NANO LETTERS

2009
Vol. 9, No. 4
1598-1603

Lighting Up Individual DNA Binding Proteins with Quantum Dots

Yuval Ebenstein,*,†,‡ Natalie Gassman,† Soohong Kim,† Josh Antelman,† Younggyu Kim,† Sam Ho,† Robin Samuel,† Xavier Michalet,† and Shimon Weiss*,†,‡,§

Department of Chemistry and Biochemistry, DOE Institute for Genomics and Proteomics, and Department of Physiology, Geffen Medical School, UCLA, Los Angeles, California 90095

Received December 18, 2008; Revised Manuscript Received February 23, 2009

ABSTRACT

The ability to determine the precise loci and occupancy of DNA-binding proteins is instrumental to our understanding of cellular processes like gene expression and regulation. We propose a single-molecule approach for the direct visualization of proteins bound to their template DNA. Fluorescent quantum dots (QD) are used to label proteins bound to DNA, allowing multicolor, nanometer-resolution localization. Protein−DNA complexes are linearly extended and imaged to determine the precise location of the protein binding sites. The method is demonstrated by detecting individual QD-labeled T7-RNA polymerases on the T7 bacteriophage genome. This work demonstrates the potential of this approach to precisely read protein binding position or, alternatively, "write" such information on extended DNA with QDs via sequence-specific molecular recognition.

Remarkable advancements to the study of transcription regulation[1−5] have occurred over recent years. New tools, and above all Chromatin immunoprecipitation (ChIP) based methods, have revealed a wealth of new information regarding the complex transcriptional networks underlying cell function. Despite these exciting methodological advancements and their capability for genome wide studies of gene expression, there is still a need for complementary approaches. New approaches should overcome some of the intrinsic drawbacks of existing methods (see Supporting Information for a brief survey of existing methods) and, more importantly, provide independent evidence regarding the state of the transcription machinery on the genome. Advances in high-resolution optical nanometer localization[6−8] and the growing toolbox for manipulating biological matter in the nanoscale open up new avenues for the study of protein−DNA interactions. In the following report, we present a single molecule approach for direct visualization of DNA binding proteins bound to their genomic targets and demonstrate the potential of this technique for genomic analysis of transcription.

In this method, optical imaging of DNA-binding proteins is utilized such that the precise location of the protein on the DNA is determined. High-resolution mapping of the DNA-binding proteins is achieved by combining far-field optics with DNA templates that have been extended with a uniform extension factor, creating arrays of DNA. This alignment simplifies the localization of objects along DNA strands and lends itself readily to automated image analysis for large scale, high-throughput measurements.[9] Two main approaches have been developed to align DNA on a solid support: molecular-combing, which utilizes the surface tension of a receding meniscus to align naked DNA anchored by its extremity on a surface,[10,11] and DNA alignment approaches that use flow and Stokes drag to perform the same feat.[12]

Optical visualization of the different proteins bound to the aligned DNA requires a robust optical contrast mechanism, such as fluorescence. Quantum dots (QDs), with their narrow, "tunable" emission spectra, provide an almost unlimited array of colors compared to organic fluorophores and therefore have ideal properties for the simultaneous observation of several components (multiplexing). Furthermore, QDs can be excited with a common excitation wavelength, reducing chromatic aberrations and simplifying microscopy. Finally, QDs are more photostable than organic fluorophores, facilitating nanometer localization of single QDs[8,13] (Figure 1in Supporting Information).

Here, we describe a technique using single QDs to identify the genomic targets of DNA-binding proteins, such as transcription factors (TFs). The principle of this approach is schematically depicted in Figure 1. First, proteins are cross-linked to DNA using standard in vivo based ChIP protocols.[2]

* Corresponding authors, uv@chem.ucla.edu and sweiss@chem.ucla.edu.
† Department of Chemistry and Biochemistry.
‡ DOE Institute for Genomics and Proteomics.
§ Department of Physiology, Geffen Medical School.

10.1021/nl803820b CCC: $40.75
Published on Web 03/16/2009

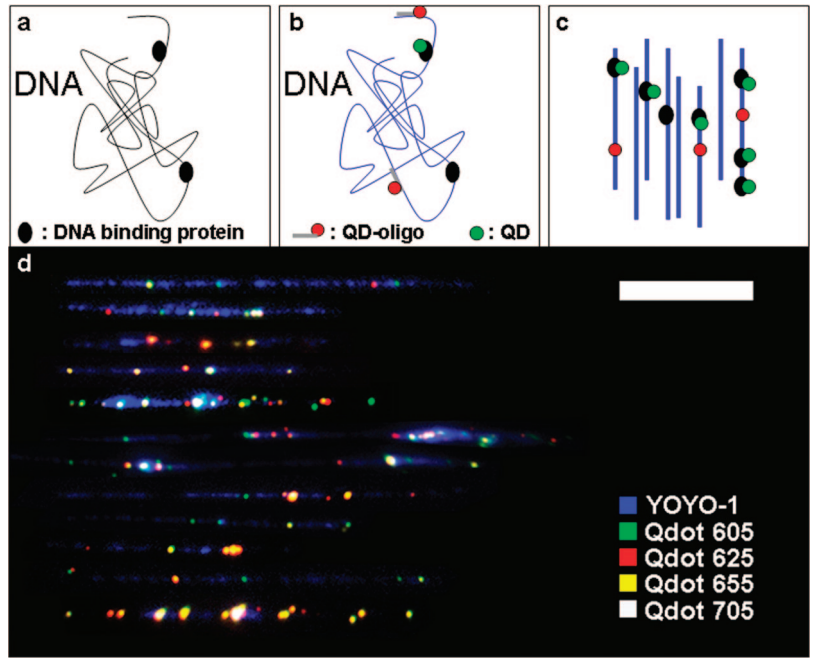

Figure 1. Experimental steps for mapping DNA binding proteins. (a) Cross-linking DNA-binding proteins to DNA. (b) Staining DNA (blue), QD labeling of bound proteins (green), and labeling of specific reference sequences on DNA with QDs (red). (c) Complexes are aligned on a glass coverslip and imaged by a fluorescence microscope. Image analysis provides information on protein location. (d) Cropped images of RNAP-biotin cross-linked to aligned DNA and bound to streptavidin QDs. DNA was stained with YOYO-1 and is displayed in blue while QDs emitting at 605, 625, 655, and 705 nm are displayed in green, red, yellow, and white, respectively. The sample was prepared as described in the Supporting Information except that RNAP was targeted by QDs of four different colors. (Scale bar 10 μm.)

Next, the cross-linked complexes are labeled with QD-conjugated antibodies against TFs of interest, while the DNA is stained with an intercalating dye. Practically, four to six different TFs can be labeled with different QDs emitting in the visible spectrum and still be easily spectrally separated, as demonstrated in Figure 1d. Additionally, one or more spectrally distinct QDs could be used to tag a reference point in the genome in a sequence specific manner, using cDNA[14,15] or peptide nucleic acid (PNA)[16,17] probes. The QD-labeled protein−DNA complexes are then aligned on a glass coverslip, resulting in linear DNA molecules decorated with QD-labeled TFs. For optical imaging, a single excitation source is sufficient to excite the fluorescence of the DNA stain and all QDs. The potential multiplexing capability of this approach and the fact that it probes multiple proteins on long stretches of DNA (10s to 100s of kb) addresses some of the limitations of traditional ChIP based methods that rely on data from an ensemble of fragmented (<1 kb) DNA molecules and thus are limited in providing information regarding cooperative binding events and long-range interactions of DNA binding proteins like transcription factors.

We demonstrate the method with a proof-of-principle experiment aimed at locating a single QD-labeled protein species bound to multiple genomic targets. As a model biological system, we studied T7-RNA polymerase (RNAP) binding to promoters of the T7 bacteriophage genome. T7 is a virus that infects most *Escherichia coli* strains. It has a fully sequenced linear double-strand DNA genome containing 39936 bp, with 17 identified promoter binding sites for T7-RNAP.[18]

To facilitate easy and selective binding of QDs to T7-RNAP, a biotinylation tag was introduced at the N-terminus of the RNAP and this tag was in vivo biotinylated prior to purification. In vitro binding of the biotinylated-RNAP to the T7 genome was performed as described.[19] Since T7-RNAP binds weakly to DNA,[20] transcription was initiated with a nucleotide subset (GTP, UTP, and CTP) to stabilize the complex. The sample was then labeled with the intercalating dye YOYO-1 and with streptavidin conjugated QDs.

We first aligned the sample by molecular combing, which reliably produces uniform stretching.[10] Combing of DNA with cross-linked RNAP using this method was successfully verified by atomic force microscopy (AFM). However, when combing the QD-labeled complexes, we could only detect one QD at the end of the observed DNA molecules (Figure 2 in Supporting Information.). We hypothesized that one of the bound QDs served to anchor the DNA to the surface, via nonspecific binding, and that the stretching force created by the receding meniscus was sufficient to detach the remaining T7-RNAP-QDs from the DNA.[21] To overcome these limitations, we utilized a flow-induced stretching technique.[12]

A 3 μL droplet containing ~5 ng of DNA was deposited at the interface between two parallel coverslips, one of which was functionalized with polylysine. Capillary force immediately drew the droplet in between the coverlips, stretching the DNA and immobilizing it on the polylysine surface, presumably by electrostatic interaction (Figure 3 in Supporting Information). We found that adding 0.1% of *n*-dodecyl-β-D-maltoside, a very mild nonionic surfactant,[22] im-

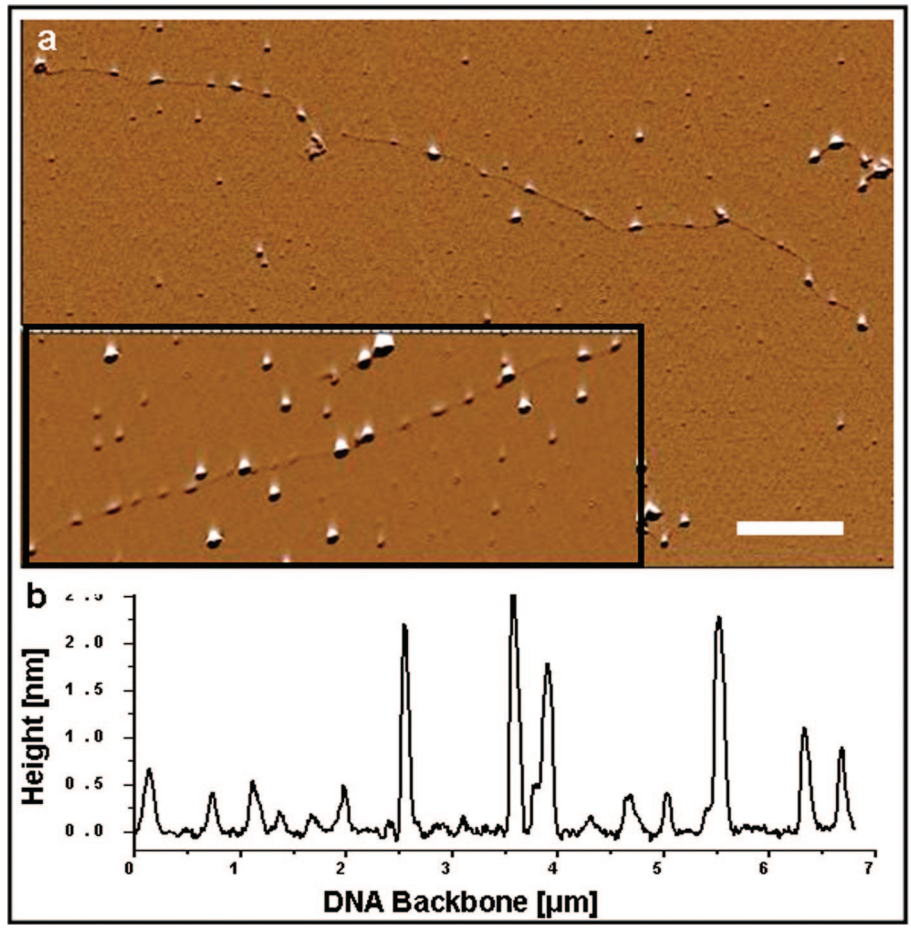

Figure 2. Evaluation of RNAP and QD binding on the T7-phage genome. (a) Amplitude AFM scans of the T7- DNA-RNAP-QD complex, before (inset) and after heparin treatment. Protrusions of two typical sizes, corresponding to the RNAP (0.5 nm) and RNAP-QD (>1.5 nm) are visible on the DNA backbone. (Scale bar 1 μm.) (b) Cross section along the DNA located in the inset, depicting the height of the various protrusions.

proved the extension uniformity and reduced nonspecific binding of free QDs to the polylysine surface.[23] The resulting fluorescence images clearly show individual QD fluorescent spots on many of the DNA molecules (Figure 3 in Supporting Information). AFM imaging of the sample on freshly cleaved mica indicates that most promoter sites were occupied by T7-RNAP (Figure 2). DNA fibers decorated with protrusions of two typical sizes were observed. We attributed smaller protrusions to bare T7-RNAP and the larger protrusions to T7-RNAP-QD.

We first evaluated the uniformity and degree of extension by measuring the length of QD-labeled DNA molecules. A histogram of 70 such molecules from several fields of view is presented in Figure 3, together with some representative images of full length genomes. We found that DNA was extended to 86 ± 6% of its B-DNA contour length (3.42 bp/nm). To determine the location of bound QDs, we measured their distance from the DNA ends and corrected for variations in extension factor by dividing each distance by the total length of the individual DNA molecule. Since this sample did not contain any marker to indicate DNA orientation, QDs are matched to known promoters by comparing their measured positions to theoretical promoter sites and choosing the orientation resulting in the best global

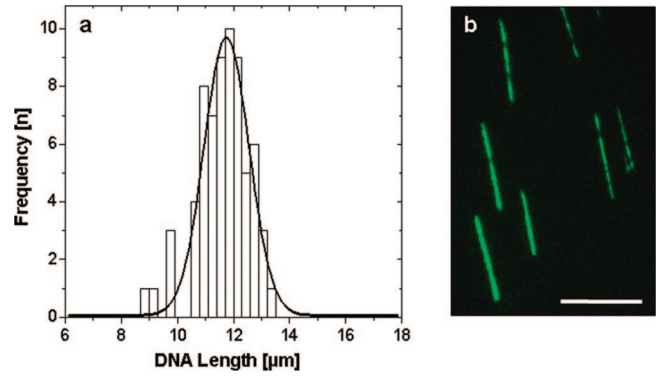

Figure 3. Evaluation of extension uniformity. (a) Histogram constructed from the end-to-end length of 70 T7- DNA-RNAP-QD complexes. A Gaussian fit to the data is centered at 11.75 μm corresponding to 86% of the B-DNA contour length. The width of the distribution is 1.6 μm corresponding to about 13%. (b) Full-length T7-phage genomes, flow-stretched on a polylysine-coated coverslip (Scale bar: 13 μm.)

match. To demonstrate the mapping accuracy of the optical measurement, we plot the mean position and standard deviation of detected QDs against the known promoter sites as shown in Figure 4. Representative genome images, scaled to span the full promoter map, are presented, and QD labels may be traced to their matching promoters. Agreement is

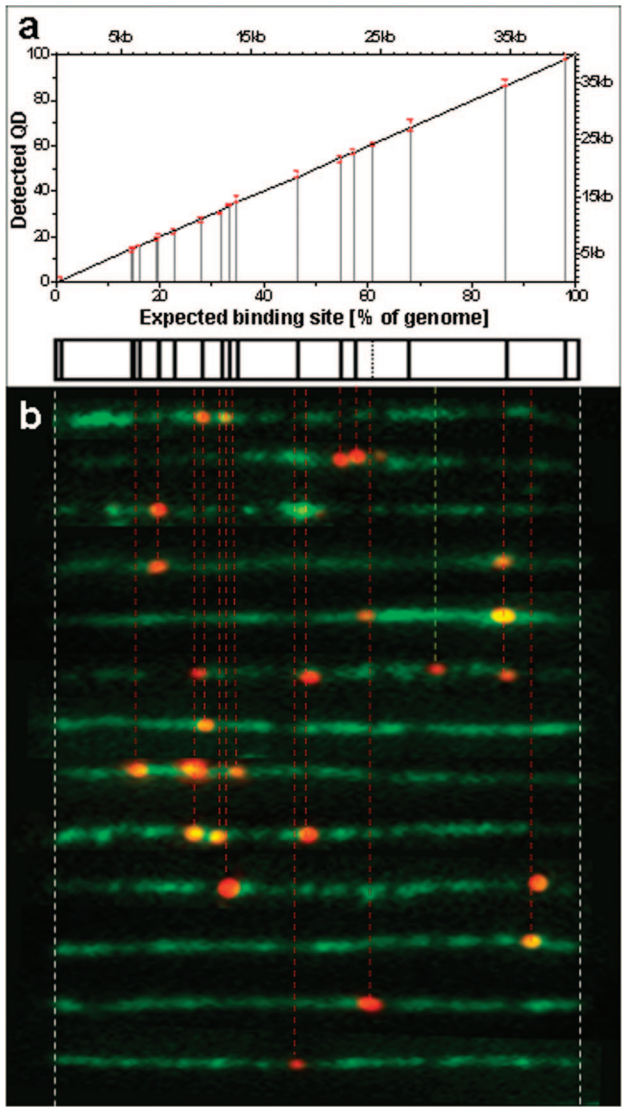

Figure 4. Evaluation of mapping precision. (a) Mean position of detected QDs (red) plotted against their assigned promoter location (dashed gray lines). The black line is a guide to the eye representing $y = x$. Error bars represent standard deviation of the mean. (b) Cropped fluorescence images of full length T7-phage genomes, scaled to span the full promoter map. White dashed lines indicate the DNA extremities. Red and yellow dashed line are guides for the eye and connect QD binding locations on different molecules to the promoter map.

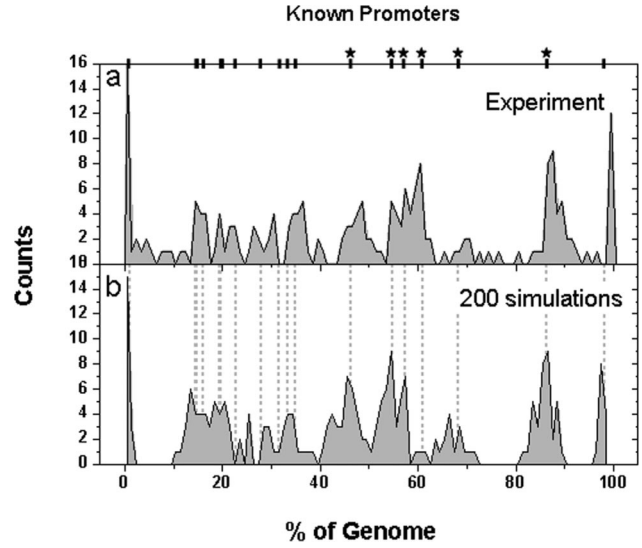

Figure 5. Evaluation of promoter occupancy. (a) Histogram of the number of binding events constructed from 150 genomes having two identifiable ends. Data are presented in normalized percentage units going from left to right in the direction of transcription. The distance between a bound QD and the DNA end was normalized to units of % of the total genome by taking the ratio of the distance and the total length. (bin = 1% [400 bp]). (b) Histogram of the number of binding events constructed from 200 simulated genomes. The simulation parameters were chosen to reproduce the experimental results as follows: extension factor variation = 5%, detection efficiency = 30%, binding efficiency of class III (consensus) and replication (end) promoters was chosen to be 3 times larger than that of class II promoters. To reproduce the experimental results, binding efficiencies 4 and 5 times larger than that of class II promoters were assigned to the high occupancy promoters at 47% and 86%, respectively (bin = 1%). Top axis and dashed gray lines indicate the location of the 17 known promoters. (Terminator and class III are labeled with a star.)

excellent, with 87% ($N = 199$) of QDs within 1kb, 50% within 398 bp and 25% within 174 bp of a promoter. This correlation indicates that QD positions are not random and represent mostly specific binding events of RNAP under our experimental conditions.

An important feature of this method is its ability to determine the relative occupancy of binding sites under given conditions. The position histogram in Figure 5 shows the frequency of QD positions along the genome relative to the positions of known RNAP promoters. The distribution is clearly discrete, with a considerable degree of correlation between promoters and QD clusters.

An encouraging finding is that 3 times more binding events were detected in regions corresponding to consensus, or tight, binding sites relative to binding sites with nonconsensus sequences. Overall, we have detected 90 QDs near the 5 class III promoters bearing the consensus sequence giving an average occupancy of 18 QD/promoter, while an average of only 5.8 QDs were detected near the 10 weaker class II promoters.[24] Most notable is the high occupancy of the region corresponding to promoters located at 86% and 46.4% of the full genome length, as well as that of a known strong terminator at 61%. A numerical simulation of our experiment showed good agreement with our observation, provided we assigned a higher binding probability for these promoters compared to other consensus promoters (Figure 5). We could find no reported explanation for this higher occupancy, suggesting that this observation could be of biological importance. This simple experiment thus indicates that our method may yield insightful results even in relatively well-known systems such as T7.

The results described above assumed that the DNA extension was uniform within each molecule. As shown by our simulations, this is probably not the case, as the observed binding site distribution is best accounted for by a nonuniform extension factor within each molecule. In principle, however, the influence of a nonuniform extension factor could be significantly reduced by introducing two or more sequence-specific labels (genomic flags) with defined spacing

serving as internal calibration. Sequence specific flags will be instrumental for the analysis of more complex samples such as bacterial or mammalian DNA. Such long genomes are likely to be fragmented and flags can serve as reference points for orientation and distance measurements. Preliminary assessment of the resolution improvement expected from such flags shows a 3−4-fold improvement in resolution upon addition of two sequence specific reference flags (as described in the Supporting Information section and presented in Table 1 in the Supporting Information). Several strategies for sequence specific labeling of DNA have been reported[17,25] and are currently tested in our laboratory for assembling QDs on double-stranded DNA in a sequence specific manner.

In conclusion, we effectively demonstrated the viability of this single molecule approach to directly visualize and map protein binding sites on DNA using QDs. At this stage, the method is applicable for the analysis of well-defined DNAs, such as intact viral genomes, linearized fosmid DNA, or uniform, clean, bacterial or mammalian genomic fragments. This simple concept has the advantage of rapid and simple sample preparation, as well as requiring orders of magnitude less material than, for instance, ChIP assays. The use of QDs offers several additional advantages as well. Since single QDs and stained DNA molecules are bright enough to be observed with the naked eye under the microscope, an inexpensive camera is sufficient for image acquisition. The use of a single excitation source simplifies data acquisition even further, allowing multicolor images to be acquired in a single shot using a color camera (Figure 3 in Supporting Information). The ability to analyze independent DNA fragments opens up a whole array of single-molecule capabilities such as: the identification of correlated or cooperative binding events on single DNA molecules or the identification of rare subpopulations masked by ensemble averaging using other techniques. Another advantage of this approach, in contrast to ChIP-on-a-chip, or other methods based on shotgun sequencing, is that it can handle DNA repeats, genomic regions that account for 30% of the human genome which are inaccessible to sequencing and array based methods.

The addition of genomic flags as well as utilization of high-throughput technologies such as microfluidics[26] and machine vision[27] could result in a high-resolution, high-throughput, single-molecule transcription factor mapping tool that may allow direct visual comparison of transcription states between genomes. Obviously, this approach could also be used to precisely and easily organize QDs on DNA molecules for nanofabrication purposes.

Acknowledgment. This work was supported by the UCLA-DOE Institute for Genomics and Proteomics (Grant DE-FC02-02ER63421) and NIH Grant R01-EB000312. Y.E. is supported by a Human Frontier Science Program Cross disciplinary Fellowship.

Note Added after ASAP Publication: This paper was published ASAP on March 16, 2009. With approval from the editorial office, Josh Antelman was removed from the Acknowledgment and added as a contributing author. The revised paper was reposted on March 24, 2009.

Supporting Information Available: A survey of existing technologies for transcription mapping, detailed methods and supporting figures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kirmizis, A.; Farnham, P. J. Genomic Approaches That Aid in the Identification of Transcription Factor Target Genes. *Exp. Biol. Med.* **2004**, *229* (8), 705–721.
(2) Massie, E. C.; Mills, G. I. ChIPping away at gene regulation. *EMBO Rep.* **2008**, *9* (4), 337–343.
(3) Elnitski, L.; Jin, V. X.; Farnham, P. J.; Jones, S. J. M. Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques. *Genome Res.* **2006**, *16* (12), 1455–1464.
(4) Simonis, M.; Kooren, J.; de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **2007**, *4* (11), 895–901.
(5) Maerkl, S. J.; Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **2007**, *315* (5809), 233–237.
(6) Michalet, X.; Lacoste, T. D.; Weiss, S. Ultrahigh-resolution colocalization of spectrally separable point-like fluorescent probes. *Methods* **2001**, *25* (1), 87–102.
(7) Lacoste, T. D.; Michalet, X.; Pinaud, F.; Chemla, D. S.; Alivisatos, A. P.; Weiss, S. Ultrahigh-resolution multicolor colocalization of single fluorescent probes. *Proc. Natl. Acad. Sci.U.S.A.* **2000**, *97* (17), 9461–9466.
(8) Thompson, R. E.; Larson, D. R.; Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **2002**, *82* (5), 2775–2783.
(9) Jing, J.; Reed, J.; Huang, J.; Hu, X.; Clarke, V.; Edington, J.; Housman, D.; Anantharaman, T. S.; Huff, E. J.; Mishra, B.; Porter, B.; Shenker, A.; Wolfson, E.; Hiort, C.; Kantor, R.; Aston, C.; Schwartz, D. C. Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules. *Proc. Natl. Acad. Sci.U.S.A.* **1998**, *95* (14), 8046–8051.
(10) Allemand, J. F.; Bensimon, D.; Jullien, L.; Bensimon, A.; Croquette, V. pH-dependent specific binding and combing of DNA. *Biophys. J.* **1997**, *73* (4), 2064–2070.
(11) Michalet, X.; Ekong, R.; Fougerousse, F.; Rousseaux, S.; Schurra, C.; Hornigold, N.; Slegtenhorst, M. v.; Wolfe, J.; Povey, S.; Beckmann, J. S.; Bensimon, A. Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* **1997**, *277* (5331), 1518–1523.
(12) Meng, X.; Benson, K.; Chada, K.; Huff, E. J.; Schwartz, D. C. Optical mapping of lambda bacteriophage clones using restriction endonucleases. *Nat. Genet.* **1995**, *9* (4), 432–438.
(13) Lagerholm, B. C.; Averett, L.; Weinreb, G. E.; Jacobson, K.; ThompsonN. L. Analysis method for measuring submicroscopic distances with blinking quantum dots. *Biophys. J.* **2006**, *91* (8), 3050–3060.
(14) Keren, K.; Berman, R. S.; Buchstab, E.; Sivan, U.; Braun, E. DNA-templated carbon nanotube field-effect transistor. *Science* **2003**, *302* (5649), 1380–1382.
(15) Bentolila, L. A.; Weiss, S. Single-step multicolor fluorescence in situ hybridization using semiconductor quantum dot-DNA conjugates. *Cell Biochem. Biophys.* **2006**, *45* (1), 59–70.
(16) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991**, *254* (5037), 1497–1500.
(17) Chan, E. Y.; Goncalves, N. M.; Haeusler, R. A.; Hatch, A. J.; Larson, J. W.; Maletta, A. M.; Yantz, G. R.; Carstea, E. D.; Fuchs, M.; Wong, G. G.; Gullans, S. R.; Gilmanshin, R. DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. *Genome Res.* **2004**, *14* (6), 1137–1146.
(18) Dunn, J. J.; Studier, F. W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **1983**, *166* (4), 477–535.
(19) Gueroui, Z.; Place, C.; Freyssingeas, E.; Berge, B. Observation by fluorescence microscopy of transcription on single combed DNA. *Proc. Natl. Acad. Sci.U.S.A.* **2002**, *99* (9), 6005–6010.
(20) Ikeda, R. A.; Richardson, C. C. Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proc. Natl. Acad. Sci.U.S.A.* **1986**, *83* (11), 3614–3618.

(21) Bensimon, D.; Simon, A. J.; Croquette, V.; Bensimon, A. Stretching DNA with a Receding Meniscus: Experiments and Models. *Phys. Rev. Lett.* **1995**, *74* (23), 4754.

(22) Stubbs, G. W.; Smith, H. G.; Litman, B. J. Alkyl glucosides as effective solubilizing agents for bovine rhodopsin. A comparison with several commonly used detergents. *Biochim. Biophys. Acta* **1976**, *426* (1), 46–56.

(23) Huang, B.; Wu, H.; Kim, S.; Zare, R. N. Coating of poly(dimethylsiloxane) with n-dodecyl-[small beta]-d-maltoside to minimize nonspecific protein adsorption. *Lab Chip* **2005**, *5* (10), 1005–1007.

(24) Smeekens, S. P.; Romano, L. J. Promoter and nonspecific DNA binding by the T7 RNA polymerase. *Nucleic Acids Res.* **1986**, *14* (6), 2811–2827.

(25) Jo, K.; Dhingra, D. M.; Odijk, T.; de Pablo, J. J.; Graham, M. D.; Runnheim, R.; Forrest, D.; Schwartz, D. C. A single-molecule barcoding system using nanoslits for DNA analysis. *Proc. Natl. Acad. Sci.U.S.A.* **2007**, *104* (8), 2673–2678.

(26) Dimalanta, E. T.; Lim, A.; Runnheim, R.; Lamers, C.; Churas, C.; Forrest, D. K.; dePablo, J. J.; Graham, M. D.; Coppersmith, S. N.; Goldstein, S.; Schwartz, D. C. A Microfluidic System for Large DNA Molecule Arrays. *Anal. Chem.* **2004**, *76* (18), 5293–5301.

(27) Lim, A.; Dimalanta, E. T.; Potamousis, K. D.; Yen, G.; Apodoca, J.; Tao, C.; Lin, J.; Qi, R.; Skiadas, J.; Ramanathan, A.; Perna, N. T.; Plunkett, G., 3rd; Burland, V.; Mau, B.; Hackett, J.; Blattner, F. R.; Anantharaman, T. S.; Mishra, B.; Schwartz, D. C. Shotgun optical maps of the whole Escherichia coli O157:H7 genome. *Genome Res.* **2001**, *11* (9), 1584–1593.

(28) He, B.; Rong, M.; Lyakhov, D.; Gartenstein, H.; Diaz, G.; Castagna, R.; McAllister, W. T.; Durbin, R. K. Rapid mutagenesis and purification of phage RNA polymerases. *Protein Expression Purif.* **1997**, *9* (1), 142–151.

(29) Ebenstein, Y.; Nahum, E.; Banin, U. Tapping mode atomic force microscopy for nanoparticle sizing: tip-sample interaction effects. *Nano Lett.* **2002**, *2* (9), 945–950.

NL803820B