Identification and Characterization of Novel Variant Major Surface Glycoprotein Gene Families in Rat *Pneumocystis carinii*

Sheng N. Huang, C. William Angus, Ross E. Turner, Vivian Sorial, and Joseph A. Kovacs

The major surface glycoprotein (MSG) is an abundant, immunodominant protein on the surface of the opportunistic pathogen *Pneumocystis carinii*. The current study identified two novel variant MSG (vMSG) gene families in rat *P. carinii* that are closely related to but distinct from MSG. These gene families encode proteins of ~90 kDa (v1MSG) and ~115 kDa (v2MSG). Compared with MSG, v1MSG is characterized by a deletion near the carboxyl terminus. The predicted v1MSG and v2MSG proteins are highly homologous to MSG at the carboxyl, but not the amino, terminus. Like MSG, they are cysteine-rich. Approximately 10% of the apparent molecular weight is due to N-linked glycosylation. Southern blotting studies demonstrated that, like MSG, v1MSG and v2MSG are the products of multicopy gene families. However, unlike MSG, each vMSG gene encodes a signal peptide, suggesting that the regulation of vMSG is different from that of MSG.

*Pneumocystis carinii* remains a major pathogen in immunocompromised patients, in particular, those infected with human immunodeficiency virus infection [1–3], and thus there is a great deal of interest in better understanding its basic biology and host interactions. *P. carinii* are coated with an abundant and highly immunogenic surface glycoprotein called the major surface glycoprotein (MSG, also called gpA) [4–6]. MSG, a highly mannosylated glycoprotein with an 
molecular weight of ~110–120 kDa, has been studied extensively because it may be a target for therapeutic intervention [4, 7, 8]. Studies have demonstrated that MSG appears to be involved in *P. carinii* attachment mechanisms via interaction with fibronectin on alveolar epithelial cells and with mannose receptors on macrophages [9, 10]. MSG can induce cellular immune responses mediated by T cells [11–13] and can elicit cytokine secretion [13, 14]. Thus, MSG may play a pivotal role not only in host-organism interactions but also in inducing host defense mechanisms.

MSG shows considerable sequence variation in preparations from different host species [15, 16]. In *P. carinii* within a single host species, MSG is encoded by a family of related but distinct genes distributed throughout the *P. carinii* genome [17–19]. These genes are clustered in tandem repeats, primarily at the telomeres of chromosomes [20, 21]. Variation of MSG gene expression is mediated by a DNA element called the upstream conserved sequence (UCS) that encodes a leader sequence common to all expressed MSGs [21–24]. Recent studies have shown that there is a single expression site for MSG, which includes the UCS and a downstream region for the variable part of the MSG gene family [22–24]. Previous studies have suggested that there is a second UCS associated with MSG expression, although a detailed evaluation has not been reported [24]. In this current study, we report the characterization of two families of genes that encode MSG variants with a mechanism for regulation of expression that is different from that of MSG.

**Materials and Methods**

*P. carinii*. Rat-derived *P. carinii* were obtained from steroid-treated rats maintained at a contract facility (Biocon, Rockville, MD) or at the National Institutes of Health (NIH, Bethesda, MD) and were partially purified by ficoll-hypaque density-gradient centrifugation as described [17]. *P. carinii* pellets were stored at −70°C for subsequent DNA and RNA extraction and immunoblot studies or used immediately for preparation of plugs for chromosomal analysis by pulsed-field gel electrophoresis (PFGE), as previously described [25].

**Library screening and Southern and Northern hybridization.** Construction of *P. carinii* cDNA libraries in lambda ZAP or a modified lambda ZAP have been described [17, 26]. For library screening, a 1.9-kb DNA fragment was amplified from *P. carinii* genomic DNA by polymerase chain reaction (PCR) with primers JK268 and JK269 (table 1), which were designed from a genomic clone that encodes an MSG-like protein called MSG99 [27]. The PCR product was gel-purified and labeled with [α-32P]dCTP with the Random Primed DNA labeling kit (Boehringer Mannheim, Indianapolis). The cDNA library was also screened with oligonucleotides JK269, JK270, and JK92 (table 1). Oligonucleotides were labeled with [γ-32P]ATP by T4 polynucleotide kinase (Ready-to-Go Molecular Biology Reagents, Pharmacia Biotechnology, Piscataway, NJ). Prehybridization and hybridization were performed...
Table 1. Sequences of oligonucleotides used in the study.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>JK270</td>
<td>ATGTTTTTATCATAATTAATAAAATA</td>
</tr>
<tr>
<td>JK268</td>
<td>GAATTTGAAAAATTC</td>
</tr>
<tr>
<td>JK269</td>
<td>TTCTTTGATGGTTTATC</td>
</tr>
<tr>
<td>JK298</td>
<td>CTTAATAAGGACCAAATT</td>
</tr>
<tr>
<td>JK299</td>
<td>TTGGGATCCTGAAATAAT</td>
</tr>
<tr>
<td>JK120</td>
<td>GCTTGGCTTAAACCGCCTGCCATGCC</td>
</tr>
</tbody>
</table>

NOTE. vMSG, variant major surface glycoprotein.

Results

Identification of genes encoding vMSGs. In examining the sequences of 3 P. carinii MSG genomic clones (PMCMG910, a clone described in [32], and AG3 in our laboratory), we noted that each contained regions of an apparently disrupted MSG gene that were highly homologous with each other. The proteins encoded by these genes were highly homologous to MSG at the carboxyl terminus but were different from MSG immediately upstream of this region. Compared with MSG, all three genes contained a similar deletion of 1000 bases. This suggested that these genes belonged to a family of expressed vMSG genes. We thus undertook studies to determine whether these genes were transcribed and translated and to characterize the encoded proteins.

In order to test our hypothesis, a 1.9-kb PCR fragment amplified from genomic DNA by primers JK268 and JK269, which contains the region immediately upstream of the deletion, was used to screen a rat P. carinii cDNA library. Under high stringency conditions, 18 of 2 million clones hybridized to this PCR fragment. All clones were excised and partially sequenced. Six of the 18 clones were found to have similar deletions, whereas 12 clones did not contain the deletion. Because the longest clone (v1MSG.9, figure 1A) was incomplete, we rescreened the library with JK269, which is located immediately upstream of the de-
Figure 1. A, Alignment of predicted amino acid sequence of 7 cDNA clones encoding variant major surface glycoproteins (vMSGs) of rat *P. carinii* (v2MSG.9, 15, 16, 2, and 7 and v1MSG.17 and 9) compared with MSG clone GP3. Boxed amino acids are identical to those in clone v2MSG.9. Predicted signal peptide cleavage site is marked by arrow. GenBank accession nos.: v2MSG.9-AF063234, v2MSG.15-AF063235, v2MSG.16-AF063236, v2MSG.2-AF063237, v2MSG.7-AF063238, v1MSG.17-AF063232, and v1MSG.9-AF063233. B, Alignment of 5 upstream untranslated region of v2MSG.9, 15, and 16 and v1MSG.17. Oligonucleotide sequences used for hybridization to *P. carinii* genomic DNA are underlined.

deleted region and appears to be specific for this vMSG gene family. JK269 hybridized to 10 of 2 million clones under high stringency conditions. Partial sequencing showed that all 10 clones had the characteristic deletion when compared with the MSG clone GP3, but only 1, which was completely sequenced, appeared to be full-length. This 2384-bp clone encoded a protein of 775 amino acids, which we have called v1MSG. When compared with GP3, v1MSG contained a deletion of ~320 amino acids immediately prior to the carboxyl terminus (figure 1A).

In order to obtain additional full-length sequences, we rescreened the cDNA library using JK270, which corresponds to the first 21 nucleotides of the coding region, and JK92, which corresponds to a variable region with little homology to MSG. JK270 hybridized to 3 clones and JK92 to 8 clones among 1.5–2 million clones screened. Two clones that hybridized to JK270 and 3 that hybridized to JK92 were sequenced. All were polyadenylated, but none contained a deletion similar to v1MSG. Three clones (2 identified by JK270 and 1 by JK92) appeared to be full-length clones of ~3500 bp encoding related proteins of ~1066 amino acids, which we have called v2MSGs. v2MSGs are highly homologous to v1MSG at the amino terminus and homologous to both standard MSGs and v1MSG at the carboxyl terminus.

To determine the size of the transcripts corresponding to the cDNA encoding the vMSGs, Northern blot analysis was performed. A single band of ~2.4 kbp was detected with JK269 (figure 2). JK92 hybridized to two bands of 2.4 and 3.5 kbp (figure 2), which correspond to v1MSG and v2MSG transcripts. Figure 2 also shows that the signal seen with probes specific for vMSGs are less intense than with an MSG-specific probe (JK120), reflecting the relative abundance of MSG compared with vMSG message.

Because we identified multiple related but unique vMSG genes and because MSGs are encoded by a multicy copy gene family, we undertook Southern hybridization studies to see if vMSGs are also encoded by multicy copy gene families. As shown in figure 3A, when JK269, JK92, and JK270 were hybridized to *P. carinii* genomic DNA that had been digested with EcoRI, *Hind*III, or *Xba*1, multiple bands were seen with each oligonucleotide, which strongly suggested the presence of multiple genes encoding vMSGs. Using JK269 and JK270 as probes, Southern blots of *P. carinii* chromosomes separated by PFGE demonstrated hybridization of each oligonucleotide to multiple chromosomes, further confirming that multiple vMSG genes are present in the *P. carinii* genome (figure 3C). Of interest, JK270 hybridized to chromosomes to which JK269 did not hybridize, suggesting that the distribution of v1MSG and v2MSG genes on *P. carinii* chromosomes is different. The presence of multiple vMSG genes in the *P. carinii* genome suggests that, similar to MSG, the vMSG cDNA sequence diversity is likely caused by the diversity of genomic vMSG genes rather than RNA editing of one gene.

To test whether the vMSG genes were translated, a rabbit polyclonal antiserum was prepared against a predicted vMSG-specific peptide (which is encoded by oligonucleotide JK92) and utilized in Western blot studies with *P. carinii* proteins. Using antibody that was affinity-purified from this serum, no reaction was seen by immunoblot with a recombinant MSG, GP3 (data not shown). Two bands, of 115 and 86 kDa (figure 4A), which correspond to the predicted sizes of v2MSG and v1MSG, were seen when using a whole *P. carinii* preparation in the immunoblot studies. Both proteins were N-glycosylated, as an ~10% decrease in *M*₄ was seen after deglycosylation with endoglycosidase F or H (figure 4B). Because MSG is solubilized by treatment with lyticase, a β-1,3 glucanase, we examined the effect of lyticase on vMSGs. Both vMSGs were solubilized by lyticase and had an apparent decrease in *M*₄ of ~10% (figure 4C).

To further confirm that the vMSGs were expressed, we per-
Figure 2. Northern blots of 2 independent preparations of rat *P. carinii* total RNA that were probed with JK269, which is specific for v1MSG (A); JK92, which is common to v1MSG and v2MSG (B); and JK120, which is specific for MSG (C). Hybridization is evident in A to 2.4-kb mRNA, in B to 2.4-kb and 3.5-kb mRNA, and in C to 3.5-kb mRNA. Blots for A and B were run in parallel on same gel; for C, membrane from B was reprobed after stripping. A and B blots were exposed for 24 h; C was exposed for 2 h. vMSG, variant major surface glycoprotein.

formed IFAs on intact rat *P. carinii* using the same affinity-purified epitope-specific antibody. A subpopulation of organisms reacted with the antibody (figure 5).

**Analysis of the features of vMSG.** The genes encoding vMSGs are rich in adenosine and thymidine (70% in v1MSG, 66% in v2MSG), similar to virtually all the genes of *P. carinii* that have been identified to date [17, 18, 33, 34]. Figure 1A shows the alignment of the deduced amino acid sequence of 2 v1MSG (v1MSG.9, 17) and 5 v2MSG clones (v2MSG.2, 7, 9, 15, 16) together with the published amino acid sequence of MSG clone GP3 [17]. The amino termini of v1MSG and v2MSG are nearly identical, but there is no homologous region in GP3 or the UCS of the MSG expression site (not shown). The initial codon of vMSGs appears to be the indicated methionine because the 5'-UTR of all the full-length cDNA clones in this study contains an in-frame stop codon at -9 (TAA) (figure 1B). Cysteine residues of v1MSG and v2MSG are numerous (5.4% in v1MSG; 5% in v2MSG), and the cysteine residues are highly conserved (figure 1A). In most cases, the cysteines were separated by 6 amino acids, as for MSG. The large number of cysteine residues would allow for complex intra- or interchain disulfide bonding. v1MSG.17 contains three potential *N*-linked glycosylation sites, and v2MSG.9 contains eight potential sites.

The predicted vMSGs have a hydrophobic amino terminus that has the characteristics of a signal peptide, which is necessary to translocate into endoplasmic reticulum for posttranslational processing. Using SignalP, the predicted signal peptide cleavage site is after amino acid 24 (TYG-NQ). The carboxy terminal of all vMSGs is nearly identical to that of GP3. A hydrophilicity profile of the deduced proteins demonstrates that this carboxyl terminus is hydrophobic, without a hydrophilic region downstream. The presence of a hydrophobic signal peptide and a hydrophobic carboxyl terminus is characteristic of proteins that have a glycosyl-phosphatidylinositol (GPI) anchor [35, 36].

**5'-UTR of vMSG.** We and others [22–24] have shown that there is only one expression site for MSG genes that encodes a conserved sequence and that regulates expression of different MSGs, presumably by a recombination event such as gene conversion. The sequence data in this study show that vMSG cDNAs do not encode this leader sequence, but all do encode a highly homologous signal peptide. As noted above, on the basis of Southern hybridization studies using oligonucleotide JK270, multiple copies of the signal peptide are present in the *P. carinii* genome. To see if a single copy or multiple copies of the vMSG 5'-UTR are present, Southern blot studies with oligonucleotides JK299 and JK298 were performed. Both oligonucleotides hybridized to multiple fragments of genomic DNA (figure 3B), indicating that, unlike MSG, multiple copies of the 5'-UTR and signal peptide of vMSG are present in the *P. carinii* genome.

**Discussion**

This study has demonstrated that two families of genes related to the MSG gene family are contained in the *P. carinii* genome and expressed by the organism. Because multiple different but related vMSGs are encoded by *P. carinii* and because the regulation of expression of these vMSGs appears to be different from that of MSG, expression of MSG-like proteins by *P. carinii* appears to be more complicated than had been previously assumed.

v1MSG genes are characterized by a deletion of ~1000 bp immediately prior to the carboxyl terminus, whereas v2MSG genes do not have this deletion. The first 100 amino acids of v1MSGs and v2MSGs, including the signal peptide, are highly conserved, as is the 5'-UTR.
Figure 3. Southern blot analysis of rat *P. carinii* DNA (20 μg/lane) digested with *Eco*RI (lane 1), *Bam*HI (lane 2), and *Xho*I (lane 3) and probed in panel *A* with JK269 (A), JK92 (B), and JK270 (C) and in panel *B* with JK298 (A) and JK299 (B). JK269 is specific for v1MSG, and JK270 and JK92 are specific for both vMSGs. JK298 and JK299 are specific for 5'-UTR of vMSG genes (figure 1B). All oligonucleotides were labeled with approximately equal numbers of counts. Blots were exposed for 48 h. Presence of multiple bands suggests that multiple copies of vMSGs are present in *P. carinii* genome. *C*, Southern blot of *P. carinii* chromosomes separated by pulsed-field gel electrophoresis and probed with JK269 (A) and JK270 (B), demonstrating hybridization with multiple *P. carinii* chromosomes and confirming that multiple copies are present in genome. Nos. at left are kbp. vMSG, variant major surface glycoprotein.

Overall, v1MSG and v2MSG are more closely related to each other than to MSG. Seventy-five percent of the amino acid residues of v1MSG.17 are conserved in v2MSG.9 versus 36% in GP3. Similarly, 45% of amino acid residues in v2MSG.9 are conserved in GP3 compared with 70% conservation between two MSGs (GP3 and GP10, GenBank accession no. U83323). The region deleted in v1MSG encodes a proline-glycine-rich as well as serine-threonine-rich region in MSGs. This serine-threonine-rich region is a candidate for *O*-linked glycosylation, as in *Saccharomyces cerevisiae*, although such glycosylation has not been demonstrated to date.

Similar to MSG genes, vMSG genes are distributed on multiple chromosomes of *P. carinii*, on the basis of oligonucleotide hybridization studies. While hybridization of the oligonucleotide probes to MSG genes cannot be excluded based on available sequence data, since all MSG genes have not been identified, the specificity of these probes for vMSG is supported by the cloning studies, in which only vMSG genes were identified despite the fact that 1% of clones in a cDNA library encode MSGs. For vMSG genes, like MSGs, *N*-linked carbohydrate residues account for ~10% of the apparent molecular weight. The identification of potential *N*-linked glycosylation sites in both v1MSG and v2MSG is consistent with this observation. Like MSG, the deduced amino acid sequence of vMSGs is predominantly hydrophilic, which is consistent with the demonstration by immunofluorescence that vMSGs are located on
A, Immunoblots of 2 samples of *P. carinii* purified from infected rat lungs (lanes 1 and 2) and reacted with affinity-purified anti-variant major surface glycoprotein (vMSG) peptide antibody. Two bands of 86 kDa and 115 kDa, corresponding to v1MSG and v2MSG, are seen in both preparations. B, Immunoblots of *P. carinii* antigens before (lane 1, 3) or after endoglycosidase F (lane 2) or endoglycosidase H (lane 4) treatment. Same first antibody was used as in A. C, Immunoblots of *P. carinii* antigens treated with lyticase. Lanes 1 and 4, whole organism extract; lanes 2 and 5, supernatant following lyticase solubilization; lanes 3 and 6, pellet following lyticase solubilization; lanes 1, 2, and 3, preimmune sera (1:2000); lanes 4, 5, and 6, hyperimmune sera (1:2000). Migration of molecular weight markers is indicated on left. Deglycosylation and lyticase solubilization each results in a loss in apparent M_r of ~10%.

The data from the current study demonstrate that vMSG genes are not regulated like MSG genes, by rearrangement of a gene in-frame at a site encoding a single leader sequence, but suggest that each vMSG gene contains the entire coding region. Whether expression is regulated by rearrangement of an entire vMSG gene downstream of one or a few promoters in a limited number of expression sites or by activation or suppression of expression at multiple sites throughout the genome, without genetic rearrangement, cannot be determined from this study. Such issues are difficult to address in an organism that cannot be cultured or cloned.

All MSG proteins have a lysine-arginine site immediately downstream of the conserved sequence encoded by the UCS. This site is a putative target for the recently identified kexin endoprotease homologue of *P. carinii* [40, 41]. Such processing would serve to eliminate the antigenically conserved leader in the mature protein, maximizing the ability of *P. carinii* to vary the MSG sequence. No similar site is present in vMSGs, but following elimination of the signal peptide, there is no region homologous to the leader, and no additional cleavage may be needed. Of note, the size of the mature MSG following leader sequence cleavage is virtually identical in size to v2MSG after the removal of the signal peptide. Given that nearly all the cysteines are conserved between v2MSG and MSG, it appears likely that the tertiary structure of the two proteins is very similar.

Antigenic variability of surface proteins is an important mechanism for the evasion of host defenses by a number of organisms, including *Trypanosoma, Borrelia,* and *Neisseria* species [36, 42–44]. MSG is an abundant surface exposed glycoprotein on *P. carinii* with the potential to interact with multiple host proteins and is a likely primary target of the immune system. It is also a potential target for therapeutic or prophylactic intervention. Similar to MSG, different MSG variants appear to be expressed on different *P. carinii* organisms, since...
Figure 5. Expression of specific epitope of variant major surface glycoprotein (vMSG) in rat-derived *P. carinii*, as evaluated by immunofluorescence using anti-vMSG-specific antibody. Double-exposed photography yields red fluorescence for organisms that react with concanavalin A–rhodamine only and yellow fluorescence for organisms that react with both concanavalin A–rhodamine and green donkey anti-rabbit fluorescein isothiocyanate. Both cyst and trophozoite forms appear to be recognized by vMSG-specific antipeptide antibody. No specific immunofluorescence was seen when preimmune sera were applied (not shown).

only a subset of organisms react with an antibody raised to an epitope that is not conserved among vMSGs. This would be consistent with the hypothesis that antigenic variation is utilized by *P. carinii* to evade host defenses. Antigenic profiles of *P. carinii* trophozoites and cysts have proved to be significantly different [45], and vMSG expression in *P. carinii* may represent stage-specific epitopes, although by immunofluorescence, it appears that both trophozoites and cysts react with an epitope-specific antibody. It is also possible that the system regulating vMSG expression represents the precursor to the very complex and precise regulatory mechanisms that have developed for MSG expression.

In summary, the sequence and molecular analysis information reported in this study has demonstrated novel MSGs that are closely related to but different from MSG. Further molecular and biochemical characterization of vMSGs should aid in understanding the role of MSG in the pathogenesis of *P. carinii* pneumonia.

References


