Variation and \textit{in vitro} splicing of group I introns in rRNA genes of \textit{Pneumocystis carinii}

Yong Liu and Michael J. Leibowitz*
Department of Molecular Genetics and Microbiology, UMDNJ – Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854-5635, USA

Received February 12, 1993; Revised and Accepted April 12, 1993

\section*{ABSTRACT}

The sequences of the rRNA genes of \textit{Pneumocystis carinii} from rat and human sources demonstrate three distinct genotypes based on the group I introns present in these genes. One rat isolate (Pc1) contains such introns in its 16S and 26S rRNA genes, while another rat isolate (Pc2) and a human isolate (Pc3) only contain an intron in the 26S rRNA gene. The four introns all catalyze their own excision from RNA transcripts, and this reaction is inhibited by the anti-pneumocystis drug pentamidine and its analogues. Although they differ in sequence, they are more similar to one another than to group I introns found in other eukaryotic microbes.

\section*{INTRODUCTION}

\textit{Pneumocystis carinii} is a yeast-like organism that infects most humans in early childhood without causing symptoms, but causes severe pneumonitis and other infections in patients with AIDS and other immunodeficiency states (1–2). Although no long-term culture system has been found for \textit{P. carinii}, the organism has been classified as a yeast based on the DNA sequence of its genes, especially those of the major rRNA operon encoding 16S (SSU), 5.8S and 26S (LSU) rRNA. The 16S rRNA gene of \textit{P. carinii} from immunosuppressed rats was found to resemble its homologue in yeast (3), although, unlike the yeast gene, it contained a group I self-splicing intron near the 3' terminus (3–7). This apparent similarity of rRNA genes of \textit{P. carinii} to their homologues in fungi was extended to the genes encoding the major rRNA operon encoding 16S (SSU), 5.8S and 26S rRNA (5). Sequence analysis of selected regions of the 26S rRNA gene of \textit{P. carinii} isolated from the lungs of two different immunosuppressed rat strains revealed them to be distinct. These two isolates have been denoted Pc1 and Pc2 (7). The isolate denoted Pc1 (7) and an independently isolated rat-derived \textit{P. carinii} strain had the same sequence previously reported (3–5, but see also GenBank entry X13687) for the intron and 3' terminal exon of the 16S rRNA gene. A group I self-splicing intron was also present in the 26S rRNA gene of isolate Pc1 (7). The other \textit{P. carinii} isolate, designated Pc2, was distinct from Pc1 at multiple sites within the 26S rRNA gene (7).

Hughes and Gigliotti have suggested that each mammalian species might be infected by a single distinct \textit{P. carinii} strain (8). Indeed, antigenic differences between \textit{P. carinii} from different host species have been demonstrated (9–11), as have differences in the sequences of mitochondrial DNA (12). However, based on DNA hybridization to a cloned DNA fragment, it is clear that different human-derived \textit{P. carinii} isolates are not identical (13). Furthermore, two distinct patterns have been observed by pulsed-field electrophoresis of chromosomal DNA from different \textit{P. carinii} isolates from rats (14–16). Thus, it is likely that different strains of \textit{P. carinii} can infect any given host species.

Group I self-splicing introns (reviewed in 17–18) are characterized by a conserved 'core' sequence (19) and the ability to catalyze their own excision from primary transcripts in the presence of guanosine, divalent cations, and in some cases, protein co-factors. Although some group I introns are relatively long and encode proteins, the simplest, including those previously described in the rRNA genes of \textit{P. carinii} (4–7), lack any protein-coding regions. The existence of group I introns in the nuclear genes of \textit{P. carinii} (4–7) and the mitochondrial genes of many fungal species (17–19) distinguishes these organisms from mammals. Thus, there is good rationale for the idea that drugs inhibiting intron splicing might be good anti-fungal agents (6, 20). The anti-pneumocystis drug pentamidine and its analogues have the ability to inhibit \textit{in vitro} splicing by the group I introns found in the 16S and 26S rRNA genes of \textit{P. carinii} type Pc1 isolated from rats (21). However, we are not aware of any reports describing group I introns in human isolates of \textit{P. carinii}. In this report, we describe the sequence and the location of the group I introns of rat-derived \textit{P. carinii} type Pc2 and of human-derived \textit{P. carinii}, which we have termed type Pc3. This is the first description of the rRNA gene organization in human-derived \textit{P. carinii}.

Since some linear excised group I introns have been shown to catalyze insertion into different sites in RNA molecules (22), related organisms may not have such introns in the same sites. Various protozoa (23), fungi (24), slime molds (25) and eukaryotic algae (26) harbor group I introns in various sites in their nuclear rRNA genes. It would be of interest to compare...
the rRNA genes of different rat and human-derived *P. carinii* for the location and sequence of group I introns. Analysis of different Tetrahymena species showed that such introns were present at a common location in the rRNA genes, although the different species showed small variations in the sequence of these introns (23).

**METHODS**

*Pneumocystis carinii* DNA

DNA was extracted as previously described (7) from *P. carinii* from the lungs of immunosuppressed rats. Pc1 denotes organisms from Sprague-Dawley rats from Sasco, Inc. (Omaha, NE); Pc2 denotes organisms from hooded rats from Harlan–Sprague-Dawley (Indianapolis, IN). DNA extracted from *P. carinii* present in broncho-alveolar lavage from five patients with AIDS complicated by *P. carinii* pneumonitis (cases 114, 134, A, W and Z), provided by J.A. Kovacs (NIH, Bethesda, MD), was prepared as previously described (27).

**Oligonucleotides**

DNA oligonucleotides (Table I) were prepared by beta-cyanoethyl phosphoramidite chemistry on Cyclone (Milligen) and 380B (Applied Biosystems) automated DNA synthesizers, and were purified by chromatography on NENsorb-Prep cartridges (NEN-DuPont).

**Amplification, cloning and sequence determination of DNA**

Specific DNA regions were amplified by polymerase chain reaction (PCR) using DNA polymerase from *Thermus aquaticus* (Perkin Elmer Cetus) under reaction conditions previously described (7). For nested PCR reactions, initial amplification was performed by 2 cycles of 94°C for 2 minutes, 56°C for 1 minute, and 72°C for 1 minute, with a final 5 minute incubation at 72°C. Amplified fragments were cloned by blunt-end ligation into *Sma*1-cleaved pUC18 DNA as described (7). Reverse transcription followed by PCR amplification was done using a Gene Amp RNA PCR kit (Perkin-Elmer-Cetus).

DNA sequence determination was performed on covalently closed superhelical DNA templates on a Genesis 2,000 Automated DNA Sequencer (DuPont), using DNA polymerase from bacteriophage T7 (Sequenase version 1.0, U.S. Biochemicals). Primers included oligonucleotides 228A and 229, which base pair to regions of pUC18 on either side of the *Sma*1 site into which amplified DNA fragments were inserted, and internal primers pairing with either strand of the intron region (Table I). The following intron-derived primers were successfully used in sequencing reactions for the cloned introns: for Pc1.LSU, oligonucleotides 4358, 4359, 4612, 4613, 4614, and 4615; for Pc2.LSU, oligonucleotides 4358, 4359, 4613 and 4614, and for Pc3.LSU oligonucleotides 4358, 4359, 4614 and 5528. In this intron nomenclature (19), the symbols before the dot indicate the organism and those after the dot (SSU, 16S and LSU, 26S rRNA) the gene in which the intron is present. Note that some of these primers had several mismatched nucleotides, yet they primed successfully and yielded sequence which was confirmed using primers 228A and 229. All reported sequences were determined at least twice on each DNA strand.

**Splicing reactions**

Regions of the 26S rRNA genes containing group I introns were PCR-amplified using primers 5573 and 4358, by 2 cycles of 94°C for 2 minutes, 55°C for 1 minute, and 72°C for 0.5 minutes, followed by 30 cycles of 94°C for 1 minute, 58°C for 1 minute, and 72°C for 1 minute, with a final 5 minute incubation at 72°C. The resulting PCR product was transcribed by bacteriophage SP6 amplification was re-amplified in the presence of a second set of primers by 2 cycles at 94°C for 2 minutes, 57°C for 1 minute, and 72°C for 1 minute, followed by 30 cycles of 94°C for 1 minute, 58°C for 1 minute, and 72°C for 1 minute, with a final 5 minute incubation at 72°C. Amplified fragments were cloned by blunt-end ligation into *Sma*1-cut pUC18 DNA as described (7).

**Table I. Oligonucleotide primers used**

<table>
<thead>
<tr>
<th>Number</th>
<th>Sequence</th>
<th>5'Coordinate</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>228A</td>
<td>AACAGCTATGACCATGAT</td>
<td>pUC polylinker</td>
<td></td>
</tr>
<tr>
<td>229</td>
<td>TTCCCAGTCACGACGTTG</td>
<td>pUC polylinker</td>
<td></td>
</tr>
<tr>
<td>1138</td>
<td>AGGAGATGGTGGGCTGCTTCCGAA</td>
<td>637(+) , 16S</td>
<td>3</td>
</tr>
<tr>
<td>1463</td>
<td>GCTGTGCTTCAAGATTAGCCATGC</td>
<td>29(+) , 16S</td>
<td>3</td>
</tr>
<tr>
<td>2892</td>
<td>GCTATCTTGGAGGAAAACCTCCGG</td>
<td>964(—), 26S</td>
<td>7</td>
</tr>
<tr>
<td>2982</td>
<td>GAGCACAATCTTATCCCGAAGTACC</td>
<td>1933(+), 26S</td>
<td>7</td>
</tr>
<tr>
<td>3175</td>
<td>GGTTGGTGCTGATGCGCCG</td>
<td>1262(+), 16S</td>
<td>3</td>
</tr>
<tr>
<td>3176</td>
<td>CCTTCCGCAAGTTCATCCAGGG</td>
<td>1796(—), 16S</td>
<td>3</td>
</tr>
<tr>
<td>3426</td>
<td>GCTCGGCTGTTACGAACGGG</td>
<td>654(—), 26S</td>
<td>7</td>
</tr>
<tr>
<td>4170</td>
<td>GCTTAAATCTCAGAAGATCC</td>
<td>3327(—), 26S</td>
<td>7</td>
</tr>
<tr>
<td>4358</td>
<td>GACGAGGCATTGGCTTAACCC</td>
<td>2267(—), 26S</td>
<td>7</td>
</tr>
<tr>
<td>4359</td>
<td>CTGCCGGAATGCCTGATGTC</td>
<td>2169(+), 26S</td>
<td>7</td>
</tr>
<tr>
<td>4360</td>
<td>GAATGCTTAAATGTCAGGGG</td>
<td>47(—), 26S</td>
<td>7</td>
</tr>
<tr>
<td>4434</td>
<td>GTACACACCGGCCGTGCC</td>
<td>1631(+), 16S</td>
<td>3</td>
</tr>
<tr>
<td>4612</td>
<td>GCAGTTACGAGCTAGAGCAGG</td>
<td>244(+), intron, 26S, Pc1</td>
<td>7</td>
</tr>
<tr>
<td>4613</td>
<td>CCCAGAGCTTAATACCCGGG</td>
<td>157(—), intron, 26S, Pc1</td>
<td>7</td>
</tr>
<tr>
<td>4614</td>
<td>CAATTGCGCGGAGTCC</td>
<td>78(+), intron, 26S, Pc1</td>
<td>7</td>
</tr>
<tr>
<td>4615</td>
<td>CCCCATATCTCTACAATACCC</td>
<td>288(—), intron, 26S, Pc1</td>
<td>7</td>
</tr>
<tr>
<td>5528</td>
<td>GTCGTTGAAAGCTGACCC</td>
<td>273(—), intron, 26S, Pc3</td>
<td>7</td>
</tr>
<tr>
<td>5573</td>
<td>ATTTAGGTGACACTATAGAAGTGGTCGATGCTGAATGCTTTT</td>
<td>2175(+), 26S</td>
<td>7</td>
</tr>
</tbody>
</table>

Table I lists all primers used for PCR amplifications and sequencing. The position of the 5' nucleotide of each primer pairing with the rRNA-coding genes is indicated as the 5' coordinate of the corresponding nucleotide on the 16S (3) or 26S (7) rRNA sequence of *P. carinii* strains classified as Pc1 (see text). The symbols (+) and (—) indicate polarity of each oligonucleotide. Sequences specific for individual strains of *P. carinii* are indicated. The underlined sequence in oligonucleotide 5573 constitutes a bacteriophage SP6 promoter.
RNA polymerase (Promega) in the presence of 7.5 μM [α-32P] UTP (80 μCi, NEN), under the recommended conditions. After 45 minutes at 37°C, 1 unit of DNase (RNase-free, RQ1, Promega) was added and DNA template was digested for 15 minutes at 37°C. Transcripts were then extracted with phenol:chloroform (1:1) and chloroform:isoamyl alcohol (24:1), ethanol precipitated, and the full-length precursor species was purified by polyacrylamide gel electrophoresis (5% acrylamide, 0.1% bis-acrylamide) run in the presence of M urea, followed by electroelution and ethanol precipitation. Splicing reactions catalyzed by this precursor were performed in the presence of 50 mM Tris–HCl (pH 7.5), 5 mM magnesium chloride, 0.4 mM spermidine, 1 unit of RNasin (Promega) and 2 μM GTP. After 10 minutes at 37°C, RNA species were separated by polyacrylamide gel electrophoresis as described above. Radioactivity in the precursor, intermediates and products of the splicing reaction were visualized by autoradiography and quantitated using an AMBiS Gel Scanner.

Antibiotics

Pentamidine was from Sigma. The pentamidine analogues BBE, DIMP and O05NAN021 (compounds 10, 7 and 9 in ref. 21) were included as molecular weight markers. Panel B shows a map consistent with the presence of the group I intron (Pc1.LSU) in Pel DNA (7). As indicated in Figure 1, Panel B (above the gene map), use of primers 4359 and 4358 resulted in amplification of a fragment of the 26S rRNA gene from Pc2 DNA whose size was consistent with the presence of a group I intron at the same location as that present in Pc1 (Panel A). When this fragment was cloned and sequenced it contained a group I intron sequence (Pc2.LSU) with similarity but not identity to Pc1.LSU (Figure 2). Sequence analysis of the flanking exon regions in this clone showed the insertion site of this intron to be identical to that reported (7) for intron Pc1.LSU. However, PCR amplification of Pc2 DNA with six pairs of primers spanning the site of the group I intron in the 16S rRNA gene in Pc1 DNA only yielded fragments of a size consistent with the absence of such an intron (Figure 1, panel B). These results were confirmed by sequence analysis of cloned fragments generated with either primers 4434 and 4360, or 3175 and 3176, which demonstrated a sequence of a rat-derived ‘variant’ P. carinii 16S rRNA gene obtained from an apparent mixed infection by both Pel and Pc2-like strains (16).

RESULTS

Location and sequence of group I introns

Figure 1 shows the results of PCR amplification of the sites in which group I introns are present in the rRNA operon from rat and human isolates of P. carinii. As indicated in Panel A, six separate pairs of primers amplified bands whose sizes were consistent with the presence of the group I intron (Pc1.SSU) in the 16S rRNA gene (4). As previously reported (7), ‘universal’ rRNA primers 3175 and 3176 also amplified a smaller DNA fragment of the size expected for a gene lacking the intron; the sequence of this fragment indicated that it was derived from the genome of the mammalian host rather than from P. carinii. Similarly, primers 4359 and 4358 amplified a fragment of the 26S rRNA gene, the size of which was consistent with the presence of the group I intron (Pc1.LSU) previously reported in Pc1 DNA (7).

As indicated in Figure 1, Panel B (above the gene map), use of primers 4359 and 4358 resulted in amplification of a fragment of the 26S rRNA gene from Pc2 DNA whose size was consistent with the presence of a group I intron at the same location as that present in Pc1 (Panel A). When this fragment was cloned and sequenced it contained a group I intron sequence (Pc2.LSU) with similarity but not identity to Pc1.LSU (Figure 2). Sequence analysis of the flanking exon regions in this clone showed the insertion site of this intron to be identical to that reported (7) for intron Pc1.LSU. However, PCR amplification of Pc2 DNA with six pairs of primers spanning the site of the group I intron in the 16S rRNA gene in Pc1 DNA only yielded fragments of a size consistent with the absence of such an intron (Figure 1, panel B). These results were confirmed by sequence analysis of cloned fragments generated with either primers 4434 and 4360, or 3175 and 3176, which demonstrated a sequence of a Pc1-related 16S rRNA gene without any intron (M. Ortiz, Y. Liu, R. Felder, and M.J. Leibowitz, unpublished). This fragment included a sequence identical to the 122-base sequence from a Pel-related 16S rRNA gene without any intron (M. Ortiz, Y. Liu, R. Felder, and M.J. Leibowitz, unpublished). This fragment included a sequence identical to the 122-base sequence from a

Figure 1. Intron location in rRNA operons of different P. carinii genotypes. Panel A shows a map of the rRNA operon of the Pc1 isolate of rat-derived P. carinii, where hatched regions code for rRNA, white bars indicate the two group I introns, and solid lines indicate internal spacers (4, 7). The bars below the map indicate the sizes of major PCR-amplified DNA products of Pel DNA in the presence of each pair of primers. The sizes of these PCR products were estimated by agarose gel electrophoresis followed by staining with ethidium bromide, using bacteriophage λ DNA digested with HindIII and φUX174 DNA digested with HaeIII (GIBCO-BRL) as molecular weight markers. Panel B shows a map consistent with the PCR products produced from rat-derived P. carinii Pc2 (7) and human-derived Pc3. The bars above this map indicate the sizes of the major products produced from Pc2 DNA with the indicated primers. The bars below the map show the sizes of products made from five isolates of human-derived P. carinii (denoted as Pc3 in the text).
Figure 3. Intron secondary structure. Panel A shows the secondary structure into which the group I intron in the gene for 26S rRNA from human-derived *P. carinii* (Pc3.LSU) can be folded. Here capital letters indicate intron sequence and lower case letters the flanking exon sequences. Arrows indicate the sites of cleavage and ligation in the intron self-splicing reaction. Helices P1–P9 are conserved among group I introns (17–18). Panel B indicates helix P10 (17–19) formed between the proximal region of the intron and the junction of the intron with the 3' exon (E2). This structure is drawn as it might exist after the first cleavage step of intron splicing has occurred, yielding a free 3' terminus on the 5' exon (E1) and a free 5' terminus on the intron to which a free guanosine nucleotide has been attached. The arrow indicates the site of cleavage 3' to the intron.

Thus, our data are consistent with the existence of two strains of *P. carinii* infecting immunosuppressed rats: Pc1, which contains group I introns in its 16S and 26S rRNA genes (3–7, 16), and Pc2, which differs from Pc1 in its 16S and 26S rRNA gene sequences (7, 16) and which only contains a group I intron in its 26S rRNA gene.

Figure 4. Comparison of structures of all group I introns. This dendrogram was generated by the 'pileup' program of the Wisconsin-GCG package indicating the sequence similarity (but not necessarily evolutionary relationships) among the eighty-two core nucleotides shared nearly by all group I introns. These are located at positions 95 to 99, 205 to 207, 210, 215 to 224, 254 to 256, 258 to 273, 278 to 283, 299 to 306, 313 to 316, 323 to 326, and 412 to 413 in intron Tt.LSU (19). Nomenclature of intron abbreviations is as in (19), where the letters and numbers before the dot indicate the organism, letters after the dot indicate the gene, and number after the comma indicates the ordinal number of the intron in genes containing more than one intron. Abbreviations and references are as cited (19) except for: As.SSU, *Ankistrodesmus stipitus* 16S rRNA (26) and Um.SSU, *Ustilago maydis* 16S rRNA (24). Note that introns Th.LSU and Trn.LSU are not listed since they are identical to Tp.LSU at all eighty-two positions. These introns include those found in bacteriophage, mitochondrial, chloroplast and eukaryotic nuclear genomes. Nuclear introns are indicated by an asterisk (*). When human-derived *P. carinii* DNA was amplified using primer pairs 4359 and 4170 or 4359 and 4358, major products were observed of the approximate size of a 26S rRNA gene containing a group I intron (Figure 1, Panel B, below the gene map). The PCR products generated from all five human-derived *P. carinii* samples with ‘universal’ primers 4359 and 4358 were
cloned and sequenced, and all yielded the same sequence denoted Pc3.LSU in Figure 2. This intron is located at the homologous site to that occupied by Pc1.LSU and Pc2.LSU in the two rat-derived P. carinii strains. All three introns in 26S rRNA genes of different strains are located between nucleotides 1921 and 1922 of these genes, numbered according to the numbering system for 26S rRNA of Escherichia coli (28).

Based on the sequence differences observed in the introns (Figure 2), we have tentatively assigned the trivial strain designation Pc3 to the five human isolates of P. carinii.

**Sequence comparison of group I introns**

The four P. carinii group I introns show considerable divergence, although several regions, including the phylogenetically conserved P, Q, R and S consensus sequences (17-18), tend to be relatively invariant (Figure 2). Figure 3 displays the likely secondary structure of intron Pc3.LSU, drawn to conserve the secondary structure proposed for group I introns (17-18), including base pairing of the 5' terminal region with both the flanking 5' exon (forming helix P1, Figure 3A) and with the flanking 3' exon (forming helix P10, Figure 3B). These base pairing interactions presumably play a role in selection of cleavage sites for intron excision (19). Therefore, it is not surprising that the 5' terminal region of Pc1.SSU, which can base-pair with phylogenetically conserved regions of 16S rRNA, is distinct from the 5' terminal sequence which is relatively conserved among the three LSU introns, which can base-pair with phylogenetically conserved regions of 26S rRNA (Figure 2). Introns Pc1.SSU (5, 6), Pc1.LSU (7) and Pc2.LSU (data not shown) can all be folded into structures resembling that shown in Figure 3.

All nuclear group I introns thus far reported are located in the small and large subunit rRNA genes. When the sequences of nuclear group I introns less than 500 bases long were compared, the introns of the four tetrahymena species (23) are much more similar to each other than are the four introns from P. carinii, although the latter introns are more similar to one another than to the introns of any other species (data not shown). When larger group I introns (some of which include extensive open reading frames) and the non-nuclear introns of mitochondrial, chloroplast and bacteriophage genomes are included, size variation between introns makes direct sequence comparison impossible. However, all intron sequences can be compared at 82 nucleotides at positions present in nearly all group I introns, referred to as core positions (19). The sequences of all 94 available group I introns are presented previously (19) plus additional sequences reported or compared in Figure 4, which graphically compares the sequences presented previously (19) plus additional sequences reported or cited here. Note that this analysis demonstrates the relative similarity of the nuclear group I introns except for introns Pp.LSU,1 and Pp.LSU,2 of Physarum polycephalum. In contrast, intron Pp.LSU,3 appears to resemble the cluster of similar nuclear group I introns, and is most closely related to the LSU introns of Tetrahymena.

**Intron-catalyzed splicing in vitro**

*In vitro* transcripts consisting of fragments of rRNA-containing introns Pc1.SSU (5–6) or Pc1.LSU (21) can catalyze intron excision and exon ligation *in vitro*. In this reaction, the primary transcript or precursor RNA first catalyzes cleavage of the 5' exon (E1)-intron (I) junction with attachment of the guanosine cofactor to the 5' end of the intron, to yield reaction intermediates E1 and guanosine-intron-3' exon (I-E2). Subsequently, the second reaction step yields free linear guanosine-intron (I) and ligated E1-E2 products.

The human-derived P. carinii intron Pc3.LSU also catalyzed its own excision *in vitro*. A linear DNA product containing the Pc3.LSU intron DNA flanked by short exon fragments linked to a bacteriophage SP6 promoter was amplified by PCR with primers 5573 and 4358 (Figure 1B). When this DNA was transcribed using bacteriophage SP6 RNA polymerase, a 454 nucleotide precursor RNA was produced, along with the various intermediates and products of the splicing reaction. The gel-purified precursor RNA ('P' in Figure 5) catalyzed intron splicing *in vitro* (Figure 5, lane 2) under conditions (21) optimal for splicing by transcripts containing Pc1.LSU. This reaction demonstrates that Pc3.LSU is, indeed, a self-splicing intron. Similar results have been obtained with intron Pc2.LSU (data not shown).

The anti-Pneumocystis drug pentamidine and a series of pentamidine analogues can act as non-competitive inhibitors of splicing by Pc1.LSU *in vitro* (21). As shown in Figure 5, pentamidine inhibits splicing by Pc3.LSU at similar concentrations to those inhibiting splicing by Pc1.LSU. As previously shown (21) for Pc1.LSU, the intron derived from human P. carinii is more sensitive to a series of pentamidine analogues, with inhibitory potency being pentamidine
Circularization of group I introns

Other group I introns have been observed to circularize after their excision as linear RNA molecules. This circularized intron is produced by linkage of the 3′ terminal G residue of the excised intron to a site near the 5′ terminus, with release of a short linear oligonucleotide from the 5′ terminus of the intron (29–31). This reaction appears to occur by a mechanism resembling the first step of the splicing reaction, with the 3′ hydroxyl group of the 3′-terminal G residue attacking a base in a new stem-and-loop formed at the 5′ end of the linear intron (stem helix denoted P') which resembles the position at the E1-I junction at which the free guanosine attacks during the first step of splicing.

When purified linear intron Pcl.LSU produced by splicing in vitro was reincubated under splicing conditions, it was converted into a band of RNA showing slower mobility on polyacrylamide gel electrophoresis (data not shown). In reactions starting with precursor RNA, increasing production of this slow-moving band was seen at concentrations of magnesium above 100 mM (21). The slow moving band produced from purified linear intron was electro-eluted from the gel and subjected to reverse transcription in the presence of primer 4613, followed by PCR amplification with primers 4612 and 4613 to yield a cDNA fragment containing the junction sequence 5′- AGUAGUGUG/UUUGAGGG-3′ connecting the 3′ terminal G with the U at position 6 of the intron Pcl.LSU (Figure 2). The complete sequence of the circularized form of the intron was confirmed by reverse transcription using primer 4615, PCR amplification with primers 4614 and 4615, cloning and sequence determination. The 5′ terminus of the linear intron can be folded into a stem-and-loop structure generated by helix P', as described above. Purified linear introns Pcl.LSU, Pc2.LSU and Pc3.LSU also yielded products with similar gel mobility to the circularized form of intron Pcl.LSU upon further reaction (data not shown). The intron sequence forming the stem-and-loop structure including helix P' is nearly identical (except for one nucleotide) in introns Pcl.LSU, Pc2.LSU and Pc3.LSU (Figure 2), while that of Pcl.SSU can form a similar helix P', although its sequence is distinct.

DISCUSSION

Taxonomy of P. carinii

Based upon the sequence similarity and homology of location between the Pc3.LSU intron and the introns in the 26S rRNA genes of P. carinii strains Pc1 and Pc2, the Pc3.LSU DNA is clearly from the P. carinii present in the two human bronchoalveolar lavage specimens. The location and sequence of group I introns in rRNA genes has allowed three distinct strains of P. carinii to identified. As previously reported (7), two distinct strains were identified in the lungs of immunosuppressed rats: Pc1 contains introns in both the 16S and 26S rRNA genes, while Pc2 only contains an intron in the gene encoding 26S rRNA. Multiple laboratories have detected DNA from both Pc1 (3–7, 16) and Pc2-like strains (7, 16). The five human isolates reported here all harbor a group I intron in the 26S rRNA gene without any evidence of an intron in the 16S rRNA gene. All five isolates have the same intron sequence, which is distinct from those of the two rat strains. We therefore have termed these human isolates as strain Pc3. Sequence analysis of additional isolates and genomic regions will be needed to determine if each strain is truly clonal in origin or if each may represent a family of non-identical but related strains.

The total number of P. carinii genotypes extant and their distribution among host species remains unknown. However, the recovery of both Pc1 and Pc2 from immunosuppressed rats indicates that each host species may harbor more than just a single host-specific strain of P. carinii. These results concur with karyotypic analysis suggesting that two strains of P. carinii infect rats (14–16), as well as with studies of surface antigens (9) and mitochondrial DNA sequences (32) of P. carinii which differ between isolates from different host species. These results contrast with the apparent identity of the 5S rRNA coding region among multiple P. carinii isolates from humans and rats (33). The relatively small size of 5S rRNA may limit its usefulness for phylogenetic studies (34–35). If multiple human isolates of P. carinii can be distinguished by the sequence of introns, variable coding regions or spacers in the rRNA operon, then this may represent a tool for study of the epidemiology of this infectious opportunistic pathogen. Although species definition for this unculturable organism is problematic, it is noteworthy that the three group I introns identified in the 26S rRNA genes of various isolates of P. carinii are more different from one another than are four introns found in 26S rRNA genes of different species of Tetrahymena (Figure 4).

The presence of a group I intron in the same site of the 26S rRNA gene of all characterized isolates of P. carinii suggests that this feature might be used as a target for diagnostic PCR amplification. Since no mammalian host or other pathogenic microorganism is known to have such an intron at this locus, primer pairs flanking the intron or pairs derived from intron and 26S rRNA sequence should amplify unique fragments from P. carinii DNA, even in the presence of contamination with DNA from host tissues, oropharyngeal flora or other opportunistic pathogens which may be found in clinical specimens. Further studies will be needed to test the utility of this approach to diagnosis of infections by P. carinii. The unique size of the PCR products derived from intron-containing P. carinii DNA might eliminate the need for hybridization with a radioactive probe, which has been used as a sensitive diagnostic method for detecting amplified DNA fragments lacking an intron (27).

It is difficult to interpret the phylogenetic implications of sequence comparisons of group I introns (19), especially considering the possibility that such introns might insert into new genes by more than one mechanism (reviewed in 36). According to the classification system of such introns (19), the group I introns of P. carinii may be grouped within group IC1.

Enzymology of group I intron splicing

The Pc3.LSU intron which we initially identified by sequence analysis has been confirmed to be a self-splicing group I intron. In the presence of magnesium, and guanosine nucleotide, it catalyzes its own excision from the 26S rRNA transcript, and further catalyzes circularization of the linear excised intron. Group I intron splicing in vitro is sensitive to inhibition by arginine and streptomycin (20, 37). We have found that tetracycline, etidium bromide, and pentamidine and its analogues inhibit this process for Pc1.LSU (21), and for the other group I introns of P. carinii including Pc3.LSU from the human clinical isolates. The apparent universality of group I introns in the rRNA operon of all isolates of P. carinii suggests the potential usefulness of antibiotics inhibiting nuclear group I intron splicing as anti-
Pneumocystis agents. The ability of the anti-Pneumocystis agent pentamidine to inhibit group I intron splicing in vitro suggests that this is a possible mechanism by which this and related compounds might act (21). This also suggests that antimicrobial agents acting against RNA targets in addition to rRNA in ribosomes can be low molecular weight compounds other than anti-sense oligonucleotides. The applicability of this approach to other ribozymes is now under investigation.

ACKNOWLEDGMENTS

Support in part by the AIDS Research Foundation, grant number 001996-14-RG from the American Foundation for AIDS Research and the New Jersey Center for Advanced Biotechnology and Medicine is gratefully acknowledged. Y.L. was supported by a fellowship from the Core Curriculum in Molecular and Cell Biology of Rutgers University and UMDNJ. Oligonucleotides were provided by L. Flores and DNA sequencing was performed by M. Rocourt and R. Felder of this department. We thank J.A. Kovacs (NIH) for providing DNA from human-derived P. carinii, J. Stringer (University of Cincinnati Medical Center) for sharing his sequence data on rat-derived P. carinii prior to publication, and R.R. Tidwell (University of North Carolina) for providing pentamidine analogues.

REFERENCES
