JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 372–381
0021-9193/91/010372-10$02.00/0

Vol. 173, No. 1

Genetic Transformation in *Streptococcus pneumoniae*: Nucleotide Sequence Analysis Shows *comA*, a Gene Required for Competence Induction, To Be a Member of the Bacterial ATP-Dependent Transport Protein Family

FRANCIS M. HUI AND DONALD A. MORRISON*

Laboratory for Molecular Biology, Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60680

Received 11 June 1990/Accepted 13 October 1990

The complete nucleotide sequence of *comA*, a gene required for induction of competence for genetic transformation in *Streptococcus pneumoniae*, was determined by using plasmid DNA templates and synthetic oligonucleotide primers. The sequence contained a single large open reading frame, ORF1, of 2,151 bp. ORF1 was included within the *comAB* locus previously mapped genetically and accounted for 50% of its extent. The predicted molecular weight of the largest polypeptide encoded within ORF1, 80,290, coincided with that measured previously (77,000) for the product of in vitro transcription-translation of the cloned *comA* locus. A Shine-Dalgarno sequence (AAAGGAG, $\Delta G = -14$ kcal) lay immediately upstream of ORF1. A sequence (TTtAat-17 bp-TAaAAT) similar to the *Escherichia coli* σ^{70} promoter consensus was located 410 bp upstream of ORF1. The deduced protein sequence of ComA showed a very strong similarity to the *E. coli* hemolysin secretion protein, HlyB, and strong similarities to other members of the family of ATP-dependent transport proteins, including the mammalian multidrug resistance P-glycoprotein. These similarities suggest that ComA functions in the transport of some molecule, possibly pneumococcal competence factor itself.

Genetic transformation in *Streptococcus pneumoniae* (pneumococcus) depends on an elaborate DNA-processing pathway characteristic of a specialized cell state termed competence. Entry into this state is accompanied by a switch of protein synthetic activity to production of a small number of competence-specific proteins (39, 40), as well as a gross metabolic shift (37). In laboratory batch cultures, competence typically appears suddenly in most or all cells of the culture at some point during exponential-phase growth, persists for 10 to 20 min, and then decays rapidly (55–58). The coordination of competence induction among the cells of a culture depends on an extracellular protein, termed competence factor (CF), which serves as a monitor of population density (55, 57, 58).

We previously reported the cloning of a locus, including genes *comA* and *comB*, involved in this regulatory circuit (5). Com$^-$ mutants were defective in competence induction and CF production but could be complemented by competent culture supernatants, implying a deficiency in elaboration of CF (41). We report here the nucleotide sequence of *comA* and the deduced sequence of the ComA protein. Homology analysis showed that ComA belonged to a family of ATP-binding proteins (21, 23) including membrane components of several bacterial transport systems (2) and eucaryotic transport proteins related to the mammalian multidrug resistance determinants (7).

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture media. Strains and plasmids used in this work were described previously (6). The plasmids used for sequencing contained portions of the *comAB* locus, as shown on the map in Fig. 1, and were generously provided by Mark S. Chandler. *Escherichia coli* cells were grown in L broth (35) or an enriched broth (53). Preparative-scale plasmid isolation followed an alkaline-lysis protocol including two cycles of CsCl gradient purification (3).

DNA sequence determination. Sequencing reactions were done with Sequenase 2.0, dGTP or dITP labeling mixes, [α-^{35}S]dATP, and circular plasmid templates. The conditions were described previously (48), except that labeling reactions were incubated for 10 min at 0°C and termination reactions were done at 37°C for 5 min. The synthetic primers used were purchased from Operon Technologies (San Pablo, Calif.) or made at the Laboratory for Molecular Biology in Chicago.

Computer analysis of sequence data. The deduced amino acid sequence of ComA was compared by the algorithm of Pearson and Lipman (47) (FASTA implementation at GenBank) with proteins in the SWISS-PROT database (release 14.0). Full-length alignment with HlyB was done by the method of Myers and Miller (44). Potential transmembrane segments were evaluated by the method of Klein et al. (28).

Nucleotide sequence accession number. The DNA sequence reported here is held in GenBank under accession number M31680.

RESULTS

DNA sequence of *comA*. The nucleotide sequence of 3,000 bp within the *comAB* locus (Fig. 1) was determined (Fig. 2). The sequencing strategy outlined in Fig. 1 focused on a region predicted previously, on the basis of mutational and expression data, to contain a gene for a 77-kDa protein oriented as shown. The observed sequence verified the presence of restriction sites previously mapped physically (6) and revealed the presence of several additional *Taq*I sites.

* Corresponding author.

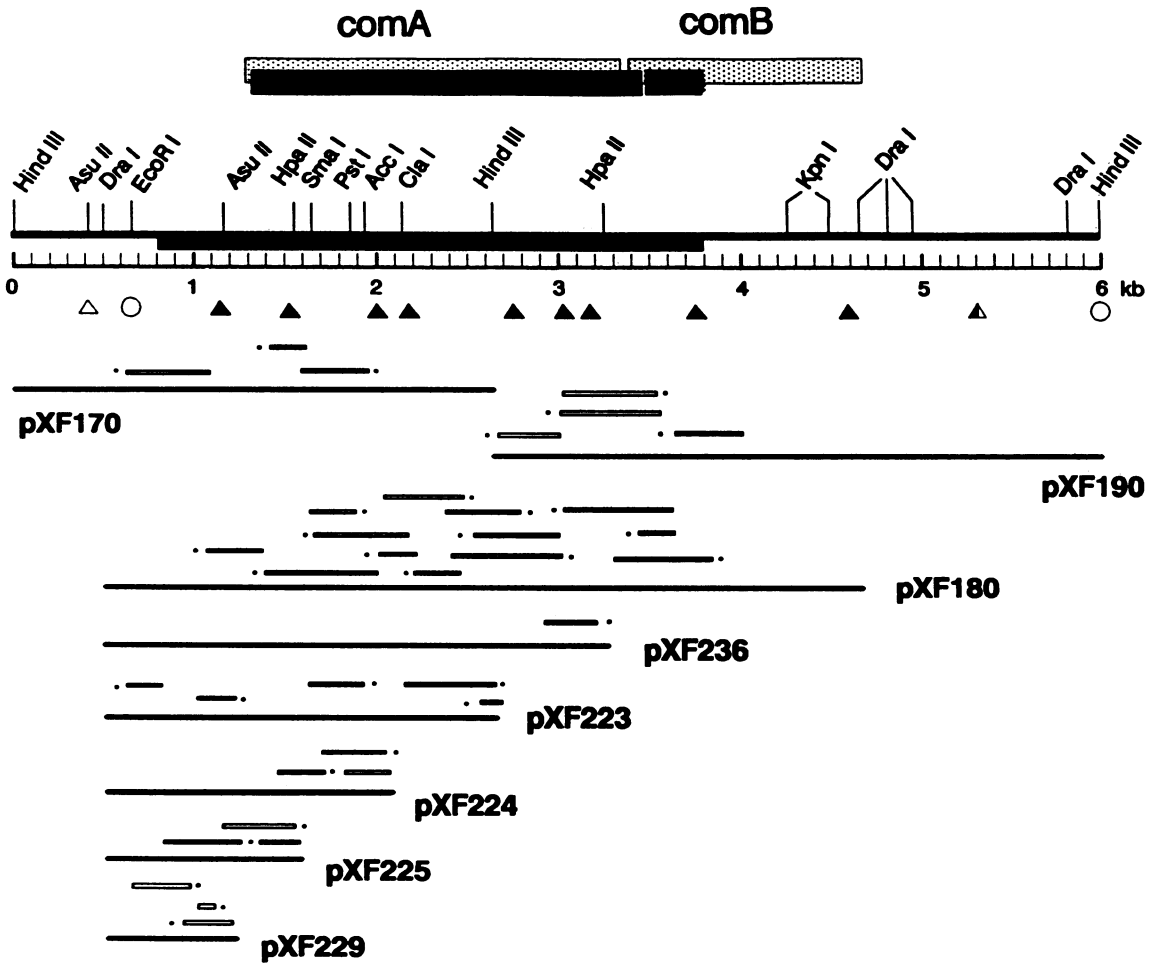

FIG. 1. Map of the *S. pneumoniae comA* locus and sequencing strategy. Stippled boxes show genes *comA* and *comB* as deduced from gene truncation studies (6). Solid boxes show genes deduced from DNA sequence. The thick bar indicates the portion of the locus sequenced in both strands for identification of ComA. Below the map, portions of the locus subcloned in the vector pKK232-8 which served as templates for sequencing reactions are indicated by thick lines. The extent of sequence determined for each primer (●) is indicated by rectangles. Triangles indicate the positions of insertion mutations, and circles indicate the outer ends of insertion-duplication mutations. Mutant phenotypes Com^- and Com^+ are indicated by closed and open symbols, respectively.

Prediction of ComA protein sequence. Examination of the sequenced region for open reading frames (ORFs) found only two larger than 200 bp. ORF1 (717 amino acids) and ORF2 (110 amino acids, extending beyond the end of the sequence shown) were separated by 15 bp. Examination of the region near the start of ORF1 for sites with potential for binding to the 3′ end of gram-positive 16S rRNA (3′-UCUUUCCUCCACUA-5′) (43, 50) showed that within ORF1, only two of the first 15 potential initiation codons (ATG, TTG, or GTG) were associated with a potential ribosome-binding site. This site (AAAGGAG) ($\Delta G = -14$ kcal [54]) at bp 501 was followed by ATG codons at bp 508 and 514, respectively. Translation starting at the second methionine (6 bp from the putative ribosome-binding site) would yield a protein of 717 amino acids and 80,290 molecular weight, which we designate as the putative ComA protein (Fig. 2). The putative gene was flanked upstream by a region of 500 bp, with no ORF larger than 138 bp, and downstream by the start of ORF2, separated from ORF1 by 15 bp. ORF2 was also preceded by a potential ribosome-binding site 5 bp from the ATG codon. Potential σ^{70}-type promoters (42) were discernible upstream at bp 13 to 40 and 75 to 113. The latter was identified by a matrix scan program similar to that of Harr et al. (19) as the site within the sequence reported here that was most similar to the canonical bacterial promoter consensus, with a 17-bp hexamer spacing and 8 of 12 bases identical to the consensus TTgAca-17 bp-TAtAAT.

Structural analysis of ComA. The predicted ComA protein consisted of two dissimilar portions (Fig. 3). The N-terminal half was largely hydrophobic in character, while the remainder was mainly hydrophilic. The algorithm of Klein et al. (28) identified six potential membrane-spanning segments, clustered in the N-terminal half of the sequence. The hydrophilic C-terminal region displayed two readily recognized motifs. One was the ATP-binding site motif A [(G, A)-x4-G-K-(S, T)] (59) represented by GISGSGKT at positions 517 to 524. The second, represented by ISGGQRQRI at positions 622 to 630, was a motif unique to proteins of the active transport family [(L, F)-S-G-G-x2-(Q, R, K)-(Q, R, K)-(I, L, V, M, A)] (8, 21, 23).

Similarity of ComA to transport protein family. Searches of

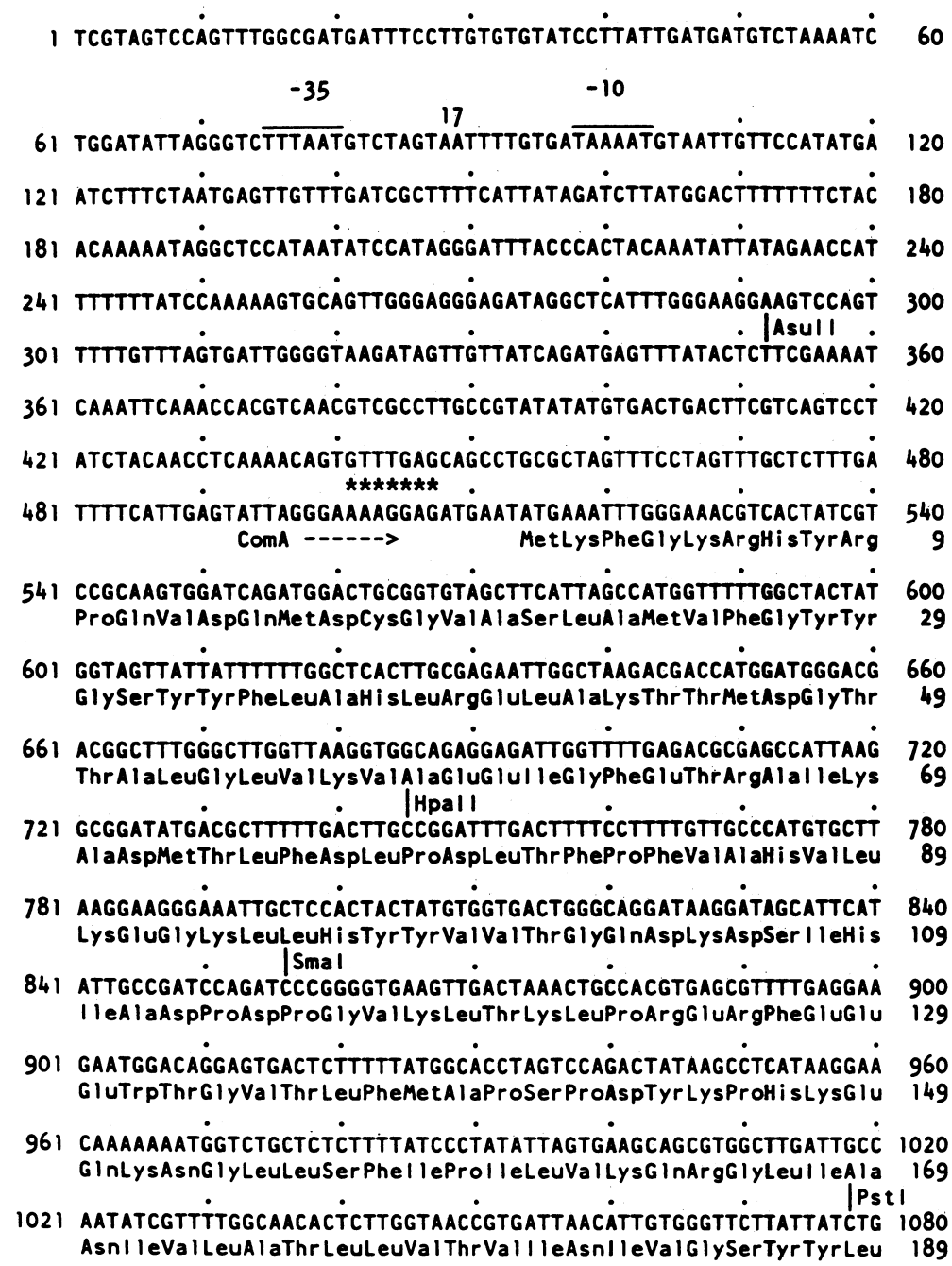

FIG. 2. Sequence of the sense strand and predicted amino acid sequence of the *S. pneumoniae comA* gene and protein. Putative promoter (−35, −10) and ribosome-binding (****) sites are indicated in the 500 bp preceding the start of the ComA ORF. A possible ribosome-binding site and second ORF following *comA* are also indicated. Landmark restriction sites are shown.

the protein sequence databases, using the entire predicted ComA sequence, revealed strong similarities between ComA and several other proteins (see, for example, Table 1). The strongest similarities involved the C-terminal portion of ComA surrounding the active transport motifs described above. The most highly similar proteins were members of a family of proteins consisting principally of several groups of ATP-binding membrane proteins, including a group of bacterial ATP-dependent transport proteins (21), a group of eucaryotic proteins conferring multidrug resistance (7, 17, 38), and the related transport proteins PMP-70 (27), CFTR (49), STE6 (32), the *Drosophila* brown and white proteins (9), and the CQR protein (12).

The highest scores in such searches were for comparisons with HlyB, an *E. coli* membrane protein required for export of the protein hemolysin A, and closely related toxin secretion proteins. The next highest scores were for matches to a set of eucaryotic membrane proteins closely related to the multidrug resistance genes. The remaining high scores were for matches to other bacterial transport system proteins in

```
1081 CAGTCTATCATTGATACCTATGTGCCAGATCAGATGCGTTCGACACTAGGGATTATTTCT 1140
     GlnSerIleIleAspThrTyrValProAspGlnMetArgSerThrLeuGlyIleIleSer  209

1141 ATTGGGCTAGTCATCGTCTACATCCTCCAGCAAATCTTGTCTTACGCTCAGGAGTATCTC 1200
     IleGlyLeuValIleValTyrIleLeuGlnGlnIleLeuSerTyrAlaGlnGluTyrLeu  229

1201 TTGCTTGTTTTGGGGCAACCGTTGTCGATTGACGTGATTTTGTCCTATATCAAGCATGTT 1260
     LeuLeuValLeuGlyGlnProLeuSerIleAspValIleLeuSerTyrIleLysHisVal  249

1261 TTTCACCTCCCTATGTCCTTCTTTGCGACACGCAGGACAGGGGAGATCGTGTCTCGTTTT 1320
     PheHisLeuProMetSerPhePheAlaThrArgArgThrGlyGluIleValSerArgPhe  269

1321 ACAGATGCTAACAGTATCATCGATGCGCTGGCTTCGACCATCCTTTCGATTTTCCTAGAT 1380
     ThrAspAlaAsnSerIleIleAspAlaLeuAlaSerThrIleLeuSerIlePheLeuAsp  289

1381 GTGTCAACGGTTGTCATTATTTCCCTTGTTCTATTTTCACAAAATACCAATCTCTTTTTC 1440
     ValSerThrValValIleIleSerLeuValLeuPheSerGlnAsnThrAsnLeuPhePhe  309

1441 ATGACTTTATTGGCGCTTCCTATCTACACAGTGATTATCTTTGCCTTTATGAAGCCGTTT 1500
     MetThrLeuLeuAlaLeuProIleTyrThrValIleIlePheAlaPheMetLysProPhe  329

1501 GAAAAGATGAATCGGGATACCATGGAAGCCAATGCGGTTCTGTCTTCTTCTATCATTGAG 1560
     GluLysMetAsnArgAspThrMetGluAlaAsnAlaValLeuSerSerSerIleIleGlu  349

1561 GACATCAACGGTATTGAGACTATCAAGTCCTTGACCAGTGAAAGTCAGCGTTACCAAAAA 1620
     AspIleAsnGlyIleGluThrIleLysSerLeuThrSerGluSerGlnArgTyrGlnLys  369

1621 ATTGACAAGGAATTTGTGGATTATCTGAAGAAATCCTTTACCTATAGTCGAGCAGAGAGT 1680
     IleAspLysGluPheValAspTyrLeuLysLysSerPheThrTyrSerArgAlaGluSer  389

1681 CAGCAAAAGGCTCTGAAAAAGGTTGCCCATCTCTTGCTTAATGTCGGCATTCTCTGGATG 1740
     GlnGlnLysAlaLeuLysLysValAlaHisLeuLeuLeuAsnValGlyIleLeuTrpMet  409

1741 GGGGCTGTTCTGGTCATGGATGGCAAGATGAGTTTGGGGCAGTTGATTACCTATAATACC 1800
     GlyAlaValLeuValMetAspGlyLysMetSerLeuGlyGlnLeuIleThrTyrAsnThr  429

1801 TTGCTGGTTTACTTTACCAATCCTTTGGAAAATATCATCAATCTGCAAACCAAGCTTCAG 1860
     LeuLeuValTyrPheThrAsnProLeuGluAsnIleIleAsnLeuGlnThrLysLeuGln  449

1861 ACAGCGCAGGTTGCCAATAACCGTCTAAATGAAGTGTATCTAGTAGCTTCTGAGTTTGAG 1920
     ThrAlaGlnValAlaAsnAsnArgLeuAsnGluValTyrLeuValAlaSerGluPheGlu  469

1921 GAGAAGAAAACAGTTGAGGATTTGAGCTTGATGAAGGGAGATATGACCTTCAAGCAGGTT 1980
     GluLysLysThrValGluAspLeuSerLeuMetLysGlyAspMetThrPheLysGlnVal  489

1981 CATTACAAGTATGGCTATGGTCGAGACGTCTTGTCGGATATCAATTTAACCGTTCCCCAA 2040
     HisTyrLysTyrGlyTyrGlyArgAspValLeuSerAspIleAsnLeuThrValProGln  509
```

FIG. 2—*Continued.*

the ATP-dependent transport protein family. An alignment of ComA and HlyB, shown in Fig. 3, revealed a strong similarity extending beyond the region of highest similarity recognized by the searching programs to include about 310 amino acids, virtually the entire C-terminal half of the proteins. The similarity was especially strong in those portions of this region generally conserved among members of the family. Of the 99 residues in HlyB identified as conserved in this family by Higgins et al. (21), 82% were conserved in ComA. The N-terminal half of ComA was also much more similar to the corresponding portion of HlyB than to any other proteins in the database (data not shown). Over the whole full-length alignment, 27% of the amino acids were identical and a total of 48% were identical or conservative replacements. In the 310-amino-acid region of high similarity, there was 37% identity and 23% more conservative replacements.

DISCUSSION

The specific role of ComA in competence regulation is unknown. The similarity of ComA to other members of the ATP-binding transport protein family (18) reported here strongly suggests, however, that its function includes a

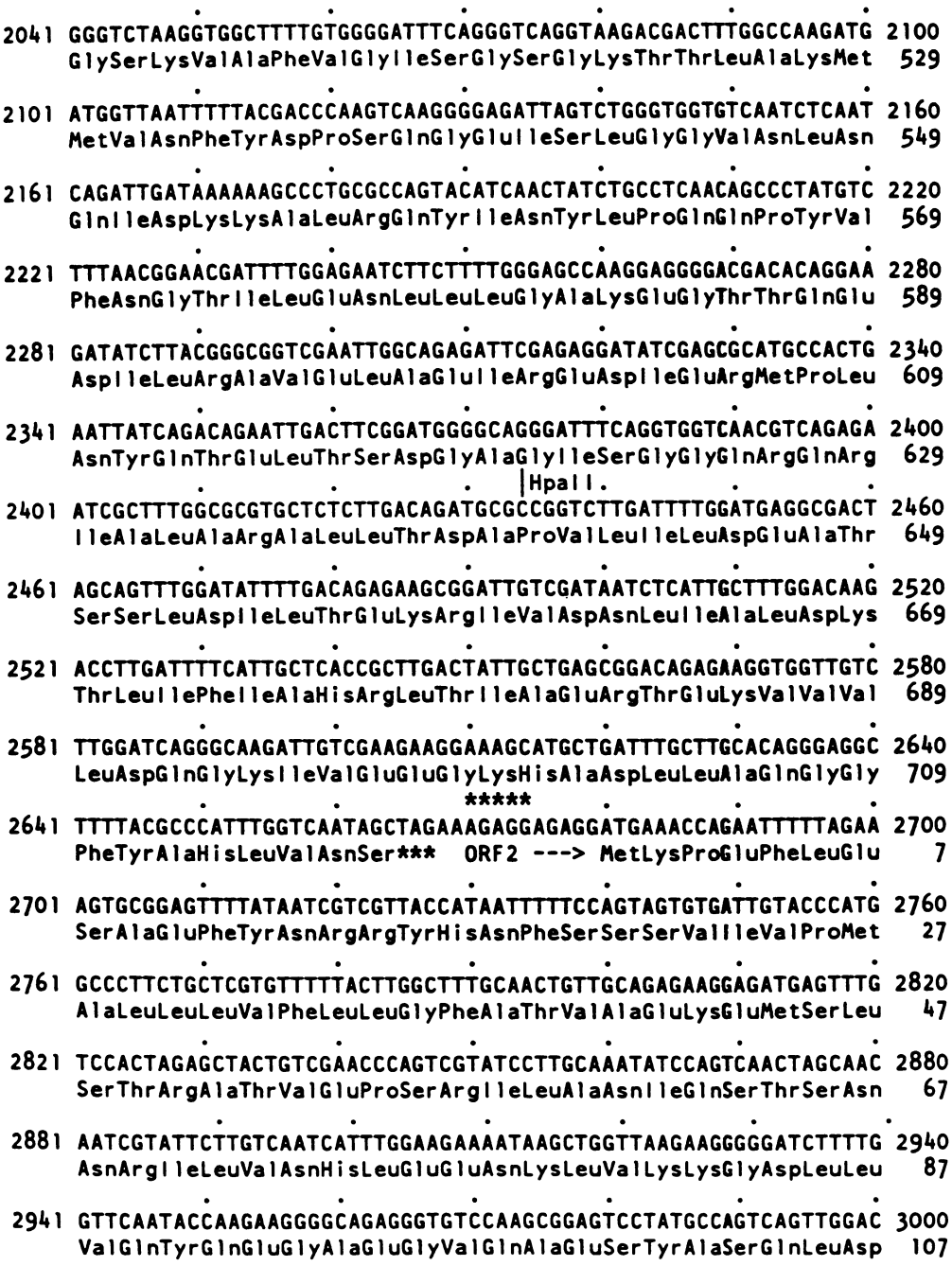

FIG. 2—*Continued.*

specific transport activity. While there is no direct evidence bearing on the nature of this possible transport activity, the observation that the strongest and most extensive similarity of ComA is with HlyB and related toxin secretion proteins may indicate that the transported molecule is a protein.

CF has been characterized as a small basic extracellular protein (55, 58), but little else is known about its biochemistry, regulation, or mode of release from the cell. The strong and extensive similarity between ComA and HlyB described here, when coupled with the deficiency in CF elaboration observed for *comA* mutants, immediately suggests a model for CF synthesis and export. In this model, CF is a specifically exported protein, analogous to HlyA, while ComA is an ATP-binding membrane component of the CF-specific transport system, analogous to HlyB. A family of bacterial protein export systems, each involving a transport protein homologous to HlyB, is represented by the hemolysins of *Proteus* species (29, 30), *Actinobacillus actinomycetemcomitans* (34), and *Actinobacillus pleuropneumoniae* (18), by the *Pasteurella haemolytica* leukotoxin (24, 52), and by the bifunctional toxin of *Bordetella pertussis* (14). As each of these other bacterial toxin genes is found in a complex locus

TABLE 1. Protein sequences similar to deduced ComA protein of *S. pneumoniae*

Gene	Database sequence matched[a]		Homology score[b]
	Identifier	Protein	
hlyB	HLYB$PROVU	Hemolysin secretion protein (30)	918
hlyB	HLYB$ECOLI	Hemolysin secretion protein, plasmid (10)	912
hlyB	HLY2$ECOLI	Hemolysin secretion protein, chromosomal (11)	858
mdr	MDR$MOUSE	Multidrug resistance protein (17)	611
mdr1	MDR$HUMAN	Multidrug resistance protein (7)	590
STE6	MDR$YEAST	Pheromone transport protein (38)	458
mdr	MDR$PLAFA	Chloroquine resistance protein (12)	322
proV	PROV$ECOLI	Proline transport protein (15)	274
CF	CFTR$HUMAN	Transmembrane conductance regulator (49)	268
glnG	GLNQ$ECOLI	Glutamine transport protein GLNQ (45)	266
fecE	FECE$ECOLI	Iron(III) dicitrate transport protein (51)	262
malK	MALK$ECOLI	Maltose transport protein (13)	240
hisP	HISP$SALTY	Histidine permease (22)	234
hisP	HISP$ECOLI	Histidine permease (31)	226
cysA	CYSA$ANANI	Sulfate permease (16)	221
nodI	NODI$RHILE	Nodulation protein I (10)	215
pstB	PSTB$ECOLI	Phosphate transport protein (1)	210
chlD	CHLD$ECOLI	Molybdenum transport protein (26)	205
fhuC	FHUC$ECOLI	Ferrichrome transport protein (4)	204
oppF	OPPF$SALTY	Oligopeptide permease protein OPPF (25)	204

[a] Gene names, identifiers, and descriptions are as in the SWISS-PROT database. Proteins with the 20 highest optimized scores (47) are listed, with original literature citations.
[b] Optimized homology score, calculated after an initial search with a ktup value of 1 (47).

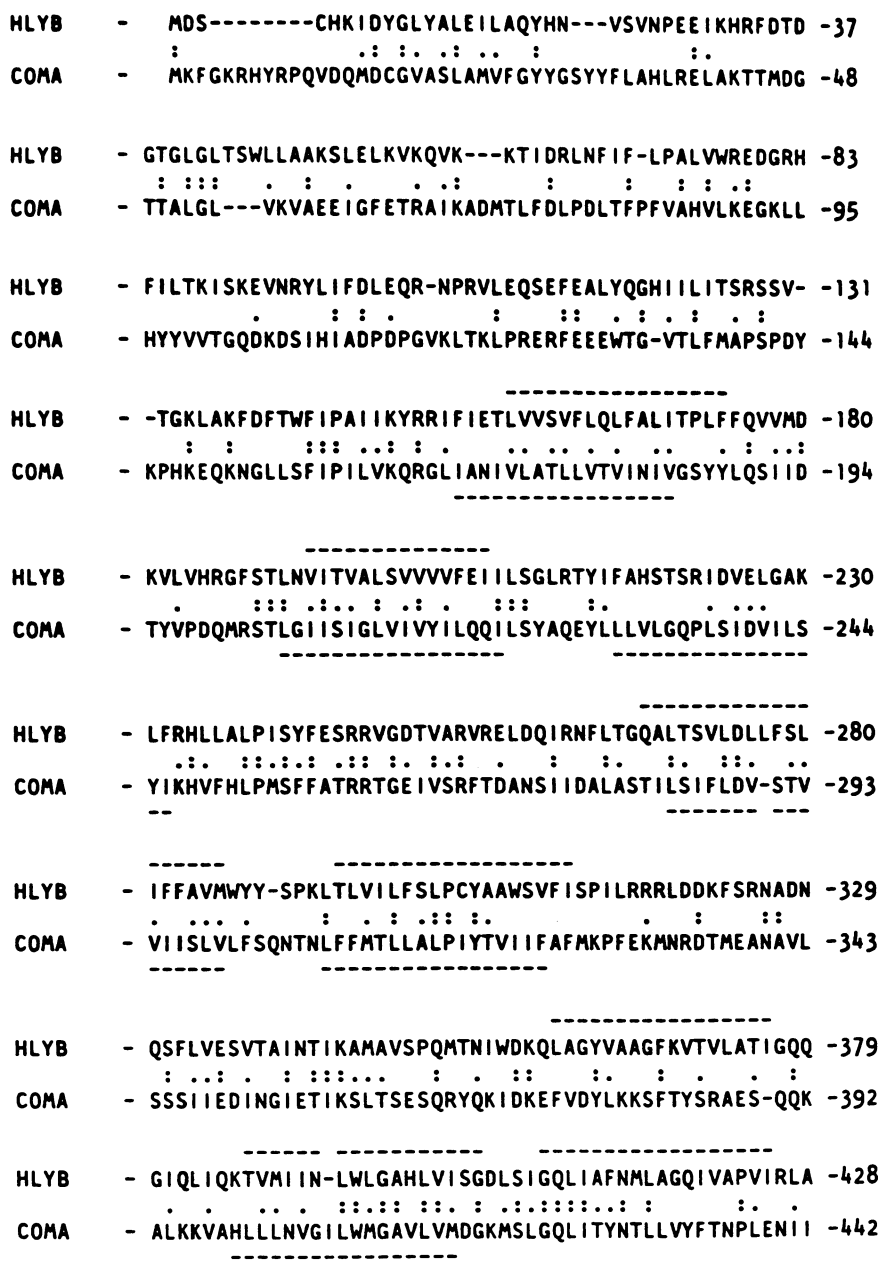

FIG. 3. Alignment of *S. pneumoniae* ComA and *E. coli* HlyB deduced protein sequences. The ComA sequence deduced from the DNA sequence as in Fig. 1 was aligned with the sequence of the *E. coli* hemolysin A secretion protein, HlyB, by the method of Myers and Miller (44), with an open gap cost of 5 and a unit gap cost of 2. NB-1 and NB-2 indicate nucleotide-binding sites (21). A colon (:) indicates identical amino acid residues. A dot (·) indicates similar residues (AST, DE, NQ, RK, ILMV, FYW). A short sequence shared only by all members of the ATP-dependent transport family is indicated by Transport. Positions in HlyB that are conserved in most members of the transport family (21) are starred (*). – – – – –, Potential membrane-spanning segments (28).

comprising genes for toxin synthesis, modification, and export, the *com* locus would, under this model, be expected to contain relevant genes in addition to *comA*, including the structural gene for CF.

A feature shared by at least three bacterial toxin secretion proteins (14, 24, 30, 52) but not by other members of the ATP-binding transport protein family is a kinase motif (G-x-G-x-x-G-17-K), occurring within the strongly conserved ATP-binding motif (denoted NB-2 in Fig. 3). In ComA, the final lysine of this motif is replaced by a proline (position 641). If the suggestion that this kinase motif is required for protein secretion (30) is verified, and if a substitution of proline for lysine disrupts kinase activity, this difference would suggest that ComA played a less direct role in CF elaboration.

Direct evidence for the above model is still absent. Indeed, even circumstantial evidence, such as induction of ComA at competence, a membrane location for ComA, or an orientation of ComA protein in the membrane appropriate

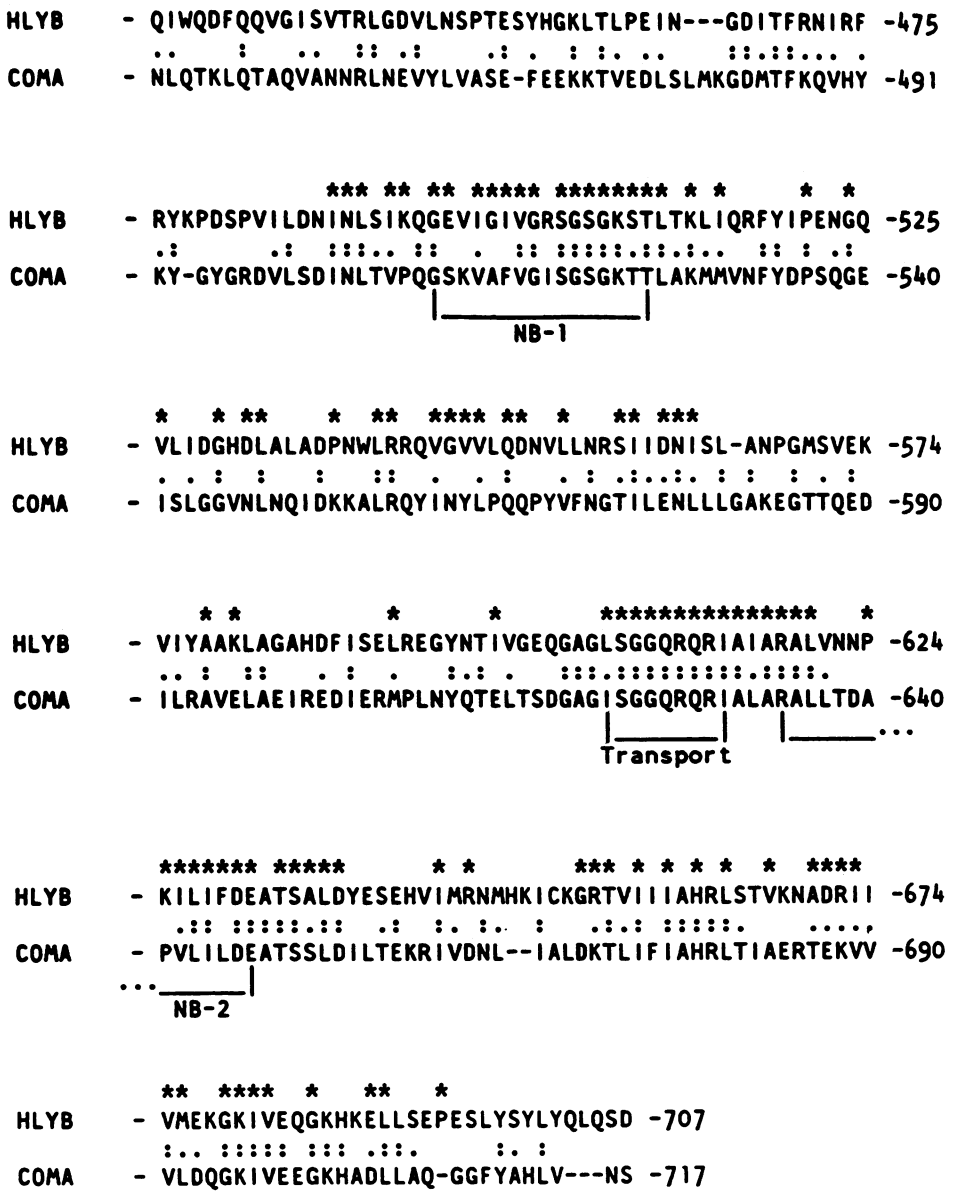

FIG. 3—*Continued.*

for export activity, is not yet available. Most known mutations affecting transformability affect specific steps of DNA uptake or intracellular processing, rather than the regulatory circuits discussed here. However, two other candidates for regulatory circuit genes do exist. In *Streptococcus sanguis*, the Wicky strain depends for competence on exogenous CF, which can be supplied from competent cultures of the spontaneously competent Challis strain; thus, Wicky appears to be a naturally occurring *com* strain (36, 46). In *S. pneumoniae*, the *trt* mutation (33), which causes cells to be constitutively competent and to retain that competence even during growth in trypsin, may represent an alteration in the CF-response circuitry which bypasses the CF activation step entirely.

ACKNOWLEDGMENT

This work was supported in part by a grant (DMB 8704310) from the National Science Foundation Genetic Biology Program.

REFERENCES

1. **Amemura, M., K. Makino, H. Shinagaw, A. Kobayashi, and A. Nakata.** 1984. Nucleotide sequence of the genes involved in phosphate transport and regulation of the phosphate regulon in Escherichia coli. J. Mol. Biol. **184**:241–250.
2. **Ames, G. F. L.** 1986. Bacterial periplasmic transport systems: structure, mechanism, and evolution. Annu. Rev. Biochem. **55**:397–425.
3. **Birnboim, H. C.** 1983. A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol. **100**:243–255.
4. **Burkardt, R., and V. Braun.** 1987. Nucleotide sequence of the fhuC and fhuD genes involved in iron (III) hydroxamate transport: domains in FhuC homologous to ATP-binding proteins. Mol. Gen. Genet. **209**:49–55.
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. **Chandler, M. S., and D. A. Morrison.** 1988. Identification of two proteins encoded by *com*, a competence control locus of *Streptococcus pneumoniae*. J. Bacteriol. **170**:3136–3141.
7. **Chen, C. J., J. E. Chin, K. Ueda, D. P. Clark, I. Pastan, M. M. Gottesman, and I. B. Roninson.** 1986. Internal duplication and homology with bacterial transport proteins in the *mdr1* (P-glycoprotein) gene from multidrug-resistant human cells. Cell **47**:381–389.
8. **Doolitle, R. F., M. J. Johnson, I. Husain, B. Van Houten, D. C. Thomas, and A. Sancar.** 1986. Domainal evolution of a prokaryotic DNA repair protein and its relationship to active transport proteins. Nature (London) **323**:451–453.
9. **Dreesen, T. D., D. H. Johnson, and S. Henikoff.** 1988. The brown protein of *Drosophila melanogaster* is similar to the white protein and to components of active transport complexes. Mol. Cell. Biol. **8**:5206–5215.
10. **Evans, I. J., and J. A. Downie.** 1986. The *nodI* gene product of *Rhizobium leguminosarum* is closely related to ATP-binding bacterial transport proteins; nucleotide sequence analysis of the *nodI* and *nodJ* genes. Gene **43**:95–101.
11. **Felmlee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163**:94–105.
12. **Foote, S. J., J. K. Thompson, A. F. Cowman, and D. J. Kemp.** 1989. Amplification of the multidrug resistance gene in some chloroquine resistant isolates of *P. falciparum*. Cell **57**:921–930.
13. **Gilson, E., H. Nikaido, and M. Hofnung.** 1982. Sequence of the *malK* gene in *E. coli* K12. Nucleic Acids Res. **10**:7449–7458.
14. **Glaser, P., H. Sakamoto, J. Bellalou, A. Ullmann, and A. Danchin.** 1988. Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*. EMBO J. **7**:3997–4004.
15. **Gowrishankar, J.** 1989. Nucleotide sequence of the osmoregulatory *proU* operon of *Escherichia coli*. J. Bacteriol. **171**:1923–1931.
16. **Green, L. S., D. E. Laudenbach, and A. R. Grossman.** 1989. A region of a cyanobacterial genome required for sulfate transport. Proc. Natl. Acad. Sci. USA **86**:1949–1953.
17. **Gros, P., J. Croop, and D. Housman.** 1986. Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology in bacterial transport proteins. Cell **47**:371–380.
18. **Gygi, D., J. Nicolet, J. Frey, and M. Cross.** 1990. Isolation of the *Actinobacillus pleuropneumoniae* haemolysin gene and the activation and secretion of the prohaemolysin by the HlyC, HlyB and HlyD proteins of *Escherichia coli*. Mol. Microbiol. **4**:123–128.
19. **Harr, R., M. Haggstrom, and P. Gustafsson.** 1983. Search algorithm for pattern match analysis of nucleic acid sequences. Nucleic Acids Res. **11**:2943–2957.
20. **Hess, J., W. Wels, M. Vogel, and W. Goebel.** 1986. Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison with a corresponding chromosomal hemolysin sequence. FEMS Microbiol. Lett. **34**:1–11.
21. **Higgins, C. F., M. P. Gallagher, M. L. Mimmack, and S. R. Pearce.** 1988. A family of closely related ATP-binding subunits from prokaryotic and eukaryotic cells. Bioessays **8**:111–116.
22. **Higgins, C. F., P. D. Haag, K. Nikaido, F. Ardeshir, G. Garcia, and G. F. L. Ames.** 1982. Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. Nature (London) **298**:723–727.
23. **Higgins, C. F., I. D. Hiles, G. P. C. Salmond, D. R. Gill, J. A. Downie, I. J. Evans, I. B. Holland, L. Gray, S. D. Buckel, and A. W. Bell.** 1986. A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature (London) **323**:448–450.
24. **Highlander, S. K., M. Chidambaram, M. J. Engler, and G. M. Weinstock.** 1989. DNA sequence of the *Pasteurella haemolytica* leukotoxin gene cluster. DNA **8**:15–28.
25. **Hiles, I. D., M. P. Gallagher, D. J. Jamieson, and C. F. Higgins.** 1987. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. J. Mol. Biol. **195**:125–142.
26. **Johann, S., and S. M. Hinton.** 1987. Cloning and nucleotide sequence of the *chlD* locus. J. Bacteriol. **169**:1911–1916.
27. **Kamijo, K., S. Taketani, S. Yokota, T. Osumi, and T. Hashimoto.** 1990. The 70-kDa peroxisomal membrane protein is a member of the Mdr (P-glycoprotein)-related ATP-binding protein superfamily. J. Biol. Chem. **265**:4534–4540.
28. **Klein, P., M. Kanehisa, and C. DeLisi.** 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta **815**:468–476.
29. **Koronakis, V., M. Cross, B. Senior, E. Koronakis, and C. Hughes.** 1987. The secreted hemolysins of *Proteus mirabilis*, *Proteus vulgaris*, and *Morganella morganii* are genetically related to each other and to the alpha-hemolysin of *Escherichia coli*. J. Bacteriol. **169**:1509–1515.
30. **Koronakis, V., E. Koronakis, and C. Hughes.** 1988. Comparison of the haemolysin secretion protein HlyB from *Proteus vulgaris* and *Escherichia coli*: site-directed mutagenesis causing impairment of export function. Mol. Gen. Genet. **213**:551–555.
31. **Kraft, R., and L. A. Leinwand.** 1987. Sequence of the complete P protein gene and part of the M protein gene from the histidine transport operon of Escherichia coli compared to that of Salmonella typhimurium. Nucleic Acids Res. **15**:8568.
32. **Kuchler, K., R. E. Sterne, and J. Thorner.** 1989. *Saccharomyces cerevisiae* STE6 gene product: a novel pathway for protein export in eukaryotic cells. EMBO J. **8**:3973–3984.
33. **Lacks, S. A., and B. Greenberg.** 1973. Competence for DNA uptake and deoxyribonuclease action external to cells in the genetic transformation of *Diplococcus pneumoniae*. J. Bacteriol. **114**:152–163.
34. **Lally, E. T., E. E. Golub, I. R. Kieba, N. S. Taichman, J. Rosenbloom, J. C. Rosenbloom, C. W. Gibson, and D. R. Demuth.** 1989. Analysis of the *Actinobacillus actinomycetemcomitans* leukotoxin gene. J. Biol. Chem. **264**:15451–15456.
35. **Lennox, E. S.** 1955. Transduction of linked genetic characters of the host by bacteriophage P1. Virology **1**:190–206.
36. **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.
37. **Lopez, A., C. Clave, R. Capeyrou, V. Lafontan, and M. C. Trombe.** 1989. Ionic and energetic changes at competence in the naturally transformable bacterium *Streptococcus pneumoniae*. J. Gen. Microbiol. **135**:2189–2197.
38. **McGrath, J. P., and A. Varshavsky.** 1989. The yeast STE6 gene encodes a homologue of the mammalian multidrug resistance P-glycoprotein. Nature (London) **340**:400–404.
39. **Morrison, D. A.** 1981. Competence-specific protein synthesis in *Streptococcus pneumoniae*, p. 39–54. *In* M. Polsinelli and G. Mazza (ed.), Transformation—1980. Cotswold Press, Oxford.
40. **Morrison, D. A., and M. Baker.** 1979. Competence for genetic transformation in pneumococcus depends on the synthesis of a small set of proteins. Nature (London) **282**:215–217.
41. **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.
42. **Mulligan, M. E., D. K. Hawley, R. Entriken, and W. R. McClure.** 1984. *E. coli* promoter sequences predict in vitro RNA polymerase selectivity. Nucleic Acids Res. **12**:789–800.
43. **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.
44. **Myers, E. W., and W. Miller.** 1988. Optimal alignments in linear space. Comput. Appl. Biosci. **4**:11–17.
45. **Nohno, T., T. Saito, and J. Hong.** 1986. Cloning and complete nucleotide sequence of the Escherichia coli glutamine permease operon (glnHPQ). Mol. Gen. Genet. **205**:260–269.
46. **Pakula, R., and W. Walczak.** 1963. On the nature of competence of transformable streptococci. J. Gen. Microbiol. **31**:125–133.
47. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.

48. **Radnis, B. A., D. K. Rhee, and D. A. Morrison.** 1990. Genetic transformation in *Streptococcus pneumoniae*: nucleotide sequence and predicted amino acid sequence of *recP*. J. Bacteriol. **172**:3669–3674.
49. **Riordan, J. R., J. M. Rommens, B. S. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, J. L. Chou, and M. L. Drumm.** 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science **245**:1066–1072.
50. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
51. **Staudenmaier, H., B. Van Hove, Z. Yaraghi, and V. Braun.** 1989. Nucleotide sequence of the *fecBCDE* genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. J. Bacteriol. **171**:2626–2633.
52. **Strathdee, C. A., and R. Y. C. Lo.** 1989. Cloning, nucleotide sequence, and characterization of genes encoding the secretion function of the *Pasteurella haemolytica* leukotoxin determinant. J. Bacteriol. **171**:916–928.
53. **Tartof, K., and C. Hobbs.** 1987. Improved media for growing plasmid and cosmid clones. Focus **9**:12.
54. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure of ribonucleic acids. Nature (London) **246**:40–41.
55. **Tomasz, A.** 1966. Model for the mechanism controlling the expression of the competent state in *Pneumococcus* cultures. J. Bacteriol. **91**:1050–1061.
56. **Tomasz, A.** 1971. Cell physiological aspects of DNA during genetic transformation in bacteria, p. 4–18. *In* L. Ledoux (ed.), Informative molecules in biological systems. North-Holland Publishing Co., Amsterdam.
57. **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.
58. **Tomasz, A., and J. L. Mosser.** 1966. On the nature of the pneumococcal activator substance. Proc. Natl. Acad. Sci. USA **55**:58–66.
59. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.