Published online 15 March 2023

Nucleic Acids Research, 2023, Vol. 51, No. 12 ***6495–6506***
https://doi.org/10.1093/nar/gkad134

Epigenetic CpG duplex marks probed by an evolved DNA reader via a well-tempered conformational plasticity

Himanshu Singh [1], Chandan K. Das[2], Benjamin C. Buchmuller [1], Lars V. Schäfer [2,*], Daniel Summerer [1,*] and Rasmus Linser [1,*]

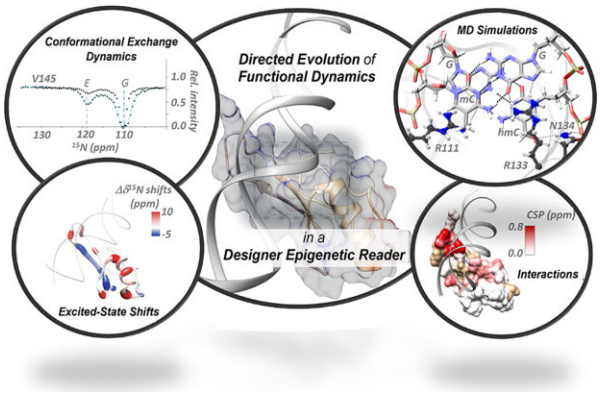
[1]Department of Chemistry and Chemical Biology, Technical University Dortmund, Otto-Hahn-Str. 4a, 44227 Dortmund, Germany and [2]Theoretical Chemistry, Ruhr University Bochum, Universitätsstr. 150, 44801 Bochum, Germany

Received November 09, 2022; Revised January 14, 2023; Editorial Decision February 12, 2023; Accepted February 15, 2023

ABSTRACT

5-methylcytosine (mC) and its TET-oxidized derivatives exist in CpG dyads of mammalian DNA and regulate cell fate, but how their individual combinations in the two strands of a CpG act as distinct regulatory signals is poorly understood. Readers that selectively recognize such novel 'CpG duplex marks' could be versatile tools for studying their biological functions, but their design represents an unprecedented selectivity challenge. By mutational studies, NMR relaxation, and MD simulations, we here show that the selectivity of the first designer reader for an oxidized CpG duplex mark hinges on precisely tempered conformational plasticity of the scaffold adopted during directed evolution. Our observations reveal the critical aspect of defined motional features in this novel reader for affinity and specificity in the DNA/protein interaction, providing unexpected prospects for further design progress in this novel area of DNA recognition.

GRAPHICAL ABSTRACT



INTRODUCTION

Cellular differentiation to stable, tissue-specific phenotypes despite identical genetic material is a prerequisite for the development of multicellular organisms. This is achieved by coordinated gene expression regulation via chromatin modification, such as the epigenetic modification of DNA nucleobases. In mammals, 5-methylation of cytosine by DNA methyltransferases (DNMTs) plays essential roles in differentiation, development, X-chromosome inactivation, and genomic imprinting; consequently, aberrant DNA methylation has been linked to multiple diseases, including cancer (1,2). Enzymatic oxidation of mC (Figure 1A) to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxycytosine (caC) is catalyzed by Ten-Eleven-Translocation (TET) dioxygenases and results in particularly high levels of oxidized mCs in embryonic stem

*To whom correspondence should be addressed. Tel: +49 231 755 3910; Email: rasmus.linser@tu-dortmund.de
Correspondence may also be addressed to Daniel Summerer. Email: daniel.summerer@tu-dortmund.de
Correspondence may also be addressed to Lars V. Schäfer. Email: lars.schaefer@ruhr-uni-bochum.de
Present address: Himanshu Singh, Department of Chemical Sciences, Indian Institute of Science Education and Research (IISER) Berhampur, Berhampur, Ganjam, Odisha 760010, India.

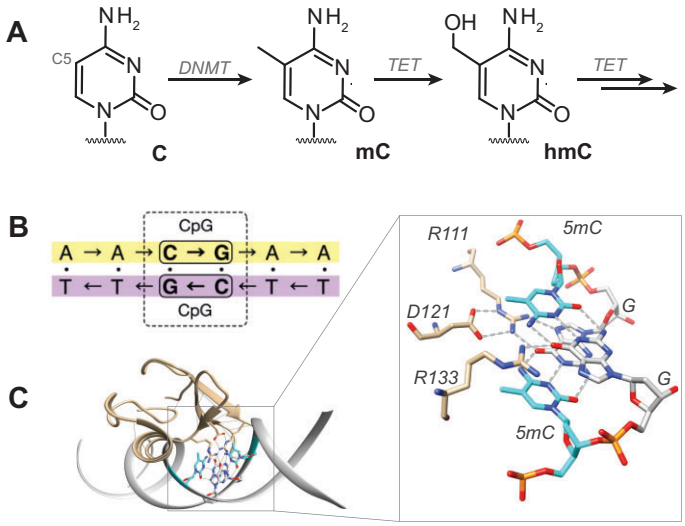


Figure 1. Natural selection of mC/mC DNA by methyl-CpG-binding domains (MBDs). (**A**) Cytosine modifications at CpG dyads, generated via DNMT and TET enzymes. (**B**) Depiction of a CpG dyad, in which cytosine can be epigenetically modified either symmetrically or asymmetrically. (**C**) Recognition of mC/mC dyads by MBDs via two H-bonds each from Arg111 and Arg133 to guanosine (16).

cells and the brain (3). These oxidized mC derivatives have been shown to exert regulatory functions in multiple contexts (4–6). Mammalian cytosine modification by DNMTs and TETs occurs predominantly in palindromic CpG dyads and can theoretically give rise to 15 different symmetric and asymmetric combinations of cytosine 5-modifications across the two CpG strands (7). However, despite the established general roles of oxidized mCs as chromatin regulators, it is poorly understood how their individual combinations in CpGs act as distinct regulatory signals, for example, by differentially modulating interactions with the large number of double-stranded DNA-binding chromatin proteins. Reader proteins that selectively interact with novel CpG duplex marks could serve as fundamental tools for studying their biological functions (8–11), but their design opens a new aspect in the field of DNA recognition that poses formidable selectivity challenges. We have recently reported the first designer reader for such a TET-associated CpG duplex mark. This protein has been evolved from a methyl-CpG-binding-domain (MBD) (12) and selectively recognizes the asymmetric combination hmC/mC in the context of all fifteen possible CpG duplex marks (13).

The 'core' MBD family proteins share a conserved domain of 70–80 residues and include the proteins MBD1–4 and methyl-CpG-binding protein 2 (MeCP2) (12). The latter represents a largely disordered DNA-binding protein for which loss-of-function mutations are associated with the neurological developmental disease Rett syndrome (RTT) (14,15). Its high-affinity interaction with mC/mC DNA hinges on two Arg fingers that both form two H-bonds to the guanosine of the CpG (Figure 1C) (16). Previous work on MBDs has shown that the stability of the three-dimensional fold is exceptionally susceptible to simple point mutations in the center of the hydrophobic core (17). Now, directed evolution of MBDs has recently been suggested as a viable path to generating reader proteins specifically targeting previously inaccessible combinations of epigenetic DNA modifications in CpG dyads, with the prospect of providing a platform for their genome-wide identification and mapping (13,18).

Directed-evolution experiments for selecting hmC/mC readers from an MeCP2 mutant library in a previous study of our labs revealed the replacement of a hydrophobic core residue (Val122) with Ala as a critical mutation, emerging in addition to a modified DNA binding interface (K109T/S134N, Figure 2A) (13). The V122A mutation in the K109T/V122A/S134N (TAN) triple mutant established high hmC/mC DNA-binding affinity (~10 nM) and specificity in electrophoretic mobility shift assays (EMSA). In contrast, a second mutant selected in those directed-evolution experiments, where Val122 was replaced with Cys (TCN), exhibited more promiscuous binding of both mC/mC and hmC/mC, even though it differed only by one nonpolar Cys residue at a core position that does not interact with DNA in the wt protein. This surprising role of a core residue in designed CpG duplex readers as potent determinant for selective target recognition reveals a fundamental lack of molecular-level understanding and unravels a considerable pitfall of structure-based approaches for the design of this novel class of epigenetic reader proteins.

MATERIALS AND METHODS

Protein expression for EMSA studies

MeCP2 variants with mutated residue 122 were generated by Quikchange site-directed mutagenesis and were expressed and purified as described earlier (13). EMSA assays were conducted as described (13). See the SI for details of these experiments (extended Materials and Methods and Supplementary Figure S1), protein and DNA constructs (Supplementary Table S1A and B), and details of fitting and error estimation (Supplementary Table S2).

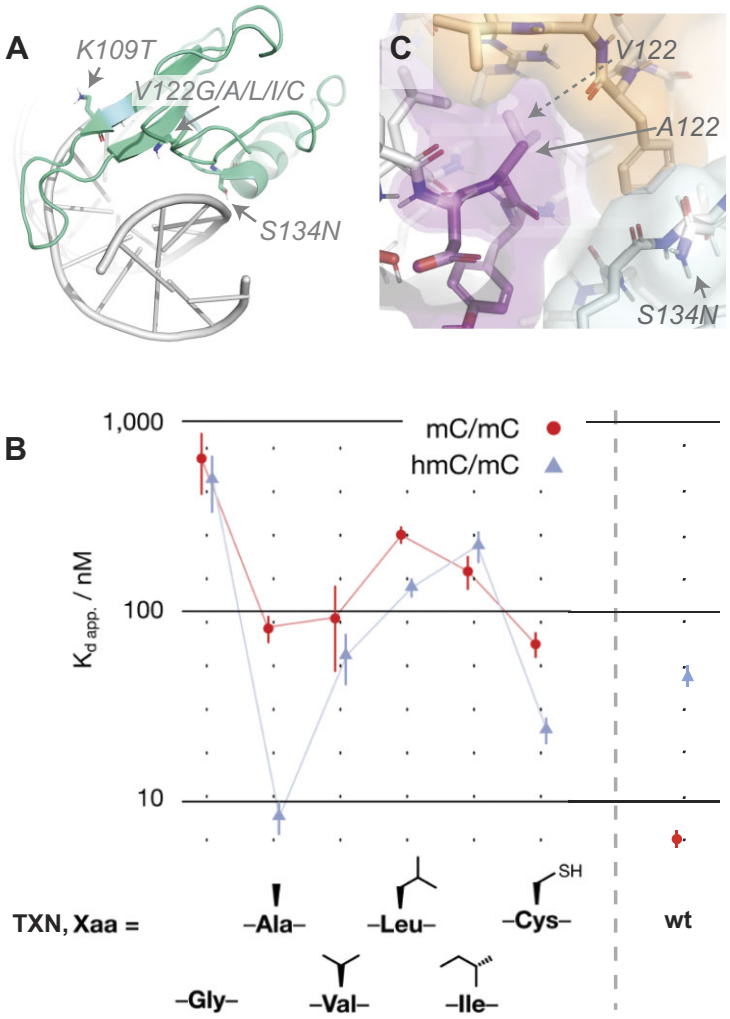


Figure 2. Engineering of hmC/mC selectivity into the natural MeCP2 reader. (**A**) Mutation sites in wt MBD (pdb: 3c2i). (**B**) Affinity for different triple mutations (K109T/V122X/S134N, 'TXN') in comparison to wt MBD (right), determined from EMSA shift essays (X = Ala, second column, represents 'TAN'; X = Val, third column ('TVN'), represents *double* mutation of the wild type). (**C**) Position 122 (the spatially more demanding Val shown in semitransparent) defines the central inter-subunit interface between extended sheet (purple, res. 104–126), helix 1 (green, 132–146), and C-terminal domain (orange, 148–161).

Protein expression and sample preparation for NMR spectroscopy

The MBD proteins were expressed and purified largely as described earlier (13) and discussed in more detail in the SI. In brief, cells were grown in ^{13}C,^{15}N-labeled M9 medium and purified using Ni affinity chromatography. After TEV cleavage and a second Ni affinity column, pure protein solutions for NMR experiments were obtained via size exclusion chromatography and concentration. For protein:DNA complex samples, the proteins were mixed in 1:1 ratio with DNA (a 12-mer with 4 nucleotides to form a hairpin, see Supplementary Table S1B) carrying central hmC/mC CpG dyads.

NMR sample preparation and assignments

Purified uniformly-^{13}C/^{15}N wild-type (KVS) and its double (TVN) and triple (TAN) mutants were prepared in a mixed solvent of 90% H_2O and 10% ^2H_2O (50 mM sodium phosphate, 50 mM NaCl, pH 6). All NMR experiments were carried out with protein concentrations of ~0.5 mM on a Bruker Avance 800 MHz NMR spectrometer using a triple-resonance cryo probe. The near-complete ^1H, ^{13}C and ^{15}N resonance assignments of MBD mutant MeCP2 protein TAN and its complex with hmC/mC DNA were deposited to the BMRB under the accession numbers 51020 and 34745, respectively. The chemical-shift perturbations were measured as $[(\Delta H)^2 + (\Delta N/10)^2]^{1/2}$, where ΔH and ΔN signify the changes in $^1H^N$ and ^{15}N chemical shifts, respectively. A suite of 3D double- and triple-resonance NMR experiments were performed for sequence-specific ^1H, ^{13}C, and ^{15}N backbone resonance assignments largely as discussed earlier (19,20). In addition, we recorded 3D HCCH-TOCSY, [^{15}N,^1H]-NOESY-HSQC, as well as aliphatic and aromatic [^{13}C,^1H]-NOESY-HSQCs for almost complete assignment of ^1H, ^{13}C and ^{15}N side-chain resonances, dihedral-angle restraints, and NOE-derived distance restrains for 3D structure calculation of the protein.

NMR spectra were processed using TopSpin4.0.8 and analyzed using CARA (21) and CCPN (22).

^{15}N relaxation experiments

Backbone ^{15}N relaxation measurements were acquired at 800 MHz and generally 291 K as described earlier (23). T_1 measurements employed eight recovery delays between 50 and 1100 ms. ^{15}N T_2 measurements were carried out using a CPMG pulse sequence (24) with relaxation delays of 5, 20, 35, 50, 70 and 90 ms. Steady-state [^{15}N, ^{1}H] heteronuclear-NOE measurements were carried out with and without proton saturation during the relaxation delay, using either 5 s of relaxation delay and 3 s of proton saturation or 8 s of relaxation delay only, respectively. Constant-time ^{15}N Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments (25) were measured at 291 K, using a constant-time delay of 40 ms and nine variable CPMG frequencies (ν_{CPMG}) ranging from 50 to 2000 Hz, in addition to a reference spectrum without delay ($\tau_{CPMG} = 0$). For each data set, the frequencies 750 and 50 Hz were repeated for estimation of errors in $R_{2,eff}$. Data for each residue (Supplementary Table S4) were analyzed individually using the NESSY software package (26) to obtain the kinetic parameters of interest, corresponding to either a no-exchange model or a two-site exchange process, dependent on the corrected Akaike information criterion (see the SI for details).

For global analysis, a two-state model was fitted to the data using the program SHEREKHAN (27). Global analysis for the wt combined residues K107, Q110, R111, F157, V159, and T160. For TAN, we combined the following residues for global exchange parameters: G103, T105, T109, Q110, R111, K112, S113, G114, S116, A117, G118, Y120, D121, Y123, I125, G129, K130, A131, F132, R133, N134, E137, L138, A140, Y141, F142, V145, G146, T148, S149 and N153. ^{15}N CEST experiments were implemented as described previously (28) at temperatures of 291, 283 and 278 K, using two sets of experiments at nominal ^{15}N B_1 fields of 15 Hz and 30 Hz and scanning 64 points across the ^{15}N spectral width from 134 to 99 ppm. CEST profiles were fit with the software Dynamics Center (version 2.7.2) to extract the exchange parameters.

NMR structure calculations

The solution structure of MBD triple mutant and its complex were determined by ARIA (29) using the following NMR constraints: *i)* dihedral angle constraints derived from TALOS-N (30) using individual $^{1}H^{N}$, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO chemical shift values as inputs (A total of 108 and 114 ϕ and ψ dihedral angle constraints were used for the apo TAN and TAN:hmC/mC complex, respectively.) and *ii)* cross peaks in NOESY spectra, identified and automatically assigned using ARIA 2.3 (29), with upper distance bounds set to 6.0 Å. The NMR structural statistics for the ensemble of 10 refined conformers of MBD apo and complex are summarized in Supplementary Table S3. The 10 lowest-energy conformers were further refined in explicit water using NMR-derived distance restraints, angle restraints and—for the apo form—25 ^{1}H/^{15}N-RDC restraints, via the inbuilt ARIA 2.3 CNS program. ^{1}H/^{15}N-RDCs were collected using Pf1 filamentous phages via IPAP HSQC experiments (31), with an alignment tensor calculated via the PALES (32) software. The program PSVS-1.4 was used to validate the quality of the selected ensemble of lowest-energy structures of TAN in apo and complex form. The 3D coordinates thus obtained were deposited in the PDB (pdb IDs: 8AJR and 8ALQ). The structure figures were prepared using Pymol and UCSF Chimera.

Molecular dynamics simulations

All MD simulations were performed using the Gromacs simulation package. Six different systems were modelled from the previously reported X-ray coordinates (PDB ID: 3C2I) (16), namely apo wt, apo double mutant (TVN), and apo triple mutant (TAN), as well as wt:mC/mC, TVN:hmC/mC and TAN:hmC/mC complexes. The ff99bsc0 Amber force field (33,34) was used for describing the proteins and nucleic acids. Each system was solvated using TIP3P water. Then, energy minimization, thermalization, and step-wise equilibrations were carried out. Subsequently, a total number of five production MD simulations for each system were initiated using different initial velocities. Each MD trajectory was 500 ns long, thus yielding a total simulation time of 2.5 μs for each of the six systems. The equations of motion were integrated using the leap-frog algorithm with a time step of 2 fs. The temperature and pressure were kept constant at 300 K and 1 bar using velocity rescaling thermostat (35) and Berendsen barostat (36), respectively. More details on MD simulations are provided in the SI.

RESULTS AND DISCUSSION

To study the peculiar influence of hydrophobic-core residues on the functional level, we generated additional mutants with different smaller (Gly) or larger (Val, Ile, Leu) residues at position 122, maintaining mutations required in the DNA binding interface (Supplementary Table S1A), in addition to TAN and TCN. We expressed the respective MBDs (residues 87–190) from human MeCP2 and measured dissociation constants (K_D) for mC/mC and hmC/mC-containing dsDNA by electrophoretic mobility shift assays (EMSA, Figure 2B, Supplementary Figure S1 and Supplementary Table S2). Only the selected TAN mutant exhibits a very high affinity (8 ± 2 nM) and selectivity for its new, hmC/mC-containing target. Strikingly, any deviation from this narrow steric space leads to a dramatic loss of selectivity, most pronounced for the TGN mutant. Importantly, both MBDs with the wild type residue Val at position 122 (TVN and wt KVS) exhibit equally poor (∼100 nM) binding of hmC/mC, but differ drastically in the binding to the canonical wild type target mC/mC, which is not bound by TVN anymore. Strikingly, this clear requirement for a specific steric demand at position 122 is observed despite the fact that it lies in a secluded element not interacting with DNA (Figure 2C).

To explore the molecular underpinnings of the modulating role of the protein core for selectivity of its interface, we conducted NMR and MD simulation studies. Wild-type MBD, the hmC/mC reader TAN, as well as the TVN mutant were overexpressed in doubly-labeled ^{15}N, ^{13}C minimal

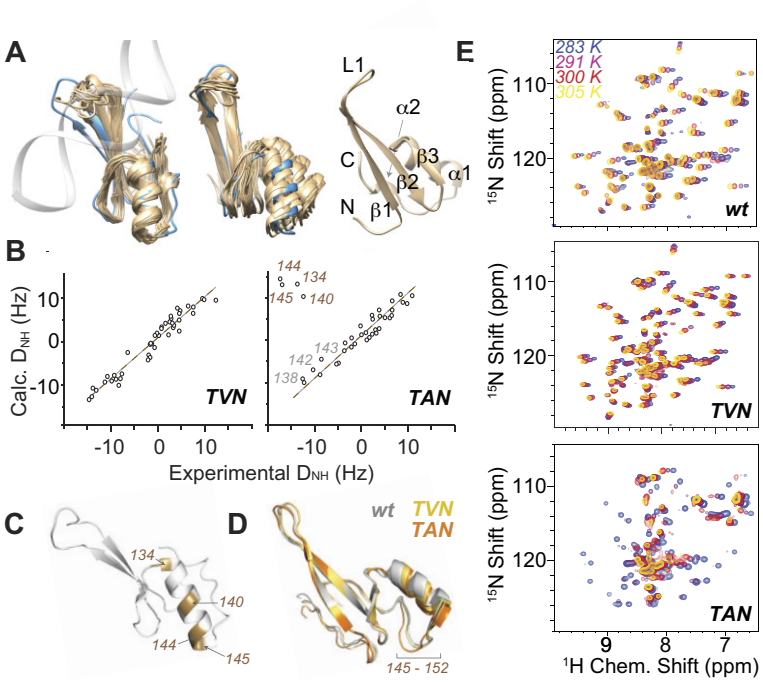


Figure 3. Assessment of structural properties of the apo hmC/mC reader. (**A**) Solution NMR structural ensemble of the hmC/mC reader TAN, showing the 20 minimal-energy structures aligned either over all residues (left, crystallographic DNA-bound structure 3c2i in blue) or with respect to the β-sheet (center), and the lowest-energy NMR structure with secondary-structural annotations (right, slightly different angle). (**B**) Residual dipolar couplings for double mutant TVN (left) and triple mutant TAN (right) correlated with theoretical values derived for the structure of the wild-type reader. (**C**) Residues with deviating RDCs depicted on 3c2i. (**D**) Average MD structures of the different apo proteins, i.e., wt (shown in gray), TVN (in yellow) and TAN (in orange), over 2.5 μs, highlighting structural variability in the helix and subsequent residues (also see Supplementary Figure S18). (**E**) Temperature dependence of wt (top), TVN (center), and TAN HSQC spectra (bottom), suggesting strong conformational exchange for TAN but not for wt and TVN. Structures in (C) and (D) from slightly different perspectives.

media and purified using affinity and size-exclusion chromatography. Backbone and sidechain ^{1}H, ^{13}C and ^{15}N resonance assignments were obtained for wt, TAN and TVN mutants, as well as the DNA-bound form of TAN via a series of triple-resonance NMR experiments in solution (see the SI). The assigned ^{15}N–^{1}H HSQC spectra of apo wt and TAN triple mutant are shown in Supplementary Figure S2. An assigned HSQC spectrum of TVN is part of Supplementary Figure S12. Resonance assignments have been deposited for the wt, TVN and TAN into the BMRB under accession codes 51 548, 51 547 and 51 020, respectively. As is described in the following, whereas wt and TVN show virtually identical behavior, TAN differs dramatically with respect to its backbone plasticity.

Ground state structure of the designed hmC/mC reader

The apo structure of the artificial hmC/mC reader TAN by NMR, determined from 698 internuclear distances from ^{15}N-edited and aliphatic or aromatic ^{13}C-edited NOESY spectra and 108 torsion angles, bears secondary-structural elements of the ordered regions (compare Supplementary Figure S3) identical to the wt reader. Backbone and hydrophobic side chains are well defined both in the helical domain and in the extended β-sheet, except for a loop region (residues 111–119, termed loop L1 in the following, compare Supplementary Figure S3) and N- and C terminal residues (Figure 3A). RDCs ($^{1}D^{NH}$) measured to specifically assess global structural features are mainly in agreement with the wt crystal structure (PDB 3c2i). However, significant deviations are observed for residues 134, 140, 144 and 145 in helix α1 (Figure 3B, C), even though RDCs for residues 138, 142 and 143 on the back of the helix are congruent. Including all RDCs into structure calculation yields poor conversion, which agrees with an ambiguous orientation of helix α1 in a dynamic ensemble (see below). Including all except those for the helical domain, we obtain an R^2 to the wt structure of 0.92 (compare Supplementary Figure S5). However, the precise inter-domain orientation eventually remains undetermined (center panel of Figure 3A). Figure 3A depicts the backbone structures of the 20 lowest-energy conformations of TAN thus obtained with a precision (root mean square deviation, RMSD) of 0.7 Å for the backbone and 1.4 Å for all heavy atoms (including residues 102–154 apart from loop L1), deposited into the PDB under accession code 8AJR. The statistics for the 20 final water-refined structures are shown in Supplementary Table S3. All of this contrasts with the double mutant TVN, which shows RDCs exactly matching the values expected from the wt mC/mC reader (Figure 3B). To better understand the different conformations and their dynamics, we complemented the NMR data with MD simulations. The average structures of apo proteins (in the absence of DNA, see Figure 3D, also compare Supplementary Figure S18B) observed in MD also point to slight deviations in the tertiary structure of the reader, corroborating the view of a

differentially plastic arrangement of the α1 helix. (See more details of the MD studies below.)

To our surprise, at a temperature of 293 K or above, many protein resonances of the hmC/mC reader TAN reversibly broaden beyond detection, indicating the presence of alternate conformational states on the μs-to-ms timescale (Figure 3E). By contrast, no significant indication of temperature-dependent exchange broadening is observed over a wide range of temperatures in the wt protein or TVN (which carries only the mutations directly involved in the DNA interactions), rendering its binding affinity to hmC/mC almost 10-fold lower than in TAN (Figure 2B and Supplementary Figure S1). Supplementary Figure S4 also shows the susceptibility of TAN amide shifts to temperature changes ($\Delta\delta N/\Delta T$ and $\Delta\delta H/\Delta T$) as well as its correlation with wt behavior. Again, the strong temperature dependence of several residues in TAN deviates from wt behavior, the most strongly deviating residues being V145, the C-terminal residue of α1 helix, and D121 next to the central mutation site. To assess the details of conformational rearrangements in the high-affinity TAN reader, we closely examined motion occurring on the ps-ns time scale as well as in the μs-ms regime, employing a large array of ^{15}N relaxation, relaxation dispersion (RD), and chemical-exchange saturation transfer (CEST) data.

The apo hmC/mC reader accesses alternate conformational states

[^{15}N, ^{1}H] heteronuclear NOE, ^{15}N R_1 and ^{15}N R_2 relaxation for TAN and wt MBD are shown in Supplementary Figures S3 and S4, respectively. The distribution of hetNOE and R_1 rates, reporting on ps–ns timescale dynamics, confirm the architecture of the domain with respect to its expected mobile N-terminus, extended C-terminus, and loop L1. More interestingly, transverse relaxation (R_2) is, in addition, sensitive to slower motions and reflects conformational-exchange processes on the μs–ms timescale. Whereas R_2 rates in wt simply mirror the fast-timescale mobility observed in R_1 and hetNOE, significant deviations in TAN reveal robust conformational exchange throughout the sequence. To assess the timescale of motion for the exchange, we carried out ^{15}N constant-time CPMG relaxation dispersion experiments. A dispersive nature in the dispersion profiles is the signature of conformational exchange on the μs–ms timescale between states with different chemical shifts. For wt, a small number of residues show the incidence of modest conformational exchange (Figure 4A and Supplementary Figure S6), with global RD on a timescale of around 200 μs (k_{ex} of 5207 ± 356 s^{-1}). In the hmC/mC reader TAN, strikingly, these exchange contributions are fourfold slower and more excessive than in the wild-type – with a timescale of about 800 μs (k_{ex} 1240 ± 10 s^{-1}) at 291 K, higher total R_2 rates up to 60 s^{-1}, and substantial exchange contributions R_{ex} up to >40 s^{-1} – widely surrounding the structural elements in loop L1, β1, β2 and β3 strands, and α1 residues. Figure 4A and B display R_{ex} mapped onto the structure and exemplary dispersion profiles (selecting the three mutation sites), respectively; Supplementary Figure S7 provides further dispersion profiles for TAN. RD data were fitted individually, assuming either a two-site exchange model or the absence of exchange, dependent on the corrected Akaike information criterion. See Supplementary Figure S8 for an overview about the residues with significant exchange contributions in both wt and TAN. Interestingly, neither V122 nor A122 backbone sites show dispersion themselves, reflecting the role of the side chain as a lever for the dynamics, the amide not being exposed to differential chemical environments itself (Figure 4A, B, center). Finally, the exchange dynamics ceases upon DNA binding. (See details regarding complex formation below.) Apparently, in the apo protein, an exchange occurs between different conformations, of which only one is relevant within the complex.

To more closely assess the conformational states sampled by the apo hmC/mC reader during the exchange, we used chemical-exchange saturation transfer (CEST) (28). Here, a weak radiofrequency field is applied to capture chemical shifts of minor conformations for each amide site. We recorded ^{15}N-based CEST on TAN at temperatures of 291, 283 and 278 K. At 283 K, we performed CEST analysis at two different saturating fields, 15 and 30 Hz (Figure 4C). A large dip represents the dominant, lower-energy ground state (GS), whose ^{15}N chemical shifts coincide with the regular NMR resonances, and a smaller dip for many residues discloses the chemical shift of an energetically excited state (ES). Amides with a significantly different minor dip, likely involved in strong structural transitions, are again distributed throughout the protein. Fitting the data to a two-state model (see the SI) indicates a dynamic equilibrium, GS ⇌ ES, on the ∼1 ms timescale, with a total exchange rate k_{ex} of 572 s^{-1} at 283 K and an excited-state population p_{ES} of around 3% (Figure 4D). ^{15}N CEST also allowed to derive forward and backward rates, k_{GE} and k_{EG}, of 17 s^{-1} and 555 s^{-1}, respectively (SI). This is in congruency with global fitting of the CPMG data at 283 K for TAN, which yielded a similar k_{ex} of 458 ± 68 s^{-1}. Further ^{15}N CEST profiles of TAN are shown in Supplementary Figure S9. An Arrhenius plot of k_{ex}, derived from CEST at three temperatures (291, 283 and 278 K, Supplementary Figure S10), provides an estimate for the activation energy of 9.5 kcal/mol, which is in line with the timescale of exchange observed in the RD data. Many residues with substantial discrepancies between ground- and excited-state chemical shifts are located in or near the binding-loop and α1 and denote changes in particular for residues just before and after the first helix (Figure 4E–G). This agrees with the contradictory RDCs for this helix and hence ambiguous relative orientation of the helical residues with respect to the β-sheet.

The absolute change in ^{15}N shifts for the excited state tends to be downfield for helical residues, in particular the beginning (N134) and end (V145) of α1, and upfield for β1, consistent with a temporary, partial release of the secondary structure at these sides (Figure 4F, G). Note that V145 also was the residue with the strongest temperature dependence (see above). Accordingly, the excited state represents a partly molten conformation, whose temporary adoption becomes possible due to altered interactions of the hydrophobic side chain in position 122 in the interface between the long β-sheet, α-helix and C-terminal residues.

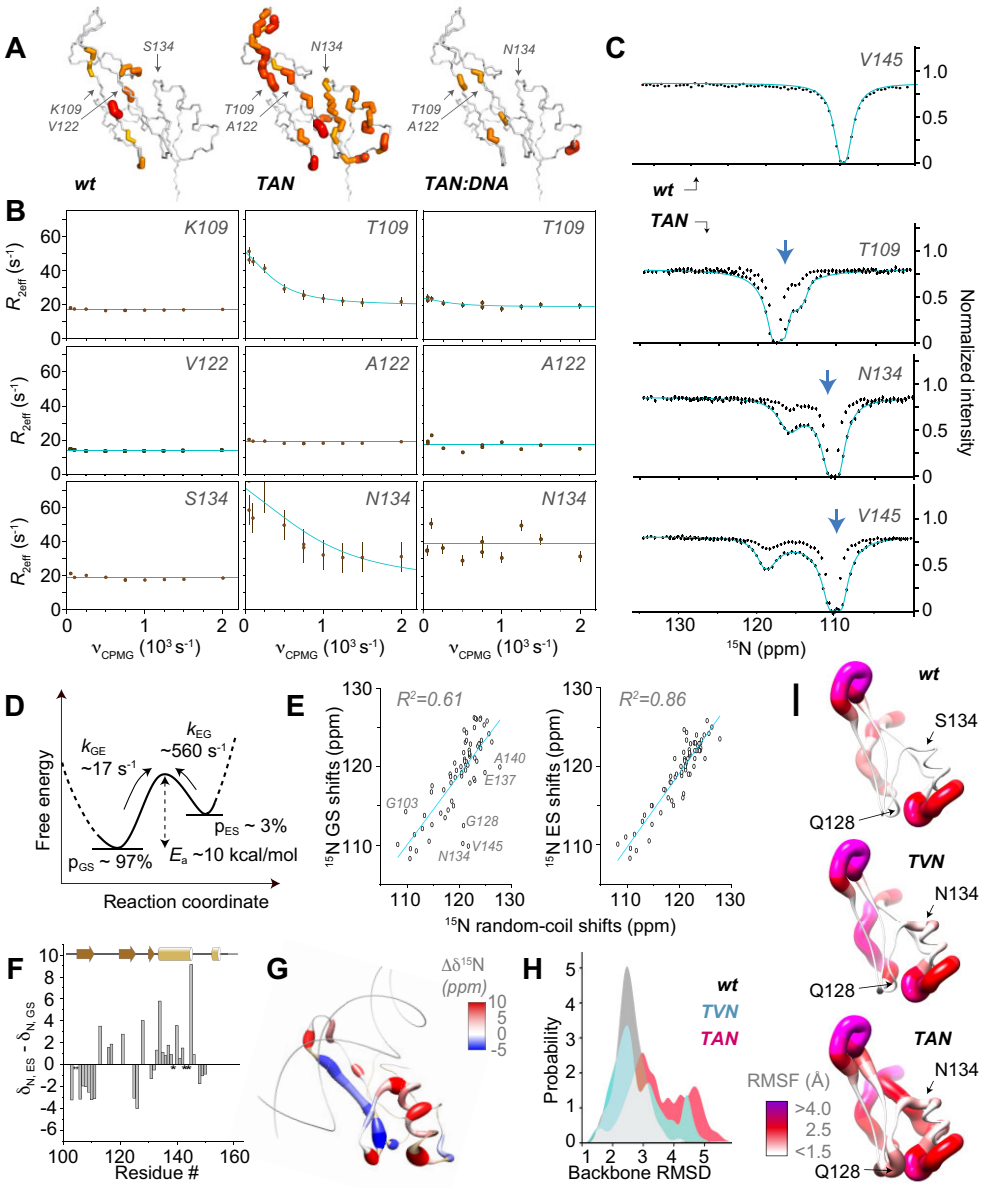


Figure 4. Conformational destabilization of the hmC/mC reader. (**A**) Comparison of exchange contributions R_{ex} from CPMG RD experiments for apo wt, apo TAN, and TAN:DNA complex, mapped onto the MBD structure (pdb 3c2i). (**B**) Constant-time ^{15}N CPMG profiles of the wt mC/mC reader (left), the hmC/mC reader TAN (middle), and the TAN:DNA complex (right) for the three mutation sites at 291 K. (**C**) ^{15}N CEST profiles (at 283 K) of apo TAN, recorded with a B_1 saturating field of 15 Hz (no fit) and 30 Hz (cyan fit) of these residues, in comparison with the wt, taking V145 as a representative residue (top). The blue arrows denote ^{15}N shifts of TAN in the DNA-bound state (at slightly higher temperature of 305 K, see more details below). (**D**) Thermodynamics and kinetics of conformational exchange as derived from CEST data and their temperature dependence. (**E**) Correlation between either ground-state (left) or excited-state chemical shifts (right) with neighbor-corrected ^{15}N random-coil shifts, revealing the tendency of unfolding in the excited state. (**F**, **G**) Difference between ground-state and excited-state ^{15}N chemical shifts in apo TAN, plotted as a function of sequence and mapped onto the structure (3c2i), respectively, with consistent trends towards partial unfolding. (**H**) Distribution of MD-observed deviations (RMSD) of the backbone from the X-ray structure, demonstrating a slight destabilization of apo TAN compared to the others. (**I**) Root-mean-square fluctuations (RMSF) witnessed in MD simulations of apo proteins, with fluctuations of the regions around residues 134 and 128 'switched on' successively going from wt to TVN to TAN (top to bottom).

On the other hand, in line with its temperature-dependent HSCQ spectra, the double mutant TVN does not show the strong conformational exchange in ^{15}N CEST or CPMG experiments (Supplementary Figures S11 and S12, respectively). The onset of vast chemical exchange by (and only by) V122A in the (otherwise identical) triple mutant protein supports the notion that the V122A mutation allows to modulate the conformational-exchange dynamics, which—when incorporated in addition to the constitutive changes in the DNA binding interface (K109T, S134N)—ultimately enables the reader to achieve high-affinity binding to the hmC/mC.

In order to shed further light on the structural fluctuations with atomic resolution, we interrogated dynamics in the apo proteins (wt, TVN and TAN) and their DNA complexes (see section on DNA binding below) in MD simulations, which provided a detailed description of motion up to 2.5 μs. Even though slower motions on the μs timescale and beyond cannot be faithfully sampled (unless enhanced-sampling techniques are used, which render the interpretation of timescales challenging (37)), the tendencies seen in the MD simulations qualitatively match those observed experimentally. Figure 4H and I shows the distribution of RMSDs to the X-ray structure (over all residues) and the residue-resolved root-mean-square fluctuations (RMSFs) of the apo proteins, respectively. Whereas the region around residues 132 and 138 shows increased plasticity over wt for both TVN and TAN mutant, TAN shows an additional systematic increase of backbone fluctuations between 119 and 132. A similar increase is observed between 103 and 109, preceding mutation K109T. Interestingly, we do observe the above-described local unfolding seen experimentally also in one of our five MD trajectories of TAN. Even though this is only a single event (due to the limited MD timescale), and hence has to be interpreted very carefully, it may shed further light on the destabilization of the structure by V122A mutation. Supplementary Figure S13 shows an overlay of the structures, where β1/β2 (harboring T109) is slightly reoriented relative to the rest of the structure, L1 loop becomes extended, and α1 is shortened by one turn (reaching only up to Y141), nicely in line with the strong shift perturbation of the C-terminal helix residues due to temperature (see above), upon DNA binding (see below), and when transitioning to the excited state (CEST data). Extended simulations, combined with enhanced-sampling methods, will be required to further characterize the nature of the putative locally unfolded state.

The structure of the hmC/mC reader in complex with DNA

In addition to the apo proteins, the complex between TAN and hmC/mC DNA was subjected to NMR investigation of structure and dynamics. To more closely investigate the binding of the hmC/mC reader to its DNA, ^{15}N-labeled TAN was titrated and equilibrated with an unlabeled, double-stranded hairpin oligonucleotide, carrying asymmetric hmC/mC modifications in the central CpG dyad, in a 1:1 molar ratio. We completed sequence-specific backbone and sidechain assignments of the TAN-DNA complex by similar experimental strategies as for the apo proteins (Supplementary Figure S14, deposited under BMRB accession code 34745). We then assessed chemical-shift perturbations (CSPs, see the SI for details) upon complex formation, which are shown as a function of sequence and depicted on the X-ray structure of the wt reader (pdb 3c2i) in Figure 5A and B, respectively. CSPs largely match the positions expected from the wt complex (pdb 3c2i), with perturbations seen in particular for the poorly structured loop L1, which slides into the DNA major groove. (Also compare Supplementary Figure S15). In addition, CSPs are found at allosteric positions distant from the interaction sites (e.g. at F142 at the end of α1 or L100), which corroborate the overall reshuffling of the dynamic conformational ensemble upon binding. The tertiary structure of the complex was elucidated using a similar set of restraints as for the apo structure; however, RDCs for the complex could not be obtained due to sample instability in the presence of alignment media. Also, due to the absence of intermolecular NOEs, we did not specifically include the DNA in structure determination. Within the precision of this assessment, all individual structural elements of the TAN:hmC/mC complex, deposited as PDB 8ALQ, seem highly reminiscent of the wt reader in complex with a symmetrical mC/mC dyad as observed in 3c2i (see an overlay of lowest-energy NMR conformers in Figure 5C, also compare Supplementary Figure S16).

We asked whether the bound form resembled either the compact or the destabilized excited state of the TAN conformational ensemble. All secondary structural features of the complex are essentially identical to the ground state apo protein (Figure 5D and E). More interestingly, knowing the residue-resolved ^{15}N chemical shifts of the protein in the complex, correlations were sought between these and ^{15}N shifts of either the ground or excited state obtained from CEST data of apo TAN (Figure 5F and G, respectively). Again, a correlation with a correlation coefficient R^2 of around 0.93 for the ground state shifts, in contrast to much larger deviations to the partially unfolded state (R^2 of 0.76), shows the ground state secondary structure of the apo reader to be reconstituted upon binding. This speaks against conformational preselection via a defined excited state as the main mechanism and instead points to facilitation of induced-fit binding of the apo protein to its target by decreased rigidity. Associated with a high energy barrier, this plasticity, however, incurs low entropic costs upon complex formation.

Increased plasticity facilitates DNA binding

We closely inspected the MD simulations of the complex to complement the experimental data on structural properties, dynamics and interactions and further elucidate the specificities of the interaction of TAN as the first hmC/mC reader with its target DNA. As an expected source of adopted selectivity in addition to the known interactions of the wt protein (38), the S134N mutation is confirmed to allow specific H-bonding between 5-hmC and reader (Figure 6A and Supplementary Figure S17). (The SI also contains a short movie depicting this interaction.) H-bonding both relates to the hmC hydroxyl group and the phosphate backbone (Figure 6B). The two-sided polar character of Asn at the same time disfavors interaction with mC/mC DNA, which in the wt hinges on a hydrophobic cluster formed by mC methyl, deoxyribose CH_2, and the aliphatic back-

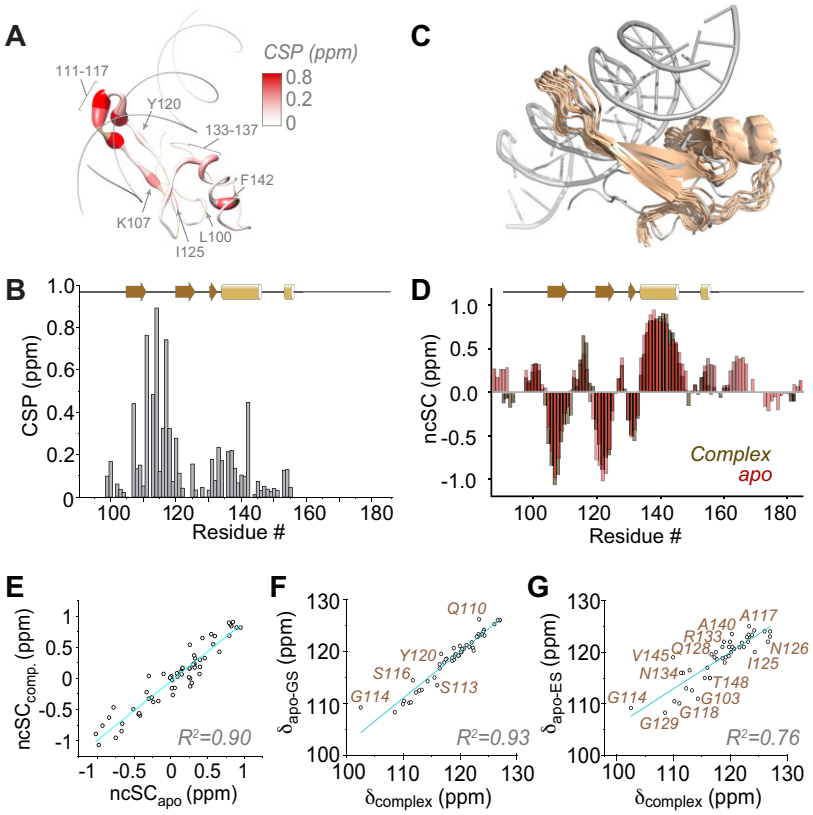


Figure 5. Experimental observations for structure and interactions in TAN complexed with its target hmC/mC DNA. (**A**, **B**) Chemical-shift perturbations (combined ^{1}H and ^{15}N shifts, see the SI for details) in the hmC/mC reader upon complex formation displayed on 3c2i and as a function of sequence, respectively. Shades of red denote CSPs above 0.2 ppm, brown color missing assignments in either apo protein or complex. Whereas residues 111–117, 107 and 133–136 are close the DNA, also more distant residues (marked) bear CSPs. (**C**) NMR structure of the TAN:hmC/mC complex, displayed as a bundle of the 10 lowest-energy conformations, overlaid with 3c2i (gray). (**D**) Overlay of neighbor-corrected secondary-structural propensities, derived from C^{α}, C^{β}, CO, N, and H^{α} shifts and their random-coil values, of TAN in apo form and in complex with hmC/mC DNA. (**E**) Correlation of secondary-structural propensities in the complex with those of the apo protein. (**F**, **G**) Correlation of ^{15}N shifts in the complex with either ground-state or excited-state shifts in apo-TAN, respectively.

side of Ser. In addition, we again inspected how complex formation is facilitated by specific structural features of the TAN mutant. Indeed, the formation of the TAN:hmC/mC DNA complex in the simulation is associated with slight conformational rearrangements. Whereas some rearrangement also occurs upon wt:mC/mC complex formation, the specific characteristics differ. Figure 6C visualizes the nature of displacement via an overlay of the average apo structure and the respective MBD:DNA complex (aligned with respect to the β1/β2 sheet). α1 in the average structure of the TAN:hmC/mC complex has slightly reduced absolute displacement (1.9 Å) from its position in apo TAN, both in comparison to the wt:mC/mC complex (2.8 Å) as well as the TVN:hmC/mC complex (3.3 Å; displacement of the helix, determined using F142 CO coordinates). However, with the slight differences adopted in its apo tertiary structure (Figure 3E and Supplementary Figure S18B), the displacement needed in TAN is a *parallel slide* of the helix, whereas for the wt, a mere *bending* is necessary. (TVN shows an intermediate behavior, where both is necessary to accommodate hmC/mC DNA.) With its decreased rigidity of the core and hence reduced strains for the architecture of its new binding interface, TAN—upon sliding into the major groove of the DNA—is able to adopt the exact structural characteristics required for the above interactions with the hmC/mC dyad to form. MD pinpoints this elevated plasticity (Figure 6D, left) to β1 (102–110, allowing to tune R111 interactions with guanosine) as well as the inter-subunit interface (118–140, which includes K130 interactions to the phosphate backbone, R133 interactions with guanosine, and N134 H-bonding to hmC), in complete agreement with the dynamics from the NMR experiments described above. This plasticity facilitates stable and tight interactions with hmC/mC DNA, upon which this mobility ceases (Figure 6D, right). Note that in contrast to the now stably formed interactions of the TAN complex, for the more restrained TVN double mutant, increased local fluctuations are now observed at residues 126–130, denoting a weakened K130-phosphate interaction, when complexing hmC/mC DNA (Figure 6D, dark blue curve, and S18A, right panel).

Two significant interactions by Arg 'fingers' 111 and 133, which can form H-bonds with the DNA backbone and hence constitute a major enthalpic driving force for complex formation, have been discussed (Figures 1C, 6E and Supplementary Figure S17) (16). These H-bonds are stably formed in our simulations (Supplementary Figure S19). The large perturbation of R111 chemical shifts upon binding

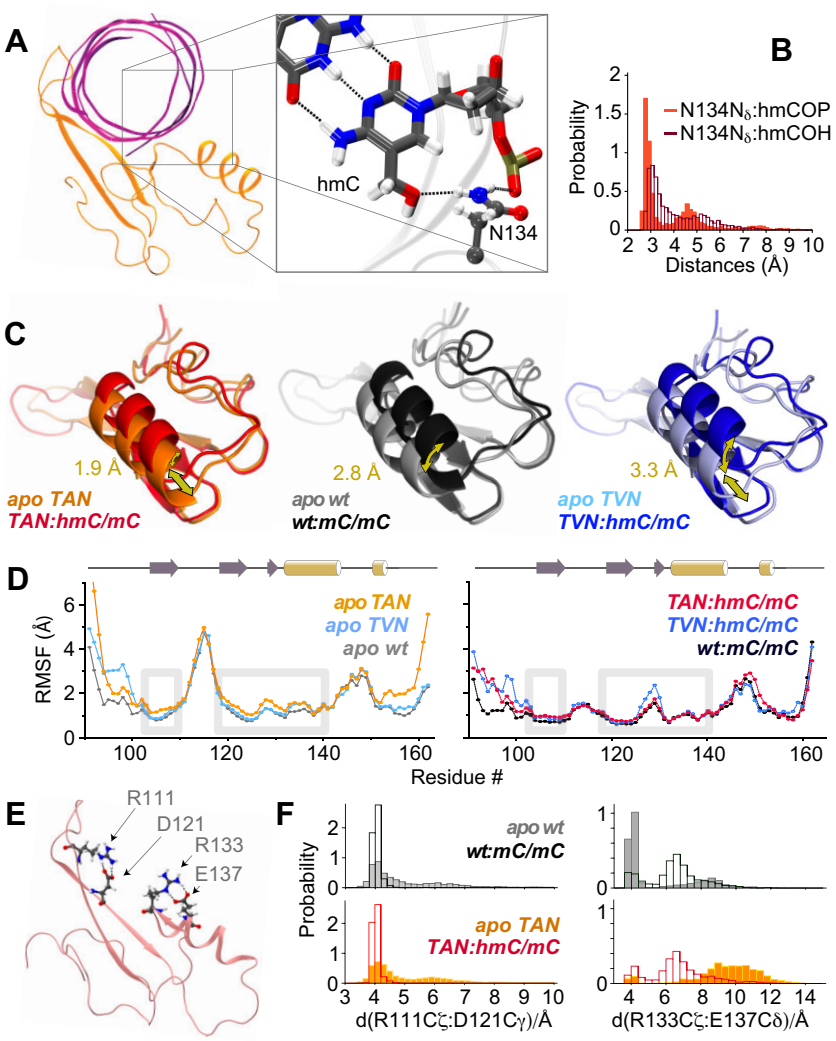


Figure 6. In-silico observations regarding complex formation between MBD and target DNA for different constructs. (**A**) Formation of the N134:hmC H-bond as the enthalpic driving force in case of the epigenetically modified CpG. (**B**) Histograms of N134 N_δ H-bonding to hmC hydroxyl and phosphate. (**C**) Average backbone structures from 2.5 μs MD simulations of wt, TAN, and TVN in apo form or in complex with their target DNAs, aligned with respect to β1/β2. The annotated arrows reflect the displacement between the F142 carbonyl (CO) group in apo proteins and complexes. (**D**) Residue-specific RMSFs in apo proteins and complexes. Gray boxes highlight β1, framing the important R111:guanosine interaction, and the area involving the inter-subunit interface (res. 122: central mutation site, 130: Lys to phosphate backbone, 133: Arg to guanosine, 134: H-bond to hmC, 137: potential R133 preorganization). (**E**) Arg residues forming a main selectivity aspect, including differentially preorganizing salt bridges. (**F**) Distance probabilities of the R111/D121 (left) and the R133/E137 salt bridge (right) of wt (top) and TAN (bottom) in apo form (solid bars) and in complex with matching target DNAs (open bars).

(see above) increases the discrepancy that its ^{1}Hε proton has in the apo state compared with other Arg sidechain moieties (Supplementary Figure S20A). This deviation, both for wt and TAN, confirms a preorganization of the salt bridge with neighboring D121 in the apo state that eventually also characterizes the complex of the wt mC/mC reader with DNA in crystals. Supplementary Figure S20B shows strips of a ^{15}N-edited NOESY experiment, where the preformed intraresidual contacts between R111 and D121 in the TAN mutant apo form are apparent. MD data for R111, which maintains a stable salt bridge with D121 both in apo and complex states of *all* readers in the simulations, is shown in Figure 6F, left panels. Importantly, in contrast to R111 and to what has been proposed in the framework of a possible selectivity mechanism (6), the salt bridge preorganizing Arg133 for its H-bond to guanosine in the complex is witnessed in the apo wt protein, but it is released in our simulations of the wt:mC/mC complex (Figure 6F, upper right panel). In the TAN mutant, by contrast, the R133/E137 salt bridge is absent both in the apo form and in the complex (Figure 6F, lower right panel, and Supplementary Figure S21). Absence of this intramolecular ionic lock in the apo protein avoids energetic penalties to open it upon complex formation, apparently without corrupting the formation of the important R133 H-bond. Finally, the precise positioning of R111 hinges on strand β1. Flexibilization of the relative position of this strand (Figures 4B and 6D, left) by Lys-to-Thr mutation of residue 109 may reduce strains negatively associated with R111 binding in the context of the modified interface.

Overall, a high affinity of the new MBD to the asymmetric, hydroxymethylated dyad seems to depend on smoothening of the free-energy landscape of the reader towards enabling conformational adaptations. This key property is fine-tuned by interactions between the three central protein secondary-structural elements (helix α1, the extended β-sheet, and the domain connecting the two)—which is defined by the central residues in the hydrophobic core—rather than the binding interface itself.

The above results show that redirecting the specificity of MBDs as a naturally existing scaffold to a new epigenetic CpG duplex modification hinges on tailored modulation of the thermodynamics of binding, derived not only from new intermolecular contacts but also tuning the characteristics of conformational plasticity in the interface. The enabling plasticity, adjusted via central residues within the protein core, derives from changes in the steric matches in the inter-domain interaction surfaces. The extent of this mobility, apparent from a tendency towards rare local unfolding, is strongly enhanced in the new hmC/mC reader. The adoption of an excited state on the μs-ms timescale motion and hence the presence of a similar tendency, albeit less pronounced, also characterizes the natural mC/mC reader. Suitably adjusted plasticity of the binding interface thus seems an important general aspect of target recognition for the MBD fold. At first glance, such 'disorder' in the apo scaffold seems like an entropically disadvantageous property for a high-affinity binder in which plasticity decreases upon binding. However, with features of this plasticity being very modestly tuned (i.e. with maintained, well-defined secondary-structural elements, tight conformational restrictions, and a high activation barrier), the penalty at physiological temperatures is minor – while still redefining a 25 Å wide binding interface – and can be largely compensated by maximized enthalpic gains due to optimized H-bond formation. The design of reader proteins that can serve as probes for the analysis of postsynthetic modifications of nucleic acids constitutes a current key aim of the soaring fields of epigenetics and epitranscriptomics. We showed that sought new properties of relevant protein-nucleic acid interfaces can be induced by directed evolution based on natural progenitors. Whereas the design and selection of mutant libraries with well-defined randomization sites guided by visual inspection of crystallographic structures that report on local interactions is the most intuitive approach, our data show that, by contrast, interrogating nucleic acids by designed reader proteins can also critically hinge on correctly adjusting protein plasticity as a modulator of selective complex formation. As such, central mutation sites far from interaction surfaces but relevant for inter-domain connectivity allosterically can enable high affinity and selectivity of readers, hence allowing for unexpected new perspectives for progress, particularly in the new field of CpG duplex mark recognition. We believe that the interrogation and understanding of dynamic networks in epigenetic readers, writers, and erasers can represent a fundamental element for designing future probes to decipher and effectively modulate the layer of epigenetic control of cell fate and function. These findings will be interesting for design problems in other contexts of target recognition, as—despite active research towards understanding of dynamic networks (39–41)—dedicated allosteric optimization of large-scale dynamics as a lever for any desired functionality still tends to escape awareness in the creation of new molecular tools.

CONCLUSION

Here, we have demonstrated that the high affinity and selectivity of the first designed epigenetic reader for oxidized CpG dyads is leveraged by well-defined conformational plasticity of the DNA binding interface, remotely orchestrated by interactions in the hydrophobic core. The observed intermediate-timescale conformational exchange towards a partial melting of secondary-structural strains, elucidated via extensive NMR and MD interrogation of protein structure and dynamics, demonstrates adapted plasticity as a lever for specific reader:DNA interactions within the vast landscape of differential epigenetic modifications. Albeit to a lower extent, exchange dynamics on the same timescale are also visible for the natural, mC/mC-specific MBD. Our study suggests that the aspect of tailored conformational plasticity may both, help understanding physiological reader selectivities and facilitate the design of novel readers as specific molecular probes for different CpG duplex marks. The findings will propel the advances in the emerging field of DNA recognition and thus in deciphering the elusive roles of individual CpG duplex modifications in chromatin regulation.

DATA AVAILABILITY

NMR chemical shifts have been deposited to the BMRB (www.bmrb.com) under accession codes 51548, 51547 and 51020. Structural data have been deposited to the PDB (www.rcsb.org) under 8AJR and 8ALQ.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Simone Eppmann for the expression and purification of MBD proteins. HS dedicates this manuscript to the late Prof KVR Chary and acknowledges the Ramalingaswami reentry fellowship by the Department of Biotechnology, New Delhi, GoI.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 27112786, 325871075, and the Emmy Noether program. Funded by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy – EXC 2033 – 390677874 – RESOLV, and EXC-114 – 24286268 – CiPS-M; Funded by the European Research Council (ERC CoG EPICODE, No. 723863 to D.S.). Funding for open access charge: Deutsche Forschungsgemeinschaft.

Conflict of interest statement. TU Dortmund University has filed a patent application for the engineered MBDs disclosed herein (PCT/EP2020/087979, pending). No further competing conflict of interests have been declared.

REFERENCES

1. Rodríguez-Paredes,M. and Esteller,M. (2011) Cancer epigenetics reaches mainstream oncology. *Nat. Med.*, **17**, 330–339.
2. Allis,C.D. and Jenuwein,T. (2016) The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.*, **17**, 487.
3. Carell,T., Kurz,M.Q., Müller,M., Rossa,M. and Spada,F. (2018) Non-canonical bases in the genome: the regulatory information layer in DNA. *Angew. Chem. Int. Ed.*, **57**, 4296–4312.
4. Raiber,E.-A., Portella,G., Cuesta,S.M., Hardisty,R., Murat,P., Li,Z., Iurlaro,M., Dean,W., Spindel,J. and Beraldi,D. (2018) 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat. Chem.*, **10**, 1258–1266.
5. Spada,F., Schiffers,S., Kirchner,A., Zhang,Y., Arista,G., Kosmatchev,O., Korytiakova,E., Rahimoff,R., Ebert,C. and Carell,T. (2020) Active turnover of genomic methylcytosine in pluripotent cells. *Nat. Chem. Biol.*, **16**, 1411–1419.
6. Ibrahim,A., Papin,C., Mohideen-Abdul,K., Gras,S.L., Stoll,I., Bronner,C., Dimitrov,S., Klaholz,B.P. and Hamiche,A. (2021) MeCP2 is a microsatellite binding protein that protects CA repeats from nucleosome invasion. *Science*, **372**, eabd5581.
7. Wu,X. and Zhang,Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*, **18**, 517–534.
8. Brinkman,A.B., Simmer,F., Ma,K., Kaan,A., Zhu,J. and Stunnenberg,H.G. (2010) Whole-genome DNA methylation profiling using MethylCap-seq. *Methods*, **52**, 232–236.
9. Serre,D., Lee,B.H. and Ting,A.H. (2010) MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.*, **38**, 391–399.
10. Booth,M.J., Raiber,E.-A. and Balasubramanian,S. (2015) Chemical methods for decoding cytosine modifications in DNA. *Chem. Rev.*, **115**, 2240–2254.
11. Buchmuller,B.C., Kosel,B. and Summerer,D. (2020) Complete profiling of methyl-CpG-binding domains for combinations of cytosine modifications at CpG dinucleotides reveals differential read-out in normal and Rett-associated states. *Sci. Rep.*, **10**, 4053.
12. Du,Q., Luu,P.-L., Stirzaker,C. and Clark,S.J. (2015) Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics*, **7**, 1051–1073.
13. Buchmuller,B.C., Dröden,J., Singh,H., Palei,S., Drescher,M., Linser,R. and Summerer,D. (2022) Evolved DNA duplex readers for strand-asymmetrically modified 5-hydroxymethylcytosine/5-methylcytosine CpG dyads. *J. Am. Chem. Soc.*, **144**, 2987–2993.
14. Amir,R.E., Van den Veyver,I.B., Wan,M., Tran,C.Q., Francke,U. and Zoghbi,H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.*, **23**, 185–188.
15. Brown,K., Selfridge,J., Lagger,S., Connelly,J., De Sousa,D., Kerr,A., Webb,S., Guy,J., Merusi,C. and Koerner,M.V. (2016) The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.*, **25**, 558–570.
16. Ho,K.L., McNae,I.W., Schmiedeberg,L., Klose,R.J., Bird,A.P. and Walkinshaw,M.D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell*, **29**, 525–531.
17. Sperlazza,M.J., Bilinovich,S.M., Sinanan,L.M., Javier,F.R. and Williams,D.C. Jr (2017) Structural basis of MeCP2 distribution on non-CpG methylated and hydroxymethylated DNA. *J. Mol. Biol.*, **429**, 1581–1594.
18. Tam,B.E., Sung,K. and Sikes,H.D. (2016) Engineering affinity agents for the detection of hemi-methylated CpG sites in DNA. *Mol. Syst. Des. Eng.*, **1**, 273–277.
19. Venters,R.A., Farmer,B.T., Fierke,C.A. and Leonard,D.S. (1996) Characterizing the use of perdeuteration in NMR studies of large proteins: 13C, 15N and 1H assignments of human carbonic anhydrase II. *J. Mol. Biol.*, **264**, 1101–1116.
20. Singh,H., Das,C.K., Vasa,S.K., Grohe,K., Schäfer,L.V. and Linser,R. (2020) The active site of a prototypical "rigid" drug target is marked by extensive conformational dynamics. *Angew. Chem., Int. Ed.*, **59**, 22916–22921.
21. Keller,R. (2005) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment. https://doi.org/10.3929/ethz-a-005068942.
22. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,P., 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.
23. Grohe,K., Patel,S., Hebrank,C., Medina,S., Klein,A., Rovó,P., Vasa,S.V., Singh,H., Vögeli,B., Schäfer,L.V. *et al.* (2020) Protein motional details revealed by complementary structural-biology techniques. *Structure*, **28**, 1024–1034.
24. Meiboom,S. and Gill,D. (1958) Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Inst.*, **29**, 688–691.
25. Mulder,F.A.A., Skrynnikov,N.R., Hon,B., Dahlquist,F.W. and Kay,L.E. (2001) Measurement of slow (μs−ms) time scale dynamics in protein side chains by 15N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.*, **123**, 967–975.
26. Bieri,M. and Gooley,P.R. (2011) Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinf.*, **12**, 421.
27. Mazur,A., Hammesfahr,B., Griesinger,C., Lee,D. and Kollmar,M. (2013) ShereKhan—calculating exchange parameters in relaxation dispersion data from CPMG experiments. *Bioinformatics*, **29**, 1819–1820.
28. Vallurupalli,P., Bouvignies,G. and Kay,L.E. (2012) Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.*, **134**, 8148–8161.
29. Rieping,W., Habeck,M., Bardiaux,B., Bernard,M., Malliavin,T.E. and Nilges,M. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, **23**, 381–382.
30. Shen,Y., Delaglio,F., Cornilescu,G. and Bax,A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
31. Ottiger,M., Delaglio,F. and Bax,A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.*, **131**, 373–378.
32. Zweckstetter,M. (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.*, **3**, 679–690.
33. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
34. Zgarbová,M., Otyepka,M., Šponer,J., Mládek,A., Banáš,P., Cheatham,T.E. and Jurečka,P. (2011) Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.
35. Bussi,G., Donadio,D. and Parrinello,M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101.
36. Berendsen,H.J.C., Postma,J.P.M., DiNola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
37. Hoffmann,F., Mulder,F.A.A. and Schäfer,L.V. (2022) How much entropy is contained in NMR relaxation parameters?*J. Phys. Chem. B*, **126**, 54–68.
38. Ohki,I., Shimotake,N., Fujita,N., Jee,J., Ikegami,T., Nakao,M. and Shirakawa,M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, **105**, 487–497.
39. van den Bedem,H. and Fraser,J.S. (2015) Integrative, dynamic structural biology at atomic resolution - it's about time. *Nat. Methods*, **12**, 307–318.
40. Hadzipasic,A., Wilson,C., Nguyen,V., Kern,N., Kim,C., Pitsawong,W., Villali,J., Zheng,Y. and Kern,D. (2020) Ancient origins of allosteric activation in a Ser-Thr kinase. *Science*, **367**, 912–917.
41. Schenkmayerova,A., Pinto,G.P., Toul,M., Marek,M., Hernychova,L., Planas-Iglesias,J., Daniel Liskova,V., Pluskal,D., Vasina,M., Emond,S. *et al.* (2021) Engineering the protein dynamics of an ancestral luciferase. *Nat. Commun.*, **12**, 3616.