JOURNAL OF BACTERIOLOGY, July 1988, p. 3136–3141
0021-9193/88/073136-06$02.00/0

Vol. 170, No. 7

# Identification of Two Proteins Encoded by *com*, a Competence Control Locus of *Streptococcus pneumoniae*

MARK S. CHANDLER[1]† AND DONALD A. MORRISON[2]*

*Laboratory for Cell, Molecular, and Developmental Biology[2] and Department of Microbiology and Immunology,[1] University of Illinois at Chicago, Chicago, Illinois 60680*

Received 30 November 1987/Accepted 20 April 1988

**The *com* locus, which controls competence for genetic transformation in *Streptococcus pneumoniae*, was analyzed by construction of a series of subclones, insertion mutations, and deletions of the cloned DNA in *Escherichia coli*. In vitro transcription-translation of these *com* plasmids revealed two neighboring genes, *comA* and *comB*, encoding proteins of 77,000 and 49,000 daltons, respectively. Their map positions and orientations were determined. Insertions in either gene eliminated the corresponding protein and had no effect on the other. In addition, a 15,000-dalton *com* protein was tentatively identified, although the exact location of this gene remains to be determined. Features of the DNA adjacent to the *com* locus are also described.**

We recently reported the cloning of *com*, a locus from *Streptococcus pneumoniae* (pneumococcus) involved in the control of competence for genetic transformation (5). The initial *com* mutant failed to release competence factor (CF) but responded normally to exogenous CF (19). *com* is the first locus involved in the control of competence to be cloned and thus offers new opportunities for identifying specific genes and studying the roles of their gene products in competence.

Competence in pneumococcal cells allows the uptake of a single-strand fragment from a molecule of DNA and the genetic recombination of that DNA with the cell chromosome (8, 13). Competence is a distinct cell state which actively growing cells enter at a particular cell density, as controlled by the small secreted protein CF. When CF reaches a threshold level, all the cells of a culture develop competence (22, 24, 25) in a response accompanied by a global shift of protein synthesis to the production of a few competence-specific proteins (17, 18). Accompanying this shift in protein synthesis is the appearance of the cell properties associated with competence, including an efficient system for processing DNA for recombination. An analogous control of the competent state has been described for other streptococcal transformation systems (9, 10, 15, 16, 20, 21, 25).

Analysis of gene functions in the control of competence in pneumococci has been limited. Although CF controls the competent state, its mechanism of action is obscure. CF has been characterized as a basic protein with a size in the neighborhood of 10,000 daltons (25). The protein does not interact with DNA in vitro, and its activity in vivo may be at the cell surface (25). The competent state usually terminates after a short time, perhaps in response to an inhibitor of activation (24). These are the only two proteins identified as having a role in the control of competence. Details about these proteins, or any other proteins involved in the control of competence, remain to be determined.

The *com* locus is more complex than would be expected for a gene encoding the small protein CF. We have shown that it is 4.2 to 5.2 kilobases (kb) in length and have tentatively identified several proteins encoded by this region (5). To identify the genes of the *com* locus, their orientations, and their protein products more precisely, we have constructed nested truncations of the cloned *com* locus. In vitro DNA-directed transcription-translation with these plasmids permitted mapping of two competence genes, which encoded 77- and 49-kilodalton (kDa) polypeptides.

## MATERIALS AND METHODS

**Bacterial strains, plasmids, and culture media.** The plasmids containing *com* DNA used in this study are described in Table 1. The plasmid vector used was the promoter selection vector pKK232-8 (2). pKK232-8 carries a promoterless chloramphenicol acetyltransferase (CAT) gene preceded by polylinker cloning sites and by translational stop codons in three reading frames and followed by a tandem duplication of the *Escherichia coli rrnB* rRNA operon transcriptional terminators. Thus, insertion of a DNA fragment containing an active promoter in this polylinker can be selected for by its activation of the CAT gene (2), while the plasmid is protected from excessive transcription by the transcriptional terminators (3). Plasmids pR29 (7) and pPS12 (6) have been described. *E. coli* DH1 (11) was used as a plasmid host. Culture conditions and selection for resistance to ampicillin, chloramphenicol, and erythromycin were as described earlier (5).

**DNA procedures.** DNA restriction and modification enzymes were used according to the recommendations of the suppliers. Restriction fragments used in cloning experiments were purified from agarose by electrophoresis onto DEAE membranes as previously described (5) or by use of Geneclean (Bio 101, Inc.). Preparation of frozen competent cells, clone analysis, and plasmid preparation were as reported earlier (5).

**Mutant phenotype determination.** The *comB* mutants obtained from plasmids pXF209 and pXF219 were tested for induction of competence on exposure to CF as described previously (5) for *comA* insertion mutations.

**Construction of pXF223.** pXF223 was constructed by isolating the 2,160-base-pair (bp) *Bam*HI-*Hin*dIII *com* fragment from pXF180 and ligating it to a *Bam*HI-*Hin*dIII digest of pKK232-8. The *Bam*HI site of pXF180 is located in the polylinker 3 bp to the left of the *Dra*I-*Sma*I ligation site (Fig.

* Corresponding author.
† Present address: Department of Molecular Biology and Genetics, Johns Hopkins University, Baltimore, MD 21205.

TABLE 1. *com* plasmids used in this study

| Plasmid | Cloning site(s) in pKK232-8[a] | Description of *com* fragment insert[b] | Source of *com* DNA | Reference |
|---|---|---|---|---|
| pXF170 | *Hin*dIII | *Hin*dIII (0–2.67) | CP1200 | 5 |
| pXF180 | *Sma*I | *Dra*I (4.67–0.51) | CP1200 | 5 |
| pXF190 | *Hin*dIII | *Hin*dIII (2.67–5.97) | CP1200 | 5 |
| pXF206 | *Sma*I | *Dra*I (4.67–0.51)Ω*ermB*(3.34::*Taq*I>) | pXF180 | 5 |
| pXF208 | *Sma*I | *Dra*I (4.67–0.51)Ω*ermB*(1.6::*Hpa*II>) | pXF180 | 5 |
| pXF209 | *Sma*I | *Dra*I (4.67–0.51)Ω*ermB*(4.65::*Hpa*II>) | pXF180 | 5 |
| pXF219 | *Hin*dIII | *Hin*dIII (2.67–5.97)Ω*ermB*(3.66::*Taq*I>) | pXF190 | 5 |
| pXF223 | *Bam*HI, *Hin*dIII | *Dra*I-*Hin*dIII (0.51–2.67) | pXF180 | This study |
| pXF224 | *Hin*dIII | *Dra*I-*Taq*I (0.51–2.03) | pXF204 | This study |
| pXF225 | *Hin*dIII | *Dra*I-*Hpa*II (0.51–1.6) | pXF208 | This study |
| pXF229 | *Bam*HI, *Acc*I | *Dra*I-*Asu*II (0.51–1.22) | pXF180 | This study |
| pXF231 | *Sma*I | *Dra*I (4.67–0.51)Δ[*Kpn*I (4.27)-*Kpn*I (4.47)] | pXF180 | This study |
| pXF232 | *Hin*dIII | *Hin*dIII (2.67–5.97)Δ[*Kpn*I (4.27)-*Kpn*I (4.47)] | pXF190 | This study |
| pXF233 | *Sma*I | *Hpa*II (3.25–4.11) | pXF180 | This study |
| pXF234 | *Sma*I | *Ava*II-*Kpn*I (5.72–4.47) | pXF190 | This study |
| pXF235 | *Sma*I | *Kpn*I-*Ava*II (4.47–5.72) | pXF190 | This study |

[a] Each *com* plasmid listed consists of pKK232-8 with an insert at the indicated polylinker site(s) that expressed promoter activity to give a Cm$^r$ phenotype. The *ermB* *Cla*I fragment from pR29 also confers Em$^r$.

[b] Restriction sites (map positions in parentheses in kilobases [Fig. 1]) at the endpoints of the pneumococcal insert. To indicate orientation of the insert, its ends are given clockwise, with the end nearer *amp* first and that nearer CAT second. For mutations constructed in vitro, the site used for *ermB* insertion and the direction of transcription of *ermB* (>; to the right on the map in Fig. 1) are also indicated. The endpoints of deletions constructed in vitro are shown in square brackets.

1), while the relevant *Hin*dIII site is at bp 2665 on the map (Fig. 1).

**Construction of pXF224 and pXF225.** pXF204 and pXF208, constructed previously (5), are derivatives of pXF180, in which a *Cla*I fragment containing an *ermB* gene was inserted at a restriction site in the *com* DNA. This *Cla*I fragment carries a single *Hin*dIII site 6 bp from one end. The *Cla*I inserts in pXF204 and pXF208 are oriented with the *Hin*dIII site on the side closest to the *Dra*I site at bp 510. A 1,550-bp *Hin*dIII fragment of pXF204 was isolated from agarose and ligated to a *Hin*dIII digest of pKK232-8 to form pXF224. A 1,120-bp *Hin*dIII fragment of pXF208 was isolated from agarose and ligated to a *Hin*dIII digest of pKK232-8 to form pXF225.

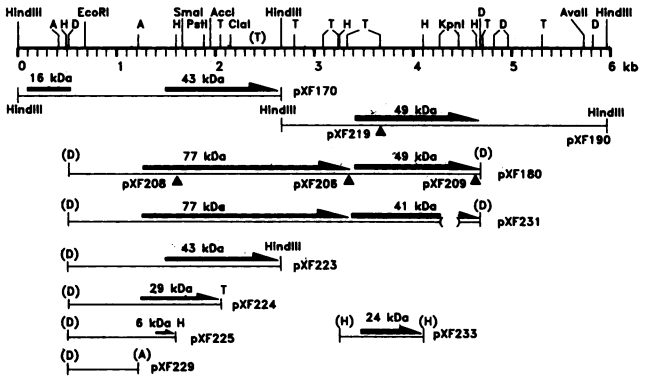


FIG. 1. Analysis of genes of the *com* locus. At the top is a physical map of the *com* locus. The lower lines represent cloned *com* fragments and indicate the extent of pneumococcal DNA present in each plasmid. Parentheses indicate a deletion. The names of additional plasmids carrying *ermB* insertions are also indicated at the site of insertion (▲) in the corresponding parental plasmid. For each subclone, the arrows indicate the approximate positions and orientations, when known, assigned to genes for the protein products observed in in vitro transcription-translation; the size of each observed polypeptide is indicated above the arrow. Restriction enzyme abbreviations: A, *Asu*II; D, *Dra*I; H, *Hpa*II; T, *Taq*I; (T), a *Taq*I site known to lie at kb 2.25 or 2.55. Restriction sites destroyed during subcloning are in parentheses. Sites absent from this locus include *Bam*HI, *Bcl*I, *Pvu*II, *Sal*I, and *Xho*I.

**Construction of pXF229, pXF231, pXF232, pXF233, pXF234, and pXF235.** The 715-bp *Bam*HI-*Asu*II fragment of pXF180 was isolated and ligated to the 5,085-bp *Bam*HI-*Acc*I fragment of pKK232-8 to form pXF229. Deletion of a *Kpn*I fragment from pXF180 and pXF190 gave rise to pXF231 and pXF232, respectively. The 855-bp *Hpa*II (kb 3.25 to 4.11) fragment of pXF180 was isolated from agarose. Its staggered ends were filled with T4 DNA polymerase; ligation to *Sma*I-digested pKK232-8 formed pXF233. pXF234 and pXF235 contain the 1,255-bp *Kpn*I-*Ava*II (kb 4.46 to 5.72) fragment of pXF190 in opposite orientations. They were constructed by isolating the *Kpn*I-*Ava*II fragment from agarose; the ends were then trimmed and filled with T4 DNA polymerase and ligated to a *Sma*I digest of pKK232-8.

**In vitro transcription-translation and electrophoresis.** Plasmids purified by ethidium bromide-CsCl centrifugation were used to direct polypeptide synthesis and labeling in a cell-free coupled transcription-translation system derived from *E. coli* (Amersham Corp.). [$^{35}$S]methionine-labeled polypeptides were separated on vertical sodium dodecyl sulfate-polyacrylamide (8 to 20% linear gradient) gels (14), followed by fluorography (1). $^{14}$C-labeled methylated-protein size standards (Amersham Corp.) were insulin (2.35 and 3.4 kDa), aprotinin (6.5 kDa), cytochrome *c* (12.5 kDa), lysozyme (14.3 kDa), soybean trypsin inhibitor (21.5 kDa), carbonic anhydrase (30.0 kDa), ovalbumin (46 kDa), bovine serum albumin (69 kDa), phosphorylase *b* (92.5 kDa), and myosin (200 kDa).

## RESULTS

**Isolation of plasmids containing nested truncations of the *com* locus.** pXF170, pXF180, and pXF190 contain three overlapping segments from the *com* locus of *S. pneumoniae* (5). Earlier results from *E. coli* maxicell experiments suggest that at least three proteins are encoded in this region (5). To analyze individual gene loci and identify their gene products, we constructed a series of nested subclones of pXF180, as

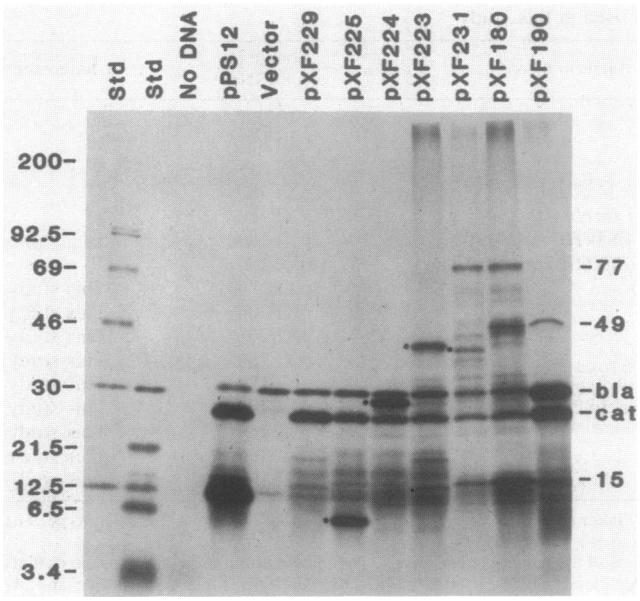


FIG. 2. In vitro transcription-translation analysis of plasmids containing nested truncations of the *com* locus. The [$^{35}$S]methionine-labeled polypeptides were separated in a sodium dodecyl sulfate-polyacrylamide (8 to 20% linear gradient) gel, followed by fluorography. The plasmid studied (or protein size standard [Std]) is indicated at the top. Vector gene products and *com* protein sizes (in kilodaltons) are indicated on the right. Sizes (in kilodaltons) and positions of protein standards are shown on the left. Dots indicate bands interpreted as truncated *com* proteins.

described in Materials and Methods and in Table 1. The *com* locus fragments carried by these subclones are illustrated in Fig. 1. pXF231 is a deletion derivative of pXF180 with the 200-bp *Kpn*I fragment removed. Except as noted in Table 1 for the pXF180 family (pXF231, pXF235, pXF206, pXF208, and pXF209), the orientation of each *com* DNA insert in the vector pKK232-8 is such that a promoter within the *com* DNA directs transcription left to right (as drawn) into the promoterless CAT gene of the vector. Preceding this promoterless CAT gene are translational stop codons in all three reading frames, lying between 4 and 70 bp from the end of the insert, depending on the cloning site and the reading frame.

**Identification of *com* polypeptides.** Truncated *com*-region plasmids were used to direct transcription-translation in an *E. coli* in vitro system. The resulting [$^{35}$S]methionine-labeled polypeptides were examined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and fluorography (Fig. 2). Our assignments of the sizes and the identities of the polypeptides produced by each plasmid are shown in Table 2. The plasmids pXF229, pXF225, pXF224, pXF223, and pXF231 were all derived from pXF180 and carry *com* DNA segments of increasing length that extend from the same (left) endpoint. Examination of the polypeptides produced by these plasmids established the positions of proteins on the *com* map as follows.

Starting from the left boundary of the *com* locus (Fig. 1), the small (715-bp) fragment of *com* DNA contained in pXF229 expressed strong promoter activity but directed synthesis of no *com*-specific polypeptides in addition to the vector products. Comparison of pXF225, pXF224, pXF223, and pXF231 showed that these produced polypeptides of increasing length, 6, 29, 43, and 77 kDa, respectively (Fig. 2). Within experimental error, these are consistent with three successive truncations of a single 77-kDa protein. A second protein, 41 kDa, was also made by pXF231, as well as one of 15 kDa. Comparison of the products of pXF231 and pXF180 showed the 41-kDa protein to be truncated in pXF231, as it was replaced by a 49-kDa product in pXF180. In addition, pXF180 made 15 and 77-kDa proteins identical in size to those made by pXF231. pXF190, which overlaps the right end of pXF180 (Fig. 1), also made a 49-kDa protein. (In pXF232, a *Kpn*I deletion derivative of pXF190, the 49-kDa protein is replaced by one that is 41 kDa [data not shown]). pXF190 may also make a 15-kDa protein; however, this region of the gel was obscured by background peptides. Further studies will be required to verify unequivocally the presence of the 15-kDa band from pXF190. In an attempt to establish the location of the 15-kDa protein on the *com* DNA, a subclone from near the right end of pXF180 was made and designated pXF233 (Fig. 1). pXF233 made a polypeptide of 24 kDa, presumably the expected truncated 49-kDa protein, but no 15-kDa protein (Fig. 3).

TABLE 2. Assignment of polypeptides produced during in vitro transcription-translation of plasmids containing *com* DNA

| Plasmid | Molecular size[a] (kDa) of polypeptides assigned to each gene | | | | | | |
|---|---|---|---|---|---|---|---|
| | *unk* | *comA* | *comB* | *com*?[b] | *bla* | *cat* | *malX* |
| pKK232-8 | | | | | 30.5 | | |
| pPS12 | | | | | 30.5 | 25 | 10 |
| pXF229 | | | | | 30.5 | 25 | |
| pXF225 | | 6 | | | 30.5 | 25 | |
| pXF224 | | 29.5 | | | 30.5 | 25 | |
| pXF223 | | 43 | | | 30.5 | 25 | |
| pXF170 | 16 | 43 | | | 30.5 | 25 | |
| pXF231 | | 77 | 41 | 15 | 30.5 | 25 | |
| pXF180 | | 77 | 49 | 15 | 30.5 | 25 | |
| pXF190 | | | 49 | 15 | 30.5 | 25 | |
| pXF233 | | | 24 | | 30.5 | 25 | |

[a] Apparent molecular sizes in polyacrylamide gradient-sodium dodecyl sulfate gels.

[b] A possible additional *com* gene product, observed with certain *com* plasmids.

Assuming that 270 bp code for 10 kDa of protein and that genes integrated at the multiple cloning site make a protein terminated within 20 amino acids of that site, the proteins encoded in each of these plasmids were assigned map locations as shown in Fig. 1. This established the location of genes for the 77- and 49-kDa proteins on the *com* DNA. The presence of the truncated proteins also allowed us to assign the directions of transcription of these two proteins. Two possible locations of the 15-kDa protein could also be inferred (vide infra). We designated the gene for the 77-kDa protein *comA* and that for the 49-kDa protein *comB*. The map locations of *comA* and *comB* deduced from the truncation studies were further tested by analysis of proteins that were disrupted by insertions in the *com* clones, as described below.

**Insertional disruption of *comA*.** pXF206, pXF208, pXF209, and pXF219 were constructed by inserting an *ermB*-containing fragment at various restriction sites within the *com* DNA (5) in pXF180 and pXF190. Analysis of the proteins produced from pXF206 and pXF208 showed that the insertions in these plasmids resulted in the loss of the 77-kDa protein but did not affect the expression of the 49- or 15-kDa protein (Table 3 and Fig. 4). As these two inserts span the region from kb 1.6 to 3.34 on the *com* map, this places stringent limits on the location of the *comA* gene (Fig. 1) that are consistent with our interpretation of the truncation data.

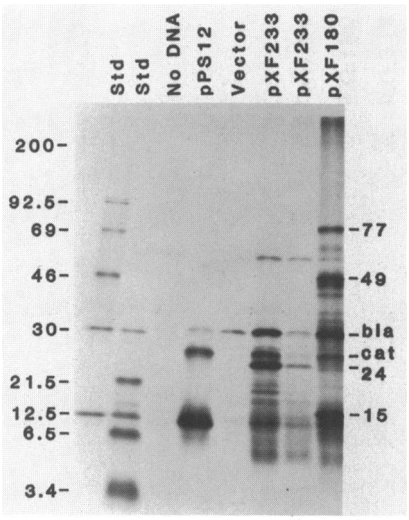


FIG. 3. In vitro transcription-translation analysis of pXF233. The labeled polypeptides were separated as in Fig. 2. The plasmid studied (or protein size standard [Std]) is indicated at the top. Vector gene products and *com* protein sizes (in kilodaltons) are indicated on the right. Sizes (in kilodaltons) and positions of protein standards are shown on the left.

**Insertional disruption of *comB*.** Analysis of the proteins produced from pXF209 and pXF219 showed that the inserts in these plasmids resulted in the loss of the 49-kDa protein (Table 3 and Fig. 4). The insert in pXF209 did not affect the expression of the 77- or 15-kDa protein (Fig. 4). (Since the insertion of pXF219 was made in pXF190 [Fig. 1], the effect of this insert on the expression of the 15-kDa protein could not be determined.) Thus, the effects of *ermB* insertions in pXF209 and pXF219 at kb 3.66 and 4.65 confirmed the position of *comB* assigned from the truncation results.

In our initial description of the *com* locus (5), we reported testing several *com* insertion mutants for response to CF. All those tested responded with high levels of competence. As the present mapping shows that all the inserts tested previously, including the original insertion-duplication mutation at this locus (*comA15*), lie within *comA*, we chose two insertions in *comB* to test for response to CF, as before. Both of the *comB* mutants (*ermB* inserts at kb 4.65 and 3.66) also responded to CF by developing normal competence (data not shown). Thus, disruption of *comB*, as well as of *comA*, caused a deficiency in competence induction that could be restored by CF.

The combined information from subclones, truncation clones, and insertion mutations shows that *comA* and *comB* map as contiguous genes, with *comA* extending from approximately kb 1.25 to 3.35. *comB* does not appear to overlap with *comA*, since the insertion which maps very near the end of *comA* does not disrupt the 49-kDa protein. Therefore, *comB* extends from approximately kb 3.40 nearly to the end of pXF180 at kb 4.665. Since the *ermB* inserts studied here disrupted either the 77- or the 49-kDa protein and since all the inserts have a Com$^-$ phenotype, we conclude that the *com* locus consists of at least two genes. Truncation products showed that *comA* and *comB* are transcribed from left to right as drawn in Fig. 1. In addition, since none of the insertions or deletions in pXF180 affected the level of expression of the undisrupted *com* proteins, there is no polar effect of these mutations.

TABLE 3. Insertional disruption of *com* proteins

| Plasmid | Location of insertion (bp)[a] | Protein disrupted | | |
|---|---|---|---|---|
| | | Name | Original size (kDa)[b] | Size(s) (kDa) of new product(s)[c] |
| pXF208 | 1600 | ComA | 77 | 6 |
| pXF206 | 3335 | ComA | 77 | 85 |
| pXF219 | 3660 | ComB | 49 | 45, 39 |
| pXF209 | 4645 | ComB | 49 | 60 |

[a] Map position of *ermB* insert made in vitro.
[b] Protein made by intact plasmid and disrupted by insertion mutation (Fig. 4).
[c] Protein(s) appearing in insertion mutant.

**Another *com* gene?** It appears that the region encoding the 15-kDa protein must overlap either *comA* or *comB*. Since none of the truncation plasmids or subclones made this protein and none of the insertions or deletions in pXF180 disrupted it, it would appear that the two possible locations for this third, putative *com* gene are at the carboxy-terminal end of *comA* or the middle of *comB*. Experiments are in progress to define the location of this gene.

**Neighboring genes and sequences.** pXF170 contains the same DNA as pXF223, with an additional 500 bp extending to the left and outside of the *com* locus (Fig. 1). Comparison of the polypeptides produced by these plasmids showed that in addition to the 43-kDa protein made by pXF223 (truncated *comA*), pXF170 produced a 16-kDa polypeptide (Fig. 5). Thus, at least part of the coding region for this polypeptide, or its regulatory sequences, must lie between the *Hin*dIII and *Dra*I sites (bp 0 and 510; Fig. 6). Since we do not have any clones extending farther to the left, we cannot deduce the size or orientation of this gene. However, as insertions in this region do not give the Com$^-$ phenotype, it is not an additional *com* gene. This protein was previously interpreted as a possible *com* protein (5).

Southern hybridization experiments were previously used to verify the structure of some of our clones (5). These experiments revealed a region of DNA between the *Dra*I and *Hpa*II sites at bp 510 and kb 1.6 that hybridized to 6 to 10 different fragments of the CP1200 chromosome (MHS in Fig. 6). Several different digests of CP1200 DNA were studied by using restriction nucleases that do not recognize sites within the region used as a probe (data not shown). The hybridizing sequences were always found distributed on 6 to 10 fragments of the chromosome. When the adjacent *Hpa*II fragment (kb 1.6 to 3.25) was used as a probe under identical conditions, it always hybridized to a single band only. This multiple hybridizing sequence could indicate the presence of an insertion sequence, which would be the first described in *S. pneumoniae*. This region could also represent multiple copies of closely related genes, e.g., rRNA genes; the fact that proteins may be transcribed within 1 kb on each side of this region argues against rRNA genes or large tRNA clusters, however. Finally, the multiple hybridizing sequence could constitute a regulatory element shared by several coordinately controlled competence-specific operons.

Studying the polypeptide products of the region near the right end of the *com* locus presented some difficulties. Clones containing this region had strong promoter activity as detected by the large amount of CAT protein made by these clones (data not shown). One of these clones (pXF230) was unstable, and for those that were stable (pXF234 and pXF235), we did not detect any insert-specific polypeptide bands in vitro. Additional studies will be required to determine if there are any other *com* genes in this region.

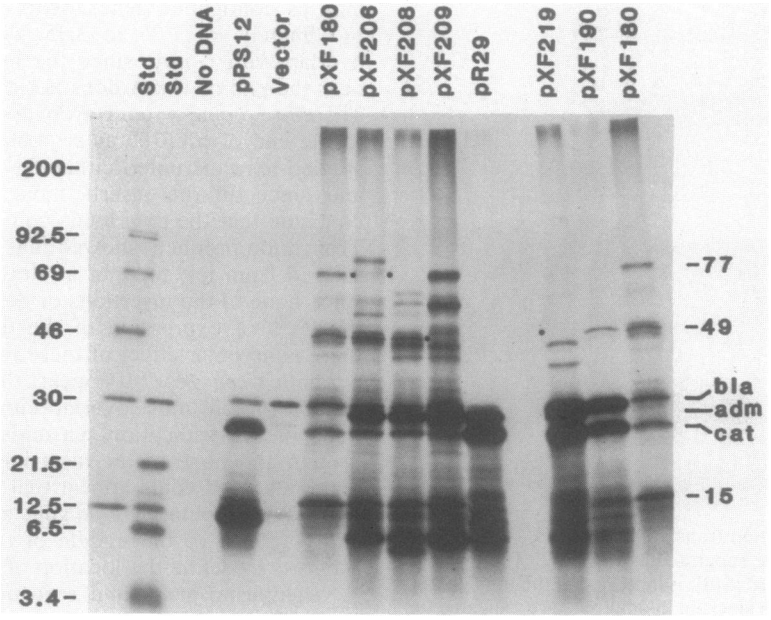


FIG. 4. In vitro transcription-translation analysis of plasmids containing insertions in subfragments of the *com* locus. The labeled polypeptides were separated as described in the legend to Fig. 2. The plasmid studied (or protein size standard [Std]) is indicated at the top. Vector gene products and *com* protein sizes (in kilodaltons) are indicated on the right. Sizes (in kilodaltons) and positions of protein standards are shown on the left. adm, adenine methylase of *ermB*. Dots indicate wild-type forms of *com* proteins disrupted by insertion.

## DISCUSSION

Identification of the protein products originating from a series of subclones, insertions, and deletions of the *com* locus allowed us to localize two nonoverlapping genes, *comA* and *comB*, coding for A (77-kDa) and B (49-kDa) proteins, respectively (Fig. 6). As it was possible to separate *comB* from *comA* on pXF233 and pXF190, and as insertions in *comA* did not affect the amount of comB and insertions in *comB* did not affect comA, we conclude that *comA* and *comB* are transcribed separately, at least in the heterologous in vitro transcription-translation system. The *com* locus may contain a third gene, coding for a 15-kDa protein.

The role of *comA* and *comB* (and a possible third *com* gene) in the control of competence is not known. Competence is initiated by the extracellular protein CF (19, 22–25). CF has at least two effects on cells. First, it stimulates the production of more CF. Second, it signals the cells to enter the competent state in preparation for transformation. If these were separate interactions, then distinct proteins could interact with CF in each process. There are also reports of an inhibitor of CF—whose production may peak at about the same time as CF (24)—which acts to bring competence to an end. *comA* or *comB* might regulate expression of the inhib-

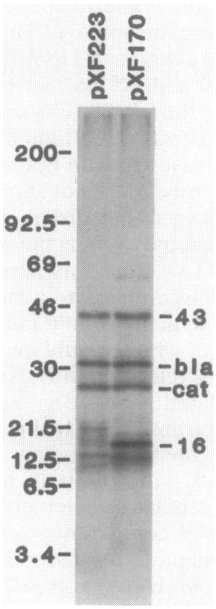


FIG. 5. In vitro transcription-translation analysis of plasmids containing segments at the left end of the *com* locus. The labeled polypeptides were separated as described in the legend to Fig. 2. The plasmid studied is indicated at the top. Vector gene products and *com* protein sizes (in kilodaltons) are indicated on the right. Sizes (in kilodaltons) and positions of protein standards are shown on the left.

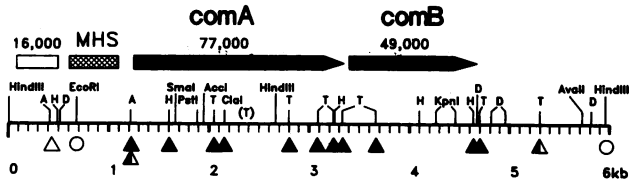


FIG. 6. *com* region of *S. pneumoniae*. The *comA* and *comB* genes, a third, unidentified coding sequence, and their protein products are schematically illustrated. The multiple hybridizing sequence (MHS) lies between *comA* and the unidentified gene. The locations and phenotypes of *ermB* insertion mutations created in vitro and characterized after transformation into the pneumococcus chromosome (5) are indicated as follows: △, $Xfo^+$; ▲, $Xfo^-$; ◭, $Xfo^{+/-}$. Also shown (○) are the outer limits for the *com* locus determined by insertion duplication mutagenesis (5). Restriction enzyme abbreviations: A, *Asu*II; D, *Dra*I; H, *Hpa*II; T, *Taq*I.

itor, since overexpression of this protein could prevent competence. Thus, there appears to be a variety of functions that these competence control proteins could serve.

Since *comA* mutants do not release CF and can be complemented by exogenous CF (5), *comA* could play a role in regulating CF production or as CF itself. The relation between the *com* locus and CF is unknown. One of the *com* proteins may be CF or a precursor of it, or this may be solely a regulatory locus.

The *Kpn*I-*Ava*II fragment (kb 4.47 to 5.72) was cloned in both orientations to identify *com* proteins at the right end of the *com* locus. This fragment spans a region downstream of *comB* which contains two sites (bp 4705 and 5300) in which insertions cause mutations affecting competence (Fig. 6). No competence-specific proteins were detected from these clones. This region could be important for the expression of the upstream *com* genes, or it may contain another *com* gene whose peptide product was not detected in our in vitro system.

All of the subclones and deletions from the *com* region described here expressed promoter activity in *E. coli*. pXF229 and pPS12 showed especially strong promoter activity (Fig. 2) in the *E. coli* in vitro system. Some subclones of the *com* locus (4) and the *mal* fragment (12) contained in pPS12 (6) have been shown to be unstable in ordinary *E. coli* vectors. These fragments could be cloned, however, in vectors incorporating strong transcriptional terminators. The results from in vitro transcription-translation support our earlier conclusions (4, 6) that these fragments of pneumococcal DNA contain sequences which act as strong promoters in *E. coli*. In addition, there are regions, especially in large fragments from the right end of the *com* locus, which are unstable even in vectors protected by transcriptional terminators. It is not clear if this is due to promoter activity, to a specific gene product, or to other structural features.

## ACKNOWLEDGMENT

This work was supported in part by Public Health Service research grant AI19875 from the National Institutes of Health.

## LITERATURE CITED

1. **Bonner, W. M., and R. A. Laskey.** 1974. A film detection method for tritium-labelled proteins and nucleic acids in polyacrylamide gels. Eur. J. Biochem. **46**:83–88.
2. **Brosius, J.** 1984. Plasmid vectors for the selection of promoters. Gene **27**:151–160.
3. **Brosius, J.** 1984. Toxicity of an overproduced foreign gene product in *Escherichia coli* and its use in plasmid vectors for the selection of transcription terminators. Gene **27**:161–172.
4. **Chandler, M. S., and D. A. Morrison.** 1987. Molecular cloning of a competence control region from *Streptococcus pneumoniae* by use of transcription terminator vectors in *Escherichia coli*, p. 193–196. *In* R. Curtiss III and J. J. Ferretti (ed.), Streptococcal genetics. American Society for Microbiology, Washington, D.C.
5. **Chandler, M. S., and D. A. Morrison.** 1987. Competence for genetic transformation in *Streptococcus pneumoniae*: molecular cloning of *com*, a competence control locus. J. Bacteriol. **169**: 2005–2011.
6. **Chen, J.-D., and D. A. Morrison.** 1987. Cloning of *Streptococcus pneumoniae* DNA in *Escherichia coli* requires vectors protected by strong transcriptional terminators. Gene **55**:179–187.
7. **Claverys, J. P., H. Prats, H. Vasseghi, and M. Gherardi.** 1984. Identification of *Streptococcus pneumoniae* mismatch repair genes by an additive transformation approach. Mol. Gen. Genet. **196**:91–96.
8. **Fox, M. S., and M. K. Allen.** 1964. On the mechanism of deoxyribonucleate integration in pneumococcal transformation. Proc. Natl. Acad. Sci. USA **52**:412–419.
9. **Gaustad, P.** 1979. Genetic transformation in *Streptococcus sanguis*. Distribution of competence and competence factors in a collection of strains. Acta Pathol. Microbiol. Scand. Sect. B **87**:123–128.
10. **Gaustad, P., J. Eriksen, and S. D. Henriksen.** 1979. Genetic transformation in *Streptococcus sanguis*. Spontaneous and induced competence of selected strains. Acta Pathol. Microbiol. Scand. Sect. B **87**:117–122.
11. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
12. **Lacks, S. A., J. J. Dunn, and B. Greenberg.** 1982. Identification of base mismatches recognized by the heteroduplex-DNA-repair system of *Streptococcus pneumoniae*. Cell **31**:327–336.
13. **Lacks, S. A., B. Greenberg, and K. Carlson.** 1967. Fate of donor DNA in pneumococcal transformation. J. Mol. Biol. **29**:327–347.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
15. **Leonard, C. G., and R. M. Cole.** 1972. Purification and properties of streptococcal competence factor isolated from chemically defined medium. J. Bacteriol. **110**:273–280.
16. **Leonard, C. G., J. M. Ranhand, and R. M. Cole.** 1970. Competence factor production in chemically defined media by noncompetent cells of group H *Streptococcus* strain Challis. J. Bacteriol. **104**:674–683.
17. **Morrison, D. A.** 1981. Competence-specific protein synthesis in *Streptococcus pneumoniae*, p. 39–53. *In* M. Polsinelli and G. Mazza (ed.), Transformation—1980. Proceedings of the 5th European Meeting on Bacterial Transformation and Transfection. Cotswold Press, Oxford.
18. **Morrison, D. A., and M. Baker.** 1979. Competence for genetic transformation in pneumococcus depends on synthesis of a small set of proteins. Nature (London) **282**:215–217.
19. **Morrison, D. A., M-C. Trombe, M. K. Hayden, G. A. Waszak, and J.-D. Chen.** 1984. Isolation of transformation-deficient *Streptococcus pneumoniae* mutants defective in control of competence, using insertion-duplication mutagenesis with the erythromycin resistance determinant of pAMβ1. J. Bacteriol. **159**:870–876.
20. **Pakula, R., and W. Walczak.** 1963. On the nature of competence of transformable streptococci. J. Gen. Microbiol. **31**:125–133.
21. **Raina, J. L., and A. W. Ravin.** 1980. Switches in macromolecular synthesis during induction of competence for transformation of *Streptococcus sanguis*. Proc. Natl. Acad. Sci. USA **77**: 6062–6066.
22. **Tomasz, A.** 1966. Model for the mechanism controlling the expression of competent state in pneumococcus cultures. J. Bacteriol. **91**:1050–1061.
23. **Tomasz, A.** 1973. Cell surface structures and the absorption of DNA molecules during genetic transformation in bacteria, p. 321–355. *In* L. Leive (ed.), Bacterial membranes and walls. Marcel Dekker, Inc., New York.
24. **Tomasz, A., and R. Hotchkiss.** 1964. Regulation of the transformability of pneumococcal cultures by macromolecular cell products. Proc. Natl. Acad. Sci. USA **51**:480–487.
25. **Tomasz, A., and J. L. Mosser.** 1966. On the nature of the pneumococcal activator substance. Proc. Natl. Acad. Sci. USA **55**:58–66.