ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Altering the Double-Stranded DNA Specificity of the bZIP Domain of Zta with Site-Directed Mutagenesis at N182

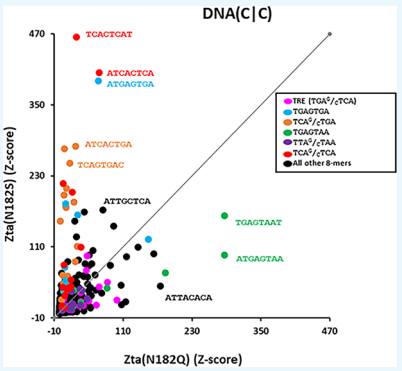
Sreejana Ray, Desiree Tillo, Nima Assad, Aniekanabasi Ufot, Aleksey Porollo, Stewart R. Durell, and Charles Vinson*

Cite This: *ACS Omega* 2022, 7, 129–139

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Zta, the Epstein–Barr virus bZIP transcription factor (TF), binds both unmethylated and methylated double-stranded DNA (dsDNA) in a sequence-specific manner. We studied the contribution of a conserved asparagine (N182) to sequence-specific dsDNA binding to four types of dsDNA: (i) dsDNA with cytosine in both strands ((DNA(C|C)), (ii, iii) dsDNA with 5-methylcytosine (5mC, **M**) or 5-hydroxymethylcytosine (5hmC, **H**) in one strand and cytosine in the second strand ((DNA(5mC|C) and DNA(5hmC|C)), and (iv) dsDNA with methylated cytosine in both strands in all CG dinucleotides ((DNA(5mCG)). We replaced asparagine with five similarly sized amino acids (glutamine (Q), serine (S), threonine (T), isoleucine (I), or valine (V)) and used protein binding microarrays to evaluate sequence-specific dsDNA binding. Zta preferentially binds the pseudo-palindrome TRE (AP1) motif ($\mathbf{T^{-4}G^{-3}A^{-2G}/_C{}^0T^2C^3A^4}$). Zta (N182Q) changes binding to $\mathbf{A^3}$ in only one half-site. Zta(N182S) changes binding to $\mathbf{G^3}$ in one or both halves of the motif. Zta(N182S) and Zta(N182Q) have 34- and 17-fold weaker median dsDNA binding, respectively. Zta(N182V) and Zta(N182I) have increased binding to dsDNA(5mC|C). Molecular dynamics simulations rationalize some of these results, identifying hydrogen bonds between glutamine and $\mathbf{A^3}$, but do not reveal why serine preferentially binds $\mathbf{G^3}$, suggesting that entropic interactions may mediate this new binding specificity.



1. INTRODUCTION

The bZIP motif occurs in a family of eukaryotic proteins with sequence-specific dsDNA binding properties.[1,2] The bZIP domain is a long bipartite alpha helix.[1] The C-terminal is a leucine zipper coiled-coil region that homodimerizes and/or heterodimerizes.[3−7] The N-terminal interacts with the major groove of dsDNA in a sequence-specific manner. Five conserved amino acids in the bZIP dsDNA-binding region interact with the nucleotides (NXXASXXCR).[8−10] We will focus on the invariant asparagine (N). Asparagine can form two hydrogen bonds with adenine.[11,12] In the bZIP motif, the asparagine[2] forms two hydrogen bonds with adjacent nucleotides on opposite strands of dsDNA.

In all X-ray structures where the bZIP dimer binds the $\mathbf{T^2C^3A^4}$ trinucleotide, including the CREB complex with the CRE dsDNA palindrome ($\mathbf{T^{-4}G^{-3}A^{-2}C^{-1}G^1T^2C^3A^4}$),[13] and GCN4,[3] Fos-Jun,[4] and Zta[10] binding the shorter pseudo-palindrome TRE ($\mathbf{T^{-4}G^{-3}A^{-2G}/_C{}^0T^2C^3A^4}$), the asparagine side-chain carbonyl oxygen accepts a hydrogen from N4 of $\mathbf{C^3}$ and the amide nitrogen donates a hydrogen to O4 of $\mathbf{T^{-4}}$ on the opposite strand.[14]

Asparagine in the bZIP domain of CEBPA binding to dsDNA recognizes the $\mathbf{A^3A^4}$ dinucleotide in the motif $\mathbf{T^{-4}T^{-3}G^{-2}C^{-1}G^1C^2A^3A^4}$.[15] Asparagine has a hydrogen bond interaction with O4 of $\mathbf{T^{-4}}$ as observed for all the CRE and TRE structures, but now, the asparagine side-chain carbonyl oxygen accepts a hydrogen from the N6 amino group of $\mathbf{A^3}$ that is stereochemically similar to N4 of $\mathbf{C^3}$.[15] The Pap1-dsDNA complex binds the $\mathbf{T^3A^4}$ dinucleotide (**TAACGT-T^{3}A**),[8] and the asparagine does not contact any of the nucleotides. Figure S1 describes the interaction of these bZIP domains with their preferred DNA motifs and illustrates the differences among them.

Mutagenesis experiments using a few dsDNA sequences have determined that the conserved asparagine contributes to sequence-specific dsDNA binding.[16] In addition, recent studies show that Zta binds dsDNA with modified cytosine. Zta binds many TRE variants with methylated cytosine (**M**) replacing thymine at two positions ($\mathbf{M^{-4}G^{-3}}$ and $\mathbf{M^2G^3}$).[10,17] Zta also binds dsDNA with the methylated C/EBP half-site ($\mathbf{GM^2AA}$) or its oxidative product 5-hydroxymethylcytosine (5hmC, **H**)[10,18] or 5-formylcytosine,[19] similar to CREB1.[20] The impact

Received: August 3, 2021
Accepted: November 23, 2021
Published: December 28, 2021

ACS Publications

© 2021 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

of the conserved asparagine on biding to methylated cytosine is unknown.

We have extended these studies and examined five mutants of the conserved asparagine (Zta(N182Q), Zta(N182S), Zta(N182I), Zta(N182V), and Zta(N182T)) binding four types of dsDNA including DNAs containing modified cytosines using protein binding microarrays (PBMs).[18,21]

2. RESULTS

2.1. PBMs with Four Types of dsDNA and Data Analysis. Protein binding microarray (PBM) experiments used Agilent DNA microarrays containing 40,330 different single-stranded DNA 60-mers[22] in 16 sectors on a glass microarray slide.[18] We examined the sequence-specific dsDNA binding of six GST-Zta chimeric constructs, Zta and five single amino acid mutants (Q, S, I, V, or T) of Zta(N182) to four types of dsDNA. T7 DNA polymerase was used to generate three types of dsDNAs: (i) dsDNA with cytosine on both strands (DNA(C|C)) and (ii, iii) dsDNA with 5mC or 5hmC in one strand and cytosine in the second strand ((DNA(5mC|C) and DNA(5hmC|C)).[18,20,23] The fourth type of dsDNA was generated by enzymatic methylation of both cytosines in all CG dinucleotides ((DNA(5mCG)).[24] Equal amounts of *in vitro* synthesized chimeric protein containing GST and bZIP domains (Figure S2) were bound to these four types of dsDNA.[18] Binding was detected using a Cy5 conjugated antibody to the GST epitope followed by the measurement of fluorescence intensities at each of the array features.

We evaluated the PBM data in five ways. First, for an overall quantitative measure of binding, we examined the fluorescence intensities of the 40,330 features in each of the 16 sectors on the glass slide. Second, we examined the dependence on nucleotide and dinucleotide composition on binding. Third, we determined a standardized score (*Z*-score)[25] for binding 8 bp long dsDNA sequences (8-mers). The *Z*-score for a given 8-mer is the number of standard deviations of the intensity of that 8-mer (computed from all features containing the 8-mer) that is from the global median intensity (computed from all array features). This measure is sensitive to changes in median binding. Fourth, we generated dsDNA logos[18] using the 50 strongest bound 8-mers. Fifth, we examined the effects of single nucleotide variants (SNVs) on binding the canonical TRE ($\mathbf{T^{-4}G^{-3}A^{-2G}/_C{}^0T^2C^3A^4}$) motif 7-mer and related 7-mers containing one or no cytosines. This allows us to determine the contribution to binding of all nucleotides, including modified cytosines, at each position in the motif.[18] We will first describe the results for each of the five methods to binding unmodified dsDNA and then binding to three dsDNAs containing modified cytosine.

2.2. Zta(N182) Mutants Binding with Unmodified dsDNA(C|C). *2.2.1. Binding to 40,330 Features Containing dsDNA(C|C).* We compared previously published data for Zta binding to unmodified dsDNA[18] and newly generated data for Zta and five Zta(N182) mutants. Figure S8A presents newly determined data for Zta (*x* axis), and Zta(N182Q) (*y* axis) binding the 40,330 dsDNA features on the array showing maximal Zta(N182Q) binding is about half less than Zta, with median binding (a measure of non-specific binding) being 17-fold weaker (Table S3). Zta(N182Q) binds some features stronger than Zta, indicative of new sequence-specific dsDNA binding (e.g., **GTGAGTAA**, **TGAGTAAT**, etc.) (Figure S8A and Figure 1A). Zta(N182S) binding is about half that of Zta, with median binding being 34-fold weaker (Figure S8B and Table S3). Again, some features are more strongly bound by Zta(N182S) (e.g., **TCACTCAT**, **ATCACTCA**, **ATGAGTGA**, **ATCACTGA**, **TCAGTGAC**, etc.) than Zta, indicating new sequence-specific dsDNA binding (Figure S8B and Figure 1B). Zta(N182T) binds unmodified dsDNA features 3.3-fold weaker than Zta but with similar specificity (Figure S8C and Figure 1C). Non-polar mutant Zta(N182I) and Zta(N182V) median binding to unmodified dsDNA are similar to and 2.6-fold stronger than Zta, respectively (Table S3). However, they do not bind with specificity (Figure S11A–D), indicative of

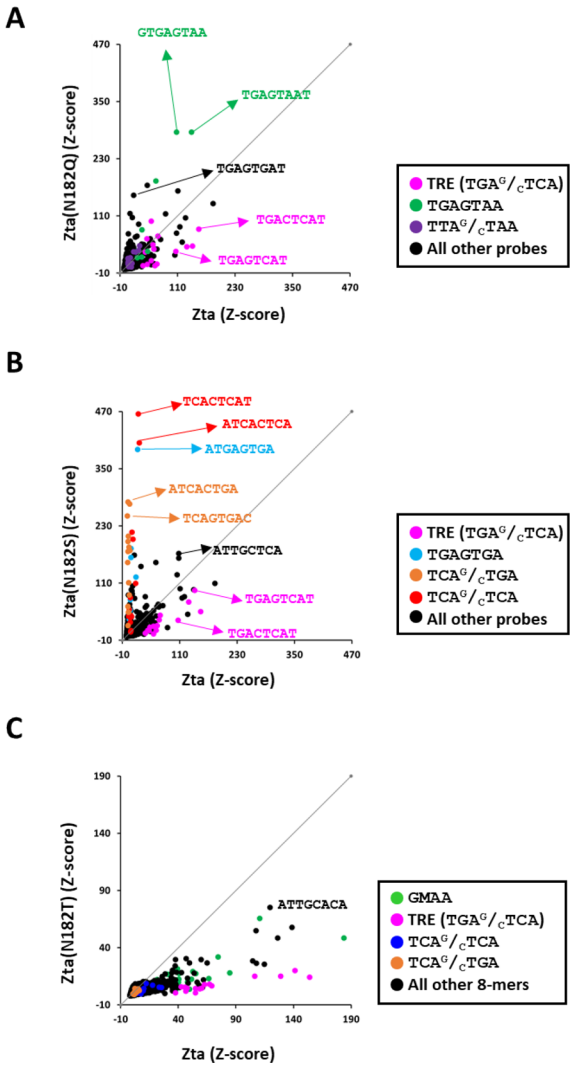


Figure 1. Zta, Zta(N182Q), Zta(N182S), and Zta(N182T) mutants binding to DNA(C|C). (A) Scatter graph comparison of PBM 8-mer *Z*-scores of Zta (*x* axis) with Zta(N182Q) (*y* axis) dsDNA features containing cytosine in both strands (DNA(C|C)). dsDNA features and 8-mers are colored into several groups: those containing the TRE ($\mathbf{TGA^G/_CTCA}$) and TRE variants changing $\mathbf{C^3}$ to $\mathbf{G^3}$ in one or both half-sites. Specific 8-mers are highlighted, and their sequences are indicated with an arrow head that is strongly bound by the Zta(N182Q) mutant and that is preferred by the wild-type Zta. (B) Same as in panel (A) but for Zta(N182S) (*y* axis). Features and 8-mers are colored into several groups: those containing the TRE and variants changing $\mathbf{C^3}$ to $\mathbf{A^3}$. Similarly, some 8-mers features are highlighted that are preferred by the Zta(N182S) mutant and that are preferred by the thr wild type. (C) Same as in panel (A) but for Zta(N182T) (*y* axis). Features and 8-mers are colored into several groups.

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

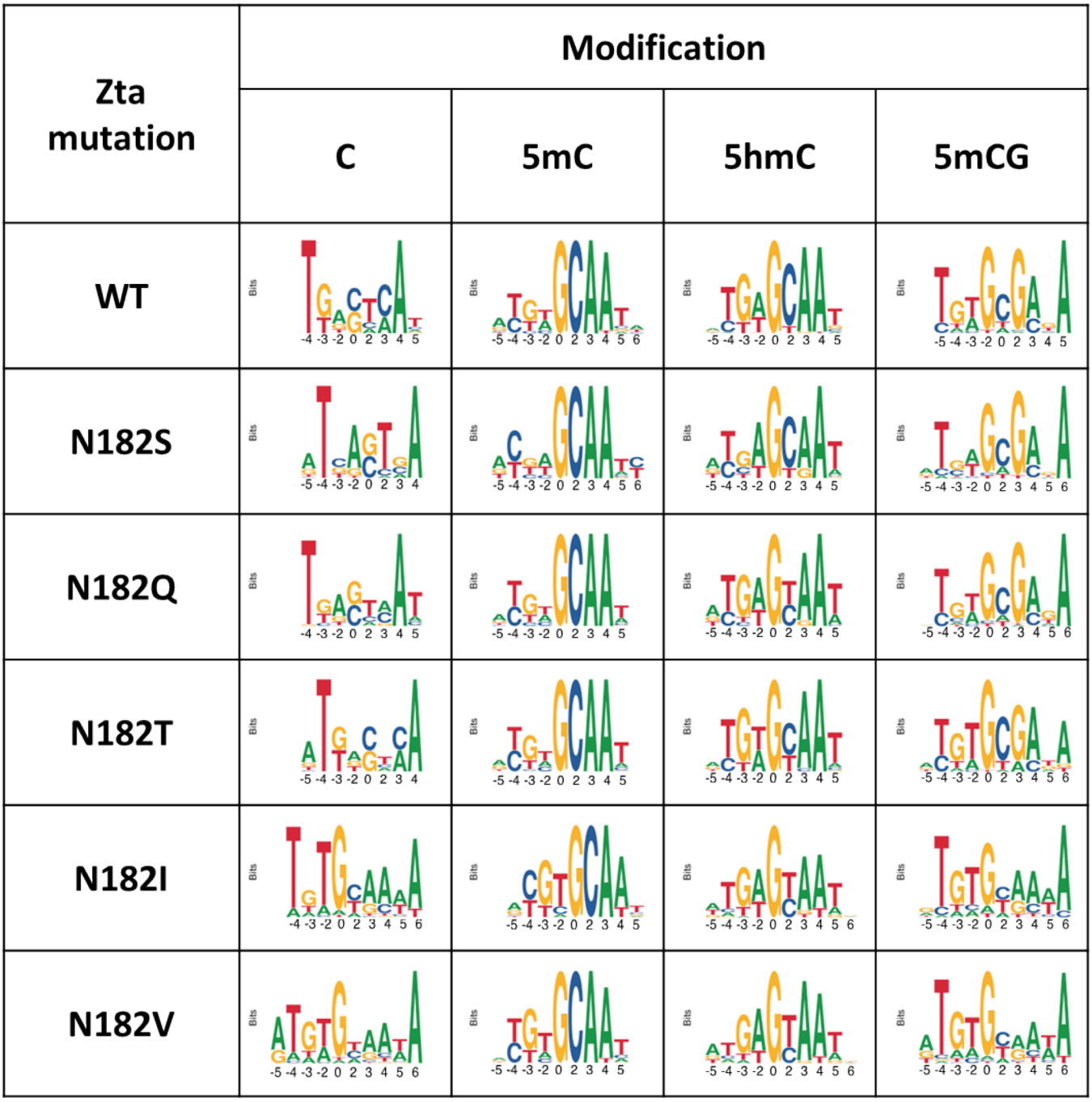


Figure 2. Sequence logos generated for well-bound 8-mers of Zta and the Zta(N182) mutant for each DNA type. Logos were generated using the top 50 8-mers (by *Z*-score) for each PBM experiment.

non-specific binding and consistent with the importance of hydrogen bonds or other electrically polarized interactions for affinity. The decreased median bindings for Zta(N182Q) and Zta(N182S) contribute to the increased skew in the distribution of binding intensities, indicating increased specific dsDNA binding (Figure S3 and Table S4).

2.2.2. Binding to Mono- and Dinucleotides. An intriguing aspect of this analysis is that median dsDNA binding is dramatically weakened for Zta(N182Q) and Zta(N1872S). Table S3 highlights that the weakening of median binding is only observed for Zta(N182Q) and Zta(N182S) binding DNA(ClC) and DNA(5mCG). To evaluate the reduction in median binding of Zta(N182Q) and Zta(N182S), we examined binding to mono- and dinucleotides, hypothesizing that asparagine may bind the adenine mononucleotide or the **CA** dinucleotide as observed in several protein-dsDNA structures.[12,26,27] Each feature contains a 60-mer where 24 nucleotides near the glass surface are common to each feature, and the remaining 36 nucleotides are variable. We counted the occurrences of each mono- and dinucleotide in the variable 36-mer for the 90% of the weakest bound features (Figures S12 and S13). For Zta, the more A:T base pairs, the stronger the binding, saturating at approximately 20 adenines in the random 36-mer. This is not observed for Zta(N182Q) and Zta-(N182S). Dinucleotides containing adenine except **AG** have similar shapes to the adenine mononucleotide (Figure S13).

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

Table 1. SNV Tables of *Z*-Scores for (A) Zta Binding to the Canonical TRE Motif (TGAG^{0}TC^{3}A), (B, C) Zta(N182Q) Binding to Two Variants of the TRE Motif, and (D, E) Zta(N182S) Binding to Two Other Variants of the TRE Motif

Zta

A

Zta TGAGTC^{3}A	-4 T	-3 G	-2 A	0 G	2 T	3 C	4 A
A	4	6	**97**	3	14	60	**97**
C	4	10	16	97	14	**97**	2
G	5	**97**	6	**97**	3	12	4
T	**97**	23	28	3	**97**	9	2

Zta(N182Q)

B

Zta(N182Q) TGAGTA^{3}A	-4 T	-3 G	-2 A	0 G	2 T	3 A	4 A
A	6	6	**61**	0	3	**61**	**61**
C	3	44	8	19	37	49	5
G	3	**61**	4	**61**	2	22	1
T	**61**	24	25	3	**61**	7	2

C

Zta(N182Q) TGACTA^{3}A	-4 T	-3 G	-2 A	0 C	2 T	3 A	4 A
A	1	4	**19**	0	5	**19**	**19**
C	0	11	1	**19**	6	49	3
G	1	**19**	4	61	8	25	1
T	**19**	24	3	3	**19**	7	3

Zta(N182S)

D

Zta(N182S) TGAGTG^{3}A	-4 T	-3 G	-2 A	0 G	2 T	3 G	4 A
A	9	6	**138**	3	2	40	**138**
C	6	196	12	44	10	32	11
G	3	**138**	9	**138**	1	**138**	1
T	**138**	35	20	1	**138**	5	8

E

Zta(N182S) TGACTG^{3}A	-4 T	-3 G	-2 A	0 C	2 T	3 G	4 A
A	1	6	**44**	3	6	12	**44**
C	2	196	1	**44**	2	32	5
G	-1	**44**	3	138	3	**44**	1
T	**44**	47	12	1	**44**	3	2

Again, this is not observed for Zta(N182Q) and Zta(N182S) helping to explain their lower median binding. Zta(N182V) and Zta(N182I) have a strengthening in binding with increased adenine, potentially reflecting the hydrophobic amino acid side interacting with the methyl group of thymine, the complement of adenine.

2.2.3. Binding to 8-mers (Z-Scores). We next compared *Z*-scores for dsDNA 8-mers, the approximate length of sequence-specific binding for the bZIP motif. Zta(N182Q) has higher *Z*-scores than Zta and binds variants of the TRE where **C^{3}** is changed to **A^{3}** in one half-site with **G^{0}** being preferred (**TGAG^{0}TA^{3}A**) (Figure S8A and Figure 1A). Zta(N182S) also has higher *Z*-scores than Zta (Figure 1B). Zta(N182S) changes binding to motifs where **C^{3}** is changed to **G^{3}** in one or both halves of the pseudo-palindromic TRE motif (**T^{-4}G^{-3}A^{-2G}/$_{C}$^{0}T^{2}C^{3}A^{4}**)[10,18,28] (Figure 1B). Zta(N182T) has little change in binding specificity (Figure 1C).

2.2.4. Sequence Logos. Sequence logos were derived from the 50 strongest bound 8-mers (Figure 2). Logos for binding to DNA(ClC) highlight that both **A^{4}** and **T^{-4}** are important for binding for both Zta and mutants.

2.2.5. Zta, Zta(N182Q), and Zta(N182S) Binding Single Nucleotide Variants (SNVs) of Five 7-mers. SNVs of five 7-mers were examined, the TRE (**TGAG^{0}TC^{3}A**) and four 7-mers where **C^{3}** is changed to either **G^{3}** or **A^{3}** in one half of the motif with either **G^{0}** or **C^{0}** at the center of the motif (Table 1A and Table S5). For the TRE that is palindromic except for the central nucleotide, if one replaces the central **G^{0}** with a **C^{0}**, then one simply switches the values in the two half-sites. Replacing **T^{-4}** or **A^{4}** is catastrophic for binding. Zta binding SNVs of the TRE highlights that binding the two half-sites is different. With **G^{0}**, there is a modest preference for **C^{3}** (*Z*-score = 97) compared to **A^{3}** (*Z*-score = 60), while on the other half of the motif, there is much greater specificity with **G^{-3}** being preferred (*Z*-score = 97) compared to **T^{-3}** (*Z*-score = 23). The four variant 7-mers highlight how **G^{0}** or **C^{0}** affects binding to 7-mers with either **G^{3}** or **A^{3}**. The central base pair has little effect on binding **TGA^{G}/$_{C}$^{0}TG^{3}A**, while **G^{0}** is preferred in binding **TGA^{G}/$_{C}$^{0}TA^{3}A** (*Z*-score = 60) compared to **C^{0}** (*Z*-score = 23).

For Zta(N182Q) (Table 1B,C and Table S6), we highlight two 7-mers, **TGAG^{0}TA^{3}A** and **TGAC^{0}TA^{3}A**, which change **C^{3}** to **A^{3}** with either **G^{0}** or **C^{0}**. The preferential binding of Zta(N182Q) for **A^{3}** only occurs with **G^{0}**. This change in binding specificity only occurs on one side of the motif. Starting with **TGAG^{0}TA^{3}A**, **G^{-3}** (*Z*-score = 61) is preferred to **T^{-3}** (*Z*-score = 24).

For Zta(N182S) (Table 1D,E and Table S7), we highlight two 7-mers, **TGAG^{0}TG^{3}A** and **TGAC^{0}TG^{3}A**. These 7-mers change the TRE **C^{3}** to **G^{3}** with either **G^{0}** or **C^{0}**. For both 7-mers, **G^{3}** is preferred. For both 7-mers, **C^{-3}** is also preferred, indicating that Zta(N182S) changes binding specificity in both

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

halves of the motif. If one half of the TRE motif changes to $\mathbf{G}^3$, then $\mathbf{G}^0$ (*Z*-score = 138) is preferred to $\mathbf{C}^0$ (*Z*-score = 44).

2.3. Zta(N182) Mutants Binding Three Modified dsDNAs. For dsDNA containing 5mC (**M**) in one strand, Zta(N182Q), like Zta, binds strongest to 8-mers containing the C/EBP half-site **GC^{2}AA**[10,18] (Table S3 and Figure S16C and Figure 3A). A few 8-mers better bound by Zta(N182Q) than Zta include **AM^{-4}GTGTAA** and **M^{-4}GTGTAAT**, which are similar to the oriLyt motif, **T^{-4}GTGTAA**[29] (Figure 3A). Strong binding to 8-mers containing **M^{-4}** has also been observed for C/EBP|ATF4 heterodimers.[24] Zta(N182Q) binding to dsDNA with 5hmC|C (**H**) in one strand or dsDNA(5mCG) is more variable (Figure S9B,C). Zta(N182Q) binds DNA(5hmC|C) less well than Zta (Figure S9B and Table S3) with several exceptions, including the 8-mer with the **CG** dinucleotide **AH^{-4}GTGTAA** and the 8-mer without the **CG** dinucleotide **ATH^{-3}AGTAA** (Figure 3B) being preferentially bound. Zta(N182Q) binding to DNA(5mCG) is weaker than Zta (Figure S9C and Table S3). meZRE2 sequences[17] are poorly bound by Zta(N182Q), and there are few 8-mers (e.g., **GM^{-4}GTGM^{2}GA** and **AT-CAGM^{2}GA**) that are well bound by Zta(N182Q) (Figure 3C and Figure S18C). Similar results occur with Zta(N182S) (Table S3 and Figures S10A−C and S16B−S18B and Figure 4A−C).

Zta(N182I) and Zta(N182V) bind DNA(5mC|C) with similar specificity but stronger than Zta (Figures S14A,D and S15A,D). Like Zta, these three mutants strongly bind 8-mers containing the methylated C/EBP half-site **GC^{2}AA** (e.g., **AMGTGM^{2}AA** and **MGTGM^{2}AAT**) (Figure S16D,E). They do not bind the other modified dsDNAs strongly (Figures S14B,C,E,F, S15B,C,E,F, S17D,E, and S18D,E). A similar result was observed for Zta(N182T) except DNA(5mCG) where some 8-mers bind dsDNA(5mCG) similar to Zta (Figure S18F).

Sequence logos derived from the 50 strongest bound 8-mers highlight the change in binding specificity for the mutants (Figure 2). Logos for binding to DNA(C|C) highlight that both **A^4** and **T^{-4}** are important for binding for Zta mutants similar to Zta. Nucleotides closer to the center of the motif are less conserved. For both 5mC and 5hmC, the tetranucleotide **GC^{2}AA** is preferred for both Zta and the mutants. For DNA(5mCG), **GC^{2}GA** is dominant for both Zta and the polar mutants. For the hydrophobic mutants, there is no **CG** dinucleotide in the logos.

2.4. Similarities and Differences of Zta and Mutants Binding Four Different dsDNAs. We next examined the similarities and differences for Zta and Zta(N182) mutations binding four types of dsDNAs using heatmaps summarizing all pairwise correlations between mutants and dsDNAs (Figures S19 and S20). The most different binding between WT Zta and the mutants is to unmodified dsDNA DNA(C|C) (Figure S19A). DNA(5mC|C) is bound similarly by Zta and all five mutants as exhibited by the high positive correlations obtained across all pairwise comparisons (Figure S19B). This can be rationalized because binding to DNA(5mC|C) is dominated by the methylation of **C^2** in 8-mers containing **GC^{2}AA**. Position 182 in Zta is not near methyl **C^2**, and changes in amino acids would not be expected to change binding to methylated **GC^{2}AA**. For DNA(5hmC|C) and DNA(5mCG) 8-mers, binding specificity aligns with the amino acid properties of the N182 mutant: the polar amino acids are similar, and the two hydrophobic amino acids are similar (Figure S19D).

When we examine each protein binding the four dsDNAs (Figure S20), DNA(5mC|C) and DNA(5hmC|C) do not change the binding specificity of the polar side-chain mutants (Figure S20C,D) while this is not the case for the hydrophobic mutants as exhibited by the low correlation between these DNA types (Figure S20F). 8-mers containing DNA(5mCG) changes binding specificity of the mutants with the polar side chains (no correlation between DNA(5mCG) and all other DNA types; Figure S20C), while it does not for the

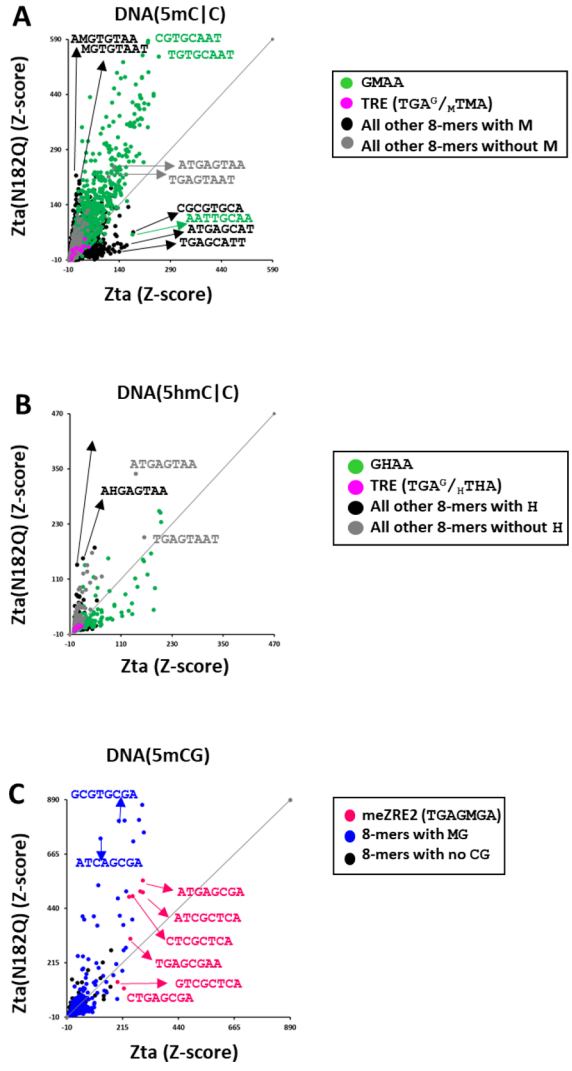


Figure 3. Zta and Zta(N182Q) mutant binding modified dsDNA. Scatter graph comparison of PBM 8-mers *Z*-scores for wild-type Zta (*x* axis) with mutant Zta(N182Q) (*y* axis) binding to DNA(5mC|C) (A). (B) Same as in panel (A) but for binding to DNA(5hmC|C). dsDNA features and 8-mers are color coded in several groups: those containing methylated or hydroxy-methylated C/EBP half-site **GC^{2}AA**, TRE (**TGA^{G}/$_{C}$TCA**), remaining all other 8-mers with cytosine and 8-mers without cytosines. (C) Same as in panel (A) but for binding to DNA(5mCG). dsDNA features and 8-mers are color coded in three groups: those containing meZRE2 (**TGAGMGA**), with methylated CG dinucleotide, and 8-mers without CG dinucleotide (black). All arrays were scanned with identical laser settings (100PMT), with the exception of DNA(5mC|C) and DNA(5hmC|C) measurements (50 and 10PMT, respectively). A selection of 8-mers is highlighted, and their sequences are provided.

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

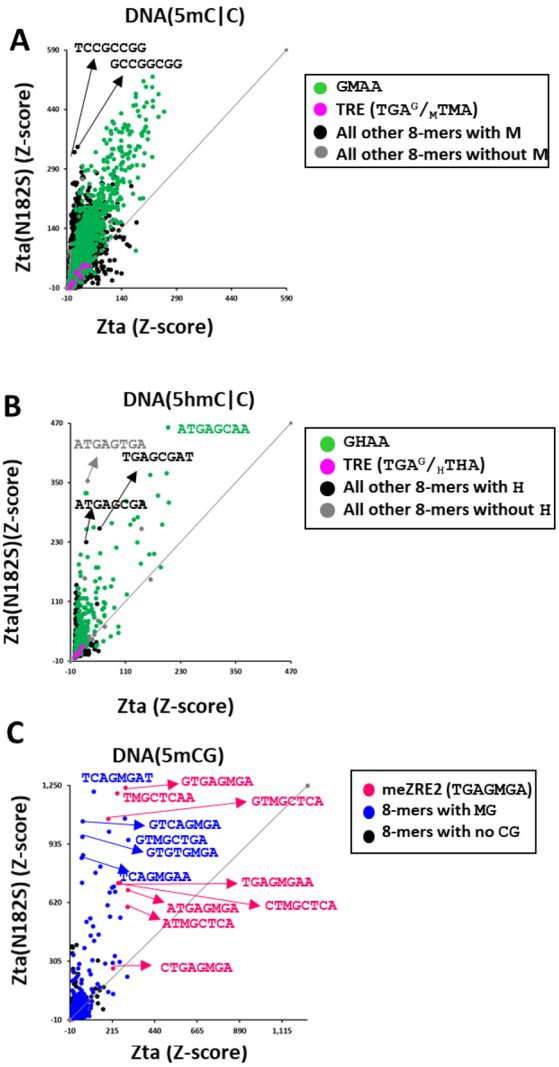


Figure 4. Zta and Zta(N182S) mutant binding modified dsDNA. Scatter graph comparison of PBM 8-mers *Z*-scores for wild-type Zta (*x* axis) with mutant Zta(N182S) (*y* axis) binding to DNA(5mC|C) (A). (B) Same as in panel (A) but for binding to DNA(5hmC|C). dsDNA features and 8-mers are color coded in several groups: those containing methylated or hydroxy-methylated C/EBP half-site **GC^{2}AA**, TRE (**TGA^{G}/$_{C}$TCA**), remaining all other 8-mers with cytosine and 8-mers without cytosines. (C) Same as in pane l (A) but for binding to DNA(5mCG). dsDNA features and 8-mers are color coded in three groups: those containing meZRE2 (**TGAGMGA**), with methylated CG dinucleotide, and 8-mers without CG dinucleotide (black). All arrays were scanned with identical laser settings (100PMT), with the exception of DNA(5mC|C) and DNA(5hmC|C) measurements (50 and 10PMT, respectively). A selection of 8-mers is highlighted, and their sequences are provided.

hydrophobic side-chain mutants (high correlation between DNA types for hydrophobic mutants; Figure S20F).

Binding comparison across the two polar and two non-polar substitutions of Zta(N182) reveals that the two polar mutants Zta(N182S) and Zta(N182Q) bind similarly with dsDNA with modified cytosines, while with DNA(C|C), their binding specificity varies (Figure S21A–D). Zta(N182V) and Zta-(N182I) bind with similar specificity to all four types of dsDNA (Figure S21E–H).

2.5. Structural Analysis. We performed a series of all-atom molecular dynamics (MD) simulations to place these binding data in a structural framework. Initially, nine simulations were performed, three proteins (Zta, Zta(N182Q), and Zta(N182S)) each binding three dsDNA sequences. We started with the X-ray structure of Zta bound to dsDNA containing the TRE (**T^{-4}G^{-3}A^{-2G}/$_{C}$^{0}T^{2}C^{3}A^{4}**) (PDB ID: 2C9N).[9] Then, we changed the dsDNA sequence to (**TT^{-3}A^{G}/$_{C}$TA^{3}A**) and (**TC^{-3}A^{G}/$_{C}$TG^{3}A**) where **C^{3}** is replaced with **A^{3}** or **G^{3}** in both halves of the pseudo-palindrome motif. The frequencies of hydrogen bonds between the N182, N182Q, and N182S side chains and the three dsDNA sequences over the microsecond trajectories are shown in Table 2. The rows are delimited by the three amino acid side chains and the columns by **C^{3}**, **A^{3}**, and **G^{3}**. Mostly, only hydrogen bonds with at least a 10% frequency in one of the three trajectories are considered.

Following the format of the crystal structure, the two protein segments are designated by Y and Z and the two single-stranded DNA (ssDNA) chains by A and B; protein Y primarily binds ssDNA chain A, and Z primarily binds B. As a control, two additional simulations that started from the end of the trajectories of the Zta(N182Q) and Zta(N182S) bound to the TRE. Despite different starting side-chain configurations from the crystal structure, the asparagine side chains of both Zta(N182Q) and Zta(N182S) quickly reorientated and regained the hydrogen bonding profile of the wild-type system.

The upper left-hand portion of Table 2 shows that asparagine in both monomers of Zta makes two hydrogen bonds: (i) the asparagine side-chain oxygen (OD1) with one of the N4-bound hydrogens (H4) of **C^{3}** and (ii) one of the side chain ND2-bound hydrogens (HD2) of asparagine with the O4 oxygen of the **T^{-4}** base on the opposite strand. These are the same hydrogen bonds that occur in the starting crystal structure, which attests to the stability of the simulation. They have also been described in an independent X-ray study of cocrystals.[10] A typical snapshot of this groove-spanning configuration is shown in Figure S22A. These two pairs of hydrogen bonds persist throughout most of the trajectory (with frequencies of 0.95 and 0.88 for one half-site and 0.68 and 0.57 for the other). An interesting observation is that the frequencies are greater for one half-site than the other. This may indirectly reflect the asymmetry at the center of the motif, with **C^{0}** in one DNA strand and **G^{0}** in the other strand.

As indicated in Figure 1 and Table 1A, the next most stable 8-mer bound by Zta is **TGAGTA^{3}A**, with a *Z*-score of 60 vs 97 for binding the TRE. As seen in Table 2, this substitution of **C^{3}** for **A^{3}** causes a change in the hydrogen bonding pattern of the asparagine from groove-spanning to a bidentate interaction with **A^{3}**. The OD1 and HD2 atoms of the asparagine side chain now bind with the H6 and N7 atoms of the adenine, respectively. A typical snapshot of this configuration is shown in Figure S22B. As seen in the top row right column of Table 2, the asparagine does not form any long-lasting hydrogen bonds to dsDNA containing **G^{3}**, consistent with the relatively low *Z*-scores of 12 and 10 for the two half-site complexes (Table 1A).

The second row in Table 2 shows the results of the simulations of Zta(N182Q) bound to three dsDNA sequences. Glutamine in Zta(N182Q) bound to the TRE (first column) forms weaker hydrogen bonds with **C^{3}** and **T^{-4}** than asparagine. Glutamine in the Zta(N182Q)/DNA-**A^{3}** complex has stronger interactions with **T^{-4}** and a strong interaction with **A^{3}** helping to explain the change in binding specificity. Interestingly, the strength of the Zta(N182Q)/DNA-**A^{3}**

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

Table 2. Hydrogen Bond Frequencies Among the N182, N182Q, and N182S Zta Proteins with C^3, A^3, and G^3 Forms of TRE dsDNA over 1 ms Molecular Dynamics Trajectories[a]

	DNA		
protein	C^3	A^3	G^3
Zta(N182)	Y OD1 − A C^3 H4 0.95	Y OD1 − A A^3 H6 0.53	Y HD2 − A G^3 N7 0.14
	Y OD1 − B C^3 H4 0.68	Z OD1 − B A^3 H6 0.44	Y HD2 − B G^3 N7 0.15
		Y HD2 − A A^3 N7 0.65	
		Y HD2 − B A^3 N7 0.64	
	Y HD2 − B T^{-4} O4 0.88	Y HD2 − B T^{-4} O4 0.10	
	Z HD2 − A G^{-4} O4 0.57	Y HD2 − A T^{-4} O4 0.01	
Zta(N182Q)	Y OE1 − A C^3 H4 0.19	Y OE1 − A A^3 H6 0.04	Y HE2 − A G^3 O6 0.13
	Z OE1 − B C^3 H4 0.16	Z OE1 − B A^3 H6 0.004	Z HE2 − B G^3 O6 0.19
		Y HE2 − A A^3 N7 0.32	Y HE2 − A G^3 N7 0.30
		Z HE2 − B A^3 N7 0.28	Z HE2 − B G^3 N7 0.41
	Y HE2 − A C^{-4} O4 0.47	Y HE2 − B T^{-4} O4 0.43	Y HE2 − B T^{-4} O4 0.53
	Z HE2 − A T^{-4} O4 0.14	Z HE2 − A T^{-4} O4 0.35	Z HE2 − A T^{-4} O4 0.67
	Y OE1 − B A^{-5} H6 0.29	Y OE1 − B A^{-5} H6 0.33	Y OE1 − B A^{-5} H6 0.36
	Y OE1 − A C^{-5} H4 0.14	Z OE1 − A C^{-5} H4 0.26	Z OE1 − A C^{-5} H4 0.55
	Y HE2 − B A^{-5} N7 0.24		
Zta(N182S)	Y OG − A C^3 H4 0.20	Y OG − A A^3 H6 0.03	
	Y OG − B C^3 H4 0.15	Y OG − B A^3 H6 0.61	
		Y HG1 − B A^3 N7 0.57	

[a]In accordance with the starting crystal structure (PDB ID: 2C9N) notation, the two protein strands are indicated by Y and Z and the two DNA strands by A and B. Except in a few cases to show the asymmetry, only hydrogen bonds with a minimum frequency of 0.1 are shown.

complex is also found to be dependent on whether the central base of the motif is G^0 or C^0 (with *Z*-scores of 61 vs 19, respectively). Like asparagine, the one methylene longer glutamine can form bidentate hydrogen bonds with adenine,[12,26,27] but in these simulations, glutamine like asparagine is forming hydrogen bonds, though less frequently, with both C^3 and T^{-4}.

We conducted an additional series of MD simulations of Zta(N182Q) in the complex with the asymmetric motifs **TGAG^{0}TA^{3}A** and **TGAC^{0}TA^{3}A** that are strongly bound. Taking into account the complementary sequences, this provided four unique half-site complexes: Zta(N182Q) with **G^{0}TA^{3}A**, **G^{0}TC^{3}A**, **C^{0}TA^{3}A**, and **C^{0}TC^{3}A**. As seen in Table S8, the results are consistent with the experimental data. In particular, the greatest hydrogen bonding residence frequencies occur for Zta(N182Q) binding the **G^{0}TA^{3}A** half-site (Y HE2-B T^{-4} O4 0.57 and Y HE2-A with A^3, N7 0.46), even greater than the symmetric sequence containing A^3 in both half-sites (strongest monomer data: Y HE2-B T^{-4} O4 0.43 and Y HE2-A with A^3, N7 0.32). Replacing the central G^0 base with C^0 results in a reduction in hydrogen bonds for both **C^{0}TA^{3}A** and **C^{0}TC^{3}A**.

The third row of Table 2 presents the simulation results for Zta(N182S) binding three dsDNA sequences. Experimentally, Zta(N182S) binds strongest to dsDNA with G^3 and G^0 at the center (i.e., the **G^{0}TG^{3}A** half-site) with a *Z*-score of 138. This is not clearly reflected in the hydrogen bonding profiles. Rather, the only significantly long-lasting interactions are two, bidentate hydrogen bonds between the serine and A^3 in one half-site. Two infrequent, single hydrogen bonds occur with C^3 in both half-sites, and no significantly resident hydrogen bonds are observed with the G^0 DNA. Given that the serine side chain is shorter than asparagine, it is not surprising that it does not reach as far into the major groove to interact with the nucleotide bases. This lack of bonding could explain why the alpha helical basic region of the Zta(N182S) protein moved away from the DNA over the course of the trajectory. This raises the possibility that the specificity is derived not from hydrogen bonds with the base pair containing G^3 but from entropy. It should be noted that the paucity of interactions seen here is in contrast to several examples of serine residues interacting with guanine and other nucleotides in the X-ray structures of other protein/DNA complexes.[12,26,27]

3. DISCUSSION

We used a PBM platform[25] to measure the bZIP domain of Zta and N182 mutants (Q, S, T, V, and I) binding to four types of dsDNA: cytosine in both strands, 5mC or 5hmC in one dsDNA strand and cytosine in the second strand,[20] and 5mC in both cytosines of all CG dinucleotides.[24] The mutants were selected to have side chains with a similar size to asparagine with either polar (serine, glutamine, and threonine) or hydrophobic (isoleucine and valine) properties. The greatest change in sequence-specific dsDNA binding was to unmodified dsDNA by Zta(N182Q) and Zta(N182S). These two mutants also had weak median binding, suggestive of increased binding specificity. We observe similarities and differences between mutants in binding DNAs containing modified cytosines.

A molecular dynamics evaluation of Zta binding dsDNA indicates that asparagine 182 forms two hydrogen bonds with nucleotides in the major groove of dsDNA, as suggested by X-ray structures.[10] Zta(N182Q) preferentially binds A^3, and the molecular dynamics simulations indicate that glutamine forms hydrogen bonds with A^3. This type of interaction has been described for both asparagine and glutamine interacting with adenine.[11] Zta(N182S) preferentially binds G^3, but no interactions are identified in our molecular dynamics simulations. Entropic forces may be dominating the preferential interaction with G^3. Calorimetric experiments indicate that the bZIP domain of GCN4 binding to the TRE (**TCA^G/$_C$^{0}TGA**) is enthalpic, while binding to the CRE **TCACGTGA** has a stronger entropic effect,[30−32] highlighting the complexity of the issue of sequence-specific dsDNA binding.

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

Potentially, the new dsDNA binding specificity is the result of avoiding interactions with the nucleotides, thus driving binding specificity. Negative design is a term that captures this avoidance property.[33,34] Negative design is a concept in protein folding in which an unfavorable interaction in the disordered state drives folding.[35,36] The use of negative design to create binding specificity can also explain some protein−protein interactions. Both attractive and repulsive interactions are observed between charged amino acids in the **e** and **g** positions of the heptads in the leucine zipper coiled coil that lie over the hydrophobic core.[5] Repulsive energies are stronger than attractive energies (coupling energy).[37] Negative design is also observed in the hydrophobic core of leucine zipper coiled coil.[38] These two negative design parameters help create the distinct interaction properties of bZIP proteins in *Drosophila*[39] and *Arabidopsis*.[40]

Sequence-specific dsDNA binding of proteins was initially suggested to involve hydrogen bonds between amino acid side chains and the nucleotides in the major groove,[41] creating a direct readout or code.[42] Subsequently, the X-ray co-crystal structure of the Trp repressor/operator did not reveal direct interactions between the amino acid side chains and nucleotides but instead interactions between the protein and the backbone of DNA that indirectly drove sequence-specific dsDNA binding.[43] A more recent work has shown that enthalpic or entropic forces can drive sequence-specific dsDNA binding.[44] The specificity of protein-dsDNA interactions has been examined for many protein domains.[45−47] These studies typically mutate the protein and probe for new dsDNA binding activity. The PBM platform allows direct examination of dsDNA binding to thousands of sequences, but repulsive interactions are hard to identify. A potential example of the negative design in protein-dsDNA interactions is the expansion of C2H2-ZF TFs in metazoans.[48] The more ancient zinc fingers have more interactions with the nucleotides, while more recent zinc fingers have more interactions with the DNA backbone. This allows base-contacting specificity residues to mutate without a catastrophic loss of DNA binding. This overdesign in binding affinity to the backbone allows interactions with the nucleotides to be repulsive, potentially generating a kaleidoscope of dsDNA binding specificity.

In summary, changing asparagine Zta(182) to either glutamine or serine creates new dsDNA binding specificity to unmodified dsDNA. These Zta(N182) mutants could be interesting biological probes to explore how different sequences in the genome are accessed.

4. EXPERIMENTAL PROCEDURES

4.1. Cloning, Expression, and Single Amino Acid Mutations of the Zta bZIP DNA-Binding Domain. The N-terminal GST fusion construct of the wild-type DNA binding domain of viral bZIP protein Zta with approximately 50 flanking amino acids was obtained in a modified pDEST15 MAGIC vector.[49] The cloned amino acid sequence of the Zta bZIP domain is shown here with the alpha helical DNA binding region in bold and N182 underlined.

Zta: STVQTAAAVVFACPGANQGQQLADIGVPQPAPV-AAPARRTRKPQQPESLEECDS**ELEIKRYKNRVASRKCR-AKFKQLLQ**HYREVAAAKSSENDRLRLLLKQMCP-SLDVDSIIPRTPDVLHEDLLNF. We chose replacement amino acids that were of comparable size to N182 (https://www.genome.jp/dbget-bin/www_bget?aaindex:CHOC760101). Mutant constructs of Zta (N182Q, N182S, N182I, N182V, and N182T) were generated by site-directed mutagenesis of the GST-fused wild-type Zta plasmid construct (GenScript, USA). Zta constructs were expressed using a PURExpress *In vitro* protein synthesis kit (NEB) previously described.[24] Similar amounts of protein synthesis were confirmed by western blot using an anti-GST-HRP conjugated antibody (Figure S2).

4.2. PBM Experiments. We used the "HK" array design (NCBI Gene Expression Omnibus (GEO) platform, GPL11260) for all PBM experiments. Single-stranded DNA 60-mers on the array were double-stranded using T7 DNA polymerase using either cytosine, 5mC (NEB), or 5hmC (Zymo Research).[20] Enzymatic methylation of CG dinucleotides on PBMs (DNA(5mCG))[24] and protein binding experiments[18,19,50−52] were performed as described previously.

4.3. Fluorescence Extraction and Data Processing. For each PBM, a microarray image was generated using an Agilent Sure Scan II scanner and analyzed using the ImaGene extraction software (BioDiscovery Inc.). Data quantification and *Z*-score calculations were performed as previously described.[18,53] All mutant proteins bound dsDNA strongly (Figure S3). Each protein was assayed with replicates in good agreement ($R > 0.8$) (Table S1 and Figures S4−S7). Arrays with the fewest saturated spots were used for further analysis. *Z*-scores for 5mC, 5mCG, and 5hmC PBM data were rescaled relative to unmodified cytosine using the slope of the line of the best fit computed from the *Z*-scores using 8-mers without cytosine (Table S2). Data (raw probe intensities and 8-mer *Z*-scores) are available at the NCBI GEO database under accession numbers GSE115351 and GSE115352.

4.4. Molecular Dynamics Simulations. All simulations were based on the PDB ID: 2C9N crystal structure.[9] Mutations of the N182 amino acid side chain and C3 nitrogenous base were done with UCSF-Chimera 1.13.1.[54] Molecular dynamics trajectories were generated with NAMD 2.13b2[55] using the CHARMM36 all-hydrogen topologies and parameters.[56] The protein/DNA complexes were centered in a periodic-boundary cell of 94 × 60 × 49 Å initial dimensions, which allowed for a minimum 12 Å gap to any wall. VMD-1.9.3[57] was used to fill the surrounding space with TIPS3P-model waters and sodium and chloride ions necessary to neutralize the system and provide a salt concentration of 150 mM. Electrostatic and VDW energy functions were calculated with a CUTOFF of 12 Å, a SWITCHDIST of 10 Å, and a PAIRLISTDIST of 14 Å. The Langevin and Langevin−Piston algorithms were used to maintain the system temperature and pressure at 300.0 °K and 1.0 ATM, respectively, with the defaults for the other parameters. Rigid bonds were used to allow for an integration timestep of 2 fs. Trajectories were run for a minimum of 1000 ns, with snapshots collected every 10 ps. Hydrogen bonds and other interactions were studied with CHARMM.[58]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c04148.

Supporting tables and figures that include PBM data sources, a summary of PBM experiments, comparison of hydrogen bonds among conserved asparagine side chains and bound DNA nucleotides, western blot of *in vitro* translation reactions, and additional data (PDF)

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

AUTHOR INFORMATION

Corresponding Author

Charles Vinson − *Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-3708-7416; Phone: 240-760-6885; Email: vinsonc@mail.nih.gov; Fax: 1-301-496-8419

Authors

Sreejana Ray − *Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-1955-1131

Desiree Tillo − *Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States; Cancer Genetics Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0003-3568-6148

Nima Assad − *Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

Aniekanabasi Ufot − *Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

Aleksey Porollo − *Center for Autoimmune Genomics and Etiology, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, United States; Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267, United States*

Stewart R. Durell − *Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-7806-0359

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c04148

Author Contributions

S.R.: Experiment designing and conducting all the experiments, data analysis, manuscript preparation, and overall editing. D.T.: Manuscript preparation and overall editing. N.A.: Data analysis and manuscript preparation. A.U.: Data analysis. A.P.: Experiment idea. S.D.: Data analysis and manuscript preparation. C.V.: Project idea, experiment designing, manuscript preparation, and overall editing.

Notes

The authors declare no competing financial interest.

Proteins used in this study: PDB ID: 2C9N, DOI: 10.2210/pdb2C9N/pdb, NDB: PD0938.

ACKNOWLEDGMENTS

We thank Matt Weirauch and Marcus Noyes for ideas about dsDNA binding specificity.

ABBREVIATION

M, 5-methylcytosine; **H**, 5-hydroxymethylcytosine; PBM, protein binding microarray; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA

REFERENCES

(1) Vinson, C. R.; Sigler, P. B.; McKnight, S. L. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **1989**, *246*, 911−916.

(2) Jindrich, K.; Degnan, B. M. The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity. *BMC Evol Biol* **2016**, *16*, 28.

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

(4) Glover, J. N. M.; Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **1995**, *373*, 257−261.

(5) Krylov, D.; Mikhailenko, I.; Vinson, C. A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J* **1994**, *13*, 2849−2861.

(6) Vinson, C.; Myakishev, M.; Acharya, A.; Mir, A. A.; Moll, J. R.; Bonovich, M. Classification of human B-ZIP proteins based on dimerization properties. *Mol. Cell. Biol.* **2002**, *22*, 6321−6335.

(7) Newman, J. R. S.; Keating, A. E. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* **2003**, *300*, 2097−2101.

(8) Fujii, Y.; Shimizu, T.; Toda, T.; Yanagida, M.; Hakoshima, T. Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nat. Struct. Biol.* **2000**, *7*, 889−893.

(9) Petosa, C.; Morand, P.; Baudin, F.; Moulin, M.; Artero, J. B.; Müller, C. W. Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. *Mol. Cell* **2006**, *21*, 565−572.

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

(11) Jakubec, D.; Laskowski, R. A.; Vondrasek, J. Sequence-Specific Recognition of DNA by Proteins: Binding Motifs Discovered Using a Novel Statistical/Computational Analysis. *PLoS One* **2016**, *11*, No. e0158704.

(12) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **2001**, *29*, 2860−2874.

(13) Schumacher, M. A.; Goodman, R. H.; Brennan, R. G. The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J Biol Chem* **2000**, *275*, 35242−35247.

(14) Konig, P.; Richmond, T. J. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **1993**, *233*, 139−154.

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

(16) Suckow, M.; Schwamborn, K.; Kisters-Woike, B.; von Wilcken-Bergmann, B.; Muller-Hill, B. Replacement of invariant bZip residues within the basic region of the yeast transcriptional activator GCN4 can change its DNA binding specificity. *Nucleic Acids Res.* **1994**, *22*, 4395−4404.

(17) Karlsson, Q. H.; Schelcher, C.; Verrall, E.; Petosa, C.; Sinclair, A. J. Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein. *PLoS Pathog.* **2008**, *4*, No. e1000005.

(18) Tillo, D.; Ray, S.; Syed, K. S.; Gaylor, M. R.; He, X.; Wang, J.; Assad, N.; Durell, S. R.; Porollo, A.; Weirauch, M. T.; Vinson, C. The Epstein-Barr Virus B-ZIP Protein Zta Recognizes Specific DNA Sequences Containing 5-Methylcytosine and 5-Hydroxymethylcytosine. *Biochemistry* **2017**, *56*, 6200−6210.

(19) Ray, S.; Tillo, D.; Ufot, A.; Assad, N.; Durell, S.; Vinson, C. bZIP Dimers CREB1, ATF2, Zta, ATF3|cJun, and cFos|cJun Prefer to Bind to Some Double-Stranded DNA Sequences Containing 5-

Formylcytosine and 5-Carboxylcytosine. *Biochemistry* **2020**, *59*, 3529−3540.
(20) Syed, K. S.; He, X.; Tillo, D.; Wang, J.; Durell, S. R.; Vinson, C. 5-Methylcytosine (5mC) and 5-Hydroxymethylcytosine (5hmC) Enhance the DNA Binding of CREB1 to the C/EBP Half-Site Tetranucleotide GCAA. *Biochemistry* **2016**, *55*, 6940−6948.
(21) Ray, S.; Tillo, D.; Assad, N.; Ufot, A.; Deppmann, C.; Durell, S. R.; Porollo, A.; Vinson, C. 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.* **2018**, *501*, 905−912.
(22) Lam, K. N.; van Bakel, H.; Cote, A. G.; van der Ven, A.; Hughes, T. R. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res.* **2011**, *39*, 4680−4690.
(23) Sayeed, S. K.; Zhao, J.; Sathyanarayana, B. K.; Golla, J. P.; Vinson, C. C/EBPbeta (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochim. Biophys. Acta* **2015**, *1849*, 583−589.
(24) Mann, I. K.; Chatterjee, R.; Zhao, J.; He, X.; Weirauch, M. T.; Hughes, T. R.; Vinson, C. CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Res.* **2013**, *23*, 988−997.
(25) Berger, M. F.; Philippakis, A. A.; Qureshi, A. M.; He, F. S.; Estep, P. W., 3rd; Bulyk, M. L. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **2006**, *24*, 1429−1435.
(26) Mandel-Gutfreund, Y.; Schueler, O.; Margalit, H. Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *J. Mol. Biol.* **1995**, *253*, 370−382.
(27) Suzuki, M. A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure* **1994**, *2*, 317−326.
(28) Farrell, P. J.; Rowe, D. T.; Rooney, C. M.; Kouzarides, T. Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos. *EMBO J* **1989**, *8*, 127−132.
(29) Schelcher, C.; Valencia, S.; Delecluse, H. J.; Hicks, M.; Sinclair, A. J. Mutation of a single amino acid residue in the basic region of the Epstein-Barr virus (EBV) lytic cycle switch protein Zta (BZLF1) prevents reactivation of EBV from latency. *J. Virol.* **2005**, *79*, 13822−13828.
(30) Berger, C.; Jelesarov, I.; Bosshard, H. R. Coupled folding and site-specific binding of the GCN4-bZIP transcription factor to the AP-1 and ATF/CREB DNA sites studied by microcalorimetry. *Biochemistry* **1996**, *35*, 14984−14991.
(31) Dragan, A. I.; Frank, L.; Liu, Y.; Makeyeva, E. N.; Crane-Robinson, C.; Privalov, P. L. Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites. *J. Mol. Biol.* **2004**, *343*, 865−878.
(32) Privalov, P. L.; Dragan, A. I.; Crane-Robinson, C. Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. *Nucleic Acids Res.* **2011**, *39*, 2483−2491.
(33) Fleishman, S. J.; Baker, D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **2012**, *149*, 262−273.
(34) Richardson, J. S.; Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 2754−2759.
(35) Jin, W.; Kambara, O.; Sasakawa, H.; Tamura, A.; Takada, S. De novo design of foldable proteins with smooth folding funnel: automated negative design and experimental verification. *Structure* **2003**, *11*, 581−590.
(36) Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* **1996**, *10*, 27−34.
(37) Krylov, D.; Barchi, J.; Vinson, C. Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* **1998**, *279*, 959−972.
(38) Acharya, A.; Rishi, V.; Vinson, C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* **2006**, *45*, 11324−11332.
(39) Fassler, J.; Landsman, D.; Acharya, A.; Moll, J. R.; Bonovich, M.; Vinson, C. B-ZIP proteins encoded by the Drosophila genome: evaluation of potential dimerization partners. *Genome Res.* **2002**, *12*, 1190−1200.
(40) Deppmann, C. D.; Acharya, A.; Rishi, V.; Wobbes, B.; Smeekens, S.; Taparowsky, E. J.; Vinson, C. Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs. *Nucleic Acids Res.* **2004**, *32*, 3435−3445.
(41) Seeman, N. C.; Rosenberg, J. M.; Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 804−808.
(42) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233−269.
(43) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **1988**, *335*, 321−329.
(44) Morgunova, E.; Yin, Y.; Das, P. K.; Jolma, A.; Zhu, F.; Popov, A.; Xu, Y.; Nilsson, L.; Taipale, J., Two distinct DNA sequences recognized by transcription factors represent enthalpy and entropy optima. *Elife* **2018**, *7*, DOI: 10.7554/eLife.32963.
(45) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **2000**, *8*, 1015−1023.
(46) Havranek, J. J.; Duarte, C. M.; Baker, D. A simple physical model for the prediction and design of protein-DNA interactions. *J. Mol. Biol.* **2004**, *344*, 59−70.
(47) Bogdanove, A. J.; Bohm, A.; Miller, J. C.; Morgan, R. D.; Stoddard, B. L. Engineering altered protein-DNA recognition specificity. *Nucleic Acids Res.* **2018**, *46*, 4845−4871.
(48) Najafabadi, H. S.; Garton, M.; Weirauch, M. T.; Mnaimneh, S.; Yang, A.; Kim, P. M.; Hughes, T. R. Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc finger DNA binding. *Genome Biol.* **2017**, *18*, 167.
(49) Sharrocks, A. D. A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase. *Gene* **1994**, *138*, 105−108.
(50) Berger, M. F.; Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat. Protoc.* **2009**, *4*, 393−411.
(51) Badis, G.; Berger, M. F.; Philippakis, A. A.; Talukder, S.; Gehrke, A. R.; Jaeger, S. A.; Chan, E. T.; Metzler, G.; Vedenko, A.; Chen, X.; Kuznetsov, H.; Wang, C. F.; Coburn, D.; Newburger, D. E.; Morris, Q.; Hughes, T. R.; Bulyk, M. L. Diversity and complexity in DNA recognition by transcription factors. *Science* **2009**, *324*, 1720−1723.
(52) Ray, S.; Ufot, A.; Assad, N.; Singh, J.; Durell, S. R.; Porollo, A.; Tillo, D.; Vinson, C. The bZIP mutant CEBPB (V285A) has sequence specific DNA binding propensities similar to CREB1. *Biochim Biophys Acta Gene Regul Mech* **2019**, *1862*, 486−492.
(53) Khund-Sayeed, S.; He, X.; Holzberg, T.; Wang, J.; Rajagopal, D.; Upadhyay, S.; Durell, S. R.; Mukherjee, S.; Weirauch, M. T.; Rose, R.; Vinson, C. 5-Hydroxymethylcytosine in E-box motifs ACAT|GTG and ACAC|GTG increases DNA-binding of the B-HLH transcription factor TCF4. *Integr Biol (Camb)* **2016**, *8*, 936−945.
(54) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139

(55) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(56) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(57) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14*, 33−38. 27-8
(58) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

https://doi.org/10.1021/acsomega.1c04148
ACS Omega 2022, 7, 129−139