The Transcription Bubble of the RNA Polymerase–Promoter Open Complex Exhibits Conformational Heterogeneity and Millisecond-Scale Dynamics: Implications for Transcription Start-Site Selection

Nicole C. Robb[1]†, Thorben Cordes[1]†, Ling Chin Hwang[1]†, Kristofer Gryte[1], Diego Duchi[1], Timothy D. Craggs[1], Yusdi Santoso[1], Shimon Weiss[2], Richard H. Ebright[3] and Achillefs N. Kapanidis[1]

1 - Biological Physics Research Group, *Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK*

2 - Department of Chemistry and Biochemistry, *University of California Los Angeles, Los Angeles, CA 90095, USA*

3 - Department of Chemistry, *Howard Hughes Medical Institute and Waksman Institute, Rutgers University, Piscataway, NJ 08854, USA*

Correspondence to Achillefs N. Kapanidis: *a.kapanidis1@physics.ox.ac.uk*
http://dx.doi.org/10.1016/j.jmb.2012.12.015
Edited by D. E. Draper

Abstract

Bacterial transcription is initiated after RNA polymerase (RNAP) binds to promoter DNA, melts ~ 14 bp around the transcription start site and forms a single-stranded "transcription bubble" within a catalytically active RNAP–DNA open complex (RP$_o$). There is significant flexibility in the transcription start site, which causes variable spacing between the promoter elements and the start site; this in turn causes differences in the length and sequence at the 5′ end of RNA transcripts and can be important for gene regulation. The start-site variability also implies the presence of some flexibility in the positioning of the DNA relative to the RNAP active site in RP$_o$. The flexibility may occur in the positioning of the transcription bubble prior to RNA synthesis and may reflect bubble expansion ("scrunching") or bubble contraction ("unscrunching"). Here, we assess the presence of dynamic flexibility in RP$_o$ with single-molecule FRET (*F*örster *r*esonance *e*nergy *t*ransfer). We obtain experimental evidence for dynamic flexibility in RP$_o$ using different FRET rulers and labeling positions. An analysis of FRET distributions of RP$_o$ using burst variance analysis reveals conformational fluctuations in RP$_o$ in the millisecond timescale. Further experiments using subsets of nucleotides and DNA mutations allowed us to reprogram the transcription start sites, in a way that can be described by repositioning of the single-stranded transcription bubble relative to the RNAP active site within RP$_o$. Our study marks the first experimental observation of conformational dynamics in the transcription bubble of RP$_o$ and indicates that DNA dynamics within the bubble affect the search for transcription start sites.

Introduction

Transcription, the synthesis of RNA from a DNA template, is the first step in gene expression and is a highly regulated process. In *Escherichia coli* and other bacteria, RNA polymerase (RNAP) initiates transcription after binding to specific sequences within promoter DNA, where binding is controlled by transcription initiation factors known as sigma (σ) factors. In typical bacterial promoters controlled by the main sigma factor σ^{70}, the RNAP-σ^{70} holoenzyme initially binds to the − 10 and − 35 elements of the promoter (reviewed in Ref. [1]), melts ~ 14 bp in the DNA surrounding the transcription start site to form a single-stranded "transcription bubble" and yields the catalytically active RNAP–DNA open complex (RP$_o$).

RNAP can initiate transcription from multiple positions within the same promoter, in both prokaryotes and eukaryotes.[2–15] In bacteria, transcription is mainly initiated by purines located within a region 4–12 bp downstream of the −10 element, which extends from position −7 to −12 relative to the +1 site. Such preferences for start sites can be used for regulation of gene expression, since variation at the 5′ ends of transcripts can affect transcript stability[16] or control the formation of secondary structures that in turn can affect translational initiation.[17,18] For example, expression of the *pyrC* gene in *E. coli* is regulated by a translational control mechanism dependent on the presence or absence of a 5′ hairpin loop structure whose synthesis depends on start-site selection.[3] In addition, different start sites can affect the extent of abortive initiation or transcriptional slippage, both of which can influence the frequency of initiation at a particular promoter.[4,19,20]

Flexibility in transcription start sites has also been observed in eukaryotes. An early study identified heterogeneity in the 5′ termini of adenovirus mRNAs, which are transcribed by cellular RNAP II.[21] In yeast, RNAP II initiates transcription at multiple start sites located 40–120 bp downstream of the TATA box,[22] presumably by actively scanning (through an ATP-driven process) for start sites downstream of the

transcription bubble, as in the case of the U4 small nuclear RNA gene SNR14.[23]

The observed flexibility in transcription start sites implies that there must be static or dynamic flexibility in the positioning of the single-stranded transcription bubble relative to the RNAP active center in RP$_o$ (Fig. 1a). The discoveries that initial transcription by RNAP involves transcription-bubble expansion ("scrunching")[24,25] and that promoter escape by RNAP involves transcription-bubble contraction ("unscrunching")[25] provide precedents for the functional importance of transcription-bubble flexibility. The discovery of scrunching and unscrunching also supported a mechanistic model[4] for the role of transcription-bubble flexibility in start-site selection: namely, transcription-bubble expansion (similar to scrunching) in RP$_o$ would place further downstream DNA in contact with the RNAP active center, resulting in a more downstream start site, and transcription-bubble contraction (similar to unscrunching) in RP$_o$ would place more upstream DNA in contact with the RNAP active center, resulting in a more upstream start site.

In this work, we studied the mechanisms for start-site heterogeneity by examining the *E. coli lac* promoter. Specifically, we have performed single-molecule FRET (*Förster resonance energy transfer*) (smFRET) measurements on individual, freely
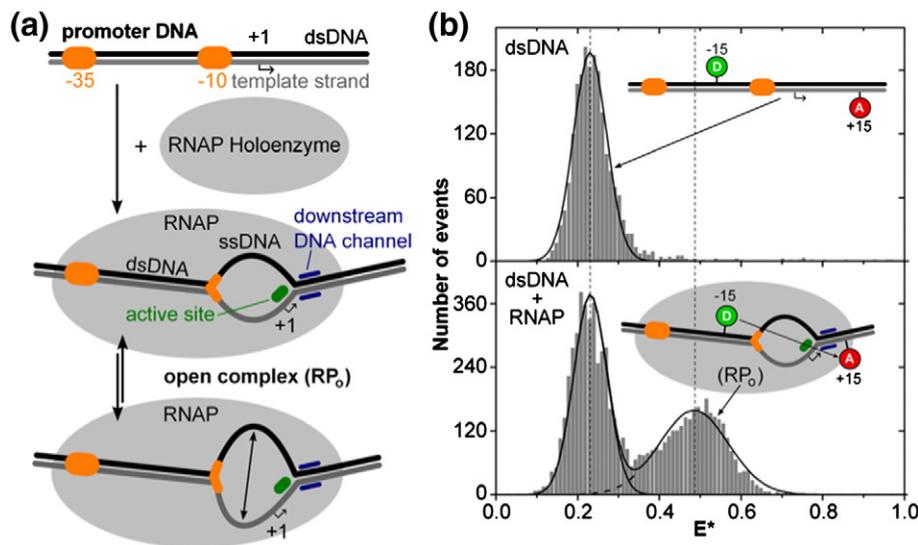
Fig. 1. Using smFRET to investigate dynamics of RNAP open complexes. (a) Schematic of the hypothesis that dynamics of single-stranded DNA in the transcription bubble of the open complex allow RNAP to sample different transcriptional start sites. The positions of the RNAP active site (green oval) and the −10 element (orange downstream region) of the DNA are assumed to be fixed with respect to each other.[36] Sampling of transcription start sites can therefore proceed via movement of single-stranded DNA within the transcription bubble, as indicated by the two representations of the open complex (RP$_o$). (b) Detecting RP$_o$ formation with smFRET. dsDNA was labeled with donor and acceptor fluorophores on either side of the transcription bubble [donor at position −15 with respect to the +1 position and acceptor at position +15; see Supplementary Fig. 1 for the DNA sequence of *lacCONS+2*(A+2C) used in (b)]. smFRET spectroscopy combined with ALEX on diffusing molecules of dsDNA alone and dsDNA with RNAP (RP$_o$) was carried out. Ratio *E** represents the uncorrected FRET efficiency, and curves were fitted with Gaussian functions to determine the center and width of the distributions.

diffusing molecules of RP_o in solution[24,26–28] in order to detect transcription-bubble flexibility, to distinguish between static and dynamic transcription-bubble flexibility and to relate transcription-bubble flexibility and dynamics to start-site selection. By measuring the FRET efficiency between fluorophore pairs probing different regions within promoter DNA, we were able to detect the formation of the transcription bubble (Fig. 1b), as well as DNA movements associated with start-site selection. Our results establish that DNA within RP_o exhibits conformational dynamics on the millisecond timescale. We further establish that the addition of different initiating nucleotides or the introduction of base-pair substitutions can reprogram start-site selection by repositioning the transcription-bubble DNA relative to the RNAP active site in RP_o. Our results are consistent with a mechanistic model[4] in which flexibility in start-site selection results from transcription-bubble expansion (scrunching)[24,25] and transcription-bubble contraction (unscrunching)[25] in RP_o.

Results

Start-site selection at *lacCONS+2* and *lacCONS+2* derivatives

The *lacCONS+2* promoter used in this work is a derivative of the *E. coli lacUV5* promoter; *lacCONS+2* differs from the *lacUV5* promoter by having a single-base-pair substitution in the −35 element (to form a consensus −35 element), a single-base-pair deletion in the spacer region between the −35 and −10 elements (to form a consensus −10/−35 spacer region) and a 2-bp insertion at position +9 in the initial transcribed region.[29] We examined the distribution of transcription start sites at *lacCONS+2* and at three *lacCONS+2* derivatives containing substitutions in the start-site region (Fig. 2). As anticipated from previous work on the start sites at the *lac* promoter and substituted *lac* promoter derivatives,[2] *lacCONS+2* exhibited a major start site at position +1, and substituted *lacCONS+2* derivatives exhibited different distributions of start sites (Fig. 2). Specifically, *lacCONS+2*(T−3A) exhibited start sites at −2, +1, +2 and +3, while *lacCONS+2*(A+2C) exhibited a start site at +1 and *lacCONS+2*(G−2T;A+1C) exhibited start sites at −1 and at +2.

Transcription-bubble flexibility in RP_o: smFRET between DNA segments upstream and downstream of the bubble

Having confirmed that RNAP can initiate transcription from multiple sites on the *lacCONS+2* derivatives, we used smFRET to investigate the dynamics of RP_o. Experiments were conducted using *lacCONS+2*(T−3A) (Fig. 2b) labeled with the donor fluorophore Cy3B and the acceptor fluorophore ATTO647N in a variety of positions (for DNA sequences and labeling positions, see Supplementary Fig. 1). In a first set of experiments, the FRET ruler ("downstream DNA ruler") monitored the distance between DNA segments upstream and downstream of the transcription bubble, by monitoring FRET between an acceptor fluorophore at position +15 and a donor fluorophore at position −15 (Fig. 3a, top panel; see Ref. 24). This FRET ruler detects RP_o formation through an increase in FRET due to opening of the transcription bubble (Fig. 3a). smFRET spectroscopy with ALEX (*alter-nating-laser excitation*)[30,31] on diffusing molecules of double-stranded DNA (dsDNA) and RP_o resulted in FRET histograms for molecules containing both donor and acceptor fluorophores (Fig. 3a; see Materials and Methods for details).

The FRET distribution of free dsDNA showed a single species with low apparent FRET (mean value of $E^* \sim 0.23$; Fig. 3a). As expected,[24] addition of RNAP to DNA to form RP_o resulted in a bimodal distribution that represents two species: dsDNA (due to dissociation of nonspecific RNAP–DNA complexes during heparin challenge; see Materials and Methods) and RP_o (Fig. 3a, third panel). The DNA distribution was centered at $E^* \sim 0.23$, whereas the RP_o distribution was centered at $E^* \sim 0.39$. The FRET distribution of RP_o was also unusually wide, both relative to dsDNA (which serves as a "static" standard; Fig. 3a, third panel) and to the expected width for a static species with a known mean FRET efficiency and photon count distribution ("shot-noise-limited" width); the latter is due to the low photon counts inherent to single-molecule fluorescence measurements.[32] The calculated shot-noise-limited width for the dsDNA peak ($\sigma = 0.038$; Fig. 3a, red Gaussian, first panel) was comparable with the actual width of the FRET distribution ($\sigma = 0.041$); in contrast, the width for the RP_o peak ($\sigma = 0.085$; Fig. 3a, black Gaussian fit, third panel) substantially exceeded its shot-noise width ($\sigma = 0.044$; Fig. 3a, red Gaussian, third panel), pointing to the presence of heterogeneity in the FRET distribution.

A possible explanation for the wide distribution of the RP_o FRET peak is the presence of conformational heterogeneity in RP_o,[26,33,34] which may exist in multiple static conformational states that do not interconvert within the millisecond-timescale transit of single RP_o complexes through the detection volume. Alternatively, RP_o may be dynamic and interconvert between different conformational states within the millisecond timescale.[26] To deconvolve static from dynamic heterogeneity, we analyzed the FRET data using burst variance analysis (BVA),[27] which detects dynamics by examining how FRET efficiency fluctuates over time within single transits of individual molecules. Essentially, molecules with
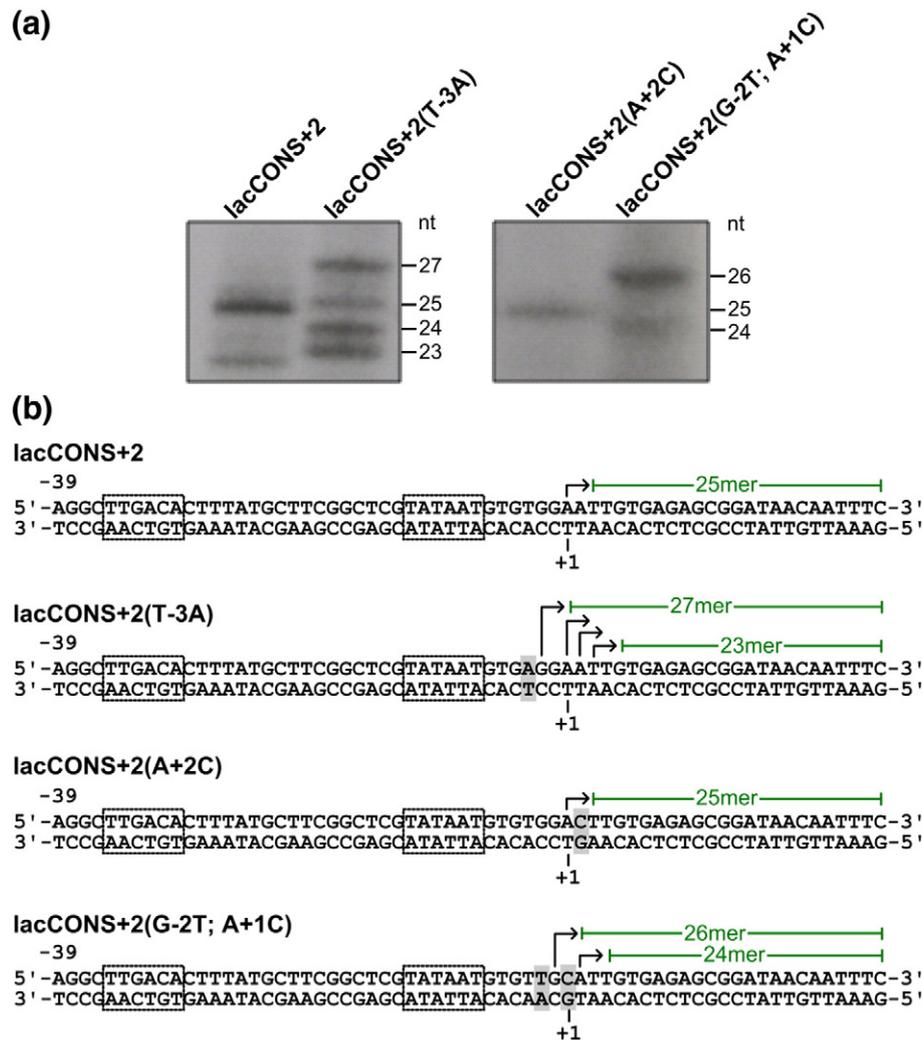
Fig. 2. Start-site selection at *lacCONS+2* and *lacCONS+2* derivatives. (a) *In vitro* transcription reactions using *lacCONS+2* promoters with base-pair substitutions; RNAP, dsDNA, ATP, UTP, CTP and [α^{32}P]GTP were incubated in transcription buffer at 37 °C for 5 min followed by heparin challenge and separation of the products on a polyacrylamide gel. Labeled RNA standards were used to determine the size of the products. (b) Sequences of the *lacCONS+2* promoter and its derivatives. Boxes are drawn around the −10 and −35 elements and the +1 position is marked. The primary RNA products observed for each sequence are labeled in green.

dynamic fluctuations in FRET are characterized by an increased FRET standard deviation compared to that expected from shot noise; BVA compares the experimentally observed standard deviations with those expected for static limit (i.e., the expected standard deviation at a certain FRET value), thus providing information on the source and timescale of any dynamics.

As expected, BVA showed that the experimental standard deviations for free dsDNA (black triangles in Fig. 3a, second panel) are close to the static limit (black continuous arc in Fig. 3a, second panel). In contrast, BVA for RP$_o$ suggests dynamic behavior as the values deviate significantly from the shot-noise expectation curve (Fig. 3a, bottom panel). In this sample, the free dsDNA population acts as an

internal control, remaining close to the static limit. These results are consistent with the hypothesis that the DNA within RP$_o$ has a dynamic component; the results further suggest that RP$_o$ can interconvert between multiple conformational states within the 0.1- to 5-ms timescale, wherein BVA is sensitive to FRET fluctuations.[27]

Transcription-bubble flexibility in RP$_o$: smFRET between nontemplate and template strands of the bubble

In a second set of experiments, the FRET ruler ("bubble DNA ruler") monitored the distance between the nontemplate strand and template strand of the transcription bubble, by monitoring FRET between a
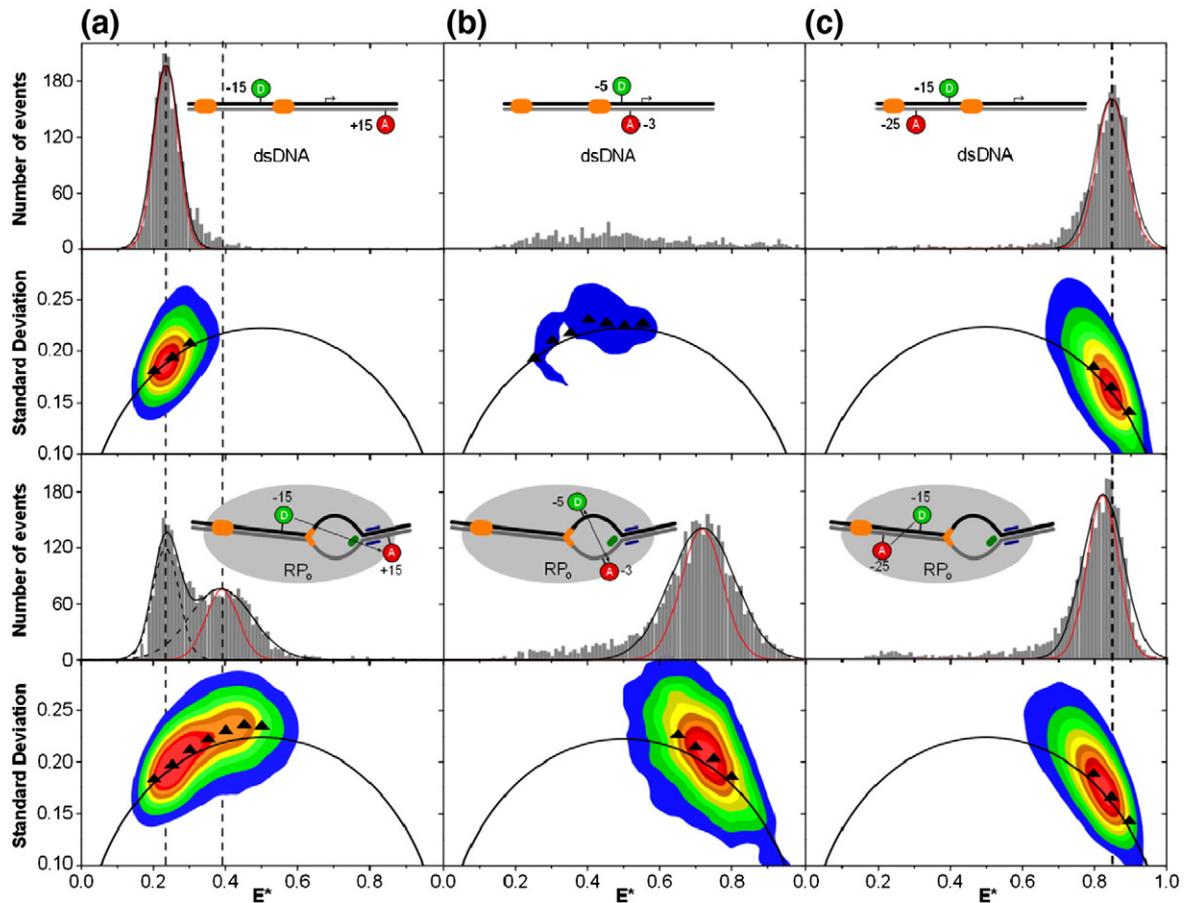
Fig. 3. Transcription-bubble flexibility in RP_o. (a) Top panel: dsDNA, *lacCONS+2*(T–3A), labeled with donor and acceptor fluorophores at positions −15 and +15, respectively (with respect to the +1 position), was analyzed using smFRET. A FRET histogram was derived from the mean values of (at least) triplicate experiments. Sizable FRET distributions were fitted with a Gaussian function (black curve) to determine the center and width of the distribution. The calculated shot-noise-limited width is shown as a red Gaussian fit. Second panel: BVA of the FRET distribution of the dsDNA. Black arc represents static limit, colored contour plots represent frequency distributions (red contour, most abundant region; blue, less abundant) and triangles represent the standard deviation of a particular part of the two-dimensional histogram of the experimental data. Third panel: samples containing dsDNA and RP_o analyzed using smFRET. Fourth panel: samples containing dsDNA and RP_o analyzed using BVA. (b) Top and second panels: dsDNA labeled with donor and acceptor fluorophores at positions −5 and −3, respectively, was analyzed by smFRET and BVA. Third and fourth panels: samples containing dsDNA and RP_o analyzed by smFRET and BVA, respectively. (c) Top and second panels: dsDNA labeled with donor and acceptor fluorophores at positions −25 and −15, respectively, was analyzed by smFRET and BVA. Third and fourth panels: samples containing dsDNA and RP_o analyzed by smFRET and BVA, respectively.

fluorophore at position −5 of the nontemplate strand and a fluorophore at position −3 of the template strand (Fig. 3b, top panel). This labeling scheme has been used previously[35] and is based on contact-mediated quenching of the two fluorophores in dsDNA followed by removal of the quenching and appearance of FRET in RP_o due to the separation of the two DNA strands in the transcription bubble. This labeling scheme allows the detection of RP_o formation with no background from free dsDNA, as the latter is not visible due to the contact-mediated quenching. As a result, free dsDNA appears as a broad, unstructured and sparsely populated FRET distribution (Fig. 3b, top

panel). Upon formation of the transcription bubble in RP_o, the fluorophores were separated to remove the contact-induced quenching and produced a high FRET distribution with mean $E^* \sim 0.72$ (Fig. 3b, third panel) and a width ($\sigma = 0.091$; Fig. 3b, black Gaussian fit, third panel) that substantially exceeds the expected shot-noise width ($\sigma = 0.057$; Fig. 3b, red Gaussian, third panel), confirming the presence of heterogeneity within RP_o. BVA analysis showed that the mean FRET standard deviation values of RP_o deviated from the static limit curve (Fig. 3b, fourth panel), consistent with the downstream ruler results that identified a dynamic component in the FRET heterogeneity.

Transcription-bubble flexibility in RP$_o$: smFRET between positions within upstream dsDNA

In a third set of control experiments, the FRET ruler ("control ruler") monitored the distance between positions within upstream dsDNA, by monitoring FRET between a fluorophore at position − 15 of the nontemplate strand and a fluorophore at position − 25 of the template strand. Since there are no known or suggested dynamics associated with this region of DNA in RP$_o$, we reasoned that the control FRET ruler would exhibit a static behavior in RP$_o$.

The apparent FRET distribution for free dsDNA was centered at mean $E^* \sim 0.85$ (Fig. 3c, top panel), whereas RP$_o$ formation led to a small decrease in the mean FRET value ($E^* \sim 0.83$; Fig. 3c, third panel). To verify that the FRET histogram for RP$_o$ actually represented a substantial amount of complex formation (typically defined as having 40–60% of the DNA being involved in RP$_o$ formation), we used fluorescence correlation spectroscopy to show that the diffusion times increased as expected for formation of RP$_o$ (Supplementary Fig. 2). The widths of the FRET distributions of free dsDNA and of RP$_o$ were similar ($\sigma = 0.047$ for dsDNA and $\sigma = 0.055$ for RP$_o$; Fig. 3c, black Gaussian fits in the first and third panels), and both distributions were close to their expected shot-noise widths ($\sigma = 0.041$ for dsDNA and $\sigma = 0.042$ for RP$_o$; Fig. 3c, red Gaussians in the first and third panels). We note that the decrease in the mean FRET efficiency between free dsDNA and RP$_o$ (from 0.85 to 0.83) suggests that part of the small increase in the FRET width (from 0.042 to 0.055) arises from the inability to resolve the free DNA and RP$_o$ distributions and, thus, is static in nature. This assessment is supported by the BVA analysis, which shows that the FRET standard deviation values for both free dsDNA and for RP$_o$ remain close to the static limit (Fig. 3c, second and fourth panels). These results suggest that the dynamic behavior observed for RP$_o$ using the downstream and bubble FRET rulers is specific to the DNA within the transcription bubble.

Reprogramming start-site selection changes the distance between DNA segments upstream and downstream of the bubble

We subsequently examined the relationship between reprogramming of start-site selection and repositioning of transcription-bubble DNA within RP$_o$. To assess the ability of initiating nucleotides to reposition transcription-bubble DNA within RP$_o$, we used a "downstream ruler" based on the *lacCONS+2*(G−2T;A+1C) promoter, which initiates transcription at positions − 1 and + 2 (confirmed by *in vitro* transcription assays; Fig. 2). We used this ruler to study the effect of the addition of initiating
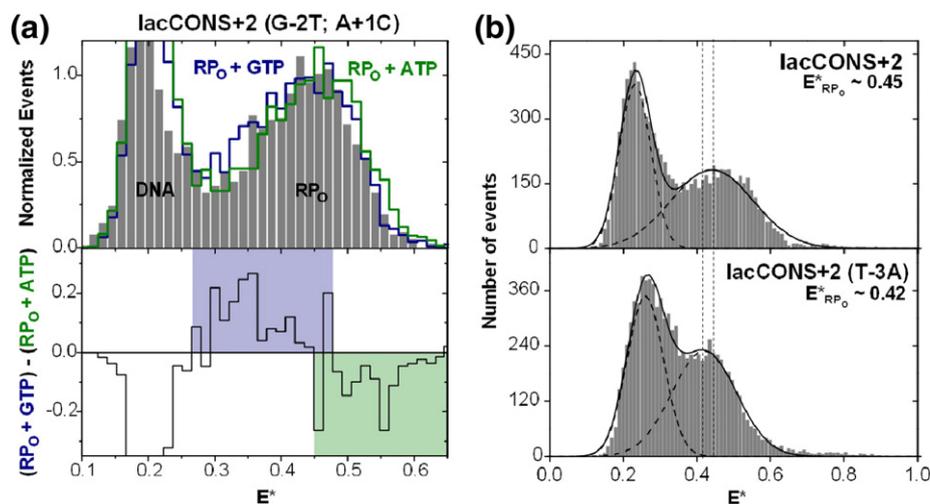
Fig. 4. Start-site reprogramming changes the distance between DNA segments upstream and downstream of the transcription bubble. (a) Reprogramming by addition of initiating nucleotides. A *lacCONS+2*(G−2T;A+1C) promoter DNA fragment that can initiate transcription at positions − 1 and + 2 was mixed with RNAP to form open complexes (RP$_o$). The dsDNA promoter was labeled with fluorophores at positions − 15 and + 15 on either side of the transcription bubble. FRET histograms of RP$_o$ alone (gray histogram), RP$_o$ with GTP (blue histogram) and RP$_o$ with ATP (green histogram) were overlaid (top panel). The difference between the FRET histograms of (RP$_o$+GTP) and (RP$_o$+ATP) was calculated (bottom panel). Both histograms were normalized to the area of the Gaussian fit function of the RP$_o$ distribution. (b) Reprogramming by base-pair substitutions in the start-site region. A *lacCONS+2*(T−3A) promoter DNA fragment that initiates transcription at a range of positions from − 2 to + 3 was incubated with RNAP to form open complexes; these complexes were compared to open complexes formed using a *lacCONS+2* promoter DNA fragment that initiates transcription from the + 1 position. Both promoter fragments were labeled with fluorophores at positions − 15 and + 15 on either side of the transcription bubble. FRET histograms of RP$_o$ of the *lacCONS+2* promoter (top panel) and *lacCONS+2*(T−3A) promoter (lower panel) were compared.

nucleotides that are complementary to, and therefore that are expected to favor, different start sites.

The FRET distribution of RP_o in the absence of nucleotides was centered at $E^* \sim 0.43$ (Fig. 4a, top panel, gray histogram). Similar to the RP_o FRET curve for the *lacCONS+2*(T–3A) promoter (Fig. 3), the RP_o FRET curve for *lacCONS+2*(G–2T;A+1C) was significantly wider than the static DNA curve, suggesting heterogeneity in RP_o. Upon addition of the initiating nucleotide GTP, which is complementary to position –1 and that therefore is expected to favor a start site at position –1, the FRET distribution shifts to slightly lower FRET values (Fig. 4a, top panel, blue histogram). In contrast, upon addition of the initiating nucleotide ATP, which is complementary to position +2 and that therefore is expected to favor a start site at position +2, the FRET distribution shifts to slightly higher FRET values upon ATP addition (Fig. 4a, top panel, green histogram). The nucleotide-dependent FRET differences are small but reproducible; a difference histogram relating the FRET histograms for RP_o+GTP and RP_o+ATP shows the FRET differences more clearly (Fig. 4a, lower panel). Analysis of a structural model of RP_o[36,37] along with estimates of the positions of the donor and acceptor fluorophores[38] (Supplementary Fig. 3) suggests that the magnitude of the observed FRET changes (from a mean $E^* \sim 0.44$ for RP_o+ATP to 0.42 for RP_o+GTP) are consistent with a small change in dye positions due to a 2-bp difference in start-site selection (from a mean donor–acceptor distance of ~72 Å for RP_o+ATP to ~76 Å for RP_o+GTP) and the corresponding changes in translocational register of transcription-bubble DNA relative to the RNAP active site.

According to our model, it may be expected that NTP binding favors a single state of RP_o, thereby making the histograms for RP_o+ATP/GTP more static and thus narrower. While the dsDNA peak in Fig. 4a always has a mean σ value of 0.03, the widths of the RP_o histograms upon nucleotide addition change slightly (for RP_o, σ=0.08; RP_o+ATP, σ=0.085; RP_o+GTP, σ=0.09). Therefore, although slight changes are observed, we do not see a decrease in the width of the FRET distributions. This may reflect the fact that the NTP concentration is not fully saturating for binding to RP_o; however, even at saturating NTP concentrations, it is entirely possible that the NTP-bound state is still dynamic, albeit biased for the NTP-based –1 or +2 translocational register, and hence, the observed widths of the histograms in our experiments may indeed represent true heterogeneity.

We also investigated the effects of reprogramming of start-site selection by base-pair substitutions in the start-site region. For this, we compared downstream rulers based on *lacCONS+2* (major start site at +1; Fig. 2) and *lacCONS+2*(T–3A) (start sites at –2, +1, +2 and +3; Fig. 2). We found that, whereas

the FRET distribution for *lacCONS+2* RP_o was centered at $E^* \sim 0.45$ (Fig. 4b, top panel), the FRET distribution for *lacCONS+2*(T–3A) RP_o was shifted to a lower FRET efficiency (mean of $E^* \sim 0.42$; Fig. 4b, lower panel). We note that this change is, in fact, more significant than the mean change in E^* suggests due to the asymmetry of the *lacCONS+2* RP_o FRET peak. We interpret this shift as reflecting a change in start-site utilization between the two promoters due to a corresponding shift in translocational register of transcription-bubble DNA relative to the RNAP active site.

Discussion

Using smFRET techniques, we have obtained evidence for the presence of DNA conformational heterogeneity and millisecond-timescale DNA conformational dynamics within the single-stranded transcription bubble of RP_o. We have observed DNA conformational heterogeneity and dynamics both in experiments assessing the apparent "length" of the transcription bubble (distances between DNA segments upstream and downstream of the transcription bubble) and in experiments assessing the apparent "width" of the transcription bubble (distances between the nontemplate and template strands of the transcription bubble). Our experiments assessing changes in the apparent length of transcription bubble upon reprogramming of start-site selection also suggest that transcription-bubble DNA conformational heterogeneity accounts for flexibility in start-site selection.

Our data support a model wherein RNAP harnesses thermally driven DNA fluctuations to access a distribution of transcription-bubble translocational registers relative to the RNAP active site, with each different translocational register corresponding to a different start site. In particular, our data support a model where transcription-bubble expansion (scrunching[24,25]) places downstream DNA in contact with the RNAP active center, facilitating the usage of downstream start sites, and transcription-bubble contraction (unscrunching[25]) places upstream DNA in contact with the RNAP active center, facilitating the usage of upstream start sites. Numerous factors, such as the DNA sequence, the availability of nucleotides and possibly the presence of transcriptional regulators, can alter the energy landscape describing the ensemble of translocation registers and therefore select one or more transcription start sites.

We have also considered whether part of the fluctuations observed may be a result of photophysical changes in the fluorophores used. We note that the proximity of a certain region of the RNAP to a fluorophore may alter the optical properties of that particular fluorophore, therefore making comparison

between dyes at different positions on the DNA difficult to interpret. However, although we cannot completely exclude the possibility that photophysics play a role in the fluctuations, the comparisons over our entire set of data and the use of control samples (such as free DNA fragments and the control ruler) indicate that the role of photophysics is likely to be minor.

In our experiments analyzing RNAP open complexes, we have assigned the first peak of the bimodal distribution to dsDNA and the second peak to RP_o. It is important to note, however, that we cannot exclude the possibility that the DNA-only peak consists of not only unbound DNA but also closed or partially closed states. Indeed, the poor fit of some of the distributions in our experiments does indicate further complexity. Multiple intermediate states in transcription initiation by σ^{70}-RNAP have been described previously (reviewed in Ref. [39]). It is therefore possible that we are detecting intermediate complexes in our analysis; additional experiments using immobilized RP_o complexes should allow these intermediate states to be studied further.

Our data suggest that at least some of the DNA dynamics occur on the timescale of milliseconds. Considering that each nucleotide addition during transcription elongation occurs on the ~30-ms timescale,[40–43] it seems likely that partial or full equilibration among transcription-bubble translocational registers may occur before the formation of the initiating dinucleotide in transcription initiation. Conclusive arguments, however, must await real-time studies of transcription-bubble DNA dynamics using immobilized complexes. In addition, methods used in this study should enable the analysis of transcription-bubble conformational heterogeneity during transcription elongation, pausing and termination.

Materials and Methods

DNA and reagents

Amino-modified oligonucleotides (IBA, Germany) were internally labeled with fluorophores Cy3B (Invitrogen, USA) and ATTO647N (ATTO-TEC, Germany), as previously described,[35] and purified using gel electrophoresis. Single-stranded DNAs were annealed in hybridization buffer [50 mM Tris–HCl (pH 8.0), 1 mM ethylenediaminetetraacetic acid and 500 mM NaCl]. Sequences of DNAs and the labeling schemes used are shown in Supplementary Fig. 1.

Formation of RNAP open complexes and initial transcribing complexes

According to published procedures,[24,35,44,45] open complexes (RP_o) were formed by mixing dsDNA (10 nM) and *E. coli* RNAP holoenzyme (50 nM; Epicentre, USA) in a total volume of 20 µl KG7 buffer [40 mM Hepes–NaOH

(pH 7), 100 mM potassium glutamate, 10 mM $MgCl_2$, 1 mM DTT, 100 µg/ml bovine serum albumin, 5% glycerol and 1 mM mercaptoethylamine] and subsequent incubation at 37 °C for 15 min. After incubation, heparin Sepharose-coated beads (1 mg/ml; GE Healthcare) were added to disrupt nonspecific RNAP–DNA complexes and to remove free RNAP. After 30 s at 37 °C, samples were centrifuged, and 13 µl of supernatant was transferred to a pre-warmed tube. Wherever indicated, ribonucleotides ATP and GTP were added to the KG7 buffer at a concentration of 1 mM after RP_o formation.

In vitro transcription assays

The *in vitro* transcription reaction mixtures were set up by adding 0.24 U RNAP (Epicentre, USA), 10 nM dsDNA promoter, 12 U RNasin (Promega, USA), 50 µM UTP, 50 µM CTP, 50 µM ATP (Fermentas, UK) and 0.3 µCi/µl [α^{32}P]GTP [10 µCi/µl (PerkinElmer)] to 1× KG7 buffer [40 mM Hepes–NaOH (pH 7), 100 mM potassium glutamate, 10 mM $MgCl_2$, 100 µg/ml bovine serum albumin, 1 mM DTT, 1 mM mercaptoethylamine and 5% glycerol] and incubated for 5 min at 37 °C. Heparin Sepharose (1 mg/ml; GE Healthcare) was added, and the reaction was allowed to continue at 37 °C for a further 55 min. Reactions were stopped by addition of 5 µl of loading dye (90% formamide, 10 mM ethylenediaminetetraacetic acid, bromophenol blue and xylene cyanol), and mixtures were incubated for 5 min at 95 °C before being loaded on a 6-M urea, 20% polyacrylamide sequencing gel and visualized by autoradiography.

Single-molecule fluorescence spectroscopy

A custom-built confocal microscope was used for smFRET experiments as previously described.[28,46] The setup was modified allowing ALEX of donor and acceptor fluorophores.[30,31] For this purpose, the fiber-coupled outputs of a green (532 nm, Samba; Cobolt, Sweden) and a red (638 nm; Cube Coherent, USA) laser were alternated with a modulation frequency of 10 kHz. Both beams were spatially filtered and coupled into an inverted confocal microscope (IX71; Olympus, Germany) equipped with an oil-immersion objective (60×, 1.35 NA, UPLSAPO 60XO; Olympus, Germany). In a typical experiment, the average excitation intensities were 250 µW at 532 nm and 60 µW at 635 nm. The same objective was used to collect the resulting fluorescence; the emission was separated from excitation light by a dichroic mirror, focused onto a 200-µm pinhole and subsequently split spectrally on two avalanche photodiodes (SPCM-AQR-14; PerkinElmer, UK) detecting the donor and acceptor fluorescence with two distinct spectral filters (green, 585DF70; red, 650LP). Custom-made LabVIEW software was used to register and evaluate the detector signal. For all experiments, the temperature of the sample was set to 37±1 °C using a custom-made heated collar attached to the objective, which was connected to a heating bath.

Data analysis

Fluorescence photons were assigned to either donor-based (D_{exc}) or acceptor-based (A_{exc}) excitation with

respect to their photon arrival time (donor detection channel, D_{em}; acceptor detection channel, A_{em}).[30,31] Two characteristic ratios, fluorophore stoichiometries S and apparent FRET efficiencies E^*, were calculated for each fluorescent burst above a certain threshold yielding a two-dimensional histogram. Stoichiometry S is the ratio between the overall green fluorescence intensity over the total green and red fluorescence intensity and describes the ratio of donor-to-acceptor fluorophores within a diffusing molecule.[30,31] The uncorrected FRET E^* efficiency [defined as $D_{exc}A_{em}/(D_{exc}A_{em} + D_{exc}D_{em})$] monitors the proximity between the two fluorophores. We selected bursts characterized by three parameters (M, T and L) from the data. In this analysis, a fluorescent signal is considered a burst if a total of L photons having M neighboring photons arrive at the detector within a time interval of T microseconds. Acceptor-containing molecules were identified by applying a burst search on $A_{exc}A_{em}$ with parameters $M=7$, $T=500\ \mu s$ and $L=12$. We additionally applied per-bin thresholds to remove spurious changes in fluorescence intensity and to select for bright donor–acceptor molecules ($A_{exc}A_{em} > 30$–100 photons). One-dimensional E^* distributions for donor–acceptor species were obtained by using a $0.45 < S < 0.8$ threshold. These E^* distributions could be fitted using a Gaussian function, yielding the mean E^* value for a certain distribution and an associated standard deviation σ. BVA analysis was performed as described previously.[27]

Fluorescence correlation spectroscopy

The same microscope and experimental configuration as described above was used for fluorescence correlation spectroscopy measurements. Excitation was at 532 nm in continuous-wave fashion (150 μW). Photon-by-photon arrival times in the donor and acceptor channels were correlated using a hardware correlator. Data in the manuscript were derived from autocorrelation in the red detection channel after green excitation to detect doubly labeled species.

Supplementary Data

† N.C.R., T.C. and L.C.H. contributed equally to this work.

Present addresses: T. Cordes, Molecular Microscopy Research Group and Single-Molecule Biophysics, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands; L. C. Hwang, Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

Abbreviations used:
RNAP, RNA polymerase; smFRET, single-molecule FRET; dsDNA, double-stranded DNA; BVA, burst variance analysis.

References

1. Haugen, S. P., Ross, W. & Gourse, R. L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat. Rev., Microbiol.* **6**, 507–519.
2. Jeong, W. & Kang, C. (1994). Start site selection at *lac*UV5 promoter affected by the sequence context around the initiation sites. *Nucleic Acids Res.* **22**, 4667–4672.
3. Liu, J. & Turnbough, C. L., Jr. (1994). Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the *pyrC* promoter in *Escherichia coli. J. Bacteriol.* **176**, 2938–2945.
4. Carpousis, A. J., Stefano, J. E. & Gralla, J. D. (1982). 5′ Nucleotide heterogeneity and altered initiation of transcription at mutant *lac* promoters. *J. Mol. Biol.* **157**, 619–633.
5. Minkley, E. G. & Pribnow, D. (1973). Transcription of the early region of bacteriophage T7: selective initiation with dinucleotides. *J. Mol. Biol.* **77**, 255–277.
6. Hoffman, D. J. & Niyogi, S. K. (1973). RNA initiation with dinucleoside monophosphates during transcription of bacteriophage T4 DNA with RNA polymerase of *Escherichia coli. Proc. Natl Acad. Sci. USA*, **70**, 574–578.
7. Hoffman, D. J. & Niyogi, S. K. (1973). Differential effects of sigma factor, ionic strength, and ribonucleoside

triphosphate concentration on the transcription of phage T4 DNA with ribonucleic acid polymerase of *Escherichia coli. Biochim. Biophys. Acta*, **299**, 588–595.

8. Di Nocera, P. P., Avitabile, A. & Blasi, F. (1975). *In vitro* transcription of the *Escherichia coli* histidine operon primed by dinucleotides. Effect of the first histidine biosynthetic enzyme. *J. Biol. Chem.* **250**, 8376–8381.

9. Dausse, J. P., Sentenac, A. & Fromageot, P. (1975). Interaction of RNA polymerase from *Escherichia coli* with DNA. Analysis of T7 DNA early-promoter sites. *Eur. J. Biochem.* **57**, 569–578.

10. Smagowicz, W. J. & Scheit, K. H. (1978). Primed abortive initiation of RNA synthesis by *E. coli* RNA polymerase on T7 DNA. Steady state kinetic studies. *Nucleic Acids Res.* **5**, 1919–1932.

11. Grachev, M. A., Zaychikov, E. F., Ivanova, E. M., Komarova, N. I., Kutyavin, I. V., Sidelnikova, N. P. & Frolova, I. P. (1984). Oligonucleotides complementary to a promoter over the region −8…+2 as transcription primers for *E. coli* RNA polymerase. *Nucleic Acids Res.* **12**, 8509–8524.

12. Learned, R. M. & Tjian, R. (1982). *In vitro* transcription of human ribosomal RNA genes by RNA polymerase I. *J. Mol. Appl. Genet.* **1**, 575–584.

13. Samuels, M., Fire, A. & Sharp, P. A. (1984). Dinucleotide priming of transcription mediated by RNA polymerase II. *J. Biol. Chem.* **259**, 2517–2525.

14. Grove, A., Adessa, M. S., Geiduschek, E. P. & Kassavetis, G. A. (2002). Marking the start site of RNA polymerase III transcription: the role of constraint, compaction and continuity of the transcribed DNA strand. *EMBO J.* **21**, 704–714.

15. Lewis, D. E. & Adhya, S. (2004). Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol. Microbiol.* **54**, 692–701.

16. Emory, S. A., Bouvet, P. & Belasco, J. G. (1992). A 5′-terminal stem–loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev.* **6**, 135–148.

17. de Smit, M. H. & van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl Acad. Sci. USA*, **87**, 7668–7672.

18. de Smit, M. H. & van Duin, J. (1990). Control of prokaryotic translational initiation by mRNA secondary structure. *Prog. Nucleic Acid Res. Mol. Biol.* **38**, 1–35.

19. Kammerer, W., Deuschle, U., Gentz, R. & Bujard, H. (1986). Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. *EMBO J.* **5**, 2995–3000.

20. Xiong, X. F. & Reznikoff, W. S. (1993). Transcriptional slippage during the transcription initiation process at a mutant *lac* promoter *in vivo*. *J. Mol. Biol.* **231**, 569–580.

21. Baker, C. C. & Ziff, E. B. (1981). Promoters and heterogeneous 5′ termini of the messenger RNAs of adenovirus serotype 2. *J. Mol. Biol.* **149**, 189–221.

22. Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**, 465–503.

23. Kuehner, J. N. & Brow, D. A. (2006). Quantitative analysis of *in vivo* initiator selection by yeast RNA polymerase II supports a scanning model. *J. Biol. Chem.* **281**, 14119–14128.

24. Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S. & Ebright, R. H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, **314**, 1144–1147.

25. Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, **314**, 1139–1143.

26. Santoso, Y., Torella, J. P. & Kapanidis, A. N. (2010). Characterizing single-molecule FRET dynamics with probability distribution analysis. *ChemPhysChem*, **11**, 2209–2219.

27. Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J. & Kapanidis, A. N. (2011). Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys. J.* **100**, 1568–1577.

28. Santoso, Y., Joyce, C. M., Potapova, O., Le Reste, L., Hohlbein, J., Torella, J. P. *et al.* (2010). Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl Acad. Sci. USA*, **107**, 715–720.

29. Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonjia, E., Ebright, Y. W. & Ebright, R. H. (2001). Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell*, **106**, 453–463.

30. Kapanidis, A. N., Laurence, T. A., Lee, N. K., Margeat, E., Kong, X. & Weiss, S. (2005). Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38**, 523–533.

31. Kapanidis, A. N., Lee, N. K., Laurence, T. A., Doose, S., Margeat, E. & Weiss, S. (2004). Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl Acad. Sci. USA*, **101**, 8936–8941.

32. Nir, E., Michalet, X., Hamadani, K. M., Laurence, T. A., Neuhauser, D., Kovchegov, Y. & Weiss, S. (2006). Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B*, **110**, 22103–22124.

33. Kalinin, S., Sisamakis, E., Magennis, S. W., Felekyan, S. & Seidel, C. A. M. (2010). On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J. Phys. Chem. B*, **114**, 6197–6206.

34. Holden, S. J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O. J. & Kapanidis, A. N. (2010). Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophys. J.* **99**, 3102–3111.

35. Cordes, T., Santoso, Y., Tomescu, A. I., Gryte, K., Hwang, L. C., Camara, B. *et al.* (2010). Sensing DNA opening in transcription using quenchable Forster resonance energy transfer. *Biochemistry*, **49**, 9171–9180.

36. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme–DNA complex. *Science*, **296**, 1285–1290.

37. Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E. & Ebright, R. H. (2012). Structural basis of transcription initiation. *Science*, **338**, 1076–1080.
38. Muschielok, A., Andrecka, J., Jawhari, A., Bruckner, F., Cramer, P. & Michaelis, J. (2008). A nano-positioning system for macromolecular structural analysis. *Nat. Methods*, **5**, 965–971.
39. Saecker, R. M., Record, M. T., Jr & Dehaseth, P. L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase–promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* **412**, 754–771.
40. Adelman, K., La Porta, A., Santangelo, T. J., Lis, J. T., Roberts, J. W. & Wang, M. D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl Acad. Sci. USA*, **99**, 13538–13543.
41. Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, **426**, 684–687.
42. Yin, H., Wang, M. D., Svoboda, K., Landick, R., Block, S. M. & Gelles, J. (1995). Transcription against an applied force. *Science*, **270**, 1653–1657.
43. Forde, N. R., Izhaky, D., Woodcock, G. R., Wuite, G. J. & Bustamante, C. (2002). Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA*, **99**, 11682–11687.
44. Kapanidis, A. N., Margeat, E., Laurence, T. A., Doose, S., Ho, S. O., Mukhopadhyay, J. *et al.* (2005). Retention of transcription initiation factor σ[70] in transcription elongation: single-molecule analysis. *Mol. Cell*, **20**, 347–356.
45. Kapanidis, A. N. & Strick, T. (2009). Biology, one molecule at a time. *Trends Biochem. Sci.* **34**, 234–243.
46. Santoso, Y. & Kapanidis, A. N. (2009). Probing biomolecular structures and dynamics of single molecules using in-gel alternating-laser excitation. *Anal. Chem.* **81**, 9561–9570.