Expression of the *Pneumocystis carinii* major surface glycoprotein epitope is correlated with linkage of the cognate gene to the upstream conserved sequence locus

Joshua K. Schaffzin† and James R. Stringer

Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA

Correspondence
James R. Stringer
stringjr@ucmail.uc.edu

The major surface glycoprotein (MSG) is a variable surface antigen of the pathogenic fungus *Pneumocystis carinii*. Many forms of MSG are encoded by a gene family. Expression of the MSG gene family is believed to be controlled in a *cis*-dependent fashion. Transcription of a given MSG gene is correlated with linkage of that gene to a unique locus called the upstream conserved sequence (UCS). These data predict that the MSG protein on a given organism will match that encoded by the MSG gene at the UCS locus in that organism. To test this hypothesis, a monoclonal antibody (mAb) that recognizes a small number of MSG isoforms was identified, and the DNA sequence encoding the mAb epitope (epitope-encoding sequence, EES) was determined. Western blotting, immunofluorescence and DNA hybridization showed that expression of the mAb epitope was associated with the presence of the EES at the UCS locus. Correlation of epitope expression and UCS linkage supports the hypothesis that expression of a particular MSG on the surface requires UCS linkage of the gene encoding it.

INTRODUCTION

Pneumocystis carinii is a yeast-like fungus that infects rat lung. *P. carinii* propagates in culture to a limited extent (Cushion, 1997). Thus, studies requiring large numbers of *P. carinii* must use organisms purified from fulminantly infected rats, which can be produced by chemically suppressing immune function.

P. carinii organisms express a surface antigen known as the major surface glycoprotein (MSG) (Walzer & Linke, 1987; Kovacs *et al.*, 1988; Linke *et al.*, 1989; Tanabe *et al.*, 1989; Linke & Walzer, 1989, 1991; Nakamura, 1998; Stringer & Keely, 2001). Multiple forms of MSG are encoded by a family of approximately 100 genes, members of which are located at 34 telomeric sites (Stringer *et al.*, 1991; Kovacs *et al.*, 1993; Sunkin *et al.*, 1994; Sunkin & Stringer, 1996, 1997; Stringer & Keely, 2001; Cornillot *et al.*, 2002; Keely

et al., 2003). Other species of *Pneumocystis*, including the species that infects humans, also have families of MSG genes (Haidaris *et al.*, 1991; Stringer *et al.*, 1993; Garbe & Stringer, 1994; Wright *et al.*, 1995; Mei *et al.*, 1998; Haidaris *et al.*, 1998; Schaffzin *et al.*, 1999; Kutty *et al.*, 2001).

We hypothesize that only one *P. carinii* MSG gene is expressed at a time in a given organism and that this control is exerted *in cis* by a unique locus named the upstream conserved sequence (UCS). The UCS encodes a tract of 365 ribonucleotides that is found at the 5′ ends of mRNAs encoding different MSG forms (Wada *et al.*, 1995; Edman *et al.*, 1996). The UCS RNA leader sequence encodes a peptide that served as a signal sequence for transport of nascent protein into the secretory apparatus when expressed in insect cells, suggesting that one role of the UCS is to direct MSG precursors into the secretory pathway (Sunkin *et al.*, 1998). Epitopes on the UCS-encoded peptide are found on MSG in whole *P. carinii* extracts, but not on MSG extracted from the *P. carinii* cell wall, suggesting that MSG is synthesized as a preprotein beginning with the UCS and that the UCS is removed proteolytically to produce the mature form of MSG found on the cell surface (Sunkin *et al.*, 1998). Thus, the UCS appears to be necessary for MSG translation and surface expression, but not for function of mature MSG proteins.

Previous studies on populations of *P. carinii* obtained from different rats showed that each population contained

---

transcripts representing different subsets of the MSG gene family. These differences allowed experiments showing that MSG genes that are linked to the UCS locus are represented as UCS-containing mRNAs, suggesting that linkage to the UCS locus is necessary and sufficient for transcription of an MSG gene (Sunkin & Stringer, 1996, 1997).

If linkage of an MSG gene to the UCS locus indeed controls surface expression of the protein encoded by that gene, then *P. carinii* populations rich in a particular MSG protein should contain many organisms that have the DNA sequence encoding that protein at the UCS locus. To test this hypothesis, we developed probes that identify a specific MSG sequence at both the protein and DNA levels. To detect specific MSG proteins, a monoclonal antibody (mAb), RA-C11, was used (Linke *et al.*, 1998). RA-C11 has been shown to recognize very few of the peptides encoded by the MSG gene family, suggesting that it recognizes an epitope that is not present in most forms of MSG. The evidence suggesting that the RA-C11 epitope is rare came from two sets of experiments. The first experiments used Western blotting to show that mAb RA-C11 reacted with MSG from only 25 % of populations examined (Vasquez *et al.*, 1996). The second set of experiments used an expression library to show that mAb RA-C11 reacted with only 1 % of the λgt11 plaques bearing MSG epitopes (Linke *et al.*, 1998).

The work described herein began by determining the sequence of the DNA that encodes the RA-C11 epitope (i.e. the epitope-encoding sequence, EES). Next, six populations of *P. carinii* that differed with respect to expression of RA-C11-reactive proteins were identified. Analysis of the fractions of organisms with the RA-C11 EES at the UCS showed that this parameter correlated directly with the abundance of RA-C11-reactive protein. These results support the hypothesis that residence of a particular MSG gene at the UCS locus is necessary and sufficient for expression the MSG protein encoded by that gene.

METHODS

Epitope identification and mapping. A *P. carinii* genomic library was constructed in λgt11 by standard methods (Sambrook *et al.*, 1989; Linke *et al.*, 1998). Approximately 3x10[6] plaques were screened with the RA-C11 mAb, and positive plaques were isolated as described previously (Sunkin *et al.*, 1994; Garbe & Stringer 1994; Schaffzin, 1999). For analysis of proteins encoded by a λgt11 clone, crude lysates were generated from bacteria (*Escherichia coli* Y1090r⁻) lysogenic for the phage (Sambrook *et al.*, 1989). Lysogenic strains were propagated in broth cultures for 12–18 h, then diluted 1:100 and grown at 32 °C to an OD_{600} of 0·5; at this point the cultures were switched to 44 °C to induce phage production. After 20 min, IPTG was added to achieve a final concentration of 10 mM, and the cultures were incubated for 1 h at 37 °C (Sambrook *et al.*, 1989). Cells were collected by centrifugation and resuspended in 0·1 ml extract buffer (50 mM Tris/HCl pH 7·5, 1 mM EDTA, 1 mM DTT, 1 mM PMSF) and stored at −20 °C. For separation of proteins on gels (8–16 % Tricine SDS-PAGE gels from Novex), cells were lysed by boiling them for 2 min in SDS-PAGE loading buffer (62·5 mM Tris/HCl pH 6·8, 10 %, v/v, glycerol, 2 % SDS, 0·0025 %

bromophenol blue, 2·5 % β-mercaptoethanol). Separated proteins were transferred to membranes, which were tested for antibody reactivity by standard methods as described previously (Linke *et al.*, 1998).

To map the RA-C11 epitope, overlapping fragments of the clone C11-S5 MSG gene were generated using PCR and subcloned into the pET30 expression vector (Novagen). A detailed description of the primers and conditions used in the PCR experiments can be found elsewhere (Schaffzin *et al.*, 1999). Inserts in recombinant plasmids were sequenced to confirm correct amplification and reading frame. Recombinant plasmids were established in strain BL21(DE3) (Novagen). Production of recombinant proteins was induced by adding IPTG (final concentration, 1 mM) to bacterial cultures. After 2 h at 37 °C, cells were collected by centrifugation, resuspended in SDS-PAGE loading buffer, boiled at 100 °C for 2 min and immediately loaded onto a 10–20 % Tricine SDS-PAGE gel (Novex). Immunoblots were prepared by standard methods (Sambrook *et al.*, 1989). Proteins produced carried an S protein tag encoded by the Pet30 vector (Novagen), so proteins of interest could be located on immunoblots using an S-tag reagent purchased from Novagen. This reagent carries an alkaline phosphatase conjugated to S protein (AP-S), which binds the S-tag. The AP-S reagent was used according to the instructions provided by the vendor. Antibodies bound to immunoblots were visualized using secondary antibodies conjugated to either AP or horseradish peroxidase. Standard methods were employed in these experiments (Sambrook *et al.*, 1989).

Measurement of the fraction of UCS-MSG junctions that contain the C11 EES. UCS-MSG junctions were obtained by amplifying *P. carinii* genomic DNA. DNA was subjected to PCR using the UCS1 primer (5′-GTTGTCCGAAGAGGTGGCA-3′), paired with MSG primer C7 (5′-GTCTTGTCCCTTTTTATAGCA-3′) under the following conditions: 95 °C for 5 min, 35 cycles of incubation at 95 °C for 60 s, 48 °C for 2 min and 72 °C for 90 s, followed by 72 °C for 10 min.

The fraction of amplified UCS-MSG junctions that contain the C11 EES was measured using an allele-specific hybridization method adapted from that described by Kren *et al.* (1997). PCR amplicons were attached to nylon membranes, and incubated with two radiolabelled oligonucleotides in succession. First, the filter was incubated with the C11 EES oligonucleotide (5′-TTGAGGGAGAAATGTTACGAATTG-AAGCG-3′) for 18 h at 50 °C in 2× SSC, 1 % SDS, 5× Denhardt's solution, and 0·2 mg ml⁻¹ sheared salmon sperm DNA (Sambrook *et al.*, 1989). Filters were washed in 1× SSPE, 0·5 % SDS at 50 °C for 30 min and immediately washed in tetramethylammonium chloride (TMAC) buffer (50 mM Tris/HCl pH 8·0, 2 mM EDTA pH 8·0, 0·1 % SDS, 3 M TMAC) at 70 °C for 1 h (Kren *et al.*, 1997). Radioactive emissions were detected by exposing the membrane to phosphor-imaging plates and reading the signal in a PhosphorImager followed by analysis with ImageQuant software (Molecular Dynamics, Amersham Biosciences). Signals obtained using the C11 EES probe were corrected for any differences in the amount of target DNA on the filter by comparing them to signals obtained with the CRJE oligonucleotide (5′-ATGGCACGGCCGGTTAAGAG-3′). This oligonucleotide corresponds to 20 of the nucleotides in the 28 bp sequence known as the conserved recombination junction element (CRJE) (Wada *et al.*, 1995). The CRJE is present between the UCS and the MSG-encoding regions of all UCS-linked MSG genes (Keely *et al.*, 2003). To perform CRJE-hybridization, the filter was stripped of the C11 EES probe by boiling in 1 % SDS for 15 min. Absence of the C11 EES probe was confirmed by exposing the membrane to phosphor-imaging plates. The stripped filter was then incubated with the CRJE probe at 42 °C in 6× SSC, 0·1 % SDS, 1× Denhardt's solution, 0·08 mg ml⁻¹ sheared salmon sperm DNA. Filters were washed at 65 °C in 0·5× SSC, 0·1 % SDS twice for 10 min. Hybridization experiments were performed in triplicate.

The C11 EES probe was shown to be specific for the target sequence by the following experiments. First, DNA molecules with a sequence known to match 29 of the 30 nucleotides in the reverse complement of the C11 EES oligonucleotide were attached to nylon membranes and tested for hybridization to the C11 EES oligonucleotide. Prior to washing in TMAC, the mismatched hybrids were detected, but the TMAC washing procedure eliminated these. By contrast, TMAC washing did not remove the signal from a control filter carrying DNA that matched the C11 EES exactly. Second, a library of plasmids carrying hundreds of different MSG genes was screened by hybridization to the C11 EES oligonucleotide. Plasmids from 10 positive colonies were subjected to DNA sequence analysis, which showed that all of them carried an exact copy of the C11 EES.

Immunoblot detection of *P. carinii* antigens. *P. carinii* organisms were extracted from the lungs of infected rats and enumerated by light microscopy as previously described (Cushion *et al.*, 1988, 1998a). For protein separation, samples containing the same number of organisms were boiled in SDS-PAGE loading buffer for 2 min and insoluble material was removed by centrifugation. SDS-soluble material was diluted by a series of six twofold dilutions in SDS-PAGE loading buffer, boiled for 1 min, and loaded onto an 8–16 % Tricine SDS-PAGE gel (Novex). Duplicate gels and Western blots were prepared and tested for reactivity with either mAb RA-C11 or mAb RA-F1 (Linke *et al.*, 1998). Signals obtained with mAb RA-F1 were used to normalize the RA-C11 signals. Developed blots were scanned digitally and the relative intensity of the 116 kDa MSG band in each lane was evaluated using ImageQuant software. The relationship between sample dilution factor and intensity of staining with mAbs was examined and this relationship generally appeared to be linear. However, linearity could not be firmly established in samples with relatively little mAb RA-C11 reactivity because it was not possible to detect bands at more than two dilutions.

Immunofluorescence. *P. carinii* organisms that had been cryogenically preserved in DMSO freezing medium (7·5 %, DMSO) were diluted in 1× PBS and passed thrice through a 10 µm filter (Millipore) (Cushion *et al.*, 2001). Organisms were prepared for immunofluorescence assay by methods described by Broomall *et al.* (1998). Briefly, organisms were applied to glass slides and fixed by air-drying followed by immersion in ice-cold acetone for 5 min. Slides were stored at −80 °C. Slides were thawed in room-temperature acetone and rehydrated in 1× PBS, then incubated in a moist chamber at 37 °C for 30 min with mAb diluted 1:40 in PBS (Sambrook *et al.*, 1989), after which they were washed in PBS containing 0·05 % Tween 20 twice for 10 min at room temperature. Slides were next incubated with FITC-conjugated goat-anti-mouse IgG (1:80 dilution, Biomeda), and washed as above. Organisms were counterstained by incubation in 0·1 mg ml^{-1} rhodamine-conjugated concanavalin A (ConA, Sigma). Control slides were incubated with either FITC goat-anti-mouse IgG plus ConA, or goat IgG alone. Organisms were visualized using a fluorescence microscope under oil immersion (1000× magnification).

Statistical analysis. Analysis of variance (ANOVA) was performed on the data from hybridization, Western blotting and immunofluorescence experiments. The Tukey procedure was used to test for significance between pairs of populations. Correlation coefficients were calculated and evaluated for significance using a one-tailed *t*-test (Stevens, 1990).

GenBank accession numbers. Sequences from previous studies were obtained from GenBank using the following accession numbers: MSGB, L29497; MSG 100, D31909; MSG 101, D31910; MSG 102 and MSG 104, D31911; MSG 103, D31912; MSG 105, D82031; GP3, L05906; PC3, L05904; PC5, L05907; C10, U83323; MSG 1, D17438; MSG 2, D17439; MSG 3, D17440; 3G5, AL592382; 22C8, AL592263.

Sequences obtained in the current study were submitted to GenBank under the following accession numbers: S5, AF164561; 60A9, AF169405; J3, AF169406.

RESULTS

RA-C11 epitope

A λgt11 expression library constructed previously (Linke *et al.*, 1998) was screened for plaques that bound mAb RA-C11. The screen yielded clone C11-S5, the insert of which contained a 2982 bp ORF in-frame with the *lacZ* gene in the λgt11 vector. This ORF encoded numerous peptide motifs common to MSG molecules (data not shown). These data indicated that the C11-S5 insert comprised the last 90 % of an MSG gene.

To identify the RA-C11 epitope-encoding sequence (EES) in the C11-S5 insert, PCR was used to generate five overlapping segments of the insert, which were each inserted into a plasmid expression vector. Recombinant proteins in lysates from bacteria carrying these plasmids were tested by Western blotting for mAb RA-C11 reactivity (Fig. 1A). As expected, only some of the plasmids produced a band of RA-C11 reactivity (see lanes 11 and 12). Positive control experiments using alkaline phosphatase-conjugated S-protein (AP-S) showed that lack of mAb RA-C11 reactivity was not due to failure of some plasmids to produce a fusion protein (lanes 1–5). The fusion proteins were also tested for reactivity with mAb RA-F1, which is an antibody to a defined epitope that is known to be present on a broad spectrum of MSGs (Linke *et al.*, 1998). Two fusion proteins reacted with mAb RA-F1 (Fig. 1A, lanes 7 and 8). The reactive proteins covered the region where the mAb RA-F1 epitope was previously mapped (Linke *et al.*, 1998).

Further mapping experiments (Fig. 1B) localized the RA-C11 epitope to a 12 amino acid sequence (SLREKCY-ELKRE) that lies upstream of the mAb RA-F1 epitope (Fig. 1C). Additional immunoblot experiments were performed using recombinant proteins from a second MSG gene, 60A9, which contained ELREKCYELKRK. This protein was recognized by mAb RA-C11 (Fig. 2). These data implicated the 10 amino acids shared by S5 and 60A9 (LREKCYELKR) in mAb RA-C11 binding.

To determine the frequency of LREKCYELKR in the MSG protein family, the GenBank database was searched. Nineteen sequences, all from different MSG genes, were identified (MSG100, PC5, MSG1, C10, 22C8, MSGB, MSG104, MSG103, MSG102, MSG5, MSG101, 3G5.3, 3G5.2, MSG2, GP3, PC3, MSG105, 3G5.1, MSG3). Another six MSG gene sequences (21H1.2, 18A9, 11A11.2, 11A11.1, 21H1.1, 11A11.3) were obtained from the Sanger Centre internet site (http://www.sanger.ac.uk/Projects/P_carinii). Only one of these 25 MSG genes (MSG100) encoded LREKCYELKR (Fig. 3). The other 24 genes encoded one of three other peptide sequences, LREgCYELKR,
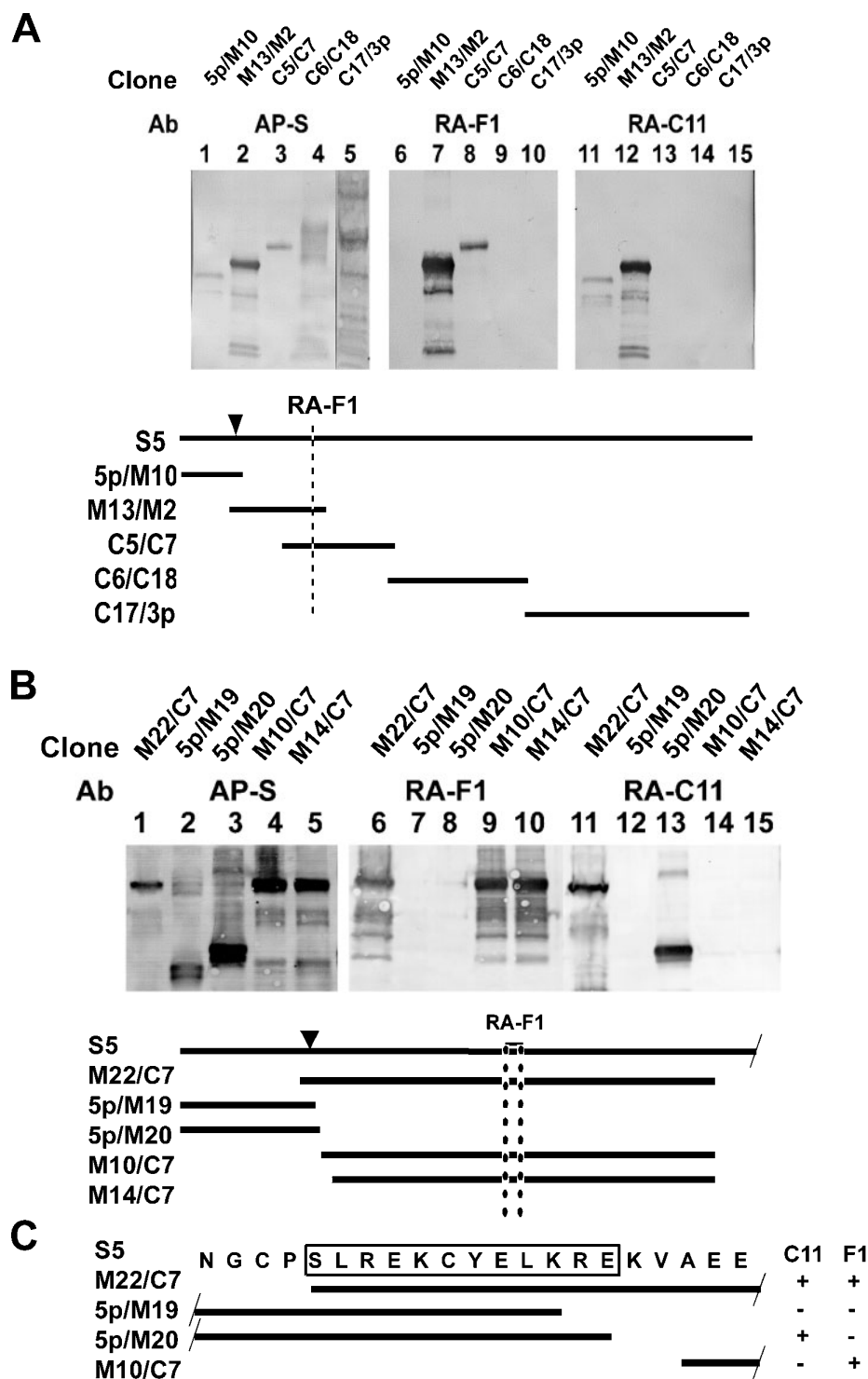
Fig. 1. Mapping the RA-C11 epitope. The insert in clone S5 C11 was divided into segments by PCR. Amplified segments were expressed in bacteria. (A, B) Triplicate Western blots tested for reactivity with either AP-conjugated S-protein (AP-S), mAb RA-F1 or mAb RA-C11. Below each set of blots are maps of the DNA segments expressed in bacteria. The arrow marks the region encoding mAb RA-C11 reactivity. (C) The sequence in the smallest region found to react with mAb RA-C11 (boxed) and the boundaries of PCR fragments that mapped the epitope to this region. Forward slashes at the ends of lines indicate that the end of the DNA segment amplified was beyond the region covered by the line.

A

B
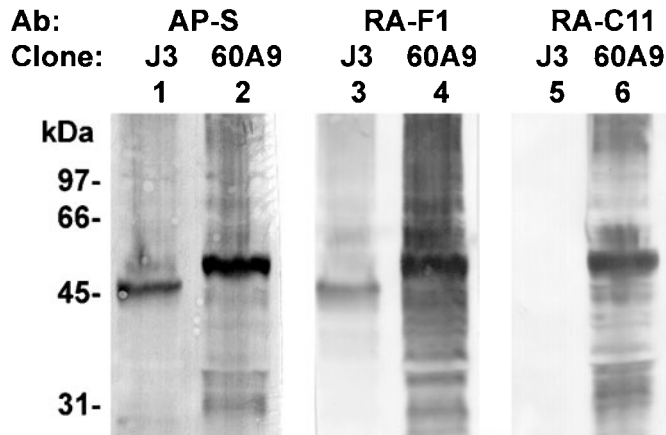
Fig. 2. Retention of mAb RA-C11 reactivity when the amino acids mapped in S5 C11 are part of a different peptide, and loss of reactivity in a peptide with a sequence that partially matches these amino acids. Clones J3 and 60A9 were generated by amplification of genomic *P. carinii* DNA and cloning into a plasmid expression vector. Triplicate Western blots were used to assay reactivity of peptides made from the two clones. (A) Lanes 1 and 2, reaction with AP-conjugated S-protein; lanes 3 and 4, reaction with mAb RA-F1; lanes 5 and 6, reaction with mAb RA-C11. (B) Sequences tested compared to those produced by expression of clone S5 C11. Clone 60A9 encoded a peptide with a sequence that matched the 10 contiguous amino acids boxed in the S5 sequence. Clone J3 encoded a peptide with a sequence that matched 8 of these 10 amino acids.

LREKCYkLKR, and LREgCYkLKR. Lack of binding of mAb RA-C11 to LREgCYELKR had been established by previous studies on the protein made from the MSGB gene (Fig. 3) (Linke *et al.*, 1998). Similar experiments showed that mAb RA-C11 did not bind to either LREKCYkLKR or LREgCYkLKR (see Fig. 2 for data obtained with MSG J3, which contained LREgCYkLKR). Results with an MSG

containing LREKCYkLKR were the same as those for MSG J3 (not shown).

The experiments with recombinant proteins produced results consistent with expectations based on previous studies in which mAb RA-C11 had been used to screen a λgt11 expression library (Linke *et al.*, 1998). In these previous studies, the mAb RA-C11 bound to only three of more than 700 phage plaques that contained the epitope recognized by mAb RA-F1. Because the epitopes for mAbs RA-C11 and RA-F1 are only 400 bp apart, and the typical gt11 chimeric phage genome carried approximately 2 kb of DNA, one would expect that most of the 700 mAb RA-F1-positive phage plaques would also carry the region in which the mAb RA-C11 epitope is located. The data in Fig. 3 suggest that nearly all of these plaques would carry LREgCYkLKR, LREKCYkLKR or LREgCYELKR. However, mAb RA-C11 did not bind to any of these plaques. Thus, all of the evidence indicated that the 10 amino acid sequence encoded by the S5 MSG gene and recognized by mAb RA-C11 is present in a small fraction of the proteins encoded by the MSG gene family. Therefore, this epitope, and the sequence that encodes it, provided markers suitable for testing the expression-site hypothesis.

S5	GCPS	**LREKCYELKR**	EKVA
MSG100	....	....	k...
PC5	k.vk	...g......	....
MSG1	n.vk	...g......	k...
C10	n.vk	...g......	k...
22C8	n.vk	...g......	k...
21H1.2	n.vk	...g......	k...
18A9	n.vk	...g......	k...
MSGB	n.vk	...g......	kr..
11A11.2	..vk	...g......	k...
11A11.1	..ve	...g......	k...
MSG104	....	......k...	q...
MSG103	....	......k...	q...
21H1.1	....	......k...	q...
MSG102	n.vn	......k...	.v..
MSG5	n.vn	......k...	....
MSG101	k.vn	......k...	....
3G5.3	d.vk	......k...	....
3G5.2	k.vk	......k...	....
11A11.3	d.vq	...g..k...	....
MSG2	n.Ie	...g..k...	....
GP3	n.Ie	...g..k...	....
PC3	n.vk	...g..k...	k...
MSG105	n.vn	...g..k...	k...
3G5.1	n.vn	...g..k...	k...
MSG3	n.vk	...g..k...	kr..

Fig. 3. The RA-C11 epitope is rare among known MSG sequences. MSG genes in GenBank and from other sources were translated *in silico* and peptides aligned. The RA-C11 epitope (in bold print on the top line) was present in only one other sequence, MSG 100.

Abundance of the RA-C11 epitope varies among populations of *P. carinii*

Previous studies on mRNA had shown that populations of *P. carinii* from individual rats tended to have transcripts encoding many different MSG isoforms, and that populations from different rats tended to differ with respect to the MSG sequences represented by mRNAs (Sunkin & Stringer, 1997). This variation suggested that some populations might express very little of the RA-C11 epitope, while

others might express much more. To determine if this were true, populations of *P. carinii* were screened for mAb-reactive proteins by immunoblotting of total proteins that had been separated by denaturing gel electrophoresis. Duplicate blots were reacted with either mAb RA-C11, to measure RA-C11 epitope abundance, or mAb RA-F1, to assess the amount of MSG. The choice of mAb RA-F1 for this purpose was based on previous work which showed that this antibody recognized 100 times as many plaques in the λgt11 expression library as mAb RA-C11 did, indicating that the RA-F1 epitope is very common among MSG forms (Linke *et al.*, 1998). In addition, the DNA sequence encoding the RA-F1 epitope is present in all known MSG genes (not shown).

The immunoblotting experiments identified three populations of *P. carinii* that differed with respect to reactivity with mAb RA-C11. Population 1160 had the highest reactivity, which was estimated to be eightfold greater than population 1153 and 44-fold greater than the population of lowest reactivity, 1140 (Fig. 4, black bars).

Abundance of the RA-C11 epitope correlates with residence of the C11-EES at the UCS locus

We next measured the relative abundance of UCS-linked C11-EES in the three populations, 1160, 1153 and 1140. To accomplish this, the UCS and attached MSG genes were amplified from genomic DNA from each of the populations. Amplification used an oligonucleotide primer specific for the UCS paired with a primer that matches a sequence conserved among MSG genes (Sunkin & Stringer, 1997; Keely *et al.*, 2003). Amplicons were characterized for size and abundance, then attached to a nylon membrane and hybridized with a radiolabelled 30 bp oligonucleotide that matched the C11 EES in clone S5-C11. To ensure specificity
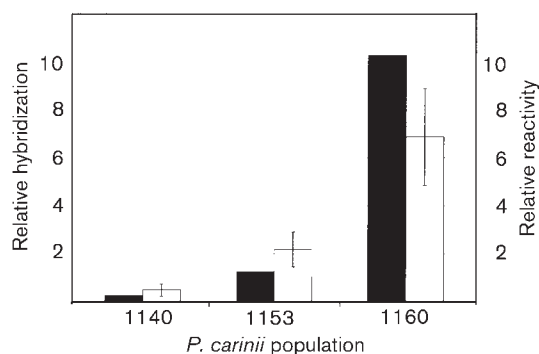

Fig. 4. mAb RA-C11 Western blot reactivity correlates with residence of the C11 EES at the UCS locus. Three populations of *P. carinii* (1140, 1153 and 1160) were analysed in two ways. Reactivity with mAb RA-C11 (black bars) was determined by Western blot of total proteins (results of single determinations are shown). Presence of the C11 EES at the UCS locus (white bars) was determined by DNA hybridization (means ± SE; three determinations).

of hybridization, tetramethylammonium chloride (TMAC) was used in the washing steps (Kren *et al.*, 1997). Control experiments showed that the 30 bp probe did not hybridize to sequences that were less than 100 % complementary to it. Bound probe was detected and quantified by phosphorimaging. To normalize the intensities of radioactive bands, C11 EES probe was removed from the membrane, which was then hybridized with an oligonucleotide identical to the 24 nucleotide CRJE, a sequence that is present at the 5′ end of every *P. carinii* MSG gene (Wada *et al.*, 1995).

Fig. 4 shows that the relative hybridization intensity (white bars) correlated with mAb RA-C11 reactivity (black bars) ($r = 0.986$). This correlation was strong enough to support the expression-site hypothesis with 90 % confidence ($P < 0.1$). Analysis of variance (ANOVA) performed on the hybridization values demonstrated a significant difference between the three populations, particularly between population 1160 and the other two ($P < 0.01$).

Residence of the C11 EES at the UCS locus correlates with surface reactivity

The expression-site hypothesis predicts that populations of *P. carinii* that exhibit high abundance of the RA-C11 epitope by immunoblotting will have more organisms with this epitope on their surface. To test this prediction, indirect immunofluoresence (IFA) experiments were attempted on organisms from populations 1140, 1153 and 1160. No staining was observed with either mAb RA-C11 or mAb RA-F1. Poor cryopreservation of the samples may have been at fault (A. G. Smulian, personal communication). In any case, it became necessary to use different populations of *P. carinii* for IFA measurements.

Samples from properly cryopreserved populations of *P. carinii* were screened by DNA hybridization (as described above) to identify those that varied with respect to representation of the C11 EES at the UCS locus. Three populations (M25-14, M54-12 and M54-27) that differed over a 10-fold range were identified. Specifically, population M25-14 exhibited the highest hybridization signal, which was approximately fourfold greater than that of population M54-12 and 10-fold greater than that of population M54-27 (Fig. 5).

To detect surface expression of the RA-C11 epitope, organisms were fixed on glass slides and treated with either mAb RA-C11 or mAb RA-F1. Organisms with antibody bound to them were visualized with FITC-conjugated anti-mouse antibody. To aid in visualization of organisms that failed to bind the FITC antibody, slides were counter-stained by treatment with rhodamine-conjugated ConA because ConA is known to bind to sugars on MSG and to label the surface of *P. carinii* (Pesanti & Shanley, 1988; Linke & Walzer, 1991). Fig. 6 shows representative fields of stained organisms. Reactivity with *P. carinii* cell wall was indicated by bright rings on cysts (example indicated by arrow in Fig. 6A). Both mAbs RA-F1 and RA-C11 co-localized with
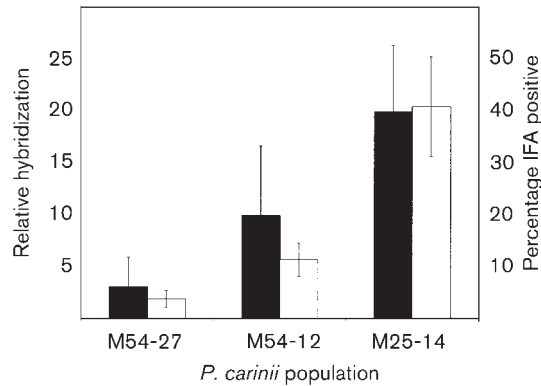
Fig. 5. Residence of the C11 EES at the UCS locus correlates with the fraction of organisms stained by mAb RA-C11. Three populations (M54-27, M54-12 and M25-14) were analysed in two ways. DNA hybridization was used to assess the fraction UCS loci that have the C11 EES (white bars). IFA determined the fraction of organisms stained with mAb RA-C11 (black bars). Means ± SE of three determinations are plotted in each case.

concanavalin A reactivity. Organisms that reacted with RA-C11 were counter-stained with DAPI, which reacts with nucleic acid and thus served as intracellular marker (Broomall *et al.*, 1998). RA-C11 and DAPI staining did not co-localize in either cysts or trophs (not shown).

The percentage of organisms stained by an antibody was determined by examining between 1000 and 3000

ConA-stained *P. carinii*. These values are depicted by the black bars in Fig. 5. There was a strong correlation between hybridization to the C11-EES and RA-C11 staining ($r = 0.975$, $P < 0.1$). ANOVA demonstrated significant differences between all three populations for both hybridization and IFA values ($P < 0.001$). These data supported the hypothesis that linkage of an MSG gene to the UCS leads to surface expression of a specific MSG protein.

DISCUSSION

Control of MSG expression is thought to be mediated by the UCS locus (Stringer & Keely, 2001). MSG transcripts match the sequences of the MSG genes that are linked to the UCS locus, suggesting that this linkage is necessary and sufficient for expression of the MSG protein encoded by the UCS-linked gene (Sunkin & Stringer, 1997). This expression-site model of transcriptional regulation predicts that the fraction of *P. carinii* organisms exhibiting a given MSG epitope will be correlated with the fraction of *P. carinii* organisms that have a gene encoding that epitope at the UCS.

Previous work had suggested that the RA-C11 epitope is rare, and therefore would serve to allow this prediction to be tested. Identification of the C11 epitope showed that 24 of the 26 known MSG isoforms do not contain it. Furthermore, preliminary experiments (not presented here), in which the C11 EES was used as a hybridization probe to screen *P. carinii* genomic libraries, suggested that only 3 of the 100 MSG genes in the *P. carinii* genome contain the DNA sequence that encodes the S5 RA-C11 epitope.
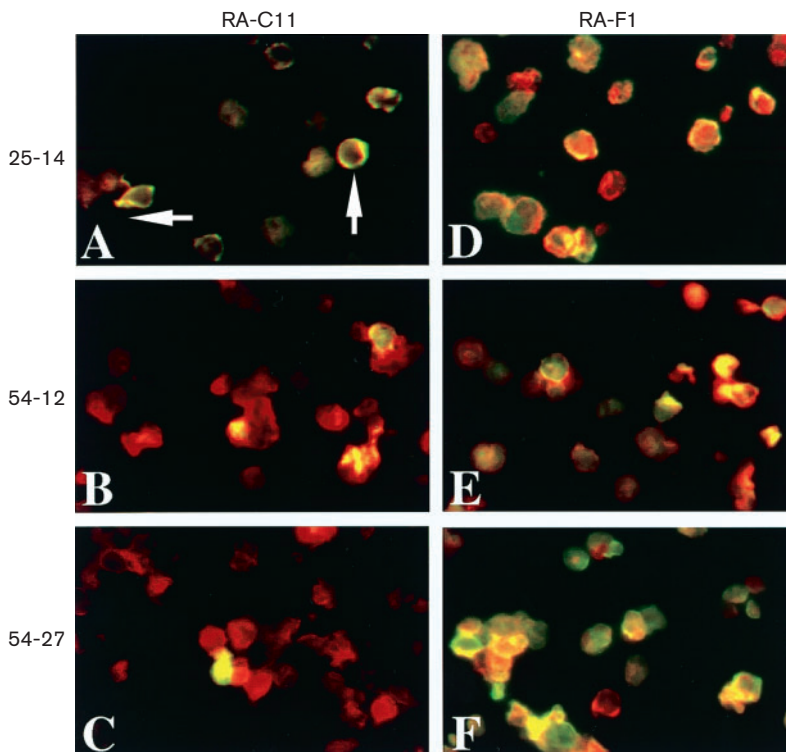

Fig. 6. IFA of organisms in three *P. carinii* populations. Organisms from populations M25-14 (A, D), M54-12 (B, E) and M54-27 (C, F) were incubated with either mAb RA-C11 (A–C) or mAb RA-F1 (D–F). Bound antibodies were visualized with FITC-conjugated anti-mouse antibody. Organisms were counterstained with rhodamine-conjugated ConA. Organisms stained by an antibody and ConA are yellow-green. Organisms stained with ConA only are red. The arrow indicates an example of a cyst exhibiting the bright ring of staining consistent with the antigen being on the surface.

The rarity of the C11 epitope suggested that its abundance might vary among populations of *P. carinii*. This was the case, and three populations that differed in this regard were readily identified by immunoblotting experiments. Subsequent analysis of the UCS locus in these populations showed that the C11 EES was linked to it in proportion to the level of RA-C11 immunoblot reactivity. To confirm the correlation, and to focus the analysis on surface antigens, DNA hybridization was used to identify a second set of *P. carinii* populations that differed with respect to the fraction of UCS-linked DNA that had the C11 EES. Subsequent analysis of these populations by IFA showed that the fraction of organisms labelled by the RA-C11 antibody was in proportion to UCS-linked C11 EES. These findings support the expression-site model of MSG regulation.

Details regarding the mechanism of regulation of expression aside, these studies show that *P. carinii* can vary with respect to expression of the RA-C11 epitope. In each of the three populations studied with IFA, some organisms did not bind the RA-C11 mAb. Hence, it appears that not all MSG isoforms are expressed in a given organism at the same time. The heterogeneity observed was not surprising for two reasons. First, heterogeneous staining of *P. carinii* populations by antibodies that recognize MSG epitopes has been reported before (Angus *et al.*, 1996). Second, heterogeneity is consistent with expectation based on previous studies of the UCS locus, which showed that heterogeneity in *P. carinii* populations is high when rats are kept in an open-air colony, where they are exposed for extended periods of time to high numbers of airborne *P. carinii* produced by other heavily infected rats in the colony (Keely *et al.*, 2003). The rats used in the study described above were from such a colony. By contrast, *P. carinii* populations that originated via proliferation of one or a few progenitors were much less heterogeneous at the UCS locus (Keely *et al.*, 2003). Such populations might also be less heterogeneous with respect to surface MSG epitopes. Studies to test this idea are under way.

Expression of one MSG on the surface at a time would mimic the strategy employed by other pathogens that establish themselves in immunocompetent mammalian hosts. Two well-studied examples of such pathogens are African trypanosomes and *Borrelia* spirochaetes (Donelson, 1995; Zhang *et al.*, 1997; Rudenko *et al.*, 1998). In each microbe, only one version of a major surface antigen is expressed per organism, and the version expressed can be changed by suppressing expression of the active gene while activating expression of an inactive gene. *P. carinii* can accomplish such a concerted change by moving a silent gene into the expression site.

The purpose of MSG expression changes might be to allow *P. carinii* to colonize immunocompetent rats. While well known as a pathogen of rats that lack a normal immune system, *P. carinii* can remain in immunocompetent rats for prolonged periods of time (Vargas *et al.*, 1995). Colonization may be the primary means by which *P. carinii*

survives. Three observations support this conjecture. First, the organism is very scarce in the environment, except where large numbers of infected rats are present (Vargas *et al.*, 1995; Wakefield, 1996). Second, outside of the rat, *P. carinii* propagates very poorly (Sloand *et al.*, 1993; Cushion, 1998b; Merali *et al.*, 1999). Third, *P. carinii* appears to be specific for rats, and not able to proliferate in other host species. This host-specificity phenotype is also exhibited by *Pneumocystis* from other host species, such as humans, mouse, rabbit and ferret (Gigliotti *et al.*, 1993; Sidman & Roths, 1994; Mazars *et al.*, 1997; Wakefield *et al.*, 1998; Demanche *et al.*, 2001; Durand-Joly *et al.*, 2002).

While the different *Pneumocystis* species and presumptive species exhibit many genetic differences, they all share the life style of habitation within the alveolar spaces of a mammalian lung (Stringer, 2002). The MSG system of *P. carinii* presumably plays a role in this lifestyle. As far as is known, all species in the genus *Pneumocystis*, including that found in humans, have families of genes that encode various forms of surface proteins (Stringer *et al.*, 1993; Garbe & Stringer, 1994; Kutty *et al.*, 2001). Understanding the MSG system in *P. carinii* should serve as a model for understanding the role of surface variation in other hosts, including humans.

ACKNOWLEDGEMENTS

REFERENCES

Angus, C. W., Tu, A., Vogel, P., Qin, M. & Kovacs, J. A. (1996). Expression of variants of the major surface glycoprotein of *Pneumocystis carinii. J Exp Med* **183**, 1229–1234.

Broomall, K. R., Morris, R. E., Walzer, P. D. & Smulian, A. G. (1998). Zymolyase treatment exposes p55 antigen of *Pneumocystis carinii. J Eukaryot Microbiol* **45**, 284–289.

Cornillot, E., Keller, B., Cushion, M. T., Metenier, G. & Vivares, C. P. (2002). Fine analysis of the *Pneumocystis carinii* f. sp. carinii genome by two- dimensional pulsed-field gel electrophoresis. *Gene* **293**, 87–95.

Cushion, M. T. (1997). *Topley and Wilson's Microbiology and Microbial Infections*, pp. 123–140, 9th edn, Edited by A. Balows & M. Sussman. London: Edward Arnold.

Cushion, M. T. (1998a). Genetic heterogeneity of rat-derived *Pneumocystis. FEMS Immunol Med Microbiol* **22**, 51–58.

Cushion, M. T. (1998b). Taxonomy, genetic organization, and life cycle of *Pneumocystis carinii. Semin Respir Infect* **13**, 304–312.

Cushion, M. T., DeStefano, J. A. & Walzer, P. D. (1988). *Pneumocystis carinii*: surface reactive carbohydrates detected by lectin probes. *Exp Parasitol* **67**, 137–147.

Cushion, M. T., Orr, S., Keely, S. P. & Stringer, J. R. (2001). Time between inoculations and karyotype forms of *Pneumocystis carinii*

f. sp. carinii influence outcome of experimental coinfections in rats. *Infect Immun* **69**, 97–107.

Demanche, C., Berthelemy, M., Petit, T., Polack, B., Wakefield, A. E., Dei-Cas, E. & Guillot, J. (2001). Phylogeny of *Pneumocystis carinii* from 18 primate species confirms host specificity and suggests coevolution. *J Clin Microbiol* **39**, 2126–2133.

Donelson, J. E. (1995). Mechanisms of antigenic variation in *Borrelia hermsii* and African trypanosomes. *J Biol Chem* **270**, 7783–7786.

Durand-Joly, I., Aliouat, el M., Recourt, C., Guyot, K., Francois, N., Wauquier, M., Camus, D. & Dei-Cas, E. (2002). *Pneumocystis carinii* f. sp. hominis is not infectious for SCID mice. *J Clin Microbiol* **40**, 1862–1865.

Edman, J. C., Hatton, T. W., Nam, M., Turner, R., Mei, Q., Angus, C. W. & Kovacs, J. A. (1996). A single expression site with a conserved leader sequence regulates variation of expression of the *Pneumocystis carinii* family of major surface glycoprotein genes. *DNA Cell Biol* **15**, 989–999.

Garbe, T. R. & Stringer, J. R. (1994). Molecular characterization of clustered variants of genes encoding major surface antigens of human *Pneumocystis carinii*. *Infect Immun* **62**, 3092–3101.

Gigliotti, F., Harmsen, A. G., Haidaris, C. G. & Haidaris, P. J. (1993). *Pneumocystis carinii* is not universally transmissible between mammalian species. *Infect Immun* **61**, 2886–2890.

Haidaris, C. G., Wright, T. W., Gigliotti, F. & Haidaris, P. J. (1991). Molecular cloning and characterization of ferret *Pneumocystis carinii* gp120. *J Protozool* **38**, 5S–6S.

Haidaris, C. G., Medzihradsky, O. F., Gigliotti, F. & Simpson-Haidaris, P. J. (1998). Molecular characterization of mouse *Pneumocystis carinii* surface glycoprotein A. *DNA Res* **5**, 77–85.

Keely, S. P., Cushion, M. T. & Stringer, J. R. (2003). Diversity at the locus associated with transcription of a variable surface antigen of *Pneumocystis carinii* as an index of population structure and dynamics in infected rats. *Infect Immun* **71**, 47–60.

Kovacs, J. A., Halpern, J. L., Swan, J. C., Moss, J., Parrillo, J. E. & Masur, H. (1988). Identification of antigens and antibodies specific for *Pneumocystis carinii*. *J Immunol* **140**, 2023–2031.

Kovacs, J. A., Powell, F., Edman, J. C., Lundgren, B., Martinez, A., Drew, B. & Angus, C. W. (1993). Multiple genes encode the major surface glycoprotein of *Pneumocystis carinii*. *J Biol Chem* **268**, 6034–6040.

Kren, B. T., Cole-Strauss, A., Kmiec, E. B. & Steer, C. J. (1997). Targeted nucleotide exchange in the alkaline phosphatase gene of HuH-7 cells mediated by a chimeric RNA/DNA oligonucleotide. *Hepatology* **25**, 1462–1468.

Kutty, G., Ma, L. & Kovacs, J. A. (2001). Characterization of the expression site of the major surface glycoprotein of human-derived *Pneumocystis carinii*. *Mol Microbiol* **42**, 183–193.

Linke, M. J. & Walzer, P. D. (1989). Analysis of a surface antigen of *Pneumocystis carinii*. *J Protozool* **36**, 60S–61S.

Linke, M. J. & Walzer, P. D. (1991). Identification and purification of a soluble species of gp120 released by zymolyase treatment of *Pneumocystis carinii*. *J Protozool* **38**, 176S–178S.

Linke, M. J., Cushion, M. T. & Walzer, P. D. (1989). Properties of the major antigens of rat and human *Pneumocystis carinii*. *Infect Immun* **57**, 1547–1555.

Linke, M. J., Sunkin, S. M., Andrews, R. P., Stringer, J. R. & Walzer, P. D. (1998). Expression, structure, and location of epitopes of the major surface glycoprotein of *Pneumocystis carinii* f. sp. carinii. *Clin Diagn Lab Immunol* **5**, 50–57.

Mazars, E., Guyot, K., Fourmaintraux, S., Renaud, F., Petavy, F., Camus, D. & Dei-Cas, E. (1997). Detection of *Pneumocystis* in European wild animals. *J Eukaryot Microbiol* **44**, 39S.

Mei, Q., Turner, R. E., Sorial, V., Klivington, D., Angus, C. W. & Kovacs, J. A. (1998). Characterization of major surface glycoprotein genes of human *Pneumocystis carinii* and high-level expression of a conserved region. *Infect Immun* **66**, 4268–4273.

Merali, S., Frevert, U., Williams, J. H., Chin, K., Bryan, R. & Clarkson, A. B., Jr (1999). Continuous axenic cultivation of *Pneumocystis carinii*. *Proc Natl Acad Sci U S A* **96**, 2402–2407.

Nakamura, Y. (1998). The major surface antigen of *Pneumocystis carinii*. *FEMS Immunol Med Microbiol* **22**, 67–74.

Pesanti, E. L. & Shanley, J. D. (1988). Glycoproteins of *Pneumocystis carinii*: characterization by electrophoresis and microscopy. *J Infect Dis* **158**, 1353–1359.

Rudenko, G., Cross, M. & Borst, P. (1998). Changing the end: antigenic variation orchestrated at the telomeres of African trypanosomes. *Trends Microbiol* **6**, 113–116.

Sambrook, J. F., Fristch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schaffzin, J. K. (1999). PhD thesis, University of Cincinnati.

Schaffzin, J. K., Garbe, T. R. & Stringer, J. R. (1999). Major surface glycoprotein genes from *Pneumocystis carinii* f. sp. ratti. *Fungal Genet Biol* **28**, 214–226.

Sidman, C. L. & Roths, J. B. (1994). *Pneumocystis carinii Pneumonia*, pp. 223–235, 2nd edn. Edited by P. D. Walzer. NY: Marcel Dekker.

Sloand, E., Laughon, B., Armstrong, M. & other authors (1993). The challenge of *Pneumocystis carinii* culture. *J Eukaryot Microbiol* **40**, 188–195.

Stevens, J. (1990). *Intermediate Statistics, a Modern Approach*, 1st edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stringer, J. R. (2002). *Pneumocystis. Int J Med Microbiol* **292**, 391–404.

Stringer, J. R. & Keely, S. P. (2001). Genetics of surface antigen expression in *Pneumocystis carinii*. *Infect Immun* **69**, 627–639.

Stringer, S. L., Hong, S. T., Giuntoli, D. & Stringer, J. R. (1991). Repeated DNA in *Pneumocystis carinii*. *J Clin Microbiol* **29**, 1194–1201.

Stringer, S. L., Garbe, T., Sunkin, S. M. & Stringer, J. R. (1993). Genes encoding antigenic surface glycoproteins in *Pneumocystis* from humans. *J Eukaryot Microbiol* **40**, 821–826.

Sunkin, S. M. & Stringer, J. R. (1996). Translocation of surface antigen genes to a unique telomeric expression site in *Pneumocystis carinii*. *Mol Microbiol* **19**, 283–295.

Sunkin, S. M. & Stringer, J. R. (1997). Residence at the expression site is necessary and sufficient for the transcription of surface antigen genes of *Pneumocystis carinii*. *Mol Microbiol* **25**, 147–160.

Sunkin, S. M., Stringer, S. L. & Stringer, J. R. (1994). A tandem repeat of rat-derived *Pneumocystis carinii* genes encoding the major surface glycoprotein. *J Eukaryot Microbiol* **41**, 292–300.

Sunkin, S. M., Linke, M. J., McCormack, F. X., Walzer, P. D. & Stringer, J. R. (1998). Identification of a putative precursor to the major surface glycoprotein of *Pneumocystis carinii*. *Infect Immun* **66**, 741–746.

Tanabe, K., Takasaki, S., Watanabe, J., Kobata, A., Egawa, K. & Nakamura, Y. (1989). Glycoproteins composed of major surface immunodeterminants of *Pneumocystis carinii*. *Infect Immun* **57**, 1363–1368.

Vargas, S. L., Hughes, W. T., Wakefield, A. E. & Oz, H. S. (1995). Limited persistence in and subsequent elimination of *Pneumocystis carinii* from the lungs after *P. carinii* pneumonia. *J Infect Dis* **172**, 506–510.

Vasquez, J., Smulian, A. G., Linke, M. J. & Cushion, M. T. (1996). Antigenic differences associated with genetically distinct *Pneumocystis carinii* from rats. *Infect Immun* **64**, 290–297.

Wada, M., Sunkin, S. M., Stringer, J. R. & Nakamura, Y. (1995). Antigenic variation by positional control of major surface glyco-protein gene expression in *Pneumocystis carinii*. *J Infect Dis* **171**, 1563–1568.

Wakefield, A. E. (1996). DNA sequences identical to *Pneumocystis carinii* f. sp. carinii and *Pneumocystis carinii* f. sp. hominis in samples of air spora. *J Clin Microbiol* **34**, 1754–1759.

Wakefield, A. E., Stringer, J. R., Tamburrini, E. & Dei-Cas, E. (1998). Genetics, metabolism and host specificity of *Pneumocystis carinii*. *Med Mycol* **36**, Suppl. 1, 183–193.

Walzer, P. D. & Linke, M. J. (1987). A comparison of the antigenic characteristics of rat and human *Pneumocystis carinii* by immuno-blotting. *J Immunol* **138**, 2257–2265.

Wright, T. W., Gigliotti, F., Haidaris, C. G. & Simpson-Haidaris, P. J. (1995). Cloning and characterization of a conserved region of human and rhesus macaque *Pneumocystis carinii* gpA. *Gene* **167**, 185–189.

Zhang, J. R., Hardham, J. M., Barbour, A. G. & Norris, S. J. (1997). Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**, 275–285.