Structural basis of DNA methylation-dependent site selectivity of the Epstein–Barr virus lytic switch protein ZEBRA/Zta/BZLF1

Florent Bernaudat[1,2,†], Montse Gustems[3,†], Johannes Günther[4,5], Mizar F. Oliva[1,6], Alexander Buschle [3], Christine Göbel[3], Priscilla Pagniez[1], Julien Lupo[1,7], Luca Signor[1], Christoph W. Müller [8], Patrice Morand[1,7], Michael Sattler [4,5], Wolfgang Hammerschmidt [3,*] and Carlo Petosa [1,*]

[1]Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), 38000 Grenoble, France, [2]European Synchrotron Radiation Facility, 71 avenue des Martyrs, 38043 Grenoble, France, [3]Research Unit Gene Vectors, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany and German Centre for Infection Research (DZIF), Partner site Munich, D-81377 Germany, [4]Institute of Structural Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany, [5]Bavarian NMR Center and Department of Chemistry, Technical University of Munich, 85748 Gaching, Germany, [6]Institut Laue-Langevin, 71 avenue des Martyrs, 38042 Cedex 9 Grenoble, France, [7]Laboratoire de Virologie, Centre Hospitalier Universitaire Grenoble Alpes, 38000 Grenoble, France and [8]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), D-69117 Heidelberg, Germany

ABSTRACT

In infected cells, Epstein–Barr virus (EBV) alternates between latency and lytic replication. The viral bZIP transcription factor ZEBRA (Zta, BZLF1) regulates this cycle by binding to two classes of ZEBRA response elements (ZREs): CpG-free motifs resembling the consensus AP-1 site recognized by cellular bZIP proteins and CpG-containing motifs that are selectively bound by ZEBRA upon cytosine methylation. We report structural and mutational analysis of ZEBRA bound to a CpG-methylated ZRE (meZRE) from a viral lytic promoter. ZEBRA recognizes the CpG methylation marks through a ZEBRA-specific serine and a methylcytosine-arginine-guanine triad resembling that found in canonical methyl-CpG binding proteins. ZEBRA preferentially binds the meZRE over the AP-1 site but mutating the ZEBRA-specific serine to alanine inverts this selectivity and abrogates viral replication. Our findings elucidate a DNA methylation-dependent switch in ZEBRA's transactivation function that enables ZEBRA to bind AP-1 sites and promote viral latency early during infection and subsequently, under appropriate conditions, to trigger EBV lytic replication by binding meZREs.

INTRODUCTION

DNA methylation in mammals is a major epigenetic modification that primarily occurs at the cytosine C5 position within CpG motifs (1). DNA methylation is commonly perceived as a repressive epigenetic mark that induces transcriptional silencing. Silencing is mediated by methyl-CpG binding proteins (MBPs) that inhibit the action of RNA polymerase II or lead to a restrictive chromatin state (2,3). In addition, CpG methylation can directly inhibit the binding of transcription factors to their DNA target sites (4,5). On the contrary, the expression of certain genes may be enhanced by DNA methylation, and a growing number of transcription factors are known to display a preference for methylated target sequences (6–13). The first of these to be identified was the EBV protein ZEBRA (also called BZLF1, Zta, Z or EB1) (6).

DNA methylation plays a pivotal role in the EBV infection cycle (14). EBV is a gamma herpesvirus that infects >90% of the world population, can cause Infectious Mononucleosis in adolescents and young adults and is associated with several epithelial and B-cell malignancies (15). EBV primarily infects B lymphocytes and has a biphasic infection cycle that alternates between latency and lytic replication (16). Upon infection, during a stage termed prelatency (17), EBV delivers its linear genomic DNA to the host cell nucleus, where multiple copies of the viral genome

---

are maintained as plasmids (also termed episomes) that are initially unmethylated. During prelatency the viral genome becomes chromatinized, histones acquire post-translational modifications, and the viral DNA becomes progressively methylated at CpG motifs (18). These changes allow EBV to repress the expression of immunodominant viral antigens and establish a strictly latent infection in memory B lymphocytes, thereby evading host immune surveillance (19). Antigen-mediated stimulation of the B-cell receptor signaling pathway can reactivate the virus in plasma cells *in vivo* (20) and induce a cascade of immediate-early, early and late lytic gene expression, leading to viral *de novo* synthesis and release of progeny. Evidence suggests that EBV lytic replication contributes to lymphomagenesis (21–23).

ZEBRA is a homodimeric protein related to the activating protein 1 (AP-1) family of bZIP transcription factors (24). ZEBRA regulates the EBV infection cycle by playing key roles both in establishing viral latency and triggering lytic replication. The transient expression of ZEBRA during prelatency when the EBV genome is unmethylated is critical for promoting the proliferation of quiescent naive and memory B cells that favors the transition to strict latency (18). During latency, when the EBV genome is methylated, the expression of ZEBRA activates a second viral transcription factor, Rta, which acts together with ZEBRA to trigger lytic replication (25,26). Underpinning ZEBRA's dual role in prelatency and lytic activation is its ability to recognize two distinct classes of DNA target sites, collectively termed ZEBRA responsive elements (ZREs) (27,28) (Figure 1A). One class comprises viral and cellular sites resembling the AP-1 consensus sequence TGAGTCA [also called TPA responsive element (TRE) (29)] recognized by cellular AP-1 proteins (24,30–32). The binding of ZEBRA to cellular AP-1 sites during prelatency is critical for promoting the proliferation of EBV-infected resting B cells (18,33). The second class comprises CpG-containing sites with the consensus TGAGCGA, which ZEBRA selectively binds when methylated. Many lytic EBV promoters have CpG-containing ZREs whose binding by ZEBRA is methylation dependent, including the Rta promoter (Rp) (6,18,27–28,28,34–36). Moreover, ZEBRA behaves like a pioneer transcription factor (PTF) that can directly bind nucleosomal DNA, recruit chromatin remodelers and enhance the local accessibility of chromatin (37). Thus, whereas host-driven methylation of the EBV genome ordinarily represses viral gene expression, ZEBRA's PTF-like behavior and ability to activate CpG-methylated viral lytic promoters allow it to overturn host-mediated epigenetic silencing.

ZEBRA's ability to preferentially bind and activate methylated viral promoters is attributed to a serine residue in its bZIP domain, Ser[186], that differs conspicuously from the corresponding alanine conserved in cellular bZIP proteins (34). Alanine substitution of Ser[186] compromises ZEBRA's ability to bind methylated ZREs and activate viral early lytic genes (34,38–42). Conversely, Ala→Ser mutations of the corresponding residues enabled a heterodimer of the cellular AP-1 proteins Fos and Jun to acquire these activities (41). The crystal structure of ZEBRA's DNA-binding domain in complex with the consensus AP-1 site revealed the unusual bZIP fold of this domain and the details of AP-1 site recognition (43). A subsequent structural study revealed how ZEBRA achieves methylated ZRE (meZRE) recognition, including a direct contact between Ser[186] and the [m]CpG motif, and related these findings to how a Jun homodimer recognizes a methylated AP-1 site (42). Recently, an additional basic motif upstream of the bZIP domain was shown to enhance the affinity of ZEBRA for a meZRE and to be required for late viral lytic gene expression (44).

Despite important advances in our understanding of ZEBRA structure and function, the molecular basis and functional implications of ZEBRA's dual specificity for AP-1 and CpG-methylated sites remain poorly understood. Here, we analyze the structure of ZEBRA bound to a CpG-methylated ZRE in a detailed comparison with the AP-1-bound structure. We investigate how the integrity of interactions that mediate [m]CpG recognition correlate with ZEBRA's functions to transactivate a methylated promoter and to activate viral lytic replication. We found that several distinct DNA-contacting residues are required for both these activities. Surprisingly, most of these residues, in particular Ser[186], contribute only to a limited degree to ZEBRA's ability to discriminate between the methylated and unmethylated forms of a CpG-containing ZRE. In striking contrast, Ser[186] is the critical determinant for ZEBRA's preference to bind a CpG-methylated ZRE versus an AP-1 sequence motif. Our data document that ZEBRA's DNA-binding interface is fine-tuned to discriminate between its two classes of ZRE target sites rather than simply to differentiate methylated from unmethylated DNA. More generally, our findings reveal how ZEBRA is capable of switching between AP-1 and CpG-methylated sites, enabling EBV's biphasic life style to establish latency upon infection and to escape from it, eventually.

MATERIALS AND METHODS

Protein expression and purification

For structure determination we used a ZEBRA construct spanning residues 175–236, which lacks the 9 C-terminal residues (res. 237–245) previously shown to reduce solubility (43) and replaced residue Cys[189] by a serine. The C189S mutation does not alter ZEBRA's binding affinity toward the meZRE2 site (Figure 5B and Table 1) but was critical for obtaining well-diffracting crystals. A pET28a vector expressing this ZEBRA construct was used to transform *Escherichia coli* strain BL21 (DE3). Cells were grown in LB medium containing kanamycin (60 μg/ml) until an OD_{600} of 0.8 was reached. Expression was induced with 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) and cells further incubated at 37°C for 5 h. Harvested cells were lysed by sonication in lysis buffer [10 mM NaCl, 20 mM TRIS/HCl pH 6.8, 5 mM β-mercaptoethanol, 1 mM PMSF, cOmplete EDTA-free (Boehringer, 1 tablet/50 ml)]. Nucleic acids were removed from the cleared lysate by polyethyleneimine (PEI) precipitation (0.3% v/v). The protein was purified by SP Sepharose chromatography (GE Healthcare) in lysis buffer using a 0.01–1 M NaCl gradient, followed by ammonium sulfate precipitation (30% w/v) and Superdex 75 chromatography (GE Healthcare) in 20 mM TRIS/HCl (pH 7.5), 150 mM NaCl, 150 mM ammonium acetate, 5 mM DTT and 0.2 mM PMSF.
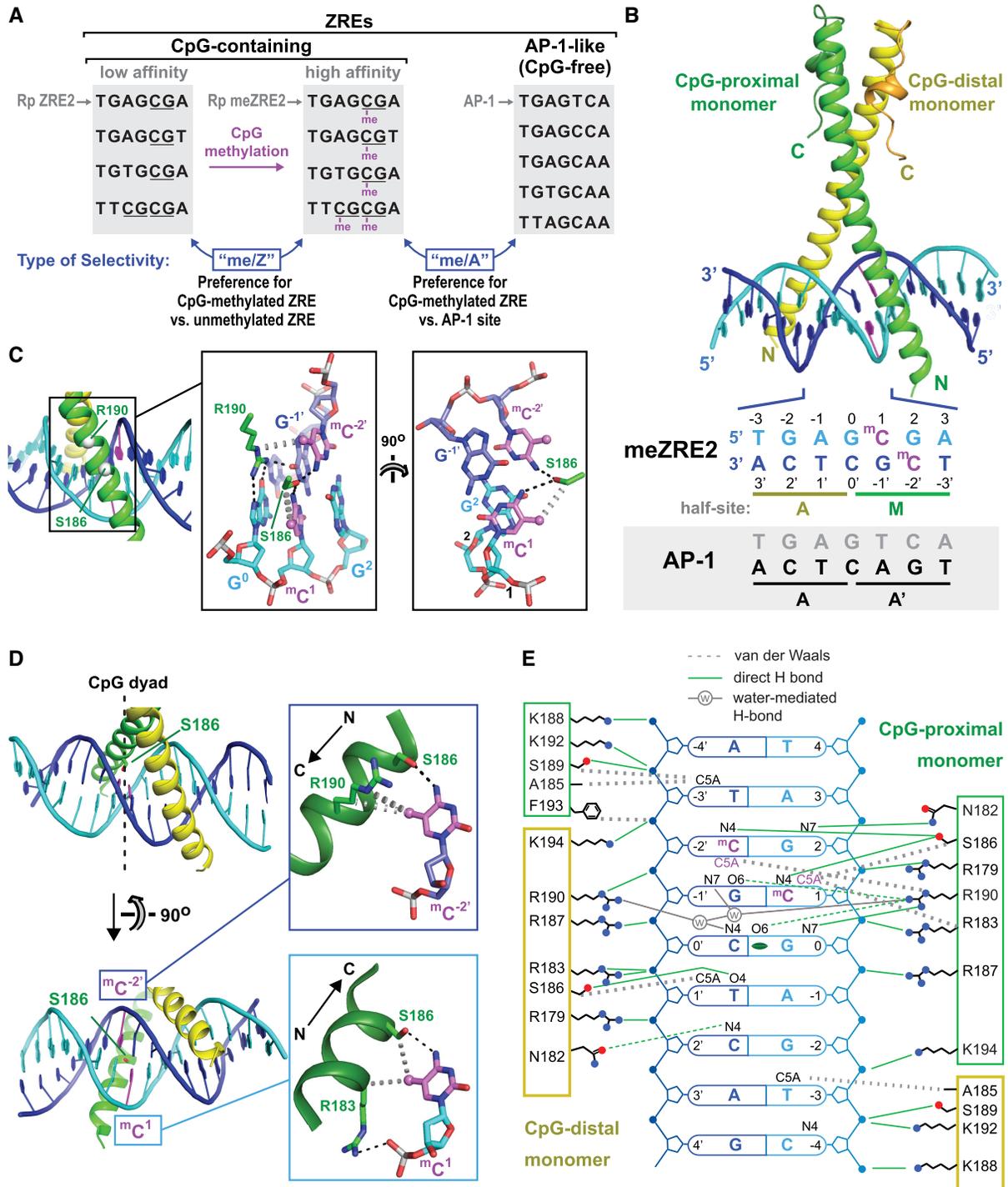
Figure 1. Structure of ZEBRA bound to meZRE2. (**A**) Representative ZRE sites. CpG-containing sites are (from top to bottom): Rp ZRE2, Nap ZRE2, BMRF1(−248) and Rp ZRE3 (6,28,35). AP-1-like sites are AP-1, Zp ZREIIIA, DSL ZRE7, DSL ZRE5 and Zp ZREIIIB (24,30-32). (**B**) Structure of ZEBRA's DNA-binding domain in complex with meZRE2. Methylcytosine bases are in magenta. The AP-1 sequence is included for comparison. The A half-site, which includes G[0], is common to both the AP-1 and meZRE2 sites. Note that the structure closely resembles that of the ZEBRA/meZRE2 complex previously reported by Hong *et al.* (42). (**C**) Details of the [m]CpG site showing interactions involving Ser[186] and Arg[190]. Black and gray dashed lines indicate H-bonds and van der Waals contacts, respectively. (**D**) The Ser[186] side chain is positioned directly on the local dyad relating the two antisense [m]CpG dinucleotides. Top inset: ZEBRA recognizes the [m]C[−2′] methyl group through van der Waals contacts with the Arg[190] side chain. Bottom inset: ZEBRA senses the methyl group on [m]C[1] through van der Waals contacts with the Arg[183] backbone Cα atom and the Ser[186] side chain methylene group. (**E**) Schematic summary of protein–DNA interactions. Residues in the CpG-distal and -proximal monomers are enclosed in yellow and green boxes, respectively. Green and gray lines represent direct and water-mediated H-bonds, respectively. A weak H-bond formed by CpG-distal Asn[182] and bifurcated H-bonds formed by CpG-proximal Arg[190] are shown as broken green lines. Dotted gray lines indicate van der Waals contacts. Protein side chains are illustrated with oxygen and nitrogen atoms shown as red and blue dots, respectively.

Table 1. Summary of binding data

			Methylated versus unmethylated sites			
Protein	DNA	$K_{d, app}$ (nM)	Selectivity relative to ZRE2[1,2]	ΔG_{app} (kcal/mol)[1]	$\Delta\Delta G_{app}$ versus ZRE2 ($\Delta\Delta G_{me/Z}$ values in bold) (kcal/mol)[1,3]	$\Delta\Delta G_{me/A}$ (kcal/mol)[1,4]
ZEBRA WT	ZRE2	1710 ± 285		-7.86 ± 0.10		
	hemi($C^{-2'}$)	667 ± 147	2.6 ± 1.0	-8.42 ± 0.14	-0.56 ± 0.24	
	hemi(C^1)	167 ± 34	10.2 ± 3.8	-9.24 ± 0.13	-1.38 ± 0.23	
	meZRE2	75.1 ± 19.8	22.8 ± 9.8	-9.71 ± 0.17	**-1.85 ± 0.27**	-0.76 ± 0.23
	AP-1	271 ± 25	6.3 ± 1.6	-8.95 ± 0.06	-1.09 ± 0.16	
S186T	ZRE2	2990 ± 208		-7.53 ± 0.04		
	meZRE2	121 ± 44	24.7 ± 10.7	-9.43 ± 0.26	**-1.90 ± 0.30**	
S186A	ZRE2	5210 ± 1010		-7.20 ± 0.12		
	meZRE2	434 ± 28	12.0 ± 3.1	-8.68 ± 0.04	**-1.48 ± 0.16**	0.69 ± 0.17
	AP-1	136 ± 29	38.3 ± 15.6	-9.36 ± 0.13	-2.16 ± 0.25	
S186C	ZRE2	7730 ± 1650		-6.97 ± 0.14		
	meZRE2	1250 ± 264	6.2 ± 2.6	-8.05 ± 0.13	**-1.08 ± 0.27**	
C189S	ZRE2	1330 ± 190		-8.01 ± 0.09		
	meZRE2	87.3 ± 17.5	15.2 ± 5.2	-9.62 ± 0.13	**-1.61 ± 0.22**	
C189A	ZRE2	2360 ± 453		-7.67 ± 0.12		
	meZRE2	118 ± 19	20.0 ± 7.1	-9.45 ± 0.10	**-1.78 ± 0.22**	
C189T	ZRE2	2470 ± 258		-7.65 ± 0.06		
	meZRE2	171 ± 22	14.4 ± 3.4	-9.23 ± 0.08	**-1.58 ± 0.14**	
N182A	ZRE2	1850 ± 313		-7.82 ± 0.10		
	meZRE2	85.3 ± 14.5	21.7 ± 7.4	-9.64 ± 0.11	**-1.82 ± 0.21**	
R183A	ZRE2	2390 ± 434		-7.67 ± 0.11		
	meZRE2	162 ± 34	14.8 ± 5.8	-9.26 ± 0.13	**-1.59 ± 0.24**	
R190A	ZRE2	5350 ± 841		-7.19 ± 0.10		
	meZRE2	1310 ± 159	4.1 ± 1.1	-8.02 ± 0.07	**-0.83 ± 0.17**	
GCN4 WT	ZRE2	3940 ± 1080		-7.37 ± 0.18		
	meZRE2	479 ± 41	8.2 ± 3.0	-8.62 ± 0.05	**-1.25 ± 0.23**	1.08 ± 0.12
	AP-1	77.9 ± 8.8	50.6 ± 19.5	-9.69 ± 0.07	-2.32 ± 0.25	
A239S	ZRE2	1600 ± 460		-7.90 ± 0.19		
	meZRE2	98.0 ± 15.6	16.3 ± 7.3	-9.56 ± 0.10	**-1.66 ± 0.29**	-0.32 ± 0.18
	AP-1	169 ± 22	9.5 ± 4.0	-9.23 ± 0.08	-1.33 ± 0.27	

			Hydroxymethylated versus methylated ZRE2			
Protein	DNA	$K_{d, app}$ (nM)	Selectivity of meZRE2 relative to site[5]	ΔG_{app} (kcal/mol)	$\Delta\Delta G_{app}$ versus meZRE2 (kcal/mol)[6]	
ZEBRA	meZRE2	69 ± 13		-9.76 ± 0.12		
	hemi($^{hm}C^1$)	113 ± 14	1.6 ± 0.5	-9.47 ± 0.08	0.29 ± 0.19	
	hemi($^{hm}C^{-2'}$)	114 ± 18	1.6 ± 0.6	-9.47 ± 0.10	0.30 ± 0.21	
	hmZRE2	216 ± 23	3.1 ± 0.9	-9.09 ± 0.06	0.68 ± 0.18	

[1] Data shown represent the mean values $\pm$ SD from three independent experiments.

[2] Calculated as K_1/K_2, where K_1 is the $K_{d, app}$ for ZRE2 and K_2 is that for the compared site. The error was determined as $\varepsilon = (\varepsilon_1/K_1 + \varepsilon_2/K_2)*(K_1/K_2)$, where ε_1 and ε_2 are the errors associated with K_1 and K_2.

[3] Calculated as $\Delta G_2 - \Delta G_1$, where ΔG_1 is the ΔG_{app} value for ZRE2 and ΔG_2 is that of the compared site. The error was determined as $\varepsilon = \varepsilon_1 + \varepsilon_2$, where ε_1 and ε_2 are the errors associated with ΔG_1 and ΔG_2, respectively. $\Delta\Delta G_{app}$ is identical to $\Delta\Delta G_{me/Z}$ when the compared site is meZRE2.

[4] Calculated as $\Delta G_2 - \Delta G_1$, where ΔG_1 is the ΔG_{app} value for AP-1 and ΔG_2 is that for meZRE2. The error was determined as $\varepsilon = \varepsilon_1 + \varepsilon_2$, where ε_1 and ε_2 are the errors associated with ΔG_1 and ΔG_2, respectively.

[5] Calculated as K_1/K_2, where K_2 is the $K_{d, app}$ for meZRE2 and K_1 is that for the compared site. The error was determined as $\varepsilon = (\varepsilon_1/K_1 + \varepsilon_2/K_2)*(K_1/K_2)$, where ε_1 and ε_2 are the errors associated with K_1 and K_2.

[6] Calculated as $\Delta G_2 - \Delta G_1$, where ΔG_2 is the ΔG_{app} value for meZRE2 and ΔG_1 is that of the compared site.

MBP-tagged constructs of the DNA-binding domains of ZEBRA (res. 175–236) and *Saccharomyces cerevisiae* GCN4 (res. 228–281) used for FP assays were expressed from a pET-M40 vector in *E. coli* strain BL21-CodonPlus (DE3)-RIL. Cells were sonicated in lysis buffer (100 mM NaCl, 20 mM TRIS/HCl pH 7.2, 5 mM β-mercaptoethanol, 1 mM PMSF, EDTA-free cOmplete inhibitor [Boehringer] 1 tablet/50 ml). The cleared lysate was incubated with amylose resin (New England Biolabs) pre-equilibrated in lysis buffer and extensively washed with high-salt buffer (1.5 M NaCl, 20 mM TRIS/HCl pH 7.2, 5 mM β-mercaptoethanol). MBP-tagged proteins were eluted with high-salt buffer containing 10 mM maltose and further purified by Superdex 200 10/300 chromatography (GE Healthcare) in phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4).

Crystallization and structure determination

DNA oligonucleotides used for crystallization (5′-AAGCACTGAG(^{m}C)GATGAAGT-3′ and 5′-TACTTCAT(^{m}C)GCTCAGTGCT-3′) were chemically synthesized (Eurofins MWG) and subsequently purified by anion-exchange chromatography using a monoQ HR10/10 (GE Healthcare) column pre-equilibrated in 10 mM NaOH. Oligonucleotides were eluted using a linear NaCl gradient (0–1 M), dialysed against deionized water, lyophilized and subsequently dissolved in deionized water. Equimolar amounts of complementary oligonucleotides were mixed at a concentration of 10 mg/ml in 0.1 M NaCl, 10 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 mM DTT and annealed in a water bath by cooling from 90 to 20°C over several hours. Hanging drop crystallization trials were carried out at 20°C by mixing equal volumes of reservoir solution and an equimolar protein:DNA mixture. Crystals grew from 22% polyethylene glycol (PEG) 4K, 18% PEG 400, 50 mM sodium acetate pH 4 and 20% isopropanol.

Diffraction data were collected from crystals flash cooled in liquid nitrogen at ESRF beamline ID23-2 ($\lambda = 0.873$ Å) on a MAR CCD 165 mm detector. Data were processed with XDS (45) and programs of the CCP4 suite (46). Molecular replacement was performed with Phaser (47) and the structure was refined with Phenix (48). Crystals contain two protein/DNA complexes in the asymmetric unit. The electron density is well defined for Complex 1 (chains A-D) but considerably poorer for Complex 2 (chains E-H), which exhibits high B factors and two-fold disorder around the DNA pseudodyad due to a lack of stabilizing crystallization contacts, explaining why R_{cryst} and R_{free} values are higher than those normally expected at this resolution. The accuracy of the structure is supported by stereochemical quality criteria (Supplementary Table S1), a high correlation coefficient (CC) with the local electron density for most residues (overall CC is 0.92 for complex 1 and 0.88 for complex 2) and low RMSD values with previously reported ZEBRA structures (Supplementary Figure S1). DNA geometry was analysed using the program 3DNA (49).

Fluorescence polarization (FP) DNA-binding assay

The following pairs of oligonucleotides were chemically synthesized (Eurofins MWG) for FP assays involving (i)

the AP-1 site: 5′-AATAAAATGACTCATAAGC-3′ and Rho-5′-AGCTTATGAGTCATTTTAT-3′ and (ii) the unmethylated, hemi-methylated and fully methylated ZRE2 sites: 5′-AATAAAATXGCTCATAAGC-3′ and Rho-5′-AGCTTATGAGXGATTTTAT-3′ where X represents either C or ^{m}C and Rho represents the rhodamine label. Complementary oligonucleotides were dissolved in 0.15 M NaCl, 10 mM TRIS/HCl pH 7.5, 1 mM EDTA and annealed in a PCR machine. MBP-ZEBRA was serially diluted in phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) containing rhodamine-labeled duplex DNA (10 nM) and unlabeled herring testes Type XIV DNA (28 ng/μl) (SIGMA D6898) and incubated in a volume of 40 μl for 30 min in a 384-well plate. Fluorescence polarization was measured at 20°C using a SYNERGY 4 plate reader (BioTek). Excitation and emission wavelengths were 530 and 580 nm, respectively, and the slit width was 5 nm in both cases. Between two and four independent experiments (three technical replicates per experiment) were performed for each protein/DNA combination. Data were fitted as $FP = FP_{min} + (FP_{max} - FP_{min}) * c^n / (c^n + K_{d,app}{}^n)$, where FP_{min} and FP_{max} are the lower baseline and upper plateau values of FP, and c is the total protein concentration. The Hill coefficient, n, was set at 2, consistent with empirical values of n derived from Hill plots that varied between 1.5 and 2.5 and in agreement with previous DNA-binding studies of bZIP proteins performed in the presence of non-specific competitor DNA (50,51). Binding curves for assays involving unmethylated ZRE2 where saturation was not fully attained could be reliably fitted because of highly reproducible values of FP_{max} across the ensemble of assays, and in several cases the results were confirmed by performing single-replicate experiments using higher protein concentrations.

LC/ESI mass spectrometry

Liquid chromatography electrospray ionization mass spectrometry (LC/ESI-MS) was performed on a 6210 LC-TOF spectrometer coupled to a HPLC system (Agilent Technologies). All solvents used were HPLC grade (Chromasolv, Sigma-Aldrich), trifluoroacetic acid (TFA) was from Acros Organics (puriss., p.a.). Solvent A was 0.03% TFA in water; solvent B was 95% acetonitrile-5% water-0.03% TFA. Just before analysis, MBP-ZEBRA samples (10 μM in phosphate-buffered saline: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) containing 0 or 20 mM DTT were diluted to a final concentration of 5 μM with water and 4 μl were injected for MS analysis. Protein samples were first desalted on a reverse phase-C8 cartridge (Zorbax 300SB-C8, 5 μm, 300 μm ID × 5 mm, Agilent Technologies) for 3 min at a flow rate of 50 μl/min with 100% solvent A and then eluted with 70% solvent B at flow rate of 50 μl/min for MS detection. MS acquisition was carried out in the positive ion mode in the 300–3200 m/z range. MS spectra were acquired and the data processed with MassHunter workstation software (v. B.02.00, Agilent Technologies) and GPMAW software (v. 7.00b2, Lighthouse Data, Denmark).

NMR

DNA oligonucleotides were purchased (Eurofins MWG) and dissolved in water. DNA duplexes were prepared by mixing both strands in equimolar amounts. For this, DNA was heated to 95°C for 5 min and then slowly cooled at RT for at least 30 min. After lyophilization, the DNA was reconstituted in a buffer containing 50 mM sodium phosphate pH 6.5, 100 mM NaCl, and 10% D_2O with a final concentration of 400 μM duplex DNA. Homonuclear $^1H,^1H$ NOESY experiments using water-flipback combined with WATERGATE for solvent suppression were carried out at 293K on a Bruker 950 MHz spectrometer equipped with z-gradient triple resonance cryoprobe. Spectra were processed using TopSpin (Bruker) and analyzed using the CCPN software suite (52).

Cells

HEK293 and Raji cells were maintained in RPMI 1640 medium with 10% fetal calf serum (FCS), 1% penicillin–streptomycin and 1% sodium pyruvate at 37°C and 5% CO_2. The ZEBRA knockout producer cell line 6169 contains an EBV genome based on the wt/B95.8 strain termed r_wt/B95.8 (6008) (53) with a stop codon after amino acid 56 of ZEBRA. The ZEBRA knockout EBV producer cells were maintained in RPMI 1640 medium with 10% FCS, 1% penicillin–streptomycin, 1% sodium pyruvate and puromycin 500 ng/ml at 37°C and 5% CO_2.

Plasmids

The DNA binding and dimerization domain of ZEBRA (residues 149–245) was cloned downstream of the tandem StrepII/FLAG-tag (54) to yield the plasmid p3928. The ZEBRA expression plasmid p509 is described elsewhere (55). All the plasmids encoding ZEBRA mutants were generated by introducing point mutations into plasmid p509. The luciferase plasmid p4376 was constructed by inserting a pentamer of a 24 bp long oligonucleotide (GGTGCTCATGAGCGAGGGCCAGAT, ZRE2 is underlined) into a basic luciferase reporter plasmid with a minimal EF1a promoter. The entire plasmid backbone of this reporter plasmid is free of CpGs (56). The plasmid p2670 is described elsewhere (57).

DNA transfection

Transfection of DNA into HEK293 and ZEBRA knockout cells was performed using PEI max (Polysciences). During the preparation of the transfection mixture, cells were switched to Optimem minimal medium (Invitrogen). The DNAs were mixed with 0.3 ml (for six-well plate) or 6 ml (for 130 mm dish) Optimem and then 6 μl PEI (1 mg/ml in water) were added per μg DNA. The mixture was incubated for 15 min at room temperature and was added to the cells for 4–5 h. Then the transfection medium was replaced by standard medium.

For the protein extracts used in EMSAs, 1×10^7 HEK293 cells per 130-mm dish were seeded the day before transfection. Each plate was transfected with 30 μg of plasmid DNA. For Western blot analysis, 8×10^5 HEK293 cells were seeded into 6-well plates the day before transfection and 0.5 μg of plasmid DNA were transfected per well. For reporter assays, 8×10^5 HEK293 cells were seeded into 6-well plates the day before transfection. Each well was cotransfected with 1 μg of reporter plasmid together with 5 ng of transactivator and 50 ng of DNA of a renilla-expressing plasmid as an internal control for data normalization. For EBV production, 8×10^5 ZEBRA knockout cells were seeded into 6-well plates the day before transfection. Each well was transfected with 0.5 μg of ZEBRA expressing plasmid (p509 encoding wt ZEBRA or ZEBRA mutant derivatives based on p509), and 0.5 μg of p2670 plasmid DNA (57) and supernatants with EBV particles were harvested three days after DNA transfection.

Electromobility shift assays

Electromobility shift assays (EMSAs) were performed with purified protein from HEK293 cells transiently transfected with Strep/FLAG:ZEBRA (p3928). Protein purification was performed as previously described (9). The oligos ZRE2for (ATAGCTTATGAGCGATTTTATC), meZRE2for (ATAGCTTATGAG^{m}CGATTTTATC), ZRE2rev (ATGATAAAATCGCTCATAAGCT), meZRE2rev (ATGATAAAAT^{m}CGCTCATAAGCT), ZREfor (ATAGCTTATGTGCAATTTTATC) and ZRErev (ATGATAAAATTGCACATAAGCT) containing the ZRE2 and ZRE5 from the BRLF1 promoter and the oriLyt, respectively, were used. EMSAs were performed as described previously (27).

Protein lysates from transiently transfected 293T cells and western blot immunostaining

To compare the steady-state protein expression of ZEBRA and ZEBRA mutants, plasmid DNAs of expression plasmids encoding ZEBRA and its nine single amino acid mutants were chemically transfected into 293T cells using polyethyleneimine. Three days after DNA transfection the cells were collected, centrifuged and washed in cold PBS and were resuspended in RIPA lysis buffer (50 mM TRIS, 150 mM NaCl, 1% NP40, 0.5% DOC, 0.1% SDS, pH 8.0) complemented with protease and phosphatase inhibitors. Cell lysates were frozen at −80°C. After thawing on ice, the lysates were mixed and centrifuged at 13 000 rpm for 10 min at 4°C. Supernatants were collected and the protein amount was determined using the Pierce BCA Protein Assay (Thermo Scientific). Protein concentrations of the lysates were adjusted using RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 1% NP40, 0.5% DOC, 0.1% SDS, pH 8.0). Lämmli buffer was added and identical protein amounts of the different samples (20 μg) were loaded on mini-Protean TGX Stain-free Precast gels from Biorad. After the runs, the gels were activated by a 45 s UV exposure and electroblotted onto nitrocellulose membranes. The membranes were blocked and incubated with the Z125 antibody (58) (1:100 of a raw hybridoma supernatant) overnight in TBS-T (25 mM Tris pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.1% Tween-20) with 5% (w/v) fat-free dry milk powder. The anti-mouse HRP (Cell signaling, #7076S) secondary antibody was used after dilution (1:10 000) in TBS-T to visualize the BZLF1

signals after adding ECL select Western Blotting Detection Reagent (Amersham). The membranes were scanned using the ChemiDoc Imaging sytem (Bio-Rad), and the images were analyzed and the signals quantitated after total cell protein normalization using the Image Lab 6.0.1 software (Bio-Rad).

In vitro DNA methylation

CpG methylation *in vitro* was performed with the *de novo* methyltransferase M.SssI and S-adenosyl methionine as described (59).

Luciferase reporter assays

Forty-eight hours post-transfection, the HEK293 cells were analyzed with the Dual-Luciferase Reporter Assay System (Promega). Luciferase activity was measured in a 96-well microplate luminometer (Orion II, Berthold).

Quantitation of viral particles in cell supernatants

Three days post-transfection the cell supernatants of ZEBRA knockout cells were collected, filtered with 1.2 μm filters and kept at 4°C. The EBV genome contained in the producer cell line ZEBRA knockout carries the *egfp* gene, and infectious units are defined with the aid of Raji cells, which turn GFP-positive upon infection, allowing the direct assessment of the concentration of infectious EBV virions as green Raji units (GRU) per milliliter by flow cytometric analysis as described earlier (60,61).

RESULTS

Structure of the ZEBRA/meZRE2 complex

We crystallized ZEBRA's DNA-binding domain in complex with a 19 base-pair (bp) DNA duplex containing the CpG-methylated ZRE2 site in the EBV promoter Rp (TGAG^{m}CGA; hereafter meZRE2) and solved the structure at 2.5 Å resolution by molecular replacement (Supplementary Table S1). Unlike the single helix of a canonical bZIP domain, whose N- and C-terminal residues bind DNA and mediate coiled-coil dimer formation, respectively, ZEBRA's C-terminal region folds back on and stabilizes an unusually short coiled coil (Figure 1B). Our crystal structure closely resembles previous ZEBRA structures bound to the AP-1 site (43) and to the Rp meZRE2 site in an alternate crystal form (42), apart from the dimerization domain which exhibits variable bending (Supplementary Figure S1a). This domain is implicated in diverse protein–protein interactions (62–67) and its flexibility may allow ZEBRA to adapt to different binding partners.

Each ZEBRA monomer recognizes one of the two meZRE2 half-sites, which we denote 'A' (half-site shared with AP-1) and 'M' (methylated half-site) (Figure 1B). Consequently, only a single (hereafter 'CpG-proximal') monomer senses the methylation state of meZRE2. In the previously reported ZEBRA/AP-1 structure, the two ZEBRA monomers interact symmetrically with the AP-1 site except with the central G^0:C$^{0'}$ base pair: residue Arg^{190} from one monomer makes base-specific contacts with the guanine whereas the same arginine from the other monomer interacts nonspecifically with the phosphate flanking the cytosine (43). The ZEBRA/meZRE2 complex preserves this asymmetry: the CpG-proximal Arg^{190} reads the G^0 base while the CpG-distal arginine contacts the DNA backbone, with water-mediated H-bonds that bridge the A and M half-sites stabilizing this configuration (Figure 1E and Supplementary Figure S2).

ZEBRA recognizes the A half-site of meZRE2 essentially as in the complex with AP-1 (apart from a minor difference described in Supplementary Figure S1b). CpG-distal residues Asn^{182}, Ser^{186} and Arg^{190} form direct or water-mediated H-bonds with the C$^{2'}$, T$^{1'}$ and C$^{0'}$ bases, respectively, while seven basic residues (Arg^{179}, Arg^{183}, Arg^{187}, Lys^{188}, Arg^{190}, Lys^{192} and Lys^{194}) mediate electrostatic interactions with DNA phosphate groups (Figure 1E). The CpG-proximal monomer recognizes DNA bases in the M half-site through direct H-bonding interactions of Asn^{182} with G^2, Ser^{186} with ^{m}C$^{-2'}$ and ^{m}C^1, and Arg^{190} with G^0 and G$^{-1'}$, while electrostatic interactions with the DNA backbone resemble those in the A half-site. Notably, ZEBRA makes more base-specific contacts with the M than with the A half-site, allowing CpG methylation to have a greater impact on specific site recognition.

AP-1 and meZRE2 site geometry deviates at the CpG site

CpG methylation induces global changes in DNA structure (68–71), raising the possibility that ZEBRA's enhanced affinity for methylated ZRE2 may reflect an altered DNA conformation. Comparing the ZEBRA-bound AP-1 and meZRE2 structures reveals nearly identical DNA geometry except at the CpG motif, where large differences are observed in base-step parameters involving the G^2:^{m}C$^{-2'}$ base pair (Figure 2A). These differences primarily reflect a displacement of the ^{m}C$^{-2'}$ base towards the CpG-proximal ZEBRA monomer by 1.5 Å relative to the corresponding G$^{-2'}$ base of AP-1 (Figure 2B), as previously observed (42). The displacement is made possible by the phosphate backbone adopting a B$_{II}$ conformation instead of the more common B$_I$ conformation of standard B-form DNA (72,73). B$_{II}$ conformations can facilitate protein–DNA interactions by increasing the exposure of DNA bases in the major groove (74). In meZRE2, the B$_{II}$ conformation allows the ^{m}C$^{-2'}$ to slip away and destack from the G$^{-1'}$ base to form a H-bond with Ser^{186} and a van der Waals contact with Arg^{190} that stabilize the shifted base. A similar B$_{II}$ conformation and base destacking was observed in the DNA-bound structure of the yeast transcription factor Ndt80 and may characterize several other structures in which an Arg residue interacts with a YpG dinucleotide motif (where Y is a pyrimidine nucleotide) (75,76). In the DNA-bound Ndt80 structure, where two TpG motifs are recognized by two Arg residues, a B$_{II}$ conformation allows each 5′ T base to destack from the 3′ G and stack onto the guanidino group of the nearby Arg residue, which forms bidentate hydrogen bonds and is coplanar with the 3′ G base. Notably, whereas the Arg residues of Ndt80 and other YpG-recognizing proteins form cation–pi interactions with the destacked Y bases, in our ZEBRA structure the position of the Arg^{190} guanidino
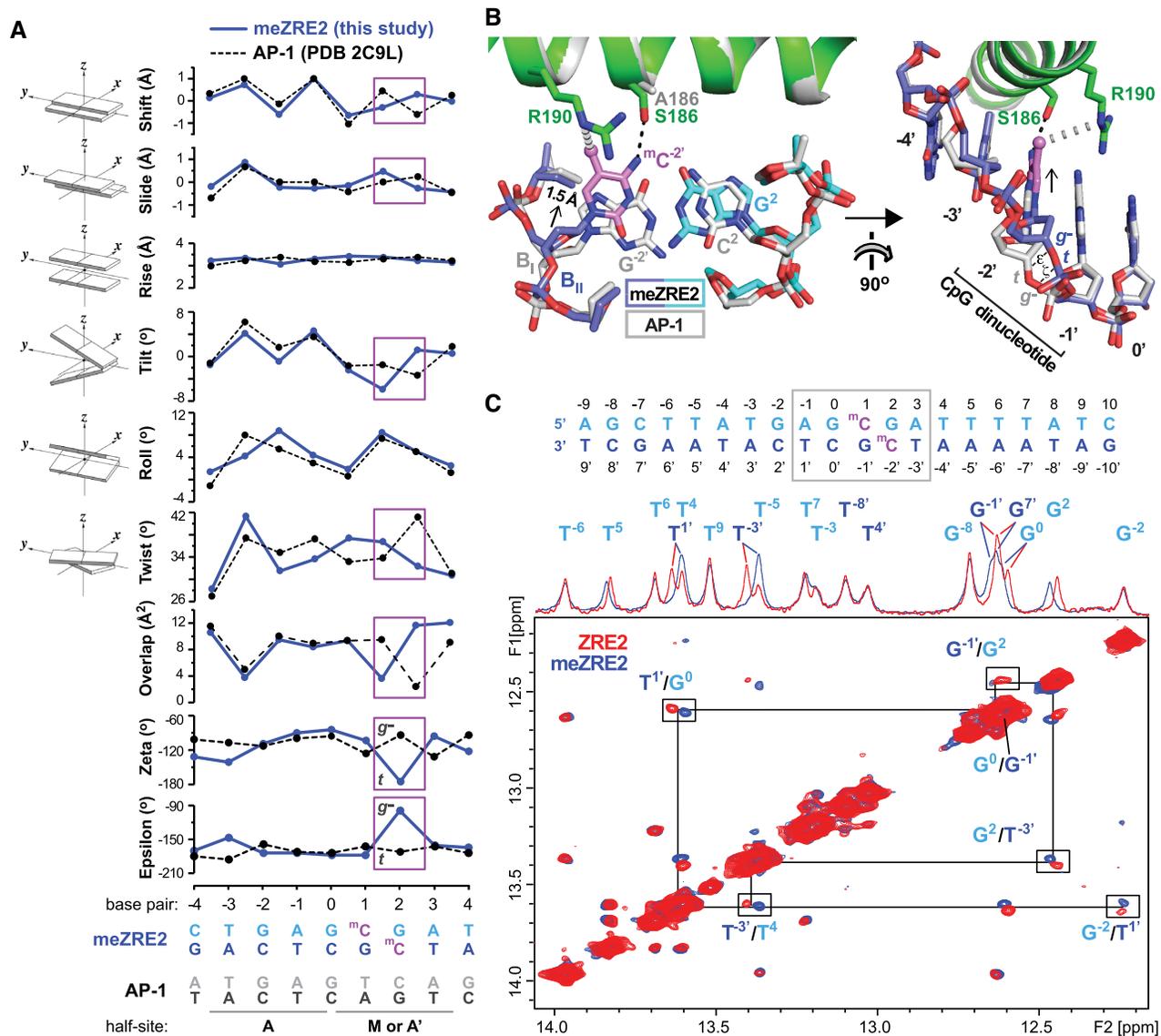
Figure 2. Deviations in AP-1 and meZRE2 site geometry localize to the methylation site. (**A**) Comparison of DNA geometry between the ZEBRA-bound AP-1 and meZRE2 sites. Base pair step parameters include translational (shift, slide, rise) and angular (tilt, roll, twist) parameters as well as the overlap between neighboring bases. Base pair numbering is that of the top (cyan) DNA strand. Base pair step parameters are plotted at the horizontal ordinate midway between the two base pairs comprising the step (e.g., values plotted at bp = 0.5 concern the base pair step $G^0C^{0'}/^mC^1:G^{-1'}$). Large deviations at base pair steps 1/2 and 2/3 are boxed in violet. Parameter values were calculated using the program 3DNA (49). Images at the left of graphs are from (49). Backbone epsilon (ε) and zeta (ζ) angles are those of the bottom (blue) DNA strand. The B_{II} conformation is characterized by ε and ζ adopting a (*gauche⁻*, *trans*) [(g^-, t)] configuration instead of the (t, g^-) configuration that characterizes B_I (72,73). (**B**) Shift of meZRE2 nucleotide $-2'$ compared to AP-1. The structure of the ZEBRA/meZRE2 complex was aligned with that of the ZEBRA/AP-1 complex (PDB 2C9L). Black and gray dashed lines indicate H-bond and van der Waals interactions, respectively. The black arrow shows the direction of the shifted base. The B_I and B_{II} backbone conformations are indicated for nucleotide $-2'$. The B_{II} conformation allows the $^mC^{-2'}$ base to interact with Ser[186] and Arg[190]. The corresponding G^{-2} base of the AP-1 site would be unstable in this position because it cannot donate a H-bond to Ser[186] (nor accept one since Ser[186] already shares its hydroxyl proton with the closer T^1 base). Right panel shows that the change in backbone geometry localizes to the linkage between nucleotides $-1'$ and $-2'$. Bonds related to torsion angles ε and ζ are indicated. (**C**) Imino region of 2D ^{1}H, ^{1}H-NOESY spectra and 1D spectra on top for ZRE2 (red) and meZRE2 (blue) DNA. The DNA sequences used and methylation sites are indicated on top. The sequential walk connecting the imino NMR signals in the central region (highlighted by a gray box in the DNA sequence) is indicated. For these base pairs notable chemical shift differences are observed upon methylation.

group relative to the $^mC^{-2'}$ base ring is too far and too greatly off-centered to form a strong cation–pi interaction. The ^mC base shift observed in our ZEBRA structure is also reminiscent of that previously described between methylated and unmethylated variants of the AP-1 site bound by homodimeric Jun (Supplementary Figure S3a) (42). However, whereas the shift in the ZEBRA complex is mediated by the $B_I \rightarrow B_{II}$ transition of a single phosphodiester bond, that in the Jun complex is achieved through small backbone adjustments that extend over several nucleotides and widen the major groove (Supplementary Figure S3b). This contrast underscores the highly localized nature of the structural changes that differentiate the ZEBRA-bound meZRE2 and AP-1 sites.

We next used solution nuclear magnetic resonance (NMR) to assess potential conformational changes induced by CpG methylation of ZRE2. Homonuclear imino NOESY spectra were recorded for 20 bp duplexes containing methylated and unmethylated ZRE2 (Figure 2C). Imino chemical shifts were readily assigned and are consistent with B-DNA geometry. NOE connectivities for the imino walk are essentially identical for methylated and unmethylated ZRE2. However, imino signals within 2 base pairs of the CpG motif show significant chemical shift changes upon methylation, with the largest differences observed for crosspeaks involving the imino groups of $G^{-1'}$ and G^2 that base pair with the methylated cytosines. These differences are consistent with the change in electronic environment caused by the spatial proximity of the two methyl groups and may also reflect a small change in the DNA helical conformation for this region. The lack of more extensive spectral changes confirms that CpG methylation does not induce large-scale changes in ZRE2 conformation. Taken together, the NMR and crystallographic data indicate that ZEBRA's selectivity for methylated over unmethylated ZRE2 does not involve changes in the overall DNA conformation but rather the recognition of structural features highly localized to the CpG site.

CpG methylation marks are read by Ser^{186} and Arg^{190}

ZEBRA binding to meZRE2 places the two CpG methylation marks in different stereochemical environments. ZEBRA recognizes the methyl groups of $^mC^1$ and $^mC^{-2'}$ through CpG-proximal residues Ser^{186} and Arg^{190}, respectively (Figure 1C). Strikingly, the Ser^{186} side chain is positioned precisely on the local dyad axis that relates the two CpG methylation marks, allowing it to hydrogen bond with both ^mC bases (Figure 1D). As previously observed (42), Ser^{186} senses the $^mC^1$ methyl group through a van der Waals contact with its side chain methylene group. The *gauche*$^+$ (g^+) rotamer observed for this side chain would be weakly populated in unbound ZEBRA since serine has a high (~85%) propensity to hydrogen bond with the helical backbone in the g^- conformation (Supplementary Figure S4a) (77). In an unmethylated ZEBRA/ZRE2 complex, the g^- rotamer of Ser^{186} would compete with the g^+ rotamer and attenuate DNA binding by reducing the number of base-specific H-bonds. By contrast, in the methylated complex the $^mC^1$ methyl group sterically selects for the g^+ rotamer,

thereby stabilizing the H-bonds with the ^mC bases (Supplementary Figure S4b).

The methylation mark on $^mC^{-2'}$ is sensed by Arg^{190} via its guanidino group (Figures 1C and 3A). This contact stabilizes the Arg^{190} side chain in a conformation that deviates slightly from that in the AP-1-bound structure (Supplementary Figure S1c). In the AP-1 complex, the Arg^{190} guanidino group forms bidentate H-bonds and is coplanar with the G^0 base, whereas in the meZRE2 complex it twists out of this plane to form a bifurcated H-bond with the $G^{-1'}$ base on the opposite strand. The resulting configuration is strikingly similar to the ^mC-Arg-G triad observed in methyl-CpG binding proteins (MBPs), whereby a conserved arginine hydrogen bonds with the G base of the CpG motif and contacts the methyl group of the adjacent ^mC base (analogous to $G^{-1'}$ and $^mC^{-2'}$ in our structure) (78–81) (Figure 3B–D). Compared to the canonical triad, ZEBRA's Arg^{190} side chain is shifted, such that it forms only a bifurcated H-bond with $G^{-1'}$ and instead forms bidentate H-bonds with G^0 on the opposite DNA strand (Figure 3E). The importance of this configuration is underscored by the observation that swapping the central G^0:$C^{0'}$ base pair for a C:G markedly destabilizes the ZEBRA/meZRE2 complex (42) and by anti-ZEBRA ChIP-seq data showing that the central G:C base pair of ZEBRA-binding sites on viral and human genomic DNA is invariable (27,82).

The two CpG methylation marks contribute unequally and independently to binding affinity

To determine the relative importance of the two CpG methylation marks for site recognition, we assessed ZEBRA's ability to bind ZRE2 sites that were either unmethylated, fully methylated or hemi-methylated on C^1 or $C^{-2'}$ (Figure 4A). An electrophoretic mobility shift assay (EMSA) showed that ZEBRA bound both hemi-methylated sites more tightly than unmethylated ZRE2 but less tightly than fully methylated ZRE2 (Figure 4B), indicating that methylation on each DNA strand has an additive effect on binding affinity. ZEBRA bound the two hemi-methylated sites with similar affinity as a viral AP-1-like site (site ZRE5 from the lytic origin of replication), although binding appeared slightly stronger when hemi-methylation was on C^1 compared to $C^{-2'}$.

To better estimate ZEBRA's affinity for differentially methylated ZRE2 sites we used a fluorescence polarization (FP) assay, in which the binding of ZEBRA to a fluorescently labeled DNA probe allows determination of the apparent dissociation constant ($K_{d,app}$). Like the EMSAs, FP assays were performed in the presence of non-specific competitor DNA so as to emulate cellular conditions, where specific ZREs compete for ZEBRA against a large background of low-affinity binding sites in the genome. ZEBRA bound the fully methylated site with >20-fold higher affinity than the unmethylated ZRE2 ($K_{d,app}$ of 75 nM versus 1.7 µM; Figure 4C and Table 1), in general agreement with previous findings (6,27,34,42). $K_{d,app}$ values lower by a factor of ~10 have been reported in the absence of competitor DNA (42). ZRE2 hemi-methylated on C^1 reduced binding by a factor of 2.2 ($K_{d,app} = 167$ nM) compared to the fully methylated site. By contrast, hemi-methylation on $C^{-2'}$ reduced binding
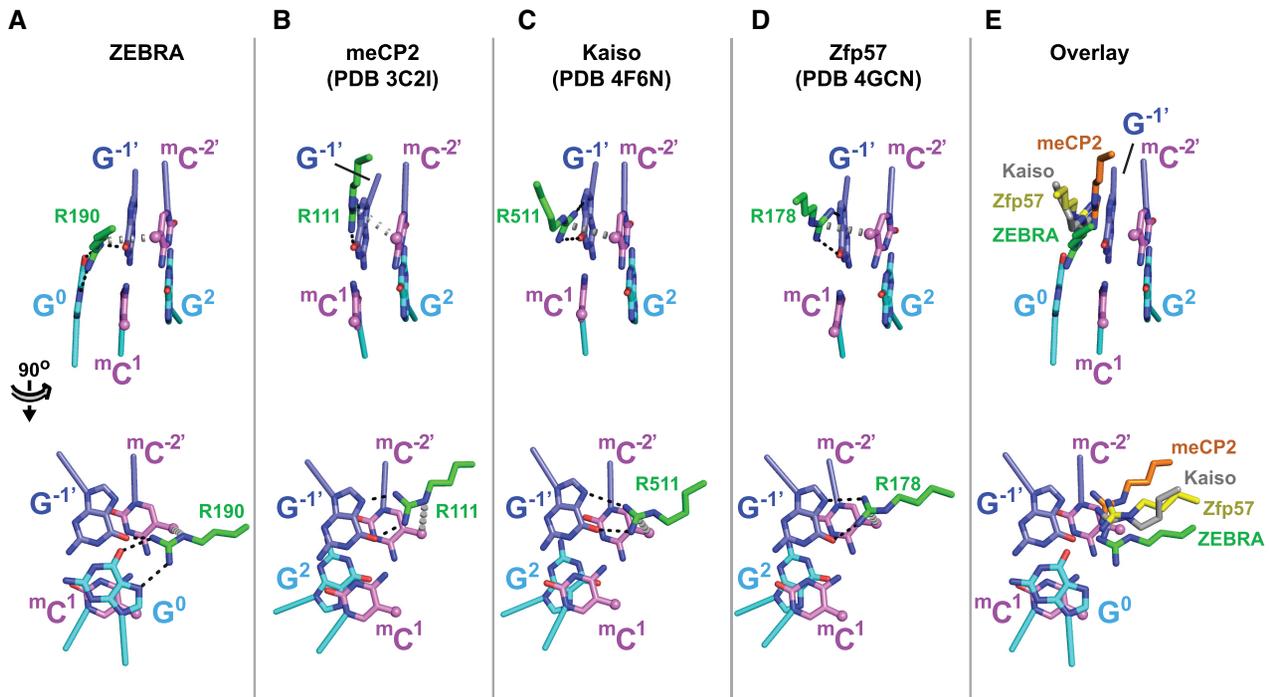
Figure 3. ZEBRA recognizes the $^{m}C^{-2'}$ methyl group through an atypical ^{m}C-Arg-G triad motif. (**A**) Recognition of the $^{m}C^{-2'}$ methyl by ZEBRA residue Arg[190]. The top and bottom panels show views perpendicular to and along the DNA helical axis, respectively. Black and gray dashed lines indicate H-bonds and van der Waals contacts, respectively. (**B–D**) ^{m}C-Arg-G triad observed in three methyl-CpG binding proteins showing that residues (B) Arg111 in meCP2 ([79]), (C) Arg511 in Kaiso ([80]) and (D) Arg178 in Zfp57 ([81]) adopt a similar orientation with respect to the methylated CpG motif. (**E**) Structural alignment of the ^{m}C-Arg-G motifs from meCP2 (orange), Kaiso (gray) and Zfp57 (yellow) with the corresponding motif from ZEBRA. For simplicity only the DNA bases from ZEBRA are shown. ZEBRA's Arg[190] side chain forms bidentate H-bonds with G^0 instead of with G^{-1} as observed in the canonical triad.

by a factor of ∼9 ($K_{d,app}$ = 670 nM), revealing that the two methylation marks contribute unequally to binding affinity, as suggested by the EMSA (Figure 4B).

Additional insights can be gained from the FP data by considering the apparent binding free energy, ΔG_{app} [calculated as $-RT\ln(1/K_{d,app})$ with temperature T and gas constant R] and how this quantity varies ($\Delta\Delta G_{app}$) between different complexes. Plotting ΔG_{app} for the above ZEBRA/DNA complexes reveals that, compared to unmethylated ZRE2, the gain in complex stability obtained by fully methylating ZRE2 ($\Delta\Delta G_{app}$ = −1.85 kcal/mol) closely matches the sum of the values obtained by hemimethylating at C^1 (−1.38 kcal/mol) and at $C^{-2'}$ (−0.56 kcal/mol) (Figure 4D and Table 1). For comparison, the energy of a neutral H-bond in water is 0.5–1.5 kcal/mol, depending on the bond strength ([83–85]). Thus, each methyl group contributes independently to the enhanced affinity of ZEBRA for the fully methylated site, with the C^1 and $C^{-2'}$ methylation marks each providing approximately 75% and 25% of the binding energy, or roughly the equivalent of a strong and weak H-bond, respectively.

Insights into the inhibitory effect of CpG hydroxymethylation

Besides CpG methylation, another epigenetic mark implicated in regulating EBV gene expression is cytosine 5-hydroxymethylation. This modification is mediated by ten-

eleven translocation (TET) dioxygenases, which convert 5-methylcytosine (^{m}C) to 5-hydroxymethylcytosine (^{hm}C) ([86]). Loss of TET2 is implicated in the development of EBV-positive nasopharyngeal carcinoma ([87]) and may play a role in EBV-positive gastric carcinoma ([88]). In EBV-infected cells ZEBRA-mediated lytic reactivation is strongly reduced by ^{hm}C modification of viral lytic promoters ([87]). *In vitro* studies have shown that ^{hm}C modification markedly inhibits the binding of ZEBRA to CpG-containing ZREs relative to the methylated ZRE site ([42,87]). In agreement with these findings, FP assays showed that ZEBRA's ability to bind meZRE2 was significantly compromised by hydroxymethylation (Supplementary Figure S5a,b). Binding was reduced to a similar degree (by a factor of ∼1.7) when either $^{m}C^1$ or $^{m}C^{-2'}$ was replaced by ^{hm}C and further reduced (by a factor of 3.1) when both modifications were made, revealing that the two hydroxymethyl marks had an additive inhibitory effect. Interestingly, ZEBRA's affinity for hydroxymethylated ZRE2 (hmZRE2) was comparable to that for the AP-1 site (Table 1), whose high abundance in the human genome might outcompete hmZRE2 for ZEBRA binding (see below).

To understand the inhibitory effect of hydroxymethylation, we modeled the structure of ZEBRA bound to a ^{hm}C-modified ZRE2 site by replacing the two ^{m}C nucleotides in our crystal structure by ^{hm}C. A survey of high-resolution ^{hm}C-containing DNA structures in the Protein Data Bank
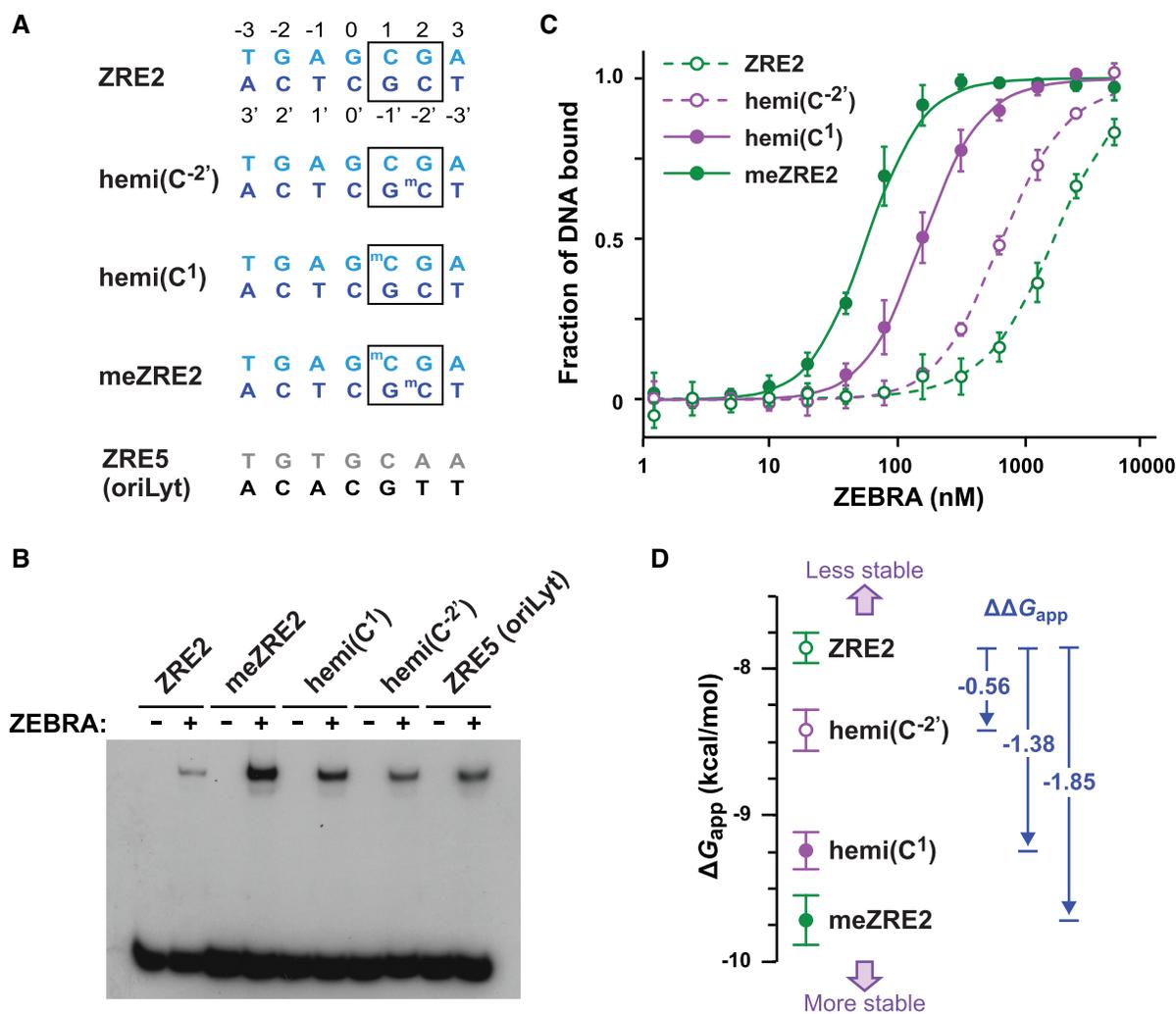
Figure 4. The two CpG methylation marks contribute unequally and independently to binding affinity. (**A**) ZRE sequences used for binding assays. Rp ZRE2 sequences (blue and cyan) were either unmethylated, hemi-methylated or symmetrically methylated as indicated. The AP-1-like sequence ZRE5 from oriLyt (black and gray) was included for comparison. (**B**) EMSA performed with affinity-purified Strep/FLAG:ZEBRA fusion protein transiently expressed in HEK293 cells. One representative experiment out of three is shown. (**C**) FP assays assessing the binding of ZEBRA to ZRE2 sites bearing 0, 1 or 2 methylation marks. (**D**) Apparent free energy of binding of ZEBRA/DNA complexes. Differences in the stability of complexes ($\Delta\Delta G_{app}$ values) are indicated.

(PDB) revealed that the ^{hm}C hydroxymethyl group preferentially adopts a *syn*-periplanar (*sp*) or (+)-clinal (+*c*) conformation (Supplementary Figure S5c,d). This rotational dimorphism is favoured by direct or water-mediated H-bonds with specific atoms of the CpG dinucleotide (89) and is consistent with energy calculations (90). Our structural model predicts that the *sp* and +*c* conformations of ^{hm}C^1 would give a strong steric clash with the backbone atoms of Arg^{183} and Asn^{182}, respectively (Supplementary Figure S5e), while the +*c* conformation of ^{hm}C$^{-2'}$ would clash with the guanidino group of Arg^{190} that interacts with the G^0 and G$^{1'}$ bases (Supplementary Figure S5f). Relieving these clashes would require an increased separation between the protein backbone and DNA bases that would disrupt the H-bonds between Ser^{186} and the C^1 and C$^{-2'}$ bases (Figure 1D) and thereby destabilize the complex.

Ser^{186} and Arg^{190} are key determinants of high-affinity meZRE2 binding

To evaluate the significance of protein contacts with the ^{m}CpG motif observed in our crystal structure, we examined the effect of single point mutations on ZEBRA's ability to bind meZRE2 in FP assays (Table 1). We first mutated Ser^{186} to either a threonine, alanine or cysteine. Whereas threonine replacement gave a modest drop in affinity (by a factor of 1.6) consistent with stereochemical considerations (detailed in Figure 5 legend), replacement by alanine caused a more pronounced reduction in affinity (by a factor of 6; Figure 5A), in line with previous findings (34,40–42) and consistent with the loss of three H-bonds that Ser^{186} makes with the M and A half-sites (Figure 1C–E). A more dramatic decrease in affinity (by a factor of >16) was observed when Ser^{186} was replaced by cysteine. This is surpris-
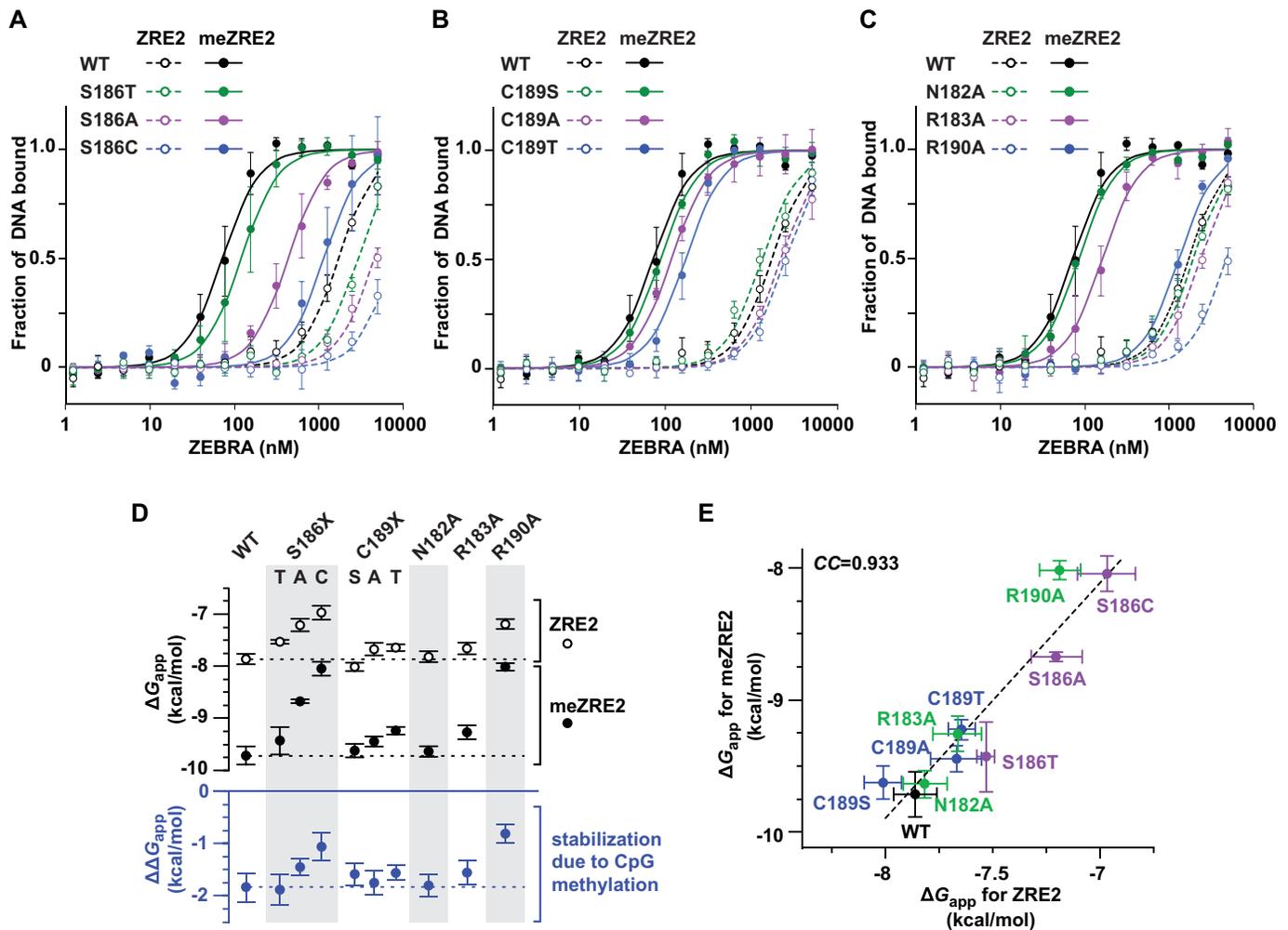
Figure 5. Effect of ZEBRA point mutations on binding affinity and selectivity for methylated ZRE2. (**A–C**) FP assays assessing the effect of (A) Ser[186] mutants, (B) Cys[189] mutants and (C) alanine substitutions of Asn[182], Arg[183] and Arg[190] on ZEBRA's ability to bind methylated and unmethylated ZRE2. The drop in affinity observed for the S186T mutant is consistent with a steric clash predicted between Arg[190] and the threonine methyl group that would hinder optimal positioning of the threonine hydroxyl group relative to the two ^mC bases. The drop in affinity observed for the C189A mutant is consistent with the loss of non-specific interactions between the Cys[189] thiol and the T^{-3} and $T^{-3'}$ phosphate groups. Modeling suggests that the decreased affinity observed for the C189T mutant is due to steric constraints involving the threonine and $T^{-3'}$ methyl groups that result in suboptimal H-bonding geometry. Effects observed for the remaining mutants are discussed in the main text. (**D**) Plot of apparent free binding energies (ΔG_{app}) of ZEBRA/DNA complexes derived from $K_{d,\,app}$ values determined in (A–C). The difference in stability between ZRE2 and meZRE2 complexes ($\Delta\Delta G_{me/Z}$) is indicated below. (**E**) Plot of ΔG_{app} values comparing the stability of meZRE2 and ZRE2 complexes; CC, correlation coefficient.

ing given the nearly isosteric cysteine and serine side chains and contrasts with the inverse serine substitution of Cys[189], which had virtually no effect (Figure 5B). We surmised that disulfide crosslink formation might explain the poor activity of the S186C mutant (Supplementary Figure S6a); however, experiments do not support this hypothesis (Supplementary Figure S6b–d). Interestingly, a dramatic loss of binding activity was also reported for a bacterial sulfate-binding protein when a Ser residue that donates a hydrogen bond to the sulfate ligand was substituted by Cys, compared to a much weaker effect when Ala or Gly was substituted (91). This loss of activity was attributed to differences in the size and preferred angles of the Cys thiol group relative to the Ser hydroxyl group and to the differential work required to polarize these groups (91,92). Similar effects might explain the poor binding activity of the ZEBRA S186C mutant. In con-

trast to serine, replacing Cys[189] by an alanine or threonine decreased binding affinity for meZRE2 by a factor of 1.6 or 2.3, respectively (Figure 5B). These results can be rationalized structurally (Figure 5 legend) and are consistent with a recent protein binding microarray study that found reduced meZRE2 binding for ZEBRA mutants C189A and C189T (93).

Finally, we investigated the effect of alanine substitutions at positions 182, 183 and 190 (Figure 5C). Asn[182] is conserved across canonical bZIP proteins and is important for specific AP-1 site recognition, with mutation to alanine resulting in the loss of high-affinity binding (94). By contrast, alanine substitution of Asn[182] had little effect on ZE-BRA's affinity for meZRE2 (Figure 5C and Table 1). This difference is partly due to the fact that, whereas the conserved Asn of bZIP proteins forms four H-bonds with the

AP-1 site ([95,96](#)), ZEBRA residue Asn[182] forms only two with meZRE2, of which one is weak (Figure [1](#)C,E and Supplementary Figure S1b), and might also reflect a tighter protein–DNA interface enabled by the smaller alanine side chain that compensates for the disrupted H-bonds. Alanine substitution of Arg[183] had only a small effect on binding affinity, consistent with this residue's non-specific interaction with the phosphate backbone. In contrast, alanine substitution of Arg[190] caused a drastic loss of binding affinity (by a factor of 17), consistent with the loss of multiple specific and non-specific interactions that this residue mediates with the M and A half-sites, respectively (Figure [3](#)A and Supplementary Figure S2).

In summary, of the five alanine substitutions tested, mutants S186A and R190A greatly reduced ZEBRA's affinity for meZRE2, whereas mutants N182A, R183A and C189A had at most only a modest effect on binding. Thus, the same residues that sense the CpG methylation status within meZRE2 are also critically required for high-affinity binding.

Selectivity for meZRE2 is robust and mostly independent of base-specific contacts

In parallel to the above experiments we also assessed point mutants for their ability to bind the unmethylated ZRE2 site. In general, mutations affected the binding of ZRE2 similarly to that of meZRE2. For example, the S186T mutation reduced the affinity for meZRE2 and ZRE2 by factors of 1.6 and 1.7, corresponding to a loss in complex stability of 0.28 and 0.33 kcal/mol, respectively (Figure [5](#)A and D). Indeed, the results for the nine mutants revealed that the ΔG_{app} values for the methylated and unmethylated complexes correlated strongly (Figure [5](#)E and Table [1](#)).

Interestingly, for certain mutations the impact on binding affinity differed significantly between ZRE2 and meZRE2. In particular, the S186A mutation decreased the binding affinity for ZRE2 by a factor of 3, compared to a factor of 6 for meZRE2, implying a 50% drop in selectivity for the methylated site. This is consistent with CpG methylation stabilizing Ser[186] in a g^+ conformation that mediates two H-bonds with the DNA (Supplementary Figure S4b), which are eliminated by the alanine substitution. In the unmethylated complex the reduced prevalence of the g^+ rotamer decreases the effective number of H-bonds disrupted by the mutation, explaining the weaker destabilizing effect. More strikingly, the R190A mutation reduced the binding affinity for meZRE2 by a factor of 17 but that for ZRE2 by only a factor of 3, hence reducing the selectivity for the methylated site by a factor of 5.5. This implies that CpG methylation has a much weaker stabilizing effect on the mutant ZEBRA/DNA complex compared to the WT ($\Delta\Delta G_{app}$ increased by >1 kcal/mol; Figure [5](#)D and Table [1](#)). This finding is consistent with the ^{m}C-Arg-G triad geometry in which the $^mC^{-2'}$ methyl group stabilizes Arg[190] in a conformation that hydrogen bonds with the G^0 and $G^{-1'}$ bases (Figure [3](#)). Truncation of this side chain renders the R190A mutant unable to sense the methylation status of the $C^{-2'}$ base, explaining why CpG methylation more weakly stabilizes the mutant complex.

Notably, no mutations were identified that abolished selectivity for methylated ZRE2 (even the R190A mutant retained 4-fold selectivity). In particular, selectivity did not require ZEBRA-specific residue Ser[186], since the alanine mutant still discriminated efficiently (12-fold) in favor of meZRE2. Taken together these observations reveal that ZEBRA's selectivity for methylated DNA is remarkably robust and only weakly depends on the integrity of individual base-specific contacts.

Ser[186] confers binding selectivity for meZRE2 over AP-1

Since ZEBRA is known to recognize two (AP-1-like and CpG-containing) classes of ZREs, we examined its affinity for methylated and unmethylated ZRE2 relative to that for the AP-1 site. As reported above, ZEBRA binds meZRE2 with >20-fold selectivity over ZRE2 ($K_{d,app}$ values of 75 and 1700 nM, respectively). FP assays showed that ZEBRA bound the AP-1 site with an affinity intermediate between these two values ($K_{d,app}$ of 270 nM; Figure [6](#)A and Table [1](#)). Consistent with this observation, the EMSA in Figure [4](#)B shows that ZEBRA's affinity for another AP-1-like site, oriLyt ZRE5, is also intermediate between that for ZRE2 and meZRE2. For comparison, we examined the site selectivity of GCN4, a canonical bZIP protein from yeast that recognizes AP-1 sites ([95](#)). As expected, GCN4 bound AP-1 tightly ($K_{d,app}$ of 78 nM) with 50-fold selectivity over ZRE2 (Figure [6](#)B). GCN4 bound meZRE2 with an intermediate affinity ($K_{d,app}$ of 480 nM) that was 8 times stronger compared to ZRE2 and 6 times weaker compared to AP-1. Thus, whereas the binding selectivity of ZEBRA follows the order meZRE2 > AP-1 > ZRE2, that of GCN4 switches the order of AP-1 and meZRE2. This difference arises because ZEBRA's affinity is both lower for AP-1 and higher for meZRE2 compared to GCN4.

Like most bZIP proteins, GCN4 has an alanine (Ala[239]) instead of ZEBRA's Ser[186] residue. Mutating the alanine residue in GCN4 to serine reduced its affinity for the AP-1 site by a factor of 2 and enhanced that for ZRE2 and meZRE2 site 2.5- and 5-fold, respectively, resulting in ZEBRA-like selectivity (meZRE2 > AP-1 > ZRE2) (Figure [6](#)D). The inverse mutation on ZEBRA (S186A) had the opposite effect: the affinity for AP-1 increased 2-fold, while that for ZRE2 and meZRE2 decreased by factors of approximately 3 and 6, respectively, yielding GCN4-like selectivity (AP-1 > meZRE2 > ZRE2; Figure [6](#)C). Comparing the stabilities (ΔG_{app} values) of all 12 protein/DNA combinations reveals a striking similarity between corresponding ZEBRA and GCN4 complexes (Figure [6](#)E). Whereas the serine-containing ZEBRA and GCN4 variants form the most stable complexes with meZRE2, the alanine substitution stabilizes the AP-1 complex (downward shift of ΔG_{app}) while destabilizing both the meZRE2 and ZRE2 complexes (upward shifts).

The above trends become evident when the differences in complex stability are expressed as $\Delta\Delta G_{app}$ values. For convenience, we denote ZEBRA's ability to discriminate meZRE2 from either ZRE2 or AP-1 as selectivity of type 'me/Z' or 'me/A' (defined as the ratio of ZEBRA's apparent binding affinity, $1/K_{d,app}$, for meZRE2 to its apparent affinity for ZRE2 or AP-1, respectively; Figure [1](#)A) and de-
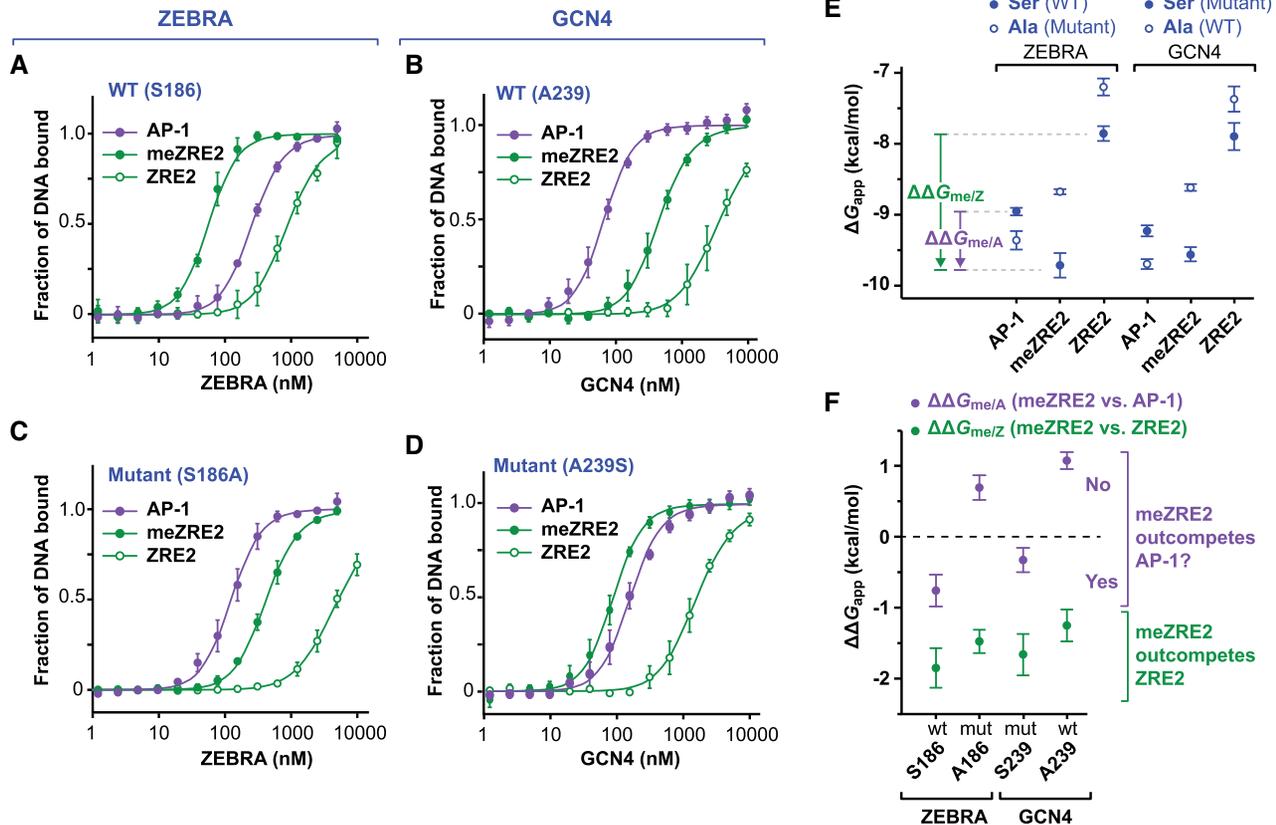
Figure 6. Ser^{186} confers binding selectivity for meZRE2 over AP-1. (A–D) FP assays assessing the relative binding affinity of (A) WT ZEBRA, (B) WT GCN4, (C) ZEBRA mutant S186A and (D) GCN4 mutant A239S for the AP-1, ZRE2 and meZRE2 sites. (E) Plot of apparent free binding energies (ΔG_{app}) of protein/DNA complexes derived from $K_{d,\ app}$ values determined in (A–D). The quantities $\Delta\Delta G_{me/Z}$ and $\Delta\Delta G_{me/A}$ are illustrated for WT ZEBRA. (F) Values of $\Delta\Delta G_{me/Z}$ and $\Delta\Delta G_{me/A}$ plotted for the indicated ZEBRA or GCN4 protein illustrating the stability of the meZRE2-bound protein relative to that of the ZRE2 ($\Delta\Delta G_{me/Z}$) or AP-1 ($\Delta\Delta G_{me/A}$) complex.

note the corresponding differences in apparent binding free energy as $\Delta\Delta G_{me/Z}$ or $\Delta\Delta G_{me/A}$, respectively. The latter quantities are illustrated for WT ZEBRA in Figure 6E and plotted for the four ZEBRA and GCN4 proteins in Figure 6F (large negative $\Delta\Delta G$ values correspond to high positive selectivity). Methylating ZRE2 induces a similar stabilization of DNA-bound ZEBRA ($\Delta\Delta G_{me/Z} = -1.85$ kcal/mol) and GCN4 (-1.66 kcal/mol) when residue 186 or 239 is serine. An alanine at this position yields a small increase ($\sim$0.4 kcal/mol) in $\Delta\Delta G_{me/Z}$, which nevertheless remains below -1.2 kcal/mol for the WT and mutant forms of both proteins, reflecting their shared high selectivity for meZRE2 over ZRE2. In contrast, the Ser→Ala substitution induces a large shift ($+1.4$ kcal/mol) in $\Delta\Delta G_{me/A}$, which flips from a negative to a positive value for both proteins. This inversion of sign corresponds to the switch in binding-site ranking described above: whereas meZRE2 outcompetes the AP-1 site for the serine-containing ZEBRA and GCN4 variants ($\Delta\Delta G_{me/A} < 0$), AP-1 outcompetes meZRE2 for the alanine variants ($\Delta\Delta G_{me/A} > 0$). Thus, the identity of the residue at or equivalent to position 186 determines which of the two binding sites these bZIP proteins preferentially bind.

Transactivation of a CpG-methylated promoter mirrors meZRE2 binding affinity

We next assessed the ability of ZEBRA mutants to transactivate a CpG-methylated promoter in a luciferase reporter assay. Pentamers of the CpG-containing ZRE site from the EBV BSLF2/BMLF1 promoter were inserted into a luciferase reporter plasmid that was otherwise devoid of CpG motifs (27,56). Following mock treatment or treatment with a *de novo* methyl transferase to introduce CpG methylation of the five ZREs, the plasmid DNA was transiently transfected into HEK293 cells together with an expression plasmid encoding WT or mutant ZEBRA protein. Quantitative western blot analysis showed that mutants were expressed at the expected size and at near WT levels (Figure 7A). As expected, transfection with the WT ZEBRA protein led to strong transactivation of the methylated promoter (97-fold higher relative to a luciferase control plasmid free of promoter elements) but yielded only background activation of the unmethylated promoter (Figure 7B). Compared to WT, the transactivation of the methylated promoter was similar or higher for three mutants (S186T, C189T and N182A), was reduced by factors of 2 to 5 for three other mutants (C189A, C189S, R183A) and was reduced by a factor of
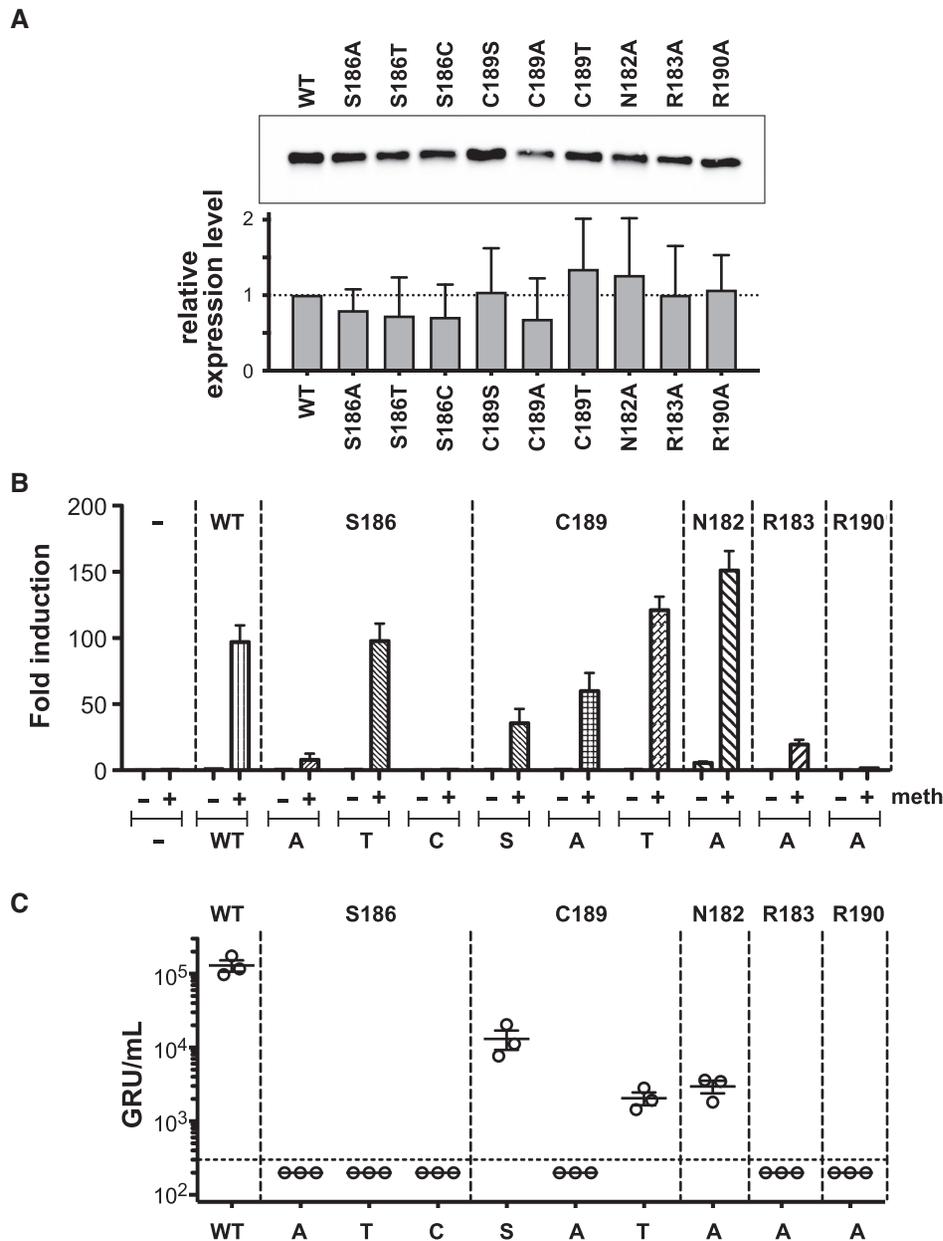
Figure 7. Effect of ZEBRA point mutations on promoter activation and viral production. (**A**) Quantitation of steady state protein levels of WT and mutant ZEBRA proteins. Top: Representative image of ZEBRA proteins after western blot immunodetection. Bottom: Statistical analysis based on six independent biological replicates and western blot analyses. Cellular total protein levels were quantified after membrane blotting and used to normalize the levels of the ZEBRA protein signals after immunostaining. WT ZEBRA signal levels were set to 1.0, and the relative expression of the ZEBRA mutants were calculated after protein normalization. 293T cells were transfected with the WT ZEBRA expression plasmid and nine plasmids encoding the ZEBRA mutants as indicated. Three days post-infection, the cells were lysed, protein lysate concentrations were determined and equal protein amounts were subjected to quantitative western blot analysis using the western blot stain-free TGX Biorad Normalization approach (Bio-Rad). Blots were probed with the Z125 monoclonal antibody ([58]). Mean and standard deviation are shown. (**B**) Luciferase reporter assays assessing the ability of different ZEBRA mutants to activate a ZRE- or meZRE-containing promoter. Unmethylated and fully CpG-methylated reporter constructs were analyzed in the presence or absence of the indicated ZEBRA expression plasmid. After data normalization to a luciferase control plasmid free of promoter elements, the *x*-fold differences were calculated. Each experiment was performed three times and the means and standard deviations are depicted. (**C**) All ZEBRA mutants are impaired to different degrees in their capacity to reactivate virus production *in vitro*. ZEBRA knockout cells were transfected with the plasmids expressing WT ZEBRA or the indicated mutants. Viral titers in the cell supernatants were analyzed by infecting Raji cells and are provided as 'green Raji units' (GRU) per ml. Results from three independent experiments, including the means and standard deviations, are shown. The horizontal dotted line represents the detection limit of our assay.

>10 for the three remaining mutants (S186A, S186C and R190A). Interestingly, the level of transcriptional activation closely mirrored the *in vitro* binding stability measured for ZEBRA mutants in complex with meZRE2 (Supplementary Figure S7a). Thus, the degree of transcriptional activation in this assay showed a gradual response commensurate with ZEBRA's affinity for the meZRE2 site.

Viral lytic activation by ZEBRA mutants suggests a threshold-like response

Next, we tested whether our ZEBRA mutants were able to induce EBV's lytic cycle. We used a HEK293 cell line stably transfected with an EBV genome encoding green fluorescent protein (GFP) and unable to express ZEBRA. Transient transfection of these cells with ZEBRA and BALF4 (encoding the viral envelope glycoprotein gp110) leads to the production of viral particles, whose concentration is assessed by infecting (and hence inducing GFP expression in) Raji cells, followed by flow cytometry to quantify the green Raji units (GRU) per ml, as previously described (61). As expected, a strong viral lytic response was observed following transfection with the WT construct (Figure 7C). In contrast, we failed to detect viral particles in the culture supernatant following transfection with six of the ZEBRA mutants (R183A, C189A, R190A and all three S186 mutants), indicating that these mutations abrogate ZEBRA's ability to activate the lytic cycle. The remaining three mutants (C189A, C189T and N182A) were able to induce viral production, although less efficiently than WT ZEBRA, reducing viral production (GRU/ml) by a factor of 10 to 50.

The above data correlate poorly with the ability of mutants to activate transcription in the luciferase reporter assay (Supplementary Figure S7b), reflecting the non-linear dependence of viral production on meZRE2-containing promoter activation. This is not surprising since EBV lytic activation is probably regulated by several genes whose expression depends on ZEBRA binding to various ZRE and meZRE sites. ZEBRA is also an essential replication factor that needs to bind the lytic origin of DNA replication to promote efficient viral DNA amplification (97,98). Interestingly, the ability of ZEBRA mutants to activate the lytic cycle was strongly associated with their ability to bind meZRE2 with a $K_{d,app}$ below ~100 nM ($\Delta G_{app} <$ −9.5 kcal/mol) in our FP assays (the only exception being mutant C189T, which activated lytic replication with a $K_{d,app}$ of 170 nM; Supplementary Figure S7c). This suggests that lytic activation involves a threshold-like response to ZEBRA/meZRE2 complex formation. This hypothesis has been recently confirmed in a model that allows a dose-dependent evaluation of ZEBRA's ability to induce the lytic phase of EBV (see Figure 8 in (99)).

DISCUSSION

In this study we investigated ZEBRA's interactions with its two classes of ZRE target sites. We show that ZEBRA residues Ser^{186} and Arg^{190} play a key role both in establishing high-affinity binding with meZRE2 and in recognizing its CpG methylation status. Methyl-CpG recognition is facilitated by a DNA backbone conformation that allows the ^mCpG motif to form an extra H-bond with Ser^{186} and to contact Arg^{190} through a non-canonical ^mC-Arg-G triad (Figures 2B and 3A). Binding assays with hemi-methylated DNA revealed that the CpG methylation mark read by Ser^{186} enhances ZEBRA's affinity for meZRE2 to a 3-fold greater extent than that read by Arg^{190} (Figure 4). Conversely, alanine point mutations revealed that the R190A mutant had a 3-fold lower meZRE2-binding affinity than the S186A mutant (Figure 5A,C and Table 1). Thus, while both residues participate in high-affinity binding and ^mCpG recognition, they make unequal and opposite contributions to these activities. Remarkably, all individual ZEBRA point mutants investigated retained a strong binding preference for meZRE2 over ZRE2 (Figure 5 and Table 1), suggesting that an inherent difference in the physico-chemical behaviour of these two sites might contribute significantly to such selectivity (denoted 'me/Z' selectivity). Indeed, computational studies have shown that methylation preferentially drives a CpG-containing DNA duplex into a protein-DNA complex compared to the unmethylated duplex (100). Consistent with this idea, the mutation that most severely compromises me/Z selectivity, R190A, is predicted to induce a large increase in methylcytosine solvent accessibility within the ZEBRA/meZRE2 complex, whereas the less compromising S186A mutation is predicted to cause only a marginal increase (Supplementary Figure S8).

ZEBRA binds more tightly to meZRE2 than to the consensus AP-1 site (Figure 6A) or to an AP-1-like site (Figure 4B). In agreement with this observation, a recent genome-wide study found that, at low levels of intracellular expression, ZEBRA predominantly associates with CpG-containing ZRE motifs and when expressed at higher levels additionally associates with AP-1-like sequences (99). What is the molecular basis for this type of selectivity (denoted 'me/A' selectivity)? Because AP-1 and meZRE2 share the A half-site in common, the answer lies in how ZEBRA interacts differently with the unique M and A' half-sites (Figure 8A). Surprisingly, these two interfaces have the same number of H-bonds. Except for Asn^{182} and Ser^{186}, all DNA-contacting residues mediate similar interactions in the two complexes. Asn^{182} makes two H-bonds with the A' half-site (with C^2 and $T^{-3'}$) but only one with the M half-site (with G^2). Conversely, Ser^{186} makes only one H-bond with the A' half-site (with T^1) but two with the M half-site (with $^mC^1$ and $^mC^{-2'}$), and so the total H-bond count is identical for both complexes [considering single and bifurcated H-bonds as equivalent (101,102); see Figure 8B legend]. Importantly, however, these H-bonds are located in different environments: the extra H-bond (between Asn^{182} and $T^{-3'}$) in the AP-1 complex is exposed to solvent, whereas the extra H-bond (between Ser^{186} and $^mC^{-2'}$) in the meZRE2 complex is buried (Figure 8A,B). Buried H-bonds are more stable (by up to 1.2 kcal/mol (103)) than those accessible to solvent since water competes for the H-bond donor and acceptor sites (104–106). The extra methyl group on $^mC^{-2'}$ also makes the M half-site more hydrophobic than the A' half-site, and so burying the larger hydrophobic surface would yield a greater entropic gain for the meZRE2 complex. Taken together, enhanced H-bond stability and a larger hydrophobic
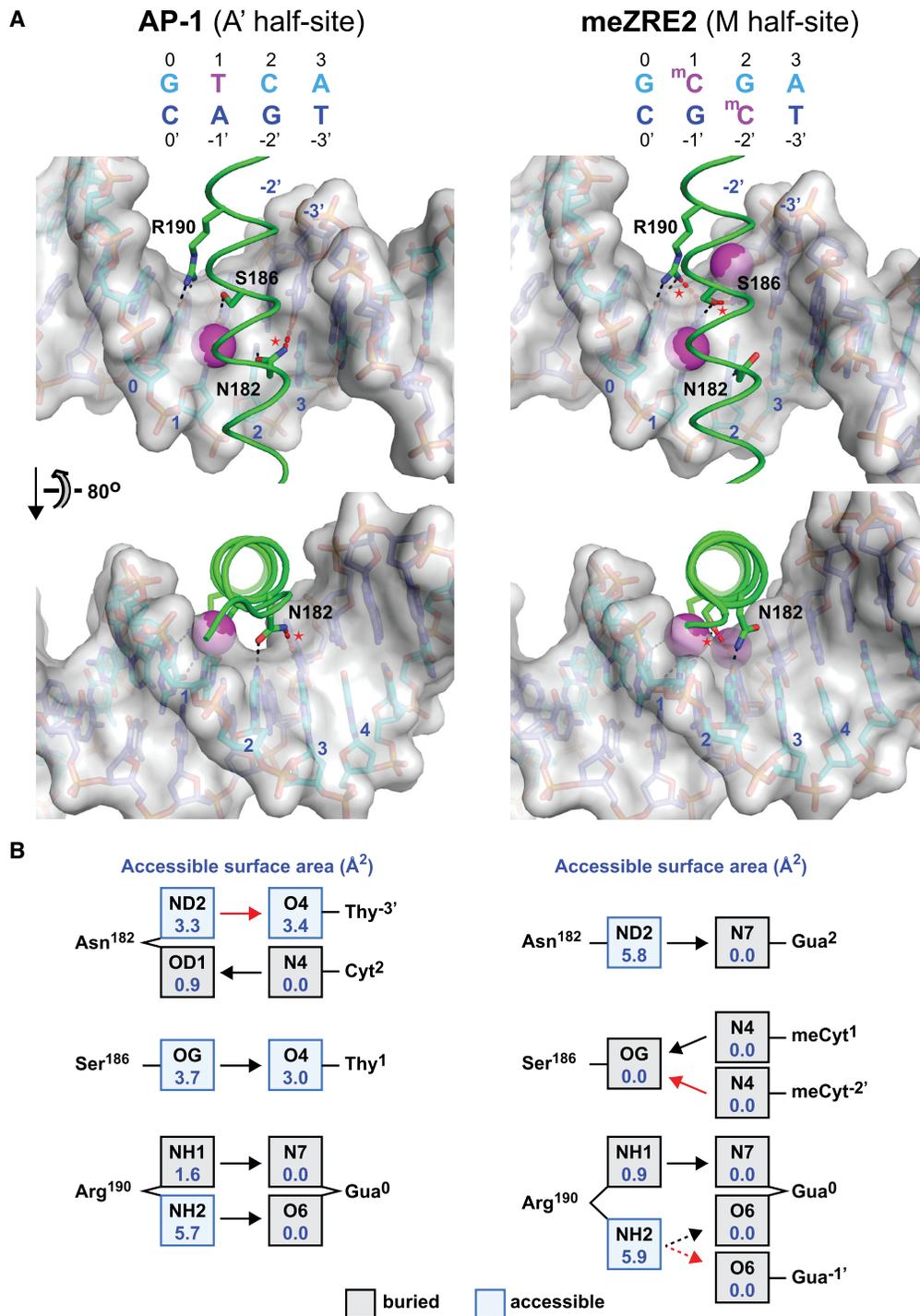
Figure 8. H-bond interactions with the unique (A' and M) half-sites of AP-1 and meZRE2. (**A**) Base-specific hydrogen bond interactions mediated by residues Asn[182], Ser[186] and Arg[190] and nucleotide bases in the A' half-site of AP-1 (left) and corresponding M half-site of meZRE2 (right). The structures of the AP-1 and meZRE2 complexes are from PDB 2C9L and this study, respectively. Because PDB 2C9L contains the S186A mutation, the Ser[186] side chain conformation for the A' half-site was taken from the A half-site of the ZEBRA/meZRE2 structure. Hydrogen bonds unique to each complex are shown in red and marked by a red asterisk. The unique H-bond in the AP-1 complex between Asn[182] and T^{-3} is accessible to solvent at the periphery of the protein/DNA interface whereas the unique H-bond in the meZRE2 complex between Ser[186] and ^{m}C$^{-2'}$ is buried in the center of the interface. Methyl groups on AP-1 base T^1 and meZRE2 bases ^{m}C^1 and ^{m}C$^{-2'}$ are shown as magenta spheres, highlighting that only the ^{m}C$^{-2'}$ methyl group is unique to meZRE2. (**B**) Accessible surface areas (ASAs) of H-bond donor and acceptor atoms. ASAs (shown in blue) were calculated using the program Areaimol of the CCP4 suite (46). Buried and solvent-accessible atoms are shown in gray and light blue squares, respectively. Arrows indicate H-bond directionality. Arrows in red correspond to H-bonds shown in red in panel (A). The dashed arrows indicate a bifurcated H-bond. The number of H-bonds is identical for the two half-sites if one considers that the bifurcated H-bond mediated by Arg[190] (a three-centered interaction in which a single proton is shared between two acceptor atoms) is comparable in strength to a canonical H-bond (101,102).

effect could reasonably account for the different stabilities of the meZRE2 and AP-1 complexes.

Unlike ZEBRA, the canonical bZIP protein GCN4 binds the AP-1 site more tightly than meZRE2 (Figure 6B). Remarkably, swapping ZEBRA's Ser[186] for an alanine and the corresponding GCN4 Ala[239] for a serine inverted the me/A selectivity of both proteins (Figure 6C–F). These findings agree with previous studies that reported enhanced affinity of the ZEBRA S186A mutant for the AP-1 site (39,41,107), decreased affinity of the same mutant for the meZRE2 site (34,35,42) and enhanced affinity of Fos and Jun for methylated ZRE sites when the corresponding Ala→Ser mutations were made (41,42). The fact that the S186A mutation only modestly reduces me/Z selectivity but dramatically inverts me/A selectivity suggests that the inability of this mutant to induce lytic gene expression and disrupt viral latency is not due to its poorer discrimination of methylated and unmethylated CpG motifs, but rather its sequestration by AP-1 sites preventing recruitment to meZREs. Since transcription factor sequestration by competing DNA binding sites can lead to a threshold-like dose-response of their target promoters (108–110), this may explain the threshold-like behaviour we observe for ZEBRA mutants in our viral lytic activation assays.

The above findings suggest a novel interaction model for understanding ZEBRA's dual transactivating functions during EBV infection that integrates both me/Z and me/A selectivity (Supplementary Figure S9). During prelatency when the incoming genomic EBV DNA is still unmethylated, ZEBRA has low affinity for the unmethylated CpG-containing ZRE sites in lytic viral promoters and preferentially binds methylated CpG-containing cellular ZREs and AP-1 sites, thereby activating genes that promote B cell proliferation and help establish latency (Supplementary Figure S9a, left). Following extensive methylation of the latent viral genome as early as two to three weeks after infection (18), CpG-containing viral ZREs surpass AP-1 sites in their binding affinity for ZEBRA, allowing ZEBRA to activate viral lytic gene expression upon its induced expression at the onset of EBV's lytic phase (Supplementary Figure S9a, right). The ZEBRA S186A mutant fails to activate lytic expression because the decreased affinity for meZRE2 results in sequestering of the mutant protein by competing DNA sequences—including specifically AP-1 sites, which are highly abundant in the human genome (111,112) and whose affinity for ZEBRA is enhanced by the mutation (Supplementary Figure S9b). Indeed, a genome-wide ChIP-seq analysis identified $>5 \times 10^5$ AP-1-like sites bound by ZEBRA when ZEBRA expression was induced in Raji cells (99). Similarly, although cellular AP-1 proteins such as Fos and Jun preferentially bind meZRE2 over ZRE2 sites, they fail to activate viral lytic genes because their affinity for AP-1 sites exceeds that for meZRE2 (Supplementary Figure S9b). This selectivity is inverted by the Ala→Ser mutations at positions equivalent to ZEBRA Ser[186], allowing these proteins to overcome AP-1 site sequestration and activate lytic gene expression (41,113) (Supplementary Figure S9a).

The hydroxyl group at residue 186 that endows ZEBRA with me/A selectivity is strikingly parsimonious from not only a structural but also a molecular evolutionary perspective. Of the six possible serine codons, a TCC codon specifies Ser[186] whereas a GCC codon specifies the corresponding alanine in several of ZEBRA's closest human orthologs (Supplementary Figure S10), suggesting that me/A selectivity may have arisen through a single G→T transition. Given that the most frequent substitution mutation of alanine is to serine (114), bZIP proteins would seem poised to evolve me/A selectivity. The fact that most have conserved the alanine indicates a strong selection pressure against accepting a mutation here. Indeed, an alignment of human bZIP proteins reveals only two exceptions where the alanine is not conserved (Supplementary Figure S11). The first is CREB3 regulatory factor (CREBRF). Like ZEBRA, CREBRF has a TCC-encoded serine (Supplementary Figures S10b and S11) and its ortholog in drosophila associates with CpG-containing motifs (115), raising the possibility that CREBRF may preferentially bind methylated CpG-containing sites, as detailed in the legend of Supplementary Figure S11.

The second exception comprises the CCAAT/enhancer-binding protein (C/EBP) family of bZIP proteins, which have a valine corresponding to Ser[186]. These proteins recognize the C/EBP site (TTGCGCAA), which is bound with enhanced affinity by C/EBPβ when methylated on the central CpG motif (116). C/EBP proteins preferentially recognize the C/EBP site over alternate sequences such as the c/AMP response element (CRE), an 8-bp motif comprising two AP-1 A' half-sites (TGACGTCA, closely resembling the 7-bp AP-1 site TGAGTCA). Mutating the unique valine to alanine greatly enhances the affinity of C/EBPα for the CRE site (117), analogous to how the S186A mutation enhances ZEBRA's affinity for the AP-1 site. Moreover, C/EBPα is sequestered to pericentromeric heterochromatin by 'natural decoy' C/EBP consensus sites located within tandem α-satellite DNA repeats. The Val→Ala mutant reduces sequestration by these decoy sites and permits binding to functional target sites, enhancing the transcriptional output from c/EBPα-responsive promoters (108). Thus, like ZEBRA, C/EBP proteins exhibit two types of site selectivity, one between different methylation states of the same site and the other between two classes of response element, with the latter selectivity altered by mutating the unique valine to alanine.

In conclusion, ZEBRA has hitherto been viewed as an unusual bZIP protein because it could preferentially bind and activate methylated promoters thanks to its unique Ser[186] residue. However, our findings reveal that the functional significance of Ser[186] is not that it enables ZEBRA to selectively bind methylated over unmethylated CpG-containing ZREs—an activity shared with other bZIP proteins—but that it enables ZEBRA to bind methylated ZREs preferentially over AP-1 sites. Knowledge of ZEBRA's two types of site selectivity clarifies our understanding of the competing molecular interactions that govern ZEBRA-dependent gene expression and should facilitate future studies aimed at unravelling ZEBRA's diverse roles in EBV infection and EBV-associated diseases.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structure has been deposited with the Protein Data

bank under accession number 7NX5. NMR data have been deposited in the BMRB database under accession numbers 50847 and 50848.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Greenberg,M.V.C. and Bourc'his,D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.*, **20**, 590–607.
2. Shimbo,T. and Wade,P.A. (2016) Proteins that read DNA methylation. *Adv. Exp. Med. Biol.*, **945**, 303–320.
3. Hodges,A.J., Hudson,N.O. and Buck-Koehntop,B.A. (2019) Cys(2)His(2) zinc finger methyl-CpG binding proteins: getting a handle on methylated DNA. *J. Mol. Biol.*, **432**, 1640–1660.
4. Tate,P.H. and Bird,A.P. (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.*, **3**, 226–231.
5. Domcke,S., Bardet,A.F., Adrian Ginno,P., Hartl,D., Burger,L. and Schübeler,D. (2015) Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*, **528**, 575–579.
6. 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.
7. 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.
8. 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.
9. 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.
10. Yin,Y., Morgunova,E., Jolma,A., Kaasinen,E., Sahu,B., Khund-Sayeed,S., Das,P.K., Kivioja,T., Dave,K., Zhong,F. *et al.* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, **356**, eaaj2239.
11. Zhu,H., Wang,G. and Qian,J. (2016) Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.*, **17**, 551–565.
12. Kribelbauer,J.F., Laptenko,O., Chen,S., Martini,G.D., Freed-Pastor,W.A., Prives,C., Mann,R.S. and Bussemaker,H.J. (2017) Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Rep.*, **19**, 2383–2395.
13. Zuo,Z., Roy,B., Chang,Y.K., Granas,D. and Stormo,G.D. (2017) Measuring quantitative effects of methylation on transcription factor-DNA binding affinity. *Sci. Adv.*, **3**, eaao1799.
14. Buschle,A. and Hammerschmidt,W. (2020) Epigenetic lifestyle of Epstein-Barr virus. *Semin. Immunopathol.*, **42**, 131–142.
15. Farrell,P.J. (2019) Epstein-Barr virus and cancer. *Annu. Rev. Pathol.*, **14**, 29–53.
16. Münz,C. (2019) Latency and lytic replication in Epstein-Barr virus-associated oncogenesis. *Nat. Rev. Microbiol.*, **17**, 691–700.
17. Jochum,S., Ruiss,R., Moosmann,A., Hammerschmidt,W. and Zeidler,R. (2012) RNAs in Epstein-Barr virions control early steps of infection. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E1396–E1404.
18. Kalla,M., Schmeinck,A., Bergbauer,M., Pich,D. and Hammerschmidt,W. (2010) AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 850–855.
19. Tao,Q. and Robertson,K.D. (2003) Stealth technology: how Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection. *Clin. Immunol.*, **109**, 53–63.
20. Laichalk,L.L. and Thorley-Lawson,D.A. (2005) Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. *J. Virol.*, **79**, 1296–1307.
21. Hong,G.K., Gulley,M.L., Feng,W.H., Delecluse,H.J., Holley-Guthrie,E. and Kenney,S.C. (2005) Epstein-Barr virus lytic infection contributes to lymphoproliferative disease in a SCID mouse model. *J. Virol.*, **79**, 13993–14003.
22. Ma,S.D., Hegde,S., Young,K.H., Sullivan,R., Rajesh,D., Zhou,Y., Jankowska-Gan,E., Burlingham,W.J., Sun,X., Gulley,M.L. *et al.* (2011) A new model of Epstein-Barr virus infection reveals an important role for early lytic viral protein expression in the development of lymphomas. *J. Virol.*, **85**, 165–177.
23. Cohen,M., Vistarop,A.G., Huaman,F., Narbaitz,M., Metrebian,F., De Matteo,E., Preciado,M.V. and Chabay,P.A. (2018) Epstein-Barr virus lytic cycle involvement in diffuse large B cell lymphoma. *Hematol. Oncol.*, **36**, 98–103.
24. 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.
25. Feederle,R., Kost,M., Baumann,M., Janz,A., Drouet,E., Hammerschmidt,W. and Delecluse,H.J. (2000) The Epstein-Barr virus lytic program is controlled by the co-operative functions of two transactivators. *EMBO J.*, **19**, 3080–3089.
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. 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 Pathog.*, **6**, e1001114.
28. Flower,K., Thomas,D., Heather,J., Ramasubramanyan,S., Jones,S. and Sinclair,A.J. (2011) Epigenetic control of viral life-cycle by a DNA-methylation dependent transcription factor. *PLoS One*, **6**, e25922.
29. Angel,P., Imagawa,M., Chiu,R., Stein,B., Imbra,R.J., Rahmsdorf,H.J., Jonat,C., Herrlich,P. and Karin,M. (1987) Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell*, **49**, 729–739.
30. Flemington,E. and Speck,S.H. (1990) Autoregulation of Epstein-Barr virus putative lytic switch gene BZLF1. *J. Virol.*, **64**, 1227–1232.
31. Lieberman,P.M. and Berk,A.J. (1990) In vitro transcriptional activation, dimerization, and DNA-binding specificity of the Epstein-Barr virus Zta protein. *J. Virol.*, **64**, 2560–2568.
32. Lieberman,P.M., Hardwick,J.M., Sample,J., Hayward,G.S. and Hayward,S.D. (1990) The zta transactivator involved in induction of lytic cycle gene expression in Epstein-Barr virus-infected lymphocytes binds to both AP-1 and ZRE sites in target promoter and enhancer regions. *J. Virol.*, **64**, 1143–1155.

33. Tsai,S.C., Lin,S.J., Chen,P.W., Luo,W.Y., Yeh,T.H., Wang,H.W., Chen,C.J. and Tsai,C.H. (2009) EBV Zta protein induces the expression of interleukin-13, promoting the proliferation of EBV-infected B cells and lymphoblastoid cell lines. *Blood*, **114**, 109–118.

34. 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.

35. Dickerson,S.J., Xing,Y., Robinson,A.R., Seaman,W.T., Gruffat,H. and Kenney,S.C. (2009) Methylation-dependent binding of the Epstein-Barr virus BZLF1 protein to viral promoters. *PLoS Pathog.*, **5**, e1000356.

36. Tillo,D., Ray,S., Syed,K.S., Gaylor,M.R., He,X., Wang,J., Assad,N., Durell,S.R., Porollo,A., Weirauch,M.T. *et al.* (2017) The Epstein-Barr Virus B-ZIP protein Zta recognizes specific DNA sequences containing 5-methylcytosine and 5-hydroxymethylcytosine. *Biochemistry*, **56**, 6200–6210.

37. Schaeffner,M., Mrozek-Gorska,P., Buschle,A., Woellmer,A., Tagawa,T., Cernilogar,F.M., Schotta,G., Krietenstein,N., Lieleg,C., Korber,P. *et al.* (2019) BZLF1 interacts with chromatin remodelers promoting escape from latent infections with EBV. *Life Sci Alliance*, **2**, e201800108.

38. 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.

39. 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.

40. Heston,L., El-Guindy,A., Countryman,J., Dela Cruz,C., Delecluse,H.J. and Miller,G. (2006) Amino acids in the basic domain of Epstein-Barr virus ZEBRA protein play distinct roles in DNA binding, activation of early lytic gene expression, and promotion of viral DNA replication. *J. Virol.*, **80**, 9115–9133.

41. 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.

42. Hong,S., Wang,D., Horton,J.R., Zhang,X., Speck,S.H., Blumenthal,R.M. and Cheng,X. (2017) Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein-Barr virus Zta. *Nucleic Acids. Res.*, **45**, 2503–2515.

43. 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.

44. Weber,E., Buzovetsky,O., Heston,L., Yu,K.P., Knecht,K.M., El-Guindy,A., Miller,G. and Xiong,Y. (2019) A noncanonical basic motif of Epstein-Barr Virus ZEBRA protein facilitates recognition of methylated DNA, high-affinity DNA binding, and lytic activation. *J. Virol.*, **93**, e00724-19.

45. Kabsch,W. (2010) Xds. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 125–132.

46. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.*, **67**, 235–242.

47. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

48. Liebschner,D., Afonine,P.V., Baker,M.L., Bunkoczi,G., Chen,V.B., Croll,T.I., Hintze,B., Hung,L.W., Jain,S., McCoy,A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.*, **75**, 861–877.

49. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.

50. Kohler,J.J., Metallo,S.J., Schneider,T.L. and Schepartz,A. (1999) DNA specificity enhanced by sequential binding of protein monomers. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 11735–11739.

51. Kohler,J.J. and Schepartz,A. (2001) Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry*, **40**, 130–142.

52. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.

53. Pich,D., Mrozek-Gorska,P., Bouvet,M., Sugimoto,A., Akidil,E., Grundhoff,A., Hamperl,S., Ling,P.D. and Hammerschmidt,W. (2019) First days in the life of naive human B lymphocytes infected with Epstein-Barr Virus. *mBio*, **10**, e01723-19.

54. Gloeckner,C.J., Boldt,K., Schumacher,A., Roepman,R. and Ueffing,M. (2007) A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *Proteomics*, **7**, 4228–4234.

55. Hammerschmidt,W. and Sugden,B. (1988) Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus. *Cell*, **55**, 427–433.

56. Klug,M. and Rehli,M. (2006) Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. *Epigenetics*, **1**, 127–130.

57. Neuhierl,B., Feederle,R., Hammerschmidt,W. and Delecluse,H.J. (2002) Glycoprotein gp110 of Epstein-Barr virus determines viral tropism and efficiency of infection. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15036–15041.

58. Mikaelian,I., Drouet,E., Marechal,V., Denoyel,G., Nicolas,J.C. and Sergeant,A. (1993) The DNA-binding domain of two bZIP transcription factors, the Epstein-Barr virus switch gene product EB1 and Jun, is a bipartite nuclear targeting sequence. *J. Virol.*, **67**, 734–742.

59. Kalla,M., Gobel,C. and Hammerschmidt,W. (2012) The lytic phase of Epstein-Barr virus requires a viral genome with 5-methylcytosine residues in CpG sites. *J. Virol.*, **86**, 447–458.

60. Dirmeier,U., Neuhierl,B., Kilger,E., Reisbach,G., Sandberg,M.L. and Hammerschmidt,W. (2003) Latent membrane protein 1 is critical for efficient growth transformation of human B cells by epstein-barr virus. *Cancer Res.*, **63**, 2982–2989.

61. Steinbruck,L., Gustems,M., Medele,S., Schulz,T.F., Lutter,D. and Hammerschmidt,W. (2015) K1 and K15 of Kaposi's sarcoma-associated herpesvirus are partial functional homologues of latent membrane protein 2A of Epstein-Barr Virus. *J. Virol.*, **89**, 7248–7261.

62. Sista,N.D., Barry,C., Sampson,K. and Pagano,J. (1995) Physical and functional interaction of the Epstein-Barr virus BZLF1 transactivator with the retinoic acid receptors RAR alpha and RXR alpha. *Nucleic Acids Res.*, **23**, 1729–1736.

63. Zhang,Q., Hong,Y., Dorsky,D., Holley-Guthrie,E., Zalani,S., Elshiekh,N.A., Kiehl,A., Le,T. and Kenney,S. (1996) Functional and physical interactions between the Epstein-Barr virus (EBV) proteins BZLF1 and BMRF1: Effects on EBV transcription and lytic replication. *J. Virol.*, **70**, 5131–5142.

64. Hong,Y., Holley-Guthrie,E. and Kenney,S. (1997) The bZip dimerization domain of the Epstein-Barr virus BZLF1 (Z) protein mediates lymphoid-specific negative regulation. *Virology*, **229**, 36–48.

65. Adamson,A.L. and Kenney,S. (1999) The Epstein-Barr virus BZLF1 protein interacts physically and functionally with the histone acetylase CREB-binding protein. *J. Virol.*, **73**, 6551–6558.

66. Wu,F.Y., Wang,S.E., Chen,H., Wang,L., Hayward,S.D. and Hayward,G.S. (2004) CCAAT/enhancer binding protein alpha binds to the Epstein-Barr virus (EBV) ZTA protein through oligomeric interactions and contributes to cooperative transcriptional activation of the ZTA promoter through direct binding to the ZII and ZIIIB motifs during induction of the EBV lytic cycle. *J. Virol.*, **78**, 4847–4865.

67. Bailey,S.G., Verrall,E., Schelcher,C., Rhie,A., Doherty,A.J. and Sinclair,A.J. (2009) Functional interaction between Epstein-Barr virus replication protein Zta and host DNA damage response protein 53BP1. *J. Virol.*, **83**, 11116–11122.

68. Marcourt,L., Cordier,C., Couesnon,T. and Dodin,G. (1999) Impact of C5-cytosine methylation on the solution structure of

d(GAAAACGTTTTC)2. An NMR and molecular modelling investigation. *Eur. J. Biochem.*, **265**, 1032–1042.

69. Pérez,A., Castellazzi,C.L., Battistini,F., Collinet,K., Flores,O., Deniz,O., Ruiz,M.L., Torrents,D., Eritja,R., Soler-López,M. *et al.* (2012) Impact of methylation on the physical properties of DNA. *Biophys. J.*, **102**, 2140–2148.

70. Lazarovici,A., Zhou,T., Shafer,A., Dantas Machado,A.C., Riley,T.R., Sandstrom,R., Sabo,P.J., Lu,Y., Rohs,R., Stamatoyannopoulos,J.A. *et al.* (2013) Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 6376–6381.

71. Rao,S., Chiu,T.P., Kribelbauer,J.F., Mann,R.S., Bussemaker,H.J. and Rohs,R. (2018) Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding. *Epigenetics Chromatin*, **11**, 6.

72. Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.*, **257**, 14686–14707.

73. Privé,G.G., Heinemann,U., Chandrasegaran,S., Kan,L.S., Kopka,M.L. and Dickerson,R.E. (1987) Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science*, **238**, 498–504.

74. Djuranovic,D. and Hartmann,B. (2004) DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers*, **73**, 356–368.

75. Lamoureux,J.S., Stuart,D., Tsang,R., Wu,C. and Glover,J.N. (2002) Structure of the sporulation-specific transcription factor Ndt80 bound to DNA. *EMBO J.*, **21**, 5721–5732.

76. Lamoureux,J.S., Maynes,J.T. and Glover,J.N. (2004) Recognition of 5′-YpG-3′ sequences by coupled stacking/hydrogen bonding interactions with amino acid residues. *J. Mol. Biol.*, **335**, 399–408.

77. Towse,C.L., Rysavy,S.J., Vulovic,I.M. and Daggett,V. (2016) New dynamic rotamer libraries: data-driven analysis of side-chain conformational propensities. *Structure*, **24**, 187–199.

78. Liu,Y., Zhang,X., Blumenthal,R.M. and Cheng,X. (2013) A common mode of recognition for methylated CpG. *Trends Biochem. Sci.*, **38**, 177–183.

79. Ho,K.L., McNae,I.W., Schmiedeberg,L., Klose,R.J., Bird,A.P. and Walkinshaw,M.D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell*, **29**, 525–531.

80. 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.

81. 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.

82. 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.

83. Fersht,A.R., Shi,J.P., Knill-Jones,J., Lowe,D.M., Wilkinson,A.J., Blow,D.M., Brick,P., Carter,P., Waye,M.M. and Winter,G. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, **314**, 235–238.

84. Williams,D.H., Searle,M.S., Mackay,J.P., Gerhard,U. and Maplestone,R.A. (1993) Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 1172–1178.

85. Pace,C.N., Fu,H., Lee Fryar,K., Landua,J., Trevino,S.R., Schell,D., Thurlkill,R.L., Imura,S., Scholtz,J.M., Gajiwala,K. *et al.* (2014) Contribution of hydrogen bonds to protein stability. *Protein Sci.*, **23**, 652–661.

86. Rausch,C., Hastert,F.D. and Cardoso,M.C. (2019) DNA modification readers and writers and their interplay. *J. Mol. Biol.*, **432**, 1731–1746.

87. 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.

88. Namba-Fukuyo,H., Funata,S., Matsusaka,K., Fukuyo,M., Rahmutulla,B., Mano,Y., Fukayama,M., Aburatani,H. and Kaneda,A. (2016) TET2 functions as a resistance factor against DNA methylation acquisition during Epstein-Barr virus infection. *Oncotarget*, **7**, 81512–81526.

89. Hardwick,J.S., Lane,A.N. and Brown,T. (2018) Epigenetic modifications of cytosine: biophysical properties, regulation, and function in mammalian DNA. *Bioessays*, **40**, 1700199.

90. Renciuk,D., Blacque,O., Vorlickova,M. and Spingler,B. (2013) Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.*, **41**, 9891–9900.

91. He,J.J. and Quiocho,F.A. (1991) A nonconservative serine to cysteine mutation in the sulfate-binding protein, a transport receptor. *Science*, **251**, 1479–1481.

92. Cannon,W.R., Briggs,J.M., Shen,J., McCammon,J.A. and Quiocho,F.A. (1995) Conservative and nonconservative mutations in proteins: anomalous mutations in a transport receptor analyzed by free energy and quantum chemical calculations. *Protein Sci.*, **4**, 387–393.

93. Ray,S., Tillo,D., Assad,N., Ufot,A., Deppmann,C., Durell,S.R., Porollo,A. and Vinson,C. (2018) 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. *Biochem. Biophys. Res. Commun.*, **501**, 905–912.

94. Tzamarias,D., Pu,W.T. and Struhl,K. (1992) Mutations in the bZIP domain of yeast GCN4 that alter DNA-binding specificity. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 2007–2011.

95. Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell*, **71**, 1223–1237.

96. 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.

97. Schepers,A., Pich,D. and Hammerschmidt,W. (1993) A transcription factor with homology to the AP-1 family links RNA transcription and DNA replication in the lytic cycle of Epstein-Barr virus. *EMBO J.*, **12**, 3921–3929.

98. Schepers,A., Pich,D. and Hammerschmidt,W. (1996) Activation of oriLyt, the lytic origin of DNA replication of Epstein-Barr virus, by BZLF1. *Virology*, **220**, 367–376.

99. Buschle,A., Mrozek-Gorska,P., Cernilogar,F.M., Ettinger,A., Pich,D., Krebs,S., Mocanu,B., Blum,H., Schotta,G., Straub,T. *et al.* (2021) Epstein-Barr virus inactivates the transcriptome and disrupts the chromatin architecture of its host cell in the first phase of lytic reactivation. *Nucleic Acids Res.*, **49**, 3217–3241.

100. Bianchi,C. and Zangi,R. (2012) How to distinguish methyl-cytosine from cytosine with high fidelity. *J. Mol. Biol.*, **424**, 215–224.

101. Baker,E.N. and Hubbard,R.E. (1984) Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.*, **44**, 97–179.

102. Rozas,I., Alkorta,I. and Elguero,J. (1998) Bifurcated hydrogen bonds: three-centered interactions. *J. Phys. Chem. A*, **102**, 9925–9932.

103. Gao,J., Bosco,D.A., Powers,E.T. and Kelly,J.W. (2009) Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat. Struct. Mol. Biol.*, **16**, 684–690.

104. Klotz,I.M. and Franzen,J.S. (1962) Hydrogen bonds between model peptide groups in solution. *J. Am. Chem. Soc.*, **84**, 3461–3466.

105. Chen,Y.W., Fersht,A.R. and Henrick,K. (1993) Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants. *J. Mol. Biol.*, **234**, 1158–1170.

106. Herschlag,D. and Pinney,M.M. (2018) Hydrogen bonds: simple after all? *Biochemistry*, **57**, 3338–3352.

107. 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.

108. Liu,X., Wu,B., Szary,J., Kofoed,E.M. and Schaufele,F. (2007) Functional sequestration of transcription factor activity by repetitive DNA. *J. Biol. Chem.*, **282**, 20868–20876.

109. Lee,T.H. and Maheshri,N. (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol. Syst. Biol.*, **8**, 576.

110. Kemme,C.A., Nguyen,D., Chattopadhyay,A. and Iwahara,J. (2016) Regulation of transcription factors via natural decoys in genomic DNA. *Transcription*, **7**, 115–120.

111. Zhou,H., Zarubin,T., Ji,Z., Min,Z., Zhu,W., Downey,J.S., Lin,S. and Han,J. (2005) Frequency and distribution of AP-1 sites in the human genome. *DNA Res.*, **12**, 139–150.

112. Kheradpour,P. and Kellis,M. (2014) Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.*, **42**, 2976–2987.

113. Lyons,D.E., Yu,K.P., Vander Heiden,J.A., Heston,L., Dittmer,D.P., El-Guindy,A. and Miller,G. (2018) Mutant cellular AP-1 proteins promote expression of a subset of Epstein-Barr virus late genes in the absence of lytic viral DNA replication. *J. Virol.*, **92**, e01062-18.

114. Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 10915–10919.

115. Tiebe,M., Lutz,M., De La Garza,A., Buechling,T., Boutros,M. and Teleman,A.A. (2015) REPTOR and REPTOR-BP regulate organismal metabolism and transcription downstream of TORC1. *Dev. Cell*, **33**, 272–284.

116. Yang,J., Horton,J.R., Wang,D., Ren,R., Li,J., Sun,D., Huang,Y., Zhang,X., Blumenthal,R.M. and Cheng,X. (2019) Structural basis for effects of CpA modifications on C/EBPβ binding of DNA. *Nucleic Acids Res.*, **47**, 1774–1785.

117. Miller,M., Shuman,J.D., Sebastian,T., Dauter,Z. and Johnson,P.F. (2003) Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J. Biol. Chem.*, **278**, 15178–15184.