JOURNAL OF BACTERIOLOGY, Apr. 1986, p. 29–33
0021-9193/86/040029-05$02.00/0

Vol. 166, No. 1

Nucleotide Sequence of the Enterotoxin B Gene from *Staphylococcus aureus*

CHRISTOPHER L. JONES AND SALEEM A. KHAN*

Department of Microbiology, Biochemistry, and Molecular Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261

Received 4 November 1985/Accepted 23 January 1986

The complete nucleotide sequence of the enterotoxin B gene from *Staphylococcus aureus*, as well as the 5′ and 3′ flanking regions, was determined. Starting from an ATG initiator codon, an open reading frame encoded the enterotoxin B precursor that consisted of 266 amino acids (M_r, 31,400). The 5′ terminal portion of the gene encodes a signal peptide 27 amino acids long. The deduced amino acid sequence matched, with a few exceptions, the published amino acid sequence of enterotoxin B. The structural gene was flanked on the 5′ side by a promoter-like sequence and on the 3′ side by a palindromic structure followed by a thymine-rich region that resembled a transcription termination signal. Downstream from the *entB* gene were two overlapping open reading frames corresponding to 134 and 185 amino acids in the opposite orientation. The signal sequence of the enterotoxin B precursor resembled that of other secreted proteins found in other bacteria.

Staphylococcus aureus produces a number of extracellular proteins that are associated with its pathogenicity. One such group of proteins is composed of the enterotoxins that are involved in staphylococcal food poisoning (1). Staphylococcal enterotoxins are classified into six serological groups: A, B, C_1, C_2, D, and E. These enterotoxins have similar structural and biological properties (1).

Staphylococcal enterotoxin B (SEB) has been studied in detail. SEB has been purified to homogeneity and consists of 239 amino acids in a single polypeptide chain with a molecular weight of 28,336 (5, 11). The complete amino acid sequence of SEB was determined by amino acid analysis (11). There is extensive homology between the amino acid sequence of SEB and that of enterotoxin C_1, which also consists of 239 amino acids (26). A precursor to SEB having an approximate molecular weight of 32,000 (34) was identified as a membrane component of SEB-producing *S. aureus* strains. A SEB precursor was also identified in the membranes of *Escherichia coli* strains carrying the cloned enterotoxin B gene (*entB*) (22). The mechanism of action of SEB is not fully understood, but it probably causes vomiting and diarrhea by its emetic action on the abdominal viscera (33).

Considerable interest has been directed toward an understanding of the mechanism of SEB production in *S. aureus*. Several studies have suggested the involvement of a small plasmid, pSN2, in SEB synthesis (6, 28–30). However, the determination of the complete nucleotide sequence of pSN2 and cloning of the *entB* gene have demonstrated that this plasmid is not involved in SEB production (12, 22).

In a previous report we have described the cloning and expression of the *entB* gene in *S. aureus* and *E. coli* (22). Studies on the expression of the *entB* gene in *E. coli* have shown that the gene is expressed only when transcribed from a strong *E. coli* promoter (22). Here we report the complete nucleotide sequence of the *entB* gene from *S. aureus* S6, including the 5′ and 3′ flanking regions. The structural gene consists of 798 nucleotides that encode the SEB precursor with 266 amino acid residues and a molecular weight of 31,400. A signal sequence of 27 amino acid residues is present at the NH_2-terminal end. The mature SEB consists of 239 amino acid residues, corresponding to a molecular weight of 28,336, in agreement with published results obtained with the purified protein (11).

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* SK155(pSK155) that contains the cloned *entB* gene from *S. aureus* S6 (22) was used in these studies. This strain was grown in L broth containing 40 μg of ampicillin per ml (18).

Preparation of plasmid DNA and restriction enzyme analysis. Plasmid DNA was prepared by CsCl-ethidium bromide density gradient centrifugation of cleared lysates (4). Restriction endonucleases were purchased from either New England Biolabs or Bethesda Research Laboratories and used as recommended by the suppliers. Restriction nuclease sites within the cloned DNA fragment were determined by using single and double digestions with various restriction enzymes. DNA fragments were analyzed by electrophoresis on either 1% agarose or 5% polyacrylamide gels. Restriction nuclease fragments for sequencing were isolated by preparative 4 or 5% slab polyacrylamide gel electrophoresis followed by crushing and elution (15).

DNA sequence determination. Restriction nuclease fragments were dephosphorylated with bacterial alkaline phosphatase (Bethesda Research Laboratories) and 5′ end labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase (Bethesda Research Laboratories) (15). Singly labeled DNA fragments were prepared either by secondary restriction enzyme digestion and subsequent separation on polyacrylamide gels or by the strand separation technique (15). DNA fragments were isolated by crushing and elution and sequenced by the method of Maxam and Gilbert (15). The samples were analyzed on 6, 8, and 20% denaturing polyacrylamide gels. The nucleotide sequence data were analyzed by an IBM PC-XT computer with the program of Schwindinger and Warner (27).

RESULTS AND DISCUSSION

DNA sequence of the *entB* gene. We have earlier reported the cloning of the *entB* gene from *S. aureus* S6 into *E. coli* by using pBR322 as the vector plasmid (22). The resulting

* Corresponding author.

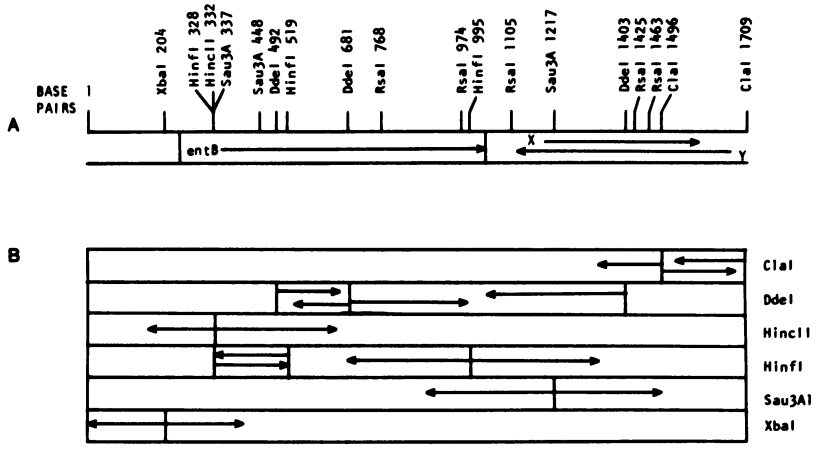

FIG. 1. Restriction nuclease map and sequencing strategy of the *entB* gene. (A) Partial restriction map of the DNA containing the *entB* gene. Nucleotides are numbered starting from the extent of sequence determination to the 5′ side of the *entB* gene (nucleotide 1) to the *Cla*I site at nucleotide 1709. The direction of transcription and translation of the *entB* gene is shown by an arrow. X and Y correspond to oppositely oriented overlapping ORFs. (B) Sequencing strategy of the DNA. Arrows indicate the direction and extent of sequence determination from the 5′ end of restriction nuclease fragments.

recombinant plasmid, pSK155, contains the *entB* gene within a 2.3-kilobase region. The nucleotide sequence of a 156-base-pair (bp) *Taq*I fragment that encodes amino acids 14 to 65 of the extracellular SEB was also described previously (22). A detailed restriction map of a 1709-bp region containing the *entB* gene and the sequencing strategy is shown in Fig. 1. About 90% of the sequence was determined by sequencing both the DNA strands. The complete nucleotide sequence of the *entB* gene, including the 5′ and 3′ flanking regions, is shown in Fig. 2. Starting from an ATG codon at nucleotide 244, there was an open reading frame (ORF) of 798 nucleotides that terminated in a TGA stop codon at nucleotide 1042. This corresponded to the SEB precursor which consisted of 266 amino acids, with a calculated molecular weight of 31,400. The SEB precursor contained a putative signal sequence of 27 amino acid residues. The calculated molecular weight of the 239-amino acid mature SEB was 28,366. This is close to the molecular weight of 28,500 calculated from the amino acid sequence of SEB (11). Since the deduced amino acid sequence corresponded to the amino acid sequence determined previously for SEB (11), we conclude that the ORF encodes the SEB precursor. There are several reasons to postulate that the ATG codon at nucleotide 244 is the translational start site. (i) This ATG is the first initiating codon upstream of the GAG codon (glutamic acid) at nucleotide 325 that encodes the NH_2-terminal residue of SEB (11). (ii) The ATG codon is preceded by a strong Shine-Dalgarno sequence, AAAGGAG, that can form seven base pairs (including three G-C pairs) with the 3′ end of *Bacillus subtilis* 16S rRNA (16). A strong ribosome-binding sequence is common to most genes from gram-positive bacteria that have been studied so far (16, 20, 25, 31, 35). In addition, the space between the last guanine in this sequence and the initiator ATG codon was seven nucleotides, which is in good agreement with the sites utilized by *B. subtilis* and *E. coli* ribosomes (9, 16). Although TTG is a common initiator codon in gram-positive bacteria (16, 35), such a codon at nucleotide 277 was not preceded by a Shine-Dalgarno sequence. (iii) The ATG codon at nucleotide 244 gave a putative signal peptide of 27 amino acids, which compares well with its estimated size reported earlier (22, 34).

The 5′ flanking region of the *entB* gene contained a possible −10 TATAT sequence which is an acceptable fit to the canonical TATAAT sequence (23). A possible −35 TTGAA sequence differed by only one nucleotide from the consensus TTGACA sequence. However, the significance of this promoter-like sequence is not clear since we have shown earlier that the *entB* gene is expressed in *E. coli* only when transcribed from a strong *E. coli* promoter (22). Forty nucleotides downstream from the TGA stop codon of the *entB* gene is a palindromic sequence that might act as a transcription terminator (23). The possible hairpin structure contained 11 complementary nucleotide pairs and one A-C mismatch. This structure was followed on the 3′ side by a thymine-rich region. These features at the 3′ end suggest that transcription terminates at this point.

Downstream from the *entB* gene were two ORFs that could potentially encode polypeptides consisting of more than 100 amino acids. The ORF X consisted of 134 codons and extended from nucleotides 1189 to 1590 on the same strand as *entB* (Fig. 2). However, no initiator ATG codons were present in this ORF, and other possible initiator codons, TTG (leucine) and GTG (valine), were not preceded by strong Shine-Dalgarno sequences that are typical of genes from gram-positive bacteria (16, 20, 25, 31, 35). The ORF Y extended from an initiator ATG codon at position 1679 to position 1123 on the opposite strand (Fig. 2). This ORF was preceded by a Shine-Dalgarno sequence at position 1688 and could potentially encode a polypeptide 185 amino acids long. Nine nucleotides downstream from the TAA stop codon of ORF Y was a palindromic sequence that could possibly act as a transcription terminator. This sequence was complementary to that shown in Fig. 2 for the possible transcription terminator of the *entB* gene. It is interesting that ORFs X and Y overlapped one another (Fig. 2). Overlapping antiparallel ORFS have been reported in several insertion sequences in procaryotes and in a few genes of bacteriophage λ (21). Most of these antiparallel ORFs have been found in the same reading frame. However, the X and Y ORFs were not in the same reading frame. The significance of these ORFs is not clear, although it has been suggested that the *entB* gene is part of a mobile genetic element like a transposon (28).

Cloning experiments were carried out to determine whether either the X or the Y ORF is involved in SEB production. A derivative of the pC194 plasmid (10) was

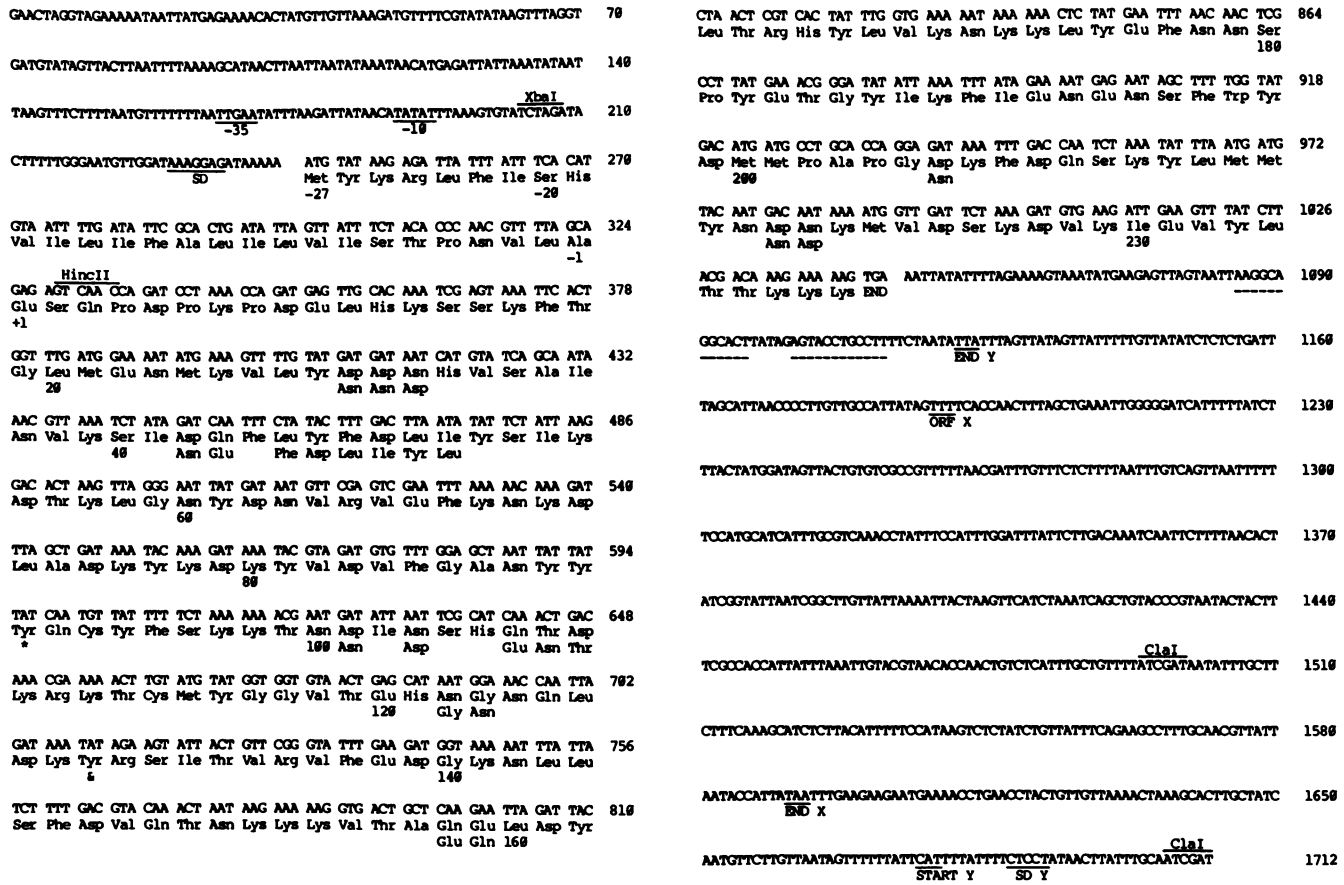

FIG. 2. Nucleotide and amino acid sequence of the *entB* gene and the 5′ and 3′ flanking regions. The sequence shown corresponds to that shown in Fig. 1. Important restriction endonuclease sites are indicated. The possible promoter sequence (−10 and −35) and the possible Shine-Dalgarno sequence (SD) of the *entB* gene are shown. A possible transcription terminator sequence is indicated by broken lines. Numbering of the amino acids starts at the NH_2 terminus of mature SEB. Amino acids −1 to −27 correspond to the signal peptide. The amino acids that are different from the reported amino acid sequence of SEB (11) are indicated. The tyrosine residue at position 91 (*) is absent in the reported amino acid sequence, while an additional tyrosine residue (&) is present at position 129. Two ORFs (ORF X and ORF Y) are indicated, as is the possible Shine-Dalgarno sequence of Y (SD Y).

constructed by inserting a 41-bp *Hin*cII-*Hae*III fragment containing the multiple cloning sites from the pUC19 plasmid (36) into the *Hin*dIII site of pC194 after the attachment of *Hin*dIII linkers (24). This plasmid was named pSK265. A 1,777-bp *entB*-containing restriction nuclease fragment extending from the *Kpn*I site (282 bp upstream from bp 1) (Fig. 1; reference 22) was isolated by polyacrylamide gel electrophoresis (15). This fragment was ligated to *Kpn*I- and *Cla*I-digested pSK265 DNA by using T4 DNA ligase. The ligation mixture was used to transform *S. aureus* RN4220 by the protoplast transformation technique (3). Chloramphenicol-resistant colonies were picked and screened for the presence of the recombinant plasmid by agarose gel electrophoresis of sheared whole-cell lysates (7). *S. aureus* SK291(pSK291) containing the 1,777-bp fragment was checked for SEB production by immunodiffusion analysis of culture supernatants as described earlier (22). This strain produced SEB (data not shown). These results show that the functional *entB* gene was contained within the 1,777-bp *Kpn*I-*Cla*I region and that the X and Y ORFs were not involved in the synthesis or secretion of SEB.

The codon usage of *entB* reflected the low G-C content of the *entB* gene. With two exceptions (UCA and UCG for serine; ACA and ACG for threonine), there was a clear preference for adenine or uracil in the third position. The G-C content of the degenerate third base of the codons in the majority of *S. aureus* chromosomal genes that have been sequenced so far is approximately 30 to 33% (20, 25, 31). This value is similar to the overall G-C content of *S. aureus* chromosomal DNA (8). In contrast, the G-C content of the degenerate third base of the codons for *entB* and staphylococcal nuclease (31) is 22 and 17%, respectively. These values are low compared with the other *S. aureus* chromosomal genes and are similar to the 22% G-C content that has been found for many *S. aureus* plasmid genes (10, 13). Thus, it is possible that *entB* may have a nonchromosomal origin, such as part of a transposon, as has been suggested earlier (28), or a phage.

Amino acid sequence. The amino acid sequence of SEB, as deduced from the DNA sequence, is shown in Fig. 2. The deduced sequence matched well with the amino acid sequence reported previously for the protein (11). The differences in the deduced and reported sequence are also indicated in Fig. 2. Both the DNA strands were sequenced in the regions involving discrepancies between the nucleotide sequence and the reported amino acid sequence. Most of these differences involved aspartic acid and asparagine, and glutamic acid and glutamine. It is likely that these differences were due to the instability of asparagine and glutamine in the Edman procedure or to the possible contamination by an

amidase in SEB preparations, as was suggested earlier (26). There is considerable homology between the amino acid sequence of SEB and enterotoxin C_1, especially in the carboxy and NH_2-terminal regions (26). The homology in the NH_2-terminal region probably accounts for the cross-reactivity of these toxins with heterologous antiserum (32). By using the best possible fit, 151 of 239 amino acids were reported to be identical between enterotoxins B and C_1 (26). In our results, this homology was increased by eight amino acid residues (159 of 239).

The signal sequence. The deduced amino acid sequence of SEB contained a putative signal sequence of 27 amino acid residues. The M_r of this peptide was calculated to be 3,034. The predicted size was consistent with an earlier report indicating that the molecular weight of the SEB precursor is greater than that of the mature SEB by about 3,500 (34). The signal sequence of SEB contained two basic amino acids in the NH_2-terminal region followed by a long stretch of uncharged, mainly hydrophobic amino acids and an alanine residue before the cleavage site (Fig. 2). These features are common to the signal peptides of most secreted proteins (16, 17, 25, 35).

The genes for a number of *S. aureus* extracellular proteins have been cloned in *E. coli* (2, 14, 16, 25, 31, 35). Most of these proteins, including protein A, enterotoxin A, β-lactamase, staphylokinase, staphylococcal nuclease, and the toxic shock syndrome exotoxin are secreted into the periplasmic fraction in *E. coli*. The signal sequences of protein A, β-lactamase, and staphylokinase resemble those of SEB (16, 25, 35). We have shown earlier that SEB produced by *E. coli* strains carrying the cloned *entB* gene is not secreted into the periplasmic space (22). Additionally, mature SEB was present in the cytoplasmic fraction of these strains. It is possible that the mature portion of SEB contains a "poison" sequence that inhibits the transport of this toxin across the membrane in *E. coli*. Such an inhibitory region has been found in β-galactosidase from *E. coli*, which is a cytoplasmic protein (19). However, we have not yet ruled out the possibility that the mature SEB is loosely attached to the outer surface of the cytoplasmic membrane in *E. coli* and is not released by the procedure used for the preparation of the periplasmic fraction (22).

ACKNOWLEDGMENTS

We thank B. Rooney for preparation of the manuscript.

This work was supported by Public Health Service grant AI 19783 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. **Bergdoll, M. S.** 1979. Staphylococcal intoxications, p. 443–493. *In* H. Riemann and F. L. Bryan (ed.), Food-borne infections and intoxications. Academic Press, Inc., New York.
2. **Betley, M. J., S. Lofdahl, B. N. Kreiswirth, M. S. Bergdoll, and R. P. Novick.** 1984. Staphylococcal enterotoxin A gene is associated with a variable genetic element. Proc. Natl. Acad. Sci. USA **81**:5179–5183.
3. **Chang, S., and S. N. Cohen.** 1979. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. Mol. Gen. Genet. **168**:111–115.
4. **Clewell, D. B., and D. R. Helinski.** 1969. Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. USA **62**:1159–1166.
5. **Dalidowicz, J. E., S. J. Silverman, E. J. Shantz, D. Stefanye, and L. Spero.** 1966. Chemical and biological properties of reduced and alkylated staphylococcal enterotoxin B. Biochemistry **5**:2375–2381.
6. **Dyer, D. W., and J. J. Iandolo.** 1981. Plasmid-chromosomal transition of genes important in staphylococcal enterotoxin B expression. Infect. Immun. **33**:450–458.
7. **Eckhardt, T.** 1978. A rapid method for the identification of plasmid deoxyribonucleic acid in bacteria. Plasmid **1**:584–588.
8. **Fasman, G. D. (ed.).** 1976. CRC Handbook of biochemistry and molecular biology, vol. 2, 3rd ed. p. 136. CRC Press, Inc., Boca Raton, Fla.
9. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translational initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
10. **Horinouchi, S., and B. Weisblum.** 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. J. Bacteriol. **150**:815–825.
11. **Huang, I.-Y., and M. S. Bergdoll.** 1970. The primary structure of staphylococcal enterotoxin B. J. Biol. Chem. **245**:3518–3525.
12. **Khan, S. A., and R. P. Novick.** 1982. Structural analysis of plasmid pSN2 in *Staphylococcus aureus*: no involvement in enterotoxin B production. J. Bacteriol. **149**:642–649.
13. **Khan, S. A., and R. P. Novick.** 1983. Complete nucleotide sequence of pT181, a tetracycline-resistant plasmid from *Staphylococcus aureus*. Plasmid **10**:251–259.
14. **Kreiswirth, B. N., S. Lofdahl, M. J. Betley, M. O'Reilly, P. M. Schlievert, M. S. Bergdoll, and R. P. Novick.** 1983. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. Nature (London) **305**:709–712.
15. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
16. **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* β-lactamase gene. J. Biol. Chem. **256**:11283–11291.
17. **Michaelis, S., and J. Beckwith.** 1982. Mechanism of incorporation of cell envelope proteins in *Escherichia coli*. Annu. Rev. Microbiol. **36**:435–465.
18. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 433. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Moreno, F., A. V. Fowler, M. Hall, T. J. Silhavy, I. Zabin, and M. Schwartz.** 1980. A signal sequence is not sufficient to lead β-galactosidase out of the cytoplasm. Nature (London) **286**:356–359.
20. **Murphy, E.** 1985. Nucleotide sequence of *ermA*, a macrolide-lincosamide-streptogramin B determinant in *Staphylococcus aureus*. J. Bacteriol. **162**:633–640.
21. **Normark, S., S. Bergstrom, T. Edlund, T. Grundstrom, B. Jaurin, P. Lindberg, and O. Olsson.** 1983. Overlapping genes. Annu. Rev. Genet. **17**:499–525.
22. **Ranelli, D. M., C. L. Jones, M. B. Johns, G. J. Mussey, and S. A. Khan.** 1985. Molecular cloning of staphylococcal enterotoxin B gene in *Escherichia coli* and *Staphyloccocus aureus*. Proc. Natl. Acad. Sci. USA **82**:5850–5854.
23. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
24. **Rothstein, R. J., L. F. Lau, C. P. Bahl, S. A. Narang, and R. Wu.** 1979. Synthetic adaptors for cloning DNA. Methods Enzymol. **65**:98–109.
25. **Sako, T., and N. Tschuchida.** 1983. Nucleotide sequence of the staphylokinase gene from *Staphylococcus aureus*. Nucleic Acids Res. **11**:7679–7693.
26. **Schmitz, J. J., and L. Spero.** 1983. The complete amino acid sequence of staphylococcal enterotoxin C_1. J. Biol. Chem. **258**:6300–6306.
27. **Schwindinger, W. F., and J. R. Warner.** 1984. DNA sequence analysis on the IBM-PC. Nucleic Acids Res. **12**:601–604.
28. **Shafer, W. M., and J. J. Iandolo.** 1979. Genetics of staphylococcal enterotoxin B in methicillin-resistant isolates of *Staphylococcus aureus*. Infect. Immun. **25**:902–911.
29. **Shafer, W. M., and J. J. Iandolo.** 1980. Transduction of staphylococcal enterotoxin B synthesis: establishment of the toxin gene in a recombination-deficient mutant. Infect. Immun. **27**:280–282.

30. **Shalita, Z., I. Hertman, and S. Sarid.** 1977. Isolation and characterization of a plasmid involved with enterotoxin B production in *Staphylococcus aureus*. J. Bacteriol. **129**:317–325.
31. **Shortle, D.** 1983. A genetic system for analysis of staphylococcal nuclease. Gene **22**:181–189.
32. **Spero, L., and B. A. Morlock.** 1979. Cross-reactions between tryptic polypeptides of staphylococcal enterotoxins B and C_1. J. Immunol. **122**:1285–1289.
33. **Sugiyama, H., and T. Hayama.** 1965. Abdominal viscera as site of emetic action for staphylococcal enterotoxin in monkey. J. Infect. Dis. **115**:330–336.
34. **Tweten, R. K., and J. J. Iandolo.** 1981. Purification and partial characterization of a putative precursor to staphylococcal enterotoxin B. Infect. Immun. **34**:900–907.
35. **Uhlen, M., B. Guss, B. Nilsson, S. Gatenbeck, L. Philipson, and M. Lindberg.** 1984. Complete sequence of the staphylococcal gene encoding protein A. J. Biol. Chem. **259**:1695–1702.
36. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**:103–119.