1774–1785 *Nucleic Acids Research, 2019, Vol. 47, No. 4* *Published online 19 December 2018*
doi: 10.1093/nar/gky1264

Structural basis for effects of CpA modifications on C/EBPβ binding of DNA

Jie Yang[1], John R. Horton[1], Dongxue Wang[2], Ren Ren[1], Jia Li[3], Deqiang Sun[3], Yun Huang[3], Xing Zhang[1], Robert M. Blumenthal[4] and Xiaodong Cheng[1,*]

[1]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA, [2]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA, [3]Center for Epigenetics & Disease Prevention, Institute of Biosciences and Technology, College of Medicine, Texas A&M University, Houston, TX 77030, USA and [4]Department of Medical Microbiology and Immunology, and Program in Bioinformatics, The University of Toledo College of Medicine and Life Sciences, Toledo, OH 43614, USA

Received September 12, 2018; Revised November 28, 2018; Editorial Decision December 04, 2018; Accepted December 08, 2018

ABSTRACT

CCAAT/enhancer binding proteins (C/EBPs) regulate gene expression in a variety of cells/tissues/organs, during a range of developmental stages, under both physiological and pathological conditions. C/EBP-related transcription factors have a consensus binding specificity of 5′-TTG-CG-CAA-3′, with a central CpG/CpG and two outer CpA/TpG dinucleotides. Methylation of the CpG and CpA sites generates a DNA element with every pyrimidine having a methyl group in the 5-carbon position (thymine or 5-methylcytosine (5mC)). To understand the effects of both CpG and CpA modification on a centrally-important transcription factor, we show that C/EBPβ binds the methylated 8-bp element with modestly-increased (2.4-fold) binding affinity relative to the unmodified cognate sequence, while cytosine hydroxymethylation (particularly at the CpA sites) substantially decreased binding affinity (36-fold). The structure of C/EBPβ DNA binding domain in complex with methylated DNA revealed that the methyl groups of the 5mCpA/TpG make van der Waals contacts with Val285 in C/EBPβ. Arg289 recognizes the central 5mCpG by forming a methyl-Arg-G triad, and its conformation is constrained by Val285 and the 5mCpG methyl group. We substituted Val285 with Ala (V285A) in an Ala-Val dipeptide, to mimic the conserved Ala-Ala in many members of the basic leucine-zipper family of transcription factors, important in gene regulation, cell proliferation and oncogenesis. The V285A variant demonstrated a 90-fold binding preference for methylated DNA (particularly 5mCpA methylation) over the unmodified sequence. The smaller side chain of Ala285 permits Arg289 to adopt two alternative conformations, to interact in a similar fashion with either the central 5mCpG or the TpG of the opposite strand. Significantly, the best-studied *cis*-regulatory elements in RNA polymerase II promoters and enhancers have variable sequences corresponding to the central CpG or reduced to a single G:C base pair, but retain a conserved outer CpA sequence. Our analyses suggest an important modification-dependent CpA recognition by basic leucine-zipper transcription factors.

INTRODUCTION

The main players for epigenetic modifications are increasingly well known, including the writers, readers and erasers of DNA and histone modification marks. In mammals, DNA methyltransferases (Dnmt3a and Dnmt3b) (1,2) methylate the 5-carbon position of cytosine, generating *de novo* 5-methylcytosine (5mC or M) that is maintained by Dnmt1 following DNA replication (3). Ten-eleven-translocation (TET) dioxygenases oxidize 5mC successively to become 5-hydroxymethylated (5hmC or H), 5-formylated (5fC), and 5-carboxylated cytosine (5caC) (4,5). The five chemical forms of cytosine (unmodified C, 5mC, 5hmC, 5fC and 5caC) are components of the 'epigenome' and exhibit the same base-pairing and protein-coding properties, but differ in their interactions with transcription factors (TFs) (6). Recent work assessing the contribution of TET enzymes have helped characterize how TETs and their enzymatic products have distinct effects on DNA modification patterns, regulatory element activity, DNA binding protein specificity and gene expression (7,8). In addition to the classic methyl-binding domain (MBD) proteins (9), there is

*To whom correspondence should be addressed. Tel: +1 713 834 6274; Email: xcheng5@mdanderson.org

a growing number of transcriptional regulators that have adapted to respond to different cytosine modification states, potentially acting as direct epigenetic sensors to instruct downstream events. Currently known examples include members of C2H2 zinc finger proteins, TFs of three families (basic-helix–loop–helix (bHLH), basic leucine-zipper (bZIP), and homeodomain), and tumor suppressor protein p53 (10–20). Among them, three different classes of DNA binding domains (MBD, C2H2 ZF and p53) use the same methyl–Arg–Gua triad to interact with 5mCpG and TpG (17,21). The bZIP proteins AP-1/Zta and homeodomain proteins HOBX13/CDX1/CDX2/LHX4 use van der Waals contacts with methyl groups of 5mC and thymine (19,20). Juxtaposition of an acidic residue against cytosine can specifically repel the negatively-charged carboxylate of 5caC as in some C2H2 ZFs (12,22) and positively-charged residues can form a basic environment electrostatically compatible with 5caC as in bHLH protein MAX (18).

Enhancer-box (E-box) sequences are DNA response elements that play a major role in regulating transcriptional activity (23). The consensus sequence of the E-box is usually CANNTG (where N can be any nucleotide), with a subset containing the palindromic sequence CACGTG (where the central CpG is the canonical DNA methylation site). The oncogenic MYC and its binding partner MAX are bHLH TFs that preferentially recognize CACGTG (24–26). MAX binds an unmodified E-box, and methylation of the central CpG greatly inhibits its binding (27), whereas 5caC restores binding to the level of unmodified C (18). The affinity of MAX for the central CpG modification of the E-box is in the order (5caC ∼ C) > 5fC > (5mC ∼ 5hmC) (18).

Human activator protein 1 (AP-1, which is homo- or hetero-dimers of Fos and Jun), and its ortholog Epstein-Barr virus Zta, are bZIP TFs that recognize three forms of 7-base pair response elements: 5′-TGAGTCA-3′, 5′-MGAGTCA-3′, and 5′-TGAGMCA-3′ (where M = 5mC) (28–32). These elements share the feature that each sequence has two methyl groups at nucleotide positions 1 and 5 from the 5′ end, resulting in four methyl groups symmetrically positioned in duplex DNA, as 5mC and thymine both have a methyl group at the 5-carbon position. These methyl groups are in van der Waals contact with a conserved Ala-Ala in AP-1, or with the corresponding Zta residues Ala-Ser (Figure 1A) in a space- and sequence-specific manner (19). AP-1, like many dimeric transcription factors (*e.g.*, Myc/Max) is a homo- or hetero-dimeric complex that comprises members of the Jun, Fos, ATF (activating transcription factor) and other bZIP protein families (33). The structure and sequence conservation among the extended family of bZIP TFs (Fos/Jun/ATF4/CREB) implies a similar pattern of DNA recognition, such as the conserved Ala-Ala dipeptide-methyl interaction (19). Interestingly, C/EBP family has a unique Ala-Val pair in the corresponding position (Figure 1A).

CCAAT/enhancer binding proteins (C/EBP) regulate gene expression in a variety of cells/tissues/organs at different developmental stages under both physiological and pathological conditions (34–36). C/EBPβ plays a role in macrophage activation and rapid granulopoiesis following the stimulation by cytokines (37,38). Here, in order to understand the effects of DNA modification on the binding of an important transcription factor, using the isolated C/EBPβ bZIP DNA binding domain, we ask how C/EBP proteins recognize 5mCpG and 5mCpA in the context of TTG-CG-CAA, and how they distinguish between 5mC and 5hmC. We elected to study human C/EBPβ because the DNA binding domains of its wild type and V285A versions have already been characterized structurally, though only in complex with unmodified DNA (PDB ID 1GU4 and 2E42).

MATERIALS AND METHODS

Protein expression and purification

We prepared C-terminal bZIP domain of human C/EBPβ (residues 269–344; UniProtKB/ Swiss-Prot: P17676) (Supplementary Figure S1). The DNA binding domain (pXC1599) and its variant V285A (pXC2027) were expressed as N-terminal 6xHis-SUMO fusion proteins (39), via a modified pET28b vector (Novagen) in *Escherichia coli* BL21-Codon Plus(DE3)-RIL (Stratagene). Bacterial cells were cultured in LB medium at 37°C up to a culture density of $A_{600\,nm} = 0.4$. Then the temperature was changed to 16°C, and about 1 h later, at $A_{600\,nm}$ of ∼0.6, 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added. After inducing overnight, cells were harvested and lysed by sonication in 20 mM sodium phosphate (pH 7.4), 500 mM NaCl, 25 mM imidazole, 5% (v/v) glycerol, 0.5 mM tris (2-carboxyethyl) phosphine (TCEP) and 0.1 mM phenylmethylsulfonyl fluoride. The lysate was treated with 0.3% (w/v) polyethylenimine (40) and centrifuged at 47 850 g for 30 min at 4°C. The supernatant was collected and subjected to a four-column chromatography protocol, conducted in a BIO-RAD NGC™ system. The sample was loaded onto a 5-ml HisTrap column (GE Healthcare), and the 6xHis-SUMO tagged protein was eluted with imidazole gradient from 25 mM to 500 mM. Eluted protein was incubated with Ulp1 protease (purified in-house) at 4°C overnight to remove the 6xHis-SUMO tag, leaving two additional N-terminal residues (His-Met) fused to the recombinant protein. The protein was further purified by 5 ml HiTrap Q-SP columns (GE Healthcare) connected in tandem (40). The protein was eluted from SP column with 100 ml linear NaCl gradient from 150 mM to 1 M. The peak fractions were confirmed by 18% SDS PAGE (Supplementary Figure S1C) and concentrated to 5 ml. The protein concentration was estimated based on Pierce™ BCA protein assay kit (ThermoFisher) (Supplementary Figure S1D). The concentrated protein was loaded onto a HiLoad Surperdex 75 (16/60) column (GE Healthcare) (Supplementary Figure S1E). The protein was collected as a single peak and concentrated to 22 mg/ml (2.44 mM) in 20 mM Tris (pH 7.5), 150 mM NaCl, 5% (v/v) glycerol and 0.5 mM TCEP. The mutant was generated by PCR and confirmed by sequencing.

Isothermal titration calorimetry (ITC)

ITC experiments were performed at 25°C using a MicroCal PEAQ-ITC automated system (Malvern instrument Ltd). Double stranded oligonucleotides (2 μM) were maintained

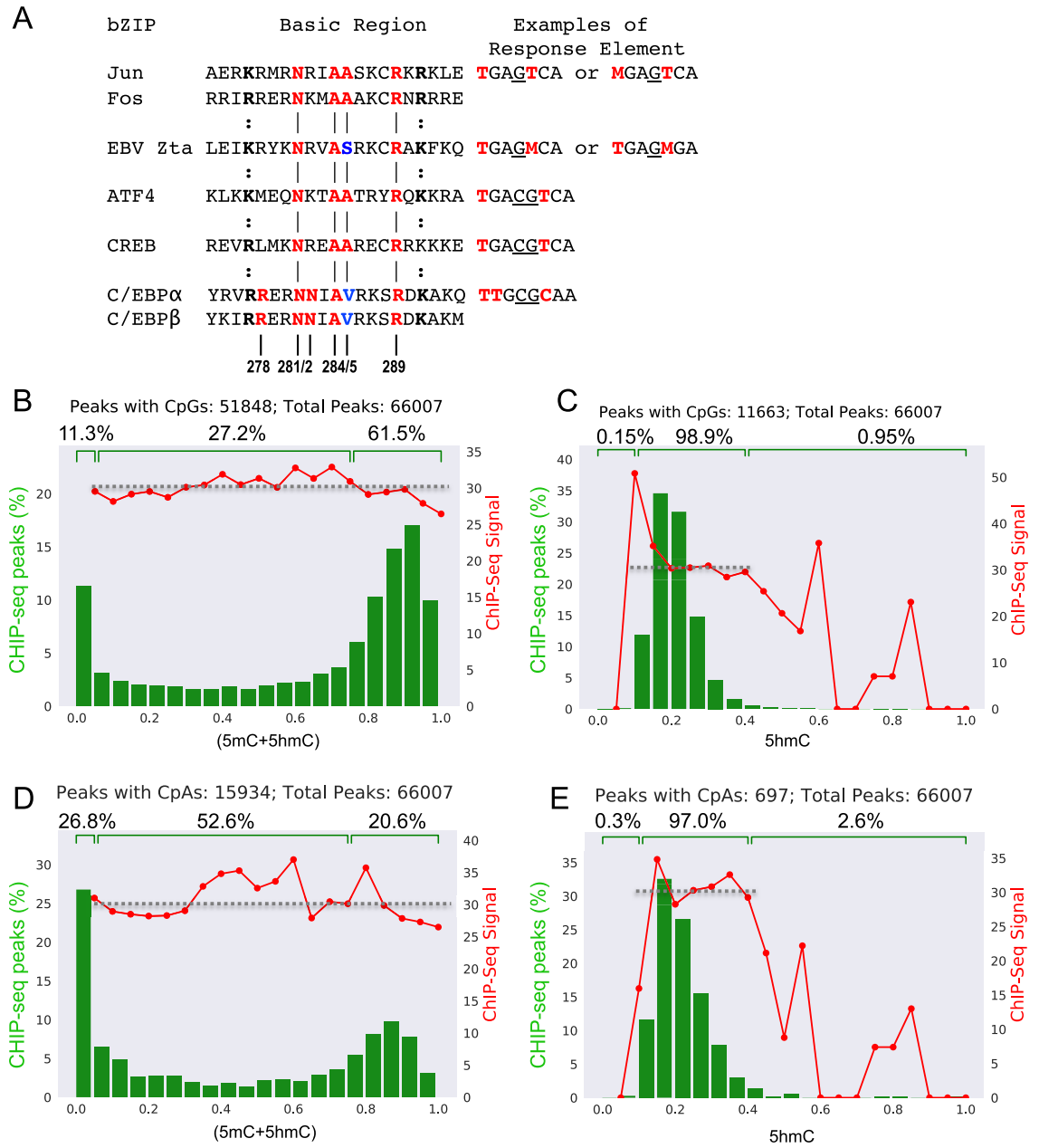

Figure 1. Conservation of C/EBP proteins and its occupancy related to 5mC. (**A**) Variations of the di-alanine motif among bZIP transcription factors. The corresponding motif is Ala-Ser in Zta and Ala-Val in C/EBP. The top two lines show Jun and Fos, homo- or hetero-dimers of which form the transcription factor AP-1. The protein residues in colors are involved in DNA base interactions (see text). The DNA nucleotides in red are the pyrimidines (T and M = 5mC), which both have a methyl group in the 5-carbon position. (B–E) Quantification of CpG (**B, C**) or CpA (**D, E**) DNA methylation levels, quantified based on WGBS data (GSM432686) for (5mC+5hmC) (green bars in B and D) and TAB-seq data (GSM882245) for 5hmC (green bars in C and E), within C/EBPβ binding sites (GSM935295), illustrated by ChIP-seq signals (FDR cutoff ≤ 0.05) in human H1 ES cells (red dots). The x-axis indicates DNA methylation or hydroxymethylation levels ranging from 0 (no modification) to 1 (complete modification). The y-axis is the percentage of C/EBPβ ChIP-seq peaks (green histogram bars) and the corresponding ChIP-seq enrichment (red dotted lines). A dashed horizontal line indicates the mean ChIP-seq signal intensity.

in the sample cell and the proteins (70–80 μM) were injected into the cell by a syringe. The amount of each injection was 2 μl under continuous stirring (750 rpm) and the reference power was set as 8 μcal/s. The duration of each injection was fixed at 4 s and the spacing time between the injections was 200 s in order to achieve equilibrium. For each oligo, a reference titration of buffer (20 mM Tris, pH 7.5 and 350 mM NaCl) without protein was subtracted from experimental data to control for heat of dilution and non-specific binding. Binding constants were calculated by fitting the data using the ITC data analysis module supplied by the manufacturer. For those binding curves that did not reach saturation (Supplementary Figure S2), the lower limit of the binding affinity was estimated.

Fluorescence-based DNA binding assay

Fluorescence polarization assays were performed using a Synergy 4 microplate reader (BioTek) to measure DNA binding affinity. The 6-carboxy-fluorescein (FAM)-labeled double strand DNA probes (5 nM) were incubated with increasing amount of proteins (monomer concentration 0.8 nM to 13.25 μM) for 15 min in 20 mM Tris (pH 7.5), 5% (w/v) glycerol, 450 mM NaCl, 0.5 mM TCEP and 0.1 mg/ml BSA. GraphPad Prism software (version 7.0) was used to do curve fitting. K_D values were calculated as [mP] = [maximum mP] × [C]/ (K_D + [C]) + [baseline mP], where mP is millipolarization and [C] is protein concentration, and ΔmP = ([mP]—[baseline mP]). The reported mean ± SEM of the interpolated K_D values were calculated from two independent experiments performed in duplicate. We note that the mP amplitudes for the oligos containing 5hmC did not reach the same level as those of C- and 5mC-containing oligos (Supplementary Figure S3), and the lower limit of the binding affinity was estimated.

Crystallography

We crystallized C/EBPβ wild type and mutant V285A in the presence of 16-bp methylated oligonucleotide (5′-TATATTGMGMAATATA-3′, where M = 5mC) by the sitting drop vapor diffusion method at 19°C. Equal amounts of purified protein (2 mM) and double-stranded oligonucleotide (2 mM) were incubated at 4°C for 30 min in 20 mM Tris (pH 7.5), 150 mM NaCl, 5% (v/v) glycerol, and 0.5 mM TCEP before crystallization. An Art Robbins Phoenix Crystallization Robot was used to set up screens with solutions from Hampton Research (PEG/Ion HT and Index HT). Crystals were observed under many conditions and X-ray diffraction data were collected from the crystals that formed in solutions of 0.1 M sodium acetate trihydrate pH 7.0, 12% (w/v) polyethylene glycol 3350 or 0.2 M sodium chloride, 0.1 M Bis–Tris pH 6.5, 25% (w/v) polyethylene glycol 3350. Crystals were cryo-protected by soaking in mother liquor supplemented with 20% (w/v) ethylene glycol before plunging into liquid nitrogen.

Two datasets of wild type in complex with methylated DNA were collected at the SERCAT beamline 22ID of Advanced Photon Source at Argonne National Laboratory. One dataset of the mutant V285A-DNA complex was collected on our local Rigaku X-ray facility equipped with a MicroMax-003 Microfocus sealed tube X-ray generator, an AFC11 partial-χ, 4-axis goniometer and an HyPix-6000HE hybrid photon counting detector. Crystallographic datasets were processed with HKL2000 (41) for SERCAT data or CrysAlis^{Pro} for Rigaku data. Molecular replacement was performed with the PHENIX PHASER module (42) by using a structure of C/EBPβ NDA binding domain (PDB ID 2E42) as a search model. PHENIX REFINE (43) was used for refinement with 5% randomly chosen reflections for validation by the R_{free} value. COOT (44) was used for model building and corrections between refinement rounds. Structure quality was analyzed during PHENIX refinements and later validated by the PDB validation server. Molecular graphics were generated using PyMol (Schrödinger, LLC).

Correlation analysis among 5mC (M), 5hmC (H) and C/EBPβ ChIP-seq peaks

Tet-assisted bisulfite sequencing (TAB-seq) data (GSM882245) (45), whole genome bisulfite sequencing (WGBS) data (GSM432686) and C/EBPβ ChIP-seq data (GSM935295) from human H1 ES cells were downloaded (46). An in-house Mmint software package was used to perform the analysis. In brief, we calculated the DNA methylation levels from WGBS data (5mC+5hmC) [Note that WBGS does not discriminate between 5mC and 5hmC (47)] or hydroxymethylation levels from TAB-seq for CpG or CpA sites within identified C/EBPβ ChIP-seq peaks. The peaks were called using MACS2 algorism (48) with false discovery rate (FDR) ≤0.05. Then we binned C/EPBβ peaks based on their average CpG or CpA DNA methylation or hydroxymethylation levels into 20 categories with 5% intervals of corresponding DNA methylation or hydroxymethylation levels ranging from 0 (no modification) to 1 (complete modification). The percentage of C/EBPβ peaks within each category is shown using green histogram bars, and the corresponding ChIP-seq enrichment using red dotted lines (Figure 1B–E).

RESULTS

C/EBPβ-bound sites had both low and high 5mC levels, but were devoid of 5hmC

Besides binding 7-bp elements (TGA-G-TCA), AP-1 Jun or Fos proteins can form heterodimers with ATF to recognize 8-bp cAMP-response elements (CRE, 5′-TGA-CG-TCA-3′), which are also recognized by CRE-binding (CREB) proteins (49,50). The difference between the two elements is the one base-pair (bp) expansion of the central C:G bp to a CpG dinucleotide (Figure 1A). This expanded central CpG dinucleotide potentially allows the CpG methylation/oxidation status to play a regulatory role. Like MAX, central CpG methylation decreases the binding affinities of CREB and ATF4 (17,51) whereas, like AP-1, replacing the outer TpG with 5mCpG leads to increased affinity of ATF4 compared to unmodified cytosine at the outmost position (17).

The C/EBP proteins, in contrast to AP-1-related proteins, accept methylation at the central CpG (52) in the sequence context of either CRE elements (TGA-CG-TCA) or C/EBP consensus sequences (TTG-CG-CAA) (15,17,53), while their binding is inhibited by 5hmC in the central CpG dinucleotide (54). This difference in binding of modified DNA corresponds well with the genome-wide occupancy of C/EBPβ in H1 human embryonic stem cells—C/EBPβ-bound sites had both low and high 5mC levels, but were devoid of 5hmC (Figure 1B–E). The two sequence elements, **TGA**-CG-**TCA** and T**TG**-CG-**CA**A, differ at the second and the third base pairs (underlined); each sequence has a CpA/TpG dinucleotide in both half sites (bold), though their spacing is different. Methylation of the CpA site, by DNMT3 (55–57), would generate a mostly-shared methylation pattern. Specifically, with 5-methylated pyrimidines capitalized, the two elements are **T**ga**C** | g**T**Ca (CRE), and **TT**g**C** | g**C**aa (C/EBP consensus), and share 5-methyl groups at positions 1, 4 and 6 (bolded).

DNA binding of C/EBPβ is enhanced by methylation

We next measured the binding of human C/EBPβ bZIP domain to double-stranded oligonucleotides (oligos) containing the C/EBP consensus recognition sequence TTGCG-CAA, in which the status of cytosine residues on both strands was either unmodified C, 5mC (M)-modified, or oxidized to 5hmC (H). We used isothermal titration calorimetry to quantitatively measure the dissociation constants (K_D). C/EBPβ DNA-binding domain bound the unmodified oligo with a K_D of 70 nM (Table 1A and Supplementary Figure S2). Under the same conditions, the 5mC-containing oligos increased binding affinity by a factor of ∼2.4, to 29 nM. In contrast, 5-hydroxymethylation reduced binding affinity by a factor of ∼36, to be more than 2.5 μM (Table 1A). Further, the modifications at CpG and CpA do not contribute equally to binding affinity. Full hydroxymethylation of the central CpG site (both strands) reduced binding affinity by a factor of ∼3.4, to 0.24 μM, while full hydroxymethylation of the two CpA sites (on opposite strands) resulted in much weaker binding (>2.5 μM), similar to the cumulative effect of 5hmC at all cytosines. This result suggests that the CpA modification has a dominant negative effect.

A different binding assay, based on fluorescence polarization, confirmed that the binding affinity of C/EBPβ with the unmodified oligo is 60–70 nM (Table 1B and C and Supplementary Figure S3), the same as the ITC result. The central CpG methylation increased binding by 3.5× (to 20 nM), while binding was decreased 5.7× (to 0.4 μM) by hydroxymethylation (Table 1B). Thus, the binding affinity is in the order 5mC (M) > C > 5hmC (H) for the CpG site. The CpA site differed in that 5mC (M) was about equal to C, while 5hmC (H) was 10-fold weaker (Table 1C).

Structure of C/EBPβ bound with methylated DNA in both CpG and CpA sites

To understand how C/EBPβ binds 5mC oligos preferentially, we next co-crystallized C/EBPβ (residues 269–344) with a 16-bp duplex oligo containing the fully-methylated central CpG dinucleotide and two methylated CpA sites (Figure 2A). This C/EBPβ-DNA complex crystallized in two space groups (C2 and $C222_1$), resulting in structures determined to resolutions of 1.75 and 1.93 Å, respectively (Supplementary Table S1). The protein components of the two structures are highly similar between the two space groups, with root-mean-square deviation of 1.4 Å across 103 pairs of Cα atoms. The differences between the two structures, from the DNA component, lie in the ends of two DNA molecules, where DNA stacking interactions occur (Supplementary Figure S4A and B), and from the protein component, lie in the helical region away from the DNA binding (Supplementary Figure S4C). We will describe the detailed methyl specific recognition in the higher resolution structure of C2 space group. The DNA molecules are largely B-form, and detailed local DNA parameters are provided in Supplementary Table S2.

As expected, the structure encompasses two long helices, creating a clamp-like basic environment ideal for electrostatic attraction to the negatively-charged DNA phosphate backbone (Figure 2A). The basic region in the N-terminal portion of the helix binds in the major groove of the DNA and recognizes half of the C/EBP element (Figure 2B), while the C-terminal portion of the leucine-zipper forms a dimer interface. The half site (TTGC) includes two T:A outer base pairs and two 5mC:G inner base pairs, with the methyl groups of 5mC bases clearly located in the omit electron density maps (Figure 2C).

From each monomer (colored blue and green in Figure 2A), six residues (Arg278, Asn281, Asn282, Ala284, Val285 and Arg289) are involved in direct base contacts with four base pairs. At the first base pair, T1:A1, Ala284 is in van der Waals contact with the 5-position methyl group of T1, while Asn281 forms a hydrogen bond (H-bond) with the *O*4 atom of the thymine (Figure 2D); there is no direct interaction with the paired adenine at A1. The second base pair, T2:A2, is engaged in the most direct contacts (Figure 2E). First, the adenine forms two H-bonds with Arg278 (via the ring *N*7 atom) and Asn281 (via the exocyclic *N*6-amino group). Second, the paired thymine at T2 forms a weak O●●●H-C type H-bond (58), and its 5-position methyl group is within van der Waals contact distance of Val285. The next base pair, G3:M3 at position 3, has many fewer direct contacts, but its interaction is focused on the 5-methyl group of 5mC, via a van der Waals contact with Val285 and a C-H●●●O type H-bond with Asn282 which, in turn, interacts with the phosphate group of the nucleotide (Figure 2F).

Among the interactions described above, two residues bridge between two neighboring base pairs. Asn281 bridges between the *O*4 atom of T1 in the top strand, and the *N*6 atom of A2 in the bottom strand, via H-bonds (Figure 2G). This is possible because asparagine can act simultaneously as an H-bond donor (to *O*4 of T1) and acceptor (from *N*6 of A2). Val285, through its two terminal methyl groups, interacts with T2 of the top strand and M3 of the bottom strand (Figure 2H). In addition, the guanidino group of Arg278 is in five-way interactions with Asn281, Asn282, Adenine A2, the phosphate group of M3 and a water molecule (Figure 2I), saturating Arg278's H-bonding and electrostatic potential. Thus in the protein-DNA interface lies a network of polar and hydrophobic interactions involving inter- and intra-molecular contacts.

The central methylated CpG dinucleotide is recognized by Arg289 from each monomer. Each arginine residue provides bidentate H-bonds to the *N*7 and *O*6 atoms of the Gua of the same half site (Figure 2J), and a van der Waals contact with the 5mC methyl group in the second half site on the same strand (cyan in Figure 2K), forming a classic 5mC-Arg-Guanine triad (21). Interestingly, the guanidino group of Arg289 is sandwiched between the hydrophobic side chain of Val285 and the methyl group of M5 (Figure 2K). The interactions with the methyl groups of M3 (via Val285 and Asn282) and M5 (via Arg289) (Figure 2F and K) would enhance the methylated DNA binding relative to unmodified cytosine (Table 1A). Superimposition of our complex structure with that of C/EBPβ bound with unmodified DNA (PDB 1GU4) showed the loss of methyl-mediated contacts in the unmodified cytosines, though the overall structures are essentially unchanged (Figure 2L).

Oxidation of 5mC to 5hmC, on either M3 or M5, would introduce an additional hydroxyl oxygen atom that would clash with the interacting amino acids, resulting in re-

Table 1. Summary of DNA binding affinities of C/EBPβ bZIP domain bound with DNA

A	5′-TATATTG**X**G**Y**AATATA-3′				
	3′-ATATAA**Y**G**X**GTTATAT-5′				
C/EBPβ	K_D (μM) (measured by ITC)[a]				
	X = Y = M	X = Y = C	X = Y = H	X = H, Y = C	X = C, Y = H
WT (V285)	0.029 ± 0.008	0.07 ± 0.02	>2.5	0.24 ± 0.05	>2.5
V285A (A285)	0.004 ± 0.001	0.36 ± 0.06	0.8 ± 0.1	>2.5	0.4 ± 0.1
B	5′-AGGATTG**X**GCAATAT-3′				
	3′-TCCTAAC**X**GTTATA-5′-FAM				
C/EBPβ	K_D (μM) (measured by FP)[b]				
	X = M	X = C	X = H		
WT (V285)	0.02 ± 0.01	0.07 ± 0.02	0.4 ± 0.1		
V285A (A285)	1.1 ± 0.2	0.5 ± 0.1	3 ± 1 (*)		
C	5′-AGGATTGCG**Y**AATAT-3′				
	3′-TCCTAA**Y**GCGTTATA-5′-FAM				
C/EBPβ	K_D (μM) (measured by FP)[b]				
	Y = M	Y = C	Y = H		
WT (V285)	0.05 ± 0.01	0.06 ± 0.01	0.6 ± 0.3 (*)		
V285A (A285)	0.007 ± 0.002	0.6 ± 0.2	0.8 ± 0.4 (*)		

[a]See Supplementary Figure S2 for original ITC data.
[b]See Supplementary Figure S3 for original FP data.
*Indication of the mP amplitudes for the H-oligos did not reach the same level as those of C- and M-containing oligos; see Supplementary Figure S3.
M = 5mC, H = 5hmC.

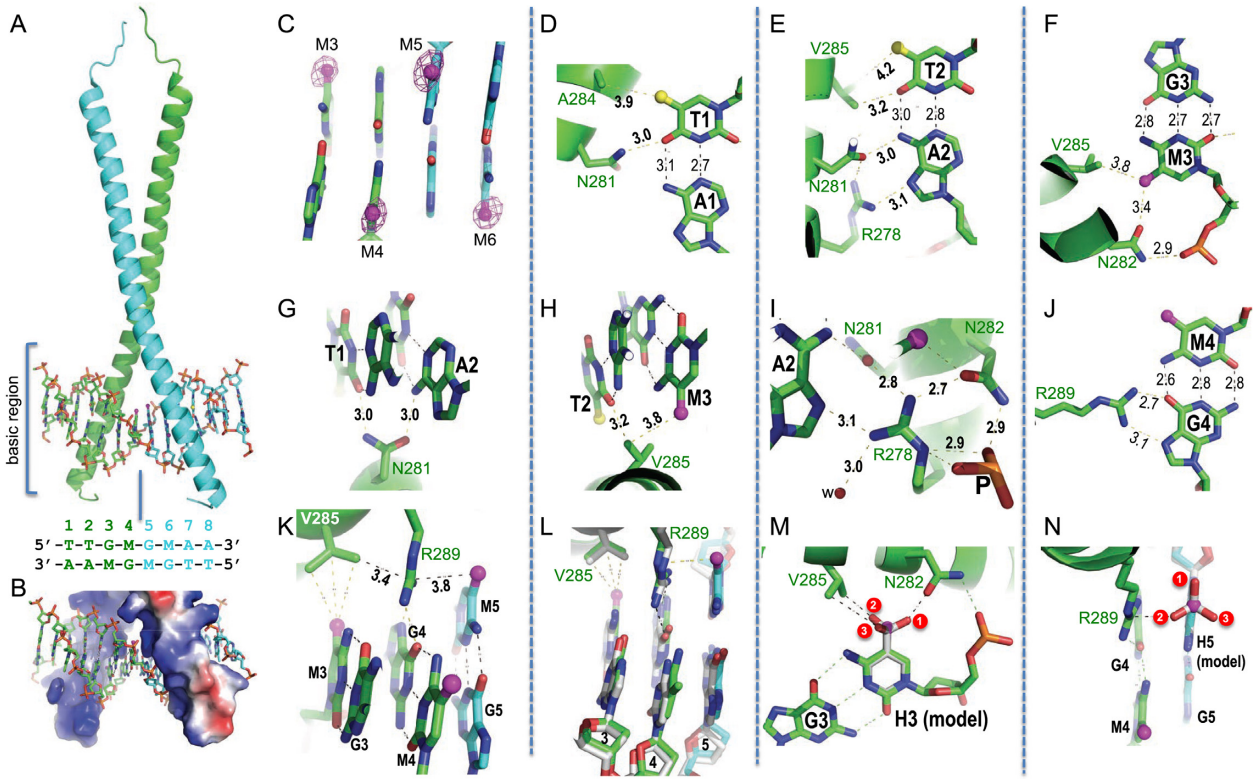

Figure 2. Structure of C/EBPβ in complex with fully methylated DNA. (**A**) Each C/EBPβ monomer (green and cyan) recognizes one half site (bp 1–4 or 5–8). (**B**) The surface charge of the C/EBPβ homodimer at neutral pH is displayed as blue for positive, red for negative, and white for neutral. (**C**) Omit electron density map (F_o—F_c) contoured at 5σ above the mean is shown for omitting the methyl groups of four 5mC bases (in magenta). The view is looking down into the DNA major groove. (**D**) Interaction with outer T1:A1 base pair. Numbers indicate the inter-atom distance in angstroms. (**E**) Interactions with T2:A2 base pair. (**F**) Interactions with G3:M3 base pair (M = 5mC). (**G**) Asn281 bridges between T1 and A2. (**H**) Val285 bridges between T2 and M3. (**I**) Arg278 is involved in 5-way interactions. (**J**) Arg289 interacts with G4. (**K**) Arg289 is sandwiched between Val285 and the methyl group of M5. (**L**) Comparison of C/EBPβ in complex with methylated DNA (colored) and unmodified DNA (in grey; PDB 1GU4). (M-N) 5hmC was modeled onto M3 and M5 positions. The hydroxyl oxygen atom of 5hmC could adopt three alternative conformations (labeled as 1, 2 and 3). All three conformations at M3 would potentially result in repulsion with Asn282 and Val285 (panel M), whereas conformation 2 at M5 would result in repulsion with Arg289 (panel N).

pulsion between protein and DNA (Figure 2M and N). We modeled the hydroxyl oxygen atom of 5hmC onto the M3 and M5 positions. The hydroxyl oxygen atom of 5hmC can rotate freely along the C5-CH_2 bond in the absence of spatial constraint, and rotating the C5-CH_2 bond every 120° generated three possible biologically-relevant conformations—all three were observed in our study of 5hmC-containing DNA bound by transcription factors WT1 and Egr1 (14). At the M3 position (the methylated CpA site), all three conformations of the modeled hydroxyl oxygen atom would interfere with the binding positions of Asn282 and Val285 (Figure 2M). At the M5 position (the methylated central CpG site), one of the three conformations would clash with Arg289, whereas the two other conformations could be accommodated towards the solvent in the major groove (Figure 2N). This modeling may explain the observation that hydroxymethylation at the CpA sites has a dominant negative effect on DNA binding strength.

The V285A variant is highly selective for methylated DNA

As noted earlier in the sequence alignment of basic regions of AP-1 related transcription factors (Figure 1A), the DNA base-contacting amino acids are highly conserved, including the asparagine (Asn281 of C/EBPβ) for bridging the first two base pairs, the di-alanine for binding methyl groups at positions 1 and 3 of each half site, and arginine (Arg289 of C/EBPβ) for recognizing central C:G base pair or central CpG dinucleotide. Unique to C/EBP family is the Ala-Val dipeptide in the place of the conserved di-alanine. We thus generated CEBPβ Val285-to-Ala (V285A) mutant and repeated the DNA binding assays using the same set of oligonucleotides (Table 1). Somewhat expectedly, the V285A protein bound unmodified oligo with affinity lowered >5x from that of wild type (WT) protein (K_D value increased from 70 to 360 nM) (Table 1A). The reduced affinity could be explained by the loss of Val285 mediated interactions with Thymine T2 and Arg289 (Figure 3A; also see Figure 2E and K).

Unexpectedly, compared to the WT, V285A affinity for 5mC increased by ∼7-fold (K_D value decreased from 29 to 4 nM) (Table 1A). The opposite effects on 5mC and C resulted in the V285A variant strongly distinguishing sequences containing 5mC from the C-containing oligonucleotide by the relatively large factor of ∼90 (comparing 4–360 nM; Table 1A). Like WT, the CpA modification has an overriding effect on the binding of V285A relative to that of the central CpG. CpA methylation alone increased the binding affinity of V285A dramatically by a factor of ∼86 (comparing 7–600 nM; Table 1C), while the change for CpG methylation was modest (∼2-fold) (Table 1B; see Discussion).

To understand the structural basis for such a huge preference for 5mC over unmodified C, we determined the co-crystal structure of V285A variant in complex with methylated DNA, to a resolution of 2.05 Å (Supplementary Table S1). We then compared this structure of V285A-5mC to that of WT-5mC and to a previously-solved structure of V285A with unmethylated DNA (PDB 1GU4). Except for the side chain of residue Arg289 (see below), the overall structure of the CEBPβ bZIP domain is essentially unchanged among these complexes, with root-mean-square deviations of just 0.2 Å across 113 pairs of Cα atoms between the WT (PDB 6MG2) and V285A (PDB 6MG3) bound with methylated DNA in the same space group of $C222_1$ (Supplementary Figure S4D), or of 0.4 Å between V285A bound to methylated (PDB 6MG3) versus unmethylated DNA (PDB 2E42).

We first compared the structures of WT (V285) and A285 variant in complex with methylated DNA, which had shown 7-fold increased binding affinity by the A285 variant. As demonstrated earlier, Val285 is engaged in a van der Waals contact with the guanidino group of Arg289, and with the methyl group of M5, which together effectively fix the conformation of Arg289 (Figure 2K). In the V285A variant, the reduced size of the Ala285 side chain allowed Arg289 to take two alternative conformations (Figure 3B). The first conformation retained the bidentate H-bonds with Guanine G4 while, in the second conformation, Arg289 moved over to the opposite strand and formed H-bonds with the Guanine G3 of the neighboring base pair (Figure 3B). In addition, each conformation of Arg289 engaged in a van der Waals contact with the methyl group of the neighboring 5′ pyrimidine (T2 or M5) of the same DNA strand (Figure 3C) and generated two methyl-Arg-Gua triads: T2-Arg-G3 of the top strand and M5-Arg-G4 of the bottom strand (Figure 3D). We termed both Arg289 conformations as 'methyl specific'. The reduced size of side chain of Ala285, relative to that of Val285, also allowed protein and DNA components to move towards one another. Of the base pairs involved in direct base–protein contacts, the first base pair (T1:A1) has the least conformational difference, while the remaining three base pairs (2–4) had an increased shift of ∼1–2 Å from outer base pair to the central base pair, due to both the reduced side-chain size of residue 285 as well as to pulling of G3 by Arg289 (Figure 3E–H). Together, these observations can explain why V285A formed a tighter complex, having increased binding affinity, with the methylated DNA.

We next compared structures of the A285 variant in complex with methylated (5mC) and unmodified C-containing oligos (Figure 3I–L). The aforementioned interactions mediated by the methyl groups of M3 and M5 vanished in the C-oligo (Figure 3K and C). Furthermore, the added T2-Arg289-G3 interaction in the A285–5mC complex did not occur in the A285 variant bound to a C-oligo. Thus, the cumulative effect of gained interactions in the A285–5mC complex (with increased affinity) and lost interactions in the A285-C complex (with decreased affinity) resulted in a variant that strongly prefers 5mC over C (by a factor of ∼90).

DISCUSSION

Conservation of C/EBP DNA-interaction region among vertebrates

The residues identified as playing key roles (in colored font in Figure 1A) are fully conserved, in all vertebrate classes save the one most distant from Mammalia—the Agnatha (jawless vertebrates such as lampreys), where no orthologs could be identified (Supplementary Figure S5). Among the other vertebrate classes, ranging from Mammalia to Chondrichthyes (Rhincodon, whale shark), the C/EBP elements responsible for recognizing DNA bases and methylation are

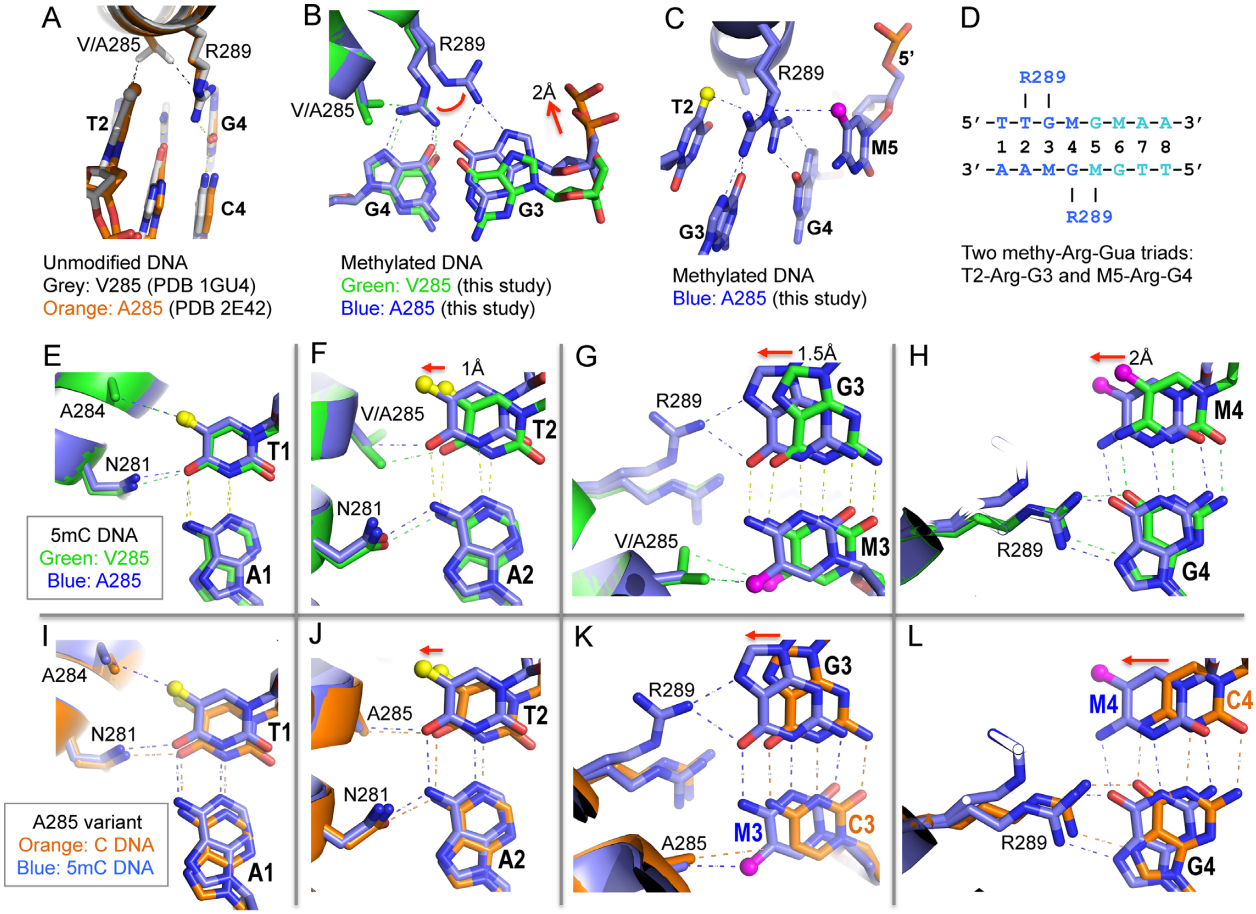

Figure 3. Structural comparisons between WT (V285) and A285 variant in binding methylated and unmodified DNA (M = 5mC). (**A**) Superimposition of WT (PDB 1GU4) and A285 (2E42) C/EBPβ in complex with unmodified DNA. (**B**) Superimposition of WT and A285 in complex with methylated DNA. Note the alternative conformations of Arg289. (**C**) Arg289 participated in two near identical interactions: M5-Arg-G4 and T2-Arg-G3. (**D**) The methyl-Arg-G interactions occur in two opposite strands and the methyl group could come from either thymine or 5mC. (**E–H**) Comparison of methylated DNA in complex with WT and A285 from outer base pair (T1:A1) to central base pair (M4:G4). (I-L) Comparison of A285 variant bound with methylated and unmodified DNA from base pairs 1 (outer) to 4 (central).

unchanged. This conservation includes the Ala-Val dipeptide that is associated in this study with relative insensitivity to the methylation status of the bound DNA (relative to the Ala-Ala dipeptide in most AP-1-related transcription factors). Excluding the Agnatha, substitutions in this region are highly focused on positions not directly involved in DNA contacts (either to bases or to the sugar-phosphate backbone).

Comparison of C/EBPβ V285A variant to other bZIP TFs containing di-alanine

As noted earlier, many family members of bZIP family have a conserved Ala-Ala dipeptide (di-alanine; Figure 1A). Some of them have been assayed for the effects of central CpG methylation within the CRE element (TGACGTCA), which decreases the binding affinities of CREB and ATF4 (17,51), as well as of MAX—a member of bHLH family which recognizes the E-box sequence (CACGTG) (27). [We note that the effect of CpA methylation has rarely been determined.]

We asked here whether a C/EBPβ V285A variant mimics the di-alanine-containing bZIP TFs in distinguishing C from 5mC of the central CpG dinucleotide, even though the residues immediately flanking the di-alanine or Ala-Val are not conserved (Figure 1A). Indeed, for the central CpG modifications, the V285A variant showed the strongest binding to unmodified cytosine; ∼2-fold weaker binding to 5mC (a small but significant effect); and >6-fold weaker binding to 5hmC (Table 1B). We suggest that Arg289, when unconstrained by Val285, can adopt a 'C-specific' conformation, and this might explain the reduced binding effects of CpG modification.

In the CREB bZIP DNA binding domain bound to unmodified DNA, the guanidine group of the corresponding arginine forms a single H-bond with the Guanine *N7* atom of the central CpG (Figure 4A) (50). In this conformation, the Arg guanidine group bridges between the Guanine *N7* atom and the phosphate group of the 5′ Cytosine, resulting in a bent conformation between the guanidine group and the aliphatic portion of the Arg side chain. The same bent conformation is observed in the corresponding arginine of Max and its binding partner Mad in recognizing the central

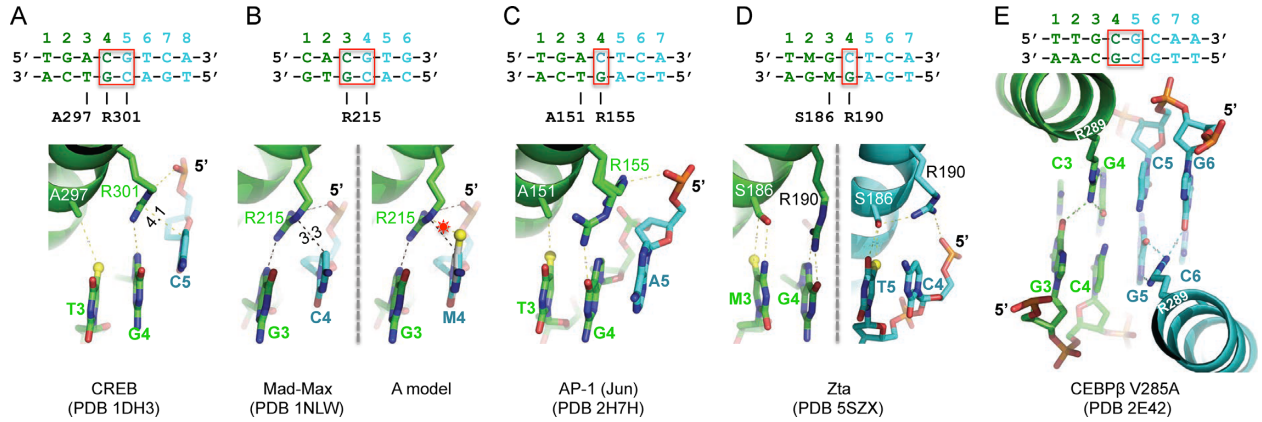

Figure 4. Comparison of C/EBPβ-related bZIP transcription factors bound with unmodified DNA. (**A**) CRE-binding (CREB) protein bound with unmodified CpG DNA (PDB 1DH3). Arg301 adopts a C-specific conformation. (**B**) The corresponding Arg in Mad/Max heterodimer adopts a similar C-specific conformation (PDB 1NLW) (left panel). Modeling a methyl group (in yellow ball) onto unmodified CpG site potentially results in repulsion (indicated by a star) with the Arg in the C-specific conformation (right panel). (**C**) AP-1 Jun dimer adopts alternative conformations in contact with the single central G:C base pair (PDB: 2H7H). (**D**) Epstein-Barr virus Zta dimer has two different conformations in each monomer. (**E**) With unmodified DNA, Arg289 of C/EBPβ connects two guanines (G4 and G3 or G5 and G6) of each half site.

CpG of the E-box hexanucleotide (Figure 4B) (59). In this conformation, the guanidine group is quite close to the *C5* atom of the CpG Cytosine (3.3–4.1 Å). Methylation at the *C5* atom is likely to sterically obstruct the Arg in this particular conformation, which we term 'C-specific' (Figure 4B, right panel), perhaps explaining the diminished CREB or Max binding to the response elements containing the central CpG modifications.

The bent 'C-specific' conformation of the conserved Arg in the bZIP family can be traced to the family members that recognize 7-bp elements with a single central G:C base pair, instead of a CpG dinucleotide. In the structure of Jun dimer bound with the pseudo-symmetric 7-bp AP-1 target site, the corresponding Arg forms two alternative conformations, one interacts with the central Guanine, and the other is in contact with the phosphate group of 5′ nucleotide (Figure 4C). In the Epstein-Barr Virus Zta, a viral transcriptional regulator that is homologous to AP-1, the corresponding Arg of each monomer adopts a different conformation, where one interacts with the central Guanine, and the other interacts with the DNA backbone phosphate group (Figure 4D) (19). In these two examples, the phosphate-interacting arginine conformation mimics the bent 'C-specific' conformation.

Interestingly, in the structure of C/EBPβ A285 variant bound with unmodified DNA (PDB: 2E42), the corresponding arginine (Arg289) did not take the bent 'C-specific' conformation; instead it bridged between the two guanines on opposite strands of each half site (between G3 and G4 for the green monomer, and between G5 and G6 for the cyan monomer) (Figure 4E). This is unique to C/EBP, because its recognition sequence has an additional G:C base pair immediately outside of the central CpG, whereas for the other examples analyzed (CREB/Max/AP-1/Zta) the immediate neighbor sequences of central CpG is A:T rich. It is possible that Arg289 might sample different conformations, and the minor 'C-specific' conformation might not be reflected in the crystal structure, in accordance with the twofold difference in binding affinity between CpG and 5mCpG. In contrast, the difference is much more pronounced in MAX (18).

Importance of recognizing CpA modification

Dnmt3A is capable of non-CpG (mostly CpA) methylation *in vitro* and *in vivo* (56,57). In fibroblasts, 99.98% of all methylation occurs at CpG dinucleotides, but in stem cells about 25% of methylated cytosines are found in non-CpG contexts (particularly in CpA) (60). The CpA methylation disappears upon induced differentiation of embryonic stem cells, and is restored in induced pluripotent stem cells due to expression of the reprogramming factors (60). It may not be a mere coincidence that the Yamanaka reprogramming factors recognize CpA-containing sequences (61). Recently, it was shown that Dnmt3A transiently binds transcribed regions of poorly-expressed genes in developing brain, initiating DNA methylation at CpA sequences, and that this methylation is inhibited by transcription (57). The 5mCpA/TpG is bound by the methyl-DNA-binding protein MeCP2 and reduces transcription, reinforcing regions of low transcriptional activity. We examined ChIP-seq data on RNA polymerase II associated transcription factors generated by the ENCODE consortium (46). Many C/EBP-related bZIP family and Max-related bHLH family transcription factors contain CpA/TpG in their consensus recognition sequences (Supplementary Figure S6), while sequences corresponding to the central CpG are variable or reduced to a single G:C base pair.

CpA/TpG could be considered an intrinsically 'hemi-methylated' DNA element. Methylation at CpA sites by Dnmt3 would then generate a 'fully methylated' 5mCpA/TpG dinucleotide. Indeed, the levels of 5mCpA/TpG undergo dynamic changes during differentiation of the germ line and in brain development from fetus to young adult (60,62,63). Furthermore, Tet enzymes are also active on thymine (5-methyluracil), generating 5-hydroxymethyluracil (5hmU) paired with adenine, which could be specifically bound by potential protein readers

(64,65). Like 5mC, the Tet-mediated hydroxylation of thymine provides an opportunity to establish a pseudo-symmetric, fully modified CpA/TpG dinucleotide. There is an intrinsic relationship between 5mCpG and TpG, as the latter can arise from the former via 5mC deamination (66) during development. Perhaps TpG dinucleotides are generated in a targeted manner (see discussion in (19)), when it is advantageous for a particular DNA sequence to be treated as if it is permanently methylated (see reviews (67–69)). On the other hand, TpG/CpA-to-CpG substitution is also possible in sperm cells with high levels of methylation (70). What we have shown here thus adds to understanding of the methylation marking/erasing/reading system, by elucidating the effects of both CpG and CpA modification (to 5mC and 5hmC) on a centrally-important transcription factor.

DATA AVAILABILITY

The X-ray structures (coordinates and structure factors) of wild-type C/EBPβ bZIP domain (PDB IDs: 6MG1 in *C*2 and 6MG2 in $C222_1$ space group) and V285A mutant (PDB 6MG3) have been deposited to Protein Data Bank.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank B. Baker of New England Biolabs for synthesizing the oligonucleotides.
Author contributions: J.Y. performed protein purification, DNA-binding experiments, mutagenesis and crystallization. J.R.H. performed X-ray data collection and structure determination; D.W. made initial expression construct, performed preliminary purification and DNA binding assays; R.R. assisted in DNA-binding assays and crystallization; J.L, D.S. and Y.H. analyzed published C/EBPβ ChIP-seq data; X.Z. and X.C. organized and designed the scope of the study; R.M.B. performed data analysis and assisted in preparing the manuscript.

FUNDING

U.S. National Institutes of Health (NIH) [GM049245-24 to X.C. and HL134780 to Y.H.]; Cancer Prevention and Research Institute of Texas [RR160029 to X.C. and RR140053 to Y.H.]; American Cancer Society [RSG-18-043-01-LIB to Y.H.]. The open access publication charge for this paper has been waived by Oxford University Press – *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
Conflict of interest statement. None declared.

REFERENCES

1. Okano,M., Xie,S. and Li,E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.*, **19**, 219–220.
2. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
3. Bestor,T., Laudano,A., Mattaliano,R. and Ingram,V. (1988) 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.*, **203**, 971–983.
4. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
5. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
6. Ren,R., Horton,J.R., Zhang,X., Blumenthal,R.M. and Cheng,X. (2018) Detecting and interpreting DNA methylation marks. *Curr. Opin. Struct. Biol.*, **53**, 88–99.
7. Scott-Browne,J.P., Cw,L. and Rao,A. (2017) TET proteins in natural and induced differentiation. *Curr. Opin. Genet. Dev.*, **46**, 202–208.
8. Wu,X. and Zhang,Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*, **18**, 517–534.
9. Lewis,J.D., Meehan,R.R., Henzel,W.J., Maurer-Fogy,I., Jeppesen,P., Klein,F. and Bird,A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, **69**, 905–914.
10. Sasai,N., Nakao,M. and Defossez,P.A. (2010) Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res.*, **38**, 5015–5022.
11. Buck-Koehntop,B.A., Stanfield,R.L., Ekiert,D.C., Martinez-Yamout,M.A., Dyson,H.J., Wilson,I.A. and Wright,P.E. (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 15229–15234.
12. Liu,Y., Toh,H., Sasaki,H., Zhang,X. and Cheng,X. (2012) An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes Dev.*, **26**, 2374–2379.
13. Liu,Y., Olanrewaju,Y.O., Zheng,Y., Hashimoto,H., Blumenthal,R.M., Zhang,X. and Cheng,X. (2014) Structural basis for Klf4 recognition of methylated DNA. *Nucleic Acids Res.*, **42**, 4859–4867.
14. Hashimoto,H., Olanrewaju,Y.O., Zheng,Y., Wilson,G.G., Zhang,X. and Cheng,X. (2014) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.*, **28**, 2304–2313.
15. Zhu,H., Wang,G. and Qian,J. (2016) Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.*, **17**, 551–565.
16. Hashimoto,H., Wang,D., Horton,J.R., Zhang,X., Corces,V.G. and Cheng,X. (2017) Structural basis for the versatile and Methylation-Dependent binding of CTCF to DNA. *Mol. Cell*, **66**, 711–720.
17. Kribelbauer,J.F., Laptenko,O., Chen,S., Martini,G.D., Freed-Pastor,W.A., Prives,C., Mann,R.S. and Bussemaker,H.J. (2017) Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Rep.*, **19**, 2383–2395.
18. Wang,D., Hashimoto,H., Zhang,X., Barwick,B.G., Lonial,S., Boise,L.H., Vertino,P.M. and Cheng,X. (2017) MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.*, **45**, 2396–2407.
19. Hong,S., Wang,D., Horton,J.R., Zhang,X., Speck,S.H., Blumenthal,R.M. and Cheng,X. (2017) Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein-Barr virus Zta. *Nucleic Acids Res.*, **45**, 2503–2515.
20. Yin,Y., Morgunova,E., Jolma,A., Kaasinen,E., Sahu,B., Khund-Sayeed,S., Das,P.K., Kivioja,T., Dave,K., Zhong,F. *et al.* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, **356**, doi:10.1126/science.aaj2239.
21. Liu,Y., Zhang,X., Blumenthal,R.M. and Cheng,X. (2013) A common mode of recognition for methylated CpG. *Trends Biochem Sci*, **38**, 177–183.
22. Liu,Y., Olanrewaju,Y.O., Zhang,X. and Cheng,X. (2013) DNA recognition of 5-carboxylcytosine by a Zfp57 mutant at an atomic resolution of 0.97 A. *Biochemistry*, **52**, 9310–9317.

23. Massari,M.E. and Murre,C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell Biol.*, **20**, 429–440.
24. Prendergast,G.C. and Ziff,E.B. (1992) A new bind for Myc. *Trends Genet.*, **8**, 91–96.
25. Torres,R., Schreiber-Agus,N., Morgenbesser,S.D. and DePinho,R.A. (1992) Myc and Max: a putative transcriptional complex in search of a cellular target. *Curr. Opin. Cell Biol.*, **4**, 468–474.
26. Blackwood,E.M. and Eisenman,R.N. (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science*, **251**, 1211–1217.
27. Prendergast,G.C., Lawe,D. and Ziff,E.B. (1991) Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell*, **65**, 395–407.
28. Tulchinsky,E.M., Georgiev,G.P. and Lukanidin,E.M. (1996) Novel AP-1 binding site created by DNA-methylation. *Oncogene*, **12**, 1737–1745.
29. Gustems,M., Woellmer,A., Rothbauer,U., Eck,S.H., Wieland,T., Lutter,D. and Hammerschmidt,W. (2014) c-Jun/c-Fos heterodimers regulate cellular genes via a newly identified class of methylated DNA sequence motifs. *Nucleic Acids Res.*, **42**, 3059–3072.
30. Paulson,E.J. and Speck,S.H. (1999) Differential methylation of Epstein-Barr virus latency promoters facilitates viral persistence in healthy seropositive individuals. *J. Virol.*, **73**, 9959–9968.
31. Bhende,P.M., Seaman,W.T., Delecluse,H.J. and Kenney,S.C. (2004) The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat. Genet.*, **36**, 1099–1104.
32. Bhende,P.M., Seaman,W.T., Delecluse,H.J. and Kenney,S.C. (2005) BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186. *J. Virol.*, **79**, 7338–7348.
33. Eferl,R. and Wagner,E.F. (2003) AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer*, **3**, 859–868.
34. Tsukada,J., Yoshida,Y., Kominato,Y. and Auron,P.E. (2011) The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. *Cytokine*, **54**, 6–19.
35. Roe,J.S. and Vakoc,C.R. (2014) C/EBPalpha: critical at the origin of leukemic transformation. *J. Exp. Med.*, **211**, 1–4.
36. Sun,C., Duan,P. and Luan,C. (2017) CEBP epigenetic dysregulation as a drug target for the treatment of hematologic and gynecologic malignancies. *Curr. Drug Targets*, **18**, 1142–1151.
37. Hirai,H., Zhang,P., Dayaram,T., Hetherington,C.J., Mizuno,S., Imanishi,J., Akashi,K. and Tenen,D.G. (2006) C/EBPbeta is required for 'emergency' granulopoiesis. *Nat. Immunol.*, **7**, 732–739.
38. Satake,S., Hirai,H., Hayashi,Y., Shime,N., Tamura,A., Yao,H., Yoshioka,S., Miura,Y., Inaba,T., Fujita,N. *et al.* (2012) C/EBPbeta is involved in the amplification of early granulocyte precursors during candidemia-induced "emergency" granulopoiesis. *J. Immunol.*, **189**, 4546–4555.
39. Lan,F., Collins,R.E., De Cegli,R., Alpatov,R., Horton,J.R., Shi,X., Gozani,O., Cheng,X. and Shi,Y. (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature*, **448**, 718–722.
40. Patel,A., Hashimoto,H., Zhang,X. and Cheng,X. (2016) Characterization of how DNA modifications affect DNA binding by C2H2 zinc finger proteins. *Methods Enzymol.*, **573**, 387–401.
41. Otwinowski,Z., Borek,D., Majewski,W. and Minor,W. (2003) Multiparametric scaling of diffraction intensities. *Acta Crystallogr. A*, **59**, 228–234.
42. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
43. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
44. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
45. Yu,M., Hon,G.C., Szulwach,K.E., Song,C.X., Zhang,L., Kim,A., Li,X., Dai,Q., Shen,Y., Park,B. *et al.* (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the Mammalian genome. *Cell*, **149**, 1368–1380.
46. Wang,J., Zhuang,J., Iyer,S., Lin,X., Whitfield,T.W., Greven,M.C., Pierce,B.G., Dong,X., Kundaje,A., Cheng,Y. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.*, **22**, 1798–1812.
47. Huang,Y., Pastor,W.A., Shen,Y., Tahiliani,M., Liu,D.R. and Rao,A. (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One*, **5**, e8888.
48. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
49. Montminy,M.R., Sevarino,K.A., Wagner,J.A., Mandel,G. and Goodman,R.H. (1986) Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 6682–6686.
50. Schumacher,M.A., Goodman,R.H. and Brennan,R.G. (2000) The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J. Biol. Chem.*, **275**, 35242–35247.
51. Iguchi-Ariga,S.M. and Schaffner,W. (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.*, **3**, 612–619.
52. McKnight,S.L. (2001) McBindall–a better name for CCAAT/enhancer binding proteins? *Cell*, **107**, 259–261.
53. Rishi,V., Bhattacharya,P., Chatterjee,R., Rozenberg,J., Zhao,J., Glass,K., Fitzgerald,P. and Vinson,C. (2010) CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20311–20316.
54. Khund Sayeed,S., Zhao,J., Sathyanarayana,B.K. and Vinson,C. (2015) C/EBPbeta (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochim. Biophys. Acta*, **1849**, 583–589.
55. Ramsahoye,B.H., Biniszkiewicz,D., Lyko,F., Clark,V., Bird,A.P. and Jaenisch,R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 5237–5242.
56. Gowher,H. and Jeltsch,A. (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J. Mol. Biol.*, **309**, 1201–1208.
57. Stroud,H., Su,S.C., Hrvatin,S., Greben,A.W., Renthal,W., Boxer,L.D., Nagy,M.A., Hochbaum,D.R., Kinde,B., Gabel,H.W. *et al.* (2017) Early-Life gene expression in neurons modulates lasting epigenetic states. *Cell*, **171**, 1151–1164.
58. Horowitz,S. and Trievel,R.C. (2012) Carbon-oxygen hydrogen bonding in biological structure and function. *J. Biol. Chem.*, **287**, 41576–41582.
59. Nair,S.K. and Burley,S.K. (2003) X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell*, **112**, 193–205.
60. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
61. Chen,X., Xu,H., Yuan,P., Fang,F., Huss,M., Vega,V.B., Wong,E., Orlov,Y.L., Zhang,W., Jiang,J. *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, **133**, 1106–1117.
62. Lister,R., Mukamel,E.A., Nery,J.R., Urich,M., Puddifoot,C.A., Johnson,N.D., Lucero,J., Huang,Y., Dwork,A.J., Schultz,M.D. *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**, 1237905.
63. Ichiyanagi,T., Ichiyanagi,K., Miyake,M. and Sasaki,H. (2013) Accumulation and loss of asymmetric non-CpG methylation during male germ-cell development. *Nucleic Acids Res.*, **41**, 738–745.
64. Pfaffeneder,T., Hackner,B., Truss,M., Munzel,M., Muller,M., Deiml,C.A., Hagemeier,C. and Carell,T. (2011) The discovery of

5-formylcytosine in embryonic stem cell DNA. *Angew. Chem.*, **50**, 7008–7012.
65. Pais,J.E., Dai,N., Tamanaha,E., Vaisvila,R., Fomenkov,A.I., Bitinaite,J., Sun,Z., Guan,S., Correa,I.R. Jr, Noren,C.J. *et al.* (2015) Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 4316–4321.
66. Wijesinghe,P. and Bhagwat,A.S. (2012) Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res.*, **40**, 9206–9217.
67. Patil,V., Ward,R.L. and Hesson,L.B. (2014) The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics*, **9**, 823–828.
68. Pinney,S.E. (2014) Mammalian Non-CpG methylation: Stem cells and beyond. *Biology (Basel)*, **3**, 739–751.
69. Jang,H.S., Shin,W.J., Lee,J.E. and Do,J.T. (2017) CpG and Non-CpG methylation in epigenetic gene regulation and brain function. *Genes (Basel)*, **8**, 148.
70. Miyahara,H., Hirose,O., Satou,K. and Yamada,Y. (2015) Factors to preserve CpG-rich sequences in methylated CpG islands. *BMC Genomics*, **16**, 144.