Dynamic Force Spectroscopy of Protein-DNA Interactions by Unzipping DNA

Steven J. Koch and Michelle D. Wang*

Department of Physics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA
(Received 18 December 2002; published 10 July 2003)

We demonstrate the first site-specific single-molecule characterization of the prominent activation barrier for the disruption of a protein-DNA binding complex. We achieved this new capability by combining dynamic force spectroscopy with unzipping force analysis of protein association and used the combination to investigate restriction enzyme binding to specific DNA sites. Analysis revealed lifetimes and interaction distances for three protein-DNA interactions. This new method is able to distinguish protein-DNA binding complexes on a site-specific, single-molecule basis.

Protein-DNA interactions are central to many major cellular processes, including transcription, replication, and packaging of DNA into chromatin. Indeed, recent sequencing of the human genome shows that $\sim15\%$ of the $\sim30\,000$ genes encode proteins which bind to nucleic acids [1]. Basic parameters of protein-DNA interactions include binding location (sequence specificity), binding affinity (equilibrium association constant), binding rate constants (on and off rates), and in some cases catalytic rate constants. While traditional biochemical (bulk) methods have been successful in elucidating some of these parameters, the new single-molecule method that we describe in this Letter provides specialized advantages and in some cases will enable measurements that thus far have been inaccessible.

This method, called unzipping force analysis of protein association (UFAPA), is a novel and versatile method for probing protein-DNA interactions [2](see Fig. 1). In this method, a single DNA double helix is unzipped [3] in the presence of DNA-binding proteins using a feedback-enhanced optical trap. One strand of the DNA is anchored to a microscope coverslip while the other strand is attached to a microsphere held in an optical trap. The DNA is unzipped as the microscope coverslip is moved away from the trapped microsphere. When the unzipping fork in a DNA reaches a bound protein, a dramatic increase in the tension in the DNA, followed by a sudden tension reduction, is detected. Analyses of the unzipping forces and lengths of the DNA tether reveal the locations of the bound proteins and the equilibrium association constants. In this Letter we add essential features to UFAPA by quantitative application of dynamic force spectroscopy (DFS) in order to determine the lifetimes (t_{off}) and characteristic interaction distances (d) of the prominent activation barriers of site-specific protein-DNA interactions. We also show that analysis of the disruption forces allows us to distinguish between different protein-DNA complexes. To demonstrate these ideas, we used this approach to study the binding of restriction enzymes to their recognition sites.

DFS is a powerful method for mechanical investigation of the interactions of biomolecules. The basic idea behind DFS is straightforward. A bound molecule will unbind under thermal agitation if given sufficient time. However, it is often not experimentally feasible to wait for unbinding in cases where the lifetime (t_{off}) of the bound state is exceedingly long. To circumvent this, DFS reduces the lifetime by tilting the energy landscape with an external force (F) to encourage unbinding. Analysis of constant force lifetimes allows determination of the lifetime under no applied force. While this is the simplest approach, the lifetime can also be determined by using a method in which the applied force is increased at a constant rate ($r = dF/dt$) to encourage unbinding [5]. In this method, as activation barriers are lowered, the unbinding force distribution from many measurements gives a measure of the natural lifetime. The unbinding force probability density function (PDF) is well defined if there is one predominant activation barrier for unbinding:

$$p(F;r) = \frac{1}{t_{off}r}\exp\left(\frac{k_BT}{t_{off}rd}\right)\exp\left\{\frac{Fd}{k_BT} - \frac{k_BT}{t_{off}rd}\exp\left(\frac{Fd}{k_BT}\right)\right\},$$

(1)

where k_BT is the thermal energy and d is the distance along the direction of the applied force between the bound state and the activation barrier peak [4]. A series of these distributions at different r values provides a relation between the most-probable unbinding force (F^*) and r:
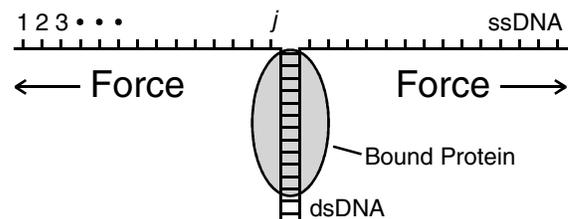

FIG. 1. Schematic of the unzipping configuration (not to scale). Applied tension unzips the two strands of the DNA molecule. The location of the unzipping fork is indicated by an unzipping index j. The presence of a DNA-binding protein at the unzipping fork is observed as an increase in the force required to separate the strands.

$$F^* = \frac{k_B T}{d} \ln\left(\frac{t_{\text{off}} r d}{k_B T}\right). \qquad (2)$$

Therefore, t_{off} and d can be determined either from $p(F; r)$ at a given r [Eq. (1)] or from the F^* vs r relation [Eq. (2)]. If there is truly a single activation barrier, these two methods should yield the same result and the plot of F^* vs $\ln(r)$ should give a straight line for all values of r. A nonlinear plot of F^* vs $\ln(r)$ then indicates the presence of more than one activation barrier, and a crossover in F^* vs $\ln(r)$ from one linear regime to another indicates a transition from one dominant activation barrier to another [4].

In order to incorporate DFS into UFAPA, we designed a novel digital loading rate clamp, $r = dF/dt = $ const, in the optical trapping setup. In this implementation, unzipping proceeded under the control of an algorithm that effectively became a loading rate clamp when the unzipping fork encountered a bound protein [6]. An example of data taken with the loading rate clamp is shown in Fig. 2. We have previously described [2] the DNA construct [7], the optical trapping instrument and calibration methods, the unzipping buffer conditions, and the proteins [8] used in these experiments. Briefly, a single DNA double helix

was unzipped in buffers containing the restriction enzymes (*Bso*BI or *Xho*I) but without Mg^{2+}, so that the enzymes would bind to the DNA without cutting it. For the data shown in Fig. 2, *Bso*BI molecules were bound to a number of sites on a pCP681-derived DNA construct. In Fig. 2(a), the presence of bound complexes is revealed by the prominent peaks seen in this force versus time graph. Stochastic unbinding events occurred when the force reached 25–50 pN and are indicated by the sudden drops in force. As Fig. 2(a) clearly shows, preceding each unbinding event the force increased linearly with time as prescribed by the loading rate clamp. In Fig. 2(b), a force ramp preceding an unbinding in Fig. 2(a) corresponds to a horizontal plateau due to the inhibition of unzipping by the bound protein complex.

As shown in Fig. 2, unbinding events are easily distinguished from the baseline unzipping forces. A novel automated event detection scheme located each event, and determined an event starting force, the disruption force, and the average loading rate during the event [9]. An example of an automatically detected event is highlighted in Fig. 2. At a given protein-DNA complex, under the action of the loading rate clamp, the actual loading rate was estimated from a linear fit of the data [Fig. 2(a)]. The observed force at any time, including Brownian noise, remained within ± 1.5 pN of the force predicted from the fit. The actual loading rates were distributed around the specified value with a standard deviation of $\sim 10\%$ of
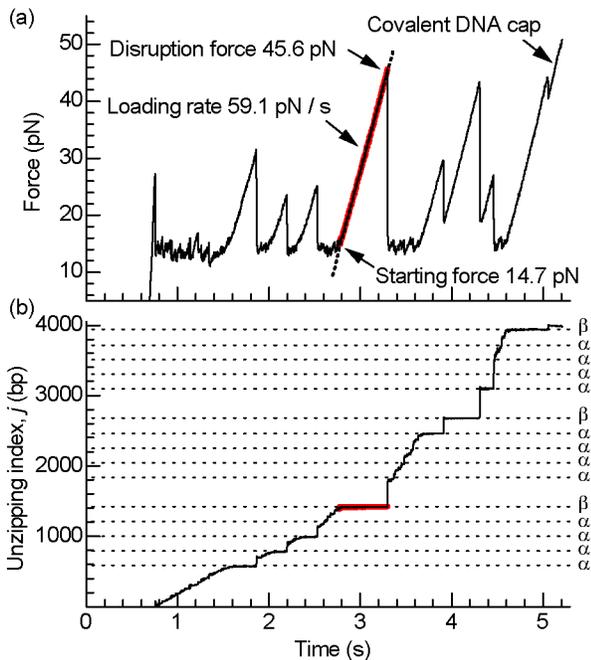

FIG. 2 (color online). Example of unzipping data taken using loading rate clamp (59 pN/s). (a) Force versus time. The graph demonstrates the uniform force loading rate for forces greater than 15 pN. One of the eight events from the automated event detection is *highlighted*. The *dotted line* shows the loading rate fit for the event. (b) Calculated unzipping index, j, vs time. Each horizontal step represents data where a restriction enzyme pins the unzipping index at a certain value until the complex disrupts. The same event as in (a) is highlighted. *Horizontal dashed lines* mark unzipping indices which are predicted *Bso*BI binding sites, of types α (ttcCTCFFFaat) and β (aaaCTCGAGact).

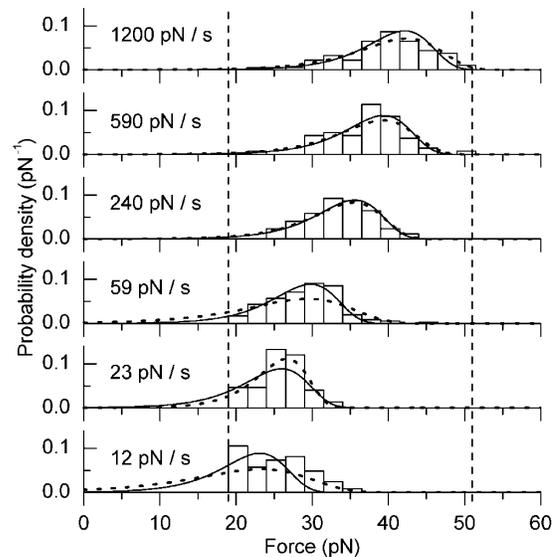

FIG. 3. Dynamic force spectroscopy for *Bso*BI unbinding from α sites ($N = 449$ events total). Each histogram shows unbinding force distribution with bins of 2.5 pN width at a given force loading rate. *Dashed curves* represent predicted force probability density functions resulting from the determination of *local* values of d and t_{off} (see text). *Solid curves* represent predicted force probability density functions resulting from the determination of *global* values of d and t_{off} (also see text). *Vertical dashed lines* designate the experimentally accessible force range of the current implementation of UFAPA.
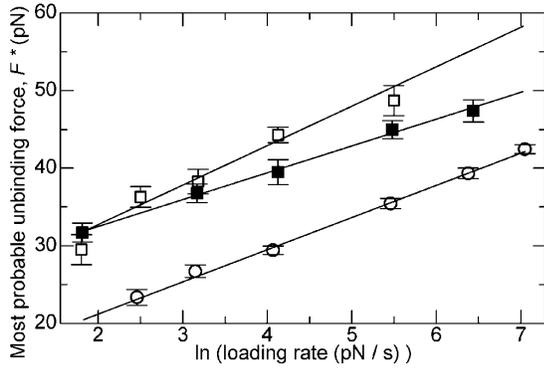
FIG. 4. Dynamic force spectroscopy for three binding species. *Data points* represent the most-probable unbinding force for a given force-loading rate, obtained from the maximum likelihood method described in the text. *Error bars* were determined using a Monte Carlo method [12]. *Open circles* represent *Bso*BI unbinding from α sites, *open squares* represent *Bso*BI unbinding from β sites, and *filled squares* represent *Xho*I unbinding from β sites. *Solid lines* are linear fits of Eq. (2) to the data. Parameters d and t_{off} obtained from the fits are listed in Table I.

the specified value. In the context of DFS, this spread is insignificant, due to the logarithmic relationship between the force-loading rate and the expected unbinding force distribution [see Eqs. (1) and (2)] [4].

In Fig. 3, we show a summary of the unbinding force distributions for *Bso*BI from α sites at various force-loading rates r. For a given r, the local fit values of d and t_{off} were obtained using the maximum-likelihood method [10], based on the assumed form of the PDF [11] [Eq. (1)]. The parameter search was expedited by the use of the Nelder-Mead simplex method. Once the best-fit PDF (a dashed curve in Fig. 3) for a given r was obtained, F^* was calculated analytically from Eq. (2). After repeating for all r, a plot of F^* vs $\ln(r)$ was generated from these results (see Fig. 4, open circles). A linear least-squares fit to this plot using Eq. (2) yielded the so-called global fit values of the parameters d and t_{off}. These values were then used to generate the global fit PDFs in Fig. 3 (solid curves). Figure 3 shows a good agreement among the measured PDF, its local fit PDF, and its global fit PDF, and Fig. 4 shows the expected linear relation for F^* vs $\ln(r)$. All these are evidence that (for the loading rates investigated), a single activation barrier dictates the behavior of *Bso*BI unbinding from α sites.

UFAPA also provides a new ability to distinguish between different protein-DNA complexes on a single-molecule site-specific basis. In addition to *Bso*BI unbinding from the α site, we also examined *Bso*BI and *Xho*I unbinding from β sites (Fig. 4). Interestingly, these two binding species also show linear but distinguishable F^* vs $\ln(r)$ plots. We used the same methods described earlier to find the best-fit d and t_{off}, and all results are summarized in Table I, along with values of the equilibrium association constant K_A for these three species, measured with UFAPA as described previously [2].

Figure 4 and Table I show that the three binding species have many potential distinguishable dynamic signatures, including the characteristics of the dominant activation barriers (d and t_{off}), the F^* vs $\ln(r)$ behavior, and even the force distributions themselves. This is true for all the species examined: between the same protein (*Bso*BI) binding to two different sites (α vs β), and between two different proteins (*Bso*BI vs *Xho*I) binding to the same DNA site (β). The former comparison is further illustrated in Fig. 5, where unbinding force distributions are shown for *Bso*BI unbinding from α (*filled bars*) and β (*lined bars*) sites at a force-loading rate of ~60 pN/s. Figure 5 shows clearly distinguishable distributions with only ~19% overlap. Therefore, under these conditions, a single measurement of the unbinding force is nearly sufficient to distinguish between the two species. For example, one could set a threshold at 33 pN such that a measurement < 33 pN is considered to correspond to an α site and a measurement > 33 pN is considered to correspond to a β site; then an assessment based on this single measurement will yield a correct conclusion 90% of the time. This capability can lead to novel assays which screen for multiple proteins and multiple binding sites simultaneously and in parallel.

We have shown that when probed with UFAPA, disruption of three protein-DNA binding species conformed well to the theory of DFS. Analysis revealed a prominent activation barrier for disruption of each site-specific protein-DNA complex. Note that it is possible that disruption of protein-DNA interactions by unzipping may not proceed along the "natural" zero-force dissociation pathways for the binding species examined. Indeed, for *Xho*I disruption from beta sites, it is likely that the natural lifetime is much lower than the apparent lifetime of 6000 s obtained here [13]. While it may be difficult to relate the apparent lifetime to the natural zero-force lifetime, it is nevertheless important to note

TABLE I. Summary of results for three different binding species. The parameters d and t_{off} are obtained from Fig. 4. K_A is the equilibrium association constant measured from the occupancy of sites at ~60 pN/s. N_{DFS} represents the total number of events included in the DFS analysis, while N_{EQ} represents the total number of sites counted for the K_A measurements.

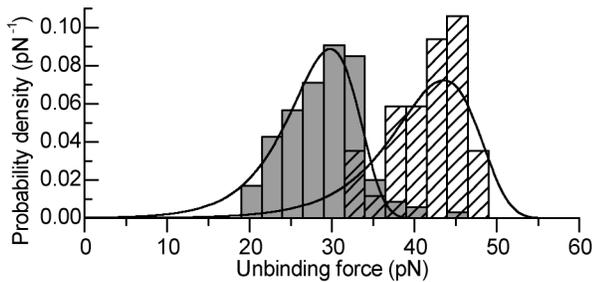| Protein	DNA binding site	d (nm)	$\ln[t_{off}$ (s)]	N_{DFS}	$\log_{10}[K_A$ (M^{-1})]	N_{EQ}
*Bso*BI	α, ttcCTCGGGaat	0.98 ± 0.04	4.53 ± 0.27	449	9.15 ± 0.07	194
*Bso*BI	β, aaaCTCGAGact	0.80 ± 0.09	6.1 ± 0.7	82	9.34 ± 0.19	30
*Xho*I	β, aaaCTCGAGact	1.18 ± 0.12	8.7 ± 0.9	47	8.94 ± 0.27	18

FIG. 5. Distinct unbinding force distributions. *Solid gray bars* represent *Bso*BI unbinding from α sites ($N = 141$), while *lined bars* represent *Bso*BI unbinding from β sites ($N = 35$). *Solid lines* are predicted distributions based on the global fit parameters from Table I. All data are for a force loading rate of ~ 60 pN/s. At this particular stretch rate, the two sites produce highly distinct unbinding signatures, as shown by both the data and the predicted PDF.

that the observed activation barrier represents an important physical aspect of protein-DNA interaction landscape. This is the first direct experimental access to site-specific protein-DNA interaction landscape with previous experiments relying on gel mobility-shift, filter-binding, or other assays which are not site specific, or inextricably include nonspecific protein-DNA interactions.

Possible future applications for UFAPA range from qualitative assays for mapping protein binding sites and distinguishing between binding species to quantitative assays which can probe the energetics of the protein-DNA interactions to access information that was previously unobtainable. Thus, UFAPA combined with DFS presents a powerful new tool for probing specific protein-DNA interactions.

*Corresponding author.
Email address: mwang@physics.cornell.edu

[1] E. S. Lander *et al.*, Nature (London) **409**, 860 (2001); J. C. Venter *et al.*, Science **291**, 1304 (2001).

[2] S. J. Koch, A. Shundrovsky, B. C. Jantzen, and M. D. Wang, Biophys. J. **83**, 1098 (2002).

[3] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).

[4] E. Evans, Annu. Rev. Biophys. Biomol. Struct. **30**, 105 (2001).

[5] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans, Nature (London) **397**, 50 (1999); X. Zhang, E. Wojcikiewicz, and V. T. Moy, Biophys. J. **83**, 2270 (2002); B. D. Brower-Toland, C. L. Smith, R. C. Yeh,

J. T. Lis, C. L. Peterson, and M. D. Wang, Proc. Natl. Acad. Sci. U.S.A. **99**, 1960 (2002).

[6] To generate a loading rate clamp, the microscope coverslip velocity (controlled by a piezostage) was modulated to produce a desired linear force-loading rate. During unzipping, the force, F, and extension were measured, and a freely jointed chain (FJC) model was used to calculate in real time the number of bases unzipped, j. The instantaneous stiffness of the ssDNA (k, pN/nm) was calculated using the FJC model and the known F and j. Throughout the experiment, the piezostretch rate (ν, nm/s) was modulated according to k in order to produce the desired linear force-loading rate (νk). The calculations were performed while still maintaining a feedback loop rate on the order of 10 kHz.

[7] The unzipping construct has 17 nearly identical repeats of ~ 200 bp. Slight variations in the sequence produced two different binding sites for the enzymes used in this report. The first, designated α sites, have the sequence CTCGGG and are bound only by *Bso*BI. The second, designated β sites, have the sequence CTCGAG and may be bound by both *Bso*BI and *Xho*I.

[8] The restriction enzymes used were *Bso*BI (350 pM) and *Xho*I (2300 pM). Experiments were performed at 23 °C in a buffer containing 50 mM sodium phosphate pH 7.0, 50 mM NaCl, 0.02% Tween-20, 10 mM EDTA.

[9] The time series data for j were converted to $\sqrt{j}$, followed by median filtering, a derivative with respect to time, and finally threshold detection of events. Disruptions show up as large positive derivatives of $\sqrt{j}$. Events with a disruption force less than 19 pN are discarded to avoid events that may be naked DNA disruption.

[10] P. R. Bevington and D. K. Robinson, in *Data Reduction and Error Analysis for the Physical Sciences*, edited by S. J. Tubb and J. M. Morris (McGraw-Hill, Inc., New York, 1992), 2nd ed., Vol. 1, Chap. 10, p. 180.

[11] This distribution assumes that all forces from 0 to ∞ are experimentally accessible, whereas in our UFAPA implementation, we have both a lower and an upper force cutoff (shown as *vertical dashed lines* in Fig. 3). To account for this, we modified Eq. (1) to set $p(F; r)$ to zero outside of the experimental range and to normalize the remaining probability so the integral remains unity. The lower force threshold depends on the starting force for the particular event, while the upper cutoff force is set to 51 pN, to prevent overstretching of the dsDNA handle.

[12] To obtain the error bars, the data points in Fig. 4 first were fit (without error bars) to estimate the global d and t_{off}. These parameters and Eq. (1) were used to perform a Monte Carlo simulation of the same number of events as were in the original data set. The simulated data set was then analyzed in the same way as the original data, and F^* was computed from Eq. (2). The simulation was repeated 1000 times, and the standard deviation of F^* was used for the error bar.

[13] Our observed K_A of about 10^9 M^{-1} would require an extremely slow on rate of about 10^5 M^{-1} s^{-1} — or, at the 1 nM concentration of protein used, an on time of about 10 000 s. Experimentally we observed an on time on the order of 100 s or less.