INFECTION AND IMMUNITY, Sept. 1988, p. 2518–2520
0019-9567/88/092518-03$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 56, No. 9

Nucleotide Sequence of Streptococcal Pyrogenic Exotoxin Type C

STEPHEN C. GOSHORN* AND PATRICK M. SCHLIEVERT

Department of Microbiology, University of Minnesota Medical School, Minneapolis, Minnesota 55455

Received 28 March 1988/Accepted 15 June 1988

The nucleotide sequence of the gene *speC*, encoding streptococcal pyrogenic exotoxin type C (SPE C), was determined. The gene encoded a mature protein of 208 amino acids, with a calculated molecular weight of 24,354. The mature amino acid sequence of SPE C was analyzed for homology with the amino acid sequences of streptococcal pyrogenic exotoxin type A, the staphylococcal enterotoxins, and toxic shock syndrome toxin-1. Of these, SPE C shared the greatest amount of homology with streptococcal exotoxin type A.

Streptococcal pyrogenic exotoxin type C (SPE C) is a member of a family of biologically and biochemically related toxins produced by *Streptococcus pyogenes* and *Staphylococcus aureus* (11a). This family includes the group A streptococcal pyrogenic toxins, staphylococcal toxic shock syndrome toxin-1 (TSST-1), and the staphylococcal enterotoxins and pyrogenic exotoxins.

The streptococcal pyrogenic toxins, classically known as the scarlet fever or erythrogenic toxins, occur in three serologically distinct forms, designated A, B, and C. In addition to being the causative agents of the symptoms associated with scarlet fever, these toxins have been associated with streptococcal toxic shock-like disease (10) and may play a role in the early events of rheumatic fever. SPE C is the most common toxin type found in recent clinical isolates, often occurring with streptococcal pyrogenic exotoxin type B (SPE B) (19). Nearly all rheumatic fever-associated strains of streptococci produce the type C toxin (19), and a majority make SPE C while not making streptococcal pyrogenic exotoxin type A (SPE A) or SPE B.

The genes for SPE A, TSST-1, and enterotoxins A, B, and C (Ent A, Ent B, and Ent C_1, respectively) have been cloned and sequenced (3–5, 7, 8, 12–14, 17, 20). Comparisons of nucleotide and amino acid sequences have shown SPE A and Ent B and Ent C_1 to be highly homologous (8, 20). Recently, it was reported that the nucleotide and amino acid sequences of Ent A shared homology with those of these toxins, although this homology was less than that seen among SPE A, Ent B, and Ent C_1 (4). Significant homology has not been found between TSST-1 and the other toxins.

We have reported the cloning of the gene for SPE C elsewhere (11a). We undertook this investigation to determine the nucleotide and amino acid sequences of SPE C and to evaluate the extent of sequence relatedness between SPE C and the other known pyrogenic toxins.

The gene encoding SPE C (*speC*) was localized to a 1.7-kilobase DNA fragment and ligated to the replicative forms of bacteriophages M13 mp18 and mp19 (15). After transformation into *Escherichia coli* JM101 (*Δlac-pro supE thi* F′ *tra D36 proAB lacI*^{q} *ZΔM15* [9]), recombinant phage were selected, and single-stranded phage DNA was prepared. Deletion subclones were obtained by using the exonuclease activity of T4 DNA polymerase in a procedure described by Dale et al. (11). Templates for dideoxy sequencing were then prepared subsequent to transformation of *E. coli* JM101 with the deletion subclones (16). Each nucleotide in the *speC*-coding sequence and flanking DNA was sequenced a minimum of three times. Fifty eight percent of the coding sequence (Fig. 1, nucleotides 28 to 440) was determined for both strands.

The nucleotide sequence of *speC* contained an open reading frame of 705 base pairs coding for 235 amino acids (Fig. 1). The probable −35 and −10 promoter regions (nucleotides −142 to −148 and −102 to −108, respectively), conformed closely to the *E. coli* consensus promoter sequences, thus facilitating recognition by the *E. coli* RNA polymerase and expression of the gene in *E. coli*. A typical Shine-Dalgarno sequence (AAGGAG) was present 6 bases 5′ of the ATG start codon.

Translation of *speC* terminated at nucleotide 705, which was succeeded 17 bases 3′ by two sets of palindromic sequences which may be able to form stem and loop structures (overlined in Fig. 1). The larger palindrome was strikingly similar to palindromic sequences found in the 3′-untranslated regions of the genes encoding Ent C_1 and SPE A. Palindromes were also found in the 3′-untranslated regions of the genes for Ent B and TSST-1, but these were not similar to the palindromic sequences in either *speC* or the other toxin genes. It is possible that these stem and loop structures function as transcription terminators (18) or to protect transcripts from degradation.

Automated amino-terminal peptide sequencing (model 470A gas-phase protein sequencer; Applied Biosystems, Foster City, Calif.) was used to determine the first 25 amino acids of the mature SPE C protein (data not shown). This analysis allowed determination of the proper reading frame and indicated that the mature protein began at residue 28. Residues 28 to 52 corresponded exactly to the sequence determined by amino-terminal sequencing. The first 27 residues, beginning with the start methionine, likely constitute a hydrophobic signal peptide which is apparently removed from the mature protein by cleavage between serine and asparagine residues (Fig. 1). The mature protein of 208 amino acids had a calculated molecular weight of 24,354, as compared with a molecular weight of 23,800 estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (11a).

The deduced amino acid sequence of mature SPE C was compared with the published amino acid sequences of mature SPE A, Ent A, Ent B, Ent C_1, and TSST-1. Monte Carlo analysis was utilized to evaluate the significance of sequence similarities (Table 1). The sequences to be compared were first optimally aligned by using the SS2 algorithm of Altschul and Erickson to produce an optimal similarity score (2). One of the sequences was then randomized a specified number of

* Corresponding author.

```
      -35                                       -10
CAACCTTGACTATTTAAATGGAACTGCCACTCCTAAAAACTAAAATATAAATACA

TTTATAAAATTTCTAAATAAACAGAAATCTGATTTTTAACTACTTACTGCTATTT
                                         SD
CATGTATTCTCGTACGAGTAATACATTTAATTAAGGAGAAAAA ATG AAA AAG   9
                                            MET Lys Lys

ATT AAC ATC ATC AAA ATA GTT TTC ATA ATT ACA GTC ATA CTG   51
Ile Asn Ile Ile Lys Ile Val Phe Ile Ile Thr Val Ile Leu

ATT TCT ACT TAT TTC ACC TAT CAT CAA AGT*GAC TCT AAG AAA   93
Ile Ser Thr Tyr Phe Thr Tyr His Gln Ser Asp Ser Lys Lys

GAC ATT TCG AAT GTT AAA AGT GAT TTA CTT TAT GCA TAC ACT   135
Asp Ile Ser Asn Val Lys Ser Asp Leu Leu Tyr Ala Tyr Thr

ATA ACT CCT TAT GAT TAT AAA GAT TGC AGG GTA AAT TTT TCA   177
Ile Thr Pro Tyr Asp Tyr Lys Asp Cys Arg Val Asn Phe Ser

ACG ACA CAC ACA TTA AAC ATT GAT ACT CAA AAA TAT AGA GGG   219
Thr Thr His Thr Leu Asn Ile Asp Thr Gln Lys Tyr Arg Gly

AAA GAC TAT TAT ATT AGT TCC GAA ATG TCT TAT GAG GCC TCT   261
Lys Asp Tyr Tyr Ile Ser Ser Glu MET Ser Tyr Glu Ala Ser

CAA AAA TTT AAA CGA GAT GAT CAT GTA GAT GTT TTT GGA TTA   303
Gln Lys Phe Lys Arg Asp Asp His Val Asp Val Phe Gly Leu

TTT TAT ATT CTT AAT TCT CAC ACC GGT GAG TAC ATC TAT GGA   345
Phe Tyr Ile Leu Asn Ser His Thr Gly Glu Tyr Ile Tyr Gly

GGA ATT ACG CCT GCT CAA AAT AAT AAA GTA AAT CAT AAA TTA   387
Gly Ile Thr Pro Ala Gln Asn Asn Lys Val Asn His Lys Leu

TTG GGA AAT CTA TTT ATT TCG GGA GAA TCT CAA CAG AAC TTA   429
Leu Gly Asn Leu Phe Ile Ser Gly Glu Ser Gln Gln Asn Leu

AAT AAC AAG ATT ATT CTA GAA AAG GAT ATC GTA ACT TTC CAG   471
Asn Asn Lys Ile Ile Leu Glu Lys Asp Ile Val Thr Phe Gln

GAA ATT GAC TTT AAA ATC AGA AAA TAC CTT ATG GAT AAT TAT   513
Glu Ile Asp Phe Lys Ile Arg Lys Tyr Leu MET Asp Asn Tyr

AAA ATT TAT GAC GCT ACT TCT CCT TAT GTA AGC GGC AGA ATC   555
Lys Ile Tyr Asp Ala Thr Ser Pro Tyr Val Ser Gly Arg Ile

GAA ATT GGC ACA AAA GAT GGG AAA CAT GAG CAA ATA GAC TTA   597
Glu Ile Gly Thr Lys Asp Gly Lys His Glu Gln Ile Asp Leu

TTT GAC TCA CCA AAT GAA GGG ACT AGA TCA GAT ATT TTT GCA   639
Phe Asp Ser Pro Asn Glu Gly Thr Arg Ser Asp Ile Phe Ala

AAA TAT AAA GAT AAT AGA ATT ATC AAT ATG AAG AAC TTT AGT   681
Lys Tyr Lys Asp Asn Arg Ile Ile Asn MET Lys Asn Phe Ser

CAT TTC GAT ATT TAT CTT GAA AAA TAATTCATCATACACAAAAAACC
His Phe Asp Ile Tyr Leu Glu Lys TER

GCCCAGAATAATCTGAGCGGTTTTGTCTTATCTCGGAGCTTTACCTCCTAATTTA
```

FIG. 1. Nucleotide sequence of *speC*. Numbering is in reference to the ATG start codon. Possible promoter (−10 and −35) and Shine-Dalgarno (SD) sequences are denoted. The deduced amino acid sequence of SPE C is given below the nucleotide sequence. The asterisk after residue 27 indicates the probable cleavage site between the signal peptide and the mature protein. Overlined nucleotides 3′ of the translation stop codon are palindromic sequences.

times. Each randomized sequence was aligned with the probe sequence and assigned a value. The optimal alignment score was subtracted from the mean of the values from the randomized comparisons and then divided by the standard deviation of the randomized scores. The resulting value or Monte Carlo score was not considered to be indicative of homology if it was less than 3.0. A score of between 3.0 and 6.0 indicated possible homology, and a score of greater than 6.0 was indicative of probable homology.

The SPE C amino acid sequence appeared to be most related to that of SPE A and less so to those of the

TABLE 1. Monte Carlo analysis of the sequence similarities between mature amino acid sequences

Probe	Monte Carlo score for compared sequence					
	SPE A	SPE C	Ent A	Ent B	Ent C_1	TSST-1
SPE A		11.0	19.4	55.8	34.6	3.72
SPE C	10.2		6.49	5.31	6.32	1.79
Ent A	12.6	5.82		20.1	10.1	3.63
Ent B	31.7	4.40	14.8		54.8	4.18
Ent C_1	37.5	4.79	15.1	64.5		3.08
TSST-1	5.43	1.29	3.06	5.20	3.23	

enterotoxins; no significant homology was found with TSST-1 (Table 1). The enterotoxin and SPE A sequences appeared to be highly interrelated, while TSST-1 had possible sequence homology with only SPE A and Ent B. Betley and Mekalanos found a higher degree of amino acid relatedness among the toxins by allowing conservative amino acid changes in the sequence alignments (4).

Considering the shared properties of the toxins, the extent of amino acid sequence divergence which has occurred was unexpected. Whereas SPE A, Ent B, and Ent C_1 are clearly related at both the amino acid and nucleotide levels, SPE C and TSST-1 share much less homology with the other toxins while retaining the properties which define pyrogenic toxins. We have previously reported that SPE A, Ent B, and Ent C_1 share epitopes (6), but we have been unable to demonstrate any antigenic relationship between those toxins and SPE C and TSST-1. Amino acid alignments of SPE C with the other toxins reveal only a few clusters of conservation (data not shown), particularly in the carboxyl halves of the proteins. Regions which are conserved between SPE C and other toxins, having limited overall homology, may represent biologically important sites or sites necessary for the structural integrity of the proteins. At present, functions have not been assigned to particular regions of the toxins. Future studies will utilize site-specific mutagenesis to analyze such regions.

Nucleotide sequence alignments of *speA*, *entB*, and *entC1* reveal large regions of similarity, particularly in the 3′ portions of the genes (8). An alignment of the 3′ portions of *speA* and *speC* also revealed highly homologous stretches of nucleotides (Fig. 2). The conservation of nucleotide sequences supports the proposition that the toxin genes arose from a common ancestral gene. At present, evolutionary relatedness cannot be established for TSST-1. It is possible that the gene for TSST-1 (*tst*) does not have a common evolutionary link with the genes for the other toxins and that its functional relatedness is due to convergent evolution. We believe, however, that it is more likely that evolutionary relatedness will be established between *tst* and the other toxin genes as more sequences become known.

The dissemination of the pyrogenic toxin genes across genera cannot readily be explained but may be related to the presence of toxin genes on mobile elements. The genes encoding SPE A and Ent A exist on bacteriophages (3, 12). We recently found that the gene specifying SPE C is also phage encoded. The gene for Ent B has been reported to be plasmid associated (1). It is possible that at one time a toxin gene crossed the genus boundary on one of these transmissible elements, but evidence of natural transfer of these genes between *S. pyogenes* and *Staphylococcus aureus* has not been reported. Considering the high degree of sequence divergence which has occurred among the pyrogenic toxin genes, evidence of transgeneric transfer (such as homologous flanking DNA) may be obscured by sequence divergence. An alternative hypothesis is that *S. pyogenes* and *Staphylococcus aureus* arose from a common ancestor organism containing the ancestral toxin gene or genes. In support of this idea, these genera share a large number of characteristics: both are gram-positive pathogenic cocci, both are approximately 70% A+T rich, and both produce several analogous gene products.

Although questions of toxin gene dissemination may not ever be fully elucidated, further study of the mobile elements associated with the toxin genes may enable some clarification of the process. Insight into toxin gene evolution may be gained by the isolation and sequencing of more toxin (and

```
                   310       320       330       340       350
speC:    ATTATTTTATATTCTTAATTCTCACACCGGTGAGTACATCTATGGAGGAATTACGCC
          :  : ::    :   ::: :  : :   :::  :  ::::: :::::
speA: TCTCTGTTATTTATGTGAAAATGCAGAAAGGAGTGCATGTATCTACGGAGGGG-------
           350       360       370       380       390

         360       370       380       390       400       410
      TGCTCAAAATAATAAAGTAAATCATAAATTATTGGGAAATCTATTTATTTCGGGAGAATC
               ::: :::  :       ::: :::  :::::    :::    :  ::
      ---------TAACAAATCATGAAGGGAATCATTTAGAAATTC--TTAAAAGATAGTCGT
                 400       410       420       430       440

         420       430       440       450
      TCAACAGAACTTAAATAACAAGATTATTCTAGAAAAGGATATC-----------------
      : ::       :  ::   :     :  :::  :   :::::
      TAAAGTATCAATCGATGGTATCCAAAGCCTATCATTTGATATTGAAACAAATAAAAAAAT
         450       460       470       480       490       500

        460       470       480       490       500       510
      -GTAACTTTCCAGGAAATTGACTTTAAAATCAGAAAATACCTTATGGATAATTATAAAAT
       ::::::   :: ::: : :::: :::: : :::::::: ::::  :::::: :  :: :
      GGTAACTGCTCAAGAATTAGACTATAAAGTTAGAAAATATCTTACAGATAATAAGCAACT
         510       520       530       540       550       560

        520       530       540       550       560       570
      TTATGACGCTACTTCTCCTTATGTAAGCGGCAGAATCGAAATTGGCACAAAAGATGGGAA
       :::     :    :: :: :           ::     ::    :: :    :    ::
      ATATACTAATGGACCTTCTAAATATGAAACTGGATATATAAAGTTCATACCTAAGAATAA
         570       580       590       600       610       620

        580       590       600       610       620       630
      ACATGAGCAAATAGACTTATTTGACTCACCAAATGAAGGGACTAGATCAGATATTTTTGC
      : :                ::: : :   :   :::   :::  :::  :   : ::
      AGAAAGTTTTTGGTTTGATTTTTTCCCTGAACCAGAATTTACTCAATCTAAATATCTTAT
         630       640       650       660       670       680

        640       650       660       670       680       690
      AAAATATAAAGATAATAGAATTATCAATATGAAGAACTTTAGTCATTTCGATATTTATCT
       : :::::::::::::  ::   :  :    :: :          :   :
      GATATATAAAGATAATGAAACGCTTGACTCAAACACAAGCCAAATTGAAGTCTACCTAAC
         690       700       710       720       730       740

        700
      TGAAAAATAA
          :: :::
      AACCAAGTAA
         750
```

FIG. 2. Nucleotide sequence homology between the 3′ halves of the *speC*- and *speA*-coding sequences. Numbering is in reference to the ATG start codons of each gene. The genes were aligned with a computer program based on the algorithm of Wilbur and Lipman (21). Matched bases are indicated by colons. Gaps introduced by the alignment program are represented by dashed lines.

toxinlike) genes, allowing firmer ancestral links to be formed in this gene family.

This work was supported by Public Health Service grant HL36611 from the National Heart, Lung, and Blood Institute. S.C.G. was supported by Public Health Service training grant 5T32 CA09138 from the National Cancer Institute.

LITERATURE CITED

1. **Altboum, Z., I. Hertman, and S. Sarid.** 1985. Penicillinase plasmid-linked genetic determinants for enterotoxins B and C_1 production in *Staphylococcus aureus*. Infect. Immun. **47:**514–521.
2. **Altschul, S. F., and B. W. Erickson.** 1986. Optimal sequence alignment using affine gap costs. Bull. Math. Biol. **48:**603–616.
3. **Betley, M. J., and J. J. Mekalanos.** 1985. Staphylococcal enterotoxin A is encoded by phage. Science **229:**185–187.
4. **Betley, M. J., and J. J. Mekalanos.** 1988. Nucleotide sequence of the type A staphylococcal enterotoxin gene. J. Bacteriol. **170:** 34–41.
5. **Blomster-Hautamaa, D. A., B. N. Kreiswirth, J. S. Kornblum, R. P. Novick, and P. M. Schlievert.** 1986. The nucleotide and partial amino acid sequence of toxic shock syndrome toxin-1. J. Biol. Chem. **261:**15783–15786.
6. **Bohach, G. A., C. J. Hovde, J. P. Handley, and P. M. Schlievert.** 1988. Cross-neutralization of staphylococcal and streptococcal pyrogenic toxins by monoclonal and polyclonal antibodies. Infect. Immun. **56:**400–404.
7. **Bohach, G. A., and P. M. Schlievert.** 1987. Expression of staphylococcal enterotoxin C_1 in *Escherichia coli*. Infect. Immun. **55:**428–432.
8. **Bohach, G. A., and P. M. Schlievert.** 1987. Nucleotide sequence of the staphylococcal enterotoxin C1 gene and relatedness to other pyrogenic toxins. Mol. Gen. Genet. **209:**15–20.
9. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2:**95–113.
10. **Cone, L. A., D. R. Woodard, P. M. Schlievert, and G. S. Tomory.** 1987. Clinical and bacteriologic observations of a toxic shock-like syndrome due to *Streptococcus pyogenes*. N. Engl. J. Med. **317:**146–149.
11. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S DNA. Plasmid **13:**31–40.

11a. **Goshorn, S. C., G. A. Bohach, and P. M. Schlievert.** 1988. Cloning and characterization of the gene, *speC*, for pyrogenic exotoxin type C from *Streptococcus pyogenes*. Mol. Gen. Genet. **212:**66–70.

12. **Johnson, L. P., and P. M. Schlievert.** 1984. Group A streptococcal phage T12 carries the structural gene for pyrogenic exotoxin type A. Mol. Gen. Genet. **194:**52–56.
13. **Jones, C. L., and S. A. Khan.** 1986. Nucleotide sequence of the enterotoxin B gene from *Staphylococcus aureus*. J. Bacteriol. **166:**29–33.
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 is not detectably transmitted by a prophage. Nature (London) **305:**709–712.
15. **Messing, J.** 1979. A multi-purpose cloning system based on the single-stranded DNA bacteriophage M13. Recombinant DNA Technical Bulletin, National Institutes of Health publication no. 79–99, p. 43–48. National Institutes of Health, Bethesda, Md.
16. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101:**20–78.
17. **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 *Staphylococcus aureus*. Proc. Natl. Acad. Sci. USA **82:**5850–5854.
18. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13:**319–353.
19. **Schlievert, P. M., K. M. Bettin, and D. W. Watson.** 1979. Production of pyrogenic exotoxin by groups of streptococci: association with group A. J. Infect. Dis. **140:**676–681.
20. **Weeks, C. R., and J. J. Ferretti.** 1986. Nucleotide sequence of the type A streptococcal exotoxin (erythrogenic toxin) gene from *Streptococcus pyogenes* bacteriophage T12. Infect. Immun. **52:**144–150.
21. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80:**726–730.