Proc. Natl., Acad. Sci. USA
Vol. 75, No. 10, pp. 4271–4275, September 1978
Biochemistry

Nucleotide sequence of the leader region of the phenylalanine operon of *Escherichia coli*

(attenuation/transcription termination/RNA secondary structure)

GERARD ZURAWSKI*, KEITH BROWN†, DAVID KILLINGLY†, AND CHARLES YANOFSKY*

* Department of Biological Sciences, Stanford University, Stanford, California 94305; and † School of Biological Sciences, University of Sydney, New South Wales 2006, Australia

Contributed by Charles Yanofsky, July 3, 1978

ABSTRACT **The *pheA* structural gene of the phenylalanine operon of *Escherichia coli* is preceded by a transcribed leader region of about 170 nucleotide pairs. *In vitro* transcription of plasmids and restriction fragments containing the *phe* promoter and leader region yields a major RNA transcript about 140 nucleotides in length. This transcript, *pheA* leader RNA, has the following features: (*i*) a potential ribosome binding site and AUG translation start codon about 20 nucleotides from its 5′ end; (*ii*) 14 additional in phase amino acid codons and a UGA stop codon after the AUG; 7 of these 14 are Phe codons; (*iii*) a 3′-OH terminus about 140 nucleotides from the 5′ end (transcription termination occurs in an A·T-rich region which is subsequent to a G·C-rich region; just beyond the site of transcription termination there is a sequence corresponding to a ribosome binding site and the AUG translation start codon of the *pheA* structural gene); (*iv*) a sequence which would permit extensive intrastrand stable hydrogen bonding. In addition to G·C-rich stem structures, highly analogous to those proposed for the leader RNAs of the tryptophan operons of *E. coli* and *Salmonella typhimurium* [Lee, F. & Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4365–4369], there is also extensive base-pairing possible between the *phe* codon region and a more distal region of the leader transcript. The roles of synthesis of the Phe-rich leader peptide and secondary structure of the leader transcript in the regulation of transcription termination at the attenuator of the *phe* operon are discussed.**

The isolation of λ transducing phages carrying the phenylalanine (*phe*) operon of *Escherichia coli* (1) has facilitated the cloning of this operon onto plasmid vectors. We have mapped the *phe* operon by cleavage with restriction endonucleases and by genetic analysis. In this paper we report that there is a transcribed leader region of about 170 base pairs (bp) immediately preceding *pheA*, the sole structural gene of the *phe* operon. The function of this region in the regulation of *phe* operon expression is discussed.

MATERIALS AND METHODS

Construction and Preparation of Plasmids. The methods used for ligation, transformation, and preparation of plasmid DNAs were essentially those described in Selker *et al.* (2).

Restriction Mapping. The methods used for restriction endonuclease digestion and analysis of DNA fragments by agarose and polyacrylamide gel electrophoresis have been described (3). A detailed restriction analysis of the *phe* operon will appear elsewhere.

***In Vitro* Transcription.** Plasmids and restriction fragments were transcribed and the ^{32}P-labeled RNA products were analyzed by electrophoresis on 10% polyacrylamide/Tris borate/EDTA gels containing 7M urea as described (4).

These studies were performed in accordance with the National Institutes of Health guidelines on recombinant DNA research.

RESULTS

Isolation and Mapping of Plasmids Carrying the Phenylalanine Operon. The fragments generated by *Eco*RI restriction of the DNA of λ*tyr*2, a transducing phage carrying the entire *phe* and *tyr* operons of *E. coli* (1), were ligated into the *Eco*RI site of the plasmid vector pMB9 (5). The mixture was used to transform a *pheA*$^-$ strain to prototrophy. All of several independent *pheA*$^+$ transformants examined carried plasmids containing a 6000-bp insert into the *Eco*RI site of pMB9. Characterization of one such plasmid, pKB45, showed that the 6000-bp insert also carried the *tyr* operon (structural genes: *aroF*, *tyrA*) (6). The physical location of *pheA* on the 6000-bp fragment (Fig. 1) was established by recloning segments of this fragment onto the plasmid vector pBR322 (7). Plasmid pKB220, which carries the 1730-bp *Pst* I-*Bam*HI fragment, has a functional *pheA* gene. The *Bam*HI site in pKB220, and the *Hae* III site 400 bp to its left, are absent in a *pheA*$^+$ plasmid, pKB472, which carries a spontaneous deletion entering from the *Bam*HI site (Fig. 1). These results show that the 1300-bp *Pst* I-*Hae* III region has a functional *pheA* gene. *pheA* should be about 1000 bp long on the basis of the estimated molecular weight of 40,000 for the *pheA* gene product, chorismate mutase–P-prephenate dehydratase (8). The *phe* operon is transcribed toward the tyrosine operon (Fig. 1) (ref. 1; G. Zurawski and K. D. Brown, unpublished results). Thus, the beginning of the *pheA* gene must lie within approximately 300 bp to the right of the *Pst* I site.

***In Vitro* Transcription of the *phe* Operon.** To locate the promoter of the *phe* operon, we isolated and transcribed DNA restriction fragments from the region to the right of the *Pst* I site (Fig. 1). As shown in Fig. 2, the *Hpa* II_{630} DNA fragment yielded an approximately 140-base (b) transcript as the major product of *in vitro* transcription. An RNA transcript about 140 b in length is also the major product when either the *Taq* I_{356} or *Eco*RI-*Hpa* I_{1000} fragments (Fig. 1), or the plasmid pKB45, is used as the DNA template (G. Zurawski, unpublished results). The synthesis of the same approximately 140-b RNA transcript from DNA templates of various sizes indicates that there is a strong transcription termination site approximately 140 bp downstream from the promoter for the *phe* operon. An RNA transcript of about 320 b is the largest product of transcription of the *Hpa* II_{630} DNA fragment (Fig. 2). If the 320-b RNA is due to transcription that terminated at the end of the DNA template rather than at the internal transcription termination site, then transcription is initiated about 150 bp to the right of

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 pair; b, base.

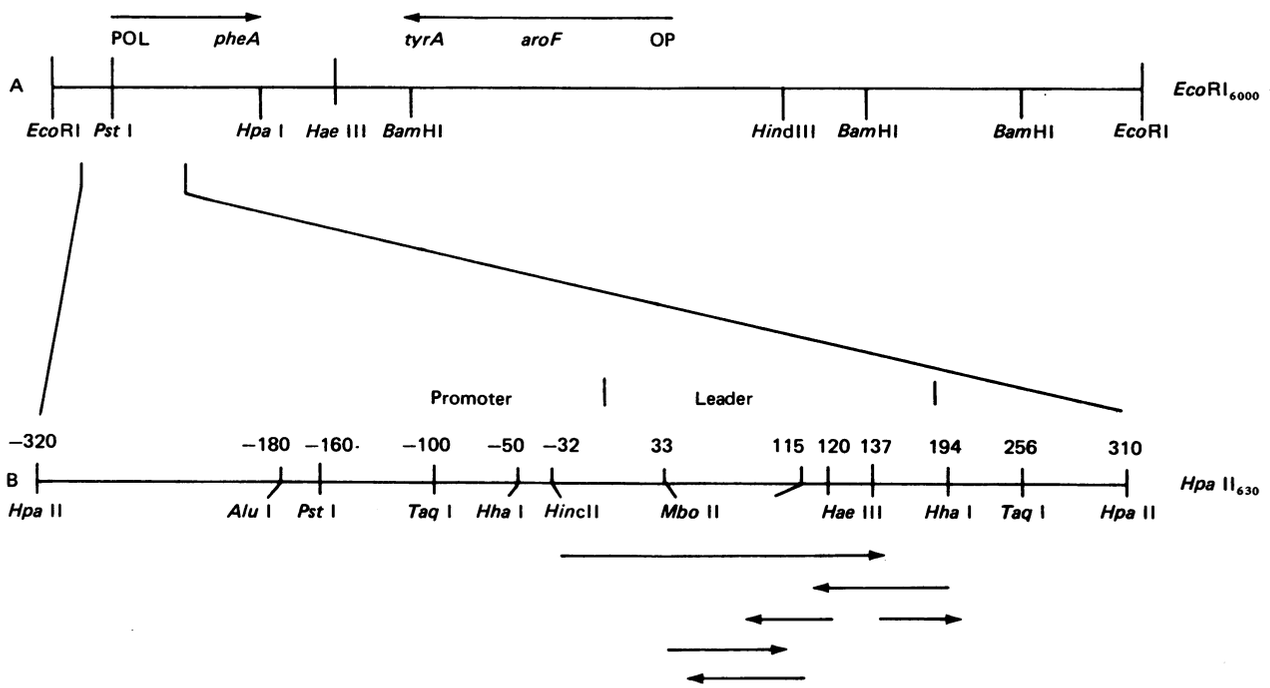

FIG. 1. (*A*) Restriction map of the 6000-bp *Eco*RI fragment of λ*tyr2* that carries the *phe* and *tyr* operons. The map is based on restriction analyses of whole plasmids (D. Killingly and K. D. Brown, unpublished results) and on the cloning described in the text. The order and orientation of the genes are in agreement with previous results (1, 6). (*B*) Restriction map of the *Hpa* II_{630} fragment carrying the *phe* promoter and leader region. The map is based on restriction analysis of ^{32}P-end-labeled restriction fragments from this region and, where indicated, is confirmed by DNA sequence analysis. The orientation of the *Hpa* II_{630} DNA fragment within the *Eco*RI_{6000} DNA fragment was deduced by restriction analysis of ^{32}P-end-labeled *Eco*RI-*Hin*cII_{400} and *Hpa* II-*Hpa* I_{700} DNA fragments. The positions of restriction sites is shown relative to position 1, the inferred transcription initiation site for the *phe* operon. There is at least one additional *Hha* I site between positions 191 and 256. *Mbo* II sites have only been determined between −32 and 194. The *Taq* I_{356} fragment is from −100 to 256. Arrows indicate the length and direction of sequence runs on which the reported DNA sequence is based. Each sequence run shown was repeated at least once.

the *Pst* I site (Fig. 1). Cleavage of the *Hpa* II_{630} fragment at the *Hin*cII site, 130 bp to the right of the *Pst* I site, prevented transcription. Thus, the promoter for the *phe* operon is located at, or close to, this *Hin*cII site. On the basis of *in vitro* transcription analyses we estimate that there is *ca.* 40% readthrough at the *phe* leader termination site. At the comparable termination sites in the *trp* operons of *E. coli* and *Samonella typhimurium* there is *ca.* 5% and 30% readthrough, respectively (4).

DNA Sequence of the Leader Region of the Phenylalanine Operon. ^{32}P-end-labeled DNA fragments from the *Hin*cII-*Hha* I_{228} region (Fig. 1) were prepared and sequenced as described by Maxam and Gilbert (10). The inferred RNA sequence from bp 1, the likely transcription initiation site for the phenylalanine operon, to bp 196 is shown in Fig. 3. The lengths of sequence runs, the extent of overlaps, and directions of sequencing are shown in Fig. 1.

DISCUSSION

Analysis of the initial transcribed segment of the *E. coli phe* operon reveals that transcription starts approximately 170 residues before the structural gene, *pheA*. A transcription termination site, which functions *in vitro* and probably *in vivo*, occurs about 25 residues before *pheA*. The features of the 170-residue leader region described below strongly suggest that the *phe* operon is regulated by translational control of transcription termination.

Transcription of the *phe* operon is initiated approximately 290 bases from one end of the *Hpa* II_{630} DNA fragment. In this region of the *Hpa* II_{630} fragment, there is a sequence of 12 base pairs that match 10 of the first 12 base pairs that are transcribed in the *trp* operon of *E. coli* (transcript sequence AAGUCACUUAAG vs. pppAAGUUCACGUAA((12). A similar correspondence exists with the *trp* operon of *S. typhimurium* (13). It is on this basis that we have assigned bp 1 in the *phe* leader region as the probable transcription start site (Fig. 3).

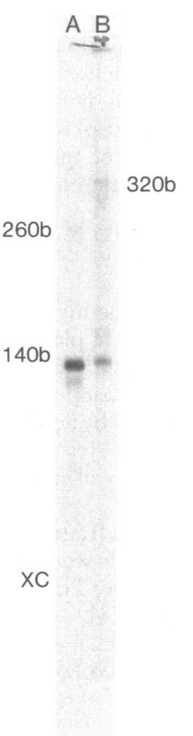

FIG. 2. RNA transcribed from either (lane A) the *Hpa* II_{570} DNA fragment carrying the regulatory region of the *E. coli trp* operon (4), or (lane B) the *Hpa* II_{630} DNA fragment carrying the regulatory region of the *E. coli phe* operon, resolved on a 10% polyacrylamide/Tris-borate/EDTA gel containing 7M urea. The major RNA in the lane A is the 140-b *trp* leader RNA. The 260-b RNA band is the result of transcription from the *trp* promoter, through the transcription termination site, to the end of the fragment (4). The lengths of these RNAs were used as standards for the estimation of the lengths of the RNAs in lane B by the method of Maniatis *et al* (9). XC indicates the location of the xylene cyanol marker dye.

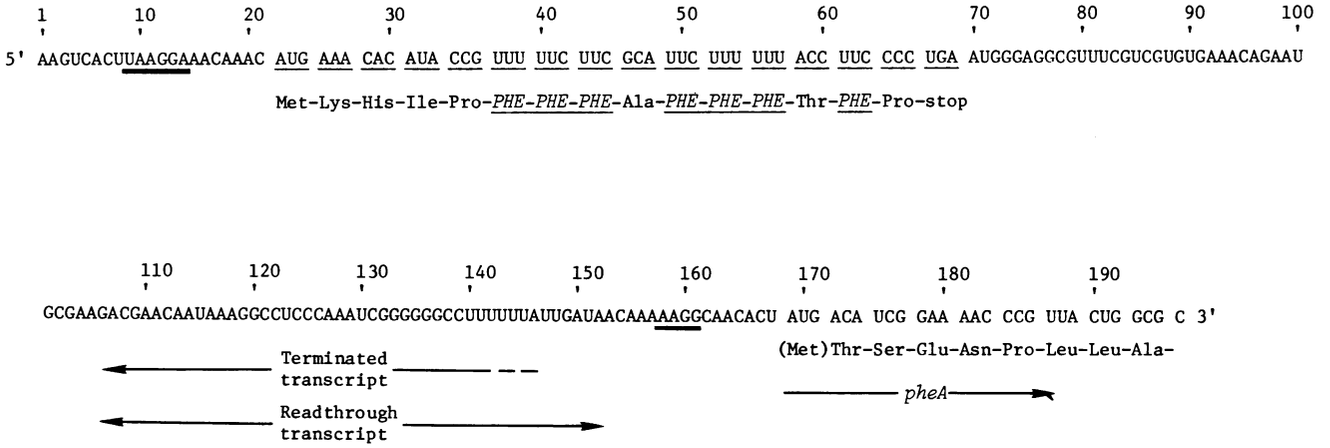

FIG. 3. The nucleotide sequence of the initial transcribed region of the *phe* operon of *E. coli*. The RNA sequence that is presented is deduced from the DNA sequence analysis of this region. The inferred site of transcription initiation is designated position 1. Contiguous bases 5′ to the translation initiation codons which are complementary to the 16S ribosomal RNA 3′ end (AUUCCUCCACUA) are underlined boldly. Amino acid codons in phase with the translation initiator codons are finely underlined. The amino acid sequence of the putative *phe* leader peptide is shown. The amino acid sequence of chorismate mutase–P-prephenate dehydratase, the *pheA* protein, is also shown. The T4e6 tryptic decapeptide of the *pheA* protein is Thr-Ser-Glu-Asn-Pro-Leu-Leu-Ala-Leu-Arg (11).

At residues 22–24 of the *phe* leader transcript there is an AUG codon in the center of a potential ribosome binding site. The purine-rich region around this presumed AUG translation initiation codon exhibits extensive homology with the ribosome binding sites for *E. coli lacZ* and *trpE* (12, 14). Also within this proposed ribosome binding site are six contiguous bases, residues 9–14, centered 11 or 12 residues 5′ to the AUG codon, which are complementary to the six 3′-terminal residues of 16S ribosomal RNA. This complementarity to 16S ribosomal RNA is a common feature of ribosome binding sites (15, 16). At approximately the same location in the *E. coli trp* leader transcript (residues 27–29) there is also an AUG translation initiation codon (12). The region around this AUG codon has been shown to bind ribosomes *in vitro* (17) and to be utilized *in vivo* for initiation of translation (18, 19).

There are 15 in phase amino acid codons in the *phe* leader transcript extending from the AUG start codon to a translation stop codon, UGA. In the *E. coli* and *S. typhimurium trp* leader transcripts there are sequences coding for 14-residue peptides (13). Although 2 of the 14 amino acids of the predicted *trp* leader peptides are tryptophan residues, it is startling to observe that out of the 15 amino acid residues in the predicted *phe* leader peptide, 7 are phenylalanine (Fig. 3). The Phe residues are clustered in two groups of three, residues 6–8 and 10–12, with the seventh Phe residue at position 14. The two tryptophan residues in the *trp* leader peptides of *E. coli* and *S. typhimurium* are adjacent, occurring at positions 10 and 11. Seven consecutive His residues occur in the predicted 16-residue *his* leader peptide of *S. typhimurium* (20). Thus, in the *phe*, *trp*, and *his* leader peptides there are multiple codons for the respective amino acids.

The *phe* leader region contains an A·T-rich region (residues 140–159) immediately preceded by a G·C-rich region (residues 119–139). This is characteristic of several transcription termination regions (21) and closely resembles those in the leader segments of the *trp* operons of *E. coli* (13), *S. typhimurium* (13), *Shigella dysenteriae* (22), and *Serratia marcescens* (G. F. Miozzari and C. Yanofsky, unpublished results). The occurrence of this termination region in the *phe* operon accounts for the approximately 140-b leader RNA transcript as the major *in vitro* transcription product of DNA carrying an intact *phe* operon promoter and leader region (Fig. 2). The precise nucleotide pair in the A·T-rich segment of the *phe* leader region at which termination occurs is not yet known. In *E. coli*, the *trp* leader transcript terminates with U 7–8 *in vitro* and U 4–8 *in vivo* (23).

The start of the *pheA* structural gene, and therefore the end of the *phe* leader region, is at base pair 169. The sequence of the 20 ribonucleotides immediately beyond the AUG codon corresponding to base pairs 169–171 correlates precisely with the sequence of the first eight amino acids of a decapeptide, T4e6 (11), derived from the *pheA* protein, chorismate mutase P-prephenate–dehydratase. There are no other possible translation start sites for the *pheA* protein since the AUG codon at position 169–171 is preceded by tandem in-phase stop codons at position 148–153 (Fig. 3). Therefore, this AUG codon must be the start codon for the *pheA* protein. Also, further 3′ to the AUG codon, we have found numerous correlations between our DNA sequence and the amino acid sequence of peptides derived from the *pheA* protein (D. Killingly, W. Stern, and K. D. Brown, unpublished results). The AUG start codon is in the center of a purine-rich potential ribosome binding site (residues 155–182). This region includes four contiguous residues (residues 158–161) complementary to the 3′ terminal sequence of 16S ribosomal RNA.

Secondary Structure of the Phenylalanine Leader Transcript. *phe* leader RNA has the potential for considerable secondary structure (Fig. 4). Much of this structure is remarkably similar to that described for the leader transcripts of the *trp* operons of *E. coli* and *S. typhimurium* (4). In particular, a G·C-rich segment (residues 119–126) is able to pair with two different regions of leader RNA. This permits two alternative stem and loop structures (Fig. 4). Residues 116–126 can pair with residues 132–142, forming a stem and loop structure immediately preceding the presumed 3′-OH end of the transcript. Stability calculations (24, 25) indicate that this structure has a free energy of $\Delta G = -23$ kcal. The analogous structure in the *E. coli trp* leader transcript ($\Delta G = -20.1$ kcal) (4), has been identified in studies of termination-relief mutants (G. Stauffer, G. Zurawski, and C. Yanofsky, unpublished results) as a region recognized in the transcription termination event. Residues 116–127 can also base-pair with residues 71–82 to form the

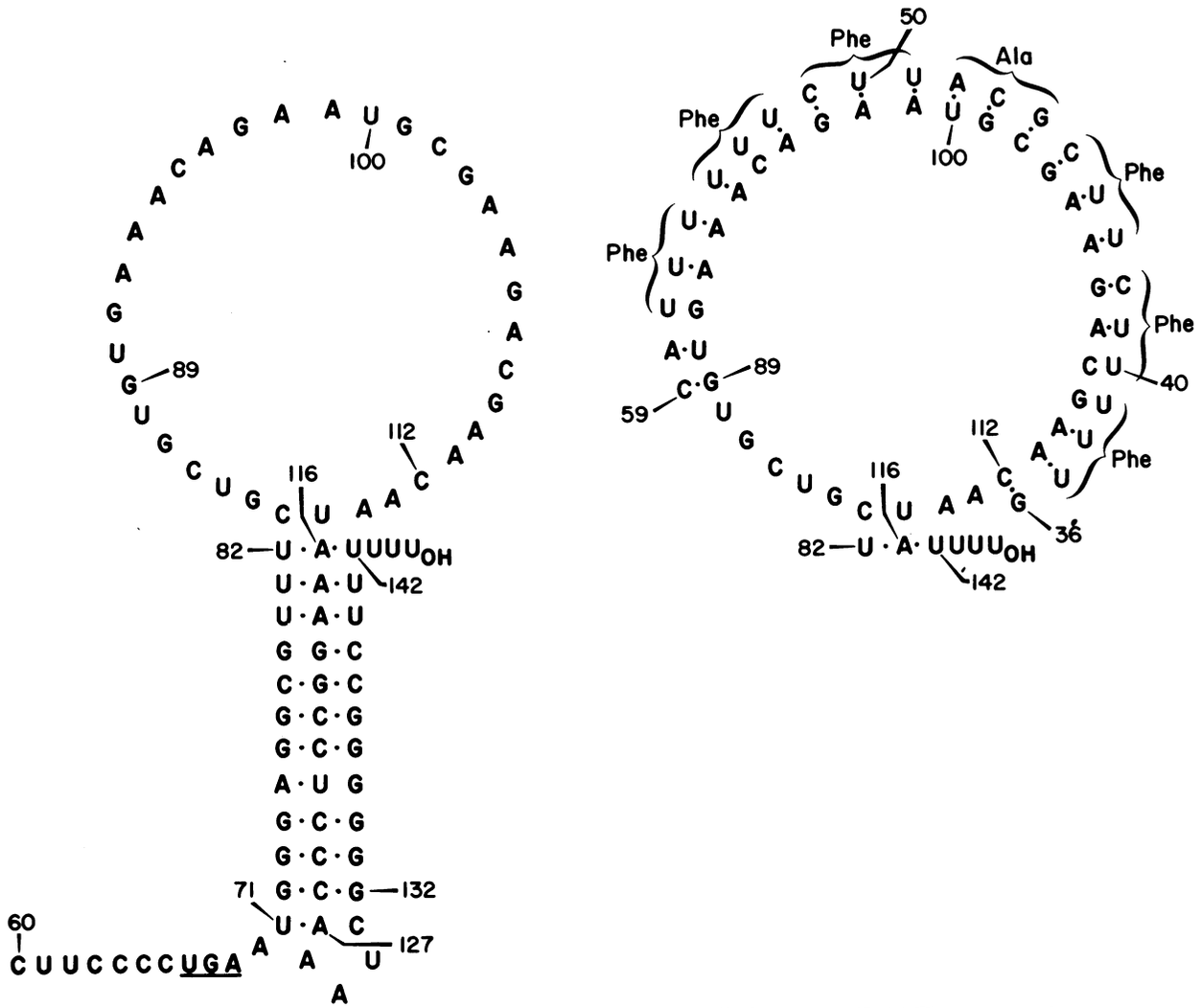

FIG. 4. Possible secondary structure in terminated leader RNA from the *E. coli phe* operon. (*Left*) The two stem-loop structures at the 3′ end of the transcript. (*Right*) The possible pairing between the Phe codon region of the transcript (residues 36–59) and the loop region from residues 89 to 112. See the text for additional explanation.

alternate stem and loop structure, with a free energy of $\Delta G = -19.3$ kcal. The analogous stem and loop structure in the *E. coli trp* leader transcript has a free energy of $\Delta G = -12.2$ kcal (4).

The similarity of the *phe* and *trp* leader regions strongly suggests that the *phe* operon is regulated by attenuation—i.e., control of transcription termination. In support of this, we have found that mutants defective in the rho termination protein (*rho*102) (26) and in a $tRNA^{Phe}$ and $tRNA^{Trp}$ modification enzyme (*trpX*) (27) have increased levels of chorismate mutase–P-prephenate dehydratase (G. P. Harper and K. D. Brown, unpublished results). These same *rho* and *trpX* mutations increase expression of the *trp* operon.

The ability of the *phe* and *trp* leader transcript to form remarkably similar stem and loop structures in the 3′-OH half of the RNA suggests that these secondary structures have an important function in attenuation. The possible roles of the secondary structure and translation of the leader transcript in the regulation of transcription termination have been discussed (4) and will be developed in more detail shortly. In particular, under *in vivo* conditions of phenylalanine (or tryptophan) excess, ribosomes would be expected to translate the full coding length of the leader peptide message and terminate translation at the leader UGA codon. A ribosome at the UGA position would mask 10 or more residues downstream to the UGA codon (16), thereby preventing the formation of the stem and loop between residues 71 and 127 in the *E. coli phe* leader transcript or residues 74 and 119 in the *E. coli trp* leader transcript. It is possible that the integrity of this stem and loop structure is an important feature of the regulation of transcription termination.

A striking feature of *phe* leader RNA is that it has the potential to form a third hydrogen bonded region with high stability (Fig. 4). This would involve base-pairing between the region coding for the seven Phe codons (residues 36–59) and residues 89–112. The resulting stem and loop has a free energy of $\Delta G = -21.7$ kcal. A comparable, though less stable stem and loop can be drawn by pairing the Trp codon region and a more distal region of *E. coli trp* leader RNA. It should be noted that the segments of *E. coli* and *S. typhimurium trp* leader RNAs that contain the tandem Trp codons are moderately resistant to ribonuclease attack *in vitro* (4). This pairing between the Phe or Trp codon region and the more distal region of the transcript may occur only when Phe or Trp codons are not translated—i.e., when the cell is deficient in charged $tRNA^{Phe}$ or $tRNA^{Trp}$. Such pairing may interfere with RNA polymerase's recognition of a termination structure in the RNA transcript, thereby permitting polymerase to continue transcription into the structural gene region of the operon.

Leila Blackman is thanked for skilled technical assistance and Bill Stern for preliminary restriction analyses. Dr. B. E. Davidson and Dr. G. Baldwin kindly provided information on CMPD peptide sequences. These studies were supported by grants from the Australian Research Grants Committee (D72/15153), from the National Science Foundation (PCM 77-24333), and from the U.S. Public Health Serice (GM 09738), and by the American Heart Association. D.K. is the recipient of a Commonwealth of Australia Postgraduate Research Award. G.Z. is a Career Investigator Fellow of the American Heart Association. C.Y. is a Career Investigator of the American Heart Association.

1. Zurawski, G. & Brown, K. D. (1976) *J. Mol. Biol.* **102,** 311–324.
2. Selker, E., Brown, K. & Yanofsky, C. (1977) *J. Bacteriol.* **129,** 388–394.
3. Brown, K. D., Bennett, G. N., Lee, F., Schweingruber, M. E. & Yanofsky, C. (1978) *J. Mol. Biol.* **121,** 153–178.
4. Lee, F. & Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4365–4369.
5. Bolivar, F., Rodriquez, R. L., Betlach, M. C. & Boyer, H. W. (1977) *Gene* **2,** 75–93.
6. Bachmann, B. J., Low, B. & Taylor, A. (1976) *Bacteriol. Rev.* **40,** 116–167.
7. Bolivar, F., Rodriquez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2,** 95–113.
8. Davidson, B. E., Blackburn, E. H. & Dopheide, T. A. A. (1972) *J. Biol. Chem.* **247,** 4441–4447.
9. Maniatas, T., Jeffrey, A. & van de Sande, H. (1975) *Biochemistry* **14,** 3787–3794.
10. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
11. Baldwin, G. (1977) Dissertation (Univ. of Melbourne, Melbourne, Australia).
12. Squires, C., Lee, F., Bertrand, K., Squires, C. L., Bronson, M. J. & Yanofsky, C. (1976) *J. Mol. Biol.* **103,** 351–381.
13. Lee, F., Bertrand, K., Bennett, G. & Yanofsky, C. (1978) *J. Mol. Biol.* **121,** 193–218.
14. Maizels, N. (1974) *Nature (London)* **249,** 647–649.
15. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
16. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4734–4738.
17. Platt, T., Squires, C. & Yanofsky, C. (1976) *J. Mol. Biol.* **103,** 411–420.
18. Schmeisser, U., Ganem, D. & Miller, J. H. (1977) *J. Mol. Biol.* **109,** 303–326.
19. Miozzari, G. F. & Yanofsky, C. (1978) *J. Bacteriol.* **133,** 1457–1466.
20. Barnes, W. M. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4281–4285.
21. Rosenberg, M., Court, D., Wulff, D. L., Shimatake, H. & Brady, C. (1977) *Nature (London)* **272,** 415–420.
22. Miozzari, G. F. & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. USA,* in press.
23. Bertrand, K., Korn, L. J., Lee, F. & Yanofsky, C. (1977) *J. Mol. Biol.* **117,** 227–247.
24. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246,** 40–41.
25. Borer, P. N., Dengler, B., Tinoco, I., Jr. & Uhlenbeck, O. C. (1974) *J. Mol. Biol.* **86,** 843–853.
26. Korn, L. J. & Yanofsky, C. (1976) *J. Mol. Biol.* **103,** 395–409.
27. Yanofsky, C. & Soll, L. (1977) *J. Mol. Biol.* **113,** 663–677.