# Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA

Y. M. Wang,[1] Robert H. Austin,[1] and Edward C. Cox[2]
[1]Department of Physics, Princeton University, Princeton, New Jersey 08544, USA
[2]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA
(Received 10 October 2005; published 27 July 2006)

We used single-molecule imaging techniques and measured the one-dimensional diffusion of LacI repressor proteins along elongated DNA to address the long-standing puzzle of why some proteins find their targets faster than allowed by 3D diffusion. Our analysis of the LacI transcription factor's diffusion yielded four main results: (1) LacI diffuses along nonspecific sequences of DNA in the form of 1D Brownian motion; (2) the observed 1D diffusion coefficients $D_1$ vary over an unexpectedly large range, from $2.3 \times 10^{-12}$ cm$^2$/s to $1.3 \times 10^{-9}$ cm$^2$/s; (3) the lengths of DNA covered by these 1D diffusions vary from 120 nm to 2920 nm; and (4) the mean values of $D_1$ and the diffusional lengths indeed predict a LacI target binding rate 90 times faster than the 3D diffusion limit.

DOI: 10.1103/PhysRevLett.97.048302 PACS numbers: 82.37.Np, 87.15.Rn, 87.15.Vv

A puzzle in gene expression has been the faster-than-diffusion binding of some proteins to their specific DNA targets [1–3]. Faster-than-diffusion binding was first observed in 1971 with LacI repressor, which binds to its *lacO* target 100 times faster than the 3D diffusion limit [1]. Many models have been proposed to explain this phenomenon [4–9]; however, ensemble measurements have been insufficient to provide conclusive evidence to support any particular model [10–14]. We used a single-molecule method approach to study LacI-DNA binding, and observed the nonspecifically bound LacI proteins performing 1D Brownian motion along DNA.

The expected association rate $k_{a(3D)}$ by which DNA-binding proteins find their specific target sequences on double-stranded DNA in a random 3D search is $4\pi D_3 l_{\text{seq}}$ per unit protein concentration, where $l_{\text{seq}}$ is the effective DNA target length, and $D_3 \doteq k_B T/3\pi\eta a = 9 \times 10^{-7}$ cm$^2$/s is the 3D diffusion coefficient of the protein in solution [1,9,15], where $k_B$ is the Boltzmann constant, $T$ is the temperature, $\eta$ is the viscosity of the solvent, and $a \approx 5$ nm is the typical diameter of the protein. With $l_{\text{seq}} \approx 3$ bp (or 1 nm), the protein-DNA association rate $k_{a(3D)}$ should be $10^8$/M/s. The original *in vitro* study on LacI-*lacO* binding by Riggs *et. al.* was with 45.5 kbp DNA of 15.5 $\mu$m in length, and the *lacO* association rate $k_{a(\text{exp})}$ was measured to be $10^{10}$/M/s, 100 times higher than the diffusion limit of $k_{a(3D)} \approx 10^8$/M/s [1] [the $10^{10}$/M/s binding rate was also reported in Refs. [9,12,16] ].

It has been proposed that such high rates can be achieved if the protein undergoes a combination of 1D diffusion along the DNA and 3D diffusion in solution, a process called facilitated diffusion. In this model, the key to faster targeting lies with the nonspecific DNA sequences that flank the target site. By 3D diffusion, a protein most likely will run into a segment of nonspecific DNA first. After nonspecific binding, the protein will diffuse along the DNA for a certain time and eventually dissociate. By doing so, the effective concentration of protein near the DNA increases, and thus the targeting rate should change. This facilitated-diffusion modified protein-target association rate $k_a$ per protein concentration has been derived by Halford and Marko [9]:

$$k_a = \left(\frac{1}{D_3 l_d} + \frac{L l_d c}{D_1}\right)^{-1} = k_{a,3\text{D}}\left(\frac{l_{seq}}{l_d} + \frac{D_3}{D_1} l_{\text{seq}} L l_d c\right)^{-1}, \quad (1)$$

where $D_1$ is the 1D diffusion coefficient of the nonspecifically bound protein along the DNA, $L$ is the total length of the DNA molecule, $l_d$ is the maximum DNA contour distance $x_{\max} - x_{\min}$ covered by the protein before dissociation, $c$ is the concentration of the target, and $(\frac{l_{\text{seq}}}{l_d} + \frac{D_3}{D_1} l_{\text{seq}} L l_d c)^{-1}$ is the acceleration factor to $k_{a,3\text{D}} = D_3 l_{\text{seq}}$. In order to check the facilitated-diffusion model directly it is necessary to know $D_1$ and $l_d$, which can only be obtained by imaging protein-DNA-binding dynamics using single-molecule measurements. In fact, if these values do not fall within a certain range, ''facilitated'' diffusion can actually slow the search times.

We used a LacI fusion-protein consisting of a green fluorescent protein (GFP)-GFP13 (S65T):*lacI-112* fusion (GFP-LacI), and stained the DNA with the dimeric cyanine dye BOBO-3. DNA constructs of Lambda Zap vector with 256 tandem copies of *lacO* ($lacO_{256}$) were used. $LacO_{256}$-DNA was 42.06 kbp long with a contour length of 14.3 $\mu$m, and the 9.22 kbp $lacO_{256}$ insertion started at 24.02 kbp. The nonspecific sequences of the DNA construct are identical to that of $\lambda$ DNA. The synthesis methods for the fusion protein and the $lacO_{256}$-DNA, and the sample preparation method are described in Ref. [17]. There were $lacO_{256}$-DNA dimers as well as monomers in the solution; the dimers were formed by the sticky-end-hybridization of two $lacO_{256}$-DNA monomers Fig. 1(a). After the LacI-DNA and BOBO-3 incubation, the GFP-LacI concentration was 50 nM and the $lacO_{256}$-DNA concentration was 11 pM (0.3 $\mu$g/ml). The DNA intercalating cyanine dyes are known to stretch DNA by 30% at 1 dye/5 bp [18], so at our concentration of

 

1 dye/10 bp, the DNA molecules were stretched by 15% to 16 $\mu$m. Since BOBO-3 produced no obvious effect on the DNA-configuration-dependent LacI-DNA specific binding [17], we expect that its effect on the less DNA-configuration-dependent LacI nonspecific binding [19] to be negligible. A catalytic oxygen scavenging solution was used to maximize dye lifetimes [17]. 1 $\mu$l of the DNA + LacI solution and 4 $\mu$l of the oxygen scavenging solution were deposited onto a fused-silica chip.

A glass cover slip was used to flatten the solvent and the edges of the cover slip were then sealed with nail polish. As the cover slip flattened the droplet, hydrodynamic flow elongated the DNA dimers, and the two LacI-$lacO_{256}$ sites stuck to the surface, creating an anchored elongated DNA molecule [Fig. 1(b) and 1(c)] stretched up to 90% of its native contour length. The tension on DNA was a few pico Newtons [20]. DNA was not observed to stick to fused-silica surfaces at our $p$H of 8.0 and bovine serum albumin concentrations, and the elongated DNA molecules were effectively suspended from the surface as evidenced by the DNA's transverse motion of $\pm 50$ nm (data not shown). Thus unbound GFP-LacI molecules interacted only with free unattached and nonspecific DNA. Note that the sticking of GFP-LacI to fused-silica surfaces occurred only at the deposition step as the air-water interface moved over the chip surface. After the cover slip was sealed the free GFP-LacI molecules ($\approx 2$ nM) diffused in the solution freely and did not stick to the surface as evidenced by observation of the freely diffusing GFP-LacI near the surface (data not shown).

The single-molecule experiments were performed using a prism-type Total Internal Reflection Fluorescence Microscopy (TIRFM) method. The laser excitation was synchronized to the 3.4 Hz data acquisition rate of the I-CCD camera. The emitted photons from BOBO-3 and GFP were collected using a 100X total internal reflection (TIRF) oil-immersion objective (NA = 1.45), went through a custom-designed dichroic mirror and emission filter set (Chroma Technology Corp, Rockingham, VT), and were recorded by an I-CCD camera (I-PentaMAX:HQ Gen III, Princeton Instruments, Trenton, NJ). The point spread function (PSF) width of the diffusion limited optical system was measured to be 280 nm, and the imaging pixel size was 117 nm. The pixel count of the camera was converted to a photon count using known conversion factors [17]. The mean 488 nm illumination intensity over the illumination areas of 30 $\mu$m $\times$ 50 $\mu$m was 1000 W/cm$^2$. The centroid location of a GFP-LacI dot was determined by fitting its 2D fluorescence intensity profile to a Gaussian. The number of detected photons per PSF per frame (between 50 and 300 photons) limited the position measurement accuracy to be between 10 nm and 50 nm [21].

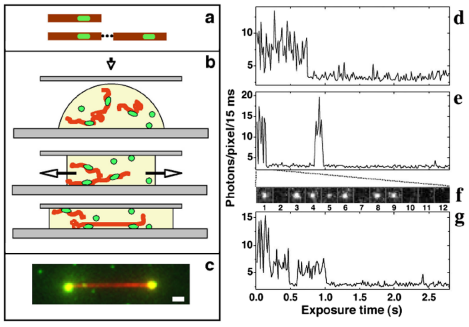


FIG. 1 (color online). (a) Schematics of a GFP-LacI (green) bound $lacO_{256}$-DNA monomer and dimer (red). (b) Elongation of the DNA. (c) Frame averaged superposed image of GFP-LacI bound to an elongated $lacO_{256}$-DNA dimer. The scale bar is 1 $\mu$m. (d) A GFP-LacI monomer of frequent blinking and unitary bleaching. (e) A GFP-LacI monomer that blinked, recovered the first bleaching in 3 s, and finally irreversibly bleached. (f) The GFP-LacI dots for the first 12 frames of (e), showing blinking at frames 2 and 7, and bleaching at frame 10. (g) A GFP-LacI dimer with two bleaching events.

Knowledge of the fluorescence characteristics of single free GFP-LacI monomers and dimers attached to fused-silica surfaces is essential in justifying the single-molecule nature of a bound protein. GFP-LacI monomers blink frequently (short fluorescence dips to near instrumental noise level), and bleach with no recovery [Fig. 1(d)]. At our excitation intensity of 1000 W/cm$^2$, mean exposure time of 10 ms, and synchronized imaging frequency of 3.4 Hz, the mean net observation time of each GFP-LacI molecule was 5 s before it bleached (giving a total laser exposure time of 0.15 s). The mean number of photons emitted by the bound GFP-LacI molecules before bleaching was $\approx 4 \times 10^4$ photons. This 5 s observation time gave the *instrumental* limit to the maximum mean distance we observed GFP-LacI motion on DNA in this experiment.

An image sequence of a single GFP-LacI molecule diffusing along DNA is shown in Fig. 2(b). This is 1 out of 70 walks that were observed, and chosen for its large net displacement. Figure 2(a) shows the frame averaged superposed image of the anchored DNA and the diffusing GFP-LacI on DNA. Time-lapse images of the diffusing protein show clear relative displacements [Fig. 2(b)], with one immobile anchoring site used as a reference point. We know that we were observing a single GFP-LacI dimer from the fluorescence time trace in Fig. 2(d), which clearly shows two bleaching steps. Both GFP-LacI monomers (80%) and dimers (20%) have been observed to diffuse on DNA. As is evident in Fig. 2(d), fluorescence time traces of bound GFP-LacI molecules were identical to that of single immobile GFP-LacI [Fig. 1(d), 1(e), and 1(g)], with the same blinking rate and characteristic bleaching time of $\approx 0.15$ s (5 s net observation time). The DNA locations of the diffusing protein at different frames are correlated and localized, thus at $D_3 = 10^8$ nm$^2$/s and our protein concentration of a few proteins/$\mu$m$^3$, the chance for two different proteins to consecutively land on the same location of DNA is 1 in 1000. Figure 2(e) plots the distribution of all relative displacements $x_i - x_{i-1}$ of the walk. This is a Gaussian of

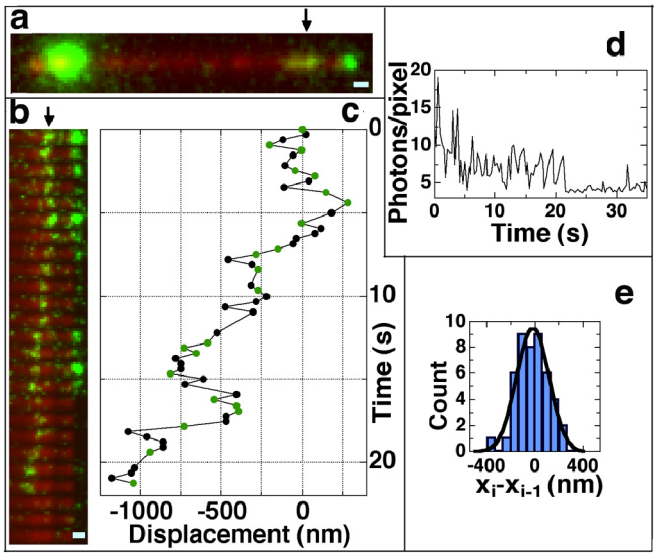


FIG. 2 (color online). (a) Frame averaged, superposed image of a GFP-LacI molecule diffusing along DNA. The two large dots at the DNA ends are LacI-$lacO_{256}$ sites, and the green segment on the nonspecific DNA (arrow) is the trace of the diffusing GFP-LacI. (b) Image series of the diffusing protein (arrow) of selected clear relative displacements corresponding to green dots in (c), the displacement vs $t$ curve of the diffusing protein. (d) Fluorescence time trace of the diffusing GFP-LacI. It is a dimer. (e) Gaussian distribution of $x_i - x_{i-1}$. The scale bar is 0.5 $\mu$m.

SD = 130 nm centered near zero, typical for Brownian motion with limited data points.

Now we discuss our analysis showing that individual protein diffusion trajectories, which consist of multiple measurements $x_i$ until the protein disassociates, are Brownian in nature, and we obtain 1D diffusion constants $D_1$ for these trajectories. Qian *et. al.* have derived an expression to tell (1) whether a single diffusion trajectory is Brownian and if so (2) obtain the diffusion constant of the trajectory [22]. This method calculates the mean square displacement $\mathrm{MSD}_{(n,N)}$ for all available time intervals of a single diffusion trajectory

$$\mathrm{MSD}_{(n,N)} = \frac{\sum_{i=1}^{N-n}(x_{i+n} - x_i)^2}{N-n} = 2D_1 n\Delta t + 2\sigma_s^2, \quad (2)$$

where $N$ is the total number of positions measured, $n$ is the measurements index going from 1 to $N$, $\Delta t$ is the time interval between two consecutive position measurements, and $\sigma_s$ is the measurement accuracy associated with each $x_i$. We can obtain $D_1$ of a single diffusion trajectory from its $\mathrm{MSD}_{(n,N)}$ to high precision by weighted linear-fitting $\mathrm{MSD}_{(n,N)}$ to $n$, taking $\mathrm{MSD}_{(n,N)}$'s variances at different $n$ into consideration. As $n$ increases, the number of available measurement points for $\mathrm{MSD}_{(n,N)}$ averaging decreases, and the variance in $\mathrm{MSD}_{(n,N)}$ increases as

$$\sigma_{n,N}^2 = (2D_1 n\Delta t)^2(2n^2+1)/[3n(N-n+1)]. \quad (3)$$

If a single trajectory is Brownian, then its $\mathrm{MSD}_{(n,N)}$ at $n$ below a cutoff $n_c$ will be a linear function of $t$, with $n_c$ determined by a set fractional $\mathrm{MSD}_{(n,N)}$ uncertainty in Eq. (3). We chose $n_c$ to be where $\sigma_{n,N}/(2D_1 n\Delta t)$ is 50%. We plot $\mathrm{MSD}_{(n,N)}$ versus $n$ for trajectories with $N > 10$, where there are at least three $\mathrm{MSD}_{(n,N)}$ values whose fractional variances are $<50\%$ [Eq. (3)]. We also used only trajectories with fewer than 5 contiguous GFP blinks. Since $\sigma_s < 50$ nm, $\mathrm{MSD}_{(n,N)}$, which is the square of the difference of two position measurements, has an offset of $2500\ \mathrm{nm}^2 < 2\sigma_s^2 < 5000\ \mathrm{nm}^2$. These photon noise offsets were subtracted in the $\mathrm{MSD}_{(n,N)}$ versus $n$ curves.

Figure 3(a) plots displacement $x$ versus time for 70 trajectories. The 15 trajectories in color are the walks for which we have obtained $D_1$, and the center black line is a stationary GFP-LacI stuck to the fused-silica surface (not DNA). Figure 3(c) plots $\mathrm{MSD}_{(n,N)}$ versus $n$ for these 15 trajectories in linear scale and Fig. 3(d) in log-log scale at low $n$ values, respectively. The log-log plots are all straight lines with the slope of 1 at low $n$, clearly indicating that the 1D trajectories are Brownian motions. The dashed line in Fig. 3(d) is a fit of Eq. (2) with weighted error [Eq. (3)] to all $n$ points below $n_c$ of the topmost trajectory. The intercepts at $n = 1$ are $2D_1\Delta t$ for each particular walk, as can be seen by inspection of Eq. (2) [Fig. 3(d)]. Thus, while all the walks are Brownian in nature, the different intercepts at $n = 1$ indicate that there is a large distribution in diffusion coefficients and there is not a unique, single value for $D_1$. We also plotted the distributions of nondegenerate relative displacements $x_i - x_{i-n}$ for the first 15 positions of all 70 trajectories for $n = 1$, 2, and 3 in Fig. 3(b); the displacements are all Gaussians centered at zero with SD increasing with $n$. This result further proves that LacI's diffusion trajectories are truly Brownian in nature, regardless of the variations in individual diffusion coefficients. A recent paper (published after this Letter was submitted) [23] also sees a similar large distribution in 1D diffusion coefficients for Rad51 on aligned DNA molecules, so our result here may be of some generality.

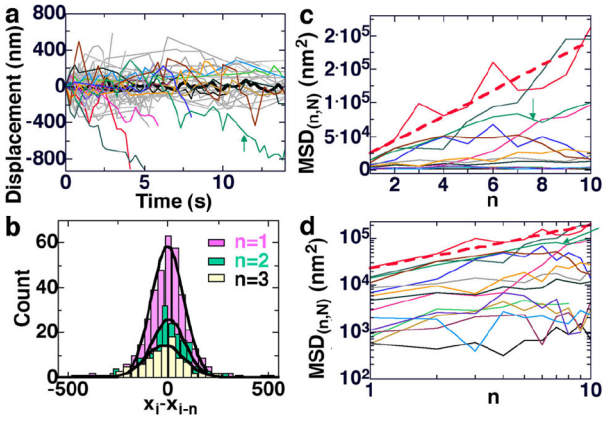


FIG. 3 (color online). (a) $x$ vs $t$ for 70 trajectories. (b) Nondegenerate $x_i - x_{i-n}$ distributions. (c) $\mathrm{MSD}_{(n,N)}$ vs $n$ for the 15 colored walks in linear scale and (d) in log-log scale. The arrows in (a), (c), and (d) denote the walk in Fig. 2. The dashed lines in (c) and (d) are the fit of Eq. (2) to the top trajectory.

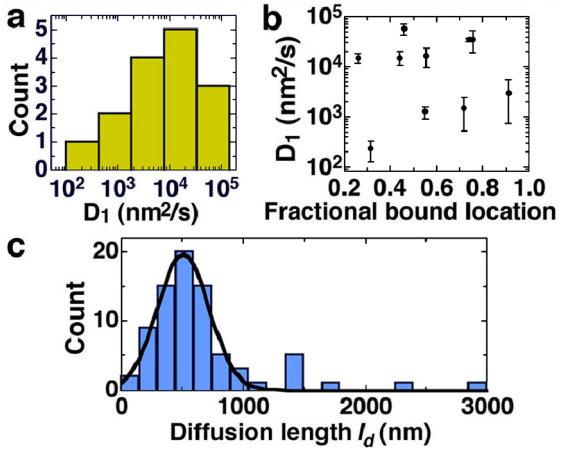


FIG. 4 (color online). (a) $D_1$ distribution of the 15 trajectories. (b) $D_1$ vs fractional bound location on the nonspecific segment of the $LacO_{256}$-DNA dimer. The error bars were obtained from the fit of $\mathrm{MSD}_{(n,N)}$ to $n$ with weighted errors at each $n$ given by Eq. (3). (c) Histogram of $x_{\max} - x_{\min} = l_d$ for the 70 trajectories. The solid line is a Gaussian fit with a mean of $500 \pm 220$ nm (mean $\pm$ SD). Values in (a) and (c) have been adjusted to DNA contour length.

Figure 4(a) shows the distribution of the 15 $D_1$ values (corrected to DNA contour length), which span a large range from $2.3 \times 10^2$ nm$^2$/s to $1.3 \times 10^5$ nm$^2$/s. Figure 4(b) shows that the different $D_1$ values are distributed randomly along the DNA, showing a lack of correlation between the $D_1$ and the regions on the $\lambda$ DNA on which the protein has diffused. Figure 4(c) shows the distribution of the $l_d$ in DNA contour length. Because $\lambda$ DNA has large sequence variance in the nonspecific region with $\pm 30\%$ difference in adenine-thymine and cytosine-guanine content, it is possible that the diffusion constants are a function of local sequence. It is also possible that the large distribution in $D_1$ is caused by conformational distributions in the protein [24]. Further experiments are needed to answer these questions.

Finally, we use our data to examine the question of the extent to which facilitated diffusion can enhance the LacI target binding rate, the biological purpose behind these measurements. Just as there is a distribution in the 1D diffusion coefficient $D_1$, there is also a distribution in the diffusion lengths $l_d$, whose value is further compromised by the mean observation time to bleaching of the GFP of 5 s. Since the final target binding is the result of many diffusion events on nonspecific DNA, we use the mean $\langle l_d \rangle$ (probably a lower bound due to bleaching) of 500 nm, and the mean diffusion coefficient of $\langle D_1 \rangle = 2.1 \times 10^{-10}$ cm$^2$/s in Eq. (1). Also using Riggs' concentration of 1 $lacO/1670\ \mu\mathrm{m}^3$, $D_3 \approx 4 \times 10^{-7}$ cm$^2$/s for LacI tetramers (a $\approx 10$ nm), and $L = 15.5\ \mu$m, the accelerating factor in Eq. (1) is $93 \pm 20$, and thus solving the 100-fold discrepancy between the theory and the experimental data. We conclude from these measurements that facilitated diffusion increases the LacI-$lacO$ binding rate well over the apparent diffusion limit. This result demonstrates that facilitated diffusion in the form of 1D Brownian motion is the mechanism responsible for the faster-than-diffusion binding of LacI to $lacO$, and quite possibly, the reason also for the observed faster-than-diffusion binding in other protein-DNA interactions.

We thank Xiao-Juan Guan and Ling Guo for sample synthesis, Jonas Tegenfeldt for design of the prism TIRF, and Ido Golding, Walter Reisner, Robert Riehn, Eric Siggia, and Henrik Flyvbjerg for discussions, and NIH Grant No. HG001506 for support.

[1] A. D. Riggs, S. Bougeois, and M. Cohn, J. Mol. Biol. **53**, 401 (1970).
[2] R. Wallis, G. R. Moore, R. James, and C. Kleanthous, Biochemistry **34**, 13 743 (1995).
[3] G. M. Dhavan, D. M. Crothers, M. R. Chance, and M. Brenowitz, J. Mol. Biol. **315**, 1027 (2002).
[4] G. Adam and M. Delbruck, *Structural Chemistry in Molecular Biology* (Freeman, San Francisco, 1968), pp. 198–215.
[5] O. G. Berg and C. Blomberg, Biophys. Chem. **4**, 367 (1976).
[6] O. G. Berg, Chem. Phys. **31**, 47 (1978).
[7] H. Kabata, O. Kurosawa, I. Arai, M. Washizu, S. A. Margarson, R. E. Glass, and N. Shimamoto, Science **262**, 1561 (1993).
[8] N. Shimamoto, J. Biol. Chem. **274**, 15 293 (1999).
[9] S. E. Halford and J. F. Marko, Nucleic Acids Res. **32**, 3040 (2004).
[10] R. B. Winter and P. H. von Hippel, Biochemistry **20**, 6948 (1981).
[11] M. Ricchetti, W. Metzger, and H. Heumann, Proc. Natl. Acad. Sci. U.S.A. **85**, 4610 (1988).
[12] M. Hsien and M. Brenowitz, J. Biol. Chem. **272**, 22 092 (1997).
[13] T. Misteli, Science **291**, 843 (2001).
[14] D. M. Gowers and S. E. Halford, EMBO J. **22**, 1410 (2003).
[15] O. G. Berg and P. H. vonHippel, Annu. Rev. Biophys. Biophys. Chem. **14**, 131 (1985).
[16] M. Barkley, Biochemistry **20**, 3833 (1981).
[17] Y. M. Wang, J. Tegenfeldt, W. Reisner, R. Riehn, X.-J. Guan, L. Guo, I. Golding, E. C. Cox, J. Sturm, and R. H. Austin, Proc. Natl. Acad. Sci. U.S.A. **102**, 9796 (2005).
[18] T. Perkins, D. E. Smith, R. G. Larson, and S. Chu, Science **268**, 83 (1995).
[19] C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, and R. Kaptein, Science **305**, 386 (2004).
[20] S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).
[21] R. E. Thompson, D. R. Larson, and W. W. Webb, Biophys. J. **82**, 2775 (2002).
[22] H. Qian, M. P. Sheetz, and E. L. Elson, Biophys. J. **60**, 910 (1991).
[23] A. Graneli, C. Yeykal, R. Robertson, and E. Greene, Proc. Natl. Acad. Sci. U.S.A. **103**, 1221 (2006).
[24] R. H. Austin, K. Beeson, L. Eisenstein, H. Frauenfelder, I. Gunsalus, and V. Marshall, Phys. Rev. Lett. **32**, 403 (1974).