Mol Gen Genet (1987) 209:15–20

MGG

Nucleotide sequence of the staphylococcal enterotoxin C1 gene and relatedness to other pyrogenic toxins

Gregory A. Bohach and Patrick M. Schlievert
University of Minnesota Medical School, Department of Microbiology, 420 Delaware St. SE, Minneapolis, MN 55455, USA

Summary. The nucleotide sequence for the structural gene *entC1* encoding staphylococcal enterotoxin C1 was determined. The gene contained 801 bp and coded for a protein of 266 amino acids. Of these, 27 comprised the signal peptide. Cleavage of the signal peptide resulted in a mature protein with 239 amino acids and a calculated molecular weight of 27496. The nucleotide sequence of *entC1* shared considerable homology (74% and 59%, respectively) with genes encoding enterotoxin B and streptococcal pyrogenic exotoxin A. A similar degree of amino acid homology was observed after alignment of the respective proteins. Thus, certain regions of these three toxin molecules possess structural similarities that may be responsible for shared biological properties.

Key words: *Staphylococcus aureus* – *Streptococcus pyogenes* – Enterotoxin C1 – Pyrogenic toxins – Homology

Introduction

Staphylococcal enterotoxin C1 (Ent C1) belongs to a family of pyrogenic toxins (PTs) produced by *Staphylococcus aureus* and *Streptococcus pyogenes*. Included in this family of toxins are the staphylococcal enterotoxins A–E, pyrogenic exotoxins A and B, toxic shock syndrome toxin-1 (TSST-1), and streptococcal pyrogenic exotoxins A–C. PTs share numerous biological properties including lymphocyte mitogenicity, immunosuppression, pyrogenicity and enhancement of susceptibility to lethal endotoxin shock (Schlievert et al. 1979; Smith and Johnson 1975; Poindexter and Schlievert 1985; Brunson and Watson 1974; Sugiyama et al. 1964). Staphylococcal enterotoxins (Ents) have the additional ability, not shared with other PTs, to induce emesis and diarrhea upon ingestion, and are responsible for staphylococcal food poisoning (Bergdoll et al. 1973; Dolman and Wilson 1938). Likewise, streptococcal pyrogenic exotoxins are unique in their abilities to induce both heart damage and cause scarlet fever (Schwab et al. 1955).

Genetic control of PT production is diverse. The genes for enterotoxin A (Ent A) and streptococcal pyrogenic exotoxin A (SPE A) are contained on bacteriophage (Johnson et al. 1980; Betley and Mekalanos 1985). In contrast, the gene encoding TSST-1 is located on the bacterial chromosome (Kreiswirth et al. 1983). Evidence has been provided for both chromosomal and plasmid determination of Ent C1 production (Altboum et al. 1985; Betley and Bergdoll 1981). Recently, the gene encoding staphylococcal Ent C1, *entC1*, was cloned into *Escherichia coli* from the bacterial chromosome (Bohach and Schlievert 1987). Regulation of other PT genes is not understood.

The primary structure of Ent C1 has been reported previously (Schmidt and Spero 1983). The protein contains 239 amino acids with a calculated molecular weight of 27500 and shares significant sequence homology with Ent B (Huang and Bergdoll 1970). Jones and Khan (1986) have determined the nucleotide sequence of the Ent B gene, *entB*, from which they have derived the amino acid sequence for the toxin. Their results are similar but not identical to those obtained by direct amino acid sequencing. Recently, the nucleotide sequence of the gene for SPE A production was reported (Weeks and Ferretti 1986; Johnson et al. 1986). The primary structure of SPE A predicted from nucleotide sequence data shares considerable homology with Ent B and Ent C1. The nucleotide and amino acid sequences of *tst* and TSST-1, respectively, are also known (Blomster-Hautamaa et al. 1986), and in spite of similarities in biological activities, TSST-1 does not have amino acid sequence homology with SPE A, Ent B or Ent C1.

This present study was undertaken to determine the nucleotide sequence of *entC1* so a comparison could be made with other PTs.

Materials and methods

Bacterial strains and growth conditions. E. coli strain JM83 (*ara, Δlac-pro, strA, thi, φ*80d*lacZΔM15*) (Messing 1979) was used as a recipient of cloned DNA in plasmid vectors. Growth and selection of recombinants were described previously (Bohach and Schlievert 1987). Bacteriophage M13 clones were propagated in *E. coli* JM101 (*Δlac-pro, supE, thi,* F′, *traD36, proAB, lacI*^{q}*ZΔM15* under standard conditions (Messing 1983).

Plasmid and phage vectors. Construction of pMIN114 was described previously (Bohach and Schlievert 1987). This recombinant plasmid contains a 1.5 kb *Hin*dIII-*Cla*I insert in pUC13 (Vieira and Messing 1982) harboring the entire *entC1* gene and codes for a protein identical to staphylococcal derived Ent C1. Deletion derivatives of *entC1* were cloned into the replicative form of M13mp8 or M13mp9 (Messing 1983) for sequence analysis.

Offprint requests to: G. Bohach

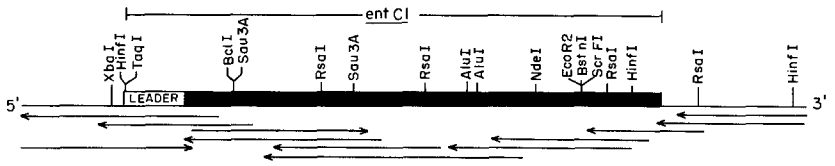

Fig. 1. Sequencing strategy and partial restriction map for *entC1*. A total of 1095 nucleotides was sequenced. The open reading frame spanning the *entC1* structural gene is located between nucleotides 1 and 801. *Arrows* indicate the direction and length of sequence determined from each Bal31-derived M13 clone

Preparation of entC1 deletions. Approximately 1 μg of *Hin*dIII- or *Bam*H1-linearized pMIN114 was digested with 1 unit of exonuclease Bal31 (New England Biolabs, Beverly, Mass). Aliquots were mixed with 40 mM EGTA and ethanol precipitated. End repair of digested DNA fragments was accomplished with Klenow fragment (International Biotechnologies, Inc. New Haven, Conn). Synthetic *Hin*dIII or *Bam*H1 linkers were added with T4 DNA ligase (Boehringer Mannheim Biochemicals, Indianapolis, Ind). Excess linkers were removed by digestion with the appropriate restriction enzymes and agarose gel electrophoresis. Plasmids were purified from the agarose, religated and transformed into *E. coli* JM83 by the method of Kushner (1978).

M13 cloning. Deletion derivatives of pMIN114 propagated in *E. coli* JM83 were harvested by the method of Holmes and Quigley (1981) and digested with *Hin*dIII and *Bam*H1. The inserts were isolated by agarose gel electrophoresis and ligated to doubly digested M13 RF DNA. Transformation of *E. coli* JM101, plating and selection of recombinant phage were done according to the methods of Messing (1983).

Nucleotide sequence determinations. Sequencing of DNA from selected M13 clones was performed using the dideoxy chain termination method of Sanger et al. (1977). All reagents were obtained from International Biotechnologies, Inc. or NEN Research Products (Boston, Mass) and used according to the manufacturers' specifications. Radiolabeled DNA fragments were separated by high resolution electrophoresis in polyacrylamide gels (6% and 8%) under denaturing conditions and visualized by autoradiography.

Amino acid sequencing. The N-terminal amino acid sequence of SPE A was determined. The toxin was purified from cultures of *S. pyogenes* strain NY-5 grown in dialyzable beef heart medium (Watson 1960). Procedures for purification of SPE A were described previously (Johnson and Schlievert 1984). Amino acid sequence analysis was performed using a Model 470A gas-phase protein sequencer interfaced to a Model 120A PTH amino acid analyzer (Applied Biosystems, Foster City, Calif). This system employs the automated Edman degradation technique of Hewick et al. (1981).

Computer analysis. Nucleotide sequences were analyzed using an IBM Personal Computer. Sequence alignments and searches were done using the algorithms of Wilbur and Lipman (1983).

Results

Nucleotide sequence of entC1

Overlapping deletions spanning the *entC1* gene were produced. The DNA fragments were cloned into M13 in both orientations for sequence analysis according to the strategy outlined in Fig. 1.

The nucleotide sequence of *entC1* is shown in Fig. 2. The putative structural gene contained an 801 bp open reading frame beginning at an ATG start codon and ending with a TAA stop codon. A possible Shine-Dalgarno sequence (AAGGAG) was located seven nucleotides upstream from the proposed ATG start codon.

Regulatory regions of entC1

The most likely −10 and −35 promoter sequences are underlined in Fig. 2. Their selection was based on their relative distances upstream from the 5′ end of *entC1* and similarities to previously described RNA polymerase recognition sites (Rosenberg and Court 1979). Palindromic sequences were located downstream from the TAA stop codon.

Primary structure of Ent C1

Figure 2 shows the predicted amino acid sequence of Ent C1 based upon nucleotide sequence data. The initial gene product contained 266 amino acids and had a molecular weight of 30511. The sequence of amino acids after alanine at residue 27 were identical to those previously identified as the N-terminus of Ent C1 (Schmidt and Spero 1983; Bohach and Schlievert 1987). Therefore, the first 27 amino acids apparently comprised the signal peptide. Cleavage of the signal peptide would produce the mature toxin containing 239 amino acids and a calculated molecular weight of 27496. The deduced amino acid sequence was similar to the sequence previously reported by Schmidt and Spero 1983. A discrepancy of four amino acids was found (Fig. 2). The nucleotide sequences in these regions were confirmed either by sequencing both strands or by sequencing numerous overlapping clones.

Homology studies

One goal of this study was to compare the nucleotide sequence of *entC1* with gene sequences from other pyrogenic toxins. These included *speA* (Weeks and Ferretti 1986; Johnson et al. 1986), *entB* (Jones and Khan 1986) and *tst* (Blomster-Hautamaa et al. 1986).

Since a number of discrepancies were noted between the two previously reported sequences for *speA* in the 5′ terminal region, the first 29 amino acids of SPE A were sequenced. Our results agreed with the sequence reported by Weeks and Ferretti (1986) (results not shown) which was subsequently used in sequence comparisons in the study.

Figure 3 shows a computer alignment of *speA*, *entB* and *entC1*. The genes *entB* and *entC1* each contain 801 bp. Of these, 589 (74%) were matched by computer alignment.

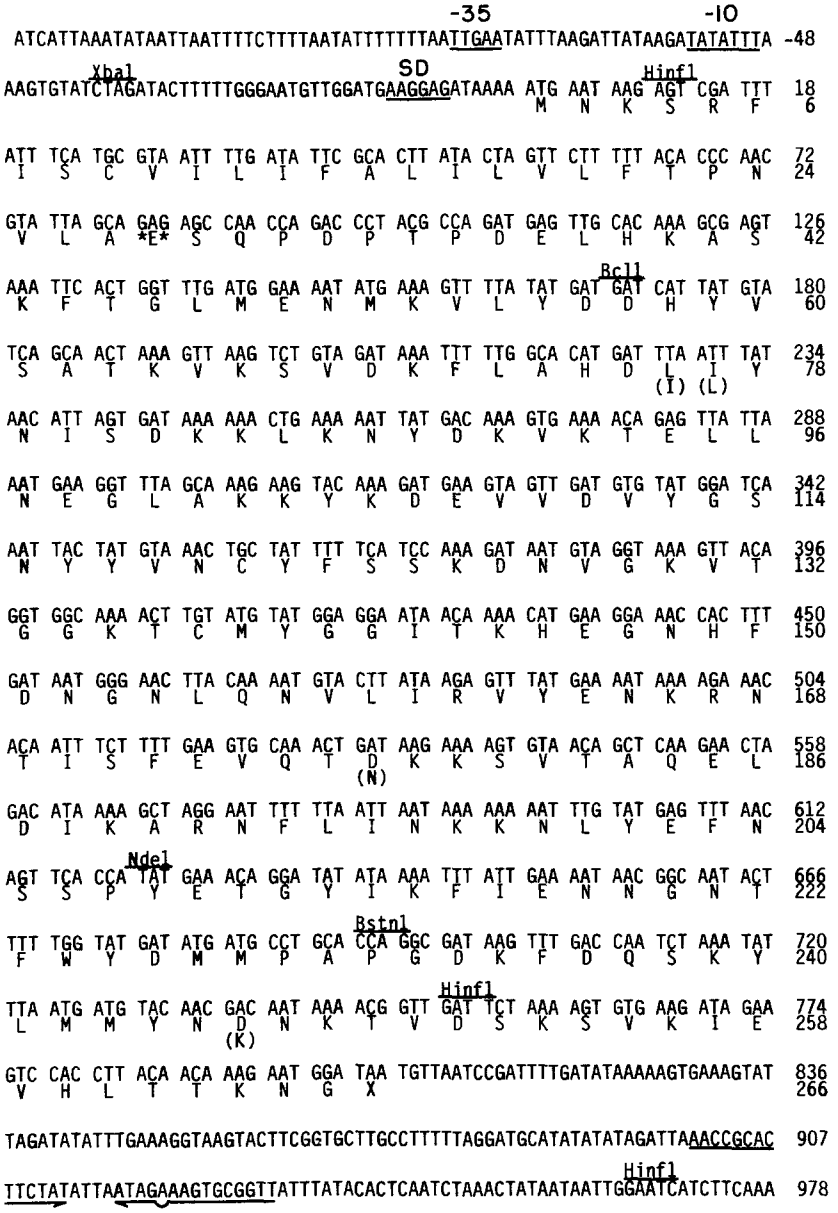

Fig. 2. Nucleotide and deduced amino acid sequences of *entC1* and enterotoxin C1 (Ent C1), respectively. Numbering is in reference to the ATG start codon and nucleotides are listed in a 5′–3′ direction. Possible Shine-Dalgarno and promoter sequences are *underlined. Arrows* indicate a downstream palindromic sequence. Important restriction endonuclease sites are shown. Amino acids 1–27 comprise the signal peptide. Residue 28 (*E*) is the N-terminal amino acid of mature Ent C1 (Schmidt and Spero 1983). Amino acids in *parentheses* are those reported by Schmidt and Spero (1983) that differ from the sequence predicted from nucleotide data

Stretches of homologous regions were present throughout the gene, especially at the 5′ and 3′ terminal regions. Of the 753 bases in *speA*, 441 (59%) matched with *entC1*. The most significant divergence was near the center of the respective genes. No significant homology was noted with TSST-1.

Several differences were noted between previously reported amino acid sequences of Ent B and Ent C1 (Huang and Bergdoll 1970; Schmidt and Spero 1983) and sequences deduced for these proteins based upon nucleotide sequence data (this study; Jones and Khan 1986). In addition, previous comparison of primary structures of several PTs (Weeks and Ferretti 1986; Johnson et al. 1986; Schmidt and Spero 1983) did not include the signal peptides of Ent B and Ent C1. We therefore compared the predicted amino acid sequences of the pyrogenic toxins. Computer alignments of SPE A, Ent C1 and Ent B are shown in Fig. 4. TSST-1 was not included because no significant homology was observed. Of 266 amino acids in each of Ent B and Ent C1, 184 (69%) were matched by computer alignment. The *speA* gene product contains 251 amino acids. Of these, 122 (49%) were homologous with Ent C1.

Figure 5 shows a comparison of nucleotide sequences upstream from the 5′ ends of *entB* and *entC1*. The two DNA sequences shared complete homology in their proposed −10 and −35 promoter regions and differed by only one base in the stretch of nucleotides between them. The upstream *Xba*I recognition site was also conserved in both genes.

Discussion

We have sequenced 1095 nucleotides encompassing *entC1*, the gene encoding staphylococcal Ent C1, and its 5′ and 3′ flanking regions. An open reading frame of 801 nucleotides, beginning with ATG was detected. This is presumed to be the correct start site for *entC1* since it is preceded with appropriate spacing by typical regulatory sequences. Seven bases upstream from the ATG start codon was a strong Shine-Dalgarno sequence (AAGGAG) that had sig-

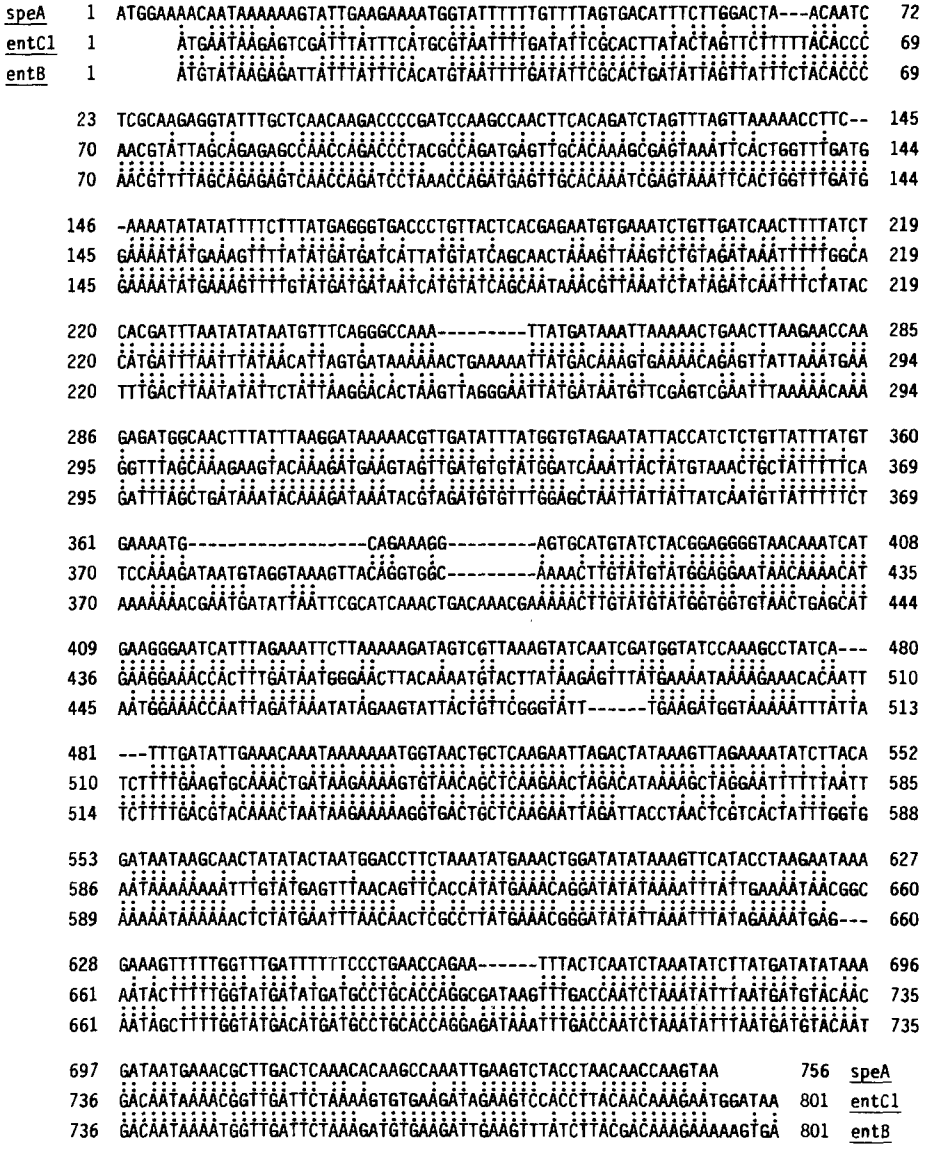

Fig. 3. Nucleotide sequence homology between *entC1, entB* and *speA*. The three structural genes were aligned with a computer program based on the algorithm of Wilbur and Lipman (1983). Matched nucleic acids (:) and gaps (----) introduced for optimal alignment are indicated. Sequence data for *entB* and *speA* were derived from reports of Jones and Khan (1986) and Weeks and Ferretti (1986), respectively

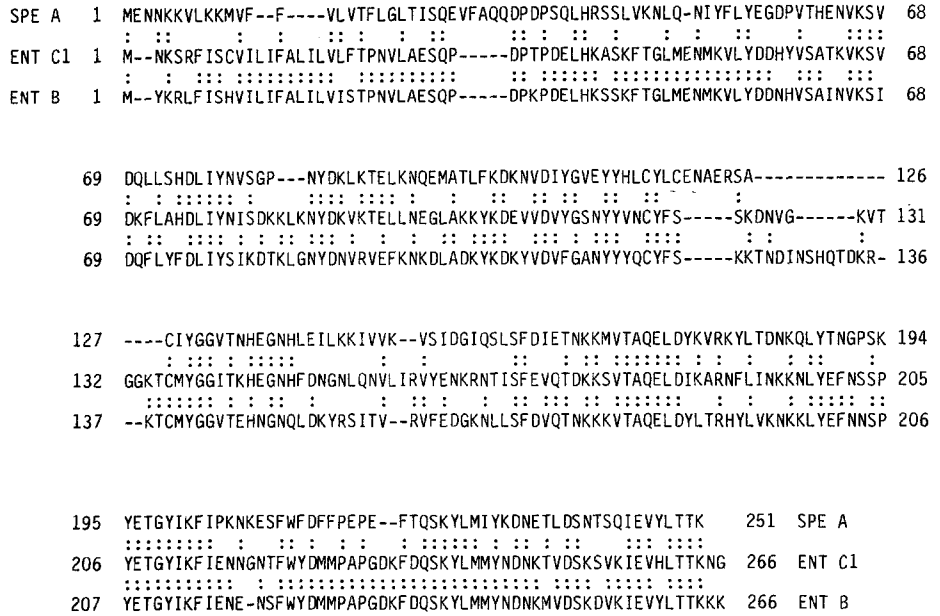

Fig. 4. Amino acid sequence homology between enterotoxin C1 (Ent C1), enterotoxin B (Ent B) and streptococcal pyrogenic exotoxin A (SPE A). Alignment was done using a computer program based on the algorithm of Wilbur and Lipman (1983). The primary structures shown are those derived from the nucleotide sequence of their respective structural genes. Matched amino acids (:) and gaps (----) introduced for optimal alignment are indicated. Data for Ent B and SPE A were derived from reports by Jones and Khan (1986) and Weeks and Ferretti (1986), respectively

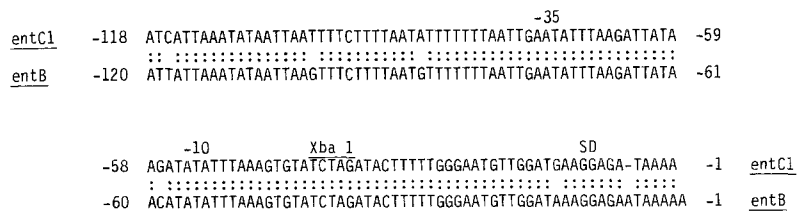

Fig. 5. Nucleotide sequence homology between regions upstream from the 5′ end of *entC1* and *entB*. Matched nucleic acids (:) and gaps (−) introduced for optimal alignment are indicated. The most likely −10 and −35 promoter sequences are shown. Sequence data for *entB* were obtained from Jones and Khan (1986)

nificant homology with the 3′ terminus of 16 S rRNA from *E. coli* and *Bacillus subtilis* (Shine and Dalgarno 1974; McLaughlin et al. 1981). Putative −10 and −35 promoter sequences were also observed upstream from *entC1*. These contain the required highly conserved nucleotides found in other promoters recognized by *E. coli* RNA polymerase (Rosenberg and Court 1979). A palindromic sequence located downstream from *entC1* may be involved in transcription termination (Rosenberg and Court 1979).

The transcribed protein predicted from the nucleotide sequence of *entC1* contained 266 amino acids and had a calculated molecular weight of 30511. Mature Ent C1, without the signal peptide, contained 239 amino acids and had a molecular weight of 27496. These results are similar to those reported previously based upon direct amino acid sequence analysis (Schmidt and Spero 1983). Only four discrepancies were found. Two of these resulted from an inversion of adjacent leucine and isoleucine residues, and the other two involved aspartic acid. Difficulties encountered in detection of aspartic acid during amino acid sequencing have been previously discussed by Schmidt and Spero (1983).

Significant nucleotide and amino acid sequence homology was observed between Ent C1, Ent B and SPE A. As expected Ent C1 and Ent B were more similar to each other than to SPE A. Although the three toxins share several biological properties, only the enterotoxins cause vomiting and diarrhea after ingestion. Unlike the enterotoxins, SPE A can predispose the host to heart damage. One can speculate that homologous regions shared by all three toxins are required for their common biological properties. In contrast, those regions in which the molecules diverge may determine their unique toxicities.

Interestingly, TSST-1 is endowed with biological properties similar to other PTs, yet it has no significant homology with Ent B, Ent C1 or SPE A. It remains to be determined if other toxins in this family are related structurally to any of these four toxins.

The similarities between the regions upstream from *entB* and *entC1* are noteworthy. Jones and Khan (1986) have found that *entB* is not expressed in *E. coli* unless transcribed from a strong *E. coli* promoter. In contrast, *entC1* is expressed efficiently by *E. coli* (Bohach and Schlievert 1987). The proposed −10 and −35 promoter sequences for both genes are identical and thus do not appear to be responsible for these observed differences.

It is interesting that staphylococcal and streptococcal PTs share sequence homology and biological properties. It is probable that these genes evolved from a common ancestor. A similar mechanism has been proposed to explain sequence homology between staphylokinase and streptokinase (Sako and Tsuchida 1983). The degree of homology between *speA, entC1* and *entB* suggests that their divergence was more recent than that predicted for staphylokinase and streptokinase.

Acknowledgements. This work was sponsored by a grant from the Kimberly-Clark Corporation. G.A.B. was supported by a fellowship from the American Heart Association, Minnesota Affiliate. Amino acid sequencing was performed by Dr. Robert Wohlhueter at the University of Minnesota Microchemical Facility. The authors thank Tim Leonard for art and photographic work and Renee Kayon for typing the manuscript.

References

Altboum Z, Hertman I, Sarid S (1985) Penicillinase plasmid-linked genetic determinants for enterotoxins B and C1 production in *Staphylococcus aureus*. Infect Immun 47:514–521

Bergdoll MS, Robbins RN, Weiss K, Borja CR, Huang IY, Chu FS (1973) The staphylococcal enterotoxins: similarities. In: Jeljaszewicz J, Hryniewicz W (eds) Contributions to microbiology and immunology, vol 1. S. Karger, Basel, p 390

Betley MJ, Bergdoll MS (1981) Staphylococcal enterotoxin type C genes not associated with extrachromosomal DNA. Abstracts of the annual meeting of the American Society for Microbiology, ASM, Washington DC D38, p 49

Betley MJ, Mekalanos JM (1985) Staphylococcal enterotoxin A is encoded by phage. Science 229:185–187

Blomster-Hautamaa DB, Kreiswirth BN, Kornblum JS, Novick RP, Schlievert PM (1986) The nucleotide and partial amino acid sequence of toxic shock syndrome toxin-1. J Biol Chem 261:15783–15786

Bohach GA, Schlievert PM (1987) Expression of staphyococcal enterotoxin C1 in *Escherichia coli*. Infect Immun 55:428–432

Brunson KW, Watson DW (1974) Pyrogenic specificity of streptococcal exotoxins, staphylococcal enterotoxin and gram-negative endotoxin. Infect Immun 10:347–351

Dolman CE, Wilson RJ (1938) Experiments with staphylococcal enterotoxin. J Immunol 35:13–30

Hewick RM, Hunkapiller MW, Hood LE, Dreyer WJ (1981) A gas-liquid solid phase peptide and protein sequenator. J Biol Chem 256:7990–7997

Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem 114:193–197

Huang IY, Bergdoll MS (1970) The primary structure of staphylococcal enterotoxin B. III. The cyanogen bromide peptides of reduced and amino ethylated enterotoxin B, and the complete amino acid sequence. J Biol Chem 245:3518–3525

Johnson LP, Schlievert PM (1984) Group A streptococcal phage T12 carries the structural gene for pyrogenic exotoxin type A. Mol Gen Genet 194:52–56

Johnson LP, Schlievert PM, Watson DW (1980) Transfer of group A streptococcal pyrogenic exotoxin production to nontoxigenic strains by lysogenic conversion. Infect Immun 28:254–257

Johnson LP, L'Italien JJ, Schlievert PM (1986) Streptococcal pyrogenic exotoxin type A (scarlet fever toxin) is related to *Staphylococcus aureus* enterotoxin B. Mol Gen Genet 203:354–356

Jones CJ, Khan SA (1986) Nucleotide sequence of the enterotoxin B gene from *Staphylococcus aureus*. J Bacteriol 166:29–33

Kreiswirth BN, Lofdahl S, Betley MJ, O'Reilly M, Schlievert PM, Bergdoll MS, Novick RP (1983) The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. Nature 305:709–712

Kushner SR (1978) An improved method for transformation of *Escherichia coli* with ColE1 derived plasmids. In: Boyer HB, Nicosia S (eds) Genetic engineering. Elsevier/North-Holland, Amsterdam, p 17

Mclaughlin JR, Murray CL, Rabinowitz JC (1981) Unique features in the ribosome binding site sequence of the gram-positive of *Staphylococcus aureus* β-lactamase gene. J Biol Chem 256:11283–11291

Messing J (1979) A multi-purpose cloning system based on the single-stranded DNA bacteriophage M13. Recombinant DNA Technical Bulletin, NIH Publication No 79–99, 2: No. 2, 43–48

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Poindexter NJ, Schlievert PM (1985) The biochemical and immunological properties of toxic-shock syndrome toxin-1 (TSST-1) and association with TSS. J Toxicol Toxin Rev 4:1–39

Rosenberg M, Court D (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Annu Rev Genet 13:319–353

Sako T, Tsuchida N (1983) Nucleotide sequence of the staphylokinase gene from *Staphylococcus aureus*. Nucleic Acids Res 11:7679–7693

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schlievert PM, Schoettle DJ, Watson DW (1979) Purification and physicochemical and biochemical characterization of a staphylococcal pyrogenic exotoxin. Infect Immun 23:609–617

Schmidt JJ, Spero L (1983) The complete amino acid sequence of staphylococcal enterotoxin C1. J Biol Chem 258:6300–6306

Schwab JH, Watson DW, Cromartie WJ (1955) Further studies of group A streptococcal factors with lethal and cardiotoxic properties. J Infect Dis 96:14–18

Shine J, Dalgarno L (1974) The 3′-terminal sequence of *E. coli* 16 S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342–1346

Smith BG, Johnson HM (1975) The effect of staphylococcal enterotoxins on the primary *in vitro* immune response. J Immunol 115:575–578

Sugiyama H, McKissic EM, Bergdoll MS, Heller B (1964) Enhancement of bacterial endotoxin lethality by staphylococcal enterotoxin. J Infect Dis 114:111–118

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268

Watson DW (1960) Host-parasite factors in group A streptococcal infections. Pyrogenic and other effects of immunologic distinct exotoxins related to scarlet fever toxins. J Exp Med 111:255–284

Weeks CR, Ferretti JJ (1986) Nucleotide sequence of the type A streptococcal exotoxin (erythrogenic toxin) gene from *Streptococcus pyogenes* bacteriophage T12. Infect Immun 52:144–150

Wilbur WJ, Lipman DJ (1983) Rapid similarity searches of nucleic acid and protein data banks. Proc Natl Acad Sci USA 80:726–730

Communicated by G.W. Lengeler

Received February 20, 1987