Biochemical and Biophysical Research Communications xxx (2014) xxx–xxx

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 Khund 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 xxxx

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.

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/

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 (PureExpress, 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**CGCAG|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**CGCAA|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: MGHMVHRPWIQDEVLSLEEKDLRDRERRMANNARERVRVR
DINEAFRELGRMCQLHLKSDKAQTKLLILQQAVQVILGLEQQVRERN
LNPKAACGGRTRIVSAHNSENEL
Tcf4: MGNNDDEDLTPEQKAEREKERRMANNARERLRVRDINEAFK
ELGRMVQLHLKSDKPQTKLLILHQAVAVILSLEQQVRERNLNPKAAC
LKRREEELHHHHHH
Tcf12: MGSTNEDEDLNPEQKIEREKERRMANNARERLRVRDINEAF
KELGRMCQLHLKSEKPQTKLLILHQAVAVILSLEQQVRERNLNPKAA
CLKRREEEL
Ascl1: MASGFGYSLPQQQPAAVARRNERERNRVKLVNLGFATLREH
VPNGAANKKMSKVETLR SAVEYIRALQQLLDEHDAVSAAFQAGVLS
PELMDYKDD DDKHMASMTGGQQMGRDP

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-mers (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).

Please cite this article in press as: J.P. Golla et al., 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, Biochem. Biophys. Res. Commun. (2014), http://dx.doi.org/10.1016/j.bbrc.2014.05.018