Single-Molecule Detection of Transcription Factor Binding to DNA in Real Time: Specificity, Equilibrium, and Kinetic Parameters

Eric A. Nalefski,* Eugene Nebelitsky, Janice A. Lloyd, and Steven R. Gullans‡

U.S. Genomics, 12 Gill Street, Suite 4700, Woburn, Massachusetts 01801

Received January 31, 2006; Revised Manuscript Received June 21, 2006

ABSTRACT: Specificity and temporal control of transcriptional machinery are encoded within sequence-specific transcription factors, of which there are thousands in mammalian genomes. Efforts to completely decipher this code will require an understanding of the DNA binding thermodynamic and kinetic properties displayed by each transcription factor, a daunting task given the current methodologies for measuring these interactions. Here, we present a novel methodology to quantify the binding of proteins to target DNA molecules based on single-molecule detection and real-time counting of individual free and bound fluorescently tagged molecules flowing past a detection device. Using this technology, we measured DNA binding by fluorescently tagged domains of four distinct transcription factors, namely, human early growth response protein Egr-1, vertebrate GATA-1, *Drosophila* GAGA factor, and λ bacteriophage Cro repressor. These proteins represent different structural classes (zinc-finger and helix−turn−helix), quaternary states (monomeric and dimeric), and relative affinities (high, intermediate, and low). Specific binding of each protein to its cognate DNA target was demonstrated at low picomolar concentrations. The equilibrium (K_d) and kinetic (k_{on} and k_{off}) constants governing DNA binding by one of these transcription factors, that of Egr-1, were measured using this approach. K_d values obtained from three different types of saturation titrations were reproducible and consistent, yielding values between 10 and 14 pM that, along with the kinetic constants, agree closely with literature values. Because this methodology offers several significant advantages over other existing approaches, namely, real-time determination, requirement for small amounts of reagents, high reproducibility, exquisite sensitivity, and amenability to high-throughput analysis, it is suitable for characterizing DNA-binding proteins as well as other interacting pairs of molecules that can be fluorescently tagged.

Transcription factors represent the largest class of proteins of known single function in the human genome (*1*). A large majority of these consists of polypeptides containing two distinct functional domains, one responsible for targeting gene-specific regulatory sites by directly binding DNA (i.e., the DNA-binding domain) and a second domain for transcriptional activation (*2*). Of all known and predicted transcription factors, the specific DNA sequences that they recognize are known for only a small fraction (i.e., <10%) (*3*). Determining the identity of the specific DNA sequences recognized by all transcription factors would greatly expand our fundamental understanding of the function of these proteins. However, an even deeper layer of understanding could be provided by characterizing the thermodynamic and kinetic constants that each transcription factor exhibits for binding its target DNA, because these parameters reflect the specificity and temporal control of the transcriptional machinery encoded within the proteins. The establishment of such a comprehensive and valuable body of knowledge regarding transcription factor binding has not been feasible to date because of limitations of current methods and technologies.

Of the several methodologies that have been developed to measure interactions between proteins and DNA, many of these, including the electrophoretic mobility shift assay (EMSA)[1] (*4*), nitrocellulose filter binding (*5*), phage display immunoassays (*6*), and microarrays (*7*, *8*), measure the binding in heterogeneous assays, where signals are generated after physical separation of free reactants from bound products. In these approaches, the time to separate free from bound species is considerable or involves other unwanted disruptions of equilibrium conditions. Other spectroscopic methodologies, such as fluorescence (*9*) and circular dichroism (*10*), can offer real-time detection of binding in homogeneous assays without physical separation of reactants and products; however, these approaches either require large amounts of materials or are not sensitive enough to quantitate high DNA-binding affinities. Surface plasmon resonance (SPR) also provides real-time detection of protein−DNA complexes, but the interpretation of data may be complicated by the requirement for immobilization of one of the reactants,

* To whom correspondence should be addressed. E-mail: enalefski@usgenomics.com.

‡ Present address: Rx-Gen, Inc., 100 Deepwood Drive, Hamden, CT 06517.

[1] Abbreviations: EMSA, electrophoretic mobility shift assay; SPR, surface plasmon resonance; ZFD, zinc-finger domain; Egr, early growth response; GFP, green-fluorescent protein; GST, glutathione *S*-transferase; BFL, BODIPY FL; EDTA, ethylenediaminetetraacetic acid; Tris, 2-amino-2-(hydroxymethyl)-1,3-propanediol; DTT, dithiothreitol; EBS, Egr-1-binding site; OR3, third right operator; BSA, bovine serum albumin; APD, avalanche photodiode.

10.1021/bi0602011 CCC: $33.50
Published on Web 10/31/2006

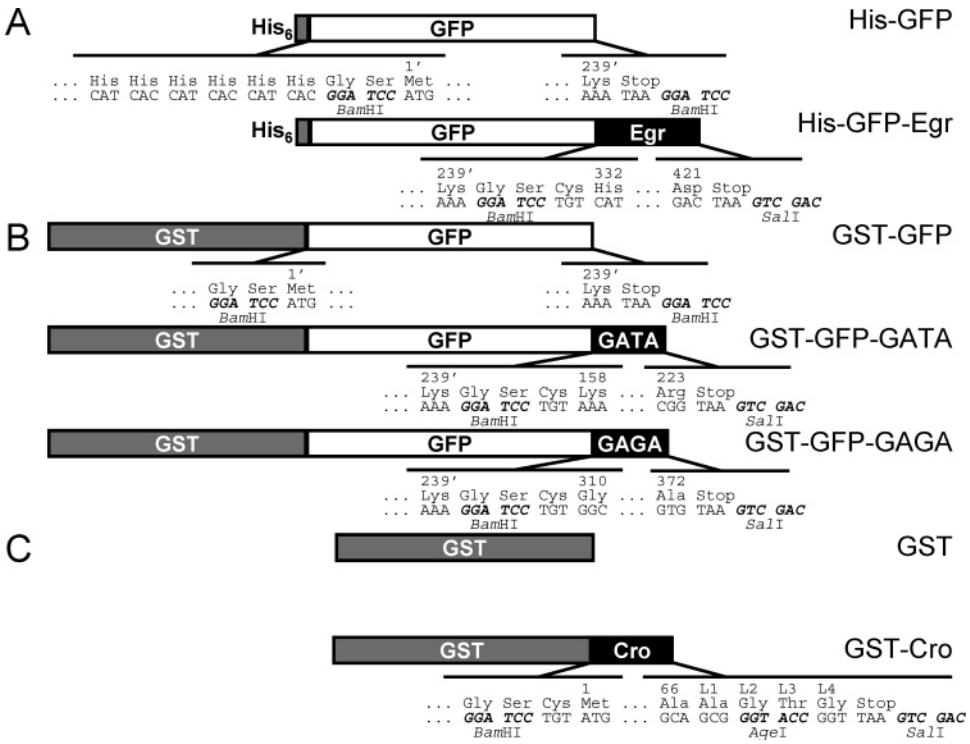

FIGURE 1: Schematic of recombinant proteins. (A) His-tagged proteins fused to GFP: the control His-GFP, 250 amino acid product of a GFP variant tagged with N-terminal His_6, and His-GFP-Egr, 343 amino acid product of zinc-finger DNA-binding domain from Egr-1 fused to an N-terminal His-GFP. (B) GST-tagged proteins fused to GFP: the control GST−GFP, 464 amino acid product of GFP variant tagged with N-terminal GST, GST−GFP−GATA, 533 amino acid product of zinc-finger DNA-binding domain from GATA-1 fused to an N-terminal GST−GFP, and GST−GFP−GAGA, 530 amino acid product of zinc-finger DNA-binding domain from the GAGA factor fused to an N-terminal GST−GFP. (C) GST-tagged proteins for chemical modification: the control GST, 239 amino acid GST, and GST−Cro, 297 amino acid product of λ Cro repressor and C-terminal artificial linker residues (L1−L4) fused to an N-terminal GST. Restriction sites used in DNA construction are shown in bold italics. GFP residues are indicated with primed numbers according to their position in the native protein; DNA-binding domain residues are indicated with nonprimed numbers from the native protein (Table 1).

Table 1: Structural and Function Properties of Transcription-Factor DNA-Binding Domains Chosen for This Study

transcription factor	domain residues[a]	structural class	quaternary state[b]	recognition target (bp)[c]	reported K_d (pM)
Egr-1	332−421	3 zinc-fingers	monomer	10	10−12[d]
					125−170[e]
GATA-1	158−223	zinc-finger	monomer	6	8300[f]
GAGA	310−372	zinc-finger	monomer	7	5200[g]
Cro repressor	1−66	helix−turn−helix	dimer	17	2.0[h]

[a] Numbering of the DNA-binding domain in the native protein. [b] Quaternary state of the protein when bound to DNA. [c] Length of recognition targets in base pairs based on the DNA footprint observed in three-dimensional structures. [d] Range of values from EMSA using Egr-1 ZFD in the *absence* of poly(dI/dC) (*28*−*30*). [e] Range of values from EMSA using Egr-1 ZFD in the *presence* of poly(dI/dC) (*18*, *37*). [f] Result from EMSA using isolated GATA-1 ZFD in low ionic-strength buffer (*19*). [g] Result from EMSA using isolated GAGA factor ZFD in intermediate ionic-strength buffer (*38*). [h] Result from the filter-binding assay using the full-length Cro repressor (*21*).

usually the DNA (*11*). Despite these shortcomings, these methodologies have been the foundation of previous efforts to discover the principles by which transcription is regulated.

Recent advances in single-molecule detection (*12*) have provided the opportunity for the development of an alternative assay technology for the analysis of protein−DNA complexes. In this paper, we employed a single-molecule counting approach to quantify interactions between several different proteins and target DNA (Table 1). The results presented indicate that this novel single-molecule-based methodology offers a sensitive, reproducible, and accurate approach for quantifying equilibrium and kinetic constants governing interactions between proteins and DNA and, by extension, between any pairs of molecules that can be fluorescently tagged.

MATERIALS AND METHODS

Design of Recombinant Proteins. A schematic representation of recombinant proteins is provided in Figure 1. DNA-binding domains from the zinc-finger domains (ZFDs) were amplified using template DNA encoding: mouse early growth response (Egr) protein Egr-1 (*13*), which encodes a protein identical to that in humans, GenBank accession number J04089, purchased from ATCC; chicken GATA-1 (*14*), I.M.A.G.E. Consortium CloneID 439605, purchased from Invitrogen; and GAGA factor (*15*), *Drosophila* cDNA clone LD30706, purchased from Invitrogen. The following sense and anti-sense primers were used: Egr-1, CGC*GGATCC*TGTCATGAACGCCCATATGCTTGC and CGCA*GTCGAC***TTA**GTCCTTCTGTCTTAAATGGAT; GATA-1, CGC*GGATCC*TGTAAACGGGCAGTCACCCAATGC

and GCGT*GTCGAC***TTA**CCGCTTCTTTTTCCCTTTGCC; GAGA factor, CGC*GGATCC*TGTGGCATTACAAGCG-GCGTCCTATCC and GCGT*GTCGAC***TTA**CACGCCGGG-TTTGGCAAAATG. In these primers, the italic font represents *Bam*HI and *Sal*I sites and the bold font corresponds to an engineered stop codon.

To fuse the Egr-1 ZFD to the green-fluorescent protein (GFP), first the amplification product of Egr-1 was digested with *Bam*HI and *Sal*I and ligated into *Bam*HI and *Sal*I sites of pQE-30 (Qiagen) downstream of His_6 codons to generate pQE-30-Egr. Second, DNA-encoding GFP was amplified from pGreenTIR (*16*), purchased from ATCC, using primers CGC*GGATCC*ATGAGTAAAGGAGAAGAACTTTTCAC (5′ GFP) and CGC*GGATCC*TTTGTATAGTTCATCCAT-GCCATGTGTAATC (3′ GFP), where the italic font corresponds to *Bam*HI sites. Finally, this amplification product was digested with *Bam*HI and ligated into the *Bam*HI site of pQE-30-Egr between His_6 and the Egr-1 ZFD to yield pQE-30-GFP-Egr. To generate a His-GFP fusion, the GFP gene was first amplified using 5′ GFP and the anti-sense primer CGC*GGATCC***TTA**TTTGTATAGTTCATCCATG-CCATGTGTAATC (3′ GFP stop), where the italic font corresponds to a *Bam*HI site and the bold font corresponds to a stop codon. This amplification product was then digested with *Bam*HI and ligated into the *Bam*HI site of pQE-30-Egr between His_6 and the Egr-1 ZFD to yield pQE-30-GFP.

GFP-tagged GATA-1 and GAGA factor fragments were fused to a schistosomal glutathione *S*-transferase (GST). First, DNA-encoding GFP was amplified using the sense primer GGA*AGATCT*ATGAGTAAAGGAGAAGAACTT-TTCAC, where the italic font corresponds to a *Bgl*II site, and the anti-sense primer 3′ GFP. Second, the amplification product was digested with *Bgl*II and *Bam*HI and ligated into the *Bam*HI site in pGEX-4T1 (Amersham Biosciences) downstream of GST to yield pGEX-4T1-GFP-*Bam*HI. Finally, the amplification products of GATA-1 and GAGA were digested with *Bam*HI and *Sal*I and ligated into the *Bam*HI and *Sal*I sites of pGEX-4T1-GFP-*Bam*HI downstream of GFP to yield pGEX-4T1-GFP-GATA and pGEX-4T1-GFP-GAGA, respectively. To construct DNA-encoding GST−GFP with an appropriate stop codon, the GFP gene was amplified using the primers 5′ GFP and 3′ GFP stop, and the amplified product was digested with *Bam*HI and ligated into the *Bam*HI site of pGEX-4T1 downstream of GST to yield pGEX-4T1-GFP.

GST and the Cro repressor fused to GST (GST−Cro) were expressed in *Escherichia coli* and served as targets for extrinsic fluorophore attachment with BODIPY FL (BFL). The Cro repressor gene was first amplified with *Pfu* DNA polymerase (Promega) from template λ DNA (Invitrogen) using the 5′ sense and 3′ anti-sense primers CGCGGATC-CTGTATGGAACAACGCATAACCCTG and GCGTGTC-GAC**TTA***ACCGGT*ACCCGCTGCTGTTGTTTTT-TTGTTACTCGGG, respectively. In these two primers, the bold font corresponds to a stop codon, underlined sequences correspond to *Bam*HI and *Sal*I sites, respectively, and the the italic font corresponds to an *Age*I site. Second, the amplification product was digested with *Bam*HI and *Sal*I and subsequently ligated into the *Bam*HI−*Sal*I sites of pGEX-4T-1 downstream of GST to yield pGEX-Cro.

The sequences of all transcription-factor DNA inserts were verified by sequencing using the ThermoSequenase Cy5.5 dye terminator cycle sequencing kit (Amersham Biosciences).

Protein Expression, Purification, and Fluorescent Labeling. DH5α cells were transformed with plasmid DNA and grown in baffled Erlenmeyer flasks at 37 °C in Luria broth containing 100 μg/mL ampicillin to an OD_{600} of 0.6. After the induction of protein expression by the addition of isopropylthio-β-galactoside to a final concentration of 1 mM, cells were grown for 4 h. Bacteria were harvested by centrifugation at 6000*g*, and expressed protein was purified at 4 °C as follows. For His_6-tagged proteins, the cell pellet was resuspended in an ice-cold buffer composed of 300 mM NaCl, 50 mM NaH_2PO_4, and 10 mM imidazole at pH 8.0 containing ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor tablets (Roche). Lysozyme was added to a final concentration of 1 mg/mL; cells were lysed by sonication on ice; and cell lysates were cleared by centrifugation at 16000*g*. The resulting supernatant was bound to nickel−nitrilotriacetic acid resin, and the adsorbed resin was washed exhaustively with lysis buffer substituted with 20 mM imidazole at pH 8.0. Protein was eluted with buffer containing 250 mM imidazole at pH 8.0, dialyzed extensively against elution buffer lacking imidazole at 4 °C, and concentrated in a stirred-cell apparatus using a YM 10 membrane (Millipore).

For GST-tagged proteins, the cell pellet was resuspended in ice-cold buffer composed of 150 mM NaCl, 20 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris) at pH 8.0, and 1 mM EDTA (STE) containing a protease inhibitor cocktail (Sigma). The resulting supernatant was bound to glutathione−agarose, and the adsorbed resin was exhaustively washed with STE buffer. The protein was eluted with STE buffer supplemented with 10 mM reduced glutathione, extensively dialyzed against STE buffer at 4 °C, and concentrated.

GST and GST−Cro were labeled by reacting 10 μM protein with 100 μM BFL maleimide (Molecular Probes) in STE buffer supplemented with 100 μM of the reductant Tris-(2-carboxyethyl)phosphine hydrochloride (Pierce). Reactions were carried out for 1 h at room temperature in the dark and quenched by the addition of 10 mM dithiothreitol (DTT). Labeled protein was separated from free fluorophore by passing the reaction mixture over a Superdex 200 HR size-exclusion column (Amersham Biosciences) run in buffer containing 200 mM NaCl, 20 mM Tris at pH 8.0, and 1 mM EDTA. Absorbance was monitored at 280 and 508 nm. Fractions exhibiting absorbances at both wavelengths were subjected to SDS−PAGE analysis, and labeled protein was visualized by placing unstained gels on a UV transilluminator. Fractions containing labeled protein devoid of visible free fluorophore were pooled, concentrated, snap-frozen, and stored at −80 °C.

Concentrations of purified proteins were estimated by comparing silver-stain intensities of samples resolved by SDS−PAGE with purified GST or GST−GFP stocks, whose concentrations were determined spectroscopically using their extinction coefficients. Small aliquots of purified proteins were snap-frozen, stored at −80 °C, and used only once immediately after thawing.

Preparation of Synthetic DNA Duplexes. DNA targets for the proteins consisted of short synthetic DNA duplexes

formed from oligonucleotides purchased from Integrated DNA Technologies. Oligonucleotides were either unmodified or 3′-labeled with the dye quencher dabcyl or the fluorophore Cy5, which exhibits spectral properties (λ_{ex} = 643 nm, and λ_{em} = 667 nm) that are suitable for detection in the red channels of our single-molecule detection device (*17*). Duplexes were comprised of the following sequences, where known target sequences are underlined: Egr-1-binding site (EBS), 27-mer artificial site AGCAGCTGAGCGTGGGCGTAGTGAGCT (*18*); GATA, 16-mer target for GATA-1 from the chicken β/ϵ-globin enhancer GTTGCAGATAAACATT (*19*); GAGA, 20-mer target for the GAGA factor from the *Drosophila* histone h3/h4 promoter AAACCCGAGAGAGTACGAAC (*20*); and the third right operator (OR3), 21-mer target for the Cro repressor of the λ phage genome TCTATCACCGCAAGGGATAAA (*21*).

To form duplexes, mixtures of complementary oligonucleotides were denatured at 95 °C for 2 min and then allowed to cool to room temperature. Duplexes were separated from single-stranded DNA by electrophoresis on 15% polyacrylamide gels run in 0.5× Tris borate EDTA (TBE buffer), excised, and electroeluted from gel slices into 1× Tris acetate EDTA buffer using a Model 422 Electro-Eluter apparatus (Bio-Rad Laboratories). Duplexes were then concentrated under vacuum, ethanol-precipitated, and resuspended in 10 mM Tris at pH 8.0 and 1 mM EDTA. Label concentrations were determined using molar extinction coefficients of $\epsilon_{454\ nm}$ equal to 32 000 cm^{-1} M^{-1} for dabcyl and $\epsilon_{643\ nm}$ equal to 250 000 cm^{-1} M^{-1} for Cy5. After we corrected for the contribution of Cy5 and dabcyl absorbance to that observed at 260 nm, duplex concentrations were determined and label stoichiometry was calculated. In all cases, dabcyl or Cy5 stoichiometries ranged between 0.7 and 1.4 moieties/duplex. All purified duplexes were demonstrated to be free of single-stranded DNA by electrophoresis on 15% polyacrylamide gels in 0.5× TBE buffer and visualization with ethidium bromide.

Measurement of Synthetic Duplex Binding by Conventional Fluorescence Spectroscopy. Bulk fluorescence measurements were carried out at 20 °C in thermostatted methacrylate cuvettes using a Fluorolog-3 (Instruments S.A., Inc.) or a QuantaMaster (Photon Technology International) fluorescence spectrophotometer. Excitation and emission slits were both set at 5.0 nm. Reactions consisted of 1 nM (final) protein and DNA duplexes in assay buffer. Concentrations of duplexes are indicated in the figure captions. For Egr-1 ZFD binding, the assay buffer consisted of 100 mM NaCl, 10 mM Tris at pH 7.5, 10 mM DTT, 10 μM $ZnSO_4$, and 250 μg/mL fatty-acid-free bovine serum albumin (BSA). For GATA and GAGA ZFDs, the latter buffer was used with NaCl reduced to 10 mM. For Cro repressor binding, the assay buffer consisted of 100 mM NaCl, 10 mM Tris at pH 8.0, 10 mM DTT, 1 mM EDTA, and 250 μg/mL BSA. Reactants were pump-mixed, and samples were allowed to equilibrate at least 15 min prior to recording data. For spectral recordings, λ_{ex} was set at 490 nm and emission was scanned from 500 to 550 nm. Assay buffer spectra were subtracted from all recordings, and spectra were normalized to the maximal fluorescence intensity observed prior to the addition of DNA.

For kinetic measurements, λ_{ex} was set at 490 nm and emission was recorded at 508 nm, respectively, using an assay buffer supplemented with 0.01% NP40. A lag time

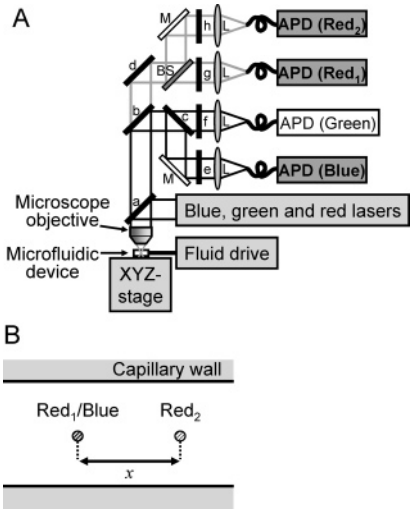

FIGURE 2: Single-molecule detection and counting system. (A) Schematic of the single-molecule counting apparatus. Three laser beams are focused through a microscope objective into the center stream line of a stage-mounted microcapillary through which fluid is driven. Blue channel (488 nm) and red_1 channel (633 nm) laser light are focused into the same spot for coincidence detection, whereas the red channel light from a second laser (red_2) is focused downstream in the microcapillary for flow velocity determination. Emitted light passing back through the objective is collected and focused onto optical cables leading to three avalanche photodiodes (APDs), corresponding to blue, red_1, and red_2, via a series of multichroics and band-pass filters (a−h), beam splitters (BS), and lenses (L). (B) Schematic view of the 50 μm wide microcapillary channel showing the direction of molecule flow serially through laser spot 1, where the blue and red_1 lasers were focused, and spot 2, where the red_2 laser was focused at a distance x downstream of spot 1.

accounting for mixing was added to the beginning of each time course. For a graphical comparison to the results obtained using single-molecule detection, the extent of fluorescence quenching was plotted against time; i.e., a decrease in quenching translates to enhanced fluorescence emission. Data were least-squares-fitted with the three-parameter monoexponential equation

$$F_{(t)} = F_0 + A_{obs}\, e^{(-k_{off}t)} \quad (1)$$

where $F_{(t)}$ represents the observed fluorescence at time t, F_0 is a fluorescence offset representing the final fluorescence, A_{obs} represents the amplitude, and k_{off} is the observed rate constant. Curve fitting was performed with Origin version 7.0 software (OriginLab Corporation).

Single-Molecule Detection. Most recordings were carried out on a prototype of the U.S. Genomics Trilogy Single Molecule Analyzer (Figure 2A), described in detail elsewhere (*17*), at 10 or 40 kHz. In this device, two laser spots were positioned in the center streamline of the microfluidic sample handling device (Figure 2B), separated by 15−20 μm. The 488 nm laser light, which excites both BFL and GFP, and 633 nm laser light, which excites Cy5, were focused into the first spot for detection in the blue and red_1 channels, respectively, to generate binding data. The 633 nm laser light was directed into the downstream spot for detection in a second red channel (red_2), which, in combination with channel red_1, enabled flow velocity determination using cross-correlation analysis as described in the Supporting Information (*17*). Where indicated, recordings were also

carried out on the commercially available U.S. Genomics Trilogy 2020 Single Molecule Analyzer at 10 kHz. This device differs from the prototype primarily in that photons are collected with a CCD camera from 10 × 1 μm laser stripes placed perpendicular to capillary flow. In this device, blue and red laser stripes were separated by 2 μm.

For equilibrium experiments, the protein was mixed with DNA in the appropriate binding buffer (see above) for at least 1 h at room temperature before taking readings. The sampling capillary was rinsed briefly with water between samples. Kinetic experiments were carried out on the Trilogy 2020 instrument, because of its shorter dead times, using an assay buffer supplemented with 0.01% NP40. For association kinetics time courses, the protein was pump-mixed with Cy5-labeled DNA and then read immediately. For dissociation time courses, the protein was first allowed to equilibrate with Cy5-labeled DNA, mixed with excess unlabeled DNA, and then analyzed. A lag time was added to the start of the recorded time courses to account for mixing. In all cases, final reactant concentrations are indicated in the figure captions.

Analysis of Single-Molecule Trace Files. Coincident event counting, which yields a linear dynamic response up to 500 pM of dual-labeled molecules, and spatial cross-correlation analysis and counting were carried out as described in the Supporting Information (*17*). Single-molecule trace files consist of fluorescence values captured in bins, where 1 bin represents data integrated over 100 μs for 10 kHz data acquisition. In event counting, the number of bound complexes is represented by the number of nonrandom coincident bins remaining after subtracting the estimated random number from the total observed and is expressed as coincident bins/s. Where necessary, temporal offsets were applied between fluorescence channels as bins, and the resulting signal responses from coincident bin counting and cross-correlation analysis were calculated. For the prototype instrument, blue and red_1 laser spots were physically overlapped, obviating the need for offsetting.

For equilibrium binding measurements, three separate data files, ranging in duration from 10 to 40 s, were averaged. From saturation titrations with 100% Cy5-labeled DNA, the number of bound protein molecules (B), recorded in bins/s, was plotted against the estimated concentration of free DNA (x) and least-squares-fitted with a standard two-parameter, single-site binding equation

$$B_{(x)} = B_{\max}\left(\frac{x}{x + K_d}\right) \tag{2}$$

where B_{max} represents the number of bound protein molecules at saturation and K_d represents the equilibrium dissociation constant. The free DNA concentration was estimated as the product of the nominal DNA concentration and the fraction of total red_1 bins attributed to DNA not observed in protein−DNA complexes. From saturation titrations with mixtures of Cy5-labeled and unlabeled DNA, the number of bound molecules was plotted against the estimated concentration of free Cy5-labeled DNA and fitted with a modified form of eq 2, where K_d is substituted with an *apparent* dissociation constant K_d^{app}. In this case, K_d^{app} corresponds to nK_d, where n equals the fraction of DNA molecules labeled with Cy5. A third fitting parameter, B_0, representing the low background coincidence signal observed in the absence of Cy5-labeled DNA (see the inset of Figure 7C), was added to eq 2 in tracer ligand titrations. Errors in calculations were propagated in a stepwise manner (*22*). For equilibrium competitive inhibitor titrations, the number of bound protein molecules was plotted against the concentration of unlabeled competitor DNA (x) and fitted with a standard two-parameter competition equation

$$B_{(x)} = B_{\max}\left(1 - \frac{x}{x + IC_{50}}\right) \tag{3}$$

where B_{max} represents the number of bound protein molecules in the absence of the competitor and IC_{50} represents the concentration of competitor that results in a 50% loss of bound molecules. The equilibrium dissociation constant of the competitive inhibitor (K_i) was evaluated by (*23*)

$$K_i = \frac{IC_{50}}{1 + \frac{[L]}{K_d}} \tag{4}$$

where [L] is the estimated concentration of free Cy5-labeled DNA and K_d is taken from direct ligand titrations.

For kinetic experiments, consecutive slices of data lasting 5 or 10 s were analyzed and the number of bound protein molecules at time t was plotted against time. Time courses were fitted with a variation of the three-parameter monoexponential decay equation, eq 1, where B_{max}, representing the number of bound protein molecules that decay, substitutes for A_{obs} and B_f, representing the final number of bound protein molecules in the presence of the competitor, substitutes for F_0. In association time courses, data were fitted with an alternative form of a two-parameter monoexponential equation

$$B_{(t)} = B_{\max}(1 - e^{(-k_{obs}t)}) \tag{5}$$

where B_{max} represents the number of bound protein molecules at the final equilibrium and k_{obs} represents an observed first-order rate constant governing the approach to equilibrium. Because the assay was carried out under pseudo-first-order conditions in which the free DNA concentration remained constant throughout the time course, k_{obs} was plotted as a function of the DNA concentration and fitted with the linear equation

$$k_{obs} \approx k_{on}[DNA] + k_{off} \tag{6}$$

where k_{on} represents the apparent second-order association rate constant and k_{off} represents the dissociation rate constant.

RESULTS

Design, Production, and Fluorescent Labeling of Recombinant Proteins. Fluorescently tagged recombinant DNA-binding polypeptides were constructed (Figure 1) to test whether their interactions with DNA could be quantified by single-molecule counting. Several members of the ZFD family of transcription factors were chosen as representative monomeric DNA-binding proteins (Table 1). The mammalian early growth response protein Egr-1 (also known as zif268, NGFI-A, and krox-24) binds DNA with relatively high

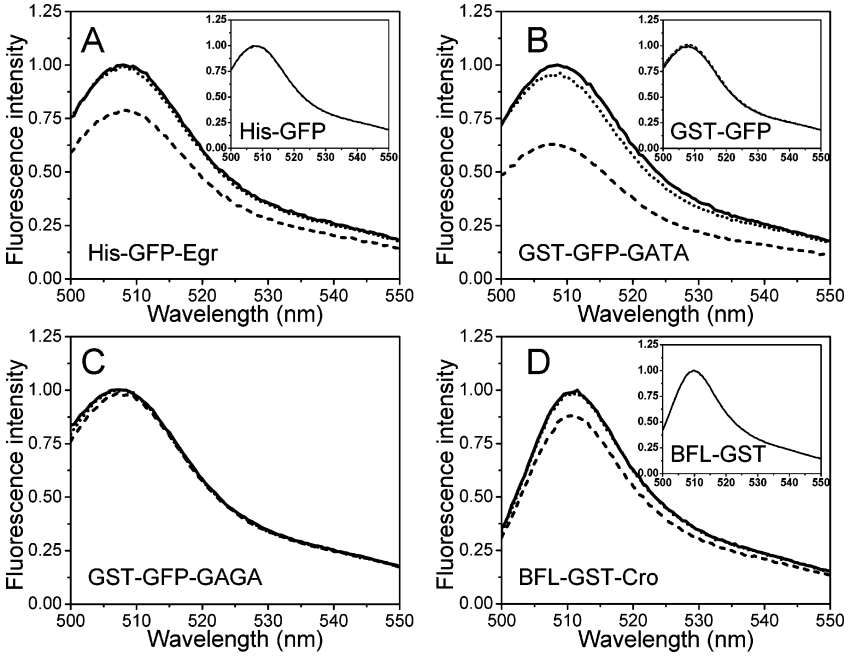

FIGURE 3: DNA-binding activity measured with conventional bulk fluorescence spectroscopy. Fluorescence emission spectra of (A) His-GFP-Egr and its control His-GFP (inset), (B) GST−GFP−GATA, (C) GST−GFP−GAGA and their control GST−GFP (inset to B), and (D) BFL-labeled GST−Cro and its control BFL−GST (inset). Spectra were collected before (—) or after the addition of the following dabcyl-labeled DNA duplexes: (A) EBS (- - -) or OR3 (···), (B and C) GATA (- - -) or GAGA (···), and (D) OR3 (- - -) or EBS (···). Buffer-subtracted spectra were normalized to the maximal emission intensity of the protein in buffer. Experimental conditions: 1 nM protein and 1 nM duplex, with $\lambda_{ex} = 490$ nm.

affinity, forming direct contacts over 10 bp of DNA through three tandem zinc-finger repeats (*24*). The DNA-binding ZFDs of chicken GATA-1 and *Drosophila* GAGA factor bind DNA with lower affinity, forming direct contacts spanning only 6−8 bp of DNA mediated through single ZFDs and C- and N-terminal basic extensions, respectively (*19*, *20*). The bacteriophage λ Cro repressor protein was chosen as a representative dimeric helix−turn−helix transcription factor, which binds DNA with very high affinity by forming direct contacts spanning 17 bp of DNA (*25*).

Two different methods for fluorescently labeling the recombinant proteins and two different affinity tags for protein purification were employed. Intrinsically fluorescent forms of the Egr-1, GATA, and GAGA ZFDs were constructed by fusing their DNA-binding domains in frame to the C terminus of a phenolate anion variant of GFP (*26*), whose spectral characteristics ($\lambda_{ex} = 488$ nm, and $\lambda_{em} = 507-509$ nm) match that of organic fluorophores detected in the blue channel of our single-molecule analysis instrument. Alternatively, the fusion construct encoding the GST-tagged Cro repressor was covalently labeled on cysteines with a maleimide conjugate of BFL, a small organic fluorophore ($\lambda_{ex} = 504$ nm, and $\lambda_{em} = 510$ nm) also detected in the blue channel. DNA-binding polypeptides were affinity-purified on the basis of adsorption to either nickel resin via a His_6 tag fused to the N terminus of GFP-tagged Egr-1 ZFD or glutathione resin via GST tags fused to N termini of the Cro repressor and GFP-tagged GATA and GAGA ZFDs.

Functional Characterization of Recombinant Proteins by Conventional Bulk Fluorescence Spectroscopy. Fluorescence spectroscopy of bulk protein solutions was used to characterize the DNA-binding activities of the purified recombinant proteins. GFP-tagged and BFL-labeled fusion proteins exhibited emission spectra characteristic of GFP and BFL, respectively (Figure 3). Binding of synthetic DNA duplexes was tested by measuring the ability of duplexes labeled with the quencher dabcyl to quench GFP or BFL emission, because the excitation spectrum of dabcyl overlaps with the emission spectrum of these fluorophores. The addition of equimolar concentrations of a dabcyl-labeled DNA containing the Egr-binding site, dabcyl-EBS, to His-GFP-Egr resulted in the quenching of GFP emission intensity by ~20% (Figure 3A). Because this quenching was not observed in the control His-tagged GST (inset of Figure 3A), the quenching of His-GFP-Egr depends upon the presence of the Egr-1 ZFD. Furthermore, the addition of a labeled duplex containing an unrelated recognition sequence, dabcyl-OR3, failed to elicit GFP quenching, indicating that DNA binding exhibited by His-GFP-Egr is specific for its recognition sequence. Because no further quenching was observed upon incubation with additional dabcyl-EBS under these conditions (data not shown), the dissociation constant for binding must be considerably less than 1 nM, indicating that the affinity of the Egr-1 ZFD for the EBS duplex is relatively high, a finding consistent with detailed equilibrium analysis (see below).

Functional properties of the intrinsically fluorescent GST−GFP-tagged GATA and GAGA ZFDs were also tested by conventional fluorescence spectroscopy. GFP emission in GST−GFP−GATA (Figure 3B) but not the control GST−GFP (inset of Figure 3B) was specifically quenched by ~40% upon addition of equimolar amounts of dabcyl-GATA duplex but only ~5% by a dabcyl-GAGA duplex, indicating that GST−GFP−GATA preferentially binds DNA containing the GATA recognition sequence. The reaction with additional dabcyl-GATA resulted in further quenching of GST−GFP−GATA, indicating that under these conditions binding had not reached saturation and that the interaction is of lower affinity than that between Egr-1 and its target (data not

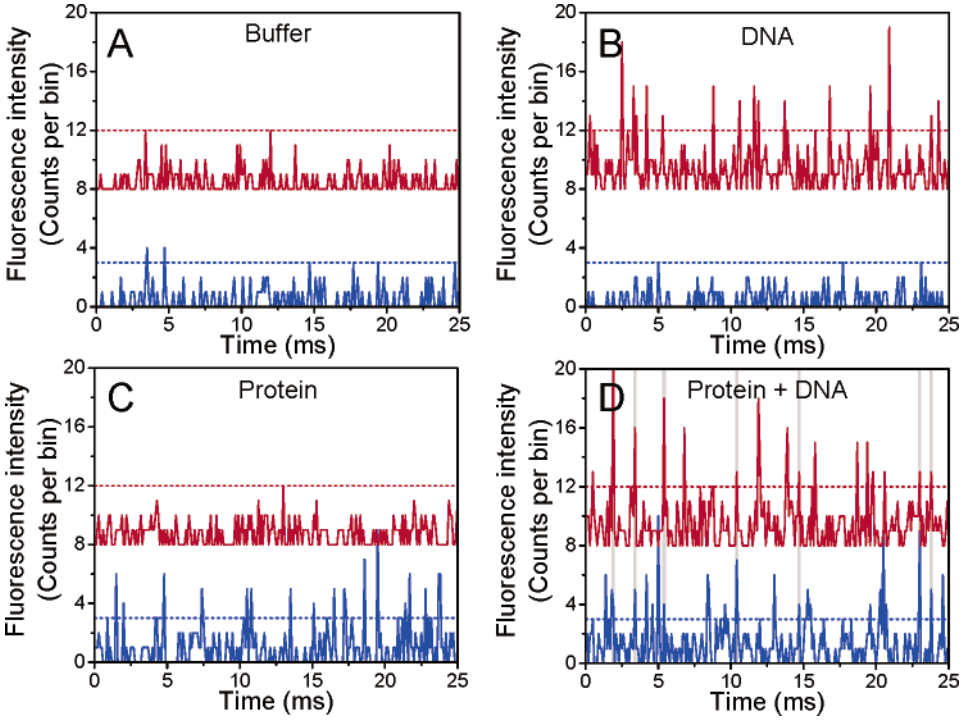

FIGURE 4: Representative single-molecule trace files. Recordings made over 25 ms for buffer only (A), Cy5-EBS DNA (25 pM) alone (B), His-GFP-Egr (~50 pM total protein) alone (C), or a mixture of His-GFP-Egr (~50 pM) and Cy5-EBS DNA (25 pM) (D). Blue channel fluorescence (protein) is shown by the blue trace, and red_1 channel fluorescence (DNA), artificially offset by eight counts, is shown by the red trace. Dashed horizontal lines indicate threshold settings employed to illustrate coincident events. Grey vertical lines in D indicate seven protein peaks detected above the threshold value that were identified by software to be coincident with DNA peaks. Data were collected at 10 kHz.

shown). On the other hand, GFP emission in GST−GFP−GAGA (Figure 3C) was not quenched to any considerable degree by either dabcyl-GAGA or dabcyl-GATA duplexes, even at higher concentrations of DNA (data not shown). These results indicate that the bulk fluorescence assay was able to demonstrate specific DNA-binding activities for the His-GFP-tagged Egr-1 ZFD and GST−GFP-tagged GATA ZFD but not for the GST−GFP-tagged GAGA ZFD.

Finally, the addition of equimolar concentrations of a dabcyl-labeled operator duplex, dabcyl-OR3, resulted in ~10% quenching of BFL emission intensity in BFL−GST−Cro (Figure 3D). Because this quenching was not observed in the control BFL-labeled protein GST (inset of Figure 3D), the quenching of this BFL-labeled Cro protein depends upon the presence of the Cro repressor domain. Furthermore, the addition of a labeled duplex containing an unrelated recognition sequence, dabcyl-EBS, failed to elicit BFL quenching, indicating that DNA binding exhibited by this Cro fusion protein is specific for the operator target sequence. Because no further quenching was observed upon incubation with additional dabcyl-OR3 under these conditions (data not shown), the dissociation constant for binding must be considerably less than 1 nM, indicating that the affinity of the Cro repressor for the OR3 duplex is relatively high, a finding consistent with detailed equilibrium analysis (Nalefski et al., manuscript in preparation).

Single-Molecule Detection of Protein−DNA Complexes. Single-molecule traces were collected from reaction mixtures containing protein and Cy5-labeled DNA target molecules, representative segments of which are shown in parts A to D of Figure 4. Thresholds (dashed horizontal lines in Figure 4), which exclude all but a small number of background fluorescence events in buffer (Figure 4A), were applied as described (*17*) to illustrate the detection of red channel events arising from Cy5-labeled EBS DNA (Figure 4B), blue channel events arising from the GFP-tagged Egr-1 ZFD (Figure 4C), and coincident events representing protein−DNA complexes (vertical lines in Figure 4D).

The single-molecule trace files were analyzed by two analytical approaches to confirm the detection of protein−DNA complexes in the binding reactions (Figure 5). First, raw fluorescence traces in both red (DNA) and blue (protein) channels were subjected to coincident event counting (top panels in Figure 5), whereby numbers of nonrandom coincident bins are estimated. The presence of a sharp rise in coincident events resulting from the reaction of the His-GFP-tagged Egr-1 ZFD with its Cy5-labeled target DNA EBS but not with an unrelated Cy5-labeled DNA OR3 (Figure 5A) indicates the detection of specific protein−DNA complexes. The position of this peak, which is centered on an offset value of 0, confirms the proper co-alignment of the red and blue laser spots. This peak was not observed when the control protein His-GFP was incubated with Cy5-labeled EBS DNA (Figure 5B), further indicating that the binding observed was due to the presence of the Egr-1 ZFD. Second, the same trace files were analyzed by spatial cross-correlation analysis (bottom panels in Figure 5), which reveals correlated spatiotemporal fluctuations in emissions from both molecules arising from bound complexes (*27*). Similar to coincident event counting, cross-correlation analysis also reveals the presence of specific protein−DNA complexes resulting from the reaction of His-GFP-tagged Egr-1 ZFD with its target DNA but not with an unrelated DNA fragment.

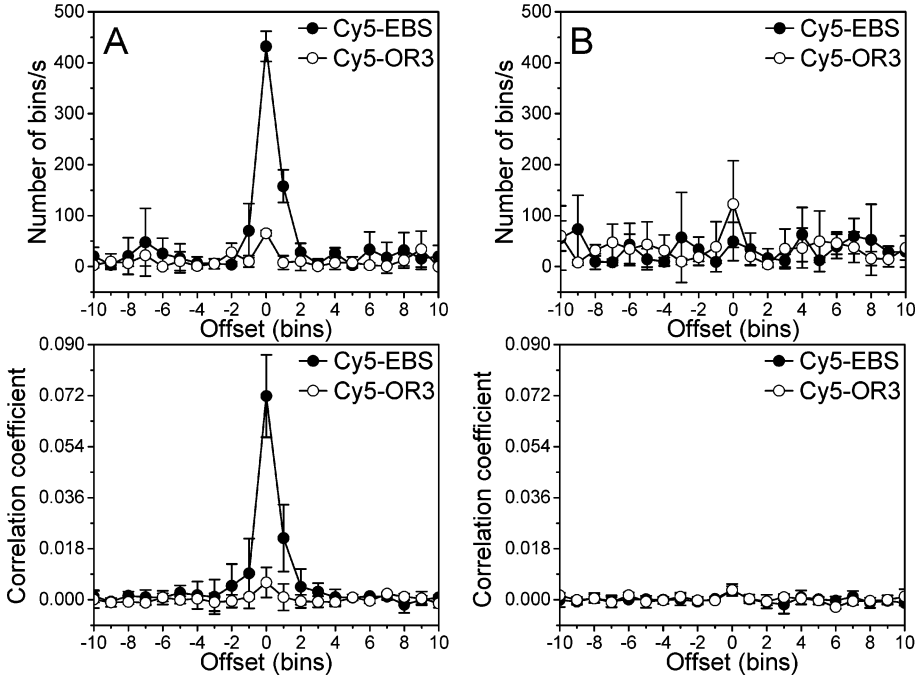

FIGURE 5: Detection of protein−DNA complexes by single-molecule event counting and cross-correlation analysis. Coincident fluorescent bins (top) and cross-correlation values (bottom) were calculated from trace files recorded on His-GFP-Egr (A) or its control His-GFP (B) reacted with the Egr-1 target Cy5-EBS (●) or nontarget Cy5-OR3 DNA (○). Data points represent the average value [±standard deviation (SD)] of signals obtained from three separate time segments, each 40 s in duration, of red_1 (DNA) and blue (protein) fluorescence channels analyzed at the indicated temporal offsets, where 1 bin translates to 100 μs. The tall spike observed in both plots in A but not B when Cy5-EBS but not Cy5-OR3 is present indicates the detection of protein−DNA complexes when only the transcription-factor DNA-binding domain is reacted with its specific target. Experimental conditions: ~50 pM total protein and 25 pM Cy5-labeled DNA. Data were collected at 10 kHz.

The novel coincident event counting approach was further tested by examining the formation of protein−DNA complexes in reaction mixtures consisting of different proteins and Cy5-labeled target DNA molecules (Figure 6). For the intrinsically fluorescent His-GFP-tagged Egr-1 ZFD (top panel in Figure 6A), binding was observed to the correct DNA target (EBS) labeled with Cy5 but not the incorrect labeled DNA (OR3). Incubation of protein with a large excess of the correct unlabeled DNA (EBS) but not with the incorrect unlabeled DNA (OR3) completely prevented binding to the Cy5-labeled target. No significant binding was observed between the control protein His-GFP and DNA (bottom panel in Figure 6A). Likewise, for the intrinsically fluorescent GST−GFP-tagged GATA and GAGA factor ZFDs (top and middle panels in Figure 6B), binding of protein to DNA was observed with the correct protein and target DNA combinations only. In both cases, binding competed with excess unlabeled DNA containing the correct target sequence and no binding to DNA was observed with the control protein GST−GFP (bottom panel in Figure 6B). Finally, for the BFL-labeled, GST-tagged Cro repressor (top panel in Figure 6C), binding to the correct Cy5-labeled DNA target (OR3) but not the incorrect Cy5-labeled DNA (EBS) was observed, and this binding was prevented with an excess of the correct unlabeled DNA (OR3) but not the incorrect DNA (EBS). No interaction with DNA was observed for the protein control BFL-labeled GST (bottom panel in Figure 6C).

Determination of Equilibrium Binding Constants. The equilibrium constant for binding of the Egr-1 ZFD to DNA was measured using the single-molecule counting approach, because this has been measured extensively using conventional approaches (Table 1). Three separate assays based on the coincident bin counting approach were applied to evaluate the equilibrium dissociation constant (K_d). In the first assay, a fixed concentration of the His-GFP-tagged Egr-1 ZFD was directly titrated with Cy5-labeled EBS DNA and the dependence of the number of protein−DNA complexes formed on the free DNA concentration was determined (Figure 7A). Nonlinear, least-squares analysis of the binding data revealed excellent fitting with a hyperbolic equation and a K_d value of 13 pM (Table 2).

A second assay was applied to independently measure the K_d value by titrating a fixed concentration of the His-GFP-tagged Egr-1 ZFD with Cy5-labeled EBS DNA diluted into unlabeled EBS DNA at a constant ratio of 1:3 (Figure 7C). In principle, this approach should decrease the free Cy5-labeled DNA concentration required for the half-maximal formation of protein−Cy5-labeled DNA complexes, as well as the maximal number of protein−DNA complexes, by a factor of 4 (see the Materials and Methods). As expected, the midpoint of the titration was shifted to considerably lower Cy5-EBS concentrations: analysis of the hyperbolic data yielded an K_d^{app} equal to 3.5 pM. This value translates to a K_d value of 14 pM, which is similar to the value obtained from direct titration of protein with Cy5-labeled DNA (Table 2).

To further probe equilibrium binding of protein to DNA, a competitive inhibition assay was applied in which a fixed amount of protein was mixed with a fixed concentration of Cy5-EBS DNA and increasing amounts of unlabeled EBS duplexes (Figure 7D). As predicted, a hyperbolic decrease

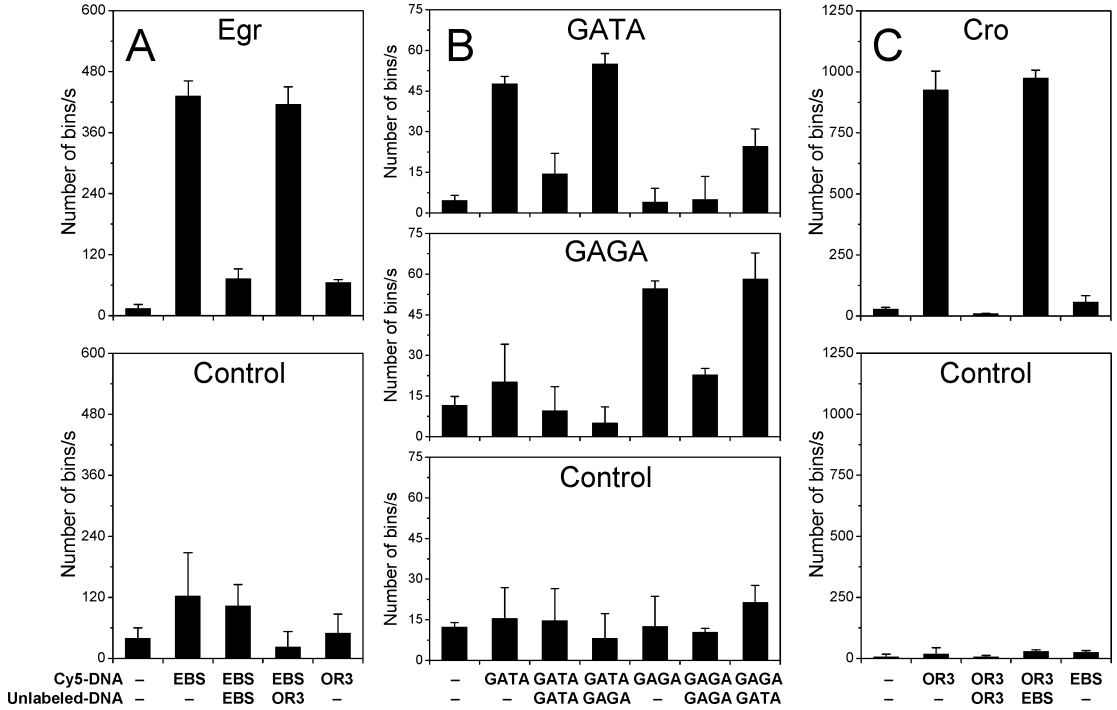

FIGURE 6: DNA-binding specificity demonstrated by single-molecule event counting. The number of bound protein−DNA complexes was determined from single-molecule trace files containing the indicated proteins and Cy5-labeled DNA and expressed as coincident bins/s. When used, unlabeled DNA was premixed with Cy5-labeled DNA prior to protein addition. (A) His-GFP-tagged Egr-1 ZFD (top) and control His-GFP (bottom); (B) GST−GFP-tagged ZFDs of GATA-1 (top), GAGA factor (middle), and control GST−GFP (bottom); and (C) BFL-labeled, GST-tagged Cro repressor (top) and control BFL−GST (bottom). Experimental conditions: ~50 pM total protein, 25 pM Cy5-labeled DNA, and 2.5 nM unlabeled DNA. Data were collected at 10 kHz, and results are expressed as the average (±SD) of analysis from three separate time segments.

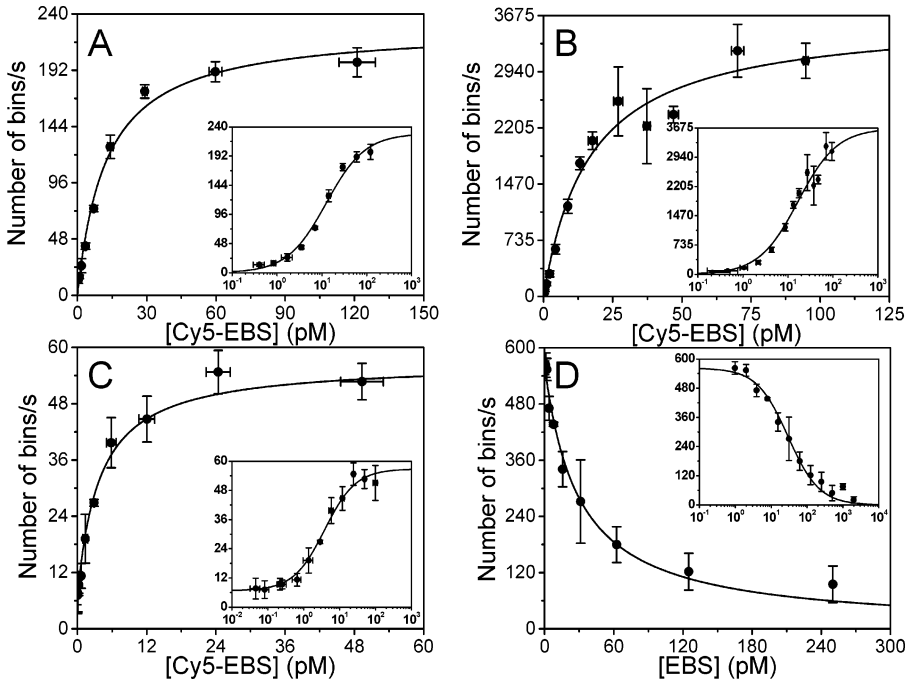

FIGURE 7: Determination of equilibrium constants for Egr-1 ZFD binding to DNA by single-molecule event counting. Representative saturation titrations of His-GFP-Egr from single-molecule recordings collected using laser spot (A, C, and D) or stripe illumination (B). Direct titrations with Cy5-EBS (A and B). Solid lines represent the best fit to a single-site binding equation (eq 2). (C) Titration of a mixture of the tracer Cy5-EBS and EBS (1:3). The solid line represents the best fit to a modified form of the single-site binding equation (eq 2). (D) Titration of the competitive inhibitor EBS, added simultaneously with 25 pM Cy5-EBS to the protein. The solid line represents the best fit to single-site competition equation (eq 3). Insets to panels present the same data plotted on a semi-log scale. Each point represents the average (±SD) of analysis from three separate time segments. Equilibrium parameters of analyses are summarized in Table 2.

in the number of bound molecules was observed as the amount of competitor increased. The fitted least-squares IC_{50} value equal to 28 pM was converted into a K_d value of 10 pM (Table 2), which is comparable to the 13−14 pM values obtained in ligand and tracer saturation titrations of protein with DNA.

Table 2: Equilibrium and Kinetic Constants for the Binding of the Egr-1 ZFD to Synthetic DNA Duplex Determined by Single-Molecule Counting[a]

Equilibrium Parameters	
assay	K_d (pM)
direct ligand[b]	13 (±3)
tracer ligand[c]	14 (±1)
competitive inhibitor[d]	10 (±3)
Kinetic Parameters	
assay	constant
association[e]	$k_{on} = 1.0\ (\pm 0.5) \times 10^9\ M^{-1}\ s^{-1}$
dissociation[f]	$k_{off} = 1.11\ (\pm 0.10) \times 10^{-3}\ s^{-1}$

[a] Average (±SD) of at least three independent determinations. Experimental conditions: 100 mM NaCl, 10 mM Tris at pH 8.0, 10 mM EDTA, 10 μM $ZnSO_4$, and 250 μg/mL BSA at room temperature. [b] Saturation titration of the protein with Cy5-EBS, as in Figure 7A. Data were fitted with eq 2. [c] Saturation titration of the protein with tracer Cy5-EBS/EBS (1:3), as in Figure 7C. Data fitted with a modified form of eq 2 yielded K_d^{app} equal to 3.5 (±0.2) pM. K_d was calculated from K_d^{app}/n, where n represents the fraction of DNA that is labeled with Cy5. [d] Saturation titration of the protein and 24 pM free Cy5-EBS with the competitor unlabeled EBS, as in Figure 7D. Data fitted with eq 3 yielded IC_{50} equal to 28 (±3) pM. K_d was calculated from IC_{50} using eq 4 and the 13 pM K_d value from the direct ligand titration assay. [e] Association time courses were initiated by the addition of Cy5-EBS to the protein. k_{on} was calculated using eq 6 from slopes of three independent determinations, as in Figure 8B. [f] The dissociation time course was initiated by the addition of a large molar excess of unlabeled EBS to the protein prebound to Cy5-EBS, as in Figure 9A. Data were fitted with monoexponential eq 1.

Finally, equilibrium measurements were conducted on a single-molecule detection apparatus that employs stripe laser illumination and captures fluorescence emission by a CCD camera. Single-molecule measurements conducted as in Figure 6 confirmed the DNA-binding specificity for the fluorescently tagged ZFDs (data not shown), demonstrating that the instrument recapitulates the findings of the prototype device. Furthermore, equilibrium measurements performed on the GFP-tagged Egr-1 ZFD using direct titrations with Cy5-labeled EBS duplexes (Figure 7B) yielded a dissociation constant of 18 (±3) pM ($n = 3$), in excellent agreement with the determinations made on the prototype instrument (Table 2).

Determination of Kinetic Rate Constants. The single-molecule approach was used to determine the association and dissociation kinetics of Egr-1 ZFD fusion protein binding to Cy5-labeled EBS duplexes. The association of protein with Cy5-EBS DNA was measured by mixing free protein with Cy5-EBS DNA and observing the formation of protein−DNA complexes at various times after mixing (Figure 8). At concentrations of Cy5-EBS DNA less than or equal to 50 pM, fitting the time course data revealed monoexponential increases in the number of protein−DNA complexes with time (Figure 8A). Under these conditions, both the observed rate constants and the number of protein−DNA complexes at the end of the time courses increased as the target DNA concentration increased, suggesting that the rate-limited step in the process involves bimolecular association between the protein and DNA. At higher concentrations, the change in the number of complexes was complete by the time recordings could be made (data not shown). Because the data were collected under pseudo-first-order conditions, analysis of the linear dependence of the observed pseudo-first-order rate constants on the target DNA concentration (Figure 8B)

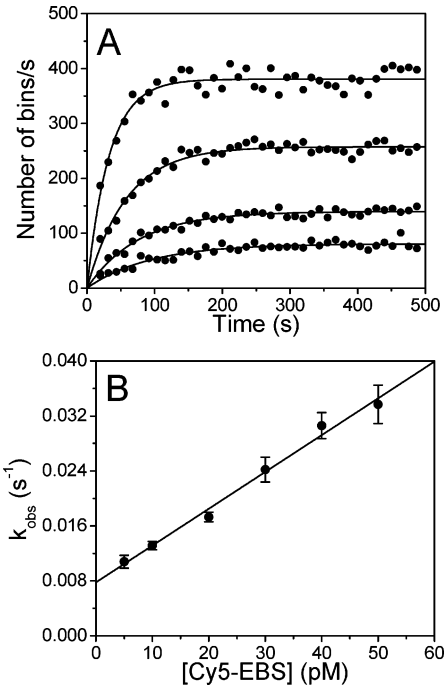

FIGURE 8: Determination of the association rate constant for Egr-1 ZFD binding to DNA by single-molecule event counting. (A) Representative time courses of Cy5-EBS binding initiated by mixing the His-GFP-Egr ZFD with 5, 10, 20, and 40 pM (final) Cy5-EBS (from bottom to top) collected on the Trilogy 2020 device. Each data point represents the analysis of 5 s segments averaged from three independent trace files; the solid line represents the best fit to a monoexponential equation (eq 5). (B) Dependence of the observed pseudo-first-order rate constants on Cy5-EBS DNA. Each data point represents the observed rate constant (±SD) for time courses generated as in A. The second-order association constant, k_{on}, was obtained from the slopes of experiments performed on three separate occasions (Table 2).

allowed for estimation of the second-order rate constant, k_{on}, from the slope, which equals $1.0 \times 10^9\ M^{-1}\ s^{-1}$ (Table 2). The first-order dissociation rate constant, k_{off}, which can be estimated from the y intercept, equaled 6.6 (± 1.4) $\times 10^{-3}\ s^{-1}$.

A more accurate determination of k_{off} was provided by direct analysis of dissociation time courses. Dissociation of protein from Cy5-labeled EBS duplexes was measured by prebinding a fixed concentration of protein with Cy5-EBS and then initiating dissociation with the addition of a vast excess of unlabeled EBS DNA and observing the loss of protein−DNA complexes at various times after mixing (Figure 9A). Excellent fitting of the time course data was achieved with a monoexponential equation and yielded an apparent first-order rate constant k_{off} equal to $1.11 \times 10^{-3}\ s^{-1}$ (Table 2). As expected, no dissociation was observed when bound protein−DNA complexes were mixed with the nontarget duplex OR3. These results were then compared to those obtained using conventional bulk fluorescence spectroscopy to follow GFP quenching as an indicator of protein binding to quencher-labeled DNA duplexes, as demonstrated above (Figure 3). After the formation of complexes between His-GFP-Egr and dabcyl-EBS duplexes, dissociation was initiated by the addition of a vast excess of unlabeled duplexes and was measured by recording the loss of GFP quenching (Figure 9B). The addition of the unlabeled target duplex EBS but not the nontarget duplex OR3 (not shown) resulted in a monoexponential decrease in GFP quenching.

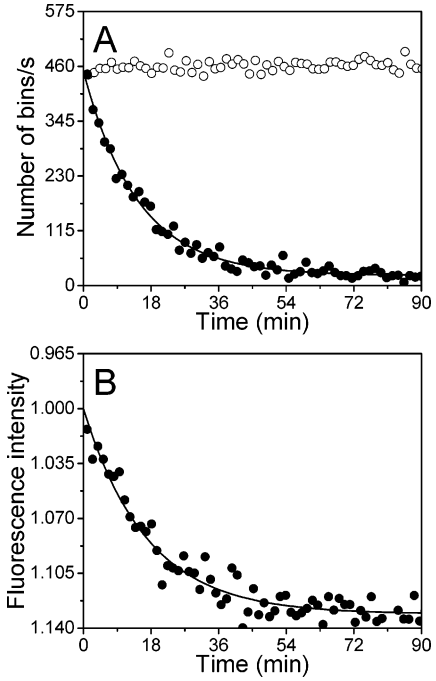

FIGURE 9: Determination of the dissociation rate constant for Egr-1 ZFD binding to DNA by single-molecule event counting and conventional bulk fluorescence. (A) Dissociation kinetics measured by single-molecule event counting on the Trilogy 2020 device. Representative time course of Cy5-EBS dissociation initiated by adding 2.5 nM EBS (●) to the His-GFP-Egr ZFD prebound to Cy5-EBS (25 pM). Each data point represents the analysis of a 10 s segment of the trace file; the solid line represents the best fit to a monoexponential equation (eq 1), yielding k_{off} (Table 2). The nontarget competitor, OR3, added at 2.5 nM (○) failed to induce dissociation. (B) Dissociation kinetics measured by bulk fluorescence. Time course of fluorescence intensity changes in the His-GFP-tagged Egr ZFD preincubated with the dabcyl-EBS duplex and then mixed with a large molar excess of unlabeled target duplex EBS. The increase in fluorescence intensity induced by unlabeled EBS, inverted on the *y* axis, indicates a specific reversal in quenching of dabcyl-EBS by the competitor. The solid line represents the best fit to a monoexponential equation (eq 1), which yielded an average (±SD) k_{off} value of 1.17 (±0.18) × 0^{-3} s^{-1} in five separate experiments. Experimental conditions: 1 nM protein, 1 nM dabcyl-EBS, and 40 nM unlabeled EBS, with $\lambda_{ex} = 490$ nm and $\lambda_{em} = 508$ nm.

As expected for a process that proceeds via first-order kinetics, the extent of quenching decreased exponentially with time. Analysis of the time course of dissociation yielded a k_{off} value of 1.17×10^{-3} s^{-1}, which is equal within error to that obtained using the single-molecule approach (Table 2).

DISCUSSION

We have employed a single-molecule detection methodology, capable of simultaneous detection of two or potentially more individual molecules tagged with distinct fluorophores flowing past a micrometer-sized detection zone, to detect, in real time, specific binding of four distinct transcription-factor DNA-binding domains to synthetic DNA duplexes containing their known recognition targets. We characterized in detail the equilibrium and kinetic DNA-binding parameters of one of these, that of the Egr-1 ZFD. Below, we relate these measurements to the published binding properties of Egr-1, validating one potential use of the new technology.

Demonstration of DNA-Binding Specificity. The single-molecule detection approach presented here revealed specific association between fusion proteins containing four distinct DNA-binding domains and their cognate target DNA duplexes at picomolar concentrations. Binding was observed for three polypeptides that normally bind as monomers (Egr-1, GATA-1, and GAGA factor) and one that binds DNA as a dimer (Cro repressor). Successful binding was achieved with fusion proteins either extrinsically labeled on cysteines with the small organic fluorophore BODIPY FL (Cro repressor) or intrinsically labeled by fusion with GFP (Egr-1, GATA-1, and GAGA factor). Among the fusions with ZFDs, binding was observed with polypeptides affixed to two different affinity purification tags (GST and His_6). Specific binding was detected with polypeptides reported to display a wide range of DNA affinities, from the high-affinity Cro repressor ($K_d \sim 2$ pM) to the lower affinity GATA-1 and GAGA factor ZFDs ($K_d \sim$ nanomolar). No cross-reactivity between the proteins and noncognate DNA targets was observed in any case tested, nor was any interaction observed between the purification tags lacking the transcription-factor domains and DNA. Together, these findings illustrate that the DNA binding observed is conferred by the transcription-factor domain and not artificially by the fusion polypeptides used for purification, GFP for intrinsic fluorescent labeling, or covalent attachment of extrinsic dyes. The protein fusion constructs and labeling approaches presented here should be applicable to a wide variety of DNA-binding domains.

Notably, in our analysis of the GAGA factor ZFD, specific DNA binding was demonstrated by the single-molecule approach but not by the bulk fluorescence approach that successfully demonstrated specific DNA binding by the other three transcription-factor fusion proteins. We postulate that the unfavorable orientation of donor and acceptor dyes in this particular protein−DNA complex might account for the failure to observe quenching of GST in the GAGA factor ZFD construct by dabcyl-labeled target DNA molecules. Unlike the bulk fluorescence assay used, the coincident event counting approach does not require changes in fluorescence intensity of one of the reacting species for generation of a binding signal. The single-molecule approach presented here might therefore succeed in detecting molecular interactions where other conventional bulk approaches fail.

Measurement of Equilibrium and Kinetic Rate Constants. Using three different experimental approaches, we measured dissociation constants ranging from 10 to 14 pM for the binding of the Egr-1 ZFD fusion protein to a synthetic DNA duplex containing an artificial 10 bp consensus sequence. In addition, very similar results were obtained on a single-molecule detection device that uses a different laser and photon-collecting technology from the platform prototype device. The measured dissociation constant closely overlaps the 10−12 pM obtained with EMSA using the isolated Egr-1 ZFD, whereby the extent of protein binding to radiolabeled DNA duplexes was measured as a function of the free protein concentration (*28*−*30*). Together, these values are 10-fold smaller than those obtained in other papers also utilizing EMSA to quantify the same binding reaction (*18*, *31*). The higher values in the latter studies are likely to result from the addition of the nonspecific DNA competitor poly(dI−dC) to the binding reactions. Although useful in reactions involving crude nuclear extracts, this reagent is known to reduce apparent affinities for purified DNA-binding proteins,

even at relatively low concentrations (*32*). Indeed, single-molecule recordings confirm that poly(dI−dC) decreases the apparent affinity of the Egr-1 ZFD for its target DNA (J. Lloyd, unpublished experiments). Other alternative experimental approaches have consistently yielded even higher values for the dissociation constant of the Egr-1 ZFD−DNA interaction, including phage display immunoassay (*33*) and SPR (*34*). Discrepancies between the present results and these alternative approaches are not unexpected because the nature and presentation of the reacting species differs fundamentally from that in the present paper.

Only limited information is available for the kinetic rate constants governing the binding of the Egr-1 ZFD to target DNA. The first-order dissociation rate constant measured in the present study, k_{off} equal to 1.11×10^{-3} s^{-1}, is the same within error as the value of 1.17×10^{-3} s^{-1} observed in bulk fluorescence measurements presented here and lies within the range of values measured using EMSA (*28*, *31*) and SPR (*34*). On the other hand, less is known about the second-order association rate constant governing the interaction between the Egr-1 ZFD and target DNA. The large value observed in the present study, k_{on} estimated to be 1.0×10^9 M^{-1} s^{-1}, is consistent with the lower limit of 7.0×10^8 M^{-1} s^{-1} projected in an EMSA study (*28*). However, a considerably smaller association rate constant, measured at 3×10^4 M^{-1} s^{-1}, has been reported in an SPR study (*34*). Although the upper limit of detectable rate constants for SPR biosensors has been reported to be approximately 10^6 M^{-1} s^{-1}, depending upon conditions, differences in rate constants measured by SPR and other methods of up to several orders of magnitude are not uncommon for identical interacting molecular systems (*35*), including protein−DNA interactions (*36*).

CONCLUSION

These results indicate that the methodology described in the present paper represents a reliable approach for quantifying the equilibrium and kinetic constants governing interactions between soluble proteins and DNA molecules in solution. Our measurements compare very favorably with other methods based on published literature, although some differences were observed in cases where the published approaches utilized binding of molecules to surfaces. From a technical perspective, our approach appears to offer a number of advantages, including a homogeneous format, real-time determination, high reproducibility, and exquisite sensitivity, able to detect bound molecular species in very dilute solutions. The only notable disadvantage is the requirement to fluorescently label the reactants. However, unlike bulk fluorescence measurement methods, this single-molecule approach does not require a fluorescence change in dyes attached to the reacting species for generation of a binding signal. We employed this methodology to determine the dissociation constant governing the binding of the Cro repressor fusion protein to a synthetic duplex containing the OR3 sequence from the λ phage genome (Nalefski et al., manuscript in preparation). The measured value is very similar to that obtained in a published filter-binding assay (Table 1), whereby the protein dependence on the binding of radiolabeled DNA was measured (*21*). When combined with single-molecule optical approaches to map binding sites of transcription factors in native DNA sequences (Nalefski et al., manuscript in preparation), this approach offers an additional experimental tool for deciphering the transcription code. Furthermore, in principle, this methodology should be able to quantify the binding of any two molecules that can be fluorescently labeled.

ACKNOWLEDGMENT

We thank Randall Burton, Eugene Chan, Martin Fuchs, Rudolf Gilmanshin, John Harris, Jonathan Larson, Jack Leonard, Ray Meyer, Mark Nadel, Jenny Rooke, Michael Shia, Gordon Wong, and Duncan Whitney for helpful and insightful discussions and Coralia Giurescu and Hai-Ping Ko for developing data analysis software.

SUPPORTING INFORMATION AVAILABLE

Description of coincident event counting and cross-correlation analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., and Yan, C., et al. (2001) The sequence of the human genome, *Science 291*, 1304−1351.
2. Pennacchio, L. A., and Rubin, E. M. (2001) Genomic strategies to identify mammalian regulatory sequences, *Nat. Rev. Genet. 2*, 100−109.
3. Bulyk, M. L. (2003) Computational prediction of transcription-factor binding site locations, *Genome Biol. 5*, 201.
4. Laniel, M. A., Beliveau, A., and Guerin, S. L. (2001) Electrophoretic mobility shift assays for the analysis of DNA−protein interactions, *Methods Mol. Biol. 148*, 13−30.
5. Stockley, P. G. (2001) Filter-binding assays, *Methods Mol. Biol. 148*, 1−11.
6. Isalan, M., and Choo, Y. (2001) Engineering nucleic acid-binding proteins by phage display, *Methods Mol. Biol. 148*, 417−429.
7. Bulyk, M. L., Gentalen, E., Lockhart, D. J., and Church, G. M. (1999) Quantifying DNA−protein interactions by double-stranded DNA arrays, *Nat. Biotechnol. 17*, 573−577.
8. Bulyk, M. L., Huang, X., Choo, Y., and Church, G. M. (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays, *Proc. Natl. Acad. Sci. U.S.A. 98*, 7158−7163.
9. Carpenter, M. L., Oliver, A. W., and Kneale, G. G. (2001) Analysis of DNA−protein interactions by intrinsic fluorescence, *Methods Mol. Biol. 148*, 491−502.
10. Carpenter, M. L., Oliver, A. W., and Kneale, G. G. (2001) Circular dichroism for the analysis of protein−DNA interactions, *Methods Mol. Biol. 148*, 503−510.
11. Buckle, M. (2001) Surface plasmon resonance applied to DNA−protein complexes, *Methods Mol. Biol. 148*, 535−546.
12. Enderlein, J., Keller, R. A., and Zander, C. (2002) in *Single Molecule Detection in Solution* (Zander, C., Enderlein, J., and Keller, R. A., Eds.) pp 1−19, Wiley-VCH, Berlin, Germany.

13. Christy, B. A., Lau, L. F., and Nathans, D. (1988) A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences, *Proc. Natl. Acad. Sci. U.S.A. 85*, 7857−7861.
14. Evans, T., and Felsenfeld, G. (1989) The erythroid-specific transcription factor Eryf1: A new finger protein, *Cell 58*, 877−885.
15. Soeller, W. C., Oh, C. E., and Kornberg, T. B. (1993) Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor, *Mol. Cell. Biol. 13*, 7961−7970.
16. Miller, W. G., and Lindow, S. E. (1997) An improved GFP cloning cassette designed for prokaryotic transcriptional fusions, *Gene 191*, 149−153.
17. D'Antoni, C. M., Fuchs, M., Harris, J. L., Ko, H.-P., Meyer, R. E., Nadel, M. E., Randall, J. D., Rooke, J. E., and Nalefski, E. A. (2006) Rapid quantitative analysis using a single molecule counting method, *Anal. Biochem. 352*, 97−109.
18. Elrod-Erickson, M., and Pabo, C. O. (1999) Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger−DNA complex, *J. Biol. Chem. 274*, 19281−19285.
19. Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., and Gronenborn, A. M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1, *Science 261*, 438−446.
20. Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M., and Clore, G. M. (1997) The solution structure of a specific GAGA factor−DNA complex reveals a modular binding mode, *Nat. Struct. Biol. 4*, 122−132.
21. Kim, J. G., Takeda, Y., Matthews, B. W., and Anderson, W. F. (1987) Kinetic studies on Cro repressor−operator DNA interaction, *J. Mol. Biol. 196*, 149−158.
22. Taylor, J. R. (1997) *An Introduction to Error Analysis*, 2nd ed., University Science Books, Sausalito, CA.
23. Cheng, Y., and Prusoff, W. H. (1973) Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50% inhibition (IC_{50}) of an enzymatic reaction, *Biochem. Pharmacol. 22*, 3099−3108.
24. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) Zif268 protein−DNA complex refined at 1.6 Å a model system for understanding zinc finger−DNA interactions, *Structure 4*, 1171−1180.
25. Albright, R. A., and Matthews, B. W. (1998) Crystal structure of λ-Cro bound to a consensus operator at 3.0 Å resolution, *J. Mol. Biol. 280*, 137−151.
26. Tsien, R. Y. (1998) The green fluorescent protein, *Annu. Rev. Biochem. 67*, 509−544.
27. Korn, K., Gardellin, P., Liao, B., Amacker, M., Bergstrom, A., Bjorkman, H., Camacho, A., Dorhofer, S., Dorre, K., Enstrom, J., Ericson, T., Favez, T., Gosch, M., Honegger, A., Jaccoud, S., Lapczyna, M., Litborn, E., Thyberg, P., Winter, H., and Rigler, R. (2003) Gene expression analysis using single molecule detection, *Nucleic Acids Res. 31*, e89.
28. Kim, J. S., and Pabo, C. O. (1998) Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants, *Proc. Natl. Acad. Sci. U.S.A. 95*, 2812−2817.
29. Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger−DNA recognition: Crystal structure of a Zif268−DNA complex at 2.1 Å, *Science 252*, 809−817.
30. Greisman, H. A., and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites, *Science 275*, 657−661.
31. Swirnoff, A. H., and Milbrandt, J. (1995) DNA-binding specificity of NGFI-A and related zinc finger transcription factors, *Mol. Cell. Biol. 15*, 2275−2287.
32. Larouche, K., Bergeron, M. J., Leclerc, S., and Guerin, S. L. (1996) Optimization of competitor poly(dI−dC)·poly(dI−dC) levels is advised in DNA−protein interaction studies involving enriched nuclear proteins, *BioTechniques 20*, 439−444.
33. Isalan, M., Choo, Y., and Klug, A. (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition, *Proc. Natl. Acad. Sci. U.S.A. 94*, 5617−5621.
34. Wu, H., Yang, W. P., and Barbas, C. F., III (1995) Building zinc fingers by selection: Toward a therapeutic application, *Proc. Natl. Acad. Sci. U.S.A.* 92, 344−348.
35. Schuck, P. (1997) Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules, *Annu. Rev. Biophys. Biomol. Struct. 26*, 541−566.
36. Bondeson, K., Frostell-Karlsson, A., Fågerstam, L., and Magnusson, G. (1993) Lactose repressor−operator DNA interactions: Kinetic analysis by a surface plasmon resonance biosensor, *Anal. Biochem. 214*, 245−251.
37. Jana, R., Hazbun, T. R., Fields, J. D., and Mossing, M. C. (1998) Single-chain λ Cro repressors confirm high intrinsic dimer−DNA affinity, *Biochemistry 37*, 6446−6455.
38. Pedone, P. V., Ghirlando, R., Clore, G. M., Gronenborn, A. M., Felsenfeld, G., and Omichinski, J. G. (1996) The single Cys2−His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding, *Proc. Natl. Acad. Sci. U.S.A. 93*, 2822−2826.

BI0602011