Biophysical Journal
Article

Biophysical Society

Dynamics and inhibition of MLL1 CXXC domain on DNA revealed by single-molecule quantification

Lin Liang,[1] Kangkang Ma,[1] Zeyu Wang,[1] Richard Janissen,[2] and Zhongbo Yu[1,*]
[1]State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy, Nankai University, Tianjin, China and [2]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft, South-Holland, The Netherlands

ABSTRACT CpG islands recruit MLL1 via the CXXC domain to modulate chromatin structure and regulate gene expression. The amino acid motif of CXXC also plays a pivotal role in MLL1's structure and function and serves as a target for drug design. In addition, the CpG pattern in an island governs spatially dependent collaboration among CpGs in recruiting epigenetic enzymes. However, current studies using short DNA fragments cannot probe the dynamics of CXXC on long DNA with crowded CpG motifs. Here, we used single-molecule magnetic tweezers to examine the binding dynamics of MLL1's CXXC domain on a long DNA with a CpG island. The mechanical strand separation assay allows profiling of protein-DNA complexes and reports force-dependent unfolding times. Further design of a hairpin detector reveals the unfolding time of individual CXXC-CpG complexes. Finally, in a proof of concept we demonstrate the inhibiting effect of dimethyl fumarate on the CXXC-DNA complexes by measuring the dose response curve of the unfolding time. This demonstrates the potential feasibility of using single-molecule strand separation as a label-free detector in drug discovery and chemical biology.

SIGNIFICANCE Epigenetic enzymes of the CXXC family interact with chromatin by recognizing CpG motifs, which govern epigenetic modifications. These modifications regulate gene expression, making CXXC a potential target for therapeutic intervention. MLL1's CXXC domain recognizes unmodified CpG and prevents DNA methylation, and mutations in CXXC and abnormal methylation patterns of CpG islands lead to mixed-lineage leukemia. We use single-molecule methods to understand the dynamics of MLL1's CXXC domain on a long DNA containing a CpG island and determine the landscape of binding energies. Our methods also reveal the inhibitory effect of a clinically used drug on the protein-DNA complexes.

INTRODUCTION

CpG islands (CGIs) recruit epigenetic enzymes of the CXXC family to modulate chromatin structure and regulate gene expression, which are potential targets for therapeutic intervention (1). For example, MLL1, as a member of the CXXC family, regulates the expression of more than 5000 genes (2–5). MLL1's CXXC domain adopts a crescent-shaped conformation and binds CpG motifs to protect the DNA from methylation. It interacts with CpG via eight conserved cysteine residues (6,7) and a positively-charged surface. Within a CGI, the CpG pattern has been proposed to govern spatially dependent collaboration among CpGs in recruiting epigenetic enzymes (8,9). However, current biophysical studies of CXXC-DNA interactions by, e.g., x-ray crystallography or isothermal titration calorimetry, have used short DNA fragments (<20 bp, for example), which cannot probe the dynamics of a CXXC protein on a long CGI DNA with crowded CpG motifs.

Amino acid motifs such as CXXC play pivotal roles in protein structure and function, and broadly serve as targets for drug design. Indeed, selective modification of a protein at a targeted motif is a prevailing tool in chemical biology and drug discovery (10). For example, the myeloablative drug busulfan converts the cysteines of a catalytic CXXC motif to dehydroalanine and lanthionine in thiol redoxin proteins (11). In another example, CD28 interacts with the CXXC motif of protein kinase Cθ, which can be targeted by a small molecule, dimethyl fumarate (DMF) (12). DMF is clinically used to treat psoriasis and multiple sclerosis. Emerging targets of DMF such as NF-κB and GAPDH (13–16) suggest it may impact tumors.

Submitted November 11, 2020, and accepted for publication March 11, 2021.

*Correspondence: zyu@nankai.edu.cn
Lin Liang and Kangkang Ma contributed equally to this work.

Editor: Gijs Wuite.
https://doi.org/10.1016/j.bpj.2021.03.045

Identification of drug targets and quantitative evaluation of the dose response are critical for optimal drug development and rely on orthogonal biophysical methods (17). However, the readout of a detection method should be robust and avoid artifacts. For example, resveratrol was reported to be a sirtuin activator, but this was later found to be an experimental artifact caused by a fluorophore modification on the peptide substrate, i.e., resveratrol activity is entirely dependent on a fluorophore covalently attached to the enzyme's substrate (18–21).

On the other hand, label-free single-molecule methods like magnetic tweezers, optical tweezers, and nanopores (22–25) have provided quantitative measurements of the dynamics of DNA unzipping, DNA supercoiling, protein-DNA interactions, and the impact of drugs on all of these without the danger of label-induced artifacts. For example, single-molecule magnetic tweezers reveal that the antitumor drug topotecan impedes DNA uncoiling by topoisomerase I (26). DNA strand separation by single-molecule tools has been used to probe protein-DNA interactions, site-specifically probing the dynamics of enzyme-targeting DNA (27–34).

Here, we used magnetic tweezers to systematically examine the dynamics of MLL1's CXXC domain binding a long CGI-containing DNA from the Hoxa9 gene. The single-molecule strand separation assay allows profiling of protein-DNA complexes and reports force-dependent hairpin unfolding times (protein binding increases the unfolding time of the hairpin). In addition, we optimize the hairpin detector to probe the unfolding time of individual CXXC-CpG complexes by designing a DNA sequence with well-isolated CpGs and an AT-only (A for adenine and T for thymine) background sequence. The rupture of individual MLL1 CXXC-CpG complexes happens in ~10 ms at a probing force of 12.5 pN. In a proof of concept, we have also demonstrated the effect of the clinically used drug DMF on the MLL1 CXXC-DNA complexes by measuring the response of the unfolding time to DMF dosage. The measured half-maximal inhibitory concentration (IC50) of DMF on the protein-DNA interactions is ~1 μM. The effect of DMF on complexes formed between the CXXC family protein of MLL1 and a CpG site is unprecedented. Our results reveal the binding landscape of MLL1's CXXC domain on a CpG island and the dissociation time upon strand separation, shedding light on the dynamics of CXXC-CpG interactions. We show that DMF inhibits the function of MLL1's CXXC domain, demonstrating the feasibility of using single-molecule strand separation as a label-free detector in drug discovery and chemical biology.

MATERIALS AND METHODS

Unless noted otherwise, all chemicals were ordered from Sigma-Aldrich, the DNA oligonucleotides (oligos) from Sangon Biotech (Shanghai, China), and the enzymes from New England Biolabs. MLL1 CXXC was synthesized by Sangon Biotech (Table S1). We purchased DMF and monomethyl fumarate (MMF) from Shanghai Darui Fine Chemicals (Shanghai, China).

Electrophoretic mobility shift assay

We used a FAM-labeled oligo, FAM-GCCACCGGTGGC, to prepare CpG dsDNA for an electrophoretic mobility shift assay (EMSA). The palindromic ssDNA forms dsDNA by self-dimerization after being heated to 95°C for 0.5 min and slowly cooled to 25°C over 2 h. In a titration buffer containing 10 mM Tris HCl (pH 7.4), 150 mM NaCl, 50 μM $ZnCl_2$, 1 mM DTT, and 6% glycerol, we titrated 100 nM CpG dsDNA with the MLL1 CXXC domain at concentrations of 0.5, 1, 2, 4, 8, 16, 32, 64, and 128 μM. The reaction ran for 30 min on ice. We then loaded 10 μL of the reaction mixture into a 5% polyacrylamide gel, which ran at room temperature and 50 V for 1 h in 0.5× TBE (Tris-borate-EDTA) buffer. In an imaging system (PXi9; Syngene, Cambridge, UK), we took gel images with 495 nm illumination and analyzed the images in ImageJ (NIH Image, Bethesda, MD). Hill equation fitting was done using nonlinear least squares in MATLAB 2017 (The MathWorks, Natwick, MA). The Hill equation is $\theta = \frac{1}{(1+(K_d/c)^n)}$, where θ represents the fraction of probe depletion, c-protein concentration, K_d the dissociation constant, and n the Hill coefficient.

DNA constructs

In our hairpin constructs, CGI hairpin and CpG hairpin, two handles serve for mechanical manipulation, whereas the stem contains sequences of interest. We made the handles by PCR from pBluescript II SK(+) (catalog no. 212205; Agilent, Santa Clara, CA) and deoxynucleoside triphosphate (dNTP) supplemented with biotin-16-deoxyuridine triphosphate(dUTP) or digoxigenin-11-dUTP (catalog no. 11093070910 or 11093088910; Roche Diagnostic, Rotkreuz, Switzerland) (PCR primers in Table S2). The handles were digested with either BbvCI or PpuMI restriction enzyme to allow ligation of the two handles to distinct ends of the hairpin junction, which is assembled from four DNA oligos (junction oligos in Table S2). We ran PCR to generate the CGI sequence for the CGI hairpin stem sealed by a 4-T loop (primers and the loop oligo in Table S2). For the CpG hairpin, we used the same handles as the CGI hairpin. DNA linkers are synthesized oligos that connect the handles and the synthesized junction (linker oligos and junction oligos in Table S2). The CpG hairpin stem and loop are made of synthesized oligos (stem and loop oligos in Table S2).

Single-molecule assays using magnetic tweezers

We used a homemade magnetic tweezers setup as previously described (35–38). In a flow cell with the bottom glass slide covered with nitrocellulose (0.1%, m/v), we injected 70 μL of antidigoxigenin antibody (0.1 mg/mL, catalog no. 11093274910; Roche Diagnostic) and incubated it for 1 h, which was followed by passivation with BSA (5 mg/mL) overnight. Next, we mixed 1 ng of DNA construct with 20 μL of streptavidin-coated superparamagnetic beads (catalog no. 65305, Dynabeads M-270 Streptavidin; Invitrogen, Waltham, MA) in 40 μL of buffer containing 10 mM Tris HCl (pH 7.4) and 100 mM NaCl. After 30 min of incubation on ice, we loaded the 40 μL of DNA-bead mixture into the flow cell. Antidig antibodies on the flow cell surface immobilize DNA constructs which are bound to beads by their biotinylated end.

We ran single-molecule mechanical assays in a magnetic tweezers (MT) buffer containing 10 mM Tris HCl (pH 7.4), 100 mM NaCl, 50 μM $ZnCl_2$, 0.003% Tween-20, and 5 mM DTT (Table S3).

Mechanical assays use a force protocol with a sampling rate of >200 Hz. At low forces (F_{low}), the DNA hairpin is closed, and the DNA extension is short. In 390 ± 30 ms (mean ± SD, N = 10), we increased the force from F_{low} to a probing force (F_{probe}) where the hairpin is fully open (long

extension). CXXC binding events can block the opening of the hairpin, causing intermediate extensions and pausing before opening. When the force increases from F_{probe} to a higher force (F_{high}), the DNA hairpin should fully open and further extend. We then decreased from F_{high} to F_{probe} for reference purposes. We finally decreased from F_{probe} to F_{low}, completing a round of the force protocol. F_{low}, F_{probe}, F_{high}, and their durations can be tuned to the target proteins and hairpin constructs, similar to published methods (27,29–31,39).

Data analysis of single-molecule assays

We analyzed the single-molecule data in MATLAB (R2017a). To examine protein-binding positions, we first corrected the zero position for each trace by subtracting the DNA extension when the hairpin is fully open (Z_{open}) at F_{probe}. After zero-correction at F_{probe}, DNA extension represents the ssDNA length released by unzipping a hairpin. Next, we built a histogram of the DNA extension for each trace. Histogram peaks in the hairpin stem region indicate the blockage of DNA unzipping caused by protein binding. When the CpG hairpin construct is bound by proteins, evenly distributed CpG sites in the stem form distinct "peak" features in the histogram, which serve as a ruler for unit conversion from nanometers to bp of dsDNA.

To convert ssDNA length at F_{probe} from nm to nt, we used the Marko-Sigga formula of a worm-like chain (WLC) model, which presents the force (F) as a function of extension (x) (40),

$$F = \frac{k_B T}{L_p^{ss}}\left(\frac{1}{4}\left(1-\frac{x}{L_c^{ss}}\right)^{-2}+\frac{x}{L_c^{ss}}-\frac{1}{4}\right), \quad (1)$$

where L_p^{ss} is the persistence length of ssDNA, L_c^{ss} the contour length of ssDNA, k_B the Boltzmann constant, and T the temperature. We used $L_p^{ss} = 0.87$ nm/nt and $L_c^{ss} = 0.69$ nm/nt at a salt condition of 100 mM NaCl (41). The WLC model is only valid for sufficiently high forces to describe the force-extension relationship of ssDNA. At forces below ~10 pN, ssDNA collapses into a compact formation in ~100 mM salt concentration (42). Therefore, the WLC model's forces applied for ssDNA should be more than 10 pN. In the study, the unzipping forces are greater than 12 pN, so the force-extension relationship of ssDNA based on the WLC model is applicable.

Because the sampling rate is more than 200 Hz, we only measured pausing times longer than 10 ms. A pause is defined as a dwell time longer than 10 ms. For the CpG hairpin construct, a protein-binding event at a CpG site is collected when it is observed within ±11 bp around the CpG site. We measured the pausing time at CpG sites using a step-fitting algorithm (43). When building pausing-time histograms for fitting heavy-tailed distributions, we followed the binning rules of the Freedman-Diaconis algorithm to suppress outlier effects. We applied exponential functions to fit the probability distribution of pausing times. We first fitted the single-exponential function of $f(x) = y_0 \times \exp\left(\frac{-x}{\tau}\right)$ to a distribution. If goodness of fit was not satisfied, we applied a double- or multiple-exponential function $f(x) = y_1 \times \exp\left(\frac{-x}{\tau_1}\right) + y_2 \times \exp\left(\frac{-x}{\tau_2}\right) + \ldots$, which is a linear combination of the single-exponential functions and suggests a distribution arising from independent Poisson processes with different rates. Detailed fitting results with statistics are in Table S4.

To estimate the IC50 from a dose response curve, we performed nonlinear least-squares fitting using a four-parameter logistic model (4PL) (44). The 4PL equation is:

$$y = \frac{a-d}{1+\left(\frac{x}{c}\right)^b}+d, \quad (2)$$

where x and y are the drug concentration and response, respectively; a and d are the lower and upper plateaus, respectively; b describes the steepness of the linear portion of the dose response curve; and c is the IC50 regarding the response midway between a and d.

RESULTS AND DISCUSSION

Single-molecule profiling of MLL1 CXXC removal from CGI by hairpin unfolding

The structure of MLL1 CXXC domain-DNA complex has been resolved (7). The MLL1 CXXC domain binds DNA perpendicularly to the DNA axis and inserts into the DNA major groove in a wedge-like manner (7) (Fig. 1 *a*, *left*). An EMSA revealed a K_d of 5.7 ± 0.2 μM (mean ± SE, N = 7) for the MLL1 CXXC domain binding to DNA with one CpG site (Fig. 1 *b*; Table S1).

To examine MLL1 CXXC binding events on a sequence with dense CpGs, we explored its real-time dynamics on single CG-rich DNA hairpins. The single-molecule experimental configuration of our magnetic tweezers is described elsewhere (35,38,45). We embedded the first CpG island (CGI) from the mouse Hoxa9 gene promoter, which is 187 bp in length (GC% = 73.8%) and consists of 20 CpGs (6), in a hairpin stem of 236 bp (Fig. 1 *a*, *right*; Table S2). Because the critical force of unfolding the CGI hairpin is 20 ± 2 pN (mean ± SD, N = 180) based on force-ramp assays at a loading rate of 3 pN/s, we observed a sudden unfolding of the entire CGI hairpin without any pauses when forces jump from 8 to 25 pN (Fig. 1 *c*). Using a force manipulation protocol with F_{low} = 8 pN, F_{probe} = 25 pN, and F_{high} = 30 pN, we unfolded the CGI hairpin and examined the pauses caused by MLL1 CXXC binding events on a sequence with dense CpGs (Figs. 1 *d* and S1).

Interactions of MLL1 CXXC and dense CpG motifs on the CGI show a complex binding landscape. In the histogram in Fig. 1 *e*, bin height represents the averaged dwell time of MLL1 CXXC at a specific site before CGI hairpin unfolding (N = 2129). Well-defined valleys divide the CGI into CpG clusters, revealing the distinct binding preferences of MLL1 CXXC. In addition, we can directly quantify the site-specific binding probabilities of MLL1 CXXC along the CGI. CpG clusters are evident from the high peaks of binding probabilities (Fig. 1 *f*, *blue*). Furthermore, we can derive the energy landscape of MLL1 CXXC on the CGI by estimating the site-specific binding energy $\Delta G_i = -k_B T ln(P_{on}^i / P_{off}^i)$ at site i, where k_B is the Boltzmann constant, T the temperature, and P the probabilities of binding (on) and unbinding (off) (Fig. 1 *f*, *red*). We can rewrite the above equation as $\Delta G_i = -k_B T ln(c/K_d^i)$. The binding energy difference between two sites i and j, $\Delta\Delta G = \Delta G_i - \Delta G_j = -k_B T ln(K_d^j / K_d^i)$, is thus independent of the protein concentration c (46). Therefore, the site-specific binding energy ΔG_i reflects the relative binding affinities of the MLL1 CXXC protein along the CGI from an energy

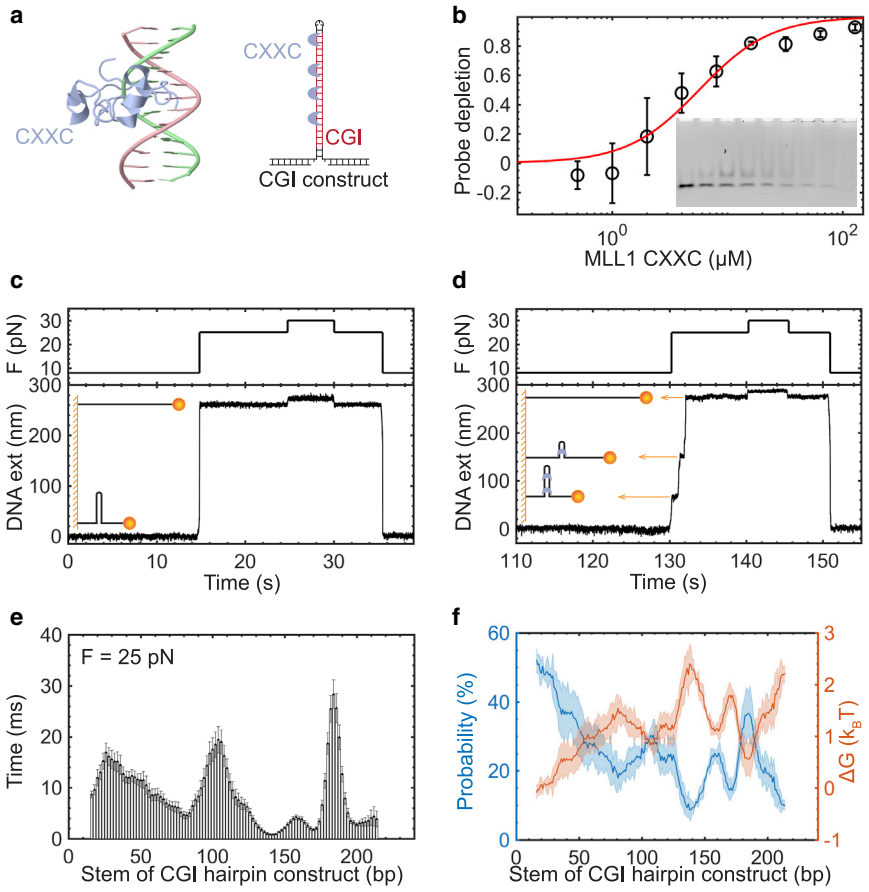

FIGURE 1 Single-molecule profiling of protein-DNA complexes on a CpG island using magnetic tweezers. (*a*) The MLL1 CXXC domain-DNA complex (*left*, Protein Data Bank: 2KKF) and a schematic of the CGI hairpin construct (*right*). The 236 bp hairpin stem contains the 187 bp CGI sequence (*red*) to which CXXC (*gray*) can bind. (*b*) EMSA of MLL1 CXXC binding a CpG DNA. Inset shows a gel of EMSA. The dsDNA concentration was 100 nM, and the MLL1 CXXC concentrations were 0.5, 1, 2, 4, 8, 16, 32, 64, and 128 μM. Data are shown as mean ± SE (N = 7). Red curve represents the fit to the Hill equation, resulting in K_d = 5.7 ± 0.2 μM (mean ± SE). (*c*) Single-molecule assays of CGI hairpin unfolding. The top panel shows a force protocol with F_{low} = 8 pN, F_{probe} = 25 pN, and F_{high} = 30 pN. The bottom panel shows corresponding DNA extensions. The cartoons illustrate a folded and an unfolded hairpin. The sampling rate was 400 Hz. (*d*) Single-molecule CGI hairpin unfolding assay in the presence of 1 μM MLL1 CXXC protein. Cartoons illustrate partial unzipping of a hairpin bound by proteins. The force protocol is the same as in (*c*). (*e*) Histogram of 2129 CGI hairpin unfolding traces from 215 molecules in the presence of MLL1 CXXC. The bin size is 2 bp. Binned counts are converted to time by dividing by the sampling rate (mean ± SE). (*f*) Averaged binding probabilities and ΔG of MLL1 CXXC on the CGI hairpin (mean ± SD, N = 3). The bin size equals 1 bp. To see this figure in color, go online.

perspective, providing an intrinsic parameter to understand epigenetic enzymes' preference for DNA motifs.

To better understand how CGI hairpin unfolding drives MLL1 CXXC unbinding, we quantified the site-specific pausing positions and times. We used a step-fitting algorithm to analyze individual pauses (43) (Fig. 2 *a*), where a pause is defined to be longer than 5 ms (half the Nyquist limit). The site-specific pausing positions reveal five CpG clusters, which is identical to the sequencing result (6) (Fig. 2 b). A minimum of three exponential functions was required to fit the pausing-time distribution of all the pauses on the CGI (Fig. 2 *c*; all fit coefficients in Table S4). To investigate MLL1 CXXC's dynamics, we further built histograms for each CpG cluster (Fig. 2 *d*). The pausing-time distributions of the five CpG clusters can be well fitted by a double-exponential function (Fig. 2 *d*; Table S4).

The NMR structure of MLL1 CXXC with a dsDNA of CCCTGCGCAGGG shows that the CXXC domain specifically recognizes the underlined CpG motif via a short 1182–1188 loop (7). At Lys1185 and Lys1186, two hydrogen bonds form among the lysines and cytosines. Moreover, two guanines form two hydrogen bonds with Lys1186 and Gln1187, respectively. In addition to base-specific contacts, the MLL1 CXXC domain can also interact with the backbone and minor groove of the DNA flanking the CpG site. Intermolecular interactions between MLL1 CXXC and DNA are mediated by Arg1150, Ser1152, Arg1154, Lys1176, Lys1178, Ile1184, Lys1185, Lys1190, Arg1192 and Lys1193, Leu1197 and Met1200. Thus, MLL1 CXXC can interact with DNA either specifically or nonspecifically.

The interactions between MLL1 CXXC and non-CpG DNA could happen when the protein searches a CpG motif via the nonspecific interactions mentioned above. We frequently observed MLL1 CXXC's binding events at non-CpG sites indicated by the high baseline (Fig. 2 *b*). Assuming the diffusion-limited binding rate to be 10^8–10^9/M/s (34,47), 1 μM MLL1 CXXC can visit a non-CpG site ~1–10 time in 10 ms. Nonspecific interactions between a neighbor CXXC protein and non-CpG DNA may help the CXXC-CpG complex resist the dsDNA strand separation, resulting in a ~10× longer dissociation time. The dissociation time distribution at cluster 5 is well fitted by a double-exponential function with the time coefficients of 10 and 70 ms, which agrees with the assumption above.

At clusters 1–4, a double-exponential function can also describe the distributions of dissociation time

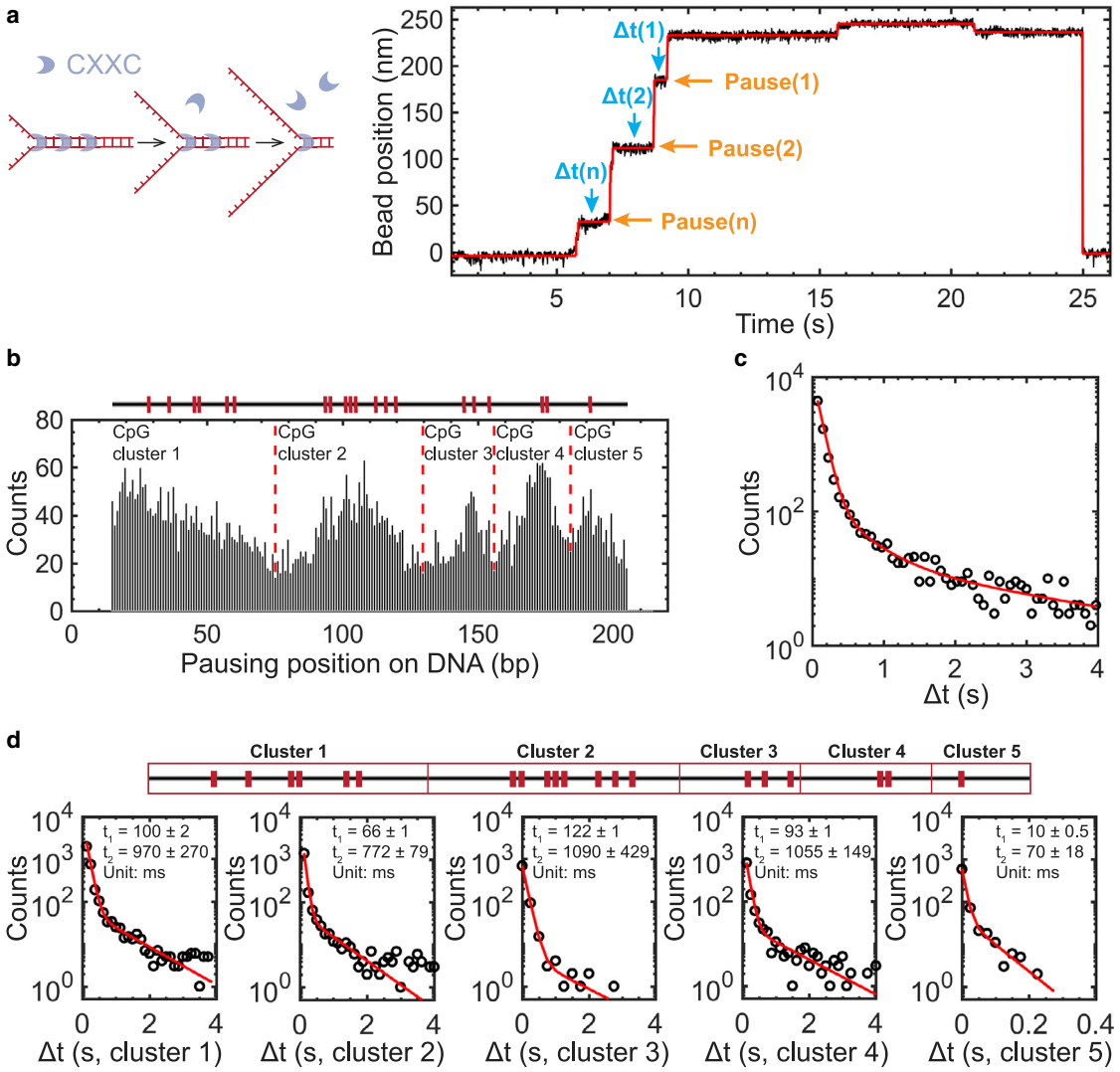

FIGURE 2 Pausing-time analysis of CGI unfolding in the presence of MLL1 CXXC. (*a*) Cartoon illustration of MLL1 CXXC being driven off the CGI by hairpin unwinding and pausing-time analysis of a trace using a step-fitting algorithm. The fitted trace is in red. (*b*) Pausing positions along the CGI revealed by step-fitting analysis. The cartoon on the top illustrates the CGI structure where the red marks indicate the CpGs. Red dashed lines indicate the boundaries of five CpG clusters. N = 8018 from 215 molecules. (*c*) The overall pausing-time distribution of the CGI hairpin. The red curve represents a triple-exponential fit. N = 8018 from 215 molecules. F_{probe} = 25 pN. (*d*) The pausing-time distributions for each CpG cluster. CpG patterns are noted above, with CpGs in red. The red curves represent exponential fits. Detailed statistics are in Table S4. F_{probe} = 25 pN. To see this figure in color, go online.

well. However, we did not observe the short time coefficient of 10 ms as that at cluster 5. The short time coefficients at clusters 1–4 are close to the long coefficient at cluster 5, ~100 ms, whereas the coefficients of a long time at clusters 1–4 are ~1000 ms, 100 times of 10 ms. Thus, we believe that the single-molecule observation of MLL1 CXXC-DNA interactions on the Hoxa9 CGI requires a three-state model to describe. In the simplest state 1, MLL1 CXXC specifically binds a CpG site and shows the shortest dissociation time of 10 ms, which is only observed at the most straightforward cluster 5 (Fig. 2 *d*; Table S4). In the scenario of state 2, nonspecific interactions between MLL1 CXXC and neighboring DNA around a CpG site help the CXXC-CpG complex to resist the dsDNA unfolding, resulting in a 10-times longer dissociation time, i.e., ~100 ms. State 2 happens to all five clusters (Fig. 2 *d*; Table S4). In the most complicated state 3, MLL1 CXXC-CpG complexes may have multivalent interactions mediated by non-CpG sequences bound by CXXC, which further extends the dissociation time to be 100 times longer, i.e., ~1000 ms (Fig. 2 *d*; Table S4). State 3 happens to clusters 1–4 but not cluster 5, which has just one CpG site and cannot form multivalent interactions. Our results showed that the CpG patterns and DNA sequence could contribute to the dynamics of

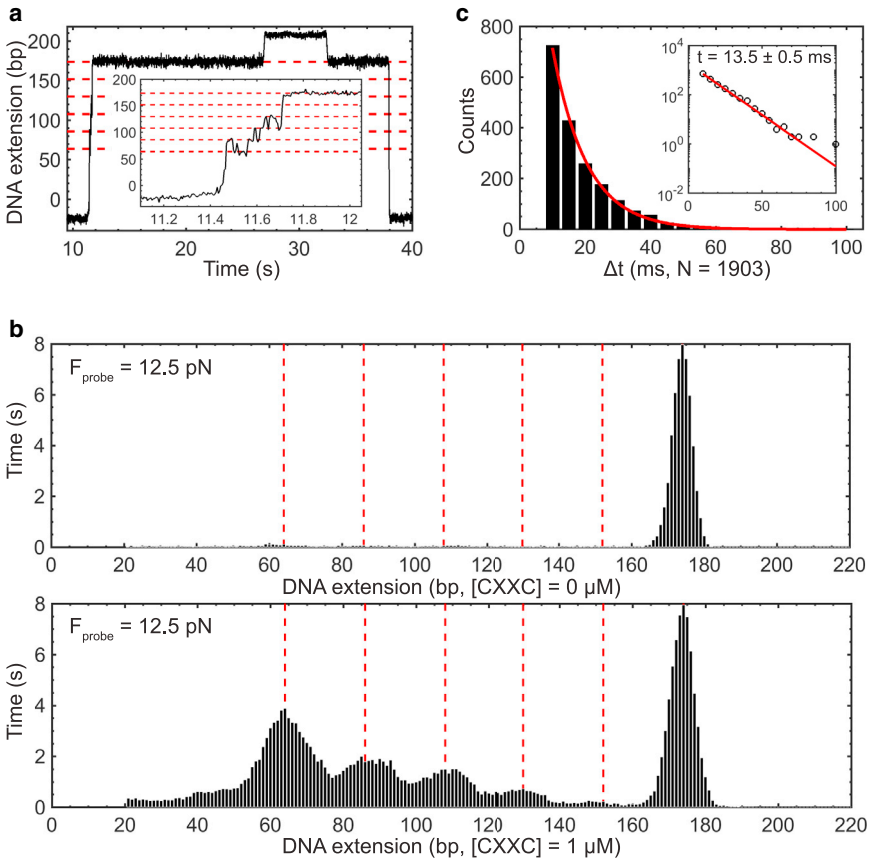

FIGURE 3 Single-molecule profiling of a CXXC-CpG-binding detector. (*a*) A single unfolding trace of a hairpin with well-isolated CpG sites at 1 μM MLL1 CXXC. Dashed lines indicate CpG positions. (*b*) Histograms of unfolding traces for the CpG hairpin without MLL1 CXXC (*top*, N = 100 from six molecules) and with MLL1 CXXC (*bottom*, N = 565 from eight molecules). Binned counts are converted to time by dividing by the sampling rate of 200 Hz. Red dashed lines indicate the CpG positions. The bin size is 1 bp. (*c*) Dwell time distribution along the CpG hairpin. The red curve shows a single-exponential fit. N = 2176 from 11 molecules. The inset is on a log scale. Detailed statistics are in Table S4. To see this figure in color, go online.

MLL1 CXXC on a CGI. Complicated CpG patterns and DNA sequence could cause multivalent protein-DNA complexes.

Label-free detector for single MLL1 CXXC-CpG-binding events

To understand the protein-DNA complex's fundamental interactions, it is necessary to examine the individual events involved in the removal of a single MLL1 CXXC protein from a CpG site. Two strategies can be applied to achieve single-molecule resolution: reduce the protein concentration or make CpG motifs well-separated on the DNA. Because single-molecule interactions rarely occur at a low protein concentration, we designed an AT-rich hairpin with five evenly distributed CpCpGpG sites in the stem. The CpCpGpG sites are separated by 18 bp spacers of AT-only sequences. The C/G-flanked CpG motif increases CXXC's binding affinity, whereas the AT-only background suppresses the nonspecific binding of CpG-binding proteins (48). For the CpG hairpin construct, the critical unfolding force is 8.5 $\pm$ 0.7 pN (mean $\pm$ SD, N = 15 from 15 molecules). When the CpG sites are occupied by CXXC, they produce features at distinct positions in the DNA unfolding landscape. Upon unfolding at F_{probe} = 12.5 pN, the five evenly distributed CpG sites produce a protein-binding fingerprint whose features can be unambiguously identified (Fig. 3 *a*).

Using a force protocol with F_{low} = 8 pN, F_{probe} = 12.5 pN, and F_{high} = 30 pN, we unfolded the CpG hairpin and collected traces for pausing analysis. In the absence of MLL1 CXXC, there are almost no pauses observed (Fig. 3 *b*, *top*). In the presence of 1 μM MLL1 CXXC, ruptures of the protein-DNA complexes produce distinct features in the spatial distribution of dwell times due to pausing (Fig. 3 *b*, *bottom*). The overall dwell time at each CpG site monotonically decreased along the hairpin unzipping direction. A possible mechanism could explain this interesting pattern. Because the CpG sites are only 18 bp apart from each other, the forward-backward fork motion is highly dynamic between free energy minimums at neighboring CpGs, biased toward the forward motion at the force of 12.5 pN. When a CpG site slows down the forward motion across itself in the presence of MLL1 CXXC, the frequency of the reverse motion to the preceding CpG sites increases, allowing rapid rebinding of CXXC from solution to the preceding sites. In 1 μM CXXC, the binding frequency could be in the order of 10^2–10^3/s, assuming the diffusion-limited binding rate to be 10^8–10^9/M/s (34,47). Thus, the overall time increases at the preceding CpG sites bound by CXXC, as observed by our experiments.

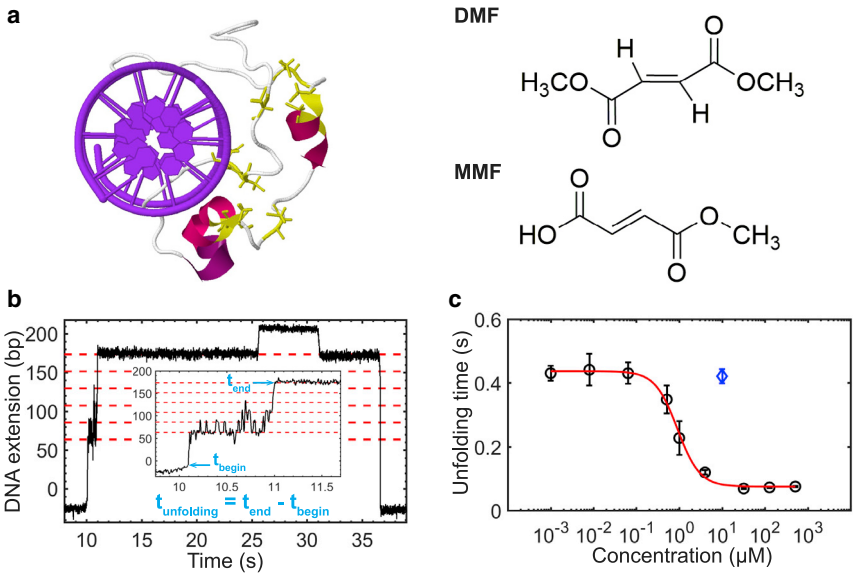

FIGURE 4 Unfolding time analysis using the single-molecule detector with drug-treated MLL1 CXXC. (*a*) Cysteines in MLL1 CXXC are targets of DMF. Protein Data Bank: 2KKF. Cysteines are highlighted with sticks. Structures of DMF and MMF are presented. (*b*) Unfolding time measurement. Beginning and ending times are noted. $F_{probe} = 12.5$ pN. (*c*) Dose response curve for CpG hairpin unfolding times with 1 μM MLL1 CXXC titrated by DMF (N > 20 from >3 molecules at each concentration). The DMF concentrations are 1, 8, 64, 512 nM, 1, 4, 32, 128, and 512 μM. The red curve is based on a 4PL for IC50 estimation (RMSE = 0.012, and R-squared = 0.997 for the goodness of fit). IC50 = 0.91 ± 0.09 μM (mean ± SE). The blue diamond shows the unfolding time in 10 μM MMF (N = 45 from three molecules). To see this figure in color, go online.

The dwell time distribution of MLL1 CXXC at the CpG sites follows a single-exponential decay function (Fig. 3 *c*). The fitted time coefficient at the CpG hairpin is 13.5 ± 0.5 ms, which is close to the short time coefficient of 10.0 ± 0.5 ms at cluster 5 of the CGI hairpin (Table S4). This suggests that the rupture of individual MLL1 CXXC-CpG complexes happens in ∼10 ms. The single-molecule detector can thus allow us to examine independent CXXC-CpG complexes. Because the detection principle employed here is the analysis of hairpin unfolding without fluorescence labeling, the measurement avoids introducing potential labeling artifacts to the interaction between a drug and its targets.

DMF damages MLL1 CXXC's binding function

To provide a proof of concept for evaluating a drug's effect on protein-DNA interactions using our label-free single-molecule detector, we tested the effect of DMF on the MLL1 CXXC-CpG-binding reaction. DMF is an electrophilic drug that targets proteins by covalently modifying more than 2400 cysteine residues (12) (Fig. 4 *a*). Although DMF can modify proteins with CXXC motifs, CXXC family proteins like MLL1 have not been identified as targets by chemical proteomic analysis.

We titrated the MLL1 CXXC-DNA reaction with DMF from 1 nM up to 512 μM in the presence of 1 μM CXXC (Fig. 4, *b and c*). At each concentration, the CXXC-DNA complexes were incubated in the presence of DMF for at least 15 min before performing the single-molecule force protocol to obtain unfolding times. We observed that DMF interrupts the protein-DNA interaction as shown by the hairpin unfolding time. After obtaining the unfolding times as a function of DMF concentration, we estimated the critical DMF dose using a 4PL to identify the relative IC50, or DMF dose that produces a response midway between the lower and upper plateaus (44) (Materials and methods). The resulting IC50 is 0.91 ± 0.09 μM (mean ± SE, n > 20 from >3 molecules at each concentration).

Because DMF is the prodrug of MMF (12), we also determined MMF's effect on the MLL1 CXXC-CpG reaction. We found that 10 μM MMF does not influence the unfolding time (Fig. 4 *c*, *blue diamond*). Orally taken DMF is rapidly metabolized to MMF, which serves as a ligand of the G-protein-coupled receptor hydroxycarboxylic acid receptor 2 (49). The null effect showed that MMF could not target the CXXC-CpG complex and does not cause side effects via the protein-DNA interactions.

DMF is used to treat autoimmune diseases such as multiple sclerosis and psoriasis. Our results reveal that DMF can target the CXXC family protein MLL1 CXXC. Because MLL1 is a transcription factor regulating the expression of more than 5000 genes via its histone methyltransferase activities (2–5), our results suggest that DMF may have an effect on processes that MLL1 is involved, such as leukemia pathogenesis. In particular, the ability of DMF to target complexes formed between a CXXC family protein and a CpG site has been overlooked by chemical proteomic studies, and its clinical effect is unexplored.

Single-molecule magnetic tweezers have been used in various drug studies, such as rapamycin-mediated protein-protein interactions between FKBP12 and FRB, formation kinetics of doxorubicin-DNA cross-link, and topoisomerase I targeted by antitumor drug topotecan (25,26,50,51). Our label-free single-molecule detector explored the concentration-dependent inhibition of DMF on the MLL1 CXXC-DNA interactions. In the future, we can further investigate the molecular mechanism by mutating the sequences of

protein or DNA. In addition, the development of an automatic buffer change method in magnetic tweezers will allow us to screen more drugs targeting the protein-DNA complexes in a high throughput fashion (52).

CONCLUSIONS

We used magnetic tweezers to examine the dynamics of MLL1's CXXC domain on a CGI of the Hoxa9 gene, revealing the landscape of binding probabilities and ΔG. Site-specific pausing-time analysis discloses the dissociation time of MLL1 CXXC along the CGI. In addition, we found that the rupture of individual MLL1 CXXC-CpG complexes happens in $\sim$10 ms at a probing force of 12.5 pN. In a proof of concept, we also demonstrated that DMF has a pronounced effect on MLL1 CXXC-DNA binding reactions, with an IC50 of $\sim$1 μM. We anticipate that our label-free single-molecule detector will benefit the development of CXXC-targeted drug discovery.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.03.045.

AUTHOR CONTRIBUTIONS

L.L., Z.W., and K.M. performed research. K.M. analyzed data and wrote the manuscript. R.J. wrote the manuscript. Z.Y. designed the research and wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by Fundamental Research Funds for Central Universities, Nankai University (035-63161138, 035-63191141, and 035-63191735 to Z.Y.). We also acknowledge support from the National Natural Science Foundation of China (Grants 31670763 and 32071227 to Z.Y.]. W. Tucker in the College of Foreign Languages at Nankai University corrected the grammar errors of the manuscript.

REFERENCES

1. Bernstein, B. E., T. S. Mikkelsen, …, E. S. Lander. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell.* 125:315–326.
2. Guenther, M. G., R. G. Jenner, …, R. A. Young. 2005. Global and Hox-specific roles for the MLL1 methyltransferase. *Proc. Natl. Acad. Sci. USA.* 102:8603–8608.
3. Muntean, A. G., and J. L. Hess. 2012. The pathogenesis of mixed-lineage leukemia. *Annu. Rev. Pathol.* 7:283–301.
4. Rao, R. C., and Y. Dou. 2015. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer.* 15:334–346.
5. Milne, T. A., J. Kim, …, C. D. Allis. 2010. Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. *Mol. Cell.* 38:853–863.
6. Erfurth, F. E., R. Popovic, …, N. J. Zeleznik-Le. 2008. MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. *Proc. Natl. Acad. Sci. USA.* 105:7517–7522.
7. Cierpicki, T., L. E. Risner, …, J. H. Bushweller. 2010. Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nat. Struct. Mol. Biol.* 17:62–68.
8. Lövkvist, C., I. B. Dodd, …, J. O. Haerter. 2016. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res.* 44:5123–5132.
9. Haerter, J. O., C. Lövkvist, …, K. Sneppen. 2014. Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.* 42:2235–2244.
10. Martín-Gago, P., and C. A. Olsen. 2019. Arylfluorosulfate-based electrophiles for covalent protein labeling: a new addition to the arsenal. *Angew. Chem. Int. Ed. Engl.* 58:957–966.
11. Scian, M., M. Guttman, …, W. M. Atkins. 2016. The myeloablative drug busulfan converts cysteine to dehydroalanine and lanthionine in redoxins. *Biochemistry.* 55:4720–4730.
12. Blewett, M. M., J. Xie, …, B. F. Cravatt. 2016. Chemical proteomic map of dimethyl fumarate-sensitive cysteines in primary human T cells. *Sci. Signal.* 9:rs10.
13. Nicolay, J. P., K. Müller-Decker, …, K. Gülow. 2016. Dimethyl fumarate restores apoptosis sensitivity and inhibits tumor growth and metastasis in CTCL by targeting NF-κB. *Blood.* 128:805–815.
14. Selman, M., P. Ou, …, J.-S. Diallo. 2018. Dimethyl fumarate potentiates oncolytic virotherapy through NF-κB inhibition. *Sci. Transl. Med.* 10:eaao1613.
15. Kornberg, M. D., P. Bhargava, …, S. H. Snyder. 2018. Dimethyl fumarate targets GAPDH and aerobic glycolysis to modulate immunity. *Science.* 360:449–453.
16. Carlström, K. E., E. Ewing, …, F. Piehl. 2019. Therapeutic efficacy of dimethyl fumarate in relapsing-remitting multiple sclerosis associates with ROS pathway in monocytes. *Nat. Commun.* 10:3081.
17. Holdgate, G. A., T. D. Meek, and R. L. Grimley. 2018. Mechanistic enzymology in drug discovery: a fresh perspective. *Nat. Rev. Drug Discov.* 17:115–132.
18. Beher, D., J. Wu, …, M. Wang. 2009. Resveratrol is not a direct activator of SIRT1 enzyme activity. *Chem. Biol. Drug Des.* 74:619–624.
19. Borra, M. T., B. C. Smith, and J. M. Denu. 2005. Mechanism of human SIRT1 activation by resveratrol. *J. Biol. Chem.* 280:17187–17195.
20. Kaeberlein, M., T. McDonagh, …, B. K. Kennedy. 2005. Substrate-specific activation of sirtuins by resveratrol. *J. Biol. Chem.* 280:17038–17045.
21. Pacholec, M., J. E. Bleasdale, …, K. Ahn. 2010. SRT1720, SRT2183, SRT1460, and resveratrol are not direct activators of SIRT1. *J. Biol. Chem.* 285:8340–8351.
22. Celaya, G., J. Perales-Calvo, …, D. Rodriguez-Larrea. 2017. Label-free, multiplexed, single-molecule analysis of protein-DNA complexes with nanopores. *ACS Nano.* 11:5815–5825.
23. Yamazaki, H., R. Hu, …, M. Wanunu. 2017. Label-free single-molecule thermoscopy using a laser-heated nanopore. *Nano Lett.* 17:7067–7074.
24. Bulushev, R. D., S. Marion, …, A. Radenovic. 2016. Single molecule localization and discrimination of DNA-protein complexes by controlled translocation through nanocapillaries. *Nano Lett.* 16:7882–7890.
25. Wang, Y., S. F. H. Barnett, …, J. Yan. 2019. Label-free single-molecule quantification of rapamycin-induced FKBP-FRB dimerization for direct control of cellular mechanotransduction. *Nano Lett.* 19:7514–7525.
26. Koster, D. A., K. Palle, …, N. H. Dekker. 2007. Antitumour drugs impede DNA uncoiling by topoisomerase I. *Nature.* 448:213–217.
27. Koch, S. J., and M. D. Wang. 2003. Dynamic force spectroscopy of protein-DNA interactions by unzipping DNA. *Phys. Rev. Lett.* 91:028103.
28. Halford, S. E., and J. F. Marko. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–3052.

29. Berghuis, B. A., D. Dulin, …, N. H. Dekker. 2015. Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. *Nat. Chem. Biol.* 11:579–585.

30. Manosas, M., J. Camunas-Soler, …, F. Ritort. 2017. Single molecule high-throughput footprinting of small and large DNA ligands. *Nat. Commun.* 8:304.

31. Gulvady, R., Y. Gao, …, J. Yan. 2018. A single molecule analysis of H-NS uncouples DNA binding affinity from DNA specificity. *Nucleic Acids Res.* 46:10216–10224.

32. Rudnizky, S., H. Khamis, …, A. Kaplan. 2018. Single-molecule DNA unzipping reveals asymmetric modulation of a transcription factor by its binding site sequence and context. *Nucleic Acids Res.* 46:1513–1524.

33. Liang, L., Z. Wang, …, Z. Yu. 2021. Single-molecule multiplexed profiling of protein-DNA complexes using magnetic tweezers. *J. Biol. Chem.* 296:100327.

34. Zhao, X., S. Guo, …, J. Yan. 2019. Single-molecule manipulation quantification of site-specific DNA binding. *Curr. Opin. Chem. Biol.* 53:106–117.

35. Yu, Z., D. Dulin, …, N. H. Dekker. 2014. A force calibration standard for magnetic tweezers. *Rev. Sci. Instrum.* 85:123114.

36. te Velthuis, A. J., J. W. Kerssemakers, …, N. H. Dekker. 2010. Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. *Biophys. J.* 99:1292–1302.

37. Lipfert, J., X. Hao, and N. H. Dekker. 2009. Quantitative modeling and optimization of magnetic tweezers. *Biophys. J.* 96:5040–5049.

38. Li, N., J. Wang, …, Z. Yu. 2019. The dynamics of forming a triplex in an artificial telomere inferred by DNA mechanics. *Nucleic Acids Res.* 47:e86.

39. Zhao, X., S. Peter, …, J. Yan. 2017. Oncofetal HMGA2 effectively curbs unconstrained (+) and (-) DNA supercoiling. *Sci. Rep.* 7:8440.

40. Marko, J. F., and E. D. Siggia. 1995. Stretching DNA. *Macromolecules.* 28:8759–8770.

41. Bosco, A., J. Camunas-Soler, and F. Ritort. 2014. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.* 42:2064–2074.

42. Fu, H., S. Le, …, J. Yan. 2013. Force and ATP hydrolysis dependent regulation of RecA nucleoprotein filament by single-stranded DNA binding protein. *Nucleic Acids Res.* 41:924–932.

43. Kerssemakers, J. W., E. L. Munteanu, …, M. Dogterom. 2006. Assembly dynamics of microtubules at molecular resolution. *Nature.* 442:709–712.

44. Sebaugh, J. L. 2011. Guidelines for accurate EC50/IC50 estimation. *Pharm. Stat.* 10:128–134.

45. Ma, X., M. Zhu, …, Z. Yu. 2019. Interactions between PHD3-bromo of MLL1 and H3K4me3 revealed by single-molecule magnetic tweezers in a parallel DNA circuit. *Bioconjug. Chem.* 30:2998–3006.

46. Wang, Y., J. Yan, and B. T. Goult. 2019. Force-dependent binding constants. *Biochemistry.* 58:4696–4709.

47. Model, M. A., and G. M. Omann. 1995. Ligand-receptor interaction rates in the presence of convective mass transport. *Biophys. J.* 69:1712–1720.

48. Lee, J. H., K. S. Voo, and D. G. Skalnik. 2001. Identification and characterization of the DNA binding domain of CpG-binding protein. *J. Biol. Chem.* 276:44669–44676.

49. Chen, H., J. C. Assmann, …, M. Schwaninger. 2014. Hydroxycarboxylic acid receptor 2 mediates dimethyl fumarate's protective effect in EAE. *J. Clin. Invest.* 124:2188–2192.

50. Kostrz, D., H. K. Wayment-Steele, …, C. Gosse. 2019. A modular DNA scaffold to study protein-protein interactions at single-molecule resolution. *Nat. Nanotechnol.* 14:988–993.

51. Pei, Y., Y. Liu, …, H. You. 2020. Detecting the formation kinetics of doxorubicin-DNA interstrand cross-link at the single-molecule level and clinically relevant concentrations of doxorubicin. *Anal. Chem.* 92:4504–4511.

52. Le, S., M. Yao, …, J. Yan. 2015. Disturbance-free rapid solution exchange for magnetic tweezers single-molecule studies. *Nucleic Acids Res.* 43:e113.

Biophysical Journal, Volume *120*

Supplemental information

Dynamics and inhibition of MLL1 CXXC domain on DNA revealed by single-molecule quantification

Lin Liang, Kangkang Ma, Zeyu Wang, Richard Janissen, and Zhongbo Yu

Supplemental information

Dynamics and Inhibition of MLL1 CXXC Domain on DNA Revealed by Single-Molecule Quantification

Running title: Dynamics of a CXXC protein on a CGI

Lin Liang[1†], Kangkang Ma[1†], Zeyu Wang[1], Richard Janissen[2], Zhongbo Yu[1]*

1 State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy, Nankai University, 38 Tongyan Road, Tianjin 300350, China

2 Department of Bionanoscience, Kavli Institute of Nanoscience, Delft, South-Holland 2629HZ, the Netherlands

* Corresponding author: Zhongbo Yu

Email: zyu@nankai.edu.cn

This Supplementary Information includes:

Figure S1. Unfolding and refolding of a CGI hairpin using single-molecule magnetic tweezers

Figure S2. Unfolding times of drug-treated MLL1 CXXC-DNA complexes.

Table S1. Sequences of MLL1 CXXC

Table S2. Oligos for constructing DNA CGI hairpins

Table S3. Buffers

Table S4. Fitting results of time distributions

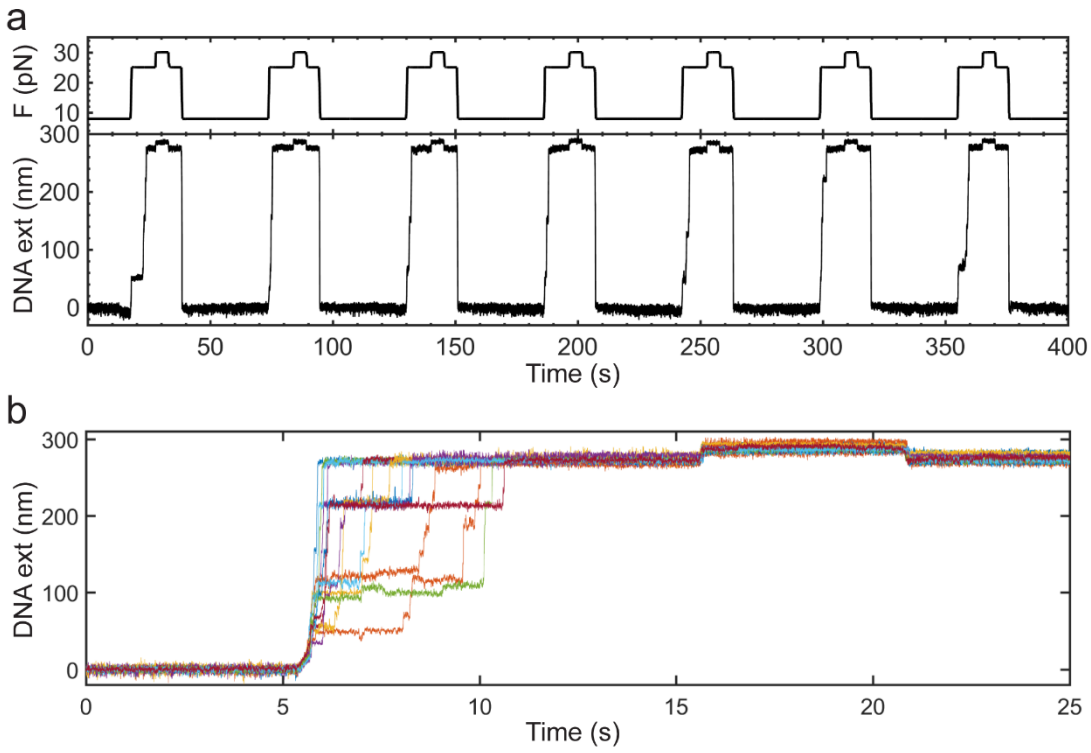

Figure S1. Unfolding and refolding of a CGI hairpin using single-molecule magnetic tweezers

(a). Repetitive assays to probe the binding of MLL1 CXXC to a CGI hairpin.

(b). Multiple overlapped force-jump traces (N = 18).

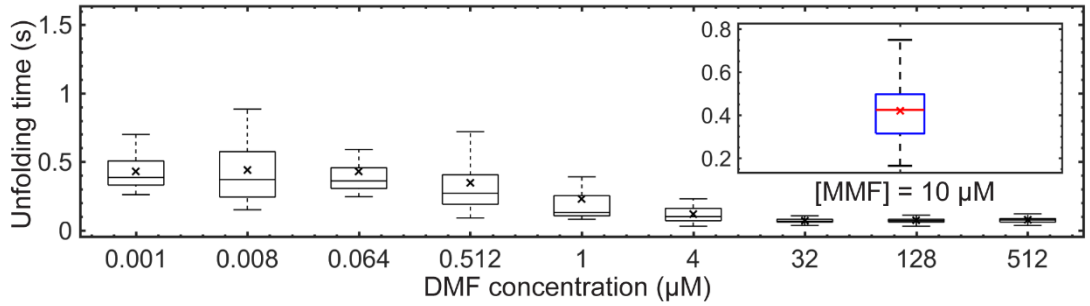

Figure S2. Unfolding times of drug-treated MLL1 CXXC-DNA complexes.

For DMF-treated reactions, unfolding times of MLL1 CXXC from the CpG hairpin (N > 20 traces from > 3 molecules for each concentration, F_{probe} = 12.5 pN) are shown. Cross: mean. Bar in a box: median. Box edges: the 25th and 75th percentiles. Whiskers: the 9th and 91st percentiles. Inset shows the data at [MMF] = 10 μM (N = 45 from 3 molecules).

Table S1. Sequence of MLL1 CXXC.

Name	Sequences of amino acids	Length
MLL1 CXXC	KKGRRSRRCGQCPGCQVPEDCGVCTNCLDKPKFG GRNIKKQCCKMRKCQNLQWMPSK	57

Notes: The lyophilized peptide is soluble in water. We prepared a stock solution with 1 mM of the peptide in a buffer containing 10 mM Tris (pH 7.4), 100 mM NaCl, 50 μM $ZnCl_2$, 1 mM DTT, and 30% glycerol. The stock solution can be further diluted with an assay buffer.

Table S2. Oligos for constructing DNA hairpins.

Name	**Sequence (5′-3′)**
Junction 1 for the CpG hairpin	GAGATGATTTGAAAAAATATGAAGAATGGTATAA TAAAAGGGTGATTTATATTTATTTATTCCGGTATTT AATTTAATTATATCCG
Junction 2 for the CpG hairpin	AATATATAACCGGATATAATTAAATTAAATACCGG AATAAATAAATATAAATCTGGGAGTAGATGTGGT TTTTGTTTGTTTG
Junction 3 for the CpG hairpin	CCCTTTTATTATACCATTCTTCATATTTTTTC
Junction 4 for the CpG hairpin	ACTCATCATTCAAACAAACAAAAACCACATCTA CTCCC
Stem 1 for the CpG hairpin	GTTATATATTTATATTTATCCGGTTTATTTATTATTT ATTTCCGGTTATTTATAATTTAATTAC
Stem 1c for the CpG hairpin	ATAATAACCGGTAATTAAATTATAAATAACCGGAA ATAAATAATAAATAAACCGGATAAATATA
Stem 2 for the CpG hairpin	CGGTTATTATATATTATTTATCCGGTATTTATTTAAT TATATTCCGGTTATTTATATATTTATATCC
Stem 2c	ATATAAATAACCGGAATATAATTAAATAAATACCG

for the CpG hairpin	GATAAATAATAT
Loop	GGTTAATATTTATTATATTTTTTTAAATATAATAAAT
for the CpG hairpin	ATTAACCGGATATAAAT
Linker 1	TCAGCAAGGAAGGAGATTTTGAAAAATTTATTTA
for the CpG hairpin	TTAGATATTGGAAATATTATTAGAG
Linker 1c	AAATCATCTCCTCTAATAATATTTCCAATATCTAAT
for the CpG hairpin	AAATAAATTTTTCAAAATCTCCTTCCTTGC
Linker 2	AATGATGAGTGTTAAAAAAAGTGGGGAAGTGAG
for the CpG hairpin	TAATGAAATTATTTTGTATGTTTTTTATATGAATTT ATTTTTTGGG
Linker 2c	GACCCCAAAAAATAAATTCATATAAAAAACATAC
for the CpG hairpin	AAAATAATTTCATTACTCACTTCCCCACTTTTTTT AAC
Junction 1	TCAGCAAGGAAGGAGATTTTGAAAAATTTATTTA
for the CGI hairpin	TTAGATATTGGAAATATTATTAGAGGAGATGATTT AAAAAAATATGAAGAATGGTATAATAAAAGGGTT TGGTTATTAGAGGACACCTA
Junction 2	GTGTCCTCTAATAACCTTTGGGAGTAGATGTGGT
for the CGI hairpin	TTTTGTTTTTTTGAATAATAAATGTTAAAAAAAGT GGGGAAGTGAGTAATGAAATTATTTTGTATGTTT TTTATATGAATTTATTTTTTGGG
Junction 3	CCCTTTTATTATACCATTCTTCATATTTTTTTAAAT
for the CGI hairpin	CATCTCCTCTAATAATATTTCCAATATCTAATAAAT AAATTTTTCAAAATCTCCTTCCTTGC
Junction 4	GACCCCAAAAAATAAATTCATATAAAAAACATAC
for the CGI hairpin	AAAATAATTTCATTACTCACTTCCCCACTTTTTTT AACATTTATTATTCAAAAAAACAAAAACCACATC TACTCCC
Loop	GTGAAGAGATTTGTAAAGTTTTCTTTACAAATCT

for the CGI hairpin	CTTCACACA
Sequence of the CGI (CpG in bold and clusters underlined)	ACCTGG**CG**GTCCTC**CG**CTAGGCCA**CGCG**TTTCC TGCT**CG**C**CG**GAGGGGGGGGGGGAACACTAGGT GGGGGAAGGGT**CGCG**GGAG**CGCGCG**CCTCAG **CG**GG**CG**GG**CG**CCTAGGAGGGAGAAGAGGGGG AGAG**CG**AG**CG**GCTG**CG**GGGAGTGAGTAGAAGA GGC**CGCG**GCCAGCCACAGGACC**CG**GCTC
Forward primer for the CGI sequence	ATATTTCCACCTACTGGACCTGGC
Reverse primer for the CGI sequence	ATATTTCCAGACAGTGGAGCCGG
Forward primer for the biotin handle	GACCGAGATAGGGTTGAGTG
Reverse primer for the biotin handle	GCATCGGCTGAGGAAAGGGAACAAAAGCTGG
Forward primer for the digoxigenin handle	ATCGTAGGGTCCTGACCGAGATAGGGTTGAGTG
Reverse primer for the digoxigenin handle	AAAGGGAACAAAAGCTGG

Table S3. Buffers.

Name	**Contents**
Titration buffer	10 mM Tris HCl (pH 7.4), 150 mM NaCl, 50 µM $ZnCl_2$, 1 mM DTT, and 6% glycerol
MT buffer	10 mM of Tris HCl (pH 7.4), 100 mM of NaCl, 50 µM ZnCl2, 0.003% Tween-20, and 5 mM of DTT

Table S4. Fitting results of time distributions.

Data source	t (ms, mean ± se)	RMSE	R-Squared	Bin size (ms)	N	Molecules
Figure 2c	t_1 = 90 ± 12 t_2 = 410 ± 45 t_3 = 2380 ± 520	88	0.98	75	8018	215
Figure 2d Cluster 1	t_1 = 100 ± 2 t_2 = 970 ± 270	10	0.99	125	3373	---
Figure 2d Cluster 2	t_1 = 66 ± 1 t_2 = 772 ± 79	2.8	1.00	125	1883	---
Figure 2d Cluster 3	t_1 = 122 ± 1 t_2 = 1090 ± 429	0.84	1.00	250	819	---
Figure 2d Cluster 4	t_1 = 93 ± 1 t_2 = 1055 ± 149	2.5	1.00	125	1194	---
Figure 2d Cluster 5	t_1 = 10.0 ± 0.5 t_2 = 70 ± 18	2.8	1.00	25	724	---
Figure 3c	t = 13.5 ± 0.5	1.5	0.98	5	2176	11