Proc. Natl. Acad. Sci. USA
Vol. 77, No. 12, pp. 7117–7121, December 1980
Biochemistry

Nucleotide sequence and expression of *Escherichia coli trpR*, the structural gene for the *trp* aporepressor

(amino acid sequence/transcription start/*trpR* operator–promoter/autoregulation)

ROBERT P. GUNSALUS AND CHARLES YANOFSKY

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

Contributed by Charles Yanofsky, August 29, 1980

ABSTRACT The nucleotide sequence of *trpR* of *Escherichia coli* was determined. This gene codes for a polypeptide (M_r 12,356) that is 108 amino acid residues in length. NH_2-terminal, COOH-terminal, and total amino acid analyses of purified aporepressor agree with the deduced amino acid sequence and establish the translation start and stop codons of the structural gene. The transcription start site for *trpR* mRNA synthesis *in vitro* was shown to be 56 base pairs prior to the translation start site. The nucleotide sequence on either side of the transcription start site is homologous to the *trp* operon operator. Purified *trp* aporepressor, when activated by L-tryptophan, protects restriction sites in this region, the presumed *trpR* operator, from cleavage by the respective restriction endonucleases. Bound RNA polymerase protects the same restriction sites. These findings and the additional observation that *trp* repressor inhibits transcription initiation *in vitro* establish that there is a functional overlap of operator and promoter sequences in the regulatory region of the *trpR* operon. These findings indicate that expression of *trpR* is autoregulatory.

The gene *trpR* of *Escherichia coli* codes for a regulatory protein, the *trp* aporepressor (1–3). When complexed with its corepressor L-tryptophan, the aporepressor forms an active repressor complex that can bind at the operator region of the *trp* operon and prevent transcription initiation (4–6). The protein is present in low concentrations in *E. coli* (7): it has been estimated that only 20–30 repressor molecules are present per cell (4). *trpR* maps at 0 min on the *E. coli* chromosome (8) and is unlinked to either the *trp* operon at 27 min or the *aroH* operon at 37 min, two operons the *trp* repressor is known to regulate. *aroH* codes for one of the three 3-deoxy-2-keto-D-*arabino*-heptulosonic acid 7-phosphate (DAHP) synthases that catalyze the first step in aromatic amino acid biosynthesis (9, 10). *trpR* was recently cloned from the *E. coli* chromosome onto phage and plasmid vectors (11, 12). The gene is present on a 1.3-kilobase *Bam*HI fragment (11, 12). A *Sal* I site was localized within either the structural gene or its promoter, because plasmids containing deletions of either the 440- or the 860-bp *Bam*HI/*Sal* I fragment failed to direct the synthesis of functional aporepressor (11). Using these *trpR* plasmids, we have determined the nucleotide sequence of *trpR* and its preceding regulatory region. In addition, we have transferred *trpR* onto a plasmid that allows a 100-fold increase in aporepressor production. Using bacterial cultures with this plasmid as a source of *trp* repressor, we have obtained this regulatory protein in pure form. We have shown by *in vitro* studies that the *trp* repressor binds to the *trpR* regulatory region and inhibits transcription initiation. This finding suggests that expression of *trpR* is autogenously regulated.

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.

MATERIALS AND METHODS

Plasmids pRPG5 and pRPG9 are *trpR*$^+$ derivatives of pRPG4 (11) and contain the 1.3-kilobase *Bam*HI *trpR*-containing fragment inserted at the *Bam*HI site of the vector pBR322. Both plasmids were constructed by ligation of *Bam*HI-cut pRB322 and the 1.3-kilobase fragment isolated by agarose gel electrophoresis from a *Bam*HI digest of pRPG4 plasmid DNA. The ligated DNA mixture was used directly for transformation of *E. coli* W3110SRT4, and colonies were selected for ampicillin resistance (40 μg/ml). Plasmid pRPG5, isolated from one of the transformants, contains two *Sal* I sites, one of which is located asymmetrically within the *Bam*HI insert distal to the vector *Sal* I site. The other plasmid type, represented by pRPG9, contains the 1.3-kilobase *Bam*HI insert in the opposite orientation. Plasmid isolation, restriction, ligation, and transformation procedures were performed as described (11). The DNA nucleotide sequence was determined by the procedures of Maxam and Gilbert (13). Polyacrylamide gel electrophoresis was performed as described by Sanger and Coulson (14). Transcription of DNA fragments, their separation by gel electrophoresis, and RNA sequence determination were carried out by using standard procedures (15–17). *trp* aporepressor was purified to homogeneity from heat-induced cells containing a plasmid, pRPG12, in which *trpR* is fused downstream from the λ P_L promoter (ref. 11; unpublished). Automated Edman degradations and amino acid analyses were performed by the Protein Structure Laboratory, University of California at Davis, Davis, California.

RESULTS

Nucleotide and Amino Acid Sequences. Plasmids pRPG5 and pRPG9 are derivatives of pBR322 containing identical 1.3-kilobase *Bam*HI inserts from *E. coli* that express *trpR*$^+$ (11). These plasmids were used as sources of DNA fragments for sequence determination and differ only in the orientation of the 1.3-kilobase *Bam*HI insert within the pBR322 vector. The strategy employed in determining the sequence of *trpR* (Fig. 1) involved determining the nucleotide sequence of DNA fragments on either side of the *Sal* I site previously shown to be within the *trpR* operon (11). The entire sequence of each DNA strand was determined in the *trpR* and adjacent regions. Sequences at or near fragment junctions were confirmed with overlapping fragments. The complete nucleotide sequence is presented in Fig. 2.

The direction of transcription of *trpR* was previously determined to be from left to right as drawn in Fig. 1 (11). Examination of the three reading frames of the nucleotide sequence revealed only one that could code for a polypeptide of significant length (Fig. 2). This potential translated sequence begins at the ATG codon at positions 57–59 and extends to the

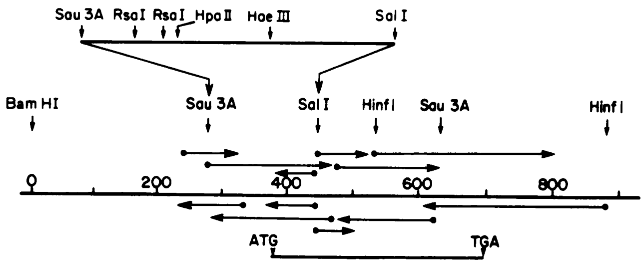

FIG. 1. Strategy employed in determining the nucleotide sequence of *trpR* of *E. coli*. The nucleotides of the 1.3-kilobase *Bam*HI fragment are numbered relative to the *Bam*HI site proximal to the *Sal* I site. Arrows above the numbered sequence show the extent of sequence runs on the antisense strand. Arrows below show sequence runs of the complementary strand. The coding region for *trpR* is indicated as well as the orientation of transcription.

termination codon TGA at positions 381–383. The polypeptide predicted from this sequence would be 108 amino acid residues in length. In-frame methionine codons also exist at residue positions 11, 42, and 66; initiation at these would yield polypeptides 98, 67, and 43 residues in length, respectively. To determine the size of the *trpR* polypeptide, the *trp* aporepressor was purified to homogeneity (unpublished data) and subjected to protein structure analyses.

Amino Acid Analyses. Amino acid analyses of acid-hydrolyzed *trp* aporepressor (Table 1) are consistent with a polypeptide length of 108 residues. The experimentally determined values for each amino acid are in good agreement with the amino acid content predicted from the nucleotide sequence (Fig. 2). The low methionine value is partially explained by NH_2-terminal sequence data (see below) indicating that 90% of our purified *trp* aporepressor lacks the NH_2-terminal methionine residue.

NH_2- and COOH-Terminal Analyses. The NH_2-terminal amino acid sequence of the *trp* aporepressor was determined by automatic Edman degradation on approximately 0.75 mg of protein. The results obtained established the NH_2-terminal sequence of the *trp* aporepressor to be Met-Ala-Glu-Glu-Ser-Pro-Tyr-Ser-Ala. These results identify the methionine codon at nucleotide positions 57–59 (Fig. 2) as the translation start codon for the *trpR* protein. Approximately 90% of the preparation of purified repressor examined had alanine as the NH_2-terminal residue; the remainder had methionine. The NH_2-terminal methionine residue is apparently largely removed *in vivo*.

On the basis of the nucleotide sequence of *trpR* we expect the COOH-terminal tryptic peptide of the *trp* aporepressor to be Ser-Asp. To isolate this acidic peptide we prepared a tryptic digest of 0.50 mg of protein and separated the peptides by two-dimensional electrophoresis/chromatography on a silica G thin-layer plate (18). Electrophoresis was performed at pH 6.5 for a very brief period because Ser-Asp would be expected to migrate rapidly. Fluorescamine staining revealed a major fast-moving negatively charged peptide that, when eluted and hydrolyzed, was found to have the composition Ser, Asp. Other peptides eluted from a similar thin-layer plate that was run for a much longer time at pH 6.5 had amino acid compositions verifying the predicted polypeptide sequence for aporepressor segments 16–27 and 64–84. On the basis of these findings we conclude that the *trpR* polypeptide is 108 residues in length and terminates in the sequence Ser-Asp-COOH (Fig. 2).

Table 1. Amino acid content of the *trp* aporepressor

Residue	Residues per molecule	
	Analysis*	Theoretical†
Ala	10.10	10
Arg	9.30	9
Asp	8.56	9
Glu	18.63	19
Gly	5.66	5
His	2.26	2
Ile	2.85	3
Leu	18.57	19
Lys	4.16	4
Met	2.52	4
Phe	0.80	1
Pro	3.81	4
Ser	5.46	6
Thr	4.03	4
Tyr	2.93	2
Val	5.28	5
Trp	—	2

* Average value for duplicate 48-hr hydrolysates.
† Calculated assuming that the *trpR* polypeptide is 108 amino acid residues in length.

Regulatory Region of the *trpR* Operon. Inspection of the nucleotide sequence preceding *trpR* reveals a region 21 base pairs long that is centered approximately 56 base pairs before the *trpR* translation initiation site and that shows extensive homology to the operators of both the *trp* and *aroH* operons (refs. 6 and 19; Figs. 2 and 3). To determine whether this DNA region can bind *trp* repressor, restriction site protection studies were carried out. A DNA fragment ($Sau3A_{186}$) spanning the presumed *trpR* operator was preincubated with *trp* repressor

```
 -50         -40        -30        -20        -10        -1 1        10         20         30         40         50
AATGGGGACGTCG TTACTGATCC GCACGTTTAT GATATGCTAT CGTACTCTTT AGCGAGTACA ACCGGGGGAG GCATTTTGCT TCCCCCGCTA ACAATGGCGA CATATT
                                                                      |---------|    |---------|
                                                                          1               1
                                    86                                         116                                 146
ATG GCC CAA CAA TCA CCC TAT TCA GCA GCG ATG GCA GAA CAG CGT CAC CAG GAG TGG TTA CGT TTT GTC GAC CTG CTT AAG AAT GCC TAC
MET ALA GLN GLN SER PRO TYR SER ALA ALA MET ALA GLU GLN ARG HIS GLN GLU TRP LEU ARG PHE VAL ASP LEU LEU LYS ASN ALA TYR

                                    176                                     206                                     236
CAA AAC GAT CTC CAT TTA CCG TTG TTA AAC CTG ATG CTG ACG CCA GAT GAG CGC GAA GCG TTG GGG ACT CGC GTG CGT ATT GTC GAA GAG
GLN ASN ASP LEU HIS LEU PRO LEU LEU ASN LEU MET LEU THR PRO ASP GLU ARG GLU ALA LEU GLY THR ARG VAL ARG ILE VAL GLU GLU

                                    266                                     296                                     326
CTG TTG CGC GGC GAA ATG AGC CAG CGT GAG TTA AAA AAT GAA CTC GGC GCA GGC ATC GCG ACG ATT ACG CGT GGA TCT AAC AGC CTG AAA
LEU LEU ARG GLY GLU MET SER GLN ARG GLU LEU LYS ASN GLU LEU GLY ALA GLY ILE ALA THR ILE THR ARG GLY SER ASN SER LEU LYS

                                    356                                         390      400
GCC GCG CCC GTC GAG CTG CGC CAG TGG CTG GAA GAG GTG TTG CTG AAA AGC GAT TGA TTTTGTA GGCCTGATAA
ALA ALA PRO VAL GLU LEU ARG GLN TRP LEU GLU GLU VAL LEU LEU LYS SER ASP END

     410        420        430        440        450        460        470        480        490        500  505
GACGTGGCGC ATCAGGCATC GTGCGACCGA ATGCCGGATG CGGCGTGAAG GCCTTATCCG TCCTACAAAT ACCCGTAATT TCAATATGTT TGGTAGGCAT GATAA
    |-----------|    |-------------|
          2                 2
```

FIG. 2. Nucleotide sequence of *trpR* of *E. coli* and the predicted amino acid sequence of its polypeptide product. The sequence is numbered relative to the experimentally determined start of transcription (+1). Translation of *trpR* begins at the methionine codon at nucleotide positions 57–59 and runs 108 codons before the stop codon at 381–383. Regions of the initial and presumed terminal segments of the transcript that could form stable secondary structures are indicated by underlining (labeled 1 and 2).

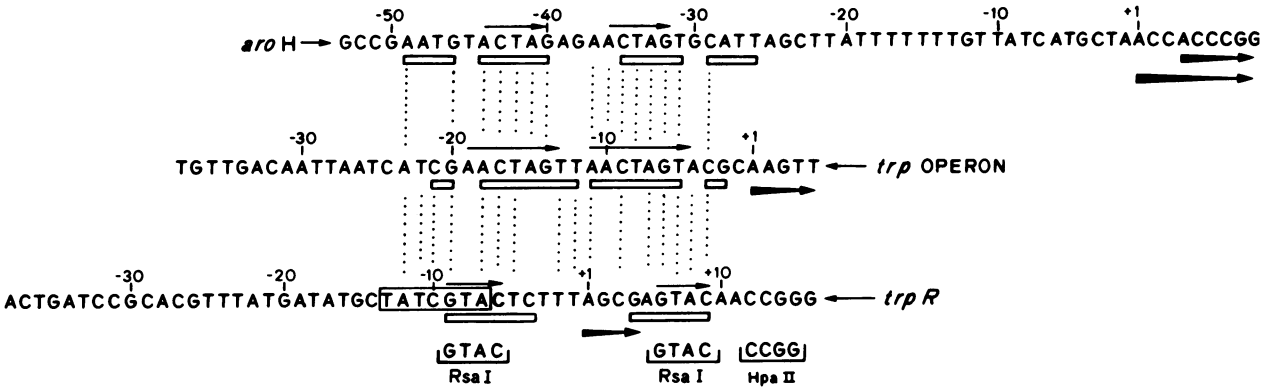

FIG. 3. The promoter-operator regions of the *aroH, trp,* and *trpR* operons. The three operators are aligned relative to their homologous sequences and are numbered from their respective transcription start sites. The vertical dotted lines indicate positions of identity. A likely Pribnow sequence in the −10 region of the *trpR* promoter is enclosed in a box. Direct sequence repeats are indicated by the thin arrows above the sequences, while regions of twofold symmetry are shown by the narrow bars immediately under the nucleotide sequences. The location of *Rsa* I and *Hpa* II restriction sites (*trpR*) that are protected by bound *trp* repressor are also indicated. Solid arrows under the sequences indicate the start and direction of transcription.

and then the restriction endonuclease *Rsa* I was added and incubation was continued. This endonuclease normally cleaves the DNA fragment at nucleotide positions −7 and +7 (Fig. 3). Fig. 4 shows that *trp* repressor protects the two *Rsa* I sites in the DNA fragment. When L-tryptophan is omitted from the incubation mixture, protection is not observed. Similarly, *trp* repressor was found to protect the *Hpa* II site at nucleotide positions +12 to +15 (Fig. 3) but not the *Hae* III site at positions +59 to +62 (data not shown). These observations show that *trp* repressor binds to the DNA region that is homologous to the operators of the *trp* and *aroH* operons.

To determine if RNA polymerase binds to the *trpR* operator region, protection experiments similar to those performed with the *trp* repressor were carried out. When the $Sau3A_{186}$ DNA fragment was preincubated with RNA polymerase prior to addition of *Rsa* I endonuclease, the two *Rsa* I sites protected by bound *trp* repressor were also protected (Fig. 4). This finding localizes the *trpR* promoter to the region containing the *trpR* operator sequence. To establish the location of the *trpR* promoter more precisely, *in vitro* transcription experiments were performed with the objective of determining the transcription start site. A *Bam*HI/*Hae* III_{390} DNA fragment (Fig. 1) containing the initial segment of *trpR* and its preceding region was transcribed *in vitro* (Fig. 5). The [α-^{32}P]CTP-labeled transcripts were separated on 10% acrylamide/7 M urea gels (20) and audioradiograms were prepared. Two major RNA transcripts

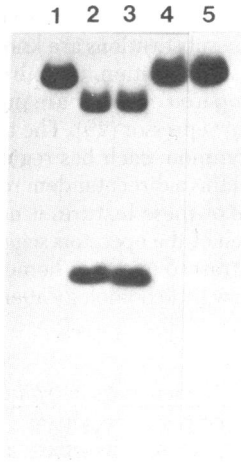

FIG. 4. Autoradiogram showing protection by RNA polymerase and *trp* repressor of the *Rsa* I restriction sites in the *trpR* regulatory region. Each reaction mixture contained [γ-^{32}P]ATP end-labeled $Sau3A_{186}$ DNA fragment in a total volume of 20 μl containing 6 mM Tris·HCl (pH 7.4), 6 mM $MgCl_2$, 10 mM NaCl, and 6 mM 2-mercaptoethanol. Where indicated, RNA polymerase (1.4 μg), L-tryptophan (50 μM), *trp* aporepressor (10 ng), and *Rsa* I endonuclease (1 unit) were added to the reaction mixture. The reaction mixtures containing RNA polymerase were preincubated for 5 min at 37°C prior to addition of restriction endonuclease. After an additional 15 min at 37°C, the samples were heated at 65°C for 10 min and 4 μl of the gel loading dye [75% (vol/vol) glycerol with 0.025% xylene cyanol blue and 0.025% bromphenol blue] was added. The samples were applied to a 5% polyacrylamide/Tris/borate slab gel (20) and electrophoresis was carried out at 200 V for 4 hr. Lane 1, undigested $Sau3A_{186}$ fragment; lane 2, $Sau3A_{186}$ fragment + *Rsa* I endonuclease; lane 3, $Sau3A_{186}$ fragment + *Rsa* I endonuclease + aporepressor; lane 4, $Sau3A_{186}$ fragment + *Rsa* I endonuclease + aporepressor + L-tryptophan; lane 5, $Sau3A_{186}$ fragment + *Rsa* I endonuclease + RNA polymerase.

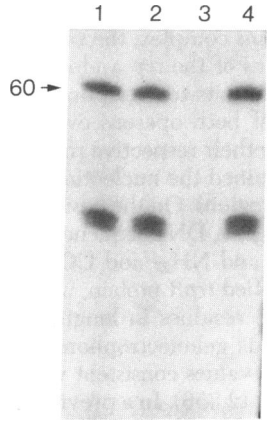

FIG. 5. Repression of transcription of *trpR in vitro*. Transcription of the *Bam*HI/*Hae* III_{390} fragment (Fig. 1) was carried out as described (15). The reaction mixture contained 300 μM UTP, 150 μM ATP, 150 μM GTP, 20 μM [α-^{32}P]CTP, 100 mM KCl, 3% (vol/vol) glycerol, 1.44 μg of RNA polymerase, 20 mM Tris acetate, 4 mM magnesium acetate, 0.1 mM dithiotheitol, and 0.1 mM EDTA. Where indicated, *trp* aporepressor (10 ng) and 50 μM L-tryptophan were added prior to incubation at 37°C. The reaction was carried out for 20 min at 37°C and then stopped by heating for 2 min at 95°C. Samples were electrophoresed on 10% polyacrylamide/Tris/borate/EDTA 7 M urea gels (15). The approximate size of the larger transcript was based on comparison with the 140-nucleotide *trp* leader transcript. This size estimate was verified by fingerprinting an RNase T1 digest of this transcript. Lane 1, DNA fragment + RNA polymerase; lane 2, DNA fragment + RNA polymerase + *trp* aporepressor; lane 3, DNA fragment + RNA polymerase + *trp* aporepressor + L-tryptophan; lane 4, DNA fragment + RNA polymerase + L-tryptophan.

were observed, a ≈40-nucleotide terminated transcript and a 60-nucleotide run-off transcript (Fig. 5). The sequence of the 60-nucleotide transcript was determined by standard RNase T1 digestion and fingerprinting of the eluted transcript followed by secondary cleavage of the T1 oligonucleotides with RNase A and identification of the digestion products (15–17). Identical analyses were carried out with [α-^{32}P]GTP-, [α-^{32}P]UTP- and [α-^{32}P]ATP-labeled transcripts. These analyses were in perfect agreement with the DNA sequence and establish the transcription start site at the nucleotide pair designated +1 in Figs. 2 and 3. The 5′ nucleotide determination was based on the observation that the triphosphate-containing oligonucleotide was labeled by A, G, or C, but not U. Fingerprints of RNase T1 digests of the 40-nucleotide transcript indicate that this transcript originates at the same start site as the 60-nucleotide transcript and ends somewhere in the vicinity of base pair 40.

Because the transcription start site was localized within the presumed *trpR* operator region (Fig. 3), *in vitro* transcription studies were performed to determine whether the *trp* repressor inhibits transcription initiation on the *trpR* operon. Preincubation of the *Bam*HI/*Hae* III$_{390}$ fragment with *trp* repressor prior to addition of RNA polymerase essentially eliminates the appearance of the 40- and 60-nucleotide *trpR* transcripts (Fig. 5). Again, omission of the corepressor, L-tryptophan, from the reaction mixture allows transcription. These observations support the restriction site protection data and indicate that both repressor and polymerase bind to the same region of the *trpR* operon. Accordingly, repressor can regulate its own synthesis.

DISCUSSION

The *trp* and *aroH* operons of *E. coli* are regulated by the polypeptide product of *trpR*, the *trp* aporepressor, in response to changes in the intracellular concentration of the corepressor, L-tryptophan (1–5, 9, 19). When the aporepressor is complexed with tryptophan, this complex, the *trp* repressor, can bind at the operator regions of the *trp* and *aroH* operons (5, 6, 19). Bound repressor inhibits transcription initiation because the operator regions of both operons overlap RNA polymerase recognition sites in their respective promoters (5, 6, 19).

We have determined the nucleotide sequence of *trpR* and its adjacent DNA regions. On the basis of cloning data that locate the structural gene, DNA sequence determinations, amino acid composition, and NH_2- and COOH-terminal sequence analyses of the purified *trpR* protein, we conclude that the *trpR* polypeptide is 108 residues in length. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis studies with the purified protein give values consistent with the calculated molecular weight (M_r 12,356). In a previous report we described the plasmid-directed synthesis *in vitro* of a polypeptide (M_r 24,000) that we considered the presumptive *trpR* polypeptide (11). We now believe that this polypeptide is the product of a fused λ–bacterial sequence in the plasmid template. An independent estimate of the size of the *trpR* polypeptide based on DNA sequence analyses and polypeptide sizing (21) is slightly lower than our determined value. The exact size of the native *trp* repressor is not known. Considering the subunit molecular weight, 12,356, previous estimates of the size of the native *trp* repressor (3, 5) would be compatible with a tetrameric structure for the protein in solution.

Immediately following the structural gene of the operon there is a G+C-rich region, with dyad symmetry, followed by A+T regions, that could function as the transcription termination site of the operon (Fig. 2). The calculated ΔG (22) for formation of the stem and loop of the corresponding transcript segment is −23.8 kcal/mol (1 kcal = 4.184 kJ). The ΔG values for the equivalent structures in the *trp* and *aroH* operons are −20 and −24 kcal/mol, respectively (22, 23).

Of the 108 amino acids residues in the *trpR* polypeptide, 14% are basic and 15% are acidic. These values are only slightly higher than those for the *E. coli trpB* and *trpA* biosynthetic polypeptides, which have 13% and 10% basic and 12% and 11% acidic amino acid residues, respectively (24). In the *trpR* polypeptide the charged residues of both types are distributed almost uniformly throughout the polypeptide chain. This arrangement is similar to that of the *lac* repressor polypeptide (25, 26) but contrasts with the structure of the λ repressor protein, in which basic residues are clustered in the NH_2-terminal portion of the polypeptide chain (27, 28).

Binding-protection studies with *trp* repressor and RNA polymerase, and *in vitro* transcription analyses, indicate that the *trpR* operon has an operator that overlaps the site of transcription initiation. This presumed operator region is homologous to the *aroH* and *trp* operon operator regions (Fig. 3). This homology has also been noted by others (21).

In addition to the observation that bound *trp* repressor inhibits transcription initiation *in vitro* on all three operons (refs. 6 and 19; this study), it should be noted that the three operators are located at different positions relative to their respective promoters. The *aroH* operator overlaps the −35 RNA polymerase recognition site, the *trp* operon operator overlaps the −10 recognition site, and the *trpR* operator is centered at the transcription initiation site and extends to the −10 polymerase recognition region (Fig. 3). The mechanism of repressor action appears to be identical for all three operons despite the different locations of the three operators in their respective promoters.

Inspection of the three homologous operator sequences reveals that there are 11 conserved base pairs. Of the 9 operator-constitutive point mutations in the *trp* operon operator for which the nucleotide substitutions are known (6), 8 occur at 3 of these 11 positions. In addition, the purines at 7 of these 11 positions in the *trp* operon operator are protected from methylation by bound *trp* repressor (29). The three operators have other features in common: each has regions of twofold symmetry and each contains a direct tandem repeat sequence (Fig. 3). The significance of these features is not known; however, their presence in each of the operators suggests some functional importance. In contrast to operator homology, the three promoter sequences show little homology at either the −10 or −35 recognition regions.

Table 2. Codon usage in *trpR* of *E. coli*

Phe TTT 1	Ser TCT 1	Tyr TAT 1	Cys TGT 0
Phe TTC 0	Ser TCC 0	Tyr TAC 1	Cys TGC 0
Leu TTA 4	Ser TCA 2	End TAA 0	End TGA 1
Leu TTG 4	Ser TCG 0	End TAG 0	Trp TGG 2
Leu CTT 1	Pro CCT 0	His CAT 1	Arg CGT 5
Leu CTC 2	Pro CCC 2	His CAC 1	Arg CGC 4
Leu CTA 0	Pro CCA 1	Gln CAA 3	Arg CGA 0
Leu CTG 8	Pro CCG 1	Gln CAG 4	Arg CGG 0
Ile ATT 2	Thr ACT 1	Asn AAT 2	Ser AGT 0
Ile ATC 1	Thr ACC 0	Asn AAC 3	Ser AGC 3
Ile ATA 0	Thr ACA 0	Lys AAA 3	Arg AGA 0
Met ATG 4	Thr ACG 3	Lys AAG 1	Arg AGG 0
Val GTT 0	Ala GCT 0	Asp GAT 3	Gly GGT 0
Val GTC 3	Ala GCC 3	Asp GAC 1	Gly GGC 3
Val GTA 0	Ala GCA 3	Glu GAA 6	Gly GGA 1
Val GTG 2	Ala GCG 4	Glu GAG 6	Gly GGG 1

Major proteins or those that must be synthesized rapidly might be expected to be coded for by transcripts rich in codons for the major isoaccepting tRNA species of the respective organism. Conversely, proteins present in trace amounts, such as repressors, might be specified by transcripts rich in minor tRNA species. Consistent with these expectations is the observation that messages for ribosomal proteins of *E. coli* are rich in codons that are translated by the major isoaccepting tRNA species (30). Comparison of codon usage in *trpR* (Table 2) and *lacI* of *E. coli* with that in *trpB* and *trpA* of the same organism does not reveal any noticeable overutilization of rarely used codons in the messages for the two repressor proteins. For example, codons that are not used or are used infrequently in the two *trp* operon genes, AUA (Ile), GUA (Val), AAG (Lys), and CGA, CGG, AGA, AGG (Arg) are correspondingly rare in *trpR* and *lacI*. Unless the use of a few rare codons is sufficient to reduce the rate of synthesis of the *trpR* polypeptide, it would appear that translational inefficiency has not been a basis of codon selection during the evolution of *trpR* of *E. coli*. The absence of a recognizable Shine–Dalgarno sequence (31) preceding *trpR* may indicate that initiation of translation in inefficient.

Transcription of the initial segment of the *trpR* operon *in vitro* yields two major transcripts, one terminating at about base pair 40, just beyond a region of dyad symmetry in DNA (Fig. 2). The corresponding segment of the transcript would form a stem and loop structure with an estimated ΔG of -26 kcal/mol (22). Both transcripts have the same 5′ end, indicating that they were initiated at the same site in the *trpR* promoter. Synthesis of the two transcripts is inhibited by *trp* repressor. The significance of the existence of the 40-base terminated transcript is unclear. It could be an artifact caused by the conditions employed in *in vitro* transcription experiments. However, if this terminated transcript is an authentic product of the operon *in vivo*, termination of transcription in the *trpR* leader region may be used to reduce expression of the operon below the level that would be dictated by polymerase and repressor action at the promoter–operator region. Alternatively, termination in the leader region could be regulated, as in other operons (15), to allow variation in the rate of synthesis of intact *trpR* messenger RNA under appropriate environmental conditions. Regardless of the significance of the terminated transcript, it is clear that expression of the *trpR* operon is autogenously regulated by the *trp* repressor.

The authors appreciate the helpful advice of Dr. Brian Nichols. These studies were supported by grants from the U.S. Public Health Service (GM09738), the National Science Foundation (PCM77-24333), and the American Heart Association (69C-15). R.P.G. is a postdoctoral fellow of the U.S. Public Health Service. C.Y. is a Career Investigator of the American Heart Association.

1. Cohen, G. & Jacob, F. (1959) *C. R. Acad. Sci. Ser. D.* **248,** 3490–3492.
2. Yanofsky, C. (1971) *J. Am. Med. Assoc.* **218,** 1026–1035.
3. Zubay, G., Morse, D. E., Schrenk, W. I. & Miller, J. H. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1100–1103.
4. Rose, J. K. & Yanofsky, C. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 3134–3138.
5. Squires, C. L., Lee, F. D. & Yanofsky, C. (1975) *J. Mol. Biol.* **92,** 93–111.
6. Bennett, G. N. & Yanofsky, C. (1978) *J. Mol. Biol.* **121,** 179–192.
7. Morse, D. E. & Yanofsky, C. (1969) *J. Mol. Biol.* **44,** 185–193.
8. Bachmann, B. J. & Low, K. B. (1980) *Microbiol. Rev.* **44,** 1–56.
9. Brown, K. D. (1968) *Genetics* **60,** 31–48.
10. Pittard, J., Camakaris, L. & Wallace, B. J. (1969) *J. Bacteriol.* **97,** 1242–1247.
11. Gunsalus, R. P., Zurawski, G. & Yanofsky, C. (1979) *J. Bacteriol.* **140,** 106–113.
12. Roeder, W. & Somerville, R. L. (1979) *Mol. Gen. Genet.* **176,** 361–368.
13. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
14. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87,** 107–110.
15. Oxender, D. L., Zurawski, G. & Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5524–5528.
16. Barrell, B. G. (1971) in *Procedures in Nucleic Acid Research,* eds. Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), Vol. 2, pp. 751–779.
17. Squires, C., Lee, F. D., Bertrand, K., Squires, C. L., Bronson, M. L. & Yanofsky, C. (1976) *J. Mol. Biol.* **103,** 351–381.
18. Wieland, T. & Georgopoulos, D. (1964) *Biochem. Z.* **340,** 476–482.
19. Zurawski, G., Gunsalus, R. P., Brown, K. D. & Yanofsky, C. (1980) *J. Mol. Biol.,* in press.
20. Maniatis, T., Jeffrey, A. & van deSande, H. (1975) *Biochemistry* **14,** 3787–3794.
21. Singleton, C. K., Roeder, W. D., Bogosian, G., Somerville, R. L. & Weith, H. L. (1980) *Nucleic Acids Res.* **8,** 1551–1560.
22. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London)* **246,** 40–41.
23. Wu, A. M. & Platt, T. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5442–5446.
24. Nichols, B. P. & Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5244–5248.
25. Farabaugh, P. J. (1978) *Nature (London)* **274,** 765–769.
26. Weber, K. & Geisler, N. (1978) in *The Operon,* eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 155–175.
27. Sauer, R. T. & Anderegg, R. (1978) *Biochemistry* **17,** 1092–1100.
28. Sauer, R. T. (1978) *Nature (London)* **276,** 301–302.
29. Oppenheim, D. S., Bennett, G. N. & Yanofsky, C. (1980) *J. Mol. Biol.,* in press.
30. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. & Dennis, P. P. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1697–1701.
31. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254,** 34–38.