Methods 48 (2009) 323–332

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Single molecule transcription elongation

Eric A. Galburt [a,b,*], Stephan W. Grill [a,b], Carlos Bustamante [c,d,e]

[a] Max Planck Institute for the Physics of Complex Systems, Nöthnitzerstraße 38, 01187 Dresden, Germany
[b] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany
[c] University of California—Berkeley, Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, 642 Stanley Hall #3220, Berkeley, CA 94720, USA
[d] University of California—Berkeley, Department of Physics, Howard Hughes Medical Institute, 642 Stanley Hall #3220, Berkeley, CA 94720, USA
[e] University of California—Berkeley, Department of Chemistry, Howard Hughes Medical Institute, 642 Stanley Hall #3220, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Accepted 30 April 2009
Available online 6 May 2009

Keywords:
Single molecule
Optical trapping
RNA polymerase II
Transcriptional pausing
Backtracking
TFIIS
Molecular motor
Diffusion

ABSTRACT

Single molecule optical trapping assays have now been applied to a great number of macromolecular systems including DNA, RNA, cargo motors, restriction enzymes, DNA helicases, chromosome remodelers, DNA polymerases and both viral and bacterial RNA polymerases. The advantages of the technique are the ability to observe dynamic, unsynchronized molecular processes, to determine the distributions of experimental quantities and to apply force to the system while monitoring the response over time. Here, we describe the application of these powerful techniques to study the dynamics of transcription elongation by RNA polymerase II from *Saccharomyces cerevisiae*.

1. Introduction

The techniques for studying single molecule eukaryotic transcription are based on the innovation of dual trap optical tweezer assays [1–4], the work of those who have advanced the *in vitro* study of RNA polymerase II (RNAP II) elongation [5–8] and the development of similar assays for the study of *Escherichia coli* RNA polymerase [9–12]. The experiments are designed to record the positions of an individual polymerase along the DNA template as a function of time. There are three advantages of collecting data from a single enzyme in the context of an optical tweezer setup: (1) the ability to observe dynamic events that would be unsynchronized if studied in bulk, (2) the measurement of distributions of observables rather than just their mean and (3) the application of a perturbing force to the system.

It has been known for more than 25 years that cellular RNA polymerases do not transcribe their templates with uniform velocities [13]. Instead, transcription elongation is interrupted by a variety of pause states [14] that play roles in many aspects of transcription regulation including promoter proximal escape [15,16], termination [17,18], nascent RNA folding [19], polyadenylation [20] and splice site selection [21]. These pause states have been studied extensively by a variety of techniques including gel electrophoresis, stop flow spectroscopy and other methods that derive the average behavior from a population of molecules. The proper analysis of these methods requires that the population be synchronized in such a way that all molecules occupy the same nucleotide position at the beginning of the experiment. What is missed by these experimental techniques is that pauses occur all over the template in a sequence biased, but stochastic manner. While strong sequence dependent pauses are the most easily studied, their behavior only represents the long time tail of the distribution of all pause times and they account for only a small fraction of the total number of pauses. Furthermore, it is clear that the cellular regulation of eukaryotic transcription elongation is not coded directly in the template sequence, but instead depends on many factors that cooperate (or compete) to determine whether or not a gene is to be transcribed. For example, weak pauses may be strengthened by the presence of factors such as nucleosomes [22] and strong pauses may be bypassed through the binding of elongation factors such as TFIIS [23]. Therefore, understanding pause dependent transcription regulation mechanisms depends on the understanding of general pause mechanisms and their inter-relatedness and requires the observation of all pauses that might occur along a template. Single molecule transcription techniques make it possible to characterize the overall distribution of pause behavior to supplement what is known about specific pause mechanisms that operate within specific sequence contexts.

* Corresponding author. Present address: Washington University School of Medicine, Department of Biochemistry and Molecular Biophysics, 660 South Euclid Avenue, Saint Louis, MO 63110, USA. Fax: +1 314 362 7183.
E-mail address: egalburt@biochem.wustl.edu (E.A. Galburt).

1046-2023/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2009.04.021

Another advantage afforded by the optical tweezers assay is the ability to apply a perturbing external force and to observe the polymerase's response. Far from being a curious detail of interest only to those studying the detailed physics of molecular motors, the performance of work and the generation of force by the polymerase are crucial for both transcription processivity and regulation within the cell. Transcription takes place in a crowded nuclear environment with many physical barriers caused by DNA template bound proteins including histones, DNA repair enzymes, DNA polymerases, other RNA polymerases, topoisomerases, and other proteins involved in DNA metabolism or transcription regulation. Furthermore, due to intervening sequences, genes in eukaryotes can stretch up to millions of base pairs requiring RNAP II to make its way through a considerable amount of traffic [24,25]. Therefore, to successfully transcribe a single complete gene, the polymerase must not only bind strongly to the DNA template, but must possess robust mechanisms to perform physical work in the face of barriers and opposing forces.

In this review of methods for the single molecule analysis of RNAP II transcription elongation we review the physics of optical trapping, describe the specific experimental protocols used to acquire and analyze single molecule data, and conclude with a short summary of the important results that have been obtained using these methods.

2. Optical trapping methods

The development of single molecule optical tweezer methods was initiated by the observation that the interaction between light in a highly focused laser beam and micron sized dielectric particles results in the formation of a three-dimensional trap with restoring forces on the order of piconewtons [1]. Here we only review the basics as full descriptions of the principles and practice of optical tweezers have been published elsewhere [26,27]. A ray optics description of light may be used to form an intuition of the interactions between the light and the particle that lead to the formation of the trap [1] (Fig. 1a and b). This description is not physically accurate unless the size of the bead is much larger than the wavelength of the light, however it serves to gain insight into the interaction between the light and the bead. In most biological applications, optical traps are used exactly in the regime where the bead size is on the order of the wavelength where these physical descriptions break down. See [27] for a detailed review of theoretical approaches to this problem.

If the index of refraction of the bead is greater than the index of refraction of the buffer, the light will be bent in the direction of the bead displacement (from the center of the trap) as it passes through the bead. Light carries momentum and thus a change in its direction results in a change of momentum. Since momentum is a conserved quantity, the bead must experience a change in momentum equal and opposite to that experienced by the light. This effect produces a force that acts to restore the bead the center of the beam (Fig. 1a). To create a three-dimensional trap a high numerical aperture lens (i.e. N.A. 1.2) is used to generate high angle rays beams that impinge on the bead (Fig. 1b). By extending the analysis of the refraction of a single ray, we see that in this arrangement, no matter which way the bead is displaced the sum of the effects from the high angle rays will result in a force pulling the bead towards the center of the trap. In practice, trap strength depends on laser power, bead size and other environmental factors and must be calibrated by comparing the strength of the trap to the thermal energy. This procedure is described in more detail in the protocol section below as it is performed on each individual bead pair before each experiment. For distances less than the diameter of the particle, the trap acts simply as a Hookean spring ($F = kx$) where F is the force, k is the trap stiffness and x is the displacement of the bead from the center of the trap. Thus, after the measurement of trap stiffness, the force on the bead may be simply calculated from its position. During an experiment both the force and the position of the individual beads relative to the trap centers are monitored at rates up to 10 kHz and in dual trap setups, the final experimental estimate of the force is obtained through an average of the force on each bead. Typically these forces are very similar and these two independent measurements of the force increase accuracy. The position measurement of the polymerase in a passive mode experiment requires that the bead–trap distances be transformed into bead–bead distances and converted into DNA contour lengths. This is done in practice using the worm-like-chain theory of DNA elasticity that describes DNA as a nonlinear spring and relates its equilibrium end-to-end length to its contour length as a function of force [28,29].

To construct a double trap tweezer, one requires a minimal set of components well aligned on an optical table in a room that minimizes environmental noise and temperature fluctuations (better than OSHA NC30 and ±0.5 °C [27]). Double trap apparatuses have the advantage over single trap tweezers in that they isolate the system from the lab frame and thus possess less noise and drift. This improvement can be crucial for the measurement of long trajectories where enzyme velocities are not very high and on the order of 1–10 nt/s. Furthermore, since data may be acquired from each trap independently, correlations between the beads may be used to reduce noise even further. Lastly, by splitting the laser into two orthogonally polarized beams, fluctuations in the direction of beam propagation that stem from the laser source itself are experienced simultaneously in both traps and do not result in relative motion of the two trapped beads [30]. Fig. 1c shows the layout of the most important features of the setup and more details may be found elsewhere [26]. Briefly, a single 5 W, 1064 nm laser (Spectra-Physics, J20-BL10-106C) is expanded with a telescope (∼4×) and then split with a polarizing beam splitter (PBS) into two separately controllable beams that each creates a single trap. A motorized $\lambda/2$-wave-plate positioned before the PBS can be used to control (and equalize) the power of each beam and thus the stiffness of each trap. A mirror mounted on a two axis piezoelectric scanner (Mad City Labs, Nano-MTA series) steers one beam (red) while an acousto-optical device (IntraAction, DTD series) may be used to steer the other beam (orange). The method of steering chosen depends on the specific needs of the instrument and can also be realized with electro-optical devices [31] and moveable lenses. After the steering components, the two beams are recombined with a second PBS, passed through another expanding telescope (∼4×) and directed through the objective lens. To create stable traps in this configuration it is necessary to expand the beam sufficiently to overfill the back plane of the objective lens to generate the highest angle rays possible to generate the force along the beam axis as described above (Fig. 1b). The flow cell (Fig. 1d) is positioned between the matched objective and condenser lenses (Nikon Plan Apo VC water immersion objectives, N.A. 1.2) and is controlled through the computer with a three axis stage (Newport, 562 series), motion controller (Newport ESP300) and motorized actuator screws (Newport, CMA series). Lastly, the light from the traps is collected, separated with a third PBS and imaged onto position sensitive detectors (Pacific Silicon, DL100). As the light influences the trapped bead, the position of the bead influences the interference pattern produced between the scattered and unscattered light leaving the trap. Using a technique known as back focal plane interferometry, the movements of the bead may be measured [32,33]. A separate optical path using a light emitting diode (LED) light source and a CCD camera (Watec, 902H2 supreme) is coupled into the objectives and used to image the beads during the experiment and may also be used for automated and independent video tracking of bead positions (Fig. 1e).

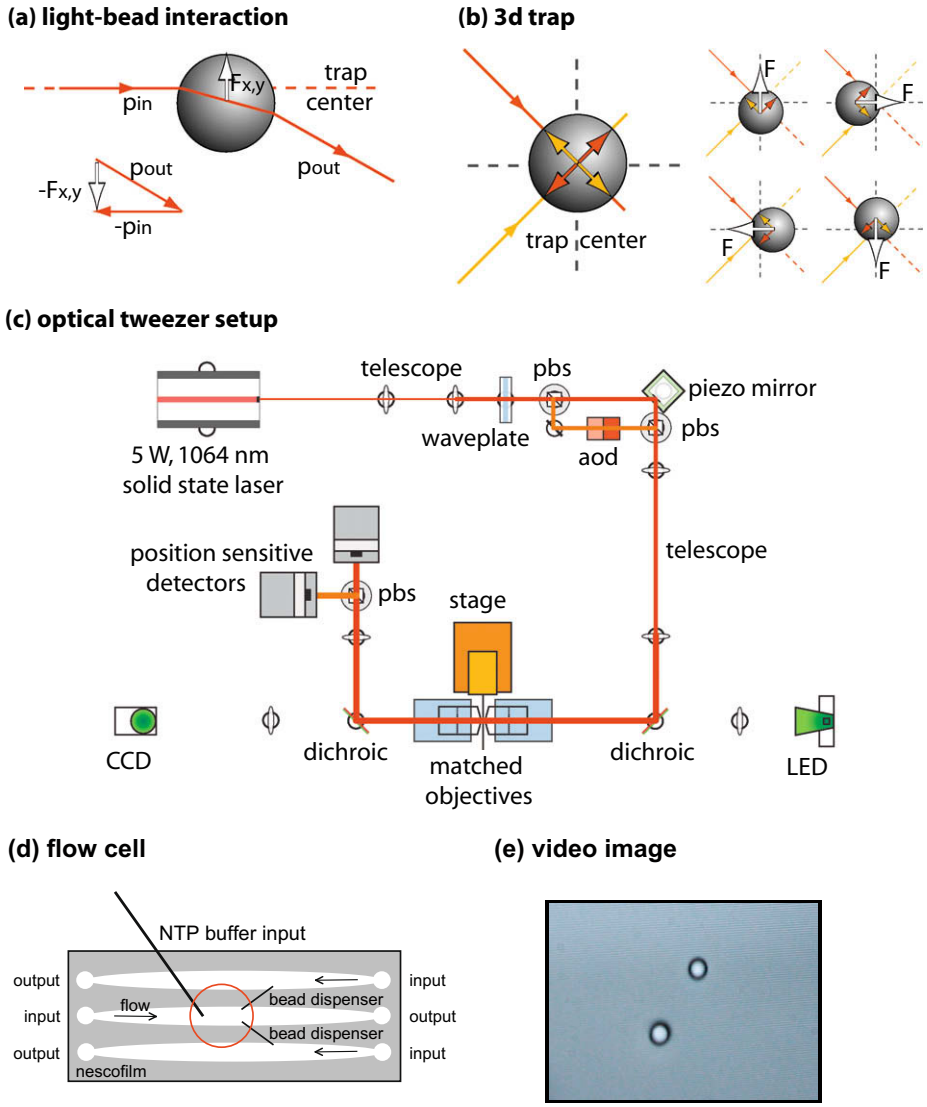

Fig. 1. Optical trapping. An optical trap can be understood by considering how a ray of light is refracted as it passes through the bead. (a) The schematic indicates rays of light (red) along with their momentum vectors (p). Since all forces are balanced with an equal an opposite one, changes in the momentum of the refracted rays are balanced by a force on the bead. If the bead is displaced from the center of the laser, the beam is deflected in the direction of the displacement. This leads to a force on the bead in the opposite direction or back towards the trap center. (b) By focusing the laser using a high numerical aperture lens, high angle rays are generated that produce forces that acts to restore the bead towards the trap center in three dimensions. (c) The basic layout of a dual trap optical tweezer setup. The path of the laser is indicated in red and passes from the source laser (upper left) through a telescope, $\lambda/2$ waveplate, polarizing beam splitter (PBS), steering elements (mirror, AOD), another pbs, another expander, objective (blue), flow cell on a stage (orange), condenser (blue) and finally through a final PBS and position sensitive detectors (gray). The orange parts of the path indicate where the orthogonal polarizations are separated by polarizing beam splitters to either steer or measure the traps independently. The imaging path uses an LED light source and a CCD camera to capture a video image. A schematic of the flow cell is shown (d) along with a sample video image (e).

3. Experimental methods

3.1. Biotin labeled RNA polymerase II

The RNA polymerases that we used for the study were purified from *Saccharomyces cerevisiae* and labeled with biotin on the carboxy-terminal tail of the Rbp3 subunit in the lab of Mikhail Kashlev at the NIH. They have unphosphorylated C-terminal domains and the full 12 subunits including Rpb4/7. The details of the purification have been exhaustively described elsewhere [5].

3.2. Digoxigenin labeled DNA template

In the beginning, we set out to understand the process of transcription elongation and the mechanism by which the polymerase converted the chemical energy stored in NTP molecules to perform the mechanical work necessary to step along the DNA track. Our strategy was to use a template DNA that was not derived from known gene sequences that potentially contain regulatory motifs that would interfere with the interpretation of our data. The design and production of the labeled template we have used in our experiments is described below, but any template may be used provided it is has a 5′ sticky end compatible with the single stranded overhang on the minimal elongation complex (see next section and Fig. 2), has an A-less (or N-less) stretch longer than the combined footprint of the polymerase and a selected restriction enzyme that ends at a unique restriction digestion site, is long enough to provide enough bead separation in the tweezers (~500 nm or 1.5 kb for 1 μm beads) and is labeled with digoxigenin or another moiety to facilitate attachment to a coated polystyrene bead (i.e. anti-digoxigenin antibody coated beads).

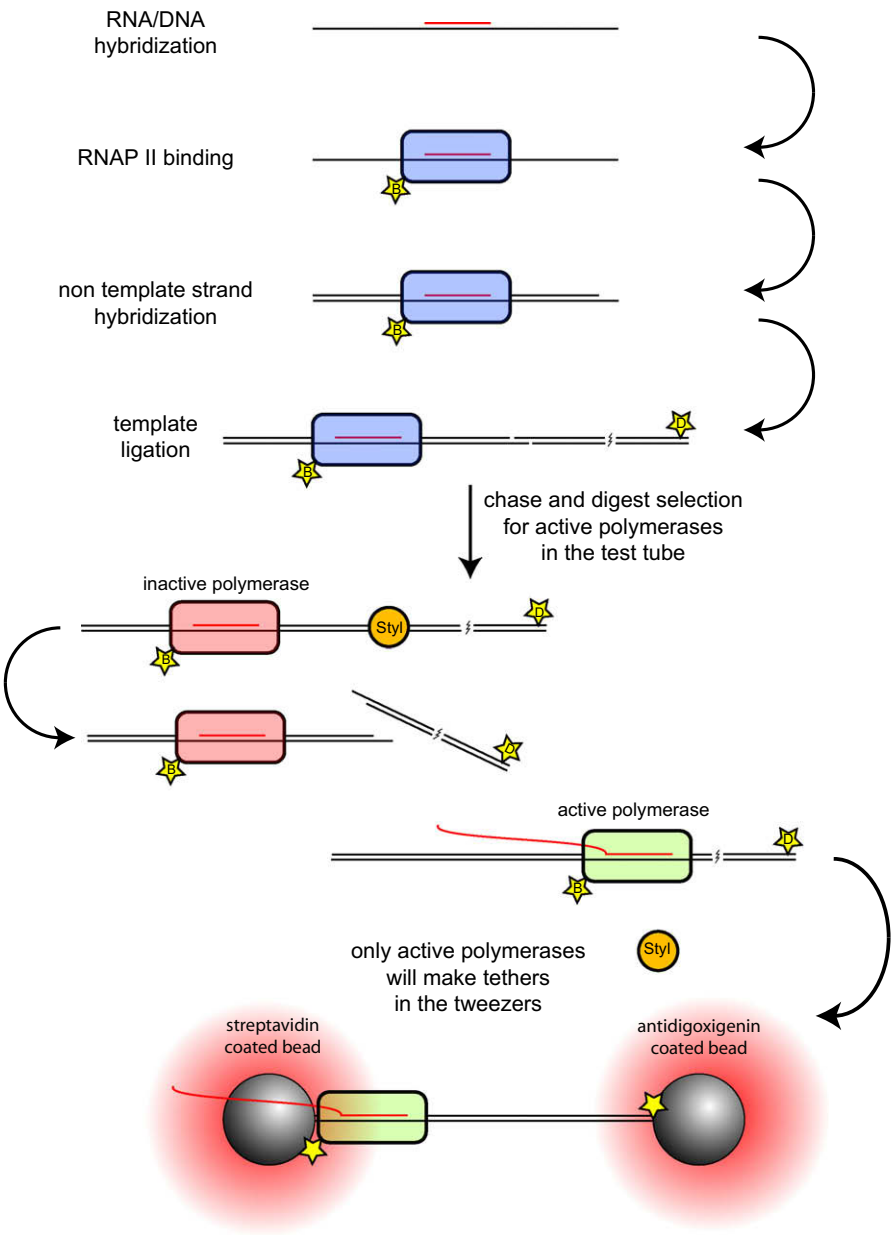

Fig. 2. Selected elongation complex construction. As described in the text, components of the EC are incubated sequentially to form the elongation complex (template ligation) as described previously [7,34]. The complexes are chased with a subset of NTPs so that active polymerases transcribe to and protect a StyI recognition site. During the subsequent digestion step, only the templates of inactive polymerases (red) will be cut while the active polymerases (green) block StyI from binding and are available to make tethers in the tweezers.

Purified pEG2 template (Qiagen Megaprep, 500 μL of 2 μg/μL or ~1 mg, see Supplementary sequence file) is digested with EcoRI (50 μL of 20 U/μL, New England Biolabs (NEB), R0101) in a reaction volume of 800 μL at 37 °C for 4 h. The restriction enzyme is heat killed at 65 °C for 30 min and the DNA template is end-labeled with digoxigenin-11-dUTP (50 μL of 1 mM, Roche, 1093088) using the Klenow fragment of DNA polymerase I (4 μL of 5 U/μL, NEB, M0210) and 100 μL of 400 mM dATP in a final volume of 1 mL at 37 °C for 1 h. The DNA polymerase is heat killed at 75 °C for 20 min. More current protocols use PCR with biotin or digoxigenin labeled oligonucleotide primers to easily generate linear DNA templates of various sizes and attachments chemistries from the same plasmid template. After phenol extraction, ethanol precipitation with 10 M ammonium acetate, and re-suspension in 1 mL TE (10 mM Tris, pH 8.0, 1 mM EDTA), the template is digested with StyI (150 μL of 10 U/μL, NEB, R0500) in a volume of 1.5 mL at 37 °C for 4 h to create the proper sticky end for ligation to the minimal elongation complexes (see next section and Fig. 2). The result of this digestion is two fragments of lengths 9.8 kb and 17 bp. After heat killing (30 min at 65 °C) and another round of phenol extraction followed by ethanol precipitation, the pellet is re-suspended in 10 mM Tris, pH 7.5, 10 mM EDTA, and 400 mM NaCl. The 9.8 kb template is separated from the 17 bp fragment via precipitation with 15% polyethylene glycol 8000 (0.3 volumes of 50% PEG 8 K) for 2 h at 4 °C. The pellet is washed with 70% ethanol, re-suspended in 100 μL TE and adjusted to ~1.5 μg/μL.

3.3. Elongation complexes (ECs)

To study biochemical processes in the optical tweezers, the first step is to arrive at a construct where the biochemical system is poised to carry out the reaction of interest once a single molecule

complex has been trapped and the measurement is ready to take place. The appropriate signal is then introduced, the process begins, and the data are collected. In the case of transcription elongation, one must first generate stalled elongation complexes (ECs) that can be restarted once a tether has been formed in the instrument. The method to generate stalled ECs *in vitro* by the sequential addition of its components was developed in Mikhail Kashlev's lab and has been described in detail [5,6,34].

The protocol for generating the stalled elongation complexes is briefly reviewed here with the specifics used to generate the complexes used in the single molecule experiments. Elongation complexes were prepared using the following oligonucleotides (Oligos Etc.) with the RNA/DNA hybrid sequence in bold although more current protocols use 10 nucleotide longer oligos that have shown increased efficiency in assembly and activity [22]: Template DNA Strand (TDS54) 5′-CAAGGGTGTCGCTTGGGTTGGCTTTTCGCCGTGT C**CCTCTCGAT**GGCTGTAAGT-3′, RNA Primer (RNA9) 5′-**AUCGAGA GG**-3′, Non-template DNA Strand (NDS50) 5′-ACTTACAGCC**ATCGAG AGG**GACACGGCGAAAAGCCAACCCAAGCGACACC-3′. Each oligo is diluted to 50 mM in TE. TDS54 (6 μL, 50 μM) is phosphorylated on the 5′ end using 1 μL T4 PNK (NEB, M0201S, 10 U/μL) in 50 μL for 1 h and diluted to 0.67 μM with 400 μL TE after heat killing the enzyme at 68 °C for 30 min. NDS50 is diluted to 13.3 μM and RNA9 to 1.33 μM. Using these stocks, 1 μL TDS54 is annealed to 1 μL of RNA9 by incubating at 45 °C for 7 min followed by decrements of two degrees every 2 min until reaching 25 °C (Fig. 2). Purified RNAP II (1 μL at 1 μM) is added to the reaction and incubated at 25 °C for 10 min followed by the addition of 1 μL of NDS50 and an incubation of 10 min at 37 °C. The 5′ end of TDS54 creates a 4s base overhang that is used to ligate the minimal EC to the 9.8 kb template DNA [7,35] (4 μL of the EC reaction is mixed the digoxigenin labeled template (1.5 μL of 1.5 μg/μL) and T4 DNA ligase (1 μL with 1 μL buffer, 2.5 μL water, NEB, M0202) at 12 °C for 1–2 h).

3.4. Selected elongation complexes (SECs)

In bulk experiments, the percentage of active complexes produced by this method (∼10% of polymerases) is sufficient to observe and analyze results based on gel electrophoresis with radioactively labeled substrates. However, in the optical tweezers we observe efficiencies around 2–3 times lower which makes single molecule data collection, where one must trap individual complexes, chase them by flowing NTPs, wait to see if activity is present and then repeat, a daunting task. To increase the chance that a stalled EC will chase after the formation of the tether, we pre-chase the ECs with a subset of NTPs in the test tube to select for active complexes. The selection is based on the overlap of the position where active polymerase will stall due to the lack of the next nucleotide and a unique restriction site (Fig. 2, chase and digest). Upon incubation with the corresponding restriction enzyme (StyI, NEB, R0500), active polymerases protect the site from digestion. Templates with inactive enzymes are not protected, leading to the cleavage of their DNA template and the separation of the biotinylated enzyme and the digoxigeninated downstream end of the DNA. Therefore, only enzymes that were initially active after EC construction will be observed as tethers in the tweezers. With this additional step, the efficiency of chasing in the tweezers increased to ∼10–20% and significantly increased the throughput of data collection.

Ligated 9.8 kb ECs (10 μL) are chased with 250 μM GTP, ATP, and CTP (Fermentas, R0481) in TB40 (40 mM KCl, 20 mM Tris, pH 7.9, 5 mM $MgCl_2$, and 1 mM β-mercaptoethanol) in a volume of 20 μL for 15 min at 25 °C. The StyI digest is performed with 5 μL (50 U) in a volume of 50 μL at 37 °C for 10 min to generate the selected EC (SEC). The SECs are kept on ice until they are incubated with functionalized polystyrene beads and loaded into the optical tweezers. The SECs exhibit a high level of activity for about a day and although one SEC reaction can maintain activity for several days, typically new reactions are performed each day to increase throughput.

3.5. Binding to functionalized polystyrene beads

Streptavidin and digoxigenin coated polystyrene beads are used to bind to the polymerase and the downstream end of the DNA template, respectively. These beads are then trapped in the tweezers to exert forces on and measure the motions of the polymerase. Both streptavidin (Spherotech, SVP-20-5, 2.0–2.4 μm) and anti-digoxigenin (Spherotech, DIGP-20-2, 2.0–2.4 μm) functionalized beads are passivated with bovine serum albumin (10× BSA, NEB, B9001S) to reduce nonspecific surface adsorption before incubating with the SECs. Three rounds of spinning the beads (6000*g* for 5 min), aspiration of buffer and re-suspension with 500 μL TB40 serve to exchange the storage buffer for transcription buffer so that BSA binding occurs under the same conditions as the experiment. The beads are incubated in 10× BSA for 30 min with constant shaking. Free BSA is removed by three more cycles of spinning, aspiration and re-suspension with TB40. At the time of the experiment, between 1 and 5 μL of the SECs are mixed with 1 μL of BSA passivated streptavidin beads at room temperature for 10 min. After diluting the binding reaction with 1 mL of TB40 the solution is kept on ice until it is introduced into the tweezers.

3.6. Tether formation

The bound streptavidin beads and the anti-digoxigenin beads now need to be introduced to the flow cell. In addition, the buffer in the cell must be exchangeable to introduce or remove NTPs from the system. To accommodate these requirements, a flow chamber with three different channels that are connected via glass micropipettes is constructed (Fig. 1d). Each channel has its own input and output ports; the upper and lower chambers are used to hold the two types of beads while the central chamber is the experimental channel where the tether is assembled and data is collected. The input to the central chamber can come from the main port carrying −NTP buffer or from the micropipette shunt which carries +NTP buffer. Another possible flow cell design strategy that will be implemented in future studies of single molecule transcription elongation is a laminar flow cell with no physical barriers between different buffers [36]. This arrangement allows for a more rapid exchange of experimental conditions and removes the need for the easily blocked micropipette connections between channels.

The flow cell is mounted on a computer controlled XYZ translation stage (Fig. 1c) so that the relative positions of the trapping lasers and the flow cell can be controlled. To assemble a single SEC between the two traps, the flow cell is moved so that the traps are close to the micropipette that supplies anti-digoxigenin antibody coated beads and a single bead is trapped in one trap. Then, the cell is moved so that the traps are close to the micropipette that supplies SEC bound streptavidin beads and a single bead is trapped in the other trap. The cell is finally moved to the center of the channel and excess beads that have flowed into the central chamber are gently flowed away from the trapped beads.

To convert the primary data of the tweezers into force data, the stiffness of each trap must be known. Since the trapping stiffness of each bead can vary due to differences in bead size, the stiffness of each trap is measured by collecting 20 s of data from the free beads in their respective traps and by fitting the corresponding frequency power spectrum. For a free bead in a symmetric trap, the mean square fluctuation of the bead position will follow a Lorentzian frequency dependence, displaying a corner frequency that is proportional to the stiffness of the trap and inversely proportional to the

friction coefficient of the bead. The power spectrum is then fit by using the friction coefficient of the bead (calculated in practice by using the viscosity of water and a bead size from the manufacturer's specifications) and fitting both the trap stiffness and the ratio of the signal of the detector (volts) to nanometers [27,37]. This step also serves as a quality control checkpoint in that beads that display distorted power spectra are rejected. A significant improvement in this procedure has been developed where an input signal of known power and frequency is introduced by oscillating the stage during the acquisition of the power spectrum data [38]. This alternative procedure results in the presence of a distinct peak in the power spectrum that can be used to experimentally measure the friction coefficient of the bead experimentally so that errors introduced by assuming a value for this parameter are eliminated.

Lastly, the two beads are repeatedly brought closer together and further apart by moving one of the steerable traps in a process called "fishing". A tether is detected when a force is generated in the traps upon separation. At this stage in the opposing force experiment, it would be advantageous to stretch the tether to high force to verify the presence of one and only one tether between the beads (since a single DNA tether has a characteristic relationship between force and extension), however it is important to keep the force as low as possible (<4 pN) as exposure to high opposing forces can prevent the successful chasing of the tether by prematurely inducing a backtracked state (supported by the unpublished observation that TFIIS increases the efficiency of chasing fivefold). In assisting force geometry, this problem would be circumvented as the applied force would serve to rescue backtracked enzymes, but in the opposing force geometry, the verification of a single tether must wait until after the transcription run, when the tether is stretched until it breaks.

3.7. Passive mode single molecule experiment

Once a tether has been formed, data collection can begin. The signals from both optical traps are recorded (to average in data analysis) so as to increase the accuracy of the measurement and only the anti-correlated signal (i.e. changes in the distance between the two beads) is analyzed [30]. In this way, noise and drift that result in the two beads moving in the same direction are eliminated from the data. The system is allowed to relax until a stable baseline signal is obtained (typically $\sim$1 min). Activity buffer (TB40 + NTP) is introduced into the flow cell via the micropipette (Fig. 1d) to initiate transcription. NTPs may be added to any desired concentration along with varying concentrations of pyrophosphate to insure well-defined concentrations for reactants and products. To study the effect of transcription elongation factors, one may also add purified factors to the activity buffer as we have done with TFIIS [35]. As the enzyme moves towards the downstream end of the DNA it shortens the tether length, thus pulling the two beads closer together against the restoring force of the traps. In the passive mode experiment, there is no active feedback on the position of the traps to keep the force constant, so as transcription continues, the force opposing the movement of the enzyme continually increases. This arrangement has the advantage that the behavior of the same enzyme may be observed under a continuum of different forces. One disadvantage is that one ends up with fewer data points overall and at any given force for data analysis. Another possible strategy that represents a mixture of the completely passive and force clamped or constant force experiments is to maintain a constant force for a certain amount of time or number of bases transcribed force and then manually jumping the force to a new value as has been reported in experiments with *E. coli* RNAP [39]. The end of the experiment is defined as when active elongation ceases for a time much longer than the characteristic times of the longest pauses that are observed (10 min). At the end of each run, the distance between the beads is ramped up until the tether breaks and the presence of only a single tether between the beads during the experiment is verified by counting the number of rips before the force on the beads returns to zero. Only experiments with a single rip may be used for further analysis since the presence of more than one tether invalidates the conversion between the end-to-end distance of the DNA and its contour length in nucleotides.

3.8. Force-jump experiment to study enzyme backtracking

In addition to studying the effect of force on transcription in general, often one wants to study how the kinetics of specific processes varies with force. Because the application of high opposing forces prohibits processive transcription due to the tendency of the enzyme to backtrack (see below), we use force-jump experiments to expose the system to high forces for short periods of time. The method is based on experiments developed to directly measure the kinetics of RNA folding over a wide range of forces [40] and has been adapted to directly study the effect of force on the kinetics of entry to the backtracked state during active transcription [35]. During a standard passive mode transcription run, the position of one trap is abruptly jumped to a position a distance away so as to jump up the force acting on the enzyme, this increased distance between the traps is maintained for a time delay (typically 1–3 s) and then returned to its original value. The enzyme position before and after the force jump are compared to ascertain whether the enzyme continued transcribing or backtracked during the force jump (Fig. 3). By performing jumps to different forces and collecting statistics, the force dependence of the probability of backtracking is measured. In addition, by dividing the distance backtracked by the time delay of the force jump, a velocity of backtracking can be measured. In the context of a model of diffusive backtracking [35,41], this velocity and its force dependence can be used to estimate the rate of an individual base pair step of the discrete random walk.

4. Data analysis methods

4.1. RNAP II elongation trajectories

The trajectory of an elongating RNA polymerase displays three important characteristics: active elongation, pauses and arrest. In particular, the polymerase exhibits runs of active elongation that are interrupted by pauses of various durations until the enzyme reaches a force against which it can no longer transcribe. During both pausing and arrest, the polymerase may spontaneously move backwards resulting in a misalignment between the 3′ end of the RNA transcript and the active site. This state is known as backtracking [6,8,42] and evidence suggests that it is important for transcript proofreading [43,44] and that it is responsible for the longest duration pauses that the polymerase experiences [35,45,46]. In the following sections, we describe the analysis of each characteristic behavior and give examples of results from previous analyses.

4.2. General data analysis

The data collected from the single molecule experiment take the form of bead-to-bead distances and forces on each bead. Since the traps act as linear springs at small deviations from their centers, these parameters are directly related by the trap stiffness that is calibrated as described above before each experiment. As the tether between the polymerase and the downstream end of the DNA gets shorter, the bead-to-bead distance decreases and the

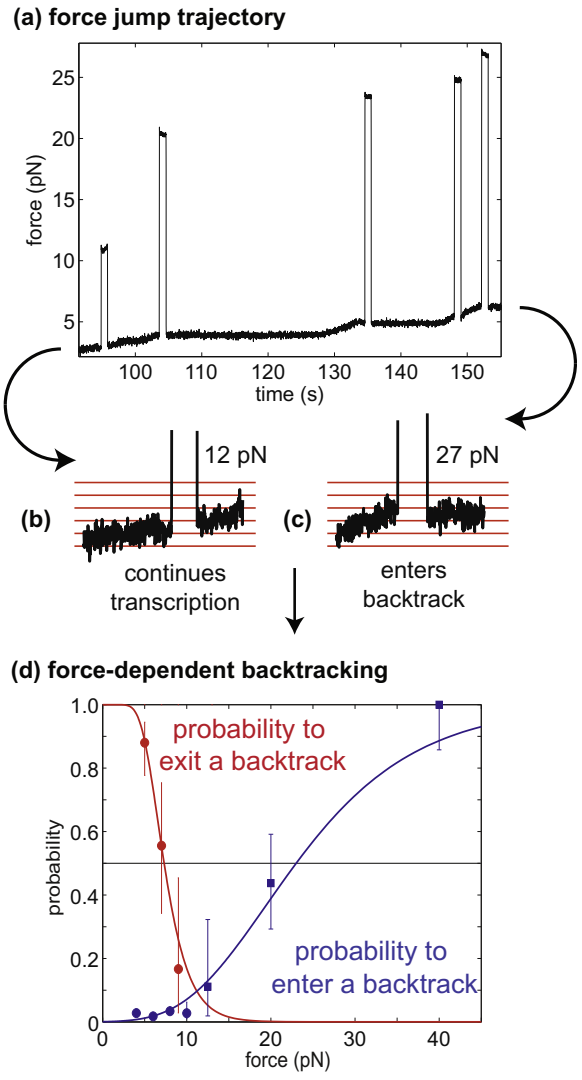

Fig. 3. Force-jump experiment. (a) An example trajectory in which five force-jumps to different forces were performed is plotted with time on the *x*-axis and force on the *y*-axis. Magnified sections of the trace before and after force jumps are shown for the force jump to ~12 pN where the enzyme continued to elongate during the jump (b) and for the jump to ~27 pN where the enzyme backtracked during the jump (c). After many force jumps were performed in this manner, the force-dependent probability of entry to the backtracked state could be plotted (d, blue line). The probability of exiting a backtrack (d, red line) was measured simply by counting the number of times an enzyme recovered from a backtrack and resumed processive elongation as a function of force. Parts of this figure have been adapted from a primary publication [35].

force on each bead increases. By subtracting the bead radii from the bead-to-bead distance, we arrive at the end-to-end distance of the tether, which must then be converted into a contour length in nucleotides using the worm-like-chain model of DNA elasticity that relates the end-to-end distance x, contour length L_0 (9830 nt), force F, persistence length P (53 nm), stretch modulus S (1200 pN-nm) and thermal energy kT at temperature T (Eq. (1)) [28,47].

$$\frac{x}{L_0} = 1 - \frac{1}{2}\left(\frac{kT}{FP}\right)^{1/2} + \frac{F}{S} \tag{1}$$

The data are collected with a 100 Hz bandwidth and are then filtered using the commonly applied 3rd order Savitzky–Golay filter with a time constant of 2.5 s [11,35,45,48,49]. Increasing the filter time constant results in greater precision of the average position of the enzyme over the time window while decreasing it increases the bandwidth at which changes in enzyme position can be observed. In practice, the specific time constant chosen depends on the kinetics of the system, the positional accuracy required and the experimental noise. Trajectories are initially selected for further analysis based on crude criteria such as the length transcribed by the enzyme (>100 nt), the noise in the trace and the confirmation of a single tether. These trajectories are then processed using computer algorithms to collect statistics and distributions of average elongation velocity, pause-free velocity, position dwell times and stall forces. The analysis of the resulting polymerase trajectories will be discussed below.

4.3. Active elongation and pausing

Active runs of elongation may be distinguished from pauses using two main strategies that are based on velocity thresholding and dwell time analysis, respectively. Velocity based methods use the time derivative of the filtered data and a threshold velocity to distinguish elongation from pausing (Fig. 4c). Raising the threshold reduces the number of pauses that are missed, but increases the number of false positives, Lowering the threshold has the opposite effect. The particular value chosen necessarily represents a compromise, but estimates of false scoring can be given based on the distributions of pause-free velocities and pause times that are obtained. Regions of the trajectory that fall under the threshold velocity may be identified as pauses and their durations may be simply calculated as the time it takes until the velocity rises above the threshold. In addition to generating a list of pause durations, this method makes it possible to remove pause regions of the trajectories and to determine the distribution of the force-dependent pause-free velocity for the enzyme that contains information regarding the type of motor mechanism employed by the enzyme during translocation [50].

In dwell time based approaches one measures how long it takes the enzyme to step a certain distance. In cases where single enzyme steps can be observed [39], this distance may be a single base pair. At high NTP concentrations (1 mM) where the enzyme has velocities of 10–20 nt/s, enzyme advances of $\geqslant$3 nt can be observed robustly over the noise in the setup we utilized [35]. A distance window of this size is moved by intervals of its size (i.e. not slid) along the data and the time elapsed for the enzyme to get from the beginning to the end of the window is logged as a single dwell time. After performing this procedure with all the data in a given condition, a list of dwell times is generated and the distribution of times may be calculated (Fig. 4b). The advantage of the dwell time approach is that no arbitrary thresholding is required. One ends up with a distribution of *all* dwell times and can first ask whether separate pause and stepping distributions even exist. Using this technique, one can clearly define pauses as dwells that are greater than the time where the active stepping distribution no longer accounts for the data. In the case of RNAP II, this cutoff time is around 1 s suggesting that dwell times longer than a second represent pauses. Reassuringly, both the velocity threshold and dwell time analyses typically result in similar pause time distributions. In the case of RNAP II, the majority of pauses are short lived, lasting less than a second while some are much longer lasting on the order of minutes [35].

4.4. Enzyme arrest and backtracking

The third feature of single molecule transcription trajectories is the eventual arrest of the enzyme. We define arrest as the point where the polymerase no longer exhibits active elongation for a period of at least 10 min. This time threshold was chosen as a time several times that of the longest pauses we observe and was tested

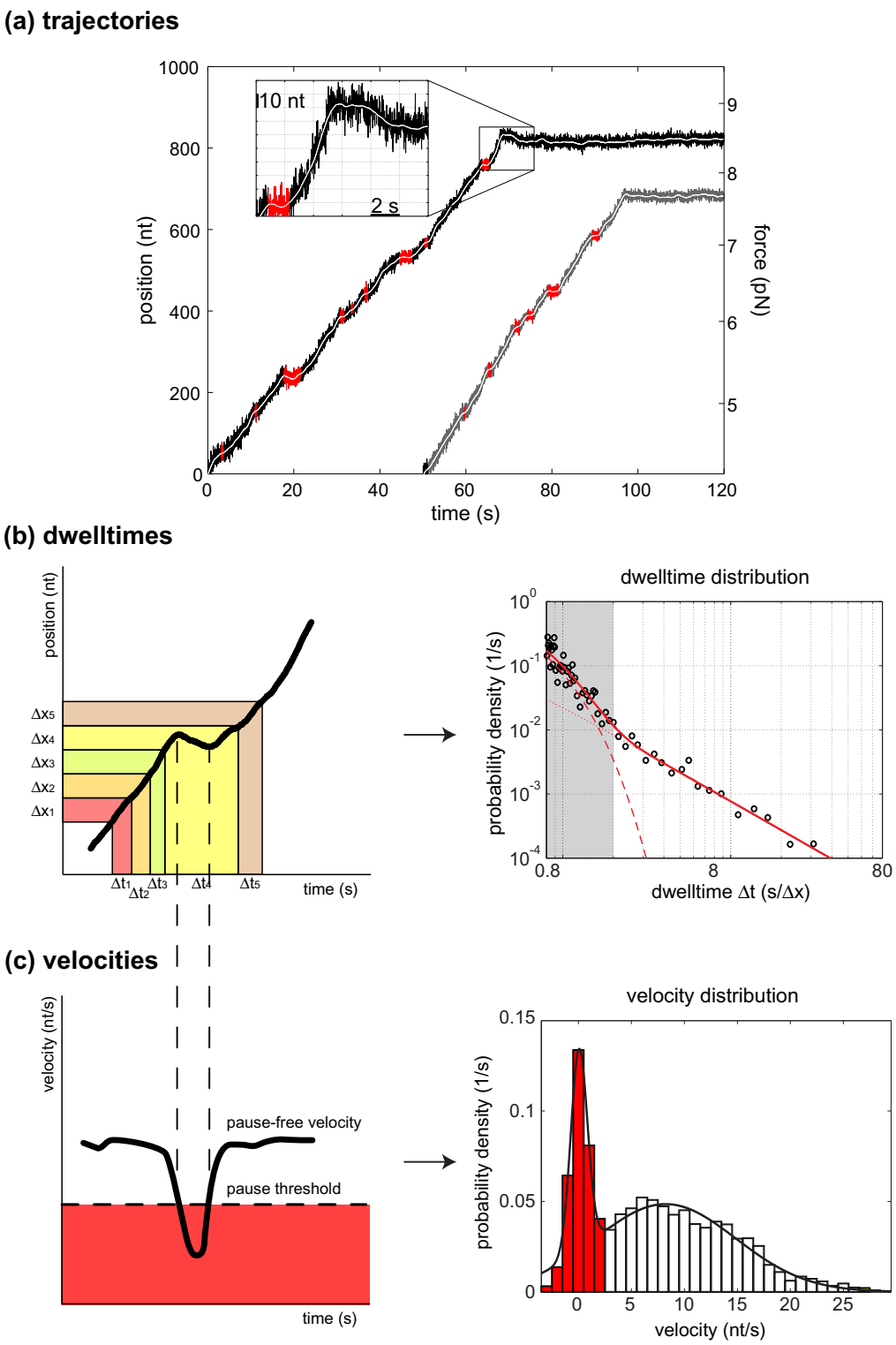

Fig. 4. Single molecule trajectories and data analysis. (a) Two single molecule trajectories of RNAP II offset in time for clarity with the raw unfiltered traces in black and gray traces overlaid with the filtered traces in white (a). The *x*-axis is time and the *y*-axes are distance (left) and force (right) since the experiments were run in passive mode. Runs of active elongation are interrupted by pauses with zero velocity (red). The ends of the runs are shown along with a close-up inset of a clear backtrack right before the end of the black run. (b) To collect dwell times, a distance window (Δx) is scanned along the trajectory and the corresponding dwell times (Δt) are collected. A probability density distribution can then be plotted indicating how likely it is for the polymerase to take a time, Δt to advance by the window Δx. (c) In the velocity analysis, the time derivative of the trajectory is calculated and a threshold (dotted line) is used to identify pause positions. The pause-free velocity can then be calculated by removing these sections of the trajectory. The velocity distribution is shown to the right with velocities likely to be pauses colored red and a fit using two Gaussian distributions (black line). Parts of this figure have been adapted from a primary publication [35].

by waiting much longer times and never observing the resumption of active elongation. At this point we record the force and use the distribution of these forces to define a "stall force" for the polymerase. Because the enzyme has (in principle) many different pause

states available to it, the stall force as we define it is not necessarily the thermodynamic stall force (i.e. the maximum possible force the motor can create based on the efficiency by which it converts the energy released by ATP hydrolysis into mechanical work) that is often reported in other single molecule experiments. Instead, the stall force so determined relates to the biasing of the system into off-pathway states from which the catalytic machinery is disabled. During approximately half of the runs, we observe backwards motion just before the point of arrest suggesting that backtracking causes the abrogation of elongation. This fact is underscored by the observation that the addition of TFIIS, an elongation factor that rescues backtracked polymerases by promoting transcript cleavage, increases the stall force from 8 pN to 17 pN [35]. To detect and quantitate backtracking, researchers studying bacterial RNAP collected and aligned pauses where backtracking was detected and analyzed the average backtrack trajectory [46]. This approach results in useful parameters such as the average distance, average velocity and average trajectory of backtracks. In addition to this type of analysis, one might ask what is the nature of a single backtrack trajectory. A theoretical framework where backtracking is considered a diffusive process in which the enzyme performs a sequence biased random walk along the template until the 3′ end of the transcript is realigned with the active site has been developed [25,35,41]. In the next section, we briefly describe this framework and an approach to fitting the dwell time distributions to gain insight into backtracking and pause mechanisms.

4.5. Dwell time distribution fitting and diffusive backtracking

The experimental dwell time distributions give an unbiased look at the processes experienced by the polymerase. The dwell times from active elongation including different rates in different sequence contexts and those from the various pause states (also sequence dependent) are all included. The task is to separate the distribution into its component parts in an unbiased way to gain insight into the number of processes and their respective prevalence. While this task can be straightforward for distributions made up of small numbers of Poissonian processes that are well separated in time, it appears that various states accessible to the polymerase (i.e. short pauses and slow elongation steps) have overlapping distributions of lifetimes making it difficult to identify them as distinct from each other. Here we discuss the theory of diffusive backtracking and its application to dwell-time fits [35,41].

We consider the trajectory of a polymerase that has entered a backtrack and lost the ability to bias its motion using the energy from NTP hydrolysis. The hypothesis is that the polymerase now acts as a passive particle that diffuses along the DNA template in steps of a single base pair with force-dependent forward (k_F) and backwards (k_B) stepping rates (Eq. (2)).

$$k_F = k_0 e^{F\delta/kT}; \quad k_B = k_0 e^{-F(1-\delta)/kT} \tag{2}$$

Now we ask how long it takes for a diffusive particle to return to the elongation competent state from the first backtracked state by performing a random walk with the above rates. We note that whether or not the first backtracked state is equivalent to the pre-translocated state of the catalytic cycle has no effect on the conclusions regarding the distribution of pause times in the model (although it is an interesting question with potentially interesting consequences). The distribution at times greater than the characteristic stepping rate is a $t^{-3/2}$ power-law that gets cutoff by an exponential (Eqs. (3) and (4)) [41]:

$$\psi(t) \propto t^{-3/2} e^{-t/\tau} \tag{3}$$

$$\tau = \left(\sqrt{k_F} - \sqrt{k_B}\right)^2 \tag{4}$$

The characteristic time of the exponential cutoff (τ) depends on the force bias allowing us to fit dwell times at different forces. Disregarding sequence effects and to a first approximation, at zero force the forward and backward rates of each step in the random walk are equal. Under this condition, the distribution is exactly proportional to $t^{-3/2}$. This long time distribution (Fig. 4b) has now been observed with both bacterial and eukaryotic polymerases [35,48]. To fit the data, we start by assuming that the longest times in the distribution are caused by backtracked pauses, a claim that is widely accepted in the field [45,51,52]. Using our expression for the distribution of backtracked pause times we can see that the amount of these long pauses simply sets the amplitude of the $t^{-3/2}$ distribution. Put another way, the number of long pauses we observe effectively sets the number of short pauses that must be due to backtracking (i.e. if we observe 10 backtracks of 100 s, we expect there to be 300 backtracks of 10 s and 10^4 backtracks of 1 s). We can now ask what percentage of short pauses are not accounted for by backtracking and attempt to attribute them to other processes. The experimental data fitted with this procedure leads us to believe that the gross majority of pauses greater than 2 s are due to backtracking [35,41].

5. Concluding remarks

The application of single molecule techniques has provided several new insights into the nature of transcription elongation. On its own, RNAP II is unable to transcribe once the force reaches 8 pN, however the addition of a single transcription factor (TFIIS) allows the polymerase to continue against forces of up to 20 pN [35]. An important detail is that the inability of the polymerase to transcribe against force is due to the off-pathway state of backtracking and no force-dependent changes in the pause-free velocity of a single enzyme have been observed. These observations are in contrast to the behavior of *E. coli* RNAP that transcribes up to 25 pN on its own and exhibits force dependence in its velocity in some assays [39,51]. While there are differences in the assays that may lead to these different results, an interesting possibility is that the eukaryotic polymerase has evolved alongside with mechanical gene regulatory processes that control the ability of the polymerase to perform work.

Many studies have commented on the nature of pause states that are experienced during transcription elongation. Work on backtracking [6,8,42], hairpin stabilized pauses [52,53], elemental pauses [54] and ubiquitous pauses [45] have all demonstrated that the enzyme experiences pauses with durations that vary from less than a second (elemental) to minutes (backtracking). However, the question as to how many different pause mechanisms exist is still a matter of active research and we have suggested that the ubiquitous pauses observed in *E. coli* RNAP may also be due, at least in part, to backtracking [35,41]. In addition, the RNA transcript structure itself can influence pausing in *E. coli* RNAP (i.e. class I or hairpin stabilized pauses [55]) although single molecule work on *E. coli* RNAP has suggested that there is no general effect when RNA structures are unfolded by force [56]. The effect of RNA structure on RNAP II remains to be studied with single molecule methods. The techniques described above must be slightly modified to allow for the attachment of the 5′ end of the transcript to a bead, but the general concepts are the same.

With the goal of augmenting our understanding the mechanical regulation of eukaryotic transcription, future single molecule studies will include the addition of other elongation factors (i.e. TFIIF, Elongin and ELL [57]) in similar assays. In the long run, more complicated transcription systems may be isolated or reconstituted to study transcription under conditions more similar to those found inside the nucleus of a living cell. The information gleaned from

single molecule elongation assays will add to the impressive amount of research and results that continue to be obtained via bulk methods to generate more complete and detailed models of transcription elongation and its regulation.

Acknowledgments

We thank Mikhail Kashlev and Maria Kireeva for making the described work possible. We also thank Anna Wiedmann, Juan Parrondo and Steve Smith for invaluable contributions along the way. The work described was funded by an NIH Grant GM-32543 (C.J.B.), a Jane Coffin Childs Postdoctoral Fellowship (E.A.G.) and by a Helen Hay Whitney Postdoctoral Fellowship (S.W.G.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ymeth.2009.04.021.

References

[1] A. Ashkin, J.M. Dziedzic, J.E. Bjorkholm, S. Chu, Opt. Lett. 11 (1986) 288.
[2] C. Bustamante, Annu. Rev. Biochem. 77 (2008) 45.
[3] W.J. Greenleaf, M.T. Woodside, S.M. Block, Annu. Rev. Biophys. Biomol. Struct. 36 (2007) 171.
[4] J.C. Meiners, S.R. Quake, Phys. Rev. Lett. 84 (2000) 5014.
[5] M.L. Kireeva, L. Lubkowska, N. Komissarova, M. Kashlev, Meth. Enzymol. 370 (2003) 138.
[6] N. Komissarova, M. Kashlev, J. Biol. Chem. 272 (1997) 15329.
[7] N. Komissarova, M.L. Kireeva, J. Becker, I. Sidorenkov, M. Kashlev, Meth. Enzymol. 371 (2003) 233.
[8] E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, Cell 89 (1997) 33.
[9] R.J. Davenport, G.J. Wuite, R. Landick, C. Bustamante, Science 287 (2000) 2497.
[10] N.R. Forde, D. Izhaky, G.R. Woodcock, G.J. Wuite, C. Bustamante, Proc. Natl. Acad. Sci. USA 99 (2002) 11682.
[11] M.D. Wang et al., Science 282 (1998) 902.
[12] H. Yin et al., Science 270 (1995) 1653.
[13] G.A. Kassavetis, M.J. Chamberlin, J. Biol. Chem. 256 (1981) 2777.
[14] R. Landick, Biochem. Soc. Trans. 34 (2006) 1062.
[15] G. Muse et al., Nat. Genet. 39 (2007) 1507.
[16] T.C. Reeder, D.K. Hawley, Cell 87 (1996) 767.
[17] I. Gusarov, E. Nudler, Mol. Cell 3 (1999) 495.
[18] W.S. Yarnell, J.W. Roberts, Science 284 (1999) 611.
[19] T. Pan, T. Sosnick, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 161.
[20] M. Yonaha, Mol. Cell 3 (1999) 593.
[21] M. de la Mata, Mol. Cell 12 (2003) 525.
[22] M. Kireeva et al., Mol. Cell 18 (2005) 97.
[23] R.N. Fish, C.M. Kane, Biochim. Biophys. Acta 1577 (2002) 287.
[24] S. Klumpp, T. Hwa, Proc. Natl. Acad. Sci. USA 105 (2008) 18159.
[25] M. Voliotis, N. Cohen, C. Molina-Paris, T. Liverpool, Biophys. J. 94 (2007) 334.
[26] J.R. Moffitt, Y. Chemla, S.B. Smith, C. Bustamante, Annu. Rev. Biochem. 77 (2008) 19.1.
[27] K.C. Neuman, S.M. Block, Rev. Sci. Instrum. 75 (2004) 2787.
[28] C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, Science 265 (1994) 1599.
[29] Y. Seol, J. Li, P.C. Nelson, T.T. Perkins, M.D. Betterton, Biophys. J. 93 (2007) 4360.
[30] J.R. Moffitt, Y.R. Chemla, D. Izhaky, C. Bustamante, Proc. Natl. Acad. Sci. USA 103 (2006) 9006.
[31] M.T. Valentine et al., Opt. Lett. 33 (2008) 599.
[32] M.W. Allersma, F. Gittes, M.J. deCastro, R.J. Stewart, C.F. Schmidt, Biophys. J. 74 (1998) 1074.
[33] F. Gittes, C.F. Schmidt, Opt. Lett. 23 (1998) 7.
[34] I. Sidorenkov, N. Komissarova, M. Kashlev, Mol. Cell 2 (1998) 55.
[35] E.A. Galburt et al., Nature 446 (2007) 820.
[36] L.R. Brewer, P.R. Bianco, Nat. Meth. 5 (2008) 517.
[37] K. Berg-Sørensen, H. Flyvbjerg, Rev. Sci. Instrum. 75 (2004) 594.
[38] S. Tolić-Norrelykke et al., Rev. Sci. Instrum. 77 (2006) 103101.
[39] E.A. Abbondanzieri, W.J. Greenleaf, J. Shaevitz, R. Landick, S.M. Block, Nature 438 (2005) 460.
[40] P.T. Li, D. Collin, S.B. Smith, C. Bustamante, I. Tinoco, Biophys. J. 90 (2006) 250.
[41] M. Depken, E.A. Galburt, S.W. Grill, Biophys. J. 96 (2009) 2189.
[42] N. Komissarova, M. Kashlev, Proc. Natl. Acad. Sci. USA 94 (1997) 1755.
[43] D. Reines, J. Biol. Chem. 267 (1992) 3795.
[44] M.J. Thomas, A.A. Platas, D.K. Hawley, Cell 93 (1998) 627.
[45] K.C. Neuman, E.A. Abbondanzieri, R. Landick, J. Gelles, S.M. Block, Cell 115 (2003) 437.
[46] J. Shaevitz, E.A. Abbondanzieri, R. Landick, S.M. Block, Nature 426 (2003) 684.
[47] T. Odijk, Macromolecules 28 (1995) 7016.
[48] Y. Mejia, H. Mao, N. Forde, C. Bustamante, J. Mol. Biol. 382 (2008) 628.
[49] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627.
[50] C. Bustamante, Y.R. Chemla, N.R. Forde, D. Izhaky, Annu. Rev. Biochem. 73 (2004) 705.
[51] K.M. Herbert et al., Cell 125 (2006) 1083.
[52] I. Toulokhonov, J. Zhang, M. Palangat, R. Landick, Mol. Cell 27 (2007) 406.
[53] C.L. Chan, D. Wang, R. Landick, J. Mol. Biol. 268 (1997) 54.
[54] I. Toulokhonov, R. Landick, Mol. Cell 12 (2003) 1125.
[55] I. Artsimovitch, R. Landick, Proc. Natl. Acad. Sci. USA 97 (2000) 7090.
[56] R.V. Dalal et al., Mol. Cell 23 (2006) 231.
[57] B.J. Elmendorf, A. Shilatifard, Q. Yan, J.W. Conaway, J. Biol. Chem. 276 (2001) 23109.