ABSTRACT A messenger RNA fragment about 220 nucleotides long has been isolated from 32P-labeled tryptophan operon mRNA of Escherichia coli. When point mutations at the end of trpB and the beginning of trpA were introduced, the resulting nucleotide changes were found; hence the mRNA fragment must include the trpB-trpA intercistronic region. Most of the nucleotide sequences can be assigned to specific locations in the structural genes, based on the amino-acid sequences of the trpB and trpA proteins. In vitro, ribosomes bind to this piece of mRNA and protect from nuclease attack a region about 40 nucleotides long, containing a central AUG codon. The triplet codons to the 3′ side of this AUG correspond to the first seven amino acids of the trpA protein; the codons to the 5′ side correspond to the last six amino acids of the trpB protein. Translation of trpB is terminated by single UGA codon, which overlaps the AUG initiation codon: UGAUG. Thus the untranslated “intercistronic” region consists of only two nucleotides. The RNA sequence spanning this region undoubtedly fulfills two functions, specifying ribosome recognition signals as well as encoding amino-acid sequences.

The nature of the “punctuation” signals governing the translation of messenger RNA is of particular importance in understanding regulation of protein synthesis. Polypeptide chain termination can be specified by any of the three “non-sense” triplet codons, UAG, UAA, or UGA (1). The principal codon used for polypeptide chain initiation is AUG (1), but it is generally believed that AUG alone is not sufficient to initiate translation. The additional signals that are required remain unknown (2). In polycistronic mRNA, “intercistronic” regions may occur between the final codon of one gene and the initiation codon of the next—several such regions from RNA bacteriophage have been sequenced (3). Studies on the histidine operon of Salmonella typhimurium indicate that at least one nucleotide occurs between the termination codon of hisD and the initiation codon of the following gene, hisC, but no upper limit could be placed on the size of this region (4).

Recently, procedures to isolate defined segments of mRNA by specific hybridizations and refined techniques of nucleotide sequencing have made it possible to determine directly the primary structure of bacterial messenger RNA labeled in vivo with 32P (5, 6). We have used these procedures to purify and sequence a segment of tryptophan (trp) operon mRNA from Escherichia coli containing the intercistronic region between trpB and trpA (the last two genes in the operon) and overlapping the structural genes on either side. The ribosome-binding approach pioneered by Steitz (7) was used to demonstrate that the “initiator” region for the trpA polypeptide is still potentially functional in this mRNA sequence. We show here that the ribosome-protected fragment spans the end of trpB as well as the beginning of trpA, and that the intercistronic region consists of only two untranslated nucleotides. By contrast, the known intercistronic regions in RNA bacteriophage vary in length from 25 to over 600 nucleotides; in these cases the intercistronic region may have specific functions (3).

MATERIALS AND METHODS

Labeling and RNA Purification. An E. coli strain merodiploid for the internal deletion trpΔLD102 (formerly trpΔED102) produces high levels of distal trp mRNA (8); we estimate that >100 copies of trp operon mRNA are present per cell under conditions that result in maximum derepression of the operon and maximum incorporation of 32P into trp mRNA (6).

A 5 ml cell suspension (OD660 = 2) was labeled with 32P and after cell lysis the cellular RNA was purified by phenol extraction, chloroform–methyl alcohol extraction, and ethanol precipitations (6). This RNA fraction was hybridized in 0.5 ml of TKE buffer (0.33 M KCl, 10 mM Tris-HCl, pH 7.3, 1 mM EDTA) to the complementary separated strands (9) of phage DNA (Fig. 1) for 18 hr at 50°C, and the DNA-RNA hybrids were collected on nitrocellulose filters. Unhybridized portions of mRNA were digested with T1 RNase (2.5 units/ml in 4 ml of TKE buffer, 15 min at 37°C), residual RNase was inactivated with iodoacetate (12), and the RNA was eluted at 95°C for 12 min into 1.5 ml of 5 mM Tris-HCl, pH 7.3, 1 mM EDTA, containing 200–300 μg of tRNA as carrier. The eluate was treated with DNase, extracted with phenol, and the RNA was precipitated with ethanol. This RNA sample was ready for either a subsequent hybridization, or for nuclease digestion and fingerprint analysis. Starting with 60 mCi of 32P, the final yield was usually about 104 cpm per nucleotide.

Ribosomes. All steps were performed at 4°C. Frozen E. coli B cells were ground with alumina (2.5 × cell weight), and extracted with HG buffer (20 mM N-2-hydroxyethylpiperazine-N’-2-ethanesulfonic acid (titrated with NH4OH), pH 7.5, 60 mM NH4Cl, 10 mM Mg(AcO)2, 0.3 mM dithiothreitol) containing 3 μg/ml of DNase (Worthington, DPFF grade). Alumina and cell debris were removed by centrifugation (30,000 × g, 30 min) and the ribosomes were pelleted from the supernatant at 165,000 × g for 2 hr. This ribosome-free “S-100” supernatant was used for tRNA charg-
mutations, trpB244 (10) and trpA38am (11), recombine with the transducing phage λ trpB87 and φ80trpA37. We purified trp mRNA from trpΔLD108 (8) and Δ[lonBtrpAB4] (11) strains by hybridization to λ trpB87 DNA and subsequently to φ80-trpA37 DNA. After each hybridization, the hybrids were treated with ribonuclease to remove unhybridized "tails," thus the final mRNA fragment is expected to correspond solely to the region of the trp operon carried in common by both phage.

ing and formylation (see below). The ribosomal pellet was resuspended in HG buffer, centrifuged again, and these "low-salt washed" ribosomal pellets were resuspended to one-fifth their original volume in HG buffer (A_260 nm = 1200) and stored in liquid nitrogen.

Charged tRNA. Stripped tRNA was prepared according to Zubay (13), since our commercial samples of tRNA had methionine-accepting activities severalfold lower than freshly purified tRNA. Charging and formylation were performed at 37° for 10 min in a reaction mixture containing one-fifth volume dialyzed "S-100" and final concentrations of 10 mM Hepes(NH_4), pH 7.2, 30 mM NH_4Cl, 9 mM Mg(ACl)_2, 0.15 mM dithiothreitol, 2 mM ATP, 80 μM each of the 20 amino acids, 8 μM stripped tRNA (4 mg/ml), and 0.3 mg/ml of calcium leucovorin. The charged tRNA was purified by phenol extraction and ethanol precipitation; its methionine-acceptor activity was 75 μmol of methionine per A_260 nm unit. About 75% of the methionine-charged species was formylated.

RESULTS

Analysis of the Purified mRNA. To establish that the RNA fragment corresponded to trp operon messenger RNA, we compared purified wild-type RNA with that from two isogenic strains of E. coli, one carrying the mutation trpB244 and the other trpA38am (Fig. 1). Two-dimensional fingerprints of the RNA obtained are shown in Fig. 2. The trpB244 mutation changes oligonucleotide A-U-A-A-A-G to A-U-A-A-U-G. In the trpB244 protein the lysine residue 16 amino acids from the carboxy-terminus is replaced by asparagine (10), thus the codon change is from AAA (lysine) to AAU (asparagine). The amber codon (UAG) of trpA38am was generated in two steps via an ochre codon (UAA) from the codon for lysine in the trpA protein (11); this alters oligonucleotide C-A-A-A-G to C-U-A-A-G-G, introducing a new Trp RNase cleavage site and changing the AAA (lysine) codon to the UAG (amber) codon as predicted by the genetic data.

Those oligonucleotides corresponding to translated trpB mRNA alone can be determined using the deletion mutant Δ[lonBtrpAB4] (Fig. 1) which lacks trpA and the very end of trpB (11). A fingerprint of mRNA isolated from this deletion strain is shown in Fig. 2. Nine of the oligonucleotides correspond to ones on the wild-type fingerprint. The tenth, oligonucleotide 27f, has a similar composition and therefore a similar mobility to the wild-type oligonucleotide 29, but its sequence is different—27f is the "fusion" oligonucleotide containing the endpoint of the AB4 deletion in trpB. It is a derivative of oligonucleotide 27 (A-U-A-U-U-U-G), having the same 5' end (A-U-A-U-U-U-G) but a different 3' end (Table 1).

The total size of the purified wild-type RNA, as judged by summation of the T_r oligonucleotides and allowing for multiple copies of the smaller ones, is 200 to 250 nucleotides (Table 1). Although the purified RNA appears highly fragmented when run on polyacrylamide gels, the largest single fragment (220 ± 30 nucleotides) yields a fingerprint identical to that of the total purified product. Therefore, all oligonucleotides present on the fingerprint must have been covalently linked as a single RNA molecule. From these results we conclude that the purified RNA is exclusively trp operon messenger RNA.
The oligonucleotide sequences are numbered to correspond to their positions on the wild-type fingerprint (Fig. 2). The sequences were determined by standard techniques (14, 15) and may be considered exact except for 14, 28, and 30, in which the absolute sequences of the pyrimidine-rich stretches could not be determined. The "fusion" oligonucleotide 27f contains the "AB4" deletion-endpoint—inits' 3' sequence was not determined. Two sequence isomers were found in spots 8 and 22, and three in spot 18. The proposed alignment of each uniquely assigned oligonucleotide with the amino-acid sequence is given. "Aa-" means an overlap from amino-acid codon 5 to 7 of the trpA protein; the trpB protein is numbered backwards from its carboxy-terminus.

and contains translated sequences from both trpB and trpA.

**Alignment of the RNA Sequences with the Structural Genes.** We attempted to align the oligonucleotides by considering all possible amino-acid sequences they could encode (reading triplets in all three "phases"), and comparing those possibilities with the known amino-acid sequences of the trpB and trpA proteins (10, 16). Many of the larger oligonucleotides generated by T1 RNase (Table 1) and pancreatic RNase (not shown) had unique locations. Moreover, after those with unique locations were assigned, most of the remainder could also be assigned uniquely (Table 1) and aligned with the amino-acid sequences of the two structural gene products (Fig. 3).

**Ribosome-Binding and the Intracistronic Region.** The alignment of oligonucleotides with amino-acid sequence did not provide any unambiguous overlaps linking the trpB and trpA cistrons. However, ribosome protection of the initiation sequence of trpA should yield the nucleotide sequences immediately preceding and following the trpA AUG initiation codon. The methods by which a ribosome-protected fragment was obtained, and the fingerprints of this fragment, p1, are given in Fig. 4 and its legend. The T1 oligonucleotides of p1 are a subset of those in Fig. 2 (WT frame). Oligonucleotide 26, A-A-U-C-U-C-U-G, which occurs at the 3' end of the ribosome-protected fragment, is in low yield because these initiation complexes were treated with pancreatic RNase and some cleavage occurred within the... U-C-U-C-U... pyrimidine
FIG. 4. The trpA ribosome-binding site. The ribosome-protected fragment, p1, was fingerprinted after T1 (left) and pancreatic (right) RNase digestion. Some overlaps, such as G-G-A-A-C with A-A-C-G, may be inferred directly. Because less than 1000 cpm of sample was used for each of these fingerprints, running times in each dimension were halved to minimize diffusion and shorten film exposure times; the actual size of each fingerprint was 10 × 10 cm. Samples of trp mRNA used for ribosome-binding contained as carrier 100 μg of uncharged tRNA and 10 μg of bacteriophage R17 RNA. Initiation complexes were formed in 50 μl, containing 20 mM Hepes (NH4) pH 7.5, 70 mM NH4Cl, 4 mM Mg(OAc)2, 0.2 mM dithiothreitol, 0.2 mM GTP, 3 A260 units of charged, formylated (total) tRNA, and 12 A260 units of low-salt washed ribosomes. After 12 min at 37°, we added 50 μl of HG buffer (to increase the magnesium concentration and stabilize the formed initiation complexes), and pancreatic RNase to 5 μg/ml. After another 10 min at 20°, the sample was chilled and layered on a 5 ml 5–20% sucrose gradient (in 20 mM Hepes(NH4), pH 7.8, 100 mM NH4Cl, 12 mM Mg(OAc)2, and 0.1 mM dithiothreitol), and centrifuged in a Beckman SW50L rotor at 40,000 rpm for 130 min (4°). The 70S peak fractions contained 3–6% of the total radioactivity, indicating that ribosomes were binding with 20–40% efficiency. These fractions were pooled, made 0.4% in sodium dodecyl sulfate, and warmed to dissociate the ribosome complexes (5 min, 50°). Following phenol extraction, the RNA was ethanol-precipitated, washed, and dried for analysis.

than 15 nucleotides to the 3′ side of the AUG codon are virtually unprotected by the ribosome, whereas RNase cleavage to the 5′ side is dependent on nucleotide concentration and temperature.

These results indicate that the chain termination signal for the trpB protein is a single UGA codon which is out-of-phase with the trpA reading frame. The final nucleotide of the UGA termination codon is actually shared as the first nucleotide of the AUG initiation codon of trpA, thus the intercistronic region between the two structural genes consists of only two untranslated nucleotides.

DISCUSSION

We report here the nucleotide sequence of an intercistronic region within a polycistronic messenger RNA from E. coli. The striking feature of this "punctuation" region between the last two genes of the trp operon is its size: only two nucleotides separate translated trpB sequence from translated trpA sequence. The UGA codon terminating translation of trpB overlaps the AUG codon initiating translation of trpA (Fig. 5). If the signal for ribosome recognition and binding is more extensive than UGAUG, then the adjacent RNA sequences perform two functions: (i) they contain ribosome recognition signals and (ii) they specify amino-acid sequence.

There are other examples of RNA sequences with dual functions. At the beginning of the replicase cistron in bacteriophage R17 RNA there are sequences to which (i) ribosomes initiating translation of the replicase polypeptide bind, and (ii) the R17 coat protein also binds (18). The MS2 coat protein gene simultaneously: (i) specifies the amino-acid sequence of the coat protein, and (ii) contains base-pairing information that permits it to be folded with a high degree of secondary structure, in one place masking the synthetase initiation codon (19). One may conclude that regulatory (protein-binding) regions, sequences determining secondary structure, and punctuation information can overlap one another or polypeptide-specifying nucleotide sequences. This provides an obvious rationalization for the degeneracy of the genetic code, since it seems unlikely that such flexibility could be achieved if each amino acid were specified by one and only one codon.

The nucleotide sequence of the trpA initiator raises the recurring question of what the ribosome and its accompanying initiation factors recognize in binding to mRNA to initiate translation. Shine and Dalgarno have proposed that a polypurine sequence found in many of the coliphage ribosome-binding sites is involved in base-pairing with the 3′ end of 16S ribosomal RNA (2). The trpA ribosome-binding site contains a similar polypurine sequence (in the trpB segment) with two uncertain assignments, G->A-G-G->A-G-A-A-A (Fig. 5). If this sequence is actually G-A-G-G-A-G-A-A-A,
then the five underlined bases may pair with their complementary sequence in \( \text{OR} \text{A-U-U-C-C-U-C-C-A-C} \) of 16S ribosomal RNA (2). This particular hydrogen-bonded structure would be more stable than some of those proposed by Shine and Dalgarno, but we cannot assess the true stability until our sequence is determined exactly.

Another important question is whether ribosome recognition of the trpA initiation sequence is affected by translation of the preceding cistron. Morse et al. have shown that translation of trpA mRNA can occur even when trpB mRNA is not being translated (20); our current results confirm that sufficient information for ribosome binding is contained within the 220 nucleotide sequence of the trpB-trpA mRNA fragment. However, when trpB mRNA is being translated, what happens to the ribosome upon release of the completed trpB polypeptide? If ribosomes dissociate from the mRNA upon termination of translation of trpB, and if they are closely packed on the mRNA while translating, it is not clear how the trpA initiator region can become accessible to an unbound ribosome. One may, therefore, ask whether the cell has evolved mechanisms to bypass some of the "normal" steps in termination and initiation, in order to translate polycistronic messenger RNA more efficiently. Alternatively, if translating ribosomes are not closely packed on trp mRNA, then ribosomes initiating translation of trpA would not have to compete with ribosomes terminating translation of trpB for the same site on the mRNA. In this connection Zalkin et al. (21) have shown that ribosomes translating in vitro cannot traverse the entire trp operon if kasugamycin, an inhibitor of initiation, is present. The ribosomes must, therefore, pass through a kasugamycin-sensitive step to initiate translation at each cistron. Kasugamycin inhibition is thought to involve interaction with the 3' end of 16S ribosomal RNA (22); thus it seems likely that this region of the ribosome must become "exposed" before or during initiation. Yet colicin E3 cleaves the 16S RNA of ribosomes in the 70S configuration (23), so the kasugamycin results do not necessarily imply that the ribosomal subunits must dissociate before initiation of the trpA polypeptide.

It has been reported that translational "readthrough" of hisD in S. typhimurium due to a late frameshift mutation has a significant inhibitory effect on the expression of the subsequent gene, hisC (4, 24), probably because of interference with the initiation of hisC translation. If this interpretation is correct, one would predict that in E. coli suppression or mutational alteration of the trpB termination codon UGA would also affect expression of trpA. Moreover, in the case of the trpB-trpA junction, it is clear that a "+1" frameshift mutation in or near the intercistronic region could shift the phase of translation so that polypeptide chain termination would not occur. Instead, translation would continue into trpA (in-phase) to produce a fused trpB-trpA polypeptide. In fact, in Neurospora crassa the trpB and trpA functions do appear to reside in a single polypeptide chain (25, 26). A similar fusion mechanism might also have given rise to the trpD and trpC proteins in E. coli, which each now possesses two separate enzymatic functions in single polypeptide chains. Gene fusion of this type has been suggested by Bonner et al. (27) as being an important mechanism in the evolution of new proteins. It follows from our findings that simple mutations in or near the small intercistronic region could fuse trpB with trpA. We think it likely that the small intercistronic region itself simply reflects the cell's economy in selecting a means of facilitating the translation of a polycistronic messenger, though how it does so remains unknown.

**Note Added in Proof.** Nearest neighbor results obtained from trp mRNA synthesized in vitro (using individual (\( ^{3} \)H)-labeled nucleoside triphosphates) are consistent with the entire nucleotide sequence shown in Fig. 3. In particular, the sequence at the end of trpB appears to be -C-A-A-C-C-U-A-C-C-C. We thank Craig Squires and Joan Steitz for expert advice, and Frank Lee, Laurence Korn, and Terry Landers for critical reading of the manuscript. This work was supported by grants from the National Science Foundation, GM36967, and the U.S. Public Health Service, GM09738. T.F. is a postdoctoral fellow of the Helen Hay Whitney Foundation; C.Y. is a Career Investigator of the American Heart Association.