Published in final edited form as:

Biochemistry. 2017 November 28; 56(47): 6200–6210. doi:10.1021/acs.biochem.7b00741.

The Epstein-Barr virus B-ZIP protein Zta recognizes specific DNA sequences containing 5-methylcytosine and 5-hydroxymethylcytosine

Desiree Tillo¹, Sreejana Ray¹, Syed Khund-Sayeed¹, Mary Rose Gaylor¹, Ximiao He^{1,4}, Jun Wang¹, Nima Assad¹, Stewart R. Durell², Aleksey Porollo³, Matthew T. Weirauch³, and Charles Vinson^{1,*}

¹Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Room 3128, Building 37, Bethesda, MD 20892

²Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Room 3128, Building 37, Bethesda, MD 20892

³Center for Autoimmune Genomics and Etiology and Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center; Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229

Abstract

The Epstein-Barr virus (EBV) B-ZIP transcription factor (TF) Zta binds to many DNA sequences containing methylated CG dinucleotides. Using protein binding microarrays (PBMs), we analyzed the sequence specific DNA binding of Zta to four kinds of double-stranded DNA (dsDNA): 1) DNA containing cytosine in both strands, 2) DNA with 5-methylcytosine (5mC) in one strand and cytosine in the second strand, 3) DNA with 5-hydroxymethylcytosine (5hmC) in one strand and cytosine in the second strand, and 4) DNA where both cytosines in all CG dinucleotides contain 5mC. We compared these data to PBM data for three additional B-ZIP proteins (CREB1 and CEBPB homodimers, and cJun/cFos heterodimers). With cytosine, Zta binds the TRE motif **TGA^C/G^CTCA** as previously reported. With CG dinucleotides containing 5mC on both strands, many TRE motif variants containing a methylated CG dinucleotide at two positions in the motif, such as

*To whom correspondence should be addressed: Tel: 1-240-760-6885, Fax: 1-301-496-8419, vinsonc@mail.nih.gov.

⁴Present address: Department of Physiology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China.

Supporting Information

Supporting information contains: Summary of datasets in this study (Table S1), Slopes for normalization of Z-scores in Figures 2,3, and 4 (Table S2), Relationship between Z-score and Fold change in binding (Table S3), Effect of the central C/G on Zta binding meZRE2 and meTRE (Table S4), list of 8-mers bound by Zta and no other B-ZIP proteins. (Table S5), Number of array features above median signal for each B-ZIP protein (Table S7), comparison of B-ZIP PBM replicates (Figure S1), Raw array intensity distributions for all B-ZIP PBM replicates (Figure S2), a comparison of B-ZIP 5mCG and C raw intensities (Figure S3), B-ZIP TF binding to DNA 8-mers containing cytosine or 5mCG (Figure S4), Zta binding known ZREs (Figures S5–S6), B-ZIP TF binding to 8-mers containing cytosine or 5mC (Figure S7), B-ZIP TF binding to DNA 8-mers containing a single cytosine or 5mC (Figure S8), B-ZIP TF binding to 8-mers containing cytosine or 5hmC (Figure S9), B-ZIP TF binding to DNA 8-mers containing a single cytosine or 5hmC (Figure S10), Zta binding to DNA with either 5mC or 5mCG for 8-mers containing one cytosine in a CG dinucleotide at several positions along the 8-mer (Figure S11), 5mC as a substitute for thymine (T) at different positions of the 8-mer (Figure S12), Enrichment of Zta-only 8-mers in Zta ChIP-seq peaks in human (Figure S13), Multiple sequence alignment of a selection of B-ZIP DNA binding domains (Figure S14), and intensity distributions for B-ZIP TFs (Figure S15).

Occurrence of “Zta only” 8-mers in other B-ZIP ChIP seq data sets (Table S6) is provided as a separate Excel (.xlsx) file.

MGAGTCA and **TGAGMGA** (where **M**=5mC) were preferentially bound. 5mC inhibits Zta binding to both TRE motif half sites **GTCA** and **CTCA**. Like the CREB1 homodimer, the Zta homodimer and the cJun/cFos heterodimer bind the C/EBP half site tetranucleotide **GCAA** stronger when it contains 5mC. Zta also binds dsDNA sequences containing 5hmC in one strand, although the effect is less dramatic than observed for 5mC. Our results identify new DNA sequences that are well-bound by the viral B-ZIP protein Zta only when they contain 5mC or 5hmC, uncovering the potential for discovery of new viral and host regulatory programs controlled by EBV.

Introduction

Zta is a B-ZIP transcription factor (TF) encoded in the Epstein-Barr Virus (EBV) genome. This TF is a key regulator of the switch between latent and lytic cycles of the virus ¹⁻³. Zta is one of the first known examples of a TF that can bind DNA sequences containing methylated CG dinucleotides resulting in activation of gene transcription in the viral and host genomes ³⁻⁵.

Zta binds many methylated DNA sequences. There are 32 known variants of these sequences that have been identified *in vivo* (Zta response elements, ZREs) ⁶. Many of these sequences contain methylated CG dinucleotides, and are similar to the pseudopalindromic *12-O-tetradecanoylphorbol-13-acetate* or TRE motif, **TGA^C/G^CTCA** ⁷ also called the AP-1 motif ⁸ where ^C/_G represents either **C** or **G** (DNA sequences are shown in bold courier font). The methylated sequences that are well-bound replace one or both thymines in the TRE with a 5mC. These include the TRE containing a methylated CG dinucleotide at the beginning of the TRE (meTRE, **MGAGTCA**, where **M**=5mC) ⁹ and a methylated CG dinucleotide at a second position in the TRE (meZRE2, **TGAGMGA**) ¹⁰.

B-ZIP domains bind DNA as homo- or hetero-dimers, with each monomer interacting with one half of the DNA sequence motif ¹¹⁻¹³. A recent structural comparison between the cJun homodimer binding meTRE and Zta homodimers binding meZRE2 highlight several amino acids important for B-ZIP recognition of methylated DNA ¹⁴. In both cJun and Zta, consecutive amino acids in each monomer (Ala265 and Ala266 in cJun, Ala185 and Ser186 in Zta) interact with the methyl group from thymine or 5mC on opposite strands of the motif separated by a nucleotide. The Ser186 residue of Zta also confers the ability to bind the asymmetric sequence meZRE2 (**TGAGMGA**), in which the hydroxyl side chain of each serine present in each Zta monomer interacts with each half of the meZRE2 sequence differently ¹⁴.

Zta has been shown to bind to a sequence containing 5mC outside of CG dinucleotides, in the sequence **TGAGMAA** ¹⁴. This is similar to what was previously observed for the B-ZIP domain of CREB1, which preferentially binds DNA sequences containing the methylated C/EBP half site, **GMAA** ¹⁵. We previously hypothesized that an alanine residue (A297) of the CREB1 protein would form a stabilizing hydrophobic interaction with the methyl group of 5mC ¹⁵.

The full range of DNA sequences bound by Zta are unknown. In this study, we used protein binding microarrays (PBMs) ^{16, 17} to characterize the DNA binding of Zta to four types of

dsDNA. In addition to binding meZRE2 and meTRE, we identify many DNA sequences that are bound by the viral B-ZIP protein Zta only when they contain 5mC or 5hmC outside of CG dinucleotides. We highlight similarities and differences of Zta binding these four types of dsDNA with three additional B-ZIP dimers, CREB1 and CEBPB homodimers, and cJun|cFos heterodimers.

Experimental Procedures

Cloning and expression of B-ZIP DNA binding domains

The B-ZIP DNA binding domains (DBDs) of Zta, mouse CREB1, cJUN, and CEBPB were obtained from Dr. Timothy Hughes, University of Toronto, Canada. CREB1, cJUN, and CEBPB were obtained as a GST construct cloned into the pETGEXCT (C-terminal GST) vector ¹⁸. Zta was obtained as a GST construct cloned into a modified pDEST15 MAGIC vector (N-terminal GST) vector ¹⁸. The DBD of cFOS was obtained as a construct cloned into a pT5 expression plasmid ¹⁹. Proteins were expressed using the PURExpress *In vitro* protein synthesis kit (NEB) according to the manufacturer's protocol in a 25µL reaction volume containing 180 ng of plasmid ²⁰. The amino acid sequences of the B-ZIP domains with the alpha helical DNA binding region in bold are shown below: Zta:

STVQTAAAVVFACPGANQGQQLADIGVPQPAPVAAPARRTRKPQPESLEECDS

ELEIKRYKNRVASRKCRAKFKQLLQ

HYREVAAAKSSENDRLRLLLKQMCPSLDVDISIIPRTPDVLHEDLLNF CREB1:

VVMASSPALPTQPAEE **AARKREVRLMKNREAARECRRKKKEYVKLEN**

RVAVLENQNKTLIEELKALKDLYCHKSD CEBPB: PPAAPAKAKAKKTVDKL

SDEYKMRRERNNIIVRKS RDKAKMRNLET

QHKVLELTAENERLQKKVEQLSRELSTLRNLFKQLPELLASAGHC cJun:

MPGETPPLSPIDMESQER **KAERKMRNRNIAASKCRKRKLERIARL**

EEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQ cFos:

AQSIGRRGKVEQLSPEEEE **KRRIRRERNKMAAAKCRNRRRELTD**

TLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPDDLGFPE

PBM Experiments

The design of the Agilent 40K array and the PBM method has been described previously ^{16, 20–22}. Specifically, we used the “HK” array design available on the NCBI Gene Expression Omnibus platform GPL11260 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL11260>). Enzymatic methylation of CG dinucleotides on the HK microarray was performed as previously described ²⁰. Double-stranding of array probes with 5mC or 5hmC was also performed as described ¹⁷ using either 5-methylcytosine (5mC, NEB) or 5-hydroxymethylcytosine (5hmC, Zymo Research). Protein binding reactions were performed as described previously ^{15, 20}. Briefly, 180 ng of plasmid containing DNA binding domains were used to express proteins using PureExpress *in-vitro* transcription translation kit (NEB) in 25 µL reaction volume as per manufacturer's instructions. The double-stranded arrays were blocked with 4% milk for 1 hour and washed with 0.1% Tween-20 in 1X PBS. Freshly synthesized protein (25 µL) was mixed with 125 µL of protein binding reaction mixture consisting of 4% milk in 1X PBS, 50 ng of salmon testes DNA, and 0.2 µg/µL bovine serum albumin and added to double-stranded array. The protein binding reactions were carried out

in hydration chamber for 1 hour followed by one wash with 0.5% Tween-20 in 1X PBS. The protein bound arrays was incubated with Alexa Fluor 647 conjugated Anti-GST antibody for 1 hour, followed by three washes with 0.05 % Tween-20. Finally array slides were washed and dried in 1X PBS and scanned using Agilent Sure Scan II scanner.

Image quantification and calculation of 8-mer Z-scores

For each PBM, image quantification and calculation of Z-scores were performed as described previously¹⁷. 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: (1) 32,896 8-mers for the enzymatic CG methylation data and associated unmethylated controls, where complementary 8-mers were combined²⁰, and (2) all 65,536 8-mers for all other PBM experiments, as complementary 8-mers are different due to the asymmetric nature of the double stranding protocol for 5mC and 5hmC PBMs. For these 8-mers we compute the Z-scores of the reverse complement of the 8-mer extracted from the array probe design. Thus, all 8-mers shown and analyzed for these experiments are taken from the strand that can contain 5mC or 5hmC. We also note that the PBM design we used contains features in which all 8-mers occur 32 times on the array when double stranded. For Z-score calculations where complements are not identical (5mC and 5hmC experiments) certain 8-mers are rare. For these experiments, we only show Z-scores for 8-mers that appear on at least 10 features, eliminating 1,148 8-mers. All modified cytosine Z-scores (5mC, 5mCG, and 5hmC) were rescaled relative to unmodified cytosine using the slope of the line of best fit computed from the Z-scores using 8-mers with no cytosine as a control. The slope of the line for the 8-mers without a cytosine are presented in Table S2. Most TFs were assayed with at least two replicates, with good agreement ($R > 0.8$) (Figure S1) and little to no saturation of spots on the arrays (Figure S2). Arrays with no saturated spots were used for further analysis. For comparison to all published B-ZIP PBM data²³, we computed a separate score (E-score)²⁴, which represents the relative rank of intensities for all 8-mers. Data (raw probe intensities and 8-mer Z-scores) are available at the NCBI GEO database under accession GSE100871.

Analysis of Zta ChIP-seq peaks

We examined a publically available Zta ChIP-seq dataset performed in human Akata cells²⁵ for the presence of 8-mers we identified that are bound only by Zta *in vitro*. To this end, we downloaded the coordinates of the peaks from the original ChIP-seq study and converted the coordinates into their corresponding DNA sequences. We then calculated the frequency of occurrence across these sequences of each of the 382 8-mers bound only by Zta in our *in vitro* PBM assays. We next compiled a panel of ChIP-seq datasets for 24 different human TFs, largely taken from ENCODE²⁶. To avoid biasing the results, we only included one dataset/cell type per TF. Where possible, we used datasets performed in cell types similar to Akata B cells, such as GM12878 (which is also an EBV-infected B cell line). We determined the frequency of each 8-mer across each of these datasets as described above for Zta, and calculated the mean and standard deviation of these values. These were used to calculate enrichment scores (observed divided by expected 8-mer frequencies) and P-values (using a standard Z-score transformation).

Results

Protein binding microarrays containing 4 types of double-stranded DNA

We generated DNA microarrays containing 4 types of dsDNA. T7 DNA polymerase can efficiently incorporate 5mC and 5hmC when double-stranding single-stranded DNA²⁷. We exploited this property to double-strand single-stranded DNA 60-mers on an Agilent microarray using cytosine, 5mC or 5hmC, creating an asymmetric distribution of 5mC and 5hmC^{15, 17}. To create the 4th type of DNA, we enzymatically methylated the two cytosines in all CG dinucleotides (one cytosine in each strand)²⁰, producing DNA that Zta is known to bind¹⁰⁶. These microarrays were used for protein binding microarray (PBM) experiments in which a GST-tagged B-ZIP DNA binding domain binds to the DNA on the microarray slide. Binding is then detected using a fluor-conjugated antibody to the GST epitope followed by measurement of bound antibody (fluorescence intensities) at each of the 40,000 array features. Using this approach, we examined the sequence specific DNA binding of Zta, a B-ZIP protein encoded in the EBV genome, and compared it to published and new data for three other B-ZIP proteins: CEBPB and CREB1 homodimers and the cJun/cFos heterodimer (Table S1).

5mC in CG dinucleotides increases Zta binding to many DNA 8-mers

We initially examined Zta binding to two types of dsDNA: DNA containing cytosine in both strands and DNA where all CG dinucleotides contain 5mC in both strands. We examined both the fluorescent intensity of the 40,000 features on the array and a calculated Z-score representing the binding to individual 8-bp long DNA sequences, or 8-mers¹⁶. Z-scores are related to fold changes in binding (Table S3). When we examined binding to the 40,000 features, CG dinucleotide methylation did not change binding to the 3,861 features without a CG dinucleotide but did increase binding to many features that contain a CG dinucleotide, as expected (Figure S3). The fluorescent intensities do not show saturated probes showing these data are within the dynamic range of the assay (Figure S2).

We next calculated a Z-score reflecting the relative binding affinity to individual DNA 8-mers. Zta has previously been shown to bind the TRE motif 7-mer $T^{-4}G^{-3}A^{-2}(C/G)^0T^2C^3A^4$ (where $(C/G)^0$ indicates that the central base at position 0 can contain either a cytosine or guanine) and variants of the TRE that contain methylated CG dinucleotides, including the meTRE ($M^{-4}G^{-3}A^{-2}G^0T^2C^3A^4$) and meZRE2 ($T^{-4}G^{-3}A^{-2}G^0M^2G^3A^4$) motifs¹⁴. We use a numbering strategy for the nucleotide positions in the TRE motif that highlights the $T^2C^3A^4$ trinucleotide that is identical to the numbering scheme used for the cyclic AMP response element (CRE) motif $T^{-4}G^{-3}A^{-2}C^{-1}|G^1T^2C^3A^4$ ²⁸.

Figure 1 summarizes molecular interactions between Zta and three DNA sequences, the TRE, meTRE, and meZRE2. Individual Zta monomers (red and blue) recognize the $T^2C^3A^4$ trinucleotide present in both halves of the TRE motif (Figure 1A,B) with the two thymines at T^{-4} and T^2 on the opposite strand (highlighted) interacting with sequential Alanine (Ala185) and Serine (Ser186) residues (Figure 1B). For simplicity, we focus on individual half-sites and highlight their interactions with an individual Zta monomer. In the case of the meTRE, 5mC substitutes for T^{-4} , generating a methylated CG dinucleotide which contains 5mC at

position -4 and position 3 of the complement (Figure 1C). Here, the $\mathbf{m}^{-4}$ interacts with Ala185. For the meZRE2 motif, 5mC is found at position 2 (which interacts with Ser186) and position -3 in the complement (Figure 1D).

Figure 2A and Figure S4A is a comparison using a scatter plot of Zta Z-scores for all 32,896 DNA 8-mers containing cytosine on both strands (x-axis) and DNA where the cytosines on both strands in all CG dinucleotides have been enzymatically methylated to 5mC (y-axis). Here, DNA 8-mer complements are averaged. We divided 8-mers into three groups, 8-mers that: (1) do not contain a CG dinucleotide (gray spots), (2) contain one CG dinucleotide (blue spots), or (3) contain 2 or more CG dinucleotides (green spots). 8-mers lacking CG dinucleotides are along the diagonal (that is, Zta DNA binding to these 8-mers is unchanged with CG methylation) and act as an internal control, indicating that differences in binding to 8-mers with CG dinucleotides is due to methylation. The best bound 8-mers without a CG dinucleotide contain the canonical pseudopalindromic TRE (Figure 1A). Some non-cytosine containing 8-mers that are similar to the TRE motif are also well-bound by Zta (for example, **TGAGTAAT**) and will be examined later.

Many 8-mers are along the vertical axis indicating they are only strongly-bound when they contain a methylated CG dinucleotide. In contrast, there are no well-bound 8-mers on the horizontal axis, indicating that CG methylation does not inhibit strong Zta binding. 8-mers where CG dinucleotide methylation increases Zta binding contain meZRE2 (e.g. **TGAGC²GAT**) and meTRE (**C⁻⁴GAGTCAT**) (Figure 2A, S4A, and Table S4). We note that the 8-mers containing meTRE are not as well-bound by Zta in our PBMs as the TRE or meZRE2 sequences. This contrasts with what was found in a recent study in which the TRE, meTRE, and meZRE2 motifs are equally strongly bound by Zta ¹⁴. We will address these differences in the discussion.

The center of TRE motif can contain either cytosine or guanine, $\mathbf{T}^{-4}\mathbf{G}^{-3}\mathbf{A}^{-2}(\mathbf{C}/\mathbf{G})^0\mathbf{T}^2\mathbf{C}^3\mathbf{A}^4$ (Figure 1B). To explore the effects of the central position on Zta binding to methylated motifs, we examined Z-scores for 8-mers containing methylated CG dinucleotides at two positions in the meTRE motif (**MGACTCAT** or **MGAGTCAT**) and the meZRE2 (**TGACMGAT** or **TGAGMGAT**) with either cytosine or guanine at the center of the motif (Table S4). The meTRE containing 8-mers **MGACTCAT** or **MGAGTCAT**, where the central cytosine or guanine is one nucleotide away from the methylated CG dinucleotide, show a similar increase in Zta binding with methylation. In contrast, the meZRE2 containing 8-mers **TGACMGAT** or **TGAGMGAT** show very different increases in DNA binding with methylation. When the CG dinucleotide is adjacent to a guanine, the increase in Zta binding is much greater with CG methylation (Z-score increases from 15 to 326 for **TGAGCGAT**) than when the CG dinucleotide is adjacent to a cytosine (the Z-score increases 4 to 60 for **TGACCGAT**), indicating that nucleotides outside of the methylated CG dinucleotide influence Zta binding. This is similar to results shown in a previous study ¹⁴. A motif logo generated from the top 50 5mCG containing 8-mers bound by Zta also highlights the preferential binding of Zta to the meZRE2 containing a central guanine (Figure 2A).

8-mers with two CG dinucleotides such as **C⁻⁴GAGC²GAT** and **AC⁻⁴GTGC²GA** (highlighted in Figure 2A, S4A), are also bound by Zta when methylated. These 8-mers, however, have

lower Z-scores than those containing the meZRE2 motif, as expected if methylation of each CG dinucleotide affected binding independently. We will later examine the contribution of each of the two 5mCs in both palindromic CG dinucleotides to the increase in Zta binding.

Effect of 5mC in CG dinucleotides on DNA binding for cJun|cFos, CREB1 and CEBPB

We next examined three additional B-ZIP domains, the cJun|cFos heterodimer (data from ¹⁹), CREB1 (data from this study), and CEBPB (data from ²⁰) binding to DNA 8-mers containing methylated CG dinucleotides (Figure 2B–D, S4B–D). For cJun|cFos heterodimers, the best bound 8-mers contain the TRE motif, as expected (Figure 2B, highlighted gray spots). The only 8-mers in which CG methylation enhances binding of cJun|cFos are the methylated TRE, **m⁻⁴GAGTCA** (Figure 2B, blue spots) ^{9, 14}. Unlike Zta, the meZRE2 motif is poorly bound by cJun|cFos (Figure 2B, S4B, **magenta spots**) ¹⁴. Binding of the cJun|cFos heterodimer to the canonical CRE motif **TGACGTCA** containing a methylated central CG is reduced, as is observed for CREB1 ³¹. For CREB1, CG dinucleotide methylation inhibits DNA binding to many 8-mers (Figure 2C, S4B). For CEBPB, CG methylation modestly enhances DNA binding for several 8-mers, including the C/EBP motif, **TTGC|GCAA**, containing a central CG dinucleotide (Figure 2D, S4D) ²⁰. Overall, CG dinucleotide methylation does not enhance DNA binding of cJun|cFos, CREB1, or CEBPB as dramatically as Zta (Figure 2B,C,D, S4B–D).

Additional methylated CG dinucleotide containing 8-mers well bound by Zta (ZRE types)

Many CG dinucleotide containing 8-mers are bound by Zta when they are methylated. We examined the three classes of Zta DNA binding sites (ZREs) defined previously ⁶ (Figure S5 and Figure S6). Class I binding sites have no CG dinucleotides and include TRE-like motifs (**TGAGTCAT**, **TGAGTAAT**) that are well-bound on our PBMs. However, many 8-mers previously identified as Class I ZREs are poorly bound, such as those containing the 7-mers **TGTGTCT**, **GTTGCAA**, and **TGGCACA** (Figure S5). Class II ZREs contain one CG dinucleotide and tend to be bound by Zta only when methylated. Our PBM experiments indicate that this class of ZREs separate into two groups with 8-mers containing **TGAGMGA** being better bound than those containing **TGAGMGC**. Class III ZREs are only bound when methylated and display a wide range of Z-scores. We note that our data do not differentiate between Class II and III ZREs, both of which are only bound when both cytosines in CG dinucleotides are methylated.

5mC in one DNA strand enhances Zta binding to 8-mers containing the C/EBP half site tetramer GM²AA

To determine the effect of individual methylated cytosines on Zta binding to DNA, we double-stranded the single-stranded DNA on the PBMs with 5mC ¹⁷. For these analyses, we compute the Z-scores for all 65,536 8-mers, as complements are structurally different, unlike the CG dinucleotide methylation data where complements are identical. Figure 3A and S7A show Zta binding to all 8-mers that contain cytosine in both DNA strands (x-axis) or 5mC in one strand and cytosine in the second strand (y-axis). There are many 8-mers that are only well-bound with 5mC, fewer 8-mers only well-bound with cytosine, and even fewer well-bound with either cytosine or 5mC. Two 8-mers that do not contain a cytosine and are well-

bound by Zta (**ATGAGTAA** and **TGAGTAAT**) are along the diagonal and function as an internal control.

We examined three motif half sites: The C/EBP half site with one cytosine (**GC²AA**), and two TRE half sites, one with one cytosine (**GTC³A**) and one with two cytosines (**C⁰TC³A**). 8-mers with the C/EBP half site **GC²AA** are well-bound only when they contain 5mC, which is also true for CREB1, another B-ZIP domain TF ¹⁵ (Figure 3C, S7C). Methylation of the cytosine in the TRE half site **GTC³A** results in poorer binding of Zta. Methylation of both cytosines in **C⁰TC³A** results in even worse binding, suggesting that methylation of C⁰ also inhibits DNA binding.

We next examined the 17,496 8-mers with only one cytosine to evaluate the contribution of individual 5mCs at different positions in the motif on the DNA binding of Zta (Figure S8A). These results recapitulate the conclusions from the previous paragraph. A few 8-mers are well-bound with either cytosine or 5mC, such as **TGAGC²AAT** and **ATGAGC²AA**, and both contain the TRE half-site **TGAG**, which is well-bound when it is unmethylated, and the C/EBP half site **GC²AA**, which is well-bound when it is methylated.

The cJun|cFos heterodimer also binds the methylated DNA tetramer **GMAA**

Previously, we showed that the B-ZIP domain of CREB1 preferentially bound the C/EBP half site tetramer **GMAA** when it contained 5mC ¹⁵ (Figure 3C, Figure S7C, S8C). We generated PBM binding data for cJun|cFos heterodimers and CEBPB homodimers to DNA sequences double stranded with 5mC. cJun|cFos heterodimers (Figure 3B, S7B, S8B) also bind 8-mers with methylated C/EBP half sites. In contrast, CEBPB does not preferentially bind **GMAA**, except for several 8-mers that contain multiple 5mCs (**MTAMGMAA** and **MTGMGMAA**, Figure 3D, S8B).

Four B-ZIPs binding to dsDNA containing 5hmC in one strand

We next examined Z-scores for Zta, CREB1 (data from ¹⁵), cJun|cFos (data from this study), and CEBPB (data from this study) binding to DNA 8-mers with 5hmC in one strand and cytosine in the second strand (Figure 4, S9, and S10). Zta binds many 8-mers with **GHAA** (where **H**=5hmC, Figure 4A, S9A), although the effect of 5hmC on binding is not as dramatic as that observed for 5mC. Similar results were observed for Zta binding 5hmC containing TRE, ZRE2, and **TGAGCAA** sequences ¹⁴. For the other TFs, CREB1 binds a few 8-mers with **GHAA** (Figure 4C, S9C) while cJun|cFos and CEBPB do not preferentially bind 8-mers with **GHAA** (Figure 4B,D, S9B,D).

Zta binding to single nucleotide polymorphisms (SNPs) and 5mC and 5hmC throughout the TRE motif

We next evaluated the effect of SNPs and 5mC and 5hmC throughout the TRE motif (Table 1) to further examine Zta DNA binding specificity. We started with the best-bound 8-mer that does not contain a cytosine: **T⁻⁴G⁻³A⁻²G⁰T²A³A⁴** (Z-score=162). In this 8-mer, the C³ from the TRE is replaced with A³. At position -4, thymine is best for Zta binding. Other nucleotides dramatically inhibiting binding with 5mC, 5hmC, and 5mCG have intermediate

binding. At position -3, guanine is the only well-bound nucleotide. At position -2, adenine is best for binding but thymine is also possible. At position 0, guanine is preferred but cytosine can be accommodated. At position 2, thymine and cytosine are well-bound, with both 5mC and 5hmC being best for Zta binding. At position 3, cytosine and adenine can be accommodated, with 5mC and 5hmC reducing binding. At position 4, only adenine is well-bound.

Contribution of each 5mC in a methylated CG dinucleotide to preferential Zta binding

Methylated CG dinucleotides in two positions of the TRE enhance DNA binding of Zta. One is the methylated TRE (meTRE) sequence $\text{M}^{-4}\text{G}^{-3}\text{A}^{-2}\text{G}^0\text{T}^2\text{C}^3\text{A}^4$ and its complement $\text{T}^{-4}\text{G}^{-3}\text{A}^{-2}\text{C}^0\text{T}^2\text{M}^3\text{G}^4$ ^{9, 14}(Figure 1). To evaluate the contribution of each 5mC in the CG dinucleotide to the increase in ZTA binding, we revisited Table 1. Enzymatic methylation of the CG dinucleotide in $\text{C}^{-4}\text{G}^{-3}\text{A}^{-2}\text{G}^0\text{T}^2\text{G}^3\text{A}^4$ produces 5mC at position -4 in one DNA strand and position 3 on the opposite strand (Figure 1). A single 5mC in position -4 stabilizes Zta binding (Z-score=82), while 5mC at position 3 is destabilizing (Z-score=7). Methylation of both cytosines in the $\text{C}^{-4}\text{G}^{-3}$ dinucleotide results in intermediate Zta binding (Z-score=43), suggesting the two 5mCs in the methylated CG dinucleotide are acting independently. These results are similar to cJun homodimers binding both versions of the hemimethylated TRE ¹⁴.

The most dramatic increase in Zta binding with CG dinucleotide methylation is for the sequence $\text{T}^{-4}\text{G}^{-3}\text{A}^{-2}\text{G}^0\text{M}^2\text{G}^3\text{A}^4$ (meZRE2), whose complement is $\text{T}^{-4}\text{M}^{-3}\text{G}^{-2}\text{C}^0\text{T}^2\text{C}^3\text{A}^4$ (Figure 1). To evaluate the contribution of each cytosine to preferential Zta binding, we started with a non-cytosine containing version of the ZRE2 sequence, $\text{T}^{-4}\text{G}^{-3}\text{A}^{-2}\text{G}^0\text{T}^2\text{G}^3\text{A}^4$ (Table 2). 5mC at position 2 enhances Zta binding (Z-score=115) compared to cytosine (Z-score=17). Examination of the -3 position finds weak binding for everything except guanine. Methylation of both cytosines in the C^2G^3 dinucleotide results in even better binding (Z-score=284) suggesting that the other 5mC in the CG dinucleotide is contributing to binding.

To evaluate additional 8-mers with a CG dinucleotide, we examined the 5,096 8-mers that contain one cytosine that is in a CG dinucleotide (Figure 5). We plot the Z-scores for Zta binding 8-mers with 5mC in one strand (x-axis) or enzymatic methylation of both cytosines in the CG dinucleotide (y-axis). These experiments were done separately but each has a paired experiment for Zta binding to DNA containing only cytosine in both strands (Figure 5, grey spots). The Z-scores for the two PBMs double stranded with cytosine are along the diagonal, indicating that the Z-scores are concordant and comparable across experiments. Some 8-mers have similar binding to either 5mC or CG dinucleotide methylation (Figure 5 and S11). These 8-mers contain the CG dinucleotide at the beginning of the sequence ($\text{M}^{-4}\text{GAGTAAT}$, $\text{AM}^{-4}\text{GAGTAA}$, and $\text{GM}^{-4}\text{GAGTAA}$). Many 8-mers are better bound when both cytosines in the CG dinucleotide contain 5mC suggesting both cytosines contribute to the increase in binding. Most of these sequences are of the form DDDGMGA (where $\text{D} = \text{A, G, or T}$), and contain a CG dinucleotide at positions C^2G^3 including the 8-mers $\text{TGAGM}^2\text{G}^3\text{AT}$ and $\text{ATGAGM}^2\text{G}^3\text{A}$ (Figure 5 and S11).

5mC is different from thymine (T)

Thymine has some structural similarity to 5mC — both contain a methyl group at carbon 5 of the pyrimidine ring. In our analysis of the meTRE (Table 1) and meZRE motifs (Table 2), 5mC and T have different effects on Zta binding depending on their position within the motif. Thymine and 5mC are both well-bound at positions −4 and 2 (Table 1). To examine how similar thymine and 5mC are for sequence-specific Zta binding to all 8-mers, we examined the 6,298 8-mers containing at least one thymine (**T**) but no cytosine (**C**) and changed each T to 5mC (**M**) producing 17,462 possible **T** to **M** substitutions (Figure 6). Of these 17,462 substitutions, 5,090 generate an **MG** dinucleotide (blue spots). Some of these **MG** dinucleotide containing 8-mers are bound by Zta either when they contain **T** or **M** at the beginning of the 8-mer (i.e. they are on or close to the diagonal), suggesting that thymine is similar to 5mC. (Figure S12A–B). Some 8-mers containing methylated CG dinucleotides are better bound than TG dinucleotides. These are meZRE2-containing 8-mers in which **TGAGM²G³A** is better bound than **TGAGT²G³A** (Figure 6), recapitulating the results in Tables 1 and 2. Many more 8-mers are better bound with 5mC rather than **T**. These 8-mers contain 5mC outside of CG dinucleotides, and these contain the methylated C/EBP half-site, **GMAA** (Figure 6, **green spots**).

Identification of 8-mers bound by Zta and no other B-ZIP proteins

Our PBM data highlight Zta as a promiscuous TF compared to the other three B-ZIPs we examined. To enumerate all the 8-mers bound by Zta that are not bound by other B-ZIP TFs, we compiled all PBM datasets available for B-ZIPs in any species ²³ encompassing 300 experiments covering 133 TFs in 16 different species. For each experiment, we identified all strongly bound 8-mers (PBM E-score ≥ 0.45), resulting in 4,247 8-mers bound by any of the 133 B-ZIP TFs. Application of the same procedure to our PBM data yielded 539 unique 8-mers bound strongly by Zta in any of the conditions we tested. Of these, 382 are unique to Zta (Table S5), many of which are non-palindromic and are bound only when they contain 5mC. Notably, these same 8-mers are also among the most frequently occurring 8-mers in a publically available Zta ChIP-seq dataset ²⁵(Figure S13), with 8-mers identified from our PBM experiments containing 5mC being significantly associated with high frequency in Zta ChIP-seq peaks *in vivo* (Wilcoxon rank sum $P=4e^{-17}$ for all “Zta only” 8-mers, $P<0.004$ for “Zta only” 8-mers containing 5mC Figure S13B,C,D,F). Further, 21 of the “Zta only” 8-mers occur significantly more often ($P < 0.01$) in the Zta ChIP-seq dataset than in a panel of 24 human B-ZIP ChIP-seq datasets (Table S6, and Methods). For example, the **ATGCACAC** 8-mer, which is strongly bound by Zta on our 5mC and 5hmC arrays, is present in 28% of the ChIP-seq peaks in the Zta Akata cell ChIP-seq dataset, which is 7.2-fold more often than the average frequency observed for this 8-mer across the 24 human B-ZIP ChIP-seq datasets ($P<10^{-4}$). Collectively, these results indicate the 8-mers we identify that are bound only by Zta *in vitro* are also likely bound exclusively by Zta *in vivo*, highlighting the unique promiscuous nature of Zta relative to other known B-ZIP proteins.

Discussion

We have examined the sequence specificity of Zta binding to four kinds of dsDNA, including dsDNA containing 5mC or 5hmC in one strand, and compared these data to

similar data for three other B-ZIP domains. 5mC and 5hmC in one DNA strand mimics what occurs biologically in several cell types including brain ³². Many DNA 8-mers are better bound by Zta when both cytosines in a CG dinucleotide contain 5mC. This contrasts with other B-ZIP TFs we and others ^{9, 20, 33, 34} have examined, in which CG dinucleotide methylation either: (1) enhances DNA binding to only a handful of sequences, as observed for cJun/cFos binding the meTRE (**MGAGTCA**) or for CEBPB binding the methylated C/EBP motif (**TTGMGCAA**), or (2) dramatically reduces binding to many 8-mers, as observed for CREB1. Like the CREB1 homodimer ¹⁵, the Zta homodimer and the cJun/cFos heterodimer bind the methylated C/EBP half site tetranucleotide **GMAA**. The effect of 5hmC on Zta binding is less dramatic than that observed for 5mC, with several 8-mers containing the hydroxymethylated **GHA**A being well-bound, similar to that observed for CREB1 ¹⁵. Our analysis also highlights two positions in the TRE motif (**T**⁻⁴ and **T**²) that increase Zta binding when they contain 5mC. Comparison of Zta binding to DNA containing 5mC and enzymatically methylated CG dinucleotides suggest both 5mCs in a methylated CG dinucleotide in position **M**²**G**³ contribute to stronger binding. Finally, we have identified many DNA sequences that are only bound by Zta when they contain 5mC or 5hmC that are likely bound by Zta *in vivo*, expanding the DNA binding repertoire of Zta which may have implications for the discovery of new viral and host regulatory programs controlled by EBV.

Several aspects of our PBM data are similar to a recent study that used a fluorescence polarization assay to show that Zta preferentially binds the methylated versions of the meTRE (**MGAGTCA**), meZRE2 (**TGAGMGA**) ¹⁴, and the sequence **TGAGMAA**. Our study expands on this work, identifying many additional 8-mers that are preferentially bound by Zta when methylated. We also find that Zta does not bind meTRE as well as the unmethylated TRE or meZRE2, whereas in the aforementioned study, there are no differences in Zta binding to these sequences (they are all equally strongly bound by Zta), which may be attributed to differences in the Zta protein used. The Hong et al. study uses a Zta protein where a cysteine is replaced with serine at position 189, whereas our PBM experiments utilizes a wild type Zta DNA binding domain. Residues in this position contact a phosphate group of the DNA backbone. While Hong et al. note that this substitution does not affect Zta binding to the meZRE or meTRE, others suggest that this mutation affects Zta binding and function *in vivo* and *in vitro* ¹⁰. It would be interesting to determine whether a Zta (C189S) mutant affects DNA binding specificity on PBMs.

5mC increases DNA binding of Zta and cJun/cFos to two sets of 8-mers: 8-mers containing the meTRE and the methylated C/EBP half site tetramer **GM**²**AA**. Our previous work has shown that the ATF4 basic region as a heterodimer with CEBPB binds the methylated CRE/CEBP chimeric 8-mer **MGAT**|**GCAA** ²⁰, which contains a methylated CG dinucleotide at the beginning of the motif, similar to the meTRE. For this sequence, methylated cytosine of the opposite strand does not contribute to DNA binding, similar to Zta and cJun homodimers and cJun/cFos heterodimers binding the meTRE ^{9, 14}.

Our previous structural analysis highlighted an alanine residue (A297) of the CREB1 protein forming stabilizing hydrophobic interactions with the methyl group of 5mC at **C**² that facilitates binding to the C/EBP half site **GM**²**AA** ¹⁵, a trait shared with the cJun (A270)|

cFos(A152) heterodimer. The corresponding residue is a serine (S186) in Zta that also forms stabilizing hydrophobic interactions with $\mathbf{m}^2$ (Figure 1 and Figure S14) ¹⁴. Interestingly, CEBPB, which does not bind the methylated C/EBP half site as well as the other B-ZIP TFs examined in this study, contains a valine (V285) at this position.

These data highlight the versatility of the B-ZIP domain in binding unmethylated and methylated DNA sequences. For example, Zta binds many more 5mC and 5hmC containing sequences compared to the other B-ZIP TFs examined (Table S5, Table S7, and Figure S15). In fact, 5mC increases the promiscuity of sequence-specific DNA binding on PBMs (Table S7). It has been hypothesized that Zta's ability to bind many DNA sequences may be due in part to increased and closer contacts with the DNA phosphate backbone ²⁹. Detailed mutagenesis experiments may help to uncover the amino acid determinants underlying the differences in specificity and binding activities observed in these four domains for methylated and unmethylated DNA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work is supported by the intramural research project of the National Cancer Institute, National Institutes of Health (NIH), Bethesda, USA. MTW was supported by NIH R21 HG008186, NIH R01 NS099068-01A1, Lupus Research Alliance "Novel Approaches", and Cincinnati Children's Hospital "Center for Pediatric Genomics" and "Trustee Award" grants. We thank Xiaoting Chen for computational support.

Abbreviations

5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
C/EBP	CCAAT-enhancer-binding protein
CRE	Cyclic AMP Response Element
CREB	Cyclic AMP Response Element Binding
dsDNA	double-stranded DNA
EBV	Epstein-Barr virus
PBM	protein binding microarray
TF	transcription factor
TRE	<i>12-O-tetradecanoylphorbol-13-acetate</i>
ZRE	Zta response element

References

1. Lieberman PM, Berk AJ. In vitro transcriptional activation, dimerization, and DNA-binding specificity of the Epstein-Barr virus Zta protein. *J Virol.* 1990; 64:2560–2568. [PubMed: 2159531]
2. Kenney, SC. Reactivation and lytic replication of EBV. In: Arvin, A.Campadelli-Fiume, G.Mocarski, E.Moore, PS.Roizman, B.Whitley, R., Yamanishi, K., editors. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Cambridge; 2007.
3. Ramasubramanyan S, Kanhere A, Osborn K, Flower K, Jenner RG, Sinclair AJ. Genome-wide analyses of Zta binding to the Epstein-Barr virus genome reveals interactions in both early and late lytic cycles and an epigenetic switch leading to an altered binding profile. *J Virol.* 2012; 86:12494–12502. [PubMed: 23015699]
4. Bhende PM, Seaman WT, Delecluse HJ, Kenney SC. The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat Genet.* 2004; 36:1099–1104. [PubMed: 15361873]
5. Tsai SC, Lin SJ, Chen PW, Luo WY, Yeh TH, Wang HW, Chen CJ, Tsai CH. EBV Zta protein induces the expression of interleukin-13, promoting the proliferation of EBV-infected B cells and lymphoblastoid cell lines. *Blood.* 2009; 114:109–118. [PubMed: 19417211]
6. Flower K, Thomas D, Heather J, Ramasubramanyan S, Jones S, Sinclair AJ. Epigenetic control of viral life-cycle by a DNA-methylation dependent transcription factor. *PLoS One.* 2011; 6:e25922. [PubMed: 22022468]
7. Angel P, Imagawa M, Chiu R, Stein B, Imbra RJ, Rahmsdorf HJ, Jonat C, Herrlich P, Karin M. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell.* 1987; 49:729–739. [PubMed: 3034432]
8. Lee W, Mitchell P, Tjian R. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell.* 1987; 49:741–752. [PubMed: 3034433]
9. Gustems M, Woellmer A, Rothbauer U, Eck SH, Wieland T, Lutter D, Hammerschmidt W. c-Jun/c-Fos heterodimers regulate cellular genes via a newly identified class of methylated DNA sequence motifs. *Nucleic Acids Res.* 2014; 42:3059–3072. [PubMed: 24371273]
10. Karlsson QH, Schelcher C, Verrall E, Petosa C, Sinclair AJ. 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.* 2008; 4:e1000005. [PubMed: 18369464]
11. Vinson CR, Sigler PB, McKnight SL. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science.* 1989; 246:911–916. [PubMed: 2683088]
12. Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, Bonovich M. Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol.* 2002; 22:6321–6335. [PubMed: 12192032]
13. Rodriguez-Martinez JA, Reinke AW, Bhimsaria D, Keating AE, Ansari AZ. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *Elife.* 2017; 6:e19272. [PubMed: 28186491]
14. Hong S, Wang D, Horton JR, Zhang X, Speck SH, Blumenthal RM, Cheng X. Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein-Barr virus Zta. *Nucleic Acids Res.* 2017; 45:2503–2515. [PubMed: 28158710]
15. Syed KS, He X, Tillo D, Wang J, Durell SR, Vinson C. 5-Methylcytosine (5mC) and 5-Hydroxymethylcytosine (5hmC) Enhance the DNA Binding of CREB1 to the C/EBP Half-Site Tetranucleotide GCAA. *Biochemistry.* 2016; 55:6940–6948. [PubMed: 27951657]
16. Berger MF, Bulyk ML. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc.* 2009; 4:393–411. [PubMed: 19265799]
17. Khund-Sayeed S, He X, Holzberg T, Wang J, Rajagopal D, Upadhyay S, Durell SR, Mukherjee S, Weirauch MT, Rose R, Vinson C. 5-Hydroxymethylcytosine in E-box motifs ACAT|GTG and ACAC|GTG increases DNA-binding of the B-HLH transcription factor TCF4. *Integr Biol (Camb).* 2016; 8:936–945. [PubMed: 27485769]
18. Sharrocks AD. A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase. *Gene.* 1994; 138:105–108. [PubMed: 8125285]

19. He X, Tillo D, Vierstra J, Syed KS, Deng C, Ray GJ, Stamatoyannopoulos J, FitzGerald PC, Vinson C. Methylated Cytosines Mutate to Transcription Factor Binding Sites that Drive Tetrapod Evolution. *Genome Biol Evol.* 2015; 7:3155–3169. [PubMed: 26507798]
20. Mann IK, Chatterjee R, Zhao J, He X, Weirauch MT, Hughes TR, Vinson C. CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Res.* 2013; 23:988–997. [PubMed: 23590861]
21. 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, Bulyk ML. Diversity and complexity in DNA recognition by transcription factors. *Science.* 2009; 324:1720–1723. [PubMed: 19443739]
22. Lam KN, van Bakel H, Cote AG, van der Ven A, Hughes TR. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res.* 2011; 39:4680–4690. [PubMed: 21321018]
23. Weirauch MT, Yang A, Albu M, Cote A, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano JC, Galli M, Lewsey M, Huang E, Mukherjee T, Chen X, Reece-Hoyes JS, Govindarajan S, Shauly G, Walhout AJM, Bouget FY, Ratsch G, Larrondo LF, Ecker JR, Hughes TR. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell.* 2014; 158:1431–1443. [PubMed: 25215497]
24. Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW 3rd, Bulyk ML. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol.* 2006; 24:1429–1435. [PubMed: 16998473]
25. Ramasubramanyan S, Osborn K, Al-Mohammad R, Naranjo Perez-Fernandez IB, Zuo J, Balan N, Godfrey A, Patel H, Peters G, Rowe M, Jenner RG, Sinclair AJ. Epstein-Barr virus transcription factor Zta acts through distal regulatory elements to directly control cellular gene expression. *Nucleic Acids Res.* 2015; 43:3563–3577. [PubMed: 25779048]
26. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Frietze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, Lacroute P, Leng JJ, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglu S, Sidow A, Farnham PJ, Myers RM, Weissman SM, Snyder M. Architecture of the human regulatory network derived from ENCODE data. *Nature.* 2012; 489:91–100. [PubMed: 22955619]
27. Chen CC, Wang KY, Shen CK. The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem.* 2012; 287:33116–33121. [PubMed: 22898819]
28. Schumacher MA, Goodman RH, Brennan RG. The structure of a CREB bZIP: somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J Biol Chem.* 2000; 275:35242–35247. [PubMed: 10952992]
29. Petosa C, Morand P, Baudin F, Moulin M, Artero JB, Muller CW. Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. *Mol Cell.* 2006; 21:565–572. [PubMed: 16483937]
30. Sagendorf JM, Berman HM, Rohs R. DNAProDB: an interactive tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.* 2017; 45:W89–W97. [PubMed: 28431131]
31. Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, Glass K, Fitzgerald P, Vinson C. CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc Natl Acad Sci U S A.* 2010; 107:20311–20316. [PubMed: 21059933]
32. 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, Ecker JR. Global epigenomic reconfiguration during mammalian brain development. *Science.* 2013; 341:1237905. [PubMed: 23828890]
33. Kribelbauer JF, Laptenko O, Chen S, Martini GD, Freed-Pastor WA, Prives C, Mann RS, Bussemaker HJ. Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. *Cell Rep.* 2017; 19:2383–2395. [PubMed: 28614722]

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

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Relationships between TRE, meTRE, and meZRE2 sequences and their interaction with Zta monomers

(A) Crystal structure (PDB: 2c9l²⁹) of the Zta homodimer binding the TRE sequence. Each monomer (red, blue) binds a half site. One TRE half site (**TGAG**) is colored in green. (B) Nucleotide-residue interaction map³⁰ generated using PDB structure 2c9l²⁹ of Zta binding the TRE sequence. Amino acids from each monomer and specific nucleotides are colored as in A. Interactions between the thymine in the TRE half site (highlighted) with the sequential Alanine (Ala185) and Serine (Ser186) interactions are indicated with bold lines. Nucleotide-residue interaction maps for the (C) meTRE and (D) meZRE2 sequences, where **m**=5mC. For (D), the interaction map was generated using the PDB structure 5szx¹⁴ binding the meZRE2 sequence.

Figure 2. 5mCG increases Zta DNA binding to many 8-mers

(A) Z-scores for Zta binding of CG-dinucleotide-containing-8-mers with just cytosine (x-axis), or with 5mCG (y-axis). The 20,349 8-mers without a CG dinucleotide are used as an internal control, and are indicated in grey. Blue spots represent the 8-mers containing one CG dinucleotide ($\# = 10,762$), and green dots ($\# = 1,785$) represent two or more CG dinucleotides, respectively. Several 8-mers are highlighted and described in the text, including two 8-mers containing the TRE sequence in magenta. Many 8-mers containing CG dinucleotides are more strongly bound upon methylation. Sequence logos generated using the top 50 bound 8-mers containing cytosine or 5mCG are shown. (B) Same as in A but for cJun|cFos heterodimers (data from ¹⁹). Two ZRE₂ containing 8-mers are highlighted in magenta. (C) Same as in (A) but for CREB1, (D) same as in (A) but for CEBPB (data from ²⁰).

Figure 3. B-ZIP TF binding to DNA 8-mers containing cytosine or 5mC

(A) 8-mer Z-scores for Zta binding to DNA 8-mers with cytosine (x-axis) or 5mC in one DNA strand (y-axis). Of the 65,536 8-mers on the array, 6,561 8-mers do not contain a cytosine, acting as an internal control (grey spots). The remaining 58,975 8-mers are divided into three groups: green spots are 8-mers containing the C/EBP half site GCAA, red and blue spots are 8-mers with TRE half sites CTC^A and GTC^A, and black spots are all other 8-mers that contain cytosine. A selection of well-bound 8-mers are highlighted and their sequences are provided. Sequence logos generated using the top 50 Zta bound 8-mers containing 5mC are shown. (B) Same as in (A) but for cJun|cFos heterodimers. (C) Same as in (A) but for CREB1 homodimers (data from ¹⁵) (D) same as in (A) but for CEBPB.

Figure 4. B-ZIP TF binding to DNA 8-mers containing cytosine or 5hmC
 8-mer Z-scores for (A) Zta homodimers, (B) cJun|cFos heterodimers, (C) CREB1 homodimers (data from ¹⁵), and (D) CEBPB homodimers binding to DNA 8-mers with cytosine (x-axis) or 5mC (y-axis) in one DNA strand. Spots are colored as described in Figure 2. A selection of well-bound 8-mers are highlighted and their sequences are provided. A sequence logo generated using the top 50 Zta bound 8-mers containing 5hmC is shown, where **H**=5hmC.

Figure 5. Zta binding to DNA 8-mers with either 5mC or 5mCG

Zta binding to DNA 8-mers with either 5mC or 5mCG. Z-scores for Zta binding to 8-mers with one cytosine in a CG dinucleotide with 5mC in one strand (x-axis), or those in which both cytosines in the CG dinucleotide are methylated (y-axis) are shown as black spots. 8-mers with the CG dinucleotide at the beginning of the 8-mer (C⁴G³DDDDD, where D= A, G, or T) are shown in blue and those with the CG dinucleotide at C²G³ (DDDDC²G³D) are shown in magenta. As a control, cytosine containing 8-mers for paired experiments in which the double-stranding reaction contained unmodified cytosines are depicted in gray. A sequence logo generated using the top 20 8-mers bound when both cytosines in the CG dinucleotide are methylated is shown.

Figure 6. 5mC as a substitute for thymine (T)

Comparison of Zta Z- scores for non-cytosine-containing 8-mers (x-axis) and 8-mers in which a single T is replaced with 5mC ($T \rightarrow M$, y-axis). T to M substitutions giving rise to a methylated CG dinucleotide ($TG \rightarrow MG$) are colored blue, and those giving rise to the methylated CEBP half-site ($GTAA \rightarrow GMAA$) are colored green. A sequence logo generated using the top 20 8-mers that are better bound by Zta when they contain 5mC than those containing T is shown.

Table 1
Effects of single nucleotide substitutions on Zta binding a non-cytosine containing version of the TRE motif, TGAGTAAAT

Shown are Z-scores for Zta binding containing the indicated nucleotides at each position. Z-scores for nucleotides that do not change the motif are in **bold**. 5mCG Z-scores are shown for substitutions giving rise to a CG dinucleotide.

<u>TGAGTAAAT</u>	-4	-3	-2	0	2	3	4
	T	G	A	G	T	A	A
A	9	11	162	2	13	162	162
C	10	14	16	46	116	145	5
G	7	162	9	162	3	26	8
T	162	36	67	4	162	12	20
M	82	7	21	7	187	51	6
H	66	6	4	3	174	13	7
5mCG	43	-	7	-	-	-	-

Table 2
Effects of single nucleotide substitutions on Zta binding a non-cytosine containing version of the ZRE2 motif, TGAGTCGAT

Shown are Z-scores for Zta binding containing the indicated nucleotides at each position. Z-scores for nucleotides that do not change the motif are in **bold**. 5mCG Z-scores are shown for substitutions giving rise to a CG dinucleotide.

<u>TGAGTCGAT</u>	-4	-3	-2	0	2	3	4
	T	G	A	G	T	G	A
A	0	2	26	2	4	162	26
C	3	1	6	17	17	145	1
G	0	26	4	26	2	26	6
T	26	4	10	2	26	12	4
M	19	-2	4	0	115	51	2
H	13	0	2	2	62	13	3
5mCG	10	-	5	-	284	-	-