Published online 1 February 2017 *Nucleic Acids Research, 2017, Vol. 45, No. 5* **2503–2515**
doi: 10.1093/nar/gkx057

Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein-Barr virus Zta

Samuel Hong[1,2,†], Dongxue Wang[1,†], John R. Horton[1,3], Xing Zhang[1,3], Samuel H. Speck[4,5], Robert M. Blumenthal[6] and Xiaodong Cheng[1,3,*]

[1]Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA, [2]Molecular and Systems Pharmacology graduate program, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA, [3]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA, [4]Department of Microbiology & Immunology, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA, [5]Emory Vaccine Center, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA and [6]Department of Medical Microbiology and Immunology and Program in Bioinformatics, The University of Toledo College of Medicine and Life Sciences, Toledo, OH 43614, USA

Received November 17, 2016; Revised January 09, 2017; Editorial Decision January 19, 2017; Accepted January 21, 2017

ABSTRACT

Activator protein 1 (AP-1) is a transcription factor that recognizes two versions of a 7-base pair response element, either 5′-TGAGTCA-3′ or 5′-MGAGTCA-3′ (where M = 5-methylcytosine). These two elements share the feature that 5-methylcytosine and thymine both have a methyl group in the same position, 5-carbon of the pyrimidine, so each of them has two methyl groups at nucleotide positions 1 and 5 from the 5′ end, resulting in four methyl groups symmetrically positioned in duplex DNA. Epstein-Barr Virus Zta is a key transcriptional regulator of the viral lytic cycle that is homologous to AP-1. Zta recognizes several methylated Zta-response elements, including meZRE1 (5′-TGAGMCA-3′) and meZRE2 (5′-TGAGMGA-3′), where a methylated cytosine occupies one of the inner thymine residues corresponding to the AP-1 element, resulting in the four spatially equivalent methyl groups. Here, we study how AP-1 and Zta recognize these methyl groups within their cognate response elements. These methyl groups are in van der Waals contact with a conserved di-alanine in AP-1 dimer (Ala265 and Ala266 in Jun), or with the corresponding Zta residues Ala185 and Ser186 (via its side chain carbon Cβ atom). Furthermore, the two ZRE elements differ at base pair 6 (C:G versus G:C), forming a pseudo-symmetric sequence (meZRE1) or an asymmetric sequence (meZRE2). *In vitro* DNA binding assays suggest that Zta has high affinity for all four sequences examined, whereas AP-1 has considerably reduced affinity for the asymmetric sequence (meZRE2). We ascribe this difference to Zta Ser186 (a unique residue for Zta) whose side chain hydroxyl oxygen atom interacts with the two half sites differently, whereas the corresponding Ala266 of AP-1 Jun protein lacks such flexibility. Our analyses demonstrate a novel mechanism of 5mC/T recognition in a methylation-dependent, spatial and sequence-specific approach by basic leucine-zipper transcriptional factors.

INTRODUCTION

DNA methylation was initially thought to inhibit transcription of some genes, whereas transcription of other genes was at most unaffected by DNA methylation (1). However, accumulating evidence shows that several families of activating transcription factors (including Epstein-Barr Virus Zta, see below) can preferentially bind methylated CpG (5mCpG) within specific sequences (2–12). A classic basic leucine-zipper (bZIP) transcription factor family, named activator protein 1 (AP-1), is substantially involved in gene regulation that controls such critical phenomena as oncogenesis, cell proliferation, and apoptosis (13,14). AP-1 is a dimeric complex that comprises members of the Jun, Fos, ATF (activating transcription factor) and MAF (musculoaponeurotic fibrosarcoma) protein families (reviewed in (14)).

Like many dimeric transcription factors (Myc/Max and NF-κB), the AP-1 complexes can form homodimers and heterodimers that mediate distinct biological outputs (13,15–18). For example, Jun/Jun homodimer and Jun/Fos

*To whom correspondence should be addressed. Tel: +1 404 727 8491; Fax: +1 404 727 3746; Email: xcheng@emory.edu
†These authors contributed equally to this work as first authors.

heterodimer activate a set of genes by binding a 7-bp 12-*O*-tetradecanoylphorbol-13-acetate (TPA)-response element known as TRE (5′-TGAGTCA-3′), as well as a methylated response element known as meTRE (5′-MGAGTCA-3′ where M = 5mC) with a methylated CpG replacing 5′ TpG dinucleotide (19–21). Like 5-methylcytosine (5mC or M), thymine (5-methyluracil) contains a methyl group at carbon-5. Thus, the TRE and meTRE elements each contain two methyl groups at nucleotide positions 1 and 5 from the 5′ end, resulting in four methyl groups symmetrically positioned at base pair (bp) positions 1, 3, 5 and 7 (Figure 1A).

Epstein-Barr Virus (EBV) is a human-specific B cell-infecting gamma-herpesvirus, associated with diseases such as infectious mononucleosis and Burkitt's lymphoma, which has both lytic and latent phases (22,23). An EBV transcription factor, Zta (also called EB1, BZLF1, or ZEBRA), was the first example of a sequence-specific transcription factor that preferentially recognizes and selectively binds methylated cytosine residues within a specific sequence, reverses epigenetic silencing and activates gene transcription (2). The EBV virion genome is unmethylated, but becomes heavily methylated during the latent stage of the virus cycle (24–26). Early lytic cycle activation depends upon the EBV Zta homodimer preferentially recognizing methylated promoters containing meZREs, a notable example of which is meZRE2 (5′-TGAGMGA-3′) (2,27–29). Significantly, this element contains two methyl groups at nucleotide positions 1 and 5 (Figure 1A).

Both human AP-1 and EBV Zta are bZIP family transcription factors that bind the classic TRE, but recognize methylated cytosine residues within different sequence contexts. Specifically, AP-1 recognizes M at position 1, whereas Zta recognizes M at position 5 of the 7-bp recognition sequences. Previous structural analysis (19,30) and sequence alignment of basic regions of AP-1 transcription factors and Zta shows that the four DNA base-contacting amino acids are highly conserved, except for Zta Ser186 which is equivalent to Jun Ala266 in AP-1 (Figure 1B). Zta Ser186 is critical for meZRE2 binding (29,31,32). Although a crystal structure of Zta bZIP homodimer-DNA complex was previously solved (PDB 2C9L), the reported structure has an S186A mutation (to mimic AP-1) in complex with oligonucleotides containing TRE, not meZRE (30). To date, it is not known how AP-1 and Zta recognize 5mC within their cognate methylated response elements.

Using isolated Jun and Zta bZIP domains, each forming a homodimer, we here report two crystal structures: a Jun/Jun-DNA complex with meTRE, and a Zta/Zta-DNA complex with meZRE2. The spatially conserved methyl groups are in contact with the conserved di-alanine (Ala265 and Ala266 in Jun) or the corresponding Zta residues Ala185 and the side chain carbon Cβ of Ser186. Further, we show that the purified Zta binds both meTRE and meZRE2 strongly and equally, while Jun binds the two elements with significantly different affinities. Our findings provide molecular explanations for the differences between the human and viral factors, and potentially help explain the behavior of other bZIP transcription factors.

MATERIALS AND METHODS

Protein expression and purification

We prepared two Jun proteins: (i) human Jun bZIP (UniProtKB/Swiss-Prot: P05412.2), residues 254–315 containing a Cys269-to-Ser mutation in the basic region, with wild-type Ala266 (Jun WT, pXC1398), and (ii) the A266S mutant of that protein (Jun A266S, pXC1440). Both were expressed as an N-terminal 6xHis-SUMO fusion (33), via a modified pET28b vector (Novagen) in *Escherichia coli* BL21-CodonPlus(DE3)-RIL (Stratagene) (Supplementary Figure S1).

Equivalent EBV Zta bZIP proteins were also prepared: (i) residues 175–236 containing a Cys189-to-Ser mutation in the basic region with wild-type Ser186 (Zta WT, pXC1416), and (ii) the S186A mutant of that protein (Zta S186A, pXC1455). These were expressed using the same vector as the Jun proteins. Published data showed that the Cys-to-Ser mutation was necessary to generate quality crystals and/or to prevent the proteins from oxidation for both Fos/Jun heterodimer and Zta/Zta homodimer (19,30). Because the Cys-to-Ser mutant did not adversely affect DNA binding affinity or specificity (34,35), for simplicity, we used the same Cys-to-Ser mutant in our biochemical and structural analyses. The altered residue, located in the basic DNA binding region, contacts a phosphate group of the DNA backbone in our structures (Supplemental Figure S2).

Bacterial cells were cultured in LB medium at 37°C, and protein expression was induced at 16°C overnight by adding 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Cells were harvested and stored at −80°C. Frozen pellets were thawed, and lysed by sonication in 20 mM sodium phosphate pH 7.4, 500 mM NaCl, 25 mM imidazole, 5% (v/v) glycerol, and 1 mM 1,4-dithiothreitol (DTT). Lysates were clarified by centrifugation at 35 000 g for 1 h, and the fusion protein was isolated on a nickel-charged HisTrap affinity column (GE Healthcare). Fractions eluted from the nickel column with a linear gradient of imidazole (25–500 mM) were pooled.

For further purification of the Jun bZIP domain, ubiquitin-like-specific protease 1 (ULP-1; purified in-house (33)) was added to the pooled nickel column fractions, followed by overnight incubation at 16°C to cleave the 6xHis-SUMO tag, resulting in the retention of two extraneous N-terminal residues (His-Met). The tag-cleaved sample was then loaded onto tandem HiTrap-Q/HiTrap-Heparin columns (GE Healthcare), followed by elution from the Heparin column using a linear gradient of NaCl (0.5–2 M). The eluted fractions were loaded onto a Superdex 200 (16/60) size exclusion column (GE Healthcare), in buffer containing 20 mM Tris–HCl pH 7.5, 500 mM NaCl, 5% (v/v) glycerol and 1 mM DTT. The final concentration of the purified Jun/Jun homodimer was estimated by Bradford protein assay (Bio-Rad no. 500-0205) (due to lack of aromatic residues).

For further purification of the Zta bZIP domain, the pooled Ni column fractions were loaded onto tandem HiTrap-Q/HiTrap-Heparin columns (GE Healthcare) in the Zta buffer (20 mM Tris–HCl pH 7.5, 150 mM ammonium acetate, 150 mM NaCl, and 1 mM DTT). The Hep-

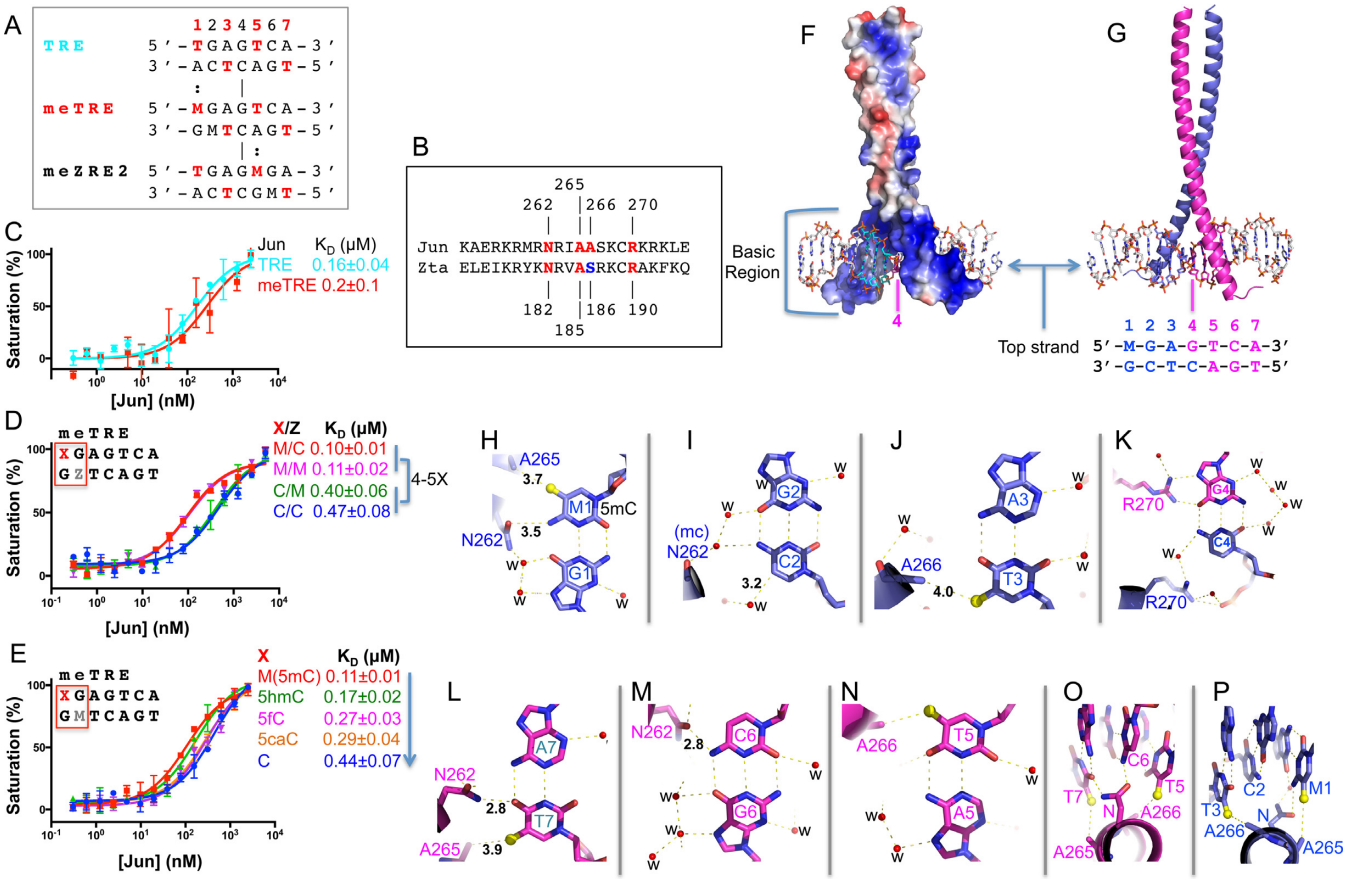

Figure 1. Jun/Jun-meTRE interactions. (**A**) Aligned DNA sequence elements of TRE, meTRE and meZRE2. (**B**) Protein sequence alignment of the basic DNA-binding regions of AP-1 (Jun) and Zta. (**C**–**E**) DNA binding affinities of the Jun bZIP domain with oligos containing variations in the first and second base pairs 'X' and 'Z' positions [M is 5mC, and can be oxidized to 5-hydroxymethylC (5hmC), 5-formylC (5fC) and 5-carboxylC (5caC)]. The reduced binding by oxidation is probably due to combined effects of steric constraints, charge differences, and/or decreased van der Waals contacts between protein and DNA. (**F**) The surface charge of Jun homodimer at neutral pH is displayed as blue for positive, red for negative, and white for neutral. (**G**) Each Jun monomer (magenta and blue) recognizes one half site. (**H**–**J**) Interactions with the first half site (base pairs 1–3). The black numbers indicate the inter-atom distance in angstroms; 'w' is water molecules; and 'mc' refers to main-chain atoms. (**K**) Interactions with the central G_4:C_4 base pair. (**L**–**N**) Interactions with the second half site (base pairs 7-5). (**O**–**P**) The similarity (di-alanine-methyl group interactions) and difference (Asn262 conformations) between Jun/Jun interactions with the two half sites. Atoms are colored dark blue for nitrogen, red for oxygen and either light blue or magenta for carbons of each half site. The methyl groups of 5mC and T are colored yellow with a small sphere.

arin column was then eluted using a linear gradient of NaCl (150 mM to 2 M). The eluted fractions were pooled and dialyzed against the Zta buffer in presence of ULP-1 in 4°C to cleave the 6xHis-SUMO tag. The dialyzed sample was then loaded onto a heparin column followed by elution as before. The eluted fractions were then dialyzed against the Zta buffer again, concentrated, and then loaded onto a Superdex 200 (16/60) size exclusion column equilibrated in the Zta buffer. Elution from the column yielded a single peak, corresponding to the expected Zta bZIP homodimer size. The final concentration of the purified homodimer was estimated by measuring absorbance at 280 nm (with an extinction coefficient of 0.395).

Fluorescence-based DNA binding assay

Fluorescence polarization assays were performed using a Synergy 4 microplate reader (BioTek) to measure DNA binding affinity, as described (6,36). The 6xHisSUMO tagged-protein and 6-caboxy-fluorescein (FAM)-labeled double strand DNA probe (5 nM) were incubated with increasing amount of proteins in 20 mM Tris–HCl, pH 7.5, 5% glycerol, and 185 mM NaCl (for Jun/Jun) or 225 mM NaCl (for Zta/Zta). The DNA sequences of the probe for Jun/Jun were 5′-GGAXGAGTCATAG-3′ and FAM-5′-CTATGACTXGTCC-3′; and for Zta/Zta were 5′-CTATGAGXGATCC-3′ and FAM-5′-GGATXGCTCATAG-3′; where **X** can be normal cytosine (C), 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), or 5-carboxylcytosine (5caC). K_D values were calculated as [mP] = [maximum mP] x [C]/(K_D + [C]) + [baseline mP], where mP is millipolarization and [C] is protein concentration (37). Curves were fit by GraphPad Prism software (version 6.0) and normalized as percentage of bound as 100 x ([mP] – [baseline mP])/([maximum mP] – [baseline mP]. The reported mean ± SEM of the interpolated K_D values were calculated from two independent experiments performed in duplicate. For those binding curves that did not reach saturation, the lower limit of the binding affinity was estimated.

Crystallography

For Jun bZIP homodimer in complex with DNA, purified Jun/Jun was mixed with annealed oligonucleotides containing the meTRE sequence (hemi-methylated CpG; Supplementary Table S1) at a molar ratio of ∼1:1. The final complex was concentrated to ∼1 mM in 20 mM Tris–HCl (pH 7.5), 100 mM NaCl, and 5% (v/v) glycerol. Initial screening was performed by the sitting-drop method using an Art Robbins Phoenix Crystallization Robot, and selected conditions were repeated and optimized by the hanging-drop method. The final rod-shaped crystals appeared at 16°C within 3 days, in mother liquor containing 50 mM citric acid, 50 mM bis–tris–propane, and 16% (w/v) polyethylene glycol 3350 at pH 5.2.

For Zta bZIP homodimer in complex with DNA, purified Zta/Zta was mixed with annealed oligonucleotides containing meZRE-2 sequence (fully methylated CpG) at a molar ratio of ∼1:1. The final complex was concentrated to ∼1 mM in 20 mM Tris–HCl (pH 7.5), 150 mM ammonium acetate, 150 mM NaCl, and 1 mM DTT. A wide range of screening conditions yielded well-diffracting crystals at 16°C within 2 months. The crystal used for data collection was harvested in mother liquor containing 0.2 M sodium phosphate monobasic monohydrate and 20% (w/v) polyethylene glycol 3350 at pH 4.7.

Crystals were examined at the SER-CAT 22ID beamline at the Advanced Photon Source, Argonne National Laboratory, and the diffraction data were processed using HKL2000 (38). Crystallographic phases for Jun/Jun-DNA and Zta/Zta–DNA complexes were determined by molecular replacement using coordinates from the protein components of previous structures (PDB 1FOS for Jun (19) and PDB 2C9L for Zta (30)). Model refinements were performed using PHENIX (39). Graphics for the figures were generated using PyMol (DeLano Scientific, LLC). Detail X-ray data collection and refinement results are summarized in Supplementary Table S1.

RESULTS

Jun homodimer has equal affinity for TRE and meTRE sequences

We first compared the binding of Jun protein to double-stranded oligonucleotides (oligos) containing the TRE (5′-**T**GAGTCA-3′) or meTRE (5′-**M**GAGTCA-3′) sequences. These elements differ only at the first base pair (bp): with T:A in TRE and 5mC:G (M:G) in meTRE (Figure 1A). As expected from previous studies (21), Jun protein bound to both sequences equally well, with dissociation constants (K_D) of ∼0.2 μM as measured by fluorescence polarization (Figure 1C). We next analyzed the effect of methylation on only one strand by replacing one of the two 5mC bases with unmodified C. The dissociation constant for the duplex oligo hemimethylated on the top strand CpG (M/C) sequence yielded the same K_D as that for the completely methylated (M/M) sequence; however the bottom hemimethylated (C/M) and completely unmodified (C/C) sequences showed comparable but 4–5× higher dissociation constants, i.e. lower binding affinities (Figure 1D). These results indicate that 5mC methylation of the CpG dinucleotide at bp-1 (top strand) enhances the binding affinity of Jun protein, whereas methylation at bp-2 (bottom strand) has a negligible effect.

Next we replaced the 5mC in the top strand of the oligo with all three oxidized forms of 5mC—5-hydroxymethylC (5hmC), 5-formylC (5fC) and 5-carboxylC (5caC)—while retaining 5mC in the bottom strand (Figure 1E). These chemical forms are generated *in vivo* by Tet dioxygenases in consecutive steps (40–42). All three oxidized forms reduced binding gradually, but K_D values were still 1.5–2.5× lower (higher affinities) than for unmodified cytosine. Overall, Jun homodimer appears to have a particularly low affinity for DNA with completely unmodified C in the first position.

Structural basis for the recognition of 5mC and T (5mU) by Jun

To understand how Jun binds meTRE (M:G) and TRE (T:A) sequences, we next co-crystallized Jun bZIP (residues 254–315) with a 18-bp duplex oligonucleotide containing the hemimethylated CpG (top strand) within the 7-bp meTRE element, plus a 5′-overhanging base on each strand (Figure 1F and G). The Jun-DNA complex structure was determined to a resolution of 1.89 Å (Supplementary Table S1). As expected, the structure is composed of two long helices, creating a clamp-like basic environment ideal for electrostatic attraction to the negatively charged DNA phosphate backbone (Figure 1F). The basic region of the N-terminal portion of the helix binds in the major groove of the DNA, recognizing half of the 7-bp meTRE element (Figure 1G). Among them, there are four G:C (or three G:C and one G:M) base pairs and three A:T base pairs. Surprisingly, except for the central G_4:C_4 base pair (see below), the remaining six base pairs do not involve the most conventional sequence-specific protein–DNA recognition, in which bidentate hydrogen-bond contacts link Gua-Arg and Ade-Asn (reviewed in (43); for example, see (44,45)).

From each monomer (colored in blue and magenta, respectively in Figure 1G), only four residues (Asn262, Ala265, Ala266 and Arg270) are involved in direct base contacts. The blue monomer interacts with the first three base pairs (Figure 1H–J). Ala265 is in van der Waals contact with the 5-position methyl group of 5mC at M_1. Asn262 forms a weak hydrogen bond with the exocyclic 4-amino group (NH_2) of 5mC, and forms a water-mediated hydrogen bond with the paired G_1 (Figure 1H). The second base pair, G_2:C_2, does not involve a direct H-bond with any side chain, instead forming water-mediated interactions with the main-chain carbonyl oxygen of Asn262 (Figure 1I), in agreement with the observation that cytosine methylation at C_2 does not influence the binding affinity (Figure 1D; M/C versus M/M or C/M versus C/C). The second alanine, Ala266, forms a corresponding van der Waals contact with the 5-position methyl group of T_3 (Figure 1J).

The central G_4:C_4 base pair does involve a classic pattern of bidentate H-bonds formed between Arg270 (in this case, from the magenta monomer) and the *N*7 and *O*6 atoms of G_4 (Figure 1K). The corresponding Arg270 of the blue monomer is in the neighborhood forming an electrostatic salt bridge with the backbone phosphate group and a water-mediated interaction with the paired C_4 (Figure 1K). If

the central G_4:C_4 base pair changes to C_4:G_4, the same Arg-Gua interaction could be maintained by the second monomer, in agreement with the ChIP-seq data in TNFα-stimulated triple-negative breast cancer cells showing that central G/C can be inverted without inhibiting Jun binding (17).

The magenta monomer interacts with the second half of the element, base pairs 7-5 in the 5′-3′ polarity of the TGA sequence (Figure 1L–N). The corresponding di-alanine, Ala265 and Ala266, forms conserved van der Waals contacts with the methyl groups of T_7 and T_5, respectively (Figure 1L and N). However, the corresponding magenta Asn262 adopts a different conformation. Rather than making indirect base contacts, as the blue Asn262 does (Figure 1H), the magenta Asn262 bridges two base pairs at positions 6 and 7, H-bonding both the *O*4 atom of T_7 and the *N*4 atom of C_6 (Figure 1O). While asparagine can act simultaneously as an H-bond donor (to *O*4 of T_7) and as an acceptor (from *N*4 of C_6), this bridging cannot occur for the first two base pairs because both are cytosines (Figure 1P). By rotating its side chain, however, Asn262 of the blue monomer accommodates the base pair at position 1 by forming a water mediated interaction involving M_1 and G_1 (Figure 1H).

Conserved T–Asn–C bridging conformation for binding 5′-TGA-3′ half sites

We compared our structure of Jun/Jun dimer-meTRE complex with the previously characterized structures of Jun/Fos heterodimer (PDB: 1FOS (19)) and Jun/Jun homodimer (PDB: 2H7H), each in complex with the unmodified TRE sequence. Regardless of the partner, the corresponding di-alanine-methyl group interactions and the T–Asn–C bridging conformation are conserved between Jun and Fos, each interacting with a 5′-TGA-3′ half site (Figure 2A). Superimposition of the two Jun/Jun homodimer complexes (PDB: 2H7H and our study) indicates the overall structure of the Jun homodimer is essentially unchanged between complexes with methylated (meTRE) or unmodified DNA (TRE), with a root-mean-squared deviation of <1 Å when comparing 124 pairs of Cα atoms (Figure 2B).

As the difference between TRE and meTRE is the first base pair, T:A in TRE and M:G in meTRE (inserted box in Figure 2B), the corresponding TGA half sites (base pairs 7-5 in the 5′-3′ polarity of the TGA sequence) are nearly identical between the two complexes (Figure 2C). However, the other half site (TGA in TRE and MGA in meTRE) involves rotation of the corresponding Asn side chain (Figure 2D). Of the base pairs involved in direct base–protein contacts, the first base pair has the largest conformational difference comparing M:G and T:A—a shift of ∼2 Å due to a lack of T–Asn–C bridging interaction involving blue subunit Asn262. Despite this shift, the Ala265-methyl distance is essentially unchanged (difference of ∼0.2 Å; Figure 2E). In summary, the change in Asn262 between M:G and T:A is consistent, whether comparing Jun/Jun complexed with TRE versus meTRE (Figure 2D and E; bp position 1), or within meTRE bp position 1 versus position 7 (Figure 1O and P).

Zta homodimer has equal affinity for meZRE1/2 but much reduced affinity for meZRE3

We next analyzed EBV Zta protein binding methylated ZRE elements. Zta binds the *BRLF1* promoter at three distinct sites (known as ZRE1-3) within the EBV genome (2,46–48) (Supplementary Figure S3). For each ZRE element, we generated a pair of duplex oligos either unmodified or methylated (Figure 3A). For ZRE1, it contains two overlapping non-CpG sites (CpC and CpA), and methylation at nucleotide position 5 of top strand (CpC) generated four methyl groups (including T methyl groups) symmetrically positioned at bp positions 1 and 5 of top stand, 3 and 7 of bottom strand (Figure 3A). [The effect of ZRE1 methylation at position 6 (CpA) is described below in the section on binding of asymmetric half sites.]

ZRE2 contains a more commonly methylated CpG site, and its methylation generates the symmetrically positioned four methyl groups, with an additional methyl group at bp-6 of the bottom strand (Figure 3A) that contributes less to the binding affinity (Supplemental Figure S4A). Zta binding affinities for meZRE1 and meZRE2 were almost identical, as were the unmodified ZRE1 and ZRE2, though in each case the methylated counterpart exhibited 5-fold stronger binding than the unmodified oligos (Figure 3B). As the methylated EBV genome can become 5-hydroxymethylated through the action of cellular TET dioxygenase (49), we also examined the effect of 5mC oxidation on binding affinity of meZRE2. As was seen for Jun protein, all three oxidized forms of 5mC within the meZRE2 sequence reduced Zta binding gradually in the order 5mC > 5hmC > 5fC > 5caC ∼C with unmodified C and 5caC showing the lowest affinity (Supplementary Figure S4B).

ZRE3 contains two CpG sites but deviates from ZRE1/2 sequences notably in lacking a TGA half site (Figure 3A). ZRE3 methylation (up to a total of seven methyl groups) does not result in the characteristic symmetric pattern of methyl groups as seen in meZRE1/2. In contrast to unmodified ZRE1 and ZRE2 (which contain three out of four spatially conserved methyl groups), Zta showed no detectable binding to unmodified ZRE3 (as shown previously (50)). Zta bound m2-ZRE3 (with two methylated cytosines only on the top strand), the best substrate of its methylated sequences (Supplementary Figure S4C) with far lower affinity than meZRE2 (by a factor of >10; Figure 3C). Among the four possible methylation sites of ZRE3, the methylation at nucleotide position 5 of the top strand, which restored one of the spatially conserved methyl groups, contributed most to the (weaker) binding affinity (Supplementary Figure S4C). Overall, these findings suggest that (i) the four methyl groups symmetrically positioned at positions 1, 3, 5 and 7 appear to be essential for Zta binding; and (ii) the sequence inversion at bp-6 between ZRE1 (C:G) and ZRE2 (G:C) had negligible effects, indicating that Zta homodimer can interact differently with the two half sites.

Structural analysis of meZRE2 binding by Zta

To understand the basis of Zta binding to two asymmetric half sites, and to explore the potential function of Ser186 (a residue unique to Zta relative to Jun), we determined the co-crystal structure of Zta in complex with meZRE2. The

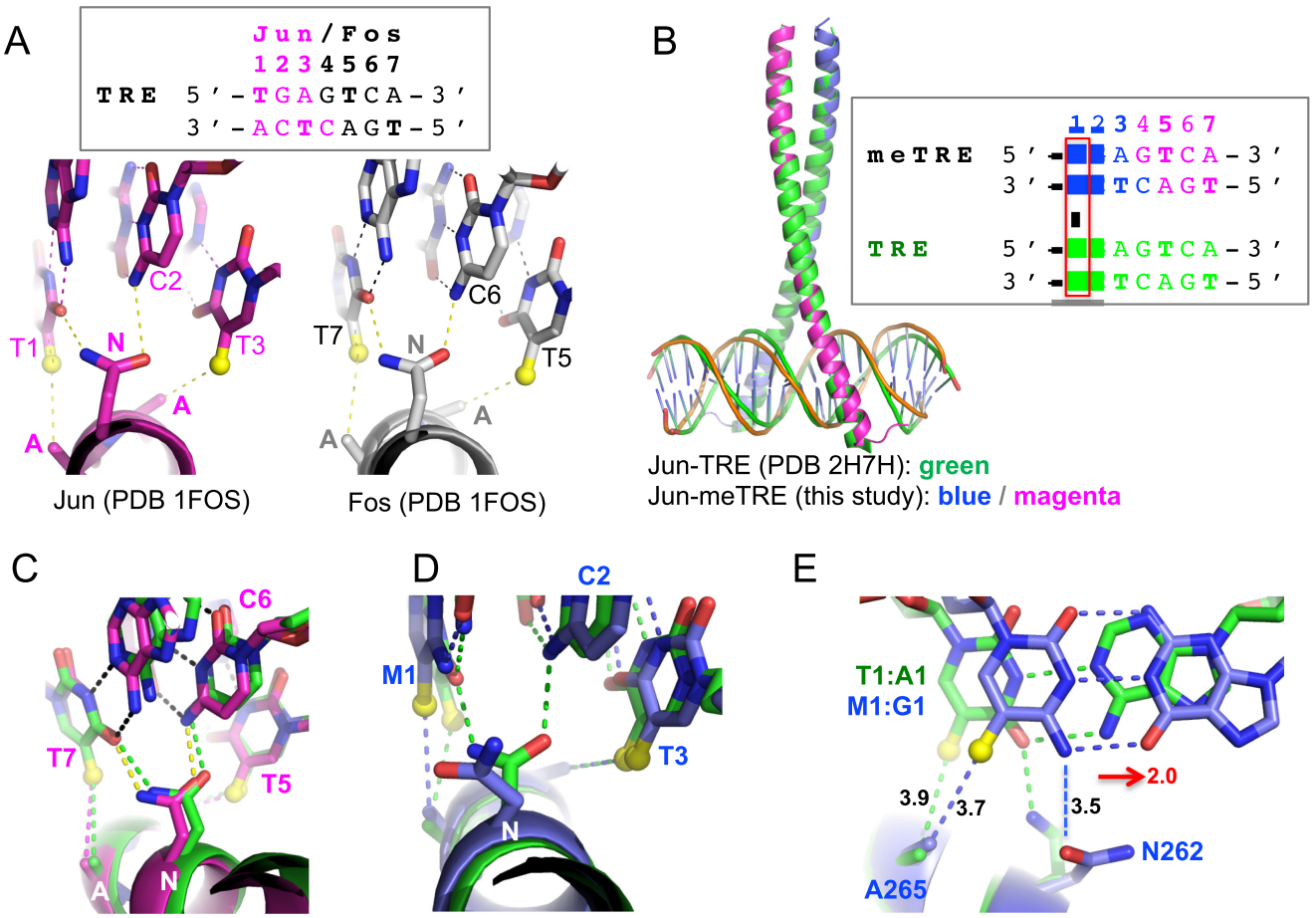

Figure 2. Comparison between TRE and meTRE binding by AP-1. (**A**) Conserved T_1-Asn-C_2 (left panel; 'N' is Asn) and T_7-Asn-C_6 (right panel) bridging interactions for binding TGA half sites in Jun/Fos heterodimer (PDB 1FOS). (**B**) Superimposition of Jun/Jun homodimer in complexes with TRE (colored in green) and meTRE (colored in magenta and blue for each monomer). (**C**) Comparison of the TGA half sites from TRE (green) and meTRE (magenta). (**D**) Comparison of the MGA half site in meTRE (blue) and the corresponding TGA half site in TRE (green). (**E**) A base pair shift observed for the first base pair between T:A in TRE (green) and M:G in meTRE (blue). Red arrow indicates the shift direction, and magnitude in angstroms.

complex was in space group C2, and was refined to a resolution of 2.3 Å (Supplementary Table S1). The half site at bp 1–3 contains a classical TGA, and involves the expected T_1-Asn182-C_2 interaction bridging the first two base pairs (Figure 3D and E). The methyl groups at T_1 and T_3 are in van der Waals contact with Ala185 and the side chain Cβ atom of Ser186 respectively (Figure 3D and F), analogous to the di-alanine of the Jun protein. In addition to the methyl group of T_3, Ser186 has two additional interactions via its hydroxyl oxygen atom: as an H-bond donor to *O*4 of T_3 (Figure 3F), and as an H-bond acceptor from the guanidinium moiety of Arg190 in the same monomer (colored magenta in Figure 3G and H). The corresponding Arg190 of the second monomer (blue) forms the classic bidentate H-bonds with the central guanine G_4 (Figure 3G).

The second meZRE2 half site has T_7–M_6–G_5 (Figure 3A, lower strand). The Zta complex shows the conserved methyl group interactions via Ala185 to the outer thymine base T_7 (Figure 3I), and from the Ser186 carbon Cβ atom to the inner M_5 base (Figure 3K). However, the interactions involving the side chains of Asn182 and Ser186 (hydroxyl oxygen atom) differ between the two half sites. The Ser186 hydroxyl oxygen atom is within H-bonding distance to the *N*4 atoms of both cytosine bases at bp 5–6 (Figure 3K, J and L). Asn182 adopts yet another different conformation by positioning its side chain within H-bonding distance of *N*7 and *O*6 of guanine G_6 (Figure 3J). This Asn–Gua interaction is unusual because neither *N*7 nor *O*6 of Gua are typically protonated, and they thus function normally as H-bond acceptors. The side chain torsion angle $\chi 2$ could rotate 180° to allow the Asn amide to form one H-bond to either the *N*7 or *O*6 of G_6. We believe that the Asn182–G_6 interaction observed here is not base specific, in agreement with the nearly identical affinity observed for ZRE1 and ZRE2 (Figure 3B). On the other hand, Asn is commonly used for adenine recognition (43) (see below). Nevertheless, the roles of Asn182 and Ser186 in the two subunits of the Zta homodimer interact differently with the two half sites. This binding adaptability stems in part from the abilities of Asn and Ser to act as an H-bond donor, acceptor, or both at the same time.

Structural comparison of Zta and Jun homodimers

To gain insights from the available structural information, we compared our Zta–meZRE2 complex structure to three

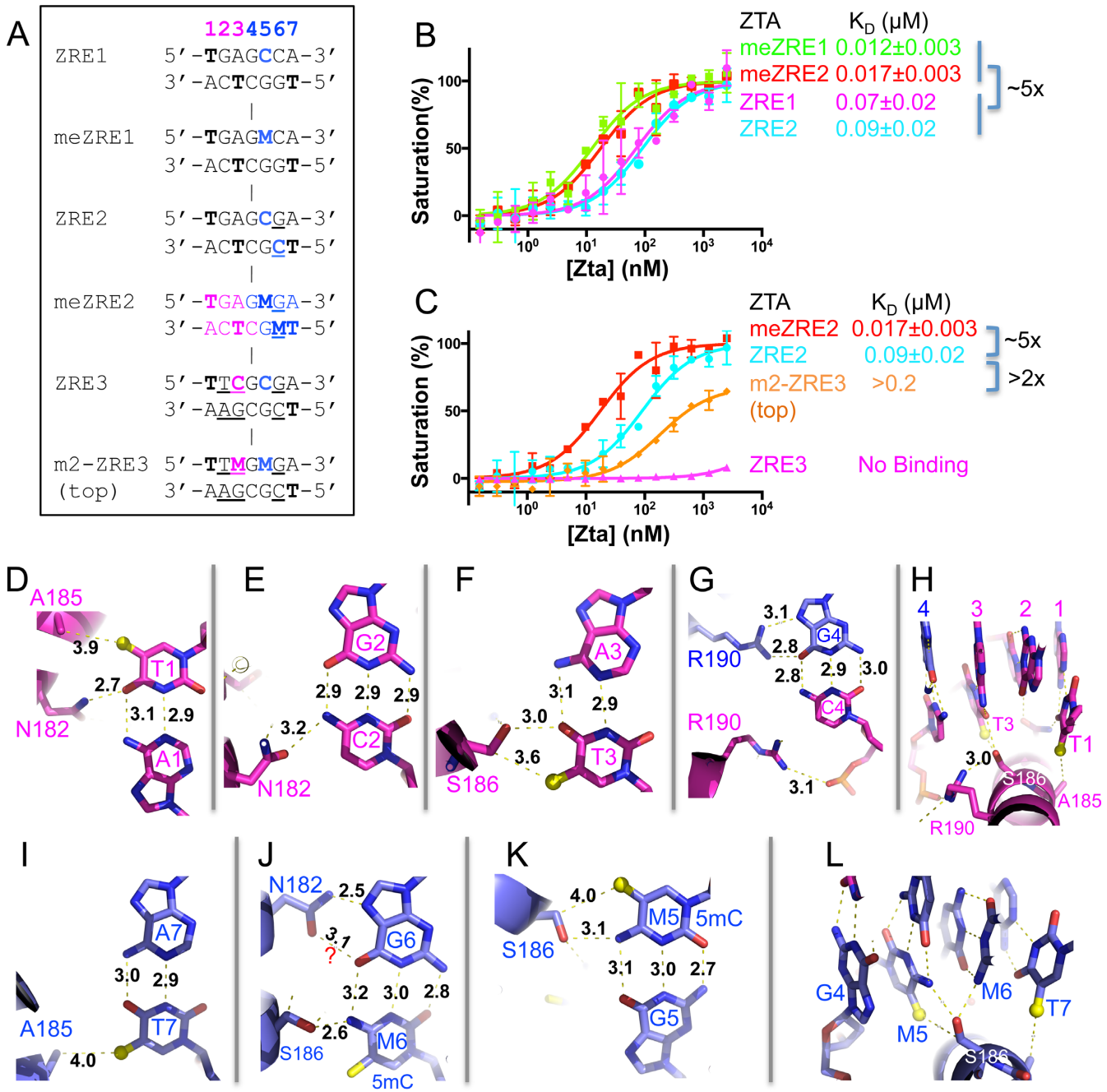

Figure 3. Zta/Zta-meZRE2 interactions. (**A**) For each ZRE element, a pair of duplex oligos either unmodified or methylated was synthesized. Sequence differences from that of ZRE1 are underlined. (**B** and **C**) DNA binding affinities of the Zta bZIP domain, with ZRE1 and ZRE2 (panel B) or ZRE2 and ZRE3 (panel C) oligos, each either methylated or unmodified. (**D**–**F**) Interactions with the first half site (base pairs 1–3) of meZRE2. The black numbers indicate the inter-atom distance in angstroms. (**G**) Interactions with the central G_4:C_4 base pair. (**H**) Ser186 interacts with Arg190 in the same monomer (colored magenta). (**I**–**K**) Interactions with the second half site of meZRE2 (base pairs 7-5). (**J**) The '?' indicates lacking a hydrogen atom between the two oxygen atoms required to form a hydrogen bond interaction. The 5mC methyl group of M_6 is colored yellow without a small sphere, to distinguish it from the spatially conserved methyl groups of M_5 and T_7. (**L**) The Ser186 hydroxyl oxygen atom is within H-bonding distance to the *N*4 atoms of both cytosine bases at positions 5–6 in meZRE2.

available structures: (i) Zta S186A mutant (mimicking AP-1) in complex with TRE (PDB 2C9L (30); Figure 4A), (ii) Jun homodimer in complex with TRE (PDB: 2H7H; Figure 4B) and (iii) Jun homodimer in complex with meTRE (Figure 4C; this study). The three pairwise comparisons revealed that the DNA-bound forms of the homodimers are highly similar in overall structures.

The difference between TRE and meZRE2 elements is two base pairs at positions 5 and 6 (Figure 4D). The superimposition of three structures of Zta, Zta S186A, and Jun revealed that substituting T:A with M:G at bp 5 had little effect (Figure 4E); while the base pair at position 6 swings towards the Zta wild-type protein but not the Zta S186A variant or Jun protein, presumably due to the Ser186-mediated H-bond pulling force (Figures 4F and G).

Superimposition of Zta-meZRE2 and Jun-meTRE complexes resulted in an alignment with three substitutions at base pairs 1, 5 and 6 (Figure 4H). A shift in the first base pair, comparing T:A in meZRE2 to M:G in meTRE (Figure 4I), is similar to what we observed for Jun proteins when comparing the binding to TRE versus meTRE (Figure 2E). While bp-5 does not move (Figure 4J), a similar swing motion is observed for bp-6 in the WT Zta complex (Figure 4K). The structural comparison suggests that (i) the shift in the first base pair, from T:A to M:G, is shared between the Jun–Jun and Jun–Zta comparisons (Figures 2E and 4I) and (ii) the swing motion in the sixth base pair is unique to WT Zta because of the pulling H-bond interaction by Ser186.

Difference between Jun and Zta in binding asymmetric half sites

Despite the observed interactions between Ser186/Asn182 and the base pair at position 6, and the different sequence at bp-6 between ZRE1 and ZRE2, Zta demonstrated nearly identical binding strength to the two oligos (Figure 3B). We asked whether Jun protein showed comparable binding flexibility. Surprisingly, Jun exhibited more than 20-fold weaker binding to meZRE2 than to TRE (Figure 5A), while Jun retained substantial affinity for meZRE1 (with K_D increased by 2-fold relative to TRE) (Figure 5C), with the two sequences differing only at bp-6 (C:G in meZRE1 and G:C in meZRE2) (Figure 5B), Like TRE, meZRE1 has a C:G at bp-6, so these results suggest that Jun requires two symmetric half sites—T/M-G-A, while Zta can accommodate symmetric as well as asymmetric half sites as evident by the nearly identical (<2-fold difference) binding affinities among TRE, meTRE, meZRE2 and meZRE1 (Figures 5D and 3B).

We suspect the ability of Zta to bind asymmetric sites probably results from the asymmetric interactions involving Ser186 (Figure 3H and L). We thus generated two reciprocal mutants, Jun Ala266-to-Ser (A266S) and Zta Ser186-to-Ala (S186A) (Supplementary Figure S5). As expected, Jun A266S now bound meZRE2 (Figure 5E), with similar affinity to that of Jun wild-type protein binding to meZRE1. Conversely, Zta S186A now bound meZRE2 (and its unmethylated counterpart) with affinity lowered by a factor of >12 (Figure 5F and Supplementary Figure S5F).

The consensus-binding element for Zta as determined by ChIP-seq suggested that an adenine could substitute at nucleotide position 6 (5′-TGA**G**-T/C-C/**A**-A-3′) (51). This raised the question of whether Asn182—which contacts G_6 in meZRE2 (Figure 3J)—could favorably interact with adenine. Asn–Ade interactions occur, for example, with Asn117 of EcoRI (52), Asn185 of EcoRV (53), Asn140 of PvuII (54), and Asn789 of human PRDM9 (45). We thus generated a modified sequence replacing the CpG in ZRE2 with CpA, as well as generating a pair of corresponding methylated sequences (Figure 5G). As expected, Zta binds the CpA sequence with 2–3-fold higher affinity than to the CpG sequence (unmodified), and binds fully methylated CpG (both strands) and methylated CpA sites equally well (Figure 5H). Analysis of the genome-wide occupancy of Zta in the Epstein-Barr virus genome (55) indicates that Zta occupies at least 17 of 21 CpA ZRE sites in the early lytic cycle (Supplementary Figure S3B and Table S2) [For comparison, Zta occupies 13 out of 15 ZRE1 sites, 10 out of 17 ZRE2 sites, and 1 out of 3 ZRE3 sites in the EBV genome].

As noted earlier that ZRE1 contains two overlapping non-CpG sites (CpC and CpA) at nucleotide positions 5 and 6 of the 7-bp element's top strand, while ZRE2 contains a CpG dinucleotide in the corresponding positions with the two cytosines located on opposite strands (Figure 5I). If both cytosines are methylated, meZRE1 will have two 5mCs on the top strand (m2ZRE1 in Figure 5I) while meZRE2 has 5mCs on the two opposite strands—the only difference occurs at position 6 (Figure 5I). We expanded the analysis of variation at position 6 to include all five possibilities: G, A, T, C and 5mC. Zta showed approximately equal binding affinity to C (meZRE1), G (meZRE2) and A (5mCpA) at position 6; ∼3-fold weaker binding to 5mC; and ∼6.5-fold weaker binding to thymine (5mU) (Figure 5J). It seems that a methyl group on the top strand at nt-6 decreased binding, while the methyl group on the bottom strand did not affect binding affinity (meZRE2).

We repeated the same experiments with Jun protein against the set of ZRE1/2-based oligos (Figure 5K). Jun showed the strongest binding with cytosine at position 6 (the symmetric sequence); ∼1.5-fold weaker binding to adenine (possible recognition by the conserved Asn262); ∼3.5-fold weaker to 5mC; and >10-fold weaker binding to guanine and thymine. Thus, the affinity of Jun for nt-6 position is C > A > 5mC >> G and T. The largest difference between Zta and Jun is that Jun protein exhibits particular discrimination against guanine at position 6 (both proteins discriminate against thymine at that position).

Difference between Jun and Zta in binding central G:C base pair

As described above, the Ser186 residue of the second Zta monomer (magenta in Figure 3H) interacts with Arg190 from the same monomer (Figure 6A). The same interaction would not occur in the S186A mutant and in Jun protein (where the corresponding residue is an Ala). We asked whether Ser186–Arg190 interaction has any functional consequence, particularly whether their interaction could prevent Arg190 from interacting with the central bp when G_4:C_4 is inverted to C_4:G_4 (see Figure 3G). We thus generated a mutant meZRE2 sequence with the central G:C substituted with C:G (Figure 6B). Quite unexpectedly, Zta

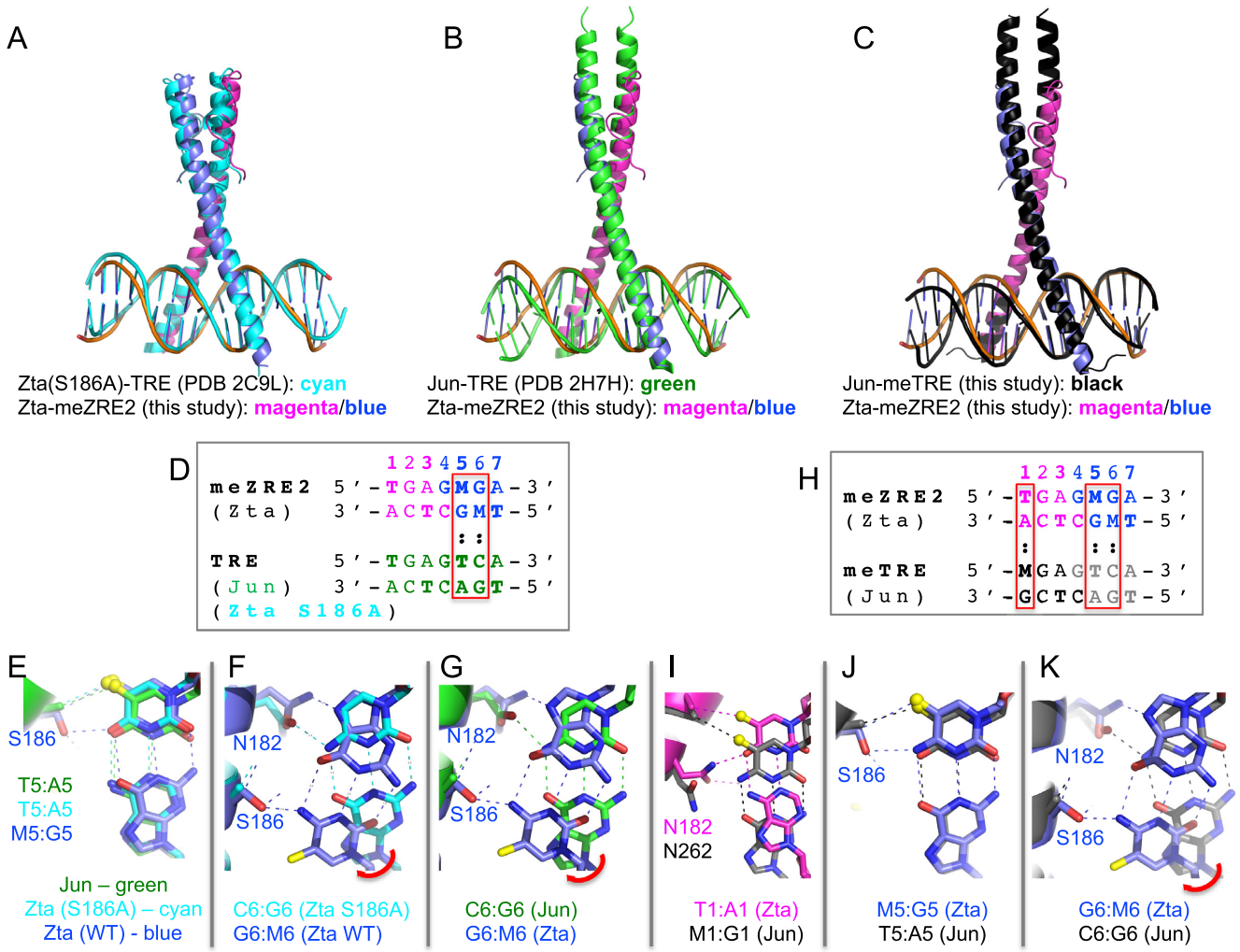

Figure 4. Structural comparison between Jun and Zta in complex with DNA. (**A**–**C**) Superimposition of Zta-meZRE2 complex onto Zta (S186A) mutant-TRE (panel A), onto Jun-TRE (panel B) and onto Jun-meTRE (panel C). (**D**) DNA sequence difference (in red box) between meZRE2 and TRE. (**E**) Superimposition of bp-5 in Jun-TRE, Zta (S186A)-TRE, and Zta-meZRE2. (**F**) Superimposition of bp-6 in Zta (S186A)-TRE and Zta-meZRE2. The 5mC methyl group of M_6 in panels F, G and K is colored yellow without a small sphere, to distinguish it from the spatially conserved methyl groups of T_1 and T_5 (or M_5). (**G**) Superimposition of bp-6 in Jun-TRE and Zta-meZRE2. (**H**) DNA sequence difference (in red boxes) between meZRE2 and meTRE. (**I** and **K**) Superimposition, for bp-1 (panel I), bp-5 (panel J) and bp-6 (panel K), of Jun-meTRE onto Zta-meZRE2.

wild-type protein showed reduced binding for the mutated sequence, yielding the same affinity as for the unmodified normal ZRE2 sequence (Figure 6C), while the Zta S186A (like Jun protein) binds mutant and normal meZRE2 sequences equally (Figure 6D). The opposite effect was seen comparing Jun wild type and A266S (Zta-like) proteins (Figure 6E and F). We note that Jun proteins bind the meZRE2 sequence only weakly under laboratory conditions, and Zta S186A (Jun-like) also has much reduced overall affinity (Figure 6D–F). Nevertheless, our *in vitro* binding data agree with the genome-wide screen of viral genes (28) and an anti-Zta ChIP-seq study across the human genome (51), both showing that the central G:C base pair (5′-TGA**G**-T/C-C/A-A-3′) is not variable in Zta binding sites.

DISCUSSION

We have focused on the role of the AP-1 di-alanine (and Zta Ala–Ser pair) in recognizing methylated DNA bases. The Ala265–Ala266 of Jun shows extraordinarily high conservation. We used NCBI BLink to identify Jun orthologs, even including hits with overall match scores of just 10% the maximum (self) match. While some of the lower-scoring records were missing the segment with the AA, there were few cases among full-length orthologs of a substitution in the di-alanine (Supplementary Figure S1B). Even the *Drosophila melanogaster* ortholog (NP_476586.1, 32% identity to human Jun) has the conserved Ala–Ala, presumably to recognize T given that *D. melanogaster* DNA has little or no 5mC (56). Among Zta orthologs the Ala–Ser shows similar conservation, and is present in even the rhesus and marmoset gamma-Herpesvirus proteins (respectively AAK95436.1, 71% identity to Zta; and NP_733896.1, 34% identity to Zta). We did find other variations among mam-

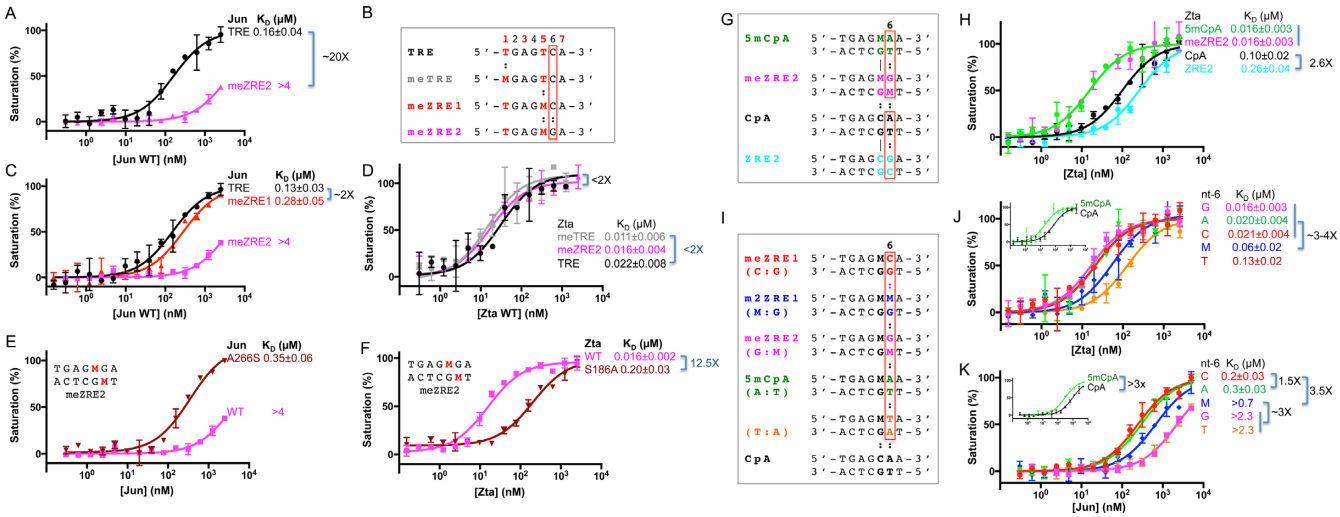

Figure 5. Difference between Jun and Zta in binding asymmetric half sites. (**A**) Jun protein binding affinity for TRE (symmetric) *vs.* meZRE2 (an asymmetric sequence). (**B**) Sequence difference at bp-6 among TRE, meTRE, meZRE1 and meZRE2. (**C**) Jun protein binding affinity for TRE and meZRE1 (both are symmetric sequences). (**D**) Similar Zta binding to three sequences (TRE, meTRE1 and meZRE2). (**E**) Jun A266S change increases binding to meZRE2 (an asymmetric sequence). (**F**) Zta S186A change reduces binding to meZRE2. (**G**) CpG and CpA containing meZREs. (**H**) Zta has increased binding on CpA sequence compared to the CpG sequence (ZRE2), and binds fully methylated CpG (both strands) and methylated CpA sequences equally well. (**I**) ZRE-based oligos with variation at bp-6. (**J**) Zta binds in approximately equal affinity to C, G and A at position 6. (**K**) Jun discriminates against a guanine at position 6. In panels J and K, inserted are controls for the binding affinities between 5mCpA and CpA oligos.

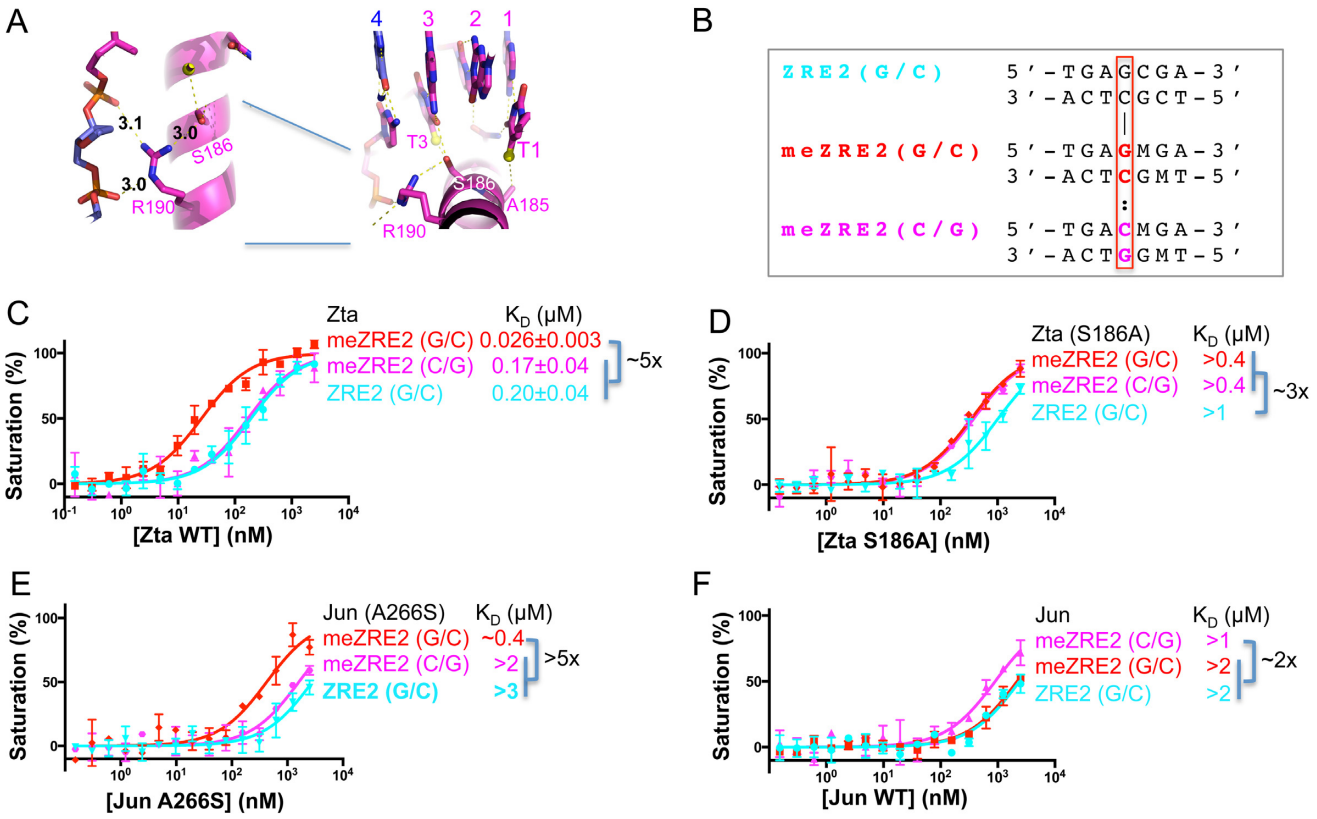

Figure 6. Difference between Jun and Zta in binding central G:C base pair. (**A**) Ser186–Arg190 interaction in one monomer of Zta. (**B**) DNA oligos with altered sequence of central base pair (G:C to C:G). (**C**) Altered central base pair (G:C to C:G) reduces Zta binding. (**D**) Zta (S186A) binding is insensitive to inversion of the central base pair. (**E**) Jun (A266S) has increased binding of meZRE2 (G:C). (**F**) Jun only weakly binds meZRE2.

malian bZIP transcription factors (Supplementary Figure S6A).

As mentioned above, the EBV virion genome becomes heavily methylated during the latent stage of the virus cycle by the cellular DNA methylation machinery (24–26), and the methyl groups are then subject to oxidation by cellular TET dioxygenases (49). Oxidized forms of 5mC within the meZRE2 sequence reduced Zta binding, with unmodified C and 5caC showing the lowest affinity (Supplementary Figure S4B). Thus, in the presence of TET, methylation of the virion EBV DNA would lead to a transient pulse of high-affinity Zta binding, that could be renewed only upon methylation of newly replicated DNA.

The *de novo* DNA methyltransferase(s) generate 5mC usually within the dinucleotide sequence context of CpG (57–59) or non-CpG (particularly CpA) (60–65). Here, we show that Zta protein binds with equal affinity meZRE sequences containing methylated CpG, CpA or CpC, at bp positions 5 and 6 within the 7-bp element (Figures 3B, 5H and J). The ability of Ser186 to act as an H-bond donor, acceptor, or both at the same time allows Zta dimer to bind both symmetric (meZRE1) and asymmetric half sites (meZRE2 and CpA meZRE). In contrast, the Zta S186A mutant and WT Jun protein—both of which have an alanine in the corresponding position and lacking the flexibility—preferentially bind symmetric sequences (TRE, meTRE and meZRE1) with a cytosine at position 6, followed by an adenine at that position. The largest difference between Zta and Jun is that Jun protein exhibits particular discrimination against meZRE2 with a guanine at position 6. This observation suggests that the inability of Zta S186A to drive gene expression from the latent viral genome (66) might be due to the loss of binding to asymmetric meZRE2 in the *BRLF1* promoter (Figure 5F). Conversely, the gain-of-function mutant associated with AP-1 alanine-to-serine (A266S in Jun), which does drive viral lytic gene expression (29) might result from a gained ability to bind asymmetric meZRE2 (Figure 5E).

Besides binding TRE elements (TGA-G-TCA), AP-1 Jun or Fos proteins can form heterodimers with ATF to recognize cAMP-response elements (CRE; TGA-CG-TCA) (reviewed in (14)), which are also recognized by CRE-binding (CREB) transcription factors (67,68). The difference between TRE and CRE elements is the one bp expansion in the central C:G bp to a CpG dinucleotide (Supplementary Figure S6B), which is recognized by the two symmetrically positioned Arg residues from each monomer (Supplementary Figure S6C). The structure and sequence conservation implies a similar pattern of DNA recognition among the extended family of bZIP transcriptional factors, such as the conserved di-alanine-methyl interactions and conserved T–Asn–C bridging conformation for the TGA half sites (Supplementary Figure S6C). This conservation suggests that the CRE-binding proteins (CREB and AP-1/ATF heterodimers) might be sensitive to DNA methylation status if C replaces either T. For example, ChIP-seq data of CREB in hippocampal neurons identified a non-canonical CRE motif (69), where the inner A:T at bp-3 position is replaced by G:C and the methylation of the cytosine at bp-3 would restore the symmetrically positioned methyl groups. In addition, the expanded central CpG dinucleotide potentially allows the CpG methylation/oxidation status to play a regulatory role (e.g. see MAX (70)).

The ability to bind either unmodified (T-containing) or modified (M-containing) elements may contribute to the diverse regulatory mechanisms of bZIP-mediated gene expression. It is interesting that spontaneous deamination of 5mC to yield thymine has long been recognized as providing hotspots for pathological mutation in the human genome (71–73). However, targeted enzymatically driven deamination of 5mC to T in human DNA (74) may provide an important regulatory mechanism. Specifically, a particular DNA sequence, outside of the germline, would effectively become permanently methylated in a given cell lineage. As sequence databases grow, it will be interesting to look for examples of lineage-specific non-germline 5mC-to-T transitions in regulatory regions (75). Though our knowledge is currently limited, the ability to predict transcription factor sensitivity to DNA modifications is becoming increasingly important (76).

ACCESSION NUMBERS

The X-ray structures (coordinates and structure factor files) of Jun bZIP domain with bound meTRE and Zta bZIP domain with bound meZRE2 have been submitted to PDB under accession number 5T01 and 5SZX, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank B. Baker of New England Biolabs for synthesizing the oligonucleotides. The Department of Biochemistry of Emory University School of Medicine supported the use of SER-CAT beamlines at the Advanced Photon Source, Argonne National Laboratory; Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science.

Author contributions: S.H. performed protein expression and purification, some DNA binding experiments, and crystallographic work; D.W. performed some DNA binding experiments; J.R.H. participated in X-ray data collection; X.Z. participated in discussion throughout and assisted in organizing and designing the study; S.H.S. provided Zta construct; R.M.B. performed data analysis and assisted in preparing the manuscript; X.C. organized and designed the scope of the study. All were involved in analyzing data and in preparing the manuscript.

FUNDING

National Institutes of Health (NIH) [GM049245-23 to X.C.]; Georgia Research Alliance Eminent Scholar [to X.C.]. The open access publication charge for this paper has been waived by Oxford University Press – *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

Conflict of interest statement. None declared.

REFERENCES

1. Bird,A.P. (1984) DNA methylation versus gene expression. *J. Embryol. Exp. Morphol.*, **83**(Suppl), 31–40.
2. Bhende,P.M., Seaman,W.T., Delecluse,H.J. and Kenney,S.C. (2004) The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat. Genet.*, **36**, 1099–1104.
3. Sasai,N., Nakao,M. and Defossez,P.A. (2010) Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res.*, **38**, 5015–5022.
4. 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.
5. Quenneville,S., Verde,G., Corsinotti,A., Kapopoulou,A., Jakobsson,J., Offner,S., Baglivo,I., Pedone,P.V., Grimaldi,G., Riccio,A. *et al.* (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell*, **44**, 361–372.
6. Liu,Y., Toh,H., Sasaki,H., Zhang,X. and Cheng,X. (2012) An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes Dev.*, **26**, 2374–2379.
7. Buck-Koehntop,B.A., Stanfield,R.L., Ekiert,D.C., Martinez-Yamout,M.A., Dyson,H.J., Wilson,I.A. and Wright,P.E. (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 15229–15234.
8. Spruijt,C.G., Gnerlich,F., Smits,A.H., Pfaffeneder,T., Jansen,P.W., Bauer,C., Munzel,M., Wagner,M., Muller,M., Khan,F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
9. Hu,S., Wan,J., Su,Y., Song,Q., Zeng,Y., Nguyen,H.N., Shin,J., Cox,E., Rho,H.S., Woodard,C. *et al.* (2013) DNA methylation presents distinct binding sites for human transcription factors. *eLife*, **2**, e00726.
10. Liu,Y., Olanrewaju,Y.O., Zheng,Y., Hashimoto,H., Blumenthal,R.M., Zhang,X. and Cheng,X. (2014) Structural basis for Klf4 recognition of methylated DNA. *Nucleic Acids Res.*, **42**, 4859–4867.
11. Hashimoto,H., Olanrewaju,Y.O., Zheng,Y., Wilson,G.G., Zhang,X. and Cheng,X. (2014) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.*, **28**, 2304–2313.
12. Spruijt,C.G. and Vermeulen,M. (2014) DNA methylation: old dog, new tricks? *Nat. Struct. Mol. Biol.*, **21**, 949–954.
13. Karin,M., Liu,Z. and Zandi,E. (1997) AP-1 function and regulation. *Curr. Opin. Cell Biol.*, **9**, 240–246.
14. Eferl,R. and Wagner,E.F. (2003) AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer*, **3**, 859–868.
15. Uluckan,O., Guinea-Viniegra,J., Jimenez,M. and Wagner,E.F. (2015) Signalling in inflammatory skin disease by AP-1 (Fos/Jun). *Clin. Exp. Rheumatol.*, **33**, S44–S49.
16. Schiefer,A.I., Vesely,P., Hassler,M.R., Egger,G. and Kenner,L. (2015) The role of AP-1 and epigenetics in ALCL. *Front. Biosci.*, **7**, 226–235.
17. Qiao,Y., He,H., Jonsson,P., Sinha,I., Zhao,C. and Dahlman-Wright,K. (2016) AP-1 is a key regulator of proinflammatory cytokine TNFalpha-mediated triple-negative breast cancer progression. *J. Biol. Chem.*, **291**, 5068–5079.
18. Papoudou-Bai,A., Hatzimichael,E., Barbouti,A. and Kanavaros,P. (2016) Expression patterns of the activator protein-1 (AP-1) family members in lymphoid neoplasms. *Clin. Exp. Med.*, doi:10.1007/s10238-016-0436-z.
19. Glover,J.N. and Harrison,S.C. (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature*, **373**, 257–261.
20. Tulchinsky,E.M., Georgiev,G.P. and Lukanidin,E.M. (1996) Novel AP-1 binding site created by DNA-methylation. *Oncogene*, **12**, 1737–1745.
21. Gustems,M., Woellmer,A., Rothbauer,U., Eck,S.H., Wieland,T., Lutter,D. and Hammerschmidt,W. (2014) c-Jun/c-Fos heterodimers regulate cellular genes via a newly identified class of methylated DNA sequence motifs. *Nucleic Acids Res.*, **42**, 3059–3072.
22. Chiu,Y.F. and Sugden,B. (2016) Epstein-Barr Virus: The Path from Latent to Productive Infection. *Annu. Rev. Virol.*, **3**, 359–372.
23. Sugden,B. (2014) Epstein-Barr virus: the path from association to causality for a ubiquitous human pathogen. *PLoS Biol.*, **12**, e1001939.
24. Paulson,E.J. and Speck,S.H. (1999) Differential methylation of Epstein-Barr virus latency promoters facilitates viral persistence in healthy seropositive individuals. *J. Virol.*, **73**, 9959–9968.
25. Fernandez,A.F., Rosales,C., Lopez-Nieva,P., Grana,O., Ballestar,E., Ropero,S., Espada,J., Melo,S.A., Lujambio,A., Fraga,M.F. *et al.* (2009) The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. *Genome Res.*, **19**, 438–451.
26. Kenney,S.C. and Mertz,J.E. (2014) Regulation of the latent-lytic switch in Epstein-Barr virus. *Semin. Cancer Biol.*, **26**, 60–68.
27. Farrell,P.J., Rowe,D.T., Rooney,C.M. and Kouzarides,T. (1989) Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos. *EMBO J.*, **8**, 127–132.
28. Bergbauer,M., Kalla,M., Schmeinck,A., Gobel,C., Rothbauer,U., Eck,S., Benet-Pages,A., Strom,T.M. and Hammerschmidt,W. (2010) CpG-methylation regulates a class of Epstein-Barr virus promoters. *PLoS Pathogens*, **6**, e1001114.
29. Yu,K.P., Heston,L., Park,R., Ding,Z., Wang'ondu,R., Delecluse,H.J. and Miller,G. (2013) Latency of Epstein-Barr virus is disrupted by gain-of-function mutant cellular AP-1 proteins that preferentially bind methylated DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 8176–8181.
30. Petosa,C., Morand,P., Baudin,F., Moulin,M., Artero,J.B. and Muller,C.W. (2006) Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. *Mol. Cell*, **21**, 565–572.
31. Francis,A., Ragoczy,T., Gradoville,L., Heston,L., El-Guindy,A., Endo,Y. and Miller,G. (1999) Amino acid substitutions reveal distinct functions of serine 186 of the ZEBRA protein in activation of early lytic cycle genes and synergy with the Epstein-Barr virus R transactivator. *J. Virol.*, **73**, 4543–4551.
32. Bhende,P.M., Seaman,W.T., Delecluse,H.J. and Kenney,S.C. (2005) BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186. *J. Virol.*, **79**, 7338–7348.
33. Lan,F., Collins,R.E., De Cegli,R., Alpatov,R., Horton,J.R., Shi,X., Gozani,O., Cheng,X. and Shi,Y. (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature*, **448**, 718–722.
34. Abate,C., Patel,L., Rauscher,F.J. 3rd and Curran,T. (1990) Redox regulation of fos and jun DNA-binding activity in vitro. *Science*, **249**, 1157–1161.
35. Wang,P., Day,L., Dheekollu,J. and Lieberman,P.M. (2005) A redox-sensitive cysteine in Zta is required for Epstein-Barr virus lytic cycle DNA replication. *J. Virol.*, **79**, 13298–13309.
36. Hashimoto,H., Liu,Y., Upadhyay,A.K., Chang,Y., Howerton,S.B., Vertino,P.M., Zhang,X. and Cheng,X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.*, **40**, 4841–4849.
37. Motulsky,H. and Christopoulos,A. (2004) *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. Oxford University Press, NY.
38. Otwinowski,Z., Borek,D., Majewski,W. and Minor,W. (2003) Multiparametric scaling of diffraction intensities. *Acta Crystallogr. A*, **59**, 228–234.
39. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
40. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
41. He,Y.F., Li,B.Z., Li,Z., Liu,P., Wang,Y., Tang,Q., Ding,J., Jia,Y., Chen,Z., Li,L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
42. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.

43. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
44. Vanamee,E.S., Viadiu,H., Kucera,R., Dorner,L., Picone,S., Schildkraut,I. and Aggarwal,A.K. (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J.*, **24**, 4198–4208.
45. Patel,A., Horton,J.R., Wilson,G.G., Zhang,X. and Cheng,X. (2016) Structural basis for human PRDM9 action at recombination hot spots. *Genes Dev.*, **30**, 257–265.
46. Packham,G., Economou,A., Rooney,C.M., Rowe,D.T. and Farrell,P.J. (1990) Structure and function of the Epstein-Barr virus BZLF1 protein. *J. Virol.*, **64**, 2110–2116.
47. Adamson,A.L. and Kenney,S.C. (1998) Rescue of the Epstein-Barr virus BZLF1 mutant, Z(S186A), early gene activation defect by the BRLF1 gene product. *Virology*, **251**, 187–197.
48. Ragoczy,T. and Miller,G. (2001) Autostimulation of the Epstein-Barr virus BRLF1 promoter is mediated through consensus Sp1 and Sp3 binding sites. *J. Virol.*, **75**, 5240–5251.
49. Wille,C.K., Nawandar,D.M., Henning,A.N., Ma,S., Oetting,K.M., Lee,D., Lambert,P., Johannsen,E.C. and Kenney,S.C. (2015) 5-hydroxymethylation of the EBV genome regulates the latent to lytic switch. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E7257–E7265.
50. Karlsson,Q.H., Schelcher,C., Verrall,E., Petosa,C. and Sinclair,A.J. (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 Pathogens*, **4**, e1000005.
51. Ramasubramanyan,S., Osborn,K., Al-Mohammad,R., Naranjo Perez-Fernandez,I.B., Zuo,J., Balan,N., Godfrey,A., Patel,H., Peters,G., Rowe,M. *et al.* (2015) Epstein-Barr virus transcription factor Zta acts through distal regulatory elements to directly control cellular gene expression. *Nucleic Acids Res.*, **43**, 3563–3577.
52. Kim,Y.C., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.
53. Winkler,F.K., Banner,D.W., Oefner,C., Tsernoglou,D., Brown,R.S., Heathman,S.P., Bryan,R.K., Martin,P.D., Petratos,K. and Wilson,K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.
54. Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) Structure of PvuII endonuclease with cognate DNA. *EMBO J.*, **13**, 3927–3935.
55. Ramasubramanyan,S., Kanhere,A., Osborn,K., Flower,K., Jenner,R.G. and Sinclair,A.J. (2012) 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.*, **86**, 12494–12502.
56. Raddatz,G., Guzzardo,P.M., Olova,N., Fantappie,M.R., Rampp,M., Schaefer,M., Reik,W., Hannon,G.J. and Lyko,F. (2013) Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 8627–8631.
57. Bestor,T., Laudano,A., Mattaliano,R. and Ingram,V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.*, **203**, 971–983.
58. Okano,M., Xie,S. and Li,E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.*, **19**, 219–220.
59. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
60. Ramsahoye,B.H., Biniszkiewicz,D., Lyko,F., Clark,V., Bird,A.P. and Jaenisch,R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 5237–5242.
61. Gowher,H. and Jeltsch,A. (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J. Mol. Biol.*, **309**, 1201–1208.
62. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
63. Lister,R., Mukamel,E.A., Nery,J.R., Urich,M., Puddifoot,C.A., Johnson,N.D., Lucero,J., Huang,Y., Dwork,A.J., Schultz,M.D. *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**, 1237905.
64. Kubo,N., Toh,H., Shirane,K., Shirakawa,T., Kobayashi,H., Sato,T., Sone,H., Sato,Y., Tomizawa,S., Tsurusaki,Y. *et al.* (2015) DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. *BMC Genomics*, **16**, 624.
65. Vlachogiannis,G., Niederhuth,C.E., Tuna,S., Stathopoulou,A., Viiri,K., de Rooij,D.G., Jenner,R.G., Schmitz,R.J. and Ooi,S.K. (2015) The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis. *Cell Rep.*, **10**, 944–956.
66. Francis,A.L., Gradoville,L. and Miller,G. (1997) Alteration of a single serine in the basic domain of the Epstein-Barr virus ZEBRA protein separates its functions of transcriptional activation and disruption of latency. *J. Virol.*, **71**, 3054–3061.
67. Montminy,M.R., Sevarino,K.A., Wagner,J.A., Mandel,G. and Goodman,R.H. (1986) Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 6682–6686.
68. Schumacher,M.A., Goodman,R.H. and Brennan,R.G. (2000) The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J. Biol. Chem.*, **275**, 35242–35247.
69. Lesiak,A., Pelz,C., Ando,H., Zhu,M., Davare,M., Lambert,T.J., Hansen,K.F., Obrietan,K., Appleyard,S.M., Impey,S. *et al.* (2013) A genome-wide screen of CREB occupancy identifies the RhoA inhibitors Par6C and Rnd3 as regulators of BDNF-induced synaptogenesis. *PLoS One*, **8**, e64658.
70. Wang,D., Hashimoto,H., Zhang,X., Barwick,B.G., Lonial,S., Boise,L.H., Vertino,P.M. and Cheng,X. (2016) MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.*, doi:10.1093/nar/gkw1184.
71. Youssoufian,H., Kazazian,H.H. Jr, Phillips,D.G., Aronis,S., Tsiftis,G., Brown,V.A. and Antonarakis,S.E. (1986) Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. *Nature*, **324**, 380–382.
72. Cooper,D.N. and Youssoufian,H. (1988) The CpG dinucleotide and human genetic disease. *Hum. Genet.*, **78**, 151–155.
73. Cooper,D.N., Mort,M., Stenson,P.D., Ball,E.V. and Chuzhanova,N.A. (2010) Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum. Genomics*, **4**, 406–410.
74. Wijesinghe,P. and Bhagwat,A.S. (2012) Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res.*, **40**, 9206–9217.
75. He,X., Tillo,D., Vierstra,J., Syed,K.S., Deng,C., Ray,G.J., Stamatoyannopoulos,J., FitzGerald,P.C. and Vinson,C. (2015) Methylated cytosines mutate to transcription factor binding sites that drive tetrapod evolution. *Genome Biol. Evol.*, **7**, 3155–3169.
76. Medvedeva,Y.A., Khamis,A.M., Kulakovskiy,I.V., Ba-Alawi,W., Bhuyan,M.S., Kawaji,H., Lassmann,T., Harbers,M., Forrest,A.R., Bajic,V.B. *et al.* (2014) Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*, **15**, 119.