Isolation and Characterization of Three *Streptococcus pneumoniae* Transformation-Specific Loci by Use of a lacZ Reporter Insertion Vector

EKATERINA V. PESTOVA* AND DONALD A. MORRISON

Laboratory for Molecular Biology, Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois

Received 26 November 1997/Accepted 9 March 1998

Although more than a dozen new proteins are produced when *Streptococcus pneumoniae* cells become competent for genetic transformation, only a few of the corresponding genes have been identified to date. To find genes responsible for the production of competence-specific proteins, a random lacZ transcripational fusion library was constructed in *S. pneumoniae* by using the insertional lacZ reporter vector pEVP3. Screening the library for clones with competence-specific β-galactosidase (β-Gal) production yielded three insertion mutants with induced β-Gal levels of about 4, 10, and 40 Miller units. In all three clones, activation of the lacZ reporter correlated with competence and depended on competence-stimulating peptide. Chromosomal loci adjacent to the integrated vector were subcloned from the insertion mutants, and their nucleotide sequences were determined. Genes at two of the loci exhibited strong similarity to parts of *Bacillus subtilis* com operons. One locus contained open reading frames (ORFs) homologous to the comEA and comEC genes in *B. subtilis* but lacked a comEB homolog. A second locus contained four ORFs with homology to the *B. subtilis* comG gene ORFs 1 to 4, but comG gene ORFs 5 to 7 were replaced in *S. pneumoniae* with an ORF encoding a protein homologous to transport ATP-binding proteins. Genes at all three loci were confirmed to be required for transformation by mutagenesis using pEVP3 for insertion duplications or an *erm* cassette for gene disruptions.

*Streptococcus pneumoniae* is a representative of a diverse group of bacteria capable of natural genetic transformation, a process that involves active binding and uptake by a bacterial cell of free extracellular DNA, either plasmid or chromosomal, and the heritable incorporation of its genetic information. Natural transformation is thought to be advantageous to this organism by allowing the acquisition of traits, for example antibiotic resistance, from genetically distinct organisms by horizontal gene transfer in natural populations (28, 52).

Natural genetic transformation in *S. pneumoniae* is a highly coordinated process. It is induced in exponentially growing cultures at a specific cell density and lasts for less than 1 h (56). This transient physiological state, in which a cell is able to bind and take up free DNA from the surrounding medium, is called competence for genetic transformation. The induction of competence is regulated by a cell-cell communication mechanism which employs a 17-residue extracellular activator peptide, the competence-stimulating peptide (CSP) (20).

The development of competence in *S. pneumoniae* is accompanied by a drastic change in protein synthesis: synthesis of most cellular proteins is switched off and the production of at least 14 competence-specific proteins is initiated (34). Following the cessation of competence, synthesis of the normal complement of cellular proteins is resumed. Although the ability to take up exogenous DNA is inducible, no competence-specific genes encoding components of the DNA binding and entry mechanisms have yet been identified. It is curious that EndA, the major *S. pneumoniae* endonuclease implicated in DNA processing and uptake, is a constitutive enzyme and is thought to be recruited by competence-specific proteins to form a DNA entry site (44). The only competence-inducible loci characterized thus far are the rec operon which includes two genes essential for transformation, recA and cinA (exp10), and the comCDE locus, which includes three quorum-sensing genes (30, 39, 41). The identification of the other competence-inducible loci is required to understand the mechanism of competence development in *S. pneumoniae* and to determine what proportion of such genes is important for transformation.

In this study, we report identification, cloning, and genetic analysis of three additional *S. pneumoniae* genetic loci, cel, cgl, and coi, the expression of which is competence specific and which include genes required for genetic transformation.

**MATERIALS AND METHODS**

**Bacterial strains and plasmids.** *S. pneumoniae* CP strains used in this study are descended from strain Rx, which was originally derived from R36A (3). Strain CP1500 (hex nov-1 bry-r str-1 ery-c2 ery-r6) was used as a source of the Nov marker in transformation assays. CP1250 (hex malM511 str-1 bgfl-1) is a CP1200 derivative with low β-galactosidase (β-Gal) activity (41). Strains CP1253, CP1254, and CP1255 are mutants bearing the chromosome-integrated pEVP3 vector. *Escherichia coli* host strains for plasmid propagation and for cloning of pneumococcal DNA were DH1, DH5α, and DH10B. Plasmids used and constructed in this study are listed in Table 1. pEVP3 is a lacZ reporter vector for insertion mutagenesis in *S. pneumoniae* (9).

**Media.** *E. coli* cells for plasmid DNA isolation were grown in either Luria-Bertani (LB) or Terrific broth (TB) (49). The solid medium was LB agar supplemented with antibiotics. For selection of plasmid-encoded drug resistance in *E. coli*, ampicillin (100 μg/ml), chloramphenicol (34 μg/ml), or erythromycin (500 μg/ml) was added. Casein hydrolysate (CAT) broth (35) was used to grow *S. pneumoniae* cultures. When CAT broth was used to grow cells in plates, 1.5% of agar (Difco Co.) was added to the medium. The medium was supplemented with antibiotics (novobiocin, 2.5 μg/ml; chloramphenicol, 2 μg/ml; or erythromycin, 2 μg/ml) as required. All the drugs were purchased from Sigma Chemical, Inc. Complete transformation medium (CTM), used to grow *S. pneumoniae* cultures for competence assay, was made by adding 1/100 volume of 0.1 M CaCl₂ (analytical grade reagent, Fisher Scientific), and 1/20 volume of 4% bovine serum albumin solution (fraction V; Sigma Co.) into complete CAT broth before use.

**General *S. pneumoniae* culture conditions.** Unless specified otherwise, *S. pneumoniae* cultures were grown in broth at 37°C without aeration. To grow isolated colonies of *S. pneumoniae*, plates containing three or four layers of CAT-based culture medium with agar were used. The second layer was inoculated with...
Pneumococcal cells at the appropriate dilution. Antibiotics for the selection of transformants were added to the fourth (top) layer in the relevant concentration (4, double-overlap method [35]). S. Bremo-choro-3-indolyl-β-D-galactopyranoside (X-Gal) indicator plates were prepared by spreading X-Gal solution (20 mg/ml in N,N-dimethylformamide) over the bottom layer of broth agar.

S. pneumoniae transformation and competence assays. Pneumococcal cultures were grown at 37°C to an OD of 0.001 (about 5 x 10^6 cells/ml) in CM broth supplemented with 9 mM HCl (initial pH about 6.6) to suppress spontaneous development of competence and grown at 37°C. When cultures reached an OD of 0.1, 2 μg of transforming DNA per ml was added to the cultures, and competence was induced by the addition of NaOH to a final concentration of 9 mM and pH of about 7.2 (36), synthetic CST (100 mg/ml; Chiron Immotopes), or both. After incubation for 1 h at 37°C to allow development of competence, uptake of DNA, and integration of new alleles, samples were diluted and plated with an appropriate antibiotic for the selection of transformants by the double-overlap method (35).

Assay of β-Gal activity. The assay of β-Gal activity was performed as described previously (35, 41). To screen large number of clones for β-Gal production under conditions specific for the induction of competence, each clone was induced to competence in CM broth essentially as described above, incubated at 37°C for 45 min, and assayed for Nov^R transformants and for β-Gal activity (by an o-nitrophenyl-β-D-galactopyranoside [ONPG] test performed in microtiter wells, using 250 μl of cell extract and 50 μl of ONPG solution). Plates were incubated either at room temperature or at 37°C for 5 h and scored for the yellow color produced.

Isolation of plasmid DNA. Small-scale plasmid DNA purification from E. coli cells was carried out by an alkaline lysis method as described previously (40). Large-scale isolation of plasmid DNA was carried out by using Wizard Maxiprep DNA Purification System (Promega Corp.) as described in the manufacturer’s manual.

Cytochrome c oxidase mutation. Mutagenic plasmids from S. pneumoniae cells, crude plasmid DNA preparations obtained by the method of Stassi et al. (53) were used for electroporation into E. coli DH5α. Typical yields were 10 to 15 μg of plasmid DNA per μl of DNA.

Construction of lacZ transcriptional fusion library in S. pneumoniae. To make the insertion library, pools of integrative mutagenic plasmids were created by the following procedure. S. pneumoniae chromosomal DNA was digested with TaqI and filled in with Klenow enzyme (New England Biolabs, Inc.) leaving 1-bp (C) 5' overhangs. peV3 vector was digested with BamH1 and partially filled in with Klenow enzyme to leave 5' G overhangs. After ligation, hybrid plasmids were transformed into E. coli DH1. Cm^R E. coli clones containing amplified mutagenic plasmids were selected and pooled for plasmid purification. Restriction analysis of selected clones showed that about 30,000 peV3-based insertion plasmids were recovered. Mixtures of purified mutagenic plasmids were transformed into S. pneumoniae CP1250, and Cm^R transformants containing peV3 vector integrated into different chromosomal loci were selected on the surface of CAT agar containing chloramphenicol. To store the resulting insertion libraries, Cm^R colonies were washed from the surface of the agar, diluted in CAT broth to an OD of about 0.005, incubated for 2 h at 37°C, and frozen at −80°C in 12% glycerol.

DNA sequencing and sequence analysis. DNA sequencing was carried out by the Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, using template plasmid DNA prepared by Wizard Maxipreps DNA Purification System or PCR-generated DNA fragments purified with Ultra- Free-MC centrifugal filter units (Millipore Corporation). DNA and protein sequence homology searches were carried out at EMBL (14) by running FASTA (40) and at the National Center for Biotechnology Information (NCBI) (U.S. National Library of Medicine, National Institutes of Health) by running Entrez (4) or BLAST (2). DNA and protein sequence alignment was carried out by the CLUSTAL V method (21), using the MegaAlign program (DNASTAR, Inc.).

The following strategy was used for sequencing of the three competence-inducible loci. Nucleotide sequences of the 2,610-bp fragment (coi) carried by the plasmid pXF511, the 2,364-bp fragment (cel) carried by pXF512, and the 3,119-bp fragment (cgf) located by pXF513 were determined by direct sequencing using plasmid DNA as a template. Primers used for the first sequencing reactions, upstream primer DAM087 (5'-ACCGGGAGCTGCAATTCTA-3') and downstream primer DAM072 (5'-CTTCACAAGTAGTTCCACCT-3'), were complementary to peV3 sequences adjacent to the insert. Primers internal to the chromosomal insert were designed for the subsequent reactions to obtain and verify the complete DNA sequence of the fragments in both strands.

To determine the sequence of a 1,482-bp chromosomal region upstream of the cloned 2,364-bp fragment of the cel locus, S. pneumoniae chromosomal DNA was digested with NheI, ligated to an XhoI-cut peV3, and PCR amplified with primers DAM087, which is complementary to peV3 upstream of the insert, and DAM117 (5'-GTCAGGGCCAATGCAGAA-3'), located in the previously sequenced region. The resulting PCR product was directly sequenced with the same primers and additional internal primers.

To determine the sequence upstream of the 3,119-bp fragment, S. pneumoniae chromosomal DNA was digested with SspI restriction enzyme and ligated to a SmaI-cut peV3. The upstream fragment was inserted in the vector in the orientation opposite that of the 3,119-bp insert in pXF513 by PCR amplification with primers DAM072, which is complementary to peV3 downstream of the insert, and DAM118 (5'-TTGGCTATTCCATGTTGGCTG-3') contained in the previously sequenced region. The resulting PCR product was sequenced with the same primers. Part (807 bp) of the new sequence obtained as a contiguous primer was combined with the previously determined sequence of the 3,119-bp chromosomal fragment.

Gene disruption mutagenesis. Two methods were used for gene disruption. One used an integrative method of peV3 insertion by genetic recombination, carried out by an internal fragment of the gene, while the second method involved either interruption or replacement of genes with the erythromycin resistance cassette, syn^R, derived from M13eryAD, ligated to an XhoI-cut peV3, and PCR amplified with primers DAM087, which is complementary to peV3 upstream of the insert, and DAM117 (5'-GTCAGGGCCAATGCAGAA-3'), located in the previously sequenced region. The resulting PCR product was sequenced with the same primers. Part (807 bp) of the new sequence obtained as a contiguous primer was combined with the previously determined sequence of the 3,119-bp chromosomal fragment.
RESULTS

Identification of competence-specific loci by screening a random transcriptional fusion library in *S. pneumoniae*. The strategy adopted for identification of competence-specific loci included making a random transcriptional fusion library in *S. pneumoniae* and identifying clones demonstrating competence-specific expression of the integrated lacZ reporter. Construction of a library in which pEVP3 is inserted at random sites of the *S. pneumoniae* chromosome involved cloning small restriction fragments of *S. pneumoniae* chromosomal DNA into the multiple cloning site of the vector to create a pool of integrative mutagenic plasmids, transforming these plasmids into *S. pneumoniae* CP1250 (malM511 str-1 bgl-1) (41), and selecting the resulting Cm<sup>r</sup> clones.

The insertion vector pEVP3 bears a chloramphenicol resistance determinant, the syn *cat* cassette, which was shown previously (9) to be autonomously expressed when inserted in the *S. pneumoniae* chromosome in a single copy. Therefore, the distribution of pEVP3 chromosomal insertions in the library was expected to be relatively random. In fact, we observed that the pEVP3 insertion library contained clones that formed colonies of different color intensities on X-Gal plates, showing that in this library the lacZ reporter gene was randomly fused to operons with different levels of expression. In total, about 20% of all colonies in the pEVP3 insertion library were blue. Assuming that in only one of two clones, the reporter gene is inserted in the same orientation as in the target gene, we conclude that 40% of plasmid inserts in the pEVP3 library were in active genes.

The transcriptional fusion library was screened for clones in which β-Gal production was activated under conditions specific for competence induction. From 200 independent clones examined individually by an ONPG assay following competence induction, three insertion mutants demonstrated significantly higher levels of β-Gal production when induced to competence. In all three, the presence of a chromosome-integrated pEVP3 vector was confirmed by Southern hybridization analysis (data not shown). The mutants were designated strains CP1253, CP1254, and CP1255.

**Initial characterization of the inducible insertion mutant strains**

CP1253, CP1254, and CP1255 and competence-specific induction of integrated lacZ reporter. To confirm that in the insertion strains CP1253, CP1254, and CP1255, the induction of β-Gal was competence specific, we examined the dependence of the expression of the lacZ transcriptional fusions on the conditions specific for competence induction: pH shift in a growing culture and the addition of the competence pheromone, CSP (41). In all three mutants, the expression of lacZ was induced to similar levels after pH shift, CSP addition, or both in combination. Accompanying induction of competence (Table 2), expression of the reporter increased 4- to 100-fold. Therefore, the peptide activator induces both competence development in the mutant strains and expression of the chromosomal loci bearing the integrated lacZ reporter gene.

The timing of this competence-specific expression was analyzed (Fig. 1). When β-Gal production was activated by induction with NaOH, the increase in enzyme activity closely coincided with the development of competence in all three mutant strains. The addition of trypsin, an inhibitor of competence induction in *S. pneumoniae* (55), prior to the induction with NaOH completely abolished both competence induction and activation of β-Gal production in all three mutant strains. Thus, the chromosomal loci bearing an integrated pEVP3 vector in mutant strains CP1253, CP1254, and CP1255 are specifically induced during competence for genetic transformation and depend on CSP for their expression.

**Cloning and nucleotide sequence analysis of loci of pEVP3**

integration from the inducible mutant strains CP1253, CP1254, and CP1255. The vector pEVP3, used to create the transcriptional fusions in the *S. pneumoniae* chromosome, also provided a convenient tool for direct cloning of the inducible loci, as recombination between the duplicated chromosomal regions generated during its integration leads to spontaneous excision of the original transforming plasmid (59, 60) bearing the same fragment of chromosomal DNA as directed the original insertion. Excised plasmids were recovered by isolation of “plasmid DNA” from the insertion strain followed by transformation into a recA mutant strain of *E. coli* with selection for the plasmid-encoded drug resistance. In the present study, we employed this method to clone fragments of the loci of pEVP3 integration from inducible mutant strains CP1253, CP1254, and CP1255.
that shared similarity with the AGGAGG consensus sequence the ORFs are preceded by potential ribosome binding sites cglB codon, except for cglC, cglD, A culture of each mutant strain was grown from 10^6 cells/ml to an OD_550 of 0.15 in CTM containing 9 mM HCl and split in two portions, one of which was induced to competence with 9 mM NaOH (large arrow). The number of Nov^R transformants and β-Gal activity (in Miller units [M.U.]) were determined at 5-min intervals. Trypsin was added to one of the two portions of each culture (2 μg/ml).

and CP1255 as plasmids pXF511, pXF512, and pXF513, respectively.

To characterize the competence-specific chromosomal loci further, we determined and analyzed the nucleotide sequences of chromosomal inserts in each of the recovered plasmids. Additional sequence information was obtained for mutant strains CP1254 and CP1255 by direct sequencing of PCR-amplified regions located upstream of the cloned fragments. The 2,610-bp chromosomal locus from CP1253 will be referred to hereafter as coi (for competence inducible). The 3,845-bp locus from CP1254 and the 3,926-bp locus from CP1255 were named cel and cgl, respectively. As discussed below, this designation was based on their homology to B. subtilis late competence operons comE and comG (cel for comE-like and cgl for comG-like). Genetic maps of coi, cel, and cgl loci are given in Fig. 2.

Nucleotide sequence analysis revealed the presence of two open reading frames (ORFs) in the coi locus, one of which was truncated on the right at the point of fusion to the pEVP3 lacZ gene. Both ORFs were preceded by putative Shine-Dalgarno sequences (51). These ORFs were designated orfl1 and coiA (Fig. 2). An additional truncated ORF, orfl2, was found immediately upstream of orfl1.

Three ORFs were identified in the cel locus. Two of these ORFs start with the ATG initiation codon and overlap by 14 nucleotides. The third ORF starts with TTG, frequently used as an initiation codon in gram-positive bacteria (29, 32), and is truncated on the left. Probable ribosome binding sites precede the first and third ORFs (37, 51). Analysis of the sequence upstream of the second ORF did not reveal any plausible ribosome binding site in the proximity of the initiation codon. A sequence with dyad symmetry that could form a stem-loop structure and possibly act as a rho-independent terminator was found 246 bp downstream from the stop codon of the second ORF.

The cgl locus contained five ORFs, designated cglA, cglB, cglC, cglD, and cglE, all of which start with an ATG initiation codon, except for cglB, which appears to start with TTG. All of the ORFs are preceded by potential ribosome binding sites that shared similarity with the AGGAGG consensus sequence (37, 51). Two putative genes, cglC and cglD, overlapped by 5 nucleotides. Although no ORFs of significant length were found downstream of cglE, analysis of this putative intergenic region revealed two sequences with dyad symmetry 94 and 168 bp downstream from the cglE stop codon, respectively, that may function as rho-independent terminators.

Genetic analysis of the coi, cel, and cgl loci by gene disruption insertion mutagenesis. To determine if at least some genes in the coi, cel, and cgl loci are essential for transformation, we analyzed these loci by gene disruption mutagenesis, using either pEVP3 for creating insertion duplications or an erythromycin resistance cassette, syn _erm_ (9), for gene interruption or replacement (Fig. 2). Since pEVP3 integration created transcriptional fusions with the lacZ reporter, this method allowed study of both the requirement of these genes for transformation and the level of their expression in mutant strains. The syn _erm_ cassette used for gene disruption contained a synthetic promoter to allow autonomous expression of the _erm_ gene in a single copy (9). Each insertion mutant was tested for competence (transformability). In addition, mutants bearing pEVP3 insertions were assayed for β-Gal activity. As evident from Table 3, only one of the genes in the coiA locus is important for competence. Both pEVP3 integration and syn _erm_ insertion in coiA caused a 100-fold decrease in transformability, while the disruption of orfl1 did not reduce the level of transformation. Furthermore, clones bearing pEVP3 integrated in orfl1, produced about 6 to 7 Miller units of β-Gal in both induced and uninduced cultures, showing that the expression of this gene is not competence specific. In contrast to orfl1, coiA is apparently induced by CSP, since the addition of the peptide resulted in a twofold increase in the expression of the coiA-fused _lacZ_ reporter.

As shown in Table 3, strain CP1253 was transformed at wild-type levels, while a coiA disruption mutant demonstrated reduced transformability. This apparent discrepancy could be explained at least in part by the duplication of the entire locus extending to orfl2 in mutant strain CP1253. The duplicated region, created by integration of pXF511, includes one complete copy of coiA and its putative upstream regulatory sequences.
Therefore, the function and regulation of this gene in mutant CP1253 were probably unaffected by plasmid integration.

Analysis of insertions in the cel locus revealed the presence of at least one gene essential for transformation, celB. The disruption of this gene by pEVP3 insertion resulted in more than a 100-fold decrease in transformability of the mutant clones, and celB::lacZ expression increased 10-fold upon induction with CSP. The disruption of orfr1, located downstream of celB, by an insertion of the syn erm cassette did not affect competence in the mutants, showing that this gene is not required for transformation. However, the orfr1::lacZ fusion induced during competence. It is possible that this induction was due to readthrough transcription from celB. Alternatively, orfr1 could be induced at competence but is not essential for transformation. The reduced transformability of mutant strain CP1254 is also puzzling. This mutant contains a duplication of a 2,364-bp fragment extending from celB to orfr1 to produce one complete copy and one truncated copy of celB. It is possible, therefore, that this duplication of celB interferes with the normal process of transformation in mutant strain CP1254.

The phenotypes of cgl mutants indicated the importance of this locus for transformation. Replacement of cglB, cglC, and cglD with the syn erm cassette resulted in complete loss of transformability (Table 3). It is possible that this mutation may cause a phenotype via a polar effect on cglE if all cgl genes form a single operon. However, since the disruption of cglE by pEVP3 insertion also resulted in transformation deficiency, it can be concluded that at least one gene in the cgl locus, cglE, is essential for competence. The cglE::lacZ fusion was induced by CSP, suggesting that the activator peptide is involved in the regulation of cglE expression. It was also noticed that a cglE disruption mutant reduced levels of transformability more severely than the originally selected mutant strain CP1255 that was created by the integration of pXF513 into the cgl locus, a difference presumably due to the duplication of the 3,119-bp fragment of cgl locus including cglB, cglC, cglD, and cglE genes in mutant strain CP1255.

**Predicted protein sequences and possible functions of the products of competence-specific loci.** Amino acid sequences of coi, cel, and cgl translation products were compared to all known proteins in the NCBI database. While the search revealed no significant homology for the putative transformation-specific gene coiA, the cel and cgl loci were found to share the highest similarity with parts of B. subtilis comE and comG operons (Fig. 3).

The cel locus includes two putative genes, celA and celB, similar to the comEA and comEC genes of a B. subtilis competence-specific locus comE (19, 25) but lacks a homolog of the second gene in that operon, comEB. A 216-residue protein
encoded by the cellA gene is about 40% identical to a B. subtilis 205-amino-acid (aa) transmembrane protein ComEA required for DNA uptake during genetic transformation. A 753-residue protein product of the cellB gene is 30% identical over the entire length of the proteins to the 776-aa B. subtilis protein ComEC that was proposed to function as a part of an aqueous channel for DNA internalization. In addition, CellB is also related to DNA transport proteins from Neisseria gonorrhoeae and Haemophilus influenzae, sharing 41% identity with N. gonorrhoeae integral membrane protein ComA (15) in a 230-aa overlap, and 39% identity with RecC2 protein of H. influenzae (31) in a 237-aa overlap (Fig. 3A).

Four of five putative genes of the S. pneumoniae cgl locus, cellA, cellB, cellC, and cellD, encode polypeptides that exhibit their highest similarity to protein products of ORFs 1 to 4 of the B. subtilis comG operon (1). The comG locus is essential for transformation, consists of seven ORFs, and is thought to encode elements of a DNA translocation machine (5, 8). As demonstrated in Fig. 3B, B. subtilis comG and S. pneumoniae cgl operons have similar organizations. However, comG gene ORFs 5 to 7 are replaced in S. pneumoniae with an ORF, cellE, which, as will be discussed later, encodes a product similar to transport ATP-binding proteins. The database search revealed also a similarity between the cgl locus and two groups of genetic loci from gram-negative bacteria, one involved in pilus assembly and the other participating in protein secretion. Among these loci, the pilus biogenesis gene clusters from Pseudomonas aeruginosa (38) and N. gonorrhoeae (17, 27), and extracellular enzyme secretion operons from Xanthomonas campestris (12) and Aeromonas hydrophila (22) demonstrated the highest similarity to the cgl genes on the levels of both protein homology and genetic organization. These pilus assembly and protein secretion gene clusters are compared to competence-specific loci cgl and comG in Fig. 3B.

The 339-aa homolog of ComG1, CglA, is similar to the ATP-binding proteins required for export of extracellular enzymes (X. campestris XpsE and A. hydrophila ExeE) and for the pilus assembly (P. aeruginosa PilB and N. gonorrhoeae PilF). The similarity appears to be especially strong in the central part of the proteins that include a region highly conserved in nucleotide triphosphate (NTP)-binding domains of various ATP-binding proteins (the so-called Walker NTP-binding motifs [61]). The predicted protein product of cglB, a 291-aa polypeptide, shares the highest sequence similarity with B. subtilis ComG2 protein and the accessory integral membrane proteins of pilus assembly (PIIC and PIIG [57]) and protein secretion (ExeF and XpsF) pathways.

The predicted translation products of cglC and cglD genes are 109- and 135-aa proteins, respectively, that share similar N-terminal hydrophobic segments. This N-terminal part of both the CglC and CglD proteins resembles a hydrophobic region that is conserved in pilins from various gram-negative bacterial species and was implicated in protein-protein interaction of pilin subunits involved in the assembly of pili (50). This region is also conserved in B. subtilis ComG3, ComG4, and ComG5 proteins. The amino acid sequence alignment of this conserved region of CglC, CglD, ComG3, ComG4, and pilins from N. gonorrhoeae (pilin E [26]), Dichelobacter nodosus (13), Moraxella nonliquefaciens (58), P. aeruginosa (46), and Neisseria meningitidis (54) is presented in Fig. 4.

Pilins, major subunits of pili, are processed by the removal of 5 or 6 aa residues from the N termini followed by N-methylation of the highly conserved terminal Phe residue. The Gly residue which immediately precedes Phe was found to be required for pilin processing (38). This Gly residue and 3 aa surrounding the cleavage site (K-G-F-T-L [baldfaced amino acids become termini after cleavage]) are also conserved in ComG pilin-like proteins, one of which, ComG3 (ComC), has been shown to be processed during export to the cell surface (8). CglC and CglD, however, lack a Gly residue in this conserved sequence, suggesting that putative S. pneumoniae pilin-like proteins may not be processed, but rather anchored to the membrane, unless Ala could substitute for Gly in the processing site.

The 198-aa protein product of the last ORF in the cgl locus, cellE, did not share any significant similarity with either of the ComG proteins or any other known competence-related proteins. However, a BLAST search revealed similarity of CglE to ATP-binding protein components of transporters which belong to the superfamily of so-called ABC transporters or traffic ATPases (10). ATP-binding proteins of seven bacterial transport systems that demonstrated the highest similarity to CglE are aligned in Fig. 5. The regions of CglE similarity with these proteins include the NTP-binding Walker B motif (61) and the highly conserved glycine-glutamine-rich sequence L-S-G-G-Q-Q-Q, named the linker peptide, that is located between the membrane-spanning helical domain and the nucleotide-binding pocket in ABC transporters, and is also found in peptide linkers that join together separate domains within various other proteins (24). In ABC transporters, this amino acid sequence is thought to be involved in the signaling mechanism that leads to ATP hydrolysis and translocation of the receptor-bound substrate through the membrane by an interaction with the hydrophobic membrane domains of the transport systems and coupling ATP-dependent conformational changes to the transport process (10). Similar to ATP-binding proteins of ABC transporters (24), the linker peptide of CglE is located between the putative hydrophobic domain (aa 46 to 75) and the Walker B motif (aa 129 to 136). However, in contrast to transport proteins, CglE does not have the conserved glycine-rich Walker A motif (G-X-G-X-G-K) essential for the ATP binding process.
binding; this may suggest that CglE is functionally different from ATP-binding components of ABC transporters.

**Putative competence-specific promoters.** Previous observations on mechanisms of competence induction in *S. pneumoniae* (34, 56), combined with our recent data, suggest that the transcription of *coi*, *cel*, and *cgl* loci is activated upon competence induction by a competence-specific regulatory mechanism. The existence of a competence-specific mechanism of transcriptional regulation, possibly a transcription factor or an alternative sigma factor, could be inferred from the examination of nucleotide sequences upstream of the putative initiation codons of *coiA*, *celA*, and *cglA*. Although no obvious *S. pneumoniae* promoters (48) were found upstream of these genes, all these genes share a conserved nucleotide motif preceded by an AT-rich region, which is located 13 to 15 bp upstream of the ATG codon. A similar nucleotide sequence was described recently by Campbell and Masure (6) as a part of competence-specific promoters. The same nucleotide motif is also present upstream of *exp10* (*cinA*), the first gene in the *rec* operon, as illustrated in Fig. 6. Interestingly, transcriptional mapping and Northern hybridization analysis of the *rec* operon have demonstrated the existence of a competence-specific promoter upstream of *exp10* (*recA*) (30, 39). It is possible, therefore, that this competence-specific promoter is represented by a conserved sequence with a consensus TTACGAATANTATAGG, preceded by an AT-rich region.

![Diagram](https://example.com/diagram.png)

**FIG. 3.** Homologs of *S. pneumoniae* cel and cgl loci. (A) Comparison of *S. pneumoniae* cel locus to *B. subtilis* comE operon (GenBank accession no. L15202), *N. gonorrhoeae* com locus (S75490), and *H. influenzae* rec2 gene (U32691). Genes encoding similar protein products are shown by the same shading; those demonstrating no similarity are not shaded. (B) Comparison of *S. pneumoniae* cgl locus to *B. subtilis* comG operon (M29691), *P. aeruginosa* pil (M32066), *N. gonorrhoeae* pil (U32688), *A. hydrophila* exe (X66504), and *X. campestris* xsp (X59079) operons. Genes encoding similar protein products are indicated by the same shading. Conserved N-terminal regions of pilins and pilin-like proteins are shown in black. Nucleotide-binding motifs that are similar in CglA, ComG1, PilB, PilF, ExeE, and XspE proteins are represented by hatched boxes.
DISCUSSION

In this study, by screening a random insertion library containing lacZ transcriptional fusions to various S. pneumoniae chromosomal loci, we identified three insertion mutants, CP1253, CP1254, and CP1255, in which β-Gal activity was induced specifically under the conditions of competence induction. In all three clones, activation of the lacZ reporter correlated with competence and depended on CSP. Inducible loci from mutants CP1253, CP1254, and CP1255 were named coi, cel, and cgl, respectively. Nucleotide sequence analysis identified the presence of three ORFs in the coi locus, three ORFs in the cel locus, and five ORFs in the cgl locus. The fact that genes at all three of the competence-specific loci, which were found on the basis of competence-linked expression, are important for formation of recombinants during transformation supports the proposal (34) that many or most of the proteins induced at competence would prove to be parts of the DNA-processing pathway of transformation.

In agreement with this hypothesis, comparison of cel translation products to all proteins in the NCBI database revealed a similarity of two putative genes of the cel locus, celA and celB, to comEA and comEC genes of B. subtilis comE operon essential for DNA uptake. In addition, celB was found to be similar to N. gonorrhoeae comA and H. influenzae rec2 genes that...
encode DNA transport proteins. This sequence similarity and the importance of cellB for transformation suggest a possible role for CelB protein as a part of the DNA uptake machinery.

As discussed above, the S. pneumoniae cgl locus demonstrated significant similarity to the B. subtilis comG operon. Four ORFs of the cgl locus, cglA to cglD, were found to be homologous to the B. subtilis comG gene ORFs 1 to 4. A significant similarity was also found between the cgl locus and operons involved in pilus biogenesis and protein secretion in gram-negative bacteria. On the basis of amino acid sequence analysis, cgl protein products were predicted to represent elements of the multicomponent DNA transport machine.

CglA and CglB proteins demonstrated homology to B. subtilis transformation-specific proteins ComG1 and ComG2 and to proteins required either for the excretion of enzymes (general secretion pathway proteins; for example, products of A. hydrophila exeE) or for the assembly of pili (products of P. aeruginosa pilB and pilC in Fig. 3) in a variety of bacterial species. On the basis of their homology to the components of protein transport systems, ComG1 and ComG2 have been suggested to participate in the assembly of a DNA translocation machine (11). It is possible that a similar role is performed by S. pneumoniae proteins CglA and CglB. This hypothesis is also supported by the fact that cglA and cglB homologs in N. gonorrhoeae, pilF and pilG, are required for genetic transformation (17, 57).

The small proteins CgIC and CgID could possibly be assembled in multimeric membrane-associated pore-like structures similar to those presumably formed by homologous pilin-like components of the general secretion pathway (43). These components, for example, protein products of A. hydrophila exeG to exeJ genes (Fig. 3), share homology with CgIC and CgID proteins. Similar pilin-like proteins in B. subtilis, ComG3 to ComG7, are thought to form a multimeric complex for the binding and uptake of transforming DNA (8, 11). One of these proteins, ComG3 (ComGC), was shown to be required for DNA binding during genetic transformation in B. subtilis and is thought to assemble with ComG4, ComG5, ComG6, and ComG7 to form a cell surface-associated structure, possibly a pore, for the binding and uptake of transforming DNA. Interestingly, N. gonorrhoeae pilin, PiIE, has been found to be required for DNA uptake during transformation (45). This involvement of pilins and pilin-like proteins in transformation may suggest similar functions for CgIC and CgID.

The product of the transformation-specific cglE gene exhibits partial homology to ATP-binding components of ABC transporters and is essential for transformation; therefore, it could be involved in DNA translocation across the cytoplasmic membrane as a part of the multicomponent protein mechanism that takes up DNA.

Genes specifically induced at competence and encoding proteins involved in DNA binding, transport, and genetic recombination can be characterized as late competence genes (11). A total of four pneumococcal late competence loci have now been identified, including three discovered in this study. Interestingly, all four loci have a similar nucleotide motif in their putative promoter regions, which may indicate the existence of a common mechanism coordinating their expression during competence. While the functions of recA, cgl, and cel loci either are demonstrated experimentally or can be predicted on the basis of protein homology, the function of coiA, the only competence-specific gene in the third locus identified in this study, is not known. No significant homology to any proteins with known function was found in databases for CoiA protein. Further studies are required to find the role of coiA in transformation, test the hypothetical function of Cel and Cgl proteins in DNA uptake, and determine whether additional components of the DNA binding and translocation complex exist, and how these components interact during DNA uptake.

REFERENCES


