

# TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine

*Cell Research* (2010) 20:1390-1393. doi:10.1038/cr.2010.156; published online 16 November 2010

## Dear Editor,

DNA methylation, which often occurs at the 5-carbon position of cytosine (5mC) located in CpG dinucleotide, is a key epigenetic hallmark and serves as a major epigenetic mechanism for establishing X-inactivation, parental imprinting and silencing retrotransposable elements during early embryogenesis in mammals. Accumulative evidence also suggests that DNA methylation plays key roles in transcriptional regulation [1]. Observations of active global loss of 5mC during early development and local loss of 5mC concurrent with gene activation have led to the search for enzymes capable of active DNA demethylation of 5mC [2].

Since the initial discovery of the first histone demethylase LSD1 [3], we have taken a candidate approach to identify potential DNA (5mC) demethylases that may be involved in the dynamic regulation of DNA methylation status. We subsequently became intrigued by the family of CXXC domain-containing enzymes, the members of which play critical roles in chromatin functioning by modifying histone or DNA methylation (Figure 1A) [4]. Members of this growing family include: myeloid-lymphoid leukemia protein, MLL (CXXC7), a histone methyltransferase, DNMT1 (CXXC9), a DNA 5mC methyltransferase, and F-box-containing and leucine-rich proteins, FBXL11/JHDM1a (CXXC8) and FBXL10/JHDM1b (CXXC2), which have been recently identified as histone demethylases. We therefore explored the possibility that an enzyme capable of oxidative 5mC demethylation was an as-of-yet unidentified member of this gene family.

Of the 11 members of the CXXC domain-containing gene family identified in the human genome by bioinformatics, 4 have been characterized as chromatin modifiers (Supplementary information, Table S1). We asked whether any of the remaining seven members (1) share homology with reported oxidoreductive enzymes; (2) contain a potential oxidative catalytic domain; or (3) con-

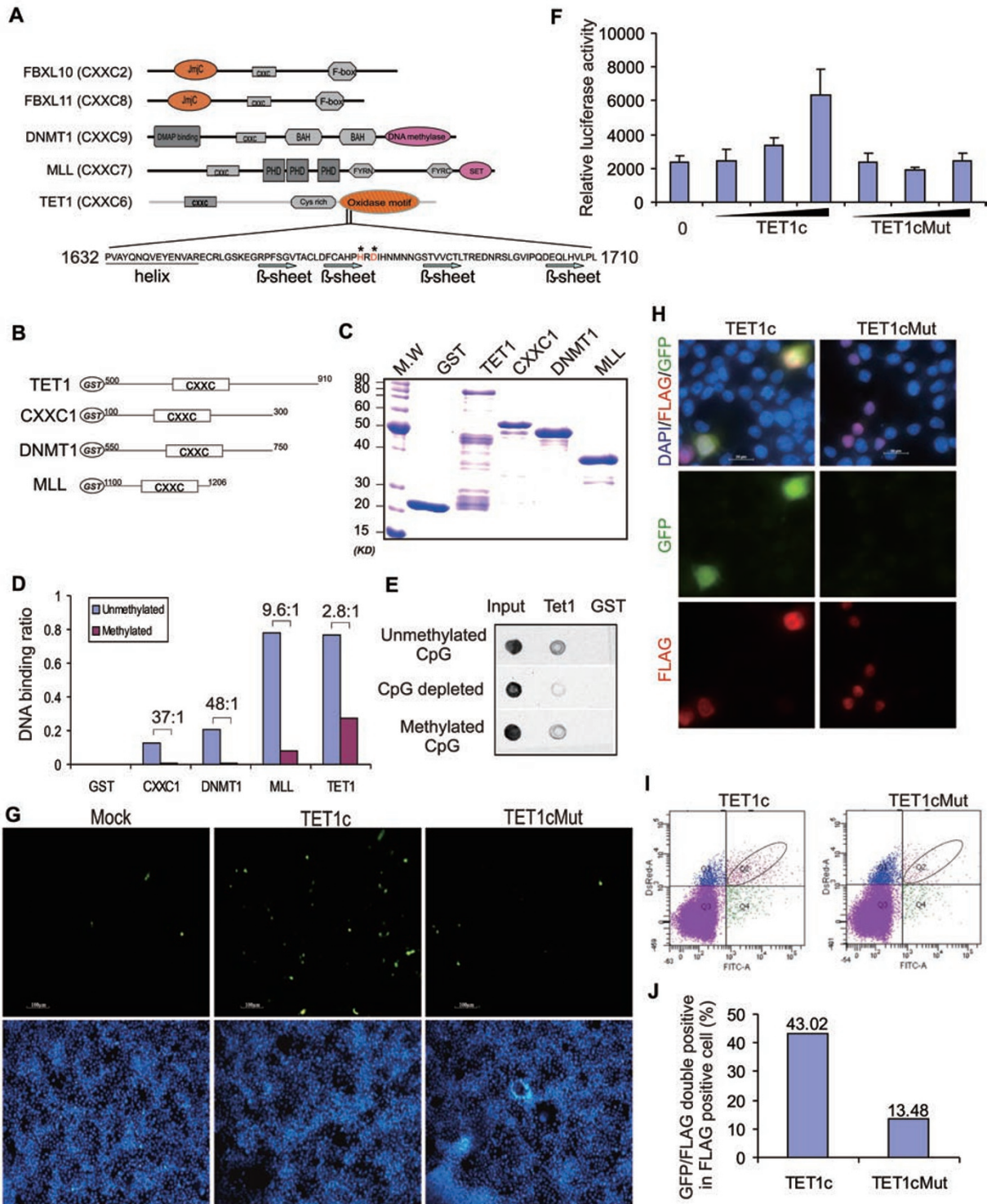
tain cofactor (i.e. NAD(P)<sup>+</sup>, FAD, iron,  $\alpha$ -ketoglutarate)-binding motifs.

Using these criteria, we decided to pursue CXXC6/TET1/LCX1 [5], which contains a unique C-terminal domain in addition to the N-terminal CXXC domain and also presents a potential iron-coordinating triad motif (HxD/E.....H..) in the C-terminal region (amino acid (aa) 1 670-2 136). Secondary structure predictions revealed a potential double-strand  $\beta$ -helix (DSBH) fold within the C-terminal region, a signature for the dioxygenase superfamily, including the jmjC family of histone demethylases [6]. There is also a cysteine-rich region immediately upstream of DSBH domain, which could potentially serve as a redox center for putative oxidoreductase activity. In summary, features within the TET1 C-terminal domain strongly suggest that TET1 may function as a non-heme iron-dependent dioxygenase, and features of its N-terminal CXXC domain predict a role in DNA or chromatin association, together making TET1 an excellent candidate as a novel DNA/chromatin-modifying enzyme.

Most of the characterized CXXC domain-containing proteins are transcription factors/cofactors, which can bind directly to unmethylated CpG dinucleotides in DNA through their CXXC domain [7]. We first investigated the ability and specificity of TET1 CXXC domain binding to DNA. The TET1-CXXC (aa 500-910) domain, along with the CXXCs from MLL, CXXC1 and DNMT1, were GST tagged (Figure 1B) and expressed in, and purified from, bacterial cultures (Figure 1C). We used GST pull-down assays to evaluate the ability of these CXXC domains to bind to double-stranded DNA oligonucleotides, either CG-unmethylated (Supplementary information, Figure S1, upper) or methylated at the 5-carbon position of cytosine in CG dinucleotides (Supplementary information, Figure S1, lower). Materials and Methods are described in Supplementary information, Data S1. The absence of binding of GST to both methylated and unmethylated oligonucleotides and the highly specific

binding of GST-CXXC1, -DNMT1 and -MLL to unmethylated oligonucleotides with minimal binding to the methylated oligonucleotides, in agreement with previous reports, validate the utility of this assay (Figure 1D, first four pairs of the columns). We found that TET1-

CXXC bound strongly to unmethylated CG-containing oligonucleotides similar to CXXC domain of its fusion partner MLL (Figure 1D, compare the last two blue columns). Strikingly, a substantial amount of methylated oligonucleotides bound to TET1-CXXC (Figure 1D, last



red column). We compared the ratio of unmethylated vs methylated oligonucleotide binding for each CXXC and observed that this ratio is significantly lower for TET1-CXXC than for the CXXC1-CXXC, DNMT1-CXXC and MLL-CXXC (2.8:1 vs 48:1, 37:1 and 9.6:1, respectively). The CXXC domain of TET1 is unique in that it lacks a typical “KFGG” motif found in the majority of characterized CXXC domains, which may lend it a structural advantage for binding methylated CpG groups [7]. Nevertheless, in concurrence with the other members of the CXXC domain family members, TET1-CXXC does not exhibit binding to CG-depleted oligonucleotides (Figure 1E). Thus, these results suggest that the CXXC domain of the TET1 protein, like that of other family members, binds specifically to CG-containing DNA sequences. Furthermore, CXXC domain of the TET1 protein exhibits greater binding to CG-methylated DNA oligonucleotides than other family members. This distinctive property of the TET1 protein supports our contention for its enzymatic action on methylated DNA.

We next examined the extent to which the TET1 C-terminal domain may catalyze 5mC demethylation by an oxidative mechanism. We hypothesized that hydroxylation of the methyl group on 5-C-CH<sub>3</sub> (5mC) and resultant formation of 5-C-CH<sub>2</sub>OH (5-hydroxyl methylcytosine (5hmC)) would be the first step in a TET1-mediated demethylation process (Supplementary information, Figure S2A).

Indeed, our *in vitro* and *in vivo* enzymatic activity assays were able to show that TET1 C-terminal domain is an active 5-methylcytosine (5mC) hydroxylase, which converts 5mC to 5-hydroxyl methylcytosine (5hmC), and that such enzymatic action *in vivo* may result in 5mC to C demethylation (Supplementary information, Figures S2 and S3).

Methylation of CpG is inversely correlated with the transcriptional activity in mammalian cells. Therefore, we next employed a luciferase reporter gene system to examine the ability of TET1 to relieve gene silencing mediated by 5mC DNA methylation. As expected, a fully *in vitro*-methylated reporter gene plasmid shows dramatically reduced activity in comparison with unmethylated plasmids (Supplementary information, Figure S4A). Strikingly, co-transfection of the TET1c constructs with fully methylated reporter gene plasmid revealed increased luciferase activity in a dosage-dependent manner for TET1c, but not for the enzymatically dead mutant TET1c (Figure 1F, compare lanes 2, 3 and 4 to 5, 6 and 7).

To confirm this putative gene activation function of TET1, we used pEGFP plasmids containing promoter elements different from the luciferase reporter gene [8]. Fully methylated pEGFP plasmid alone yielded almost no signal in transfected cells (Figure 1G, left panels), compared with unmethylated control plasmids (Supplementary information, Figure S4B). However, co-transfection of the fully methylated pEGFP plasmid with TET1c resulted in a greater number of GFP-positive cells (Figure 1H, middle panels), while co-transfection of the TET1c mutant showed only minimal number of the GFP-positive cells (Figure 1G, right panels). To demonstrate that the observed increase in GFP expression was due to the gene reactivation subsequent to expression of TET1 protein, we fluorescently sorted transfected cells (green-GFP reporter signal and red-TET1flag tagged signal; Figure 1H and 1I). Three-fold more GFP-positive cells were isolated with TET1c-positive cells than that isolated with TET1cMut-positive cells (Figure 1J), suggesting that the enzymatic activity of TET1 may be responsible for relieving the DNA methylation-mediated gene silencing.

In summary, in this study, we have identified and

**Figure 1** DNA-binding property and transcriptional activity of CXXC/TET1. **(A)** Schematic diagram of representative CXXC family members including FBXL10, FBXL11, DNMT1, MLL and TET1. TET1 contains a CXXC motif in its N-terminal domain, and an oxidase-like motif in its the C-terminal domain. Additionally, an iron-binding motif (HxD), indicated by asterisks, is located within the oxidase motif region. **(B,C)** Diagram of GST-tagged CXXC domains from TET1, CXXC1, DNMT1 and MLL **(B)** and purified recombinant proteins separated in PAGE gel **(C)**. **(D)** Binding property of CXXC domain protein to unmethylated (blue columns) and methylated DNA oligonucleotides (red columns). The relative activity was represented by the ratio of bound DNA to input DNA. **(E)** Dot blot analysis confirming the binding specificity of TET1-CXXC to unmethylated and methylated CpG-containing DNA oligonucleotides, but not to CpG-depleted oligonucleotides. **(F)** Relative luciferase activity of fully methylated Gal4-TK-luc reporter gene co-transfected with TET1c or TET1cMut. **(G)** GFP reporter gene assay. GFP (green) signals are readily observed in cells in which the fully methylated pEGFP reporter is co-transfected with TET1c (middle panels), but hardly seen in cells co-transfected with TET1cMut (right panels). DAPI staining indicates the position of the nuclei of the cells (lower panels). 10× magnification. **(H)** Signal from fully methylated pEGFP reporter (GFP, green signal) was readily seen in TET1c-expressing cells (Flag, red signal, left panels), but was hardly seen in TET1cMut-expressing cells (Flag, red signal, right panels). 60× magnification. **(I)** FACS analysis of GFP and Flag-TET1c (or TET1cMut) double-positive cells. There are significantly more GFP (x axis, FITC) and Flag (y axis, DsRed) double-positive cells in the methylated GFP reporter gene /TET1c co-transfected cells (Q2, left panel) than the TET1cMut co-transfected cells (Q2, right panel). **(J)** The percentage of GFP/Flag (red) double-positive cells calculated from FACS analysis.

characterized a new CXXC domain family member, CXXC6/TET1. It binds both to methylated and unmethylated DNA likely through its CXXC domain, and its C-terminal portion possesses a 5'-hydroxylase activity, which oxidizes 5mC in DNA to generate an atypical nucleotide 5-hydroxymethylcytosine (5hmC) both *in vitro* and *in vivo*. Moreover, overexpression of active TET1 in cells not only results in the accumulation of 5hmC, but also promotes DNA demethylation and reactivates DNA methylation-silenced reporter genes (Supplementary information, Figures S2 and S3; Figure 1F and 1G).

Recently, the 5mC hydroxylase activity of TET1 family members that converts 5mC to 5hmC has been independently reported [9, 10]. Furthermore, the functional role of TET1 as well as its product 5hmC in gene regulation has been proposed for ES cell renewal and lineage determination [9, 10]. However, neither of these reports directly showed that the hydroxylation activity is important for TET1-mediated transcription regulation. In this current study, we used two independent reporter gene systems to show that the enzymatically inactive TET1 was impaired in transcriptional regulation, thereby, for the first time, providing the direct link between TET1's hydroxylase activity and transcription regulation. We also report that TET1's N-terminal domain containing CXXC domain binds to both methylated and unmethylated CpG-rich DNA showing, for the first time, dual binding activities of the CXXC domain, thus providing additional insight into the mechanism for how TET1 may be targeted to exert influence in various regions of the genome.

Taken together, we provide evidence to suggest that TET1 is likely a novel DNA-binding transcriptional regulator involved in the dynamic regulation of DNA methylation; it promotes gene activation likely through erasure (alteration) of the gene silencing epigenetic mark, 5mC. Thus, these findings may provide a novel mechanism for dynamic regulation of DNA methylation and eukaryotic gene transcription.

## Acknowledgments

We acknowledge Dr G P Pfeifer for providing us with the methylated pEGFP-N1 plasmid for our initial experiment. We are very thankful for the helpful discussions and technical supports provided by Drs E Fox and C Li at Harvard Medical School. We also appreciate Drs Stephen P Sugrue and Ursula Kaiser for their critical reading of the manuscript. Y Geno Shi is a PEW scholar. This

work was supported by NIH Grants GM078458 and DK077036.

Haikuo Zhang<sup>1</sup>, Xin Zhang<sup>1</sup>, Erin Clark<sup>1</sup>,  
Michelle Mulcahey<sup>2</sup>, Stephen Huang<sup>2</sup>,  
Yujiang Geno Shi<sup>1</sup>

<sup>1</sup>Division of Endocrinology, Diabetes and Hypertension, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Avenue, Boston, MA 02115, USA; <sup>2</sup>Division of Endocrinology, Children's hospital Boston, Harvard Medical School, Boston, MA 02115, USA

Correspondence: Yujiang Geno Shi  
Tel: +617-525-8097, Fax: +617-582-6193  
E-mail: yujiang\_shi@hms.harvard.edu

## References

- 1 Tate PH, Bird AP. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* 1993; **3**:226-31.
- 2 Gehring M, Reik W, Henikoff S. DNA demethylation by DNA repair. *Trends Genet* 2009; **25**:82-90.
- 3 Shi Y, Lan F, Matson C, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 2004; **119**:941-53.
- 4 Katoh M, Katoh M. Identification and characterization of human *CXXC10* gene *in silico*. *Int J Oncol* 2004; **25**:1193-9.
- 5 Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res* 2002; **62**:4075-4080.
- 6 Iyer LM, Tahiliani M, Rao A, Aravind L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 2009; **8**:1698-710.
- 7 Allen MD, Grummitt CG, Hilcenko C, et al. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *EMBO J* 2006; **25**:4503-12.
- 8 Jin S-G, Guo C, Pfeifer GP. GADD45A does not promote DNA demethylation. *PLoS Genet* 2008; **4**:e1000013.
- 9 Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009; **324**:930-5.
- 10 Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 2010; **466**:1129-1133.

(Supplementary information is linked to the online version of the paper on *Cell Research* website.)