Localized Frameshift Mutation Generates Selective, High-Frequency Phase Variation of a Surface Lipoprotein Encoded by a Mycoplasma ABC Transporter Operon

PATTY THEISS AND KIM S. WISE*

Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri-Columbia, Columbia, Missouri 65212

Received 10 February 1997/Accepted 14 April 1997

The wall-less mycoplasmas have revealed unusual microbial strategies for adaptive variation of antigenic membrane proteins exposed during their surface colonization of host cells. In particular, high-frequency mutations affecting the expression of selected surface lipoproteins have been increasingly documented for this group of organisms. A novel manifestation of mutational phase variation is shown here to occur in Mycoplasma fermentans, a chronic human infectious agent and possible AIDS-associated pathogen. A putative ABC type transport operon encoding four gene products is identified. The 3' distal gene encoding P78, a known surface-exposed antigen and the proposed substrate-binding lipoprotein of the transporter, is subject to localized hypermutation in a short homopolymeric tract of adenine residues located in the N-terminal coding region of the mature protein. High-frequency, reversible insertion/deletion frameshift mutations lead to selective phase variation in P78 expression, whereas the putative nucleotide-binding protein, P63, encoded by the most 5' gene of the operon, is continually expressed. Mutation-based phase variation in specific surface-exposed microbial transporter components may provide an adaptive advantage for immune evasion, while continued expression of other elements of the same transporter may preserve essential metabolic functions and confer alternative substrate specificity. These features could be critical in mycoplasmas, where limitations in both transcriptional regulators and transport systems may prevail. This study also documents that P63 contains an uncharacteristic hydrophobic sequence between predicted nucleotide binding motifs and displays an amphiphilic character in detergent fractionation. Both features are consistent with an evolutionary adaptation favoring integral association of this putative energy-transducing component with the single mycoplasma membrane.

With an ability to establish persistent or chronic infections in its mammalian host (30, 54), Mycoplasma fermentans typifies many of the nearly 100 species in the genus Mycoplasma (44). This agent has been implicated as a human pathogen associated with fulminant respiratory distress syndrome (31, 33) and pathologic lesions in immunocompromised individuals with AIDS (7, 32). Like all mycoplasmas, this species is evolutionarily related to the low-G+C-content branch of gram-positive eubacteria (61) and contains a small genome (20, 23). Mycoplasmas offer interesting and useful models as minimal prokaryotic pathogens. Notably, these organisms endure an extracellular existence exposed to host defenses, yet they possess only a single limiting plasma membrane without the additional protective outer membrane system or cell wall matrix characteristic of other eubacteria. Consequently, critical systems for transport and other metabolic pathways associated with “periplasmic” functions in both gram-negative and gram-positive eubacteria (4, 36) are on the membrane surface of mycoplasmas. How such components are concealed or otherwise evade host defenses remains a key question in understanding the adaptation of mycoplasmas as infectious agents and pathogens. For example, even though simple binding of antibody (Ab) may directly inhibit mycoplasma metabolic activities, growth, and survival (6, 10, 11, 14, 22), these organisms effectively maintain cell interactions, scavenge nutrients, and propagate during in vivo colonization of host cell surfaces in compartments accessible to the humoral system.

A general solution to mycoplasmal survival may lie in the operation of mutational systems to diversify the antigenic and structural surface phenotypes of these organisms in propagating populations. Adaptive surface variation through error-prone mutational systems linked to multigene families (63, 67) has been discussed as a general mycoplasmal strategy for generating high-frequency size and phase variations in coat proteins (5, 12, 35, 69, 70). Recent studies continue to delineate analogous new systems of surface protein variation in pathogenic mycoplasma species (34), strengthening this concept as a general means for mycoplasmal adaptation.

The potentially major role of surface variation in mycoplasmal adaptation is also consistent with new lines of evidence emerging from the entire contiguous chromosomal sequences determined for two mycoplasma species, Mycoplasma genitalium (20) and Mycoplasma pneumoniae (26). The 580-kb M. genitalium chromosome contains fewer than 500 predicted genes, with a striking sparsity of homologs for known regulatory elements (e.g., a single sigma factor was identified, and no two-component sensor elements were apparent). Similar limitations have been noted for the 816-kb genome of M. pneumoniae. Several anabolic and metabolic pathways, including those for cell wall synthesis, are also lacking, as anticipated from earlier evidence of biochemical and metabolic deficiencies in mycoplasma (37). On the other hand, the predominance of motifs suggesting general transport systems (configured as operons) is consistent with active exchange of metabolites with
the environment. However, the actual number of such systems reported for these two species may be insufficient to accommodate several critical substrates for which de novo synthesis is precluded (20, 26). Compared to larger, conventional eubacteria, mycoplasmas have limited genomic capabilities, which are compatible with (i) a parasitic lifestyle requiring nutrient acquisition to compensate for the absence of major biosynthetic pathways, (ii) a dynamic cell surface, in terms of transport function and variable surface components mediating interactions with the host, and (iii) an underrepresentation of classic gene regulation systems, including those affecting structural and functional features of surface-associated components of eubacterial cells (20, 26, 41). In this context, mutations affecting the characteristics and function of assorted surface components represent interesting alternatives to classic gene regulation and sensing systems as an adaptive strategy for mycoplasmas.

To better understand surface structure and dynamics in mycoplasma host adaptation, we have investigated features of M. fermentans membrane constituents. In particular, we have identified several surface lipoproteins in this species that are subject to variation. Some display size differences among strains (64), while others show high-frequency phase variation in propagating populations, either by differential presentation of epitopes on the surface (57, 58) or by varied expression of the surface lipoprotein itself (58). While patterns of variation in colonies suggest heritable and reversible phenotypic switching, the underlying mechanisms are not fully understood. One protein undergoing phase variation in expression is the P78 lipoprotein (58). This protein can be metabolically labeled with palmitate, is defined by a specific surface-binding monoclonal antibody (MAb), and can be detected in the human host at infected lesions (31).

In this report we describe characteristics of the P78 lipoprotein as determined by its gene sequence and delineate its probable role as a surface-exposed component of a membrane ABC transporter encoded within an operon. We also demonstrate the mutational basis by which P78, but not all products of the operon, undergoes high-frequency phase variation in expression. The possible ramifications of such localized mutational activities for the surface adaptation of mycoplasmas and other prokaryotic pathogens are discussed. In the course of this study, an unusual hydrophobic domain was also identified within the predicted nucleotide-binding protein encoded by this operon; this domain could mediate integral association with the plasma membrane.

MATERIALS AND METHODS

Bacterial strains, plasmids and culture conditions. M. fermentans PG187 has been characterized and was grown in modified Hayflick medium (66) as previously described (58, 64). The Escherichia coli strains used were KW251 (Promega Biotech), for chromosomal library construction and propagation of bacteriophage isolates, and DH5α MCR (BRL Life Technologies, Inc.), for propagation of plasmids and expression of plasmid-encoded recombinant proteins. KW251 cells were grown with shaking at 37°C in NZCYM (50) broth supplemented with 0.2% maltose. DH5α MCR cells were grown with shaking at 37°C in 2× yeast-tryptone medium supplemented with ampicillin at 100 μg per ml or with kanamycin at 40 μg per ml, as appropriate for plasmid selection. Plasmids pMPFZ-3, pMPFZ-4, and pMPFZ-7 were derived from pGEM-11Z (Promega). Plasmid pMPFZ-20 was derived from pGEM-7Z (Promega). Plasmid constructs placed M. fermentans DNA inserts under the control of the T7 promoter present in the vectors. Induction of expression was provided by cotransformation with pGP1-2 (56), which confers kanamycin resistance and which contains the heat-inducible T7 RNA polymerase gene. Conditions for the induction and analysis of recombinant mycoplasma proteins are described elsewhere (57). Competent E. coli DH5α MCR cells were transformed according to the manufacturer’s instructions (BRL Life Technologies). Plasmid DNA was extracted from E. coli with a Wizard Miniprep DNA purification system (Promega).

Chemicals and enzymes. Restriction enzymes, avian myeloblastosis virus (AMV) reverse transcriptase (RT), and T4 ligase were obtained from Promega, Boehringer Mannheim Biochemicals, and BRL Life Technologies and used as recommended by the manufacturers. AmpliTaq (Perkin-Elmer) or Vent (New England Biolabs) DNA polymerases were used for PCR amplifications. Oligonucleotide primers were synthesized by the Molecular Biology Program DNA Core Facility at the University of Missouri with an ABI model 390A DNA synthesizer (Applied Biosystems, Inc.).

Antibodies and immunoblotting. MAB F202C21A to P78 has been described elsewhere (64). Mouse antisera to synthetic peptides were generated and used as described previously (13). Peptides included NH2-CSKSFGVIKANQDINI (beginning at nt 5925), PHG18T (beginning at nt 5633) and 5′-ATTCGCTTCATTTGCAGC-3′ (beginning at nucleotide [nt] 5633) and 5′-CTTCTGTAATCAGC-G3′ (beginning at nt 5826), respectively. (All nucleotide positions refer to the mDNA sequence deposited in GenBank accession no. AF003088, listed below.) The PCR mixture for labeling all DNA probes in this study contained approximately 50 ng of template DNA, 1× AmpliQ buffer, 5 μl of AmpliQ DNA polymerase, 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP, 0.35 mM DIG, and 1 nmol of each primer in a total volume of 100 μl. PCRs were performed under oil in a 30 ml microtiter plate. After PCR amplification, the PCR product was purified using a 2× wash solution (2× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 1% SDS) for 5 min, followed by a 0.5× wash (0.5× SSC, 0.1% SDS), and the blot was developed by colorimetric detection.

PCR amplification of the p78 sequences from M. fermentans strains. Chromosomal DNA from M. fermentans PG187, Inconitus, K-12, SK 5, and MT-2; three others selected from isolates (provided by Shih-Ching Lo of the Armed Forces Institute of Pathology); and Mycoplasma hominis 1620 (71) were used as templates for PCR. Forward and reverse primers, representing sequences within the p78 gene, were 5′-GTATCTGTAATCAGCG-3′ (beginning at nt 5633) and 5′-ATTCGCTTCATTTGCAGC-3′ (beginning at nt 5633), respectively. A PCR mixture containing approximately 50 ng of DNA, 1× AmpliQ buffer, 5 μl of AmpliQ DNA polymerase, 20 mM (each) deoxynucleoside triphosphate (dNTP), and 1 nmol of each primer in a total volume of 100 μl. The cycling profile consisted of denaturing at 95°C for 1.5 min, annealing at 55°C for 1 min, and extension at 72°C for 1.5 min for 25 cycles. The predicted fragment (about 760 bp) was amplified from each sample and hybridized to
a DIG-labeled probe generated with forward primer 5'-CTCGTGTAGGCAAAGGCAAGAAG-3' (beginning at nt 5910) and reverse primer 5'-CCATTGAAACTCTGCTTCAATCTG-3' (beginning at nt 6146), which corresponds to a 237-bp sequence within the predicted 760-bp amplified product.

A partial sequence encoding ORF 3 (beginning at nt 2139) and reverse 5'-GCGACACAGAAG-3' (beginning at nt 2519) was cloned to generate M. fermentans ABC transporters. An M. fermentans chromosomal insert from a second recombinant phase isolate expressing a P78 epitope (see above) was cloned to generate pMFpZ-7. Sequencing 5' from the p78 gene toward the end of this 12-kb insert revealed open reading frame (ORF) 3 (encoding 327 amino acids [aa]) adjacent to the p78 gene, ORF 2 (encoding 538 aa) immediately upstream of ORF 3, and another ORF upstream of ORF 2, which extended to the 5' end of the pMFpZ-7 construct. To identify sequences further upstream of the insert in pMFpZ-7 a DIG-labeled probe corresponding to a 95-bp sequence near the 5' end of ORF 4 (encoding 312 aa) was used. The forward and reverse primers were 5'-CTTCACTTAA-3' (beginning at nt 292) and reverse primer 5'-GGATATTTTTAATAGC-3' (beginning at nt 7320). Reaction mixtures contained approximately 1.9 kb. A plasmid library of size-selected genomic EcoRI fragments was generated in pGEM7-Z and screened with the same probe, yielding pMFpZ-20 (see Fig. 1B). The confirmed DNA sequence of this EcoRI fragment of approximately 1.9 kb. A plasmid library of size-selected genomic EcoRI fragments was generated in pGEM7-Z and screened with the same probe, yielding pMFpZ-20 (see Fig. 1B). The confirmed DNA sequence of this EcoRI fragment of approximately 1.9 kb.

Two independent strategies were used to identify and circumvent a deletion of 232 nt to obtain the correct genomic sequence and to confirm that it encoded a single ORF product (P63) with a high degree of sequence similarity to ATPases over its entire length. First, primers in each ORF that flank the region in question were used to amplify a PCR product from genomic DNA of M. fermentans PG18T and Incognitus. The PCR mixture contained approximately 10 ng of chromosomal DNA from M. fermentans PG18T or Incognitus, 1× AmpliTaq buffer, 5 U of AmpliTaq DNA polymerase, 20 mM (each) dNTP, and 1 nmol of each primer in a total volume of 100 l. The forward and reverse primers were 5'-CTTCAACACCATCATAC-3' (beginning at nt 79) and 5'-GG ACTAACACCAAAAAATGGT-3' (beginning at nt 1556), respectively. The cycling profile was 94°C for 1 min, 54°C for 1.5 min, and 72°C for 2 min for a total of 25 cycles. Products were analyzed on 0.8% agarose gels. Single bands were excised and precisely sized from all bands. DNA sequences from two independently obtained PCR amplifications from each strain were identical and revealed a continuous ORF including the 232-bp fragment deleted from pMFpZ-20.

To further confirm the product of the ORF, PAs were generated against an N-terminal peptide epitope of the predicted P63 product (described above). Western immunoblots of proteins from M. fermentans PG18T stained with the antiserum revealed a specific protein product of approximately 63 kDa (shown also in Fig. 2B). Preimmune serum failed to detect the protein, and competitive inhibition of Ab staining with the corresponding synthetic peptide confirmed the specificity of the Ab for this protein product (data not shown). Immunodepletion of P63 from antiserum to protein sequence encoded by the deduced ORF verified the presence of a continuous gene sequence encoding P63.

PCR amplification of the 5' region of the p78 gene from clonal isolates that vary in expression of P78. Organisms from 200-ml broth cultures of mid-logarithmic-phase M. fermentans clonal isolates selected from a line of organisms in which the expression of P78 were centrifuged and rinsed with phosphate-buffered saline. Cells were resuspended in 70 l of water and heated at 100°C for 10 min. The same PCR mixture and cycling profile were used as that used to detect the p78 gene in other M. fermentans strains (28). It was possible to detect the forward primer was 5'-GCGACACAGAAG-3' (beginning at nt 4818) and the reverse primer was 5'-CATGCTAATTTTGTAC-3' (beginning at nt 5363). Products of the predicted size (approximately 545 bp) were excised from a 0.8% agarose gel, purified with the Gene Clean system (Bio 101, Inc.), and directly sequenced with primers corresponding to flanking sequences (Fig. 1A). The deduced P78 sequence is described below.

Amplification of the sequence encoding the predicted transmembrane domain in P63. The forward and reverse primers used to amplify the transmembrane domain of P63 were 5'-GGGGAATGGAACCATACCGAATG-3' (beginning at nt 1171) and 5'-CAGTGTCCCTTCTTTTACAC-3' (beginning at nt 1897), respectively. The PCR mixture contained approximately 10 ng of chromosomal DNA from M. fermentans PG18T or Incognitus, 1× AmpliTaq buffer, 5 U of AmpliTaq DNA polymerase, 20 mM (each) dNTP, and 1 nmol of each primer in a total volume of 100 l. The cycling profile was 94°C for 1 min, 54°C for 1.5 min, and 72°C for 2 min for a total of 25 cycles. Products were confirmed to be the predicted size (approximately 725 bp) and were excised and directly sequenced.

RT-PCR. Total RNA was obtained from a 10-ml mid-logarithmic-phase culture of M. fermentans PG18T (displaying a positive P78 phenotype [58]) by the hot phenol extraction method (1). To generate cDNA, duplicate sets of reaction mixtures containing RNA and 4 pmol of each reverse primer were prepared. Reaction mixtures were incubated at 65°C for 5 min and then placed on ice. To each of one set of mixtures, 20 mM (each) dNTP, 1 l of AMV RT, and AMV RT buffer (1×) were added, to a final volume of 30 l. Identical control reaction mixtures, but without AMV RT, were prepared in parallel. Reaction mixtures were placed at 37°C for 2 h. Then, 1 l of stop buffer (above) was added. To each of one set of reactions, 10 l of E. coli Tris, 10 mM EDTA, pH 7.5, was then added. Five microliters of each reaction mixture was removed and used as a template for PCR. The PCR mixture contained 5 l of template, 20 mM (each) dNTP, 1× AmpliTaq buffer, 1 l of AmpliTaq, approximately 4 pmol of each forward and reverse primer in a set, and water to a final volume of 100 l. The primers, sets designed to span the junctions of specific ORFs in the p78 operon, were as follows: PS1, forward 5'-GCGACCAATTTTTGTAC-3' (beginning at nt 2139) and reverse 5'-ACCGGATGAAACGCTTAAACC-3' (beginning at nt 2600), spanning the ORF 1–ORF 2 junction; PS2, forward 5'-CAGTTAGTCACTAC-3' (beginning at nt 3569) and reverse 5'-CTCAGAAATTTTTGTAC-3' (beginning at nt 4226), spanning the ORF 2–ORF 3 junction; PS3, forward 5'-GCAAGCATACGGGGCC-3' (beginning at nt 4818) and reverse 5'-CATGCTATTTTGTAC-3' (beginning at nt 5363), spanning the ORF 3–ORF 4 junction; and PS4, forward 5'-CTGCTAACAAAAATGGTGAT-3' (beginning at nt 6800) and reverse 5'-GCGACACAGAAG-3' (beginning at nt 7320), spanning the 3' end of ORF 4 and the stem-loop structure downstream of this terminal gene in the operon. The location of each primer set is shown in Fig. 1C. The cycling profile consisted of denaturing at 95°C for 2 min, annealing at 55°C for 2 min, and extension at 72°C for 1 min for a total of 35 cycles. Products were analyzed on 0.8% agarose gel. Control reaction mixtures prepared without RT were used in reactions performed under identical conditions.

DNA sequencing and computer analysis. DNA sequencing reactions were performed in the Molecular Biology Program DNA Core Facility with Taq DyeDeoxy terminators and a model 373A automated DNA sequencer (Applied Biosystems Inc.). DNA and protein sequences were analyzed by the Genetics Computer Group sequence analysis software package from the University of Wisconsin, Madison, Wis. (16).

Nucleotide sequence accession number. The contiguous sequence containing the p78 operon and its flanking regions has been deposited in the GenBank database under accession no. AF000380.

RESULTS
Characterization of the p78 gene. As part of a general strategy to identify genes encoding antigenic proteins of M. fermentans, a phage expression library was constructed in Lambda-GEM-11 with chromosomal DNA of M. fermentans PG18T and screened with mouse antiserum generated against whole organisms or TX-114 phase proteins of M. fermentans PG18T (data not shown). After restriction enzyme mapping of the insert to identify additional sites for subcloning the p78 gene, a 5.7-kb SacI-XhoI fragment was subsequently cloned to generate pMFpZ-4, which also expressed the full-length P78 detected by the MAb against this protein (Fig. 1A). The pMFpZ-4 insert was fully sequenced with primers corresponding to flanking sequences in the vector and by primer walking. The deduced P78 sequence is described below.

Southern blot analysis of M. fermentans PG18T chromosomal DNA with a PCR-generated 293-bp probe within the putative p78 gene (see Materials and Methods for all primer sequences and precise locations) identified single fragments of similar intensity and predicted size in digests prepared with EcoRI (approximately 0.55 kb) or NspV (approximately 2.3 kb) (data not shown). These results are consistent with there being a single copy of the p78 gene. To assess the presence of the p78 gene in additional strains and isolates, an internal sequence of approximately 760 bp from the middle of the ORF was amplified by PCR from M. fermentans Incognitus, SK 5, K-2, and MT-2 and from three isolates of M. fermentans from humans (see Materials and Methods). These PCR products were of the predicted size and hybridized to a 237-bp probe corresponding to sequences contained within the 760-bp amplified product, con-
The mature protein is hydrophilic, further suggesting that the enzyme limiting membrane of the organism. The sequence of the mature P78 protein is anchored by its lipid moiety in the lipid acylation of this Cys residue and cleavage of the signal, signal peptidase II homolog (13, 55). It is predicted that after cleavage site (SSSC), consistent with recognition by a N terminus, a hydrophobic core, and a tetrapeptide recognition terminator. The P78 sequence contains a typical prokaryotic prolipoprotein signal peptide (9, 13) with a positively charged N terminus. The mature P78 protein is external to the membrane. This was consistent with its known accessibility to proteolysis and Ab binding on intact cells (64).

The P78 lipoprotein is encoded within a four-gene operon. No ORFs (greater than 42 codons) were found on either strand downstream of the p78 gene for approximately 300 nt. However, an ORF occurred upstream of the p78 gene and extended to the 5’ end of the insert in pMFPZ-4. This ORF (later identified as part of the ORF for P35) had the same orientation as p78 and terminated 18 nt upstream of the ATG start codon for P78. This short intervening sequence contained the putative RBS for P78 translation but no stem-loop structure, raising the possibility that P78 was expressed from a polycistronic message. Subsequent sequencing of a region approximately 5 kb upstream of the p78 gene revealed three additional ORFs organized in a manner suggesting that the added ORFs are not part of the p78 operon. The organization of the operon is indicated in Fig. 1C, which shows the following similarly oriented encoded products (from 5’ to 3’): P63 (ORF 1; 600 aa), P58 (ORF 2; 536 aa), P35 (ORF 3; 326 aa), and P78 (ORF 4; 680 aa), respectively. The strategy for sequencing the 5’ portion of this putative operon upstream of
the insert contained in pMFZ-4 is outlined in Fig. 1B and is described in detail in Materials and Methods.

No ORFs (greater than 77 codons) were found upstream of ORF 1 on either strand for 481 nt, suggesting that p63 is the most 5' gene of the operon. Characteristic of operon organization, each ORF upstream of P78 contained an ATG start codon with an appropriately spaced RBS, and these ORFs displayed adjacent or overlapping start and termination codons. No stem-loop structures were associated with ORFs other than the one 3' of ORF4 (P78).

The orientation of genes, the spacing of ORFs, and the presence of a single stem-loop structure following the 5' p78 gene imply expression of these genes from a polycistronic message. However, Northern blot analysis with double-stranded, PCR-generated probes representing each of the four gene sequences failed to detect any transcript under conditions used previously to monitor mycoplasma gene expression (12). Therefore, RT-PCR was performed as a more sensitive method to assess the presence of mRNA spanning the junctions of ORFs in the operon. Primer sets consisted of a forward and reverse primer, each set spanning the junctions between two successive genes. A control primer set spanned the 3' end of the p78 structural gene and the stem-loop structure immediately following the termination codon of the gene (Fig. 1C). The cDNA template for PCR was generated by annealing each reverse primer to mRNA and extending with RT. Thus, the PCR product subsequently amplified by opposing primer pairs would result only if the mRNA used to generate the template spanned the junction of each gene. Amplified PCR products of the predicted sizes were obtained from the three primer sets that spanned each ORF junction (Fig. 1C). No products were generated under identical reaction conditions when RT was omitted, thereby ensuring that the PCR products did not result from amplification of contaminating genomic DNA in the RNA preparation. The control primer set spanning the 3' end of the p78 gene and the stem-loop structure downstream of the p78 gene failed to amplify a product, as expected, indicating that no stable transcript was synthesized immediately downstream of the proposed operon. These results were consistent with a genomic configuration of genes in which surface lipoprotein P78 was expressed from the 3' distal gene of a four-gene operon, designated here as the p78 operon.

The p78 operon has characteristic features of ABC transporters. Based on sequence similarities and structural motifs identified in searches of databases, proteins encoded by the p78 operon are predicted to include components of a bacterial ABC transporter (19, 24). Each of the two proteins encoded by ORFs immediately upstream of the ORF for the P78 lipoprotein, P58 (ORF 2) and P35 (ORF 3), contains eight putative membrane-spanning domains, as predicted from the sequence by using the algorithm of Klein et al. (29). Although these proteins showed no overall sequence similarities to other proteins in available databases, their predicted features are compatible with the formation of a solute channel in the membrane similar to the characterized transmembrane proteins in other ABC transport operons (24, 25, 53). Direct evidence that one transmembrane domain of P35 can promote translocation across the membrane was obtained in a recent study (13) of random TnphoA mutagenesis of M. fermentans chromosomal fragments in E. coli, which identified an exported sequence within P35 (the p35::TnphoA fusion occurred in the Glu codon at nt 4307).

P63 (ORF 1) shows considerable protein sequence similarity to other ATP-binding proteins of ABC transporters. The highest degree of homology was to the MglA protein of the methylgalactoside transport system of E. coli (38, 68), with 65% similarity (and 30% identity) over 368 aa. P63 contains several regions characteristic of ATPases, including two nucleotide binding domains, each containing Walker box A and Walker box B motifs (38, 60). Figure 2 shows the alignment of P63 with the three products giving the highest amino acid sequence similarities retrieved from databases (using the FAST-A search program). Each is an ATPase from a known ABC transporter.

Although P63 has features similar to those of other ATPases and although its gene is organized within a genetic structure characteristic of those of ABC transporter or permease operons, one feature of this protein is quite distinctive. Typically, ATPases encoded by ABC transport operons are found to be hydrophilic by sequence analysis and are relatively soluble in aqueous environments (19, 24). However, P63 contains a highly unusual 55-aa hydrophobic region predicted to form a transmembrane domain (29, 49) (Fig. 2; indicated by the large gap in the three homologs shown). PCR-derived sequences of this region from M. fermentans PG18T and Incognitus (see Materials and Methods) confirmed the presence of this hydrophobic region of P63 in these two strains of the organism. The presence of this hydrophobic region in P63 raised the possibility that this protein might be tightly associated with the membrane. TX-114 phase fractionation of proteins from M. fermentans (Fig. 2B), where it initially partitioned into the denser detergent phase with other amphiphilic membrane proteins but subsequently precipitated during the fractionation procedure (lane P). Selective precipitation of some TX-114 phase proteins has been documented in mycoplasmas, including the P95 protein of M. fermentans (65). This component also was evident (Fig. 2B) after the Western blot was restained with a combination of MAbs specific for M. fermentans membrane proteins. Other TX-114 phase components, including P150, P61, and the P78 lipoprotein, remained comparatively soluble in the detergent phase. The partitioning of P63 into the detergent phase suggests that this putative nucleotide-binding protein may have inherent amphiphilic properties that could directly promote interaction with a membrane.

In conjunction with the characteristic nucleotide binding motifs of P63 and the predicted polytopic membrane topology of P58 and P35 (capable of forming a transmembrane channel), the P78 lipoprotein itself is also characteristic of an increasing number of ABC transporters. Although the p78 sequence showed no significant similarity to other sequences in the database, it has striking analogies to other characterized transport operons of gram-positive organisms (55) and mycoplasmas (18). These encode lipid-modified proteins that have been shown or are proposed to be substrate binding components of ABC transporters. The use of anchored proteins for this function in cells lacking a confined periplasmic compartment has been discussed previously (9, 55). These proteins are also the only components of the transporters predicted to be substantially exposed on the external side of the plasma membrane.

Mutational basis for selective, high-frequency phase variation of the P78 surface lipoprotein. In addition to its possible function as a substrate binding protein exposed on the cell surface, P78 is subject to another important mechanism operating in M. fermentans, that of high-frequency phase variation in expression. The sequence of the p78 operon provided a means to assess the molecular basis of phase variation, both of the P78 product per se and of other products encoded by this
A clonal lineage oscillating in the expression of P78 was established by colony immunoblotting with the MAb to P78 (Fig. 3A). Western immunoblots of three variant isolates stained with this MAb (Fig. 3B, left) showed that P78 varied in its expression from on to off to on, corresponding to the surface phenotypes observed on colony immunoblots.

Mutations operating at several different levels might alter P78 expression. These include (i) mutations affecting the transcription of the entire p78 operon, (ii) mutations affecting the transcription or translation of the p78 gene, and (iii) DNA rearrangements affecting expression of the MAb-defined P78 epitope. We first addressed the possibility that high-frequency phase variation is generated by mutations affecting expression of the entire operon. Western immunoblots of the phase-variable P78 isolates were stained with the antipeptide Ab defining P63, the product of the 5' gene in the p78 operon. P63 was expressed in all isolates of the lineage, regardless of the expression state of P78 (Fig. 3B, right). Thus, the first gene product of the operon was constitutively expressed in these populations, arguing against mutationally altered transcription as a mechanism controlling P78 phase variation.

Since P78 was encoded by the 3' gene of the operon and appeared to selectively undergo variation in expression, mutational events affecting the translation of P78 were investigated.
Western blots of proteins from the same lineages described above were screened with an antipeptide Ab generated against an N-terminal epitope of P78 (the location of this peptide is indicated in Fig. 3C). The same pattern of P78 expression was demonstrated with this Ab (Fig. 3B, right) as with the MAb to P78 (Fig. 3B, left). This result suggested that any mutational event affecting translation might occur upstream of the N-terminal peptide sequence, and therefore led to the localization of a region in P78 that was further examined for mutational differences in phase variants. To assess mutations in this region, the sequence of the p78 gene encoding the N-terminal region of P78 was obtained by PCR from four isolates in a lineage oscillating in P78 expression. The primers were positioned to include the sequence upstream of the p78 gene and downstream of the sequence encoding the N-terminal peptide epitope in P78. In the populations which express P78 (Fig. 3B, lanes 1 and 3) the sequence in this region corresponds to an ORF encoding a full-length P78 (Fig. 3C, top). This was identical to the genomic sequence obtained from phase isolates expressing P78. However, in sequences obtained from the two clonal populations which did not express P78 (Fig. 3B, lane 2 shows one such isolate) a single base deletion in a short tract of adenine (A) residues was found (nts 5178 to 5184). This mutation creates a frameshift which would result in the truncation of P78 due to a TAA stop codon immediately downstream of the sequence encoding the N-terminal peptide (Fig. 3C, bottom). These data indicate that a high-frequency insertion/deletion event in a short homopolymeric poly(A) tract generates phase variation of P78. The predicted product in these P78-negative isolates would be approximately 10 kDa. However, no product of this size was detected in a negative variant by Western immunoblots stained with the Ab generated against the N-terminal epitope or by autoradiography of membrane proteins metabolically labeled as previously described (58) with [35S]Cysteine (data not shown).

**DISCUSSION**

This study revealed several important aspects of phenotypic variation involving the P78 lipoprotein of *M. fermentans*. The p78 gene appears to be a member of a four-gene operon based on the close spacing of each ORF and the presence of only one stem-loop structure following the 3′ most distal gene (p78). The detection by RT-PCR of mRNA spanning the junction of each gene in the putative P78 operon is also consistent with the presence of a polycistronic message. Sequence data indicated that P78 contains a prokaryotic prolipoprotein signal peptide sequence (9), which is consistent with previous results indicating that P78 can be metabolically labeled with palmitate (64). The epitope recognized by the MAb to P78 (Fig. 3B, left) was identical to the genomic sequence obtained from phage isolates expressing P78. However, no product of this size was detected in a negative variant by Western immunoblots stained with the Ab generated against the N-terminal epitope or by autoradiography of membrane proteins metabolically labeled as previously described (58) with [35S]Cysteine (data not shown).

FIG. 3. Mutational basis for generating phase variation in P78 expression. (A) Colony immunoblots stained with the MAB to P78 to establish clonal lineages oscillating in the expression of P78. The left immunoblot shows a clonally derived population displaying a negative phenotype, with a revertant positive colony indicated by an arrow. Restaining of this immunoblot with a PAb generated against *M. fermentans* proteins (64) revealed several colonies in this region of the immunoblot (data not shown). The colony immunoblot on the right shows the subsequent progeny population (derived by clonal isolation from the population shown on the left) with a positive phenotype. A revertant negative sector of a colony within this population is indicated by the arrow. (B) SDS-PAGE and immunoblots of total protein preparations from three isolates in a clonal lineage of *M. fermentans* PG18T stained with the MAB to P78 (left) or an Ab generated against a synthetic peptide representing an epitope in the N-terminal region of P78 (right). The location and sequence of this peptide is described in Materials and Methods. The P78 phenotypes (+ or −) of isolates in the lineage are indicated. Progeny and progenitor populations are identified by arrows above the lanes. Total protein in lanes 2 and 3 was from the two populations shown in the left and right portions, respectively, of panel A. The immunoblot on the right was restained with an Ab to a synthetic peptide representing an epitope within P63. The location and sequence of the peptide are described in Materials and Methods. The positions of P78 and P63 are indicated by arrows. (C) Features of two versions of the P78 protein determined from PCR-derived sequences (see Materials and Methods for locations and sequences of primers) of isolates shown in lanes 1 and 2 (P78 on to P78 off) in panel B. The location of the N-terminal peptide within P78, used to generate an Ab, is shown. Below each version of the protein is the sequence resulting in its expression, either the sequence resulting in full-length P78 (top) or the sequence, reflecting the deletion of a single A residue, resulting in a frameshift and truncation of the P78 sequence (bottom). Expression states of P78 are indicated. (D) Comparison of sequences spanning the poly(A) tract in the p78 genes from four clonal variants which oscillate in the expression of P78 from − to + to − to +. Sequences are aligned, with a dash indicating the gap in the poly(A) tract.
increasing number of lipoproteins have been implicated as substrate binding elements of ABC transporters. For example, OppA (43) from *Bacillus subtilis* was the first lipoprotein biochemically shown to be membrane associated and to transport an oligopeptide. In addition, AcrA (2) and MalX (21) from *Streptococcus pneumoniae* and RsbB (68) from *B. subtilis* are well-characterized substrate-binding lipoproteins of other ABC transporters. Several lipoproteins of gram-positive organisms (55) and the P37 mycoplasma lipoprotein of *Mycoplasma hyorhinzis* (18) are speculated to function in a similar manner, although their substrate specificities are unknown. Similarly, many of the putative ABC type transporter genes identified in the complete genome sequences of *M. genitalium* (20) and *M. pneumoniae* (26) encode lipoprotein components. By analogy, P78 of *M. fermentans* may be a substrate-binding protein, although its substrate binding specificity cannot be formally established by sequence homology. Studies of transporter function in mycoplasmas may be generally impaired by the lack of auxotrophic genetic markers and by the possibility raised by genome sequencing analysis (20, 26) and this study (see below) that degenerate transporter specificity may be employed to accommodate multiple substrates.

Features of the three other proteins encoded by the p78 operon support the identification of an ABC transporter and have noteworthy characteristics. Based on its high degree of sequence similarity to the ATPases of other ABC transporters and the presence of conserved nucleotide binding motifs, P63 is predicted to be the ATP-binding protein of the transporter. Previous studies suggest that proteins of this type are typically not integrally associated with membranes but rather are localized in the cytoplasm, with predicted interactions occurring with components of the translocase structure (19, 24, 42, 49). Interestingly, the P63 sequence includes an extended hydrophobic region of 55 aa capable of forming a transmembrane domain. Notably, this region separates motifs associated with the cytoplasmic function of nucleotide binding. The region also contains a central Pro residue (Fig. 2A) bounded by residues that could create an exterior loop (48, 49). This feature of the model would provide for the presence of two transmembrane domains, placing sequences both N-terminal and C-terminal of the extended hydrophobic stretch on the cytoplasmic side of the membrane. Notably, this model also adheres to the “positive inside” rule for predicting topologies of integral membrane proteins (49, 59), since multiple basic residues occur immediately before and after the extended hydrophobic region (Fig. 2A). P63 partitions into the detergent phase during fractionation with TX-114 further suggests the amphiphilic properties of this protein, as predicted by its large hydrophobic domain. P63 may be the first known example of an ATPase encoded in an ABC transport operon which is integrally bound to a membrane. However, the possibility cannot be formally ruled out that the interaction of P63 with other membrane proteins results in the partitioning of an amphiphilic complex into the detergent phase. Analogous proteins with the hydrophobic structural features of P63 have not been identified in other mycoplasma genomes analyzed to date (20, 26). Recently, the ATP-binding MalK protein in *E. coli*, overexpressed from plasmids, was shown to associate with the membrane even though the sequence of MalK does not predict a hydrophobic domain. This was proposed to occur by insertion of a conserved helical loop domain between the two transmembrane proteins of the transporter (MalG and MalF) which shield MalK from the hydrophobic nature of the membrane (52). In contrast, P63 may be able to form an integral membrane association, as distinct from the model proposed for MalK.

The components forming the translocating channel characteristic of many ABC transporters typically comprise polytopic membrane proteins containing 12 membrane-spanning segments. These may be encoded within one protein or two proteins, each with six transmembrane domains (15, 51). However, variations in this structure are known. MalF and MalG have eight and six transmembrane domains, respectively (53), and RsbC and RsbD of *B. subtilis* contain eight domains each (68). P58 and P35 of the p78 operon contain eight potential transmembrane domains each, unlike many transporters but nevertheless consistent with the formation of a channel.

Important functional outcomes of P78 phase variation were suggested by defining the underlying genetic basis of high-frequency switching in expression of this surface constituent. Changes in the length of a short poly(A) tract within the sequence coding for the region near the N terminus of the mature lipoprotein correlated precisely with the expression state of P78. Seven residues in this specific tract encoded full-length P78, which was detected by Western immunoblots and by surface-binding MAbs on colony immunoblots. Frameshift mutations resulting in a reduction to six residues in this tract would produce a prematurely truncated product, resulting in a negative surface phenotype. Notably, other short poly(A) tracts identified during sequencing of the region of the p78 gene encoding the N-terminal region of P78 remained unchanged during these phase transitions. This suggests the presence of a specific regional context within this gene that appears to be preferentially subject to insertion/deletion mutations. Examples of high-frequency insertion/deletion mutations involving homopolymeric tracts linked to specific genes have been identified in several organisms, such as in the *opc* gene in *Neisseria meningitidis* (40), the *vlp* genes in *M. hyorhinzis* (12, 69), and the *fim* gene in *Bordetella pertussis* (62), where transcriptional effects result. Several documented systems also entail frameshift mutations in ORFs of surface proteins, as reviewed in reference 45. However, this report may provide the first evidence of such a mutation affecting the expression of a primary gene product involved in an ABC transporter complex.

The expression of genes in ABC transport operons is often tightly regulated by the need and availability of a specific substrate (24). In contrast to this form of classically regulated gene expression (17), several systems in which adaptation to changing environments appears to occur through high-frequency mutations generating functional diversity within populations have been characterized. It is hypothesized that, in these systems, having elevated mutational rates in a specific subset of genes allows for particular phenotypic traits to be rapidly selected during unpredictable changes in the (in-host) environment (39, 45). For the p78 ABC transport operon, P78 is subject to a high-frequency, random mutational event that independently dictates its expression state. Although it is possible that additional transcriptional regulation of the entire operon occurs, there is currently no evidence for this. In fact, genomic sequence analysis of mycoplasmas has revealed an extremely limited repertoire of known transcriptional regulators (20, 26, 41). In this context, perhaps the most intriguing theoretical aspect of this system is the notion that the operon, by selective mutational switching of P78 expression, may encode and express multiple or alternative transport functions which could be (i) independent of the P78 product, (ii) independent of this product, or (iii) contingent on interactions with other surface proteins. An important consequence of such degeneracy would be to provide alternative transport capabilities to an organism through the use of subsets of the same genes. Although speculative at this time, mutational switching
of gene subsets associated with transport-related operons may indeed be an important evolutionary solution for mycoplasmas, in which a limited number of genes could be used for the maximal flexibility in adapting to a parasitic lifestyle that may require the exchange of unusually broad ranges of metabolites with their environment.

Finally, the immunological consequences of phase variation of surface components such as P78 may also be a significant feature of mycoplasmal adaptation. As parasitic organisms that exist in compartments accessible to the humoral immune system, mycoplasmas may have evolved through pressures dictated by immune recognition of surface components. Recognition of mycoplasma surface antigens by host Abs generated during infection can lead to growth inhibition (11), and strategies for evasion of Ab recognition through the masking of antigens by phase-variable surface components have been documented for *M. fermentans* (57) and *M. hyorhinis* (11). Proteins involved in critical surface metabolic functions, including exposed transporter components, are likely targets of deleterious immune recognition. Thus, a key function of phase variation in expression of these targets may be evasion of immune recognition through transient changes in surface-exposed targets within a propagating population. In this regard, ongoing studies in this laboratory indicate that P78 is a target antigen of human serum Abs (27). It is also possible that proposed intracellular niches for *M. fermentans* (3) may provide a protected environment which might be more effectively occupied by variant populations expressing antigenic, extracellular proteins such as P78 that confer optimal nutritional efficiency in that environment. Overall, selective phase variation of P78 may therefore have consequences both for immune evasion and for adaptation of transport function in diverse in-host environments.

**ACKNOWLEDGMENTS**

We thank Mary F. Kim for critical experimental contributions to this work.

This study was supported in part by Public Health Service grants AI32219 (K.S.W.) and T32 AI07276 (P.T.) from the National Institute of Allergy and Infectious Diseases.

**REFERENCES**