Cloning, sequence, and expression of the lysostaphin gene from *Staphylococcus simulans*

(preproenzyme/tandem repeats/extracellular processing/bacteriocin/plasmid-encoded)

PAUL A. RECSEI*, ALEXANDRA D. GRUSS, AND RICHARD P. NOVICK

Department of Plasmid Biology, The Public Health Research Institute of the City of New York, New York, NY 10016

Communicated by Esmond E. Snell, October 17, 1986 (received for review August 7, 1986)

ABSTRACT A 1.5-kilobase-pair fragment of DNA that contains the lysostaphin gene from *Staphylococcus simulans* and its flanking sequences has been cloned and completely sequenced. The gene encodes a preproenzyme of Mr, 42,000. The NH₂-terminal sequence of the preproenzyme is composed of a signal peptide followed by seven tandem repeats of a 13-amino acid sequence. Conversion of prolystaphin to the mature enzyme occurs extracellularly in cultures of *S. simulans* and involves removal of the NH₂-terminal portion of the proenzyme that contains the tandem repeats. The high degree of homology of the repeats suggests that they have arisen by duplication of a 5-amino-base-pair sequence. In *S. simulans*, the lysostaphin gene is present on a large β-lactamase plasmid.

Lysostaphin is a cell wall-degrading enzyme secreted by a single known strain of *Staphylococcus simulans* (NRRL B-2628) isolated by Schindler and Schuhardt (1, 2). The enzyme lyases practically all known staphylococcal species but it is inactive against bacteria of all other genera (1, 3, 4). Although its catalytic properties are not well-characterized, lysostaphin apparently hydrolyzes polypeptide cross-links present in the peptidoglycan of the staphylococcal cell wall (5). The enzyme is a monomer of Mr ≈ 25,000 and is reported to contain zinc (6). Lysostaphin production occurs in stationary-phase cultures grown under certain conditions and appears to be coordinated with production of other extracellular enzymes, including a protease and a hexosaminidase (7). Producing cultures are resistant to the enzyme, while cultures grown under at least some nonproducing conditions are sensitive (8). It is not clear whether resistance to lysostaphin in *S. simulans* results from the action of an immunity product(s) or whether it occurs naturally under certain conditions. Alterations in sensitivity to the enzyme may be due to changes in the amino acid composition of the peptidoglycan (8, 9).

In this paper, we show that the lysostaphin gene is present on a large Staphylococcus plasmid and encodes a preproenzyme of Mr ≈ 42,000. Conversion of prolystaphin to the mature enzyme occurs extracellularly in cultures of *S. simulans* and involves removal of the NH₂-terminal portion of the proenzyme, which contains seven tandem repeats of a 13-amino acid sequence.

MATERIALS AND METHODS

Materials. Restriction enzymes, T4 DNA ligase, *Escherichia coli* DNA polymerase, and ribonuclease were from Boehringer Mannheim; M13 pentadecamer primer was from New England Biolabs; goat antibodies to rabbit IgG were from Miles-Yeda (Rehovot, Israel); calf intestine alkaline phosphatase and 5-bromo-4-chloroindolyl phosphate were from Sigma; and lysostaphin was from Mead Johnson.

Preparation of DNA. *S. simulans* grown to midlogarithmic phase on 0.5 liter of CAA medium (8) was harvested by centrifugation, washed with 50 mM Tris-HCl/50 mM EDTA, pH 7.8, and suspended in 100 ml of this buffer containing lysostaphin (50 μg/ml) and lysozyme (0.5 mg/ml). After 2 hr at 37°C, Pronase (1 mg/ml) and NaDodSO₄ (0.6%) were added and the suspension was incubated for 2 hr at 37°C. The lysate was then extracted twice with an equal volume of phenol. Nucleic acid was precipitated by addition of 2 vol of ethanol, collected by centrifugation, dissolved in 10 ml of TE (10 mM Tris-HCl/1 mM EDTA, pH 8.0), and digested with pancreatic RNase (30 μg/ml) and T1 RNase (2 units/ml) for 2 hr at 37°C. The DNA was precipitated with ethanol and dissolved in TE. The yield was 1.5 mg. Chromosomal and plasmid DNA were obtained by CsCl density-gradient centrifugation. Plasmid DNA was isolated from *E. coli* by alkaline NaDodSO₄ extraction of cell lysates (10).

Cloning and DNA Sequencing. Cloning was carried out using pUC8 as the vector and *E. coli* JM105 as the host (11). *S. simulans* DNA was partially digested with MboI and fractionated by centrifugation through a 12-ml 10–30% sucrose gradient at 35,000 rpm for 2 hr. Fragments (10 μg) from 5 to 15 kilobase pairs (kbp) were pooled and ligated to BamHI-digested pUC8 (2 μg). About 80% of the transformants obtained using the ligated DNA contained recombinant plasmids, as indicated by inactivation of lacZ, the truncated β-galactosidase gene from *E. coli* present on the pUC plasmids. DNA sequences were determined by the dideoxy-chain-termination method (12) using the phages vectors M13mp10 and M13mp11 (13).

Lysostaphin Assays. *Staphylococcus aureus* RN492 (14), a constitutive β-lactamase producer that is relatively resistant to ampicillin, was used as the indicator strain. *E. coli* colonies grown on L agar containing ampicillin (50 μg/ml) were exposed to chloroform vapor for 30 min and overlaid with GL top agar (15) containing a 0.1% (vol/vol) suspension of *S. aureus* RN492 that had been grown to stationary phase on C medium (15). Liquid samples (5 μl) were added to wells in 1% agarose containing 0.1 M NaCl and 0.05 M potassium phosphate (pH 7.2) and were overlaid in the same way. The area of cell lysis was proportional to the amount of lysostaphin in the range from 1 to 500 ng.

Immunoblots. Rabbit antibodies to lysostaphin were prepared and purified by affinity chromatography as described (16). Goat antibodies to rabbit IgG were cross-linked to alkaline phosphatase with glutaraldehyde (17). *E. coli* JM105 (pRG5) cells grown to late-logarithmic phase in 20 ml of LB medium (18) containing ampicillin (50 μg/ml) were harvested by centrifugation, washed with 10 mM Tris-HCl/30 mM

*To whom reprint requests should be addressed.
formylmethionine for of ampicillin plates. E. coli colonies degrade ampicillin and permit staphylococcal cells to grow around the colonies. Clear zones at the center of the staphylococcal cells are due to the production of lysostaphin by the E. coli colonies.

NaCl, pH 8.0, suspended in 1 ml of this buffer, and sonicated for 2 min at 0°C using a Branson S 125 sonicator. The culture supernatant was concentrated 20-fold by ultrafiltration using an Amicon YM-10 membrane. Samples were subjected to NaDodSO4/polyacylamid gel electrophoresis (19) and transferred to nitrocellulose at 1.5 A for 2 hr. Immunoreactive proteins were detected as described by Blake et al. (20).

Southern Blots. DNA was fractionated by electrophoresis on 1% agarose in 50 mM Tris-HCl/50 mM boric acid/1 mM EDTA, pH 8.0, at 80 V for 4 hr and transferred to nitrocellulose (21). Filters were prehybridized at 65°C for 2 hr in a solution containing 0.8 M NaCl, 0.08 M sodium citrate, 0.5% NaDodSO4, denatured salmon sperm DNA (50 μg/ml), and 5× Denhardt's solution (22). Filters were hybridized at 65°C for 16 hr with 2 × 10^6 cpmp of 32P-labeled probe (23), washed as described (24), and exposed to Fuji RX film for 24 hr at −70°C.

Protein Sequencing. Automated Edman degradation was performed on a Beckman 890C liquid-phase sequencer. Phenylthiohydantoin-derivatized amino acids were identified by high-pressure liquid chromatography as described by Terr (25).

**FIG. 2.** Strategy for sequencing the 1.5-kbp Hpa II/HindIII fragment containing the lysostaphin gene from S. simulans. Only the restriction sites used for sequencing are shown. Arrows indicate the direction and length of determined sequences. Greater than 95% of the sequence was determined from both strands.

**FIG. 3.** Nucleotide sequence of the lysostaphin gene from S. simulans and amino acid sequence of the encoded gene product. The NH2-terminal formylmethionine of preprolystaphin is designated as residue −143 and the NH2-terminal alanine of mature lysostaphin is designated as residue +1.
RESULTS

Isolation of Lysostaphin-Producing Clones of E. coli. S. simulans DNA was partially digested with Mbo I and fragments with an average size of 10 kbp were isolated by sucrose gradient centrifugation and ligated to BamHI-digested pUC8 (11). The ligated DNA was used to transform E. coli JM105 and ampicillin-resistant colonies were overlaid with a suspension of S. aureus RN422 to screen for lysostaphin production. Nine of ≈1000 E. coli clones containing recombinant plasmids (amp*, lacZ−) were lysostaphin producers (Fig. 1). Assuming that the S. simulans chromosome is ≈2000 kbp (26), the probability of cloning a chromosomal gene at this frequency is ≈0.4 (27).

Restriction Analysis, Subcloning, and Sequencing of the Lysostaphin Gene. Lysostaphin-producing transformants contained recombinant plasmids with inserts of either 6.0, 6.5, or 8.0 kbp. Restriction analysis showed that these inserts were present in either orientation with respect to the vector and contained a 4.3-kbp fragment in common. The lysostaphin gene was localized within a 1.5-kbp Hpa II/HindIII fragment, which appeared to contain a promoter that was functional in E. coli, since clones containing the insert in either orientation with respect to the lacZ− gene (i.e., in either pUC8 or pUC9) produced similar amounts of enzyme. The nucleotide sequence of the cloned DNA was determined as shown in Fig. 2 and is given in Fig. 3. A 1167-nucleotide open reading frame extends from a UUG initiation codon (nucleotides 245–247) to a UGA termination codon (nucleotides 1412–1414). The inferred ribosome-binding site AGGAGG (nucleotides 231–237) is separated from the initiation codon by 7 bp and shows complete complementarity to the postulated mRNA binding site of the 16S ribosomal RNA (28). A presumed promoter with −35 and −10 regions at nucleotides 89–95 and 110–119, respectively, is highly homologous to ρ7 promoters from Bacillus subtilis (29). No additional open reading frames are present in the cloned fragment.

The Lysostaphin Gene Encodes a Preproenzyme. The lysostaphin gene encodes a protein composed of 389 amino acids with Mr 42,205 (Table 1). The NH2-terminal sequence of the encoded product contains a cluster of four positively charged residues followed by an uncharged largely hydrophobic sequence and, therefore, has the properties characteristic of a signal peptide (30). The Ala-Ser bond at position −121 and −120 or that at position −108 and −107 (Fig. 4) is the likely signal cleavage site. The sequence from alanine −95 through arginine −5 is composed of seven tandem repeats of a 13-amino acid sequence (Fig. 4). The first six of these repeats are identical, while the seventh contains 3 amino acid substitutions. This portion of the molecule is highly ionic (33% of the residues are acidic or basic) and has a net negative charge. The 39-bp nucleotide sequences that encode these repeats are also highly homologous (Fig. 4). Repeats two,

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Preprolysostaphin predicted</th>
<th>Lysostaphin Predicted</th>
<th>Observed*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>31</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Arg</td>
<td>9</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Asn</td>
<td>27</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>(Asx)</td>
<td>(35)</td>
<td>(23)</td>
<td>(24)</td>
</tr>
<tr>
<td>Asp</td>
<td>8</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gin</td>
<td>12</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>(Glx)</td>
<td>(39)</td>
<td>(15)</td>
<td>(15)</td>
</tr>
<tr>
<td>Glu</td>
<td>27</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>38</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>His</td>
<td>10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Ile</td>
<td>16</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Leu</td>
<td>16</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Lys</td>
<td>29</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Met</td>
<td>9</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Phe</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Pro</td>
<td>19</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Ser</td>
<td>31</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Thr</td>
<td>40</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Trp</td>
<td>8</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Tyr</td>
<td>19</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Val</td>
<td>32</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

Total residues: \( M_r = 42,205 \)

Values are expressed as mol per mol of enzyme.

*Data from Trayer and Buckley (6).

a signal peptide (30). The Ala-Ser bond at position −121 and −120 or that at position −108 and −107 (Fig. 4) is the likely signal cleavage site. The sequence from alanine −95 through arginine −5 is composed of seven tandem repeats of a 13-amino acid sequence (Fig. 4). The first six of these repeats are identical, while the seventh contains 3 amino acid substitutions. This portion of the molecule is highly ionic (33% of the residues are acidic or basic) and has a net negative charge. The 39-bp nucleotide sequences that encode these repeats are also highly homologous (Fig. 4). Repeats two,
three, and four are identical; repeats one and six (which are identical) and five contain a single substitution in the third base of a codon with no effect on the amino acid coded; and repeat seven contains six substitutions.

The NH$_2$-terminal sequence of lysostaphin as determined by Edman degradation (data not shown) is Ala-Ala-Thr-His-Glu, corresponding to residues 1–5 of the encoded gene product (Fig. 4). Predicted values for the amino acid composition and molecular weight of the enzyme, calculated assuming that the lysostaphin sequence extends from Ala-1 to the COOH-terminal Lys-246 of the encoded product, are in excellent agreement with those determined experimentally (6), as shown in Table 1. The only significant difference is in tryptophan content, which is not unexpected, however, as the reaction used for tryptophan determination is often not quantitative (31). The mature enzyme has $M_r$ 26,921 as calculated from the sequence, contains no cysteine, and has more basic and amidated residues (13% and 11% of the total, respectively) than acidic residues (5% of the total).

**Synthesis of Precursor and Mature Forms of Lysostaphin in E. coli and S. simulans.** Late-logarithmic-phase cultures of E. coli JM105 containing pRG5 (the recombinant plasmid containing the 1.5-kbp Hpa II/HindIII fragment of S. simulans DNA cloned into the Acc I/HindIII sites of pUC8) show lysostaphin activity in the supernatant, periplasmic, and cytoplasmic fractions (65%, 15%, and 20%, respectively, of the total of 3 μg of enzyme per ml). Protein blotting using affinity-purified antibodies to lysostaphin showed the presence of the mature enzyme in the supernatant (Fig. 5, lane 8). The E. coli cellular fraction contains smaller amounts of the mature enzyme and larger amounts of a cross-reactive protein with an apparent $M_r$ of ~64,000 (lane 7). A cross-reactive protein with identical electrophoretic mobility is also present in the supernatant of lysostaphin-producing cultures of S. simulans (lane 6). This protein is most likely prolysostaphin since it precedes lysostaphin in appearance and then disappears as lysostaphin accumulates (lanes 1–4). It is not present in nonproducing cultures of S. simulans or E. coli.

**Lysostaphin Is Plasmid-Encoded.** S. simulans contains several plasmids, including a large plasmid of ~40 kbp (Fig. 6, lane 2). DNA blot analysis using the cloned lysostaphin gene as probe showed that the gene is present on the large plasmid (lane 3). S. simulans is resistant to penicillin (2), a trait that is often associated with large plasmids in S. aureus (32). Hybridization with a restriction fragment derived from the β-lactamase gene of the S. aureus plasmid pI258 (14) showed that the large S. simulans plasmid also carries a β-lactamase determinant.

**DISCUSSION**

The lysostaphin gene from S. simulans encodes a preproenzyme of $M_r$ 42,205 with an NH$_2$-terminal sequence that is composed of a signal peptide followed by seven tandem repeats of a 13-amino acid sequence. Conversion of the proenzyme to the mature enzyme involves cleavage of the bond between arginine —1 and alanine +1 with removal of the NH$_2$-terminal portion of the proenzyme, which contains the repeated sequences (Fig. 4). Immunoblot analysis using antibodies to lysostaphin shows that a cross-reactive protein with an apparent $M_r$ of ~64,000 accumulates in the supernatant of early stationary-phase cultures of S. simulans and then disappears as lysostaphin accumulates. Evidence that this cross-reactive protein is in fact prolysostaphin has been obtained by showing that the purified protein is converted to mature lysostaphin in vitro (R. Zhou and P.A.R., unpublished results). Overestimation of the molecular weight of prolysostaphin by NaDodSO$_4$/polyacrylamide gel electrophoresis is probably the result of below-average binding of NaDodSO$_4$ to the protein because of the high content of glutamyl residues in the tandem repeats. It is known, for example, that esterification of single glutamyl residues of the E. coli chemotaxis proteins results in an increase in migration rate on NaDodSO$_4$/polyacrylamide gels (33). The high degree of homology of the repeats suggests that they have arisen by duplication of a 39-bp sequence of DNA. Their role, if any, remains to be established. It is interesting to note that the mature forms of two proteins associated with the cell envelope of Gram-positive bacteria, protein A from S. aureus (34), and M protein from Streptococcus pyogenes (35), also contain tandemly repeated peptides. The use of UUG as an initiation codon, described here for the lysostaphin gene, has also been observed for several other staphylococcal genes (36). Alkaline protease and neutral protease from B. subtilis (29, 37, 38) and Bacillus amyloliquefaciens (39, 40) are also synthesized as preproenzymes. While the Bacillus proproteases differ from prolysostaphin in structural organization and do not contain repeated sequences, the proenzyme cleavage sites of these proteins are similar.

Lysostaphin and other bacterial proteins with bactericidal activity are known as bacteriocins (41). Plasmids of various sizes have been associated with bacteriocin production and immunity in Staphylococcus (42–44). Additional studies will be required to determine whether the lysostaphin plasmid encodes an immunity product(s) that protects the host against...
lysostaphin. Hybridization analyses show that the \( \beta \)-lactamase determinants from \( S. \) aureus and \( S. \) simulans are homologous, and it seems likely that the lysostaphin plasmid will show other similarities to the \( S. \) aureus \( \beta \)-lactamase plasmids. Lysostaphin itself may be related to autolytic enzymes, which appear to be widely distributed in staphylococci (45). Although little is known about their properties, these enzymes are apparently associated with the cell wall of the producing organism and may play an essential role in reshaping the wall during cell growth and division.

We thank Dr. Steve Projan and Dr. James M. Manning for help and advice. We thank Donna Atherton of the Rockefeller University Protein Sequencing Facility for protein sequence analysis. This work was supported by a grant from the Ambrose Monell Foundation and by National Institutes of Health Grant GM-14372 (R.P.N.).