Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4163–4167, October 1977
Biochemistry

5′-Terminal nucleotide sequence of *Escherichia coli* lactose repressor mRNA: Features of translational initiation and reinitiation sites

(*lac* operon regulatory gene/*in vitro* transcription/RNA sequencing/RNA secondary structure)

DEBORAH A. STEEGE*

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510

Communicated by Charles Yanofsky, June 17, 1977

ABSTRACT In a sequence of 214 nucleotides at the 5′ terminus of the *I* gene mRNA, which codes for the lactose repressor protein of *Escherichia coli*, (*i*) an untranslated leader sequence of 28 residues precedes the repressor coding region; (*ii*) a GUG initiates synthesis of the wild-type repressor; (*iii*) GUG and AUG are the functional initiators for the synthesis of restart polypeptides activated by early *I* gene amber mutations, confirming previous assignments for these residues based on protein sequencing data; and (*iv*) sequences complementary to 16S ribosomal RNA provide stronger potential mRNA·16S rRNA interaction at the wild-type initiation site than at the restart sites. When *I* mRNA is used to direct the formation of initiation complexes *in vitro*, ribosomes bind only to the wild-type initiator region.

A striking feature of the *I* mRNA sequence is the presence of a number of in-phase GUGs that have not been observed to serve as initiation signals *in vivo* in the nonsense mutant strains examined. The selective use of potential initiator triplets in the *I* mRNA leads to the following conclusions. First, when presented with several neighboring initiator triplets at the wild-type initiator region, ribosomes select the one preceded by the strongest appropriately positioned complementarity to the 16S 3′ end. Second, ribosomes do not restart after termination simply by moving to the next available initiator codon. Third, the formation of stable secondary structures predicted for the untranslated *I* mRNA beyond chain-terminating nonsense mutations may prevent ribosome access to some potential reinitiation sites.

The lactose repressor protein, encoded by the *lac I* gene, exerts negative control over the expression of the lactose operon of *Escherichia coli*. The possibility of using the *I* gene to obtain information concerning the specificity of translation initiation was raised several years ago by the discovery that nonsense mutations early in the gene activate reinitiation of repressor synthesis beyond the position of the chain terminating codon. At least three sites within that portion of the mRNA corresponding to the first 62 amino acids of the wild-type repressor protein direct the synthesis of COOH-terminal restart polypeptides which accumulate at 10% of the wild-type repressor level (1–3).

The amino acid sequences of the wild-type repressor (4, 5) and the restart polypeptides from amber mutant strains (1–3) predict which *I* mRNA codons are utilized as reinitiation signals. Whereas restart proteins are not detected in wild-type cells, two restart polypeptides are found in strains carrying amber mutations at amino acid positions 7, 12, and 17. The larger one initiates at a valine codon (presumed to be GUG) in position 23; the smaller one initiates at a methionine codon (AUG) in position 42. When the amber block is moved beyond the first restart site to position 26, only the polypeptide initiated at methionine-42 is made. Finally, an amber mutation at position 60 activates initiation of a third restart protein at leucine-62. From the fact that leucine-62 mutates to UAG with a frequency characteristic of a single base change, UUG was deduced to be the triplet for this reinitiation event (6).

The apparent use of GUG and UUG as well as AUG for reinitiation in the *I* gene was the first *in vivo* evidence supporting earlier *in vitro* observations that several codons in addition to AUG could stimulate *N*-formylmethionine-dependent initiation of protein synthesis (7, 8). This, plus the very frequent occurrence of reinitiation sites early in the *I* gene, prompted the suggestion that the ribosome may use every AUG, GUG, and UUG for reinitiation. On the other hand, the appearance of specific conserved sequences in the vicinity of initiator triplets at the beginnings of cistrons has led to the current notion that other elements may be as important to ribosome recognition as the presence of the start codon itself (9, 10). Therefore, it seemed important to establish whether additional features are also present in mRNA reinitiation signals.

Both to examine the *lac I* gene mRNA sequences that participate in the initiation of protein synthesis and to increase our understanding of the functioning of this regulatory gene, I determined the nucleotide sequence of the first 214 bases of the repressor mRNA (*I* mRNA). This report describes the *in vitro* synthesis of specific *I* mRNA fragments initiated at the I^Q promoter and the binding of these fragments to ribosomes *in vitro*. Features of the mRNA primary and secondary structures that relate to translational initiation are considered. A detailed description of the nucleotide sequence analysis and a discussion of other aspects of the mRNA will be presented elsewhere.

RESULTS

Sequence of the *lac I* mRNA. To obtain α-^{32}P-labeled *I* mRNA for nucleotide sequence analysis, I produced messenger fragments 100–200 nucleotides long by kinetically controlled transcription of a *lac* transducing phage DNA (λc_{I857} h80S$^-$t68d*lac* I^Q) and selectively purified the *I* mRNA by the hybridization scheme shown in Fig. 1. The synthesis conditions were developed based on the following rationale. The I^Q promoter was used because it increases *I* gene expression 10-fold over the low level observed in wild-type I^+ strains (11). Transcription from promoters on the $\lambda\phi$80d*lac* I^Q template other than those controlled by the λc_I repressor was achieved by selecting an ionic strength (80 mM KCl) that maximized RNA synthesis in the presence of a purified λ repressor preparation (data not shown). Finally, I reasoned that the relative repre-

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

Abbreviation: *I* mRNA, *lac* repressor mRNA.

* Present address: Department of Biochemistry, Duke University Medical Center, Durham, NC 27710.

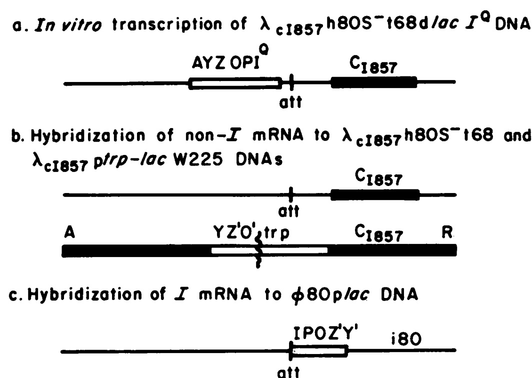

FIG. 1. Scheme used to obtain *lac I* mRNA fragments, showing genetic maps of bacteriophage strains used for *in vitro* transcription and hybridization. ——, ϕ80 genetic material; ■, λ genetic material; ▭, *E. coli* genetic material.

sentation of RNA from weak promoters might be increased if the RNA polymerase molecules could be synchronized beyond the stage of promoter recognition and DNA binding in the form of reaction intermediates initiated with the two triphosphates used for RNA starts. Accordingly, RNA polymerase was preincubated for 5 min at 37° with the template in the presence of Mg^{2+}, GTP, and ATP; the temperature was then shifted to 20°; CTP, UTP, and rifampicin were added, and the initiated RNA chains were extended to the desired length (Fig. 2). With the use of the GTP and ATP preincubation, *lac I* mRNA transcription became detectable.

Phage RNAs and any lac Z RNA were removed from the transcription product by successive hybridization to DNAs from the nontransducing parent phage $\lambda c_{I857}h80S^{-}t68$ and from λc_{I857}*ptrp-lac*W225, in which fusion of the *trp* and *lac* operons has deleted the *I* gene (see Fig. 1). *I* mRNA transcripts were finally recovered from the remaining nonhybridizable radioactivity by annealing to ϕ80p*lac* DNA. The material routinely released from the ϕ80p*lac* hybrids comprised only 3.5% of the total transcription product.

Standard nearest-neighbor phosphate transfer methods (12, 13) were used to determine the sequence of the first 214 bases of the *I* mRNA fragments purified in this way (Fig. 3). Oligonucleotides were ordered within the mRNA sequence by using information obtained from limited digestion with T1 RNase and from fingerprints of *I* mRNA fragments of various sizes produced by premature termination or "pausing" (14) of the RNA polymerase at discrete sites and were subsequently fractionated on a 7 M urea/12% acrylamide slab gel (14). In the case of the 5′-terminal 30 nucleotides, I used the ability of this region of the mRNA to bind to ribosomes *in vitro* in order to isolate a discrete fragment for sequence analysis (see below).

Nearest-neighbor data and overlap between T1 and pancreatic RNase digestion products leave a (GUG,AAU) ambiguity in the ribosome-protected fragment and several unordered residues in amino acid positions 53–55. In these cases, amino acid sequences of a *trp-I* gene fusion protein, in which the *I* mRNA leader sequence is actually translated in phase (15), and of the *lac* repressor protein, respectively, suggest that the nucleotide sequences are as written. The presence of several UCG and CUG trinucleotides in the vicinity of leucine-62 precluded unambiguous assignment of GUX(G) to valine-52 and UUG(C) to leucine-62. Otherwise, the independently determined *I* mRNA sequence corresponds exactly to the NH_2-terminal repressor protein sequence (4, 5); protein synthesis is initiated at a GUG codon succeeding a nontranslated leader sequence of 28 nucleotides (Fig. 3).

Ribosomes Bind to a Single Site *In Vitro*. Ribosomes recognize at least four regions within the first 214 bases of the *I* gene mRNA as initiation signals. Although active translation of the wild-type cistron may mask internal reinitiation signals, in *in vitro* initiation reactions, in which ribosomes remain bound at the position of the initiator codon, one might expect to detect recognition of the restart sites. If this were the case, then a substantial fraction of the RNA sequence should be protected against ribonuclease digestion. Contrary to this prediction, when unfractionated *I* mRNA preparations were used to direct the formation of initiation complexes *in vitro*, a fingerprint of the material recovered from the 70S ribosomes contained only a subset of the *I* mRNA digestion products. Oligonucleotides representing only the leader region and amino acid positions 1–5 predominated (Fig. 2, compare *b* with *a*); RNA sequences adjacent to the valine-23 and methionine-42 reinitiation codons (T1 RNase products 19, 20b, 22a, and 24, Table 1) were not observed. These results suggest that the restart

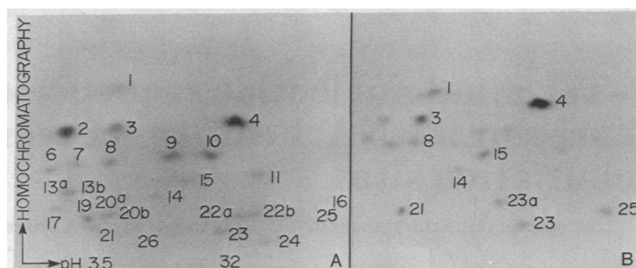

FIG. 2. T1 RNase fingerprints of the [α-^{32}P]GTP-labeled unfractionated *I* mRNA (*a*) and ribosome-protected fragment (*b*). The first dimension was electrophoresis on cellulose acetate (Cellogel) strips in 7 M urea/pyridinium acetate, pH 3.5; the second was homochromatography on 20 × 20 cm PEI-cellulose thin-layer plates (Brinkmann) in homochromatography mixture C of Barrell (13). Oligonucleotides are numbered to correspond with the sequences given in Table 1. Only the numbered oligonucleotides in (*b*) appeared reproducibly in good yield. *I* mRNA fragments labeled with 2–4 mCi of one α-^{32}P-labeled ribonucleoside triphosphate were produced by transcription of $\lambda c_{I857}h80S^{-}t68d$*lac* I^Q DNA in mixtures described by Musso *et al.* (29) with the following modifications. RNA polymerase (14 μg/ml) was preincubated with the template DNA (50 μg/ml) for 5 min at 37° in 20 mM Tris·HCl, pH 7.9/80 mM KCl/0.1 mM dithiothreitol/0.1 mM EDTA/10 mM $MgCl_2$/bovine serum albumin (0.5 mg/ml)/40 μM (each) ATP and GTP. At 40 sec after shifting the reaction to 20°, elongation of RNA chains was begun with the addition of 40 μM (each) UTP and CTP and rifampicin (10 μg/ml). Reactions were terminated after 50–75 sec by adding 50 mM EDTA and 100 μg of unfractionated *E. coli* tRNA. After phenol extraction, the RNA fragments were separated from unincorporated mononucleoside triphosphates by chromatography on a 1.1 × 30 cm column of Sephadex G-50 (fine) in 0.1 M NaCl/10 mM Tris·HCl, pH 7.5. After precipitation with 2 volumes of ethanol, the RNA was resuspended in 10 mM Tris·HCl, pH 7.8/10 mM $MgCl_2$/10 mM $CaCl_2$, treated with RNase-free DNase (Worthington) at 25 μg/ml for 15 min at 4°, phenol extracted, and reprecipitated with ethanol. The *I* gene transcripts were purified by the hybridization steps outlined in Fig. 1. $\lambda c_{I857}h80S^{-}t68$ and λc_{I857}p *trp-lac*W225 DNAs were used at 50- and 30-fold molar excess, respectively, over the template DNA. A 10-fold excess of ϕ80p*lac* DNA was used to recover the *I* mRNA fragments. After elution from the nitrocellulose filters and DNase treatment, the *I* mRNA was stored at −20° in distilled water. Previously reported hybridization procedures were used (30), except that RNase treatment of the hybrids was omitted. *I* mRNA samples were used to direct formation of translation initiation complexes *in vitro*, as described (31), under conditions that gave a ribosome/mRNA ratio of 5. The average efficiency of ribosome binding to the wild-type initiator region was 30%. In some experiments, including that shown in *b*, additional oligonucleotides (spots 6, 7, 9, 10, and 11) corresponding to *I* mRNA sequences for amino acid positions 8–14 appeared in the fingerprints of the ribosome-protected material. Whether this represents ribosome binding at an out-of-phase AUG codon in this region or heterogeneity in the positioning of the ribosomes at the wild-type initiation site is not known.

Table 1. *I* mRNA T1 oligonucleotides

1 G	16 UUUCUG(C)
2 CG	17 CCACG(U)
3 AG	19 AACCAG(G)
4 UG	20a AAAACG(C)
5 UUG(C)	20b AAAAAG(U)
6 CCG(G)	21 AAACCAG(U), CAAACAG(U)
7 CAG(A)	22a UUUCCCG(C)
8 AAG(A), AAG(C)	22b UUAUACG(A)
9 UCG, CUG	23 UCAAUUCAG(G)
10 AUG(U), AUG(G)	23a AAUUCAG(G)
11 UAUG(C)	24 UCUCUUAUCAG(A)
13a CCAG(C)	25 pGp
13b ACCG(U)	26 C(AC,AAC)AACUG(G)
14 UAACG(U)	32 AAUUACAUUCCCAACCG(C)
15 AAUG(U)	

The T1 RNase digestion products are numbered to correspond with the fingerprints shown in Fig. 2; 3′ nearest neighbors of unique products are enclosed in parentheses. Partial digestions with spleen phosphodiesterase resolved sequence ambiguities in T1 products 23 and 24. The presence of two isomeric sequences in spot 21 was indicated by pancreatic RNase and U2 RNase digestion data obtained from material labeled with [α-^{32}P]ATP and [α-^{32}P]CTP. A tetraphosphorylated moiety derived from the 5′ terminus was not detected; spot 25, however, contained pGp. Degradation of the 5′-terminal pppGp to pGp may have occurred during the hybridization reactions at 67°. When *I* mRNA preparations were used to direct the formation of translation initiation complexes *in vitro* (see *text*), the following subset of T1 RNase digestion products was obtained: G(G), AG(A), AG(U), UG(G), UG(A), AAG(A), UAACG(U), AAUG(U), AAACCAG(U), UCAAUUCAG(G), and pGp. An additional oligonucleotide, spot 23a, gave nearest-neighbor data identical to those for spot 23. It had the same mobility, however, as the spleen phosphodiesterase partial digestion product AAUUCAG(G) generated from oligonucleotide 23. Spot 23a presumably originated during the initial trimming of the ribosome-bound RNA with pancreatic ribonuclease.

sites, which are used less efficiently than the wild-type initiator *in vivo*, are not bound detectably *in vitro*.

However, one might argue that, because premature termination by RNA polymerase during the synthesis reaction would reduce the proportion of RNA from the two restart sites, preferential ribosome binding to the 5′ region was simply the consequence of its predominance in the mixture of transcripts. To determine whether the restart sites could be recognized when present in longer mRNA fragments, an [α-^{32}P]GTP-labeled *I* mRNA preparation was fractionated on a 7 M urea/12% acrylamide slab gel (14). The RNA fragments extracted (16) from several regions of the gel were used separately in initiation reactions. The T1 RNase fingerprints (not shown) of the ribosome-protected material from all samples were identical; even with RNA chains >200 nucleotides long, only the wild-type initiator region was observed. Thus, the ribosome-binding data support the conclusion that the initiator for the wild-type repressor protein is used at least 10-fold more efficiently than the first two restart sites *in vitro* as well as *in vivo* (1–3).

DISCUSSION

Several conclusions regarding initiation triplet usage and the specificity of initiation site selection in *I* mRNA can be drawn from inspection of the nucleotide sequence (Fig. 3). The wild-type repressor protein is initiated with GUG at nucleotide 29. This demonstrates that, at a normal initiation site in an *E. coli* mRNA, GUG can specify *N*-formylmethionine; previously, GUG was identified as the initiator triplet for the bacteriophage MS2 A protein (17). The *I* mRNA sequence also confirms the earlier assignment of GUG and AUG to amino acid positions 23 and 42, where they function as initiators for the first two restart polypeptides. Although my analysis could not verify the UUG assignment inferred from genetic data (6) for the restart codon at position 62, this is supported by direct DNA sequence data (P. J. Farabaugh, personal communication).

A most striking aspect of the *I* mRNA is the large number of initiation triplets that do occur in the repressor reading frame. There are 6 GUG triplets, 1 AUG, and 1 UUG among the first 214 nucleotides, In addition, the GUX at valine-52 has been shown to be GUG (P. J. Farabaugh, personal communication); of these, only four have been observed to serve for initiation or reinitiation in the amber mutants analyzed thus far. Although early *in vitro* experiments had suggested that CUG might function as an initiation triplet (7, 8), it is the UUG at position 62, rather than the CUG at position 56, that is read as *N*-formylmethionine *in vivo* in response to a nonsense mutation at position 54 or 55 (2, 18).

Thus, based on the amber mutants examined, clearly not every potential initiation triplet yields a restart polypeptide *in vivo*. Moreover, at the wild-type initiation site, where several in-phase initiator triplets are clustered together, only one appears to specify *N*-formylmethionine. Can such initiation site selectivity be explained in terms of our current understanding of mRNA·ribosome interaction?

Initiation and mRNA·rRNA Complementarity. Shine and Dalgarno (9) have suggested that base pairing between an mRNA sequence 5′ to the initiator codon and nucleotides at the 3′ end of the *E. coli* 16S rRNA occurs during the initiation step of protein biosynthesis. Regions complementary to 16S RNA centered at about 10 bases 5′ to the initiator triplet appear in the known ribosome-binding sites from bacteriophage and *E. coli* mRNAs (19). In two cases—for the bacteriophage R17 A protein initiator region (10) and for a bacteriophage λ P_R RNA ribosome binding site (16)—mRNA·rRNA complexes have been demonstrated experimentally.

From Fig. 3, it is clear that the mRNA·rRNA complementarity is most substantial at the initiation site for the wild-type repressor, consistent with its more efficient functioning relative to the reinitiation sites *in vivo* and its unique protection by ribosomes *in vitro*. By contrast, the sequences preceding the valine-23, methionine-42, and leucine-62 restart sites have only two, three, and three bases, respectively, complementary to the 3′-oligonucleotide of 16S RNA.

The extent of mRNA·rRNA complementarity appears to provide a plausible explanation for the selective use of one of several clustered potential initiators at the normal initiation site. Here there are three GUG codons, only one of which specifies *N*-formylmethionine *in vivo*. In addition, a mutant repressor carrying a threonine (ACG) to methionine (AUG) change at amino acid position 5 has been observed to initiate normally (3). Thus, of the potential initiator codons grouped in this region, the GUG that is preceded by a longer and more appropriately positioned mRNA·rRNA complementarity is used *in vivo*. This pattern has been noted for the selection of initiator codons in many other bacterial and phage ribosome-binding sites (19).

Reinitiation and *I* mRNA Secondary Structure. Another factor known to influence ribosome recognition of mRNA is the sequestering of potential initiator regions by stable RNA secondary and tertiary structures (20). Although active translation of a wild-type cistron may normally preclude the formation of hairpin loops and prevent reinitiation events as well, mRNA sequences distal to nonsense codons might be expected to assume thermodynamically preferred conformations. Hence, all possible *I* mRNA secondary structure interactions were tabulated and examined for their relevance to reinitiation.

The most stable double-stranded region that might be expected to form beyond the position of ribosome release at the

```
pppG.GAA.GAG.AGU.CAA.UUC.AGG.GUG(GUG.AAU)GUG.AAA.CCA.GUA.ACG.UUA.UAC.GAU.GUC.GCA.GAG.UAU
       8   3   3      23         4     4   15    4     21       14    22b      10   9   7   3   11

                                                met-lys-pro-val-thr-leu-tyr-asp-val-ala-glu-tyr
                                                 1                                       10

GCC.GGU.GUC.UCU.UAU.CAG.ACC.GUU.UCC.CGC.GUG.GUG.AAC.CAG.GCC.AGC.CAC.GUU.UCU
 6   4        24        13b       22a   2   4   4    19      13a  17   16

ala-gly-val-ser-tyr-gln-thr-val-ser-arg-val-val-asn-gln-ala-ser-his-val-ser
                             20          23 restart              30

GCG.AAA.ACG.CGG.GAA.AAA.GUG.GAA.GCG.GCG.AUG.GCG.GAG.CUG.AAU.UAC.AUU.CCC.AAC.CGC
 2   20a     2       20b    4     8   2   2   10   2   3   9          32          2

ala-lys-thr-arg-glu-lys-val-glu-ala-ala-met-ala-glu-leu-asn-tyr-ile-pro-asn-arg
                             40          42 restart                       50

GUX.GC(A.C,AA.C)AA.CUG.GCG.GGC.AAA.CAG.UCG.UUG *
          26              2        21    9   5

val—ala—gln—gln-leu-ala-gly-lys-gln-ser-leu
                             60          62 restart
```

FIG. 3. The 5′-terminal 214 residues of the *lac I* mRNA, aligned with the repressor amino acid sequence (4). The nucleotide sequence is arranged such that the initiator codons encoding *N*-formylmethionine for the three restart proteins are directly under the wild-type initiator sequence; the UUG assignment (*) for the leucine-62 restart codon is that deduced previously from genetic data (6). All in-phase AUG, GUG, and UUG codons appear in bold type. A dashed line indicates the ribosome-protected fragment. Contiguous bases 5′ to the initiator codons which are complementary to the 16S rRNA 3′ oligonucleotide (AUCACCUCCUUA$_{OH}$) are underlined, with potential U·G base pairs shown by dots. Centered under the appropriate sequences are the numbers corresponding to T1 RNase oligonucleotides (Table 1).

I gene nonsense codons in amino acids 7, 12, and 17 which activate reinitiation at valine-23 and methionine-42 is illustrated in Fig. 4 *left*. The hairpin loop is closed by a helical stem of 10 base pairs which has a calculated melting temperature of 97° (21, 22). Its formation would explain several features of *lac I* reinitiation in these amber mutant strains. First, the GGGA complementarity to rRNA that appears four nucleotides prior to the GUG at valine-38 is buried in the double-stranded stem; this could account for the previously observed lack of reinitiation at this site. Second, this structure might provide a plausible rationale for the curious observation that the valine-23 restart polypeptide always occurs together with the methionine-42 restart protein (3). Note that the valine-23 codon and the eight bases preceding it are base-paired. However, ribosome binding to the more accessible methionine-42 site should destabilize the structure, because a portion of the stem falls within the region normally protected from nuclease digestion by ribosomes. Hence, the valine-23 site would be available only as a result of a position 42 restart.

If we now consider nonsense mutations later in the *I* gene, we see that the most stable hairpin loop beyond ribosome release at a position 26 amber codon would be that shown in Fig. 4 *right*. The calculated melting temperature for this stem is 78° (21, 22). Again, the sequestering of nucleotides preceding the potential initiator at valine-38 is consistent with the exclusive use of the methionine-42 AUG in this repressor mutant. On the other hand, in strains carrying an ochre mutation at position 33 or 37, in which the presence of ribosomes translating sequences immediately adjacent to valine-38 might prevent *I* mRNA self-interaction here, use of the valine-38 GUG codon might be detected. Intriguingly, an unusually high level of restart material is found in such strains; these NH_2-terminal

FIG. 4. The most stable hydrogen-bonded regions postulated for *I* mRNA beyond the position of ribosome release at nonsense codons corresponding to repressor amino acid residues tyrosines 7, 12, and 17 (*Left*) and glutamine-26 (*Right*) (see *text*). Potential initiator codons in the repressor reading frame are boxed; numbers refer to the wild-type repressor protein sequence (4). The thermodynamic data of Crothers and associates (21, 22) give a calculated free energy of formation (ΔG) of −19.1 kcal/mol and a melting temperature (T_m) of 97° for the 10-base-pair helical stem (*Left*); a ΔG of −10.7 kcal/mol and a T_m of 78° are predicted for the 6-base-pair structure (*Right*). The calculated values are based on measurements obtained in 1 M NaCl; however, the T_ms are comparable to those observed in physiological buffers containing 10 mM Mg^{2+} (32).

protein sequences, however, have not been determined (J. Miller, personal communication).

The Problem of Translational Reinitiation. My analysis of the primary and secondary structures of the first 214 nucleotides of the *I* gene mRNA suggests plausible explanations for several features of the documented pattern of *lac* repressor restarts. However, with only these data in hand, certain reservations must be kept in mind, especially when evaluating the importance of possible mRNA·rRNA pairing in ribosome selection of reinitiation sites, which function at only 10% efficiency. Moreover, the statement that only three internal sites signal reinitiation is based on characterization of restart proteins from *I* gene nonsense mutant strains representing only 17 of the 20 available positions within the first 62 amino acids of the repressor cistron (1–3). The implications of the mRNA secondary structure compel study of additional mutants to determine if other in-phase GUG or possibly CUG codons function under some circumstances. Another unknown factor is the intactness of mRNA distal to polypeptide chain termination in the various mutant strains. It is possible, for example, that the unexplained absence of detectable restart material in strains with nonsense mutations at positions 2, 3, 5, and 6 (J. Miller, personal communication) is a result of rapid degradation of the message exposed to nucleases after ribosome dissociation.

On a more general level, it is clear that mRNA·rRNA interaction and the negative influence of mRNA secondary structure cannot entirely explain initiation specificity in *E. coli*. Ribosomal proteins and initiation factors make equally important contributions (for a discussion, see ref. 23); their role in reinitiation has not yet been explored. Furthermore, the evidence which suggests that ribosomes discharge and reattach at intercistronic termination and initiation sites when translating a polycistronic mRNA (24, 25) does not establish unequivocally whether a single ribosome has the capacity, upon termination, to reinitiate at a new site downstream on the same mRNA molecule. The occurrence of frequent reinitiation sites in the *lac I* mRNA originally prompted the hypothesis that the production of COOH-terminal restart polypeptides results from nonspecific ribosome binding to the next available initiator codon beyond the nonsense termination codon. Although the restart data cannot provide molecular information about ribosome-loading mechanisms at natural versus restart sites, the evidence that a pair of proteins is produced by some mutant strains and that not all in-phase GUG codons signal reinitiation argues for independent selection of initiator regions rather than simple movement to the next available AUG, GUG, or UUG.

Reinitiation after chain termination at nonsense codons has been shown to occur in the *E. coli lac Z* gene (26) and in the bacteriophage T4rIIb cistron (27) as well as in the *lac I* gene. In yeast, although reinitiation as it is seen in these cases has not been observed, Sherman and coworkers (28) have shown that, when the wild-type initiator codon for the iso-1 cytochrome *c* protein is altered by mutation, other AUGs mutationally induced in this region of the gene can function as initiation codons. The information that the study of *lac I* reinitiation has provided concerning how mRNA primary and secondary structures may function as determinants of ribosome initiation site recognition encourages one to anticipate that an examination of initiation mutants in yeast and other eukaryotic organisms may similarly shed light on the mechanisms by which eukaryotic ribosomes select initiator regions.

I thank J. Steitz, J. Miller, and T. Platt for their enthusiastic interest in this project and comments on the manuscript. This work was supported by National Institutes of Health Grant AI-10243 to J. Steitz and an American Cancer Society Postdoctoral Fellowship to D.A.S.

1. Platt, T., Weber, K., Ganem, D. & Miller, J. H. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 897–901.
2. Ganem, D., Miller, J. H., Files, J. G., Platt, T. & Weber, K. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3165–3169.
3. Files, J. G., Weber, K. & Miller, J. H. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 667–670.
4. Beyreuther, K., Adler, K., Geisler, N. & Klemm, A. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3576–3580.
5. Platt, T., Files, J. G. & Weber, K. (1973) *J. Biol. Chem.* **248**, 110–121.
6. Files, J. G., Weber, K., Coulondre, C. & Miller, J. H. (1975) *J. Mol. Biol.* **95**, 327–330.
7. Clark, B. F. C. & Marcker, K. A. (1966) *J. Mol. Biol.* **17**, 394–406.
8. Ghosh, H. P., Söll, D. & Khorana, H. G. (1967) *J. Mol. Biol.* **25**, 275–298.
9. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
10. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4734–4738.
11. Müller-Hill, B., Crapo, L. & Gilbert, W. (1968) *Proc. Natl. Acad. Sci. USA* **59**, 1259–1264.
12. Billeter, M. A., Dahlberg, J. E., Goodman, H. M., Hindley, J. & Weissman, C. (1969) *Cold Spring Harbor Symp. Quant. Biol.* **34**, 635–646.
13. 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.
14. Maizels, N. M. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3585–3589.
15. Sommer, H., Schmitz, A., Schmeissner, U., Miller, J. H. & Wittmann, H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, in press.
16. Steitz, J. A. & Steege, D. A. (1977) *J. Mol. Biol.*, in press.
17. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymakers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., Nolf, F. & Van Montagu, M. (1975) *Nature* **256**, 273–278.
18. Miller, J. H., Ganem, D., Lu, P. & Schmitz, A. (1977) *J. Mol. Biol.* **109**, 275–301.
19. Steitz, J. A., Sprague, K. U., Steege, D. A., Yuan, R. C., Laughrea, M., Moore, P. B. & Wahba, A. J. (1977) in *Nucleic Acid-Protein Recognition*, ed. Vogel, H. J. (Academic Press, New York), pp. 491–508.
20. Steitz, J. A. (1975) in *RNA Phages*, ed. Zinder, N. D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 319–352.
21. Gralla, J. & Crothers, D. M. (1973) *J. Mol. Biol.* **73**, 497–511.
22. DeLisi, C. & Crothers, D. M. (1971) *Biopolymers* **10**, 1809–1827.
23. Steitz, J. A., Wahba, A. J., Laughrea, M. & Moore, P. B. (1977) *Nucleic Acids Res.* **4**, 1–15.
24. Webster, R. E. & Zinder, N. D. (1969) *J. Mol. Biol.* **42**, 425–439.
25. Zalkin, H., Yanofsky, C. & Squires, C. L. (1974) *J. Biol. Chem.* **249**, 465–475.
26. Michels, C. A. & Zipser, D. (1969) *J. Mol. Biol.* **41**, 341–347.
27. Sarabhai, A. & Brenner, S. (1967) *J. Mol. Biol.* **27**, 145–167.
28. Sherman, F. & Stewart, J. (1975) in *Organization and Expression of the Eukaryotic Genome, Biochemical Mechanisms of Differentiation in Prokaryotes and Eukaryotes.*, eds. Bernardi, G. & Gros, F. (American Elsevier, New York), pp. 175–191.
29. Musso, R. E., deCrombrugghe, B., Pastan, I., Sklar, J., Yot, P. & Weissman, S. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4940–4944.
30. Kramer, R. A. & Rosenberg, M. (1976) *Nucleic Acids Res.* **3**, 24–26.
31. Steege, D. A. (1977) *J. Mol. Biol.*, in press.
32. Gralla, J., Steitz, J. A. & Crothers, D. M. (1974) *Nature* **248**, 204–207.