Proc. Nat. Acad. Sci. USA
Vol. 70, No. 12, Part I, pp. 3585–3589, December 1973

The Nucleotide Sequence of the Lactose Messenger Ribonucleic Acid Transcribed from the UV5 Promoter Mutant of *Escherichia coli*

(*in vitro* transcription/operator/dinucleotide priming/polymerase pausing/protein initiation sequence)

NANCY M. MAIZELS

Committee on Biophysics, Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Communicated by J. D. Watson, August 9, 1973

ABSTRACT I have sequenced the first 63 bases of mRNA transcribed *in vitro* from the UV5 promoter mutant of the *E. coli* lactose operon. Sonic fragments of DNA, 1000 base pairs long and purified to contain only the *lac* operator–promoter region, were used as template. The UV5 promoter mutation allows transcription of the *lac* operon in the absence of catabolite activator protein and cAMP; *lac* repressor controls the synthesis of this RNA. I find that during synthesis, RNA polymerase pauses at particular sites along the DNA, naturally generating several discrete sizes of RNA that provide overlaps useful for sequencing. The UV5 *lac* mRNA initiates within the *lac* operator and copies the operator sequence. The AUG initiator codon for β-galactosidase occurs at position 39 of the message. The sequence is:

pppA-A-U-U-G-U-G-A-G-C-G-G-A-U-A-A-C-A-A-U-U-U-C-A-C-A-C-A-G-G-A-A-A-C-A-G-C-U-A-U-G-A-C-C-A-U-G-A-U-U-A-C-G-G-A-U-U-C-A-C-U-G-G.

The initial sequence of a messenger RNA molecule from a bacterial operon should have signals for a variety of controls: signals unique to that operon as well as signals common to many. The message from the *Escherichia coli* lactose operon is particularly useful for studying such control signals, because there are available many mutations in the controlling region of the *lac* operon, both mutant promoters (binding or initiation sites for the RNA polymerase and CAP protein) and mutant operators (binding sites for the *lac* repressor). These mutations will appear as landmarks if they are transcribed, revealing what specific features of the *lac* operon appear in the mRNA. In this paper I report the sequence of the beginning of the messenger transcribed *in vitro* from the UV5 mutant promoter of the *lac* operon. The 5′ end of this *lac* messenger begins just inside the *lac* operator region, and includes a copy of the operator sequence.

Control of transcription of the *lac* operon

The two control sites of the *E. coli lac* operon, the promoter (*p*) and the operator (*o*), lie between the repressor gene (*i*) and the genes for β-galactosidase (*z*), lactose permease (*y*), and transacetylase (*a*). RNA polymerase recognizes the promoter and transcribes *z*, *y*, and *a* in the presence of cAMP, CAP protein, and an inducer of the *lac* operon. In the absence of inducer, the *lac* repressor (the product of gene *i*) interacts specifically with the operator region to block transcription. The natural inducer is allolactose, an isomer of lactose made by β-galactosidase (1). Mutations in the operator (o^c or operator constitutive) or in the repressor protein allow transcription of the *lac* genes in the absence of an inducer.

The *lac* operon responds to the general cellular control mediated by CAP and cAMP, as well as to the operon-specific control exerted by *lac* repressor. CAP and cAMP regulate all "catabolite repressible" operons of *E. coli*. Transcription of such operons occurs only in the absence of the preferred metabolite, glucose (2, 3). CAP stimulates transcription only when it has been activated by cAMP; in the presence of glucose, the cellular cAMP level drops, CAP is not activated, and a variety of operons for alternative energy sources cannot be transcribed (4–6). There are probably separate binding or recognition sites in the promoter for RNA polymerase and CAP protein, since one class of promoter mutations is known to abolish *lac* transcription, while another class allows transcription to occur independently of CAP and cAMP (7, 8). In my experiments I have used a mutation of the latter class, which was isolated as a lac^+ revertant of L8, a presumed point mutation in the CAP-binding site (9). In this double mutant, UV5, the *lac* genes are transcribed in the absence of CAP and cAMP both *in vivo* (7) and *in vitro* (10).

Lac transcription *in vitro* mimics transcription *in vivo:* it requires a functional promoter, CAP, and cAMP. *Lac* repressor inhibits transcription, and IPTG (isopropyl-β,D-thiogalactoside), a synthetic inducer of the operon, alleviates repression (11). Purified *lac* repressor binds tightly and specifically to purified DNA containing the *lac* operator region [$K_d = 10^{-13}$ M (12)], binds moderately to IPTG and other inducers [$K_m = 10^{-6}$ M (13)], and when bound to IPTG will not bind the *lac* operator region.

In vitro synthesis of UV5 *lac* mRNA

Sonic Fragments of lac *DNA.* In order to obtain pure *lac* mRNA for sequencing, I first isolate DNA fragments containing only a portion of the *lac* region and use them as template for transcription *in vitro*. Since the only promoter on such fragments is the *lac* promoter, they should in principle support synthesis of only one species of RNA–*lac* mRNA. To prepare these fragments, I sonicate whole *lac* transducing phage (either λ*h*80c I857 *t*68d*lac* or λ*h*80c I857 *S*7d*lac*UV5), purify the DNA, then use *lac* repressor to bind only fragments containing the *lac* operator to a nitrocellulose filter, and finally release those fragments with IPTG (14, 15). The sonic fragments are 1000 base-pairs long, or about 2% the length of whole phage DNA. Starting with 100 mg of *lac* transducing phage, I recover 0.2–0.4 mg of *lac* fragments, equal to about 0.5 nmol of *lac* operators. [The phage growth and other procedures are described by Gilbert and Maxam (16).] Most

Abbreviations: IPTG, isopropyl-β,D-thiogalactoside; CAP, catabolite activator protein.

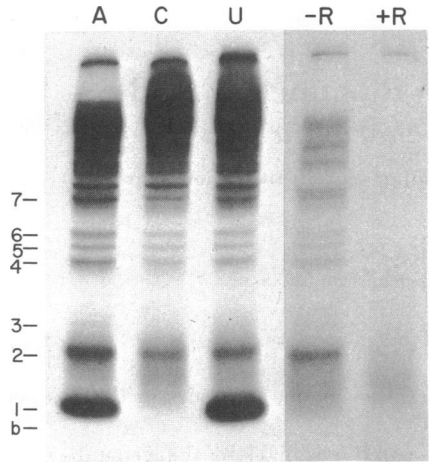

FIG. 1. The three columns on the *left* display the products of A-, C-, and U-labeled syntheses subjected to electrophoresis on a 12% polyacrylamide gel. The *numbers* indicate the gel band numbering used in the text; *b* indicates the position of a bromphenol blue-dye marker. The two columns on the *right* show the products of A-labeled reactions run with (+*R*) and without (−*R*) *lac* repressor. Discrete bands are absent in the repressed reaction. A T1 fingerprint of repressed material revealed no characteristic oligonucleotides, indicating that the diffuse background seen on gels is the product of nonspecific synthesis. Template DNA is isolated from *lac* transducing phage, purified from heat-induced lysogens as described by Gilbert and Maxam (16), except that these phages are defective and must be grown in a lysogen that also carries a nondefective helper phage. The denser *lac* phage are separated from helper on a CsCl equilibrium gradient in a Spinco angle 40 rotor, centrifuged for 20–24 hr at 22,000 rpm. Large quantities of λ*h*80c I857*t*68d*lac* phage are prepared from strain V5009 [RV(λ*h*80c I857*t*68, λ*h*80c I857*t*68d*lac*) $\phi 80^R$] from E. Signer. RV is *lac izya del* X74. Starting with λ*h*80c I857d*lac*UV5 phage, grown from strain CA8222 from R. Arditti, I constructed an *S7* derivative and made a double lysogen, NMM115 [RV(λ*h*80c I857*S7*, λ*h*80c I857*S7*d*lac*UV5) $\phi 80^R$], which was used for preparation of λ*h*80c I857*S7*d*lac*UV5 phage. In a typical reaction, 5 μg of *lac* operator–promoter fragments and 40 μg of RNA polymerase (purified as described in ref. 29, through the glycerol gradient) are preincubated for 6 min at 25° in a 100-μl mixture containing: 30 mM Tris·HCl (pH 8.0), 0.1 M KCl, 5 mM $MgCl_2$, 4 mM EDTA, 0.1 mM $PO_4^{=}$, 0.1 mM dithiothreitol, 10% glycerol, 0.05% bovine-serum albumin, 0.2 mM GpA (Sigma), and 5 μM triphosphates, all four or only one triphosphate labeled in the alpha position with ^{32}P (New England Nuclear Corp., specific activity >100 Ci/mmol). Absence of free divalent cations inhibits initiation of RNA synthesis until addition of 1/10 volume of 0.15 M $MgCl_2$ + 95 μg/ml of rifamycin; reactions done in the presence of repressor or CAP factor are preincubated with 15 mM $MgCl_2$ in the absence of triphosphates, and synthesis is initiated by adding triphosphates and rifamycin. The rifamycin reduces background synthesis very slightly, but since the polymerase/promoter ratio is low (0.4/1 assuming 10% active polymerase) and the reaction very quick, the background without rifamycin is not significant. After synthesis for 3–9 min at 25°, the reaction is quenched by addition of 1/10 volume of 0.8 M EDTA. After 20–40 μg of carrier tRNA (Schwarz) is added, ^{32}P-labeled RNA is separated from unincorporated label either by: (*i*) chromatography on a 2-ml Sephadex G-100 column in 50 mM NH_4HCO_3 (incorporated label, about 3–8% of the counts, elutes in the void volume and unincorporated label is retarded), followed by phenol extraction, ether extraction to remove the phenol, and concentration by lyophilization; or by (*ii*) increasing the reaction volume to 0.4 ml, extracting with phenol, transferring the aqueous phase to a silicated disposable tube containing 1/10 volume of 20% sodium acetate, precipitating with 2 volumes of 95% ethanol, pelleting the RNA for 30 min at 20,000 × *g*, and finally washing with 1 ml of 95% EtOH. These two ethanol precipitations remove over 99% of the unincorporated label, and are done rapidly by chilling the suspension at −70° for 15 min rather than for several hours at −20°. The RNA is boiled for 3 min in 7 M urea, quick chilled, and subjected to electrophoresis on a 20 × 20 × 0.2-cm 7 M urea–12% acrylamide slab gel in 90 mM Tris-borate (pH 8.3)–2.5 mM EDTA, for 2.5 hr at 24 mA.

of these fragments should contain the *lac* promoter as well as the operator, and will therefore serve as template for synthesis of *lac* mRNA.

I have shown by several criteria that transcription from these DNA fragments produces *lac* mRNA. (*i*) The transcript hybridizes asymmetrically to the correct strand of two different *lac* transducing phages: to the *l* strand of λp*lac*5 and to the *r* strand of $\phi 80$p*lac*. (These phages are described in ref. 17.) To be sure that hybridization assayed only for the *lac* region, I constructed the phage i^{80}hyλp*lac*5. This hybrid phage carries $\phi 80$ genes wherever the template has λ genes, and vice versa, and carries the *lac* region derived from λp*lac*5. RNA synthesized from either λ*h*80d*lac*UV5 or λ*h*80d*lac* fragments hybridizes asymmetrically to the *l* strand of i^{80}hyλp*lac*5. (*ii*) *Lac* transcription from fragments with the wild-type *lac* promoter requires CAP and cAMP, but transcription from those with the UV5 promoter mutation does not. (*iii*) *Lac* repressor represses transcription from the fragments. These results parallel those with intact DNA from *lac* transducing phage as template (9, 11, and my unpublished data).

Primed Initiation of Transcription. In order to synthesize RNA with low concentrations of triphosphates, I initiate transcription with a dinucleotide primer. RNA polymerase will not initiate at low triphosphate concentrations (5 μM); however, if a dinucleotide primer is added, the polymerase will use it to start chains at sequences within the initiation site that are complementary to the dinucleotide (18). Initiation with a dinucleotide affords two advantages. First, it allows synthesis at low triphosphate concentration, so labeled triphosphates are incorporated efficiently and the rate of elongation of RNA chains decreases, simplifying manipulations. Second, if there were several promoters present on the template DNA, a specific dinucleotide primer would select for initiation at one promoter rather than the others. For instance, specific dinucleotides selectively stimulate synthesis of specific T7 mRNAs (19). At low triphosphate concentration, only the dinucleotide GpA stimulates transcription from the UV5 *lac* operator–promoter fragments in the absence of CAP protein and cAMP. *Lac* repressor specifically represses this RNA synthesis.

The Polymerase Pauses. I attempted to synthesize short pieces of RNA representing the 5′ end of the *lac* messenger by starting and stopping synthesis synchronously. After preincubation of 10 pmol of UV5 *lac* operator–promoter fragments and 40 pmol of RNA polymerase in the reaction mixture without triphosphates described in the legend to Fig. 1, I started synthesis by adding triphosphates and rifamycin (20). [Alternatively, after preincubation in the absence of divalent cations, magnesium was added to initiate synthesis (21).] After 3–9 min, synthesis was stopped by addition of EDTA. Electrophoresis on a 12% acrylamide–7 M urea slab gel at pH

8.3 resolves the products of either synthesis into the pattern shown in Fig. 1. The striking aspect of such syntheses is that they are neither synchronous nor random, but rather the reaction produces a discrete set of RNA molecules, ranging in size from 7 to over 100 bases long. In the presence of *lac* repressor, about 25% as much RNA is synthesized, and this material electrophoreses as a smear, indicating that the bands correspond to products of repressible synthesis—that is, they are *lac* mRNA. The relative amounts of RNA in different bands varies with the length of the RNA synthesis reaction. After a chase with a high concentration of unlabeled triphosphates, all the material moves to a higher molecular weight. This finding implies that the bands correspond to RNA synthesized by polymerase molecules which are pausing when the reaction is terminated, rather than by polymerase molecules which have quit synthesizing before the end of the reaction. If these gel bands represent a set of RNA molecules initiating at the same site and pausing at specific regions along the DNA template, the reaction has naturally generated the overlaps and ordering necessary for sequencing.

Sequence of UV5 *lac* mRNA

I eluted RNA from the smallest gel band (gel band *1*), and determined its sequence to be G-A-A-U-U-G-U_{OH}. This sequence contains no C, and that is why the gel pattern of C-labeled RNA lacks this band (Fig. 1). The second band includes the material in gel band *1*, as well as additional information: G-A-A-U-U-G-U-G-A-G-C-G-G-A-U-A-A-C_{OH}. The sequences of the gel bands are shown in Table 1. Sequencing successively longer gel bands showed that the bands represent a nested set of RNA molecules, all initiating at the same site with GpA and pausing at different places along the DNA template.

Table 2 shows the oligonucleotide fragments produced by pancreatic and T1 digestion of each gel band. I was able to determine the sequence of all but one of the oligonucleotides from nearest-neighbor data and by digestion with RNases A, T1, or U2, and alkaline hydrolysis, without using partial digestion of fragments. The 17-base long T1 fragment from position *13* to *29*, "T12," posed the only special problem. Since it elutes poorly from DEAE–paper, I isolated it by gel electrophoresis of a T1 digest of total RNA synthesized in a short reaction. From the sequence of gel band *2*, I knew that the sequence of the 5′ end of T12 was A-U-A-A-Cp. Analysis of RNase A and U2 digestion products of T12 singly labeled with A, C, G, or U did not distinguish between the sequences A-U-A-A-C-A-C-A-A-U-U-U-C-A-C-A-G-(G) and A-U-A-A-C-A-A-U-U-U-C-A-C-A-C-A-G-(G). I determined the sequence to be the latter by analysis of the products of a partial venom digest (22) of an A-labeled fragment; and also by pancreatic digestion of a 24-base long minor gel band of A-labeled RNA (which should overlap the first 10 bases of T12): this contained the oligonucleotides G-A-A-Up, G-A-G-Cp, G-G-A-Up, A-A-Cp, and A-A-Up, but no A-Cp.

Table 2 lists the 3′-OH ends which I recovered and sequenced from some of the gel bands. The molar yields of 3′-OH ends relative to the other oligonucleotides is rarely greater than 0.5, implying that several pauses may occur at different sites in the same short region of DNA. In some cases I did not observe any fragment with a 3′-OH terminus, presumably

TABLE 1. *Sequences of paused transcripts from UV5 lac DNA*

Gel band	
1	GAAUUGU$_{OH}$
2	GAAUUGUGAGCGGAUA$^{A_{OH}}_{AC_{OH}}$
3	GAAUUGUGAGCGGAUAACAAU...
4	GAAUUGUGAGCGGAUAACAAUUUCACACAGGAAACAGCUA$_{OH}$
5	GAAUUGUGAGCGGAUAACAAUUUCACACAGGAAACAGCUAUGACC$^{A_{OH}}_{AU_{OH}}$
6	GAAUUGUGAGCGGAUAACAAUUUCACACAGGAAACAGCUAUGACCAUGAUU...
7	GAAUUGUGAGCGGAUAACAAUUUCACACAGGAAACAGCUAUGACCAUG(AUUACGG,AUUCACUGG)...

Individual bands were eluted from a 7 M urea–12% acrylamide slab gel (16) and sequenced by standard fingerprinting techniques described by Barrell (30) with modifications (16). For all two-dimensional fingerprints, the first dimension of electrophoresis was on cellulose-acetate strips in 7 M urea–0.2% pyridine–5% acetic acid–1 mM EDTA (pH 3.5) and the second dimension was on DEAE–paper in 7% formic acid. I attempted to order the last two long T1 fragments in band *7* [A-U-U-A-C-G-(G) and A-U-U-C-A-C-U-G-(G)] by a T1 fingerprint of material electrophoresing with slightly greater mobility than band *7*. That region included more A-U-U-A-C-G-(G) than A-U-U-C-A-C-U-G-(G), but the fingerprint was not clean enough to absolutely determine the order. These two T1 fragments are written in the order implied by that fingerprint, which agrees with the amino-acid sequence of β-galactosidase. Hyphens indicating phosphodiester bonds are omitted from oligonucleotide sequences in the tables.

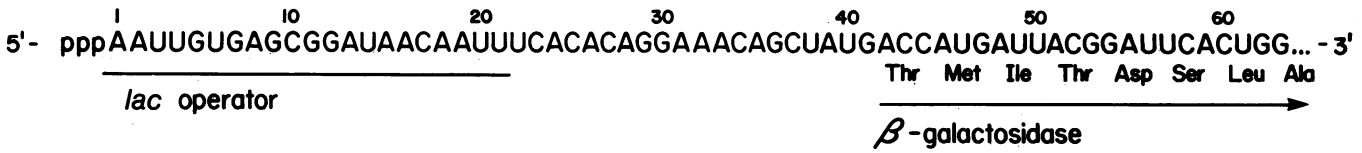

FIG. 2. The first 63 bases of the messenger RNA transcribed from the UV5 *lac* promoter. The *lac* operator sequence is copied into bases *1–21*. β- Galactosidase initiates at the AUG codon at position *39*, and the amino-terminal sequence of β-galactosidase (23) is matched to the triplets on the messenger which code for those amino acids. About 15% of the messenger initiates at the G before the first A, as pppG-A-A-U-U-G.

because of the low yield due to multiple pauses. No preferred sequence is evident at the 3'-OH ends of the pausing RNA.

The Natural Initiation Site. Does the GpA primer stimulate synthesis at a natural initiation site of the mRNA? I synthesized RNA in the absence of GpA and in the presence of a high concentration (200 μM) of ATP and GTP. This produces molecules with 5'-terminal triphosphates. Gel electrophoresis of this RNA shows that the lengths of these molecules are not perceptibly different from the lengths of molecules initiated with GpA, implying that under both conditions synthesis initiates at or near the same site. To identify the sequence at the 5' terminus of triphosphate-initiated mRNA, I did an indirect experiment, described in the legend to Table 3. Synthesis of RNA labeled with [α-^{32}P]-UTP was primed by a high concentration of ATP and GTP. I divided the product in half and treated one fraction with bacterial alkaline phosphatase before gel electrophoresis and RNase digestion. Oligonucleotides with triphosphate ends smear and disappear from my fingerprints; I could thus identify them as those spots absent from a fingerprint unless the RNA had been treated with alkaline phosphatase. In the T1 fingerprint of band *3*, an A-A-U-U-Gp spot appeared which was missing from the control with an intact triphosphate end. A new A-A-Up sequence appears in the pancreatic RNase digest of band *2*, and is absent from the control. A small amount of G-A-A-Up also appears on the pancreatic RNase fingerprint of the phosphatase-treated RNA; it is absent if the triphosphate end is left intact. Thus I infer that the UV5 *lac* mRNA usually starts pppA-A-U-U-G..., although about 15% of the molecules start pppG-A-A-U-U-G.... Reducing the concentration of either purine triphosphate to 5 μM while maintaining the other at 200 μM does not reduce the efficiency of initiation, indicating that in the absence of high ATP the polymerase initiates readily with GTP.

TABLE 2. *Oligonucleotides appearing in successively longer gel bands*

Band	Pancreatic Fragments	T1 Fragments
1	GAAU(U), U(G), GU_{OH}	AAUUG(U)
2	GU(G), GAGC(G), GGAU(A), AA_{OH} & AAC_{OH}	UG(A), AG(C), CG(G), G(A), $AUAA_{OH}$ & $AUAAC_{OH}$
3	AAC(A), AAU(U)	
4	U(U), U(C), C(A), 2AC(A), AGGAAAC(A), AGC(U)	AUAACAAUUUCACACAG(G), G(A), AAACAG(C), CUA_{OH}
5	U(A), AU(G), GAC(C)	CUAUG(A), $ACCA_{OH}$ & $ACCAU_{OH}$
6	C(A), AU(G), GAU(U)	ACCAUG(A)
7	U(A), AC(G), GGAU(U), U(C), C(A), AC(U), U(G) ...	AUUACG(G), G(A), AUUCACUG(G)

This table lists oligonucleotides found in pancreatic and T1 digests of successively longer gel bands. A numbered gel band contains all fragments listed above it in the table, (except for 3'OH ends) as well as those oligonucleotides listed at the band number. Fingerprints of RNA synthesized in reactions containing only a single labeled triphosphate yielded nearest-neighbor data. The 3' nearest-neighbor nucleotide is shown in *parentheses*.

TABLE 3. *Moles of oligonucleotides in digests of triphosphate-initiated RNA, with and without alkaline phosphatase treatment*

Pancreatic fingerprint, gel band 2

fragment:	GAAU(U)	AAU(U)	GU(G)	GGAU(A)
+ phosphatase	0.15	0.85	1.0	0.8
- phosphatase	--	--	1.0	0.8

T1 fingerprint, gel band 3

fragment:	AAUUG(U)	AUAACAAUUUCACACAG(G)
+ phosphatase	1.0	1.0
- phosphatase	0.15	1.0

I synthesized RNA in a reaction mix like that described for Fig. 1, except that GpA was omitted and included were 200 μM ATP + GTP, 5 μM CTP, and 5 μM [α-^{32}P]UTP. After synthesis for 1 min at 25° (the rate of elongation is faster at high triphosphate concentration), I split the RNA product into two portions, and treated one portion with bacterial alkaline phosphatase (Worthington), 4 μg/10 μg of carrier tRNA, in 50 mM NH_4HCO_3 (pH 8.8) for 1 hr at 37°. Addition of an equal volume of 90% glacial acetic acid + 10 mM EDTA followed by incubation for 30 min at 37° inactivated the phosphatase (31). After lyophilization, addition of 1/10 volume of 1 M Tris (pH 8.8) neutralized the phosphatase-treated sample before gel electrophoresis on a 7 M urea–12% acrylamide slab gel in parallel with the untreated sample. Parallel gel bands were eluted, digested with either RNase A or RNase T1, and fingerprinted.

The UV5 lac *mRNA codes for β-Galactosidase.* The amino-terminal sequence of β-galactosidase, product of the *lac z* gene, begins Thr-Met-Ile-Thr-Asp-Ser-Leu-Ala (Charlotte Hering in this laboratory; and ref. 23). As shown in Fig. 2, the first AUG initiator codon in the mRNA occurs at position *39*. The triplets following this AUG code for a protein of the sequence Thr-Met-Ile-Thr-Asp-Ser-Leu, the first seven amino acids of β-galactosidase.

Features Shared with Untranslated Regions of Other RNA Molecules. The UV5 *lac* mRNA presumably has a site for initiation of protein synthesis at the AUG at position *39*. Ribosomes will bind to initiator or initiator-like regions of RNA and single-stranded DNA and protect these regions from nuclease digestion (24–26); the sequences of several such ribosome-binding sites have been determined. Sequences common to the untranslated region of many of these ribosome binding sites (summarized in ref. 26) include: Pu-Pu-U-U-U-Pu-Pu; Py-A-G-G-A; and a nonsense codon within several bases of the AUG initiator codon. The UV5 *lac* mRNA includes the sequence A-A-U-U-U-C-A at positions *18–24*, which, with C substituted for a purine, resembles the common U-U-U sequence. The sequence C-A-G-G-A occurs at positions *27–31* of the UV5 *lac* mRNA. While no nonsense codons occur immediately before the initiator AUG, two nonsense codons occur out of phase after the initiator triplet, and one occurs in phase and one occurs out of phase in the untranslated region [at positions *6–8* (UGA), *14–16* (UAA), *40–42* (UGA), and *46–48* (UGA)]. In some ribosome-binding sites a stable hairpin loop could form around the initiator AUG, but such structures in the UV5 mRNA initiator region would not be stable.

The operator is transcribed

Gilbert and Maxam (16) have shown that the sequence of the region of the *lac* operator protected by *lac* repressor from DNase digestion is:

T-G-G-A-A-T-T-G-T-G-A-G-C-G-G-A-T-A-A-C-A-A-T-T
A-C-C-T-T-A-A-C-A-C-T-C-G-C-C-T-A-T-T-G-T-T-A-A

The UV5 *lac* mRNA includes 21 bases of the operator sequence. The operator sequence in the UV5 mRNA must be copied from the same operator involved in *lac* repressor binding, and not some identical sequence elsewhere on the DNA, because two spontaneous mutations to the o^c phenotype introduce single base changes in the operator region of the mRNA synthesized from the UV5o^c *lac* fragment template (27).

Symmetric regions in the operator sequence on the mRNA, A-A-U-U-G-U at positions *1–6* and A-C-A-A-U-U at positions *16–21*, can pair to form a marginally stable hairpin loop at the 5′ end of the messenger. This is the only stable configuration formed by hydrogen bonding between complementary base pairs in the UV5 mRNA. The mRNA does contain another group of eight pairs of symmetrically arranged bases, centered around position *26.5* and including bases *13–15*, *17*, *20–24*, and their reflections.

Since the UV5 mRNA begins within the operator region—the region of DNA protected by *lac* repressor from DNase digestion—the polymerase must initiate transcription at a site hidden by *lac* repressor in uninduced cells. This sequencing data supports the results of Eron and Block (10) and Chen *et al.* (28), who have suggested that polymerase competes with repressor for the same functional site on DNA containing CAP-independent promoters. However, Chen *et al.* claim that in the wild-type control region the polymerase initiation site does not overlap the repressor binding site.

Methods similar to those I have described for determining the sequence of the UV5 *lac* messenger should be useful for determining the sequence of other RNAs synthesized *in vitro*. The wild-type *lac* mRNA will be of particular interest.

I thank Allan Maxam for patient instruction in RNA sequencing, and Wally Gilbert for his unflagging interest and advice. I also thank Jim Files, Ron Ogata, Terry Platt, and Tom Ward for *lac* repressor, Charlotte Hering for some of the RNA polymerase, Ricardo Block for CAP, and Debbie Sutcliffe for preparing this manuscript. This work was supported by USPHS Grant GM 09541 from the National Institute of General Medical Sciences. I have been supported by NIGMS Biophysics Training Grant TOI GM 782-15.

1. Jobe, A. & Bourgeois, S. (1972) *J. Mol. Biol.* **69**, 397–408.
2. Emmer, M., de Crombrugghe, B., Pastan, I. & Perlman, R. L. (1970) *Proc. Nat. Acad. Sci. USA* **66**, 480–487.
3. Zubay, G., Schwartz, D. & Beckwith, J. (1970) *Proc. Nat. Acad. Sci. USA* **66**, 104–110.
4. Ullman, A. & Monod, J. (1968) *FEBS Lett.* **2**, 57–60.
5. Perlman, R. L. & Pastan, I. (1968) *J. Biol. Chem.* **243**, 5420–5427.
6. Makman, R. S. & Sutherland, F. W. (1965) *J. Biol. Chem.* **240**, 1309–1314.
7. Silverstone, A. E., Arditti, R. R. & Magasanik, B. (1970) *Proc. Nat. Acad. Sci. USA* **66**, 773–779.
8. Beckwith, J., Grodzicker, T. & Arditti, R. (1972) *J. Mol. Biol.* **69**, 155–160.
9. Arditti, R. R., Scaife, J. G. & Beckwith, J. R. (1968) *J. Mol. Biol.* **38**, 421–426.
10. Eron, L. & Block, R. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 1828–1832.
11. de Crombrugghe, B., Chen, B., Anderson, W. B., Nisiley, P., Gottesman, M., Pastan, I. & Perlman, R. L. (1971) *Nature New Biol.* **231**, 139–142.
12. Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970) *J. Mol. Biol.* **48**, 67–83.
13. Gilbert, W. & Müller-Hill, B. (1966) *Proc. Nat. Acad. Sci. USA* **56**, 1891–1898.
14. Bourgeois, S. & Riggs, A. D. (1970) *Biochem. Biophys. Res. Commun.* **38**, 348–354.
15. Gilbert, W. (1972) in *Polymerization in Biological Systems*, Ciba Foundation Symposium (ASP, Amsterdam), Vol. 7 (new series), pp. 245–256.
16. Gilbert, W. & Maxam, A. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3581–3585.
17. Shapiro, J., MacHattie, L., Eron, L., Ihler, G., Ippen, K. & Beckwith, J. (1969) *Nature* **224**, 768–774.
18. Downey, K. M., Jurmark, B. S. & So, A. G. (1971) *Biochemistry* **10**, 4970–4975.
19. Minkley, E. G. & Pribnow, D. (1973) *J. Mol. Biol.* **77**, 255–277.
20. Sippel, A. & Hartmann, G. (1968) *Biochim. Biophys. Acta* **157**, 218–219.
21. Blattner, F. R. & Dahlberg, J. E. (1972) *Nature New Biol.* **237**, 227–232.
22. Min Jou, W. & Fiers, W. (1969) *J. Mol. Biol.* **40**, 187–201.
23. Zabin, I. & Fowler, A. V. (1972) *J. Biol. Chem.* **247**, 5432–5435.
24. Steitz, J. A. (1969) *Nature* **224**, 957–964.
25. Hindley, J. & Staples, D. H. (1969) *Nature* **224**, 964–967.
26. Robertson, H. D., Barrell, B. G., Weith, H. L. & Donelson, J. E. (1973) *Nature New Biol.* **241**, 38–40.
27. Gilbert, W., Maizels, N. & Maxam, A. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38**, in press.
28. Chen, B., de Crombrugghe, B., Anderson, W. B., Gottesman, M. E. & Pastan, I. (1971) *Nature New Biol.* **233**, 67–70.
29. Berg, D., Barrett, K. & Chamberlin, M. (1971) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 21D, pp. 506–519.
30. Barrell, B. G. (1971) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper and Row, New York), Vol. 2, pp. 751–779.
31. Sneider, T. W. (1971) *Anal. Biochem.* **44**, 658–669.