Title
Building a novel single-molecule system to study readers of DNA methylation with high resolution reveals binding preferences of MBD proteins

Permalink
https://escholarship.org/uc/item/0965973k

Author
Strauskulage, Luke

Publication Date
2023

Peer reviewed|Thesis/dissertation

Building a novel single-molecule system to study readers of DNA methylation with high resolution reveals binding preferences of MBD proteins

by
Luke Strauskulage

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Danica Galonic Fujimori

82967A5AE18B4D6...

Danica Galonic Fujimori
_____
Chair

DocuSigned by:

Jeremy Reiter

DocuSigned by:4B3...

Jeremy Reiter
_____

Geeta Narlikar

DocuSigned by:43B...

Geeta Narlikar
_____

Barbara Panning

B18F20197C95417...

Barbara Panning
_____

_____
Committee Members

Dedicated to my aunt, Theresa Strauskulage (June 7, 1958 - June 19, 2018), and

my Nana, Anne Zigmont (March 21, 1936 - March 14, 2023).

Acknowledgements

I want to thank all of my collaborators and supporters for helping me throughout my training and development at UCSF. I would not have succeeded in completing my PhD without their assistance. I am also grateful for all of the thankless work that so many people have done to help ensure my success. I want to call attention to those vital, unglamorous efforts which often make all the difference without ever getting any of the credit.

Everyone will warn you about working with joint thesis advisors while pursuing your PhD. Despite that, I eagerly selected Barbara Panning and Sy Redding as my co-PIs. I am very glad that I had both of them supporting me throughout my time in graduate school. Not everyone can trust that their supervisors are truly looking out for their best interests, but I never had to question that with Barbara and Sy. I am grateful for all the time they dedicated to my personal growth and the opportunities they tirelessly tried to provide for me.

Additionally, I decided that two PIs just was not enough for me, so I selected three more to serve on my thesis committee. I need to thank Jeremy Reiter, Danica Fujimori, and Geeta Narlikar for their support over the years, but especially when I was a young graduate student. Danica kindly made it possible for me to move to San Francisco early and rotate in her lab before my program had even started, because I was far too earnest and three rotations would have never been enough for me. I need to thank my rotation mentor from the Fujimori lab, Lindsey Pack, for all of her patience and guidance during that time. Lindsey's support really helped me successfully transition to life in grad school and research at this level.

lucky to have Joel Hrit as another fantastic rotation mentor who helped me grow both as a scientist and as a person. I will always fondly remember our beer tasting adventures. Leeanne Goodrich and Betsy Martin also became fierce advocates for me, and really shaped my growth as a scientist in the lab and as an individual outside of the lab.

Since I became a member of the Panning lab, we were lucky to have some fantastic scientists join our team as well. Myles Hochman was a talented scientist and a great roommate. Eric Simental (another roommate of mine) and Amanda Chung have each worked hard to develop new systems in the Panning lab, and have been supportive friends both in and out of the lab. The most recent member to join the Panning lab, Katie Augspurger, has supported me through so many obstacles in recent years. I would not have been able to finish my PhD following the beginning of the pandemic if it were not for Katie's constant and unwavering support. Katie is a fantastic colleague who reminded me what it is like to work on a team. Additionally, none of us would get anything done if it were not for Christine Wong working to make sure that the Panning lab (as well as the entire neighborhood) has all the reagents and materials needed to do science. Additionally, I need to thank Dale Talbot for always being willing to quietly pick up the slack to keep things running in the Panning lab.

I need to thank Barbara for giving me the opportunity to work with Jesse Hurst in the Panning lab. Jesse was an undergraduate volunteer who commuted to UCSF from UC Berkeley to help me with my experiments in the lab. I need to thank Jesse for her dedication and enthusiasm to work with me on this project. It was an important mentorship opportunity that helped me grow as a scientist and simultaneously pushed my project forward. Additionally, I need to thank Sy for allowing me to work with Olga

work done. I am happy to see them all moving on to better things, and I am eternally grateful that I was still able to rely on their support ever after they left the Redding lab.

Discussion of the community in the Redding lab would not be complete without acknowledging our friends in the Al-Sady lab. Our labs met routinely for years, and the community we built together helped ensure our success. I need to thank Bassem for being a mini-third PI for me. I need to thank Rachel Greenstein and Nick Sanchez for their leadership as well as their companionship. When Eric Simental, Dana Kennedy, and Henry Ng all joined the Al-Sady lab, they reinvigorated all of us with their potential and enthusiasm. My training would not have been nearly as thorough or rewarding if it were not for working so closely with members of the Al-Sady lab for years.

Additionally, I had the opportunity to work with amazing scientists from beyond UCSF. I need to thank Sy for the opportunity to collaborate with Liam Holt and Srinjoy Sil. Conducting experiments with Srinjoy was a special time in my PhD and working alongside him was a very refreshing experience that came at a time when I really needed it.

I also need to thank Barbara for ensuring that I had the opportunity to collaborate with Steve Jacobsen and members of his lab from UCLA. I am so glad that Brandon Boone and Lucia Ichino virtually approached me at the Cold Spring Harbor Meeting on Epigenetics and Chromatin in 2020. The opportunity to work with them came at a pivotal moment in my PhD where I really needed additional support and community. Collaborating with Lucia and Brandon was not only productive, but it was also a lot of fun. I am very fortunate that I had the opportunity to publish some of this work with them. In particular I want to thank Steve Jacobsen and the entire MBD team, including

Brandon Boone, Lucia Ichino, Suhua Feng, Shuya Wang, and Ranjith Pappareddy, for trusting me to work with their proteins and helping me with my own experiments. Some of the data from this collaboration included in this thesis has already been previously published in the report "MBD5 and MBD6 couple DNA methylation to gene silencing through the J-domain protein SILENZIO." *Science* 372. 6549 (2021): 1434-1439.

Lastly, I need to thank my friends and family. I feel very lucky that this list overlaps with many of the people I have already acknowledged. First, I need to thank Ramesh Raina for encouraging me to pursue research as an undergraduate and giving me the opportunity to work in his lab. There I worked alongside Snigdha Chatterjee, who supported me throughout my entire time in graduate school. I am very happy that she followed me to the Bay Area for her own PhD, which she completed significantly faster than me. Additionally, as undergraduates, Snigdha and I worked closely with our postdoc mentor, Pratibha Choudhary. I need to thank Pratibha for teaching me the importance of being a genuine researcher.

After moving to San Francisco for my PhD, I have been lucky enough to find myself surrounded by both new and old friends. I am lucky to have friends from my past lives, like Kristine Swarts and Maya Kramer, both in the Bay Area, even if I do not see them as much as I should. I am thrilled to live in a place where I am surrounded by a community of some of my closest friends, including Ady Steinbach, Donovan Trinidad, Chris Carlson, Varun Bhadkamkar, Elise Muñoz, Francesca Del Frate, and Haley Gause. It has been lovely to go out more often with friends like Jonathan Kupfer, Andrew Santore, Eddie Carrillo, and Adam Longhurst. And it has been just as much fun to stay in and play games with friends like Kyle Lopez, Liz Bond, Ben Herken, and

Mitchell Lopez. I am so thankful for the support I have received from these friends and, many more over the years, to help me achieve a better work-life balance than when I started graduate school early for an unrequired summer rotation.

I also have to thank the friends that support me from afar. I am lucky to still have so many supportive friends that I get to see whenever I go home to Pennsylvania. I need to thank Chris Wilbur, Sarah Hosie, Emily Gillette, Michael Santarsiero, Tori DeGiosio, Brooke Walker, Nicole Zientek MacDonald, and Zach MacDonald for their never-ending support. Additionally, I am lucky to have the love and support of friends across the country including Lillian Byl, Sylvie Alusitz, Lynda Brady, and a slew of old Quidditch friends who have all supported me through the years of my PhD.

Lastly, I would like to thank my family who constantly reminds me that I am loved even though I left them to move across the country to pursue this degree and sometimes only get to see them once a year. I am lucky to have a loving family who support me no matter what I do, including my brother, my dad, my mom, my godmother Aunt Ellen, my Aunt Paula and Uncle Gary, my Aunt Mare, and my Uncle Mark and Grandpa Bob. I am grateful for all the ways our relationships have grown stronger over the years. Being away from my family for so long has made me appreciate them even more, and I need to thank them for continuing to love and support me despite the distance. I am lucky enough to know that they are proud of me, and I am proud of myself for all the work that I dedicated to this project.

Building a novel single-molecule system to study readers of DNA methylation with high resolution reveals binding preferences of MBD proteins

Luke Strauskulage

Abstract

DNA methylation is a prominent epigenetic modification modulating proper access to and storage of genomic information in many eukaryotic cells. The distribution of methylated CpG dinucleotides (mCGs) throughout the genome is dynamic and proper maintenance is required throughout the life of an organism. The density of DNA methylation is often correlated with transcriptional outputs, but the underlying mechanism of how mCG density is readout by the cell is not fully understood. One feature of densely methylated regions of the genome is there ability to recruit reader proteins containing Methyl-CpG Binding Domains (MBDs), which help regulate gene expression. Here we study MBD binding to densely methylated DNA using a single molecule fluorescent imaging system called DNA curtains. These techniques are inherently limited by the diffraction of light, complicating our ability to map MBD-binding events to local regions of DNA methylation. In this study, we developed a high resolution system using dCas9 technology to map binding events on DNA curtains with high accuracy. Our novel single molecule approach confirmed that MBDs selectively bind densely methylated DNA and revealed special features within these regions that may be important to MBD recruitment throughout the genome.

Table of Contents

List of Figures

List of Tables

1. Introduction

Epigenetic modifications act as genomic context clues to facilitate temporal and spatial regulation of gene expression. From plants to mammals, DNA methylation is an important epigenetic signal in many eukaryotic genomes (1,2). In mammals, 5-methylcytosine is the predominant DNA modification and primarily occurs at CpG dinucleotides (3,4). Highlighting the importance of this modification, the distribution of methylated CpGs (mCGs) across the genome is tightly regulated. While CpGs are under-represented throughout mammalian genomes, they frequently occur in dense clusters called CpG islands, which are often found in cis-regulatory regions (5,6). Methylation of CpG islands is dynamic and dense methylation at promoters often correlates with transcriptional repression (7).

To control transcription, CpG methylation within sequence-specific binding sites can block recognition by transcription factors (8,9). However, in many CpG islands, methylation of specific CpGs is not required for regulation; instead, the overall methylation density correlates with transcriptional status (10). This lack of site specificity raises the question of how mCG density is readout to control transcription.

In addition to inhibition of transcription factor binding, mCGs can act as a platform which is recognized and specifically bound by mCG reader proteins. Methyl-CpG Binding Domains (MBDs) recruit MBD-family proteins to sites of DNA methylation (11,12). Like DNA methylation, MBDs are conserved across many eukaryotic genomes, highlighting their importance in coordinating the proper readout of DNA methylation (13,14,15). MBD-family proteins are often members of large chromatin regulatory complexes, which frequently inhibit transcription (16).

1

Like the underlying distribution of mCGs, the distribution of MBDs throughout the genome is highly specific (11,17). MBDs must identify and stably bind the correct regulatory mCGs in a genome containing many off-target binding sites. Additionally, plant MBDs have the added difficulty of differentiating between potential 5-methylcytosine binding sites which occur in a variety of sequence contexts (1,14,15,18). In order to properly bind mCGs throughout the genome, a common feature shared by many MBDs is that they are more frequently found at densely methylated regions in the genome (17). Despite this conserved binding behavior, *in vitro* different MBDs have the ability to differentially bind the same methylated DNA substrate (19). This distinct recognition highlights that mCGs provide contextual information which must be integrated by MBDs to orchestrate the correct genomic output.

Unpacking the dynamics of MBD binding across a complex distribution of many methylated sites is challenging. Single molecule studies offer the ability to monitor individual proteins interacting with DNA in real-time (20). This advantage makes them powerful approaches to characterize and understand the biophysical basis of MBD binding. These methods have previously identified novel behaviors of MBDs on DNA, including sliding, bending, and looping (21,22). We previously employed a single molecule approach called DNA curtains to study MBDs. DNA curtains utilize fluorescent microscopy to analyze many individual DNA binding events along kilobase pairs of flow-stretched DNA (23,24,25). We used DNA curtains to characterize the binding preference of *Arabidopsis thaliana* MBD6 (AtMBD6) on methylated DNA curtains. Our analysis identified a correlation between the distribution of mCGs and AtMBD6 binding, verifying that AtMBD6 is a bona fide mCG binding protein in plants (26).

Traditionally, many single molecule studies relying on fluorescent microscopy, including DNA curtains, often sacrifice resolution in order to directly observe DNA-protein interactions in real-time (27). This limited resolution makes it difficult to study how cues, like mCG density or DNA sequence, affect MBD binding. In order to better understand how these factors can influence MBD binding, we aimed to develop a single molecule system with higher resolution.

In this report, we develop a system to increase the resolution of DNA curtains and use it to study MBD binding on a complex DNA substrate containing thousands of mCGs. Fluorescent imaging of MBDs on DNA curtains revealed that they generally prefer binding to densely methylated DNA. However, further quantifying this recruitment was difficult at low resolution. Utilizing dCas9 technology, we generate a high resolution method to map binding events with high accuracy, which can be utilized to study any DNA-protein interaction. Using this method, we validate that MBD recruitment to DNA is more complicated than the presence or absence of DNA methylation. With increased accuracy, we were able to measure MBD preferentially binding to regions of densely methylated DNA. Within these regions, we find that MBDs have an intrinsic preference for sites that fall approximately one helical turn away from neighboring mCGs, and we show that this is a conserved preference across multiple MBDs. Despite this similarity, our analysis reveals that different MBDs are encoded to create unique binding distributions on the same underlying methylated DNA sequence.

2. Results

2.1 Visualization of MBD binding on densely methylated DNA

MBD family proteins must differentiate between specific regulatory mCGs competing for recognition with many off-target sites of DNA methylation. To better understand MBD discrimination of methylated DNA, we investigated the binding of MBDs on DNA curtains. DNA curtains are a high-throughput, single molecule assay for studying protein-DNA interactions (23,24,25). In DNA curtain assays, biotinylated-DNA is tethered to a fluid lipid bilayer using streptavidin. The captured DNA molecules remain mobile and are aligned at microfabricated chrome barriers within a microfluidic flowcell (**Figure 1**). The DNA in a curtain are extended in solution by buffer flow, visualized with an intercalating dye called YOYO-1, and imaged using TIRF microscopy. Because all of the DNA are tethered in the same, specific orientation, positions along the DNA in a curtain directly correspond to sequence.
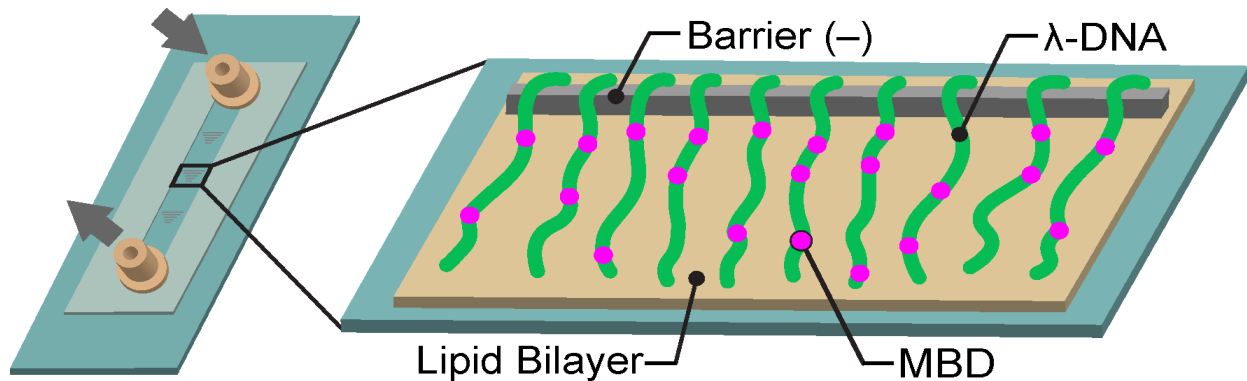
Figure 1. DNA curtains setup. This illustration depicts the setup of a DNA curtains experiment. **Left**, the cartoon indicates how a microfluidic flowcell is created on a microscope slide. **Right**, the cartoon shows how DNA molecules (green) attached to a fluid lipid bilayer are aligned at nanofabricated chrome barriers for imaging. Fluorescent proteins (magenta) can then be imaged interacting with DNA in real-time: in this case, MBDs interacting with methylated DNA.

Unlike many traditional biochemical assays, which often rely on short DNA molecules to study DNA binding proteins *in vitro*, DNA curtains are amenable to using relatively long DNA substrates, typically DNA from bacteriophage λ (λ-DNA). Across the nearly 50 kilobase pair λ-phage genome, 3113 CpGs occur in a variety of local sequence contexts, offering a uniquely complex distribution of potential MBD binding sites. We aimed to methylate all potential mCG sites using the CpG-specific methyltransferase M.SssI. First, we validated that all CpGs were methylated by checking for protection from BstUI, a restriction enzyme blocked by mCGs. We were unable to detect any BstUI activity on M.SssI-methylated λ-DNA, while untreated λ-DNA was readily digested (**Figure 2A**). Second, we performed Enzymatic Methyl-sequencing (EM-seq) (28), which revealed an average methylation probability of 97.6% across all CpG sites (**Figure 2B**). Throughout this report, we use the mCG density as determined by EM-seq as a benchmark for M.SssI DNA methylation.
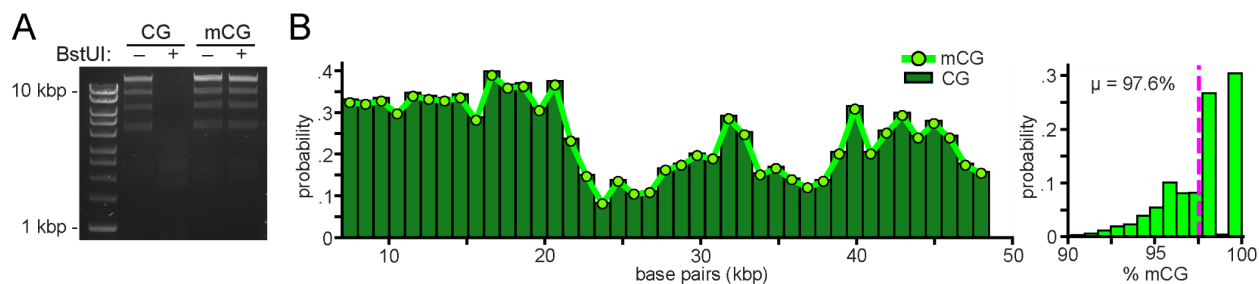
Figure 2. Validation of λ-DNA methylation by M.SssI. (**A**) λ-DNA was incubated with mCG-sensitive restriction enzyme BstUI before and after treatment with DNA methyltransferase M.SssI. Samples were then run out in a 0.67% agarose gel and visualized with SYBR-safe. (**B**) **Left**, the dark green histogram shows the distribution of CpG dinucleotides across the length of the λ-genome. The light green line indicates the distribution of methylated CpGs after treatment with M.SssI as quantified by EM-seq. **Right**, the light green histogram shows the methylation percentage of all CpG dinucleotides in the λ-genome as measured by EM-seq. The average methylation percentage for all CpGs is 97.6%.

We purified the MBD domain of mouse MBD1 with an N-terminal 6xHis-tag (aa 2-75; MBD1-MBD). Constructs were designed to express MBDs with and without an additional C-terminal LPETG motif + Strep-tag for sortase labeling MBDs to visualize their binding on DNA curtains (29). Sortase is a transpeptidase from bacteria, used to cleave target proteins in a site-specific manner. At the cleavage site, the sortase enzyme can be used to covalently attach new molecules, like fluorescent dyes, to the target protein via formation of a new peptide bond (**Figure 3A**). After purifying recombinant sortase and optimizing an efficient labeling strategy (**Figure 3B**), we found this approach and the subsequent clean-up to be unwieldy and unnecessarily time-consuming for labeling and visualizing multiple MBDs. This led us to seek more efficient labeling strategies, described below.

To validate specific binding of these 6xHis-tagged MBD1-MBD constructs to methylated DNA, we first performed classic *in vitro* binding assays: electrophoretic mobility shift assays (EMSAs) or gel shifts. In EMSAs, protein-bound DNA is visualized by its impaired migration during gel electrophoresis (30). Short DNA molecules are often used as substrates, and while studying proteins that specifically recognize mCGs, these DNAs often contain only a single symmetrically methylated CpG dinucleotide (18,32). While EMSAs have been pivotal for both developing and enhancing our understanding of MBDs and many other DNA-binding proteins, gel shifts are notoriously finicky assays. The ability to achieve a successful gel shift is not only dependent on the strength and stability of a particular DNA-protein interaction, but also the specific conditions used to capture and visualize binding.
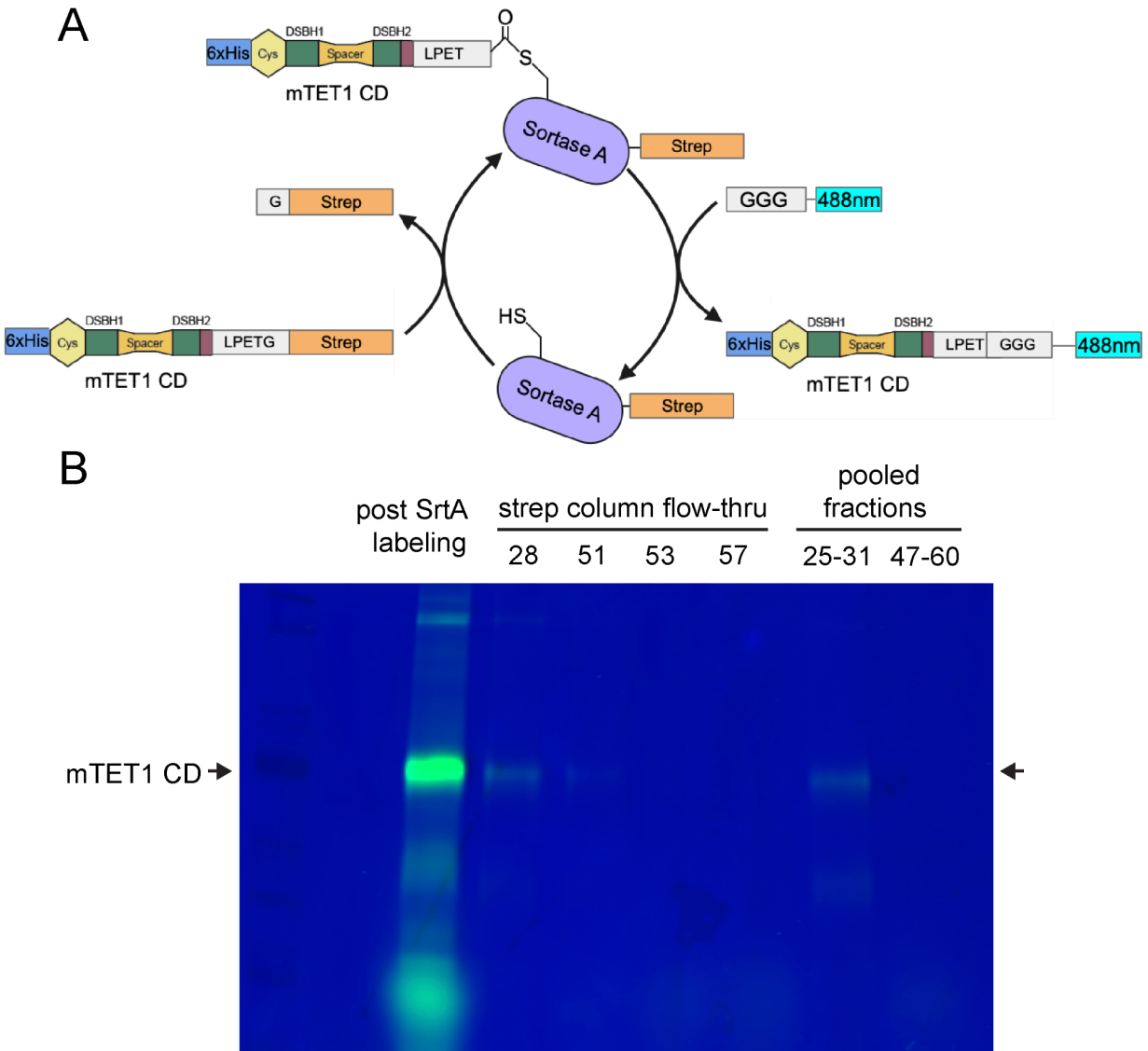
Figure 3. Overview of Sortase labeling. (**A**) Schematic overview of Sortase A (SrtA) labeling strategy, using the catalytic domain from mouse TET1 (mTET1 CD, ~75kDa) as an example. SrtA will first cleave target proteins at the LPETG motif, releasing the C-terminal Strep-tag. Fluorescently labeled peptides can then be covalently attached to the C-terminal end of target proteins. (**B**) SDS-PAGE gel depicting successful sortase labeling of mTET1 CD using a molar ratio of **0.667** SrtA : **1** mTET1 : **13.33** peptide. The overnight labeling reaction was then cleaned up over a Strep column, and subsequent lanes show which fractions contained dilute sortase-labeled mTET1 CD. Pooled fractions containing purified sortase-labeled protein were concentrated and stored for later use.

Historically, EMSAs have provided novel insights into MBD binding, but these experiments often contained a large unshifted population of methylated DNA (14, 31, 32). This trend suggests that MBDs can be particularly sensitive to binding conditions and may even indicate that MBD binding to certain methylated DNAs can be inherently unstable. Initially, we struggled to visualize specific binding of 6xHis-tagged MBD1-MBD to a short DNA molecule containing a single mCG (**Figure 4**). This result could indicate that this MBD1 construct has impaired binding to methylated DNA or that we are using suboptimal binding conditions.
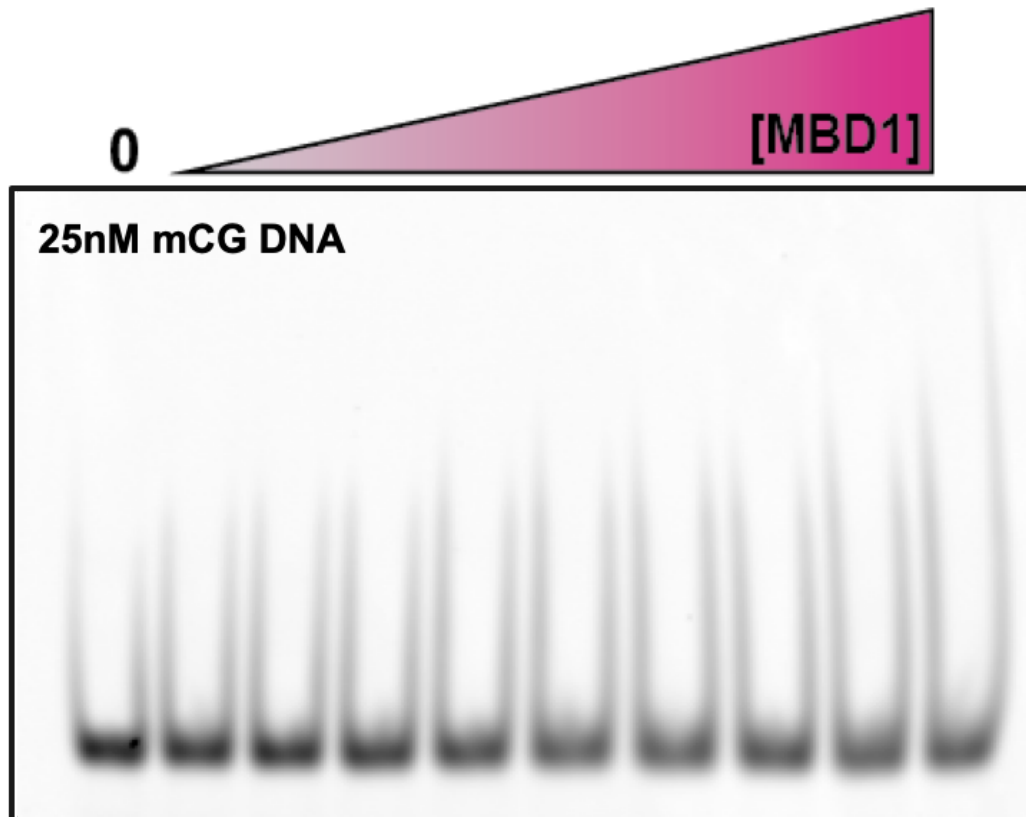
Figure 4. MBD1-MBD gel shift with a single mCG. Image depicting gel shift of DNA containing a single mCG after incubation with MBD1-MBD. FAM-labeled DNAs were 32 base pairs long with a central, symmetrically methylated CpG. Increasing amounts of MBD1-MBD were added up to a molar excess of 1:400. At MBD1-MBD concentrations above this, we began to observe non-specific binding to unmethylated DNA under these conditions.

However, we also failed to observe specific binding to DNA containing only a single mCG when using fluorescence polarization, another *in vitro* binding assay (33). The lack of bound DNA molecules in these assays may actually indicate that robust MBD recruitment to DNA is not solely predetermined by the mere presence of a potential mCG binding site.

Since it is well established that MBD binding correlates with dense clusters of mCGs (17), we further tested 6xHis-tagged MBD1-MBD binding to a longer DNA substrate with multiple mCGs. We prepared a 215 base pair DNA fragment amplified from the multi-cloning site of pUC19 (pUC-MCS). This relatively long DNA molecule contains 14 CpGs, more closely resembling densely methylated regions that MBDs occupy *in vivo*. We methylated all 14 CpGs using M.SssI, and validated complete methylation by protection from digestion by restriction enzyme, SmaI.

In an EMSA, we could observe specific binding of 6xHis-tagged MBD1-MBD to methylated pUC-MCS (**Figure 5**). At low concentrations, we only observed binding of MBD1-MBD to methylated pUC-MCS, confirming that this MBD1-MBD construct is capable of specifically binding mCGs. As expected for DNA binding proteins, we began to observe non-specific binding to unmethylated pUC-MCS at higher concentrations of MBD1-MBD. In both cases, MBD1-MBD binding to DNA appeared as smears rather than discrete band shifts, which is most likely caused by DNA-protein complex falling apart while migrating through the gel. This instability can be a result of our specific assay conditions, or it can indicate that transient, dynamic binding may be a feature of MBDs interacting with this kind of DNA substrate. Regardless, this gel shift validates that recombinant 6xHis-tagged MBD1-MBD maintains its specificity for methylated DNA.
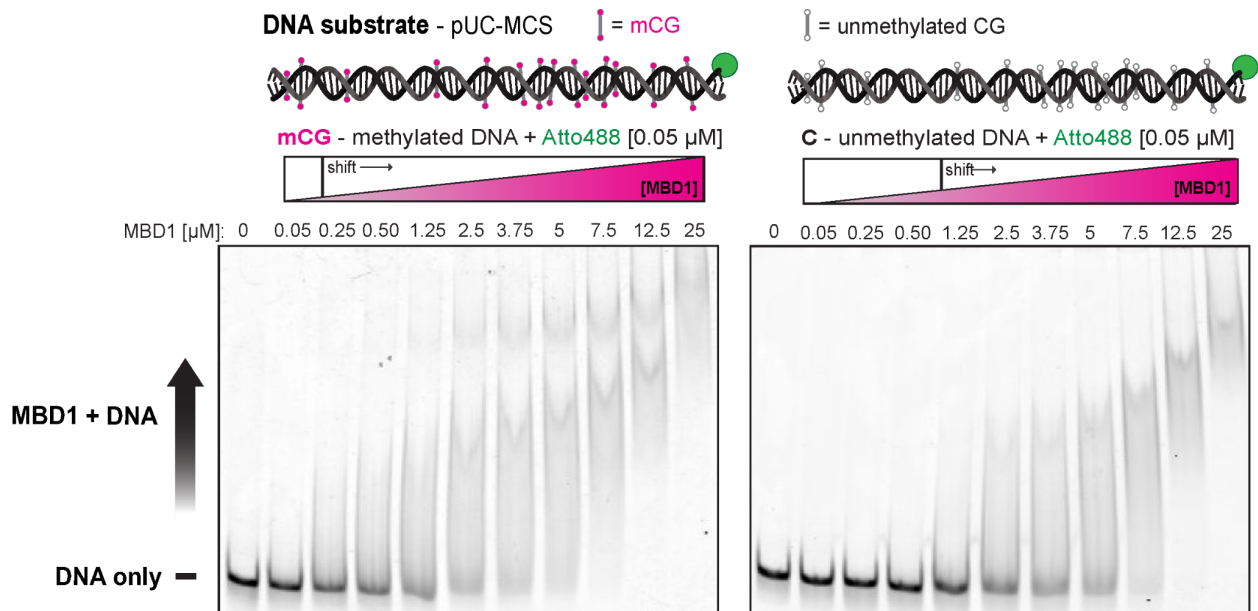
Figure 5. MBD1-MBD gel shift with multiple mCGs. Images depicting gel shift of DNA containing multiple potential binding sites of DNA methylation after incubation with MBD1-MBD. Increasing amounts of MBD1-MBD were added up to 1:500 molar excess. Atto488-labeled DNAs were amplified from pUC19 and contained 14 CpGs which were either left unmethylated or methylated by M.SssI. Gels were imaged using a Typhoon laser scanner.

Rather than further optimizing binding conditions for gel shifts, we moved forward with testing MBD1-MBD binding on DNA curtains, which would allow us to further probe how MBDs interact with many potential mCG binding sites. Unlike EMSAs, where we are only able to directly observe aggregate binding to relatively short DNAs with a limited number of mCGs, DNA curtains allow us to simultaneously image MBD occupancy at many sites along a distribution containing thousands of mCGs.

To visualize 6xHis-tagged MBD1-MBD binding to methylated DNA curtains, we used commercially available fluorescent antibodies. While this is not a direct labeling strategy, we found that using fluorescent antibodies was far more practical than sortase labeling and allowed us to efficiently test multiple MBDs and binding conditions. While the standard quality control assays for these antibodies (immunofluorescence, western

blots, etc.) do not guarantee that they will perform well in a single molecule experiment, we found that anti-6xHis-Alexa555 (clone 4E3D10H2/E3, Invitrogen: MA1-135-A555) reliably labeled all 6xHis-tagged proteins we tested on DNA curtains.

Despite struggling to capture stable binding events in a gel shift, 6xHis-tagged MBD1-MBD readily and stably bound methylated DNA curtains when injected into the flowcell (**Figure 6**). Under the same conditions, we were unable to detect any MBD1-MBD binding to unmethylated λ-DNA. The absence of off-target binding events indicates that 6xHis-tagged MBD1-MBD binding in solution to DNA curtains is highly specific to mCGs. Moving forward, we relied on DNA curtains to test recombinant MBD constructs for specific binding to methylated DNA.
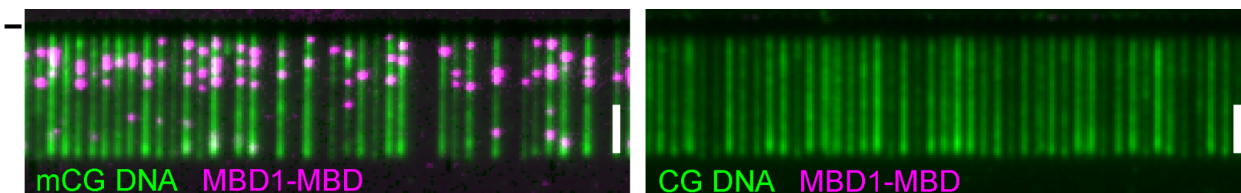
Figure 6. MBD1-MBD binding to methylated DNA curtains. Representative images depicting YOYO-1 stained λ-DNA (green) bound by MBD1-MBD (magenta). **Left**, image shows stable MBD1-MBD binding events on λ-DNA after methylation by M.SssI. **Right**, image shows that no non-specific binding of MBD1-MBD can be seen on unmethylated λ-DNA. (—) indicates chrome diffusion barriers. Scale bar - 5 µm.

To uncover conserved cues that may generally govern DNA binding of multiple MBDs rather than any specific family member, we tested several different MBDs on methylated DNA curtains. In addition to mouse MBD1-MBD, we also purified N-terminal 6xHis-tagged MBDs from mouse MeCP2 (aa 90-162; MeCP2-MBD) as well as mouse MBD3 (aa 1-69; MBD3-MBD). We utilized the 6xHis-tag and anti-6xHis-Alexa555 to visualize these additional mouse MBDs. Like MBD1, MeCP2 is a mammalian MBD family member known to bind methylated DNA (11,34). However, multiple studies have

also reported that MeCP2 is capable of more promiscuous DNA binding beyond mCGs (35, 36). Similarly, MBD3 has a controversial reputation among MBD family members. Unlike many of the other mammalian MBDs, MBD3 contains a mutation in the conserved MBD, which has been shown to interrupt mCG recognition (37). Despite this finding, multiple other studies have reported conflicting binding activity of MBD3 to differentially modified DNA (38, 39).

Beyond multiple mammalian MBDs from the same species, we also studied a more distantly related MBD from plants on DNA curtains. We tested the MBD from *Arabidopsis thaliana* MBD6 (11-81aa, AtMBD6-MBD) purified with an N-terminal MBP-tag and a C-terminal 6xHis-tag. AtMBD6-MBD was directly labeled using Cy3-NHS-ester fluorescent dyes to visualize binding on DNA curtains. In plant genomes, multiple forms of DNA methylation are much more prevalent than in mammalian genomes. Thus, plant MBDs must differentiate between canonical mCG binding sites, as well as prevalent mCHGs and mCHHs (where H ≠ G) (1,14,15,18). We can partially test AtMBD6-MBD specificity for mCGs since λ-DNA contains an underlying level of bacterial DNA methylation. Previous studies have shown that Dam and Dcm DNA methyltransferases modify λ-DNA before it is packaged in the bacteriophage capsid (40). Dcm sites (5'-CC(A/T)GG-3') are modified at the C5 position on the second cytosine, creating multiple sites of CHG methylation along λ-DNA, which could serve as potential binding sites for AtMBD6-MBD.

We found that all MBDs we tested on DNA curtains bound methylated DNA. Like mouse MBD1-MBD, both mouse MeCP2-MBD and MBD3-MBD, as well as AtMBD6-MBD, were capable of generating stable, long-lived binding events on

methylated DNA curtains (**Figure 7**). When tested on unmethylated DNA curtains, maintaining the same underlying distribution of bacterial DNA methylation, we could not observe any non-specific binding of mouse MeCP2-MBD or AtMBD6-MBD. This finding indicates that binding of both mouse MBD1-MBD and MeCP2-MBD is highly specific to mCGs at the single molecule level on DNA curtains. Additionally, AtMBD6-MBD also specifically and stably bound mCGs on DNA curtains, further confirming that it may recognize these sites over other forms of DNA methylation in plants.
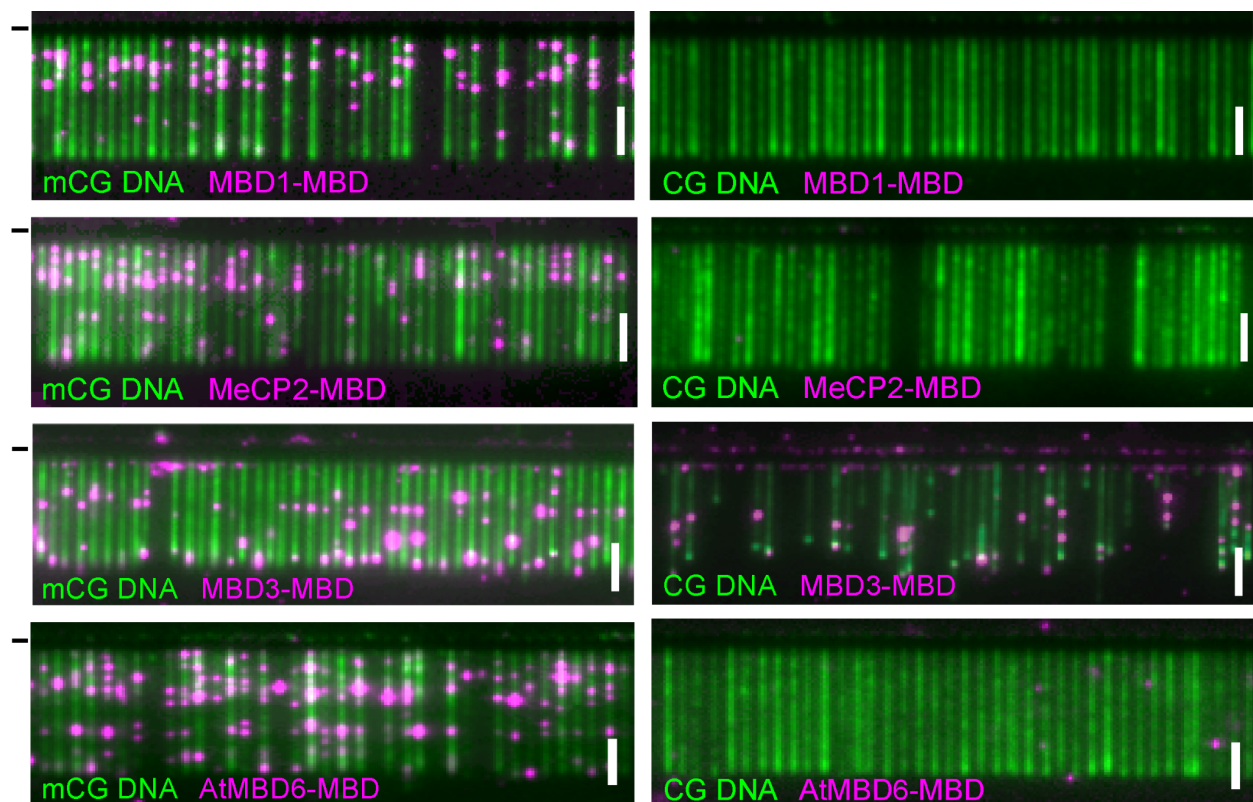
Figure 7. Variable MBD domains binding to methylated DNA curtains. Representative images depicting YOYO-1 stained λ-DNA (green) bound by a variety of MBDs (magenta). **Left**, image shows stable MBD binding events on λ-DNA after methylation by M.SssI. **Right**, images show whether non-specific binding of MBDs can be observed on unmethylated λ-DNA curtains. (—) indicates chrome diffusion barriers. Scale bar - 5 μm.

However, we did observe binding of mouse MBD3-MBD to unmethylated DNA curtains, which could suggest that the mutation in mouse MBD3-MBD does impact its ability to specifically bind methylated DNA. The binding events of MBD3-MBD were largely localized to the free end of the DNA, and even appeared to compact λ-DNA molecules. This behavior suggests that many of the observed MBD3-MBD binding events may be non-specific. In the future, more work will need to be done to test whether this potential activity is reproducible and further characterize MBD3-MBD binding at the single molecule level. Moving forward in this study, we focused on further investigating the binding behavior of MBDs which specifically recognize mCGs.

To begin analyzing the behavior of MBDs, we mapped each binding event back to a position on methylated DNA curtains. First, we align the two fluorescent channels from our microscope, described in detail below (see section **2.3 Sources of error on DNA curtains**). Next, we developed automated software to segment the fluorescent signal of the YOYO-1 stained DNA and return its coordinates in the microscope image. We used those positions to identify which fluorescent MBDs are bound to that DNA molecule. Since the tethered end of the λ-DNA is obscured by the chrome barrier, we measure the relative position of each MBD from the free-end of the DNA. Then, assuming a uniform extension of the DNA, we converted this pixel distance to a base pair position (**Figure 8**). This approach allows us to automatically analyze the position of thousands of MBD binding events along methylated λ-DNA.
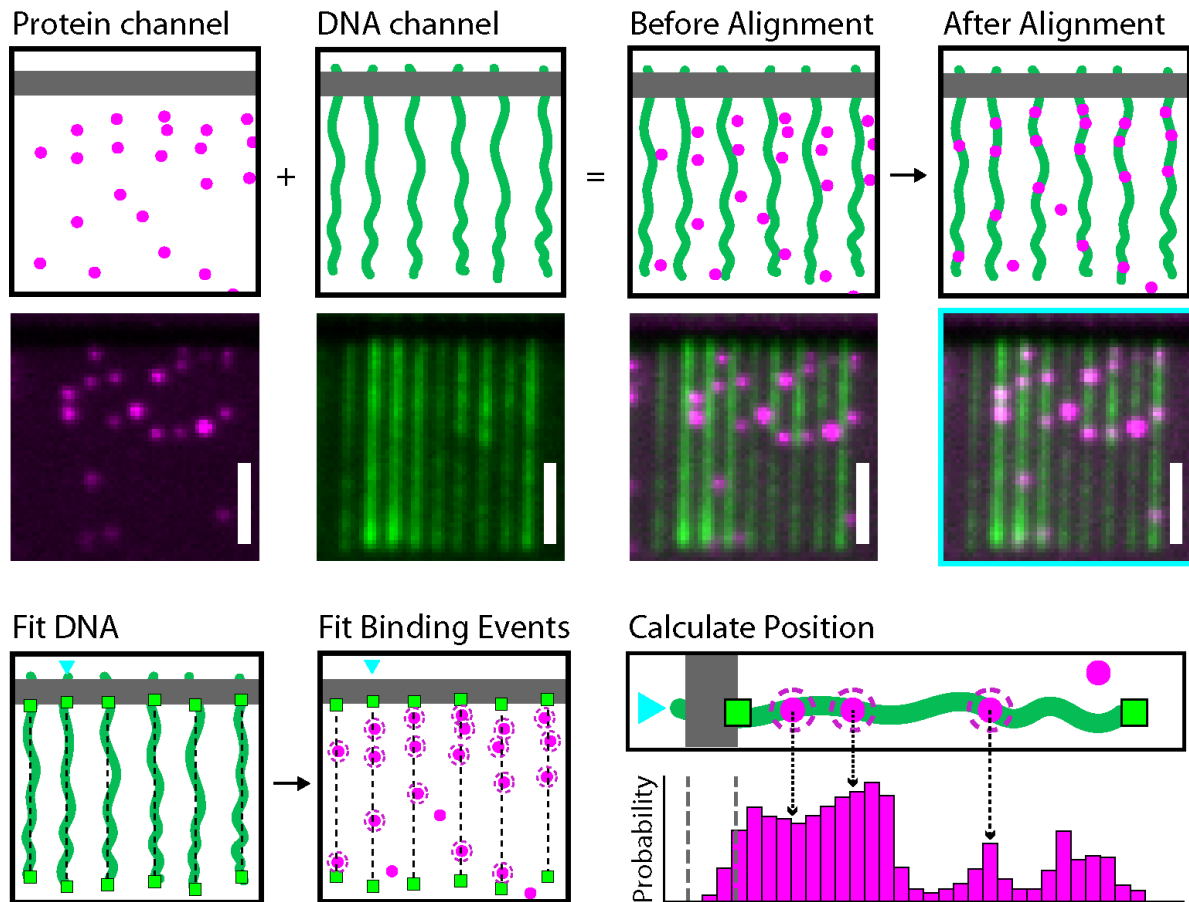
Figure 8. Overview of baseline image analysis. Diagram showing how images from two fluorescent channels are overlaid with one another and analyzed to calculate the position of DNA binding events. Briefly, custom software is used to automatically identify the ends of YOYO-1 stained DNA molecules. These bounds are used to determine which fluorescent proteins are likely bound to the DNA. Additionally, the position of the top and bottom of the DNA can be used to determine the average number of DNA base pairs per pixel. A linear estimate of the extension is then used to calculate the position of binding events based on their distance from the free-end of the DNA.

Overall, measured binding distributions of MBDs which specifically bound methylated DNA (MBD1-MBD, MeCP2-MBD, and AtMBD6-MBD) were largely consistent with the underlying distribution of mCGs, as measured by EM-seq (**Figure 9A/B/D**). In the case of MBD3-MBD, the majority of analyzable binding events were localized at the free-end of the methylated DNA **(Figure 9C)**.
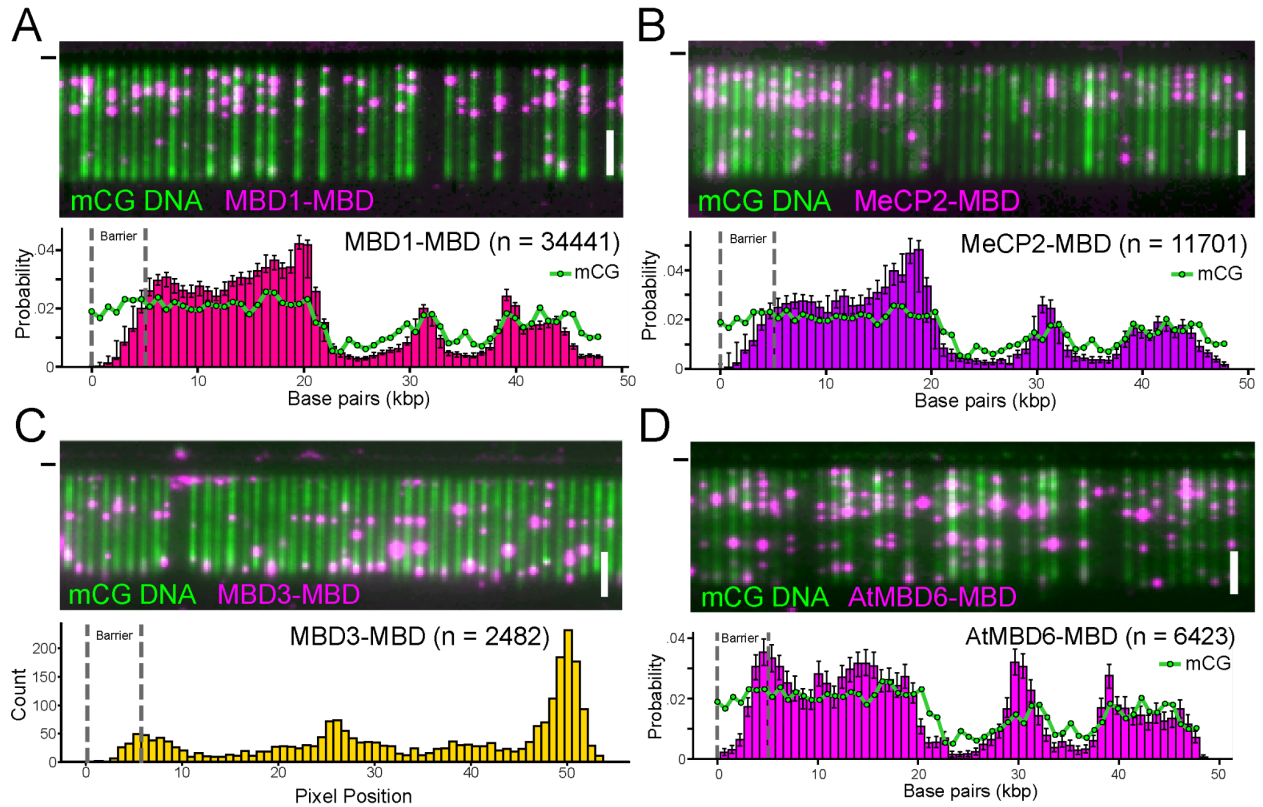
Figure 9. Variable MBD distributions on DNA curtains using baseline analysis. Representative images of variable fluorescent MBDs (magenta) binding to methylated DNA curtains (green). (—) indicates chrome diffusion barriers. Scale bar - 5 μm. Histograms show the measured binding distributions of MBDs based on their distance from the free-end of the DNA. Where applicable, the underlying distribution of mCGs is indicated by the green line. Vertical grey dotted lines indicate where DNA is obscured by barriers. (**A**) mouse MBD1-MBD, (**B**) mouse MeCP2-MBD, (**C**) mouse MBD3-MBD, and (**D**) AtMBD6-MBD.

To further validate using methylated DNA curtains as a platform for studying MBD binding, we found that MBD1-MBD and MeCP2-MBD binding distributions did not vary dramatically regardless of the concentration of the reader, the orientation of the λ-DNA, or the presence of non-CpG bacterial DNA methylation (**Figure 10**).
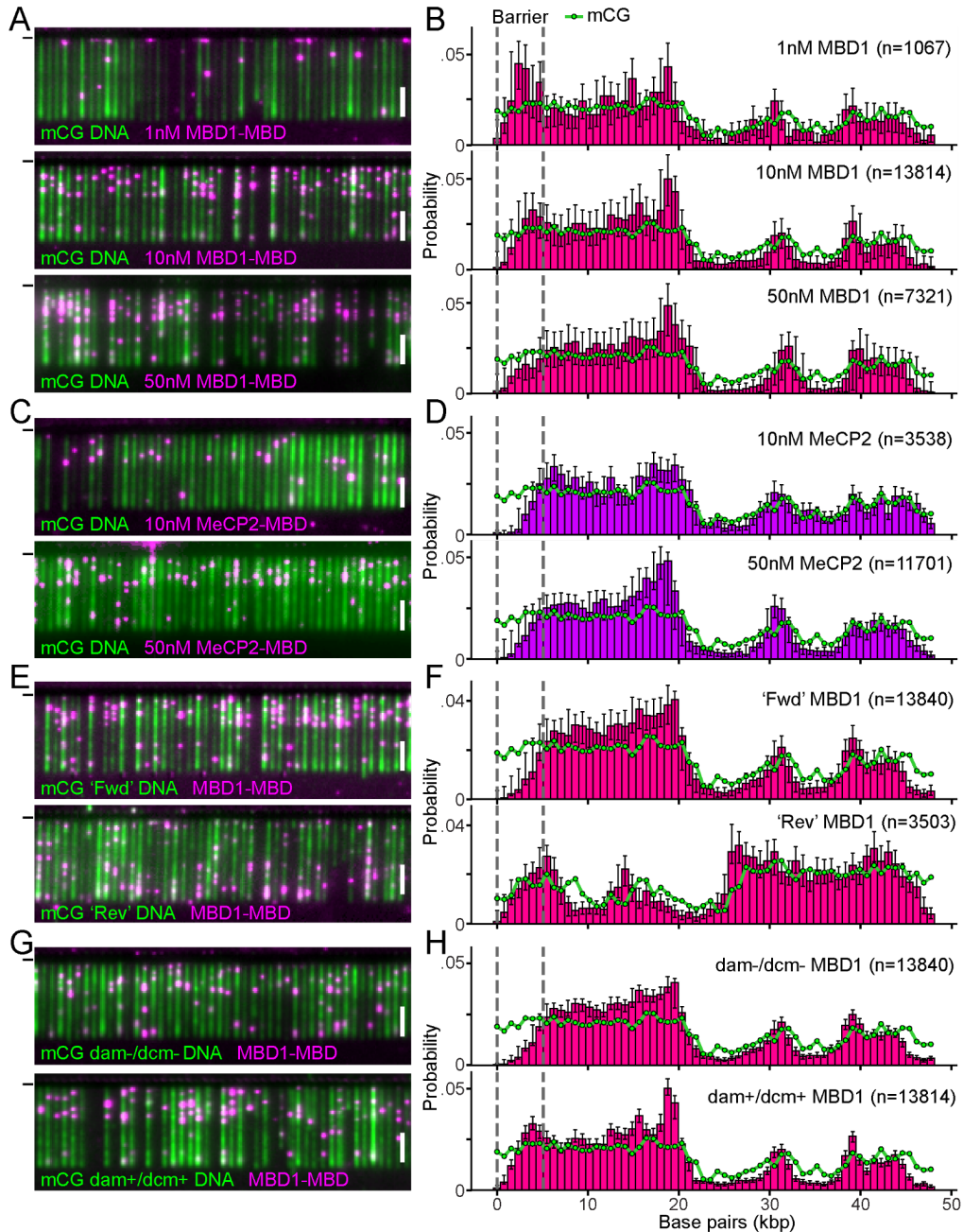
Figure 10. MBD1-MBD and MeCP2-MBD binding under variable conditions. Representative images of MBDs (magenta) bound to methylated λ-DNA (green). (─) indicates chrome diffusion barriers. Scale bar - 5 µm. Histograms indicate binding distributions measured from the free-end of the DNA plotted alongside the mCG distribution (green line). (**A**) representative images and (**B**) binding distributions of MBD1-MBD at variable concentrations. (**C**) representative images and (**D**) binding distributions of MeCP2-MBD at variable concentrations. (**E**) representative images and (**F**) binding distributions of MBD1-MBD on DNA curtains tethered from either end. (**G**) representative images and (**H**) binding distributions of MBD1-MBD on DNA curtains with and without underlying bacterial DNA methylation.

Similarly, we used AtMBD6-MBD to confirm that putative MBD binding events, imaged using TIRF microscopy, were truly bound to methylated DNA instead of nonspecifically stuck to the surface of the flowcell (**Figure 11**). When buffer flow was halted, MBD proteins left the field of view as tethered λ-DNA molecules retracted and began diffusing in solution. MBD binding events then returned to the field of view when buffer flow was resumed and λ-DNA molecules were stretched back over the barriers.
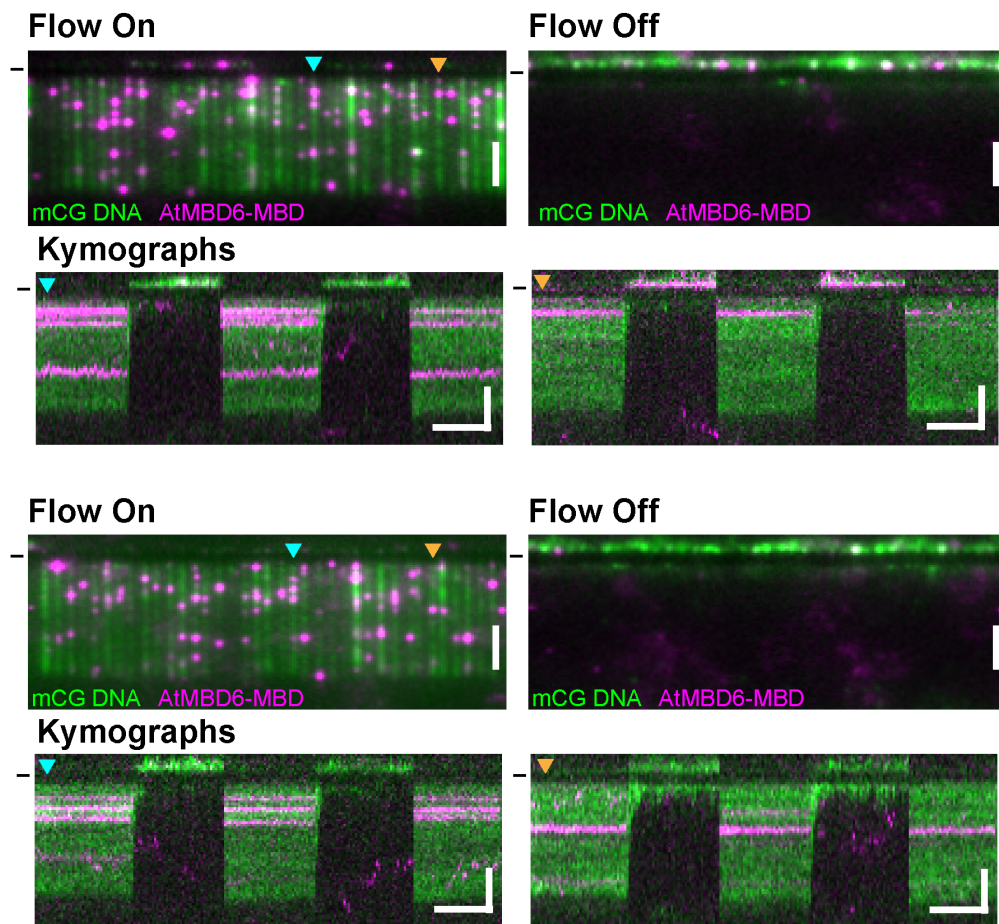
Figure 11. Kymographs of AtMBD6-MBD bound to methylated DNA curtains. Representative images show the same fields of view before and after halting buffer flow. AtMBD6-MBD binding events (magenta) only remain visible by TIRF in a flow dependent manner, indicating that they are localized to methylated DNA (green). Triangles indicate the DNA molecules represented by kymographs, where we can see the same slice over time. Vertical scale bars indicate 5 µm, and horizontal scale bars mark approximately 2.5 seconds of imaging. (─) indicates chrome diffusion barriers.

The measured binding distributions of MBD1-MBD, MeCP2-MBD, and AtMBD6-MBD all correlated with the underlying distribution of mCGs along the length of λ-DNA (Pearson's r > 0.7, **Figure 12A**). Despite this overall correlation between MBDs and mCGs, all measured binding distributions did vary noticeably from the underlying distribution of DNA methylation. This consistent skew left certain regions of mCGs under-represented or over-represented by MBD binding. Relatively dense regions of DNA methylation were more likely to be bound by MBD1-MBD, MeCP2-MBD, and AtMBD6-MBD (**Figure 9A/B/D**). Similarly, areas with relatively few mCGs were less likely to be bound by any MBD. When plotting the correlation between MBDs and mCGs, this consistent skew resulted in a steeper, exponential appearance (**Figure 12A**). Even though these MBDs come from different species, they all maintained a conserved preference to stably bind densely methylated regions of DNA methylation.
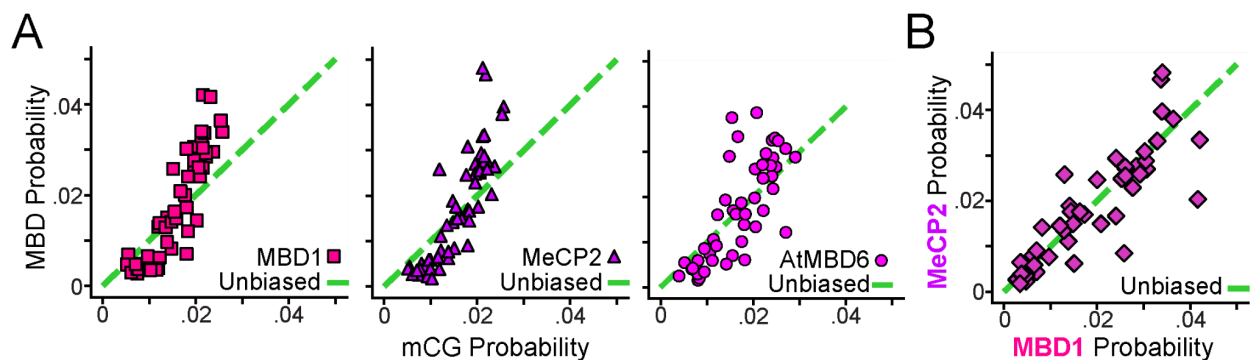
Figure 12. Correlation plots of MBD binding distributions on methylated λ-DNA. (**A**) Correlation plots of each MBD binding with the distribution of mCGs in λ-DNA. Green lines represent a linear trend, if all mCGs were seen equally by MBDs. (**B**) Correlation between mouse MBD1-MBD and mouse MeCP2-MBD binding along methylated λ-DNA. Green line represents a linear trend, if both MBDs bind to the same regions with the same frequency.

Next, we wanted to further investigate similarities and differences between different MBDs. While a few subtle differences existed, the binding of mouse MBD1-MBD and MeCP2-MBD largely correlated with one another at this resolution (**Figure 12B**). At low salt, mouse MBD1-MBD and AtMBD6-MBD had very similar binding distributions, suggesting that even distantly related MBDs may read the same distribution of DNA methylation similarly (**Figure 13A**). However, their binding distributions were markedly different from one another under high salt conditions (**Figure 13B**). Both mouse and plant MBDs maintained mCG-density dependent binding when imaging with high salt on DNA curtains. However, this mCG-density dependent binding manifested in different ways for each MBD.
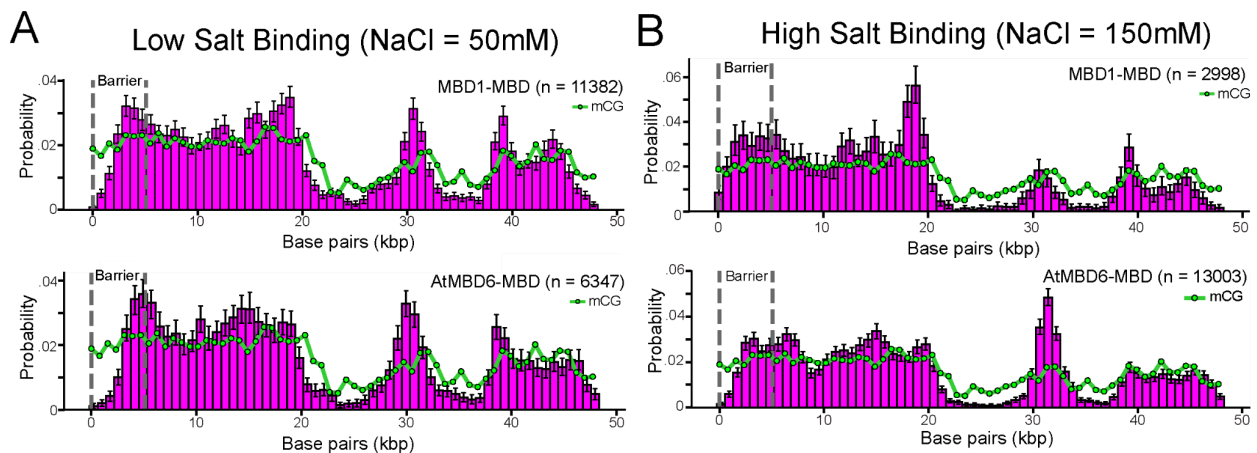
Figure 13. MBD1-MBD and AtMBD6-MBD binding distributions at low and high salt. (**A**) Binding distributions showing MBD1-MBD (top) and AtMBD6-MBD (bottom) enrichment across methylated DNA curtains when imaged under low salt conditions, NaCl=50mM. Histograms are created by measuring the position of binding events based on their distance from the free-end of the DNA. MBD binding probability is plotted alongside the underlying mCG distribution (green line). (**B**) Histograms showing how the binding distributions of MBD1-MBD (top) and AtMBD6-MBD (bottom) change under high salt conditions, NaCl=150mM. Binding distributions are measured from the free-end of the DNA plotted alongside the mCG distribution (green line).

MBD1-MBD enrichment was primarily conserved within a single area of the CG-rich region near the top end of λ-DNA. Meanwhile, AtMBD6-MBD binding was primarily localized in a dense region of DNA methylation in the center of λ-DNA, flanked by two AT-rich tracks. This differential binding reveals that differences within small, conserved MBD domains are significant enough to lead to novel enrichment at unique sites on the same underlying distribution of DNA methylation.

This binding trend, visualized on gel shifts and DNA curtains, further validates previous studies showing that MBDs frequently bind densely methylated clusters of mCGs. Additionally, our work shows that the MBD domain alone is sufficient to bind methylated DNA in a mCG-density dependent manner.

2.2 Regulation of mCG-density dependent binding by MBDs

In the cell, MBD domains do not exist as independent functional units. Instead, any underlying binding behavior of these domains must be orchestrated in the context of the full-length protein. Additionally, MBDs are frequently found as members of large nuclear complexes containing many proteins, and MBD binding must operate in the context of these binding partners (16). To explore how the mCG-density dependent binding of MBD domains could be regulated, we tested MBD binding in the context of larger protein constructs.

First, we tested binding of full-length *Arabidopsis thaliana* MBD6 (AtMBD6-FL) once again purified with an N-terminal MBP-tag and a C-terminal 6xHis-tag (26). To visualize binding of this construct to DNA curtains, AtMBD6-FL was directly labeled using Cy3-NHS-ester. Like AtMBD6-MBD, we observed stable binding of AtMBD6-FL on methylated DNA (**Figure 14A**).

In the context of the full-length protein, we still saw no off-target binding of AtMBD6-FL to unmethylated DNA, even in the presence of bacterial DNA methylation. Mapping the position of AtMBD6-FL binding events on methylated DNA revealed a binding distribution that closely followed with the underlying distribution of DNA methylation, even under high salt conditions (**Figure 14B**).
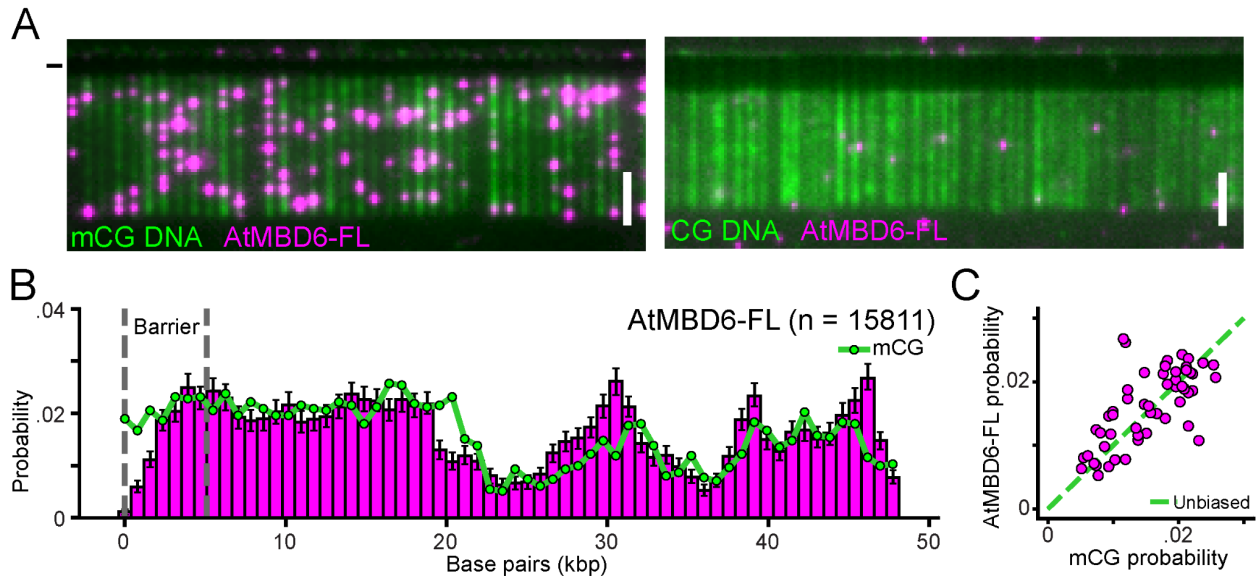
Figure 14. AtMBD6-FL binding to methylated DNA curtains. (**A**) Representative images of AtMBD6-FL binding events (magenta) on methylated and unmethylated DNA (green). (—) indicates chrome diffusion barriers. Scale bar - 5 μm. (**B**) Histogram showing the binding distribution of AtMBD6-FL binding as measured from the free-end of λ-DNA molecules plotted alongside the underlying mCG distribution (green line). (**C**) Correlation plot of AtMBD6-FL binding with the distribution of mCGs in λ-DNA. Green line indicates a theoretical linear trend.

The enrichment of AtMBD6-MBD at densely methylated regions was not seen in the binding distribution of AtMBD6-FL. Additionally, AtMBD6-FL did not appear depleted from sparsely methylated regions unlike AtMBD6-MBD. Furthermore, the correlation plot of AtMBD6-FL more closely followed a linear trend with the underlying mCG distribution (**Figure 14C**). This result shows that the mCG-density dependent binding of AtMBD6-MBD was suppressed in the context of the full-length protein. This finding is at

odds with the tendency of full-length MBDs to localize at densely methylated regions throughout the genome (17). This duality suggests that this observed inhibition of mCG-density dependent binding can be relieved in cells. Future studies will be needed to better understand how MBDs function in the context of full-length proteins and how the cell regulates their binding activity.

Next, we wanted to test MBD binding in the context of an even larger protein complex. We repurified the MBD domains of mouse MBD1 and MeCP2 with N-terminal 1xFLAG-tags or 3xFLAG-tags. In order to simulate MBD binding in the context of a large nuclear protein complex, we labeled FLAG-tagged MBDs with Quantum dots (Qdots) conjugated to anti-FLAG antibodies. Qdots are large semiconductor nanocrystals which fluoresce at a specific wavelength determined by their physical size (41). We labeled FLAG-tagged MBDs with Qdots that emit far-red light, 705 nm. Qdot705 particles are around 20 nm in diameter, which is significantly larger than the average globular protein, mimicking a large nuclear complex.

Upon initially testing FLAG-tagged MBDs on methylated DNA curtains, we frequently observed a shorter apparent lifetime for MBD binding events. This instability made it impractical to image MBD binding distributions on methylated DNA curtains. To troubleshoot binding of FLAG-tagged MBDs, we repurified FLAG-tagged constructs while retaining an N-terminal 6xHis-tag, normally cleaved off by TEV protease after purification. This preparation left us with 6xHis-TEV-1xFLAG-MBD and 6xHis-TEV-3xFLAG-MBD constructs, which were frequently more stable on DNA curtains than TEV-cleaved FLAG-tagged constructs. We used these tandem-tagged MBDs, primarily FLAG-tagged MBD1-MBDs, to optimize binding of FLAG-tagged MBD

constructs to DNA curtains. Additionally, retaining both tags allowed us to more easily investigate binding of the same MBD under different labeling conditions.

On DNA curtains, we saw that FLAG-tagged MBD1-MBDs maintained their specificity for methylated DNA and did not display any non-specific binding to unmethylated DNA curtains (**Figure 15A/D**). However, we found that the number of visualizable binding events was dependent on the number of FLAG-tags. When imaging Qdot-labeled 1xFLAG-MBD1-MBD, we observed a number of binding events comparable to 6xHis-tagged MBD1-MBD labeled with anti-6xHis-Alexa555 (both imaged at 10nM, **Figure 15A-C**). When imaging Qdot-labeled 3xFLAG-MBD1-MBD, we saw significantly more labeled protein binding to DNA curtains, forcing us to use a 40-100x lower working concentration of 3xFLAG-MBD1-MBD to image the same number of binding events as in previous experiments (as low as 100pM, **Figure 15D-F**). Regardless of concentration, both Qdot705-labeled MBD1-MBDs created binding distributions which lost their mCG-density dependence, no longer binding more frequently in densely methylated regions (**Figure 15B/E**). Compared to 6xHis-tagged MBD1-MBD, the binding of Qdot-labeled 1xFLAG-MBD1-MBD and 3xFLAG-MBD1-MBD correlated better with the distribution of mCGs in λ-DNA, resulting in a more linear trend between MBD1-MBD binding and mCGs (**Figure 15C/F**).

To further verify that Qdot-labeled MBD1-MBDs have suppressed mCG-density dependent binding, we analyzed FLAG-tagged MBD1-MBDs on methylated DNA curtains under different binding conditions. We found that Qdot-labeled MBD1-MBD maintained a largely unbiased binding distribution even in the presence of high salt, regardless of TEV cleavage, and at variable concentrations of MBD1-MBD (**Figure 16**).
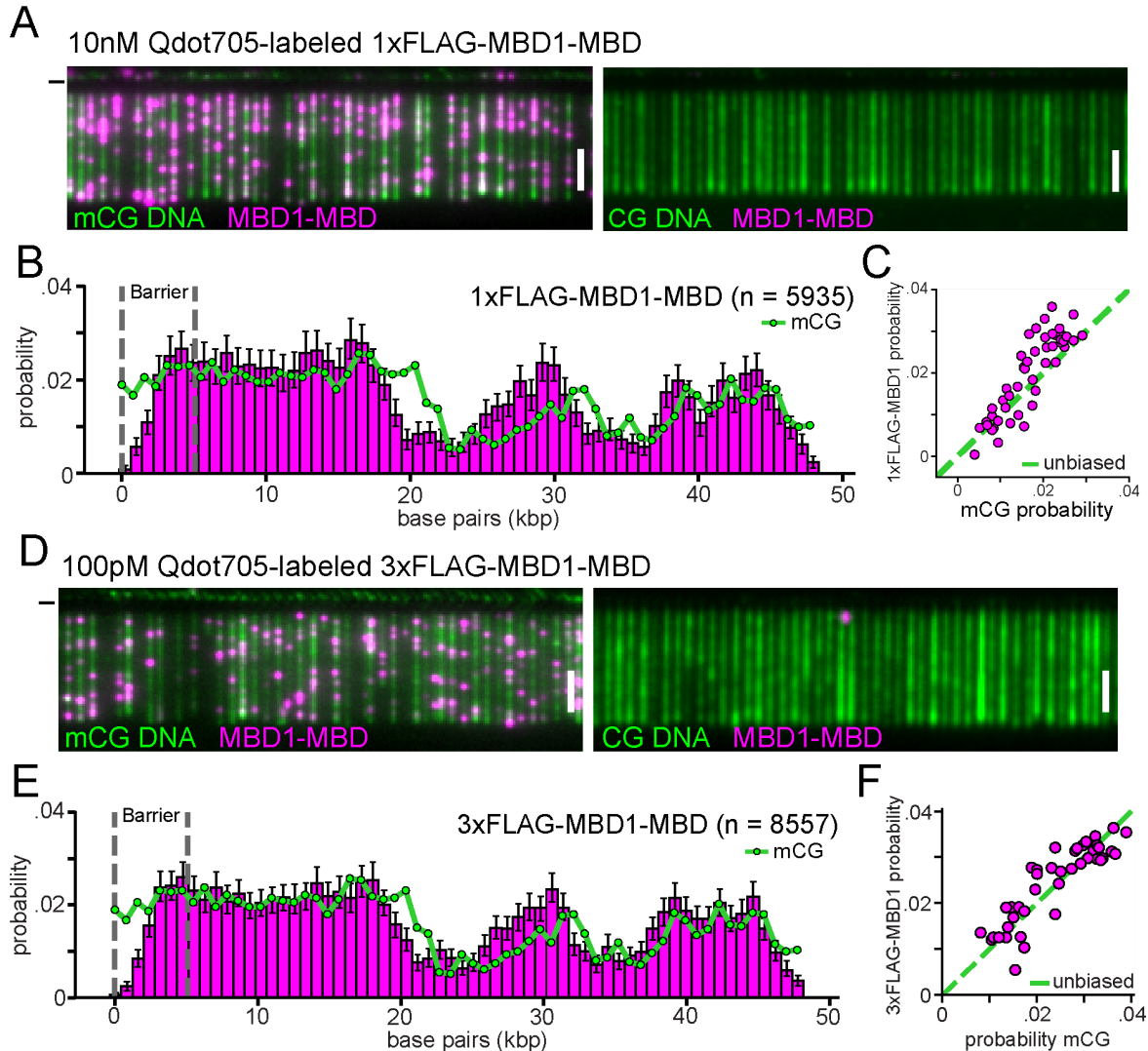
Figure 15. Qdot-labeled FLAG-tagged MBD1-MBD binding to DNA curtains.
(**A**) Representative images of 1xFLAG-MBD1-MBD (magenta) specifically binding to DNA curtains (green) when methylated. Like all previous experiments, 10nM MBD1-MBD was used to generate an analyzable binding distribution on DNA curtains. (─) indicates chrome diffusion barriers. Scale bar - 5 µm. (**B**) Histogram showing the binding distribution of Qdot-labeled 1xFLAG-MBD1-MBD measured from the free-end of λ-DNA molecules. MBD binding is plotted alongside the distribution of mCGs (green line). (**C**) Correlation plot of 1xFLAG-MBD1-MBD binding with the distribution of mCGs in λ-DNA. Green line indicates a theoretical linear trend. (**D**) Representative images showing that stable binding events of 3xFLAG-MBD1-MBD were possible at concentrations as low as ≤100pM. Images show 3xFLAG-MBD1-MBD binding events (magenta) on methylated DNA (green). (─) indicates chrome diffusion barriers. Scale bar - 5 µm. (**E**) Measured binding distribution of Qdot-labeled 3xFLAG-MBD1-MBD based on the distance from the free-end of λ-DNA molecules. Histogram is plotted alongside the underlying distribution of mCGs (green line). (**F**) Correlation plot of 3xFLAG-MBD1-MBD binding with the distribution of mCGs in λ-DNA. Green line indicates a theoretical linear trend.

These results show that Qdot-labeled MBD1-MBDs no longer displayed clear mCG-density dependent binding to methylated DNA curtains at this resolution. Instead, FLAG-tagged MBD1-MBDs consistently displayed non-discriminatory binding to methylated DNA curtains when labeled with Qdot705 at this resolution.
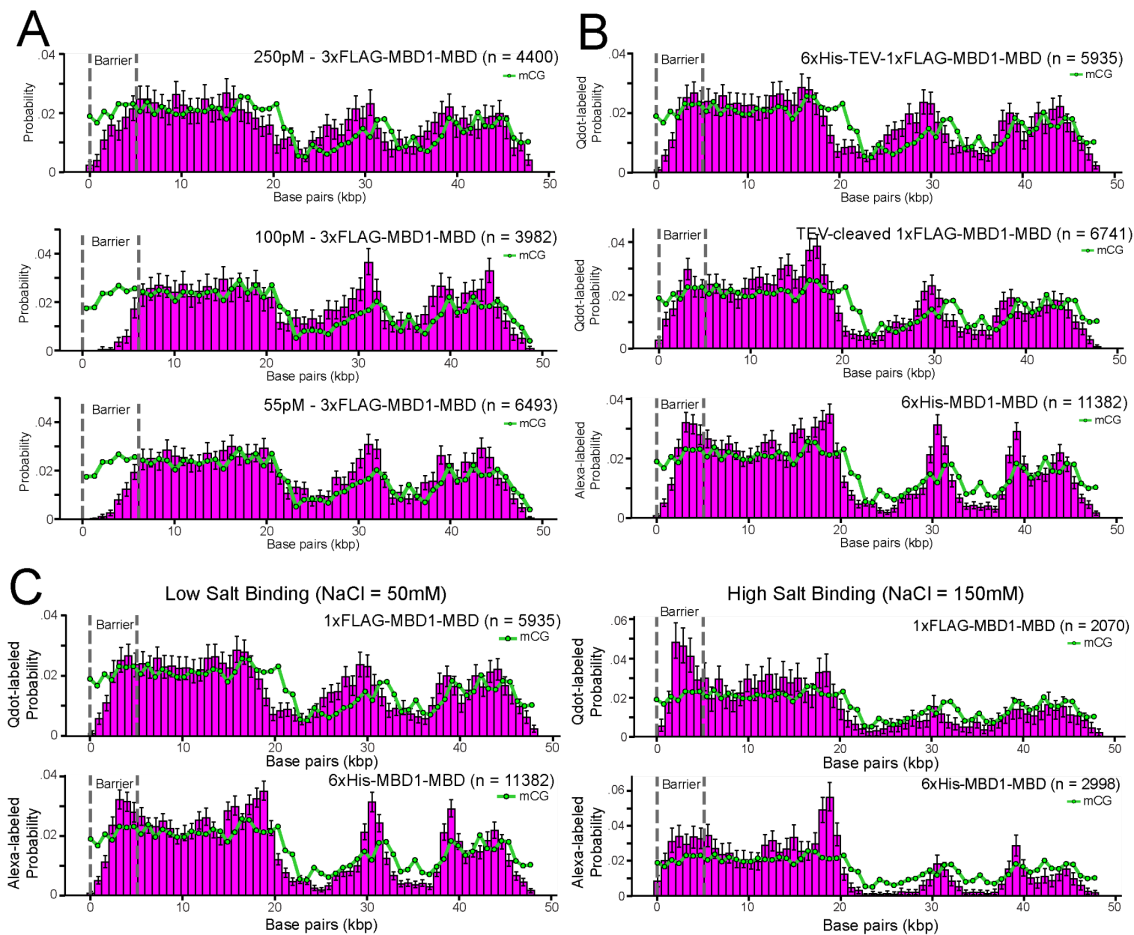
Figure 16. Qdot-labeled MBD1-MBD binding under variable conditions.
(**A**) Distributions showing effect of concentration on Qdot-labeled 3xFLAG-MBD1-MBD binding. Binding distribution is measured based on the distance of binding events from the free-end of the DNA and plotted alongside the underlying mCG distribution (green line). (**B**) Binding distributions showing how Qdot-labeled 1xFLAG-MBD1-MBD binds methylated DNA before and after TEV cleavage. Additionally, representative binding distribution of Alexa555-labeled 6xHis-MBD1-MBD is included for comparison. Histograms show the distribution of binding events measured from the free-end of the DNA alongside the underlying distribution of mCGs (green line). (**C**) Histograms show the effect of salt on the binding of Alexa-labeled 6xHis-MBD1-MBD and Qdot-labeled 3xFLAG-MBD1-MBD. Low and high salt distributions are measured based on the distance of binding events from the free-end of the DNA and plotted alongside the mCG distribution (green line).

To validate that this change in MBD binding behavior could be attributed to the size of the Qdot, we tested 3xFLAG-MBD1-MBD binding with a smaller fluorescent antibody. We used commercially available anti-FLAG-Cy3 antibodies which allowed us to visualize 3xFLAG-MBD1-MBD on methylated DNA curtains without Qdot705. However, the labeling efficiency of 3xFLAG-MBD1-MBD using anti-FLAG-Cy3 was relatively low, forcing us to use concentrations more than 100x higher than Qdot-labeled 3xFLAG-MBD1-MBD. Under these labeling conditions, Cy3-labeled 3xFLAG-MBD1-MBD once again showed mCG-density dependent binding in the absence of Qdot705 (**Figure 17**). These results suggest that both size and protein concentration can regulate an MBD's ability to cluster in densely methylated regions.
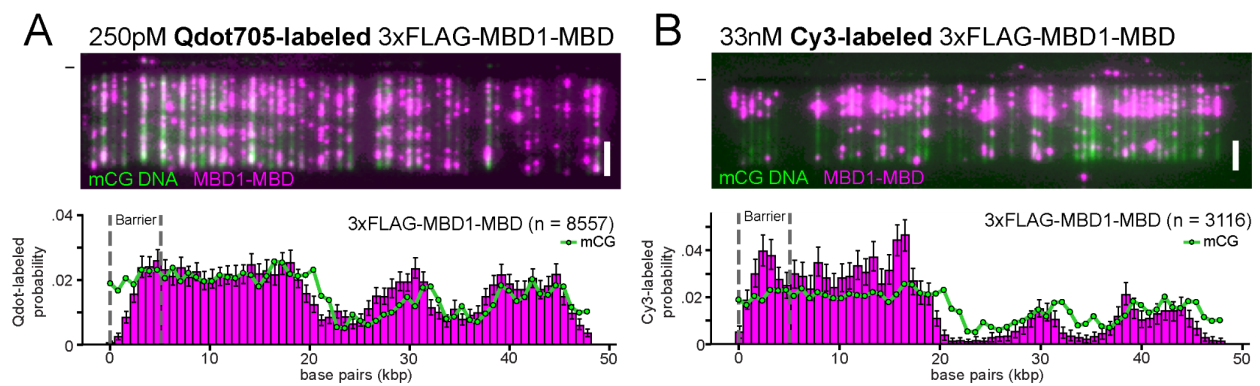
Figure 17. Differentially labeled 3xFLAG-MBD1-MBD binding to DNA curtains.
Top, representative images depicting 3xFLAG-MBD1-MBD binding events (magenta) on methylated DNA curtains (green). (─) indicates chrome diffusion barriers. Scale bar - 5 μm. **Bottom**, histograms showing binding distributions of 3xFLAG-MBD1-MBD on methylated DNA. Green line indicates the underlying mCG distribution. (**A**) Image and binding distribution of Qdot-labeled 3xFLAG-MBD1-MBD. (**B**) Image and binding distribution of Cy3-labeled 3xFLAG-MBD1-MBD.

To investigate how concentration can affect an MBD binding distribution, we investigated Qdot705-labeled 3xFLAG-MBD1-MBD binding at a higher concentration. Since the labeling efficiency of MBD1-MBD was so high under these conditions, we imaged a low concentration of Qdot-labeled 3xFLAG-MBD1-MBD (100pM) in the

presence of excess dark 3xFLAG-MBD1-MBD (10nM) to match the total concentration of MBD1-MBD we normally use when we observe mCG-density dependent binding. When imaging Qdot705-labeled 3xFLAG-MBD1-MBD in the presence of excess dark MBD1-MBD, we observed that the binding distribution shifted closer to densely methylated regions (**Figure 18**).
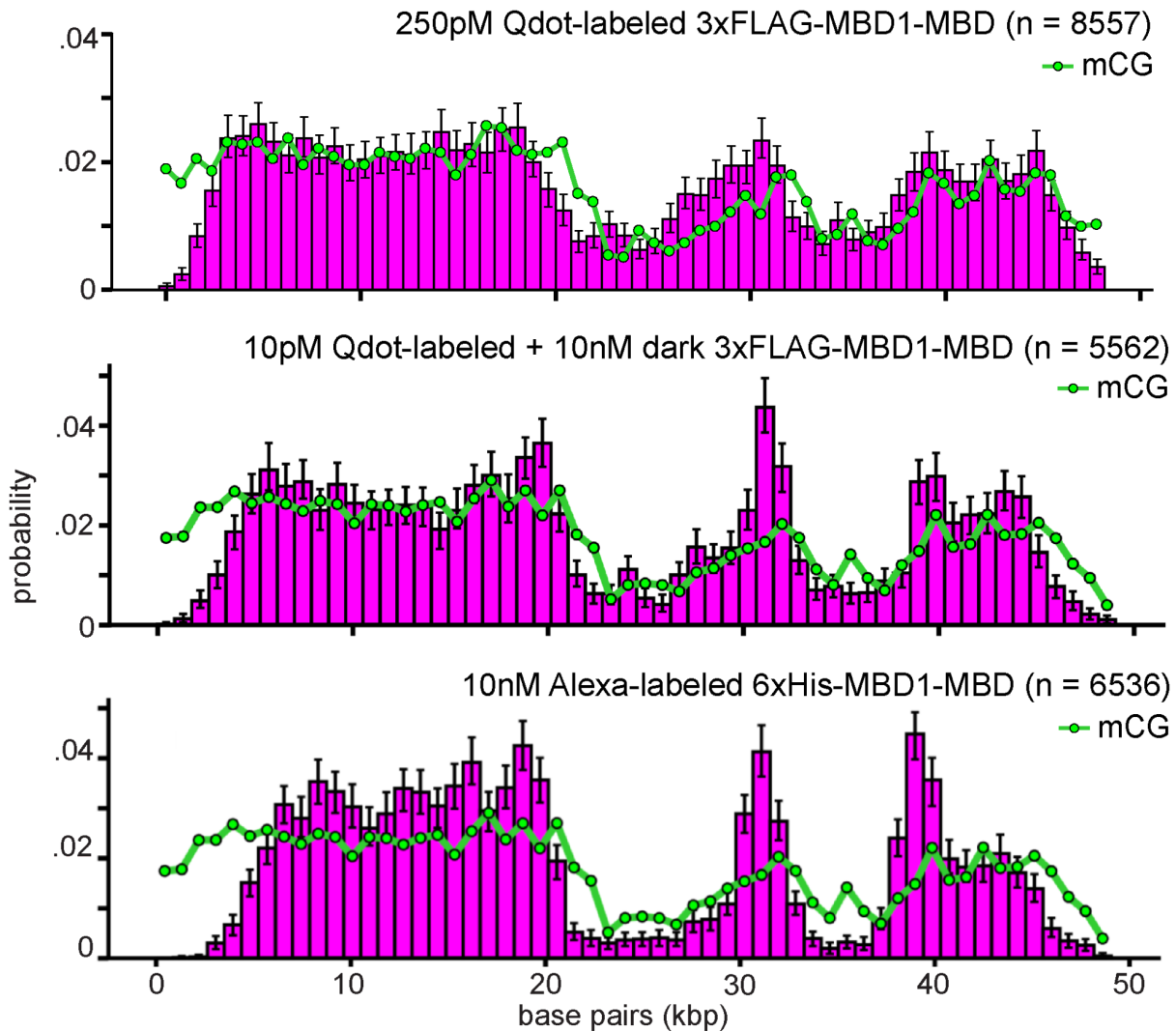
Figure 18. Qdot-labeled MBD1-MBD binding with excess dark MBD1-MBD. Representative histograms of differentially labeled MBD1-MBD binding to methylated DNA curtains. Distributions show Qdot-labeled 3xFLAG-MBD1-MBD binding in the presence and absence of excess dark MBD1-MBD, compared to a representative distribution of Alexa555-labeled MBD1-MBD. Binding distributions were created based on their distance from the free-end of the DNA. Green lines indicate the underlying mCG distribution.

This result suggests that a critical concentration of MBD1-MBD may be required for enrichment at densely methylated regions. However, mCG-density dependent binding still appeared suppressed in the presence of Qdot705, suggesting that this activity is at least partially dependent on label size.

To directly test differentially labeled MBD1-MBD binding at the same concentration, we utilized tandem-tagged MBD1-MBD, presenting both an N-terminal 6xHis-tag and 1xFLAG-tag. Previously we showed that labeling MBDs via either one of these tags creates analyzable binding distributions using the same MBD concentration, 10nM. On methylated DNA curtains, we observed that imaging the same tandem-tagged MBD1-MBD construct could produce different binding distributions depending on the label used to visualize the MBD, even while maintaining the same concentration of MBD1-MBD.

Once again, Qdot-labeled MBD1-MBD showed uniform binding along methylated DNA curtains, which generally matched the underlying mCG distribution (**Figure 19A**). However, like in previous experiments, Alexa555-labeled MBD1-MBD maintained mCG-density dependent binding (**Figure 19B**). Even while using the same MBD1-MBD construct, Alexa-labeled MBD1-MBD did not match Qdot-labeled MBD1-MBD binding to methylated DNA curtains. This result is consistent with Qdot-dependent suppression of mCG-density dependent binding on DNA curtains at this resolution.
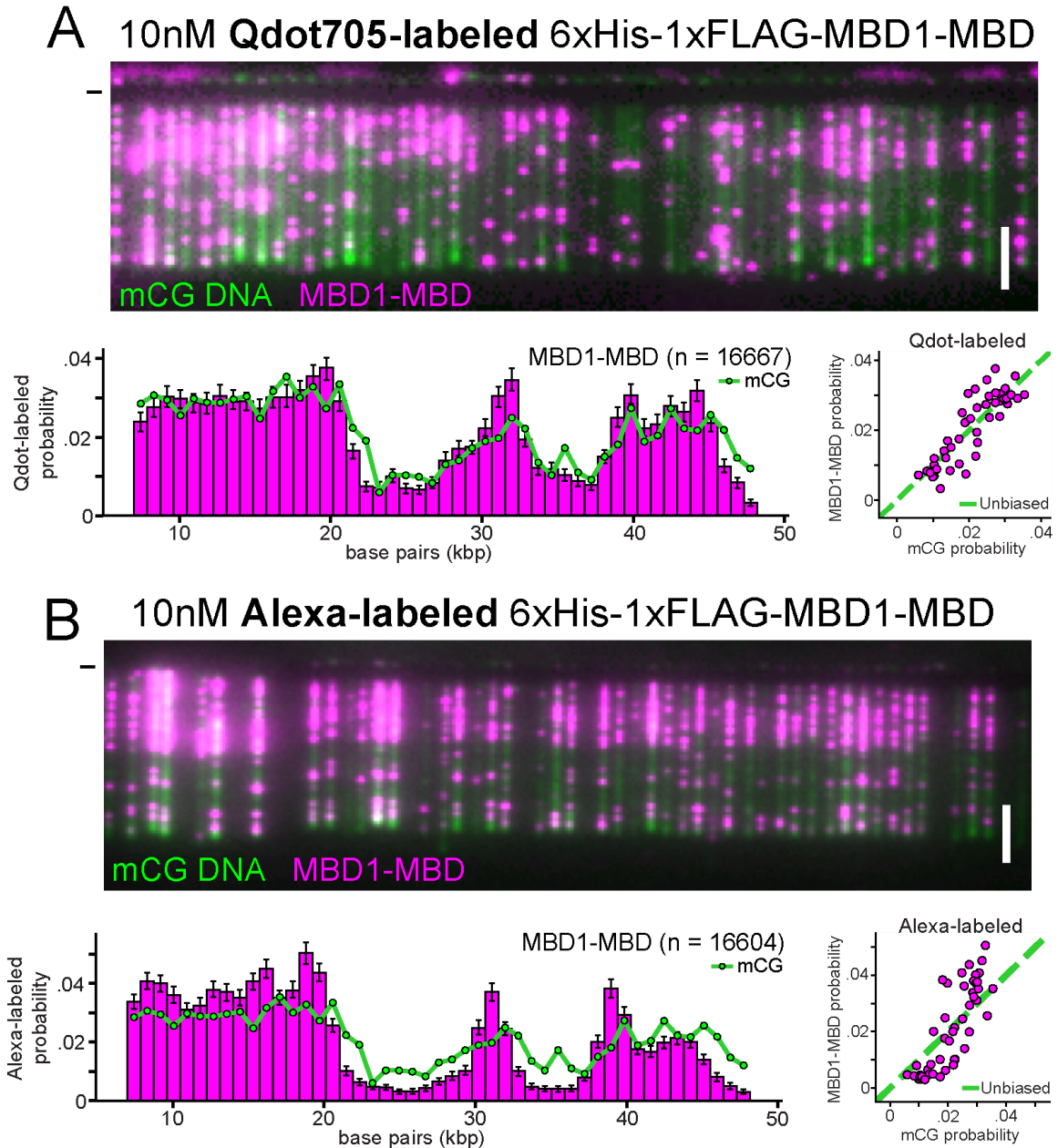
Figure 19. Imaging tandem-tagged MBD1-MBD on DNA curtains. Representative images show methylated DNA (green) bound by differentially labeled MBD1-MBD (magenta). (—) indicates chrome diffusion barriers. Scale bar - 5 μm. Histograms show the analyzed binding distributions under different labeling conditions compared to the mCG distribution (green line). Correlation plots compare differentially labeled MBD1-MBD binding with the mCG distribution. Green dashed lines indicate a linear trend between MBD binding and mCG. (**A**) Binding of Qdot-labeled tandem-tagged MBD1-MBD on methylated DNA curtains. (**B**) Binding of Alexa-labeled tandem-tagged MBD1-MBD on methylated DNA curtains.

Directly comparing the binding distributions of Alexa-labeled MBD1-MBD and Qdot-labeled MBD1-MBD validated their unique binding preferences. Densely methylated regions were more frequently bound by Alexa-labeled MBD1-MBD, while sparsely methylated regions were more frequently bound by Qdot-labeled MBD1-MBD (**Figure 20**). These results further support that labeling MBDs with Qdot705 is sufficient to suppress their mCG-density dependent binding, independent of MBD concentration.
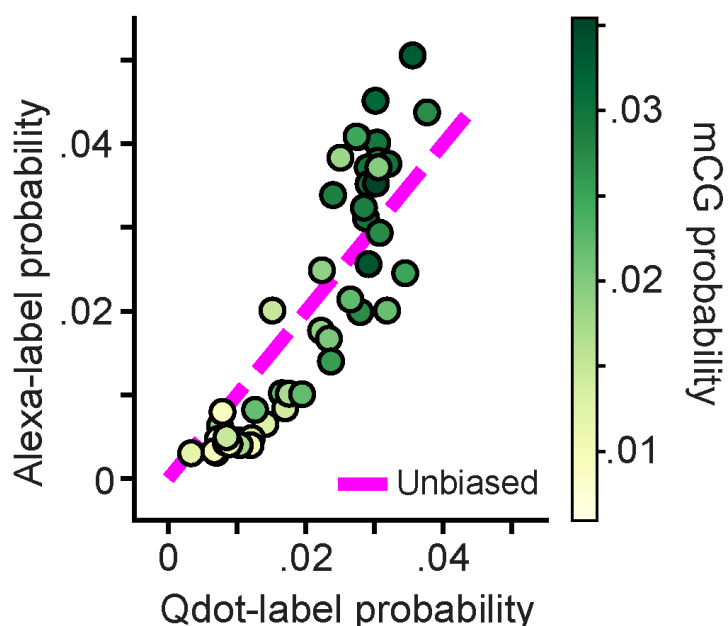
Figure 20. Correlation between differentially labeled MBD1-MBD binding. Correlation plot comparing the binding probability of Alexa-labeled MBD1-MBD with the binding probability of Qdot-labeled MBD1-MBD. Each point is color coded by the underlying mCG density within that region. An unbiased, linear trend is represented by the dashed magenta line. Points falling off the line are differentially bound by the same MBD1-MBD construct under different labeling conditions.

Indirectly labeling mouse MBD1-MBD with a Qdot705-labeled antibody, or even an Alexa555-labeled antibody, could be a relatively large perturbation to the relatively compact MBD domain (<15kDa). We repurified and directly labeled tandem-tagged 1xFLAG-MBD1-MBD using Cy3-NHS-ester to verify the binding behavior of a minimal mouse MBD1-MBD construct (Cy3-MBD1-MBD). This MBD construct is smaller than all other MBDs that we have imaged, which either include a MBP-tag (>40kDa) or a fluorescent antibody (>150kDa).

Despite numerous attempts, we struggled to get active and efficiently labeled Cy3-MBD1-MBD using NHS-ester labeling under these conditions. This consistent outcome suggests that MBD1-MBD may be particularly sensitive to this kind of chemistry. Furthermore, leftover free dye made it difficult to quantify the labeled population of Cy3-MBD1-MBD. When imaging Cy3-MBD1-MBD on DNA curtains, we were forced to use up to 10x higher concentrations than previous experiments, suggesting that our NHS-ester labeling efficiency was low. Regardless, we were able to visualize stable Cy3-MBD1-MBD binding on methylated DNA curtains. Unlike AtMBD6-FL or Qdot-labeled MBD1-MBD, directly labeled Cy3-MBD1-MBD bound mCGs in a mCG-density dependent manner (**Figure 21**). Compared to previous experiments with 6xHis-tagged MBDs, recruitment of directly labeled Cy3-MBD1-MBD was even more dependent on the density of mCGs. This result further suggests that MBD concentration may play a role in regulating mCG-density dependent binding.

Together, these results show that MBD binding to methylated DNA is context-dependent. A single mCG is not necessarily sufficient for stable recruitment of MBDs, and all mCGs in a distribution of DNA methylation are not seen equally. An MBD's sensitivity to mCG density loosely correlates with MBD concentration and size. This trend suggests that a critical concentration of local MBD may be required to stably cluster together in densely methylated regions. Additionally, these results are consistent with MBD domains having an inherent preference for densely methylated DNA, which can be regulated in the context of a full-length protein in a dynamic nuclear environment.
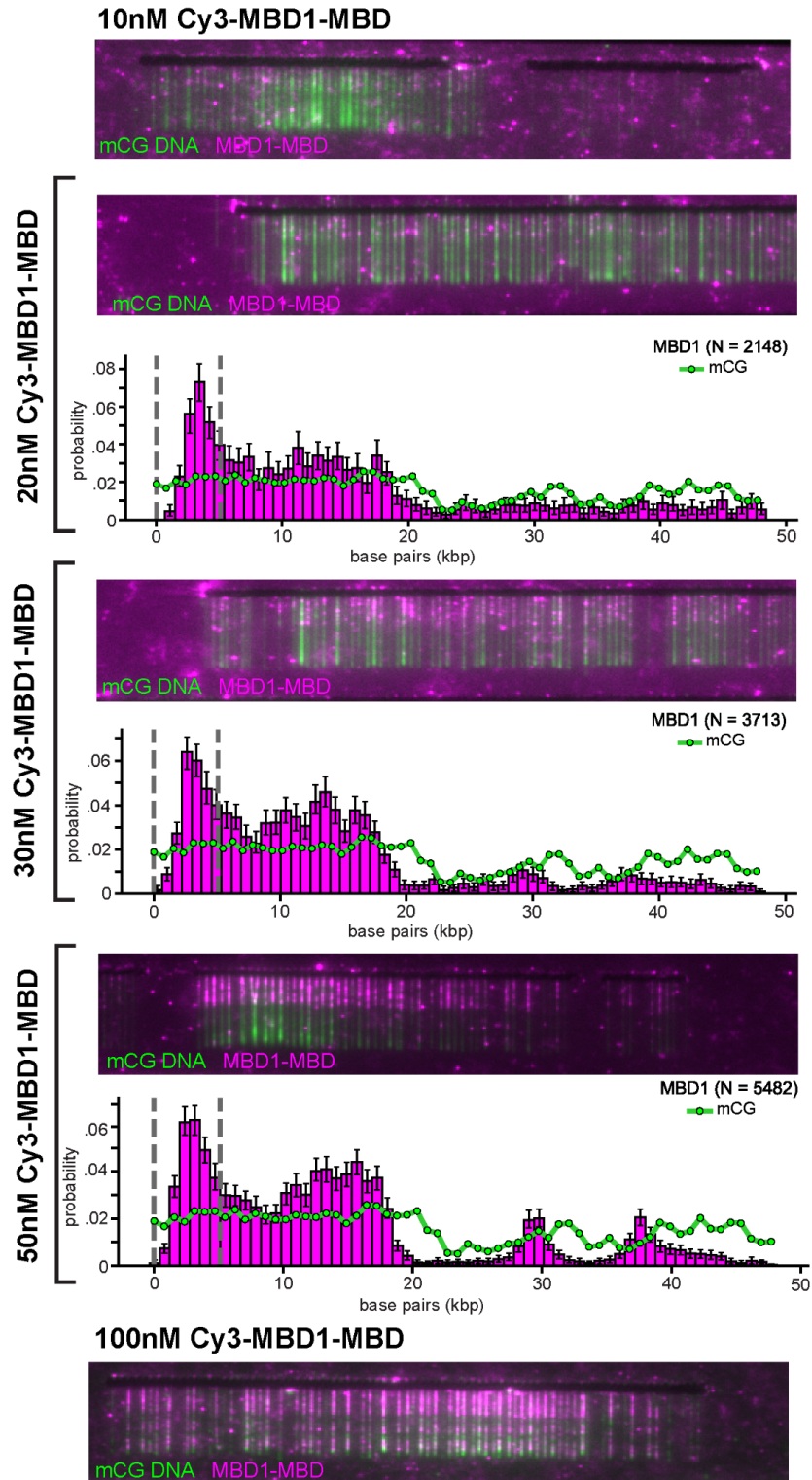
Figure 21. Directly-labeled MBD1-MBD binding to methylated DNA curtains.
Representative images and binding distributions of Cy3-NHS-ester labeled MBD1-MBD (magenta) on methylated DNA curtains (green). Histograms show measured MBD1-MBD binding events alongside the underlying distribution of mCGs (green line).

2.3 Sources of error on DNA curtains

Despite the power of DNA curtains and other single molecule imaging techniques, there are multiple sources of error that can limit the precision of these measurements. Error can be introduced while imaging MBD binding to DNA curtains on the microscope. Additionally, error in our measurements can come from how we analyze the images that we collect. These limitations can substantially limit our ability to analyze MBD binding at the single molecule level.

When imaging MBDs on DNA curtains, a low labeling efficiency can skew our measured binding distributions. A poorly labeled population can lead to sampling error if we fail to visualize a significant proportion of binding events. This possibility is consistent with some of the most mCG-density dependent MBD constructs requiring higher concentrations of MBD to image on DNA curtains. However, with sufficiently high n, we would still be able to accurately measure a binding distribution even if the majority of binding events were dark. All of our measured distributions consist of thousands of binding events, which should provide an accurate understanding of the underlying distribution, even if a large amount of the population is not visible (**Figure 22**). With sufficiently high n, low labeling efficiency cannot explain the density dependent binding of MBDs that we see on DNA curtains, especially for MBD binding distributions enriched at one specific region (**Figure 13**).
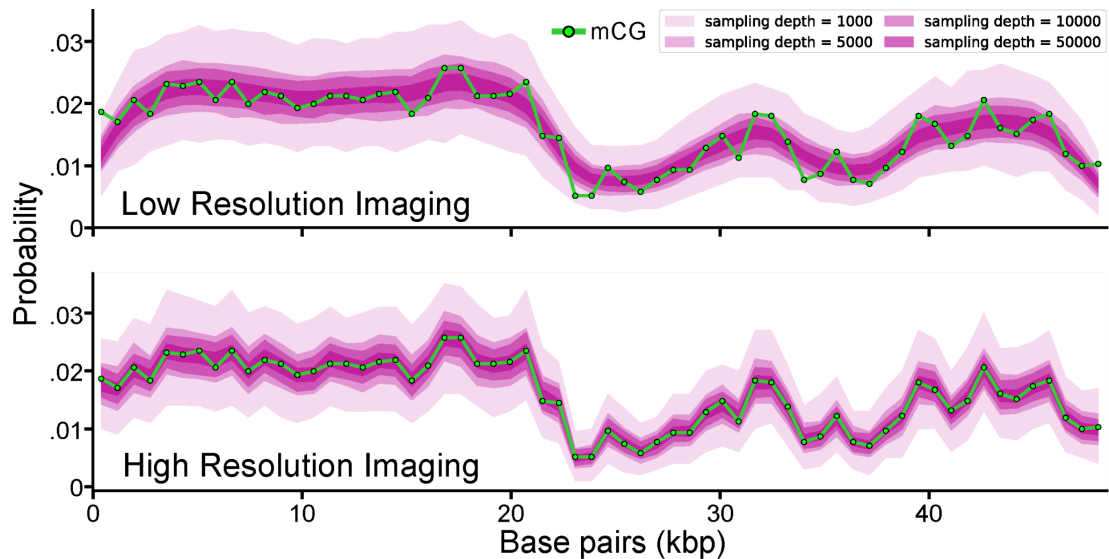
Figure 22. Effect of sampling depth and error on a measured binding distribution. Plots showing how sampling depth and error affects our ability to measure a distribution, using the known distribution of mCGs in λ-DNA. Each shaded region represents the range of probabilities a different sample size could create. The green line indicates the true distribution of mCGs. **Top**, random sampling of mCGs at low resolution, like measuring binding events based on their distance from the free-end of the DNA. Whenever an mCG was sampled, its position was adjusted based on an error rate of approximately ± 1000 bp. **Bottom**, random sampling of mCGs at much higher resolution, emphasizing the effect of sampling depth. Whenever an mCG was sampled, its position was modified by a much lower error rate of approximately ± 50 bp.

The most likely way that low labeling efficiency could be skewing our measured distributions would be if labeling MBDs artificially skewed their binding preferences. In this case, a measurement of visualizable MBDs would not be an accurate representation of an underlying dark population, which would more closely resemble the natural binding behavior of MBDs. Since DNA curtains require fluorescently labeling any protein of interest, it is impossible to image MBD binding events without some modification to these domains. Of all the binding conditions that we imaged, the two directly-labeled MBD domains (NHS-ester labeled AtMBD6-MBD and Cy3-MBD1-MBD) both showed mCG-density dependent binding to DNA curtains, suggesting that this is an inherent behavior of MBDs. In fact, the majority of conditions we used to image MBD

binding to DNA curtains all consistently show mCG-density dependent binding of MBDs to methylated DNA. The only conditions in which an MBD did not show density dependent binding was when AtMBD6-MBD was imaged in the context of the full-length protein and when we attached a large semiconductor nanocrystal to MBD1-MBD (**Figure 23**). These exceptions demonstrate that binding conditions can skew the behavior of MBDs but most likely in a way that masks mCG-density dependent binding of MBDs. Further exploration of how biologically relevant conditions (like concentration, salt, and size) can affect MBD binding will be informative to understand how cells regulate MBD binding throughout the genome.
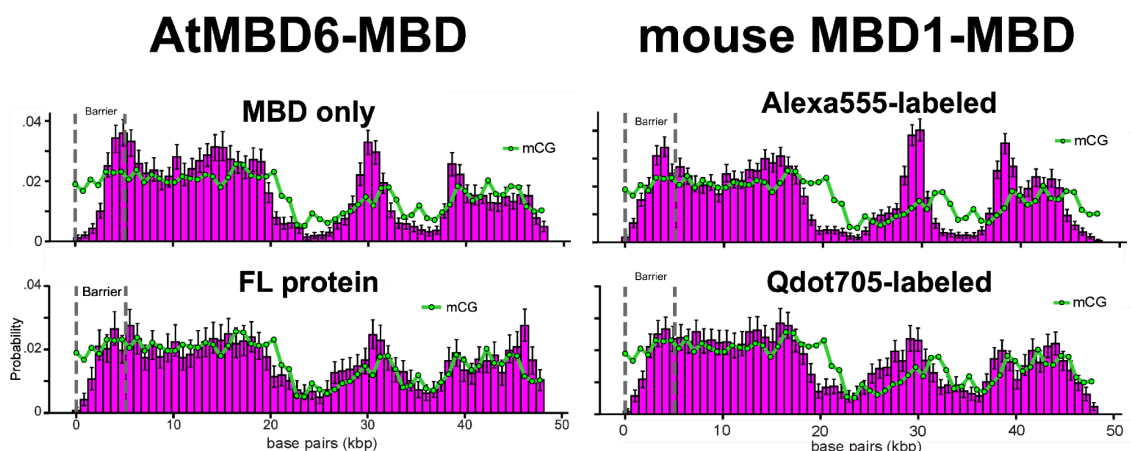
Figure 23. Regulation of mouse MBD1-MBD and AtMBD6-MBD binding. Representative binding distributions of AtMBD6-MBD and mouse MBD1-MBD under different conditions, showing that mCG-density dependent binding is suppressed in the context of AtMBD6-FL and Qdot705-labeled MBD1-MBD.

After collecting images of MBDs on DNA curtains, we can further introduce error during image analysis. Our ability to correctly position binding events on DNA is a function of three factors: calculating the position of each MBD1-MBD within its own fluorescent channel, aligning that signal with the corresponding DNA in the second fluorescent channel, and determining the dimensions of DNA molecules in the flowcell.

When localizing fluorescent MBDs, the point spread function (PSF) of these point sources can be fit to determine their mean position within 1nm (42). Thus, localizing MBD binding events within their own fluorescent channel does not contribute a large amount of error to our measurements. We also rely on the ability to accurately position point sources to align the two fluorescent channels. We image commercially available beads which fluoresce at multiple wavelengths in both channels. We then use the offsets between each image to align our fluorescent channels (**Figure 24**). Generally, we found that shifting the entire field of view using the average (x,y) offset was sufficient to align our microscope images. Once we began imaging with Qdot705, we found that this approach was error prone under these imaging conditions. Moving forward, we began using the measured offsets to perform a piecewise affine transformation to differentially shift individual regions across our field of view (43). This new approach allows us to localize our channels to within several nanometers.
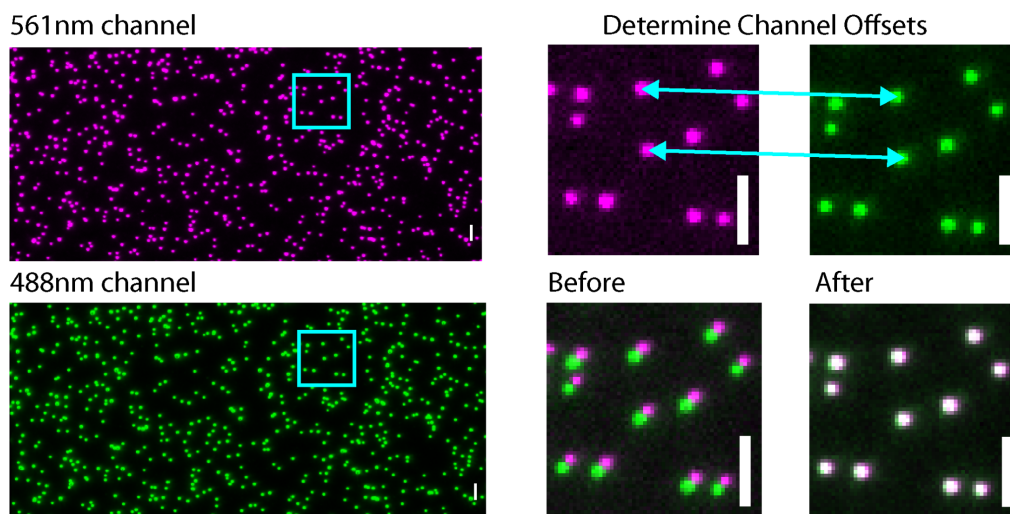
Figure 24. Overview of fluorescent channel alignment. The representative images simultaneously show the position of tetraspeck beads, which fluoresce at multiple wavelengths, in both microscope channels. Point-fitting the position of the same beads in both channels allows us to determine the offset between each channel across the field of view. These offsets can then be used to align images of DNA molecules in one channel with their corresponding binding events in the other channel. Scale bar - 5 μm.

The largest source of error in our ability to assign binding events to base pair positions arises from fitting the DNA molecule under flow. Unlike fluorescent point sources, which appear as diffraction-limited spots, the fluorescent signal of the entire λ-genome is distributed over many pixels. Identifying the beginning of each λ-DNA molecule is difficult because the fluorescent signal of the tethered-end is occluded by the chrome barrier. The free-end of the DNA is visible, but the fluorescent signal is diffuse because of large lateral fluctuations in solution.

Furthermore, when extended under buffer flow, tethered DNA is differentially stretched along its length (**Figure 25**). Positions near the tethered-end experience larger forces than those near the free-end. This difference results in greater extension near the tethering point, and consequently, fewer DNA base pairs per pixel. Meanwhile, positions near the free-end are under less force, resulting in more base pairs of DNA contained within each pixel (44,45,46).
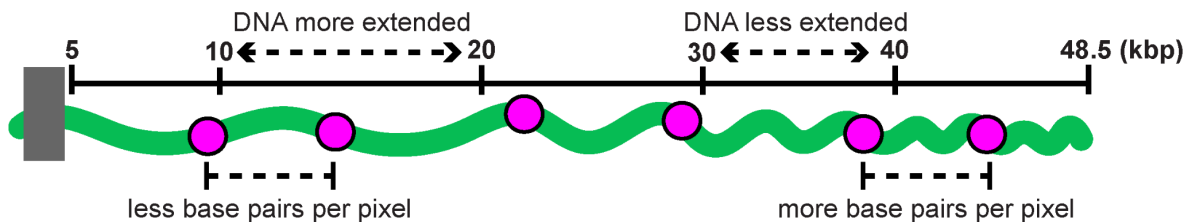
Figure 25. Non-uniform extension of tethered DNA molecules under flow. This cartoon shows how DNA molecules are differentially extended in the flowcell. The grey rectangle represents the chrome diffusion barrier where λ-DNA (green) is tethered and stretched under buffer flow. This results in non-uniform extension of λ-DNAs, impeding our ability to accurately measure the position of DNA binding events (magenta).

This phenomenon is why many low resolution binding distributions measured on DNA curtains appear stretched when compared to the underlying distribution of DNA methylation (**Figure 15, 17**). Even if we could reliably identify the sub-pixel position of each end of the DNA, the non-uniform extension of DNA under flow will still hinder our

accuracy in assigning base pair positions to binding events. However, if we could account for this non-uniform extension, we could generate high-resolution distributions of MBDs, or any other protein, on DNA curtains.

2.4 Using dCas9s to measure and lower error of analysis

To minimize the error of our measurements, we sought a more accurate way to analyze proteins bound to DNA curtains. Additionally, we required a tool to quantify the localization error of our measurements. To achieve these goals, we used fluorescent dCas9s as fiduciary landmarks by targeting them to 8 unique sites evenly spaced along the length of λ-DNA (**Figure 26A**). Using fluorescent dCas9s to localize DNA molecules removes our reliance on fitting the extended and variable YOYO-1 fluorescence. By allowing us to use several point sources that are fit with high precision, dCas9s additionally provide a way to directly measure the non-uniform extension of DNA under flow. Furthermore, this approach allows us to quantify the error of our measurements by comparing the measured positions of dCas9s to their target sites.
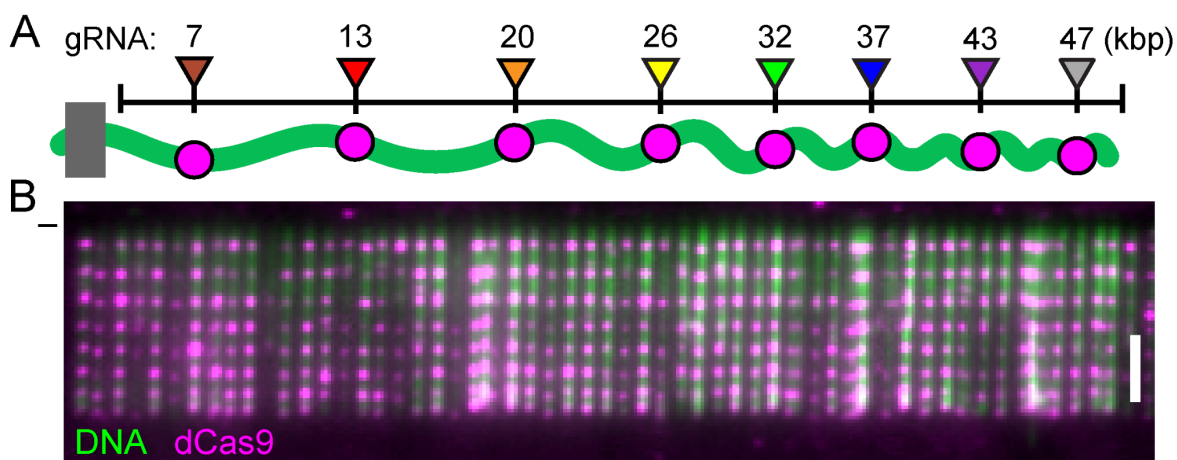
Figure 26. Targeting dCas9s along λ-DNA. (**A**) This cartoon shows how dCas9 molecules (magenta) were targeted approximately every 5-7 kbp along the length of λ-DNA (green). Color-coded triangles indicate the target site of each dCas9. (**B**) Representative image showing 8 dCas9s (magenta) stably bound to YOYO-1 stained λ-DNA (green). (—) indicates chrome diffusion barriers. Scale bar - 5 µm.

We found that commercially available 6xHis-tagged recombinant dCas9 labeled with anti-6xHis-Alexa555 was the most efficient way to reliably target and visualize multiple dCas9s along λ-DNA (Alt-R S.p. dCas9 protein V3, **Figure 26B**). We initially used these dCas9s to quantify the localization error of our analysis when fitting binding events based on a uniform extension from the free-end of the DNA. We identified 8 unique dCas9 populations along the length of λ-DNA. However, the majority of these populations were frequently mislocalized from their target sites. To quantify the error of this approach, we directly compared the distribution of each measured dCas9 population to the true target site.
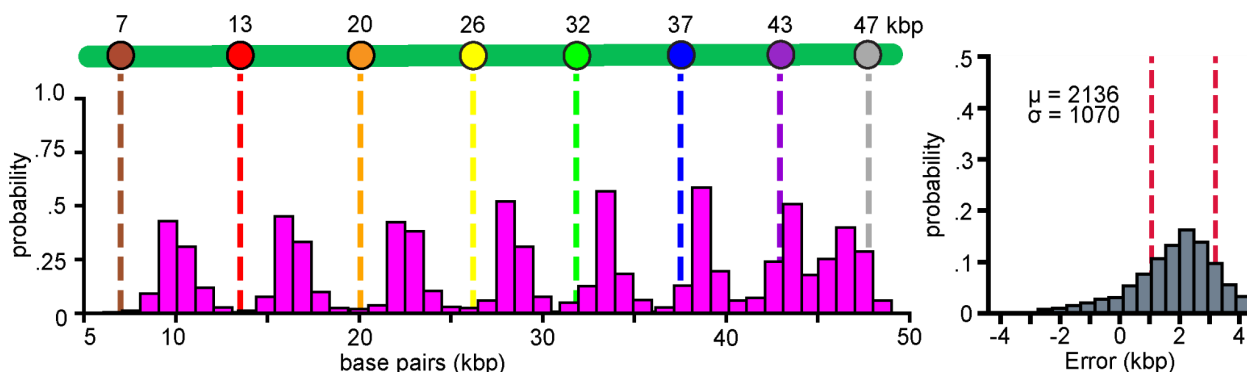
Figure 27. Measured Binding distribution of dCas9s using baseline analysis.
Left, histogram showing the analyzed distribution of dCas9 positions based on their distance from the free-end of the DNA. Multi-colored lines indicate the true target position of each population. **Right**, histogram displaying the average global error for all dCas9s measurements. The mean and standard deviation were determined by fitting a gaussian to the distribution.

The average mean position of all dCas9s was shifted 2136 base pairs upstream from the target site and each measurement fell within a standard deviation of 1070 base pairs from the measured mean (**Figure 27**). This error demonstrates the effect of assuming a uniform extension of differentially extended DNA molecules and describes the range that we can expect any single measurement to fall within when using this approach. Additionally, the mean offset was greater for dCas9s targeted to sites farther

away from the free end of the DNA. This skew suggests that this analysis underestimated the number of DNA base pairs contained within each pixel, further increasing error for any binding event farther away from the free-end of the DNA.

To refine this analysis, we used dCas9s to directly measure the average extension of DNA molecules in the flow cell. We calculated an average of 874 base pairs of DNA contained within each pixel, which is consistently higher than our previous estimation (**Figure 28**). Using this average measurement, we can more accurately fit binding events based on their pixel distance from the free-end of the DNA.
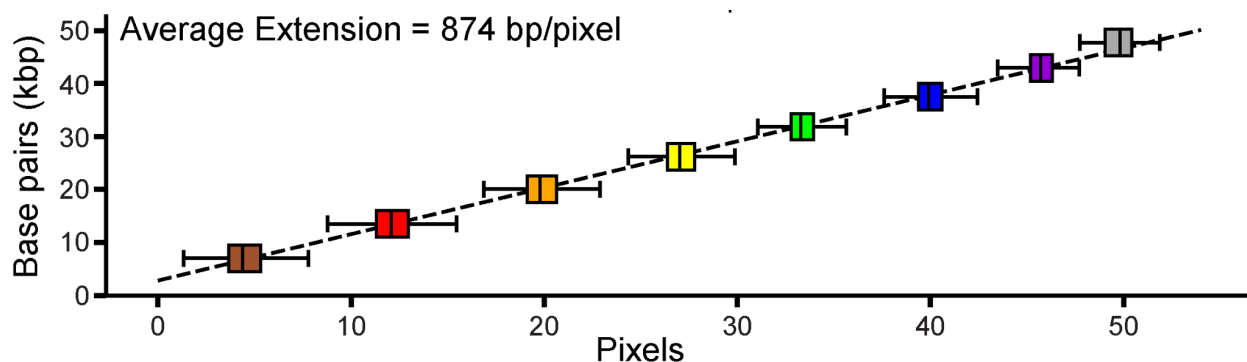
Figure 28. Calculating the average extension of DNA curtains using dCas9s. The box-and-whisker plots each represent the range of pixel positions measured for each population of dCas9 molecules. The dotted-line is the best linear fit of the extension of λ-DNA molecules as measured by dCas9s.

Applying this measured average to dCas9s dramatically shifted all populations closer to their true target site. The average mean position of all dCas9s was now only 134 base pairs away from the true target site. However, each measurement still fell within a standard deviation of 904 base pairs away from the measured mean (**Figure 29**). The majority of low-resolution MBD binding distributions presented here, as measured from the free-end of the DNA, were all calculated using this global average linear extension.
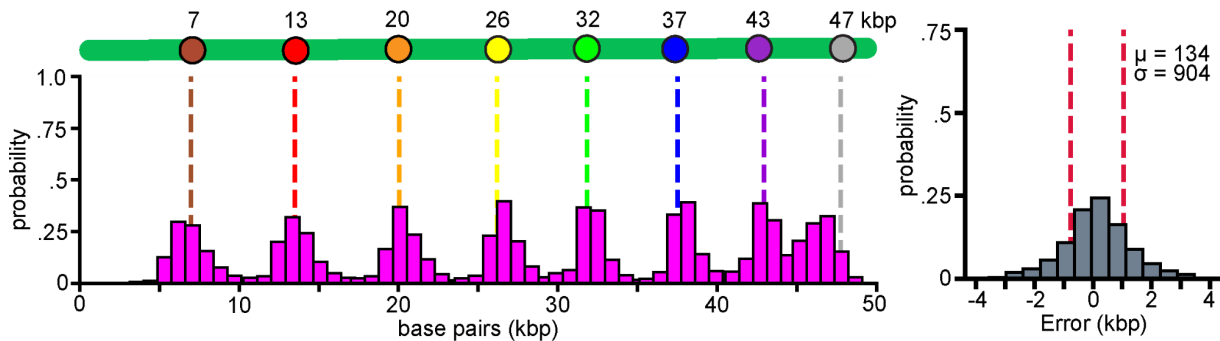
Figure 29. Binding distribution of dCas9s using global linear extension.
Left, histogram showing the analyzed distribution of dCas9 positions based on their distance from the free-end of the DNA using the global measurement of 874 bp per pixel. Multi-colored lines indicate the correct target position of each population. **Right**, histogram displaying the average global error for all dCas9s using this analysis. The mean and standard deviation were determined by fitting a gaussian to the distribution.

Overall, we find that errors arising from using fluorescently labeled DNA to identify the positions of binding events are substantial enough to mislocalize a DNA bound protein to a site hundreds of base pairs away from its actual location. This relatively large error is not unexpected for diffraction-limited measurements. Additionally, such a large amount of error would have a smoothing effect on measured binding distributions (**Figure 22**). Despite this error rate, our ability to observe mCG-density dependent binding of MBDs under these conditions further highlights that this is a driving factor for MBD recruitment to methylated DNA.

To improve the resolution of DNA curtains, we once again repurposed these dCas9s as fiduciary landmarks that identify true locations of specific base pair positions along each DNA molecule within a curtain. Rather than relying on a global estimate of DNA extension, we tried using pairs of dCas9 molecules to directly measure the extension of individual DNA molecules. This would allow us to use a more accurate measurement of local DNA extension to fit nearby binding events rather than relying on a global average which propagates error over great distances.

First, we tested the feasibility of this approach by using any possible flanking pair of dCas9 molecules to fit internal binding events, even if the pair was separated by great distances. After measuring the extension of a bound DNA molecule using a flanking pair of dCas9s, we used this factor to calculate the position of internal dCas9 binding events based on their distance from the flanking pair. Despite including distant pairs of dCas9s in this analysis, using this measurement of DNA extension allowed us to more accurately predict dCas9 target sites (**Figure 30A**). Despite an overall improvement, this approach did not dramatically shift the average mean of all dCas9s, which remained 176 base pairs away from the target site. However, we saw a dramatic reduction in the error of our fits since each measurement now fell within a standard deviation of 234 base pairs from the measured mean. This improvement matches previous measurements using proteins as specific markers of DNA sequence (47).

Next, we restricted this analysis by limiting ourselves to the local DNA extension as determined by the closest possible flanking pair of dCas9s. While still relying on a linear estimate of DNA extension, these measurements will only be used within a local region of DNA (10-15kbp), where fluctuations in extension should be much smaller than over the length of the entire molecule. After measuring the local extension of the DNA using the closest possible flanking pair of dCas9s, we used this measured extension to calculate the position of internal dCas9 binding events based on their distance from the flanking dCas9s. When using this approach, the average mean position of all dCas9s was further reduced to 42 base pairs and each measurement fell within a standard deviation of 190 base pairs from the measured mean (**Figure 30B**), significantly improving our ability to measure the position of binding events on DNA curtains.
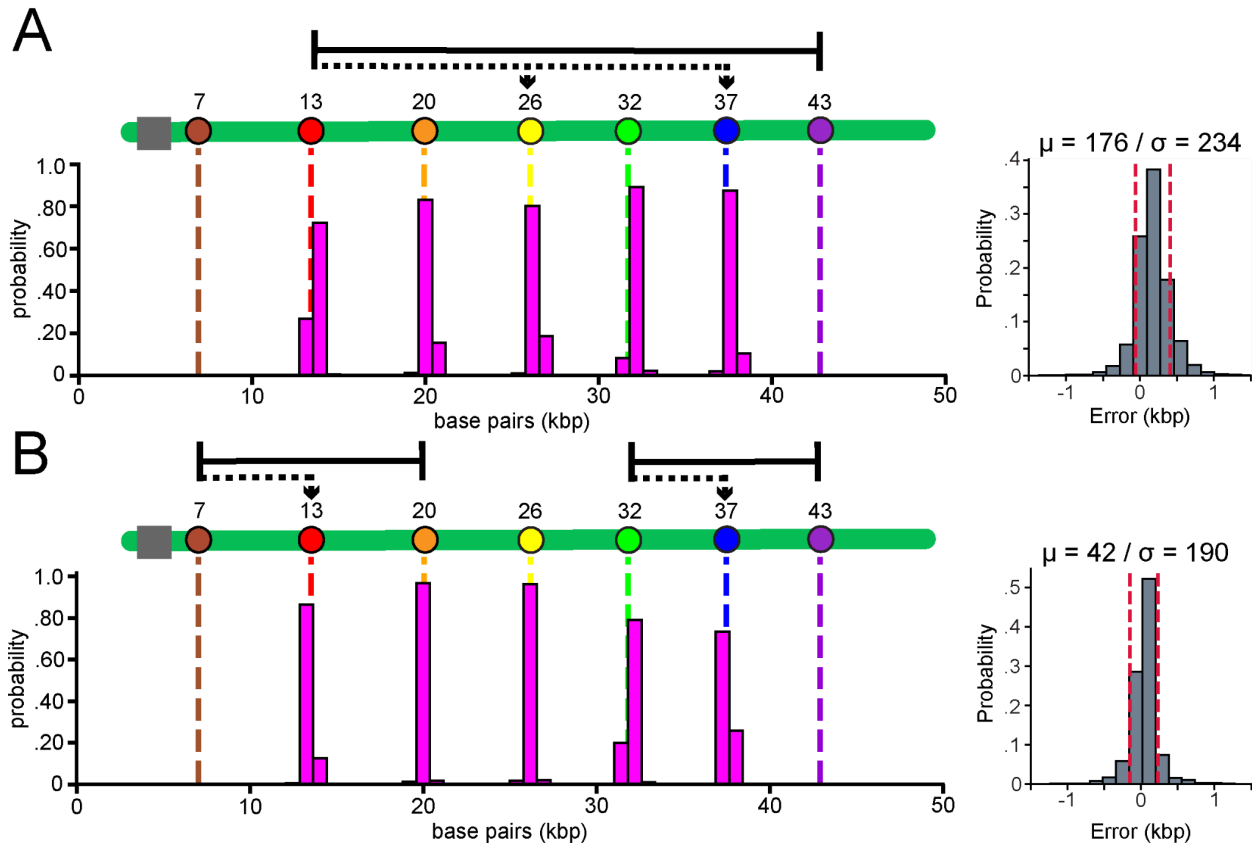
Figure 30. Using flanking pairs of dCas9s to measure binding events.
(**A**) **Left**, histogram showing the analyzed distribution of dCas9 positions based on the DNA extension as measured by any flanking pair of dCas9s. Multi-colored lines indicate the correct target position of each population. **Right**, histogram displaying the average global error for all dCas9s fit by any possible flanking pair. (**B**) **Left**, histogram showing the analyzed distribution of dCas9 positions based on the local DNA extension as measured by the closest flanking pair of dCas9s. Multi-colored lines indicate the correct target position of each population. **Right**, histogram displaying the average global error for all dCas9s using the measured local DNA extension as determined by the closest flanking pair of fiducial dCas9s.

Despite this increase in resolution, using pairs of dCas9s to fit nearby binding events is still limited by relying on a linear estimate of DNA extension. We hypothesized that we could further improve our analysis by using fiducial dCas9s to develop a method to fit the DNA in a way that better accounts for non-uniform extension. First, we created custom software that uses template matching to locate each DNA molecule by way of the fluorescent signal from bound dCas9s (**Figure 31A**). Then, we adapted a function

describing the extension of a worm-like chain under flow to capture the behavior of tethered λ-DNA within a DNA curtain (45,46). We use this function to fit extension curves that relate the pixel position of dCas9 molecules within an image to their targeted position in base pairs. Once the fitting parameters are established, any protein bound at an unknown target site can also be converted from a pixel position in the image to a base pair position on the DNA.
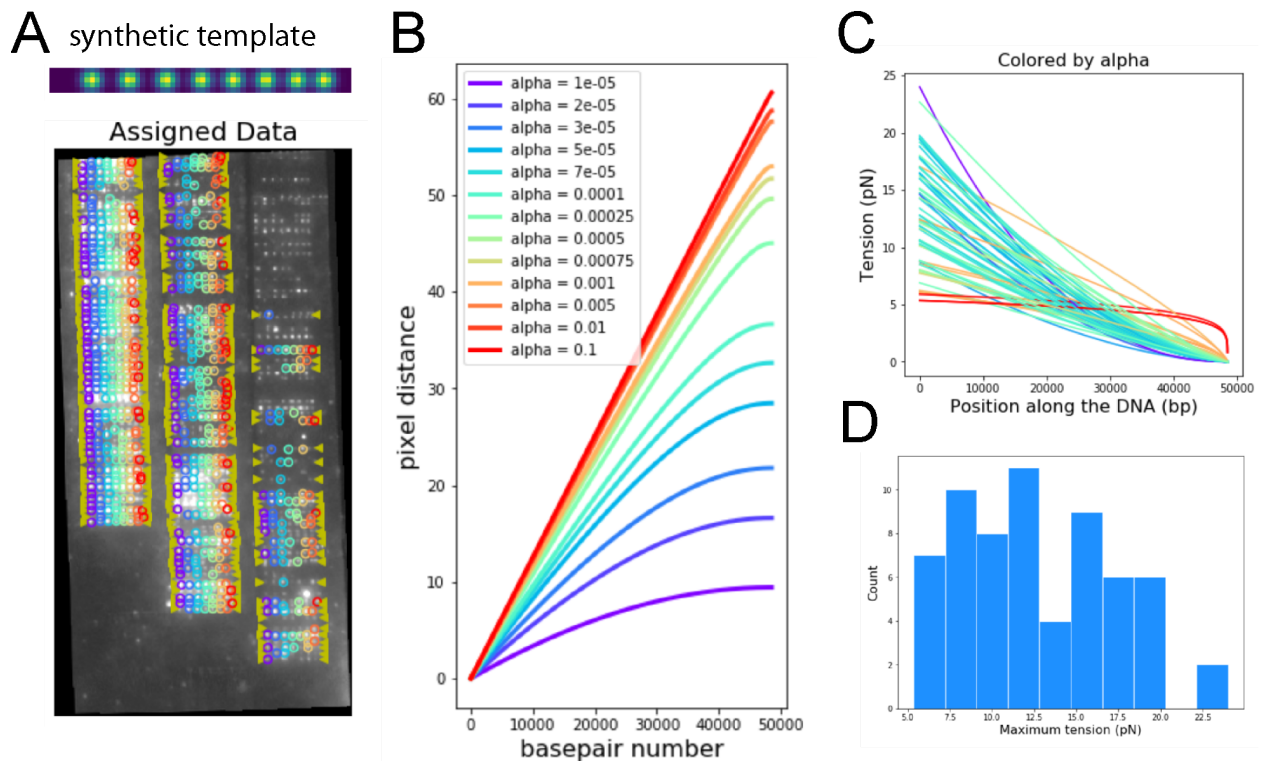
Figure 31. Using dCas9s to measure non-uniform extension of DNA. (**A**) Overview of using pattern recognition to identify DNA molecules without YOYO-1 fluorescence. **Top**, image of synthetic template used to identify dCas9 bound DNAs. Templates are generated based on the average fluorescent signal of dCas9s in each experiment. **Bottom**, representative image showing how dCas9 populations and bound DNAs are identified using pattern recognition. (**B**) Plot modeling how the extension of tethered λ-DNA molecules changes as force (represented by alpha) increases. Under higher forces, the extension of λ-DNA becomes more uniform. (**C**) Plot showing how tension along measured λ-DNA molecules decreases closer to the free-end of the DNA. (**D**) Histogram showing the maximum tension calculated for representative DNAs, all falling within a normal range (pN) for tethered DNA molecules extended by buffer flow.

The critical parameter for these fits is the force acting on the DNA at each base pair position along its length. The average force we estimate by this method is consistent with previous measurements of tethered DNA under flow (**Figure 31B-D**). Additionally, we find measured forces to be distributed in a narrow range across multiple DNA curtains, and populations of dCas9 to be clustered tightly around their expected positions (**Figure 32**).
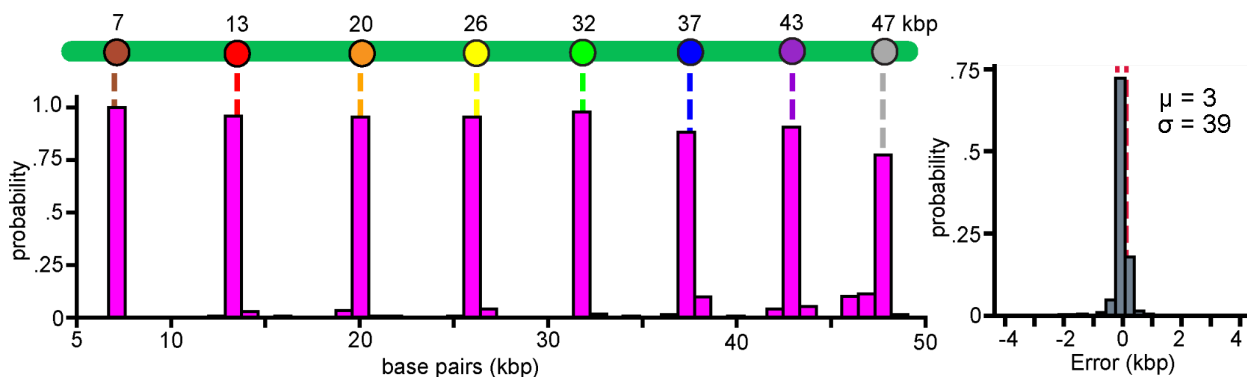
Figure 32. Binding distribution of dCas9s determined by non-uniform extension. **Left**, histogram showing the analyzed distribution of dCas9 positions based on the measured non-uniform extension of each DNA molecule. Multi-colored lines indicate the correct target position of each population. **Right**, histogram displaying the average global error for all dCas9s using this analysis. The mean and standard deviation were determined by fitting a gaussian to the distribution.

However, we found that the last dCas9 population near the free-end of the λ-DNA (47,752 bp) was not described as well by the same parameters that fit the remaining dCas9s. This result is possibly due to the large DNA fluctuations in this region and the likelihood of interacting with adjacent DNA in the curtain, and is why we did not include this dCas9 in our pairwise analysis above (**Figure 29**). Moving forward, we excluded this dCas9 population and removed any binding events fit to this region on the DNA after our analysis.

This new approach will allow us to use dCas9 molecules to more accurately measure the position of DNA bound proteins. To quantify the error of this analysis, we

imaged the remaining 7 fiducial dCas9s labeled with anti-6xHis-Alexa555 to calculate the non-uniform extension of DNAs in one channel. Then we used that information to calculate the positions of 3 experimental dCas9 populations imaged with anti-6xHis-Alexa488 in a second channel (**Figure 33A**). Despite our success utilizing anti-6xHis-Alexa555, we struggled to find another commercially available fluorescent anti-6xHis antibody that worked well at the single molecule level. We settled on using anti-6xHis-Alexa488 (clone HIS.H8, Invitrogen: MA1-21315-A488), which had a relatively poor labeling efficiency under our imaging conditions. Despite having to use the same tag to label distinct dCas9 populations, we were able to minimize dCas9s being labeled with the wrong antibody by subsequently injecting the Alexa488-labeled populations after the more efficiently labeled anti-6xHis-Alexa555 populations were already bound to DNA curtains (**Figure 33B**).

Using this analysis, we were able to identify all 3 populations of dCas9s with high accuracy (**Figure 33C**). On average, the difference between the measured mean positions of dCas9s and their true target sites was approximately 41 base pairs and did not display a consistent offset. Furthermore, the average standard deviation of these populations was 160 base pairs, lower than any other analysis we tried and a full order of magnitude below the error measured using our original approach (**Figure 33D**). Together, these data demonstrate that using dCas9s as fiduciary landmarks makes it possible to achieve similar resolution seen in genomic assays within an active single-molecule experiment.
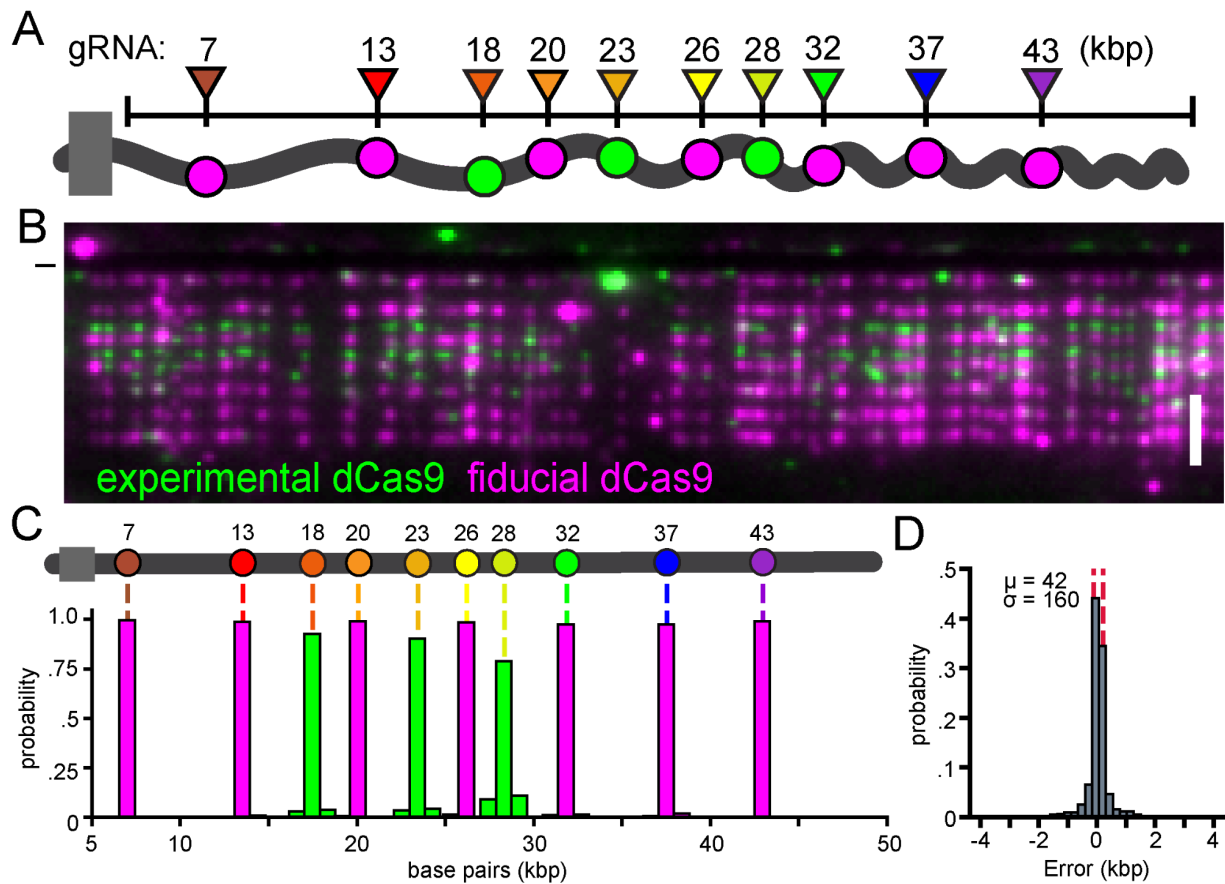
Figure 33. Using dCa9s to fit binding events in a second channel. (**A**) Cartoon indicating how dark DNA molecules (grey) can be simultaneously visualized by fiducial dCas9s (magenta) and used to fit experimental dCas9s (green) imaged in a second fluorescent channel. (**B**) Representative image showing both fluorescent dCas9 populations bound to dark DNA curtains. Fiducial dCas9s (magenta) can be used to measure the non-linear extension of DNA and subsequently calculate the target site of experimental dCas9s (green) in the second channel. (**C**) Histogram showing the analyzed distribution of experimental dCas9 populations (green) based on the measured non-uniform extension of each DNA molecule via dCas9s in a second channel (magenta). Multi-colored lines indicate the correct target position of each population. (**D**) Histogram displaying the average global error for all 3 populations of experimental dCas9s using this analysis. The mean and standard deviation were determined by fitting a gaussian to the distribution.

2.5 Measuring MBD1-MBD binding with increased accuracy

Our low resolution binding distributions were useful for analyzing general trends that could govern MBD binding to methylated DNA. However, the inherent error in low resolution distributions limited our ability to make definitive claims about MBD binding preferences. To further investigate MBD binding at the single molecule level, we applied this new non-uniform extension analysis to query MBD1-MBD binding at high resolution.

Since we could only reliably label and image 6xHis-tagged dCas9s, we moved forward with Qdot-labeled 3xFLAG-tagged MBD1-MBD to further investigate MBD binding. We used fluorescent dCas9s to visualize methylated DNA molecules in one channel and to localize fluorescent MBD1-MBD binding events in a second channel (**Figure 34**). By calculating the position of MBD1-MBDs based on the non-uniform extension of the DNA, we generated a map of Qdot705-labeled MBD1-MBD binding at 10x higher resolution than our initial efforts.
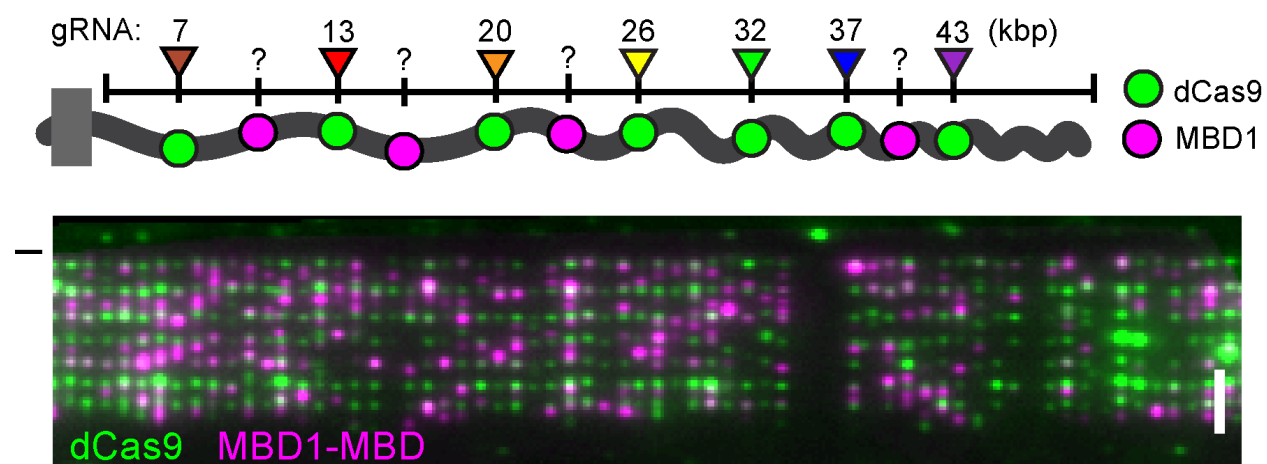
Figure 34. Targeting dCas9s alongside MBD1-MBD on DNA curtains. (**A**) This cartoon shows how dCas9 molecules (green) were targeted alongside MBD1-MBD (magenta) on dark methylated DNA curtains. Color-coded triangles indicate the target site of each fiducial dCas9. (**B**) Representative image showing dark methylated DNA bound by Qdot-labeled 3xFLAG-MBD1-MBDs (magenta) and Alexa555-labeled dCas9s (green). (—) indicates chrome diffusion barriers. Scale bar - 5 μm.

In the absence of YOYO-1 and presence of dCas9s, this new binding distribution of MBD1-MBD still correlated with mCG density, confirming that DNA methylation is the driving factor for stable MBD1-MBD recruitment to DNA (Pearson's r > 0.7, **Figure 35**).
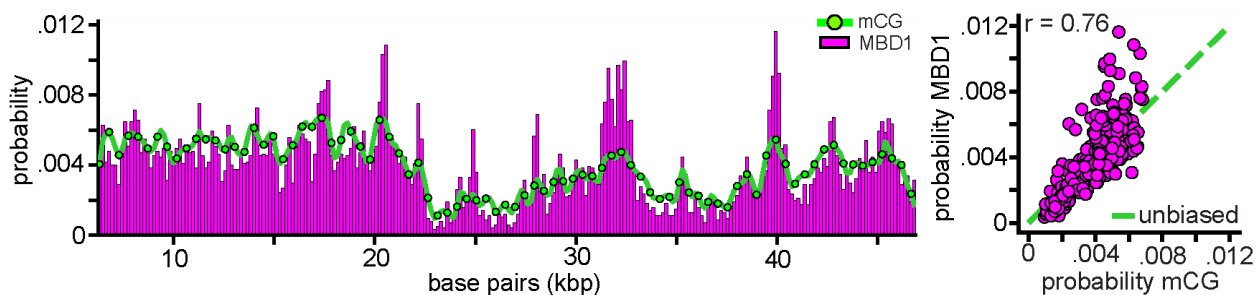
Figure 35. Binding distribution of MBD1-MBD measured using dCas9s.
Left, histogram showing the calculated positions of MBD1-MBD binding events based on the non-uniform extension of each DNA molecule, n = 11440. The light green line indicates the distribution of mCGs across λ-DNA. Data is binned based on the measured error rate of this analysis, σ = 160bp. **Right**, correlation plot showing how the probability of Qdot-labeled MBD1-MBD binding compares to the distribution of mCGs along λ-DNA.

At low resolution, 100pM Qdot-labeled MBD1-MBD appeared to lose mCG-density dependent binding, linearly correlating with the underlying mCG distribution (**Figure 15**). However, the overall correlation with the distribution of mCGs was actually lower when Qdot-labeled MBD1-MBD binding was measured with higher accuracy, and regions of enriched MBD1-MBD binding became clear again (**Figure 35**). This result suggests that the underlying mCG-density dependent binding of MBD1-MBD (**Figures 9/12/13**) was not fully eliminated in the presence of Qdot705 but only impaired, making it difficult to observe this behavior at low resolution. Instead, higher resolution was required to visualize mCG-density dependent binding at this low concentration of Qdot-labeled MBD1-MBD. This finding validated that we could further investigate this aspect of MBD binding using Qdot-labeled MBDs at high resolution.

To explore this mCG-density dependence further, we aimed to identify signals that are found within regions where MBD binding is enriched on DNA curtains. We took this opportunity to further develop our method by assigning measured binding events to specific mCG sites. Because we were unable to observe MBD1-MBD binding on unmethylated DNA curtains, we can assume that each visualized binding event corresponds to an MBD1-MBD bound at a specific mCG. This additional information reduces the number of potential binding sites on λ-DNA from 48502 to just 3113. In practice, we distribute the probability from each MBD1-MBD binding event among local mCG sites in a manner that reflects the localization uncertainty of this approach, previously measured using dCas9s, $\sigma = 160$ (**Figure 33**).

The result of this analysis is a relative occupation probability of MBD1-MBD at each mCG in λ-DNA. When comparing this result to the methylation levels measured by EM-seq, we could see two potential outcomes. First, if MBD1-MBD actually has an equal probability to bind to all mCGs along λ-DNA, then we expect our measured occupancy to equal the true methylation levels. Alternatively, if MBD1-MBD binding integrates additional signals—like mCG density.—we expect to see our measured occupancy deviate from the methylation levels measured by EM-seq.

To differentiate between these possibilities, we numerically determined 95% confidence intervals for occupancy of each mCG in a manner that reflects the true methylation level and our experimental sampling. In a dataset of this size, we can expect that ~10 sites would fall outside of the 95% confidence intervals due to sampling error; however, we find that over half of our measured occupation probabilities fall outside this range (**Figure 36**). In fact, we identify mCG sites that are both over- and

under-represented by MBD1-MBD binding. As a result, we conclude that MBD1-MBD does not bind to all mCGs equally, suggesting the presence of an mCG alone is not necessarily enough to stably recruit MBD1-MBD to DNA.
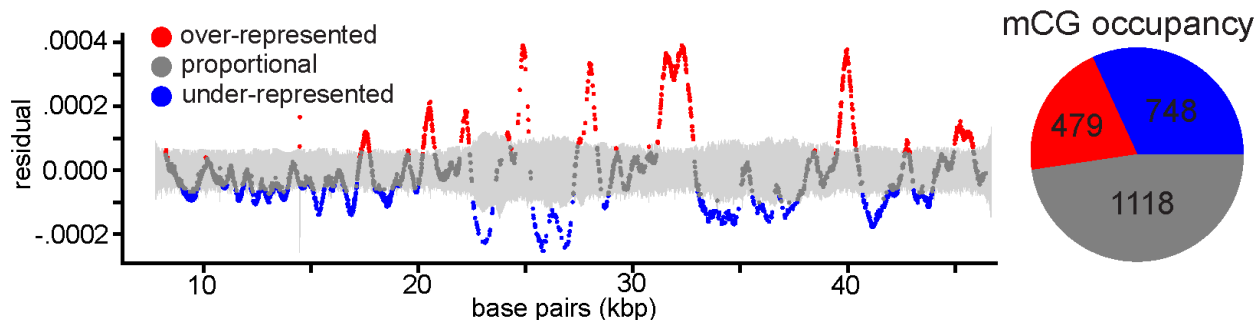
Figure 36. Site-specific probability of MBD1-MBD binding. Left, plot showing the residual probability of MBD1-MBD binding to each individual mCG. The light grey region represents 95% confidence intervals for how likely each site would be bound if all sites were seen equally. Each dot represents an mCG and the probability that MBD1-MBD was found at that site within our measured data set. ● = sites over-represented by MBD1-MBD binding, ● = sites under-represented by MBD1-MBD binding, and ● = sites proportionally represented by MBD1-MBD binding. **Right**, pie chart showing the number of mCGs bound by MBD1-MBD at different frequencies. 2345 out of 3113 mCGs were included after removing mCGs that occur too close to either end of the DNA for efficient visualization..

Based on our low resolution binding distributions, we hypothesized that local mCG density may influence recruitment of MBD1-MBD to DNA. To test this hypothesis, we looked at the flanking DNA sequence surrounding MBD1-MBD binding sites. For this analysis, we calculated the frequency of the four canonical DNA bases as well as methylated cytosines within mCGs. By pooling all of the weighted data, we generated an aggregate map of the average MBD1-MBD binding site within λ-DNA (**Figure 37**). In the whole data set, there was a correlation between more frequent MBD1-MBD binding and neighboring DNA methylation. Beyond analyzing the average signal of all mCG binding sites in λ-DNA, we specifically investigated sequences flanking mCGs either over-represented or under-represented by MBD1-MBD binding.

Average MBD1 Binding Site
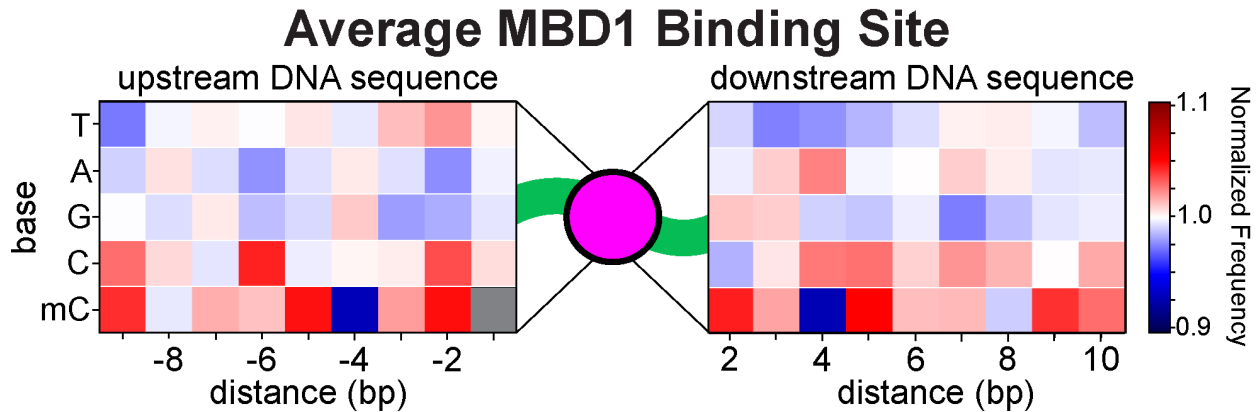
Figure 37. Average flanking sequence of mCGs bound by MBD1-MBD. Heat-map depicting the average frequency of different DNA bases surrounding mCGs bound by MBD1-MBD. mCG sites and their flanking sequences were weighted based on their MBD1-MBD binding frequency. All flanking sequences were normalized by their underlying frequency in λ-DNA. Positions 0 and 1 correspond to the bound mCG.

When we examined the mCGs that were over-represented for MBD1-MBD binding, we found that they were more frequently surrounded by additional mCGs. Similarly, we saw that mCGs under-represented for MBD1-MBD binding were less likely to have neighboring mCGs (**Figure 38A/B**).

Beyond the presence of neighboring sites of DNA methylation, specific spacings between mCGs also correlated with MBD1-MBD binding frequency. mCGs over-represented by MBD1-MBD binding were more frequently flanked by mCGs both 5 and 9 base pairs away. In contrast, sites over-represented by MBD1-MBD binding were actually depleted for mCGs 4 base pairs apart (**Figure 38A/C**). We saw the opposite trends in sites under-represented by MBD1-MBD binding, which were enriched for mCGs 4 base pairs away and less frequently had neighboring mCGs 5 or 9 base pairs away (**Figure 38B/C**). Based on this high-resolution analysis, we were able to map a distribution of MBD1-MBD binding events back to specific sites to uncover additional cues that correlate with higher or lower MBD1-MBD binding frequencies.
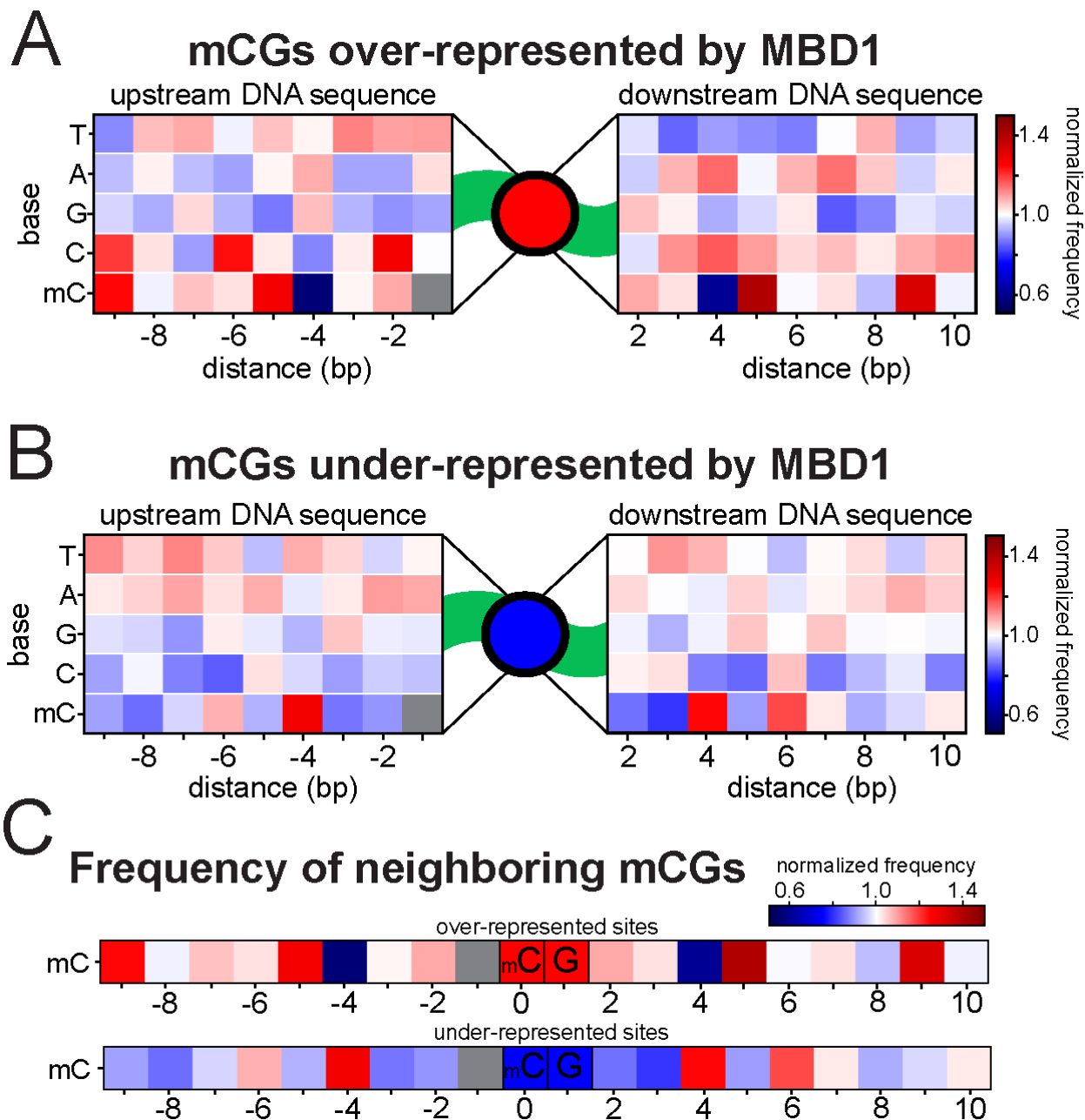
Figure 38. mCG sites under- and over-represented by MBD1-MBD. (**A**) Heat-map depicting the average frequency of different DNA bases surrounding mCGs over-represented by MBD1-MBD binding. (**B**) Heat-map depicting the average frequency of DNA bases surrounding mCGs under-represented by MBD1-MBD binding. (**C**) Heat map showing the frequency of neighboring mCGs at MBD1-MBD sites. **Top**, shows the frequencies of mcGs surrounding mCGs sites over-represented by MBD1-MBD binding. **Bottom**, shows the frequencies of mcGs surrounding mCGs sites under-represented by MBD1-MBD binding.

2.6 MBD1-MBD binding to sparsely methylated DNA

In the mammalian genome, mCGs do not only exist in dense clusters, but MBDs also encounter regions of intermediate and sparsely methylated mCGs. In order to query MBD1-MBD behavior on DNA with more variable mCG density, we tested MBD1-MBD binding on DNA curtains with unique distributions of mCGs. We used λ-DNA methylated with HhaI (5'-GCGC-3') or HpaII (5'-CCGG-3') methyltransferases. Of the 3113 CpGs methylated by M.SssI, only 328 are HpaII sites and only 215 are HhaI sites. To validate full methylation, we tested for protection from HhaI and HpaII restriction enzymes and quantified DNA methylation using bisulfite sequencing (48). Methylation levels were approximately ~95% at all HhaI and HpaII sites.

First, we tested MBD1-MBD binding in the absence of dCas9s and saw stable recruitment along the length of HpaII-methylated DNA (**Figure 39A**). Next, we combined MBD1-MBDs with dCas9s to generate a high-resolution binding distribution of MBD1-MBD on HpaII-DNA (**Figure 39B**). Consistent with our previous experiments, MBD1-MBD binding generally correlated with mCGs. However, once again MBD1-MBD binding was more frequent at relatively dense HpaII sites and less frequent at others (**Figure 39C**). This further supports that MBD1-MBD is differentially recruited to specific regions in a mCG-density dependent manner. Some regions enriched for MBD1-MBD binding on M.SssI-DNA appeared conserved on HpaII-DNA, while other areas of enrichment were lost when only HpaII sites were methylated. Similarly, some relatively dense regions of HpaII sites were not enriched for MBD1-MBD binding, suggesting that cues within dense regions of DNA methylation can further influence MBD1-MBD recruitment.
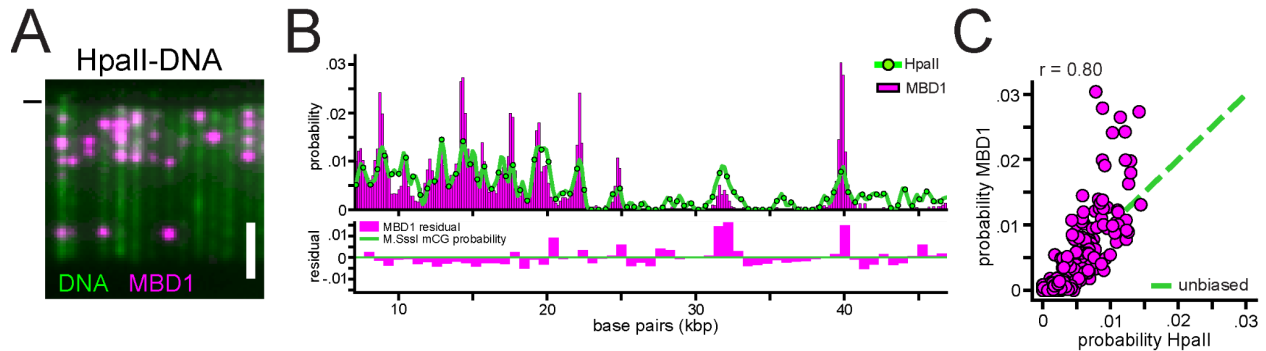
Figure 39. MBD1-MBD binding on HpaII-DNA as measured by dCas9s.
(**A**) Representative image of MBD1-MBD binding events (magenta) on YOYO-1 stained
λ-DNA methylated with HpaII (green). (—) indicates chrome diffusion barriers. Scale bar
- 5 µm. (**B**) Binding distribution of MBD1-MBD on HpaII-methylated DNA as measured
by dCas9s. Light green line indicates distribution of methylated HpaII sites. Relative
MBD1-MBD enrichment on M.SssI-DNA is plotted below to compare binding
distributions. (**C**) Correlation plot of MBD1-MBD binding events with methylated HpaII
sites, Pearson's r = 0.80. Green line indicates a putative 1:1 correlation between both.

We reduced the number of potential binding sites further by testing MBD1-MBD

binding on HhaI-methylated DNA. Unlike the robust binding we see on other mCG

distributions, MBD1-MBD binding events were largely absent on sparsely methylated

HhaI-DNA (**Figure 40A**). We combined MBD1-MBDs with dCas9s on HhaI-DNA to

measure this skewed binding at high-resolution. Only a single region of mCGs were

enriched for MBD1-MBD binding, while all other sites were under-represented for

MBD1-MBD binding (**Figure 40B**). This further confirms that MBD1-MBD is stably

recruited to DNA in an mCG-density dependent manner. The binding distribution on

HhaI-methylated DNA was so heavily biased that MBD1-MBD binding no longer

correlated with mCGs (**Figure 40C**). This high-resolution map identified that

MBD1-MBD enriches at a ~200 base pair region of λ-DNA containing some of the

densest clusters of HhaI sites. MBD1-MBD binding on HhaI-methylated DNA suggests

that changing the distribution of mCGs alone may be sufficient to generate novel

enrichment of MBD1-MBD at specific sites along the same underlying DNA sequence.
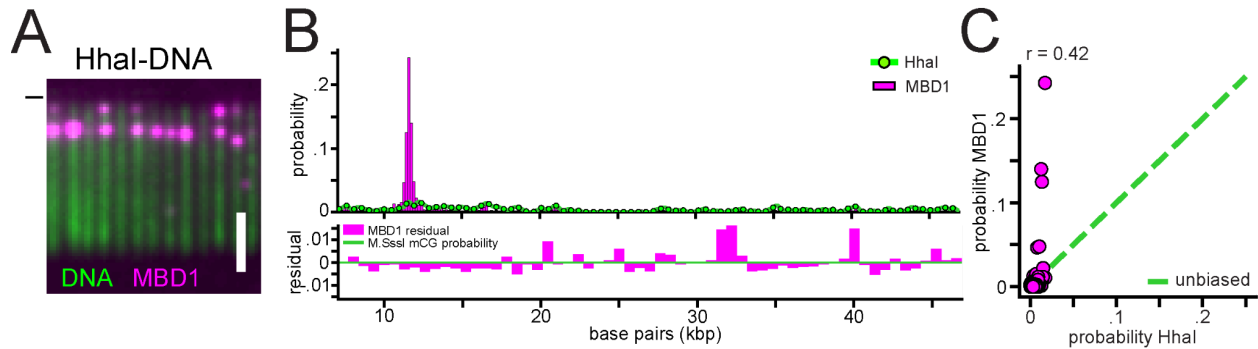
Figure 40. MBD1-MBD binding on HhaI-DNA as measured by dCas9s.
(**A**) Representative image of MBD1-MBD binding events (magenta) on YOYO-1 stained
λ-DNA methylated with HhaI (green). (—) indicates chrome diffusion barriers. Scale bar
- 5 µm. (**B**) Histogram showing binding distribution of MBD1-MBD on HhaI-methylated
DNA as measured by dCas9s. Light green line indicates distribution of methylated HhaI
sites along λ-DNA. Relative MBD1-MBD enrichment on M.SssI-DNA is plotted below to
compare binding distributions. (**C**) Correlation plot of MBD1-MBD binding events
compared to methylated HhaI sites, Pearson's r = 0.42. Green line indicates a putative
1:1 correlation between MBD1-MBD binding and methylated HhaI sites.

The region of MBD1-MBD enrichment on HhaI-methylated DNA was not enriched

on M.SssI-methylated DNA, where the same sites are also methylated. We

hypothesized that this difference is driven by a lack of favorable binding sites at other

positions on sparsely methylated HhaI-DNA which are present on densely methylated

M.SssI-DNA. To test this hypothesis, we investigated MBD1-MBD binding on DNA

which was methylated at both HhaI and HpaII sites. Unlike HhaI-only methylated DNA,

MBD1-MBD was once again robustly recruited along the length of the more densely

methylated HhaI+HpaII-DNA (**Figure 41A**). When measured using dCas9s, the binding

distribution of MBD1-MBD once again correlated with mCGs (**Figure 41B/C**).

MBD1-MBD still bound relatively dense sites at a higher frequency on

HhaI+HpaII-methylated DNA, but the single peak of MBD1-MBD binding seen on

HhaI-only methylated DNA was lost when additional sites were methylated. Despite the

presence of potential HhaI binding sites, the distribution of MBD1-MBD on both

HhaI+HpaII-methylated DNA was largely indistinguishable from MBD1-MBD binding on

HpaII-methylated DNA. This result suggests that, while MBD1-MBD is capable of stably

binding HhaI sites, their appearance within λ-DNA is particularly unfavorable for

MBD1-MBD binding.
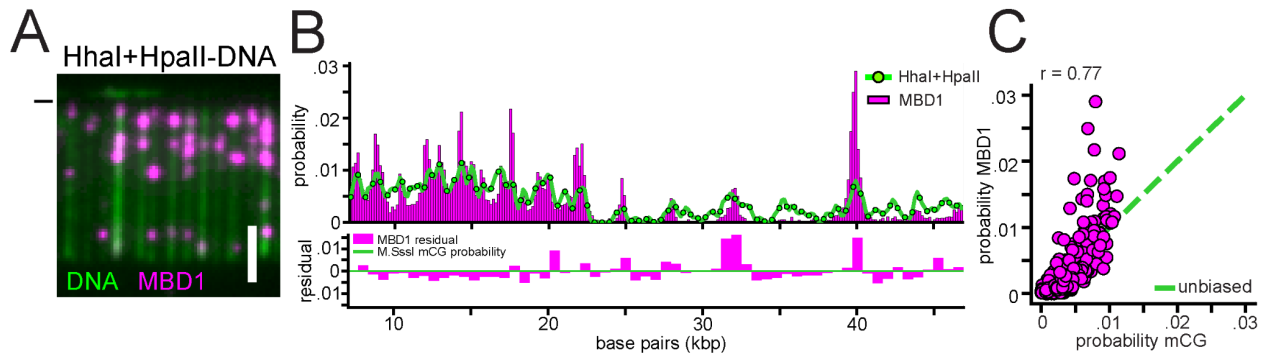
Figure 41. MBD1-MBD binding on HhaI+HpaII-DNA as measured by dCas9s.
(**A**) Representative image of MBD1-MBD binding events (magenta) on YOYO-1 stained
λ-DNA methylated with both HhaI and HpaII (green). (—) indicates chrome diffusion
barriers. Scale bar - 5 µm. (**B**) Binding distribution of MBD1-MBD on
HhaI+HpaII-methylated DNA as measured by dCas9s. Light green line indicates
distribution of methylated HhaI and HpaII sites. Relative MBD1-MBD enrichment on
M.SssI-DNA is plotted below to compare binding distributions. (**C**) Correlation plot of
MBD1-MBD binding events compared with methylated HhaI and HpaII sites, Pearson's r
= 0.77. Green line indicates a putative 1:1 correlation between MBD1-MBD binding and
methylated HhaI and HpaII sites.

Our data from multiple unique DNA methylation distributions shows that the

presence of CpG methylation is not the sole factor leading to MBD1-MBD recruitment

on DNA curtains. Instead, enrichment of MBD1-MBD binding consistently occurs at

relatively dense clusters of mCGs.

2.7 High resolution analysis reveals similar binding of MeCP2-MBD

With a reliable method for high-resolution mapping of binding events on DNA curtains in place, we wanted to explore if any binding cues governing MBD1-MBD binding were generalizable to other MBD family members. Using Alexa555-labeled dCas9s to measure non-uniform extension of DNA limited us to imaging MBD binding events with Qdot705. We returned to imaging 3xFLAG-tagged mouse MeCP2-MBD binding on M.SssI-methylated DNA curtains using anti-FLAG-Qdot705.
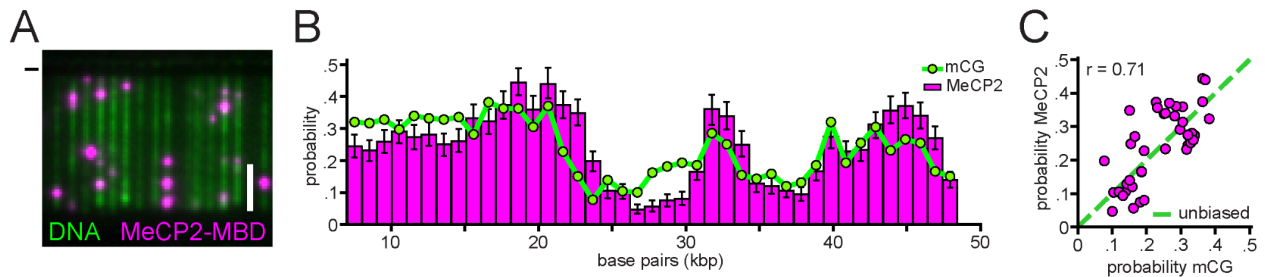
Figure 42. Qdot-labeled MeCP2-MBD low resolution binding on DNA curtains.
(**A**) Representative images of M.SssI-methylated DNA (green) bound by Qdot-labeled 3xFLAG-MeCP2-MBD (magenta). (—) indicates chrome diffusion barriers. Scale bar - 5 μm (**B**) Histogram showing the distribution of MeCP2-MBD binding events measured from the free-end of bound DNA molecules. Light green line indicates the mCG distribution. (**C**) Correlation plot comparing MeCP2-MBD binding distribution with the underlying mCG distribution. Green dashed line indicates a theoretical linear trend.

First, we observed stable binding of Qdot-labeled MeCP2-MBD along the length of M.SssI-methylated DNA stained with YOYO-1 (**Figure 42A**). We calculated the position of MeCP2-MBD binding events based on their distance from the free-end of the DNA to generate a low resolution distribution of Qdot-labeled MeCP2-MBD, which largely correlated with the underlying mCG distribution (**Figure 42B/C**). However, compared to Qdot-labeled MBD1-MBD, the mCG-density dependent binding of MeCP2-MBD (**Figure 9/12**) seemed less suppressed when labeled with Qdot705 and imaged at low resolution since some variability from the underlying mCG distribution

remained evident. Regardless, we anticipated any mCG-density dependent binding would be clearer when analyzing MeCP2-MBD binding at high resolution.
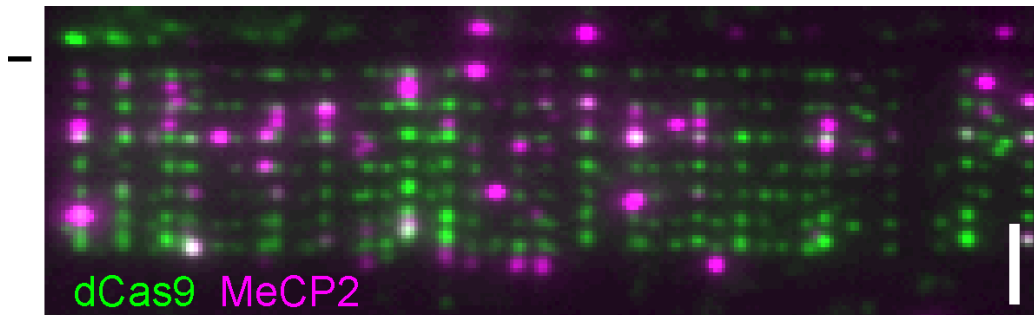
Figure 43. Targeting MeCP2-MBD to methylated DNA curtains alongside dCas9s. Representative image depicting dark M.SssI-methylated DNA bound by Qdot-labeled MeCP2-MBD (magenta) alongside fiducial dCas9s (green). (—) indicates chrome diffusion barriers. Scale bar - 5 µm

Next, we successfully targeted Alexa555-labeled dCas9s to dark methylated DNA curtains alongside Qdot-labeled MeCP2-MBD (**Figure 43**). Using the non-uniform extension of DNA to measure Qdot-labeled MeCP2-MBD binding events created a distribution which revealed clear enrichment of MeCP2-MBD at densely methylated regions along methylated DNA (**Figure 44A**).
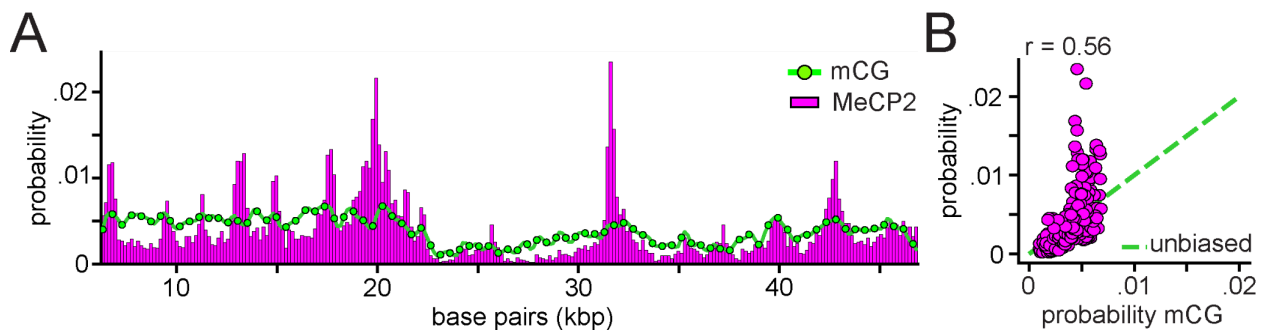
Figure 44. Binding distribution of MeCP2-MBD measured using dCas9s.
(**A**) Histogram showing the calculated positions of MeCP2-MBD binding events based on the non-uniform extension of each DNA molecule as measured by dCas9s, n = 9247. The light green line indicates the distribution of mCGs across λ-DNA. Data is binned based on the measured error rate of this analysis, σ = 160bp. (**B**) Correlation plot showing how the probability of Qdot-labeled MBD1-MBD binding compares to the distribution of mCGs along λ-DNA. Green line indicates a theoretical linear trend.

The distribution of MeCP2-MBD actually deviated more significantly from the underlying DNA methylation distribution than MBD1-MBD. In fact, MeCP2-MBD binding only weakly correlated with the distribution of DNA methylation since many sites appeared enriched or depleted by MeCP2-MBD binding (Pearson < 0.7, **Figure 44B**). Despite having very similar binding distributions at low-resolution (**Figure 12B**), MeCP2-MBD and MBD1-MBD create distinct binding distributions when imaged at high-resolution with dCas9s. Even with these differences, both readers share an intrinsic preference for some of the same densely methylated regions, further demonstrating that mCG-density dependent binding is a conserved feature across multiple MBD domains.
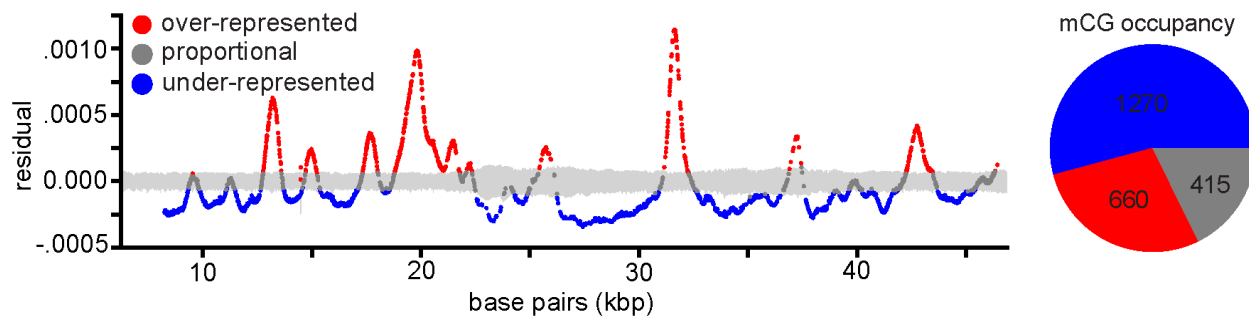
Figure 45. Site-specific probability of MeCP2-MBD binding. Left, plot showing the residual probability of MeCP2-MBD binding to each mCG. The light grey region represents 95% confidence intervals for how likely each site would be bound if all sites were seen equally. Each dot represents an mCG and the probability that MeCP2-MBD was found at that site. ● = sites over-represented by MeCP2-MBD binding, ● = sites under-represented by MeCP2-MBD binding, and ● = sites proportionally represented by MeCP2-MBD binding. **Right**, pie chart showing the number of mCGs bound by MeCP2-MBD at different frequencies. 2345 out of 3113 mCGs were included after removing mCGs that occur too close to either end of the DNA.

To explore the additional cues that lead to MeCP2-MBD enrichment, we created a weighted distribution by assigning these binding events to specific mCGs, like we did for MBD1-MBD. This analysis revealed that the vast majority of mCGs were either over-represented or under-represented by MeCP2-MBD binding (**Figure 45**). Next, we

analyzed the frequencies of flanking sequences at MeCP2-MBD binding sites. Like

MBD1-MBD, on average, there was a general correlation between more frequent

MeCP2-MBD binding and neighboring DNA methylation (**Figure 46**).
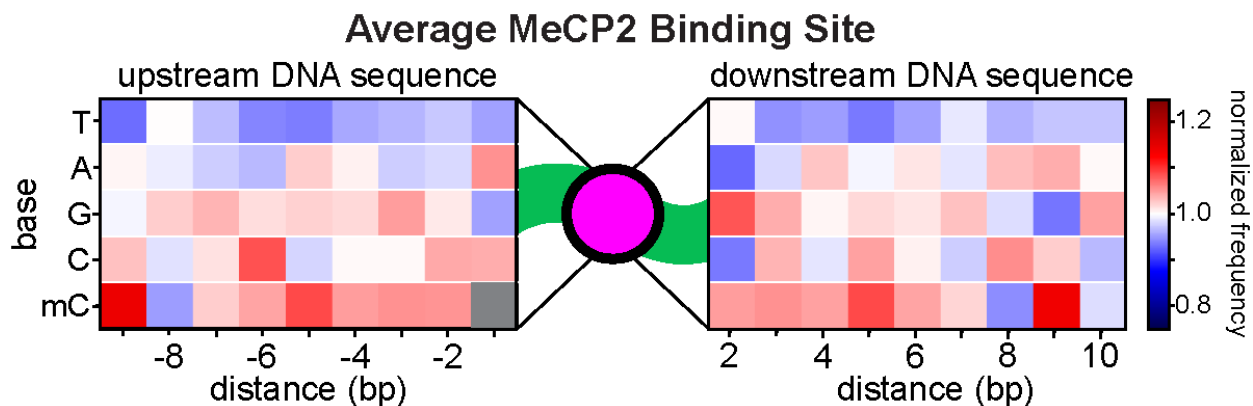
Figure 46. Average flanking sequence of mCGs bound by MeCP2-MBD. Heat-map depicting the average frequency of different DNA bases surrounding mCGs bound by MeCP2-MBD. mCG sites and their flanking sequences were weighted based on their MeCP2-MBD binding frequency. All flanking sequences were normalized by their underlying frequency in λ-DNA. Positions 0 and 1 correspond to the bound mCG.

Sites over-represented by MeCP2-MBD binding were more frequently

surrounded by additional mCGs, while mCGs under-represented by MeCP2-MBD were

less likely to have neighboring mCGs (**Figure 47A/C**). Once again we found that

specific spacings of neighboring mCGs correlated with MBD binding frequencies. Unlike

MBD1-MBD, no particular spacing between neighboring mCGs appeared negatively

correlated with MeCP2-MBD binding. However, we found that MeCP2-MBD binding was

strongly correlated with sites that had neighboring mCGs 5 and 9 base pairs away,

which was also true for MBD1-MBD. Sites under-represented by MeCP2-MBD binding

were not enriched for any particular mCG spacing, but they were rarely flanked by a

neighboring mCG 9 base pairs away (**Figure 47B/C**).
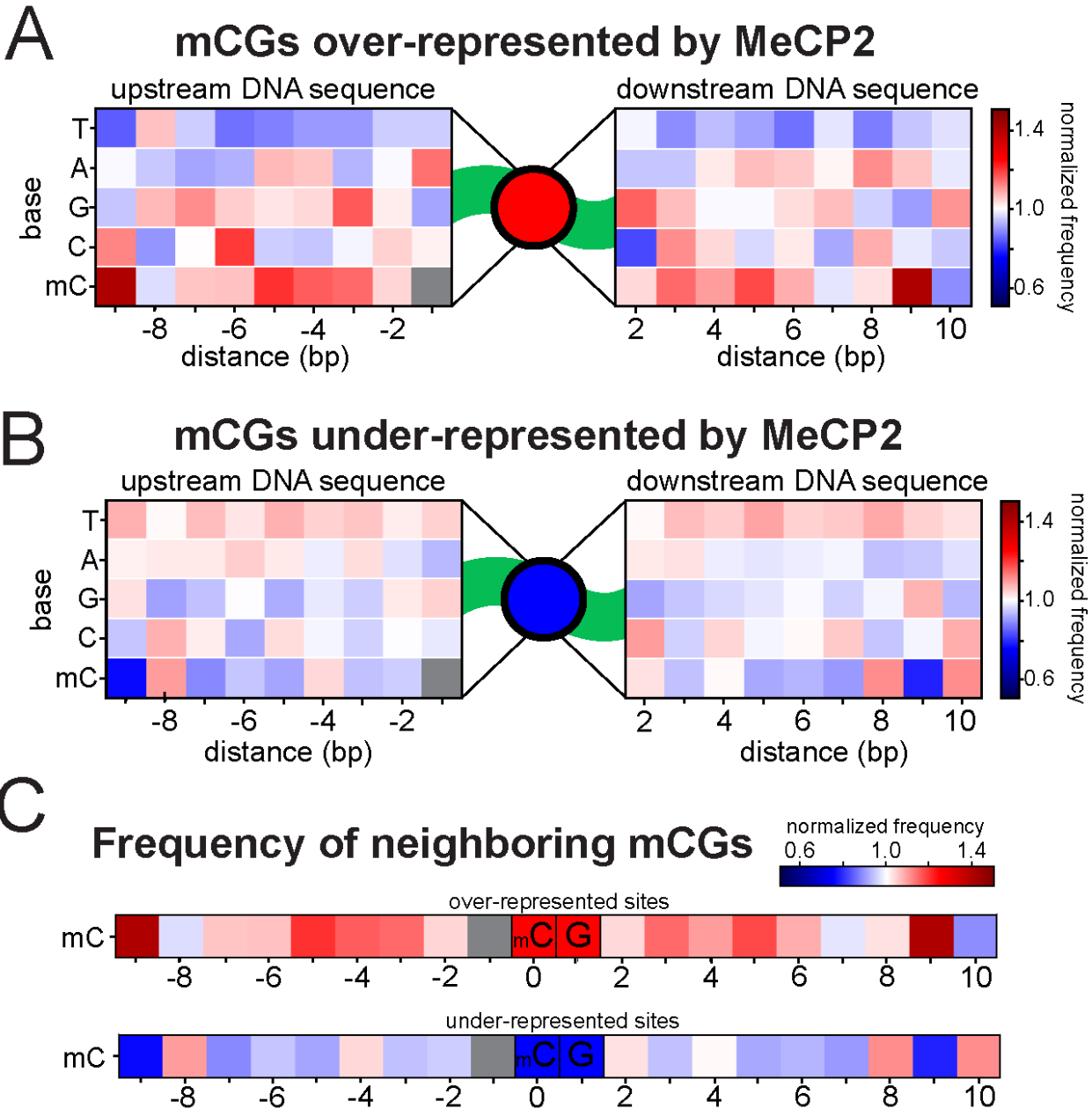
Figure 47. Average flanking sequence of mCGs under- and over-represented by MeCP2-MBD. (**A**) Heat-map depicting the average frequency of different DNA bases surrounding mCGs over-represented by MeCP2-MBD binding. (**B**) Heat-map depicting the average frequency of DNA bases surrounding mCGs under-represented by MeCP2-MBD binding. (**C**) Heat map showing the frequency of neighboring mCGs at MeCP2-MBD sites. **Top**, shows the frequencies of mcGs surrounding mCGs sites over-represented by MeCP2-MBD binding. **Bottom**, shows the frequencies of mcGs surrounding mCGs sites under-represented by MeCP2-MBD binding.

This high-resolution analysis shows that both MBDs from MBD1 and MeCP2 can create distinct binding distributions on the same methylated DNA substrate. While both MBDs appear to respond to unique cues, they share the ability to preferentially bind densely methylated regions at sites with specific neighboring mCGs. In particular, both MBD1-MBD and MeCP2-MBD frequently bound sites 9 base pairs away from neighboring mCGs, suggesting that this may be a conserved signal influencing binding of MBD family members.

2.8 Measuring MBD binding to differentially spaced mCGs

To further test this conserved signal, as well as how other mCG spacings affect MBD binding, we returned to EMSAs, which are useful for directly comparing MBD binding between specific conditions. We initially tested 3xFLAG-MBD1-MBD and 3xFLAG-MeCP2-MBD binding to DNA hairpins that contained two neighboring mCGs (**Figure 48**). While we kept the flanking sequence of each mCG conserved, we altered the number of DNA base pairs in between each binding site.
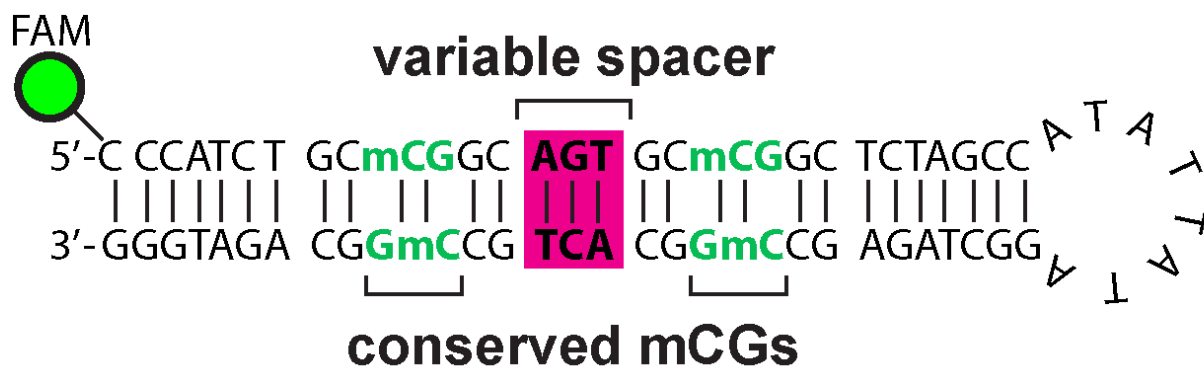
Figure 48. DNA hairpin substrate for MBD binding in gel shifts. Diagram depicts the sequence of DNA used to form dsDNA hairpins. Upon annealing to itself, this DNA molecule will create two symmetrical mCGs, with conserved flanking bases (green). DNA can be added or removed to the internal region to vary the spacing between neighboring mCGs (magenta). DNA hairpins were FAM-labeled to visualize gel shifts.

We hypothesized that the measured preference for mCGs 9 base pairs apart may arise from two mCGs approximately one helical turn away from one another. To explore this possibility, we tested MBD1-MBD's ability to stably bind DNA with pairs of mCGs falling within a helical turn (<9 base pairs), around a helical turn (9-10 base pairs), and beyond a helical turn (>10 base pairs).

We initially observed that 3xFLAG-MBD1-MBD strongly favored binding to mCGs that were positioned 9 or 10 base pairs away from one another (**Figure 49A**). However, despite robust recruitment on methylated DNA curtains, we were unable to observe stable binding of 3xFLAG-MBD1-MBD to any other pair of mCGs or DNA containing a single mCG site. Similarly, we could not visualize any recruitment of 3xFLAG-MeCP2-MBD to methylated DNA in a gel shift, even when two neighboring mCGs were separated by 9 base pairs of DNA (**Figure 49B**).

To confirm that 3xFLAG-MBD1-MBD and 3xFLAG-MeCP2-MBD are both truly capable of binding mCGs, we performed a super-shift using the same FLAG antibody that we used to label MBDs on DNA curtains. We hypothesized that anti-FLAG may relieve potential inhibition from the negatively charged 3xFLAG-tag to rescue mCG binding. We found that antibody bound 3xFLAG-MBD1-MBD and 3xFLAG-MeCP2-MBD were able to stably bind and shift DNA containing all pairs of mCGs (**Figure 50A/B**). Additionally, in a super-shift, both 3xFLAG-tagged MBDs stably bound DNA with a single mCG but did non-specifically bind unmethylated DNA. However, under these binding conditions, we could not measure a clear preference for mCGs 9 or 10 base pairs apart in an EMSA. Additionally, these super-shifts produced multiple shifted species which we could not properly identify, complicating any analysis.
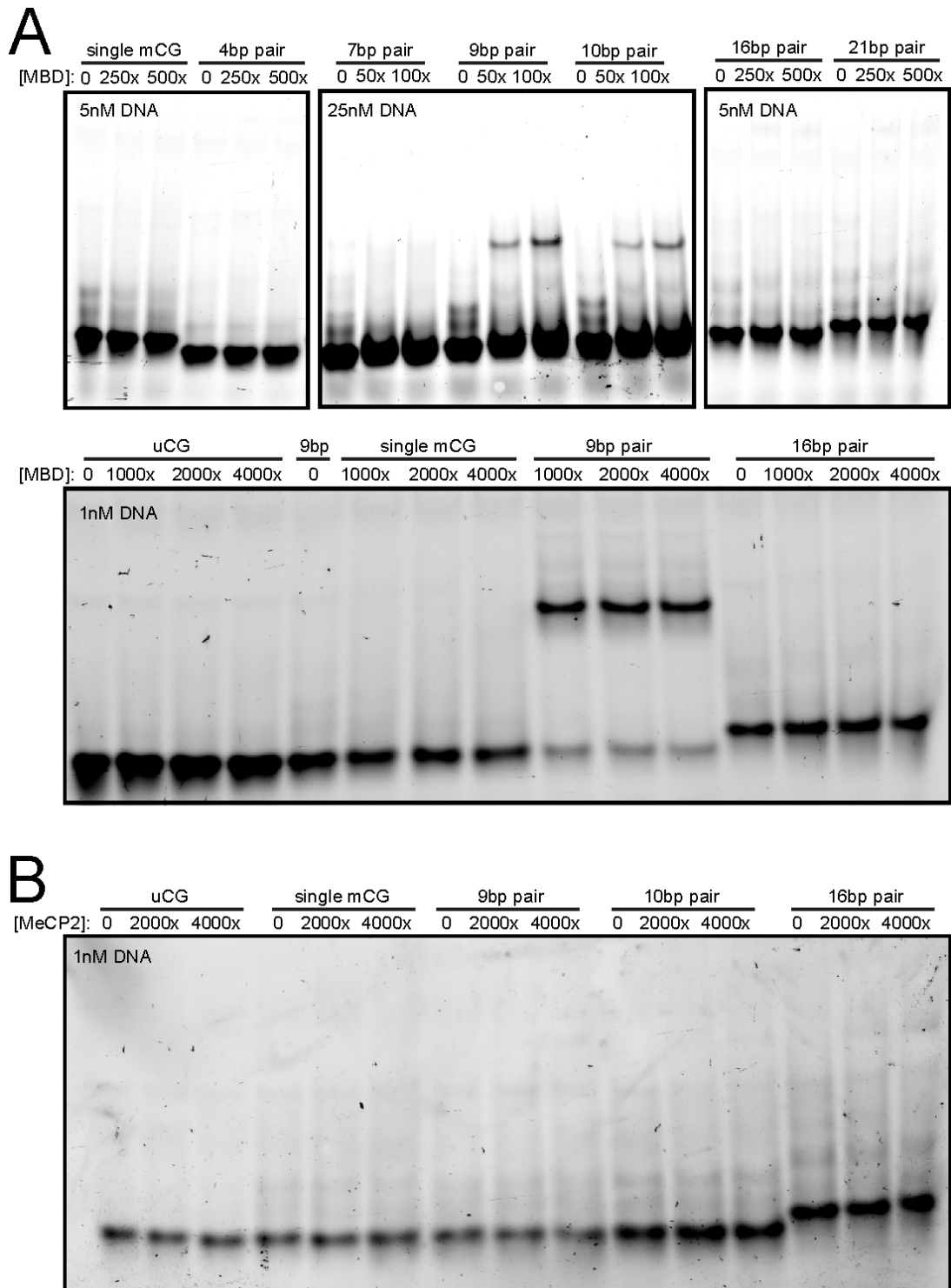
Figure 49. EMSAs with 3xFLAG-tagged MBD1-MBD and 3xFLAG-MeCP2-MBD.
(**A**) Representative images of EMSAs run on differentially modified hairpin DNAs incubated with molar excess 3xFLAG-MBD1-MBD:[DNA]. (**B**) Representative gel shift of differentially modified hairpin DNAs after incubation with molar excess 3xFLAG-MeCP2-MBD:[DNA].
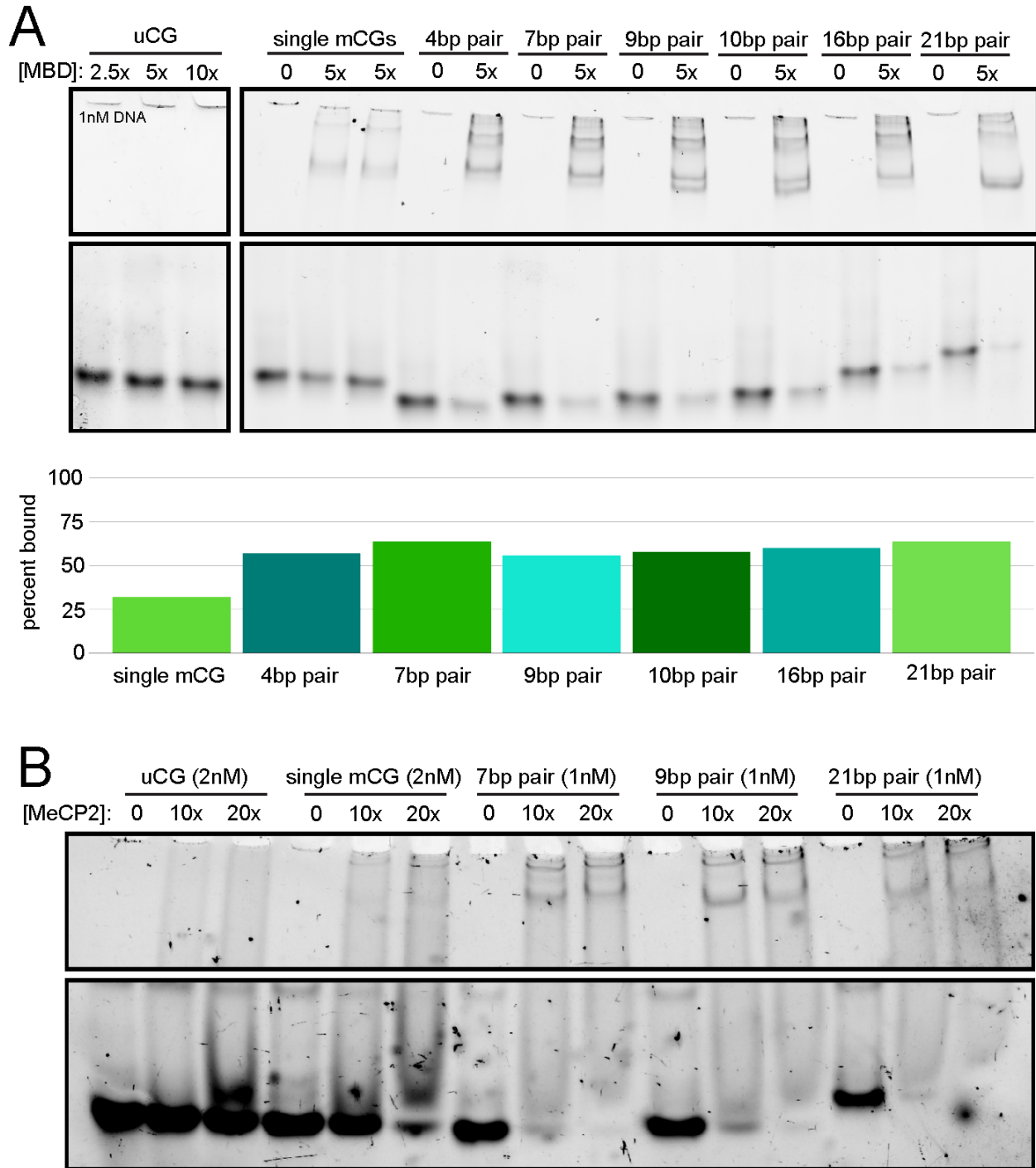
Figure 50. 3xFLAG-tagged MBD1-MBD and MeCP2-MBD supershifts.
(**A**) Representative image of 3xFLAG-MBD1-MBD supershift on differentially modified DNA hairpin substrates. Bar graph shows quantification of DNA bound after incubating with molar excess antibody-labeled MBD1-MBD:[DNA]. (**B**) Representative image of 3xFLAG-MeCP2-MBD supershift on differentially modified hairpin DNA. Molar excess indicates [MeCP2-MBD]:[mCG].

Next, we tested 1xFLAG-MBD1-MBD in an EMSA and found that it was able to weakly bind DNA with a single mCG (**Figure 51A**), further suggesting that the FLAG-tags do negatively impact MBD binding under these conditions. We hypothesized that a super-shift with 1xFLAG-MBD1-MBD may also enhance binding to methylated DNA. Furthermore, we reasoned that this supershift may not produce multiple shifted bands since there are no additional FLAG-tags that can recruit multiple antibodies. In a supershift, we found that 1xFLAG-MBD1-MBD robustly bound DNA containing all pairs of mCGs without non-specifically binding to unmethylated DNA (**Figure 51B**). However, binding to a single mCG remained unstable in these conditions, resulting in smears.
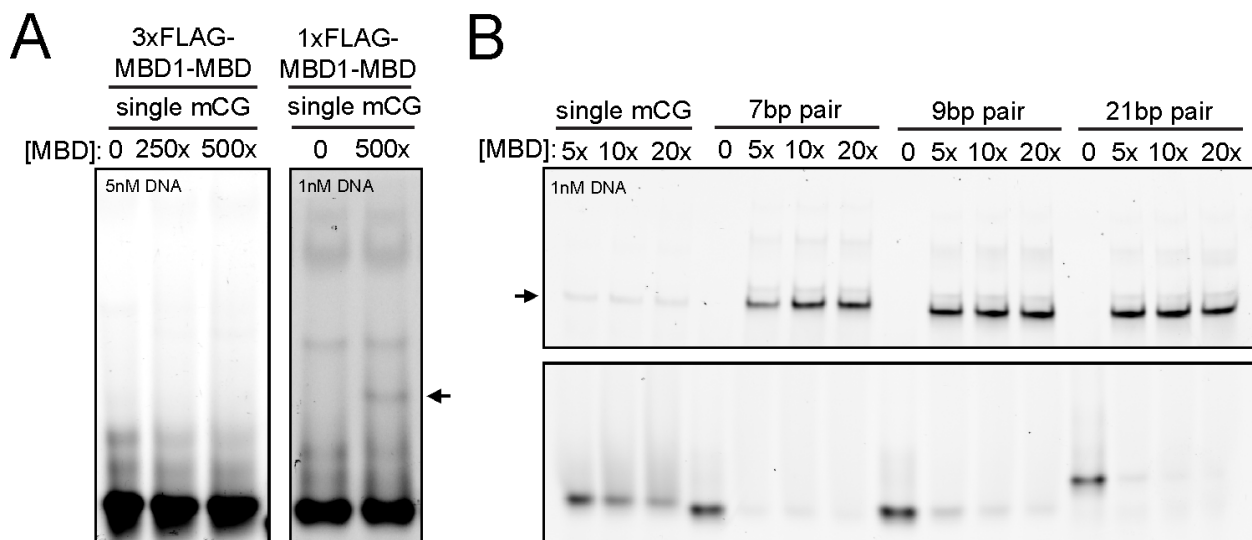
Figure 51. 1xFLAG-tagged MBD1-MBD EMSAs. (**A**) Representative image of 3xFLAG-MBD1-MBD and 1xFLAG-MBD1 both incubated with DNA containing a single symmetrical mCG binding site. Arrow indicates weak binding by 1xFLAG-MBD1-MBD. Molar excess indicates [MBD1-MBD]:[DNA]. (**B**) Representative images showing 1xFLAG-MBD1-MBD supershift on differentially modified DNA hairpins. Arrow indicates gel shift from stable binding at a single mCG. Molar excess = [MBD1-MBD]:[DNA].

The fact that MBD1-MBD binding was stable on DNA with two mCGs suggests that a second mCG at any distance can have a stabilizing effect since there was no observable preference for mCGs 9 or 10 base pairs apart. Additionally,

1xFLAG-MBD1-MBD supershifts produced a single dominant shifted species on all methylated DNA substrates. This result indicates that a second MBD1-MBD binding event is not likely stabilizing MBD binding at neighboring mCGs.

To further test whether MBD1-MBD does prefer binding to mCGs that occur a helical turn away from one another, we utilized 6xHis-tagged MBD1-MBD, which previously bound methylated DNA in a gel shift (**Figure 5**). Like on DNA curtains, we only saw stable recruitment of 6xHis-MBD1-MBD to DNA when it was methylated (**Figure 52**). However, compared to the concentration of MBDs required for supershifts, 6xHis-MBD1-MBD would only weakly bind methylated DNA in these EMSAs
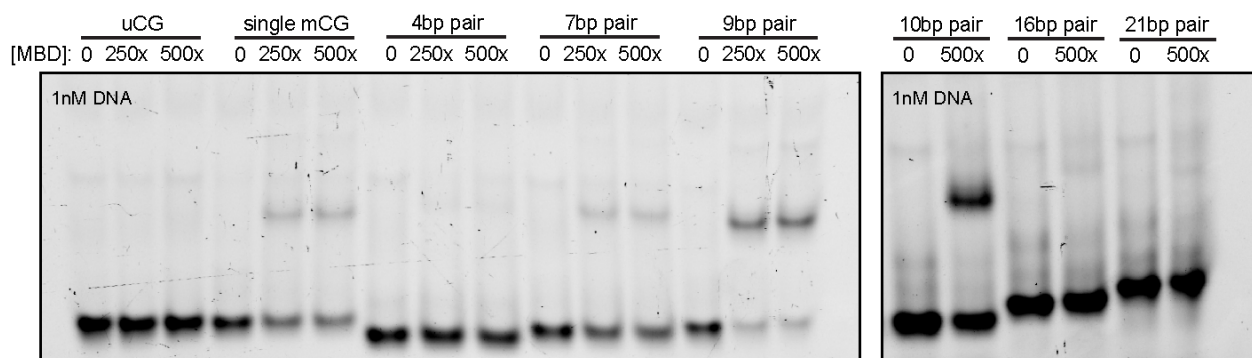
Figure 52. 6xHis-tagged MBD1-MBD EMSAs on DNA hairpins. Representative images showing gel shifts of differentially methylated DNA hairpins after incubation with 6xHis-tagged MBD1-MBD. In both cases, molar excess [MBD1-MBD]:[DNA]. On the left gel, 1nM DNA is achieved by mixing 500pM FAM-DNA with 500pM dark DNA. For all pairs of mCGs, 500pM methylated FAM-DNA is combined with 500pM of uCG DNA.

Like 3xFLAG-tagged MBD1-MBD, 6xHis-MBD1-MBD did not bind all DNAs with two mCGs equivalently. 6xHis-MBD1-MBD shifted a higher proportion of DNA with mCGs separated by 9 or 10 base pairs than other distances (**Figure 52**). This preference suggests that MBD1-MBD binding is not only sensitive to the amount of DNA methylation, but also the context in which it appears. Like what we observed in 1xFLAG-MBD1-MBD supershifts, the migration of DNA bound by 6xHis-MBD1-MBD did not change whether one or two mCGs were present. This lack of higher order species

suggests that the preference for two neighboring mCGs is independent of cooperativity between multiple MBD proteins.

By testing MBD1-MBD binding in bulk, we showed that the measured correlation between MBD occupancy and mCGs separated by 9 base pairs can lead to enhanced MBD1-MBD binding (**Figure 49/52**). However, this result was inconsistent across different binding conditions (**Figure 50/51**). Future work should be done to further validate that MBD spacing can affect MBD recruitment to DNA. When we were able to observe this behavior in gel shifts, MBD's intrinsic preference for neighboring mCGs was not unique to sites that are explicitly 9 base pairs apart, but extended to other distances around one helical turn of DNA. Additionally, we observed that enhanced MBD1-MBD binding at neighboring mCGs can be achieved without recruitment of multiple MBDs, suggesting something about this DNA substrate may create a uniquely favorable binding interaction for a single MBD. This kind of spacing may be an underlying signal recognized by MBDs in densely methylated regions throughout the genome.

3. Discussion

As previously published, we have shown that DNA curtains could be used to study MBD preferences for methylated DNA (26). MBD binding to DNA curtains shows a conserved preference for relatively dense regions of DNA methylation. To further study this binding behavior, we developed a single molecule system to generate high-resolution maps of DNA-protein interactions using fluorescent microscopy. We utilized this new technique to study how MBDs interrogate a complex distribution of DNA methylation. By using dCas9 technology, we were able to map a binding distribution of MBDs to specific sites of DNA methylation. This advancement made it possible to identify novel cues that lead to stable binding of MBD proteins.

3.1 Increasing resolution by using dCas9s as fiducial markers on DNA

Traditionally, constraints from the diffraction limit of light impede our ability to make high resolution measurements using fluorescent microscopy. To overcome this limitation, we used dCas9s to provide additional spatial information. We showed that we could robustly target dCas9s to label specific sites along the length of a DNA curtain. Additionally, we leveraged previous work defining the non-uniform extension of tethered DNA molecules in solution. Combining the site specificity of dCas9s with the known physical parameters of flow-stretched DNA allowed us to fit thousands of binding events with nanometer accuracy (**Figures 31/32/33**). While we applied this analysis to readers of DNA methylation, it can be used to study any stable DNA-protein interaction. Using this method to analyze a DNA-binding protein without any well-known binding preferences will allow us to better identify signals that lead to stable recruitment to DNA.

Additionally, we showed how known recruitment signals, like MBDs' dependence on mCGs, can be leveraged to refine a measured binding distribution. This additional step was possible because the site specificity of dCas9s allowed us to quantify the error rate of this approach. This parameter enabled us to distribute the signal of each measured binding event among all local mCGs, allowing us to create a list of binding probabilities for MBD occupancy at each potential mCG site (**Figures 36/45**). While this statistical power is limited by the number of potential binding sites as well as the resolution required to resolve them from one another, this method could be applied to any protein with known binding preferences to identify additional DNA features that may enhance or inhibit stable binding. The data generated from this kind of analysis could make it possible to identify how proteins may differentiate between many seemingly similar sites throughout the genome.

Additionally, this high-resolution mapping could be further utilized to study other aspects of DNA methylation. Here, we used a known distribution of DNA methylation to uncover unknown binding preferences of MBDs. Alternatively, employing an MBD with known binding preferences would allow us to map an unknown distribution of DNA methylation on DNA curtains. We would not be the first to use MBDs to visualize a distribution of DNA methylation (49-53). In fact, mutations which allow MBDs to bind methylated DNA without any bias have already been developed for other biotechnological applications (54). However, analyzing the binding of these MBDs on DNA curtains using our non-uniform extension analysis would allow us to map an underlying distribution of DNA methylation with high accuracy using fluorescent microscopy. Fitting fluorescent MBDs with high precision would also allow us to monitor

how readers or writers create and modify distributions of DNA methylation on DNA curtains. Furthermore, if we could create a DNA curtain using genomic material from biological samples, we could even use this technique and fluorescent MBDs to visually sequence the distribution of DNA methylation across a population of cells. With the development of this method, future applications of this analysis will provide a deeper understanding of many binding distributions on DNA curtains.

3.2 Limitations

While the high resolution analysis we established here is a powerful tool for improving fluorescence imaging of DNA-protein interactions, several constraints do restrict its implementation. Many other single molecule assays utilize fluorescence to image protein bound DNA, and using dCas9s as markers of specific sequences could improve the resolution of these approaches. However, the non-uniform extension analysis we developed relies on the force-extension of tethered DNA molecules extended under buffer flow. This specific condition limits the widespread use of this approach in other fluorescence imaging systems. Other setups may not require overcoming non-uniform extension of DNA like we had to on DNA curtains. However, applying this approach to these methods, which rely on unique ways of capturing and imaging DNA, may require generating entirely new models to analyze DNA molecules in solution with this approach.

Unlike traditional sequencing experiments which destroy the underlying DNA in order to map binding events and modifications, our approach preserves the underlying DNA in the context of an active single molecule experiment. This advantage allows us to image binding distributions on the same molecule over time and under different

conditions. By doing so with dCas9s, we can achieve the same resolution as traditional

genomic sequencing assays without destroying the DNA. However, gaining this

resolution in this context comes with several trade-offs complicating the downstream

analysis. While DNA curtains are a comparatively high-throughput single molecule

approach, we cannot feasibly obtain millions of reads which are obtained in sequencing

datasets. This limitation comparatively restricts the statistical power of our analysis.

While the error of our fits approaches the same resolution of techniques like

ChIP-seq, the error is uniquely propagated in our approach. Despite the majority of

measurements being localized somewhere within ± 160bp of their true binding site, this

assumption will frequently not be true. Based on the standard deviation of our

distributions, nearly a third of the time, our measurements will be mislocalized at an

even greater distance. This means that many of the ~160bp windows where we

measure binding events actually have none, increasing the noise of our measured

distributions. This kind of false positive is very similar to non-specific binding that affects

approaches like ChIP-seq, which will increase noise in those datasets as well. However,

in traditional genomic sequencing assays, this background is filtered out during analysis

and peak calling algorithms measure binding above this noise. We are unable to filter

out severely mislocalized binding events from our datasets, meaning our signal must be

even stronger in order to distinguish it from noise.

Additionally, we showed that measured binding distributions can be enhanced

given our known error rate, allowing us to map individual binding events back to specific

sites. Unfortunately, this step is only possible if DNA-binding proteins respond to very

specific underlying DNA sequences that govern their recruitment to DNA, like MBDs

binding to methylated CpG dinucleotides or dCas9s binding to sites followed by PAMs. Additionally, even if such a signal exists for a particular DNA-binding protein, the context of the underlying distribution will restrict our ability to measure occupancy at individual sites. This constraint comes from too few or too many potential binding sites underlying the measured binding distribution.

If there are too many potential binding sites, they may be clustered together so tightly (<160bp) that resolving them from one another becomes difficult. In order to accurately compare binding at many nearby sites, it would require even greater numbers of measured binding events. Studying a protein with fewer potential binding sites can also be problematic. While fewer binding sites would be farther apart and more easily resolved from one another, it would limit our ability to measure a conserved binding signal only found in a small subset of sites. For example, we were not able to predict the gRNAs of dCas9 populations fit using fiducial dCas9s (**Figure 53**). Despite the overall accuracy of each measured distribution, bootstrapping each dCas9 population to identify the underlying PAM at the true target site was not possible. This outcome demonstrates that many bound sites with a conserved signal are required to overcome the noise of other potential binding sites along the DNA. This limitation is why we did not map MBD binding back to specific mCGs on HpaII-DNA or HhaI-DNA.

On M.SssI-methylated DNA, there were enough sites to measure a conserved binding signal. However, the shear density of mCGs made us question our ability to accurately map MBDs on this distribution of DNA methylation. To validate that this analysis worked, we analyzed the distribution of mCGs within MBD1-MBD enriched regions without mapping each binding event back to a specific site.
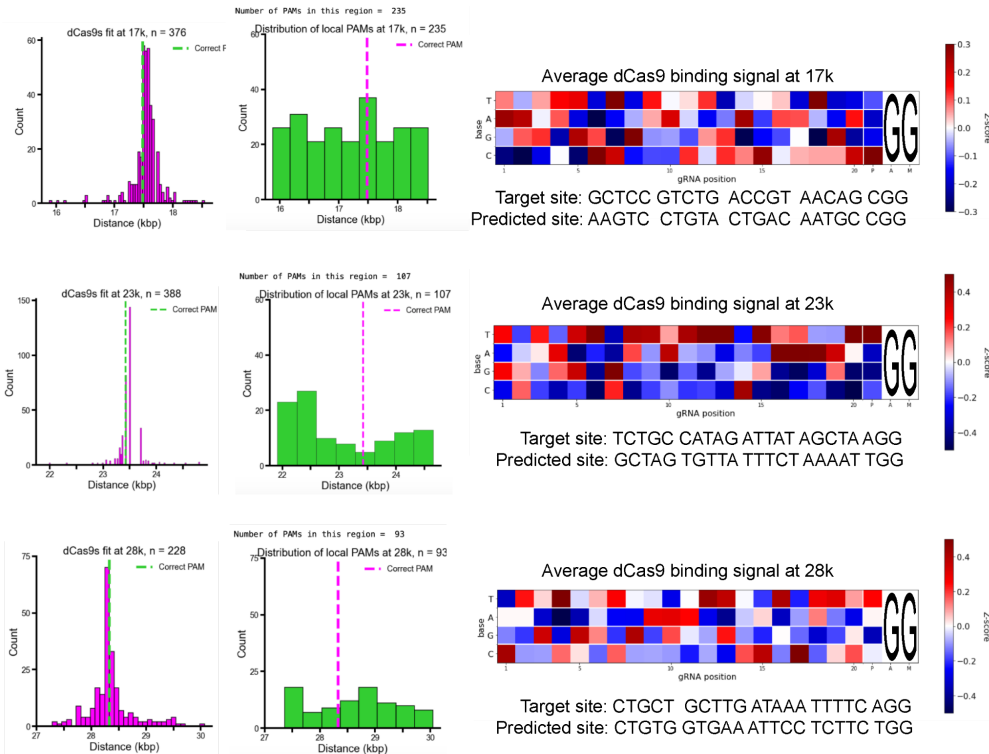
Figure 53. Prediction of dCas9 gRNAs using measured error rate. Left, histograms showing the distribution of measured dCas9 binding events measured around their target site (magenta). Histograms showing the distribution of all local PAMs around the target site (green). **Right**, heat map of the average measured binding site for dCas9s after bootstrapping the fit population of binding events using the measured error rate of 160bp to distribute the signal among local PAMs (NGG).

Within bins over-represented by MBD1 binding, we found some of the same

mCG spacings identified by mapping our binding distribution back to specific sites

(**Figure 54A**). Additionally, we analyzed flanking sequences of mCGs without trying to

differentiate between potential binding sites that occur in bins over-represented by

MBD1-MBD. Instead, all mCGs within a bin were weighted equally. This approach

generated a flanking sequence which was comparable to the average measured binding

site of MBD1-MBD after mapping binding events to specific mCGs (**Figure 54B**). This

consistency validates that this method can be used to analyze the distribution of

DNA-binding proteins under the right circumstances to identify potential signals that

may influence DNA-protein interactions.
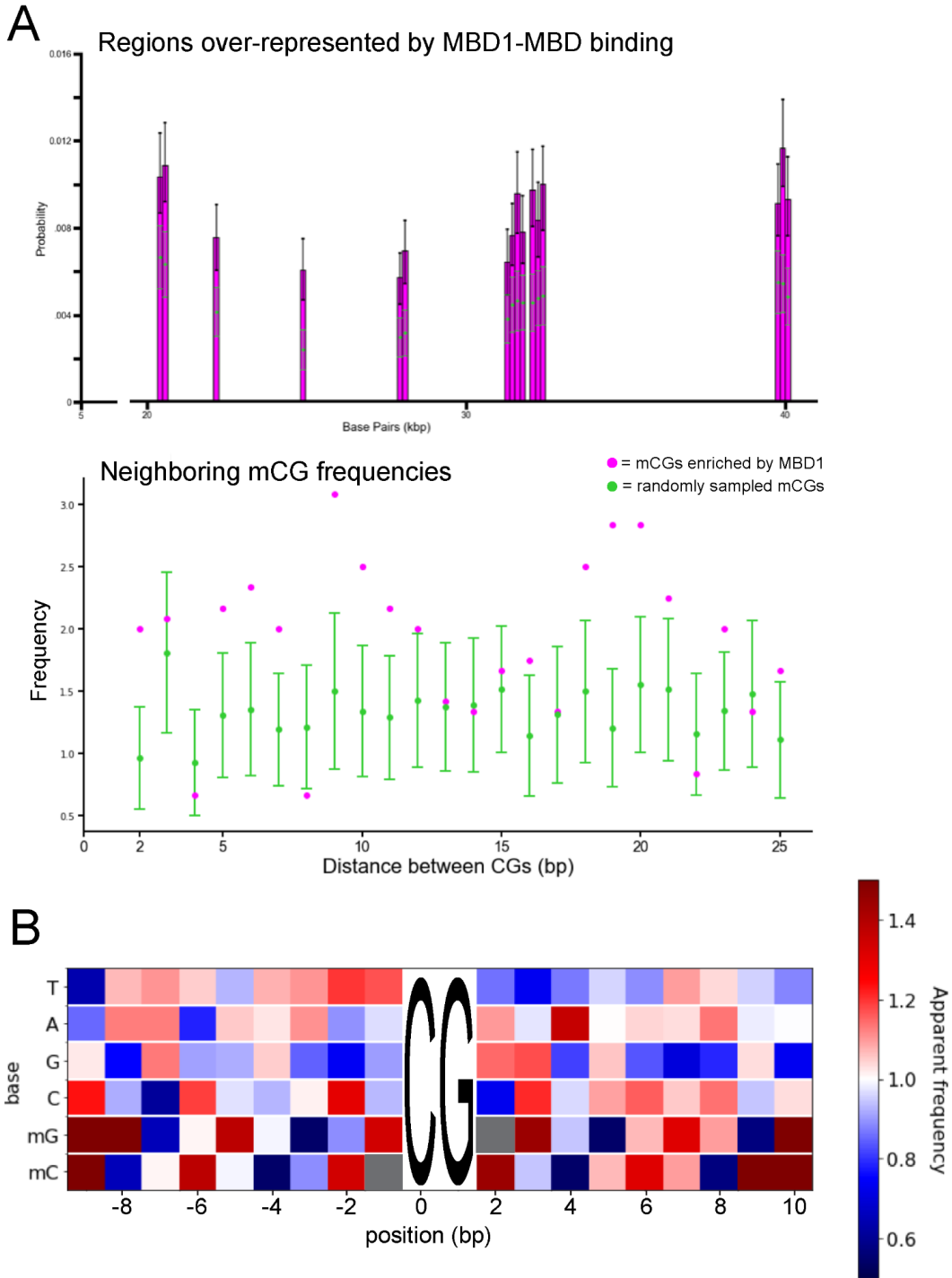
Figure 54. Frequency of mCGs from binned MBD1 enrichment. (**A**) **Top**, over-represented bins from high resolution MBD1-MBD binding distribution (magenta). **Bottom**, plot showing frequency that specific neighboring mCGs appeared in over-represented bins (magenta) alongside the frequency that specific spacings appeared when bins were randomly selected (green). (**B**) Heat-map showing the frequency of sequences flanking mCGs from over-represented bins. Here 'mG' represents a G base paired with a methylated mC on the reverse strand.

3.3 MBD binding is impacted by mCG-density and spacing

While it is well established that MBDs are recruited to methylated DNA, we used this new method to uncover novel signals that may regulate MBD binding. Similarly, previous studies have identified unique cues that enhance or inhibit MBD binding to methylated DNA (32,55). Many of these works utilize full-length proteins or larger constructs that extend beyond the conserved MBD domain. The amount of discrimination encoded within the conserved MBD itself has remained largely unknown. Here we show that minimal, recombinant MBD domains do not bind all mCGs equally; instead, they are able to differentiate between potential binding sites. Like previous studies, we found that MBDs bind more frequently to densely methylated DNA. This behavior led to regions of enriched MBD binding on densely methylated M.SssI-methylated DNA. This mCG-density dependence was so great that it was generally apparent in low resolution binding distributions (**Figures 9/12/13**), and later confirmed using high resolution analysis (**Figures 35,44**). This preference was more extreme on sparsely methylated HpaII-DNA and HhaI-DNA, where MBD1-MBD binding was largely limited to regions which contained relatively dense clusters of mCGs (**Figures 39,40**). Similarly, in gel shifts, stable MBD binding to DNA substrates was often only possible in the presence of multiple mCGs (**Figures 5,49,51**).

In cells, genomic assays have revealed that full-length MBDs are frequently bound at densely methylated regions in the genome (17). This behavior has also been seen in biotechnological applications, like MBD-seq, where MBD domains are used to selectively bind and isolate methylated DNA for sequencing. MBD domains tend to enrich densely methylated DNA more frequently than sparsely methylated DNA

fragments (56,57). On DNA curtains, we saw that mouse MBD1-MBD, mouse MeCP2-MBD, and AtMBD6-MBD all preferred dense regions of DNA methylation where binding sites were surrounded by additional mCGs. Despite the consistency of MBD mCG-density dependent binding, we identified biologically relevant conditions where this binding preference was suppressed (**Figures 14,15**), suggesting that this activity can be regulated in cells. Future studies are required to better understand how protein size and concentration, as well as other cues, can affect MBD binding preferences.

Using our high resolution analysis, we were able to compare thousands of potential mCG binding sites, which allowed us to unpack additional signals within densely methylated regions that may influence MBD recruitment. We identified that both MBD1-MBD and MeCP2-MBD frequently bound mCGs flanked by a second mCG 9 base pairs away on the DNA (**Figures 38,47**). By further testing this spacing in gel shifts, we were able to identify that this signal did not only correlate with MBD1-MBD binding on DNA curtains, but was sufficient to increase MBD1-MBD's recruitment to methylated DNA under certain conditions (**Figures 49,52**). However, tagged MBD binding generally seemed weak in most gel shift conditions we tested. This caveat prevented us from convincingly testing whether any mCG pairs were refractory to MBD binding. However, enhanced MBD1-MBD binding in EMSAs was also possible when neighboring mCGs were 10 base pairs away, suggesting that mCGs separated by nearly a helical turn of DNA may be a specific cue leading to enriched binding in densely methylated regions of the genome. Previous studies have shown that there is an enrichment of mCGs 10 base pairs apart within the human genome (58), and mCGs separated by 10 base pairs of DNA are enriched on nucleosomal DNA in plants (59).

These trends further suggest that this specific spacing may be an important feature of the DNA methylome that regulatory proteins, like MBDs, may recognize.

Future studies are required to better understand how much influence neighboring mCGs can have on MBD binding. One possibility is that two mCGs falling so close together may promote cooperative binding between two nearby MBDs to stabilize their association with DNA. In fact, previous work has shown that MeCP2-MBD may bind methylated DNA as a dimer (60). Alternatively, MBD binding may be enhanced at neighboring mCGs because of a unique physical feature of DNA with two mCGs at specific distances. In our gel shifts, the migration of MBD-bound DNA was consistent whether MBD1-MBD was incubated with DNA containing a single mCG or two mCGs (**Figures 51,52**). This result suggests that the additional stability seen on DNA with two mCGs was not coming from multiple MBD binding events. Instead, something about two mCGs separated by 9-10 base pairs of DNA stabilized recruitment of a single MBD with the DNA. Given MBD's sensitivity to the position of the second mCG, pure density of mCGs may not explain this enhanced MBD binding. Instead, specific pairs of mCGs may create a uniquely favorably substrate for MBD binding by altering DNA shape. Previous work has shown that methylated bases do affect the structure of DNA by altering physical properties like major groove width (61). These differences could be sufficient to create a particularly good binding interaction between an MBD and an mCG at sites with this spacing throughout the genome.

3.4 Conclusion

In this study we show that high-resolution and high-throughput measurements can be made using DNA curtains. Applying such a sensitive technique to monitor molecular interactions in real-time can provide new molecular insights into how different DNA-protein interactions take place. We found that binding of MBDs to DNA is not solely predetermined by the mere presence of a methylated CpG dinucleotide. Instead, this conserved domain has the intrinsic ability to differentiate between methylated sites that occur in different sequence contexts. Although this general behavior is conserved between MBDs from plants and mammals, we found conditions where every MBD we tested was able to create distinct binding distributions on the same methylated DNA. This result indicates that slight differences within this conserved domain are sufficient to modulate a seemingly redundant binding activity. Beginning to understand these differences reveals how these related but distinct proteins can recognize the same signal in different ways to carry out unique functions throughout the genome.

4. Materials & Methods

4.1 Baseline DNA curtains setup

DNA curtains were performed as previously described with minimal adjustments (23,24,25). Briefly, chrome diffusion barriers were nanofabricated on quartz microscope slides with holes drilled through them to enable use of microfluidics. Flowcells were created on nanofabricated slides by creating a small chamber using double-sided tape and coverslips, which were baked on using a vacuum oven. On the back side of the slide, NanoPorts (IDEX Part # N-333) were hot glued over the holes to attach tubing for microfluidics. To begin an experiment, a flowcell was equilibrated in Lipids Buffer (10mM Tris-HCl, pH 7.5 and 100mM NaCl). Then a lipid bilayer composed of DOPC, PEG-2000 DOPE, and biotinylated DOPE lipids (Avanti Polar Lipids 850375, 880130, and 870273) was deposited on the surface of the slide. Streptavidin, diluted in DNA curtains Blocking Buffer (0.5mg/mL BSA, 0.5% Pluronic F-127, 25mM Tris-HCl pH 7.5, 150mM NaCl, 1mM MgCl, and 1mM DTT), was injected into the flowcell to functionalize the biotinylated lipid heads. λ-DNA molecules were tethered to individual lipid heads via a biotin-streptavidin-biotin linkage and visualized using Imaging Buffer (Blocking Buffer supplemented with 10pM YOYO1, ThermoFisher Y3601). Buffer flow was used to align λ-DNA molecules at diffusion barriers and keep DNA in an extended conformation while imaging. We perform our experiments on a custom-built prism-type total internal reflection fluorescence (TIRF) microscope using both a 488nm and 561nm solid state lasers (Coherent). Emitted light was split using an Optosplit 2 split at 561nm (Semrock Di03-R561-t1-25x36) collected on the left and right sides of an emCCD camera (Andor) to generate microscope images.

4.2 Preparing λ-DNA for methylated DNA curtains

First λ-DNA (NEB Catalog # N3011 or Sigma SKU:D3654 for dam⁻/dcm⁻ λ-DNA) was methylated using CpG-specific methyltransferase (MTase) M.SssI (NEB Catalog # M0226M), HpaII MTase (NEB Catalog # M0214S), and/or HhaI MTase (NEB Catalog # M0217S). These DNA methylation reactions were set up with modifications according to NEB's protocol "Recommended Protocol for Methylation of Genomic DNA" and pushing the methylation of all potential sites to completion. Briefly, every 1 µg of λ-DNA is combined with at least 4 units of MTase in 5 µL of reaction buffer (1x NEBuffer 2.0 supplemented with 640µM SAM for M.SssI). Concentrated stocks of λ-DNA are briefly heated prior to pipetting with wide-bore tips to prevent shearing since many λ-DNA molecules can anneal to one another via cos sites. Generally ~25 µg of λ-DNA is used to generate long-term working stocks for DNA curtains. The methylation reaction is started by addition of MTase and incubated at 37°C for 4 hours.

The reaction is moved to 65°C to heat kill MTases. At this time, differentially modified oligos purchased from IDT are spiked-in to functionalize the ends of λ-DNA with "handles." Excess /5Phos/-AGGTCGCCGCCC/Bio/-3' along with /5Phos/GGGCGGCGACCT/3DiG_N/ are added to molar ratios of 1:240. After 20 minutes, the reaction is slow-cooled to room temperature to allow the modified oligos to anneal to the λ-DNA cos sites. The methylation reaction buffer is then supplemented with 10 mM ATP-NaOH, pH=8.0 and 80 units of T4 DNA ligase (NEB Catalog # M0202) are added per µg of λ-DNA. The ligation is incubated overnight at room temperature.

The handled λ-DNA is then cleaned up to remove protein and excess DNA oligos. Methylated / handled λ-DNA was effectively purified from reaction components

using size exclusion chromatography (SEC) or PEG precipitation. SEC was performed using an archaic AKTApurifier and frac-920 fractionator. The overnight ligation reaction was run over an sephacryl S-1000 SF resin (GE Product # 17047601) packed in a HiLoad 16/60 Superdex 75 PrepGrade column (GE Product Code # 17-1068-01) at 0.5mL/min in TE150 (10mM Tris-HCl, pH=7.4, 1mM EDTA, and 150mM NaCl). 1mL fractions were collected. Purified λ-DNA in TE150 should elute from this column in 10-15mL. Fractions were verified to contain λ-DNA using DNA gel electrophoresis in a low percentage (≤ 1%) agarose gel. These dilute fractions are sufficient for DNA curtains and were kept at 4°C for long-term storage.

For PEG precipitation of handled λ-DNA, an equal volume of 40% PEG8000 and 10mM $MgCl_2$ was added to the overnight ligation. The solution was left nutating at RT for at least 60 minutes to form a small pellet of λ-DNA. The precipitated λ-DNA was collected by centrifuging in a tabletop centrifuge at 14,0000xg at 4°C for 10 minutes. The supernatant was removed, and the precipitated DNA was washed with 0.5-1mL of 70% EtOH while nutating at RT. The pellet was then spun down again at 14,0000xg at 4°C for 10 minutes. After 3 washes with 70% EtOH, the pellet was dried and resuspended in TBS (150mM NaCl and 50mM Tris-HCl, pH = 7.5) to a working concentration of approximately 100ng/μL. Successful recovery of λ-DNA was verified using DNA gel electrophoresis in a low percentage (≤ 1%) agarose gel. Resuspended λ-DNA was moved to 4°C for long-term storage.

4.3 Validation and quantification of DNA methylation

Initially, methylation of λ-DNA was confirmed by protection from restriction enzyme digestion. For M.SssI-methylated stocks, 1 ug of λ-DNA was combined with 10 units of BstUI (NEB Catalog # R0518) in 50 uL of 1x NEB CutSmart buffer. The reaction was incubated at 60°C for 60 minutes. Digestion or protection was visualized by running out the reaction in a 0.67% agarose gel with 1x SYBR Safe dye. Similarly, λ-DNA stocks methylated with HhaI and/or HpaII MTases were incubated with the corresponding restriction enzymes, HhaI (NEB Catalog # R0139) and HpaII (NEB Catalog # R0171), for 60 minutes at 37°C prior to gel electrophoresis.

Additionally, methylation levels were further quantified using sequencing methods. For M.SssI-methylated DNA, methylation of CGs was quantified using NEBNext Enzymatic Methyl-seq (EM-seq) module paired with NEBNext Ultra II FS library prep kit using the company's protocols. For HhaI- and HpaII-methylated DNA stocks, bisulfite sequencing (BS-seq) was used to quantify mCG, mCHG, and mCHH levels. The total number of mC vs C reads from both the forward and reverse DNA strands (n = 6226) were used to determine the methylation percentage of symmetrically methylated mCG sites (n = 3113). When plotting the mCG methylation alongside MBD binding distributions, the measured methylation percentage for each symmetrical site was used to determine the mCG probability for each binned region of DNA. Any mCG sites which were absent from the sequencing data or poorly sequenced (low read count) were assigned the average methylation percentage for that stock.

4.4 Cloning and purification of recombinant MBDs

Initially, pCDFDuet and pACYCDuet vectors were modified in-house to facilitate N-terminal and C-terminal sortase labeling, respectively. N-terminal sortase labeling is achieved by ensuring N-terminal glycines can be exposed following cleavage with TEV protease. C-terminal sortase labeling requires an additional LPXTG (LPETG) motif at the C-terminus. Additionally, we included a C-terminal Strep-tag downstream from the sequence encoding the LPETG motif to facilitate clean-up following a C-terminal sortase labeling reaction. These vectors were initially used to clone mouse MBD domains for expression in E. coli. In addition to the sortase labeling motifs, both of these vectors maintained an N-terminal 6xHis-tag used for purification, and eventually utilized to fluorescently label MBDs on methylated DNA curtains.

Mouse MBD1-MBD (aa 2-75) and MBD3-MBD (aa 1-69) were cloned using cDNA generated from mouse embryonic stem cells. MBD4-MBD (aa 42-117) was also successfully cloned from mouse cDNA and expressed in E.coli, but crashed out during dialysis following purification. Both mouse MeCP2-MBD (aa 90-162) and MBD2-MBD (aa 148-216) failed to amplify from cDNA under the conditions we were using. MBD2-MBD was never cloned, and MeCP2-MBD was purchased as a gene block from IDT. This MeCP2-MBD construct encodes the conserved MBD domain, but does not include ~10 additional N-terminal amino acids historically included in the MeCP2-MBD crystal structure and many subsequent *in vitro* studies (62). Initially, purifying mouse MBDs for C-terminal sortase labeling was prioritized, but eventually we focused on using MBDs cloned for N-terminal sortase labeling since it was the more minimal

construct design. We never observed a binding difference on DNA curtains between MBDs cloned and purified for N-terminal or C-terminal sortase labeling.

FLAG-tagged MBD1-MBD and MeCP2-MBD were purchased from GenScript. The coding sequence for these genes were cloned into the pET28a(+)-TEV vector with the addition of an N-terminal 1x-FLAG-tag or 3xFLAG-tag encoded downstream from the TEV cleavage site. Additionally, a longer FLAG-tagged MeCP2-MBD construct was also purchased in pET28a(+)-TEV. This long MeCP2-MBD construct was designed to include the additional residues found in the MeCP2-MBD crystal structure (aa 78-162), but it was never purified and tested on DNA curtains.

Recombinant AtMBD6-MBD and AtMBD6-FL were prepared and gifted to us from the Jacobsen lab at UCLA. These constructs included the addition of an N-terminal MBP-tag and C-terminal 6xHis-tag. AtMBD6-MBD and AtMBD6-FL were directly labeled with Cy3-NHS-ester for visualizing their binding on methylated DNA curtains.

All mouse MBDs were expressed and purified in-house. These constructs were transformed into BL21(DE3) competent E. coli (NEB Catalog # C2527). Large expression cultures (generally 1-2 L of LB or TB broth) were grown in baffled flasks under antibiotic selection at 37°C until OD=0.6-0.8. At this time, expression of MBDs was induced by addition of freshly prepared 0.5 mM IPTG. We never saw an appreciable difference in induction of protein expression with different IPTG concentrations. Similarly, we did not observe a significant difference in protein yield after letting expression proceed for 3 hours at 37°C or lower temperatures overnight. Generally, MBD expression cultures were incubated at 200-250rpm overnight at 18°C.

After expression, bacterial cultures were spun down at 4,000 x g at 4 degrees. The cell pellets were pooled and resuspended in Lysis Buffer containing 40mM Tris-HCl, pH=7.5, 500mM NaCl, 1mM DTT, 0.1% Triton X-100, and 1mM PMSF. (Approximately 10mL of lysis buffer was used per liter of expression culture.) The resuspended pellets were flash frozen in liquid nitrogen and stored at -80°C until purification. Cell pellets were also frequently stored dry, and thawed into Lysis Buffer with no appreciable effect on protein yield. Cell pellets were thawed and sonicated on ice using a blunt tip (such as Sonics, Vibra-Cell, CV33). The amplitude was set at 40% and sonicated for 10 seconds with a 60 second rest on ice in between subsequent sonications. This approach was repeated for at least 1 total minute of sonication. The lysate was then clarified by centrifuging (Beckman-Coulter, Optima L-90k UltraCentrifuge, Ti45 rotor) at 30,000 x g for 30 minutes at 4°C.

Regardless of construct design, all MBDs were purified using N-terminal 6xHis-tags and Ni-NTA resin. Purifications were performed by using FPLC with 1-5mL HisTrap columns (GE Product # 17-5248 or 17-5247) or gravity columns filled with Ni-NTA Agarose (Protino REF 745400). If using FPLC, lysate supernatants were further clarified using 0.22µm PES filters (Millipore SLGP033RS). Prior to binding, Ni-NTA resins were all equilibrated in Wash Buffer containing 40mM Tris-HCl, pH=7.5, 500mM NaCl, 1mM DTT, and 40mM imidazole. To bind HisTrap columns, the clarified supernatant was loaded into a Superloop, and slowly injected over the HisTrap column at a low flow rate for at least 30 minutes. To bind free Ni-NTA resin, equilibrated resin was added to the supernatant and left nutating in a cold room for at least 45 minutes.

After binding Ni-NTA resin, non-specifically bound proteins were washed off with Wash Buffer containing 40mM Tris-HCl, pH=7.5, 500mM NaCl, 1mM DTT, and 40mM imidazole for at least 10-20 column volumes. To elute MBDs and maintain the 6xHis-tag for visualizing binding on methylated DNA curtains, bound protein was collected using Elution Buffer containing 40mM Tris-HCl, pH=7.5, 250mM NaCl, 500mM imidazole, and 1mM DTT.

To purify FLAG-tagged MBDs without the additional N-terminal 6xHis-tag, MBDs were eluted by performing an on-column cleavage reaction with HaloTEV protease (Promega Cat. # G6602). This purification was also done using a gravity column packed with Ni-NTA agarose. The MBD-bound resin was resuspended to a 30% slurry in TEV Cleavage Buffer containing 50mM Tris-HCl, pH = 7.5, 150mM NaCl, 1mM DTT, and 0.005% IGEPAL CA-650. Then 150 units of HaloTEV protease was added per 1mL of resuspended bound resin — 240µL in 8mL total. This cleavage reaction was incubated overnight while nutating in a cold room. TEV-cleaved FLAG-tagged MBDs were then collected with the flow-through from this reaction. While this purification strategy was successful, the on-column cleavage step was usually inefficient, lowering our yield. In fact, no free HaloTEV was observable along with cleaved MBDs collected in the flow-through. This suggests that Promega HaloTEV may also be labeled with a 6xHis-tag, trapping it on the Ni-NTA resin, which would potentially reduce its cleavage efficiency in this setup. In the future, other sources of TEV protease should be explored to optimize this purification strategy.

To isolate MBDs from any remaining contaminants and buffer exchange it for storage, MBDs were additionally purified using SEC. First, MBD elutions were

concentrated down to ≤ 500 µL prior to injecting them over a Superdex 75 Increase 10/300 GL column (S75) equilibrated in Sizing Buffer containing 40mM Tris-HCl, pH=7.5, 150mM NaCl, 1mM DTT. Since all mouse MBD constructs are relatively small (approximately 10-15kDa) centrifugal protein concentrators with a 3 kDa cutoff were used to concentrate elutions (Amicon Ultracel – REF UFC800324). Concentrated MBD elutions were injected over the S75 column at a flow rate of 0.4mL/min, and the flow-through was collected in 0.5mL fractions. Purified MBDs generally passed through the S75 column after ~13mL of Sizing Buffer. Relevant fractions were pooled and concentrated in a centrifugal protein concentrator (3kDa cutoff) for storage. Beer's law was used to quantify the concentration of purified MBDs. After sizing and concentrating, purified stocks of MBDs were supplemented with 10-25% glycerol was added as a cryoprotectant and aliquots were flash frozen in liquid nitrogen and stored at -80°C.

4.5 Visualizing MBDs on DNA curtains

After realizing that sortase labeling multiple MBDs would be an unwieldy approach, we found that indirectly labeling recombinant MBDs with fluorescent antibodies was the most efficient way to visualize MBD binding on DNA curtains.

6xHis-tagged MBDs were most efficiently labeled with anti-6xHis-Alexa555 (clone 4E3D10H2/E3, Invitrogen: MA1-135-A555). We struggled to find another commercially available fluorescent anti-6xHis antibody that worked well at the single molecule level. Surprisingly we saw poor labeling efficiency using the same anti-6xHis antibody clone labeled with Alexa488 (clone 4E3D10H2/E3, Invitrogen: MA1-135-A488). Similarly, another Alexa488-labeled anti-6xHis antibody clone also failed to robustly label binding distributions on DNA curtains (clone HIS.H8, Invitrogen: MA1-21315-A488). Additionally,

we had mixed results using a Penta-His-AlexaFluor488 conjugate (Qiagen), which was only available at 0.2 mg/mL. Similarly, many other fluorescent antibodies were only available at concentrations below 1 mg/mL, which complicated efficiently labeling MBDs at relatively high concentrations. Our inability to find alternative commercially available fluorescent anti-6xHis antibodies limited us to imaging 6xHis-tagged MBDs with Alexa555.

To label FLAG-tagged MBDs, we relied on name-brand monoclonal anti-FLAG M2 antibodies (Sigma, SKU F3165). Initially we tried utilizing commercially available fluorescent anti-FLAG antibodies. We found that Cy3-labeled anti-FLAG M2 antibodies (Sigma, SKU A9594) allowed us to visualize MBDs on DNA curtains, but required higher concentrations of MBDs, suggesting that it was not a robust labeling strategy. Additionally, we found that FITC-labeled anti-FLAG M2 antibodies (Sigma, SKU 4049) seemed to efficiently label MBDs, but the high rate of photo-bleaching for FITC made it unsustainable as a labeling strategy at the single molecule level. Ultimately, we generated Qdot-labeled anti-FLAG antibodies to visualize FLAG-tagged MBDs on DNA curtains. We conjugated Qdot705 to anti-FLAG M2 antibodies using a SiteClick Antibody Labeling Kit (Invitrogen, REF S10454). After conjugation and clean-up according to the company's protocol, we stored anti-FLAG-Qdot705 antibodies at 4°C in Tris-Buffered Saline (TBS).

Before an experiment, MBDs were diluted in Imaging Buffer and labeled with fluorescent antibodies prior to injection into the flowcell. To preserve fluorescent antibody stocks, we used the lowest molar excess possible while maintaining efficient labeling of MBDs. We typically combined MBDs with fluorescent antibodies at 1:2 molar

ratio and let them incubate on ice in the dark for at least 15 minutes to equilibrate. Using a greater molar excess of fluorescent antibodies did not increase the amount of visualizable MBDs on DNA curtains, suggesting that a 1:2 molar ratio was sufficient to maximize the labeling efficiency of this approach. Similarly, giving antibodies more time to bind tagged-MBDs or allowing antibodies to bind at a higher temperature did not lead to an observable increase in the number of visualizable binding events. To prevent dissociation of antibody-bound MBDs, subsequent dilutions were performed immediately before injection into the flowcell.

Additionally, to directly visualize MBDs on DNA curtains, in-house NHS-ester labeling reactions were performed using uncleaved 6xHis-TEV-1xFLAG-MBD1-MBDs. Cy3-NHS-ester labeled AtMBD6 constructs were prepared and gifted to us by the Jacobsen lab from UCLA (26). To label MBD1-MBDs, Cy3-NHS-ester was resuspended in DMSO to a concentration of 10mM. Labeling reactions were carried out with 100µM MBD1-MBD in Sizing Buffer, modified for NHS-ester labeling (50mM HEPES, pH = 6.5, 150mM NaCl, and 1mM DTT). Initially, we tried labeling MBD1-MBD with a 10:1 molar excess of Cy3-NHS-ester for approximately 5 minutes at RT before quenching the reaction with 10mM glycine in 1M Tris-HCl, pH = 7.5. However, we kept noticing a large amount of precipitate forming during NHS-ester labeling reactions. The most successful labeling conditions that we found were labeling 100µM MBD1-MBD with less Cy3-NHS-ester (down to as low as 2:1 molar excess) for approximately 5 minutes at RT before quenching the reaction with 1M Tris-HCl, pH = 7.5. Any precipitate was spun down before cleaning-up the supernatant using HiTrap Desalting columns equilibrated

in Sizing Buffer (buffered with Tris-HCl, pH = 7.5). Fractions were pooled and concentrated before flash freezing aliquots in 25% glycerol and storing at -80°C.

During a DNA curtains experiment, fluorescent MBDs were loaded into a 100µL loop and injected into the flow cell under constant, low flow (0.2mL/min) to bind methylated or unmethylated DNA curtains. While collecting images for analysis, buffer flow was increased to 1.0mL/min with a 30mL syringe (BD REF 302832) and laser power was set to 5mW. Images were collected with a 100 msec interval and 100 msec exposure and EMCCD gain was set at 500-750. Each DNA curtain was imaged for a total of 20 seconds (200 images total – 100 for each of our two fluorescent channels).

4.6 Alignment of fluorescent channels

To properly align each fluorescent channel, we calibrate our system by simultaneously imaging the same fluorescent particles in both channels. Images for alignment are taken after the completion of an experiment to account for the exact position of each channel during a particular experiment. TetraSpeck Microspheres (0.1µm, Invitrogen, REF T-7279) are mixed thoroughly and diluted at least 1:50 in Lipids Buffer (10mM Tris-HCl and 100mM NaCl). 10µL of this dilution is spotted on a SuperFrost Slide (VWR Catalog # 48311-703) and placed under a 22 x 22mm coverslip (Fisher Scientific, Catalog # 12-544-10). Images from at least 10 different positions are collected in both channels by continuous illumination.

We use the offsets of these particles to correctly align the channels using custom software, available on the Redding lab GitHub (https://github.com/ReddingLab). Image stacks from each channel are z-projected to create an average image of the position of TetraSpeck beads in each fluorescent channel. Point-fitting is used to determine the

precise location of the same fluorescent beads as they appear in both channels. By comparing these measurements from many images, we can directly measure the offset between all TetraSpeck beads in both fluorescent channels.

At this point, we can use the average offset between channels to align images of DNA curtains. All pixels from images of one channel can be modified by the average (x,y) offset to bring it into register with the second channel. While this approach generally worked well to align images of DNA curtains, we found that it was not robust enough for reliable alignment across all images. In particular, images of Qdot705 were consistently misaligned using this approach.

To improve our alignment, we modified our approach to perform a piecewise affine transformation to align our channels. This image analysis has been previously established, allowing us to adopt this method to align our images (43). Briefly, we still used TetraSpeck microspheres to determine the offsets between each channel. Instead of measuring the average offsets for the entire channel, we determine how the offsets vary over the entire field of view. The piecewise affine transformation uses this information to differentially shift different regions of the image to account for variations in misalignment across the field of view. Although more computationally intensive, the transformed images are better aligned than simply using the average offset to uniformly shift the entire field of view.

4.7 Baseline analysis of binding events on DNA curtains

In traditional DNA curtains experiments, λ-DNA is visualized using the signal of YOYO-1 fluorescence. After collecting images and aligning channels as described

above, we calculate the position of fluorescent protein binding events using the dimensions of the DNA as measured by YOYO-1 fluorescence.

This analysis is described at length on the Redding lab GitHub (https://github.com/ReddingLab). Briefly, to automate this analysis, edge finding software is used to automatically fit the fluorescent signal of DNA. The precise pixel positions corresponding to the beginning and end of visible DNA are determined by measuring the loss of YOYO-1 fluorescence at either end of the molecule. In the second channel, after automatically point-fitting potential DNA binding events, the bounds of each DNA molecule are used to determine binding events that fall on each DNA molecule. The pixel distance from the free-end of the DNA is then used to calculate the position of each binding event.

For our baseline analysis, we used a linear estimation of DNA extension to convert this pixel distance into a base pair position. Generally, we used the average global extension of DNA curtains as measured by dCas9s (874 bp / pixel) to convert relative pixel distances to base pairs.

Alternatively, this conversion can be done using a unique extension for each DNA molecule using the top and bottom positions as measured by YOYO-1 fluorescence. However, for this approach to reliably work, it requires knowing where the YOYO-1 fluorescence actually ends at the top end of the DNA. Chrome barriers obscure an undetermined amount of DNA at the tethered end of the molecule. Additionally, variability in nanofabrication results in differentially sized barriers across different slides and even across different regions of the same slide, further complicating our ability to estimate how much DNA is blocked by barriers.

Generally, we know the pixel size of barriers and that dCas9s targeted to sites around 5kbp are partially blocked by barriers, suggesting that the beginning of visible DNA is around that position. Thus, when fitting binding events based on the linear extension as measured by YOYO-1 fluorescence, we assumed the top position was around 5kbp. However, based on estimations using dCas9s (**Figure 27**), this approach was generally less reliable than using the global average as measured with dCas9s (**Figure 29**).

Measuring binding events based on their distance from the free-end of the DNA was used to generate all low resolution MBD binding distributions. To create histograms accurately reflecting the resolution of this analysis, binding data was plotted in ~874bp bins, based on the average number of base pairs per pixel. Infrequently, low resolution binding distributions were plotted using 1000bp bins, comparable to the error rate of this analysis. Error bars represent 95% confidence intervals as determined by performing bootstrap analysis.

4.8 Targeting dCas9s to DNA curtains

Recombinant dCas9 protein was purchased from IDT (Alt-R S.p. dCas9 protein V3, Catalog # 1081067) and used to visualize specific sequences along DNA curtains. The dCas9 protein was loaded with gRNA according to IDT's "Alt-R CRISPR-Cas9 system - *in vitro* cleavage of target DNA with RNP complex" protocol with slight modifications for DNA curtains.

crRNA:tracrRNA duplexes were used to target dCas9 to DNA curtains. Specific crRNAs were purchased from IDT (see **Table 1**) and resuspended to 100µM in IDTE Buffer. crRNAs were diluted to 10µM and annealed to IDT tracrRNA (Catalog #

1072533) at a 1:1 ratio in IDT's Nuclease-Free Duplex Buffer. After denaturing the RNAs for 5 minutes at 95°C, the reaction was slow cooled to room temperature to allow for gRNA complex formation. The gRNA was then aliquoted and stored at -20 degrees.

Alt-R S.p. dCas9 protein as supplied from IDT is diluted to 1µM before use in IDT's Cas9 Buffer (150mM KCl and 30mM HEPES, pH = 7.5). dCas9 protein is diluted to ~200 nM and loaded with excess gRNA at a 1:10 ratio in Cas9 buffer and incubated at room temperature for at least 10 minutes to allow for RNP complex formation. To visualize the dCas9 proteins, we generally label dCas9s molecules with anti-HisAlexa555 (clone 4E3D10H2/E3, product number MA1-135-A555). In the case of two-color dCas9 experiments, several dCas9s were labeled with anti-HisAlexa488 (clone HIS.H8, product number MA1-21315-488). To label dCas9s, excess antibody is added to the dCas9 loading reaction at a 1:2 ratio and then incubated in the dark at room temperature for at least 15 minutes. For Alexa488-labeled dCas9s, a higher molar excess of antibody, up to a 1:6 molar ratio, was used since the labeling efficiency with this antibody was generally low.

The fluorescent RNP complex is then diluted to ~4nM in modified IDT Cas9 Nuclease Reaction Buffer (0.5mg/mL BSA, 0.5% PF127, 40mM Tris-HCl, pH=7.5, 100 mM NaCl, 5mM $MgCl_2$, 0.1mM EDTA, 1mM DTT) and a 100µL loop was used to inject dCas9 into the flow cell. The fluorescent RNP complex was incubated in the flow cell with no buffer flow for 5-10 minutes before excess dCas9 was washed out. To remove any nonspecifically bound apo-dCas9, the flow cell was washed with several 100µL injections of Imaging Buffer containing approximately 100µg/mL heparin. In the case of two-color dCas9 experiments, robustly labeled Alexa555-dCas9s were injected into the

flow cell first, followed by Alexa488-dCas9s. Since the labeling efficiency of anti-6xHis-Alexa488 was so poor, this order of addition minimized the chance of DNA bound Alexa555-labeled dCas9s from being mislabeled by excess anti-6xHis-Alexa488. Similarly, when combining dCas9s and MBDs on DNA curtains, dCas9s were targeted first. DNA curtains could then be washed with heparin before injecting MBDs.

Table 1. dCas9 gRNA sequences	
Position	**Target site sequence**
1k (1010-1029bp)	5'- CACCCTTATCTGGTTGCCGACGG - 3'
5k (5039-5058bp)	5'- GAAATCACGCTGATTTACAGCGG - 3'
6k (5988-6007bp)	5'- TCCGTTATGAGGATGTGCTCTGG -3'
7k (6986-7005bp)	5'- CTATGGCTGCATTCAGGATGCGG -3'
13k (13501-13520bp)	5'- GAAATCGACCTGACAGAGGTCGG -3'
18k (17463-17482bp)	5'- GCTCCGTCTGACCGTAACAGCGG -3'
20k (20048-20067bp)	5'- TGCCAGTGCATCAGCTGCTCAGG -3'
23k (23403-23422bp)	5'- TCTGCCATAGATTATAGCTAAGG -3'
26k (26189-26208bp)	5'- GCTGGGTCAGGTTGTTCTTTAGG -3'
28k (28306-28325bp)	5'- CTGCTGCTTGATAAATTTTCAGG -3'
32k (31827-31846bp)	5'- CGGTGCAAACCTCAGCAAGCAGG -3'
37k (37509-37528bp)	5'- AAGCCAGAATGCAGAATCACTGG -3'
43k (42937-42956bp)	5'- CAAACCATAACCGCCATCGCTGG -3'
47k (47752-47771bp)	5'- ATCTGCTGATGATCCCTCCGTGG - 3'

4.9 dCas9 analysis on YOYO-1 stained DNA

Fluorescent dCas9 images were generally analyzed as described above. Channels were aligned and point-fitting was used to identify dCas9 molecules in the field of view. These binding events were assigned to DNAs and fit using a linear estimate of the bound DNA's extension, assuming that 5000 - 48502 bp of DNA were visible. This low resolution distribution of dCas9s revealed that this estimate of DNA extension did a particularly bad job fitting binding distributions.

Moving forward, we used dCas9 pixel positions to further analyze DNA dimensions. First, we used the pixel positions of all dCas9s to calculate an average of 874 base pairs per pixel across all DNA molecules measured. Using this linear estimate of DNA extension allowed us to better fit binding events based on their distance from the free-end of the DNA. Almost all low resolution MBD binding distributions were measured from the free-end using this global linear extension.

Furthermore, we began using dCas9s to measure the specific extension of individual DNA molecules to more accurately fit nearby binding events, rather than doing so by relying on the global average extension. We restricted this approach by using flaking pairs of dCas9s to fit dCas9 binding events that fall in between them. The measured DNA extension determined by the flanking pair should be most accurate over the internal region of the DNA between them. By using the pixel positions of the flanking pair and their known target sites, we can accurately measure a unique extension for individual DNAs. Using this measured extension, the position of internal dCas9 binding events were then calculated based on their pixel distance from the flanking pair.

To further refine this pairwise analysis, we restricted our calculations to closest flanking pairs of dCas9s. This means that dCas9 binding events could only be fit using a local extension measured by their closest neighboring dCas9s. This restriction meant that dCas9 pairs used to measure extension were all ~10kbp apart. While still relying on a linear estimate of DNA extension, fluctuations in DNA extension should be relatively minimal over these distances. Fitting internal binding events with this local measurement of DNA extension reduced the error of our calculations even further.

4.10 Using dCas9s to measure non-uniform extension and binding events

In the absence of YOYO-1 fluorescence, dCas9s can be used to image and fit the dimensions of each DNA molecule in a DNA curtain. The process of using pattern recognition to identify DNA molecules is described on the Redding lab GitHub (https://github.com/ReddingLab).

Briefly, images of dark DNA, labeled with dCas9s, are collected, z-projected, and aligned as usual. While first validating this approach, we performed this analysis on dCas9 binding events in the presence of YOYO-1 fluorescence in the second channel. Regardless, point fitting is used to first identify dCas9 binding events in their respective microscope channel. The average fluorescent signal of all dCas9s and their average distance from one another is used to generate a synthetic image of targeted dCas9s binding DNA. This template is used as a mask to automatically identify DNA strands in the field of view based on the signal of dCas9 binding events alone. By aligning dCas9 binding events with this template, point-fit dCas9s are sorted into populations based on their target sites. Additionally, this idealized template of dCas9 binding events is used to rotate images to ensure that DNAs are linear to facilitate subsequent image analysis.

Next, we use the pixel positions of dCas9 binding events as measurements of their known target sites to directly measure the extension of individual DNAs using as many as 7 individual points. The non-uniform extension of tethered DNA molecules under flow has been modeled, yielding equations that describe the relationship between force and extension (45,46). We adapted a function describing the extension of a worm-like chain under flow to quantify the extension of tethered λ-DNA using the positional information provided by dCas9s. Using this function, we generated extension curves that compare the pixel position of dCas9s within an image to their targeted position in base pairs.

Plotting the extension of DNA molecules allowed us to calculate the force acting on the DNA at each base pair position along its length. In order to fit our molecules using this previously described function, we had to add an additional gamma factor to further account for differential force along the length of DNA molecules in a curtain. Ultimately, we found that the average force that we estimate by this method is consistent with previous measurements of tethered DNA under flow. Once the fitting parameters of this function are established for tethered DNAs, any protein bound at an unknown target site can also be converted from its pixel position in the image to a base pair position bound on the DNA. Any DNAs with force-extension curves that are not described well by this function are excluded from analysis. Once the dimensions of a dark DNA molecule were determined using dCas9s, they could be used to identify binding events that fall on the same DNA molecules in the second fluorescent channel. Based on the force-extension curves generated for each DNA molecule, we can calculate the base pair position of binding events in the second channel based on their

pixel position without relying on an assumed linear extension. Initially, using this analysis to predict the target site of dCas9 molecules in the second channel, allowed us to quantify the error of this approach, σ = 160. Histograms displaying binding distributions measured with this high resolution analysis were generated using 160 bp bins. Additionally, this statistic enabled us to further refine our measured MBD binding distributions on DNA curtains.

4.11 Mapping high resolution MBD binding events back to mCG sites

In order to compare the position of MBD binding events with the distribution of mCGs, we first determined the probability that any particular mCG would appear in our measured datasets. We completed this analysis using the known error rate of our high resolution analysis (σ = 160) as measured by calculating the position of dCas9s in the second channel using the non-uniform extension of the DNA, described above.

To map MBD binding events back to specific mCG sites, we use bootstrap analysis to calculate confidence intervals for the probability that any mCG could appear bound in our binding distributions. Briefly, we randomly sampled all possible mCGs, and then we shifted the position of each sampled mCG based on our measured error rate. Based on this new position, we identified the closest mCG which would represent the 'measured' mCG in our dataset. We sampled all possible mCG binding sites like this to generate a data set that matched the depth of our measured MBD binding events, and repeated this analysis at least 300 times. This allowed us to generate 95% confidence intervals representing how frequently or infrequently any mCG could be bound in a dataset the size of our measured MBD binding distributions after accounting for the error rate of our analysis. If the MBD binding probability at a particular mCG site falls

outside this range, it suggests that MBD occupancy at the site cannot be explained by random chance alone.

Next, we used this same error rate to map MBD binding events back to mCG sites. Rather than just assigning binding events to the nearest mCG, we distributed the signal of each measured MBD binding event to all local mCGs in a weighted manner based on the error rate of this high resolution analysis. This allowed us to generate a more accurate probability for MBD occupancy at any given mCG based on all the measured binding events in our datasets. Once each site had an associated probability, we compared this to our confidence intervals to determine which mCGs were bound more or less frequently than expected.

To generate heatmaps of mCG sites throughout λ-DNA, sites were weighted based on MBD binding frequency as well as measured methylation percentages. Their frequencies were further normalized based on the frequency of all possible flanking sequences throughout the entire λ-genome.

4.12 Electrophoretic mobility shift assays (EMSAs)

Short FAM-labeled DNA substrates were ordered from IDT and annealed in IDT Duplex Buffer. DNA oligos were annealed by heating to 95°C for 5-10 minutes and then slow cooling to room temperature. Longer DNA substrates were PCR amplified from pUC19 using an Atto488-labeled primer. These PCR products were cleaned-up, methylated using M.SssI, and then cleaned-up again. Methylation was validated based on protection from SmaI digestion visualized using DNA gel electrophoresis. **Table 2** contains a list of DNA sequences used for EMSAs.

Table 2. EMSA DNA substrates		
DNA	**Figure(s)**	**Sequence**
32mer DNA	4	Fwd: 5'- CCT AGC TGT AAC TAG mCG CAT GCA GTA CTT GAG -3'
		Rev: 5'- CTC AAG TAC TGC ATG mCG CTA GTT ACA GCT AGG -3'
pUC19 primers	5	Fwd: 5'-GGG ATG TGC TGC AAG GC-3'
		Rev: 5'- TAT GCT TCC GGC TCG TAT GT-3'
DNA hairpin - 4bp	49, 50, 52	5'- CCC ATC TGC mCG GC mCG GCT CTA GCC ATAT TAT AGG CTA GAG C mCG GC mCG GCA GAT GGG -3'
DNA hairpin - 7bp	49, 50, 51, 52	5'- CCC ATC TGC mCG GGC AGC mCG GCT CTA GCC ATAT TAT AGG CTA GAG C mCG GGC TGC mCG GCA GAT GGG -3'
DNA hairpin - 9bp (uCG and single mCGs)	49, 50, 51, 52	5'- CCC ATC TGC mCG GGC AGT GC mCG GCT CTA GCC ATAT TAT AGG CTA GAG C mCG GGC ACT GC mCG GCA GAT GGG -3'
DNA hairpin - 10bp	49, 50, 52	5'- CCC ATC TGC mCG GGC AGT TGC mCG GCT CTA GCC ATAT TAT AGG CTA GAG C mCG GGC AAC TGC mCG GCA GAT GGG -3'
DNA hairpin - 16bp	49, 50, 52	5'- CCC ATC TGC mCG GGC AGT CTA GAG TGC mCG GCT CTA GCC ATAT TAT AGG CTA GAG C mCG GGC ACT CTA GAC TGC mCG GCA GAT GGG -3'
DNA hairpin - 21bp	49, 50, 51, 52	5'- CCC ATC TGC mCG GGC AGT CAT AGG TAG AGT GC mCG GCT CTA GCC ATAT TAT AGG CTA GAG C mCG GGC ACT CTA CCT ATG ACT GC mCG GCA GAT GGG -3'

Binding reactions were set up using DNA curtains Blocking Buffer supplemented with 5% glycerol to facilitate loading the gel. We found that adding glycerol before or after equilibrating MBDs with methylated DNA had no effect on binding. Fluorescent DNA was diluted to a final concentration of 1-50nM, depending on the EMSA. We found that observable gel shifts were possible at all DNA concentrations in this range. DNA was then incubated in the presence of a molar excess of MBD, usually in a volume of 10µL. While performing supershifts, MBDs were incubated with a molar excess of anti-FLAG M2 antibody at a molar ratio of 1:2 for at least 15 minutes. We found that stable binding required variable amounts of MBD. Under stable binding conditions, we could visualize a large gel shift with as little as 5x molar excess MBD. However, under many conditions, an appreciable gel shift required 100x molar excess MBD or more. Binding reactions were allowed to equilibrate in the dark at room temperature for at least 15 minutes. To visualize gel shifts, binding reactions were loaded in a 5% Native polyacrylamide gel first equilibrated in 0.5x TBE for at least 60 minutes at 240V. Samples were run out in the dark in a cold room for at least 60 minutes at 80 V (approximately 5 mAmps). Gel shifts were then imaged using an Amersham Typhoon laser scanner and quantification was performed using ImageJ.

5. References

1. Schmitz, Robert J., Zachary A. Lewis, and Mary G. Goll. "DNA methylation: shared and divergent features across eukaryotes." *Trends in Genetics* 35.11 (2019): 818-827.

2. Moore, Lisa D et al. "DNA methylation and its basic function." *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* vol. 38,1 (2013): 23-38. doi:10.1038/npp.2012.112

3. Jones, Peter A., and Daiya Takai. "The role of DNA methylation in mammalian epigenetics." *Science* 293.5532 (2001): 1068-1070.

4. Iurlaro, Mario et al. "DNA methylation homeostasis in human and mouse development." *Current opinion in genetics & development* vol. 43 (2017): 101-109. doi:10.1016/j.gde.2017.02.003

5. Antequera, Francisco. "CpG Islands and DNA Methylation." *eLS* (2007): 1-7.

6. Bird, A et al. "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA." *Cell* vol. 40,1 (1985): 91-9. doi:10.1016/0092-8674(85)90312-5

7. Kass, S U et al. "How does DNA methylation repress transcription?." *Trends in genetics : TIG* vol. 13,11 (1997): 444-9. doi:10.1016/s0168-9525(97)01268-7

8. Medvedeva, Yulia A., et al. "Effects of cytosine methylation on transcription factor binding sites." *BMC genomics* 15.1 (2014): 1-12.

9. Hark, A T et al. "CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus." *Nature* vol. 405,6785 (2000): 486-9. doi:10.1038/35013106

10. Baubec, Tuncay, and Dirk Schübeler. "Genomic patterns and context specific interpretation of DNA methylation." *Current opinion in genetics & development* 25 (2014): 85-92.

11. Du, Qian, et al. "Methyl-CpG-binding domain proteins: readers of the epigenome." *Epigenomics* 7.6 (2015): 1051-1073.

12. Roloff, Tim C et al. "Comparative study of methyl-CpG-binding domain proteins." *BMC genomics* vol. 4,1 (2003): 1. doi:10.1186/1471-2164-4-1

13. Hendrich, Brian, and Susan Tweedie. "The methyl-CpG binding domain and the evolving role of DNA methylation in animals." *TRENDS in Genetics* 19.5 (2003): 269-277.

14. Zemach, Assaf, and Gideon Grafi. "Characterization of Arabidopsis thaliana methyl-CpG-binding domain (MBD) proteins." *The Plant journal : for cell and molecular biology* vol. 34,5 (2003): 565-72. doi:10.1046/j.1365-313x.2003.01756.x

15. Coelho, Fernanda Silva, et al. "Methyl-CpG binding proteins (MBD) family evolution and conservation in plants." *Gene* 824 (2022): 146404.

16. Ballestar, E, and A P Wolffe. "Methyl-CpG-binding proteins. Targeting specific gene repression." *European journal of biochemistry* vol. 268,1 (2001): 1-6. doi:10.1046/j.1432-1327.2001.01869.x

17. Baubec, Tuncay et al. "Methylation-dependent and -independent genomic targeting principles of the MBD protein family." *Cell* vol. 153,2 (2013): 480-92. doi:10.1016/j.cell.2013.03.011

18. Mahana, Yutaka et al. "Structural Insights into Methylated DNA Recognition by the Methyl-CpG Binding Domain of MBD6 from *Arabidopsis thaliana*." *ACS omega* vol. 7,4 3212-3221. 19 Jan. 2022, doi:10.1021/acsomega.1c04917

19. Fraga, Mario F et al. "The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties." *Nucleic acids research* vol. 31,6 (2003): 1765-74. doi:10.1093/nar/gkg249

20. Hilario, Jovencio, and Stephen C. Kowalczykowski. "Visualizing protein–DNA interactions at the single-molecule level." *Current opinion in chemical biology* 14.1 (2010): 15-22.

21. Pan, Hai, et al. "CpG and methylation-dependent DNA binding and dynamics of the methylcytosine binding domain 2 protein at the single-molecule level." *Nucleic acids research* 45.15 (2017): 9164-9177.

22. Liu, Ming et al. "DNA looping by two 5-methylcytosine-binding proteins quantified using nanofluidic devices." *Epigenetics & chromatin* vol. 13,1 18. 16 Mar. 2020, doi:10.1186/s13072-020-00339-7

23. Fazio, Teresa, et al. "DNA curtains and nanoscale curtain rods: high-throughput tools for single molecule imaging." *Langmuir* 24.18 (2008): 10524-10531.

24. Greene, Eric C., et al. "DNA curtains for high-throughput single-molecule optical imaging." *Methods in enzymology*. Vol. 472. Academic Press, 2010. 293-315.

25. Soniat, Michael M., et al. "Next-generation DNA curtains for single-molecule studies of homologous recombination." *Methods in enzymology*. Vol. 592. Academic Press, 2017. 259-281.

26. Ichino, Lucia et al. "MBD5 and MBD6 couple DNA methylation to gene silencing through the J-domain protein SILENZIO." *Science (New York, N.Y.)*, eabg6130. 3 Jun. 2021, doi:10.1126/science.abg6130

27. Moerner, William E. "New directions in single-molecule imaging and analysis." *Proceedings of the National Academy of Sciences* 104.31 (2007): 12596-12602.

28. Feng, Suhua et al. "Efficient and accurate determination of genome-wide DNA methylation patterns in Arabidopsis thaliana with enzymatic methyl sequencing." *Epigenetics & chromatin* vol. 13,1 42. 7 Oct. 2020, doi:10.1186/s13072-020-00361-9

29. Guimaraes, Carla P., et al. "Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions." *Nature protocols* 8.9 (2013): 1787-1799.

30. Hellman, Lance M., and Michael G. Fried. "Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions." *Nature protocols* 2.8 (2007): 1849-1861.

31. Jørgensen, Helle F., Ittai Ben-Porath, and Adrian P. Bird. "Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains." *Molecular and cellular biology* 24.8 (2004): 3387-3395.

32. Clouaire, Thomas, et al. "Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA." *Nucleic acids research* 38.14 (2010): 4620-4634.

33. LiCata, Vince J., and Andy J. Wowor. "Applications of fluorescence anisotropy to the study of protein–DNA interactions." *Methods in cell biology* 84 (2008): 243-262.

34. Meehan, Richard, Joe D. Lewis, and Adrian P. Bird. "Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA." *Nucleic acids research* 20.19 (1992): 5085-5092.

35. Lei, Ming, et al. "Plasticity at the DNA recognition site of the MeCP2 mCG-binding domain." *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1862.9 (2019): 194409.

36. Kinde, Benyam, et al. "Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2." *Proceedings of the National Academy of Sciences* 112.22 (2015): 6800-6806.

37. Saito, Motoki, and Fuyuki Ishikawa. "The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2." *Journal of Biological Chemistry* 277.38 (2002): 35434-35439.

38. Yildirim, Ozlem, et al. "Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells." *Cell* 147.7 (2011): 1498-1510.

39. Liu, Ke, et al. "Structural analyses reveal that MBD 3 is a methylated CG binder." *The FEBS Journal* 286.16 (2019): 3240-3254.

40. Korba, Brent E., and John B. Hays. "Partially deficient methylation of cytosine in DNA at CCTAGG sites stimulates genetic recombination of bacteriophage lambda." *Cell* 28.3 (1982): 531-541.

41. Kairdolf, Brad A., et al. "Semiconductor quantum dots for bioimaging and biodiagnostic applications." *Annual review of analytical chemistry* 6 (2013): 143-162.

42. Mortensen, Kim I., et al. "Optimized localization analysis for single-molecule tracking and super-resolution microscopy." *Nature methods* 7.5 (2010): 377-381.

43. Niekamp, Stefan, et al. "Nanometer-accuracy distance measurements between fluorophores at the single-molecule level." *Proceedings of the National Academy of Sciences* 116.10 (2019): 4275-4284.

44. Ladoux, Benoit, et al. "Direct imaging of single-molecules: from dynamics of a single dna chain to the study of complex dna-protein interactions." *Science Progress* 84.4 (2001): 267-290.

45. Zimm, Bruno H. "Extension in flow of a DNA molecule tethered at one end." *Macromolecules* 31.18 (1998): 6089-6098.

46. Perkins, Thomas T., et al. "Stretching of a single tethered polymer in a uniform flow." *Science* 268.5207 (1995): 83-87.

47. Kim, Soohong, et al. "Enzymatically Incorporated Genomic Tags for Optical Mapping of DNA‐Binding Proteins." *Angewandte chemie international edition* 51.15 (2012): 3578-3581.

48. Gouil, Quentin, and Andrew Keniry. "Latest techniques to study DNA methylation." *Essays in biochemistry* 63.6 (2019): 639-648.

49. Shim, Jiwook, et al. "Detection and quantification of methylation in DNA using solid-state nanopores." *Scientific reports* 3.1 (2013): 1389.

50. Fang Lim, Shuang, et al. "DNA methylation profiling in nanochannels." *Biomicrofluidics* 5.3 (2011): 034106.

51. Cipriany, Benjamin R., et al. "Single molecule epigenetic analysis in a nanofluidic channel." *Analytical chemistry* 82.6 (2010): 2480-2487.

52. Okamoto, Yukihiro, et al. "Identification of single molecular DNA methylation points by microfluidic DNA molecule stretching and quantum dot detection." *16th International Conference on Miniaturized Systems for Chemistry and Life Sciences, MicroTAS 2012*. Chemical and Biological Microsystems Society, 2012.

53. Heck, Christian, et al. "Analytical epigenetics: single-molecule optical detection of DNA and histone modifications." *Current opinion in biotechnology* 55 (2019): 151-158.

54. Heimer, Brandon W., Brooke E. Tam, and Hadley D. Sikes. "Characterization and directed evolution of a methyl-binding domain protein for high-sensitivity DNA methylation analysis." *Protein Engineering, Design & Selection* 28.12 (2015): 543-551.

55. Klose, Robert J., et al. "DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG." *Molecular cell* 19.5 (2005): 667-678.

56. Harris, R. Alan, et al. "Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications." *Nature biotechnology* 28.10 (2010): 1097-1105.

57. Neary, Jennifer L., et al. "Comparative analysis of MBD-seq and MeDIP-seq and estimation of gene expression changes in a rodent model of schizophrenia." *Genomics* 109.3-4 (2017): 204-213.

58. Lövkvist, Cecilia, et al. "DNA methylation in human epigenomes depends on local topology of CpG sites." *Nucleic acids research* 44.11 (2016): 5123-5132.

59. Chodavarapu, Ramakrishna K., et al. "Relationship between nucleosome positioning and DNA methylation." *Nature* 466.7304 (2010): 388-392.

60. Khrapunov, Sergei, et al. "Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity." *Biochemistry* 53.21 (2014): 3379-3391.

61. Rao, Satyanarayan, et al. "Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein–DNA binding." *Epigenetics & chromatin* 11 (2018): 1-11.

62. Ho, Kok Lian, et al. "MeCP2 binding to DNA depends upon hydration at methyl-CpG." *Molecular cell* 29.4 (2008): 525-531.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Luke Strauskulage

6A61ECBBE7A049F...   Author Signature

3/20/2023

Date