

Single-molecule tracking of the transcription cycle by sub-second RNA detection

Zhengjian Zhang^{1*†}, Andrey Revyakin^{1**†}, Jonathan B Grimm¹, Luke D Lavis¹, Robert Tjian^{1,2,3}

¹Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, United States; ²Li Ka Shing Center for Biomedical and Health Sciences, University of California, Berkeley, Berkeley, United States; ³Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, United States

Abstract Transcription is an inherently stochastic, noisy, and multi-step process, in which fluctuations at every step can cause variations in RNA synthesis, and affect physiology and differentiation decisions in otherwise identical cells. However, it has been an experimental challenge to directly link the stochastic events at the promoter to transcript production. Here we established a fast fluorescence in situ hybridization (fastFISH) method that takes advantage of intrinsically unstructured nucleic acid sequences to achieve exceptionally fast rates of specific hybridization ($\sim 10^7 \text{ M}^{-1}\text{s}^{-1}$), and allows deterministic detection of single nascent transcripts. Using a prototypical RNA polymerase, we demonstrated the use of fastFISH to measure the kinetic rates of promoter escape, elongation, and termination in one assay at the single-molecule level, at sub-second temporal resolution. The principles of fastFISH design can be used to study stochasticity in gene regulation, to select targets for gene silencing, and to design nucleic acid nanostructures.

DOI: 10.7554/eLife.01775.001

*For correspondence: zhangzh@janelia.hhmi.org (ZZ); revyakina@janelia.hhmi.org (AR)

[†]These authors contributed equally to this work

[‡]Present address: University of Leicester, Leicester, United Kingdom

Competing interests: See page 17


Funding: See page 18

Received: 24 October 2013

Accepted: 06 December 2013

Published: 28 January 2014

Reviewing editor: James T Kadonaga, University of California, San Diego, United States

 Copyright Zhang et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Transcription is the first, and frequently the most regulated step in the flow of genetic information from DNA to protein. Transcription is a dynamic, multi-step process in which RNA polymerase (RNAP) (i) binds to the promoter to form the closed complex (RP_c); (ii) melts the promoter to form the open complex (RP_o); (iii) performs several abortive cycles of synthesis and release of short 2–12 nt RNA products as the initial transcribing complex (RP_{itc}); (iv) escapes the promoter; (v) undergoes some promoter-proximal pausing; (vi) forms an elongating complex (RD_e) whose processivity can either be interrupted by more pauses or stimulated by trailing RNAPs and elongation factors; and, finally (vii) terminates at the end of the transcription unit (reviewed in *DeHaseth et al., 1998; Murakami and Darst, 2003; Cheung and Cramer, 2012*). Due to the low copy number of genes in a cell (usually one in prokaryotes and two in eukaryotes), molecular fluctuations at any of the above steps may cause large cell-to-cell variability in the amount of the final RNA transcript produced in populations of otherwise genetically identical cells grown under identical conditions, and thus can affect gene expression, and cell physiology (*Ogawa, 1993; Raj and Van Oudenaarden, 2009; Yamanaka, 2009; Gupta et al., 2011; Lionnet and Singer, 2012*). Therefore, understanding how molecular fluctuations at different steps of the transcription cycle alter transcriptional outcomes is required to dissect the mechanism of gene regulation.

Single-molecule techniques are ideally suited to directly monitor molecular fluctuations in multi-step reactions in real-time, without averaging out their inherent stochasticity (*Weiss, 1999*), and have provided important insights into the dynamics of transcription, unattainable by conventional ensemble

eLife digest The body produces proteins by transcribing DNA (genes) to make messenger RNA, which is then translated to make a protein. Transcription begins when an enzyme called RNA polymerase binds to the DNA and catalyzes the process by which genetic information from the double helix is copied to a complementary RNA transcript, which subsequently becomes the messenger RNA.

Because a living cell usually contains only one or a few copies (alleles) of a given gene, molecular fluctuations play a crucial role in cellular transcription. Therefore, studying transcription kinetics at the level of single molecules may provide critical insights into how cells deal with—or even take advantage of—molecular fluctuations. A number of different single-molecule techniques can be used to follow transcription, but these techniques are often relatively slow compared to transcription in living cells, or they suffer from other problems such as only being able to study one step in the transcription process.

Now, Zhang, Revyakin et al. have systematically devised a technique called ‘fastFISH’ that is fast enough to track the production of single RNA molecules directly and instantaneously. FastFISH builds on an existing technique called FISH—short for fluorescence in situ hybridization—in which fluorescent molecules are attached to single strands of DNA or RNA. These single strands pair with specific regions of complementary DNA or RNA molecules, and they can be visualized with a fluorescence microscope. However, conventional FISH is a ‘snap-shot’ technique that is not suitable for making real-time observations under physiological conditions.

FastFISH relies on single strands of fluorescently labeled DNA and RNA that bind to complementary strands of DNA or RNA extremely quickly, even under physiological conditions, because they contain only three of the four ‘regular’ nucleotides that make up DNA or RNA. As a proof of principle, Zhang, Revyakin et al. used fastFISH to study the kinetics of transcription by the bacteriophage T7 RNA polymerase and were able to measure multiple stages of the transcription cycle in a single-molecule experimental setup.

By allowing each stage of transcription to be tracked in real-time at the level of single-molecules, fastFISH will permit a more in-depth analysis of the factors that regulate how genes are expressed as proteins in our cells. Moreover, the ability to design single-strand probes that bind rapidly to DNA and RNA targets could have many additional applications, including new strategies for more efficient gene silencing.

DOI: [10.7554/eLife.01775.002](https://doi.org/10.7554/eLife.01775.002)

methods (Bai et al., 2006). For instance, single-molecule assays based on optical nanomanipulation have revealed pausing and backtracking of RD_e (Neuman et al., 2003; Shaevitz et al., 2003; Galburt et al., 2007) and measured the force and the torque generated by RD_e (Wang et al., 1998; Ma et al., 2013). Methods based on magnetic nanomanipulation and single-molecule fluorescence resonance energy transfer have probed conformational changes in DNA and RNAP within RP_o , and RP_{ic} (Kapanidis et al., 2006; Revyakin et al., 2006; Tang et al., 2009; Chakraborty et al., 2012; Robb et al., 2013). Finally, single-molecule localization studies have probed the dynamics of the initial promoter search by RNAP (Wang et al., 2012; Friedman et al., 2013). However, most current single-molecule methods focus on only one step of transcription, and are not well suited to relate protein-DNA complex assembly and dynamics during multiple stages of the transcription cycle to RNA production. In addition, most methods do not measure RNA production directly, but rather infer it from changes in DNA conformation or movement of RNAP along the DNA. Most recently, single-molecule tracking of key protein-DNA interactions coupled with detection of the RNA production has been demonstrated in the bacterial (Friedman and Gelles, 2012) and human transcription systems (Revyakin et al., 2012). The former study achieved time resolution for RNA detection on the order of ~ 10 s, and thus provided a dynamic, quantitative view of the full transcription cycle of bacterial RNAP. However, time scales of many events in transcription are on the order of ~ 1 s, particularly under physiological conditions (for instance, the residence time of transcription factors on DNA (reviewed in Hager et al., 2009), the rate of initiation and promoter clearance by RNAPs (Revyakin et al., 2006; Tang et al., 2009), and the expected time delay between cooperatively elongating RNAPs molecules (Epshtein and Nudler, 2003)). Thus, a real-time method for nascent transcript detection at ~ 1 s time scales would significantly enhance our

ability to dissect the dynamics of the transcription process, and to correlate stochastic fluctuations at different steps with transcriptional outcomes.

Currently, the most sensitive and specific methods for detecting RNA transcripts continue to rely on complementary nucleic acid hybridization. However, oligonucleotide probes typically hybridize several orders of magnitude slower than diffusion under physiological conditions (effective rate constant k_{on} , less than $10^5 \text{ M}^{-1} \text{ s}^{-1}$, [Chan et al., 1995 and references therein]) and, as a result, do not permit real-time nascent RNA detection with common single-molecule setups, such as total internal reflection fluorescence (TIRF) microscopy. Here we present a fast fluorescence in situ hybridization method (fastFISH) to detect synthesis of nascent RNA transcripts at sub-second time resolution, at the single-molecule level. The method takes advantage of our finding that single-stranded nucleic acid probes with sequences comprised of just three of the four bases (A, U and C for RNA probes, and A, T, and G for the complementary DNA targets) are intrinsically unstructured and, as a result, hybridize at exceptionally fast rates ($\sim 10^7 \text{ M}^{-1} \text{ s}^{-1}$) without compromising sequence specificity. As a proof of concept, we applied fastFISH to probe the production of nascent RNA transcripts by the speedy bacteriophage T7 RNA Polymerase (T7 RNAP) in vitro. Furthermore, by using fastFISH in combination with fluorescently labeled RNAP, we dissected the full T7 RNAP transcription cycle (promoter binding, promoter escape, elongation, and termination). FastFISH can be used to study transcription by multi-subunit prokaryotic and eukaryotic RNA polymerases at the level of stochastic molecular interactions. The rules for generating fastFISH probe-target pairs can also be used for gene silencing, gene profiling and bottom-up assembly of nucleic acid nanostructures (Rothmund, 2006).

Results

Design and characterization of fastFISH hybridization probes

To achieve real-time nascent RNA detection, we set out to find a general rule for designing RNA-probe pairs that hybridize at the fastest possible rates. Hybridization of two single-stranded nucleic acid fragments requires unfolding of the fragments into unstructured coils, which then anneal to form an intermolecular base paired helix (Lima et al., 1992 and references therein). In support of this notion, nucleic acid probes with less stable secondary structures show faster hybridization rates to complementary nucleic acid targets (Lima et al., 1992; Kushon et al., 2001; Wang and Drlica, 2004; Gao et al., 2006; Yilmaz et al., 2006). Likewise, decreasing the length of hairpins in conventional molecular beacons increases their hybridization rates (Tsourkas et al., 2003). Thus, we reasoned that intrinsically unstructured nucleic acid oligonucleotide pairs of optimal lengths should have the fastest hybridization kinetics under physiological conditions without compromising specificity.

To systematically examine the propensity of RNA sequences to form base-paired secondary structures, we used the nucleic acid structure prediction tool Mfold (Zuker, 2003) to calculate the free energy of self-folding, $\Delta G_{37^\circ\text{C}}$, of randomly selected short RNA sequences ($N = 338,417$). We chose RNA sequence with length of 19 nt and GC content between 0.4 and 0.6 as representative of a typical oligonucleotide primer. We found that $\Delta G_{37^\circ\text{C}}$ values were mostly negative and widely distributed ($-1.7 \pm 1.7 \text{ kcal mol}^{-1}$), indicating that an average random 19-mer RNA sequence is mostly structured (Figure 1A). Next, we closely examined a subset of 19-mer RNA sequences that had positive $\Delta G_{37^\circ\text{C}}$ ($>+2 \text{ kcal mol}^{-1}$, $N = 2,768$, $<1\%$ of the total pool of random sequences), and found that these mostly unstructured sequences were composed predominantly of only three bases A, U, and C ($\sim 84\%$ had only 2 G residues or less, significantly less than the 4.75 G residues expected on average, Figure 1—figure supplement 1). The bias towards the lower G-content in the unstructured sequences was not surprising, because guanine is the most potent in base-pairing interactions: it forms a triple hydrogen bond with cytosine and a wobble pair with uracil. We then calculated $\Delta G_{37^\circ\text{C}}$ for randomly picked 19-mers composed of only A, U, and C ($N = 99,777$, GC content between 0.4 and 0.6, Figure 1A), and found that these AUC-sequences had mostly positive $\Delta G_{37^\circ\text{C}}$, with a much narrower distribution ($+2.5 \pm 0.6 \text{ kcal mol}^{-1}$, Figure 1A). Analysis of other three-base-derived RNA sequences (AUG, CAG, CUG) indicated that AUC-sequences were unique in their largely positive $\Delta G_{37^\circ\text{C}}$ (Figure 1A). Calculation of $\Delta G_{37^\circ\text{C}}$ for random DNA 19-mers (that could be used as complementary probes for RNA targets) indicated that ATG-based and ATC-based DNA 19-mers were also significantly less structured than their four-base counterparts (Figure 1A). Therefore, we hypothesized that the use of the AUC alphabet for RNA targets, and ATG alphabet for DNA probes should allow the fastest annealing reactions under physiological temperature of 37°C .

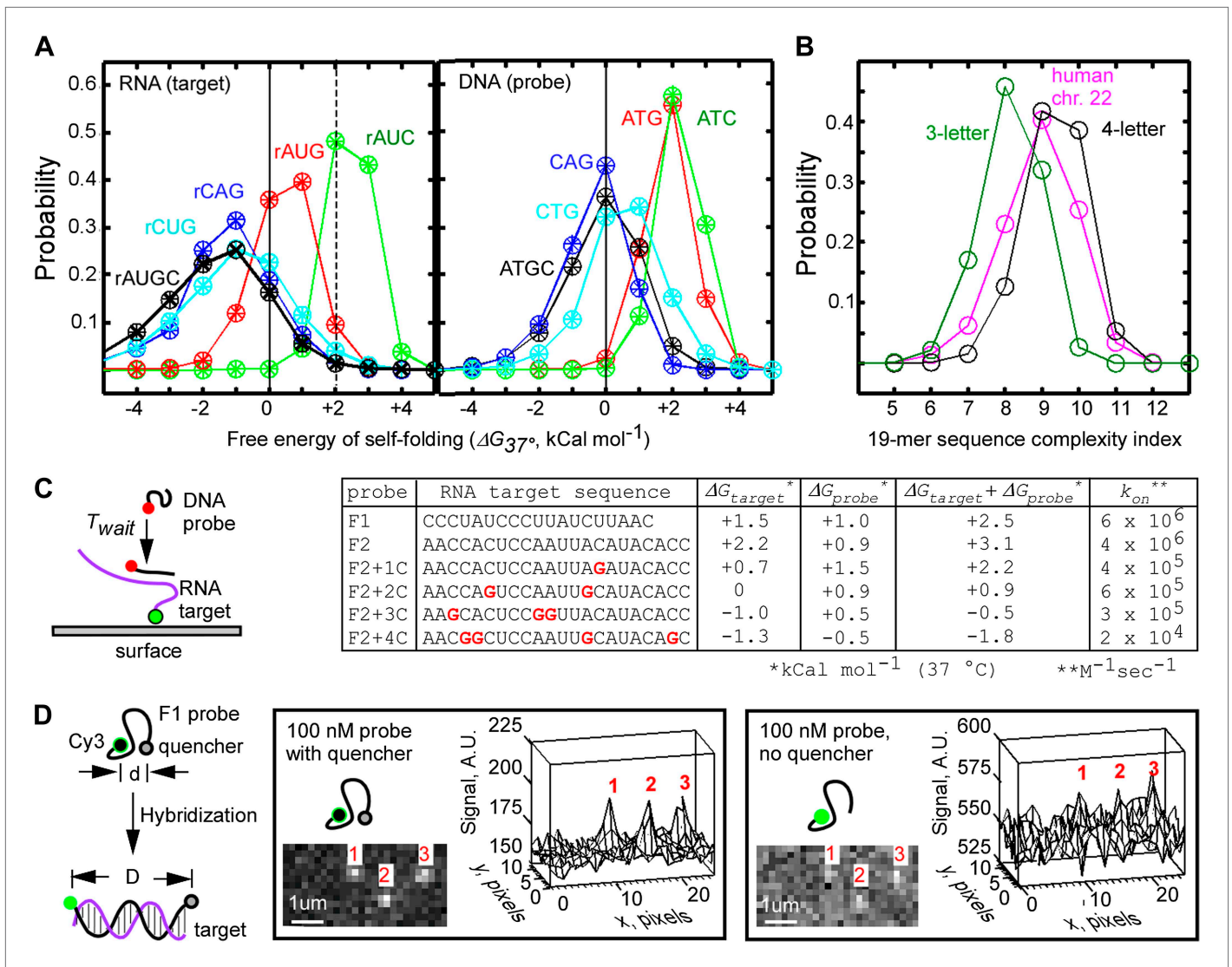


Figure 1. Design of fastFISH probe-target pairs. **(A)** Probability distribution of Mfold-calculated free energies of self-folding of randomly selected, single-stranded 19-mer RNA (left) and DNA (right) oligonucleotides, composed of three or four bases. Results of analysis of three independent sets are shown as '+', 'x', and 'o'. About 100,000 three-letter sequences, and about 300,000 four-letter sequences were analyzed in each set. **(B)** Lempel-Ziv complexity analysis of three-letter 19-mer oligonucleotides (one set of ~100,000 AUC sequences), four-letter 19-mer oligonucleotides (one set of ~300,000 AUGC sequences), and all tiling 19-mers from the exome of the human chromosome 22. **(C)** Single-molecule measurements of hybridization rates of fastFISH probe-target pairs, and the effect of G-residues in the targets on the rates. Left: schematic of experiment. Cy3-labeled 90-base RNA oligonucleotides containing a single target sequence were immobilized on a glass surface through a biotin moiety at the 3' end. Atto633-labeled DNA probes were injected into the imaging flow cell, and their hybridization was detected using TIRF/CoSMoS to obtain the probe arrival time T_{wait} . Right: table of RNA target sequences, Mfold-calculated free energies of self-folding of RNA targets (ΔG_{target}), DNA probes (ΔG_{probe}), combined energies of targets and probes, and on-rates calculated based on probe T_{wait} and concentrations. **(D)** Self-quenching approach to reduce fluorescence background from unbound DNA probes in TIRF imaging of probe-target hybridization. Left: schematic of experiment. A quencher (e.g., Iowa Black FQ) is placed on one end of a DNA probe, and a fluorophore (e.g., Cy3) is placed on the opposite end of the DNA probe. The short persistence length of single-stranded DNA ($l_p \sim 0.8$ nm, **Smith et al., 1996; Dessinges et al., 2002**) enables quenching of Cy3. Upon hybridization to the target, the distance between Cy3 and Iowa Black FQ increases due to the larger $l_p \sim 50$ nm of the DNA-RNA duplex, leading to an increase of Cy3 fluorescence. Middle: representative TIRF image and a corresponding three-dimensional plot of target-hybridized F1 probes acquired in the presence of unbound, self-quenched F1 probe at 100 nM. Right: same set of molecules imaged in the presence of unbound, unquenched F1 probe at 100 nM.

DOI: 10.7554/eLife.01775.003

The following figure supplements are available for figure 1:

Figure supplement 1. Sample sequences of the unstructured 19-mer RNA oligos.

DOI: 10.7554/eLife.01775.004

Figure 1. Continued on next page

Figure 1. Continued

Figure supplement 2. Single-molecule measurements of fastFISH probe-target hybridization rates using TIRF/CoSMoS.

DOI: [10.7554/eLife.01775.005](https://doi.org/10.7554/eLife.01775.005)

Figure supplement 3. Reduction of fluorescence background of free fastFISH probes by self-quenching: additional representative images.

DOI: [10.7554/eLife.01775.006](https://doi.org/10.7554/eLife.01775.006)

To ensure that the use of the three-base alphabet did not compromise the specificity of probe-target hybridization, we calculated the textual complexity of randomly picked three-letter 19-mers using the algorithm of Lempel and Ziv ([Ziv and Lempel, 1976](#); [Kaspar and Schuster, 1987](#); [Orlov and Potapov, 2004](#)). This algorithm, commonly used in lossless data compression programs, calculates the minimal number of operations required to reconstruct a sequence of symbols by copying and inserting segments of an existing sub-sequence. Thus, nucleic acid sequences that contain repetitive elements would have a lower LZ complexity (and would be less specific in hybridization) ([Wright and Church, 2002](#)). We found that three-letter 19-mer sequences had complexity indices of 8.1 ± 0.8 ($N = 99,777$), while four-letter 19-mers had complexity indices of 9.3 ± 0.8 ($N = 338,417$, [Figure 1B](#)). As a comparative reference, we calculated complexity indices for tiling 19-mers in all exons of human chromosome 22 to be 8.9 ± 1.0 ([Figure 1B](#)). Importantly, a significant fraction of three-letter 19-mers (~31%) had complexity indices of 9 and higher, which matched or exceeded the average complexity of human exons. These calculations indicate that a significant fraction of the random probe sequences composed of only three bases nevertheless retain sequence complexity and specificity that match their physiological, four-base derived counterparts.

We applied the AUC rule and the complexity filter to generate two candidate probe-target pairs (referred to as F1 and F2). The calculated $\Delta G_{37^\circ\text{C}}$ for the F1 and F2 pairs were +1.5 and +2.2 kcal mol⁻¹, respectively, for the AUC-based RNA targets, and +1.0 and +0.9 kcal mol⁻¹ for their ATG-based complementary DNA probes ([Figure 1C](#)), suggesting that the pairs are mostly unstructured and likely to be fast-hybridizing. Indeed, the F1 and F2 probes annealed to their surface-immobilized RNA targets at exceptionally fast rates— $6 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ and $4 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, respectively, as measured by TIRF-based Colocalization Single-Molecule Spectroscopy (CoSMoS, [Friedman et al., 2006](#)) ([Figure 1C](#), [Figure 1—figure supplement 2](#)). These on-rates were at least 100-fold faster than typical four-base nucleic acid probes of similar lengths reported in literature ([Lima et al., 1992](#); [Kushon et al., 2001](#); [Wang and Drlica, 2004](#); [Friedman et al., 2006](#); [Gao et al., 2006](#); [Yilmaz et al., 2006](#)). Consistent with the AUC rule, introducing back one or more G residues into the F2 RNA target sequence decreased the rate of its hybridization to the complementary DNA probe (10-fold reduction for one G residue and 300-fold reduction for four G residues), which correlated with the progressively lower free energies of self-folding ([Figure 1C](#), [Figure 1—figure supplement 2D](#)). We conclude that the target sequences designed using the AUC rule, combined with the complexity filter, achieved our goal of generating superior hybridization kinetics for fast transcript detection.

Although the F1 and F2 probes were exceptionally fast, detection of RNA on the sub-second time scales would require the use of fluorescent probe concentrations above 100 nM ([Figure 1—figure supplement 2D](#)). Such concentrations are generally incompatible with common single-molecule detection setups such as TIRF microscopy, due to high fluorescence background from the freely diffusing, unbound probe molecules. We overcame this problem by attaching a quencher to the single-stranded probe at the end opposite of the fluorophore ([Marras et al., 2002](#)). This self-quenching strategy enabled the use of free probes at concentrations up to 1 μM ([Figure 1D](#), [Figure 1—figure supplement 3](#)). We believe that, in the absence of the secondary structure, the self-quenching effect was likely mediated by random polymer motion and/or contact quenching ([Johansson et al., 2002](#); [Marras et al., 2002](#)). We refer to our method for fast nucleic acid detection using unstructured, sequence specific, self-quenched fluorescent probes as 'fastFISH'.

Real-time single-molecule detection of transcription with fastFISH

To demonstrate that fastFISH can detect nascent transcripts in real time at single-molecule resolution, we used the bacteriophage T7 RNAP. This single-subunit RNAP is an excellent test case for fastFISH: at physiological temperature (37°C) it initiates transcripts at an effective rate of at least 1 s⁻¹ ([Martin and Coleman, 1987](#)), and elongates transcripts at ~240 nt s⁻¹ ([Golomb and Chamberlin, 1974](#); [Bonner et al., 1994](#)). As a model template, we used a fluorescently labeled linear DNA fragment

containing the consensus promoter for T7 RNAP (Milligan *et al.*, 1987) and the F1 target sequence (positioned between +28 and +46 from the transcription start site, +1), and ending at +295. We surface-immobilized the Cy3-labeled DNA via a biotin moiety attached to the upstream end of the template (position -75) (Figure 2A). In this configuration, RNAP molecules were expected to initiate transcription at the promoter, elongate the nascent RNA in the direction away from the surface, and run off the free, untethered template end at +295, together with their nascent RNA products.

We supplied unlabeled RNAP, NTPs and the fluorescent, self-quenched F1 probe to the imaging flow-cell, and monitored the interactions of the F1 probe with the DNA loci by TIRF/CoSMoS (Friedman *et al.*, 2006; Revyakin *et al.*, 2012). We reasoned that the F1 target sequence in the nascent RNA would become available for hybridization once the active site of elongating RNAP (RD_e) reaches position +60 (taking into account the ~14 bases of nascent RNA protected by RNAP [Huang and Sousa, 2000]). The RD_e complex would then remain on the DNA until it runs off at +295. At the expected average RNAP elongation rate of 240 nt s^{-1} , the F1 probe would have a brief ~1 s window of opportunity to 'catch' the nascent transcript at the DNA molecule of origin. Probe hybridization to the RNA product would then be observed as a fluorescent spot co-localizing with the DNA locus.

Figure 2B shows the co-localization analysis of probe-DNA interactions in a typical single-molecule fastFISH experiment. This analysis counts all DNA molecules that co-localized with a probe 'spot' for more than five cumulative frames (at 0.4 s/frame) during the incubation (~15 min), and plots the $\Delta x, \Delta y$ displacements between the probe and DNA molecules as a two-dimensional histogram (Revyakin *et al.*, 2012). Specific probe-DNA interactions were observed, as indicated by a prominent peak at position (0, 0) of the co-localization histogram. In a typical experiment, 10–30% of DNA templates in a

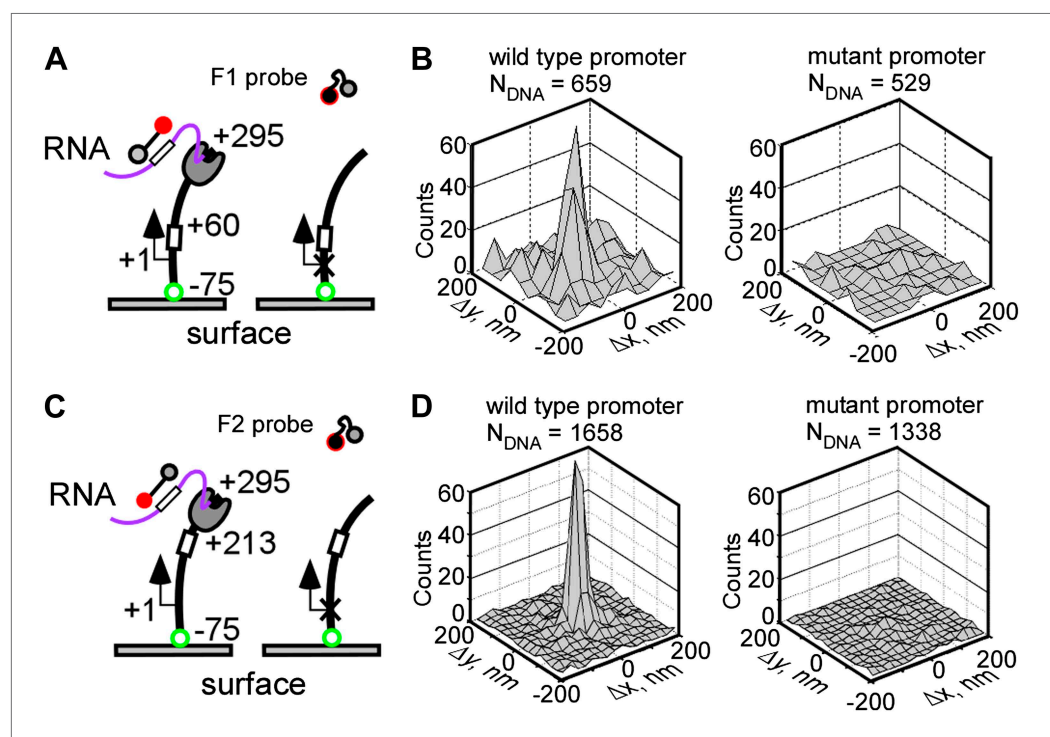


Figure 2. Real-time single-molecule detection of transcription by T7 RNAP using fastFISH. (A) Schematic of experiment. DNA templates containing a single consensus promoter for T7 RNAP (+1), or a single null mutant promoter (cross), were immobilized on a surface, with the promoter directing transcription towards the free end (+295). The templates contained the F1 fastFISH target sequence downstream from the promoter (from +28 to +46) which was expected to become available for hybridization in the nascent RNA after the RNAP active site reaches position +60. (B) Co-localization analysis of F1 probe-DNA interactions in a representative experiment. Left: wild type promoter. Right: mutant promoter. (C) and (D) Same as panels (A) and (B), but for the F2 target and probe. The templates contained the F2 fastFISH target sequence (from +181 to +199) which was expected to become available for hybridization in the nascent RNA after the RNAP active site reaches position +213.

DOI: 10.7554/eLife.01775.007

field of view co-localized with a probe. This apparent incomplete template utilization was not due to inefficient probe hybridization (see **Figure 4** and 'Discussion'), but, instead, was likely limited by the accessibility of the DNA template due to surface interference. No probe-DNA interactions were observed in a control experiment with a DNA template containing a null mutant promoter (**Raskin et al., 1993**), indicating that the co-localization was due to promoter-specific transcription. We also performed real-time single-molecule transcription experiments using the other unstructured probe, F2, and obtained nearly identical DNA-probe co-localization results (**Figure 2C–D**). Taken together, these findings suggest that fastFISH can detect production of nascent transcripts in real-time.

Single-molecule dynamics of T7 RNAP-DNA interactions

To measure the efficiency of real-time detection of nascent transcripts by fastFISH, we needed a reference to define the start and end of each productive single-molecule transcription event. We reasoned that monitoring the interactions between RNAP molecules and the DNA templates during promoter binding and run-off can serve this purpose. Therefore, we fluorescently labeled RNAP with Cy5 using HaloTag (**Figure 3—figure supplement 1**), supplied Cy5-RNAP and NTPs to an imaging flow cell containing immobilized DNA templates, and monitored Cy5-RNAP interactions with the DNA molecules (**Figure 3**). We observed transient RNAP binding events that lasted between 0.08 and 8 s, and were promoter-specific (**Figure 3B–D**). The events were comprised of two distinct populations: short-lived events whose dwell times fit well to a single exponential distribution (mean dwell time $T_0 \sim 0.14$ s); and a long-lived, bell-shaped population skewed towards long-lived events (peak dwell time $T_1 \sim 1.7$ s). The durations of observed interactions were not limited by the lifetime of the Cy5 label before photobleaching (**Figure 3—figure supplement 2**). We interpret the short-lived, stochastic interactions to be promoter-specific but non-productive RP_c and RP_o , because no interactions were observed with templates containing a null promoter sequence (at frame rates of 12.5 Hz, **Figure 3B**) and similar short-lived interactions were observed in the absence of NTPs (single exponential $T_0 \sim 0.3$ s, **Figure 3D**).

We interpret the long-lived, bell-shaped population of events to be full, productive transcription cycles, because (i) the long-lived events were not observed in the absence of NTPs (**Figure 3D**); (ii) the peak dwell time of the long-lived events, T_1 , increased nearly linearly with the length of the transcribed DNA segment ($T_1 \sim 1.7$ s, ~ 2.7 s, and ~ 3.7 s for DNA segments with $l = 295$ bp, 633 bp, and 910 bp, respectively, **Figure 3E–G**); and (iii) the RNAP fluorescence signal, on average, decayed towards the end of long-lived events (**Figure 3C,H**), consistent with RNAP initiating transcription at the promoter located close to the surface (75 bp, or ~ 25 nm), and then elongating towards the free, untethered DNA end, down the gradient of the TIRF evanescent field.

We conclude that the dwell time distributions of RNAP-DNA interactions define productive transcription events which can be used as a reference to determine the efficiency of nascent RNA detection by fastFISH, and to dissect the full transcription cycle by RNAP. We note that the slope of the plot of long-lived RNAP dwell times vs DNA length, $(\sim 3.2 \pm 0.6) \times 10^{-3} \text{ s nt}^{-1}$ (**Figure 3G**), gives an estimate of RNAP elongation rate of $\sim 300 \text{ nt s}^{-1}$. The intercept of the plot with the time axis ($l = 0$) gives an estimate of the net time that RNAP spends on the DNA template without elongating, $T_{stationary} = 0.7 \pm 0.3$ s, which includes the net duration of promoter opening and abortive cycling (RP_c , RP_o and RP_{its}), and, possibly, the time RNAP idles at the free DNA end before run-off.

Single-molecule dissection of the T7 RNAP transcription cycle with fastFISH

To determine the efficiency of nascent RNA detection by fastFISH, and to demonstrate the use of fastFISH in dissecting the kinetics of the full transcription cycle, we simultaneously monitored RNAP-DNA interactions and the production of transcripts in real-time, using Cy5-labeled RNAP together with the Cy3-labeled probe F1 (**Figure 4A**). In this two-color experiment, we expected to observe the following sequence of events: (i) initial appearance of an RNAP spot at the DNA locus, corresponding to promoter binding (RP_c); (ii) continued occupancy of the DNA by RNAP, corresponding to promoter opening (RP_o), abortive cycling (RP_{its}), elongation to +60 (RD_e) and the time required for probe hybridization; (iii) appearance of an F1 probe spot; (iv) co-occupancy of the DNA by RNAP and probe corresponding to elongation (RD_e); and (v) simultaneous disappearance of the RNAP and probe spots corresponding to run-off at +295. As shown in a representative video montage, precisely this sequence of events was observed (**Figure 4B,C**). Typically, the probe arrived at the DNA within 1 s after arrival of RNAP.

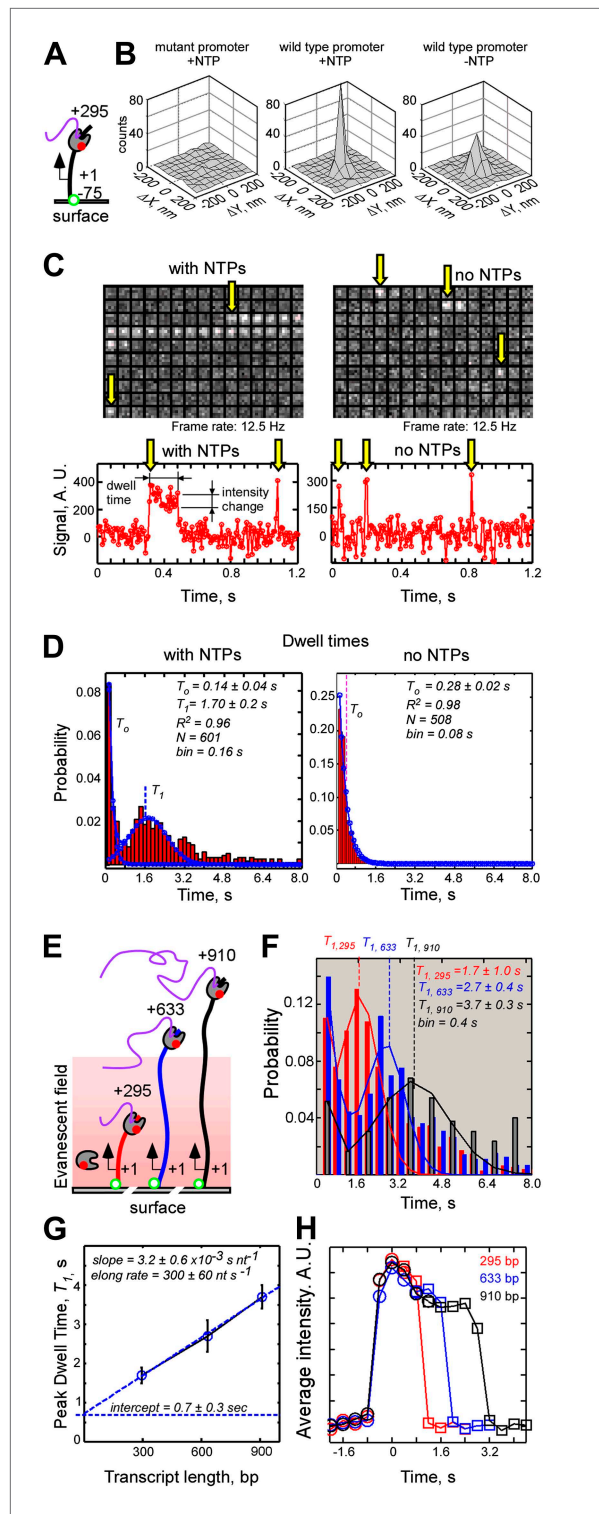


Figure 3. Single-molecule dynamics of T7 RNAP-DNA interactions. (A) Schematic of experiment. (B) Co-localization analysis of RNAP-DNA interactions. Left: null promoter DNA template in the presence of NTPs. Center: wild type promoter DNA template in the presence of NTPs. Right: wild type promoter DNA template in the absence of NTPs. (C) Representative data. Top: video montages of RNAP
 Figure 3. Continued on next page

Three lines of evidence suggested that the temporal co-occurrence of the F1 probe and RNAP corresponded to specific capture of nascent RNA in RD_e. First, no binding of F1 after RNAP binding was observed in the absence of NTPs (data not shown), or in the presence of only ATP and GTP, which limited initial transcription to positions from +1 to +6 (Figure 4—figure supplement 1), indicating that the probe bound specifically to the nascent RNA, and not to RNAP or to the DNA template. Second, no F1 probe binding was observed in the absence of the target sequence in the transcribed segment of the template, indicating that hybridization was specific for the F1 target (Figure 4—figure supplement 2). Third, and most importantly, when we moved the F1 target sequence from position +28 to position +181, we found that binding of the probe was delayed by an additional ~0.5 s (Figure 4—figure supplement 3), consistent with the extra time RD_e needs to transcribe the extra 153 bp DNA segment at 300 nt s⁻¹ (Figure 3G).

To estimate the efficiency of nascent RNA detection by fastFISH, we post-synchronized (Blanchard et al., 2004) all RNAP-DNA and probe-DNA interactions by RNAP run-off at the end of productive transcription events (defined as RNAP interactions that lasted ≥0.8 s), and super-imposed them to generate heat maps of all RNAP-DNA and probe-DNA interactions (Figure 4E). We found that ~81% of RD_es contained an RNA probe immediately before run-off (Figure 4F). The residual ~19% of undetected RD_es were mostly likely due to the presence of dark fluorophores in the chemically synthesized probes (Revyakin et al., 2012), and the inherent stochasticity of probe hybridization. The heat map analysis of all RNAP-DNA and probe-DNA interactions also showed that, in the majority of cases (~72%), the probe and RNAP dissociated from the DNA simultaneously upon run-off (ΔT_{off} ~0). In the remaining ~28% of cases, the RNA probe persisted on the DNA for ≥0.4 s after RNAP run-off (Figure 4F); the nature of these events remains to be determined.

To extract the kinetic information on early stages of transcription, we post-synchronized all RNAP-DNA and probe-DNA interactions by RNAP binding during

Figure 3. Continued

interactions with template containing consensus promoter for a 5×5 pixel region of interest centered at a single, photobleached DNA molecule (1 pixel = 200 nm, imaged at 12.5 Hz). Bottom: fluorescence time traces corresponding to the montages shown on top. Baseline of zero intensity indicates no binding. Left: experiment carried out in the presence of NTPs. Right: experiment carried out in the absence of NTPs. Yellow arrows indicate the first frames of RNAP binding events. **(D)** Dwell time probability distributions of RNAP-DNA binding events. Left: experiment carried out in the presence of NTP. Fitting to a sum of single exponential and Gaussian functions is shown in blue. Right: experiment carried out in the absence of NTPs. Fitting to a single exponential function is shown in blue. **(E)** Dependence of the peak dwell time of RNAP-DNA interactions on the length of the transcribed DNA segment: schematic of experiment. DNA templates containing transcribed segments spanning from +1 to +295 (red), +633 (blue), or +910 (black) were separately immobilized, and interactions of labeled RNAP were recorded at 2.5 Hz. **(F)** Dwell time probability distributions of RNAP-DNA interactions for the three DNA templates shown in **(E)**. $N = 749$ for the +1...+295 template (red), $N = 716$ for the +1 ... +633 template (blue), and $N = 213$ for the +1 ... 910 DNA template (black). The peak dwell times were calculated by fitting the distributions to a sum of single exponential and Gaussian functions. **(G)** Plot of the peak dwell time of RNAP-DNA interactions vs the length of the transcribed DNA segment. **(H)** Decay of intensity of RNAP fluorescence signal during productive RNAP-DNA interactions as an indicator of elongation by RNAP. All RNAP-DNA interactions having dwell times longer than 0.8 s (experiments in **F**) were post-synchronized by RNAP binding ($t = 0$, circles) and by RNAP run-off (squares), and weight-averaged plots of RNAP binding and run-off were plotted for DNA templates having transcribed segments of different lengths (red – 295 bp; blue – 633 bp; black – 910 bp). Time offsets between RNAP binding and run-off were set at peak dwell lifetimes, T_i , for the respective DNA templates measured in **(F)**.

DOI: [10.7554/eLife.01775.008](https://doi.org/10.7554/eLife.01775.008)

The following figure supplements are available for figure 3:

Figure supplement 1. Labeling and biochemical characterization of T7 RNAP activity.
DOI: [10.7554/eLife.01775.009](https://doi.org/10.7554/eLife.01775.009)

Figure supplement 2. Measurements of photobleaching times of fluorophores used in this study.
DOI: [10.7554/eLife.01775.010](https://doi.org/10.7554/eLife.01775.010)

productive transcription (RNAP-DNA interactions that lasted ≥ 0.8 s) (**Figure 4D**). Analysis of the resulting heat-map of probe and RNAP binding gave the average time delay between the RNAP and probe arrival, ΔT_{on} of 0.7 ± 0.2 s. We interpret ΔT_{on} as the net duration of promoter opening (T_{open}), abortive cycling ($T_{abortive}$), elongation from position +13 to position +60 ($T_{elongation}$), and the time for probe hybridization ($T_{hybridization}$). Because promoter opening by RNAP is generally very fast (~ 30 s $^{-1}$, **Stano et al., 2002**), T_{open} is negligible. Therefore:

$$T_{abortive} = \Delta T_{on} - (T_{elongation} + T_{hybridization})$$

Based on our measurement of the RNAP peak elongation rate (300 ± 60 nt s $^{-1}$, **Figure 3**), $T_{elongation} = (60-13)/300 = \sim 0.16$ s. Based on our measurement of the F1 probe hybridization rate (6×10^6 M $^{-1}$ s $^{-1}$, **Figure 1C**), $T_{hybridization} = \sim 0.33$ s (at the 500 nM probe concentration used in the experiment). Thus, $T_{abortive} = 0.7 - (0.16 + 0.33) = \sim 0.2$ s. This estimate of $T_{abortive}$ is within the upper limit set by our measurement of $T_{stationary} = 0.7 \pm 0.3$ s (the net time RNAP spends on the DNA template without elongating, **Figure 3G**). The fact that $T_{stationary}$ is larger than $T_{abortive}$, suggests that RNAP might dwell at the end of the DNA template for ~ 0.5 s before run-off. Pausing/arresting at free DNA ends has been previously observed with multi-subunit bacterial and eukaryotic RNA polymerases in ensemble assays, and likely involved backtracking by these RNAPs (lifetimes of >1 min, **Arndt and Chamberlin, 1988; Izban et al., 1995**). It remains to be tested whether T7 RNAP dwells at the end of the DNA via a similar mechanism, and whether the pause is part of a natural termination process.

Discussion

Here we present a strategy to design fast probe-target hybridization pairs for quantitative single-molecule detection of nascent RNA transcripts at sub-second time resolution under physiological conditions (fastFISH). The hybridization rates of fastFISH probes were near $\sim 10^7$ M $^{-1}$ s $^{-1}$ which exceeded annealing rates of probe-target sequences reported in literature by up to three orders of magnitude. We followed three simple rules in fastFISH probe-target design. First, the targets and the probes should be unstructured, which we achieved using sequences composed of only three bases—A, U, and C for RNA targets, and A, T, and G for DNA probes. We note that the three-base rule of fast hybridization also applies to DNA-DNA pairs (data not shown). Second, a complexity filter should be applied to ensure specificity and 1-to-1 stoichiometry of hybridization. This is essential for single-molecule counting, and cannot be achieved by using repetitive sequences such as poly (A/T). Third, fastFISH probes should be self-quenched. This reduces the fluorescence background from the free probe, and enables the use of near-micromolar probe concentrations

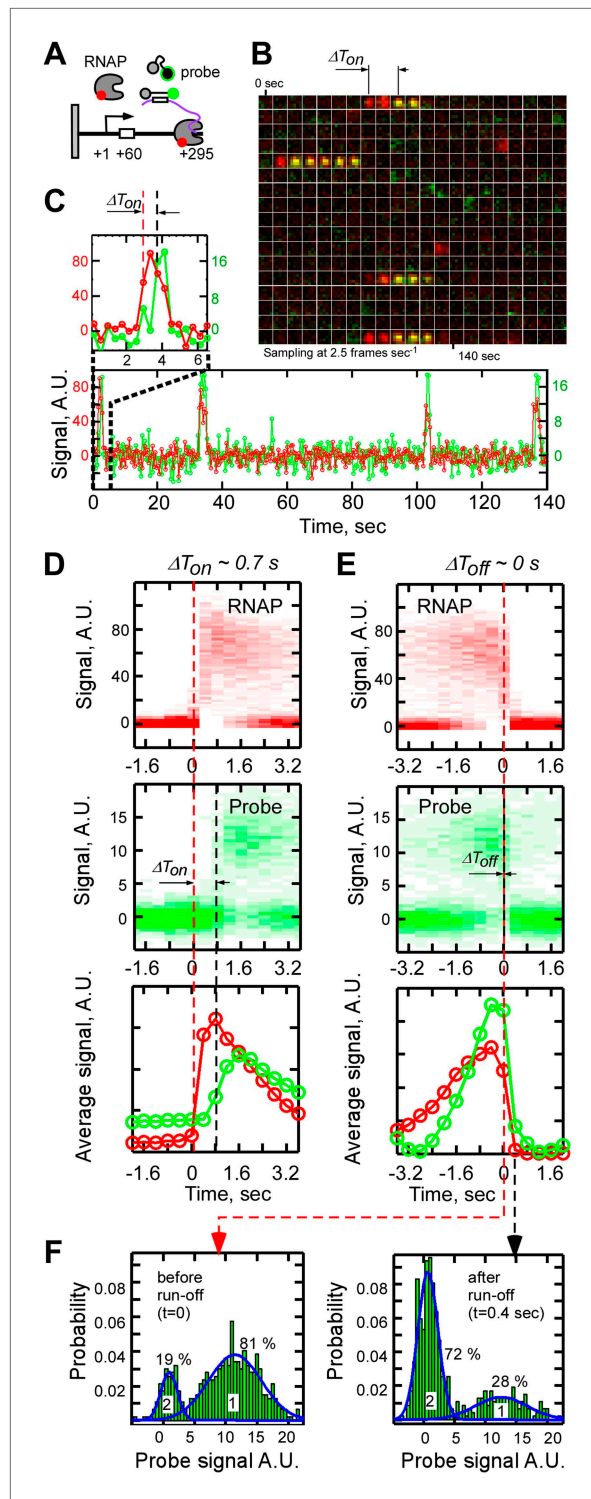


Figure 4. Single-molecule dissection of the T7 RNAP transcription cycle using fastFISH. **(A)** Schematic of experiment. **(B)** Merged video montages (5×5 pixel region of interest) centered at a single, previously photobleached DNA molecule. Interactions of RNAP (false colored in red) and probe (false colored in green) with the DNA locus were recorded. **(C)** Signal (A.U.) vs. Time (sec) plot showing the interaction of RNAP (red) and probe (green) with the DNA locus. **(D)** Heatmaps and line graphs showing the average signal (A.U.) vs. Time (sec) for RNAP (red) and Probe (green) interactions. **(E)** Heatmaps and line graphs showing the average signal (A.U.) vs. Time (sec) for RNAP (red) and Probe (green) interactions. **(F)** Histograms showing the probability of probe signal (A.U.) before and after run-off. *Figure 4. Continued on next page*

for real-time single-molecule imaging. Self-quenching of fastFISH probes likely occurs through random polymer motion and/or contact quenching (Johansson et al., 2002; Marras et al., 2002), and does not rely on hairpin formation (which is known to compromise hybridization rates [Tsourkas et al., 2003]). The choice of probe length of ~ 20 nucleotides was based on the end-to-end distance of a 20-mer nucleic acid duplex, expected to separate the fluorophore and the quencher beyond typical Forster distance (4–6 nm) upon hybridization. Shorter probes in combination with a short-range quencher may be used (Zhu et al., 2005) if the potential faster off-rate of the probe is not a concern. We note that probe hybridization rates could be further increased using peptide nucleic acid (PNA) probes, because the neutral charge of the PNA backbone is expected to reduce the electrostatic repulsion between the probe and the target (Kuhn et al., 2002).

To use fastFISH, the three-base target sequence must be inserted within the context of an existing four-base (i.e., ‘structured’) RNA sequence, which, in principle, might sequester the fastFISH target from hybridization. However, in our experience, the presence of adjacent four-base sequences usually does not interfere with co-transcriptional RNA detection, likely because four-letter sequences become kinetically trapped within their own secondary structures. Nevertheless, the use of a secondary structure prediction tool is recommended to ensure that flanking sequences do not interfere with the fastFISH target.

As a proof of concept, we used fastFISH to dissect the full transcription cycle of the prototypical single-subunit T7 RNAP, and obtained kinetic estimates of promoter search (upper limit of non-specific RNAP-DNA interactions: $T_{search} \sim 80$ ms), promoter binding, promoter escape, elongation and termination by RNAP (summarized in Figure 5). We recorded several differences between our measurements and previous studies. First, the measured rate of stochastic RP_c dissociation ($T_D \sim 0.3$ s, $k_{off} \sim 3$ s $^{-1}$) was higher than what had been reported in previous ensemble and single-molecule studies (k_{off} from 0.4 s $^{-1}$ to 1 s $^{-1}$ [Tang and Patel, 2006; Tang et al., 2009]). Second, the measured rate of promoter escape ($T_{abortive} \sim 0.2$ s, $k_{escape} \sim 5$ s $^{-1}$) was higher than in previous

Figure 4. Continued

imaged simultaneously at 2.5 Hz. Four representative transcriptional events are shown. (C) Fluorescence time traces corresponding to the montage shown in Panel B (RNAP and probe are shown in red and green, respectively). The first transcriptional event is zoomed in and shown on top. (D) Calculation of the time delay of probe binding with respect to RNAP binding (ΔT_{on}) from heat maps of all RNAP-DNA (top) and all probe-DNA binding events (middle) post-synchronized by RNAP binding (T_{on} , $N_{DNA} = 112$, $N_{events} = 469$). Baseline of zero intensity indicates no RNAP/probe binding. Time point $t = 0$ corresponds to the frame immediately before RNAP binding (red dashed line). On the bottom are the weight-averaged, normalized signal intensities of RNAP and probe binding calculated based on the heat-maps of all RNAP-DNA (red) and probe-DNA (green) binding events. (E) Calculation of the time delay of probe dissociation with respect to RNAP run-off (ΔT_{off}) from heat maps of all RNAP-DNA and probe-DNA dissociation events post-synchronized by RNAP dissociation (T_{off}). Time point $t = 0$ corresponds to the frame immediately before RNAP dissociation (red dashed line). (F) Calculation of the efficiency of real-time RNA detection (left), and of the fraction of events in which RNA was released from DNA upon RNAP run-off (right). Productive RNAP-DNA interactions were post-synchronized by RNAP dissociation as shown in Panel E, and probability distribution of probe signal intensity was plotted for the time point immediately before RNAP run-off (left, $t = 0$, for the efficiency of RNA detection) and the time point immediately after RNAP run-off (right, $t = 0.4$ s, for the fraction of RNA released upon run-off). Fits of distributions to sums of two Gaussian functions are shown in blue ($R^2 = 0.92$). The higher mean-value Gaussian (peak 1) corresponds to the events with the probe signal present and the lower mean value (peak 2) corresponds to the events without the probe signal present.

DOI: [10.7554/eLife.01775.011](https://doi.org/10.7554/eLife.01775.011)

The following figure supplements are available for figure 4:

Figure supplement 1. Control single-molecule transcription experiments in the presence of an incomplete set of NTPs.

DOI: [10.7554/eLife.01775.012](https://doi.org/10.7554/eLife.01775.012)

Figure supplement 2. Sequence-specific control of real-time RNA detection by fastFISH.

DOI: [10.7554/eLife.01775.013](https://doi.org/10.7554/eLife.01775.013)

Figure supplement 3. Effect of the distance between the promoter and the probe target sequence on the time delay between probe and RNAP binding, ΔT_{on} .

DOI: [10.7554/eLife.01775.014](https://doi.org/10.7554/eLife.01775.014)

ensemble and single-molecule studies (k_{escape} from 0.25 s^{-1} to 1 s^{-1} , [Skinner et al., 2004; Tang et al., 2009]). Third, the measured peak elongation rate ($300 \pm 60 \text{ nt s}^{-1}$) was somewhat higher than in previous ensemble measurements ($\sim 240 \text{ nt s}^{-1}$, Bonner et al., 1994; Golomb and Chamberlin, 1974), and significantly higher than in previous single-molecule measurements (from 40 nt s^{-1} to 120 nt s^{-1} , Thomen et al., 2008; Skinner et al., 2004). The overall faster rates measured in our study are likely due to the use of optimal physiological temperatures (37°C). Indeed, the catalytic rate of the T7 RNAP measured in steady-state conditions is known to increase up to 10-fold in the 20°C – 37°C temperature range (Maslak and Martin, 1993). Another potential contributing factor to the faster observed rates could be that our method to dissect the RNAP transcription cycle does not require the use of force, or covalent modification of the DNA templates.

We note that the dwell time distribution of productive RNAP-DNA interactions was skewed towards longer-lasting events, likely due to spontaneous RNAP pausing. Our observation of pausing might explain previously reported premature dissociation of colliding elongating RNAPs, observed at high RNAP:template ratios ('bumping', Zhou and Martin, 2006).

FastFISH detection of nascent RNA, coupled to single-molecule detection of protein-DNA interactions will be particularly useful to study the dynamics of initiation, escape, elongation, and termination by multi-subunit RNAPs, and the coupling of the transcriptional steps to downstream RNA-processing events. The guidelines for designing fastFISH probe-target pairs can be used to efficiently detect endogenous RNAs in situ and in vivo, to silence gene expression, and to build novel nucleic acid nanostructures.

Materials and methods

Prediction of secondary structure, and calculation of complexity of random oligonucleotides comprised of different alphabets

Random 19-mer DNA and RNA sequences were generated using a home-written Matlab routine. Briefly, built-in matlab routine `randperm` was used to pick a random set of k integers from a pool of n integers, where $n = 3^{19}$ (for 3-base sequences) or 4^{19} (for 4-base sequences), and $k = 3^{12}$ (531441). A different seed for the Matlab random number generator was used in independent trials to ensure independent sampling. (Sets of 3-base 19-mers were used, because a 19-mer is representative of commonly used hybridization probes, and analysis of $k = 3^{12}$ 19-mer sequences could be performed

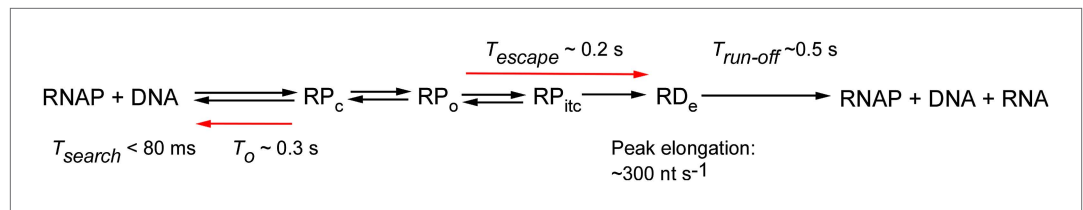


Figure 5. Summary of RNAP transcriptional kinetics measured in this work.
DOI: 10.7554/eLife.01775.015

on a modern portable computer.) Built-in Matlab routine `dec2base` was then used to convert sets of k integers into strings of numeric characters of base 3 (0, 1, and 2—for 3-base sequences) or base 4 (0, 1, 2, and 3—for 4-base sequences). The numeric characters were then replaced with A, U, C, and G, and sequences having GC-content ≥ 0.4 and ≤ 0.6 were saved as ASCII files, to be processed with the secondary structure prediction tool Mfold. Calculation of free energies of self-folding for DNA and RNA sequences was performed on a personal computer (routine `hybrid-ss-min`, release 3.8, available at <http://mfold.rna.albany.edu>, Zuker, 2003). For DNA, folding parameters –‘sodium = 0.15 –magnesium = 0.01’ were used to match the composition of hybridization buffer used in single-molecule experiments on probe on-rate measurements. For RNA, default folding parameters were used. The free energies of self-folding were extracted from Mfold output files using a home-written Matlab parsing routine, and plotted as normalized probability histograms with 1 kcal mol^{-1} binning. At least three independently randomized sets of 3^{12} sequences were analyzed for each alphabet (rAUC, rAUG, rCAG, rCUG, rAUGC, ATC, ATG, CAG, CTG, and ATGC) and results from independent sets were consistent to within 0.5%.

Sequence complexity was calculated using Matlab routine `kolmogorov` (provided by Stephen Faul, Radisens Diagnostics, code available at the Mathworks website), modified to output the absolute number of Lempel–Ziv operations, not normalized for the sequence length (Kaspar and Schuster, 1987). Per the modified `kolmogorov` routine, the complexity measure of a sequence comprised of 19 unique symbols (e.g., ‘abcdefghijklmnopqrs’) is 19, and the complexity measure of a sequence comprised of 19 identical symbols (e.g., A_{19}) is 2. To calculate the complexity of all tiling 19-mers of the human chromosome 22, the genomic sequence (length: 51,304,560 bp) and the annotation file indicating the locations of exons of known genes were downloaded from the UCSC genome browser, the sequences of exons in the forward strand were assembled into one contiguous string, and the complexities were calculated using a 19-base sliding window.

Sequences of fastFISH probes and their variants

The sequences of the F1 probes were as follows (synthesized by Integrated DNA technologies, IDT, Coralville, IA):

- F1: 5'-Cy3-GTT AAG ATA AGG GAT AGG G-3' (not self-quenched);
- F1: 5'-Cy3-GTT AAG ATA AGG GAT AGG G-FQ-3' (self-quenched with Iowa Black FQ)
- F1: 5'-Atto633-GTT AAG ATA AGG GAT AGG G-RQ-3' (self-quenched with Iowa Black RQ)

The sequences of self-quenched F2 probes and the F2+1C, F2+2C, F2+3C, and F2+4C variants (these variants were selected as representative of their categories based on the free energies of self-folding) were as follows (synthesized by Fidelity Systems, Gaithersburg, MD):

- F2: 5'-BHQ-3-GGT GTA TGT AAT TGG AGT GGT T-C6-Atto633-3';
- F2+1C: 5'-BHQ-3-GGT GTA TCT AAT TGG AGT GGT T-C6-Atto633-3';
- F2+2C: 5'-BHQ-3-GGT GTA TGC AAT TGG ACT GGT T-C6-Atto633-3';
- F2+3C: 5'-BHQ-3-GGT GTA TGT AAC CGG AGT GCT T-C6-Atto633-3';
- F2+4C: 5'-BHQ-3-GCT GTA TGC AAT TGG AGC CGT T-C6-Atto633-3'.

In the above sequences, BHQ-3 is Black Hole 3 quencher, C6 is a six-carbon linker.

The sequences of RNA targets used in single-molecule TIRF/CoSMoS measurements of probe hybridization rates were as follows:

- F1F2: 5'- CCC UAU CCC UUA UCU UAA CCA CUC CAA UUA CAU ACA CCC UUC AAA CUU
CAA ACU UCA AAG CAC AAG UUU UAU CCG GCC UUU AUU CAC AUU-bio-3'

F1F2+1G: 5'- CCC UAU CCC UUA UCU UAA *CCA CUC CAA UUA* **GAU** ACA CCC UUC AAA
 CUU CAA ACU UCA AAG CAC AAG UUU UAU CCG GCC UUU AUU CAC AUU-bio-3'
 F1F2+2G: 5'- CCC UAU CCC UUA UCU UAA *CCA* **GUC** CAA *UUG* CAU ACA CCC UUC AAA
 CUU CAA ACU UCA AAG CAC AAG UUU UAU CCG GCC UUU AUU CAC AUU-bio-3'
 F1F2+3G:5'- CCC UAU CCC UUA UCU UAA *GCA CUC* **CGG** UUA CAU ACA CCC UUC AAA
 CUU CAAACU UCA AAG CAC AAG UUU UAU CCG GCC UUU AUU CAC AUU-bio-3'
 F1F2+4G:5'- CCC UAU CCC UUA UCU UAA *CGG CUC* CAA *UUG* CAU ACA **GCC** UUC AAA
 CUU CAA ACU UCA AAG CAC AAG UUU UAU CCG GCC UUU AUU CAC AUU-bio-3'

In the above sequences, regions that correspond to the F1 and F2 targets are underlined and italicized, respectively, and the variations within the F2 target region to introduce 1-4 G residues are highlighted in bold. Note that the F1 and F2 targets overlap by three bases. The biotin residue on the 3' end enabled immobilization of the RNA targets on surface for single-molecule imaging by TIRF/CoSMoS. To fluorescently label the RNA targets, a Cy3-labeled DNA oligonucleotide (sequence: 5'-Cy3-AAT GTG AAT AAA GGC CGG ATA AAA CTT GTG C-3' IDT) was annealed to the 3' of the RNA targets.

Single-molecule imaging instrument

Single-molecule measurements of probe-target hybridization rates and T7 RNAP transcription were performed using a home-built multi-color total-internal-reflection microscope as described ([Revyakin et al., 2012](#)). In brief, fluorescence was excited and imaged through a 60 × 1.49 NA objective lens (Olympus, Center Valley, PA). The objective lens and the sample holder were pre-heated to 37°C (Bioprotech, Butler, PA). Fluorescence was excited with 532 nm (Coherent Verdi G2, intensity 100–300 W/cm², Coherent, Santa Clara, CA) and 640 nm (Coherent Cube 100, intensity 100 W/cm²) lasers, split into two imaging channels (band centered at 580 nm, width 60 nm; and band centered at 675 nm, width 45 nm), and imaged using two separate, synchronized EMCCD cameras (512 × 512 pixels, 1 pixel = 200 nm, conventional acquisition at 2.5 Hz or electron-multiplied acquisition at 12.5 Hz) (Andor, Belfast, United Kingdom). The position of the sample with respect to the objective lens was actively stabilized in (x,y,z). To that end, 2.8 μm-diameter magnetic beads (Life Technologies, Carlsbad, CA) decorating the imaging surface were illuminated with an infrared (IR) light-emitting diode (851 nm, Roithner LaserTechnik, Vienna, Austria), tracked with an IR camera using real-time image-processing ([Gosse and Croquette, 2002](#); [Revyakin et al., 2012](#)) at 30 Hz, and the position of the bead in the field of view was adjusted in real-time at 1–5 Hz using a 3-axes nanopositioning stage (Physik Instrumente, Karlsruhe, Germany). The IR camera and framegrabber for bead tracking were from 1stVision (Andover, MA).

Preparation of glass flow cells for single-molecule measurements of probe-target hybridization rates

Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Borosilicate coverslips were placed in ceramic racks (Thomas Scientific, Swedesboro, NJ), cleaned with piranha solution (a mixture of three parts concentrated H₂SO₄ and one part 30% H₂O₂, which is extremely corrosive and explosive) twice for 30 min, rinsed copiously with deionized water, and hydroxylated with 0.5 M KOH solution for 1 hr with sonication. The coverslips were then rinsed copiously with deionized water, rinsed twice by dipping in acetone (Chromasolv-Plus) for 10 s, and placed into a 3% solution of aminopropyltriethoxylane (APTES, Thermo Fischer Scientific, Waltham, MA) in acetone for 45 min, with gentle shaking. APTES-treated coverslips were then rinsed copiously with deionized water, and blow-dried with nitrogen. 50 μl drops of solution of 10% PEG-succinimidyl valerate (PEG-SVA, Mw = 5,000, Laysan) and 0.05% biotin-PEG-SVA (Mw = 5,000, Laysan) in 0.45M K₂SO₄ and 0.1M NaHCO₃ (pH 9.0) was squeezed between pairs of coverslips, and the coverslip pairs were incubated with PEG at 30°C for 30 min. PEG-treated coverslips were then rinsed copiously with deionized water, blow-dried with nitrogen, and the unreacted amine groups were end-capped by squeezing a 50 μl drop of solution of 2 mg sulfo-succinimidyl acetate (Pierce) in 0.1M NaHCO₃ (Pierce) for 10 min. The PEG-treated and amine-capped coverslips were rinsed copiously with deionized water, blow-dried with nitrogen, and stored dry at –80°C. (We found that end-capping of unreacted APTES amine groups was essential to eliminate non-specific interactions of fluorescently labeled probes with the glass surfaces at high probe concentrations.) An imaging flow cell containing seven reaction channels (volume of each channel: ~25 μl) and side injection ports was constructed using two PEG-treated coverslips and double-sided adhesive tape (3M VHB 4095). The flow cell was mounted on the microscope, and the inner surface was decorated with the 2.8 μm magnetic beads for active stage stabilization (~1–3 beads per 100 × 100 micron field of

view) and functionalized with streptavidin (by incubating 5 $\mu\text{g/ml}$ solution of streptavidin in phosphate-buffered saline for 30 s in a channel immediately before use).

Measurement of probe-target hybridization rates with TIRF/CoSMoS

Biotinylated, fluorescently labeled synthetic RNA targets were captured in the flow cell at a density $\sim 1000\text{--}2000$ molecules per field of view ($\sim 100 \times 100 \mu\text{m}$) by injecting 20 pM solution of RNA target in loading buffer (phosphate-buffered saline containing 0.08% Tween-20). The position of the microscope stage was then actively stabilized in (x,y,z) (see above), and a 'mapping' video of diffraction-limited spots indicating the positions of RNA target molecules was acquired at 2.5 Hz until all RNA spots were photobleached. Single-molecule imaging here and below was conducted in the presence of 0.9 mM Trolox, 2.2 mM protocatechuic acid (PCA), and 10 $\mu\text{g/ml}$ protocatechuate dehydrogenase (Toyobo), further purified to remove contaminating nuclease activity (Jonas Korch, Pacific Biosciences, personal communication). After mapping of the RNA targets, the loading buffer was replaced with hybridization buffer (loading buffer plus 10 mM MgCl_2) and a solution of fluorescently labeled, self-quenched probes in hybridization buffer at concentrations in the 2–500 nM range was injected into the flow cell at flow rate $\sim 50 \mu\text{l s}^{-1}$. A real-time video of probe binding events (observed as appearance of diffraction-limited fluorescent spots, **Figure 1—figure supplement 2**) was acquired at 2.5 Hz. For most accurate measurements of the hybridization rates, the concentrations of probes were adjusted empirically to achieve average probe arrival times of 10–100 s. Data analysis was essentially as described (Revyakin et al., 2012). Briefly, single-molecule spots were identified in every frame of the target and probe videos using the software Insight (Huang et al., 2010), and each spot was localized in (x,y) by fitting a 5×5 pixel spot intensity distribution to a 2D Gaussian. The (x,y) coordinates of all identified spots in all video frames were grouped into clusters ('blobs') using a 2D histogram (1 pixel binning), and the average (x,y) of every blob (indicating the locations of target and probe molecules) was calculated. Then, for every target blob, the closest probe blob was found, the deviations $\Delta x \Delta y$ were calculated for all such probe-target pairs, and plotted as a 2D histogram ('the co-localization plot'). RNA targets that were located within two standard deviations of the central peak at position (0,0) of the co-localization plot (~ 40 nm) were selected. Such 'active' RNA molecules typically comprised $\sim 30\%$ of all RNA molecules. Time series of probe-target binding was generated for every active RNA (using the mean fluorescence intensity from a 5×5 pixel region of interest (ROI), with the mean intensity of the ROI perimeter pixels subtracted as background). Each time series was examined, and the video frame corresponding to the probe arrival at the target was identified (**Figure 1—figure supplement 2**). The dataset containing all probe arrival times (T_{wait} , usually 100–500 events) was binned, with the center of the first bin corresponding to the minimal observed T_{wait} , the center of the last bin corresponding to the maximal observed T_{wait} and the number of bins $N_{\text{bins}} = \sqrt[3]{(2N_{\text{events}})}$ (Scott, 1992). The first bin was discarded to account for the discontinuity, and for the lag time due to the buffer exchange during probe injection. The probability histogram of T_{wait} (**Figure 1—figure supplement 2**) was fit to a single exponential function using the Matlab built-in routine `nlinfit`, which gave an estimate of the characteristic hybridization time (95% confidence interval), and the goodness of fit (typically, $R^2 > 0.98$). On-rates for probes F1, F2, F2+1C, F2+2C, F2+3C and F2+4C were calculated as $k_{\text{on}} = 1/(T_{\text{wait}} \times [\text{probe}])$.

Promoter DNA constructs for single-molecule transcription assays

Promoter DNA constructs were amplified by PCR. To attach and image the DNA constructs, the PCR primer complementary to the sequence upstream from the promoter was labeled with biotin and Cy3/Cy5/Atto633 at the 5'-end. In most experiments, the PCR-amplified DNA template spanned from -75 to $+295$ with respect to the transcription start site ($+1$). In the DNA template sequence shown below, the underlined sequences are, respectively: the upstream PCR primer (5' base at -75), the consensus T7 RNAP promoter, the F1 probe target, the F2 probe target, and the downstream PCR primer (5' base of the complementary sequence at $+295$). The $+1$ base is shown in bold.

(-75) TTATGTATCATACACATACGATTAGGTGACACTATAGAACTCGAGCAGCTGGATCC
 T A A T A C G A C T C A C T A T A **G G G** A G A C C A C A A C G G T T T C C C T C T A G A
CCCTATCCCTTATCTTAACGAATTGTGAGCGCTCACAATTCAAACCTTCAAACCTTCAA
 CTTCAAACCTTCAAACCTTCAAACCTTCAAACCTTCAAACCTTCAAACCTTCAA
 ACTTCAAACCTTCAAACCTTCAAACCTTGAATTCTTCAAAAACACTCCAATTACATACACC
 TTTCAAACCTTCAAACCTTCAAACCTTGAATTCTTCAAAAACACTCCAATTACATACACC
 TTTAAAGACCGTAAAGAAAATAAGCACAAGTTTTATCCGGC ($+295$)

For PCR-amplify DNA templates containing longer transcribed sequences (ending at +633 and +910), the same upstream primer, and primers with 5' ends corresponding to +633 and +910 were used. The sequence for the +910 DNA template is shown below, with the regions corresponding to the +633 and +910 primers underlined, and the sequence upstream from +296 identical to the sequence of the -75...+295 template:

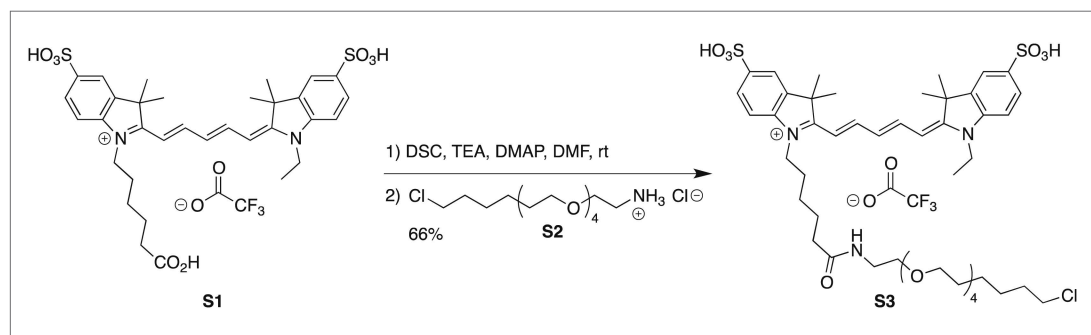
(+296) CTTTATTCACATTCTTGCCCGCTGATGAATGCTCATCCGGAATTCCGATGGCAA
 TGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTACCCCTTGTTACACCGTTTT
 CCATGAGCAAACCTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGC
 AGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTT
 CCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTC
 ACCAGTTTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTTT
 CACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTC
 AGGTTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAAT
 GAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTT
 AAGGCAGTTATTGTTGCCCTTAAACGCCTGGTGCTACGCCTGAATAAGTGATA
 ATAAGCGGATGAATGGCAGAAATTCGCCGGATCTTTGTGAAGGAACCTTA
 CTTCTGTGGTGTGACATA (+910)

In the DNA template containing the null T7 RNAP promoter mutation (experiments in **Figures 2 and 3**), the wild type promoter sequence was replaced with TAA TAA CC ACT CAC TAT AGG G (the mutation is underlined). In the DNA template containing deleted F1 target (experiment in **Figure 4—figure supplement 2A**), the sequence of the F1 target was replaced with CAA ACT TCA AAC TTC AAA C. In the DNA template containing deleted F2 probe target (experiment in **Figure 4—figure supplement 2B**), the sequence of the F2 target was replaced with CAA ACT TCA AAC TTC AAA C.

We note that, in the context of the full transcript, the 3-base unstructured RNA target sequences might be potentially trapped in secondary structures formed by 4-base sequences present in the same transcript. Therefore, we used Vienna RNA-folding software (**Gruber et al., 2008**) to calculate the secondary structure of the full transcripts to ensure that 4-base sequences do not hybridize to the 3-base targets, and, instead, tend to self-hybridize locally.

Synthesis of Cy5-PEG₄-HaloTag Ligand (S3)

Commercial reagents were obtained from reputable suppliers and used as received. All solvents were purchased in septum-sealed bottles stored under an inert atmosphere. Reactions were monitored by thin layer chromatography (TLC) on precoated TLC glass plates (silica gel 60 F₂₅₄, 250 μm thickness) or by LC/MS (4.6 mm × 150 mm 5 μm C18 column; 5 μl injection; 10–95% or 50–95% CH₃CN/H₂O, linear gradient, with constant 0.1% vol/vol TFA additive; 20 min run; 1 ml/min flow; ESI; positive ion mode; UV detection at 254 nm). High-resolution mass spectrometry was performed by the Mass Spectrometry Center in the Department of Medicinal Chemistry at the University of Washington. NMR spectra were recorded on a 400 MHz spectrometer. ¹H and ¹³C chemical shifts (δ) were referenced to TMS or residual solvent peaks, and ¹⁹F chemical shifts (δ) were referenced to CFCl₃. Data for ¹H NMR spectra are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublets, m = multiplet), coupling constant (Hz), integration.



Cy5-PEG₄-HaloTag Ligand (S3)

Cyanine dye **S1** (10 mg, 13.0 μmol, 1.33 eq; prepared as described in (**Mujumdar et al., 1993**)) was combined with *N,N'*-disuccinimidyl carbonate (3.3 mg, 13.0 μmol, 1.33 eq) and DMAP (0.16 mg,

1.30 μmol , 0.133 eq) in DMF (1 ml), and triethylamine (7.2 μl , 51.9 μmol , 5.33 eq) was added. The reaction was stirred at room temperature for 1 hr, at which point LC/MS analysis indicated complete conversion to the NHS ester. HaloTag ligand (O4)amine **S2** (3.4 mg, 9.76 μmol , 1 eq) in DMF (0.5 ml) was then added, and the reaction was stirred at room temperature for 18 hr while shielded from light. The crude reaction mixture was directly purified by reverse phase HPLC (10–95% MeCN/H₂O, with constant 0.1% vol/vol TFA additive) to provide **S3** (6.9 mg, 66%, TFA salt) as a blue-purple solid. ¹H NMR (MeOD, 400 MHz) δ 8.318 (t, J = 13.0 Hz, 1H), 8.313 (t, J = 13.0 Hz, 1H), 7.97–7.85 (m, 4H), 7.34 (dd, J = 8.1, 2.3 Hz, 2H), 6.68 (t, J = 12.5 Hz, 1H), 6.35 (d, J = 13.6 Hz, 1H), 6.34 (d, J = 13.6 Hz, 1H), 4.24–4.07 (m, 4H), 3.66–3.49 (m, 16H), 3.45 (t, J = 6.5 Hz, 2H), 3.36–3.33 (m, 2H), 2.22 (t, J = 7.3 Hz, 2H), 1.93–1.63 (m, 6H), 1.76 (s, 12H), 1.57 (p, J = 6.9 Hz, 2H), 1.52–1.23 (m, 9H); ¹⁹F NMR (MeOD, 376 MHz) δ –75.66 (s); Analytical HPLC: >99% purity (4.6 mm \times 150 mm 5 μm C18 column; 5 μl injection; 10–95% CH₃CN/H₂O, linear gradient, with constant 0.1% vol/vol TFA additive; 20 min run; 1 ml/min flow; ESI; positive ion mode; detection at 254/633 nm); HRMS (ESI) calculated for C₄₇H₆₉ClN₃O₁₁S₂ (M)⁺ 950.4062, found 950.4078.

Purification and labeling of T7 RNAP

T7 RNAP was purified as recombinant, N-terminal 6xHis and 6xHis-Halo fusions from the *E. coli* strain BL21 (DE3) using standard Ni-NTA chromatography per recommendations of the Ni-NTA resin manufacturer (Qiagen, Hilden, Germany). Purified proteins were dialyzed against storage buffer (contains 20 mM potassium phosphate pH 7.5, 100 mM NaCl, 50% glycerol, 10 mM DTT, 0.1 mM EDTA and 0.2% sodium azide) and stored at –20°C. To label Halo-RNAP, 20 μM of protein was mixed with 60 μM Cy5-PEG4-HaloLigand in reaction buffer (contains 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT, and 0.005% NP40), incubated at room temperature for 1 hr, and then at 4°C overnight. To remove unreacted HaloLigand, the reaction mixture was passed through a PD-10 column (GE Healthcare, Waukesha, WI) pre-equilibrated with the storage buffer. Protein labeling efficiencies were >90% based on UV-Vis absorption spectroscopy. Proteins were examined using SDS-polyacrylamide gel electrophoresis and analysis of Cy5 fluorescence with Typhoon Trio+, (GE Healthcare), followed by Coomassie Brilliant Blue staining.

Ensemble transcription by T7 RNAP

All transcription reaction mixture contained the following ingredients unless otherwise specified: 40 mM Tris–HCl pH 7.9, 6 mM MgCl₂, 2 mM Spermidine, 0.5 mM each nucleotide triphosphate (NTP), 4 ng/ μl yeast tRNA (Sigma, further purified with protease K treatment and phenol/chloroform extraction), 100 ng/ μl bovine serum albumin (New England Biolabs, Beverly, MA), 2 μM carrier DNA oligonucleotide (sequence: 5'-GTA TTG AGT CTT CAT TCT GTA T-3', IDT), 0.08% Tween 20 (EMD Millipore, Gibbstown, NJ), 0.9 mM Trolox, 2.2 mM PCA, and 0.3 U/ μl RNasin (Promega Madison, WI). In ensemble measurements, 10 nM DNA template (same PCR fragment as used in single-molecule assays), 0.17 $\mu\text{Ci}/\mu\text{l}$ α -³²P-ATP (Perkin Elmer, Waltham, MA) and different concentrations of RNAP were also included in the reaction. After incubation for 10 min at 37°C, 2 μl of the transcription reaction was loaded onto a 6% denaturing polyacrylamide gel. The gel was dried under vacuum at 80°C for 1 hr, exposed to a PhosphorImager screen and scanned (Typhoon Trio+, GE Healthcare).

Single-molecule transcription by T7 RNAP

The protocol for preparation of passivized glass surfaces for single-molecule transcription measurements will be described in detail elsewhere. Briefly, borosilicate coverslips (VWR) were cleaned with piranha solution as described above, blow-dried with nitrogen, and spin-coated to create a 50-nm layer of polystyrene containing 1% azide-terminated polystyrene on the glass surface. An imaging flow cell was assembled as described above, and the imaging surface was functionalized with biotin-PEG-alkyne as described (Presolski *et al.*, 2011). The flow cell was then mounted on the microscope holder, and the objective lens and the flow cell were pre-heated to 37°C. The imaging surface was blocked by incubation with loading buffer (phosphate-buffered saline containing 0.08% Tween-20) for 5 min, and functionalized by incubation with the loading buffer containing 5 $\mu\text{M}/\text{ml}$ streptavidin for 30 s. Biotinylated, fluorescently labeled DNA templates were captured and 'mapped' in the field of view of the microscope as described for the probe-target hybridization experiments. The position of the microscope stage was actively stabilized in (x, y, z), and 100 μl of transcription reaction mixture (see above) supplemented with 10 $\mu\text{g}/\mu\text{l}$ protocatechuate dehydrogenase, 5 nM Cy5-Halo-labeled or (His)₆ tagged, unlabeled RNAP, and/or 500 nM of fastFISH probe was injected into the flow cell. Videos of real-time interactions of RNAP and/or probe with the mapped, photobleached DNA loci were

recorded with synchronized EMCCD cameras at 2.5 Hz, unless otherwise specified. Single-molecule transcription reactions typically lasted for 15 min.

Analysis of RNAP and probe interactions with active DNA loci

Diffraction-limited RNAP spots were identified in every video frame, localized by 2D Gaussian fitting, and co-localized with the previously mapped DNA loci as described above for the probe-target hybridization experiments. Time series of RNAP and probe fluorescence intensities were generated for every 'active' DNA molecule using an ROI of 3×3 pixels, with 1-pixel perimeter background subtraction. A DNA locus was defined as 'active' if it co-localized with an RNAP spot in 5–9 cumulative frames (user-defined parameter) during the 15-min reaction. Videos frames in which RNAP bound to DNA (T_{on}) and dissociated from DNA (T_{off}) were identified in the time series manually, and the dwell times of RNAP-DNA interactions ($T_{off} - T_{on}$) were calculated. The probability distributions of RNAP dwell times were fit to a single exponential function, or to a sum of single exponential and Gaussian functions using Matlab routine `nlinfit`.

To measure the average time delay between binding of RNAP and the probe during productive RNAP-DNA interactions (ΔT_{on}), RNAP-DNA interactions that had dwell times ≥ 0.8 s were selected, post-synchronized by their T_{on} , and a two-dimensional histogram of all RNAP binding events (a heat map of signal intensity vs frame number [Figure 4, and Figure 4—figure supplements 1–3]) was generated (routine `histcn`, Bruno Luong, Fogale Nanotech, code available on the Mathworks website). The same dataset of RNAP T_{on} was then used as a reference to calculate a heat map of all probe binding events in the respective synchronized probe videos. For every frame of the heat map, the centers of mass of RNAP and probe signal intensity distributions were calculated, and normalized by setting the maximum intensity to 1, and weight-averaged profiles of all RNAP and probe binding events were generated. The rising edges of the averaged RNAP and probe profiles were found by fitting the maxima of the respective first derivatives to second-degree polynomial, and ΔT_{on} was calculated by subtracting the edge for probe from the edge for RNAP. The same approach was used to calculate ΔT_{off} (Figure 4E), except that video frames in which RNAP dissociated from the DNA (T_{off}) were used as references to post-synchronize RNAP- and probe-DNA dissociation events to generate respective heat maps, and to calculate weight-averaged profiles of all RNAP and probe dissociation events.

To calculate the efficiency of nascent RNA detection by the probe (Figure 4F), the probe dissociation heat map was used to calculate the distribution of the probe signal intensity at the last frame before RNAP dissociation. The distribution was then fit to a sum of two Gaussians using the Matlab routine `nlinfit`, and the ratio of the areas under the two Gaussians was calculated to give the percentage of productive RNAP-DNA interactions that contained the probe before run-off (i.e., the efficiency of RNA detection).

To calculate the fraction of RNAP run-off events in which the RNA remained at the DNA locus after RNAP dissociation, the probe dissociation heat map was used to calculate the distribution of the probe signal intensity in the frame immediately after RNAP dissociation (+0.4 s). The distribution was then fit to a sum of two Gaussians, and the ratio of the areas under the two Gaussians was calculated.

To assess the decay of RNAP fluorescence signal during elongation out of the TIRF evanescent field (Figure 3H), the weight-averaged profiles of all RNAP binding and dissociation events (obtained for templates having transcribed segments of different lengths—295, 633, and 910 bp) were plotted, and the edges of the RNAP dissociation profiles were manually offset in time from the edges of the RNAP binding profiles by the nearest-integer number of frames corresponding to the peak dwell time for the respective templates (Figure 3F).

Acknowledgements

We thank Yan Li for molecular cloning of the designed DNA templates and technical assistance; Brian English for suggestions on data analysis; Bo Huang, Stephen Faul, and Bruno Luong for software for data analysis; Sean Eddy, Don Rio, and William Dynan for critical reading of the manuscript; and Sarah Moorehead for administrative support. Robert Tjian is a Howard Hughes Medical Institute Investigator.

Additional information

Competing interests

RT: Robert Tjian is President of the Howard Hughes Medical Institute (2009–present), one of the three founding funders of *eLife*. The other authors declare that no competing interests exist.

Funding

Funder	Author
Howard Hughes Medical Institute	Luke D Lavis, Robert Tjian

The funder had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

ZZ, AR, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article, Contributed unpublished essential data or reagents; JBG, LDL, Analysis and interpretation of data, Drafting or revising the article, Contributed unpublished essential data or reagents; RT, Conception and design, Drafting or revising the article

References

- Arndt KM**, Chamberlin MJ. 1988. Transcription termination in *Escherichia coli*. Measurement of the rate of enzyme release from Rho-independent terminators. *Journal of Molecular Biology* **202**:271–285. doi: [10.1016/0022-2836\(88\)90457-3](https://doi.org/10.1016/0022-2836(88)90457-3).
- Bai L**, Santangelo T, Wang M. 2006. Single-molecule analysis of RNA polymerase transcription. *Annual Review of Biophysics and Biomolecular Structure* **35**:343–360. doi: [10.1146/annurev.biophys.35.010406.150153](https://doi.org/10.1146/annurev.biophys.35.010406.150153).
- Blanchard SC**, Kim HD, Gonzalez RL Jr, Puglisi JD, Chu S. 2004. tRNA dynamics on the ribosome during translation. *Proceedings of the National Academy of Sciences of the United States of America* **101**:12893–12898. doi: [10.1073/pnas.0403884101](https://doi.org/10.1073/pnas.0403884101).
- Bonner G**, Lafer EM, Sousa R. 1994. Characterization of a set of T7 RNA polymerase active site mutants. *Journal of Biological Chemistry* **269**:25120–25128.
- Chakraborty A**, Wang D, Ebright YW, Korlann Y, Kortkhorjia E, Kim T, Chowdhury S, Wigneshweraraj S, Irschik H, Jansen R, Nixon BT, Knight J, Weiss S, Ebright RH. 2012. Opening and closing of the bacterial RNA polymerase clamp. *Science* **337**:591–595. doi: [10.1126/science.1218716](https://doi.org/10.1126/science.1218716).
- Chan V**, Graves DJ, McKenzie SE. 1995. The biophysics of DNA hybridization with immobilized oligonucleotide probes. *Biophysical Journal* **69**:2243–2255. doi: [10.1016/S0006-3495\(95\)80095-0](https://doi.org/10.1016/S0006-3495(95)80095-0).
- Cheung ACM**, Cramer P. 2012. A movie of RNA polymerase ii transcription. *Cell* **149**:1431–1437. doi: [10.1016/j.cell.2012.06.006](https://doi.org/10.1016/j.cell.2012.06.006).
- DeHaseth PL**, Zupancic ML, Record MT. 1998. RNA polymerase-promoter interactions: the comings and goings of RNA Polymerase. *Journal of Bacteriology* **180**:3019–3025.
- Dessinges M-N**, Maier B, Zhang Y, Peliti M, Bensimon D, Croquette V. 2002. Stretching single stranded DNA, a model polyelectrolyte. *Physical Review Letters* **89**:248102. doi: [10.1103/PhysRevLett.89.248102](https://doi.org/10.1103/PhysRevLett.89.248102).
- Epshtein V**, Nudler E. 2003. Cooperation between RNA polymerase molecules in transcription elongation. *Science* **300**:801–805. doi: [10.1126/science.1083219](https://doi.org/10.1126/science.1083219).
- Friedman LJ**, Chung J, Gelles J. 2006. Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. *Biophysical Journal* **91**:1023–1031. doi: [10.1529/biophysj.106.084004](https://doi.org/10.1529/biophysj.106.084004).
- Friedman LJ**, Gelles J. 2012. Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. *Cell* **148**:679–689. doi: [10.1016/j.cell.2012.01.018](https://doi.org/10.1016/j.cell.2012.01.018).
- Friedman LJ**, Mumm JP, Gelles J. 2013. RNA polymerase approaches its promoter without long-range sliding along DNA. *Proceedings of the National Academy of Sciences of the United States of America* **110**:9740–9745. doi: [10.1073/pnas.1300221110](https://doi.org/10.1073/pnas.1300221110).
- Galburt EA**, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M, Bustamante C. 2007. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**:820–823. doi: [10.1038/nature05701](https://doi.org/10.1038/nature05701).
- Gao Y**, Wolf LK, Georgiadis RM. 2006. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Research* **34**:3370–3377. doi: [10.1093/nar/gkl422](https://doi.org/10.1093/nar/gkl422).
- Golomb M**, Chamberlin M. 1974. Characterization of T7-specific ribonucleic acid polymerase: IV. Resolution of the major in vitro transcripts by gel electrophoresis. *Journal of Biological Chemistry* **249**:2858–2863.
- Gosse C**, Croquette V. 2002. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophysical Journal* **82**:3314–3329. doi: [10.1016/S0006-3495\(02\)75672-5](https://doi.org/10.1016/S0006-3495(02)75672-5).
- Gruber AR**, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008. The Vienna RNA websuite. *Nucleic Acids Research* **36**:W70–W74. doi: [10.1093/nar/gkn188](https://doi.org/10.1093/nar/gkn188).
- Gupta PB**, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES. 2011. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**:633–644. doi: [10.1016/j.cell.2011.07.026](https://doi.org/10.1016/j.cell.2011.07.026).
- Hager GL**, McNally JG, Misteli T. 2009. Transcription dynamics. *Molecular Cell* **35**:741–753. doi: [10.1016/j.molcel.2009.09.005](https://doi.org/10.1016/j.molcel.2009.09.005).
- Huang B**, Bates M, Zhuang X. 2010. Super resolution fluorescence microscopy. *Annual Review of Biochemistry* **78**:993–1016. doi: [10.1146/annurev.biochem.77.061906.092014](https://doi.org/10.1146/annurev.biochem.77.061906.092014).

- Huang J, Sousa R. 2000. T7 RNA polymerase elongation complex structure and movement. *Journal of Molecular Biology* **303**:347–358. doi: [10.1006/jmbi.2000.4150](https://doi.org/10.1006/jmbi.2000.4150).
- Izban MG, Samkurashvili I, Luse DS. 1995. RNA polymerase II ternary complexes may become arrested after transcribing to within 10 bases of the end of linear templates. *Journal of Biological Chemistry* **270**:2290–2297. doi: [10.1074/jbc.270.5.2290](https://doi.org/10.1074/jbc.270.5.2290).
- Johansson MK, Fidder H, Dick D, Cook RM. 2002. Intramolecular dimers: a new strategy to fluorescence quenching in dual-labeled oligonucleotide probes. *Journal of the American Chemical Society* **124**:6950–6956. doi: [10.1021/ja025678o](https://doi.org/10.1021/ja025678o).
- Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebricht RH. 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**:1144–1147. doi: [10.1126/science.1131399](https://doi.org/10.1126/science.1131399).
- Kaspar F, Schuster H. 1987. Easily calculable measure for the complexity of spatiotemporal patterns. *Physical Review A* **36**:842–848. doi: [10.1103/PhysRevA.36.842](https://doi.org/10.1103/PhysRevA.36.842).
- Kuhn H, Demidov VV, Coull JM, Fiandaca MJ, Gildea BD, Frank-Kamenetskii MD. 2002. Hybridization of DNA and PNA molecular beacons to single-stranded and double-stranded DNA targets. *Journal of the American Chemical Society* **124**:1097–1103. doi: [10.1021/ja0041324](https://doi.org/10.1021/ja0041324).
- Kushon SA, Jordan JP, Seifert JL, Nielsen H, Nielsen PE, Armitage BA. 2001. Effect of secondary structure on the thermodynamics and kinetics of PNA hybridization to DNA hairpins. *Journal of the American Chemical Society* **123**:10805–10813. doi: [10.1021/ja016310e](https://doi.org/10.1021/ja016310e).
- Lima WF, Monia BP, Ecker DJ, Freier SM. 1992. Implication of RNA structure on antisense oligonucleotide hybridization kinetics. *Biochemistry* **31**:12055–12061. doi: [10.1021/bi00163a013](https://doi.org/10.1021/bi00163a013).
- Lionnet T, Singer RH. 2012. Transcription goes digital. *EMBO Reports* **13**:313–321. doi: [10.1038/embor.2012.31](https://doi.org/10.1038/embor.2012.31).
- Ma J, Bai L, Wang M. 2013. Transcription under torsion. *Science* **340**:1580–1583. doi: [10.1126/science.1235441](https://doi.org/10.1126/science.1235441).
- Marras SAE, Kramer FR, Tyagi S. 2002. Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Research* **30**:e122.
- Martin CT, Coleman JE. 1987. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry* **26**:2690–2696. doi: [10.1021/bi00384a006](https://doi.org/10.1021/bi00384a006).
- Maslak M, Martin CT. 1993. Kinetic analysis of T7 RNA polymerase transcription initiation from promoters containing single-stranded regions. *Biochemistry* **32**:4281–4285. doi: [10.1021/bi00067a017](https://doi.org/10.1021/bi00067a017).
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research* **15**:8783–8798. doi: [10.1093/nar/15.21.8783](https://doi.org/10.1093/nar/15.21.8783).
- Mujumdar RB, Ernst LA, Mujumdar SR, Lewis SJ, Waggoner AS. 1993. Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjugate Chemistry* **4**:105–111. doi: [10.1021/bc00020a001](https://doi.org/10.1021/bc00020a001).
- Murakami KS, Darst SA. 2003. Bacterial RNA polymerases: the whole story. *Current Opinion in Structural Biology* **13**:31–39. doi: [10.1016/S0959-440X\(02\)00005-2](https://doi.org/10.1016/S0959-440X(02)00005-2).
- Neuman KC, Abbondanzieri EA, Landick R, Gelles J, Block SM. 2003. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**:437–447. doi: [10.1016/S0092-8674\(03\)00845-6](https://doi.org/10.1016/S0092-8674(03)00845-6).
- Ogawa K. 1993. Differentiation and proliferation of hematopoietic stem cells. *Blood* **81**:2844–2853.
- Orlov YL, Potapov VN. 2004. Complexity: an internet resource for analysis of DNA sequence complexity. *Nucleic Acids Research* **32**:W628–W633. doi: [10.1093/nar/gkh466](https://doi.org/10.1093/nar/gkh466).
- Presolski SI, Hong VP, Finn MG. 2011. Copper-catalyzed azide-alkyne click chemistry for bioconjugation. *Current Protocols in Chemical Biology* **3**:153–162. doi: [10.1002/9780470559277.ch110148](https://doi.org/10.1002/9780470559277.ch110148).
- Raj A, Van Oudenaarden A. 2009. Single-molecule approaches to stochastic gene expression. *Annual Review of Biophysics* **38**:255–270. doi: [10.1146/annurev.biophys.37.032807.125928](https://doi.org/10.1146/annurev.biophys.37.032807.125928).
- Raskin CA, Diaz GA, McAllister WT. 1993. T7 RNA polymerase mutants with altered promoter specificities. *Proceedings of the National Academy of Sciences of the United States of America* **90**:3147–3151. doi: [10.1073/pnas.90.8.3147](https://doi.org/10.1073/pnas.90.8.3147).
- Revyakin A, Liu C, Ebricht RH, Strick TR. 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**:1139–1143. doi: [10.1126/science.1131398](https://doi.org/10.1126/science.1131398).
- Revyakin A, Zhang Z, Coleman RA, Li Y, Inouye C, Lucas JK, Park SR, Chu S, Tjian R. 2012. Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes & Development* **26**:1691–1702. doi: [10.1101/gad.194936.112](https://doi.org/10.1101/gad.194936.112).
- Robb NC, Cordes T, Hwang LC, Gryte K, Duchi D, Craggs TD, Santoso Y, Weiss S, Ebricht RH, Kapanidis AN. 2013. The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *Journal of Molecular Biology* **425**:875–885. doi: [10.1016/j.jmb.2012.12.015](https://doi.org/10.1016/j.jmb.2012.12.015).
- Rothmund PWK. 2006. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**:297–302. doi: [10.1038/nature04586](https://doi.org/10.1038/nature04586).
- Scott DW. 1992. *Multivariate density estimation: theory, practice, and visualization (Wiley series in probability and statistics)*. Hoboken, NJ: John Wiley and Sons.
- Shaevitz JW, Abbondanzieri EA, Landick R, Block SM. 2003. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426**:684–687. doi: [10.1038/nature02191](https://doi.org/10.1038/nature02191).
- Skinner GM, Baumann CG, Quinn DM, Molloy JE, Hoggett JG. 2004. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-molecule view of the transcription cycle. *Journal of Biological Chemistry* **279**:3239–3244. doi: [10.1074/jbc.M310471200](https://doi.org/10.1074/jbc.M310471200).
- Smith SB, Cui Y, Bustamante C. 1996. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**:795–799. doi: [10.1126/science.271.5250.795](https://doi.org/10.1126/science.271.5250.795).

- Stano NM**, Levin MK, Patel SS. 2002. The +2 NTP binding drives open complex formation in T7 RNA polymerase. *Journal of Biological Chemistry* **277**:37292–37300. doi: [10.1074/jbc.M201600200](https://doi.org/10.1074/jbc.M201600200).
- Tang G-Q**, Patel SS. 2006. Rapid binding of T7 RNA polymerase is followed by simultaneous bending and opening of the promoter DNA. *Biochemistry* **45**:4947–4956. doi: [10.1021/bi052292s](https://doi.org/10.1021/bi052292s).
- Tang G-Q**, Roy R, Bandwar RP, Ha T, Patel SS. 2009. Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **106**:22175–22180. doi: [10.1073/pnas.0906979106](https://doi.org/10.1073/pnas.0906979106).
- Thomen P**, Lopez PJ, Bockelmann U, Guillerez J, Dreyfus M, Heslot F. 2008. T7 RNA polymerase studied by force measurements varying cofactor concentration. *Biophysical Journal* **95**:2423–2433. doi: [10.1529/biophysj.107.125096](https://doi.org/10.1529/biophysj.107.125096).
- Tsourkas A**, Behlke MA, Rose SD, Bao G. 2003. Hybridization kinetics and thermodynamics of molecular beacons. *Nucleic Acids Research* **31**:1319–1330. doi: [10.1093/nar/gkg212](https://doi.org/10.1093/nar/gkg212).
- Wang F**, Redding S, Finkelstein IJ, Gorman J, Reichman DR, Greene EC. 2012. The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion. *Nature Structure Molecular Biology* **20**:174–181. doi: [10.1038/nsmb.2472](https://doi.org/10.1038/nsmb.2472).
- Wang J-Y**, Drlica K. 2004. Computational identification of antisense oligonucleotides that rapidly hybridize to RNA. *Oligonucleotides* **14**:167–175. doi: [10.1089/oli.2004.14.167](https://doi.org/10.1089/oli.2004.14.167).
- Wang MD**, Schnitzer MJ, Yin H, Landick R, Gelles J, Block SM. 1998. Force and velocity measured for single molecules of RNA polymerase. *Science* **282**:902–907. doi: [10.1126/science.282.5390.902](https://doi.org/10.1126/science.282.5390.902).
- Weiss S**. 1999. Fluorescence spectroscopy of single biomolecules. *Science* **283**:1676–1683. doi: [10.1126/science.283.5408.1676](https://doi.org/10.1126/science.283.5408.1676).
- Wright M**, Church GM. 2002. An open-source oligomicroarray standard for human and mouse. *Nature Biotechnology* **20**:1082–1083. doi: [10.1038/nbt1102-1082](https://doi.org/10.1038/nbt1102-1082).
- Yamanaka S**. 2009. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**:49–52. doi: [10.1038/nature08180](https://doi.org/10.1038/nature08180).
- Yilmaz LS**, Okten HE, Noguera DR. 2006. Making all parts of the 16S rRNA of *Escherichia coli* accessible in situ to single DNA oligonucleotides. *Applied and Environmental Microbiology* **72**:733–744. doi: [10.1128/AEM.72.1.733-744.2006](https://doi.org/10.1128/AEM.72.1.733-744.2006).
- Zhou Y**, Martin CT. 2006. Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision-induced “bumping”. *Journal of Biological Chemistry* **281**:24441–24448. doi: [10.1074/jbc.M604369200](https://doi.org/10.1074/jbc.M604369200).
- Zhu P**, Clamme J-P, Deniz AA. 2005. Fluorescence quenching by TEMPO: a sub-30 Å single-molecule ruler. *Biophysical Journal* **89**:L37–L39. doi: [10.1529/biophysj.105.071027](https://doi.org/10.1529/biophysj.105.071027).
- Ziv J**, Lempel A. 1976. On the complexity of finite sequences. *IEEE Transactions on Information Theory* **IT-22**:75–81.
- Zuker M**. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**:3406–3415. doi: [10.1093/nar/gkg595](https://doi.org/10.1093/nar/gkg595).