MSG Gene Cluster Encoding Major Cell Surface Glycoproteins of Rat Pneumocystis carinii

Miki WADA and Yoshikazu NAKAMURA*

Department of Tumor Biology, The Institute of Medical Science, The University of Tokyo,

P.O. Takanawa, Tokyo 108, Japan

(Received 29 June 1994)

Abstract

The MSG genes of Pneumocystis carinii encode major cell surface glycoproteins of multi-gene families and play crucial roles in the pathobiology of P. carinii pneumocystosis. The 11,370-bp chromosomal DNA encoding MSG was cloned and sequenced. It contained three open reading frames in tandem repeat, MSG99, MSG100 and ORF-3. MSG99 and MSG100 shared 39% protein sequence identity and belonged to the same MSG family, regardless of the internal ca. 1,000-bp deletion in MSG99. ORF-3 encodes a putative 47,820-Da protein and had no sequence similarities to MSG molecules or any known proteins. However, the ORF-3 protein was rich in proline residues and highly hydrophilic except for the C-terminal region, which seemed to be an anchoring sequence in the cell membrane. Northern and Southern blot hybridization analyses revealed a 3-kb ORF-3 transcript as well as multiple DNA fragments carrying the MSG and ORF-3 sequences. These results strongly suggested that diverse MSG genes are clustered in tandem repeats and these clusters involve ORF-3. The possible association of ORF-3 protein in the cell membrane is discussed.

Key words: Pneumocystis carinii; major surface glycoproteins; genomic MSG cluster

1. Introduction

Pneumocystis carinii is an opportunistic pathogen that often causes fatal pneumonia in immunosuppressed or immunodeficient patients with conditions or treatments such as AIDS, cancer chemotherapy or organ transplantation. Although the incidence of the disease has been increasing, little is known about the molecular pathobiology of P. carinii pneumocystosis. P. carinii is a eukaryotic microbe that is classified with the 'protista fungi' or 'fungi' group and infects many mammalian hosts. P. carinii derived from rats has an unusually abundant and highly immunogenic surface glycoprotein(s) named MSG (formerly called gp120 or PI15) with apparent masses ranging from 100 to 125 kDa. Several lines of evidence indicate that the MSG antigen play a crucial role in host-parasite interaction during P. carinii pneumocystosis. Gigliotti and Hughes found that passive immunization with an anti-MSG monoclonal antibody partly protects against the progression of P. carinii pneumonia in animal models. Fisher et al. demonstrated a specific T-cell response to MSG molecules after immunization with natural infection. Ezekowitz et al. showed that uptake of P. carinii by alveolar macrophages is mediated by the mannose receptor which interacts with the mannose moiety of MSG. Pottratz et al. and Haidaris et al. provided evidence that MSG participates in the attachment of the organism via interaction with fibronectin on alveolar epithelia and macrophages. Cloning of cDNAs encoding MSG polypeptides has been performed independently by three groups. Haidaris et al. have reported the partial cloning and sequence of a cDNA encoding a gpA antigen, analogous to MSG, of P. carinii from ferrets. Kovacs et al. and Wada et al. have reported the cloning and sequencing of cDNAs encoding the MSG polypeptides of rat P. carinii. A full-length MSG-cDNA, referred to as MSG5, has been recently cloned by Kitada et al. from a rat expression library of rat-derived P. carinii cDNAs. MSG5 encodes a 120,765-Da protein composed of 1,076 amino acids. Sequence analyses of cloned MSG-cDNAs revealed an MSG-gene family with ~70% protein sequence identity between subtypes. The family of MSG proteins is rich in cysteine residues and these cysteines are highly conserved in all MSG subtypes regardless of the sources with different species specificity, suggesting structural and/or functional importance. Previous analyses by genomic Southern, polymerase chain reaction (PCR) and karyotype hybridization suggested that MSG-gene-family members are scattered throughout most of the P. carinii chromosomes in tandem repeat structures.

In this study, we have cloned and analyzed the genomic
DNAs of rat *P. carinii*, which encode MSG antigens. The data clearly demonstrated that MSG genes are clustered in tandem repeats. Interestingly, a novel protein gene is also associated in the MSG cluster.

2. Materials and Methods

2.1. *P. carinii* organisms

*P. carinii* was cultivated in nude rats and the trophozoites were separated from broncho-alveolar lavage essentially as described previously.5

2.2. *P. carinii* libraries and screening

*P. carinii* DNA was isolated according to standard methods as described previously.13 A *P. carinii* genomic library was constructed by digesting *P. carinii* DNA with BamHI and ligating the DNA into the same site of λ EMBL3 DNA. The recombinant phage DNA was packaged with Gigapack II Gold packaging extract (Stratagene, La Jolla, CA) and plated on *E. coli* LE392 cells.15 The library was screened by plaque hybridization using the 2.4-kb *MSG1*-cDNA probe. DNA blot hybridization was conducted according to standard methods16 except that hybridization signals were detected by the nonradioisotope method using ECL Direct Nucleic Acid Labelling and Detection Systems (Amersham Laboratories, Buckinghamshire, UK) according to the manufacturer’s instructions. Eighteen positive clones were selected among 7,500 recombinant plaques. One of the positive clones, λMW111, was subjected to DNA sequence analysis.

2.3. RNA blot hybridization

RNA was isolated from *P. carinii* organisms using the RNA extraction kit (Amersham Laboratories) according to the manufacturer’s instructions as described previously.13,14 The RNA was subjected to agarose-gel electrophoresis in the presence of formaldehyde, and blotted onto Hybond™-N+ filters (Amersham Laboratories) as described by Sambrook et al.16 Hybridization probe was 620-bp *Xbal-Hin&11l* DNA fragment generated from pMW106 plasmid (a subclone derivative from AMW111; this paper), which carries an internal coding sequence of ORF-3 at positions 9245–9869 (Fig. 1). The filter was hybridized to the ORF-3 probe, washed according to standard procedures,16 and exposed to X-ray film.

2.4. Other DNA procedures

Double-stranded or single-stranded DNAs were sequenced by the dideoxynucleotide chain-termination method17 with fluorescent primers by an automatic DNA sequencer (A.L.F.™ DNA Sequencer; Pharmacia,

---

**Figure 1.** Nucleotide sequence of the MSG gene cluster. DNA and deduced amino acid sequence of the genomic MSG DNA cloned in λMW111 (single letter code) are shown. The number of the nucleotide position is counted from the 5’ end (BamHI) of the original clone. The first, second and third ORFs are referred to as *MSG99*, *MSG100* and ORF-3, respectively. *MSG99* carried an internal one-fourth deletion compared with the *MSG3* sequence at nucleotide positions ca. 2400–3400.

**Figure 2.** Hydropathicity profile of the ORF-3 protein. The DNA strider program based on that of Kyte and Doolittle18 was used. The characteristic protein domains are indicated.
Uppsala, Sweden). DNA blot hybridization and plaque hybridization were conducted according to standard methods except that hybridization signals were detected by the nonradioisotope method using ECL Direct Nucleic Acid Labelling and Detection Systems (Amersham Laboratories) according to the manufacturer's instructions.

3. Results and Discussion

3.1. Structure of the MSG gene cluster

To clarify the structure of genomic MSG genes, the genomic DNAs were cloned from a EMBL3 library of rat P. carinii DNA by plaque hybridization with the MSG-cDNA probe, and subjected to DNA sequence analysis. AMW111 contained an 11,370-bp P. carinii DNA and its nucleotide sequence carried three open reading frames (ORFs) in tandem repeat (Fig. 1; DDBJ/EMBL/GenBank accession no. D31909 and D17441). The first ORF encoded an 89,300-Da protein, composed of 774 amino acids, and shared 37% protein sequence identity (and 61% protein sequence similarity) with MSG5, thereby referred to as MSG99. However, MSG99 was 326 amino acids shorter than MSG5 in the C-terminal region.

The second ORF contained 1,083 amino acids, with a deduced mass of 120,921 Da. The protein shared 76% protein sequence identity with MSG5, thereby designated MSG100, but not identical to any MSG sequences so far reported or analyzed, thus providing again a new member of the MSG family. It carried an entire coding sequence analogous to MSG5. Analysis of the hydropathicity profile demonstrates a single hydrophobic region common to the MSG5 protein at the C-terminus, which may be necessary for membrane anchorage of this surface protein. MSG99 and MSG100 are spaced for 311 bp, which was consistent with our previous findings that 350-500 bp DNA fragments can be amplified by PCR using primers designed to detect DNA segments between adjacent MSG genes.13

It was not immediately obvious if the MSG99 and MSG100 genes are expressed in P. carinii. However, a class of cDNAs whose sequences are highly homologous to the coding sequence of MSG99 and its 5′ untranslated
Figure 5. DNA blot hybridization of the *P. carinii* MSG gene clusters. Fifteen *P. carinii* genomic DNAs cloned independently into λ EMBL3 (Fig. 1) were digested with *Bam*H1/EcoRI, fractionated through 1.0% agarose, and subjected to DNA blot hybridization using probes specific for *MSG1* (A) or *ORF-3* (B) as described in Materials and Methods and in Fig. 4. The same filter blot was used for hybridization to the second probe after washing. Genomic clones: 1, AMW101; 2, AMW102; 3, AMW103; 4, AMW106; 5, AMW107; 6, AMW111; 7, AMW112; 8, AMW113; 9, AMW114; 10, AMW115; 11, AMW116; 12, AMW117; 13, AMW118; 14, AMW119; 15, AMW120.

region (5' UTR) sequence, was isolated from the cDNA library (unpublished data). Interestingly, the cDNA sequence revealed a 42-bp intron located at the N-terminal of *MSG99* (Fig. 1). On the other hand, an MSG-cDNA species which carries a 5' UTR sequence homologous to the spacer sequence between *MSG99* and *MSG100* has not been detected, regardless of the fact that *MSG100* is highly homologous to *MSG5*-cDNA. These findings suggest that *MSG99* is expressed while *MSG100* is silent or, alternatively, splicing or recombination of diverse segments may generate a chimeric coding sequence from these two genes.

### 3.2. A novel gene associated in the MSG cluster

The third ORF, *ORF-3*, was located 2,789-bp downstream of *MSG100*, and encoded a potential 47,820-Da protein. It shared no sequence homologies with MSG and any known protein sequences in the data base. However, the predicted protein has several characteristic features: it is rich in proline residues and is highly hydrophilic except for the C-terminus (Fig. 2). There is also a C-terminal region rich in serine and threonine residues (ST-rich domain). Northern blot hybridization detected a 3.0-kb RNA, which is approximately 1.5-kb larger than the minimum coding capacity for *ORF-3* (Fig. 3). We assumed that the *ORF-3* gene is not highly expressed compared with MSG because the intensity of the *ORF-3* transcript was much lower than that of the MSG transcript appeared in the same filter blot (data not shown).

The *ORF-3* DNA probe hybridized to multiple restriction fragments in Southern blot hybridization (Fig. 4B). Similarly complex hybridization profiles were found by probing with the *MSG* cDNA (Fig. 4A). These results
suggested that ORF-3 is associated in a diverse and multiple gene family and, presumably, localized within the MSG clusters. In fact, of the 15 independent genomic MSG DNAs so far tested, all carried ORF-3 (Fig. 5). Most of the cloned DNAs showed different restriction and hybridization patterns except for the three clones, AMW111, 112 and 114 (Fig. 5, lanes 6, 7 and 9), which seemed to be identical. Taking these results into consideration, we concluded that chromosomes MSG clusters involved ORF-3 genes and are highly diverse. The hydrophobic anchor at the C-terminus suggests that the ORF-3 protein may be membrane-bound.

The present work clearly demonstrated the presence of multiple MSG genes in tandem repeat on the P. carinii chromosomes. These findings are consistent with our previous PCR data which suggested MSG gene clusters, and again emphasize the potential for antigenic variability of the organism and how this variability is controlled at the level of gene expression.

Acknowledgments: We thank Kazuhiro Kitada for preparation of P. carinii organism, and Koichi Ito for helpful discussion and reading of the manuscript. This work was supported by grants from The Ministry of Education, Science and Culture, Japan, Uehara Memorial Foundation, Tokyo Biochemical Research Foundation, and Japan Health Sciences Foundation.

References

1. Selik, R. M., Starcher, E. T., and Curran, J. W. 1987, Opportunistic diseases reported in AIDS patients: frequencies, associations, and trends, AIDS, 1, 175-182.
2. Edman, J. C, Kovacs, J. A., Masur, H., Santi, D. V., Elwood, H. J., and Sogin, M. L. 1988, Ribosomal RNA sequence shows Pneumocystis carinii to be a member of the Fungi, Nature, 334, 519-522.
3. Watanabe, J., Hori, H., Tanabe, K., and Nakamura, Y. 1989, Phylogenetic association of Pneumocystis carinii with the 'Rhizopoda/Myxomycota/Zygomycota group' indicated by comparison of SS ribosomal RNA sequences, Mol. Biochem. Parasitol., 32, 163-168.
4. Radding, J. A., Armstrong, M. Y. K., Ullu, E., and Richards, F. F. 1989, Identification and isolation of a major cell surface glycoprotein of Pneumocystis carinii, Infect. Immun., 57, 2149-2157.
5. Tanabe, K., Takasaki, S., Watanabe, J., Kobata, A., Egawa, K., and Nakamura, Y. 1989, Glycoproteins composed of major surface immunodeterminants of Pneumocystis carinii, Infect. Immun., 57, 1363-1368.
6. Gigliotti, F. and Hughes, W. T. 1988, Passive immunoprophylaxis with specific monoclonal antibody confers partial protection against Pneumocystis carinii pneumonitis in animal models, J. Clin. Invest., 81, 1666-1668.
7. Fisher, D. J., Gigliotti, F., Zauderer, M., and Harmsen, A. G. 1991, Specific T-cell response to a Pneumocystis carinii surface glycoprotein (gp120) after immunization and natural infection, Infect. Immun., 59, 3372-3376.
8. Ezekowitz, R. A. B., Williams, D. J. et al. 1991, Uptake of Pneumocystis carinii mediated by the macrophage mannose receptor, Nature, 351, 155-158.
9. Pottratz, S. T., and Martin II, W. J. 1990, Mechanism of Pneumocystis carinii attachment to cultured rat alveolar macrophages, J. Clin. Invest., 86, 1678-1683.
10. Pottratz, S. T., Paulsrud, J., Smith, J. S., and Martin II, W. J. 1991, Pneumocystis carinii attachment to cultured lung cells by pneumocystis gp120, a fibronectin binding protein, J. Clin. Invest., 88, 403-407.
11. Haidaris, P. J., Wright, T. W., Gigliotti, F., and Haidaris, C. G. 1992, Expression and characterization of a cDNA clone encoding an immunodeterminant surface glycoprotein of Pneumocystis carinii, J. Infect. Dis., 166, 1113-1123.
12. Kovacs, J. A., Powell, F., Edman, J. C. B. et al. 1993, Multiple genes encode the major surface glycoprotein of Pneumocystis carinii, J. Biol. Chem., 268, 6034-6040.
13. Wada, M., Kitada, K., Saito, M., Egawa, K., and Nakamura, Y. 1993, cDNA sequence diversity and genomic clusters of major surface glycoprotein genes of Pneumocystis carinii, J. Infect. Dis., 168, 979-985.
14. Kitada, K., Wada, M., and Nakamura, Y. 1994, Multigene family of major surface glycoproteins of Pneumocystis carinii: full-size cDNA cloning and expression, DNA Res., 1, 57-66.
15. Frischau, A. M., Lehrach, H., Pousta, A., and Murray, N. 1983, Lambda replacement vectors carrying polylinker sequences, J. Mol. Biol., 170, 827-842.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
17. Sanger, F., Nicklen, S., and Coulson, A. R. 1977, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
18. Kyte, J. and Doolittle, R. F. 1982, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol., 157, 105-132.