GAGA Factor Overcomes 1D Diffusion Barrier by 3D Diffusion in Search of Nucleosomal Targets

Xinyu A. Feng[1,2], Kaitlin M. Ness[1], Celina Liu[1], Ishtiyaq Ahmed[2,†], Gregory D. Bowman[2], Taekjip Ha[2,3,4,5,*], Carl Wu[1,6,*]

[1] Department of Biology, Johns Hopkins University, Baltimore, Maryland, USA.
[2] Department of Biophysics, Johns Hopkins University, Baltimore, Maryland, USA.
[3] Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.
[4] Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.
[5] Howard Hughes Medical Institute, Baltimore, Maryland, USA.
[6] Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.
[†] Project affiliation.
[*] Corresponding authors. E-mails: wuc@jhu.edu (W.C.); tjha@jhu.edu (T.H.)

Abstract

The eukaryotic chromatin landscape plays important roles in DNA metabolism and is characterized by positioned nucleosomes near regulatory DNA, nucleosome-depleted regions and supranucleosomal organization. Nucleosome core histones limit DNA accessibility by structurally blocking half of the DNA surface and altering its topology, but how nucleosomes affect target search by sequence-specific transcription factors (TFs) remains enigmatic. Here, we used multi-color smFRET to investigate how *Drosophila* GAGA Factor (GAF) locates its targets. On free DNA, GAF rapidly diffuses in 1D to a single cognate motif but escapes after subsecond transient association. Nucleosomes effectively block 1D diffusion into its core, but GAF can bind, with surprisingly prolonged residence, at internal cognate sites by direct association from 3D. Our findings demonstrate the occlusive power of nucleosomes to 1D sliding and reveal that a combination of 1D and 3D diffusion by a zinc finger TF enables efficient target search on chromatin.

Introduction

In eukaryotic organisms, DNA is packaged into chromatin to regulate accessibility and recruitment of DNA-binding proteins involved in DNA replication, transcription, and genome maintenance. Sequence-specific transcription factors, hereafter referred to as TFs, access specific DNA motifs (cognate motifs) on the genome to regulate transcription of target genes. Bacterial transcription factors such as LacI searches for its target on the relatively naked bacterial chromosome using a combination of 1-dimensional (1D) and 3-dimensional (3D) diffusion, referred to as facilitated diffusion[1–3]. How transcription factors search for their targets on the eukaryotic chromatin is less understood due to the packing of DNA into nucleosome particles, higher order chromatin structure on the kbp and Mbp scale, and the scarcity of experimental investigations[4].

Chromatin can limit the accessibility of TFs at specific genomic regions by positioning nucleosomes over the cognate DNA motif[5], thus regulating transcription by structural occlusion However, prior *in vitro* studies showed that different TFs may prefer to bind different DNA locations within a nucleosome[6–8], and some TFs can invade nucleosomes by unwrapping DNA at the

nucleosome edge or disrupting higher order chromatin organization[9–11]. Nonetheless, the mechanisms by which eukaryotic TFs perform the search for their targets on chromatin, bind selectively to certain nucleosomal sites over others, and exhibit stable association is poorly understood. Here, we address these questions for Drosophila GAGA Factor (GAF) – a single-zinc-finger TF protein that has been studied extensively by genetic, biochemical and single-molecule imaging approaches. Multi-functional GAF facilitates not only in 'pioneering' the chromatin accessibility of cognate promoter and enhancer sites for the benefit of neighboring TFs by nucleosome binding and recruitment of chromatin remodeling enzymes, but also promoter-promoter and promoter-enhancer looping, assembly of the transcription preinitiation complex and establishment of the paused RNA polymerase II machinery[12–22]. Cys_2-His_2 Zinc finger (ZF) TFs constitute the largest family of eukaryotic TFs, with ~500 out of 1600 human TFs containing one or multiple ZF DNA binding domain(s) (DBD)[23]. The NMR structure of GAF-DBD shows the ZF surrounded by short basic regions[24]. Despite its single ZF, full-length GAF forms a range of multimeric (on average hexameric) complexes through the N-terminal POZ domain[25] and preferentially binds to clusters of 'GAGAG' cognate motifs[26]. [2721]

Here, we used multi-color single-molecule fluorescence resonance energy transfer (smFRET)[28] to study the target search process of GAF-DBD. Using FRET as a readout for cognate-motif-specific binding, we found that the DNA sequence specificity of GAF-DBD is kinetically defined by the rate of association as well as dissociation. During target search, GAF-DBD undergoes 1D diffusion on free DNA with transient entrapment at cognate motifs. While still remaining on linear DNA, GAF-DBD escapes from transient entrapment to locate a neighboring cognate site. Sites at the nucleosome edge but not deeper sites in the nucleosome are found by 1D invasion from free linker DNA. Inner nucleosomal motifs require direct association of GAF-DBD from solution by 3D diffusion and form a more stable complex than on free DNA. Together, our results provide the first mechanistic insights on how a eukaryotic TF combines 1D and 3D diffusion to search for and locate its cognate site on free DNA and nucleosomes.

Results

Visualizing Sequence-Specific DNA-Binding Using smFRET

To investigate how GAF searches for its cognate motif on DNA, we placed a FRET donor (Cy3) at the N-terminus of GAF-DBD (by one-pot N-terminal cysteine labeling[29]; Extended Data Fig. 1) and a FRET acceptor (Cy5) on the DNA, 5 bp from the 'GAGAGAG' motif (Cy5 Cognate DNA). The 90 bp DNA was biotinylated on one end and immobilized for smFRET imaging (Fig. 1a, left). When Cy3-GAF-DBD is specifically bound to the motif, the donor-acceptor proximity results in acceptor emission via FRET (Fig. 1a, middle). In the control cognate motif deletion, GAF-DBD is nonspecifically bound (Cy5 Non-Cognate DNA; Fig. 1a, right). The greater average distance between the fluorophores does not allow detectable FRET; therefore, we expect primarily donor emission. Under this experimental design, Cy3-fluorescence with FRET signifies motif-specific binding, and Cy3-fluorescence without FRET signifies nonspecific binding.

GAF-DBD Efficiently Locates Cognate Motif on Linear DNA

When the surface-immobilized DNA was incubated with Cy3-GAF-DBD (0.4 nM), we observed single GAF-DBD binding events on DNA as an abrupt appearance of fluorescence signal above background (Fig. 1b). Motif-specific binding events showed strong acceptor emission during binding because of FRET (Fig. 1b, Cy5 intensity). For example, in the single-molecule fluorescence time trajectory in Fig. 1b (third trajectory), a GAF-DBD molecule associates specifically with the cognate motif at ~ 9 s, causing acceptor emission, and dissociates at 14 s;

another GAF-DBD is also specifically bound at ~ 31 s but quickly dissociates; two more GAF-DBD molecules are nonspecifically bound at ~ 48 s and 78 s, showing donor emission only. Out of all DNA binding events, GAF-DBD successfully locates the motif (FRET) in 68% ± 5% of binding events. This success rate is much higher than 2% (motif length divided by entire length of DNA) – the expected probability if GAF-DBD target search solely relies on random 3D collisions until successful collision with its target motif. Thus, we hypothesized GAF-DBD undergoes one-dimensional (1D) sliding on DNA to search for its target motif (hereafter, 1D sliding is used interchangeably with 1D diffusion and includes both helically-coupled sliding as well as 1D hopping). Under this hypothesis, the time delay between the increase in donor and increase in acceptor fluorescence would define the period between GAF-DBD landing anywhere on the DNA and successful recognition of the target motif. In most binding events, the time delay is shorter than the resolution of our instrument (35 ms), but we did occasionally detect such a delay (Fig. 1c). In addition, we observed transient donor-only fluorescence within the duration of some binding events, which suggests GAF-DBD transiently slides off before quickly before returning to motif (Fig. 1d, asterisks). The donor-only dwell time is inversely correlated with ionic strength, suggesting that 1D diffusion involves some hopping along the DNA[30,31] (Extended Data Fig. 2). At lower salt concentrations, GAF-DBD is more frequently found at nonspecific sites (Extended Data Fig. 2c,d), consistent with documentation that sequence specificity can be lost in low-salt buffers[32].

DNA Binding Kinetics Defines GAF-DBD Sequence Specificity

We observed only the donor emission on a mutant DNA sequence replacing the GAF motif (Fig. 1e). Because of this complete absence of FRET signal, we could reliably use FRET as a reporter for motif-specific binding. Detectable binding attempts on non-cognate DNA (0.0018 s^{-1} or 1 binding event every 560 s) are 16-fold less frequent than on cognate DNA (0.028 s^{-1} or 1 binding event every 36 s; Fig. 1f). The GAF-DBD dwell time on non-cognate DNA ($1/k_{off}$ = 0.33 s) is 9-fold shorter than on cognate DNA (3.1 s; Fig. 1g), indicating that the cognate motif transiently traps GAF-DBD to provide sequence-specificity. These two-color FRET results demonstrate that sequence specificity of GAF-DBD is defined by both association (k_{on}) and dissociation (k_{off}) kinetics.

Three-Color FRET Assay Reveals GAF-DBD Sliding on Naked DNA

On the *Drosophila melanogaster* genome, GAF preferentially binds to closely spaced cognate motifs. To mimic a native GAF target, we used a 187-bp segment from the well-studied *Drosophila* hsp70 promoter sequence, retaining the two longest GAF motifs while several other short GA-elements were mutated (Fig. 2a). A Cy7 fluorophore, which is also a FRET acceptor for Cy3-GAF-DBD[33], was placed 3 bp away from Motif 1 'GAGAGGGAGAGA'; a Cy5 fluorophore was placed 5 bp away from Motif 2 'GAGAGAG'; the two motifs flank 57 bp of intervening DNA. If Cy3-GAF-DBD is bound to the Cy7-Motif 1, we expect Cy7 emission from Cy3-Cy7 FRET. Likewise, we expect Cy3-Cy5 FRET when GAF-DBD is bound to Motif 2. If the protein is nonspecifically bound, we expect Cy3 emission only (Fig. 2a).

In the single-molecule trajectory of Fig. 2b, expanded in Fig. 2d, a Cy3-GAF-DBD molecule lands on the DNA molecule at ~17 s, emitting Cy3 fluorescence. Almost immediately, we also observe Cy3-Cy7 FRET, suggesting that GAF-DBD locates Motif 1. Cy7 intensity drops and Cy5 intensity increases 0.25 s later, suggesting GAF-DBD slides away from Motif 1 to reach Motif 2. 0.25 s later, GAF-DBD slides back to re-engage with Motif 1 (Fig. 2b). In a typical DNA binding event, Cy7 fluorescence fluctuates between zero and positive intensities for the duration

of GAF-DBD binding; Cy5 also fluctuates in anti-correlation with Cy7 fluorescence except for nonspecific binding. We interpret this as GAF-DBD undergoing 1D diffusion between Motif 1 and Motif 2 on the DNA molecule. To further analyze the dwell time at each motif and at nonspecific sites, we assigned binding locations over time based on the fluorescence intensities of Cy3, Cy5 and Cy7 (colored tracks in Fig. 2b-d). The blue and red interspersed with green bars are observed for all stable binding events (Fig. 2c). On average, dwell times on Motif 1 (blue, τ_1 = 0.25 s) are longer than Motif 2 (red, τ_2 = 0.11 s) (Fig. 2e), possibly because Motif 1 (12 bp) has a longer cognate sequence than Motif 2 (7 bp).

Nonspecific binding by GAF-DBD dominates at low ionic strength, exhibiting an extended dwell time in the Cy3-only state (from 0.3 to 1.4 s with decrease of $MgCl_2$). (Fig. 2e, Extended Data Fig. 3a, b), consistent with results from single-motif DNA (Extended Data Fig. 2). When Motif 1 (Cy7) is replaced with a random sequence, Cy3-Cy7 FRET becomes transient or undetectable, as expected, while Cy3-Cy5 FRET remains stable (Extended Data Fig. 3c-e). This confirms that the alternating FRET between Motifs 1 and 2 is a consequence of 1D sliding on DNA rather than artifactual fluorophore binding or conformational changes of the DBD itself. Hence, we conclude that GAF-DBD slides back and forth in 1D during its search (Fig. 2f), and importantly, that this 1D diffusion has a functional outcome in target location.

On the native hsp70 promoter, Motif 1 and Motif 2 are located on different strands of the DNA. To mimic the native configuration, we swapped Motif 1 with the complementary sequence 'CTCTCCCTCTCT'. Thus, GAF-DBD must flip 180 degrees orthogonal to the DNA axis for motif recognition. Interestingly, GAF-DBD still slides between Motifs 1 and 2 (Extended Data Fig. 3f-h), with a similar overall dwell time on DNA (3.6 +/- 0.05 s compared to 3.34 +/- 0.08 s). These results indicate that GAF-DBD can flip on DNA during 1D diffusion to accommodate opposite motif orientations.

Nucleosome Blocks 1D Sliding into its Core Beyond SHL5

We next asked if GAF-DBD uses 1D diffusion to locate GAF motifs inside a nucleosome. We tested three 40-N-40 mono-nucleosome constructs, each containing a Cy5-labeled 'GAGAGA' motif on the linker DNA and a second Cy7-labeled 'GAGAGA' motif at distinct superhelical locations (SHL) 7, 5 or 3 inside a nucleosome positioned by the Widom 601 Sequence[34] (the major groove of these motifs faces the histone core) (Fig. 3a). Fig. 3b-d show representative single-molecule trajectories of GAF-DBD on these constructs. The binding event in Fig. 3b shows alternating Cy5 and Cy7 fluorescence, a signature of 1D sliding between two binding sites. This type of trace is common for the SHL7-motif nucleosome, which suggests Cy3-GAF-DBD slides between the Cy5-linker DNA motif and the Cy7-SHL7 motif (Fig. 3e). Among 485 binding events, 75% shows sliding between the SHL7-motif and linker DNA, 19% shows stable residence on the linker DNA motif (but short range off-motif sliding would be undetected), and 7% shows residence on the SHL7 motif only (Fig. 3h). In contrast, motifs placed at SHL5 or SHL3 are rarely visited: ~90% of binding events shows Cy3-Cy5 FRET only, suggesting binding only to the linker DNA motif but not at SHL5 or SHL3. (Fig. 3c, d, f, g, h) In conclusion, nucleosome organization allows 1D sliding of GAF-DBD to SHL7 for motif recognition, but effectively blocks sliding to SHL5 or deeper locations in the nucleosome core.

Linker DNA is the Preferred Landing Site

Among binding events leading to sliding on the SHL7-motif nucleosome, GAF-DBD initially lands on the linker motif 62% of the time (Fig. 3i, 'Linker to Nuc') compared to 38% for the nucleosomal SHL7 motif. This indicates that the linker motif has greater accessibility (1.6x) than

at SHL7. Less common are Cy3-Cy5 FRET only (19%, 'Linker motif Only') throughout the trace (e.g., Fig 2e, row 2, constant red), but still more frequent than Cy3-Cy7 FRET only (7%, 'Nuc motif Only'; constant blue), a further indication that the linker motif is more accessible (Fig.3i). Occasionally, a Cy3-only state appears before FRET with either Cy5 or Cy7 ('Nonspecific to Linker' or 'Nonspecific to Nuc'), consistent with the aforementioned, short interval between initial binding to a nonspecific site and motif engagement. We also observed that, on average, GAF-DBD has a longer dwell time on the linker motif (0.25 s) than on the SHL7-motif (0.1 s), indicating that motif location at the nucleosome edge reduces the dwell time of GAF-DBD. (Fig. 3j). Hence, nucleosome organization affects both the on-rate and off-rate for binding to the SHL7 motif.

3D Diffusion Dominates 1D Sliding in Nucleosomal Target Search

The highly positioned 601 nucleosome almost completely blocks 1D and 3D access to the GAF cognate motif at SHL3 and SHL5, possibly because stable histone-DNA interactions of the 601 sequence inhibits local conformational flexibility to accommodate GAF-DBD binding. On native promoters such as *Drosophila hsp70*, multiple GAGA elements over multiple phases along the DNA helical axis potentially increases 3D accessibility[13,35,36]. Accordingly, we designed and constructed by linker ligation[37], a 0-N-40 *hsp70* promoter nucleosome (nucleotides -209 to -23, positioned to contain two native GAGA elements (the hsp70 sequence is identical to that shown in Extended data Fig. 3f). As shown in Fig. 4A, one element 'CTCTCCCTCTCT' near the nucleosome dyad was labeled with Cy7, while the other, 'GAGAGAG,' at the nucleosome edge was Cy5-labeled; we used this 'natural' double-labeled nucleosome to explore the potential for target search by 3D diffusion.

Based on the inaccessibility of GAGA motifs embedded deep in the positioned 601 nucleosome (Fig. 3), we hypothesized that GAF-DBD should locate the Cy7-Dyad Motif predominantly through three-dimensional (3D) diffusion. The Cy5-edge motif is on SHL7 of the hsp70 nucleosome, so GAF-DBD may locate the motif by either 1D sliding (Fig. 3b, e, h) or 3D diffusion. Fig. 4b shows idealized traces of a GAF-DBD searching for the Cy7-dyad motif by 3D diffusion: the first binding event directly targets the Cy7-dyad motif via 3D diffusion, exhibiting only Cy3-Cy7 FRET throughout the binding event; the second event binds to the Cy5-edge motif via 1D or 3D diffusion, showing only Cy3-Cy5 FRET. Note that if GAF-DBD is unable to slide between the Cy5-edge and Cy7-dyad motifs, there should be no alternating Cy5 and Cy7 FRET within one GAF-DBD binding event (Extended data Fig. 3g, 3h).

We observed four types of binding events for GAF-DBD (N=160): A) Binding to the edge motif (48%, Fig. 4c); B) Binding to the dyad motif (distal to Cy7), with FRET to both acceptor fluorophores (36%, Fig. 4d); C) Binding to the dyad motif (proximal to Cy7), with FRET to Cy7 only (13%, Fig. 4e) ; D) Binding to residual non-reconstituted free DNA with alternating FRET to Cy5 or Cy7 (3%, Fig. 2b). Remarkably, 97% of binding events show no anticorrelated changes in Cy5 and Cy7 FRET emission as shown earlier for binding to free DNA, indicating that GAF-DBD remains bound to one motif without sliding to the other on the same nucleosome (Fig. 4f). Hence, we conclude that 3D diffusion is the major search mode for locating inner nucleosomal motifs.

Although we observed three nucleosome binding locations, one binding location is usually preferred for each individual nucleosome (Extended Data Fig. 4a). For events following a type A binding, 88% are also type A (type A-A; Extended Data Fig. 4b), likewise for type B-B (84%) and type C-C events (88%). This suggests that the reconstituted hsp70 nucleosome population has at least three stable, distinctly phased configurations, each presenting a different cognate motif favorable for accessibility by GAF-DBD from 3D diffusion. We occasionally observed transient Cy3-signal without FRET within the duration of nucleosomal motif-specific binding (Fig. 4d, 4e: black asterisks), suggesting that GAF-DBD may undergo ultra-short range (<10 bp) 1D diffusion

on the nucleosome. Measured the dwell times of GAF-DBD bound to the two cognate motifs at the nucleosome edge and dyad fit well to a single-component exponential decay and are strikingly 10-fold longer than on naked DNA (Fig. 4g). This indicates that nucleosome organization better traps GAF-DBD at a specific cognate motif on chromatin than on naked DNA and should help GAF prioritize nucleosomal over naked DNA sites during target search.

Discussion

3-color FRET is an increasingly popular approach for studying complex DNA processes, including LacI target search[3], chromatin remodeling[3,38,39],and DNA repair[40]. Inspired by the study on LacI's search for the Lac operator on bacterial DNA[3], we used multi-color FRET to explore whether a eukaryotic TF also relies on 1D diffusion for target search on free DNA and nucleosomes. Our work reveals that the DBD of transcription factor GAF searches for its cognate DNA and nucleosome targets by a combination of 1D and 3D diffusion. 1D diffusion over several hundred bp is the major search mode used to locate GA-rich motifs on nucleosome-free DNA and at the accessible edge (SHL7) of nucleosomes. But 1D diffusion cannot reach motifs at SHL5 or locations deeper into the nucleosome core, making this search mode less effective at locating inner nucleosomal sites. 3D diffusion, however, allows GAF-DBD to find motifs inside a nucleosome by direct association from solution (Fig. 5). Thus, GAF-DBD combines 1D sliding for accessible DNA and 3D diffusion for nucleosomes during target search, achieving specificity for cognate sites and selective retention at nucleosomal motifs.

We postulate that full-length GAF (FL-GAF) harboring multiple zinc finger DBDs may also combine 1D and 3D search modes to efficiently locate its target on chromatin (Extended Data Fig. 4). Unlike GAF-DBD which is monomeric, FL-GAF forms a series of multimeric complexes (on average, hexamers) via the POZ domain and preferentially binds DNA containing a cluster of cognate motifs distributed across ~100 bp[25,26]. *In vivo*, the stably bound population of FL-GAF resides on chromatin for 2 min[21] whereas monomeric GAF-DBD only dwells a few seconds on DNA. This is possibly because dissociation of the FL-GAF multimer requires multiple DBDs to simultaneously dissociate from all engaged motifs – a low probability event compared to a single DBD dissociating from one site. Interestingly, the *in vivo* residence time of the transiently bound FL-GAF population (3.7 s)[21] is similar to the *in vitro* dwell time of GAF-DBD. Thus, transient binding *in vivo* may be attributed to either nonspecific binding of a FL-GAF multimer[21], and/or, as this study suggests, specific binding between one DBD (in a monomeric or multimeric FL-GAF) and an isolated cognate motif.

Our results indicate that in nucleosome-free chromatin regions, GAF would locate the cognate motif via 1D sliding after initial landing on DNA by 3D diffusion. Nucleosomes act as barriers to sliding, in which GAF may slide up to the edge of a nucleosome (SHL7) but not further into its core. Therefore, motifs at inner nucleosome locations must be found by 3D diffusion. However, our placement of these motifs at SHL5 and SHL3 on the 601 nucleosome shows little or no GAF-DBD binding, suggesting that the rotational phase and accessibility is unfavorable at these positions, where the cognate motif faces core histones. Interestingly, binding to motifs on *Hsp70* promoter nucleosomes is more efficient (Figs. 3h, 4f). This could possibly be due to differences in the motif sequences or, more likely, to phasing of the two cognate motifs at *Hsp70* [7,41] where the DNA major groove of at least one cognate motif faces away from core histones towards the solvent. In addition, ultra-short range (bp-scale) DNA mobility of the native *Hsp70* sequence relative to the histone octamer may allow different rotational phases to be sampled for better motif recognition[42,43].

Pioneer TFs are broadly defined by their ability to recognize cognate motifs on nucleosomes and initiate remodeling of the local chromatin landscape. Prior evidence has

established that GAF has pioneering properties: 1) It is necessary for the biogenesis of open chromatin at inducible, developmental and house-keeping *Drosophila* gene promoters *in vivo* and *in vitro*[13,16]; 2) It recruits ATP-dependent chromatin remodelers to create DNaseI-hypersensitive sites on chromatin[15,44]; and 3) It exhibits an uncommonly long residence time and high target occupancy loci in live-cell single-molecule imaging studies[21]. Our findings that the target search mechanism of GAF depends on the chromatin environment, where free DNA motifs are mainly located by 1D diffusion and inner nucleosome motifs by 3D diffusion. Intuitively, 3D search is less efficient than 1D because of the need for stochastic collision between TF and cognate motif, whereas 1D search could start anywhere on a stretch of accessible DNA. Accordingly, we and others show that TFs have a lower k_{on} for nucleosome substrates[5,10]. In contrast, dissociation from nucleosomal DNA motifs is ~10 times slower (k_{off}) than from free DNA, because GAF-DBD diffusion is confined by histone-DNA contacts to only ~5 bp of accessible DNA, and potentially also by adjacent stabilizing histone-DNA contacts. Our observation of less frequent but more stable binding to nucleosomal cognate motifs is consistent with the dissociation rate compensation mechanism recently reported for yeast Reb1 and Cbf1[10].

In vitro, GAF uses 1D scanning to locate cognate sites on free or accessible DNA and 3D diffusion to bind to nucleosome targets. *In vivo*, DNA accessibility is dynamic[45,46], where a particular cognate DNA sequence can switch between nucleosome-bound and nucleosome-free states for transcription regulation. GAF is necessary for creating nucleosome-free regions at gene promoters during *Drosophila* development[16]. Our results suggest that mechanistically, GAF can initially locate an inaccessible region by 3D diffusion to bind the nucleosome core or by 1D invasion into the nucleosome edge (Extended Data Fig. 4a). Upon successful target recognition, GAF may then recruit ATP-dependent chromatin remodelers (NURF, PBAP)[47–49] to remove nucleosomes from the cognate motif (Extended Data Fig. 4b,c), for the benefit of neighboring TFs lacking such ability. Once the cognate site becomes nucleosome-free, GAF can re-associate with the site by 1D diffusion and remain bound for minutes[21]. A dissociated GAF is rapidly replaced by another GAF multimer to achieve high occupancy over any given period (Extended Data Fig. 4d). This sequence of events would establish an accessible promoter and maintain its accessibility by essentially constant GAF occupancy despite on-off kinetics, allowing downstream events in the transcription initiation process (Extended Data Fig. 4d)[50–53].

1D diffusion has been observed for many DNA and RNA binding proteins, including bacterial TFs[3,54–59] and eukaryotic TF p53[60]. Here, we report the rapid back-and-forth, bp-scale 1D DNA diffusion during a eukaryotic TF's search for chromatin targets. The rate of a single GAF-DBD scanning from one cognate motif to another is strikingly higher than bacterial TF LacI by 4 orders of magnitude[3]. This suggests that the interaction between GAF-DBD and cognate sequence is much less stable than observed for LacI, providing an explanation for the selective occupancy of full-length GAF with multiple DBDs on a cluster of cognate sites[25,26]. Our findings raise the question whether 1D sliding is a general property of other eukaryotic TFs within and beyond the zinc finger TF family.

Furthermore, a requirement for 3D diffusion to robustly associate with nucleosome targets, to our knowledge, is unprecedented. Because free DNA only makes up 5-30% of chromatin in the gene body depending on nucleosome repeat lengths in different organisms[61], an optimal balance between 1D and 3D search modes has important implications for subclasses of TFs, especially those that are involved in pioneering chromatin alterations by efficient targeting of nucleosome-embedded cognate sites. On the other hand, a non-pioneer TF's intrinsic preference for a 1D search mode due to poor nucleosome association would conveniently restrict its search space to nucleosome-free regions. The kinetic interplay between TFs, chromatin remodelers and nucleosome mobility in creating dynamically accessible chromatin provides exciting future opportunities for deeper understanding of transcription regulation.

Materials and Methods

GAF-DBD protein expression and purification

GAF-DBD (Supplemental Information) was cloned into pET SUMO plasmid (Invitrogen K300-01) using Gibson Assembly (NEB E5520S) to contain an additional N-terminal cysteine for fluorophore labeling. The pET 6xHis-SUMO-Cys-GAF-DBD plasmid was transformed into *E. coli* (Novagen Rosetta 2 DE3, Sigma-Aldrich 71400) for protein expression. On the first day, a starter culture was grown from a single colony in 30 mL Terrific Broth (TB; Sigma-Aldrich T9179) and incubated overnight on a 37°C shaker. One the second day, the starter culture was added to 1 L TB and incubated at 37°C until OD reached 0.8. IPTG was added to 0.4 mM final concentration to induce protein expression. Cell pellets were harvested at 5 hr after IPTG induction, resuspended in 30 mL Lysis Buffer (50 mM Tris pH 7.4, 300 mM NaCl, 10% glycerol, 0.05% Triton X-100), and stored in -80°C.

6xHis-SUMO-Cys-GAF-DBD protein was first purified on Ni-NTA agarose (Qiagen 30210). Briefly, Triton X-100 (0.05% final concentration), protease inhibitor (Roche 4693132001), and β-mercaptoethanol (2 mM final conc.) were added to the thawed cell suspension. Cells were lysed by sonication (amplitude level 50, 10 s on, 50 s off, 3 min process time) on ice. Lysate was clarified by centrifugation at 18,000 g for 45 min at 4°C. The supernatant was transferred to 1.25 mL pre-equilibrated Ni-NTA agarose (2 mL slurry) in a 15 mL tube and incubated at 4°C for 1 hr with gentle shaking. Then agarose beads were washed with 10 mL of Wash Buffer 1 (WB1; 50 mM Tris pH 7.4, 500 mM NaCl, 10% glycerol, 0.05% Triton X-100, 2 mM BME, 20 mM Imidazole pH 8), centrifuged down, resuspended in another 10 mL of WB1, and transferred to a pre-chilled 2.5 cm gravity flow column. Beads were washed three times with 5 mL of WB1, followed by two times with 5 mL of Wash Buffer 2 (50 mM Tris pH 7.4, 300 mM NaCl, 10% glycerol, 0.05% Triton X-100). To elute the protein, 2 mL Elution Buffer (WB2 with 250 mM Imidazole) was added to the beads and incubated for 10 min. 1 mL elution fractions were collected immediately after each addition of elution buffer, and another 1 mL was collected after 5 min incubation. Typically, 10 fractions were collected, and most protein eluted in the first 2-3 fractions. Peak fractions were pooled and then further purified by cation exchange chromatography (Cytiva HiTrap SP HP 17115101) on a Fast Protein Liquid Chromatography (FPLC) instrument.

Labeling GAF-DBD with Cy3 fluorophore

We followed the "one-pot" reaction protocol by Jiang et al. [29] to cleave off the 6XHis-SUMO using SUMO protease (Invitrogen K300-01) and label the N-terminal cysteine with Cy3 in a site-specifically manner. Briefly, 6xHis-SUMO-Cys-GAF-DBD was buffer exchanged into Labeling Buffer (100 mM HEPES pH 6.9, 0.5 mM TCEP, 300 mM NaCl) via 24 hr dialysis. In parallel, the transesterification reaction of Cy3-NHS (Cytiva PA13105) was performed by incubating 0.2 mg of Cy3-NHS in 50 μL of 100 mM HEPES pH 6.9, 0.5 mM TCEP and 500 mM MESNa (Sigma-Aldrich PHR1570) for 6 hr at room temperature, to yield Cy3-MESNa. The "one-pot" labeling and protease cleavage reaction consists of 100 μL of dialyzed protein, 10 μL of SUMO protease, and 12 μL of 6 mM Cy3-MESNa. The reaction proceeded for 36 hr at room temperature. The cleaved 6XHis-SUMO and 6XHis-tagged SUMO protease were removed by Ni-NTA agarose. The labeling efficiency of GAF-DBD was measured to be 92%.

2-color TIRF microscope

See Poyton *et al.* Methods: Two-Color Single-Molecule FRET Microscope Instrumentation[39].

2-color smFRET imaging of Cy3-GAF-DBD binding to Cy5-labeled DNA

Two complementary single-stranded DNA – one containing an internal Cy5 fluorophore, another containing a 5' biotin – were annealed to form Cy5-labeled DNA constructs (Supplemental Information, Section 2). On a PEG-passivated and sparsely biotinylated flow channel (Nano Surface Sciences, PEG passivated and biotin functionalized quartz slide and glass coverslip pair), inject the reagents in order: 100 μL of T50 Buffer (10 mM Tris-HCl pH 8, 50 mM NaCl), 40 uL of 0.2 mg/mL Neutravidin in T50 (1 min incubation; Thermo Fisher 31000), 100 uL of T50, 50 μL of 12.5 pM Cy5-DNA (3 min incubation), 100 μL of T50, 50 uL of 0.2 nM Cy3-GAF-DBD in Imaging Buffer (50 mM NaCl, 50 ug/mL BSA (Roche 10711454001), 0.05% NP40 (Sigma I8896), 12.5 mM HEPES–KOH pH 7.6, 0.05 mM EDTA, 6.25 mM MgCl2, 5% glycerol, 0.8% w/v dextrose, 2 mM Trolox, 1 mg/ml glucose oxidase (Sigma-Aldrich G2133) and 500 U/ml catalase (Sigma-Aldrich C3155)). Cy3-GAF-DBD was pre-diluted to 20 nM using Storage Buffer (50 mM Tris pH 7.4, 300 mM NaCl, 10% glycerol, 0.05% Triton X-100) to prevent aggregation. The channel was imaged under a 2-color TIRF microscope with 10 Hz frame rate (unless stated otherwise). Laser excitation was programmed to be 10 frames of Cy5 excitation followed by 990 frames of Cy3 excitation for a 100-second movie.

2-color *smFRET data analysis*

2-color movies were converted to single-molecule fluorescence intensity time trajectories using custom-written IDL scripts and analyzed using custom-written MATLAB scripts. Dwell time data were manually collected by recording the start and end time of each event of interest. The 1-CDF (cumulative distribution function) of the dwell time histogram was fit to a single-exponential decay function (ExpDec1 function: $y = y_0 + e^{-x/\tau}$) in OriginPro. Binding frequency data were manually collected by counting the total number (*N*) of binding events in each movie and dividing by the product of movie length (*t* seconds) and number of trajectories (*k*) in the movie: frequency = *N* / (*t* * *k*).

Labeling DNA oligo with Cy5 or Cy7

Single-stranded DNA oligos were site-specifically labeled with Cy5-NHS (Cytiva PA15100) or Cy7-NHS (Sulfo-Cyanine7-NHS, Lumiprobe 25320) via an amino group attached to an internal thymine through a 6-carbon linker (Integrated DNA Technologies, /iAmMC6T/). A standard 62.5 μL labeling reaction contains 160 μM amino-modified oligo, 200 mM freshly dissolved $NaHCO_3$, 8 mM NHS-dye and nuclease-free water. The reaction was incubated with gentle mixing for 4 hr at room temperature, then overnight at 4°C. The labeled oligo was purified by ethanol precipitation to remove excess dye. Cy7-labeled oligo typically requires two rounds of ethanol precipitation to eliminate free dye. If the labeling efficiency was lower than 70%, a second round of labeling reaction would be performed on the labeled oligo. The final labeling efficiencies were typically 80-90%.

Cy5 Cy7 dual-labeled DNA construction

Making a DNA construct dual-labeled at our desired positions was an engineering challenge because a one-step PCR reaction would require an internally labeled DNA oligo to be longer than 100 bases. Such DNA oligos were not produced by IDT at the time. (Nowadays it can be purchased as an Ultramer.) To overcome this challenge, a shorter dual-labeled DNA was made

by PCR, restriction digested to produce a sticky overhang, and then ligated with a biotinylated DNA fragment to form the complete construct. Briefly, perform a 2.4 mL PCR reaction using GoTaq Buffer (Promega M7921), 200 μM dNTPs, 1 μM forward primer, 1 μM reverse primer, 100 ng or less template DNA and 24 μL Taq DNA polymerase (NEB M0273L). Purify and concentrate the PCR DNA product by ethanol precipitation. Digest the DNA with DraIII-HF restriction enzyme (NEB R3510L) for 3 hr at 37°C. Purify the digested DNA product by anion exchange chromatography (Cytiva 17115301) on an FPLC instrument, followed by ethanol precipitating DNA from the peak fractions. Simultaneously prepare the biotinylated DNA fragment by annealing two single-stranded DNA oligos. Ligate the Cy5 Cy7 dual-labeled DNA with the biotinylated fragment using T4 DNA ligase (Thermo Fisher EL0011) and purify the ligated DNA by agarose gel extraction or FPLC in case of a large-scale production.

Cy5 Cy7 dual-labeled 601 nucleosome reconstitution

See Poyton *et al.* Methods: Nucleosome Reconstitution[39].

Cy5 Cy7 dual-labeled hsp70 nucleosome reconstitution

The DNA sequence of the hsp70 nucleosome is a modified 187-bp fragment of the *Drosophila melanogaster* hsp70 promoter. According to previously mapped nucleosome positions on a longer hsp70 promoter fragment, our 187 bp DNA is estimated to form a 40-N-0 nucleosome, where the 147 bp nucleosomal DNA (N) contains all 9 near-cognate and cognate motifs (GAG, GAGA, GAGAG, etc.) on the DNA. To mutate 7 of the motifs, leaving only two GAF motifs intact, we swapped the 7 motifs with the corresponding nucleotides on the Widom 601 sequence[34] to preserve nucleosome-positioning property of the DNA. To ensure that the nucleosome forms at the desired position on DNA, we followed the nucleosome ligation protocol in Huh et al.[37] Briefly, we reconstituted the nucleosome core particle (NCP) using a 150 bp + 3 nt DNA, heat-shifted the NCP at 55°C for 30 min for betting positioning, and then used T4 ligase to attach the 37 bp + 3 nt biotinylated linker DNA by incubating at 4°C for 12 hr.

3-color TIRF microscope

See Poyton *et al.* Methods: Three-Color Single-Molecule FRET Microscope Instrumentation[39].

3-color imaging of Cy3-GAF-DBD binding to Cy5 & Cy7-labeled DNA or nucleosome

On a PEG-passivated and sparsely biotinylated flow channel, flow in the following reagents in order: 100 μL of T50 Buffer, 40 uL of 0.2 mg/mL Neutravidin in T50 (1 min incubation), 100 uL of T50, 50 μL of 50 pM Cy5-Cy7-labeled-nucleosome or DNA (3 min incubation), 100 μL of T50, 50 uL of 0.1 nM (unless stated otherwise) Cy3-GAF-DBD in Imaging Buffer. The channel was imaged under a 3-color TIRF microscope with 28.6 Hz frame rate (35 ms exposure time per frame) for 601 nucleosomes with GAF motifs (Fig. 3) or 10 Hz frame rate (100 ms exposure time) for hsp70 nucleosome (Fig. 4). Laser excitation was programmed to be 10 frames Cy7, 10 frames Cy5, 960 frames Cy3,10 frames Cy7, 10 frames Cy5.

3-color smFRET data analysis

3-color movies were converted to single-molecule fluorescence intensity time trajectories using custom-written IDL scripts, and analyzed using custom-written MATLAB scripts (https://github.com/ashleefeng/singlemolecules/tree/master/smfret3color). Binding event dwell times were manually collected by recording the start and end time of each event of interest. Single-molecule trajectories were converted into colored tracks using custom-written MATLAB scripts for visualization and motif dwell time analysis. The 1-CDF of binding dwell time or motif dwell time histogram was fit to ExpDec1 function in Origin. Binding events on 601 nucleosomes were manually categorized into nucleosome and linker DNA binding (FRET alternates between Cy5 and Cy7 in the same binding event), linker DNA binding only (Cy3-Cy5 FRET only), or nucleosome binding only (Cy3-Cy7 FRET only) (Fig. 4h). The landing site of GAF-DBD on SHL7-601 (Fig. 4i) nucleosome was analyzed by manually categorizing the initial FRET pattern of binding events into 1) landing on linker DNA and sliding to nucleosome (Cy3-Cy5 to Cy3-Cy7 FRET), 2) landing on nucleosome and sliding to linker DNA (Cy3-Cy7 to Cy3-Cy5 FRET) , 3) linker DNA binding only (Cy3-Cy5 FRET throughout), 4) nucleosome binding only (Cy3-Cy7 FRET throughout), 5) nonspecific binding and sliding to linker DNA (Cy3 only to Cy3-Cy5 FRET), and 6) nonspecific binding and sliding to nucleosome (Cy3 only to Cy3-Cy7 FRET).

Acknowledgements

We thank Richard He and Eric Lin for assistance with molecular cloning. This work was supported by the US National Institutes of Health (GM122569 to T.H., GM145844 to C.W.), Johns Hopkins University Discovery Award (C.W., G.B. and T.H.), Howard Hughes Medical Institute (T.H.), and Bloomberg Distinguished Professorships (T.H. and C.W.).

References

1. Von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
2. Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
3. Marklund, E. *et al.* DNA surface exploration and operator bypassing during target search. *Nature* **583**, 858–861 (2020).
4. Suter, D. M. Transcription Factors and DNA Play Hide and Seek. *Trends Cell Biol.* **30**, 491–500 (2020).
5. Luo, Y., North, J. A., Rose, S. D. & Poirier, M. G. Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res.* **42**, 3017–3027 (2014).
6. Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
7. Michael, A. K. *et al.* Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science (80-.).* **368**, 1460–1465 (2020).
8. Michael, A. K. *et al.* Cooperation between bHLH transcription factors and histones for DNA access. *Nature* **619**, 385–393 (2023).
9. Donovan, B. T. *et al.* Basic helix-loop-helix pioneer factors interact with the histone octamer to invade nucleosomes and generate nucleosome-depleted regions. *Mol. Cell* **83**, 1251-1263.e6 (2023).
10. Donovan, B. T., Chen, H., Jipa, C., Bai, L. & Poirier, M. G. Dissociation rate compensation mechanism for budding yeast pioneer transcription factors. *Elife* **8**, (2019).
11. Mivelaz, M. *et al.* Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. *Mol. Cell* **77**, 488-500.e9 (2020).
12. Adkins, N. L., Hagerman, T. A. & Georgel, P. GAGA protein: A multi-faceted transcription

factor. in *Biochemistry and Cell Biology* vol. 84 559–567 (2006).
13. Tsukiyama, T., Becker, P. B. & Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525–532 (1994).
14. Fuda, N. J. *et al.* GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. *PLoS Genet.* **11**, e1005108 (2015).
15. Tsukiyama, T. & Wu, C. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**, 1011–1020 (1995).
16. Gaskill, M. M., Gibson, T. J., Larson, E. D. & Harrison, M. M. GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. *Elife* **10**, (2021).
17. Chetverina, D., Erokhin, M. & Schedl, P. GAGA factor: a multifunctional pioneering chromatin protein. *Cell. Mol. Life Sci.* **78**, 4125–4141 (2021).
18. Wilkins, R. C. & Lis, J. T. DNA distortion and multimerization: novel functions of the glutamine-rich domain of GAGA factor. *J. Mol. Biol.* **285**, 515–525 (1999).
19. Mahmoudi, T., Katsani, K. R. & Verrijzer, C. P. GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *EMBO J.* **21**, 1775–81 (2002).
20. Li, X. *et al.* GAGA-associated factor fosters loop formation in the Drosophila genome. *Mol. Cell* **83**, 1519-1526.e4 (2023).
21. Tang, X. *et al.* Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. *Nat. Struct. Mol. Biol.* **29**, 665–676 (2022).
22. Duarte, F. M. *et al.* Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev.* **30**, 1731–1746 (2016).
23. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
24. Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M. & Clore, G. M. The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat. Struct. Biol.* **4**, 122–132 (1997).
25. Katsani, K. R., Hajibagheri, M. A. N. & Verrijzer, C. P. Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J.* **18**, 698–708 (1999).
26. Van Steensel, B., Delrow, J. & Bussemaker, H. J. Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2580–2585 (2003).
27. Batut, P. J. *et al.* Genome organization controls transcriptional dynamics during development. *Science (80-.).* **375**, 566–570 (2022).
28. Ha, T. *et al.* Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6264–6268 (1996).
29. Jiang, H., D'Agostino, G. D., Cole, P. A. & Dempsey, D. R. Selective protein N-terminal labeling with N-hydroxysuccinimide esters. in *Methods in Enzymology* vol. 639 333–353 (Academic Press Inc., 2020).
30. Bonnet, I. *et al.* Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* **36**, 4118–4127 (2008).
31. Mirny, L. *et al.* How a protein searches for its site on DNA: The mechanism of facilitated diffusion. *J. Phys. A Math. Theor.* **42**, 434013 (2009).
32. Moll, J. R., Acharya, A., Gal, J., Mir, A. A. & Vinson, C. Magnesium is required for specific DNA binding of the CREB B-ZIP domain. *Nucleic Acids Res.* **30**, 1240–1246 (2002).
33. Lee, S., Lee, J. & Hohng, S. Single-molecule three-color FRET with both negligible spectral overlap and long observation time. *PLoS One* **5**, e12270 (2010).
34. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).

35. Hamiche, A., Sandaltzopoulos, R., Gdula, D. A. & Wu, C. ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* **97**, 833–842 (1999).
36. Georgel, P. T. Chromatin potentiation of the hsp70 promoter is linked to GAGA-factor recruitment. *Biochem. Cell Biol.* **83**, 555–565 (2005).
37. Huh, J.-W. *et al.* Multivalent di-nucleosome recognition enables the Rpd3S histone deacetylase complex to tolerate decreased H3K36 methylation levels. *EMBO J.* **31**, 3564–3574 (2012).
38. Sabantsev, A., Levendosky, R. F., Zhuang, X., Bowman, G. D. & Deindl, S. Direct observation of coordinated DNA movements on the nucleosome during chromatin remodelling. *Nat. Commun.* **10**, 1–12 (2019).
39. Poyton, M. F. *et al.* Coordinated DNA and histone dynamics drive accurate histone H2A.Z exchange. *Sci. Adv.* **8**, 2021.10.22.465479 (2022).
40. Stinson, B. M., Moreno, A. T., Walter, J. C. & Loparo, J. J. A Mechanism to Minimize Errors during Non-homologous End Joining. *Mol. Cell* **77**, 1080-1091.e8 (2020).
41. Li, S., Zheng, E. B., Zhao, L. & Liu, S. Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell Rep.* **28**, 2689-2703.e4 (2019).
42. Rudnizky, S., Khamis, H., Malik, O., Melamed, P. & Kaplan, A. The base pair-scale diffusion of nucleosomes modulates binding of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12161–12166 (2019).
43. Tan, C. & Takada, S. Nucleosome allostery in pioneer transcription factor binding. *Proc. Natl. Acad. Sci.* **117**, 20586–20596 (2020).
44. Judd, J., Duarte, F. M. & Lis, J. T. Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. *Genes Dev.* **35**, 147–156 (2021).
45. Calderon, D. *et al.* The continuum of Drosophila embryonic development at single-cell resolution. *Science (80-.).* **377**, (2022).
46. Kubik, S. *et al.* Opposing chromatin remodelers control transcription initiation frequency and start site selection. *Nat. Struct. Mol. Biol.* **26**, 744–754 (2019).
47. Okada, M. & Hirose, S. Chromatin Remodeling Mediated by Drosophila GAGA Factor and ISWI Activates fushi tarazu Gene Transcription In Vitro. *Mol. Cell. Biol.* **18**, 2455–2461 (1998).
48. Nakayama, T., Shimojima, T. & Hirose, S. The PBAP remodeling complex is required for histone H3.3 replacement at chromatin boundaries and for boundary functions. *Development* **139**, 4582–4590 (2012).
49. Xiao, H. *et al.* Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* **8**, 531–543 (2001).
50. Badenhorst, P., Voas, M., Rebay, I. & Wu, C. Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.* **16**, 3186–3198 (2002).
51. Boija, A. *et al.* CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II. *Mol. Cell* **68**, 491-503.e5 (2017).
52. Kuroda, M. I., Kang, H., De, S. & Kassis, J. A. Dynamic Competition of Polycomb and Trithorax in Transcriptional Programming. *Annu. Rev. Biochem.* **89**, 235–253 (2020).
53. Wolle, D. *et al.* Functional Requirements for Fab-7 Boundary Activity in the Bithorax Complex. *Mol. Cell. Biol.* **35**, 3739–3752 (2015).
54. Carcamo, C. C. *et al.* ATP binding facilitates target search of SWR1 chromatin remodeler by promoting one-dimensional diffusion on DNA. *Elife* **11**, (2022).
55. Gowers, D. M., Wilson, G. G. & Halford, S. E. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci.* **102**, 15883–15888 (2005).
56. Ragunathan, K., Liu, C. & Ha, T. RecA filament sliding on DNA facilitates homology

search. *Elife* **2012**, e00067 (2012).
57. Globyte, V., Lee, S. H., Bae, T., Kim, J. & Joo, C. CRISPR /Cas9 searches for a protospacer adjacent motif by lateral diffusion . *EMBO J.* **38**, (2019).
58. Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., Macrae, I. J. & Joo, C. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell* **162**, 96–107 (2015).
59. Kim, J. M. *et al.* Dynamic 1D Search and Processive Nucleosome Translocations by RSC and ISW2 Chromatin Remodelers Summary Introduction. (2023).
60. Tafvizi, A. *et al.* Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* **95**, L01–L03 (2008).
61. Perišić, O., Collepardo-Guevara, R. & Schlick, T. Modeling Studies of Chromatin Fiber Structure as a Function of DNA Linker Length. *J. Mol. Biol.* **403**, 777–802 (2010).
62. Xiao, H. *et al.* Dual Functions of Largest NURF Subunit NURF301 in Nucleosome Sliding and Transcription Factor Interactions. *Mol. Cell* **8**, 531–543 (2001).

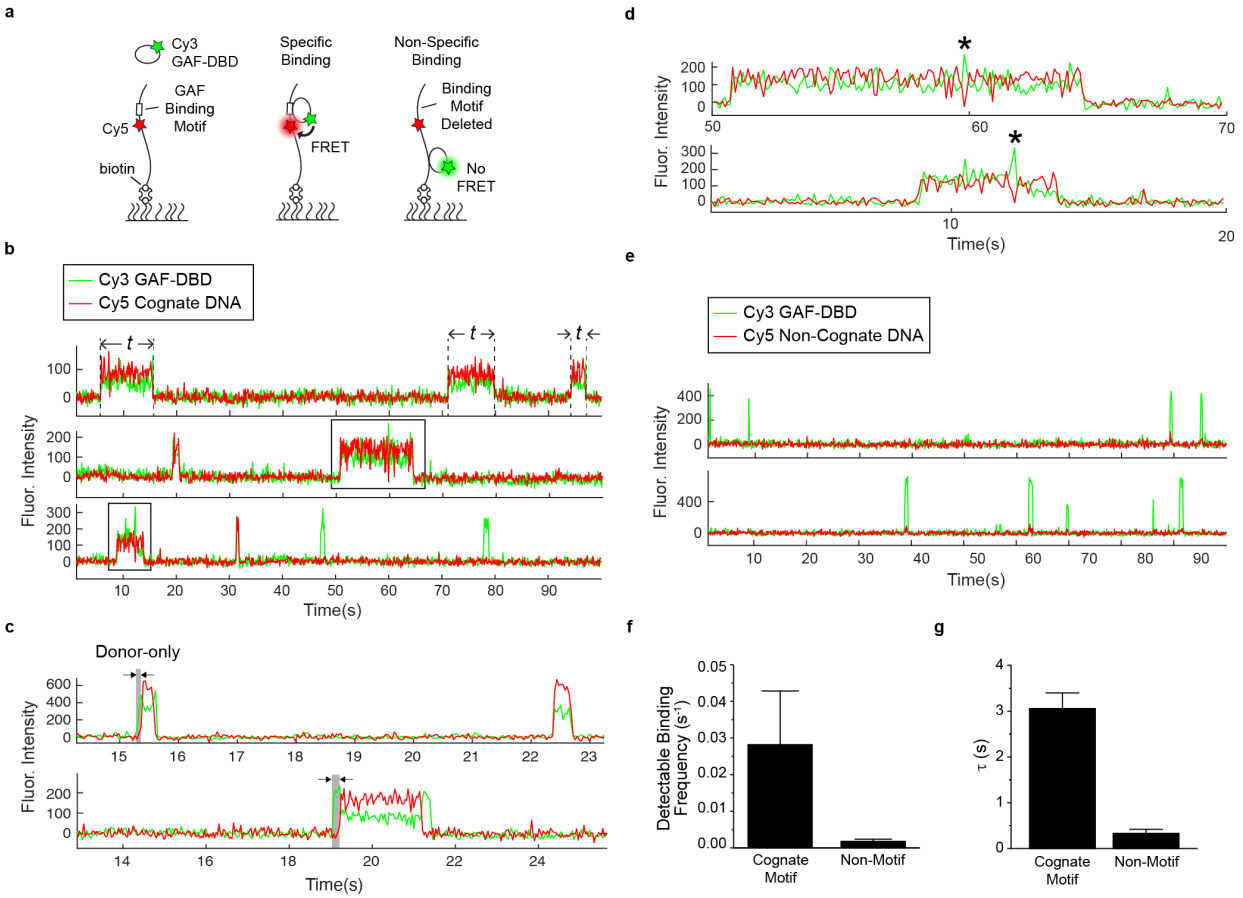


Figure 1. DNA sequence specificity of GAF-DBD is kinetically defined. **a**, Schematics of 2-color smFRET experiment to measure cognate motif binding of GAF-DBD. **b**, Single-molecule trajectories showing Cy3-labeled GAF-DBD binding to Cy5-labeled DNA containing a GAF cognate motif. **c**, Representative single-molecule trajectories of Cy5-cognate DNA bound by Cy3-GAF-DBD showing an initial donor-only period before acceptor signal increases. The donor-only period is highlighted in grey. **d**, Zoomed-in view of binding events boxed in **b**. "*" indicates transient donor-only fluorescence within a binding event. **e**, Single-molecule trajectories showing Cy3-labeled GAF-DBD nonspecifically binding to Cy5-labeled non-cognate DNA. **f**, Binding frequency of GAF-DBD to cognate or non-cognate DNA. **g**, Dwell time of GAF-DBD on cognate or non-cognate DNA from fitting 1-CDF to an exponential decay function. Error bars show standard deviation of three technical replicates.

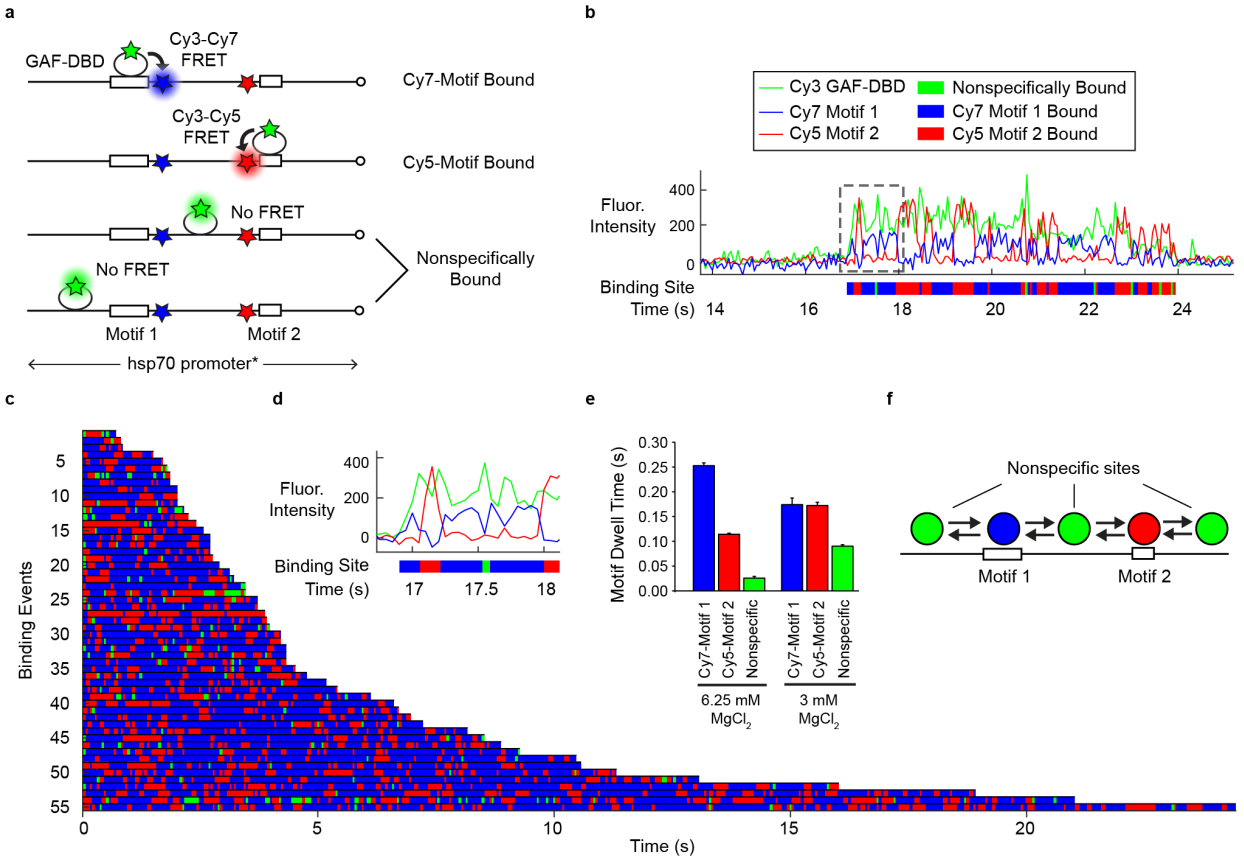


Figure 2. GAF-DBD explores free DNA by 1-dimensional diffusion. **a**, 3-color smFRET distinguishes whether GAF-DBD is bound to Cy7-labeled Motif 1, Cy5-labeled Motif 2, or a nonspecific site on the *D. melanogaster* hsp70 promoter DNA. **b**, A single-molecule trajectory on a DNA molecule shows GAF-DBD sliding back-and-forth on the DNA between two cognate motifs. Binding location over time is shown below the trajectory as a colored track. **c**, 55 binding events shown as colored tracks. **d**, Zoomed-in view of the boxed region in **b**. **e**, GAF-DBD dwell times on Motif 1, Motif 2, or a nonspecific site on the DNA at regular (6.25 mM) or low (3 mM) MgCl2 concentration. **f**, Schematic of GAF-DBD sliding on DNA. Circles represent GAF-DBD. Green circles indicate GAF-DBD binding to a nonspecific site; the blue circle indicates binding to Motif 1; the red circle indicates binding to Motif 2.

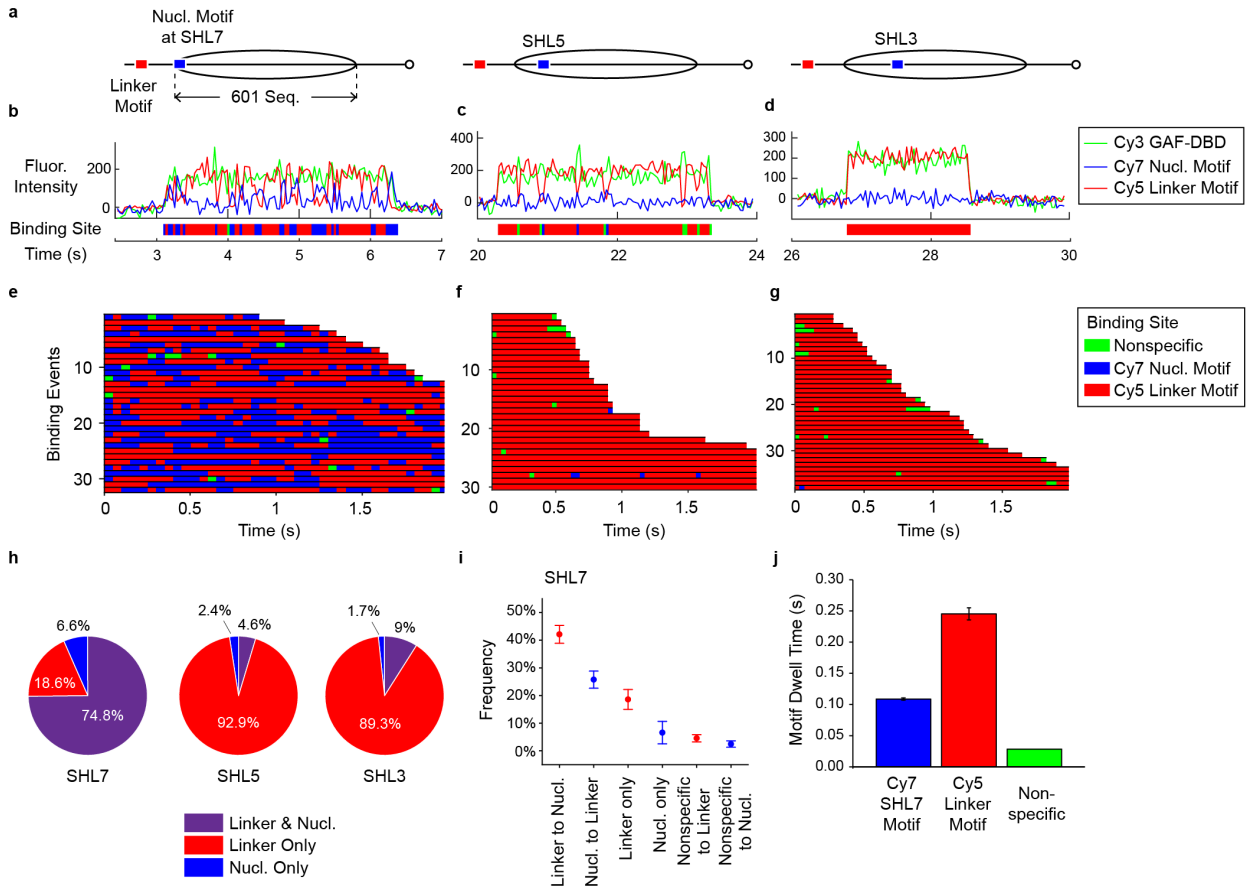


Figure 3. Nucleosome blocks GAF-DBD 1D sliding to SHL5 and deeper locations into the nucleosome core. **a**, Schematics of nucleosome constructs where the nucleosomal Cy7-labeled cognate motif (in blue) is placed at SHL7, SHL5, or SHL3 location and the other Cy5-labeled cognate motif (in red) is on the linker DNA. The nucleosome is positioned by the Widom 601 sequence flanked by 40 bp linker DNA on both sides. **b-d**, Representative single-molecule trajectories of Cy3-GAF-DBD binding to linker or nucleosomal motifs on the SHL7 (**b**), SHL5 (**c**) and SHL3 motif (**d**) nucleosome constructs. **e-g**, Binding site locations for GAF-DBD binding events on nucleosome constructs where Cy7-labeled motif is located at SHL7 (**e**), SHL5 (**f**) or SHL3 (**g**) location. **h**, GAF-DBD binding behavior categories on each nucleosome construct. **i**, Categories of 1D diffusion behaviors on the SHL7 construct. **j**, GAF-DBD dwell times on the Cy7-SHL7 motif, Cy5-linker motif and nonspecific sites on the SHL7 construct.

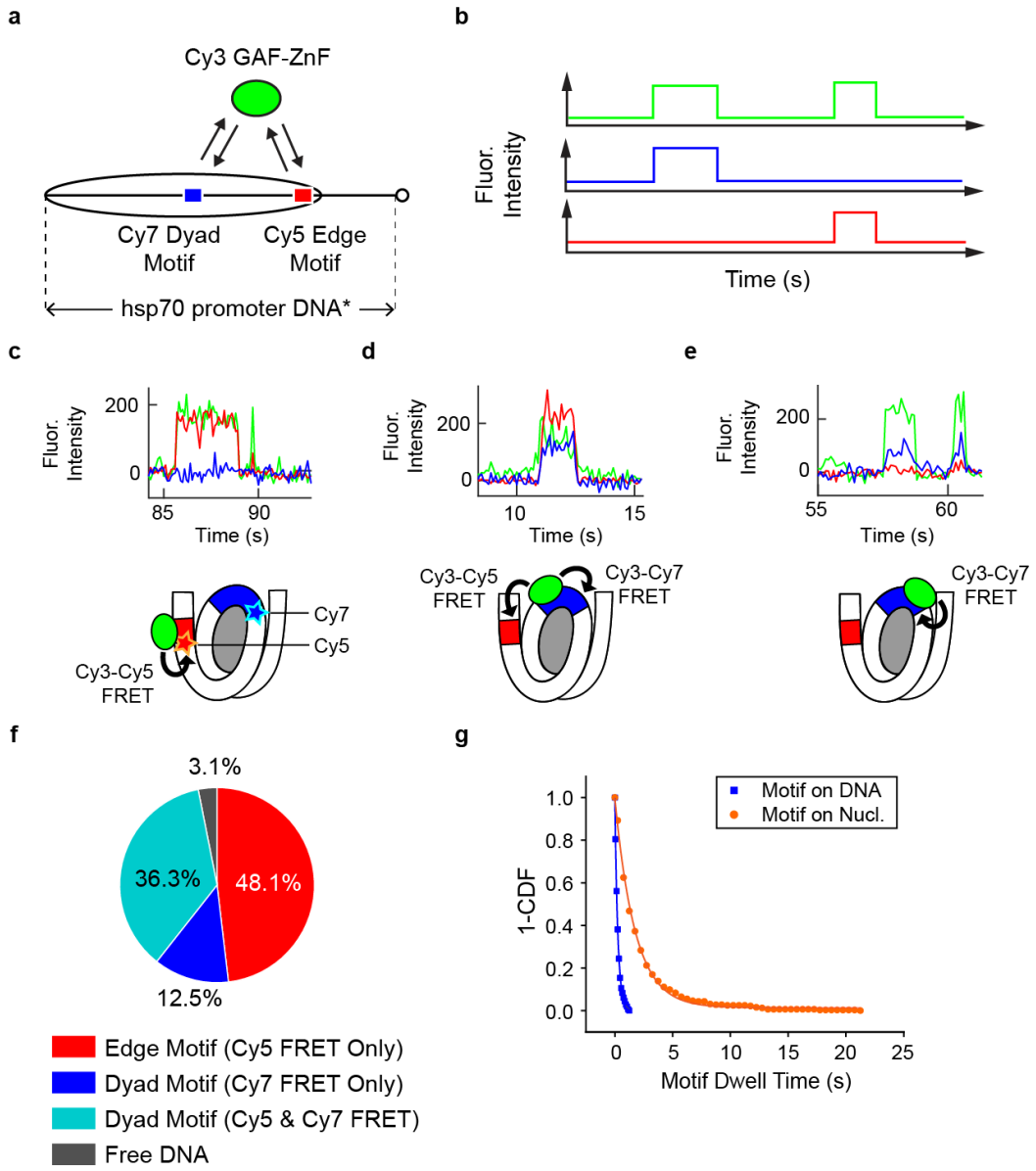


Figure 4. 3D diffusion dominates inner nucleosomal target search. **a**, *D. melanogaster hsp70* nucleosome construct for investigating GAF-DBD target search when both cognate motifs are embedded in the nucleosome. **b**, Idealized single-molecule trajectories for GAF-DBD locating cognate motifs on the hsp70 nucleosome. **c-e**, Representative single-molecule trajectories on the hsp70 nucleosome, with corresponding FRET schematic shown below the trajectory. **f**, Binding event categories on the hsp70 promoter nucleosome construct. **g**, 1-CDF of motif dwell times on nucleosome (orange) compared with free DNA (blue). Free DNA data are from Cy7-Motif 1 dwell time at 6.25 mM $MgCl_2$ (Fig. 2e).

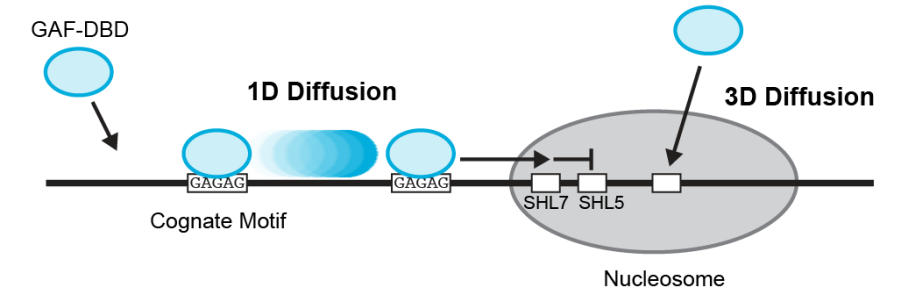
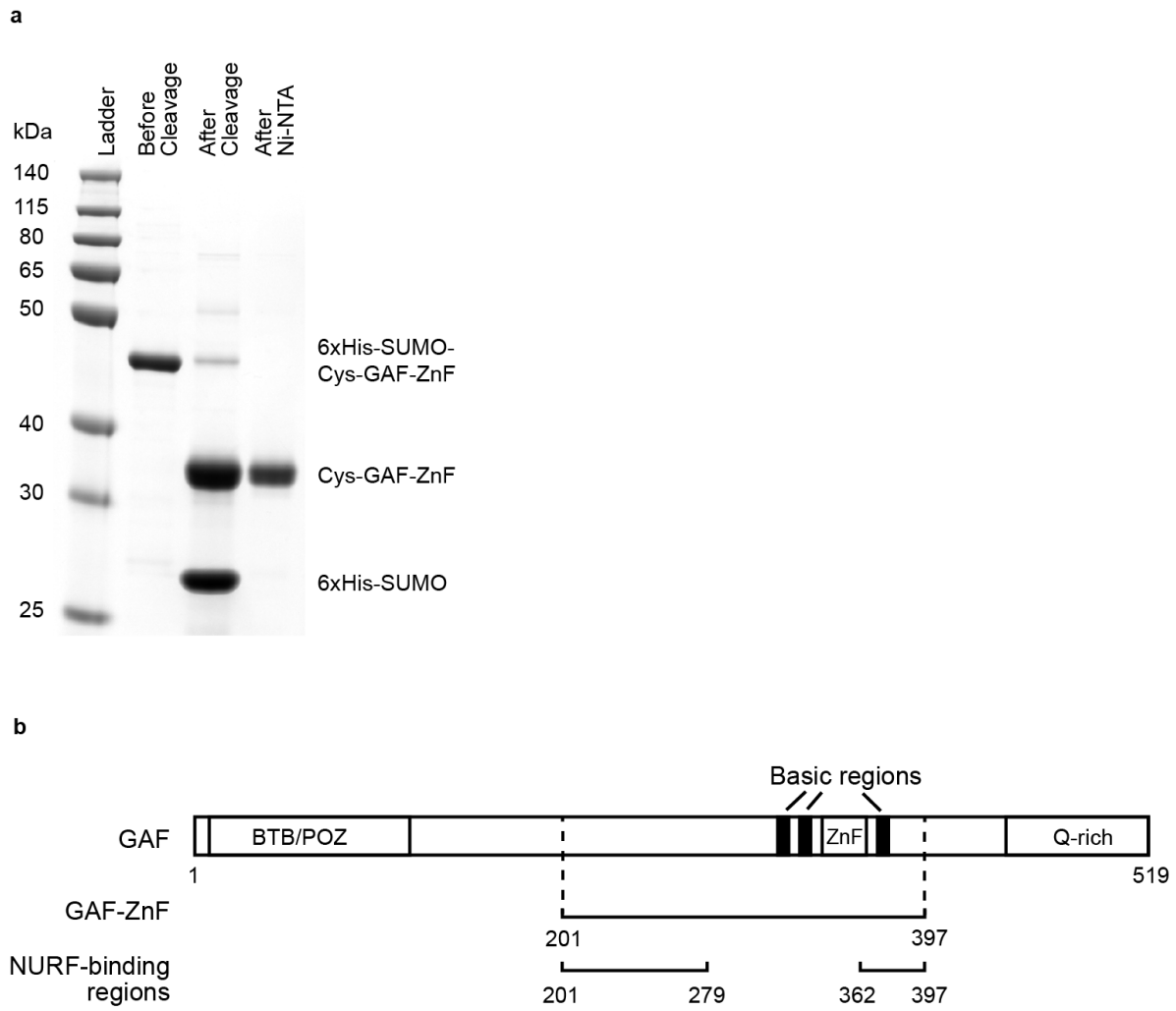
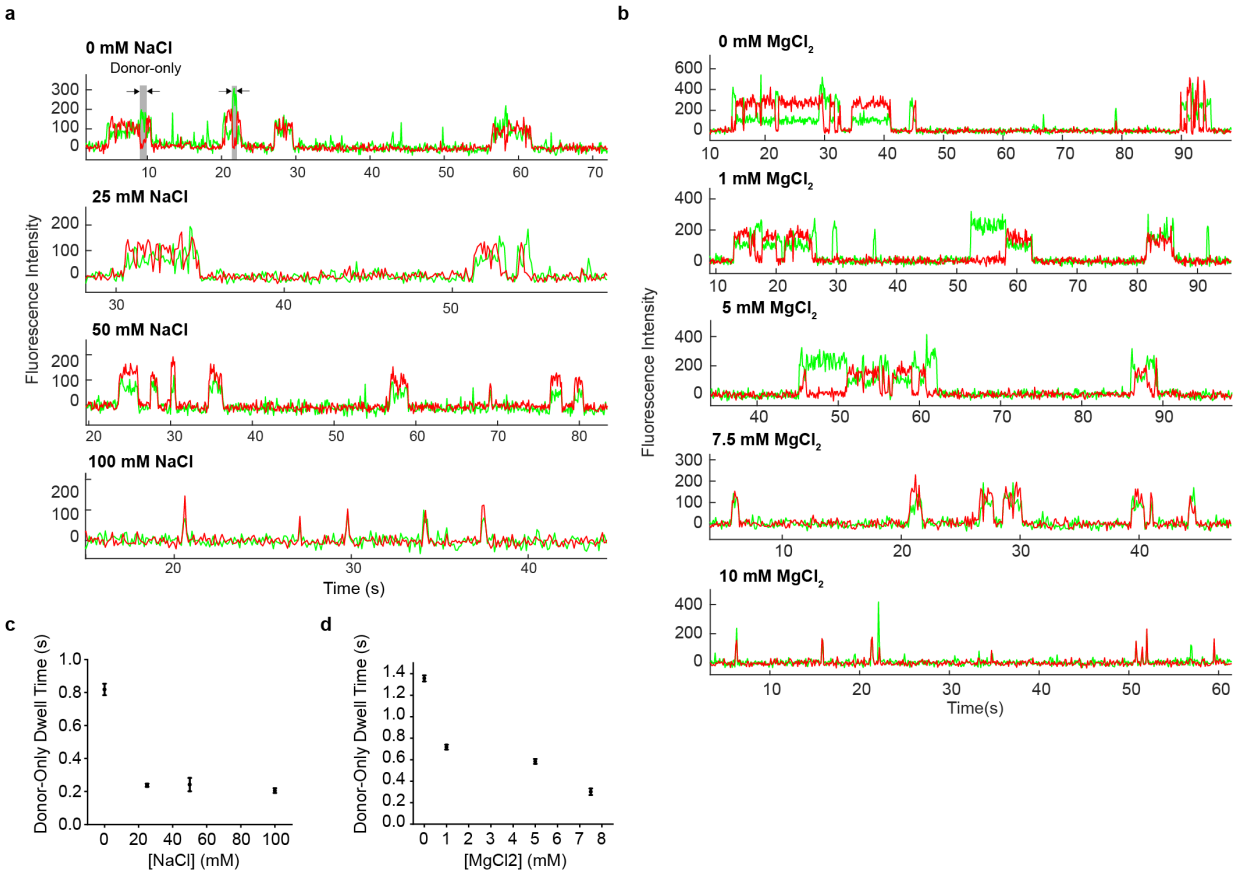


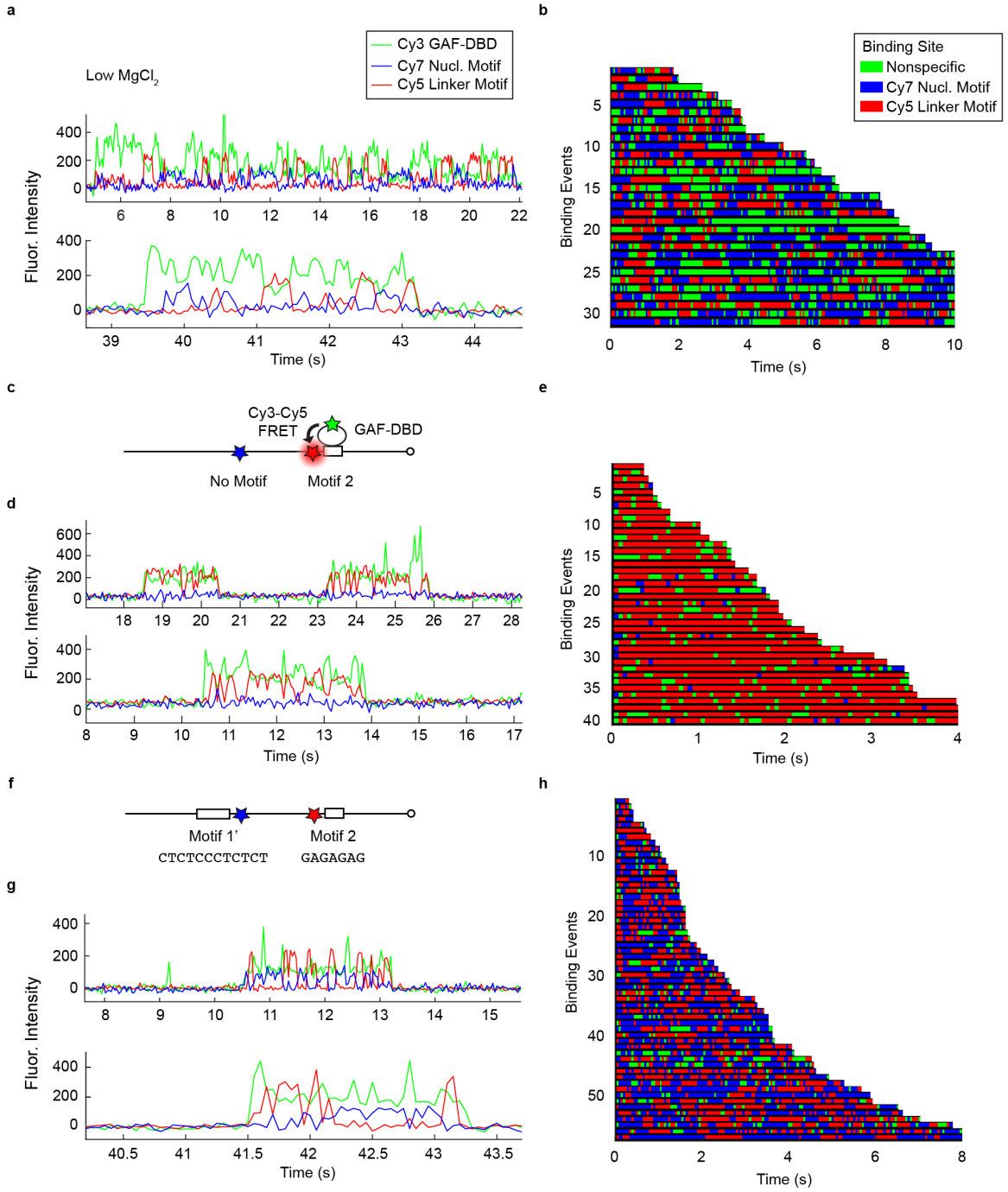
Figure 5. Model for GAF-DBD target search on chromatin. GAF-DBD uses two search modes to locate its target on chromatin. In the 1D sliding mode, GAF-DBD lands on an off-target location on free DNA, then slides on DNA to locate the adjacent GAF motif ("GAGAG"). It can escape the cognate site to locate another cognate site nearby. 1D search allows GAF-DBD to invade into the nucleosome edge and locate cognate motifs at SHL7. Alternatively, GAF-DBD can also directly associate with a cognate motif in the nucleosome core from 3D space. This 3D search mode allows GAF to effectively target nucleosomal motifs that are inaccessible by 1D sliding.



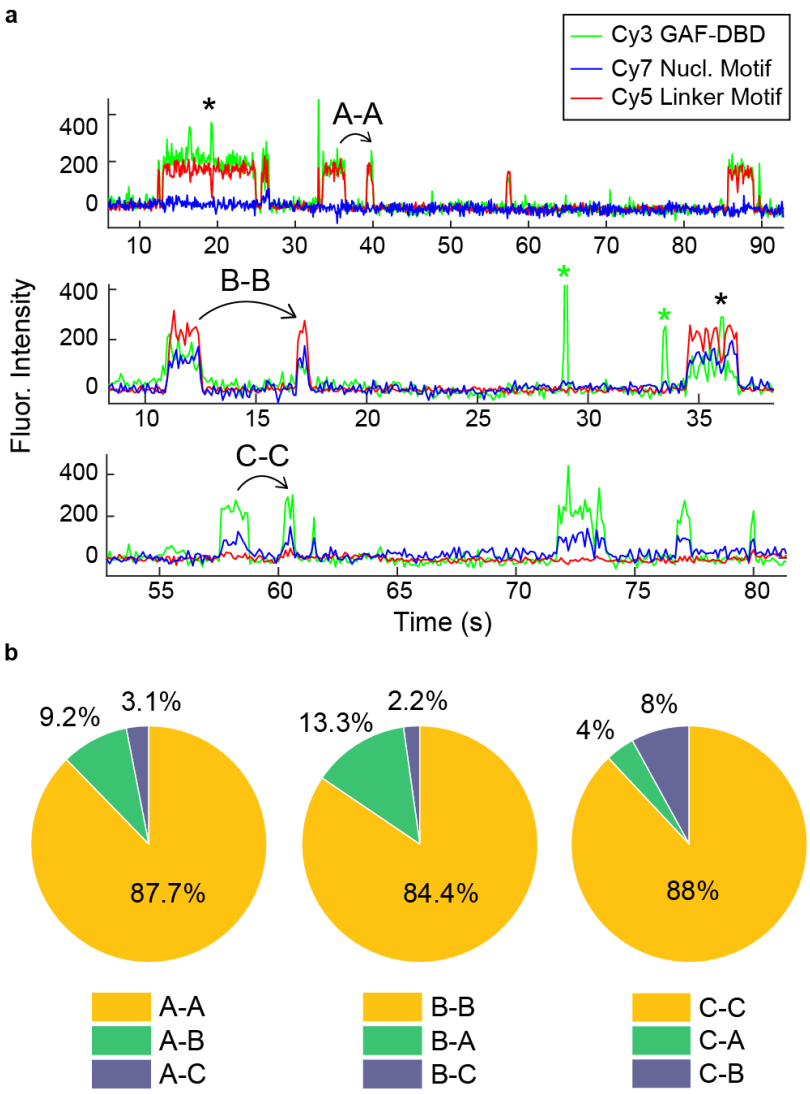
Extended Data Figure 1. Purification of GAF-DBD protein. **a**, Subtractive Ni-NTA purification after "one-pot" reaction where 6xHis-SUMO was cleaved off and the N-terminal cysteine is exposed and labeled with Cy3. The calculated molecular weights are 34.4 kDa for 6xHis-SUMO-Cys-GAF-DBD, 21.0 kDa for Cys-GAF-DBD and 13.4 kDa for 6xHis-SUMO. **b**, Schematics of full-length GAF short isoform, GAF-DBD and NURF-binding regions[62] to scale.



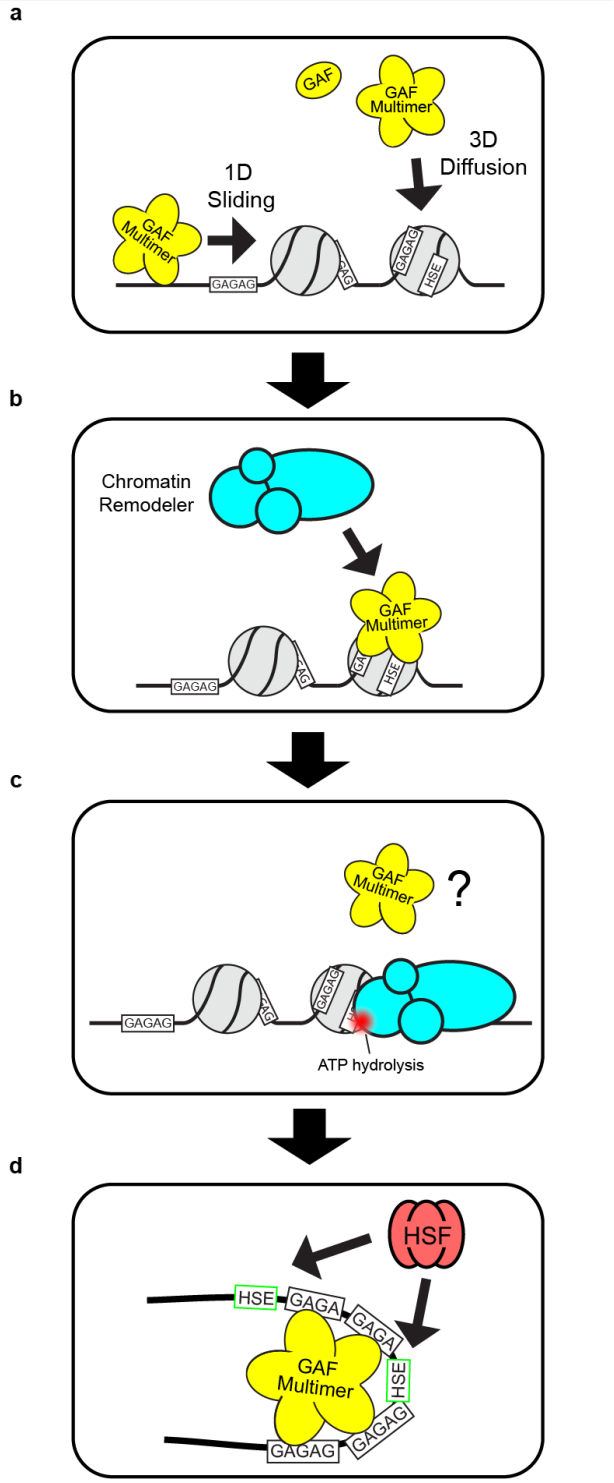
Extended Data Figure 2. Donor-only dwell time is inversely correlated with ionic strength. **a**, Representative single-molecule trajectories when Cy5-cognate DNA is bound by Cy3-GAF-DBD in 0, 25, 50 or 100 mM NaCl. Grey highlighted durations indicate donor-only dwell times. **b**, Representative trajectories when Cy5-cognate DNA was bound by Cy3-GAF-DBD in 0, 1, 5, 7.5, or 10 mM $MgCl_2$. **c**, Donor-only dwell time (τ from fitting 1-CDF to a single-exponential decay; see Material and Methods) depends on NaCl concentration. **d**, Donor-only dwell time also depends on $MgCl_2$ concentration. Error bars are standard error from fitting.



Extended Data Figure 3. GAF sliding kinetics on Cy5 & Cy7 dual-labeled DNA depends on salt concentration and cognate motif. **a**, Representative single-molecule trajectories of Cy5 & Cy7 DNA bound by Cy3-GAF-DBD when the buffer contains low $MgCl_2$ (3 mM). **b**, 31 colored tracks of Cy5 & Cy7 DNA bound by Cy3-GAF-DBD at 3 mM $MgCl_2$. **c**, Schematic of Motif 2 Only construct where Motif 1 was mutated to a non-cognate sequence. **d**, Representative single-molecule trajectories of Motif 2 Only DNA bound by Cy3-GAF-DBD. **e**, 40 colored tracks of Motif 2 Only DNA bound by Cy3-GAF-DBD.



Extended Data Figure 4. GAF-DBD preferentially visits the same cognate motif on the same hsp70 nucleosome. **a**, Representative single-molecule fluorescence intensity time trajectories showing repetitive visits to the same binding site on a single nucleosome. Upper trace shows repetitive visits to binding site A; middle trace shows repetitive visits to binding site B; lower trace shows repetitive visits to binding site C. Black asterisks mark transient Cy3-only fluorescence within a binding event, potentially caused by short-range 1D diffusion on the nucleosome. Green asterisks indicate binding events to non-cognate sites on the nucleosome. **b**, For all binding events on site A (left pie chart, N = 65), B (middle, N = 45) or C (right, N = 25), the binding location of the following binding event.



Extended Data Figure 5. Hypothetical model of full-length GAF pioneering chromatin opening at a target promoter.