Repeated DNA in *Pneumocystis carinii*

SAUNDRA L. STRINGER, SUNG-TAE HONG,[†] DAVID GIUNTOLI, AND JAMES R. STRINGER*

Department of Molecular Genetics, Biochemistry, and Microbiology, College of Medicine,
University of Cincinnati, Cincinnati, Ohio 45267-0524

Received 3 December 1990/Accepted 18 March 1991

A 16-kb DNA fragment designated Rp3-1 and cloned from the genome of rat-derived *Pneumocystis carinii* was found to contain sequences that were repeated approximately 70 times per genome. The repeated sequences in Rp3-1 spanned at least 10.4 kb. Sequences in Rp3-1 were present on each of the 16 *P. carinii* chromosomes resolved by field inversion gel electrophoresis. Most of the *P. carinii* genomic sequences homologous to those in the Rp3-1 clone appeared to be as long as those in the Rp3-1 clone but were highly polymorphic with respect to restriction enzyme cleavage sites. The Rp3-1 DNA fragment appears to be a member of a family of large, degenerate, dispersed repeats.

Pneumocystis carinii is an opportunistic pathogen that has been a major cause of morbidity and mortality among AIDS patients (13, 17). The organism had been widely considered to be a protozoan, but phylogenetic analysis based on rRNA sequences placed the pathogen among the fungi (8, 29, 32). *P. carinii* was originally defined by the histopathological characteristics of pneumocystis pneumonia, and parasites displaying these histopathological characteristics, although from diverse mammalian hosts, are all classified under this rubric. Analysis of surface antigens has suggested that pneumocystis pneumonia may not be caused by a single genetic entity (14, 31), but this idea has been difficult to test, primarily because the parasite does not grow as a sustainable culture in vitro.

Recently, molecular genetic tools that promise to provide the means to understand better the genetic nature of *P. carinii* and the spectrum of genetic variability among organisms classified as *P. carinii* on the basis of histopathology have become available. In previous work, comparative analysis of *P. carinii* chromosomes separated by electrophoretic means showed that organisms isolated from different rat colonies displayed distinct chromosomal band patterns, suggesting the existence of distinct strains of *P. carinii* in nature (11). To understand better the genetic variability among *P. carinii* organisms, we have been studying repeated DNA in rat-derived *P. carinii*.

Repeated DNA has been found in most eucaryotes examined and has proved to be a useful marker for epidemiological studies (25). Some repeated DNAs serve structural roles in chromosomes, such as the sequences that cap the ends of chromosomes (2), while others arise from various forms of mobile genetic elements (10, 33). Some repeated DNAs resemble retrovirus proviruses. Examples of this group include the copia element in *Drosophila* spp. (18) and the Ty element in *Saccharomyces cerevisiae* (3, 4). A second group of repeated DNA elements is represented by the retroposons, which include mammalian long interspersed repetitive sequences (27), *Cin4* elements in maize (26), and the jockey element in *Drosophila* spp. (21). Like retroviruses, retroposons appear to move via an RNA intermediate (33). Members of a third group of repeated DNA elements, such

as *Drosophila melanogaster* P elements, move without the mediation of RNA transcripts (19). A fourth group is represented by the invertrons, which resemble the adenovirus genome. Like adenovirus DNA, these elements are typically quite large and have complex inverted-repeat sequences at their ends. Several elements of this type have been described in fungi (24).

In this report, we describe a family of repeated DNA sequences that compose a significant fraction of the genome of rat-derived *P. carinii*. The *P. carinii* repeat occurs in forms that exceed 10 kb in length and is present on each of the chromosomes resolvable by pulsed-field gradient electrophoresis. The repeat family contains as many as 100 members per genome, and there is a high degree of sequence heterogeneity among the repeat family members.

MATERIALS AND METHODS

Either Sprague-Dawley (Harlan Industries, Madison, Wis.) or Lewis (Charles River Breeding Laboratories, Wilmington, Mass.) rats were immunosuppressed with weekly injections of 4 mg of methylprednisolone acetate (Upjohn Laboratories, Kalamazoo, Mich.) and monitored for infection, and organisms were harvested as described previously (6).

P. carinii DNA was isolated as described previously (11). In brief, for each preparation of DNA, 2×10^{10} filtered, DNase-treated *P. carinii* organisms obtained from infected rat lung homogenates were lysed in 0.25 M EDTA containing 0.25 mg of proteinase K (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) per ml and 1% Sarkosyl at 55°C. The lysates were incubated at 55°C for 4 h. Digested lysates were extracted twice with phenol-chloroform and twice with chloroform alone, and the DNA was precipitated with ethanol. Approximately 20 µg of DNA was obtained. Rat DNA was isolated from cultured rat cells and purified as described previously (28).

The *P. carinii* genomic library was constructed in Lambda Fix (Stratagene Inc., La Jolla, Calif.) by standard procedures (15). Specifically, 3 µg of *P. carinii* DNA obtained from four Lewis rats was partially cleaved with *Sau3A*. The ends of the DNA were partially filled by treatment with Klenow polymerase in the presence of dGTP and dATP. The DNA was fractionated by electrophoresis through a 0.7% agarose gel. DNA migrating between size markers 23 and 16 kb was excised, extracted from the agarose, and ligated to 1

* Corresponding author.

[†] Present address: Department of Parasitology, Seoul National University College of Medicine, Seoul 110-460, Korea.

μ g of the Lambda Fix vector (Stratagene). The ligated DNA was packaged in vitro with Gigapack (Stratagene), and approximately 60,000 plaques were produced by infection of *Escherichia coli* LE392. Approximately 30% of these plaques were chimeric phage, as judged by plaque formation on *E. coli* P2 392 and by hybridization to a radiolabeled probe prepared by nick translation of a sample of the agarose gel-purified DNA used in the ligation to the Lambda Fix vector.

Methods for plasmid and phage DNA purification, restriction mapping, DNA labeling, DNA hybridization, Southern blotting, and cloning into plasmids were as described previously (15, 16). Southern blotting used Nytran membranes (Schleicher & Schuell, Keene, N.H.). Restriction enzymes were purchased from New England BioLabs, Inc. (Beverly, Mass.), and were used in accordance with vendor specifications. The plasmid vector used in subcloning fragments from Rp3-1 was pIBI30 (International Biotechnologies Inc., New Haven Conn.). The *P. carinii* thymidylate synthase gene was isolated from the lambda library by probing with a 57-base oligonucleotide containing a sequence from position 46 to position 102 of the gene described by Edman et al. (9).

Field-inversion gel electrophoresis (FIGE) was performed as follows. *P. carinii* organisms were suspended in 0.6% low-gelling-temperature agarose (Sigma, St. Louis, Mo.)–0.125 M EDTA at a density of 5×10^8 nuclei per ml. Gel-embedded cells were treated with 0.25 mg of proteinase K per ml–1% Sarkosyl at 55°C for 18 h. Digested agarose plugs were stored in 0.5 M EDTA (pH 9.0) at 4°C. FIGE gels were made of 1% agarose (Sigma) in 0.5 $\times$ TBE (45 mM Tris-HCl, 45 mM boric acid, 0.125 M EDTA [pH 7.4]) and run at 14 to 16°C in 0.5 $\times$ TBE. FIGE analysis was performed at 105 V for 120 h, 50 s forward and 25 s backward (5, 11). Southern blots were prepared from FIGE gels as described previously (11).

RESULTS

Isolation of a cloned copy of a *P. carinii* repeat. A library of chimeric lambda phage carrying segments of the genome from rat-derived *P. carinii* (11) was screened by hybridization to radioactive *P. carinii* DNA. About 20% of the chimeric phage plaques emitted a very strong radioactive signal, suggesting that these phage contained DNA fragments that were present in multiple copies in the *P. carinii* genome. The Rp3-1 phage was chosen for further analysis. The DNA in Rp3-1 did not hybridize to rat DNA (see Fig. 3).

Distribution of the DNA sequence in Rp3-1 among *P. carinii* chromosomes. The distribution of Rp3-1 sequences in the genome of *P. carinii* was examined by hybridization of radioactive Rp3-1 DNA to *P. carinii* chromosomes that had been resolved by FIGE. As shown in lane 1 of Fig. 1, Rp3-1 hybridized to each of the chromosomal bands resolved by FIGE. Some bands emitted a stronger signal than did others. In most cases, the difference in the radioactive signal intensity corresponded to a difference in the mass of DNA present in each band, as judged by staining with ethidium bromide. However, there were some exceptions, such as that seen by comparison of the three bands migrating at 525, 505, and 485 kb. Each of these bands was stained with the same intensity by ethidium bromide, but the 485-kb band was much darker on the radioautograph than were the two larger bands. This result suggested that the 485-kb band contained more copies of the repeat than did the two larger bands.

Location of repeated sequences within the 16-kb Rp3-1 fragment. The location of repeated sequences within Rp3-1

FIG. 1. Hybridization of Rp3-1 and fragments of Rp3-1 to *P. carinii* chromosomes. *P. carinii* chromosomes were separated by FIGE and transferred to Nytran filters. Identical filters were hybridized to various radiolabeled probes at 68°C in 6 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5 $\times$ Denhardt's solution–10 μ g of salmon sperm DNA per ml–0.2% sodium dodecyl sulfate. After hybridization, filters were washed in 0.2 $\times$ SSC–0.1% sodium dodecyl sulfate at 68°C. Radiolabeled probes were as follows: lane 1, Rp3-1; lane 2, *EcoRI*-E fragment; lane 3, *EcoRI*-C fragment; lane 4, *EcoRI*-G fragment; lane 5, *Sac*-E2 fragment. The locations of the fragments used as hybridization probes are shown in Fig. 2. The autoradiogram shown was obtained by exposure to X-ray film for 20 h at –70°C with one intensifying screen by a previously described method (15). Molecular weight markers were *S. cerevisiae* chromosomes.

was determined by hybridization of radioactive subcloned fragments derived from Rp3-1 to *P. carinii* chromosomes resolved by FIGE. Figure 2 shows the positions of restriction enzyme cleavage sites in the 16-kb Rp3-1 insert. Maps were determined by partial digestion of end-labeled DNA and confirmed by subcloning pieces of the phage insert into plasmids as described in Materials and Methods. The lambda vector had no sites for *EcoRI*; consequently, all five fragments excised from Rp3-1 by digestion with *EcoRI* were presumed to be from the *P. carinii* genome. All but *EcoRI*-F were subcloned. Three *SacI* fragments, *Sac*-D, *Sac*-E2, and *Sac*-F, were also subcloned. Because the lambda vector contained *SacI* sites flanking the cloning site, *SacI* fragments from the ends of the *P. carinii* insert terminated at sites in the vector. The seven subcloned fragments were used to map the repeated sequences in Rp3-1.

Figure 1 shows that each of the fragments tested hybridized to multiple chromosomal bands, indicating that a repeated sequence was present on each subcloned segment of Rp3-1. (Results obtained with *Sac*-F [not shown] were identical to those obtained with *Sac*-E2.) These results indicated that the repeated sequences were not localized to one part of Rp3-1 but were distributed throughout the length of the 16-kb fragment. The repeated sequences spanned at least 10.4 kb of DNA, the distance from *Sac*-F to *EcoRI*-E

A. **B.**

FIG. 2. Restriction enzyme maps of Rp3-1. Maps represent the 16-kb *P. carinii* DNA fragment present in Rp3-1. (A) Restriction enzyme cleavage sites for enzymes that cut the DNA several times (vertical lines). (B) Cleavage sites for enzymes that cut only once or twice in the Rp3-1 sequence.

(Fig. 2). Finding repeated sequences distributed throughout the Rp3-1 insert raised the possibility that the Rp3-1 insert may be internally redundant, with the same repeated sequence present in each of the subcloned Rp3-1 fragments. To examine this possibility, we hybridized radioactive subcloned fragments to restriction enzyme-digested Rp3-1 DNA fragments separated by gel electrophoresis. The experiment showed little or no cross-hybridization between most of the subcloned fragments. The only notable exception was *Sac*-F, which cross-hybridized with *Sac*-G (see Fig. 4). Therefore, it appeared that internal redundancy could not explain the results shown in Fig. 1. Further control hybridization experiments with plasmid vector sequences alone showed no hybridization to *P. carinii* chromosomes (data not shown), while a fragment derived from the *P. carinii* thymidylate synthase gene hybridized to a single band, as did a probe from the rRNA genes of *P. carinii* (data not shown).

It is interesting to note that while each subcloned DNA fragment hybridized to multiple *P. carinii* chromosomes, the band intensities obtained with each probe were not identical. The most obviously distinct band pattern was produced by *Eco*RI-E, which originated from the left end of Rp3-1. The differences in band patterns produced by the other subcloned probes were more subtle. Two examples of such differences can be seen by comparison of probes *Eco*RI-C

and *Eco*RI-G (Fig. 1, lanes 3 and 4). With *Eco*RI-C as a probe, the 575-kb band was darker than the 570-kb band and the 320-kb band was darker than the 330-kb band. With *Eco*RI-G as a probe, the opposite relative band intensities were observed.

Structure of Rp3-1-related sequences in the *P. carinii* genome. The results of the experiments described above suggested two possibilities regarding the relationship between the structure of the insert in phage Rp3-1 and the structure and organization in the *P. carinii* genome of sequences related to Rp3-1. Either Rp3-1 contained several different short sequences, each of which was repeated in the *P. carinii* genome but which were ordinarily not linked to one another, or Rp3-1 contained all or part of a copy of one member of a degenerate family of large (on the order of 10 kb) repeated elements. To distinguish these possibilities, we analyzed *P. carinii* genomic sequences related to those in Rp3-1. Three approaches were taken. In the first approach, we used plaque hybridization to determine whether sequences found in Rp3-1 were generally linked to one another in the *P. carinii* genome. We next used Rp3-1 and Rp3-1 subclones as probes with which to characterize fragments of the *P. carinii* genome containing sequences similar to those found in Rp3-1. Finally, we analyzed several chimeric phage that carried DNA fragments related to Rp3-1.

Sequences in Rp3-1 were often linked in the *P. carinii*

TABLE 1. Linkage and copy number of sequences in the *P. carinii* genome that are related to segments of Rp3-1

Class ^a	Result with probe ^b :						No. of plaques in class	% of Rp3-1-positive plaques ^c	% of total plaques ^d
	<i>EcoRI</i> -E	<i>EcoRI</i> -C	<i>EcoRI</i> -G	<i>EcoRI</i> -D	<i>Sac</i> -F	Thymidylate synthase			
A	+	+	+	+	+	—	100	47	11.8
B	—	+	+	+	+	—	45	21	5.3
C	—	—	+	+	+	—	11	5	1.3
D	—	—	+	—	+	—	2	0.9	0.24
E	—	—	—	+	+	—	6	3	0.71
F	—	—	—	—	+	—	3	1.4	0.35
G	+	—	—	—	—	—	41	19	4.9
H	+	+	—	—	—	—	3	1.4	0.35
I	+	+	—	+	+	—	1	0.5	0.12
J	+	—	—	—	+	—	1	0.5	0.12
K	—	+	—	—	—	—	0	0	0
L	—	—	+	—	—	—	0	0	0
M	—	—	—	+	—	—	0	0	0
TS	—	—	—	—	—	+	3	0	0.35

^a Plaques were classified according to the probes that produced a positive hybridization signal. For example, class A plaques hybridized to all probes, class B plaques hybridized to all probes except *EcoRI*-E, and class G plaques hybridized to only probe *EcoRI*-E.

^b Probes *EcoRI*-E, *EcoRI*-C, *EcoRI*-G, and *EcoRI*-D are shown in Fig. 2. The thymidylate synthase gene probe was used to estimate the copy number of RP3-1 related sequences relative to this single-copy gene. +, positive hybridization; —, negative hybridization.

^c Classes A through M comprised 213 plaques.

^d 843 chimeric phage plaques were analyzed.

genome. Rp3-1 fragments were used to screen the lambda library containing segments of the *P. carinii* genome. Nearly half of the plaques that hybridized to one Rp3-1 subclone also hybridized to all the other subclones (class A in Table 1), indicating that the repeated DNA sequences in the Rp3-1 clone were typically tightly linked in other members of this repeat family. The only subclone that appeared to be often unlinked to the others was *EcoRI*-E, which came from the left end of Rp3-1. Consistent with the plaque hybridization data, *EcoRI*-E also produced the most distinctive band pattern when hybridized to *P. carinii* chromosomes (Fig. 1).

The plaque screening experiment also provided an estimate of the number of copies of Rp3-1-related sequences per *P. carinii* genome. Of 843 chimeric phage plaques, 213 hybridized to at least one subcloned segment from Rp3-1. By contrast, a thymidylate synthase gene probe hybridized to only 3 of 843 plaques. Hence, Rp3-1-related sequences were about 70-fold more abundant in the phage library than were thymidylate synthase gene sequences. If the copy number of the thymidylate synthase gene is taken to be 1 per haploid genome, Rp3-1-related sequences would appear to number about 70 per haploid genome.

Restriction fragment length polymorphism in genomic DNA related to Rp3-1. Plaque hybridization indicated that sequences in each of the subcloned segments of Rp3-1 were usually linked in the *P. carinii* genome, suggesting that the Rp3-1 insert was not an oddity in which several otherwise autonomous repeats happened to be clustered but instead represented a member of a family of large repeated elements. If this were true, one might expect to find multiple copies of the entire Rp3-1 structure in the genome of *P. carinii*. Since repeated DNA is often hypervariable, one might also expect to find in the *P. carinii* genome variants of the sequence cloned in phage Rp3-1. To examine these possibilities, we used Rp3-1 and subclones of Rp3-1 to identify Rp3-1-related DNA fragments of the *P. carinii* genome produced by digestion with a battery of restriction enzymes. In these experiments, aliquots of *P. carinii* DNA were digested with restriction endonucleases, and the resulting fragments were separated by electrophoresis through an agarose gel. All *P.*

carinii DNA digestions were verified to be complete by analysis of reactions which were run in parallel and in which phage DNA was mixed with *P. carinii* genomic DNA. The separated fragments of *P. carinii* DNA were transferred to a nylon membrane, which was hybridized to a radioactive probe prepared from either the entire Rp3-1 sequence (Fig. 3) or subcloned fragments derived from Rp3-1 (Fig. 4).

Figure 3 shows that each of the five restriction enzymes produced a large number of bands with sequence homology to Rp3-1. Some bands were of sizes expected from the physical map of Rp3-1, and some bands were not. Rat DNA, run in lanes 2, 4, 6, 8, and 10, did not hybridize to Rp3-1. No homology between rat DNA and Rp3-1 could be detected, even after overexposure of the Southern blot to X-ray film (data not shown). Lanes 1 and 12 contained Rp3-1 phage DNA digested with *EcoRI* and *SacI*, respectively. In the case of *EcoRI*, all of the *P. carinii*-derived bands present in the phage appeared to be present in the *P. carinii* genomic DNA. The *P. carinii* genome also contained *SacI* fragments that comigrated with *Sac*-C, *Sac*-E, and *Sac*-G. *Sac*-D and *Sac*-F fragments would not be expected to occur in the *P. carinii* genome, because these two fragments end at *SacI* sites present in the phage vector (Fig. 2).

Figure 4 shows that each subcloned Rp3-1 fragment also hybridized to a complex array of *P. carinii* genomic fragments. The *EcoRI*-E fragment produced the simplest pattern, a result reminiscent of that obtained when this subclone was hybridized to *P. carinii* chromosomes (Fig. 1). The other subclones produced more complex collections of bands, within which it was usually possible to find bands that comigrated with bands produced from the cloned copy of Rp3-1, loaded in the odd-numbered lanes. Examination of the lanes containing Rp3-1 indicated that cross-hybridization between subclones occurred to various extents. The strongest cross-hybridization occurred between *Sac*-F and *Sac*-G, while *EcoRI*-D and *Sac*-E2 showed little cross-hybridization to other segments of Rp3-1. The extent of cross-hybridization between *Sac*-F and *Sac*-G depended on the stringency at which the blot was washed, indicating that

FIG. 3. Hybridization of R3-1 to restriction enzyme fragments of *P. carinii* genomic DNA. Five 5- μ g aliquots of *P. carinii* DNA (lanes 3, 5, 7, 9, and 11) and five 25- μ g aliquots of rat DNA (lanes 2, 4, 6, 8, and 10) were digested with the restriction enzymes shown at the top of the figure. The digested samples were electrophoresed through 0.75% agarose. DNA was transferred to Nytran, and the filter was hybridized to radiolabeled R3-1 DNA. Hybridization and washing conditions and autoradiographic detection were as described in the legend to Fig. 1. Lanes 1 and 12 contained 50 ng of R3-1 cleaved with *Eco*RI and *Sac*I, respectively.

these two segments were related in sequence but were not identical.

The data in Fig. 3 and 4 showed that the genome of *P. carinii* contained a complex collection of sequences related to those found in R3-1. It was possible to find bands that comigrated with bands produced from the cloned copy of R3-1, as well as other bands that were homologous to R3-1 but were of sizes not predicted from the locations of restriction enzyme cleavage sites in R3-1. These data are consistent with what would be expected if the genome of *P. carinii* contained both R3-1 replicas and R3-1 variants. An alternative explanation of the data would be that some segments of R3-1 also exist as small repeated elements, separate from the rest of R3-1. If this were the case, some of the fragments in Fig. 3 and 4 could have been derived from cleavage at restriction endonuclease cleavage sites lying in unique-sequence DNA flanking these small inserts. This possibility seems unlikely because, with the exception of *Eco*RI-E, R3-1-related sequences in the genome are rarely found disassociated from other R3-1 related sequences (Table 1).

Analysis of other chimeric phage carrying fragments related to R3-1. To further explore the idea that the *P. carinii* genome harbors variants of the R3-1 sequence that contain most or all of the sequences in R3-1 but that differ from the cloned R3-1 insert in the locations of restriction endonuclease cleavage sites, we characterized 10 chimeric phage that hybridized to *P. carinii* sequences from phage R3-1. DNA was isolated from each phage and digested with *Eco*RI and *Sac*I, and the resulting DNA fragments were sized by gel electrophoresis. Two phages contained fragments that comigrated with R3-1 fragments *Sac*-E, *Sac*-F, and *Eco*RI-G. The other eight phage genomes bore no resemblance to

FIG. 4. Hybridization of R3-1 subclones to restriction enzyme-digested *P. carinii* genomic DNA. The six panels show results obtained when six replicate Southern blots were hybridized to six different probes. *P. carinii* DNA was fragmented by incubation with the restriction enzymes *Eco*RI, *Bgl*II, *Hind*III, *Pvu*II, *Bam*HI, and *Sac*I (lanes 1, 3, 5, 7, 9, and 11, respectively). Similarly, R3-1 was fragmented with the same restriction enzymes and run in adjacent lanes (lanes 2, 4, 6, 8, 10, and 12). Each replicate filter was hybridized to a radiolabeled probe and washed as described in the legend to Fig. 1. Radiolabeled probes were as follows: panel A, R3-1; panel B, *Eco*RI-E fragment; panel C, *Eco*RI-C fragment; panel D, *Eco*RI-D fragment; panel E, *Sac*-E2 fragment; panel F, *Sac*-F fragment. The locations of the fragments used as hybridization probes are shown in Fig. 2. The *Bgl*II digest of R3-1 in lane 4 was a partial digest.

FIG. 5. Analysis of two other members of the Rp3-1 family. (A) Radioactive phage Rp4-1 DNA hybridized to *P. carinii* chromosomes resolved by FIGE. Experimental conditions were as described in the legend to Fig. 1. (B) Radioactive phage Rp4-1 DNA hybridized to restriction enzyme-digested *P. carinii* genomic DNA. (C) Radioactive phage Rp20-1 DNA hybridized to restriction enzyme-digested *P. carinii* genomic DNA. Lane assignments and experimental conditions for experiments in panels B and C were as described in the legend to Fig. 4.

Rp3-1, as determined by restriction mapping, but were homologous to parts of Rp3-1. Several phages were used as hybridization probes in experiments such as those described above. Figure 5 shows the results obtained with phage Rp4-1, which shared sequence homology with the *Eco*RI-D fragment of Rp3-1 (panels A and B), and phage Rp20-1, which shared sequence homology with the *Sac*-E, *Sac*-F, and *Sac*-G fragments of Rp3-1 (panel C). Rp4-1 hybridized to all the chromosomal bands of *P. carinii* resolved by FIGE and also hybridized to a complex array of fragments in restriction enzyme-digested *P. carinii* DNA. The pattern of *P. carinii* fragments homologous to Rp4-1 resembled the pattern seen when Rp3-1 *Eco*RI-D was used as a probe (Fig. 4). Rp20-1 also hybridized to many fragments in the *P. carinii* genome.

These data are consistent with the idea that most of the fragments seen in Fig. 3 and 4 were produced from Rp3-1 variants.

DISCUSSION

The experiments described above addressed two questions. First, what is the structure of the segment of the *P. carinii* genome carried in the chimeric phage Rp3-1? Second, what are the structures of Rp3-1 replicas and variants and how are these sequences organized in the *P. carinii* genome? With regard to the first question, analysis of the chimeric phage Rp3-1 established the following facts. (i) Rp3-1 con-

tained DNA that was repeated in the *P. carinii* genome. (ii) This repeated DNA was found on all of the *P. carinii* chromosomes resolvable by FIGE. (iii) The repeated sequences in phage Rp3-1 were distributed across restriction fragments that spanned at least 10.4 kb of DNA. (iv) The Rp3-1 sequence was not significantly internally redundant. The second question was more difficult to answer, but structural analysis of Rp3-1-related sequences in the *P. carinii* genome suggested that Rp3-1 is representative of a complex family of dispersed repeated elements, many of which are as large as Rp3-1 but are not exactly the same in sequence. The complexity and apparent degeneracy of the Rp3-1 family make it difficult to determine the size of a typical family member, but two lines of evidence support the idea that many repeats are at least 10 kb in length. First, hybridization to chimeric phage plaques showed that 47% of phages carrying sequences homologous to any part of Rp3-1 carried sequences homologous to all of Rp3-1. Second, consistent with these hybridization screening data, cloned chimeric phage that were homologous to Rp3-1 typically contained sequences that were homologous to most of the DNA in Rp3-1. These arguments do not rule out the existence of smaller repeats with homology to Rp3-1 sequences, but Rp3-1 appears to be representative of the family in terms of size. It is also possible that the repeated sequence in Rp3-1 is part of a repeated structure that is even larger than

the 16-kb Rp3-1 insert. We cannot rule out this possibility until the ends of the repeat are located.

It is also difficult to determine exactly how many copies of the repeat exist in the *P. carinii* genome, but the frequency of Rp3-1 sequences relative to that of thymidylate synthase sequences suggests that there are at least 70 copies of Rp3-1 family members per haploid genome. These estimates lead to the conclusion that the genome of *P. carinii* derived from infected rats is rich in repeated DNA. The genome of the parasite has been estimated from electrophoretic karyotyping to be 10^7 bp (11). Therefore, if the average Rp3-1-like sequence is 10 kb in length, 70 copies of such a sequence would contain 7×10^5 bp of DNA, which would compose 7% of the rat-derived *P. carinii* genome.

The function, if any, of Rp3-1 repeats is unknown. It is possible that Rp3-1 repeats are important structural components of *P. carinii* chromosomes, such as centromeres or telomeres (2). Alternatively, it is possible that Rp3-1 repeats are autonomous genetic elements that can propagate themselves and move about in the genome.

With regard to a possible structural role for Rp3-1, telomeres are the least likely candidate, for several reasons. First, telomeres are difficult to clone and would not be expected to be present in our genomic library. Second, the Rp3-1 family was not distributed equally among *P. carinii* chromosomes. Third, telomeres typically display a distinctive simple repeat sequence motif, such as the $C_{2-3}A(CA)_{1-3}$ motif found at the ends of *S. cerevisiae* chromosomes (30). The restriction enzyme cleavage patterns of Rp3-1 and its relatives are more complex than would be expected for such a simple repeat sequence, and partial sequence analysis of segments of Rp3-1 showed no telomeric-like repeat sequences. While it seems unlikely that Rp3-1 came from within a telomere, it is possible that Rp3-1 sequences are situated adjacent to telomeres. In *S. cerevisiae*, a sequence called the Y' element lies next to many, if not all, telomeres (7). The Y' element is 6.7 kb and is present in one to four copies per telomere. Telomere-adjacent elements similar to Y' have also been found in a number of other organisms, including *Drosophila* spp. (23, 34), *Xenopus laevis* (20), cereal grasses (1), nematodes (22), and slime mold (12).

It is also possible that Rp3-1 is a mobile genetic element. Rp3-1 is larger than the best-studied eucaryotic movable elements, such as Ty of *S. cerevisiae* and the P elements of *D. melanogaster*, but other movable elements, such as invertrons (24), which are known to occur in fungi, and large retroposons, such as long interspersed repetitive sequences (21, 26, 27), are often larger than the minimum size estimate for Rp3-1. With respect to the possible role of retrotransposition, it is interesting to note that Rp3-1 family members more often shared sequences with the right-most two-thirds of Rp3-1. This situation resembles that seen in retroposon families. Because retroposons are derived by reverse transcription of mRNAs, which is subject to premature termination, most retroposon family members contain sequences derived from the 3' end of the RNA, but only the largest members contain sequences derived from the 5' end. However, the frequent occurrence of phage clones that were homologous to only the left-most one-third of Rp3-1 is difficult to reconcile with the idea that Rp3-1 is a retroposon (Table 1).

Further structural analysis of Rp3-1 and its relatives should help determine whether Rp3-1 fits any of the categories of repeated DNA seen in other organisms. It will be particularly important to locate the ends of repeat family members, to determine whether repeats are located at spe-

cific chromosomal loci, and to determine if part or all of a repeat is present as RNA.

Whatever its origin or function, Rp3-1 should be useful as a target for DNA-based detection of *P. carinii* and as a means to distinguish strains of the organism. The high copy number of Rp3-1 would provide an enhanced signal in procedures such as DNA hybridization and the polymerase chain reaction. DNA fingerprinting with Rp3-1 repeats has potential for detecting strain variation and assessing strain relatedness among *P. carinii* isolates.

ACKNOWLEDGMENTS

This work was supported by grant AI 28471 from the Public Health Service in conjunction with a Public Health Service program project grant in mycology (PO1 AI 28392).

We thank M. T. Cushion for help with isolating *P. carinii* and for helpful comments on the manuscript.

REFERENCES

1. Bedbrook, J. R., J. Jones, M. O'Dell, R. D. Thompson, and R. B. Flavell. 1980. A molecular description of telomeric heterochromatin in *Secale* species. *Cell* 19:545-560.
2. Blackburn, E. H., and J. W. Szostak. 1984. The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem.* 53:163-194.
3. Boeke, J. D., D. J. Garfinkel, C. A. Sytles, and G. R. Fink. 1985. Ty elements transpose through an RNA intermediate. *Cell* 40:491-500.
4. Cameron, J. R., E. Y. Loh, and R. W. Davis. 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell* 16:739-751.
5. Carle, G. F., M. Frank, and M. V. Olsen. 1986. Electrophoretic separation of large DNA molecules by periodic inversion of the electric fields. *Science* 232:65-68.
6. Cushion, M. T., J. J. Ruffolo, and P. D. Walzer. 1988. Analysis of the developmental stages of *Pneumocystis carinii* in vitro. *Lab. Invest.* 58:324-331.
7. Dunn, B., P. Szauter, M. L. Pardue, and J. W. Szostak. 1984. Transfer of yeast telomeres to linear plasmids by recombination. *Cell* 39:191-201.
8. Edman, J. C., J. A. Kovacs, H. Masur, D. V. Santi, H. J. Elwood, and M. L. Sogin. 1988. Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the fungi. *Nature (London)* 334:519-522.
9. Edman, U., J. C. Edman, B. Lundgren, and D. V. Santi. 1989. Isolation and expression of *Pneumocystis carinii* thymidylate synthase gene. *Proc. Natl. Acad. Sci. USA* 86:6503-6507.
10. Haber, J. 1983. Mating type genes in *Saccharomyces cerevisiae*, p. 559-650. In J. A. Shapiro (ed.), *Mobile genetic elements*. Academic Press, Inc., New York.
11. Hong, S.-T., P. E. Steele, M. T. Cushion, P. D. Walzer, S. L. Stringer, and J. R. Stringer. 1990. *Pneumocystis carinii* karyotypes. *J. Clin. Microbiol.* 28:1785-1795.
12. Johnson, E. M. 1980. A family of inverted repeat sequences and specific single-strand gaps at the termini of the *Physarum* rDNA palindrome. *Cell* 22:875-886.
13. Kovacs, J. A., J. W. Hiemenz, A. M. Macher, D. Stover, H. W. Murray, J. Shelhamer, H. C. Lane, C. Urmacher, C. Honig, D. L. Longo, M. M. Parker, C. Natanson, J. E. Parrillo, A. S. Fauci, P. A. Pizzo, and H. Masur. 1984. *Pneumocystis carinii* pneumonia: a comparison between patients with the acquired immunodeficiency syndrome and patients with other immunodeficiencies. *Ann. Intern. Med.* 100:663-671.
14. Linke, M. J., M. T. Cushion, and P. D. Walzer. 1989. Properties of the major rat and human *Pneumocystis carinii* antigens. *Infect. Immun.* 57:1547-1555.
15. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*, p. 194-195. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. Meade, J. C., J. Shaw, S. Lemaster, G. Gallagher, and J. R. Stringer. 1987. Structure and expression of a tandem gene pair

- in *Leishmania donovani* that encodes a protein structurally homologous to eucaryotic cation-transporting ATPases. *Mol. Cell. Biol.* 7:3937-3946.
17. Mills, J. 1986. Pneumocystis carinii and Toxoplasma gondii infections in patients with AIDS. *Rev. Infect. Dis.* 8:1001-1011.
 18. Mount, S. M., and G. M. Rubin. 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol. Cell. Biol.* 5:1630-1638.
 19. O'Hare, K., and G. M. Rubin. 1983. Structure of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* 34:25-35.
 20. Pardue, M. L. 1974. Localization of repeated DNA sequences in *Xenopus* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38:475-482.
 21. Priimagi, F., L. J. Mizrokhi, and Y. V. Ilyin. 1988. The drosophila mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. *Gene* 70:253-262.
 22. Roth, G. E., and K. B. Moritz. 1981. Restriction enzyme analysis of the germ line limited DNA of *Ascaris suum*. *Chromosoma* 83:169-190.
 23. Rubin, G. 1978. Isolation of a telomeric DNA sequence from *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* 42:1041-1046.
 24. Sakaguchi, K. 1990. Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. *Microbiol. Rev.* 54:66-74.
 25. Scherer, S., and D. A. Stevens. 1988. A *Candida albicans* dispersed, repeated gene family and its epidemiologic applications. *Proc. Natl. Acad. Sci. USA* 85:1452-1456.
 26. Schwarz-Sommer, Z., L. Leclercq, E. Gobel, and H. Saedler. 1987. Cin4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons. *EMBO J.* 6:3873-3880.
 27. Singer, M. F. 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* 28:433-434.
 28. Stringer, J. R., R. M. Kuhn, J. L. Newman, and J. C. Meade. 1985. Unequal homologous recombination between tandemly arranged sequences stably incorporated into cultured rat cells. *Mol. Cell. Biol.* 5:2613-2622.
 29. Stringer, S. L., J. R. Stringer, M. A. Blase, P. D. Walzer, and M. T. Cushion. 1989. Pneumocystis carinii: sequence from ribosomal RNA implies a close relationship with fungi. *Exp. Parasitol.* 68:450-461.
 30. Szostak, J. W., and E. H. Blackburn. 1982. Cloning yeast telomeres on linear plasmid vectors. *Cell* 29:245-255.
 31. Walzer, P. D., and M. J. Linke. 1987. A comparison of the antigenic characteristics of rat and human *Pneumocystis carinii* by immunoblotting. *J. Immunol.* 138:2257-2265.
 32. Watanabe, J., H. Hori, K. Tanabe, and Y. Nakamura. 1989. Phylogenetic association of *Pneumocystis carinii* with the 'Rhizopoda/Myxomycota/Zygomycota' indicated by comparison of 5S ribosomal RNA sequences. *Mol. Biochem. Parasitol.* 32:163-168.
 33. Weiner, A. M., P. L. Deininger, and A. Efstratiadis. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the inverse flow of genetic information. *Annu. Rev. Biochem.* 55:631-661.
 34. Young, B. S., A. Pession, K. L. Traverse, C. French, and M. L. Pardue. 1983. Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. *Cell* 34:85-94.