

# Long-range interactions between transcription factors

Yan Mei Wang<sup>1</sup>, Jonas O Tegenfeldt<sup>1</sup>, Jim Sturm<sup>2</sup> and Robert H Austin<sup>1</sup>

<sup>1</sup> Department of Physics, Princeton University, Princeton, NJ 08544, USA

<sup>2</sup> Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA

Received 28 September 2004, in final form 7 June 2005

Published 3 August 2005

Online at [stacks.iop.org/Nano/16/1993](http://stacks.iop.org/Nano/16/1993)

## Abstract

We discuss a method for analysing the number of GFP–LacI fusion transcription factors bound to a construct of 256 contiguous LacI binding sites using photon bleaching statics. We show by using a combination of imaging of the construct in nanochannels, photon statistics and addition of IGFP that the binding coefficient of the LacI decreases with increasing occupation of the construct, with a binding coefficient of  $10^{-6}$  M when only 15 of the 256 possible sites are occupied. We model this effect by assuming that the GFP–LacI dimer introduces elastic strain into the helix by generalized deformations, and that this strain propagates over distances at least as large as the persistence length.

(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

There are three levels of protein–DNA specificity. The first level is the formation of chromatin by the binding of proteins to DNA. While not exquisitely sequence dependent, chromosomes do form very ordered structures and there certainly is a basic sequence dependence to how proteins are guided to form chromosomes during mitosis. At the next level of specificity and complexity are proteins such as restriction enzymes which cut DNA at certain sites. These proteins do not have the on–off control that transcription factors have but they do certainly cut at precision sites, of great biological importance. Finally we have the transcription factors, which show the highest specificity of all: binding constants to certain sites approach  $10^{-14}$  molar, and are  $10^7$  times higher than non-specific binding constants to dsDNA.

There is an amazing amount of highly specific control that transcription factors must exert on gene expression in order for an organism to successfully adapt to changing conditions. An excellent introduction to this subject can be found in a review article by Luscombe *et al* [1]. Generally, transcription factors can be grouped into four basic classes, although there are many exceptions to the following list: (1) helix–turn–helix, (2) zinc finger, (3) leucine, (4) helix–loop–helix. For all four cases the basic structure of the transcription factor complex consists of two DNA binding sections which are separated by a linker region. The linker region does not contact the DNA helix. Why is this particular structure so

predominantly chosen? We can guess that perhaps part of the recognition process involves some sort of distortion (straining) of the helix in the non-contacted region. When such strain is introduced into the helix, the associated strain energy becomes part of the free energy of the binding process and can strongly affect the binding association. If one looks at only single transcription factors binding to a single site it is difficult to identify the extent to which induced elastic strain in the helix has influenced the binding coefficient of the factor, other than by looking at changes in the coefficient when there is a different basepair sequence [2]. However, if multiple binding sites are sequentially aligned on a section of DNA then it is possible to see how the binding coefficients change, if they do, with increasing occupation number, and from this deduce the possible role of protein-induced strain and strain propagation in the binding process.

There is a connection between the thermally induced bending and twisting dynamics of a dsDNA molecule and the elastic moduli of dsDNA. Thermal energy stores  $\frac{k_B T}{2}$  of energy per degree of freedom in a system, where  $k_B$  is Boltzmann's constant and  $T$  is the temperature in kelvins. This thermal energy results in a bending and twisting persistence length  $p_B$  and  $p_T$ , respectively. These lengths are basically the average radius of curvature due to thermally induced bending and the average distance over which the helix twists through an RMS angle  $\langle \phi^2 \rangle^{1/2} \sim 2\pi$ . The bending persistence length is

$$p_B = \frac{E I_a}{k_B T} \quad (1)$$

while the twisting persistence length is

$$p_T = \frac{GI_p}{k_B T} \quad (2)$$

where  $E$  is the Young's modulus of dsDNA and  $G$  is the shear modulus of ssDNA. The bending persistence length is quite easy to observe since it results in the deformation of the backbone of the dsDNA molecule, giving rise to a random-walk aspect to the contour of the molecule which can be directly measured as the radius of gyration  $R_g$  of the polymer.  $R_g$  is basically the radius of the glowing blob that a genomic length dsDNA molecule appears as in a microscope. For a non-self-avoiding random walk,  $R_g$  is given for a polymer of contour length  $L$  and bending persistence length  $p_B$  by

$$R_g = \left[ \frac{L p_B}{6} \right]^{1/2}. \quad (3)$$

The twisting modulus is rather more difficult to measure directly, although topological considerations [3] make it possible to measure  $G$  by observing the braiding of DNA closed circles [4]. Although both  $E$  and  $G$  are undoubtedly a function of basepair sequence [2], as a rough rule of thumb we can say that  $E \sim G \sim 10^9$  dyne  $\text{cm}^{-2}$  for dsDNA at 300 K and 100 mM salt.

If a protein complex when it binds to DNA strains the dsDNA molecule by introducing a localized bend of radius  $R$  or a torsional twist angle of magnitude  $\Theta_o$  into the dsDNA we can compute the strain energy. In the case of a rod which is bent into a radius of curvature  $R$  in a length  $L$ , we have

$$U_{\text{tot}}(R) = \frac{LEI_s}{R^2} \quad (4)$$

where  $I_s$  is called the surface moment of inertia:

$$I_s = \int x^2 dx dy. \quad (5)$$

If the rod is twisted through an angle  $\Theta$ , we have

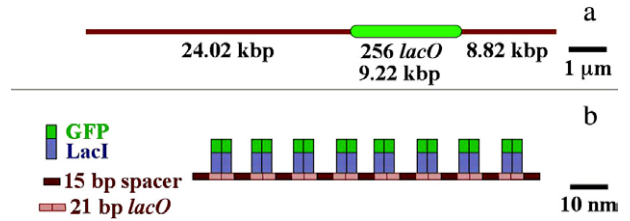
$$U_{\text{tot}}(\Theta) = \frac{I_p G}{L} \theta^2 \quad (6)$$

where  $G$  is the shear modulus and  $I_p$  is the polar moment of inertia:

$$I_p = \int r^2 dA. \quad (7)$$

Thus, given knowledge of the elastic moduli of dsDNA it is in principle possible to calculate the elastic strain stored in a deformed dsDNA molecule when a protein binds to and distorts the dsDNA. The calculation is possible if certain simplifying assumptions are made. For example, if we assume, as was done in the original phage 434 repressor work [2], that the strain is strictly confined to two contacting regions then the strained region is well defined and occurs over a region  $L$  which is much less than the persistence length of the dsDNA molecule.

These strain energies  $U$  then affect the binding coefficient of the protein via the Boltzmann equation, a fundamental result from statistical physics. The Boltzmann equation relates the probability for a system to have an energy  $U$  under conditions of thermal equilibrium at temperature  $T$ . To calculate the



**Figure 1.** (a) Schematic diagram of the  $\lambda$  DNA construct with 256 tandem copies of the *LacO* binding units. *LacO*<sub>256</sub>-DNA is 42.06 kbp long, and the 9.22 kbp *LacO*<sub>256</sub> insertion starts at 24.02 kbp. (b) Scale model of the specific *LacO* sites with the sequence of 5'-AATTGTGAGCGGATAACAATT-3', the spacer sequences between them, and the GFP-LacI proteins bound to full occupancy. The dimensions of the LacI and GFP molecules are 3 nm  $\times$  6 nm [10] and 3 nm  $\times$  4 nm [11], respectively.

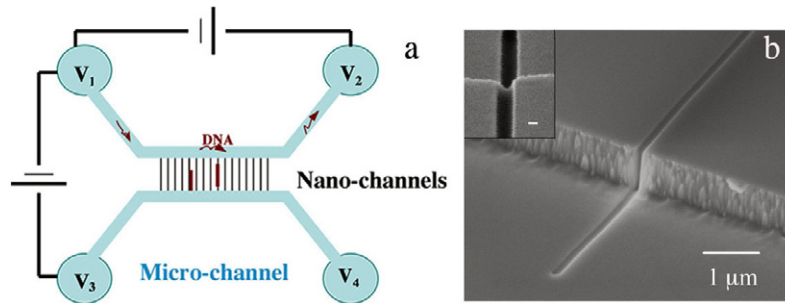
change in binding strength we assume that a transcription factor has a fixed conformation which demands that a certain amount of strain energy  $U$  be invested in the bound complex, in addition to the chemical energies of interaction with the DNA bases and charges. If the sequence denoted by  $n$  changes that binding energy by an amount  $\Delta U_n$  then the binding coefficient  $K_n$  will be decreased by the Boltzmann factor relative to the no-strain case  $K_0$ :

$$K_n = K_0 \exp[-\Delta U/k_B T]. \quad (8)$$

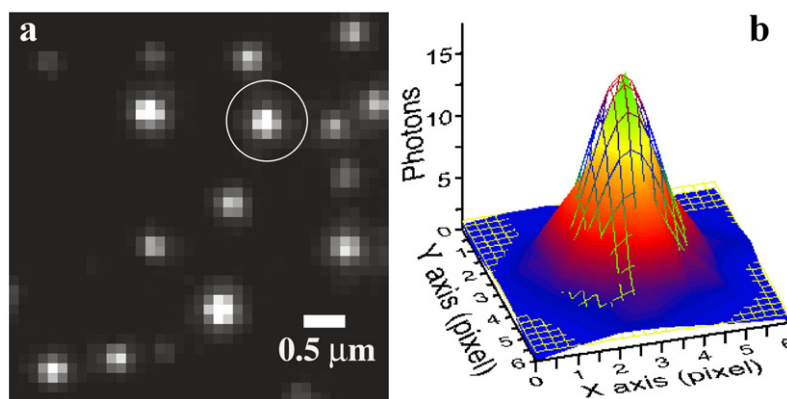
We live on a rather cold planet and dsDNA is rather stiff (the Young's modulus  $E$  of dsDNA is of the order of that of nylon), so it is not surprising how little strain one needs to impose on a section of DNA before the elastic energy terms become similar in magnitude to  $k_B T = 1/40$  eV for  $T = 300$  K. For example, the normal bending radius of DNA due to thermal fluctuations is 50 nm, a very large radius of curvature compared to the length  $L$  where binding of transcription factors occur, perhaps 10 basepairs (bp).

However, equation (8) as written applies only for single transcription factors binding to single isolated sites. One can ask what would happen if multiple transcription factors bind to a contiguous stretch of binding sites if the strain propagates *outside* the local binding region. Since in the linear response regime the strain energies increase quadratically with strain (the bend angle  $R$  or the twist angle  $\Theta$ ), if two transcription factors bind next to each other and *if* the strains add linearly then the strain energy increases by a factor of 4 for two nearby sites. However, we really do not know to what extent deformations induced by a transcription factor actually propagate outwards from the binding sites since most x-ray crystallography is done on simple protein-DNA complexes.

For continuum mechanics in fact it is impossible for there to be a strain (a deformation of an object) in the absence of a stress, an applied force. Thus, any strain induced at the site of a transcription factor binding event should not in linear continuum mechanics result in strain 'reaching out' to a remote distance. In our case, however, we are dealing with a molecular system where localized strain can result in a deformation of the molecular structure which can propagate through the molecule to affect processes somewhere else. This 'non-localized' kind of effect is very well known in biology: it is the essence of allosteric effects in proteins [5]. Allosteric response is expected in a protein which is a complex three-dimensional object, but such an adaptive structural change is perhaps less



**Figure 2.** (a) Schematic diagram of the micro- and nanofluidic device. Blue regions are microchannels and the bridging black lines are nanochannels where the DNA molecules (red) are elongated. DNA molecules are guided consecutively into micro- and nanochannels using electrophoresis. (b) SEM images of a  $120 \text{ nm} \times 150 \text{ nm}$  channel made using FIB. The inset is the top slanted view ( $52^\circ$ ) of the nanochannel, showing its dimensions. The scale bar for the inset is  $100 \text{ nm}$ .



**Figure 3.** (a) Frame-averaged image of single GFP-LacI molecules attached to a quartz surface taken from a  $3.7 \text{ Hz}$  measurement with an exposure time of  $0.1 \text{ s}$ . This image is the average of 10 frames. The pixel size is  $154 \text{ nm}$  using a  $100\times$  objective and a  $1.5\times$  projection lens. The brightness variation of the frame-averaged GFP-LacI fluorescence dots indicates variation in the number of photons emitted. (b) A 2D Gaussian fit (lines) to the number of emitted photons from the circled GFP-LacI dot in (a). The standard deviations in the  $X$  and  $Y$  directions are  $\sigma_X = 125 \text{ nm}$  and  $\sigma_Y = 131 \text{ nm}$ . The optical resolution is the FWHM of the point spread function,  $\sigma_X \times 2.356 = 295 \text{ nm}$ .

expected in dsDNA, which is a simpler linear biomolecule. We go no further here with speculation but instead return to data.

In this paper, we show using a combination of photon counting statistics and nanochannel elongation studies that there is evidence of strain propagation outwards from transcription factor binding sites which strongly limits the number of contiguously bound transcription factors.

## 2. Methods and materials

To study the specific binding of LacI to DNA, we synthesized and cloned LacI-green fluorescent protein (GFP-LacI) fusion proteins and constructed  $\lambda$  DNA with LacO inserts. We employed the widely used fluorescent marker GFP (S65T mutant) to label LacI monomer proteins [6–8]. In nature LacI repressor binds DNA as dimers, and two dimers bound to two different DNA molecules can tetramerize and thus induce DNA aggregation. To avoid DNA aggregation in our experiments, we removed the carboxy-terminus of the LacI proteins which is necessary for tetramerization. We expect to observe GFP-LacI monomers in solution and GFP-LacI dimers upon specific binding to DNA. For our LacI-DNA measurements, we observed no aggregation.

DNA with 256 tandem copies of LacO was constructed as shown schematically in figure 1. DNA constructs with LacO

inserts were cloned using the bacteria phage  $\lambda$ . The LacO repeats were liberated from plasmid pAFS59 [8, 9] (provided by Dr James Broach's lab) by digesting with BamHI. The liberated LacO repeats were then ligated to the BamHI site of Lambda DASH II vector (Stratagene). The LacO<sub>256</sub>-DNA molecules constructed using this method are  $42.06 \text{ K}$  basepair (bp) long with a contour length  $L_c \approx 14.3 \mu\text{m}$ . Each unit of the insert is  $36 \text{ bp}$  long with  $22 \text{ bp}$  long LacO spaced  $14 \text{ bp}$  apart. The tandem LacO<sub>256</sub> is  $9216 \text{ bp}$  long and is located off-centre. As a result, LacO<sub>256</sub>-DNA constructs are asymmetric, with a longer ( $24.02 \text{ kbp}$ ) and a shorter arm ( $8.82 \text{ kbp}$ ). GFP-LacI protein ( $60 \text{ nM}$ ) was incubated with DNA in  $0.5\times$  TBE buffer and  $100 \mu\text{g ml}^{-1}$  bovine serum albumin (BSA) for  $1 \text{ h}$ . The ratio of LacI monomers to LacO binding sites was  $6/1$ . After the binding of LacI to DNA, the LacI-DNA molecules were intercalated with BOBO-3 dye (Molecular Probes) at a dye-to-DNA ratio of 1 dye molecule per 5 basepairs of DNA.

We used two methods to quantitatively analyse our LacI-DNA molecules: The first method is a total photon counting method: it gives the number of proteins that bind to DNA; the second method utilizes nanofluidic channels and electrophoresis to elongate DNA molecules, and thus enables localization of the bound proteins. In the first method,  $5 \mu\text{l}$  of LacI-DNA-TBE was sandwiched between a quartz chip and a thin quartz cover slip, which was sealed with nail polish.

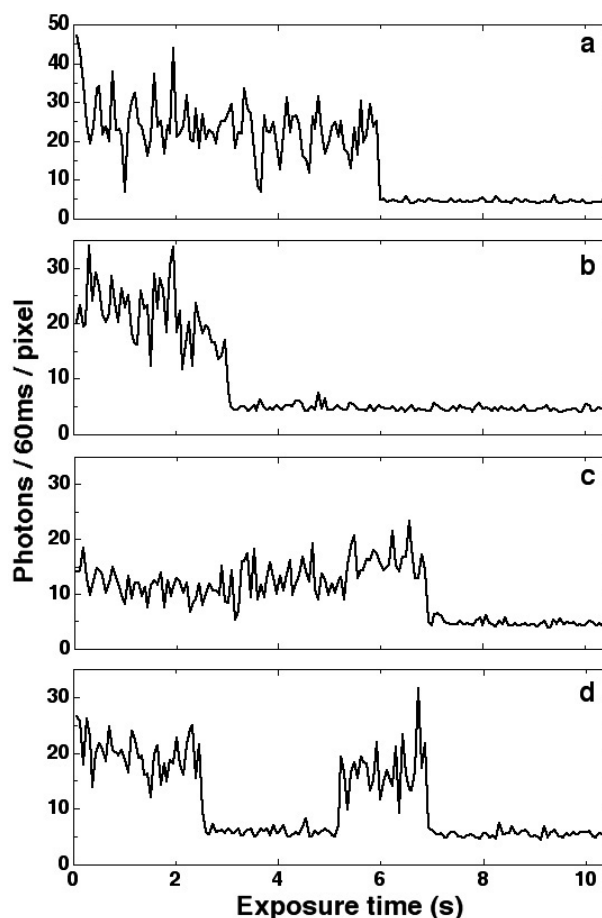
A number of protein molecules attached themselves to the quartz surface in addition to the DNA molecules, giving rise to a background of nonspecifically bound protein molecules in the image. By imaging the stuck proteins using total internal reflection fluorescence (TIRF) microscopy, we obtained the total number of photons emitted by the stuck proteins before bleaching occurs. The total photon count gives the number of proteins bound to DNA when normalized by the total number of photons emitted by a GFP–LacI dimer. Since this method sums the photons until bleaching occurs it is relatively insensitive to the laser illumination level from run to run as long as the optical components are unchanged.

In the second method, a platform was developed utilizing nanofabrication and TIRF microscopy to localize the bound protein along DNA and verify the relative uniformity of the binding. Microchannels one micron deep were fabricated using photolithography and reactive-ion etching; the nanochannels were fabricated using a focused ion beam (FIB) milling method into fused silica (amorphous quartz). Figure 2(b) shows a scanning electron microscopy (SEM) image of a 120 nm × 150 nm nanochannel. Holes were sand-blasted at the end of the microchannels (disc regions) for inputting the DNA. After the etching of micro- and nanochannels and the hole-drilling, the device was sealed with a thin fused-silica cover slip by a quartz–quartz bonding method. Reservoirs were then affixed over the holes at the four ends of the microchannels and a prism was placed on top of the nanochannel region for TIRF microscopy.

We first discuss the imaging and occupation determination of non-stretched DNA molecules, and begin with single GFP–LacI fusion proteins. Figure 3 shows a TIRF image of GFP–LacI fusion proteins attached to a quartz surface, in the form of fluorescence dots. This image is averaged over 10 frames. GFP–LacI was first suspended in 0.5 × TBE and 100 μg ml<sup>-1</sup> BSA to a final concentration of 4 nM. Then 5 μl LacI–TBE solution was sandwiched between a quartz chip and a thin quartz cover slip, which was sealed with nail polish. Some proteins stuck to the quartz surface and we imaged the stuck proteins using TIRF microscopy. The excitation power for this image is 300 W cm<sup>-2</sup>; the illumination area is 50 μm × 100 μm.

To determine the number of photons emitted by a single GFP–LacI molecule, the fluorescence counts/pixels received and read by the camera were converted into number of emitted photons. The total fluorescence count of a protein dot, which is typically a few pixels across and includes counts from both the protein and background, was first obtained. Then the background fluorescence count measured from non-fluorescent pixels adjacent to the dot was subtracted. Protein fluorescence counts were then converted into emitted photons using the photoelectron-to-digital unit-conversion factor of 16.72 count/e<sup>-</sup> (manufacturer specification) and the collection efficiency of our microscope/camera system of 6.4%.

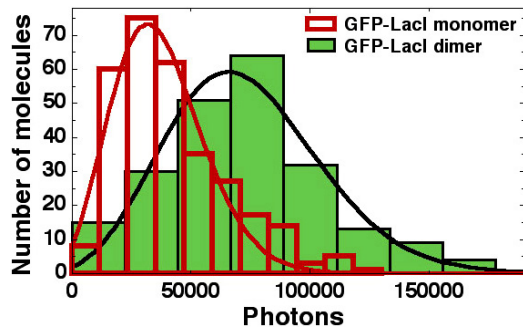
Representative fluorescence time traces of GFP–LacI monomer dots taken from 16.4 Hz measurements with continuous illumination are shown in figure 4. We observed frequent blinking events, which are sudden fluorescence dips to near noise level that usually last for the order of 100 ms, seen previously for GFP molecules [6, 12]. Each photon count value was the peak photon count of a fluorescence



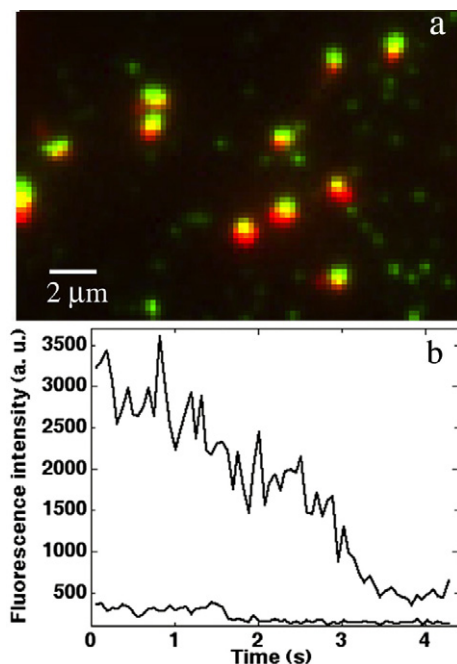
**Figure 4.** Peak fluorescence intensity versus time for single GFP–LacI monomers. The continuous illumination mode was used. The illumination intensity was 300 W cm<sup>-2</sup>. Each peak photon count value was the pixel-averaged value of the selected centre bright pixels of a fluorescence dot. The typical frequent blinking and slight decline in intensity followed by irreversible bleaching are shown in (a) and (b). Some molecules exhibited a fluorescence intensity increase after the initial decline, as in (c). Several molecules showed recovery from the photo-bleaching in seconds, and then eventually photo-bleached again, this time irreversibly (d). The background has a mean of 4.7 photons/pixel per 60 ms of exposure time.

dot. Analysis of approximately 100 GFP–LacI monomer dots shows that the majority of the molecules exhibited frequent blinking, and a slow decline in fluorescence intensity, followed by irreversible photo-bleaching as shown in figures 4(a), (b). Some molecules exhibited a fluorescence intensity increase after the initial decline (figure 4(c)). Several molecules showed recovery from photo-bleaching in seconds, and then eventually photo-bleached again, this time irreversibly (figure 4(d)).

Figure 5 compares the number of photons emitted by GFP–LacI monomers and dimers. Both distributions follow a Poisson distribution. The mean number of photons emitted by monomers is  $3.69 \times 10^4$ , and this agrees with the reported value of  $\approx 10^5$  photons [13, 14]. This value is the same with or without oxygen scavenging. The mean number of photons emitted by GFP–LacI dimers is  $7.32 \times 10^4$ , which is exactly twice that of GFP–LacI monomers. Thus it is clear that GFP–GFP fluorescence self-quenching is negligible in these experiments. From this result, we infer that the number of



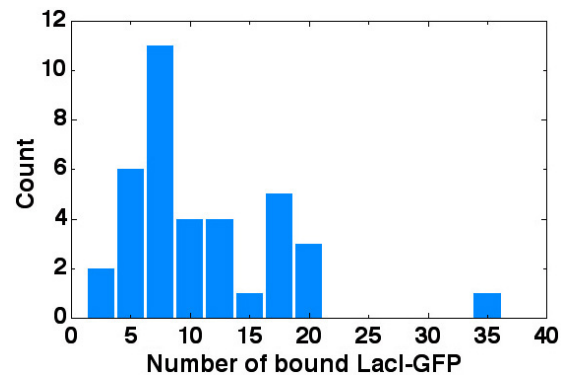
**Figure 5.** Histogram comparing total photons emitted in the lifetime of single GFP-LacI monomers and dimers. The lines are Poisson fits to the photon distributions: the values of the total emitted photons are  $(3.69 \pm 1.88) \times 10^4$  (mean  $\pm$  s.d.) for monomers and  $(7.32 \pm 3.19) \times 10^4$  for dimers. GFP-LacI dimers emit exactly twice as many photons as monomers, indicating that the GFP-GFP fluorescence self-quenching, if any, is negligible.



**Figure 6.** (a) GFP-LacI bound to LacO<sub>256</sub>-DNA attached to a quartz surface. Superimposed green and red dots are bound LacI-DNA molecules, while independent green dots are single unbound GFP-LacI molecules. (b) Fluorescence-intensity time traces of a bound GFP-LacI dot (solid line) and an unbound GFP-LacI dot (dashed line) measured using the two brightest pixels of each molecular image on the CDD. The intensity of the bound GFP-LacI dot declines continuously, obscuring the photo-bleaching events of individual bound GFP-LacI molecules. The scale represents the relative fluorescence intensity of the dots to that of the background, which is denoted as one in the figure. The represented GFP-LacI dot has  $\approx 20$  bound proteins.

photons emitted by GFP multimers scales linearly with the number of GFP molecules, and we use  $3.69 \times 10^4$  photons as the yield per GFP-LacI monomer in subsequent calculations.

GFP-LacI molecules bound to dsDNA were characterized in the same way as free GFP-LacI molecules, in which LacI-DNA solution was sandwiched between a quartz chip and a thin cover slip. Bound GFP-LacI stick to a quartz surface, whereas



**Figure 7.** Histogram for the number of bound GFP-LacI to LacO<sub>256</sub>-DNA collected from dots in the measurement of figure 6. The number of bound proteins ranges from 3 to 40.

DNA molecules do not, and thus the DNA arms tethered to the stuck GFP-LacI were seen diffusing freely. Time-averaged images of GFP-LacI and DNA were false coloured to be green and red, respectively, and superimposed so that the overlapping green and red dots indicate a bound LacI-DNA molecule. Figure 6 shows (a) a colour image of GFP-LacI bound to LacO<sub>256</sub>-DNA attached to a quartz surface and (b) fluorescence intensity time traces of a bound GFP-LacI dot and an unbound GFP-LacI dot. The fluorescence intensity versus time pattern of the bound GFP-LacI dot is different from that of an unbound one; it is a continuous declining curve devoid of sudden photo-bleaching events. The dot of bound GFP-LacI proteins in figure 6 corresponds to  $\approx 20$  bound proteins.

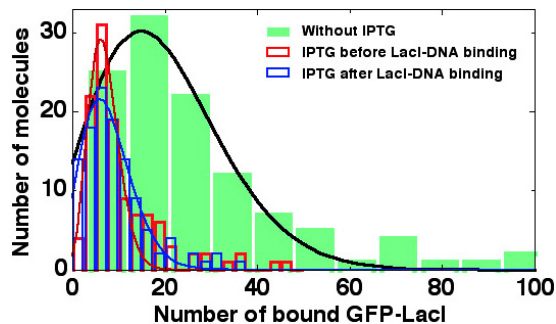
The number of bound GFP-LacI molecules was obtained using the integrated photon counting to bleach method discussed above, and the distribution of observed GFP-LacI molecules bound as observed for a sample of DNA molecules is plotted in figure 7. The concentration of GFP-LacI molecules in solution was 50 nM, and the concentration of binding sites, computed using the assumption that there are 512 monomer binding sites and 0.02 nM of dsDNA molecules in solution, was 10 nM. The number of bound proteins ranges from 4 up to order 100 with a mean of  $13 (\pm 6, \text{s.d.})$ , which yields a binding coefficient  $k_D$  of  $2.5 \mu\text{M}$ , far less than the expected binding  $k_D$  for dimers of approximately 10 nM [15, 16].

Figure 8 shows the actual occupation of binding sites observed at 10 nM GFP-LacI concentrations, and is to be compared to figure 1. When the GFP-LacI monomer/LacO ratio was increased to 15/1 by tripling the protein concentration so that the concentration of GFP-LacI was 150 nM, the number of bound GFP-LacI also increased approximately three times, indicating that we were far from saturating the stretch of specific binding sites, in spite of the known single transcription binding coefficient of 10–11 M.

We have also studied the effect of inducer IPTG on LacI-DNA interaction, in part to verify that the observed bindings are specific and also to test the heterogeneity of the binding as a function of occupation. It is generally believed that IPTG binds to LacI and releases it from its associated DNA. We measured the number of bound GFP-LacI molecules after adding 1 mM IPTG to the LacI-DNA-BOBO-3 solution; the mean number decreased by  $\approx 60\%$  from 19 to 8 (figure 9). This mean number of molecules is significantly higher than expected for



**Figure 8.** Actual occupancy level observed: only 2.5% of the available sites were bound.

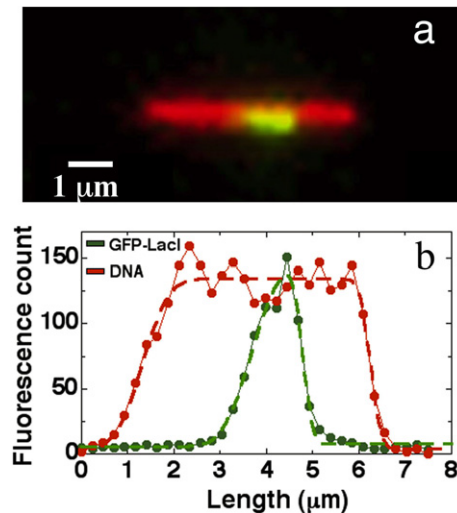


**Figure 9.** The number of GFP-LacI bound to *LacO*<sub>256</sub>-DNA for three different conditions: without IPTG, with 1 mM IPTG added after the LacI-DNA-BOBO-3 binding, and with 1 mM IPTG pre-incubated with GFP-LacI for one hour before adding DNA and BOBO-3 dye. The numbers of bound GFP-LacI are  $19 \pm 13$  (mean  $\pm$  s.d.),  $8 \pm 5$ , and  $7 \pm 3$ , respectively.

the 1 mM concentration of IPTG, where the bound proteins should mostly if not all be dissociated. This mean value is the same for measurements performed a few minutes to a few hours after adding IPTG. Premixing 1 mM IPTG with GFP-LacI for one hour before adding DNA gave a similar mean of 7 bound proteins. One possible explanation for this result of a high number of bound proteins is that the binding coefficient of LacI to a tandem array of binding sites is a function of the number of bound proteins  $N$  as we outlined at the start of this paper.

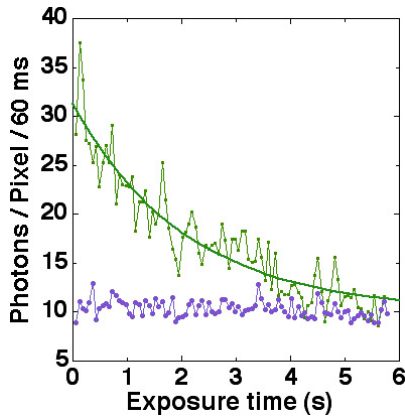
Thus, the data would indicate that the binding coefficient of the transcription factors at a 2% occupation factor is only of the order of  $10^{-6}$ , three orders of magnitude down from the single-factor binding coefficient. The nanochannels can provide significantly more information about what can be happening. For example, we know from the nanochannels that the transcription factors are basically uniformly distributed over the cassette region. Figure 10(a) is a time-averaged image of GFP-LacI bound to *LacO*<sub>256</sub>-DNA elongated in a nanochannel. There are  $\approx 20$  GFP-LacI bound to the operator sequence in this image. The fit for the DNA length gives the DNA length  $L_{zD} = 4.9 \mu\text{m}$ . The fractional centre location of the LacI-*LacO*<sub>256</sub> is 0.40, and agrees with that of *LacO*<sub>256</sub> of 0.32. The fractional length of the LacI-*LacO*<sub>256</sub> is  $L_{zp}/L_{zD} = 0.22$ , and agrees with that of *LacO*<sub>256</sub> of  $L_{c}/L_c = 9.2 \text{ kbp}/42.05 \text{ kbp} = 0.22$ . This result indicates that the 20 GFP-LacI proteins were distributed across the entire *LacO*<sub>256</sub> sequence. The length of the *LacO* segment is unaffected by the bound proteins, and thus we infer that the binding of 20 LacI does not obviously affect the global DNA properties, such as persistence length, since this would result in a change in the length of the section when elongated in a 100 nm nanochannel [17]. Further, since the length is unchanged we also predict that the transcription factors twist rather than bend the DNA.

The nanochannels can also provide another significant piece of information. The initial assessment of the occupation



**Figure 10.** (a) Time-averaged image of GFP-LacI bound to *LacO*<sub>256</sub>-DNA elongated in a  $150 \text{ nm} \times 200 \text{ nm}$  channel. There are 20 GFP-LacI molecules bound to this *LacO*<sub>256</sub>-DNA molecule. The right half of the protein-bound region is brighter; it contains ‘2.5’ more GFP-LacI than the left half. This molecule travels from right to left into the nanochannel driven by an electric field of  $5 \text{ V}/50 \mu\text{m}$ . (b) Fluorescence intensity profiles for the DNA and the bound GFP-LacI.

of the binding of transcription factors was done for DNA-protein complexes bound to a quartz surface. When two objects are fixed to a third immovable surface it is of course possible to introduce strain between these two fixed sites. Thus, one can ask if the observed low occupation numbers only occur when the DNA is fixed via protein binding to the quartz surface at several points, and additional proteins then introduce twist which is locked between these points. Note that this scenario does not negate the point of this paper that the LacI transcription factor twists the DNA when it binds, but it does propose that the decreased binding coefficient will not occur if the DNA is unconstrained. To test this, we carried out a photon counting measurement of the number of transcription factors bound to the DNA construct shown in figure 10. We again integrated the total photon counts until bleaching and divided by the photons/LacI to determine the number of bound LacI transcription factors bound to the *LacO*<sub>256</sub> region. Figure 11 shows the integrated photon counts for the *LacO*<sub>256</sub> region elongated in a  $120 \text{ nm} \times 120 \text{ nm}$  nanochannel. Figure 12 shows a histogram of the number of DNA molecules which had  $N$  bound transcription factors. We obtain, as we did for the case of unconstrained DNA molecules simply bound to the quartz surfaces, that the number of bound GFP-LacI molecules is only of the order of 10, far below the expected number. Since the only constraint for these molecules is a simple entropic constraint, it is unlikely that the earlier result for the quartz bound molecules is some artefact of surface binding.



**Figure 11.** The integrated bleaching of bound GFP-LacI molecules to LacO<sub>256</sub> in a  $120 \times 120 \text{ nm}^2$  channel. The lower trace is the background level.

Although we are faced with a puzzle, in that we do not understand how the strain is transmitted between remote transcription factor sites, we can still model a possible angle for the induced twist using the linear elasticity formalism we developed at the start of this paper. From equation (8), a decrease in binding coefficient  $\Delta k$  of  $10^{-3}$  requires an increase in the strain energy  $\Delta U$  if we assume each transcription factor has a fixed twist strain  $\Delta\Theta_o$ . At the addition of the  $n$ th factor, the added energy  $\Delta U$  for a spring constant  $k$  becomes

$$\Delta U = \frac{k}{2} [(n+1)\Delta\Theta_o]^2 - [n\Delta\Theta_o]^2 = k \left( n + \frac{1}{2} \right) \Theta_o^2. \quad (9)$$

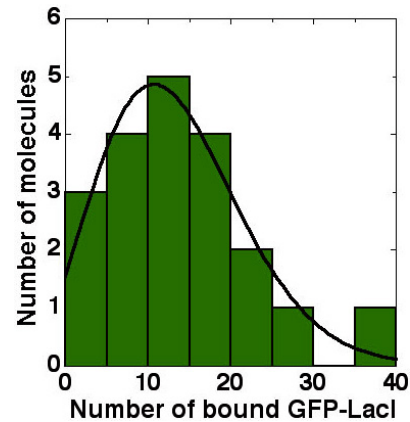
Thus, the binding constant  $K_n$  for the  $n$ th factor is

$$K_n = K_o \exp \left[ \frac{-k(n + \frac{1}{2})\Theta_o^2}{k_B T} \right]. \quad (10)$$

For  $n = 10$  we have  $\ln[K_n/K_o] = -7 = 11k\theta_0^2/k_B T$ . It is hard to go much further here without some speculation. If we assume that the effective length  $L$  of the torsional spring is given by the length of the  $\times 256$  cassette of binding LacI binding sites on the dsDNA, and assume that the shear modulus  $G$  is about  $10^9 \text{ dynes cm}^{-2}$ , we find that the predicted torsional strain  $\Delta\Theta_o$  due to the binding of a transcription factor is  $90^\circ$ . That is a rather high number, but not totally crazy. Of course, we had to make many assumptions here that are questionable, but perhaps the basic idea is clear from this exercise.

### 3. Conclusions

We have developed a new analytical method, the integration of total photons to bleach a single molecule. This technique, coupled with the analysis of dsDNA stretched by confinement in nanochannels, has allowed us to study the number of contiguously bound transcription factors in a sequence of 256 sites embedded in a genomic length of DNA. We have further shown that the binding coefficient of the contiguously bound transcription decreases with increasing occupation. Our results indicate that at 2% occupation of the transcription binding sites the binding coefficient has decreased by a factor close to 1000 from the single-factor binding coefficient. We



**Figure 12.** Histogram of the number of DNA molecules which have  $N$  bound GFP-LacI molecules to LacO<sub>256</sub> in a  $120 \times 120 \text{ nm}^2$  channel as determined from the integrated bleaching curves of figure 11.

have modelled this effect by assuming that the transcription factors twist the dsDNA as they bind, and that the addition of additional twist with each factor binding adds an elastic strain which decreases the binding coefficient. A simple linear strain model has been shown to fit the data by assuming that each duplex transcription factor adds  $90^\circ$  of twist as it binds. The fact that the effect is seen in DNA within nanochannels implies that some sort of a conformation change is being propagated along a significant length of DNA, of the order of 50 nm.

### References

- [1] Luscombe N, Austin S, Berman H and Thornton J 2000 *Genome Biol.* **1** 1–37
- [2] Hogan M, Legrange J and Austin B 1983 *Nature* **304** 752–4
- [3] White J H and Bauer W R 1986 *J. Mol. Biol.* **189** 329–41
- [4] Hearst J E and Hunt N G 1991 *J. Chem. Phys.* **95** 9322–8
- [5] Rousseau F and Schymkowitz J 2005 *Curr. Opin. Struct. Biol.* **15** 23–30
- [6] Pierce D W, Hom-Booher N and Vale R D 1997 *Nature* **388** 338
- [7] Gordon G S, Sitnikov D, Webb C D, Teleman A, Straight A, Losick R, Murray A W and Wright A 1997 *Cell* **90** 1113–21
- [8] Straight A F, Belmont A S, Robinnett C C and Murray A W 1996 *Curr. Biol.* **6** 1599–608
- [9] Robinnett C C, Straight A, Li G, Wilhelm C, Sudlow G, Murray A and Belmont A S 1996 *J. Cell Biol.* **135** 1685–700
- [10] Lewis M, Chan G, Horton N C, Kercher M A, Pace H C, Schumacher M A, Brennan R G and Lu P 1996 *Science* **271** 1247–54
- [11] Yang F, Moss L G and Phillips G N 1996 *Nat. Biotechnol.* **14** 1246–51
- [12] Dickson R M, Cubitt A B, Tsien R Y and Moerner W E 1997 *Nature* **388** 355–8
- [13] Kubitschek U, Kuckmann O, Kues T and Peters R 2000 *Biophys. J.* **78** 2170–9
- [14] Chiu C-S, Kartalov E, Unger M, Quarke S and Lester H A 2001 *J. Neurosci. Methods* **105** 55–63
- [15] Levandoski M M, Tsodikov O V, Frank D E, Melcher S E, Saecker R M and Record T M Jr 1996 *J. Mol. Biol.* **260** 697–717
- [16] Fumin D, Stefanie S, Olav Z, Brigitte K-W, Benno M-H and Andrew B 1999 *J. Mol. Biol.* **290** 653–66
- [17] Tegenfeldt J O *et al* 2004 *Proc. Natl Acad. Sci. USA* **101** 10979–83