APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 1987, p. 969–973
0099-2240/87/050969-05$02.00/0

Vol. 53, No. 5

Nucleotide and Deduced Amino Acid Sequences of the *Staphylococcus aureus* Phospho-β-Galactosidase Gene

FREDERICK BREIDT, JR., AND GEORGE C. STEWART*

Department of Microbiology, University of Kansas, Lawrence, Kansas 66045

Received 22 December 1986/Accepted 26 February 1987

We sequenced the *Staphylococcus aureus* phospho-β-galactosidase gene. The protein product of this gene consisted of 470 amino acids, giving a molecular weight of 54,557. This gene appears to be transcribed as the terminal sequence on a polycistronic message.

Lactose metabolism in gram-positive bacteria is generally initiated with the phosphorylation of lactose upon uptake by the phosphoenolpyruvate-dependent phosphotransferase system. The lactose phosphate which appears intracellularly is converted to glucose and galactose-6-phosphate by phospho-β-galactosidase. In the lactic streptococci, this enzyme plays a pivotal role in lactose catabolism, with the eventual production of lactic acid. Because these bacteria are used as starter cultures for dairy fermentations, much attention has been focused on phospho-β-galactosidase (11).

The gene for phospho-β-galactosidase has been shown to be chromosomally located in *Staphylococcus aureus* (1, 5) and *Streptococcus sanguis* (20) or plasmid borne, as in *Lactobacillus casei* (9), *Streptococcus cremoris* (7), and *Streptococcus lactis* (10, 19). The gene for phospho-β-galactosidase has been cloned from *S. aureus* (1), *L. casei* (9), *S. lactis* (4, 10), and *S. cremoris* (7). The products of these cloned determinants are 56,000-, 56,000-, 58,000-, and 40,000-dalton proteins, respectively. To date, however, no DNA or protein sequence information has been reported for these phospho-β-galactosidase determinants or for their products.

Here we report the nucleotide sequence of the staphylococcal phospho-β-galactosidase gene and the deduced amino acid sequence of the protein.

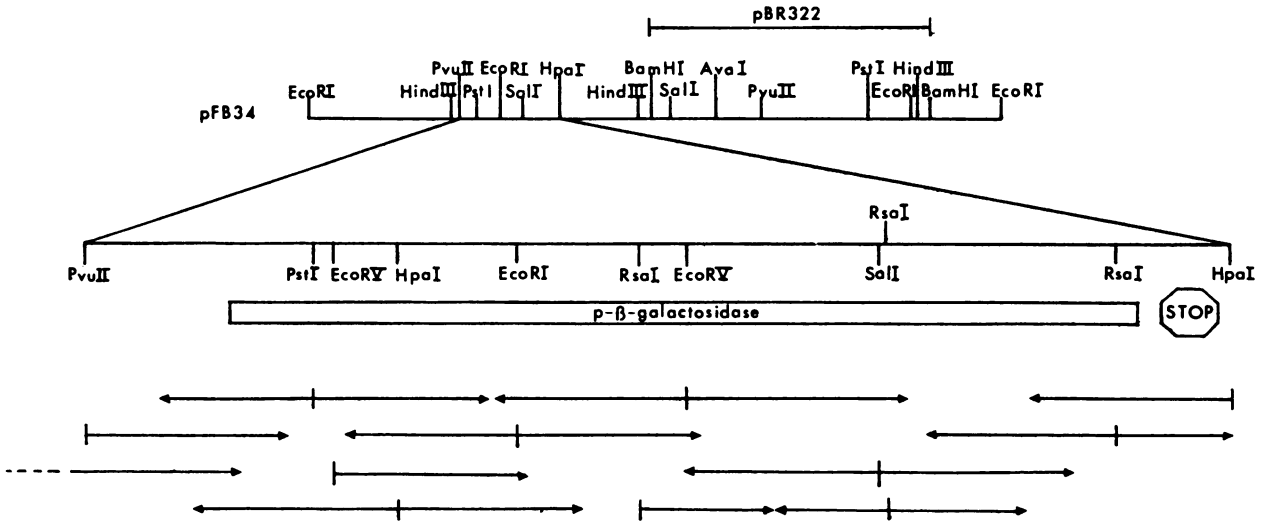

FIG. 1. Sequencing strategy for the phospho-β-galactosidase gene. A physical map of the 6.1-kilobase insert of pFB34 (1) is shown with the 1.8-kilobase *Pvu*II-*Hpa*I fragment, whose sequence is reported here, presented in expanded form. Sites for the indicated restriction endonucleases are also given in Fig. 2. Restriction fragments were subcloned into phage vectors M13mp10 and M13mp11. The arrows indicate the origin, direction, and extent of the individual sequencing reactions. Because each fragment was subcloned and sequenced in a shotgun fashion, each DNA fragment indicated was independently sequenced multiple times.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophage. Plasmid pFB34 is a pBR322 derivative which carries the staphylococcal phospho-β-galactosidase gene on a 5.1-kilobase *Eco*RI-*Bam*HI DNA fragment (1). Phage M13mp10 and M13mp11 and recombinant derivatives were propagated in *Escherichia coli* JM101 (13). DNA manipulations were as previously reported (1).

Determination of nucleotide sequence. Restriction fragments from the phospho-β-galactosidase region of pFB34 or

* Corresponding author.

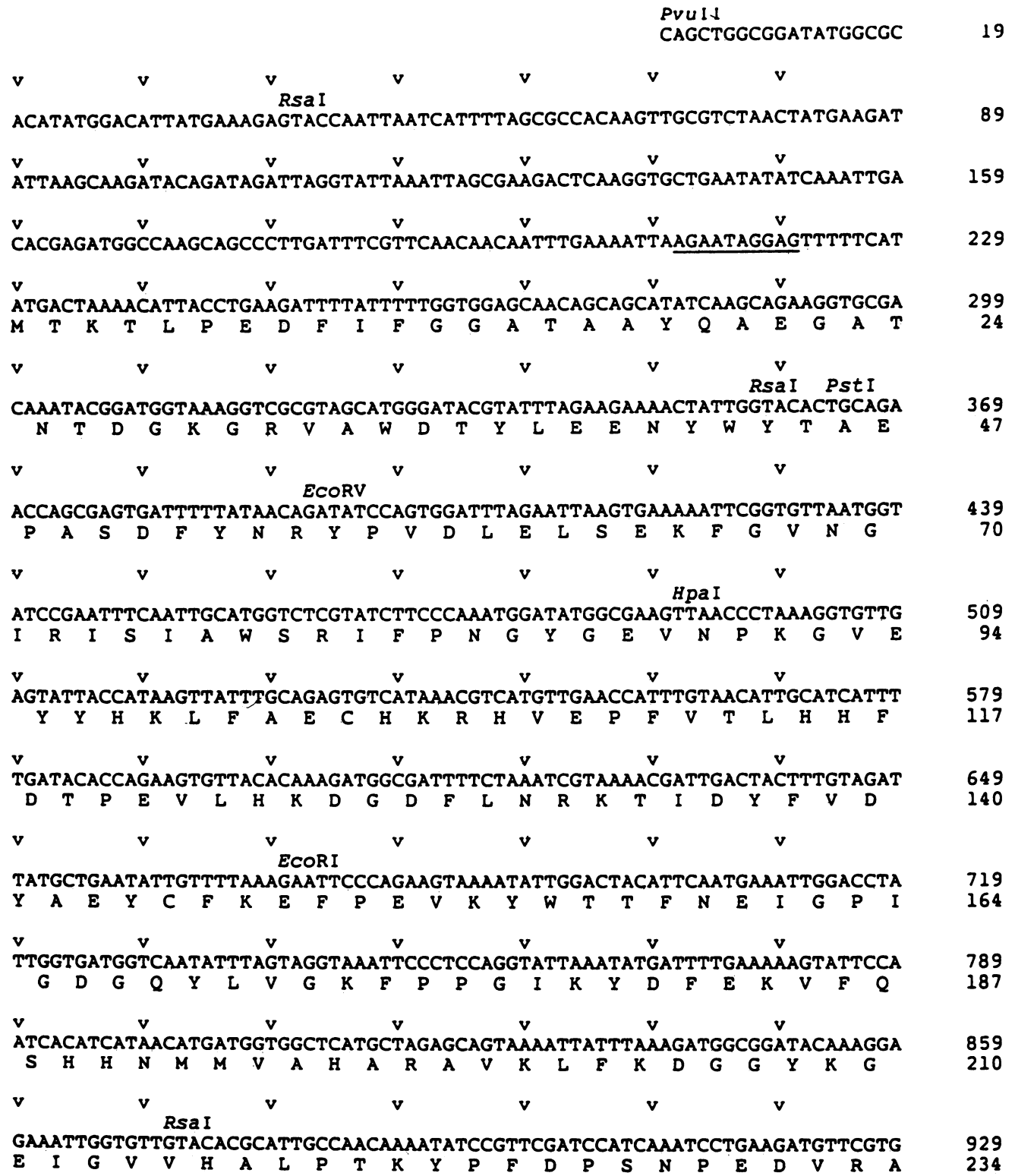

FIG. 2. Nucleotide sequence of the nonsense strand of the *Pvu*II-*Hpa*I fragment containing the phospho-β-galactosidase gene and the corresponding amino acid sequence of the translation product. The ribosome-binding site sequence is underlined. The indirectly repeated sequence capable of forming the stem structure of a potential transcription terminator is indicated (*). Amino acids are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

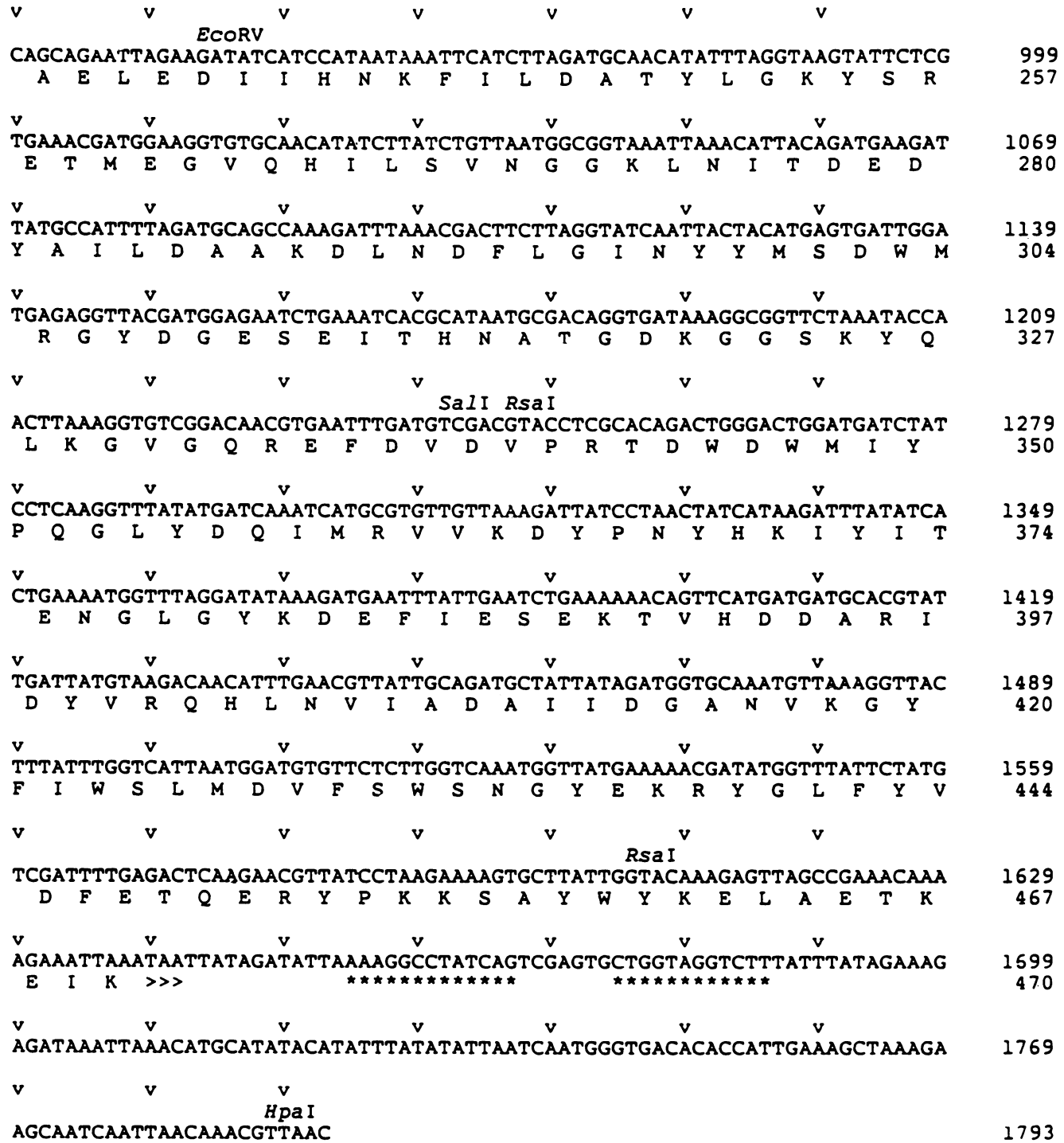

FIG. 2 (*Continued*)

its subclones (1) were cloned into the M13 phage vectors and sequenced by the dideoxy chain termination method of Sanger et al. (17) with electrophoresis in 8% polyacrylamide–7 M urea gels. Restriction enzymes, T4 DNA ligase, and the Klenow fragment of DNA polymerase I were obtained from Pharmacia, Inc. and Promega Biotech. [α-^{32}P]dATP was obtained from New England Nuclear Corp.

RESULTS AND DISCUSSION

Nucleotide sequence of the phospho-β-galactosidase gene. The restriction map of and sequencing strategy for the phospho-β-galactosidase gene are shown in Fig. 1. Previous results (1) indicated that the structural gene lay between the *Pvu*II and *Hpa*I sites (Fig. 1). The nucleotide and amino acid

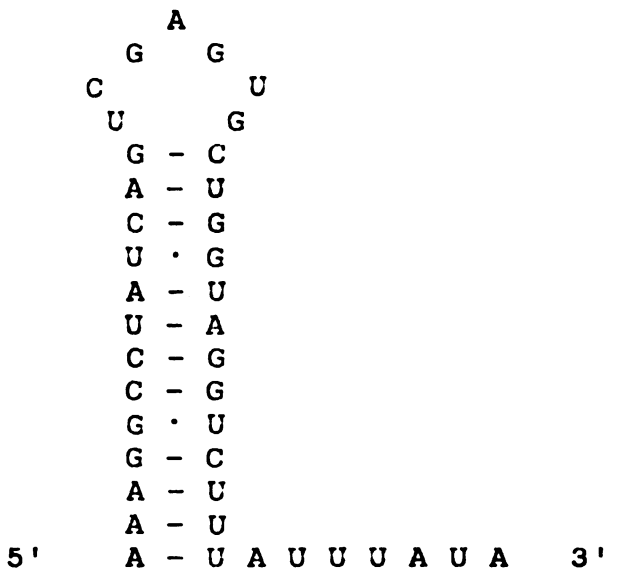

FIG. 3. Predicted secondary structure of the phospho-β-galactosidase mRNA corresponding to the potential transcription terminator sequence indicated in Fig. 2. The calculated ΔG is −15.2 kcal/mol (ca. −63.6 J/mol) (18).

sequences are shown in Fig. 2. A 1,416-base-pair open reading frame exists from the initiation codon ATG (position 230 to 232) to the ochre termination codon TAA (position 1640 to 1642). This open reading frame encodes a 470-amino-acid polypeptide with a calculated molecular mass of 54,557 daltons. This value is consistent with the reported molecular weights of the protein purified from *S. aureus* (6) or expressed in *E. coli* maxicells (1). A potential ribosome-binding site sequence (AGAATAGGAG) is found seven bases 5′ to the initiation codon. Because staphylococcal proteins are expressed when their genes are introduced into *Bacillus subtilis* (12, 15), the staphylococcal and *B. subtilis* 16S rRNAs must have similar 3′ termini. Based upon the published sequence for the 3′ end of *B. subtilis* 16S rRNA (3), the predicted free energy of base pairing (21) of the putative ribosome-binding site with the rRNA was found to be −15.4 Kcal (ca. −64.4 J), similar to values obtained for *B. subtilis* genes (15).

The phospho-β-galactosidase open reading frame is followed by a sequence with the potential to form a stem-loop structure (position 1659 to 1691; Fig. 3). This structure is similar to transcription terminators described for *B. subtilis* (14, 16). The calculated ΔG value (21) for this stem-loop structure is −15.2 kcal (ca. −63.6 J).

Although the open reading frame is preceded by a ribosome-binding site sequence, there are no good candidates for promoter sequences immediately upstream of the open reading frame. Preliminary DNA sequence analysis indicates that this open reading frame is preceded by at least three additional substantial open reading frames (F. Breidt, Jr., Ph.D. thesis, University of Kansas, Lawrence, 1987). Thus, it appears that the phospho-β-galactosidase message is the terminal protein-encoding sequence on a polycistronic mRNA. A genetic study of the staphylococcal *lac* genes has indicated that the phospho-β-galactosidase gene is tightly linked to and coordinately regulated with the genes for the lactose-specific components of the phosphotransferase system, i.e., enzyme II^{lac} and factor III^{lac} (5). The polycistronic nature of the staphylococcal *lac* mRNA raises the possibility that these genes are part of an operon in *S. aureus*.

The N-terminal amino acid reported for phospho-β-galactosidase purified from *S. aureus* is threonine (6). There is a threonine codon immediately following the ATG initiation codon in the sequence. This suggests that the mature enzyme is processed by having its N-formyl methionine removed.

Codon usage analysis. The amino acid composition and codon usage frequencies for the *S. aureus* phospho-β-galactosidase gene are presented in Table 1. The codon usage is biased toward A- or U-rich codons, reflective of the relatively low G+C content of *S. aureus* DNA (32 to 35 mol%). However, some codons, such as AUA (Ile), which are rarely used in highly expressed genes in *E. coli* (2), are infrequently used in this *S. aureus* message (3.4% of Ile codons), despite their A+U-rich nature. This is consistent with the codon usage seen with the *S. aureus* protein A gene, whereby AUA represents only 7.1% (1 of 14) of the Ile codons (22). However, two other chromosomal genes from *S. aureus* do not show this codon bias. The AUA codon represents 36.4% (4 of 11) and 25% (8 of 32) of the Ile codons of the staphylococcal nuclease (*nuc*) and lipase (*geh*) genes, respectively (8, 18). Therefore, the codon bias may indeed be related to expression levels and may not simply be a reflection of relative levels of isoaccepting tRNA species in the cell.

Phospho-β-galactosidases have been studied in a number of gram-positive bacteria. The primary structure of the *S. aureus* enzyme should prove to be of interest for comparison with the streptococccal enzymes when their amino acid or DNA sequences become available.

TABLE 1. Amino acid composition and codon usage

Amino acid (no.)	Mol%	Codons used
Ala (28)	5.96	GCU, 5; GCC, 3; GCA, 17; GCG, 3
Cys (2)	0.43	UGU, 2; UGC, 0
Asp (39)	8.30	GAU, 34; GAC, 5
Glu (37)	7.87	GAA, 33; GAG, 4
Phe (26)	5.53	UUU, 15; UUC, 11
Gly (39)	8.30	GGU, 26; GGC, 5; GGA, 8; GGG, 0
His (16)	3.40	CAU, 14; CAC, 2
Ile (29)	6.17	AUU, 17; AUC, 11; AUA, 1
Lys (35)	7.45	AAA, 31; AAG, 4
Leu (26)	5.53	UUA, 21; UUG, 3; CUU, 1; CUC, 0; CUA, 1; CUG, 0
Met (9)	1.91	AUG, 9
Asn (21)	4.47	AAU, 13; AAC, 8
Pro (19)	4.04	CCU, 9; CCC, 0; CCA, 9; CCG, 1
Gln (10)	2.13	CAA, 10; CAG, 0
Arg (17)	3.62	CGU, 9; CGC, 2; CGA, 2; CGG, 0; AGA, 4; AGC, 0
Ser (16)	3.40	UCU, 7; UCC, 0; UCA, 5; UCG, 0; AGU, 4; AGC, 0
Thr (23)	4.89	ACU, 5; ACC, 0; ACA, 13; ACG, 5
Val (30)	6.38	GUU, 12; GUC, 3; GUA, 10; GUG, 5
Trp (10)	2.13	UGG, 10
Tyr (38)	8.09	UAU, 28; UAC, 10

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI21574 from the National Institutes of Health and by University of Kansas General Research allocation 3376-XO-0038.

LITERATURE CITED

1. **Breidt, F., Jr., and G. C. Stewart.** 1986. Cloning and expression of the phospho-β-galactosidase gene of *Staphylococcus aureus* in *Escherichia coli*. J. Bacteriol. **166**:1061–1066.
2. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalogue usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.
3. **Green, C. J., G. C. Stewart, M. A. Hollis, B. S. Vold, and K. F. Bott.** 1985. Nucleotide sequence of the *Bacillus subtilis* ribosomal RNA operon *rrnB*. Gene **37**:261–266.
4. **Harlander, S. K., L. L. McKay, and C. F. Schachtele.** 1984. Molecular cloning of the lactose-metabolizing genes from *Streptococcus lactis*. Appl. Environ. Microbiol. **48**:347–351.
5. **Hengstenberg, W., W. K. Penberthy, K. L. Hill, and M. L. Morse.** 1968. Metabolism of lactose by *Staphylococcus aureus*. J. Bacteriol. **96**:2187–2188.
6. **Hengstenberg, W., W. K. Penberthy, and M. L. Morse.** 1970. Purification of the staphylococcal 6-phospho-β-D-galactosidase. Eur. J. Biochem. **14**:27–32.
7. **Inamine, J. M., L. N. Lee, and D. J. LeBlanc.** 1986. Molecular and genetic characterization of lactose-metabolic genes of *Streptococcus cremoris*. J. Bacteriol. **167**:855–862.
8. **Lee, C. Y., and J. J. Iandolo.** 1986. Lysogenic conversion of staphylococcal lipase is caused by insertion of the bacteriophage L54a genome into the lipase structural gene. J. Bacteriol. **166**:385–391.
9. **Lee, L.-J., J. B. Hansen, E. K. Jagusztyn-Krynicka, and B. M. Chassy.** 1982. Cloning and expression of the β-D-phosphogalactoside galactohydrolase gene of *Lactobacillus casei* in *Escherichia coli*. J. Bacteriol. **152**:1138–1146.
10. **Maeda, S., and M. J. Gasson.** 1986. Cloning, expression and location of the *Streptococcus lactis* gene for phospho-β-D-galactosidase. J. Gen. Microbiol. **132**:331–340.
11. **McKay, L. L.** 1980. Regulation of lactose metabolism in dairy streptococci, p. 153–18. *In* R. Davies (ed.), Developments in food microbiology, vol. 1. Applied Science Publishers, Ltd., Essex, United Kingdom.
12. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol. Chem. **256**:11283–11291.
13. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
14. **Mongkolsuk, S., E. J. Duvall, and P. S. Lovett.** 1985. Transcription termination signal for the *cat*-86 indicator gene in a *Bacillus subtilis* promoter-cloning plasmid. Gene **37**:83–90.
15. **Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186**: 339–346.
16. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
18. **Shortle, D.** 1983. A genetic system for analysis of staphylococcal nuclease. Gene **22**:181–189.
19. **Snook, R. J., and L. L. McKay.** 1981. Conjugal transfer of lactose-fermenting ability among *Streptococcus cremoris* and *Streptococcus lactis* strains. Appl. Environ. Microbiol. **42**:904–911.
20. **St. Martin, E. J., L. N. Lee, and D. J. LeBlanc.** 1982. Genetic analysis of carbohydrate metabolism in streptococci, p. 232–233. *In* D. Schlessinger (ed.), Microbiology—1982. American Society for Microbiology, Washington, D.C.
21. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
22. **Uhlén, M., B. Guss, B. Nilsson, S. Gatenbeck, L. Philipson, and M. Lindberg.** 1984. Complete sequence of the staphylococcal gene encoding protein A: a gene evolved through multiple duplications. J. Biol. Chem. **259**:1695–1702.