https://doi.org/10.1093/nar/gkab843

Single molecule characterization of the binding kinetics of a transcription factor and its modulation by DNA sequence and methylation

Hadeel Khamis[1,2], Sergei Rudnizky[1], Philippa Melamed[1] and Ariel Kaplan[1,3,4,*]

[1]Faculty of Biology, Technion – Israel Institute of Technology, Haifa 32000, Israel, [2]Faculty of Physics, Technion – Israel Institute of Technology, Haifa 32000, Israel, [3]Faculty of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel and [4]Russell Berrie Nanotechnology Institute, Technion – Israel Institute of Technology, Haifa 32000, Israel

Received June 15, 2021; Revised August 04, 2021; Editorial Decision September 07, 2021; Accepted September 24, 2021

ABSTRACT

The interaction of transcription factors with their response elements in DNA is emerging as a highly complex process, whose characterization requires measuring the full distribution of binding and dissociation times in a well-controlled assay. Here, we present a single-molecule assay that exploits the thermal fluctuations of a DNA hairpin to detect the association and dissociation of individual, unlabeled transcription factors. We demonstrate this new approach by following the binding of Egr1 to its consensus motif and the three binding sites found in the promoter of the *Lhb* gene, and find that both association and dissociation are modulated by the 9 bp core motif and the sequences around it. In addition, CpG methylation modulates the dissociation kinetics in a sequence and position-dependent manner, which can both stabilize or destabilize the complex. Together, our findings show how variations in sequence and methylation patterns synergistically extend the spectrum of a protein's binding properties, and demonstrate how the proposed approach can provide new insights on the function of transcription factors.

INTRODUCTION

Already from the seminal work of Jacob and Monod (1), a central regulatory role was recognized for the equilibrium occupancy of response elements by transcription factors (TFs), which depends on their concentration, affinity, and, in eukaryotes, the chromatin accessibility of their binding sites (2,3). However, recent studies have highlighted the importance of transient TF–DNA interactions, and indicate that the *kinetics* of TF binding and dissociation play an important role in regulating transcription *in vivo* (4–11). In particular, the importance of the TF residence time in determining the transcriptional burst duration (6,10,11) and size (6,10) has been demonstrated. It has also been shown that association and dissociation are complex processes (12–19), that cannot be described as simple second- and first-order reactions, respectively. TFs find their targets by coupling 3D diffusion in the solution with one-dimensional diffusion while bound nonspecifically to DNA (20), and most are unable to bind nucleosomal DNA, so the association time is affected by the genomic context of the binding site (21,22) and by a kinetic competition with nucleosomes (23). On the other side, broad distributions in the residence time of individual TFs were observed *in vivo* (8) and also higher-order phenomena, such as dissociation facilitated by other TFs (13,15) or nucleosomes (24). Hence, given that TF binding to promoters and enhancers is a central event in the initiation of transcription, whose perturbation is linked to different disease states (25), quantitative measurements of the binding and dissociation times of TFs are of utmost importance. However, the complexity described above can be obscured by the ensemble averaging limitations of traditional biochemical methods, stressing the importance of following the binding reaction at the single-molecule level to reveal the full distribution of binding and dissociation times. Indeed, much of the evidence for the importance of the binding kinetics, and measurements of kinetic rates, are based on single-molecule tracking experiments, which can follow the motions of nuclear factors in living cells (26,27). Although very powerful, these experiments also suffer from important limitations (28): First among them is are the need to label the protein of interest, which can result in a perturbation of its diffusional and binding properties (29,30). In addition, their typical resolution is much larger than the size of the protein and its binding site. It is also challenging to uncouple the dynamics originating from the chromatin itself

*To whom correspondence should be addressed. Tel: +972 778871907; Fax: +972 778871908; Email: akaplanz@technion.ac.il

and that from the chromatin-binding protein. And, photobleaching and motions out of the focal plane prevent measurements of long trajectories (12). In vitro single-molecule methods based on fluorescence detection allow for a controlled environment (31), and have revealed important aspects of the binding of proteins to chromatin (24,32–34), but also require labeling and are limited to short DNA sequences.

Here, we introduce a new and complementary approach to characterize the binding kinetics of unlabeled TFs in a single molecule biophysical assay that can also gradually accommodate the complexity of the chromatin in real genes. Using DNA unzipping with optical tweezers, we exploit the thermal fluctuations of DNA as a sensor to follow, in real-time, the binding and dissociation of individual TFs to a binding site in its native sequence environment. We demonstrate our approach with Egr1, an inducible transcription factor responsible for regulating a variety of genes (35), and the *Lhb* gene, which encodes for the Luteinizing Hormone beta subunit and is expressed in gonadotrope cells (36). The DNA binding domain of Egr1 contains 3 zinc fingers (ZFs), each interacting with three DNA base pairs, making a 9 bp binding motif (37). Genome-wide studies obtained a consensus motif (CGCCCACGC) to which Egr1 binds with high affinity (38). However, most functional Egr1 binding sites in the genome deviate from this consensus, and these differences are gene-specific and evolutionarily conserved. This is also the case for *Lhb,* which contains 3 different binding sites in its proximal promoter. In a previous work (39), we showed that their specific sequence, and the conserved sequences flanking the binding sites, modulate the binding affinity and the conformation of the bound protein, suggesting that a functionally important and sequence-specific spectrum of conformations exist for Egr1. Here, we study how the sequence of the different sites modulates the binding kinetics. We found that the DNA sequence, at the core binding site and flanking it, modulates both binding and dissociation rates. Next, given that hypermethylation of Egr1 binding sites is associated with developmental defects (40–43) and that the proximal promoter of *Lhb* is hypomethylated in gonadotrope cells (44), we monitored how methylation of the *Lhb* sites, each with a distinct CpG methylation pattern, affects the binding kinetics. Our data show that methylation of CpGs modulates only the protein's residence time on the DNA. This modulation depends on the sequence of the DNA and the position of the methylation site and can both stabilize and destabilize the complex. Together, our results demonstrate the strength of the proposed method and highlight the versatility of DNA sequence and CpG methylation as a mechanism for TF binding regulation.

MATERIALS AND METHODS

Reagents

The original plasmid for the DNA binding domain of Egr1, from Dr Scot Wolfe, was kindly provided by Dr Amit Meller. The protein was expressed and purified as previously described (39).

Optical tweezers

Experiments were performed in a custom-made double-trap optical tweezers apparatus, as previously described (39,45,46). Briefly, the beam from an 852 nm laser (TA PRO, Toptica) was coupled into a polarization-maintaining single-mode optical fiber. The collimated beam out of the fiber, with a waist of $w_0 = 4$ mm, was split by a polarizing beam splitter (PBS) into two orthogonal polarizations, each directed into a mirror and combined again with a second PBS. One of the mirrors is mounted on a nanometer scale mirror mount (Nano-MTA, Mad City Labs). A X2 telescope expands the beam, and also images the plane of the mirrors into the back focal plane of the focusing microscope objective (Nikon, Plan Apo VC 60×, NA/1.2). Two optical traps are formed at the objective's focal plane, each by a different polarization, and with a typical stiffness of 0.3–0.5 pN/nm. A second, identical objective collects the light, the two polarizations separated by a PBS, and imaged onto two Position Sensitive Detectors (First Sensor). The beads' position relative to the center of the trap is determined by back focal plane interferometry (47). Calibration of the setup was done by analyzing the thermal fluctuations of the trapped beads (48), which were sampled at 100 kHz. Experiments were conducted in a laminar flow chamber (Lumicks), which was passivated following a published protocol (49) with some modifications. Briefly, we washed the chamber twice by flushing alternately 1M NaOH and Liquinox 1% for 10 min each. Casein (1%) was sonicated and filtered, diluted to 0.2%, flushed into the chamber and incubated in it for 40 min. After the incubation, the system was washed using the working buffer until all the free Casein was flushed. Using this system, we were able to work with very low Egr1 concentrations, ∼1 nM.

Molecular construct for single-molecule experiments

The constructs for single-molecule experiments (Supplementary Table S9) were generated as described previously (46), with a number of changes. Briefly, TF binding segments containing the *Lhb* promoter sequence were amplified by PCR from mouse genomic DNA, and segments for the non-native contexts (the 601 nucleosome positioning sequence and the *Cga* gene promoter) were amplified from a plasmid (a gift from Daniela Rhodes (MRC, Cambridge, UK)) and mouse genomic DNA, respectively (50). Primers used for the amplification reactions are listed in Supplementary Table S7. The constructs were digested using DraIII-HF (New England Biolabs) overnight according to the manufacturer's instructions. A 10 bp hairpin (Sigma) was ligated to the construct using T4 DNA ligase (New England Biolabs), in a reaction with 1:5 molar excess of the hairpin, at 16 °C. The construct was subsequently digested overnight with BglI (New England Biolabs). Methylation of the constructs was done using CpG Methyltransferase (M.SssI) and s-adenosylmethionine (sam) from New England Biolabs. The methylation efficiency was tested in a restriction reaction with FauI (New England Biolabs), whose activity is abolished for methylated DNA (Supplementary Figure S5).

We generated two ∼2000-bp DNA handles, each incorporating a specific tag (double digoxygenin and biotin), using commercially purchased 5′ modified primers (Supplementary Table S8) in a standard PCR reaction, using bacteriophage lambda DNA as the template. The other two primers (Supplementary Table S8) were designed to contain repeats of three DNA sequences recognized by single-strand nicking enzymes: Nt.BbvCI and Nb.BbvCI (both from New England Biolabs) on the biotin-tagged handle and the digoxygenin-tagged handle, respectively. The nicking enzymes generated 29 nt complementary overhangs on each handle. Handles were mixed at equal molar ratios for DNA annealing, creating a ∼4000 bp fragment of annealed DNA handles. A ∼350 bp dsDNA alignment segment with the sequence of the 601 DNA was prepared using commercially purchased primers (Supplementary Table S8) in a standard PCR reaction, ligated to the handles, and gel-purified (QIAquick 28706, Qiagen). Binding segments were ligated to DNA handles using a rapid ligase system (Promega) in a 3:1 molar ratio, 30 min at room temperature. The full construct (i.e. handles + alignment segment + TF binding segment) was incubated for 15 min on ice with 0.8 μm polystyrene beads (Spherotech), coated with anti-Digoxigenin (anti-DIG). The reaction was then diluted 1000-fold in binding buffer (10 mM Tris·Cl pH 7.4, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 3% v/v glycerol and 0.01% BSA). Tether formation was performed *in situ* (inside the optical tweezers' experimental chamber) by trapping a DNA-bound anti-DIG bead in one trap, trapping a 0.9 μm streptavidin-coated polystyrene beads in the second trap, and bringing the two beads into close proximity to allow binding of the biotin tag in the DNA to the streptavidin in the bead.

Data analysis

Data were sampled and stored at 2500 Hz. All further processing and analysis were done using Matlab functions. dsDNA was modeled with an Extensible Worm Like Chain model (eWLC) (51). The persistence length of dsDNA was experimentally determined by analyzing the extension versus force data in the region below the unzipping transition (Supplementary Figure S9A). The data was fitted to an eWLC model (51) with fixed Young modulus (1200 pN) and inter-phosphate distance (0.34 nm) and allowing the fitting algorithm to determine the persistence length (Supplementary Figure S9A, red curve). This resulted in a distribution of values for the persistence length, whose mean was equal to 45 nm, which is consistent with previous reports (52). All the data was then analyzed with this value. For ssDNA, we used a WLC model (51) with a persistense persistence length of 0.82 nm, determined by extrapolating the data from Bosco *et al*. (53) to our salt concentrations. Using the portion of the force versus extension curves above the unzipping transition, we fitted a WLC equation (Supplementary Figure S9A, green curve) to find the ssDNA inter-phosphate distance. The distribution of the results was centered around 0.65, which we then used to analyze all the data. As our experiments are done without any feedback loop on the force or the tether extension, changes in the tether's contour length, such as those occurring when a segment of dsDNA unzips, are accompanied by changes in force (Supplementary Figure S9B, top). These changes then affect the extension via the finite compliance of the system (Supplementary Figure S9B, middle). Hence, the number of bps unzipped (Supplementary Figure S9B, bottom) was calculated from the extension measurements by first subtracting the force-dependent extension of the handles and then dividing by twice the extension of a single ssDNA base at the given tension.

Full unzipping experiments

In full unzipping experiments (Supplementary Figures S1A, blue, S7A, S10) the position of the mirror, and hence the distance between the traps, was linearly increased. When the tension reaches 17–18 pN, the long (∼400 bp) DNA sample is gradually unzipped, as evidenced by a sequence-dependent and reproducible pattern of events that include a sudden decrease in force together with an increase in extension. Each of these events represents the cooperative opening of 20–30 bp of DNA. A protein bound to DNA inhibits the progression of the unzipping fork, as the force required to disrupt protein-DNA interactions is significantly higher than those needed to disrupt DNA alone (39,54–56). Detection of bound proteins in these experiments (Supplementary Figures S1A, blue, S7A, S10) was based on an increase in the force of more than 0.5 pN relative to the median force at the same position for experiments in the absence of Egr1. Events were classified as belonging to a specific site if the breaking event was located within a 20 bp window relative to the binding site's expected center. The significance of differences in the breaking forces was assessed with a two-tail Student's *t*-test. Applying the same criteria for the data obtained without Egr1 resulted in no binding events detected. No non-specific binding (i.e. binding to positions other than the known binding sites for Egr1) was detected (Supplementary Figure S10, blue curves). No binding to ssDNA was detected in relaxation (rezipping) curves (Supplementary Figure S10, red curves).

Kinetics experiments

To measure the dissociation and association times of Egr1 (Figures 1,2,4,5), DNA was unzipped only partially, until reaching the vicinity of a known Egr1 binding site. The next 20–30 bp segment in the unzipping curve can then thermally fluctuate between an 'open' and 'closed' state. The applied force was then finely adjusted to result in a given probability of closed versus open configuration. The fluctuations in extension were measured for 1 min, after which the construct was moved to a region in the laminar flow chamber containing Egr1. Binding of a protein stabilizes the closed state (Supplementary Figure S11). Detection of closed and open states, and hence the time durations for each state, were done using a hidden Markov model, implementing the HAMMY algorithm usually applied to smFRET data (57). While some differences were observed for the DNA breathing kinetics between different sequences and methylation states (see for example Supplementary Figure S12B, C), in all cases they were much faster than the protein's binding kinetics, and the protein's binding was identified by a sudden

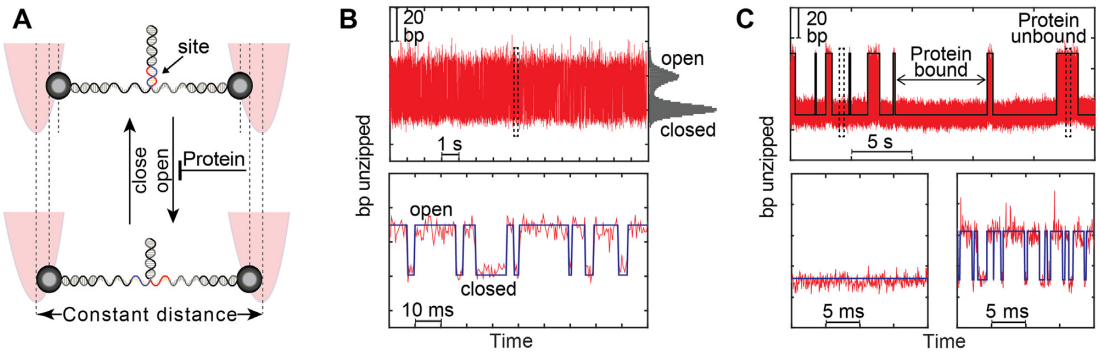

Figure 1. Real-time measurements of binding and dissociation. (**A**) Schematic description of the experiment. The DNA construct is attached from both sides to polystyrene beads trapped in optical traps. The distance between the traps is gradually increased, unzipping the DNA until reaching the vicinity of a binding site for Egr1, and then kept at a constant value. A segment of ∼25–30 bp in the fork, which includes the binding site, thermally fluctuates between an open and a closed state. (**B**) Breathing fluctuations in the number of base pairs unzipped. The histogram on the right side corresponds to the probability of populating the two states, open and closed. The lower panel shows a close up look at the dashed region in the upper one. (**C**) Breathing fluctuations in the presence of Egr1. Binding of the protein suppresses the DNA fluctuations and is detected as a long and stable closed state. The lower panel provides a closer look into the bound (left dashed box in the upper panel) and unbound (right dashed box) protein states, respectively. Data in (B, C) is shown unfiltered, as sampled at 2.5 kHz.

transition to a 'closed' state that lasted more than 150 ms. No binding events were detected in the absence of protein (Supplementary Figure S1C, D), or in the absence of a binding site. The residence time was defined as the time lapsed until the reappearance of the fluctuations. The binding time was calculated by multiplying the time between events by the instantaneous value of P_{closed}. Cumulative distribution function (CDFs) for the residence and binding times were fitted to exponential functions to extract the association and dissociation rates, k_{on} and k_{off}, respectively. The rates are presented as the value obtained from the fit ±95% confidence intervals, obtained from bootstrap by resampling the data 100 times, with replacement.

While measurements of dissociation were very consistent between experiments, we observed some variability in the association times between different days (Supplementary Figure S2), likely because k_{on} is a sensitive function of the protein's concentration. Hence, to reliably compare between the binding kinetics to different DNA constructs, we exploited different channels in our laminar flow cell for different samples (Supplementary Figure S14). The constructs were successively and repetitively moved, after tether formation, into a common channel containing the protein to measure the binding kinetics. This allows us to minimize any potential error caused by the effective concentration of active proteins (Supplementary Figure S2), by effectively 'dipping' the different DNA molecules into the same protein solution. In our work, we only compare data taken from the same session (∼3 h), i.e. between experiments where the kinetics was measured together and therefore the protein's concentration is, by design, identical.

Kinetics under force

In experiments characterizing the kinetics on a template under force (Figure 3), the DNA was unzipped until reaching the vicinity of the Egr1 binding site, and the distance between the traps was fixed to observe extension fluctuations between the open and the closed states. A drop in the standard deviation of the force, calculated in real-time in 100 ms windows, indicates a protein binding event. Upon detection of the binding event, the voltage in the mirror was instantaneously increased, moving quickly (150 ± 50 ms response-time) the steerable trap to increase the applied force to a predetermined value. The distance between the two traps was then held constant until the force drops, which indicates dissociation of the protein and reveals the bound state's lifetime at the specific force probed. Next, the mirror voltage, and hence the trap position, were restored to their original values to allow a successive round. Using this protocol (Figure 3A), we were able to explore the residence time at an extended force range (Figure 3B). Taking into account the known response time of the system ($t_{DT} =$ 150) we calculated the typical residence time (τ) from the measured average residence time ($\langle t \rangle_{meas}$), by solving the equation $\langle t \rangle_{meas} = \tau^{-1} \int_{t_{DT}}^{\infty} t \exp(-t/t_{DT})$. This results in $\langle t \rangle_{meas} = (t_{DT} + \tau) \exp(-t/t_{DT})$, from which the approximate solution $\tau \approx \frac{1}{2}(\langle t \rangle_{meas} + \sqrt{\langle t \rangle_{meas}^2 + 4t_{DT}^2})$ was obtained.

Electrophoretic mobility shift assays (EMSA)

'Top' and 'bottom' DNA oligonucleotides of 12, 22 and 32 bp length were purchased from IDT. The top strands were 5′-end labeled with ^{32}P using T4 Polynucleotide Kinase (New England Biolabs). The labeled strands were then annealed, in an excess concentration, with their corresponding complementary DNA strands and purified from a gel. The DNA constructs were incubated for 3 hours on ice (∼4°C) with a gradient of concentrations of Egr1 in binding buffer. The samples were loaded on a native 12% polyacrylamide in 1× TBE gel while running, to ensure fast entry of the complex to the gel's wells. Gels were run for 45 min at 200 V, ensuring that the temperature of the running buffer is maintained below 10°C. Gels were dryed (BioRad, 583) and imaged using a GE Typhoon FLA7000 phosphoimager. The

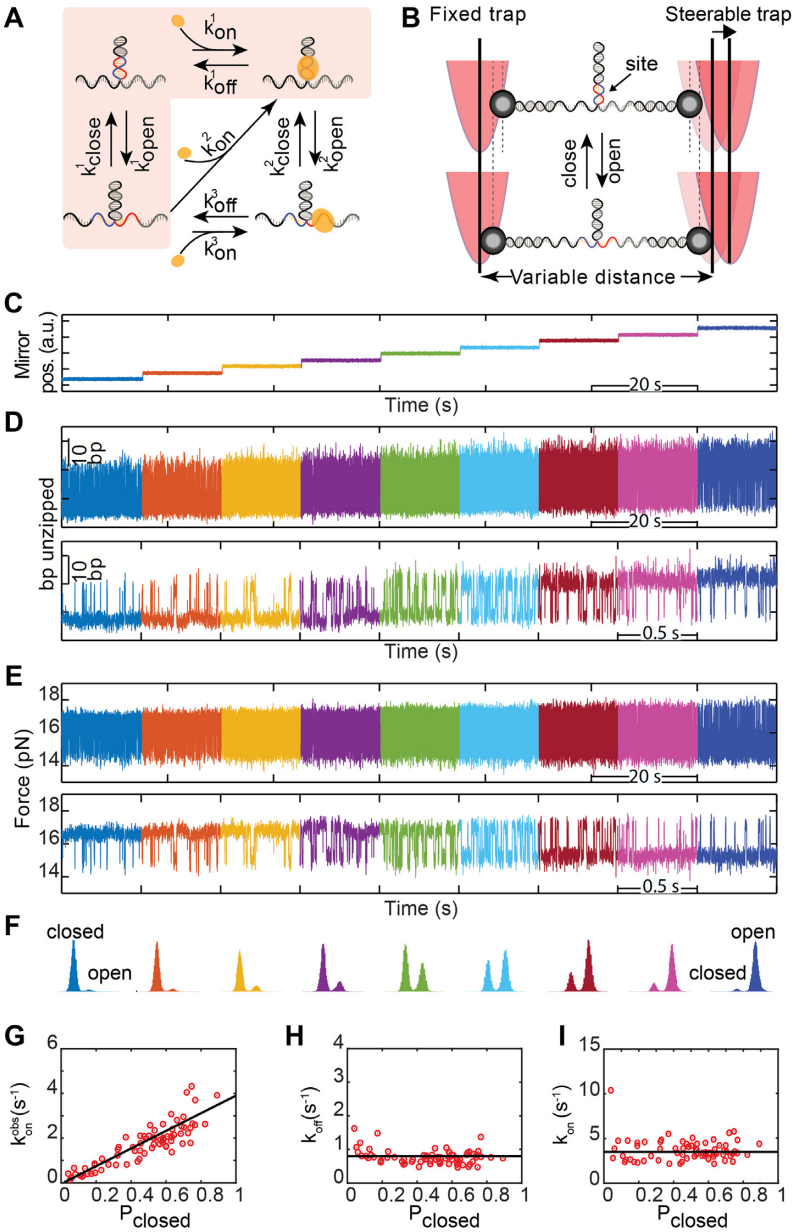

Figure 2. Egr1 binds passively to DNA. (**A**) Possible association paths of Egr1. (**B**) Illustration of the experiment. The position of the steerable trap is gradually shifted in a controlled manner to adjust the tension on the tether. (**C**) The mirror position as a function of time. (**D**, **E**) The number of base pairs unzipped (D) and the force (E) as a result of DNA breathing. Each color corresponds to a different distance between the two traps (C), and is characterized by a different probability to occupy the closed state. The lower panels in (D, E) are a zoom-in to 0.5 s at each condition from the same experiment. Data are shown as sampled, at 2.5 kHz. (**F**) Histograms of the extension for each condition. The lower value in the histogram corresponds to the closed state and the higher value to the open state. (**G**) Observed association rates as a function of p_{closed}. (**H**) Observed dissociation rate as a function of p_{closed}. (**I**) Intrinsic association rate, obtained by dividing the observed association rate by p_{closed}, as a function of p_{closed}. Data in (E–G) correspond to site –1 in the *lhb* proximal promoter. $N = 11$ DNA molecules, 1818 and 1790 detected times, for binding and residence, respectively.

fraction of bound protein was estimated using ImageLab. Curves for the fraction bound vs. Egr1 concentration were fitted to a hyperbolic binding equation to extract the dissociation constant (Supplementary Figure S6F). To validate that [DNA]≪[Egr1], we conducted experiments at different DNA concentrations that resulted in the same binding curves. The results shown are an average of multiple gels, and the error bars are the standard error of the mean (Supplementary Figure S6G).

RESULTS

Real-time measurements of the binding and dissociation of individual TFs

To follow the binding and dissociation of a single, unlabeled Egr1 protein to its binding site, we use optical tweezers to monitor the spontaneous fluctuations of a partially unzipped DNA molecule. Each of the strands of a ∼400 bp DNA 'sample' containing the consensus binding site for

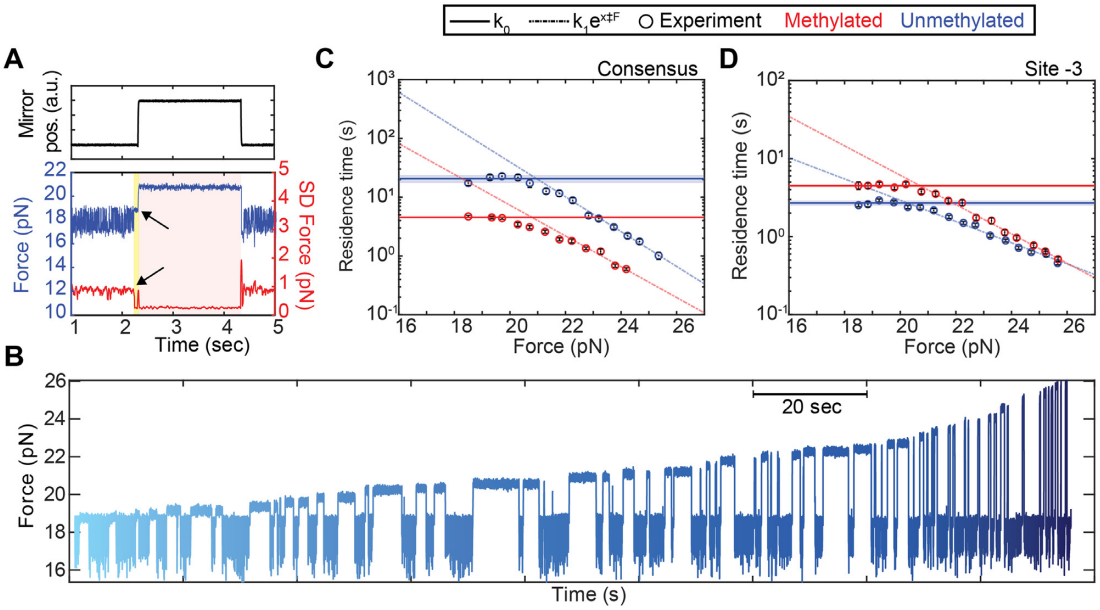

Figure 3. Two dissociation pathways exist for Egr1. (**A**) Illustration of the dissociation under force assay. Top panel: Mirror position versus time. Bottom panel: external force (left Y-axis, blue), and standard deviation of the force (right Y-axis, red) versus time. The experiments response time is shaded in yellow. Times at which Egr1 is bound are shaded in red. Data were filtered to 200 Hz (Butterworth). (**B**) Representative binding events at various applied forces (different shades of blue). Data shown as sampled, at 2.5 kHz. (**C**, **D**) Residence time vs. applied force for unmethylated (blue) and methylated (red) DNA for the consensus site (C) and site -3 in *Lhb* (D). The average of data points at forces below 20 pN is shown as a horizontal line ± standard deviation. An exponential decay fit for the data higher than 20 pN is shown in dashed lines.

Egr1, is attached to a long (∼2000 bp) dsDNA 'handle' (Figure 1A). The other end of each handle is bound to a polystyrene bead held in one of the traps of a dual trap optical tweezers setup (Figure 1A). By steering one of the traps away from the other, creating a tension above 17–18 pN, the DNA sample is unzipped in a sequence-dependent pattern, consisting of successive drops in force that are concomitant with increases in the overall extension (Supplementary Figure S1A), which represent the cooperative openings of successive DNA segments (Supplementary Figure S1B). When the unzipping fork reaches the vicinity of the TF binding site, we stop the movement of the traps and hold them at a constant distance from each other. The DNA sample, which is under tension, exhibits spontaneous 'breathing' dynamics (58), i.e. the thermally driven opening and closing of a short (25–30 bp) DNA segment (Figure 1B), which we follow with a time resolution below 1 ms. In general, due to the differences in binding energy between CG and AT pairs, these fluctuations populate only two states, which we term 'open' and 'closed'. Next, exploiting a laminar flow cell, the fluctuating DNA molecule is moved to a different region where it is exposed to a solution containing Egr1 (10 nM, unless specified otherwise). Binding of the protein to its binding site stabilizes the closed state, momentarily suppressing the breathing, which reappears upon dissociation of the protein (Figure 1A, C). Since the typical time the oscillations remain suppressed is much longer than the time spent in the closed state by a protein-free DNA molecule (Supplementary Figure S1C, D), we are able to set a threshold (t_{TH} = 150 ms) and reliably identify the protein-bound state as a stable closed state whose duration is longer than t_{TH}. No bound events are identified in the absence of proteins (Figure 1B, Supplementary Figure S1C) or for DNA sequences that do not include a binding site for Egr1 (Supplementary Figure S13). Identification of multiple bound and free states allows us to characterize the meantime the protein stays bound to DNA and the time in between those events, i.e. the *observed* binding and residence times, and their reciprocals, the observed binding and dissociation rates, k_{on}^{obs} and k_{off}^{obs}.

Egr1 binds passively to dsDNA, and exhibits two dissociation regimes

Notably, the ability to control the force applied on the DNA opens also the possibility to shed light on the *mechanism* of binding. In principle, Egr1 binding can occur through one of three pathways (Figure 2A): first, it can bind passively to the already closed DNA. Next, it can bind to the open ssDNA and catalyze closing of the duplex. Finally, the protein may bind to open ssDNA and wait for the DNA to close spontaneously. Assuming that the DNA thermal fluctuations are in rapid equilibrium relative to Egr1 binding, each one of these pathways predicts a different functional dependence on the probability P_{closed} of the DNA being closed (Supplementary Discussion). The first pathway suggests that k_{on}^{obs} will increase linearly, while the other two suggest k_{on}^{obs} will decrease linearly with P_{closed}. In order to elucidate which one is the dominant binding pathway, we exploit the fact that changing the distance between the traps (Figure 2B,C), and hence the applied tension on the fork, results in a very sensitive modulation of P_{closed}, from ∼1 to ∼0 in a ∼1 pN range (Figure 2D–F). These measurements reveal a linear dependence of k_{on}^{obs} on P_{closed} (Figure 2G), which is

an indication that Egr1 binds only to the closed DNA, exploiting a momentary and thermally driven closing of the fork (Figure 2A). No binding of Egr1 to single-stranded DNA was observed in full unzipping-rezipping curves (Supplementary Figure S10), consistent with this interpretation. Interestingly, the inhibition of binding by thermal fluctuations, together with the suppression of these fluctuations by a bound protein, offer an interesting mechanism for binding cooperativity between different TFs. Binding of one TF to its response element may stabilize the thermal breathing of a region that contains a response element of a second TF, thus increasing the binding rate, and effectively the affinity, of the second protein.

In contrast to the association rate, the observed dissociation rate is independent of the tension on the fork (Figure 2H). However, since our method is based on monitoring the DNA fluctuations, and thus requires $0 < P_{closed} < 1$, the range of forces we can probe is very narrow (Figure 2E). To widen this force range and elucidate whether the residence times we measure are affected by the applied tension, we develop a new assay (Figure 3A). First, we detect Egr1 binding via the suppression of the breathing fluctuations. Once a binding event is detected, we rapidly (with a 150 ± 50 ms response-time) move one of the traps in a single step to increase the applied tension. The distance between the two traps is then held constant until the force drops, indicating protein dissociation and revealing the bound state's lifetime at the specific force probed. Finally, the trap is restored to its original position allowing for additional measurement rounds. Using this approach, we can explore the residence time at an extended force range (Figure 3B, C). Theoretical models for such dynamic force spectroscopy experiments (59,60) predict an exponential decrease in the dissociation kinetics as a function of the force, allowing the extraction of the transition state's location and energy. However, our data do not show a simple exponential decay, but rather a turnover from a force-independent range to a force-dependent one that fits well to an exponential function (Figure 3C). This behavior seems to be independent of the sequence of the binding site (Figure 3D). This type of turnover has been observed in studies probing protein unfolding under force (61) and suggests the existence of two pathways for dissociation: a force-independent one and a forced-induced dissociation pathway with a different rate. Notably, given that unzipping the DNA partially mimics the progression of polymerases and helicases on the DNA, it is possible that these different dissociation pathways play a role in ensuring stable binding of the protein to DNA in the absence of a translocating enzyme, but efficient and rapid dissociation in their presence.

Since the turnover takes place at forces that are higher than those used in the fluctuation's measurements of Figure 1, these experiments reveal the inherent, or force-free association kinetics. With the typical concentrations used in this study, hundreds of binding and dissociation events can be identified in a single trace and used to calculate the cumulative distribution function (CDFs) for the bound state duration (Supplementary Figure S1E). In addition, we calculate the CDF for the *inherent* association time (for simplicity, the binding time), obtained by multiplying the observed binding times by the instantaneous P_{closed} (Figure 2I, Supplementary Figure S1F). While, in general, the shape of these distributions may reveal the existence of several underlying processes controlling association and dissociation and therefore may not be characterized by a single parameter, in this specific case we find that both are well-fitted by exponential functions, which allows us to extract the inherent, force-free binding and dissociation rates, k_{on} and k_{off}, respectively. Notably, although the measured k_{off} is highly consistent between different experimental sessions (Supplementary Figure S2B), we do observe some variability in the measured k_{on}, likely as a result of small changes in the effective concentration (Supplementary Figure S2A). To minimize this effect, we exploited a laminar flow cell where different samples flow in different non-mixing channels and are moved, alternatively, into a common channel containing the protein. We compare in what follows only values of k_{on} that were obtained in the same experimental session. With this approach, we can now ask how these rates are modulated.

The sequence and context of the binding site modulate the binding kinetics

In a previous work (39), using DNA unzipping to disrupt Egr1–DNA complexes at equilibrium, we reported significant differences in affinity between the consensus site and the three evolutionary-conserved binding sites at the proximal promoter of the *Lhb* gene. With the approach described here, we are now able to characterize the more informative *kinetics* of binding. In other words, while the previous study indicates that k_{on}/k_{off} is affected by the DNA sequence, we can now probe the effect that the sequence has on k_{on} and k_{off} separately. Measurements of the *Lhb* sites in their native genomic context, and the consensus site inserted in the 601 nucleosome positioning sequence (50) (Figure 4A) reveal that both the dissociation and association kinetic rates are modulated by the identity of the binding site (Figure 4B–D). To uncouple between the potential effect of the core binding sequence and that of the sequences surrounding the binding sites, we also probe the *Lhb* sites when inserted in the 601 context and the proximal promoter of the *Cga* gene (46) (Supplementary Figure S3A). Differences in k_{off}, which are also affected by the flanking sequences (Supplementary Figure S3B), are consistent with a modulation in the structure of the complex (39). Differences in k_{on} (Supplementary Figure S3C), suggest that binding is not diffusion-limited but includes an additional step of transition into a tightly bound complex, and that this step occurs at a different rate for each site. Interestingly, this 'isomerization' step is also affected by the sequences flanking the binding site, but differences between the sites are present even when inserted in a common context (Supplementary Figure S3C), indicating that both the core and flanking sequences can affect isomerization. A simple two-state model (Supplementary Figure S4) allows estimating their relative contributions (Supplementary Tables S1–S5), and reveals a non-additive effect for mutations in the core binding site, in contrast to previous studies (62) (Supplementary Discussion). Overall, our data stress the importance of the genomic environment as a modulator of binding kinetics.

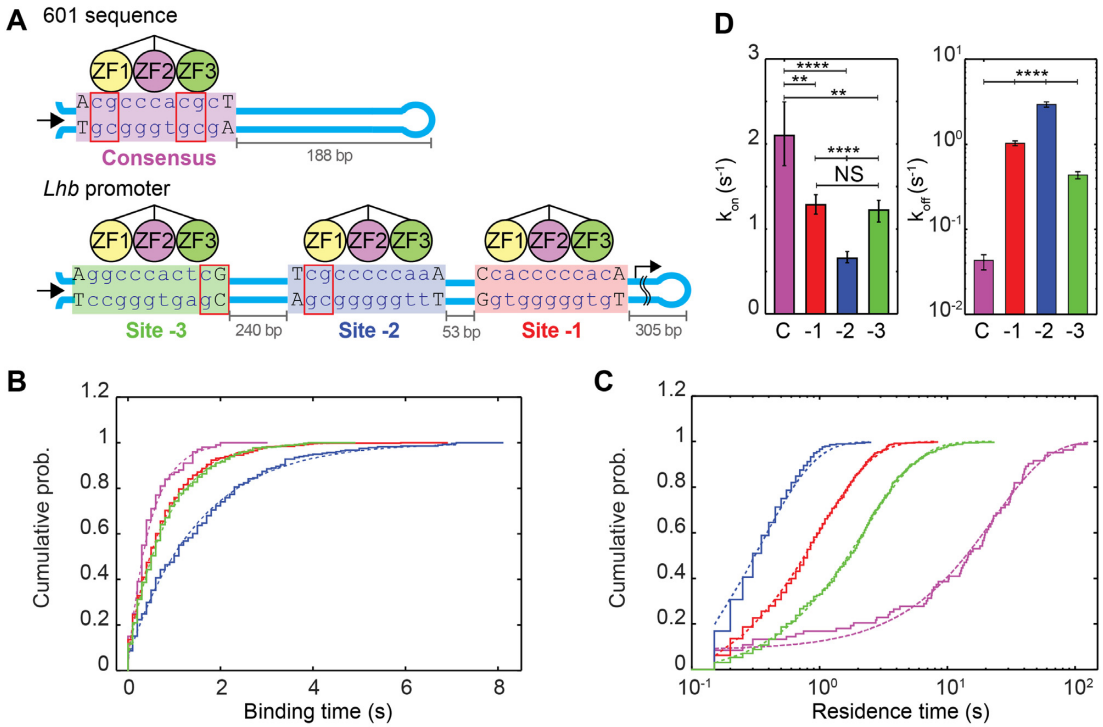

Figure 4. The sequence of Egr1 binding site modulates both its association and dissociation kinetics. (**A**) DNA constructs containing the consensus binding site inserted in the 601 DNA sequence (top) and the *Lhb* proximal promoter with its three identified sites (bottom). The orientation of Egr1 and the ZF contacts are illustrated. CpG sites are highlighted with red squares. The unzipping direction is indicated as a black arrow. (**B**, **C**) Cumulative probability distribution functions for the binding (B) and the residence (C) times, for the consensus (magenta), site –3 (green), site –2 (blue) and site –3 (red) at [Egr1] = 10 nM. A bin size of 0.2 s was used. The data presented were obtained in a single session. $N = 6$ molecules, 100 and 83 binding and dissociation events, respectively, for the consensus; 5, 511 and 508 for site –1; 4, 278, 296 for site –2; and 4, 359, 353 for site –3. A single-exponential fit is presented as a dashed line. A semi-log plot was used in (C) to accommodate the large differences in residence time between the consensus and the other sites. (**D**) Association (left) and dissociation (right) rates. The data in D show the rate obtained from the fit ±95% confidence interval obtained by bootstrapping (resampling the data 100 times, with replacement). $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{*}P < 0.05$, Kolmogorov–Smirnov's test of the distributions. The data shown in the figure was collected at the same Egr1 concentration and at the same session for accurate comparison.

Methylation induces a sequence- and position-dependent modulation of the binding kinetics

The conserved differences in the core sequences of the *Lhb* binding sites also imply that they harbor different potential CpG methylation sites, which can add an additional, and differential, layer of regulation to Egr1 binding to the different sites. Hence, we ask whether methylation can also modulate the binding kinetics, and if so, which of the kinetic rates is affected. We first probe a construct containing the consensus motif (Material and methods, Supplementary Figure S5), which contains two methylation sites, one on each side of the motif (Figure 4A), and characterize the distributions of binding and residence times (Figure 5A, B). Interestingly, while methylation by M.SssI methyltransferase does not affect k_{on} (Figure 5A), we observe a remarkable sensitivity of k_{off} for the methylation state (Figure 5B), which represents a 5-fold decrease in transcription factor residence time, and a reduction in the binding energy of ∼1 kcal/mol (Supplementary Discussion, Supplementary Table S6). This is in stark contrast with previous works (63,64), although likely the result of using a long DNA segment in our study (Supplementary Discussion, Supplementary Figure S6). Interestingly, an irreversible unzipping experiment, where we continuously increase the force to induce dissociation of a bound protein (39) reveals no differences in the binding position upon methylation, but significant differences in the breaking force (Supplementary Figure S7), which suggests a different conformation of the protein when bound to methylated DNA.

Next, when we repeat our assay using the *Lhb* sites, no significant effect is observed upon treatment with M.SssI for the binding time of any of the sites, consistently with the results for the consensus site (Figure 5C). The residence times for site –1, which has no CpG sites in its core binding site or near flanking sequences, and site –2, which has a single methylation site, are also insensitive to the methylation reaction (Figure 5D). However, site –3, which also contains a single CpG site but on an opposite orientation to that of site –2 (Figure 4A), shows a 2-fold increase in residence time, indicating that methylation makes the complex more stable, by –0.4 ± 0.2 kcal/mol (Figure 5D, Supplementary Table S6). Interestingly, the modulation provided by the methyl group in this case is opposite to that of the consensus motif, indicating that methylation of Egr1 binding sites can have both stabilizing and destabilizing effects.

To clarify whether the differences observed for site –3 and the consensus (stabilization and destabilization, respectively) are the result of their different core sequence, or their different flanking context, we measure the kinetics of site

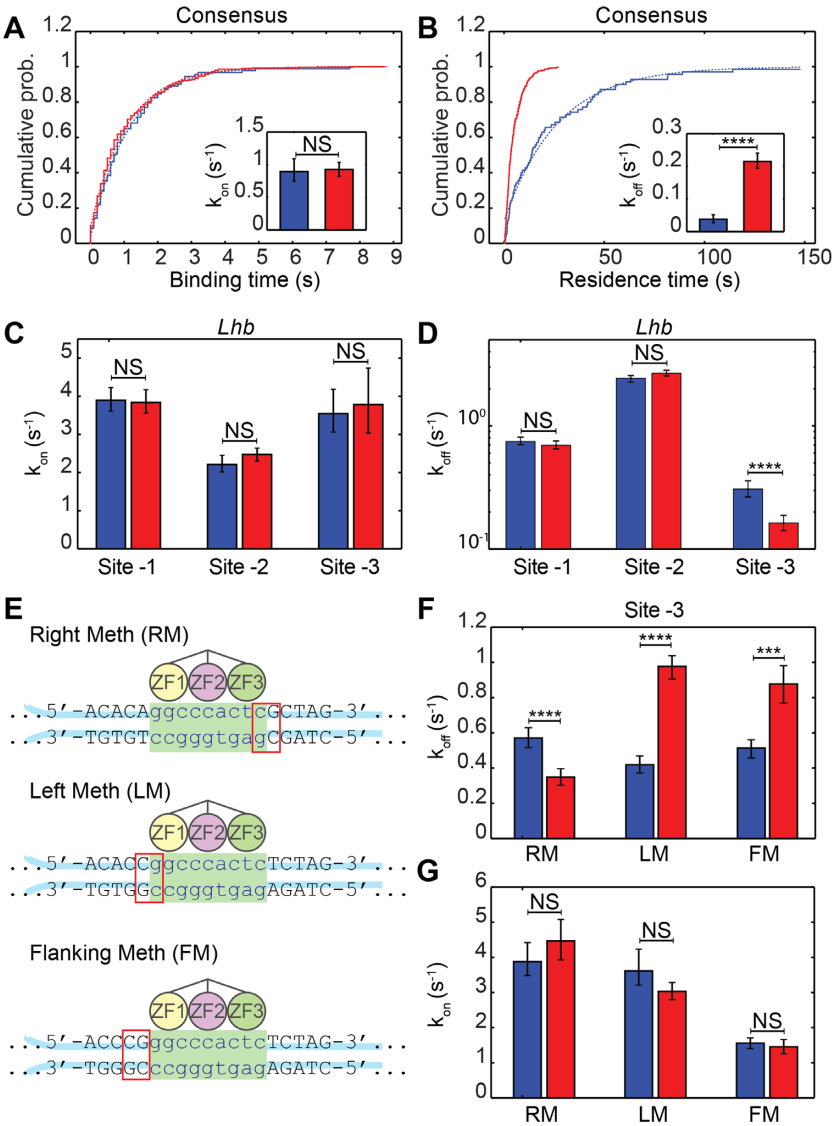

Figure 5. Methylation modulates the binding kinetics in a sequence- and position-depending manner. (**A**, **B**) Cumulative distributions of the binding (A) and residence (B) times for unmethylated (blue) or methylated (red) consensus binding site. The dashed line shows a single exponential fit. The inserts show the rate obtained from the fit ±95% confidence interval obtained by bootstrapping (resampling the data 100 times, with replacement). The experiment is done at 1 nM of Egr1. $N = 6$ molecules, 91 binding events, and 70 dissociation events for the unmethylated case; 4, 304, 299 for the methylated one. (**C**, **D**) Association (C) and dissociation (D) rates for the unmethylated (blue) and methylated (red) three sites in the *Lhb* promoter, calculated and presented as in (A) and (B). The experiment is done at 10 nM Egr1. $N = 3$ molecules, 544 and 535 detected binding and dissociation events, respectively, for site –1 unmethylated; 4, 759, 744 for site –1 methylated; 3, 449, 460 for site –2 unmethylated; 5, 707, 716 for site –2 methylated; 3, 210, 197 for site –3 unmethylated; 5, 124, 112 for site –3 methylated. (**E**) Sequence for site –3 from the *Lhb* promoter, integrated into a new context with a CpG inside the binding site, in its left (LM) and right (RM) sides, or a CpG outside the 9bps core of the binding site (FM). Egr1 binding orientation and ZFs contacts base pairs are illustrated. (**F**, **G**) Dissociation (F) and binding (G) rates for the constructs on (E), calculated and presented as in (A) and (B). The experiment is done at 10 nM Egr1 for constructs RM & LM, and at 20 nM for FM. $N = 4$ molecules, 263 and 264 detected binding and dissociation events, respectively, for unmethylated RM; 3, 202, 196 for methylated RM; 3, 276, 269 for unmethylated LM; 4, 489, 490 for methylated LM; 5, 302, 302 for unmethylated FM; 3, 173, 178 for methylated FM. ****$P < 0.0001$, ***$P < 0.001$, NS $P > 0.05$, Kolmogorov–Smirnov's *t*-test. Each methylated (red) and unmethylated (blue) pair was collected at the same Egr1 concentration and at the same session for accurate comparison

–3 inserted into the same context as the consensus motif (Figure 5E–G, construct RM). Consistent with the previous result, a significant decrease in k_{off} and no effect on k_{on} are observed upon methylation (Figure 5F,G), suggesting that the core composition, but not the flanking context is responsible in this case for the observed effect of DNA methylation. Since Egr1 binds asymmetrically to its binding motif, and demonstrates a greater sensitivity to a perturbation from the side of ZF3 (39), we hypothesize that the position of the methyl group relative to the ZFs may dictate the stabilizing or destabilizing effect of the methylation. Hence, we eliminate the CpG adjacent to ZF3 in site –3 and insert a new one adjacent to ZF1, without changing the 9 bp core sequence (Figure 5E, construct LM). Remarkably, the methylation effect on the stability is now reversed, resulting in destabilization of the complex (Figure 5F) and provid-

ing support for our hypothesis. Finally, shifting the methylation site on the LM construct to be outside the core binding site, but immediately adjacent to ZF1 (Figure 5E, construct FM), results in a similar destabilization effect upon methylation (Figure 5F). Hence, the same binding site can exhibit different and sometimes opposite responses to methylation, depending on the position of the CpG site, both in the core and flanking sequences. Together, the large but static modulation range provided by differences in core and flanking sequence, and the dynamic control of dissociation rate by CpG methylation, likely contribute to the broad functionality of a single TF.

DISCUSSION

Here, we demonstrate a single molecule biophysical assay that allows to follow, in real-time, the binding and dissociation of individual, unlabeled TFs to DNA. Our method is based on monitoring the thermal fluctuations of DNA, which are suppressed upon protein binding. With the ability to obtain the full probability distribution for the binding and residence times of a TF to a known site, while at the same time offering the possibility to gradually increase the complexity of the system and follow binding in the vicinity of nucleosomes (65) and chromatosomes (66), we expect this method to offer important mechanistic insights to complement the information obtained from live cell measurements.

We demonstrate our approach by studying binding of Egr1 to the *Lhb* proximal promoter. In line with our previous work, which showed that the equilibrium properties of binding are modulated by the conserved deviations found in the sequence of the *Lhb* sites and the conserved sequences around them, here we show that both the kinetic rates, k_{on} are k_{off}, are separately modulated by these factors. Different binding sites within the same DNA segment, measured at the same Egr1 concentration, exhibit differences in k_{on}, which characterizes the typical duration needed for Egr1 to bind in a stable conformation to its site and inhibit its fluctuations. On the other hand, k_{on} depends also on the concentration (Supplementary Figure S8), suggesting that the binding process has more than one step, with only the first one involving diffusion of the TF to the vicinity of the binding site. Notably, although closing of the DNA segment containing the binding site is a prerequisite for protein binding, the DNA breathing kinetics is much faster than the estimated rate for the second step in binding, suggesting that the existence of this additional step is not a result of our experimental method. Previous studies showed that Egr1 is able to perform 1D diffusion on the DNA (67,68), thus increasing the speed at which it finds its binding site by 'facilitated diffusion' (20). It does so by exploiting two different modes of DNA binding: the 'search' mode, with two ZFs attached to the DNA, has low binding energy and allow fast and efficient DNA scanning, while the 'recognition mode', where all three ZFs are bound to the DNA, has higher binding energy (67). The existence of these two conformations enables Egr1 to overcome the 'speed-stability' paradox, i.e. the contradicting needs for both efficient sliding on non-specific DNA and stable binding to a specific site (69). Since our experiments only detect stably bound proteins, one possible interpretation for the context-dependence of k_{on} is that the observed isomerization step involves, at least partially, 1D diffusion into the binding site. However, differences in k_{on} were observed also for different sites inserted in a common context, indicating that this effect may only partially explain the results. While it is also possible that a *transition* between the search mode and the recognition mode is modulated by the sequence of the site itself, a thorough elucidation of the nature of the different steps involved in reaching the binding site and stably binding to it will require additional studies. Remarkably, differences in residence time as large as 40-fold were observed as a result of 3 bp differences in the binding site, which is consistent with studies showing a broad, power-law distribution in the residence time for other transcription factors *in vivo* (12). Moreover, the high degree of sequence conservation observed for the binding sites of Egr1 suggests an important role played by these differences in residence time in the differential regulation of the many genes controlled by Egr1.

Different sequences correspond to different potential methylation patterns, and while CpG methylation does not affect the kinetics of Egr1 binding, our data shows that it modulates the dissociation of the protein. Notably, while previous structural studies predicted a steric clash by the addition of a methyl group (70), and therefore a reduced stability, a more recent one reported that the structure of Egr1 'adapts' its conformation and displays the same affinity even when the target sequence is completely methylated (64), in contrast with our results. The observed modulation of the dissociation is sensitive to the position of the CpG in the binding site or its vicinity and can result in both stabilization and destabilization of the complex. Interestingly, a structural study of CTCF, which contains an array of 11 ZFs, also revealed sensitivity for CpG methylation in a position-dependent manner (71). How does a methyl group in the binding site affect the stability of the complex? While one possible mechanism is steric hindrance by the methyl group, the effect observed for a methylated CpG outside the core binding site, and the fact that methylation can stabilize as well as destabilize the bound complex, suggest that part of the effect is related to changes in the local structure of DNA. Although similar crystal structures were reported for methylated and unmethylated DNA (72), methylation was shown to affect the flexibility (73,74), hydration structure (75), and sugar pucker conformations (76) of DNA. Molecular dynamics simulations showed that methylation increases, locally, the propensity of DNA toward different values of roll and propeller twist, and that the position of the modification and the local sequence context has significant effects on the amount of structural variation observed (77). The formation of protein-DNA complexes involves changes, or distortions, in the structure of DNA, and these are clearly accompanied by an energetic cost. A recent study highlighted this by showing that, in many cases, the affinity of TFs for DNA that includes mismatches, and is therefore distorted *a priori*, is higher than that for non-perturbed DNA (78). This increase in affinity correlates with distortions that resemble those that are observed in the DNA when it is part of a bound complex. A similar mechanism of 'pre-paying' the energetic cost of binding by distorting the DNA, may explain the observed effect of the methylation on the binding of Egr1. In addi-

tion, DNA molecules are polymorphic and dynamic, locally exploring different conformations, that are also sequence-dependent (79–81), and methylation may affect the dynamic equilibrium between different structural states (82–85), so it is possible that increasing the relative occupancy of the properly distorted state in the ensemble, or modifying the rates of interconversion between different structures, contribute to our results.

Interestingly, we observe a high degree of destabilization upon methylation for the consensus sequence. A very long residence time may be necessary to fulfill a specific function, e.g. functioning as a pioneer transcription factor (32) or driving cell differentiation (86). However, such a long residence time may also be toxic for the cell in other circumstances. Hence, methylation of the consensus site may serve as an important switch to prevent this toxicity, while allowing long binding 'on-demand'.

In summary, our study demonstrates a novel single-molecule approach that allows characterizing the different factors that affect the binding and dissociation kinetics of a TF, or other DNA binding proteins. With Egr1 as a model, we demonstrate that the sequence of DNA (at the core binding site and around it) and its methylation state are powerful and versatile modulators of these kinetics. While our approach is not devoid of limitations (Supplementary Discussion), the main one being the requirement of a clear separation of timescales between DNA breathing and protein binding, we expect our method to be widely applicable. In a broader context, previous studies have shown that significant genetic variants are located in non-coding regions, and that complex phenotypes are the result of altered binding of TFs (87). Together with previous results showing the importance of the binding kinetics in tailoring the transcriptional response, our results demonstrating that these kinetics are sensitive to small differences in core sequence, environment, and methylation state, may help explain the mechanism by which these phenotypes arise.

DATA AVAILABILITY

All data are available in the main text or the supplementary materials.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Israel Science Foundation [1782/17, 937/20 to A.K. and 1850/17 to P.M.]. Funding for open access charge: Israel Science Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Jacob,F. and Monod,J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.
2. Polach,K.J.J. and Widom,J. (1995) Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.*, **254**, 130–149.
3. Klemm,S.L., Shipony,Z. and Greenleaf,W.J. (2019) Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.*, **20**, 207–220.
4. Lickwar,C.R., Mueller,F., Hanlon,S.E., McNally,J.G. and Lieb,J.D. (2012) Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*, **484**, 251–255.
5. Clauß,K., Popp,A.P., Schulze,L., Hettich,J., Reisser,M., Torres,L.E., Uhlenhaut,N.H., Gebhardt,J.C.M., Christof,J., Gebhardt,M. *et al.* (2017) DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res.*, **45**, 11121–11130.
6. Stavreva,D.A., Garcia,D.A., Fettweis,G., Gudla,P.R., Zaki,G.F., Soni,V., McGowan,A., Williams,G., Huynh,A., Palangat,M. *et al.* (2019) Transcriptional bursting and co-bursting regulation by steroid hormone release pattern and transcription factor mobility. *Mol. Cell*, **75**, 1161–1177.
7. Popp,A.P., Hettich,J.M. and Gebhardt,J.C. (2021) Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. *Nucleic Acids Res.*, **49**, 6249–6266.
8. Hipp,L., Beer,J., Kuchler,O., Reisser,M., Sinske,D., Michaelis,J., Gebhardt,J.C.M. and Knöll,B. (2019) Single-molecule imaging of the transcription factor SRF reveals prolonged chromatin-binding kinetics upon cell stimulation. *Proc. Natl. Acad. Sci.*, **116**, 880–889.
9. Loffreda,A., Jacchetti,E., Antunes,S., Rainone,P., Daniele,T., Morisaki,T., Bianchi,M.E., Tacchetti,C. and Mazza,D. (2017) Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. *Nat. Commun.*, **8**, 313.
10. Senecal,A., Munsky,B., Proux,F., Ly,N., Braye,F.E., Zimmer,C., Mueller,F. and Darzacq,X. (2014) Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep.*, **8**, 75–83.
11. Donovan,B.T., Huynh,A., Ball,D.A., Patel,H.P., Poirier,M.G., Larson,D.R., Ferguson,M.L. and Lenstra,T.L. (2019) Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *EMBO J.*, **38**, e100809.
12. Garcia,D.A., Fettweis,G., Presman,D.M., Paakinaho,V., Jarzynski,C., Upadhyaya,A. and Hager,G.L. (2021) Power-law behavior of transcription factor dynamics at the single-molecule level implies a continuum affinity model. *Nucleic Acids Res.*, **49**, 6605–6620.
13. Erbaş,A. and Marko,J.F. (2019) How do DNA-bound proteins leave their binding sites? The role of facilitated dissociation. *Curr. Opin. Chem. Biol.*, **53**, 118–124.
14. Suter,D.M. (2020) Transcription factors and DNA play hide and seek. *Trends Cell Biol.*, **30**, 491–500.
15. Chen,T.Y., Cheng,Y.S., Huang,P.S. and Chen,P. (2018) Facilitated unbinding via multivalency-enabled ternary complexes: new paradigm for protein-DNA interactions. *Acc. Chem. Res.*, **51**, 860–868.
16. Izeddin,I., Récamier,V., Bosanac,L., Cissé,I.I., Boudarene,L., Dugast-Darzacq,C., Proux,F., Bénichou,O., Voituriez,R., Bensaude,O. *et al.* (2014) Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *Elife*, **3**, e02230.
17. Hansen,A.S., Amitai,A., Cattoglio,C., Tjian,R. and Darzacq,X. (2020) Guided nuclear exploration increases CTCF target search efficiency. *Nat. Chem. Biol.*, **16**, 257–266.
18. Normanno,D., Boudarène,L., Dugast-Darzacq,C., Chen,J., Richter,C., Proux,F., Bénichou,O., Voituriez,R., Darzacq,X. and Dahan,M. (2015) Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat. Commun.*, **6**, 7357.
19. McNally,J.C., Müller,W.G., Walker,D., Wolford,R. and Hager,G.L. (2000) The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science*, **287**, 1262–1265.
20. Kolomeisky,A.B. (2011) Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.*, **13**, 2088–2095.
21. Hettich,J. and Gebhardt,J.C.M. (2018) Transcription factor target site search and gene regulation in a background of unspecific binding sites. *J. Theor. Biol.*, **454**, 91–101.
22. Dror,I., Rohs,R. and Mandel-Gutfreund,Y. (2016) How motif environment influences transcription factor search dynamics: finding a needle in a haystack. *Bioessays*, **38**, 605–612.

23. Tims,H.S., Gurunathan,K., Levitus,M. and Widom,J. (2011) Dynamics of nucleosome invasion by DNA binding proteins. *J. Mol. Biol.*, **411**, 430–448.
24. Luo,Y., North,J.A., Rose,S.D. and Poirier,M.G. (2014) Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res.*, **42**, 3017–3027.
25. Deplancke,B., Alpern,D. and Gardeux,V. (2016) The genetics of transcription factor DNA binding variation. *Cell*, **166**, 538–554.
26. Liu,Z. and Tjian,R. (2018) Visualizing transcription factor dynamics in living cells. *J. Cell Biol.*, **217**, 1181–1191.
27. Elf,J. and Barkefors,I. (2019) Single-molecule kinetics in living cells. *Annu. Rev. Biochem.*, **88**, 635–659.
28. Lionnet,T. and Wu,C. (2021) Single-molecule tracking of transcription protein dynamics in living cells: seeing is believing, but what are we seeing? *Curr. Opin. Genet. Dev.*, **67**, 94–102.
29. Landgraf,D., Okumus,B., Chien,P., Baker,T.A. and Paulsson,J. (2012) Segregation of molecules at cell division reveals native protein localization. *Nat. Methods*, **9**, 480–482.
30. Wisniewski,J., Hajj,B., Chen,J., Mizuguchi,G., Xiao,H., Wei,D., Dahan,M. and Wu,C. (2014) Imaging the fate of histone Cse4 reveals de novo replacement in S phase and subsequent stable residence at centromeres. *Elife*, **3**, e02203.
31. Luo,Y., North,J.A. and Poirier,M.G. (2014) Single molecule fluorescence methodologies for investigating transcription factor binding kinetics to nucleosomes and DNA. *Methods*, **70**, 108–118.
32. Donovan,B.T., Chen,H., Jipa,C., Bai,L. and Poirier,M.G. (2019) Dissociation rate compensation mechanism for budding yeast pioneer transcription factors. *Elife*, **8**, e43008.
33. Mivelaz,M., Cao,A.M., Kubik,S., Zencir,S., Hovius,R., Boichenko,I., Stachowicz,A.M., Kurat,C.F., Shore,D. and Fierz,B. (2020) Chromatin fiber invasion and nucleosome displacement by the Rap1 transcription factor. *Mol. Cell*, **77**, 488–500.
34. Kilic,S., Bachmann,A.L., Bryan,L.C. and Fierz,B. (2015) Multivalency governs HP1α association dynamics with the silent chromatin state. *Nat. Commun.*, **6**, 7313.
35. Thiel,G. and Cibelli,G. (2002) Regulation of life and death by the zinc finger transcription factor Egr-1. *J. Cell. Physiol.*, **292**, 287–292.
36. Melamed,P., Haj,M., Yosefzon,Y., Rudnizky,S., Wijeweera,A., Pnueli,L. and Kaplan,A. (2018) Multifaceted targeting of the chromatin mediates gonadotropin-releasing hormone effects on gene expression in the gonadotrope. *Front. Endocrinol. (Lausanne).*, **9**, 58.
37. Pavletich,N. and Pabo,C. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809–817.
38. Christy,B. and Nathans,D. (1989) DNA binding site of the growth factor-inducible protein Zif268. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 8737–8741.
39. Rudnizky,S., Khamis,H., Malik,O., Squires,A.H., Meller,A., Melamed,P. and Kaplan,A. (2018) Single-molecule DNA unzipping reveals asymmetric modulation of a transcription factor by its binding site sequence and context. *Nucleic Acids Res.*, **46**, 1513.
40. Chen,M., Xiong,F. and Zhang,L. (2013) Promoter methylation of Egr-1 site contributes to fetal hypoxia-mediated PKCε gene repression in the developing heart. *Am. J. Physiol. Integr. Comp. Physiol.*, **304**, R683–R689.
41. Lawrence,J., Chen,M., Xiong,F., Xiao,D., Zhang,H., Buchholz,J.N. and Zhang,L. (2011) Foetal nicotine exposure causes PKCε gene repression by promoter methylation in rat hearts. *Cardiovasc. Res.*, **89**, 89–97.
42. Ogishima,T., Shiina,H., Breault,J.E., Terashima,M., Honda,S., Enokida,H., Urakami,S., Tokizane,T., Kawakami,T., Ribeiro-Filho,L.A. *et al.* (2005) Promoter CpG hypomethylation and transcription factor EGR1 hyperactivate heparanase expression in bladder cancer. *Oncogene*, **24**, 6765–6772.
43. Romens,S.E., McDonald,J., Svaren,J. and Pollak,S.D. (2015) Associations between early life stress and gene methylation in children. *Child Dev.*, **86**, 303–309.
44. Yosefzon,Y., David,C., Tsukerman,A., Pnueli,L., Qiao,S., Boehm,U. and Melamed,P. (2017) An epigenetic switch repressing Tet1 in gonadotropes activates the reproductive axis. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 10131–10136.
45. Malik,O., Khamis,H., Rudnizky,S., Marx,A. and Kaplan,A. (2017) Pausing kinetics dominates strand-displacement polymerization by reverse transcriptase. *Nucleic Acids Res.*, **45**, 10190–10205.
46. Rudnizky,S., Bavly,A., Malik,O., Pnueli,L., Melamed,P. and Kaplan,A. (2016) H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. *Nat. Commun.*, **7**, 12958.
47. Gittes,F. and Schmidt,C.F. (1998) Interference model for back-focal-plane displacement detection in optical tweezers. *Opt. Lett.*, **23**, 7–9.
48. Berg-Sørensen,K., Flyvbjerg,H., Introduction,I., Berg-Sørensen,K. and Flyvbjerg,H. (2004) Power spectrum analysis for optical tweezers. *Rev. Sci. Instrum.*, **75**, 594–612.
49. Brouwer,I., King,G.A., Heller,I., Biebricher,A.S., Peterman,E.J.G. and Wuite,G.J.L. (2017) Probing DNA–DNA interactions with a combination of quadruple-trap optical tweezers and microfluidics. In: *Optical Tweezers: Methods and Protocols*. Humana Press, NY, pp. 275–293.
50. Lowary,P and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19–42.
51. Marko,J.F. and Siggia,E.D. (1995) Stretching DNA. *Macromolecules*, **28**, 8759–8770.
52. Seol,Y., Li,J., Nelson,P.C., Perkins,T.T. and Betterton,M.D. (2007) Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6-7 microm. *Biophys. J.*, **93**, 4360–4373.
53. Bosco,A., Camunas-Soler,J. and Ritort,F. (2014) Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.*, **42**, 2064–2074.
54. Koch,S.J. and Wang,M.D. (2003) Dynamic force spectroscopy of protein-DNA interactions by unzipping DNA. *Phys. Rev. Lett.*, **91**, 028103.
55. Koch,S.J., Shundrovsky,A., Jantzen,B.C. and Wang,M.D. (2002) Probing protein-DNA interactions by unzipping a single DNA double helix. *Biophys. J.*, **83**, 1098–1105.
56. Jiang,J., Bai,L., Surtees,J.A., Gemici,Z., Wang,M.D. and Alani,E. (2005) Detection of high-affinity and sliding clamp modes for MSH2-MSH6 by single-molecule unzipping force analysis. *Mol. Cell*, **20**, 771–781.
57. McKinney,S.A., Joo,C. and Ha,T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.*, **91**, 1941–1951.
58. von Hippel,P.H., Johnson,N.P. and Marcus,A.H. (2013) Fifty years of DNA 'breathing': reflections on old and new approaches. *Biopolymers*, **99**, 923–954.
59. Bell,G.I. (1978) Models for the specific adhesion of cells to cells. *Science*, **200**, 618–627.
60. Dudko,O., Hummer,G. and Szabo,A. (2006) Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, **96**, 108101.
61. Best,R.B., Paci,E., Hummer,G. and Dudko,O.K. (2008) Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J. Phys. Chem. B*, **112**, 5968–5976.
62. Chattopadhyay,A., Zandarashvili,L., Luu,R.H. and Iwahara,J. (2016) Thermodynamic additivity for impacts of base-pair substitutions on association of the Egr-1 zinc-finger protein with DNA. *Biochemistry*, **55**, 6467–6474.
63. Hashimoto,H., Olanrewaju,Y.O., Zheng,Y., Wilson,G.G. and Zhang,X. (2014) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.*, **28**, 2304–2313.
64. Zandarashvili,L., White,M.A., Esadze,A. and Iwahara,J. (2015) Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1. *FEBS Lett.*, **589**, 1748–1753.
65. Rudnizky,S., Khamis,H., Malik,O., Melamed,P. and Kaplan,A. (2019) The base pair-scale diffusion of nucleosomes modulates binding of transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 12161–12166.
66. Rudnizky,S., Khamis,H., Ginosar,Y., Goren,E., Melamed,P. and Kaplan,A. (2021) Extended and dynamic linker histone-DNA Interactions control chromatosome compaction. *Mol. Cell*, **81**, 3410–3421.
67. Zandarashvili,L., Vuzman,D., Esadze,A., Takayama,Y., Sahu,D., Levy,Y. and Iwahara,J. (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E1724–E1732.

68. Zandarashvili,L., Esadze,A., Vuzman,D., Kemme,C.A., Levy,Y. and Iwahara,J. (2015) Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5142–E5149.
69. Veksler,A. and Kolomeisky,A.B. (2013) Speed-selectivity paradox in the protein search for targets on DNA: is it real or not? *J. Phys. Chem. B*, **117**, 12695–12701.
70. Elrod-Erickson,M., Benson,T.E. and Pabo,C.O. (1998) High-resolution structures of variant Zif268–DNA complexes: implications for understanding zinc finger–DNA recognition. *Structure*, **6**, 451–464.
71. Hashimoto,H., Wang,D., Horton,J.R., Zhang,X., Corces,V.G., Cheng,X., Kunde-Ramamoorthy,G., Nosavanh,L.M., Shen,L., Mishiro,T. *et al.* (2017) Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell*, **40**, 2747–2758.
72. Renciuk,D., Blacque,O., Vorlickova,M. and Spingler,B. (2013) Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.*, **41**, 9891–9900.
73. Ngo,T.T.M.M., Yoo,J., Dai,Q., Zhang,Q., He,C., Aksimentiev,A. and Ha,T. (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.*, **7**, 10813
74. Zaichuk,T. and Marko,J.F. (2021) Single-molecule micromanipulation studies of methylated DNA. *Biophys. J.*, **120**, 2148–2155.
75. Teng,X. and Hwang,W. (2018) Effect of methylation on local mechanics and hydration structure of DNA. *Biophys. J.*, **114**, 1791–1803.
76. Banyay,M. and Gräslund,A. (2002) Structural effects of cytosine methylation on DNA sugar pucker studied by FTIR. *J. Mol. Biol.*, **324**, 667–676.
77. Rao,S., Chiu,T.P., Kribelbauer,J.F., Mann,R.S., Bussemaker,H.J. and Rohs,R. (2018) Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding. *Epigenetics Chromatin*, **11**, 6.
78. Afek,A., Shi,H., Rangadurai,A., Sahay,H., Senitzki,A., Xhani,S., Fang,M., Salinas,R., Mielko,Z., Pufall,M.A. *et al.* (2020) DNA mismatches reveal conformational penalties in protein–DNA recognition. *Nature*, **587**, 291–296.
79. Maehigashi,T., Hsiao,C., Kruger Woods,K., Moulaei,T., Hud,N. V. and Williams,L.D. (2012) B-DNA structure is intrinsically polymorphic: even at the level of base pair positions. *Nucleic Acids Res.*, **40**, 3714–3722.
80. Fraser,J.S., Clarkson,M.W., Degnan,S.C., Erion,R., Kern,D. and Alber,T. (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature*, **462**, 669–673.
81. Nikolova,E.N., Kim,E., Wise,A.A., O'Brien,P.J., Andricioaei,I. and Al-Hashimi,H.M. (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature*, **470**, 498–504.
82. Behe,M. and Felsenfeld,G. (1981) Effects of methylation on a synthetic polynucleotide: the B–Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 1619–1623.
83. Frederick,C.A., Saal,D., Van Der Marel,G.A., Van Boom,J.H., Wang,A.H.-J. and Rich,A. (1987) The crystal structure of d(GGm5CCGGCC): the effect of methylation on A-DNA structure and stability. *Biopolymers*, **26**, S145–S160.
84. Nishimura,Y., Torigoe,C. and Tsuboi,M. (1986) Salt induced B — a transition of poly(dG).poly(dC) and the stabilization of A form by its methylation. *Nucleic Acids Res.*, **14**, 2737–2748.
85. Geahigan,K.B., Meints,G.A., Hatcher,M.E., Orban,J. and Drobny,G.P. (2000) The dynamic impact of CpG methylation in DNA. *Biochemistry*, **39**, 4939–4946.
86. Gurdon,J.B., Javed,K., Vodnala,M. and Garrett,N. (2020) Long-Term association of a transcription factor with its chromatin binding site can stabilize gene expression and cell fate commitment. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 15075–15084.
87. Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A., Abecasis,G.R. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.