Biochemical and Biophysical Research Communications 449 (2014) 248–255

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold

Jaya Prakash Golla [a], Jianfei Zhao [a], Ishminder K. Mann [a], Syed K. Sayeed [a], Ajeet Mandal [a], Robert B. Rose [b], Charles Vinson [a,*]

[a] Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Room 3128, Building 37, Bethesda, MD 20892, United States
[b] Department of Biochemistry, North Carolina State University, 128 Polk Hall, Raleigh, NC 27695, United States

ARTICLE INFO

Article history:
Received 1 May 2014
Available online 14 May 2014

Keywords:
Carboxylation
E-Box motif
CG dinucleotide
Basic-helix-loop-helix
DNA binding
Tcf3|Ascl1 heterodimer

ABSTRACT

Three oxidative products of 5-methylcytosine (5mC) occur in mammalian genomes. We evaluated if these cytosine modifications in a CG dinucleotide altered DNA binding of four B-HLH homodimers and three heterodimers to the E-Box motif CGCAG|GTG. We examined 25 DNA probes containing all combinations of cytosine in a CG dinucleotide and none changed binding except for carboxylation of cytosine (5caC) in the strand CGCAG|GTG. 5caC enhanced binding of all examined B-HLH homodimers and heterodimers, particularly the Tcf3|Ascl1 heterodimer which increased binding ∼10-fold. These results highlight a potential function of the oxidative products of 5mC, changing the DNA binding of sequence-specific transcription factors.

Published by Elsevier Inc.

1. Introduction

In mammals, ∼60–80% of the cytosines in the CG dinucleotide are methylated in somatic cells, particularly in the CG poor regions of the genome [1]. The biological consequences of 5mC in the CG dinucleotide vary [2–4]. Methylation can inhibit the DNA binding of transcription factors (TFs) involved in housekeeping functions like ETS (CCGGAA), SP1 (CCCGCC), and NRF-1 (CGCCTGCG) [5] suggesting a mechanistic link between hypermethylation of CG islands and gene suppression that is observed in some cancers [6]. Alternatively, CG dinucleotide methylation can increase DNA binding of TFs [7] resulting in repression [8] and/or activation of nearby genes [9]. For example, C/EBP family members preferentially bind methylated DNA sequences and are critical for activation of tissue specific promoters during differentiation [7].

Recently, the TET family of dioxygenases was identified that iteratively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), then 5-formylcytosine (5fC), and finally 5-carboxylcytosine (5caC) [10]. Both 5fC and 5caC can be removed by mammalian thymine DNA glycosylase (TDG) and replaced with cytosine (C) to complete the demethylation of 5mC, which occurs when cells differentiate. The abundance of different cytosine forms varies dramatically within cells and between cell types suggesting a potential biological function [10–12].

The effect of 5hmC, 5fC, and 5caC on DNA binding of TFs is only now being investigated [13]. In the present study, we used the Electrophoretic mobility shift assay (EMSA) to examine the DNA binding of four B-HLH homodimers and three heterodimers to 25 double-stranded DNA 28-mers (dsDNA) containing the E-Box 8-mer CGCAG|GTG with different cytosine forms of the CG dinucleotide and observe that 5caC enhances DNA binding. These results were confirmed circular dichroism (CD) thermal denaturation.

2. Materials and methods

2.1. Protein binding microarrays

The 40,000 feature array design consists of 60-mer DNAs, 35-bps are unique DNA sequences connected to a common 25-bp

Abbreviations: C, cytosine; B-HLH, basic-helix-loop-helix; E-Box, Enhancer Box; Tcf3, transcription factor 3; Tcf4, transcription factor 4; Tcf12, transcription factor 12; Ascl1, Achaete-scute homolog 1; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine; EMSA, electrophoretic mobility shift assay; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA.

* Corresponding author. Fax: +1 301 496 8419.
E-mail address: vinsonc@mail.nih.gov (C. Vinson).

http://dx.doi.org/10.1016/j.bbrc.2014.05.018
0006-291X/Published by Elsevier Inc.

sequence used for double stranding [14]. CG dinucleotides were enzymatically methylated and the effect on DNA binding was determined [9].

2.2. In vitro transcription & translation

Protein synthesis was performed using in vitro translation kit (PURExpress, NEB) and the resulting reaction was diluted to a ratio of 1:5 with CD buffer (150 mM KCl, 12.5 mM K_2HPO_4–KH_2PO_4, pH 7.4, 1 mM DTT, 0.25 mM EDTA). 2 μL of diluted reaction mixture containing Tcf3|Ascl1 heterodimer was used in EMSA assays described below.

2.3. DNA oligonucleotides

Twenty single-stranded DNA 28-mer (ssDNA) cartridge-purified oligonucleotides were purchased from W.M. Keck Oligonucleotide Synthesis Facility at Yale to examine two DNA sequences. Five 28-mer DNAs (CTGACCGATA**C**GCAG|GTGCCTGACTGAC) termed the sense strand (a) contained different versions of the cytosine in bold (C, 5mC, 5hmC, 5fC, 5caC). The strong E-Box motif is underlined and the center of the dyad is marked. Five 28-mer DNAs termed the anti-sense strand (b) (GTCAGTCAGGCAC|CTG**C**GTATCGGTCAG) contained different versions of the cytosine in bold. The weak E-box 28-mer on the sense-strand (a) is CTGACCCATA**C**GCAA|ATGTCTGACTGAC. The anti-sense strand was end-labeled with γ-^{32}P ATP (specific activity 5000 Ci/mmol, MP Biomedicals) using T4 polynucleotide kinase (NEB), and was purified by ProbeQuant G-50 micro column (GE Healthcare Biosciences). dsDNA probes were generated by annealing the labeled anti-sense strand and unlabeled sense strand.

2.4. EMSA

The binding of B-HLH proteins to 25 dsDNAs with all possible modifications of cytosine in a CG dinucleotide was analyzed by EMSA [15]. For Tcf3|Ascl1 heterodimer made by IVT, 2 μL of diluted protein was used. For EMSA with purified B-HLH domains, 10 μM dimer was heated at 65 °C for 15 min in the presence of 1 mM DTT, followed by cooling at room temperature for 5 min. Protein dimers and ^{32}P-labeled dsDNA (7 pM) were then added to the EMSA binding buffer (CD buffer containing 0.5 mg/mL BSA, 10% glycerol, 0.02 μg/μL poly dIdC, 10 mM $MgCl_2$) in a final volume of the reaction 20 μL.

2.5. Protein expression and purification

The DNA binding B-HLH domains of Tcf3, Tcf4, Tcf12, and Ascl1 were expressed from a T7 expression vector named pT5 plasmid [16] in *Escherichia coli* BL21 DE3 (LysE) cells. Cells were grown, induced, and collected by centrifugation at +4 °C for 15 min at 6000×*g*. The pellet was resuspended in 4 mL lysis buffer (50 mM Tris–HCl, pH 8.0; 1 mM EDTA; 1 mM DTT; 0.2 mM PMSF), frozen on dry ice and lysed at room temperature in the presence of 1.3 M KCl. The lysate was centrifuged at 30,000 rpm for 30 min in the Beckman L8-80 ultracentrifuge in the 60Ti rotor. The pellet was brought to 4 M urea, sonicated, heated at 65 °C for 15 min, and centrifuged at 5400×g for 10 min [17]. The supernatant was dialyzed to a low salt buffer (20 mM Tris–HCl, pH 8.0, 10 mM KCl, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF) using the Amicon Ultra-15 column (catalog # UFC901024, EMD Millipore), and then loaded to a SP Sepharose column (catalog # 17-0729-01, GE Healthcare Biosciences). The protein was then eluted off the column using 300 mM and 1000 mM KCl and purified by HPLC.

Tcf4 has a C-terminal His tag (HHHHHH) and Ascl1 has a C-terminal Flag φ10 tag (MDYKDDDDKHMASMTGGQQMGRDP). The amino acid sequences for the proteins are:

Tcf3: MGHMVHRPWIQDEVLSLEEKDLRDRERRMANNARERVRVRDINEAFRELGRMCQLHLKSDKAQTKLLILQQAVQVILGLEQQVRERNLNPKAACGGRTRIVSAHNSENEL
Tcf4: MGNNDDEDLTPEQKAEREKERRMANNARERLRVRDINEAFKELGRMVQLHLKSDKPQTKLLILHQAVAVILSLEQQVRERNLNPKAACLKRREEELHHHHHH
Tcf12: MGSTNEDEDLNPEQKIEREKERRMANNARERLRVRDINEAFKELGRMCQLHLKSEKPQTKLLILHQAVAVILSLEQQVRERNLNPKAACLKRREEEL
Ascl1: MASGFGYSLPQQQPAAVARRNERERNRVKLVNLGFATLREHVPNGAANKKMSKVETLRSAVEYIRALQQLLDEHDAVSAAFQAGVLSPELMDYKDDDDKHMASMTGGQQMGRDP

2.6. CD spectroscopy

CD spectroscopy was performed using a Jasco J-720 spectropolarimeter and thermal denaturation curves were fitted [18]. The sum line in Fig. 3A is twice the concentration.

2.7. Crystal structure of transcription factor E47 (Tcf3) homodimer

The image of the X-ray structure of the E47 homodimer bound to DNA [19] was generated using the program Chimera http://www.cgl.ucsf.edu/chimera/.

3. Results

3.1. Protein binding microarrays

We used protein binding microarrays [14] to determine the DNA binding specificities of the Tcf3|Ascl1 heterodimer binding to unmethylated and enzymatically methylated CG dinucleotides using Agilent microarrays containing 40,000 features. The Tcf3|Ascl1 heterodimer bound the E-box motif 8-mer **C**GCAG|GTG well when both cytosines in the CG dinucleotide were either unmethylated (C) or 5mC (Fig. 1A). Methylation of a CG dinucleotide in the center of E-Box (CGCA**C**|GTG) inhibits binding (Table 1).

3.2. In vitro translated Tcf3 & Ascl1 proteins binding 25 different dsDNA with modified CG dinucleotides

The five ssDNA 28-mer (CTGACCGATA**C**GCAG|GTGCCTGACTGAC) with different cytosines were annealed with the complementary ssDNA to make 25 dsDNAs with different chemical forms of the CG dinucleotide. Fig. 1B is an EMSA with 25 DNAs shows that the Tcf3|Ascl1 mixture bound five DNAs and all contain 5caC for the C in bold (**C**GCAG|GTG). Other cytosine modifications did not affect dramatically DNA binding.

3.3. Four B-HLH homodimers binding 25 dsDNAs

To quantify the contribution of 5caC to Tcf3|Ascl1 binding, we used pure B-HLH domains. Fig. 1C–F presents an EMSA using two DNAs, unmodified DNA and DNA with two 5caCs in the CG dinucleotide in CGCAG|GTG. A half-log dilution from 1000 nM to 10 nM of four B-HLH homodimers shows 5caC increases binding of all four homodimers with Ascl1 showing the largest increase in binding by ~6-fold. Next, we examined homodimer binding to 25 dsDNAs with different CG dinucleotides. All four homodimers at 300 nM preferentially bound the five DNAs containing 5caC in the CG dinucleotide in CGCAG|GTG (Table 2).

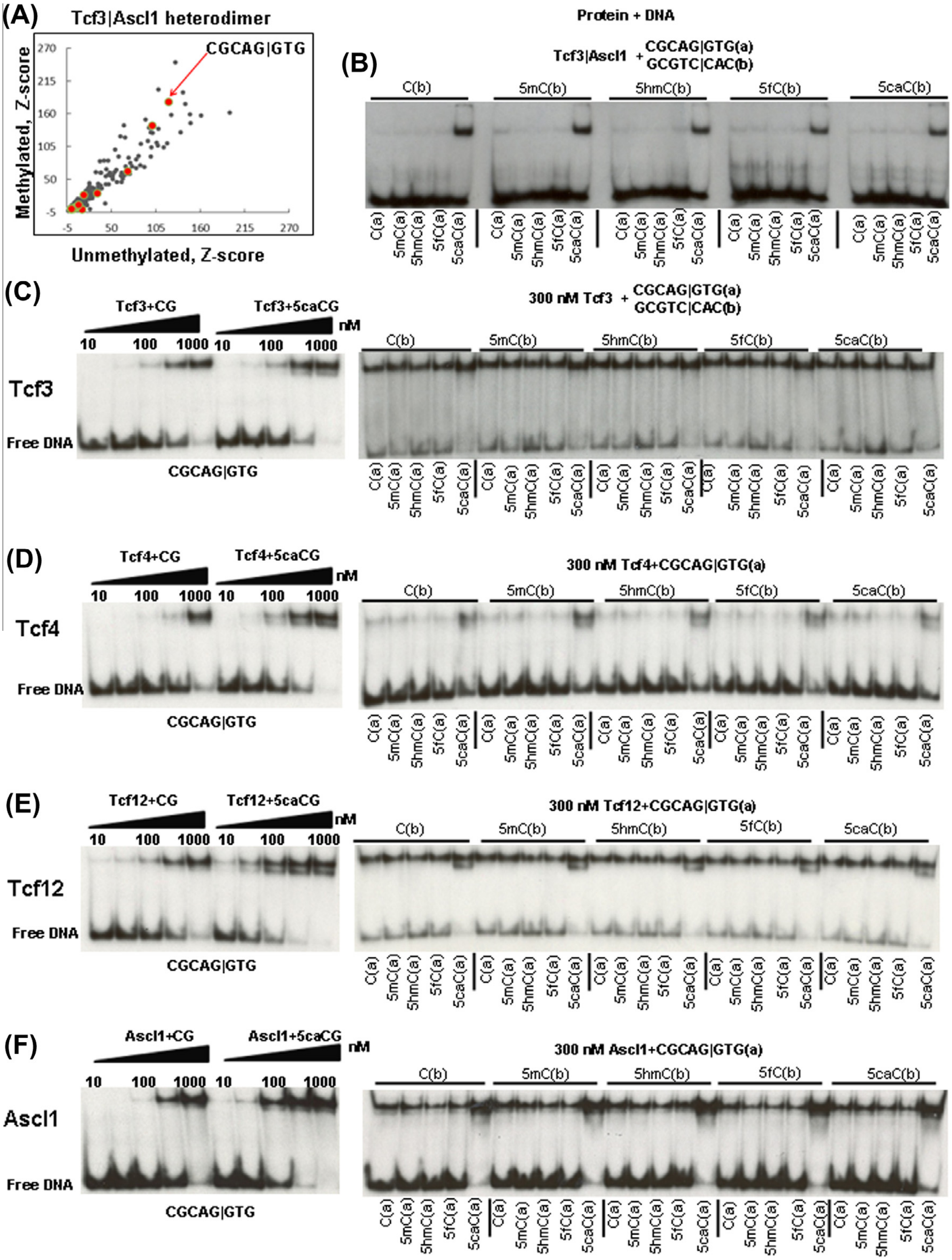

Fig. 1. 8-mers bound by Tcf3|Ascl1 heterodimer and EMSA with pure Tcf3 B-HLH domain. (A) *Z*-scores for 8-mer DNA binding by the Tcf3|Ascl1 heterodimer calculated from protein binding arrays that contained either unmethylated or methylated cytosine in the CG dinucleotide [9]. All 16 E-Box sequences CGCAN|NTG are in red. (B) EMSA of Tcf3 & Ascl1 mixture produced by in vitro transcription translation reaction binding to 25 dsDNA 28-mers containing CGCAG|GTG with different chemical forms of the CG dinucleotide. The heterodimer only binds when 5caC is in the CG dinucleotide in the 8-mer CGCAN|NTG. EMSA with pure Tcf3 B-HLH domain: (C) a half-log dilution from 1000 nM to 10 nM for the Tcf3 homodimer binding two DNA 28-mers showing preferential binding to DNA containing 5caCs in the CG dinucleotide. The right panel shows 300 nM Tcf3 homodimer binding to 25 DNAs with different cytosine forms of the CG dinucleotide. (D) Tcf4, (E) Tcf12, and (F) Ascl1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Tcf3|Ascl1 heterodimer binds CGCAG|GTG ~10-fold better when the CG contains 5caC

We next examined an equimolar mixture of purified Tcf3 and Ascl1 B-HLH domains binding to modified CG dinucleotides. Fig. 2A presents an EMSA of the Tcf3|Ascl1 heterodimer with a half-log dilution from 30 nM to 0.3 nM binding to the two DNAs described previously, unmodified and 5caC containing DNA. The mixture of Tcf3 and Ascl1 binds better (K_d = 10–30 nM) than either the Tcf3 homodimer (K_d = 300–1000 nM) or the Ascl homodimer

Table 1
Z-scores for the E-Box motifs (CGCAN|NTG) bound by Tcf3|Ascl1 heterodimer on methylated and unmethylated arrays.

E-Box motif 8-mer	Unmethylated	Methylated
CGCAG|GTG	120.7	180.0
CGCAC|CTG	101.0	140.0
CGCAG|CTG	70.5	63.7
CGCA|GATG	32.9	27.4
CGCAT|CTG	15.8	24.2
CGCAC|ATG	9.4	8.2
CGCAT|ATG	7.7	5.6
CGCAT|GTG	5.6	3.7
CGCAA|GTG	1.5	2.2
CGCAA|CTG	2.3	2.0
CGCAG|TTG	3.3	1.4
CGCAA|TTG	0.3	1.2
CGCAC|TTG	1.7	1.0
CGCAA|ATG	0.6	0.7
CGCAC|GTG	13.8	0.2
CGCAT|TTG	0.7	0.2

Table 2
~K_d ranges for B-HLH homodimers and heterodimers binding dsDNA 28-mers with the E-Box CGCAG|GTG, as estimated by EMSA. Fold increase in binding caused by 5caC.

B-HLH domains	CGCAG|GTG K_d (nM)	~fold increase in binding caused by 5caC
Tcf3	300–1000	~2
Tcf4	300–1000	~3
Tcf12	100–300	~4
Ascl1	300–1000	~6
Tcf3|Ascl1	10–30	~10
Tcf4|Ascl1	60–200	~6
Tcf12|Ascl1	60–200	~10

(K_d = 300–1000 nM) indicating that the mixture is forming Tcf3|Ascl1 heterodimers as expected [20]. Two 5caCs in the CG dinucleotide increases DNA binding ~10-fold to a K_d between 1 and 3 nM. We next examined 3 nM Tcf3|Ascl1 heterodimer binding to 25 dsDNAs. Only DNA containing 5caC in the cytosine in bold **C**CGCAG|GTG (a) is well bound. A modest inhibition of binding is observed with 5hmC in the CGCAG|GTG 8-mer.

3.5. The Tcf3|Ascl1 heterodimer binds the weak E-box CGCAA|ATG ~10-fold better when the CG contains 5caC

Tcf3|Ascl1 binding the weak E-Box CGCAA|ATG (Table 1) is enhanced by 5caC. Fig. 2B presents an EMSA with a half-log dilution from 1000 nM to 10 nM of the Tcf3|Ascl1 heterodimer binding unmodified and 5caC containing DNA. Binding to CGCAA|ATG is ~100-fold weaker than CGCAG|GTG. Again, 5caC in the CG dinucleotide is bound ~10-fold better. When 300 nM of protein was used, of the 25 dsDNAs that were examined, only the 5 probes containing 5caC for the cytosine in bold (**C**GCAA|ATG) are well bound.

3.6. The Tcf12|Ascl1 and Tcf4|Ascl1 heterodimers

Fig. 2C and D presents an EMSA of Tcf12|Ascl1 and Tcf4|Ascl1 heterodimers from 200 nM to 2 nM binding the two DNA probes with CGCAG|GTG discussed previously. Both Tcf12|Ascl1 and Tcf4|Ascl1 mixtures bind unmodified DNA better than the either homodimer indicating heterodimer formation. 5caC increases binding of Tcf12|Ascl1 and Tcf4|Ascl1 heterodimers ~10-fold and ~6-fold respectively. Tcf12|Ascl1 and Tcf4|Ascl1 heterodimers binding to 25 dsDNAs show binding to the five DNAs containing 5caC in **C**CGCAG|GTG).

3.7. CD spectra and stability of the Tcf3|Ascl1 heterodimer bound to 4 dsDNAs

CD spectroscopy was also used to examine DNA binding. Thermal stability of Tcf3, Ascl1, and Tcf3|Ascl1 mixture was measured at 222 nm to determine the α-helical content of the dimers (Fig. 3A). With heating, the Tcf3 homodimer cooperatively looses ellipticity at 222 nm as observed for other B-HLH homodimers [21,22]with a T_m of 63.2 °C. The denaturation is well fit using a two-state model of α-helical dimers becoming unhelical monomers. The Ascl1 homodimer is less stable and we do not observe a low temperature baseline. The mixture of Tcf3 and Ascl1 has a T_m of 55.4 °C which is more than the sum of the two homodimer denaturations suggesting heterodimer formation. Addition of DNA increased both the ellipticity and stability of Tcf3|Ascl1 heterodimer. The Tcf3|Ascl1 heterodimer bound to DNA is less stable than dsDNA alone suggesting that the loss of ellipticity at 222 nm is not a consequence of the melting of the DNA but that the heterodimer denatures upon heating in the presence of dsDNA, a wavelength were the ellipticity of DNA does not change when dsDNA is denatured [7].

3.8. CD spectra and thermal stability of DNA containing 5caC

Next, we determined if 5caC in a CG dinucleotide changes the stability of both strong (CGCAG|GTG) and weak E-Box (CGCAA|ATG) motif dsDNA. The CD spectra from 200 nm to 300 nm of 4 dsDNAs at 6 °C, unmodified cytosine, 5caC on one strand, and 5caC on both strands, is similar (Fig. 3B and C), with a minimum at 245 nm and maximum between 270 nm and 280 nm, traits of B-form DNA [23]. Thermal stability at 245 nm shows that, for both strong and weak E-Box motif the unmodified DNA is most stable of the 4 DNAs, denaturing at 74 °C and 70 °C respectively. The two DNAs containing one 5caC in the strong E-Box are less stable (68 °C and 69 °C) while the DNA with two 5caC is the least stable (67 °C) (Table 3, Fig. 3D). Similarly, for weak E-Box the two DNAs with one 5caC are less stable and the DNA with two 5caC is the least stable, as seen for the 28-mers with CGCAG|GTG in the center (Fig. 3E, Table 4). The thermal denaturation of Tcf3|Ascl1 heterodimer bound to dsDNA containing CGCAA|ATG did not produce a clear two-state transition indicative of poor binding as observed with EMSA.

3.9. CD spectra and stability of the Tcf3|Ascl1 heterodimer bound to modified dsDNAs

The thermal stability of the Tcf3|Ascl1 heterodimer monitored at 222 nm is greater when bound to the two DNAs containing 5caC in the CG dinucleotide in CGCAG|GTG (T_m = 64.4 and 64.2 °C) compared to the two DNAs containing cytosine (T_m = 61.4 and 61.8 °C) (Fig. 3F, Table 3). The thermal stability of the Ascl1 homodimer shows similar traits, stability is higher when bound to two DNAs that contain 5caC in the CG dinucleotide in CGCAG|GTG (T_m = 48.5 and 47.8 °C) compared to the two DNA containing unmodified cytosine (T_m = 43.6 and 41.2 °C) (Fig. 3G, Table 3).

3.10. Crystal structure of transcription factor E47 (Tcf3) homodimer bound to DNA

Most crystal structures of B-HLH domains bound to DNA do not show a protein interaction with the base in the position analogous to the 5caC [24]. However, the E47 (Tcf3) homodimer bound to E-Box DNA shows an arginine interacting with an adenine (Fig. 4A) that is in the same position as the 5caC that increases Tcf3|Ascl1 binding [19] suggesting a direct protein-DNA interaction. Fig. 4B presents the amino acid sequence of DNA binding region of the four B-HLH proteins used in this study. The arginine in the E47

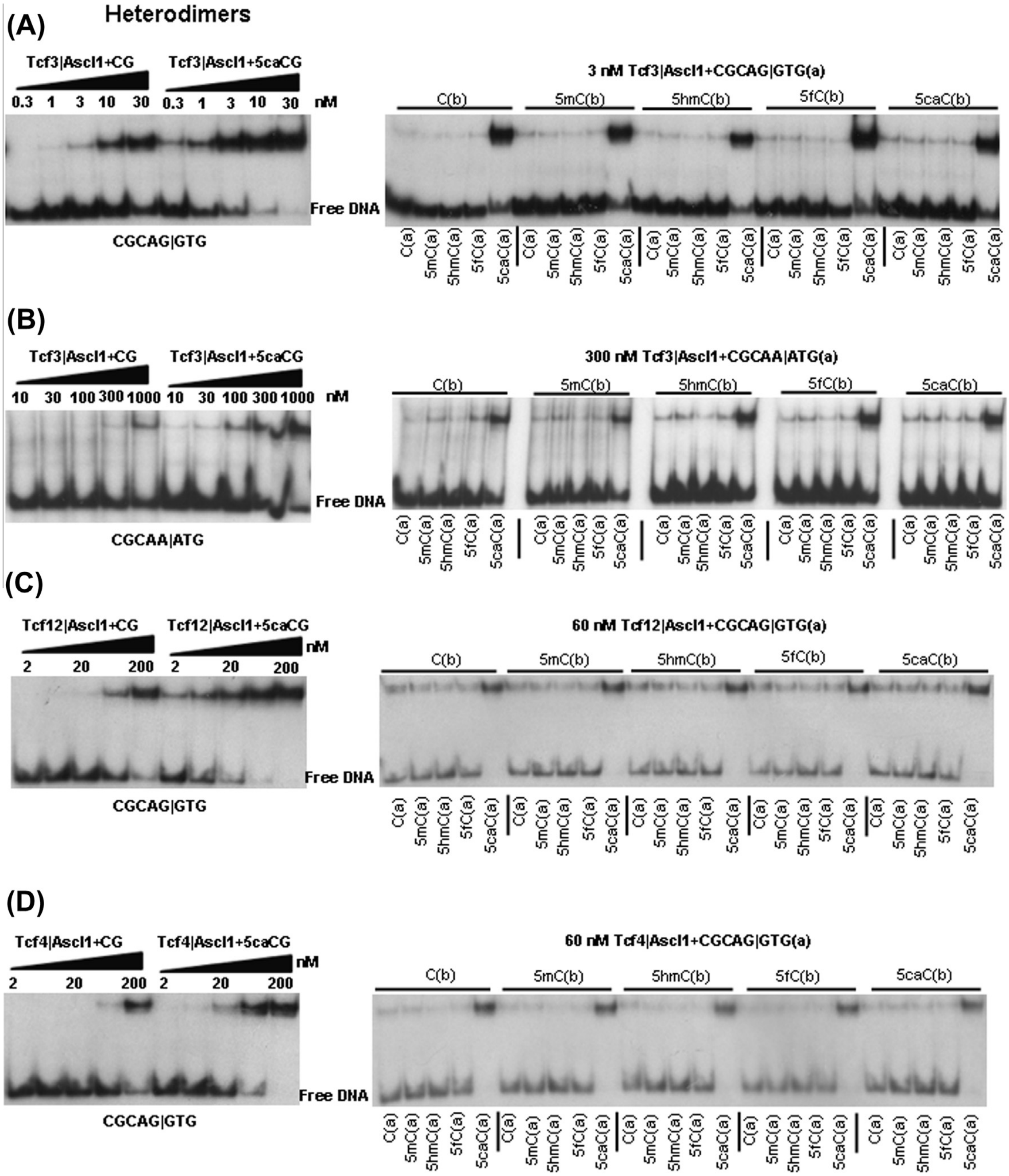
(A)
Heterodimers
Tcf3|Ascl1+CG
Tcf3|Ascl1+5caCG
0.3 1 3 10 30 0.3 1 3 10 30 nM
Free DNA
CGCAG|GTG
3 nM Tcf3|Ascl1+CGCAG|GTG(a)
C(b) 5mC(b) 5hmC(b) 5fC(b) 5caC(b)
C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a)
(B)
Tcf3|Ascl1+CG
Tcf3|Ascl1+5caCG
10 30 100 300 1000 10 30 100 300 1000 nM
Free DNA
CGCAA|ATG
300 nM Tcf3|Ascl1+CGCAA|ATG(a)
C(b) 5mC(b) 5hmC(b) 5fC(b) 5caC(b)
C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a)
(C)
Tcf12|Ascl1+CG
Tcf12|Ascl1+5caCG
nM
2 20 200 2 20 200
Free DNA
CGCAG|GTG
60 nM Tcf12|Ascl1+CGCAG|GTG(a)
C(b) 5mC(b) 5hmC(b) 5fC(b) 5caC(b)
C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a)
(D)
Tcf4|Ascl1+CG
Tcf4|Ascl1+5caCG
nM
2 20 200 2 20 200
Free DNA
CGCAG|GTG
60 nM Tcf4|Ascl1+CGCAG|GTG(a)
C(b) 5mC(b) 5hmC(b) 5fC(b) 5caC(b)
C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a) C(a) 5mC(a) 5hmC(a) 5fC(a) 5caC(a)

Fig. 2. Tcf3|Ascl1, Tcf12|Ascl1, and Tcf4|Ascl1 heterodimers binding modified CGs in two 8-mer: CGCAG|GTG and CGCAA|ATG. (A) EMSA showing a half-log dilution from 30 nM to 0.3 nM for the Tcf3|Ascl1 heterodimer binding two dsDNA 28-mers containing CGCAG|GTG, unmodified and containing 5caC in the CG dinucleotide. The right panel shows 3 nM Tcf3|Ascl1 heterodimer binding to 25 DNAs (see Fig. 1). (B) EMSA showing a dilution from 1000 nM to 10 nM for the Tcf3|Ascl1 heterodimer binding two dsDNA 28-mers containing a weak E-Box (CGCAA|ATG). The right panel shows binding of 300 nM Tcf3|Ascl1 heterodimer to 25 DNAs. (C) EMSA showing a dilution from 200 nM to 2 nM for the Tcf12|Ascl1 heterodimer binding two dsDNAs 28-mers containing CGCAG|GTG. The right panel shows 60 nM Tcf12|Ascl1 heterodimer binding to 25 dsDNAs. (D) Tcf4|Ascl1.

homodimer structure that interacts with the base 5-bp from the center of the dyad is conserved in four B-HLH proteins examined and may explain preferentially bind of 5caC.

4. Discussion

Three oxidative products of 5mC have recently been identified in mammalian genomes and their biological significance is being investigated [11]. Their abundance varies in tissues [10] suggesting they are regulated intermediates with the potential to have biological functions. A potential function of the three oxidative products of 5mC is to change the sequence-specific DNA binding of TFs. We used EMSA and CD spectroscopy to examine the effect of five cytosine nucleotides (C, 5mC, 5hmC, 5fC, and 5caC) on binding of 4 B-HLH homodimers and 3 B-HLH heterodimers both binding to DNA. We examined 25 DNAs containing different combinations of modified C on the two Cs in the CG dinucleotide of the ‘‘flank’’ of the E-box 8-mer **C**GCAG|GTG. When the cytosine in bold is carboxylated (5caC), binding in increased for all 4 homodimers and 3 heterodimers. The Tcf3|Ascl1 heterodimer showed the strongest preferential of ~10-fold, more than either homodimer. However, we do not know which monomer in the Tcf3|Ascl

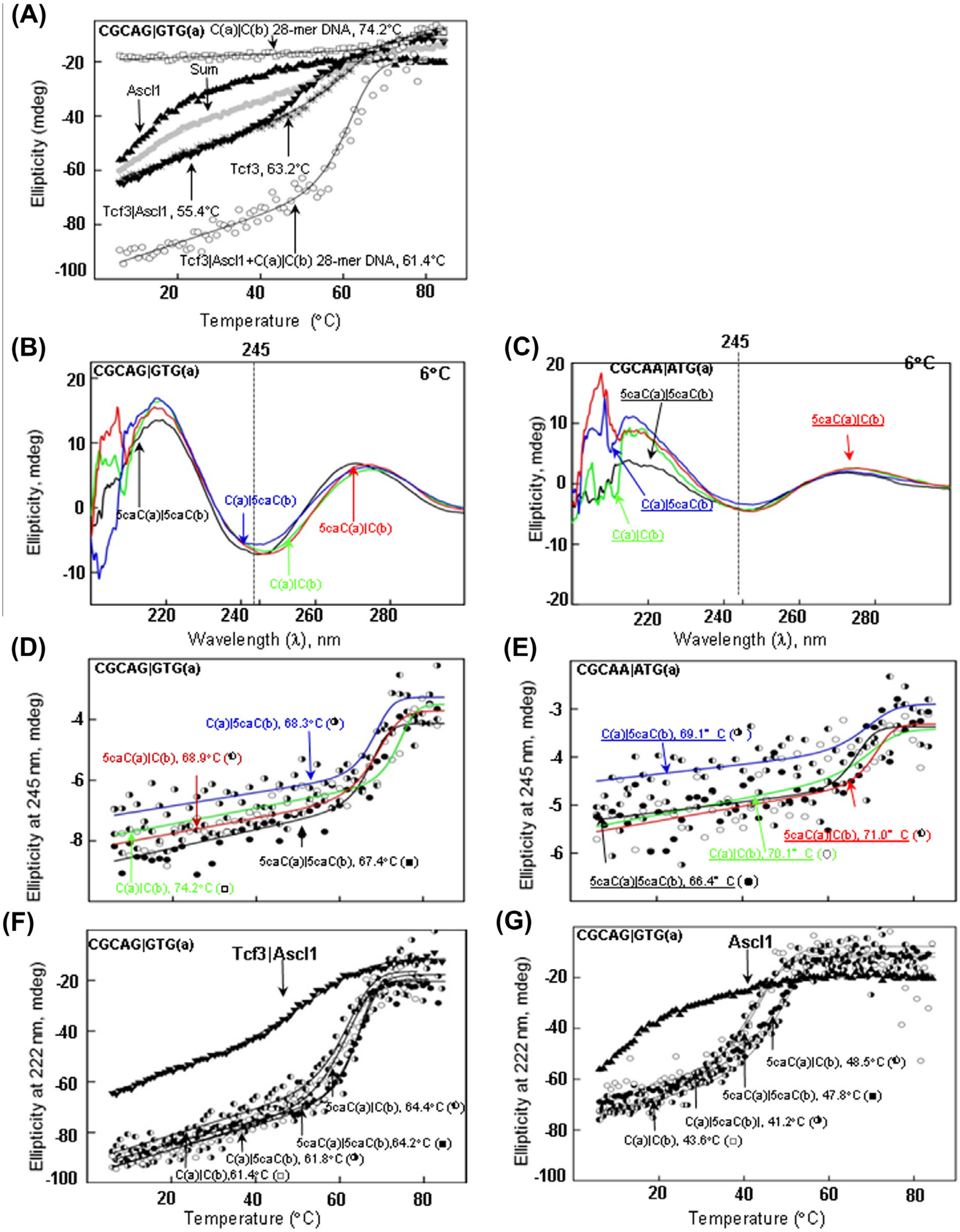

Fig. 3. (A) Circular dichroism spectra and thermal denaturation of DNA, Protein, and DNA–Protein complex. CD at 222 nm of the thermal stability of 2 μM Tcf3 homodimer (*), Ascl1 homodimer (▲), and Tcf3Ascl1 heterodimer in the absence (▼), or presence of dsDNA 28-mer containing the unmodified E-Box CGCAG|GTG (○). The grey circles show the sum of the Tcf3 and Ascl1 thermal denaturation curves. Thermal denaturation at 245 nm of 2 μM dsDNA 28-mer containing CGCAG|GTG (□). (B and C) CD spectra from 200 nm to 300 nm at 6 °C for four dsDNAs that vary 5caC in the CG dinucleotide (2 μM) containing strong E-box (CGCAG|GTG) and weak E-Box (CGCAA|ATG). (D and E) Thermal stability of the four dsDNA 28-mers described in B monitored by circular dichroism at 245 nm. A fitted curve to a two-state transition is shown. (F and G) Thermal stability of 2 μM Tcf3|Ascl1heterodimer and Ascl1 homodimer monitored at 222 nm in the absence (▼) or presence of the four DNAs described above.

heterodimer is binding 5caC. The heterodimer could exist as an ensemble of two states, one where theTcf3 monomer is interacting with 5caC and the second where the Ascl1 monomer is interacting with 5caC. Thus, changing the monomer which is ∼20 Å away from 5caC can change the preferential binding of the second monomer in the dimer. Elucidating the allosterical mechanisms acting over long distances would be interesting to unravel.

Table 3
The CD thermal denaturation monitored at 245 nm for four dsDNA 28-mers containing CGCAG|GTG (a) where the two cytosines in the CG dinucleotide are either C or 5caC. Thermal denaturation of Tcf3|Ascl1 heterodimer and Ascl1 homodimer monitored at 222 nm bound to four DNAs.

DNA (E-Box)	CGCAG|GTG (°C)		
	DNA (°C ± S.E.)	DNA + Tcf3|Ascl1 (°C ± S.E.)	DNA + Ascl1 (°C ± S.E.)
C(a)|C(b) (○)	74.2 ± 0.27	61.4 ± 0.15	43.6 ± 0.65
C(a)|5caC(b) (◑)	68.3 ± 0.27	61.8 ± 0.18	41.2 ± 0.51
5caC(a)|C(b) (◐)	68.9 ± 0.30	64.4 ± 0.18	48.5 ± 0.38
5caC(a)|5caC(b) (●)	67.4 ± 0.28	64.2 ± 0.07	47.8 ± 0.15

Table 4
The CD thermal denaturation monitored at 245 nm for four dsDNA 28-mers containing CGCAA|ATG (a) where the cytosine in the CG dinucleotide are either C or 5caC.

DNA (E-Box)	CGCAA|ATG (°C)	
	Exp # 1 (°C ± S.E.)	Exp # 2 (°C ± S.E.)
C(a)|C(b) (○)	70.1 ± 1.32	72.9 ± 0.75
C(a)|5caC(b) (◑)	69.1 ± 1.04	69.1 ± 1.18
5caC(a)|C(b) (◐)	71.0 ± 0.55	69.9 ± 0.76
5caC(a)|5caC(b) (●)	66.4 ± 0.50	67.3 ± 0.60

B-HLH TFs recognize E-box sequences (CAN|NTG) [20] (for clarity, we place a vertical line in the center of B-HLH dyad). Tcf3 (aka E12, E47) heterodimerizes with different B-HLH proteins in different tissues in mouse [25] and human [20]. For example, Tcf3 heterodimerizes with myoD to drive muscle differentiation [26] and NeuroD to drive neuron differentiation [27].

There are over 60 members of the B-HLH family of transcription factors or proteins dimerize as homodimers and heterodimers and binds to E-Box like sequences (CAN|NTG) [20]. Some B-HLH members, e.g. Myc|Max heterodimers, drive cell growth [21] while other members, e.g. Tcf3|MyoD heterodimers, drive cell differentiation [28]. The various cytosine modifications in E-Box motifs may add an additional layer to DNA binding specificity of this family of proteins. We propose that arginine in the E47 homodimer (Tcf3) that is interacting with the same base as the critical 5caC may mediate the preferential binding to 5caC [19]. This arginine is conserved in Tcf family members and their dimerization partners but not for the B-HLH proteins involved in cell growth like Myc and Max. Potentially when 5caC is produced during the demethylation of tissue specific enhancers [3], Tcf3 and its various heterodimer partners bind these DNA sequences and shift the cell toward differentiation and away from B-HLH dimers involved in cell growth that do not have the arginine hypothesized to bind 5caC.

The abundance of oxidation products of 5mC in cells can be modulated by either activating TET enzymes or inactivating thymine DNA glycosylase (TDG)-mediated base excision repair [29]. Determining if B-HLH proteins bind to 5caC containing CGCAN|NTG in cells is difficult because their occurrences in genome are rare [10]. If experimental systems can be identified where these modification are more abundant, it may become feasible. Methods have been developed to determine the occurrence of 5hmC [30] and 5fC [31] at single CG dinucleotide resolution. Development of methods to determine 5caC in the genome at CG

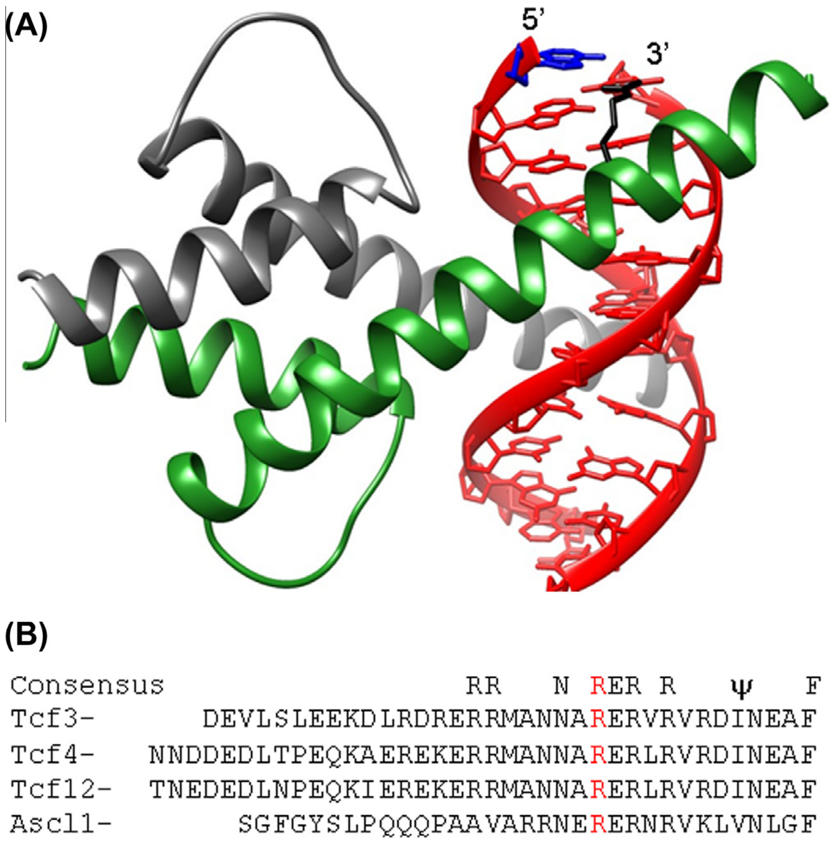
(A)
5'
3'
(B)
Consensus RR N RER R ψ F
Tcf3- DEVLSLEEKDLRDRERRMANNARERVRVRDINEAF
Tcf4- NNDDEDLTPEQKAEREKERRMANNARERLRVRDINEAF
Tcf12- TNEDEDLNPEQKIEREKERRMANNARERLRVRDINEAF
Ascl1- SGFGYSLPQQQPAAVARRNERERNRVKLVNLGF

Fig. 4. Crystal structure of transcription factor E47 (Tcf3) homodimer bound to DNA. The DNA is in red. One monomer is in grey, the second monomer is in green, the adenine in the same position as the critical 5caC is in blue and the invariant arginine is in black. (B) Amino acid sequence for the DNA regions of the B-HLH transcription factors [28] of Tcf3, Tcf4, Tcf12, and Ascl1. The invariant arginine that interacts with the adenine in the E47 homodimer|DNA complex is in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dinucleotide resolution is necessary. In summary, some B-HLH proteins preferentially bind the E-Box motif (CAN|NTG) with a CG dinucleotide on the flank when it contains 5caC.

Funding

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

Acknowledgments

We thank Peter Schuck, NIBIB and Grzegorz Piszczek, NHLBI, NIH for advice on CD spectroscopy.

References

[1] A. Bird, M. Taggart, M. Frommer, et al., A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA, Cell 40 (1985) 91–99.

[2] C. Vinson, R. Chatterjee, P. Fitzgerald, Transcription factor binding sites and other features in human and Drosophila proximal promoters, Subcell Biochem. 52 (2011) 205–222.

[3] R. Chatterjee, C. Vinson, CpG methylation recruits sequence specific transcription factors essential for tissue specific gene expression, Biochim. Biophys. Acta 1819 (2012) 763–770.

[4] C. Vinson, R. Chatterjee, CG methylation, Epigenomics 4 (2012) 655–663.

[5] J.M. Rozenberg, A. Shlyakhtenko, K. Glass, et al., All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues, BMC Genomics 9 (2008) 67.

[6] S.B. Baylin, P.A. Jones, A decade of exploring the cancer epigenome – biological and translational implications, Nat. Rev. Cancer 11 (2011) 726–734.

[7] V. Rishi, P. Bhattacharya, R. Chatterjee, et al., CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 20311–20316.

[8] Y. Liu, H. Toh, H. Sasaki, et al., An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence, Genes Dev. 26 (2012) 2374–2379.

[9] I.K. Mann, R. Chatterjee, J. Zhao, et al., CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo, Genome Res. 23 (2013) 988–997.

[10] S. Ito, L. Shen, Q. Dai, et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science 333 (2011) 1300–1303.

[11] W.A. Pastor, L. Aravind, A. Rao, TETonic shift: biological roles of TET proteins in DNA demethylation and transcription, Nat. Rev. Mol. Cell Biol. 14 (2013) 341–356.

[12] A. Inoue, L. Shen, Q. Dai, et al., Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development, Cell Res. 21 (2011) 1670–1676.

[13] C.G. Spruijt, F. Gnerlich, A.H. Smits, et al., Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, Cell 152 (2013) 1146–1159.

[14] K.N. Lam, H. van Bakel, A.G. Cote, et al., Sequence specificity is obtained from the majority of modular C_2H_2 zinc-finger arrays, Nucleic Acids Res. 39 (2011) 4680–4690.

[15] C.R. Vinson, T. Hai, S.M. Boyd, Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design, Genes Dev. 7 (1993) 1047–1058.

[16] S. Ahn, M. Olive, S. Aggarwal, et al., A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos, Mol. Cell. Biol. 18 (1998) 967–977.

[17] M. Olive, D. Krylov, D.R. Echlin, et al., A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis, J. Biol. Chem. 272 (1997) 18586–18594.

[18] D. Krylov, I. Mikhailenko, C. Vinson, A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions, EMBO J. 13 (1994) 2849–2861.

[19] T. Ellenberger, D. Fass, M. Arnaud, et al., Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer, Genes Dev. 8 (1994) 970–980.

[20] C. Murre, G. Bain, M.A. van Dijk, et al., Structure and function of helix-loop-helix proteins, Biochim. Biophys. Acta 1218 (1994) 129–135.

[21] D. Krylov, K. Kasai, D.R. Echlin, et al., A general method to design dominant negatives to B-HLHZip proteins that abolish DNA binding, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 12274–12279.

[22] V. Rishi, J. Gal, D. Krylov, et al., SREBP-1 dimerization specificity maps to both the helix-loop-helix and leucine zipper domains: use of a dominant negative, J. Biol. Chem. 279 (2004) 11863–11874.

[23] J. Kypr, I. Kejnovska, D. Renciuk, et al., Circular dichroism and conformational polymorphism of DNA, Nucleic Acids Res. 37 (2009) 1713–1725.

[24] P.C. Ma, M.A. Rould, H. Weintraub, et al., Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation, Cell 77 (1994) 451–459.

[25] A. Lazorchak, M.E. Jones, Y. Zhuang, New insights into E-protein function in lymphocyte development, Trends Immunol. 26 (2005) 334–338.

[26] H. Weintraub, R. Davis, S. Tapscott, et al., The myoD gene family: nodal point during specification of the muscle cell lineage, Science 251 (1991) 761–766.

[27] M.H. Farah, J.M. Olson, H.B. Sucic, et al., Generation of neurons by transient expression of neural bHLH proteins in mammalian cells, Development 127 (2000) 693–702.

[28] C.R. Vinson, K.C. Garcia, Molecular model for DNA recognition by the family of basic-helix-loop-helix-zipper proteins, New Biol. 4 (1992) 396–403.

[29] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, Nature 502 (2013) 472–479.

[30] M. Yu, G.C. Hon, K.E. Szulwach, et al., Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome, Cell 149 (2012) 1368–1380.

[31] C.X. Song, K.E. Szulwach, Q. Dai, et al., Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming, Cell 153 (2013) 678–691.