Published in final edited form as:

Biochem Biophys Res Commun. 2018 July 02; 501(4): 905–912. doi:10.1016/j.bbrc.2018.05.080.

Replacing C189 in the bZIP domain of Zta with S, T, V, or A changes DNA binding specificity to four types of double stranded DNA

Sreejana Ray¹, Desiree Tillo¹, Nima Assad¹, Aniekanabasi Ufot¹, Christopher Deppmann², Stewart R. Durell³, Aleksey Porollo⁴, and Charles Vinson^{1,*}

¹Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

²Departments of Biology, Cell Biology, and Biomedical Engineering, University of Virginia, Charlottesville, Virginia 22903

³Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

⁴Center 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

Zta is a bZIP transcription factor (TF) in the Epstein-Barr virus that binds unmethylated and methylated DNA sequences. Substitution of cysteine 189 of Zta to serine (Zta(C189S)) results in a virus that is unable to execute the lytic cycle which was attributed to a change in binding to methylated DNA sequences. To learn more about the role of this position in defining sequence-specific DNA binding, we mutated cysteine 189 to four other amino acids producing Zta(C189S), Zta(C189T), Zta(C189A), and Zta(C189V) mutants. Zta and mutants were used in protein binding microarray (PBM) experiments to evaluate sequence-specific DNA binding to four types of double-stranded DNA (dsDNA): 1) with cytosine in both strands (DNA(CIC)), 2) with 5-methylcytosine (5mC) in one strand and cytosine in the second strand (DNA(5mCIC)), 3) with 5-hydroxymethylcytosine (5hmC) in one strand and cytosine in the second strand (DNA(5hmCIC)), and 4) with both cytosines in all CG dinucleotides containing 5mC (DNA(5mCG)). Zta(C189S) and Zta(C189T) bound the TRE (AP-1) motif (TGAG^G/CTCA) more strongly than wild-type Zta, while binding to other sequences, including the C/EBP half site GCAA was reduced. Binding of Zta(C189S) and Zta(C189T) to DNA containing modified cytosines (DNA(5mCIC), DNA(5hmCIC), and DNA(5mCG)) was reduced compared to Zta. Zta(C189A) and Zta(C189V) had higher non-specific binding to all four types of DNA. Our data suggests that position C189 in Zta impacts sequence-specific binding to DNA containing modified and unmodified cytosine.

*Corresponding author vinsonc@mail.nih.gov.

Keywords

Zta; bZIP; transcription factor; cytosine methylation; DNA binding

Introduction

The Epstein-Barr virus transcription factor (TF) Zta is a key regulator of the switch from the latent to lytic phases of infection [1]. Zta contains a bZIP domain that interacts with DNA in a sequence-specific manner [2]. Zta homodimers can bind several unmethylated and methylated DNA sequences called Zta response elements or ZREs [3]. These include the unmethylated AP-1 or *12-O-Tetradecanoylphorbol-13-acetate (TPA)-responsive element* (TRE) motif [4] ($\text{T}^{-4}\text{G}^{-3}\text{A}^{-2}(\text{G/C})^0\text{T}^2\text{C}^3\text{A}^4$) and two variants containing 5-methylcytosine (5mC, **M**) in CG dinucleotides: the methylated TRE (meTRE, $\text{M}^{-4}\text{G}^{-3}\text{A}^{-2}\text{G}^0\text{T}^2\text{C}^3\text{A}^4$) [5], and the methylated ZRE2 (meZRE2, $\text{T}^{-4}\text{G}^{-3}\text{A}^{-2}\text{G}^0\text{M}^2\text{G}^3\text{A}$) [6]. In both meTRE and meZRE2, 5mCs replace thymines (**T**) in one strand of the TRE. A recent X-ray structure of mutant Zta(C189S) binding meZRE2 highlights how the methyl group on 5mC is similar to the analogous methyl group in thymine [7].

Extending this work, we recently examined the bZIP domain of Zta binding to four types of double-stranded DNA (dsDNA) using protein binding microarrays (PBMs) [8]. We found that many TRE variants containing methylated cytosines at the aforementioned two positions ($\text{M}^{-4}\text{G}^{-3}$ and M^2G^3) were well bound by Zta, with motifs containing a methylated CG dinucleotide at M^2G^3 being among the strongest bound sequences. Unlike previous findings [7], our data indicates that Zta does not bind meTRE ($\text{M}^{-4}\text{G}^{-3}$) as strongly as the TRE ($\text{C}^{-4}\text{G}^{-3}$) or meZRE2 (M^2G^3) sequences. We and others have also shown Zta can bind DNA sequences with 5mC outside of CG dinucleotides, in the case of the methylated C/EBP half-site **GMAA** [7, 8], similar to that observed for another bZIP domain, CREB1 [9]. In addition, the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC, **H**) can increase Zta binding to hydroxymethylated C/EBP half-sites (**GHAA**) [8].

Several X-ray structures of bZIP homodimers in complex with dsDNA, including Zta [7, 10], CREB1 [11], and cJun-cFos (AP-1) [12] have mutated the naturally occurring cysteine to serine to allow for crystal formation. Cysteine in one molecule can form a disulfide bond with the cysteine in the second molecule resulting in a homodimer that compromises crystallization [10]. In Zta, this cysteine is in position 189.

Zta(C189S) results in a virus that is defective in DNA replication and is unable to execute the virus lytic cycle [6, 13]. However, it is unclear whether this replication defect in Zta(C189S) mutants is due to altered DNA binding specificity, which may be possible as amino acids in position 189 are in close proximity to the phosphate 5' of thymines (T^{-4}) in the meZRE2 and TRE motifs [7]. Van der Waals interactions between the beta carbon of S189 in Zta(C189S) with T^{-4} of both the TRE and meZRE2, and with methylated cytosine (M^{-3}) of the opposite strand of meZRE2 are also possible [7]. C189S mutants retain sequence-specificity in binding to several sequences, such as the meZRE2 [13], oriLyt (**TGTGTAA**) [14], and to the meTRE and TRE motifs [7]. One report indicates Zta(C189S)

loses binding specificity to methylated DNA sequences critical for *in vivo* function, such as meZRE2 and meZRE3 (**TTMGMGA**) [6]. Additionally, the consequence of changing Zta to Zta(C189S) on binding other DNA sequences has not been determined.

In this study, we examine the DNA binding specificity of four mutations of Zta(C189) – C189S, C189T, C189A, and C189V – to four types of dsDNA using PBMs [15–18]. By evaluating binding to all DNA 8-mer sequences, we are able to examine the differences in sequence-specific DNA binding between Zta and Zta(C189) mutants.

Material and methods

Cloning and expression of mutant bZIP DNA binding domains.

The wild-type construct of Zta (bZIP DNA binding domain (DBD) plus 50 flanking amino acids) was obtained as a N-terminal GST construct cloned into a modified pDEST15 MAGIC vector [19]. Mutant constructs of Zta (C189S, C189T, C189V, and C189A) were generated via site-directed mutagenesis of the wild-type construct (GenScript). Proteins were expressed using the PURExpress *In vitro* protein synthesis kit (NEB) as described previously [17]. Protein expression was similar for Zta and mutants as monitored using an anti-GST antibody (Fig S1). The amino acid sequence of the Zta bZIP domain with the DBD (bold) and the mutated residue (underlined) is shown below:

STVQTAAAVVFACPGANQGQQLADIGVPQPAPVAAPARRTRKPQPESLEECDSELE
IKRYKNRVASRKCRAKFKQLLQHYREVAAAKSSSENDRLRLLKQMCPSLDVDSIIP
 RTPDVLHEDLLNF

PBM Experiments.

The PBM method has been described previously [15–17]. We used the “HK” array design described in the NCBI Gene Expression Omnibus (GEO) platform GPL11260. Single-stranded oligonucleotides on each array were made double-stranded by primer extension using cytosine, 5-methylcytosine (5mC, NEB) or 5-hydroxymethylcytosine 5hmC (5hmC, Zymo Research) in the primer extension reaction as previously described [18]. Enzymatic methylation of CG dinucleotides on the HK microarray was performed as described in [17].

Image quantification and calculation of 8-mer Z-scores.

For each PBM, image quantification and calculation of 8-mer Z-scores were performed as described previously [8, 18]. All proteins were assayed with at least two replicates (representing individual IVT reactions and arrays), with good agreement between replicates and previously published PBM data for Zta [8] (R = 0.8, Table S1, Fig S2). Z-scores for 5mC, 5mCG, and 5hmC PBM data were rescaled relative to unmodified cytosine using the slope of the line of best fit computed from the Z-scores using 8-mers without cytosine (Table S2). Arrays with the fewest saturated spots were used for further analysis. Raw and processed data are available at the NCBI GEO database under accessions GSE108272 (reviewer token: cxqziqgudlapzid) and GSE108273 (reviewer token: sjuhkkgwbzqfbal).

Results

Generation of four types of ds DNA on protein binding microarrays

Single-stranded DNA 60-mers on microarrays were double stranded with: 1) cytosine, producing dsDNA with cytosine in both strands (DNA(C|C)), 2) 5mC (**M**), producing dsDNA with 5mC in one strand and cytosine in the second strand (DNA(5mC|C)), or 3) 5hmC (**H**), producing DNA(5hmC|C) [8, 18, 20]. A fourth type of dsDNA was generated by enzymatic methylation of the two cytosines in all CG dinucleotides (one cytosine on each strand) [17]. These microarrays were used for PBM experiments using GST-tagged Zta or Zta(C189) mutants to examine sequence-specific DNA binding. Binding is detected using a fluor-conjugated antibody to the GST epitope followed by measurement of bound antibody (fluorescence intensities) at each of the 40,330 array features.

Zta and Zta(C189) mutants binding to DNA(C|C)

We initially examined the fluorescent intensities of the 40,330 features containing DNA(C|C) for Zta and the four Zta(C189) mutations (Fig 1A-D, Table S3,S4). Examination of fluorescent intensities allows for a quantitative measure of binding to different features on the array, which invariably contain a preferentially bound motif. Essentially, we are examining the binding of the best bound motif in a random background. In general, binding of the four mutants to most features is decreased relative to Zta. Protein concentrations in the *in vitro* translation reactions used for all experiments show similar amounts of protein (Fig S1), and similar results are obtained across different protein preparations and arrays (Fig S2), suggesting that weaker binding observed is due to different sequence-specific DNA binding properties. Features containing the C/EBP half-site **GC²AA** are better bound by Zta, while features containing the TRE motif are preferentially bound by the Zta(C189S), and Zta(C189T) mutants (Table S4).

To examine sequence-specific binding, we computed a standardized score (Z-score) for all possible 8-bp sequences (8-mers) for each PBM experiment [15]. The Z-score for a given 8-mer is the number of standard deviations that the median intensity of that 8-mer (computed from all array features containing the 8-mer) is from the global median intensity (computed from all array features). This measure is sensitive to changes in global median intensity, representative of non-specific DNA binding. Fig 1E is a scatterplot of Z-scores for Zta (x-axis) and Zta(C189S) (y-axis) binding all 32,896 8-mers of DNA(C|C), where complements are combined. 8-mers containing the canonical TRE (**TGA^C/G^TTCA**) are more specifically bound (higher Z-scores) by Zta(C189S). 8-mers containing the TRE variant **TGAGTA³A** are better bound by Zta, suggesting that C189 affects Zta binding to nucleotides at position 3. Greater variability is observed in comparing Zta and Zta(C189S) than their respective duplicates ($R=0.85$ vs. $R=0.98$ for duplicates, Z-test $p<2e-308$, Fig S2), indicating the observed differences in binding 8-mers are real.

We next examined Zta(C189T) binding to 8-mers containing DNA(C|C). Like Zta(C189S), 8-mers containing the TRE motif are preferentially bound relative to Zta (Fig 1F). Zta(C189A) and Zta(C189V) bind the TRE sequence, but not as strongly as Zta (Fig 1G-H). Zta(C189S) and Zta(C189T), and Zta(C189A) and Zta(C189V) bind DNA(C|C) 8-mers

similarly, as may be expected as these pairs of amino acids have similar physico-chemical properties (Fig S3A-B). Here, DNA binding specificity correlates with the polarity gradient in the mutant side chains: a pair of polar mutants, Zta(C189S) and Zta(C189T), have a similar DNA binding profile, which is distinct from the one elicited by the two hydrophobic mutants, Zta(C189A) and Zta(C189V).

Zta and Zta(C189) mutants binding to DNA(5mCIC)

We next compared Zta to C189 mutants binding to DNA(5mCIC) (Fig 2). All 65,536 8-mers are examined as complements are not identical since only one strand contains 5mC. Zta, Zta(C189S), and Zta(C189T) bind many features containing DNA(5mCIC), and median and average fluorescent intensity for all features increases relative to DNA(CIC) (Table S2). For Zta and Zta(C189S), many well-bound DNA(5mCIC) 8-mers contain the C/EBP half-site **GM²AA** (Fig 2A,S4) as observed previously for Zta [7, 8] and CREB1 [9]. While most **GM²AA**-containing 8-mers are better bound by Zta compared to Zta(C189S), several 8-mers are strongly bound by both proteins (e.g. **ATGAGM²AA** and **ATGTGM²AA**). In addition, several non-**GM²AA** containing 8-mers are more strongly bound by Zta(C189S), including those that contain the subsequence **TGTGM²A** (Fig 2A, labelled). Zta(C189T) binds similar DNA(5mCIC) 8-mers as Zta(C189S) (Fig 2B and Fig S3C). Zta(C189A) and Zta(C189V) bind most 8-mers weakly as observed for binding DNA(CIC) with the exception of the 8-mers **ATGAGM²AA** and **TGAGM²AAT** (Fig 2C-D, S3D).

Zta and Zta(C189) mutants binding to DNA(5hmCIC)

Fig 3A presents a comparison of Z-scores for binding DNA(5hmCIC) 8-mers for Zta (x-axis) [8] and Zta(C189S) (y-axis). Zta(C189S) binds several 8-mers containing the hydroxymethylated C/EBP half-site **GH²AA**, similar to Zta (e.g. **TGTGH²AAT** and **ATGAGH²AA**, Fig S5A-B). However, many 8-mers are weakly bound compared to Zta (e.g. **ATTAGH²AA**). Again, Zta(C189T) has similar binding properties to Zta(C189S) (Fig 3B, S3E). 5hmC abrogates Zta(C189A) and Zta(C189V) binding (Fig S5D-E) with all 8-mers being weakly bound relative to wild-type Zta (Fig 3C-D,S3F).

Zta and Zta(C189) mutants binding DNA(5mCG)

Previously, we showed that Zta binds many 8-mers containing DNA(5mCG) (Fig S6A) [8]. We compared the 8-mer binding of Zta to Zta(C189) mutants (Fig 4) to DNA(5mCG). As seen with DNA(CIC) (Fig 1), the TRE that does not contain a CG dinucleotide is more specifically bound by all mutants than by Zta. Zta(C189S) and Zta(C189T) mutants bind several DNA(5mCG) 8-mers, including those that contain the meZRE2 and the meZRE2-like sequence **TGTGMGA** (Fig 4A-B,S6B-C; Table S5A). Most DNA(5mCG) 8-mers, however, are not as strongly bound by Zta(C189S) and Zta(C189T) compared to Zta, including 8-mers containing the meZRE3 sequence (Table S5B), consistent with previously reported results [6]. In contrast, Zta(C189A) and Zta(C189V) bind most DNA(5mCG) 8-mers worse than Zta (Fig 4C-D; Fig S6D-E). The meTRE (**M⁻⁴G⁻³AGTGA**) is modestly bound by all mutants similar to our results for Zta [8] (Fig S7, Table S5C), indicating that C189 does not affect Zta binding to the meTRE.

Zta and mutants binding to single nucleotide polymorphisms (SNPs), 5mC, and 5hmC throughout the TRE motif variant TGAGTA³A

We evaluated the effect of SNPs and modified cytosines to further examine DNA binding specificity of Zta(C189) mutants to the TRE (Table 1 and S7). This was achieved by examining the Z-scores of SNPs in a TRE motif without a cytosine, **TGAGTA³A**. Most SNPs result in similar changes in binding in Zta and Zta(C189) mutants, including increases in binding specificity for all four C189 mutants when **T²** is replaced with **M²** as observed previously for Zta [8] and CREB1 [18]. One exception is in position 3. Here, when **A³** is replaced with **C³**, the Z-score is not greatly affected for Zta (decreases from 162 to 145) (Table 1A), whereas binding increases for Zta(C189S) (Table 1B) and all other mutants (Table S7). For example, Zta(C189S) has a Z-score of 52 for binding **A³** and a Z-score of 211 for **C³** (Table 1B).

C189 does not greatly affect cooperativity between methylated cytosines in binding to meZRE2

Previously, we showed that methylation of both cytosines in the **C²G³** dinucleotide of the meZRE2 motif contributes to stronger Zta binding [8]. We next examined the effects of the C189 mutations on binding the individual methylated cytosines in the meZRE2 motif **T⁻⁴G⁻³A⁻²G⁰M²G³A⁴** (complement: **T⁻⁴M⁻³G⁻²C⁰T²C³A⁴**). Table 2 presents Z-scores for Zta and Zta(C189S) binding a non-cytosine containing ZRE2 motif variant **TGAGT²GA**. Here, methylation of both cytosines in the **C²G³** dinucleotide results in higher Z-scores (5mCG Z-score=289) compared to when individual cytosines are methylated (Z-score for **M²**=168, Table 2A). Increases in Z-score are also seen for the all Zta(C189) mutants (Table 2B,S6). Additional 8-mers are more strongly bound by Zta(C189S) and Zta(C189T) mutants when both cytosines are methylated at **C²G³** (Fig S8). This increase in binding is not as dramatic as observed for wild-type Zta, suggesting that the cooperativity between individual methylated cytosines in meZRE2 is mediated only in part by the residue at position 189.

Structural differences in Zta(S189) and Zta(C189) on binding dsDNA

To examine the structural basis of the differences in DNA binding between Zta and C189 mutants, we examined nucleotide-residue contact maps [21] for Zta(C189S) (PDB: 5szx [7]) (Fig S9A) binding the canonical meZRE2 motif (**TGAGM²GA**, complement: **TM⁻³GCTCA**) and a structure where serine is changed to cysteine (Fig S9B). As summarized in the introduction, the serine in Zta(C189S) of each Zta monomer interacts via hydrogen bonds (bold red lines, Fig S9A) with the phosphate backbone upstream of the thymines at each end of the meZRE2 (**T⁻⁴** and **T⁻⁴'**), as well as additional van der Waals interactions with **T⁻⁴** and the methylated cytosine of the opposite strand of meZRE2 (**M⁻³'**) (bold black lines, Fig S9A and [7]). Examination of the nucleotide-residue contact map for Zta(C189) shows similar interactions, with an additional interaction with the DNA phosphate backbone just outside of the meZRE2 at position -5 (Fig S9B, circled). Combined with the PBM results, the structural analysis suggests that the main driving force in the Zta-DNA interaction at position 189 is hydrogen bonding (Fig S9C). As long as high polarity of the X^γ-H bond is maintained (X = O in serine or threonine), this residue contributes to stronger DNA binding with a phosphate group of the DNA backbone. With the less polar bond (X =

S or C in cysteine and valine, respectively) or lack thereof (alanine), binding to DNA is reduced.

Discussion

We examined the effect of changing cysteine at position 189 in the bZIP domain of Zta on DNA binding specificity. Here, we utilized a PBM platform to evaluate the sequence-specificity of Zta and four mutants Zta(C189S), Zta(C189T), Zta(C189A), and Zta(C189V) to four types of dsDNA containing either cytosine on both strands, 5mC or 5hmC on one DNA strand, or 5mCG. The mutants were selected to have side chains with similar size to cysteine to prevent additional steric effects and to present a gradient in hydropathy from polar (serine and threonine) to hydrophobic (alanine and valine).

Zta(C189S) and Zta(C189T) bound the TRE motif (**TGA^G/C^TTCA**) stronger than Zta while binding to all other sequences was reduced. Zta(C189S) and Zta(C189T) also bound DNA(5mC/C), DNA(5hmC/C), DNA(5mCG) similarly to Zta, with reductions in binding to several 8-mers containing these modified cytosines, suggesting that C189 plays a general role in stabilizing or promoting binding to methylated DNA sequences. Zta(C189A) and Zta(C189V) bound DNA but had reduced sequence specificity to all four types of dsDNA examined. Our structural analysis indicates that additional hydrogen bonds between polar groups of the side chains of serine or threonine with the DNA backbone stabilize Zta(C189S) or Zta(C189T) binding, explaining the increased binding of these mutants to preferentially bound sites, such as the TRE.

Examination of the amino acid sequence of the bZIP domain across the six main eukaryotic lineages highlights the diversity and variability of the amino acids found at the position analogous to C189 in Zta [22]. bZIPs in plant, amoeba, heterokont, and excavate lineages primarily have amino acids containing uncharged polar side chains such as tyrosine or serine. All plants possess serine at this position [22, 23]. Cysteine, however, is only observed in ophisthokont (metazoan and fungal) bZIP proteins and is most prevalent in metazoan bZIPs. The emergence of cysteine in the bZIP domain may have evolved to enable responsiveness to oxidative stress produced by demanding physiological conditions. For example, in the case of c-JUN/c-FOS heterodimers (AP-1), oxidative stress results in disulfide bond formation between individual bZIP monomers, and loss of DNA binding [24]. Vertebrate bZIP proteins involved in modulating cell growth (FOS family members: c-FOS, c-JUN, CREB1, and ATF2) contain cysteine, whereas those involved in differentiation (e.g. C/EBP and PAR family members) have serine.

The differences between Zta and Zta(C189S) in binding DNA underscores the ability of the PBM platform to reveal changes in sequence-specific DNA binding. We suggest that polarity in the side chain at position 189 (serine and threonine) to retain sequence-specific binding of Zta to the TRE while hydrophobic side chains (alanine and valine) compromise its sequence-specificity. Further, serine and threonine reduce sequence-specificity to several methylated sequences such as meZRE2 which may mediate the *in vivo* function of Zta. More sophisticated molecular modeling may allow for the elucidation of the physical mechanisms underlying these observations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the intramural research project of the National Cancer Institute, National Institutes of Health (NIH).

References

- [1]. Miller G, El-Guindy A, Countryman J, Ye J, Gradoville L, Lytic cycle switches of oncogenic human gammaherpesviruses, *Adv Cancer Res*, 97 (2007) 81–109. [PubMed: 17419942]
- [2]. Vinson CR, Sigler PB, McKnight SL, Scissors-grip model for DNA recognition by a family of leucine zipper proteins, *Science*, 246 (1989) 911–916. [PubMed: 2683088]
- [3]. 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*, 6 (2011) e25922. [PubMed: 22022468]
- [4]. 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*, 49 (1987) 729–739. [PubMed: 3034432]
- [5]. 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*, 42 (2014) 3059–3072. [PubMed: 24371273]
- [6]. 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*, 4 (2008) e1000005. [PubMed: 18369464]
- [7]. 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*, 45 (2017) 2503–2515. [PubMed: 28158710]
- [8]. Tillo D, Ray S, Syed KS, Gaylor MR, He X, Wang J, Assad N, Durell S, Porollo A, Weirauch MT, Vinson C, The Epstein-Barr virus B-ZIP protein Zta recognizes specific DNA sequences containing 5-methylcytosine and 5-hydroxymethylcytosine, *Biochemistry*, (2017).
- [9]. 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*, 55 (2016) 6940–6948. [PubMed: 27951657]
- [10]. 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*, 21 (2006) 565–572. [PubMed: 16483937]
- [11]. 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*, 275 (2000) 35242–35247. [PubMed: 10952992]
- [12]. Glover JN, Harrison SC, Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA, *Nature*, 373 (1995) 257–261. [PubMed: 7816143]
- [13]. Wang P, Day L, Dheekollu J, Lieberman PM, A redox-sensitive cysteine in Zta is required for Epstein-Barr virus lytic cycle DNA replication, *J Virol*, 79 (2005) 13298–13309. [PubMed: 16227252]
- [14]. Schelcher C, Valencia S, Delecluse HJ, Hicks M, Sinclair AJ, Mutation of a single amino acid residue in the basic region of the Epstein-Barr virus (EBV) lytic cycle switch protein Zta (BZLF1) prevents reactivation of EBV from latency, *J Virol*, 79 (2005) 13822–13828. [PubMed: 16227304]
- [15]. 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*, 24 (2006) 1429–1435. [PubMed: 16998473]

- [16]. 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*, 39 (2011) 4680–4690. [PubMed: 21321018]
- [17]. 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*, 23 (2013) 988–997. [PubMed: 23590861]
- [18]. 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)*, 8 (2016) 936–945. [PubMed: 27485769]
- [19]. Sharrocks AD, A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase, *Gene*, 138 (1994) 105–108. [PubMed: 8125285]
- [20]. Sayeed SK, Zhao J, Sathyanarayana BK, Golla JP, Vinson C, C/EBPbeta (CEBPB) protein binding to the C/EBP/CRE DNA 8-mer TTGC/|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide, *Biochim Biophys Acta*, 1849 (2015) 583–589. [PubMed: 25779641]
- [21]. Sagendorf JM, Berman HM, Rohs R, DNAProDB: an interactive tool for structural analysis of DNA-protein complexes, *Nucleic Acids Res*, 45 (2017) W89–W97. [PubMed: 28431131]
- [22]. Jindrich K, Degnan BM, The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity, *BMC Evol Biol*, 16 (2016) 28. [PubMed: 26831906]
- [23]. Deppmann CD, Alvania RS, Taparowsky EJ, Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks, *Mol Biol Evol*, 23 (2006) 1480–1492. [PubMed: 16731568]
- [24]. Abate C, Patel L, Rauscher FJ, 3rd, Curran T, Redox regulation of fos and jun DNA-binding activity in vitro, *Science*, 249 (1990) 1157–1161. [PubMed: 2118682]

Fig 1. Zta and mutants binding DNA(CIC).

(A-D) Comparison of PBM fluorescence intensities for Zta (x-axis) and (A) Zta(C189S), (B) Zta(C189T), (C) Zta(C189A), or (D) Zta(C189V) (y-axis) binding to 40,330 features containing DNA(CIC). Features are divided into three groups: those containing the TRE site TGA^G/C^GTCA (magenta spots), those containing the C/EBP half-site GCAA (green), and all other features (black). (E-H) Scatterplots comparing relative binding (Z-scores) for Zta (x-axis) and (E) Zta(C189S), (F) Zta(C189T), (G) Zta(C189A), or (H) Zta(C189V) (y-axis) to 32,986 8-mers containing DNA(CIC). 8-mers are divided into three groups as in panels (A)-(D). A selection of 8-mers are highlighted and their sequences provided.

Fig 2. Zta and mutants binding DNA(5mC|C).

Scatterplots comparing Zta DNA(5mC|C) 8-mer Z-scores (x-axis) to those of (A) Zta(C189S), (B) Zta(C189T), (C) Zta(C189A), or (D) Zta(C189V) (y-axis). Spots are colored as in Figure 1.

Fig 3. Zta and mutants binding DNA(5hmC|C).

Scatterplots comparing Zta DNA(5hmC|C) 8-mer Z-scores [8] (x-axis) to those of (A) Zta(C189S), (B) Zta(C189T), (C) Zta(C189A), or (D) Zta(C189V) (y-axis). 8-mers are colored as in Figure 1.

Fig 4. Zta and mutants binding DNA(5mCG).

Scatterplots comparing Zta DNA(5mCG) 8-mer Z-scores (x-axis) to those of (A) Zta(C189S), (B) Zta(C189T), (C) Zta(C189A), or (D) Zta(C189V) (y-axis). 8-mers are divided into three groups based on the presence (blue) or absence (black) of methylated CG dinucleotides (MG), and those that contain the TRE motif (magenta).

Table 1.

Z-scores for (A) Zta and (B) Zta(C189S) binding the indicated SNPs of a TRE motif variant, TGAGTAA.

(A) Zta							
<u>TGAGTAA</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	83	5	23	3	434*	70	3
H	66	6	4	3	173	13	7
5mCG	43	-	7	-	-	-	-
(B) Zta(C189S)							
<u>TGAGTAA</u>	-4	-3	-2	0	2	3	4
	T	G	A	G	T	A	A
A	5	4	52	1	7	52 *	52
C	5	8	6	27	59	211*	4
G	3	52	6	52	2	18	7
T	52	6	36	1	52 *	8	13
M	19	-1	13	0	370*	77	4
H	13	2	3	3	177	18	4
5mCG	14	-	5	-	-	-	-

M=5mC, H=5hmC. Z-scores for nucleotides that do not change the motif are in bold. DNA(5mCG) Z-scores are shown for substitutions that result in a CG dinucleotide.

Asterisks(*) highlight substitutions that are mentioned in the text.

Table 2.

Z-scores for (A) Zta and (B) Zta(C189S) binding the indicated SNPs of a meZRE2 variant, TGAGTGA.

(A) Zta								
TGAGTGA	-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	17	-7	14	-2	168*	70	-4	
H	13	0	2	2	62	13	3	
5mCG	10	-	5	-	283*	-	-	
(B) Zta(C189S)								
TGAGTGA	-4	-3	-2	0	2	3	4	
	T	G	A	G	T	G	A	
A	1	2	18	0	4	52	18	
C	3	0	4	21	10	211	2	
G	3	18	2	18	4	18	8	
T	18	2	7	1	18	8	5	
M	11	-3	6	-2	160*	77	4	
H	5	0	1	2	67	18	3	
5mCG	6	-	4	-	158*	-	-	

See legend for Table 1.