Published in final edited form as:
J Mol Biol. 2008 October 17; 382(4): 843–858. doi:10.1016/j.jmb.2008.07.060.

DNA sequences in *gal* operon override transcription elongation blocks

Dale E. A. Lewis[1], **Natalia Komissarova**[2], **Phuoc Le**[1], **Mikhail Kashlev**[3], and **Sankar Adhya**[1]

1*Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-4264, USA*

2*Section on Microbial Genetics, Laboratory of Molecular Genetics, National Institute, of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892-2785, USA*

3*National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201, USA*

Summary

The DNA loop that represses transcription from the *gal* promoters is infrequently formed in stationary phase cells because the concentration of the loop architectural protein, HU, is significantly low at that state resulting in expression of the operon in the absence of the *gal* inducer, D-galactose. Unexpectedly, transcription from the *gal* promoters, under these conditions, overrides the physical block because of the presence of the Gal repressor (GalR) bound to an internal operator (O_I) located downstream of the promoters. We have shown here that although a stretch of pyrimidine residues (UUCU) in the RNA:DNA hybrid located immediately upstream of O_I weaken the RNA-DNA hybrid and favors RNA polymerase (RNAP) pausing and backtracking, a stretch of purines (GAGAG) in the RNA present immediately upstream of the pause sequence in the hybrid acts as an anti-pause element by stabilizing the RNA:DNA duplex and preventing backtracking. This facilitates forward translocation of RNAP including overriding of the DNA-bound GalR barrier at O_I. When the GAGAG sequence is separated from the pyrimidine sequence by a 5-bp DNA insertion, RNAP backtracking is favored from a weak hybrid to a more stable hybrid. RNAP backtracking is sensitive to Gre factors, D-galactose and antisense oligonucleotides. The ability of a native DNA sequence to override transcription elongation blocks in the *gal* operon uncovers a previously unknown way to regulate galactose metabolism in *E. coli*. It also explains the synthesis of *gal* enzymes in the absence of inducer for biosynthetic reactions.

Keywords

GalR; backtracking; roadblock; pausing; transcription

Introduction

Although gene transcription is primarily regulated at the level of initiation of RNA synthesis, transcription is also regulated at the levels of elongation and termination. The galactose (*gal*)

Address correspondence to: Dale Lewis, 37 Convent Drive, Room 5138, Bethesda, MD 20892-4264, Tel: (301) 496-9129, Fax: (301) 496-2212, Email: aed@helix.nih.gov.

operon of *Escherichia coli* is transcribed from two overlapping promoters, *P2* (transcription start point (*tsp*) as +1) and *P1* (*tsp* as +6) (Figure 1(a)), which are regulated by several transcription factors including Gal repressor (GalR). [1; 2; 3; 4; 5] GalR binding to its operators, O_E at position −55.5 and O_I at position +58.5, in the presence of the histone-like protein, HU, and supercoiled DNA causes DNA looping due to DNA-bound GalR dimer-dimer interactions and represses transcription initiation from both *P1* and *P2*. [4; 5; 6] In the absence of HU, GalR binding to O_E stimulates *P2* and represses *P1* by promoting and inhibiting open complex formation, respectively. [7; 8; 9] *In vivo*, *P2* is activated approximately 6-fold in the absence of HU.[6] Unexpectedly, GalR binding to O_I, the internal operator located within the transcribed region, does not hinder transcription elongation in the absence of DNA looping. When the DNA sequence upstream of O_I was interfered by a 5-bp insertion, GalR binding to O_I blocked about 60% of the elongating RNA polymerase (RNAP), resulting in short transcripts *in vitro* [10] (Results shown below). We studied the nature of roadblock formation by O_I-bound GalR and tested a series of mutant templates to determine what makes the wild type (wt) *gal* DNA resistant to roadblock during elongation. We found that RNAP pausing produces short transcripts transiently in the wild type but presistently in the mutant DNA. We also identified regulatory DNA sequences that potentiate or suppress RNAP backtracking, depending on their locations with respect to the nascent RNA 3' end, after encountering GalR.

Results

A 5-bp insertion in wt *gal* DNA inhibits *gal* transcription

Previously, we showed that the spatial relationship (113-bp) of both operators was critical for DNA looping repression. [10] We inserted a 5-bp (GATCT) sequence at position +37 of *P2* transcription start point (*tsp*) to change the orientation of O_I by 5-bp with respect to O_E and this prevent DNA looping. We found that *P2* was repressed in the absence of looping in this construct by an unknown mechanism. In this paper, we investigated this mode of repression. Figures 1(a) and (b) show DNA templates and result of 10 min transcription on wt *gal* template and a mutant template with a 5-bp (GATCT) sequence inserted 13-bp upstream of O_I at position +37. A strong *rho*-independent transcription terminator, t_{rpoC}, terminated *P1* and *P2* transcripts, resulting in 125- and 130-nt transcripts, respectively. [11] GalR repressed *P1* and somewhat activated *P2* in the wt DNA, as expected (Figure 1(b), lanes 1, 2). Surprisingly, in the 5-bp insertion template (Figure 1(a), ∇5), *P2* transcripts were reduced 2.8-fold by GalR (Figure 1(b), lanes 3, 4). This reduction of *P2* activity was not due to DNA looping since both operators were out of phase by 5-bp and HU was absent from the reaction. [10] We found that the repression of *P2* was accompanied by accumulation of short transcripts (lane 4). GalR blocked 60% of *P2* transcription, generating short transcripts referred to as "road-blocked transcript" (RBT). RBT was observed on both supercoiled and linear templates containing the 5-bp insertion (data not shown).

Roadblock is promoter independent

The above results demonstrated RBT formation from *P2* under conditions where *P1* was repressed (Figure 1(b)). We investigated whether GalR causes RBT from the *P1* promoter. The O_E sequence was mutated to a non-operator sequence in the *P1* template to prevent GalR from repressing *P1* (template *P1* ∇5, see Materials and Methods). To study individual promoters, we also introduced base-changes to create templates with one or the other promoter mutated, $P2^{-}P1^{+}$ *or* $P2^{+}P1^{-}$ in ∇5 (see plasmids section in Materials and Methods; Figure 1(c), lanes 1, 3). GalR blocked 76% full-length transcripts from *P2* ∇5 and 64% from *P1* ∇5 (lanes 2, 4). The difference in RBT formation in Figure 1(b) and (c) was due to a different GalR preparation used in the two reactions. In addition, we tested a DNA template in which the *gal* DNA segment containing the GATCT site was fused to a heterologous bacteriophage *T7A1* promoter at position +24 (*T7A1* ∇5) as indicated in Figure 1(a). The O_I-bound GalR blocked 63% of

T7A1 transcription (Figure 1(d)), demonstrating that roadblock formation was independent of the promoter and dependent on the insert-containing *gal* sequence from +24 to O_I. The amount of RBT formation observed in our *in vitro* conditions is unlikely to be affected by multiple rounds of transcriptions. First, only 2 rounds of transcription were observed in the absence of heparin (Figure S1, compare full-length *P2* in lanes 1 and 3). Second, almost the same amount of RBT, 89% and 83%, were observed in the presence and absence of heparin, respectively (Figure S1, lanes 2, 4).

Mapping of the 5' and 3' ends of RBT

To map the 5' end of RBT, unlabeled RBTs were isolated from an *in vitro* transcription reaction performed with the *P2* ∇5 template and used in reverse transcriptase assays with a primer annealing to the RNA region from +22 to +37 (Figure 2(a), lane 5). The same primer was used to sequence the template DNA strand, which is complimentary to the nontemplate strand (lanes 1–4). The results showed that RBT from *P2* ∇5 originated at the normal *tsp* of wt *P2* (+1).

To determine the 3' end of RBT, the terminating substrates 3'-O-methylcytosine 5'-triphosphate, 3'-O-methylguanine 5'-triphosphate and 3'-O-methyluridine 5'-triphosphate were used in a transcription reaction on the *P1* ∇5 template in the absence of GalR to generate an RNA size ladder (Figure 2(b), lanes 1–3). [11] In a separate reaction, RBT obtained on the same template in the presence of GalR (lane 4). The 3' end of RBT from *P1* ∇5 was mapped at position +47U, demonstrating that the length of RBT was 42-nt since *P1* transcribes from +6. The 3' end of RBT was 5-bp downstream from the GATCT sequence and 8-bp upstream from the O_I sequence. Identical results were obtained for *P2* ∇5. Its RBT was 47-nt long (data not shown), showing that elongating RNAPs from *P1* ∇5 and *P2* ∇5 halted at the same site. In addition, we used *T7A1* ∇5 DNA and walked histagged RNAP in the presence of a subset of 5 μM NTPs to the desire positions as indicated (Figure 2 (c) and (d)). The length of the RBT was 48-nt (Figure 2(d), lane 8). However, the *T7A1* ∇5 DNA was initiated with a tetraribonucleotide (CAUC) at position -1 and labeled at position 13 (EC13) with [α-^{32}P]CTP. Therefore, the actual length of RBT was 47-ntThe result confirmed that the 3' end of RBT was mapped at position +47U. In what follows, we report detailed studies on the mechanisms of RBT formation mostly using the*2* ∇5 DNA template.

Requirement of the GalR-O_I complex for RBT formation

Figure 1 shows that RBT formation on templates containing the 5-bp insertion required GalR. To confirm that the binding of GalR to O_I is responsible for RBT formation, the O_I sequence was mutated to a non-operator sequence to prevent GalR from binding to it. Indeed, while 64% RBT was observed on O_I^+ template (Figure 3(a), lanes 1, 2), there was no reduction of full-length transcripts from *P2* and the formation of RBT on O_I^- template (lanes 3, 4), indicating that the binding of GalR to O_I is necessary for RBT formation.

RBT formation is due to a paused RNAP complex

The formation of RBT could be due to a paused, irreversibly arrested or released RNAP. [12; 13; 14] To discriminate between these models, D-galactose that dissociates GalR from O_I was added to the roadblocked elongation complex (EC). Removal of GalR from DNA is not expected to have any effect on an irreversibly inactivated or a released complex. If the inactivation were intrinsically reversible, GalR removal from O_I would cause the paused EC to resume elongation and synthesize full-length transcripts. Figure 3(b) shows the amount of RBT before and after D-galactose addition. RBT was observed at 0.7 min after transcription initiation and its amount gradually increasing thereafter (lanes 1–8). When D-galactose was added to the reaction 10 min after initiation of transcription (lanes 9–20), the amount of RBT decreased and the amount of full-length *P2* transcripts increased, indicating a paused complex. Interestingly, RNAP could read-through O_I-bound GalR even in the absence of D-galactose

over extended time periods (data not shown), perhaps because of spontaneous dissociation of GalR from O_I.

Enhancement of an intrinsic pause helps RBT formation

To understand why RBT was detected only on the templates containing the GATCT insert, we compared the kinetic behavior of RNAP as it approaches O_I-bound GalR on the *P2* and *P2* ∇*5* templates. Figure 3(c) shows that at 30 sec the same fraction of RBT were made on both templates. On the wild type template, the RBT fraction diminished rapidly with a half-life of 5 min, being converted to full-length transcripts. In contrast, on the *P2*∇5 template, the amount of RBT decreased slowly and was detectable even after 90 min of incubation. We observed that RNAP tends to pause 8-bp upstream from O_I both in *P2* and *P2*∇5 templates at the $^{+44}$-UUCU-$^{+47}$ RNA sequence. Whereas, the pause is very brief in *P2*, RNAP needs longer time to overcome the pause in *P2*∇5. This suggests that certain element(s) of the wild type sequence which help RNAP escapes from the pause state is altered by the 5-bp insertion.

Next, we compared the kinetics of transcription on the two templates in the absence of GalR (Figure 3(d)). To enhance pausing, we limited the concentration of GTP, the next base after the pause site $^{+44}$-UUCU**G**-$^{+48}$, to 1 μM, but kept the other NTPs at 100 μM. The paused transcripts occurred at the same position on the templates where RBT were formed in the presence of GalR. Furthermore, the pause was much longer in the *P2* ∇*5* template than in *P2* template (lanes 2–7) in agreement with the results shown in Figure 3(c), confirming that an intrinsic pause site in *gal* DNA was intensified by the 5-bp insertion. The presence of GalR at O_I makes the latter pause generate much higher level of RBT as shown in Figure 1(b).

Antisense oligonucleotides suppress roadblock

It was shown previously that when a Lac operator was placed at an internal position of a transcription unit, Lac repressor bound to it, stopped elongating RNAP and caused reversible backtracking of the enzyme. [15] During RNAP backtracking after a pause, the catalytic center of RNAP disengages from the 3' end of the transcript causing inactivation of EC. The enzyme, together with the transcription bubble and the RNA:DNA hybrid, relocates upstream along the DNA and RNA. [16] The backtracked complex is transcriptionally inactive (arrested complex). Such an arrested complex is reactivated either spontaneously by the return of RNAP to the original position or by the action of GreB protein (see below). [16; 17] It was shown previously that a U-rich RNA sequence, which forms a weak RNA:DNA hybrid, favors backtracking. [17; 18; 19]

In the active complex, but not in the backtracked one, the rear edge of RNAP extends to about 14-nt of RNA upstream of the 3' end of RNA. [20] The annealing of short oligonucleotides to nascent RNA 14–16 nt from the 3' end of the transcripts prevents RNAP from backtracking and stabilizes the 3' end of the transcript in the catalytic center of RNAP. [15; 16; 18] This action of oligonucleotides suppresses RBT if the latter is indeed caused by backtracking.

Figure 4(a) shows the results of transcription on *P2*∇5 template in the presence of GalR and antisense oligonucleotides. Oligo-24, which anneal to the nascent RNA at a distance of 24-nt upstream from the 3' end of RBT, had no effect on roadblock formation (lanes 2,3). Oligo-18 had marginal effect (lane 4). Oligo-12, which anneals to the RNA immediately behind RNAP, suppressed RBT (lane 5), indicating that RBT in *gal* may originate from RNAP backtracking. Oligo-6 had no effect on RBT probably because it did not hybridize to the RNA since its target sequence was covered by RNAP (lane 6). Another factor potentially capable of preventing backtracking, a strong secondary structure formed by nascent RNA immediately behind RNAP, was not observed in the *gal* sequence; the mfold program did not generate any stable structure

for *gal* RNA from +1 to +37-nt, which represents the region of the transcript upstream of RNAP at the pause site encountering GalR. [21]

GreA and GreB effect on RBT

GreA and GreB are transcription factors known to influence elongation by inducing an endoribonuclease activity of RNAP catalytic center. [22; 23; 24; 25; 26; 27] GreB cleaves off both short (2-nt) and long (up to 24-nt) 3' proximal RNA fragments in backtracked ECs since in these complexes the catalytic center of the enzyme is aligned with an internal position of the RNA. [23; 28] The cleavage rescues arrested complexes by generating a new 3' RNA end in the catalytic center. [22; 23; 29] Unlike GreB, GreA cleaves off only short (2–3 nt) RNA fragments and only in active ECs that briefly travel in the reverse direction by 2–3 nt. To test whether GreA and/or GreB affect RBT, the factors were added at 50 nM concentration to the transcription assays before RBT was formed. Figure 4(b) shows that GreA suppressed RBT formation slightly in the *P2*∇5 template (lanes 2, 3) while GreB suppressed it completely by converting RBT to full-length transcripts (lanes 2, 4). Since the dissociation constant (Kd) for GreA is in the range of 800–1000 nM, and that for GreB is approximately 80–100 nM, [25 30; 31; 32] we also used 800 nM GreA and 80 nM GreB in other reactions to test whether the low level of GreA (50 nM) was responsibled for the partial suppression of RBT. We obtained similar result (data not shown). The sensitivity of the nascent RNA to GreB and partial resistance to GreA cleavage strongly supports the idea that RNAP backtracks along the DNA after it encounters the O_I-bound GalR.

To determine the distance of RNAP backtracking, the roadblocked EC at *P2*∇5 was purified from NTPs by using RNA spin columns before the addition of Gre factors. The removal of NTPs prevents elongation of the cleaved products. While GreA did not show much RBT cleavage (Figure 4(c), lane 3) GreB generated a 43-nt RNA by cleaving 4 nts from the RBT of *P2* ∇5 (lane 4). The RBT of *P1*∇5 (42-nt) was used as a marker to map the length of the cleaved products (lane 5). In some of these experiments, 40–42 nts products were observed (data not shown). Truncation of RBT by 4 and 7 nts shows that RNAP catalytic center relocates 4- or 7-nt upstream from the original site. In this experiment, GreA did not cleave RBT because RNAP, after 15 min incubation, backtracked by 4–7 nts thus making the RNA insensitive to GreA.

The transcription bubble in the roadblocked complex

We mapped the "transcription bubble" in the roadblocked EC by using potassium permanganate ($KMnO_4$), which oxidizes distorted, unpaired, thymidine nucleotides. [33; 34] Taq DNA polymerase is unable to read through the oxidized nucleotides, which allows detecting them in the modified templates by PCR amplification. [33] $KMnO_4$ was added to the *P2*∇5 DNA in the presence of GalR and RNAP with or without NTPs. In the absence of NTPs, −2 to +4 region on the nontemplate strand and −11 to −8 region on the template strand were modified by $KMnO_4$ revealing open complex formation (Figure 5, lanes 1, 7). In the presence of NTPs, the bubble translocated downstream from the initiation site toward the termination site as shown by the reduction of the signal around the −11 to +4 region (lanes 2, 8). Incidentally, the residual open complexes found at the promoter region in the presence of NTPs may reflect the nonproductive complexes described previously. [35] $KMnO_4$-sensitive sites were found from +30 to +42 on the nontemplate strand (lane 2) and from +28 to +32 on the template strand (lane 8). The 3' end of RBT is located at position +47 on the *P2* ∇5 transcription unit. Given the location of the EC, the transcription bubble with a typical size of 12–13 nt [36] would have spread between +35 and +48 positions with a strong $KMnO_4$ cleavage signal from T residues in the +44 to + 47 segment of the nontemplate strand. We did not detect such a signal (lane 2), but observed a 15-nt long transcription bubble spreading between +28 and +42 positions. The increased length and upstream boundary of the transcription bubble is in agreement with the

idea that in the roadblocked complex, RNAP backtracks by up to 7-nt. We think that in the roadblocked complex, RNAP oscillates between the stalled position and a location 4- or 7-nt upstream. Oscillation by a backtracked RNAP has been reported previously. [17]

Presence of anti-pause sequence element in *gal* DNA

As discussed above, the 5-bp insertion alters an intrinsic anti-pause signal encoded in wt *gal* sequence, around position +37. This model predicts that other changes in this region (anti-pause DNA element) would also generate RBT. To test this hypothesis, we investigated whether deleting 5-bp (from + 36 to +41) of the wt DNA (*P2*Δ5) enhances RBT formation as *P2*∇5 does. Whereas the full-length transcripts were unaffected by GalR with no observation of RBT in the wild type *P2* template (Figure 6(a), lanes 1, 2), 47% of transcription was blocked by O_I-bound GalR in the *P2*Δ5 DNA (lanes 3, 4). The RBT observed in *P2*Δ5 DNA was 10-nt shorter than that observed in *P2*∇5 DNA (data not shown), indicating that RNAP paused at the identical position in both cases. This suggests that it is not the sequence of the 5-bp insertion or deletion, but rather an alteration of an "anti-pause" sequence encoded around the +37 region in wild type *gal* that is responsible for RBT formation by an arresting RNAP. In addition, the specific sequence around position +37 instead of RNA length makes RNAP pause on the insertion-containing templates.

Location of the anti-pause element

To map the element(s) in the wt *gal* that prevents arrest of RNAP, we scanned the DNA by inserting the 5-bp (GATCT) sequence at five other positions (+14, +28, +32, +35 and +47) (Figure 6(b)). In the presence of GalR, (i) insertions at +14 (∇5A) and +28 (∇5B) sites did not generate any RBT (lanes 1–6), (ii) insertions at +32 (∇5C) and +35 (∇5D) generated 11% of RBT (lanes 7–10), and (iii) insertions at +37 (∇5) and +47 (∇5F) generated 42% and 54% of RBT, respectively (lanes 11–14). These data show that the insertion has to be in close proximity to O_I in order to generate substantial amount of RBT. Interestingly, RBT observed with the insertion at +47 was 1–2 nt longer than that observed with the insertion at +37 (lanes 12 vs 14). Perhaps, the sequence context around the pause site determines the exact 3' end of RBT. In the +37 (∇5) case, the 3' end is located at position +47 (∇5F) between T and G (gagagatctgttc**tg**g); however in the +47 case, the 3' end is located at position +48 between G and A or at position +49 between A and T (gagagttctggtta**gat**ctcc).

To further localize the proposed anti-pause signal of wild type *gal*, we inserted a TGAGA sequence at the same +37 site of *P2* (Figure 7(a), designated *P2 ∇5G*). The new template, despite the insertion, preserved the same 10-bp sequence ($^{+37}$-TGAGAGTTCT-$^{+47}$) upstream of the RBT 3' end as in the wild type template; the sequence farther upstream was different. The TGAGA insertion shows significantly less RBT (16%) compared to the GATCT insertion (62% RBT) (Figure 7(b), lanes 2, 4), showing that the 10-bp sequence favors RNA polymerase elongation. We also tested the effect of a 10-bp sequence (AATTATGAGA) at position +37 (*P2 ∇10C*), which preserves the 15-bp sequence upstream of the RBT 3' end ($^{+37}$-AATTATGAGAGTTCT-$^{+52}$). The latter sequence completely prevented RBT formation (lanes 5, 6). These results indicate that the anti-pause signal is located within the sequence 15-bp upstream of RBT 3' end.

Role of RNA:DNA hybrid in RNAP pausing

We investigated the mechanism by which an anti-pause element identified above prevents RBT by applying a computational approach. Our objective was to understand how the presence of the anti-pause element makes RNAP favors transcription elongation or the absence of it makes RNAP favors pausing (backtracking) at the UUCU pause site. RNA:DNA hybrid is the major determinant of lateral stability of the EC that determines forward (elongation) or backward (backtracking) movement of RNAP. [16; 18] The hybrid is believed to be 8–9 bp long. [37; 38]

For a number of templates (Table 1), we calculated the predicted $\Delta G°_{37}$ of 9-bp long RNA:DNA hybrid [39] in the EC at the pause site and the site backtracked by 4- and 7-nt. In wild type *gal* sequence, ΔG of RNA:DNA hybrid in the EC at the pause site is −6.8 kcal mol^{-1}, and in the complex backtracked by 4- and 7-nt, ΔG is −6.4 and −4.7 kcal mol^{-1}, respectively (Table 1, line 1). This difference in energy makes backtracking by 4- or 7-nt unfavorable. In the GATCT ($\overline{/5}$) insertion, ΔG of RNA:DNA hybrid at the pause site is −5.5 kcal mol^{-1}, and in the complex backtracked by 4-nt and 7-nt, the corresponding ΔG becomes −8.0 and −7.8 kcal mol^{-1} , respectively (line 11). The latter energy distribution makes backtracking by 4- or 7-nt favorable in $\overline{/5}$. Consistently, we observed no RBT in the first case, and 60% RBT in the second case. Table 1 shows that with a few exceptions (lines 19–22), when the RNA:DNA hybrid is more stable at the pause site than at the backtracked site, elongation is favored with RNAP overcoming the roadblock as well (lines 1–6). Conversely, when the hybrid is less stable at the pause site than at the backtracked site, backtracking and RBT formation are favored (lines 7–18).

It appears that a purine rich (GAGAG) sequence is an important factor in determining the stability of the hybrid and therefore the location of the RNAP. The GATCT insertion at +37 in effect shifts the GAGAG sequence 5-bp upstream and makes it a part of the hybrid in the backtracked EC. Next we changed the Gs to Ts in this sequence to decrease the stability of the upstream hybrid and study their effects on RBT (Figures 7(c); Table 1, lines 6, 9, 10). The RBT decreased from 57% to 33% on the 34T template (Figures 7(c), lanes 1–4; Table 1, line 10), to 20% on the 36T template (lanes 5, 6; line 9), and was eliminated on the 34T36T template (lanes 7, 8; line 6). This result is consistent with the idea that the stability of RNA:DNA hybrid determines whether or not RNAP reads through the GalR roadblock.

Discussion

In the absence of an inducer, the *gal* operon is partially derepressed when the cells are in stationary phase because of a 4-fold reduction of HU levels that decreases the chance of DNA looping. [40] In the absence of HU, *P2* transcripts are derepressed 6-fold *in vivo*.[6] Previous studies showed that a stable O_I-bound GalR complex does not block an elongating RNAP in *gal*. [6; 8] Since the enzymes of the *gal* operon are needed for biosynthetic gylcosylation reactions, it is important that *gal* enzymes are made in the absence of D-galactose, and O_I-bound GalR complex does not block transcription elongation at the internal operator. The latter results were unexpected given that in several cases a strong DNA binding protein placed in the transcribed region produces RBT. [41] Our current results confirmed that the O_I-bound GalR does not block transcription from the *gal* promoters *in vitro*. Our finding that an intrinsic sequence arrangement in *gal* overcomes the barrier explains the mechanism of *gal* transcription in the absence of inducer during stationary phase. Although encounter with GalR causes a detectable pause of EC on wild type *gal* DNA, transcription is not blocked. The ability of wild type sequence to support transcription through DNA-bound GalR is easily compromised by mutations immediate upstream of O_I leading to the formation of stable roadblocked complex.

RBT formation on mutant templates containing a 5-bp insertion (13-nt upstream of O_I) is inhibited by GreB and partially by GreA, as well as by oligonucleotides complementary to the transcript upstream of RNAP. These results, together with $KMnO_4$ footprinting of the transcription bubble, show that the EC roadblocked by GalR backtracks along the RNA and DNA by 4- to 7-nt, carrying the catalytic center of the enzyme to an internal position of the transcript, and inhibiting RNA elongation. Occasionally, RNAP returns to the original location and can regain its ability to transcribe DNA when GalR is either removed by D-galactose or spontaneously dissociated from O_I. In the roadblocked complex, RNAP stops just 8-bp upstream of O_I. This is a remarkably short distance, since, according to hydroxyl radical footprinting and to X-ray crystallography of the EC, RNAP covers at least 13-bp of DNA

downstream from the catalytic center. [36; 42] GalR was shown to cover an extra 3-bp on either side of the O_I sequence, [4; 43] suggesting that RNAP transcribes at least 8-bp against an opposing force. Using yeast RNAP II as a model, it was recently shown that the return of the backtracked complex to the 3' end of the transcript can be prevented by a lower opposing force compared to the force required to stop RNAP during normal elongation. [44] The addition of the TFIIS cleavage factor, a GreB homologue, substantially increased the ability of the enzyme to transcribe against the applied force. [44] These results argue that the ability of RNAP to spontaneously overcome an obstacle is impaired. This correlates with our observation that although RNAP initially pauses in both wild type and ▽5 templates as it collides with GalR, it is the ability to escape from the pause in the wild type that can be impaired by mutations describes above.

We argue that the relative strength of the regional RNA:DNA hybrid affects the fate of RNAP in wild type and mutant templates. In both templates, four 3' proximal nucleotides in RNA (UUCU) weakly pair with the template thereby favoring backtracking (Figure 8). Since rPy:dPu base pairs are weaker than dPy:dPu base pairs, [45] the re-formation of DNA duplex in front of a backtracked RNAP is favorable at the UUCU pause sequence. These properties of 3'-proximal sequence favor pausing of RNA polymerase in both wild type and ▽5 templates (Figure 8). However, in wild type DNA, these properties are counteracted by strong base pairing of the purine rich (GAGAG) sequence in this upstream part of the RNA:DNA hybrid before backtracking, and by weak base pairing of the AAUUAU sequence (rAU) in the upstream part of the backtracked RNA:DNA hybrid (Figure 8, a1 and a2). Note, that the GAGAG sequence produces rPu:dPy pairs, which are stronger than dPu:dPy. [45] We believe that these properties of the wild type sequence make the backtracked location of the RNA-DNA hybrid unfavorable for RNAP, and help RNAP escape the pause. The inserted GATCT sequence weakens the upstream part of RNA:DNA hybrid at the original location of the EC and re-locates the strong GAGAG containing hybrid in the upstream segment of the backtracked hybrid making the latter location preferable for RNAP in the mutant template (Figure 8, b1 and b2). Therefore, in the ▽5 template RNA polymerase stays longer at the upstream location in the absence of GalR and cannot escape the roadblock in the presence of GalR bound to O_I.

In summary, we have identified a new anti-pause element, GAGAG, located 12-bp upstream of O_I, which has a dual role in modulating RNAP elongation. When the GA-rich sequence is located immediately upstream of the pause site, UUCU, it stabilizes the RNA:DNA hybrid, favors elongation and overrides operator bound GalR. In contrast, when the GA-rich is shifted 5-bp upstream of the pause site, elongating RNAP pause and backtrack. The GA-rich sequence stabilizes the backtracked hybrid causing failure of RNAP to override the operator-bound GalR.

Materials and Methods

Reagents

All restriction endonucleases and alkaline phosphatase were purchased from New England Biolabs; T4 DNA ligase and Max efficiency DH5α™ competent cells from Invitrogen; Recombinant RNasin® Ribonuclease Inhibitor (40 U/μl) from Promega; SequaGel sequencing system from National Diagnostics; nucleotide triphosphates (NTPs) (100 mM) and 3'-O-methylguanosine 5'-triphosphate from Amersham Pharmacia Biotech Inc.; primers from BioServe Biotechnologies and Sigma Genosys; antisense oligonucleotides from Oligos Etc.; XL PCR and DNA sequencing kits from Applied Biosystem; [α-^{32}P]UTP and [α-^{32}P]CTP (specific activity = 3000 Ci/mmol, 10 μCi/μl) and [γ-^{32}P]ATP (specific activity = 7000 Ci/mmol, 167 μCi/μl) from MP Biomedicals, formerly ICN Biochemicals, Inc., [α-^{32}P]GTP (specific activity = 3000 Ci/mmol, 10 μCi/μl) from GE Healthcare.

Other proteins and strains

E. coli RNAP (specific activity: 1.85 – 2.5 × 10^3U/mg, concentration: 1U/μl) was from USB Corporation. GreA and GreB were gifts from Sergei Borukhov (UMDNJ, Stratford, NJ). GalR and CRP were purified as described. [46; 47]

Plasmids

The plasmids used in this study are listed in Table 2. The wt plasmid pSA850 ($O_E^{G+}P2^+P1^+O_I^{G+}$), which contains the *gal* operator/promoter region from −70 to + 96, was used as a PCR template to construct plasmids. [10; 11] Plasmid pSA859 [$O_E^{G+}P2^+P1^+O_I^{G+}$(+5−bp)], *wt ∇5*, contains a 5-bp (GATCT) insertion at position +37 with respect to the *tsp* of *P2* (+1). Plasmid pDL225 [$O_E^{G+}P2^+P1^-O_I^{G+}$], *P2*, contains two mutations (a G to A change at position −9 and a C to A change at position −8), which convert the −10 element of *P2* to the consensus −10 element of σ^{70} promoter (TATGCT to TATAAT). Such changes inactivate *P1* by destroying the extended −10 of *P1*, resulting in maximal expression of *P2*. [48] Plasmid pDL1004 (*P2 ∇5*) is a derivative of pDL225 with a 5-bp (GATCT) insertion at position +37. Most plasmids used in this study are derivatives of pDL1004. Plasmid pDL994 [$O_E^{N}P2^-P1^+O_I^{G+}$(+5−bp)], or *P1 ∇5*, contains several mutations: a single base pair change from −14G to T inactivates *P2* by destroying the extended −10 of *P2*; [49; 50] a double base pair change from −4G and −3G to A's, which convert the −10 element of *P1* to the consensus −10 element of σ^{70} promoter, results in maximal expression of *P1*. Finally, the O_E (GTGTAAACGATTCCAC) is replaced by a non-operator sequence (CACTATGGCGAACGTC). Plasmid pDL571, or *T7A1 ∇5*, contains *T7* bacteriophage A1 promoter DNA (−56 to +24) fused with *gal* wt∇5 DNA downstream from +24. The plasmids were sequenced on an ABI Prism 310 Genetic Analyzer to verify the intended base-pair substitutions, insertions or deletions.

In vitro transcription assays

In vitro transcription reactions were performed as described. [11] Briefly, supercoiled DNA templates (2–4 nM) and RNAP (20 nM) were preincubated at 37°C for 5 min in transcription buffer (20 mM Tris acetate, 10 mM magnesium acetate, 200 mM potassium glutamate) supplemented with 1 mM DTT, 1 mM ATP, 0.8 U rRNasin® in a total reaction volume of 50 μl. When used, 200 nM GalR was present unless indicated otherwise. To initiate transcription, NTPs were added to a final concentration of 0.1 mM GTP, 0.1 mM CTP, 0.01 mM UTP and 5 μCi [α-^{32}P]UTP (3000 Ci/mmol) unless indicated otherwise. The reactions were incubated for an additional 10 min before terminated by the addition of an equal volume (50 μl) of loading dye (90% formamide, 10 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue). Transcripts were separated on 8% polyacrylamide sequencing gels. The *RNAI* transcripts driven by GalR independent promoter present in the plasmids (108-nt) were used as internal controls to quantify the relative amount of *gal* transcripts. [51] The *P1* and *P2* full-length transcripts contain 36 and 39 U residues, respectively. The roadblock transcripts (RBT) from *P1* and *P2* contain 12 and 15Us, respectively. In mutant templates, number of Us depended on the sequence. To calculate the percentage of RBT that was blocked by O_I-bound GalR, full-length and RBT were normalized to the *RNAI* transcripts and corrected for U content according to the formula %RBT = B/(A+B) × 100, where A=area of full-length/(area of RNAI × number of Us in full-length) and B= area of RBT/(area of RNAI × number of Us in RBT). When GTP was used as a labeled nucleotide (experiments of Figure 3(c) & (d)), RBT was corrected for G content.

Use of antisense oligonucleotides

In vitro transcription assays were conducted as described except that 50 μM oligonucleotides were present during 5 min preincubation at 37°C before NTPs were added to initiate transcription.

GreA and GreB induced cleavage

To study the effect of GreA and GreB on RBT, the transcription assays were performed in the presence of GalR and Gre factors. GreA and GreB were used in the amounts discussed in the text.

To study the cleavage pattern of nascent transcripts by RNAP in the presence of GreA or GreB, NTPs were removed from the roadblock reactions by using RNA mini quick spin columns (Roche Applied Science). Next, GreA or GreB was added as described above and the cleavage pattern of the RNA was detected on a 24 % sequencing gel (19:1).

Mapping of the 5' ends of RBT

In vitro transcription assays were carried out as described above using cold UTP for the experimental reactions and labeled [α-^{32}P]UTP for the control experiments. Unlabeled RBT were excised from a 10% sequencing gel using labeled RBT as a marker. The RNA was eluted from the gel slice as outlined for the Electro-elutor (Model 422-BioRad). The eluted RNA was concentrated on a speed vacuum centrifuge. Next, an equal volume of phenol/chloroform/ isoamyl alcohol (25:24:1) was added to the RNA. The aqueous layer (RNA) was removed and loaded on an RNA mini quick spin column. Primer extension of the RNA was performed according to the protocol for AMV reverse transcriptase (Promega) using [γ-^{32}P] labeled primer (5'-TCTCATAATTCGCTCC-3'), which mapped from +37 to +22. The DNA was sequenced by PCR using the same primer according to the protocol for *fmol*® DNA cycle sequencing (Promega).

Mapping of the 3' ends of RBT with 3-O-methylguanosine

RBT of *P1* was obtained as described on supercoiled template in the presence of GalR. Linear DNA template (355-bp) generated by PCR from *P1*∇5 plasmid was used to map the 3' ends of RBT as described. [11] Cyclic AMP receptor protein (CRP, 50 nM) and cAMP (100 μM) were added to the transcription reaction to maximize the strength of *P1* promoter (in the absence of cAMP-CRP complex, *P1* activity on linear template is very weak). To generate a ladder of nucleotides, transcription was carried out for 10 min in the presence of 0.25 μM 3-O-methylcytosine-5'-triphosphate, 3-O-methylguanosine-5'-triphosphate or 3-O-methyluridine-5'-triphosphate containing 1.0 mM ATP, 0.01 mM UTP and 0.1 mM GTP and CTP. Products of the reactions were separated on a 12% polyacrylamide sequencing gel.

The 3' end of RBT was also mapped by mobilizing histagged RNAP (0.12 μg) on Ni-NTA-agarose and walking along the DNA (1 μg) as described[16; 17]. Stable complex was stalled at position 12 (EC12) after initiating with 1mM CUAU (a tetraribonucleotide primer), 50 μM ATP and GTP at 37°C for 5 min. Complexes were washed 3–5 times in 1 ml transcription buffer, TB (20 mM Tris-HCl, pH 7.9, 40 mM KC, 5 mM $MgCl_2$ and 1 mM β-mercaptoethanol. The RNA was labeled at EC13 with 3 μl of [α-^{32}P]CTP (3000 Ci/mmol) and incubated at room temperature for 5 min. The complexes were washed as above and a subset of NTPs (5 μM) was added to walk RNAP to the desire position; the process of walking and washing were repeated.

Mapping of transcription bubble by $KMnO_4$ footprinting

The sequences of wt and *P2*∇5DNA templates for the potassium permanganate ($KMnO_4$) assay were determined according to the *fmol*® DNA cycle sequencing system protocol (Promega). The non-template stand was sequenced using [γ-^{32}P]XbaI-5 primer (5'-TCAACGGAGCTCGTCG– 3'), which mapped from −108 to −93. The template strand was sequenced using [γ-^{32}P]BamHI-10 primer (5'-GCGGATCCCTAAACTC-3'), which mapped from +161 to +141. Approximately 10 μg/μl DNA was sequenced using "profile 1" from *fmol*® DNA cycle sequencing system protocol. The $KMnO_4$ footprinting was performed as described. [33; 52] *In vitro* transcription assays were performed as described above with 10 nM supercoiled DNA. At the end of the transcription reactions, 10 mM $KMnO_4$ was added for 2–3 min at 37°C. The reactions were quenched with β-mercaptoethanol (0.5 M). An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added to the reactions, which were heated at 80°C for 2 min before chilling on ice. Next, the reactions were centrifuged for 1 min before the aqueous DNA layers were transferred to DNA mini quick spin columns. For primer extension assay, 18 μl DNA, 20 μl PCR Master mix (Promega) and 2 μl (2 pmoles) [γ-^{32}P] BamHI-10 or [γ-^{32}P]XbaI-5 primers were used to amplify the product using the *fmol*® DNA cycle sequencing. The products were analyzed on an 8% sequencing gel.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Cancer Institute and the Center for Cancer Research. We thank S. Borukhov, M. Liu, M. Soukhodolets, and T. Soares for the gift of purified proteins, and V. Zhurkin for critical discussions.

References

1. Adhya S, Miller W. Modulation of the two promoters of the galactose operon of Escherichia coli. Nature 1979;279:492–494. [PubMed: 221830]
2. Musso RE, Di Lauro R, Adhya S, de Crombrugghe B. Dual control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. Cell 1977;12:847–854. [PubMed: 200371]
3. Aiba H, Adhya S, de Crombrugghe B. Evidence for two functional gal promoters in intact Escherichia coli cells. J Biol Chem 1981;256:11905–11910. [PubMed: 6271763]
4. Aki T, Adhya S. Repressor induced site-specific binding of HU for transcriptional regulation. Embo J 1997;16:3666–3674. [PubMed: 9218807]
5. Aki T, Choy HE, Adhya S. Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. Genes Cells 1996;1:179–188. [PubMed: 9140062]
6. Lewis DE, Geanacopoulos M, Adhya S. Role of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at gal promoters in Escherichia coli. Mol Microbiol 1999;31:451–461. [PubMed: 10027963]
7. Goodrich JA, McClure WR. Regulation of open complex formation at the Escherichia coli galactose operon promoters. Simultaneous interaction of RNA polymerase, gal repressor and CAP/cAMP. J Mol Biol 1992;224:15–29. [PubMed: 1312605]
8. Choy HE, Park SW, Aki T, Parrack P, Fujita N, Ishihama A, Adhya S. Repression and activation of transcription by Gal and Lac repressors: involvement of alpha subunit of RNA polymerase. Embo J 1995;14:4523–4529. [PubMed: 7556095]
9. Roy S, Garges S, Adhya S. Activation and repression of transcription by differential contact: two sides of a coin. J Biol Chem 1998;273:14059–14062. [PubMed: 9603899]

10. Lewis DE, Adhya S. In vitro repression of the gal promoters by GalR and HU depends on the proper helical phasing of the two operators. J Biol Chem 2002;277:2498–2504. [PubMed: 11700313]
11. Lewis DE. Identification of promoters of Escherichia coli and phage in transcription section plasmid pSA850. Methods Enzymol 2003;370:618–645. [PubMed: 14712680]
12. Landick R. RNA polymerase slides home: pause and termination site recognition. Cell 1997;88:741–744. [PubMed: 9118216]
13. Landick R. RNA polymerase clamps down. Cell 2001;105:567–570. [PubMed: 11389826]
14. Mooney RA, Artsimovitch I, Landick R. Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. J Bacteriol 1998;180:3265–3275. [PubMed: 9642176]
15. Toulme F, Guerin M, Robichon N, Leng M, Rahmouni AR. In vivo evidence for back and forth oscillations of the transcription elongation complex. Embo J 1999;18:5052–5060. [PubMed: 10487757]
16. Komissarova N, Kashlev M. Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc Natl Acad Sci U S A 1997;94:1755–1760. [PubMed: 9050851]
17. Komissarova N, Kashlev M. RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. J Biol Chem 1997;272:15329–15338. [PubMed: 9182561]
18. Palangat M, Landick R. Roles of RNA:DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II. J Mol Biol 2001;311:265–282. [PubMed: 11478860]
19. Velikodvorskaya T, Natalia Komissarova N, Sen R, King RA, Banik-Maiti S, Weisberg RA. Inhibition of a Transcriptional Pause by RNA Anchoring. 2008
20. Komissarova N, Kashlev M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. Proc Natl Acad Sci U S A 1998;95:14699–14704. [PubMed: 9843952]
21. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003;31:3406–3415. [PubMed: 12824337]
22. Borukhov S, Polyakov A, Nikiforov V, Goldfarb A. GreA protein: a transcription elongation factor from Escherichia coli. Proc Natl Acad Sci U S A 1992;89:8899–8902. [PubMed: 1384037]
23. Borukhov S, Sagitov V, Goldfarb A. Transcript cleavage factors from E. coli. Cell 1993;72:459–466. [PubMed: 8431948]
24. Orlova M, Newlands J, Das A, Goldfarb A, Borukhov S. Intrinsic transcript cleavage activity of RNA polymerase. Proc Natl Acad Sci U S A 1995;92:4596–4600. [PubMed: 7538676]
25. Koulich D, Orlova M, Malhotra A, Sali A, Darst SA, Borukhov S. Domain organization of Escherichia coli transcript cleavage factors GreA and GreB. J Biol Chem 1997;272:7201–7210. [PubMed: 9054416]
26. Sparkowski J, DAS A. The Nucleotide-Sequence of Grea, a Suppressor Gene That Restores Growth of an Escherichia-Coli Rna-Polymerase Mutant at High-Temperature. Nucleic Acids Research 1990;18:6443–6443. [PubMed: 2243801]
27. Sparkowski J, DAS A. Location of a New Gene, Grea, on the Escherichia-Coli Chromosome. Journal of Bacteriology 1991;173:5256–5257. [PubMed: 1885509]
28. Lee DN, Feng G, Landick R. GreA-induced transcript cleavage is accompanied by reverse translocation to a different transcription complex conformation. J Biol Chem 1994;269:22295–22303. [PubMed: 8071356]
29. Toulme F, Mosrin-Huaman C, Sparkowski J, Das A, Leng M, Rahmouni AR. GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. Embo J 2000;19:6853–6859. [PubMed: 11118220]
30. Borukhov S, Laptenko O, Lee J. Escherichia coli transcript cleavage factors GreA and GreB: functions and mechanisms of action. Methods Enzymol 2001;342:64–76. [PubMed: 11586920]
31. Rutherford ST, Lemke JJ, Vrentas CE, Gaal T, Ross W, Gourse RL. Effects of DksA, GreA, and GreB on transcription initiation: insights into the mechanisms of factors that bind in the secondary channel of RNA polymerase. J Mol Biol 2007;366:1243–1257. [PubMed: 17207814]

32. Loizos N, Darst SA. Mapping interactions of Escherichia coli GreB with RNA polymerase and ternary elongation complexes. J Biol Chem 1999;274:23378–23386. [PubMed: 10438515]
33. Sasse-Dwight S, Gralla JD. KMnO4 as a probe for lac promoter DNA melting and mechanism in vivo. J Biol Chem 1989;264:8074–8081. [PubMed: 2722774]
34. Kashlev M, Komissarova N. Transcription termination: primary intermediates and secondary adducts. J Biol Chem 2002;277:14501–14508. [PubMed: 11856750]
35. Kubori T, Shimamoto N. A branched pathway in the early stage of transcription by Escherichia coli RNA polymerase. J Mol Biol 1996;256:449–457. [PubMed: 8604130]
36. Zaychikov E, Denissova L, Heumann H. Translocation of the Escherichia coli transcription complex observed in the registers 11 to 20: "jumping" of RNA polymerase and asymmetric expansion and contraction of the "transcription bubble". Proc Natl Acad Sci U S A 1995;92:1739–1743. [PubMed: 7878051]
37. Nudler E, Mustaev A, Lukhtanov E, Goldfarb A. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell 1997;89:33–41. [PubMed: 9094712]
38. Sidorenkov I, Komissarova N, Kashlev M. Crucial role of the RNA:DNA hybrid in the processivity of transcription. Mol Cell 1998;2:55–64. [PubMed: 9702191]
39. Sugimoto N, Nakano S, Katoh M, Matsumura A, Nakamuta H, Ohmichi T, Yoneyama M, Sasaki M. Thermodynamic Parameters to Predict Stability of Rna/DNA Hybrid Duplexes. Biochemistry 1995;34:11211–11216. [PubMed: 7545436]
40. Azam TA, Iwata A, Nishimura A, Ueda S, Ishihama A. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. Journal of Bacteriology 1999;181:6361–6370. [PubMed: 10515926]
41. Deuschle U, Gentz R, Bujard H. Lac Repressor Blocks Transcribing Rna-Polymerase and Terminates Transcription. Proceedings of the National Academy of Sciences of the United States of America 1986;83:4134–4137. [PubMed: 3520567]
42. Vassylyev DG, Vassylyeva MN, Zhang JW, Palangat M, Artsimovitch I, Landick R. Structural basis for substrate loading in bacterial RNA polymerase. Nature 2007;448163-U4
43. Majumdar A, Adhya S. Probing the structure of gal operator-repressor complexes. Conformation change in DNA. J Biol Chem 1987;262:13258–13262. [PubMed: 3308875]
44. Galburt EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M, Bustamante C. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. Nature 2007;446:820–823. [PubMed: 17361130]
45. Hung SH, Yu Q, Gray DM, Ratliff RL. Evidence from Cd-Spectra That D(Purine)Center-Dot-R (Pyrimidine) and R(Purine)Center-Dot-D(Pyrimidine) Hybrids Are in Different Structural Classes. Nucleic Acids Research 1994;22:4326–4334. [PubMed: 7937162]
46. Majumdar A, Rudikoff S, Adhya S. Purification and properties of Gal repressor:pL-galR fusion in pKC31 plasmid vector. J Biol Chem 1987;262:2326–2331. [PubMed: 2950087]
47. Ryu S, Kim J, Adhya S, Garges S. Pivotal role of amino acid at position 138 in the allosteric hinge reorientation of cAMP receptor protein. Proc Natl Acad Sci U S A 1993;90:75–79. [PubMed: 8380500]
48. Busby S, Truelle N, Spassky A, Dreyfus M, Buc H. The selection and characterisation of two novel mutations in the overlapping promoters of the Escherichia coli galactose operon. Gene 1984;28:201–209. [PubMed: 6376286]
49. Bingham AH, Ponnambalam S, Chan B, Busby S. Mutations that reduce expression from the P2 promoter of the Escherichia coli galactose operon. Gene 1986;41:67–74. [PubMed: 3516794]
50. Johnston F, Ponnambalam S, Busby S. Binding of Escherichia coli RNA polymerase to a promoter carrying mutations that stop transcription initiation. J Mol Biol 1987;195:745–748. [PubMed: 3309340]
51. Tomizawa J, Itoh T, Selzer G, Som T. Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. Proc Natl Acad Sci U S A 1981;78:1421–1425. [PubMed: 6165011]
52. Liu M, Tolstorukov M, Zhurkin V, Garges S, Adhya S. A mutant spacer sequence between -35 and -10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. Proc Natl Acad Sci U S A 2004;101:6911–6916. [PubMed: 15118087]

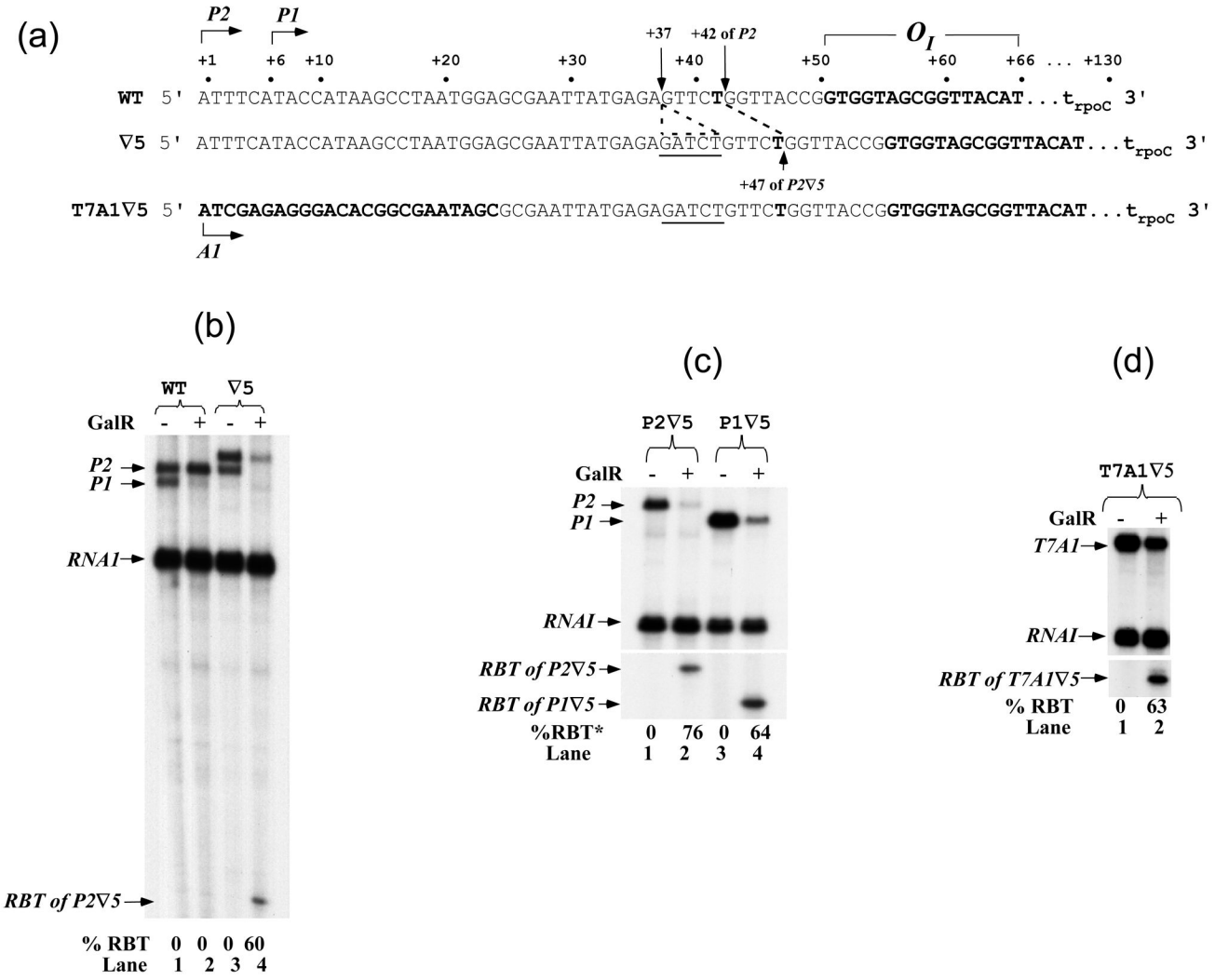

Figure 1. RNAs made from *gal* and *T7A1* promoters

(a) The regulatory region of the *gal* operon. The arrows indicate the position of the transcription start points of *P2* (+1) and *P1* (+6) promoters. In this report, the DNA co-ordinates are referred with respect to the transcription start point of *P2* as +1. GalR binds to O_I (bold letters) from +51 to +66. A *rho*-independent terminator, t_{rpoC} generates transcripts from *P1* and *P2* of 125- and 130-nt, respectively. The 5-bp (GATCT, ∇5) is inserted at position +37 in wt DNA. The 3' ends of RBT in wt and ∇5 are located at positions +42 and +45, respectively. *T7A1* promoter sequence (+1 to +24, bold letters) is fused to the downstream *gal* DNA containing the 5-bp insertion. (b) A 5-bp insert caused short transcript formation. The arrows marked *P2* and *P1* are RNAs made from the two *gal* promoters and terminated at the *rpoC* terminator in the presence (+) and absence (−) of 200 nM GalR; the arrow marked *RNA1* shows the RNA made from a plasmid promoter (*rep*) served as an internal control; the arrow marked RBT of *P2* ∇5 indicates the RNA roadblock from *P2* by O_I-bound GalR. The percentage of *P2* transcripts that was roadblocked by O_I-bound GalR is showed below each lane. Lanes 1, 2 represent wt DNA and lanes 3, 4 represent ∇5 DNA. (c) RBT is formed on both *P1* and *P2* promoters. *Gal* RNAs made *in vitro* from *P2*∇5 (lanes 1, 2) and *P1*∇5 (lane 3, 4) templates. *The percent RBT shown represent average of this and two other experiments. (d) RBT is formed on *T7A1* promoter. RNAs made from *T7A1*∇5 DNA. *T7A1* represents the full-length transcripts

and RBT of *T7A1 ∇5* represents roadblock transcripts. The same nomenclature is used in the following figures.

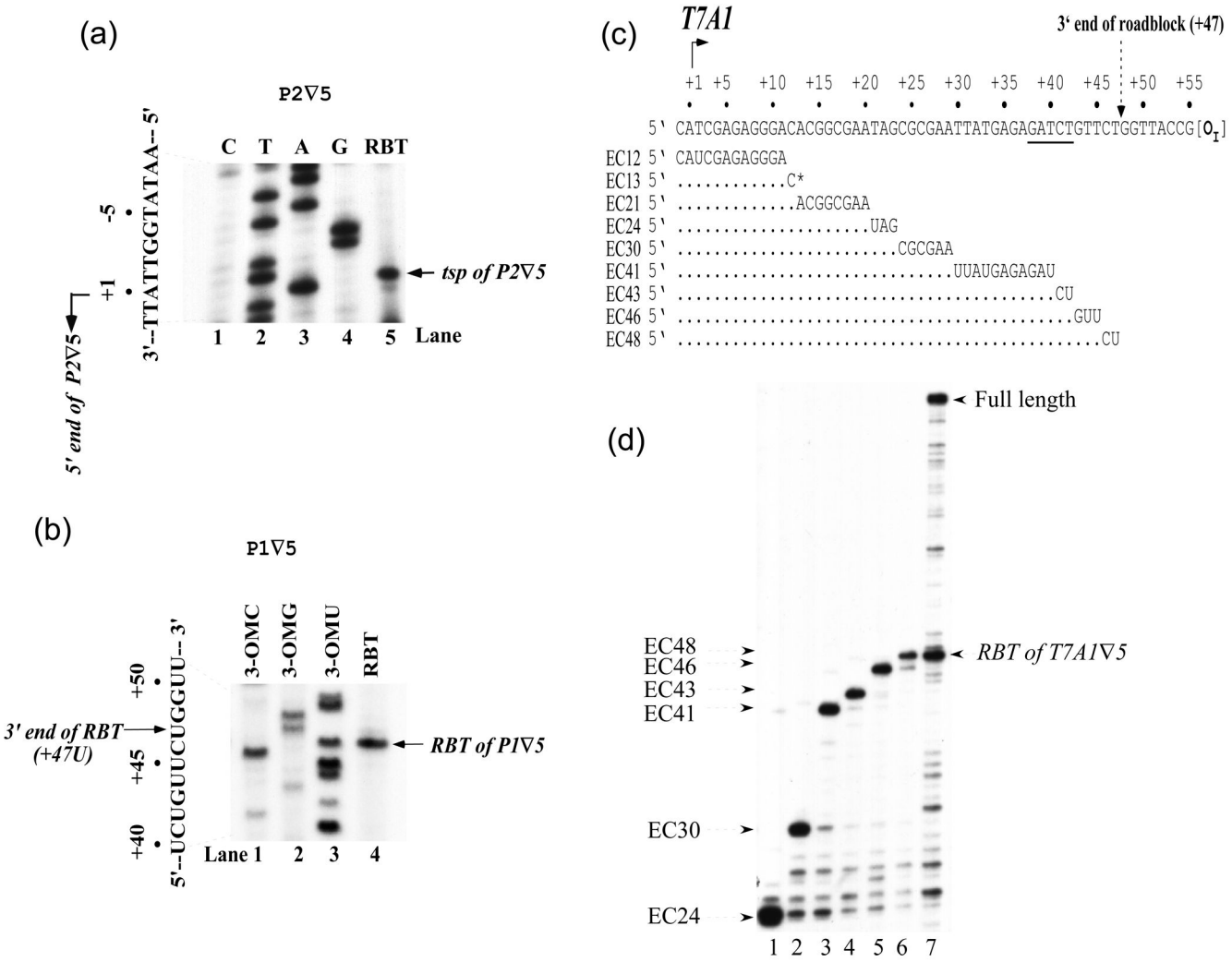

Figure 2. Mapping of 5' and 3' ends of RBT

(a) Mapping of the 5' ends of RBT. The sequence of the nontemplate strand of plasmid *P2*∇5 is shown in lanes 1–4 and to the left of the gel. Note that in the lanes marked C, T, A and G reflect sequencing reactions using dideoxy G, A, T and C, respectively. The product of reverse transcription of RBT obtained on *P2*∇5 is shown in lane 5. The arrow indicates the position of the 5' ends (*tsp*) of *P2*∇5 (+1). (b) Mapping of the 3' ends of RBT. *P1*∇5 linear template was transcribed in the presence cAMP, CRP and 3'-O-methylcytosine-5'-triphosphate or 3'-O-methylguanosine-5'-triphosphate or 3'-O-methyluridine-5'-triphosphate to generate RNA ladders (lanes 1–3). The cAMP-CRP complex enhances transcription from *P1*. Lane 4 shows RBT from supercoiled *P1*∇5 DNA obtained in the presence of GalR as in Figure 1(c), lane 4. The sequence of the RBT from +40 to +51 is shown to the left of the gel. The arrow indicates the position of the 3' end of RBT. (c) Walking of histagged RNAP on *T7A1*∇DNA to the desire positions as indicated by EC# in the presence of a subset of NTPs. EC13 was labeled with [α-^{32}P]CTP. (d) RNA made on *T7A1*∇DNA by walking RNAP; Lane 1: EC24, lane 2: EC30, lane 3: EC41, lane 4: EC43, lane 5: EC46, lane 6: EC48, and lane 7: EC13 in the presence of GalR and all four NTPs.

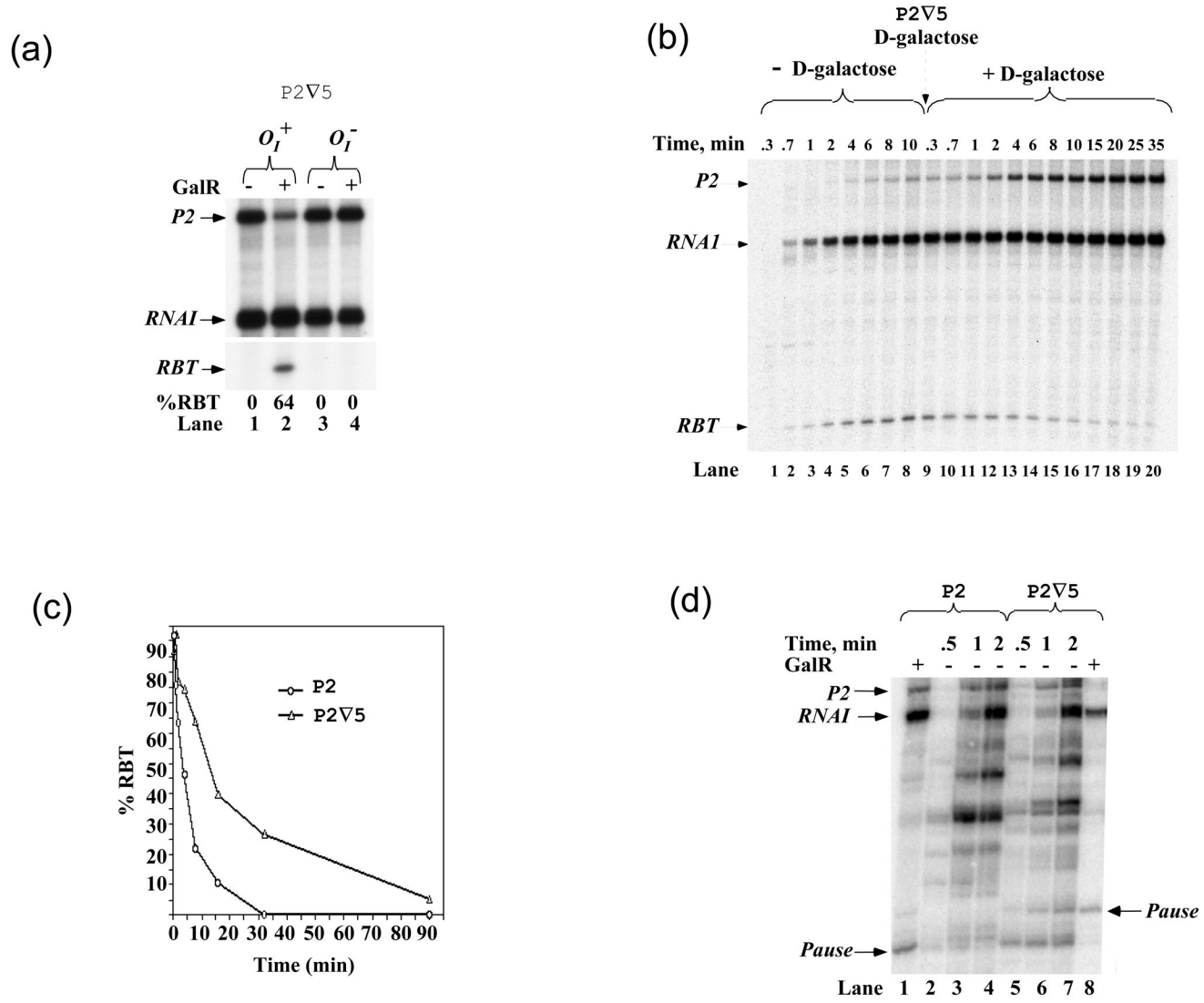

Figure 3. Effect on RBT formation and disappearance

(a) Binding of GalR to O_I is required for RBT formation. RNAs made from wt O_I^+ (lanes 1, 2) and mutant O_I^- (lanes 3, 4) in *P2*∇5 in the presence and absence of 200 nM GalR. (b) RBT belongs to a paused complex. *In vitro* transcription was carried out in the presence of GalR for indicated time periods. After 10 min, D-galactose (broken arrow) was added to a final concentration of 0.8% (v/v) and incubation continued as indicated. (c) Rate of *gal* transcription in the presence of GalR. Transcription was performed on *P2* and *P2*∇5 templates in the presence of GalR for indicated time periods; RBT fraction was calculated. This experiment was done in the presence of 10 μM GTP and 100 μM other NTPs. Similar, when 10 μM UTP and 100 μM other NTPs were used, RBT was also observed in both templates at 2 min following divergent kinetics. (d) Rate of *gal* transcription in the absence of GalR. Transcription was performed on *P2* (lanes 2–4) and *P2*∇5 (lanes 5–7) templates in the absence of GalR for indicated time periods, in the presence of 1 μM GTP and 100 μM other NTPs. Lanes 1 and 8 were RBT markers obtained in the presence of GalR on both templates.

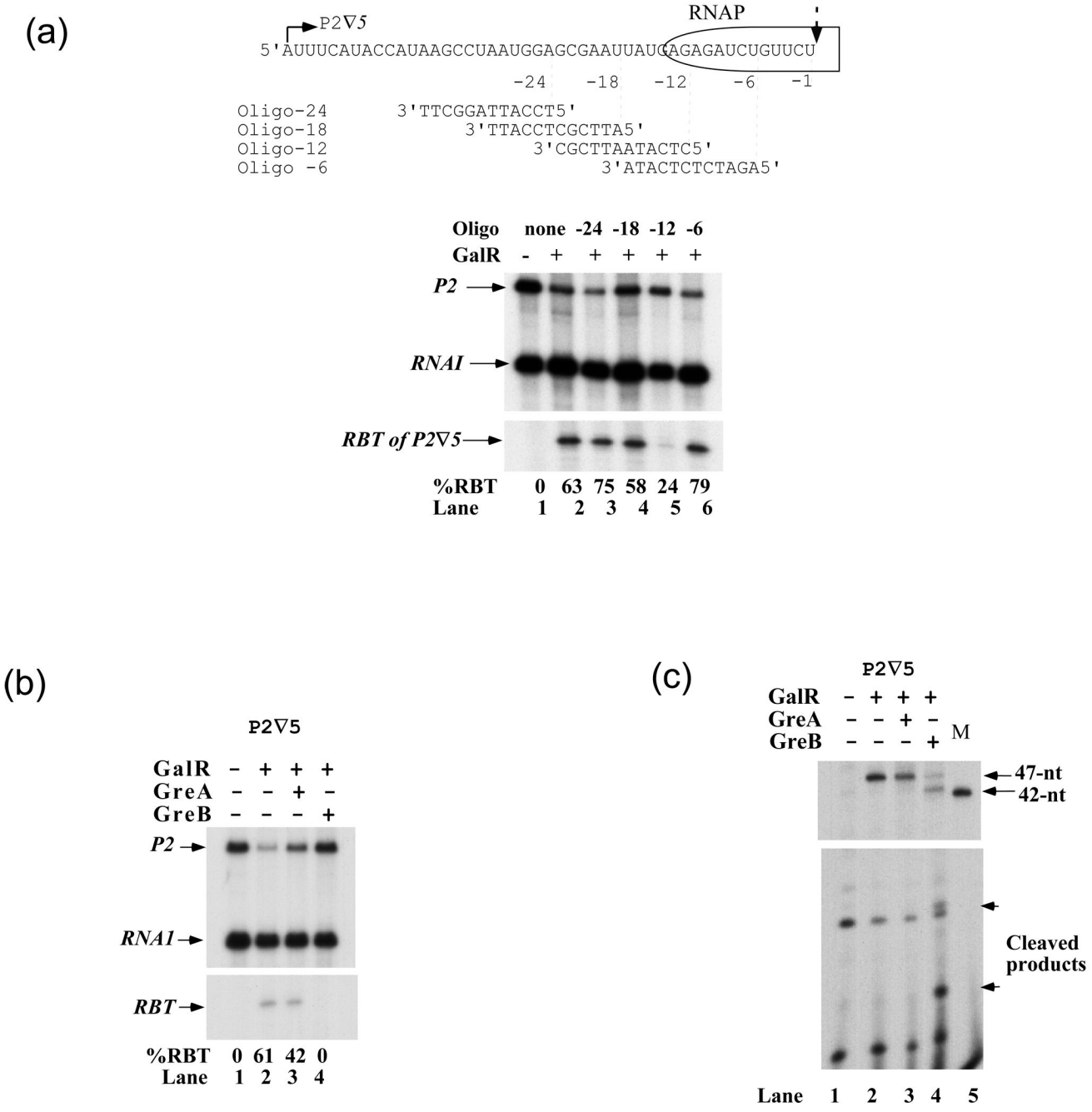

Figure 4. RBT results from RNAP backtracking

(a) The effect of antisense oligonucleotides on RBT. The autoradiogram shows RNAs made on *P2*∇5 DNA in the presence of GalR and oligonucleotides. Schematic represents EC containing RBT (RNA sequence and rear RNAP boundary are shown). The broken arrow indicates the 3' end of RBT (+47U), which is considered as −1 on the RNA, the nucleotides upstream from this site have negative values. The 12-mer oligonucleotides are numbered according to the annealing position of their 5' ends on the RNA. (b) Effect of GreA (50 nM) and GreB (50 nM) on RBT in the presence of NTPs. GreA (lane 3) or GreB (lane 4) was added at the start of the reaction before RBT was formed on *P2*∇5 template as described in Materials

and Methods. The reactions were run on an 8% gel. (c) Cleavage of RBT by GreA and GreB in the absence of NTPs. RBT was made on *P2*∇5 DNA. The NTPs were removed from the reactions by RNA mini spin columns before 50 nM GreA (lane 3) or 50 nM GreB (lane 4) was added for 15 min. The RBT of *P1* was used as a marker lane 5). The reactions were run on an 8% gel (top) and a 24% gel (bottom).

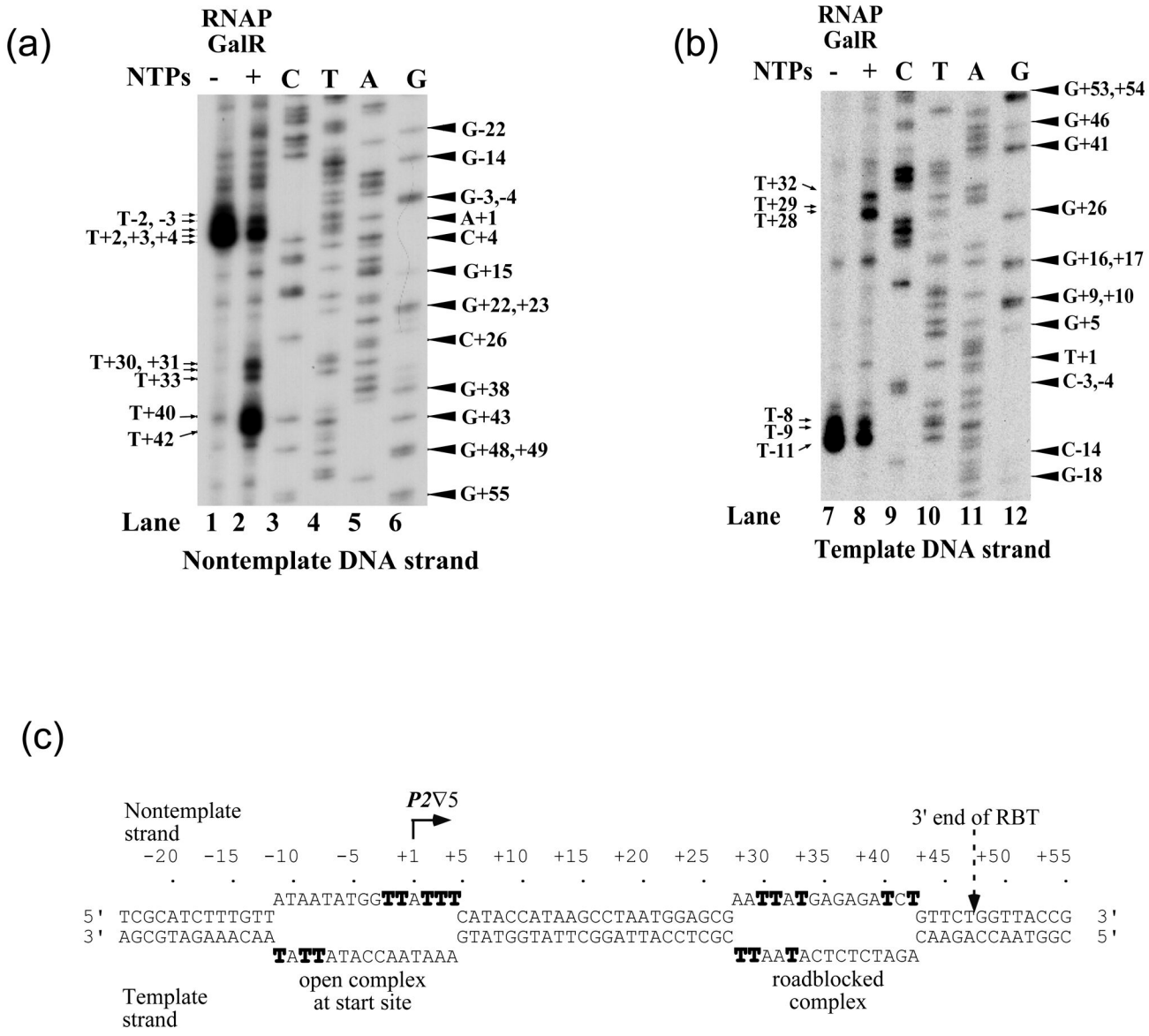

Figure 5. $KMnO_4$ probing of roadblocked complex reveals upstream translocation of transcription bubble

(a) and (b) Open complex was formed on *P2*∇5 DNA in the presence of GalR and RNA poymerase in the presence (lanes 2, 8) and absence (lanes 1, 7) of NTPs. Reactions were treated with $KMnO_4$. DNA was purified and amplified using primer complementary to either the nontemplate (a) or template (b) strand. The same primers were used for sequencing the nontemplate (lanes 3–6) and template (9–12) strands. Note that in the lanes marked C, T, A and G reflect sequencing reactions using dideoxy G, A, T and C, respectively. Arrows on the left of the gels indicate the sites sensitive to $KMnO_4$. (c) The sequence of the *P2*∇5 DNA shows $KMnO_4$-sensitive residues in bold.

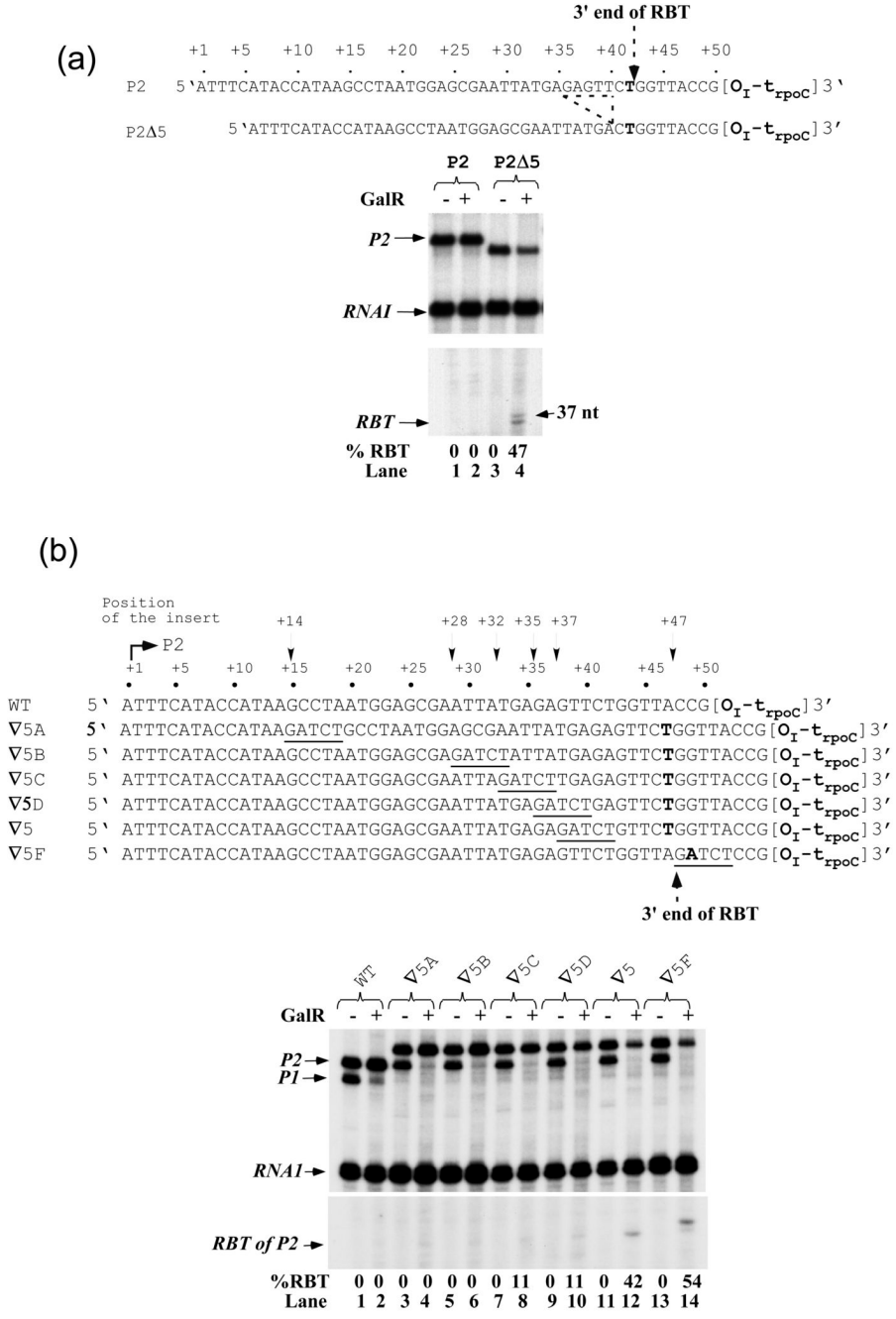

Figure 6. Alteration of *wt gal* sequence within 23-nt of O_I causes RBT formation

(a) Effect of a 5-bp deletion on RBT. The sequence comparison of *P2* and *P2 Δ5* DNAs shows the location of the 5-bp deletion (broken line). RNAs made from *P2* (lanes 1, 2) and *P2 Δ5* (lanes 3, 4). (b) *P2* template derivatives containing GATCT insertion at various positions indicated by arrows were transcribed as described in Materials and Methods except that 2 nM DNA and 80 nM GalR were used. Wt (lanes 1, 2), ∇5A (lanes 3, 4), ∇5B (lanes 5, 6), ∇5C (lanes 7, 8), ∇5D (lanes 9, 10), ∇5 (lanes 11, 12), and ∇5F (lanes 13, 14).

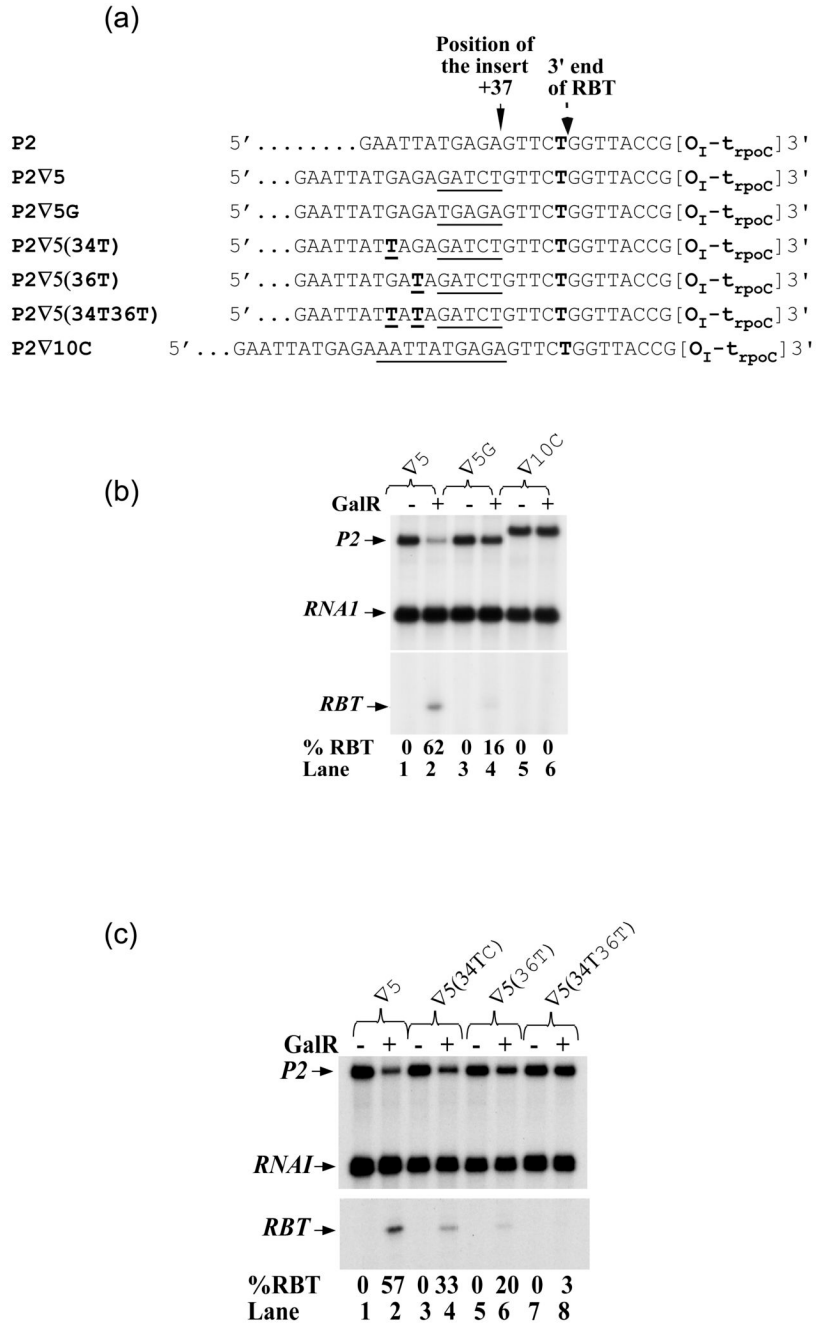

Figure 7. Weakening RNA:DNA hybrid in backtracked EC enhances RBT formation

(a) *P2* template derivatives containing various insertions (underlined) at position +37 or G to T substitutions (bold underlined T), some of which mimic wt sequence. (b) RNA made from templates containing 5- or 10-bp insertion (∇5 (lanes 1, 2), ∇5G (lanes 3, 4) and ∇10C (lanes 5, 6)). (c) RNA made from templates containing G to T substitution at position +34 and/or +36 (∇5 (lanes 1, 2), ∇34T (lanes 3, 4), ∇36T (lanes 5, 6), and ∇34T36T (lanes 7, 8).

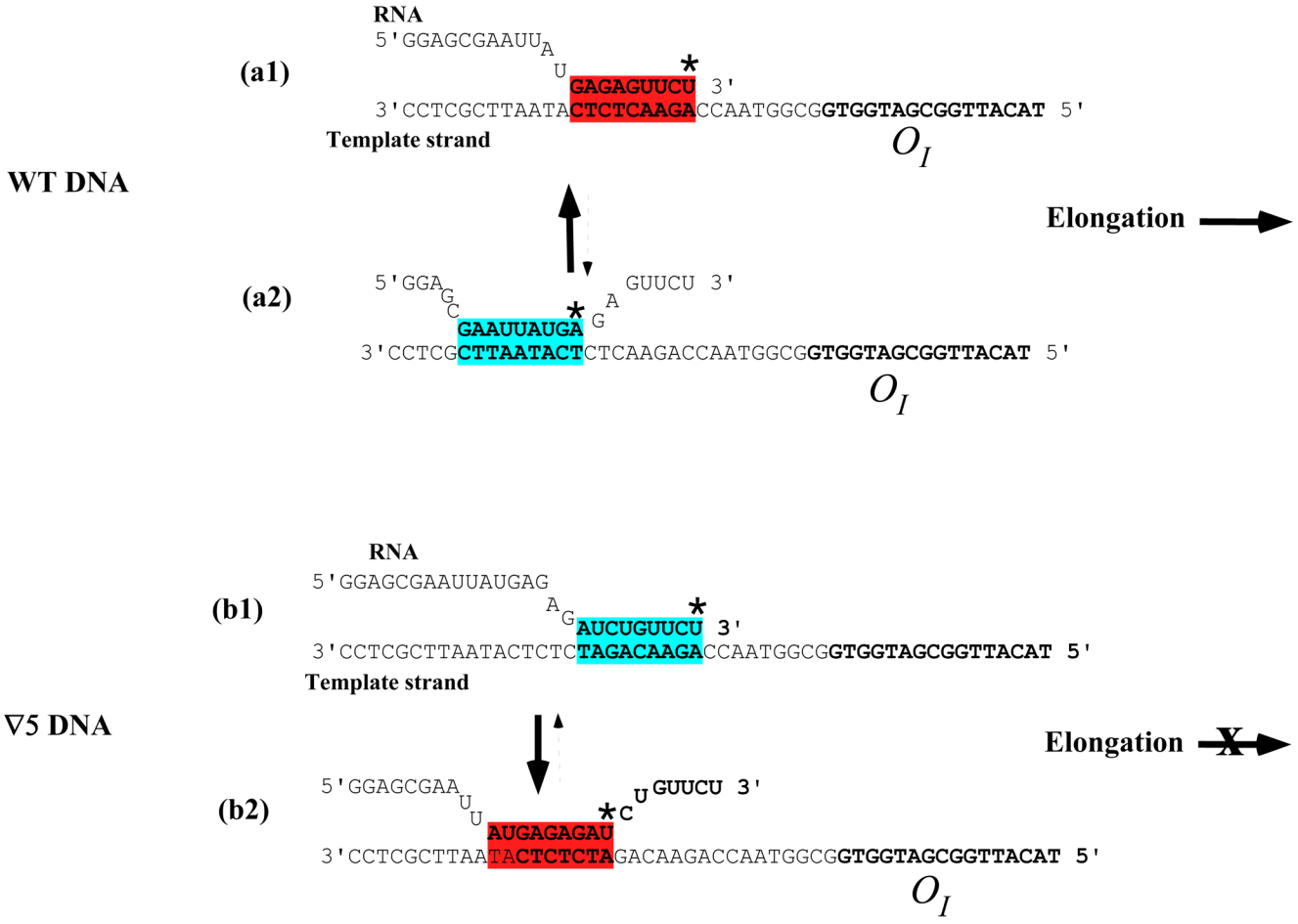

Figure 8. Intrinsic DNA sequence determining whether RNAP elongation or backtracking

The RNA:DNA hybrid at the pause site (a1) and site backtracked by 7-nt (a2) in wild type DNA. The figures show a 9-bp hybrid. The red star represents the base at the catalytic center of RNAP. (a1) When the purine rich sequence (GAGAG) is immediately upstream of the pyrimidine rich sequence (UUCU) as in wild type DNA, the hybrid maintained the 3' end of the RNA in the catalytic center due to a strong RNA:DNA hybrid (red) and favors elongation. (a2) The upstream pyrimidine sequence (AAUUAU) counteracts backtracking of RNAP in wild type DNA due to a weak RNA:DNA hybrid (blue). The solid arrow indicates that RNAP would remain more at the strong hybrid instead of the weak hybrid. The broken arrow signifies the opposite. (b1) and (b2) When a 5-bp (GATCT) DNA separated the GAGAG site from the UUCU pause site, RNAP backtracks from a weak hybrid (blue) to a stable hybrid (red), favoring RNAP arrest as indicated by the blocking of elongation by the red X.

Table 1

Calculations of predicted ΔG_{37} of 9-bp RNA:DNA hybrid. RNA sequence comparison of wt and mutant templates upstream of the 3' end of RBT. Predicted ΔG_{37} of RNA:DNA hybrid at pause sites and sites backtracked by 4- and 7-nt. The percentage of RBT for each template is shown. The pyrimidine-rich pause site (UUCU) and the upstream AU-rich (AAUUAU) sequence are outlined in blue, purine-rich (GAGAG) is in red. Gs to Us changes are in green. All sequences that were analyzed contain a UUCU sequence at the 3' end of the hybrid. The AU-rich sequence placed upstream of the hybrid does not alter backtracking from such a RNA:DNA hybrid. The GAGAG sequence strengthens the hybrid at the pause site in the wild type as an anti-pause element, promoting RNAP elongation. When a 5-bp is inserted between the GAGAG site and the pause site, RNAP backtracked from a weak hybrid to a stable hybrid, and hinders elongation.

#	Plasmids	RNA sequence upstream of RBT 3' end	Predicted ΔG_{37} (kcal mol^{-1})			% RBT
			Pause site	Backtracked by 4-nt	Backtracked by 7-nt	
1	WT	GGAGCGAAUUAUGAGAGUUCU	-6.8	-6.4	-4.7	0
2	∇5Δ5B	GGAGCGAAUUAUGAGAGAUCU	-7.7	-6.4	-4.7	0
3	∇5Δ5C	GGAGCGAAUUAGAUCUGUUCU	-5.5	-5.7	-4.9	0
4	∇5Δ5D	GGAGCGAAUUAUAUCUGUUCU	-5.5	-4.1	-3.3	0
5	∇10C	AUGAGAAAUUAUGAGAGUUCU	-6.8	-6.4	-4.2	0
6	∇5(+34T36T)	GAAUUAUUAUAGAUCUGUUCU	-5.5	-6.4	-4.1	0
7	∇5D	GAAUUAUGAGAUCUGAGUUCU	-6.2	-8.0	-7.1	11
8	∇5G	GAAUUAUGAGAUGAGAGUUCU	-6.8	-8.7	-7.6	16
9	∇5(+36T)	GAAUUAUGAUAGAUCUGUUCU	-5.5	-6.4	-6.2	20
10	∇5(+34T)	GAAUUAUUAGAGAUCUGUUCU	-5.5	-7.1	-5.7	33
11	∇5	GAAUUAUGAGAGAUCUGUUCU	-5.5	-8.0	-7.8	60
12	Δ5	CUAAUGGAGCGAAUUAUUUCU	-5.4	-6.4	-12.5	47
13	∇5Δ5A	GGAGCGAGAGAGAUCUGUUCU	-5.5	-8.0	-8.4	42
14	∇5I	GAAUUAUGAGAUAUUGGUUCU	-6.2	-7.1	-6.2	48
15	∇5II	GAAUUAUGAGACGGCUGUUCU	-8.7	-11.9	-10.5	32
16	∇5III	GAAUUAUGAGAGAUCUGUUCU	-6.5	-10.6	-9.8	50
17	∇5IV	GAAUUAUGAGAAAGUUGUUCU	-5.3	-6.7	-7.6	29
18	∇5V	GAAUUAUGAGAGUGAUGUUCU	-6.0	-8.3	-8.3	32
19	∇10B	AUGAGAUGAGAGAUCUGUUCU	-5.5	-8.0	-7.8	38
20	∇10A	AUGAGAGAAUUAGAGAGUUCU	-6.8	-7.3	-6.4	8
21	∇5C	GAAUUAGAUCUUGAGAGUUCU	-6.8	-7.3	-6.4	11
22	∇5VI	GAAUUAUGAGAUUGGCGUUCU	-9.5	-10.0	-6.5	16

Table 2

List of plasmids

Plasmid	Local name	Features	Source
pSA850	wt	a 166-bp *gal* DNA fragment (−75 to +91)	Lewis & Adhya, 2002
pDL907	wt∇5A	pSA850 ∇ GATCT @+14 of *P2* tsp	This work
pDL511	wt∇5B	pSA850 ∇ GATCT @+28	This work
pDL512	wt∇5C	pSA850 ∇ GATCT @+32	This work
pDL513	wt∇5D	pSA850 ∇ GATCT @+35	This work
pSA859	wt∇5E	pSA850 ∇ GATCT @+37	Lewis & Adhya, 2002
PDL908	wt∇5F	pSA850 ∇ GATCT @+47	This work
pDL229	*P1*con	pSA850 (−15T, − 4A & −3A)	This work
pDL994	*P1*con∇5	pDL229 ∇ GATCT @+37	This work
pDL225	*P2*con	pSA850 (−8A & −9A)	This work
pDL1022	*P2*conΔ5	pDL225ΔGAGTT (+36 to +40)	This work
pDL1004	*P2*con∇5	pDL225 ∇ GATCT @+37	This work
pDL1039	*P2*con∇5A	pDL225 ∇ TATTG @+37	This work
pDL1040	*P2*con∇5B	pDL225 ∇ CGGCT @+37	This work
pDL1041	*P2*con∇5C	pDL225 ∇ GGACT @+37	This work
pDL1042	*P2*con∇5D	pDL225 ∇ AAGTT @+37	This work
pDL1043	*P2*con∇5E	pDL225 ∇ GTGAT @+37	This work
pDL1044	*P2*con∇5F	pDL225 ∇ TTGGC @+37	This work
pDL1060	*P2*con∇5G	pDL225 ∇ TGAGA @+37	This work
pDL1061	*P2*con∇10C	pDL225 ∇ AATTATGAGA @+37	This work
pDL1062	*P2*con∇10A	pDL225 ∇ GATCTTGAGA @+37	This work
pDL1063	*P2*con∇10B	pDL225 ∇ TGAGAGATCT @+37	This work
pDL1011	*P2*con∇5Δ5A	pDL1004 (ΔATTAT (+29 to +33))	This work
pDL1012	*P2*con∇5Δ5B	pDL1004 (ΔGTTCT (+43 to +47))	This work
pDL1079	*P2*con∇5Δ5C	pDL1004 (ΔTGAGA (+33 to +37))	This work
pDL1080	*P2*con∇5Δ5D	pDL1004 (ΔGAGAG (+34 to +38))	This work
pDL999	*P2*con∇5O^N_1	pDL1004 (O^N_1)	This work
pDL571	*T7A1*∇5	pSA859 (*T7A1* promoter DNA (−56 to + 24)	This work