Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DNA Sequence and Substrate Concentration Shape TET2-Mediated Oxidation of CpG and Non-CpG Methylcytosines

Aninda Sundar Dey
University of Missouri-Kansas City
Chayan Bhattacharya
University of Missouri-Kansas City
Vidit Minda
University of Missouri-Kansas City
Gauresh Shivji
University of Missouri-Kansas City
Mridul Mukherji
mukherjim@umkc.edu
University of Missouri-Kansas City

Article

Keywords: TET2 dioxygenase, 5-methylcytosine oxidation, Flanking sequence preference, Non-CpG methylation, Epigenetic regulation, DNA demethylation

Posted Date: October 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10974122/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DNA Sequence and Substrate Concentration Shape TET2-Mediated Oxidation of CpG and Non-CpG Methylcytosines

Aninda Sundar Dey, Chayan Bhattacharya, Vidit Minda, Gauresh Shivji and Mridul Mukherji*

Division of Pharmacology & Pharmaceutical Sciences, University of Missouri-Kansas City, USA

*Correspondence: Division of Pharmacology & Pharmaceutical Sciences, University of Missouri-Kansas City, MO 64108, USA; e-mail: mukherjim@umkc.edu

Abstract

Ten-eleven translocation (TET) dioxygenases initiate active DNA demethylation by sequentially oxidizing 5-methylcytosine (5mC) to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. We used quantitative LC-MS/MS to investigate how local DNA sequence and substrate concentration influence human TET2-mediated oxidation of CpG (mCG) and non-CpG (mCA) methylcytosines. TET2 exhibited modest flanking-sequence preferences for mCG but substantially stronger sequence dependence for mCA. Systematic analysis identified CCGmCGGC and CTGmCACC as the most efficiently oxidized mCG and mCA substrates, respectively, among sequences examined. Sequence-dependent differences were greater for higher oxidation products than for 5-hydroxymethylcytosine, and relative preferences were retained in a distinct DNA backbone and with full-length TET2. Strikingly, substrate concentration altered these preferences: substrates preferred at higher concentrations became less favored as concentration decreased, shifting relative preference toward AT-rich substrates in both mCG and mCA contexts. Preferred and less-preferred substrates exhibited similar equilibrium binding affinities, indicating that DNA-binding affinity alone does not explain catalytic discrimination. Preferred mCA contexts also overlapped with CAC and CAG contexts associated with DNMT3A and DNMT3B activity. Together, these findings establish local DNA sequence and substrate concentration as interconnected determinants of TET2-mediated methylcytosine oxidation.

Keywords

TET2 dioxygenase, 5-methylcytosine oxidation, Flanking sequence preference, Non-CpG methylation, Epigenetic regulation, DNA demethylation

Introduction

DNA methylation at the 5-position of cytosine (5mC) is a major epigenetic modification in vertebrates that contributes to regulation of gene expression, embryonic development, X-chromosome inactivation, genomic imprinting, and transposon silencing [1, 2]. In mammals, a substantial proportion of CpG dinucleotides are methylated, with methylated CpG (mCG) sites distributed throughout repetitive and intergenic regions as well as gene bodies and promoter regions [3, 4]. DNA methylation patterns are established and maintained primarily by the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B [5, 6]. In contrast, active DNA demethylation is initiated by the Ten-Eleven Translocation (TET) family of Fe(II)/2-oxoglutarate-dependent dioxygenases, TET1, TET2, and TET3, which sequentially oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) [7, 8]. These higher oxidation products can contribute to restoration of unmodified cytosine, most notably through thymine DNA glycosylase (TDG)-initiated base excision repair [9, 10] (Figure S1).

Although most mammalian DNA methylation occurs at CpG dinucleotides, substantial non-CpG methylation (mCH, where H = A, C, or T) is also present in embryonic stem cells (ESCs), neurons, oocytes, and other cell types [3, 11-13]. Non-CpG methylation is developmentally regulated and has been implicated in transcriptional regulation, neurodevelopment, and disease [3, 11-13]. Among non-CpG contexts, mCA is especially prominent in neurons and constitutes a major component of the neuronal non-CpG methylome [11, 12]. Previous studies from our laboratory and others have established that TET dioxygenases can oxidize mCH substrates *in vitro*, producing 5hmCH, 5fCH, and 5caCH, although oxidation is generally less efficient than at mCG sites [14, 15]. Thus, both CpG and non-CpG methylcytosines are substrates for TET

oxidation, raising the question of whether local DNA sequence differentially influences TET activity in these two contexts.

Recent work by Adam *et al.* [16] demonstrated that the DNA sequence surrounding a modified cytosine can strongly influence TET1- and TET2-mediated oxidation. Using a high-throughput Deep Enzymology approach with randomized DNA substrate libraries analyzed by bisulfite conversion coupled to next-generation sequencing (NGS), they identified pronounced flanking-sequence effects extending from the −3 to +2 positions surrounding the modified cytosine. Their analysis revealed a strong preference for adenine and a disfavor for guanine at the −1 position, together with disfavor for cytosine at the +1 position. Flanking-sequence preferences were broadly similar between TET1 and TET2 and between 5mC and 5hmC substrates, although the magnitude of the effects was generally greater for TET2, and stronger sequence dependence was also observed at non-CpG sites [16]. Selected favorable and unfavorable substrates were subsequently examined by LC-MS, supporting the sequence-dependent differences observed in the high-throughput analysis. The comprehensive sequence-preference profiles were therefore derived primarily from the randomized-library/NGS analysis, with LC-MS providing validation of selected substrates.

Adam *et al.* further reported that the overall flanking-sequence preferences of TET enzymes differed markedly from sequence preferences previously identified for DNMT1, DNMT3A, and DNMT3B and proposed that these differences may reflect distinct mechanisms of DNA recognition [16]. This apparent divergence is particularly interesting because DNMT3 and TET enzymes catalyze opposing processes in the establishment and removal of cytosine methylation and together contribute to the dynamic regulation of DNA methylation. An important question, therefore, is whether these enzyme families exhibit entirely distinct

sequence preferences or whether their preferred substrates overlap in local sequence context. Because apparent enzyme specificity can depend on reaction conditions, substrate concentration represents a potentially important variable. The randomized DNA libraries used in the Deep Enzymology study were examined at substantially lower substrate concentrations than those typically used in biochemical assays with defined DNA substrates, raising the question of whether the same flanking-sequence preferences persist across concentration regimes.

To address these questions, we systematically examined the effects of flanking sequence on TET2-mediated oxidation using defined mCG and mCA substrates and quantitative LC-MS/MS measurement of individual 5mC oxidation products [8, 17]. We focused on TET2 because Adam *et al.* observed particularly pronounced flanking-sequence effects for this enzyme with both 5mC and 5hmC substrates [16], and because TET2 has a major role in initiating active DNA demethylation and is frequently altered in human disease [18]. We compared TET2 activity toward defined substrates under high-micromolar conditions used for systematic sequence analysis and then examined how relative substrate preference changed as substrate concentration was decreased toward the nanomolar range.

Our results show that TET2 flanking-sequence preference depends on both substrate context and substrate concentration. Sequence effects were relatively modest for mCG but substantially greater for mCA, with pronounced discrimination at the −1 and +2 positions. Under high-micromolar conditions, GC-rich sequence contexts were preferentially oxidized, and the favored mCA contexts included mCAC and mCAG, which overlap with sequence contexts associated with DNMT3A and DNMT3B activity, respectively [19, 20]. Most importantly, decreasing substrate concentration progressively shifted TET2 preference from

GC-rich toward AT-rich substrates, providing a potential explanation for differences between our findings and the low-concentration Deep Enzymology results of Adam *et al.* [16]. Sequence-dependent differences were particularly evident in formation of 5fC and 5caC, whereas preferred and less-preferred substrates exhibited similar equilibrium DNA-binding affinities. Together, these findings establish TET2 flanking-sequence specificity as a context-dependent catalytic property influenced by local sequence, substrate identity, substrate concentration, and the stage of oxidation being measured.

Materials and Methods

Chemicals and Reagents: All analytical-grade chemicals were obtained from Thermo Fisher Scientific (Waltham, MA) or MilliporeSigma (Burlington, MA). Molecular biology kits were obtained from Superior Scientifics (Lenexa, KS). Nucleoside standards were obtained from Biosynth Carbosynth (San Diego, CA), and microbiological growth media were obtained from Difco Laboratories (Franklin Lakes, NJ).

Cloning, Expression, and Purification of the TET2 Catalytic Domain and Full-Length Myc-Tagged hTET2: The C-terminal catalytic domain of human TET2 (TET2$^{1129–1936, \Delta 1481–1843}$, i.e. residues 1129–1936, del 1481–1843), corresponding to the minimal catalytically active dioxygenase domain, was cloned and expressed as previously described [21]. Recombinant TET2 was purified by cation-exchange chromatography according to previously established procedures [15, 22]. Full-length human TET2 containing an N-terminal Myc tag was expressed and purified as previously described [15, 23].

Design of DNA Substrates: Methylated CpG (mCG) and CpA (mCA) substrates were designed within a common 25-mer double-stranded DNA (dsDNA) backbone containing a

single 5-methylcytosine (5mC) residue on one strand. Nucleotides at positions −3, −2, −1, +2, and +3 relative to the modified cytosine were systematically varied to evaluate the contribution of local flanking sequence to TET2-mediated oxidation. The +1 nucleotide was fixed as G for mCG substrates and A for mCA substrates. The reference methylated strands for mCG-containing substrates were 5′-GCGCCGGTCGTAmCGGCCGCTCCCGC-3′ while for mCA containing substrates were 5′-GCGCCGGTCGTAmCAGCCGCTCCCGC-3′. For each methylated strand, an unmethylated complementary oligonucleotide was synthesized and annealed to generate a dsDNA substrate containing a single 5mC residue. This design maintained the same number of modified cytosines per duplex for direct comparison of mCG and mCA substrates. A complete list of the DNA sequences used in this study is provided in Table S1. All oligonucleotides were synthesized by Integrated DNA Technologies (IDT, Chicago, IL).

TET2-Mediated Oxidation of 5mC and Substrate-Concentration Studies: Unless otherwise indicated, TET2 oxidation reactions were performed using $TET2^{1129–1936,\ \Delta1481–1843}$ in a total volume of 100 µL containing 4 nmol of 5mC-containing dsDNA substrate (40 µM DNA, equivalent to 40 µM 5mC), 12.96 pmol of purified TET2 ($hTET2^{1129–1936,\ \Delta1481–1843}$ or full-length hTET2), 75 µM $FeSO_4$, 1 mM 2-oxoglutarate, 5 mM ascorbate, and 50 mM HEPES (pH 8.0). Each dsDNA substrate contained a single 5mC residue on the methylated strand, whereas the complementary strand remained unmethylated. Reactions were incubated at 37 °C for 1 h and quenched by addition of 5 µL of 500 mM EDTA. Unless otherwise stated, reactions were performed in triplicate. TET2 activity and substrate preference were evaluated by quantitative LC-MS/MS measurement of the individual oxidation products 5hmC, 5fC, and 5caC, as described below.

To determine whether TET2 flanking-sequence preference was influenced by substrate concentration, representative GC-rich and AT-rich 25-mer dsDNA substrates identified from the systematic sequence analysis were examined at 0.04, 0.07, 0.1, 0.2, and 0.4 μM. This concentration range was selected to resolve the substrate-concentration regime over which the relative sequence preference shifted from GC-rich to AT-rich substrates. Corresponding GC-rich and AT-rich substrates were examined in both the mCG and mCA sequence contexts. For these concentration-dependence experiments, substrate concentration was varied while the TET2 concentration, reaction buffer, cofactors, incubation temperature, and reaction time were maintained as described for the standard oxidation assay. Following incubation, reactions were quenched, processed, and analyzed by LC-MS/MS using the same procedures as the standard reactions. Relative oxidation of the GC-rich and AT-rich substrates was compared at each concentration to define the concentration-dependent shift in TET2 flanking-sequence preference.

Purification and Enzymatic Digestion of DNA: Following TET2-mediated oxidation, dsDNA was purified from the reaction mixture using the Superior Oligo Purification Kit (Superior Scientifics, Lenexa, KS) according to the manufacturer's instructions. Purified DNA was eluted in a final volume of 20 μL. The recovered DNA was enzymatically digested at 37 °C for 12 h in a 40-μL reaction containing 2 units of DNase I and 60 units of S1 nuclease to generate nucleoside monophosphates. Subsequently, 2 units of calf intestinal alkaline phosphatase (CIAP) were added, and samples were incubated for an additional 4 h at 37 °C to remove phosphate groups and generate free nucleosides. A 35-μL aliquot of each digested sample was subjected to LC-MS/MS analysis.

MS/MS Parameters for Nucleosides: LC-MS/MS analysis was performed using a SCIEX 3200 QTRAP mass spectrometer (Foster City, CA) coupled to a Shimadzu UFLC LC-20 system (Columbia, MD). Instrument control and data acquisition were performed using Analyst software version 1.6.2. Nucleoside separation was achieved using two reversed-phase C18 columns, depending on the ionization mode. For positive ion mode, a 150 × 2 mm C18 column with a 5 μm particle size and 100 Å pore size was used, whereas for negative ion mode, a 100 × 2 mm C18 column with a 5 μm particle size and 60 Å pore size was used. Both columns were obtained from Keystone Scientific (now Thermo Fisher Scientific). Detailed mass spectrometric parameters and ion transitions used for nucleoside quantification are described in our previous publications [15, 22].

Liquid Chromatographic Conditions for LC-MS/MS Studies: Chromatographic separation in negative ion mode was performed using a neutral solvent system. Solvent A consisted of 10 mM ammonium acetate (pH 6.5), while solvent B was composed of 80% acetonitrile and 20% 10 mM ammonium acetate (pH 6.5). The elution gradient was as follows: 0% B from 0-2 min, 0-20% B from 2-5 min, 20-60% B from 5-9 min, 60-0% B from 9-10 min, followed by a 5-minute post-equilibration with solvent A. The flow rate was maintained at 0.3 mL/min. For positive ion mode, a water/methanol-based solvent system was used. Solvent A was water adjusted to pH 3.5 with formic acid, and solvent B was methanol adjusted to pH 3.0 with formic acid. The gradient elution was: 0% B from 0-1 min, 0-2% B from 1-12 min, 2-30% B from 12-17 min, held at 30% B from 17-18 min, 30-0% B from 18-18.5 min, followed by a 3.5-minute post-equilibration with solvent A. The flow rate was consistently 0.3 mL/min. Standard curves, limits of detection (LOD), lower limits of quantification (LLOQ), and matrix effects for eight nucleosides were determined using our previously established LC-MS/MS method. Detailed

mass spectrometric parameters and standard curves for modified cytosines have been reported in our earlier work [15, 22].

The LC-MS/MS peak areas for the TET2 oxidation products were background-corrected by subtracting the peak area measured in the corresponding negative-control reaction from that measured in the TET2-containing reaction. The resulting background-corrected peak areas are denoted as ΔA_{5hmC}, ΔA_{5fC}, and ΔA_{5caC}. Each dsDNA substrate contained a single 5mC residue on the methylated strand, which served as the substrate for TET2 and was sequentially oxidized to 5hmC, 5fC, and 5caC during the reaction. In contrast, each substrate contained multiple unmodified cytosine residues that were not modified during the TET2 reaction and therefore provided an internal reference for normalization. Accordingly, the LC-MS/MS peak area of deoxycytidine (A_C) in each sample was used as the internal normalization signal. Because the DNA substrates differed in the number of unmodified cytosine residues, A_C was divided by the number of unmodified cytosines in the corresponding DNA substrate (#C), yielding the deoxycytidine peak area per unmodified cytosine (A_C / #C). The single 5mC residue was not included in #C because it did not contribute to the deoxycytidine peak. Each background-corrected oxidation-product peak area was then normalized to A_C / #C. This internal normalization helped account for differences in DNA recovery and sample injection and minimized the effects of run-to-run and day-to-day variation in mass-spectrometer sensitivity, thereby facilitating quantitative comparison of oxidation-product formation across different DNA substrates and independent experiments.

Total Product Formation (TPF), representing the sum of the normalized formation of the three TET2 oxidation products, was calculated as:

$$\mathrm{TPF} = \frac{\Delta A_{5hmC}}{A_C/\#C} + \frac{\Delta A_{5fC}}{A_C/\#C} + \frac{\Delta A_{5caC}}{A_C/\#C}$$

Because sequential TET2-mediated oxidation requires one oxidation event to convert 5mC to 5hmC, two cumulative oxidation events to generate 5fC, and three cumulative oxidation events to generate 5caC (Figure S2), Total Oxidation Events (TOE) was calculated by weighting each oxidation product according to the number of oxidation events required for its formation, as described previously [24]:

$$\mathrm{TOE} = 1 \times \frac{\Delta A_{5hmC}}{A_C/\#C} + 2 \times \frac{\Delta A_{5fC}}{A_C/\#C} + 3 \times \frac{\Delta A_{5caC}}{A_C/\#C}$$

where $A_C$represents the LC-MS/MS peak area of deoxycytidine and #C represents the number of cytosines in the corresponding DNA sequence. Standard deviations were calculated from three independent enzymatic assays and are reported in the corresponding figures.

Fluorescence-Polarization Analysis of TET2–DNA Binding: The equilibrium DNA-binding affinity of TET2 was examined by fluorescence polarization using 3′-FAM-labeled 25-mer dsDNA substrates. The preferred and less-preferred mCG and mCA substrates identified from the oxidation experiments were selected for binding analysis. A FAM label was incorporated at the 3′ end of the methylated strand, which was annealed to its corresponding unmethylated complementary strand using a stepwise thermal annealing protocol. Binding reactions contained 10 mM HEPES (pH 8.0), 100 mM NaCl, 5 nM FAM-labeled dsDNA, and twofold serial dilutions of purified TET2 ranging from 0.0125 to 8 μM. Samples were incubated at 4 °C for 30 min before fluorescence-polarization measurements were obtained using a SpectraMax M5 plate reader (Molecular Devices, San Jose, CA). Each condition was measured in triplicate. Fluorescence-polarization data were analyzed using SoftMax Pro software version 6.4.2, and binding curves were fitted using an equilibrium binding model to determine the apparent DNA dissociation constant (K_d). Binding results were obtained from independent experimental measurements and are reported as mean ± standard deviation.

Results

TET2 Activity at Physiologically Relevant 5mC Concentrations

To provide a physiological framework for selecting substrate concentrations for our biochemical studies, we considered published estimates of genomic 5mC abundance and mammalian nuclear volume. Approximately 4% of cytosine residues in the human genome have been reported to occur as 5mC [25]. Nuclear volume varies substantially among mammalian cell types, ranging from approximately 20–70 μm^3 in resting lymphocytes to 75–2,700 μm^3 in non-lymphoid G0 cells from tissues including liver, kidney, brain, and heart [26]. The haploid human genome contains approximately 3.2×10^9 base pairs [27], and the genomic GC content is approximately 41%, corresponding to approximately 20–21% cytosine in double-stranded DNA [28]. Based on these parameters, the bulk-equivalent concentration of genomic DNA-bound 5mC is estimated to be in the order of tens to hundreds of micromolar or higher, depending primarily on nuclear volume and global methylation level. For example, assuming a diploid genome of approximately 6×10^9 base pairs, 20–21% cytosine content, 4% of cytosines present as 5mC, and a relatively large nuclear volume of 2,700 fL yields a bulk-equivalent 5mC concentration of approximately 60 μM. Smaller nuclear volumes would yield correspondingly higher estimates. These calculations are intended to provide an order-of-magnitude framework rather than a precise concentration experienced by TET2 *in vivo*, because chromatin organization, local 5mC density, genomic accessibility, nuclear volume, and methylation levels vary among cell types and genomic regions.

We therefore used this estimated range to guide the selection of substrate concentrations for our *in vitro* experiments rather than assigning a single value as the physiological nuclear 5mC concentration. TET2-mediated oxidation was initially compared using defined 25-mer

dsDNA substrates at 40 and 200 μM. Because each DNA duplex contained a single 5mC residue, the molar concentration of DNA substrate corresponded directly to the concentration of 5mC in the reaction. Comparable relative flanking-sequence preferences were observed at 40 and 200 μM. We therefore selected 40 μM 5mC-containing dsDNA for the systematic substrate-preference studies described below, unless otherwise indicated. Thus, 40 μM represents an experimentally tractable high-micromolar substrate concentration informed by the bulk-equivalent estimates above rather than a precise intranuclear 5mC concentration.

Notably, the randomized dsDNA substrates used in the Deep Enzymology study of Adam *et al.* were present at 122 nM [16], approximately 328-fold lower than the 40 μM concentration used for our systematic sequence analysis. In addition to this substantial difference in substrate concentration, the two studies employed different analytical approaches. Adam *et al.* derived comprehensive flanking-sequence preferences primarily from randomized DNA libraries analyzed by bisulfite conversion coupled to next-generation sequencing, with selected substrates subsequently validated by LC-MS [16]. In the present study, individual defined substrates were analyzed by quantitative LC-MS/MS following complete enzymatic digestion to nucleosides, allowing 5hmC, 5fC, and 5caC to be measured individually and substrate oxidation to be evaluated using both TPF and TOE. These differences in experimental design are important when comparing the sequence-preference profiles obtained in the two studies and prompted us to examine subsequently whether TET2 flanking-sequence preference itself changes with substrate concentration.

During initial LC-MS/MS experiments at 40 μM substrates, we observed reproducible sequence-dependent differences in TET2-mediated oxidation among 25-mer dsDNA substrates containing the same centrally positioned mCG site but differing in the surrounding DNA

sequence (Table S1A). Both TPF, representing the combined normalized formation of 5hmC, 5fC, and 5caC, and TOE, which additionally weights these products according to the number of successive oxidation steps required for their formation, were higher for substrate A1 than for substrate A2 (Figure S3). This initial observation prompted us to systematically examine the contribution of individual flanking nucleotides by varying positions from −3 to +3 surrounding the centrally positioned 5mC within an otherwise common 25-mer dsDNA backbone. To permit direct comparison of CpG and non-CpG substrates, each duplex contained a single 5mC residue on only one strand, thereby maintaining an equivalent 5mC concentration across all substrates.

Flanking-Sequence Effects on TET2-Mediated mCG Oxidation

To determine how local DNA sequence influences TET2-mediated oxidation of 5mC in the CpG context, we first systematically varied the nucleotides immediately flanking the mCG site at the −1 and +2 positions. All four canonical nucleotides were examined at each position (N−1mCGN+2), generating 16 (4^2) 25-mer dsDNA substrates that differed only at these two positions (Table S1B). TET2-mediated oxidation was quantified by LC-MS/MS, and relative substrate preference was evaluated using Total Product Formation (TPF) and Total Oxidation Events (TOE).

TET2 exhibited relatively modest differences in oxidation among the 16 N−1mCGN+2 substrates (Figures S4 and S5), indicating that individual nucleotide substitutions immediately flanking mCG exert only subtle effects under these conditions. The effects of nucleotide identity at the −1 and +2 positions were largely independent, with no strong dependence of the preference at one position on the nucleotide present at the other. Nevertheless, consistent positional trends were apparent across the sequence set. At the −1 position, G was generally

favored, whereas C was comparatively less favored; at the +2 position, G was favored, whereas A was comparatively less favored. Differences among the intermediate nucleotides were smaller. Among all 16 combinations, GmCGG exhibited the highest activity, whereas CmCGA exhibited the lowest activity. GmCGG and CmCGA were therefore carried forward as the preferred and less-preferred starting sequence contexts, respectively, for analysis of the more distal flanking positions.

Because simultaneous variation of the −3, −2, −1, +2, and +3 positions would require analysis of 4^5 (1,024) sequence combinations, the −2, −3, and +3 positions were examined sequentially. At each step, all four nucleotides were substituted at the position under investigation while retaining the preferred and less-preferred sequence contexts identified at the preceding step. At the −2 position, nucleotide substitutions were examined within the N−2GmCGG and N−2CmCGA contexts (Table S1C; Figures S6A–S6B and S7A–S7B). Within the preferred GmCGG background, C at −2 produced the highest activity, yielding CGmCGG. Conversely, within the less-preferred CmCGA background, A at −2 produced the lowest activity, yielding ACmCGA. C and A were therefore carried forward at the −2 position in the preferred and less-preferred sequence contexts, respectively.

The −3 position was subsequently varied within the N−3CGmCGG and N−3ACmCGA contexts (Table S1D; Figures S6C–S6D and S7C–S7D). Within the preferred background, C at −3 produced the highest activity, yielding CCGmCGG. Within the less-preferred background, A at −3 produced the lowest activity, yielding AACmCGA. C and A were therefore retained at −3 along the preferred and less-preferred sequence trajectories, respectively. Finally, the +3 position was varied within the CCGmCGGN+3 and AACmCGAN+3 contexts (Table S1E; Figures S6E–S6F and S7E–S7F). Within the preferred

background, C at +3 generated the most efficiently oxidized sequence, CCGmCGGC. Conversely, A at +3 generated the least efficiently oxidized sequence within the less-preferred background, AACmCGAA.

Collectively, these results demonstrate that individual flanking-nucleotide substitutions exert relatively modest effects on TET2-mediated mCG oxidation, but that their cumulative combination progressively increases the distinction between preferred and less-preferred sequence contexts. The systematic −1/+2 analysis identified G as modestly favored at both positions, with C comparatively less favored at −1 and A less favored at +2. Subsequent stepwise analysis of the more distal positions selected C at −2, −3, and +3 along the preferred sequence trajectory and A at each of these positions along the less-preferred trajectory. This progressive selection ultimately identified the GC-rich CCGmCGGC motif as the most efficiently oxidized substrate (TOE = 8.13E−02 ± 2.20E−03) and the AT-rich AACmCGAA motif as the least efficiently oxidized substrate (TOE = 6.04E−02 ± 9.65E−04) among the sequences examined. Examination of the individual oxidation products further revealed that 5hmC formation was comparable between these two substrates, whereas formation of the higher oxidation products 5fC and 5caC was approximately 30% greater for the preferred GC-rich substrate (Figure 1).

Flanking-Sequence Effects Are More Pronounced for mCA Than mCG

Because mCA represents a major form of non-CpG methylation in several mammalian cell types, particularly neurons and embryonic stem cells [11-13, 29, 30], we next examined whether local sequence similarly influences TET2-mediated oxidation in the mCA context. Using the same 25-mer dsDNA backbone employed for the mCG analysis, we systematically varied the nucleotides immediately flanking the mCA site at the −1 and +2 positions. All four canonical

nucleotides were examined at each position (N−1mCAN+2), generating 16 (4^2) possible nucleotide combinations (Table S1G). TET2-mediated oxidation was quantified by LC-MS/MS, and relative substrate preference was evaluated using TPF and TOE.

In contrast to the relatively modest sequence dependence observed for mCG substrates, TET2-mediated oxidation of mCA showed substantially greater variation among the 16 N−1mCAN+2 sequence contexts (Figures S8 and S9). The effects of nucleotide identity at the −1 and +2 positions were largely independent, with the preference at either position generally retained across the different nucleotide contexts at the other position. At the −1 position, G and C were favored over T and A, with the overall preference following G ≥ C ≫ T ≥ A. At the +2 position, C and G were favored over T and A, with an overall preference of C ≥ G ≫ T ≥ A. Thus, in contrast to the relatively subtle positional effects observed for mCG, the immediate flanking positions of mCA showed a clear preference for G/C-containing over A/T-containing sequence contexts. Among all 16 combinations, GmCAC exhibited the highest activity, whereas AmCAA exhibited the lowest activity. GmCAC and AmCAA were therefore carried forward as the preferred and less-preferred starting sequence contexts, respectively, for analysis of the more distal flanking positions.

As in the mCG analysis, the −2, −3, and +3 positions were subsequently examined sequentially. At each step, all four nucleotides were substituted at the position under investigation while retaining the preferred and less-preferred sequence contexts identified at the preceding step. At the −2 position, nucleotide substitutions were examined within the N−2GmCAC and N−2AmCAA contexts (Table S1H; Figures S10A–S10B and S11A–S11B). Within the preferred GmCAC background, T at −2 produced the highest activity, yielding TGmCAC. Conversely, within the less-preferred AmCAA background, A at −2 produced the

lowest activity, yielding AAmCAA. T and A were therefore carried forward at −2 along the preferred and less-preferred sequence trajectories, respectively.

The −3 position was subsequently varied within the N−3TGmCAC and N−3AAmCAA contexts (Table S1I; Figures S10C–10D and S11C–S11D). Within the preferred background, C at −3 produced the highest activity, yielding CTGmCAC. Within the less-preferred background, T at −3 produced the lowest activity, yielding TAAmCAA. C and T were therefore retained at −3 along the preferred and less-preferred sequence trajectories, respectively. Finally, the +3 position was varied within the CTGmCACN+3 and TAAmCAAN+3 contexts (Table S1J; Figures S10E–S10F and S11E–S11F). Within the preferred background, C at +3 generated the most efficiently oxidized sequence, CTGmCACC. Conversely, A at +3 generated the least efficiently oxidized sequence within the less-preferred background, TAAmCAAA.

Collectively, these results demonstrate that TET2-mediated oxidation is more strongly influenced by local flanking sequence in the mCA context than in the mCG context. The systematic −1/+2 analysis revealed the strongest positional preferences, with G/C-containing contexts favored over A/T-containing contexts at both positions. Subsequent stepwise analysis of the more distal positions selected T at −2, C at −3, and C at +3 along the preferred sequence trajectory, whereas A at −2, T at −3, and A at +3 were selected along the less-preferred trajectory. This progressive selection ultimately identified CTGmCACC as the most efficiently oxidized mCA substrate (TOE = 4.04E−02 ± 9.74E−04) and TAAmCAAA as the least efficiently oxidized mCA substrate (TOE = 2.44E−02 ± 9.82E−04) among the sequences examined. Examination of the individual oxidation products further revealed that 5hmC formation was comparable between these two substrates, whereas the preferred CTGmCACC substrate produced approximately 2.2-fold more 5fC and 1.7-fold more 5caC than

TAAmCAAA (Figure 2). Thus, the influence of flanking sequence was particularly pronounced in the formation of the higher oxidation products.

TET2 Flanking-Sequence Preferences Are Retained Across DNA Backbones

To determine whether the relative flanking-sequence preferences identified in the 25-mer substrates were dependent on the surrounding DNA backbone, we transferred the preferred GC-rich CCGmCGGC and less-preferred AT-rich AACmCGAA motifs to a distinct 12-mer dsDNA backbone previously used in structural studies of TET2 [31, 32]. The 12-mer substrates reproduced the relative oxidation pattern observed with the corresponding 25-mer substrates (Table S1F). Formation of 5hmC was comparable between the two sequences, whereas the preferred GC-rich substrate generated approximately 30% more 5fC and 5caC than the less-preferred AT-rich substrate (Figure 3). Thus, the relative preference for the GC-rich mCG sequence was maintained despite alteration of the surrounding DNA backbone.

We next asked whether this sequence preference was specific to the minimal TET2 catalytic domain used for the systematic sequence analysis. The same 12-mer mCG substrates were therefore examined using full-length, Myc-tagged TET2. Full-length TET2 produced slightly more 5hmC from the preferred GC-rich substrate and approximately 30% more 5fC and 5caC than from the less-preferred AT-rich substrate (Figure 3). Thus, the relative mCG flanking-sequence preference was reproduced with full-length TET2, indicating that the observed sequence dependence is not restricted to the catalytic-domain construct.

A similar pattern was observed for mCA substrates. The preferred GC-rich CTGmCACC and less-preferred AT-rich TAAmCAAA motifs were transferred to the 12-mer dsDNA backbone (Table S1K). Consistent with the corresponding 25-mer substrates, 5hmC

formation was comparable between the two 12-mer substrates, whereas the preferred CTGmCACC substrate produced approximately 2.2-fold more 5fC and 1.7-fold more 5caC than TAAmCAAA. The relative preference was also retained with full-length, Myc-tagged TET2. The preferred GC-rich mCA substrate produced slightly more 5hmC and approximately 50% more 5fC and 5caC than the less-preferred AT-rich substrate.

Collectively, these experiments demonstrate that the relative flanking-sequence preferences identified for both mCG and mCA substrates were retained following transfer to a distinct DNA backbone and were not restricted to the minimal TET2 catalytic-domain construct used for the initial sequence analysis. Across both substrate classes, differences between preferred and less-preferred sequences were consistently more pronounced for the higher oxidation products, 5fC and 5caC, than for 5hmC. The magnitude of the flanking-sequence effect remained substantially greater for mCA than for mCG, although overall TET2-mediated oxidation was greater for mCG than for mCA substrates, consistent with previous studies [14, 15].

Decreasing Substrate Concentration Shifts TET2 Preference from GC-Rich to AT-Rich Sequences

Our systematic studies of flanking-sequence effects were performed at 40 μM DNA substrate and identified preferential oxidation of GC-rich sequences in both mCG and mCA contexts. At this concentration, the preferred GC-rich 25-mer mCG substrate exhibited substantially greater TOE than the less-preferred AT-rich substrate. Similarly, the preferred GC-rich 25-mer mCA substrate showed substantially greater TOE than its less-preferred AT-rich counterpart. These relative preferences were also retained upon transfer of the motifs to a distinct 12-mer DNA backbone and with full-length TET2, as described above. Thus, the 40 μM experiments

established a consistent preference for the GC-rich substrates in both mCG and mCA contexts. To determine whether this relative sequence preference was maintained as substrate concentration decreased, we compared the representative GC-rich and AT-rich 25-mer dsDNA substrates at 0.04, 0.07, 0.1, 0.2, and 0.4 μM. As mentioned in the Materials and Methods section this concentration range was selected to resolve the substrate-concentration regime over which the relative sequence preference shifted from GC-rich to AT-rich substrates. Both mCG- and mCA-containing substrates were examined using the TET2 catalytic domain, and oxidation was quantified as TOE, which weights formation of 5hmC, 5fC, and 5caC according to the one, two, and three cumulative oxidation events required for their formation, respectively.

A pronounced concentration-dependent reversal in relative sequence preference was observed for the mCG substrates (Figure 4). At 0.4 μM, TET2 preferentially oxidized the GC-rich substrate, consistent with the GC-rich preference observed at 40 μM, and the GC-rich substrate remained favored at 0.2 μM. In contrast, at 0.1 μM the relative preference reversed, with the AT-rich substrate producing greater TOE than the GC-rich substrate. The AT-rich preference was maintained at 0.07 μM and became particularly pronounced at 0.04 μM. Thus, although the GC-rich mCG substrate was preferred at 40 μM and remained favored at 0.4 and 0.2 μM, decreasing substrate concentration to 0.1 μM and below reversed the relative preference in favor of the AT-rich substrate.

A remarkably similar concentration-dependent reversal was observed for the mCA substrates. At 0.4 μM, TET2 preferentially oxidized the GC-rich mCA substrate, consistent with the GC-rich preference observed at 40 μM, and the GC-rich substrate remained favored at 0.2 μM. At 0.1 μM, however, the relative preference reversed, with greater TOE observed for the AT-rich substrate. This AT-rich preference persisted at 0.07 μM and became

particularly pronounced at 0.04 μM. Thus, as observed for mCG, the GC-rich mCA substrate was preferred at 40 μM and at 0.4 and 0.2 μM, whereas the AT-rich substrate became preferred at 0.1 μM and below. Together, these results demonstrate that the apparent flanking-sequence preference of TET2 is strongly dependent on substrate concentration. At 40 μM, the GC-rich mCG and mCA substrates were preferentially oxidized, and the same relative preference was observed at 0.4 and 0.2 μM. At 0.1, 0.07, and 0.04 μM, however, the relative preference was reversed, with greater oxidation of the corresponding AT-rich substrates. Strikingly, this reversal occurred over the same concentration interval for both substrate classes, with the transition in relative preference occurring between 0.1 and 0.2 μM. These findings establish that TET2 flanking-sequence preference is not invariant but changes with substrate availability, such that sequence preferences determined under one substrate-concentration regime may not predict relative catalytic activity under another.

Preferred and Less-Preferred TET2 Substrates Exhibit Similar Equilibrium Binding Affinities

To determine whether the flanking-sequence preferences observed in the oxidation assays could be explained by differences in equilibrium DNA-binding affinity, we performed fluorescence-polarization assays using the preferred and less-preferred mCG and mCA substrates. Binding of the purified TET2 catalytic domain to FAM-labeled 25-mer dsDNA substrates (Table S1L) was measured, and the resulting binding curves were used to determine apparent equilibrium K_d (Figure 5). For the mCG substrates, the preferred and less-preferred sequences yielded K_d values of 1.11 and 1.06 μM, respectively. Similarly, the preferred and less-preferred mCA substrates yielded K_d values of 1.04 and 1.01 μM, respectively. Thus, despite clear differences in TET2-mediated oxidation between the preferred and less-preferred substrates, their equilibrium binding affinities were very similar. These results indicate that the

observed differences in catalytic activity cannot readily be explained by differences in equilibrium DNA-binding affinity alone.

Discussion

The present study demonstrates that TET2-mediated oxidation of methylcytosine is shaped by both local DNA sequence and substrate concentration. Using defined mCG and mCA substrates and quantitative LC-MS/MS analysis of individual oxidation products, we found substantially stronger flanking-sequence effects for mCA than for mCG, with GC-rich contexts preferentially oxidized under the high-micromolar conditions used for systematic sequence analysis. Most importantly, decreasing substrate concentration progressively shifted relative TET2 preference from GC-rich toward AT-rich substrates. Together, these findings demonstrate that apparent TET2 flanking-sequence specificity is not an invariant property of the enzyme but varies with substrate context and availability.

The concentration dependence provides a potential explanation for the apparent difference between our results and those of Adam *et al.* [16]. Using randomized DNA libraries at a total substrate concentration of 122 nM, Adam *et al.* identified pronounced TET1/TET2 flanking-sequence preferences, including preference for A and disfavor for G at the −1 position. In contrast, our systematic analysis at 40 μM identified preferential oxidation of GC-rich contexts. Importantly, however, these observations were not mutually exclusive: the GC-rich preference was maintained at 40–200 μM, whereas progressive reduction of substrate concentration shifted TET2 toward the AT-rich preference observed by Adam *et al.* Thus, the two studies may capture different concentration-dependent manifestations of TET2 substrate preference rather than fundamentally incompatible sequence specificities. The analytical approaches are also complementary. Whereas the comprehensive profiles of Adam *et al.* were

derived primarily from randomized-library/bisulfite-NGS measurements with selected substrates validated by LC-MS, our LC-MS/MS approach directly resolved 5hmC, 5fC, and 5caC in individually defined substrates. Together, these findings emphasize that apparent enzyme sequence specificity can depend not only on the DNA sequence examined but also on substrate concentration and on how different stages of sequential oxidation are captured by the analytical approach.

The mechanistic basis for the concentration-dependent shift remains unresolved. Attempts to characterize the transition kinetically were limited at low substrate concentrations because TET2 performs successive oxidation reactions and 5hmC, 5fC, and 5caC are formed and consumed at different stages of the reaction. At the lowest concentrations, all oxidation products could not be simultaneously quantified with sufficient reliability by LC-MS/MS, preventing robust determination of complete product profiles. The present data therefore establish concentration-dependent differential utilization of GC-rich and AT-rich substrates but do not define the kinetic mechanism underlying this transition. Resolving this behavior will require approaches capable of quantitatively following individual oxidation states at low substrate concentrations and, ideally, over time.

A second major finding is that flanking-sequence dependence was considerably stronger for mCA than for mCG. Individual substitutions surrounding mCG produced relatively modest effects, although their cumulative combination distinguished the GC-rich CCGmCGGC substrate from the AT-rich AACmCGAA substrate. In contrast, mCA oxidation showed pronounced discrimination at the immediate −1 and +2 positions, with G/C favored over A/T, ultimately identifying CTGmCACC and TAAmCAAA as the preferred and less-preferred sequences among those examined. Previous biochemical studies established that

mammalian TET enzymes can oxidize 5mC in non-CpG contexts [14, 15, 24]; our results extend these observations by demonstrating that TET2-mediated mCA oxidation is particularly sensitive to the surrounding DNA sequence. This may be biologically relevant because non-CpG methylation is abundant and developmentally regulated in pluripotent cells and neurons [3, 11-13], with mCA constituting a major component of neuronal non-CpG methylation [11, 12]. The mCAC context is particularly noteworthy because it represents a prominent target of MeCP2 recognition in the brain and contributes to MeCP2-dependent transcriptional regulation [12, 30, 33, 34]. Thus, the strong sequence dependence of TET2 toward mCA raises the possibility that local sequence contributes to differential turnover of biologically important non-CpG methylation contexts. Whether these *in vitro* preferences translate into preferential oxidation at genomic mCA sites *in vivo* remains to be established.

The mCA results also reveal an intriguing overlap with the sequence preferences of the DNMT3 methyltransferases. DNMT3A and DNMT3B catalyze *de novo* methylation in both CpG and non-CpG contexts [35-37] and exhibit substantial flanking-sequence dependence [19, 20, 38]. Particularly relevant to the present findings, Lee *et al.* reported that DNMT3A- and DNMT3B-dependent non-CpG methylation is enriched in CAC and CAG contexts, respectively [19], while Mao *et al.* demonstrated that DNMT3A/B sequence preferences contribute to characteristic genome-wide methylation signatures [20]. DNMT3B has also been shown to exhibit pronounced flanking-sequence effects in both CpG and non-CpG substrates [38]. Our finding that TET2 preferentially oxidizes mCA when C or G occupies the +2 position therefore corresponds directly to mCAC and mCAG contexts, respectively. Particularly notable is the preferred extended TET2 substrate CTGmCACC, which contains the mCAC context associated with DNMT3A-dependent non-CpG methylation. These observations identify an overlap between local sequence contexts associated with DNMT3-mediated methylation and those

preferentially oxidized by TET2, despite the opposing catalytic functions of the two enzyme families. They do not establish coordinated activity or evolutionary convergence *in vivo*, but raise the possibility that local DNA sequence contributes to both the establishment of non-CpG methylation and its subsequent TET-mediated oxidation. Importantly, because TET2 preference shifts toward AT-rich substrates as substrate concentration decreases, the extent of this overlap with DNMT3 sequence preferences may itself be condition dependent.

The concentration dependence observed here also provides a framework for relating these biochemical preferences to genomic 5hmC patterns. Adam *et al.* found that the AT-rich flanking preferences identified in their low-concentration randomized-library experiments were reflected in genome-wide 5hmC distributions and in methylation changes following TET depletion [16], supporting the biological relevance of intrinsic TET sequence preferences. In particular, genomic sites with higher 5hmC/5mC ratios were enriched for sequence features favored by TET1 and TET2 in their biochemical assays [16]. Our findings do not conflict with these genomic observations; rather, they show that TET2 sequence preference is not restricted to a single class of flanking contexts. Under high-micromolar conditions, TET2 preferentially oxidized the GC-rich substrates examined here, whereas decreasing substrate concentration progressively shifted relative preference toward AT-rich substrates. Thus, TET2 can efficiently oxidize methylated cytosines embedded in both GC-rich and AT-rich sequence environments, with their relative utilization depending on substrate availability. In the genomic setting, where effective substrate availability is further shaped by chromatin accessibility, local methylation density, TET recruitment, and interacting proteins, such context dependence may broaden the range of methylated sequence environments accessible to TET2 oxidation.

Direct quantification of the individual oxidation products provides an additional mechanistic clue. TET2 catalyzes an iterative pathway in which 5mC is sequentially converted to 5hmC, 5fC, and 5caC [8]. Structural and biochemical studies further indicate that 5mC, 5hmC, and 5fC are not equivalent TET2 substrates and are accommodated and oxidized with distinct catalytic efficiencies [32]. In our experiments, differences between preferred and less-preferred mCG and mCA substrates were considerably greater for 5fC and 5caC than for 5hmC. The relatively small sequence-dependent difference in 5hmC formation, together with the substantially larger differences in 5fC and 5caC, therefore raises the possibility that flanking sequence exerts a greater influence on one or more steps following the initial oxidation of 5mC. This distinction may have functional consequences because 5fC and 5caC can be excised by thymine DNA glycosylase (TDG), initiating base-excision repair and restoration of unmodified cytosine [39, 40]. Sequence-dependent differences in progression toward these higher oxidation states could therefore influence how efficiently a methylated site proceeds toward active DNA demethylation. However, our endpoint measurements cannot distinguish among effects on individual catalytic rates, productive substrate positioning, enzyme dissociation and rebinding, or product release and should not be interpreted as direct evidence for altered TET2 processivity.

Our fluorescence-polarization measurements further constrain the possible mechanisms underlying sequence-dependent TET2 activity. Preferred and less-preferred mCG and mCA substrates exhibited very similar equilibrium K_d values despite substantially larger differences in their oxidation profiles. Thus, the observed catalytic differences cannot readily be explained by differences in equilibrium DNA-binding affinity alone. Importantly, K_d and K_m are not interchangeable: K_d describes equilibrium binding affinity, whereas K_m is a composite kinetic parameter determined by multiple rate constants and therefore need not equal the dissociation

constant. Accordingly, the fluorescence-polarization measurements do not permit relative K_m values to be inferred for the GC-rich and AT-rich substrates. Nevertheless, the concentration-dependent reversal in substrate preference raises the possibility that these substrates differ in their kinetic parameters. Under a simple Michaelis–Menten framework, preferential oxidation of the AT-rich substrate at low substrate concentration would be consistent with a greater catalytic efficiency (k_{cat}/K_m), whereas preferential oxidation of the GC-rich substrate at higher concentration could reflect a greater maximal catalytic rate. Such behavior could arise from differences in k_{cat}, K_m or both, and therefore does not by itself establish that the AT-rich substrate has a lower K_m. The combination of similar equilibrium binding affinities and divergent, concentration-dependent oxidation profiles instead suggests that sequence discrimination may involve steps following initial DNA association. Structural studies have shown that TET2 recognition involves extrusion of the target 5mC from the DNA duplex and its positioning within the catalytic cavity [31, 32]. Local sequence could therefore influence DNA geometry or deformability, the energetic requirements for base flipping, productive positioning of the modified cytosine within the active site, individual oxidation steps, or product release and rebinding, without substantially altering overall equilibrium binding affinity. Because TET2 catalyzes sequential oxidation of 5mC to 5hmC, 5fC, and 5caC, however, this simple kinetic framework should be regarded as a mechanistic hypothesis rather than a quantitative description of the present endpoint measurements. Distinguishing among these possibilities will require kinetic and structural approaches capable of resolving the sequential oxidation reactions across the substrate-concentration range in which the sequence-preference shift occurs.

In summary, our findings identify local DNA sequence and substrate concentration as interconnected determinants of apparent TET2 substrate specificity. The concentration-

dependent shift from GC-rich toward AT-rich substrates provides a potential reconciliation between our findings and the low-concentration Deep Enzymology results of Adam *et al.* [16]. At higher substrate concentrations, the overlap between preferred TET2 mCA contexts and the CAC/CAG environments associated with DNMT3A and DNMT3B [19, 20] reveals an intriguing intersection between the sequence preferences of enzymes responsible for opposing cytosine modifications. Together with the similar equilibrium binding affinities of preferred and less-preferred substrates, these findings support a model in which TET2 sequence discrimination reflects the interplay among local DNA sequence, substrate availability, and productive catalytic events following initial binding. Thus, TET2 sequence specificity should be viewed not as a single invariant property of the enzyme, but as a context-dependent feature of its catalytic interaction with methylated DNA.

Acknowledgements

The authors thank Dr. William Gutheil for his critical reading of the manuscript and helpful comments. ChatGPT (OpenAI) was used to assist with language editing and refinement of the manuscript.

Funding

This research was supported by the U.S. Department of Defense through an Idea Award (W81XWH-13-1-0174), the Aplastic Anemia & MDS International Foundation, and a University of Missouri Research Board (UMRB) grant awarded to M.M.

Author contributions

A.S.D. performed the experiments, analyzed the data, and wrote the original draft of the manuscript. C.B. performed the experiments and analyzed the data. V.M. and G.S. performed

the experiments. M.M. conceived the study, supervised the research, and reviewed and finalized the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).
2. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* **33 Suppl**, 245-254 (2003).
3. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).
4. Laurent, L. et al. Dynamic changes in the human methylome during differentiation. *Genome Res* **20**, 320-331 (2010).
5. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).
6. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).
7. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935 (2009).
8. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303 (2011).
9. He, Y.F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307 (2011).
10. Iwan, K. et al. 5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo. *Nature chemical biology* **14**, 72-78 (2018).
11. Lister, R. et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
12. Guo, J.U. et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature neuroscience* **17**, 215-222 (2014).
13. Ziller, M.J. et al. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* **7**, e1002389 (2011).
14. DeNizio, J.E. et al. TET-TDG Active DNA Demethylation at CpG and Non-CpG Sites. *J Mol Biol* **433**, 166877 (2021).
15. Dey, A.S., Bhattacharya, C., Guan, Y., Jha, B.K. & Mukherji, M. Demethylation of Non-CpG Sites in DNA Is Initiated by TET2 5-Methylcytosine Dioxygenase. *DNA* **1**, 26-36 (2021).
16. Adam, S. et al. Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic 5hmC patterns. *Commun Biol* **5**, 92 (2022).
17. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423-434 (2011).
18. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843 (2010).
19. Lee, J.H., Park, S.J. & Nakai, K. Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. *Sci Rep* **7**, 11295 (2017).
20. Mao, S.Q., Cuesta, S.M., Tannahill, D. & Balasubramanian, S. Genome-wide DNA Methylation Signatures Are Determined by DNMT3A/B Sequence Preferences. *Biochemistry* **59**, 2541-2550 (2020).

21. Bhattacharya, C., Dey, A.S., Ayon, N.J., Gutheil, W.G. & Mukherji, M. Efficient Purification and LC-MS/MS-based Assay Development for Ten-Eleven Translocation-2 5-Methylcytosine Dioxygenase. *J Vis Exp* (2018).
22. Dey, A.S., Ayon, N.J., Bhattacharya, C., Gutheil, W.G. & Mukherji, M. Positive/negative ion-switching-based LC–MS/MS method for quantification of cytosine derivatives produced by the TET-family 5-methylcytosine dioxygenases. *Biology Methods and Protocols* **5** (2020).
23. Guan, Y. et al. Context dependent effects of ascorbic acid treatment in TET2 mutant myeloid neoplasia. *Commun Biol* **3**, 493 (2020).
24. DeNizio, J.E., Liu, M.Y., Leddin, E.M., Cisneros, G.A. & Kohli, R.M. Selectivity and Promiscuity in TET-Mediated Oxidation of 5-Methylcytosine in DNA and RNA. *Biochemistry* **58**, 411-421 (2019).
25. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **8**, 24 (2015).
26. Suciu, D. Morphometric study of the interphase nucleus in some radiosensitive and radioresistant mammalian cells. *J Theor Biol* **113**, 599-609 (1985).
27. Makalowski, W. The human genome structure and organization. *Acta Biochim Pol* **48**, 587-598 (2001).
28. Piovesan, A. et al. On the length, weight and GC content of the human genome. *BMC Res Notes* **12**, 106 (2019).
29. Stroud, H. et al. Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell* **171**, 1151-1164 e1116 (2017).
30. Gabel, H.W. et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89-93 (2015).
31. Hu, L. et al. Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545-1555 (2013).
32. Hu, L. et al. Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118-122 (2015).
33. Lagger, S. et al. MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genet* **13**, e1006793 (2017).
34. Kinde, B., Gabel, H.W., Gilbert, C.S., Griffith, E.C. & Greenberg, M.E. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 6800-6806 (2015).
35. Gowher, H. & Jeltsch, A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J Mol Biol* **309**, 1201-1208 (2001).
36. Arand, J. et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet* **8**, e1002750 (2012).
37. Liao, J. et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature genetics* **47**, 469-478 (2015).
38. Dukatz, M. et al. Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. *Nucleic Acids Res* **48**, 11495-11509 (2020).

39. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *The Journal of biological chemistry* **286**, 35334-35338 (2011).
40. Shen, L. et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692-706 (2013).

Figures

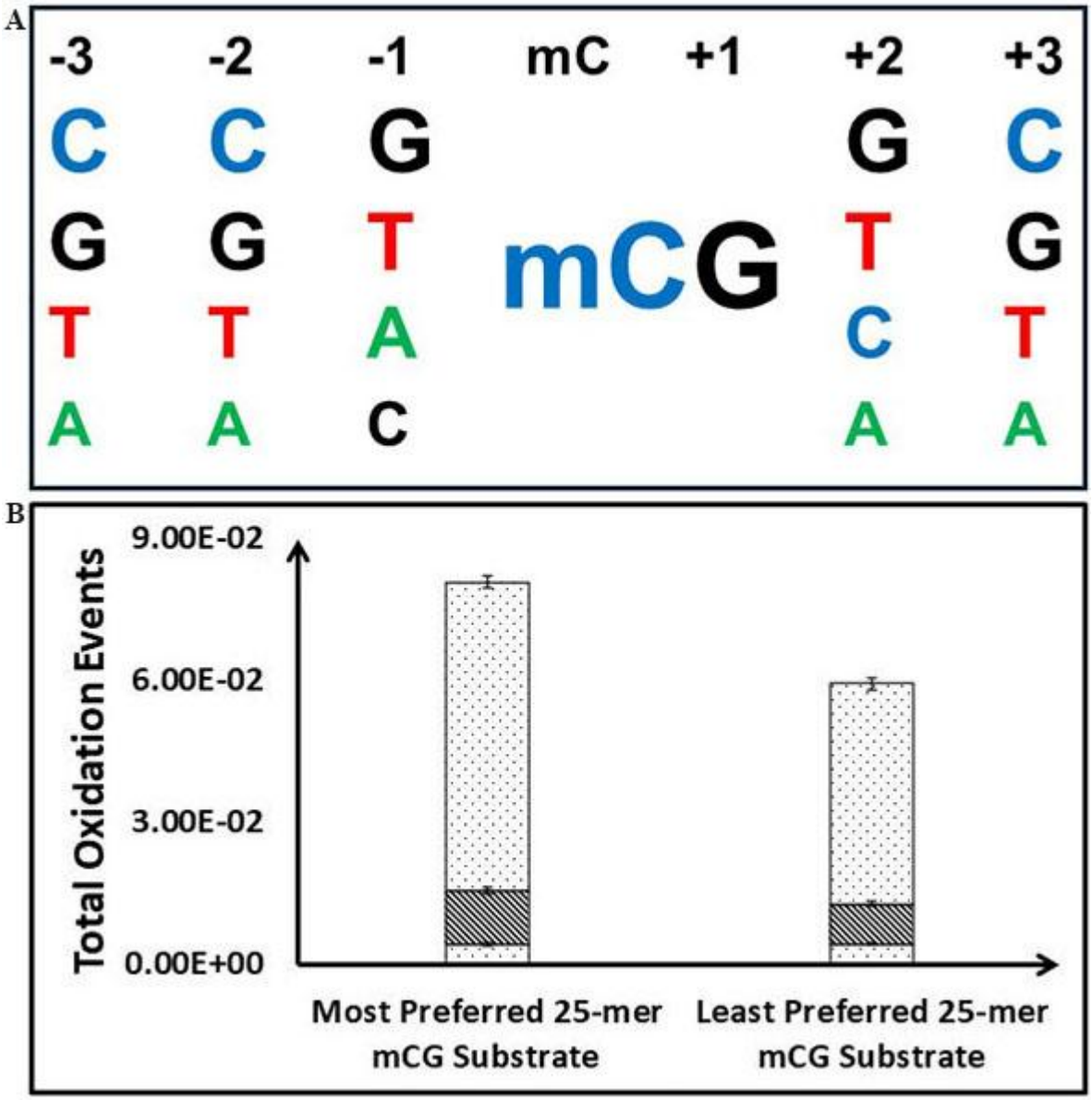


Figure 1. Flanking-sequence preferences of TET2-mediated mCG oxidation. (A) Summary of nucleotide preferences at positions −3 to +3 surrounding mCG; larger letters indicate greater relative preference. (B) Comparison of TOE for the most efficiently oxidized CCGmCGGC and least efficiently oxidized AACmCGAA 25-mer dsDNA substrates. Data represent mean ± SD from three independent experiments.

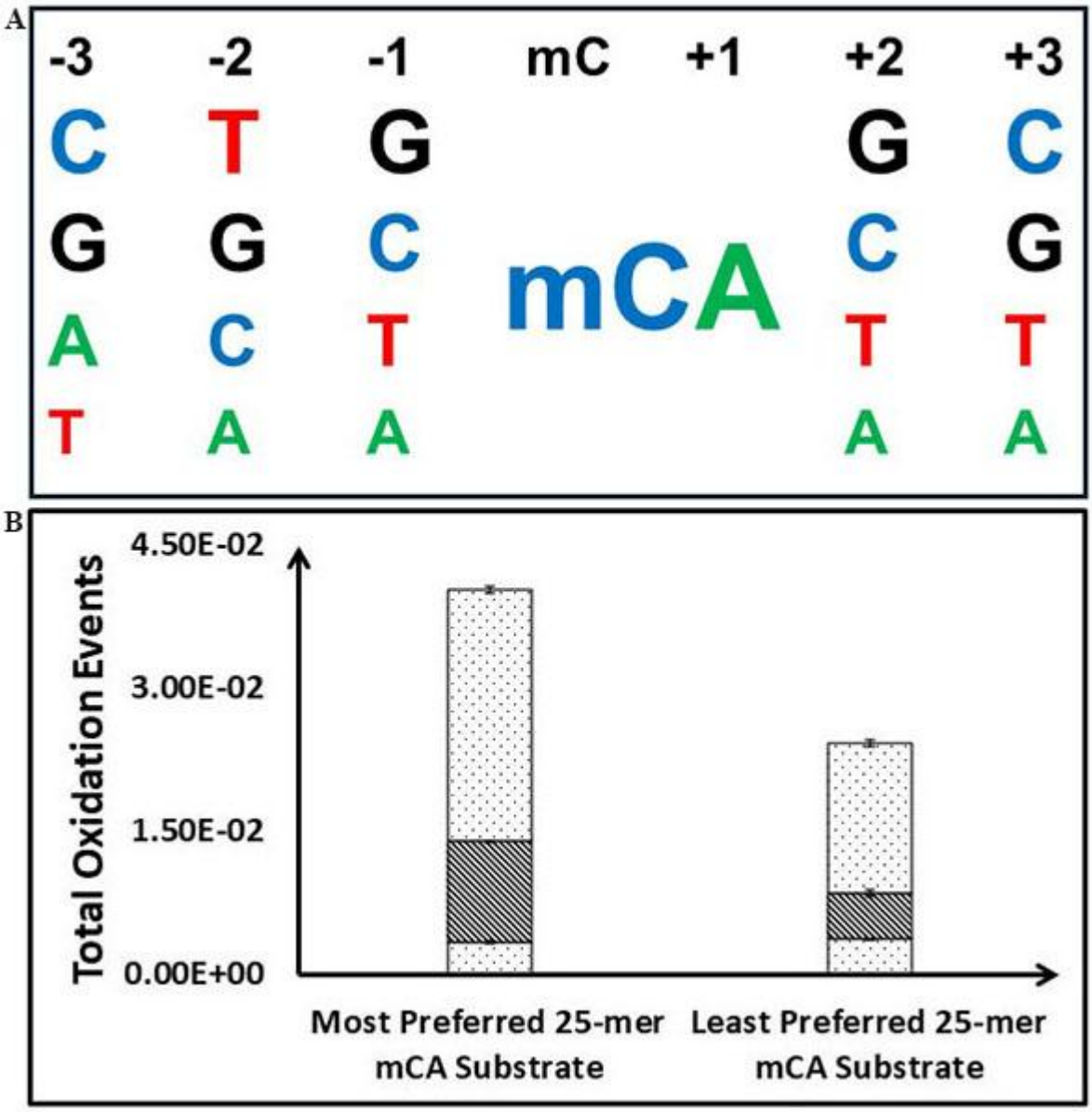


Figure 2. Flanking-sequence preferences of TET2-mediated mCA oxidation. (A) Summary of nucleotide preferences at positions −3 to +3 surrounding mCA; larger letters indicate greater relative preference. (B) Comparison of TOE for the most efficiently oxidized CTGmCACC and least efficiently oxidized TAAmCAAA 25-mer dsDNA substrates. Data represent mean ± SD from three independent experiments.

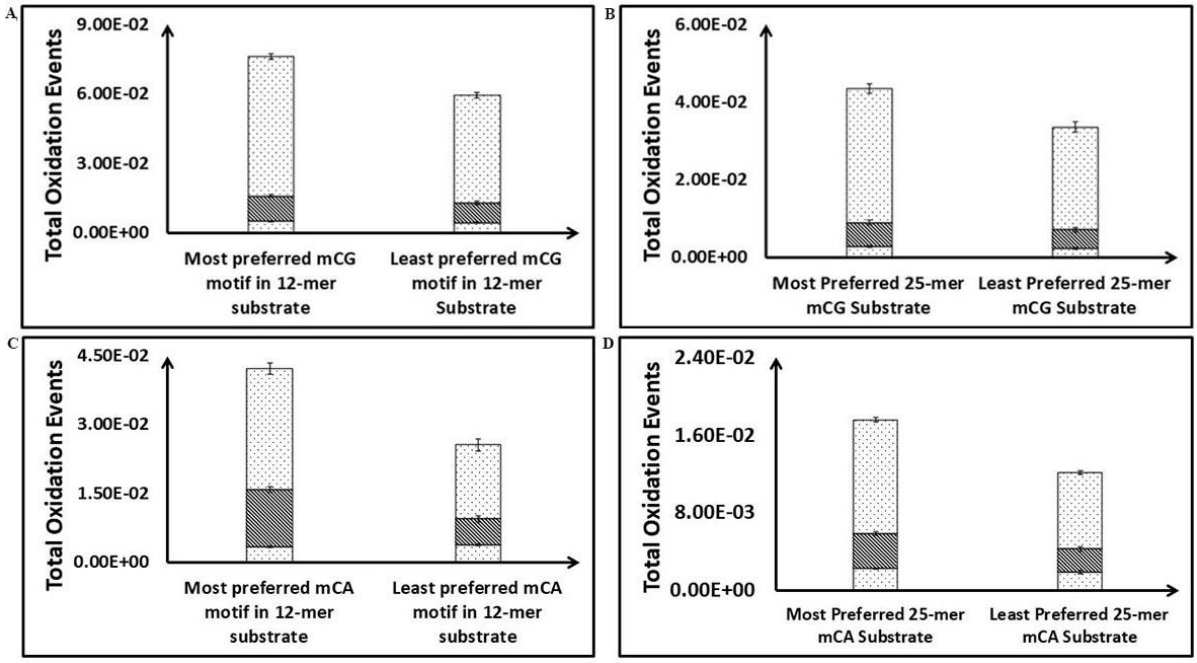


Figure 3. TET2 flanking-sequence preferences are retained across DNA backbones and with full-length TET2. TOE comparison of preferred and less-preferred mCG substrates in the 12-mer backbone (A) and with full-length TET2 (B), and corresponding mCA substrates in the 12-mer backbone (C) and with full-length TET2 (D). Data represent mean ± SD from three independent experiments.

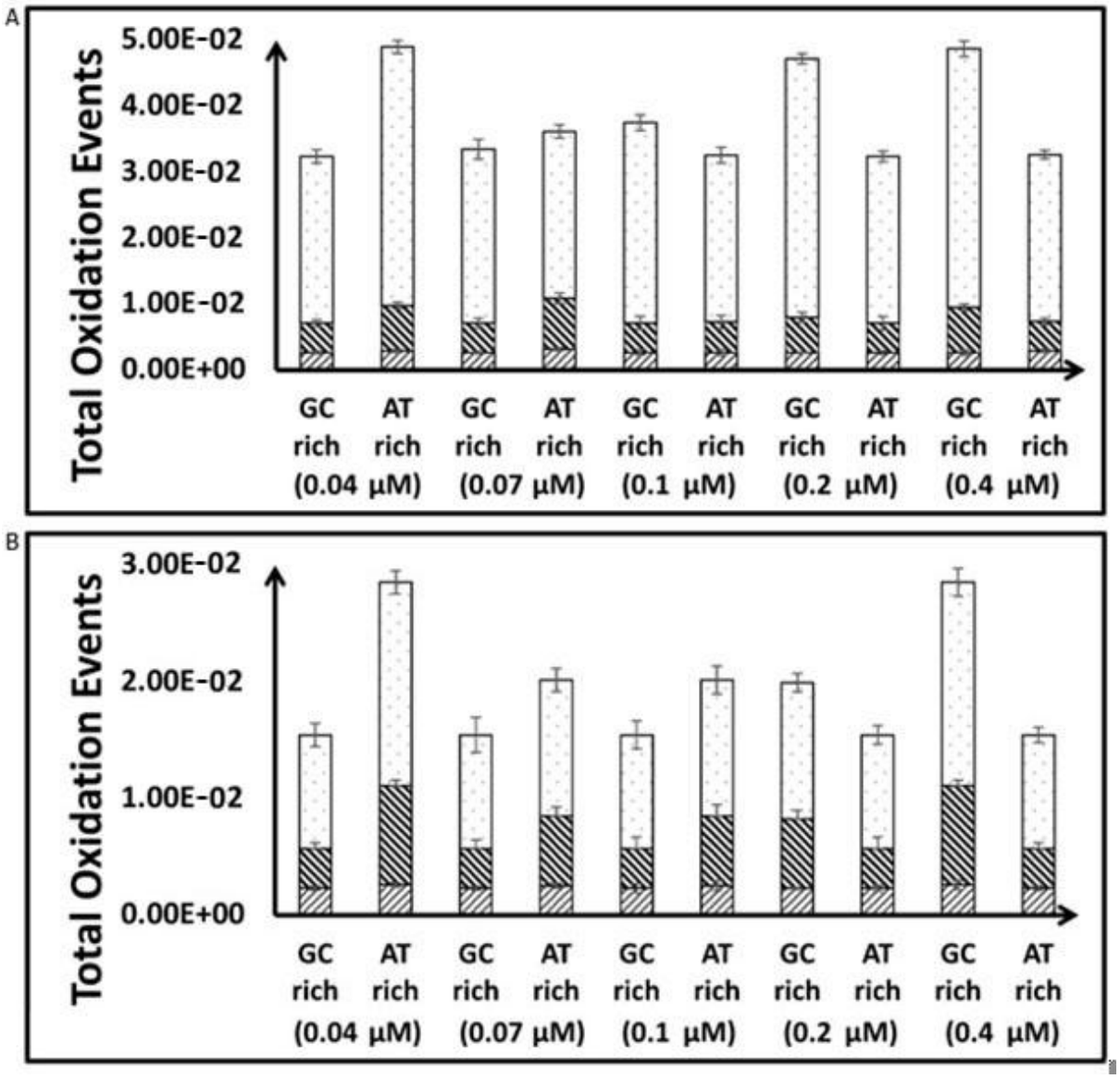


Figure 4. Substrate concentration alters TET2 flanking-sequence preference. TOE comparison of GC-rich and AT-rich 25-mer dsDNA substrates at 0.04–0.4 µM for mCG (A) and mCA (B). TET2 substrate preference shifts from AT-rich to GC-rich at ≥0.07–≤0.1 µM for mCG and at ≥0.1–≤0.2 µM for mCA, respectively. Data represent mean ± SD from three independent experiments.

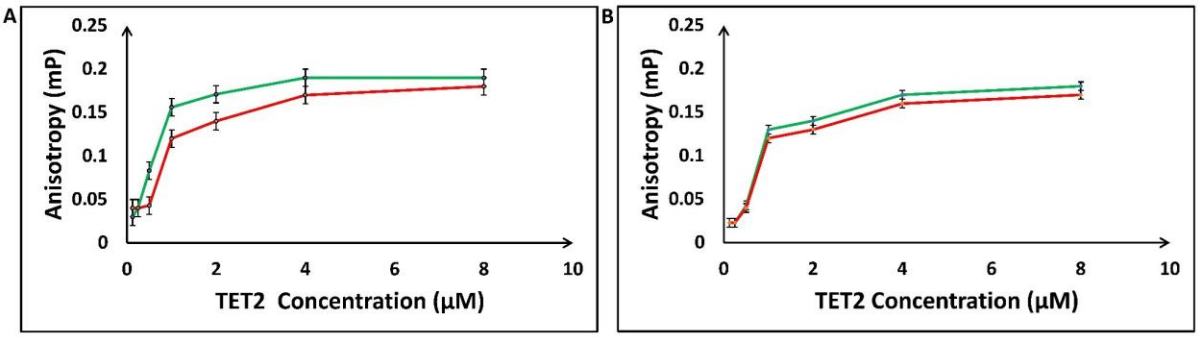


Figure 5. Preferred and less-preferred TET2 substrates exhibit similar equilibrium binding affinities. Fluorescence-polarization binding curves for preferred and less-preferred 25-mer dsDNA substrates containing mCG (A) or mCA (B). Apparent K_d values were 1.11 and 1.06 μM for the preferred and less-preferred mCG substrates and 1.04 and 1.01 μM for the corresponding mCA substrates, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DeyetalSI.pdf