

# Identification of Chemical Probes Targeting MBD2

Diane Erdmann, Jean Contreras, Rémy Le Meur, Bruno Vitorge, Vincent Saverat, Ambre Tafit, Corinne Jallet, Véronique Cadet-Daniel, Corentin Bon, Phannarath Phansavath, et al.

▶ **To cite this version:**

Diane Erdmann, Jean Contreras, Rémy Le Meur, Bruno Vitorge, Vincent Saverat, et al.. Identification of Chemical Probes Targeting MBD2. ACS Chemical Biology, 2022, Published as part of the ACS Chemical Biology special issue “Epigenetics 2022”., 17 (6), pp.1415-1426. 10.1021/acschembio.1c00959 . pasteur-03688129

**HAL Id: pasteur-03688129**
**https://pasteur.hal.science/pasteur-03688129**

Submitted on 16 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Identification of chemical probes targeting MBD2

Diane Erdmann[1,2], Jean Contreras[1,3], Rémy A. Le Meur [4], Bruno Vitorge[4], Vincent Saverat[1], Ambre Tafit[1], Corinne Jallet[1], Véronique Cadet-Daniel[1], Corentin Bon[1], Phannarath Phansavath[5], Virginie Ratovelomanana-Vidal[5], Albert Jeltsch[6], Sophie Vichier-Guerre[1], J. Iñaki Guijarro[4], Paola B. Arimondo[1,*]

[1]*Epigenetic Chemical Biology, Department of Structural Biology and Chemistry, Institut Pasteur, Université Paris Cité, CNRS UMR3523, 75015 Paris, France*

[2] *Ecole Doctorale MTCI, Université Paris Cité, Sorbonne Paris Cité, Paris, France*

[3] *Ecole Doctorale FIRE, Université Paris Cité, Sorbonne Paris Cité, Paris, France*

[4] *Biological NMR and HDX-MS Technological Platform, Institut Pasteur, Université Paris Cité, CNRS UMR3528, 75015 Paris, France*

[5] *PSL University, Chimie ParisTech, Institute of Chemistry for Life & Health Sciences, CNRS UMR8060, 75005 Paris, France*

[6] *Department of Biochemistry, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, D-70569 Stuttgart, Germany*

* Correspondence: should be addressed to Paola B. Arimondo, EpiCBio, Institut Pasteur, Université *Paris Cité*, CNRS UMR3523, email: paola.arimondo@cnrs.fr

**Abstract**

Epigenetics has received a lot of attention in the last decade. Many insights on epigenetic (dys)regulation in diseases have been obtained and clinical therapies targeting them are in place. However, the readers of the epigenetic marks are lacking enlightenment behind this revolution, and it is poorly understood how DNA methylation is being read and translated to chromatin function and cellular responses. Chemical probes targeting the methyl-CpG readers, as the Methyl-CpG Binding Domain proteins (MBD), could be used to study this mechanism. We have designed analogues of 5-methylcytosine to probe the MBD domain of human MBD2. By setting-up a protein thermal shift assay and an AlphaScreen®-based test, we were able to identify three fragments that bind MBD2 alone and disrupt the MBD2-methylated DNA interactions. 2D NMR experiments and virtual docking gave valuable insights into the interaction of the ligands with the protein showing that the compounds interact with residues that are important for the DNA recognition. These constitute the starting point for the design of potent chemical probes for MBD proteins.

Running title: Chemical targeting of MBD2

## Introduction

Epigenetics is a widely explored biological field relevant for the understanding of cellular differentiation, embryogenesis and development, in complementary to genetics; (1,2) and aberrant epigenetic landscapes are found in human diseases. (3,4) Still, several questions remain to be answered, even concerning well-known epigenetic modifications such as DNA methylation in mammals, which is the most stable and studied epigenetic modification. (5) DNA methylation appears on gene regulatory sequences where it represses transcription, but it also occurs on gene bodies where it can be an activator of expression. (6,7) Three main actors of this modification are known: the writers, called the DNA-methyltransferases (DNMTs) (8), the erasers, the Ten-Eleven-Translocation enzymes (TETs) (9), and the readers, the Methyl-CpG Binding Proteins (MBPs) (10). On the one hand, the enzymatic mechanism of DNMTs, which add the methyl group on the DNA cytosine (11), and the process of active demethylation, involving the TETs through several oxidation steps and the intervention of the Base Excision Repair (BER) machinery, (12) are well characterized. On the other hand, the MBPs interpret the methylation profiles as an intermediate between the methylation patterns and gene regulation. They constitute a hub of signaling for DNA methylation and hydroxymethylation, being involved in large protein complexes that influence other epigenetic modifications, transcription factors or DNA accessibility. (13) How this hub works and its mechanism are still underexplored. Furthermore, DNA methylation is dysregulated in diseases, with methylation occurring on sequences where it is not supposed to, or on the contrary it can be missing. (14) Therefore, in addition to targeting the DNMTs, it might be of interest to explore the effects of the chemical targeting of MBPs in diseases, such as neurological pathologies and cancer. (15) Furthermore, the identification of chemical probes that target the MBPs in cells would help understand the processes, in which the DNA methylation readers are involved (16). The MBPs are divided in three families: the Methyl-CpG Binding Domain (MBDs), the Kaiso proteins and the SET RING finger Associated domain (SRA). (17) Each of these families has its own specificities. In this work, we focused on the MBDs. The five founder members (MeCP2, MBD1, MBD2, MBD3, MBD4) of the MBDs have been structurally characterized and some important regulatory pathways, in which they are involved, have been identified. (18) There are, however, few studies relating the identification of chemical probes targeting MBDs. Importantly, Wyhs *et al.* developed a time-resolved TR-FRET screening assay to identify compounds that disrupt the interaction between the MBD domain of hMBD2 and methylated DNA. (19) Two compounds were identified ($IC_{50}$ of 290 nM for NF449 and 2.7 nM for aurintricarboxylic acid), but-both compounds lack selectivity, bear reactive moieties and interfere with multiple processes. (20)(21) Few years before, the same laboratory identified, by using an adapted ELISA assay, compound KCC-07 able to inhibit the DNA-MBD2 interaction. (22) Specificity towards MBD2 compared to MeCP2 was observed, but binding ability to MBD2 alone without DNA was not studied. Interestingly, compound KCC-07 showed *in vivo* activity extending the lifespan of medulloblastoma xenografted mice by 6 days and restoring Brain-specific

Angiogenesis Inhibitor (BAI1) expression. (23) These studies paved the way to the identification of chemical probes for MBDs with the potential to help understand the role of these proteins. Here, we chose to develop two novel screening assays to identify compounds that bind to MBD alone and compounds that interfere with the binding of MBD to methylated DNA (MeDNA), respectively (**Figure 1**). The first screening assay selects molecules that interact with the MBD even with weak affinity, and the second one highlights the compounds that can disrupt the MBD-MeDNA interaction. We chose the MBD domain of human MBD2 (hMBD2) as target, because it has the strongest affinity for DNA's 5-methylcytosine (5mC). *MBD2* has been identified as a non-essential gene causing no major phenotype change upon knock-out, with possible compensation by others MBDs. (24–26) In the cellular context, MBD2 is involved in the Nucleosome Remodeling and Deacetylase (NuRD) complex, as the recruiter of the complex to DNA targets for gene repression. (27) In cancer, MBD2 is a key repressor of tumor suppressor genes by binding to their promoters, as shown in colorectal cancer. (28,29) Hence we expect that novel MBD2 binding inhibitors might be suitable at restoring epigenetically silenced genes, (25,26) as KCC-07. (23)

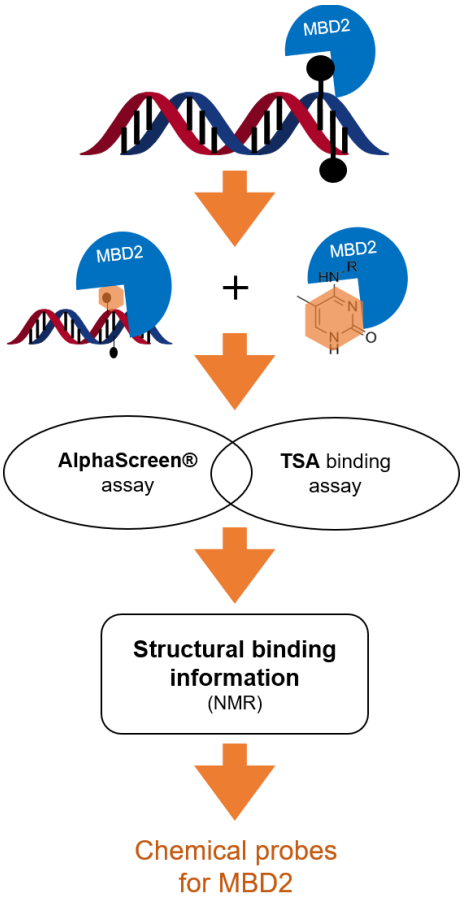



**Figure 1**- *Methodology to identify chemical probes of hMBD2*. Nine derivatives of 5mC substituted on position C4 were synthetized and screened in two assays: a protein thermal shift (TSA) binding assay and an AlphaScreen® assay to evaluate the binding of the compounds to MBD2 or their ability to disrupt the MBD2-MeDNA complex, respectively. NMR studies were carried out on selected compounds.

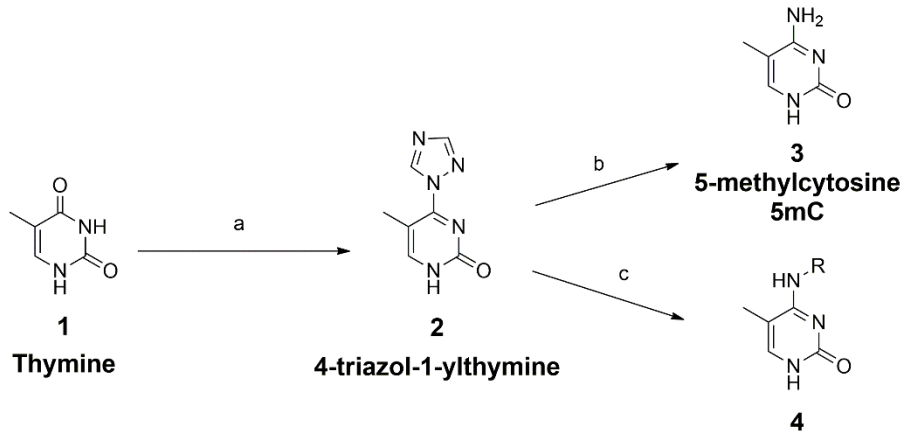
MBD2 is represented in blue, the 5mC by the black lollipop on the DNA and synthesized 5mC derivatives by orange hexagons.

Unlike Wyhs *et al.*, we opted for a focused chemical library and designed nine compounds that could potentially bind directly to MBD2. This should allow a better specificity regarding other DNA binding proteins. To design these compounds, we opted for a fragment-like approach based on derivatives of 5mC as MBD2 strongly binds to 5mC in DNA. We substituted 5mC on the amine group in the C4 position of the pyrimidine core (**Figure 1**), in order to keep on the C5 position the methyl group for the selectivity and on the N1 position the option to add the deoxyribose and synthetize the corresponding nucleoside.

We then performed NMR and modelling studies on the active compounds from the two screening assays and obtained structural insights on the interactions and identified some pharmacophoric features important for the interactions. After identifying the best compound, we synthesised the nucleoside analogue and submitted it to our assays. These results constitute the basis for the development of more potent probes.

## Results and Discussion

*Chemical design and synthesis of 5-methylcytosine derivatives.* We designed derivatives of 5mC by adding various substituents on the amine group in the C4 position of the pyrimidine core. Position N1 was not modulated to eventually couple the sugar on the best derivative. For the synthesis, we used thymine (**1**) as a starting point (**Scheme 1**). By nucleophilic substitution, a triazole is added to activate the C4 position (30) that can thus undergo a reaction of aromatic substitution with various primary amine derivatives. (31) The 4-triazol-1-ylthymine (**2**) obtained is the building block used to synthesize the 5mC as control compound, **3**, and $N^4$ derivatives **4**.




***Scheme 1****- Synthetic pathway to obtain $N^4$-substituted-5-methylcytosine derivatives.* (a) 1,2,4-triazole, $POCl_3$, triethylamine, acetonitrile, RT, overnight. (b) $NH_3$, RT, overnight. (c) R-$NH_2$, triethylamine, *N*,*N*-dimethylformamide, RT, overnight.

*Selected 5-methylcytosine derivatives.* First, we explored the addition of alkylphenyl groups on the $N^4$ position by varying the length of the linker between the pyrimidine and the phenyl group (**Figure 2**, compounds **4a** to **4e**). Then, various substituents were chosen: a pyridine was added to the linker with one carbon (**4f**) and two carbons (**4i**); and two chlorophenethyl substituents to the two carbons linker (the chloride atom is in para in compound **4g** and in meta in compound **4h**).

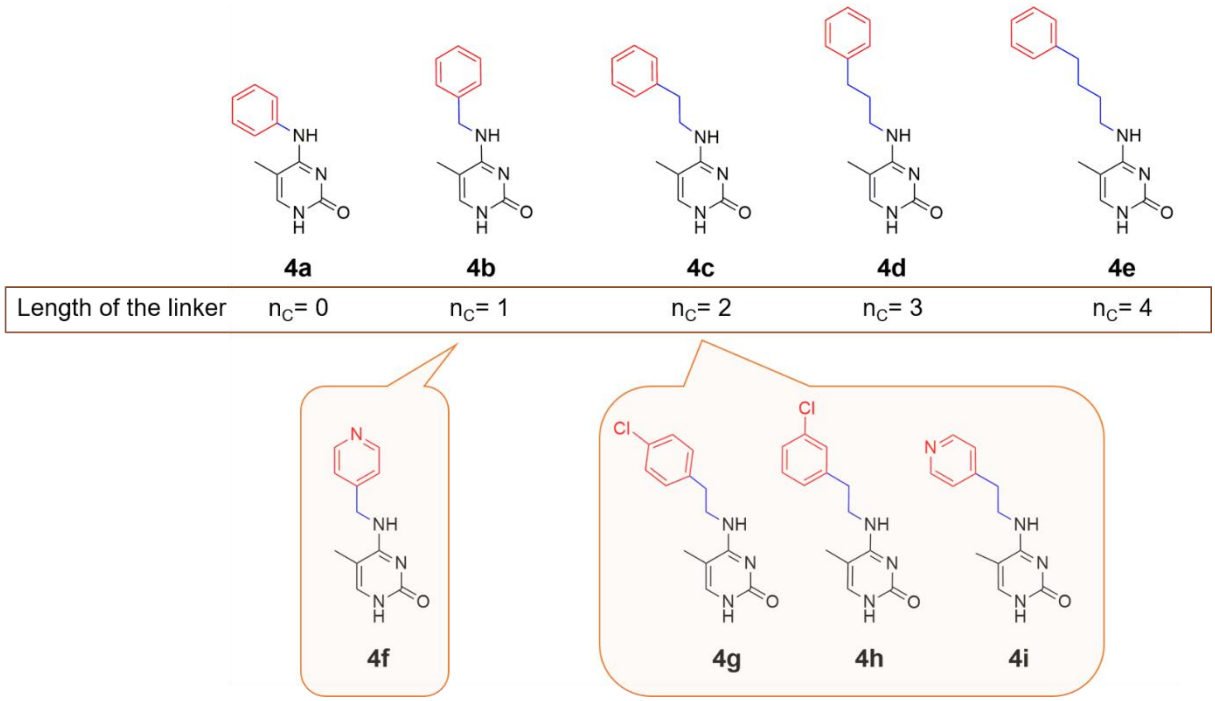



***Figure 2**- Structure of the nine 5mC derivatives **(4)**.* The length of the linker arm ($n_C$) is indicated by the number of carbon atoms. The pyridine- and chlorophenethyl-containing molecules are encircled in orange.

*Biophysical assays.* To evaluate the ability of the compounds to bind to MBD2, we optimized a thermal shift assay (TSA). The fluorophore Sypro™ Orange dye was used as it is sensitive to the environment and its fluorescence increases when it binds to the hydrophobic regions of a protein upon denaturation. MBD2 is heated from 4 °C to 95 °C, monitoring the fluorescence signal of the dye (**Figure 3A,** left panel). The melting temperature ($T_M$) of the protein is calculated from the derivative of the denaturation curve (**Figure 3A,** right panel). The assay was calibrated using three 12mer duplexes of same DNA sequence but differing in affinity towards MBD2: non-methylated, hemi-methylated and methylated DNA (**Figure 3A**). When MBD2 binds to methylated DNA there is a shift in the $T_M$ of 20°C relative to the free protein, of 16°C when hemimethylated DNA is added and of 5°C with non-methylated DNA. In agreement with the literature (32), these results confirm the preference of MBD2 for methylated DNA and validate the assay.

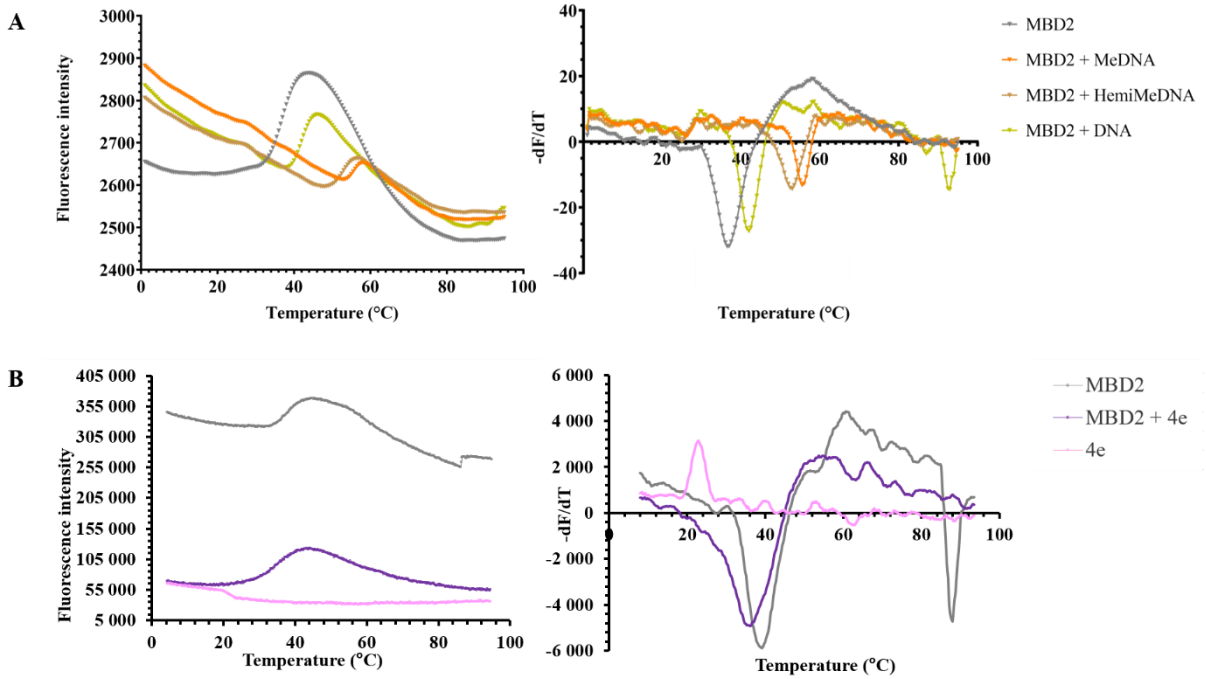
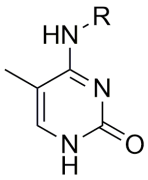


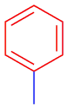
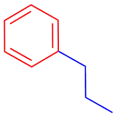
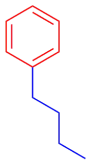
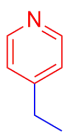

***Figure 3***- *TSA screening assay.* **A**. Validation of the assay. MBD2 (10 µM) was incubated with Sypro™ Orange (10X) alone (grey line) or in the presence of three DNA duplexes (12 µM) with different methylation states, non-methylated DNA (DNA) (light green line), hemi-methylated DNA (hemiDNA) (brown line) and methylated DNA (MeDNA) (orange line). On the left, the denaturing curves are shown and, on the right panel, the derivative to calculate the $T_M$. **B**. MBD2 (10 µM) with 4% DMSO was incubated with Sypro™ Orange (10X) (grey line) or in the presence of compound **4e** (2 mM) (purple line). The denaturation and the derivative curves are shown also for compound **4e** alone (pink line).

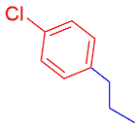
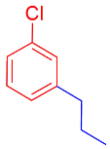
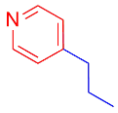
Next, we screened the nine 5mC derivatives at 200-fold the MBD2 concentration (**Table 1**). In this test, the binding of a compound to MBD2 induces a change in the denaturation curve of the protein, as observed for compound **4e** that induces a difference in $T_M$ ($\Delta T_M$) of -3 °C (**Figure 3B**). None of the compounds emitted a fluorescence signal at the wavelength of interest (data not shown). Compounds **4c**, **4e** and **4f** showed the highest shifts in $\Delta T_M$, indicating that the compounds interact with MBD2.

**Table 1**- *Screening results*. The compounds are described by the number of carbons of the linker and the substituent R on the amine group in C4 position of the 5mC. The melting temperature shift ($\Delta T_M$, °C) is reported for each compound as measured by TSA with a ratio compound/protein (Cpd/P) of 200. The concentration of MBD2 is 10 µM. The results of the AlphaScreen® assay are reported as percentage of DNA-MBD2 binding remaining at a ratio of Cpd/P = 10,000. The concentrations of DNA and MBD2 are 50 nM.




**4**

| Compound | Linker length (number of carbon $n_C$) | R | Binding MBD2 | Disruption MBD2/MeDNA |
|---|---|---|---|---|
| | | | $\Delta T_M$ (°C) (ratio Cpd/P = 200) | % MBD2 linked to DNA (ratio Cpd/P = 10000) |
| **4a** | $n_C = 0$ |  | 1.0 ± 0 °C | 72 ± 23% |
| **4b** | $n_C = 1$ |  | 0.9 ± 0.6 °C | 84 ± 14% |
| **4c** | $n_C = 2$ |  | 2.3 ± 1.3 °C | 78 ± 17% |
| **4d** | $n_C = 3$ |  | 1.7 ± 2 °C | 78 ± 7% |
| **4e** | $n_C = 4$ |  | 3.0 ± 1.4 °C | 64 ± 27% |
| **4f** | $n_C = 1$ |  | 2.8 ± 2.9 °C | 71 ± 21% |

| | | | | |
|---|---|---|---|---|
| **4g** | $n_C = 2$ |  | 1.0 ± 0.4 °C | 83 ± 11% |
| **4h** | $n_C = 2$ |  | 0.5 ± 0.7 °C | 86 ± 12% |
| **4i** | $n_C = 2$ |  | 0.5 ± 0.7 °C | 81 ± 26% |

In parallel, we screened the compounds for their ability to disrupt the MBD2-MeDNA complex. We set-up an AlphaScreen® assay using a methylated 12mer DNA duplex bearing in 5' a biotin (biot-MeDNA) that binds to the streptavidin donor beads of the AlphaScreen® technology. The histidine-tagged MBD2 domain is recognized by the nickel chelate acceptor beads of the assay and is added at a 1:1 ratio to biot-MeDNA (Supplementary Fig. S1A). The excitation of the donor beads at 680 nm generates singlet oxygen that excites the acceptor beads only if the donor and acceptor beads are at a maximal distance of 200 nm. Thus, only the binding of MBD2 to biot-MeDNA results in a fluorescence signal at 570 nm. Competition experiments with unlabeled methylated, hemi-methylated and non-methylated DNA validated the assay. (Supplementary Fig. S1B). The assay was also validated with compound KCC-07 that it is described to interfere with the binding of MBD2 to methylated DNA (23) (Supplementary Fig. S2A)

The nine molecules were tested at 10,000-fold the MBD2-MeDNA concentration (**Table 1**). The most potent compound was **4e**, followed by compounds **4f** and **4a**. Interestingly, compounds **4e** and **4f** gave a positive signal in both assays. Of note, compounds **4c**, **4d**, **4e**, and **4f** did not inhibit DNMT3A/3L in an enzymatic assay (Supplementary Fig. S3A) nor demethylated a methylated promoter in cells, as monitored by the reactivation of the luciferase gene under the control of a methylated CMV-promoter (Supplementary Fig. S3B). Finally, the compounds were not cytoxic in HEPG2 cells, an hepatocarcinoma cell line (Supplementary Fig. S4). We thus pursued the study of the interaction of the compounds with MBD2 by NMR.

*Assignment, dynamics and secondary structure of MBD2 in solution.* Backbone (H, N, C, CA) and CB NMR chemical shifts of MBD2 were assigned using $^{1}H$-$^{15}N$ correlation spectra to monitor the binding and map the interaction interface of the ligands on the MBD2 structure. $^{1}H$-$^{15}N$ correlation spectra of

MBD2 showed 68 out of the 71 expected backbone amide signals (Supplementary Fig. S5). However, twelve of these resonances were very large, did not show correlations in $^{1}H/^{15}N/^{13}C$ 3D spectra and could thus not be assigned. Moreover, 6 resonances gave very weak peaks in $^{1}H$-$^{15}N$ correlation spectra but could be assigned from the corresponding correlations in 3D experiments. The observed line broadening most likely arose from conformational exchange at the µs-ms time scale, hampering the assignment of 15 amide resonances (79% coverage) and revealing important protein dynamics of the free protein. These results are in agreement with the literature describing that the binding to methylated DNA stabilizes and structures MBD2 (33) and with the absence of structure of unbound MBD2 in the PDB depository. To compare the structure of isolated MBD2 to the one in complex with methylated DNA, and to evaluate its internal dynamics on the ns-ps time scale, we determined the secondary structure of MBD2 and its order parameters ($S^2$) from the backbone and CB chemical shifts using Talos-N and the random coil index (RCI) method (Supplementary Fig. S6). Unbound MBD2 displayed the three-stranded ß-sheet (B1: 160-165, B2: 173-181; B3: 184-187) and the α-helix as in complex with DNA, although strands B1 and B3 were shorter according to the available chemical shifts. In solution, MBD2 showed moderate to high amplitude motions in the N-terminal region (146-159), ß-strands B1 (161-164) and B3 (184-186), the B1-B2 loop (167-173) that interact with methylated DNA, and in the C-terminal tail (204-220). Most of the unassigned amides were in these dynamic regions that contain basic residues that interact with DNA. Together, these results indicated that in solution MBD2, effectively, displays the same secondary structure that is observed in complex with DNA, but is highly dynamic, with internal motions in the ns-ps time scale and conformational exchanges on the µs-ms time scale.

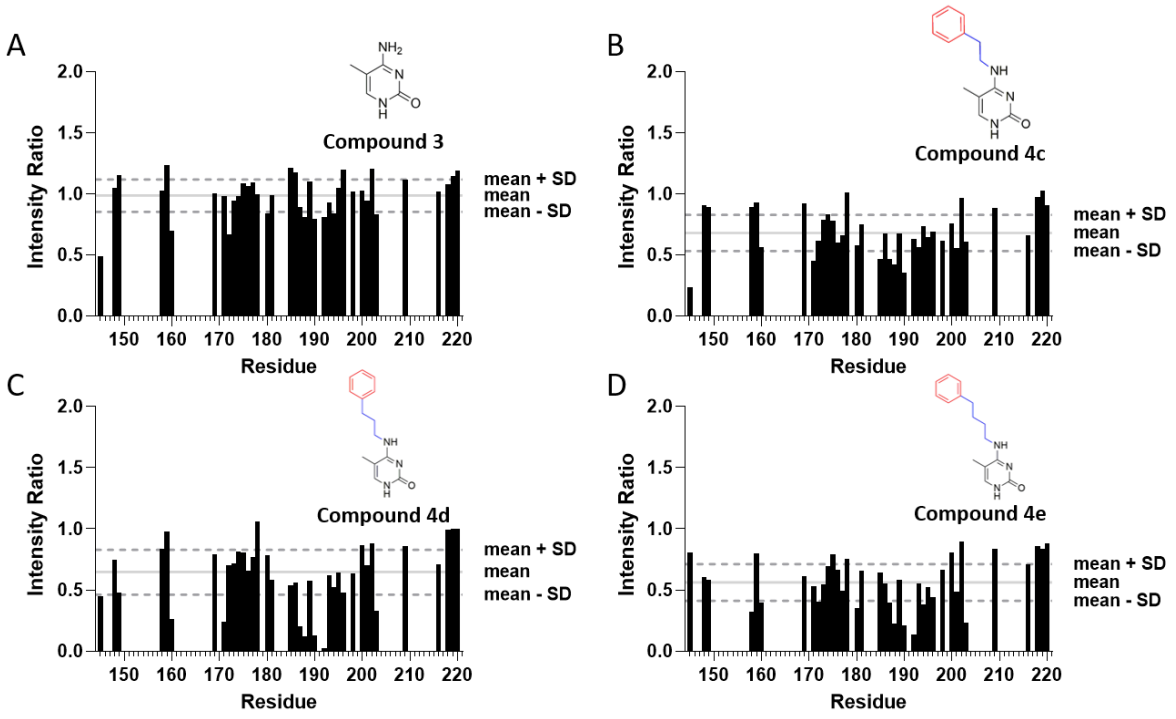



***Figure 4**- Relative intensities in $^{1}H$-$^{15}N$ (SOFAST) correlation spectra of MBD2 as a function of residue,* in the presence of compound **3** (**A**), **4c** (**B**), **4d** (**C**) and **4e** (**D**).

*Binding assays monitored by NMR.* The $^{1}H$-$^{15}N$ correlation spectra of MBD2, which can be viewed as the fingerprint of the structure and dynamics of a protein, were used to monitor the interaction of the compounds (Supplementary Fig. S7). The interaction of compounds **3**, **4c**, **4d** and **4e** were detected by NMR and we analyzed the effect of the linker arm and identified the binding sites. The residues involved in the interaction were identified monitoring the chemical shift perturbations (CSPs) (Supplementary Fig. S8) and the changes in relative intensity induced by the presence of the compounds (80-fold excess**)** in MBD2 $^{1}H$-$^{15}N$ SOFAST spectra **(Figure 4** and **Table 2).** The effects of the compounds were modest for CSP and somewhat larger for the relative intensities. The comparison of CSP values and relative intensities suggests that **4e** is the best binder. Interestingly, mostly the same residues are affected in their CSP and relative intensity values by the addition of the compounds. These residues are mainly localized in a cleft binding one of the 5mCs of the DNA. Additionally, a single residue is found in the B2-B3 loop that interacts with the backbone of 5mCG in DNA (**Figure 5**). As this loop is far away from the main interaction site detected by NMR in the known MBD2:DNA structures, this result suggests that either the compounds can -bind in two different locations of the protein, or the B2-B3 loop could be closer to the main identified binding region or an allosteric effect.

**Table 2**- *Binding analysis by NMR and docking*. The effects caused by binding of the compounds are reported as mean intensity ratios ([MBD2+compound]/MBD2) and mean CSP. The affinity estimate for the most stable binding conformation is established by *in silico* docking.

| Compound | Mean intensity ratio | Mean CSP (ppm) | Affinity (Docking) (Kcal/mol) |
|---|---|---|---|
| **3** | 0.99 | 0.0048 | -4.15 |
| **4c** | 0.68 | 0.0095 | -5.63 |
| **4d** | 0.65 | 0.0086 | -5.52 |
| **4e** | 0.56 | 0.0090 | -5.65 |

To gather further insights on the interaction of the compounds with MBD2, we carried out unconstrained docking and visualized the best binding conformers for the four compounds in the structure of human MBD2 in complex with methylated DNA (PDB code 6CNQ) (**Figure 5**). In the best docking conformers, compounds **4c**, **4d** and **4e** are located in a long cleft at the interface of the MBD2-MeDNA complex. This is in agreement with the main compound-MBD2 interaction interface mapped by NMR. Nevertheless, the orientation of the 5mC moiety in the compounds and in the DNA are different.

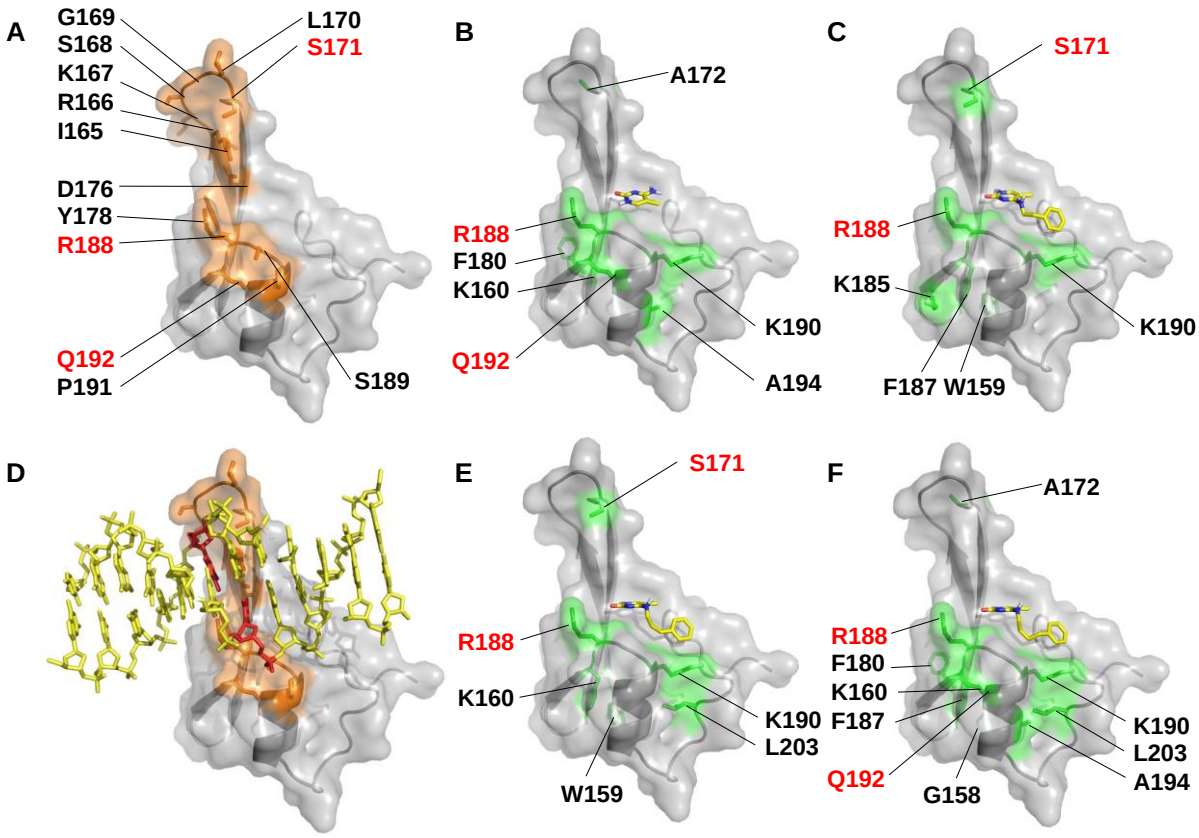



***Figure 5****-Structure of MBD2 (PBD: 6CNQ), NMR interface mapping and docking of the compounds* (**A,D)**: Structure of MBD2 in complex with MeDNA. For visualization purposes, in (**A**), the structure of MBD2 is shown without DNA. In **D**, the structure of DNA is shown in yellow, the two 5mCs are displayed in red and the surface of MBD2 close to the two 5mCs is highlighted in orange. (**B, C, E, F):** Best-energy docking pose and MBD2 residues (green) showing significant intensity loss in NMR upon compound addition (< mean - SD intensity ratio) for compounds **3** (**B**), **4c** (**C**), **4d** (**E**) and **4e** (**F**) Residues close to 5mC or showing significant intensity loss in NMR are labelled in black, or highlighted in red, if they are both close to 5mC and showed significant intensity loss in NMR upon compound addition.

*Nucleoside derivative of compound **4e***. As mentioned in the introduction, we left the possibility to synthetise the corresponding nucleoside of the most potent fragment identified in the assays. Thus, the nucleoside analogue **5e** of compound **4e** was synthetized by enzymatic glycosylation using *N*-deoxyribosyltransferase (NDT) from *L. leichmannii* as described in (34) (**Figure 6A**).

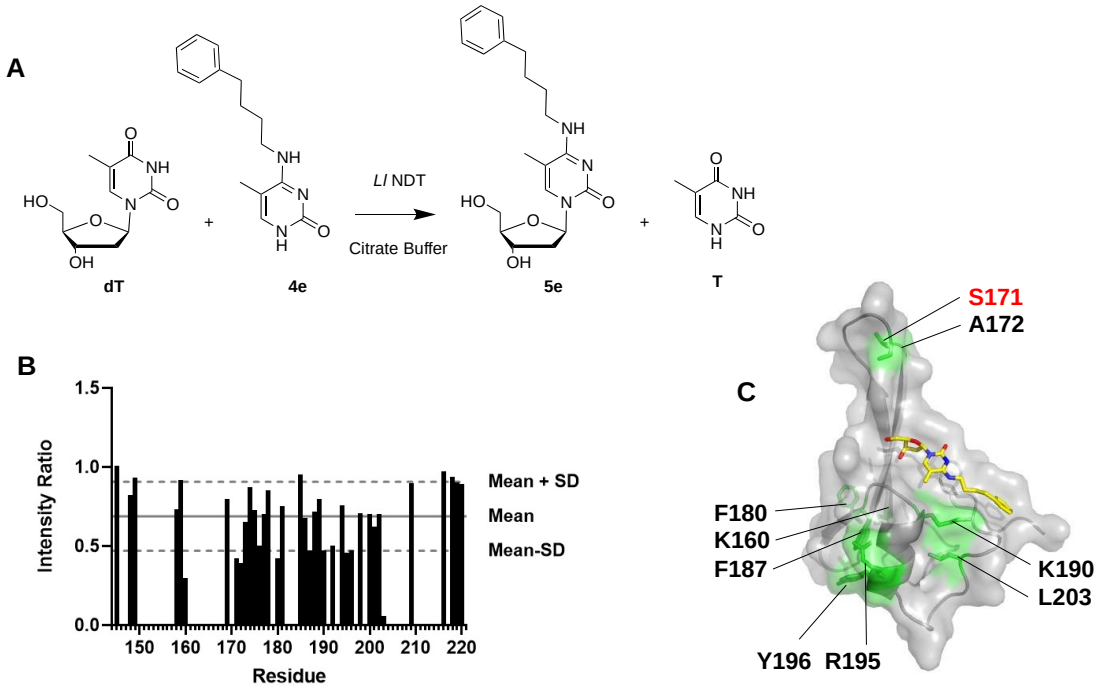



***Figure 6***- (**A)** *Enzymatic synthesis of compound* **5e**: the compound was obtained by enzymatic glycosylation using *N*-deoxyribosyltransferase (NDT) from *L. leichmannii*. (**B)** *NMR interface mapping of compound* **5e:** Relative intensities in $^{1}H$-$^{15}N$ (SOFAST) correlation spectra of MBD2 as a function of the residue in the presence/absence of compound **5e**. (**C)** *Docking of compound* **5e** in the structure of MBD2 (PDB: 6CNQ) in complex with MeDNA. Best-energy docking pose of the compound (yellow) are reported together with MBD2 residues (green) showing significant intensity loss in NMR upon compound **5e** addition (< mean - SD intensity ratio). Residues showing significant intensity loss in NMR signal are labelled in black, or highlighted in red, if they are close to 5mC.

Compound **5e** did not show an improvement in the interaction with MBD2 (≙Tm= 1.7°C± 0.8) and lost the ability to interfere with the MBD2/MeDNA complex (9.8%± 4.5 decrease of signal in the AlphaScreen® assay). Although compound **5e** shows a high effect on the intensity of some residue signals, the signal loss induced in the MBD2 $^{1}H$-$^{15}N$ correlation spectra (**Figure 6 B** and **C**) is overall milder than the one observed for the corresponding modified base **4e**, as shown by the respective mean intensity ratios of 0.69 ± 0.22 (**5e**) and 0.56 ± 0.21 (**4e**). Importantly, residues R188 and Q192, which are in the vicinity of MeC in the MBD2/MeDNA complex structure, show a significant signal loss in the presence of compound **4e** but not of compound **5e**, suggesting that **4e** is better positioned than **5e** to interfere with DNA binding. In contrast, residues R195 and Y196 showed significant intensity loss, comparable to K190, highlighted for compounds **4b, c, d** and **e**. 2D NMR and docking of **4e** and **5e** suggest that the phenyl ring (on the $N^4$) anchors the compounds in a cleft close to the DNA binding site by pi-cation interaction. Interestingly, the best docking poses for **5e** is located in the same cleft as **4e** (**Figure 6C**) and shows an equivalent binding energy (-6.5 vs -5.7 kCal/mol),
Thus, the addition of the deoxyribose on **4e** did not bring the expected increase in the binding to MBD2 and the NMR experiments suggests that compound **5e** adopts a configuration that less interferes with the binding to DNA, as also observed in the AlphaScreen assay.

## Conclusions

In conclusion, we set-up two orthogonal screening assays, one monitoring the interaction of the compounds with MBD2 and one the ability of the compounds to disrupt the binding of MBD2 to methylated DNA. The TSA is the first screening that allows to identify compounds that directly bind to the protein. The Z' factor, evaluated according to *Zhang et al*., (35) is 0.53 that indicates an optimized assay efficient for screening a small chemical library. To monitor the interference with the MBD2/MeDNA complex we developed an AlphaScreen® assay based on a robust technology providing

better sensitivity, dynamic range and is less prone to potential fluorescence interference that can come from the compounds. (36) The assay confirmed that compound KCC-07, used in cancer cells in the literature, (23) interacts with the binding of MBD2 to methylated DNA.

We synthesized nine derivatives of 5mC, which are substituted on the amine group in position C4. We confirmed a weak interaction of MBD2 with ~~the~~ 5mC (**3**) alone. Different alkyl linker arms for the phenyl group were explored by comparing **4c**, **4d** and **4e**. Although no major difference was observed, the trend is the longer the linker, the stronger the effects, as indicated by the relative intensities in NMR and the AlphaScreen® assay. Two compounds were active in both screening assays: **4e** and **4f**. The most potent was compound **4e** bearing a butylphenyl group on the amine at the C4 position of the 5mC.

To validate the data, we used for the first time NMR experiments to follow the interaction of the compounds with MBD2 and identify the amino acid residues of the protein involved in the interaction. First, we established by NMR that although unbound MBD2 is highly dynamic, its secondary structure is the same as the one observed in complex with DNA (Supplementary Fig. S4). Second, we identified the residues of MBD2 that are affected upon compound binding. Most interesting, the residues affected by the compounds partially overlap one region involved in the interaction with 5mC in DNA. Independent docking without constraints was in agreement with the main interaction sites detected by NMR.

The results validate our hypothesis that derivatives of 5mC are good fragments for the design of chemical probes for MBD2. The NMR experiments suggest that the compounds interact with MBD2 in the cleft and the loop where the two 5mC of DNA are positioned. The comparison by NMR of compound **4e** and its nucleoside analogue **5e** suggest that a good strategy to follow is to synthesize compounds that anchor into the cleft, are longer and less flexible to better bind to MBD2 and compete with DNA. These compounds constitute a good starting point for the development of-ligands of MBD2 and further studies are ongoing. A chemical probe binding MBD2 will be of great use to understand its role in the interpretation of aberrant DNA methylation profiles in pathologies. The chemical targeting of MBPs is a new avenue that could lead to a more specific effect (16) than targeting DNMTs in methylation-dependent pathologies.

## Materials and Methods

### *Chemistry*

Reagents and solvents were purchased from commercial sources and used without further purification. All anhydrous reactions were conducted under an argon atmosphere. Purifications by preparative HPLC were carried out on Agilent 1100 Series system (with a diode-array detector and automatic fraction collector) on a C18 reverse phase column (Kromasil, 5 μm, 100 Å, 10 × 250 mm) using a flow rate of 4.0 mL min−1 and an isocratic or linear gradient of acetonitrile in 10 mM triethylammonium acetate buffer (TEAA) over 20 min. Analytical HPLC was carried out on an Agilent system (1100 series) using a C18 reverse phase column (Kromasil, 5 μm, 100 Å, 4.6 × 150 mm for analytical analysis) at a flow rate of 1 mL min−1 and a linear gradient of acetonitrile in 10 mM triethylammonium acetate buffer over 20 min. Yields refer to chromatographically and spectroscopically pure (>95%) compounds. NMR experiments to characterize the synthesized molecules were recorded on an Agilent DirectDrive 500 spectrometer (Agilent Technologies, Santa Clara) with a proton resonating frequency of 499.8 MHz. Spectra were recorded using VnmrJ 4.2A (Agilent Technologies). Coupling constants (J) are in hertz (Hz), and signals are designated as follows: s, singlet; d, doublet; t, triplet; m, multiplet; br, broad singlet, etc. Chemical shifts are given in ppm (δ). High-resolution mass spectra were recorded on a Waters Q-TOF micro MS instrument under electrospray ionization in the positive ionization mode using a mobile phase of acetonitrile/water with 0.1% formic acid.

Class II nucleoside 2-deoxyribosyltransferase from *L. leichmannii* (LlNDT-II or NDT) was produced and purified as described. (37)

Synthesis of **4-triazol-1-ylthymine (2):**

To a solution of 1,2,4-triazol (13.9 g, 202 mmol, 7 eq.), 6 mL of $POCl_3$ (66.4 mmol, 2.3 eq.) were added at 0 °C followed by a dropwise addition of 28 mL of triethylamine (202 mmol, 7 eq.). The reaction mixture was stirred at 0 °C for 30 min, then at room temperature for 30 min. Thymine (3.64 g, 28.8 mmol) was added and the heterogeneous mixture was stirred at room temperature overnight. Then the mixture was refluxed for 5 h. The mixture was diluted with water (100 mL) and filtered. The precipitate was washed three times with water and once with a minimum of EtOH. The white product was recovered with 52% yield (2.6 g). **HRMS-ESI** (m/z) calculated for $[C_7H_7N_5O+ H]^+$: 178.0684; found: 178.0723. **$^1$H NMR** (500MHz, DMSO-$d_6$): $\delta$ 2.25 (s, 3H), 8.10 (s, 1H), 8.34 (s, 1H), 9.27 (s, 1H), 12.27 (br. s, 1H, 1 NH). **$^{13}$C NMR** (125 MHz, DMSO-$d_6$): $\delta$ 153.9, 152.2, 145.8, 144.1, 132.3, 104.4, 15.8.

Synthesis of **5-methylcytosine (3):**

To a solution of 7 N ammonia in methanol (25 mL), 4-triazol-1-ylthymine (1.46 g, 8.28 mmol) were added and the reaction mixture was stirred at room temperature for 3.5 h. The mixture was refluxed at 90 °C for 2 h and stirred again at room temperature overnight. After concentration, the residue was dissolved in hot EtOH (the solid was filtered out). Cold acetone (30 mL) was then added at room temperature and the formed precipitate collected by filtration to obtain compound **3** as a yellowish powder (632 mg, 61%). **HRMS-ESI** (m/z) calculated for $[C_5H_8N_3O + H]^+$: 126.0623; found: 126.0664. **$^1$H NMR** (500 MHz, DMSO-$d_6$): $\delta$ 7.16 (d, $J$ = 1.3 Hz, 1H, H-6 cytosine), 3.17 (s, 2H, $NH_2$), 1.79 (d, $J$ = 1.1 Hz, 3H, $CH_3$). **$^{13}$C NMR** (125 MHz, DMSO-$d_6$): $\delta$ 166.4, 156.9, 139.9, 99.6, 12.9.

Synthesis of compound ***N*$^4$-phenyl-5-methylcytosine (4a):**

To a suspension of NaH (56 mg, 2.1 mmol, 2.2 eq.) in dry DMF (5 mL) was added aniline (270 µL, 3.8 mmol, 4 eq.) and the suspension was stirred until the bubbling stopped, then 4-triazol-1-ylthymine **2** (170 mg, 0.96 mmol) was added and the mixture was heated at 80 °C for 33 h. After quenching with a few drops of water, volatiles were evaporated and the crude was purified on silica gel chromatography to give compound **4a** (150 mg, 78%) as a light pink powder. **HPLC** (gradient 0-60) $t_R$= 9.45 min. **HRMS-ESI** (m/z) calculated for $[C_{11}H_{12}N_3O + H]^+$: 202.0936; found: 202.0975. **$^1$H NMR** (500 MHz, DMSO-*d*6): $\delta$ 10.55 (s, 1H, NH), 8.39 (s, 1H, H 6-cytosine), 7.77 – 7.68 (m, 2H, H-phenyl), 7.36 – 7.28 (m, 3H, H-phenyl), 7.08 (tt, $J$ = 7.4, 1.2 Hz, 1H, NH), 2.00 (d, $J$ = 1.1 Hz, 3H, $CH_3$). **$^{13}$C NMR** (126 MHz, DMSO-*d*6): $\delta$ 162.6, 156.2, 140.7, 139.0, 128.1, 123.6, 122.8, 100.8, 13.1.

General procedure for the synthesis of compounds **4** (except **4a**):

To a solution of **4-triazol-1-ylthymine** (**2**) (100 mg, 0.7 mmol) in dry DMF (3 mL) under argon, triethylamine (119 µL, 0.85 mmol, 1.5 eq) and the desired amine derivative (0.85 mmol, 1.5 eq.) were added. The mixture was stirred at room temperature overnight then concentrated to dryness. The residue was purified by flash chromatography on silica gel using a linear gradient of MeOH in DCM (0-10%) to obtain the desired product.

Compound ***N*$^4$-benzyl-5-methylcytosine (4b)**:

Obtained as a white powder (82 mg, 67%, from benzylamine). **HPLC** (gradient 0-40) $t_R$= 13.66 min. **HRMS-ESI** (m/z) calculated for $[C_{12}H_{14}N_3O + H]^+$: 216.1137; found 216.1134. **$^1$H NMR** (500 MHz, DMSO-*d*6): $\delta$ 10.19 (s, 1H, NH), 8.14 (s, 1H, NH), 7.51 (d, $J$ = 5.6 Hz, 1H, H-4 benzyl), 7.35 – 7.19 (m, 4H, H-2, H-3, H-5, H-6 benzyl), 7.16 (s, 1H, H-6 cytosine), 4.55 (d, $J$ = 4.9 Hz, 2H, $CH_2$), 1.86 (s,

3H, $CH_3$). **$^{13}$C NMR** (126 MHz, DMSO-*d*6): $\delta$ 163.8, 163.0, 156.5, 139.7, 139.2, 128.1, 127.0, 126.5,100.0, 42.9, 12.7.

Compound **5-methyl-*N*$^4$-(phenethyl)cytosine (4c)**:

Obtained as white powder (82 mg, 63%, from phenethylamine). **HPLC** (gradient 0-60) $t_R$= 12.19 min. **HRMS-ESI** (m/z) calculated for $[C_{13}H_{16}N_3O + H]^+$: 230.1249; found: 230.1288. **$^1$H NMR** (500 MHz, DMSO-*d*6): $\delta$ 10.18 (s, 1H, NH), 7.30 (t, $J$ = 7.5 Hz, 2H, H-2, H-6 phenyl), 7.25 – 7.16 (m, 3H, H-3, H-4, H-5 phenyl), 7.12 (d, $J$ = 1.3 Hz, 1H, H-6 cytosine), 7.04 (t, $J$ = 5.7 Hz, 1H, NH), 3.51 (ddd, $J$ = 9.1, 7.6, 5.8 Hz, 2H, $CH_2$), 2.84 (dd, $J$ = 8.6, 6.5 Hz, 2H, $CH_2$), 1.78 (d, $J$ = 1.1 Hz, 3H, $CH_3$). **$^{13}$C NMR** (126 MHz, DMSO-*d*6): $\delta$ 163.8, 156.7, 139.7, 138.9, 128.6, 128.3, 126.0, 100.0, 41.6, 34.6, 12.7.

Compound **5-methyl-*N*$^4$-(3-phenylpropyl)cytosine (4d)**:

Obtained as a white powder (53 mg, 39%, from 3-phenyl-1-propylamine). **HPLC** (0-60) $t_R$= 13.61 min. **HRMS-ESI** (m/z) calculated for $[C_{14}H_{18}N_3O + H]^+$: 244.1450; found 244.1449. **$^1$H NMR** (500 MHz, DMSO-*d*6): $\delta$ 8.27 (s, 1H, NH), 7.28 (t, $J$ = 7.5 Hz, 2H, H-2, H-6 phenyl), 7.24 – 7.19 (m, 2H, H-3, H-5 phenyl), 7.19 – 7.14 (m, 1H, H-1 phenyl), 7.09 (d, $J$ = 1.4 Hz, 1H, H-6 cytosine), 6.91 (t, $J$ = 5.6 Hz, 1H, NH), 3.35 – 3.28 (m, 2H, $CH_2$), 2.61 (t, 2H, $J$ = 7.7 Hz, $CH_2$), 1.84 (tt, $J$ = 8.6, 6.6 Hz, 2H, $CH_2$), 1.79 (d, $J$ = 1.1 Hz, 3H, $CH_3$). **$^{13}$C NMR** (126 MHz, DMSO-*d*6): $\delta$ 163.8, 156.6, 141.8, 138.7, 128.2, 125.7, 100.0, 48.6, 32.7, 30.6, 30.3, 12.7.

Compound **5-methyl-*N*$^4$-(4-phenylbutyl)cytosine (4e)**:

Obtained as a white powder (105 mg, 72%, from 4-phenylbutylamine). **HPLC** (gradient 0-60) $t_R$= 15.14 min **HRMS-ESI** (m/z) calculated for $[C_{15}H_{20}N_3O + H]^+$: 258.1562; found: 258.1601. **$^1$H NMR** (500 MHz, DMSO-*d*6): $\delta$ 10.09 (s, 1H, NH), 7.30 – 7.24 (m, 2H, H-2, H-6 phenyl), 7.23 – 7.18 (m, 2H, H-3, H-5 phenyl), 7.18 – 7.13 (m, 1H, H-4 phenyl), 7.09 (s, 1H, H-6 cytosine), 6.93 (t, $J$ = 5.7 Hz, 1H, NH), 3.31 (q, $J$ = 6.5 Hz, 2H, $CH_2$), 2.60 (t, $J$ = 7.3 Hz, 2H, $CH_2$), 1.78 (d, $J$ = 1.1 Hz, 3H, $CH_3$), 1.61 – 1.49 (m, 4H, $CH_2$-$CH_2$). **$^{13}$C NMR** (126 MHz, DMSO-*d*6) $\delta$ 163.8, 156.7, 142.2, 138.7, 128.3, 125.6, 100.0, 39.7, 34.9, 28.5, 28.4, 12.7.

Compound **5-methyl-*N*$^4$-(4-methylpyridine)cytosine (4f):**

Obtained as white powder (116 mg, 94%, from 4-(aminomethyl)pyridine). **HPLC** (gradient 0-20) $t_R$= 10.50 min. **HRMS-ESI** (m/z) calculated for $[C_{11}H_{13}N_4O + H]^+$: 217.1045; found: 217.1084. **$^1$H NMR** (500 MHz, DMSO-*d*6): $\delta$8.50 – 8.46 (m, 2H, H-2, H-6 pyridine), 7.61 (t, $J$ = 6.1 Hz, 1H, NH), 7.28 – 7.23 (m, 2H, H-3, H-5 pyridine), 7.19 (s, 1H, H-6 cytosine), 4.55 (d, $J$ = 5.9 Hz, 2H, $CH_2$), 1.89 (d, $J$ = 1.1 Hz, 3H, $CH_3$). **$^{13}$C NMR** (126 MHz, DMSO-*d*6) δ 164.0, 156.5, 149.4, 148.8, 139.5, 122.0, 100.0, 42.2, 12.7.

Compound ***N*$^4$-(4-chlorophenethyl)-5-methylcytosine (4g)**:

Obtained as a white powder (82 mg, 55%, from 2-(4-chlorophenyl)ethylamine). **HPLC** (gradient 0-60) $t_R$= 14.30 min. **HRMS-ESI** (m/z) calculated for $[C_{13}H_{15}ClN_3O + H]^+$: 264.0904; found 264.0900. **$^1$H NMR** (500 MHz, DMSO-*d*6): $\delta$ 10.19 (s, 1H, NH), 7.40 – 7.29 (m, 2H, H-3, H-5 phenyl), 7.28 – 7.17 (m, 2H, H-2, H-6 phenyl), 7.12 (s, 1H, H-6 cytosine), 7.02 (t, $J$ = 5.7 Hz, 1H, NH), 3.50 (dt, $J$ = 7.9, 6.0 Hz, 2H, $CH_2$), 2.83 (t, $J$ = 7.4 Hz, 2H, $CH_2$), 1.77 (s, 3H, $CH_3$). **$^{13}$C NMR** (126 MHz, DMSO-*d*6): δ 163.8, 156.7, 139.0, 138.7, 130.7, 130.5, 128.2, 100.1, 41.3, 33.8, 12.7.

Compound ***N*⁴-(3-chlorophenethyl)-5-methylcytosine (4h)**:

Obtained as a white powder (34 mg, 23%, from 2-(3-chlorophenyl)ethylamine). **HPLC** (0-60) $t_R$= 14.12 min. **HRMS-ESI** (m/z) calculated for $[C_{13}H_{15}ClN_3O + H]^+$: 264.0829; found: 264.0898. **$^1$H NMR** (500 MHz, DMSO-$d_6$): $\delta$ 10.18 (s, 1H, NH), 7.32 (d, $J$ = 7.6 Hz, 1H, H-6 phenyl), 7.29 – 7.25 (m, 2H, H-2, H-5 phenyl), 7.18 (dt, $J$ = 7.5, 1.4 Hz, 1H, H-4 phenyl), 7.12 (s, 1H, H-6 cytosine), 7.02 (t, $J$ = 5.7 Hz, 1H, NH), 3.51 (dt, $J$ = 7.8, 6.0 Hz, 2H, $CH_2$), 2.85 (t, $J$ = 7.3 Hz, 2H, $CH_2$), 1.77 (d, $J$ = 1.1 Hz, 3H, $CH_3$). **$^{13}$C NMR** (125 MHz, DMSO-$d_6$): $\delta$ 163.8, 156.6, 142.3, 139.0, 132.9, 130.1, 128.5, 127.4, 126.0, 100.0, 41.2, 34.0, 12.6.

Compound ***N*⁴-(4-ethylpyridine)-5-methylcytosine (4i)**:

Obtained as a white powder (82 mg, 63%, from 4-(2-aminoethyl)pyridine). **HPLC** (0-20) $t_R$= 13.54 min. **HRMS-ESI** (m/z) calculated for $[C_{12}H_{15}N_4O + H]^+$: 231.1201; found: 231.1240. **$^1$H NMR** (500 MHz, DMSO-$d_6$): δ 11.80 (s, 1H, NH), 8.74 – 8.69 (m, 2H, H-2, H-6 pyridine), 7.72 - 7.67 (m, 2H, H-3, H-5 pyridine), 7.62 (s, 1H, H-6 cytosine), 3.79 (q, $J$ = 6.8 Hz, 2H, $CH_2$), 3.06 (t, $J$ = 7.2 Hz, 2H, $CH_2$), 1.87 (d, $J$ = 1.1 Hz, 3H, $CH_3$). **$^{13}$C NMR** (125 MHz, DMSO-*d6*): $\delta$ 158.7, 154.5, 149.7, 144.6, 141.9, 126.3, 101.4, 41.5, 33.6, 12.3.

Compound **5-methyl-*N*⁴-(4-phenylbutyl)-deoxycytidine (5e)**:

Enzymatic glycosylation of 4e (17 mg, 0.07 mmol) was performed with 67 mg (0.28 mmol, 4 eq) of thymidine in 6 mL of citrate buffer (10 mM) and 5% of DMSO. The reaction was started by adding NDT (200 µL at 5 µg µL$^{-1}$) and run at 50 °C overnigh. was monitored by analytical reverse phase HPLC. After 18 h the reaction reached equilibrium. The reaction was stopped by heating the reaction mixture at 80 °C for 5 min. After purification by HPLC, compound **5e** was obtained as a white powder (17.2 mg, 69% from 4e). **HPLC** (0-60) $t_R$= 14.47 min. **HRMS-ESI** (m/z) calculated for $[C_{20}H_{28}N_3O_4 + H]^+$: 374.2074; found: 374.2074. **$^1$H NMR** (400 MHz, DMSO-$d_6$) $\delta$: 7.56 (d, $J$ = 1.0 Hz, 1H, H cytosine), 7.30-7.24 (m, 2H, H-2, H-6 phenyl), 7.23-7.13 (m, 3H, H-3, H-4, H-5 phenyl), 7.10 (t, $J$ = 8.8 Hz, 1H, NH), 6.17 (dd, $J$ = 6.2, 7.6 Hz, 1H, H-1'), 5.14 (d, $J$ = 4.2 Hz, 1H, OH-3'), 4.95 (t, $J$ = 5.4 Hz, 1H, OH-5'), 4.20 (ddd, $J$ = 3.7, 6.4, 9.8 Hz, 1H, H-3'), 3.75 (dd, $J$ = 3.9, 7.1 Hz, 1H, OH-5'), 3.62-3.50 (m, 2H, H-5', H-5"), 3.36-3.30 (m, 2H, $\underline{CH_2}$NH), 2.60 (t, $J$ = 7.4 Hz, 2H, $\underline{CH_2}$-phenyl), 2.06 (ddd, J = 2.6, 6.0, 12.9 Hz, 1H, H-2"), 1.95 (ddd, J = 6.0, 7.7, 12.9 Hz, 1H, H-2″), 1.84 (d, $J$ = 0.9 Hz, 3H, $CH_3$), 1.64-1.48 (m, 4H, 2 $CH_2$); **$^{13}$C NMR** (151 MHz, DMSO-$d_6$) $\delta$: 163.1, 155.5, 142.6, 137.8, 128.8, 128.7, 126.1, 102.0, 87.5, 85.0, 70.9, 61.9, 40.6, 35.4, 28.9, 28.8, 13.6.

***Expression and purification of MBD of MBD2***

*GST-MBD2 fusion*

The methyl-CpG binding domain sequence of MBD2 (670-889 bp) was inserted in the pGEX-6p-2 vector from Addgene into the BamH1 and Xho1 sites. In addition to MBD2, the fusion protein contains a N-terminal Gluthatione S-Transferase (GST) tag for purification followed by a PreScission protease site. For expression, *E. coli* BL21 (CodonPlus(DE3) – Agilent) RIL cells transformed with the pGEX-6p-2-GST-MBD2 plasmid, were grown in Luria-Bertani medium supplemented with 100 µg/mL ampicillin at 37 °C to an $OD_{600}$ of 0.6 and induced overnight with 1 mM IPTG at 20 °C. The collected cells were resuspended in 20 mM HEPES pH 7.5, 500 mM KCl, 0.2 mM DTT, 1 mM EDTA and 10% glycerol and disrupted by sonication. The supernatants were passed through glutathione Sepharose 4B (Protino® GST/4B 5 mL, Macherey-Nagel) and the bound proteins were eluted with 40 mM glutathione

and dialyzed against the cleavage buffer (50 mM Tris.HCl pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) for 2 h. Cleavage was performed by adding 0.3 units of PreScission (GE Healthcare) enzyme to 7.3 mg of fusion protein and incubated overnight at 4 °C. The cleaved proteins were passed through a cationic exchange column (HiPrep SP PH 16/10 from GE Healthcare) and the bound proteins were eluted with a NaCl gradient from 150 mM to 1 M. The purified MBD protein was dialyzed overnight at 4 °C against 20 mM ammonium bicarbonate pH 8, lyophilized and kept at 4 °C. Expression and final purity of samples were checked by SDS-PAGE. The resulting protein consists of residues 146-220 with a 4 residue N-terminal extension (GPLG) due to cloning.

The protocol to express and purify $^{15}N$ or $^{15}N$ and $^{13}C$ labelled MBD2 was similar but bacteria were grown in minimum M9 media supplemented with $^{15}NH_4Cl$ or/and $^{13}C$ glucose as the only source of nitrogen and carbon, respectively.

*MBD2-His*

The methyl-CpG binding domain sequence of MBD2 (670 - 889 bp) was inserted in the pET43.1a vector from Genscript into the HindIII and Nde1 sites. In addition to MBD2, the fusion protein contains a C-terminal Histidine (His) tag for purification and AlphaScreen® assay. For expression, *E. coli* BL21 (CodonPlus(DE3) – Agilent) RIL transformed with plasmid pET43.1a-MBD2-His were grown in Luria-Bertani medium supplemented with 100 µg/mL ampicillin at 37 °C to an $OD_{600}$ of 0.6, then induced overnight with 1 mM IPTG at 20 °C. The collected cells were resuspended in 50 mM $Na_2PO_4$ pH 7.5, 300 mM NaCl, 20 mM imidazole, 0.2 mM DTT and 10% glycerol and disrupted by sonication. The supernatant was passed through a HisTrap FF (5 mL, GE Healthcare) column and the bound proteins were eluted by a gradient of imidazole up to 500 mM. The his-tagged protein was further purified by gel filtration using a HiPrep 16/60 Sephacryl S-100 High Resolution (GE Healthcare Life Sciences) column equilibrated and eluted in the assay buffer (20 mM Tris pH 7.4, 125 mM NaCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 0.5 mM DTT and 4% glycerol). The purified MBD2 protein was flash frozen with dry ice and ethanol and stored at −20 °C. Expression and final purity of samples were checked by SDS-PAGE.

***Biophysical and structural assays***

*Hybridization of the T4-MeHairpin DNAs*

Hairpin forming oligonucleotides with the sequence 5'-GCCTTmCGGTGGCTTTTGCCACmCGAAGGC-3′ named MeDNA, or 5′-biotin-GCCTTmCGGTGGCTTTTGCCACmCGAAGGC-3′ named biot-MeDNA were annealed by heating to 95 °C for 5 min and cooling slowly to room temperature.

*Protein Thermal Shift Assay (TSA)*

Melting curves were determined in transparent 96 well plates using a temperature increment of 0.05 ˚C from 4 ˚C to 95 ˚C on the Applied Biosystems QuantStudio™ 6 Flex system. Sypro™ Orange dye (Invitrogen) was used to label MBD2 protein at a final 10x dilution (from 5,000x stock solution in DMSO). Ligand were tested at 200-fold excess concentration (2 mM) of protein (10 µM) in 20 mM HEPES pH 8, 120 mM KCl, 1 mM $MgCl_2$, 0.2 mM DTT, 1 mM EDTA, 35 ng/µL Carrier RNA, 4% DMSO. Twenty-five microliters of each sample were loaded on a 96 wells plate (AB0900 semi skirted, Thermo Scientific). Each reaction was repeated on at least two different plates. MBD2 alone and MBD2 with MeDNA (12 µM) were included in each plate as controls. The $T_M$ was determined at the lowest point of the first derivative of the melting curve of a sample.

*AlphaScreen® screening assay*

MBD2-His and biot-MeDNA complex interaction inhibition was assayed by the AlphaScreen® (amplified luminescent proximity homogenous assay) technology using an AlphaScreen® no-wash assay

kit containing Streptavidin Donor beads and nickel chelate (Ni-NTA) AlphaScreen® Acceptor beads (PerkinElmer Inc. ref: 6760619C). The assay was performed in 384-wells black opaque plates (general Corning™, ref: 3575) utilizing a final volume of 25 µL/well in 50 mM HEPES pH 8, 120 mM KCl, 1 mM $MgCl_2$, 1 mM EDTA, 0.2 mM DTT, 0.1 mg/mL BSA, 4% glycerol, 0.05% Tween 20. In each well 5 µL of assay buffer, 5 µL of compound (2.5 mM) and 5 µL of MBD2-His (250 nM) were added. The plate was covered with a dark lid and incubated at 4 °C in the dark for 15 min. Then 5 µL of Biotin-hairpinMeDNA (250 nM) and of a mix of Streptavidin Donor beads and Nickel chelate Acceptor beads were added (25 µg/mL of each bead). The alpha signal was read after 30 min of incubation at room temperature in the dark on a PerkinElmer EnVision® multimode microplate reader with shaking (speed 200 rpm). In every test plate, the following controls were assayed: duplicate samples of MBD2-His (50 nM) and biot-MeDNA (50 nM) and the two beads (5 µg/mL for each bead), duplicate samples of MBD2-His (50 nM) and biot-MeDNA (50 nM) and the two beads (5 µg/mL for each bead) with DMSO (1% or 2%), one well containing the Streptavidin donor beads (5 µg/mL), one the Nickel chelate acceptor beads (5 µg/mL), and one well containing the assay buffer only. The percentage of biot-MeDNA-MBD2-His complex formation was calculated from the ratio of the fluorescence average value of the tested compound over the one of the MeDNA-MBD2 complex.

*NMR assignment and ligand binding studies*

NMR experiments were run on an 800 MHz (Avance NEO) or a 600 MHz (Avance III HD) spectrometers with 18.8 or 14.1 Tesla magnetic fields, respectively. The spectrometers were equipped with a cryogenically cooled triple resonance $^{1}H[^{13}C/^{15}N]$ probe. All experiments were performed at 20°C, Resonance assignment experiments were performed with an $^{15}N/^{13}C$-labelled MBD2 sample at 0.5 mM concentration in 10 mM sodium phosphate buffer pH 6.5, 120 mM NaCl 1 mM TCEP 5% $D_2O$. Ligand binding experiments were performed with MBD2 samples concentrated at 50 µM in 50 mM HEPES pH 7.0, 120 mM NaCl, 1 mM TCEP, 5% $D_2O$. Spectra were recorded using TopSpin 4.07 or 3.6.3 (Bruker BioSpin) and analyzed with CCPNMR 2.5. (38)

MBD2 backbone and CB assignments were performed using 2D $^{1}H$-$^{15}N$ and $^{1}H$-$^{13}C$ HSQCs (39) CBCA(CO)NH (40) and the BEST versions of the 3D HNCA, HNCO, HN(CA)CO and HNCACB pulse sequences implemented from NMRLib. (41) The secondary structure of MBD2 was determined from the N, C, CA, and CB chemical shifts with TALOS-N (42) and compared to the secondary structure of the protein in interaction with methylated DNA (PDB entry 6CNQ) determined with DSSP (43). The internal dynamics (ns-ps time scale) of the protein was obtained from the chemical shifts using the RCI method. (44)

Ligand binding was assessed by exchange broadening and chemical shift perturbation (CSP), which were determined using the $^{1}H$-$^{15}N$ correlation SOFAST-HMQC (SOFAST) experiment on $^{15}N$-labelled MBD2 (50 µM) samples in the absence and presence of 4 mM of the analysed ligand. All samples contained 5% to 10% DMSO in NMR buffer. We checked by NMR ($^{1}H$-$^{15}N$ HSQCs) that 5% or 10% DMSO did not change the structure or internal dynamics of the protein at 20°C. Exchange broadening was determined from the ratio of intensities (I/Io) in the presence (I) or absence (Io) of ligands. The CSP induced by ligand binding was determined from the differences of $^{15}N$ ($\Delta\delta_N$) and $^{1}HN$ ($\Delta\delta H_N$) chemical shifts of MBD2 in the presence or absence of ligands the free and bound forms using the formula:

$$CSP = \sqrt{(0.15 \times \Delta\delta_N)^2 + \Delta\delta_{HN}^2}$$

*Docking*

Automatic docking of the ligands into MBD2 was performed with Smina (45) and AutoDock Vina (46) on a Conda environment. Ligand structures were extracted from ChemDraw (Perkin Elmer) using

obabel, energy minimized with RDKit (MMFF94S force field) and prepared for flexible docking with the mk_prepare_ligand.py script. For MBD2, we used the PDB structure 6CNQ of human MBD2 in complex with methylated DNA (32). After extracting the coordinates of the protein from the PDB file, the structure of MBD2 was prepared for flexible docking as described in (47), setting residues R166, K167, S168, R188, S189 as flexible. Actual docking was run with Smina with a 50 $Å^3$ box centered on the ligand's 5mC moiety. We retrieved the nine best poses. In house scripts were used to automate ligand preparation, to configure and run the docking procedure, as well as to create pdb structures with the poses and with DNA. Structures were visualized and analyzed with Pymol (Schrödinger LLC); ligand-MBD2 contacts were established with Ligplot+ (48) and PLIP (49).

**Supplementary Information**

Supplementary information is available free of charge at http://pubs.acs.org and contains additional data reported in six figures, NMR spectra and HPLC chromatograms of the compounds.

**Acknowledgements**

MBD2-His plasmid was kindly given by Prof Srinivasan and Prof Nelson from Sidney Kimmel Comprehensive Cancer Center in the University of Medicine Johns Hopkins (Baltimore, MD, USA) that we thank. We thank Ludovic Halby for the design of synthetic pathway, and for the fruitful discussions for the project conception.

**Funding sources**

The 800-MHz NMR spectrometer of the Institut Pasteur was partially funded by the Région Ile de France (SESAME 2014 NMRCHR grant no 4014526). The project was financed by the French PPR Antiobioresistance ANR-20-PAMR-0011 – TheraEPI (to PBA).

## References


1. Nicoglou A, Merlin F. Epigenetics: A way to bridge the gap between biological fields. Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci. 2017 Dec 1;66:73–82.

2. Russo VEA (Vincenzo EA), Martienssen RA, Riggs AD. Epigenetic mechanisms of gene regulation [Internet]. Cold Spring Harbor Laboratory Press; 1996 [cited 2020 Apr 28]. Available from: http://agris.fao.org/agris-search/search.do?recordID=US201300304567

3. Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. Cancer epigenetics: Moving forward. PLoS Genet. 2018;14(6):e1007362.

4. Bertogliat MJ, Morris-Blanco KC, Vemuganti R. Epigenetic mechanisms of neurodegenerative diseases and acute brain injury. Neurochem Int. 2020 Feb 1;133:104642.

5. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Genome Res. 1975;14(1):9–25.

6. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007 Apr;39(4):457–66.

7. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010 Jul;466(7303):253–7.

8. Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol. 1988 Oct 20;203(4):971–83.

9. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science. 2009 May 15;324(5929):930–5.

10. Shimbo T, Wade PA. Proteins That Read DNA Methylation. In: Jeltsch A, Jurkowska RZ, editors. DNA Methyltransferases - Role and Function [Internet]. Cham: Springer International Publishing; 2016 [cited 2020 Apr 22]. p. 303–20. (Advances in Experimental Medicine and Biology; vol. 945). Available from: http://link.springer.com/10.1007/978-3-319-43624-1_13

11. Klimasauskas S, Kumar S, Roberts RJ, Cheng X. Hhal methyltransferase flips its target base out of the DNA helix. Cell. 1994 Jan 28;76(2):357–69.

12. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol. 2010 Sep;11(9):607–20.

13. Du Q, Luu PL, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics. 2015 Sep;7(6):1051–73.

14. Ehrlich M. DNA hypermethylation in disease: mechanisms and clinical relevance. Epigenetics. 2019 Dec 2;14(12):1141–63.

15. Rausch C, Hastert FD, Cardoso MC. DNA Modification Readers and Writers and Their Interplay. J Mol Biol. 2020 Mar 13;432(6):1731–46.

16. Readers of DNA methylation, the MBD family as potential therapeutic targets [Internet]. [cited 2020 Mar 30]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6438182/

17. Buck-Koehntop BA, Defossez PA. On how mammalian transcription factors recognize methylated DNA. Epigenetics. 2013 Feb;8(2):131–7.

18. Defossez PA, Stancheva I. Chapter 12 - Biological Functions of Methyl-CpG-Binding Proteins. In: Cheng X, Blumenthal RM, editors. Progress in Molecular Biology and Translational Science [Internet]. Academic Press; 2011 [cited 2021 Jul 13]. p. 377–98. (Modifications of Nuclear DNA and its Regulatory Proteins; vol. 101). Available from: https://www.sciencedirect.com/science/article/pii/B9780123876850000123

19. Wyhs N, Walker D, Giovinazzo H, Yegnasubramanian S, Nelson WG. Time-Resolved Fluorescence Resonance Energy Transfer Assay for Discovery of Small-Molecule Inhibitors of Methyl-CpG Binding Domain Protein 2. J Biomol Screen. 2014 Aug;19(7):1060–9.

20. El-Ajouz S, Ray D, Allsopp RC, Evans RJ. Molecular basis of selective antagonism of the P2X1 receptor for ATP by NF449 and suramin: contribution of basic amino acids in the cysteine-rich loop. Br J Pharmacol. 2012;165(2):390–400.

21. Roos A, Dhruv HD, Mathews IT, Inge LJ, Tuncali S, Hartman LK, et al. Identification of aurintricarboxylic acid as a selective inhibitor of the TWEAK-Fn14 signaling pathway in glioblastoma cells. Oncotarget. 2017;8(7):12234–46.

22. Reichert Z. Targeting Epigenetic Repression by interfering with methyl-binding domain protein function. PhD dissertation, John Hopkins University, March 2010. UMI number: 3410047. ProQuest, USA.

23. Zhu D, Osuka S, Zhang Z, Reichert ZR, Yang L, Kanemura Y, et al. Suppresses Medulloblastoma Formation by Protecting p53 from Mdm2-Mediated Degradation. Cancer Cell. 2018;33(6):1004-16.e5.

24. Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 2001 Mar 15;15(6):710–23.

25. Campbell PM, Bovenzi V, Szyf M. Methylated DNA-binding protein 2 antisense inhibitors suppress tumourigenesis of human cancer cell lines in vitro and in vivo. Carcinogenesis. 2004 Apr 1;25(4):499–507.

26. Wood KH, Zhou Z. Emerging Molecular and Biological Functions of MBD2, a Reader of DNA Methylation. Front Genet. 2016;7:93.

27. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999 Jan 8;13(15):1924–35.

28. Martin V, Jørgensen HF, Baur Chaubert AS, Berger J, Barr H, Shaw P, et al. MBD2-Mediated Transcriptional Repression of the *p14*$^{ARF}$ Tumor Suppressor Gene in Human Colon Cancer Cells. Pathobiology. 2008;75(5):281–7.

29. Devailly G, Grandin M, Perriaud L, Mathot P, Delcros JG, Bidet Y, et al. Dynamics of MBD2 deposition across methylated DNA regions during malignant transformation of human mammary epithelial cells. Nucleic Acids Res. 2015 Jul 13;43(12):5838–54.

30. Buchini S, Leumann CJ. 2′-O-Aminoethyl Oligoribonucleotides Containing Novel Base Analogues: Synthesis and Triple-Helix Formation At Pyrimidine/Purine Inversion Sites. Eur J Org Chem. 2006 Jul;2006(14):3152–68.

31. Halby L, Marechal N, Pechalrieu D, Cura V, Franchini DM, Faux C, et al. Hijacking DNA methyltransferase transition state analogues to produce chemical scaffolds for PRMT inhibitors. Philos Trans R Soc B Biol Sci. 2018 Jun 5;373(1748):20170072.

32. Liu K, Xu C, Lei M, Yang A, Loppnau P, Hughes TR, et al. Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. J Biol Chem. 2018 May 11;293(19):7344–54.

33. Scarsdale JN, Webb HD, Ginder GD, Jr DCW. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. 2017;12.

34. Vichier-Guerre, S.; Dugue, L.; Bonhomme, F.; Pochet, S.,. Expedient and generic synthesis of imidazole nucleosides by enzymatic transglycosylation. Org Biomol Chem. 2016;14(14):3638–53.

35. Zhang, J-H; Chung,TDY; Oldenburg, KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen. 1999;4(2):67–73.

36. Glickman JF, Wu X, Mercuri R, Illy C, Bowen BR, He Y, Sills M. A comparison of ALPHAScreen, TR-FRET, and TRF as assay methods for FXR nuclear receptors. J Biomol Screen. 2002;7(1):3–10.

37. Kaminski, P. A. Functional cloning, heterologous expression, and purification of two different N-deoxyribosyltransferases from Lactobacillus helveticus. J Biol Chem. 2002;277:14400–7.

38. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins Struct Funct Bioinforma. 2005;59(4):687–96.

39. Stanek J, Schubeis T, Paluch P, Güntert P, Andreas LB, Pintacuda G. Automated Backbone NMR Resonance Assignment of Large Proteins Using Redundant Linking from a Single Simultaneous Acquisition. J Am Chem Soc. 2020 Mar 1;142(12):5793–9.

40. Muhandiram DR, Kay LE. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. J Magn Reson B. 1994 Mar 1;103(3):203–16.

41. Favier A, Brutscher B. NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. J Biomol NMR. 2019 May 1;73(5):199–211.

42. Shen Y, Bax A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR. 2013 Jul 1;56(3):227–41.

43. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577–637.

44. Berjanskii MV, Wishart DS. A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts. J Am Chem Soc. 2005 Nov 1;127(43):14970–1.

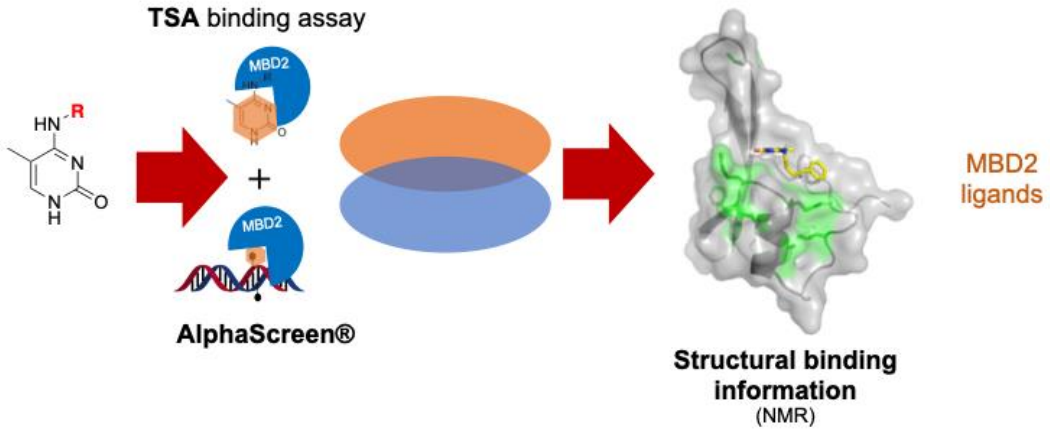
45. Koes DR, Baumgartner MP, Camacho CJ. Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. J Chem Inf Model. 2013 Aug 26;53(8):1893–904.

46. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455–61.

47. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc. 2016 May;11(5):905–19.

48. Laskowski RA, Swindells MB. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. J Chem Inf Model. 2011 Oct 24;51(10):2778–86.

49. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein–ligand interaction profiler. Nucleic Acids Res. 2015 Jul 1;43(W1):W443–7.


TSA binding assay
MBD2
MBD2
AlphaScreen®
Structural binding information
(NMR)
MBD2 ligands