INFECTION AND IMMUNITY, Oct. 1996, p. 3967–3974 Vol. 64, No. 10
0019-9567/96/$04.00+0

Identification, Cloning, and Sequencing of the Immunoglobulin A1 Protease Gene of *Streptococcus pneumoniae*

JAVAID H. WANI,[1] JOANNE V. GILBERT,[2] ANDREW G. PLAUT,[2] AND JEFFREY N. WEISER[1,3]*

Department of Pediatrics, Children's Hospital of Philadelphia,[1] and Department of Microbiology, University of Pennsylvania School of Medicine,[3] Philadelphia, Pennsylvania, and Department of Medicine, Tufts University-New England Medical Center, Boston, Massachusetts[2]

Received 6 March 1996/Returned for modification 5 June 1996/Accepted 2 July 1996

The pneumococcus expresses a protease that hydrolyzes human immunoglobulin A1 (IgA1). A gene for IgA1 protease was identified from a plasmid library of pneumococcal DNA because of the effect of its overexpression on the colony morphology of *Streptococcus pneumoniae*. The deduced 1,964-amino-acid sequence is highly homologous to that of the IgA1 protease from *Streptococcus sanguis*. The similarity to the *S. sanguis* enzyme and the presence of a putative zinc-binding site suggest that the pneumococcal enzyme is a metalloprotease. The two streptococcal sequences differ in a hydrophilic region with 10 tandem repeats of a 20-mer in *S. sanguis*, which is replaced by a similar but less repetitive sequence in *S. pneumoniae*. Antiserum reactive with the pneumococcal IgA1 protease was used to demonstrate that the majority of the protein is cell associated. The expression and function of this gene were confirmed by insertional mutagenesis. Interruption of the chromosomal gene resulted in loss of expression of an approximately 200-kDa protein and complete elimination of detectable IgA1 protease activity.

Streptococcus pneumoniae, the pneumococcus, efficiently colonizes the human nasopharynx. Colonization of the mucosal surface is the initial step in the pathogenesis of respiratory tract infection, for which the pneumococcus is a leading etiologic agent (40). This organism appears to be highly adapted to its human host, since carriage by other host species is unusual. An important factor in its ability to become established on mucosal surfaces is the effective evasion of local host defenses. The pneumococcus is one of many pathogens colonizing the human respiratory tract that produce a proteolytic enzyme which specifically cleaves the heavy chain of human immunoglobulin A1 (IgA1), the predominant class of immunoglobulin present on mucosal membranes (11, 12, 18). IgA1 protease has been characterized for a number of bacteria, including *Neisseria gonorrhoeae* (28), *Neisseria meningitidis* (13, 27), *Haemophilus influenzae* (4, 29), *Bacteroides melanogenicus* (21), and *Streptococcus sanguis* (7, 8). Although each of these enzymes hydrolyzes human IgA1, these proteases represent several classes of molecules with different catalytic mechanisms (serine protease, thiol-activated protease, and metalloprotease) and distinct sites of cleavage within the heavy-chain hinge region. The independent evolution of several distinct classes of enzymes with a similar biological function points to the importance of inactivation of IgA1 for the colonization of mucosal pathogens.

There is little known about the IgA1 protease in the pneumococcus. A recent survey of 114 isolates of *S. pneumoniae* revealed that at least 94% demonstrated the ability to hydrolyze human IgA1 (17). Like *S. sanguis* and *Streptococcus mitior*, *S. pneumoniae* cleaves a single peptide bond between Pro-227 and Thr-228 in the heavy chains of human and great ape IgA1 molecules, the only known substrates for this enzyme (7, 30). The protease in *S. pneumoniae* demonstrates marked antigenic variation, since at least 17 antigenic forms have been identified (17). The gene for the IgA1 protease from another streptococcus (*S. sanguis*), a metalloprotease, has been recently identified and characterized (7, 8). In this paper we report the identification and characterization of the IgA1 protease gene of the pneumococcus and describe some of its salient features.

* Corresponding author. Mailing address: University of Pennsylvania School of Medicine, Department of Microbiology, 802A Abramson Research Building, 34th St. and Civic Center Blvd., Philadelphia, PA 19104. Phone: (215) 573-3511. Fax: (215) 573-9068. Electronic mail address: Weiser@mail.med.upenn.edu.

MATERIALS AND METHODS

Bacterial strains, growth conditions, and media. Strains R6x and P110 are derived from strain R6 and have been described previously (31, 38). Pneumococci were grown in semisynthetic medium (C+Y medium, pH 8.0) at 37°C without shaking (16). Broth cultures were plated on tryptic soy plates containing 1% agar onto which 100 μl of catalase (5,000 U) (Worthington Biochemical Co., Freehold, N.J.) was added. Cultures were grown at 37°C overnight in a candle extinction jar, which provided an atmosphere of increased CO_2 necessary for optimal growth on this medium. *Escherichia coli* was grown in Luria-Bertani broth with or without 1% agar. Colony morphology was assessed with a stereozoom microscope with oblique transmitted illumination as described previously (44). When required, antibiotics were added to the media at the following concentrations: kanamycin, 50 μg/ml; erythromycin, 500 μg/ml (for *E. coli*) or 1 μg/ml (for *S. pneumoniae*).

Reagents and antibodies. Restriction endonucleases and DNA modification enzymes were purchased from New England Biolabs, Beverly, Mass. [α-^{32}P]dCTP was obtained from Amersham, Arlington Heights, Ill. All chemicals were purchased from Sigma Chemical Co., St. Louis, Mo., unless stated otherwise. Polyclonal antiserum to formalin-killed whole pneumococci from strain P6, a serotype 9V clinical isolate, was raised in rabbits. To obtain antiserum with higher specificity for the pneumococcal IgA1 protease, the antiserum was absorbed six times for 24 h at 4°C with the *iga* mutant P262.

Construction and screening of a genomic library. Pneumococcal genomic DNA was prepared by a published method (24). Chromosomal DNA from P110 was partially digested with *Sau*3AI and size fractionated on a 10 to 40% sucrose density gradient by centrifugation at 100,000 × *g* for 22 h (32). The fractions containing DNA fragments ranging in size from 3 to 8 kb were pooled, ethanol precipitated, and used for ligation into *Bam*HI-digested shuttle plasmid pMU1328 (1). The ligation mixture was then transformed into *E. coli* DH5α and selected on Luria-Bertani agar containing erythromycin (500 μg ml^{-1}). Competent R6x cells were transformed with DNA from pooled plasmids as described previously (16). Erythromycin-resistant transformants were screened for a colony morphology that differed from that of the R6x parent. Plasmids conferring altered colony morphology were extracted from *S. pneumoniae* along with genomic DNA and transformed into *E. coli* for further analysis.

Recombinant DNA methods. DNA sequencing was performed by the dideoxy chain termination method (33). Sequence analysis was carried out with the Genetics Computer Group software package from the University of Wisconsin (5). Databases were searched by using the BLAST program through the National

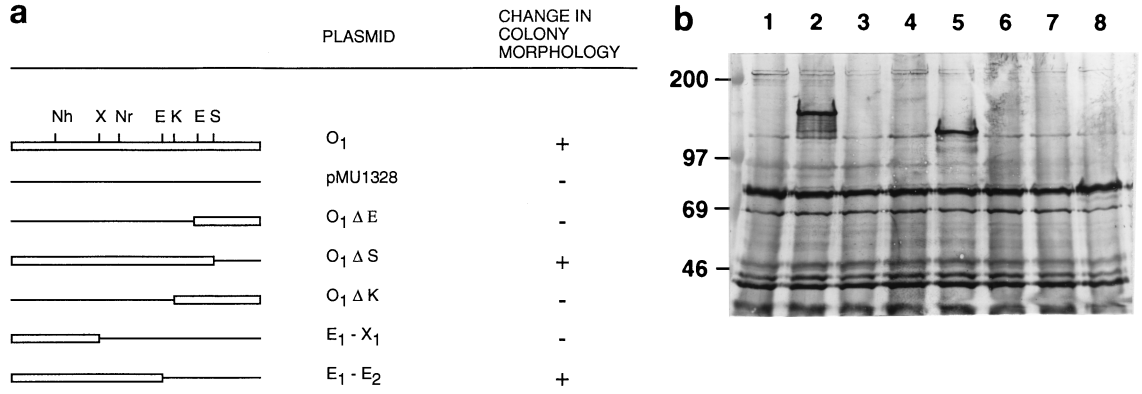

FIG. 1. (a) The top line shows the restriction endonuclease map of the 3.15-kb fragment from strain P110 cloned in plasmid pMU1328, which confers altered colony morphology on strain R6x (O_1). Plasmids in which portions of this insert were deleted were also tested for their ability to change (+) or not change (−) colony morphology when transformed into strain R6x. DNA fragments deleted from this 3.15-kb region are indicated by single lines. Restriction endonucleases: E, *Eco*RI; K, *Kpn*I; Nh, *Nhe*I; Nr, *Nru*I; S, *Sal*I; X, *Xba*I. (b) Western blot analysis of whole-cell lysates of pneumococcal strain R6x transformed with pMU1328 carrying various fragments of the 5′ region of the *iga* gene. An antiserum raised against heat-killed pneumococci was used to detect proteins expressed by plasmids which were not present in the host strain. The restriction maps of the plasmid constructs tested in lanes 2 to 8 are shown in panel a. Lanes: 1, R6x (host strain); 2, R6x/O_1; 3, R6x/pMU1328 (no insert); 4, R6x/O_1ΔE; 5, R6x/O_1ΔS; 6, R6x/O_1ΔK; 7, R6x/E_1-X_1; 8, R6x/E_1-E_2. Plasmids O_1, O_1ΔS, and E_1-E_2 show overexpression of a protein not seen in the host strain. Molecular mass standards (in kilodaltons) are shown on the left.

Center for Biotechnology Information (2). The 3′ end of the *iga* gene of strain P110 was obtained by PCR. A fragment of the pneumococcal genome from an unrelated strain with IgA1 protease activity had been previously cloned and partially sequenced, including the 3′ end of the gene. A primer (5′-TCCAAAG CGACAGTAACC-3′) from within the sequence of the P110 *iga* gene and a primer (5′-CTTGGTTTGTTCTTCATCAC-3′) which is 3′ to the termination codon on the opposite strand were used to obtain a 3.3-kb PCR product from the P110 chromosome. The amplified product was cloned into the vector pCRII (Invitrogen, San Diego, Calif.) according to the manufacturer's instructions. The sequence was then extended beyond the primer 3′ to the termination codon by inverse PCR (32).

Western blotting (immunoblotting). Bacterial cells were adjusted to equal cell densities on the basis of optical density at 620 nm. The cell pellet was washed with phosphate-buffered saline (pH 8.0) and resuspended in 1/10 the original culture volume in loading buffer consisting of Tris-HCl (0.0625 M; pH 8.0), glycerol (10%), sodium dodecyl sulfate (SDS) (2%), 2-mercaptoethanol (5%), and bromophenol blue (0.00125%). The culture supernatant was mixed with 5× loading dye consisting of glycerol (50%), SDS (15%), 2-mercaptoethanol (15%), and bromophenol blue (1.5%). A 20-fold-concentrated culture supernatant was obtained by centrifugation in a Centriprep-50 (Amicon, Beverly, Mass.), which retains proteins of greater than 50,000 Da. The cell fraction and the corresponding amount of the supernatant fraction were heated to 100°C for 5 min prior to separation by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on 7.5% acrylamide separating gels. Electrotransfer of proteins onto Immobilon-P (Millipore Co., Bedford, Mass.) and Western blotting were carried out as described previously except that the membrane was immersed in methanol prior to being wetted (39). Immunoblotting of membranes was carried out as follows. Membranes were first washed twice with Tris-saline blocking buffer (TSBB) (10 mM Tris-HCl [pH 8.0], 0.5 M NaCl, 0.5% Tween 20, and 0.02% sodium azide) and incubated for 16 h with a 1/5,000 dilution of antiserum. After five washes in Tris-saline blocking buffer, goat anti-rabbit immunoglobulin G antibody conjugated to alkaline phosphatase (Bio-Rad Laboratories, Hercules, Calif.) was added, and the membrane was incubated for another 2 h. Reactivity was visualized after an additional five washes with Tris-saline blocking buffer as described previously (32).

Mutagenesis of IgA1 protease gene. The *ermC* gene obtained from pIM13, isolated as a *Cla*I-*Hin*dIII fragment, was cloned into pBR322 digested with *Cla*I-*Hin*dIII (20). The construct, pE169, served as a source of the erythromycin resistance cassette. A 1.9-kb *Eco*RI fragment encompassing the 5′ portion of the *iga* gene was cloned into the *Eco*RI site in pHSG399 (37). An internal 777-bp *Nhe*I-*Nru*I fragment within this *Eco*RI insert was deleted, and the erythromycin resistance cassette from pE169 digested with *Nhe*I and *Sca*I was inserted into the deletion. The construct was linearized by digestion with *Bam*HI, which cleaves within the vector sequence, and transformed into R6x with transformants selected on erythromycin. The mutant with the deletion-insertion mutation in the R6x chromosome was designated P262.

Southern hybridization. Interruption of the chromosomal gene in P262 was confirmed by Southern hybridization. Electrophoretically separated chromosomal DNAs from R6x and P262 were digested with *Eco*RI or *Hin*dIII and transferred bidirectionally to Hybond-N (Amersham). Hybridization under high-stringency conditions with an *Eco*RI-*Xba*I fragment from the *iga* gene labeled with [α-^{32}P]dCTP by nick translation was carried out as described previously (32).

Assay of IgA1 protease activity. The pneumococci were grown overnight in Todd-Hewitt broth supplemented with 0.5% yeast extract (Difco Laboratories, Detroit, Mich.). The clarified supernatants were subjected to 60% ammonium sulfate precipitation, which resulted in 25-fold concentration of proteins. The salt was removed by thorough dialysis against Tris-HCl buffer (50 mM; pH 8.0). IgA1 protease activity was assayed by incubating the dialyzed supernatant with a human monoclonal immunoglobulin A1 substrate in Tris buffer containing 20 mM $CaCl_2$ and 20 mM $MgCl_2$. After 3 h of incubation at 37°C, the mixtures were subjected to SDS-PAGE on a 10% acrylamide separating gel, and protein bands were visualized by Coomassie brilliant blue staining.

Nucleotide sequence accession number. The nucleotide sequence from which the amino acid sequence shown in Fig. 3 was deduced has been submitted to GenBank and assigned accession number U47687.

RESULTS

Identification of the IgA1 protease gene. A genomic library of *S. pneumoniae* DNA in a streptococcal-*E. coli* shuttle plasmid, pMU1328, was screened for clones capable of affecting the colony morphology of the pneumococcus. The source of DNA for the library, strain P110, is a spontaneous, very transparent variant of strain R6 (1, 31). In screening transformants of strain R6x to identify a genomic fragment conferring the very transparent phenotype of the donor DNA, transformants with an opaque phenotype were noted in addition to the expected transparent transformants. These opaque colonies, however, differed from spontaneous opaque variants previously described by our group because they appeared umbilicated as a result of autolysis (31, 47). Several colonies with this altered phenotype from different transformation experiments were chosen. Genomic DNA was extracted from these transformants and used to retransform R6x (backcross). Plasmids capable of conferring the altered phenotype when transformed back into the recipient pneumococcal strain were then isolated for further analysis by transformation into *E. coli*. Restriction endonuclease mapping of four such plasmids indicated that the inserts were overlapping over a span of 3.1 kb.

The region within this 3.1 kb that affected colony morphology was defined by testing a series of internal deletions within the plasmid insert for their ability to alter colony characteristics when transformed into R6x (Fig. 1a). The minimum insert fragment capable of conferring altered colony morphology was

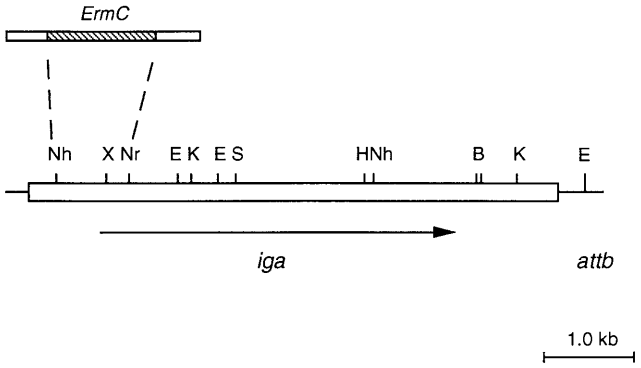

FIG. 2. Restriction map of a 6.4-kb region (open box) from strain P110 which spans the entire *iga* gene. The orientation of the gene is indicated by an arrow. Above the restriction map is the *Eco*RI fragment from the 5′ end of the gene interrupted by insertion of the *ermC* gene. Mutagenesis of the chromosomal *iga* gene was carried out by transforming the construct with the *ermC* gene flanked by fragments of the *iga* gene into strain R6x. Restriction endonucleases: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; Nh, *Nhe*I; Nr, *Nru*I; S, *Sal*I; X, *Xba*I.

1.9 kb in length at the 5′ end of the 3.1-kb region. Altered colony morphology was associated with overexpression of a protein requiring this 5′ region (Fig. 1b). Deletions involving the 5′ end of the 3.1-kb insert resulted in a loss of the ability of the insert to confer altered colony morphology. The minimum fragment of 1.9 kb required for overexpression of this protein was sequenced (Fig. 1b, lane 8). Analysis of this sequence revealed the 5′ end of a single long open reading frame of 1.6 kb encoding a putative N terminus of 543 amino acids. This N-terminal fragment was extended to 964 amino acids by extending the sequence to the 3′ end of the 3.1-kb insert. The 964-amino-acid N-terminal sequence showed similarity (64% identity) to the N-terminal 854 amino acids of the 1,878 amino acids of the IgA1 protease of *S. sanguis* (7). This suggested that expression of an N-terminal 543 amino acids of this protein, a putative IgA1 protease of *S. pneumoniae*, was sufficient to affect the colony morphology of strain R6x when expressed on the multicopy plasmid pMU1328.

Cloning of the complete IgA1 protease gene. The 3′ end of the IgA1 protease gene of *S. pneumoniae* was cloned separately as follows. A DNA fragment encoding the IgA1 protease activity from an unrelated pneumococcal strain had been cloned and partially sequenced (unpublished data). In order to clone the remainder of the gene from strain P110, sequence from the 3′ end of the gene from the other strain was used to construct a PCR primer to amplify P110 chromosomal DNA. The 3′ end of the P110 gene on a 3.3-kb PCR fragment was cloned and sequenced. The overlapping nucleotide sequences of the two clones from strain P110 put together had 6,415 bp containing a single open reading of 5,891 bp (Fig. 2), which has been designated *iga* on the basis of the similarity of its deduced amino acid sequence to the *S. sanguis* IgA1 protease sequence.

Characteristics of the pneumococcal *iga* gene. Near the 5′ end of the open reading frame was an ATG codon located four bp from the sequence GGAGG, which could function as a ribosome binding site (36). The open reading frame would encode 1,964 amino acids from this initiation codon (Fig. 3). Following the TAG termination codon, another 262 bp of nucleotide sequence was obtained by inverse PCR. This sequence overlaps the sequence adjacent to the unique attachment site for conjugative transposon Tn*5252*, suggesting that this feature is 3′ to the *iga* gene (42). The sequence following the *iga* gene did not contain inverted repeat sequences that resemble transcriptional terminators.

The entire 1,964 amino acids of the *S. pneumoniae iga* product and the IgA1 protease of *S. sanguis* were compared (Fig. 4). The deduced amino acid sequence of the pneumococcal gene, *iga*, is 75% similar and 64% identical to the sequence of the 1,878-amino-acid IgA1 protease of *S. sanguis* (8). Only three regions in the pneumococcal sequence, from amino acids 193 to 292 and 394 to 674 and the C terminus from position 1942, showed relatively less similarity to the sequence from *S. sanguis*. The first region of divergence, in which the pneumococcal sequence contains an additional approximately 60 residues, encompasses the N terminus of the *S. sanguis* protein when expressed in *E. coli* (8). The second region of divergence encompasses the region in *S. sanguis* with 10 nearly identical tandem repeats of a 20-mer of unknown function. Although an exact copy of these repeats is not found in the pneumococcal *iga* gene, the corresponding region contains 11 sequences matching as many as 10 of 20 amino acids in the *S. sanguis* consensus repetitive sequence (Fig. 5). A valine residue in position 1, a proline in position 3, and a glycine between positions 10 and 18 are particularly conserved in this region in the pneumococcal protease. The Kyte-Doolittle hydropathy plot of the amino acid sequence deduced from the entire open reading frame of the *iga* gene of *S. pneumoniae* showed that the region containing this repetitive feature is hydrophilic compared with the remainder of the protein (Fig. 6) (14).

There were three hydrophobic regions near the N terminus, from amino acids 10 to 45, 103 to 125, and 133 to 162. The first hydrophobic region could function as a typical prokaryotic signal sequence (43). The second of the hydrophobic domains is located between the sequence L-P-N-T-G-S and a highly charged, lysine-rich region, K-K-K-V-K-N-K. The sequence L-P-N-T-G-S followed by a hydrophobic stretch of 24 residues and charged tail resembles the anchoring domain common to many gram-positive surface proteins (6, 34). However, this feature, which is also found in the corresponding region in *S. sanguis*, occurs near the N terminus rather than in its more typical location at the C terminus. In addition, there is a single hydrophobic region in the remainder of the sequence from amino acid 1453 to 1473, which could serve as a transmembrane domain. The pneumococcal protease, like the *S. sanguis* protease, has the sequence H-E-M-T-H (positions 1605 to 1609 in the pneumococcal protease) with a highly conserved glutamic acid 20 amino acids carboxy terminal to the histidine. This motif matches the internal zinc-binding consensus sequence (H-x-x-T-H) found in bacterial metalloproteases (10, 41). The pneumococcal gene product, but not the *S. sanguis* gene product, includes the sequence, G-Y-I-K-E-G-K-T, from amino acid 187 to 194, which resembles the P-loop ATP/GTP-binding consensus sequence (GxxxxGKT) (19). The deduced amino acid sequence of the pneumococcal gene did not demonstrate significant homology with IgA1 proteases of bacterial species other than *S. sanguis*.

Mutagenesis of the IgA1 protease gene. Mutagenesis of the pneumococcal *iga* gene was carried out to confirm that it is expressed and that it encodes an IgA1 protease. The *iga* gene was interrupted by replacing an internal 777-bp *Nhe*I-*Nru*I fragment located 234 bp 3′ to the beginning of the open reading frame with a 1.2-kb erythromycin resistance cassette (*ermC*) (Fig. 2). This construct on a plasmid which is unable to replicate in pneumococcus was linearized and transformed into pneumococcal strain R6x. Allelic exchange resulting from a crossover event on both sides of the marker gene was obtained by selecting for resistance to erythromycin. Interruption of the *iga* gene near its 5′ end in the R6x chromosome was

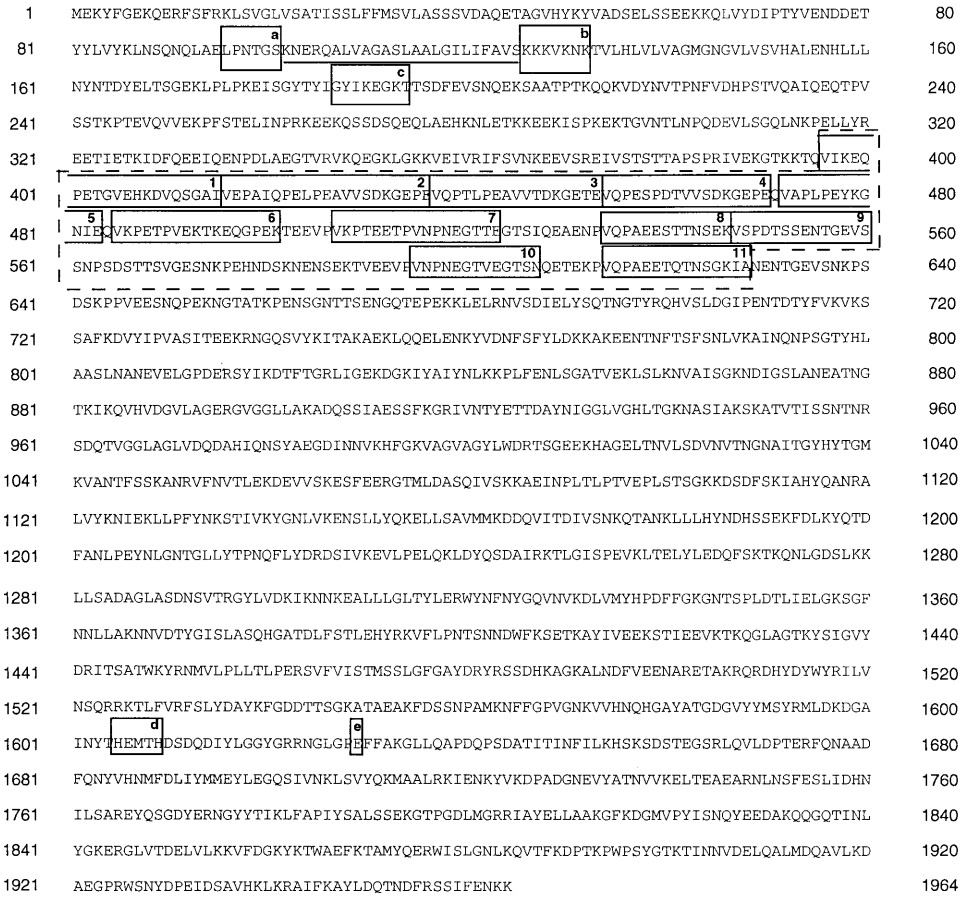

FIG. 3. The 1,964-amino-acid sequence of the IgA1 protease of *S. pneumoniae* P110 deduced from the nucleotide sequence of the *iga* gene. A region near the N terminus that resembles the C-terminal anchor for many other gram-positive surface proteins is divided into three contiguous components, an LPNTGS motif (box a), a hydrophobic domain (underlined), and a lysine-rich charged sequence (box b). An ATP/GTP-binding consensus sequence is shown in box c. The sequence in box d and the downstream glutamic acid residue in box e constitute a zinc-binding domain. A hydrophilic region with similarity to the multiple tandem repeats in the IgA1 protease of *S. sanguis* is shown (dashed box), with each sequence with similarity in boxes 1 through 11.

confirmed by Southern blot analysis with probes spanning the deleted fragment (data not shown). This construct would result in the formation of a truncated gene product after translation of the first approximately 75 of 1,964 amino acids. A single deletion-insertion mutant, P262, was selected for further analysis.

Western blot analysis of whole-cell lysates was used to demonstrate that the *iga* gene on the pneumococcal chromosome is expressed (Fig. 7). Antiserum raised to whole pneumococci reacted with an approximately 200-kDa band in R6x. This is close to the 219-kDa predicted size of the gene product based on its deduced amino acid sequence. Pneumococcal mutant strain, P262, in which the *iga* gene was interrupted, did not express this band, indicating that it represents the *iga* gene product. The specificity of the antiserum for this protein was further improved by serial absorption with P262.

The ability to detect the *iga* gene product with the absorbed antiserum was used to assess whether the protein is exported, as is the case for other known IgA1 proteases (25). Following growth to mid-log phase, cultures divided between cell and supernatant fractions were examined in Western blots. Although the 200-kDa protein could be identified in the supernatant fraction, the overwhelming majority of the protein remained cell associated. Concentration of proteins in the culture supernatant fraction confirmed that the *iga* gene product is present, albeit in a low concentration in this fraction. The lack of the protein in the concentrated culture supernatant from P262 demonstrated that the absence of cell-associated protein in the mutant was not caused by altered secretion.

The ability of the *iga* gene product to hydrolyze its natural substrate, human IgA1, was determined by comparing the deletion-insertion mutant and its parent strain (Fig. 8). Culture supernatant was incubated with human monoclonal immunoglobulin A1 substrate, and hydrolysis of the heavy chain was determined by resolving its proteolytic products by SDS-PAGE. Interruption of the *iga* gene in P262 was associated with a complete loss of human IgA1-specific proteolytic activity in culture supernatants. This result confirmed that the *iga* gene encodes an IgA1 protease and that this gene is responsible for the ability of the pneumococcus to cleave the heavy chain of human IgA1.

The *iga* gene was identified because of its effect on colony morphology when expressed on a multicopy plasmid in *S. pneumoniae*. The colony morphology of strain P262 was compared with that of its parent strain to determine whether interruption of the chromosomal gene also affected colony morphology. There was, however, no apparent difference in colony morphology between R6x and P262. This suggests that the pneumococcal IgA1 protease is not a significant factor in the ex-

FIG. 4. Comparison of the deduced amino acid sequences of the *iga* gene product of *S. pneumoniae* P110 (upper sequence) and the IgA1 protease of *S. sanguis* (lower sequence). The Bestfit program of the Genetics Computer Group package was used, and symbols (|, :, and .) represent degrees of similarity based on the Dayhoff PAM-250 matrix.

pression of varied colony phenotypes described by this and other groups (44).

DISCUSSION

This report describes the identification, cloning, and sequencing of a gene encoding a pneumococcal cell surface protein, the IgA1 protease. This gene was identified by an unusual method involving the screening of a genomic library for sequences mediating observable changes in colony morphology. Our group has demonstrated that expression of a number of cell surface components can alter colony morphology as observed on transparent surfaces with oblique, transmitted illumination (45, 47). Colony morphology is a function of how visible light passing through a colony is affected by the packing of organisms within the colony (3). Some cell surface products may affect the ability of cells to associate in a colony and thereby contribute to the optical properties of the colony. The ability of cloned DNA to alter the optical properties of the cell has been used as a sensitive means of isolating a number of genes through the screening of individual transformants (22, 31, 46). This approach, however, has been applied exclusively to genes which encode cell surface components that contribute to the colony phenotype. The identification of the pneumococ-

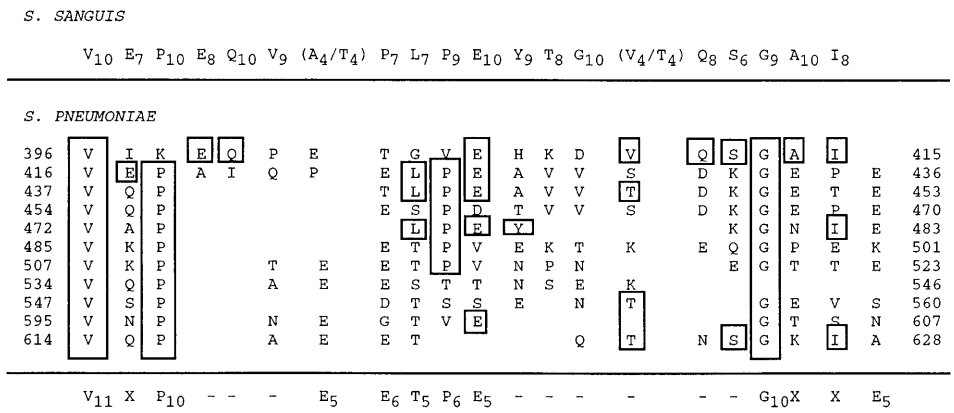

FIG. 5. The consensus sequence for the 10 nearly identical tandem repeats of 20 amino acids in *S. sanguis* is shown at the top. The frequency of occurrence of the most common amino acid in each position is indicated by a subscript. Eleven sequences in the *iga* product of *S. pneumoniae* between amino acids 396 and 627 that resemble the *S. sanguis* repeat region are listed below the *S. sanguis* sequence, with their positions shown. Residues in the pneumococcal *iga* product that match the *S. sanguis* repeat region consensus above are boxed, and the most highly conserved amino acids in this *S. pneumoniae* region are listed at the bottom with their frequencies indicated by subscripts.

cal IgA1 protease gene described in this paper demonstrates that genes encoding surface products which do not appear to contribute to colony morphology may also be isolated by screening transformants for differences in colony characteristics. In this study the colony morphology of strain R6x was modified by introduction of the 5′ half of the IgA1 protease gene, even though the chromosomal copy of this gene was shown to have no discernible effect on colony morphology. The ability of the gene fragment to affect colony morphology was most likely because of the expression of the cell surface gene product on a multicopy plasmid at levels far greater than that for the chromosomal gene. The *iga* gene fragments which altered colony morphology were associated with overexpression of proteins of the predicted size on Western analysis (Fig. 1b).

The screening technique employed in this study led to the isolation of a partial gene. The function of this gene was suggested by its sequence similarity to the IgA1 protease gene of another streptococcal species. The IgA1 protease gene of *S. pneumoniae* had not previously been characterized. Because of the likely importance of this gene in the pathogenesis of pneumococcal infection, the remainder of the gene was isolated from the same strain and the entire 6.0-kb gene was sequenced. Its biological function was confirmed by mutagenesis of the chromosomal gene, which completely abolished IgA1 protease activity present in culture supernatant.

The IgA1 protease in *S. sanguis* had been classified as a metalloprotease because its activity was inhibited by EDTA (15, 26). On the basis of the sequence homology between the *S. sanguis* and *S. pneumoniae* gene products it is likely that the pneumococcal IgA1 protease is also a metalloprotease. Further evidence that the pneumococcal enzyme is a metalloprotease comes from the identification of a consensus zinc-binding site (10). This sequence is conserved in both streptococcal IgA1 proteases, and site-directed mutagenesis of this sequence in *S. sanguis* resulted in the loss of catalytic activity, although the protein was full length (8). X-ray crystallography of other bacterial metalloproteases has shown that this pentapeptide and the glutamic acid residue 20 residues downstream constitute the zinc-binding site (10, 41).

A highly repetitive region found in the *S. sanguis* protein is particularly poorly conserved in the *S. pneumoniae* IgA1 protease. Highly repetitive regions are common among gram-positive surface proteins and often exhibit strain-to-strain variation in length and amino acid sequence (9). Differences in this hydrophilic region in *iga* may account for the lack of immunological cross-reactivity between streptococcal enzymes and may contribute to the marked antigenic variation within the pneumococcal IgA1 protease (17).

Western analysis demonstrated that the *iga* gene in *S. pneumoniae* expresses a protein that is predominately cell associ-

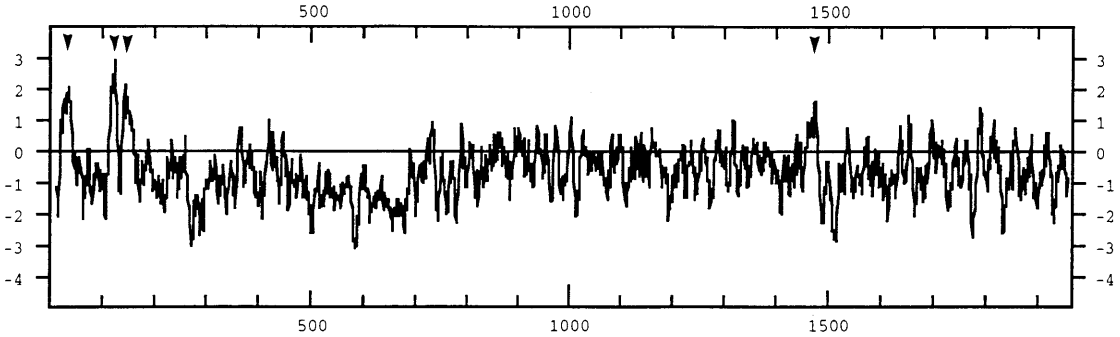

FIG. 6. Kyte-Doolittle hydropathy plot of the deduced 1,964-amino-acid sequence of the *iga* gene of *S. pneumoniae*. Hydrophobic regions are above the line, and the hydrophobic index is indicated on the vertical axis. Numbers on the horizontal axis refer to the position in open reading frame beginning with the putative initiation codon. Four hydrophobic regions referred to in the text are indicated by arrowheads.

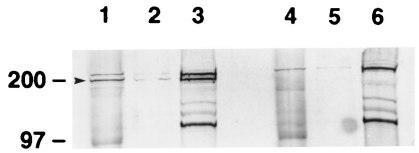

FIG. 7. Western blot analysis comparing high-molecular-mass proteins reacting with an antiserum raised against whole pneumococci. A band of approximately 200 kDa (arrowhead) in a whole-cell lysate of the parent strain, R6x (lane 1), is not detected in whole-cell lysates of the deletion-insertion mutant P262 (lane 4). The corresponding culture supernatant samples (R6x, lane 2; P262, lane 5) showed minimal amounts of this protein in the parent strain. The protein was apparent in a 20-fold-concentrated culture supernatant fraction of the parent strain (R6x, lane 3; P262, lane 6). Molecular mass standards (in kilodaltons) are shown on the left.

ated but is also present in smaller amounts in the culture supernatant. The cell-associated and secreted proteins have the same apparent approximate molecular mass of 200 kDa. Since this is close to the predicted size based on the deduced amino acid sequence, there does not appear to be extensive processing of the protein as occurs in the IgA1 proteases of other pathogens (28). This is also close to the size of the mature protein in *S. sanguis* (7). Expression of the *S. sanguis* gene in *E. coli*, however, resulted in a smaller protein with the N terminus of M-D-K-E-A-L-N-Q-N, which is 210 residues from the first methionine in the open reading frame. No similar sequence is found in the corresponding region of the *S. pneumoniae* protein. The amino acid sequence beginning with the upstream methionine, however, is highly conserved between the two species. The conservation of the deduced upstream sequence indicates that it is most likely included in the expressed enzyme in streptococci. In both species this upstream region contains the sequence L-P-x-T-G-x, followed by a 24-amino-acid hydrophobic stretch which is followed by a lysine-rich, highly charged domain. This feature matches the cell wall anchor common to many gram-positive bacterial surface proteins, although these are always located at the C terminus (23, 35). It has been hypothesized that the N-terminal part of the protein at threonine in the hexapeptide becomes attached to the peptidoglycan, while the remainder of the protein remains cytoplasmic (34). In the case of the streptococcal IgA1 proteases, this would leave 1,865 amino acids within the cell, which does not appear to be the case. Therefore, the significance of this feature in this atypical location in the protein is unclear. It was observed, however, that the IgA1 protease expressed in the pneumococcus is predominately cell associated. The specific mechanism by which the protein associates with the cell will be addressed in subsequent investigations.

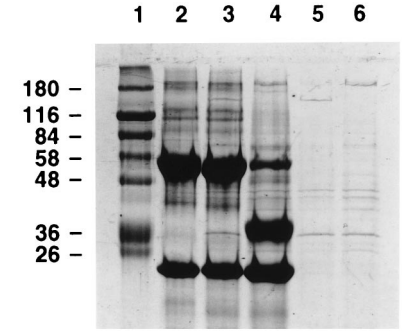

FIG. 8. Assay of IgA1 protease activity in culture supernatants from the parent strain, R6x, and the *iga* deletion-insertion mutant P262. Lanes: 1, molecular mass markers (in kilodaltons); 2, IgA1 plus buffer; 3, IgA1 plus P262 culture supernatant; 4, IgA1 plus R6x culture supernatant; 5, P262 culture supernatant; 6, R6x culture supernatant. In lane 4, the approximately 36-kDa band is the Fc fragment of the cleaved IgA1, while a Fd fragment (α chain section in the Fab product) comigrates with the light chains.

ACKNOWLEDGMENTS

J.N.W. is a Lucille P. Markey Charitable Trust Scholar. This work was supported by grants from the Lucille P. Markey Charitable Trust and from NIH (AI38446) (to J.N.W.). The work at Tufts-New England Medical Center Hospital was supported by NIH grant NIDR DEO9677, GRASP Digestive Disease Core Research Center grant P30 DK34928, and the Nutricia Corporation, Zoetermeer, The Netherlands.

REFERENCES

1. **Achen, M. G., B. E. Davidson, and A. J. Hillier.** 1986. Construction of plasmid vectors for the detection of streptococcal promoters. Gene **45:**45–49.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Braun, W.** 1965. Bacterial genetics. W. B. Saunders Co., Philadelphia.
4. **Bricker, J., M. H. Mulks, A. G. Plaut, E. R. Moxom, and A. Wright.** 1983. IgA proteases of *Haemophilus influenzae*: cloning and characterization in *E. coli* K-12. Proc. Natl. Acad. Sci. USA **80:**2681–2685.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
6. **Fischetti, V. A., V. Pancholi, and O. Schneewind.** 1990. Conservation of a hexapeptide sequence in the anchor region of surface proteins from Gram-positive cocci. Mol. Microbiol. **4:**1603–1605.
7. **Gilbert, J. V., A. G. Plaut, Y. Fishman, and A. Wright.** 1988. Cloning of the gene encoding streptococcal immunoglobulin A protease and its expression in *Escherichia coli*. Infect. Immun. **56:**1961–1966.
8. **Gilbert, J. V., A. G. Plaut, and A. Wright.** 1991. Analysis of the immunoglobulin A protease gene of *Streptococcus sanguis*. Infect. Immun. **59:**7–17.
9. **Hollinshead, S., V. A. Fischetti, and J. R. Scott.** 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. **207:**196–203.
10. **Jongeneel, C. V., J. Bouvier, and A. Bairoch.** 1989. A unique signature identifies a family of zinc-dependent metallopeptidases. FEBS Lett. **242:** 196–203.
11. **Kett, K., P. Brandtzaeg, J. Radl, and J. J. Haaijman.** 1986. Different subclass distribution of IgA-producing cells in human lymphoid organs and various secretory tissues. J. Immunol. **136:**3631–3635.
12. **Kilian, M., J. Mestecky, and R. E. Schrohenloher.** 1979. Pathogenic species of the genus *Haemophilus* and *Streptococcus pneumoniae* produce immunoglobulin A1 protease. Infect. Immun. **26:**143–149.
13. **Koomey, J. M., and S. S. Falkow.** 1984. Nucleotide sequence homology between the immunoglobulin A1 protease genes of *Neisseria gonorrhoeae*, *N. meningitidis*, and *Haemophilus influenzae*. Infect. Immun. **43:**101–107.
14. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
15. **Labib, R. S., N. J. Calvanico, and T. B. J. Tomasi.** 1978. Studies on the extracellular proteases of *Streptococcus sanguis*: purification and characterization of a human IgA1 specific protease. Biochim. Biophys. Acta **526:**547–559.
16. **Lacks, S., and R. D. Hotchkiss.** 1960. A study of the genetic material determining an enzyme activity in pneumococcus. Biochim. Biophys. Acta **39:**508–517.
17. **Lomholt, H.** 1995. Evidence of recombination and an antigenically diverse immunoglobulin A1 protease among strains of *Streptococcus pneumoniae*. Infect. Immun. **63:**4238–4243.
18. **Male, C. L.** 1979. Immunoglobulin A1 production by *Haemophilus influenzae* and *Streptococcus pneumoniae*. Infect. Immun. **62:**254–261.
19. **Matti, S., P. R. Sibbald, and A. Wittinghofer.** 1990. The P-loop—a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. **15:**430–434.
20. **Monod, M., C. Denoya, and D. Dubnau.** 1986. Sequence and properties of pIM13, a macrolide-lincosamide-streptogramin B resistance plasmid from *Bacillus subtilis*. J. Bacteriol. **167:**138–147.
21. **Mortensen, S. B., and M. Kilian.** 1984. Purification and characterization of an immunoglobulin A1 protease from *Bacteroides melaninogenicus*. Infect. Immun. **45:**550–557.
22. **Moxon, E. R., R. A. Deich, and C. Connelly.** 1984. Cloning of chromosomal DNA from *Haemophilus influenzae*: its use for studying the expression of the type b capsule and virulence. J. Clin. Invest. **73:**298–306.
23. **Navarre, W. W., and O. Schneewind.** 1994. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in Gram-positive bacteria. Mol. Microbiol. **14:**115–121.
24. **Pearce, B. J., Y. B. Yin, and H. R. Masure.** 1993. Genetic identification of

exported proteins in *Streptococcus pneumoniae*. Mol. Microbiol. **9:**1037–1050.
25. **Plaut, A. G.** 1983. The IgA1 proteases of pathogenic bacteria. Annu. Rev. Microbiol. **37:**603–622.
26. **Plaut, A. G., R. J. Genco, and T. B. Tomasi, Jr.** 1974. Isolation of an enzyme from *Streptococcus sanguis* which specifically cleaves IgA. J. Immunol. **113:** 289–291.
27. **Plaut, A. G., J. V. Gilbert, M. S. Artenstein, and J. D. Capra.** 1975. *Neisseria gonorrhoeae* and *N. meningitidis*: production of an extracellular enzyme which cleaves human IgA. Science **190:**1103–1105.
28. **Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer.** 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature (London) **325:**458–462.
29. **Poulsen, K., J. Brandt, J. P. Hjorth, H. C. Thøgersen, and M. Kilian.** 1989. Cloning and sequencing of the immunoglobulin A1 protease gene (*iga*) of *Haemophilus influenzae* serotype b. Infect. Immun. **57:**3097–3105.
30. **Qiu, J., G. P. Brackee, and A. G. Plaut.** 1996. Analysis of the specificity of bacterial immunoglobulin A (IgA) proteases by comparative study of ape serum IgAs as substrate. Infect. Immun. **64:**933–937.
31. **Saluja, S. K., and J. N. Weiser.** 1995. The genetic basis of colony opacity in *Streptococcus pneumoniae*: evidence for the effect of box elements on the frequency of phenotypic variation. Mol. Microbiol. **16:**215–227.
32. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
34. **Schneewind, O., A. Fowler, and K. F. Faull.** 1995. Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. Science **268:**103–106.
35. **Schneewind, O., P. Model, and V. A. Fischetti.** 1992. Sorting of protein A to the staphylococcal cell wall. Cell **70:**267–281.
36. **Shine, J., and L. Dalgano.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
37. **Takeshita, S., M. Sato, M. Toba, W. Masahashi, and T. Hashimoto-Gotoh.** 1987. High-copy-number and low-copy-number plasmids for *lacZ*α-complementation and chloramphenicol- or kanamycin-resistance selection. Gene **61:**63–74.
38. **Tiraby, J. G., and M. S. Fox.** 1973. Marker discrimination in transformation and mutation of pneumococcus. Proc. Natl. Acad. Sci. USA **70:**3541–3545.
39. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76:**4350–4354.
40. **Tuomanen, E. I., R. Austrian, and H. R. Masure.** 1995. Pathogenesis of pneumococcal infection. N. Engl. J. Med. **332:**1280–1284.
41. **Valle, B. L., and D. S. Auld.** 1990. Active-site zinc ligands and active H_2O of zinc enzymes. Proc. Natl. Acad. Sci. USA **87:**220–224.
42. **Vijayakumar, M., and S. Ayalew.** 1993. Nucleotide sequence analysis of the termini and chromosomal locus involved in the site-specific integration of the streptococcal conjugative transposon Tn*5252*. J. Bacteriol. **175:**2713–2719.
43. **von Heijne, G.** 1986. A new method for predicting signal cleavage sites. Nucleic Acids Res. **14:**4683–4690.
44. **Weiser, J. N., R. Austrian, P. K. Sreenivasan, and H. R. Masure.** 1994. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. Infect. Immun. **62:**2582–2589.
45. **Weiser, J. N., and S. T. H. Chong.** Phenotypic variation in *Streptococcus pneumoniae* correlates with quantitative differences in cell wall associated phosphorylcholine. Submitted for publication.
46. **Weiser, J. N., S. T. H. Chong, D. Greenberg, and W. Fong.** 1995. Identification and characterization of a cell envelope protein of *Haemophilus influenzae* contributing to phase variation in colony opacity and nasopharyngeal colonization. Mol. Microbiol. **17:**555–564.
47. **Weiser, J. N., Z. Markiewicz, E. I. Tuomanen, and J. H. Wani.** 1996. Relationship between phase variation in colony morphology, intrastrain variation in cell wall physiology and nasopharyngeal colonization by *Streptococcus pneumoniae*. Infect. Immun. **64:**2240–2245.

Editor: V. A. Fischetti