

# Probing Protein-DNA Interactions by Unzipping a Single DNA Double Helix

Steven J. Koch, Alla Shundrovsky, Benjamin C. Jantzen, and Michelle D. Wang

Department of Physics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853 USA

**ABSTRACT** We present unzipping force analysis of protein association (UFAPA) as a novel and versatile method for detection of the position and dynamic nature of protein-DNA interactions. A single DNA double helix was unzipped in the presence of DNA-binding proteins using a feedback-enhanced optical trap. When the unzipping fork in a DNA reached a bound protein molecule we observed a dramatic increase in the tension in the DNA, followed by a sudden tension reduction. Analysis of the unzipping force throughout an unbinding “event” revealed information about the spatial location and dynamic nature of the protein-DNA complex. The capacity of UFAPA to spatially locate protein-DNA interactions is demonstrated by noncatalytic restriction mapping on a 4-kb DNA with three restriction enzymes (*Bso*BI, *Xho*I, and *Eco*RI). A restriction map for a given restriction enzyme was generated with an accuracy of  $\sim 25$  bp. UFAPA also allows direct determination of the site-specific equilibrium association constant ( $K_A$ ) for a DNA-binding protein. This capability is demonstrated by measuring the cation concentration dependence of  $K_A$  for *Eco*RI binding. The measured values are in good agreement with previous measurements of  $K_A$  over an intermediate range of cation concentration. These results demonstrate the potential utility of UFAPA for future studies of site-specific protein-DNA interactions.

## INTRODUCTION

Protein-DNA interactions are essential to cellular processes. In replication, transcription, recombination, DNA repair, and DNA packaging proteins bind to DNA as activators or repressors, to recruit other proteins, or to carry out various catalytic activities. These DNA-binding proteins include polymerases, helicases, nucleases, isomerases, ligases, histones, and others. Because of their great importance, protein-DNA interactions have justifiably drawn much attention from biochemical researchers over the last half-century. More recently, the application of single-molecule mechanical techniques to the interactions of proteins and DNA has attracted great interest, in particular for the study of molecular motors such as RNA polymerases, DNA polymerases, helicases, and topoisomerases (Yin et al., 1995; Wang et al., 1998; Wuite et al., 2000; Bianco et al., 2001; Dohoney and Gelles, 2001; Strick et al., 2000), as well as the investigation of chromatin structure (Cui and Bustamante, 2000; Bennink et al., 2001; Brower-Toland et al., 2002).

Critical parameters for protein-DNA interactions include location, specificity, and strength of interaction. Many biochemical techniques exist that provide information about these parameters, but none provide all of them at once on a molecule-by-molecule basis. We describe here a single molecule technique for the analysis of protein-DNA interactions. It is based on unzipping a single DNA double helix in the presence of bound proteins. We term this technique unzipping force analysis of protein association (UFAPA). We show that UFAPA is a powerful approach for locating

specific binding sites for a given protein on a DNA molecule, and for probing the energetics of the protein-DNA interactions.

Previously, it was demonstrated that the force required to unzip naked DNA depends strongly on the local nucleotide sequence (Bockelmann et al., 1997, 1998; Essevez-Roulet et al., 1997). Furthermore, this force could be predicted from a simple quasi-equilibrium model accounting only for the energies of A-T versus G-C basepairs and the series compliance of the system. Our work extends this unzipping technique to the study of protein-DNA interactions.

The restriction endonucleases compose a well-studied class of DNA-binding proteins. *Eco*RI and other restriction endonucleases have been important tools in the development of modern molecular biology, and have also served as useful models for other protein-DNA interactions. As a proof of principle, this report presents 1) detection of *Eco*RI binding at two canonical sites separated by 11 bp; 2) noncatalytic restriction mapping of DNA using *Bso*BI, *Xho*I, and *Eco*RI; and 3) determination of the cation concentration dependence of the equilibrium association constant of *Eco*RI binding.

## MATERIALS AND METHODS

The experimental configuration is shown in Fig. 1 *A*. One strand of a double-stranded (ds) DNA molecule to be unzipped was attached to the surface of a microscope coverslip while the other strand, which originated from the same end, was attached to a polystyrene microsphere. To unzip the dsDNA, the two single strands of the DNA molecule were pulled apart by moving the coverslip while holding the microsphere in a fixed position with a feedback-enhanced optical trap. The number of unzipped basepairs is referenced by an unzipping index  $j$ . This configuration is a combination of those used by Bockelmann et al. (1997) and Wang et al. (1998).

Submitted October 2, 2001, and accepted for publication April 22, 2002.

Address reprint requests to Michelle D. Wang, 518 Clark Hall, Ithaca, NY 14853. Tel.: 607-255-6414; Fax: 607-255-6428; E-mail: mdw17@cornell.edu.

© 2002 by the Biophysical Society

0006-3495/02/08/1098/08 \$2.00

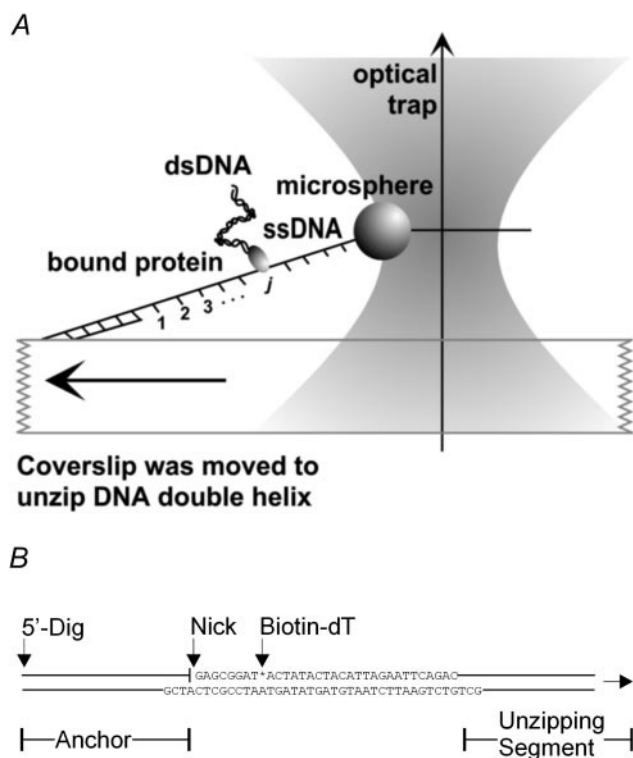


FIGURE 1 Experimental configuration. (A) Cartoon of the unzipping configuration. The two strands of the DNA molecule are unzipped with a feedback-enhanced optical trap. Unzipping proceeds rather smoothly until a DNA-bound protein is encountered, and additional force is required to unzip through it. The location of the unzipping fork is indicated by an unzipping index  $j$ , which is the number of basepairs unzipped from the coverslip-bound end of the DNA molecule. (B) Schematic of the DNA molecule, *not to scale*. The complete sequences are shown for the two oligonucleotides composing the insert duplex. The locations of the *digoxigenin* and *biotin* labels, and the *nick* are shown.

## Biochemical materials

### DNA molecules

The DNA molecule used for unzipping was adapted from Bockelmann et al. (1997) and is shown in Fig. 1 B. One end of the DNA was labeled with a digoxigenin (dig) for attachment to a coverslip via anti-digoxigenin (Roche Molecular Biochemicals, Indianapolis, IN). The nicked strand, 1.1 kb distant from the dig-labeled end, was labeled with a biotin 8 bp away from the nick for attachment to a streptavidin-coated 0.48  $\mu\text{m}$ -diameter microsphere (Bangs Laboratories, Inc., Fishers, IN). Therefore, when the DNA was unzipped by  $j$  bases, there were  $N_{\text{ss}} = 2j - 8$  bases in the ssDNA. The two insert oligos whose complete sequences are shown in Fig. 1 B allowed for coupling the dig-labeled anchoring segment to the unzipping segment, via a 3' overhang and a 5' overhang on the bottom strand of the duplex.

The anchoring double-stranded segment (1120 bp) was derived from the *rpoB* gene contained in pRL574 (kindly provided by R. Landick; template no. 5 in Schafer et al., 1991). The dig label was the result of PCR with a dig-labeled primer. After PCR, the segment was digested with *Bst*XI (NEB), gel extracted, and ligated to the ATCG-3' overhang of the insert duplex.

The repetitive unzipping segment used in most of the experiments was derived from pCP681 (kindly provided by C. Peterson) consisting of a

sequence of 17 head-to-tail segments of the form xxxxyzxxxxxyxxxxxy derived from 5S rRNA genes (see Logie and Peterson, 1997, for the corresponding 11 head-to-tail segments of the form xxxxyzxxxxxy). pCP681 was digested with *Eco*RI (New England BioLabs, Beverly, MA; NEB), the 4.1-kb fragment was gel-extracted, and then ligated to the 5'-GCT overhang of the insert duplex. For further *Eco*RI studies, a different unzipping segment was ligated via the same 5'-GCT overhang. This unzipping segment was the large fragment of an *Eco*RI digest of pBR322, resulting in a single *Eco*RI site  $\sim 2.4$  kb downstream from the nick. Unzipping constructs were attached to the dig surface by incubating at DNA concentrations  $\leq 30$  pM. Given a maximum DNA tethering efficiency of 10% (unpublished results), this is equivalent to a solution concentration of  $\leq 3$  pM.

### Unzipping buffer conditions

Experiments with *Bso*BI and *Xho*I were performed at room temperature (23°C) in a buffer containing 50 mM sodium phosphate buffer pH 7.0, 50 mM NaCl, 0.02% Tween-20, 10 mM EDTA. To facilitate better comparisons of future *Eco*RI results with previously reported results, our mapping and equilibrium constants studies of *Eco*RI were performed at room temperature (23°C) in a buffer containing 10 mM Hepes, pH 7.6, 1 mM EDTA, 50  $\mu\text{M}$  DTT, 100  $\mu\text{g}/\text{ml}$  BSA, 500  $\mu\text{g}/\text{ml}$  Blotting Grade Blocker (Bio-Rad, Hercules, CA) and NaCl added to produce total  $\text{Na}^+$  concentrations of 106 to 262 mM. All buffers did not contain  $\text{Mg}^{2+}$ , which is required for catalytic activity of the restriction endonucleases. The Hepes buffer is similar to the buffer used by Ha et al. (1989) for temperature dependence studies of *Eco*RI binding to its site on pBR322. Because of the tendency for *Eco*RI to aggregate at lower ionic strengths (Jen-Jacobson et al., 1983), the *Eco*RI mapping data were taken at 131 mM total  $\text{Na}^+$  concentration.

### Enzymes

All enzymes were commercial grade, purchased from NEB, and used without further purification.

*Eco*RI. To determine the molar concentration of actively binding *Eco*RI, we performed an agarose gel mobility shift assay (data not shown). Various concentrations of *Eco*RI were incubated for 1 h at room temperature with 10 nM of a 33 bp synthesized DNA duplex containing a single *Eco*RI binding site in 10 mM Hepes, pH 7.6, 1 mM EDTA, 50  $\mu\text{M}$  DTT, 100  $\mu\text{g}/\text{ml}$  BSA, 156–194 mM total  $\text{Na}^+$  concentration during incubation. These samples were then run on a 2.4% agarose gel at 4°C to determine the fraction of DNA bound. Using this assay, we determined the concentration of actively binding *Eco*RI molecules in the undiluted stock to be 300 nM. This is  $\sim 40\%$  of the expected 800 nM based on NEB's reported activity for this lot (62 kD dimer;  $2 \times 10^6$  U/mg specific activity; 100,000 U/ml stock concentration). The difference between our measured activity and that reported by NEB may reflect degradation of enzymatic activity or errors in NEB's reported unit concentration and specific activity. *Eco*RI equilibrium constant measurements were performed with NEB enzyme at concentrations from 50 to 6000 pM with the actual concentration chosen for maximum expected counting precision (see Results).

*Bso*BI and *Xho*I. The concentrations of *Bso*BI and *Xho*I were determined from the company's reported unit concentration and specific activity of each enzyme; the actual active binding fraction was not determined. *Bso*BI and *Xho*I were used at respective concentrations of 0.7 nM (72 kD dimer;  $4 \times 10^6$  U/mg specific activity; 200 U/ml working concentration) and 4.6 nM (52 kD dimer;  $1.7 \times 10^6$  U/mg specific activity; 400 U/ml working concentration).

### Instrumentation and calibration

The measurements were obtained using a single-beam optical trapping microscope. After passing through a single-mode optical fiber (Oz Optics,

Carp, ON) and an acousto-optic deflector (NEOS Technologies, Inc., Melbourne, FL), 1064 nm laser light (Spectra-Physics Lasers, Inc. Mountain View, CA) was focused onto the sample plane using a 100 $\times$ , 1.4 NA, oil immersion objective on an Eclipse TE200 DIC microscope (Nikon USA, Melville, NY). After interacting with a trapped microsphere, the laser light was collected by a 1.4 NA oil immersion condenser and projected onto a quadrant photodiode (Hamamatsu, Bridgewater, NJ). The photocurrents from each quadrant of the photodiode were amplified and converted to voltage signals using a position detection amplifier (On-Trak Photonics, Inc., Lake Forest, CA). The position of the optical trap relative to the sample was adjusted with a servo-controlled 1-D piezoelectric stage (Physik Instrumente GmbH & Co, Waldbronn, Germany). Analog voltage signals from the position detector and stage position sensor were anti-alias filtered at 5 kHz (Krohn-Hite, Avon, MA) and digitized at 7 to 13 kHz for each channel using a multiplexed analog to digital conversion PCI board (National Instruments Corporation, Austin, TX).

The calibration and data conversion methods of the instrument were adapted from those used by Wang et al. (1997, 1998). In brief, the first step of the calibration determined the position of the trap center relative to the beam waist and the height of the trap center relative to the coverslip. The second step of the calibration determined the position detector sensitivity and trap stiffness. The third step of the calibration located the anchor position of the DNA on the coverslip, and was performed before each unzipping measurement by stretching a DNA at low load (<5 pN, not sufficient to unzip). These calibrations were subsequently used to convert data into force and extension for an actual unzipping experiment.

### Determination of the force-extension relations

Elastic parameters of both dsDNA and single-stranded (ss) DNA are necessary for the interpretation of the data (see Results). The elastic parameters of dsDNA were obtained from Wang et al. (1997), who used an extensible worm-like-chain model (Marko and Siggia, 1995): the contour length per base 0.338 nm, the persistence length of DNA 43.1 nm, and the stretch modulus 1205 pN. To obtain the elastic parameters of ssDNA, a modified version of the DNA molecule was constructed that included a capped end on the double-stranded part that was to be unzipped (Bockelmann et al., 1997). First, this DNA was completely unzipped (forces 12–17 pN). This resulted in a rather extended molecule with dsDNA (at the coverslip anchor) and ssDNA in series. This unzipped DNA was then stretched to a higher force up to 50 pN to obtain the force-extension curve, which reflects elastic contributions from both the dsDNA and ssDNA. Given the elastic parameters of dsDNA, this curve allowed the determination of the elastic properties of ssDNA using an extensible freely jointed-chain model (Smith et al., 1996): a contour length per base of 0.539 nm, a persistence length of 0.796 nm, and a stretch modulus of 580 pN.

### Unzipping data acquisition

To unzip a DNA double helix as shown in Fig. 1 *A*, the coverslip was moved relative to the trapped microsphere with a piezoelectric stage to stretch the DNA under either a velocity clamp or a proportional velocity clamp. Both of these clamps were implemented with digital feedback, with an average rate for a complete feedback cycle of 7–13 kHz. In the velocity clamp mode, the coverslip was moved at a constant velocity  $v_s$  (in nm/s) relative to the trapped microsphere, whose position was kept constant by modulating the light intensity (trap stiffness) of the trapping laser. Unzipping, during which dsDNA was converted to ssDNA, was observed as a reduction in the tension of the DNA. In the proportional velocity clamp mode the coverslip was moved at a velocity  $v_s$  that was proportional to the number of unzipped bases,  $N_{ss}$ , calculated at real time, while the position of the microsphere was kept constant by modulating the light intensity (trap stiffness) of the trapping laser. In other words, in the proportional velocity clamp mode  $v_s/N_{ss}$ , rather than  $v_s$ , was held constant. The method of

computing the number of unzipped bases is discussed in the Results section. Unzipping was observed as a reduction in the tension of the DNA and a corresponding increase in the velocity of stretching. The velocity clamp is rather straightforward as a method of stretching and was used in some of the experiments, whereas the proportional velocity clamp is an enhancement to account for the increasing compliance of the ssDNA as the construct is unzipped. The proportional velocity clamp will allow future UFAPA studies to quantitatively analyze the forces of unbinding events at different locations on the DNA.

## RESULTS AND DISCUSSION

A number of experiments were carried out to demonstrate the capability of the UFAPA approach to locate DNA-binding sites and to assess the dynamic signatures of protein-DNA interactions. The DNA-binding proteins used here were restriction enzymes (*Bso*BI, *Xho*I, and *Eco*RI). As shown in Fig. 1 *A*, tethered DNA was incubated with a restriction enzyme in the absence of  $Mg^{2+}$ , which allowed the restriction enzyme to bind to its cognate site without cutting the DNA molecule. Before unzipping, the DNA and protein were incubated together for  $\sim$ 15 min to allow them to come to equilibrium. Longer incubation times did not increase the fraction of detectable bound complexes.

### Detection of bound proteins

As the DNA was unzipped, the tension (force) and extension of the DNA were monitored continuously. An example of data is shown in Fig. 2 *A*, which is a plot of the force-extension relation for an unzipping process that used a velocity clamp at 700 nm/s. The force-extension curve in the presence of *Bso*BI (*red curve*) differs dramatically from that of naked DNA (*black curve*). The unzipping force for naked DNA was rather uniform (12–17 pN), whereas unzipping in the presence of *Bso*BI produced a series of dramatic increases in force (up to 40 pN) with each increase followed by a rapid relaxation. The high force events observed for unzipping of DNA in the presence of *Bso*BI rise from a baseline that corresponds to the force curve obtained from unzipping of naked DNA. These high-force events presumably correspond to the resistance of *Bso*BI to unzipping and its subsequent unbinding from the DNA double helix.

To determine where a protein binds, the unzipping index  $j$  (see Fig. 1) must be converted from force-extension curves. This conversion relies on the elastic parameters (see Materials and Methods) of the stretched DNA, which in our configuration was composed of both ssDNA and dsDNA. It uses a method similar to that used by Wang et al. (1998) to compute the DNA tether length during a single molecule transcription experiment. The converted data from Fig. 2 *A* are shown in Fig. 2 *B*, where  $j$  is plotted as a function of time. Compared with the naked DNA curve, the *Bso*BI curve shows a pronounced staircase pattern at each protein disruption event due to clamping of the helix by the bound

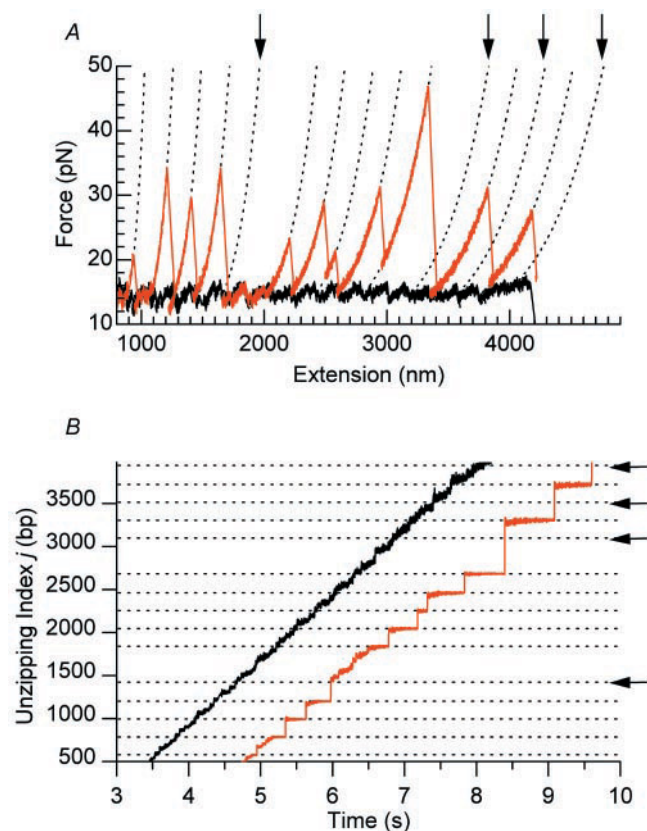


FIGURE 2 Comparison of DNA unzipping data in the absence and presence of binding proteins. (A) Force versus extension for two identical DNA molecules unzipped in the absence (black lines) or presence (red lines) of *BsoBI* (700 pM) using a velocity clamp at 700 nm/s. The resistance to unzipping by *BsoBI* resulted in distinctive peaks that were not present with the naked DNA. The dotted black curves represent calculated force-extension relations for DNA molecules as shown in Fig. 1 A, where  $j$  is set to the starting index for each of the *BsoBI* binding sites. See Methods for explanation of ssDNA and dsDNA modeling. (B) Unzipping index  $j$  versus time. The unzipping index  $j$  was calculated from data shown in A. The origin of the time axis is arbitrary. Horizontal dotted lines indicate the expected binding sites, corresponding to the dotted curves in Fig. 2 A. A peak in Fig. 2 A that resulted from *BsoBI* resistance became a large plateau because the unzipping index  $j$  remained unchanged until *BsoBI* unbound. At the concentration of *BsoBI* used some sites remain unoccupied, as shown at sites ~1400 bp, 3100 bp, 3500 bp, and 4000 bp (indicated by horizontal arrows).

*BsoBI*. The locations of the plateaus clearly indicate the locations of the *BsoBI* binding sites on the DNA sequence. These measured binding sites agree well with the expected sites, which are indicated by the dotted horizontal lines in the plot.

Fig. 3 illustrates the high resolution of the unzipping technique for ascertaining the location of one bound protein relative to another. Fig. 3 A shows force versus unzipping index  $j$  for unzipping carried out in the presence of *EcoRI* using a velocity clamp at 280 nm/s. The DNA molecule contains two expected closely spaced *EcoRI* sites (vertical bars under the horizontal axis) differing by only 11 bp within each repeat of the tandem repeat sequence. Bound

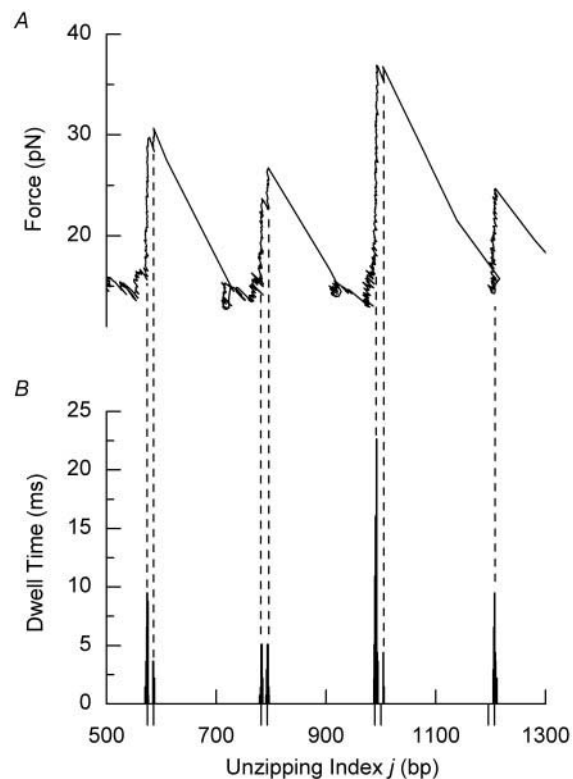


FIGURE 3 Detection of protein binding sites. (A) Force versus unzipping index  $j$  in the presence of *EcoRI* ( $[EcoRI] = 83$  pM,  $[Na^+] = 131$  mM) using a velocity clamp at 280 nm/s. The resistance to unzipping by *EcoRI* resulted in distinctive peaks at the locations of bound *EcoRI*. Each peak was followed by a sudden reduction of force after *EcoRI* unbound. (B) Dwell time versus unzipping index  $j$  for forces  $>20$  pN (threshold force for inclusion of data). The vertical dashed lines indicate peaks in the dwell time distributions. The vertical bars below the unzipping index axis indicate the predicted binding locations of *EcoRI* based on the known recognition sequence on the pCP681-derived construct.

*EcoRI* was detected by a sudden rise in the force for unzipping. When *EcoRI* binding to one of these sites was disrupted, the DNA double helix unzipped and the tension dropped until it reached the level characteristic of that for unzipping naked DNA or until another bound *EcoRI* was encountered by the unzipping fork. As demonstrated by the doublet peaks around  $j = 600, 800,$  and  $1000$  bp in Fig. 3 A, binding sites that differ by as little as 11 bp can be readily resolved. To facilitate location of binding sites, a plot of dwell time versus unzipping index  $j$  is shown in Fig. 3 B. Only data corresponding to forces  $>20$  pN are included in this plot, and the bin size for unzipping index is 1 bp. The standard deviation of a peak is 3 bp, the resolution limit for the determination of the location of one bound protein relative to another.

### Mapping of bound proteins

Restriction mapping was used to illustrate one of the important applications of this technique: accurate and precise

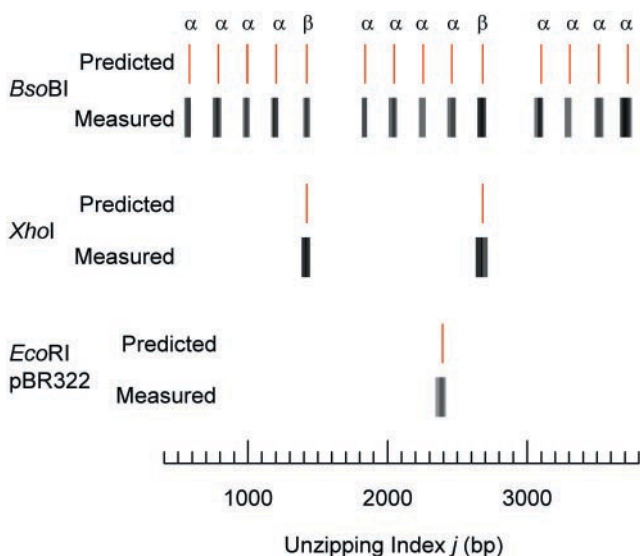


FIGURE 4 Noncatalytic restriction mapping. This is a summary of data from multiple DNA molecules for the three restriction enzymes studied. For each enzyme, *red bars* mark the expected recognition sites. For *BsoBI*,  $\alpha$  and  $\beta$  represent two different canonical binding sites (see text). The *gray-scale intensity* represents the binding frequency in log scale determined from unzipping experiments (see text). Data for *BsoBI* ( $[BsoBI] = 700$  pM; 7 DNA molecules unzipped) and *XhoI* ( $[XhoI] = 4.6$  nM; 4 DNA molecules unzipped) are for binding to the repetitive DNA molecule, while data for *EcoRI* ( $[EcoRI] = 300$  pM; 16 DNA molecules unzipped) are for binding to the pBR322-derived DNA molecule, for which there is one well-known binding site (see Methods).

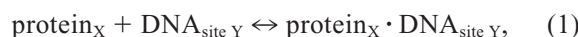
mapping of bound proteins. Restriction maps were created for three restriction enzymes complexed with the unzipping DNA molecule (Fig. 4, either repetitive or pBR322-derived DNA molecules). *EcoRI*, *BsoBI*, and *XhoI* were disrupted using a proportional velocity clamp at 0.24–0.59 nm nt<sup>-1</sup> s<sup>-1</sup>. *BsoBI* is known to recognize the sequence CYCGRG, where Y is any pyrimidine and R is any purine (Ruan et al., 1996; van der Woerd et al., 2001). The DNA used to produce Fig. 4 had two different canonical recognition sequences for *BsoBI*, referred to here as  $\alpha$  (tcCTCGGaat) and  $\beta$  (aaaCTCGAGaga). The unzipping index axis has been subdivided into bins of 12 bp width, comparable to the footprint of *EcoRI* and *BsoBI* as estimated from their crystal structures. Data were combined from stretching several different DNA molecules, so that the *grayscale intensity* represents the binding fraction of a given bin, i.e., the fraction of the DNA molecules that had unzipping force >25 pN in the bin. These maps show excellent agreement with the expected restriction maps: over a 4 kb DNA molecule, this technique can locate restriction binding sites with an accuracy of ~25 bp and a precision of ~30 bp. This resolution is not expected to decrease appreciably for much longer DNA molecules (for example, DNA molecules of a few Mbp).

With suitable progress toward automation and parallelism, UFAPA could have future applications in the field of

genome mapping and sequencing. A recent innovation, optical mapping, is a single-molecule restriction mapping technique that preserves site ordering information (Schwartz et al., 1993; Cai et al., 1998). Although not suitable for small-scale mapping, the technique was successfully automated to complete a whole-genome shotgun map of a 3-megabase organism (Lin et al., 1999). UFAPA shares many of the advantages of optical mapping, including site order preservation and other advantages inherent to a single-molecule technique. Furthermore, UFAPA has two other potential advantages—it is a non-imaging technique, allowing for better basepair location resolution, and UFAPA is noncatalytic, allowing for reversible rapid screening of multiple enzymes on a single molecule.

### Determination of equilibrium association constants

The affinity of a protein for its DNA binding site is characterized by the equilibrium association constant. Because UFAPA can directly detect protein-DNA binding, it allows for direct and site-specific measurements of equilibrium association constants,  $K_{A,XY}$ :



and

$$K_{A,XY} = \frac{1}{[\text{protein}_X]} \frac{[\text{protein}_X \cdot \text{DNA}_{\text{site } Y}]}{[\text{DNA}_{\text{site } Y}]}. \quad (2)$$

For a given protein X concentration, the ratio of bound to unbound Y sites,  $r = [\text{protein}_X \cdot \text{DNA}_{\text{site } Y}]/[\text{DNA}_{\text{site } Y}]$ , gives a measure of the equilibrium association constant  $K_{A,XY}$ . As in a gel-mobility shift assay, and unlike filter-binding assays, measurements at a single protein concentration are sufficient to determine  $K_A$ , although in general, titration of protein could be useful to reveal deviations from the above relation and to determine binding stoichiometry. In our studies, the free *EcoRI* concentration was the same as the total protein concentration (50–6000 pM) due to the low effective DNA concentration ( $\leq 3$  pM, see Methods).

The distribution of the number of bound sites follows a binomial distribution; the relative uncertainty in  $K_A$  is as follows:

$$\frac{\sigma_{K_A}}{K_A} = \frac{1+r}{\sqrt{r(N-1)}}, \quad (3)$$

where  $\sigma_{K_A}$  is the standard deviation and  $N$  is the number of measurements. Equation 3 shows that for a given number of measurements, the best precision is obtained when  $r = 1$ . For the present study,  $r$  was kept near 1 to minimize the error; other than increased uncertainty, no differences in the mean values of  $K_A$  were observed when  $r \cong 0.1$  or  $r \cong 10$  (data not shown).

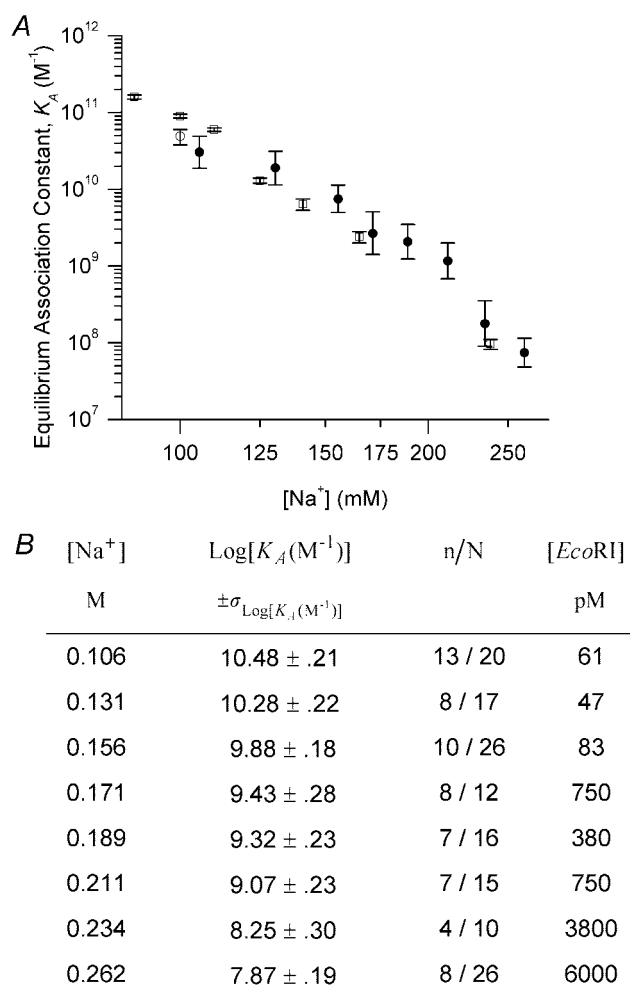


FIGURE 5  $Na^+$  concentration dependence of the equilibrium association constant for *EcoRI* binding to its site on pBR322, plotted on a log-log scale. (A) *Solid circles* represent our measured association constant,  $K_A$ , as described in the text (23°C, 10 mM Hepes, pH 7.6, variable  $Na^+$  concentration). The *open circle* represents filter binding data from Ha et al. (1989) for binding to the pBR322 at 21.1°C (10 mM Hepes, pH 7.6 at 20°C, NaCl added to  $Na^+$  concentration of 100 mM). *Open squares* represent Terry et al. (1983) filter binding data for binding to the pBR322 site at 37°C (20 mM Tris-HCl, pH 7.6, variable  $Na^+$  concentration). (B) Tabulation of the UFAPA data from Fig. 5 A.  $N$  represents the number of DNA molecules probed for binding, and  $n$  represents the number of binding sites found to be occupied.  $[EcoRI]$  represents the actual *EcoRI* concentration used in pM.

Fig. 5 A shows  $K_A$  values that were determined for *EcoRI* binding to its canonical site on pBR322 at various  $Na^+$  concentrations (*solid dots*). Our data in Fig. 5 A are also tabulated in Fig. 5 B. For a given DNA molecule, a site was considered bound if the unzipping force exceeded 20 pN within 100 bp of the expected site. By making measurements at a given site on multiple DNA molecules, the ratio of bound to unbound sites,  $r$ , was obtained for that site. In Fig. 5 both the  $K_A$  values and their error bars are shown on a logarithmic scale. For  $N$  DNA molecules probed and  $n$

bound enzymes detected, the standard deviation in  $\log(K_A)$  is given by

$$\sigma_{\log(K_A)} = \log(e) \left[ (N-1) \left( 1 - \frac{n}{N} \right) \frac{n}{N} \right]^{-1/2}$$

based on a binomial distribution.

For comparison, Fig. 5 A also shows  $K_A$  values for *EcoRI* binding to its canonical site on pBR322 at various  $Na^+$  concentrations from Ha et al. (1989) (*open circle*) and Terry et al. (1983) (*open squares*). The exact conditions for these measurements are given in the figure caption. Our buffers are essentially the same as those used by Ha et al. (1989) and have the same pH as those used by Terry et al. (1983). Our measurement temperature (23°C) was similar to that of Ha et al. (1989) (21.1°C), but somewhat different from that of Terry et al. (1983) (37°C).

Our UFAPA measurements in Fig. 5 overlap the values of Terry et al. (1983) over most of the  $[Na^+]$  ranging from 131 to 262 mM. Deviation from the data of Terry et al. (1983) at the lower salt condition,  $[Na^+] = 106$  mM, is most likely due to protein aggregation, which is known to occur for *EcoRI* at low ionic strength, and in the absence of saturating DNA binding sites (Jen-Jacobson et al., 1983). Terry et al. found the slope of their data to be commensurate with eight ion pairs involved in the binding of *EcoRI* to the pBR322 site: a result consistent with that of Jen-Jacobson et al. (1983). Over the  $[Na^+]$  range of 131 to 234 mM, UFAPA data follow a similar slope to Terry et al. (1983). It is currently unknown whether the apparent increase in slope magnitude around 262 mM  $Na^+$  is due to technical difficulties of the UFAPA method, although the filter binding assay from Terry et al. (1983) shows a similar effect.

Determination of  $K_A$  using UFAPA offers a number of new features compared with traditional bulk equilibrium methods. 1) UFAPA is direct and site-specific, reducing possible complications from nonspecific DNA binding sometimes encountered in bulk studies. 2) A single UFAPA measurement is fast, avoiding possible dissociation of the protein-DNA complex before a measurement is obtained, as may occur in bulk studies (for example, while the sample is entering a gel). Future studies may elucidate whether UFAPA is useful to probe  $K_A$  values with particularly fast dissociation rates. 3) Values of  $K_A$  can be determined simultaneously for multiple protein-DNA interactions at different binding sites on the DNA.

Determination of  $K_A$  using UFAPA also has some limitations. 1) The principal limitation of the UFAPA approach is the lack of commercially available low-cost instrumentation with suitable automation for precise counting statistics. As shown in Fig. 5, counting binding from a few molecules results in a large uncertainty: 10 counts produce at best 67% precision, and 400 counts are required to achieve at best 10% precision for a level consistent with the results of Terry et al. (1983). Established biochemical assays also have the

advantage of running many samples in parallel: up to as many lanes or filter ports as are available on the apparatus. 2) A further point to consider is the current inability to easily titrate the concentration of DNA binding sites. In the current implementation, the DNA is surface-tethered and there is no ability to perform assays under saturating DNA conditions. Future enhancements may allow the immobilization of the DNA-binding protein and subsequent titration of DNA against various protein surface densities. These enhancements would allow determination of binding activity, oligomeric state of binding protein, and other information.

As with traditional biochemical studies of  $K_A$ , UFAPA also has a range of accessible  $K_A$  values. The lower limit will depend on the solubility of the particular protein, and the ability to have enough protein in solution to have appreciable ratio of bound sites. The upper limit for UFAPA can in principle be raised higher than the  $\sim 10^{11} \text{ M}^{-1}$  measured in this report. The current implementation requires that there be significantly fewer DNA binding sites than protein molecules so that the DNA's alteration of the free protein concentration is negligible. As  $K_A$  increases, the amount of surface-tethered DNA should also be decreased. At some point, the reduction in surface tether density would make the current practice unmanageable, but values of at least  $10^{12} \text{ M}^{-1}$  are expected to be measurable by lowering the effective DNA concentration ( $< 0.1 \text{ pM}$ ). These potential limits are not strictly defined and future enhancements of UFAPA could expand the accessible  $K_A$  range, although it is not clear whether this range would exceed that spanned by established bulk assays.

## CONCLUSIONS

The UFAPA technique presented here is a novel and general tool for detection of protein-DNA interactions. It is a single molecule technique that yields the locations of bound proteins and the equilibrium association constants for the protein-DNA interactions. As further enhancements are made we anticipate broad applications of UFAPA in the study of protein-DNA interactions, from simple binding site detection and DNA sequence analysis to the determination of previously unknown protein binding sites on DNA and the detection of previously unknown DNA-binding proteins.

We thank Richard C. Yeh for participation in the construction of the optical trapping setup, and Dr. Robert M. Fulbright, Brent D. Brower-Toland, Dr. Arthur LaPorta, and Dr. Karen Adelman for helpful technical advice and scientific discussions. M.D.W. was supported by grants from the NIH, the Damon Runyon Scholar Award, the Beckman Young Investigator Award, the Alfred P. Sloan Research Fellow Award, and the Keck Foundation's Distinguished Young Scholar Award. S.J.K. was supported by Cornell University DOETG. S.J.K. and A.S. were supported by Cornell University NIHTG. B.C.J. was supported by Cornell's Nanobiotechnology Center.

## REFERENCES

- Bennink, M. L., S. H. Leuba, G. H. Leno, J. Zlatanova, B. G. de Grooth, and J. Greve. 2001. Unfolding individual nucleosomes by stretching single chromatin fibers with optical tweezers. *Nat. Struct. Biol.* 8:606–610.
- Bianco, P. R., L. R. Brewer, M. Corzett, R. Balhorn, Y. Yeh, S. C. Kowalczykowski, and R. J. Baskin. 2001. Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature.* 409:374–378.
- Bockelmann, U., B. Essevez-Roulet, and F. Heslot. 1997. Molecular stick-slip motion revealed by opening DNA with piconewton forces. *Phys. Rev. Lett.* 79:4489–4492.
- Bockelmann, U., B. Essevez-Roulet, and F. Heslot. 1998. DNA strand separation studied by single molecule force measurements. *Phys. Rev. E.* 58:2386–2394.
- Brower-Toland, B. D., C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson, and M. D. Wang. 2002. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc. Natl. Acad. Sci. USA.* 99:1960.
- Cai, W., J. Jing, B. Irvin, L. Ohler, E. Rose, H. Shizuya, U. J. Kim, M. Simon, T. Anantharaman, B. Mishra, and D. C. Schwartz. 1998. High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping. *Proc. Natl. Acad. Sci. USA.* 95:3390–3395.
- Cui, Y., and C. Bustamante. 2000. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proc. Natl. Acad. Sci. USA.* 97:127–132.
- Dohoney, K. M., and J. Gelles. 2001. Chi-sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature.* 409:370–374.
- Essevez-Roulet, B., U. Bockelmann, and F. Heslot. 1997. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. USA.* 94:11935–11940.
- Ha, J. H., R. S. Spolar, and M. T. Record, Jr. 1989. Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J. Mol. Biol.* 209:801–816.
- Jen-Jacobson, L., M. Kurpiewski, D. Lesser, J. Grable, H. W. Boyer, J. M. Rosenberg, and P. J. Greene. 1983. Coordinate ion pair formation between EcoRI endonuclease and DNA. *J. Biol. Chem.* 258:14638–14646.
- Lin, J., R. Qi, C. Aston, J. Jing, T. S. Anantharaman, B. Mishra, O. White, M. J. Daly, K. W. Minton, J. C. Venter, and D. C. Schwartz. 1999. Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. *Science.* 285:1558–1562.
- Logie, C., and C. L. Peterson. 1997. Catalytic activity of the yeast SWI/SNF complex on reconstituted nucleosome arrays. *EMBO J.* 16:6772–6782.
- Marko, J. F., and E. D. Siggia. 1995. Stretching DNA. *Macromolecules.* 28:7016–7018.
- Ruan, H., K. D. Lunnen, M. E. Scott, L. S. Moran, B. E. Slatko, J. J. Pelletier, E. J. Hess, J. Benner II, G. G. Wilson, and S. Y. Xu. 1996. Cloning and sequence comparison of Aval and BsoBI restriction-modification systems. *Mol. Gen. Genet.* 252:695–699.
- Schafer, D. A., J. Gelles, M. P. Sheetz, and R. Landick. 1991. Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature.* 352:444–448.
- Schwartz, D. C., X. Li, L. I. Hernandez, S. P. Ramnarain, E. J. Huff, and Y. K. Wang. 1993. Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science.* 262:110–114.
- Smith, S. B., Y. Cui, and C. Bustamante. 1996. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science.* 271:795–799.
- Strick, T. R., V. Croquette, and D. Bensimon. 2000. Single-molecule analysis of DNA uncoiling by a type II topoisomerase. *Nature.* 404:901–904.

- Terry, B. J., W. E. Jack, R. A. Rubin, and P. Modrich. 1983. Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences. *J. Biol. Chem.* 258: 9820–9825.
- Wang, M. D., M. J. Schnitzer, H. Yin, R. Landick, J. Gelles, and S. M. Block. 1998. Force and velocity measured for single molecules of RNA polymerase. *Science*. 282:902–907.
- Wang, M. D., H. Yin, R. Landick, J. Gelles, and S. M. Block. 1997. Stretching DNA with optical tweezers. *Biophys. J.* 72:1335–1346.
- Wuite, G. J., S. B. Smith, M. Young, D. Keller, and C. Bustamante. 2000. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*. 404:103–106.
- Yin, H., M. D. Wang, K. Svoboda, R. Landick, S. M. Block, and J. Gelles. 1995. Transcription against an applied force. *Science*. 270: 1653–1657.
- van der Woerd, M. J., J. J. Pelletier, S. Xu, and A. M. Friedman. 2001. Restriction enzyme bsoBI-DNA complex. A tunnel for recognition of degenerate DNA sequences and potential histidine catalysis. *Structure. (Camb.)*. 9:133–144.