Nucleic Acids Research, Vol. 19, No. 13 3499–3506

Distinctive patterns of translational reinitiation in the *lac* repressor mRNA: bridging of long distances by out-of-frame translation and RNA secondary structure, effects of primary sequence

Rhonda J.Matteson+, Sorojini J.Biswas and Deborah A.Steege*
Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

Received May 17, 1991; Accepted May 29, 1991

ABSTRACT

In the early region of the *Escherichia coli lac* repressor mRNA, translational reinitiation events triggered by nonsense codons occur over long distances and in a distinctive pattern not explained by simple use of the next available initiator triplet. Defined fusions of the restart sites to the *lacZ* coding region have been used to explore the basis for these reinitiation patterns and to ask whether the sites can function in independent initiation at the 5′ end of an mRNA. The results obtained confirm earlier indications that the restart sites may have little or no inherent capacity for binding free 30S ribosomes. The data also add to growing evidence that primary sequence elements are important determinants of reinitiation efficiency. On the basis of the reinitiation activities for nonsense sites throughout the early region of the mRNA, we suggest that out-of-frame restarts and RNA secondary structure bridge long distances between the point of termination and downstream restart codons. Such bridging mechanisms could serve more generally as a means of propagating translational activity across long polycistronic mRNAs.

INTRODUCTION

Premature termination of translation resulting from nonsense mutations within a gene can lead to reinitiation of protein synthesis with N-formylmethionine by bacterial ribosomes at nearby AUG, GUG or UUG triplets not normally used for initiation. Translational reinitiation has been observed in a number of bacterial and phage genes, but has been studied most extensively in the rIIB gene of bacteriophage T4 (1,2) and the *lacI* gene of *Escherichia coli* (3–8). A number of fundamental questions have been addressed through studies of this process: to which codons N-formylmethionyl-tRNA responds, how ribosomes select specific sites for reinitiation, what factors determine reinitiation efficiency, and how the process compares mechanistically to initiation complex assembly at the 5′ end of a messenger RNA. Recently, it has become clear that certain translationally coupled initiation sites have the same properties as restart sites (9). Determining the rules for reinitiation will therefore provide information on the requirements for coupling in this class of initiation sites and more generally, on ribosome behavior at intercistronic junctions.

In most instances reinitiation is restricted to initiator triplets in the immediate vicinity of the chain-terminating nonsense mutation. However, beginning with the earliest work analyzing the N-terminal sequences of the *lac* repressor restart polypeptides (3–5) and the sequence of the *lacI* gene mRNA (6), reinitiation patterns in the early region of the *lac* repressor mRNA have appeared quite different. Here, reinitiation events occurred over long distances and in a distinctive pattern reflecting selective use of potential initiator triplets rather than simple movement to the next available AUG, GUG or UUG in the mRNA (Fig. 1). Nonsense mutations at repressor amino acid positions 2, 3, 5 and 6 did not result in detectable reinitiation, at least as assayed by production of restart polypeptides which negatively complemented repressor activity (J. Miller, unpublished observations). Mutations at positions 7, 12 or 17 led to the synthesis of not one, but a pair of restart polypeptides initiated at the in-frame GUG in the position of Val23 and at the AUG encoding Met42 (5). An amber mutation at position 26 gave only the restart protein initiated at the Met42 AUG (3), despite the fact that a GUG at the position of Val38 was the first in-frame initiator triplet downstream from the point of chain termination (6). Finally, an amber mutation at position 60 activated reinitiation at a nonstandard initiator triplet, the UUG codon at the position of Leu62 (4). The inactivity of the Val38 GUG in this group of nonsense mutants was explained by later studies, which showed that the Val38 GUG directs synthesis of a restart protein only if the position of the nonsense mutation is such that ribosomes on the mRNA prevent interfering secondary structure from forming (7). This result, along with direct evidence for structure from nuclease mapping experiments (7), indicated that local RNA secondary structure is a major factor dictating restart site selection in the rather GC-rich early region of the *lac* repressor mRNA.

* To whom correspondence should be addressed

+ Present address: Henderson Family Practice, Vance Medical Arts Building, Henderson, NC 27536, USA

Certain additional features of translational restarting in the *lac* repressor mRNA emerged when the nonsense mutations at positions 7, 26, 36 and 39 were placed in *lacI-Z* fusions so that β-galactosidase activity could be used as an assay for reinitiation at the *lacI* restarts (8). Distance was found to be a significant determinant of reinitiation efficiency: restarts at the Val38 GUG and Met42 AUG sites 3 or 15 nucleotides downstream from a nonsense codon were more efficient than restarts at the same sites 45 or 102 nucleotides downstream of a nonsense codon. Nevertheless, it still remained unclear why reinitiation occurred over such long distances in the early region of the *lacI* mRNA and generally does not in other systems (2). Evidence for reinitiation in the +1 reading frame of the *lacI* sequence was also obtained (repressor reading frame defined as the 0 frame). This raised the possibility that out-of-frame translation on the message between a nonsense codon and restart triplet could either block ribosome access to in-frame initiator triplets or, alternatively, could potentiate use of in-frame triplets by bridging the distance to these restarts.

Other results from these studies suggested that the *lacI* restart sites lack independent activity in primary initiation events requiring *de novo* assembly of components (7,8). Despite the fact that reinitiation efficiencies at the Val38 GUG and Met42 AUG restart sites approached half the efficiency with which the wild-type *lac* repressor protein is initiated, the restart sites were completely inactive in *lacZ* fusions when they were placed at the 5′ end of the *lacZ* mRNA. In *in vitro* ribosome binding assays, which also test for independent initiation, ribosome protection of the restart sites was barely detectable. Relative to binding at the true start site of the *lacI* gene, ribosome binding efficiencies for the restart sites were more than 100-fold lower. However, these sequences in the early region of the *lacI* mRNA assume stable structures that sequester the restart codons. Thus the alternative explanation for these results, that structure was the basis for the apparent lack of independent initiation activity, could not be ruled out.

Our present understanding of how distance, mRNA structure, the presence of out-of-frame initiator triplets, and primary sequence determine reinitiation patterns in the early region of the *lac* repressor mRNA is based on analysis of only 4 nonsense sites. A number of the questions about the basis for the observed reinitiation patterns and about the function of the restart sites in reinitiation and primary initiation remain unanswered.

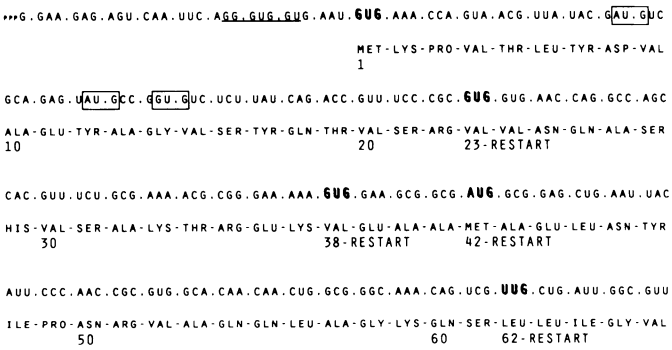

Figure 1. Sequence of the early region of the *lac* repressor mRNA and the corresponding protein sequence (6,10,11). In-frame codons used for initiation and reinitiation (5,7) are shown in large bold type; out-of-frame triplets implicated in reinitiation are boxed. The Shine-Dalgarno complementarity (12) for the wild-type initiator region is underlined.

Accordingly, we have extended our analysis of *lacI-Z* fusions to a number of additional nonsense sites spanning the early region of the *lac* repressor mRNA and have directly tested the +1 reading frame for its involvement in in-frame reinitiation events. In addition, the *lacI-Z* fusions in which the Val38 GUG and Met42 AUG restart sites are present by themselves have been subjected to a deletion analysis to ask if removing the sequences required to form RNA secondary structure unmasks independent initiation activity.

MATERIALS AND METHODS

Enzymes and biochemicals

Carrier-free (^{32}P)orthophosphoric acid and [α-^{35}S]dATP were purchased from New England Nuclear. Restriction enzymes were obtained from commercial sources and used according to suppliers' recommendations. Other enzyme sources were as follows: RNase A, Worthington; *Bal*31 nuclease (slow form), International Biotechnologies, Inc.; DNA polymerase I Klenow fragment, Bethesda Research Laboratories; Amplitaq DNA polymerase, United States Biochemical Corp.; and mung bean nuclease, Pharmacia LKB Biotechnology. Materials for DNA sequence analysis by dideoxy methods (13) were from U.S. Biochemicals. Cephaloridine was purchased from Sigma.

Bacterial strains

The *lacI*$^-$ nonsense mutations used in this study were obtained as derivatives of the strain GM1, which carries the *lacI* gene on the episome F′ *pro lac*, I^Q, *L8* (14). Each mutation was transferred from the episome to plasmid pMC7, a pMB9 derivative bearing the $lacI^Q$ gene (15). The F′ was first transferred to a Rec$^+$ recipient harboring pMC7 to allow recombination between the episome and the plasmid population. Plasmids isolated from this strain were then introduced by transformation (16) into a LacI$^-$Z$^+$ recipient, and those carrying the *lacI*$^-$ nonsense mutation were detected on Lac indicator medium. Plasmids in the resulting collection were maintained in the *lac* deletion strain DS70, F$^-$ Δ(*lac*) *trpR* λ^s (17). The protease-deficient strain BNN103, Δ(*lacIPOZYA*)*U169* *proA*$^+$ Δlon araD139 strA thi hflA150[chr::Tn*10*] (18) was obtained from V. Burdett, Duke University.

Plasmids and plasmid constructions

Standard techniques for DNA preparation and cloning were used in plasmid constructions and analysis (19). Each construct was subjected to DNA sequence analysis to verify the plasmid-insert junctions and the position of the nonsense mutation (13,20). The two *lacZ* fusion vectors were obtained from M. Casadaban, University of Chicago. Plasmid pMC1403 contains the *lacZ* coding region without an upstream promoter or translational initiator region; pMC1513 contains the complete *lacZ* gene with the *lacUV5* promoter, but a linker places the translation initiation site and first 8 codons for β-galactosidase out-of-frame with the remainder of the gene (21).

Fusions designed to measure reinitiation efficiency were generated from pMC1403 by cloning fragments from the *lacI* gene nonsense mutant collection harbored as derivatives of plasmid pMC7 into the unique *Sma*I site of pMC1403 just upstream of the *lacZ* coding region. The standard fragment from the *lacI* gene was a 291-bp *Hin*cII-*Sau*96 fragment which includes the $lacI^Q$ promoter and the early region of the gene. When the staggered ends of this fragment are repaired using the Klenow

fragment of DNA polymerase I, the repressor reading frame meshes with that for β-galactosidase. Slightly shorter fragments (*Hin*cII-*Hae*III, 156 bp) were used to generate the fusions bearing UAG7 and UAG17 which lack sequences downstream of the Val23 restart site. In-frame alignment of these fragments was achieved by ligation to vector DNA linearized by *Xma*I digestion and repaired to give blunt ends.

Constructs made to test the requirement for translation in the +1 reading frame of the *lacI* sequence were generated from plasmid UAG7 as follows. The region of UAG7 between codon 5 of *lacI* and the junction with downstream *lacZ* sequences was copied using PCR. The desired mutations in the +1 frame AUG and GUG triplets to ACG and GCG, respectively, were generated at the same time by specifying base substitutions at the appropriate positions of the upstream PCR primer. Primers were designed to give products that, when digested with restriction enzymes, were appropriate for directed cloning between the two segments of the *lacZ* gene in pMC1513. Cleavage at the 3′ end of PCR products with *Bam*H1 permitted 0 frame alignment of *lacI* with the downstream segment of the *lacZ* coding region. Cleavage of 5′ ends with *Sca*I generated a blunt end that could be linked in either the *lacI* 0 or +1 frame to the beginning of the *lacZ* gene by appropriate cleavage and repair of pMC1513 sequences.

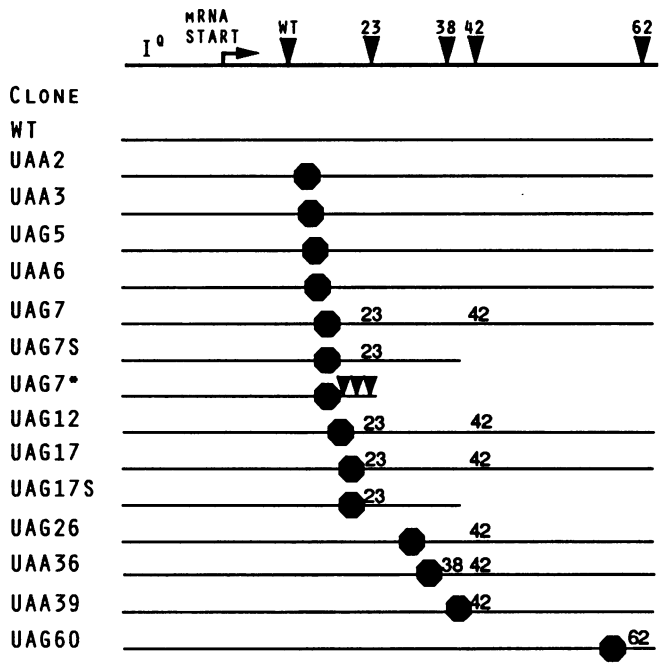

Figure 2. Schematic diagrams of the *lacI-Z* fusion plasmids used to measure translational reinitiation efficiency. The upper line represents the region encompassing a *Hin*cII-*Sau*96 DNA fragment, which contains the I^Q promoter, the start point for *lacI* transcription, the initiator region for the wild-type repressor (WT), and the four in-frame *lacI* restart sites (23, Val23 GUG; 38, Val38 GUG; 42, Met42 AUG; and 62, Leu62 UUG) within the early region. This fragment or the smaller subfragments indicated were cloned as described in Materials and Methods into plasmid pMC1403 (21). The resulting fusions are designated by the nonsense mutation present, with the suffix S denoting clones in which less of the *lacI* gene sequence is present downstream of the restart codon. The wild-type fusion, UAG7, UAG7*, UAG26, UAA36, and UAA39 have been described previously (8). In the plasmid designated UAG7*, the +1 frame of the *lacI* gene sequence is fused to the *lacZ* coding region. The position of the chain-terminating nonsense mutation in each plasmid is denoted by an octagon. The points at which translation reinitiates, based on previous analyses of the restart polypeptides produced in *lacI*⁻ nonsense mutant strains (3–5,7), are indicated by numbers. Arrowheads indicate the AUG and GUG triplets in the +1 frame.

The reference plasmid for this series of constructs was pMC1513-IF, in which the two segments of the *lacZ* coding region in pMC1513 have been linked in-frame (9).

The parent plasmid used to test the *lacI* restart sites for primary initiation was pKC8 (8). It contains the $lacI^Q$ promoter on a 68-bp *Hin*cII-*Hin*fI fragment, which was inserted with *Eco*RI linkers into the *Eco*RI site of pMC1403. The reference plasmid pKCWT contains the translation initiation site from the beginning of the wild-type *lacI* gene inserted into the *Sma*I site of pKC8 as a 59-bp *Hin*fI-*Msp*I fragment (8). Derivatives of pKC8 which contain either the Val38 site alone (44-bp fragment) or the Val38 and Met42 sites together (56-bp fragment) in the *Sma*I site (8) will be referred to throughout as pKC38 and pKC38/42. Deletions directed 5′→3′ into the *lacI* sequences in these plasmids were made as follows. The plasmids were linearized upstream of the *lacI* sequence at the *Eco*RI site, and then treated with the slow form of *Bal*31 nuclease (0.25U/μg DNA, 30°C, 2–45 min). Following repair of the resulting ends, the DNA was digested with *Bam*HI to excise the fragment of interest. The fragments were separated from the plasmid by polyacrylamide gel electrophoresis and then directed into pKC8 by ligation to vector DNA digested with *Sma*I and *Bam*HI.

Enzyme assays

β-Galactosidase levels were determined as described by Miller (22), using bacteria cultured in LB broth containing 0.5% glycerol. Relative plasmid copy number in plasmid-bearing strains was assessed by β-lactamase levels. Mid-log phase cultures were sonicated, and β-lactamase levels in the crude extracts were determined by measuring the rate of decrease in absorbance at 255 nm of a 0.1 mM cephaloridine solution in 0.1 M sodium phosphate, pH 7.0 (23).

RESULTS

Reinitiation activity in the *lacI* gene early region

In previous studies (8) we found that defined fusions of the *lacI* gene early region to the *lacZ* coding region present in plasmid pMC1403 provided a sensitive assay for the reinitiation events in *lacI* sequences. With the $lacI^Q$ promoter driving transcription, β-galactosidase activities ranging from ~50–6000 units were obtained. β-Galactosidase activity appeared to reflect translational efficiency, because differences in plasmid copy number were not observed, and the more than 100-fold range in β-galactosidase values were not accounted for by variations in *lacI*-specific mRNA levels. As will be shown in this work, differences in the susceptibility of the fusion proteins to proteolysis are also not evident.

Fig. 2 diagrams the segments of the *lacI* gene present in the *lacI-Z* fusions constructed for this and the previous study (8). Named by the nonsense mutation present, each *lacI* DNA fragment contains the $lacI^Q$ promoter, a nonsense mutation, and one or more restart sites. A plasmid containing the same fragment from a wild-type $lacI^Q$ gene serves as the reference (WT). In each fusion, translation begins at the initiation site for the wild-type repressor protein and proceeds to the position of the chain-terminating nonsense mutation, indicated by an octagonal symbol. The points at which translation is known to reinitiate in the *lac* repressor mRNA (3–5,7) are indicated by numbers corresponding to repressor amino acid positions. The series includes a number of nonsense mutations not previously assayed in *lac* fusions. The ochre and amber mutations at positions 2,

3, 5 and 6 are the ones which, based on assays of negatively complementing repressor activity and isopropyl-thiogalactoside-binding material, did not appear to trigger reinitiation downstream (J. Miller, unpublished observations). The amber mutation at position 60 activates the restart at the Leu62 UUG triplet (4). The series also includes two plasmids in which a shorter fragment was used to omit the Met42 restart site (UAG7S and UAG17S) and one in which the β-galactosidase reading frame is aligned with the +1 reading frame of the *lacI* DNA sequence (UAG7*).

Table 1 lists the β-galactosidase levels measured for the *lacI-Z* fusions harbored in two *lac* deletion hosts, DS70 (*lon*$^+$) and BNN103 (*lon*$^-$). For the wild-type construct, the number is a measure of initiation at the beginning of the *lac* repressor coding region, and as such it indicates the level of translation in the repressor reading frame occurring upstream of each restart. For constructs bearing the nonsense codons indicated, the activities represent reinitiation efficiencies. β-Galactosidase activities are quite similar in the two strains, with no increases in the levels observed in the *lon*$^-$ hosts. This suggests that the restart proteins produced from the *lacI-Z* fusions do not show marked differences in stability that would give aberrant estimates of reinitiation efficiency.

Table 1. Translational reinitiation efficiencies in the *lacI* gene early region

Plasmid	DS70 (*lon*$^+$) β-Galactosidase Units	%	BNN103 (*lon*$^-$) β-Galacctosidase Units	%
WT	4830	100	4080	100
UAA2	5	0.1	7	0.2
UAA3	15	0.3	11	0.2
UAG5	5	0.1	4	0.1
UAA6	4	0.1	4	0.1
UAG7	69	1.4	51	1.2
UAG7S	22	0.4	26	0.7
UAG7*	728	15	478	11
UAG12	93	1.9	84	2.0
UAG17	116	2.4	85	2.0
UAG17S	20	0.4	10	0.2
UAG26	157	3.2	99	2.2
UAA36	1500	32	1060	26
UAA39	1940	40	1230	29
UAG60	171	3.5	108	2.5

Cultures of DS70 (*lon*$^+$) and BNN103 (*lon*$^-$) derivatives containing the indicated plasmids were assayed for β-galactosidase activity. Plasmid designations are as in Fig. 2. In this and subsequent tables, β-galactosidase units are expressed as *o*-nitrophenol absorbance units at A_{420} per min per A_{600} unit of cells. Here the background values for the strain lacking a plasmid have been subtracted. For DS70, these values averaged 25 units, and for BNN103, 2 units. Enzyme levels are also expressed relative to the level resulting from initiation at the beginning of the *lacI* gene in the wild-type construct (WT). All values are the average of at least 4 independent determinations.

The data reveal that even with a more sensitive assay, the nonsense mutations at positions 2, 3, 5 and 6 do not appear to activate detectable reinitiation in the repressor reading frame. β-Galactosidase levels for the constructs bearing these mutations are 0.1−0.3% of primary initiation at the beginning of the *lacI* gene. When the position of the UAG is moved one codon farther to position 7, there is a sharp increase to 1.4% in in-frame reinitiation, which previous data indicate occurs at the Val23 GUG and Met42 AUG (5). This level increases modestly to 1.9% for UAG12, and to 2.4% for UAG17. The values for UAG7S and UAG17S, in which the Met42 AUG restart site is not present, suggest that activity from the in-frame restart at the Val23 GUG is quite low. By far the highest levels of in-frame reinitiation are in response to UAA36, which has been shown to give a mixture of the Val38 and Met42 restart polypeptides (7), and UAA39, which gives the Met42 restart polypeptide (7). In contrast, reinitiation at the Leu62 UUG in response to UAG60 is much lower, 3.5%, despite a spacing of only 3 nucleotides between the termination and restart codons. Finally, note that the level of reinitiation measured for UAG7* in the +1 reading frame of the *lacI* sequence is 10-fold higher than the level of reinitiation in the repressor reading frame (UAG7), indicating that out-of-frame translational activity is a major event occurring on this segment of the mRNA.

Translation in the +1 frame is required for the in-frame restarts from UAG7

The substantial level of reinitiation activity observed for UAG7* focused attention on the +1 frame of the *lacI* sequence and its potential involvement in the in-frame reinitiation events that produce the Val23 and Met42 restart polypeptides in response to the UAG7 nonsense mutation. Downstream of UAG7, there are in the +1 frame two AUGs and one GUG between repressor amino acid positions 8 and 15. These define a short open reading frame that terminates at a UGA immediately beyond the in-frame GUG restart codon at Val23 (Fig. 1). To facilitate analysis of this region, the polymerase chain reaction was exploited to selectively copy the segment of plasmid UAG7 between repressor

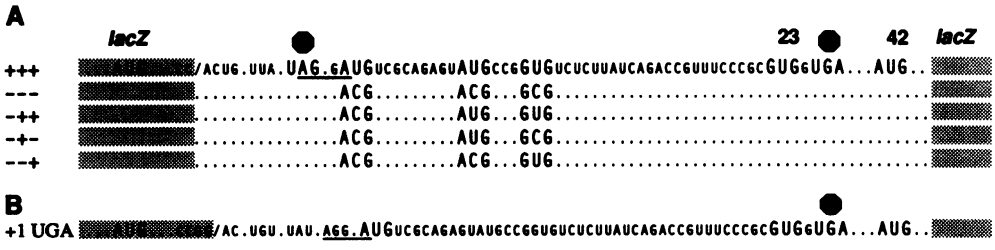

Figure 3. Diagram of the RNA sequences from *lacI-Z* fusions constructed to examine translation and reinitiation involving the +1 frame between UAG7 and the Val23 GUG. (A) requirement for reinitiation in the +1 frame for subsequent reinitiation at in-frame *lacI* restarts. (B) direct linkage of the *lacI* gene +1 frame to the source of primary initiation as a direct assay of reinitiation in response to the +1 frame UGA stop codon. The plasmids are derivatives of pMC1513 (stippled bar) in which primary initiation begins at the *lacZ* AUG and proceeds to UAG7 (A) or to the +1 frame UGA (B). The 0 frame fusion of the *lacI* gene early region to the *lacZ* coding region is as in the pMC1403 *lacI-Z* fusions of Fig. 2. This region of the *lacI* sequence was copied from plasmid UAG7 using PCR. Mutations were introduced into the AUG and GUG triplets in the +1 frame by making the appropriate base substitutions in upstream DNA primers. The first two nucleotides 3′ to the upstream junction with the pMC1513 vector sequence (/) are substitutions made in the *lacI* sequence to generate a *ScaI* restriction site. Octagons indicate the positions of the UAG7 nonsense codon and the +1 frame UGA. The Shine-Dalgarno sequence generated by the UAG7 mutation is underlined.

amino acid position 5 and the 3′ *lacI-Z* junction. By placing the region of interest at the 5′ end of the DNA to be amplified, mutations in the +1 frame AUG and GUG triplets could then be made simply by specifying the appropriate base substitutions in the upstream PCR primers. As shown in Fig. 3, the PCR products were cloned in the linker between the two segments of the *lacZ* gene in pMC1513 to generate a 0 frame fusion of the *lacI* gene early region to *lacZ* which is equivalent to that present in the pMC1403 *lacI-Z* series. In the pMC1513 derivatives, primary initiation begins at the *lacZ* AUG and proceeds into the *lacI* sequence in either the repressor reading frame (A) or the +1 frame (B). These plasmids are designated by the status of the three potential initiator codons in the +1 frame, which are indicated wild-type (+) or mutant (−).

Table 2 reports the reinitiation activities in the repressor reading frame for the pMC1513 derivatives of Fig. 3. For this series the reference plasmid measuring the level of primary initiation upstream of the restarts is pMC1513-IF, in which the two segments of the *lacZ* gene have been linked in-frame (9). The value for the parent vector pMC1513 indicates the level of background activity. The activities for the two plasmids with no base substitutions introduced into the sequence (+++ and +1 UGA) provide reassurance that in the pMC1513 context,

Table 2. Translation in the +1 frame as a requirement for in-frame reinitiation

Plasmid	β-Galactosidase (units)	Reinitiation efficiency (%)
pMC1513-IF	13,400	100
pMC1513	16	0.12
+++	213	1.6
−−−	21	0.16
−++	175	1.3
−+−	61	0.46
−−+	39	0.29
+1 UGA	276	2.1

Cultures of DS70 containing the indicated plasmids were assayed for β-galactosidase activity. Plasmid designations are as in Fig. 3. Reinitiation efficiencies are expressed relative to the level of upstream translation defined by pMC1513-IF. All values are the average of at least 4 independent determinations.

reinitiation efficiencies relative to the upstream translation level are similar to those measured in the pMC1403 series. In-frame reinitiation in response to UAG7 (+++) occurs at 1.6% efficiency, a value very close to the 1.4% reported for plasmid UAG7 in Table 1. The value for plasmid +1 UGA, in which the +1 frame of *lacI* has been linked directly to upstream *lacZ* in order to quantitate in-frame reinitiation in response to the +1 frame UGA stop codon, is 2.1%, again in the same range as in-frame reinitiation in response to UAG17 and UAG26.

The data for the mutants reveal that translation in the +1 frame of the *lacI* sequence is required for subsequent in-frame reinitiation events downstream. If the AUG and GUG codons are all mutated to ACG and GCG, respectively, in-frame reinitiation is abolished. The first AUG is spaced only 1 nucleotide 3′ to UAG7 and might have been expected to serve as the major restart codon. Changing this AUG to ACG results in a reproducible decrease in in-frame reinitiation, but only by 20%. Of the two remaining initiator codons in the +1 frame, it is the AUG that when present as the only available initiator triplet gives substantial levels of in-frame reinitiation above background. Note that this second AUG is the most appropriately spaced of the three triplets from a 4-base Shine-Dalgarno sequence AGGA, which was generated by the same base change that produced the UAG7 nonsense mutation. As judged from the information obtained here, it is likely that the two AUGs contribute most of the reinitiation activity occurring in the +1 frame of the *lacI* sequence.

Do the *lacI* restart sites function as independent ribosome binding sites?

When the Val38 GUG and Met42 AUG sites were tested earlier in *lac* fusions for their activity in independent initiation, β-galactosidase activities were at background levels (8). These fusions were under control of the $lacI^Q$ promoter, and they contained small fragments bearing the two most efficient restart sites: the Val38 GUG restart site alone (44 bp, pKC38) and the Val38 GUG and Met42 AUG sites together (56 bp, pKC38/42). The negligible β-galactosidase values obtained for both constructions containing *lacI* restart sequences suggested that

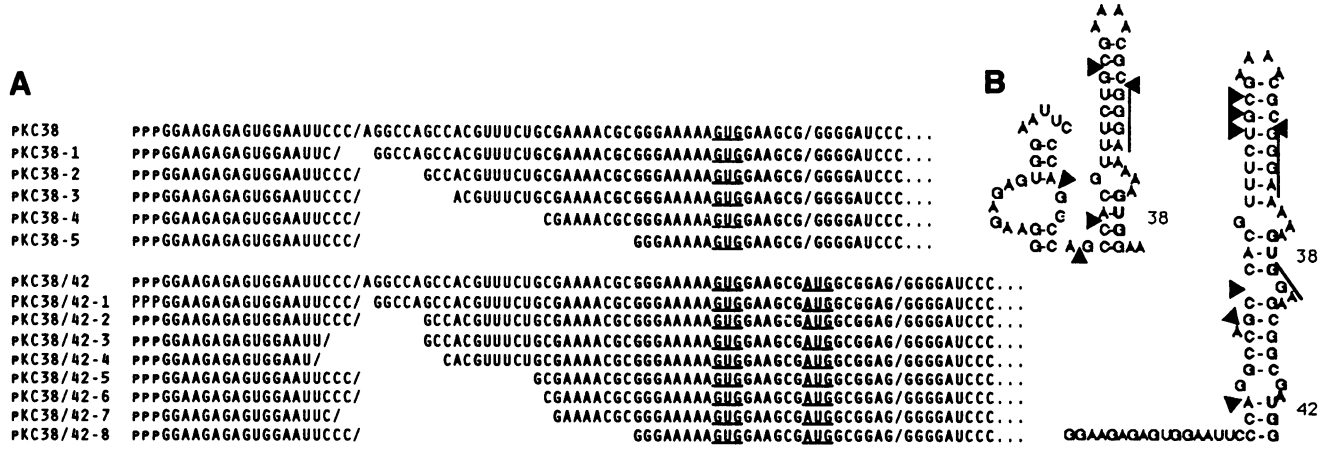

Figure 4. The 5′ terminal RNA sequences (A) and predicted structures (B) for *lacI-Z* fusions testing the Val38 and Met42 restart sites for primary initiation activity. The two parent plasmids contain small fragments bearing the indicated restart sites which previously were cloned into the *Sma*I site of pKC8, a pMC1403 derivative in which the I^Q promoter has been inserted into the *Eco*RI site (8). The deletion sets were generated from pKC38 and pKC38/42 by digestion with *Bal*31 nuclease as described in Materials and Methods. On the sequences (A), restart codons are underlined, and plasmid junctions are indicated by (/). On the structures (B), arrowheads indicate the positions of the deletion breakpoints and lines indicate Shine-Dalgarno complementaries. The structures are those with the minimum free energies predicted by the RNAFLD program of Zuker and Stiegler (24) using the energy parameters in Cech *et al.* (25).

neither the Val38 GUG site alone nor the Val38 GUG and Met 42 AUG sites in combination function as primary initiators at the 5′ end of a messenger RNA. However, these sequences have a high GC-content, and they had been shown to form stable secondary structures which influenced restart site selection (7). Thus an alternative explanation for inactivity of the restart sites was simply that their secondary structures prevented ribosome binding. To address this possibility, deletions of the *lacI* sequences in pKC38 and pKC38/42 have been made in the 5′→3′ direction to remove the sequences constituting the ascending stems of the structures. The sequences of the deletion variants obtained are shown in Fig. 4A, and the deletion endpoints are indicated in Fig. 4B on the predicted RNA structures. Table 3 lists the β-galactosidase activities for the deletion derivatives. Plasmid pKCWT, which contains a 59-bp fragment encoding the wild-type *lacI* gene initiator region, serves as the reference. The parent plasmid lacking any inserted initiation or reinitiation sites (pKC8) indicates the background level of activity. In *lac* deletion hosts harboring the constructs, β-galactosidase activity is consistently low. Although the potential for structure is not absent entirely in the deletion derivatives, it varies considerably. In a number of cases the predicted base-pairings (24) are weak or leave the initiator triplet and Shine−Dalgarno complementarities in regions of the RNA predicted to be single-stranded. Thus, if the restart sites had independent initiation activity, one would have expected to see activity in at least one of the sequence contexts provided by the deletion series. In other cases of structurally occluded initiator regions, the strategy used here has been successful in unmasking initiation activity (26−28).

Table 3. Assays of independent initiation from the Val38 and Met42 restart sites

Plasmid	β-Galactosidase (units)	Initiation efficiency (%)
pKCWT	1800	100
pKC8	9	0.5
pKC38	11	0.6
pKC38-1	8	0.4
pKC38-2	9	0.5
pKC38-3	13	0.7
pKC38-4	4	0.2
pKC38-5	4	0.2
pKC38/42	8	0.4
pKC38/42-1	14	0.8
pKC38/42-2	14	0.8
pKC38/42-3	16	0.9
pKC38/42-4	10	0.5
pKC38/42-5	12	0.7
pKC38/42-6	19	1.0
pKC38/42-7	7	0.4
pKC38/42-8	5	0.3

Cultures of DS70 derivatives bearing the indicated plasmids were assayed for β-galactosidase activity. Plasmid designations are as in Fig. 4. Plasmid pKC8 is the parent plasmid lacking a translation initiation site upstream of the *lacZ* coding region. pKCWT contains the initiator region for the wild-type *lac* repressor protein. Each value is the average of at least 3 independent determinations. Relative initiation efficiency is compared to pKCWT, which is set at 100%.

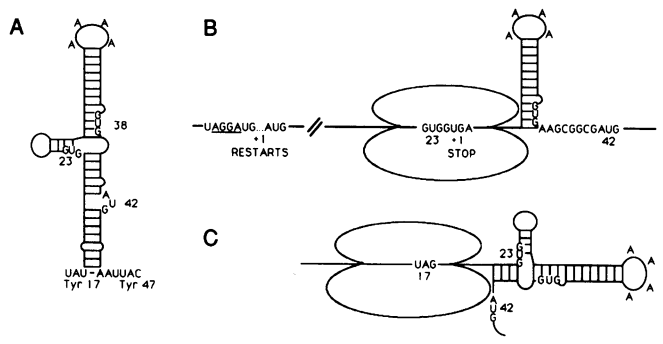

Figure 5. Scheme showing predicted RNA structures and translation in the early region of the *lac* repressor mRNA. (A) RNA beyond the position of ribosome release at UAG7, not undergoing translation. (B) RNA beyond the position of ribosome release at UAG7, undergoing translation of the open reading frame defined by the + 1 frame AUG and UGA codons shown. (C) RNA beyond the position of ribosome release at UAG17. The potential for structure was analyzed using the suboptimal folding program of Zuker (29) with the energy parameters of Turner *et al.* (30). The particular regions of the sequence folded were selected assuming ribosomes protect ~9 nucleotides beyond the position of the stop codon (31,32). The foldings shown in A and C are the optimal structures predicted. For case B, the base-paired stem-loop shown was consistently present in the optimal as well as 8 of 9 suboptimal folds. The sequences in the immediate vicinity of the Met42 AUG showed significantly less and quite variable potential for involvement in secondary structure. The AUG occurred in the loop of short hairpins, in single-stranded RNA, or in the positions of interior or bulge loops within irregular stems.

DISCUSSION

Out-of-frame translation and RNA structure bridge reinitiation events

From the outset, two features of reinitiation in the early region of the *lac* repressor mRNA have been particularly puzzling: restarts occur quite far from the termination codon, and both the Val23 GUG and Met42 AUG restart polypeptides are made in response to UAG7, UAG12, and UAG17 (5). The mRNA sequence and the most stable stem-loop structure predicted for this segment of the mRNA by early programs offered a possible explanation for the occurrence of both restarts together (6): the base-paired stem was positioned so as to mask the Val23 GUG restart site but leave the Met42 AUG and its Shine−Dalgarno sequence accessible just downstream. Ribosomes were thus proposed to bind to the Met42 AUG and, by destabilizing the structure, potentiate binding to the Val23 GUG. This view became less tenable when more sophisticated RNA folding programs (24,29) and nuclease mapping data (7) indicated

Site		Space	Rel. Eff.	Fit
lacI +1 AUGs	UGAAUGUGAAACCAGUAACGUUAUAGGAUGUCGCAGAGUAUGCC	1/13	15	26
lacI Val23	UUAUCAGACCGUUUCCCGCGUGGUGAACCAGGCCAGCCACGUUU	-1	2.7	34
lacI Val38/42	GCCACGUUUCUGCGAAAACGCGGUAAAAAGUGGAAGCGGCGAUG	3/15	32	40
lacI Met42	CUGCGAAAACGCGGGAAAAAGUGUAAGCGGCGAUGGCGGAGCUG	6	40	49
lacI Leu62	UGGCACAACAACUGGCGGGCAAAUAGUCGUUGCUGAUUGGCGUU	3	3.5	17
f1 gene VII	GUCUGCGCCUCGUUCCGGCUAAGUAACAUGGAGCAGGUCGCGGA	1	11	17

Figure 6. Sequences of translational reinitiation sites, aligned on the closest termination codon known to activate each site. Included along with the early *lacI* restarts is the phage f1 gene VII initiation site, which has all the properties of a restart site (9). Termination and restart codons are in boldface type; potential Shine-Dalgarno complementarities are underlined. *Space* denotes the number of nucleotides between stop and restart triplets, with the 1-nucleotide space between the Val23 GUG and the +1 frame UGA downstream indicated in the opposite sense as −1. *Rel. Eff.* is reinitiation efficiency expressed relative to the translation level upstream of each restart, which is set at 100%. *Fit* is the extent in percent to which the region of 35 nucleotides including the stop and restart codons matches a consensus sequence derived from a group of initiation sites with independent activity (M. Ivey-Hoyle and D. Steege, unpublished results). The restart sites were aligned with this consensus sequence by placing the stop codons in the position of the codon which would lie in the ribosomal A site.

extensive basepairing of sequences throughout the early region, which effectively buries all of the in-frame restart sites (Fig. 5A). Also, from the lack of independent activity displayed by the restart sites (8), it appeared that reinitiation would eventually prove not to be explained in terms of ribosome binding, but would instead involve reuse of 30S subunits already present on the mRNA from upstream translation.

A more plausible explanation for the atypical pattern of reinitiation in the early region of the *lacI* gene sequence that concurs with the mRNA's potential for structure is provided by present experiments which have looked in more detail at events in both the 0 frame and the +1 frame. The data confirm that in-frame reinitiation is at background levels for nonsense mutations at repressor amino acid positions 2, 3, 5, and 6. In-frame reinitiation then becomes detectable when a C→G change generates both a UAG at position 7 as well as a 4-base Shine–Dalgarno sequence. Reinitiation in the +1 frame in response to UAG7 is occurring at 10-fold higher levels than restarts in the repressor reading frame, however, and there are three potential initiator triplets in this frame between UAG7 and the Val23 GUG. These results suggest that the UAG7 nonsense mutation has actually activated reinitiation events in the +1 frame which are needed for the subsequent in-frame restarts at Val23 GUG and Met42 AUG. Our finding that in-frame reinitiation in response to UAG7 is abolished if the +1 frame initiator triplets are removed by point mutations indicates that translation in the +1 frame is indeed a requirement for downstream restarts in the *lac* repressor reading frame.

Reinitiation in the +1 frame would be expected to do two things to potentiate the in-frame restarts, as shown in Fig. 5B. Translation in the +1 frame will take ribosomes to a UGA stop codon spaced one nucleotide downstream from the in-frame Val23 GUG, thereby positioning ribosomes to restart at this GUG. At the same time, ribosomal movement on the mRNA through the +1 reading frame will prevent the lower two-thirds of the structure shown in Fig. 5A from forming. Basepairing of the sequences beyond the position of the UGA stop will still occur, however, and will keep the Met42 AUG in close proximity to ribosomes terminating at the +1 frame UGA. A reinitiation event at the Met42 AUG would then only require accommodating and somehow bypassing the protruding structure. RNA structure as a factor in bridging distance would also explain why the more distal nonsense mutations at UAG12 and UAG17 activate the same two in-frame restarts. The positions of these nonsense mutations allow ribosomes translating the beginning of the gene to move closer to the two restarts (Fig. 5C), and during the fraction of time nucleotides at the base of the stem are not annealed, the restarts will be accessible. Thus, singly or in combination, out-of-frame translation and basepairing of noncontiguous sequences would provide a means of propagating translational activity over substantial distances in the messenger RNA.

There is not as yet direct evidence that the 30S subunit from a termination reaction can remain bound to the mRNA and reinitiate translation, let alone reinitiate translation beyond intervening RNA secondary structure. However, the experiments of Martin and Webster (33) showed that the 30S subunit persists on the mRNA following release of the 50S subunit. Also, from the fact that elongating ribosomes can in certain instances bypass sequences in mRNA to translate more distal codons, it is apparent that the mRNA binding site on the ribosome must be able to accommodate protruding bulges and structures. Some frameshifts result from hops of 5 and 6 nucleotides (34), but the most remarkable case of hopping is phage T4 gene 60, complete translation of which requires that elongating ribosomes bypass an interruption of 50 nucleotides (35).

Primary sequence and reinitiation efficiency

With data for the reinitiation events activated by only a few nonsense mutations, our previous analysis (8) left open entirely the question of how primary sequence elements in the restart sites determine reinitiation efficiency. On the basis of the present results, there are now several more cases in which the restart codon falls within the domain of the terminating ribosome, and therefore in which primary sequence effects should predominate over the influence of RNA secondary structure. This group of restart sites is illustrated in Fig. 6. The sequences have been aligned on the functional stop codon. Indicated for each is the number of nucleotides in the space between the stop and restart codons and the site's efficiency relative to the level of upstream translation. The efficiency given for reinitiation at the Val23 GUG in UAG7S was calculated relative to the level of upstream translation in the +1 frame which would terminate at the UGA just beyond the GUG (Fig. 4B). That for the Val38 GUG is some fraction of 32%, since a mixture of the Val38 and the Met42 restart proteins are produced in response to UAA36 (7,8). Finally, as a measure of the degree to which the sequence contexts presented to the terminating ribosome are similar to a ribosome binding site, the fit to a consensus sequence we have derived for primary initiation sites (M. Ivey-Hoyle and D. Steege, unpublished results) is indicated.

Also included in the list of Fig. 6 is the initiation site for phage f1 gene VII, which like the *lacI* restarts, apparently lacks the information required for recognition by 30S ribosomes (9). Translation of gene VII occurs only via coupling to upstream gene V, but is 10-fold lower than gene V translation even at the normal 1-nucleotide spacing from the gene V stop codon. Activity from the VII site is extremely sensitive to stop codon position, dropping by 80% if the spacing is increased from 1 to 5 nucleotides. In all respects the VII site behaves as a reinitiation site (9), and it provides an additional example of a restart spaced 1 nucleotide downstream from the point of termination.

The data in Fig. 6 add to growing evidence arguing that the primary sequences of restart sites and coupled initiation sites are important determinants of efficiency. The group of sites shown here exhibit a wide range of efficiencies despite relatively close spacing to the stop codon, and those with the closest spacings do not have the highest activities. None of the sites with Shine–Dalgarno complementarities of less than 3 nucleotides has substantial activity, a pattern concurring with other data demonstrating this element's importance. Das and Yanofsky (36) showed that coupled expression at the *trpB-trpA* intercistronic junction falls dramatically when deletions remove the Shine–Dalgarno sequence in the *trpA* ribosome binding site. Spanjaard and van Duin (37) found that reinitiation efficiency in the absence of the Shine–Dalgarno sequence AGGAGGU is about 10% of that in its presence. On the other hand, the very low activity seen here for the *lacI* Leu62 restart argues that the presence of a recognizable Shine–Dalgarno complementarity does not guarantee efficient reinitiation. With a suboptimal UUG initiator triplet and an overall sequence quite dissimilar to a prototypic independent initiation site, the Leu62 restart site is clearly missing needed determinants. Perhaps the most intriguing question raised by the present group of sites is why reinitiation

at the Met42 AUG spaced 6 nucleotides downstream of UAA39 is by far the most efficient restart. Two features of note distinguish this site from the others. First, the overall sequence context presented to the terminating ribosome shows the highest similarity to an independent ribosome binding site. Second, in addition to the UAAG Shine–Dalgarno complementarity upstream of the AUG, there is another complementarity, GGGA, spaced 7 nucleotides upstream from the stop codon. In light of evidence from studies of frameshifts indicating that the 3′ end of 16S rRNA scans the message during elongation (38), it is tempting to suggest that as the ribosome undergoes termination, an interaction between the 16S rRNA 3′ end and the GGGA sequence comes into play. By slowing dissociation of the 30S subunit, such an interaction would conceivably increase the probability of a subsequent restart at the Met42 AUG. Clearly, systematic mutational analysis of individual sites will be required to test this idea and to identify other potential contacts between the mRNA and the 30S ribosome that determine the frequency of translational reinitiation following a termination event.

ACKNOWLEDGEMENTS

The authors are indebted to J.Miller for bacterial strains and for permission to cite unpublished information from his laboratory. They thank K.Cone, P.Pavco, M.Ivey-Hoyle and A.Kendall for their assistance with this study, and R.Weiss for helpful suggestions. This work was supported by a grant from the National Institute of General Medical Sciences (GM33349).

REFERENCES

1. Sarabhai,A. and Brenner,S. (1968) *J. Mol. Biol.* **27**, 145–162.
2. Napoli,C., Gold,L. and Singer,B.S. (1981) *J. Mol. Biol.* **149**, 433–449.
3. Platt,T., Weber,K., Ganem,D. and Miller,J.H. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 897–901.
4. Ganem,D., Miller,J.H., Files,J.G., Platt,T. and Weber,K. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3165–3169.
5. Files,J.G., Weber,K. and Miller,J.H. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 667–670.
6. Steege,D.A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4163–4167.
7. Cone,K.C. and Steege,D.A. (1985a) *J. Mol. Biol.* **186**, 725–732.
8. Cone,K.C. and Steege,D.A. (1985b) *J. Mol. Biol.* **186**, 733–742.
9. Ivey-Hoyle,M. and Steege,D.A. (1989) *J. Mol. Biol.* **208**, 233–244.
10. Farabaugh,P.J. (1978) *Nature* **274**, 765–769.
11. Beyreuther,K. (1978) In Miller,J.H. and Reznikoff,W.S. (eds.) The Operon. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 123–154.
12. Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
13. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Miller,J.H., Ganem,D., Lu,P. and Schmitz,A. (1977) *J. Mol. Biol.* **109**, 275 –301.
15. Calos,M.P. (1978) *Nature* **274**, 762–765.
16. Hanahan,D. (1983) *J. Mol. Biol.* **166**, 557–580.
17. Cone,K.C., Sellitti,M.A. and Steege,D.A. (1983) *J. Biol. Chem.* **258**, 11296–11304.
18. Young,R.A. and Davis,R.W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
19. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
20. Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.* **65**, 499–560.
21. Casadaban,M., Martinez-Arias,A., Shapira,S.K. and Chou,J. (1983) *Methods Enzymol.* **100**, 293–308.
22. Miller,J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
23. Ross,G.W. and O'Callaghan,C.H. (1975) *Methods Enzymol.* **43**, 69–85.
24. Zuker,M. and Stiegler,P. (1981) *Nucleic Acids Res.* **9**, 133–145.
25. Cech,T.R., Tanner,N.K., Tinoco,I.,Jr., Weir,B.R., Zuker,M. and Perlman,P.S (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3903–3907.
26. Kastelein,R.A., Berkhout,B. and van Duin,J. (1983) *Nature* **305**, 741–743.
27. Berkhout,B. and van Duin,J. (1985) *Nucleic Acids Res.* **13**, 6955–6967.
28. Schmidt,B.F., Berkhout,B., Overbeek,G.P., van Strien,A. and van Duin,J. (1987) *J.*†Mol. Biol. **195**, 505–516.
29. Zuker,M. (1989) *Science* **244**, 48–52.
30. Turner,D.H., Sugimoto,N. and Freier,S.M. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 167–192.
31. Steitz,J.A. (1969) *Nature* **224**, 957–964.
32. Hartz,D., McPheeters,D.S., Traut,R. and Gold,L. (1988) *Methods Enzymol.* **164**, 419–425.
33. Martin,J. and Webster,R.E. (1975) *J. Biol. Chem.* **250**, 8132–8139.
34. Weiss,R.B., Dunn,D.M., Atkins,J.F. and Gesteland,R.F. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 687–693.
35. Huang,W.M., Ao,S.-Z., Casjens,S., Orlandi,R., Zeikus,R., Weiss,R., Winge,D. and Fang,M. (1988). *Science* **239**, 1005–1012.
36. Das,A. and Yanofsky,C. (1984) *Nucleic Acids Res.* **12**, 4757–4768.
37. Spanjaard,R.A. and van Duin,J. (1989) *Nucleic Acids Res.* **17**, 5501–5507.
38. Weiss,R.B., Dunn,D.M., Dahlberg,A.E., Atkins,J.F. and Gesteland,R.F. (1988) *EMBO J.* **7**, 1503–1507.