ACS OMEGA

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

Article

REL Domain of NFATc2 Binding to Five Types of DNA Using Protein Binding Microarrays

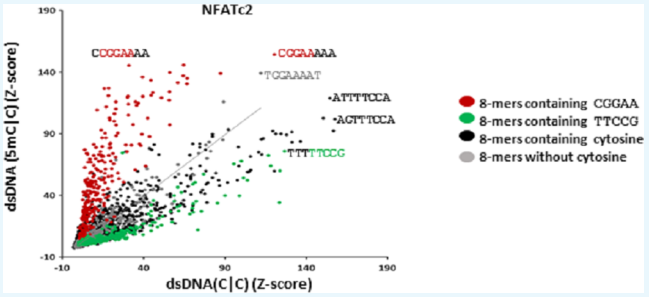
Sreejana Ray, Desiree Tillo, Stewart R. Durell, Syed Khund-Sayeed, and Charles Vinson*

Cite This: *ACS Omega* 2021, 6, 4147−4154

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: NFATc2 is a DNA binding protein in the Rel family transcription factors, which binds a **CGGAA** motif better when both cytosines in the **CG** dinucleotide are methylated. Using protein binding microarrays (PBMs), we examined the DNA binding of NFATc2 to three additional types of DNA: single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) with either 5-methylcytosine (5mC, **M**) or 5-hydroxymethylcytosine (5hmC, **H**) in one strand and a cytosine in the second strand. **ATTTCCAC**, the complement of the core **GGAA** motif, is better bound as ssDNA compared to dsDNA. dsDNA containing the 5-mer **CGGAA** with either 5mC or 5hmC in one DNA strand is bound stronger than **CGGAA**. In contrast, the reverse complement **TTCCG** is bound weaker when it contains 5mC. Analysis of the available NFATc2:dsDNA complexes rationalizes these PBM data.



1. INTRODUCTION

*N*uclear *F*actor of *A*ctivated *T*-cells *c*ytoplasmic 2, or NFATc2[1] (also known as NFAT1, NFAT1a, or NFATP)[2] is a member of the calcium-responsive NFAT family of transcription factors (TFs). NFATc2 is expressed in many somatic tissues including immune and endothelial cells[2−5] and is involved in the regulation of cellular processes, including cell cycle regulation, T-cell differentiation and activation, and development.[2,6−8] Because of its importance in development, dysregulation of NFAT results in malignancies and other pathologies.[3,9,10]

NFAT contains two separate functional domains, the NFAT-homology region (NHR) that is involved in calcium binding and subcellular localization and the REL-homology region (RHR) that is a sequence-specific DNA binding domain.[11] The RHR is composed of two immunoglobulin folds, RHR-N involved in DNA binding and RHR-C that mediates homo- or heterodimerization.[12] All NFAT TFs bind DNA as monomers to the core NFAT DNA binding motif $\mathbf{G^1G^2A^3A^4A^5A^6}$.[11] NFAT TFs can bind cooperatively with other TFs including AP-1,[13] GATA4,[14] IRF4,[15,16] FOXP3,[17] and MEF2.[18] For clarity, we present DNA sequences using the bold font.

Several structures of monomeric NFATc2 bound with double-stranded DNA (dsDNA) have been solved[12,19,20] and amino acids that interact with both strands in the major groove of dsDNA have been highlighted. Two arginines in the loop regions of RHR-N (R421 and R430) form bidentate hydrogen bond interactions in the major groove with $\mathbf{G^1}$ and $\mathbf{G^2}$ of the NFAT DNA motif.[21] The amide side chain of Q571 forms hydrogen bonds with $\mathbf{A^3}$ of the core motif.[22] Several amino acids' N-terminal of R421, R430, and Q571 interacts with the thymines in the complementary strand $\mathbf{T^{6'}T^{5'}T^{4'}T^{3'}C^{2'}C^{1'}}$ (nucleotide bases in the complementary strand are denoted with a "′"). Y424 interacts with $\mathbf{T^{3'}}$ and $\mathbf{T^{4'}}$, and R572 interacts with $\mathbf{T^{5'}}$ and $\mathbf{T^{6'}}$ via van der Waals contacts, and both R572 and Y424 form hydrogen bonds with the DNA backbone.[11]

It was shown that NFATc2 can bind strongly to sequences containing 5-mer $\mathbf{C^{-1}G^1G^2A^3A^4}$ when a cytosine in both strands of the $\mathbf{C^{-1}G^1}$ dinucleotide is methylated (e.g., $\mathbf{M^{-1}G^1G^2A^3A^4}$ in one strand and $\mathbf{T^{4'}T^{3'}C^{2'}M^{1'}G^{-1}}$ in the second strand),[23] expanding the DNA sequences bound by NFATc2. To further explore the range of DNA binding of NFATc2, we performed protein binding microarray (PBM) experiments[24] with microarrays containing three different types of DNA, single-stranded DNA (ssDNA), dsDNA with 5mC in one strand and a cytosine in the second strand (dsDNA(5mC|C)), and dsDNA with 5hmC in one strand and a cytosine in the second strand (dsDNA(5hmC|C)),[25] and compared these data to previous data of NFATc2 binding to dsDNA with cytosine in both strands (dsDNA(C|C)) and where both cytosines in all CG dinucleotides contain 5mC (dsDNA-(5mCG)).[23] Previously, we have examined several bZIP family TFs binding with these modified dsDNAs and identified the critical roles of conserved amino acids in their sequence-

Received: August 23, 2020
Accepted: December 25, 2020
Published: February 2, 2021

https://dx.doi.org/10.1021/acsomega.0c04069
ACS Omega 2021, 6, 4147−4154

specific DNA binding ability.[26,27] Here, the contribution of 5mC and 5hmC in one strand of dsDNA to NFATc2 binding was examined.

2. RESULTS

2.1. Protein Binding Microarrays with Four Types of DNA. Previously, we examined NFATc2 binding to two types of dsDNA using PBMs, dsDNA where both strands contain a cytosine (dsDNA(C|C)) and dsDNA where both cytosines in all CG dinucleotides are methylated (dsDNA(5mCG)).[23] NFATc2 preferentially bound many 8-mers containing a methylated **CG** dinucleotide in 5-mer **CGGAA**.[23] To learn more about NFATc2 binding to different types of DNA, we examined NFATc2 binding to three additional forms of DNA using PBMs. The 16 grids of the Agilent HK design 40k DNA microarray[28,29] were divided into four chambers containing four grids each using a gasket slide from Agilent (Figure S1). One chamber was left as ssDNA and the three other chambers were double-stranded using either a cytosine producing dsDNA(C|C), or 5mC producing dsDNA(5mC|C), or 5hmC producing dsDNA(5hmC|C).[30−33] NFATc2 was bound to two grids in each of the four sectors. Examination of binding to all four types of DNA on a single microarray slide allows for comparisons between DNA types. These new data for NFATc2 binding dsDNA(C|C)[23] agree with our previously published data (R = 0.8, Figures S2 and S3E).

2.2. Rel Domain of NFATc2 Binding to Four Types of DNA. We evaluated NFATc2 binding to four types of DNA by examining the median and average binding to the 40,000 DNA features on each sector of the Agilent microarray [Figure 1 (replicate data 1), Figure S4 (replicate data 2), and Table S1]. The difference reflects sequence-specific DNA binding. The median binding of NFATc2 to ssDNA (2256, Figure 1) is higher than the three types of dsDNA, indicating that NFATc2 preferentially binds ssDNA. However, the average binding intensities for dsDNA(C|C) (2830) and dsDNA(5mC|C) (3404) are higher compared to ssDNA (2351), indicating more sequence-specific binding. This is observed in the longer right tail of all the three dsDNA distributions (Figure 1 and Table S1), indicating that NFATc2 is binding specifically to some dsDNA. Both average and median binding is highest for dsDNA with 5mC, followed by a cytosine, and finally 5hmC. A search for the most enriched DNA motifs in the best-bound 1% of the 40,000 PBM features for all the four DNA types finds the core motif 5-mer **TTTCC** as the best bound sequence in both ssDNA and dsDNA(C|C). The preference for binding ssDNA **TTTCC** suggests that these interactions are stronger than the interactions with the opposite strand **GGAAA**. The best bound features for dsDNA(5mC|C) and dsDNA(5hmC|C) contain the 5-mer **MGGAA** and **HGGAA** (Table S2).

2.2.1. NFATc2 Binding to ssDNA and dsDNA(C|C). We next calculated the relative binding affinities (Z-scores)[29] for NFATc2 binding to all 8-mers. While other 8-mer measures have been developed such as the rank-based E-score,[34,35] we prefer to use Z-scores as they are directly associated with binding affinity and give an idea of the specificity of the protein under investigation. Figure 2 (Figure S5A for replicate data) presents a scatter plot comparing NFATc2 Z-scores to all DNA 8-mers for dsDNA(C|C) (x-axis) or ssDNA (y-axis). The 8-mers are divided into two groups, those that contain 5-mer **TTTCC**, which is enriched in the top bound array features for both ssDNA and dsDNA(C|C), and all other 8-mers. For

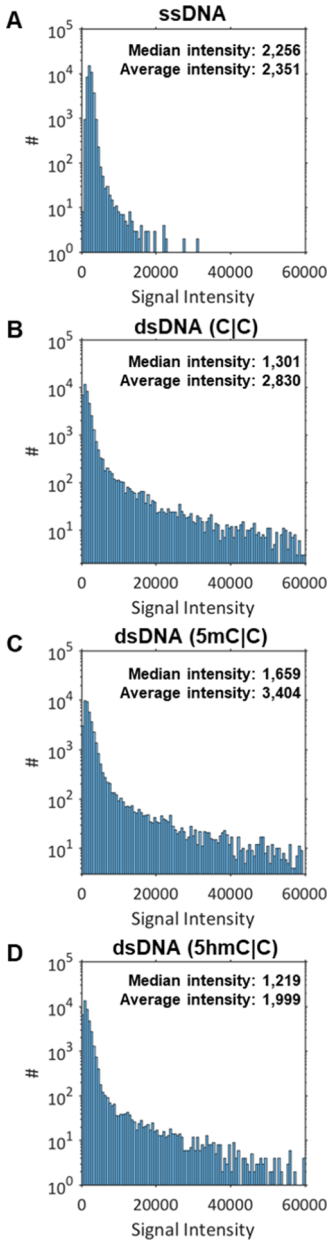


Figure 1. NFATc2 binding four types of DNA. Histogram of fluorescence intensities from PBM experiments representing NFATc2 binding to 40,000 DNA features containing (A) ssDNA, (B) dsDNA, and (C) dsDNA with 5mC in one strand and cytosine in the other strand (5mC|C) and (D) dsDNA with 5hmC in one strand and cytosine in another strand (5hmC|C). Values above 60,000 intensity units are not shown. The median and average binding intensities of NFATc2 with four types of DNA are indicated on the plots.

dsDNA(C|C), reverse complements are equivalent (i.e., **TTTCC** is equivalent to **GGAAA**), but with ssDNA, complements are not equivalent. We obtain lower Pearson correlation values for the 8-mer Z-scores for ssDNA replicate experiments (Figures S2B and S3) compared to those obtained for the raw intensity values (Figure S2A), highlighting the lower sequence specificity for ssDNA. Nevertheless, among the top bound 8-mers bound by NFAT as ssDNA are those containing **TTTCC** and **GTTCC** (Figures 2, S5A-C, Table S3).

2.2.2. NFATc2 Binding to dsDNA(5mC|C). Recently, it was shown that the Rel domain of NFATc2 binds 8-mers

https://dx.doi.org/10.1021/acsomega.0c04069
ACS Omega 2021, 6, 4147−4154

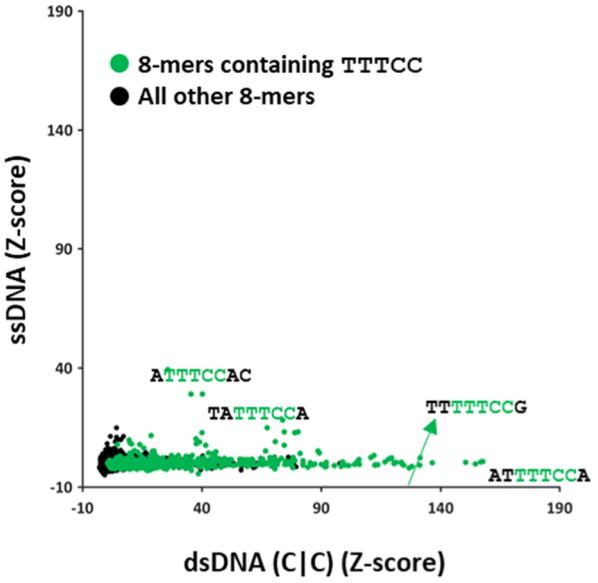


Figure 2. NFATc2 binding to single-stranded DNA 8-mers. Z-scores for NFATc2 binding to double-stranded DNA 8-mers (*x*-axis) or single-stranded DNA 8-mers (*y*-axis). 8-mers are divided into two groups: 8-mers containing **TTTCC** (green) and all other 8-mers (black). A selection of 8-mers is indicated with their sequences provided.

containing the 5-mer **CGGAA** stronger when both cytosines in the CG dinucleotide are 5mC.[23] We examined binding when only one strand contains 5mC. Figure 3A shows the scatter plot comparing 8-mers bound by NFATc2 to dsDNA(C|C) (*x*-axis) and dsDNA(5mC|C) (*y*-axis). For this comparison, we divided 8-mers into four groups: (1) those that contain the NFATc2 canonical 5-mer **CGGAA**, (2) those that contain the reverse complementary 5-mer **TTCCG**, (3) all other 8-mers with a cytosine, and (4) 8-mers with no cytosine. NFATc2 binding to 8-mers without a cytosine is along the diagonal and acts as an internal control. The best bound 8-mer in both dsDNA(5mC|C) and dsDNA(C|C) in the non-cytosine 8-mer group is **TGGAAAAT**; thymine and 5mC both contain a methyl group at the same position of the pyrimidine ring. In general, 8-mers with **CGGAA** are better bound with dsDNA(5mC|C), while 8-mers containing the reverse complement **TTCCG** are better bound with dsDNA(C|C). Examination of NFATc2 binding to 8-mers containing a single cytosine or 5mC (Figure 3B) shows a similar pattern to Figure 3A. Here, all 8-mers containing the **MGAAA** 5-mer are well bound, suggesting that methylation of a single cytosine at position −1 in the **M**$^{-1}$**GGAA** 5-mer is sufficient for the stronger binding of NFATc2 (Figure 3B).

2.2.2.1. Contribution of Each 5mC in a Methylated CG Dinucleotide to Preferential NFATc2 Binding. We compared the dsDNA(5mC|C) data to previous data where both cytosines in all CG dinucleotides are 5mC (dsDNA-(5mCG)).[23] We examined the 5096 8-mers that contain one cytosine that is in a CG dinucleotide (Figure 3C). 8-mers containing the **CGGAA** 5-mer are along the diagonal, indicating that methylation of the single cytosine in **CGGAA** is energetically similar to methylation of both cytosines in the **CG** dinucleotide. It also suggests that methylation of the cytosine on the other strand (e.g., **TTCMG**) does not change NFATc2 binding. While our experiments do not allow for different combinations of unmethylated and methylated

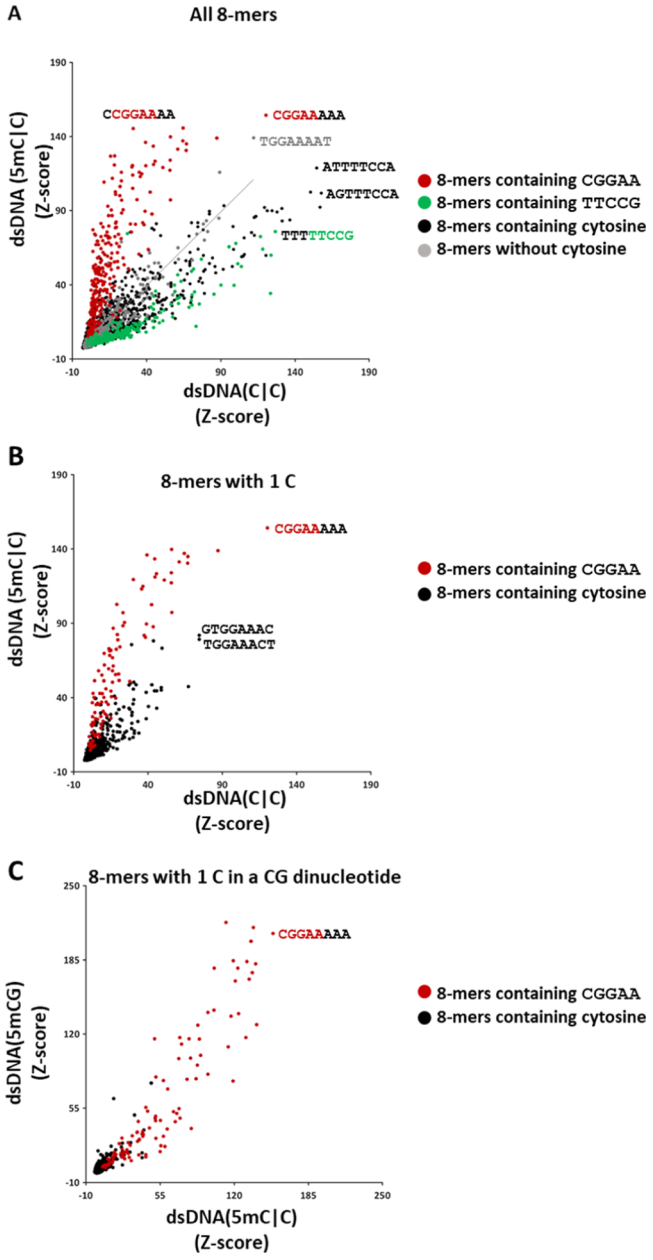


Figure 3. NFATc2 binding to double-stranded DNA 8-mers with 5mC in one strand and cytosine in another strand. (A) 8-mer Z-score comparison of NFATc2 binding to dsDNA(C|C) (*x*-axis) and dsDNA(5mC|C) (*y*-axis). The 8-mers are divided into four groups: those containing **CGGAA** (red), containing **TTCCG** (green), all other cytosine containing 8-mers (black), and those without a cytosine (gray). The sequences of several 8-mers are indicated. A line of best fit (gray) is shown for the 8-mers without a cytosine. (B) Same as in (A) but for 8-mers containing one cytosine. (C) Z-scores for Zta binding to dsDNA(5mC|C) 8-mers (*x*-axis) or 8-mers in which all cytosines in all CG dinucleotides are 5mC (dsDNA(5mCG)) (*y*-axis).[23] Only data for 8-mers containing a single CG in the CG dinucleotide are shown. 8-mers containing the core NFAT motif **CGAAA** are shown as red and all other 8-mers are shown in black.

cytosines on the same strand (i.e., we cannot measure binding of **TTCM**$^{1'}$**G**), the reduction of NFATc2 binding to the sequence **TTMMG** observed in Figure 3A suggests that 5mC at position 2′ is responsible for the reduced binding to the complement motif.

https://dx.doi.org/10.1021/acsomega.0c04069
ACS Omega 2021, 6, 4147−4154

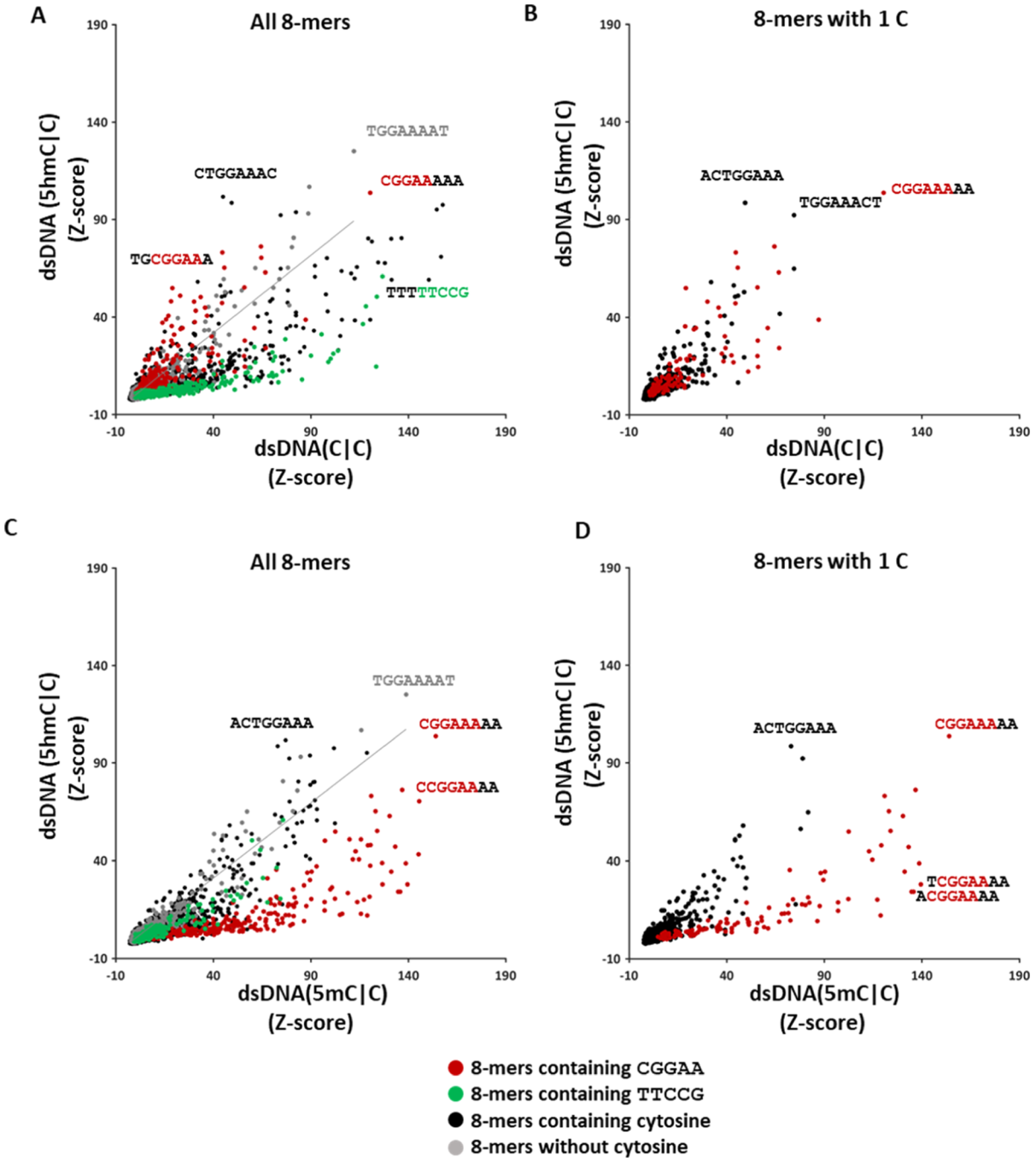


Figure 4. NFATc2 binding to dsDNA 8-mers with 5hmC in one strand. (A) 8-mer Z-score comparison of NFATc2 binding to dsDNA(C|C) (*x*-axis) and dsDNA(5hmC|C) (*y*-axis). (B) Same as in (A) but for 8-mers containing one cytosine. (C) 8-mer Z-score comparison of NFATc2 binding to dsDNA(5mC|C) *x*-axis and dsDNA(5hmC|C) (*y*-axis). 8-mers are colored as shown in Figure 3, with (D) same as in (C) but for 8-mers containing 1 cytosine. For all panels, 8-mers are colored as shown in Figure 3, with the sequences of several 8-mers indicated.

2.2.3. NFATc2 Binding to dsDNA(5hmC|C). 5hmC is an oxidative product of 5mC that is associated with active demethylation of DNA *in vivo*.[36] We compared NFATc2 binding to dsDNA(C|C) and dsDNA(5hmC|C) (Figure 4A). The effect of 5hmC on NFATc2 binding is not as dramatic as observed with 5mC. Similar to dsDNA(5mC|C), all 8-mers containing the reverse complement (**TTCCG**) are preferentially bound to dsDNA(C|C) compared to dsDNA(5hmC|C). Few 8-mers with **CGGAA** are moderately better bound with 5hmC (e.g., **TGHGGAAA**). Examination of 8-mers with only one cytosine (Figure 4B) highlights a single 8-mer, with **CGGAA** (**CGGAAAAA**) being the best bound 8-mer with either 5hmC or C. 8-mers containing 5hmC outside of the core motif (e.g., **AHTGGAAA**) are also more strongly bound by NFATc2, indicating that 5hmC at positions outside of the core motif can influence NFATc2 binding.

Figure 4C,D compares NFATc2 binding to 8-mers for dsDNA(5mC|C) (*x*-axis) and dsDNA(5hmC|C) (*y*-axis). Most 8-mers containing **MGGAA** are better bound than **HGGAA** (e.g., **TMGGAAAA** and **AMGGAAA**), with the exception being 8-mer $^{M}/_{H}$**GGAAAAA**, which is similarly strongly bound by NFATc2 when it contains either 5mC or

https://dx.doi.org/10.1021/acsomega.0c04069
ACS Omega 2021, 6, 4147−4154

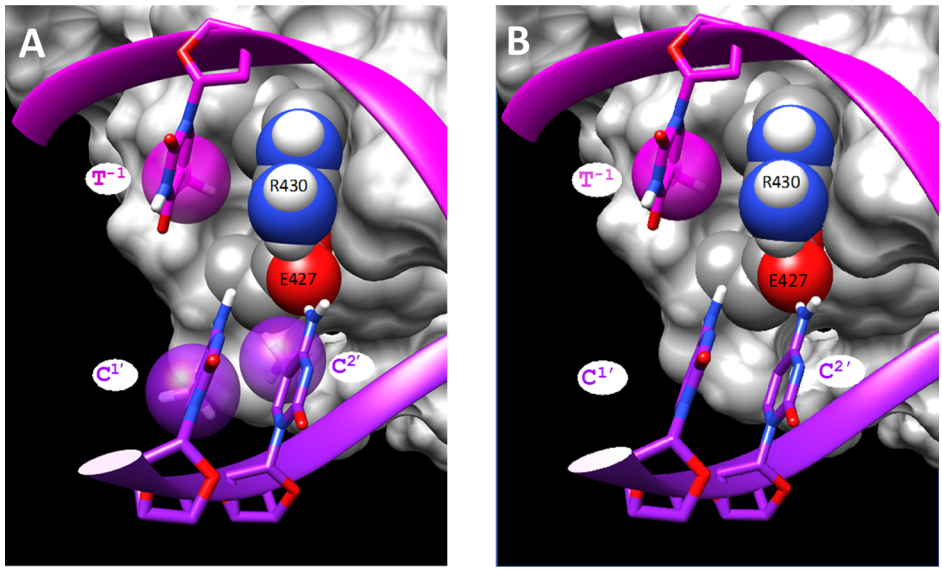


Figure 5. Models of NFATc2 binding methylated and unmethylated dsDNA. (A) Model of NFATc2 binding the $\mathbf{T^{-1}GGAA}$ motif and the methylated complement $\mathbf{TTM^{2'}M^{1'}A}$. The model was developed from segments H, V, and Z of the PDB: 1pzu crystal structure (ref 20) by simply adding methyl groups to the C5 atoms of the cytosine bases. Only $\mathbf{T^{-1}}$ of the top DNA strand and $\mathbf{M^{1'}}$ and $\mathbf{M^{2'}}$ of the bottom strand are shown for clarity. Pertinent residues of the protein are shown as CPK spheres, and the rest of the protein is represented as a gray surface. The bases and sugars of the DNA are shown as sticks, and the backbones are represented by ribbons. The atom color code is nitrogen—blue, oxygen—red, hydrogen—white, protein carbon—gray, and DNA carbon and backbone ribbons—magenta and purple, respectively, to distinguish the two strands. The methyl groups of $\mathbf{T^{-1}}$, $\mathbf{M^{1'}}$, and $\mathbf{M^{2'}}$ are represented by transparent spheres. (B) Model of NFATc2 binding the $\mathbf{T^{-1}GGAA}$ motif and the unmethylated complement $\mathbf{TTC^{2'}C^{1'}A}$.

5hmC. NFATc2 modestly prefers 8-mers with **TTMMG** to **TTHHG**. Some 8-mers containing cytosines outside the core **CGGAA** 5-mer are better bound when they contain 5hmC compared to 5mC or cytosine (e.g., the aforementioned **AHTGAAA**), highlighting that 5hmC at positions outside of the NFAT core motif can promote NFATc2 binding to dsDNA.

2.3. Structural Analysis of NFATc2 Binding Methylated dsDNA. We investigated the available NFAT/dsDNA structural complexes[20,21] in order to understand the physical causes of the effects of cytosine modifications on NFATc2 dsDNA binding. Figure 5A examines the N-terminal segment of the dsDNA binding domain of NFAT1 (RHR-N) binding the beginning of the $\mathbf{T^{-1}GGAA}$ consensus element. In particular, it is seen that the methyl group of the thymine at the −1 position ($\mathbf{T^{-1}}$) occupies a partially hydrophobic pocket formed by the side chain of R430 and the protein backbone. This is consistent with thymine being a preferred nucleotide at this position for complex formation, and why 5mC, for which the added methyl group is at the same location as in thymine, is preferred over cytosine. On the complementary strand, it is seen that the added methyl group on the cytosine at the 1 position ($\mathbf{C^{1'}}$) is distant from the protein, consistent with our suggestion that methylation at this position has no effect on the dsDNA binding affinity of the complex. Finally, it is seen that for the second cytosine on this strand ($\mathbf{C^{2'}}$), the added methyl group butts against the side chain of E427, which is not seen for the unmethylated version of the consensus (Figure 5B). This suggestion of a potential steric clash is consistent with the finding that methylation of this cytosine weakens binding. This is further supported by an NMR structure involving only the RHR-N domain of NFAT2 (PDB 1A66;[21]). In this case, $\mathbf{C^{2'}}$ and E427 are in close proximity, and methylation causes an obvious steric clash (Figure S6).

3. DISCUSSION

Previously, we showed that NFATc2 binds stronger to the 5-mer $\mathbf{C^{-1}G^{1}G^{2}A^{3}A^{4}}$ when both cytosines in the CG dinucleotide are methylated (i.e., $\mathbf{M^{-1}G^{1}G^{2}A^{3}A^{4}}$ in one strand and $\mathbf{T^{4'}T^{3'}C^{2'}M^{1'}G^{-1}}$ in the second strand).[23] Here, we examine how NFAT2c binds to three additional kinds of DNA: ssDNA, dsDNA(5mC|C), and dsDNA(5hmC|C). ssDNA 8-mers containing **TTTCC** or **GTTCC** are among the best bound 8-mers by NFAT, suggesting that interactions with the strand containing the **TTTCC** consensus may dominate interactions on the opposite strand. We evaluated how 5mC in either strand of dsDNA, one in the $\mathbf{C^{-1}GGAA}$ and two in the complementary strand $\mathbf{TTC^{2'}C^{1'}G}$, affects NFATc2 binding. 5mC in **CGGAA** and 5mC in both the cytosines of the **CG** dinucleotide in **CGGAA** are similarly bound by NFATc2, suggesting that 5mC in the opposite strand ($\mathbf{TTC^{2'}C^{1'}G}$) at $\mathbf{C^{1'}}$does not change binding. However, 5mC at position 2 in the complement $\mathbf{TTC^{2'}C^{1'}G}$ inhibits DNA binding, which we attribute to a steric clash occurring between the methyl group of 5mC at position 2′ of the core motif and Glu427 of the RHR-N domain of NFATc2. 5hmC is similar to cytosine for binding **CGGAA** and weakens binding to **TTCCG** more than 5mC.

The strong binding to certain ssDNA 8-mers may be biologically important. It is becoming increasingly clear that non-B-form DNA structures are bound in a sequence-specific manner and take part in gene regulation by selective binding of different TFs and small molecules.[37−39] Sequence-specific binding to ssDNA has regulatory roles in eukaryotic transcription.[40] The stronger binding to ssDNA may reflect the conformational flexibility of ssDNA[40,41] and indicate that sequence-specific binding is primarily on one strand of DNA.

Our PBM data and structural analysis suggest a similarity between 5mC and thymine on NFATc2 binding (Figures 3

https://dx.doi.org/10.1021/acsomega.0c04069
ACS Omega 2021, 6, 4147−4154

and 5). However, some 8-mers with **MGAAA** are better bound than those containing **TGAAA** (Figure S7A,B), suggesting that they are not equivalent. Many of these 8-mers differ from each other at locations outside of the core NFAT motif (e.g., **AATMGGAA**, p-val = 0.006, Figure S7B). In addition, several thymine to 5mC substitutions also reduce binding of NFATc2, including those outside of the core motif (e.g., **TGGAAAAT** is better bound than **TGGAAAAM**, p-val = 0.04, Figure S7B). We currently cannot provide a physical explanation for NFATc2 binding these 8-mers. Our data highlight the effects of cytosine and modified cytosines inside and outside of the core motif on NFATc2 dsDNA binding, providing a richer description of the dsDNA binding specificity of NFATc2. The DNA binding domain of all NFAT members is highly conserved (64−72% sequence identity),[42] suggesting that other NFAT family members may have similar changes in dsDNA binding with 5mC and 5hmC.

4. MATERIALS AND METHODS

4.1. Cloning and Expression of Human NFATc2 DNA Binding Domain. The construct containing human NFATc2 is an N-terminal GST construct cloned into a Gateway system pDEST15 vector.[23] The protein was expressed using the PURExpress *In vitro* protein synthesis kit (NEB) as per the manufacturer's protocol[33] in a 12.5 μL reaction volume containing 90 ng of plasmid. The amino acid sequence of NFATc2 (https://www.ncbi.nlm.nih.gov/protein/Q13469.2) with the Rel homology DNA binding domain (RHR-N) is shown below in bold:

LVPPTWPKPLVPAIPICSIPVT**ASLPPLEWPLSSQSG-SYELRIEVQPKPHHRAHYETEGSRGAV-KAPTGGHPVVQLHGYMENKPLGLQIFIGTADERILK-PHAFYQVHRITGKTVTTTSYEKIVGNTKVLEI-PLEPKNNMRATIDCAGILKLRNADIELRKGET-DIGRKNTRVRLVFRVHIPESSGRIVSLQTASNPIECSQR-SA**HELPMVERQDTDSCLVYGGQQMILTGQNFT-SESKVVFTEKTTDGQQIWEMEATVDKDKSQPNMLFVEI-PEYRNKHIRTPVKVNFYVINGKRKRSQPQHFTYHPV-PAIKTEPTDEYDPTL.

4.2. Design of 40k Feature PBM and Double Stranding of the Microarray. The 40,000 array feature PBM design contains 16 sectors, each containing a DNA grid with 40,000 features. Each feature consists of a single-stranded DNA 60-mer probe with a 35bp long variable and 25bp invariable sequence.[24,43] The variable sequence is designed in such a way that all 10bp sequences (10-mers) are represented once on the array, and all 8-mers (including complements) are represented 32 times. T7 DNA polymerase was used to incorporate cytosine (NEB) or 5mC (NEB) or 5hmC (Zymo Research) into the DNA, respectively, in the probes during double stranding, creating unmethylated, hemi-methylated, or hemi-hydroxymethylated dsDNA.[30,33,44] To monitor the double-stranding efficiency, the double-stranding reaction mixture was spiked with Cy3-dCTP (4%).[30,33,44]

4.3. DNA-Protein Binding Reaction and Data Analysis. DNA protein binding reactions were performed using these four types of DNA.[45] The image generated from the Agilent Surescan microarray scanner was quantified using ImaGene 9 software (BioDiscovery Inc.), and the extracted probe intensity values were used for the calculation of Z-scores.[44] In previous studies, complementary 8-mers were combined,[33] but because of the asymmetric nature of the double-stranding protocol for 5mC and 5hmC on PBMs, complementary 8-mers are different. Therefore, the Z-scores of the reverse complement 8-mer were only considered and extracted from the array probe design as these represent the sequences containing the modified cytosine in dsDNA. Z-score is the measure of standard deviations for each 8-mer intensity from the global array median intensity.[34] All proteins were assayed at least twice with good agreement ($R > 0.85$) between replicates (Figures S2 and S3). For all visualizations in the main figures, we used data from representative experiments (indicated in Figure S2). Data (raw probe intensities and 8-mer Z-scores) have been deposited at the NCBI Gene Expression Omnibus (GEO) database under accession GSE10463.

4.4. Motif Enrichment. Motif enrichment analysis in the top 1% (403) of array probes ranked by intensity was performed using the MEME tool of the MEME software suite version 4.11.2.[46] Motifs were searched on the forward strands for the ssDNA PBMs and the reverse strand for the dsDNA(5mC|C) and dsDNA(5hmC|C) experiments. Enriched motifs were searched for on both strands of the dsDNA(C|C) experiments.

4.5. Structural Modeling. Models of the NFATc2 protein binding the sequence **TGGAAA** and the methylated complement **TTTMMA** were developed from the crystal structure of the unmethylated complex (ref 20; PDB: 1pzu) with the UCSF Chimera software package.[47] Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c04069.

Number of array features above median signal intensity with two replicates, motif enrichment, top 0.1% of 8-mers bound by NFAT, Schematics description of experimental design, summary of NFATc2 correlations, comparison of 8-mer Z-scores of replicates data sets, NFATc2 binding to four types of DNA (replicate 2), NFATc2 binding to ssDNA 8-mers, highlight of the interaction of NFATc2 with 5mC, and effect of thymine to 5mC substitutions (PDF)

■ 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: 1-240-760-6885; 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;* orcid.org/0000-0003-3568-6148

Stewart R. Durell − *Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

https://dx.doi.org/10.1021/acsomega.0c04069
ACS Omega 2021, 6, 4147−4154

Syed Khund-Sayeed − *Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c04069

Funding

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

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

M, 5-methylcytosine; **H**, 5-hydroxymethylcytosine; PBM, Protein Binding Microarray; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA

■ REFERENCES

(1) Li, X.; Ho, S. N.; Luna, J.; Giacalone, J.; Thomas, D. J.; Timmerman, L. A.; Crabtree, G. R.; Francke, U. Cloning and chromosomal localization of the human and murine genes for the T-cell transcription factors NFATc and NFATp. *Cytogenet. Cell Genet.* **1995**, *68*, 185−191.

(2) Rao, A.; Luo, C.; Hogan, P. G. TRANSCRIPTION FACTORS OF THE NFAT FAMILY:Regulation and Function. *Annu. Rev. Immunol.* **1997**, *15*, 707−747.

(3) Mancini, M.; Toker, A. NFAT proteins: emerging roles in cancer progression. *Nat. Rev. Cancer* **2009**, *9*, 810−820.

(4) Crabtree, G. R.; Clipstone, N. A. Signal transmission between the plasma membrane and nucleus of T lymphocytes. *Annu. Rev. Biochem.* **1994**, *63*, 1045−1083.

(5) Shaw, J.; Utz, P.; Durand, D.; Toole, J.; Emmel, E.; Crabtree, G. Identification of a putative regulator of early T cell activation genes. *Science* **1988**, *241*, 202−205.

(6) Crabtree, G. R. Generic Signals and Specific Outcomes. *Cell* **1999**, *96*, 611−614.

(7) Ranger, A. M.; Grusby, M. J.; Hodge, M. R.; Gravallese, E. M.; de la Brousse, F. C.; Hoey, T.; Mickanin, C.; Baldwin, H. S.; Glimcher, L. H. The transcription factor NF-ATc is essential for cardiac valve formation. *Nature* **1998**, *392*, 186−190.

(8) Graef, I. A.; Mermelstein, P. G.; Stankunas, K.; Neilson, J. R.; Deisseroth, K.; Tsien, R. W.; Crabtree, G. R. L-type calcium channels and GSK-3 regulate the activity of NF-ATc4 in hippocampal neurons. *Nature* **1999**, *401*, 703−708.

(9) Jauliac, S.; López-Rodriguez, C.; Shaw, L. M.; Brown, L. F.; Rao, A.; Toker, A. The role of NFAT transcription factors in integrin-mediated carcinoma invasion. *Nat. Cell Biol.* **2002**, *4*, 540−544.

(10) Robbs, B. K.; Cruz, A. L. S.; Werneck, M. B. F.; Mognol, G. P.; Viola, J. P. B. Dual roles for NFAT transcription factor genes as oncogenes and tumor suppressors. *Mol. Cell. Biol.* **2008**, *28*, 7168−7181.

(11) Chen, L.; Glover, J. N. M.; Hogan, P. G.; Rao, A.; Harrison, S. C. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **1998**, *392*, 42−48.

(12) Stroud, J. C.; Chen, L. Structure of NFAT bound to DNA as a monomer. *J. Mol. Biol.* **2003**, *334*, 1009−1022.

(13) Jain, J.; McCaffrey, P. G.; Valge-Archer, V. E.; Rao, A. Nuclear factor of activated T cells contains Fos and Jun. *Nature* **1992**, *356*, 801−804.

(14) Molkentin, J. D.; Lu, J.-R.; Antos, C. L.; Markham, B.; Richardson, J.; Robbins, J.; Grant, S. R.; Olson, E. N. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **1998**, *93*, 215−228.

(15) Hu, C.-M.; Jang, S. Y.; Fanzo, J. C.; Pernis, A. B. Modulation of T cell cytokine production by interferon regulatory factor-4. *J. Biol. Chem.* **2002**, *277*, 49238−49246.

(16) Rengarajan, J.; Mowen, K. A.; McBride, K. D.; Smith, E. D.; Singh, H.; Glimcher, L. H. Interferon regulatory factor 4 (IRF4) interacts with NFATc2 to modulate interleukin 4 gene expression. *J. Exp. Med.* **2002**, *195*, 1003−1012.

(17) Wu, Y.; Borde, M.; Heissmeyer, V.; Feuerer, M.; Lapan, A. D.; Stroud, J. C.; Bates, D. L.; Guo, L.; Han, A.; Ziegler, S. F.; Mathis, D.; Benoist, C.; Chen, L.; Rao, A. FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell* **2006**, *126*, 375−387.

(18) Youn, H.-D.; Chatila, T. A.; Liu, J. O. Integration of calcineurin and MEF2 signals by the coactivator p300 during T-cell apoptosis. *EMBO J.* **2000**, *19*, 4323−4331.

(19) Giffin, M. J.; Stroud, J. C.; Bates, D. L.; von Koenig, K. D.; Hardin, J.; Chen, L. Structure of NFAT1 bound as a dimer to the HIV-1 LTR κB element. *Nat. Struct. Biol.* **2003**, *10*, 800−806.

(20) Jin, L.; Sliz, P.; Chen, L.; Macián, F.; Rao, A.; Hogan, P. G.; Harrison, S. C. An asymmetric NFAT1 dimer on a pseudo-palindromic κB-like DNA site. *Nat. Struct. Biol.* **2003**, *10*, 807−811.

(21) Zhou, P.; Sun, L. J.; Dötsch, V.; Wagner, G.; Verdine, G. L. Solution structure of the core NFATC1/DNA complex. *Cell* **1998**, *92*, 687−696.

(22) 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.

(23) Yin, Y.; Morgunova, E.; Jolma, A.; Kaasinen, E.; Sahu, B.; Khund-Sayeed, S.; Das, P. K.; Kivioja, T.; Dave, K.; Zhong, F.; Nitta, K. R.; Taipale, M.; Popov, A.; Ginno, P. A.; Domcke, S.; Yan, J.; Schubeler, D.; Vinson, C.; Taipale, J. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **2017**, *356*, No. eaaj2239.

(24) Philippakis, A. A.; Qureshi, A. M.; Berger, M. F.; Bulyk, M. L. Design of compact, universal DNA microarrays for protein binding microarray experiments. *J. Comput. Biol.* **2008**, *15*, 655−665.

(25) 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.

(26) 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.

(27) 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.

(28) 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.

(29) 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.

(30) Khund-Sayeed, S.; He, X.; Holzberg, T.; Wang, J.; 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.* **2016**, *8*, 936.

(31) He, X.; Tillo, D.; Vierstra, J.; Syed, K.-S.; Deng, C.; Ray, G. J.; Stamatoyannopoulos, J.; FitzGerald, P. C.; Vinson, C. Methylated Cytosines Mutate to Transcription Factor Binding Sites that Drive Tetrapod Evolution. *Genome Biol. Evol.* **2015**, *7*, 3155−3169.

(32) 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.
(33) 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.
(34) 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.
(35) Weirauch, M. T.; Yang, A.; Albu, M.; Cote, A. G.; Montenegro-Montero, A.; Drewe, P.; Najafabadi, H. S.; Lambert, S. A.; Mann, I.; Cook, K.; Zheng, H.; Goity, A.; van Bakel, H.; Lozano, J.-C.; Galli, M.; Lewsey, M. G.; Huang, E.; Mukherjee, T.; Chen, X.; Reece-Hoyes, J. S.; Govindarajan, S.; Shaulsky, G.; Walhout, A. J. M.; Bouget, F.-Y.; Ratsch, G.; Larrondo, L. F.; Ecker, J. R.; Hughes, T. R. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **2014**, *158*, 1431−1443.
(36) Branco, M. R.; Ficz, G.; Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **2011**, *13*, 7−13.
(37) Ray, S.; Tillo, D.; Boer, R. E.; Assad, N.; Barshai, M.; Wu, G.; Orenstein, Y.; Yang, D.; Schneekloth, J. S., Jr; Vinson, C. Custom DNA Microarrays Reveal Diverse Binding Preferences of Proteins and Small Molecules to Thousands of G-Quadruplexes. *ACS Chem. Biol.* **2020**, *15*, 925−935.
(38) Wu, G.; Tillo, D.; Ray, S.; Chang, T. C.; Schneekloth, J. S., Jr; Vinson, C.; Yang, D. Custom G4 Microarrays Reveal Selective G-Quadruplex Recognition of Small Molecule BMVC: A Large-Scale Assessment of Ligand Binding Selectivity. *Molecules* **2020**, *25*, 3465.
(39) Reina, C.; Cavalieri, V. Epigenetic Modulation of Chromatin States and Gene Expression by G-Quadruplex Structures. *Int. J. Mol. Sci.* **2020**, *21*, 4172.
(40) Tomonaga, T.; Levens, D. Activating transcription from single stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5830−5835.
(41) Bao, L.; Zhang, X.; Jin, L.; Tan, Z.-J. Flexibility of Nucleic Acids: From DNA to RNA ArXiv e-prints [Online]. 2015, http://adsabs.harvard.edu/abs/2015arXiv150906450B (accessed September 1, 2015).
(42) Mognol, G. P.; Carneiro, F. R. G.; Robbs, B. K.; Faget, D. V.; Viola, J. P. B. Cell cycle and apoptosis regulation by NFAT transcription factors: new roles for an old player. *Cell Death Dis.* **2016**, *7*, No. e2199.
(43) Berger, M. F.; Badis, G.; Gehrke, A. R.; Talukder, S.; Philippakis, A. A.; Peña-Castillo, L.; Alleyne, T. M.; Mnaimneh, S.; Botvinnik, O. B.; Chan, E. T.; Khalid, F.; Zhang, W.; Newburger, D.; Jaeger, S. A.; Morris, Q. D.; Bulyk, M. L.; Hughes, T. R. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **2008**, *133*, 1266−1276.
(44) 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.
(45) 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.* **2016**, *8*, 936−945.
(46) Bailey, T. L.; Johnson, J.; Grant, C. E.; Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39−W49.
(47) 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://dx.doi.org/10.1021/acsomega.0c04069
ACS Omega 2021, 6, 4147−4154