Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments

Dulin, D; Berghuis, BA; Depken, SM; Dekker, NH

DOI
10.1016/j.sbi.2015.08.007

Publication date
2015

Document Version
Accepted author manuscript

Published in
Current Opinion in Structural Biology

Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments

David Dulin[2], Bojk A. Berghuis[1], Martin Depken[1] & Nynke H. Dekker[1*]

[1]Department of Bionanoscience, Kavli institute of Nanoscience, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
[2]Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

* Correspondence to: n.h.dekker@tudelft.nl

Abstract
Single-molecule experiments provide a unique means for real-time observation of the activity of individual biomolecular machines. Through such techniques, insights into the mechanics of e.g. polymerases, helicases, and packaging motors have been gleaned. Here we describe the recent advances in single-molecule force spectroscopy instrumentation that have facilitated high-throughput acquisition at high spatiotemporal resolution. The large datasets attained by such methods can capture rare but important events, and contain information regarding stochastic behaviors covering many orders of magnitude in time. We further discuss analysis of such data sets, and with a special focus on the pause states described in the general literature on RNA polymerase pausing we compare and contrast the signatures of different reaction pathways.

Introduction
The correct readout, maintenance, repair, and replication of genomic information involves a stunning variety of carefully coordinated and regulated molecular machines, including key players such as RNA polymerases (RNAp) and replisomes. The progression of these machines is highly dynamic, as their activity is frequently and stochastically interrupted by numerous co-factors or intrinsic catalytic events. For example, during transcription, RNAp progression is interrupted by various types of pauses, including backtracking pauses resulting from RNAp diffusing upstream on the DNA template (Galburt et al., 2007; Shaevitz et al., 2003), and regulatory pauses allowing for transcription-translation synchronization and RNA co-transcriptional folding (Herbert et al., 2006; Larson et al., 2014; Zamft et al., 2012). During replication, the replisome also frequently halts, e.g. for the formation of primers on the Okazaki fragment (Lee et al., 2006), due to DNA polymerase exchange (Geertsema et al., 2014), and possibly also due to proofreading (Ibarra et al., 2009; Manosas et al., 2012; Wuite et al., 2000).

To obtain quantitative insight into the functioning of molecular machines, probing at the single-molecule level with nanometre-scale spatial and millisecond temporal resolution has proved to be very successful (Dulin et al., 2013). This success results from the ability to detect transient intermediates or rare events that are masked when ensemble techniques are used. Single-molecule methods may be broadly grouped into optical methods visualizing individual molecules using fluorescence-based microscopy (Huang et al., 2009; Joo et al., 2008; Lichtman and Conchello, 2005), and force-based methods such as atomic force microscopy, flow-based stretching, and optical and magnetic tweezers (Neuman and Nagy, 2008). While both approaches have provided invaluable insight into the functioning of molecular machines, we here focus on the latter. Motivated by the sub-nanometer elementary length scales of biological substrates such as DNA, much effort has gone into developing instruments with high spatiotemporal resolution (Dulin et al., 2013). Provided

sufficiently low enzyme kinetic rates (~1 bp/s), optical tweezers have been developed with single DNA base pair (bp) resolution (Moffitt et al., 2008). Such tweezers have allowed detailed mechanistic studies of e.g. RNAp (Abbondanzieri et al., 2005), the Φ29 packaging motor (Chemla et al., 2005), and the Hepatitis C viral helicase NS3 (Cheng et al., 2011).

Though biological systems that catalyze chemical reactions along a single pathway (Chemla et al., 2005; Chistol et al., 2012; Moffitt et al., 2009) have successfully been characterized with high-resolution optical tweezers, approaches suitable for analysis of more complicated pathways displayed remain a challenge due to the low yield of such technique. We here review recent approaches to parallelized single-molecule force-spectroscopy methods using widefield (e.g. camera-based) detection to simultaneously track the action of a large number of independent molecules and substrates (**Fig. 1**). For the case of polymerizing enzymes with an internal pause dynamics, we further discuss an analysis approach that capitalizes on large data sets to reveal information about the enzyme translocation pathway, and review the signatures of various pause types described in the literature.

Single-molecule approaches go parallel

The simplest high-throughput single-molecule approach, typically denoted flow–stretch experiments, involves the use of liquid flow to exert drag forces on tethered beads (**Fig. 1a**). Flow-stretch experiments have the advantage of being relatively simple while still allowing for the observation of enzymatic activity on long DNA templates (> 40 kb) at about a 100 nm resolution. This approach has been extensively used to study the dynamics of replication in the bacteriophage T7 model system (Duderstadt et al., 2014; Geertsema et al., 2015). A more recent approach, denoted nanophotonic standing-wave array trap (nSWAT), uses microfluidics combined with waveguiding to create an array of optical traps (Soltani et al., 2014) (**Fig. 1b**). Being integrated into a micro-fabricated chip, the nSWAT assay substantially reduces the complexity of optical tweezers and the influence of mechanical noise. Another recent approach is acoustic force spectroscopy (AFS), which uses a flow-cell integrated piezo element to apply acoustic pressure to micron-sized beads (Sitters et al., 2015) (**Fig. 1c**). AFS is ideal for high-throughput experiments, as it generates a homogenous force over distances much larger than the sizes of the force transducers. For the same reasons, magnetic tweezers (MT) are also inherently suitable for parallelization (Cnossen et al., 2014) (**Fig. 1d**). The capabilities of a widefield camera-based imaging approach combined with MT were first demonstrated by following either tens of molecules in real-time (Ribeck and Saleh, 2008), or hundreds with post-processing image analysis (De Vlaminck et al., 2011). The potential for multiplexing is further expanded by a recent increases in camera resolution (**Fig. 1d**), and the use of graphics processing units interfaced with the CUDA parallel computing framework. It is now possible to simultaneously follow up to 800 beads in real-time, with a resolution of 1 nm at 25 Hz (Cnossen et al., 2014).

Further development of high-throughput experiments will greatly help the mapping of single-molecule events that are either very long-lived or rare. We now discuss how to visualize such data, and illustrate the two main scenario where multiplexing becomes important by discussing two recent studies utilizing high-throughput MT approaches to study DNA binding proteins and polymerase activity.

Dwell-time distributions

Dwell-time analysis has a long history of being applied to single-molecule experiments where it is possible to directly measure the duration (the dwell-time) of the event of interest (Roy et al., 2008). With the increase in the type of events that can be studied with statistical significance in multiplexed force-spectroscopy experiments, it becomes important to be able

to visualize the data over many orders of magnitude in time and probability. For such visualizations it is especially useful to collect dwell times in a weighted histogram, where the score in each bin is normalized with the bin width and total number of dwell times recorded. Such weighted histogram represents empirical dwell-time distributions (DTD). The basic shape of an empirical DTD is insensitive to the binning used, which makes it possible to consistently visualize the data over several orders of magnitude by using various binning schemes (e.g. using log-scaled bins).

Magnetic tweezers based study of DNA binding proteins capitalizing on improved statistics
High throughput proved essential in a recent study probing *E.coli* replication fork arrest induced by the tight interaction between the DNA binding protein Tus and its cognate DNA-binding sequence *Ter* (Berghuis et al., 2015). It had previously been argued that the tight interaction was dependent on specific interactions between Tus and the replisome (Mulugu et al., 2001). By using a multiplexed MT DNA hairpin assay (**Fig. 2a**) to mimicking fork progression without the proteins of the repliosome, it was shown that protein-protein interactions were not necessary for tight Tus-*Ter* interactions. Gathering the necessary dwell-time statistics would have been very demanding in time without multiplexing, as strand separation frequently remained blocked at the Tus–*Ter* site for hundreds of seconds (**Fig. 2b**).

Magnetic tweezers-based study of polymerase activity capitalizing on improved statistics
A recent study of viral mutagenesis used high-throughput magnetic tweezers to examine rare nucleotide misincorporation events during replication (Dulin et al., 2015) **(Fig. 1d)**. In this work, the progression of a viral RNA-dependent RNA polymerase (RdRp) was monitored at the single-molecule level (**Fig. 2c**). Different traces taken during the same experiment differ remarkably (**Fig. 2d**)**,** and to characterize rare long pauses large data sets were recorded at many experimental conditions. To analyze the data, a new approach based on the direct fitting of mechano-chemical models to complete data sets was used. We conclude by illustrating the basic ingredients of this approach, and comment on the signatures of the polymerase pause types found in the literature.

Dwell-time based analysis for untangling molecular motor translocation pathways
Standard approaches used to analyze single-molecule traces require either extensive pre-processing of traces in the form of pause-picking algorithms (Abbondanzieri et al., 2005; Herbert et al., 2006), or discards important information about pause durations through considering local velocities (Neuman et al., 2003; Zamft et al., 2012). Here we use simulations to outline a recent extension of dwell-time analysis that can be used to fit complete data sets directly to mechano-chemical models of translocating motors (Dulin et al., 2015). In this approach, the traces are not segmented with fixed time intervals as is done in velocity-based analysis methods, but is segmented with fixed translocation distances—the dwell-time windows—to produce dwell-time statistics. As one often lacks the resolution to gather dwell-time statistics for single translocation steps, the data is instead collected over dwell-time windows that can span multiple translocation steps. We will here focus on pause analysis, and not discuss the delicate and largely unresolved influence of noise on the shorter time events, such as pause-free translocation. For illustrative purposes, we will assume that on the time scale of pauses it is possible to smooth the trace sufficiently to suppress the effect of noise.

In **Figure 3a** we show a simulated example trace with a dwell-time window of 10 bp, and the series of scored dwell-times indicated as t_1, t_2, etc. Given the rates in any specific translocation scheme, one can in principle use first-passage time analysis (Redner, 2007) to

calculate the DTD $P_1(t)$ of times t for taking one step. The DTD for taking n steps is then the n-fold convolution $P_n(t)= [P_1*P_1*\ldots* P_1](t)$ of the single-step distribution. The DTD can in principle be arbitrarily complicated, but we here focus on a few of the simplest important cases that result from translocation schemes suggested in the literature.

First order process. In the simplest scenario, each step along the elongation pathway is a first order process with a constant rate k (**Figure 3b**, inset). The one-step DTD is then an exponential distribution, and the n step distribution is a Gamma distribution of order n and characteristic rate k. From **Figure 3b** it is clear that the width of the Gamma distribution depends on the number of substeps, a fact that can be used to estimate n.

Off-pathway pauses. Polymerase elongation can be temporarily halted by thermally induced changes in enzyme structure (Abbondanzieri et al., 2005; Herbert et al., 2006; Neuman et al., 2003), inducing what is referred to as off-pathway pauses. At its simplest, the structural change occurs with a first-order rate k_p, and is reversed with the rate k_e (**Figure 3c**, inset). Though it is straightforward to calculate the DTD for arbitrary rates, again for simplicity, we focus on the case where pauses are entered infrequently in each dwell-time window, and typically last much longer than it takes to cross the dwell-time window without pausing. In this case, the first order process of the main elongation pathway captures most dwell-time windows, resulting in a Gamma distribution for short dwell times (**Figure 3c**). For longer dwell times, infrequent entrance into the pause state results in exponentially distributed dwell-times with a lifetime $1/k_e$. The probabilistic weight under the short-time Gamma distribution and the long-time exponential distribution reports on the probability to move through a dwell-time window without pausing and the probability to enter at least one pause. Importantly, this type of pause will not have a lifetime that depends on nucleotide concentration, as it is exited through the reversal of the thermal fluctuation that triggered the pause. The probability of entering a pause will depend on nucleotide concentration though, as the pause entry directly competes with nucleotide addition.

On-pathway pauses. Another type of pause observed in the literature results from a stochastic and drastic reduction of the nucleotide addition speed, e.g. after a non-cognate base has been inserted (Dulin et al., 2015; Tsai and Johnson, 2006; Yang et al., 2012). This leads to an on-pathway pause, which has a DTD similar to the off-pathway pause but with a pause-escape rate that is sensitive to nucleotide concentrations and an entrance probability that is independent on overall nucleotide concentration. Based on the different nucleotide dependencies of on- and off-pathway pauses, it is possible to distinguish them using nucleotide concentrations sweeps.

Hybrid pauses. A recent study (Dulin et al., 2015) suggests that there exist off-pathway pauses, but where the paused state retains a degree of enzymatic activity. For such pauses, both pause probability and pause exit rate will depend on the overall nucleotide concentration.

Backtracking pauses. Another important pause is the backtracking pause, which results from the polymerase diffusively moving backwards and misaligning the product terminal with the active site. To reverse such a pause, the polymerase has to diffuse back to its original position (Galburt et al., 2007; Shaevitz et al., 2003). The diffusive return induces pauses with a distinct -3/2 power-law distributed dwell-times over a large time range (Depken et al., 2009; Galburt et al., 2007; Voliotis et al., 2008) (**Figure 3d**, inset). The width of this region is bounded by a lower corner time and upper cutoff time, both set by the hopping rates in the diffusive backtrack (**Fig. 3d**) (Depken et al., 2009).

Composite distributions. Experimental DTDs are generally more complicated than any of the above scenarios, both in that the timescales might not be well separated and that several types of pauses can be present at once. Still, the large data sets now becoming available through multiplexed experiments makes it possible to distinguish a large number of features (**Fig. 4)**. Details regarding the precise reaction scheme can be extracted not only by noting that pause mechanisms can be differentiated by how their probabilities and lifetimes depend on nucleotide concentration, but also from force dependence, and from the introduction of non-native nucleotide analogs (Dulin et al., 2015). In **Fig. 4** we show the empirical DTD based on successively larger data sets collected from of a viral RdRP (Dulin et al., 2015) with several pause types. An increasingly significant and rich structure pertaining to long-time and rare events is seen as the data sets are enlarged.

To fit models to DTD it is often convenient to determine the number of pause states through the Bayesian Information Criterion (BIC) (Schwarz, 1978), and fit out model parameters by Maximum Likelihood Estimation (MLE) (Cowan, 1998) applied directly to the dwell-times. At present, this approach requires at least a partial analytic solution for the first-passage time of the corresponding reaction scheme (Dulin et al., 2015), though future developments of numerical approaches will likely eliminate this requirement.

Outlook

The approaches discussed here can be applied to many different types of molecular motors. For example, replication is regularly interrupted by rare and slow events, such as proofreading activity on the nascent DNA and polymerase exchange. A recent study has also demonstrated the utility of combining high-resolution single-molecule techniques with next generation sequencing technology (Larson et al., 2014) to report on the specific DNA sequence responsible for a regulatory transcriptional pause. With further progress in multiplexing and the ability to localize the absolute position of a molecular machine on its template, it should soon become possible to investigate sequence-dependent pausing patterns on a global scale.

Acknowledgments

We thank our current and former colleagues in the Nynke Dekker and Martin Depken groups of the Department of Bionanoscience, TU Delft for useful discussions, in particular Jelmer Cnossen, Behrouz Eslami-Mossallam, and Theo van Laar. We also acknowledge Elio Abbondanzieri for discussions on RNA polymerases, and Eric Sniider, Clara Posthuma, and Craig Cameron for discussions on RNA-dependent RNA polymerases. M.D. acknowledges early discussions with Stephan Grill and Eric Galburt regarding dwell-time distributions and polymerases. This work was financed by a TU Delft startup grant to M.D. and by VICI and TOP grants from the Netherlands Organisation for Scientific Research to N.H.D.

References

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature *438*, 460-465.

Berghuis, B.A., Dulin, D., Xu, Z.Q., van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N.E., Depken, M., and Dekker, N.H. (2015). Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. Nature chemical biology *11*, 579-585.

Chemla, Y.R., Aathavan, K., Michaelis, J., Grimes, S., Jardine, P.J., Anderson, D.L., and Bustamante, C. (2005). Mechanism of force generation of a viral DNA packaging motor. Cell *122*, 683-692.

Cheng, W., Arunajadai, S.G., Moffitt, J.R., Tinoco, I., Jr., and Bustamante, C. (2011). Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. Science (New York, NY *333*, 1746-1749.

Chistol, G., Liu, S., Hetherington, C.L., Moffitt, J.R., Grimes, S., Jardine, P.J., and Bustamante, C. (2012). High degree of coordination and division of labor among subunits in a homomeric ring ATPase. Cell *151*, 1017-1028.

Cnossen, J.P., Dulin, D., and Dekker, N.H. (2014). An optimized software framework for real-time, high-throughput tracking of spherical beads. The Review of scientific instruments *85*, 103712.

Cowan, G. (1998). Statistical Data Analysis (Oxford University Press).

De Vlaminck, I., Henighan, T., van Loenhout, M.T., Burnham, D.R., and Dekker, C. (2012). Magnetic forces and DNA mechanics in multiplexed magnetic tweezers. PloS one *7*, e41432.

De Vlaminck, I., Henighan, T., van Loenhout, M.T., Pfeiffer, I., Huijts, J., Kerssemakers, J.W., Katan, A.J., van Langen-Suurling, A., van der Drift, E., Wyman, C*., et al.* (2011). Highly parallel magnetic tweezers by targeted DNA tethering. Nano letters *11*, 5489-5493.

Depken, M., Galburt, E.A., and Grill, S.W. (2009). The origin of short transcriptional pauses. Biophysical journal *96*, 2189-2193.

Duderstadt, K.E., Reyes-Lamothe, R., van Oijen, A.M., and Sherratt, D.J. (2014). Replication-fork dynamics. Cold Spring Harbor perspectives in biology *6*.

Dulin, D., Lipfert, J., Moolman, M.C., and Dekker, N.H. (2013). Studying genomic processes at the single-molecule level: introducing the tools and applications. Nature reviews Genetics *14*, 9-22.

Dulin, D., Vilfan, I.D., Berghuis, B.A., Hage, S., Bamford, D.H., Poranen, M.M., Depken, M., and Dekker, N.H. (2015). Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. Cell Rep.

Galburt, E.A., Grill, S.W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E., Kashlev, M., and Bustamante, C. (2007). Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. Nature *446*, 820-823.

Geertsema, H.J., Duderstadt, K.E., and van Oijen, A.M. (2015). Single-molecule observation of prokaryotic DNA replication. Methods Mol Biol *1300*, 219-238.

Geertsema, H.J., Kulczyk, A.W., Richardson, C.C., and van Oijen, A.M. (2014). Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery. Proceedings of the National Academy of Sciences of the United States of America *111*, 4073-4078.

Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., and Block, S.M. (2006). Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell *125*, 1083-1094.

Huang, B., Bates, M., and Zhuang, X. (2009). Super-resolution fluorescence microscopy. Annual review of biochemistry *78*, 993-1016.

Ibarra, B., Chemla, Y.R., Plyasunov, S., Smith, S.B., Lazaro, J.M., Salas, M., and Bustamante, C. (2009). Proofreading dynamics of a processive DNA polymerase. Embo Journal *28*, 2794-2802.

Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., and Ha, T. (2008). Advances in single-molecule fluorescence methods for molecular biology. Annual review of biochemistry *77*, 51-76.

Larson, M.H., Mooney, R.A., Peters, J.M., Windgassen, T., Nayak, D., Gross, C.A., Block, S.M., Greenleaf, W.J., Landick, R., and Weissman, J.S. (2014). A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science (New York, NY *344*, 1042-1047.

Lee, J.B., Hite, R.K., Hamdan, S.M., Xie, X.S., Richardson, C.C., and van Oijen, A.M. (2006). DNA primase acts as a molecular brake in DNA replication. Nature *439*, 621-624.

Lichtman, J.W., and Conchello, J.A. (2005). Fluorescence microscopy. Nature methods *2*, 910-919.

Lipfert, J., Hao, X., and Dekker, N.H. (2009). Quantitative modeling and optimization of magnetic tweezers. Biophysical journal *96*, 5040-5049.

Manosas, M., Spiering, M.M., Ding, F., Bensimon, D., Allemand, J.F., Benkovic, S.J., and Croquette, V. (2012). Mechanism of strand displacement synthesis by DNA replicative polymerases. Nucleic acids research *40*, 6174-6186.

Moffitt, J.R., Chemla, Y.R., Aathavan, K., Grimes, S., Jardine, P.J., Anderson, D.L., and Bustamante, C. (2009). Intersubunit coordination in a homomeric ring ATPase. Nature *457*, 446-450.

Moffitt, J.R., Chemla, Y.R., Smith, S.B., and Bustamante, C. (2008). Recent advances in optical tweezers. Annual review of biochemistry *77*, 205-228.

Mulugu, S., Potnis, A., Shamsuzzaman, Taylor, J., Alexander, K., and Bastia, D. (2001). Mechanism of termination of DNA replication of Escherichia coli involves helicase-contrahelicase interaction. Proceedings of the National Academy of Sciences of the United States of America *98*, 9569-9574.

Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., and Block, S.M. (2003). Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell *115*, 437-447.
Neuman, K.C., and Nagy, A. (2008). Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nature methods *5*, 491-505.
Redner, S. (2007). A Guide to First-Passage Processes (Cambridge University Press).
Ribeck, N., and Saleh, O.A. (2008). Multiplexed single-molecule measurements with magnetic tweezers. The Review of scientific instruments *79*, 094301.
Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. Nature methods *5*, 507-516.
Schwarz, G. (1978). Estimating Dimension of a Model. AnnStat *6*, 461-464.
Shaevitz, J.W., Abbondanzieri, E.A., Landick, R., and Block, S.M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. Nature *426*, 684-687.
Sitters, G., Kamsma, D., Thalhammer, G., Ritsch-Marte, M., Peterman, E.J., and Wuite, G.J. (2015). Acoustic force spectroscopy. Nature methods *12*, 47-50.
Soltani, M., Lin, J., Forties, R.A., Inman, J.T., Saraf, S.N., Fulbright, R.M., Lipson, M., and Wang, M.D. (2014). Nanophotonic trapping for precise manipulation of biomolecular arrays. Nature nanotechnology *9*, 448-452.
Tsai, Y.C., and Johnson, K.A. (2006). A new paradigm for DNA polymerase specificity. Biochemistry *45*, 9675-9687.
Voliotis, M., Cohen, N., Molina-Paris, C., and Liverpool, T.B. (2008). Fluctuations, pauses, and backtracking in DNA transcription. Biophysical journal *94*, 334-348.
Wuite, G.J., Smith, S.B., Young, M., Keller, D., and Bustamante, C. (2000). Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. Nature *404*, 103-106.
Yang, X., Smidansky, E.D., Maksimchuk, K.R., Lum, D., Welch, J.L., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2012). Motif D of Viral RNA-Dependent RNA Polymerases Determines Efficiency and Fidelity of Nucleotide Addition. Structure.
Yu, Z., Dulin, D., Cnossen, J., Kober, M., van Oene, M.M., Ordu, O., Berghuis, B.A., Hensgens, T., Lipfert, J., and Dekker, N.H. (2014). A force calibration standard for magnetic tweezers. The Review of scientific instruments *85*, 123114.
Zamft, B., Bintu, L., Ishibashi, T., and Bustamante, C. (2012). Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. Proceedings of the National Academy of Sciences of the United States of America *109*, 8948-8953.

Papers of special interest

Chemla, Y.R., Aathavan, K., Michaelis, J., Grimes, S., Jardine, P.J., Anderson, D.L., and Bustamante, C. (2005). Mechanism of force generation of a viral DNA packaging motor. Cell *122*, 683-692.
Early demonstration of high spatiotemporal resolution optical tweezers

Cnossen, J.P., Dulin, D., and Dekker, N.H. (2014). An optimized software framework for real-time, high-throughput tracking of spherical beads. Rev Sci Instrum *85*, 103712.
First CUDA based online tracking software capable of tracking hundreds of beads simultaneously

Depken, M., Galburt, E.A., and Grill, S.W. (2009). The origin of short transcriptional pauses. Biophysical journal *96*, 2189-2193.
Use of dwell-time analysis to analyze the frequency of backtracking pauses

Huhle, A., Klaue, D., Brutzer, H., Daldrop, P., Joo, S., Otto, O., Keyser, U.F., and Seidel, R. (2015). Camera-based three-dimensional real-time particle tracking at kHz rates and Angstrom accuracy. Nature communications *6*, 5885.
First high resolution magnetic tweezers, opening up the possibility of multiplexing data acquisition at high resolution

Larson, M.H., Mooney, R.A., Peters, J.M., Windgassen, T., Nayak, D., Gross, C.A., Block, S.M., Greenleaf, W.J., Landick, R., and Weissman, J.S. (2014). A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science (New York, NY *344*, 1042-1047.
First combination of single-molecule optical tweezers with new generation sequencing

Moffitt, J.R., Chemla, Y.R., and Bustamante, C. (2010). Methods in statistical kinetics. Methods in enzymology *475*, 221-257.
Summary of the analysis approaches for on-pathway reactions illustrated with single-molecule optical tweezers data on phi29 bacteriophage packaging motors.

Papers of outstanding interest

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature *438*, 460-465.
First demonstration of high spatiotemporal resolution optical tweezers

Dulin, D., Vilfan, I.D., Berghuis, B.A., Hage, S., Bamford, D.H., Poranen, M.M., Depken, M., and Dekker, N.H. (2015). Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. Cell reports.
First application of highly multiplexed magnetic tweezers together with dwell time analysis to single molecule enzyme kinetics

Berghuis, B.A., Dulin, D., Xu, Z.Q., van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N.E., Depken, M., and Dekker, N.H. (2015). Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. Nature chemical biology *11*, 579-585.
First application of highly multiplexed magnetic tweezers together with dwell time analysis to characterize highly stable protein-DNA complex

Figure captions

Figure 1. Recent developments in high-throughput single-molecule force spectroscopy methods (a) A flow-stretch experiment where drag forces are applied to a DNA molecule by a sustained fluid flow, and the progression of leading strand replication is monitored through the length of the DNA tether between a bead and the surface. With this method it has been possible to observe e.g. primase induced pausing in lagging strand DNA synthesis (Lee et al., 2006) and coordination and exchange of DNA polymerases (Geertsema et al., 2014). Figure panel adapted from (Duderstadt et al., 2014) **(b)** In a nanophotonic standing-wave array trap assay, interference between two counter-propagating lasers beams produces a standing wave inside a wave guide and a linear array of optical traps in the juxtaposed flow chamber. By placing two waveguides in close proximity, it is possible to simultaneously trap and exert tension on an array of 'dumb-bell' geometries. **(c)** In the acoustic force spectroscopy assay, a piezo plate on the flow cell vibrates at MHz frequency, generating an acoustic wave that homogeneously pushes on beads tethered to the top surface of the flow-cell. The field of view is imaged on a CMOS camera, where thousands of beads can be followed with ~ 5 nm resolution. Figure panel adapted from (Sitters et al., 2015). **(d)** A magnetic tweezers experiment that uses a pair of magnets to apply a force or torque to a tethered molecule. The magnetic tweezers assay is readily parallelized, as shown in the right image of hundreds of simultaneously tracked magnetic beads (Cnossen et al., 2014) in a homogenous force field (De Vlaminck et al., 2012; Lipfert et al., 2009; Yu et al., 2014).

Figure 2. High-throughput single-molecule assays and the resulting multiplexed datasets. (a) In the Tus-*Ter* experimental assay, the DNA-protein lock is formed by force-induced DNA hairpin strand separation. The interaction strength is quantified by measuring the lifetime of the lock at a constant force. **(b)** This yields a dataset of dwell-times until rupture events (i.e. full opening of the hairpin). The resulting dwell-time distribution shows evidence multiple exponential states, revealing the intermediate steps towards lock formation (inset). **(c)** A pair of magnets is used to apply a force to a double-stranded RNA molecule while an RdRp transcribes the RNA. As progression of the fork converts double- to single-stranded RNA in the tether, transcription can be followed by monitoring the lateral position of the bead. Figure panel adapted from (Dulin et al., 2015). **(d)** Three traces of the transcription activity of an RdRp (blue points) acquired during the same experiment at an applied force of 20 pN and 1 mM, filtered with a 0.5 Hz low pass filter (black).

Figure 3. Dwell-time distributions for translocating systems (a) Position vs. time trace generated through a Gillespie simulation of a stochastic stepper with a simple pause (reaction scheme in inset of panel (c)). The times it takes to cross equal-sized dwell-time windows (t_1, t_2, t_3, ...) are scored as dwell-times. **(b)** The distribution of dwell-times generated from the translocation scheme sketched in the inset is given by the Gamma distribution. Keeping the average time τ it takes to cross the dwell-time window constant, the width of the distribution decreases with the number of steps needed to cross the window. **(c)** For a translocation scheme with a rare and long off-pathway pauses (reaction scheme in the inset), the dwell-time distribution will for short times follow a Gamma distribution centered on the average time it takes to translocate through a dwell-time window by n non-pausing steps ($n/(k+k_p)$). At long dwell-times, the distribution will be dominated by the exponentially distributed pause-escape times (with the average $1/k_e$). The probabilistic weight of the Gamma and exponential-pause regions can be used to calculate the pause frequency. The basic shape of the dwell-time distribution of on-pathway pauses is the same, while its dependence on e.g. nucleotide concentrations differ (see main text). **(d)**

For a translocation scheme including rare backtracks (reaction scheme in the inset), we again have an initial pause-free region that follows a Gamma distribution. At later times, there is an algebraically distributed region decaying with the power -3/2, and bounded by the timescales $t_0=(k_{in}\ k_{out})^{-1/2}$ and $t_1=(k_{in}^{1/2}-\ k_{out}^{1/2})^{-1/2}$ (Depken et al., 2009). The red dots in panel (c) and (d) are generated by Gillespie simulations of the translocation schemes shown in the insets in the respective panels.

Figure 4. Probability density distribution of the dwell times as a function of the number of P2 RdRp activity traces. In the leftmost panel, the dwell-time distribution is composed of more than 1000 dwell times derived from 5 independent traces using a 10 bp dwell-time window. In this case, the short-time Gamma distribution, which represents around 95% of the events, can be clearly identified. In the center and right panels, the number of independent traces is increased to 20 (yellow data), and 45 (red data), respectively. Here, one can readily observe new features at long dwell times with increased statistics. These detailed probability density distributions can be described by a minimal composite model with two exponential pauses (dominating dwell times between 1 and 20 s), and a power law originating in a backtracked pause (dominating dwell times above 20 s) (Dulin et al., 2015). In all three panels, the dwell times are log binned, and the error bars on each bin represent the standard deviation as calculated from 1000 bootstrapped data sets.

Figures

Figure 1

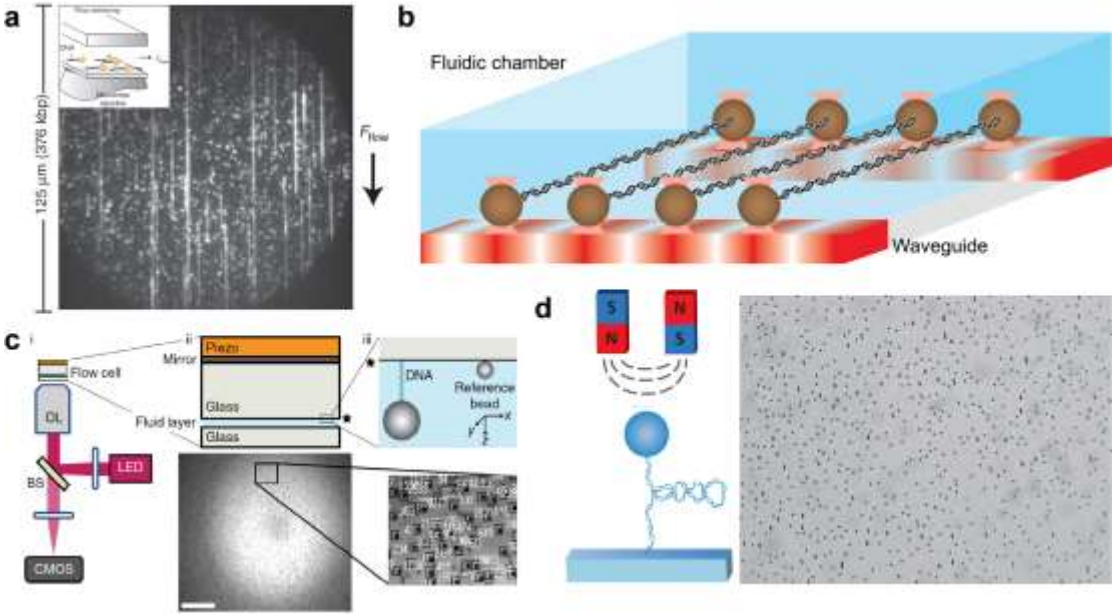


Figure 2

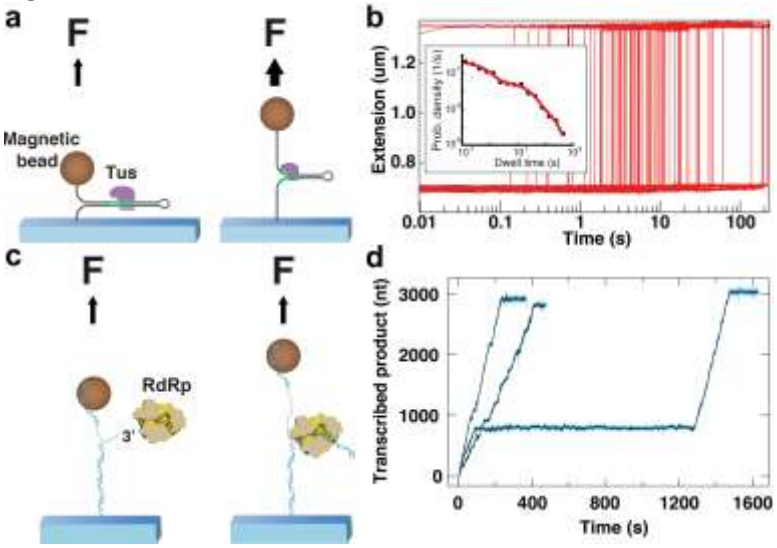

a
F
F
Magnetic bead
Tus
b
Extension (um)
1.2
1.0
0.8
0.01
0.1
1
10
100
Time (s)
c
F
F
RdRp
3'
d
Transcribed product (nt)
3000
2000
1000
0
0
400
800
1200
1600
Time (s)

Figure 3

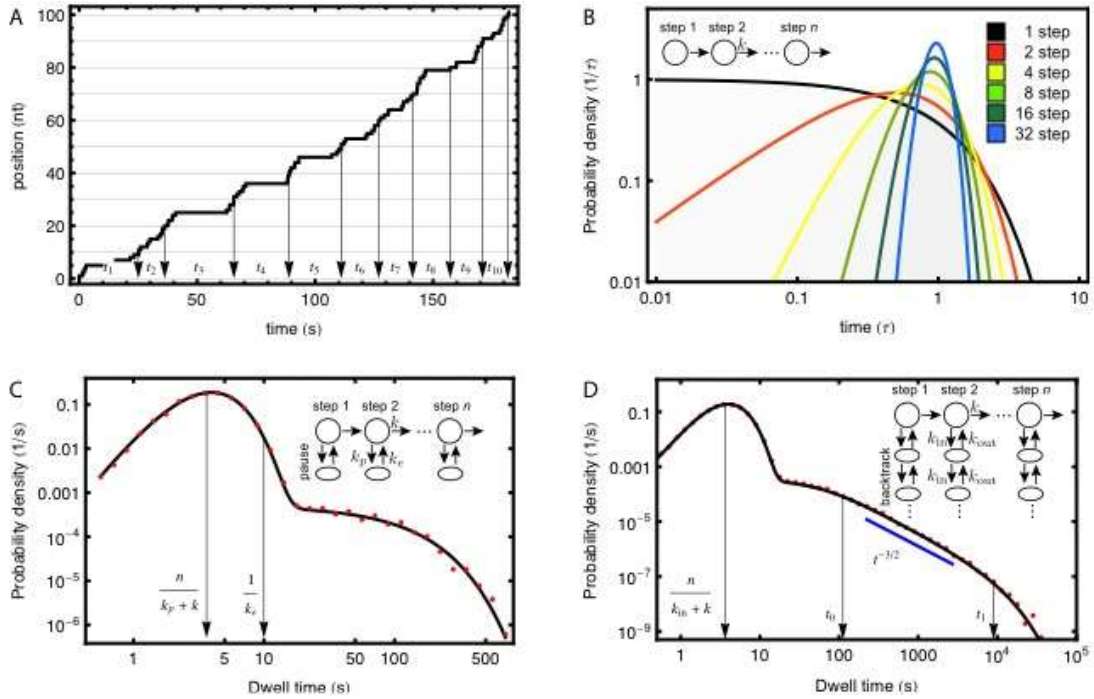
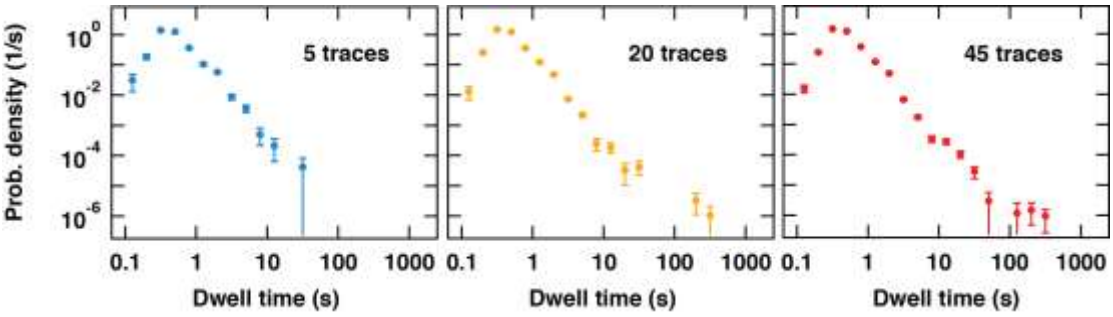

A
position (nt)
time (s)
B
step 1
step 2
step n
1 step
2 step
4 step
8 step
16 step
32 step
Probability density (1/τ)
time (τ)
C
Probability density (1/s)
pause
Dwell time (s)
D
backtrack
Probability density (1/s)
Dwell time (s)

Figure 4


Prob. density (1/s)
10^0
10^-2
10^-4
10^-6
5 traces
20 traces
45 traces
0.1
1
10
100
1000
Dwell time (s)