An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex

Megha A. Desai[1], Heather D. Webb[2], Leander M. Sinanan[3], J. Neel Scarsdale[4], Ninad M. Walavalkar[3], Gordon D. Ginder[5],* **and David C. Williams, Jr[3],***

[1]Department of Human and Molecular Genetics and Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA, [2]Department of Pathology and Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA, [3]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, [4]Institute of Structural Biology and Drug Design, Center for the Study of Biological Complexity, and Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA and [5]Departments of Internal Medicine, Human and Molecular Genetics, and Microbiology and Immunology and Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA

ABSTRACT

The MBD2-NuRD (Nucleosome Remodeling and Deacetylase) complex is an epigenetic reader of DNA methylation that regulates genes involved in normal development and neoplastic diseases. To delineate the architecture and functional interactions of the MBD2-NuRD complex, we previously solved the structures of MBD2 bound to methylated DNA and a coiled-coil interaction between MBD2 and p66α that recruits the CHD4 nucleosome remodeling protein to the complex. The work presented here identifies novel structural and functional features of a previously uncharacterized domain of MBD2 (MBD2$_{IDR}$). Biophysical analyses show that the MBD2$_{IDR}$ is an intrinsically disordered region (IDR). However, despite this inherent disorder, MBD2$_{IDR}$ increases the overall binding affinity of MBD2 for methylated DNA. MBD2$_{IDR}$ also recruits the histone deacetylase core components (RbAp48, HDAC2 and MTA2) of NuRD through a critical contact region requiring two contiguous amino acid residues, Arg^{286} and Leu^{287}. Mutating these residues abrogates interaction of MBD2 with the histone deacetylase core and impairs the ability of MBD2 to repress the methylated tumor suppressor gene *PRSS8* in MDA-MB-435 breast cancer cells. These findings expand our knowledge of the multi-dimensional interactions of the MBD2-NuRD complex that govern its function.

INTRODUCTION

Epigenetic regulation comprises heritable changes in gene expression most commonly brought about by DNA methylation and histone modifications. The predominant form of DNA methylation in mammals involves addition of a methyl group to the C5 carbon of the cytosine residue in a cytosine-guanine dinucleotide (CpG) through the enzymatic activity of DNA methyl-transferases, DNMT1, DNMT3A and DNMT3B (1). Regions of high CpG density are often associated with gene promoters (2), which remain unmethylated except for a subset of tissue-specific genes involved in normal differentiation and development (3–6). Aberrant hypermethylation of tumor suppressor gene promoters is associated with oncogenesis in a wide array of tissues (7). The methyl-CpG binding domain (MBD) family proteins recognize this methylated mark and repress the associated genes by recruiting different co-repressor complexes. The MBD family of proteins include the first identified MeCP2 (8) and MBD1, MBD2, MBD3 and MBD4 (9). With the exception of mammalian MBD3, all MBD proteins bind to methylated DNA although with varying affinities (10–13).

MBD2 binds densely methylated CpG islands and represses transcription of the associated genes through recruitment of the Nucleosome Remodeling and Deacetylase (NuRD) co-repressor complex (14). The MBD2-NuRD co-

repressor complex from both cell lines and primary cells has been characterized and is comprised of at least one copy each of the MTA1/2/3, HDAC1/2, RbAp46/48, p66α/β, and CHD3/4 and MBD2 proteins (5,14–16). Recent studies have evaluated the stoichiometry of protein interactions in the NuRD complex (15,16), however much remains to be explored about the nature and assembly of protein–protein interactions within this complex. Previous work in our laboratory identified MBD2 as a silencer of the chicken ρ-globin gene (5,17,18) as well as murine and human embryonic and fetal β-type globin genes in adult erythroid cells (4,18,19). In addition, MBD2 has been implicated in aberrant silencing of methylated tumor suppressor genes in carcinogenesis (20–26).

The MBD2 protein consists of an N-terminal glycine-arginine repeat region (GR), a methyl-binding domain (MBD) which binds *in vivo* to densely methylated DNA (27), an uncharacterized domain of MBD2 ($MBD2_{IDR}$) and a coiled-coil domain. In previous work we showed that the C-terminal coiled-coil of MBD2 binds to the p66α component of NuRD, which contributes to the recruitment of CHD4 and gene silencing. Consistent with these findings, the p66α coiled-coil domain peptide can bind to native MBD2 in cells and relieve MBD2-mediated repression of target genes such as the embryonic and fetal β-type globin genes in adult erythroid cell culture systems (18). This proof-of-concept study underscored the biological significance of functional disruption of the MBD2-NuRD co-repressor complex and led us to pursue characterization of other MBD2 mediated interactions within the NuRD complex.

Intrinsically disordered proteins (IDPs) are a rapidly advancing area of research due to their importance in human biology. Although IDPs lack a stable three-dimensional structure under physiological conditions, they can serve as hubs of multi-protein interactions for diverse cellular functions including transcription regulation, chromatin remodeling and cell signaling because the intrinsic disorder permits transient, low affinity but high specificity protein–protein and nucleic acid-protein interactions (28). Among the MBD protein family members, the transcription repression domain of MBD1 (29) and 60% of full-length MeCP2 protein have been shown to be intrinsically disordered (30), even in the presence of their binding partners.

Having previously determined the structures of the MBD and coiled-coil domains of MBD2, we present here the unique structural and functional features of the previously uncharacterized $MBD2_{IDR}$. We show that this region is intrinsically disordered in isolation and in the context of the full-length protein bound to DNA, we identify its role in modifying kinetics and affinity of DNA-binding, and map the critical sites needed for $MBD2_{IDR}$ to recruit the histone deacetylase core complex within the context of the intact MBD2 protein in cells. We anticipate these results will facilitate efforts for further biochemical and structural characterization of the MBD2-NuRD complex and open up avenues to target co-repressor activities of MBD2-NuRD via disruption of $MBD2_{IDR}$-mediated interactions with the NuRD complex.

## MATERIALS AND METHODS

Protein expression and purification

Various MBD2 constructs (Figure 1a) were cloned into a modified pET32a vector (31) including: the $MBD2_{IDR}$ of human (amino acids 238–356) and mouse (amino acids 241–359) MBD2; a human single-chain coiled-coil domain construct (scMBD2-p66α) comprised of the MBD2 coiled-coil domain (amino acids 361–393), a short GGSG linker, and the p66α coiled-coil domain (amino acids 137–178); human $MBD2_{MBD}$ construct comprised of the isolated MBD (amino acids 150–214), and a slightly longer construct that incorporates the MBD through the first twenty amino acids from the $MBD2_{IDR}$ (amino acids 150–260), and a full-length human MBD2 single-chain construct (MBD2FLsc) comprised of MBD2 (amino acids 150–393) and the scMBD2-p66α. Note that the full-length construct does not incorporate the N-terminal GR repeat region unique to mammalian MBD2 proteins. The resulting plasmids were transformed into the Rosetta II (DE3) (Invitrogen) *E. coli* strain, grown in either Luria Bertani medium (unlabeled) or M9 minimal media (^{13}C, ^{15}N, $±^2$H labeled) and induced with 1mM isopropyl-β-d-thiogalactopyranoside at an $A_{600} \sim 0.8$ for 2 or 4 h, respectively. Bacterial pellets from 1L of growth media were lysed with 30 ml of the B-PER reagent (Thermo Scientific) and the expressed fusion protein purified by nickel affinity chromatography. For analytical ultracentrifugation studies and DNA binding affinity studies, the fusion protein was further purified by gel filtration chromatography (Superdex-75 26/60, GE Healthcare). For circular dichroism (CD) and nuclear magnetic resonance (NMR) studies, the $MBD2_{IDR}$ and scMBD2-p66α were separated from the thioredoxin fusion domain by thrombin cleavage at room temperature overnight, and further purified by gel filtration (Superdex-75 26/60, GE Healthcare) and ion exchange (MonoS 10/100, GE Healthcare) chromatography. The final protein concentration was determined by UV absorbance at 280 nm.

MBD2FLsc:DNA complex preparation

Complementary 17 base pair oligonucleotides containing a central methylated CpG dinucleotide (5′-GAGGCGCT(^{me}C)GGCGGCAG-3′) were purchased from Integrated DNA Technologies (IDT), annealed and purified by ion exchange chromatography (MonoQ 10/100, GE Healthcare) as described previously (32). The concentration of purified double-stranded oligonucleotide was determined by UV absorbance at 260 nm and the DNA added in 10% molar excess to purified uniform ^{2}H, ^{13}C, ^{15}N-labeled MBD2FLsc. The resulting complex was concentrated and buffer exchanged into NMR buffer (see below).

Surface plasmon resonance (SPR) analysis

3′ biotinylated forward oligonucleotide was purchased from IDT, annealed to the reverse strand and purified as described above. Protein and DNA samples were buffer exchanged into 10 mM HEPES pH 7.4, 50 mM NaCl, 3 mM
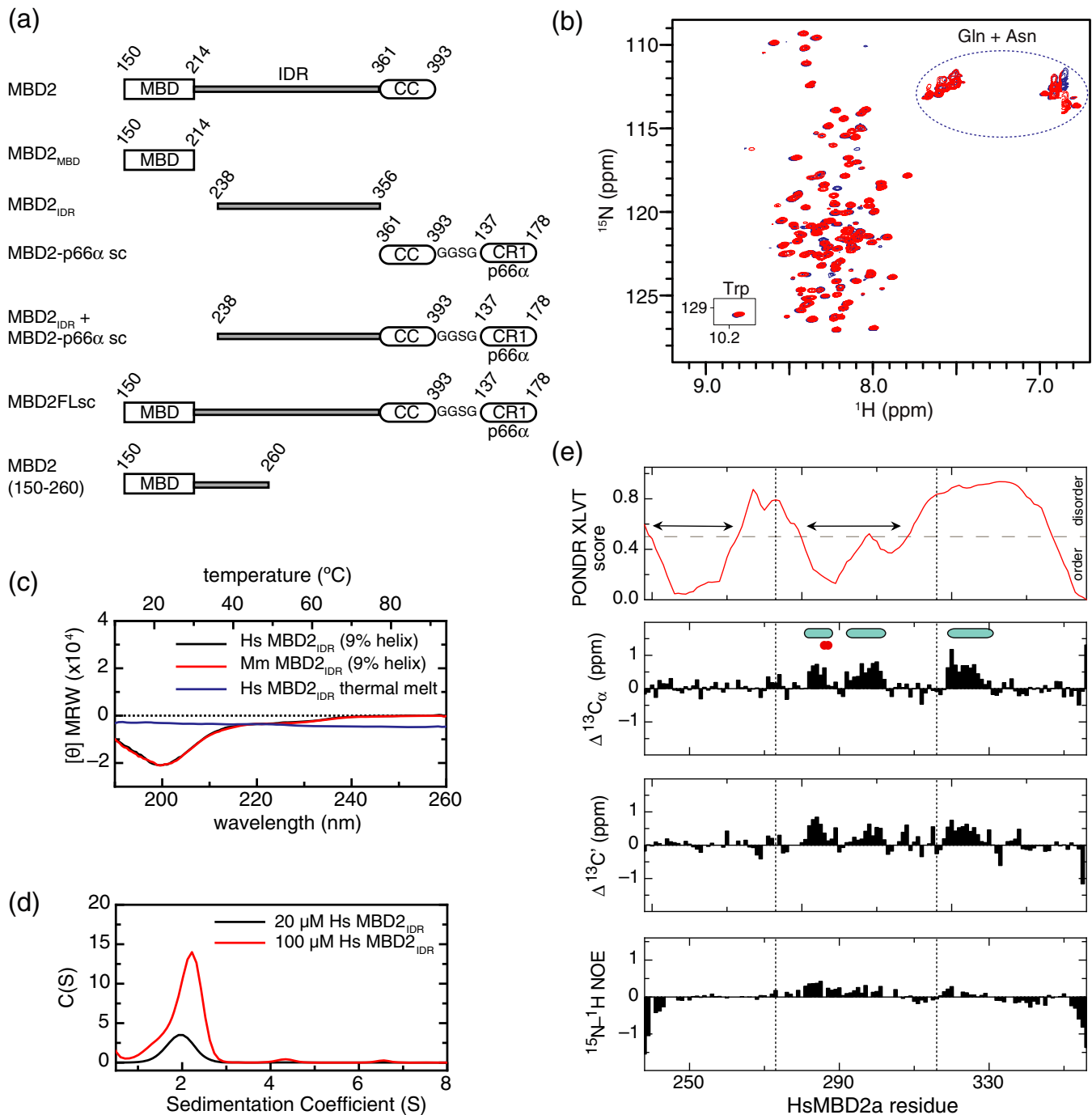
Figure 1. The MBD2$_{IDR}$ of MBD2 is intrinsically disordered in isolation. (**a**) A schematic diagram shows the domain organization of the different constructs used in these studies including the methyl-cytosine binding domain (MBD), intrinsically disordered region (IDR), and coiled-coil (CC) domains as well as a single chain (sc) construct that includes a short GGSG linker and the coiled-coil domain (CR1) of p66α. (**b**) An overlay of ^{15}N-HSQC spectra for the MBD2$_{IDR}$ of *Hs*MBD2 (red) and *Mm*MBD2 (blue) shows a lack of chemical shift dispersion with highly degenerate Gln and Asn ^{15}N-^{1}H$_2$ (blue circle) and Trp ^{15}N-^{1}Hε (inset) resonances. (**c**) CD spectra of the MBD2$_{IDR}$ at 25°C for *Hs*MBD2 (black) and *Mm*MBD2 (red) show little evidence of secondary structure formation while a thermal melt for *Hs*MBD2 MBD2$_{IDR}$ (blue) fails to reveal a cooperative transition characteristic of folded domains. (**d**) Fitting analytical ultracentrifugation sedimentation velocity data to a continuous size distribution (SEDFIT software (35)) indicates that the *Hs*MBD2$_{IDR}$ is largely monomeric at 20 μM (black) but does show evidence of slight oligomerization and aggregation at 100 μM (red) concentration. (**e**) PONDR® VLXT analysis of the MBD2$_{IDR}$ is plotted in the upper panel which shows two sub-regions of low disorder propensity (indicated by horizontal double-headed arrows, amino acids 240–262 and 280–308) within the IDR. Secondary chemical shifts were calculated for the MBD2$_{IDR}$ as described by Kjaergaard and Poulsen (41) and plotted for alpha (^{13}C$_α$) and carbonyl (^{13}C') resonances in the second and third panels. Regions with positive secondary chemical shifts consistent with α-helix formation are indicated with light blue ovals. Large red dots indicate the location of mutations that disrupt binding to the NuRD HDCC. Heteronuclear ^{15}N-^{1}H NOE for the MBD2$_{IDR}$ are plotted in the lower panel showing that most of the domain is highly dynamic with values < 0.5. The vertical dotted lines in all four panels indicate the three different sub-regions used for mapping the interaction with the NuRD HDCC. The second sub-region contains a stretch of positive heteronuclear NOE values consistent with a propensity to form α-helices.

MgCl$_2$, 0.1 mM EDTA, 1 mM DTT. SPR analysis of binding was performed on a Biacore T100 (GE Healthcare) as described previously (10,33). The purified DNA was bound to a Sensor SA chip (10 ng/ul DNA, 10 ul/min flow rate, 100 s) until a final relative response of approximately 100 U. Kinetic binding analyses were carried out at a flow rate of 30 ul/min (10mM HEPES, 50mM NaCl, 3mM MgCl$_2$, 0.1mM EDTA, 1mM DTT, 0.05% polysorbate 20, pH 7.4) and the data were fit using the manufacturer's software. As described previously, the rapid on and off-rates for the isolated MBD2$_{MBD}$ precluded analysis of binding kinetics. Therefore, the binding constant for MBD2$_{MBD}$ was determined by steady-analysis (10,33).

Fluorescence polarization (FP) analysis

3′ FAM labeled reverse strand oligonucleotide was purchased from IDT, annealed to the forward strand and purified as described above. Protein and DNA samples were buffer exchanged into 10 mM HEPES pH 7.5, 3 mM MgCl2, 0.1 mM EDTA, 1 mM DTT, and 0.02% sodium azide with either 50 mM NaCl (low salt buffer) or 150 mM NaCl (high salt buffer). Serial dilutions of protein were added to 10 mM DNA and fluorescence polarization measured on a PHERAstar FS microplate reader (BMG Labtech). The dissociation constant (K_D) was determined for each by fitting the observed polarization ([mP]) to a general equation for two state binding:

$$[mp] = [mP]_{DNA} + \frac{\left([D] + K_D + [P] - \sqrt{([D] + K_D + [P])^2 - 4*[P]*[D]}\right)}{2*[D]} * ([mP]_{sat} - [mP]_{DNA}),$$

in which [D] and [P] are the DNA and protein concentration and [mP]$_{DNA}$ and [mP]$_{sat}$ are the polarization for free and protein saturated DNA, respectively. The data were fit and plotted using Pro Fit software (QuantumSoft).

Analytical ultracentrifugation

Protein was buffer exchanged into 20mM Tris pH 8.0, 150 mM NaCl and sedimentation velocity analyzed at 40 000 rpm, 20°C on a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with a four and eight-position AN-60Ti rotor. Sedimentation profiles were recorded for 20 µM and 100 µM samples using UV absorption (280 nm) and interference scanning optics. The sample partial specific volume, buffer density and viscosity were calculated with the SEDNTERP (34) software and the effective molecular weight determined by fitting the data to a continuous size distribution with the SEDFIT (35) software.

Circular dichroism

The MBD2$_{IDR}$ was buffer exchanged into 10 mM NaPO$_4$, pH 6.5 at a final concentration of ∼100 µg/ml protein. CD spectra were collected from 185 to 260 nm (0.5 nm interval, 24 nm/min, 0.1 cm path length, 25°C) on a Chirascan^{TM}-plus CD spectrometer (Applied Photophysics). CD spectra were normalized to provide mean residue molar ellipticity ([θ] MRW) in degrees cm^{-2} dmol^{-1} residue^{-1}. Helical content for each peptide was calculated as described previously (36,37). The temperature dependence of circular dichroism (thermal melt) was followed by measuring ellipticity at 222 nm at 1°C intervals over a temperature range of 10–94°C with a heating rate of 1°C /min.

Nuclear magnetic resonance

Uniform ^{13}C, ^{15}N labeled protein was buffered exchanged into 10 mM NaPO$_4$, pH 6.5, 0.02% sodium azide, 1mM dithiothreitol, and 10% ^{2}H$_2$O and concentrated to 0.5–1 mM. NMR spectra were collected on a Bruker Avance III 700 MHz instrument at 25°C, and data were processed and analyzed with NMRPipe (38) and CcpNmr (39), respectively. Standard double and triple resonance experiments (^{15}N-HSQC, HNCO, HN-CACB, CBCA(CO)NH, HBHA(CO)NH, HN(CA)NNH, HNCACO, CCH-TOCSY, (H)CC(CO)NH, ^{15}N-NOESY-HSQC) were collected and the backbone and sidechain H$_{\alpha/\beta}$ resonances assigned for mouse MBD2$_{IDR}$. The ^{1}H chemical shifts were referenced to a separate sample of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) prepared in NMR buffer, while ^{13}C and ^{15}N chemical shifts were indirectly referenced using the IUPAC-IUB recommended ratios (40). Given the sequence identity and similar ^{15}N-HSQC spectra, only HNCO, HNCACB, CBCA(CO)NH, HBHA(CO)NH and ^{15}N-NOESY_HSQC triple resonance spectra were necessary to propagate backbone and sidechain H$_{\alpha/\beta}$ resonance assignments to human MBD2$_{IDR}$. For secondary chemical shift analyses, random coil chemical shifts for the MBD2$_{IDR}$ were calculated using the webserver at http://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/ (41).

^{1}H/^{15}N heteronuclear steady-state NOE spectra for human MBD2$_{IDR}$ were collected with and without 3 s of initial proton saturation in an interleaved fashion with spectral widths of 9800 Hz over 626 complex points in ω2 (^{1}H) and 1560 Hz over 256 complex points in ω1 (^{15}N).

^{15}N-TROSY-HSQC, TROSY-HNCO and ^{15}N-NOESY-HSQC spectra were collected for a ∼1mM sample of uniform ^{2}H, ^{13}C, ^{15}N-MBD2FLsc bound to methylated DNA. Once corrected for the spin state selective J-coupling offset, the HNCO spectra were very similar between MBD2$_{IDR}$ and MBD2FLsc such that backbone amide resonances could be assigned for most of the IDR region of MBD2FLsc (88 out of 110 amide resonances). Based on these assignments, the weighted chemical shift distance in ppm was calculated as $\Delta = \sqrt{(0.1 * \Delta\delta_N)^2 + \Delta\delta_{H_N}^2}$.

Cell culture

Human embryonic kidney 293T cells were maintained in Dulbecco's modified Eagle's Medium (DMEM) containing 10% heat inactivated fetal bovine serum (Hyclone), 2mM L-glutamine and 100 U/ml penicillin and streptomycin. MDA-MB-435 breast cancer cells were maintained in DMEM supplemented with 10% heat inactivated fetal bovine serum and 100 U/ml penicillin and streptomycin. Cells were cultured at 37°C and 5% CO2.

Co-immunoprecipitation

Various MBD2 constructs were cloned into the pCMV-Tag2B (Stratagene) vector in frame with an N-terminal flag-tag sequence. HEK 293T cells were transfected with the resulting vectors (18 μg plasmid DNA) by calcium phosphate precipitation method (42) and harvested after 48 h. Cells were lysed and immunoprecipitated with anti-flag M2 antibody (Sigma) and mouse IgG (Santa Cruz) controls according to the Sigma Flag-IPT kit protocol (Sigma-Aldrich, Inc., St. Louis, MO, USA). The precipitated proteins were then analyzed for different components of the MBD2-NuRD complex by western blot using antibodies against RbAp48 (Abcam ab79416), HDAC2 (Millipore #05–814) and MTA2 (Santa Cruz sc-28731).

For full-length MBD2 pull downs, the cells were lysed in micrococcal nuclease (MNase) digestion buffer (25mM HEPES-KOH pH 7.6, 100mM NaCl, 5mM MgCl2, 3mM CaCl2, 10% glycerol, 0.2% NP40 and 1X EDTA-free protease inhibitor cocktail (Roche)) followed by MNase digestion using 1500 U/ml of MNase (Worthington Biochemical, Lakewood, NJ, USA) for 2 h on ice. Ethidium bromide was then added to the lysate at 300 μg/ml followed by a spin at 10 000xg for 15 mins at 4°C. The supernatant was then subjected to immunoprecipitation as previously described.

Site-directed mutagenesis

Mutant oligonucleotides targeting the conserved residues of the 1st and 2nd region of MBD2$_{IDR}$ were designed using the QuikChange Primer Design Program (www.agilent.com/genomics/qcpd). Primers used for mutagenesis are listed in Supplementary Table S1. Mutagenesis was carried out per manufacturer's protocol using the Quikchange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, Inc., Santa Clara, CA, USA). Clones were verified by sequencing and used for transfections.

Sequence alignments

Protein-coding sequences of MBD2 for different species were obtained from the NCBI Protein database and aligned using the PRALINE software (http://www.ibi.vu.nl/programs/pralinewww/).

Local isoelectric point calculation

Each residue (*i*) in the protein was assigned a local isoelectric point (pI) calculated from the amino acid sequence (43) for residues spanning $i \pm 7$. The pI was calculated using the SeqUtils.IsoelectricPoint module in the Bio package from the Biopython Project (www.biopython.org). For those residues less than seven from the N- or C-termini, the isoelectric point was calculated from the amino acid sequence for residues spanning $i \pm n$ where n is the number of residues between *i* and the N- or C- terminus.

Lentiviral knockdown and expression of MBD2

The shMBD2 (5′ - GGGTAAACCAGACTTGAA - 3′) sequence was cloned into the pRRL.H1.shRNA vector. Full-length wild-type and mutant MBD2 gene sequences were

cloned in the pLV203 vector with a flag-tag added to the C-terminus (Genecopoeia, Rockville, MD, USA). Both full-length MBD2 genes contain a single silent mutation introduced into the MBD2 shRNA target sequence to confer shRNA resistance. The vectors were packaged into a lentivirus by calcium phosphate transfections of HEK 293T cells. MDA-MB-435 breast cancer cells were transduced with packaged virus and grown *in vitro* for 7 days post-transduction before quantitative polymerase chain reaction and western blot analyses. Protein concentrations of cell lysates were measured using the BioRad *DC*^{TM} Protein Assay. The MBD2 protein was quantified on western blots with the *LI-COR* Odyssey® Infrared Imaging System using the Image Studio Software based on Chemi-luminescence imaging. Primers used for qPCR analysis are listed in Supplementary Table S2. Antibodies used for MBD2 and Flag epitope protein detection were from Santa Cruz (sc1244) and Abcam (ab1162), respectively. All lentiviral vector infection experiments were repeated independently at least three times and a paired Student's *t*-test used to compare repression of gene expression.

RESULTS

The MBD2$_{IDR}$ region is intrinsically disordered

We cloned, expressed and purified from bacteria the MBD2$_{IDR}$ from both the human (amino acids 238–356) and mouse (amino acids 241–359) MBD2 proteins (Figure 1a). The 2D ^{15}N-HSQC spectra of each (Figure 1b) show features characteristic of a disordered peptide. The amide backbone resonances are quite sharp, the ^{1}H(N) chemical shifts fall between 7.5 and 8.5 ppm, and the sidechain amide resonances are highly degenerate. In particular, the ^{1}H-^{15}N$_\epsilon$ resonances of two Trp residues completely overlap (inset, Figure 1b), while seven Asn and ten Gln sidechain ^{15}N-^{1}H$_2$ resonances overlap extensively (blue dashed oval, Figure 1b). These findings indicate that the MBD2$_{IDR}$ domain remains largely disordered in solution.

We further characterized this region by CD and analytical ultracentrifugation. The CD spectra are consistent with a highly disordered peptide comprised of only 9% helix at 25°C (Figure 1c) based on mean residue molar ellipticity ([θ] MRW) at 222 nm. The temperature dependence for CD at 222 nm did not show a cooperative transition typical of a folded domain (blue curve, Figure 1c). Sedimentation velocity by analytical ultracentrifugation analysis is consistent with a monomeric protein at 20 μM with a tendency to aggregate at higher concentrations (Figure 1d).

Although the backbone resonances of the MBD2$_{IDR}$ lack significant chemical shift dispersion, we were able to assign these resonances using standard 3D heteronuclear techniques. The mouse domain was particularly well behaved, such that we first assigned mouse MBD2$_{IDR}$ and then carried those assignments over to the human protein. Based on these assignments, secondary chemical shifts were calculated for the MBD2$_{IDR}$ (41) and plotted for alpha (^{13}C$_\alpha$) and carbonyl (^{13}C') carbons in Figure 1e. Three separate segments of the MBD2$_{IDR}$ have positive secondary chemical shifts indicative of a propensity to form α-helices (light blue ovals, Figure 1e). For comparison, the disorder propensity for the MBD2$_{IDR}$ as calculated by PONDR®

VLXT (Predictor of Naturally Disordered Regions) (44,45) is plotted in the first panel of Figure 1e. This latter analysis predicts that two sub-regions within MBD2$_{IDR}$ (amino acids 212–273 and 274–316) have low disorder propensity (indicated by double-headed arrows in Figure 1e). The first two of the potential helical segments fall within the second region of low disorder propensity, supporting the idea that this region can adopt transient helical secondary structure in isolation. Heteronuclear NOE values (Figure 1e, lower panel) remain below 0.5 throughout the MBD2$_{IDR}$ indicating a high degree of internal dynamic motion; yet, the same region with low disorder propensity and large secondary chemical shifts (amino acids 212–316) also show the largest heteronuclear NOE values. Therefore, these data demonstrate that while the MBD2$_{IDR}$ remains largely unstructured in isolation, the central portion has a tendency to form α-helices that may fold into a more stable structure either in the context of full-length protein or upon binding other components of NuRD.

To test whether the MBD2$_{IDR}$ depends on surrounding regions of the protein to adopt a regular structure, we sought to analyze this region in the context of the surrounding MBD and coiled-coil domains. However, a polypeptide including the MBD, MBD2$_{IDR}$ and coiled-coil domains of MBD2 has proven difficult to purify from bacterial expression systems. We recently designed and tested a remarkably stable and well-behaved single chain construct of the coiled-coil domains from MBD2 and p66α (Supplementary Figure S1a). This single-chain construct (scMBD2-p66α) comprises the MBD2 coiled-coil domain (amino acids 361–393), a short GGSG linker, and the p66α coiled-coil domain (amino acids 137–178). The scMBD2-p66α retains helical content as measured by CD analysis up to 95°C (Supplementary Figure S1b) and the 2D ^{15}N-HSQC spectrum remains dispersed even at 80°C (Supplementary Figure S1c). Adding the scMBD2-p66α to the MBD and MBD2$_{IDR}$ of MBD2 (MBD2FLsc, amino acids 212–356) stabilizes the protein for bacterial expression and purification.

A 2D ^{15}N-HSQC spectrum of the MBD2FLsc protein bound to methylated DNA contains a few broadened and dispersed as well as many sharp and highly degenerate resonances (Figure 2a). An overlay of the 2D ^{15}N-HSQC spectra (Figure 2b) for MBD2FLsc bound to methylated DNA (red) and MBD2$_{IDR}$ (blue), plotted at a higher contour level, shows that the sharp and degenerate resonances from MBD2FLsc are similar to those of the MBD2$_{IDR}$. Given the similarity in chemical shifts and the sharpness of these peaks, we were able to carry over assignments for 80% of the amide resonances of the IDR in the MBD2FLsc:DNA complex. Plotting ^{1}H,^{15}N chemical shift differences between resonances from the isolated MBD2$_{IDR}$ and the same region in the MBD2FLsc:DNA (Figure 2d) confirm that the chemical shifts are very similar throughout the region (Δ < 0.05ppm).

Comparing the broadened and more dispersed resonances (Figure 2c) with those for the isolated MBD2$_{MBD}$ bound to methylated DNA (teal) and coiled-coil domains (purple) (18) shows that these peaks are attributable to the MBD and coiled-coil domains. In previous studies we found that the chemical shifts for Gly^{169} and Ala^{172} are quite different when bound to either unmethylated or methylated

DNA: Gly^{169} shifts from 107 ppm to 102 ppm in ^{15}N while the Ala^{172} shifts from 7.0 ppm to 7.3 ppm in ^{1}H$_N$, respectively (10). As can be seen in Figure 2c, Gly^{169} and Ala^{172} show very similar chemical shifts for both MBD2FLsc and MBD2$_{MBD}$ indicating a similar methylation specific binding mode. Together these findings show that the MBD2$_{IDR}$ region remains largely disordered even in the context of full-length protein and does not appear to affect the structure or DNA binding mode of the MBD.

The MBD2$_{IDR}$ modifies DNA binding kinetics and overall binding affinity

To determine if the MBD2$_{IDR}$ alters the affinity of MBD2 for DNA, we measured binding to methylated DNA by surface plasmon resonance. As described previously (33), the isolated MBD shows rapid on and off-rates, which requires steady state analysis to determine the overall binding affinity (K_D = 330 nM, Figure 3a). In contrast, the MBD2FLsc shows an ∼100-fold increase in affinity (K_D = 2 nM) as compared to the MBD2$_{MBD}$. Inspection of the MBD2$_{IDR}$ reveals that the region just C-terminal to the MBD contains positively charged Arg and Lys residues, which could contribute to DNA binding through non-specific electrostatic interactions. Therefore, we calculated a theoretical isoelectric point for a sliding window of 15 amino acids along the entire length of each protein. The results of this analysis for human (*Hs*) MBD2 paralogs (MBD2, MBD3, MBD3-L1 and MBD3-L2) as well as *Bombyx mori (Bm)* and *Amphimedon queenslandica (Aq)* MBD2 orthologs are plotted for the region homologous to the MBD2$_{IDR}$ of each protein in Figure 3b. Each of the MBD2 orthologs contains a positively charged region just following the MBD and at the N-terminus of the MBD2$_{IDR}$ (Figure 3b). To test whether this region is sufficient to modify DNA binding, we measured binding kinetics by surface plasmon resonance for an MBD2 construct that includes the MBD and the first 20 amino acids of the MBD2$_{IDR}$ (residues 150–260). This construct incorporates amino acids from the N-terminal portion of the IDR (red bar, Figure 3b) that show evolutionary conservation of a net positive charge. Similar to MBD2FLsc, MBD2(150–260) binds with increased affinity (K_D = 20 nM) compared to the isolated MBD (Figure 3a).

To validate these findings, we measured DNA binding affinity using a fluorescence polarization based assay (Figure 3c). Similar to the results from SPR studies, MBD2FLsc and MBD (150–260) bind methylated DNA with ∼100-fold greater affinity than MBD2$_{MBD}$. In addition, the binding affinity for all three constructs decreases by ∼10-fold with an increase in NaCl concentration from 50 mM to 150 mM. The latter observation is consistent with non-specific charge interactions contributing to the overall binding affinity. Therefore, the MBD2$_{IDR}$ contains a positively charged region that modifies binding kinetics and affinity despite a lack of regular structure formation. Similar positively charged disordered regions adjacent to DNA binding domains have been described previously and referred to as 'fuzzy' domains (46–48).
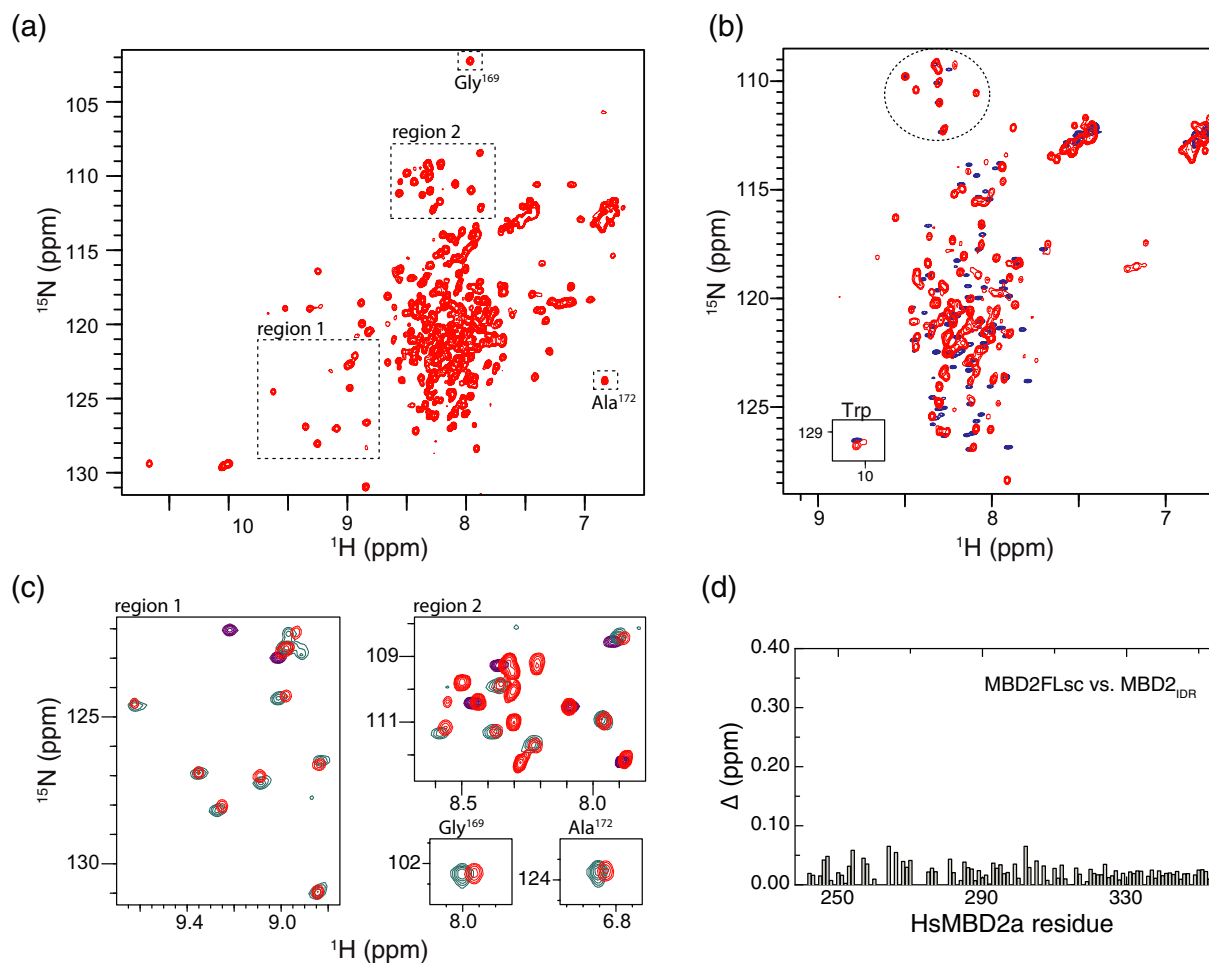
Figure 2. NMR analysis of intrinsically disordered MBD2$_{IDR}$. (**a**) A 2D ^{15}N-HSQC spectra of MBD2FLsc (red) bound to methylated DNA shows both sharp and highly degenerate as well as a few broadened and dispersed resonances. (**b**) An overlay of 2D ^{15}N-HSQC spectra for MBD2FLsc bound to methylated DNA (red) and the MBD2$_{IDR}$ are plotted at higher contours which shows that the sharp and degenerate resonances from the MBD2FLsc are similar to those from the MBD2$_{IDR}$. This similarity is most easily appreciated when comparing the more isolated peaks from glycine residues (highlighted with a dashed circle). (**c**) Expanded regions from the 2D ^{15}N-HSQC show that the broadened and dispersed resonances from MBD2FLsc (red) align with resonances from either the MBD2$_{MBD}$ bound to methylated DNA (teal) or the MBD2-p66α coiled-coil complex (purple). Likewise, two key resonances from the MBD2$_{MBD}$ that show unique chemical shifts when bound to methylated DNA (Gly169 and ALa172) show similar chemical shifts in the context of MBD2FLsc. (**d**) Plotting the chemical shift distance (Δ) between resonances from the isolated MBD2$_{IDR}$ and the same region of MBD2FLsc when bound to methylated DNA shows only small chemical shift differences (< 0.05 ppm) throughout the region.

The MBD2$_{IDR}$ binds the histone deacetylase core complex of NuRD

In previous studies we showed that the coiled-coil domain of MBD2 recruits the p66α and CHD4 proteins to NuRD. To investigate whether the MBD2$_{IDR}$ and coiled-coil domains interact with the remaining core components of NuRD (MTA2, HDAC2 and RbAp48), the MBD2$_{IDR}$ was expressed in 293T cells as a flag-tagged construct, with and without the scMBD2-p66α coiled-coil domains (Figure 4a and d, respectively). Cell lysates were immunoprecipitated with an anti-flag antibody followed by western blot analysis to identify the NuRD components interacting with the flag-tagged MBD2$_{IDR}$. Figure 4a shows that the MBD2$_{IDR}$ of MBD2 strongly interacts, either directly or indirectly, with the RbAp48, HDAC2 and MTA2 components of NuRD while the scMBD2-p66α does not. The MBD2$_{IDR}$ does not interact with p66α and CHD4 proteins (Supplementary Figure S2) as expected, which uniquely maps the interaction of MBD2 with the histone deacetylase and chromatin remodeling components of the NuRD complex into two distinct functional domains, the MBD2$_{IDR}$ and the MBD2 coiled-coil domain, respectively.

Segments of intrinsically disordered proteins can fold upon interacting with target proteins to adopt regular secondary structure. These segments, referred to as molecular recognition features (MoRFs)(48–50), may be identified as regions with low disorder propensity by disorder prediction algorithms. To identify the minimal region necessary for binding to the histone deacetylase core complex of NuRD, the MBD2$_{IDR}$ was divided into three separate sub-regions (amino acids 212–273, 274–316 and 317–360, Figure 1e) in order to isolate the central portion with helical and structural propensity from the N and C-terminal regions. We tested whether each of these regions, separately and in combination, could immunoprecipitate HDCC components. As
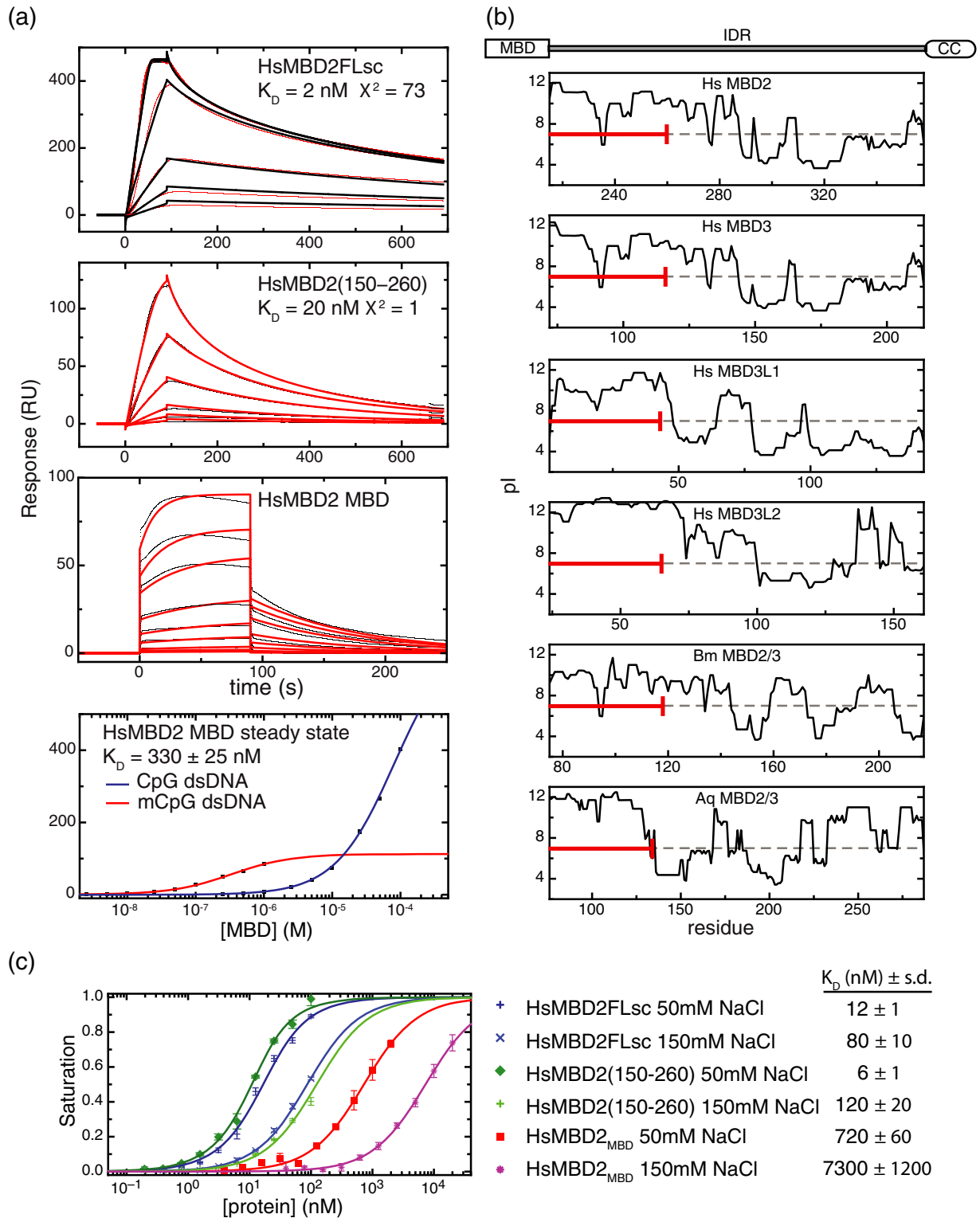
Figure 3. The MBD2$_{IDR}$ modifies DNA binding kinetics and affinity. (**a**) Surface plasmon analysis of DNA binding shows that both MBD2FLsc (upper panel), and MBD2(150–260) (second panel) bind with higher affinity than the isolated MBD (third panel). Steady-state analysis of the isolated MBD binding to methylated (red) and unmethylated (blue) DNA is shown in the fourth panel. (**b**) The local isoelectric point, calculated as described in the text, is plotted for the region homologous to the MBD2$_{IDR}$ in *Hs*MBD2, *Hs*MBD3, *Hs*MBD3-L1, *Hs*MBD3-L2, *Bm*MBD2/3 and *Aq*MBD2/3. A red bar indicates the region from the N-terminus of the MBD2$_{IDR}$ included in the MBD2(150–260) construct, which shows that the positive charge is conserved across the animal kingdom. (**c**) To validate the SPR studies, binding affinity (K_D) of MBD2FLsc, MBD2(150–260) and MBD2$_{MBD}$ for methylated DNA was determined by a fluorescence polarization assay. Fitting the resulting binding curves to a general two-state binding model shows that both MBD2FLsc and MBD2(150–260) bind with ~100-fold greater affinity than the isolated MBD2$_{MBD}$. Furthermore, the binding affinity of each shows a similar dependence on salt concentration such that increasing NaCl from 50 to 150 mM reduces the affinity by ~10-fold.
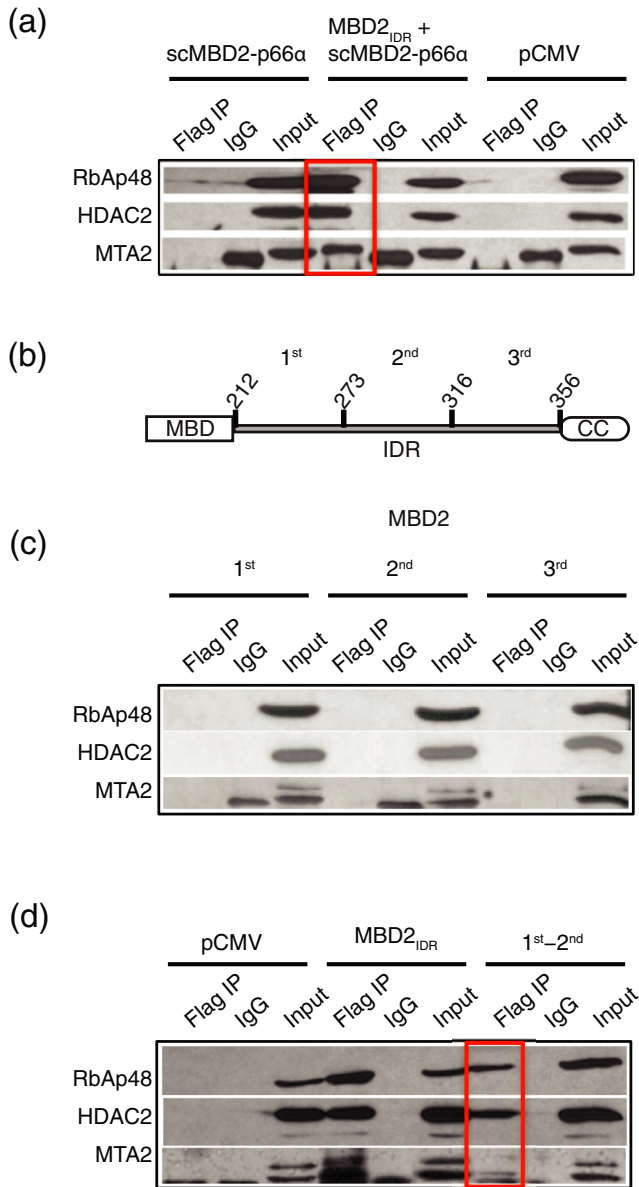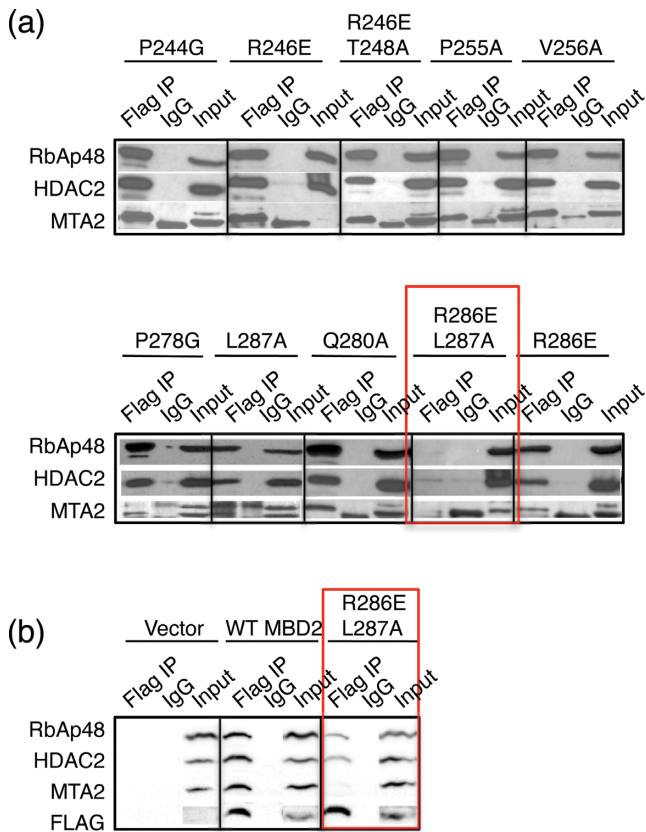
Figure 4. The MBD2$_{IDR}$ binds the histone deacetylase core complex of NuRD. (**a**) The MBD2$_{IDR}$ with scMBD2-p66α was transfected in high-transfection-efficiency HEK 293T cells. Immunoprecipitation and western blot analysis of the transfected cells indicates that the MBD2$_{IDR}$ binds the histone deacetylase core components RbAp48, HDAC2 and MTA2, whereas the scMBD2-p66α construct does not. The flag-IP lane shows pull down of the histone deacetylase core components by immunoprecipitation using an anti-flag antibody directed against the flag-tagged MBD2$_{IDR}$. IgG and expression vector pCMV serve as negative controls while the input lane shows 2% of the input. (**b**) The MBD2$_{IDR}$ was divided into three sub-regions to test whether we could isolate critical binding region(s) for the core complex components. (**c**) The three sub-regions of MBD2$_{IDR}$ were expressed individually in HEK 293T cells and immunoprecipitated, but failed to bind to either of the histone deacetylase core complex components. Note that a non-specific band appears in the IgG lane when blotted with anti-MTA2 that runs just below the MTA2 protein in the input and flag IP lanes. (**d**) The region of MBD2$_{IDR}$ from amino acids 212–316 comprising the first and second ordered sub-regions in combination can bind to RbAp48, HDAC2 and MTA2, although a weaker interaction with MTA2 was observed.

can be seen in Figure 4c, none of these three sub-regions interact with the core complex in isolation. However, the first and second sub-regions of MBD2$_{IDR}$ in combination (amino acids 212–316) were sufficient to bind RbAp48, HDAC2 and MTA2, albeit with a somewhat weaker interaction with MTA2 as compared to the entire MBD2$_{IDR}$ (Figure 4d).

Key residues within minimal MBD2$_{IDR}$ are evolutionarily conserved: full-length protein sequences of HsMBD2 orthologs from across the animal kingdom (Aq, Trichoplax adhaerans (Ta), Bm, Danio rerio (Dr), Gallus gallus (Gg), Mus musculus (Mm)), and Hs paralogs MBD3, MBD3-like 1 and 2 (MBD3-L1, MBD3-L2) were obtained from the NCBI protein database. The protein sequences of individual domains of MBD2 were aligned using the default settings of the PRALINE online server. An alignment of the MBD (Supplementary Figure S3a) for orthologs (Aq, Bm, Dr, Gg and Mm) and a paralog (MBD3) that contain a recognizable MBD shows a high degree of conservation from sponge to human (AqMBD2 and HsMBD2 proteins share 64% identity in their MBDs). Critical residues known to influence DNA binding affinity and methylation selectivity (Tyr^{178} and Lys^{174} of HsMBD2) [10,12,13] are highly conserved indicating that methylation specificity likely developed with the first multi-cellular organisms. Notably, several homologs (Ta MBD2/3, Hs MBD3-L1 and L2) lack an MBD, yet previous studies have shown that both the MBD3-L1 and L2 paralogs interact with NuRD and compete with MBD2 and MBD3 for complex formation (51–53). Therefore, NuRD complex recruitment and DNA binding represent distinct and separable functions of MBD2.

An alignment of the MBD2$_{IDR}$ and coiled-coil regions (Supplementary Figure S3b), shows much less conservation with large insertions (i.e. TaMBD2/3 and AqMBD2/3) and smaller deletions as compared to HsMBD2. Nonetheless, specific individual residues are absolutely conserved across species suggesting that these amino acids are critical to function. Highly conserved residues were identified within the first two ordered regions of MBD2$_{IDR}$ implicated in binding to the NuRD histone deacetylase core complex. We mutated these conserved residues individually or in pairs as follows: (1) P244G, (2) R246E, (3) T248A, (4) P255A, (5) V256A, (6) P278G, (7) Q280A, (8) W283A, (9) R286E, (10) L287A, (11) L290A, (12) R246E/T248A and (13) R286E/L287A (see red stars above the sequence alignment in Supplementary Figure S3b). Flag-tagged MBD2$_{IDR}$ mutants were expressed in 293T cells followed by immunoprecipitation and blotting for components of the histone deacetylase core complex of NuRD. Of the eleven conserved residues tested, only two contiguous amino acids from the second ordered region of MBD2$_{IDR}$, Arg^{286} and Leu^{287}, when mutated, markedly decreased binding to the core complex (Figure 5a). Interestingly neither of these mutations in isolation nor any of the other mutations disrupted binding to the histone deacetylase core components. Next, we introduced the double mutations, R286E/L287A, into the full-length human MBD2 gene sequence by site-directed mutagenesis. Since MBD2-NuRD contains several potential DNA binding proteins, the immunoprecipitation protocol for full length MBD2Flag constructs was modi-

Figure 5. Two conserved residues from minimal MBD2$_{IDR}$ are necessary to bind the histone deacetylase core of NuRD. (**a**) Mutating five residues from the first region of minimal MBD2$_{IDR}$ (P244G, R246E, P255A, V256A and R246E/T248A) does not affect immunoprecipitation of the histone deacetylase core complex. Similarly, mutating four residues from the second ordered region of minimal MBD2$_{IDR}$ (P278G, Q280A, R286E and L287A) does not affect immunoprecipitation of the histone deacetylase core complex. However, combined mutation of two adjacent residues (R286E/L287A) significantly abrogates the ability of MBD2$_{IDR}$ to recruit the histone deacetylase core components. Strikingly, the individual mutants R286E and L287A have no effect on interaction of the MBD2$_{IDR}$ with the histone deacetylase core. Three additional mutations (T248A, W283A and L290A) do not affect binding of MBD2$_{IDR}$ to the histone deacetylase core complex (data not shown). (**b**) Full length flag-tagged MBD2 carrying the double mutation R286E/L287A (Double Mutant) also displays a disrupted interaction with RbAp48, HDAC2 and MTA2. The interaction of Double Mutant MBD2 with MTA2 is completely lost whereas its interaction with RbAp/HDAC shows significant reduction.

fied to eliminate contamination through non-specific DNA binding. The cell lysate from transfected cells was treated with micrococcal nuclease, which cleaves both single- and double-stranded DNA and RNA, and with ethidium bromide, which intercalates between stacked bases in the double helix. As expected, immunoprecipitation of the NuRD complex by the mutant full length MBD2 revealed impaired interaction with the histone deacetylase core components: RbAp48, HDAC2 and MTA2, with the effect being most pronounced for MTA2 (Figure 5b). Together these observations indicate that Arg[286] and Leu[287] comprise a critical interaction surface necessary but not sufficient for recruiting the histone deacetylase core complex of NuRD.

To determine if interaction with the histone deacetylase core components is necessary for methylation dependent

gene silencing by MBD2, we designed a repression assay to study the functional effects of the R286E/L287A (Double Mutant) MBD2 on expression of an endogenous gene. The Double Mutant MBD2Flag construct was expressed in a background of partial knockdown of endogenous MBD2 in MDA-MB-435 breast cancer cells. The percentage of repression of the endogenous tumor suppressor gene *PRSS8* was used to measure MBD2-NuRD mediated transcriptional repression. We previously showed that 75% stable knockdown of MBD2 by shRNA in MDA-MB-435 cells increased *PRSS8* expression, demonstrating that MBD2 mediates endogenous transcriptional repression of this gene in these cells ([20]).

Concomitant knockdown of endogenous MBD2 and enforced expression of wild-type or Double Mutant MBD2Flag was achieved by simultaneous lentivirus mediated delivery of anti-MBD2 shRNA (SH) and wild-type (WT) or Double Mutant MBD2Flag mRNA resistant to knockdown ([20]). Western blotting with anti-MBD2 antibody (Figure 6a), which recognizes both endogenous (lower band) and lentivirally expressed MBD2Flag (top band), shows about 50% knockdown of the endogenous MBD2 protein in SH+WT and SH+Double Mutant cells as compared to negative control cells simultaneously infected with both a scramble short RNA sequence (SC) expression vector and an empty lentivirus (LV) expression vector. Despite the ability of this MBD2 shRNA construct to knock down MBD2 levels by greater than 75% (Supplementary Figure S4a) ([20]), we were unable to achieve the same level of knockdown when cells were concomitantly infected with lentiviral MBD2 expression constructs, regardless of whether cells were infected with both constructs simultaneously or sequentially (in either order). Detection of the exogenously expressed MBD2Flag proteins with an anti-flag antibody demonstrates similar expression levels in SH+WT and SH+Double Mutant transduced cells, as determined by quantitative chemiluminescence (Figure 6b). Taking into account both knockdown of endogenous MBD2 and expression of MBD2Flag proteins, the total amount of MBD2 protein in both groups (SH+WT and SH+Double Mutant) remains constant and represents more total MBD2 protein than in the negative control cells. Partial knockdown of endogenous MBD2 and enforced expression of wild type MBD2, results in approximately 70% more repression of the *PRSS8* gene than in negative control cells. However, enforced expression of the Double Mutant MBD2 results in a smaller increase in repression of the *PRSS8* gene as compared to the negative control cells (Figure 6c). Hence the amount of repression of the *PRSS8* gene is consistently ≥2-fold higher (*P*-value<0.01) in SH+WT cells than in SH+Double Mutant cells. We attribute this difference to the inability of Double Mutant MBD2 to effectively recruit the histone deacetylase core components of the NuRD complex. The suggestion of residual repressive activity by the Double Mutant MBD2 may reflect that it retains an intact coiled-coil domain. We have shown previously that the interaction of the coiled-coil domain with p66α and CHD4 mediates at least part of the repressive activity of MBD2-NuRD ([18]).
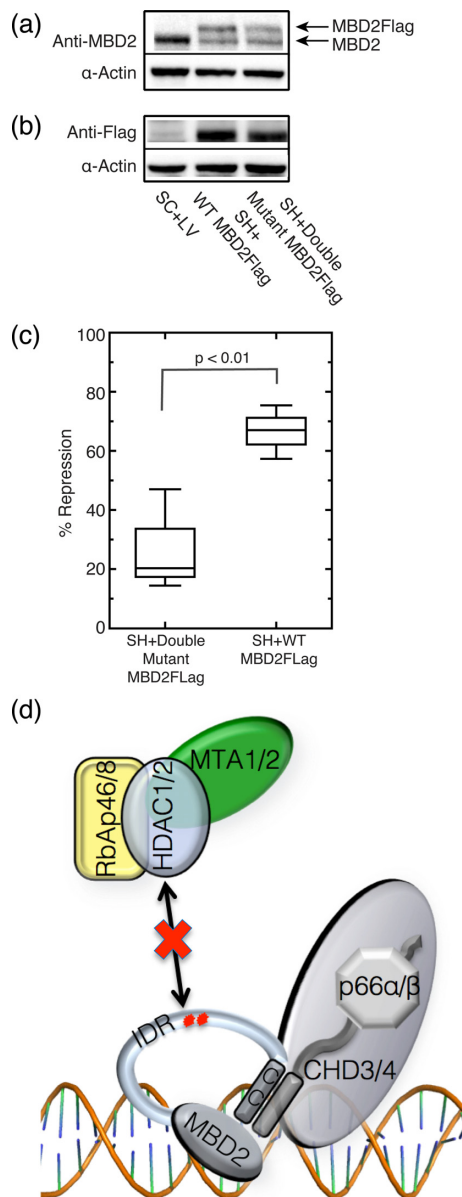
Figure 6. Double Mutant MBD2 shows reduced transcriptional repression of its methylated target gene *PRSS8*. Western blot quantitative chemiluminescence analysis of MDA-MB-435 cells infected with lentivirus for MBD2 knockdown (SH) and expression of wild-type (WT) and R286E/L287A (Double Mutant) MBD2 shows (**a**) ~50% and ~40% knockdown of endogenous full-length MBD2 in SH+WT and SH+Double Mutant cells, respectively, and (**b**) equivalent expression of WT and Double Mutant MBD2Flag in SH+WT and SH+Double Mutant cells, respectively. SH+Double Mutant cells contain ~60% mutant MBD2Flag as compared to SH+WT cells. (**c**) Transcript levels of *PRSS8* were measured by qPCR, normalized to endogenous transcript levels of *GAPDH*, and plotted as percent repression compared to SC+LV control cells. SH+WT cells show significantly increased repression of *PRSS8* as compared to SH+Double Mutant cells (p < 0.01). Box plots represent three independent experiments and a paired Student's *t*-test used to compare repression of *PRSS8* gene expression. (**d**) A working model of the architecture of the MBD2-NuRD co-repressor complex maps interactions between MBD2 and the histone deacetylase core subunits to the MBD2$_{IDR}$, wherein mutation of the two contiguous residues comprising the critical interaction site (shown as red asterisks in the IDR) impairs binding. The previously characterized (CC) coiled-coil domain of MBD2 recruits the chromatin remodeler CHD4 through p66α and the MBD recruits the co-repressor complex to sites of dense CpG methylation.

DISCUSSION

Research over the past two decades has challenged the traditional structure-function paradigm of proteins. Instead of adopting a well-folded three-dimensional structure in solution, many proteins are entirely disordered or contain functionally important intrinsically disordered regions (IDR). Estimates indicate that from 15% to 45% of eukaryotic proteins have long (>30 amino acid) regions of disorder (28). These regions can function either by folding upon binding to a target ligand or through dynamic interactions involving rapidly exchanging conformations of the disordered state (28,46,48).

Here we show that a large segment (~145 amino acids) between the conserved MBD and coiled-coil domains of MBD2 behaves as an intrinsically disordered region in isolation and in the context of the full-length protein. This region does not adopt a regular structure even when bound to methylated DNA. Nonetheless incorporating the MBD2$_{IDR}$ affects the binding kinetics and overall affinity for DNA, functioning as a 'fuzzy' domain (46). The N-terminal portion of the MBD2$_{IDR}$ has a net positive charge that is conserved across evolution and likely contributes to this fuzzy interaction with the DNA (see Figure 3b). IDRs flanking functional DNA-binding domains have been shown to influence DNA recognition and potentially rates of association and lifetime of the DNA-protein interactions (54,55).

IDRs are often found in proteins that serve as hubs of one-to-many protein interactions (56–58). In this report we demonstrate that the MBD2$_{IDR}$ of MBD2 binds the histone deacetylase core complex of NuRD comprised of the MTA2, HDAC2 and RbAp48 proteins. To map this interaction, we divided the MBD2$_{IDR}$ into three potential MoRFs based on NMR analyses of residual structure and the disorder propensity. We found that the first two segments of the MBD2$_{IDR}$ are necessary and sufficient to bind the histone deacetylase core complex (Figure 4d). Interestingly, this region incorporates both the positively charged segment that contributes to DNA binding as well as the region with helical propensity. We were unable to identify sub-regions of MBD2$_{IDR}$ that selectively bind to different components of the histone deacetylase core complex, which suggests cooperative all-or-none formation of this complex (Figure 4c). This observation indicates that either the MBD2$_{IDR}$ forms a large contact surface for binding each of the proteins cooperatively or that the MBD2$_{IDR}$ binds to one of the core components which in turn interacts with other components. In support of the latter interpretation, recent structural analyses have shown that the ELM2 and SANT domains of MTA1 bind directly to HDAC1 (59) and a small peptide from the C-terminal region of MTA1 binds RbAp48 (60). The MTA proteins may provide a common point of contact such that MBD2$_{IDR}$ binding directly to either of RbAp, HDAC, or MTA would be sufficient to recruit all of the histone deacetylase core components.

Intrinsic disorder in MBD2 appears to be under selective pressure as evidenced by conservation of numerous residues within the MBD2$_{IDR}$ across several paralogs and distantly related orthologs, despite large insertions and sequence variability (Supplementary Figure S3). Using this

evolutionary conservation as a guide, we designed point mutations in the first and second segments of the $MBD2_{IDR}$ and found that two highly conserved residues (Arg^{286} and Leu^{287}) located within the helical region of the $MBD2_{IDR}$ are critical for binding to the histone deacetylase core complex. Mutating these residues even in the context of full length MBD2 diminishes interaction with the histone deacetylase core complex. A recent analysis of MBD2 alternative splice variants revealed that dominant isoform in embryonic stem cells, MBD2c, lacks most of the C-terminus including the $MBD2_{IDR}$ and the coiled-coil domains. Consistent with the role of these two domains in recruiting CHD4 and the histone deacetylase core complex, MBD2c fails to interact with any of the NuRD components and lacks repressive activity ([61]).

MBD2-NuRD has been established as a transcriptional repressor of embryonic and fetal β-type globin genes in primary adult erythroid cells across different species ([4,5,17,18]) and of tumor suppressor genes in breast cancer cells ([20,26]). We hypothesized that the Arg^{286} and Leu^{287} mutations of MBD2 that impair recruitment of the histone deacetylase core components of the NuRD complex would in turn abrogate repression of methylated target genes. To test this hypothesis, we assayed the expression of a previously identified highly methylated MBD2 target gene, *PRSS8,* in MDA-MB-435 breast cancer cells ([20]). *PRSS8* encodes for prostasin, a member of the trypsin family of serine proteases that has been implicated in inhibition of metastasis of both breast and prostate cancer cells ([62,63]). A CpG-rich region in the promoter and exon 1 of *PRSS8* is heavily methylated when this gene is repressed in MDA-MB-435 breast cancer cells ([62]), and demethylation by 5-aza-2′-deoxycytidine combined with histone deacetylase inhibitor treatment reactivates expression of *PRSS8*. In our studies, Double Mutant MBD2 protein exerts diminished repression of *PRSS8* when compared to an equivalent amount of wild-type MBD2 protein, (Figure 6c) thus confirming the importance of these two amino acid residues in MBD2-NuRD function.

Our results can be summarized into a dynamic working model for the architecture of the NuRD complex wherein the subunits carrying the chromatin remodeling enzymatic activities unique to the NuRD complex can be mapped to two separate domains of MBD2: the $MBD2_{IDR}$ which recruits the histone deacetylase core components and the coiled-coil domain which recruits the chromatin remodeling subunit CHD4 through its interaction with p66α (Figure 6d). Delineating the multi-protein interactions within the MBD2-NuRD complex should enable future attempts to decouple the distinct enzymatic activities of the complex, in isolation or combination, to disrupt the repressive functions of each and dissect their potentially independent roles in regulating specific sets of genes.

Understanding the role of IDRs in the assembly of multi-protein complexes like MBD2-NuRD adds valuable insight to their function in epigenetic regulation. IDRs have been well recognized as hubs for interaction of many proteins due to their malleability and fluctuating structures. MBD1-c was recently shown to selectively interact with different binding partners through its intrinsically disordered transcriptional repression domain ([29]). Structural characterization of intrinsically disordered regions of other proteins such as p53 ([64,65]), PTP1B ([66]), androgen receptor ([67]), α-synuclein ([68,69]) and oncoprotein c-myc ([70,71]) has elucidated the function of these proteins in pathological pathways and lead to successful efforts developing small molecule inhibitors for therapeutic drug targeting. The work presented here sets the stage for further biochemical and functional characterization of the MBD2-NuRD complex and raises the possibility of selectively abrogating NuRD dependent silencing of specific sets of genes by disrupting the $MBD2_{IDR}$ mediated interactions with the HDAC complex. Exploitation of such mechanistic findings could ultimately lead to improved treatment of hemoglobin disorders and cancer.

ACCESSION NUMBERS

The NMR assignments have been deposited in the Biological Magnetic Resonance Bank (BMRB accession: 25426).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

FUNDING

REFERENCES

1. Kim,J.K., Samaranayake,M. and Pradhan,S. (2009) Epigenetic mechanisms in mammals. *Cell Mol. Life Sci.*, **66**, 596–612.
2. Illingworth,R.S. and Bird,A.P. (2009) CpG islands–'a rough guide'. *FEBS Lett.*, **583**, 1713–1720.
3. Straussman,R., Nejman,D., Roberts,D., Steinfeld,I., Blum,B., Benvenisty,N., Simon,I., Yakhini,Z. and Cedar,H. (2009) Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.*, **16**, 564–571.
4. Rupon,J.W., Wang,S.Z., Gaensler,K., Lloyd,J. and Ginder,G.D. (2006) Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 6617–6622.
5. Kransdorf,E.P., Wang,S.Z., Zhu,S.Z., Langston,T.B., Rupon,J.W. and Ginder,G.D. (2006) MBD2 is a critical component of a methyl cytosine-binding protein complex isolated from primary erythroid cells. *Blood*, **108**, 2836–2845.

6. Berger,J., Sansom,O., Clarke,A. and Bird,A. (2007) MBD2 is required for correct spatial gene expression in the gut. *Mol. Cell. Biol.*, **27**, 4049–4057.

7. Jones,P.A. and Baylin,S.B. (2007) The epigenomics of cancer. *Cell*, **128**, 683–692.

8. Meehan,R.R., Lewis,J.D., McKay,S., Kleiner,E.L. and Bird,A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, **58**, 499–507.

9. Hendrich,B. and Bird,A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.

10. 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.

11. 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.

12. 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.

13. Fraga,M.F., Ballestar,E., Montoya,G., Taysavang,P., Wade,P.A. and Esteller,M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.*, **31**, 1765–1774.

14. Feng,Q. and Zhang,Y. (2001) The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev.*, **15**, 827–832.

15. Kloet,S.L., Baymaz,H.I., Makowski,M., Groenewold,V., Jansen,P.W., Berendsen,M., Niazi,H., Kops,G.J. and Vermeulen,M. (2014) Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS J.*, doi:10.1111/febs.12972.

16. Smits,A.H., Jansen,P.W., Poser,I., Hyman,A.A. and Vermeulen,M. (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.*, **41**, e28.

17. Singal,R., Wang,S.Z., Sargent,T., Zhu,S.Z. and Ginder,G.D. (2002) Methylation of promoter proximal-transcribed sequences of an embryonic globin gene inhibits transcription in primary erythroid cells and promotes formation of a cell type-specific methyl cytosine binding complex. *J. Biol. Chem.*, **277**, 1897–1905.

18. 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) p66{alpha}-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.

19. Rupon,J.W., Wang,S.Z., Gnanapragasam,M., Labropoulos,S. and Ginder,G.D. (2011) MBD2 contributes to developmental silencing of the human epsilon-globin gene. *Blood Cells Mol. Dis.*, **46**, 212–219.

20. Mian,O.Y., Wang,S.Z., Zhu,S.Z., Gnanapragasam,M.N., Graham,L., Bear,H.D. and Ginder,G.D. (2011) Methyl-binding domain protein 2-dependent proliferation and survival of breast cancer cells. *Mol. Cancer Res.*, **9**, 1152–1162.

21. Magdinier,F. and Wolffe,A.P. (2001) Selective association of the methyl-CpG binding protein MBD2 with the silent p14/p16 locus in human neoplasia. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4990–4995.

22. Sansom,O.J., Berger,J., Bishop,S.M., Hendrich,B., Bird,A. and Clarke,A.R. (2003) Deficiency of Mbd2 suppresses intestinal tumorigenesis. *Nat. Genet.*, **34**, 145–147.

23. Stefanska,B., Suderman,M., Machnes,Z., Bhattacharyya,B., Hallett,M. and Szyf,M. (2013) Transcription onset of genes critical in liver carcinogenesis is epigenetically regulated by methylated DNA-binding protein MBD2. *Carcinogenesis*, **34**, 2738–2749.

24. Yu,S., Khor,T.O., Cheung,K.L., Li,W., Wu,T.Y., Huang,Y., Foster,B.A., Kan,Y.W. and Kong,A.N. (2010) Nrf2 expression is regulated by epigenetic mechanisms in prostate cancer of TRAMP mice. *PLoS One*, **5**, e8579.

25. Lopez-Serra,L., Ballestar,E., Ropero,S., Setien,F., Billard,L.M., Fraga,M.F., Lopez-Nieva,P., Alaminos,M., Guerrero,D., Dante,R. *et al.* (2008) Unmasking of epigenetically silenced candidate tumor suppressor genes by removal of methyl-CpG-binding domain proteins. *Oncogene*, **27**, 3556–3566.

26. Esteller,M. (2008) Epigenetics in cancer. *N. Engl. J. Med.*, **358**, 1148–1159.

27. Brackertz,M., Gong,Z., Leers,J. and Renkawitz,R. (2006) p66alpha and p66beta of the mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res.*, **34**, 397–406.

28. Tompa,P. (2012) Intrinsically disordered proteins: A 10-year recap. *Trends Biochem. Sci.*, **37**, 509–516.

29. Hameed,U.F., Lim,J., Zhang,Q., Wasik,M.A., Yang,D. and Swaminathan,K. (2014) Transcriptional repressor domain of MBD1 is intrinsically disordered and interacts with its binding partners in a selective manner. *Sci. Rep.*, **4**, e4896.

30. Adams,V.H., McBryant,S.J., Wade,P.A., Woodcock,C.L. and Hansen,J.C. (2007) Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem.*, **282**, 15057–15064.

31. Cai,M., Williams,D.C. Jr, Wang,G., Lee,B.R., Peterkofsky,A. and Clore,G.M. (2003) Solution structure of the phosphoryl transfer complex between the signal-transducing protein IIAGlucose and the cytoplasmic domain of the glucose transporter IICBGlucose of the escherichia coli glucose phosphotransferase system. *J. Biol. Chem.*, **278**, 25191–25206.

32. 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.

33. 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.

34. Lebowitz,J., Lewis,M.S. and Schuck,P. (2002) Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci.*, **11**, 2067–2079.

35. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.

36. 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.

37. Kumita,J.R., Smart,O.S. and Woolley,G.A. (2000) Photo-control of helix content in a short peptide. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 3803–3808.

38. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

39. 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.

40. Markley,J.L., Bax,A., Arata,Y., Hilbers,C.W., Kaptein,R., Sykes,B.D., Wright,P.E. and Wuthrich,K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *J. Biomol. NMR*, **12**, 1–23.

41. Kjaergaard,M. and Poulsen,F.M. (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the ramachandran distribution. *J. Biomol. NMR*, **50**, 157–165.

42. Kingston,R.E., Chen,C.A. and Okayama,H. (2003) Calcium phosphate transfection. *Curr. Protoc. Cell. Biol.*, **19**, 20.3.1–20.3.8.

43. Bjellqvist,B., Hughes,G.J., Pasquali,C., Paquet,N., Ravier,F., Sanchez,J.C., Frutiger,S. and Hochstrasser,D. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023–1031.

44. Li,X., Romero,P., Rani,M., Dunker,A.K. and Obradovic,Z. (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.*, **10**, 30–40.

45. Romero,Obradovic, and Dunker,K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform. Ser. Workshop Genome Inform.*, **8**, 110–124.
46. Fuxreiter,M. (2012) Fuzziness: linking regulation to protein dynamics. *Mol. Biosyst*, **8**, 168–177.
47. Fuxreiter,M. and Tompa,P. (2012) Fuzzy complexes: a more stochastic view of protein function. *Adv. Exp. Med. Biol.*, **725**, 1–14.
48. Uversky,V.N. (2011) Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.*, **43**, 1090–1103.
49. Vacic,V., Oldfield,C.J., Mohan,A., Radivojac,P., Cortese,M.S., Uversky,V.N. and Dunker,A.K. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.*, **6**, 2351–2366.
50. Mohan,A., Oldfield,C.J., Radivojac,P., Vacic,V., Cortese,M.S., Dunker,A.K. and Uversky,V.N. (2006) Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.*, **362**, 1043–1059.
51. Jiang,C.L., Jin,S.G., Lee,D.H., Lan,Z.J., Xu,X., O'Connor,T.R., Szabó,P.E., Mann,J.R., Cooney,A.J. and Pfeifer,G.P. (2002) MBD3L1 and MBD3L2, two new proteins homologous to the methyl-CpG-binding proteins MBD2 and MBD3: characterization of MBD3L1 as a testis-specific transcriptional repressor. *Genomics*, **80**, 621–629.
52. Jiang,C., Jin,S. and Pfeifer,G.P. (2004) MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex. *J. Biol. Chem.*, **279**, 52456–52464.
53. Jin,S., Jiang,C., Rauch,T., Li,H. and Pfeifer,G.P. (2005) MBD3L2 interacts with MBD3 and components of the NuRD complex and can oppose MBD2-MeCP1-mediated methylation silencing. *J. Biol. Chem.*, **280**, 12700–12709.
54. Ghosh,R.P., Nikitina,T., Horowitz-Scherer,R.A., Gierasch,L.M., Uversky,V.N., Hite,K., Hansen,J.C. and Woodcock,C.L. (2010) Unique properties and interactions of the domains of the methylated DNA binding protein 2 (MeCP2). *Biochemistry*, **49**, 4395–4410.
55. Pontius,B.W. (1993) Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem. Sci.*, **18**, 181–186.
56. Patil,A. and Nakamura,H. (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.*, **580**, 2041–2045.
57. Haynes,C., Oldfield,C.J., Ji,F., Klitgord,N., Cusick,M.E., Radivojac,P., Uversky,V.N., Vidal,M. and Iakoucheva,L.M. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol.*, **2**, e100.
58. Dunker,A.K., Cortese,M.S., Romero,P., Iakoucheva,L.M. and Uversky,V.N. (2005) Flexible nets. the roles of intrinsic disorder in protein interaction networks. *FEBS J.*, **272**, 5129–5148.
59. Millard,C.J., Watson,P.J., Celardo,I., Gordiyenko,Y., Cowley,S.M., Robinson,C.V., Fairall,L. and Schwabe,J.W. (2013) Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol. Cell*, **51**, 57–67.
60. Alqarni,S.S., Murthy,A., Zhang,W., Przewloka,M.R., Silva,A.P., Watson,A.A., Lejon,S., Pei,X.Y., Smits,A.H., Kloet,S.L. *et al.* (2014) Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex. *J. Biol. Chem.*, **289**, 21844–21855.
61. Lu,Y., Loh,Y.H., Li,H., Cesana,M., Ficarro,S.B., Parikh,J.R., Salomonis,N., Toh,C.X., Andreadis,S.T., Luckey,C.J. *et al.* (2014) Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. *Cell. Stem Cell.*, **15**, 92–101.
62. Chen,L.M. and Chai,K.X. (2002) Prostasin serine protease inhibits breast cancer invasiveness and is transcriptionally regulated by promoter DNA methylation. *Int. J. Cancer*, **97**, 323–329.
63. Chen,L.M., Hodge,G.B., Guarda,L.A., Welch,J.L., Greenberg,N.M. and Chai,K.X. (2001) Down-regulation of prostasin serine protease: a potential invasion suppressor in prostate cancer. *Prostate*, **48**, 93–103.
64. Dawson,R., Muller,L., Dehner,A., Klein,C., Kessler,H. and Buchner,J. (2003) The N-terminal domain of p53 is natively unfolded. *J. Mol. Biol.*, **332**, 1131–1141.
65. Wells,M., Tidow,H., Rutherford,T.J., Markwick,P., Jensen,M.R., Mylonas,E., Svergun,D.I., Blackledge,M. and Fersht,A.R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 5762–5767.
66. Krishnan,N., Koveal,D., Miller,D.H., Xue,B., Akshinthala,S.D., Kragelj,J., Jensen,M.R., Gauss,C.M., Page,R., Blackledge,M. *et al.* (2014) Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat. Chem. Biol.*, **10**, 558–566.
67. Myung,J.K., Banuelos,C.A., Fernandez,J.G., Mawji,N.R., Wang,J., Tien,A.H., Yang,Y.C., Tavakoli,I., Haile,S., Watt,K. *et al.* (2013) An androgen receptor N-terminal domain antagonist for treating prostate cancer. *J. Clin. Invest.*, **123**, 2948–2960.
68. Uversky,V.N., Li,J. and Fink,A.L. (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.*, **276**, 10737–10744.
69. Toth,G., Gardai,S.J., Zago,W., Bertoncini,C.W., Cremades,N., Roy,S.L., Tambe,M.A., Rochet,J.C., Galvagnion,C., Skibinski,G. *et al.* (2014) Targeting the intrinsically disordered structural ensemble of alpha-synuclein by small molecules as a potential therapeutic strategy for parkinson's disease. *PLoS One*, **9**, e87133.
70. Berg,T., Cohen,S.B., Desharnais,J., Sonderegger,C., Maslyar,D.J., Goldberg,J., Boger,D.L. and Vogt,P.K. (2002) Small-molecule antagonists of Myc/Max dimerization inhibit myc-induced transformation of chicken embryo fibroblasts. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 3830–3835.
71. Hammoudeh,D.I., Follis,A.V., Prochownik,E.V. and Metallo,S.J. (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-myc. *J. Am. Chem. Soc.*, **131**, 7390–7401.