INFECTION AND IMMUNITY, Aug. 1994, p. 3092–3101
0019-9567/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 62, No. 8

Molecular Characterization of Clustered Variants of Genes Encoding Major Surface Antigens of Human *Pneumocystis carinii*

THOMAS R. GARBE* AND JAMES R. STRINGER

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

Received 9 February 1994/Returned for modification 9 April 1994/Accepted 3 May 1994

A 13-kb genomic fragment from human *Pneumocystis carinii* was cloned as repetitive DNA. The fragment contains a cluster of three related genes, each 3 kb in size, and the 5′ end of a fourth gene. The predicted polypeptide of the first gene in the cluster comprises 1,030 amino acid residues with a total molecular mass of 116 kDa. The gene's predicted amino acid sequence bears 32% identity to predicted sequences of recently described gene fragments of ferret *P. carinii*, which encode an immunodominant surface glycoprotein (*gpA*) (P. J. Haidaris, T. W. Wright, F. Gigliotti, and C. G. Haidaris, J. Infect. Dis. 166:1113–1123, 1992), and 36% identity to the predicted sequence of a rat *P. carinii* major surface glycoprotein gene (*msg*) (J. A. Kovacs, F. Powell, J. C. Edman, B. Lundgren, A. Martinez, B. Drew, and C. W. Angus, J. Biol. Chem. 268:6034–6040). DNA hybridization showed that sequences related to the cloned *msg* genes reside on at least 12 chromosomes of human *P. carinii* at various degrees of multiplicity and/or homology. Affinity-purified antibodies with specificity to a fusion protein made from the human *P. carinii msg*I gene recognized two bands on a Western immunoblot containing total human *P. carinii* protein; they also recognized fusion proteins derived from the other two genes of the cluster. Monoclonal antibodies with reactivity to Msg of human *P. carinii* recognized fusion proteins produced from two *msg* genes. Fusion proteins were also recognized by sera from healthy humans and from patients. The *msg* genes are candidates for the development of immunotherapy and subunit vaccines for the treatment and prevention of *P. carinii* pneumonia.

Pneumocystis carinii infections are ubiquitous in humans. Serologic studies with children showed that 80 to 100% of all healthy 4-year-olds had immunoglobulins specific to *P. carinii* (27, 31). In immunocompromised individuals and 50% of all AIDS patients (4), the organism can multiply to extraordinary levels, mainly in the cavities of pulmonary alveoli, causing life-threatening pneumonia (28, 46). Chemotherapy is usually effective but frequently leads to adverse reactions (25). Development of immunotherapy for *P. carinii* pneumonia is a credible possibility. It has been shown that the number of *P. carinii* organisms in *scid* mice could be drastically reduced and survival could be increased by continued administration of hyperimmune antiserum (36) and that *P. carinii* pneumonia in ferret and rat animal models could be mitigated by administration of a monoclonal antibody (MAb) to a 116-kDa major surface antigen of *P. carinii* (13). The 116-kDa antigen is now known to be a family of polypeptides encoded by a large number of heterogeneous genes in *P. carinii* from both rats and ferrets (15, 21, 41). In rat *P. carinii* this family is known as the major surface glycoprotein (Msg) (21, 41), and in ferret *P. carinii* it is known as gpA (15). Human *P. carinii* is genetically (6, 38, 41, 49, 50) and serologically (12, 20, 47) distinct from *P. carinii* of rats and ferrets. Therefore, it is essential to isolate the human *P. carinii* antigen gene(s) for molecular characterization and production of the antigen(s).

We report here on the binding of immunoglobulins in sera from healthy humans and patients to fusion gene products from three different human *P. carinii msg* genes, which we have cloned as repetitive DNA and subcloned into the λgt11 expression vector. The complete nucleotide sequence of one gene and partial sequences of three other genes have been determined.

* Corresponding author. Mailing address: Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati, P.O. Box 670524, 231 Bethesda Ave., Cincinnati, OH 45267-0524. Phone: (513) 558-0097. Fax: (513) 558-8474.

MATERIALS AND METHODS

DNA manipulation. Standard DNA manipulations were performed as described by Sambrook et al. (37). The DNA molecular size marker was the 1-kb ladder supplied by Gibco-BRL. Southern hybridizations were carried out according to the Blotto protocol (37) with Hybond N membrane (Amersham) for transfers. Pulsed-field gel electrophoretic separation and transfer of *P. carinii* chromosomes onto Hybond N^+ membrane (Amersham) was as described previously (18); the blot shown in Fig. 3 was a gift of M. Cushion and was hybridized according to Denhardt's protocol (37). Stringent Southern hybridization washes were done twice in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS) at 68°C for 90 min (37). Washes at reduced stringency were done twice in 2× SSC–0.1% SDS at 42°C for 60 min (10). Removal of labeled probe DNA from Southern blots prior to reuse was performed by incubating the blot in 0.4 M NaOH at 50°C for 1 h, following the recommendations of Amersham.

DNA. Agarose-embedded rat and human *P. carinii* organisms (16) were a generous gift of M. Cushion. Human DNA was extracted from HeLa cells (supplied by the American Type Culture Collection) by the λ phage DNA extraction protocol (37). Plasmid pBlur8 (35) carrying repetitive human *Alu* sequence was a gift of A. Menon. The rat *msg* gene used for the Southern hybridization shown in Fig. 4 was present in two Bluescript subclones, each carrying one of the two consecutive *Sac*I restriction fragments E2 and F of clone Rp3-1 (42, 42a), which were independently labeled and mixed together for the hybridization. Plasmid vector pBluescript SKII(+) was from Stratagene and was used for the subcloning of *Xba*I and

*Hin*dIII fragments of clone TG/61, using phenolized total digests for ligation and *Escherichia coli* DH5α (37) as the transformation recipient.

DNA labeling with ^{32}P was performed by using the Random-Primed DNA-Labeling Kit supplied by Boehringer Mannheim. For the labeling of genomic human *P. carinii* DNA, DNA was electrophoresed from the agarose-embedded organisms (16) into 0.3% low-melting-point agarose (SeaPlaque) at 20 V overnight. Gel that contained high-molecular-weight DNA was melted at 70°C and digested by overnight incubation at 37°C with 1 U of agarase (Calbiochem); DNA was not precipitated. Isolation of DNA restriction fragments for labeling was analogous, except that 0.7 to 1.0% agarose gels were used.

Construction and screening of λ replacement and λgt11 expression libraries. Human *P. carinii* DNA in melted 0.3% low-melting-point agarose (see above) was digested with *Sau*3AI (Bethesda Research Laboratories) in the presence of the recommended buffer in order to generate fragments in the size range 15 to 20 kb (19). After heat inactivation of the restriction enzyme at 70°C for 10 min, the agarose was hydrolyzed by overnight incubation at 37°C in the presence of 1 U of agarase. DNA ends were dephosphorylated by a 30-min treatment with 10 U of alkaline phosphatase (Boehringer Mannheim). To inactivate the enzyme, the sample was brought to 20 mM ethylene glycol-bis(β-aminoethyl ether)-*N,N,N′,N′*-tetraacetic acid (EGTA) (pH 7) and heated at 70°C for 15 min. After phenolization and ethanol precipitation in the presence of 100 mM NaCl, the DNA was dissolved in TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). One microgram of *Sau* 3A-digested *P. carinii* DNA was ligated to 1 μg of *Bam*HI-digested λ dash DNA (Stratagene), using buffer and T4 DNA ligase from Bethesda Research Laboratories in a final volume of 5 μl. Ligations were packaged with λ packaging extract Gigapack I Gold (Stratagene) and amplified as recommended, using *E. coli* indicator strain 392(P2) (Stratagene) of the Spi$^-$ phenotype, to select phages containing foreign DNA (19). Plaques were lifted from three 100-mm-diameter plates with 30,000 plaques each (37), and the DNA was fixed to Nytran membranes (Schleicher & Schuell) by UV cross-linking (Stratagene). Large-scale λ dash lysates were prepared as described previously (37), using *E. coli* SRB, supplied with the λ dash vector. λ stocks with a titer of $\geq 10^{11}$ PFU/ml were purified by consecutively sedimenting and floating the virions through CsCl block gradients (5); DNA was liberated by the formamide method (5).

For the construction of a λgt11 subclone expression library, 2 μg of TG/61 DNA, in a volume of 100 μl of TE, was sheared to an average size of 3 kb by sonication under conditions precalibrated with λ DNA (Gibco-BRL). Further procedures were as described previously (17). Ten presumptive recombinant plaques (17) were tested for inserts by PCR, using primers and conditions as described by Thole et al. (43); eight plaques each produced a band, with an average size of 2 kb. The original library contained 10^5 recombinant plaques and was amplified (17), prior to screening. Serologic screening of the library was done as described previously (17), using 150-mm-diameter L-broth agar plates with 10^5 plaques per plate. High-titer phage stocks from λgt11 clones were obtained from one 150-mm L-broth agar plate each (5); purification of phage stocks and DNA liberation were as described for λ dash above.

Nucleotide sequencing. DNA was sequenced with the cycle sequencing system from Bethesda Research Laboratories at the suggested temperature cycling program. The program consisted of 20 cycles of a denaturing step at 95°C for 30 s, an annealing step at 55°C for 30 s, and an extension-termination step at 70°C for 60 s, followed by 10 cycles of a denaturation step at 95°C for 30 s and an extension-termination step at 70°C for 60 s. The λgt11 forward primer was primer no. 499 described by Thole et al. (43). The reverse primer 5′d(TTGA CACCAGACCAACTGGTAATG)3′ was from New England Biolabs. T3 and T7 primers for the sequencing of Bluescript plasmids and of λ dash (TG/61) were from Stratagene. Gene-specific primers for sequencing were custom made. Second-strand sequencing was partially performed by using the automated Applied Biosystems 370A DNA sequencer with the Taq Dye Deoxy Terminator Cycle Sequencing Kit. Alignment of nucleotide and amino acid sequences was performed with the ALIGN program, version 1.02 (Scientific and Educational Software, 1989). The amino acid computer search was performed at the National Center for Biotechnology Information with the Blast network service (1).

Sera. Hyperimmune serum against whole human *P. carinii* was a gift of M. Linke and P. Walzer. It was produced in rabbits as described previously (22) by using complete Freund's adjuvant (Pierce) for the first injection and incomplete adjuvant (Pierce) for all succeeding booster injections (21a). We thank G. Smulian and P. Walzer for the provision of human sera and MAbs. Sera from healthy donors were from Hoxworth Blood Center, Cincinnati, Ohio. MAb 2A1 was raised against purified human Msg; MAb 4-4E7 was raised against whole rat *P. carinii*. By Western blot (immunoblot) analysis, both MAbs recognized a band of human *P. carinii* protein with a mobility equal to that of the band recognized by MAb 85-1-5E12 (12) (39a), which is specific to Msg from human *P. carinii*. Affinity purification of hyperimmune rabbit serum against human *P. carinii* was done as described previously (2), using 30 ml of a 1/200 dilution on plaque lifts from TG6107. Horseradish peroxidase-conjugated antibodies, enzyme-linked immunosorbent assay grade, were from Bio-Rad; anti-human immunoglobulin G–horseradish peroxidase conjugate was from Pierce.

Protein analysis and Western immunoblotting. Human *P. carinii* organisms were a gift of M. Linke and P. Walzer. For the expression of fusion protein from the λgt11 subclones, 5 ml of an exponential-phase culture of *E. coli* Y1089 lysogens in LB was thermally induced at 42 to 45°C for 15 min, made 10 mM with 1 M isopropyl-β-D-thiogalactopyranoside (IPTG) (17), incubated for 90 min at 37°C, spun at 8,000 × *g* and 4°C for 10 min, and suspended in 400 μl of water containing 50 μg of phenylmethylsulfonyl fluoride (Sigma) per ml for the inhibition of sensitive proteinases. For polyacrylamide gel electrophoresis (PAGE) under denaturing conditions, 10-μl samples were boiled for 3 min in a water bath in the presence of 0.062 M Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, 5% β-mercaptoethanol, and 0.001% bromophenol blue. SDS-PAGE molecular weight standards were either broad or high range and were provided by Bio-Rad. Electrophoresis was in discontinuous 8% acrylamide slab gels (37), using the Mighty Small equipment supplied by Hoefer Scientific Instruments. Proteins were transferred to Millipore Immobilon-P membranes, where they were stained by various sera by techniques described previously (44).

Nucleotide sequence accession number. The sequence reported for human *msg*I has been entered into GenBank under accession number 27092.

RESULTS

Clustered variants of the major surface glycoprotein gene. It is known that the major surface glycoprotein (*msg*) genes of rat *P. carinii* are repetitive (21, 45). Therefore, we reasoned that the homologous genes in human *P. carinii* might be accessible by cloning repetitive DNA. Rat *P. carinii* clone Rp3-1, which

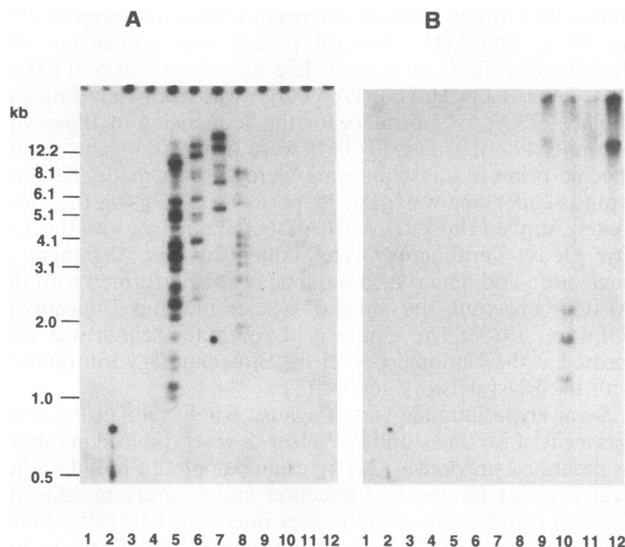

FIG. 1. Clone TG/61 is specific to human *P. carinii*. Rat *P. carinii* DNA (1 μg) (lanes 1 to 4), human *P. carinii* DNA (2 μg) (lanes 5 to 8), and human HeLa cell DNA (10 μg) (lanes 9 to 12) were each digested with *Hin*dIII, *Eco*RI, *Sal*I, and *Xba*I, respectively, electrophoresed into a 0.7% agarose gel, and blotted onto Hybond N membrane by using a Stratagene pressure blotter. Hybridization to labeled DNA probes was followed by high-stringency washes (37) and autoradiography at −70°C for 3 days. (A) Samples probed with TG/61; (B) samples probed with TG/91.

was cloned as repetitive DNA (42), was subsequently found to carry two *msg* genes (41, 42a). Accordingly, we screened an amplified λ dash replacement library of human *P. carinii* with human *P. carinii* (donor) DNA. Since donor DNA and the library were likely to contain host DNA, we hybridized duplicate plaque lifts with the recombinant *Hin*dIII fragment isolated from pBlur8 (35), which contains the repetitive *Homo sapiens Alu* sequence. Of 29 plaques (of 10^5 recombinant plaques) that signaled with *P. carinii* donor DNA, three did not signal with the human *Alu* sequence. One of these clones, TG/61, contained *P. carinii* DNA. Figure 1A shows that labeled TG/61 DNA hybridized to human *P. carinii* DNA (lanes 5 to 8) but not to human HeLa cell DNA or to rat *P. carinii* DNA. The complexity of the band patterns indicated that TG/61-related sequences were present in multiple copies. After removal of the TG/61 label, the same blot was hybridized to another of the three clones (TG/91), which hybridized solely to human HeLa DNA (Fig. 1B, lanes 9 to 12).

To examine the distribution of TG/61 sequences in the genome, pulsed-field gel electrophoretically separated chromosomes of *P. carinii*, isolated from three different hosts, were subjected to stringent Southern hybridization analysis using the *P. carinii* DNA fragment from λgt11 subclone TG6106 of TG/61 (Table 1 and Fig. 2) as a probe. The result is shown in Fig. 3. All but two human *P. carinii* chromosomes hybridized (lane 3), showing that the sequence within the 2.9-kb fragment carried on TG6106 represents a copy of a family of sequences that are represented throughout the human *P. carinii* genome (41). Under the stringent washing conditions used, the chromosomes of neither rat *P. carinii* (lane 1) nor ferret *P. carinii* (lane 2) hybridized. We next examined TG/61 DNA for evidence of internal sequence redundancy by using the ^{32}P-labeled 1.63-kb *Xba*I restriction fragment of TG/61 (Fig. 2) to probe TG/61 restriction fragments in a Southern hybridization

TABLE 1. Recombinant λgt11 subclones of TG/61[a]

Clone	*msg* gene of origin	Insert length (kb)	Encoded amino acid residues that align with MsgI[b]	Size of fusion protein (kDa)[c]	
				Predicted	Serologic stain
TG6101[d]	II	1.72	**430–830**	180	130
TG6103[d]	II	0.14	**270–310**	120	120
TG6104	III	3.05	**80–1030**	220	180–130
TG6105	I	3.40	**75**–1030	220	190–140
TG6106	I	2.87	**50–1000**	210	190–140
TG6107	I	0.64	**260**–480	140	145
TG6108	I	0.70	**1–225**	140	NS[e]
TG6109	III	2.14	**125**–840	200	200–130
TG6111	I	3.50	**130**–1030	230	160–130
TG6111	II	3.50	1–**60**	NI[f]	NI
TG6113	III	0.80	**145–450**	150	165
TG6115	III	3.54	**70**–1030	215	195–145
TG6115	IV	3.54	1–**65**	NI	NI
TG6116[d]	II	0.14	**270–310**	120	120
TG6117	I	2.30	**1–750**	200	NS
TG6118	I	0.48	**260–420**	135	140
TG6119	II	3.12	**50–1030**	225	200–140
TG6120	III	1.88	**280–950**	190	200–120
TG6121	I	0.67	**240**–465	140	160

[a] A λgt11 subclone library was constructed from clone TG/61 and screened with hyperimmune rabbit serum raised against human *P. carinii*. DNAs of the 17 isolated clones were analyzed by digestion with *Eco*RI to estimate the length of the recombinant inserts and by DNA sequencing.

[b] Boldface indicates that the insert terminus was directly sequenced rather than extrapolated from the insert length. Determination of insert lengths was performed by digestion of phage DNA with *Eco*RI followed by electrophoresis in 0.7% agarose.

[c] The predicted size of the fusion proteins is the sum of the molecular mass of 116 kDa for β-galactosidase and the molecular mass of the recombinant amino acid residues deduced from the length of the insert and the start of the amino terminus, assuming an average molecular mass of 112.6 Da per amino acid residue. Serologic staining for Western blot analysis was done by hyperimmune rabbit serum against human *P. carinii* at a dilution of 1/1,000.

[d] Clones TG6101, -03, and -16 were assigned to *msg*II on the basis of DNA sequencing, which showed that they could not have come from *msg*I or *msg*III.

[e] NS, no signal.

[f] NI, not investigated.

experiment. The result, as shown in Fig. 4 (lane 2), revealed that the sequence on the 1.63-kb *Xba*I fragment was repeated almost throughout clone TG/61, suggesting the presence of repetitive sequence. Only two *Xba*I fragments (1.7 and 1.84 kb) failed to hybridize; both are located next to the right λ vector arm (Fig. 2). To determine if TG/61 contained sequences related to *msg* genes from rat *P. carinii*, we used Rp3-1 rat *msg* genes as a hybridization probe. Following reduced-stringency washes, the rat *msg* gene probe displayed strong hybridization to both a 1.63- and a 3.9-kb *Xba*I fragment; three more fragments produced a weaker hybridization signal, leaving the 1.70-kb fragment located next to the right λ vector arm the only nonhybridizing fragment (Fig. 4, lane 3).

The major surface glycoprotein gene of human *P. carinii* should produce a protein recognizable by antisera to human *P. carinii*. To determine if TG/61 carried sequences encoding a human *P. carinii* antigen, we constructed a λgt11 subclone expression library. After screening of plaques with hyperimmune rabbit serum against human *P. carinii*, over 150 plaques produced a positive donut-shaped signal. Seventeen λgt11 subclones (Table 1) of various signal strengths were isolated and subjected to nucleotide sequence analysis. Three distinct genes homologous to rat *msg* were identified and located on clone TG/61 by restriction mapping (Fig. 2). Since the 1.70-kb *Xba*I fragment did not signal in a Southern hybridization with

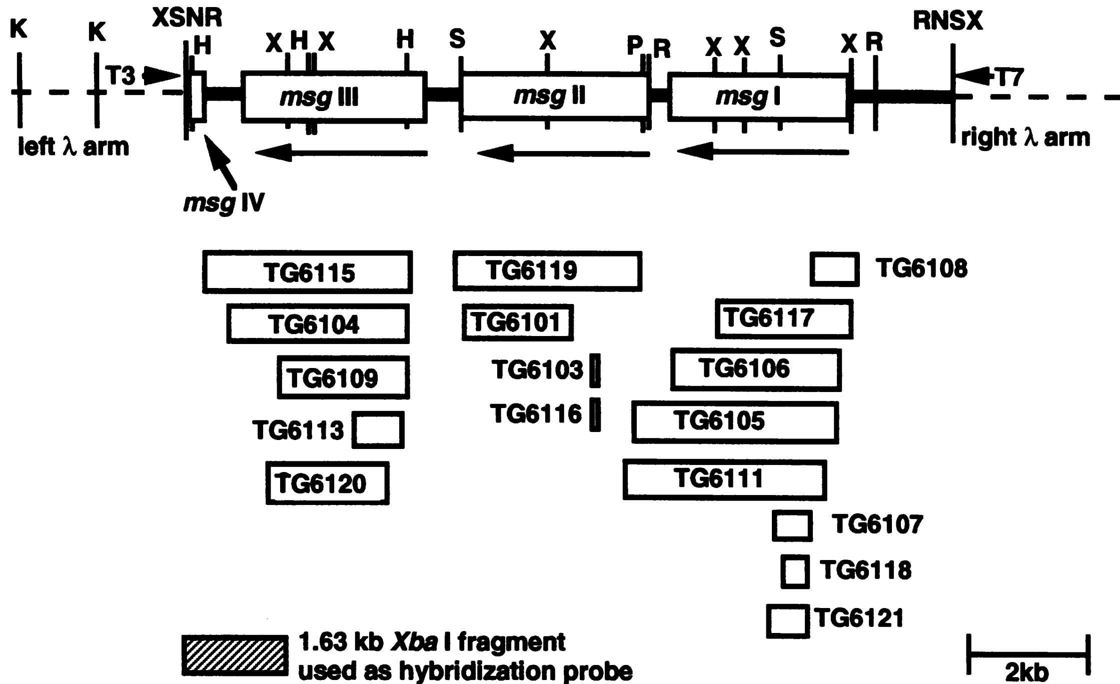

FIG. 2. Restriction map of clone TG/61 and locations of λgt11 subclones. Abbreviations and symbols: H, *Hin*dIII; K, *Kpn*I; N, *Not*I; P, *Spe*I; R, *Eco*RI; S, *Sal*I; X, *Xba*I; arrows labeled T3 and T7, primer sites on λ dash vector; open boxes labeled *msg*I, *msg*II, *msg*III, and *msg*IV, open reading frames. The direction of transcription of each gene is from right to left, as indicated by arrows underneath the boxes.

the 1.63-kb *Xba*I fragment or with the rat *msg* gene (Fig. 4), it is unlikely that this fragment carries any *msg*-related sequence. Therefore, *msg*I appeared to be the first gene in the cluster. The beginning of a fourth *msg*-homologous gene (*msg*IV) (Fig. 2) was found at the end of TG6115 by sequence analysis (see below). The four *msg* genes are linked in a head-to-tail fashion. Partial sequences from *msg*I and *msg*III were reported previously as h*msg*12 and h*msg*4 (41).

Nucleotide sequence analysis. Both strands of *msg*I were sequenced, and these sequences were compared with partial sequences derived from the other three genes in TG/61. The *msg*I gene contains an open reading frame spanning 3,093 bp (Fig. 5A), suggesting that it is a functional gene. A predicted amino acid sequence for MsgI, starting at a GUG codon, is shown. Several features of this GUG codon suggest that it may be used to initiate translation. (i) All four putative *msg* genes

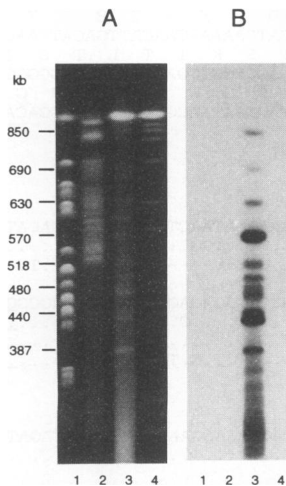

FIG. 3. Variants of *msg*I are represented on most chromosomes of human *P. carinii*. (A) Ethidium bromide-stained pulsed-field gel electrophoretically separated chromosomes from *P. carinii*. (B) Autoradiograph of hybridization to labeled recombinant *Eco*RI fragment of TG6106 (see Fig. 2). Lanes: 1, rat *P. carinii*; 2, ferret *P. carinii*; 3, human *P. carinii*; 4, yeast chromosomal markers (Bio-Rad).

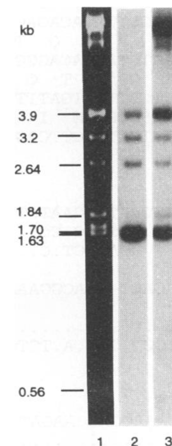

FIG. 4. TG/61 contains repetitive DNA which is related to a rat *P. carinii msg* gene. Clone TG/61 DNA was digested with *Xba*I, and the fragments were separated by agarose gel electrophoresis (lane 1). Lane 2, results after hybridization to the 1.63-kb *Xba*I fragment (Fig. 2) followed by stringent washes. Lane 3, results after hybridization to two *msg* genes from rat *P. carinii* (42a) followed by washes at reduced stringency (10).

A)

```
AAAAACTACCTTTAAAACTTTCAATATTCAACATCTTAATGAAGACAATATAAAATAGAAACAGCTTTAAAAGTCTGCCCACACACCCAACAAGTACGCT  100
TTCTCTCACATTAATGTCCTCTAGATTTTGGCGCGGGCGGTGGCGCGGGCGGTCAAGCGGCAGGTAGCAGGGGTAAAGAATAATGAAGCTGAAGAACGTC  200
                                    V  A  R  A  V  K  R  Q  V  A  G  V  K  N  N  E  A  E  E  R
TTTTTGCTTTGATTATGAGAGCAGATTATAAAGATGAAAGTAAATGTAAAAATAAAATAAAAGAATATTGCGATGGGTTAAAAAATGCTTCACTAACATC  300
L  F  A  L  I  M  R  A  D  Y  K  D  E  S  K  C  K  N  K  I  K  E  Y  C  D  G  L  K  N  A  S  L  T  S
AGAAGAAGTACATAAAGAATTAAAAGATTTTTGCAAAGATGGAAGTCAAGGAAAAAAATGCGAAGAACTAAAAAAAAATGTTGAAGCAAAATGCAATAAT  400
 E  E  V  H  K  E  L  K  D  F  C  K  D  G  S  Q  G  K  K  C  E  E  L  K  K  N  V  E  A  K  C  N  N
TTTAAAACAAAATTGGAAGGATTGGTGAAAAAAGATGCCTCAGGCTTGACAAATGATGATTGTAAAGAGAATGAACGACAATGCCTATTTTTGGAGGGAG  500
 F  K  T  K  L  E  G  L  V  K  K  D  A  S  G  L  T  N  D  D  C  K  E  N  E  R  Q  C  L  F  L  E  G
CATGCCCTGATCTTGTAGAGGATTGTAGCAAACTAAGGAATCTATGTTACCAGAAAAAACGTGAAGGAGTAGCAGAAGAAGTCCTTTTGAGGGCACTTCG  600
A  C  P  D  L  V  E  D  C  S  K  L  R  N  L  C  Y  Q  K  K  R  E  G  V  A  E  E  V  L  L  R  A  L  R
TGGGGATTTAGGAAATAAAACTGAATGCGAAAAGAAAATAAAGGATGTTTGCCCAAAAATAGGCCAAGAAAGTGATGAGTTGACATTGTTGTGTCTTGAT  700
 G  D  L  G  N  K  T  E  C  E  K  K  I  K  D  V  C  P  K  I  G  Q  E  S  D  E  L  T  L  L  C  L  D
CAGAAGAAAACATGCACGAATCTTATGACTGCGCGAGACAAAAAGTGTAATACACTTGAGGAAGATGTTAAAAAAGCACTTGAAAACAAAAACAACTTGT  800
 Q  K  K  T  C  T  N  L  M  T  A  R  D  K  K  C  N  T  L  E  E  D  V  K  K  A  L  E  N  K  N  N  L
TAGGAAAATGTTTGCCATTACTTGAGCATGCTACTTTCACAGAGGGGACTGCAAAAAAAGCATCACAGTGCACTCCAAATAAAGACTGCGAGGACTATCT  900
L  G  K  C  L  P  L  L  E  H  A  T  F  T  E  G  T  A  K  K  A  S  Q  C  T  P  N  K  D  C  E  D  Y  L
ACCAAAGTGTGATGAATTGGCAGAAGAATGTGGAAAAAAGGGCATCATTTACATACATCCGGGACCCGATTTTGATCCAACTAAGCCAGAACCTACAGTA 1000
 P  K  C  D  E  L  A  E  E  C  G  K  K  G  I  I  Y  I  H  P  G  P  D  F  D  P  T  K  P  E  P  T  V
GCAGAGGACATAGGGCTGGAAGAGCTTTATAAAAAGGCCGCAGAAGATGGTGTTCATATCGGAAAGCCTCCTGTAAGAGACGCAACTGCTCTACTGGCAC 1100
 A  E  D  I  G  L  E  E  L  Y  K  K  A  A  E  D  G  V  H  I  G  K  P  P  V  R  D  A  T  A  L  L  A
TTTTGATTCAGAATCCAGATCCTAAAATTCAAGCGAACGAAAAAGAAAAATGCAAAAAAGTTCTTGAAAATAAGTGTAAAGAGTTAAAAAAACATGAGGT 1200
L  L  I  Q  N  P  D  P  K  I  Q  A  N  E  K  E  K  C  K  K  V  L  E  N  K  C  K  E  L  K  K  H  E  V
TTTGGGAGATCTTTGTAATCAAAATGCAGCTAGTCAAAGTGGAACTAAAAAGTGTGAAGAACTAGAGAAGGAATTAGCAAACAGTACTAAAATTCTTTCT 1300
 L  G  D  L  C  N  Q  N  A  A  S  Q  S  G  T  K  K  C  E  E  L  E  K  E  L  A  N  S  T  K  I  L  S
GAAAAAATAAAGAATAAACATCTCTCTGGATCCGGAGAAACCATTCCATGGTATAAGTTGTCGACATTTCTTAGTGACAGTGACTGTGCAAGGTTAGAAT 1400
 E  K  I  K  N  K  H  L  S  G  S  G  E  T  I  P  W  Y  K  L  S  T  F  L  S  D  S  D  C  A  R  L  E
CAGACTGTTTTTATTTTGCTCAAGATAAAGATCCTCTTAAAAAAGAATGCAAGAATGTTAAGGCAGCATGTTATAAGAGAGGGCTTGATGCACGGGCAAA 1500
S  D  C  F  Y  F  A  Q  D  K  D  P  L  K  K  E  C  K  N  V  K  A  A  C  Y  K  R  G  L  D  A  R  A  N
TAAAGTGCTGCAAGAAAATATGCGTGGGTTATTACGTGGTTCAAATCAAAGTTGGCTTAAGAAATTTCAACAAGAATTAGTAAAAGTATGTGAGAAACTG 1600
 K  V  L  Q  E  N  M  R  G  L  L  R  G  S  N  Q  S  W  L  K  K  F  Q  Q  E  L  V  K  V  C  E  K  L
AAAGAAGAAAATAAAGGAAGTTTTTCAAACGATGAATTATTTGTTCTGTGTGTACAGCCAGCAAAAGCAGCCCGGTTGCTTACACATGATCTTCGAATGA 1700
 K  E  E  N  K  G  S  F  S  N  D  E  L  F  V  L  C  V  Q  P  A  K  A  A  R  L  L  T  H  D  L  R  M
CAACTATCTTTTTACGACAACAGTTGGATCAAAAGCGAGATTTTCCGACAGTAAAGACTGCCAAGGGAATTAGGGAGAAATGCCAAGATTTAGGGAAAGG 1800
T  T  I  F  L  R  Q  Q  L  D  Q  K  R  D  F  P  T  V  K  T  A  K  G  I  R  E  K  C  Q  D  L  G  K  G
ATTCCAAAAAGAAATTACGTGGCCATGTCATACACTGGAGCAGCAATGCAATCGCTTGGGGACTACAGAAATTTTGAAGCAGGTTTTACTAAATGAACAC 1900
 F  Q  K  E  I  T  W  P  C  H  T  L  E  Q  Q  C  N  R  L  G  T  T  E  I  L  K  Q  V  L  L  N  E  H
AAGGATACTTTGAAAACTCATGAAAACTGTGTAACATATTTAAAGGAAAAGTGTAATAAATGGTCTAGAAGGGGCGATGACCGTTTCTCTTTTGTATGTG 2000
 K  D  T  L  K  T  H  E  N  C  V  T  Y  L  K  E  K  C  N  K  W  S  R  R  G  D  D  R  F  S  F  V  C
TTTTCCAAAACGCTACGTGTAAGCTGATGGTAAAAGACGTACAAGATAGGTGCAAAATATTCAAAGAAAATATAAAAGTTTCAGAAATTGTTGATTTTCT 2100
V  F  Q  N  A  T  C  K  L  M  V  K  D  V  Q  D  R  C  K  I  F  K  E  N  I  K  V  S  E  I  V  D  F  L
TAAAAATAATACGAATAACATAACAACACTGGAAAGAAACTGTCCCTCCTGGCATACGTACTGCAATAGATTTTCATCTAATTGTCCAGATTTTTCAAAA 2200
 K  N  N  T  N  N  I  T  T  L  E  R  N  C  P  S  W  H  T  Y  C  N  R  F  S  S  N  C  P  D  F  S  K
AAGAATCCTTGTACAAAAATTAAAAACAATTGTAAGCCATTTTATGAAAGAAAAGCCTTGGAAGACGCTCTCAAAGTAGAGCTTCGAGGAAAGCTGAGTG 2300
 K  N  P  C  T  K  I  K  N  N  C  K  P  F  Y  E  R  K  A  L  E  D  A  L  K  V  E  L  R  G  K  L  S
ATGAAAACAAATGTACTGCAGCATTGAAGGGATATTGTACACTAGCGGGAAATGTAAATAACGCATCAGTCAGAAGTTTATGCAAAGATAACACTCAAGG 2400
D  E  N  K  C  T  A  A  L  K  G  Y  C  T  L  A  G  N  V  N  N  A  S  V  R  S  L  C  K  D  N  T  Q  G
TAGCAATAAAAAGACTGACGAAAAAGTGGTAGAAGAGCTTTGTAAGAAATTGATGGAAGAAGTAAAAGAGCAATGCGAGACATTACCAGCAGAATTAAAA 2500
 S  N  K  K  T  D  E  K  V  V  E  E  L  C  K  K  L  M  E  E  V  K  E  Q  C  E  T  L  P  A  E  L  K
CAACCGGCAGACGATCTAGAAAAAGATGTTAAGACATATGAGGAACTTAAGGAAGAGGCAAAGAAAGCAATGAACAAGTCCAGCCTTGTTTTATCATTCG 2600
 Q  P  A  D  D  L  E  K  D  V  K  T  Y  E  E  L  K  E  E  A  K  K  A  M  N  K  S  S  L  V  L  S  F
TTAAGAAAGACGGAAATAATACACCGAAAAATAATAGCAAAAGCGAAGATAAGAATGTCGTTTCAAATGAAAAAGATACCATAAAACATGTGAAAATACT 2700
V  K  K  D  G  N  N  T  P  K  N  N  S  K  S  E  D  K  N  V  V  S  N  E  K  D  T  I  K  H  V  K  I  L
ACGGAGAGGAGTTAAGGATGTATTAGTAACAGAATTAGAAGCCAAGGCGTTTGACTTAGCAGCAGAAGTGTTTGGCAGATATGTAGATTTAAAGGAAAGA 2800
 R  R  G  V  K  D  V  L  V  T  E  L  E  A  K  A  F  D  L  A  A  E  V  F  G  R  Y  V  D  L  K  E  R
TGCGAGAAATTGACCTTGGATTGTGGGATTAAAGACGATTGCGATGGTTTAAAAGGTGTGTGTGGAAAGATTAAGAAGAAATGTCGCGATCTGAAGCCTC 2900
 C  E  K  L  T  L  D  C  G  I  K  D  D  C  D  G  L  K  G  V  C  G  K  I  K  K  K  C  R  D  L  K  P
TGGAGGTGAAGTCGCACGAAATAGTCACAGAAAGCACAACGACGACCACAACGACAACAACGACCGTTACCGATCCGAAGGCAACAGAATGCAAATCCTT 3000
L  E  V  K  S  H  E  I  V  T  E  S  T  T  T  T  T  T  T  T  V  T  D  P  K  A  T  E  C  K  S  L
ACAGACAACAGACACATGGGTTACACAGACATCGACACACACAAGCACATCTACCATCACATCTACGATTACATCAAAAATAACATTGACATCAACAAGG 3100
 Q  T  T  D  T  W  V  T  Q  T  S  T  H  T  S  T  S  T  I  T  S  T  I  T  S  K  I  T  L  T  S  T  R
CGGTGCAAACCAACCAAGTGTACGACAGGGGATGATGCAGAAGACGTGAAGCCAAGTGAAGGCTTGAGGGTGAGCGGGTGGAATGTGATGAGGGGGGCAA 3200
 R  C  K  P  T  K  C  T  T  G  D  D  A  E  D  V  K  P  S  E  G  L  R  V  S  G  W  N  V  M  R  G  A
TAGTAGCAATGGTTATTTCGTTCATGATTTAGAAAGCAAATGAGCAAGATTTAATGATGAGCATGTATATTTAGTAATAACGAGCAGGGTAGGACAGGAA 3300
I  V  A  M  V  I  S  F  M  I -*-
AGAGTAAAGTAGATTGAGTCATCTCTTCAGCATCAATTGAAGTGCATTACA
```

B)

```
msgII    ACTGAATTTCCATAAGAATCTTGGAATTAAC---AAAGAATTCAATAGTTAATTTCAGTATGAAAACATAACTCTAACATCTCACATGTAG
msgIV    -..A.G..A.TTG..C.CC...AA......TTAG...T....T.....C.....A.-------TG..T...GG.................C.A
msgI                  .A..CT.C...--------T---...ACT.......T.C..CA..TTA.....G...---------------.T..AA.A

msgII    TCAAAACACACT--AACGGAAACTCAAACTCAAAACACACACCACAGCGCAG--CCACCCCCGTCAGGCACAGCTTGGCGCGGGCGGTG
msgIII                                                          ..T.....G.........-.......
msgIV    C.G.......T.--.................-.GC................AAC.....G.............................
msgI     .AG.....GCT.TA..A.TCTG.C..C..A.CC....AGT..GCTTT.T.TC--A..TTAAT...CTCT.G.TT...............
```

C)

```
msgI     TAGAAAGCAAATGAGCAAGATTTAATGATGAGCATGTATATTTAGTAAT--AACGAGCAGGGTAGGACAGGAAAGAGTAAAGTAGATTGAG
msgII    ...........................................ATA.......................................

msgI     TCATCTCTTCAGCATCAATTGAA-GTGCATTAC
msgII    .......................A.........
```

FIG. 5. Nucleotide sequence of *msg*I and comparison of flanking regions with the other *msg* sequences on TG/61. Sequence determination of *msg*I was from both strands. All other sequences were determined from a single strand. Sequence identity is indicated by dots; dashes indicate gaps. (A) Nucleotide sequence of *msg*I and its predicted amino acid sequence. (B) Nucleotide sequence alignment of the upstream regions of the four *msg* genes on TG/61. Putative GTG start codons are underlined. (C) Nucleotide sequence alignment of the downstream regions of *msg*I and *msg*II. TAG stop codons are underlined.

```
MsgI     VARAVKRQ-VAGVKNNEA------EE--RLFALIMRADYKDESKCKNKIKEYCDGLKNASLTSEEVHKELKDFCKDGSQGKKCEELKKNVEAKCNNFKTK   91
MsgII    .......G-AQAA.GTKGTSVYDD..--Y.L...-LK.AM.Q....E.LEK..KE..DSN.NK..I.EK.....GV.KADT..QD..VK.TK..TD...-   95
MsgIII   ..A......--.Q.TGTQD------..EV..L...VGK..--NV...ENLEK..KE....T.EAK..D.K.ES...--NEKE...K..DKIQN...TL.EE   88
MsgIV    .........-.T.ASGVDE------..V-..L...LKE.S..DK..EE.LEKH.KE.SE.N..P.Q..EK.....DSKKRD...K.......K..GD...E   92
rat Msg  M..P....A.Q.AQDEID------.K--H.L.F.VKDK..E.Q...EELEK..KE..E.DKNL.N.DDKV.GL.D.KKRDE..KDV..K..DELKD.---   89

MsgI     LEGLVKKDASGLTN---DDCKENERQCLFLEGACPD-LVED-CSKLRNLCYQKKREGVAEEVLLRALRGDLGNKTECEKKIKDVCPKIGQESDELTLLCL  186
MsgII    ------..LEPT...PSD.N.....QE.........NE.K.K-......K......DK....A.....SN...SD.T.K.G.EN..LALAE..N..MQK.F  189
MsgIII   .KKILP.GI.A.KD---K.....................-.....-.N........R..D.............SIED.NK.KEEL.KI..VL.K......N...  183
MsgIV    ..E....E..N.K.---.E.TK..Q...--------------------------------------------------------------------  118
rat Msg  -.EELQ.VLNNIKD---EN.EKY.EK.IL..ETDY.-VIK.N.IE..EG..KL...K....L.....G..AKEEAK.KG.MNT...VLSR.....MSF..  184

MsgI     DQKKTCTNLMTARDKKCNTLEEDVKKALENKNNLLGKCLPLLEHATFTEGTAKKASQCTPNKDCEDYLP-KCDELAEECGKKGIIYIHPGPDFDPTKPEP  285
MsgII    .T.SL..S.----------------------------------------------------------------------------------------  198
MsgIII   N.EE..KD.-V.EG......KT..QG..SG..K.QE...S...QCY.YI.------------N..NNDMI..I..G.K.QEQN.V..P..........A  270
rat Msg  .SA...GD.----K..LG.VC.PL..E.-KD.E.AE..HER..KCH.YGEACDD.----------------..KKFE.Q.KG.N...KA.ES.LS.V..RA  264

MsgI     TVAEDIGLEELYKKAAEDGVHIGKPPVRDATAL--------LALLIQNPDPKIQANEKEKCKKVLENKCKELKKHEVLGDLCNQNAASQSGTKKCEELEK  377
MsgIII   .L............-..KMSFMSTCERCNSFVG--------I..--------KKKH..QE.I.T.QKS.ENPHE..A.EN..KA.TL.DY.KE....FKN  353
rat Msg  SLLRS...DDV..N.EKH.II...SG.DLPRKSGTNFCKIS.L..SRDE.K.---EPDK..T.A..KCDASKYLNTE.EK..-------KDGN.N.KCK.  354

MsgI     ELANSTKILSEKIKNKHLSGSGETIP-------WYKLSTFLSDSDCARLESDCFYF--AQDKDPLKKECKNVKAACYKRGLDARANKVLQENMRGLLRGS  468
MsgIII   DVNKTC...TS.VIDNR.FDPVNSKVVG-----.G..P....KEE.T....Y....--KESCPDAE.A.M..............................  446
rat MSG  I.DVKERCTNL.L.LYLKGL.T.YDDQESDPLS.GQ.P..FIKGE..E...E...LEK.CKDNNID.A.Q.AR.....K.Q.RML..FF.KELK.K.---  451

MsgI     NQSWLKKFQQELVKVCEKLKEENKGSFSNDELFVLCVQPAKAARLLTHDLRMTTIFLRQQLDQKRDFPTVKTAKGIREKCQDLGKGFQKEITWPCHTLEQ  568
rat Msg  ----GHVRFYSDP.D.K.YVV..CTKLDKKY.-PR.LY.KELCYG.SN.IFLQSKE.SAL..DQ....LK.DCVELK...-.ELSSDSLLNLEK.I..KR  545

MsgI     QCNRLGTTEILKQVLLNEHKDTLKTHENCVTYLKEKCNKWSRRGDDRFSFVCVFQNATCKLMVKDVQDRCKIFKENIKVSEIVDFL----KNNTNNITTL  664
rat Msg  R.EYFRVS.GFRN.F.EKKD.S.M.QD..TKA.H...HQLY..RKNS..VS.ALPEE..SY..FHTSQD.SSL.V...NEK.LEKIGEEI.KANK.EALV  645

MsgI     ERNCPSWHTYCNRFSSNCPDFSKKNP--------CTKIKNNCKPFYERKALEDALKVELRGKLSDENKCTAALKGYCT-LAGNVNNASVRSLCKDNTQGS  755
rat Msg  .EL.TT.GRH.HQLME....DL..KENGNGNDHN.EALQEK.NKTF.KLK..EE.SHL.K.S.K.-D..KE..GKR..E.EK.EAFKTLYGK.D...---  741

MsgI     NKKTDEKVVEELCKKLMEEVKEQCETLPAELKQPADDLEKDVKTYEELKEEAKKAMNKSSLVLSFVKKDGNNTPKNNSKSEDKNVVSNEKDTIKHVKILR  855
rat Msg  --------K.NV....VDK..KR.P..KD..ENAKKE.T.MKNE.DD..KA.E.STEAAK.L..RPRQTVMPNAQ.G.D.TLVPPPPQAPAGPPPPGSPP  833

MsgI     ----------------------------------------------RGVKDVLVTELEAKAFDLAAEVFGRYVDLKERCEKLTLDCGIKDDCDGLKGV  907
rat Msg  PPPSQNGTPGTPGGETGASGGTPGTPGTPGTPGTPGGMMKYAKLGLVK.TYV.GG.S.V.V....ATTIALEL.LE...E.KA.E....F.E..PDT.QA  933

MsgI     CGKIKKKCRDLKPLEVKSH--EIVTESTTTTTTTTTTVT--DPKATE---------------------CKSLQTTDTWVTQTSTHTSTSTITSTITSKIT  982
MsgII    -------------------------------------------------------------------------------------V..R..
MsgIII   .VADE...H........P.--.T..........------------------------------------------------------------
rat Msg  .EN.DTL.-K.E...I.P.HT.KI..TK.E.K.E.K.E.KT.G..D.KTVEKTVTETKSVGGGKVTEE.TMI........S..L....T.S...V..TV. 1032

MsgI     LTSTRRCKPTKCTTG--D--------DAEDVKPSEGLRVSGWNVMRGAIVAMVISFMI                                           1030
MsgII    ...............EG.--------.......................VLL........
rat Msg  ...M.K.........D--SSKETQKEE.D.E...N..MKIRVPDMIKIMLLGVIVMG.M                                          1088
```

FIG. 6. Alignment of deduced amino acid sequence data from *msg*I to *msg*IV and rat *msg*. Sequence identity is indicated by dots; dashes indicate gaps. The rat Msg sequence is from reference 21. Cysteine residues of MsgI (C) and potential glycosylation sites (NXS/T) are in boldface.

in TG/61 contain GUG at this position (Fig. 5B). (ii) The sequence preceding these GUG codons is conserved in all four *msg* genes (Fig. 5B) and resembles sequences preceding the presumptive AUG start codons of rat *P. carinii msg* genes (21). (iii) The predicted amino acid sequences immediately following the GUG codons show high identity to each other and to a rat *P. carinii* Msg sequence (Fig. 6). (iv) Both of the two AUG codons upstream of the GUG of *msg*I are out of frame. An in-frame AUG is located 75 bp further downstream of the GUG, but this AUG is not present in the other three *msg* genes in TG/61 (Fig. 6).

The *msg*I gene contains 62 mol% adenosine and thymidine, with 69 mol% A or T in the third codon position, which conforms to the A/T bias seen in all known rat *P. carinii* gene sequences (8, 21). Sequence divergence among the first 350 nucleotides of *msg*I and the remaining three genes is between 25 and 33%. The deduced amino acid sequences of the first 350 nucleotides of the four *msg* genes in TG/61 are approximately twice as divergent (42 to 58%) as the gene sequences (Fig. 6). This suggests that accumulation of missense mutations in *msg* genes has been beneficial to the species and that positive selection of missense mutations in duplicated *msg* genes allowed the acquisition of a large repertoire of different Msgs. Since Msgs are surface antigens, it is possible that one benefit of acquiring a large number of different Msgs has been to increase the capacity of the organism to evade the immune system, as is known to happen in other pathogenic microbes that possess a family of genes encoding different forms of surface proteins (3).

The *msg*I gene and a rat Msg cDNA (21) have 56% nucleotide sequence identity. The complete predicted amino acid sequence of MsgI is 58 amino acids shorter than the predicted sequence of the one complete rat *P. carinii* Msg described (21). Much of this size difference seems due to the absence in MsgI of a proline/glycine-rich module, which starts at residue 817 in the rat *P. carinii* Msg shown in Fig. 6. The generality of this difference between rat and human *P. carinii* is not yet known. The computer-generated optimal alignment in

Fig. 6 matched MsgI to the rat Msg sequence at 35% of residues. The identity increases to 42% if the proline/glycine-rich region is not counted. As in other Msg molecules, cysteine residues (shown in boldface in Fig. 6) of the four human Msg molecules are numerous (5.4% in MsgI) and, as far as sequence data are available, are completely conserved. The degree of cysteine conservation is also high when human and rat Msg molecules are compared. Of the 56 cysteine residues in MsgI, only four are not in rat Msg (replaced by either leucine, aspartic acid, or asparagine) (Fig. 6). Cysteine residues are also conserved between human Msg and the two gpA cDNA fragments from ferret *P. carinii* (15) (alignment not shown).

Human MsgI has 10 potential N-glycosylation sites (NXS/T), which are printed in boldface in Fig. 6. The rat *P. carinii* Msg in Fig. 6 contains five NXS/T sequences, only one of which lies in the same region as a site in MsgI (residue 831). Comparison of the 5′ ends of human *P. carinii* Msgs shows that two of the first three sites in MsgI are also present in at least one other Msg from human *P. carinii*. Potential sites for O-linked glycosylation include the threonine-rich module near the carboxyl terminus (residues 930 to 986 of Msg I), which is conserved in two human *P. carinii* Msgs as well as in several Msgs from rat *P. carinii* (21, 45).

Regions upstream of genes *msg*II and *msg*IV are highly conserved, but the region upstream of *msg*I is different (Fig. 5B). The regions downstream of the stop codons from genes *msg*I and *msg*II are nearly identical (Fig. 5C). Conservation of sequences upstream and downstream of *msg* open reading frames seems to be a common feature of *msg* genes in both rat and human *P. carinii* (21, 42a, 45). The degree of conservation of untranslated sequences can exceed that of translated sequences.

Serologic characterization of MSG fusion proteins. The antigen-expressing λgt11 subclones of TG/61 (Table 1) were isolated by using serum from a rabbit that had been immunized with human *P. carinii*. Of the 17 λgt11 subclones, 8 were from *msg*I, 5 were from *msg*III, and 4 were from *msg*II (Table 1 and Fig. 2). All subclones contained gene segments starting within the first 840 nucleotides downstream of the putative initiation codon of gene *msg*I.

Recombinant proteins produced by the λgt11 clones are usually fused to β-galactosidase, which adds 116 kDa to the molecular mass of the antigenic moiety (17). From the predicted open reading frames of the 5′ termini of the *msg* genes and the lengths of the inserts, one can predict the size of the fusion protein. Figure 7A shows an analysis of fusion proteins from λgt11 subclones by Western blotting with hyperimmune rabbit serum. The rabbit antiserum did not stain β-galactosidase (Fig. 7A, lane 1), but it stained multiple bands in each of the lanes containing Msg fusion proteins (Fig. 7A, lanes 2 to 14). In most cases, the major bands had much smaller molecular weights than expected from the sizes of the respective recombinant inserts. The presence of minor bands with higher molecular weights in the lanes containing the fusion proteins TG6105, -06, -15, and -19 (Fig. 7A, lanes 4, 5, 10, and 12, respectively) suggests that the major bands with smaller-than-expected sizes were not produced by translational termination due to the presence of stop codons within the DNA segments inserted in λgt11 (Fig. 7A and Table 1). The complex band patterns may have been produced by partial proteolysis. Two clones, TG6108 and TG6117, did not stain on immunoblots with hyperimmune rabbit serum (data not shown). Interestingly, both of these clones contained inserts that start upstream of the putative initiation codon of *msg*I (Table 1), suggesting that the two antigenic proteins were expressed as free proteins, highly susceptible to proteolysis.

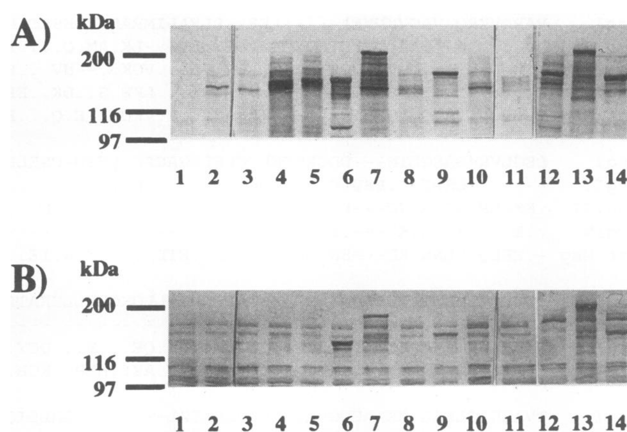

FIG. 7. Recognition of Msg fusion proteins by hyperimmune rabbit serum and serum from a healthy human subject. Polypeptides of induced *E. coli* Y1089 bacteria lysogenized with the λgt11 subclones of TG/61 (Table 1) were separated by SDS–8% PAGE, blotted, and immunostained with hyperimmune rabbit serum against human *P. carinii* at a dilution of 1/1,000 (A) or human serum from a healthy subject at a dilution of 1/50 (B). The rabbit serum was adsorbed to an *E. coli* Y1090 lysate; the human serum was adsorbed to a sonicate of induced *E. coli* Y1089(λgt11) lysogen as described by Sambrook et al. (37). Lanes: 1, λgt11 control; 2, TG6101; 3, TG6104; 4, TG6105; 5, TG6106; 6, TG6107; 7, TG6109; 8, TG6111; 9, TG6113; 10, TG6115; 11, TG6118; 12, TG6119; 13, TG6120; 14, TG6121.

Approximately 30% of all human sera recognize polypeptides which are believed to belong to the Msg complex of human *P. carinii* (30, 40). To determine if fusion proteins made from the λgt11 subclones were recognized by human sera, we analyzed these proteins by Western blotting with sera positive for Msg from healthy individuals and from patients with AIDS-related *P. carinii* pneumonia (39). Figure 7B shows the result obtained by staining fusion proteins with a serum from a healthy donor. The serum did not react with β-galactosidase produced by the λgt11 vector control (lane 1), but it recognized bands in lanes 6, 7, 9, 10, 12, 13, and 14, which contained fusion proteins from λgt11 subclones TG6107 (*msg*I), TG6109 (*msg*III), TG6113 (*msg*III), TG6115 (*msg*III), TG6119 (*msg*II), TG6120 (*msg*III), and TG6121 (*msg*I), respectively. The lack of reactivity with TG6105 and TG6106 (Fig. 7B, lanes 4 and 5, respectively) was unexpected, because TG6105 and TG6106 each contain the sequence present in TG6107 and TG6121, which were well stained (Fig. 7B, lanes 6 and 14, respectively). Perhaps the epitopes on TG6107 and -21 that were recognized by the human serum were located in a region of the fusion protein that was lost from fusion proteins TG6105 and -06 because of proteolysis. Since the Western blot analysis with both sera (Fig. 7 A and B, respectively) yielded identical results in repeated and simultaneous experiments, the presence of an artifact that might have arisen from continuous protein degradation during storage can be ruled out. Thus, the difference between the data in lanes 4 and 5 of Fig. 7A and B indicates that the hyperimmune rabbit serum recognized some epitopes that the human serum failed to recognize.

Western immunoblots with nine other human sera showed that only one serum failed to recognize any of the fusion proteins. The other eight sera all recognized MsgIII fusion protein TG6120 (data not shown). Sera from patients with AIDS-related *P. carinii* pneumonia also recognized MsgIII fusion proteins. Figure 8 shows results obtained with sera (designated A through F) from six different patients. All six

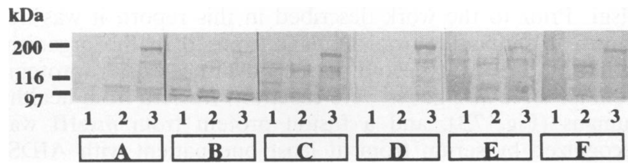

FIG. 8. Sera from AIDS-related *P. carinii* pneumonia patients recognize Msg fusion proteins. Proteins of induced lysogens of *msg*I clone TG6107 (lanes 2) and *msg*III clone TG6120 (lanes 3) were analyzed by Western blotting with sera from six patients with AIDS-related pneumocystis pneumonia (sera A to F). Lanes 1, protein of an induced Y1089(λgt11) lysogen. Patient sera were adsorbed to a sonicate of induced Y1089(λgt11) lysogen as described by Sambrook et al. (37) and used at a dilution of 1/50.

sera reacted with MsgIII fusion protein TG6120 (lanes 3), and all but D reacted with MsgI fusion protein TG6107 (lanes 2). However, sera E and F reacted strongly with β-galactosidase (lanes 1), and sera A, B, and C also weakly stained β-galactosidase, which could account for the binding to the Msg fusion proteins. The result with serum A was suggestive of reactivity to MsgIII fusion protein, and that with serum C was suggestive of reactivity to both Msg fusion proteins, but both results were not as clear as with serum D because of faint recognition of β-galactosidase (lanes 1).

Epitopes encoded by MsgI and MsgIII are present on proteins expressed in human *P. carinii*. To determine if antibodies with specificity to MsgI would recognize a polypeptide from *P. carinii* protein preparations, we first used affinity-purified antibodies for a Western blot analysis of *P. carinii* proteins. The affinity-purified antibodies with specificity to TG6107 were cross-reactive with fusion proteins originating from *msg*II and *msg*III (Fig. 9A, lanes 2 and 3) and reacted with two bands from human *P. carinii* with apparent molecular masses of 120 and 114 kDa (lanes 1 and 4). Both bands had sizes near the predicted size of MsgI (116 kDa). Several things could have caused the apparent difference between the predicted size of MsgI and the observed bands. First, the gel electrophoresis conditions may have caused aberrant migration of the proteins. Second, although the reading frame was open between the presumptive start and stop codons, we cannot rule out the presence of small introns in the *msg*I gene. Removal of such introns would reduce the size of the predicted protein. Posttranslational modifications such as proteolytic cleavage or glycosylation also could have occurred. Alternatively, it is possible that neither of the bands recognized by the affinity-purified antiserum contained MsgI because the multiple copies of variant *msg* genes endow *P. carinii* populations with the potential to express many different Msg polypeptides that would be expected to be antigenically related. While it is not presently possible to determine which Msg molecules are actually expressed, these data show that human *P. carinii* contains proteins that are antigenically related to MsgI.

The crude hyperimmune rabbit serum stained both the 120- and 114-kDa bands equally (Fig. 9A, lane 5), but the affinity-purified antiserum stained the 120-kDa band more prominently, suggesting that affinity purification of the serum enriched for immunoglobulins specific to one or more polypeptides in the 120-kDa band. The relationship between the 120- and 114-kDa bands was explored by Western blot analysis by using two MAbs specific to human *P. carinii* Msg (39a). Figure 9A shows that MAb 2A1 recognized MsgI fusion proteins TG6107 and TG6121 (lanes 8 and 9) and both the 120- and 114-kDa bands (lane 6). However, in contrast to the immunoaffinity-purified antiserum, MAb 2A1 stained the 114-kDa band more prominently, suggesting that the antigen composition of the 114-kDa band was different from the composition of the 120-kDa band. Experiments with a second MAb also indicated that epitopes encoded by the cloned *msg* genes are present on multiple proteins of various sizes. MAb 4-4E7, which recognized MsgIII fusion proteins TG6109, -13, and -20 (Fig. 9B, lanes 11 to 13), faintly stained a 114-kDa band and a smaller band in the lane containing human *P. carinii* polypeptides (Fig. 9B, lane 6) and more prominently stained a doublet band, the smaller of which comigrated with the 120-kDa band of human *P. carinii* (Fig. 9B, lanes 5 and 6). Taken together, the results obtained with immunoaffinity-purified antiserum and MAbs suggest that human *P. carinii* contains at least three size classes of proteins that contain epitopes encoded by at least two of the cloned *msg* genes. We noticed that in our hands the MAbs recognized *P. carinii* polypeptide bands with apparent molecular weights that

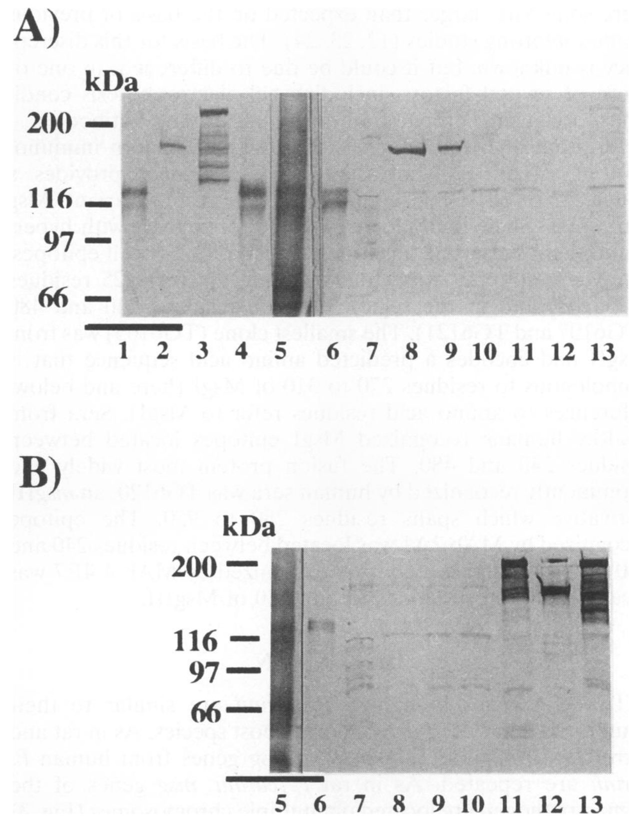

FIG. 9. Immunoglobulins specific to MsgI and MAbs recognize polypeptides from human *P. carinii*. Protein samples were separated by SDS–6% PAGE, blotted, and immunostained with affinity-purified hyperimmune rabbit serum against human *P. carinii* (lanes 1 to 4) or crude hyperimmune rabbit serum against human *P. carinii* (panels A and B, lanes 5), MAb 2A1 (panel A, lanes 6 to 13), and MAb 4-4E7 (panel B, lanes 6 to 13). Affinity purification was performed on membrane-bound plaques of TG6107 by using 30 ml of hyperimmune rabbit serum at a dilution of 1/200 as described previously (2). That the affinity purification was effective can be seen by comparing lanes 4 and 5 (A). Crude rabbit serum was used at a 1/1,000 dilution, anti-rabbit–horseradish peroxidase conjugate was used at a 1/3,000 dilution, and both MAbs and the anti-mouse–horseradish peroxidase conjugate were used at a 1/1,000 dilution. Protein samples other than total human *P. carinii* were fusion proteins. Lanes: 1, human *P. carinii*, 0.6 μg; 2, TG6119; 3, TG6120; 4 to 6, human *P. carinii*, 1.0 μg; 7, λgt11; 8, TG6107; 9, TG6121; 10, TG6119; 11, TG6109; 12, TG6113; 13, TG6120.

were somewhat larger than expected on the basis of previous immunoblotting studies (12, 23, 24). The basis for this discrepancy is unknown, but it could be due to differences in one or more of several factors, including gel electrophoresis conditions, organisms, preparation of organisms, and antibodies.

Location of B-cell epitopes. Analysis by Western immunoblotting of antigen-expressing λgt11 subclones provides a rough map of B-cell epitopes on the three variant Msg molecules. Since their cloning relied on reactivity with hyperimmune rabbit serum, all clones carried rabbit B-cell epitopes. On MsgI, epitopes were found among the first 225 residues (TG6108) and in the region between residues 240 and 480 (TG6107 and TG6121). The smallest clone (TG6103) was from MsgII and encodes a predicted amino acid sequence that is homologous to residues 270 to 310 of MsgI (here and below, references to amino acid residues refer to MsgI). Sera from healthy humans recognized MsgI epitopes located between residues 240 and 480. The fusion protein most widely and prominently recognized by human sera was TG6120, an *msg*III derivative which spans residues 280 to 950. The epitope recognized by MAb 2A1 was located between residues 240 and 480 on MsgI, and the epitope recognized by MAb 4-4E7 was located between residues 280 and 450 of MsgIII.

DISCUSSION

The *msg* genes in human *P. carinii* are similar to their counterparts in *P. carinii* from other host species. As in rat and ferret *P. carinii* (15, 21, 42a, 45), *msg* genes from human *P. carinii* are repeated. As in rat *P. carinii*, *msg* genes of the human pathogen are located on multiple chromosomes (Fig. 3) and can occur as tandem arrays (21, 42a, 45). Genes within the TG/61 array are all about the same size but are heterogeneous in sequence, as is the case for the Rp3-1 *msg* gene array from rat *P. carinii* (42a). The repertoire of multiple related *msg* genes may be differentially expressed and may be involved in a strategy to evade the host's immune response by antigenic variation (3). In this regard, it is interesting that serum D from a patient with AIDS-related *P. carinii* pneumonia recognized an MsgIII fusion protein but did not recognize a fusion protein derived from *msg*I (Fig. 8D, lanes 2 and 3). Further experiments will be necessary to determine if such differences reflect differential Msg expression or recognition during infection.

Compared with a rat Msg molecule, human MsgI lacks a module composed of 48 amino acid residues (see residue 854 of MsgI in Fig. 6). This module varies in length among rat Msg molecules and is rich in proline and glycine residues (21), bearing similarity to collagen and collagen-related protozoal antigens (18, 48). It is possible that this collagen-related region is responsible for the reported collagenase sensitivity of some Msg molecules from rodents but not from humans (12).

A function that can be expected of a major surface protein of an extracellular pathogen is host cell adherence (32), and Msg has been implicated in this function by binding to fibronectin (33), surfactant protein A (51), and mannose-binding protein (7). An interesting feature in this regard is the threonine-rich module present in all Msg molecules characterized so far, which is a feature shared by mucin-related proteins in two pathogenic protozoa (29, 34) and by the highly glycosylated glue protein of *Drosophila* pupa (11). The threonine residues of Msg molecules may be carbohydrate associated (11), although O-linked glycosylation of Msg molecules has not yet been demonstrated. However, N-linked glycosylation has been demonstrated by biochemical studies on Msgs from different *P. carinii* (13, 22, 24). Consistent with these findings is the presence of 10 potential sites for N-linked glycosylation in MsgI. Prior to the work described in this report, it was not known if antibodies from humans exposed to the organism would bind to Msg devoid of carbohydrate. Fusion proteins from all three *msg* genes were reactive with sera from healthy humans (Fig. 7B), and a fusion protein from *msg*III was recognized by serum from at least one patient with AIDS-related *P. carinii* pneumonia (Fig. 8). Since there is no evidence for glycosylation of *E. coli* proteins (9), these data show that *P. carinii* polypeptide epitopes on Msg elicit antibody production during natural exposure to the organism.

Subunit vaccines (26) against *P. carinii*, which may be based on Msg, could be of benefit in the mitigation or prevention of *P. carinii* pneumonia. Likewise, Msg is an obvious target for immunotherapy of clinical *P. carinii* pneumonia (14, 36). Our data provide the beginning of an understanding in molecular terms of the structure and antigenicity of this complex group of proteins in human *P. carinii.*

ACKNOWLEDGMENTS

This work was supported by grant R01 AI 28471 in conjunction with a Public Health Service program project grant in mycology (P01 AI28932) from the National Institutes of Health.

We thank L. Frohlich, D. Maley, M. Ryan, and M. Kaselis for technical assistance, D. Bintzler for oligonucleotide synthesis and automated DNA sequencing, and S. L. Stringer, M. J. Linke, A. G. Smulian, and P. D. Walzer for discussions and critical evaluation of the manuscript.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment tool. J. Mol. Biol. **215:**403–410.
2. **Beall, J. A., and G. F. Mitchell.** 1986. Identification of a particular antigen from a parasite cDNA library using antibodies from selected portions of western blots. J. Immunol. Methods **86:**217–223.
3. **Borst, P.** 1991. Molecular genetics of antigenic variation. Immunol. Today **12:**A29–A33.
4. **Centers for Disease Control.** 1992. HIV/AIDS surveillance, year-end edition. Centers for Disease Control, Atlanta.
5. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
6. **Edlind, T. D., M. S. Bartlett, G. A. Weinberg, G. N. Prah, and J. W. Smith.** 1992. The β-tubulin gene from rat and human isolates of *Pneumocystis carinii*. Mol. Microbiol. **6:**3365–3373.
7. **Ezekowitz, R. A. B., D. J. Williams, H. Koziel, M. Y. K. Armstrong, A. Warner, F. F. Richards, and R. M. Rose.** 1991. Uptake of *Pneumocystis carinii* mediated by the macrophage mannose receptor. Nature (London) **351:**155–158.
8. **Fletcher, L. D., L. C. Berger, S. A. Peel, R. S. Baric, R. D. Tidwell, and C. C. Dykstra.** 1993. Isolation and identification of six *Pneumocystis carinii* genes utilizing codon bias. Gene **129:**167–174.
9. **Garbe, T., D. Harris, M. Vordermeier, R. Lathigra, J. Ivanyi, and D. Young.** 1993. Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation. Infect. Immun. **61:**260–267.
10. **Garbe, T., C. Jones, I. Charles, G. Dougan, and D. Young.** 1990. Cloning and characterization of the *aroA* gene from *Mycobacterium tuberculosis*. J. Bacteriol. **172:**6774–6782.
11. **Garfinkel, M. D., R. E. Pruitt, and E. M. Meyerowitz.** 1983. DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. **168:**765–789.
12. **Gigliotti, F.** 1992. Host species-specific antigenic variation of a mannosylated surface glycoprotein of *Pneumocystis carinii*. J. Infect. Dis. **165:**329–336.
13. **Gigliotti, F., L. R. Ballou, W. T. Hughes, and B. D. Mosley.** 1988. Purification and initial characterization of a ferret *Pneumocystis carinii* surface antigen. J. Infect. Dis. **158:**848–854.
14. **Gigliotti, F., and W. T. Hughes.** 1988. Passive immunoprophylaxis with specific monoclonal antibody confers partial protection

against *Pneumocystis carinii* pneumonitis in animal models. J. Clin. Invest. **81**:1666–1668.

15. **Haidaris, P. J., T. W. Wright, F. Gigliotti, and C. G. Haidaris.** 1992. Expression and characterization of a cDNA clone encoding an immunodominant surface glycoprotein of *Pneumocystis carinii*. J. Infect. Dis. **166**:1113–1123.
16. **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.
17. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Construction and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning, vol. I. IRL Press, Oxford.
18. **Iams, K., J. R. Young, V. Nene, J. Desai, P. Webster, O. K. ole-MoiYoi, and A. J. Musoke.** 1990. Characterization of the gene encoding a 104-kilodalton microneme-rhoptry protein of *Theileria parva*. Mol. Biochem. Parasitol. **39**:47–60.
19. **Kaiser, K., and N. E. Murray.** 1985. The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries, p. 1–48. *In* D. M. Glover (ed.), DNA cloning, vol. I. IRL Press, Oxford.
20. **Kovacs, J. A., J. L. Halpern, B. Lundgren, J. C. Swan, J. E. Parrillo, and H. Masur.** 1989. Monoclonal antibodies to *Pneumocystis carinii*: identification of specific antigens and characterization of antigenic differences between rat and human isolates. J. Infect. Dis. **159**:60–70.
21. **Kovacs, J. A., F. Powell, J. C. Edman, B. Lundgren, A. Martinez, B. Drew, and C. W. Angus.** 1993. Multiple genes encode the major surface glycoprotein of *Pneumocystis carinii*. J. Biol. Chem. **268**: 6034–6040.

21a. **Linke, M. J.** Personal communication.

22. **Linke, M. J., M. T. Cushion, and P. D. Walzer.** 1989. Properties of the major antigens of rat and human *Pneumocystis carinii*. Infect. Immun. **57**:1547–1555.
23. **Lundgren, B., M. Lebech, K. Lind, J. O. Nielsen, and J. D. Lundgren.** 1993. Antibody response to a major human *Pneumocystis carinii* surface antigen in patients without evidence of immunosuppression and in patients with suspected atypical pneumonia. Eur. J. Clin. Microbiol. Infect. Dis. **12**:105–109.
24. **Lundgren, B., G. Y. Lipschik, and J. A. Kovacs.** 1991. Purification and characterization of a major human *Pneumocystis carinii* surface antigen. J. Clin. Invest. **87**:163–170.
25. **Masur, H., H. C. Lane, J. A. Kovacs, C. J. Allegra, and J. C. Edman.** 1989. *Pneumocystis* pneumonia: from bench to clinic. Ann. Intern. Med. **111**:813–826.
26. **Mekalenos, J. J.** 1992. Bacterial mucosal vaccines, p. 43–50. *In* J. E. Cicardi, J. R. McGhee, and J. M. Keith (ed.), Genetically engineered vaccines. Plenum Press, New York.
27. **Meuwissen, J. H. E. T., I. Tauber, A. D. E. M. Leeuwenberg, P. J. A. Beckers, and M. Sieben.** 1977. Parasitologic and serologic observations of infection with *Pneumocystis* in humans. J. Infect. Dis. **136**:43–49.
28. **Murray, J. F., and J. Mills.** 1990. Pulmonary infectious complications of human immunodeficiency virus infection. Am. J. Respir. Dis. **141**:1582–1598.
29. **Murray, P. J., and T. W. Spithills.** 1991. Variants of *Leishmania* surface antigen derived from multigenic family. J. Biol. Chem. **266**:24477–24484.
30. **Peglow, S. L., A. G. Smulian, M. J. Linke, C. L. Pogue, S. Nurre, J. Crisler, J. Phair, J. W. M. Gold, D. Armstrong, and P. Walzer.** 1990. Serologic responses to *Pneumocystis carinii* antigens in health and disease. J. Infect. Dis. **161**:296–306.
31. **Pifer, L. L., W. T. Hughes, S. Stagnow, and D. Woods.** 1978. *Pneumocystis carinii* infection: evidence for high prevalence in normal and immunosuppressed children. Pediatrics **61**:35–41.
32. **Pottratz, S. T., and W. J. Martin II.** 1994. Mechanisms of *Pneumocystis carinii* attachment to lung cells, p. 237–250. *In* P. D. Walzer (ed.), *Pneumocystis carinii* pneumonia, 2nd ed. Marcel Dekker Inc., New York.
33. **Pottratz, S. J., J. Paulsrud, J. S. Smith, and W. J. Martin II.** 1991. *Pneumocystis carinii* attachment to cultured lung cells by Pneumocystis gp 120, a fibronectin binding protein. J. Clin. Invest. **88**:403–407.
34. **Reyes, M. B., G. D. Pollevick, and A. C. Frasch.** 1994. An unusually small gene encoding a putative mucin-like glycoprotein in *Trypanosoma cruzi*. Gene **140**:139–140.
35. **Rinehart, F. P., T. G. Rich, P. L. Deininger, and C. W. Schmidt.** 1981. Renaturation rate studies of a single family of interspersed repeated sequences in human deoxyribonucleic acid. Biochemistry **20**:3003–3010.
36. **Roths, J. B., and C. L. Sidman.** 1993. Single and combined humoral and cell-mediated immunotherapy of *Pneumocystis carinii* pneumonia in immunodeficient SCID mice. Infect. Immun. **61**: 1641–1649.
37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
38. **Sinclair, K., A. E. Wakefield, S. Banerji, and J. M. Hopkin.** 1991. *Pneumocystis carinii* organisms derived from rat and human hosts are genetically distinct. Mol. Biochem. Parasitol. **45**:183–184.
39. **Smulian, G. A.** Personal communication.

39a. **Smulian, G. A., and M. J. Linke.** Personal communication.

40. **Smulian, G. A., D. W. Sullivan, M. J. Linke, N. A. Halsey, T. C. Quin, A. P. MacPhail, M. A. Hernandez-Avila, S. T. Hong, and P. D. Walzer.** 1993. Geographic variation in the humoral response to *Pneumocystis carinii*. J. Infect. Dis. **167**:1243–1247.
41. **Stringer, S. L., T. R. Garbe, S. M. Sunkin, and J. R. Stringer.** 1993. Genes encoding antigenic surface glycoproteins in *Pneumocystis* from humans. J. Eukaryot. Microbiol. **40**:821–826.
42. **Stringer, S. L., S. T. Hong, D. Giuntoli, and J. R. Stringer.** 1991. Repeated DNA in *Pneumocystis carinii*. J. Clin. Microbiol. **29**: 1194–1201.

42a. **Sunkin, S. M., S. L. Stringer, and J. R. Stringer.** 1994. A tandem repeat of rat-derived *Pneumocystis carinii* genes encoding the major surface glycoprotein. J. Eukaryot. Microbiol. **41**:288–296.

43. **Thole, J. E. R., R. Schoeningh, A. A. M. Janson, T. Garbe, J. E. Clark-Curtiss, T. H. M. Ottenhoff, R. R. P. de Vries, and C. Abou-Zeid.** 1992. A major secreted antigen of *Mycobacterium leprae*. Mol. Microbiol. **6**:153–163.
44. **Towbin, H. T., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
45. **Wada, M., K. Kitada, M. Saito, K. Egawa, and Y. Nakamura.** 1993. cDNA sequence diversity and genomic clusters of major surface glycoprotein genes of *Pneumocystis carinii*. J. Infect. Dis. **168**:979–985.
46. **Walzer, P. D.** 1994. Pathogenic mechanisms, p. 251–265. *In* P. D. Walzer (ed.), *Pneumocystis carinii* pneumonia, 2nd ed. Marcel Dekker Inc., New York.
47. **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.
48. **Werner, C., G. I. Higashi, J. A. Jates, and T. V. Rajan.** 1989. Differential recognition of two cloned *Brugia malayi* antigens by antibody class. Mol. Biochem. Parasitol. **35**:209–218.
49. **Yong, L., and M. J. Leibowitz.** 1993. Variation and in vitro splicing of group I introns in rRNA genes of *Pneumocystis carinii*. Nucleic Acids Res. **21**:2415–2421.
50. **Zhang, J., M. T. Cushion, and J. R. Stringer.** 1993. Molecular characterization of a novel repetitive element from *Pneumocystis carinii* from rats. J. Clin. Microbiol. **31**:244–248.
51. **Zimmerman, P. E., D. R. Voelker, F. X. McCormack, J. R. Paulsrud, and W. J. Martin II.** 1992. 120-kD surface glycoprotein of *Pneumocystis carinii* is a ligand for surfactant protein A. J. Clin. Invest. **89**:143–149.