JOURNAL OF BACTERIOLOGY, Apr. 1993, p. 2475–2478
0021-9193/93/082475-04$02.00/0

Vol. 175, No. 8

Cloning and Sequence Determination of the *valS* Gene, Encoding Valyl-tRNA Synthetase in *Lactobacillus casei*

BRIAN V. TAYLOR, JEFFREY TOY, TIM L. SIT, AND ANDREW L. BOGNAR*

Department of Microbiology, University of Toronto, 150 College Street, Toronto, Ontario, Canada M5S 1A8

Received 20 November 1992/Accepted 3 February 1993

The DNA sequence of the *valS* gene from *Lactobacillus casei* and the predicted amino acid sequence of its valyl-tRNA synthetase product have been determined. An open reading frame coding for a protein of 901 amino acids was found. A clone containing the intact *L. casei valS* gene functionally complemented the temperature-sensitive growth of the *valS* mutant strain 236c of *Escherichia coli*. The *valS* gene and the downstream folylpolyglutamate synthetase gene are transcribed in the same direction but are separated by a putative transcription terminator.

Amino acyl-tRNA synthetases catalyze the attachment of amino acids with their cognate tRNAs. The structures of these enzymes are quite diverse, but they can be divided into two classes which are distinguished by consensus amino acid sequences (6–8). Class I synthetases have the consensus sequences HIGH and KMSKS, associated with the Rossmann fold for binding ATP (16). These enzymes aminoacylate the 2′ OH of the terminal nucleotide of the tRNA. Class II synthetases have up to three sequence motifs in common, distinct from those of class I enzymes, and they aminoacylate the 3′ OH of the terminal nucleotide (6). Within the class I synthetases, valyl-tRNA synthetases (VRS) have been shown to have additional sequence similarity beyond the general consensus sequences to isoleucyl-tRNA synthetases and, to a lesser extent, to the leucyl- and methionyl-tRNA synthetases (2, 12). Of these last two proteins, a high-resolution X-ray structure has been determined for the methionyl-tRNA synthetase (5).

We have recently cloned the *Lactobacillus casei* gene encoding folylpolyglutamate synthetase (FPGS) (23). To determine whether there are coregulated genes or genes encoding enzymes with related functions, we have sequenced over 2.7 kb of the genomic DNA located upstream of the FPGS gene. A long open reading frame was found in this region. Comparison of the DNA and predicted amino acid sequences with those found in the GenBank data base revealed that the gene product has a high degree of homology to the product of the VRS gene of *Bacillus stearothermophilus* (VRSBS) and a lesser degree of homology to the product of the *Escherichia coli* VRS gene (VRSEC). In this study, we report the cloning, nucleotide sequence, and expression of the *L. casei valS* (VRSLC) gene and comparison of the predicted amino acid sequence with those of other VRS.

Cloning and sequencing of the *valS* gene. We have previously reported the DNA sequence of the *L. casei* FPGS gene, as well as 650 bp of upstream sequence (23). This sequence was obtained from plasmid pGT3, a 3.7-kbp *Bam*HI-*Pst*I subclone of pGT1. Plasmid pGT1 contains 5 kbp of *L. casei* genomic DNA (Fig. 1) (23). The DNA sequence of the entire pGT1 plasmid insert was determined by the dideoxynucleotide termination method described by Sanger et al. (21) with α-^{35}S-dATP (Amersham) and T7 DNA polymerase (Sequenase) and with double-stranded plasmid DNA templates prepared by the procedure described by Kraft et al. (14). Residues 1453 to 3261, indicated in Fig. 2, are derived from pGT1. Analysis of this sequence revealed an open reading frame of 1.7 kb, and comparison of the predicted amino acid sequence with those in the GenBank data base, performed with the MacVector program (IBI), showed a high degree of homology with VRSBS (2). Comparison of the two sequences predicted that the *L. casei* gene was incomplete, lacking 950 bp at the 5′ end. To isolate the remainder of the gene, we electroeluted a 400-bp fragment from the 5′ end of pGT1 (corresponding to the *Nco*I-*Bam*HI fragment shown in Fig. 1) from a polyacrylamide gel for use as a probe. The probe was labelled with [α-^{32}P]dCTP or [α-^{32}P]dATP by the random-primer procedure (9), with the Klenow fragment of DNA polymerase I, and the probe was used to screen our *L. casei* genomic library (23) by colony blot hybridization (18). DNA was transferred to nitrocellulose or nylon membranes and hybridized as described by Southern (22) at 65°C. One positive clone was identified (pGT5; Fig. 1). This clone was analyzed by restriction mapping and sequencing from both ends of the insert. An *Eco*RI-*Nco*I fragment from pGT5 was used to rescreen the *L. casei* genomic library. Two additional clones were obtained (pD5 and pC10; Fig. 1), and of these, pD5 was found to contain the full-length *valS* gene, as well as the full-length FPGS gene.

The strategy used to determine the DNA sequence of the *valS* gene is shown in Fig. 1b. Sequence information was obtained from the ends of the plasmids described above as well as from subclones obtained by deletion at the restriction sites indicated in Fig. 1. All sequences shown were obtained for both strands, except bases 1 to 156, corresponding to the 5′ untranslated region of the *valS* gene, which were determined for only one strand with pC10 as the template. The open reading frame corresponding to the *valS* gene contains 2,703 bp and predicts a protein product of 901 amino acids (Fig. 2) with a predicted molecular mass of 103,143 Da. Of the sequence shown, 481 bp at the 3′ end have been reported previously (23).

Analysis of plasmid-encoded gene products in maxicells. The full-length clone, pD5, and pUC13 (24) were transformed into *E. coli* DR1984 (*recA uvrB*) (19) by the method described by Hanahan (11), and the plasmid-dependent

* Corresponding author.

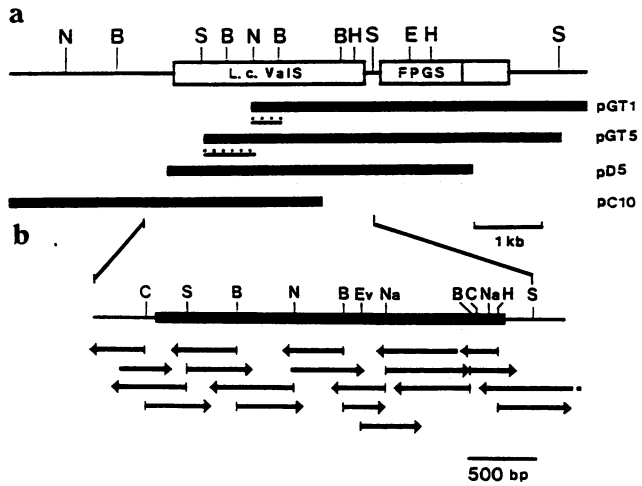

FIG. 1. Cloning, restriction map, and sequencing strategy of the *valS* gene. (a) The upper section is a restriction map of the segment of *L. casei* genomic DNA contained in the overlapping clones shown below as solid bars. The large open boxes represent open reading frames, and names are given inside the boxes for those products whose functions are known. The narrow bars with broken lines above them represent DNA fragments of the clones shown above that were used as probes to isolate the clones shown below. (b) The region of *L. casei* DNA that was sequenced is shown with an expanded restriction map. The solid bar represents the open reading frame of the *valS* gene. The solid arrows show stretches of sequence determined from single templates. Arrows beginning at vertical bars represent sequences determined from plasmids deleted at the restriction site denoted by the bar. If the vertical bar is not present, the sequence was determined from the end of a subclone (upper section) originating at a *Sau*3AI site not shown on the map. The arrow at right originating at a solid square represents sequence determined from a specific primer which was previously described (23). Restriction enzyme abbreviations are as follows: N, *Nco*I; B, *Bam*HI; S, *Sph*I; H, *Hin*dIII; E, *Eco*RI; Na, *Nar*I; Ev, *Eco*RV; C, *Cla*I.

proteins were labelled with [^{35}S]methionine and cysteine (Translabel; ICN Pharmaceuticals) according to the method described by Sancar et al. (20). Gel electrophoresis with 13% polyacrylamide gels was performed according to the method described by Laemmli (15), and the protein bands were visualized by fluorography. A labeled protein band with an M_r of 100,000 was observed (Fig. 3, lanes 3 and 4), in good agreement with the value predicted for VRSLC; in addition, bands with M_r values of 43,000 (FPGS) and 30,000 (β-lactamase) were observed.

Complementation studies. Plasmid pD5 was transformed into *E. coli* strain 236c, which is temperature sensitive for VRS and is unable to grow at 42°C (3), and into *E. coli* SF4 (*folC strA recA* Tn*10*::*srlC*), which is deficient in folylpolyglutamate synthetase-dihydrofolate synthetase (1, 10), to test for complementation of *valS* and *folC* deficiencies, respectively. Strain 236c/pD5 was able to grow at 42°C, indicating that the plasmid expressed a functional VRS. Strain SF4/pD5 grew on Vogel-Bonner minimal medium (25) lacking methionine, indicating that FPGS was also expressed. The direction of transcription of both *L. casei* genes in this plasmid was opposite to that of the *lac* promoter, suggesting that both genes are transcribed from endogenous promoters. A putative promoter upstream of the FPGS gene has been previously identified by primer extension experiments, although it lacks a recognizable −35 sequence (23). Figure 2 shows that a canonical −10 promoter sequence is found at −278 to −273 relative to the predicted ATG start codon of the *valS* gene. The sequence TTGACC at −301 to −296 contains five of the six consensus bases for the −35 promoter sequence, with an optimal spacing of 17 bp between this sequence and the −10 promoter sequence. These consensus sequences suggest a strong promoter, but we have not tested these predictions experimentally, except for the expression studies in *E. coli* described above. A region of dyad symmetry predicting a stem-loop structure followed by seven T residues strongly resembling a *rho*-independent transcription terminator is located 77 to 45 bp upstream of the predicted start codon for the *valS* gene, between the putative promoter and the start codon. This may attenuate the expression of the *valS* gene. We have not investigated this possible regulation, but no open reading frames that might correspond to short leader peptides are present in the 5′ untranslated region overlapping this stem-loop structure. A second stem-loop structure followed by six T residues is located 39 bp downstream of the predicted stop codon for the *valS* gene and may constitute the transcription terminator for the gene. Our results suggest that the *valS* gene is a single transcriptional unit, although we have not tested for termination experimentally.

Amino acid similarity comparisons. We have compared the predicted amino acid sequence of VRSLC with those of previously reported genes encoding VRSBS (2), VRSEC (12), and VRSSC, the product of the *VAS1* gene from *Saccharomyces cerevisiae* (13), by using the MacVector program (IBI) and manual adjustment. Alignments were also based on a previously published comparison of VRS (12). Those residues of the VRSLC which are identical in all four VRS proteins are shown as shaded residues in Fig. 2. The overall amino acid identity among all four synthetases is 25%. The highest degree of similarity is 60% identical amino acid residues with VRSBS, compared with 41% identity with VRSEC and 38% with VRSSC. If functionally conservative changes are considered, the similarity is 69% for VRSBS, 54% for VRSEC, and 49% for VRSSC (data not shown). The VRSLC is clearly more closely related to the VRSBS than to the other VRS proteins, both from the overall higher number of identical residues and from the fact that no deletions or insertions need be postulated to align the proteins, compared with several that are required to align them with the *E. coli* and yeast enzymes (data not shown).

The Rossmann nucleotide binding fold is the characteristic structural feature of class I amino acyl-tRNA synthetases, which include the VRS enzymes (8). The X-ray crystal structure of the large fragment of *E. coli* methionyl-tRNA synthetase (MRSEC) bound to ATP (5) shows that the nucleotide binding fold consists of five parallel β-strands surrounded by five α-helices. The signature sequences, HIGH and KMSKS, are both located on loops that contact the binding fold. The VRSEC and MRSEC enzymes have an overall amino acid identity of 20% and a functional similarity of about 40% per unit length (12). The VRS enzymes have blocks of conserved residues, which can be used to align them with homologous regions of MRSEC. The loop containing the HMGH sequence of the VRSLC, which is homologous to the HIGH consensus sequence, is preceded by residues with a high probability of forming a β-sheet and is followed by an α-helix and another β-sheet in the same relative positions as similar structures found in MRSEC. The α-helix and the loop contain conserved amino acid residues that can be aligned with MRSEC; however, the β-sheets do not, although these β-sheets are found in identical positions. This suggests that the α and β structures making up the

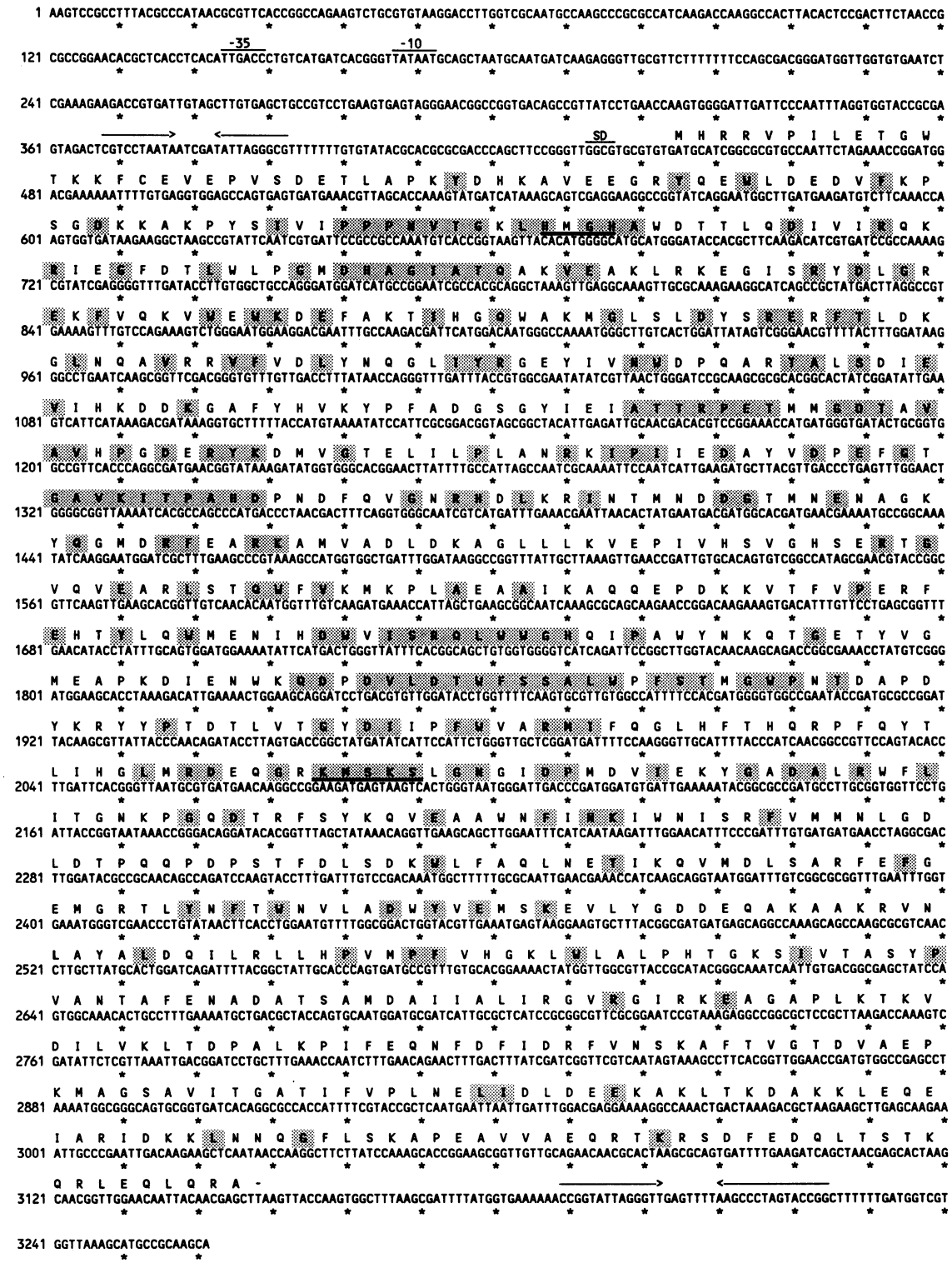

FIG. 2. Nucleotide sequence of the *L. casei valS* gene and flanking regions. The deduced amino acid sequence of the VRS is shown above the nucleotide sequence. The consensus −10 and −35 putative promoter sequences 5′ to the open reading frame are overlined. The deduced Shine-Dalgarno sequence is indicated by overlining. Potential terminator sequences in the 5′ and 3′ noncoding regions are indicated by converging arrows. The termination codon is indicated with a dash. Amino acid residues that are identical to those in the other three reported *valS* genes, VRSBS, VRSEC, and VRSSC, are shaded. The consensus HIGH and KMSKS sequence motifs for class I tRNA synthetases are double underlined.

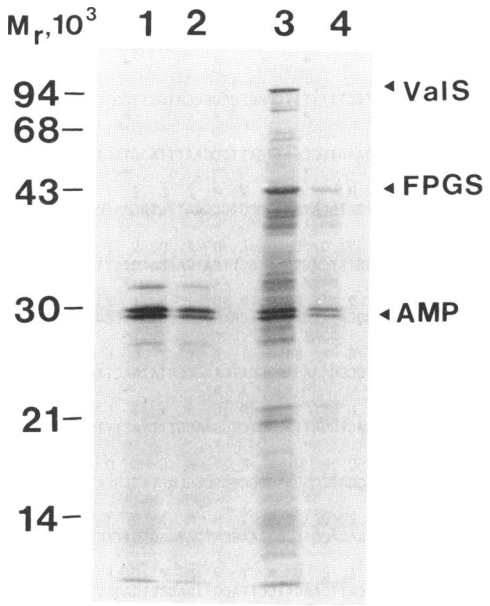

FIG. 3. Expression of plasmid-dependent proteins in maxicells. *E. coli* DR1984 cells were transformed with plasmid D5 (lanes 3 and 4) or pUC13 (lanes 1 and 2), and extracts were analyzed by electrophoresis on sodium dodecyl sulfate-13% polyacrylamide gels and by autoradiography. The positions of the gene products encoded by the plasmids are indicated by arrowheads at right. The positions of coelectrophoresed standards are indicated by their molecular weights at left.

nucleotide binding fold of the two enzymes are conserved. The crystal structures of two other class I synthetases, the tyrosyl- and glutamyl-tRNA synthetases, have been determined and display the same structure around the HIGH sequences (4, 17). The KMSKS-containing loop in MRSEC is flanked by β-strands. The predicted structure at the downstream β-strand position is conserved in VRSLC, although there is no amino acid similarity. However, large blocks of well-conserved sequence in the VRS enzymes are absent from the MRSEC, such as that between residues 250 and 420 of VRSLC, suggesting that there are significant differences in the overall structures of the two proteins.

We thank A. R. Fersht for the gift of *E. coli* 236c. We thank J. B. Campbell for the use of computer programs.

This work was supported by grant MT-9822 from the Medical Research Council of Canada.

REFERENCES

1. **Bognar, A. L., C. Osborne, B. Shane, S. Singer, and R. Ferone.** 1985. *Escherichia coli* folylpolyglutamate synthetase-dihydrofolate synthetase, gene cloning and high expression, purification and properties. J. Biol. Chem. **260**:5625–5630.
2. **Borgford, T. J., N. J. Brand, T. E. Gray, and A. R. Fersht.** 1987. The valyl-tRNA synthetase from *Bacillus stearothermophilus* has considerable sequence homology with the isoleucyl-tRNA synthetase of *Escherichia coli*. Biochemistry **26**:2480–2486.
3. **Brand, N., and A. R. Fersht.** 1986. Molecular cloning of the gene encoding the valyl-tRNA synthetase from *Bacillus stearothermophilus*. Gene **44**:139–142.
4. **Brick, P., and D. M. Blow.** 1987. Crystal structure of a deletion mutant of tyrosyl-tRNA synthetase complexed with tyrosine. J. Mol. Biol. **194**:287–297.
5. **Brunie, S., C. Zelwer, and J.-L. Risler.** 1990. Crystallographic study at 2.5 Å resolution of the interaction of methionyl-tRNA synthetase from *Escherichia coli* with ATP. J. Mol. Biol. **216**:411–424.
6. **Cusak, S., C. Berthet-Colominas, M. Hartlein, N. Nassar, and R. Leberman.** 1991. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. Nature (London) **347**:249–255.
7. **Cusak, S., M. Hartlein, and R. Leberman.** 1991. Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. Nucleic Acids Res. **19**:3489–3498.
8. **Eriani, G., M. Delarue, O. Poch, J. Gangloff, and D. Moras.** 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature (London) **347**:203–206.
9. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
10. **Ferone, R., S. C. Singer, M. H. Hanlon, and S. Roland.** 1983. Isolation and characterization of an *E. coli* mutant affected in dihydrofolate- and folylpolyglutamate-synthetase, p. 585–589. *In* J. A. Blair (ed.), Chemistry and biology of pteridines. de Gruyter, Berlin.
11. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
12. **Heck, J. D., and G. W. Hatfield.** 1989. Valyl-tRNA synthetase gene of *Escherichia coli*. Primary structure and homology within a family of aminoacyl-tRNA synthetases. J. Biol. Chem. **263**:868–877.
13. **Jordana, X., B. Chatton, M. Paz-Weisshaar, J.-M. Buhler, F. Cramer, J. P. Ebel, and F. Fasiolo.** 1987. Structure of yeast valyl-tRNA synthetase (VAS1) and the homology of its translated amino acid sequence with *Escherichia coli* isoleucyl-tRNA synthetase. J. Biol. Chem. **262**:7189–7194.
14. **Kraft, R., J. Tardiff, K. S. Krauter, and L. A. Lienwand.** 1988. Using mini-prep plasmid DNA for sequencing double-stranded templates with Sequenase. BioTechniques **6**:544–546.
15. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
16. **Rossmann, M. G., A. Liljas, C. V. Branden, and L. J. Banaszak.** 1976. Evolutionary and structural relationships among dehydrogenases, p. 61–102. *In* P. D. Boyer (ed.), The enzymes, 3rd ed., vol. 11. Academic Press, New York.
17. **Rould, M. A., J. J. Perona, D. Söll, and T. A. Steitz.** 1989. Structure of *Escherichia coli* glutamyl-tRNA synthetase complexed with tRNA (Gln) and ATP at 2.8 Å resolution. Science **246**:1135–1142.
18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
19. **Sancar, A., and W. D. Rupp.** 1979. Cloning of the *uvrA*, *lexC*, and *ssb* genes of *Escherichia coli*. Biochem. Biophys. Res. Commun. **90**:123–129.
20. **Sancar, A., R. P. Wharton, S. Seltzer, B. M. Kasinsky, N. D. Clarke, and W. D. Rupp.** 1981. Identification of the *uvrB* gene product. J. Mol. Biol. **148**:45–62.
21. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
22. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
23. **Toy, J., and A. L. Bognar.** 1990. Cloning and expression of the gene encoding *Lactobacillus casei* folyl-γ-glutamate synthetase in *Escherichia coli* and determination of its primary structure. J. Biol. Chem. **265**:2492–2499.
24. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
25. **Vogel, H. J., and D. M. Bonner.** 1956. Acetyl ornithine of *Escherichia coli*: partial purification and some properties. J. Biol. Chem. **218**:97–106.