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 exotoxins 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 C1, 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 C1 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 C1 (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 lacPZ ΔM15 [9]), recombinant phage 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 dyeoxy 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 by 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 C1 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 C1, 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.
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 C1 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 C1 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 (tsst) 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 likely that evolutionary relatedness will be established between tsst 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

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

<table>
<thead>
<tr>
<th>Probes</th>
<th>Monte Carlo score for compared sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPE A</td>
</tr>
<tr>
<td>SPE A</td>
<td>11.0</td>
</tr>
<tr>
<td>SPE C</td>
<td>10.2</td>
</tr>
<tr>
<td>Ent A</td>
<td>12.6</td>
</tr>
<tr>
<td>Ent B</td>
<td>31.7</td>
</tr>
<tr>
<td>Ent C1</td>
<td>35.8</td>
</tr>
<tr>
<td>TSST-1</td>
<td>5.43</td>
</tr>
</tbody>
</table>
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 sequences 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 ST32 CA09138 from the National Cancer Institute.

LITERATURE CITED


