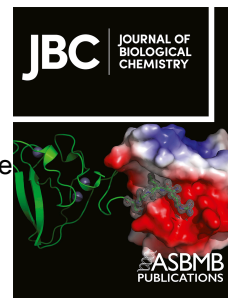


# Journal Pre-proof



Densely methylated DNA traps Methyl-CpG Binding Domain protein 2 but permits free diffusion by Methyl-CpG Binding Domain protein 3

Gage O. Leighton, Elizabeth Marie Irvin, Parminder Kaur, Ming Liu, Changjiang You, Dhruv Bhattaram, Jacob Piehler, Robert Riehn, Hong Wang, Hai Pan, David C. Williams, Jr.

PII: S0021-9258(22)00871-7

DOI: <https://doi.org/10.1016/j.jbc.2022.102428>

Reference: JBC 102428

To appear in: *Journal of Biological Chemistry*

Received Date: 25 March 2022

Revised Date: 19 August 2022

Accepted Date: 20 August 2022

Please cite this article as: Leighton GO, Irvin EM, Kaur P, Liu M, You C, Bhattaram D, Piehler J, Riehn R, Wang H, Pan H, Williams DC Jr, Densely methylated DNA traps Methyl-CpG Binding Domain protein 2 but permits free diffusion by Methyl-CpG Binding Domain protein 3, *Journal of Biological Chemistry* (2022), doi: <https://doi.org/10.1016/j.jbc.2022.102428>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.

**Densely methylated DNA traps Methyl-CpG Binding Domain protein 2 but permits free diffusion by Methyl-CpG Binding Domain protein 3**

Gage O. Leighton<sup>1,†</sup>, Elizabeth Marie Irvin<sup>2,†</sup>, Parminder Kaur<sup>3,4</sup>, Ming Liu<sup>3</sup>, Changjiang You<sup>5</sup>, Dhruv Bhattaram<sup>6</sup>, Jacob Piehler<sup>5</sup>, Robert Riehn<sup>3</sup>, Hong Wang<sup>2,3,4</sup>, Hai Pan<sup>3,7\*</sup>, David C. Williams Jr<sup>1,\*</sup>

<sup>1</sup>Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA

<sup>2</sup>Toxicology Program, <sup>3</sup>Department of Physics, <sup>4</sup>Center for Human Health and the Environment, North Carolina State University, Raleigh, North Carolina, NC 27695, USA

<sup>5</sup>Department of Biology and Center for Cellular Nanoanalytics (CellNanOs), Universität Osnabrück, Osnabrück 49076, Germany.

<sup>6</sup>Department of Biomedical Engineering, Georgia Institute of Technology & Emory University of Medicine, Atlanta, Georgia, USA

<sup>7</sup>Current address: Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou 325001, P.R. China

† Contributed equally

\*To whom correspondence may be addressed. Tel: 919-843-9949; Fax: 919-966-6718

Email: david\_willjr@med.unc.edu. Correspondence may also be addressed to: Email: hpan@ucas.ac.cn

***Running Title***

Methylation-dependent DNA binding dynamics of MBD2 and MBD3

***Keywords:***

DNA methylation, epigenetics, DNA binding protein, gene regulation, single-molecule biophysics, atomic force microscopy (AFM)

**Abstract**

The methyl-CpG binding domain 2 and 3 proteins (MBD2 and MBD3) provide structural and DNA-binding function for the Nucleosome Remodeling and Deacetylase (NuRD) complex. The two proteins form distinct NuRD complexes and show different binding affinity and selectivity for methylated DNA. Previous studies have shown that MBD2 binds with high affinity and selectivity for a single methylated CpG dinucleotide while MBD3 does not. However, the NuRD complex functions in regions of the genome that contain many CpG dinucleotides (CpG islands). Therefore, in this work we investigate the binding and diffusion of MBD2 and MBD3 on more biologically relevant DNA templates that contain a large CpG island or limited CpG sites. Using a combination of single-molecule and biophysical analyses, we show that both MBD2 and MBD3 diffuse freely and rapidly across unmethylated CpG-rich DNA. In contrast, we found methylation of large CpG islands traps MBD2 leading to stable and apparently static binding on the CpG island while MBD3 continues to diffuse freely. In addition, we demonstrate both proteins bend DNA, which is augmented by methylation. Together, these studies support a model in which MBD2-NuRD strongly localizes to and compacts methylated CpG islands while MBD3-NuRD can freely mobilize nucleosomes independent of methylation status.

## Introduction

The methyl-CpG binding domain (MBD) family of proteins binds methylated DNA through a conserved domain that recognizes the symmetrically related methylcytosines in a cytosine-guanosine dinucleotide (CpG) (1). The structure of this domain bound to a single methylated CpG site has been determined for most members of the MBD family (2–9). However, biologically relevant differential DNA methylation occurs within regions of the genome that contain tens to hundreds of CpG sites (CpG islands) (10–14). Furthermore, methylation of CpG islands in promoters and enhancers correlates with nucleosome occupancy, chromatin compaction, and associated gene silencing. Hence, we have investigated how MBD proteins bind and diffuse along these CpG islands to better understand the functional consequences in a more biologically relevant context.

In the current work, we focus on the structure and dynamics of the MBD2 and MBD3 proteins. These two highly homologous proteins arose from a duplication of the ancestral MBD present across the animal kingdom (1, 15). They contribute to the structure and function of the Nucleosome Remodeling and Deacetylase (NuRD) (16) complex that can reposition nucleosomes, deacetylate histones, and modify gene expression. The NuRD complex (17–20) consists of a least six additional proteins, each of which has multiple paralogs that provide histone deacetylase activity (HDAC1/2), histone binding, and chromatin remodeling function (CHD3/4), and protein-protein interactions (GATAD2A/B, RBBP4/7, MTA1/2/3, CDK2AP1). The MBD2 and MBD3 proteins form distinct NuRD complexes that appear to have unique functional roles (16, 21, 22).

The two proteins show different levels of selectivity for methylated CpGs attributable primarily to a single amino acid change from tyrosine (MBD2) to phenylalanine (MBD3) within the DNA binding site (Fig. 1) (1, 6, 23, 24). MBD2 shows up to 100-fold selectivity for a fully methylated CpG dinucleotide compared to an unmethylated CpG (1, 4, 25, 26). In contrast, MBD3 binds DNA with an overall much lower affinity and shows no or slight (3-5 fold) selectivity

for a methylated CpG (6, 26). Consistent with this binding difference, previous genomic localization studies found that MBD2-NuRD predominantly binds at heavily methylated CpG islands associated with silenced genes, whereas MBD3 localizes to methylated and unmethylated CpG islands associated with expressed genes (23, 27–29). However, recent data suggests the alternative interpretation that both MBD2 and MBD3 depend on methylation for proper localization across the genome (30). Therefore, how methylation selectivity of the MBD2 and MBD3 proteins impacts this localization and function remains an open question in the field.

In previous work, we used single-molecule analyses to study the behavior of the isolated MBD from MBD2 on various DNA substrates (31). Consistent with NMR and bulk biochemical studies, we found a remarkable difference in DNA bending and sliding of the MBD from MBD2 (MBD2<sub>MBD</sub>) on methylated and unmethylated DNA containing CpG islands. MBD2<sub>MBD</sub> is mostly restricted to the methylated CpG islands, while it freely diffuses across unmethylated CpG islands. Furthermore, we also uncovered a novel role for the intrinsically disordered region (IDR) of MBD2 in DNA bending (25). The DNA bending angle induced by both the MBD and a small portion of the IDR (MBD2<sub>MBD+IDR</sub>) on unmethylated CpG-rich DNA is larger than observed for CpG-free and further increases upon binding methylated CpG-rich DNA.

In separate structural studies of MBD3, we found that the MBD from MBD3 shows only weak selectivity for a single mCpG within a small (17 bp) dsDNA fragment (6). Based on a combination of chemical shift analyses, mutagenesis, and residual dipolar coupling measurements, we showed that MBD3 exchanges rapidly between CpG specific and nonspecific binding modes, leading to chemical shift averaging between these two states. Hence, MBD3 recognizes a methylated CpG site, as evidenced by significant chemical shift changes, but does not strongly localize to this site when bound to DNA. This difference between the DNA binding dynamics of MBD2 and MBD3 correlates with the prior localization studies that show both MBD2 and MBD3 localize to unmethylated CpG islands, while MBD2 more exclusively localizes to methylated CpG islands (23, 27–29).

Despite these recent studies, we do not know how the remaining, largely unstructured regions of MBD2 and MBD3 influence diffusion along DNA. Furthermore, it is unclear how reduced selectivity and binding affinity of MBD3 modifies its distribution and sliding on methylated and unmethylated CpG islands, which contain many CpG sites. To address these questions, in the current studies, we use a combination of biophysical techniques, including AFM imaging (32–34), DNA tightrope assays (35–37), and NMR (5, 6) to measure the binding and sliding of MBD2 and MBD3 on methylated and unmethylated DNA substrates.

## Results

### ***MBD2<sub>sc</sub> carries out unbiased 1D diffusion on CpG-free-rich DNA and subdiffusion on CpG-free DNA***

In the DNA tightrope assay, DNA molecules are stretched under hydrodynamic flow inside a flow cell. Anchoring of stretched DNA between poly-L-lysine coated silica microspheres leads to the formation of DNA tightropes at an elongation of ~90% of the DNA contour length (**Fig. 2A**). The spatial resolution of the DNA tightrope assay was estimated to be 16 nm (37). Uniquely, DNA tightropes created using tandemly ligated DNA allow us to directly correlate DNA binding events with the underlying specific DNA sequences or structures such as three-stranded R-loops (31, 32, 37, 38). To study MBD proteins diffusion, we ligated linear DNA fragments to form DNA tightropes with CpG free or alternating CpG-free and CpG-rich regions (**Fig. 2B and Fig. S8A**). The results from our previous study revealed that the isolated MBD from MBD2, with or without a small portion of the adjacent intrinsically disordered region (MBD2<sub>MBD</sub> and MBD2<sub>MBD+IDR</sub>), carry out unbiased 1D diffusion on CpG-rich DNA but undergoes subdiffusion on CpG-free DNA. In contrast, both proteins stably and statically bind to methylated CpG (mCpG) regions.

In this study, we purified a construct that contains almost the entire length of the MBD2b isoform plus the coiled-coil region from GATAD2A (MBD2sc) to investigate how these additional regions impact DNA binding and diffusion (**Experimental Procedures, Fig. 1A**). We previously found that the MBD2sc binds DNA with an approximately 100x higher affinity than MBD2<sub>MBD</sub> (25). Consistent with these results, fluorescence anisotropy experiments showed that MBD2sc binds to DNA containing unmethylated, methylated, or no CpG site with equilibrium dissociation constants of 2.8 ( $\pm 0.1$   $\mu\text{M}$ ), 0.007 ( $\pm 0.002$   $\mu\text{M}$ ), or 9.1 ( $\pm 0.4$   $\mu\text{M}$ ), respectively (**Fig. 2C**). Hence, we questioned whether the additional binding affinity provided by the intrinsically disordered region in MBD2 would modify DNA binding and sliding. We directly addressed this question using the DNA tightrope assay. For the DNA tightrope assay, we first generated DNA substrates by tandemly ligating linear DNA fragments containing CpG-free or unmethylated CpG-free-rich sequences (**Fig. 2B**). Further, we conjugated His-tagged MBD2sc to streptavidin-coated quantum dots (SAv-QDs) through the <sup>B</sup>Ttris-NTA linker (**Fig. 2A**) (39). Following the formation of the DNA tightropes, we introduced QD-labeled MBD2sc into the flow cell. Analysis of MBD2sc on DNA tightropes revealed two populations (**Table 1**): apparently immobile throughout data acquisition and mobile molecules (**Fig. 2D**). To exclude that this apparently immobile population reflects aggregation, we categorized particles as a single protein or cluster based on their individual QD blinking rate (Table S1) which shows that only a small fraction comprises multiple proteins. To obtain diffusion coefficients for mobile MBD2sc on DNA tightropes, we tracked the position of MBD2sc-QDs on DNA by Gaussian fitting to kymographs (particle position versus time plots) (36, 40). We obtained diffusion coefficients and alpha exponents by fitting the MSD versus time. An alpha exponent of 1 indicates an unbiased random walk, and a value less than 1 indicates subdiffusion (41). The diffusion coefficients displayed by MBD2sc on the CpG-free DNA tightrope were significantly slower than those on DNA tightropes containing CpG sites (CpG-free-rich) (**Fig. 2E**). In addition, on the CpG-free-rich tightropes, MBD2sc displayed alpha exponents close to 1 ( $1.0 \pm 0.2$ ), indicating largely unbiased 1D diffusion on DNA (**Fig. 2F**,

**Table 2**). As compared to CpG-free-rich DNA, the alpha exponents for MBD2sc on CpG-free DNA tightropes were slightly ( $p < 0.001$ ) reduced ( $0.8 \pm 0.2$ , **Fig. 2F**, **Table 2**). Overall, MBD2sc displayed slightly different diffusion ranges on CpG-free DNA tightropes compared to CpG-free-rich DNA tightropes containing multiple CpG sites (**Fig. S3**). In summary, MBD2sc shows more rapid and extensive 1D diffusion on CpG-free-rich DNA as compared to CpG-free sequences.

### ***MBD2sc statically binds to methylated CpG regions***

To evaluate how DNA methylation affects the dynamics of MBD2sc on DNA, we imaged QD-labeled MBD2sc on the CpG-free-rich DNA tightropes after methylation. Linear DNA substrates were methylated before ligation using CpG Methyltransferase (M.SssI) with S-adenosine methionine (SAM) as a cofactor (**Experimental Procedures**). We confirmed methylation of the linear CpG-free-rich DNA substrate by digestion with the methylation-sensitive HpaII restriction endonuclease (**Fig. S1**). We then tandemly ligated the methylated CpG-free-rich (mCpG-free-rich, **Fig. 3A**) and used it to form DNA tightropes between silica beads. Compared to unmethylated CpG-free-rich and CpG-free DNA, the binding density of MBD2sc increased approximately 4x on mCpG-free-rich DNA tightropes (**Fig. 3B**, **Fig. S4A**). This result is consistent with the preferential binding of MBD2sc to mCpG sites, as supported by binding affinity measurements (**Fig. 2C**). Notably, while the majority of MBD2sc observed on unmethylated CpG-free-rich DNA tightropes was mobile (77%), on mCpG-free-rich tightropes the majority of the protein bound was apparently immobile (96%) throughout data acquisition (**Fig. 3B** and **Table 1**). Furthermore, the distance between adjacent proteins on the mCpG-free-rich DNA tightropes is Gaussian distributed, with the peak centered at  $2.3 (\pm 0.3 \mu\text{m})$  (**Fig. 3C**). This spacing matches the calculated distances between adjacent mCpG-rich regions on DNA tightropes (**Fig. 3A**), considering that they are stretched to ~90% of their contour length. Taken together, fluorescence imaging of MBD2sc on DNA tightropes establishes that MBD2sc recognizes methylated CpG islands through stable and apparently static binding.



### ***MBD2sc bends unmethylated and methylated DNA upon binding***

To evaluate whether MBD2sc affects the DNA conformation upon binding, we applied AFM imaging in air to visualize the MBD2sc-DNA complexes on linear unmethylated or methylated CpG-free-rich DNA (**Fig. 4**). Based on the central location of the CpG-rich region, we identify proteins bound to this region when they are located between 38-50% from either end of the linearized DNA. The heights of MBD2sc on DNA ( $0.73 \pm 0.05$  nm), was significantly ( $p < 0.001$ ) greater than that of dsDNA itself ( $0.31 \pm 0.03$  nm), allowing us to unambiguously identify the protein-DNA complexes and determine whether they are located within the CpG-rich or CpG-free regions. The binding position analysis revealed that MBD2sc preferentially binds to the CpG-rich region on both the unmethylated and methylated DNA substrates (**Fig. 4C&D**). Similar to what we discovered previously for the isolated MBD2<sub>MBD</sub>, MBD2sc induced DNA bending when localized to the CpG-free and CpG-rich regions (**Fig. 4E**). The DNA bending angles caused by MBD2 at the CpG-rich region ( $77 \pm 27^\circ$ ) were slightly larger than observed on the CpG-free region ( $49 \pm 28^\circ$ , **Table 3**). In contrast, MBD2sc induced significantly ( $p < 0.001$ ) larger bending angles ( $91 \pm 27^\circ$ ) at the mCpG-rich region (**Fig. 4F** and **Table 3**). In summary, MBD2sc bends CpG-free and unmethylated CpG-rich DNA substrates, promoting additional DNA bending when interacting with mCpG DNA.

### ***MBD3 diffuses freely on methylated CpG DNA substrates***

Previous studies showed that MBD3 lacks a high affinity for CpG islands regardless of methylation status. Our recent study by NMR showed that MBD3 binds DNA with low affinity and shows only a slight preference for methylated DNA (6). Fluorescence anisotropy experiments demonstrated that MBD3sc binds to DNA containing unmethylated, methylated, or no CpG sites with equilibrium dissociation constants of  $8.1 (\pm 0.4 \mu\text{M})$ ,  $5.9 (\pm 0.4 \mu\text{M})$ , and  $14.2 (\pm 0.3 \mu\text{M})$ , respectively (**Fig. 2C**). To investigate how MBD3 differs from MBD2 in binding to CpG islands, we analyzed its binding to DNA tigtropes using the same set of DNA substrates (CpG-free, CpG-free-rich, and mCpG-free-rich, **Fig. 5A**). Similar to MBD2sc, MBD3sc showed

both static and mobile populations on these DNA substrates (**Table 1**). Compared to MBD2sc, MBD3sc static binding rates appeared unaffected by CpG island presence ( $47 \pm 7\%$  on CpG-free-rich and  $54 \pm 8\%$  on CpG-free) and methylation status ( $52 \pm 10\%$  on mCpG-free-rich) (**Table 1**). MBD3sc shows a much lower fraction of static binding events on mCpG-free-rich tigtropes than MBD2sc.

We fit the MSD versus time to obtain the diffusion coefficients and alpha exponents of mobile MBD3sc on different DNA substrates. MBD3sc showed slower diffusion on CpG-free-rich and CpG-free DNA tigtropes than MBD2sc (**Fig. 5** and **Table 2**). Furthermore, the alpha exponents of MBD3sc on CpG-free-rich and CpG-free DNA were slightly less than 1, suggesting constrained and subdiffusive motion. Surprisingly, MBD3sc showed faster diffusion on methylated CpG-free-rich tigtropes and an alpha exponent of  $1.0 (\pm 0.3)$ . This faster diffusion of MBD3sc on mCpG-free-rich DNA contrasts with the apparently static binding by MBD2sc. Additionally, we measured the diffusion ranges of MBD3sc on these DNA tigtropes (**Fig. S3B**). MBD3sc appears to diffuse over a more extensive range on mCpG-free-rich than on CpG-free and CpG-free-rich DNA substrates ( $p < 0.001$ ). It is worth noting that the diffusion range of MBD3sc on the methylated DNA tigtrope was increased to  $1.6 (\pm 0.8 \mu\text{m})$ . Taken together, MBD3sc appears to exhibit subdiffusion on unmethylated DNA substrates. Moreover, the majority of MBD3sc on mCpG-free-rich DNA is mobile and moves through unbiased 1D diffusion.

### ***Unmethylated and methylated DNA bending upon MBD3sc binding***

To further investigate the DNA conformational changes upon MBD3sc binding, we used AFM to visualize MBD3sc-DNA complexes on different substrates (**Fig. 6A&B**). The binding position analysis showed that MBD3sc displayed a slight preference for the CpG (42%, **Fig. 6C**) and mCpG (45%) DNA regions (38% to 50% of DNA length) (**Fig. 6D**), though it was weaker than MBD2sc. Similarly, the bending angles induced upon MBD3sc binding at the CpG-rich region ( $68 \pm 28^\circ$ ) and mCpG-rich region ( $82 \pm 30^\circ$ ) were significantly larger ( $p < 0.001$ ) than on

CpG-free region ( $47 \pm 26^\circ$  and  $36 \pm 26^\circ$ ) (**Fig. 6E&F** and **Fig. S5**). Though the DNA bending angle induced by MBD3sc at the mCpG-rich region was similar to MBD2sc, it is distinctly smaller at the CpG-free region than MBD2sc (**Fig. S5**). To sum up, MBD3sc bends CpG-free and CpG-free-rich DNA upon binding. Similar to MBD2sc, MBD3sc promotes additional bending when binding to the mCpG-rich region.

### ***MBD3sc forms DNA-DNA pairing tracts upon binding on methylated DNA substrates***

Although MBD3sc bends DNA similarly to MBD2sc, we observed DNA-DNA pairing tracts formed upon MBD3sc binding (8% of 1059 DNA molecules), but not MBD2sc (**Fig. 7A**). Approximately half of the DNA-DNA pairing tracts mediated by MBD3sc (300 nM) were formed within the methylated CpG-rich region (48%, N=48). The average tract length was  $0.08 (\pm 0.02 \mu\text{m})$ , which was much shorter than the length of the CpG-rich region (1697 bps,  $\sim 0.56 \mu\text{m}$ ). To confirm that DNA-DNA pairing tract formation is due to MBD3sc binding, we increased the MBD3sc protein concentration (from 300 nM to  $1.2 \mu\text{M}$ ) but kept the DNA concentration ( $\sim 0.5 \text{ ng}/\mu\text{l}$ ) constant (**Fig. 7B**). As a result, the percentage of tract formation increased from 8% to 25% (N=278 out of 1130 DNA molecules), and 52% of the DNA-DNA pairing tracts (N=226, **Fig. 7B**) formed in the mCpG-rich region (38% to 50% from DNA end), with a significant increase of the tract length to  $0.13 (\pm 0.04 \mu\text{m})$  (**Fig. 7B**). In comparison, MBD2sc formed large clusters, not individual DNA-DNA pairing tracts, at the same protein concentration ( $1.2 \mu\text{M}$ ) and reaction conditions (**Fig. S6**). Hence, MBD3sc shows a unique capacity to induce DNA-DNA pairing tracts, the majority of which form within the mCpG-rich region and the length of these tracts expands with increasing protein concentration.

### ***MBD2 exchanges rapidly between neighboring mCpG sites***

While the DNA tightrope assays demonstrate apparently static binding of MBD2sc to methylated CpG islands on the seconds to minutes timescale, in previous work, we found that the isolated MBD from MBD2 and MBD4 can exchange between two closely spaced mCpG sites on fast NMR timescales (sub-millisecond) (5). To determine whether MBD2sc likewise

exchanges rapidly between two neighboring mCpGs, we collected nuclear magnetic resonance (NMR) spectra of MBD2sc bound to DNA substrates with CpG sites separated by 14 bps (**Fig. 8A, Experimental Procedures**). This spacing is larger than but within one standard deviation of the average spacing in the CpG-rich DNA (**Fig. S2B**). Additionally, this spacing places the mCpG sites on nearly opposite sides of B-form DNA.

As shown in **Figure 8A and S7**, select amide resonances corresponding to the MBD2 DNA binding domain (V22, I23, G27, K43) in the 2D  $^{15}\text{N}$ -HSQC spectrum demonstrate distinct chemical shifts when bound to DNA with either one or the other of the CpG sites methylated. Importantly, these chemical shift differences allow us to interrogate the rate of exchange between the two sites. When both sites are methylated, these same amides show a single peak at a position that falls between those for the singly methylated DNA. Hence, MBD2sc shows chemical shift averaging consistent with rapid exchange between the two adjacent mCpG sites. This observation indicates the exchange rates are much faster than the difference in chemical shift between the two states ( $k \gg \delta_A - \delta_B$ ), which is in the millisecond to sub-millisecond timescale. This rapid exchange contrasts with the apparent static binding for seconds to minutes as observed by single-molecule fluorescence on DNA tightropes. Therefore, we conclude that MBD2sc exchanges rapidly between neighboring CpGs, but the distance between these sites on the mCpG-rich substrate (10 bps, **Fig. S2B**) is below the spatial and temporal resolution of the fluorescence microscope. The high number of CpG sites within the island effectively trap MBD2 resulting in an apparently static binding behavior.

To further explore this hypothesis, we constructed a DNA tightrope substrate with far fewer CpG sites (CpG-mini, **Fig. 8B**) which would not be considered a CpG island as statistically defined (10, 42). Importantly, the relative spacing between CpG sites within the CpG-mini is comparable to that of the CpG-rich region (**Fig. S2B**). We observed slightly fewer apparently static binding events when MBD2sc bound to this methylated CpG-mini DNA tightrope (90%) compared to the mCpG-free-rich substrate (96%, **Table 1**). Additionally, MBD2sc showed

transient pausing (59% percent mobile) and transition events (41% percent) between separate regions (**Fig. 8C**), which were not observed on the mCpG-free-rich DNA (**Fig. 3B**). The distance of MBD2sc between pausing events was measured at 1.6 ( $\pm$  0.5  $\mu$ m) (**Fig. 8D**), which is within error of the estimated distance between CpG-mini segments (1.9  $\mu$ m, **Fig. 8B**). Lastly, MBD3sc diffusion on the methylated CpG-mini DNA tightropes exhibited unbiased free 1D diffusion (**Fig. S8**) over a range comparable to the mCpG-free-rich tightropes. These observations demonstrate that the size of methylated CpG islands may impact the frequency of apparently static binding by MBD2sc and its ability to transition between CpG-rich regions. While MBD2sc can rapidly exchange between closely spaced mCpG sites, it cannot escape large methylated CpG islands. In contrast, MBD3sc diffuses rapidly across both small and large methylated CpG regions.

## Discussion

The MBD2 and MBD3 proteins provide structural and DNA binding functionality to the NuRD complex. The two proteins form mutually exclusive NuRD complexes, but only MBD2 shows strong selectivity for methylated CpG dinucleotides. The intrinsically disordered region and coiled-coil domains of each protein bridge between the histone deacetylase core and chromatin remodeling sub-complexes of NuRD. In these studies, we used protein constructs that incorporate the MBD N-terminal DNA binding domain, the central intrinsically disordered region, and its C-terminal coiled-coil domain fused to its native binding partner, the coiled-coiled domain from GATAD2A (**Fig 1A**). In previous work, we showed that the coiled-coil domains of MBD2 and MBD3 form stable heterodimeric complexes with the coiled-coil domain of GATAD2A (43, 44). In addition, we found that connecting the GATAD2A and MBD2 coiled-coil domains with a short linker generates a highly stable monomeric domain, such that these single-chain constructs (MBD2sc and MBD3sc) are more stable in solution. Notably, including the disordered and coiled-coil regions increases the binding affinity of MBD3 for methylated DNA from a  $K_D \sim$

50 mM (6) to ~ 5.9 mM (**Fig. 2C**), which facilitates single-molecule studies under dilute conditions. Furthermore, the MBD2sc construct binds methylated DNA with approximately 100-fold greater affinity than the isolated MBD domain (25). These highly homologous constructs allow us to compare the binding and diffusion of MBD2 and MBD3 on different unmethylated and methylated DNA substrates, probing the relationship between DNA binding behavior and the length and density of the CpG-rich region.

We find that MBD2sc statically binds to the methylated CpG-free-rich DNA tightropes over the 30-second timescale of these experiments with a physical spacing ( $2.2 \pm 0.3 \mu\text{m}$ ) consistent with the distance between adjacent mCpG-rich regions ( $2.36 \mu\text{m}$ ). We further confirm that MBD2sc localizes to the mCpG-rich region by AFM (**Fig. 4**). These data show that MBD2sc preferentially and stably binds to the CpG-rich DNA with 2-fold higher occupancy than CpG-free regions. As expected, methylating the CpG-rich DNA leads to a further increase in occupancy on the CpG-rich region. In addition, localization to the methylated CpG-rich region corresponds with an increase in DNA bending (**Fig. 4**).

In contrast, MBD3sc undergoes unbiased 1D diffusion even when bound to methylated CpG-rich DNA. Moreover, the rate of diffusion of MBD3sc increases on methylated DNA (**Fig. 5B and Table 2**). Previous NMR studies demonstrated that MBD3 rapidly exchanges between CpG-specific and nonspecific binding modes (6). The current studies show that this CpG-specific binding mode promotes rapid diffusion along the CpG-island. Consistent with this hypothesis, both MBD2sc and MBD3sc show reduced and more restricted diffusion on CpG-free DNA (**Figs. 2 and 5**). Hence, the interaction with unmethylated CpG-rich DNA permits free diffusion for both proteins, whereas methylation restricts diffusion by MBD2 while allowing free diffusion by MBD3.

Since NMR analysis shows that MBD2sc rapidly exchanges between neighboring mCpGs, we questioned if reducing the CpG island's size or the number of CpGs would decrease the static binding on DNA tightropes. The 140 bps CpG-mini DNA substrate does not meet the

minimal length of a statistically defined CpG island (42) when either fully methylated (11 mCpGs) or partially methylated (2 mCpGs). Nonetheless, MBD2sc shows mainly static binding on the methylated and partially methylated CpG-mini (**Fig. 8**). Interestingly, we noticed transitions between pausing events, indicating the diffusion of MBD2sc from one CpG-mini region to another across the CpG-free region (**Fig. 8C**). Therefore, short methylated CpG regions may not sufficiently trap the MBD2-NuRD complex for biological function.

Together, these data support a model for the localization and distinct functional roles of the two proteins. MBD2 strongly localizes to large and heavily methylated CpG islands, where it can recruit NuRD to drive nucleosome positioning, histone deacetylation, and ultimately chromatin compaction. Conversely, MBD3 remains mobile on methylated DNA, allowing NuRD to freely reposition nucleosomes across or perhaps even moving them away from methylated CpG islands. This model correlates with genomic localization studies of MBD2 and MBD3 which have shown that MBD2 strongly localizes to methylated CpG islands associated with silenced genes, while MBD3 localizes to unmethylated CpG islands associated with open chromatin and expressed genes (27–29). We propose that trapping of MBD2 in heavily methylated CpG islands drives MBD2-NuRD to preferentially move nucleosomes into these same regions. Importantly, MBD2 does not bind statically to a single mCpG, but instead continues to move nucleosomes within the methylated island, ultimately stabilizing a nucleosome-rich and compacted state. In contrast, MBD3 does not get trapped by CpG-rich regions, either methylated or unmethylated, such that MBD3-NuRD can move nucleosomes both into and out of these regions which does not drive a compacted state.

Furthermore, our model suggests that MBD3-NuRD opposes the function of MBD2-NuRD by freely mobilizing nucleosomes across methylated regions. This role of MBD3-NuRD may prevent aberrant silencing by MBD2-NuRD until a minimal threshold of methylation density across a large region has been reached. Hence, the MBD2-NuRD and MBD3-NuRD remodeling



complexes have distinct functional roles that, at least in part, reflect the dynamics and distribution of MBD2 and MBD3 on methylated and unmethylated DNA.

We previously identified that MBD2 induced bending of various DNA substrates. In the studies here, we find that including the disordered and coiled-coil regions induce additional bending. We propose that the positively charged unstructured regions help neutralize the phosphate backbone, releasing water and ions which contributes to bending of the DNA (45, 46). Of note, structural studies of MBD2 and MBD3 bound to small DNA fragments have not shown evidence of DNA bending (4, 8, 9), suggesting that the bending we observe does not involve intercalation of the DNA by the protein and is more apparent over longer length-scales. In a chromatin, DNA exists as various pre-bent shapes such as DNA-wrapped nucleosomes or 3D loops. Therefore, preferential binding to bent DNA could promote, or at least allow localization of MBD2-NuRD to nucleosome rich and compacted regions of the genome. Furthermore, the small footprint of MBD on DNA (4) along with preferentially binding to bent DNA may allow MBD2 and MBD3 to bind DNA wrapped around nucleosomes. Structural (45, 46) and biophysical analyses of MBD2 bound to nucleosomes could help clarify these possibilities.

Tract formation by MBD3sc also suggests that at high concentrations, MBD3 may still promote chromatin compaction at methylated CpG islands. Furthermore, alternative interaction partners, such as associated transcription factors (47–53), could play a dominant role in localizing MBD3-NuRD complexes. Finally, these data confirm that a methylated region of only 11 CpGs is sufficient to cause MBD2sc to pause for a significant period (tens of seconds). Therefore, a few methylated CpGs, perhaps in conjunction with transcription factor recruitment, could be sufficient for gene silencing by the MBD2-NuRD complex.

Hence, we propose that the gene duplication event that generated MBD2 and MBD3 paralogues in vertebrates permitted the sub-specialization of these two related complexes. MBD2 demonstrates a much higher affinity for and likely more effectively compacts chromatin of



methylated CpG islands. In contrast, MBD3 has lost methylation selectivity and likely more efficiently maintains open chromatin at unmethylated and transcriptionally active CpG islands. Therefore, MBD2-NuRD contributes to gene silencing at methylated promoters while MBD3-NuRD helps maintain open chromatin at unmethylated, and possibly methylated, CpG island-associated promoters and enhancers.

## **Experimental procedures**

### ***Protein expression and purification***

Human MBD2 (amino acids 150-393) and MBD3 (amino acids 1-249) were cloned into a modified pET32a vector with N-terminal thioredoxin, hexahistidine, and TEV protease sites. The constructs incorporated the coiled-coil domain of GATAD2A (amino acids 137-178) at the C-terminus to create a single-chain fusion, as described previously (25). After transforming the plasmids into Rosetta2 (DE3) *Escherichia coli*, the cells were grown at 37°C in Luria broth (unlabeled) or M9 minimal media (<sup>2</sup>H-, <sup>15</sup>N-labeled) to an A<sub>600</sub> ~0.8 and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 2.5 or 4 hours, respectively. Pelleted cells were either frozen at -20°C or immediately lysed in 20 mM Tris pH 8.0, 1 M NaCl, and cComplete Protease Inhibitor Cocktail (Roche) by sonication. Using standard procedures, we clarified the lysate by centrifugation at 16,000x g for 30 minutes before purifying the protein by nickel affinity chromatography. For NMR studies, we digested the uniformly <sup>2</sup>H- and <sup>15</sup>N-labeled protein with TEV protease overnight at 4°C and removed the thioredoxin fusion tag by a second nickel affinity chromatography step. We then purified both labeled and unlabeled samples by gel filtration (Superdex-75, GE Life Sciences) and verified the purity by SDS-PAGE.

### ***Nuclear magnetic resonance spectroscopy***

We purchased complementary 38 base pairs (bps) DNA oligonucleotides from Integrated DNA Technologies (IDT) with two CpG sites separated by 14 bps. Three double-stranded DNA samples were prepared with either the first, second, or both sites symmetrically methylated

(GGAGGCGCT(mC)GGCGGCAGCCTGGAA(mC)GGAATTCTTCTA). The DNA was annealed and purified by anion exchange chromatography (MonoQ 10/100, GE Healthcare), with the concentration determined by UV before adding at 10% molar excess to  $^2\text{H}$ - and  $^{15}\text{N}$ - labeled MBD2sc. The protein:DNA complexes were buffer exchanged into 10 mM  $\text{NaPO}_4$ , pH 6.5, 150 mM NaCl, 0.02% sodium azide, 1 mM dithiothreitol, and 10%  $^2\text{H}_2\text{O}$  at a final protein concentration of 0.3 mM.  $^{15}\text{N}$ -HSQC spectra were collected on a Bruker Avance III 850 MHz magnet at 25°C, processed with NMRPipe, and analyzed in CcpNMR (54, 55).

### Fluorescence anisotropy

Protein binding to a 17-bp 6-FAM labeled DNA (IDT DNA) was performed in 50 mM HEPES pH 7.5, 100 mM NaCl, and 1 mM  $\text{MgCl}_2$ . Complementary single strand DNA were annealed and purified on a SOURCE 15Q anion exchange column (GE Life Sciences). Methylated DNA was ordered with an internal CpG symmetrically methylated once annealed (5'-FAM-CTGCCGC(mC)GAGCGCCTC-3'). A DNA substrate devoid of CpG sites (CpG-free) was identically annealed and purified (5'-FAM-CTGGCCCCAGGGCCCTC-3'). Protein was serially diluted with DNA (10 nM) before measuring polarization on a CLARIOstar microplate reader (BMG Labtech). Binding isotherms were fit in DataGraph (56) by a standard sigmoidal equation as previously described (25).

### Surface plasmon resonance

Binding affinities of MBD2sc and MBD3sc to biotinylated DNA matching our FP substrates was performed on a Biacore 8K (Cytiva). We used a buffer of 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM  $\text{MgCl}_2$ , 1 mM TCEP, and 0.05% TWEEN-20 flowing at 30  $\mu\text{L}/\text{min}$ . DNA was immobilized to a SA sensor chip by injecting 100 nM of each substrate for 60 seconds resulting in a response of ~350. Protein was rapidly injected at consecutively increasing concentrations, single-cycle kinetics (57), each lasting 120 seconds and a final dissociation of 600 seconds. Response curves were fit using the Biacore Insight Evaluation software (Figure S9).

### ***DNA substrates for single-molecule imaging***

To generate a DNA substrate incorporating a known target sequence for MBD2, we subcloned a portion (3837 bps) of the death-associated protein kinase 1 (DAPK1) promoter (chromosome 9, bases 87497573 to 87501409), which includes CpG-rich (4689 bps) and CpG-poor (2150 bps) regions. This sequence was then sub-divided into clones that contain only the CpG-rich region (CpG-rich: 4705 bps) or CpG-rich plus poor regions (CpG-rich-poor: 6839 bps) within the pGEM backbone. For a CpG-free DNA substrate, we purchased the pCpGfree-vitroNmcs plasmid (5488 bps) from InvivoGen, which does not contain any CpG dinucleotides. The CpG-rich region from the DAPK1 promoter was then sub-cloned into the pCpGfree-vitroNmcs plasmid at the *ScaI*/*NcoI* restriction sites to generate a 1697 bps CpG-rich region within a 5466 bps CpG-free backbone (CpG-free-rich DNA: 7163 bps). In addition, we subcloned a much smaller CpG rich region derived from the DAPK1 CpG-rich sequence into the *ScaI*/*NcoI* restriction sites in the pCpGfree-vitroNmcs plasmid (CpG-mini: 140 bps). Notably, the *StuI* restriction site in pCpGfree-vitroNmcs is equidistant from both ends of the cloned fragment such that *StuI* digestion generates a linear fragment with the various inserts located in the middle.

We treated purified plasmids with CpG methyltransferase (M.SssI) and S-adenosine methionine (SAM) cofactor at 37°C for 2 hours to generate methylated DNA substrates. Restriction digest with *HpaII* confirmed complete DNA methylation (**Fig. S2**). A total of 30 *HpaII* sites are present in the CpG-rich substrate. For DNA tightropes, we linearized DNA substrates with *StuI* (New England BioLabs) digestion and ligated the DNA using the Quick Ligation™ Kit (New England BioLabs) at room temperature overnight. Finally, we purified the ligated DNA samples by phenol-chloroform extraction.

### ***Protein-QD conjugation***

We purchased streptavidin-conjugated quantum dots (SAv-QDs-655) from Invitrogen. For

QD labeling of N-terminal His<sub>6</sub>-tagged MBD2sc and MBD3sc, we incubated 0.7  $\mu$ L of SAv-QDs-655 (1  $\mu$ M) with 1.5  $\mu$ L of the multivalent chelator tris-nitrilotriacetic acid (<sup>BT</sup>tris-NTA, 2  $\mu$ M) for 10 minutes at room temperature (39). We then added the His-MBD proteins (0.5  $\mu$ L of MBD2 and 1.0  $\mu$ L of MBD3, each 4  $\mu$ M) to the SAv-QD-NTA solution and incubated for an additional 10 minutes at room temperature. Finally, we diluted all samples 100x before injecting them into the flow cell in the imaging buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 1 mM MgCl<sub>2</sub>, and 1 mg/mL BSA).

### ***Fluorescence imaging and analysis***

The oblique angle total internal reflection microscopy-based particle tracking of QD-labeled proteins on DNA tightropes was performed as described previously (37). Briefly, we collected images with an inverted microscope (Nikon Ti-E) with a 100x objective (APO TIRF, Nikon). Red (655 nm) QD-protein complexes were excited at 488 nm by a solid-state laser (Sapphire DPSS). The signal was split into two channels using a dichroic mirror (T605LPXR, Chroma) and passed through an emission filter (ET655/40nm, Chroma). We assembled flow cells as described in previous studies (37). We immobilized poly-L-lysine (2.5 mg/ml, MW>300 KDa, Wako Chemicals) treated silica beads onto a coverslip surface with PEGylation and then introduced ligated DNA substrates into the flow cell with a syringe pump at a flow rate of 300  $\mu$ l/min. DNA tightrope length was controlled by adjusting the silica bead coverage resulting in an average length of 12.5  $\mu$ m (Fig S4B). The quality of the DNA tightropes is determined after each experiment by adding YOYO-1 Iodide dye (Thermo Fisher) to the flow cell at a concentration of 0.4  $\mu$ M. Any datasets with an observable relaxed or flexible tightrope appearance are discarded.

All images were collected using an EMCCD (iXon DU897, Andor Technology) at a 50 ms/frame time resolution.

The mean square displacement (MSD) as a function of time is given by:

$$MSD(n\Delta t) = \frac{1}{N-n} \sum_{i=1}^{N-n} [(x_{i+n} - x_i)^2] \quad (1)$$

where  $N$  is the total number of frames in the trajectory,  $n$  is the number of frames for different time intervals,  $\Delta t$  is the time between frames, and  $x_i$  is the position of the protein-QD in the frame  $i$ . We determined the 1-D diffusion coefficient ( $D$ ) and alpha exponent ( $\alpha$ ) by a custom routine developed in LabView Software based on the following equation :

$$MSD = 2Dt^\alpha \quad (2)$$

We categorized a protein as mobile if the diffusion coefficient was greater than  $1 \times 10^{-5} \mu\text{m}^2/\text{s}$  and the  $R^2$  value from data fitting (Equation 2) exceeded 0.9. We analyzed the diffusion range using a custom MATLAB script.

### ***AFM imaging and image analysis***

All DNA and protein samples were pre-incubated for 20 minutes at room temperature, diluted 10x in AFM buffer (25mM NaOAc, 25 mM HEPES–KOH (pH 7.5), and 10 mM  $\text{Mg}(\text{OAc})_2$ ), and deposited onto a freshly cleaved mica surface (SPI Supply). The samples were washed with purified water (MilliQ®) and dried with nitrogen gas. The final concentration of substrates deposited onto mica was  $\sim 0.5 \text{ ng}/\mu\text{l}$  and 30 nM for DNA and protein, respectively. All images were obtained using the tapping-mode in air on an MFP-3D-Bio AFM (Asylum Research). We used cantilevers (PPP-FMR, Nanosensors) with spring constants at  $\sim 2.8 \text{ N/m}$  and collected images at a scan size of  $3 \mu\text{m} \times 3 \mu\text{m}$ , a scan rate of 1–2 Hz, and a resolution of  $512 \times 512$  pixels. The DNA bending angle was analyzed using Asylum software.

### ***Statistical Analysis***

The statistical significance level based on one-way ANOVA with Tukey's test for *post-hoc* analysis was set at  $p < 0.05$  (SPSS version 27, IBM).

**Data availability**

All data presented is contained within this manuscript and is available from the authors upon request.

*Supporting information*—This article contains supporting information.

*Acknowledgements*— We would like to thank the Weninger group at NCSU for technical support.

*Funding and additional information*—This work was supported by the National Institutes of Health [R01GM107559 to H. W. and R. R., R01GM098264 and R01DK115563 to D. W.], and by the Deutsche Forschungsgemeinschaft [YO 166/1-1 to C. Y.]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

*Conflict of interest*—The authors have no conflicts of interest to declare regarding the publication of this article.

*Abbreviations*—The abbreviations used are: MBD, methyl-CpG binding domain; MBD2, methyl-CpG binding domain protein 2; MBD3, methyl-CpG binding domain protein 3; CpG, cytosine-guanosine dinucleotide; NuRD, nucleosome remodeling and deacetylase complex; IDR, intrinsically disordered region; SAV-QD, streptavidin-coated quantum dots; QD, quantum dot; NTA, nitrilotriacetic acid; AFM, atomic force microscopy; NMR, nuclear magnetic resonance; bp, base pair.

## References

1. Hendrich, B., and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538–6547
2. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell.* **105**, 487–497
3. Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., and Walkinshaw, M. D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell.* **29**, 525–531
4. Scarsdale, J. N., Webb, H. D., Ginder, G. D., and Williams, D. C., Jr (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.* **39**, 6741–6752
5. Walavalkar, N. M., Cramer, J. M., Buchwald, W. A., Scarsdale, J. N., and Williams, D. C., Jr (2014) Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. *Nucleic Acids Res.* **42**, 11218–11232
6. Cramer, J. M., Scarsdale, J. N., Walavalkar, N. M., Buchwald, W. A., Ginder, G. D., and Williams, D. C., Jr (2014) Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *J. Biol. Chem.* **289**, 1294–1302
7. Otani, J., Arita, K., Kato, T., Kinoshita, M., Kimura, H., Suetake, I., Tajima, S., Ariyoshi, M., and Shirakawa, M. (2013) Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J. Biol. Chem.* **288**, 6351–6362
8. Liu, K., Xu, C., Lei, M., Yang, A., Loppnau, P., Hughes, T. R., and Min, J. (2018) Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. *J. Biol. Chem.* **293**, 7344–7354
9. Liu, K., Lei, M., Wu, Z., Gan, B., Cheng, H., Li, Y., and Min, J. (2019) Structural analyses reveal that MBD3 is a methylated CG binder. *FEBS J.* **286**, 3240–3254
10. Deaton, A. M., and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022
11. Portela, A., Liz, J., Nogales, V., Setien, F., Villanueva, A., and Esteller, M. (2013) DNA methylation determines nucleosome occupancy in the 5'-CpG islands of tumor suppressor genes. *Oncogene.* **32**, 5421–5428
12. Collings, C. K., and Anderson, J. N. (2017) Links between DNA methylation and nucleosome occupancy in the human genome. *Epigenetics Chromatin.* **10**, 18
13. Collings, C. K., Waddell, P. J., and Anderson, J. N. (2013) Effects of DNA methylation on nucleosome stability. *Nucleic Acids Res.* **41**, 2918–2931
14. Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., Miller, J., Schlaeger, T., Daley, G. Q., and Feinberg, A. P. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **41**, 1350–1353
15. Cramer, J. M., Pohlmann, D., Gomez, F., Mark, L., Kornegay, B., Hall, C., Siraliev-Perez, E., Walavalkar, N. M., Sperlazza, M. J., Bilinovich, S., Prokop, J. W., Hill, A. L., and Williams, D. C., Jr (2017) Methylation specific targeting of a chromatin remodeling complex from sponges to humans. *Sci. Rep.* **7**, 40674
16. Le Guezennec, X., Vermeulen, M., Brinkman, A. B., Hoeijmakers, W. A., Cohen, A., Lasonder, E., and Stunnenberg, H. G. (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* **26**, 843–851



17. Torrado, M., Low, J. K. K., Silva, A. P. G., Schmidberger, J. W., Sana, M., Sharifi Tabar, M., Isilak, M. E., Winning, C. S., Kwong, C., Bedward, M. J., Sperlazza, M. J., Williams, D. C., Jr, Shepherd, N. E., and Mackay, J. P. (2017) Refinement of the subunit interaction network within the nucleosome remodelling and deacetylase (NuRD) complex. *FEBS J.* **284**, 4216–4232
18. Allen, H. F., Wade, P. A., and Kutateladze, T. G. (2013) The NuRD architecture. *Cell. Mol. Life Sci.* **70**, 3513–3524
19. Zhang, W., Aubert, A., Gomez de Segura, J. M., Karuppasamy, M., Basu, S., Murthy, A. S., Diamante, A., Drury, T. A., Balmer, J., Cramard, J., Watson, A. A., Lando, D., Lee, S. F., Palayret, M., Kloet, S. L., Smits, A. H., Deery, M. J., Vermeulen, M., Hendrich, B., Klenerman, D., Schaffitzel, C., Berger, I., and Laue, E. D. (2016) The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. *J. Mol. Biol.* **428**, 2931–2942
20. Zhang, Y., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **13**, 1924–1935
21. Leighton, G., and Williams, D. C., Jr (2020) The Methyl-CpG-Binding Domain 2 and 3 Proteins and Formation of the Nucleosome Remodeling and Deacetylase Complex. *J. Mol. Biol.* **432**, 1624–1639
22. Yu, X., Azzo, A., Bilinovich, S. M., Li, X., Dozmorov, M., Kurita, R., Nakamura, Y., Williams, D. C., Jr, and Ginder, G. D. (2019) Disruption of the MBD2-NuRD complex but not MBD3-NuRD induces high level HbF expression in human adult erythroid cells. *Haematologica.* **104**, 2361–2371
23. Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell.* **153**, 480–492
24. Saito, M., and Ishikawa, F. (2002) The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.* **277**, 35434–35439
25. Desai, M. A., Webb, H. D., Sinanan, L. M., Scarsdale, J. N., Walavalkar, N. M., Ginder, G. D., and Williams, D. C., Jr (2015) An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res.* **43**, 3100–3113
26. Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X., and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40**, 4841–4849
27. Menafrá, R., Brinkman, A. B., Matarese, F., Franci, G., Bartels, S. J., Nguyen, L., Shimbo, T., Wade, P. A., Hubner, N. C., and Stunnenberg, H. G. (2014) Genome-wide binding of MBD2 reveals strong preference for highly methylated loci. *PLoS One.* **9**, e99603
28. Shimbo, T., Du, Y., Grimm, S. A., Dhasarathy, A., Mav, D., Shah, R. R., Shi, H., and Wade, P. A. (2013) MBD3 localizes at promoters, gene bodies and enhancers of active genes. *PLoS Genet.* **9**, e1004028
29. Gunther, K., Rust, M., Leers, J., Boettger, T., Scharfe, M., Jarek, M., Bartkuhn, M., and Renkawitz, R. (2013) Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res.* **41**, 3010–3021
30. Hainer, S. J., McCannell, K. N., Yu, J., Ee, L.-S., Zhu, L. J., Rando, O. J., and Fazio, T. G. (2016) DNA methylation directs genomic localization of Mbd2 and Mbd3 in embryonic stem cells. *Elife.* 10.7554/eLife.21964
31. Pan, H., Bilinovich, S. M., Kaur, P., Riehn, R., Wang, H., and Williams, D. C., Jr (2017) CpG and methylation-dependent DNA binding and dynamics of the methylcytosine binding domain 2 protein at the single-molecule level. *Nucleic Acids Res.* **45**, 9164–9177



32. Benarroch-Popivker, D., Pisano, S., Mendez-Bermudez, A., Lototska, L., Kaur, P., Bauwens, S., Djerbi, N., Latrick, C. M., Fraisier, V., Pei, B., Gay, A., Jaune, E., Foucher, K., Cherfils-Vicini, J., Aeby, E., Miron, S., Londono-Vallejo, A., Ye, J., Le Du, M. H., Wang, H., Gilson, E., and Giraud-Panis, M. J. (2016) TRF2-Mediated Control of Telomere DNA Topology as a Mechanism for Chromosome-End Protection. *Mol. Cell.* **61**, 274–286
33. Wang, H., Tessmer, I., Croteau, D. L., Erie, D. A., and Van Houten, B. (2008) Functional characterization and atomic force microscopy of a DNA repair protein conjugated to a quantum dot. *Nano Lett.* **8**, 1631–1637
34. Wang, H., Yang, Y., Schofield, M. J., Du, C., Fridman, Y., Lee, S. D., Larson, E. D., Drummond, J. T., Alani, E., Hsieh, P., and Erie, D. A. (2003) DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14822–14827
35. Lin, J., Countryman, P., Chen, H., Pan, H., Fan, Y., Jiang, Y., Kaur, P., Miao, W., Gurgel, G., You, C., Piehler, J., Kad, N. M., Riehn, R., Opresko, P. L., Smith, S., Tao, Y. J., and Wang, H. (2016) Functional interplay between SA1 and TRF1 in telomeric DNA binding and DNA-DNA pairing. *Nucleic Acids Res.* **44**, 6363–6376
36. Kad, N. M., Wang, H., Kennedy, G. G., Warshaw, D. M., and Van Houten, B. (2010) Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single-molecule imaging of quantum-dot-labeled proteins. *Mol. Cell.* **37**, 702–713
37. Lin, J., Countryman, P., Buncher, N., Kaur, P., E, L., Zhang, Y., Gibson, G., You, C., Watkins, S. C., Piehler, J., Opresko, P. L., Kad, N. M., and Wang, H. (2014) TRF1 and TRF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres. *Nucleic Acids Res.* **42**, 2493–2504
38. Pan, H., Jin, M., Ghadiyaram, A., Kaur, P., Miller, H. E., Ta, H. M., Liu, M., Fan, Y., Mahn, C., Gorthi, A., You, C., Piehler, J., Riehn, R., Bishop, A. J. R., Tao, Y. J., and Wang, H. (2020) Cohesin SA1 and SA2 are RNA binding proteins that localize to RNA containing regions on DNA. *Nucleic Acids Res.* **48**, 5639–5655
39. Reichel, A., Schaible, D., Al Furoukh, N., Cohen, M., Schreiber, G., and Piehler, J. (2007) Noncovalent, site-specific biotinylation of histidine-tagged proteins. *Anal. Chem.* **79**, 8590–8600
40. Dunn, A. R., Kad, N. M., Nelson, S. R., Warshaw, D. M., and Wallace, S. S. (2011) Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.* **39**, 7487–7498
41. Saxton, M. J. (1997) Single-particle tracking: the distribution of diffusion coefficients. *Biophys. J.* **72**, 1744–1753
42. Illingworth, R. S., and Bird, A. P. (2009) CpG islands--'a rough guide'. *FEBS Lett.* **583**, 1713–1720
43. Gnanapragasam, M. N., Scarsdale, J. N., Amaya, M. L., Webb, H. D., Desai, M. A., Walavalkar, N. M., Wang, S. Z., Zu Zhu, S., Ginder, G. D., and Williams, D. C., Jr (2011) p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7487–7492
44. Walavalkar, N. M., Gordon, N., and Williams, D. C., Jr (2013) Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66alpha). *J. Biol. Chem.* **288**, 3419–3427
45. Privalov, P. L., Dragan, A. I., and Crane-Robinson, C. (2009) The cost of DNA bending. *Trends Biochem. Sci.* **34**, 464–470
46. Beckwitt, E. C., Kong, M., and Van Houten, B. (2018) Studying protein-DNA interactions using atomic force microscopy. *Semin. Cell Dev. Biol.* **73**, 220–230
47. Moody, R. R., Lo, M. C., Meagher, J. L., Lin, C. C., Stevers, N. O., Tinsley, S. L., Jung, I., Matvekas, A., Stuckey, J. A., and Sun, D. (2018) Probing the interaction between the

- histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. *J. Biol. Chem.* **293**, 2125–2136
48. Lejon, S., Thong, S. Y., Murthy, A., AlQarni, S., Murzina, N. V., Blobel, G. A., Laue, E. D., and Mackay, J. P. (2011) Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48-FOG-1 complex. *J. Biol. Chem.* **286**, 1196–1203
  49. Ivanochko, D., Halabelian, L., Henderson, E., Savitsky, P., Jain, H., Marcon, E., Duan, S., Hutchinson, A., Seitova, A., Barsyte-Lovejoy, D., Filippakopoulos, P., Greenblatt, J., Lima-Fernandes, E., and Arrowsmith, C. H. (2018) Direct interaction between the PRDM3 and PRDM16 tumor suppressors and the NuRD chromatin remodeling complex. *Nucleic Acids Res.* 10.1093/nar/gky1192
  50. Hong, W., Nakazawa, M., Chen, Y. Y., Kori, R., Vakoc, C. R., Rakowski, C., and Blobel, G. A. (2005) FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J.* **24**, 2367–2378
  51. Robbe, Z. L., Shi, W., Wasson, L. K., Scialdone, A. P., Wilczewski, C. M., Sheng, X., Hepperla, A. J., Akerberg, B. N., Pu, W. T., Cristea, I. M., Davis, I. J., and Conlon, F. L. (2022) CHD4 is recruited by GATA4 and NKX2-5 to repress noncardiac gene programs in the developing heart. *Genes Dev.* 10.1101/gad.349154.121
  52. Waldron, L., Steimle, J. D., Greco, T. M., Gomez, N. C., Dorr, K. M., Kweon, J., Temple, B., Yang, X. H., Wilczewski, C. M., Davis, I. J., Cristea, I. M., Moskowitz, I. P., and Conlon, F. L. (2016) The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Dev. Cell.* **36**, 262–275
  53. Miller, A., Ralser, M., Kloet, S. L., Loos, R., Nishinakamura, R., Bertone, P., Vermeulen, M., and Hendrich, B. (2016) Sall4 controls differentiation of pluripotent cells independently of the Nucleosome Remodelling and Deacetylation (NuRD) complex. *Development.* **143**, 3074–3084
  54. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* **59**, 687–696
  55. Skinner, S. P., Goult, B. T., Fogh, R. H., Boucher, W., Stevens, T. J., Laue, E. D., and Vuister, G. W. (2015) Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 154–161
  56. Adalsteinsson, D., and Schultz, P. (2022) *The DataGraph Manual*
  57. Karlsson, R., Katsamba, P. S., Nordin, H., Pol, E., and Myszka, D. G. (2006) Analyzing a kinetic titration series using affinity biosensors. *Anal. Biochem.* **349**, 136–147

### Figure Legends

**Figure 1: MBD2 and MBD3 DNA binding domains.** (A) Cartoon representation of the MBD2 (blue) and MBD3 (green) protein constructs, which include the MBD and coiled-coil domains (CC) and a fusion to the coiled-coil domain from its native binding partner GATAD2A (CR1, orange). We used the MBD2b isoform, which lacks a glycine-arginine rich N-terminal region (dashed line) unique to the MBD2a isoform in mammals. (B) A cartoon diagram depicts an alignment of the methyl-CpG binding domains from MBD2 (blue) (8) and MBD3 (green) (9) bound to DNA (yellow sticks). The methyl carbons in the two methyl-cytosine bases of the CpG dinucleotide are shown as yellow spheres. The two arginine (R166 and R188) and one tyrosine (Y178) amino acids in MBD2 critical for selectively binding methylated DNA are shown in sticks. (C) A sequence alignment of the two DNA binding domains shows that the two arginine residues are conserved (highlighted in blue), but a phenylalanine replaces tyrosine in MBD3 (highlighted in red).

### Figure 2: Binding and diffusion of MBD2sc on unmethylated DNA substrates.

(A) A schematic drawing of the DNA tightrope assay shows the flow-cell with an expanded representation of red (655 nm) QD-conjugated MBD2sc. <sup>B</sup>Tris-NTA links the His<sub>6</sub>-tag on the MBD2sc proteins to streptavidin coated-QD. (B) A cartoon drawing depicts ligated DNA substrates for the DNA tightrope assay: CpG-free and CpG-free-rich. (C) The binding affinity of MBD2sc or MBD3sc for unmethylated or methylated DNA as measured by fluorescence polarization. (D) Representative kymographs of QD-labeled MBD2sc on CpG-free and CpG-free-rich tightropes. Diffusion coefficients (E) and alpha exponents (F) of MBD2sc on CpG-free, and CpG-free-rich DNA tightropes. \*\*\*:  $p < 0.001$ . See Table 1 for detailed analysis.

**Figure 3. MBD2sc becomes static at the methylated CpG-rich regions on DNA tightropes.**

(A) A Cartoon drawing of the CpG-free-rich DNA fragment (top panel) and ligated DNA substrate containing alternating mCpG-rich and CpG-free sequences (mCpG-free-rich, bottom panel) for the DNA tightrope assay. (B) Kymograph of MBD2sc on the ligated mCpG-free-rich DNA tightrope. (C) Histogram of the distance between adjacent MBD2sc-QDs on mCpG-free-rich DNA tightropes. The solid lines represent Gaussian fit to the data ( $R^2 > 0.98$ ) with a peak centered at  $2.2 (\pm 0.3 \mu\text{m})$  ( $N = 461$ , the error bars represent mean  $\pm$  SD).

**Figure 4. MBD2sc induces DNA bending upon binding to CpG-free-rich and mCpG-free-rich DNA substrates.**

Representative AFM images of MBD2sc on the linear CpG-free-rich (A) and on mCpG-free-rich DNA (B). The XY scale bar is 200 nm. Inset: an expanded 3-D image of the indicated region. (C and D) Analysis of the binding position of MBD2sc on the linear CpG-free-rich (C) and mCpG-free-rich (D) substrates. Over 49% of MBD2sc ( $N = 95$  out of 192) binds to the CpG-rich region (38% to 50%) (C) and 60% of MBD2sc ( $N = 131$  out of 230) binds to the mCpG-rich region (D). (E and F) DNA bending angles induced by MBD2sc upon binding to the CpG-free region ( $49^\circ \pm 28^\circ$ , mean  $\pm$  SD,  $N = 97$ ) and CpG-rich region ( $77^\circ \pm 27^\circ$ ,  $N = 95$ ) on the CpG-free-rich DNA, and upon binding to the CpG-free region ( $49^\circ \pm 30^\circ$ ,  $N = 99$ ) and m-CpG-rich region ( $91^\circ \pm 27^\circ$ ,  $N = 131$ ) on the mCpG-free-rich DNA (F).

**Figure 5. DNA tightrope assay of MBD3sc on unmethylated and methylated DNA substrates.**

(A) Representative kymographs of MBD3sc on different DNA tightropes. (B and C) Diffusion coefficients (B) and alpha exponents (C) of MBD3sc on CpG-free, CpG-free-rich, and mCpG-free-rich DNA tightropes. \*:  $p < 0.05$ ; \*\*:  $p < 0.005$ ; \*\*\*:  $p < 0.001$ . See Table 1 for detailed analysis.

**Figure 6. MBD3sc induces DNA bending upon binding to CpG-free-rich and mCpG-free-rich DNA substrates.** (A and B) Representative AFM images of MBD3sc on the linear CpG-free-rich (A) and on mCpG-free-rich DNA (B). The XY scale bar is 200 nm. Inset: An expanded 3-D image of the indicated region. (C and D) Analysis of the binding position of MBD3sc on the linear CpG-free-rich and mCpG-free-rich substrates. Over 41% of MBD3sc (N = 119 out of 287) binds to the CpG-rich region (38% to 50%) (C) and 45% of MBD3sc (N = 138 out of 306) binds to the mCpG-rich region (D). (E and F) DNA bending angles induced upon MBD3sc binding to the CpG-free region ( $47^\circ \pm 26^\circ$ , mean  $\pm$  SD, N = 168) and CpG-rich region ( $68^\circ \pm 28^\circ$ , N = 119) on CpG-free-rich DNA (E), and upon binding to the CpG-free region ( $36^\circ \pm 26^\circ$ , N = 168) and mCpG-rich region on mCpG-free-rich DNA ( $82^\circ \pm 30^\circ$ , N = 138) on the mCpG-free-rich DNA (F).

**Figure 7. DNA-DNA pairing tracts formed upon MBD3sc binding on the methylated CpG-free-rich DNA substrate.** (A and B) Representative AFM images of a DNA-DNA pairing tract (circle, left panels), tract positions (middle panels), and tract lengths (right panels) upon MBD3sc binding on the linear mCpG-free-rich DNA at a lower (A) and higher (B) MBD3sc concentration. The XY scale bar is 200 nm. 48% (N = 48) and 52% (N = 226) of the tracts formed in the methylated CpG-rich region at the low and high MBD3sc concentration, respectively. The length of tracts formed upon MBD3sc binding on the linear mCpG-free-rich DNA is  $0.08 \pm 0.02 \mu\text{m}$  and  $0.13 \pm 0.04 \mu\text{m}$  (mean  $\pm$  SD) at the low and high MBD3sc concentration, respectively.

**Figure 8. MBD2sc exchanges between neighboring mCpG sites demonstrated by NMR and the DNA tightrope assay.** (A) Overlay of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of select amide resonances corresponding to the MBD2 DNA binding domain (V22, I23, G27, K43) upon binding to DNA with either the first (orange), second (blue), or both (black) CpG sites methylated. With both sites methylated, we observe an averaging of the peak indicating rapid exchange on the order of  $\leq 1$ ms. (B) Schematic drawing of the ligated mCpG-mini DNA substrate used for the DNA tightrope assay. (C) Representative fluorescence image (top) and kymograph (bottom) of MBD2sc on the mCpG-mini DNA tightrope. Arrows at the bottom pointing to pausing events. (D) Histogram of the distance between transient pausing events on mCpG-mini DNA tightropes (N = 83).

**Table 1. Fraction of statically bound MBD2sc and MBD3sc on unmethylated- and methylated-DNA tightropes.**

DNA	MBD2sc		MBD3sc	
	Static binding (%)	N	Static binding (%)	N
CpG-free-rich	23 ± 6	147	47 ± 7	123
CpG-free	19 ± 6	333	54 ± 8	240
mCpG-free-rich	96 ± 3	246	52 ± 10	345
mCpG-mini	90 ± 1	495	50 ± 8	158

The values represent mean ± SD from 2 to 3 experiments for each data set.

**Table 2. Diffusion coefficient of MBD2sc and MBD3sc on different DNA substrates.**

DNA	MBD2sc			MBD3sc		
	D ( $\mu\text{m}^2/\text{s}$ )	$\alpha$ exponent	N	D ( $\mu\text{m}^2/\text{s}$ )	$\alpha$ exponent	N
CpG-free	$0.04 \pm 0.02$	$0.8 \pm 0.2$	99	$0.04 \pm 0.03$	$0.8 \pm 0.3$	199
CpG-free-rich	$0.15 \pm 0.05$	$1.0 \pm 0.2$	100	$0.04 \pm 0.01$	$0.9 \pm 0.2$	144
mCpG-free-rich	NA.	NA.		$0.09 \pm 0.02$	$1.0 \pm 0.2$	174
mCpG-mini	NA.	NA.		$0.05 \pm 0.02$	$0.9 \pm 0.1$	70

The values represent mean  $\pm$  SD from 2 to 4 experiments for each data set.



**Table 3. Summary of DNA bending angles induced by MBD2sc and MBD3sc binding.**

DNA	MBD2sc		MBD3sc		
	Bending Angle (°)	N	Bending Angle (°)	N	
Unmethylated	CpG free region	49 ± 28	97	47 ± 26	168
	CpG rich region	77 ± 27	95	68 ± 28	119
Methylated	CpG free region	49 ± 30	99	36 ± 26	168
	CpG rich region	91 ± 27	131	82 ± 30	138

The values represent mean ± SD from 2 to 3 experiments for each data set. The significance values regarding the difference among the data sets are reported in Supplementary Fig. 7.

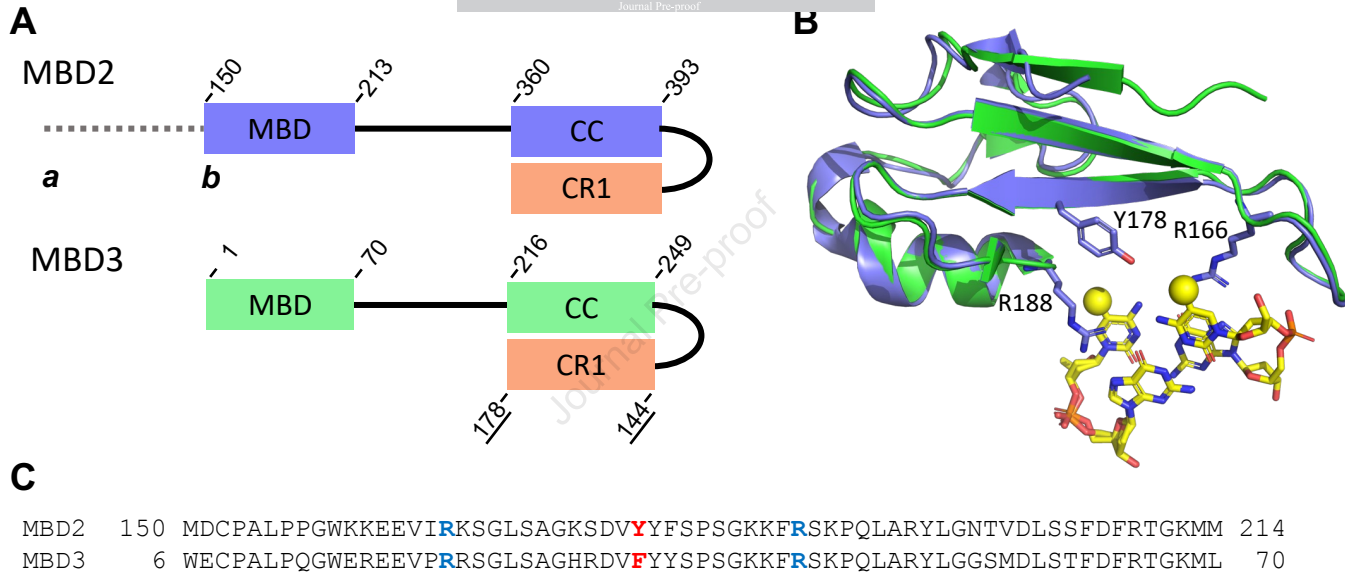


Figure 1

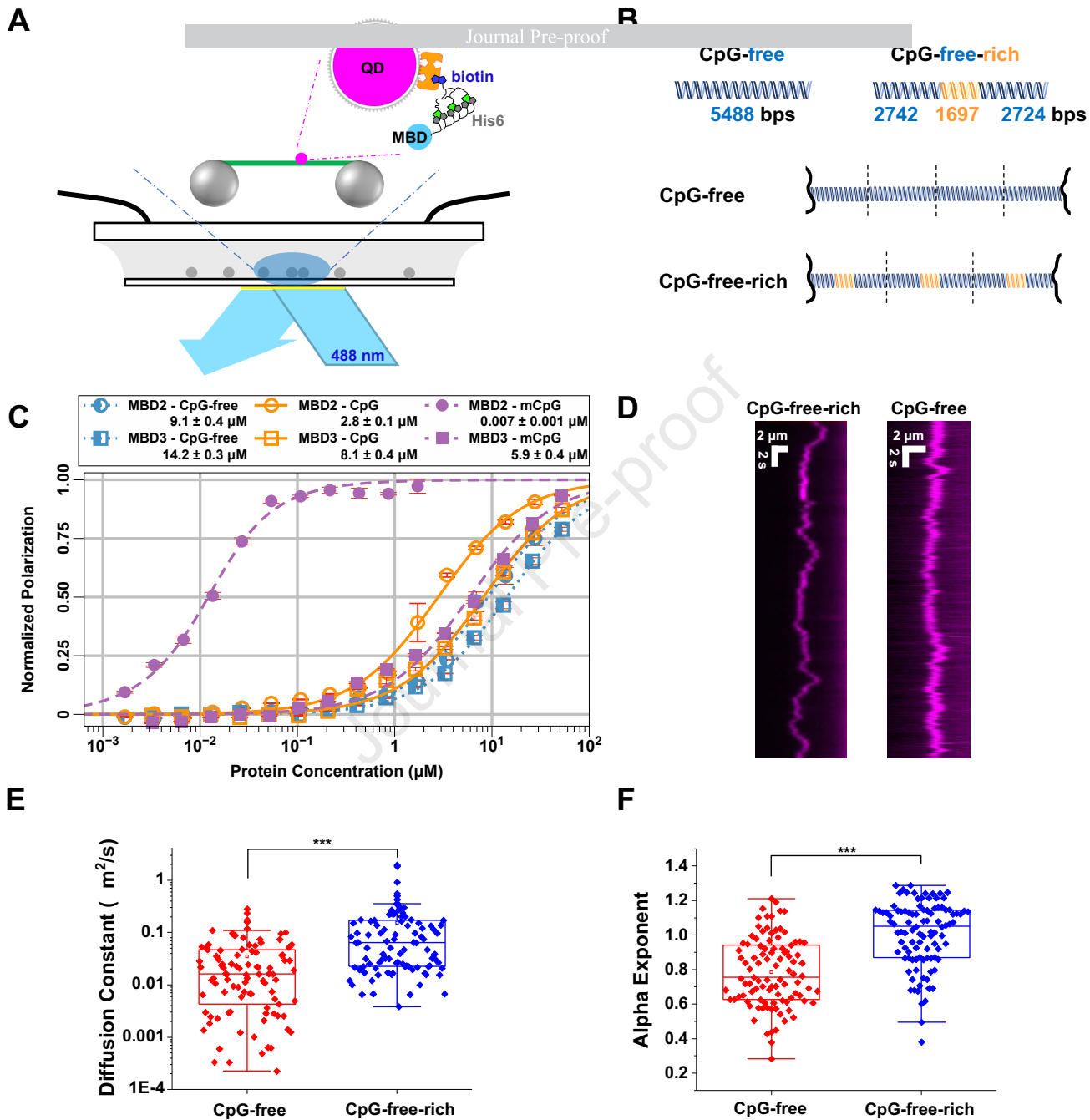
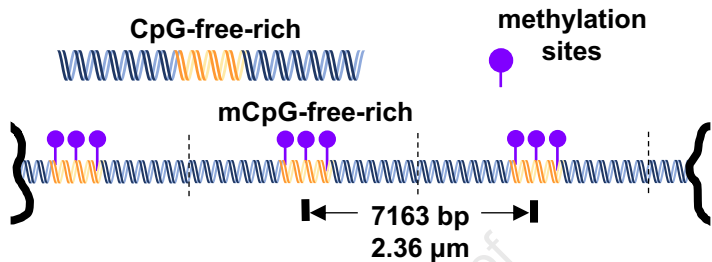
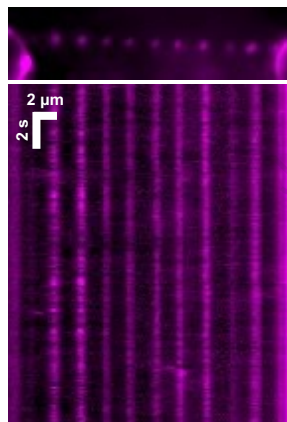


Figure 2

A



B



C

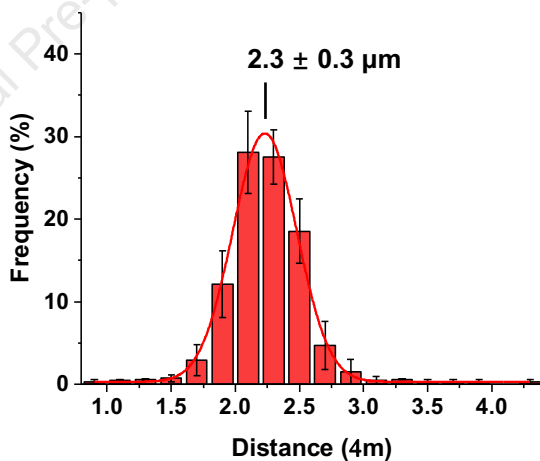
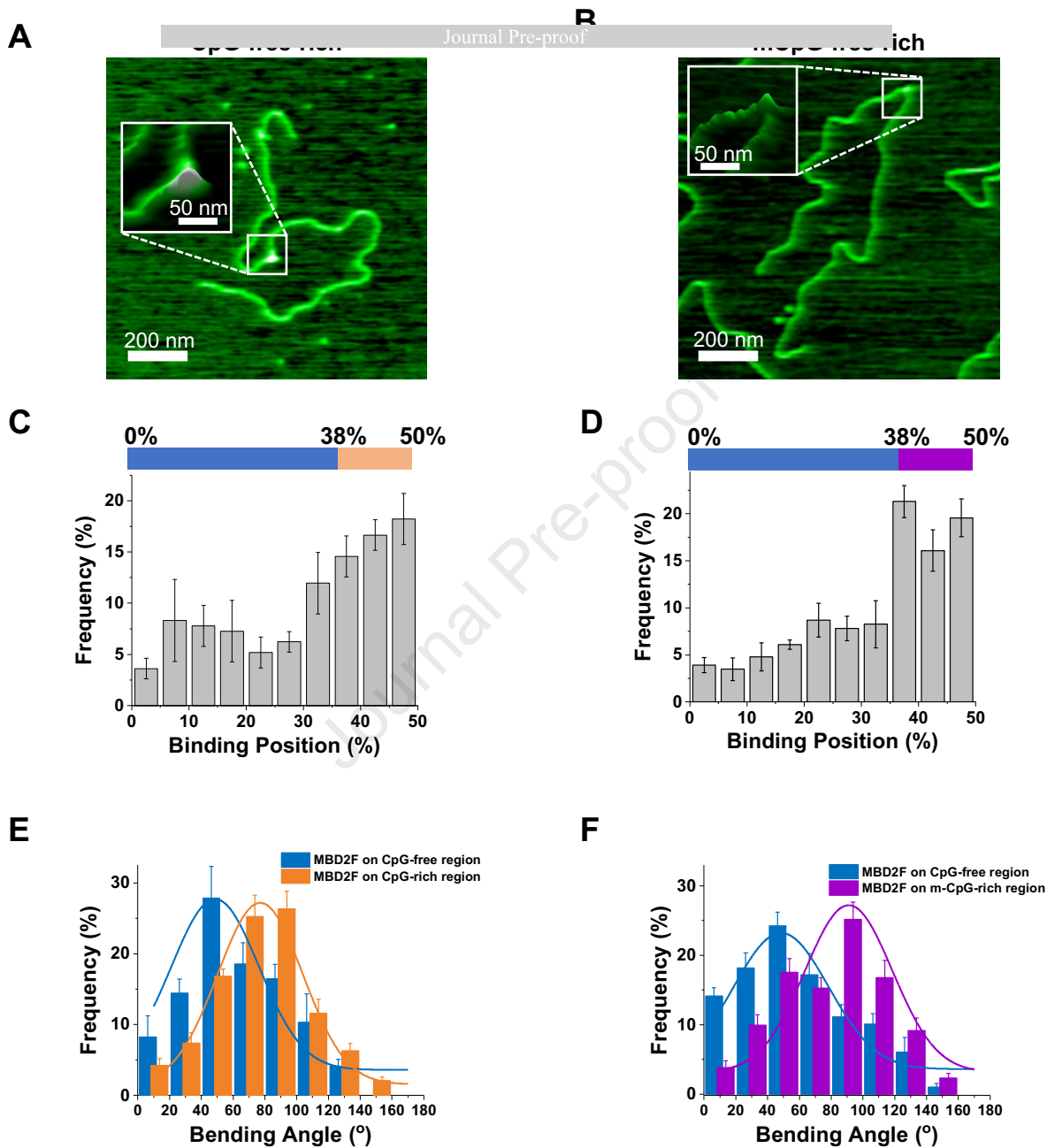
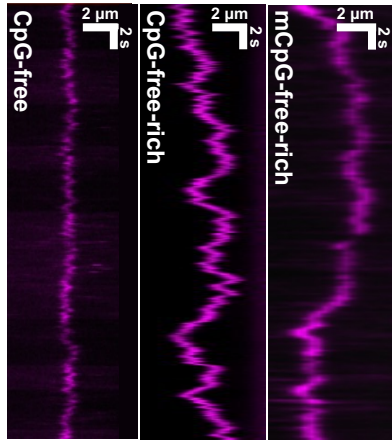
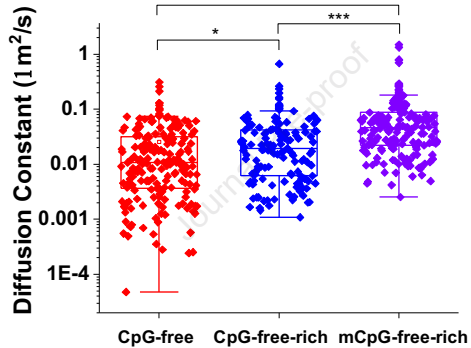
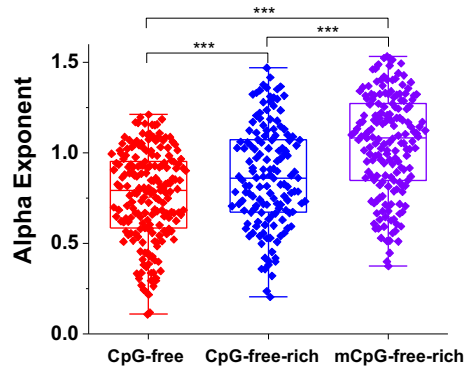
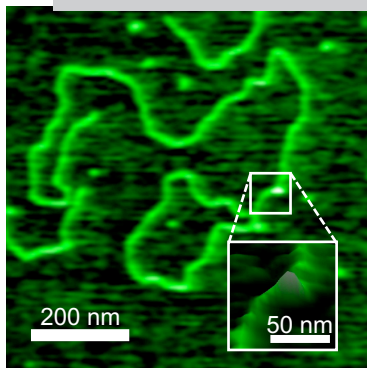
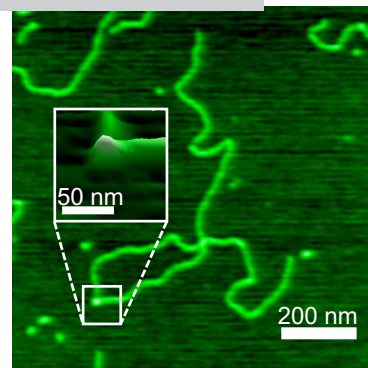
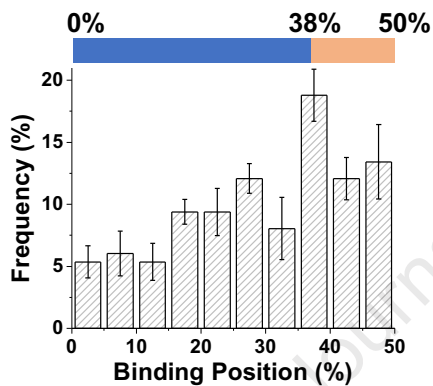
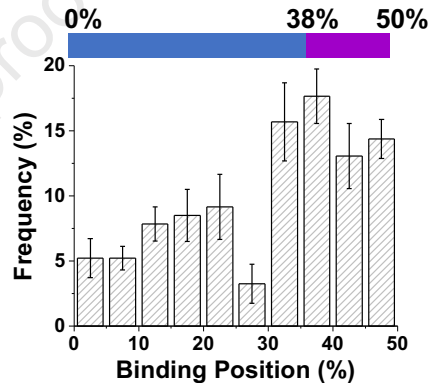
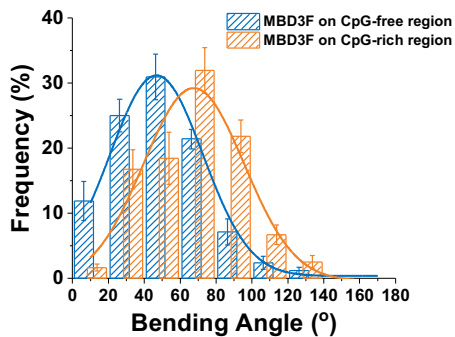
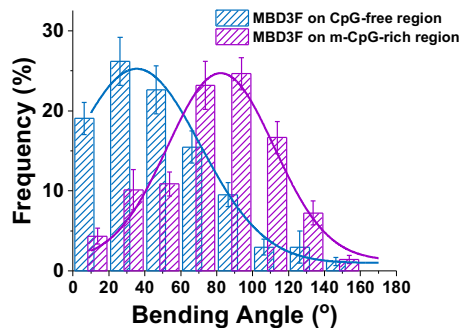


Figure 3



**A****B****C****Figure 5**

**A****B****C****D****E****F****Figure 6**

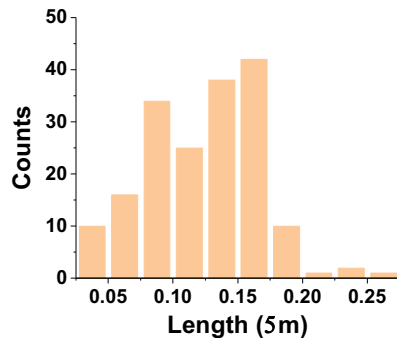
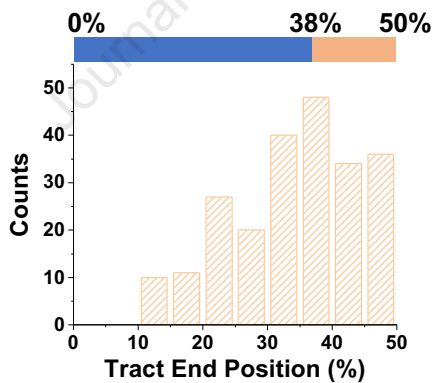
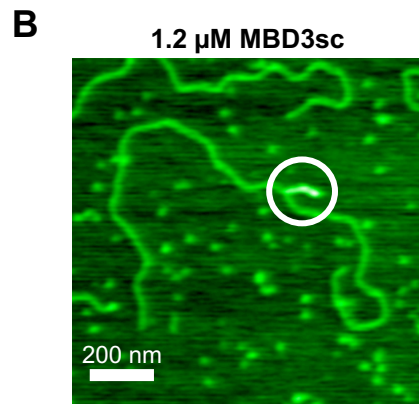
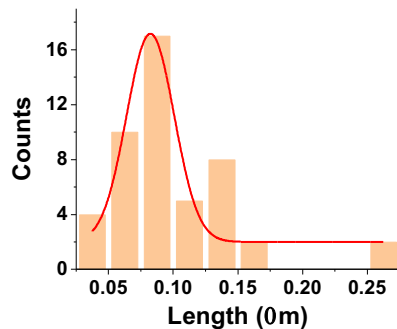
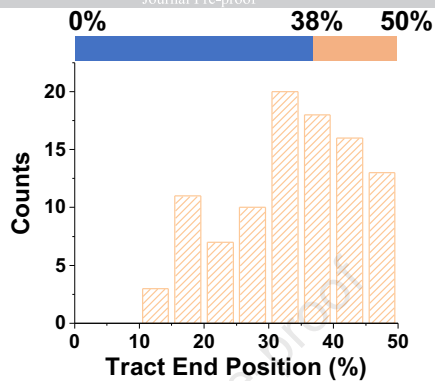
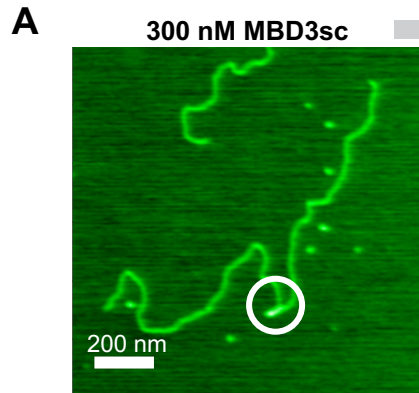
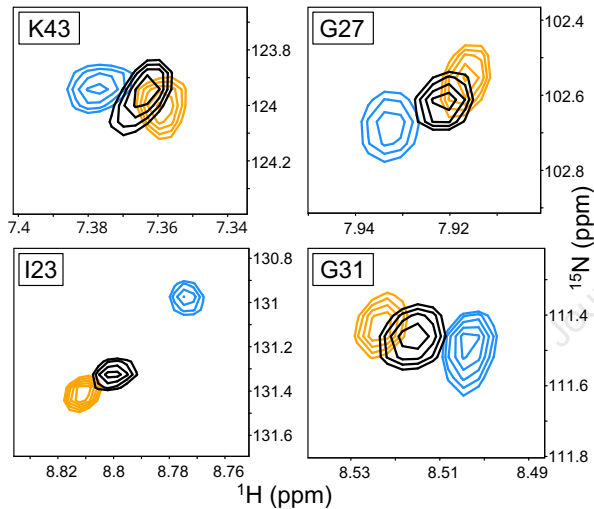
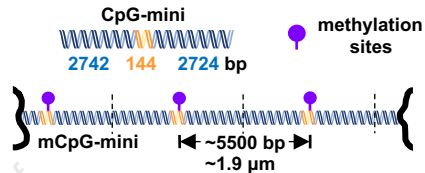
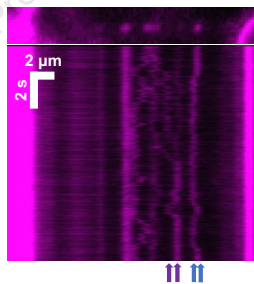
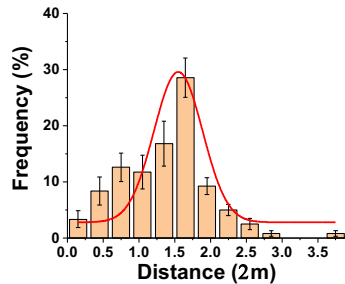


Figure 7



**A****B****C****D****Figure 8**

*Author contributions-* G.L., H.W., H.P, and D.W. conceptualization; G.L, E.I., M.L., C.Y., D.B., and H.P. investigation; D.B and J.P. resources; R.R., H.W., H.P., and D.W. supervision; G.L. and E.I. writing – original draft; H.W., H.P., and D.W. writing – reviewing and editing.

Journal Pre-proof