HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2021 January 04.

Published in final edited form as:

Biochemistry. 2016 December 13; 55(49): 6940–6948. doi:10.1021/acs.biochem.6b00796.

5-Methylcytosine (5mC) and 5-Hydroxymethylcytosine (5hmC) Enhance the DNA Binding of CREB1 to the C/EBP Half-Site Tetranucleotide GCAA

Khund Sayeed Syed[†], **Ximiao He**[†], **Desiree Tillo**[†], **Jun Wang**[†], **Stewart R. Durell**[‡], **Charles Vinson**[*,†]

[†]Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Room 3128, Building 37, Bethesda, Maryland 20892, United States

[‡]Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Room 3128, Building 37, Bethesda, Maryland 20892, United States

Abstract

In human and mouse stem cells and brain, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) can occur outside of CG dinucleotides. Using protein binding microarrays (PBMs) containing 60-mer DNA probes, we evaluated the effect of 5mC and 5hmC on one DNA strand on the double-stranded DNA binding of the mouse B-ZIP transcription factors (TFs) CREB1, ATF1, and JUND. 5mC inhibited binding of CREB1 to the canonical CRE half-site |GTCA but enhanced binding to the C/EBP half-site |GCAA. 5hmC inhibited binding of CREB1 to all 8-mers except TGAT|GCAA, where binding is enhanced. We observed similar DNA binding patterns with ATF1, a closely related B-ZIP domain. In contrast, both 5mC and 5hmC inhibited binding of JUND. These results identify new DNA sequences that are well-bound by CREB1 and ATF1 only when they contain 5mC or 5hmC. Analysis of two X-ray structures examines the consequences of 5mC and 5hmC on DNA binding by CREB and FOS|JUN.

Graphical Abstract

[*]**Corresponding Author:** Telephone: 1-301-496-8783. Fax: 1-301-496-8419. vinsonc@mail.nih.gov.

Author Contributions

K.S.S. and C.V. conceptualized the study and analyzed the results. K.S.S. and J.W. performed the PBM experiments. X.H. wrote the bioinformatics scripts for PBM data analysis and performed calculations for motif enrichment in DHSs and GO analysis. S.R.D. performed the structural modeling of proteins bound to DNA. All authors contributed to writing of the manuscript.

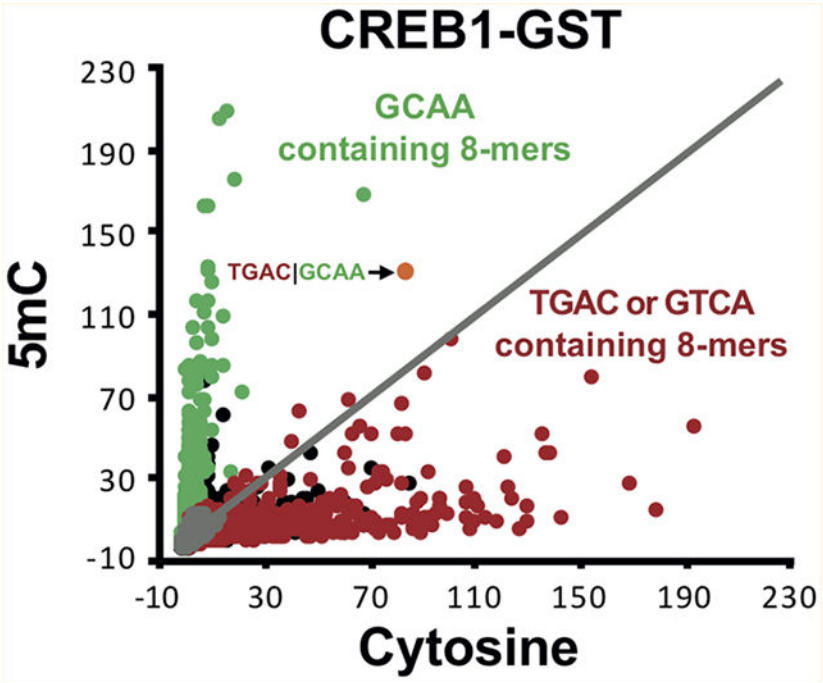
The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.bio-chem.6b00796.

A schematic of the modified PBM approach (Figure S1), comparisons of PBM replicates (Figures S2–S4), motif enrichment analysis in DHS (Figure S5), and promoter GO analysis (Figure S6) (PDF)



Sequence-specific DNA binding of transcription factors (TFs) is the foundation of regulated gene expression. Two recent observations have expanded the lexicon of sequence-specific DNA binding of mammalian TFs. The first is that methylated cytosines (5mC) are iteratively oxidized by the ten-eleven-translocation (TET) family of dioxygenases to 5-hydroxymethylcytosine (5hmC),[2] 5-formylcytosine (5fC), and 5-carboxycytosine (5caC).[3] 5fC and 5caC are removed by thymine DNA glycosylase (TDG), completing the cytosine demethylation cycle.[4,5] 5hmC accumulates in several tissues, suggesting it may have a regulatory function.[6] The second finding is that 5mC can occur in non-CG dinucleotides, initially being observed in stem cells[7] and later in the brain.[8] 5mC and 5hmC outside of CG dinucleotides have not been identified as the predominant modification in a population of cells, though future studies may identify situations in which they are more abundant.

5mC on both cytosines in CG dinucleotides can inhibit or enhance DNA binding of B-ZIP proteins.[9–12] The cAMP response element-binding protein (CREB1) regulates expression of cellular genes by binding to the consensus CRE sequence (TGAC|GTCA)[13,14] and its variants.[15,16] The vertical line in the middle of the 8-mer represents the center of the DNA motif dyad. Methylation of both cytosines in the CG dinucleotide at the center of the CRE motif inhibits CREB1 binding.[9,10,17] In contrast, methylation of the central CG dinucleotide in the C/EBP motif TTGC|GCAA enhances DNA binding of C/EBP family members.[9,10,18] While the effect of 5mC and 5hmC in CG dinucleotides on sequence-specific DNA binding of B-ZIP and other TFs has been examined,[10,19–21] their effect on DNA binding to non-CG dinucleotide-containing sequences has not been examined.

The B-ZIP domain is a long α helix with a bipartite structure. The C-terminal leucine zipper region mediates dimerization, and the N-terminal basic region mediates sequence-specific DNA binding.[22] The amino acid determinants of leucine zipper-mediated B-ZIP

dimerization specificity and stability have been investigated in detail.[15,23–28] However, the relationship between the amino acid sequence of the B-ZIP basic region and sequence-specific DNA binding, including DNA sequences containing 5mC and 5hmC, has not been examined.

Previously, we used 5mC or 5hmC to double-strand the oligonucleotides on Agilent DNA microarrays, followed by a protein binding microarray experiment,[29] to evaluate the effects of these cytosine modifications on sequence-specific DNA binding of two B-HLH TFs.[30] In this study, we used the same approach to examine the sequence-specific DNA binding of three B-ZIP TFs (CREB1, ATF1, and JUND) to microarrays containing cytosine, 5mC, or 5hmC.

EXPERIMENTAL PROCEDURES

Cloning and Expression of Mouse B-ZIP DNA-Binding Domains.

The B-ZIP DNA-binding domains (DBDs) of mouse CREB1, ATF1, and JUND were obtained from T. Hughes (University of Toronto, Toronto, ON) as a GST construct cloned into the pETGEXCT (C-terminal GST) vector.[31] All three proteins were individually expressed using the PureExpress *in vitro* protein synthesis kit (NEB),[10] in a 25 μL reaction volume containing 180 ng of plasmid.

Amino Acid Sequences of All Three Proteins Used in This Study.

The basic region DNA-binding domains are shown in bold in Chart 1.

Double-Stranding the Agilent 40K Array with Cytosine, 5mC, or 5hmC.

The specifics of the Agilent 40K array design and use of PBMs have been described elsewhere.[10,29,32,33] Briefly, these arrays comprise ~40000 single-stranded 60-mer oligonucleotides, each containing a variable probe sequence that is 35 bases long and a common 25-base sequence near the glass surface, which is complementary to the primer sequence used in DNA double-stranding. The 35-mer variable region is based on deBruijn sequences, which allow for all possible 8-mer sequences to be represented 32 times on the array. Specifically, we used the “HK” array design available on the NCBI Gene Expression Omnibus (GEO) platform GPL11260 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL11260).

To assess the effect of modified cytosines in a non-CG sequence context on TF binding, we modified the double-stranding procedure[30,32] by using either 5-methylcytosine (5mC, NEB) or 5-hydroxymethylcytosine (5hmC, Zymo Research). The resulting double-stranded DNA on the array will contain either cytosine on both strands or 5mC or 5hmC on one strand and cytosine on the second strand. This results in a hemimethylated or hemihydroxymethylated state. DNA double-stranding was performed as previously described.[10,34] The double-stranding reaction mixture was spiked with Cy3-dCTP (4%) to monitor the double-stranding efficiency.[10]

Protein Binding Reaction.

Protein binding reactions were performed as described previously.[10] Briefly, 180 ng of plasmid containing DNA-binding domains of CREB1, ATF1, or JUND was used to express proteins using the PureExpress *in vitro* transcription translation kit (NEB) in a 25 μL reaction volume following the manufacturer's instructions. The double-stranded arrays were blocked with 4% milk for 1 h and washed with 0.1% Tween 20 in 1× PBS. Freshly synthesized protein (25 μL) was mixed with 125 μL of a protein binding reaction mixture consisting of 4% milk in 1 × PBS, 50 ng of salmon testes DNA, and 0.2 μg/μL bovine serum albumin and added to the double-stranded array. The protein binding reactions were performed in the hydration chamber for 1 h followed by one wash with 0.5% Tween 20 in 1× PBS. The protein bound array was incubated with the Alexa Fluor 647-conjugated anti-GST antibody for 1 h, followed by three washes with 0.05% Tween 20. Finally, array slides were washed and dried in 1 × PBS and scanned using an Agilent Sure Scan II scanner.

Image Quantification and Analysis of *Z*-Scores.

For each protein-bound microarray, image quantification and calculation of *Z*-scores were performed as described previously.[30] Microarray images were analyzed using ImaGene (BioDiscovery Inc.), and the extracted data (probe intensity values) were used for further analysis. The probe median intensities were used to calculate the *Z*-score for all 65536 8-mers. In previous studies, complementary 8-mers were combined,[10] but because of the asymmetric nature of the double-stranding protocol, complementary 8-mers are different. We then consider only the *Z*-score of the reverse complement of the 8-mer extracted from the array probe design as this represents the sequences that contain the modified cytosines in our double-stranding reaction. Thus, all 8-mers shown are the strands that can contain 5mC or 5hmC. In Figures 1 and 2, *Z*-scores on the *y*-axis were rescaled using the slope of the line of best fit computed from the *Z*-scores using 8-mers with no cytosine as a control. All proteins were assayed in duplicate with good agreement between replicates (Figures S2–S4). Data (raw probe intensities and 8-mer *Z*-scores) are available at the NCBI GEO database under accession code GSE88897.

Genomic Analysis of 8-mers in Regulatory Regions.

Human DNaseI hypersensitive site (DHS) data from 125 tissue and cell lines are from the ENCODE project[35] and were downloaded from the University of California Santa Cruz Genome Bioinformatics Web site (http://genome.ucsc.edu/).[36]

To determine the enrichment of each 8-mer in the DHSs, we calculated an enrichment score for each 8-mer[37] across the human genome (UCSC build hg19). For each 8-mer M of length L [e.g., the C/EBP motif (TTGCGCAA); $L = 8$], we denote $M(x_{\text{start}}{:}x_{\text{end}})$ to record the positions where the motif starts and ends: $x_1{:}x_1 + L - 1$, $x_2{:}x_2 + L - 1$,... $x_N{:}x_N + L - 1$, N being the total number of motifs in the genome. For each position $x_i{:}x_i + L - 1$, if it overlapped with the examined regions (DHSs), $x_i = 1$; otherwise, $x_i = 0$.

For all the DHSs, the observed (OCC_{obs}) and expected (OCC_{exp}) occurrences of the 8-mer are calculated as $OCC_{obs} = \sum_{i=1}^{N} x_i$ and $OCC_{exp} = N(L_r/L_g)$, respectively, where N is the total number of that 8-mer in the genome, L_r is the total length of base pairs in the examined

regions (DHSs), and L_g is the length of the genome. The enrichment score (E) for 8-mer M is then calculated with the equation $E = OCC_{obs}/OCC_{exp}$, where OCC_{obs} is the observed number of occurrences and OCC_{exp} is the expected number of occurrences of 8-mer M in the examined regions (DHSs).

Gene ontology (GO) analysis of 8-mer-containing promoters was performed using DAVID.[38]

Structural Modeling.

Two crystal structures are available for investigation of the effects of cytosine modification on the binding of the proteins studied here. These are the structures of homodimeric CREB1 bound to palindromic CRE ($T^{-4}G^{-3}$-$A^{-2}C^{-1}|G^{1}T^{2}C^{3}A^{4}$)[39] (PDB entry 1DH3) and heterodimeric cFos|cJun binding to the canonical AP1 or TRE motif ($T^{-4}G^{-3}A^{-2C}/_{G}{}^{0}T^{2}C^{3}A^{4}$) (PDB entry 1FOS).[40] Molecular models were developed with the CHARMM software package.[41] Figures were generated with the UCSF Chimera package.[42]

RESULTS

Protein Binding Microarrays (PBMs) for Evaluating the Effect of 5mC and 5hmC on DNA Binding.

DNA polymerases can incorporate 5mC and 5hmC into DNA when double-stranding single-stranded DNA.[43] Previously, we exploited this property to double-strand the single-stranded DNA on an Agilent microarray using 5mC or 5hmC.[30] This generates double-stranded arrays with 5mC or 5hmC on one DNA strand, mimicking what occurs biologically in several cell types, including brain.[8,12,54] These arrays are then used in protein binding microarray (PBM) experiments (Figure S1). We are using an Agilent design termed the HK array, containing 40000 features, which include all possible eight-base sequences (8-mers) represented 32 times on the array when complementary 8-mers are combined. We use a standardized score (Z-score) from PBM data that reflects the binding affinity of a given TF for a particular 8-mer.[29] Previously, we combined complementary 8-mers,[10] but here we present Z-scores for all 65596 8-mers as complementary 8-mers can be different because of the double-stranding protocol. All shown 8-mers are those that can contain 5mC or 5hmC.

5mC Enhances DNA Binding of CREB1 and ATF1 to 8-mers Containing the C/EBP Half-Site Tetranucleotide |GCAA.

Figure 1A shows a scatter plot of Z-scores for CREB1-GST binding to DNA 8-mers (see Experimental Procedures) containing cytosine on both DNA strands (x-axis) or 5mC on one strand and cytosine on the second strand (y-axis). As an internal control, we plot 8-mers with no cytosines in gray. There is no 8-mer without a cytosine that is well-bound by CREB1-GST. For 8-mers containing cytosine, two groups are evident: those that are well-bound by CREB1-GST only when they contain cytosine and those that are well bound only when they contain 5mC on one strand. The well-bound 8-mers with cytosine contain the CRE half-site |GTCA with the best bound 8-mer being the canonical CRE 8-mer[13] containing two cytosines at positions −1 and 3 of the motif $T^{-4}G^{-3}A^{-2}C^{-1}|G^{1}T^{2}C^{3}A^{4}$. Nucleotides are numbered relative to their position from the center of the motif dyad as in

the crystal structure coordinate file,[39] and this pattern of numbering has been followed for all other 8-mers described in subsequent sections. The 8-mers containing 5mC on one DNA strand that are well-bound by CREB1-GST contain the C/EBP half-site |GC^{2}AA with 5mC at position 2 in the motif.

We next examined 8-mers with only one cytosine to evaluate the contribution of individual 5mCs at different positions in the motif to the DNA binding of CREB1-GST (Figure 1B). Three 8-mers are highlighted, TGAC^{-1}|GTAA, TGAT|GTC^{3}A, and TGAT|GC^{2}AA with cytosine in three positions in the motif. 8-mers containing one cytosine that are at other positions in the motif are not well-bound. 5mC inhibits binding for the 8-mer TGAC^{-1}|GTAA with a cytosine at position −1. This is consistent with the previous observation that methylation of both cytosines in the central CG dinucleotide of the CRE motif inhibits CREB1 binding.[9] 5mC also inhibits CREB1-GST binding for the 8-mer TGAT|GTC^{3}A containing a cytosine at position 3, indicating that 5mC at this position inhibits DNA binding. The 8-mer TGAT|GC^{2}AA, with a cytosine at position 2 and containing the C/EBP half-site |GCAA, is well bound only when it contains 5mC. A few 8-mers are well bound with either cytosine or 5mC, including the CRE|C/EBP chimeric 8-mer TGAC^{-1}|GC^{2}AA (Figure 1A),[10,15,16,44] at positions that inhibit (C^{-1}) and enhance (C^{2}) binding.

We next examined how cytosine and 5mC affected the DNA binding of the closely related ATF1-GST[25] and obtained qualitatively similar results (Figure 1C,D). While the trends were the same, the magnitudes of *Z*-scores for ATF1 were lower than those of CREB1 (compare the *y*-axis of Figure 1C to the *y*-axis of Figure 1A, and Figures S1 and S2).

The binding results for JUND are shown in Figure 1E. As expected, the best bound 8-mer with cytosine was the canonical AP-1 motif TGAGTCAT.[45] In contrast to that of CREB1 and ATF1, DNA binding of JUND was uniformly inhibited by 5mC.

5hmC Enhances DNA Binding of CREB1 and ATF1 to 8-mers Containing the C/EBP Half-Site |GCAA.

We next examined the effect of 5hmC on DNA binding of CREB1, ATF1, and JUND (Figure 2). As observed with 5mC, 5hmC inhibited DNA binding of CREB1 and ATF1 to 8-mers containing the CRE half-site |GTC^{3}A (Figure 2A,B). With the exception of a few 8-mers, 5hmC obliterated DNA binding of CREB1 and ATF1 to most 8-mers that were well bound by 5mC, with the TGAT|GC^{2}AA 8-mer being best bound (for CREB1, compare the *y*-axis of Figure 1A to the *y*-axis of Figure 2A; for ATF1, compare Figure 1C to Figure 2B). As with 5mC, the magnitudes of *Z*-scores for ATF1 were lower than those of CREB1. 5hmC inhibited DNA binding of JUND to all cytosine-containing 8-mers as observed for 5mC (Figure 2C).

Structural Analysis of the Effects of 5mC and 5hmC on B-ZIP TF Binding.

We analyzed two X-ray structures, CREB1 binding to a canonical CRE motif T^{-4}G^{-3}A^{-2}C^{-1}|G^{1}T^{2}C^{3}A^{4}[39] (PDB entry 1DH3) and a cFos|cJun heterodimer binding to the AP-1 motif T^{-4}G^{-3}A^{-2}C/$_{g}$^{0}T^{2}C^{3}A^{4} (PDB entry 1FOS)[40] as a proxy for the JunD homodimer, to provide a physical explanation for the observed effects of 5mC and 5hmC on sequence-specific DNA binding. Both cFos and cJun are homologous to JUND. Figure 3A shows the main region of

contact of the CREB1 protein with the nucleotides of the CRE motif ($C^{-1}|G^1T^2C^3$). Panels B and C of Figure 3 show the results of modeling-in methyl groups to represent the 5mC modifications of C^{-1} and C^3. For C^{-1}, the added methyl group packs next to the side chain of R301, with a small degree of atomic overlap. At C^3, the added methyl significantly overlaps the NH_2 group of the side chain of N293. Similar steric conflicts occur for the 5hmC modification. We suggest these steric clashes explain the decrease in CREB1 binding affinity that occurs with 5mC- and 5hmC-modified CRE DNA (Figures 1 and 2).

Panels D and E of Figure 3 show the interactions of cJun with 5mC for C^{-1} and C^3, respectively. This could be modeled because in the crystal the cJun|cFos dimer straddles the DNA helix in both possible orientations. These demonstrate significant steric clashes of the added methyl group on C^1 with the side chain of R279, and on C^3 with N271, the same asparagine that causes a steric clash with C^3 in the CREB1 DNA structure (Figure 3C). Similar clashes would occur for the 5hmC modifications, and these are likely what cause the drastic reduction in binding affinity observed for JUND (Figures 1E and 2C).

Figure 4A shows a model of CREB1 binding the CRE|C/EBP motif ($TGAT|G^1C^2A^3A^4$), which was generated to examine how 5mC enhances binding at C^2. The change from the CRE ($|G^1T^2C^3A^4$) to C/EBP ($|G^1C^2A^3A^4$) half-sites involves a switch from T^2C^3 to C^2A^3. While there is a gap between the base of the new C^2 and the side chain of A297, the new A^3 packs well, forming hydrogen bonds with the side chain of N293. Figure 4B shows the same model with 5mC-modifed DNA at C^2. The added methyl group to C^2 fills the gap to form a stabilizing hydrophobic bond with the side chain of A297. This recapitulates the interaction between the methyl group of T^2 with A297 in the CREB1/CRE structure (Figure 3A). Thus, 5mC plays the same structural role that thymine does. The added stabilization of this hydrophobic bond is seen in the increase in *Z*-scores for the 8-mer $TGAT|GC^2AA$ with 5mC (Figure 1A,C). Figure 4C shows that for the 5hmC modification, the added hydroxyl group is able to form another stabilizing interaction with A297, a hydrogen bond with the backbone carbonyl oxygen, supporting our finding that 5hmC modification enhances CREB1 binding to the same $TGAT|GC^2AA$ sequence (Figure 2A). Why the JunD homodimer does not bind $|GC^2AA$ is not obvious from the examination of the cFos|cJun heterodimer structure. Potentially, the structure is a poor proxy for JunD homodimer-binding DNA, or sequence-specific DNA binding is subtle on an atomic level.

$TGAT|GC^2AA$ Is Enriched in Regulatory Regions in Brain and Promoters of Genes Involved in Brain Function.

We examined human regulatory regions [DNasel hypersensitive sites (DHS)][46] and promoters to illuminate the potential biological function of $TGAT|GC^2AA$, the 8-mer in which CREB1 binding is enhanced when it is modified by 5hmC (Figure S5). We selected two negative control motifs with a base composition similar to that of the TGAT|GCAA sequence but not bound by any known TF. These negative control sequences are not enriched in any of the DHS examined, whereas C/EBP, CRE, and CRE|C/EBP chimeric motifs are enriched in DHS in all 125 samples. $TGAT|GC^2AA$ is enriched in DHSs from fewer samples, including DHSs derived from all brain cell types considered (Figure S5). Additional gene ontology analysis indicates that $TGAT|GC^2AA$-containing promoters are

enriched for genes involved in biological processes related to brain function, including cognition, neurological system processes, and sensory perception (Figure S6).

DISCUSSION

In brain and stem cells, 5mC and 5hmC occur outside of CG dinucleotides, broadening the landscape of sequence-specific DNA binding of TFs. We have examined how 5mC and 5hmC on one strand of double-stranded DNA alters the DNA binding specificity of three B-ZIP homodimers using Agilent DNA microarrays. 5mC and 5hmC enhanced binding to some 8-mers containing the C/EBP half-site |GC^{2}AA Similar results were obtained for the closely related ATF1. Both 5mC and 5hmC inhibit the binding of JUND to DNA. This is the first report describing enhanced DNA binding for CREB1 when DNA contains 5mC or 5hmC. These observations suggest that cytosine modifications differentially affect the DNA binding activities of three TFs containing the same DNA-binding domain, the B-ZIP domain.

The structural analysis of CREB1 and cFos|cJun bound to DNA containing 5mC and 5hmC identified a steric clash when our experimental data indicated that the modification inhibited DNA binding. This included CREB1 binding a CRE motif (T^{-4}G^{-3}A^{-2}C^{-1}|G^{1}T^{2}C^{3}A^{4}), where C^{3} contains 5mC or 5hmC and Jun binding the cytosines in the AP-1 motif. The preferential binding of CREB1 to the 5mC-containing C/EBP half-site |GC^{2}AA highlights how thymine and 5mC are structurally similar, possessing a methyl group on the same carbon in the pyrimidine ring. This provides an explanation for why CREB1 can bind the CRE half-site |GT^{2}CA when it contains unmodified cytosines and |GC^{2}AA when it contains 5mC.

We note that in the CREB/CRE structure, a conserved cysteine is replaced with a serine to prevent disulfide bond formation,[39] and the consequence for the crystal structure and DNA binding is unclear. Furthermore, in both the Fos-Jun| DNA[40] and the viral B-ZIP-containing protein ZTA|DNA structures, the conserved cysteine residues are replaced with serines.[47] In the case of the ZTA protein, this mutation compromises its function,[48] suggesting a change in DNA binding specificity.

A limitation of our experimental system is that we cannot explore the effects of binding of TF to 8-mers containing different combinations of modifications. However, our PBM experiments and structural analyses identify positions in the motif in which modified cytosines can inhibit (e.g., C^{-1} and C^{3}) and promote (e.g., C^{2}) binding by CREB1 or ATF1, allowing predictions regarding the binding affinities of novel sequences to be made. For example, the 8-mer TGAC^{-1}|GC^{2}AA with cytosine at C^{-1} and 5mC at C^{2} would be better bound than any 8-mer we identified.

A dramatic result is that CREB1 binds only one 8-mer well when it contains 5hmC. This sequence, TGAT|GC^{2}AA, is enriched in mammalian genomes and arises when the methylated CRE|C/EBP chimeric 8-mer TGAC^{-1}|GC^{2}AA spontaneously deaminates.[34] This deamination product is 10 times more abundant than its parental sequence. Our genomic analysis indicates that this 8-mer is modestly enriched in DHSs in a subset of cell types,

including brain, unlike the C/EBP, CRE, or CRE|CEBP consensus motifs, which are highly enriched in DHSs of all cell types we examined. This suggests that while most occurrences of C/EBP, CRE, or CRE|C/EBP consensus motifs are associated with regulatory function in all cell types, only a subset of TGAT|GC^{2}AA occurrences may function in a cell- or condition-specific manner, such as those that contain 5mC and 5hmC in brain cells. TGAT|GC^{2}AA is bound by other TFs, such as the ATF4|CEBPB heterodimer, which is involved in the stress response[10] when it does not contain 5mC or 5hmC. We speculate that different cytosine modifications in this 8-mer would recruit different TFs (ATF4|CEBPB heterodimers binding the unmodified 8-mer and CREB1 binding the modified 8-mer), allowing for the activation of different transcriptional networks and biological processes.

CREB1 is known to function in the brain that contains these modified cytosines, suggesting that these modifications may be biologically relevant. The role of CREB1 has been extensively studied in memory formation.[49] Dynamic changes in DNA methylation and demethylation in postmitotic neurons are necessary for long-term memory formation.[50,51] Fear memory upregulates DNMT3a and DNMT3b in hippocampus but not DNMT1,[52] and TET1 deletion leads to impairment of hippocampal neurogenesis and memory deficits.[53] Importantly, non-CG methylation is abundant in brain tissues compared to other tissues;[8,54] however, their roles are not yet understood.

The biological importance of CREB1 binding DNA containing 5mC and 5hmC is difficult to evaluate. Cytosine modifications outside of CG dinucleotides are rare and never become the prominent cytosine modification in a population of cells, making it very cumbersome to biochemically examine them *in vivo.* We are hopeful that biological systems in which cytosines outside of CG dinucleotides are predominantly 5mC or 5hmC will be identified, allowing one to examine their function. Alternatively, it may be possible to design alleles that do not bind to unmodified DNA but still bind to 5hmC containing DNAs. If these alleles have biological activity, it would suggest that 5hmC binding is biologically important.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work is supported by intramural research grants from the National Cancer Institute, National Institutes of Health.

ABBREVIATIONS

5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
CREB	cyclic AMP response element-binding protein
DHS	DNaseI hypersensitive site

PBM	protein binding microarray
PDB	Protein Data Bank
TF	transcription factor

REFERENCES

(1). Ptashne M (1988) How eukaryotic transcriptional activators work. Nature 335, 683–689. [PubMed: 3050531]

(2). Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, and Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935. [PubMed: 19372391]

(3). Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, and Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333, 1300–1303. [PubMed: 21778364]

(4). Kohli RM, and Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502, 472–479. [PubMed: 24153300]

(5). He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, and Xu GL (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333, 1303–1307. [PubMed: 21817016]

(6). Kriaucionis S, and Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324, 929–930. [PubMed: 19372393]

(7). Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, and Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315–322. [PubMed: 19829295]

(8). Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, and Ecker JR (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341, 1237905. [PubMed: 23828890]

(9). 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. [PubMed: 21059933]

(10). Mann IK, Chatterjee R, Zhao J, He X, Weirauch MT, Hughes TR, and Vinson C (2013) CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. Genome Res. 23, 988–997. [PubMed: 23590861]

(11). Gustems M, Woellmer A, Rothbauer U, Eck SH, 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. [PubMed: 24371273]

(12). Spruijt CG, and Vermeulen M (2014) DNA methylation: old dog, new tricks? Nat. Struct. Mol. Biol. 21, 949–954. [PubMed: 25372310]

(13). Montminy MR, and Bilezikjian LM (1987) Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. Nature 328, 175–178. [PubMed: 2885756]

(14). Mayr B, and Montminy M (2001) Transcriptional regulation by the phosphorylation-dependent factor CREB. Nat. Rev. Mol. Cell Biol. 2, 599–609. [PubMed: 11483993]

(15). Vinson CR, Hai T, and Boyd SM (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. Genes Dev. 7, 1047–1058. [PubMed: 8504929]

(16). Flammer JR, Popova KN, and Pflum MK (2006) Cyclic AMP response element-binding protein (CREB) and CAAT/enhancer-binding protein beta (C/EBPbeta) bind chimeric DNA sites with high affinity. Biochemistry 45, 9615–9623. [PubMed: 16878996]

(17). Iguchi-Ariga SM, 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. [PubMed: 2545524]

(18). He X, Syed KS, Tillo D, Mann I, Weirauch MT, and Vinson C (2015) GABPalpha Binding to Overlapping ETS and CRE DNA Motifs Is Enhanced by CREB1: Custom DNA Microarrays. G3: Genes, Genomes, Genet. 5, 1909–1918.

(19). Golla JP, Zhao J, Mann IK, Sayeed SK, Mandal A, Rose RB, and Vinson C. (2014) Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. Biochem. Biophys. Res. Commun. 449, 248–255. [PubMed: 24835951]

(20). Khund Sayeed S, Zhao J, Sathyanarayana BK, Golla JP, 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, Gene Regul Mech. 1849, 583–589.

(21). Zhu H, Wang G, and Qian J (2016) Transcription factors as readers and effectors of DNA methylation. Nat. Rev. Genet. 17, 551–565. [PubMed: 27479905]

(22). Vinson CR, Sigler PB, and McKnight SL (1989) Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science 246, 911–916. [PubMed: 2683088]

(23). Krylov D, Mikhailenko I, and Vinson C (1994) A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. EMBO J. 13, 2849–2861. [PubMed: 8026470]

(24). Moitra J, Szilak L, Krylov D, and Vinson C (1997) Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. Biochemistry 36, 12567–12573. [PubMed: 9376362]

(25). Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, and Bonovich M (2002) Classification of human B-ZIP proteins based on dimerization properties. Mol. Cell. Biol. 22, 6321–6335. [PubMed: 12192032]

(26). Acharya A, Ruvinov SB, Gal J, Moll JR, and Vinson C (2002) A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. Biochemistry 41, 14122–14131. [PubMed: 12450375]

(27). Newman JR, and Keating AE (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science 300, 2097–2101. [PubMed: 12805554]

(28). Reinke AW, Baek J, Ashenberg O, and Keating AE (2013) Networks of bZIP protein-protein interactions diversified over a billion years of evolution. Science 340, 730–734. [PubMed: 23661758]

(29). Berger MF, and Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat. Protoc. 4, 393–411. [PubMed: 19265799]

(30). Khund-Sayeed S, He X, Holzberg T, Wang J, Rajagopal D, Upadhyay S, Durell SR, Mukherjee S, Weirauch MT, Rose R, and Vinson C (2016) 5-Hydroxymethylcytosine in E-box motifs ACAT|GTG and ACAC|GTG increases DNA-binding of the B-HLH transcription factor TCF4. Integr Biol. 8, 936–945.

(31). Sharrocks AD (1994) A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase. Gene 138, 105–108. [PubMed: 8125285]

(32). Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, and Bulyk ML (2009) Diversity and complexity in DNA recognition by transcription factors. Science 324, 1720–1723. [PubMed: 19443739]

(33). Lam KN, van Bakel H, Cote AG, van der Ven A, and Hughes TR (2011) Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. Nucleic Acids Res. 39, 4680–4690. [PubMed: 21321018]

(34). He X, Tillo D, Vierstra J, Syed KS, Deng C, Ray GJ, Stamatoyannopoulos J, FitzGerald PC, and Vinson C (2015) Methylated Cytosines Mutate to Transcription Factor Binding Sites that Drive Tetrapod Evolution. Genome Biol. Evol. 7, 3155–3169. [PubMed: 26507798]

(35). The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. [PubMed: 22955616]

(36). Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AF, Speir ML, Zweig AS, Haussler D, Kuhn RM, and Kent WJ (2015) The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 43, D670–681. [PubMed: 25428374]

(37). Chatterjee R, He X, Huang D, FitzGerald P, Smith A, and Vinson C (2014) High-resolution genome-wide DNA methylation maps of mouse primary female dermal fibroblasts and keratinocytes. Epigenet. Chromatin 7, 35.

(38). Huang DW, Sherman BT, and Lempicki RA (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

(39). Schumacher MA, Goodman RH, and Brennan RG (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. [PubMed: 10952992]

(40). Glover JN, and Harrison SC (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. Nature 373, 257–261. [PubMed: 7816143]

(41). Brooks BR, Brooks CL 3rd, Mackerell AD Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, and Karplus M (2009) CHARMM: the biomolecular simulation program. J. Comput. Chem. 30, 1545–1614. [PubMed: 19444816]

(42). Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF Chimera-a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. [PubMed: 15264254]

(43). Chen CC, Wang KY, and Shen CK (2012) The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. J. Biol. Chem. 287, 33116–33121. [PubMed: 22898819]

(44). Moll JR, Acharya A, Gal J, Mir AA, and Vinson C (2002) Magnesium is required for specific DNA binding of the CREB B-ZIP domain. Nucleic Acids Res. 30, 1240–1246. [PubMed: 11861917]

(45). Lee W, Mitchell P, and Tjian R (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell 49, 741–752. [PubMed: 3034433]

(46). Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, Stehling-Sun S, Sabo PJ, Byron R, Humbert R, Thurman RE, Johnson AK, Vong S, Lee K, Bates D, Neri F, Diegel M, Giste E, Haugen E, Dunn D, Wilken MS, Josefowicz S, Samstein R, Chang KH, Eichler EE, De Bruijn M, Reh TA, Skoultchi A, Rudensky A, Orkin SH, Papayannopoulou T, Treuting PM, Selleri L, Kaul R, Groudine M, Bender MA, and Stamatoyannopoulos JA (2014) Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science 346, 1007–1012. [PubMed: 25411453]

(47). Petosa C, Morand P, Baudin F, Moulin M, Artero JB, and Muller CW (2006) Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. Mol. Cell 21, 565–572. [PubMed: 16483937]

(48). Karlsson QH, Schelcher C, Verrall E, Petosa C, and Sinclair AJ (2008) Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein. PLoS Pathog. 4, e1000005. [PubMed: 18369464]

(49). Kandel ER (2012) The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. Mol. Brain 5, 14. [PubMed: 22583753]

(50). Heyward FD, and Sweatt JD (2015) DNA Methylation in Memory Formation: Emerging Insights. Neuroscientist 21, 475–489. [PubMed: 25832671]

(51). Alaghband Y, Bredy TW, and Wood MA (2016) The role of active DNA demethylation and Tet enzyme function in memory formation and cocaine action. Neurosci. Lett. 625, 40–46. [PubMed: 26806038]

(52). Miller CA, and Sweatt JD (2007) Covalent modification of DNA regulates memory formation. Neuron 53, 857–869. [PubMed: 17359920]

(53). Zhang RR, Cui QY, Murai K, Lim YC, Smith ZD, Jin S, Ye P, Rosa L, Lee YK, Wu HP, Liu W, Xu ZM, Yang L, Ding YQ, Tang F, Meissner A, Ding C, Shi Y, and Xu GL (2013) Tet1 regulates adult hippocampal neurogenesis and cognition. Cell Stem Cell 13, 237–245. [PubMed: 23770080]

(54). Wen L, Li X, Yan L, Tan Y, Li R, Zhao Y, Wang Y, Xie J, Zhang Y, Song C, Yu M, Liu X, Zhu P, Li X, Hou Y, Guo H, Wu X, He C, Li R, Tang F, and Qiao J (2014) Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol. 15, R49. [PubMed: 24594098]

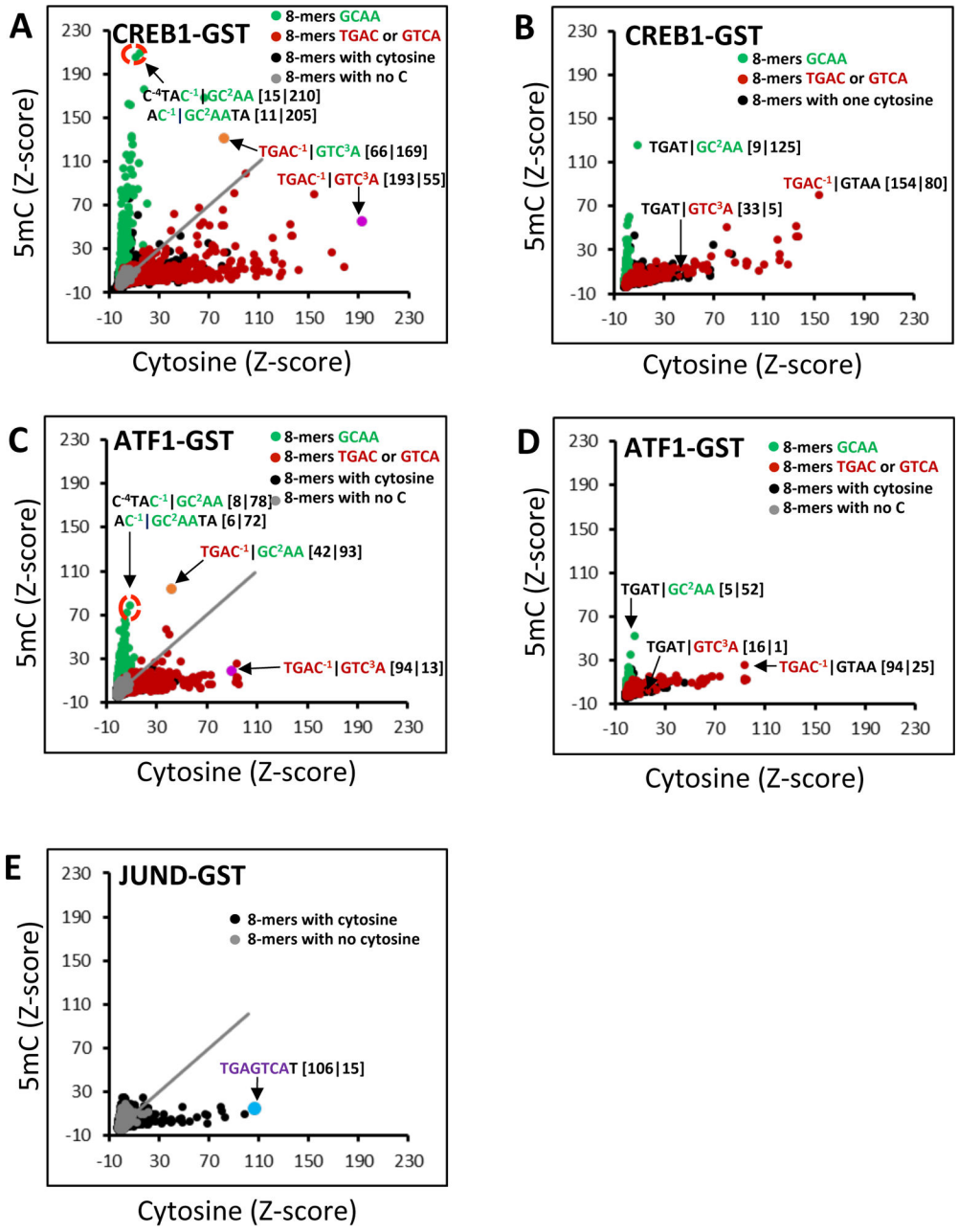


Figure 1.

CREB1, ATF1, and JUND binding to DNA 8-mers containing cytosine or 5mC. (A) *Z*-Scores for binding of CREB1-GST to DNA 8-mers with cytosine (*x*-axis) or 5mC (*y*-axis) on one DNA strand. 8-mers that do not contain a cytosine appear as gray spots, and those containing cytosine are shown as black spots. 8-mers containing the C/EBP half-site | GC^2AA (# = 1279) are colored green and those with CRE half-sites $TGAC^{-1}$| or | GTC^3A (# = 2554) red. Four 8-mers are highlighted. For the sake of clarity, the CRE half-site is colored red and the C/EBP half-site green. The *Z*-scores for binding to cytosine|5mC are given in brackets. (B) CREB1-GST binding to DNA 8-mers containing one cytosine (*x*-axis) or 5mC (*y*-axis). Three 8-mers are highlighted. (C) Same as panel A, but for ATF1-GST. (D) Same

as panel B, but for ATF1-GST. (E) Same as panels A and C, but for JUND-GST. The canonical AP-1 site TGAGTCAT bound by JUND is shown as a blue spot.

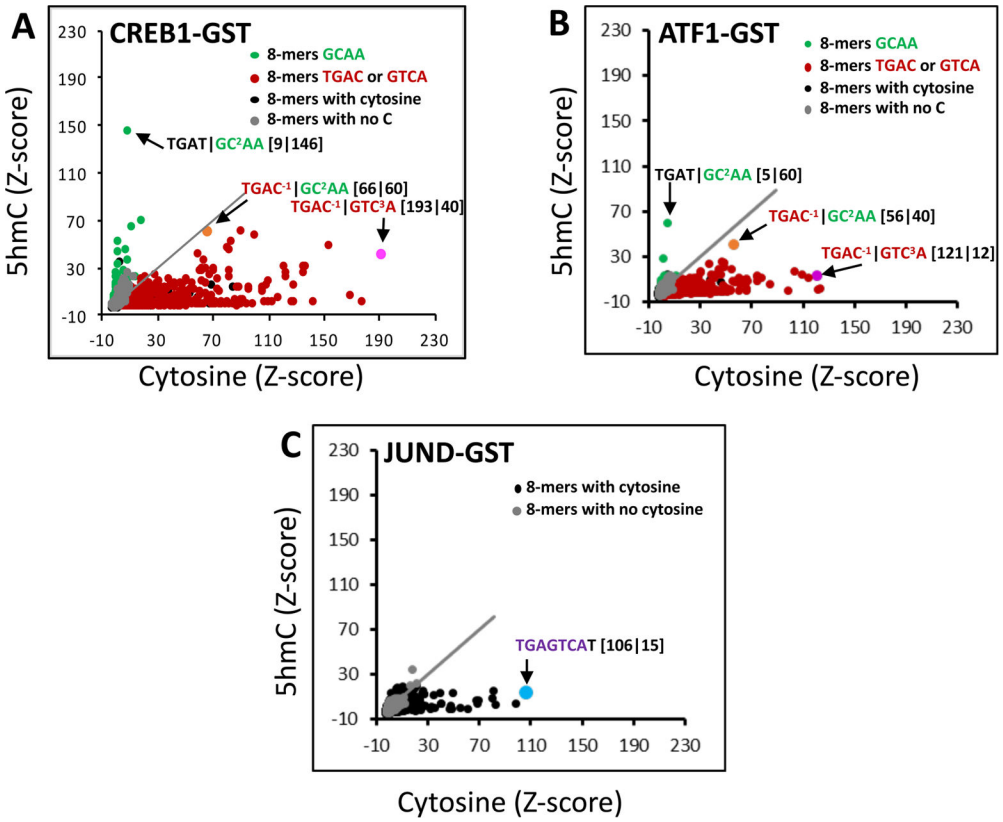


Figure 2.

CREB1, ATF1, and JUND binding to DNA 8-mers containing cytosine or 5hmC. (A) Scatter plot of *Z*-scores for CREB1-GST binding to 8-mers containing cytosine (*x*-axis) or 5hmC (*y*-axis). (B) Same as panel A, but for ATF1-GST. (C) Same as panels A and B, but for JUND-GST.

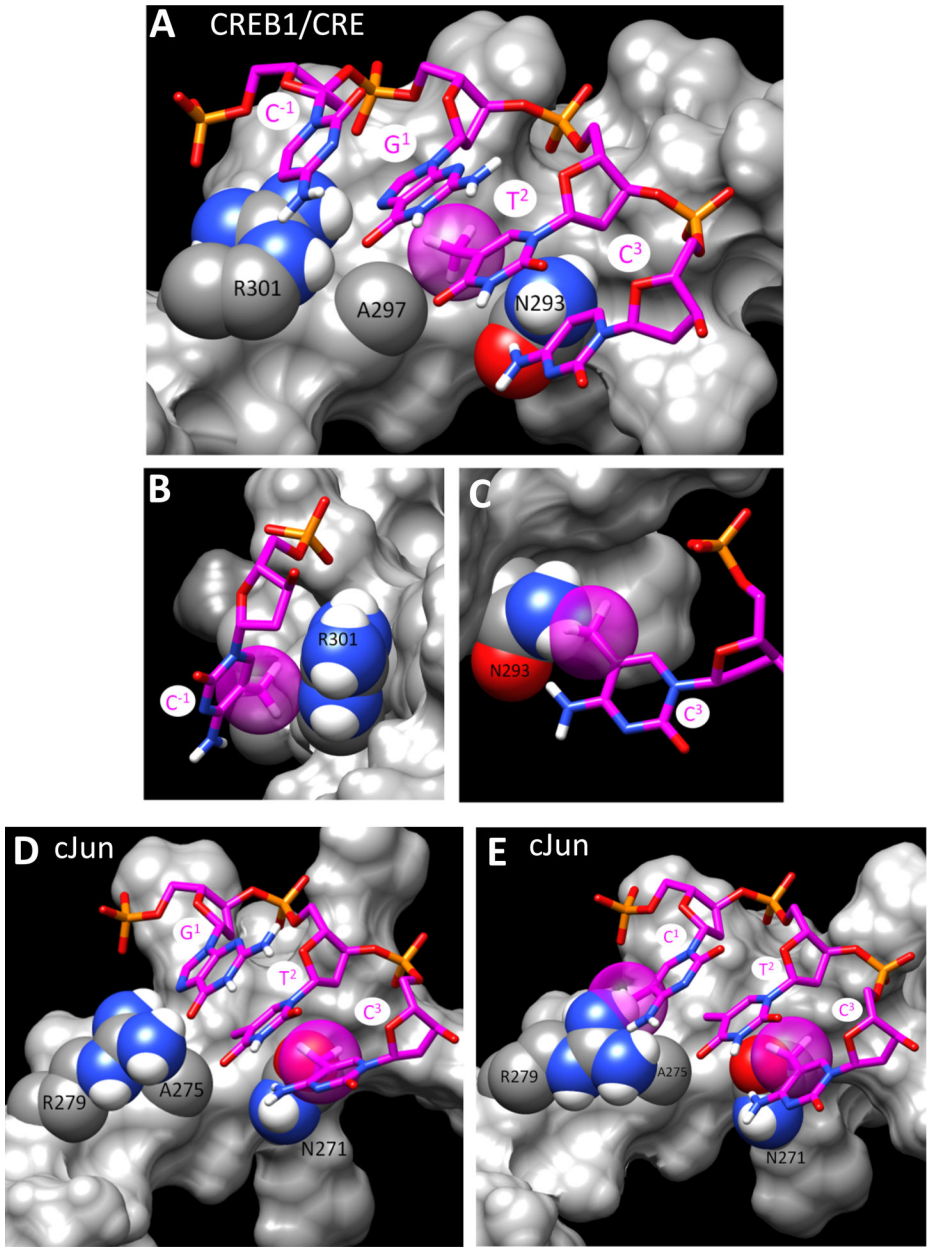


Figure 3.

Steric clashes explain the decrease in the level of binding of B-ZIP TFs to modified cytosines. (A) Crystal structure of CREB1 bound to four bases of the canonical CRE ($C^{-1}|G^1T^2C^3$). The protein is shown as a gray surface and highlights the nucleotide base and protein side chain interactions. Specific, interacting residues are shown as van der Waals spheres: gray for carbon, red for oxygen, blue for nitrogen, and white for hydrogen. The DNA is shown as a stick model, with the additional colors of magenta for the carbons and orange for the phosphorus atoms. The methyl group of the T^2 base is shown as a transparent sphere to highlight contact with the side chain of A297. (B and C) Locations of modeled-in methyl groups of the 5mC-modifed C^{-1} and C^3 nucleotides, respectively. The carbons of the added methyl groups are shown as transparent spheres. (D and E) Crystal structures of cJun

binding both strands of the AP-1 response element (|$G^1T^2C^3$ and $C^1T^2C^3$, respectively) with modeled-in 5mC methyl groups. The display format and color code are similar to those used in panel A.

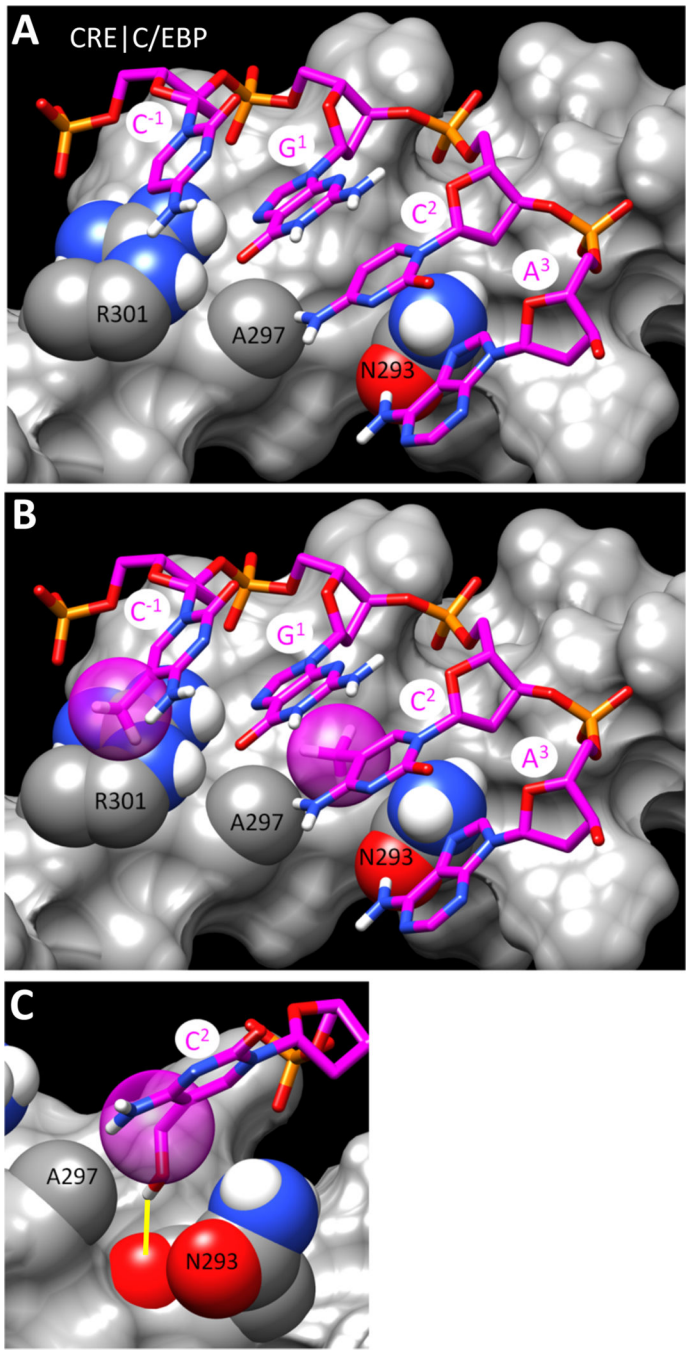


Figure 4.

Models of CREB1 bound to the C/EBP half-site C|GCAA The display format and color code are similar to those used in Figure 3: (A) complex with unmodified DNA, (B) same model with 5mC modifications, and (C) highlight of the hydroxyl group of 5hmC-modified C^2 forming a hydrogen bond with the carbonyl oxygen of N293 (yellow line).

CREB1: VVMASSPALPTQPAEE**AARKREVRLMKNREAARECRRKKKEYVKCLEN**RVAVLENQNKTLIEELKALKDLYCHKSD

ATF1: MTSPVTLTSQTTKTDD**PQLRREIRLMKNREAARECRRKKKEYVKCLEN**RVAVLENQNKTLIEELKTLKDLYSHKSV

JUND: SFGDSPPLSPIDMDTQ**ERIKAERKRLRNRIAASKCRKRKLERISRLEE**KVKTLKSQNTELASTASLLREQVAQLKQKVLSHVNSGCQLLP

Chart 1.