Characterization of Major Surface Glycoprotein Genes of Human *Pneumocystis carinii* and High-Level Expression of a Conserved Region

QIN MEI,† ROSS E. TURNER, VIVIAN SORIAL, DIANE KLIVINGTON, C. WILLIAM ANGUS, AND JOSEPH A. KOVACS*

Critical Care Medicine Department, Clinical Center, National Institutes of Health, Bethesda, Maryland 20892

Received 12 March 1998/Returned for modification 7 May 1998/Accepted 23 June 1998

To facilitate studies of *Pneumocystis carinii* infection in humans, we undertook to better characterize and to express the major surface glycoprotein (MSG) of human *P. carinii*, an important protein in host-pathogen interactions. Seven MSG genes were cloned from a single isolate by PCR or genomic library screening and were sequenced. The predicted proteins, like rat MSGs, were closely related but unique variants, with a high level of conservation among cysteine residues. A conserved immunodominant region (of approximately 100 amino acids) near the carboxy terminus was expressed at high levels in *Escherichia coli* and used in Western blot studies. All 49 of the serum samples, which were taken from healthy controls as well as from patients with and without *P. carinii* pneumonia, were reactive with this peptide by Western blotting, supporting the hypothesis that most adult humans have been infected with *P. carinii* at some point. This recombinant MSG fragment, which is the first human *P. carinii* antigen available in large quantities, may be a useful reagent for investigating the epidemiology of *P. carinii* infection in humans.

*Pneumocystis carinii* remains an important life-threatening opportunistic pathogen of immunocompromised patients, especially those with human immunodeficiency virus (HIV) infection. The major surface glycoprotein (MSG; also called glycoprotein A) is the most abundant protein expressed on the surface of *P. carinii*, as assessed by Coomassie blue staining (22, 29, 36), and appears to play a critical role in the pathogenesis of pneumocystosis, possibly by acting as an attachment ligand to lung cells (7, 28, 45). MSG is also a target of both humoral and cellular immune responses by the host (8, 11, 22, 37–39). Previously, we reported that multiple genes encode humoral and cellular immune responses by the host (8, 11, 22, 33). While animal models can provide important information about the biology of *P. carinii* species (10, 16, 17, 33), these studies suggest that rat MSG variants (5, 35, 42, 43). These studies suggest that *P. carinii* has developed an elaborate system for antigenic variation, presumably to evade host defense mechanisms.

Molecular and immunological studies have clearly demonstrated that *P. carinii* organisms isolated from different host species are distinct organisms and may in fact be separate species (10, 16, 17, 33). While animal models can provide important information about the biology of *P. carinii*, studies examining human interactions with *P. carinii* need to use human *P. carinii*-derived reagents. The cloning of human *P. carinii* (Pre*merocystis carinii* f. sp. *hominis*) MSG genes has recently been reported (9, 34). Since only one full-length sequence was reported, the present study was undertaken to further characterize the *P. carinii* f. sp. *hominis*-derived family of MSG genes. In addition, we undertook to express these genes, since *P. carinii* f. sp. *hominis* cannot be cultured and there is no reliable source of organisms for purifying large amount of antigens or other biologically relevant proteins.

---

**MATERIALS AND METHODS**

**DNA preparation.** DNA was isolated from an autopsy lung sample from an HIV-infected patient with *P. carinii* pneumonia according to standard methods, using sodium dodecyl sulfate (SDS) and proteinase K (0.5 μg/ml), followed by phenol-chloroform extraction and ethanol precipitation (3). A genomic library using the same DNA cloned into the XhoI site of the lambda GEM12 vector (Promega, Madison, Wis.) was commercially prepared (Lofstrand Labs Limited, Gaithersburg, Md.).

**PCR and subcloning.** Primers to amplify full-length human *P. carinii* genes were designed on the basis of published data (9). The sense primer, JK151 (5′-TTT CAT ATG GCC CGG GCG GTC AGG CCG CAG-3′), corresponds to nucleotides 155 to 175 of a published MSG sequence (GenBank accession no. L27092), and the antisense primer, JK152 (5′-CTA CAT GAT GAA CGA AAT AAC TAC TGC TAC TGC TAC-3′), is complementary to nucleotides 3215 to 3242. An *Nde*I site, which substitutes a methionine for the valine of the original sequence, was created at the beginning of JK151 in order to facilitate subcloning and expression. For amplification, 1 μg of genomic DNA was added to a 50-μl reaction mixture containing primers (25 pM each), deoxynucleoside triphosphates (0.2 mM), 5 U of AmpliTag (Perkin-Elmer), and MgCl₂ (2.5 mM). DNA amplification was performed on a Perkin-Elmer Cetus DNA thermal cycler. An initial denaturation cycle (1 min at 98°C) was followed by 36 cycles of denaturation at 95°C for 1 min, annealing at 50°C for 2 min, and extension at 72°C for 2 min, followed by a final extension after the last cycle at 72°C for 10 min.

Primers for amplifying the conserved carboxyl-terminal region of the human *P. carinii* MSG gene were designed on the basis of the alignment of five new MSG genes as well as the published sequence. The sense primer was JK451 (5′-GAA TTC GAT CTG AAG CTT CGG TCT CTC-3′), and the antisense primer was JK452 (5′-TTC TAG AAA CCC ACT CAT CTG CGA AAT-3′). An EcoRI site was added to the antisense primer, and an Xhol site, which encodes an in-frame stop codon, was added to the sense primer, to facilitate subcloning. One microgram of plasmid DNA was used for PCR amplification under the conditions described above. PCR amplification products were subjected to electrophoresis in 1% agarose gel in 1× Tris-borate-EDTA buffer; then PCR products were directly subcloned into PCR II (Invitrogen, Carlsbad, Calif.) according to the manufacturer’s instructions.

**Southern hybridization and library screening.** For Southern hybridization with a radioactive probe, DNA was treated with restriction enzymes, separated by agarose gel electrophoresis, and transferred to Hybond N+ membranes (Amersham Life Science, Arlington Heights, Ill.) with 0.4 M NaOH. DNA was...
probed with an approximately 600-bp XbaI fragment of the human P. carinii MSG III gene (9) (a gift from James R. Stringer, University of Cincinnati, Cincinnati, Ohio) that had been labeled with [α-32P]dATP or [α-32P]dCTP by using a random priming kit (Boehringer Mannheim). Filters were prehybridized for 4 h and then hybridized overnight at 55°C in 6× SSPE (1× SSPE is 0.18 M NaCl, 10 mM Na2HPO4, and 1 mM EDTA) at pH 7.7) and 5× SDS and 5× Denhardt’s solution. Blots were washed in 6× SSPE-0.5% SDS at room temperature for 10 min and then in 0.5× SSPE-0.5% SDS at 55°C twice for 30 min each time. The genomic library was screened by using a purified full-length fragment of human P. carinii MSG 11 under the conditions described above. One clone that hybridized strongly to the probe was cloned into the BamHI site of pBluescript II (Stratagene, La Jolla, Calif.).

Nucleotide sequencing. Sequencing was performed with an automated sequencer (model 373 or 377; Applied Biosystems, Perkin-Elmer, Foster City, Calif.) either in our laboratory or by contract (San Diego State University, San Diego, Calif.). The nucleotide sequence and deduced amino acid sequence data were analyzed by Factura (Gene Codes Corp., Ann Arbor, Mich.), MacVector (Scientific Imaging Systems, New Haven, Conn.), ClustalW (40), and GeneWorks (IntelliGenetics, Mountain View, Calif.).

Construction and expression of recombinant human P. carinii MSG. The full-length human P. carinii MSG 32 gene was inserted into pBlueBacHis2A (Invitrogen) at the EcoRI site for expression in a baculovirus insect cell system. Correct insertion was confirmed by restriction mapping and sequencing. Isolation of recombinant virus plaque purification, and amplification of higher titer virus stocks were performed according to the manufacturer’s protocols (Invitrogen). PCR amplification with gene-specific primers was used to confirm the presence of the gene in the virus. Sf9 cells were grown at 27°C in SFII-900 medium (GIBCO BRL, Grand Island, N.Y.) with 5% fetal calf serum to a density of 2.0×10^6 cells/ml. Cells were infected at a multiplicity of infection of 5. Seventy-two hours after infection, cells were harvested by centrifugation, washed with phosphate-buffered saline (PBS) supplemented with phenylmethylsulfonyl fluoride (PMSF) (1 mM), and then resuspended in 10 mM Tris-HCl, pH 8, with 1 mM PMSF and sonicated. The cell lysate was analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting.

For expression of a conserved region of the MSG in Schizochytrium coltii, the 306-bp PCR product of this region was ligated in frame into pET28A (Novagen, Madison, Wis.) at the EcoRI site. pET28A is an expression vector in which a histidine tag precedes the insertion site. Restriction mapping and sequencing were performed to confirm correct insertion. Expression was induced in E. coli BL21(DE3) using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Recombinant protein was solubilized with 6 M urea and purified by affinity chromatography using a nickel column according to the manufacturer’s instructions (Novagen). The sample was eluted with elution buffer without urea, dialyzed by using 0.5× PBS to eliminate imidazole, and lyophilized for storage. Recombinant protein was analyzed by SDS-PAGE and Western blotting.

SDS-PAGE and Western blotting. SDS-PAGE and Western blotting were performed by standard techniques, as previously described (18). Electrophoresis was carried out in preprepared discontinuous 8% and 14% acrylamide-Tris-glycine gels (Novex, San Diego, Calif.). Proteins were stained with Coomassie blue or silver (26). Anti-peptide antisera to a peptide specific for MSG 32 (KMYGLFYGSGKE) (Research Genetics, Huntsville, Ala.), anti-Xpress monoclonal antibody (TITSTITSKITLTST, corresponding to amino acids 968 to 981) (Invitrogen), and anti-rabbit-IgG, alkaline phosphatase-conjugated goat anti-rabbit IgG and goat anti-mouse IgG (all from GIBCO BRL), or horseradish peroxidase-conjugated goat anti-rat IgG, anti-rat IgG, and anti-mouse IgG (Jackson ImmunoResearch Laboratories, Inc., West Grove, Pa.) were used as second antibodies in Western blotting.

Nucleotide sequence accession numbers. The sequences of the five PCR-generated human P. carinii MSG clones and the 12,792-bp clone from the genomic library have been deposited in GenBank under accession no. AF033208 through AF033322 and AF038556 respectively.

RESULTS

Although the human P. carinii MSG gene has recently been identified, only one full-length clone has been reported (9, 34). Since multiple variant copies of the MSG gene are present in the rat P. carinii genome, and available data support a similar organization in human P. carinii (9, 34), we wanted to obtain sequence information on additional full-length human P. carinii MSG clones in order to better characterize this family of genes. Since in rat P. carinii the S′ and 3′ ends of the MSG genes are highly conserved, primers based on the published sequences of these regions of the human P. carinii MSG gene (with modifications to facilitate subcloning and expression) were used in an attempt to amplify additional full-length genes. By using DNA isolated from the lung of an AIDS patient with P. carinii pneumonia, a band of the correct size (approximately 3.1 kb) was amplified and subcloned. Five clones that differed in their restriction mapping and hybridization patterns were identified and sequenced (GenBank accession no. AF033208 to AF033212). All clones encoded MSG variants that were clearly related but differed from each other. The coding regions of the clones varied in length from 3,054 to 3,087 bases, encoding proteins of 1,008 to 1,028 amino acids with predicted molecular sizes of 114 to 117 kDa. When pairs of clones are compared, they are 74 to 91% identical at the nucleotide level and 63 to 88% identical at the amino acid level. Overall, approximately 50% of the amino acids are conserved in all five clones. The clones are more closely related to each other than to rat P. carinii MSG genes. There is approximately 60% identity at the DNA level and 40% identity at the amino acid level between human P. carinii MSG and rat P. carinii MSG GP3.

To obtain additional P. carinii f. sp. hominis MSG sequences, a genomic library was screened with one of these clones, and one clone that hybridized strongly was subcloned and sequenced. This 12,792-bp clone (GenBank accession no. AF038556) contained three full-length MSG sequences and one partial MSG sequence in a head-to-tail tandem arrangement, similar to that previously reported (9, 34). One of the full-length MSG sequences did not have a complete open reading frame due to a frame shift between bases 6290 and 6347. The codon corresponding to a methionine at the beginning of rat P. carinii MSG clones encoded a valine in all the open reading frames, consistent with earlier observations (9, 34).

Figure 1 shows an alignment of the predicted proteins encoded by these genes, together with a rat P. carinii MSG sequence. Among the human P. carinii MSG sequences, there is substantial variability downstream of the amino terminus, while the region near the carboxyl terminus is highly conserved. For example, there is 63% identity in the last 100 amino acids among all the genes (excluding the region encoded by the PCR primer), which is about 5 times as high as the conservation among the first 100 amino acids (13% excluding the primer region). Like most known genes of P. carinii, all human P. carinii MSG genes show a strong AT bias, especially in the third position (approximately 70% A or T) (4, 9, 19, 41). As in other MSG molecules, cysteine residues of the human P. carinii MSG molecules are relatively numerous (5.7 to 5.9%) and are highly conserved: 96% of all the cysteine residues present in the human P. carinii MSG clones are conserved in all the clones. In a comparison between human MSG 11 and rat P. carinii MSG clone GP3, 94% of cysteine residues are conserved. The cysteine residues are unevenly distributed in two main regions and often show a pattern of two cysteines separated by six to seven amino acids, similar to the pattern seen in rat P. carinii (19). There is no predictable pattern to the intervening amino acids. All MSG proteins share a highly conserved amino acid domain rich in threonine and serine residues near the carboxyl terminus. Seven to 13 potential N-linked glycosylation sites [NX(S,T)] were identified in the MSGs. A premature stop codon was seen in MSG 32 after residue 1008, most probably due to a PCR artifact resulting in a point mu-
FIG. 1. Alignment of the deduced amino acid sequences encoded by two of the human *P. carinii* MSG genes contained in the genomic clone (HMSGp1 and HMSGp3) and the five genes generated by PCR (HMSGG1, HMSGG4, HMSGG3, and HMSGG5), together with a published sequence (GBHMSG) and a rat *P. carinii* MSG sequence (RMSGGP3) (GenBank accession no. L05906). Alignment was performed with ClustalW. Amino acids that are conserved in at least six of the sequences are boxed. A methionine was substituted for valine at position 1 in the PCR clones during PCR to facilitate expression and thus is excluded from the alignment. The peptides that were synthesized and used to generate antipeptide antibodies are underscored with a solid box (conserved epitope) or a diagonally striped box (human MSG 32-specific epitope). The arrows flank the conserved region that was expressed in pET28a.
**FIG. 2.** Immunoblots of recombinant nearly full-length human 
*P. carinii* MSG 32 expressed in a baculovirus system. Human 
*P. carinii* MSG 32 has a stop codon at amino acid 1010 that is likely a PCR artifact but results in elimination of the 
hydrophobic tail. For both blots, the lanes are as follows: lanes 1, recombinant 
baculovirus-derived rat *P. carinii* MSG GP3; lanes 2, recombinant baculovirus-
derived human *P. carinii* MSG 32; lanes 3, Sf9 cells alone. (A) The reactivity of 
Xpress monoclonal antibody, which was generated against a peptide tag encoded by 
the vector, can be seen with a protein of approximate 116 kDa (arrow). (B) The reactivity of an antipeptide antibody generated against a peptide specific for human 
*P. carinii* MSG 32 can be seen with a protein of the same molecular size (arrow).

**FIG. 3.** Time course of expression of a conserved region of human *P. carinii* 
MSG 33, as evaluated by SDSPAGE (14% gel) and Coomassie blue staining. 
(A) Expression of recombinant protein (arrow) at different time points after 
induction with IPTG (1 mM). Maximal expression of the protein can be reached 
at 2 h (arrow); (B) The pET28A vector alone under the same conditions. No 
equivalent band is seen.

**FIG. 4.** Purification of a conserved region of human *P. carinii* 
MSG 33 (arrow). The purified protein was obtained from a 1-liter culture of *E. coli*. The 
identity of the protein was confirmed by immunoblotting using both the T7-tag monoclonal antibody and a polyclonal anti-
epitope antibody generated in rabbits against an epitope (TTTSTTSKITTST) contained within the recombinant carboxy-
terminal fragment (Fig. 5A). No reactivity was seen with pre-
immune rabbit serum, with uninduced *E. coli* extracts (data not shown), or with the second antibody alone (Fig. 5C).

To evaluate the utility of this recombinant peptide as a tool for investigating the seroepidemiology of *P. carinii* infection, 
immunoblotting studies with a variety of human serum samples 
(diluted 1:100) were undertaken. Samples included those from 
11 immunosuppressed patients with recent or acute *P. carinii* 
pneumonia but without HIV infection, from 5 patients with HIV 
infection and *P. carinii* pneumonia, from 17 patients with HIV 
infection but without *P. carinii* pneumonia, from 3 patients 
with neither HIV infection nor *P. carinii* pneumonia, and from 
13 healthy laboratory workers. All 49 samples reacted by im-
munoblotting with the recombinant peptide (Fig. 5B). Because the 
recombinant peptide included a region that was vector 
derived, a subset of four samples was simultaneously evaluated by 
immunoblotting for reactivity with recombinant β-galac-
tosidase expressed in the same vector. None of the samples 
reacted with the recombinant β-galactosidase (data not shown), 
demonstrating that the reactivity seen was against the *P. carinii*-
derived peptide region. In addition, little or no reactivity 
was seen when rat, mouse, or cat serum was used (Fig. 5C).

**DISCUSSION**

In recent years, a great deal of interest has focused on the 
MSG of *P. carinii*, since it is both a likely virulence factor and

---

**FIG. 5.** Purification of a conserved region of human *P. carinii* 
MSG 33. Lane 1, whole *E. coli* extract following a 2-h induction with IPTG (1 mM) and sol-
bilization in binding buffer containing 6 M urea; lane 2, flowthrough after binding 
of the recombinant protein to a nickel column; lane 3, eluate following elution with-
out urea; lane 4, threefold-concentrated purified protein; lane 5, eluate following elu-
tion with 3 M urea. Elution with urea leads to a broader band on SDSPAGE gels.
alkaline phosphatase-conjugated second antibody was used. For each of the remaining lanes, the appropriate recombinant peptide was used. (A) Lane 1, reactivity of the T7-tag monoclonal antibody, which reacts with a pET28A vector-derived epitope that precedes the MSG peptide (arrow); lane 2, reactivity with a polyclonal anti-epitope antibody generated against a conserved epitope contained within the recombinant MSG fragment; lanes 3 and 4, reactivity with serum samples from healthy humans (diluted 1:100). (B) Reactivities of a variety of human serum samples (diluted 1:100). Lanes 1, 4 and 6, serum samples from healthy humans; lanes 2 and 7, serum samples from HIV-infected patients with a history of P. carinii pneumonia; lanes 3, 5, and 8, serum samples from patients with a history of P. carinii pneumonia but without HIV infection; lane 9, serum from an immunosuppressed patient without P. carinii pneumonia or HIV infection. All samples reacted by immunoblotting with the recombinant peptide (arrow). (C) Reactivity with no first antibody (lane 1) or reactivity with serum samples from healthy humans (diluted 1:100). (B) Reactivity was not seen with a control protein expressed in the same bacterial expression system and was easily purified via a six-histidine tag encoded in the vector. By immunoblotting, all 49 human serum samples tested were reactive with this peptide, regardless of their immune status or history of P. carinii pneumonia, supporting the hypothesis that most humans have been exposed to P. carinii at some point (22, 25). Reactivity was not seen with a control protein expressed in the same plasmid. Little or no reactivity was seen when serum from a rat, mouse, or cat was used (Fig. 5), suggesting that the responses are specific for human P. carinii and not a result of exposure to an irrelevant environmental antigen. The low level of reactivity seen with the rat serum may also represent cross-reactivity due to prior exposure to P. carinii. Lane 1, sp. carinii, given that there is homology between rat and human P. carinii MSG sequences in this region (Fig. 1).

Previous studies evaluating the serological responses of humans to P. carinii have relied on a variety of assays, including fixed-tissue staining, immunofluorescence, enzyme-linked immunosorbent assays, and Western blotting using either rat or human P. carinii organisms or purified proteins as antigens. Results have been conflicting; some studies have shown that a high proportion of humans have serological reactivity with P. carinii antigens, while others have shown low response rates (2, 6, 15, 21–23, 26, 32). In previous studies using purified human P. carinii MSG, we noted response rates of 34% to 66% (23). Western blot studies have shown similar rates of reactivity with MSG (26). Differences between these previous results and the present results may be related to the sensitivity of the techniques, the integrity of the human P. carinii antigens used in prior studies (given that the antigens were derived from autopsy samples), or other methodological differences. Thus, this peptide is the first human P. carinii-specific antigen available in sufficient quantities for large-scale studies. The preliminary studies reported here suggest that this antigen may be a useful tool for seroepidemiologic investigation of P. carinii infection in humans, for example, to identify the period of seroconversion in humans, although more extensive studies are needed to verify this utility.

While antibodies appear to play a role in clearing P. carinii, T-cell response appears to be of primary importance in clearance (12, 30). Additional studies are needed to determine if this recombinant fragment contains T-cell epitopes in addition to B-cell epitopes. If so, evaluation of proliferative responses to this antigen may provide useful prognostic information, for example, about the risks of developing P. carinii pneumonia during HIV infection. In addition, given that it is a highly conserved region of the highly variable MSG, this recombinant
peptide is a good candidate for evaluation as a vaccine for the prevention of P. carinii pneumonia.

REFERENCES

2. Peptide is a good candidate for evaluation as a vaccine for the prevention of P. carinii pneumonia.

ACKNOWLEDGMENTS

We thank James Stringer and Saundra Stringer for providing a fragment of a human P. carinii MSG gene.

REFERENCES

2. Peptide is a good candidate for evaluation as a vaccine for the prevention of P. carinii pneumonia.