Vol. 181, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 27, 1991 Pages 138-144

Nucleotide sequence of leukocidin S-component gene (*lukS*) from methicillin resistant *Staphylococcus aureus*

Arifur Rahman[1], Kazuo Izaki[1], Iwao Kato[2], and Yoshiyuki Kamio[1]*

[1]Department of Agricultural Chemistry, Faculty of Agriculture, Tohoku University, Sendai 981, Japan

[2]Second Department of Microbiology, Chiba University School of Medicine, Chiba 280, Japan

Received September 25, 1991

SUMMARY: The nucleotide sequence of *lukS* gene encoding S-component of Staphylococcal leukocidin from methicillin resistant *Staphylococcus aureus* (MRSA) was determined. The structural gene of *lukS* consisted of 857 base pairs. An open reading frame that could encode a 35,556 dalton polypeptide consisting of 315 amino acids was assigned. The molecular size of the polypeptide predicted from the amino acid composition was close to the value of pre-matured S-component determined in DNA-directed transcription/translation system. Inspection of the amino acid sequence deduced from nucleotide sequence of *lukS* and that from S-component of leukocidin clarified that pre-matured S-component contains a typical signal sequence at the NH_2 terminus. The amino acid sequence of predicted matured S-component correlated exactly with the known N-terminal 50 amino acid sequence of S-component from MRSA and *S. aureus* V8. The molecular size of the predicted matured protein was also close to the value of S-component determined in both MRSA and *S. aureus* V8. The nucleotide sequence of the 5'-flanking region showed the presence of the consensus sequence of ribosome binding site, Pribnow box and the RNA polymerase recognition site in *Escherichia coli*.

Staphylococcal leukocidin, known to be important in the pathogenicity of certain staphylococcal diseases consists of two protein components (S and F) that act synergistically to induce cytotoxic changes in human and rabbit polymorphonuclear leukocytes (1). Most of our knowledge about the mode of action of this toxin comes from the work of Woodin (1) and further extensive investigations have been reported by Noda et al. (2) as a cell membrane damaging toxin. However, there is no detail information about the relation between the chemical structure and function of these components on

*To whom correspondence should be addressed.

0006-291X/91 $1.50

the overall mechanism of leukocidin. Recently, we have tested 21 MRSA strains collected from hospital patients for producer of leukocidin. All the strains are found to be low or high producer of leukocidin (A. Rahman, K. Kitano, K. Izaki and Y. Kamio, submitted paper). In an unpublished paper, we have described the cloning and expression of the gene of S-component (*lukS*) from MRSA strain No.4 in *Escherichia coli* (3). In the present study, we determined the complete nucleotide sequence of the *lukS* gene and its flanking region to study the mechanism of the expression of *lukS* gene and the chemical structure of S-component of leukocidin.

MATERIALS AND METHODS

Bacterial strains and plasmids. *Escherichia coli* DH5α {*recA1*, λ^-, *Δ(lac-proAB), endA1, gyrA96, thi, hsdR17, relA1, supE44,[F', traD36, proAB, lac* I^q *ZΔM15]*} was used as a host strain for recombinant plasmids. Plasmid pUC119 and pUC118 were used as cloning vectors. *E. coli* MV1184 {*ara, Δ(lac-pro), strA, thi, (φ80 lacZΔM15), Δ(srl-recA)306::Tn10*(tet^r), *[F', traD36, proAB, lacI*^{q}, *ZΔM15]*}, and M13KO7 were used as a host and a helper phage, respectively for preparation of single stranded DNA.

Plasmid constructions. Plasmid pSRK91-1: Plasmid pSRK91 (Fig.1) was digested with *Bam*HI and *Kpn*I, following digestion with exonuclease III and mung bean nuclease according to the procedures described by Henikoff (4). After self-ligation, plasmid pSRK91-1 was obtained.

DNA sequencing. A series of deletion derivatives were obtained from plasmid pSRK91-1 (Fig.1) and subcloned into both pUC119 and pUC118. Resulting plasmids were introduced into *E. coli* MV1184 to isolate ss-DNA using bacteriophage M13K07 as described previously (5). DNA sequencing was performed by cycle sequencing of DNA with Dye primer -21M13 (Applied biosystem) using DNA sequencer (Model 373A, DNA sequencing System, Applied biosystem).

RESULTS AND DISCUSSION

Nucleotide sequence of the *lukS* gene from pSRK91-1. The 2.2 kb *Hin*dIII-*Xba*I fragment in pSRK91 contains the *lukS* gene of MRSA strain No.4 (3). We have determined the nucleotide sequence of the fragment containing the *lukS* and its 5'- and 3'-flanking regions according to the sequencing strategy shown in Fig.1. The nucleotide sequence of the fragment was comprising 1,466 bp as shown in Fig.2. Within this sequence, we can identify an open reading frame which begins with an ATG codon at position 374 and terminates with TGA at position 1,319. The open reading frame encoded a protein of 315 amino acids with a calculated molecular weight of 35,556, which coincides

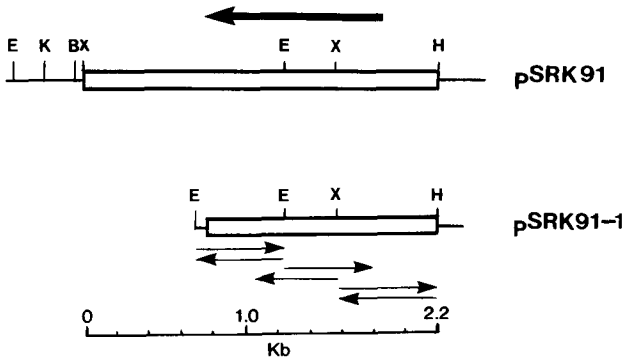

Fig.1. Restriction endonuclease map of 2.2 kb *Hin*dIII-*Xba*I fragment and sequencing strategy. The S-component of leukocidin coding region is indicated by a horizontal arrow. Double and single lines are inserted 2.2 kb fragment and polylinker region of vector plasmid pUC119, respectively. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; X, *Xba*I.

with that determined by SDS-PAGE of the product of pSRK91 DNA-directed transcription/translation system (3). The N-terminal 50 amino acid residues of the purified S-component from MRSA and from *S. aureus* V8 were identical to the deduced 50 amino acid residues from 30th N-terminal amino acid residue to 79th one. The total amino acid composition of matured S-component of MRSA strain No.4 deduced from the nucleotide sequence (from 461 to 1,318 base) almost corresponded with that of the matured S-protein of *S. aureus* V8 except the number of serine residue in the matured S-component from *S. aureus* V8 is 20 and that of deduced from nucleotide sequence was 30 (Table 1). Moreover, the molecular weight calculated from the deduced amino acid sequence was also found to be same as the purified S-protein from MRSA and that of expressed protein from *E. coli*(pSRK91). Therefore, the position of Ala^{29}-Ala^{30} of the pre-matured S-component might be processed by signal peptidase to secrete as a matured protein. In an unpublished paper, we have shown that the S-component was secreted into

Fig.2. Nucleotide sequence of *lukS* gene and the deduced amino acid sequence of S-component of leukocidin. The sequence is arranged so that 1 is the first nucleotide in the recognition sequence for *Hin*dIII and 1,466 is the last nucleotide sequenced in pSRK91-1 DNA. The strand shown is in the 5' to 3' direction. A putative ribosomal binding site and promoter sites are indicated by double underlines and rectangular boxes, respectively. Three inverted repeats downstream of the translation termination codon are indicated by horizontal arrows. A signal sequence is indicated by single underline. Vertical arrow represents the processing site of the pre-matured S-component.

```
        10         20         30         40         50         60         70
GAAGCTTAAATAAATATTCTCAAGTTAATAAATAATTAACTTTTAGATGGATTCATCAAAAATCGTAAAA

        80         90        100        110        120        130        140
AGAACAATTTGTATTTTACAAACATTAATTAAAAATAAAAGCAAGACATTCGTGCAATCGGTTACCTTAA

       150        160        170        180        190        200        210
ATTGTTTACAACTGTCAACAATACCAAGGTTTTATTAACTATATTTCTCACAAAATTAGCTTTTAGCATT

       220        230        240        250        260        270        280
CCAAACAAAAAAGGTTAAATTGAACGGAATTATGGCATTTTTAACTTAATTGTAAAAAAAGTTGATAATG

       290        300        310        320        330        340        350
GTCAATTGTTAATGAACAGTTAATTATAATAACGTCCAAAATATATTATTATTTAATTAAGTTAAATAAA

       360        370        380        390        400        410        420
ATTATAGAAAGACAGTGAAACTTATGCTTAAAAATAAAATATTAGCTACAACTTTATCTGTGAGCTTACT
                              M  L  K  N  K  I  L  A  T  T  L  S  V  S  L  L

       430        440        450        460        470        480        490
TGCCCCTCTTGCCAATCCGTTATTAGAAAATGCTAAAGCTGCCAACGATACTGAAGACATCGGTAAAGGA
 A  P  L  A  N  P  L  L  E  N  A  K  A  A  N  D  T  E  D  I  G  K  G

       500        510        520        530        540        550        560
AGCGATATAGAAATTATCAAAAGGACAGAAGATAAAACAAGTAATAAATGGGGCGTGACTCAAAATATTC
S  D  I  E  I  I  K  R  T  E  D  K  T  S  N  K  W  G  V  T  Q  N  I  Q

       570        580        590        600        610        620        630
AATTTGATTTTGTGAAGGATACAAAATATAACAAAGATGCTCTGATATTAAAGATGCAAGGATTCATTAG
  F  D  F  V  K  D  T  K  Y  N  K  D  A  L  I  L  K  M  Q  G  F  I  S

       640        650        660        670        680        690        700
CTCTAGAACAACATATTACAACTATAAAAAAACTAATCATGTTAAAGCTATGCGATGGCCATTCCAATAT
 S  R  T  T  Y  Y  N  Y  K  K  T  N  H  V  K  A  M  R  W  P  F  Q  Y

       710        720        730        740        750        760        770
AATATTGGTTTAAAAACAAATGATAAATATGTTTCTTTAATTAATTATTTACCTAAAAATAAAATTGAAT
N  I  G  L  K  T  N  D  K  Y  V  S  L  I  N  Y  L  P  K  N  K  I  E  S

       780        790        800        810        820        830        840
CTACAAACGTGAGTCAGACATTAGGATACAATATCGGTGGTAATTTCCAATCAGCCCCATCACTCGGTGG
  T  N  V  S  Q  T  L  G  Y  N  I  G  G  N  F  Q  S  A  P  S  L  G  G

       850        860        870        880        890        900        910
TAATGGATCATTTAACTATTCTAAATCGATTAGCTATACACAACAAAATTATGTAAGTGAAGTAGAACAA
 N  G  S  F  N  Y  S  K  S  I  S  Y  T  Q  Q  N  Y  V  S  E  V  E  Q

       920        930        940        950        960        970        980
CAAAACTCAAAAAGTGTTTTATGGGGCGTCAAAGCGAATTCATTCGCCACTGAATCAGGTCAAAAATCAG
Q  N  S  K  S  V  L  W  G  V  K  A  N  S  F  A  T  E  S  G  Q  K  S  A

       990       1000       1010       1020       1030       1040       1050
CATTTGATAGCGATTTATTTGTAGGCTACAAACCTCATAGTAAAGATCCTAGAGATTATTTCGTTCCAGA
  F  D  S  D  L  F  V  G  Y  K  P  H  S  K  D  P  R  D  Y  F  V  P  D

      1060       1070       1080       1090       1100       1110       1120
CAGTGAGTTACCTCCTCTTGTACAAAGTGGATTTAACCCTTCATTTATCGCCACAGTATCTCATGAAAAA
 S  E  L  P  P  L  V  Q  S  G  F  N  P  S  F  I  A  T  V  S  H  E  K

      1130       1140       1150       1160       1170       1180       1190
GGTTCAAGCGATACAAGCGAATTTGAAATTACTTACGGAAGAAACATGGATGTCACTCATGCCATTAAAA
G  S  S  D  T  S  E  F  E  I  T  Y  G  R  N  M  D  V  T  H  A  I  K  R

      1200       1210       1220       1230       1240       1250       1260
GATCAACGCATTATGGCAACAGTTATTTAGACGGACATAGAGTCCATAATGCATTCGTAAATAGAAACTA
  S  T  H  Y  G  N  S  Y  L  D  G  H  R  V  H  N  A  F  V  N  R  N  Y

      1270       1280       1290       1300       1310       1320       1330
TACTGTTAAATACGAGGTCAATTGGAAGACTCATGAAATCAAGGTGAAAGGACAGAATTGATATGAAAAT
 T  V  K  Y  E  V  N  W  K  T  H  E  I  K  V  K  G  Q  N                 -

      1340       1350       1360       1370       1380       1390       1400
GAATAAATTAGTCAAATCATCCGTTGCTACATCTATGGCATTATTATTACTTTCTGGTACTGCTAATGCT
      1410       1420       1430       1440       1450       1460
GAGGTAAAATAACACCAGTCAGCGTAAAAAAGTCGATGACAAGGTTACTTTATACAAACCACCAGC
```

Table 1. Amino acid compositions of matured and pre-matured S-component of leukocidin of *S. aureus*

Amino acid(s)	No. of amino acid residues		
	Purified S-component[a]	Matured S-component[b]	Pre-matured S-component[c]
Lys	25	27	30
Arg	8	8	8
His	8	8	8
Asp	39	16	16
Asn		26	29
		42	45
Glu	28	13	14
Gln		13	13
		26	27
Thr	16	20	22
Ser	20	30	32
Gly	21	20	20
Ala	12	10	15
Val	18	19	20
Ile	15	16	17
Leu	12	12	20
Cys	0	0	0
Trp	2	4	4
Met	3	3	4
Tyr	18	18	18
Phe	14	14	14
Pro	10	9	11
Total residue	269	286	315
Calculated Mr	30,585	32,525	35,525

[a] Determined by the amino acid analysis of the purified S-component from *S. aureus* V8.
[b] Calculated from the predicted amino acid sequence (residues 30-315) of the matured S-component (Fig.2).
[c] Calculated from the predicted amino acid sequence (residues 1-315) of the pre-matured S-component (Fig.2).

periplasm in *E. coli* DH5α(pSRK91) as a matured protein with molecular weight of 32 kDa (3). No cysteine residue was detected from deduced amino acid sequence of MRSA and chemical analysis of matured S-component from *S. aureus* V8 (6). Although it was reported earlier (6) that the molecular weight of S-component from *S. aureus* V8 is 31 kDa calculated on the basis of amino acid composition, in our present report we found that molecular weight of S-component from both V8 and MRSA are 32 kDa. This difference might be due to under-calculation of serine residue. These findings indicate that the open reading frame encoded the pro-S-component of leukocidin, and that the N-terminus of the pro-form was processed in its conversion into the active form.

The polarity and hydropathy of S-component according to Capaldi and Vanderkooi (7) and Kyte and Doolittle (8) respectively revealed that polarity was 25.2% and presence of no significant hydrophobic segment in matured S-component.

Analysis of the putative signal sequence. Sequence analysis of the N-terminal region of the *lukS* revealed a putative 29-amino acid signal sequence that showed several features characteristic of signal peptide (9). A cluster of positively charged amino acids adjacent to the methionine residue at the N-terminus (Met-Leu-Lys-Asn-Lys) was followed by a region rich in hydrophobic amino acids (Ile-Leu-Ala-Thr-Thr-Leu-Ser-Val-Ser-Leu-Leu-Ala-Pro-Leu-Ala-Asn-Pro-Leu-Leu). There were two Ala residues at positions -3 and -1 with respect to the cleavage site. These results indicate the presence of a single signal peptide.

Codon usage. The codon usage for the *lukS* gene was computed, and the following observations were made concerning the strong bias in the usage. (a) Codon AAA is used by 23 out of 27 Lys residues and GAT is used by 13 out of 16 Asp residues. (b) Neither AGA nor AGG is used for Arg at all in the gene. In additiion, although 8 His residues are present in the protein, none of them used codon CAC. These characteristics in the codon usage are similar to those of codon used in *pnl* gene from *Erwinia carotovora* Er which is a phytopathogenic species causing soft-rot (10). The percentage of G + C content in the structural gene of *lukS* was 9.57%. It was found that the whole sequenced data from 1-1466 bp shares low G + C content.

The nucleotide sequence of the 5'- and 3'-flanking regions. A presumed ribosome binding site (GAAAGA) was found 12 bp upstream of the most likely ATG translation initiation codon at position 374, which may correspond to the Shine-Dalgarno sequence (11). The sequence TAAATT at -143 to -148 bp from the tanslation initiation codon of the gene assumed to be -10, share the homology of the bases of *E. coli* -10 sequence with a conserved initial TA and the final T as Pribnow box (12). Another sequence TTAGCA at -166 to -171 bp position from the initiation codon and 17 bp apart from the -10 sequence was presumed to be -35 promoter sequence. Three palindromic

sequences were found between the stop codon of the *lukS* gene and the last codon sequenced (Fig.2). In an unpublished paper (3), it was shown that the *lukS* gene is expressed in *E. coli* cells by its own promoter.

Amino acid homology study. We computed our data for homology of amino acid with the gene bank (Bio-Database, 14-2 GENETYX). But no significant homology came out. It was also found that there was no homology of amino acid sequence deduced from *lukS* gene with that of Pseudomonal cytotoxin (13). Previous study of leukocidin S of *S. aureus* V8 revealed that the receptor of S-component of the leukocidin is G_{M1} on the surface of leukocyte (6). Our sequence data will be helpful for the detail study of the role of S-protein in the toxic mechanism of leukocidin.

REFERENCES

1. Woodin, A.M. (1972) In "The Staphylococci" (J.O. Cohen. Ed.) pp. 281-299, Wily Interscience, New York.
2. Noda, M., Hirayama, T., Matsuda, F. and Kato, I. (1985) Infect Immun. 50, 142-145.
3. Rahman, A., Izaki, K., Noda, M., Kato, I, and Kamio, Y. (1991) unpublished paper.
4. Henikoff, S. (1984) Gene, 28, 351-359.
5. Vieira, J. and Messing, J. (1987) Methods in Enzymology, 153, 3-11.
6. Noda, M., Hirayama, T., Kato. I. and Matsuda, F. (1980) Biochim. Biophys. Acta. 633, 33-44.
7. Capaldi, A.R. and Vanderkooi, G. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 930-932.
8. Kyte, J. and Doolittle, F. R. (1982) J. Mol. Biol. 157, 106-132.
9. von Heijne, G. (1983) Eur. J. Biochem. 133, 17-21.
10. Ohnishi, H., Nishida, T., Yoshida, A., Kamio, Y. and Izaki, K. (1991) Biochem. Biophys. Res. Commun. 176, 321-327.
11. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
12. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
13. Hayashi, T., Kamio, Y., Hishinuma, F., Usami, Y., Titani, K. and Terawaki, Y. (1989) Mol. Microbiol. 3, 861-868.