JOURNAL OF BACTERIOLOGY, Sept. 1981, p. 787-796
0021-9193/81/090787-10$02.00/0

Vol. 147, No. 3

Nucleotide Sequence of the Structural Gene for Tryptophanase of *Escherichia coli* K-12

MICHAEL C. DEELEY AND CHARLES YANOFSKY*

Department of Biological Sciences, Stanford University, Stanford, California 94305

Received 23 March 1981/Accepted 1 June 1981

The tryptophanase structural gene, *tnaA*, of *Escherichia coli* K-12 was cloned and sequenced. The size, amino acid composition, and sequence of the protein predicted from the nucleotide sequence agree with protein structure data previously acquired by others for the tryptophanase of *E. coli* B. Physiological data indicated that the region controlling expression of *tnaA* was present in the cloned segment. Sequence data suggested that a second structural gene of unknown function was located distal to *tnaA* and may be in the same operon. The pattern of codon usage in *tnaA* was intermediate between codon usage in four of the ribosomal protein structural genes and the structural genes for three of the tryptophan biosynthetic proteins.

Tryptophanase is found in bacteria exclusively and is primarily present in those species indigenous to the intestinal tracts of animals (10). Although the importance of tryptophanase to bacteria is not completely understood, what is known about its catalytic activities and formation suggests that it is a catabolic enzyme. Tryptophanase has many enzymatic activities. In addition to catalyzing tryptophan degradation (21, 28), resulting in stochiometric production of indole, pyruvate, and ammonia, tryptophanase can synthesize tryptophan from indole and serine (21, 28, 32). It is this latter reaction that provides the basis for the genetic selection of *tnaA*$^+$ (33); i.e., in mutant strains that cannot produce functional tryptophan synthetase β chains, tryptophanase is the only enzyme the bacterium can form that allows utilization of indole as a substitute for tryptophan. In *Escherichia coli*, synthesis of tryptophanase is inducible by tryptophan (30). Genetic and physiological studies have established that cyclic AMP and the catabolite gene activator protein (CAP) are required for tryptophanase expression (4, 31). To study the structure and regulation of *tnaA*, we cloned and sequenced this structural gene and its surrounding regions.

For a complete review of the properties of tryptophanase, see the article by Snell (28).

MATERIALS AND METHODS

Media. The following medium allows growth of *trpB* strains that are *tnaA*$^+$ (33): Vogel and Bonner minimal salts (30)-10 μg of indole per ml-50 μg of DL-5-methyltryptophan per ml-0.2% glycerol. Addition of 0.5% acid-hydrolyzed casein to this medium promotes rapid growth. Difco antibiotic medium no. 3 and nutrient broth were also used.

Bacterial strains. The bacterial strains employed in these studies are listed in Table 1.

Procedure for in vivo cloning of *tnaA*$^+$, using Mu-mediated transposition onto conjugative plasmid RP4. Cultures of strains 678(RP4) and MD42 were grown overnight at 30 and 37°C, respectively, in antibiotic medium no. 3, diluted at least 10-fold in the same medium, and grown to a density of about 5×10^8 cells per ml. One milliliter of the Trp$^+$ donor culture [678(RP4)] and 2 ml of the Trp$^+$ recipient culture (MD42) were mixed and collected by filtration on a sterile 0.45-μm membrane filter (Millipore Corp.). The filter was then placed on a prewarmed (42°C) nutrient agar plate and incubated for 2 h. The mating cells were eluted and washed twice with 0.8% saline and then plated on medium selective for *tnaA*$^+$ and containing 20 μg of kanamycin per ml or 15 μg of tetracycline per ml. The resulting Trp$^+$ colonies were isolated and tested, as described in Results, to determine which harbored *tnaA*$^+$ plasmids and which harbored *trpB*$^+$ plasmids.

Purification of RP4::Mu *c*(Ts)::*tnaA*$^+$ plasmids. Isolation of the DNA of RP4::Mu *c*(Ts)::*tnaA*$^+$ plasmids proved difficult, presumably owing to their large size. The following plasmid isolation procedure was found satisfactory. One liter of cells was grown in *tnaA*$^+$ selective medium plus 15 μg of tetracycline per ml at 37°C to the stationary phase. Cells were harvested at 4°C by centrifugation, and cleared lysates were prepared by the method of Katz et al. (17) with the following modifications. Spheroplasts were lysed with Triton X-100 at 37°C for 30 to 45 min before centrifugation; the cleared lysate (diluted equally with water) was treated with 100 μg of proteinase K per ml for 1 h at 37°C to degrade protein; the DNA was precipitated overnight with polyethylene glycol (14). Equilibrium centrifugation of the dissolved precipitate in a cesium chloride-ethidium bromide density gra-

TABLE 1. *Bacterial strains*

Strain[a]	Relevant genotype	Source or reference
MD33	W3110 F⁻ Δ(*trpEA*)*2 tnaA2*	C. Yanofsky
MD35	W3110 F⁻ Δ(*trpEA*)*2 tnaA2 ilv*::Tn*10*[b]	Transductant of MD33 with phage grown on MC4100
MC4100	F⁻ *ara ilv*::Tn*10*[b]	B. Bachmann
MD50	F⁻ Δ(*trpEB*)*9 thr leu thi*	C. Yanofsky
678(RP4)	F⁻ *his arg purE aroD rpsL fep*::Mu *c*(Ts) (RP4)[c]	Laird et al. (19)
MD37	W3110 F⁻ Δ(*trpEA*)*2 tnaA2* (Mu⁺)	Lysogeny of MD33 with Mu⁺ phage
MD39	W3110 F⁻ Δ(*trpEA*)*2 tnaA2* (Mu⁺) *rpsL*	UV mutagenesis of MD37 followed by selection for streptomycin resistance
MD42	W3110 F⁻ Δ(*trpEA*)*2 tnaA2* (Mu⁺) *rpsL gyrA recA1*	Recombinant from cross MD39 × KL163
MD54	F⁻ Δ(*trpEB*)*9 thr-1 leu-6 thi-1 tnaA2 ilv*::Tn*10*[b]	Transductant of MD50 with phage grown on MD35
MD55	F⁻ Δ(*trpEB*)*9 thr-1 leu-6 thi-1 tnaA2 ilv*::Tn*10*[b] *gyrA recA1*	Recombinant from cross MD54 × KL163
MD56	F⁻ Δ(*trpEB*)*9 thr-1 leu-6 thi-1 tnaA2*	Transductant of MD54 with phage grown on MD33
KL163	Hfr *gyrA thyA25 deoB recA1*	B. Bachmann

[a] All strains are *E. coli* K-12.

[b] Tn*10* confers resistance to tetracycline.

[c] RP4 confers resistance to tetracycline, ampicillin, and kanamycin.

dient resolved two distinct DNA bands. The lower band was removed, extracted with isopropyl alcohol saturated with 10 mM Tris-hydrochloride (NaCl saturated; pH 7.4 at 4°C) containing 0.1 mM EDTA (TE) to remove ethidium bromide, and dialyzed against TE overnight at 4°C to remove salts. DNA concentrations were estimated spectrophotometrically at 260 nm.

Cloning procedures. The procedures used for restriction enzyme digestion, ligation, bacterial transformation, isolation of plasmid DNAs smaller than RP4, and isolation of DNA fragments from polyacrylamide gels have been described previously (25).

Coupled transcription-translation analysis of plasmids. The coupled system of Zubay et al. (34) as described by Gunsalus et al. (13) was used for transcription-translation analysis of the plasmids. [^{35}S]methionine was used to label the polypeptides encoded by the template DNAs and produced in the S30 extracts.

DNA sequence determination. DNA fragments were end-labeled with [γ-^{32}P]ATP (>2,500 Ci/mmol; ICN Pharmaceuticals) and polynucleotide kinase (P-L Biochemicals) as described by Maxam and Gilbert (20). Labeled fragments were denatured in 0.3 N NaOH–0.1 mM EDTA in 10% glycerol at 95°C for 30 s, immediately chilled in ice water for 30 s, and rapidly loaded onto 45 mM Tris-borate (pH 8.1 at 20°C)–1 mM EDTA–5 to 8% polyacrylamide gels. Electrophoresis at either 4 or 20°C resolved single-stranded DNA (20). Five separate sequencing reactions were performed on the labeled single-stranded DNAs: G, A>G, A>C, C, and C+T. The polyacrylamide-urea gel system of Sanger and Coulson (24) was used to display sequence ladders. DNA sequences were analyzed in part with the aid of the computer program of Korn et al. (18). Programs and facilities were courteously provided by the Stanford Molgen Project and the National Institutes of Health SUMEX-AIM facility.

Assays. Tryptophanase assays of crude cell extracts employed the artificial substrate *S-o*-nitrophenyl-L-cysteine, as outlined by Suelter et al. (29). This substrate was synthesized by the method of Boyland et al. (5). The following procedure was used to measure tryptophanase specific activities. Each cell pellet was suspended in 2 ml of lysis buffer (0.1 M KPO_4 [pH 7.8 at 4°C], 50 μM pyridoxal 5-phosphate, 10 mM 2-mercaptoethanol, 2 mM EDTA), sonicated, and centrifuged for 30 min at 9,000 × *g*. Five- to 50-μl portions of the resulting supernatants were added to 0.95 ml of the incubation mixture (100 mM KPO_4 [pH 7.8 at 30°C], 100 μM pyridoxal 5-phosphate, 20 mM 2-mercaptoethanol, 1 mg of bovine serum albumin per ml). The volume was then adjusted to 1 ml with lysis buffer. *S-o*-nitrophenyl-L-cysteine degradation by tryptophanase was initiated by the addition of 1 ml of 0.66 mM *S-o*-nitrophenyl-L-cysteine in water (at 30°C), and the reaction was allowed to proceed for 10 min at 30°C. The reaction was terminated by the addition of 0.2 ml of 1 N NaOH. The extent of tryptophanase action was estimated by measuring the increase in absorbance at 420 nm. One unit of tryptophanase is defined as the amount of enzyme that produces an increase in absorbance of 1.0 at 470 nm in 10 min at 30°C. Protein concentrations were determined by the Bio-Rad protein assay procedure (6). Bovine serum albumin was used as the protein standard.

RESULTS

tnaA was initially cloned in vivo by using the method of Laird et al. (19). The recipient bacterium in these experiments, MD42, is Trp⁻ and unable to utilize indole as an L-tryptophan source. The Trp⁺ donor bacterium 678(RP4) does not survive the temperature induction of lysogenic Mu *c*(Ts), which it contains, whereas MD42, lysogenic for Mu⁺, does survive. When these strains are mated at 42°C, only MD42 cells which acquire an RP4 derivative bearing *tnaA*⁺ or *trpB*⁺ will survive in a medium containing an antibiotic to which RP4 confers resistance and indole as the sole source of L-tryptophan. When these strains were mated under the conditions described above and the exconjugants were plated upon *tnaA*⁺ selective medium containing 15 μg of tetracycline per ml, colonies appeared at a frequency of $5/10^9$ recipient cells. Twenty-five colonies from these matings were picked and purified by being streaked on the same selective medium. We distinguished cloned *tnaA*⁺ isolates from *trpB*⁺ isolates by testing each isolate's abilities to grow on two media.

Strains which contain normally regulated *tnaA*$^+$ require tryptophan or 5-methyltryptophan for induction of tryptophanase activity (2, 33). The addition of glucose to the medium severely represses tryptophanase induction (4, 31, 33). *trpB*$^+$ *tnaA* strains do not exhibit these regulatory characteristics. Therefore, isolates requiring 5-methyltryptophan for growth on medium containing indole as a source of L-tryptophan and which are unable to grow when glucose is substituted for glycerol as a carbon source have *tnaA*$^+$ rather than *trpB*$^+$ on the plasmid (2, 4, 31, 33). Four isolates with the correct characteristics were retained.

To prove that *tnaA*$^+$ was indeed inserted into RP4, plasmid DNA was isolated from the four strains and used to transform Trp$^-$ *tnaA* strain MD42 to Trp$^+$. Each of the DNAs gave multiply drug-resistant Trp$^+$ transformants. These transformants exhibited the same phenotypic characteristics as the original isolates.

*Eco*RI fragments of one of the *tnaA*$^+$ plasmids were ligated into the single *Eco*RI site of plasmid pACYC184. Transformation of Trp$^-$ *tnaA* strain MD56 with this ligation mixture yielded 12 colonies that were tetracycline resistant and Trp$^+$. When the plasmids in these transformants were examined, all possessed an *Eco*RI fragment approximately 8,500 base pairs (bp) in length, in addition to the 4-kilobase vector. DNA of one of the recombinant plasmids, designated pMD1, and pACYC184 DNA were used in subsequent transformations of strain MD56 to verify the association of the Trp$^+$ phenotype with the 8,500-bp insert. Plasmid DNA was prepared from a pMD1 transformant, and a restriction map was constructed (Fig. 1).

To localize *tnaA* within the cloned *Eco*RI fragment, *Hin*dIII and *Bam*HI fragments of pMD1 were ligated into pBR322 vectors, respec-

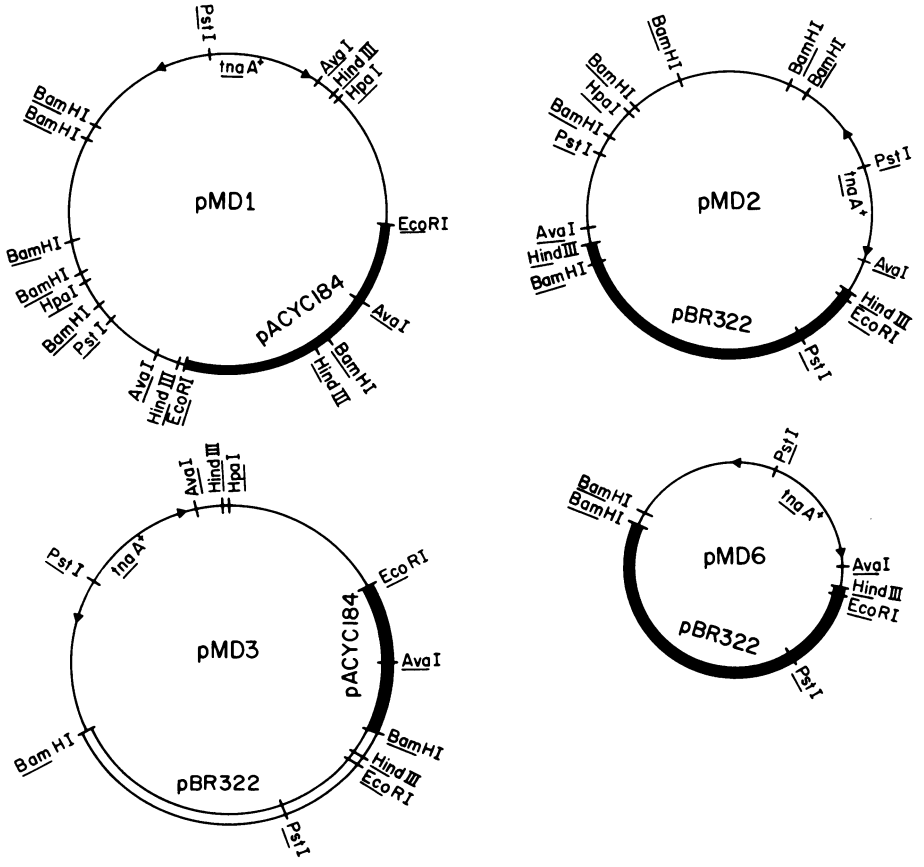

FIG. 1. *Restriction maps of plasmids pMD1, pMD2, pMD3, and pMD6. The restriction endonuclease sites are indicated. The heavy lines represent vector sequences; the light lines correspond to chromosomal DNA. The location of tnaA is indicated by arrowheads. pMD2 was derived from pMD1 by digestion with HindIII and ligation into the single HindIII site of pBR322. pMD3 was derived from pMD1 by digestion with BamHI and ligation into the single BamHI site of pBR322. pMD6 was derived from pMD1 by digestion with HindIII plus BamHI and ligation into identically restricted pBR322.*

tively restricted with *Hin*dIII and *Bam*HI. The mixtures were used to transform Trp⁻ *tnaA recA* strain MD55 to Trp⁺. Two plasmids containing *Hin*dIII fragments were found that conferred the Trp⁺ phenotype. Each had the same 6,100-bp fragment insert. They differed only in the orientation of the insert in the vector. They were designated pMD2 and pMD2R. pMD3, a plasmid containing a 5,600-bp *Bam*HI segment from pMD1, was also found to confer the Trp⁺ phenotype. Neither *Pst*I fragment from pMD1, when recloned into pBR322, conferred the Trp⁺ phenotype, suggesting that there is a *Pst*I site within *tnaA*.

Inspection of the restriction maps of pMD2 and pMD3 localized the tryptophanase structural gene to a 3,200-bp *Hin*dIII/*Bam*-HI segment of pMD1 (Fig. 1). As predicted, a *Pst*I site was present within this region. To construct a plasmid containing only the *tnaA*⁺ segment from pMD1, pMD1 and pBR322 were digested with *Hin*dIII and *Bam*HI, a mixture of these fragments was ligated, and the resulting products were used to transform Trp⁻ MD55 to Trp⁺. When the plasmid DNAs from these Trp⁺ transformants were examined, each was found to contain the 3,200-bp segment from pMD1.

To determine whether *tnaA*⁺ on pMD6 was regulated normally, strain MD55(pMD6) was tested for its ability to produce tryptophanase when exposed to increasing levels of L-tryptophan in glycerol medium and for its sensitivity to glucose in a medium containing high levels of tryptophan (120 μg/ml). Tryptophanase specific activity increased with increasing inducer (L-tryptophan) concentration (Fig. 2). Tryptophanase expression in this strain was severely repressed by glucose (Fig. 2).

To find the *tnaA*-reading frame within pMD6, we prepared a detailed restriction map of the plasmid and sequenced two *Hin*fI segments around the single *Pst*I site within the insert (Fig. 3). A comparison of the deduced amino acid sequences from the determined DNA sequences with the protein structure data of the *E. coli* B tryptophanase (15) allowed us to identify the *tnaA*-reading frame. As predicted, a *Pst*I recognition sequence existed within *tnaA*, 620 bp downstream from the first codon.

Figure 3 presents the sequencing strategy employed in the determination of the entire *tnaA* sequence and its flanking regions. The nucleotide sequence of each strand of DNA was determined at least once. Figure 4 presents the nucleotide sequence of *tnaA*, together with the deduced amino acid sequence. The corresponding sequence of the tryptophanase of *E. coli* B (15) is also presented. The equivalent of four large K-12 tryptic fragments are unidentified in the *E. coli* B tryptic fragment catalog. Table 2 compares the deduced amino acid composition of the K-12 tryptophanase to that experimentally determined with the enzyme of the B strain. Table 3 summarizes codon usage in *tnaA* of *E.*

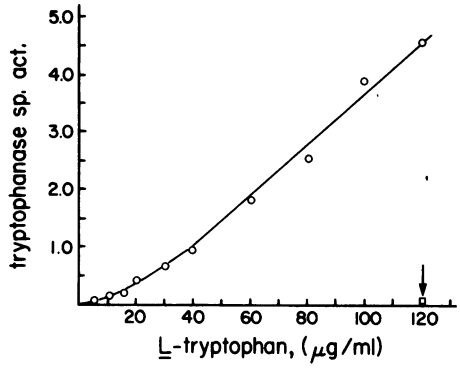

FIG. 2. *Induction of tryptophanase in MD55-(pMD6) grown with various* L*-tryptophan concentrations. An overnight culture of MD55(pMD6) grown in the presence of 10 μg of* L*-tryptophan per ml was diluted 100-fold and used to inoculate a flask containing 200 ml of minimal medium plus 0.2% glycerol, 0.05% acid-hydrolyzed casein,* L*-tryptophan (5 to 120 μg/ml), and 30 μg of ampicillin per ml. (The arrow indicates the value obtained with cells grown with 120 μg of* L*-tryptophan per ml and supplemented as mentioned above except that 0.2% glucose replaced glycerol.) Cells were harvested at a density of 4 × 10⁸ to 6 × 10⁸/ml, chilled, and washed once with 0.8% saline. Assays were performed as described in the text. Specific activity (sp. act.) was measured as the increase in absorbance at 470 nm/10 min per mg of protein.*

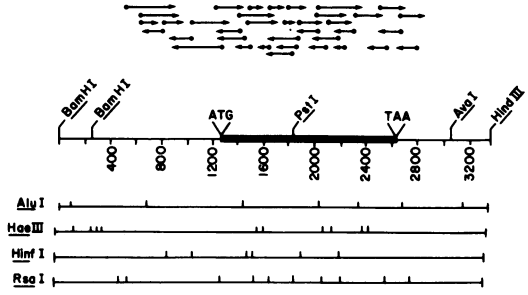

FIG. 3. *Restriction map of the pMD6 insert and sequencing strategy. The location of restriction sites for four of the restriction endonucleases used are displayed in the bottom four lines. The numbers on the top line refer to the number of nucleotide pairs from the BamH1 end of the insert. The heavy line represents tnaA. The start and stop codons indicate the orientation of tnaA along the restriction map of the pMD6 insert. Labeled ends and the extent of useful sequence information obtained from each fragment are indicated by the dots and arrows above the maps.*

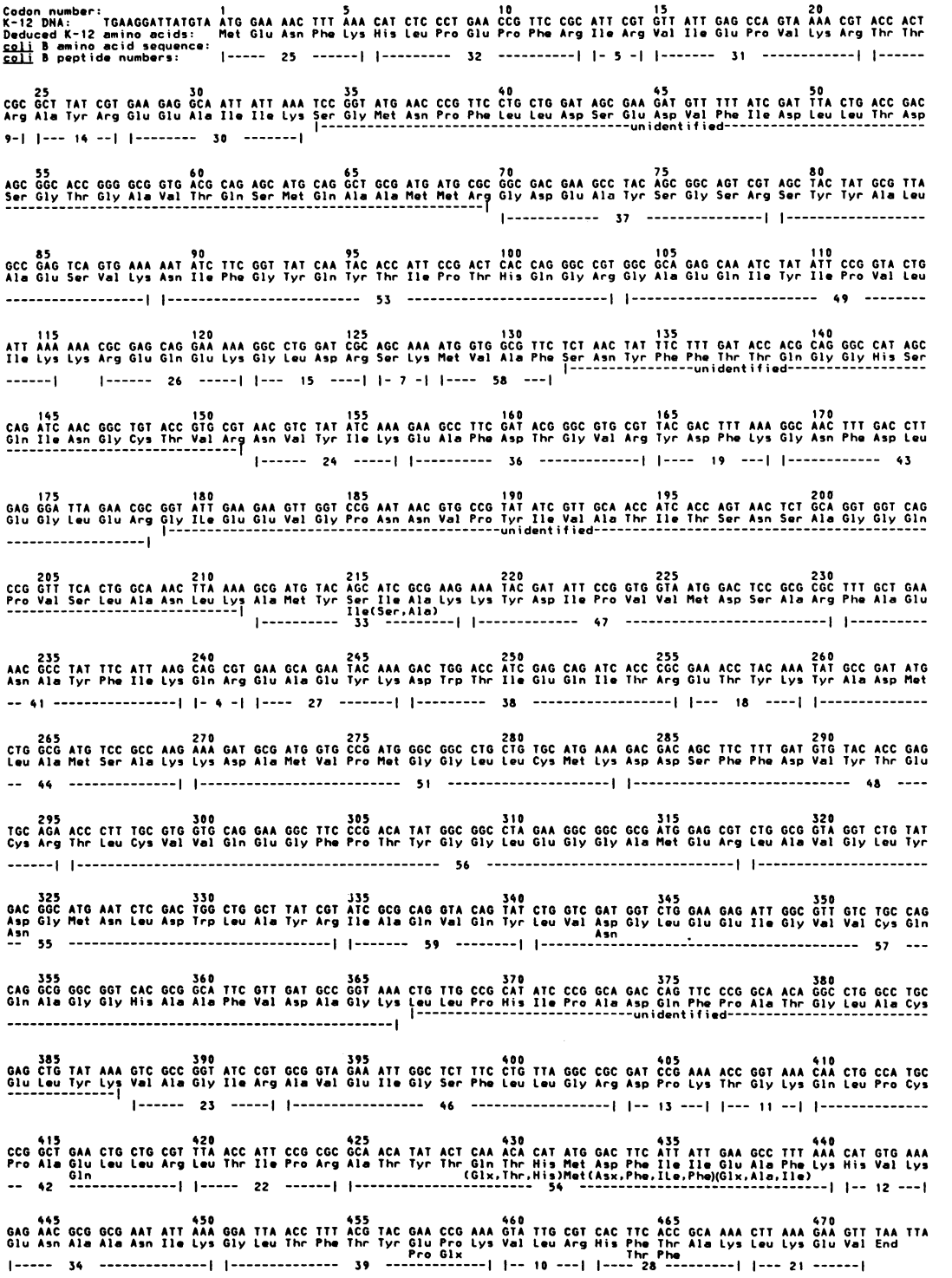

FIG. 4. *Tryptophanase structural gene DNA and amino acid sequences. The nucleotide sequence of tnaA of the K-12 strain of E. coli is given, together with the deduced amino acid sequence. All of the identified tryptic fragments isolated and sequenced from the B strain (15) are presented and are aligned relative to the deduced sequence of the E. coli K-12 protein. The tryptic fragment numbers refer to the catalog numbers assigned to each peptide, as stated in reference 15. Note that the E. coli K-12 equivalent of four large tryptic fragments are unidentified in the E. coli B tryptic fragment catalog.*

TABLE 2. *Amino acid composition of tryptophanase of E. coli K-12 and B*

amino acid	Number of residues[a] K12	B	amino acid	Number of residues[a] K12	B
Asp	25	} 40	Met	16	16
Asn	15		Ile	29	28
Thr	28	27	Leu	36	37
Ser	19	19	Trp	2	2
Glu	34	} 56	Tyr	25	24
Gln	19		Phe	23	25
Pro	20	19	Lys	29	29
Gly	38	37	His	8	8
Ala	43	46	Arg	24	24
Cys	7	6	<Total>	471	474
Val	31	31	Calculated mass (daltons)	52,242	52,595

[a] The number of residues of each amino acid in the K-12 enzyme is deduced from the nucleotide sequence. These data are compared with the experimentally determined values for the B enzyme (16). Standard three-letter abbreviations for the amino acids are used.

coli K-12.

Figure 5 presents the nucleotide sequence of the region immediately upstream from the *tnaA* start codon. The absence of a long, open reading frame in this region suggests that it is not translated and, therefore, that it may contain the promoter and regulatory region controlling tryptophanase production. In this regard, there is a 16-bp sequence exhibiting extensive dyad symmetry and homology with the CAP-binding site in the *lac* operon (27). This site is centered 378 bp upstream from the *tnaA* start codon (Fig. 5). Since 10 of the symmetrical 16 bp are identical to those of the *lac* CAP-binding site, it is likely that this region of the *tnaA* operon is the site of CAP recognition. There are several sequences homologous to known RNA polymerase-binding regions downstream from the potential CAP-binding site. These are not indicated in Fig. 5. Experiments are in progress to determine the exact location of the *tnaA* promoter.

The nucleotide sequence of the 850 bp immediately downstream from *tnaA* was determined. A continuous reading frame was found beginning 90 bp from the end of *tnaA* (Fig. 6). A sequence complementary to the 3′ end of 16S rRNA of *E. coli*, a possible ribosome recognition region (26), was present 7 bp before the first ATG codon of the presumed gene. These findings suggest that there is a structural gene, in the same orientation, just beyond *tnaA*.

Plasmids pMD1, pMD2, pMD2R, pMD3, pMD6, and pMD13 were used as templates in an in vitro coupled transcription-translation system (34). The labeled polypeptide products were analyzed by sodium dodecyl sulfate-polyacrylamide gel (12.5%) electrophoresis. Plasmids pMD1, pMD2, and pMD2R directed the synthesis of two intensely labeled polypeptides, 52,000 and 60,000 to 70,000 in molecular weight (Fig. 7). The 52,000-dalton protein migrated identically to pure *E. coli* B tryptophanase (data not shown). Plasmid pMD13, which contains an insertion of 260 bp of DNA from the *E. coli* biotin operon at the *Pst*I site of *tnaA*, did not confer the Trp^+ phenotype. This plasmid, when used as a template in the S30 system, failed to direct the synthesis of the 52,000-dalton protein (Fig. 7), but did direct the synthesis of a 20,000-dalton polypeptide. If translation of the *tnaA*-reading frame were to halt within the inserted 260-bp *Pst*I fragment, a polypeptide of about 20,000 to 28,000 daltons would be produced. This suggests that the 20,000-dalton protein is the truncated fusion polypeptide translated from the start of the *tnaA*-reading frame of plasmid pMD13. Plasmids pMD3 and pMD6, which lack 3,500 bp of the region upstream from *tnaA*, did not direct the production of the 60,000- to 70,000-dalton polypeptide. This suggests that the structural gene encoding this protein is upstream from *tnaA*.

Although a polypeptide 25,000 daltons or larger should be encoded by the reading frame found downstream from *tnaA*, a gene product of this size was not found when $tnaA^+$ plasmids containing the 3,500 bp downstream from *tnaA* (pMD1 and pMD3) were used as templates in

TABLE 3. *Codon usage in tnaA of E. coli K-12*

amino acid	codon	number of codons	percentage in tnaA	percentage in trpEBA	percentage in rplAKJL	amino acid	codon	number of codons	percentage in tnaA	percentage in trpEBA	percentage in rplAKJL
Phe	UUU	9	39	49	38	Glu	GAA	22	65	76	85
	UUC	14	61	51	62		GAG	12	35	24	15
Leu	UUA	7	19	8	2	Ser	UCU	3	16	18	43
	UUG	2	6	12	2		UCC	3	16	23	32
	CUU	3	8	8	2		UCA	2	11	9	0
	CUC	2	6	11	2		UCG	0	0	15	4
	CUA	1	3	5	0		AGU	2	11	11	4
	CUG	21	58	56	92		AGC	9	47	24	18
Ile	AUU	16	55	56	29	Pro	CCU	1	5	15	0
	AUC	13	45	42	71		CCC	0	0	15	0
	AUA	0	0	2	0		CCA	2	10	22	9
							CCG	17	85	42	91
Met	AUG	16	100	100	100	Thr	ACU	3	11	16	53
Val	GUU	8	26	25	48		ACC	17	61	53	44
	GUC	4	13	19	4		ACA	4	14	12	3
	GUA	7	23	14	38		ACG	4	14	19	0
	GUG	12	38	42	10						
						Ala	GCU	5	12	17	53
Tyr	UAU	15	60	61	13		GCC	10	23	31	6
	UAC	10	40	39	87		GCA	11	25	16	27
							GCG	17	40	35	14
End	UAA	1	100	50	100						
	UAG	0	0	0	0	Cys	UGU	1	14	40	0
	UGA	0	0	50	0		UGC	6	86	60	100
His	CAU	5	63	52	0	Trp	UGG	2	100	100	0
	CAC	3	37	48	100						
						Arg	CGU	13	54	43	57
Gln	CAA	4	21	33	25		CGC	10	42	48	43
	CAG	15	79	67	75		CGA	0	0	4	0
							CGG	0	0	3	0
Asn	AAU	4	27	40	10		AGA	1	4	3	0
	AAC	11	73	60	90		AGG	0	0	0	0
Lys	AAA	26	90	85	87	Gly	GGU	12	32	41	51
	AAG	3	10	15	13		GGC	23	60	42	49
							GGA	2	5	8	0
Asp	GAU	13	52	65	25		GGG	1	3	19	0
	GAC	12	48	35	75						

[a] Codon frequencies are expressed in percentages, i.e., how often a particular codon is used to encode a particular amino acid. The average frequencies for *trpE*$^+$-, *trpB*$^+$-, and *trpA*$^+$-coding regions of *E. coli* K-12 (9, 22) and the average values for four of the *E. coli* K-12 ribosomal protein-coding regions (23) are given for comparison. A total of 471 *tnaA* codons, 658 *rpl* codons, and 1,185 *trp* codons have been included.

```
1         10        20        30        40        50        60        70        80        90        100       110
•         •         •         •         •         •         •         •         •         •         •         •
GTAAACCGCGCATACAGCCGCATTCTGACTGTCAGATGCGCTTCGCTTCATTGTTACCGCTCCTGTTATTCCTCAACCCTTTTTTTAAACATTAAAATTCTTACGTAATT

          120       130       140       150       160       170       180       190       200       210       220
          •         •         •         •    ***** *  **** *    •         •         •         •         •         •
TATAATCTTTAAAAAAAGCATTTAATATTGCTCCCCGAACGATTGTGATTCGATTCACATTTAAACAATTTCAGAATAGACAAAAACTCTGAGTGTAATAATGTAGCCTC

          230       240       250       260       270       280       290       300       310       320       330
          •         •         •         •         •         •         •         •         •         •         •
GTGTCTTGCGAGGATAAGTGCATTATGAATATCTTACATATATGTGTTGACCTCAAAATGGTTCAATATTGACAATAAAATTGTCGATCACCGCCCTTGATTTGCCCTTC

          340       350       360       370       380       390       400       410       420       430       440
          •         •         •         •         •         •         •         •         •         •         •
TGTAGCCATCACCAGAGCCAAACCGATTGATTCAATGTGTTCTATTTGTTTGCTATATCTTAATTTTGCCTTTTGCAAAGGTCATCTCTCGTTTATTTACTTGTTTTAGT

          450       460       470       480       490       500       510       520       530       540       550
          •         •         •         •         •         •         •         •         •         •         •
AAATGATGGTGCTTGCATATATATCTGGCGAATTAATCGGTATAGCAGATGTAATATTCACAGGGATCACTGTAATTAAAATAAATGAAGGATTATGTAATGGAA-----
```

FIG. 5. *Presumed tnaA promoter-regulatory region. The nucleotide sequence of the 540 bp immediately upstream from the tnaA translation start codon is presented. The sequence is numbered arbitrarily. The nucleotides complementary to the 3' end of E. coli K-12 16S rRNA are underlined. The nucleotides exhibiting dyad symmetry in the presumed CAP-binding region are underlined by dashes. Residues homologous to the symmetric sequence of the lac CAP-binding site (27) are indicated by asterisks.*

the coupled system. The absence of the presumed gene product was not due to the lack of methionine residues in the presumed polypeptide, as there are seven ATG codons present in the first 254 codons of the open reading frame.

DISCUSSION

Using the Mu::RP4 transposition method (11) described by Laird et al. (19) to clone fragments of the bacterial chromosome, we selected strains

```
Codon number:    463          466          469          472

                                                                    1
DNA sequence:    CAC TTC ACC GCA AAA CTT AAA GAA GTT TAA TTAATACTACAGAGTGGCTATAAGGATGTTAGCCACTCTCTTACCC
                                                      1          2
Amino acids:     His Phe Thr Ala Lys Leu Lys Glu Val End

                                                 1          4          7          10         13
     2
TACATCCTCAATAACAAAAATAGCCTTCCTCTAAAGGTGGCATC ATG ACT GTT CAA GCT GAA AAA AAG CAC TCT GCA TTT TGG ------
                                             Met Thr Val Gln Ala Glu Lys Lys His Ser Ala Phe Trp ------
```

FIG. 6. *Presumed intercistronic region between tnaA$^+$ and an assumed adjacent structural gene. The nucleotide sequence corresponding to the carboxy terminus of tryptophanase, the presumed 90-bp punctuation region, and the presumed amino terminus of a second protein are shown. The deduced amino acid sequences are presented. Each half of the two regions of dyad symmetry are underlined and numbered. Dashed lines indicate the nucleotides complementary to the 3′ end of E. coli 16S rRNA (26).*

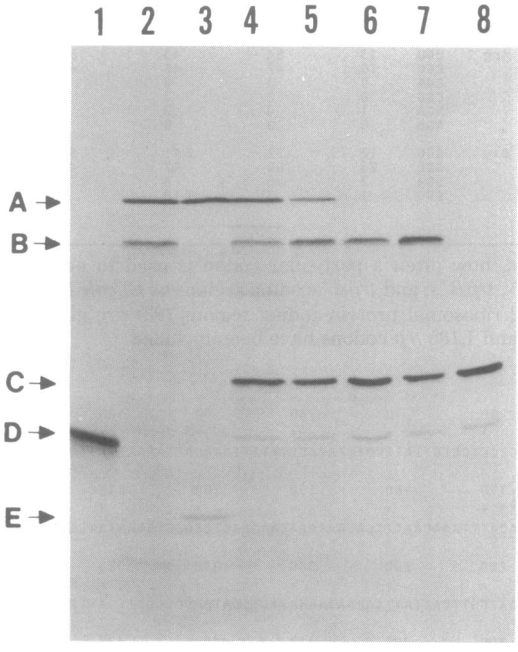

FIG. 7. *Coupled in vitro transcription-translation. Lanes 1 through 8 present the labeled polypeptides produced in an S30 extract when the following plasmids were used as templates: (1) pACYC184; (2) pMD1; (3) pMD13; (4) pMD2; (5) pMD2R; (6) pMD3; (7) pMD6; (8) pBR322. Electrophoresis was from top to bottom. The lettered arrows indicate the polypeptides encoded by: (A) the structural gene located upstream from tnaA; (B) tnaA; (C) the β-lactamase structural gene; (D) the chloramphenicol resistance gene; (E) the tnaA-reading frame of plasmid pMD13, containing an insertion of 260 bp at codon number 218. Protein was labeled with [^{35}S]methionine.*

harboring RP4::Mu *c*(Ts)::*tnaA*$^+$ plasmids. By subcloning DNA fragments from one of these plasmids, we constructed plasmid pMD6, which contain all of *tnaA*. This plasmid also has the intact regulatory region controlling tryptophanase production and part of what may be a second structural gene beginning 90 bp downstream from *tnaA*.

The *tnaA* nucleotide sequence was determined with pMD6 DNA and was used to deduce the tryptophanase amino acid sequence (Fig. 4). On the basis of protein structure data for the *E. coli* B enzyme, Kagamiyama et al. (15, 16) have predicted a polypeptide length of 474 amino acid residues. The polypeptide length deduced from the K-12 DNA sequence is 471 residues, in excellent agreement with their data. The amino acid compositions of the two tryptophanases are quite similar (Table 2), and tryptic fragments identified from the *E. coli* B enzyme closely match the deduced amino acid sequence of the enzyme from K-12 (Fig. 4).

Two patterns of codon usage, evident in *E. coli* structural genes, were reflected in the set of four ribosomal protein genes (*rplA*, *rplK*, *rplJ*, and *rplL*) and the set of three tryptophan biosynthetic protein genes (*trpE*, *trpB*, and *trpA*). The codons preferentially used in the ribosomal protein genes are those that are efficiently recognized by the most abundant tRNA species (12, 23). Preference for these codons has not been observed in the genes for the tryptophan biosynthetic proteins (9, 22). Codon usage in *tnaA* resembled each of these classes, depending on which amino acid was considered (Table 3). Thus, the pattern of codon usage for leucine, valine, histidine, threonine, alanine, tyrosine, and glutamic acid residues paralleled the codon usage in the *trp* genes. However, codon usage for proline residues resembled codon usage in the *rpl* genes. There was no significant difference in the two classes in the choice of codons for phenylalanine, lysine, isoleucine, arginine, and glutamine residues. Codon usage in *tnaA* for glycine,

asparagine, cysteine, and aspartic acid residues appeared intermediate between the patterns evident in the *rpl* and *trp* genes. The interesting feature of codon usage in *tnaA*, thus, is that the pattern was intermediate between those of highly expressed and occasionally expressed genes (12).

The existence of an open reading frame beginning 90 bp past the stop codon of *tnaA* and continuing at least 750 bp suggests that there may be a second gene coregulated with and immediately downstream from *tnaA* (Fig. 6). There is a potential "Shine and Dalgarno" sequence −11 to −7 base pairs ahead of the presumed start codon (26). Although the identity of this second gene is uncertain, three lines of evidence suggest that it may encode the low-affinity tryptophan permease. First, some mutants selected for their inability to use tryptophan as the sole source of carbon were not only tryptophanase deficient, but also were deficient in the low-affinity tryptophan permease (8). Second, the low-affinity permease has been shown to share the same regulatory characteristics as *tnaA*; that is, both activities are tryptophan inducible and sensitive to glucose repression (2–4, 8, 31, 33). Third, phage P1 transduction data indicate that mutations affecting inducible tryptophan permease production are closely linked to *tnaA* (D. Oxender, personal communication). These observations suggest that expression of the low-affinity permease is associated with expression of tryptophanase, a finding consistent with the interpretation that the two structural genes may be in the same operon.

ACKNOWLEDGMENTS

We appreciate the help of Robert P. Gunsalus with the S30 experiments and the advice of Brian P. Nichols on sequencing techniques.

M.C.D. is a predoctoral trainee of the U.S. Public Health Service. C.Y. is a career Investigator of the American Heart Association. These studies were supported by Public Health Service grant GM09738 from the National Institutes of Health and by National Science Foundation grant PCM77-24333.

LITERATURE CITED

1. **Bennett, G. N., and C. Yanofsky.** 1978. Sequence analysis of operator constitutive mutants of the tryptophan operon of *Escherichia coli*. J. Mol. Biol. **121**:179–192.
2. **Bilezikim, J. P., R. O. R. Kaempfer, and B. Magasanik.** 1967. Mechanism of tryptophanase induction in *Escherichia coli*. J. Mol. Biol. **27**:495–506.
3. **Boezi, J. A., and R. D. DeMoss.** 1960. Properties of a tryptophan transport system in *Escherichia coli*. Biochim. Biophys. Acta **49**:471–84.
4. **Botsford, J. L., and R. D. DeMoss.** 1971. Catabolite repression of tryptophanase in *Escherichia coli*. J. Bacteriol. **105**:303–312.
5. **Boyland, E., D. Manson, and R. Nery.** 1962. The reaction of phenylhydroxylamine and 2-naphthylhydroxylamine with thiols. J. Am. Chem. Soc. **56**:658–666.
6. **Bradford, M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**: 248.
7. **Burns, R. O., and R. D. DeMoss.** 1962. Properties of tryptophanase from *Escherichia coli*. Biochim. Biophys. Acta **65**:233–244.
8. **Burrous, S., and R. D. DeMoss.** 1963. Studies on tryptophan permease in *Escherichia coli*. Biochim. Biophys. Acta **73**:623–637.
9. **Crawford, I. P., B. P. Nichols, and C. Yanofsky.** 1980. Nucleotide sequence of the *trpB* gene in *E. coli* and *S. typhimurium*. J. Mol. Biol. **142**:489–502.
10. **DeMoss, R. D., and K. Moser.** 1969. Tryptophanase in diverse bacterial species. J. Bacteriol. **98**:167–171.
11. **Faelen, M., A. Toussaint, M. Van Montagu, S. Van den Elsacker, G. Engler, and J. Schell.** 1977. In vivo genetic enginnering: the Mu-mediated transposition of chromosomal DNA segments onto transmissible plasmids, p. 521–530. *In* A. I. Bukhari, J. A. Shapiro, and S. L. Adhya (ed.), DNA insertion elements, plasmids, and episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressitivity. Nucleic Acids Res. **9**:r43–r74.
13. **Gunsalus, R. P., G. Zurawski, and C. Yanofsky.** 1979. Structural and functional analysis of cloned deoxyribonucleic acid containing the *trpR-thr* region of the *Escherichia coli* chromosome. J. Bacteriol. **140**:106–113.
14. **Humphreys, G. O., G. A. Willshaw, and E. S. Anderson.** 1975. A simple method for the preparation of large quantities of pure plasmid DNA. Biochim. Biophys. Acta **383**:457–463.
15. **Kagamiyama, H., H. Matsubara, and E. E. Snell.** 1972. The chemical structure of tryptophanase from *Escherichia coli*. III. Isolation and amino acid sequence of the tryptic peptides. J. Biol. Chem. **247**:1576–1585.
16. **Kagamiyama, H., H. Wada, H. Matsubara, and E. E. Snell.** 1972. The chemical structure of tryptophanase from *Escherichia coli*. II. The structure of tryptophanase monomer. J. Biol. Chem. **247**:1571–1575.
17. **Katz, L., D. T. Kingsbury, and D. R. Helinski.** 1973. Stimulation by cyclic adenosine monophosphate of plasmid deoxyribonucleic acid replication and catabolite repression of the plasmid deoxyribonucleic acid-protein relaxation complex. J. Bacteriol. **114**:577–591.
18. **Korn, L. J., C. L. Queen, and M. N. Wegman.** 1977. Computer analysis of nucleic acid regulatory sequences. Proc. Natl. Acad. Sci. U.S.A. **74**:4401–4405.
19. **Laird, A. J., D. W. Ribbons, G. C. Woodrow, and I. G. Young.** 1980. Bacteriophage Mu-mediated gene transposition and *in vitro* cloning of the enterochelin gene cluster of *Escherichia coli*. Gene **11**:347–357.
20. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:540–564.
21. **Newton, W. A., and E. E. Snell.** 1964. Catalytic properties of tryptophanase, a multifunctional pyridoxal phosphate enzyme. Proc. Natl. Acad. Sci. U.S.A. **51**: 382–389.
22. **Nichols, B. P., M. van Cleemput, and C. Yanofsky.** 1980. The nucleotide sequence of *Escherichia coli trpE*: anthranilate synthetase component I contains no tryptophan residues. J. Mol. Biol. **146**:45–54.
23. **Post, L. E., G. D. Strycharz, M. Nomura, H. Lewis, and P. P. Dennis.** 1979. Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit β in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **76**:1697–1701.
24. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**: 107–110.
25. **Selker, E., K. Brown, and C. Yanofsky.** 1977. Mitomycin C-induced expression of *trpA* of *Salmonella ty-*

phimurium inserted into plasmid ColE1. J. Bacteriol. **129**:388–394.

26. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosomal binding sites. Proc. Natl. Acad. Sci. U.S.A. **71**:1342–1346.
27. **Simpson, R. B.** 1980. Interaction of the cAMP receptor protein with the *lac* promotor. Nucleic Acids Res. **8**: 759–766.
28. **Snell, E. E.** 1975. Tryptophanase: structure, catalytic activities, and mechanisms of action. Adv. Enzymol. **42**: 287–333.
29. **Suelter, C. H., J. Wang, and E. E. Snell.** 1976. Direct spectrophotometric assay of tryptophanase. FEBS Lett. **66**:230–232.
30. **Vogel, H. J., and D. M. Bonner.** 1956. Acetylornithinase of *Escherichia coli*: partial purification and some properties. J. Biol. Chem. **218**:97–106.
31. **Ward, D. F., and M. D. Yudkin.** 1976. Mutations in *Escherichia coli* that relieve catabolite repression of tryptophanase synthesis. Tryptophanase promotor-like mutations. J. Gen. Microbiol. **92**:133–137.
32. **Watanabe, T., and E. E. Snell.** 1972. Reversibility of the tryptophanase reaction: synthesis of tryptophan from indole, pyruvate, and ammonia. Proc. Natl. Acad. Sci. U.S.A. **69**:1086–1090.
33. **Yudkin, M. D.** 1976. Mutations in *Escherichia coli* that relieve catabolite repression of tryptophanase synthesis. Mutations distant from the tryptophanase gene. J. Gen. Microbiol. **92**:125–132.
34. **Zubay, G., D. E. Morse, W. J. Schrenk, and J. H. Miller.** 1972. Detection and isolation of the repressor protein for the tryptophan operon of *Eshcerichia coli*. Proc. Natl. Acad. Sci. U.S.A. **69**:1100–1103.