The EMBO Journal (2012) 31, 630–639 | © 2012 European Molecular Biology Organization | Some Rights Reserved 0261-4189/12
www.embojournal.org

THE EMBO JOURNAL

EMBO open

Factor-independent transcription pausing caused by recognition of the RNA–DNA hybrid sequence

Aleksandra Bochkareva[1,3], Yulia Yuzenkova[1,3], Vasisht R Tadigotla[2] and Nikolay Zenkin[1,*]

[1]Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK and [2]Department of Physics, Boston University, Boston, MA, USA

Pausing of transcription is an important step of regulation of gene expression in bacteria and eukaryotes. Here we uncover a factor-independent mechanism of transcription pausing, which is determined by the ability of the elongating RNA polymerase to recognize the sequence of the RNA–DNA hybrid. We show that, independently of thermodynamic stability of the elongation complex, RNA polymerase directly 'senses' the shape and/or identity of base pairs of the RNA–DNA hybrid. Recognition of the RNA–DNA hybrid sequence delays translocation by RNA polymerase, and thus slows down the nucleotide addition cycle through 'in pathway' mechanism. We show that this phenomenon is conserved among bacterial and eukaryotic RNA polymerases, and is involved in regulatory pauses, such as a pause regulating the production of virulence factors in some bacteria and a pause regulating transcription/replication of HIV-1. The results indicate that recognition of RNA–DNA hybrid sequence by multi-subunit RNA polymerases is involved in transcription regulation and may determine the overall rate of transcription elongation.
The EMBO Journal (2012) **31,** 630–639. doi:10.1038/emboj.2011.432; Published online 29 November 2011
Subject Categories: chromatin & transcription
Keywords: elongation complex; RNA–DNA hybrid; RNA polymerase core; transcription pausing; translocation

Introduction

The first step of gene expression, transcription, in all living organisms is accomplished by multi-subunit RNA polymerases (RNAPs). Elongation by catalytic core of RNAP is highly processive and can proceed for thousands of base pairs without dissociation from the template. RNAP core is thought to be mostly indifferent to the sequence of the nucleic acids to allow processive and uninterrupted transcription elongation. This was suggested to be one of the reasons why RNAPs require accessory proteins to recognize specific sequences in DNA during initiation of transcription (Zenkin and Severinov, 2008). However, control of the rate of transcription elongation is now recognized as an important step in regulation of gene expression. This regulation is accomplished through pausing of transcription.

Several types of signals were shown to cause transcription elongation pausing: misalignment of the incoming nucleotide (Kireeva and Kashlev, 2009), sequences resembling promoter elements that are recognized by σ-factor that fails to dissociate from RNAP (Ring *et al*, 1996; Brodolin *et al*, 2004; Nickels *et al*, 2004), hairpin formation in the nascent RNA behind transcribing RNAP (Chan and Landick, 1989), or by thermodynamic instability of the elongation complex (EC) that causes backtracking (Bai *et al*, 2004; Tadigotla *et al*, 2006; Maoileidigh *et al*, 2011). RNAP pausing is an important mechanism of transcriptional regulation both in prokaryotes and eukaryotes (Landick, 2006). Promoter-proximal σ-dependent pausing is crucial for the regulation of phage development, and was also recently suggested to be a common way of regulation of bacterial gene expression (Hatoum and Roberts, 2008; Perdue and Roberts, 2011). Regulation of expression of bacterial operons *his* (Chan and Landick, 1989), *trp* (Landick *et al*, 1987), *pyr* (Donahue and Turnbough, 1994), and some other operons (Yanofsky, 1981) in response to changes in supply of nutrients is mediated by hairpin-dependent transcription pausing. Coupling of transcription and translation (Gong and Yanofsky, 2003), co-transcriptional RNA folding (Wong *et al*, 2007), binding of elongation regulators (Borukhov *et al*, 2005) and both ρ-dependent (Ciampi, 2006) and intrinsic (Farnham and Platt, 1981) termination require RNAP pausing. It was shown that RNAP II pausing may assist co-transcriptional splicing in *Saccharomyces cerevisiae* (Alexander *et al*, 2010).

All transcription pauses are thought to be off pathway (a non-obligatory) events (Herbert *et al*, 2008; Landick, 2009). This means that, upon pause signal recognition, the EC branches off the nucleotide addition cycle (NAC). NAC is the elemental step of transcription elongation. NAC starts with formation of the phosphodiester bond between nascent transcript and NTP bound in the i + 1 site of RNAP. After NMP incorporation, EC oscillates in a thermal equilibrium between pre- and post-translocated states (Figure 1). The post-translocated state is stabilized by binding of the incoming NTP in the vacated i + 1 site, which ensures unidirectional movement of RNAP. This step is followed by the repetition of the NAC. RNAP can turn away from the pathway of the NAC, thus leading to a pause of transcription. This can happen through backtracking of the EC, when the 3′ end of RNA disengages from the RNAP active centre, or via a conformational change in the RNAP active centre, which slows down catalysis (Landick, 2009). In fact, backtracking by 1 bp (base pair), referred here as 'backstepping' (Cramer, 2006), is believed to be a stable translocation state and co-exist in equilibrium with pre- and post-translocated states (Wang *et al*, 2009; Roghanian *et al*, 2011). Pause signals are thought to cause kinetic partitioning of RNAPs between

*Corresponding author. Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Baddiley-Clark Building, Richardson Road, Newcastle upon Tyne NE2 4AX, UK.
Tel.: + 44 1912083227; Fax: + 44 1912083205;
E-mail: n.zenkin@ncl.ac.uk
[3]These authors contributed equally to this work

Received: 13 September 2011; accepted: 7 November 2011; published online: 29 November 2011

 ©2012 European Molecular Biology Organization

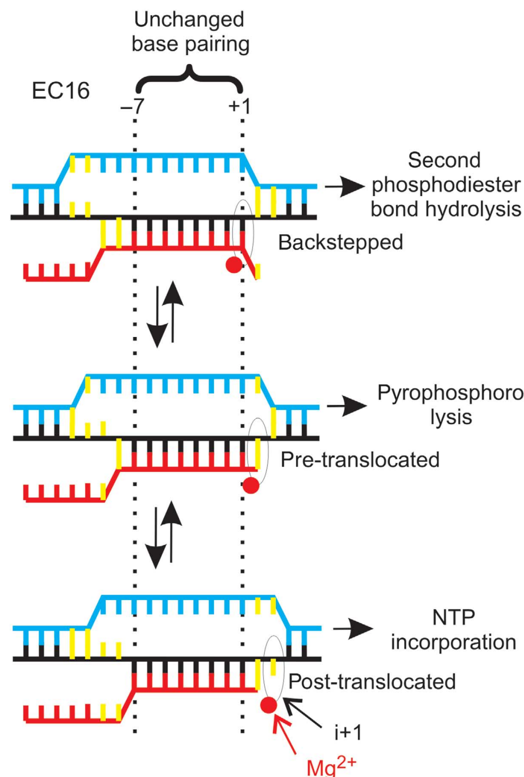

Figure 1 Translocation oscillation of the elongation complex. Scheme of RNAP oscillation in translocation equilibrium and the architecture of the nucleic acids scaffold of the elongation complex in post-translocated, pre-translocated and backstepped states, as follows from our results (Supplementary Figure 1). The template DNA, the non-template DNA and the RNA are black, blue and red, respectively. Base pairs that break or form during oscillation between these translocation states are shown in yellow. Catalytic Mg^{2+} ions and the $i+1$ site of the RNAP active centre are shown by a red circle and an oval, respectively. Significant backtracking further than 1 bp was not observed in our experiments (Supplementary Figure 1) and is omitted from the scheme.

active and paused states, as opposed to affecting all transcribing RNAPs uniformly (Landick, 2009). This means that some RNAPs would stay *in* the pathway and not respond to the pause signals. 'In pathway' pauses that are caused by delays in the steps within the NAC have not been demonstrated so far.

Pre-translocated (in pathway) pauses, however, are intuitively acceptable and were proposed to exist by kinetic modelling (Bai *et al*, 2004). An example of a translocation pause is the σ-dependent pause. It however requires an accessory factor (σ). Here we show that RNAP core is able to recognize the sequence throughout the length of the RNA–DNA hybrid, and that this may determine the rate of translocation and, as a result, cause an 'in pathway' pause of transcription. We show that this phenomenon is conserved among multi-subunit RNAPs and may be involved in regulation of some physiological processes such as production of virulence factors in some bacteria and transcription and replication of HIV-1.

Results

Analysis of translocation equilibrium of the EC

The aim of our study was to test if there exist sequences that may cause pausing of transcription via recognition by RNAP core. We argued that such sequences, if they exist, must change translocation equilibrium of the EC. Thermodynamics modelling (Tadigotla *et al*, 2006) may subsequently exclude sequences that affect translocation equilibrium through thermodynamics of nucleic acids scaffold, rather than via sequence recognition by RNAP. Monitoring RNAP reactions that depend on the translocation equilibrium is a sensitive method for search of such sequences (Kashkina *et al*, 2006; Kent *et al*, 2009). We chose analysis of the rate of second (penultimate from the 3′ end of RNA; Figure 1) phosphodiester bond hydrolysis, which proceeds only from backstepped (1 bp backtracked) state, and which is relatively slow, being suitable for manual measurements (compared with NTP incorporation and pyrophosphorolysis (see below)). The backstepped state was shown to co-exist in equilibrium with the pre- and post-translocated states, and the rate of second phosphodiester bond hydrolysis was shown to directly depend on the translocation equilibrium (Zenkin *et al*, 2006; Kent *et al*, 2009; Roghanian *et al*, 2011).

We used ECs assembled with *Escherichia coli* RNAP core (Figure 2A). Complexes were similar to those used in our and other studies previously (Sidorenkov *et al*, 1998; Yuzenkova *et al*, 2010), and are based on the sequence of the transcribed region of the T7A1 promoter (the initial sequence will be referred to as 'wild type', wt). Complexes were assembled with 13 nucleotide long RNA (EC13) and then walked by three positions to form EC16 (to allow free translocational oscillation and backtracking). RNA was labelled by incorporation of α-[^{32}P]-NMP at the 3′ end, allowing for subsequent monitoring of the cleavage products (Figure 2A).

Translocation equilibrium strongly depends on base pairing at the edges of the RNA–DNA hybrid and transcription bubble (Zenkin *et al*, 2006; Kent *et al*, 2009). To exclude these effects from our analysis, we decided to focus on the region within the length of the RNA–DNA hybrid where, during translocational oscillation, base pairs remain either melted or paired, and where the strength of base pairing is not expected to influence translocation equilibrium significantly. Therefore, first, we determined the lengths of the transcription bubble and the RNA–DNA hybrid in the EC. To do that, we systematically introduced mismatches in the nucleic acids of the EC ($EC16^{WT}$) and measured the rate of second phosphodiester bond hydrolysis (Supplementary Figure 1). Mismatches in base pairs that break or form during RNAP oscillation (yellow in Figure 1) would affect the translocation equilibrium, while mismatches in positions that remain paired or unpaired during oscillation would not. The scheme of the EC structure based on our results (Supplementary Figure 1) is shown in Figure 1. As can be seen, 12–13 bp of the DNA duplex are melted, while the length of the RNA–DNA hybrid in the backstepped state is 10 bp. Importantly, the obtained results are fully consistent with the current understanding of the structure of the transcription EC (Kashkina *et al*, 2007; Vassylyev *et al*, 2007a, b), thus validating the experimental system used in our study. Note, that no significant backtracking further than 1 base pair was observed (as judged by the absence of cleavage products longer than dinucleotide), though it is not restricted by the structure of the scaffold. This is consistent with observation that backstepped complex is energetically stable translocation state (Wang *et al*, 2009).

As follows from the above results (Figure 1), the base pairing in the regions −7 to +1 (relative to the position of the RNAP active centre in the backstepped state) does not change during translocational oscillation. To confirm that the changes of the sequence in this region are not expected to cause any significant thermodynamic effects on translocation equilibrium, we calculated relative probabilities of RNAP backstepping in ECs bearing substitutions in the region of the RNA–DNA hybrid (substitutions were made simultaneously in the template and non-template strands and the RNA so that no mismatches were introduced) using previously described thermodynamics model of the EC (Tadigotla *et al*, 2006). This model takes into account the energies of base pairing and stacking interactions within the nucleic acids scaffold. Relative probabilities of backstepping are shown in logarithmic scale in Figure 2B, top (normalized to the probability of backstepping in $EC16^{WT}$, which was taken as 1). Changes of sequence in the region between positions −7 and +1 indeed had no or very little effect on probability of RNAP backstepping, comparing, for example, to the changes in position −8. Strong base pairing in position −8 is expected to favour backstepping, as the RNA–DNA hybrid has to form at this position for the shift into the backstepped state to occur.

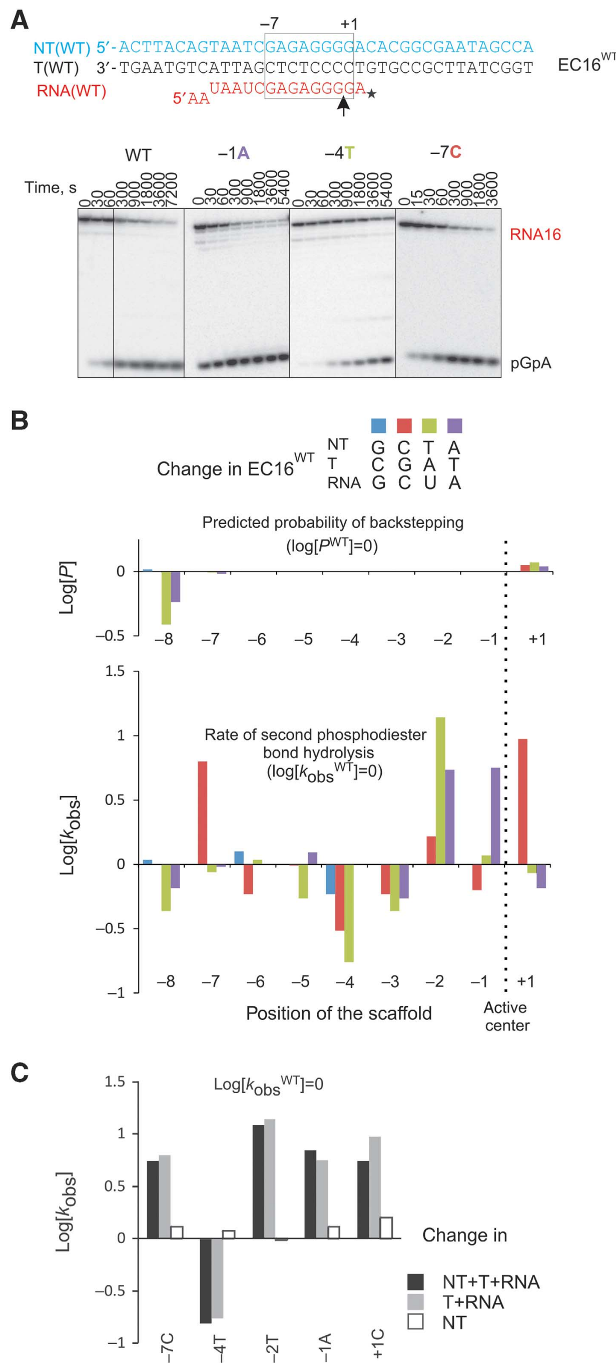

Figure 2 RNAP recognizes the sequence of the RNA–DNA hybrid. (**A**) Representative gels of hydrolytic reactions in elongation complexes bearing changes in the region of the RNA–DNA hybrid (here and after, substitutions are depicted according to the sequence of the non-template DNA). Sequences of the nucleic acids of the $EC16^{WT}$ are shown at the top of the panel (see Supplementary Methods). The arrow shows the second phosphodiester bond hydrolysed in the backstepped state, and the frame shows the region of the nucleic acids scaffold used in the substitution analysis. Substitutions in $EC16^{WT}$ were introduced in the RNA, template and non-template DNA strands simultaneously, so that no mismatches were generated. (**B**) The top histogram shows probabilities of RNAP occurrence in the backstepped state in EC16s bearing substitutions, calculated by using a thermodynamic model of the elongation complex (normalized to the probability of backstepping in $EC16^{WT}$, which was taken as 1). The bottom histogram shows the rates of second phosphodiester bond hydrolysis, k_{obs} (as a measure of RNAP backstepping), in EC16s bearing substitutions (normalized to the rate of the reaction in $EC16^{WT}$, which was taken as 1). (**C**) RNAP recognizes sequence of the RNA–DNA hybrid but not the sequence of the non-template DNA. Rates of second phosphodiester bond hydrolysis (presented as in panel **B**) in elongation complexes bearing substitutions either in both, the RNA–DNA hybrid and the non-template strand (black), or in the RNA–DNA hybrid alone (grey), or in the non-template strand alone (white).

Sequence of the RNA–DNA hybrid can influence translocation equilibrium

In our *in vitro* analysis we focused on the region between positions −7 and +1 of the nucleic acids scaffold, where sequence does not influence translocation equilibrium according to theoretical predictions, and used position −8 (which influences equilibrium according to thermodynamics modelling) as a positive control. The rates of second phosphodiester bond hydrolysis in EC16s bearing systematic substitutions are presented in logarithmic scale in Figure 2B, bottom (normalized to the rate of phosphodiester bond cleavage in $EC16^{WT}$, which was taken as 1). As expected from thermodynamics modelling, strong base pairing at position −8 increased the rate of second phosphodiester bond cleavage (note the match of theoretical and experimental results), while the most of the changes between positions −7 and +1 affected the reaction relatively weakly (less than two-fold as compared with $EC16^{WT}$). However, in contrast to theoretical predictions, the *in vitro* analysis revealed that several substitutions between positions −7 and +1 led to a significant increase (up to 13-fold) or decrease (up to 6-fold) of the rate of the reaction (Figure 2B, bottom). Substitutions +1C, −1A, −2A, −2T, −7C (here and after sequences of nucleic acids in the EC are given according the sequence of the non-template DNA strand) strongly stabilize, and −3T, −4T, −4C destabilize the backstepped state of the $EC16^{WT}$ (see also Supplementary Figure 2A). However, as mentioned

©2012 European Molecular Biology Organization

above, these effects cannot be explained by changes in thermodynamic properties of the EC (Figure 2B, top).

In the crystal structure of the EC, amino acids of the RNAP main channel are in intimate contacts with the nucleic acids along the length of the RNA–DNA hybrid (Vassylyev *et al*, 2007a). Our results, therefore, indicate that RNAP may 'sense' or 'recognize' the sequence of the nucleic acids scaffold (chemical groups of the bases and/or small differences in the geometries of base pairs; see Discussion and Supplementary Figure 2B), which leads to a shift in translocation equilibrium and, as a result, to an altered rate of hydrolysis of the second phosphodiester bond. Note also that sequence recognition may affect backstepping stronger than thermodynamic stability of the EC (Figure 2B, bottom; compare effects of changes at position −8 to those in positions −7, −4, −2, −1 and +1).

From the above it is unclear if RNAP recognizes the sequence of the non-template strand or the sequence of the RNA–DNA hybrid of the EC. Therefore, we analysed hydrolysis in complexes bearing substitutions in either the non-template strand alone or the RNA–DNA hybrid alone. We tested substitutions that had the most pronounced effect on the rate of hydrolysis (−7C, −4T, −2T, −1A, +1C). As seen from Figure 2C, substitutions in the non-template strand alone had no effect on the rate of second phosphodiester bond cleavage. In contrast, the rates of the reaction in complexes bearing substitutions only in the RNA–DNA hybrid were very similar to the rates in complexes with simultaneous substitutions in both, the non-template strand and the RNA–DNA hybrid (Figure 2C). Therefore, this result indicates that RNAP senses the sequence of the RNA–DNA hybrid, but not that of the non-template DNA strand, of the EC. Supplementary Figure 2A shows graphical representation of the RNA–DNA hybrid sequences recognized by RNAP.

Recognition of the RNA–DNA hybrid sequence determines the rate of reactions catalysed by RNAP

The above results indicate that, if, due to recognition of the RNA–DNA hybrid sequence, EC16 is stabilized in the backstepped state, ECs EC15 and EC14, on the same sequence, should be stabilized in the pre-translocated and the post-translocated states, respectively. To test that, we constructed a nucleic acids scaffold containing all changes that stabilize EC16 in the backstepped state, but do not affect the thermodynamic stability of the EC according to thermodynamics modelling (8-*CGAA*G*TAC*-15 (numbering from the 5′ end of the RNA)—stabilization sequence, ST (changes to WT sequence are in *Italic*); Supplementary Figure 2A). ECs $EC14^{ST}$, $EC15^{ST}$, $EC16^{ST}$ formed on ST scaffold were analysed for the rates of second phosphodiester bond hydrolysis, pyrophosphorolysis and NTP addition, which proceed exclusively from backstepped, pre-translocated and post-translocated states, respectively, and thus can give direct information about translocation state of the EC. As seen from Figure 3A–C, $EC14^{ST}$ (only NTP addition is fast), $EC15^{ST}$ (only pyrophosphorolysis is fast) and $EC16^{ST}$ (only second phosphodiester bond hydrolysis is fast) are indeed strongly stabilized in post-translocated, pre-translocated and backstepped states, respectively. In fact, $EC15^{ST}$ was frozen in the pre-translocated state so strongly that it could not hydrolyse the second phosphodiester bond, but instead performed hydrolysis of the first phosphodiester bond (Figure 3A), even though the second one is a preferred substrate for hydrolysis (Yuzenkova and Zenkin, 2010). Also, $EC16^{ST}$ was so strongly stabilized in the backstepped state that pyrophosphorolysis in it was outcompeted by second phosphodiester bond hydrolysis (Figure 3B) even at lower pH (pH 7), which is favourable for pyrophosphorolysis (Sosunov *et al*, 2003) but strongly slows down hydrolysis (Yuzenkova and Zenkin, 2010).

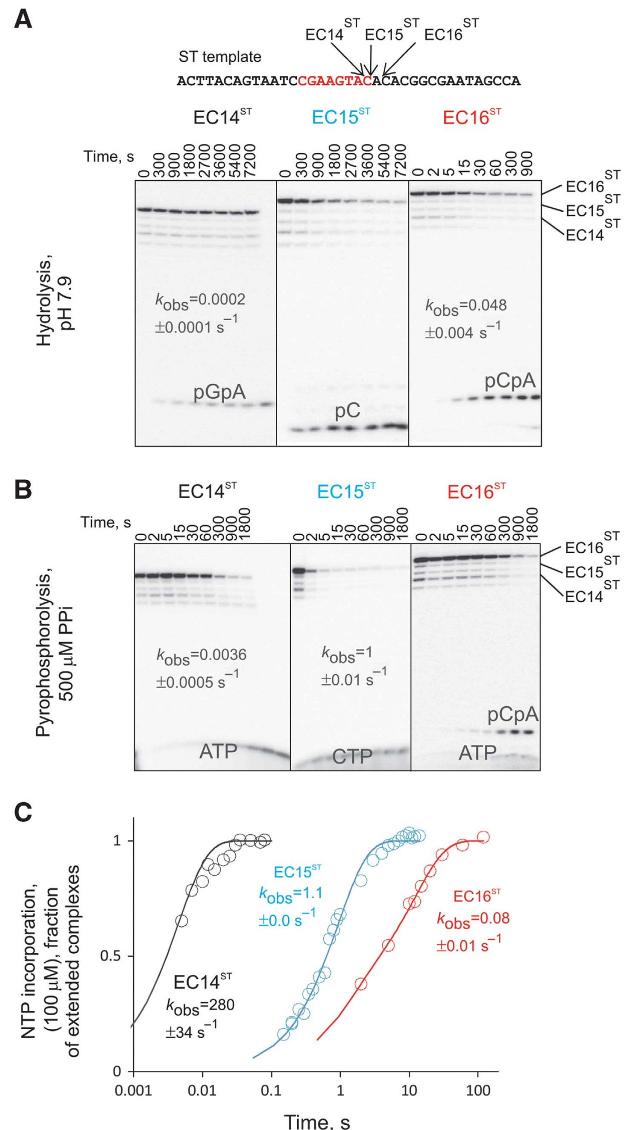

Figure 3 Recognition of the RNA–DNA hybrid influences translocation equilibrium of the elongation complex. (**A**) Kinetics of RNA hydrolysis in $EC14^{ST}$, $EC15^{ST}$ and $EC16^{ST}$, formed on the ST template shown at the top (ST sequence in red). RNA in the complexes was labelled at the 3′ end NMP to allow monitoring of the cleavage products. Note that $EC15^{ST}$ performs cleavage of the first, rather than the second, phosphodiester bond. (**B**) Kinetics of pyrophosphorolysis in $EC14^{ST}$, $EC15^{ST}$ and $EC16^{ST}$. RNA in the complexes was labelled at the 3′ end NMP, to allow monitoring of formation of an NTP. Note that in $EC16^{ST}$ hydrolysis of the second phosphodiester bond overtakes pyrophosphorolysis. (**C**) Kinetics of NTP incorporation in $EC14^{ST}$, $EC15^{ST}$ and $EC16^{ST}$, shown in logarithmic scale. The solid lines are fits of the data (circles) into a single (for $EC14^{ST}$ and $EC15^{ST}$) or a double (for $EC16^{ST}$) exponential equation. For $EC16^{ST}$, the rate of the slow fraction is shown (see text for details). Data were normalized to predicted amplitudes of the fits, which were taken as 1.

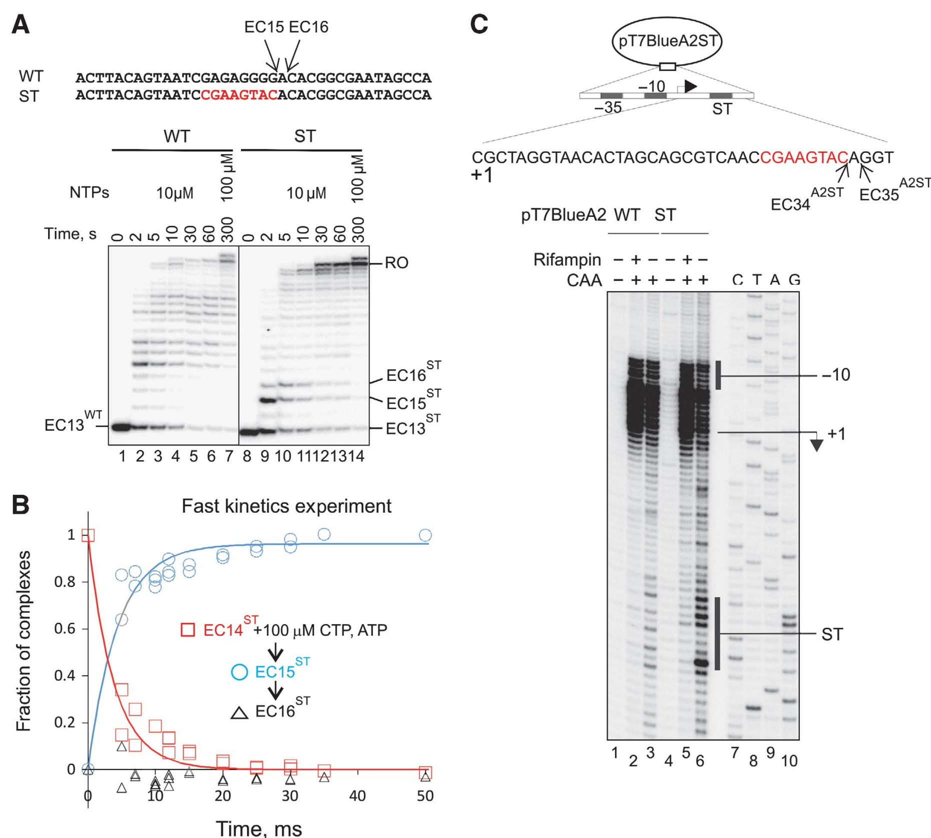

Figure 4 Recognition of the RNA–DNA hybrid sequence causes transcription pauses *in vitro* and *in vivo*. (**A**) Kinetics of elongation on WT and ST templates (shown at the top of the panel with ST sequence in red). (**B**) Recognition of the RNA–DNA hybrid sequence causes 'in pathway' pause of transcription. Kinetics of the extension of $EC14^{ST}$ (red rectangles) supplied with 100 μM ATP and CTP, which allow transcription through position 15 (blue circles) to position 16 (black triangles). Kinetics data were fitted into a single exponential equation (solid lines) and normalized to the predicted maximal extension of $EC14^{ST}$ (which was taken as 1). Note that all RNAPs that escaped from $EC14^{ST}$ pause at position 15. (**C**) Probing of pausing upon ST sequence recognition *in vivo*. The scheme of the construct used is shown above the gel, and paused complexes observed *in vitro* (Supplementary Figure 3B) are designated below the sequence. Cells carrying plasmids with A2WT or A2ST templates were treated with CAA in the presence (lanes 2 and 5) or absence (lanes 3 and 6) of rifampicin. Sequencing reaction on the A2WT template is loaded as a marker. Positions of promoter −10 region, transcription start site and the ST sequence are shown as bars and an arrow.

No such correlation of the rates of hydrolysis, pyrophosphorolysis and NTP incorporation was observed for $EC14^{WT}$, $EC15^{WT}$ and $EC16^{WT}$ formed on WT sequence (Supplementary Table 1). Fast pyrophosphorolysis in $EC15^{ST}$ was not due to the identity of the 3′ end NMP of the RNA: substitution of CMP at the 3′ end of RNA to GMP slowed down pyrophosphorolysis just 2.5 times. Note also that the structures of the nucleic acids scaffold in all complexes did not restrict shift to any of the translocation states (Figure 2A), indicating that the observed effects were due to the sensing of the RNA–DNA hybrid sequence by RNAP.

As follows from Supplementary Figure 2A, every position of the RNA–DNA hybrid influences (to various extents) the translocation equilibrium of the EC. This suggests that virtually any sequence of the RNA–DNA hybrid can determine the translocation equilibrium of the EC and, as a result, the rates of reactions catalysed by RNAP (see also below).

'In pathway' mechanism for pausing of transcription

We hypothesized that if recognition of the RNA–DNA hybrid sequence can stabilize RNAP in the pre-translocated state, it should cause a pause of transcription by restricting translocation into elongation-competent post-translocated state. To test this hypothesis, we analysed transcription through the ST sequence (Figure 4A). ECs $EC13^{WT}$ and EC13ST formed on WT and ST sequences, respectively, were allowed to transcribe to the end of the template. In agreement with the above hypothesis, the ST sequence resulted in a pause of transcription at position 15, where pre-translocated $EC15^{ST}$ is formed (Figure 4A). No pause in this register was observed on WT template (Figure 4A). In contrast to some previously described pauses (Kireeva and Kashlev, 2009), the pause at position 15 on the ST sequence was independent of the identity of dNMP in position 16 of the template (Supplementary Figure 3A). The result indicates that recognition of the RNA–DNA hybrid sequence may cause pausing of transcription by delaying translocation. The ST sequence also caused a pause at position 16, which corresponds to the backstepped $EC16^{ST}$, suggesting that recognition of the RNA–DNA hybrid sequence during transcription may also stimulate backtracking and as a result pausing of transcription (see also below). Note that introduction of the ST sequence strongly reduces pauses in positions 17–24 (Figure 4A). These effects may be caused by different factors. For example, reduction of the pause at position 17 may be caused by introduction of T:A base pair five positions upstream of the pause site, which, according to our results (Figure 2B, Supplementary Figure 2A), would destabilize $EC17^{WT}$.

©2012 European Molecular Biology Organization

In contrast, reduction of the pause at position 22 of WT template may be due to thermodynamic effect of weakening of the rear edge of the RNA–DNA hybrid (in the similar manner as weak base pairs in position −8 destabilize backstepped state in Figure 2B). It is also possible that these pauses on WT template are caused by yet unknown interaction of RNAP with nucleic acids scaffold (such as interactions of RNA exit channel with the transcript), which is altered by introduction of the ST sequence.

RNAP passes through the pre-translocated state at every NAC. Stabilization of $EC15^{ST}$ in the pre-translocated state suggests that the pause at position 15 is caused solely by slow translocation. The pause at position 15 thus may be a first experimental evidence for a factor-independent translocation pause, that is, of an 'in pathway' pause. This hypothesis suggests that there should be no kinetic partitioning of the ECs entering the pause, and all complexes should respond to the pause signal. We measured the proportion of RNAPs that pause at position 15 of the ST template ($EC15^{ST}$) before being extended further. RNAPs were allowed to transcribe from $EC14^{ST}$ past $EC15^{ST}$ in 100 μM NTPs. Reaction in nearly saturating concentration of NTPs proceeds in millisecond range and was therefore measured using a Quench Flow device. As seen from Figure 4B, all RNAPs that have escaped from $EC14^{ST}$ paused at position 15. In the timescale used, no read-through to position 16 was observed even after full extension of $EC14^{ST}$. This result indicates that the pause observed at position 15 is an obligatory event, that is, is an 'in pathway' translocation pause. Note that recognition of the RNA–DNA hybrid sequence may participate not only in the 'in pathway' pausing: for example, it stabilizes $EC16^{ST}$ (and HIV-1 pause; see below) in the backstepped state, although backstepping is an 'off pathway' event. Consistently, in contrast to $EC15^{ST}$, the kinetics of NTP incorporation in $EC16^{ST}$ contains fast and slow phases (Figure 3C), suggesting the existence of kinetic partitioning of the complexes, which is a characteristic of an 'off pathway' pause (Kireeva and Kashlev, 2009; Landick, 2009).

Recognition of the RNA–DNA hybrid sequence causes pausing of transcription in vivo

To test if recognition of the sequence of the RNA–DNA hybrid by RNAP can cause a pause of transcription *in vivo*, we introduced ST sequence into initially transcribed region of A2 promoter of T7 bacteriophage from position 28 to position 35. We confirmed that ST sequence indeed caused pre-translocated pause on this template *in vitro* (Supplementary Figure 3B), indicating that recognition of the RNA–DNA hybrid sequence and pausing caused by it do not depend on sequences upstream or downstream of the RNA–DNA hybrid. Next, *E. coli* cells carrying plasmids containing either WT T7A2 template (pT7BlueA2WT) or T7A2 template with ST sequence (pT7BlueA2ST) were treated with chloroacetaldehyde (CAA). CAA, upon entering the cell, modifies single-stranded regions of DNA, present in transcription initiation and ECs. CAA-modified positions are then revealed by primer extension on purified plasmids. As seen from Figure 4C, open promoter complex formed on both pT7BlueA2WT and pT7BlueA2ST (lanes 3 and 6). However, only pT7BlueA2ST contained a region of stably open DNA, which corresponds to the position of the ST sequence (Figure 4C, lanes 3 and 6). This stably open region was eliminated by the addition of rifampicin, which prevents escape from the promoter into elongation (Figure 4C, lanes 2 and 5). This result indicates that the stable opening of DNA in the region of ST sequence was due to a paused EC.

As mentioned above, all positions of the RNA–DNA hybrid may influence the translocation equilibrium of EC (Supplementary Figure 2A) to a different degree. Accordingly, combinations of some individual positions of the ST sequence were also causing pausing of transcription, although weaker than that caused by the full ST sequence (Supplementary Figure 3C). Therefore, it is possible that there exists a large number of configurations of the RNA–DNA hybrid sequence that may disfavour translocation. This suggests that recognition of the RNA–DNA hybrid sequence may influence the overall rate of transcription elongation.

RNA–DNA hybrid sequence recognition participates in regulation of transcription

Some transcriptional pauses are known to be species-specific (Artsimovitch *et al*, 2000; Kireeva and Kashlev, 2009), that is, to be recognized by multi-subunit RNAPs from some organisms, but not from the others. To test if recognition of the RNA–DNA hybrid sequence is conserved among multi-subunit RNAPs, we compared pausing on ST sequence between *E. coli* RNAP, RNAP from distantly related *Streptococcus pneumoniae*, and three eukaryotic RNAPs: pol I, pol II and pol III. Intriguingly, we observed similar pausing at positions 15 and 16 by all RNAPs tested (Figure 5A). The apparent differences in pause efficiencies can be explained by the various elongation properties of these RNAPs (such as hydrolysis efficiency, elongation rate and so on) as well as by variations in amino acids contacting the RNA–DNA hybrid and potentially participating in the hybrid sequence recognition (Supplementary Table 2). Note also the existence of some species-specific pauses further downstream the template.

We investigated two physiologically important hairpin-independent pauses of transcription. Bacterial *ops* pause is required for binding of a transcription factor RfaH to the paused RNAP, making RNAP able to processively transcribe downstream sequences of some bacterial operons (Bailey *et al*, 1997; Artsimovitch and Landick, 2000; Santangelo and Roberts, 2002). A pause of human immunodeficiency virus, HIV-1, is required for a proper co-transcriptional folding of the nascent RNA, which is needed for binding of protein Tat, the regulator of viral transcription and replication (Palangat *et al*, 1998). The ECs paused on *ops* and HIV-1 signals were shown to be stabilized in the pre-translocated and backstepped states, respectively (Artsimovitch and Landick, 2000; Palangat and Landick, 2001). We hypothesized that these pauses are caused by RNA–DNA hybrid recognition. To test this hypothesis, we introduced sequences of the *ops* signal (11 positions upstream and 12 positions downstream of the pause site) and of the HIV-1 signal (12 positions upstream and 19 positions downstream of the pause site) into the transcribed parts of the nucleic acids scaffold (*ops* and HIV-1 scaffolds, respectively; Figure 5B and C, respectively). As seen from Figure 5B and C, respectively, *E. coli* RNAP paused in position 22 on the *ops* template ($EC22^{ops}$), while Pol II formed a paused EC at position 22 of HIV-1 template ($EC22^{HIV-1}$), both corresponding to the previously determined pause sites (Palangat *et al*, 1998; Artsimovitch and Landick, 2000). First, we noted, that, in

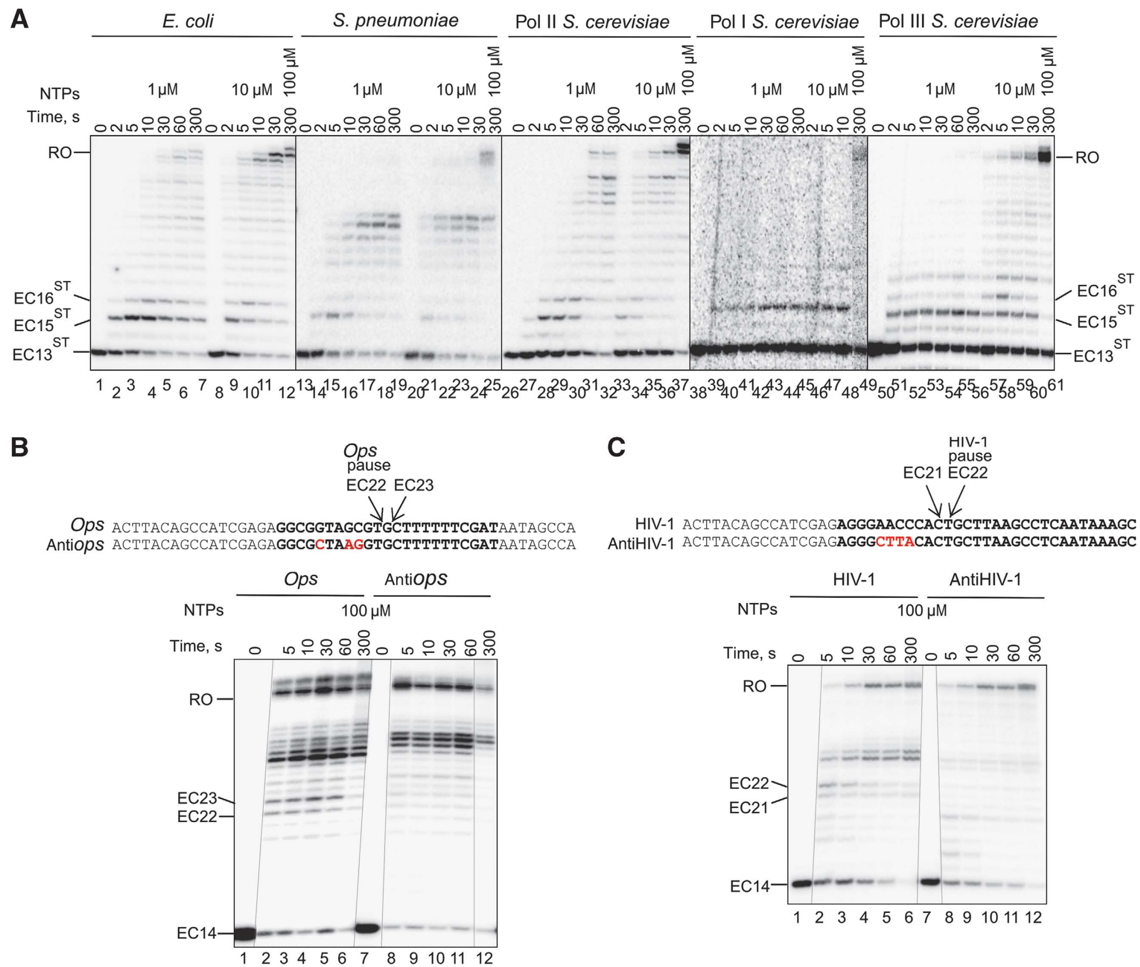

Figure 5 Recognition of RNA–DNA hybrid sequence participates in regulation of transcription. (**A**) Kinetics of transcription on the ST template by RNAPs from *E. coli, S. pneumonia* and pol I, pol II and pol III from *S. cerevisiae*. (**B**, **C**) Kinetics of transcription on the *ops,* anti*ops* (panel **B**) HIV-1 and antiHIV-1 (panel **C**) sequences, depicted above the gels. *RfaQ-ops* and HIV-1 sequences inserted into WT sequence (Figure 2A) are shown in bold, and the changes that according to our results should prevent pausing are in red. The sites of *ops* and HIV-1 pauses, consistent with the previously reported (Palangat *et al*, 1998; Artsimovitch and Landick, 2000), are shown with arrows. Black lines separate parts of one gel that were brought together.

agreement with the above hypothesis, pre-translocated $EC22^{ops}$ was followed by an additional pause, while backstepped $EC22^{HIV-1}$ was preceded by an additional pause; both patterns resembling pre-translocated $EC15^{ST}$ and backstepped $EC16^{ST}$ formed on the ST template (Figure 4A).

Next, we introduced substitutions in the region of the RNA–DNA hybrids of paused complexes that, according to our results, should not favour stabilization of $EC22^{ops}$ in the pre-translocated state (anti*ops* template, Figure 5B) and $EC22^{HIV-1}$ in the backstepped state (antiHIV-1 template, Figure 5C). Three substitutions in the *ops* template and four substitutions in the HIV-1 template significantly decreased efficiency of corresponding pauses (Figure 5B and C). These substitutions do not influence translocation equilibrium through changes in the thermodynamic stability of the EC (Figure 2B, top). Note that the RNA–DNA hybrid on the antiHIV-1 template is not stronger than that on the HIV-1 template, indicating that the pause is not determined by a weak hybrid as was proposed earlier (Palangat and Landick, 2001). Introduction of the above substitutions in the non-template strands of scaffolds did not affect pauses significantly (Supplementary Figure 4A and B), suggesting that *ops* and HIV-1 pauses are caused by sensing of the RNA–DNA hybrid sequence by RNAP. Also, in agreement with the above hypothesis, a pause following the *ops* pause and a pause preceding the HIV-1 pause were also significantly decreased by changes in the RNA–DNA hybrid. It should be noted that, earlier, *E. coli ops* pause was shown to be efficiently recognized by distantly related *Bacillus subtilis* RNAP (Artsimovitch *et al*, 2000), suggesting that the mechanism of pausing at *ops* site is conserved, which is consistent with the conservation of the pausing caused by the RNA–DNA hybrid sequence recognition. Note that some pauses downstream of HIV-1 and *ops* signals were altered/reduced by the introduced changes. These effects can be explained by alteration of the hybrid recognition, thermodynamics effects or yet unknown interaction of RNAP with nucleic acids (see also above).

Discussion

In this work, we showed that RNAP core can sense the identity of base pairs at most of the positions of the RNA–DNA hybrid of the EC. This recognition restricts the oscillation of the EC between translocation states, and if translocation equilibrium is shifted towards the pre-translocated state the recognized sequence may cause a pause of

©2012 European Molecular Biology Organization

transcription. A pause upon the RNA–DNA hybrid recognition happens if the rate of translocation slows down the NAC, that is, is limiting relative to the rates of NTP binding and phosphodiester bond formation. On most sequences, translocation is presumably faster than the chemistry step of the NAC. This is supported by our finding that $EC14^{ST}$, stabilized in the post-translocated state, incorporated NTP only slightly faster than complexes $EC14^{WT}$, 15^{WT}, 16^{WT} formed on WT sequence, which do not have preference to any of the translocation states (compare Figure 3C and Supplementary Table 1). However, even minor changes to the sequence of the RNA–DNA hybrid may influence translocation equilibrium (Supplementary Figure 2A) and cause a relatively strong pausing of elongation (Supplementary Figure 3C). Note also that short delays in translocation that are not limiting for the NAC may increase probability of 'off pathway' events, such as backtracking. Furthermore, it is possible that recognition of the RNA–DNA hybrid sequence may also stabilize 'off pathway' pauses caused by different signals. Our results are supported by and are consistent with the earlier proposition that the entire RNA–DNA hybrid may be a signal for pausing and backtracking by pol II, and that pol II may somehow 'feel' it rather than respond to thermodynamics of nucleic acids (Hawryluk *et al*, 2004).

Promoter escape is known to be influenced by the sequence of the initially transcribed region downstream of the transcription start site (Hsu *et al*, 2006), although the mechanism of that is not understood. Results presented here open a possibility that the abortive transcription, the efficiency of promoter clearance and promoter-proximal pausing may in part be determined by the sensing of the sequence of the RNA–DNA hybrid formed during synthesis through the initially transcribed region. Most recently, promoter proximal σ-dependent pausing by bacterial RNAP was shown to, in part, depend on the sequence downstream of the elements recognized by σ-factor: the sequence corresponding to the RNA–DNA hybrid of the paused complex (Perdue and Roberts, 2010). This sequence causes backtracking by RNAP thus stabilizing the pause. Our results indicate that this phenomenon may involve recognition of the RNA–DNA hybrid sequence by RNAP.

RNAP core was proposed to 'sense' the sequences of downstream (Ederth *et al*, 2002) and upstream DNA (Yuzenkova *et al*, 2011) during elongation. Taken together with our results, this suggests that recognition of the sequences of the nucleic acids of the EC by RNAP, rather than thermodynamics of the nucleic acids, may determine the overall rate of transcription elongation. This may also explain the response of RNAP to transcription factors, such as NusA, NusG, RfaH, which may allosterically change some determinants of sequence recognition by RNAP core and thus influence response to pauses and/or termination signals.

At present, we cannot establish whether RNAP recognizes chemical groups of the RNA–DNA hybrid or its shape. Most of the amino acids of the main channel of RNAP interact with the phosphate backbone of the RNA–DNA hybrid, and thus, likely, may sense deviations in the geometry of base pairs, rather than chemical groups of the bases. Some amino acids may possibly interact with the bases of the hybrid. We tested several mutant *E. coli* RNAPs bearing alanine substitutions of conserved amino acids of the main channel (βN568, βD1240, β′P427, β′Q465 and β′M932) for their sensitivity to changes in the RNA–DNA hybrid sequence (Supplementary Figure 2B). βD1240, β′Q465 and β′M932 seem not to sense base pairs in positions that are close to them in the EC (Vassylyev *et al*, 2007b). We however found that βN568 and β′P427 may be involved in sensing of base pairs in positions −3 and −2 (βN568) and −1 (β′P427) of EC16 (Supplementary Figure 2B).

Materials and methods

Proteins and templates

RNAPs from *E. coli*, and *S. pneumoniae* were purified as described (Sidorenkov *et al*, 1998). *S. cerevisiae* pol I, pol II and pol III were kindly provided by Soren Nielsen (Newcastle University). Sequences of oligonucleotides used for the ECs assembly and other templates are shown in Supplementary Methods. T7A2-based templates for promoter-specific transcription were obtained by PCR. For *in vivo* experiments, T7A2-based templates were cloned in pT7BlueT7A2 (Naryshkina *et al*, 2006).

Transcription assays

ECs were assembled as described (Yuzenkova *et al*, 2010; Roghanian *et al*, 2011) in transcription buffer TB (20 mM Tris–HCl, pH 7.9, 40 mM KCl). For pyrophosphorolysis TB with pH 7 was used. Complexes of *E. coli* RNAP were immobilized on Ni^{2+}-NTA agarose beads (Qiagen), while complexes of *S. pneumonia* RNAP and pol I, II and III on streptavidin-agarose beads (Fluka) through biotin on the 5′ end of the non-template strand. All complexes were assembled with 13 nucleotide long RNA (EC13), which was either ^{32}P labelled at the 5′ end before complex assembly or labelled at the 3′ end after EC assembly by incorporation of the 3′ end [α-^{32}P]NMP, with subsequent removal of the unincorporated nucleotide through washing of the beads with TB. Reactions were started by addition of 10 mM (final) $MgCl_2$ with or without NTPs or PPi at concentrations specified in figures, incubated for various times indicated in the figures at 37°C (except for pol I, II and III, which were incubated at 30°C) and stopped by addition of a loading buffer containing formamide. Transcription on T7A2 templates was initiated with 100 μM dinucleotide CpG and 40 μM CTP, ATP and GTP, one of which was [α-^{32}P] radiolabelled (7.5 Ci/mmol). After 5 min, resultant complexes were immobilized on Ni^{2+}-NTA agarose beads, washed to remove nucleotides as described (Zenkin *et al*, 2007) and chased in the presence of all NTPs. Paused complexes for Supplementary Figure 3B were isolated by washing chase reaction after 10 s with TB. Pyrophosphorolysis and NTP addition in paused complexes were performed as above. Products of all reactions were resolved by denaturing (8 M Urea) 10, 20 or 23% PAGE, revealed by PhosphorImaging (GE Healthcare), and analysed using ImageQuant software (GE Healthcare). A proportion of complexes that undergone reaction was plotted against time and fitted to a single exponential equation by using nonlinear regression (Yuzenkova *et al*, 2010). Fast kinetic experiments were performed and analysed essentially as described (Yuzenkova *et al*, 2010), except for assembled EC13 were walked by one position by addition of 1-μM NTP specified by the template, before mixing them with 10 mM $MgCl_2$ (final concentration) and NTPs (at concentrations specified in figures) in a Quench Flow regime.

In vivo footprinting

DH5α cells transformed with plasmids carrying derivatives of T7A2 template were grown in 10 ml of M9 media with casamino acids to OD_{600} 0.6. A volume of 600 μl CAA was added for 20 min at 37°C. Cells were spin down and plasmids were immediately purified. Analysis of the DNA modifications was performed by primer extension and products were analysed by 6% PAGE, as above.

Thermodynamic modelling of transcription pausing

The probability calculations are based on a model of the EC as described in Tadigotla *et al* (2006). Configuration of transcription EC for calculations was as follows: 10 base pairs RNA–DNA hybrid, 2 unpaired bases of DNA upstream and 0 unpaired bases of DNA downstream of the hybrid (Figure 1). We also restricted the number of translocation states to pre-translocated, post-translocated and backstepped (backtracking by 1 bp) based on our experimental results. The probabilities of the backstepped state given by the

©2012 European Molecular Biology Organization

Boltzmann distribution were normalized to the probability of backstepping of EC16^{WT}, which was taken as 1.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Kenn Gerdes for fruitful discussion, Jan-Willem Veening for *S. pneumoniea* strain and Soren Nielsen for proteins. This work was supported by UK Biotechnology and Biological Sciences Research Council, BBSRC under SysMO initiative and the European Research Council [ERC-2007-StG 202994-MTP] to NZ.

Author contributions: AB and YY performed experiments and analysed data; VT performed mathematical modelling; NZ designed experiments, analysed data and wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alexander RD, Innocente SA, Barrass JD, Beggs JD (2010) Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* **40:** 582–593

Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* **97:** 7090–7095

Artsimovitch I, Svetlov V, Anthony L, Burgess RR, Landick R (2000) RNA polymerases from *Bacillus subtilis* and *Escherichia coli* differ in recognition of regulatory signals *in vitro*. *J Bacteriol* **182:** 6027–6035

Bai L, Shundrovsky A, Wang MD (2004) Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J Mol Biol* **344:** 335–349

Bailey MJA, Hughes C, Koronakis V (1997) RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. *Mol Microbiol* **26:** 845–851

Borukhov S, Lee J, Laptenko O (2005) Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol* **55:** 1315–1324

Brodolin K, Zenkin N, Mustaev A, Mamaeva D, Heumann H (2004) The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. *Nat Struct Mol Biol* **11:** 551–557

Chan CL, Landick R (1989) The Salmonella typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins. *J Biol Chem* **264:** 20796–20804

Ciampi MS (2006) Rho-dependent terminators and transcription termination. *Microbiology* **152:** 2515–2528

Cramer P (2006) Molecular biology. Self-correcting messages. *Science* **313:** 447–448

Donahue JP, Turnbough Jr CL (1994) Nucleotide-specific transcriptional pausing in the pyrBI leader region of *Escherichia coli* K-12. *J Biol Chem* **269:** 18185–18191

Ederth J, Artsimovitch I, Isaksson LA, Landick R (2002) The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. *J Biol Chem* **277:** 37456–37463

Farnham PJ, Platt T (1981) Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription *in vitro*. *Nucleic Acids Res* **9:** 563–577

Gong F, Yanofsky C (2003) A transcriptional pause synchronizes translation with transcription in the tryptophanase operon leader region. *J Bacteriol* **185:** 6472–6476

Hatoum A, Roberts J (2008) Prevalence of RNA polymerase stalling at *Escherichia coli* promoters after open complex formation. *Mol Microbiol* **68:** 17–28

Hawryluk PJ, Ujvari A, Luse DS (2004) Characterization of a novel RNA polymerase II arrest site which lacks a weak 3′ RNA-DNA hybrid. *Nucleic Acids Res* **32:** 1904–1916

Herbert KM, Greenleaf WJ, Block SM (2008) Single-molecule studies of RNA polymerase: motoring along. *Annu Rev Biochem* **77:** 149–176

Hsu LM, Cobb IM, Ozmore JR, Khoo M, Nahm G, Xia L, Bao Y, Ahn C (2006) Initial transcribed sequence mutations specifically affect promoter escape properties. *Biochemistry* **45:** 8841–8854

Kashkina E, Anikin M, Brueckner F, Lehmann E, Kochetkov SN, McAllister WT, Cramer P, Temiakov D (2007) Multisubunit RNA polymerases melt only a single DNA base pair downstream of the active site. *J Biol Chem* **282:** 21578–21582

Kashkina E, Anikin M, Tahirov TH, Kochetkov SN, Vassylyev DG, Temiakov D (2006) Elongation complexes of Thermus thermophilus RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Res* **34:** 4036–4045

Kent T, Kashkina E, Anikin M, Temiakov D (2009) Maintenance of RNA-DNA hybrid length in bacterial RNA polymerases. *J Biol Chem* **284:** 13497–13504

Kireeva ML, Kashlev M (2009) Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc Natl Acad Sci USA* **106:** 8900–8905

Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34:** 1062–1066

Landick R (2009) Transcriptional pausing without backtracking. *Proc Natl Acad Sci USA* **106:** 8797–8798

Landick R, Carey J, Yanofsky C (1987) Detection of transcription-pausing *in vivo* in the trp operon leader region. *Proc Natl Acad Sci USA* **84:** 1507–1511

Maoileidigh DO, Tadigotla VR, Nudler E, Ruckenstein AE (2011) A unified model of transcription elongation: what have we learned from single-molecule experiments? *Biophys J* **100:** 1157–1166

Naryshkina T, Kuznedelov K, Severinov K (2006) The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA-DNA hybrid. *J Mol Biol* **361:** 634–643

Nickels BE, Mukhopadhyay J, Garrity SJ, Ebright RH, Hochschild A (2004) The sigma 70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. *Nat Struct Mol Biol* **11:** 544–550

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

Palangat M, Meier TI, Keene RG, Landick R (1998) Transcriptional pausing at +62 of the HIV-1 nascent RNA modulates formation of the TAR RNA structure. *Mol Cell* **1:** 1033–1042

Perdue SA, Roberts JW (2010) A backtrack-inducing sequence is an essential component of *Escherichia coli* sigma(70)-dependent promoter-proximal pausing. *Mol Microbiol* **78:** 636–650

Perdue SA, Roberts JW (2011) sigma(70)-dependent transcription pausing in *Escherichia coli*. *J Mol Biol* **412:** 782–792

Ring BZ, Yarnell WS, Roberts JW (1996) Function of *E. coli* RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. *Cell* **86:** 485–493

Roghanian M, Yuzenkova Y, Zenkin N (2011) Controlled interplay between trigger loop and Gre factor in the RNA polymerase active centre. *Nucleic Acids Res* **39:** 4352–4359

Santangelo TJ, Roberts JW (2002) RfaH, a bacterial transcription antiterminator. *Mol cell* **9:** 698–700

Sidorenkov I, Komissarova N, Kashlev M (1998) Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol Cell* **2:** 55–64

Sosunov V, Sosunova E, Mustaev A, Bass I, Nikiforov V, Goldfarb A (2003) Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J* **22:** 2234–2244

Tadigotla VR, O Maoiléidigh D, Sengupta AM, Epshtein V, Ebright RH, Nudler E, Ruckenstein AE (2006) Thermodynamic and kinetic modeling of transcriptional pausing. *Proc Natl Acad Sci USA* **103:** 4439–4444

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007a) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448:** 157–162

©2012 European Molecular Biology Organization

Vassylyev DG, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R (2007b) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448:** 163–168

Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, Kornberg RD (2009) Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324:** 1203–1206

Wong TN, Sosnick TR, Pan T (2007) Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci USA* **104:** 17995–18000

Yanofsky C (1981) Attenuation in the control of expression of bacterial operons. *Nature* **289:** 751–758

Yuzenkova Y, Bochkareva A, Tadigotla VR, Roghanian M, Zorov S, Severinov K, Zenkin N (2010) Stepwise mechanism for transcription fidelity. *BMC Biol* **8:** 54

Yuzenkova Y, Tadigotla VR, Severinov K, Zenkin N (2011) A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J* **30:** 3766–3775

Yuzenkova Y, Zenkin N (2010) Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proc Natl Acad Sci USA* **107:** 10878–10883

Zenkin N, Kulbachinskiy A, Yuzenkova Y, Mustaev A, Bass I, Severinov K, Brodolin K (2007) Region 1.2 of the RNA polymerase sigma subunit controls recognition of the −10 promoter element. *EMBO J* **26:** 955–964

Zenkin N, Severinov K (2008) RNA polymerase—the third class of primases. *Cell Mol Life Sci* **65:** 2280–2288

Zenkin N, Yuzenkova Y, Severinov K (2006) Transcript-assisted transcriptional proofreading. *Science* **313:** 518–520

***The EMBO Journal* is published by *Nature Publishing Group* on behalf of *European Molecular Biology Organization*. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License. [http://creativecommons.org/licenses/by-nc-sa/3.0/]**