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

QIN MEI,† ROSS E. TURNER, VIVIAN SORIA, DIANE KLI VINGTON, 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, 23, 37–39). Previously, we reported that multiple genes encode the MSG of rat *P. carinii*, and we demonstrated that different MSGs can be expressed in the lung of a rat infected with *P. carinii* (1, 19). Similarly, multiple genes encode the MSG of *P. carinii* infecting ferrets and mice (13, 14, 44). Additional studies have shown that there is a single genomic site for expression of 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* (*Pneumocystis 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.

(This work was presented in part at the 5th International Workshops on Opportunistic Protists and the 5th General Meeting of the European Concerted Action on Pneumocystis Research, September, 1997, Lille, France [31].)

**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, by using sodium dodecyl sulfate (SDS) and protease K (0.5 μg/ml), followed by phenol-chloroform extraction and ethanol precipitation (3). A genomic library using the same DNA cloned into the *Xho* 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 TTC GTA TAA AGA-3'), corresponds to nucleotides 153 to 175 of a published MSG sequence (GenBank accession no. L27092), and the antisense primer, JK152 (5'-CTA AAC CAT GAA CGA AA-3'), is complementary to nucleotides 3215 to 3244. An *NdeI* 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 μM each), deoxynucleoside triphosphates (0.2 mM), 5 U of AmpliTaq (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 genes 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 CTT GAC ACC AAC CCG CTT CCA CAG-3'), and the antisense primer was JK452 (5'-TTC TAG AAA CCC ACT CAT CTT CAA-3'). An *EcoRI* site was added to the sense primer, and an *XhoI* site, which encoded an in-frame stop codon, was added to the antisense 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 the 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

---

* Corresponding author. Mailing address: Building 10, Room 7D43, 10 Center Dr., MSC 1662, Bethesda, MD 20892-1662. Phone: (301) 496-9907. Fax: (301) 402-1213. E-mail: jkovacs@nih.gov.
† Present address: Merck Research Laboratories, Dept. of Drug Metabolism, West Point, PA 19486-0004.
probe with an approximately 600-bp XbaI fragment of the human \textit{P. carinii} MSG III gene (9) (a gift from James R. Stringer, University of Cincinnati, Cincinnati, Ohio) that had been labeled with [a-32P]dATP or [a-32P]dCTP by using a random priming kit (Boehringer Mannheim). Filters were prehybridized for 4 h and then hybridized overnight at 37°C in 6x SSPE (1x SSPE is 0.18 M NaCl, 10 mM NaH2PO4, and 1 mM EDTA [pH 7.7])-0.5% SDS and 5x Denhardt's solution. Blots were washed in 6x SSPE-0.5% SDS at room temperature for 10 min and then in 0.5x 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 \textit{P. carinii} MSG 11 under the conditions described above. One clone that hybridized strongly to the probe was subcloned 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 and AutoAssembler (both from Applied Biosystems). Sequencer (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 \textit{P. carinii} MSG. The full-length human \textit{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 high-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 recombinant virus, plaque purification, and amplification of high-titer virus stocks. Since multiple variant copies of the MSG gene are present in \textit{P. carinii}, the \textit{MSG} gene (9) (a gift from James R. Stringer, University of Cincinnati, Ohio) that had been labeled with [a-32P]dATP or [a-32P]dCTP by using a random priming kit (Boehringer Mannheim). Filters were prehybridized for 4 h and then hybridized overnight at 55°C in 6x SSPE (1x SSPE is 0.18 M NaCl, 10 mM NaH2PO4, and 1 mM EDTA [pH 7.7])-0.5% SDS and 5x Denhardt's solution. Blots were washed in 6x SSPE-0.5% SDS at room temperature for 10 min and then in 0.5x 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 \textit{P. carinii} MSG 11 under the conditions described above. One clone that hybridized strongly to the probe was subcloned into the BamHI site of pBluescript II (Stratagene, La Jolla, Calif.).

**Characterization of Human \textit{P. carinii} MSG**

**Figures**

- **Figure 1** shows an alignment of the predicted proteins encoded by these genes, together with a rat \textit{P. carinii} MSG sequence. Among the human \textit{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 \textit{P. carinii}, all human \textit{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 \textit{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 \textit{P. carinii} MSG clones are conserved in all the clones. In a comparison between human MSG 11 and rat \textit{P. carinii} MSG clone GP3, 94% of cysteine residues are conserved. The cysteine residues are unevenly distributed in four main regions and often show a pattern of two cysteine residues separated by six to seven amino acids, similar to the pattern seen in rat \textit{P. carinii} (19). There is no predictable pattern to the internal cysteine 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 observed 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 (HMSG11, HMSG14, HMSG32, HMSG33, and HMSG35), 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 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). (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 approximate 116 kDa (arrow). (B) The pET28A vector alone under the same conditions. No equivalent band is seen.

FIG. 3. Time course of expression of a conserved region of human P. carinii MSG 33, as evaluated by SDS-PAGE (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). Lane 1, whole E. coli extract following a 2-h induction with IPTG (1 mM) and solubilization 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 without urea; lane 4, threefold-concentrated purified protein; lane 5, eluate following elution with 3 M urea. Elution with urea leads to a broader band on SDS-PAGE gels.

Characterization of Human P. carinii MSG

A major goal of these studies was to produce agents derived from P. carinii f. sp. hominis that could be used to investigate immune responses to human P. carinii, a major effort was made to express a full-length P. carinii f. sp. hominis MSG gene at a high level. By using a baculovirus-insect cell system, a nearly full-length clone, MSG 32 (which contains the premature stop codon), was expressed. A time course showed that maximal expression occurred after 60 to 72 h of infection. The identity of the recombinant protein was confirmed by Western blotting using both an antibody against a peptide tag present in the vector and an antipeptide antibody raised against a peptide specific for MSG 32 (Fig. 2). No reactivity was seen when Sf9 cells alone or recombinant baculovirus-derived rat MSG GP3 was used as the target. The multiple bands seen in the Western blots, especially when the MSG-specific antipeptide antibody was used, likely represent protein degradation products, or possibly modification of the recombinant protein.

Although rat MSG GP3 could be produced at a high level in a baculovirus system and was easily purified by affinity chromatography using a nickel column (24), prolonged attempts to produce and purify high levels of human P. carinii MSG were unsuccessful. We then focused on expressing a highly conserved region at the carboxyl terminus of human P. carinii MSGs (Fig. 1). Preliminary data obtained by epitope mapping have demonstrated that this region is highly immunogenic for antibody production in both rats and humans (1a), suggesting that a peptide encompassing this region could be used in seroepidemiological studies.

PCR was used to amplify this conserved region without the carboxyl-terminal hydrophobic tail, since this hydrophobic tail could potentially interfere with expression and purification. A fragment of approximately 300 bp was obtained by PCR amplification using primers JK451 and JK452, with MSG 33 as a template. The PCR product was subcloned into pET28A, and expression was induced by culturing in 1 mM IPTG. High-level expression was observed within 2 h (Fig. 3A); no equivalent band was seen when pET28A was used without an insert under the same conditions (Fig. 3B). The presence of a six-histidine sequence in the expressed portion of the vector preceding the insert allowed rapid, one-step purification of the recombinant protein (Fig. 4). Although the yield was variable from experiment to experiment, during a typical study about 7 mg of pu-
sequences have confirmed that the Lys-Arg sequence located 5 positions downstream of the valine is conserved in all *P. carinii* f. sp. *homininis* MSGs identified to date. This is consistent with the postulated role of this sequence in all *P. carinii* strains as a target for cleavage by *P. carinii* kexin (20, 31), which would eliminate the conserved leader from the surface-expressed form of the MSG variants and maximize antigenic variation.

A second major focus of the present study was to express *P. carinii* f. sp. *homininis* MSG at a high level in order to provide reagents for studies of human-*P. carinii* interaction. While a nearly full-length MSG could be expressed in a baculovirus system, we were unable to induce high-level expression, perhaps because of the high proportion of cysteine residues. The presence of a hydrophobic tail may also interfere with expression, as the only full-length clone we were able to express had a premature stop codon prior to this hydrophobic tail.

However, an immunodominant region at the carboxyl terminus (minus the hydrophobic tail) could be expressed at a high level in a 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* f. sp. *homininis* variants, 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 responses appear 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.

ACKNOWLEDGMENTS

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

REFERENCES


Characterization of Human P. carinii MSG.