Isoform Diversity and Tandem Duplication of the Glycoprotein A Gene in Ferret *Pneumocystis carinii*

Terry W. Wright, Terrence Y. Bissendial, Constantine G. Haidaris, Francis Gigliotti, and Patricia J. Simpson Haidaris*

Departments of Microbiology and Immunology, Medicine/Hematology Unit, Dental Research, Pediatrics, and Pathology and Laboratory Medicine, University of Rochester School of Medicine and Dentistry, Box 610, 601 Elmwood Ave. Rochester, NY 14642, USA

Received 14 March 1995; revised 24 April 1995

Abstract

Two ferret *P. carinii* gpA cDNA clones were identified that reacted identically with a panel of anti-gpA monoclonal antibodies, although their nucleotide sequences were 22% divergent. Each clone hybridized to a single mRNA species of 3,600 nucleotides only in *P. carinii*-infected lung mRNA, but RT-PCR analysis demonstrated that these cDNA clones were derived from two distinct gpA mRNA transcripts. Further PCR analysis demonstrated that the ferret *P. carinii* genome contains at least two gpA genes lying in tandem on a single chromosome separated by a 329-bp intergenic region. Based on the terminal gene sequences of this tandem repeat and the cDNA clones, a composite full-length ferret *P. carinii* gpA coding sequence was constructed. The intergenic region immediately downstream of the stop codon of the first gpA gene contains three putative polyadenylation signals, and constitutes the 3' untranslated region (UTR) of the gpA mRNA. Primer extension of the gpA mRNA resulted in products extending 74 and 244 nucleotides into the 5' UTR. However, the intergenic region lying greater than 25 nucleotides upstream of the first methionine of the second gpA gene was found to be absent from the 5' UTR.

Key words: cDNA; gene organization; polyadenylation; *Pneumocystis carinii* pneumonia; oligonucleotide

1. Introduction

AIDS patients are particularly susceptible to recurrent episodes of *Pneumocystis carinii* pneumonia (PCP). Since *P. carinii* cannot be cultured in vitro, steroid-treated ferrets and rats, as well as severe combined immunodeficient (SCID) mice, are commonly used models of PCP. Although the *P. carinii* organisms infecting these different hosts are morphologically similar, several studies have demonstrated that they consistently display host species-specific antigenic, karyotypic, and genotypic differences. Furthermore, experimental transmission of *P. carinii* isolated from one host to a different immunodeficient host species is restricted.

Glycoprotein A (gpA) of *P. carinii* has been implicated as one of the ligands participating in the adherence of *P. carinii* to host alveolar epithelial cells. GpA, also called MSG, gp120 or P115, is found on the surface of both the trophozoite and cyst forms of *P. carinii*, and is a target of the humoral and cellular immune systems of the host. Passive immunization with an anti-gpA monoclonal antibody partially protects ferrets and rats against the development of PCP. Comparison of *P. carinii* gpA molecules isolated from different host species revealed variation in the antigenic properties and apparent molecular weights of gpA, which are related to the host species of origin. Furthermore, DNA sequence data obtained from ferret, rat, mouse, and human *P. carinii* gpA genes revealed that this host species-specific variation originated in the nucleotide sequence of the genes encoding gpA. Genetic variation in the large-subunit mitochondrial rRNA gene of human *P. carinii* organisms suggests that genetically distinct *P. carinii* organisms can infect humans. Cushion et al., demonstrated genetically distinct rat *P. carinii* strains on the basis of their karyotypes, and that two of these strains can coinfect a single rat. Liu and Leibowitz described two rat *P. carinii* strains based on the presence or absence of a group I intron in the 16S rRNA gene. Genetic variation in *P. carinii* from the same host extends to the genes encoding gpA. For example, several similar, but distinct, cDNA clones encoding gpA have been isolated from rat *P. carinii* expression libraries, indicating that more than one gpA gene is expressed. Whether this phenomenon...
Multiple Genes Encode Ferret *P. carinii* gpA

Table 1. *Ferret P. carinii* gpA-specific oligodeoxyribonucleotides.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys-AS</td>
<td>5’-ACA(C/T)T(T/G)CTCTTCAA(C/T)TCAACACA-3’</td>
<td>Fig. 6, pos. 1701–1787</td>
</tr>
<tr>
<td>307-S</td>
<td>5’-GCTGTGATAAGACATTTGGAATGC-3’</td>
<td>Fig. 6, pos. 1950–1973</td>
</tr>
<tr>
<td>903-S</td>
<td>5’-CCAGAAGGAATAGATCTACTGAAGTG-3’</td>
<td>Ref. 26, Fig. 1, pos. 445–468</td>
</tr>
<tr>
<td>904-AS</td>
<td>5’-GCTGATATCCGTTACAG-3’</td>
<td>Ref. 26, Fig. 1, pos. 910–990</td>
</tr>
<tr>
<td>906-AS</td>
<td>5’-CTAATAAGAGCTCTTCACCC-3’</td>
<td>Ref. 26, Fig. 1, pos. 643–622</td>
</tr>
<tr>
<td>909-S</td>
<td>5’-GAATATAGTAGAGACTGTC-3’</td>
<td>Ref. 26, Fig. 1, pos. 1185–1296</td>
</tr>
<tr>
<td>1003-S</td>
<td>5’-GGTTGTAAAGGACTTTACAG-3’</td>
<td>Fig. 6, pos. 1081–1102</td>
</tr>
<tr>
<td>1004-AS</td>
<td>5’-GGTTGTATCTATGTCCTAATTGG-3’</td>
<td>Fig. 6, pos. 909–888</td>
</tr>
<tr>
<td>MoPc10-AS</td>
<td>5’-GACAACGCTCTTTCTACAGTTTTTA</td>
<td>Fig. 6, pos. 1049–1013</td>
</tr>
<tr>
<td>gpA 06-AS</td>
<td>5’-CAYTGCACTTTATCTGCTG-3’</td>
<td>Fig. 6, pos. 334–317</td>
</tr>
<tr>
<td>gpA 08-AS</td>
<td>5’-CCACTTTATCTGCTG-3’</td>
<td>Fig. 6, pos. 36–19</td>
</tr>
<tr>
<td>gpA 09-S</td>
<td>5’-CCCTACATGTAATCTTGC-3’</td>
<td>Fig. 6, pos. 2716–2736</td>
</tr>
<tr>
<td>IG-S</td>
<td>5’-CGTAGATAGACATCACCC-3’</td>
<td>Fig. 6, pos. 3405–3424</td>
</tr>
<tr>
<td>Pre-Met-S</td>
<td>5’-CTAGTAAACACTCCTCAATAAGAATGGAC-3’</td>
<td>Fig. 6, pos. 3461–3491</td>
</tr>
<tr>
<td>IG-AS</td>
<td>5’-GCACACACATCTCACATGGAACG-3’</td>
<td>Fig. 6, pos. 3194–3173</td>
</tr>
</tbody>
</table>

is the result of infection by multiple strains each expressing a single gpA isoform or a single *P. carinii* organism expressing more than one gpA gene remains unknown. PCR analyses have demonstrated that at least some of the gpA genes of rat *P. carinii* are arranged in tandem repeats on a single chromosome,35,41 thereby giving a single organism the potential to express more than one gpA gene. Multiple gpA genes have also been identified in the genomes of mouse19 and human42,43 *P. carinii* as well, although it is unknown if more than one gene is expressed. To determine whether isoform diversity and tandem duplication of gpA genes is conserved in ferret *P. carinii* and to begin to map the 5’ and 3’ UTRs of gpA transcripts, cloning of portions of two contiguous ferret *P. carinii* gpA genes and characterization of their corresponding mRNAs were performed.

2. Materials and Methods

2.1. Characterization of isoform variants of ferret *P. carinii* gpA cDNA

Recombinant phage PoPc1 through PoPc9 and MoPc10 through MoPc23 previously selected and plaque purified from a *P. carinii*-infected ferret lung cDNA clone bank26 were further characterized. Recombinant phage were plated at a low density on NCZYM (Life Technologies, Gaithersburg, MD) agar plates in the presence of 0.3 mM IPTG to induce fusion protein expression. Four individual plaque plugs from each antibody-positive clone were pooled and boiled for 10 min in 100 µl of SDS-PAGE sample buffer containing 5% (v/v) β-mercaptoethanol. The samples were run on a 7.5% polyacrylamide gel and electroblotted to nitrocellulose. The clones PoPc1 through PoPc9 were subsequently screened by Western blotting with a pool of monoclonal antibodies (MAbs) and clones MoPc10 through MoPc23 were subsequently screened by Western blotting with the anti-gpA polyclonal antiserum.26 For nucleotide sequencing, the cDNA inserts were subcloned into the EcoRI site of the plasmid vector pGEM3Z (Promega, Madison, WI) and sequence analysis was performed as described.26 Sequence data were analyzed using MacVector (IBI-A Kodak Company, New Haven, CT) and the University of Wisconsin Genetics Computer Group (UWGCG) software package.44,45

2.2. Northern blot analysis

Northern blot analysis was performed on *P. carinii*-infected and uninfected ferret lung poly(A)+ mRNA26 using recombinant phage DNA from clone MoPc10 as a probe.

2.3. Polymerase chain reaction

Genomic DNA and poly(A)+ mRNA was isolated from *P. carinii*-infected and uninfected ferret lung.26 PCR, followed by agarose gel electrophoresis of the amplified DNA products, was performed.19 For reverse transcriptase (RT)-PCR, first-strand cDNA synthesis from 0.5 µg of *P. carinii*-infected ferret lung poly(A)+ mRNA was primed with oligo (dT).26 The gpA-specific oligodeoxynucleotide primers were purchased either from the Department of Pediatrics Synthetic Oligomer Facility at the University of Rochester or from Genosys Biotechnologies, Inc. (The Woodlands, TX) (Table 1). The β-actin-sense (S) (5’-GCTGTGCTATGTTGCCCTAGAC-3’) and β-actin-asense (AS) (5’-ACA(C/T)T(T/G)CTCTTCAA(C/T)TCAACACA-3’) were used as internal controls.
Figure 1. Western blot analysis of the β-galactosidase fusion proteins encoded by the 23 ferret *P. carinii* gpA cDNA clones. Panel A: The fusion proteins of clones PoPc1-9 (lanes 1-9), *E. coli* β-galactosidase (lane 10), and Agt11 β-galactosidase (lane 11), screened with the pool of anti-gpA monoclonal antibodies. Panel B: The fusion proteins of clones MoPc10-23 (lanes 2-15), *E. coli* β-galactosidase (lane 16), and Agt11 β-galactosidase (lane 1), screened with rabbit anti-gpA polyclonal antiserum. Panel C: Western blot analysis of purified *E. coli* β-galactosidase and β-galactosidase from an IPTG-induced lysate of Agt11-infected *E. coli* cells, as detected by anti-β-galactosidase antiserum (lanes 1 and 2, respectively). The migration pattern of the molecular weight markers is indicated in the right margin.

Figure 2. Nucleotide and deduced amino acid sequence alignments of two ferret *P. carinii* gpA cDNA clones. Panel A: The nucleotide sequence of clone MoPc10 was aligned with the previously described cDNA clone, PoPc9, using the UWGCG gap program. Identical nucleotides are indicated by vertical lines, and gaps in the sequence alignment are denoted by hyphens. The nucleotide sequence of clone MoPc10 has been deposited with GenBank under accession number U19870. Panel B: The deduced amino acid sequences of clones MoPc10 and PoPc9 were aligned using the UWGCG gap program. Identical amino acids are indicated by dark circles, and gaps in the alignment are denoted by hyphens. Conserved cysteine residues are denoted with asterisks.
Multiple Genes Encode Ferret P. carinii gpA

2.4. Primer extension of gpA transcripts

Primer extension of the gpA transcript was performed as described.\textsuperscript{47} Oligodeoxyribonucleotide primer gpA08-AS (Table 1) was end-labeled with [\(\gamma^{32}P\)]ATP (Du Pont/NEN, Boston, MA) using T4 polynucleotide kinase (Life Technologies).\textsuperscript{48} Radiolabeled gpA08-AS primer (1.3 ng) and 30 \(\mu\)g of \(P.\) carinii-infected ferret lung total RNA were combined, denatured at 90°C for 5 min then annealed for 90 min at 65°C in 15 \(\mu\)l of hybridization buffer (0.15 M KCl, 0.01 M Tris-HCl pH 8.3, 1 mM EDTA). After hybridization, the RNA/primer mixture was slowly cooled to 42°C, and the following was added: 0.9 \(\mu\)l of 1 M Tris-HCl pH 8.3, 0.9 \(\mu\)l of 0.5 M MgCl\(_2\), 0.25 \(\mu\)l of 1 M DTT, 6.75 \(\mu\)l of 1 mg/ml actinomycin D, 1.33 \(\mu\)l of 5 mM dNTPs, and 20 \(\mu\)l dH\(_2\)O. Five units of Avian Myeloblastosis Virus RT (Life Technologies) were then added to extend the primer to the 5' end of the gpA mRNA. Reverse transcription was performed for 1 hr at 42°C, followed by RNase digestion.\textsuperscript{47} After RNase digestion, the end-labeled negative strand cDNA products were separated on an 8% denaturing polyacrylamide sequencing gel.

3. Results

3.1. Characterization of cDNA clones encoding ferret \(P.\) carinii gpA

Twenty-three recombinant \(\lambda\)gt11 clones that reacted with both anti-gpA polyclonal antiserum and a pool of anti-gpA MAbs were identified. Western blot analysis revealed that the \(\beta\)-galactosidase fusion proteins of these 23 clones ranged in size from 130–163 kDa (Fig. 1A, lanes 1–9, Fig. 1B, lanes 2–15). All gpA cDNA clones selected with the polyclonal anti-gpA antiserum (PoPc1-9) were also positive with the pool of 10 different anti-gpA MAbs; similarly, all of the gpA cDNA clones selected with the pool of anti-gpA MAbs (MoPc10-23) were positive with the polyclonal anti-gpA antiserum.\textsuperscript{26} Based on the reactivity of their \(\beta\)-galactosidase fusion proteins with individual anti-gpA MAbs, and the nucleotide sequence of the cDNA inserts, the 23 clones were grouped into three classes represented by cDNA clones PoPc3, PoPc9, and MoPc10. Characterization of PoPc9 and PoPc3 indicated that they are nonoverlapping cDNA clones encoding the majority of ferret \(P.\) carinii gpA.\textsuperscript{26} Although the fusion protein encoded by MoPc10 reacted with the same individual anti-gpA MAbs as clone PoPc9, a comparison of the DNA sequences revealed 22% divergence (Fig. 2A). The deduced amino acid sequences are 80% similar to each other, and each is derived from a single open reading frame (ORF) (Fig. 2B). Furthermore, all of the cysteine residues except one in the overlapping region are conserved.\textsuperscript{19}

3.2. Northern hybridization

Clone MoPc10 hybridized under conditions of high stringency to a single mRNA species of ~3600 nucleotides (nts) only in \(P.\) carinii-infected ferret lung mRNA (Fig. 3A, lane 1). This was consistent with earlier Northern blot analysis demonstrating that the gpA-specific clones PoPc3 and PoPc9 hybridize to a single mRNA species of ~3600 nts.\textsuperscript{26} Both PCR and Southern blot analysis were used to confirm that MoPc10 was derived from a \(P.\) carinii-specific transcript, and not from a host lung transcript induced by infection or steroid treatment (data not shown). These data suggest that clone MoPc10 represents a new isoform of ferret \(P.\) carinii gpA.
Figure 4. Relationship of ferret *P. carinii* gpA cDNA clones PoPc3, PoPc9, and MoPc10 as determined by RT-PCR. Panel A: *P. carinii*-infected ferret lung poly(A)+ RNA was PCR amplified either after reverse transcription with MMLV RT (+RT) or without prior reverse transcription (−RT). The reactions were as follows: +RT and primers 909-S/Cys-AS (lane 1), −RT and primers 909-S/Cys-AS (lane 2), +RT and primers 1003-S/Cys-AS (lane 3), −RT and primers 1003-S/Cys-AS (lane 4), +RT and primers 903-S/904-AS (lane 5), +RT and primers 1003-S/904-AS (lane 6), −RT and primers 903-S/904-AS (lane 7), +RT and primers 903-S/904-AS (lane 8), −RT and primers β-actin-S/β-actin-AS (lane 9), and −RT and primers β-actin-S/β-actin-AS (lane 10). Thirty-five PCR cycles of 94°C for 90 s, 60°C for 90 s, and then 72°C for 120 s were performed, and the reaction products were resolved on a 1.8% agarose gel. The molecular size standards are in 100-bp increments (lanes M). The black arrows denote the darker staining 600-bp molecular size standard. Panel B: Diagram showing the relationship of the three *P. carinii* gpA cDNA clones as determined by RT-PCR and DNA sequence analysis. The arrows denote the positions of the primers used to amplify the clone junctions (S=sense; AS=antisense). The open arrow marks the previously unknown DNA segment lying between clones PoPc9 and PoPc3. The size of the amplification product expected using 909-S/Cys-AS is 702-bp, 1003-S/Cys-AS is 620-bp, 903-S/904-AS is 465-bp, and β-actin-S/β-actin-AS is 450-bp.

3.3. Determination of the interrelationship of ferret *P. carinii* gpA clones PoPc3, PoPc9 and MoPc10 by RT-PCR

RT-PCR analysis using primer combinations (described in figure legends and Table 1) from clones PoPc3, PoPc9, and MoPc10 was performed to determine their relationship to one another. The gpA mRNA segments encoding clones PoPc9 (Fig. 4A, lane 1) and MoPc10 (Fig. 4A, lane 3) were found to be upstream of and co-linear with the gpA mRNA segment encoding clone PoPc3; however, the gpA mRNA segments encoding clones PoPc9-PoPc3 and MoPc10-PoPc3 (gpA9-3 and gpA10-3, respectively) are distinct gpA transcripts (Fig. 4A, lanes 1 and 3 and Fig. 4B). GpA10-3 contains an endogenous EcoRI restriction site separating the two clones, whereas, the corresponding sequence spanning gpA9-3 does not contain an EcoRI site. For each primer pair tested, no DNA product was amplified from the mRNA template when it was not first reverse transcribed (Fig. 4A, lanes 2, 4, 8 and 10). *P. carinii*-infected ferret lung poly(A)+ RNA was analyzed by RT-PCR using primer combinations from PoPc9 and MoPc10. The absence of RT-PCR products was expected and further supported the conclusion that clones MoPc10 and PoPc9 are derived from different gpA-specific transcripts expressed in the population of *P. carinii* organisms infecting the ferret (Fig. 4B).

3.4. Tandem repeat of gpA genes on a single ferret *P. carinii* chromosome

To determine whether the gpA9-3 and gpA10-3 gpA genes were located on the same chromosome, PCR analysis was performed. When primer 307-S was used in combination with primer MoPc10-AS, a 2600-bp DNA product was amplified from *P. carinii*-infected, but not normal ferret lung DNA (Fig. 5, lane 1 compared to lane 2, respectively). Similarly, when primer 307-S was used in combination with primer 906-AS, a 2200-bp DNA product was amplified from *P. carinii*-infected, but not normal ferret lung DNA (Fig. 5A, lane 3 compared to lane 4, respectively). To confirm that the PCR fragments spanned two adjacent gpA genes on a single *P. carinii* chromosome, the nucleotide sequence of the larger 2600-bp PCR product was determined. DNA sequence analysis revealed that the cloned PCR product encoded the 3' end of one ferret *P. carinii* gpA gene, 329 bp of intergenic sequence, and the 5' end of another gpA gene. A schematic of the genomic organization of ferret *P. carinii* gpA genes, including the PCR primers used, is shown in Fig. 5B. A single ORF encoding 400 amino acids ending with a TAG stop codon at the 3' end of the first gpA
The unmodified gpA protein has a predicted molecular stream of 9 amino acids of the first gpA gene at nt 3486. The gpA gene of the tandem repeat is located 330 nts downstream from the stop codon and ending with the TAG stop codon (Fig. 6). The gene encodes 1051 amino acids beginning with the ATG start codon and ending with the TAG stop codon in the intergenic region at nts 3178, 3182, and 3186 (Fig. 6). The first ATG codon of the second gene of the tandem repeat is present. Three potential polyadenylation signals are present downstream from the stop codon, but upstream of the 3' UTR of the gpA mRNA. The expected 479-bp product was also amplified by conventional PCR from genomic DNA isolated from the same infected ferret lung (Fig. 7A, lane 3). Similarly, to determine whether the intergenic region was related to the 5' UTR of the gpA mRNA, RT-PCR was performed on poly(A)+ mRNA derived from a single infected ferret lung using an antisense primer located in the intergenic region (IG-AS) paired with a sense primer located near the 3' end of the gpA coding sequence (gpA 09-S). The IG-AS primer was located downstream of the stop codon, but upstream of the three potential polyadenylation signals. This primer pair yielded the expected 479-bp reaction product (Fig. 7A, lane 1), indicating that at least a portion of the intergenic region is related to the 3' UTR of the gpA mRNA. The expected 479-bp product was also amplified by conventional PCR from genomic DNA isolated from the same infected ferret lung (Fig. 7A, lane 3). Similarly, to determine whether the intergenic region was related to the 5' UTR of the gpA mRNA, RT-PCR was performed on the same infected ferret lung poly(A)+ mRNA. Using a sense primer 61 bp upstream of the start codon of the second gene of the tandem repeat (IG-S) paired with an antisense primer near the 5' end of the coding region of the gpA gene (gpA 08-AS), no product was amplified by RT-PCR (Fig. 7A, lane 4). Using the same primer pair to amplify genomic DNA isolated from the same infected ferret lung by PCR, the expected 117-bp product was observed (Fig. 7A, lane 6). These data indicated that the intergenic region which lies between nts 3405 and 3424 downstream of the stop codon of the first gpA gene of the tandem repeat, indicated that the intergenic region may be transcribed as the 3' untranslated region (UTR) of the gpA mRNA. To test this hypothesis, RT-PCR was performed on poly(A)+ mRNA derived from a single infected ferret lung using an antisense primer located in the intergenic region (IG-AS) paired with a sense primer located near the 3' end of the gpA coding sequence (gpA 09-S). The IG-AS primer was located downstream of the stop codon, but upstream of the three potential polyadenylation signals. This primer pair yielded the expected 479-bp reaction product (Fig. 7A, lane 1), indicating that at least a portion of the intergenic region is related to the 3' UTR of the gpA mRNA. The expected 479-bp product was also amplified by conventional PCR from genomic DNA isolated from the same infected ferret lung (Fig. 7A, lane 3). Similarly, to determine whether the intergenic region was related to the 5' UTR of the gpA mRNA, RT-PCR was performed on the same infected ferret lung poly(A)+ mRNA. Using a sense primer 61 bp upstream of the start codon of the second gene of the tandem repeat (IG-S) paired with an antisense primer near the 5' end of the coding region of the gpA gene (gpA 08-AS), no product was amplified by RT-PCR (Fig. 7A, lane 4). Using the same primer pair to amplify genomic DNA isolated from the same infected ferret lung by PCR, the expected 117-bp product was observed (Fig. 7A, lane 6). These data indicated that the intergenic region which lies between nts 3405 and 3424 (Fig. 6) upstream of the gpA start codon on the genomic DNA is not included in the 5' UTR of the gpA transcript. To determine whether the intergenic sequence immediately upstream of the ATG start codon at nt 3586 (Fig. 6) upstream of the gpA start codon on the genomic DNA is not included in the 5' UTR of the gpA transcript.
Figure 6. Nucleotide and predicted amino acid sequence of the composite full-length ferret *P. carinii* gpA gene. A single full-length gpA gene was constructed using the nucleotide sequences of clones MoPc10 (nts 653-1213) and PoPc3 26 (nts 1214-2323), and the 5' and 3' terminal sequences derived from the gpA tandem repeat PCR product. The sequence of the genomic intergenic region (nts 3156-3486) is included between the stop codon of the first gpA gene and the start codon of the second gpA gene. The three positive polyadenylation signals in the intergenic region are bracketed above the AATAAA sequences. The start codon of the second gpA gene in the tandem repeat is included, with the guanines at positions -3 and +4 marked by closed circles. Amino acids are indicated by the one letter code with the stop codon denoted by an asterisk. Potential N-linked glycosylation sites are underlined.
Multiple Genes Encode Ferret \textit{P. carinii} gpA

6) is included in the mature mRNA encoding gpA, a pre-Met sense primer corresponding to nts 3460–3491 (Table 1 and Fig. 6) was synthesized. RT-PCR using this primer coupled with an antisense primer located at the 5' end of the coding region resulted in amplification of a product of the expected size (Fig. 7A, lane 7). To determine the length of the gpA 5' UTR, primer extension analysis was performed using primer gpA-08-AS. Two distinct primer-extended products of 110 and 280 nts were obtained in the primer extension reaction (Fig. 7B). These results suggest that two different +1 sites for transcription initiation may be found for the mature gpA transcript, resulting in 5' UTRs of 74 and 244 nts, respectively (Fig. 7B). Alternatively, the longer reaction product may represent the true gpA mRNA start site while the shorter reaction product may result from termination of the reverse transcription step during primer extension. Taken together, these data support the conclusion that only a small portion of the intergenic sequence immediately adjacent to the start codon is included in the gpA 5' UTR, while the majority of the intergenic region is not transcribed and processed into the 5' UTR of a mature gpA transcript (Fig. 7C).

4. Discussion

In the present study, we isolated cDNA and genomic clones of ferret \textit{P. carinii} gpA and characterized their gene organization and the approximate map positions of the 5' and 3' UTRs of the mature gpA transcripts. Two cDNA clones were derived from very similar, but distinct, gpA mRNA transcripts showing divergence in the sequence of ~20%. Each ferret \textit{P. carinii} gpA cDNA clone hybridized to a single mRNA species of approximately 3600 nts, which is similar in size to the mRNAs for rat and mouse \textit{P. carinii} gpA.\textsuperscript{19,27,35} Thus, ferret \textit{P. carinii} organisms express more than one distinct gpA mRNA, as do rat \textit{P. carinii}.\textsuperscript{27,35} The question of whether the host is infected with a single \textit{P. carinii} organism expressing multiple gpA genes simultaneously or with multiple strains of \textit{P. carinii} each expressing a different isoform of the gpA gene remains unanswered. It will be difficult to answer this question definitively until the methods to propagate clonal populations of \textit{P. carinii} organisms are readily available.

A single ferret \textit{P. carinii} organism appears to contain multiple gpA genes scattered throughout its genome. PCR analyses demonstrated that two gpA genes were arranged in a tandem repeat on an individual ferret \textit{P. carinii} chromosome, while Southern blot analysis demonstrated that an internal fragment of the ferret \textit{P. carinii} gpA gene hybridized, under high stringency conditions, to each ferret \textit{P. carinii} chromosome (Wright et al. unpublished data). These data suggest that a single ferret-derived organism has a genomic repertoire sufficient to express more than one gpA gene. The tandem chromosomal repeat of gpA genes has also been demonstrated in rat and human \textit{P. carinii}, indicating that this is likely a genomic structure common to all \textit{P. carinii}.\textsuperscript{41,43} It remains unknown whether all of the gpA-related sequences in the \textit{P. carinii} genome are complete, intact genes. There is a possibility that some of these sequences may be either pseudogenes, or incomplete fragments of gpA genes.

The 329-bp intergenic region separating the tandem repeat of ferret \textit{P. carinii} gpA genes contained consensus polyadenylation signals downstream from the stop codon of the first gpA gene, but did not contain obvious consensus eukaryotic promoter regions upstream of the start codon of the second gpA gene. Consistent with these observations, RT-PCR data demonstrated that the intergenic region immediately downstream of the first gpA gene in the tandem repeat constitutes the 5' UTR of the ferret \textit{P. carinii} gpA mRNA, while the intergenic region lying more than 25 nts upstream of the start codon of the second gpA gene of the tandem repeat appears to be unrelated to the 5' UTR of the gpA mRNA (Fig. 7). Primer extension analysis of the gpA mRNA indicated that the 5' UTR extends at least 50–200 nts upstream of the nucleotide at position -25 (nt 3460, Fig. 6). These data are consistent with recent work demonstrating that the intergenic region of tandem rat \textit{P. carinii} gpA genes is also closely related to the 3' UTR of the rat \textit{P. carinii} gpA mRNA.\textsuperscript{41} Furthermore, the intergenic region upstream of the first methionine codon of the second rat \textit{P. carinii} gpA gene in the tandem repeat shows little resemblance to the 5' UTR of the mRNA, except for the 14 nts preceding the first codon for methionine.\textsuperscript{41} Recent data by Wada et al.\textsuperscript{52} suggest that a conserved DNA element called the upstream conserved sequence (UCS), which maps to a single chromosome, mediates the expression of \textit{P. carinii} gpA/MSG genes. This element is attached to most expressed MSG genes; however, different organisms in the population have different MSG genes attached to the UCS,\textsuperscript{53} which explains, at least in part, the genetic heterogeneity of gpA molecules. GpA genes may be transcribed as polycistronic mRNAs, which are then rapidly processed into the steady state population of unit length mature gpA transcripts. This suggests a role for cis-splicing of expressed transcripts from tandemly arrayed gpA genes;\textsuperscript{52} a gene organization conserved in gpA genes of \textit{P. carinii} derived from all host species studied. Alternatively, genomic copies of gpA genes may need to be moved into an expression locus for transcription into a precursor RNA molecule, which is then segregated into individual mRNAs by polyadenylation and trans-splicing as occurs in \textit{Trypanosoma}.\textsuperscript{53} Taken together, these data indicate that transcription does not initiate at the intergenic regions between tandem repeats of the gpA gene.

Nucleotide sequencing of the PCR-generated fragment
Figure 7. RT-PCR showing the location of the intergenic region on the mRNA encoding gpA. Panel A: *P. carinii*-infected ferret lung total RNA with (lane 1), or without (lane 2) reverse transcription and *P. carinii*-infected ferret lung genomic DNA (lane 3) were amplified with the gpA 09-S and IG-AS primers. *P. carinii*-infected ferret lung total RNA with (lane 4), or without (lane 5) reverse transcription and *P. carinii*-infected ferret lung genomic DNA (lane 6) were amplified using the IG-S and gpA 08-AS primers. *P. carinii*-infected ferret lung total RNA with (lane 7), or without (lane 8) reverse transcription and *P. carinii*-infected ferret lung genomic DNA (lane 9) were amplified with the Pre-Met-S and gpA 06-AS. *P. carinii*-infected ferret lung total RNA with (lane 10) or without (lane 11) reverse transcription was amplified using the 903-S and 904-AS primers. Thirty-five PCR cycles of 94°C for 90 s, 50°C for 90 s, and then 72°C for 120 s were performed, and the reaction products were resolved on a 1.8% agarose gel. The molecular size standards are in 100-bp increments (lanes M). Panel B: Autoradiograph of 8% denaturing polyacrylamide gel of end-labeled primer extension products. Primer extension of *P. carinii*-infected ferret lung total RNA was performed using the gpA-specific primer gpA 08-AS (lane 1). Denatured, radiolabeled molecular weight markers in 100-nt increments are shown in lane M. The 200 nucleotide size standard is marked with a black arrowhead. Panel C: Schematic diagram depicting the tandem repeat of gpA genes on a *P. carinii* chromosome, and the relationship of the 5' and 3' UTRs of the gpA mRNA to the intergenic sequence. The gpA coding sequence is depicted by open boxes, while the 5' and 3' UTRs of the gpA mRNA are indicated by single lines. For purposes of illustration, the mature mRNA transcript is shown schematically to emphasize that the 3' UTR is derived from the intergenic sequence, while only a portion of the 5' UTR is derived from the genomic intergenic sequence. The relevant primers and their positions are indicated by dark arrows.
containing the 3’ end of one gpA gene and the 5’ end of another allowed the construction of a composite full-length ferret \textit{P. carinii} gpA gene. The resulting, full-length gpA gene may not actually exist in the genome of ferret \textit{P. carinii}. However, this composite full-length gene is representative of ferret \textit{P. carinii} gpA and may be used to examine the structural characteristics of gpA, as the gpA genes of \textit{P. carinii} isolated from the same host species are very similar.\textsuperscript{19,27,35,43} The deduced amino acid sequence of the full-length ferret \textit{P. carinii} gpA gene is 1051 amino acids, smaller than the 1088,\textsuperscript{27} 1076\textsuperscript{40} or 1211\textsuperscript{52} amino acid sequences reported for full-length rat \textit{P. carinii} gpA, but larger than the 1030 amino acids reported for human \textit{P. carinii} gpA.\textsuperscript{43} The amino acid sequence of the full-length ferret \textit{P. carinii} gpA is 35% identical and 54% similar to rat \textit{P. carinii} gpA,\textsuperscript{27} and 37% identical and 54% similar to human \textit{P. carinii} gpA.\textsuperscript{43} The high degree of diversity observed in the coding regions of the ferret, rat and human \textit{P. carinii} gpA genes are also observed in the intergenic regions lying between tandem genomic copies of these genes. Despite the diversity in these gpA molecules, there are obvious similarities. Each of the published rat,\textsuperscript{27} human,\textsuperscript{43} and ferret \textit{P. carinii} gpA amino acid sequences contain numerous putative \textit{N}-linked glycosylation sites, consistent with previous reports documenting \textit{N}-linked glycosylation of gpA.\textsuperscript{30,31,34} Thirteen sites were present in ferret \textit{P. carinii} gpA, as compared to 10 sites in human\textsuperscript{43} and 5 sites in rat\textsuperscript{27} \textit{P. carinii} gpA. A higher density of potential glycosylation sites correlates with Western blot data suggesting that ferret \textit{P. carinii} gpAs may be more heavily glycosylated than rat or human \textit{P. carinii} gpAs.\textsuperscript{11} One region of strong homology was identified at amino acid residues 969–1013 of ferret \textit{P. carinii} gpA (Fig. 6), residues 1002–1046 of rat \textit{P. carinii} gpA,\textsuperscript{27} and residues 952–996 of human \textit{P. carinii} gpA.\textsuperscript{43} These regions are very rich in threonine residues and shared greater than 70% identity over the 45 amino acid stretch. Threonine-rich regions have been identified in the mucin-like surface glycoproteins of the protozoan parasites \textit{Trypanosoma} and \textit{Leishmania}.\textsuperscript{54,55} Furthermore, at least one \textit{Trypanosoma cruzi} mucin-like surface glycoprotein has been shown to bind mammalian host cells.\textsuperscript{55} Thus, the possibility exists that the threonine-rich region in gpA may be a mucin-like region of O-linked glycosylation which plays an analogous role in the binding of \textit{P. carinii} to host cells. In each of the gpA molecules examined thus far, the C-terminal amino acids form a strongly hydrophobic tail. In ferret \textit{P. carinii} gpA, the 14 hydrophobic C-terminal amino acids are immediately preceded by a five-amino acid hydrophilic stretch, which is preceded by three amino acids with small side groups (Fig. 6). This general C-terminal amino acid conformation is characteristic of other eukaryotic proteins, such as the \textit{Trypanosoma brucei} VSGs,\textsuperscript{56} which are anchored to the cell by glycosyl phosphatidyl-inositol membrane anchors.\textsuperscript{51}

Particularly obvious upon alignment of these three gpA molecules, was the large number of conserved cysteine residues, which has been noted previously.\textsuperscript{19} Ferret \textit{P. carinii} gpA contains 63 (6%) cysteine residues, rat \textit{P. carinii} gpA contains 60 (5.5%) cysteine residues,\textsuperscript{27} and human \textit{P. carinii} gpA contains 56 (5.4%) cysteine residues,\textsuperscript{43} the majority of which are conserved among these three gpA molecules. This suggests a similar secondary structure, and likely, a similar function for the gpAs of \textit{P. carinii} isolated from different host species. Despite the high degree of diversity observed between these molecules, several conserved features indicate that the gpAs of \textit{P. carinii} isolated from different host species are homologous molecules.

**Acknowledgments:** This investigation was supported by in part by USPHS awards HL49610, HL30616, HL50615, AI-23302 and AI-07362 from the National Institutes of Health, and a grant from the Strong Children’s Research Center.

**References**

antigens and characterization of antigenic differences between rat and human isolates, J. Infect. Dis., 159, 60–70.


Multiple Genes Encode Ferret *P. carinii* gpA

Res., 1, 57–66.


