Available online at www.sciencedirect.com
ScienceDirect
Fungal Genetics and Biology 44 (2007) 905–919
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Pneumocystis murina MSG gene family and the structure of the locus associated with its transcription

Scott P. Keely [a], Michael J. Linke [b,c], Melanie T. Cushion [b,c], James R. Stringer [a,*]

[a] *Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati College of Medicine, Cincinnati, OH 45220, USA*
[b] *Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, OH 45220, USA*
[c] *Department of Veterans Affairs Medical Center, Research Service, Cincinnati, OH 45220, USA*

Received 20 October 2006; accepted 3 January 2007
Available online 10 January 2007

Abstract

Analysis of the *Pneumocystis murina* MSG gene family and expression-site locus showed that, as in *Pneumocystis carinii*, *P. murina* MSG genes are arranged in head-to-tail tandem arrays located on multiple chromosomes, and that a variety of MSG genes can reside at the unique *P. murina* expression site. Located between the *P. murina* expression site and attached MSG gene is a block of 132 basepairs that is also present at the beginning of MSG genes that are not at the expression site. The center of this sequence block resembles the 28 basepair CRJE of *P. carinii*, but the block of conserved sequence in *P. murina* is nearly five times longer than in *P. carinii*, and much shorter than in *P. wakefieldiae*. These data indicate that the *P. murina* expression-site locus has a distinct structure.

Keywords: Pneumocystis; Mouse; Antigen; Variation; Gene family; Gene expression

1. Introduction

The fungal genus *Pneumocystis* contains multiple species including the causative agent of human *Pneumocystis* pneumonia, which afflicts individuals with impaired immune system function, such as Acquired Immunodeficiency Syndrome (AIDS) patients (Thomas and Limper, 2004). Studying *Pneumocystis* organisms is difficult because they do not proliferate well in culture (Walzer et al., 2001). Therefore, animal models have been the main source of organisms (Armstrong and Cushion, 1994; Dei-Cas et al., 1998; Larsen et al., 2002).

Pneumocystis murina, the species of *Pneumocystis* found in laboratory mice (Keely et al., 2004), is of interest because the laboratory mouse provides an advanced animal model of host response to *Pneumocystis* infection (Wright et al., 2001; Lund et al., 2003; An et al., 2003; Qureshi et al., 2003; McAllister et al., 2004; Empey et al., 2004; Qureshi et al., 2005; Linke et al., 2005, 2006a,b). *P. murina* is a close relative of *Pneumocystis carinii* and *Pneumocystis wakefieldiae*, both found in rats, and a more distant relative of the human pathogen, *P. jirovecii* (Frenkel, 1976, 1999; Stringer et al., 2001, 2002; Keely and Stringer, 2005; Redhead et al., 2006). All four of these species feature an abundant surface protein called Major Surface Glycoprotein (MSG) (see Table 1), which has also been observed in *Pneumocystis* from ferrets, where it is known as gpA (Linke et al., 1989; Tanabe et al., 1989; Haidaris et al., 1991, 1992, 1998; Nakamura et al., 1991; Gigliotti, 1992; Stringer et al., 1993; Garbe and Stringer, 1994; Wright et al., 1994; Kovacs et al., 1993, 1994; Wright et al., 1995).

Studies on other *Pneumocystis* species, primarily *P. carinii*, suggest that *P. murina* may use a family of MSG genes to produce antigenic variation in populations of the fungus dwelling in mice. *P. carinii* MSG is encoded by a multigene family, members of which are arranged as head-to-tail repeats located near the telomeres of all 17 chromosomes (Wada et al., 1993; Kovacs et al., 1993; Kitada et al., 1994; Wada et al., 1995; Edman et al., 1996; Stringer and Keely, 2001; Cornillot et al., 2002; Stringer, 2003, 2005; Keely et al., 2005).

* Corresponding author. Fax: +1 513 558 8474.
E-mail address: stringjr@ucmail.uc.edu (J.R. Stringer).

1087-1845/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2007.01.004

Table 1
Abbreviations referring to *P. murina* MSG gene family and its expression site

Abbreviation	Meaning	Defining characteristics
MSG	Major Surface Glycoprotein	The predominant protein found on the surface of *Pneumocystis* organisms. MSGs are approximately 120 kDa, are glycosylated and cysteine rich. There are many different isoforms of MSG, each encoded by a member of the MSG gene family.
UCS	Upstream Conserved Sequence	The conserved sequence found at the 5-prime end of every messenger RNA molecule encoding an MSG protein.
CRJE	Conserved Recombination Junction Element	The conserved sequence found at the beginning of every open reading frame encoding an MSG protein.
ES	Expression Site	The locus to which an MSG gene must be attached in order to be transcribed. The ES includes the UCS locus and a presumptive promoter.

All indications are that only one MSG gene is expressed per *P. carinii* organism. Control of *P. carinii* MSG expression involves a single-copy locus called the expression site, or ES (Table 1). The majority of organisms in *P. carinii* populations appear to be haploid (Wyder et al., 1998). Because only one MSG gene can occupy the expression site at a time, the expression site system restricts transcription of the family to one gene per organism (Wada et al., 1995; Edman et al., 1996; Sunkin and Stringer, 1996, 1997; Keely et al., 2003). A large number of different MSG genes have been observed at the *P. carinii* expression site, suggesting that MSG genes can be moved to this locus by DNA recombination (Sunkin and Stringer, 1997; Keely et al., 2003). The mechanism of recombination is not known, but the possibility of a site-specific recombinase has been raised by the presence of a 28 basepair sequence called the Conserved Recombination Junction Element (CRJE) (Table 1). There is a copy of the CRJE at the junction between the expression site and adjacent MSG coding region. In addition to this copy, every MSG gene not at the expression site begins with a copy of the CRJE. Thus it is possible that recombination between the expression site and MSG genes at other loci is mediated by interactions between two copies of the CRJE. Whatever the mechanism, switching the MSG gene at the expression site discontinues expression of the previous resident MSG gene and activates expression of the new expression-site-linked MSG gene. Studies with antibodies that bind to few isoforms of MSG have shown that the protein encoded by the MSG gene at the expression-site locus is present in or on *P. carinii* cells (Schaffzin and Stringer, 2004). Other studies using antibodies detected antigen variation between clusters of *P. carinii* in a rat lung (Angus et al., 1996). Hence it would appear that *P. carinii* populations in the lung develop antigenic variation by switching the MSG gene that is at the expression site.

The *P. carinii* expression site encodes most of the 410 basepair Upstream Conserved Sequence (UCS) (Table 1), so named because it was found at the five prime ends of mRNAs encoding diverse MSGs (Wada et al., 1995; Edman et al., 1996; Sunkin and Stringer, 1996, 1997; Sunkin et al., 1998; Stringer and Keely, 2001; Kutty et al., 2001; Keely et al., 2003; Schaffzin and Stringer, 2004; Stringer, 2003, 2005). The last 28 basepairs of the UCS are encoded by the CRJE. The UCS serves as the translation initiation site for production of an MSG precursor peptide (Sunkin et al., 1998). The peptide encoded by the UCS appears to serve to send a precursor of mature MSG into the secretory pathway for transport to the cell surface (Sunkin et al., 1998). The MSG on the surface lacks the amino acids encoded by the UCS, suggesting that these are removed by proteolysis in the endoplasmic reticulum (Sunkin et al., 1998).

Given the similarities between species in the genus *Pneumocystis*, *P. murina* would be expected to contain a family of MSG genes that are expressed via an expression-site mechanism. However, this species might also be expected to have its own properties in this regard. Therefore, structural analysis of the expression-site locus of *P. murina* was necessary. Previously, a sequence that is approximately 62% identical to the *P. carinii* UCS was described at the 5′ end of a *P. murina* mRNA encoding an MSG (Haidaris et al., 1998). Part of this sequence was shown to map to one chromosome of *P. murina*, suggesting that there is a single expression-site locus in this species (Haidaris et al., 1998). The data presented herein characterize the locus encoding the *P. murina* UCS and provide more information about the MSG gene family in this species. These data show that the UCS locus of *P. murina* is present once per haploid genome and that a population of *P. murina* organisms can contain different MSG genes attached to the UCS locus. Between the genomic sequence encoding the UCS and the attached MSG gene, there is a sequence resembling the 28 bp CRJE of *P. carinii*, but sequence conservation extends substantially beyond the 28 bp CRJE-like sequence to encompass a block nearly five times longer than the CRJE of *P. carinii*. Multiple *P. murina* MSG genes were detected by hybridization, PCR and cloning. All but one chromosome contained MSG-related sequences. Head-to-tail tandem arrays of *P. murina* MSG genes were detected by PCR but no evidence of other tandem arrangements emerged. Sequence analysis detected 26 different MSG genes, which is only about one third as many as the number of MSG genes in *P. carinii*. Quantitative real-time PCR supported the hypothesis that the *P. murina* MSG gene family is smaller than the *P. carinii* MSG gene family.