl-CaA synthetase of *Escherichia coli. Biochem* 1985; **24**: 6245–52.
30 Zahn K. Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. *J Bact* 1996; **178**: 2926–33.
31 Lahti CJ, D'Oliveira CE, Johnson PJ. β-Succinyl-coenzyme A Synthetase from *Trichomonas vaginalis* is a soluble hydrogenosomal protein with an amino-terminal sequence that resembles mitochondrial presequences. *J Bact* 1992; **174**: 6822–30.
32 Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* 1981; **78**: 3824–8.
33 Janin J. Surface and inside volumes in globular proteins. *Nature* 1979; **277**: 491–2.
34 Bhaskaran R, Ponnuswamy PK. Positional flexibilities of amino acid residues in globular proteins. *Int J Peptide Protein Res* 1988; **32**: 241–55.
35 Eisenberg D, Schwarz E, Komavomy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* 1984; **179**: 125–42.
36 Wang AL, Wang CC. Viruses of the protozoa. *Ann Rev Microbiol* 1991; **45**: 251–63.
37 Katiyar SK, Visvesvara GS, Edlind TD. Comparisons of ribosomal RNA sequences from amitochondrial protozoa: implications for processing, mRNA binding and paromomycin susceptibility. *Gene* 1995; **152**: 27–33.
38 Ryan DG, Bridger WA. Bacterial expression of rat liver succinyl-CoA synthetase α-subunit. Factors that contribute to blocked translation of transcripts encoding a mitochondrial signal sequence. *J Mol Biol* 1991; **219**: 165–74.
39 Lentz TL. The recognition event between virus and host cell receptor: a target for antiviral agents. *J Gen Virol* 1990; **71**: 751–66.
40 Williams RK, Jiang GS, Snyder SW, *et al*. Purification of the 110-kilodalton glycoprotein receptor for mouse hepatitis virus (MHV)-A59 from mouse liver and identification of a nonfunctional, homologous protein in MHV-resistant SJL/J mice. *J Virol* 1990; **64**: 3817–23.
41 Diamond LS. The establishment of various trichomonads of animals and man in axenic cultures. *J Parasitol* 1957; **43**: 488–90.
42 Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.
43 Chomzynski P, Sacchl N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987; **162**: 156–9.
44 Alderete JF. Antigen analysis of several pathogenic strains of *Trichomonas vaginalis. Infect Immun* 1983; **39**: 1041–7.