Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*

Lori A.Wagner[1]*, Robert B.Weiss[1,2], Robert Driscoll[3], Diane S.Dunn[1,2] and Ray F.Gesteland[1,2]
[1]Department of Human Genetics, 743 Wintrobe Building, University of Utah, Salt Lake City, UT 84132, [2]Howard Hughes Medical Institute, 743 Wintrobe Building, University of Utah, UT 84132 and [3]Biology Department, Wilson Hall, University of North Carolina, Chapel Hill, NC 27514, USA

Received March 16, 1990; Accepted May 16, 1990

ABSTRACT

A run of 11 adenine or thymine residues at the 5′ end of an out-of-frame *lacZ* gene causes a high level of β-galactosidase expression in *E. coli*. This effect was not observed for a run of guanine residues. Reverse transcription of mRNA isolated from *E. coli* containing the run of 11 A's reveals heterogeneity of transcript length while reverse transcription of mRNA isolated from *S.cerevisiae* containing the same gene shows no heterogeneity. Protein sequencing of the β-galactosidase molecules derived from the out-of-frame construct containing a run of adenines reveals the addition of a lysine at the run. A new method was developed where messages small enough to allow resolution of single nucleotide differences on an acrylamide gel are electrophoresed, electroblotted onto nylon and probed. This confirmed the reverse transcription results and showed that additional residues can be added to transcripts derived from DNA containing 10 or 11 thymine residues. A mechanism for slippage is discussed where the A-U rich RNA-DNA hybrid can denature during elongation and rehybridize in an offset position, causing the addition of extra residues to the transcript.

INTRODUCTION

The mechanistic aspects of transcription allow *E. coli* RNA polymerase to maintain an error frequency of approximately 10^{-5} during in vivo transcription.[1,2,3] Several changes leading to productive elongation occur when the nascent transcript is 8−9 bases long.[4] During this time: the initiation factor sigma is lost from the holoenzyme, the RNA-DNA hybrid reaches its mature length of approximately 12 base pairs, the transcription bubble reaches a length of about 11−17 denatured DNA base pairs and the RNA polymerase DNA footprint decreases from 30 to 23 basepairs.[5,6,7,8,9] To maintain accuracy, the correct ribonucleoside triphosphate must be incorporated into the growing RNA chain and the RNA polymerase must not lose its position on the template. For accurate transcription to occur, bases should not be added to or subtracted from the templated information during elongation. Incorporation of incorrect bases during transcription has recently been shown to be monitored by an editing function intrinsic to the RNA polymerase.[10]

The accuracy of transcription has certain exceptions: addition of non-templated residues to the mRNA has been reported in eukaryotic viruses. Late transcripts in vaccinia virus contain 35 adenine residues at their 5′ ends although only 3 adenines are encoded in the DNA.[11,12,13] Vesicular stomatitis virus (VSV) transcripts are extensively polyadenylated at their 3′ ends, however only 7 thymine residues exist at the corresponding site in the DNA.[14,15,16] A mutant VSV polymerase affects the extent of polyadenylation.[17,18] Reiterative copying of the thymine residues in the DNA by slippage of the viral RNA polymerase has been proposed as a mechanism for polyadenylation in both these viruses.

Evidence indicates that *E.coli* RNA polymerase is also capable of transcriptional slippage. In vitro transcription of calf thymus DNA by *E. coli* RNA polymerase yields a transcript of polyadenylic acid when ATP is the only ribonucleoside triphosphate added to the reaction.[19] The length of the transcript is 5−10 times longer than the template and the addition of any or all of the other 3 ribonucleoside triphosphates results in inhibition of polyadenylic acid production. A mechanism was proposed where the RNA polymerase reiteratively transcribes short thymidylic acid regions in the template by repeated cycles of melting the RNA-DNA hybrid, slippage of the two strands relative to each other, rehybridization and subsequent elongation. A recent paper showed that the 5′ ends of in vitro transcripts derived from *tet* promoters which encode 4 adenines at the mRNA start site are heterogeneous, containing 1−12 adenine residues.[20] *E. coli* RNA polymerase slippage during transcription initiation was proposed to explain the result.

We show that high level addition of variable numbers of uracil or adenine residues arise from transcripton of plasmid DNA in *E. coli* containing runs of 10 or more thymine or adenine residues. We propose RNA polymerase slippage during elongation as a mechanism to account for this phenomenon.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Constructions

Oligonucleotides were synthesized by an Applied Biosystems 380A or B DNA synthesizer and the crude material cloned into a pBR322 derived vector containing an engineered derivative of the *lacZ* gene. The out-of-frame vectors were created by inserting oligonucleotides into a previously described vector cut with ApaI/Hind III.[21] The vector used in the mini-message experiment is as follows: a *lacZ* fragment containing a HindIII site in the 5th codon of the gene was cut by EcoRI/DraI releasing a fragment from a pUC9 polylinker to the DraI site in *lacY*. This fragment was cloned into the EcoRI site and the HindIII site (blunted) on pKK223-3 (P.L. Biochemico Inc.). The translational start of *lacZ* was modified by inserting an oligonucleotide between the EcoRI site and the HindIII sites at the 5′ end of the *lacZ* gene. A synthetic Tac promoter was created with a unique SalI site between the −10 region in the promoter and the transcriptional start by cloning a EagI/HindIII oligo into the EagI site in the tet^r gene of pBR322 and the HindIII site on the vector. The mini-message construct was created by cloning the oligonucleotide, 5′AGCTTGTAAGCAAGCCCGCCTAATGAGCGGGCTTTTTTTT3′ into the HindIII site as a transcriptional and translational terminator.[22,23] An oligo containing 11T's with a T or C mixed at the 7th position in the run was then cloned into the HindIII site (5′AGCTAGATCTTTTTT(T/C)TTTTA). Appropriate oligos containing runs of 7, 8, 9, or 10 T's were cloned into the same position. Inserts were verified by dideoxy sequencing of the plasmid DNA.[24] The bacterial strain used was *E. coli* K-12 SU1675 del- *prolac, recA56*/F′IQ.[25] The yeast constructs were created by cloning into a modified YEp13 vector in which the HindIII and SstI sites had been filled in.[26] A SalI/PstI partial digest allowed a 7 kb SalI/PstI fragment from pLG669-Z to be inserted.[27] The *lacZ* fragment contained a BamHI linker containing a HindIII site. Sequences from the out-of-frame constructs were cloned into the yeast vector by HindIII/SacI cuts and insertion of HindIII/SacI *lacZ* fragments from the out-of-frame constructs.

β-galactosidase activity measurements and protein sequencing

Whole cell β-galactosidase assays and activity measurements for *E. coli* were done as previously described.[21] Cells were grown at 37°C in L Broth with 2mM IPTG and 100 μg/ml of ampicillin. Yeast β-galactosidase assays were done on strain SX50-1C (leu2-3, 2-112, his3, ura3-52, trp1-289 am) of *Saccaromyces cerevisiae* grown at 30°C to an OD_{600} of .2−.4 in YNB-leu, 2% raffinose.[28] Purification and amino-terminal sequencing of β-galactosidase was as previously described.[21]

RNA isolation and sequencing

1.0 ml of log-phase *E. coli* growing at 37°C in LB (RNA sequencing) or M9 supplemented with 0.4% casamino acids, 1 μg/ml B1, and 0.4% glucose (primer extension) was quickly added to 1.0 ml of boiling 20 mM NaAcetate pH 5.2, 0.3M sucrose, 2% SDS (the salt was not included if the culture was grown in M9). The mixture was boiled for 30 seconds, then mixed with an equal volume of phenol equilibrated with 100mM Tris pH 8.0 that had been heated to 55°C. The mixture was vortexed vigorously, centrifuged, extracted and reextracted with a phenol/chloroform solution. After EtOH precipitation the RNA was resuspended in 20mM $NaHPO_4$, 1mM EDTA and quantified by OD_{260}. 40 micrograms of this RNA was used in a 15 microliter annealing mix (50 mM Tris HCl pH 8.3, 60 mM NaCl, 10mM DTT) with a ^{32}P end labelled primer for 15 minutes at 37°C. $Mg(OAc)_2$ was added to 6 mM and the mRNA was sequenced with AMV reverse transcriptase (Life Sciences). Yeast RNA was prepared by resuspending 2×10^8 cells (grown in YEPD, 30°C) in 0.2 ml of 0.5 M NaCl, 0.2M Tris pH 7.5, 10mM Na_2EDTA, 1% SDS with 0.4 g of acid washed glass beads (0.3 mm). 0.2 ml of phenol/CHCL3 was added and the tube vortexed for 2.5 min., 0.3 ml of buffer was added, mixed, the aqueous phase was removed, re-extracted with phenol/$CHCl_3$, and precipitated with 2.5 vols of EtOH. 80 micrograms of this RNA prep was used in a sequencing reaction.

Electroblots

Four micrograms of RNA were boiled 2 min. in loading buffer (deionized formamide, 0.25% bromphenol blue, 0.25% xylene cyanol) then electrophoresed in a 7.5% polyacrylamide gel. The gel was removed from siliconized plates by placing Whatman 3mm paper over it and then electroblotted onto Biodyne A or Genescreen in 0.5× TBE for 30 min. at 1 or 1.5 Amps. The wet nylon membrane was UV crosslinked, then baked at 120°C for 10 min. Hybridization occurred at 45°C overnight in 15 mls of 40mM Na_2HPO_4, 130mM NaCl, 13% PEG, 5% SDS and 20 ng 60mer probe tailed with α^{32}ATP. The blot was washed at room temp. with solution that had been heated to 37°C. The first 3 washes were in 130mM NaCl, 40mM Na_2HPO_4 pH 7.2, 5% SDS and 1mM EDTA. In the last 3 washes the SDS was reduced to 1%. The blot was exposed to film for 20 min.

RESULTS

Plasmids were constructed in which the *lacZ* gene was in the zero, +1 or −1 reading frame and a run of 11 nucleotides was cloned into the 5′ end of the coding region in each construct (Figure 1). It was expected that only the zero frame construct would result in a high level of β-galactosidase expression since the ribosomes translating the messages in the +1 or the −1 frame would encounter stop codons soon after initiation. Instead we observed that a run of 11 adenine residues in any frame resulted in a high level of β-galactosidase expression. The zero frame construct containing the run of 11 A's, p1801, (Figure 1) gave a β-galactosidase activity of 42,344 units while the constructs in the +1 and −1 frames, 4p1802 and 2p1802, had an activity level of 13,110 and 11,434 units, respectively. A +1 frame construct containing 11 thymine residues (4p1901) also exhibited a high β-galactosidase level of 10,678 units.

Neither the run of 11 A's nor the run of 11 T's retained their high activity level when the homopolymeric run was interrupted by a single nucleotide. The +1 construct containing 11 A's interrupted by a G at the 6th nucleotide (4p1803) and the +1 construct in which a run of 11 T's was interrupted by an adenine (4p1902) both had approximately 100 fold lower levels of β-galactosidase than the constructs containing an uninterrupted run. A run of 11 G's inserted into the 5′ end of the coding region of the +1 construct (4p2001) resulted in a low β-galactosidase level of 101 units, unlike the constructs containing the adenine or thymine runs. Several possible explanations for this phenomenon include: ribosomal frameshifting, transcriptional events where residues are added to or subtracted from the message, or a DNA replication effect. DNA replication errors seem unlikely as large regions of eukaryotic DNA containing adenine or thymine runs have often been unambiguously sequenced using *E. coli* as a vector. Therefore the most plausible

Series	frame	slip site	ß-galactosidase activity (whole cell units)
p1801	(0)	AUG- - -UU A AAA AAA AAA A CG- - - lacZ	42,344 +/- 3457
4p1802	(+1)	A AAA AAA AAA A	13,110 +/- 345
2p1802	(-1)	A AAA AAA AAA A	11,434 +/- 365
4p1803	(+1)	A AAA AGA AAA A	125 +/- 2
4p1901	(+1)	UUA GA U UUU UUU UUU U CG	10,678 +/- 779
4p1902	(+1)	U UUU UAU UUU U	100 +/- 2
4p2001	(+1)	AGC UU G GGG GGG GGG G AG	101 +/- 3

Figure 1. Series indicates the name of the construct from which the mRNA pictured is transcribed. Numbers in parentheses indicate the reading frame. β-galactosidase can be synthesized from a +1 frame construct by ribosomal frameshifting, transcriptional addition of 2 nucleotides or the transcriptional subtraction of one nucleotide. Likewise in the −1 frame constructs, one nucleotide may be added or two subtracted to make an in frame protein product. The 4p1800 series is designed to monitor three frames by contraction of the initial 4p-window to a p-window via an EcoRI cut and ligation (second EcoRI site borders *lacZ* codon 6) or a Bam HI fill-out to change the 4p-into a 2p- window[21]. The host is SU1675 F'IQ, 37°C in LB as described for β-galactosidase assays in materials and methods.

explanations for the high level β-galactosidase expression observed in out-of-frame constructs are ribosomal frameshifting or a transcriptional effect. The following experiments address this issue.

To investigate the possibility of transcriptional heterogeneity, RNA was isolated from *E. coli* containing the plasmid, 4p1802 or 4p1803, and the 5′ ends of the two messages were sequenced with reverse transcriptase (Figure 2A). The sequence ladders representing the 3′ region of the mRNA at the bottom of the autoradiogram for both the 4p1802 and 4p1803 transcripts are similar. 5′ of the region containing the run of 11 A's in 4p1802 and the interrupted run in 4p1803, the comparative clarity of the two ladders is markedly different. The ladder for 4p1803 remains readable throughout the autoradiogram but in 4p1802 heterogeneity is evident 5′ of the adenine run. DNA sequence of the plasmid 4p1802 using reverse transcriptase as the polymerase does not reveal heterogeneity 5′ of the adenine run, indicating that the DNA is not the source of the heterogeneity observed in the transcripts (unpublished results). These results could be explained if a heterogeneous population of RNA molecules were transcribed from the plasmid 4p1802. An alternative explanation is that the apparent sequence heterogeneity is arising during reverse transcription of the run of 11 A's in the mRNA.

To address the question of whether a reverse transcription artifact was occurring, a +1 frame yeast vector, y4p1802, was constructed which is identical within the coding region to the *E.coli* vector, 4p1802. This allowed us to ask whether the high β-galactosidase levels and heterogeneous mRNA sequences seen in *E. coli* containing plasmids with adenine runs was intrinsic to that organism or if these effects were also seen in yeast. In *Saccharomyces cerevisiae*, y4p1802 exhibits a β-galactosidase activity of 25 units which is only 0.8% of the in-frame activity level (Table 1). y4p1803, which is identical to y4p1802 except for a replacement of the 6th A in the run with a C, has an activity level of 1.5 units. RNA was isolated from *S. cerevisiae* containing the y4p1802 vector and sequenced. The sequence ladder which resulted was homogeneous (Figure 2A), implying that a heterogeneous population of messages arose in *E. coli* from 4p1802 but not from the equivalent vector, y4p1802, in yeast.

Table 1

Series		Activity	%
y4p1802	AUG... AGC UUA AAA AAA AAA ACG	25.0 units	0.8%
y4p1803	A AAA ACA AAA A	1.5 units	0.05%

The constructs pictured are both in the +1 frame and are equivalent to the *E. coli* constructs of the same name. β-galactosidase activity in *S. cerevisiae* was measured as in materials and methods. The percent column refers to the percent of β-galactosidase activity compared to the in-frame yeast β-galactosidase activity.

The protein sequence of the amino terminal region of β-galactosidase isolated from *E. coli* containing the +1 frame plasmid, 4p1802, is summarized in Figure 2B. This region was sequenced in order to find the site at which the reading frame is restored to zero and which amino acid was inserted to allow this restoration. 80% of the amino acid signal reveals 4 lysines following the leucine while 3 lysines appear in 20% of the same protein sequence. Following the lysines are an arginine and an isoleucine which represent translation of *lacZ* in the zero frame. The +1 frame was translated prior to the run of adenine residues, therefore, the zero frame has been restored by an event in the vicinity of the adenine run. This event could be ribosomes reading 4 A's as a lysine, resulting in a frameshift or alternatively,

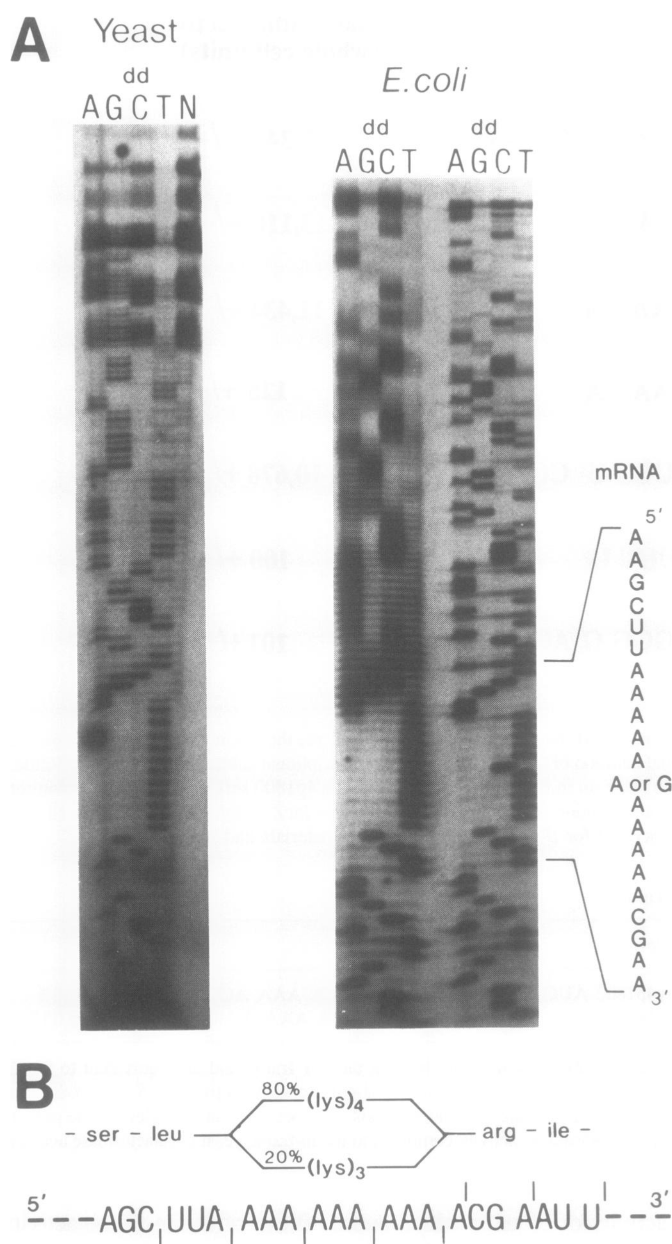

Figure 2. The dideoxy sequencing ladders generated from *lacZ* mRNA in *E. coli* (SU1675, constructs 4p1802 and 4p1803) and *S. cerevisiae* (y4p1802) are shown in A. 4p1802 mRNA, which is derived from a plasmid containing a run of 11 adenines, has a heterogeneous sequence ladder while neither mRNA derived from 4p1803, a plasmid containing an interrupted run of adenines, nor mRNA derived from y4p1802, a yeast plasmid containing a run of 11 A's, exhibit sequence heterogeneity. The NH_2- terminal protein sequence of 4p1803 is shown in B., illustrating a composite signal of 4 and 3 lysine residues decoded from the A string.

subtraction of an adenine residue during transcription. Both these events would produce a protein containing 3 lysines. The β-galactosidase molecules containing 4 lysines could be due to the transcriptional addition of two adenines or a -2 frameshifting event by the ribosome or both. Amino acid levels less than 20% of the total signal at each position cannot be accurately detected in this protein sequence.

To investigate transcript heterogeneity in detail an alternative technique was adopted where a mini-message, which is small enough to allow resolution of single base differences on an acylamide gel, is electrophoresed, electroblotted onto a nylon membrane and probed. In this fashion, transcripts differing in size by a single base can be visualized without subjecting the RNA to in vitro enzymatic manipulation. The mini-message was created by cloning a strong transcriptional terminator into the 5' end of the *lacZ* gene (C_1).[22,23] An oligonucleotide coding for a run of 11 uracils (U_{11}) was cloned into a restriction site in the region of the plasmid encoding the mini-message (Figure 3A). Three controls were used in the experiment to determine the average heterogeneity in length of a mini-message: the first control, C_1, is the original mini-message without the run of 11 U's, the second control, C_2, has a slightly longer coding sequence than C_1. The third control is identical to U_{11} except for a U to C change at the 7th base in the uracil run, creating a run of only 6 bases (U_6-Figure 3A). U_{11} mini-message uninduced by IPTG is shown in the far right lane (Figure 3B). The autoradiogram shown in Figure 3B shows the existence of several bands in the control lanes while the U_{11} transcript is markedly more heterogeneous in length. The lane containing the U_{11} transcript has 7 bands above those in the U_6 message and contains a total of at least 16 bands compared to 7 for the U_6 message. The length heterogeneity that exists in the controls is attributed to multiple transcription start and termination sites.

Primer extension analysis of the mini-messages confirmed the electroblotting results (Figure 3C). The lane containing the U_{11} mini-message reveals multiple bands in the region expected for a full length transcript. The control messages have several bands at their 5' ends which may be due to alternative transcriptional start sites.

To determine the number of consecutive thymine residues needed to cause transcript heterogeneity, a series of plasmids coding for mini-messages containing runs of 7, 8, 9, 10 and 11 uracil residues were transformed into *E.coli* and the resulting RNA used in electroblots. The lanes containing U_7 and U_8 (Figure 4) show bands similar in number to the controls shown in Figure 3B. U_9 appears to have one band corresponding to a transcript larger than those seen in U_7 and U_8. The U_{10} lane has a heterogeneous number of transcript sizes, all larger than those seen in U_7 and U_8. A run of approximately 10 thymine residues is required for transcript heterogeneity to occur.

DISCUSSION

Our results indicate that 10–11 consecutive adenine or thymine residues encoded in *E. coli* plasmid DNA cause a heterogeneous number of nucleotides to be incorporated into the transcript at the site of the homopolymeric run. Neither interrupted runs of A's or T's, nor strings of 11 guanine residues cause this transcriptional effect to occur. A yeast vector containing a run of 11 A's did not cause length heterogeneity in the resulting message. Slippage events during DNA replication have been proposed to explain mutational hotspots flanked by direct repeats in *E. coli*, as well as polymorphic G-T tracts and variable poly A tracts within Alu sequences in humans.[29,30,31,32,33] However, *E.coli* plasmid DNA containing a run of 11 adenine residues does not exhibit heterogeneity when sequenced. These results, as well as the length heterogeneity seen in electroblots of transcripts derived from plasmids containing runs of 10–11 T's, confirms that nucleotide addition to *E. coli* transcripts takes place in vivo.

The stability of the RNA-DNA hybrid during transcriptional elongation may provide the explanation for these results. The

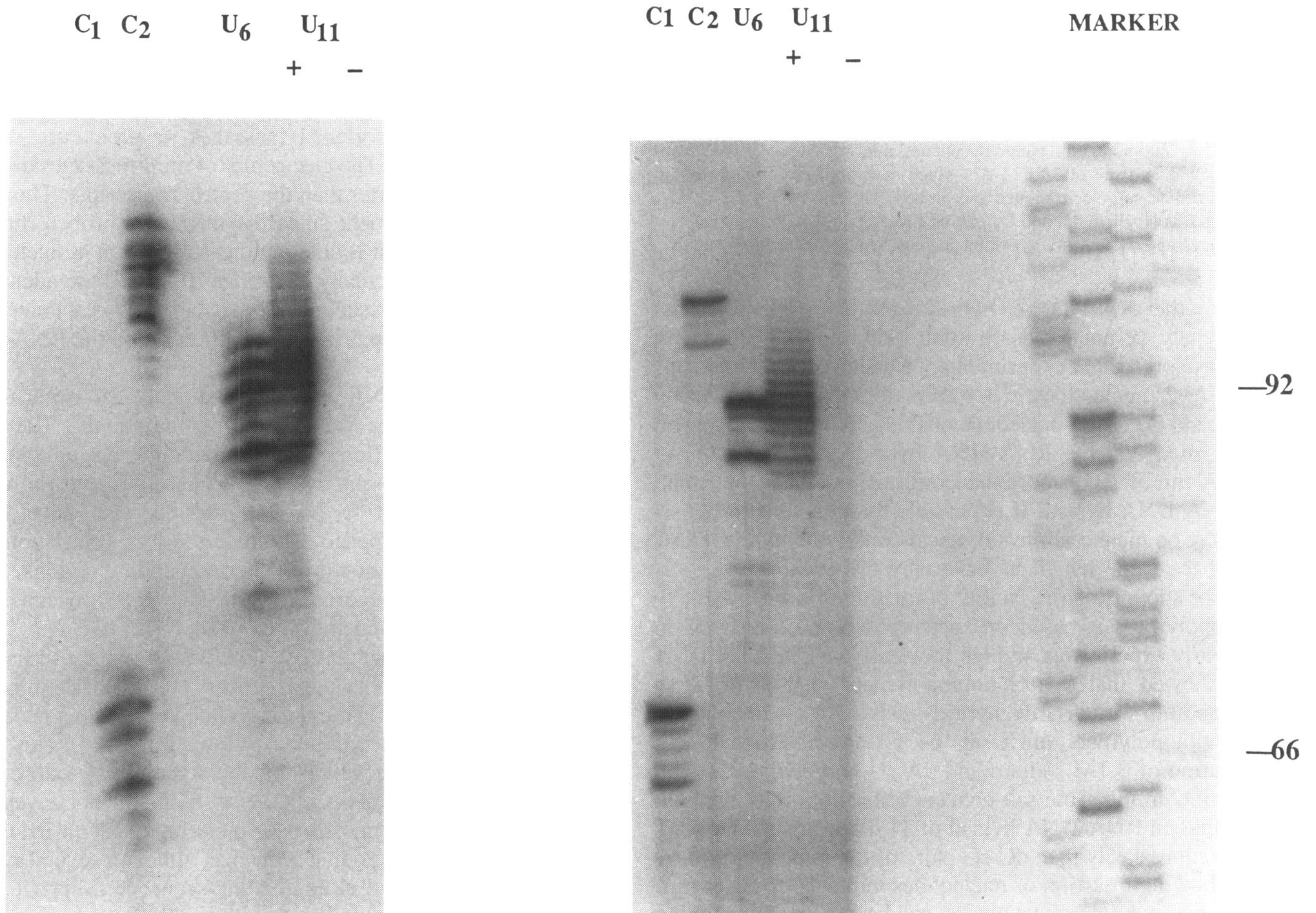

Figure 3. A. A schematic of the mini-messages U_6 and U_{11} is shown. The 5′ leader is represented by a line to the left of the AUG. The coding sequence is in non-highlighted letters while the run or the interrupted run is highlighted by a block. A strong stem-loop followed by several uridines serves as a transcription terminator. B. The first lane contains Control 1, which is the mini-message pictured above without the highlighted block (94 nucleotides). Control 2 does not contain the highlighted block but it has a much longer coding region making the mini-message 144 nucleotides in length. U_6 is in the third lane while U_{11} induced and uninduced (+/−) by IPTG are in the 4th and 5th lanes. The order of the lanes is identical in the reverse transcription in C. except for the addition of an arbitrary sequencing marker to the right.

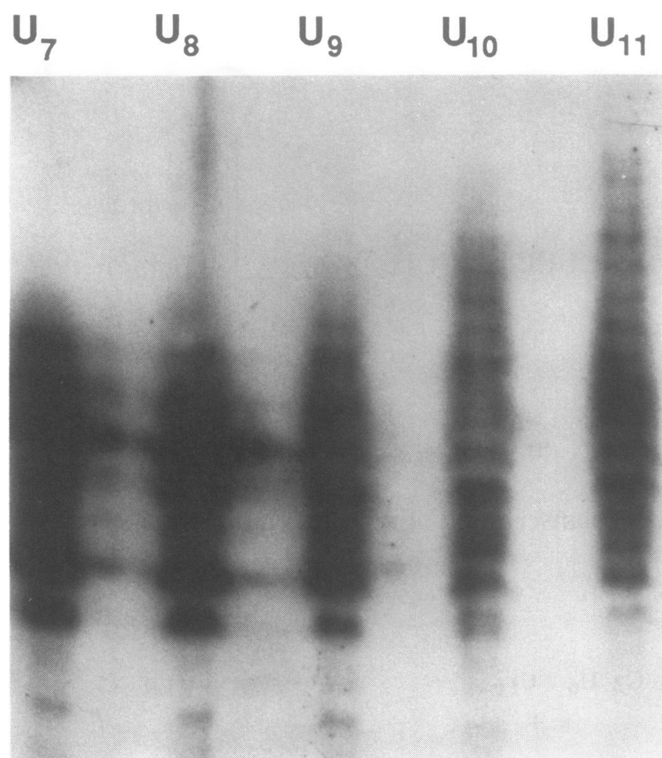

Figure 4. The autoradiogram pictured contains mini-messages derived from plasmids containing a run of 7 to 11 A's which have been electrophoresed on a 7.5% acrylamide gel, electroblotted and probed. The lanes show the U_7-U_{11} mini-messages as labelled. U_7 and U_8 exhibit the average length heterogeneity for a mini-message while U_9-U_{11} exhibit increasing heterogeneity in length.

length of the RNA-DNA hybrid has been identified as approximately 12 nucleotides by two independent methods. In vitro RNase protection experiments yield short RNA fragments of 12 +/−2 nucleotides.[34] Crosslinking experiments using gamma azido GTP as a nucleotide analog and crosslinking agent has confirmed that the RNA-DNA hybrid is 12 basepairs in length.[8] A run of 10 or more nucleotides approaches the length of the RNA-DNA hybrid. It is possible that this length of A-U hybrid may be melted at physiological conditions, while a G-C hybrid of the same length is less easily disrupted.

Evidence supporting the facility of melting of dA-rU and dT-rA homopolymers is based on spectrophotometric analysis in vitro.[35] Early experiments on high molecular weight RNA-DNA hybrids showed that dG:rC homopolymers melt at 62°C in 0.001M sodium while rG:dC hybrids melt at 89°C. In contrast, rA:dT homopolymers melt at 64°C at the higher salt concentration of 0.1M sodium and dA:rU homopolymers melt at only 45°C in the same salt concentration. It is reasonable to suppose that an RNA-DNA hybrid of 11 dA-rU or dT-rA pairs may melt more easily than dC-rG pairs of the same length. Our results, showing insertion of nucleotides into transcripts at runs of A's or T's, but not at runs of G's, can be expained by melting of an RNA-DNA hybrid consisting of dA-rU or dT-rA pairs during elongation. If rehybridization occurs in an offset position, nucleotides could be added or subtracted from the transcript during elongation.

Alternate re-pairing of the RNA-DNA hybrid, after melting, can easily occur in a homogeneous run of any kind. Although runs of U's interrupted by a single A may be melted as easily as runs of 11 U's, re-hybridization can not occur in an offset position, because of the mismatch this creates in the middle of the RNA-DNA hybrid. This could explain the low β-galactosidase activity of 4p1902 (Figure 1) which contains a run of U's interrupted by a single A.

Provided the RNA-DNA hybrid can melt and rehybridize in an offset position, elongation must be able to occur in this altered context. Addition of nucleotides to the RNA suggests that the RNA moves towards the 3′ end of the DNA strand to which it is hybridized, leaving the 3′ base of the RNA, where synthesis occurs, paired with the DNA in the normal fashion. Based on the electroblot data, primer extension results and protein sequencing, addition of nucleotides and thus the slippage of RNA towards the 3′ end of the DNA, is substantially favored over slippage in the reverse direction. Slippage of the message in the opposite direction, towards the 5′ end of the DNA, creates a mismatch between the RNA and the DNA at the 3′ end of the message. This mismatch could be tolerated by the RNA polymerase, or the offending base could be removed by an editing function, allowing elongation to continue[10]. It is also possible for elongation to be aborted at this point. Subtraction of nucleotides by RNA slippage in the direction described is supported by the existence of 3 lysines in 20% of the amino acid signal from β-galactosidase sequence (Figure 2B), however, this experiment cannot prove that subtraction occurred transcriptionally. Primer extension of the mini-messages reveals bands shorter than the control (Figure 3C). It is unknown if these bands are due to shortened messages or premature reverse transcriptional stops. The electroblot of mini-messages does not reveal any bands shorter than the control transcripts. This is the most reliable experiment since the mRNA is probed directly, therefore we believe it is unlikely that subtraction of nucleotides from the transcript actually occurs at 10 or 11 base adenine or thymine runs although subtraction may be possible at longer runs due to increased basepairing possibilities within the RNA-DNA hybrid.

It is notable that RNA polymerase slippage can occur during elongation in vivo as well as during initiation[20]. The RNA polymerase is in a different conformation during initiation, the sigma factor is present and the RNA-DNA hybrid is not completely formed[4,8,34]. Most transcripts also have leader sequences where slippage would not affect translation if it occurred during initiation or in other non-coding regions. These factors may make it more likely for slippage to occur during initiation or in non-coding regions in natural *E.coli* genes. Slippage during elongation potentially affects the translational reading frame. It is unlikely that slippery adenine or thymine runs would be tolerated in natural *E. coli* coding regions because 3 reading frames are represented following the run. This could confer a selective disadvantage on the organism. A search of the Genbank EMBL database release # 61.0 did not reveal any adenine or thymine runs 10 base pairs or longer in natural *E. coli* or bacteriophage coding sequence, although several runs of this length were found in non-coding sequence. Slippage of *E. coli* RNA polymerase implies that it can also jump several bases within the denatured DNA transcription bubble onto an identical stretch of heteropolymeric A-T rich template. This is somewhat analagous to coronavirus transcription where short primer RNA sequences hybridize to several places in the RNA genome at UCUAAAC repeats.[36] The presence of a short region of denatured DNA during transcription in *E. coli* limits the area where single stranded RNA could hybridize to. This factor and the stability of the RNA-DNA hybrid plays an important role

in ensuring transcriptional accuracy during elongation in *E. coli*.

REFERENCES

1. Blank, A., Gallant,J.A., Burgess, R.R., and Loeb, L.A. (1986) Biochem. 25, 5920–5928.
2. Rosenberger, R.F. and Hilton, J. (1983) Mol. Gen. Genet. 191, 207–212.
3. Springate, C.F. and Loeb,L.A. (1975) J. Mol. Biol. 97, 577–591.
4. Hansen, U.M. and McClure, W.R. (1980) J.Biol.Chem. 255, 9564–9570.
5. Siebenlist, U. (1979) Nature 279, 651–652.
6. Gamper,H.B. and Hearst,J.E. (1982) Cell 29, 81–90.
7. Kirkegaard,K., Buc,H., Spassky,A. and Wang, J.C. (1983) PNAS USA 80, 2544–2548.
8. Hanna, M.M. and Meares, C.F. (1983) PNAS USA 80, 4238–4242.
9. Metzger,W., Schickor,P. and Heumann,H. (1989) EMBO 8, 2745–2754.
10. Libby,R.T., Nelson,J.L., Calvo,J.M. and Gallant,J.A. (1989) EMBO, 8, 3153–3158.
11. Ahn,B. and Moss,B. (1989) J.Virol. 63, 226–232.
12. Schwer,B. and Stunnenberg,H.G. (1988) EMBO 7, 1183–1190.
13. Schwer,B., Visca,P., Vos,J.C., and Stunnenberg,H.G. (1987) Cell 50, 163–169.
14. Rose,J.K. (1980) Cell 19, 415–421.
15. McGeoch,D.J. (1979) Cell 17, 673–681.
16. Schubert,M., Keene,J.D., Herman, R.C. and Lazzarini, R.A. (1980) J. Virol. 34, 550–559.
17. Hunt,D.M., Smith,E.F. and Buckley, D.W. (1984) J.Virol. 52, 515–521.
18. Hunt,D.M., Mehta,R. and Hutchinson, K.,L. (1988) J. Gen. Virol. 69, 2555–2561.
19. Chamberlin,M. and Berg,P. (1964) J.Mol.Biol. 8, 708–726.
20. Harley,C.B., Lawrie,J., Boyer,H.W. and Hedgpeth, J. (1989) Nucl. Acids Res. 18, 547–552.
21. Weiss,R.B., Dunn,D.M., Atkins,J.F. and Gesteland,R.F. (1987) CSH Symp.Quant.Biol. 52, 687–693.
22. Christie,G.E., Farnham,P.J., and Platt, T. (1981) PNAS USA 78, 4180–4184.
23. Wu,A.M., and Platt,T. (1978) PNAS USA 75, 5442–5446.
24. Chen, E.Y. and Seeburg, P.H.(1985) DNA 4, 165.
25. Miller,J. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
26. Crowley,J.C. and Kaback,D.(1984) J.Bact. 159, 413–417.
27. Guarente,L. and Ptashne,M. (1981) PNAS USA 78, 2199–2203.
28. Rose, M., Casadaban,M.J., and Botstein,D. (1981) PNAS USA 78, 2460–2464.
29. Tautz, D.(1989) Nucl. Acids Res. 17, 6463–6471.
30. Economou, E.P., Bergen,A.W., Warren, A.C., and Antonarakis, S.E.(manuscript in prep.)
31. Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz,R.A., DeRiel,J.K., Forget,B.G., Weissman,S.M., Slightom,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.C. and Proudfoot,N.J. (1980) Cell 21, 653–668.
32. Streisinger,G., Okada,Y., Emrich,J., Newton,J., Tsugita,A., Terzaghi,E., and Inouye,M. (1966) CSH Symp. Quant. Biol. 31, 77–84.
33. Farabaugh,P.J. and Miller,J.H. (1978) J. Mol. Biol. 126, 847–863.
34. Kumar,S.A., and Krakow,J.S. (1975) J. Biol. Chem. 250, 2878–2884.
35. Chamberlin,M.J.(1965) Fed.Proc. 24, 1446–1457.
36. Lai,M.M.C., Makino,S., Soe,L.H., Shieh,C.K., Keck,J.G. and Fleming,J.O. (1987) CSH Symp. Quant. Biol. 52, 359–365.