JOURNAL OF BACTERIOLOGY, May 1998, p. 2701–2710 Vol. 180, No. 10
0021-9193/98/$04.00+0

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* transcriptional 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-r1 bry-r str-r1 ery-r2 ery-r6*) was used as a source of the Nov^r marker in transformation assays. CP1250 (*hex malM511 str-1 bgl-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 $CaCl_2$ (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

* Corresponding author. Present address: Department of Pathology, Northwestern University, Chicago, IL 60611. Phone: (312) 503-0951. Fax: (312) 503-8240. E-mail address: e-pestova@nwu.edu.

TABLE 1. Plasmids used and constructed in this study

Plasmid	Relevant genotype or phenotype	Description or purpose in this study	Source or reference
M13eryAD	M13mp19::syn *erm*	Source of the syn *erm* cassette	11
pCR™2.1	Amp^r Km^r	*E. coli* cloning vector; direct cloning of PCR products	Invitrogen, Inc.
pEVP3	Cm^r; pLGW300Δ*kanJH1*::syn *cat*	*S. pneumoniae* reporter and insertion vector	This study (9)
pXF511	pEVP3::′*orfl1 orfl2 coiA*′	Original clone of *coi* locus from CP1253	This study
pXF512	pEVP3::′*celB orfr1*′	Original clone of the *cel* locus from CP1254	This study
pXF513	pEVP3::′*cglA cglB cglC cglD cglE*	Original clone of the *cgl* locus from CP1255	This study
pXF522	pEVP3::′*orfl1*′ (fragment of *coi orfl1*)	Insertional mutagenesis of the *coi* locus	This study
pXF523	pEVP3::′*coiA*′	Insertional mutagenesis of the *coi* locus	This study
pXF524	pXF511 *coiA*::syn *erm*	Insertional mutagenesis of the *coi* locus	This study
pXF525	pEVP3::′*celB*′	Insertional mutagenesis of the *cel* locus	This study
pXF526	pXF512 *orfr1*::syn *erm*	Insertional mutagenesis of the *cel* locus	This study
pXF527	pEVP3::′*cglE*′	Insertional mutagenesis of the *cgl* locus	This study
pXF528	pXF513Δ*cglB-cglD*::syn *erm*	Mutagenesis of the *cgl* locus	This study

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-overlay method [35]). 5-Bromo-4-chloro-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 initiated at an optical density at 550 nm (OD_{550}) of 0.001 (about 10^6 cells/ml) in CTM 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_{550} of 0.1, 2 μg of transforming DNA per ml was added to the culture, and competence was induced by the addition of NaOH to a final concentration of 9 mM and pH of about 7.2 (36), synthetic CSP (100 ng/ml; Chiron Mimotopes), 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-overlay method (35).

Assay of β-Gal activity. The assay of β-Gal activity was performed as described previously (33, 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 CTM 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 (49). Large-scale isolation of plasmid DNA was carried out by using Wizard Maxipreps DNA Purification System (Promega Corp.) as described in the manufacturer's manual.

To recover chromosome-integrated 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* DH10B. Typical yields were 10 to 15 chloramphenicol-resistant clones per μg of DNA.

Construction of random *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 *Taq*I and filled in with Klenow enzyme (New England Biolabs, Inc.) leaving 1-bp (C) 5′ overhangs. pEVP3 vector was digested with *Bam*HI 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 pEVP3-based insertion plasmids were recovered. Mixtures of purified mutagenic plasmids were transformed into *S. pneumoniae* CP1250, and Cm^r transformants containing pEVP3 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_{550} of about 0.05, 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 at 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 Ultrafree-MC centrifugal filter units (Millipore Corporation). DNA and protein sequence analysis was performed with the DNASIS program (Hitachi America, Ltd.) and EditSeq program (DNASTAR, Inc.). 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 MegAlign program (DNASTAR, Inc.).

The following strategy was used for sequencing of the three competence-inducible loci. Nucleotide sequences of the 2,600-bp fragment (*coi*) carried by the plasmid pXF511, the 2,364-bp fragment (*cel*) carried by pXF512, and the 3,119-bp fragment (*cgl*) locus carried by pXF513 were determined by direct sequencing using plasmid DNA as a template. Primers used for the first sequencing reactions, upstream primer DAM087 (5′-ACCCGGGAGCTCGAATTCTA-3′) and downstream primer DAM072 (5′-CTTCCACAGTAGTTCACCACCT-3′) were complementary to pEVP3 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-kb fragment of the *cel* locus, *S. pneumoniae* chromosomal DNA was digested with *Nhe*I, ligated to an *Xba*I-cut pEVP3, and PCR amplified with primers DAM087, which is complementary to pEVP3 upstream of the insert, and DAM117 (5′-GTCAAGCCCAATCGCAAGA-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-kb fragment, *S. pneumoniae* chromosomal DNA was digested with *Ssp*I restriction enzyme and ligated to a *Sma*I-cut pEVP3. The upstream fragment inserted in the vector in the orientation opposite that of the 3,119-bp insert in pXF513 was PCR amplified with primers DAM072, which is complementary to pEVP3 sequence downstream of the insert, and DAM118 (5′-TTGCTCATTCCACTTGGCTG-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 from these primers 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 method employed integration of pEVP3 into the gene of interest targeted by an internal fragment of the gene, while the second method involved either interruption or replacement of genes with the erythromycin resistance cassette, syn *erm*, derived from M13eryAD (9) by PCR amplification of a 1,657-bp fragment bearing syn *erm* flanked by convenient restriction sites. *S. pneumoniae* CP1250 was used for mutagenesis. Mutants were selected on CAT agar containing either chloramphenicol or erythromycin. The structure of each pEVP3 insertion was verified by PCR amplification of a junction fragment using primer DAM072, which is complementary to pEVP3 sequences, and a second primer located upstream of the integrated vector.

To target integration of pEVP3 into *orfl1* of the 2,610-bp *coi* locus, a mutagenic plasmid was made by subcloning a 506-bp *Sph*I-*Nhe*I fragment of this gene, derived from pXF511, into *Sph*I-*Xba*I-digested pEVP3. Both pEVP3 and a syn *erm* cassette were employed for disruption of *coiA*. Mutagenic plasmid pXF523 was used to insert pEVP3 into this gene. The plasmid was constructed by subcloning a 314-bp *Eco*RI-*Xba*I *coiA* fragment, derived from pXF511 and filled in with Klenow enzyme on the *Eco*RI end, into *Sma*I-*Xba*I-digested pEVP3. To disrupt *coiA* with the syn *erm* cassette, pXF524 was constructed with syn *erm* inserted into a *coiA* fragment. The PCR-amplified syn *erm* cassette was digested with *Sma*I and subcloned into the pXF511 *Sac*II site treated with T4 DNA polymerase to remove 3′ overhangs. The resulting plasmid, pXF524, was linearized with *Sal*I and transformed into *S. pneumoniae* CP1250 to disrupt the chromosomal *coiA* gene by homologous recombination.

Two genes, *celB* and *orfr1*, were disrupted in the *cel* locus. To create pEVP3 insertion disruption of *celB*, a mutagenic plasmid, pXF525, was created by subcloning a 1,283-bp *Sph*I-*Nhe*I fragment of *celB* derived from pXF512 into *Sph*I-*Xba*I-digested pEVP3. *orfr1* was disrupted by insertion of the syn *erm* cassette by using pXF526, a pXF512 derivative that contained syn *erm* subcloned into the *orfr1* fragment. The plasmid was constructed by insertion of the syn *erm* cassette on a *Sma*I-digested PCR fragment of M13eryAD into a T4 polymerase-treated

TABLE 2. Competence and β-Gal induction in independent clones of mutant strains CP1253, CP1254, and CP1255

Strain	Clone no.	Competence (Nov^r/μl)[a]				β-Gal activity (M.U.)[b]			
		Not induced	Induced by:			Not induced	Induced by:		
			pH shift	CSP	CSP + pH shift		pH shift	CSP	CSP + pH shift
CP1253	1	0	1,780	792	2,460	0.6	6.5	5.2	8.2
	2	0	2,140	1,560	2,400	0.5	7.2	6.1	8.9
CP1254	1	0	34	22	42	0.8	3.8	3.2	3.3
	2	0	48	18	59	0.4	3.0	2.6	3.5
CP1255	1	0	272	47	288	0.3	34.2	31.5	35.9
	2	0	304	68	344	3.1	37.5	31.4	38.8
CP1250		0	2,540	605	2,200	0.2	0.3	0.2	0.2

[a] Two independent clones of each inducible insertion mutant were grown in CTM containing 9 mM HCl to an OD_{550} of 0.1. DNA was added to 2 μg/ml, and the cultures were incubated without further treatment or were induced by adding NaOH (shifting the pH), CSP, or both. Incubation was continued for 45 min. Competence was measured as the number of Nov^r transformants in 1 μl of transformed culture.

[b] Activity of β-Gal (in Miller units [M.U.]) was measured in crude cell extracts from the same cultures 45 min after competence induction. Part (100 μl) of each extract was used for the ONPG assay.

*Bsm*I site of pXF512. To replace the chromosomal *orfr1* with the disrupted copy, pXF526 was linearized with *Sal*I and transformed into *S. pneumoniae* CP1250.

A 895-bp fragment of the *cgl* chromosomal locus that included the 3′ end of *cglB*, the complete *cglC* gene, and the 5′ end of *cglE*, was replaced with the syn *erm* cassette, using pXF528. To make pXF528, an *Xba*I-*Fsp*I fragment of PCR-amplified syn *erm* was inserted into pXF513 digested with *Pml*I and *Spe*I. *cglE* was disrupted by an insertion of pEVP3 derivative pXF527, created by subcloning a 351-bp *Pst*I-*Sph*I fragment of *cglE* from pXF513 into *Nsi*I and *Xba*I sites of pEVP3.

Nucleotide sequence accession numbers. The *coi*, *cel*, and *cgl* nucleotide sequences are available at GenBank as accessions AF052207, AF052208, and AF052209, respectively.

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^r 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 inserts 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,

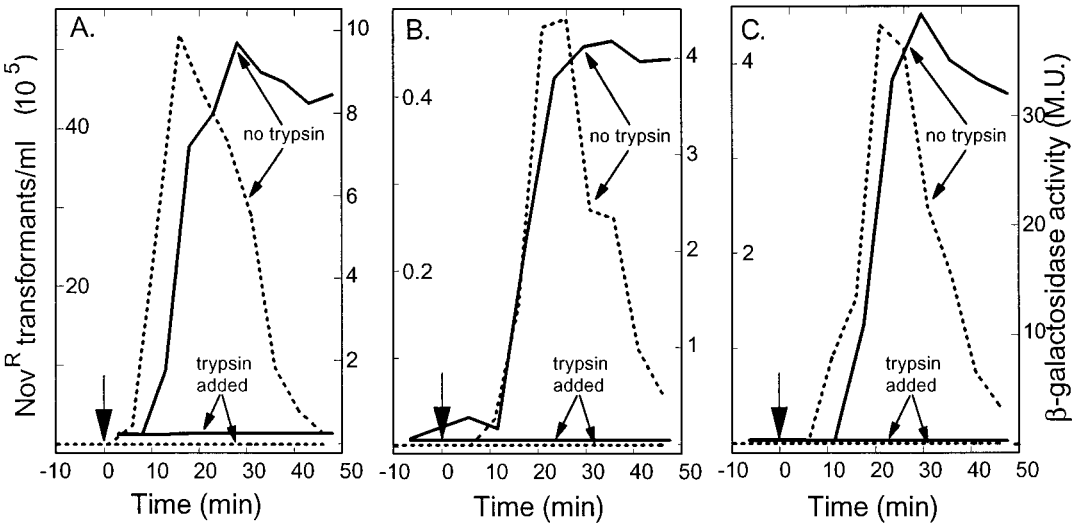

FIG. 1. Competence and β-Gal production in mutant strains after induction with the pH shift. Mutant strains CP1253 (A), CP1254 (B), and CP1255 (C) are shown. A culture of each mutant strain was grown from 10^6 cells/ml to an OD_{350} of 0.15 in CTM containing 9 mM HCl and split in two portions, one of 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 (42) 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.

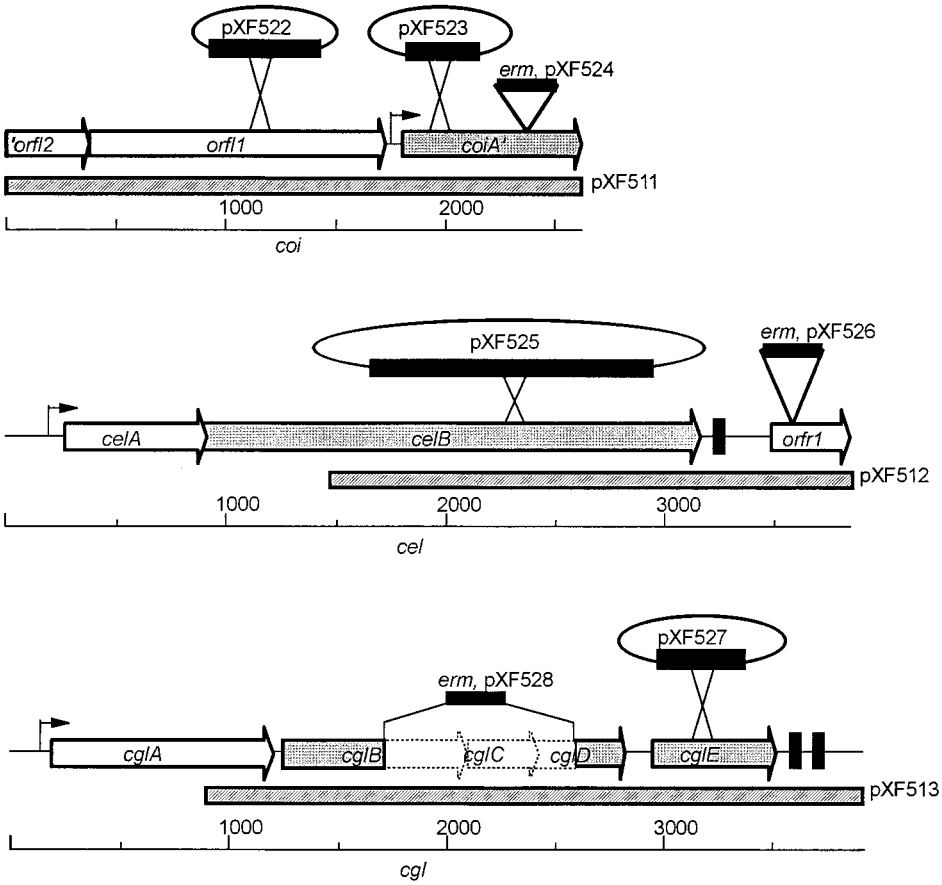

FIG. 2. Organization and gene disruption mutagenesis of *coi*, *cel*, and *cgl* loci. The genetic map of each locus indicates ORFs and putative promoter (small arrows) and terminator (wide vertical line) regions. The insertion plasmids pXF522, pXF523, pXF525, and pXF527 are shown as ellipses above the restriction map of each locus, with targeting fragments as boxes. Plasmids pXF524 and pXF526 carry syn *erm* inserted into *coiA* and *cel orfr1*, respectively. In pXF528, genes *cglB* to *cglD* are replaced with syn *erm*. Syn *erm* carried by pXF524, pXF526, and pXF528 is indicated by a black box. The portion of the chromosomal *cgl* locus replaced with syn *erm* is shown by broken lines. Shaded genes are those for which disruption results in transformation deficiency. The targeting fragments of pXF511, pXF512, and pXF513 plasmids are shown below the map line (in base pairs).

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 created by the integration of pXF512 (Fig. 2) was 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

TABLE 3. Competence and β-Gal production in independent clones of different insertion strains

Site of gene disruption	Donor plasmid or strain	Clone no.	Competence ($Nov^r/\mu l$)[a,c]		β-Gal activity (M.U.)[b,c]	
			Not induced	Induced by CSP	Not induced	Induced by CSP
coi orfl1	pXF522	1	0	2,880	6.1	6.2
		2	0	2,550	6.8	6.7
coiA	pXF523	1	0	21	1.5	4.4
		2	3	24	2.6	4.4
coiA	pXF524	1	1	26		
		2	0	23		
celB	pXF525	1	0	15	1.6	14.4
		2	0	18	1.2	19.3
cel orfr1	pXF526	1	1	2,680		
		2	0	2,960		
cglB to *cglD*	pXF528	1	0	0		
		2	0	0		
cglE	pXF527	1	0	2	2.2	12.3
		2	0	12	1.5	12.7
None	CP1250		0–2	2,620–2,400	0.2–0.5	0.2–0.6

[a] Independent clones of insertion mutants were grown in 10 ml of CTM 9 mM HCl. At an OD_{550} of 0.1, DNA was added to 2 μg/ml. Cultures were split into two 5-ml portions, one of which was induced by adding CSP to 100 ng/ml. Competence was measured as the number of Nov^r transformants in 1 μl of culture at 30 min after induction.

[b] Activity of β-Gal was measured in crude cell extracts from the cultures of pEVP3 insertion strains. M.U., Miller units.

[c] Values for CP1250 control cultures assayed in nine experiments are given as ranges.

encoded by the *celA* 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 *celB* 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, CelB 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 Rec2 protein of *H. influenzae* (31) in a 237-aa overlap (Fig. 3A).

Four of five putative genes of the *S. pneumoniae cgl* locus, *cglA*, *cglB*, *cglC*, and *cglD*, 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, *cglE*, 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 (PilC and PilG [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 [boldfaced 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, *cglE*, 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-X-G-X-G-K) essential for the ATP

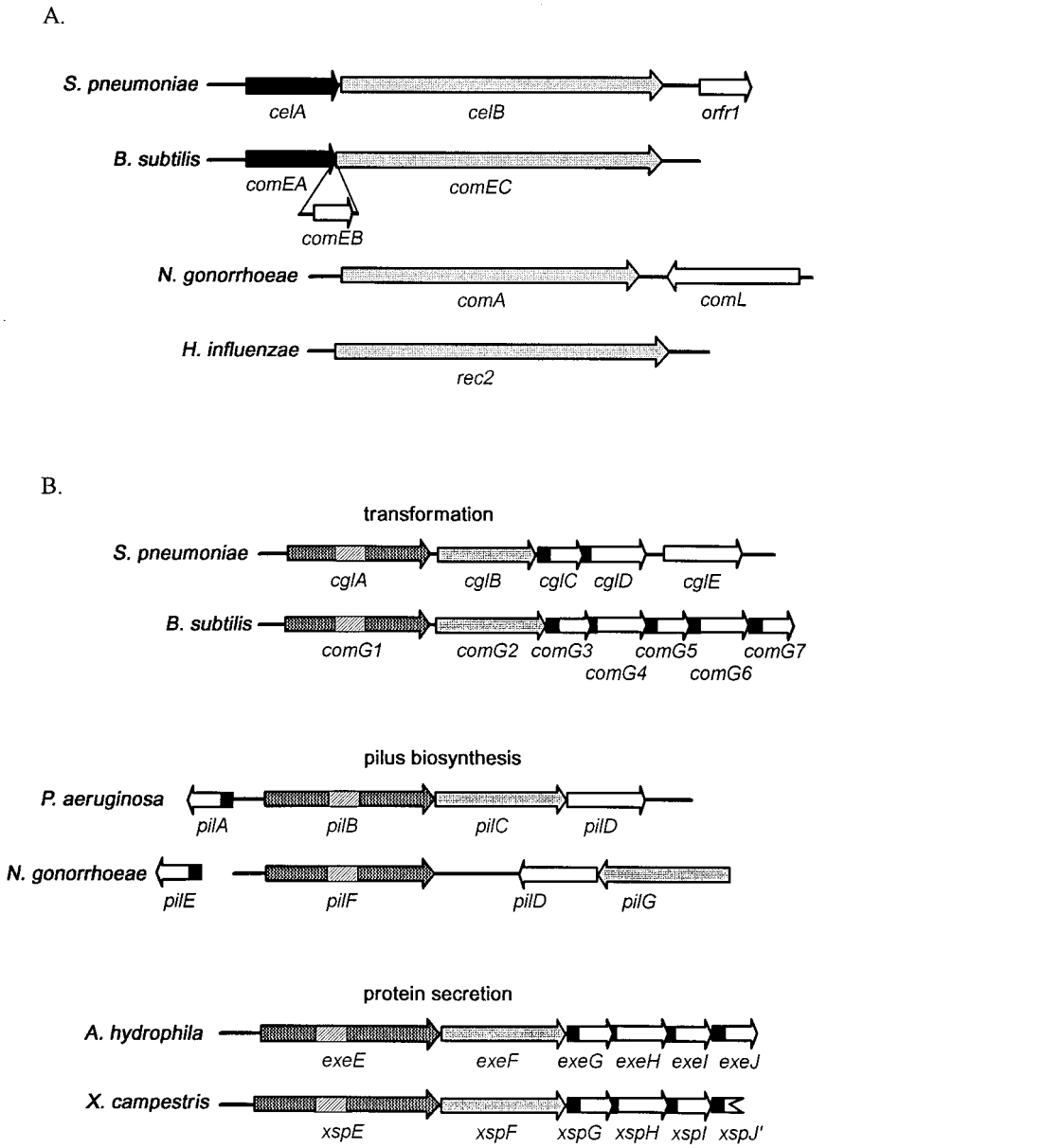

FIG. 3. Homologs of *S. pneumoniae cel* and *cgl* loci. (A) Comparison of *S. pneumoniae cel* locus to *B. subtilis comE* operon (GenBank accession no. L15202), *N. gonorrhoeae com* locus (S75490), and *H. influenzae rec2* gene (U32691). Genes encoding similar protein products are shown by the same shading; 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* (U32588), *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.

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 TTACGAATANTAT AGG, preceded by an AT-rich region.

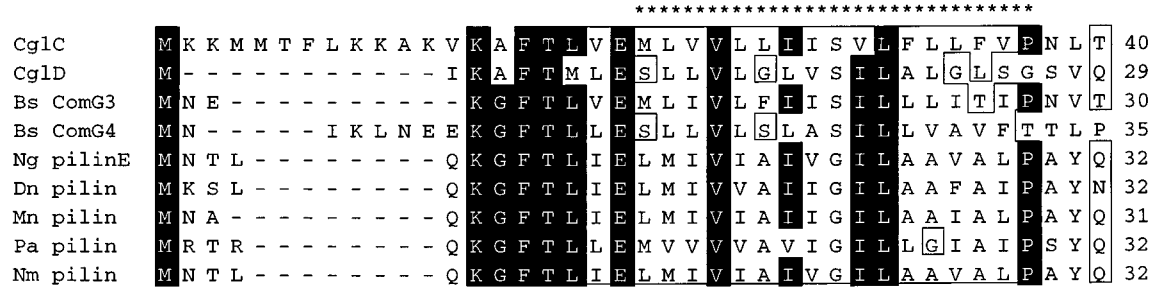

FIG. 4. Similarity of CglC and CglD to pilins and *B. subtilis* pilin-like proteins ComG3 and ComG4. Multiple-sequence alignment of N-terminal portions of *S. pneumoniae* CglC and CglD; *B. subtilis* (Bs) pilin-like proteins ComG3 and ComG4; and pilins from *Neisseria gonorrhoeae* (Ng) (pilin E; accession no. Z69262), *Dichelobacter nodosus* (Dn) (M92188), *Moraxella nonliquefaciens* (Mn) (P09829), *P. aeruginosa* (Pa) (L76605), and *Neisseria meningitidis* (Nm) (Z49820). Amino acid residues that are similar in the majority of the proteins (six of nine) are boxed. Amino acids that are identical in the majority of the proteins are indicated by white letters on black background. The hydrophobic domain following the K-G-F-T-L processing consensus sequence is indicated by the asterisks. Gaps introduced to maximize alignment are indicated by dashes.

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

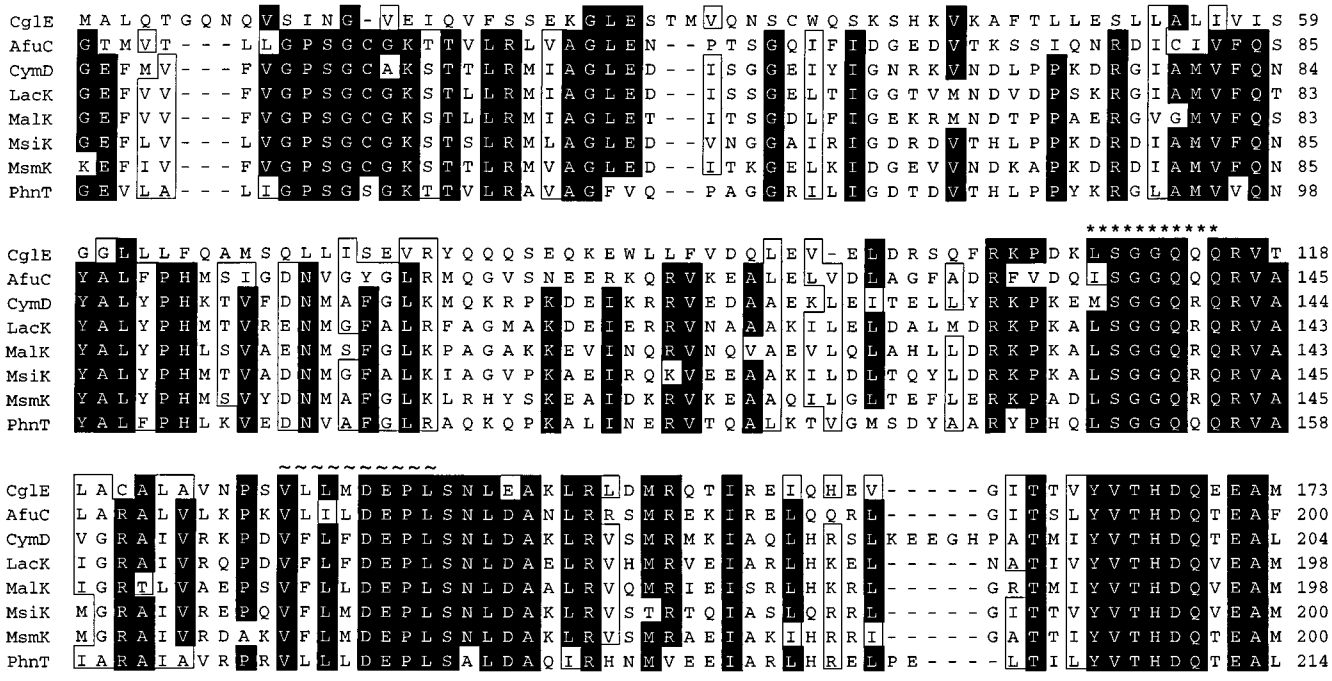

FIG. 5. Similarity of *S. pneumoniae* CglE to ATP-binding components of traffic ATPases. Multiple-sequence alignment of regions of similarity between *S. pneumoniae* CglE (residues 1 to 173), *Actinobacillus pleuropneumoniae* AfuC protein involved in iron transport (accession no. U04954 [7], residues 30 to 200), *Klebsiella oxytoca* CymD component of cyclodextrin transporter (S55406 [16], residues 29 to 204), *Agrobacterium radiobacter* LacK protein involved in lactose transport (Q01937 [62], residues 28 to 198), *E. coli* maltose-transporter component MalK (P02914 [18], residues 28 to 198), cellobiose/xylobiose transporter component MsiK of *Streptomyces lividans* (U12007 [23], residues 30 to 200), multiple-sugar-transport ATP-binding protein MsmK of *Streptococcus mutans* (Q00752 [47], residues 30 to 200), and *Salmonella typhimurium* LT2 2-aminoethylphosphate transporter component PhnT (U69483 [unpublished], residues 43 to 214). Amino acid residues that are similar in the majority of the proteins (six of nine) are boxed. Amino acids that are identical in the majority of the proteins are indicated by white letters on black background. The linker peptide with the consensus sequence L-S-G-G-Q-Q (***) and the Walker B nucleotide binding motif (~~~) are indicated. Gaps introduced to maximize alignment are indicated by dashes.

FIG. 6. Putative promoter region of competence-specific operons. Alignment of nucleotide sequences located upstream of *coiA*, *celA*, *cglA*, and *exp10* initiation codons. Regions of identity are shown by the black background, while residues similar in three of four sequences are boxed. The consensus sequence is represented above the aligned sequences. Ribosome binding sites and ATG start codons are shaded. Gaps introduced to maximize alignment are indicated by dashes.

encode DNA transport proteins. This sequence similarity and the importance of *celB* 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* and *exeF* genes represented in Fig. 3) 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 be 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 CglC and CglD 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 CglC and CglD 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, PilE, 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 CglC and CglD.

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

1. **Albano, M., R. Breitling, and D. Dubnau.** 1989. Nucleotide sequence and genetic organization of the *Bacillus subtilis comG* operon. J. Bacteriol. **171:** 5386–5404.
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. **Avery, O. T., M. MacLeod, and M. McCarty.** 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. **79:**137–158.
4. **Boguski, M. S.** 1994. Bioinformatics. Curr. Opin. Genet. Dev. **4:**383–388.
5. **Breitling, R., and D. Dubnau.** 1990. A pilin-like membrane protein is essential for DNA binding by competent *Bacillus subtilis*. J. Bacteriol. **172:**1499–1508.
6. **Campbell, E. A., and H. R. Masure.** 1996. Competence induced promoters in *S. pneumoniae*, abstr. H109, p. 502. *In* Abstracts of the 96th General Meeting of the American Society for Microbiology. American Society for Microbiology, Washington, D.C.
7. **Chin, N., J. Frey, C. F. Chang, and Y. F. Chang.** 1996. Identification of a locus involved in the utilization of iron by *Actinobacillus pleuropneumoniae*. FEMS Microbiol. Lett. **143:**1–6.
8. **Chung, Y. S., and D. Dubnau.** 1995. ComC is required for the processing and translocation of ComGC, a pilin-like competence protein of *Bacillus subtilis*. Mol. Microbiol. **15:**543–551.
9. **Claverys, J.-P., A. Dintilhac, E. V. Pestova, B. Martin, and D. A. Morrison.** 1995. Construction and evaluation of new drug-resistance cassettes for gene disruption mutagenesis in *Streptococcus pneumoniae*, using an *ami* test platform. Gene **164:**123–128.
10. **Doige, C. A., and G. F.-L. Ames.** 1993. ATP-dependent transport systems in bacteria and humans: relevance to cystic fibrosis and multidrug resistance. Annu. Rev. Microbiol. **47:**291–319.
11. **Dubnau, D.** 1993. Genetic exchange and homologous recombination, p. 555–584. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. Biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
12. **Dums, F., J. M. Dow, and M. J. Daniels.** 1991. Structural characterization of protein secretion genes of the bacterial phytopathogen *Xanthomonas campestris* pathovar *campestris*: relatedness to secretion systems of other gram-negative bacteria. Mol. Gen. Genet. **229:**357–364.
13. **Elleman, T. C., D. J. Stewart, K. G. Finney, P. A. Hoyne, and C. W. Ward.** 1990. Pilins from the B serogroup of *Bacteroides nodosus*: characterization, expression and cross-protection. Infect. Immun. **58:**1545–1551.
14. **Emmert, D. B., P. J. Stoehr, G. Stoesser, and G. N. Cameron.** 1994. The European bioinformatics institute (EBI) databases. Nucleic Acids Res. **22:** 3445–3449.
15. **Facius, D., and T. F. Meyer.** 1993. A novel determinant (ComA) essential for natural transformation competence in *Neisseria gonorrhoeae* and the effect of a *comA* defect on pilin variation. Mol. Microbiol. **10:**699–712.
16. **Fiedler, G., M. Pajatsch, and A. Bock.** 1996. Genetics of a novel starch utilization pathway present in *Klebsiella oxytoca*. J. Mol. Biol. **256:**279–291.
17. **Freitag, N. E., H. S. Seifert, and M. Koomey.** 1995. Characterization of the

pilF-pilD pilus-assembly locus of *Neisseria gonorrhoeae*. Mol. Microbiol. **16:** 575–586.
18. **Gilson, E., H. Nikaido, and M. Hofnung.** 1982. Sequence of the *malK* gene in *E. coli* K12. Nucleic Acids Res. **10:**7449–7458.
19. **Hahn, J., G. Inamine, Y. Kozlov, and D. Dubnau.** 1993. Characterization of *comE*, a late competence operon of *B. subtilis* required for the binding and uptake of transforming DNA. Mol. Microbiol. **10:**99–111.
20. **Håvarstein, L. S., G. Coomaraswamy, and D. A. Morrison.** 1995. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. Proc. Natl. Acad. Sci. USA **92:** 11140–11144.
21. **Higgins, D. G., A. J. Bleasby, and P. M. Sharp.** 1992. ClustalV: improved software for multiple sequence alignment. CABIOS **8:**189–191.
22. **Howard, S. P., J. Critch, and A. Bedi.** 1993. Isolation and analysis of eight *exe* genes and their involvement in extracellular protein secretion and outer membrane assembly in *Aeromonas hydrophila*. J. Bacteriol. **175:**6695–6703.
23. **Hurtubise, Y., F. Shareck, D. Kluepfel, and R. Morosoli.** 1995. A cellulase/xylanase-negative mutant of *Streptomyces lividans* 1326 defective in cellobiose and xylobiose uptake is mutated in a gene encoding a protein homologous to ATP-binding proteins. Mol. Microbiol. **17:**367–377.
24. **Hyde, S. C., P. Emsley, M. J. Hartshorn, M. M. Mimmack, U. Gileadi, S. R. Pearce, M. P. Gallagher, D. R. Gill, R. E. Hubbard, and C. F. Higgins.** 1990. Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature **346:**362–365.
25. **Inamine, G. S., and D. Dubnau.** 1995. ComEA, a *B. subtilis* integral membrane protein required for genetic transformation, is needed both for DNA binding and transport. J. Bacteriol. **177:**3045–3051.
26. **Jonsson, A. B., D. Ilver, P. Falk, J. Pepose, and S. Normak.** 1994. Sequence changes in the pilus subunit lead to tropism variation of *Neisseria gonorrhoeae* to human tissue. Mol. Microbiol. **13:**403–416.
27. **Lauer, P.** 1993. Conservation of genes encoding components of a type IV pilus assembly two-step protein export pathway in *Neisseria gonorrhoeae*. Mol. Microbiol. **8:**357–368.
28. **Lorenz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. **58:**563–602.
29. **Malke, H.** 1986. Codon usage in streptococci. J. Basic Microbiol. **26:**587–595.
30. **Martin, B., P. Garcia, M.-P. Castanie, and J.-P. Claverys.** 1995. The *recA* gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction. Mol. Microbiol. **15:**367–379.
31. **McCarthy, D.** 1989. Cloning of the *rec2* locus of *Haemophilus influenzae*. Gene **75:**135–143.
32. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* beta-lactamase gene. J. Biol. Chem. **256:**11283–11291.
33. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Morrison, D. A., and M. F. Baker.** 1979. Competence for genetic transformation in pneumococcus depends on synthesis of a small set of proteins. Nature **282:**215–217.
35. **Morrison, D. A., S. A. Lacks, W. R. Guild, and J. M. Hageman.** 1983. Isolation and characterization of three new classes of transformation-deficient mutants of *Streptococcus pneumoniae* that are defective in DNA transport and genetic recombination. J. Bacteriol. **156:**281–290.
36. **Morrison, D. A., B. Mannarelli, and M. N. Vijayakumar.** 1982. Competence for transformation in *Streptococcus pneumoniae*: an inducible high-capacity system for genetic exchange, p. 136–138. *In* D. Schlessinger (ed.), Microbiology—1982. American Society for Microbiology, Washington, D.C.
37. **Murray, C. L., and J. C. Rabinowitz.** 1982. Nucleotide sequences of transcription and translation initiation regions in bacillus phage phi-29 early genes. J. Biol. Chem. **257:**1053–1062.
38. **Nunn, D., S. Bergman, and S. Lory.** 1990. Products of three accessory genes, *pilB*, *pilC*, and *pilD*, are required for biogenesis of *Pseudomonas aeruginosa* pili. J. Bacteriol. **172:**2911–2919.
39. **Pearce, B. J., A. M. Naughton, E. A. Campbell, and H. R. Masure.** 1995. The *rec* locus, a competence-induced operon in *Streptococcus pneumoniae*. J. Bacteriol. **177:**86–93.
40. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
41. **Pestova, E. V., L. S. Håvarstein, and D. A. Morrison.** 1996. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. Mol. Microbiol. **21:**853–862.
42. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55:**339–351.
43. **Pugsley, A. P.** 1996. Multimers of the precursor of a type IV pilin-like component of the general secretory pathway are unrelated to pili. Mol. Microbiol. **20:**1235–1245.
44. **Puyet, A., B. Greenberg, and S. Lacks.** 1990. Genetic and structural characterization of EndA, a membrane-bound nuclease required for transformation of *S. pneumoniae*. J. Mol. Biol. **213:**727–738.
45. **Rudel, T., D. Facius, R. Barten, I. Sheuerpflug, E. Nonnenmacher, and T. F. Meyer.** 1995. Role of pili and the phase-variable PilC protein in natural competence for transformation of *Neisseria gonorrhoeae*. Proc. Natl. Acad. Sci. USA **92:**7986–7990.
46. **Russel, M. A., and A. Darzins.** 1994. The *pilE* gene product of *Pseudomonas aerugenosa*, required for pili biogenesis, shares amino acid sequence identity with the N-termini of type 4 prepilin proteins. Mol. Microbiol. **13:**973–985.
47. **Russel, R. R., J. Aduse-Opoku, I. C. Sutcliffe, L. Tao, and J. J. Ferretti.** 1992. A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. J. Biol. Chem. **267:**4631–4637.
48. **Sabelnikov, A. G., B. Greenberg, and S. A. Lacks.** 1995. An extended −10 promoter alone directs transcription of the *Dpn* II operon of *Streptococcus pneumoniae*. J. Mol. Biol. **250:**144–145.
49. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
50. **Schoolnik, G. K., R. Fernandez, J. Y. Tai, J. Rothbard, and E. C. Gotschlich.** 1984. Gonococcal pili. Primary structure and receptor binding domain. J. Exp. Med. **159:**1351–1370.
51. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature **254:**34–38.
52. **Solomon, J. M., and A. D. Grossman.** 1996. Who's competent and when: regulation of natural genetic competence in bacteria. Trends Genet. **12:** 150–155.
53. **Stassi, D. L., P. Lopez, M. Espinosa, and S. A. Lacks.** 1981. Cloning of chromosomal genes in *Streptococcus pneumoniae*. Proc. Natl. Acad. Sci. USA **78:**7028–7032.
54. **Taha, M. K., D. Giorgini, and X. Nassif.** 1996. The *pilA* regulatory gene modulates the pilus-mediated adhesion of *Neisseria meningitidis* by controlling the transcription of *pilC1*. Mol. Microbiol. **19:**1073–1084.
55. **Tomasz, A.** 1966. Model for the mechanism controlling the expression of competent state in pneumococcus cultures. J. Bacteriol. **91:**1050–1061.
56. **Tomasz, A., and R. D. Hotchkiss.** 1964. Regulation of the transformability of pneumococcal cultures by macromolecular cell products. Proc. Natl. Acad. Sci. USA **51:**480–486.
57. **Tonjum, T., N. E. Freitag, E. Namork, and M. Koomey.** 1995. Identification and characterization of *pilG*, a highly conserved pilus-assembly gene in pathogenic Neisseria. Mol. Microbiol. **16:**451–464.
58. **Tonjum, T., C. F. Marrs, F. Rozsa, and K. Bovre.** 1991. The type 4 pilin of *Moraxella nonliquefaciens* exhibits unique similarities with the pilins of *Neisseria gonorrhoeae* and *Dichelobacter* (*Bacteroides*) *nodosus*. J. Gen. Microbiol. **137:**2483–2490.
59. **Vasseghi, H., and J.-P. Claverys.** 1983. Amplification of a chimeric plasmid carrying an erythromycin-resistance determinant introduced into the chromosome of *Streptococcus pneumoniae*. Gene **21:**285–292.
60. **Vasseghi, H., J.-P. Claverys, and A. M. Sicard.** 1981. Mechanism of integrating foreign DNA during transformation of *Streptococcus pneumoniae*, p. 137–154. *In* M. Polsinelli and G. Mazza (ed.), Transformation—1980. Proceedings of the 5th European Meeting on Bacterial Transformation and Transfection. Cotswold Press Ltd., Oxford, England.
61. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
62. **Williams, S. G., J. A. Greenwood, and C. W. Jones.** 1992. Molecular analysis of the *lac* operon encoding the binding-protein-dependent lactose transport system and beta-galactosidase in *Agrobacterium radiobacter*. Mol. Microbiol. **6:**1755–1768.