Unexpected thermodynamic signature for the interaction of hydroxymethylated DNA with MeCP2

David Ortega-Alarcon[1], Rafael Claveria-Gimeno[1,2†], Sonia Vega[1], Olga C. Jorge-Torres[3], Manel Esteller[3,4,5,6], Olga Abian[1,2,7,8*] and Adrian Velazquez-Campoy[1,2,3,7,8*]

[1] Institute of Biocomputation and Physics of Complex Systems (BIFI), Joint Units GBsC-CSIC-BIFI and ICVV-CSIC-BIFI, Universidad de Zaragoza, Zaragoza, 50018, Spain
[2] Instituto de Investigación Sanitaria Aragón (IIS Aragón), 50009, Zaragoza, Spain
[3] Josep Carreras Leukaemia Research Institute (IJC), 08916, Badalona, Barcelona, Spain
[4] Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), 28029, Madrid, Spain
[5] Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010, Barcelona, Spain
[6] Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), 08907, l'Hospitalet de Llobregat, Barcelona, Spain
[7] Centro de Investigación Biomédica en Red en el Área Temática de Enfermedades Hepáticas y Digestivas (CIBERehd), 28029, Madrid, Spain
[8] Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, 50009, Zaragoza, Spain

* To whom correspondence should be addressed. Tel: +34 976 762996; Email: adrianvc@unizar.es. Correspondence may also be addressed to Olga Abian. Email: oabifra@unizar.es
Present Address: Rafael Claveria-Gimeno, CerTest Biotec S.L., 50840, San Mateo de Gallego, Zaragoza, Spain

ABSTRACT

Hydroxymethylated cytosine (5hmC) is a stable DNA epigenetic mark recognized by methyl-CpG binding protein 2 (MeCP2), which acts as a transcriptional regulator and a global chromatin-remodeling element. Because 5hmC triggers a gene regulation response markedly different from that produced by methylated cytosine (5mC), both modifications must affect DNA structure and/or DNA interaction with MeCP2 differently. MeCP2 is a six-domain intrinsically disordered protein (IDP) with two domains responsible for dsDNA binding: methyl-CpG binding domain (MBD) and intervening domain (ID). Here we report the detailed thermodynamic characterization of the interaction of hmCpG-DNA with MeCP2. We find that hmCpG-DNA interacts with MeCP2 in a distinctly different mode with a particular thermodynamic signature, compared to methylated or unmethylated DNA. In addition, we find evidence for Rett syndrome-associated mutations altering the interaction of MeCP2 with dsDNA in a cytosine modification-specific manner which may correlate with disease onset time and clinical severity score.

KEYWORDS

Methyl-CpG binding protein 2 (MeCP2); Rett syndrome; protein-DNA interaction; hydroxymethylated DNA; calorimetry; ligand-induced protein stabilization

INTRODUCTION

Epigenetics consists of the transient or long-lasting modification of gene transcription without altering the genetic code imprinted within the DNA sequence. Chemical modifications and local/global architecture of DNA as well as associated chromatin proteins govern the transcriptional status of genes. This pervasive mechanism constitutes the molecular basis for the ability of differentiated cells to express only those genes that are necessary for their own activity. Just as allosterism, a high-level regulation of protein function based on the interplay of protein conformation, association and interaction, was termed “the second secret of life” [1], epigenetics can be considered “the third secret of life”.

Epigenetic information encrypted through chemical marks is maintained and modified by writers/erasers, and interpreted and integrated by readers. Cytosine modification within a CpG context is one of such regulatory elemental pieces of epigenetic information, for which DNA-methyltransferases, DNA-demethylases, and methyl-CpG binding proteins collectively operate in a coordinated way.

Modifications of carbon-5 of cytosine regulate transcription by sequence- and context-dependent effects on the interactions of epigenetic readers such as DNA transcription factors [2]. Cytosine methylation (5mC) is an epigenetic mark with a direct biological function usually consisting on transcriptional repression. 5mC is vital for many important processes: embryogenesis, parental imprinting, X-chromosome inactivation, silencing of endogenous retroviruses, regulation of gene expression, and splicing. Oxidation of 5mC leads to hydroxymethylated cytosine (5hmC) which is involved in regulation of pluripotency, senescence, neuronal development, and tumorigenesis [3–5]. 5hmC was identified in viruses [6] and vertebrates [7], but it has only drawn attention in the last decade. The problematic discrimination between 5mC and 5hmC obscured the discernment of the biological effect of both modifications.

5hmC has been identified as an important epigenetic modification in mammalian cells [3]. 5hmC is enriched in embryonic stem cells and its levels decrease as cells differentiate, except in the brain, with 10-fold higher levels than in peripheral tissues and stem cells [3,8,9]. In brain cortex, about 4% of all cytosines are 5mC, and about 1% are 5hmC (i.e., 25% of 5mC) [10,11]. In particular, in Purkinje cells 5hmC ~~level is~~ represents 40% of 5mC level [3].

It has been found that 5hmC is critical for brain development: abundant in developing neurons, with higher levels than in neuronal progenitor cells, and specifically localizing within gene bodies, enhancers and promoters (but not at transcription start sites, which results in gene inactivation) in important genes for neuronal differentiation and synaptic function, whereas 5mC is strongly depleted within those regions [12,13].

Several 5hmC readers have been identified (e.g., methyl-CpG-binding protein 3 or MBD3, ubiquitin like with PHD and ring finger domains 2 protein or Uhrf2, and chromatin target of PRMT1 or CHTOP-methylosome complex) [14,15], but methyl-CpG binding protein 2 (MeCP2) is the major 5hmC-binding protein in the brain and the protein with highest affinity for 5hmC, *in vitro* and *in vivo* [16].

MeCP2 is a transcriptional regulator involved in early stages of neuronal development, differentiation, maturation, and synaptic plasticity control [17], but it also binds massively to chromatin acting as a chromatin architecture remodeling factor [18–20]. MeCP2 point mutations or deletions causing activity

loss or deficiency are associated with Rett syndrome (RTT, OMIM#312750). Although a rare disease (1:10000 births), RTT is the main cause of mental retardation in females [21–23].

MeCP2 binds 5hmC- and 5mC-containing DNA with comparable affinity [14,16]. Although the binding to 5mC and 5hmC is not sequence-specific, as higher affinity for 5mC has been reported [24,25]. From molecular modeling, a reduction of 0.6 kcal/mol in Gibbs energy of binding (i.e., 3-fold less favorable binding affinity) has been estimated for 5hmC compared to 5mC [26]. According to electrophoretic mobility shift and surface plasmon resonance assays, the RTT-associated mutation R133C in MeCP2 preferentially inhibited 5hmC binding, but affinities were not quantitated [16].

5mC and 5hmC in CpG-island promoters exert different effects on gene expression: 5mC is associated with stably repressed gene state within compact heterochromatin, whereas 5hmC is associated with increased cis-gene expression within euchromatin [27–29]. Then, MeCP2 may facilitate transcription by binding to 5hmC in neuronal cell types, and repress transcription by binding to 5mC. Interestingly, some 5mC recognition proteins, such as MBD1, MBD2 or MBD4, do not bind 5hmC. This suggests 5hmC might represent an exclusion mechanism for some MBDs [30–32], and 5hmC might have a dual role: 1) net demethylation hindering 5mC binders; and 2) stable DNA modification with effector activity by recruiting 5hmC binders. Therefore, 5hmC activate genes by its own effect or by decreasing 5mC levels and diminishing the binding of silencing and inhibiting repressor proteins.

5hmC-enriched DNA is released readily from chromatin by micrococcal nuclease (MNase) digestion, whereas 5mC-containing chromatin is more resistant to digestion. This is consistent with 5mC promoting compact chromatin [33]. MeCP2 regulates the accessibility of 5hmC-containing DNA to MNase, supporting a model in which MeCP2 binding to 5hmC within highly expressed genes may widely facilitate transcription through its effects on chromatin organization [16].

5hmC also modify strand separation propensity, thus, controlling gene expression by facilitating the action of the transcription machinery on DNA [34]. It has been reported that 5hmC decreases DNA stability [35,36], while 5mC increases DNA stability by enhancing the hydrophobicity of the cytosine [37,38]. 5mC and 5hmC are stiffer than unmethylated cytosine, resulting in poorer circularization efficiencies and lower ability to form nucleosomes [26]. Overall, 5mC causes changes in chromatin reorganization due to increased stiffness and MBD proteins binding, whereas modification to 5hmC further increases stiffness, promotes opening of DNA by the transcription machinery, and provokes functional changes due to distortion of repression/activation associated to differential binding of MBD proteins [26]. Therefore, the effect of cytosine modifications will be the combined effect on DNA structure and effectors interactions.

Experimental evidence indicates differential roles for 5mC and 5hmC in gene expression regulation in different brain cell types, with implications for the etiology of human brain diseases [39]. Because levels in 5hmC and MeCP2 increase during neuron development and maturation, and mutations in MeCP2 result in varied RTT onset time and disease burden and severity, RTT-associated mutations might impact MeCP2 and its interactions in a cytosine modification-dependent manner. In an attempt to unravel distinctive features of the interaction of MeCP2 with 5hmC, compared to 5C and 5mC, we have performed a thorough thermodynamic characterization of such interaction by a combination of biophysical techniques.

MATERIAL AND METHODS

Protein expression and purification

All protein variants (MBD, ID, NTD-MBD-ID, and full-length MeCP2, as well as NTD-MBD-ID R106W and R133C mutants) were expressed and purified following the same procedure. These constructs correspond to MeCP2 isoform e2 (Figure S1). Transformed BL21 (DE3) Star *E. coli* cultures were grown at 37 °C overnight in Luria-Bertani (LB) medium supplemented with 50 μg/mL kanamycin. Afterwards, 4 L of LB supplemented with 25 μg/mL kanamycin were inoculated (1:100 dilution) and incubated until reaching an OD of 0.6 at a 600 nm wavelength. Protein expression was induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) at 18 °C overnight. Cells were broken by sonication in ice and 20 U/mL benzonase (Merck-Millipore, Madrid, Spain) was added to degrade nucleic acids. Proteins were purified using metal affinity chromatography in a HiTrap TALON column (GE-Healthcare Life Sciences, Barcelona, Spain) with two washing steps: 1) 50 mM sodium phosphate, pH 7, 300 mM NaCl, and 2) 50 mM sodium phosphate, pH 7, 800 mM NaCl (to remove potential DNA contamination), before a 10-150 mM imidazole elution gradient. Purity was checked by SDS-PAGE.

Removal of the histidine-tag was performed by GST-tagged PreScission Protease processing in cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) at 4 °C for 4 hours. Progress of the protease processing was monitored by SDS-PAGE. Finally, proteins were further purified using a combination of two affinity chromatographic steps to remove the histidine-tag or uncleaved protein (HiTrap TALON column, from GE-Healthcare Life Sciences, Barcelona, Spain) and the GST-tagged PreScission Protease (GST TALON column, from GE-Healthcare Life Sciences, Barcelona, Spain). Purity and homogeneity were checked by SDS-PAGE and size-exclusion chromatography. The proteins were stored in 50 mM Tris pH 7.0 at –80 °C. The identity of all proteins was checked by mass spectrometry (4800plus MALDI-TOF/MS, from Applied Biosystems - Thermo Fisher Scientific, Waltham, MA, USA). Potential DNA contamination was always assessed by determining the 260/280 UV absorption ratio. An extinction coefficient of 11460 M^{-1} cm^{-1} at 280 nm was employed for all variants (a single tryptophan is located in MBD), except for the R106W mutant (16960 M^{-1} cm^{-1}) and ID (which does not have aromatic residues).

DNA interaction calorimetric assays were performed in different buffers (50 mM Tris pH 7, 0-150 mM NaCl; 50 mM Pipes, pH 7; 50 mM Phosphate, pH 7). Thus, when required, buffer exchange was carried out using a 3 or 10 kDa-pore size ultrafiltration device (Amicon centrifugal filter, Merck-Millipore, Madrid, Spain) at 4000 rpm and 4 °C.

Double-stranded DNA

HPLC-purified hydroxymethylated 45-bp single-stranded DNA (ssDNA) oligonucleotides corresponding to the promoter IV of the brain-derived neurotrophic factor (BDNF) were obtained from Integrated DNA Technologies. Two complementary pairs of DNA were used for DNA interaction assays: forward hmCpG, 5'-GCCATGCCCTGGAA(5-hMe)CGGAACTCTCCTAATAAAAGATGTATCATTT-3'; reverse hmCpG, 5'-AAATGATACATCTTTTATTAGGAGAGTTC(5-hMe)CGTTCCAGGGCATGGC-3'.

The oligonucleotides were annealed to obtain 45-bp double-stranded symmetrically hydroxymethylated DNA (hmCpG-DNA). Once dissolved to obtain a 0.5 mM ssDNA solution for each oligonucleotide, they were mixed at an equimolar ratio and annealed using a Stratagene Mx3005P qPCR real-time thermal cycler (Agilent Technologies, Santa Clara, CA, USA). The thermal annealing profile consisted of four steps: 1) equilibration at 25 °C for 30 s; 2) heating ramp up to 99 °C; 3) equilibration at 99 °C for 60 s; and 4) 3-hour cooling process down to 25 °C at a rate of 1 °C/180 s.

Fluorescence spectroscopy

Thermal unfolding assays were performed in a Cary Eclipse fluorescence spectrophotometer (Varian – Agilent, Santa Clara, CA, USA) using a 1 cm path-length quartz cuvette (Hellma Analytics, Müllheim, Germany). The temperature was controlled by a Peltier unit and monitored using a temperature probe. Thermal stability assays were performed in 50 mM Tris pH 7, at a heating rate of 1 °C/min using excitation and emission wavelengths of 290 nm and 330 nm (maximal spectral change for intrinsic tryptophan fluorescence emission with temperature), with 5 nm bandwidths. Thermal unfolding of the protein was reversible and the experiments were analyzed considering a two-state unfolding model (for MBD and NTD-MBD-ID):

$$F = \frac{(A_N + B_N T) + (A_U + B_U T) exp\left(-\frac{\Delta G(T)}{RT}\right)}{1 + exp\left(-\frac{\Delta G(T)}{RT}\right)}$$
$$\Delta G(T) = \Delta H(T_m)\left(1 - \frac{T}{T_m}\right) + \Delta C_P\left(T - T_m - T\ln\frac{T}{T_m}\right) \quad (1)$$

or a three-state unfolding model (for full-length MeCP2):

$$F = \frac{(A_N + B_N T) + (A_I + B_I T) exp\left(-\frac{\Delta G_1(T)}{RT}\right) + (A_U + B_U T) exp\left(-\frac{\Delta G_1(T) + \Delta G_2(T)}{RT}\right)}{1 + exp\left(-\frac{\Delta G_1(T)}{RT}\right) + exp\left(-\frac{\Delta G_1(T) + \Delta G_2(T)}{RT}\right)}$$
$$\Delta G_i(T) = \Delta H_i(T_{m,i})\left(1 - \frac{T}{T_{m,i}}\right) + \Delta C_{P,i}\left(T - T_{m,i} - T\ln\frac{T}{T_{m,i}}\right) \quad (2)$$

where F is the fluorescence signal, R is the ideal gas constant, T is the absolute temperature, T_m is the unfolding temperature, $\Delta H_i(T_m)$ is the unfolding enthalpy for the transition i, $\Delta C_{P,i}$ is the unfolding heat capacity for the transition i, and ΔG_i is the stabilization Gibbs energy for the transition i. The adjustable parameters A_N, B_N, A_I, B_I, A_U, and B_U define the intrinsic fluorescence signal of the distinguishable conformational states (N: native; I: intermediate; U: unfolded) and their temperature dependence. The stabilizing effect upon dsDNA interaction was assessed performing thermal denaturations of the different proteins (at 5 µM) in the presence of 5hmC-containing dsDNA (hmCpG-DNA) (at 10 µM) under the same conditions.

Dynamic light scattering (DLS)

Dynamic light scattering measurements were performed in a DynaPro Plate Reader III (Wyatt Technology, Santa Barbara, CA) using a 384-multiwell plate (Aurora Micro-plates, Whitefish, MT). All solutions were filtered using 0.2-µm membranes and protein stocks were centrifuged in microtubes for 2 minutes at maximum speed to prevent protein aggregates from interfering with the DLS measurements. For each measurement, 5 acquisitions of 5 seconds were taken, and the apparent

hydrodynamic radius was estimated from the experimental diffusion coefficient, obtained by the cumulant fit of the translational autocorrelation function, assuming a Rayleigh sphere model. Experiments were performed in Tris 50 mM pH 7.0, at 20 °C, with protein solutions at 20 μM concentration in the absence and the presence of dsDNA (unmethylated dsDNA, CpG-DNA; 5mC-containing dsDNA, mCpG-DNA; and 5hmC-containing dsDNA, hmCpG-DNA) at 20 μM.

Differential scanning calorimetry (DSC)

Thermal unfolding assays were also performed in a high-precision differential scanning calorimeter (DSC). The excess partial molar heat capacity of the protein in solution was measured as a function of temperature in an Auto-PEAQ-DSC (MicroCal, Malvern-Panalytical, Malvern, UK). Experiments were performed in 50 mM Tris pH 7, with a 20 μM protein solution and 20 μM dsDNA solution (unmethylated dsDNA, CpG-DNA; 5mC-containing dsDNA, mCpG-DNA; and 5hmC-containing dsDNA, hmCpG-DNA), scanning from 15 to 95 °C at a scan rate of 60 °C/h.

The thermal unfolding cooperativity can be quantitated by applying a model-free data analysis based on the van 't Hoff unfolding enthalpy. From the thermogram (excess molar heat capacity of the protein as a function of the temperature, $\Delta C_P(T)$), the calorimetric unfolding enthalpy, ΔH_{cal}, the unfolding temperature, T_m, and the maximal unfolding heat capacity, $\Delta C_{P,max}$, can estimated, from which it is possible to calculate the van 't Hoff enthalpy, ΔH_{vH}:

$$\Delta H_{vH} = \frac{4RT_m^2 \Delta C_{P,max}}{\Delta H_{cal}} \quad (3)$$

The ratio $\Delta H_{cal}/\Delta H_{vH}$ can be considered as a cooperativity index: 1) if $\Delta H_{cal}/\Delta H_{vH} = 1$, the thermogram would reflect a single transition and the protein unfolds according to a two-state model (i.e., the macromolecule contains a single energetic domain); 2) if $\Delta H_{cal}/\Delta H_{vH} > 1$, the thermogram would reflect two or more partially overlapping transitions and the protein unfolds according to a non-two-state model (i.e., the macromolecule contains two or more domains or unfolding units that unfold individually); and 3) if $\Delta H_{cal}/\Delta H_{vH} < 1$, the thermogram would reflect an unfolding transition coupled to subunit dissociation (i.e., several protein subunits associate into an oligomeric unfolding entity that unfolds into isolated monomers). The $\Delta H_{cal}/\Delta H_{vH}$ ratio represents a geometric parameter comparing the area and the width of unfolding trace: for a given area (calorimetric unfolding enthalpy) the narrower the transition, the higher the unfolding cooperativity.

Isothermal titration calorimetry (ITC)

The interaction between the different proteins and hmCpG-DNA was studied using a high-sensitivity Auto-iTC200 isothermal titration calorimeter (MicroCal, Malvern-Panalytical, UK). Experiments were performed in different buffers: 50 mM Tris pH 7, 0-150 mM NaCl; 50 mM Pipes, pH 7; 50 mM Phosphate, pH 7. Protein solution in the calorimetric cell at 3-5 μM was titrated with hmCpG-DNA solution at 50 μM in the injecting syringe. A sequence of 2 μL-injections of titrant solution evenly spaced by 150 s was programmed with a stirring speed of 750 rpm and a reference thermal power of 10 μcal/s. The association constant, K_B, and the apparent enthalpy of binding, $\Delta H_{B,obs}$, were estimated through non-linear least-squares regression data analysis of the experimental data employing a single ligand binding

site model (1:1 protein:dsDNA stoichiometry) for MBD or ID interacting with hmCpG-DNA, or a two ligand binding sites model (1:2 protein:dsDNA stoichiometry) for NTD-MBD-ID interacting with hmCpG-DNA, employing user-defined fitting routines implemented in Origin (OriginLab, Northampton, MA):

$$[P]_{T,j} = [P]_0\left(1-\frac{v_{inj}}{V_0}\right)^j, [L]_{T,j} = [L]_0\left(1-\left(1-\frac{v_{inj}}{V_0}\right)^j\right)$$
$$[P_iL]_j = [P]_{T,j}\frac{K_{Bi}[L]_j}{1+K_{Bi}[L]_j}$$
$$[L]_{T,j} = [L]_j + \sum_{i=1}^{N}[P]_{T,j}\frac{K_{Bi}[L]_j}{1+K_{Bi}[L]_j} \quad (4)$$
$$Q_j = \frac{1}{v_{inj}[L]_0}\sum_{i=1}^{N} V_0\Delta H_{B,obs,i}\left([P_iL]_j - [P_iL]_{j-1}\left(1-\frac{v_{inj}}{V_0}\right)\right) + Q_d$$

where Q_j is the ligand injected-normalized heat effect after injection j, v_{inj} is the injection volume, V_0 is the cell volume, Q_d is the background injection heat (the so-called "heat of dilution"), P_iL is the protein-dsDNA complex in which binding site i is occupied, K_{Bi} is the association constant for complex P_iL, $\Delta H_{B,obs,i}$ is the binding enthalpy for complex P_iL, $[P]_0$ is the initial protein concentration in the cell, $[L]_0$ is the ligand concentration in the syringe, $[P]_{T,j}$ is the total concentration of protein in the cell after injection j, $[L]_{T,j}$ is the total concentration of ligand after injection j, $[P]_j$ is the concentration of free protein after injection j, and $[L]$ is the concentration of free ligand after injection j. For a single binding site $N = 1$, and for two binding sites $N = 2$. A detailed description of those models applied in ITC can be found elsewhere [40,41]. The dissociation constant K_d for each binding site was calculated as the inverse of K_B, and the binding Gibbs energy and entropy were calculated applying standard well-known relationships: $\Delta G_B = -RT\ln K_B$, $\Delta G_B = \Delta H_B - T\Delta S_B$. The binding heat capacity, ΔC_{PB}, was determined by performing titrations at different temperatures and estimating the derivative of the observed enthalpy with temperature through linear regression in the enthalpy vs. temperature plot.

The association binding constant K_B is not affected by the possible buffer ionization as long as the pK_a of the buffer is close to the experimental pH. However, the observed binding enthalpy (and, therefore, the entropic contribution also) will contain an additional contribution from buffer ionization. Therefore, the buffer selected would potentially affect the partition of the Gibbs enthalpy into enthalpy and entropy. The extrinsic contributions from the buffer can be removed by estimating the buffer-independent binding enthalpy, ΔH_B, according to [42,43]:

$$\Delta H_{B,obs} = \Delta H_B + \Delta n_H \Delta H_{buffer} \quad (5)$$

where Δn_H is the number of protons exchanged between the protein-dsDNA complex and the bulk solution upon dsDNA binding, and ΔH_{buffer} is the ionization enthalpy of the selected buffer. Therefore, titrations were performed in buffers with different ionization enthalpies (Tris, 11.35 kcal/mol; Pipes, 2.67 kcal/mol; and phosphate, 0.86 kcal/mol) [44] in order to estimate the buffer-independent thermodynamic parameters (ΔH and Δn_H) from linear regression using equation 5. Knowing the binding Gibbs energy and the buffer-independent binding enthalpy, the buffer-independent binding entropy can be readily calculated. The parameter Δn_H may be non-zero when ligand binding results in changes in the proton dissociation constant of certain ionizable residues (in the protein or the ligand) as a consequence of changes in their microenvironment upon complex formation.

RESULTS

hmCpG-DNA induces a smaller stabilization compared to mCpG-DNA

The thermal denaturation of MBD and NTD-MBD-ID in the presence of hmCpG-DNA followed by intrinsic tryptophan fluorescence emission could be analyzed by applying the two-state unfolding model (Figure 1A). The observed differences in the unfolding parameters, T_m and $\Delta H(T_m)$, for the three dsDNA types (CpG-DNA, mCpG-DNA, and hmCpG-DNA) are larger than the experimental error (Figure 1A and Tables 1 and S1). An induced stabilization effect intermediate between that of CpG-DNA and mCpG-DNA (in terms of unfolding temperature and enthalpy) was observed for hmCpG-DNA.

Interestingly, contrary to the results reported for CpG-DNA and mCpG-DNA [45], the thermal unfolding of full-length MeCP2 in the presence of hmCpG-DNA could not be satisfactorily fitted considering a single transition. Thus, a two-transition model (i.e., three-state model) had to be considered, with T_m's separated by about 10 °C (Figure 1A and Tables 1 and S1). This suggests that spectroscopic unfolding traces might not always reveal the subtle intricacies associated with the unfolding process. In other words, the unfolding of full-length MeCP2 in the presence of CpG-DNA-bound or mCpG-DNA-bound full length MeCP2 followed by intrinsic fluorescence may overlook some events, yielding a single apparent transition. In fact, depending on the difference in T_m and $\Delta H(T_m)$, and depending on the intrinsic fluorescence signal for the three conformational states involved in a three-state model, the overall unfolding trace may show a single broad thermal transition (i.e., with a small apparent unfolding enthalpy typical of low unfolding cooperativity) or two distinguishable sequential thermal transitions.

The apparent hydrodynamic radius of the different protein-dsDNA complexes was measured (Figure 1B). The size histograms showed average and standard deviation values that varied according to the type of dsDNA bound. The histograms exhibited positive skewness, likely reflecting the heterogenous, biased folding landscape of MeCP2 towards disordered conformations. MBD and NTD-MBD-ID showed a hydrodynamic radius close to that predicted from its molecular weight and considering their partially unfolded nature. The presence of dsDNA with different methylation modifications caused a change in the average hydrodynamic radius of the complexes: for MBD, the interaction with dsDNA increased the size of the complex, as expected, but hmCpG-DNA caused the largest increment; for NTD-MBD-ID, the interaction with CpG-DNA and mCpG-DNA slightly reduced the size compacting the overall structure, but hmCpG-DNA caused an increase in the complex molecular size again (Figure 1B). Flexibility and plasticity of the protein and the dsDNA are key factors in this observation, and it is apparent that the interaction of MBD and NTD-MBD-ID with dsDNA resulted in a more compact conformation of dsDNA, corresponding to hmCpG-DNA the higher structural rigidity, as expected [26].

In order to gain more insight on the stabilization effect of dsDNA on MeCP2 constructs, thermal denaturation experiments were carried out by DSC (Figure 1C-D). A key advantage in DSC is that no spectroscopic labels are required and any processes taking place with non-zero enthalpy (as in the unfolding) will be reflected in the experimental trace, providing a global signature. On the contrary, the fluorescence unfolding trace only revealed the behavior of the environment surrounding the single tryptophan in MBD, which may result in overlooking important unfolding steps as commented above.

The detailed analysis of the DSC curves was hampered by several factors, but, mainly: two thermal transitions occurring in NTD-MBD-ID, an enthalpically dominant dsDNA denaturation process, and the

excessive number of unfolding parameters required for characterizing the overall transition in a meaningful manner. However, the thermal unfolding of MBD monitored by DSC was in fair qualitative agreement with the spectroscopic results: larger extent of the stabilization effect for mCpG-DNA. Noticeably, the unfolding of dsDNA occurred at much higher temperature than that of the protein. Therefore, although a DNA-induced stabilization effect on MBD could be observed, a first event consisting of the dissociation of the protein from the dsDNA and subsequent unfolding was observed, and then a second event consisting of the unfolding of the dsDNA in a second event. Very importantly, because of the uncoupling between protein and dsDNA unfolding, no differences in the thermal stability of CpG-DNA, mCpG-DNA and hmCpG-DNA could be observed (Figure S2). This contrasts with the previous observation of a different thermal stability of CpG-DNA, mCpG-DNA, and hmCpG-DNA [35–38,46].

In the case of NTD-MBD-ID, the unfolding of protein and dsDNA occurred within a common temperature range, through partially overlapping transitions. According to the free DNA transition (Figure S2), the first transition should correspond to protein dissociation and unfolding. The thermal unfolding monitored by DSC was again in fair qualitative agreement with the spectroscopic results: larger stabilization extent for mCpG-DNA and intermediate stabilization extent for hmCpG-DNA. Noticeably, the global unfolding enthalpy of the complexes was much larger than that for the free dsDNA, reflecting a major contribution from the protein dissociation process and a lesser contribution from the protein unfolding itself. As commented below, the interaction of NTD-MBD-ID with dsDNA is characterized by a very large, negative enthalpy at 20 °C, and, therefore, the protein-dsDNA complex dissociation occurring at a high temperature would be accompanied by a considerably large, positive enthalpy contributing to the overall unfolding of the complex. In the case of the MBD, because the protein-dsDNA interaction is characterized by a very small enthalpy and the protein-dsDNA complex dissociates in an uncoupled manner at a rather low temperature, the excess enthalpy (with regard to the intrinsic enthalpy of the dsDNA) is negligible. In the case of CpG-DNA and hmCpG-DNA two apparent transitions could be distinguished (with a difference in unfolding temperatures smaller than 5 °C), but for mCpG-DNA that effect was less pronounced, pointing to a higher unfolding cooperativity. In fact, an apparent unfolding cooperativity index can be calculated through the enthalpy ratio (using equation 3), with $\Delta H_{cal}/\Delta H_{vH}$ values of 8.5, 9.5, and 10.7 for mCpG-DNA, CpG-DNA, and hmCpG-DNA, respectively. Therefore, the lower structural cooperativity of the protein-dsDNA complex corresponds to hmCpG-DNA.

hmCpG-DNA interacts with MeCP2 with lower affinity compared to mCpG-DNA

DNA-induced stabilization effects confirmed that NTD-MBD-ID behaves similarly to full-length MeCP2, as reported previously [47]. Therefore, NTD-MBD-ID is an appropriate construct to study MeCP2 interaction with DNA, exhibiting better properties (higher purification yield, higher solubility and lower aggregation/degradation propensity) than full-length MeCP2. The DNA interaction assays were carried out with MBD and NTD-MBD-ID for comparison.

MBD interacted with hmCpG-DNA with a similar, albeit lower, binding affinity to that for CpG-DNA or mCpG-DNA, with a dissociation constant in the submicromolar range (Figure 2A and Tables 2 and S2). Performing calorimetric titrations employing buffers with different ionization enthalpies allowed

estimating the buffer-independent intrinsic binding enthalpy (Figure 2C and Tables 2 and S2). Strikingly, the binding enthalpy for hmCpG-DNA was very favorable (negative contribution to the binding Gibbs energy): –21.6 kcal/mol, compared to a slightly positive, unfavorable enthalpy for CpG-DNA (0.8 kcal/mol) and mCpG-DNA (1.5 kcal/mol), as a remarkable example of enthalpy-entropy compensation. However, the net number of protons n_H exchanged with the bulk solution upon protein-DNA complex formation was smaller: 1.5 protons released for hmCpG-DNA compared to 2.4 and 2.1 protons released upon CpG-DNA and mCpG-DNA binding, respectively. In addition, performing experiments at different temperatures allowed estimating the binding heat capacity $\Delta C_{P,B}$ (Figure 2C and Tables 2 and S2), providing a larger, negative heat capacity value for hmCpG-DNA (–3.1 kcal/K·mol) compared to CpG-DNA (–2.4 kcal/K·mol) and mCpG-DNA (–2.1 kcal/K·mol).

Although the experiments carried out with MBD provided valuable information, as indicated above, the NTD-MBD-ID construction represents a better protein scaffold for the DNA binding studies. As it happened with CpG-DNA and mCpG-DNA, NTD-MBD-ID exhibited two binding sites for hmCpG-DNA with different binding affinities. The hmCpG-DNA high-affinity binding site showed an intermediate binding affinity between those for CpG-DNA and mCpG-DNA, with a dissociation constant in the low nanomolar range (Figure 2B and Tables 2 and S2), whereas the low-affinity binding site showed a binding affinity similar to that for mCpG-DNA. Performing calorimetric titrations employing buffers with different ionization enthalpies allowed estimating the buffer-independent intrinsic binding enthalpies for both DNA binding sites (Figure 2D and Tables 2 and S2). The binding enthalpy for the high-affinity binding site was very favorable, although not as favorable as that for CpG-DNA or mCpG-DNA: –40.6 kcal/mol, compared to –54.6 kcal/mol and –48.4 kcal/mol for CpG-DNA and mCpG-DNA, respectively. Again, enthalpy-entropy compensation was responsible for a similar binding affinity for the three DNA types but different enthalpy-entropy partition. The binding enthalpy for the low-affinity binding site was small and favorable (–2.0 kcal/mol), similar to that for mCpG-DNA (–2.1 kcal/mol) and slightly different to that for CpG-DNA (–7.6 kcal/mol). However, while the net number of protons exchanged with the bulk solution upon hmCpG-DNA binding to the low-affinity site (2.5 protons released) was similar to that for CpG-DNA (2.9 protons released) but larger than that for mCpG-DNA (1.3 protons), the value for the high-affinity site was much larger: 2.3 protons released for hmCpG-DNA compared to 0.1 protons released upon CpG-DNA and mCpG-DNA binding. In addition, performing experiments at different temperatures allowed estimating the binding heat capacity (Figure 2D and Tables 2 and S2) for both binding sites. In the high-affinity binding site, the heat capacity value for hmCpG-DNA (–3.1 kcal/K·mol) was larger than that for CpG-DNA (–2.7 kcal/K·mol) and mCpG-DNA (–2.2 kcal/K·mol); whereas in the low-affinity binding site the heat capacity value was similar to that for CpG-DNA and mCpG-DNA (slightly lower than –1 kcal/K·mol).

hmCpG-DNA interaction is strongly affected by Rett syndrome-associated mutations

In order to shed light into the impact of RTT-associated mutations on the energetics of DNA binding, calorimetric experiments were performed with two mutant constructions: NTD-MBD-ID R106W and NTD-MBD-ID R133C. The comparison of these two mutations is interesting because they correspond to a substitution of arginine by an aminoacid of different size and polarity (tryptophan vs. cysteine)

located at two different protein regions (back of the DNA binding site and within the folding core of MBD vs. DNA binding site surface, respectively) with very different pathological outcomes (severe vs. mild clinical severity score, respectively).

The R106W mutation had a detrimental effect on the binding to hmCpG-DNA in a similar way to the result observed for CpG-DNA and mCpG-DNA, and that deleterious effect was larger in the high-affinity binding site compared to the low-affinity binding site (Figure 3A and Tables 3 and S3). However, the R133C mutation altered the binding of hmCpG-DNA differentially: R133C substitution hardly affected the binding affinity of CpG-DNA and mCpG-DNA, but largely reduced the binding affinity of hmCpG-DNA accompanied by a considerable partial enthalpy-entropy compensation phenomenon (Figure 3B and Tables 3 and S3).

hmCpG-DNA interacts with MeCP2 with distinctive features

The ITC experiments allowed the direct determination of the binding affinity for the interaction of MeCP2 constructions with the different cytosine-modified dsDNAs. Titrations performed employing different buffers and different temperatures allowed the determination of the buffer-independent enthalpic and entropic contributions to the Gibbs energy of binding, the net number of protons exchanged upon complex formation, and the heat capacity of binding (Tables 2 and S2), revealing unexpected peculiarities for the interaction with hmCpG-DNA, pointing to a different mode of interaction: either different network of interactions between dsDNA and protein (including the water molecules trapped within the binding interface) and/or different local/global conformational changes accompanying the interaction constrained by the hydroxymethylated cytosine.

In addition, experiments performed at high ionic strength (buffer supplemented with 150 mM NaCl) unveiled an intriguing feature: the interaction of MBD with any of the three dsDNAs was abolished by high ionic strength and the interaction affinity of NTD-MBD-ID with CpG-DNA and mCpG-DNA was markedly diminished (~1000-fold factor) [45], whereas the interaction affinity of NTD-MBD-ID with hmCpG-DNA in the high-affinity binding site was slightly strengthened (dissociation constant of 0.5 nM, compared to 1.2 nM with low ionic strength) with a significant reduction in enthalpy (–27.6 kcal/mol compared to –66.9 kcal/mol in Tris buffer), as well as in the low-affinity binding site (dissociation constant of 38 nM compared to 71 nM at low ionic strength) (Figure 4).

Although the interaction of hmCpG-DNA with the NTD-MBD-ID low-affinity binding site was characterized by roughly similar binding parameters to those of CpG-DNA and mCpG-DNA interaction, the interaction of hmCpG-DNA with the NTD-MBD-ID high-affinity site showed differential features with regard to the other cytosine-modified dsDNAs. Summarizing:

- Larger, negative binding heat capacity
- Larger, negative net number of protons exchanged (i.e., released) upon binding
- Smaller, less favorable binding enthalpy (and, therefore, smaller, less unfavorable entropic contribution)
- Enhanced affinity at (quasi)physiological ionic strength

As discussed below, the different thermodynamic binding signature for hmCpG-DNA suggests a different interaction mode compared to the other two cytosine-modified dsDNAs.

DISCUSSION

Because the hydroxymethylation of cytosine in DNA is not only a demethylation intermediate, but a stable epigenetic mark, it is important to understand how this modification affects the interaction of epigenetic readers, such as MeCP2, with DNA and its consequences on the subsequent downstream regulatory and functional events.

Here we have studied the interaction of symmetrically hydroxymethylated dsDNA with MeCP2, compared to symmetrically methylated or unmethylated dsDNA. The relevance of that comparison lies in the fact that the majority of 5mC occurs at CpG dinucleotide motifs in a symmetric manner. In addition, dominant levels of symmetric 5fC among total 5fC sites during in vitro TET-dependent 5mC oxidation have been reported, suggesting that 5hmC usually appears in a symmetric manner [48].

The binding affinity of hmCpG-DNA interacting with MBD and NTD-MBD-ID was rather similar to that of CpG-DNA and mCpG-DNA, but major differences were observed in the enthalpic and entropic contributions to the Gibbs energy of binding: the interaction of hmCpG-DNA was enthalpically much less favorable and entropically much less unfavorable (Tables 2 and S2). In a previous study employing the MBD and a shorter 19-bp DNA sequence from the same BDNF promoter, a larger binding affinity reduction (13-fold factor) induced by cytosine hydroxymethylation, compared to methylation, was reported; however, the experimental conditions were rather different (experiments done at pH 6 and 37 °C) [49]. A larger affinity reduction was experimentally determined through fluorescence polarization assays using MBD: K_d of 0.01 μM and 0.26 μM for 5hmC and 5mC, respectively [50]. But in this case, not only the experimental conditions but also the dsDNA sequence was different. As an additional piece of information, the isolated ID was able to interact with hmCpG-DNA with similar affinity to that of CpG-DNA or mCpG-DNA (Figure S3) [47]; still, the binding affinity was slightly lower (K_d of 0.4 μM for hmCpG-DNA, compared to about 0.15 μM for CpG-DNA or mCpG-DNA binding to ID).

Despite the similarity in binding affinity, the thermal denaturations monitored by intrinsic tryptophan fluorescence showed a stabilization effect induced by hmCpG-DNA on MBD and NTD-MBD-ID constructions intermediate between those caused by CpG-DNA and mCpG-DNA (Figure 5 and Tables 1 and S1). The dsDNA-induced excess energetic stabilization effect can be quantitated by the well-known relationship [51]:

$$\Delta\Delta G = \frac{\Delta H(T_m)}{T_m}\Delta T_m \qquad (6)$$

providing again an intermediate value for hmCpG-DNA compared to CpG-DNA and mCpG-DNA (Figure 5). Therefore, MeCP2 seem to interact differently with the three dsDNAs, effectively discriminating among them in terms of overall structural stability of the complex. The results obtained with hmCpG-DNA reinforce the consideration of NTD-MBD-ID as an appropriate proxy for the full-length MeCP2 regarding structural stability and dsDNA interaction features (Figure 5) which is in fair agreement with *in vivo* studies reporting that, when protein levels were reduced in MeCP2-truncated mice, Rett-like symptoms were relieved to some extent with the MBD-ID construction [52].

The thermal denaturations revealed that the cytosine modification also affects the unfolding cooperativity. A single unfolding transition was observed spectroscopically in all complexes (MBD and NTD-MBD-ID bound to any dsDNA); however, hmCpG-DNA decouples the unfolding of the two dsDNA

binding sites in full-length MeCP2, because two unfolding transitions had to be considered in the analysis. This suggests that the local conformation of the DNA binding sites and/or the global protein arrangement and interdomain contacts may be somewhat different in those complexes. In fact, considerable molecular size differences for NTD-MBD-ID in the absence and presence of DNA have been determined (Figure 1B). This is a reflection of the plasticity and structural adaptability of modular IDPs (such as MeCP2) to minor changes in the interacting partners, as well as the relative flexibility induced on the dsDNA by the different cytosine modifications.

The lower dsDNA binding affinity of MBD compared to NTD-MBD-ID triggered a smaller stabilization effect, as observed by spectroscopic and calorimetric thermal denaturations. The calorimetric traces provided more information, since not only changes within the environment near the tryptophan are monitored, but any other processes taking place (for example, the unfolding of the dsDNA). In the case of MBD complexes, non-overlapping transitions for protein and dsDNA could be distinguished, from which the dsDNA-induced stabilization effect on the protein (reflecting the coupling between complex dissociation and protein unfolding) could be estimated. The three different dsDNAs showed similar unfolding traces (Figures 1 and S2), indicating that methylation and hydroxymethylation did not alter the stability of the 45-bp dsDNA, contrary to reported 5hmC-induced destabilization and 5mC-induced stabilization in shorter DNA sequences [35–38,46]. Nevertheless, those methylation-dependent differences were experimentally observed in the context of a 12-bp dsDNA fragment (Dickerson dodecamer: $(CGCGAATTCGCG)_2$), whereas in this work we have employed a much longer 45-bp dsDNA fragment, where the effect of the symmetrically modified cytosine may get diluted.

In the case of the NTD-MBD-ID complexes, the protein and the dsDNA transitions overlapped (reflecting the coupling between complex dissociation and protein/dsDNA unfolding), from which the stabilization effect on the protein was more difficult to estimate, but still in fair qualitative agreement with the spectroscopic experiments. A very large global unfolding enthalpy was observed, and, because the intrinsic unfolding enthalpy of the protein is rather small [53], the major contribution to the overall unfolding enthalpy should stem from the protein-dsDNA dissociation (interactions broken plus any accompanying conformational changes). In fact, the large negative protein-dsDNA binding enthalpy at 20 °C (about –50 kcal/mol) together with a large negative binding heat capacity (about –3 kcal/mol) for the high-affinity binding site, plus the additional interaction and stabilization from the low-affinity binding site, would yield a very large, positive dissociation enthalpy for the protein-dsDNA complex at high temperature.

Despite the similarity in stability against thermal denaturation for the three types of dsDNA, pointing to a similar unfolding propensity for transcription, according to the DLS measurements the complex with the hmCpG-DNA showed the larger apparent hydrodynamic radius when forming a complex with MeCP2, suggesting a lower circularization propensity and compactness in relation with a more accessible chromatin expected *in vivo*.

Two relevant crystallographic structures for a MeCP2-dsDNA complex are available concerning our study: MBD bound to symmetrically methylated dsDNA (PDB 3c2i) [54] and hemi-hydroxymethylated dsDNA bound to MBD (PDB 6yww) [55]. Despite some limitations (symmetric and non-symmetric cytosine modification, and DNAs with different sequence), valuable information can be obtained from

their comparison. The key interactions involving the main cytosine in each complex are depicted in Figure 6: with adjacent nucleobases, R133, S134, and hydrogen-bonded water molecules for the mCpG-DNA complex; and with adjacent nucleobases, R133, K135, and hydrogen-bonded water molecules for the hmCpG-DNA complex. Substantial dissimilarities involving the main cytosine interactions can be noticed. In the mCpG-DNA complex there are five water molecules establishing eight hydrogen bonds, and the 5mC is establishing a hydrogen bond with a water molecule through the nitrogen-4, and other hydrogen bonds with adjacent nucleobases; interestingly, R133 is not interacting directly with 5mC. However, in the hmCpG-DNA complex there are only three water molecules establishing three hydrogen bonds, and the 5hmC is establishing a hydrogen bond with a water molecule and two hydrogen-bonds with R133 through the hydroxyl group, and other hydrogen bonds with adjacent nucleobases. The reduction in the number of hydrogen-bonds with water molecules in the case of hmCpG-DNA can be related to the less favorable binding enthalpy compared to that for mCpG-DNA (Tables 2 and S2). In addition, the larger binding heat capacity for hmCpG-DNA may be explained considering those structural differences: the reduction in the number of water molecules trapped within the protein-dsDNA binding interface suggest a larger dehydration upon hmCpG-DNA binding, resulting in a more negative binding heat capacity and a less unfavorable global entropy (because of a more favorable desolvation entropy) (Tables 2 and S2). Moreover, the burial of the hydroxyl group in 5hmC upon binding would further contribute to a more negative binding heat capacity, as concluded from the structural parameterization of the folding and binding energetics: hydroxyl groups behave similarly to non-polar groups regarding the binding heat capacity [43,56].

One of the more striking features of the interaction of hmCpG-DNA with NTD-MBD-ID is the lack of sensitivity to the ionic strength, whereas CpG-DNA and mCpG-DNA experienced a dramatic affinity loss (1000-fold reduction) at high ionic strength [45]. Considering the linkage relationship:

$$\frac{\partial \ln K_B}{\partial \ln[X]} = \Delta n_X \qquad (6)$$

where K_B (=1/K_d) is the association constant for the binding of a primary ligand (e.g., dsDNA), X is a secondary ligand (e.g., salt ion), and Δn_X is the difference in the saturation fraction for the ligand X between the protein complex and the free protein and free primary ligand [57]. Thus, Δn_X is the net number of ligand X molecules released to (if negative) or captured from (if positive) the bulk solution upon complex formation with the primary ligand. These results indicate that there is an almost zero net number of salt ions released to the bulk solution upon protein-DNA complex formation in the case of hmCpG-DNA, but at least three salt ions are released upon complex formation with CpG-DNA or mCpG-DNA. This is another piece of evidence supporting the different mode of interaction of hmCpG-DNA compared to the other dsDNAs.

As expected, the larger changes in the thermodynamic binding profile caused by the RTT-associated MBD mutations were observed in the high-affinity binding site of NTD-MBD-ID, which is the site located in the MBD. The main effect of R106W is the reduction in affinity of the high-affinity binding site for all types of dsDNA (CpG-DNA, mCpG-DNA, and hmCpG-DNA), and the main effect of R133C is the reduction in affinity of the high-affinity binding site specifically for hmCpG-DNA. This is in agreement with the structural information: 5hmC is coordinated with R133 through two hydrogen bonds, which are absent for 5mC (Figure 6). Therefore, the substitution of arginine by cysteine would have a greater

impact for 5hmC than for 5mC. These results indicate that R106W, a substitution of a polar residue by a hydrophobic residue inside the MBD core far from the DNA binding site that distorts the global architecture of the MBD (ΔT_m = –3.1 °C, and $\Delta\Delta H(T_m)$ = –14 kcal/mol, corresponding to more than 20% reduction, –0.5 kcal/mol, in its conformational stabilization energy) [58], affected the interaction with dsDNA irrespective of the cytosine modification. In fact, a similar 20-30-fold reduction in binding affinity induced by the R106W was observed for the three dsDNA types, accompanied by a remarkable reduction in enthalpic interactions ($\Delta\Delta H$ = +60-70 kcal/mol), partially counterbalanced by an improved (more favorable) entropic contribution to the Gibbs energy of binding (Tables 3 and S3). However, R133C, a substitution of a polar residue by another polar residue within the dsDNA binding site that hardly alters MBD conformation (ΔT_m = –0.4 °C, and $\Delta\Delta H(T_m)$ = –4 kcal/mol, corresponding to less than 2% reduction, –0.04 kcal/mol, in its conformational stabilization energy) [58], affected specifically the interaction with hmCpG-DNA, as envisaged from the crystallographic structure [55]. The reduction of affinity caused by R133C results from a less favorable enthalpy ($\Delta\Delta H$ = +27 kcal/mol) and a less unfavorable entropic contribution ($-T\Delta\Delta S$ = –25 kcal/mol) (Tables 3 and S3). The alterations induced by these two mutations (local/global effects, structural stability changes, cytosine modification specificity) correlate with their clinical severity score: R106W (global structural distortion and cytosine-methylation independent effect) is associated with a phenotypic score more severe than that of R133C (local structural distortion and 5hmC-specific effect).

The hydroxylation of methylated cytosines in DNA, albeit a minor modification, is a stable epigenetic mark with important functional and structural consequences. In particular, as shown here, MeCP2 interacts with hmCpG-DNA in a different manner compared to CpG-DNA or mCpG-DNA. Very likely, the different interaction mode could be reflected in a different conformation of MeCP2 within the complex, with different spatial interdomain organization of the domains and/or local intradomain arrangements. This would affect the architectural configuration of the chromatin and further interactions with transcription machinery elements.

RTT-associated MBD mutations may affect MeCP2 structure and function in different ways: distorting overall or local structure in folded regions, constraining structural propensities in disordered regions, and hampering dsDNA interaction in a general fashion or specifically affecting cytosine-modified dsDNA. In particular, as shown here, R106W mutation affects globally the structure of the MBD folded core and diminish the binding affinity of dsDNA, whereas R133C mutation exclusively reduces the binding affinity of hydroxymethylated dsDNA. Thus, some RTT mutations (e.g., R106W) may elicit a global deleterious effect, irrespective of the methylation level of DNA, the methylation/hydroxymethylation ratio, or the cellular developmental stage; but other mutations (e.g., R133C) may cause an impact which is dependent on the presence of certain DNA modifications and the cellular/tissue/organism developmental stage that is correlated with regulated levels of those specific DNA modifications (e.g., 5hmC levels raise during development as neurons mature from progenitor cells). This might contribute to explaining the differences in phenotypic severity and disease onset time linked to different RTT mutations.

From a translational point of view, this insight on Rett-associated mutations effect on DNA binding that we propose in this work might open new and more rational therapeutical pathways. In the case of

patients with R133C mutation, where a specific decrease of the affinity for hmCpG-DNA has been measured, strategic transient targeting the TET system could lead to a regularization of MeCP2-binding to CpG islands in the brain by lowering the 5hmC/5mC ratio; nevertheless, the global effect of this approach should be assessed, because of the biological role of hmCpG-DNA. Differently, patients with the R106W mutation, where a loss in protein structure leads to an impaired interaction with DNA in a global manner (independently of the epigenetic modification), they would benefit more from the development of molecular chaperones recovering the functional MeCP2 structure and thus rescuing its ability to bind DNA.

The experiments have been performed under non-physiological conditions, with recombinant protein and naked dsDNA, for which the extrapolation of absolute values of the properties (e.g., binding affinities or conformational stabilities) to physiological conditions is not straightforward. However, our goal was to accomplish a comprehensive biophysical characterization of the effect of the 5hmC compared to 5mC or 5C, and the observed differences may be relevant physiologically.

It has been reported previously that the different N-terminal domains in e1 and e2 isoforms modulate many aspects in MeCP2: dsDNA binding affinity, structural stability, turn-over rate, binding dynamics, response to neuronal depolarization, circadian oscillations, interacting protein partners, and specifically regulated target genes [59]. Because our aim was to compare with previous results, the experiments reported in this work have been performed with protein constructs corresponding to MeCP2 isoform e2. Additional structural studies employing the NTD-MBD-ID construction, as a better scaffold than MBD for dsDNA interactions, introducing RTT-associated mutations and differently modified cytosines are required in order to get further insight at molecular/atomic detail into still controversial aspects of MeCP2 function, such as the ability to discriminate between different cytosine modifications and its impact on MeCP2 and dsDNA conformation, or the structural details on the concurrent molecular recognition of dsDNA molecules by MBD and ID.

DATA AVAILABILITY

All data produced or analyzed for this study are available from the corresponding author upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at XXX online.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

David Ortega-Alarcon: data curation; formal analysis; investigation; software; validation; visualization; writing – review & editing

Rafael Claveria-Gimeno: data curation; formal analysis; investigation; software; validation; visualization; writing – review & editing

Sonia Vega: investigation; writing – review & editing

Olga C. Jorge-Torres: investigation; validation; writing – review & editing
Manel Esteller: conceptualization; resources; writing – review & editing
Olga Abian: conceptualization; formal analysis; funding acquisition; methodology; resources; software; supervision; project administration; writing – original draft preparation; writing – review & editing
Adrian Velazquez-Campoy: conceptualization; formal analysis; funding acquisition; methodology; resources; software; supervision; project administration; writing – original draft preparation; writing – review & editing

DECLARATION OF COMPETING INTEREST

None.

ACKNOWLEDGEMENT

This work was supported by the Spanish Ministry of Economy and Competitiveness and European ERDF Funds (MCIU/AEI/FEDER, EU) [BFU2016-78232-P to A.V.C., BES-2017-080739 to D.O.A.]; Miguel Servet Program from Instituto de Salud Carlos III [CPII13/00017 to OA]; Fondo de Investigaciones Sanitarias from Instituto de Salud Carlos III and European Union (ERDF/ESF, “Investing in your future”) [PI18/00349 and PI21/00394 to O.A.]; Diputación General de Aragón [Protein Targets and Bioactive Compounds Group E45_20R to A.V.C., and Digestive Pathology Group B25_20R to O.A.]; and the Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd).

REFERENCES

[1] J. Monod, Chance and Necessity: Essay on the Natural Philosophy of Modern Biology, Penguin Books, UK, 1977.

[2] L. Lercher, M.A. McDonough, A.H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown, C.J. Schofield, Structural insights into how 5-hydroxymethylation influences transcription factor binding, Chem. Commun. (Camb.) 50 (2014) 1794–1796.

[3] S. Kriaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, Science 324 (2009) 929–930.

[4] S. Ito, L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, C. He, Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science 333 (2011) 1300–1303.

[5] Y.F. He, B.Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.X. Song, K. Zhang, C. He, G.L. Xu, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, Science 333 (2001) 1303–1307.

[6] G. Wyatt, S. Cohen, A new pyrimidine base from bacteriophage nucleic acids. Nature 170 (1952) 1072–1073.

[7] N.W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid, Biochem. J. 126 (1972) 781–790.

[8] L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, Y. Wang, J. Xie, Y. Zhang, C. Song, M. Yu, X. Liu, P. Zhu, X. Li, Y. Hou, H. Guo, X. Wu, C. He, R. Li, F. Tang, J. Qiao, Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain, Genome Biol. 15 (2014) R49.

[9] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, Quantification of the sixth DNA base hydroxymethylcytosine in the brain, Angew. Chem. Int. Ed. Engl. 49 (2010) 5375–5377.

[10] S.G. Jin, Y. Jiang, R. Qiu, T.A. Rauch, Y. Wang, G. Schackert, D. Krex, Q. Lu, G.P. Pfeifer, 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations, Cancer Res. 71 (2011) 7360–7365.

[11] P. Tognini, D. Napoli, T. Pizzorusso, Dynamic DNA methylation in the brain: a new epigenetic mark for experience-dependent plasticity, Front. Cell. Neurosci. 9 (2015) 331.

[12] M.A. Hahn, R. Qiu, X. Wu, A.X. Li, H. Zhang, J. Wang, J. Jui, S.G. Jin, Y. Jiang, G.P. Pfeifer, Q. Lu, Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis, Cell Rep. 3 (2013) 291–300.

[13] T. Khare, S. Pai, K. Koncevicius, M. Pal, E. Kriukiene, Z. Liutkeviciute, M. Irimia, P. Jia, C. Ptak, M. Xia, R. Tice, M. Tochigi, S. Moréra, A. Nazarians, D. Belsham, A.H. Wong, B.J. Blencowe, S.C. Wang, P. Kapranov, R. Kustra, V. Labrie, S. Klimasauskas, A. Petronis, 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary, Nat. Struct. Mol. Biol. 19 (2012) 1037–1043.

[14] C.G. Spruijt, F. Gnerlich, A.H. Smits, T. Pfaffeneder, P.W. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H.C. Eberl, A. Mensinga, A.B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, Cell 152 (2013) 1146–1159.

[15] H. Takai, K. Masuda, T. Sato, Y. Sakaguchi, T. Suzuki, T. Suzuki, R. Koyama-Nasu, Y. Nasu-Nishimura, Y. Katou, H. Ogawa, Y. Morishita, H. Kozuka-Hata, M. Oyama, T. Todo, Y. Ino, A. Mukasa, N. Saito, C. Toyoshima, K. Shirahige, T. Akiyama, 5-hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex, Cell Rep. 9 (2014) 48–60.

[16] M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system, Cell 151 (2012) 1417-1430.

[17] J. Ausio, A. Martinez de Paz, M. Esteller, MeCP2: The long trip from a chromatin protein to neurological disorders, Trends Mol. Med. 20 (2014) 487–498.

[18] N.L. Adkins, P.T. Georgel, MeCP2: Structure and function, Biochem. Cell Biol. 89 (2011) 1–11

[19] X. Nan, F.J. Campoy, A. Bird, MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin, Cell 88 (1997) 471–481.

[20] J.C. Hansen, R.P. Ghosh, C.L. Woodcock, Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin, IUBMB Life 62 (2010) 732–738.

[21] R.E. Amir, I.B. Van den Veyver, M. Wan, C.Q. Tran, U. Francke, H.Y. Zoghbi, Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2, Nat. Genet. 23 (1999) 185–188.

[22] J.L. Neul, The relationship of Rett syndrome and MECP2 disorders to autism, Dialogues Clin. Neurosci. 14 (2012) 253–262.

[23] A.J. Sandweiss, V.L. Brandt, H.Y. Zoghbi, Advances in understanding of Rett syndrome and MECP2 duplication syndrome: Prospects for future therapies, Lancet Neurol. 19 (2020) 689-698.

[24] M. Bostick, J.K. Kim, P.O. Estève, A. Clark, S. Pradhan, S.E. Jacobsen, UHRF1 plays a role in maintaining DNA methylation in mammalian cells, Science 317 (2007) 1760–1764.

[25] V. Valinluck, H.-H. Tsai, D.K. Rogstad. A. Burdzy, A. Bird, L.C. Sowers, Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2), Nucleic Acid Res. 32 (2004) 4100–4108.

[26] F. Battistini, P.D. Dans, M. Terrazas, C.L. Castellazzi, G. Portella, M. Labrador, N. Villegas, I. Brun-Heath, C. González, M. Orozco, The impact of the HydroxyMethylCytosine epigenetic signature on DNA structure and function, PLoS Comput. Biol. 17 (2021) e1009547.

[27] R.S. Illingworth, A.P. Bird, CpG islands--'a rough guide', FEBS Lett. 583 (2009) 1713–1720.

[28] B.M. Colquitt, W.E. Allen, G. Barnea, S. Lomvardas, Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity, Proc. Natl. Acad. Sci. USA 110 (2013) 14682–14687.

[29] M. Münzel, D. Globisch, T. Carell, 5-hydroxymethylcytosine, the sixth base of the genome, Angew. Chem. Int. Ed. Engl. 50 (2011) 6460–6468.

[30] J. Boyes, A. Bird, DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein, Cell 64 (1991) 1123–1134.

[31] S.-G. Jin, S. Kadam, G.P. Pfeifer, Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine, Nucleic Acid Res. 38 (2010) e125.

[32] P.A. Wade, Methyl CpG-binding proteins and transcriptional repression, Bioessays, 23 (2001) 1131–1137.

[33] M.A. Karymov, M. Tomschik, S.H. Leuba, P. Caiafa, J. Zlatanova, DNA methylation-dependent chromatin fiber compaction in vivo and in vitro: Requirement for linker histone, FASEB J. 15 (2001) 2631–2641.

[34] P.M. Severin, X. Zou, K. Schulten, H.E. Gaub, Effects of cytosine hydroxymethylation on DNA strand separation, Biophys. J. 104 (2013) 208–215.

[35] A. Thalhammer, A.S. Hansen, A.H. El-Sagheer, T. Brown, C.J. Schofield, Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation, Chem. Commun. 47 (2011) 5325–5327.

[36] C.M. Lopez, A.J. Lloyd, K. Leonard, M.J. Wilkinson, Differential effect of three base modifications on DNA thermostability revealed by high resolution melting, Anal. Chem. 84 (2012) 7336–7342.

[37] C.M. Rodriguez Lopez, B. Guzman Asenjo, A.J. Lloyd, M.J. Wilkinson, Direct detection and quantification of methylation in nucleic acid sequences using high-resolution melting analysis, Anal. Chem. 82 (2010) 9100–9108.

[38] M. Wanunu, D. Cohen-Karni, R.R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M.L. Klein, M. Drndic, Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules, J. Am. Chem. Soc. 133 (2010) 486–492.

[39] A. Kozlenkov, J. Li, P. Apontes, Y.L. Hurd, W.M. Byne, E.V. Koonin, M. Wegner, E.A. Mukamel, S. Dracheva, A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons, Sci. Adv. 4 (2018) eaau6190.

[40] S. Vega, O. Abian, A. Velazquez-Campoy, A unified framework based on the binding polynomial for characterizing biological systems by isothermal titration calorimetry, Methods 76 (2015) 99–115.

[41] E. Freire, A. Schön, A. Velazquez-Campoy, Isothermal titration calorimetry: General formalism using binding polynomials, Methods Enzymol. 455 (2009) 127–155.

[42] H.J. Hinz, D.D.F. Shiao, J.M. Sturtevant, Calorimetric investigation of inhibitor binding to rabbit muscle aldolase, Biochemistry 10 (1971) 1347–1352.

[43] J. Gomez, E. Freire, Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin, J. Mol. Biol. 252 (1995) 337–350.

[44] R.N. Goldberg, N. Kishore, R.M. Lennen, Thermodynamic quantities for the ionization reactions of buffers, J. Phys. Chem. Ref. Data 31 (2002) 231–370.

[45] R. Claveria-Gimeno, P.M. Lanuza, I. Morales-Chueca, O.C. Jorge-Torres, S. Vega, O. Abian, M. Esteller, A. Velazquez-Campoy, The intervening domain from MeCP2 enhances the DNA affinity of the methyl binding domain and provides an independent DNA interaction site, Sci. Rep. 7 (2017) 41635.

[46] Y.P. Bhavsar-Jog, E. Van Dornshuld, T.A. Brooks, G.S. Tschumper, R.M. Wadkins, Epigenetic modification, dehydration, and molecular crowding effects on the thermodynamics of i-motif structure formation from C-rich DNA, Biochemistry 53 (2014) 1586–1594.

[47] D. Ortega-Alarcon, R. Claveria-Gimeno, S. Vega, O.C. Jorge-Torres, M. Esteller, O. Abian, A. Velazquez-Campoy, Influence of the disordered domain structure of MeCP2 on its structural stability and dsDNA interaction, Int. J. Biol. Macromol. 175 (2021) 58–66.

[48] L. Xu, Y.-C. Chen, J. Chong, A. Fin, L.S. McCoy, J. Xu, C. Zhang, D. Wang, Pyrene-based quantitative detection of the 5-formylcytosine loci symmetry in the CpG duplex content during TET-dependent demethylation, Angew. Chem. Int. Ed. Engl. 53 (2014) 11223–11227.

[49] M.J. Sperlazza, S.M. Bilinovich, L.M. Sinanan, FR. Javier, D.C.Jr. Williams, Structural basis of MeCP2 distribution on non-CpG methylated and hydroxymethylated DNA, J. Mol. Biol. 429 (2017) 1581–1594.

[50] H. Hashimoto, Y. Liu, A.K. Upadhyay, Y. Chang, S.B. Howerton, P.M. Vertino, X. Zhang, X. Cheng, Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, Nucleic Acids Res. 40 (2012) 4841–4849.

[51] W.J. Becktel, J.A. Schellman, Protein stability curves, Biopolymers 26 (1987) 1859–1877.

[52] R. Tillotson, J. Selfridge, M.V. Koerner, K.K.E. Gadalla, J. Guy, D. De Sousa, R.D. Hector, S.R. Cobb, A. Bird, Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. Nature 550 (2017) 398–401.

[53] D. Ortega-Alarcon, R. Claveria-Gimeno, S. Vega, O.C. Jorge-Torres, M. Esteller, O. Abian, A. Velazquez-Campoy, Stabilization effect of intrinsically disordered regions on multidomain proteins: The case of the methyl-CpG protein 2, MeCP2, Biomolecules 11 (2021) 1216.

[54] K.L. Ho, I.W. McNae, L. Schmiedeberg, R.J. Klose, A.P. Bird, M.D. Walkinshaw, MeCP2 binding to DNA depends upon hydration at methyl-CpG, Mol. Cell 29 (2008) 525–531.

[55] A. Ibrahim, C. Papin, K. Mohideen-Abdul, S. Le Gras, I. Stoll, C. Bronner, S. Dimitrov, B.P. Klaholz, A. Hamiche, MeCP2 is a microsatellite binding protein that protects CA repeats from nucleosome invasion, Science, 372 (2021) eabd5581.

[56] E. Freire, Thermodynamics of protein folding and molecular recognition, Pure & Appl. Chem. 89 (1997) 2245–2252.

[57] J.Jr. Wyman, Linked functions and reciprocal effects in hemoglobin: A second look. Adv. Protein Chem. 19 (1964) 223–286.

[58] D. Ortega-Alarcon, R. Claveria-Gimeno, S. Vega, O.C. Jorge-Torres, M. Esteller, O. Abian, A. Velazquez-Campoy, Molecular context-dependent effects induced by Rett syndrome-associated mutations in MeCP2, Biomolecules 10 (2020) 1533.

[59] A. Martínez de Paz, L. Khajavi, H. Martin, R. Claveria-Gimeno, S.T. Dieck, M.S. Cheema, J.V. Sanchez-Mut, M.M. Moksa, A.Carles, N.I. Brodie, T.I. Sheikh, M.E. Freeman, E.V. Petrotchenko, C.H. Borchers, E.M. Schuman, M. Zytnicki, A. Velazquez-Campoy, O. Abian, M. Hirst, M. Esteller, J.B. Vincent, C.E. Malnou, J. Ausió, MeCP2-E1 isoform is a dynamically expressed, weakly DNA-bound protein with different protein and DNA interactions compared to MeCP2-E2, Epigenetics Chromatin, 12 (2019) 63.

TABLE AND FIGURES LEGENDS

Table 1. Thermal stability of MeCP2 MBD, NTD-MBD-ID and full-length MeCP2 constructs (5 μM) in the presence of hydroxymethylated (hmCpG-) dsDNA (10 μM) at pH 7.

	T_m (°C)	$\Delta H(T_m)$ (kcal/mol)
MBD	51.6 ± 0.3	41 ± 2
NTD-MBD-ID	68.4 ± 0.2	63 ± 4
NTD-MBD-ID-TRD-CTDα-CTDβ	57.5 ± 0.2 67.6 ± 0.2	28 ± 1 74 ± 4

Table 2. Buffer-independent binding parameters for the interaction of the different MeCP2 variants with hydroxymethylated (hmCpG-) dsDNA at pH 7 and 20°C.

	K_d (nM)	ΔG_B (kcal/mol)	ΔH_B (kcal/mol)	$-T\Delta S_B$ (kcal/mol)	ΔC_{PB} (kcal/K·mol)	Δn_H
MBD	630	–8.3	–21.6	13.3	–3.1	–1.5
NTD-MBD-ID	1.2	–12.0	–40.6	28.6	–3.3	–2.3
	71	–9.6	–2.0	–7.6	–1.0	–2.5

Two dissociation constants K_d (for high-affinity and low-affinity sites) are provided for NTD-MBD-ID because it has two dsDNA binding sites.

Relative error in K_d is 15%, absolute error in ΔG_B is 0.1 kcal/mol, absolute error in ΔH_B and $-T\Delta S_B$ is 0.5 kcal/mol, error in ΔC_{PB} is 0.4 kcal/K·mol, and absolute error in Δn_H is 0.2.

Table 3. Buffer-independent binding parameters for the interaction of the different Rett syndrome-associated MeCP2 mutants with hydroxymethylated (hmCpG-) dsDNA at pH 7 and 20°C.

	K_d (nM)	ΔG_B (kcal/mol)	ΔH_B (kcal/mol)	$-T\Delta S_B$ (kcal/mol)
NTD-MBD-ID	1.2	–12.0	–40.6	28.6
	71	–9.6	–2.0	–7.6
NTD-MBD-ID R106W	21	–10.3	19.1	29.4
	360	–8.6	–15.3	6.7
NTD-MBD-ID R133C	24	–10.2	–13.6	3.4
	110	–9.3	–10.5	1.2

Relative error in K_d is 15%, absolute error in ΔG_B is 0.1 kcal/mol, absolute error in ΔH_B and $-T\Delta S_B$ is 0.5 kcal/mol.

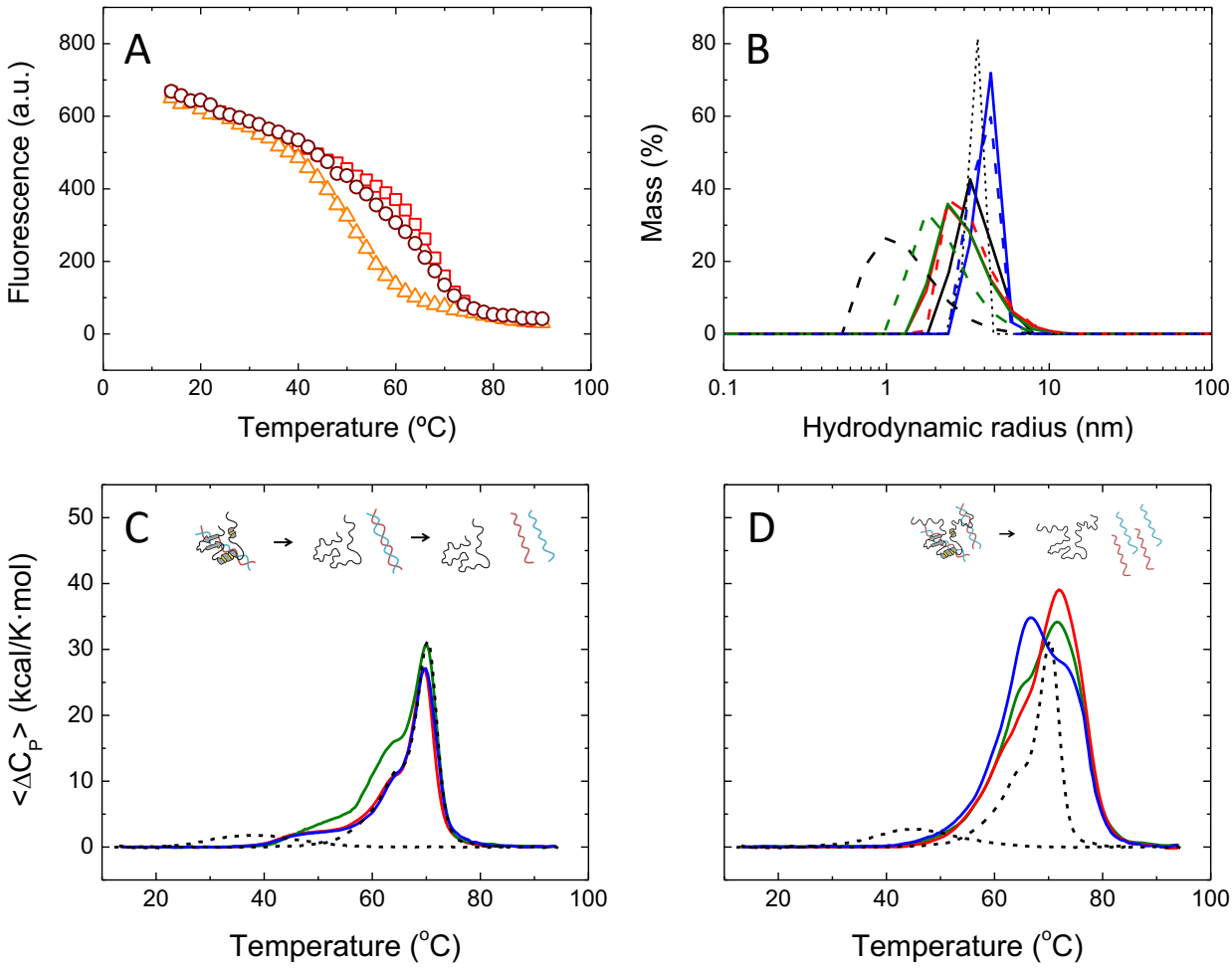

Figure 1. Structural stability monitored by spectroscopy and DSC. (A) Fluorescence thermal denaturations for MBD (orange triangles), NTD-MBD-ID (red squares), and NTD-MBD-ID-TRD-CTDα-CTDβ or full-length MeCP2 (dark red circles) at pH 7 at a concentration of 5 μM in the presence of 10 μM hmCpG-DNA. Unfolding traces could be fitted (continuous lines) employing a single transition unfolding model according to equation 1 (MBD and NTD-MBD-ID) or a two-transition unfolding model according to equation 2 (full-length MeCP2). (B) Hydrodynamic radius distribution for MBD (dashed lines) and NTD-MBD-ID (continuous lines) interacting with dsDNA: no DNA (black), CpG-DNA (green), mCpG-DNA (red), and hmCpG-DNA (blue). The thin dotted black line corresponds to protein-free dsDNA. (C, D) Calorimetric thermal denaturations of MBD (C) and NTD-MBD-ID (D) at pH 7 at a concentration of 20 μM in absence of DNA (dotted line, low temperature transition), and in the presence of 20 μM CpG-DNA (green), mCpG-DNA (red), and hmCpG-DNA (blue). The thermal denaturation of free CpG-DNA is also shown (dotted line, high temperature transition). A drawing has been included in each DSC plot to illustrate the molecular events taking place driven by temperature, either sequentially or simultaneously, for each protein construct (MBD and NTD-MBD-ID in cartoon representation, and dsDNA in light blue and tile color).

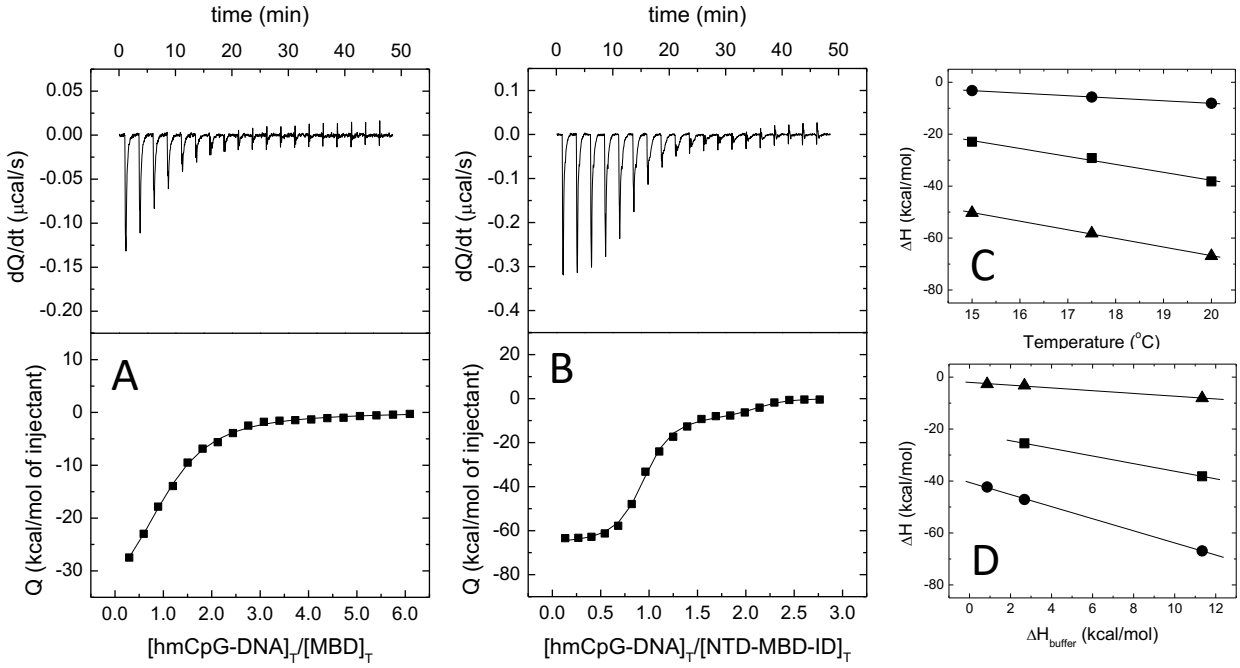

Figure 2. Interaction of MBD and NTD-MBD-ID with hmCpG-DNA by ITC. Interaction of hmCpG-dsDNA with MBD (A) and NTD-MBD-ID (B) assessed by isothermal titration calorimetry (ITC) at pH 7. Upper plots show the thermogram (raw thermal power as a function of time) and the lower plots show the binding isotherm (ligand-normalized heat effects as a function of the molar ratio). Non-linear least-squares analysis using a single-site or a two-site binding model allowed estimating the observed binding affinity and enthalpy (continuous line). (C) The heat capacity of binding for the interaction with the single binding site in MBD (squares) and the two binding sites in NTD-MBD-ID (circles and triangles) was determined by performing experiment at different temperatures and calculating the slope of the linear regression in the enthalpy vs. temperature plot). (D) The buffer-independent binding enthalpy and the net number of protons exchanged for the interaction with the single binding site in MBD (squares) and the two binding sites in NTD-MBD-ID (circles and triangles) was estimated employing equation 5 applying linear regression (intercept with y-axis and slope, respectively).

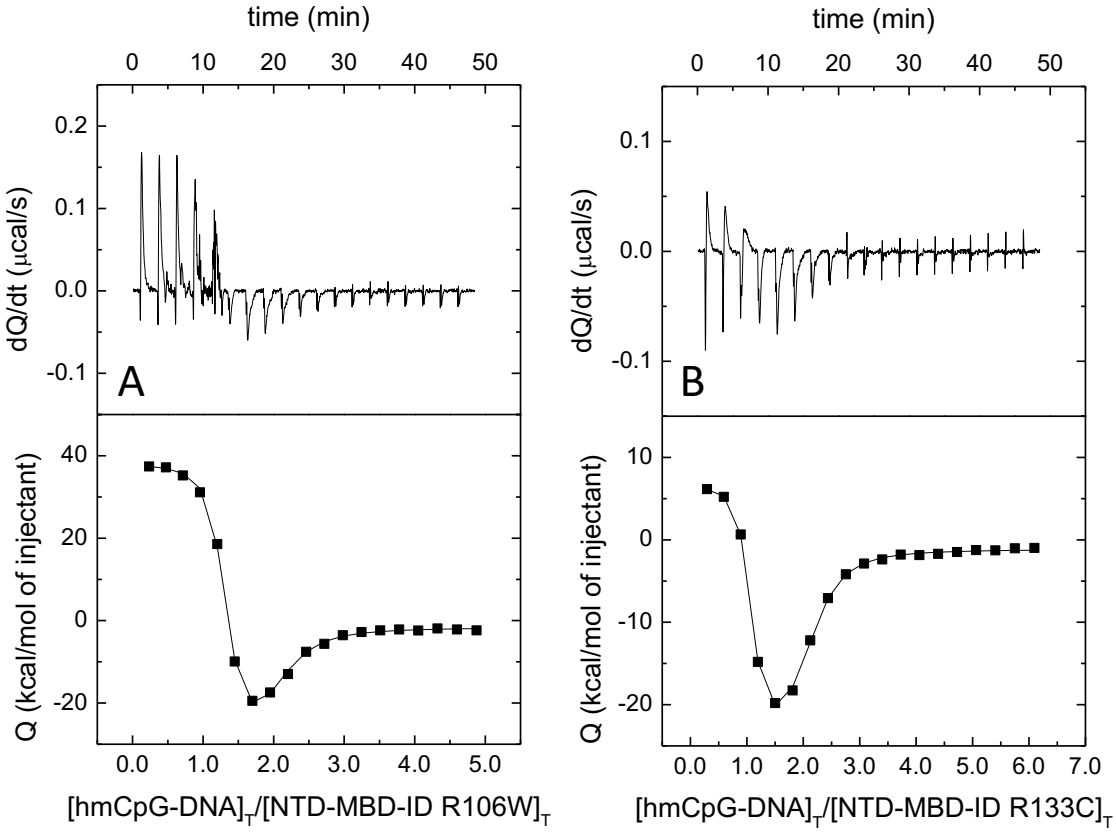

Figure 3. Interaction of NTD-MBD-ID Rett syndrome mutants with hmCpG-dsDNA by ITC. Interaction of hmCpG-dsDNA with NTD-MBD-ID R106W (A) and R133C (B) assessed by isothermal titration calorimetry (ITC) at pH 7. Upper plot shows the thermogram (raw thermal power as a function of time) and the lower plot shows the binding isotherm (ligand-normalized heat effects as a function of the molar ratio). Non-linear least-squares analysis using a two-site binding model allowed estimating the observed binding affinity and enthalpy for the high-affinity and low-affinity binding sites (continuous line).

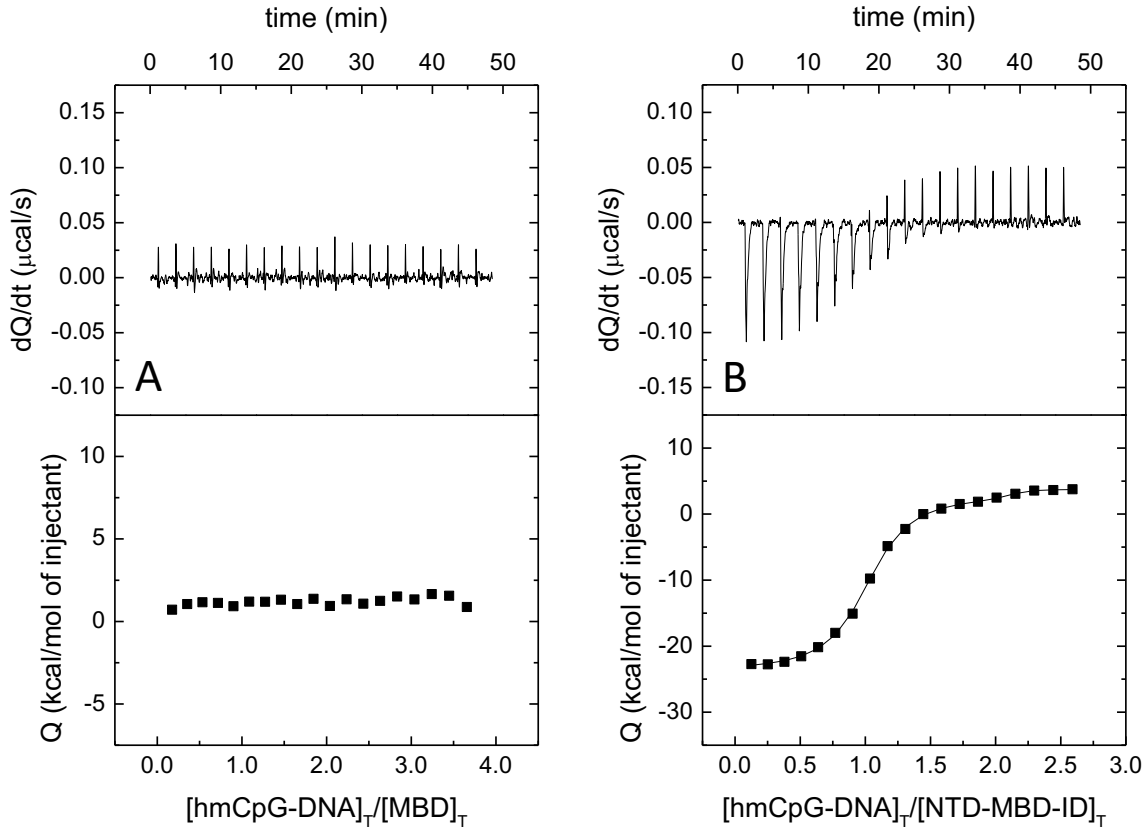

Figure 4. Influence of ionic strength on hmCpG-DNA binding by ITC. Interaction of MBD (A) and NTD-MBD-ID (B) with hmCpG-dsDNA assessed by isothermal titration calorimetry (ITC) at pH 7 and 150 mM NaCl. Upper plots show the thermogram (raw thermal power as a function of time) and the lower plots show the binding isotherm (ligand-normalized heat effects as a function of the molar ratio). Non-linear least-squares analysis using the two-site binding model (for NTD-MBD-ID) allowed estimating the observed binding affinity and enthalpy (continuous lines). The estimated dissociation constants for the high-affinity and low-affinity sites are 0.5 nM and 38 nM, respectively.

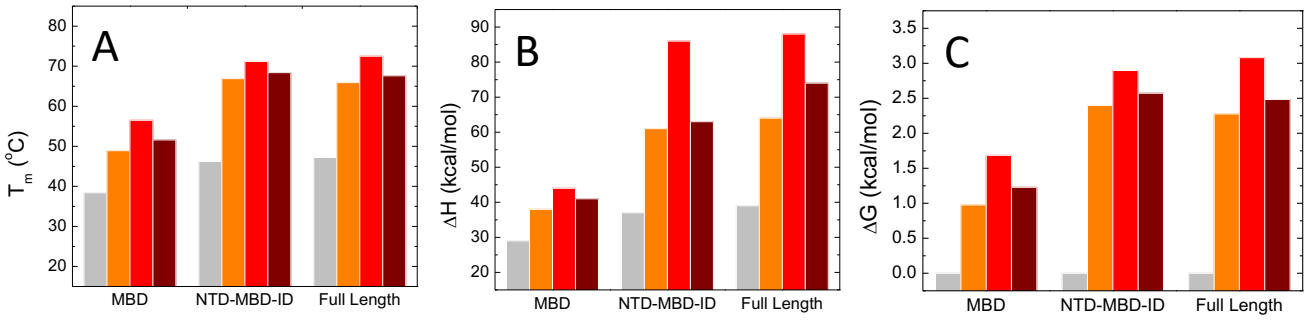

Figure 5. Structural stability parameters for MeCP2 constructions. Unfolding temperature (A), unfolding enthalpy (B), and excess dsDNA-induced Gibbs energy stabilization (C) for MBD, NTD-MBD-ID, and full-length MeCP2 in the absence of dsDNA (grey) and in the presence of CpG-DNA (orange), mCpG-DNA (red), and hmCpG-DNA (dark red).

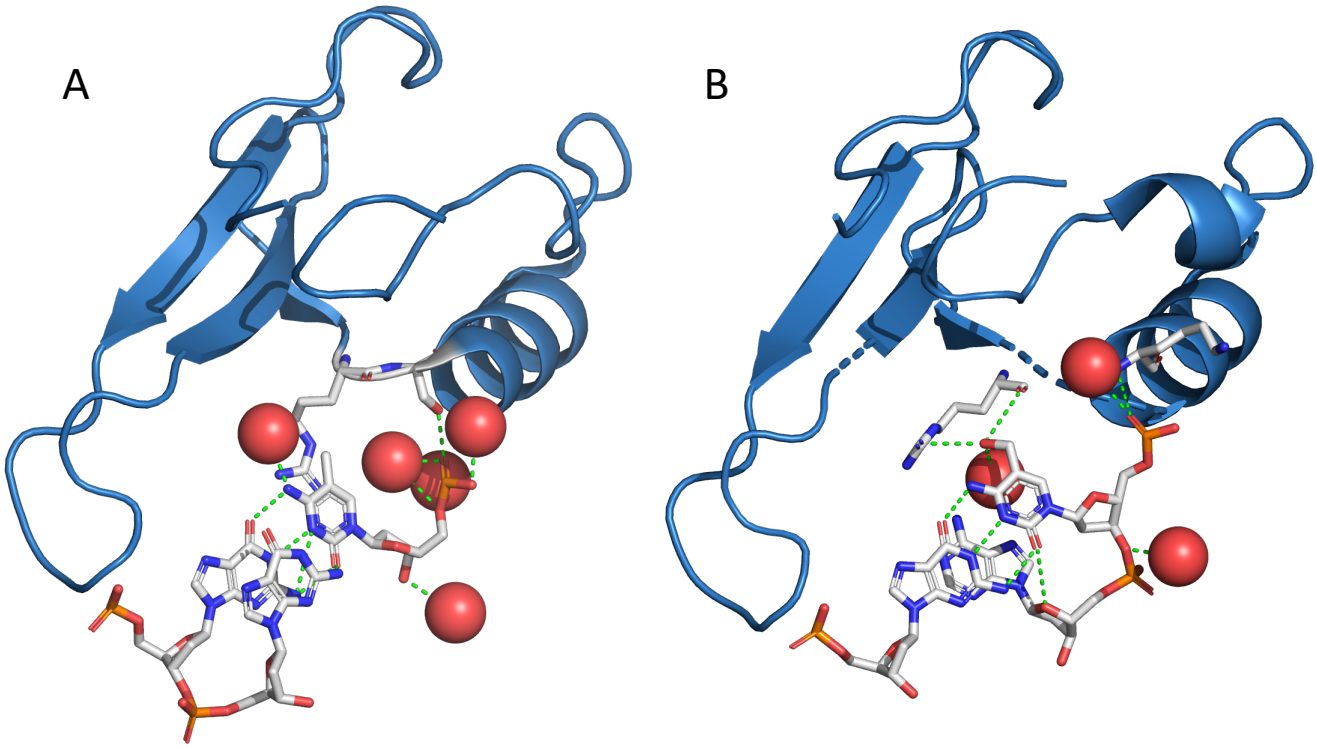

Figure 6. Schematic depiction of the interaction of the interaction of 5mC and 5hmC with MeCP2 MBD. Main interactions between the modified cytosines with MBD residues and water molecules in the available crystallographic structures: PDB 3c2i for mCpG-DNA complex (A), and PDB 6yww for hmCpG-DNA complex (B). Water molecules are shown as red spheres and key MBD residues interacting with 5mC and 5hmC are highlighted. Hydrogen bonds are shown as green dotted segments.

SUPPLEMENTARY DATA

Unexpected thermodynamic signature for the interaction of hydroxymethylated DNA with MeCP2

David Ortega-Alarcon[1], Rafael Claveria-Gimeno[1,2†], Sonia Vega[1], Olga C. Jorge-Torres[3], Manel Esteller[3,4,5,6], Olga Abian[1,2,7,8*] and Adrian Velazquez-Campoy[1,2,3,7,8*]

[1] Institute of Biocomputation and Physics of Complex Systems (BIFI), Joint Units GBsC-CSIC-BIFI and ICVV-CSIC-BIFI, Universidad de Zaragoza, Zaragoza, 50018, Spain
[2] Instituto de Investigación Sanitaria Aragón (IIS Aragón), 50009, Zaragoza, Spain
[3] Josep Carreras Leukaemia Research Institute (IJC), 08916, Badalona, Barcelona, Spain
[4] Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), 28029, Madrid, Spain
[5] Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010, Barcelona, Spain
[6] Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), 08907, l'Hospitalet de Llobregat, Barcelona, Spain
[7] Centro de Investigación Biomédica en Red en el Área Temática de Enfermedades Hepáticas y Digestivas (CIBERehd), 28029, Madrid, Spain
[8] Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, 50009, Zaragoza, Spain

* To whom correspondence should be addressed. Tel: +34 976 762996; Email: adrianvc@unizar.es. Correspondence may also be addressed to Olga Abian. Email: oabifra@unizar.es

Table S1. Thermal stability of MeCP2 MBD, NTD-MBD-ID and full-length MeCP2 constructs (5 μM) in the presence of unmethylated (CpG-), methylated (mCpG-), and hydroxymethylated (hmCpG-) dsDNA (10 μM) at pH 7.

Table 2. Buffer-independent binding parameters for the interaction of the different MeCP2 variants with unmethylated (CpG-), methylated (mCpG-), and hydroxymethylated (hmCpG-) dsDNA at pH 7 and 20°C.

Table 3. Buffer-independent binding parameters for the interaction of the different Rett syndrome-associated MeCP2 mutants with unmethylated (CpG-), methylated (mCpG-), and hydroxymethylated (hmCpG-) dsDNA at pH 7 and 20°C.

Figure S1. MeCP2 sequence colored by domains.

Figure S2. Thermal unfolding of CpG-DNA, mCpG-DNA, and hmCpG-DNA.

Figure S3. Interaction of ID with hmCpG-DNA by ITC.

Table S1. Thermal stability of MeCP2 MBD, NTD-MBD-ID and full-length MeCP2 constructs (5 μM) in the presence of unmethylated (CpG-), methylated (mCpG-), and hydroxymethylated (hmCpG-) dsDNA (10 μM) at pH 7.

		T_m (°C)	$\Delta H(T_m)$ (kcal/mol)
MBD	no DNA[a]	38.4 ± 0.3	29 ± 1
	CpG-dsDNA[a]	48.9 ± 0.3	38 ± 2
	mCpG-dsDNA[a]	56.5 ± 0.3	44 ± 2
	hmCpG-DNA[b]	51.6 ± 0.3	41 ± 2
NTD-MBD-ID	no DNA[a]	46.2 ± 0.2	37 ± 1
	CpG-dsDNA[a]	64.5 ± 0.1	60 ± 2
	mCpG-dsDNA[a]	71.2 ± 0.2	86 ± 4
	hmCpG-dsDNA[b]	68.4 ± 0.2	63 ± 4
NTD-MBD-ID-TRD-CTDα-CTDβ (full-length MeCP2)	no DNA[a]	47.2 ± 0.2	39 ± 1
	CpG-dsDNA[a]	65.9 ± 0.1	64 ± 2
	mCpG-dsDNA[a]	72.5 ± 0.2	88 ± 2
	hmCpG-dsDNA[b]	57.5 ± 0.2	28 ± 1
		67.6 ± 0.2	74 ± 4

[a] Previous work (45).

[b] This work.

Table S2. Buffer-independent binding parameters for the interaction of the different MeCP2 variants with unmethylated (CpG-), methylated (mCpG-), and hydroxymethylated (hmCpG-) dsDNA at pH 7 and 20°C.

		K_d (nM)	ΔG_B (kcal/mol)	ΔH_B (kcal/mol)	$-T\Delta S_B$ (kcal/mol)	ΔC_{PB} (kcal/K·mol)	Δn_H
MBD	CpG-DNA[a]	450	–8.5	0.8	–9.3	–2.3	–2.4
	mCpG-DNA[a]	240	–8.9	1.5	–10.4	–1.9	–2.1
	hmCpG-DNA[b]	630	–8.3	–21.6	13.3	–3.1	–1.5
NTD-MBD-ID	CpG-DNA[a]	1.9	–11.7	–54.6	42.9	–2.7	–0.1
		250	–8.9	–7.6	–1.3	–0.9	–2.9
	mCpG-DNA[a]	0.56	–12.4	–48.4	36.0	–2.2	–0.1
		62	–9.7	–2.1	–7.6	–0.9	–1.3
	hmCpG-DNA[b]	1.2	–12.0	–40.6	28.6	–3.3	–2.3
		71	–9.6	–2.0	–7.6	–1.0	–2.5

Two dissociation constants K_d (for high-affinity and low-affinity sites) are provided for NTD-MBD-ID because it has two dsDNA binding sites.

Relative error in K_d is 15%, absolute error in ΔG_B is 0.1 kcal/mol, absolute error in ΔH_B and $-T\Delta S_B$ is 0.5 kcal/mol, error in ΔC_{PB} is 0.4 kcal/K·mol, and absolute error in Δn_H is 0.2.

[a] Previous work (45).

[b] This work.

Table S3. Buffer-independent binding parameters for the interaction of the different Rett syndrome-associated MeCP2 mutants with unmethylated (CpG-), methylated (mCpG-), and hydroxymethylated (hmCpG-) dsDNA at pH 7 and 20°C.

		K_d (nM)	ΔG_B (kcal/mol)	ΔH_B (kcal/mol)	$-T\Delta S_B$ (kcal/mol)
NTD-MBD-ID	CpG-DNA[a]	1.9	–11.7	–54.6	42.9
		250	–8.9	–7.6	–1.3
	mCpG-DNA[a]	0.56	–12.4	–48.4	36.0
		62	–9.7	–2.1	–7.6
	hmCpG-DNA[b]	1.2	–12.0	–40.6	28.6
		71	–9.6	–2.0	–7.6
NTD-MBD-ID R106W	CpG-DNA[a]	27	–10.2	15.6	–25.8
		320	–8.7	–25.6	16.9
	mCpG-DNA[a]	23	–10.2	22.6	–32.8
		430	–8.5	–16.7	8.2
	hmCpG-DNA[b]	21	–10.3	19.1	29.4
		360	–8.6	–15.3	6.7
NTD-MBD-ID R133C	CpG-DNA[a]	2.5	–11.5	–49.1	37.6
		83	–9.5	–11.8	2.3
	mCpG-DNA[a]	2.1	–11.6	–45.0	33.4
		62	–9.7	–10.0	0.3
	hmCpG-DNA[b]	24	–10.2	–13.6	3.4
		110	–9.3	–10.5	1.2

Relative error in K_d is 15%, absolute error in ΔG_B is 0.1 kcal/mol, absolute error in ΔH_B and $-T\Delta S_B$ is 0.5 kcal/mol.

[a] Previous work (45,56).

[b] This work.

```
1      MVAGMLGLRE EKSEDQDLQG LKDKPLKFKK VKKDKKEEKE GKHEPVQPSA HHSAEPAEAG
61     KAETSEGSGS APAVPEASAS PKQRRSIIRD RGPMYDDPTL PEGWTRKLKQ RKSGRSAGKY
121    DVYLINPQGK AFRSKVELIA YFEKVGDTSL DPNDFDFTVT GRGSPSRREQ KPPKKPKSPK
181    APGTGRGRGR PKGSGTTRPK AATSEGVQVK RVLEKSPGKL LVKMPFQTSP GGKAEGGGAT
241    TSTQVMVIKR PGRKRKAEAD PQAIPKKRGR KPGSVVAAAA AEAKKKAVKE SSIRSVQETV
301    LPIKKRKTRE TVSIEVKEVV KPLLVSTLGE KSGKGLKTCK SPGRKSKESS PKGRSSSASS
361    PPKKEHHHHH HHSESPKAPV PLLPPLPPPP PEPESSEDPT SPPEPQDLSS SVCKEEKMPR
421    GGSLESDGCP KEPAKTQPAV ATAATAAEKY KHRGEGERKD IVSSSMPRPN REEPVDSRTP
481    VTERVS
```

Figure S1. MeCP2 sequence colored by domains. The different domains in MeCP2 e2 are shown in grey (NTD), red (MBD), blue (ID), green (TRD), purple (CTD-α), and orange (CTD-β). The two residues mutated (Rett-associated mutations: R106W, R133C) are shown in bold.

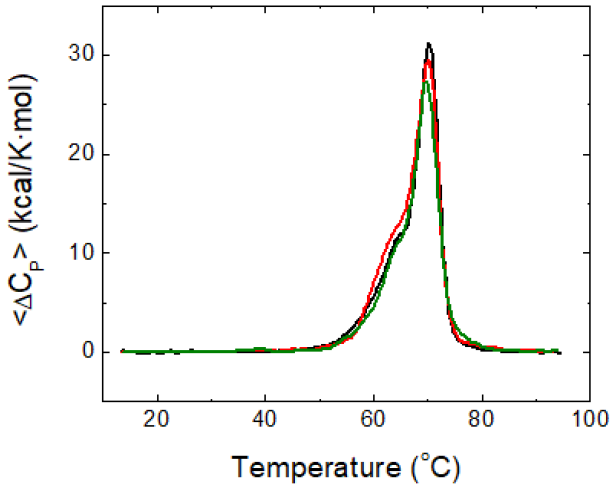

Figure S2. Thermal unfolding of CpG-DNA, mCpG-DNA, and hmCpG-DNA by DSC. Thermal unfolding of the three cytosine-modified dsDNA: CpG-DNA, mCpG-DNA, and hmCpG-DNA by differential scanning calorimetry (DSC) at pH 7. The apparent unfolding temperature T_m is almost coincident in the three cases (ca. 70.0 °C).

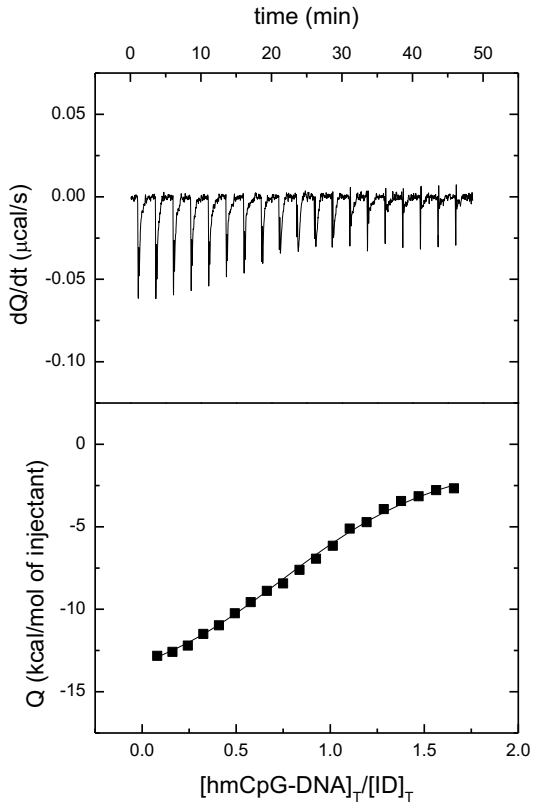

Figure S3. Interaction of ID with hmCpG-DNA by ITC. Interaction of hmCpG-dsDNA with ID assessed by isothermal titration calorimetry (ITC) at pH 7. Upper plot shows the thermogram (raw thermal power as a function of time) and the lower plot shows the binding isotherm (ligand-normalized heat effects as a function of the molar ratio). Non-linear least-squares analysis using a single-site binding model allowed estimating the observed binding affinity and enthalpy (continuous line). The estimated dissociation constant K_d was 0.43 µM.