Volume 11 Number 22 1983 Nucleic Acids Research

Nucleotide sequence of the staphylokinase gene from *Staphylococcus aureus*

Tomoyuki Sako and Nobuo Tsuchida

Yakult Central Institute for Microbiological Research, 1796 Yaho, Kunitachi, Tokyo 186, Japan

Received 11 October 1983; Accepted 31 October 1983

ABSTRACT

We have determined the entire nucleotide sequence of a 1.4-kilobase segment containing the staphylokinase gene, sak, molecularly cloned from the bacteriophage SØ-C genome of Staphylococcus aureus. The probable coding region is 489 base pairs long and these base pairs are translated into a polypeptide of 163 amino acid residues (M_r= 18,490) with a presumed signal sequence of 27 amino acid residues at the NH_2-terminal end. In regions adjacent to the sak structural gene a possible promotor sequence and three possible terminator sequences for transcription were found about 100 base pairs upstream from the initiation codon and about 300, 400, and 500 base pairs downstream from the termination codon, respectively; they are active in an in vitro transcription system using Escherichia coli RNA polymerase. The immunoactive 18,500-dalton and 15,500-dalton proteins corresponding to a precursor form before secretion and a mature form after secretion of the sak gene products, respectively, were identified by the E. coli maxicell system.

INTRODUCTION

Most secretory proteins are synthesized as a larger precursor having a signal sequence at their NH_2-terminal end (for review, see ref. 1, 2). The signal sequence of prokaryotes and eukaryotes have a common polar structure consisting of a stretch of hydrophobic amino acids following an NH_2-terminal basic segment (2, 3, 4). Secretory proteins derived from Gram-positive bacteria also have signal sequences, some of which have been demonstrated to function in the Gram-negative organism Escherichia coli as well (5, 6, 7, 8). However, several differences in the structure of signal sequences and the mechanism of protein secretion between Gram-positive and Gram-negative bacteria have been reported . The signal sequences of Gram-positive bacteria are relatively long (4, 9, 10), sequential processing of the signal sequence attended by the translocation of the protein is detected only in Gram-positive bacteria (5, 11), and the cleavage site of the signal sequence of the protein which can be expressed and secreted is apparently different in the two kinds of organisms (7, 12). A relatively large amount of data is available for understanding the mechanism of protein secretion in E. coli. On the other

hand attempts to clarify the mechanism in Gram-positive organisms have just been begun.

Recently we cloned the structural gene of staphylokinase, which is an extracellular protein of Staphylococcus aureus and is one of the plasminogen activators, from the temperate bacteriophage SØ-C of S. aureus (13). The gene, sak, was stably expressed in E. coli and the active product with a molecular weight and an antigenicity indistinguishable from those of authentic staphylokinase was efficiently secreted into the periplasmic space. Thus it can be a useful tool to investigate the mechanism of protein secretion in Gram-positive as well as in Gram-negative organisms. In this study we have determined the nucleotide sequence of the sak gene and found that the gene codes for a polypeptide of 163 amino acid residues with a presumed signal sequence of 27 amino acid residues at the NH_2-terminal end. The control regions for transcription and translation of the sak gene have also been identified in the adjacent regions. In addition we compared the amino acid sequence of the sak gene product with that of streptokinase, which is produced by various strains of streptococci and activates plasminogen non-proteolytically like staphylokinase.

MATERIALS AND METHODS

Bacterial strains and plasmids

E. coli K12 strains WA802 (14) and N1790 (15) were used as bacterial hosts. N1790 was the generous gift of Dr. H. Ogawa of Osaka University. The plasmid vector used was pBR322 (16).

DNA technology

Restriction endonucleases, T4 DNA ligase, and T4 DNA polymerase were obtained from commercial sources and used as described previously (13). The methods for purification of plasmid DNAs, agarose gel electrophoresis and transformation of E. coli cells were described previously (13).

In vitro construction of recombinant plasmids

Recombinant plasmids constructed as described below as well as pSAK361 and pSAK-HP2 (13) are schematically presented in Fig. 1.

a) pTS365: The 4.9-kb DNA fragment derived from phage SØ-C (17) was isolated by agarose gel electrophoresis after cleavage of pSAK361 (13) DNA with HindIII followed by electroelution. The DNA fragment thus isolated was cleaved with PstI and HaeIII to produce the 1.4-kb HaeIII-PstI fragment, and ligated to pBR322 DNA which had been cleaved with PstI and HincII. The resulting plasmid pTS365 carried the 1.4-kb HaeIII-PstI fragment in place of

the HincII-PstI segment in the bla gene.

b) pTS372: The 4.9-kb DNA fragment was cleaved with AvaII and the resulting 2.0-kb DNA fragment between the $HindIII_1$ and AvaII sites was purified electrophoretically. Both staggered ends of the fragment were repaired with T4 DNA polymerase in the presence of four deoxyribonucleoside triphosphates and ligated to pBR322 DNA which had been cleaved with HincII. The recombinant plasmids thus obtained unexpectedly had vast deletions. One of them, pTS372, carried the 2.0-kb $HindIII_1$-AvaII fragment but had a deletion in pBR322 of a region between a point near the HincII site in the tet gene and a point near the PvuII site. The precise points of deletion have not been determined. Moreover, pTS372 had another insert of about 0.5 kb in length within the region of the largest HinfI segment of pBR322, although the origin of the insert has not been determined. Thus pTS372 was 5.2 kb in length.

c) pTS373: The 4.9-kb DNA fragment was cleaved with AccI and the resulting 2.8-kb DNA fragment between the two AccI sites was purified electrophoretically. Repairing its staggered ends and ligation to pBR322 were done as described above. The recombinant plasmid pTS373 thus obtained also had a deletion of a region covering the tet gene between a point near the HindIII site and a point near the PvuII site, like pTS372 and carried the 2.8-kb AccI fragment in place of the region. It was therefore 4.9 kb in total length.

d) pTS376 and pTS377: The 3.5-kb DNA fragment between the $HindIII_1$ and EcoRI sites was partially digested with Sau3A and ligated to pBR322 DNA which had been cleaved with HindIII and BamHI. Two of the resulting plasmids, pTS376 and pTS377, carried the 1.3-kb and 2.8-kb fragments, respectively, flanked by the HindIII and Sau3A sites in place of the HindIII-BamHI segment of pBR322.

e) pTS378 and pTS379: The 1.4-kb $AccI_1$-AvaII DNA fragment was isolated electrophoretically and both staggered ends of the fragment were repaired as described above and ligated to pBR322 DNA which had been digested with HindIII and repaired with T4 DNA polymerase. The recombinant plasmids pTS378 and pTS379 thus obtained carried the 1.4-kb fragment at the HindIII site in directions opposite to each other.

Analysis of plasmid-encoded proteins

The maxicell system described by Horii et al. (15), which was originally reported by Sancer et al. (18), was generally followed to label the plasmid-encoded proteins. Cells of E. coli N1790 uvrA recA harboring each of the recombinant plasmids were grown in 5 ml of M9 medium supplemented with 0.5%

casamino acid, 40 μg of tryptophan per ml and an appropriate antibiotic (50 μg of ampicillin or 10 μg of tetracycline per ml) at 37°C to about 2 x 10^8 cells/ml and irradiated with 180 J of ultraviolet light per m^3. The irradiated cells were shaken at 37°C for 16 hr. Then the cells were harvested by centrifugation, washed twice, suspended with a half volume of M9 medium supplemented with 40 μg of tryptophan per ml, and cultured at 37°C for 1 hr to starve amino acids. [^{14}C]Amino acid mixture (1.75 mCi/mg; CEA, Cèdex, France) was added to the culture at a final concentration of 20 μCi/ml and the culture was shaken at 37°C for an additional 1 hr. The cells were pelleted by centrifugation, washed twice, and suspended in 0.2 ml of 50 mM Tris-HCl (pH 8.0)-15 mM Na_2-EDTA. Twenty μl of a 5 mg/ml solution of lysozyme was added to the suspension, which was then frozen and thawed three times. To the lysate was added 50 μl of 5 x concentrated lysing buffer (312.5 mM Tris-HCl, pH 6.8, 10% SDS, 25% mercaptoethanol, 50% glycerol and 0.05% bromophenol blue). The lysate was boiled for 3 min and electrophoresed on 13.5% SDS/polyacrylamide gel according to the method of Laemmli (19). The gel was then fixed, treated with an autoradiography enhancer (EN^{3}HANCE; New England Nuclear, Boston, Mass.), dried and exposed to a Fuji XR film (Fuji, Tokyo, Japan) at -70°C.

Immunoprecipitation

Five to ten μl of the maxicell lysate was diluted with 500 μl of Triton buffer (50 mM Tris-HCl (pH 8.0), 0.15 M NaCl, 0.1 mM Na_2-EDTA and 2% Triton X-100). The diluent was added to 1 μl of a solution of the monoclonal antibody against the staphylokinase produced by *E. coli* and held at room temperature for about 15 hr. The method of preparation and the characteristics of the monoclonal antibody against the staphylokinase will be published elsewhere (T. Sako and M. Ohwaki, in preparation). After the reaction, 100 μl of 10% formalin-fixed *S. aureus* cells (Protein-A bacterial adsorbent; Miles Laboratories, Elkhart, Ind.) was added and the mixture was incubated for 2 hr on ice. *S. aureus* cells with antigen-antibody complexes were pelleted by centrifugation, washed three times with 300 μl of Triton buffer and suspended with 30 μl of the lysing buffer. The suspension was boiled for 3 min and allowed to cool to room temperature. The bacteria were removed by centrifugation. A sample of the supernatant was analyzed by SDS/polyacrylamide gel electrophoresis as described above.

DNA sequencing

DNA was sequenced by the procedure of Maxam and Gilbert (20). All restriction fragments were 5'-labeled with [r-^{32}P]ATP (5,000 Ci/mmole; Amersham, Buckinghamshire, England) and T4 polynucleotide kinase (New England

Biolabs, Beverly, Mass.).

In vitro RNA synthesis

The 1.4-kb $AccI_1$-AvaII fragment or its digests with a restriction endonuclease (about 0.1 µg) were incubated at 37°C for 15 min with 0.2 units of E. coli RNA polymerase (New England Biolabs) in 20 µl of reaction mixture containing 20 mM Tris-HCl (pH 7.9), 100 mM KCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.5 mM spermidine and 50 µM each of ATP, GTP, CTP and UTP with [α -^{32}P]UTP (400 Ci/mmole, Amersham). The reaction was stopped by adding 1 µl of 0.25 M Na_2-EDTA (pH 8.0) and 2 µl of 1% SDS. Ethanol precipitation was performed twice and the precipitate was dissolved in 15 µl of 5 M urea containing 1mM Na_2-EDTA (pH 8.0), 0.05% bromophenol blue and 0.05% xylene cyanol FF. The mixture was heated at 80°C for 3 min, quickly chilled on ice and loaded on 5% polyacrylamide gel containing 8 M urea. After the electrophoresis the gel was exposed to a Kodak X-Omat AR film (Eastman Kodak, Rochester, N.Y.) at -70°C.

RESULTS

Localization of the sak gene on the Sø-C genome

As reported in the preceding paper (13), the plasmid pSAK361 which carries the 4.9-kb HindIII fragment from the Sø-C DNA, directs the synthesis of active staphylokinase in E. coli cells. To determine the region necessary for the expression of the sak gene we constructed a precise restriction map of the 4.9-kb HindIII fragment and subcloned various restriction fragments onto pBR322. The recombinant plasmids constructed as described in Materials and Methods are schematically presented in Fig. 1 and the staphylokinase activity produced by the E. coli clone harboring each of them is also shown in Fig. 1. The plasmids pTS373 and pTS372, both of which directed the synthesis of staphylokinase, determined boundaries of the left and right sides of the region necessary to express the sak gene, at the $AccI_1$ and AvaII sites, respectively. Thus we inserted the 1.4-kb $AccI_1$-AvaII fragment into pBR322 and obtained the two plasmids, pTS378 and pTS379, which carried the 1.4-kb fragment in directions opposite to each other (Fig. 1). The two plasmids directed the synthesis of staphylokinase in similar amounts. Fig. 1 also shows that E. coli cells carrying the plasmid pTS365 or pTS376 did not synthesize staphylokinase. This indicates that the information essential for the expression of sak is present in the two regions between the $AccI_1$ and HaeIII sites and between the $Sau3A_1$ and AvaII sites. These results suggest that the 0.2-kb segment flanked by the HaeIII and $Sau3A_1$ sites which overlapped between the inserted fragments on pTS365 and pTS376 includes a

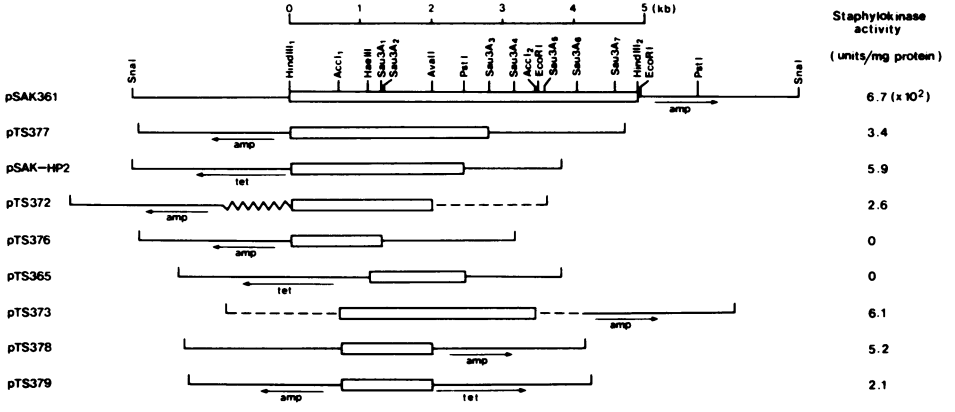

Fig. 1. Physical structures of the plasmid pSAK361 and of the plasmids carrying various regions of the 4.9 kb HindIII fragment. All the plasmids are linearized at the SnaI site on the pBR322 segment. Open bars, solid lines, broken lines and the wavy line represent the segments derived from the SØ-C DNA, those from pBR322 DNA, deleted regions and an insert, respectively. The antibiotic resistance determined by each plasmid is also shown with the orientation of the gene indicated by an arrow. The activity of the staphylokinase produced by the E. coli clone harboring each of the plasmids was determined as described previously (13).

portion of the structural gene and/or the promotor of sak.

Detection of the sak gene product

Plasmid-encoded proteins were analyzed by using the maxicell system (15). UV-irradiated and pre-incubated E. coli N1790 recA uvrA cells harboring each of the recombinant plasmids described in Fig. 1 were labeled with [^{14}C]amino acid mixture and portions of the maxicell lysates were subjected to SDS/polyacrylamide gel electrophoresis followed by fluorography. Fig. 2-a shows a representative fluorogram of the gel. Four proteins of 29 K, 18.5 K, 15.5 K and 14.3 K daltons were specified by the 4.9-kb HindIII fragment on the plasmid pSAK361. Two of them, the 18.5 K and 15.5 K proteins, were commonly synthesized from all the staphylokinase-positive plasmids which carried at least the AccI$_1$-AvaII segment. On the other hand, the 29 K and 14.3 K proteins were coded within the HindIII-Sau3A$_1$ segment (Fig. 2-a lane 5). To identify the sak gene product the maxicell lysates were immunoprecipitated with a monoclonal antibody against the staphylokinase produced by E. coli and then analyzed by SDS/polyacrylamide gel electrophoresis. As shown in Fig 2-b, the two proteins of 18.5 K and 15.5 K daltons were immunoactive, indicating that they are both the products of the sak gene. The molecular weight of the

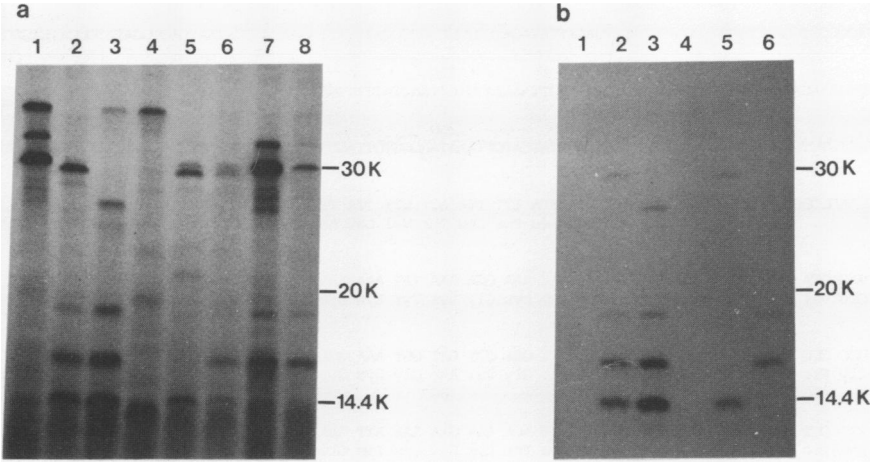

Fig. 2-a and b. Analyses of plasmid-encoded proteins in maxicells.
(a) The maxicell lysate of E. coli N1790 carrying (1) pBR322, (2) pSAK361, (3) pSAK-HP2, (4) pTS365, (5) pTS376, (6) pTS377, (7) pTS378 or (8) pTS379 was electrophoresed on 13.5% SDS/polyacrylamide gel followed by fluorography. Molecular weight standards (30K, carbonic anhydrase; 20K, soybean trypsin inhibitor; 14.4K, α-lactalbumin) were purchased from Pharmacia Fine Chemicals (Uppsala, Sweden).
(b) The maxicell lysate of E. coli N1790 carrying (1) pBR322, (2) pSAK361 (3) pSAK-HP2, (4) pTS365, (5) pTS376 or (6) pTS378 was treated with anti-staphylokinase monoclonal antibody and then bound to formalin-fixed S. aureus cells. The bound proteins were analyzed as described above.

latter agreed well with that of the active staphylokinase (15.3 K daltons) reported previously (13). Apparently the remaining two proteins of 29 K (25 K in case of pSAK-HP2) and 14.3 K daltons were also immunoactive. However, it

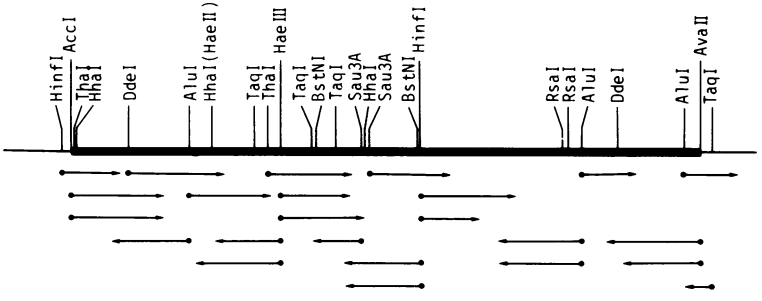

Fig. 3. Restriction map and sequencing strategy of the 1.4-kb AccI-AvaII fragment. The arrows indicate the direction of sequencing and the length of the sequence determined. The region whose sequence is given in Fig. 4 is drawn with a bold line.

```
AccI
                                                    50                                                  100
GTATACGCGCTGGAACATTAATATATGTGTTTGAAATTATAGATGGTTGGTGTCGCATTTATTGGAACAATCATAATGAGTGGATATGGCATGAGAGATT

                                                   150                                                  200
GATTGTGAAAGAAGTGTTTTAATTCTTAGGTTAAAATGTTAAATATTTGTTAATTATTTTTGAATGTAAGTTTAGTTTCTTTTAATATTTTATTGATTTT

                                                   250                                                  300
TAATATTTTCTCAATATAAAATGAAGTTGTTGATATTTATCATCTTAAATAAGGGTGTTAGCTATAAAAAGAGATAAATAAAAACAAATATATTATATTT

GGAGGAAGCGCC ATG CTC AAA AGA AGT TTA TTA TTT TTA ACT GTT TTA TTG TTA TTA TTC TCA TTT TCT TCA ATT ACT
             fMet Leu Lys Arg Ser Leu Leu Phe Leu Thr Val Leu Leu Leu Leu Phe Ser Phe Ser Ser Ile Thr
               1                                  10                                  20

AAT GAG GTA AGT GCA TCA AGT TCA TTC GAC AAA GGA AAA TAT AAA AAG GGC GAT GAC GCG AGT TAT TTT GAA CCA
Asn Glu Val Ser Ala Ser Ser Ser Phe Asp Lys Gly Lys Tyr Lys Lys Gly Asp Asp Ala Ser Tyr Phe Glu Pro
                           30                                  40

ACA GGC CCG TAT TTG ATG GTA AAT GTG ACT GGA GTT GAT GGT AAA GGA AAT GAA TTG CTA TCC CCT CAT TAT GTC
Thr Gly Pro Tyr Leu Met Val Asn Val Thr Gly Val Asp Gly Lys Gly Asn Glu Leu Leu Ser Pro His Tyr Val
        50                                  60                                  70

GAG TTT CCT ATT AAA CCT GGG ACT ACA CTT ACA AAA GAA AAA ATT GAA TAC TAT GTC GAA TGG GCA TTA GAT GCG
Glu Phe Pro Ile Lys Pro Gly Thr Thr Leu Thr Lys Glu Lys Ile Glu Tyr Tyr Val Glu Trp Ala Leu Asp Ala
                           80                                  90

ACA GCA TAT AAA GAG TTT AGA GTA GTT GAA TTA GAT CCA AGC GCA AAG ATC GAA GTC ACT TAT TAT GAT AAG AAT
Thr Ala Tyr Lys Glu Phe Arg Val Val Glu Leu Asp Pro Ser Ala Lys Ile Glu Val Thr Tyr Tyr Asp Lys Asn
        100                                 110                                 120

AAG AAA AAA GAA GAA ACG AAG TCT TTC CCT ATA ACA GAA AAA GGT TTT GTT GTC CCA GAT TTA TCA GAG CAT ATT
Lys Lys Lys Glu Glu Thr Lys Ser Phe Pro Ile Thr Glu Lys Gly Phe Val Val Pro Asp Leu Ser Glu His Ile
                           130                                 140

AAA AAC CCT GGA TTC AAC TTA ATT ACA AAG GTT GTT ATA GAA AAG AAA TAAAACAAAATAGTTGTTTATTATAGAAAGTAATGT
Lys Asn Pro Gly Phe Asn Leu Ile Thr Lys Val Val Ile Glu Lys Lys
        150                                 160
          850                                 900
CTTGATTGAATATGTGTAGTGAAATTATCTTTCATCAAATTCTCATTCATGCACGAATGGTTCTGCCCCACCTAATCAGATATTAGGTGACTTATGGGGA

          950                                1000
GAAATCAGTTAGAATGACATAGTCATGTCTATTTAAGCAGGTGCGTTACACACCTGATGTCTATTTACATTTAAAGATAAAATGTGCTATTATTTTACTA

         1050                                1100
GAACTTTTTAACATTTCTCTCAAGATTTAAATGTAGATAACAGGCAGGTACTACGGTACTTGCCTGTTTTTTTATGTTATAGCTAGCCTTCGGGCAGTTT

         1150                                1200
TTGTTATGATGCGTTACACACGCATCAACTATTCACACCTATCTTTGTTCACCTAAGCATGTCACTGGGTGTTTTTTTCTTACGATAGAGAGCATAGTTT

         1250                                1300
TCATACTACTCCCCGTAGTATATATGACTTTAGCATTCCCGTATAACAGTTTACGGGGTGCTTTTTATGTTATACTTACTTTTATATAGTAGGAGTGGAC
                                  AvaII
         1350                      |
TATATAGCTGGTCAGAGGCTGTATATCTGACTGTTGGTCC
```

Fig. 4. Nucleotide sequence of the 1.4-kb AccI-AvaII fragment. The nucleotides are numbered from the left end of the AccI recognition sequence. The amino acid sequence of the sak gene product is shown. Possible promotor sequences (-35 sequence and -10 sequence) and a possible Shine-Dalgarno sequence are underlined.

was revealed that they specifically bound to the formalin-fixed S. aureus cells without antibody (data not shown).

The DNA sequence of the sak gene

The DNA sequence of the region between the $AccI_1$ and AvaII sites in which the sak structural gene and its promotor sequence should be included was

determined by the method of Maxam and Gilbert (20). The strategy of sequencing with the detailed physical map of the AccI$_1$-AvaII fragment is shown in Fig. 3. About 90% of the region was sequenced at least twice or for both strands. The resultant DNA sequence of 1,377 bp is given in Fig. 4. Only one open reading frame which could code for a polypeptide of significant length was found. No other polypeptide longer than 60 amino acid residues could be coded. The coding region included the unique HaeIII site and two adjacent Sau3A sites which were thought to be located within the sak gene (Fig. 1). The predicted polypeptide which starts at the ATG codon at positions 313-315 and extends to the termination codon TAA at positions 802-804 is composed of 163 amino acid residues and has a calculated molecular weight of 18,490 which is consistent with that of one of the staphylokinase-related proteins detected in the maxicell system (Fig. 2-b). From these observations, we concluded that the sak gene codes for an 18,490-dalton protein.

The amino acid sequence of the sak gene product deduced from the DNA sequence (Fig. 4) contains a putative signal sequence which consists of an uncharged, mainly hydrophobic stretch of 27 amino acid residues except for glutamic acid at residue 24 following a basic segment containing lysine and arginine at the NH_2-terminal end.

The amino acid composition of the 15.5 K sak gene product

The 15.5 K active staphylokinase was purified from the periplasmic fraction of the E. coli cells by two cycles of CM-cellulose column chromatography. The purity of the protein was at least 95% as judged from the pattern of a SDS/polyacrylamide gel by staining with Coomassie blue (data not shown). The purified protein (0.1 mg) was hydrolyzed with 3 N mercaptoethanesulfonic acid (Pierce, Rockford, Ill.) at 110°C for 24 hr and analyzed with an amino acid analyzer (JLC-6AH; Nippon Denshi, Tokyo, Japan). The result is shown in Table 1 together with the amino acid composition deduced from the DNA sequence. The amino acid composition of the purified enzyme was in good agreement with that of the putative mature form of the sak gene product deduced from the DNA sequence if it was cleaved at alanine at residue 27.

In vitro RNA synthesis

To determine the initiation site for transcription of the sak gene active in E. coli cells, RNA was synthesized in vitro on the 1.4-kb AccI$_1$-AvaII fragment using E. coli RNA polymerase and the products were analyzed by gel electrophoresis. Three major transcripts 900, 225 and 150 nucleotides in length were synthesized from the fragment (Fig. 5 lanes 1 and 7). When the

Table 1 Amino acid composition

	amino acid analysis	from DNA sequence[a)]
Asp Asx	11.4	8
Asn		6 (5)
Thr	10.7	12 (10)
Ser	9.4	13 (8)
Glu Glx	13.65	15 (14)
Gln		0
Pro	10.8	9
Gly	12.2	9
Ala	5.4	6 (5)
Val	10.4	14 (12)
Cys	0	0
Met	1.1	2 (1)
Ile	6.7	8 (7)
Leu	8.6	16 (8)
Tyr	9.1	9
Phe	7.15	10 (7)
Lys	18.7	21 (20)
His	1.54	2
Arg	0.81	2 (1)
Trp	trace	1
Total	138	163 (136)

a) The values in the parentheses are the number of individual amino acid residues changed when the first 27 amino acid residues from the NH_2-terminus are omitted. The remaining residues did not change.

template DNA was cleaved at one or two sites with HinfI, Sau3A, BstNI, HaeIII or ThaI, only the 900-nucleotide transcript disappeared and instead a new transcript of about 530, 410, 320, 225 or 210 nucleotides in length, respectively, which seems to be a run-off product, was synthesized (Fig. 5 lanes 2-6). The difference between the length of the DNA fragment containing the NH_2-terminal portion of the sak gene and that of the newly-synthesized RNA transcript was constantly about 230 nucleotides, suggesting that the newly-synthesized transcripts were all initiated at a fixed origin at position about 230 or about 80 nucleotides upstream from the initiation codon and extended toward the sak gene. As the initial 900-nucleotide transcript was the only major product that disappeared when the template DNA was cleaved, this transcript was thought to be initiated at the same fixed origin and terminated within the 1.4-kb AccI-AvaII fragment in the in vitro transcription system using E. coli RNA polymerase.

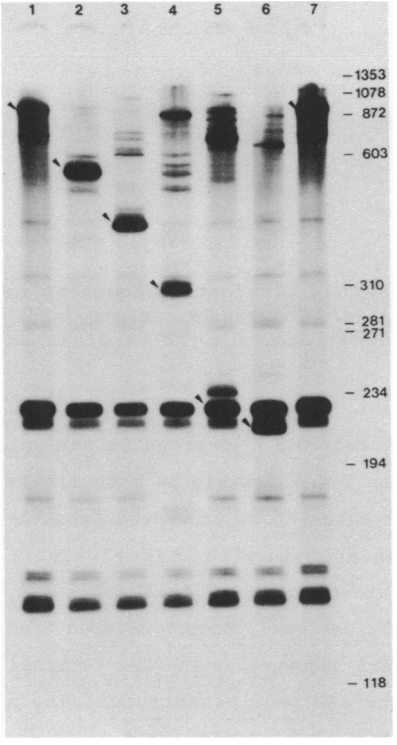

Fig. 5. Analysis of RNA transcripts synthesized *in vitro* by *E. coli* RNA polymerase. RNA was synthesized *in vitro* by *E. coli* RNA polymerase and electrophoresed on 5% polyacrylamide gel containing 8 M urea followed by autoradiography. The template DNAs used are the *Acc*I-*Ava*II fragment (1 and 7) and its digests with (2) *Hin*fI, (3) *Sau*3A, (4) *Bst*NI, (5) *Hae*III or (6) *Tha*I. The RNAs indicated by arrowheads are the 900-nucleotide transcript (lanes 1 and 7) and the newly-synthesized ones (lanes 2-6). The molecular weight standards used are the *Hae*III digests of øX-174 DNA.

DISCUSSION

In the previous paper (13) we reported the cloning and expression of the *sak* gene from the temperate bacteriophage Sø-C of *S. aureus*. We have now restricted the region necessary to express the *sak* gene in *E. coli* within the *Acc*I-*Ava*II segment of about 1.4 kb in length and determined the DNA sequence of the region. We found only one open reading frame coding for a polypeptide 163 amino acid residues in length which begins at the ATG codon at positions 313-315 and terminates at the TAA codon at positions 802-804. The coding sequence is unambiguously the *sak* structural gene because (a) the canonical Shine-Dalgarno ribosome binding sequence GGAGG which is commonly observed in Gram-positive ribosome binding sites (12) precedes 7 nucleotides upstream from the initiation codon, (b) the 0.2-kb region between the unique *Hae*III site and the *Sau*3A$_2$ site, which is essential for the expression of *sak*, is included in the sequence, and (c) the calculated molecular weight of the predicted polypeptide (M_r=18,490) is consistent with that of one of the *sak*-related proteins (18.5 K protein) detected in the maxicell lysate.

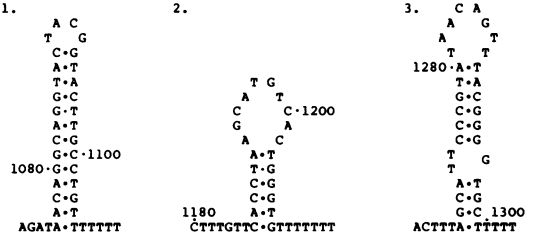

Fig. 6. Secondary structures of the three possible terminator sequences in the region adjacent to the 3'-end of the sak gene. The stem structures of 1, 2 and 3 begin at 272 bp, 383 bp and 465 bp downstream from the termination codon, respectively.

We identified the two proteins of 18.5 K and 15.5 K daltons as the sak gene products by using the maxicell system. The 18.5 K protein is most likely to be the precursor and the 15.5 K protein the mature form processed from the precursor, since we detected a 15.5 K-dalton protein with staphylokinase activity which was secreted into the periplasmic space of E. coli cells (13, this report), and the amino acid sequence of the sak gene product deduced from the nucleotide sequence contains a typical signal sequence at the NH_2-terminal region. The predicted signal sequence of the sak gene product contains a short basic segment followed by a stretch of hydrophobic amino acid residues, resembling many identified signal sequences of E. coli periplasmic and membrane proteins (2). The molecular weight and the amino acid composition of the purified staphylokinase extracted from the periplasmic space of E. coli cells suggest that the signal sequence of the sak gene product is cleaved at or near alanine at residue 27, where many other precursors of exported proteins are cleaved (2, 4, 7), though the exact cleavage site of the product in E. coli as well as in S. aureus is not yet determined.

The sak gene was expressed from the recombinant plasmids carrying the AccI-AvaII fragment inserted at the HindIII site of pBR322 in both directions. Thus the transcription of the sak gene is assumed to initiate within the AccI-AvaII segment. From the analyses of RNA synthesized in vitro from the AccI-AvaII fragment by E. coli RNA polymerase, we identified a transcript which starts at a unique position about 80 nucleotides upstream from the initiation codon of sak and reads through the sak gene. The transcript was labeled with [r-^{32}P]GTP but not with [r-^{32}P]ATP (unpublished data), showing that RNA synthesis initiates with GTP but not with ATP. We cannot rule out the possibility of UTP or CTP initiation which is not commonly used in E. coli

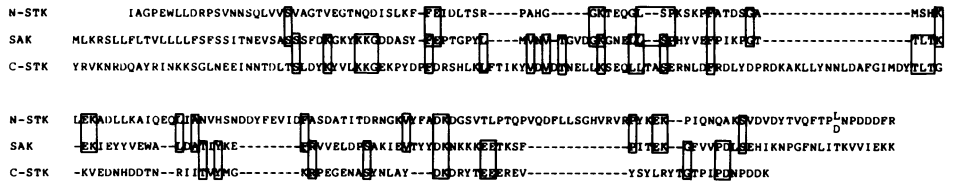

Fig. 7. Comparison of the amino acid sequences of staphylokinase and streptokinase. The amino acid sequences of staphylokinase (SAK) and the NH_2-terminal (N-STK) and COOH-terminal (C-STK) regions of streptokinase were compared. The one-letter abbreviations are those recommended by the IUPAC-IUB Commission on Biochemical Nomenclature. The sequences are rearranged to obtain maximal sequence homology by introducing breaks or gaps in either sequence.

(21). One possible initiation region can be located in the DNA sequence at positions 226-232 as GTTGTTG about 80 bp upstream from the initiation codon. In fact the potential RNA polymerase recognition site TTGATT at positions 193-199 and Pribnow box TAAAAT at positions 217-223 (Fig. 4), which are homologous with the promotor sequences of S. aureus β-lactamase gene (12) and of many E. coli and B. subtilis genes (22, 23), are found at the expected positions about 10 and 35 bp upstream, respectively, from the predicted transcription initiation site. These facts suggest that the potential promotor sequence described above functions actually in S. aureus as well as in E. coli. One notable feature in the upstream 170-bp region from the Shine-Dalgarno sequence is that the region is unusually abundant in A-T pairs. The content of A-T pairs of the region is 85.3% whereas that in the overall AccI-AvaII segment is 68.1%. Several direct and inverted repeats are also found, although the function of the region is not understood. The upstream AT-rich sequence of the sak gene may promote efficient transcription by S. aureus RNA polymerase, as mentioned by Moran et al. (23) in the case of B. subtilis genes, or it may play a critical role in regulation of expression of the sak gene.

The longest transcript synthesized in vitro from the AccI-AvaII fragment by E. coli RNA polymerase covering the sak gene was about 900 nucleotides in length and may function as an mRNA in vivo. This indicates that the RNA synthesis can terminate within the AccI-AvaII fragment. Three potential terminator sequences composed of a GC-rich inverted repeat followed by a stretch of Ts are found in the region between 300 and 500 nucleotides downstream from the termination codon. These sequences are homologous with rho-independent terminators identified in E. coli (21). Their possible secondary structures are shown in Fig. 6. Such sequences are also detected in

several genes from the Gram-positive organism B. subtilis (4, 9, 24), but the importance of these sequences in transcription termination in Gram-positive bacteria has not yet been established.

Finally, we compared the amino acid sequence of the sak gene product with that of the streptokinase reported by Jackson and Tang (25). They assumed that these two nonprotease plasminogen activators are derived from a common ancestral gene and that the latter is formed by the duplication and fusion of this gene. As shown in Fig. 7, the amino acid sequence of the sak gene product appears to be homologous with both the NH_2-terminal and COOH-terminal domains of the streptokinase defined by Jackson and Tang (25). However, the sequence homology of the predicted mature protein of sak with the NH_2-terminal and COOH-terminal regions of the streptokinase is only about 14% and 18%, respectively, the smallest values for functionally related proteins (26). Moreover the sequences of the streptokinase homologous with those of the staphylokinase are neither within the sequences homologous with serine proteases nor within the internal homologous sequences of the streptokinase (25). Thus we assume that staphylokinase is distantly related to streptokinase, and that if they share a commom ancestor, they would have diverged from each other a fairly long time ago. It is possible that staphylokinase activates or binds to plasminogen by a different mechanism from that of streptokinase.

ACKNOWLEDGEMENTS

We would like to express our sincere thanks to Ms. S. Sawaki for excellent technical assistance, to Dr. T. Sakurai, in whose laboratory this work was initiated, for support and criticism of the manuscript, and to Dr. H. Ohmori for introducing to T. S. the DNA sequencing technique. We also wish to thank Dr. H. Ogawa for the gift of E. coli K12 N1790, Dr. A. Iwabuchi for analyzing the amino acid composition, and Ms. T. Ichinose for typing the manuscript. This work was supported in part by a grant from Ministry of Science and Technology of Japan.

REFERENCES

1. Emr, S. D., Hall, M. N. and Silhavy, T. J. (1980) J. Cell Biol. 86, 701-711
2. Michaelis, S. and Beckwith, J. (1982) Ann. Rev. Microbiol. 36, 435-465.
3. Davis, B. D. and Tai, P.-C. (1980) Nature 283, 433-438.
4. Palva, I. Petterssen, R. F., Kalkkinen, N., Lehtovaara, P., Sarvas, M. and Soderlund, H. (1981) Gene 15, 43-51.
5. Gray, O. and Chang, S. (1981) J. Bacteriol. 145, 422-428.
6. Yang, M., Galizzi, A. and Henner, D. (1983) Nuc. Acids Res. 11, 237-249.

7. Lofdahl, S., Guss, B., Uhlen, M., Philipson, L. and Lindberg, M. (1983) Proc. Natl. Acad. Sci. USA. 80, 697-701.
8. McLaughlin, J. R., Murray, C. L. and Rabinowitz, J. C. (1981) J. Biol. Chem. 256, 11283-11291.
9. Neugebauer, K., Sprengel, R. and Schaller, H. (1981) Nuc. Acids Res. 9, 2577-2589.
10. Kroyer, J. and Chang, S. (1981) Gene 15, 343-347.
11. Nielsen, J. B. K. and Lampen, J. O. (1982) J. Biol. Chem. 257, 4490-4495.
12. Duggleby, C. and Jones, S. A. (1983) Nuc. Acids Res. 11, 3965-3076.
13. Sako, T., Sawaki, S., Sakurai, T., Itoh, S., Yoshizawa, Y. and Kondo, I. (1983) Mol. Gen. Genet. 190, 271-277
14. Wood, W. B. (1966) J. Bacteriol. 16 118-133.
15. Horii, T., Ogawa, T., Ogawa, H. (1981) Cell 23, 689-697.
16. Boliver, F. Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynecker, H. L., Boyer, H. W., Crosa, J. H. and Falkow, S. (1977) Gene 2, 95-113.
17. Kondo, I., Itoh, S. and Yoshizawa, Y. (1981) in Staphylococci and Staphylococcal Infections, Zbl. Bakt. Suppl. 10, ed. Jelijaszewicz, J., pp 357-362, Gustav Fischer Verlag, Stuttgart and New York.
18. Sancer, A., Hack, A. M. and Rupp, W. D. (1977) J. Bacteriol. 137, 692-693.
19. Laemmli, U. K. (1970) Nature (London) 277, 680-685.
20. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
21. Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319-353.
22. Hawley, D. K. and McClure, W. R. (1983) Gene 11, 2273-2255.
23. Moran, C. P., Lang, N., LeGrice, S. F., Lee, G., Stephans, M., Sonenshein, A. L., Pero, J. and Losick, R. (1982) Mol. Gen. Genet. 186, 339-346.
24. Shimotsu, H., Kawamura, F., Kobayashi, Y. and Saito, H. (1983) Proc. Natl. Acad. Sci. USA. 80,658-662.
25. Jackson, K. W. and Tang, J. (1982) Biochemistry 21, 6620-6625.
26. Dayhoff, M. O., Barker, W. C. and Hunt, L. T. (1983) Methods Enzymol. 91, 524-545.