

# DNA Sequences in *gal* Operon Override Transcription Elongation Blocks

Dale E. A. Lewis<sup>1\*</sup>, Natalia Komissarova<sup>2</sup>, Phuoc Le<sup>1</sup>, Mikhail Kashlev<sup>3</sup> and Sankar Adhya<sup>1</sup>

<sup>1</sup>Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-4264, USA

<sup>2</sup>Section on Microbial Genetics, Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892-2785, USA

<sup>3</sup>Frederick Cancer Research and Development Center, National Cancer Institute, Frederick, MD 21702-1201, USA

Received 16 April 2008;  
received in revised form  
12 June 2008;  
accepted 23 July 2008  
Available online  
27 July 2008

Edited by R. Ebright

The DNA loop that represses transcription from galactose (*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 physical block because of the presence of the Gal repressor 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  $O_I$  weakens 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 antipause element by stabilizing the RNA:DNA duplex and preventing backtracking. This facilitates forward translocation of RNAP, including overriding of the DNA-bound Gal repressor 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 of regulating *gal* metabolism in *Escherichia coli*. It also explains the synthesis of *gal* enzymes in the absence of inducer for biosynthetic reactions.

Published by Elsevier Ltd.

Keywords: GalR; backtracking; roadblock; pausing; transcription

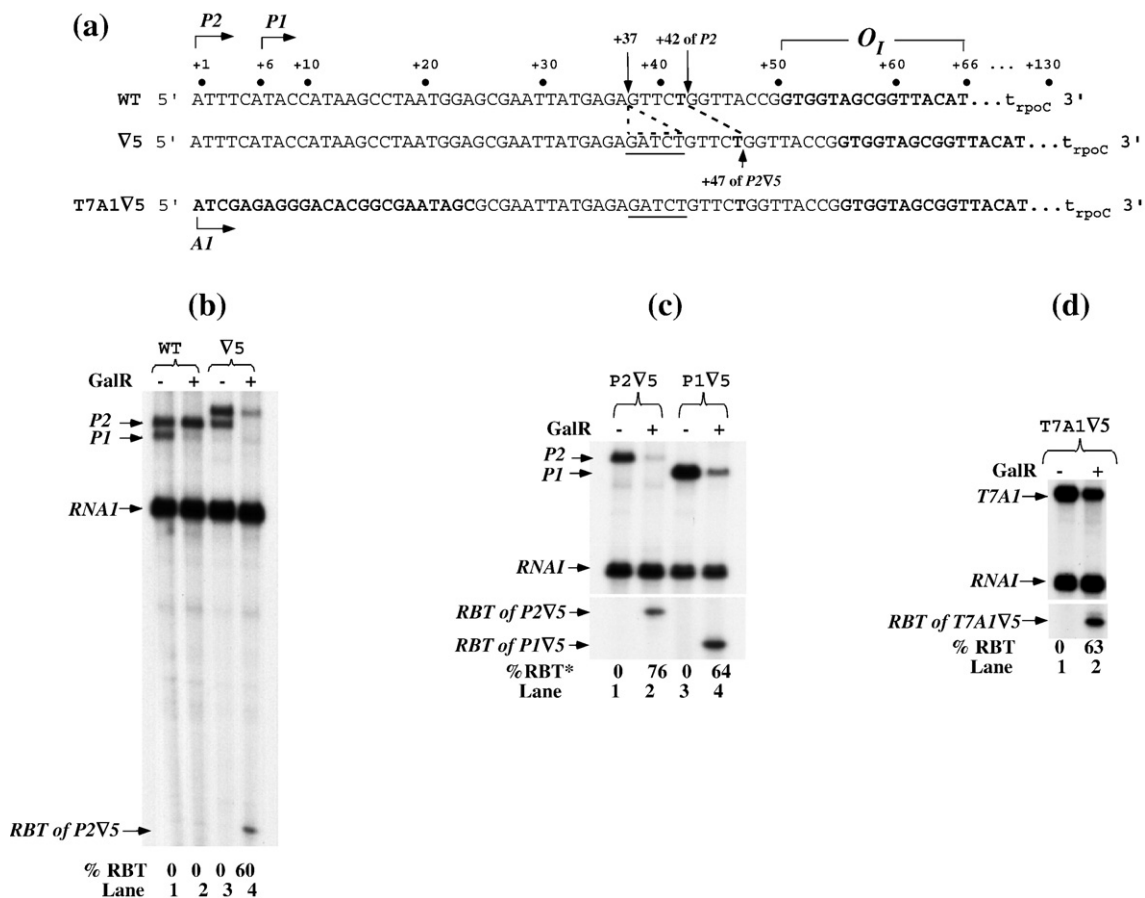
## Introduction

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*) operon of *Escherichia coli* is transcribed from two overlapping promoters,  $P_2$  [transcription start point (*tsp*) as +1] and  $P_1$  (*tsp* as +6) (Fig. 1a), which are regulated by several transcription factors, including Gal repressor

(GalR).<sup>1–5</sup> 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  $P_1$  and  $P_2$ .<sup>4–6</sup> In the absence of HU, GalR binding to  $O_E$  stimulates  $P_2$  and represses  $P_1$  by promoting and inhibiting open complex formation, respectively.<sup>7–9</sup> *In vivo*,  $P_2$  is activated approximately sixfold in the absence of HU.<sup>6</sup> 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*<sup>10</sup> (also 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

\*Corresponding author. E-mail address: aed@helix.nih.gov.

Abbreviations used: *gal*, galactose; RNAP, RNA polymerase; *tsp*, transcription start point; GalR, Gal repressor; wt, wild type; RBT, roadblock transcript; NTP, nucleotide triphosphate; CTP, cytidine triphosphate; EC, elongation complex; GTP, guanosine triphosphate; UTP, uridine triphosphate; CRP, cyclic AMP receptor protein.



**Fig. 1.** RNAs made from *gal* and *T7A1* promoters. (a) The regulatory region of the *gal* operon. The arrows indicate the position of the *tsf* of *P2* (+1) and *P1* (+6) promoters. In this report, the DNA coordinates are referred to with respect to the *tsf* 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. Five base pairs (*GATCT*,  $\nabla 5$ ) are inserted at position +37 in wt DNA. The 3' ends of RBT in wt and  $\nabla 5$  are located at positions +42 and +47, 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 a 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 in the absence (-) of 200 nM GalR; the arrow marked *RNAI* shows that the RNA made from a plasmid promoter (*rep*) served as internal control; the arrow marked RBT of *P2V5* 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 shown below each lane. Lanes 1 and 2 represent wt DNA, and lanes 3 and 4 represent  $\nabla 5$  DNA. (c) RBT is formed on both *P1* and *P2* promoters. *gal* RNAs made *in vitro* from *P2V5* (lanes 1 and 2) and *P1V5* (lanes 3 and 4) templates. \*The percent RBT shown represents the average of this and two other experiments. (d) RBT is formed on *T7A1* promoter. RNAs made from *T7A1V5* DNA. *T7A1* represents the full-length transcripts, and RBT of *T7A1V5* represents RBTs. The same nomenclature is used in the following figures.

the wild-type (wt) *gal* DNA resistant to roadblock during elongation. We found that RNAP pausing produces short transcripts transiently in the wt, but persistently 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.<sup>10</sup> We inserted a 5-bp (*GATCT*) sequence at position +37 of *P2 tsf* to change the orientation of  $O_I$  by 5 bp with respect to  $O_E$ , preventing DNA looping. We found that *P2* was repressed in the absence of looping in this construct by an unknown mechanism. In this study, we investigated this mode of repression. Figure 1a and b shows DNA templates and results of 10-min transcription on wt *gal* template and on 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.<sup>11</sup> GalR repressed *P1* and partially activated *P2* in the wt DNA, as expected (Fig. 1b, lanes 1 and 2). Surprisingly, in the 5-bp insertion template (Fig. 1a,  $\nabla 5$ ), *P2* transcripts were

reduced 2.8-fold by GalR (Fig. 1b, lanes 3 and 4). This reduction in *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.<sup>10</sup> 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 “roadblock 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 (Fig. 1b). We investigated whether GalR causes RBT from the *P1* promoter. The  $O_E$  sequence was mutated to a nonoperator sequence in the *P1* template to prevent GalR from repressing *P1* (template  $P1\nabla 5$ ; 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  $\nabla 5$ ; see Plasmids; Fig. 1c, lanes 1 and 3). GalR blocked 76% full-length transcripts from  $P2\nabla 5$  and 64% full-length transcripts from  $P1\nabla 5$  (lanes 2 and 4). The difference in RBT formation in Fig. 1b 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 ( $T7A1\nabla 5$ ), as indicated in Fig. 1a. The  $O_I$ -bound GalR blocked 63% of *T7A1* transcription (Fig. 1d), 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 two rounds of transcription were observed in the absence of heparin (Supplementary Fig. S1, compare full-length *P2* in lanes 1 and 3). Second, almost the same amounts of RBT (89% and 83%) were observed in the presence and in the absence of heparin, respectively (Supplementary Fig. S1, lanes 2 and 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  $P2\nabla 5$  template and used in reverse transcriptase assays, with a primer annealing to the RNA region from +22 to +37 (Fig. 2a, lane 5). The same primer was used to sequence the template DNA strand, which is complimentary to the non-template strand (lanes 1–4). The results showed that RBT from  $P2\nabla 5$  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

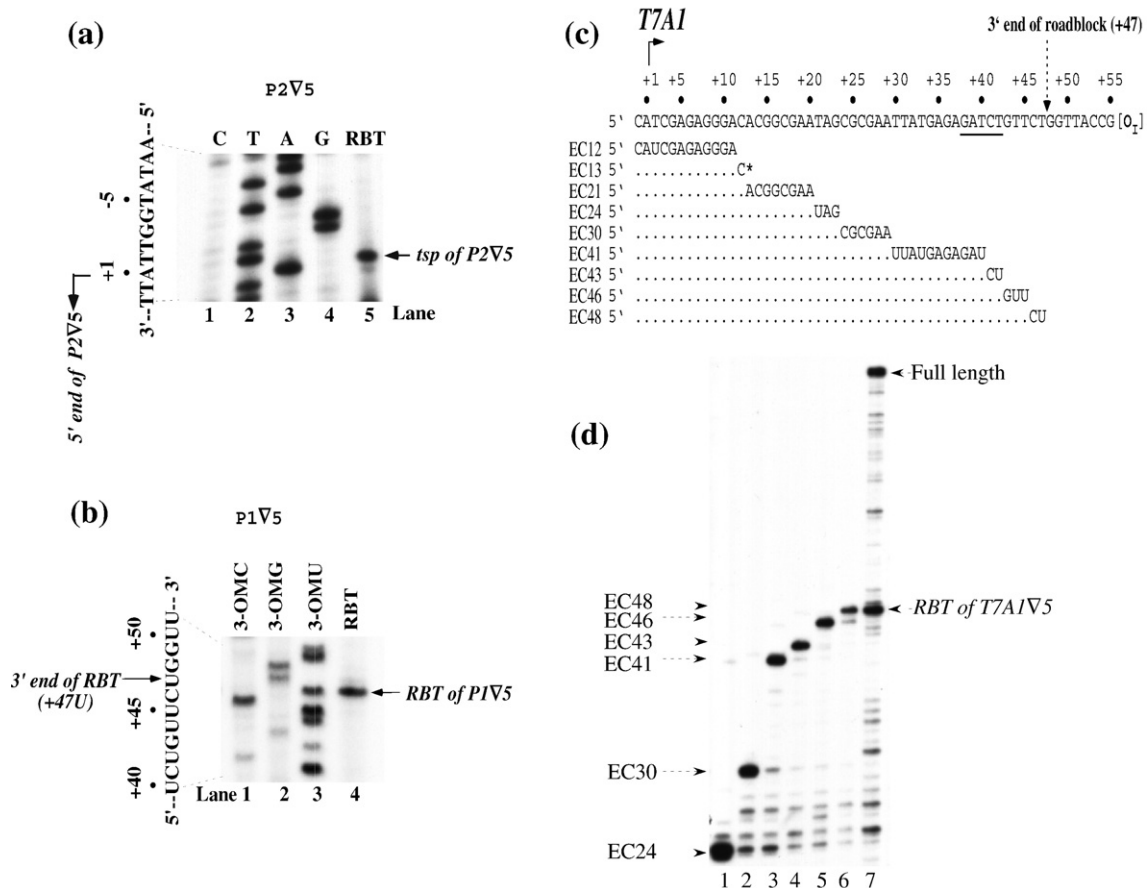
reaction on the  $P1\nabla 5$  template in the absence of GalR to generate an RNA sized-ladder (Fig. 2b, lanes 1–3).<sup>11</sup> In a separate reaction, RBT was obtained on the same template in the presence of GalR (lane 4). The 3' end of RBT from  $P1\nabla 5$  was mapped at position +47U, demonstrating that the length of RBT was 42 nt since *P1* transcribes from +6. The 3' end of RBT was 5 bp downstream of the GATCT sequence and 8 bp upstream of the  $O_I$  sequence. Identical results were obtained for  $P2\nabla 5$ . Its RBT was 47 nt long (data not shown), showing that elongating RNAPs from  $P1\nabla 5$  and  $P2\nabla 5$  halted at the same site. In addition, we used  $T7A1\nabla 5$  DNA and walked His-tagged RNAP in the presence of a subset of 5  $\mu$ M nucleotide triphosphates (NTPs) to the desired positions as indicated (Fig. 2c and d). The length of the RBT was 48 nt (Fig. 2d, lane 8). However, the  $T7A1\nabla 5$  DNA was initiated with a tetranucleotide (CAUC) at position –1 and labeled at position 13 (EC13) with [ $\alpha$ -<sup>32</sup>P]cytidine triphosphate (CTP). Therefore, the actual length of RBT was 47 nt. The 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  $P2\nabla 5$  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 nonoperator sequence to prevent GalR from binding to it. Indeed, while 64% RBT was observed on  $O_I^+$  template (Fig. 3a, lanes 1 and 2), there was no reduction of full-length transcripts from *P2* and the formation of RBT on  $O_I^-$  template (lanes 3 and 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.<sup>12–14</sup> 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 complex or a released complex. If the inactivation were intrinsically reversible, GalR removal from  $O_I$  would cause the paused EC to resume elongation and to synthesize full-length transcripts. Figure 3b shows the amount of RBT before and after D-galactose addition. RBT was observed at 0.7 min after transcription initiation, with 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



**Fig. 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 *P2V5* is shown in lanes 1–4 and to the left of the gel. Note that the lanes marked C, T, A, and G reflect sequencing reactions using dideoxy G, A, T, and C, respectively. The product of the reverse transcription of RBT obtained on *P2V5* is shown in lane 5. The arrow indicates the position of the 5' ends (*tsp* of *P2V5* (+1)). (b) Mapping of the 3' ends of RBT. *P1V5* linear template was transcribed in the presence of cAMP, CRP, and 3'-*O*-methylcytosine-5'-triphosphate, 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 *P1V5* DNA obtained in the presence of GalR as in Fig. 1c (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 His-tagged RNAP on *T7A1V5* DNA to the desired positions as indicated by EC number in the presence of a subset of NTPs. EC13 was labeled with [ $\alpha$ -<sup>32</sup>P]CTP. (d) RNA made on *T7A1V5* 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.

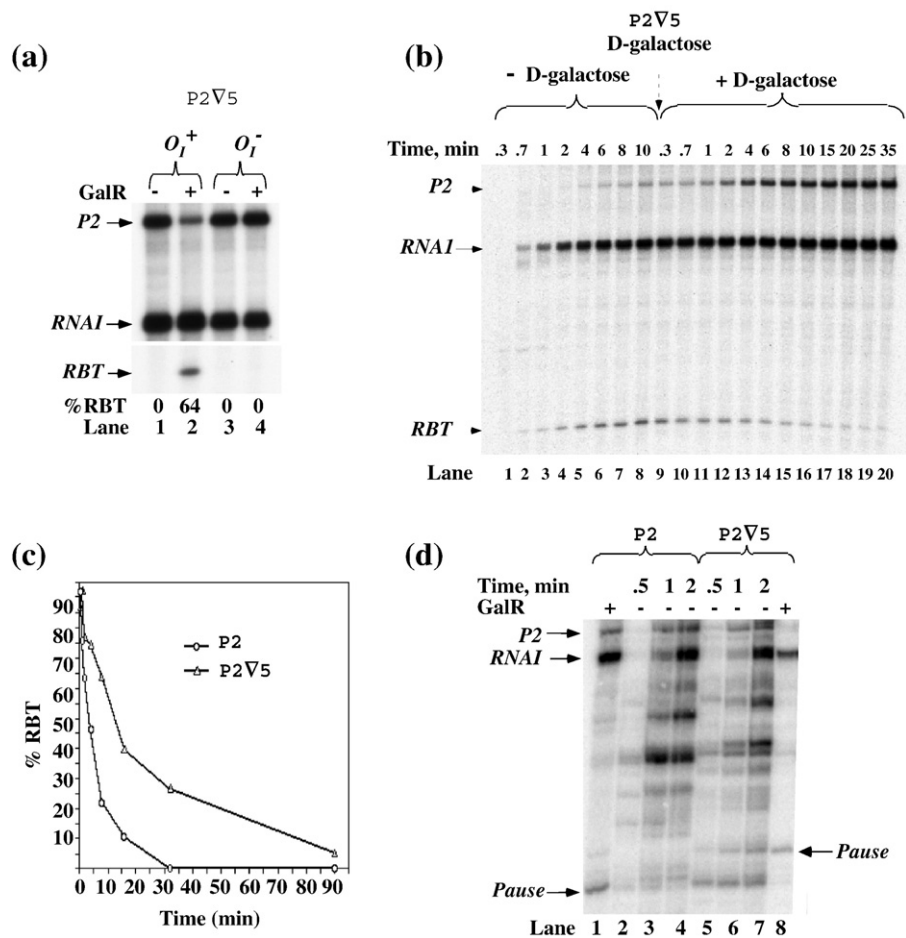
through  $O_1$ -bound GalR even in the absence of D-galactose over extended time periods (data not shown), perhaps because of spontaneous dissociation of GalR from  $O_1$ .

### 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 behaviors of RNAP as it approaches  $O_1$ -bound GalR on the *P2* and *P2V5* templates. Figure 3c shows that at 30 s, the same fraction of RBT was made on both templates. On the wt template, the RBT fraction diminished rapidly with a half-life of 5 min upon conversion into full-length transcripts. In contrast, on the *P2V5* 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 of  $O_1$  both in *P2* and in *P2V5* templates at the <sup>+44</sup>-UUCU-<sup>+47</sup> RNA sequence. Whereas the pause was very brief in *P2*, RNAP needed longer time to overcome the pause in *P2V5*. This suggests that a certain element(s) of the wt sequence that helps RNAP escape from the pause state was altered by the 5-bp insertion.

Next, we compared the kinetics of transcription on the two templates in the absence of GalR (Fig. 3d). To enhance pausing, we limited the concentration of guanosine triphosphate (GTP; the next base after the pause site <sup>+44</sup>-UUCUG-<sup>+48</sup>) to 1  $\mu$ M, but kept the other NTPs at 100  $\mu$ M. The paused transcripts occurred at the same position on the templates where RBTs were formed in the presence of GalR. Furthermore, the pause was much longer in the *P2V5* template than in the *P2* template (lanes 2–7), in agreement with the results shown in Fig. 3c confirming that an intrinsic pause site in *gal* DNA was intensified by the



**Fig. 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 and 2) and mutant  $O_I^-$  (lanes 3 and 4) in *P2∇5* in the presence and in the 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% (vol/vol), 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 performed in the presence of 10  $\mu$ M GTP and 100  $\mu$ M other NTPs. Similarly, when 10  $\mu$ M UTP and 100  $\mu$ M 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  $\mu$ M GTP and 100  $\mu$ M other NTPs. Lanes 1 and 8 were RBT markers obtained in the presence of GalR on both templates.

5-bp insertion. The presence of GalR at  $O_I$  makes the latter pause generate a much higher level of RBT as shown in Fig. 1b.

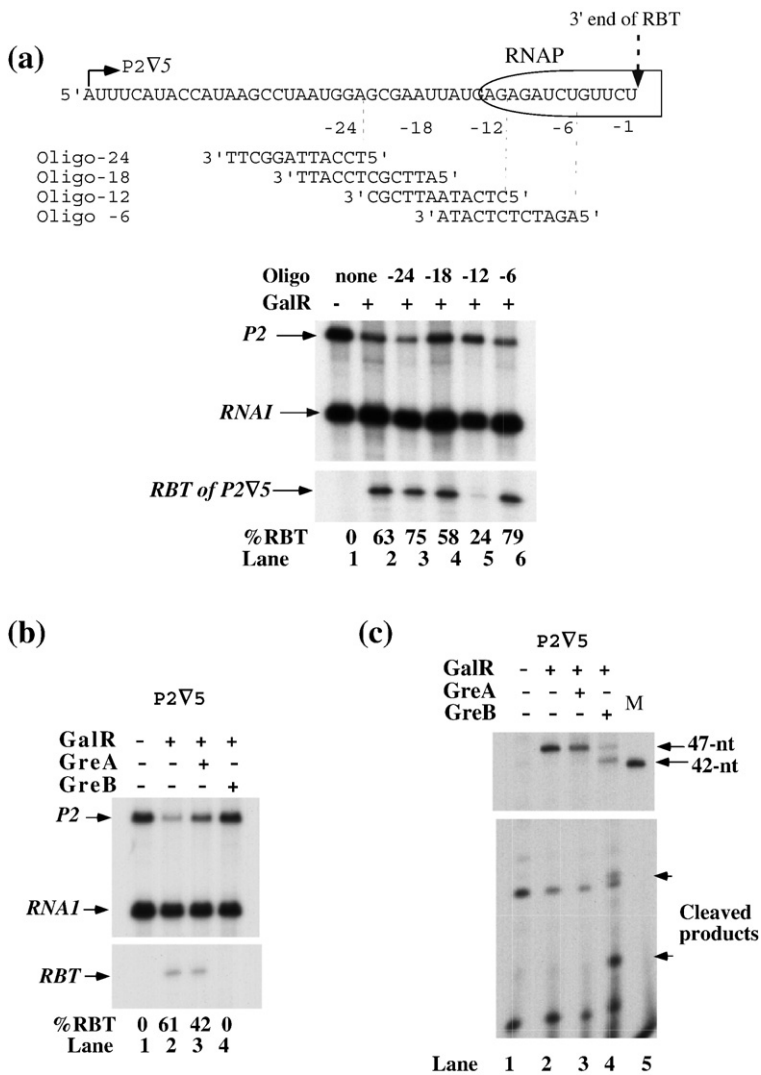
#### Antisense oligonucleotides suppress roadblock

It was shown previously that when a Lac operator was placed at an internal position of a transcription unit, a Lac repressor bound to it, stopped the elongating RNAP, and caused reversible backtracking of the enzyme.<sup>15</sup> 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.<sup>16</sup> 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).<sup>16,17</sup> It was shown previously that a U-rich RNA sequence, which forms a weak RNA:DNA hybrid, favors backtracking.<sup>17–19</sup>

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.<sup>20</sup> 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.<sup>15,16,18</sup> This action of oligonucleotides suppresses RBT if the latter is indeed caused by backtracking.

Figure 4a shows the results of transcription on *P2∇5* template in the presence of GalR and antisense oligonucleotides. Oligo-24, which anneals to the nascent RNA at a distance of 24 nt upstream of



**Fig. 4.** RBT results from RNAP backtracking. (a) The effect of anti-sense oligonucleotides on RBT. The autoradiogram shows RNAs made on P2V5 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 of 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 P2V5 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 P2V5 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 marker (lane 5). The reactions were run on an 8% gel (top) and on a 24% gel (bottom).

the 3' end of RBT, had no effect on roadblock formation (lanes 2 and 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 that is 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.<sup>21</sup>

### GreA and GreB effects on RBT

GreA and GreB are transcription factors known to influence elongation by inducing an endoribonuclease activity of the RNAP catalytic center.<sup>22–27</sup> 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

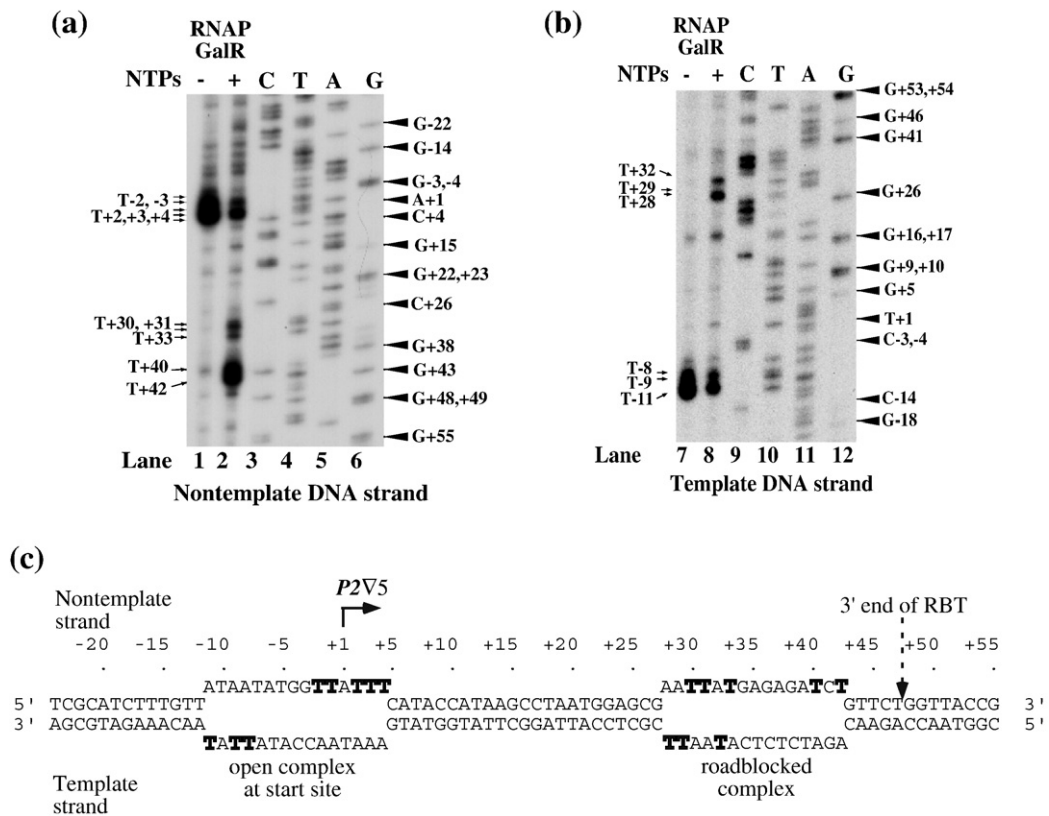
of the enzyme is aligned with an internal position of the RNA.<sup>23,28</sup> The cleavage rescues arrested complexes by generating a new 3' RNA end in the catalytic center.<sup>22,23,29</sup> Unlike GreB, GreA cleaves off only short (2–3 nt) RNA fragments and only in active ECs that briefly travel in the reverse direction by 2–3 nt. To test whether GreA and/or GreB affects RBT, the factors were added at 50 nM concentration to the transcription assays before RBT was formed. Figure 4b shows that GreA suppressed RBT formation slightly in the P2V5 template (lanes 2 and 3), while GreB suppressed it completely by converting RBT into full-length transcripts (lanes 2 and 4). Since the dissociation constant ( $K_d$ ) for GreA is within the range of 800–1000 nM, and that for GreB is approximately 80–100 nM,<sup>25,30–32</sup> we also used 800 nM GreA and 80 nM GreB in other reactions to test whether the low level of GreA (50 nM) was responsible for the partial suppression of RBT. We obtained similar results (data not shown). The sensitivity of the nascent RNA to GreB and partial resistance to GreA cleavage strongly support the idea that RNAP backtracks along the DNA after it encounters the O<sub>I</sub>-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 (Fig. 4c, lane 3) GreB generated a 43-nt RNA by cleaving 4 nt 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- to 42-nt products were observed (data not shown). Truncation of RBT by 4 and 7 nt shows that the RNAP catalytic center relocates 4 or 7 nt upstream of the original site. In this experiment, GreA did not cleave RBT because RNAP, after 15 min of incubation, backtracked by 4–7 nt, thus making the RNA insensitive to GreA.

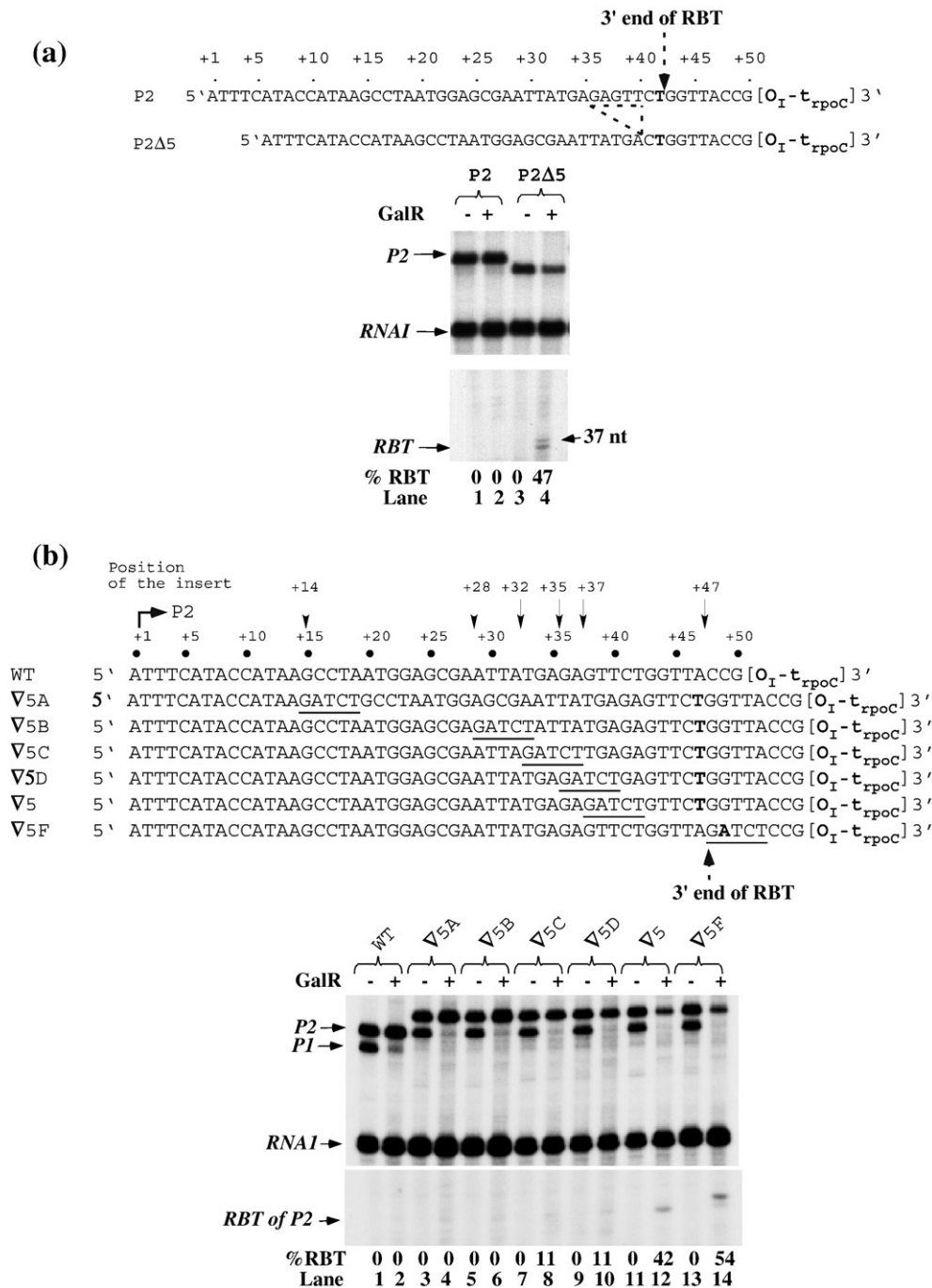
**The transcription bubble in the roadblocked complex**

We mapped the “transcription bubble” in the roadblocked EC by using potassium permanganate (KMnO<sub>4</sub>), which oxidizes distorted, unpaired thymidine nucleotides.<sup>33,34</sup> *Taq* DNA polymerase is unable to read through the oxidized nucleotides,

allowing their detection in the modified templates by PCR amplification.<sup>33</sup> KMnO<sub>4</sub> was added to the *P2∇5* DNA in the presence of GalR and RNAP with or without NTPs. In the absence of NTPs, regions –2 to +4 on the nontemplate strand and regions –11 to –8 on the template strand were modified by KMnO<sub>4</sub>, revealing open complex formation (Fig. 5, lanes 1 and 7). In the presence of NTPs, the bubble translocated downstream of the initiation site toward the termination site, as shown by the reduction of the signal around the regions –11 to +4 (lanes 2 and 8). Incidentally, the residual open complexes found at the promoter region in the presence of NTPs may reflect the nonproductive complexes described previously.<sup>35</sup> KMnO<sub>4</sub>-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<sup>36</sup> would have spread between positions +35 and +48, with a strong KMnO<sub>4</sub> cleavage signal from T residues in segments +44 to +47 of the nontemplate strand. We did not detect such a signal (lane 2), but observed a 15-nt-long transcription bubble spreading between



**Fig. 5.** KMnO<sub>4</sub> 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 polymerase in the presence (lanes 2 and 8) and in the absence (lanes 1 and 7) of NTPs. Reactions were treated with KMnO<sub>4</sub>. DNA was purified and amplified using a primer complementary to either the nontemplate (a) and template (b) strands. The same primers were used for sequencing the nontemplate (lanes 3–6) and template (9–12) strands. Note that 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<sub>4</sub>. (c) The sequence of the *P2∇5* DNA shows KMnO<sub>4</sub>-sensitive residues in bold.



**Fig. 6.** Alteration of wt *gal* sequence within 23 nt of O<sub>I</sub> causes RBT formation. (a) Effect of a 5-bp deletion on RBT. The sequence comparison of P2 and P2Δ5 DNAs shows the location of the 5-bp deletion (broken line). RNAs made from P2 (lanes 1 and 2) and P2Δ5 (lanes 3 and 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. Lanes 1 and 2, wt; lanes 3 and 4, ∇5A; lanes 5 and 6, ∇5B; lanes 7 and 8, ∇5C; lanes 9 and 10, ∇5D; lanes 11 and 12, ∇5; lanes 13 and 14, ∇5F.

positions +28 and +42. The increased length and upstream boundary of the transcription bubble are 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.<sup>17</sup>

#### Presence of antipause sequence element in *gal* DNA

As discussed above, the 5-bp insertion alters an intrinsic antipause signal encoded in wt *gal* sequence around position +37. This model predicts that other changes in this region (antipause DNA element) would also generate RBT. To test this

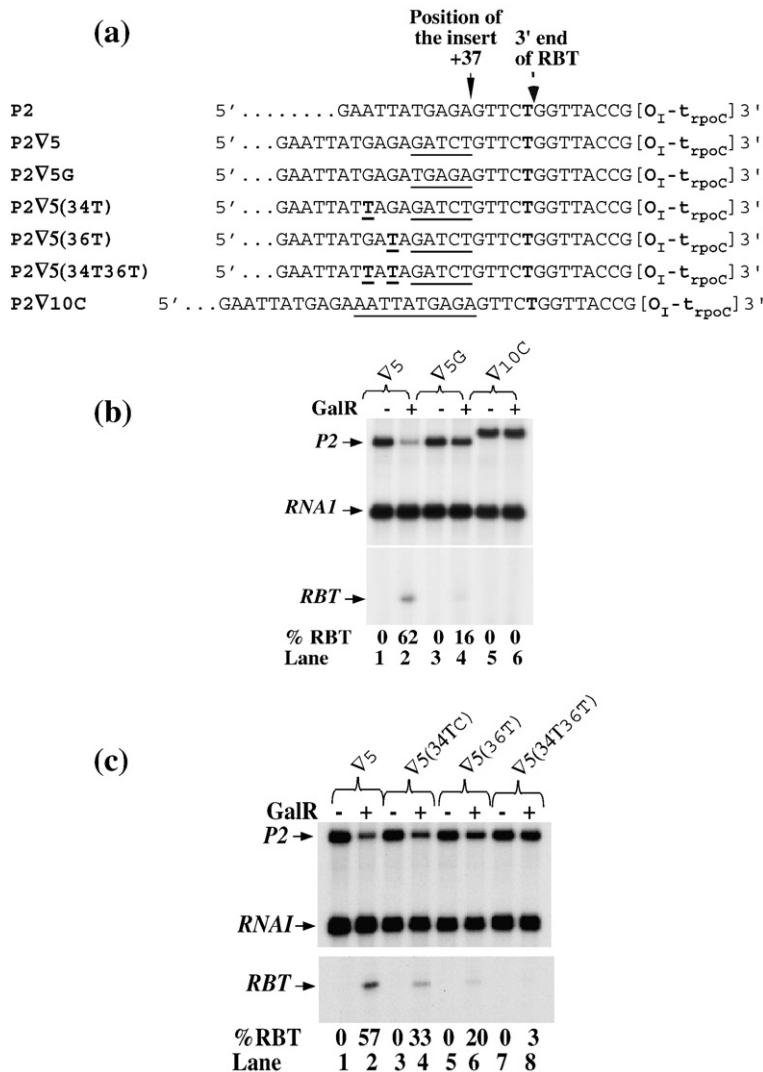
hypothesis, we investigated whether deleting 5 bp (from +36 to +41) of the wt DNA (*P2Δ5*) enhances RBT formation as *P2∇5* does. Whereas the full-length transcripts were unaffected by GalR, with no observation of RBT in the wt *P2* template (Fig. 6a, lanes 1 and 2), 47% of transcription was blocked by *O*<sub>I</sub>-bound GalR in the *P2Δ5* DNA (lanes 3 and 4). The RBT observed in *P2Δ5* 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 “antipause” sequence encoded around the +37 region in wt *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 antipause 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) (Fig. 6b). In the presence of GalR: (i) insertions at +14 (∇5A) and +28 (∇5B) 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*<sub>I</sub> in order to generate a 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 (lane 12 *versus* lane 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 (gagagatctgttct**gg**); however, in the case of +47, the 3' end is located at position +48 between G and A, or at position +49 between A and T (gagagtctgtggtta**gatctcc**).

To further localize the proposed antipause signal of wt *gal*, we inserted a TGAGA sequence at the same +37 site of *P2* (Fig. 7a, designated *P2∇5G*). The new template, despite the insertion, preserved the same 10-bp sequence (+<sup>37</sup>-TGAGAGTTCT-<sup>+47</sup>) upstream of the RBT 3' end as in the wt template;



**Fig. 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 (lanes 1 and 2, ∇5; lanes 3 and 4, ∇5G; lanes 5 and 6, ∇10C). (c) RNA made from templates containing G-to-T substitution at position +34 and/or position +36 (lanes 1 and 2, ∇5; lanes 3 and 4, ∇34T; lanes 5 and 6, ∇36T; lanes 7 and 8, ∇34T36T).

the sequence farther upstream was different. The TGAGA insertion shows significantly less RBT (16%) compared to the GATCT insertion (62% RBT) (Fig. 7b, lanes 2 and 4), showing that the 10-bp sequence favors RNAP elongation. We also tested the effect of a 10-bp sequence (AATTATGAGA) at position +37 ( $P2\triangledown10C$ ), which preserves the 15-bp sequence upstream of the RBT 3' end ( $^{+37}$ -AATTATGAGAGTTCT- $^{+52}$ ). The latter sequence completely prevented RBT formation (lanes 5 and 6). These results indicate that the antipause signal is located within the sequence 15 bp upstream of the RBT 3' end.

### Role of RNA:DNA hybrid in RNAP pausing

We investigated the mechanism by which an antipause element identified above prevents RBT by applying a computational approach. Our objective was to understand how the presence of the antipause element makes RNAP favor transcription elongation or how its absence makes RNAP favor pausing (backtracking) at the UUCU pause site. RNA:DNA hybrid is the major determinant of the lateral stability of the EC that determines forward (elongation) or backward (backtracking) movement of RNAP.<sup>16,18</sup> The hybrid is believed to be 8–9 bp long.<sup>37,38</sup> For a number of templates (Table 1), we calculated the predicted  $\Delta G_{37}^0$  of the 9-bp-long RNA:DNA hybrid<sup>39</sup> in the EC at the pause site, and the site backtracked by 4 and 7 nt. In wt *gal* sequence, the  $\Delta G$  value of the RNA:DNA hybrid in the EC at the pause site is  $-6.8$  kcal/mol; in the complex backtracked by 4 and 7 nt,  $\Delta G$  is  $-6.4$  and  $-4.7$  kcal/mol, respectively (Table 1, line 1). This difference in energy makes backtracking by 4 or 7 nt unfavorable. In the GATCT ( $\triangledown 5$ ) insertion,  $\Delta G$  of the RNA:DNA hybrid at the pause site is  $-5.5$  kcal/mol; in the complex backtracked by 4 and 7 nt, the corresponding  $\Delta G$  becomes  $-8.0$  and  $-7.8$  kcal/mol, respectively (line 11). The latter energy distribution makes backtracking by 4 or 7 nt favorable in  $\triangledown 5$ . Consistently, we observed no RBT in the first case, and we observed 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 to study their effects on RBT (Fig. 7c; Table 1, lines 6, 9, and 10). The RBT decreased from 57% to 33% on the 34T template (Fig. 7c, lanes 1–4; Table 1, line 10), to 20%

on the 36T template (lanes 5 and 6; line 9), and was eliminated on the 34T36T template (lanes 7 and 8; line 6). This result is consistent with the idea that the stability of the RNA:DNA hybrid determines whether 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 fourfold reduction in HU levels that decreases the chance of DNA looping.<sup>40</sup> In the absence of HU, *P2* transcripts are derepressed sixfold *in vivo*.<sup>6</sup> Previous studies showed that a stable  $O_1$ -bound GalR complex does not block an elongating RNAP in *gal*.<sup>6,8</sup> Since the enzymes of the *gal* operon are needed for biosynthetic glycosylation reactions, it is important that *gal* enzymes are made in the absence of D-galactose, and  $O_1$ -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.<sup>41</sup> Our current results confirmed that the  $O_1$ -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 wt *gal* DNA, transcription is not blocked. The ability of wt sequence to support transcription through DNA-bound GalR is easily compromised by mutations immediately upstream of  $O_1$ , leading to the formation of stable roadblocked complex.

RBT formation on mutant templates containing a 5-bp insertion (13 nt upstream of  $O_1$ ) is inhibited by GreB and partially inhibited 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 4–7 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  $O_1$ . In the roadblocked complex, RNAP stops just 8 bp upstream of  $O_1$ . This is a remarkably short distance since, according to hydroxyl radical footprinting and X-ray crystallography of the EC, RNAP covers at least 13 bp of DNA downstream of the catalytic center.<sup>36,42</sup> GalR was shown to cover an extra 3 bp on either side of the  $O_1$  sequence,<sup>4,43</sup> 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

**Table 1.** Calculations of the predicted  $\Delta G_{37}$  values of the 9-bp RNA:DNA hybrid

#	Plasmids	RNA sequence upstream of RBT 3' end	Predicted $\Delta G_{37}$ (kcal mol <sup>-1</sup> )			% RBT
			Pause site	Backtracked by 4-nt	Backtracked by 7-nt	
1	WT	GGAGCGAAUUAUGAGAGUUCU	-6.8	-6.4	-4.7	0
2	V5 $\Delta$ 5B	GGAGCGAAUUAUGAGAGUUCU	-7.7	-6.4	-4.7	0
3	V5 $\Delta$ 5C	GGAGCGAAUUAUGAUCUGUUCU	-5.5	-5.7	-4.9	0
4	V5 $\Delta$ 5D	GGAGCGAAUUAUUCUGUUCU	-5.5	-4.1	-3.3	0
5	V10C	AUGAGAAUUAUGAGAGUUCU	-6.8	-6.4	-4.2	0
6	V5(+34T36T)	GAAUUAUUAUGAUCUGUUCU	-5.5	-6.4	-4.1	0
7	V5D	GAAUUAUGAGAUUCUAGAUUCU	-6.2	-8.0	-7.1	11
8	V5G	GAAUUAUGAGAUAGAGAUUCU	-6.8	-8.7	-7.6	16
9	V5(+36T)	GAAUUAUGAUAGAUUCUGUUCU	-5.5	-6.4	-6.2	20
10	V5(+34T)	GAAUUAUAGAGAUUCUGUUCU	-5.5	-7.1	-5.7	33
11	V5	GAAUUAUGAGAGAUUCUGUUCU	-5.5	-8.0	-7.8	60
12	$\Delta$ 5	CUAAUGGAGCGAAUUAUUCU	-5.4	-6.4	-12.5	47
13	V5 $\Delta$ 5A	GGAGCGAGAGAGAUUCUGUUCU	-5.5	-8.0	-8.4	42
14	V5I	GAAUUAUGAGAUUAUUGGUUCU	-6.2	-7.1	-6.2	48
15	V5II	GAAUUAUGAGACGGCUGUUCU	-8.7	-11.9	-10.5	32
16	V5III	GAAUUAUGAGAGAUUCUGUUCU	-6.5	-10.6	-9.8	50
17	V5IV	GAAUUAUGAGAAAGUUGUUCU	-5.3	-6.7	-7.6	29
18	V5V	GAAUUAUGAGAGUGAUGUUCU	-6.0	-8.3	-8.3	32
19	V10B	AUGAGAUAGAGAUUCUGUUCU	-5.5	-8.0	-7.8	38
20	V10A	AUGAGAGAAUUAUGAGAGUUCU	-6.8	-7.3	-6.4	8
21	V5C	GAAUUAUGAUUCUGAGAGUUCU	-6.8	-7.3	-6.4	11
22	V5VI	GAAUUAUGAGAUUGGCGUUCU	-9.5	-10.0	-6.5	16

RNA sequence comparison of wt and mutant templates upstream of the 3' end of RBT. Predicted  $\Delta G_{37}$  values of RNA:DNA hybrid at pause sites and sites backtracked by 4 and 7 nt. The percent RBT for each template is shown. The pyrimidine-rich pause site (UUCU) and the upstream AU-rich (AAUUAU) sequence are outlined in blue; the purine-rich (GAGAG) sequence is outlined in red. G-to-U changes are in green. All sequences 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 wt as an antipause element, promoting RNAP elongation. When 5 bp is inserted between the GAGAG site and the pause site, RNAP backtracked from a weak hybrid to a stable hybrid and hindered elongation. (For interpretation of the references to colour in this table legend, the reader is referred to the web version of this article.)

stop RNAP during normal elongation.<sup>44</sup> The addition of the TFIIIS cleavage factor, a GreB homologue, substantially increased the ability of the enzyme to transcribe against the applied force.<sup>44</sup> 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 wt and  $\nabla$ 5 templates as it collides with GalR, it is the ability to escape from the pause in the wt that can be impaired by mutations as described above.

We argue that the relative strength of the regional RNA:DNA hybrid affects the fate of RNAP in wt and mutant templates. In both templates, four 3' proximal nucleotides in RNA (UUCU) weakly pair with the template, thereby favoring backtracking (Fig. 8). Since rPy:dPu base pairs are weaker than dPy:dPu base pairs,<sup>45</sup> reformation 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 RNAP in both wt and  $\nabla$ 5 templates (Fig. 8). However, in wt 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 (Fig. 8a1 and a2). Note that the GAGAG sequence produces rPu:dPy pairs, which are stronger than dPu:dPy pairs.<sup>45</sup> We believe that these properties of the wt 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 relocates the strong GAGAG-containing hybrid in the upstream segment of the backtracked hybrid, making the latter location preferable for RNAP in the mutant template (Fig. 8b1 and b2). Therefore, in the  $\nabla$ 5 template, RNAP stays longer at the upstream location in the absence of GalR and cannot escape the roadblock in the presence of GalR bound to  $O_1$ .

In summary, we have identified a new antipause element, GAGAG, located 13 bp upstream of  $O_1$  that 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, an elongating RNAP pauses and backtracks. 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 DH5 $\alpha$ <sup>TM</sup>-competent cells were obtained from Invitrogen. Recombinant RNasin<sup>®</sup> Ribonuclease Inhibitor (40 U/ $\mu$ l) was obtained from Promega. SequaGel sequencing system was obtained from National Diagnostics. NTPs (100 mM) and 3'-O-methylguanosine 5'-triphosphate were obtained from Amersham Pharmacia Biotech, Inc. 3'-o-methylcytidine 5'-triphosphate and 3'-o-methyluridine 5'-triphosphate were obtained from TriLink Biotechnologies. Primers were obtained from BioServe Biotechnologies and Sigma Genosys. Antisense oligonucleotides were obtained from Oligos Etc. XL PCR and DNA sequencing kits were obtained from Applied Biosystem.

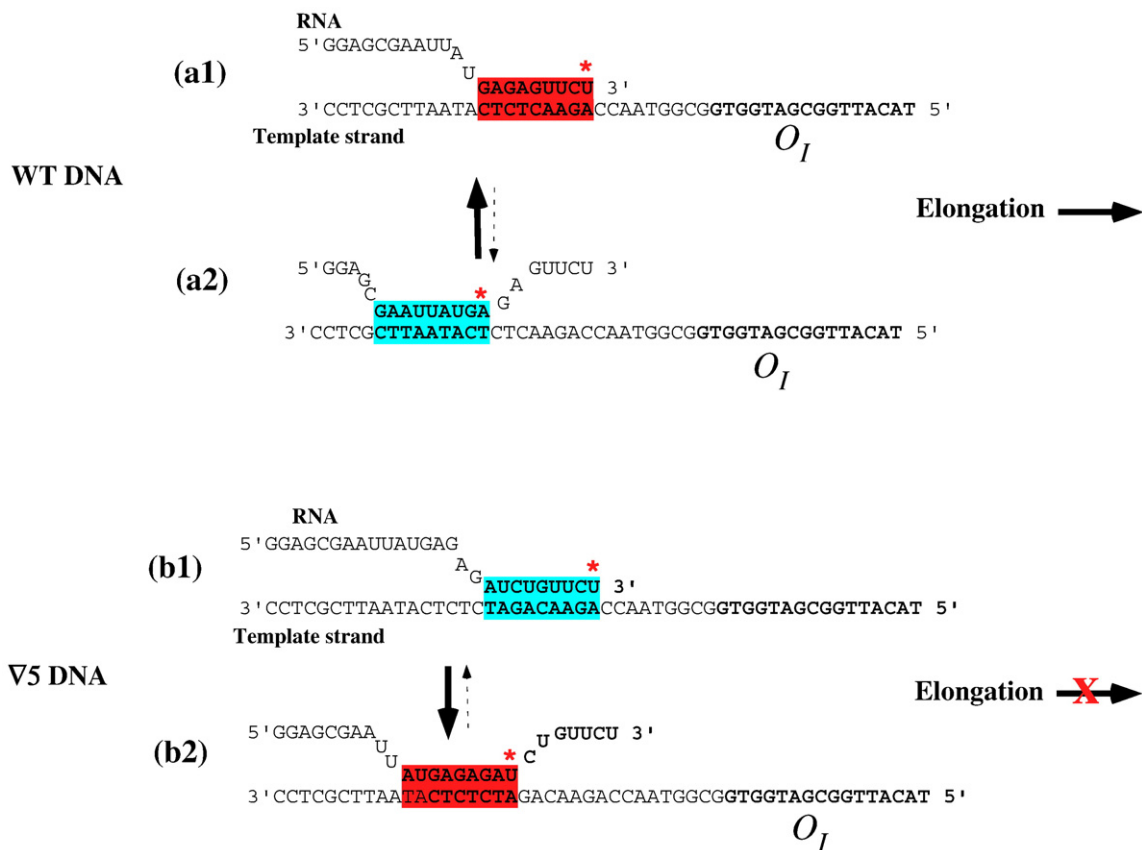
[ $\alpha$ -<sup>32</sup>P]Uridine triphosphate (UTP) and [ $\alpha$ -<sup>32</sup>P]CTP (specific activity: 3000 Ci/mmol; 10  $\mu$ Ci/ $\mu$ l) and [ $\gamma$ -<sup>32</sup>P]ATP (specific activity: 7000 Ci/mmol; 167  $\mu$ Ci/ $\mu$ l) were obtained from MP Biomedicals (formerly ICN Biochemicals, Inc.). [ $\alpha$ -<sup>32</sup>P]GTP (specific activity: 3000 Ci/mmol; 10  $\mu$ Ci/ $\mu$ l) was obtained from GE Healthcare.

### Other proteins and strains

*E. coli* RNAP (specific activity:  $1.85 \times 10^3$  to  $2.5 \times 10^3$  U/mg; concentration: 1 U/ $\mu$ l) was obtained from USB Corporation. GreA and GreB were gifts from Sergei Borukhov (University of Medicine and Dentistry of New Jersey, Stratford, NJ). GalR and cyclic AMP receptor protein (CRP) were purified as described.<sup>46,47</sup>

### Plasmids

The plasmids used in this study are listed in Table 2. The wt plasmid pSA850 ( $O_E^+P2^+P1^+O_1^+$ ), which contains the *gal* operator/promoter region from -70 to +96, was used as a PCR template to construct plasmids.<sup>10,11</sup> Plasmid pSA859 [ $O_E^+P2^+P1^+O_1^+$  (+5 bp)], or wt $\nabla$ 5, contains a 5-



**Fig. 8.** Intrinsic DNA sequence determining RNAP elongation or backtracking. The RNA:DNA hybrid at the pause site (a1) and site backtracked by 7 nt (a2) in wt 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 wt DNA, the hybrid maintains 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 wt DNA due to a weak RNA:DNA hybrid (blue). The solid arrow indicates that RNAP would remain more at the strong hybrid instead of at 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 backtracked from a weak hybrid (blue) to a stable hybrid (red), favoring RNAP arrest as indicated by the blocking of elongation by the red X.

bp (GATCT) insertion at position +37 with respect to the *tsp* of *P2* (+1). Plasmid pDL225 [ $O_E^+P2^+P1^+O_I^+$ ], or *P2*, contains two mutations (a G-to-A change at position -9 and a C-to-A change at position -8) that convert the -10 element of *P2* to the consensus -10 element of  $\sigma$ [70] promoter (TATGCT to TATAAT). Such changes inactivate *P1* by destroying the extended -10 of *P1*, resulting in maximal expression of *P2*.<sup>48</sup> Plasmid pDL1004 (*P2* $\nabla$ 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 [ $O_E^N P2^+ P1^+ O_I^+$  (+5 bp)], or *P1* $\nabla$ 5, contains several mutations: a single base-pair change from -14G to T inactivates *P2* by destroying the extended -10 of *P2*;<sup>49,50</sup> a double base-pair change from -4G and -3G to A, which converts the -10 element of *P1* to the consensus -10 element of  $\sigma$ [70] promoter, results in maximal expression of *P1*. Finally, the  $O_E$  (GTGTAAC-GATTCCAC) is replaced by a nonoperator sequence (CACTATGGCGAACGTC). Plasmid pDL571, or *T7A1* $\nabla$ 5, contains *T7* bacteriophage A1 promoter DNA (-56 to +24) fused with *gal* wt $\nabla$ 5 DNA downstream of +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.<sup>11</sup> 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, and 200 mM potassium glutamate) supplemented with 1 mM DTT, 1 mM ATP, and 0.8 U of rRNasin® in a total reaction volume of 50  $\mu$ l. 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  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P] UTP (3000 Ci/mmol), unless indicated otherwise. The

reactions were incubated for an additional 10 min before being terminated by the addition of an equal volume (50  $\mu$ l) of loading dye (90% formamide, 10 mM ethylenediamine-tetraacetic acid, 0.1% xylene cyanol, and 0.1% bromophenol blue). Transcripts were separated on 8% polyacrylamide sequencing gels. The *RNAI* transcripts driven by a GalR-independent promoter present in the plasmids (108 nt) were used as internal controls to quantify the relative amount of *gal* transcripts.<sup>51</sup> The *P1* and *P2* full-length transcripts contain 36 and 39 U residues, respectively. The RBTs from *P1* and *P2* contain 12 and 15 U residues, respectively. In mutant templates, the number of Us depended on the sequence. To calculate the percentage of RBT that was blocked by  $O_I$ -bound GalR, full-length transcripts and RBT were normalized to the *RNAI* transcripts and corrected for U content according to the formula: %RBT =  $B/(A+B) \times 100$ , where  $A$  = area of full length / (area of *RNAI*  $\times$  number of Us in full length) and  $B$  = area of RBT / (area of *RNAI*  $\times$  number of Us in RBT). When GTP was used as a labeled nucleotide (experiments of Fig. 3c and d), RBT was corrected for G content.

### Use of antisense oligonucleotides

*In vitro* transcription assays were conducted as described, except that 50  $\mu$ M oligonucleotides were present during 5 min of preincubation at 37 °C before NTPs were added to initiate transcription.

### GreA- and GreB-induced cleavage

To study the effects 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

**Table 2.** List of plasmids used in this study

Plasmid	Local name	Features	Source
pSA850	wt	a 166-bp <i>gal</i> DNA fragment (-75 to +91)	Lewis and Adhya <sup>10</sup>
pDL907	wt $\nabla$ 5A	pSA850 $\nabla$ GATCT at +14 of <i>P2</i> <i>tsp</i>	This work
pDL511	wt $\nabla$ 5B	pSA850 $\nabla$ GATCT at +28	This work
pDL512	wt $\nabla$ 5C	pSA850 $\nabla$ GATCT at +32	This work
pDL513	wt $\nabla$ 5D	pSA850 $\nabla$ GATCT at +35	This work
pSA859	wt $\nabla$ 5E	pSA850 $\nabla$ GATCT at +37	Lewis and Adhya <sup>10</sup>
pDL908	wt $\nabla$ 5F	pSA850 $\nabla$ GATCT at +47	This work
pDL229	<i>P1</i> con	pSA850 (-15T, -4A, and -3A)	This work
pDL994	<i>P1</i> con $\nabla$ 5	pDL229 $\nabla$ GATCT at +37	This work
pDL225	<i>P2</i> con	pSA850 (-8A and -9A)	This work
pDL1022	<i>P2</i> con $\Delta$ 5	pDL225 $\Delta$ GAGTT (+36 to +40)	This work
pDL1004	<i>P2</i> con $\nabla$ 5	pDL225 $\nabla$ GATCT at +37	This work
pDL1039	<i>P2</i> con $\nabla$ 5A	pDL225 $\nabla$ TATTG at +37	This work
pDL1040	<i>P2</i> con $\nabla$ 5B	pDL225 $\nabla$ CGGCT at +37	This work
pDL1041	<i>P2</i> con $\nabla$ 5C	pDL225 $\nabla$ GGACT at +37	This work
pDL1042	<i>P2</i> con $\nabla$ 5D	pDL225 $\nabla$ AAGTT at +37	This work
pDL1043	<i>P2</i> con $\nabla$ 5E	pDL225 $\nabla$ GTGAT at +37	This work
pDL1044	<i>P2</i> con $\nabla$ 5F	pDL225 $\nabla$ TTGGC at +37	This work
pDL1060	<i>P2</i> con $\nabla$ 5G	pDL225 $\nabla$ TGAGA at +37	This work
pDL1061	<i>P2</i> con $\nabla$ 10C	pDL225 $\nabla$ AATTATGAGA at +37	This work
pDL1062	<i>P2</i> con $\nabla$ 10A	pDL225 $\nabla$ GATCTTGAGA at +37	This work
pDL1063	<i>P2</i> con $\nabla$ 10B	pDL225 $\nabla$ TGAGAGATCT at +37	This work
pDL1011	<i>P2</i> con $\nabla$ 5 $\Delta$ 5A	pDL1004 [ $\Delta$ ATTAT (+29 to +33)]	This work
pDL1012	<i>P2</i> con $\nabla$ 5 $\Delta$ 5B	pDL1004 [ $\Delta$ GTTCT (+43 to +47)]	This work
pDL1079	<i>P2</i> con $\nabla$ 5 $\Delta$ 5C	pDL1004 [ $\Delta$ TGAGA (+33 to +37)]	This work
pDL1080	<i>P2</i> con $\nabla$ 5 $\Delta$ 5D	pDL1004 [ $\Delta$ GAGAG (+34 to +38)]	This work
pDL999	<i>P2</i> con $\nabla$ 5 $O_I^N$	pDL1004 ( $O_I^N$ )	This work
pDL571	<i>T7A1</i> $\nabla$ 5	pSA859 [ <i>T7A1</i> promoter DNA (-56 to +24)]	This work

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 [ $\alpha$ - $^{32}$ P]UTP for the control experiments. Unlabeled RBTs 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 in accordance with the protocol for AMV reverse transcriptase (Promega) using  $\gamma$ - $^{32}$ P-labeled primer (5'-TCTCATAATTCGCTCC-3'), which mapped from +37 to +22. The DNA was sequenced by PCR using the same primer in accordance with the protocol for *fmol*<sup>®</sup> 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. A linear DNA template (355 bp) generated by PCR from *P1*∇5 plasmid was used to map the 3' ends of RBT as described.<sup>11</sup> CRP (50 nM) and cAMP (100  $\mu$ M) were added to the transcription reaction to maximize the strength of *P1* promoter (in the absence of a cAMP-CRP complex, *P1* activity on the linear template is very weak). To generate a ladder of nucleotides, transcription was carried out for 10 min in the presence of 0.25  $\mu$ M 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 His-tagged RNAP (0.12  $\mu$ g) on Ni-NTA agarose and walking along the DNA (1  $\mu$ g) as described.<sup>16,17</sup> A stable complex was stalled at position 12 (EC12) after initiation with 1 mM CUAU (a tetranucleotide primer), 50  $\mu$ M ATP, and 50  $\mu$ M GTP at 37 °C for 5 min. Complexes were washed three to five times in 1 ml of transcription buffer, TB (20 mM Tris-HCl pH 7.9, 40 mM KCl, 5 mM MgCl<sub>2</sub>, and 1 mM  $\beta$ -mercaptoethanol). The RNA was labeled at EC13 with 3  $\mu$ l of [ $\alpha$ - $^{32}$ P]CTP (3000 Ci/mmol) and incubated at room temperature for 5 min. The complexes were washed as described above, and a subset of NTPs (5  $\mu$ M) was added to walk RNAP to the desired position; the process of walking and washing was repeated.

### Mapping of transcription bubble by KMnO<sub>4</sub> footprinting

The sequences of wt and *P2*∇5 DNA templates for the KMnO<sub>4</sub> assay were determined in accordance with the *fmol*<sup>®</sup> DNA cycle sequencing system protocol (Promega). The nontemplate strand was sequenced using [ $\gamma$ - $^{32}$ P]XbaI-5 primer (5'-TCAACGGAGCTCGTCC-3'), which mapped from -108 to -93. The template strand was sequenced

using [ $\gamma$ - $^{32}$ P]BamHI-10 primer (5'-GCGGATCCC-TAAACTC-3'), which mapped from +161 to +141. Approximately 10  $\mu$ g/ $\mu$ l DNA was sequenced using "profile 1" from *fmol*<sup>®</sup> DNA cycle sequencing system protocol. The KMnO<sub>4</sub> footprinting was performed as described.<sup>33,52</sup> *In vitro* transcription assays were performed as described above with 10 nM supercoiled DNA. At the end of the transcription reactions, 10 mM KMnO<sub>4</sub> was added for 2–3 min at 37 °C. The reactions were quenched with  $\beta$ -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  $\mu$ l of DNA, 20  $\mu$ l of PCR Master mix (Promega), and 2  $\mu$ l (2 pmol) of [ $\gamma$ - $^{32}$ P]BamHI-10 or [ $\gamma$ - $^{32}$ P]XbaI-5 primers were used to amplify the product using *fmol*<sup>®</sup> DNA cycle sequencing. The products were analyzed on an 8% sequencing gel.

## Acknowledgements

This work was supported by the Intramural Research Program of the National Institutes of Health, the 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.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2008.07.060](https://doi.org/10.1016/j.jmb.2008.07.060)

## References

- Adhya, S. & Miller, W. (1979). Modulation of the two promoters of the galactose operon of *Escherichia coli*. *Nature*, **279**, 492–494.
- Musso, R. E., Di Lauro, R., Adhya, S. & de Crombrughe, B. (1977). Dual control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. *Cell*, **12**, 847–854.
- Aiba, H., Adhya, S. & de Crombrughe, B. (1981). Evidence for two functional gal promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* **256**, 11905–11910.
- Aki, T. & Adhya, S. (1997). Repressor induced site-specific binding of HU for transcriptional regulation. *EMBO J.* **16**, 3666–3674.
- Aki, T., Choy, H. E. & Adhya, S. (1996). Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. *Genes Cells*, **1**, 179–188.
- Lewis, D. E., Geanakopoulos, M. & Adhya, S. (1999). Role of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at gal promoters in *Escherichia coli*. *Mol. Microbiol.* **31**, 451–461.
- Goodrich, J. A. & McClure, W. R. (1992). 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.* **224**, 15–29.
8. Choy, H. E., Park, S. W., Aki, T., Parrack, P., Fujita, N., Ishihama, A. & Adhya, S. (1995). Repression and activation of transcription by Gal and Lac repressors: involvement of alpha subunit of RNA polymerase. *EMBO J.* **14**, 4523–4529.
  9. Roy, S., Garges, S. & Adhya, S. (1998). Activation and repression of transcription by differential contact: two sides of a coin. *J. Biol. Chem.* **273**, 14059–14062.
  10. Lewis, D. E. & Adhya, S. (2002). *In vitro* repression of the gal promoters by GalR and HU depends on the proper helical phasing of the two operators. *J. Biol. Chem.* **277**, 2498–2504.
  11. Lewis, D. E. (2003). Identification of promoters of *Escherichia coli* and phage in transcription section plasmid pSA850. *Methods Enzymol.* **370**, 618–645.
  12. Landick, R. (1997). RNA polymerase slides home: pause and termination site recognition. *Cell*, **88**, 741–744.
  13. Landick, R. (2001). RNA polymerase clamps down. *Cell*, **105**, 567–570.
  14. Mooney, R. A., Artsimovitch, I. & Landick, R. (1998). Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.* **180**, 3265–3275.
  15. Toulme, F., Guerin, M., Robichon, N., Leng, M. & Rahmouni, A. R. (1999). *In vivo* evidence for back and forth oscillations of the transcription elongation complex. *EMBO J.* **18**, 5052–5060.
  16. Komissarova, N. & Kashlev, M. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc. Natl Acad. Sci. USA*, **94**, 1755–1760.
  17. Komissarova, N. & Kashlev, M. (1997). RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J. Biol. Chem.* **272**, 15329–15338.
  18. Palangat, M. & Landick, R. (2001). Roles of RNA: DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II. *J. Mol. Biol.* **311**, 265–282.
  19. Komissarova, N., Velikodvorskaya, T., Sen, R., King, R. A., Banik-Maiti, S. & Weisberg, R. A. (2008). Inhibition of a transcriptional pause by RNA anchoring. *Mol. Cell*, In the press. doi:10.1016/j.molcel.2008.06.019.
  20. Komissarova, N. & Kashlev, M. (1998). Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl Acad. Sci. USA*, **95**, 14699–14704.
  21. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.
  22. Borukhov, S., Polyakov, A., Nikiforov, V. & Goldfarb, A. (1992). GreA protein: a transcription elongation factor from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **89**, 8899–8902.
  23. Borukhov, S., Sagitov, V. & Goldfarb, A. (1993). Transcript cleavage factors from *E. coli*. *Cell*, **72**, 459–466.
  24. Orlova, M., Newlands, J., Das, A., Goldfarb, A. & Borukhov, S. (1995). Intrinsic transcript cleavage activity of RNA polymerase. *Proc. Natl Acad. Sci. USA*, **92**, 4596–4600.
  25. Koulich, D., Orlova, M., Malhotra, A., Sali, A., Darst, S. A. & Borukhov, S. (1997). Domain organization of *Escherichia coli* transcript cleavage factors GreA and GreB. *J. Biol. Chem.* **272**, 7201–7210.
  26. Sparkowski, J. & Das, A. (1990). The nucleotide-sequence of *greA*, a suppressor gene that restores growth of an *Escherichia coli* RNA-polymerase mutant at high-temperature. *Nucleic Acids Res.* **18**, 6443.
  27. Sparkowski, J. & Das, A. (1991). Location of a new gene, *greA*, on the *Escherichia coli* chromosome. *J. Bacteriol.* **173**, 5256–5257.
  28. Lee, D. N., Feng, G. & Landick, R. (1994). GreA-induced transcript cleavage is accompanied by reverse translocation to a different transcription complex conformation. *J. Biol. Chem.* **269**, 22295–22303.
  29. Toulme, F., Mosrin-Huaman, C., Sparkowski, J., Das, A., Leng, M. & Rahmouni, A. R. (2000). GreA and GreB proteins revive backtracked RNA polymerase *in vivo* by promoting transcript trimming. *EMBO J.* **19**, 6853–6859.
  30. Borukhov, S., Laptenko, O. & Lee, J. (2001). *Escherichia coli* transcript cleavage factors GreA and GreB: functions and mechanisms of action. *Methods Enzymol.* **342**, 64–76.
  31. Rutherford, S. T., Lemke, J. J., Vrentas, C. E., Gaal, T., Ross, W. & Gourse, R. L. (2007). 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.* **366**, 1243–1257.
  32. Loizos, N. & Darst, S. A. (1999). Mapping interactions of *Escherichia coli* GreB with RNA polymerase and ternary elongation complexes. *J. Biol. Chem.* **274**, 23378–23386.
  33. Sasse-Dwight, S. & Gralla, J. D. (1989). KMnO<sub>4</sub> as a probe for lac promoter DNA melting and mechanism *in vivo*. *J. Biol. Chem.* **264**, 8074–8081.
  34. Kashlev, M. & Komissarova, N. (2002). Transcription termination: primary intermediates and secondary adducts. *J. Biol. Chem.* **277**, 14501–14508.
  35. Kubori, T. & Shimamoto, N. (1996). A branched pathway in the early stage of transcription by *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **256**, 449–457.
  36. Zaychikov, E., Denissova, L. & Heumann, H. (1995). 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. USA*, **92**, 1739–1743.
  37. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. (1997). The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, **89**, 33–41.
  38. Sidorenkov, I., Komissarova, N. & Kashlev, M. (1998). Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol. Cell*, **2**, 55–64.
  39. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T. *et al.* (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.
  40. Azam, T. A., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* **181**, 6361–6370.
  41. Deuschle, U., Gentz, R. & Bujard, H. (1986). Lac repressor blocks transcribing RNA-polymerase and terminates transcription. *Proc. Natl Acad. Sci. USA*, **83**, 4134–4137.
  42. Vassylyev, D. G., Vassylyeva, M. N., Zhang, J. W., Palangat, M., Artsimovitch, I. & Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, **448**, 163; U4.
  43. Majumdar, A. & Adhya, S. (1987). Probing the structure of gal operator–repressor complexes.

- Conformation change in DNA. *J. Biol. Chem.* **262**, 13258–13262.
44. Galburt, E. A., Grill, S. W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E. *et al.* (2007). Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature*, **446**, 820–823.
  45. Hung, S. H., Yu, Q., Gray, D. M. & Ratliff, R. L. (1994). Evidence from CD-spectra that D(purine)·R(pyrimidine) and R(purine)·D(pyrimidine) hybrids are in different structural classes. *Nucleic Acids Res.* **22**, 4326–4334.
  46. Majumdar, A., Rudikoff, S. & Adhya, S. (1987). Purification and properties of Gal repressor:pL-galR fusion in pKC31 plasmid vector. *J. Biol. Chem.* **262**, 2326–2331.
  47. Ryu, S., Kim, J., Adhya, S. & Garges, S. (1993). Pivotal role of amino acid at position 138 in the allosteric hinge reorientation of cAMP receptor protein. *Proc. Natl Acad. Sci. USA*, **90**, 75–79.
  48. Busby, S., Truelle, N., Spassky, A., Dreyfus, M. & Buc, H. (1984). The selection and characterisation of two novel mutations in the overlapping promoters of the *Escherichia coli* galactose operon. *Gene*, **28**, 201–209.
  49. Bingham, A. H., Ponnambalam, S., Chan, B. & Busby, S. (1986). Mutations that reduce expression from the P2 promoter of the *Escherichia coli* galactose operon. *Gene*, **41**, 67–74.
  50. Johnston, F., Ponnambalam, S. & Busby, S. (1987). Binding of *Escherichia coli* RNA polymerase to a promoter carrying mutations that stop transcription initiation. *J. Mol. Biol.* **195**, 745–748.
  51. Tomizawa, J., Itoh, T., Selzer, G. & Som, T. (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Natl Acad. Sci. USA*, **78**, 1421–1425.
  52. Liu, M., Tolstorukov, M., Zhurkin, V., Garges, S. & Adhya, S. (2004). A mutant spacer sequence between –35 and –10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. *Proc. Natl Acad. Sci. USA*, **101**, 6911–6916.