Proc. Natl. Acad. Sci. USA
Vol. 76, No. 4, pp. 1706–1710, April 1979
Biochemistry

Regulation of the threonine operon: Tandem threonine and isoleucine codons in the control region and translational control of transcription termination

(DNA sequencing/attenuation/RNA secondary structure)

JEFFREY F. GARDNER

Department of Microbiology, University of Illinois, Urbana, Illinois 61801

Communicated by Charles Yanofsky, January 31, 1979

ABSTRACT **The DNA sequence of 178 base pairs preceding the first structural gene of the threonine operon of *Escherichia coli* has been determined. A region of perfect 2-fold rotational symmetry, involving 28 base pairs, precedes the first structural gene. The structural similarity of this sequence to known RNA polymerase termination sites suggests that this region is the termination site of the threonine operon leader RNA. Moreover a mutation (*thr* 79-20), which confers a derepressed, constitutive phenotype, was sequenced and found to be a G·C insertion in the putative terminator. A potential coding region for a 21-amino acid leader peptide ends approximately 18 base pairs before the terminator. This peptide contains eight threonine and four isoleucine codons. Eleven of these codons are in tandem. A model for threonine operon regulation, involving alternative secondary RNA structures and translation of leader RNA, is discussed.**

The threonine (*thr*) operon of *Escherichia coli* is composed of four structural genes, arranged in the order *thr*A_1, A_2, B, C, (1), and a regulatory region which maps before *thr*A_1 (2, 3). The operon is an example of a system that is regulated by multivalent repression. When the intracellular levels of threonine and isoleucine are in excess, the synthesis of the enzymes controlled by the *thr* operon is repressed (4). To date, no regulatory mutants corresponding to apo-repressor mutations have been identified. However, mutations that result in derepressed, constitutive synthesis of the *thr* operon enzymes have been isolated and mapped in the *thr* regulatory region (2, 3).

Recently, by use of a secondary site λ lysogen located in the *thr* regulatory region (5), it has been possible to construct λ*pthrspi* transducing phages that carry the *thr* operon from both the wild type or a derepressed, constitutive mutant (*thr*79-20) (6). Restriction endonuclease analyses of these phages have made it possible to identify and to isolate DNA from the region preceding *thr*A_1 (6). In this communication, I report the DNA sequence of this region and present evidence that translational control of transcription termination (attenuation) is involved in *thr* regulation. The mechanism of control appears similar to that proposed for other operons involved in amino acid biosynthesis (7, 8).

MATERIALS AND METHODS

DNA Sources and Sequencing. A 1700-base pair (bp) *Hae* III restriction fragment, carrying DNA from the region immediately preceding and including approximately 100 bp of *thr*A_1, has been identified and mapped (6). Fig. 1 shows a restriction map of this region. pVH51 (mini-ColE1) plasmids carrying insertions of this fragment in the *Hin*dII site (9) were used as sources of DNA. Details of plasmid constructions will be published elsewhere. Procedures using recombinant DNA were performed in accordance with the National Institutes of Health guidelines. DNA sequence determination was carried out by the procedure of Maxam and Gilbert (10).

Enzymes and Electrophoresis. *Taq* I was prepared by the method of Sato *et al.* (11). *Hae* III, *Hha* I, and *Tha* I were purchased from Bethesda Research Laboratories (Rockville, MD). T4 polynucleotide kinase was purchased from New England BioLabs. Restriction fragments used for DNA sequencing were isolated from polyacrylamide gels (6, 10).

RESULTS

Wild-Type Sequence. Fig. 1 shows the scheme utilized for DNA sequencing of the region preceding *thr*A_1 and Fig. 2 shows some representative sequencing gels. These restriction sites have been mapped previously (6). Except for the sequence from the *Taq* I site into *thr*A_1 and for the *thr*79-20 mutation, both strands of the DNA were sequenced and continuity of the sequence was ensured by sequencing across restriction sites. All sequences were confirmed by at least two independent experiments.

One problem encountered in interpreting the sequencing data was determining the exact number of A·T base pairs from positions −27 to −35 and −58 to −63 (see Figs. 2 and 3). For example, the sequence reading from the *Taq* I site includes nine As from positions −27 to −35 and five Ts from −58 to −63 (Fig. 2*A*) whereas reading the opposite strand sequence from the *Tha* I site showed eight Ts from −27 to −35 and six As from −58 to −63 (data not shown). Because these sequences lie in regions of extensive symmetry, it is probable that secondary structure effects on the mobilities of DNA fragments created this problem. Consequently, the sequences from these regions were assigned from regions of the gels where secondary structures should not affect the mobilities of the fragments. Thus, the run of nine As at −27 to −35 was assigned from the sequence originating at the *Taq* I site and the sequence of six Ts at −58 to −63 was assigned from the sequence originating at the *Tha* I site.

The position of *thr*A_1 was determined by matching the DNA sequence with the codon assignments of the amino-terminal amino acids. The sequence of the seven amino-terminal residues of *thr*A_1 has been reported to be Met-Arg-Val-Leu-Lys-Gly-Gly (12).* With the exception of the Phe (between Lys and Gly),

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: bp, base pairs

* See *Note Added in Proof*.

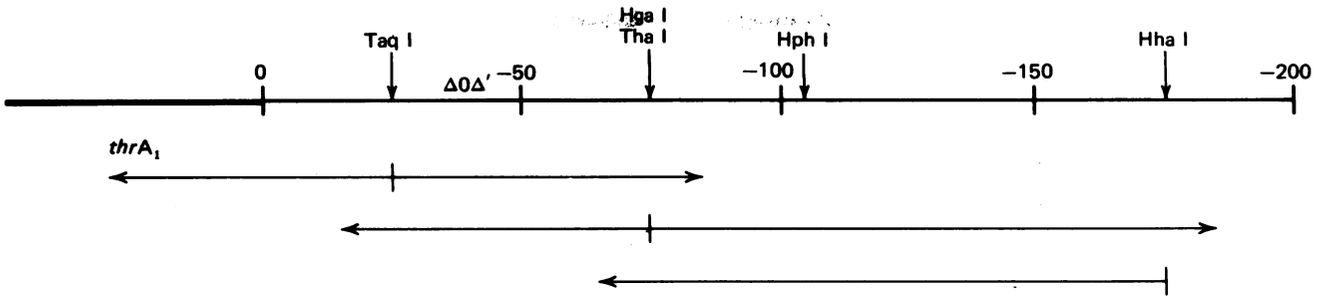

FIG. 1. Restriction map of *thr* regulatory region map based on analyses of 1700-bp *Hae* III restriction fragment that carries this region (6). Coordinates are given in bp from the initiation codon of $thrA_1$. *att*ΔΟΔ', a secondary λ attachment site in the *thr* regulatory region, has been mapped (6) within the *Taq* I–*Hha* I fragment. Horizontal arrows show the regions of DNA sequenced in this report. Restriction sites used for 5'-^{32}P-end labeling (10) were the *Taq* I, *Tha* I, and *Hha* I sites.

my sequence matches the codon assignments of the amino acids with the nucleotide sequence (Fig. 3). The sequence 5'-G-A-G-G-T-3', which is complementary to the 16S ribosomal RNA 3' end (13, 14), occurs at positions −9 to −12 in the sequence and is probably the ribosome binding sequence for $thrA_1$.

***thr*79-20 Sequence.** *In vivo*, cells bearing the *thr* 79-20 mutation exhibit derepressed, constitutive synthesis of the *thr* operon enzymes (6). The sequence change in *thr* 79-20 was determined by sequencing from the *Taq* I site (see Fig. 1) and was identified as a G·C insertion at position −37 (See Figs. 2*C* and 3). This is consistent with previous work (2, 6) which has shown that this mutation maps in the region preceding $thrA_1$.

DISCUSSION

Features of the Sequence and Implications for *thr* Regulation. Fig. 3 shows the DNA sequence of the 178 bp preceding $thrA_1$ and the first 30 bp of the $thrA_1$ gene. Several features of the sequence are similar to those found for analogous regulatory regions of the *trp* (7), *phe* (8), and *his* (15, 16) operons. First,

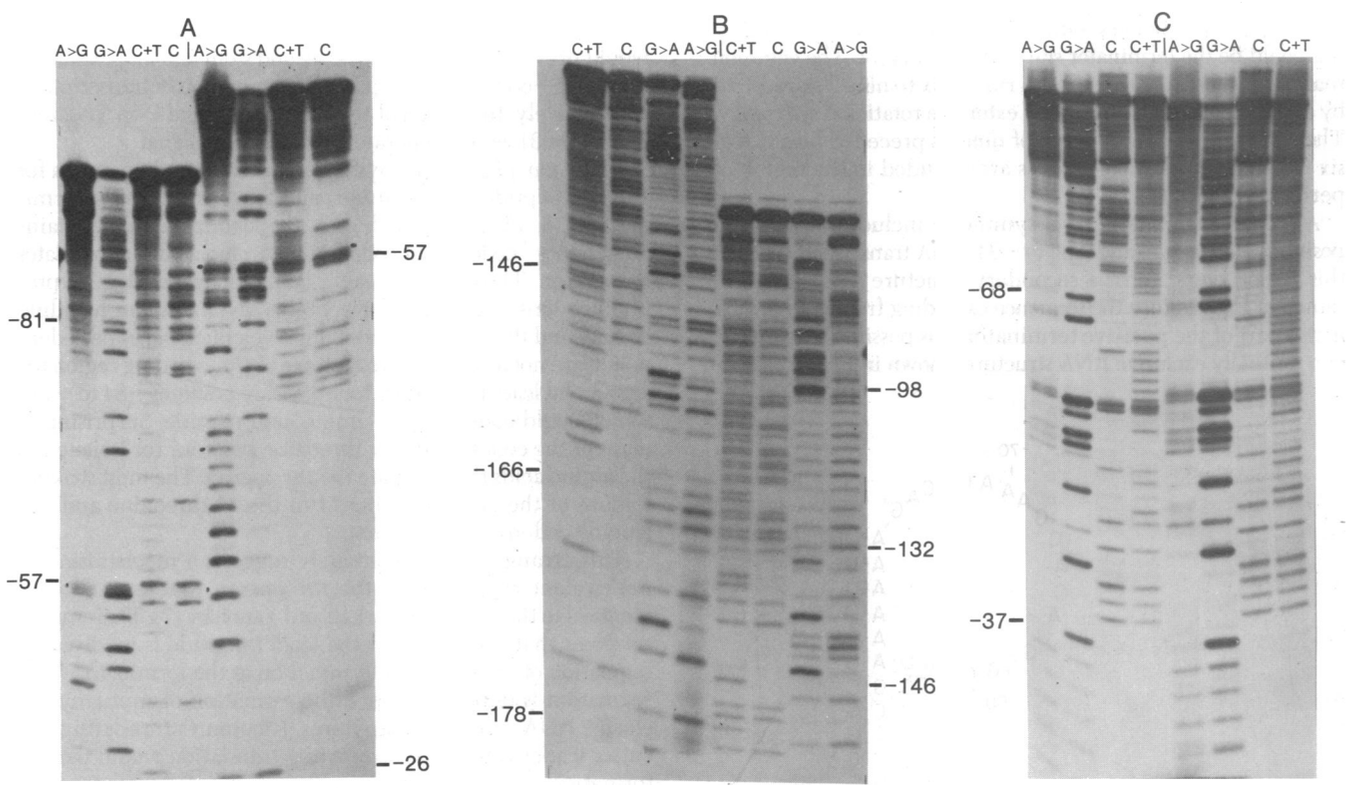

FIG. 2. Representative sequence data. The numbers refer to nucleotides in the sequence (see Fig. 3). (*A*) Sequence from −26 to −75, originating from the *Taq* I site. Gel shows results of two loadings (10) done 12 hr apart. The sequence from −75 to −90 was obtained by extending the electrophoresis time (data not shown). (*B*) Sequence from −178 to −98 originating from the *Hha* I site. The run of tandem threonine and isoleucine codons from −100 to −132 is clearly seen. By extending the electrophoresis time, it was possible to sequence across the *Tha* I site. Sequencing of the opposite strand from the *Tha* I site toward the *Hha* I site confirmed the sequence (data not shown). (*C*) Comparison of the sequences of the wild type (*Left*) and *thr*79-20 mutation (*Right*) originating from the *Taq* I site. Both sets of sequencing reactions were run on the same gel. The *thr*79-20 mutation shows an insertion of a C at position −37 to −39. The run of six Ts in the *thr*79-20 sequence (positions −58 to −63), as opposed to the run of five Ts at the same position in the wild-type sequence, is probably due to destabilization of secondary structure caused by the G-C insertion rather than by the occurrence of a double mutation. Consistent with this interpretation is the observation that sequencing of the opposite strand from the *Tha* I site showed a run of six As in the wild-type sequence.

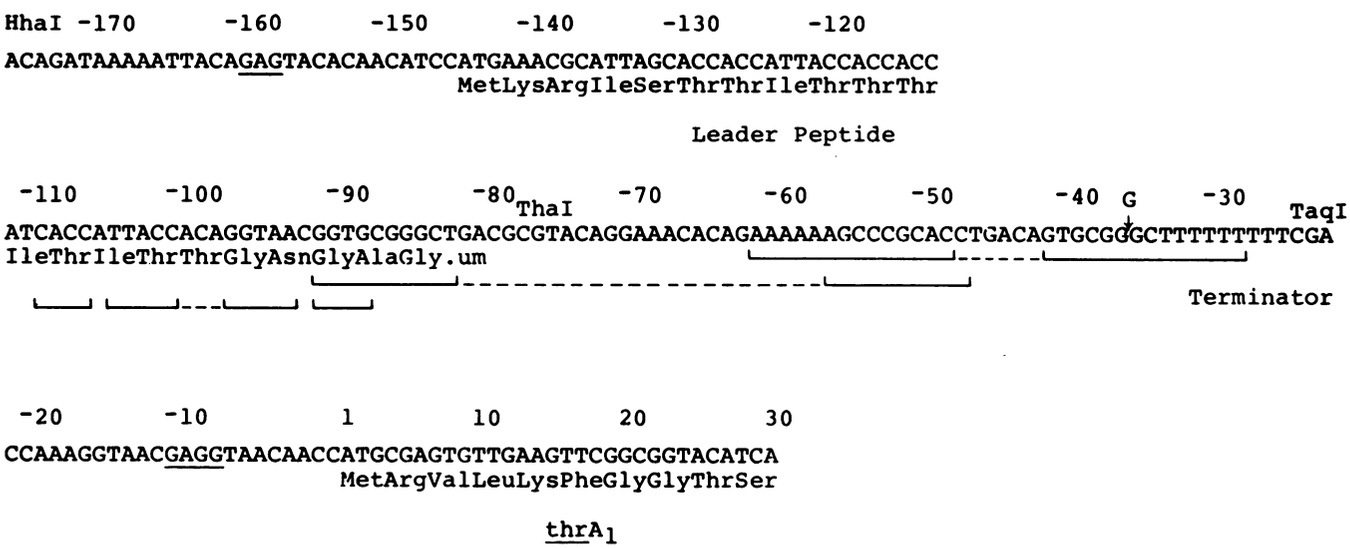

FIG. 3. DNA sequence of the proposed *thr* operon control region. Only one DNA strand is shown. Coordinates are given in bp and are assigned relative to the first nucleotide of the *thr*A_1 protein. The *Hha* I site lies three to five bp to the left of position −178. Brackets connected by dashed lines show regions of rotational symmetry. A possible leader peptide, from positions −147 to −85, containing eight threonine and four isoleucine codons is shown. The sequence GAG at −160 to −162 is complementary to the 3′ end of 16S ribosomal RNA (13, 14) and may be the ribosome binding sequence for the leader peptide. It is also possible that the coding region extends beyond position −178. The sequence G-A-G-G-T from −12 to −9 is probably the ribosome binding sequence for *thr*A_1 (13, 14). The proposed terminator lies at positions −27 to −63. The *thr*79-20 mutation is indicated as an insertion of a G at position −37 to −39.

there is a region of perfect 2-fold rotational symmetry, involving 28 bp, from positions −30 to −43 and −50 to −63 (see Fig. 3). By analogy to sequences found in the above regulatory regions, this could be the terminator site of the attenuator. The common feature of these terminators is a run of six to nine Ts preceded by a high G-C region containing extensive rotational symmetry. The *thr* operon contains a run of nine Ts preceded by a run of six G-C base pairs. Six of the Ts are included in the region of perfect symmetry.

A second region of perfect symmetry includes 20 bp from positions −48 to −58 and −84 to −94. RNA transcribed through this region could form a secondary structure similar to that shown in Fig. 4. Since the sequence extending from −48 to −58 is also part of the putative terminator, it is possible to form the two mutually exclusive RNA structures shown in Fig. 4. Similar

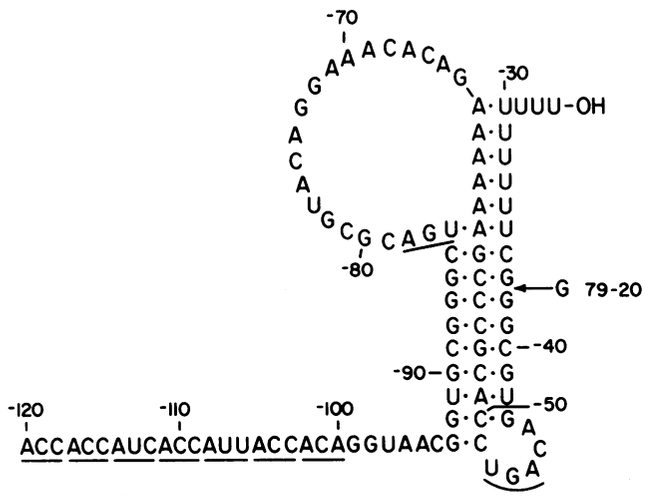

FIG. 4. Proposed alternative secondary structures of RNA transcribed from the *thr* leader region. Tandem threonine and isoleucine codons as well as in-phase termination codons are underlined. The insertion of a G in the *thr*79-20 mutation is indicated (see text for further details).

structures can be found in the *trp* (7), *phe* (8), and *his* (15, 16) operons. From the model proposed by Lee and Yanofsky (7), formation of the first stem and loop (positions −48 to −93) would prevent formation of the second stem and loop structure (positions −63 to −30) and prevent termination of transcription. Alternatively, formation of the second stem and loop (positions −30 to −63) would generate a termination signal.

In the *trp*, *phe*, and *his* systems, a potential coding region for a leader peptide, 14–16 amino acids long, precedes the terminator (7, 8, 15, 16). In each case, the leader peptide contains "regulatory" codons that code for the amino acid that regulates the operon. Thus, the *trp* leader contains two tandem tryptophan codons, the *his* leader contains seven tandem histidine codons, and the *phe* leader contains seven phenylalanine codons which are not all tandem. In the case of *thr*, a coding region for a potential leader peptide is found within positions −84 to −147 which could code for a 21-amino acid peptide. Surprisingly, eight of the codons code for threonine and four for isoleucine, the amino acids that regulate the *thr* operon. The most striking feature of the peptide is that 11 of the 12 threonine and isoleucine codons are in tandem.

This arrangement immediately suggests a mechanism for multivalent regulation of the *thr* operon that is essentially identical to that proposed by Lee and Yanofsky (7). Their model proposes that translation of the leader peptide is involved in regulation of transcription termination at the terminator. This regulation is dependent upon the formation of mutually exclusive RNA secondary structures. Ribosomes translating the entire leader RNA and terminating translation would permit formation of the terminator secondary structure, whereas ribosomes stalling at the regulatory codons would favor formation of the alternate structure and thereby preclude formation of the terminator secondary structure. Because translation of the putative *thr* operon leader peptide would be extremely sensitive to intracellular levels of threonine *and* isoleucine, it is possible that translation of the leader determines whether RNA polymerase terminates transcription at the terminator or proceeds into the structural genes. Under conditions in which threonine or isoleucine, or both, is limiting, ribosomes would stall at the

regulatory codons. This would allow formation of the RNA secondary structure involving positions −49 to −59 and −83 to −93. Formation of this structure would exclude formation of the second stem and loop, which may be important in signaling transcription termination, and RNA polymerase would not terminate. Under conditions such that threonine and isoleucine are in excess, translation of the leader would proceed so that the ribosomes would physically prevent formation of the first stem and loop, because RNA involved in secondary structure is translated. In this case, the second stem and loop structure could form and RNA polymerase would terminate transcription. This mechanism differs slightly from that proposed for *trp* and *phe* in that in these systems ribosomes do not actually translate the region of secondary structure, although they can interfere with secondary structure formation. Thus, the sequence of tandem threonine and isoleucine codons in the putative *thr* leader region provides a mechanism whereby transcription termination is regulated by the intracellular levels of threonine and isoleucine. To observe multivalent repression (which may actually be transcription termination), both amino acids (or the charged aminoacyl tRNAs) must be present in excess whereas a deficiency in either amino acid (or charged aminoacyl tRNA) would result in ribosome stalling and prevent transcription termination.

In the *trp* and, especially, *phe* leader regions (7, 9) it is possible to draw additional stable secondary RNA structures that involve base pairing of the coding sequences of the leader with a more distal sequence preceding the terminator. Because this base pairing involves RNA containing the codons of the regulatory amino acids (tryptophan or phenylalanine) it has been proposed that this secondary structure forms when ribosomes stall and that these base pairings may allow the "starvation response" (transcription through the terminator) to be specific for some of the codons on the leader RNA (17). An analogous stable structure is not present in the *thr* leader region.

The position of *thr*P has not yet been determined. If the *thr* operon is similar to the *trp* and *phe* operons (7, 8), *thr*P should lie at least 140 bp to the left of the attenuator and would not be present in the sequence presented here. One would also predict that it should be possible to detect the leader RNA *in vitro* and *in vivo*. Preliminary *in vitro* transcription studies, using restriction fragment templates, have detected a 6S RNA that hybridizes to the correct separated strand λ*thr* DNA (unpublished data). If this RNA is the *in vitro* terminated transcript, the position of *thr*P would lie immediately to the left of the *Hha* I site (see Fig. 3) in the sequence.

***thr*79-20 Mutation and Transcription Termination.** Although transcription termination has been observed at both the *trp* (7) and *phe* (8) terminators, the exact nature of the termination signal is unknown. However, Stauffer *et al.* (18) have recently isolated and sequenced "termination relief" mutations in the *trp* attenuator. These mutations show an increased frequency of both *in vivo* and *in vitro* readthrough transcription beyond the terminator. The common feature of these mutations is that they decrease the stability of the potential RNA secondary structure which may be important in signalling termination. Base pairing consistent with the existence of this secondary structure has been observed *in vitro* (7) and it has been proposed that the integrity of this structure is necessary for efficient termination.

The *thr*79-20 mutation, which maps in the analogous region of the proposed *thr* terminator, may act by a similar mechanism because the insertion would decrease the stability of the RNA secondary structure that may be involved in transcription termination. Thus, the *thr*79-20 mutation may exhibit the derepressed, constitutive phenotype because RNA polymerase no longer terminates efficiently at the *thr* terminator regardless of the state of translation of the leader region.

Other Considerations. It is known that cells that carry some mutations in Thr-tRNA synthetase (*thr*S) and would be expected to contain low intracellular levels of charged Thr-tRNA, are derepressed for the *thr* operon enzymes (19). This is consistent with the model presented above because a deficiency of charged Thr-tRNA would promote ribosome stalling at the regulatory codons. Also consistent with the model is the observation that the antibiotic borrelidin (20), which inhibits charging of Thr-tRNA (21), derepresses the *thr* operon in wild-type cells (20, 21). It is interesting to note that the threonine, and possibly the isoleucine, codon sequences do not appear to be random. Seven of the eight threonine codons are ACC and three of the four isoleucine codons are ATT. Thus, regulation of *thr* may be (largely) dependent upon the extent of charging of just two tRNA species.

At the present time, it is not possible to determine the role other components might play in *thr* regulation. Although the *thr*S and *ile*S gene products may play an indirect role, via charging of tRNAs, it is also possible that they participate directly in some aspect of regulation. Similarly, it is not known whether *thr* regulation is mediated via repressor–operator control in addition to the probable attenuation control described in this report. Saint-Girons (22) has recently isolated mutations (*thr*X) that are unlinked to the *thr* operon, *thr*S, or *ile*S. The *thr*X locus could code for a regulatory protein (apo-repressor or antiterminator?) or possibly a tRNA or tRNA modification enzyme. Further characterization of these mutations should prove useful in elucidating the mechanism of *thr* regulation.

Note Added in Proof. It has just come to my attention that Cohen and Véron published (23) a more recent sequence (Met-Arg-Val-Leu-Lys-Phe-Gly-Gly) for the amino terminus of *thr*A_1. This sequence agrees completely with the sequence shown in Fig. 3.

I am deeply indebted to William Reznikoff for helpful discussions and suggestions on the phage and plasmid constructions. I also thank Kevin Bertrand and Wayne Barnes for helpful discussions, Samuel Kaplan and Abigail Salyers for criticizing the manuscript, Charles Yanofsky for furnishing unpublished information, and Steven Lynn for expert technical assistance. This work was supported by a National Science Foundation Grant PCM 78-07931.

1. Thèze, J. & Saint-Girons I. (1974) *J. Bacteriol.* **118,** 990–998.
2. Gardner, J. F. & Smith, O. H. (1975) *J. Bacteriol.* **124,** 161–166.
3. Saint-Girons, I. & Margarita, D. (1975) *J. Bacteriol.* **124,** 1137–1141.
4. Freundlich, M. (1963) *Biochem. Biophys. Res. Commun.* **10,** 277–282.
5. Gardner, J. F., Smith, O. H., Fredricks, W. & McKinney, M. A. (1974) *J. Mol. Biol.* **90,** 613–631.
6. Gardner, J. F. & Reznikoff, W. S. (1978) *J. Mol. Biol.* **126,** 241–258.
7. Lee, F. & Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4365–4369.
8. Zurawski, G., Brown, K., Killingly, D. & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4271–4275.
9. Hershfield, V., Boyer, H., Chow, L. & Helinski, D. R. (1976) *J. Bacteriol.* **126,** 447–453.
10. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
11. Sato, S., Hutchison, C. A. & Harris, J. I. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 542–546.
12. Falcoz-Kelley, F., Janin, J., Saari, J. C., Véron, M., Truffa-Bachi, P. & Cohen, G. N. (1972) *Eur. J. Biochem.* **28,** 507–519.

13. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
14. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4734–4738.
15. DiNocera, P. P., Blasi, F., DiLauro, R., Frunzio, R. & Bruni, C. B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4276–4280.
16. Barnes, W. M. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4281–4285.
17. Zurawski, G., Elseviers, D., Stauffer, G. V. & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5988–5992.
18. Stauffer, G. V., Zurawski, G. & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4833–4837.
19. Johnson, E. J., Cohen, G. N. & Saint-Girons, I. (1977) *J. Bacteriol.* **129,** 66–70.
20. Hütter, R., Poralla, K., Zachau, H. G. & Zähner, H. (1966) *Biochem. Z.* **344,** 190–196.
21. Nass, G., Poralla, K. & Zähner, H. (1969) *Biochem. Biophys. Res. Commun.* **34,** 84–91.
22. Saint-Girons, I. (1978) *Mol. Gen. Genet.* **162,** 95–100.
23. Cohen, G. N. & Véron, M. (1974) in *Metabolic Interconversions of Enzymes,* 3rd International Symposium, Seattle, WA, June 1973, ed. Fischer, E. H. (Springer, New York), pp. 335–347.