Communication

Exploiting Substrate Promiscuity to Develop Activity-Based Probes for TET Family Enzymes

Uday Ghanty, Jamie E. DeNizio, Monica Yun Liu, and Rahul M. Kohli

J. Am. Chem. Soc., **Just Accepted Manuscript** • Publication Date (Web): 06 Dec 2018

Exploiting Substrate Promiscuity to Develop Activity-Based Probes for TET Family Enzymes

Uday Ghanty[†,‡], Jamie E. DeNizio[†,‡], Monica Yun Liu[†,‡], Rahul M. Kohli[†,‡,*]

†Department of Medicine, ‡Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104.

Supporting Information Placeholder

ABSTRACT: TET enzymes catalyze the repeated oxidation of 5-methylcytosine in genomic DNA. Due to the challenges of track-ing reactivity within a complex DNA substrate, chemical tools to probe TET activity are limited, despite these enzyme's crucial role in epigenetic regulation. Here, building on precedents from related Fe(II)/α-ketoglutarate-dependent dioxygenases, we show that TET enzymes can promiscuously act upon cytosine bases with unnatural 5-position modifications. Oxidation of 5-vinylcytosine (vC) in DNA results in the predominant formation of a 5-formylmethylcytosine product that can be efficiently labeled to provide an endpoint read-out for TET activity. The reaction with 5-ethynylcytosine (eyC), moreover, results in the formation of a high-energy ketene intermediate that can selectively trap any active TET isoform as a covalent enzyme-DNA complex, even in the complex milieu of a total cell lysate. Exploiting substrate promiscuity therefore offers a new and needed means to directly track TET activity *in vitro* or *in vivo*.

Methylation of cytosine bases in a CpG context is an important epigenetic modification involved in gene silencing, imprinting, and cellular differentiation. While 5-methylcytosine (mC) represents the most prevalent modification, the discovery of ten-eleven translocation (TET) family enzymes dramatically expanded the complexity of the epigenetic landscape (Figure 1A).[1] TET enzymes can oxidize mC to 5-hydroxymethylcytosine (hmC), which can be further oxidized to 5-formylcytosine (fC) and 5-carboxylcytosine (caC), and these oxidized mC bases (ox-mCs) can both serve as intermediates in DNA demethylation and play independent epigenetic roles.[1-5] The importance of ox-mCs is evident in loss-of-function studies that highlight a host of pathologies, including a strong association with hematologic malignancies.[6]

In spite of the significance of TET enzymes, studying their activity has remained a significant challenge. Most assays employed are of limited throughput, relying upon indirect analysis with coupling enzymes or degradation of DNA into nucleosides, followed by LC/MS/MS or TLC analysis.[1-5,7-9] Similarly, while genetic approaches have provided functional insights, understanding when and where TET enzymes are active has been limited by the lack of any activity-dependent chemical probes.

TET enzymes are Fe(II)/α-ketoglutarate (α-KG)-dependent dioxygenases that act via generation of a reactive Fe(IV)-oxo intermediate that drives proton abstraction and subsequent rearrangements that result in substrate oxidation.[10] The fact that mC, hmC, fC, and even rarely T, can each be substrates for oxidation suggests a degree of intrinsic promiscuity.[11-14] Tolerance to substrate variations has been previously observed with another Fe(II)/α-KG-dependent dioxygenase, thymine hydroxylase (TH), that acts on the free thymine nucleobase; Stubbe and coworkers showed that various unnatural modifications at the 5-position of uracil could be oxidized by TH.[15,16] We reasoned that if TET enzymes are similarly active on unnatural 5-modified cytosine bases built into DNA, then those alternative substrates might be useful for tracking TET activity.

Given the steric constraints evident in the DNA-bound structure of a truncated, active version of human TET2 (TET2-CS) (Figure S1),[12] we focused on exploring TET reactivity with two substituents of comparable size to ox-mCs, 5-vinylcytosine (vC) and 5-ethynylcytosine (eyC) (Figure 1). To this end, we first generated DNA containing the desired modifications. A 20-bp oligonucleotide substrate (S20) was synthesized containing 5-iodocytosine (iC) at the 5′-end. This terminal base was then converted to vC (S20-vC) or eyC (S20-eyC), respectively, by a modified Suzuki-Miyaura cross-coupling reaction[17,18] or a modified Sonogashira reaction[19] performed on-resin (Figure S2). The purified oligonucleotides were hybridized to an unmodified complement strand and reacted either with or without TET2-CS, under conditions optimized for activity. When analyzed by ESI-MS, we readily detected new

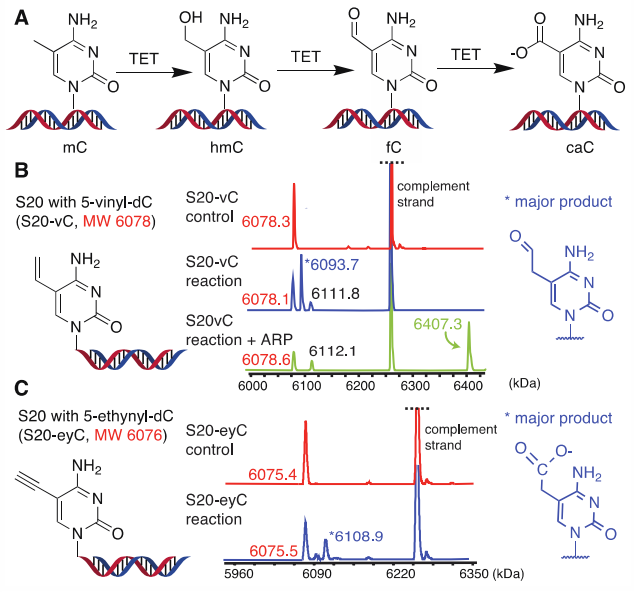

Figure 1. Promiscuity of TET activity on unnatural 5-modified cytosine bases in DNA. (A) TET enzymes can accept three different potential substrates (mC, hmC and fC), generating three associated products (hmC, fC and caC). (B) S20-vC (50 nM) was reacted with or without TET2-CS (0.6 μM) and the purified DNA analyzed by ESI-MS. The masses listed are consistent with the substrate oligonucleotide strand (red), the major product (blue), minor product (black), and the major product derivatized with hydroxylamine aldehyde reactive probe (ARP) (green). In concert with reduction experiments (Figure S3), ARP conjugation identifies the major product as 5-formylmethylcytosine. (C) S20-eyC was similarly reacted and analyzed by ESI-MS. The substrate oligonucleotide mass is shown in red and the noted product mass in blue, corresponding to 5-carboxymethylcytosine.

products, suggesting that TET2-CS is able to accommodate these unnatural substrates for oxidation (Figure 1B, 1C). For S20-vC, the major product showed a mass increase consistent with a single oxidation event and a heavier minor product, that could be consistent with a diol. For S20-eyC, the recovered product detected was consistent with a carboxymethyl product; however, we consistently also noted a decreased recovery of DNA after the reaction.

We sought to better characterize the reaction products resulting from S20-vC oxidation. According to the TET reaction mechanism, the Fe(IV)-oxo intermediate could potentially abstract a proton from either carbon of the vinyl group. Subsequent hydroxyl radical rebound and tautomerization could generate either a 5-acetyl or 5-formylmethyl product (Figure S3). To confirm that one of these singly-oxidized species represented the major product, we incubated S20-vC reaction product with an aldehyde/ketone reactive hydroxylamine derivative ($ARP\text{-}ONH_2$) and observed a quantitative shift in the major product (Figure 1B). Then, to differentiate between the proposed possibilities, the S20-vC product was incubated with $Na(OAc)_3BH$ prior to labeling with $ARP\text{-}ONH_2$ (Figure S3A). A mass shift was no longer observed, and the ESI-MS was consistent with reductive amination (Figure S3B), which would only be possible if 5-formylmethylcytosine is the major product. To also probe the minor product (~5:1 ratio, Figure S3B) identity, we synthesized 5-(1,2-dihydroxyethyl)-C-containing DNA, which could plausibly arise via hydrolysis of an epoxide intermediate, as previously observed for the reaction of 5-vinyluracil with TH;[15,16] 2D-TLC co-spotting supported the identification of the diol as the minor product (Figure S4).

We were intrigued that, unlike mC, which can be oxidized multiple times, S20-vC appears to only be oxidized once, generating the aldehyde. Structural modeling suggested that while vC can be accommodated, the oxidation product might be sterically excluded from further oxidation (Figure S1). Consistent with this hypothesis, we generated a bulkier 5-propenylcytosine containing substrate (S20-pC) and did not observe product formation by 1D-TLC (Figure S5).

Having established that TET enzymes show promiscuity towards these unnatural 5-modified substrates, we next aimed to exploit this reactivity in the development of new chemical probes for TET activity. To facilitate the generation of DNA probes, we synthesized nucleotide triphosphate versions of 5-vinylcytosine and 5-ethynylcytosine (Figure S6). A 90-bp DNA template was designed containing ten central CpG pairs (S90) and lacking other cytosines outside of

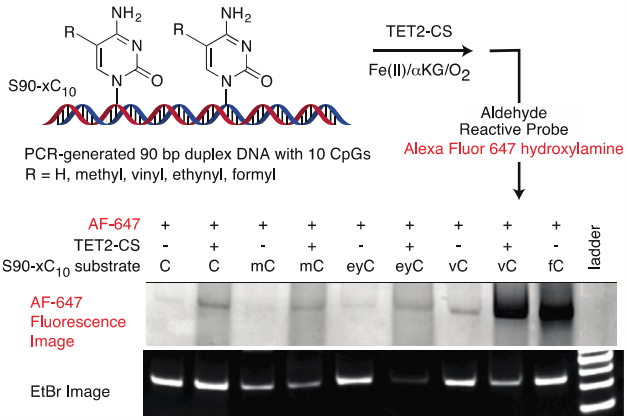

Figure 2. 5-vinylcytosine offers an endpoint assay for TET activity. PCR using modified dCTPs was used to generate 90-mer DNA substrates with ten CpG modifications (S90-xC_{10}). The S90-xC_{10} substrates (50 nM) were reacted with or without TET2-CS (0.6 μM) and then derivatized with the fluorescent aldehyde-reactive Alexa Fluor 647 hydroxylamine (AF-647). The reactions were analyzed on a 5% TBE-PAGE gel and imaged for AF-647 fluorescence and then stained with ethidium bromide (EtBr) to detect total DNA. S90-fC10 without TET2-CS oxidation was used as a positive control for AF-647 labeling.

the primers. The template was amplified with modified dCTPs resulting in symmetrically modified DNA (S90-xC_{10}) with vC or eyC at the central CpG sites. Control DNA was also synthesized with unmodified C, mC or fC. The S90-xC_{10} substrates were incubated in the presence or absence of TET2-CS. As aldehyde-reactive probes have proven utility with fC detection,[20] the reaction products were then incubated with the Alexa Fluor 647 hydroxylamine (AF-647). The reactions were run on a polyacrylamide gel and then imaged for AF-647 fluorescence (Figure 2). Across the substrate series, we observed AF-647 labeling of S90-vC_{10} treated with TET2-CS, and not with S90-C_{10}, S90-mC_{10}, or S90-eyC_{10}. The labeling is comparable to that of an S90-fC_{10} control without TET reaction, suggesting efficient conversion. To compare reactivity, we also performed a titration of TET2-CS against S90-xC_{10} substrates and assessed substrate consumption by LC/MS/MS (Figure S7). Consumption of vC and eyC was, respectively, only ~1.7-fold or ~3.5-fold reduced relative to mC, suggesting robust reactivity. In the AF-647 labeling experiment, it is particularly notable that although mC oxidation can generate fC among its products, under these conditions no labeling of S90-mC_{10} was observed, likely due to conversion to caC. The fact that vC generates the aldehyde as the major and terminal product makes this unnatural substrate particularly useful as a means to assay for TET endpoint activity.

In the reaction of S90-eyC_{10} with TET2-CS (Figure 2), we were struck by one added observation: as with S20-eyC, the recovery of S90-eyC_{10} was diminished after reaction. In studying 5-ethynyluracil, Stubbe and colleagues had previously demonstrated that oxidation by TH led to the generation of a highly-reactive ketene intermediate that reacted with and covalently inhibited TH.[16] In a similar vein, Cravatt and colleagues employed aryl acetylenes as activity-based probes of P450 enzymes.[21] Recognizing these precedents, we considered whether a ketene intermediate that could yield either a carboxymethyl product or a stable protein-DNA complex could explain the eyC-DNA loss (Figure 3A).

To explore this hypothesis, we generated the S90-xC_{10} substrates with a 5′-Cy5 labeled primer and reacted either with or without TET2-CS. The reaction products were analyzed by SDS-PAGE and imaged first for Cy5 fluorescence to detect the DNA and then stained to detect protein. With the S90-eyC_{10} alone, and not with S90-C_{10} or S90-mC_{10}, we observed the formation of a new species containing Cy5-fluorescence running at a molecular weight higher than the DNA substrate or TET2 alone (Figure 3B). Eliminating (α-KG, an essential co-substrate, from the reaction results in disappearance of the new species, implying a requirement for TET activity, and formation of the species tracked with pH and salt conditions preferred for optimal TET reactivity (Figure S8). To better identify the species as an enzyme-DNA complex, we generated smaller 59-bp Cy5-labeled DNA substrates containing less CpGs (S59-eyC_3; S59-eyC_1). We also purified the full catalytic domains of TET1, TET2 and TET3. We observed that all combinations of TET enzymes with either S90-eyC_{10} or S59-eyC_3 resulted in activity-dependent complex formation (Figure S9A). The complexes ran at different masses that correlated with the sizes of each isozyme or DNA substrate. As further validation, Proteinase K eliminated detection of the enzyme-DNA complex (Figure S9B). When excess enzyme was reacted with S59-eyC_1, intact DNA was about equivalent to enzyme-DNA complex, suggesting complex formation is at least as efficient as carboxymethyl generation (Figure S10).

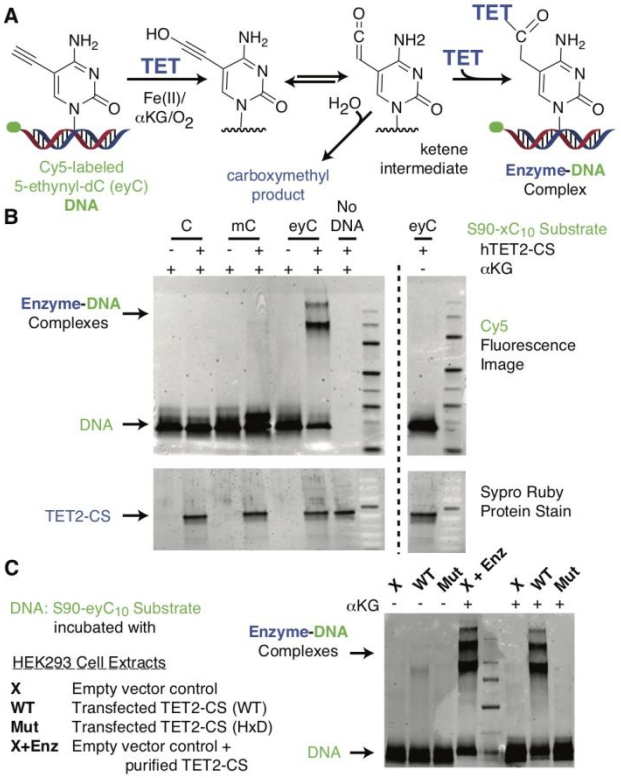

Figure 3. 5-ethynylcytosine is an activity-dependent probe forming enzyme-DNA crosslinks. (A) Oxidation of eyC could result in the formation of a high-energy ketene intermediate that could either yield the carboxymethyl product or an enzyme-DNA complex. (B) Cy5 labeled S90-xC_{10} substrates (50 nM) containing C, mC or eyC were reacted with TET2-CS (0.6 μM). The reactions were analyzed by PAGE and the gels imaged for Cy5 fluorescence or stained for protein (Sypro Ruby). The formation of a new species with eyC was dependent upon activity in the presence of the co-substrate α-KG. (C) Cell extracts from HEK293T cells transfected with empty vector (X), TET2-CS (WT), or catalytically inactive TET2 (Mut) were incubated with S90-eyC_{10} substrate. The enzyme-DNA crosslinked species are identical to those observed when purified TET2-CS was added to extracts from cells transfected with the empty vector (X + Enz).

As a demonstration of the utility of eyC in biological settings, we examined the activity-dependent probe in cell lysates. HEK293T cells were transfected with a wild-type TET2-CS expression plasmid, a catalytically-inactive mutant (HxD), or an empty plasmid control. 24 hrs after transfection, the cells were collected, lysed and the total cell lysate incubated with S59-eyC$_{10}$ under TET reaction conditions. As predicted, lysates from cells transfected with mutant TET2 or empty vector did not result in a new Cy5-labeled species. By contrast, cells expressing TET2-CS yielded a new enzyme-DNA complex (Figure 3C), whose identity was further confirmed by both Western blotting and by comparison to the complex formed with purified TET2-CS (Figure S11). This complex was not formed with the HxD mutant and required α-KG (Figure 3C), demonstrating that eyC can indeed function as an activity-based probe even in a complex cell lysate.

In summary, our work suggests that TET enzymes have a broad substrate tolerance that includes unnatural 5-position modifications on cytosine bases. Given the rapidly expanding interest in the role of ox-mC bases in epigenetic regulation, these unnatural modified bases offer needed chemical tools for probing TET biology. Specifically, vC, which is associated with one major product rather than three sequential substrate/product pairs, offers a more facile means to detect product formation in a high throughput manner. Also, eyC, due to its crosslinking reactivity, can be applied to understand when and where TET isozymes are active in given cellular niches or to identify TET-like enzymes in other species, such as the TET-like enzyme from the amoeba *Naegleria gruberi*, which readily reacts with the probe (Figure S12). Analogous probes may also be useful for other Fe(II)/α-KG-dependent enzymes such as histone demethylases, which have been recently shown to accept N$^{\varepsilon}$-alkylated-Lys analogs.[22] Transfecting probe-containing oligonucleotides or feeding nucleoside analogs for genomic incorporation[23] could offer a powerful means to localize TET enzymes or otherwise track epigenome dynamics in future work.

ASSOCIATED CONTENT

Supporting Information. The Supporting Information is available free of charge on the ACS Publications website. Supporting Information includes methods, synthetic schemes, and additional substrate and product characterization.

AUTHOR INFORMATION

Corresponding Author

*Rahul M. Kohli: rkohli@pennmedicine.upenn.edu

Notes

No competing financial interests have been declared.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health R01-GM118501. J.E.D was supported by T32-GM07729 and the NSF; M.Y.L by F30-CA196097. We thank J. Salvino and F. Mohammed for experimental assistance.

REFERENCES

(1) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930-935.

(2) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300-1303.

(3) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, *333*, 1303-1307.

(4) Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem. Int. Ed Engl.* **2011**, *50*, 7008-7012.

(5) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929-930.

(6) Bowman, R. L.; Levine, R. L. *Cold Spring Harb Perspect. Med.* **2017**, *7*, 10.1101/cshperspect.a026518.

(7) Liu, M. Y.; DeNizio, J. E.; Kohli, R. M. *Methods Enzymol.* **2016**, *573*, 365-385.

(8) Crawford, D. J.; Liu, M. Y.; Nabel, C. S.; Cao, X. J.; Garcia, B. A.; Kohli, R. M. *J. Am. Chem. Soc.* **2016**, *138*, 730-733.

(9) Globisch, D.; Munzel, M.; Muller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Bruckl, T.; Biel, M.; Carell, T. *PLoS One* **2010**, *5*, e15367.

(10) Lu, J.; Hu, L.; Cheng, J.; Fang, D.; Wang, C.; Yu, K.; Jiang, H.; Cui, Q.; Xu, Y.; Luo, C. *Phys. Chem. Chem. Phys.* **2016**, *18*, 4728-4738.

(11) Hu, L.; Lu, J.; Cheng, J.; Rao, Q.; Li, Z.; Hou, H.; Lou, Z.; Zhang, L.; Li, W.; Gong, W.; Liu, M.; Sun, C.; Yin, X.; Li, J.; Tan, X.; Wang, P.; Wang, Y.; Fang, D.; Cui, Q.; Yang, P.; He, C.; Jiang, H.; Luo, C.; Xu, Y. *Nature* **2015**, *527*, 118-122.

(12) Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y. *Cell* **2013**, *155*, 1545-1555.

(13) Liu, M. Y.; Torabifard, H.; Crawford, D. J.; DeNizio, J. E.; Cao, X. J.; Garcia, B. A.; Cisneros, G. A.; Kohli, R. M. *Nat. Chem. Biol.* **2017**, *13*, 181-187.

(14) Pfaffeneder, T.; Spada, F.; Wagner, M.; Brandmayr, C.; Laube, S. K.; Eisen, D.; Truss, M.; Steinbacher, J.; Hackner, B.; Kotljarova, O.; Schuermann, D.; Michalakis, S.; Kosmatchev, O.; Schiesser, S.; Steigenberger, B.; Raddaoui, N.; Kashiwazaki, G.; Muller, U.; Spruijt, C. G.; Vermeulen, M.; Leonhardt, H.; Schar, P.; Muller, M.; Carell, T. *Nat. Chem. Biol.* **2014**, *10*, 574-581.

(15) Thornburg, L. D.; Lai, M. T.; Wishnok, J. S.; Stubbe, J. *Biochemistry* **1993**, *32*, 14023-14033.

(16) Thornburg, L. D.; Stubbe, J. *Biochemistry* **1993**, *32*, 14034-14042.

(17) Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G. *Angew. Chem. Int. Ed Engl.* **2013**, *52*, 10553-10558.

(18) Knapp, D. M.; Gillis, E. P.; Burke, M. D. *J. Am. Chem. Soc.* **2009**, *131*, 6961-6963.

(19) Filichev, V. V.; Astakhova, I. V.; Malakhov, A. D.; Korshun, V. A.; Pedersen, E. B. *Chemistry* **2008**, *14*, 9968-9980.

(20) Raiber, E. A.; Beraldi, D.; Ficz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, 13, R69.

(21) Wright, A. T.; Cravatt, B. F. *Chem. Biol.* **2007**, *14*, 1043-1051.

(22) Hopkinson, R. J.; Langley, G. W.; Belle, R.; Walport, L. J.; Dunne, K.; Munzel, M.; Salah, E.; Kawamura, A.; Claridge, T. D. W.; Schofield, C. J. *Chem. Commun. (Camb)* **2018**, *54*, 7975-7978.

(23) Guan, L.; van der Heijden, G. W.; Bortvin, A.; Greenberg, M. M. *Chembiochem* **2011**, *12*, 2184-2190.

Table of Contents Artwork