Published in final edited form as:

Biochemistry. 2019 February 05; 58(5): 411–421. doi:10.1021/acs.biochem.8b00912.

Selectivity and Promiscuity in TET-mediated oxidation of 5-methylcytosine in DNA and RNA

Jamie E. DeNizio[#1,2,3], **Monica Yun Liu**[#1,2,3], **Emmett M. Leddin**[4], **G. Andrés Cisneros**[4], and **Rahul M. Kohli**[2,3,†]

[1]Graduate Group in Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States.

[2]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States.

[3]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States.

[4]Department of Chemistry, University of North Texas, Denton, Texas 76201, United States.

[#] These authors contributed equally to this work.

Abstract

Enzymes of the ten-eleven translocation (TET) family add diversity to the repertoire of nucleobase modifications by catalyzing the oxidation of 5-methylcytosine (5mC). TET enzymes were initially found to oxidize 5-methyl-2'-deoxycytidine in genomic DNA, yielding products that contribute to epigenetic regulation in mammalian cells, but have since been found to also oxidize 5-methylcytidine in RNA. Considering the different configurations of single- and double-stranded DNA and RNA that co-exist in a cell, defining the scope of TET's preferred activity and the mechanisms of substrate selectivity is critical to better understand the enzymes' biological functions. To this end, we have systematically examined the activity of human TET2 on DNA, RNA, and hybrid substrates *in vitro*. We found that, while ssDNA and ssRNA are well tolerated, TET2 is most proficient at dsDNA oxidation and discriminates strongly against dsRNA. Chimeric and hybrid substrates containing mixed DNA and RNA character helped reveal two main features by which the enzyme discriminates between substrates. First, the identity of the target nucleotide alone is the strongest reactivity determinant, with a preference for 5-methyldeoxycytidine, while both DNA or RNA are relatively tolerated on the rest of the target strand. Second, while a complementary strand is not required for activity, DNA is the preferred partner, and

[†] Correspondence should be addressed to R.M.K. (rkohli@pennmedicine.upenn.edu).

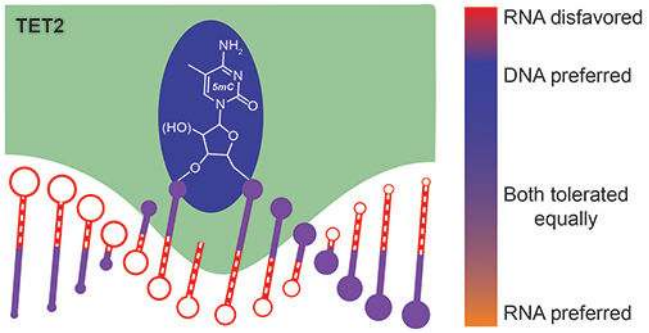
Supporting Information. The Supporting Information is available free of charge on the ACS Publications website. Figures and captions provide confirmation of substrate integrity, nucleoside standard curves for LC-MS/MS analysis, additional *in vitro* TET reaction data, dynamic, energy, conformational, and H-bonding differences from MD simulations of 5mdC- and 5mrC-containing models (PDF).

Accession Codes
Human TET1 Q8NFU7
Human TET2 Q6N021
Human TET3 O43151

The authors declare no competing financial interest.

complementary RNA diminishes reactivity. Our biochemical analysis, complemented by molecular dynamics simulations, provides support for an active site optimally configured for dsDNA reactivity but permissive for various nucleic acid configurations, suggesting a broad range of plausible roles for TET-mediated 5mC oxidation in cells.

Graphical Abstract



INTRODUCTION

Long thought to be the only significant DNA modification in mammalian genomes, 5-methyl-2'-deoxycytidine (5mdC) is now known to be a substrate for further oxidation. Oxidative modifications to 5mdC, largely in the context of cytosine guanine dinucleotides (CpGs), are catalyzed by enzymes of the ten-eleven translocation (TET) family, which are Fe(II) and α-ketoglutarate (α-KG) dependent dioxygenases. Step-wise oxidation of 5mdC by TET enzymes can yield 5-hydroxymethyl-2'-deoxycytidine (5hmdC), 5-formyl-2'-deoxycytidine (5fdC), and 5-carboxyl-2'-deoxycytidine (5cadC)[1–3]. These DNA modifications can function as potential intermediates in the long-sought pathway for erasure of 5mdC, referred to as demethylation, and can also serve independent functions regulating gene expression[4]. For example, 5hmdC is thought to activate loci silenced by methylation[5], and intragenic 5hmdC in DNA may also impact RNA splicing[6]. The distinctive localization patterns of 5hmdC, 5fdC and 5cadC in sequencing studies further fuels speculation about additional independent roles in epigenetic regulation[7].

Just as 2'-deoxycytidine (dC) modifications are widely prevalent in DNA, modifications to the 5-position in cytidine (rC) in RNA are also found across all domains of life[8,9]. 5-methylcytidine (5mrC) and its oxidized analogs (5hmrC, 5frC, and 5carC) are detectable in many types of RNA, including mRNA, tRNA, rRNA, and ncRNA[8–10]. Levels of 5mrC in total RNA vary across tissues and organisms but can approach levels of 5mdC in DNA; for example, in mouse brain, 5mrC comprises approximately 1% of total cytosines in RNA[11,12], while 5mdC is 2–5% of cytosines in DNA[13,14]. By contrast, hydroxymethylation is approximately 2–3 orders of magnitude lower in RNA than in DNA[13,14]. Unlike for 5mdC in DNA, the function of 5mrC in RNA is not well established, although potential roles in nuclear export of mRNA have been demonstrated[15]. In Drosophila, 5hmrC was found to be enriched in the coding regions and may favor translation of mRNA transcripts, as more ribosomes were bound to mRNA containing 5hmrC[16]. With regards to demethylation,

pathways for 5mrC "removal" have not been found; however, the loss of 5hmrC has been shown to increase 5mrC levels, suggesting the possibility of functional parallels[17].

Although DNA and RNA have unique methyltransferases involved in methylation[18], TET enzymes appear to be involved in oxidation of both DNA and RNA. In mammalian systems, all three TET enzymes (TET1, UniProtKB Q8NFU7; TET2, UniProtKB Q6N021; TET3, UniProtKB O43151) exhibited activity on 5mrC *in vitro* and in transfected cells[11]. TET1 and TET2 were also detected in an unbiased screen for RNA-binding nuclear proteins in mouse embryonic stem cells[19]. Further support for TET oxidation in both DNA and RNA comes from studies demonstrating that activity on both substrates can be inhibited by the same means: Mutations in isocitrate dehydrogenase (IDH) which result in accumulation of the competitive inhibitor 2-hydroxyglutarate, previously known to impede DNA oxidation, have also been shown to reduce RNA oxidation[20]. Phylogenetic analysis also highlights a broader role for TET enzymes in acting on different nucleic acids[21,22]. Drosophila again provide a compelling example in this regard; they lack the canonical 5mdC substrate in their DNA, but have a TET enzyme homologue that appears to act on RNA as a bona fide substrate[16,23].

The observation of both DNA and RNA oxidation, as well as the diversity of potential functions of 5mdC and 5mrC oxidation products, makes it a priority to understand how the nature of the nucleic acid impacts TET activity. No systematic comparison of TET activity on various configurations of DNA versus RNA has yet been performed, but previous biochemical experiments and crystal structures provide a strong starting point for speculating about the mechanistic features involved in DNA versus RNA oxidation[24,25]. In the structure of a truncated active human TET2 variant (TET2-CS) bound to a 12-mer double-stranded (ds) DNA substrate, one of the most striking features is that 5mdC on the target strand is flipped out of the duplex and into the active site of TET, raising questions about the conformational flexibility required for dC- versus rC-based substrates. Outside of the target nucleotide, key contacts on the target DNA strand are largely confined to the phosphate backbone, suggesting the possibility of promiscuity towards nucleic acid identity. Regarding the impact and necessity of a complement strand, the structure and biochemistry present conflicting data. Mutations of TET2-CS that could disrupt contacts with the complementary strand decrease activity on dsDNA, but the effect on single-stranded (ss) DNA has not been tested in parallel[24]. One study showed that dsDNA is more reactive than ssDNA[11], while another suggested that ssDNA is more reactive[26]. The relative activities of TET enzymes on various DNA and RNA configurations are even less well-established. Early evidence suggested that the catalytic domain of mouse TET1 greatly prefers dsDNA to a sequence-matched ssRNA substrate. TET-mediated oxidation to 5frC and 5carC has also been demonstrated and could potentially occur in dsRNA, though oxidation to these forms appeared to occur at very low efficiency[11,27].

To better understand the breadth of TET enzyme capabilities, here we have performed a comprehensive examination of the impact of nucleic acid identity on TET activity, by systematic variation of the target base, target strand, and complementary strand of the substrates. Our results show a dominant role for the target nucleotide in dictating the efficiency of oxidation. Outside of this preference, TET2 has a surprising tolerance for most

substrate configurations, except for dsRNA, arguing that TET enzyme promiscuity makes these enzymes suited to diverse biological roles. By placing our biochemical analysis in the context of targeted molecular dynamics simulations, we further identify structural features that support this promiscuity as well as potential mechanisms that can explain the observed spectrum of activity on DNA versus RNA substrates.

MATERIALS AND METHODS

Oligonucleotide preparation.

DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). RNA oligonucleotides and DNA/RNA chimeras were purchased from Trilink. All oligonucleotides were HPLC purified and the masses confirmed. The oligonucleotides were resuspended in 10 mM Tris-Cl pH 7.5 and quantified by UV absorbance spectroscopy, using the extinction coefficient at 260 nm for each oligonucleotide. All substrates were diluted to 10 μM (of reactive, methylated substrate, whether single- or double-stranded) in 10 mM Tris-HCl pH 7.5 and 50 mM NaCl. Duplexed oligos were annealed by heating to 95 °C for 30 s and cooling at 1 °C increments, for 15 s per step, to 4 °C. In order to confirm that the substrates were not degraded by the annealing process, 5 pmol of oligonucleotides were analyzed on a pre-warmed Tris-Borate-EDTA/7 M Urea/20% polyacrylamide gel before or after annealing, and visualized by SYBR Gold (Invitrogen) staining (Figure S1).

Purification of TET enzymes from Sf9 insect cells.

The crystalized human TET2 variant (TET2-CS) includes residues 1129–1936 Δ1481–1843[24]. The full catalytic domains of TET2 (TET2-CD, residues 1129–2002), TET1 (TET1-CD, 1418–2136), and TET3 (TET3-CD, 689–1660) were also purified. These constructs, with an N-terminal FLAG tag, were subcloned into a pFastBac1 vector for Sf9 insect cell expression, as described previously[28]. Expression was carried out for 24 hr. Cells from a 1 L culture were collected and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 0.2% (v/v) NP-40) containing cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche, 1 tablet/10 mL). Cells were lysed by three passes through a microfluidizer at 15,000 psi. The lysate was cleared by centrifugation at 20,000*g* for 30 min. The supernatant was then passed two to three times over a 500 μL or 1 mL packed column of anti-FLAG M2 affinity resin (Sigma), prepared according to the manufacturer's instructions. The column was washed three times with 10 mL of wash buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 15% (v/v) glycerol). To elute the bound protein, one column volume of elution buffer (wash buffer containing 100 μg/mL of 3× FLAG peptide (Sigma)) was incubated on the column for 10 min, and serial elutions were collected until no more protein was detected by Bio-Rad Protein Assay and SDS-PAGE. The three most concentrated fractions were pooled, aliquoted, and stored at −80 °C. Protein concentrations were measured using a Qubit 4 Fluorometer (Invitrogen) prior to use in *in vitro* activity assays.

TET reactions on DNA and RNA substrates.

Substrates were diluted to the designated concentrations in reaction buffer (50 mM HEPES, pH 6.5, 100 mM NaCl, 1 mM α-KG, 1 mM DTT, and 2 mM sodium ascorbate). Fresh ammonium iron(II) sulfate (Sigma) was added to 75 μM prior to initiation, and purified

enzyme was added last to start the reaction (t = 0), bringing the total volume to 25 μL. This mixture was incubated at 37 °C for 30 min. Reactions were quenched by addition of pre-mixed quenching solution (25 μL H_2O, 100 μL Oligo Binding Buffer (Zymo), and 400 μL ethanol). Oligonucleotide products were purified using the Zymo Oligo Clean & Concentrator kit, eluted in 10 μL of Millipore water, and analyzed by LC-MS/MS.

Quantification of reaction products by LC-MS/MS.

The purified TET reaction products were degraded to component DNA and/or RNA nucleosides using 1 μL of Nucleoside Digestion Mix (New England Biolabs) in 10 μL total volume for 4 hours at 37 °C. The mixture was diluted 10-fold into 0.1% formic acid, and approximately 2 pmol was loaded onto an Agilent 1200 Series HPLC equipped with a 5 μm, 2.1 × 250 mm Supelcosil LC-18-S analytical column (Sigma) equilibrated to 45 °C in Buffer A (0.1% formic acid). The nucleosides were separated in a gradient of 0–10% Buffer B (0.1% formic acid, 30% (v/v) acetonitrile) over 8 min at a flow rate of 0.5 mL/min. Tandem MS/MS was performed by positive ion mode ESI on an Agilent 6460 triple-quadrupole mass spectrometer, with gas temperature of 225 °C, gas flow of 12 L/min, nebulizer at 35 psi, sheath gas temperature of 300 °C, sheath gas flow of 11 L/min, capillary voltage of 3,500 V, fragmentor voltage of 70 V, and delta EMV of +1,000 V. Collision energies were 10 V for all bases, aside from C and 5hmC for which the collision energies were 25V. MRM mass transitions were {dC 228.1→112.1 m/z; dT 243.1→127.1; 5mdC 242.1→126.1; 5hmdC 258.1→124.1; 5fdC 256.1→140.0; 5cadC 272.1→156.0; rC 244.1→112.1; 5mrC 258.1→126.1; 5hmrC 274.1→124.1; 5frC 272.1→140.0; 5carC 288.1→156.0}. Standard curves were generated from standard deoxyribonucleosides (Berry & Associates) and ribonucleotide triphosphates (Trilink), first digested to nucleosides using the Nucleoside Digestion mix under the same conditions as above prior to serial dilution, from 1,250 fmol to 0.614 fmol (Figure S2); sample peak areas were fit to the standard curve to determine amounts of each modified cytosine in the sample. Results are expressed as the percentage of each modified cytosine out of all modified DNA or RNA bases detected. Alternatively, to quantify total oxidation events, we considered that TET enzymes catalyze one oxidation reaction to generate 5hmC, two reactions to generate 5fC, and three reactions to generate 5caC. Therefore, we expressed total oxidation events as 1×(% 5hmC) + 2×(% 5fC) + 3×(% 5caC)[29].

Molecular dynamics simulations.

Molecular dynamics (MD) simulations were performed using the ff99SB force field with the pmemd.cuda program from the AMBER16 software suite[30,31]. The 5mrC system was created from a crystal structure of the human TET2-DNA complex (PDBID: 4NM6)[24]. Similar to our previous simulations[32], the current system includes magnesium as a surrogate for iron in the active site, coordinated His residues protonated on ND1, and a linker inserted using Modeller[33,34].

The previously published deoxyribonucleotide parameters for cytosine derivatives, α-ketoglutarate, Mg(II), and Zn were used[32], and ribonucleotide parameters for 5mrC were generated using PyRED[35–38]. The system with 5mrC was generated using the LEaP module[39] and neutralized to a net charge of zero with potassium ions and then solvated in a

truncated octahedron of TIP3P water extending at least 12 Å from the complex surface[40]. SHAKE was used for all bonds involving hydrogen, and long-range Coulomb interactions were treated using the smooth particle mesh Ewald method with a 9 Å cutoff[41]. Simulations were carried out with the Berendsen thermostat in the NVT ensemble with a 2 fs time-step for 100 ns[42].

All dsDNA 5mdC reference calculations were performed with Langevin dynamics. Two tests were performed in order to validate that meaningful comparisons could be made between the 5mrC system and the previous work with the 5mdC system. These tests included a single 100 ns run of the 5mrC system with Langevin dynamics (minimization in the NVT ensemble and production in the NPT ensemble with the Berendsen barostat; collisional frequency of 1.0 ps^{-1}) and a 1 fs time-step, and a single 100 ns run of the 5mdC system with the Berendsen thermostat in the NVT ensemble and a 2 fs time-step used for the dynamics reported here.

Upon confirming compatibility, each simulation was carried out three times and data were averaged across replicates. The cpptraj program in AMBER was used to analyze trajectories for correlations, hydrogen bonding, atomic fluctuations, root mean square deviations, residue distances, and nucleotide angular parameters[43–45]. Energy decomposition analysis (EDA) was performed using an in-house FORTRAN90 program to investigate the Coulomb and van der Waals interactions between individual residues throughout the simulation[46–48]. VMD and the ProDy interface were used in principal component (PCA) and normal mode (NMA) analyses[49,50]. The first 100 PCA modes were calculated from each trajectory using 5,000 snapshots. UCSF Chimera was used to generate images[51].

RESULTS

Design of Substrate Series.

To probe TET's capacity for reacting with various configurations of deoxyribo- versus ribonucleic acids, we designed oligonucleotides that permitted systematic analysis of substrate requirements for TET activity. Two different substrate series were generated for these experiments, each containing a 16-bp all-DNA oligonucleotide, as well as a matched all-RNA oligonucleotide, which contain uridine in place of thymidine. These substrates contain a single 5-methylcytosine (5mC) in a CpG sequence context: 5mdC in the all-DNA substrate (designated as D(M)) or 5mrC in the all-RNA substrate (designated as r(m)) (Figure 1A). Given the potential relevance of the sugar identity of the target base, which is extruded into the active site during oxidation, we also designed chimeric substrates that could distinguish between the impact of the identity of the target 5mC and that of other nucleotides on the target strand. In the DNA substrate, 5mdC was replaced with 5mrC (yielding D(m)), and vice versa for the RNA substrate (yielding r(M)). Thus, Series 1 and 2 each contain four total oligonucleotides: D(M), r(m), D(m), and r(M).

Notably, we designed Series 1 and 2 to be complementary to one another. For Series 2, we generated two additional oligonucleotides that contained unmodified cytosine, rather than 5mC, in the all-DNA or all-RNA context (D(C) or r(c)). Comparing Series 1 substrates in the absence or presence of either D(C) or r(C) complementary strands could therefore permit

us to examine single-stranded versus double-stranded nucleic acids. Furthermore, complexing Series 1 and 2 substrates that both contain reactive 5mC sites could permit an analysis of activity on different strands in a duplex.

TET activity on 5mdC versus 5mrC.

TET's activity was first analyzed on the simplest iteration of the series, the all single-stranded substrates. *In vitro* reactions were performed using recombinant TET2-CS[24], and the reacted substrates were purified, digested to nucleosides and analyzed by LC-MS/MS (Figure 1A, Figure S2). We first examined the product distribution at a 2:1 ratio of substrate to enzyme. By displaying the data as the relative percentages of each modified cytosine, the amount of remaining 5mC reflects the substrate consumption, while the flux through the oxidative pathway is given by the relative distribution of the three forms of TET-oxidized products. Under these conditions with Series 1, the DNA substrate ($D(M)_1$) is entirely reacted, while nearly all the RNA substrate ($r(m)_1$) is converted (7% 5mrC remaining), indicating that the RNA is a relatively proficient substrate. The DNA and RNA substrate reactions do differ, however, in the *extent* to which they are oxidized: Roughly 80% of the DNA substrate is converted to the higher oxidized modifications (5fdC and 5cadC), while the RNA substrate is primarily converted to the first oxidative product, 5hmrC (65%) (Figure 1B). With the Series 2 substrates, the general trends all hold, with the exception that TET less proficiently oxidizes $r(m)_2$, with approximately 40% of the 5mrC substrate unreacted, relative to $D(M)_2$ where 5mdC is completely consumed (Figure 1D).

Given that driving conditions could obscure the comparison between relative reactivities of the all-DNA and all-RNA substrates, we also examined reactivity at an 8:1 ratio of substrate:enzyme. Under these conditions, there remains 2.6 and 5.2-fold as much RNA as DNA substrate for Series 1 and Series 2, respectively (Figure S3A,B). To see how far we could push the oxidation of RNA, we also analyzed the reactivity at a 1:1 ratio of substrate to enzyme. For both Series 1 and 2, while roughly 60% of 5mdC is oxidized to 5cadC, there is very little 5carC generated from 5mrC (12% for Series 1 and none detectable for Series 2) despite near complete consumption of 5mrC (Figure S3A,B). Interestingly, while only 12% of 5mrC remains unreacted for $r(m)_2$, it is almost exclusively converted to 5hmrC (81%), suggesting a greater barrier to oxidation from 5hmC to 5fC for RNA than for DNA.

To facilitate a more rigorous comparison of the overall reactivity of substrates, we also analyzed the data for total oxidation events for each substrate. Considering that 5fC and 5caC both represent multiple catalytic events, total oxidation can be defined as the sum of the turnover events, namely: (fraction of 5hmC) + 2×(fraction of 5fC) + 3×(fraction of 5caC)[29]. Analyzing the data this way, the preference for ssDNA over ssRNA is evident and maintained across different conditions. Specifically, we observed an average of 1.7 and 3.3-fold more turnovers, for Series 1 and Series 2, respectively, with D(M) substrates relative to r(m) substrates across the different concentration ratios (Figure 1C,E).

We focused our initial analysis on TET2-CS given that the variant is well-expressed and more active than full catalytic domain variants of TET2 (TET2-CD). To test the generality of the preference for ssDNA over ssRNA across the mammalian TET enzymes, we also examined the reactivity of $D(M)_1$ and $r(m)_1$ with TET1-CD, TET2-CD and TET3-CD.

Consistent with prior studies, all three TET enzymes are capable of oxidizing both the ssDNA and ssRNA substrates, and a consistent preference for oxidation of 5mdC in $D(M)_1$ over 5mrC in $r(m)_1$ was observed. The proportion of reacted 5mdC, *i.e.* sum of the percentages of 5hmC, 5fC, and 5caC, in $D(M)_1$ is 6.6-fold that of 5mrC in $r(m)_1$ for TET1-CD, 1.5-fold for TET2-CD, and 5-fold for TET3-CD (Figure S3C). Thus, activity can reliably be detected with both DNA and RNA, across two substrate series, a range of substrate:enzyme ratios, and different TET variants, and ssDNA appears to be consistently preferred over ssRNA for human TET enzymes.

To further explore the mechanistic basis for discrimination between ssDNA and ssRNA, we next focused on the nature of the target nucleotide relative to the flanking residues of the substrates. After reacting TET2-CS with $D(m)_1$, which differs from $D(M)_1$ only by the addition of a single 2'-hydroxyl at the target base, we quantified 5mrC and the 5hmrC, 5frC, and 5carC oxidation products. With this substrate, the reactivity decreases significantly, with a turnover that averaged 2.4-fold less across the substrate:enzyme ratios examined relative to $D(M)_1$, and is more comparable to the all-RNA substrate, $r(m)_1$ (Figure 1C). The reciprocal change in the $r(M)_1$ substrate, which differs from $r(m)_1$ only by the removal of a single 2'-hydroxyl at the target base, shows a rescue of activity, with $r(M)_1$ showing largely comparable activity to the all-DNA substrate, $D(M)_1$. Like the Series 1 substrates, the Series 2 substrates display a similar pattern, where the distribution of oxidized products from $r(M)_2$ more closely resembles that of $D(M)_2$, while that of $D(m)_2$ is more similar to $r(m)_2$ (Figure 1D). Thus, analysis of these chimeric substrate oligonucleotides indicates that the difference in activity between ssDNA and ssRNA is largely attributable to the target base identity, rather than the DNA or RNA character of the surrounding bases.

Impact of complement strand on reactivity.

We next sought to evaluate the impact of a complement strand and its nucleic acid identity on TET activity. The Series 1 substrates, including the all-DNA strand, all-RNA strand, and the two chimeric oligonucleotides, were each annealed to an unmethylated complement, either the all-DNA strand, $D(C)_2$, or the all-RNA strand, $r(c)_2$. Given the wide range of substrate reactivity observed across these permutations, the substrates were compared at a 1:1 substrate:enzyme ratio using TET2-CS. Under these conditions, the dsDNA ($D(M)_1$:$D(C)_2$) substrate reacts to near completion, with almost complete conversion of 5mdC to 5cadC (Figure S4A). Similarly, single-stranded $D(M)_1$ shows complete consumption of 5mdC, and total oxidation events are 92% of those observed with dsDNA (Figure 2A), as oxidation to 5cadC is incomplete (Figure S4A). This small decrease in activity on ssDNA compared to dsDNA is also be observed at a lower substrate:enzyme ratio where the total oxidation events of ssDNA are also 92% of dsDNA (Figure S4B), even though the oxidized products of both the ss- and dsDNA substrates are predominantly 5hmdC and 5fdC (Figure S4B). Notably, with the full catalytic domains, TET2-CD shows a similar small reactivity preference for dsDNA over ssDNA, while TET1-CD and TET3-CD are essentially agnostic to the presence or absence of the complementary strand (Figure S5). As prior studies have hinted at a length dependence to substrate reactivity[25,26], we generated two additional 5mdC-containing substrates containing the embedded $D(M)_1$ 16-mer sequence in a 27-mer or a 60-mer oligonucleotide and compared ssDNA versus dsDNA

reactivity. Although reactivity of these substrates does decrease with length, the observation that dsDNA is more reactive than ssDNA is maintained across all three substrates in the series (Figure 2B, Figure S6). The observation that ssDNA reactivity is slightly enhanced by a DNA complement also carries forward to the chimeric substrate $D(m)_1$, though notably the target RNA base in $D(m)_1$ has the greater effect of lowering overall activity on both single and double stranded substrates (Figure 2A).

In contrast to the improved reactivity when duplexing DNA with DNA, duplexing any target strand with RNA dampens reactivity. TET2-CS showed reactivity on the 5mdC of the DNA:RNA hybrid duplex ($D(M)_1$:$r(c)_2$), but total oxidation events are decreased relative to the dsDNA substrate $D(M)_1$:$D(C)_1$ (23% reduction); in comparison, $D(m)_1$ experiences an even greater reduction in reactivity (68%) when complexed to RNA (Figure 2A). Interestingly, for the chimeric $r(M)_1$ substrate, whose reactivity rivaled the all-DNA $D(M)_1$ substrate when single-stranded, reactivity decreases when duplexing this substrate to the DNA complement and decreases further when duplexing to the RNA complement. The disfavored nature of an RNA complementary strand is most evident with the all-RNA substrate, $r(m)_1$, where reactivity is nearly eliminated on the dsRNA substrate, contrary to prior suggestions that dsRNA may be a substrate[27]. Thus, although the target base identity is a critical determinant of reactivity with single-stranded substrates, these preferences can be overridden by disfavored duplex structures in the rest of the substrate.

Activity on symmetrically methylated duplexes.

Our analysis of duplex substrates initially focused on a single reactive 5mC site in the target strand, but we recognized that these substrates are well suited to examining a more complex question. Considering that most CG:CG dinucleotide pairs in genomic DNA are symmetrically methylated[52], TET enzymes likely encounter duplex substrates where both strands contain reactive 5mC nucleobases. The DNA:RNA hybrids, $D(M)_1$:$r(m)_2$ and $D(M)_2$:$r(m)_1$, as well as the chimera hybrids, $D(m)_1$:$r(M)_2$ and $D(m)_2$:$r(M)_1$, are instructive experimental substrates, as in each case, activity on each strand can be assessed independently, given that one strand contains 5mdC, while the other contains 5mrC. With $D(M)_1$:$r(m)_2$ and $r(m)_1$:$D(M)_2$, we observed nearly symmetrically-reciprocal results, with high activity on the all-DNA strand and less activity on the all-RNA strand in both cases (Figure 3A). Notably, comparing these two DNA:RNA hybrids with the dsDNA duplex (5mC on both strands) shows a greater proportion of the 5mdC converted to 5cadC for the all-DNA substrate (Figure 3B). This result suggests that having an RNA complement strand, whether it is reactive or non-reactive as above, decreases activity on the DNA strand. With the chimeric substrates, $D(m)_1$:$R(M)_2$ and $r(M)_1$:$D(m)_2$, the RNA strand containing 5mdC is more reactive than the DNA strand containing 5mrC (Figure 3). This observation is consistent with a dominant influence for the identity of the target nucleotide over that of the flanking nucleic acids in the target strand.

Modeling of 5mrC in TET-dsDNA structure.

We have used classical molecular dynamics (MD) simulations to examine the TET2-CS structure complexed to dsDNA[32] in order to understand the mechanisms that could contribute to nucleic acid selectivity. Given the importance of the target site identity revealed

by our biochemical analysis, we converted 5mdC to 5mrC by inserting a single 2'-hydroxyl at the target 5mdC base and then examined its impact on non-bonded interactions and H-bond dynamics. The MD results provide a qualitative assessment that help explain the tolerance of TET2 toward 5mrC at the target site, as well as the preference for oxidation of 5mdC over 5mrC.

The mutation of 5mdC to 5mrC results in both global and local changes in the structure and dynamics of the system that help explain the decreased reactivity of 5mrC (Figure 4, Figure S7–9). At a global level, the inclusion of the 2'-OH results in an overall destabilizing interaction of the DNA with the protein ($\Delta E_{5mdC\ system\ -\ 5mrC\ system}$=−47.4 kcal/mol) (Fig 4A). Additionally, the normal mode analysis of both systems suggests that the dynamics are also affected by the inclusion of the 2'-OH (Figure S7A,B). Interestingly, the substitution of 5mdC with 5mrC in the target strand creates a partial shift from B-form DNA, most distinctly shown by the change in shift and slide toward A-form character in two base pairs upstream (5') of the target base (Figure S8), as well as altered intermolecular interactions spanning the R1260-R1269 region of the enzyme (Figure S9). The A-form is common in RNA and DNA-RNA duplexes and in protein-DNA complexes for polymerases and endonucleases[53]. These alterations could partially explain the decreased reactivity of the 5mrC chimeric substrates as well as the more profound defect in oxidation of dsRNA that was observed.

A more local view of the active site sheds light on differences that can explain the decreased reactivity of 5mrC, but also the preservation of core elements that permit 5mrC to be oxidized. The 2'-OH of 5mrC does not introduce any evident steric conflicts, but results in the formation of an H-bond with the phosphate backbone of G7, the downstream base in the CpG dyad (Figure 4, Figure S9, Table S1). This DNA structural change alters the interactions of the substrate in the active site, including an increase in the predominance of the H-bonds between G7 and S1290 and K1299 with the simultaneous loss of H-bonds from T1259 and S1286 to G7. Residues R1261 and R1262, which form key interactions with the target base backbone (Figure S9) and with the αKG analog, respectively, are stabilized via additional Coulomb and van der Waals interactions in the 5mrC system. Nonetheless, despite these interactions with the backbone of the target strand, there are changes suggesting decreased engagement of the nucleobase to the enzyme: In addition to loss of an H-bond between S1290 and the nucleobase, the H-bond interactions on the Watson-Crick face of the nucleobase from H1904 are reduced by a third of the overall simulation time in the 5mrC-containing structure compared with the 5mdC system. Notably, other core components of the active site are largely intact. In particular, we have previously demonstrated the importance of an active site scaffold in modeling and mutagenesis experiments, whereby the active site T1372 hydrogen bonds with Y1902, which stacks with the target cytosine nucleobase to position it optimally for oxidation[32]. The introduction of 5mrC does not have a noticeable impact on the intermolecular interactions of the scaffold.

DISCUSSION

Given the expanded recognition of the role of oxidized 5mC bases in DNA and RNA biology, here we employed a systematic approach to explore the nucleic acid determinants of

TET reactivity. Our results offer a view of the selectivity and promiscuity that characterizes TET-mediated oxidation and allow for substantiation of a model (Figure 5) that links reactivity to the nature of the target nucleotide, the flanking nucleotides on the target strand, and the complementary strand.

Our results uncover a strong association between the nucleic acid identity of the target nucleotide and TET reactivity. We find that TET enzymes are tolerant to both ssDNA and ssRNA substrates, but ssDNA substrates are preferred, consistent with past studies[11]. Strikingly, the diminished reactivity of the ssRNA can be rescued by removing the single 2'-OH from the target nucleotide. Correspondingly, the addition of a single 2'-OH at the target nucleotide decreases the reactivity of ssDNA to the level observed with the ssRNA substrate. These results align with findings from an *in vitro* study showing that TET can oxidize substrates containing 2'-(R)-fluorinated modified cytosines, albeit to a lesser extent than 5mdC-containing DNA[54]. In support of our biochemical findings, molecular dynamic simulations suggest that the addition of a single 5mrC in the active site of a TET2-dsDNA structure can be accommodated, allowing for key nucleobase-orienting interactions to be maintained, but results in significant energetic and conformational changes surrounding the target base relative to 5mdC which could explain its altered reactivity. These findings are interesting to consider in the context of other DNA/RNA modifying enzymes. Some enzymes, such as the base excision repair enzyme uracil DNA glycosylase, potently discriminate against a ribonucleotide target using steric discrimination against the 2'-OH[55]. For other enzymes, such as AID/APOBEC family DNA deaminase enzymes, a preference for a deoxynucleotide over ribonucleotide target appears linked to the preferred sugar puckers accessible to the DNA substrates[56]. The modeling with TET2-CS suggests an alternative mechanism that can explain tolerance for 5mrC, but a preference for 5mdC. The 5mrC substrate can be accommodated from a steric and conformational perspective, but the interactions of the reactive base appear disfavored relative to 5mdC.

Our biochemical data also inform our understanding of the role of the flanking and complementary strand nucleotides on activity. dsDNA is the preferred substrate, but DNA:RNA hybrids are also tolerated when the DNA is the substrate (Figure 5). The only conformation that appears to be strongly discriminated against is dsRNA. While a single 5mdC can alleviate some of the negative effects of the RNA duplex, it is still strongly disfavored, possibly due to the discrimination against A-form conformations. The proficiency for oxidation of ssRNA suggests that the conformational flexibility available to single-stranded substrates, but not dsRNA, facilitates promiscuous oxidation of 5mC. In considering the impact of the nucleotides outside of the target site, TET2 offers an interesting comparison to another base-flipping enzyme, the RNA methyltransferase, DNMT2. When methylation activity was examined on tRNA substrates, a chimera that contains a dC in place of rC in the loop region of the tRNA substrate was more, rather than less, efficiently methylated than the all-RNA natural analog[57]. Therefore, rather than having selectivity primarily dictated by the target nucleotide as we observe for TET2, the flanking nucleotides and their secondary structure determines reactivity for DNMT2.

The evident promiscuity in TET family enzymes raises important points with regards to their physiological function. dsDNA, ssDNA, ssRNA, and DNA:RNA hybrids co-exist in

mammalian cells and represent four of the five most preferred substrates across our tested series, which lends biochemical credence to the possibility that TET oxidation products could be generated in any of these settings. At the same time, the observed promiscuity *in vitro* raises questions about how TET-mediated oxidation could be targeted *in vivo*. For functions when activity on DNA predominates, spatial regulation is likely to play an important role. TET1 and TET2 have been shown to primarily localize to the nucleus in both overexpression studies and in cells, while TET3 localizes to both the cytoplasm and the nucleus[58–61]. Regulation could also involve the non-catalytic domains and/or partner proteins. Our studies focused on the crystal structure version of TET2 and the catalytic domains of TET1–3, but full-length TET1 and TET3 have CXXC-domains that can mediate targeting to non-methylated CpG islands in DNA, while TET2 is thought to utilize its ancestral gene neighbor, IDAX, for this purpose[62]. For functions where activity on RNA is thought to predominate, major questions remain open regarding the location and timing of cytosine methylation in RNA and how selectivity for TET-mediated oxidation could be achieved.

More broadly, beyond the mammalian TET1–3 enzymes, it will be interesting to see if promiscuity towards DNA and ssRNA is a shared feature of the TET enzymes found across the phylogenetic tree, especially in organisms that offer a restricted or unusual selection of nucleic acid substrates[21,22]. TET homologs in some species have been associated with transposons and retroelements, suggesting the possibility that their activity could be associated with less conventional nucleic acid intermediates associated with mobile genetic elements[63]. Among the species with distinct TET homologs, Drosophila stand out in particular since they lack methylated DNA yet have retained a TET homolog that appears to act on RNA. Whether the proposed impact on protein translation applies to other species, or whether this represents a unique branch point in the evolution of TET activity, remains to be explored. Our findings are also interesting to consider in the context of the broader nucleic acid dioxygenase family, which also includes AlkB family enzymes. The promiscuity of AlkB enzymes towards DNA and RNA has been thought broaden their function in nucleic acid repair[64] and may have been exploited in the evolution of diverse physiological functions potentially associated with TET enzymes.

The observed promiscuity of TET activity adds to the challenges of studying DNA and RNA cytosine modifications. In knockouts, phenotypes attributed to activity on 5mdC in DNA, may instead be related to altered oxidation of 5mrC in RNA or vice versa. Engineering of TET2-CS has been able to yield variants with altered selectivity[32,65]. Our MD simulations suggest that it may similarly be possible to find mutants with altered selectivity by manipulating the interactions that differentiate 5mdC from 5mrC engagement in the active site. For example, mutations of key residues such as S1290 or K1299 in TET2 could potentially enhance or diminish the activity on DNA versus RNA. Indeed, beyond improving our understanding of the molecular determinants of enzyme selectivity, building enhanced specificity into TET enzymes could offer tools to better explore their biological roles as well as expand the power of CRISPR/dCas9-based epigenetic editing biotechnology[66,67].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors would like to thank the UNT Chemistry Department for computing time (via NSF CHE-1531468), Dr. A. R. Walker for assistance with the parameterization of the 5mrC nucleotide and meaningful discussions, and Dr. Clementina Mesaros for providing mass spectrometry support and guidance. We also thank Tong Wang for editorial assistance.

Funding Sources

This work was supported by the National Institutes of Health (R01-GM118501 to R.M.K.) and the National Science Foundation (DGE-1321851 to J.E.D.).

ABBREVIATIONS

5mC	5-methylcytosine
5mdC	5-methyl-2'-deoxycytidine
CpG	cytosine guanine dinucleotides
TET	Ten-eleven translocation
α-KG	α-ketoglutarate
5hmdC	5-hydroxymethyl-2'-deoxycytidine
5fdC	5-formyl-2'-deoxycytidine
5cadC	5-carboxyl-2'-deoxycytidine
dC	2'-deoxycytidine
rC	cytidine
5mrC	5-methylcytidine
5hmrC	5-hydroxymethylcytidine
5frC	5-formylcytidine
5carC	5-carboxylcytidine
IDH	isocitrate dehydrogenase
dsDNA/RNA	double-stranded DNA/RNA
ssDNA/RNA	single-stranded DNA/RNA
MD	molecular dynamics
EDA	energy decomposition analysis
PCA	principal component analysis

NMA	normal mode analysis

REFERENCES

1. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, and Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 324, 930–935. [PubMed: 19372391]
2. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, and Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 333, 1300–1303. [PubMed: 21778364]
3. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, and Xu GL (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 333, 1303–1307. [PubMed: 21817016]
4. Liu MY, DeNizio JE, Schutsky EK, and Kohli RM (2016) The expanding scope and impact of epigenetic cytosine modifications. Curr. Opin. Chem. Biol 33, 67–73. [PubMed: 27315338]
5. Mellen M, Ayata P, and Heintz N (2017) 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. Proc. Natl. Acad. Sci. U. S. A 114, E7812–E7821. [PubMed: 28847947]
6. Marina RJ, Sturgill D, Bailly MA, Thenoz M, Varma G, Prigge MF, Nanan KK, Shukla S, Haque N, and Oberdoerffer S (2016) TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. EMBO J. 35, 335–355. [PubMed: 26711177]
7. Wu H, and Zhang Y (2015) Charting oxidized methylcytosines at base resolution. Nat. Struct. Mol. Biol 22, 656–661. [PubMed: 26333715]
8. Motorin Y, Lyko F, and Helm M (2010) 5-methylcytosine in RNA: Detection, enzymatic formation and biological functions. Nucleic Acids Res. 38, 1415–1430. [PubMed: 20007150]
9. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, and Preiss T (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. 40, 5023–5033. [PubMed: 22344696]
10. Huang W, Lan MD, Qi CB, Zheng SJ, Wei SZ, Yuan BF, and Feng YQ (2016) Formation and determination of the oxidation products of 5-methylcytosine in RNA. Chem. Sci 7, 5495–5502. [PubMed: 30034689]
11. Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, Cai Q, Ji D, Jin SG, Niedernhofer LJ, Pfeifer GP, Xu GL, and Wang Y (2014) Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J. Am. Chem. Soc 136, 11582–11585. [PubMed: 25073028]
12. Huber SM, van Delft P, Mendil L, Bachman M, Smollett K, Werner F, Miska EA, and Balasubramanian S (2015) Formation and abundance of 5-hydroxymethylcytosine in RNA. ChemBioChem. 16, 752–755. [PubMed: 25676849]
13. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, and Carell T (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One. 5, e15367. [PubMed: 21203455]
14. Wagner M, Steinbacher J, Kraus TF, Michalakis S, Hackner B, Pfaffeneder T, Perera A, Muller M, Giese A, Kretzschmar HA, and Carell T (2015) Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. Angew. Chem. Int. Ed Engl. 54, 12511–12514. [PubMed: 26137924]
15. Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, Li A, Wang X, Bhattarai DP, Xiao W, Sun HY, Zhu Q, Ma HL, Adhikari S, Sun M, Hao YJ, Zhang B, Huang CM, Huang N, Jiang GB, Zhao YL, Wang HL, Sun YP, and Yang YG (2017) 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. Cell Res. 27, 606–625. [PubMed: 28418038]
16. Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, Calonne E, Hassabi B, Putmans P, Awe S, Wetzel C, Kreher J, Soin R, Creppe C, Limbach PA, Gueydan C, Kruys V, Brehm A, Minakhina S, Defrance M, Steward R, and Fuks F (2016) Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 351, 282–285. [PubMed: 26816380]

17. Shen Q, Zhang Q, Shi Y, Shi Q, Jiang Y, Gu Y, Li Z, Li X, Zhao K, Wang C, Li N, and Cao X (2018) Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. Nature. 554, 123–127. [PubMed: 29364877]

18. Schapira M (2016) Structural chemistry of human RNA methyltransferases. ACS Chem. Biol 11, 575–582. [PubMed: 26566070]

19. He C, Sidoli S, Warneford-Thomson R, Tatomer DC, Wilusz JE, Garcia BA, and Bonasio R (2016) High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. Mol. Cell 64, 416–430. [PubMed: 27768875]

20. Xu Q, Wang K, Wang L, Zhu Y, Zhou G, Xie D, and Yang Q (2016) IDH1/2 mutants inhibit TET-promoted oxidation of RNA 5mC to 5hmC. PLoS One. 11, e0161261. [PubMed: 27548812]

21. Iyer LM, Tahiliani M, Rao A, and Aravind L (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle. 8, 1698–1710. [PubMed: 19411852]

22. Iyer LM, Zhang D, Burroughs AM, and Aravind L (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res. 41, 7635–7655. [PubMed: 23814188]

23. Raddatz G, Guzzardo PM, Olova N, Fantappie MR, Rampp M, Schaefer M, Reik W, Hannon GJ, and Lyko F (2013) Dnmt2-dependent methylomes lack defined DNA methylation patterns. Proc. Natl. Acad. Sci. U. S. A 110, 8627–8631. [PubMed: 23641003]

24. Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, Shi YG, Zhu J, Wang P, and Xu Y (2013) Crystal structure of TET2-DNA complex: Insight into TET-mediated 5mC oxidation. Cell. 155, 1545–1555. [PubMed: 24315485]

25. 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, and Xu Y (2015) Structural insight into substrate preference for TET-mediated oxidation. Nature. 527, 118–122. [PubMed: 26524525]

26. Kizaki S, and Sugiyama H (2014) CGmCGCG is a versatile substrate with which to evaluate tet protein activity. Org. Biomol. Chem 12, 104–107. [PubMed: 24162071]

27. Basanta-Sanchez M, Wang R, Liu Z, Ye X, Li M, Shi X, Agris PF, Zhou Y, Huang Y, and Sheng J (2017) TET1-mediated oxidation of 5-formylcytosine (5fC) to 5-carboxycytosine (5caC) in RNA. ChemBioChem. 18, 72–76. [PubMed: 27805801]

28. Liu MY, DeNizio JE, and Kohli RM (2016) Quantification of oxidized 5-methylcytosine bases and TET enzyme activity. Methods Enzymol. 573, 365–385. [PubMed: 27372762]

29. Crawford DJ, Liu MY, Nabel CS, Cao XJ, Garcia BA, and Kohli RM (2016) Tet2 catalyzes stepwise 5-methylcytosine oxidation by an iterative and de novo mechanism. J. Am. Chem. Soc 138, 730–733. [PubMed: 26734843]

30. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, and Walker RC (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh ewald. J. Chem. Theory Comput. 9, 3878–3888. [PubMed: 26592383]

31. Case DA, Cerutti SD, Cheatham TE, 3rd, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein D, Merz KM, Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Simmerling CL, Botello-Smith MW, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, York MD, and Kollman PA (2017) Amber 2017. University of California, San Francisco.

32. Liu MY, Torabifard H, Crawford DJ, DeNizio JE, Cao XJ, Garcia BA, Cisneros GA, and Kohli RM (2017) Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. Nat. Chem. Biol 13, 181–187. [PubMed: 27918559]

33. Sali A, and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol 234, 779–815. [PubMed: 8254673]

34. Fiser A, Do RK, and Sali A (2000) Modeling of loops in protein structures. Protein Sci. 9, 1753–1773. [PubMed: 11045621]

35. Vanquelef E, Simon S, Marquant G, Garcia E, Klimerak G, Delepine JC, Cieplak P, and Dupradeau FY (2011) R.E.D. server: A web service for deriving RESP and ESP charges and building force

field libraries for new molecules and molecular fragments. Nucleic Acids Res. 39, W511–7. [PubMed: 21609950]

36. Wang F, Becker J-P, Cieplack P, and Dupradeau F-Y (2013) R.E.D. python: Object oriented programming for amber force fields. Université de Picardie-Jules Verne, Sanford Burnham Prebys Medical Discovery Institute.
37. Dupradeau FY, Pigache A, Zaffran T, Savineau C, Lelong R, Grivel N, Lelong D, Rosanski W, and Cieplak P (2010) The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. Phys. Chem. Chem. Phys 12, 7821–7839. [PubMed: 20574571]
38. Bayly CI, Cieplak P, Cornell W, and Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. J. Phys. Chem 97, 10269–10280.
39. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich A, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery J,J. A, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Keith T, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, and Fox DJ (2016) Gaussian 09, Revision A.02 Gaussian, Inc., Wallingford, CT.
40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, and Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J. Chem. Phys 79, 926.
41. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG (1995) A smooth particle mesh ewald method. J. Chem. Phys 103, 8577–8593.
42. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, and Haak JR (1984) Molecular dynamics with coupling to an external bath. J. Chem. Phys 81, 3684–3690.
43. Roe DR, and Cheatham TE, 3rd. (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J. Chem. Theory Comput. 9, 3084–3095. [PubMed: 26583988]
44. Babcock MS, Pednault EP, and Olson WK (1994) Nucleic acid structure analysis. mathematics for local cartesian and helical structure parameters that are truly comparable between structures. J. Mol. Biol 237, 125–156. [PubMed: 8133513]
45. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung CS, Westhof E, Wolberger C, and Berman HM (2001) A standard reference frame for the description of nucleic acid base-pair geometry. J. Mol. Biol 313, 229–237. [PubMed: 11601858]
46. Graham SE, Syeda F, and Cisneros GA (2012) Computational prediction of residues involved in fidelity checking for DNA synthesis in DNA polymerase I. Biochemistry. 51, 2569–2578. [PubMed: 22397306]
47. Elias AA, and Cisneros GA (2014) Computational study of putative residues involved in DNA synthesis fidelity checking in thermus aquaticus DNA polymerase I. Adv. Protein Chem. Struct. Biol 96, 39–75. [PubMed: 25443954]
48. Dewage SW, and Cisneros GA (2015) Computational analysis of ammonia transfer along two intramolecular tunnels in staphylococcus aureus glutamine-dependent amidotransferase (GatCAB). J. Phys. Chem. B 119, 3669–3677. [PubMed: 25654336]
49. Humphrey W, Dalke A, and Schulten K (1996) VMD: Visual molecular dynamics. J. Mol. Graph 14, 33–38. [PubMed: 8744570]
50. Bakan A, Meireles LM, and Bahar I (2011) ProDy: Protein dynamics inferred from theory and experiments. Bioinformatics. 27, 1575–1577. [PubMed: 21471012]
51. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF chimera--a visualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612. [PubMed: 15264254]

52. Schubeler D (2015) Function and information content of DNA methylation. Nature. 517, 321–326. [PubMed: 25592537]
53. Kulkarni M, and Mukherjee A (2017) Understanding B-DNA to A-DNA transition in the right-handed DNA helix: Perspective from a local to global transition. Prog. Biophys. Mol. Biol 128, 63–73. [PubMed: 28576665]
54. Schröder AS, Parsa E, Iwan K, Traube FR, Wallner M, Serdjukow S, and Carell T (2016) 2'-(R)-fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. Chem. Commun. (Cambridge, U.K.). 52, 14361–14364.
55. Pearl LH (2000) Structure and function in the uracil-DNA glycosylase superfamily. Mutat. Res 460, 165–181. [PubMed: 10946227]
56. Nabel CS, Lee JW, Wang LC, and Kohli RM (2013) Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. Proc. Natl. Acad. Sci. U. S. A 110, 14225–14230. [PubMed: 23942124]
57. Kaiser S, Jurkowski TP, Kellner S, Schneider D, Jeltsch A, and Helm M (2017) The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. RNA Biol. 14, 1241–1251. [PubMed: 27819523]
58. Arioka Y, Watanabe A, Saito K, and Yamada Y (2012) Activation-induced cytidine deaminase alters the subcellular localization of Tet family proteins. PLoS One. 7, e45031. [PubMed: 23028748]
59. Muller T, Gessi M, Waha A, Isselstein LJ, Luxen D, Freihoff D, Freihoff J, Becker A, Simon M, Hammes J, Denkhaus D, zur Muhlen A, Pietsch T, and Waha A (2012) Nuclear exclusion of TET1 is associated with loss of 5-hydroxymethylcytosine in IDH1 wild-type gliomas. Am. J. Pathol 181, 675–683. [PubMed: 22688054]
60. Di Stefano B, Sardina JL, van Oevelen C, Collombet S, Kallin EM, Vicent GP, Lu J, Thieffry D, Beato M, and Graf T (2014) C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature. 506, 235–239. [PubMed: 24336202]
61. Huang Y, Wang G, Liang Z, Yang Y, Cui L, and Liu CY (2016) Loss of nuclear localization of TET2 in colorectal cancer. Clin. Epigenetics 8, 9, eCollection 2016. [PubMed: 26816554]
62. Ko M, An J, Bandukwala HS, Chavez L, Aijo T, Pastor WA, Segal MF, Li H, Koh KP, Lahdesmaki H, Hogan PG, Aravind L, and Rao A (2013) Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature. 497, 122–126. [PubMed: 23563267]
63. Iyer LM, Zhang D, de Souza RF, Pukkila PJ, Rao A, and Aravind L (2014) Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. Proc. Natl. Acad. Sci. U. S. A 111, 1676–1683. [PubMed: 24398522]
64. Falnes PØ, Bjørås M, Aas PA, Sundheim O, and Seeberg E (2004) Substrate specificities of bacterial and human AlkB proteins. Nucleic Acids Res. 32, 3456–3461. [PubMed: 15229293]
65. Sudhamalla B, Wang S, Snyder V, Kavoosi S, Arora S, and Islam K (2018) Complementary steric engineering at the protein-ligand interface for analogue-sensitive TET oxygenases. J. Am. Chem. Soc 140, 10263–10269. [PubMed: 30028600]
66. Komor AC, Badran AH, and Liu DR (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell. 169, 559.
67. DeNizio JE, Schutsky EK, Berrios KN, Liu MY, and Kohli RM (2018) Harnessing natural DNA modifying activities for editing of the genome and epigenome. Curr. Opin. Chem. Biol 45, 10–17. [PubMed: 29452938]

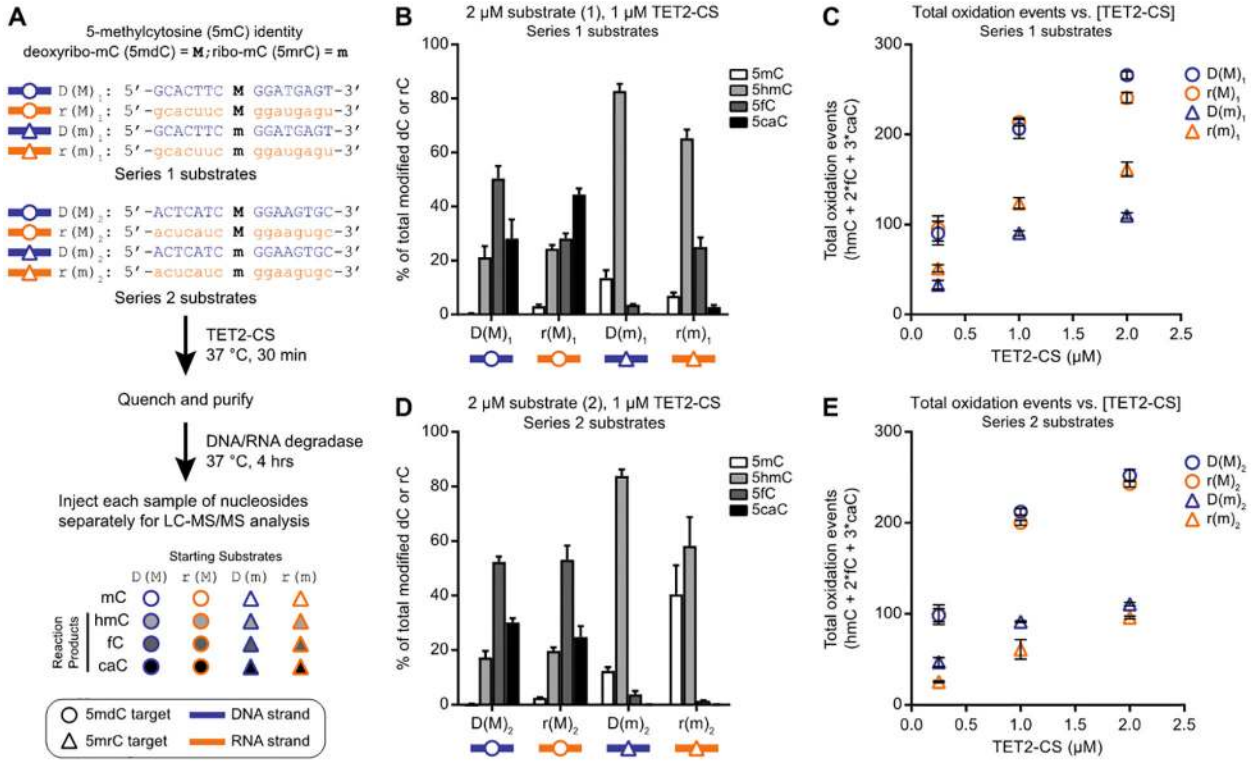


Figure 1.

TET2 discriminates mostly based on the sugar identity of the target nucleotide rather than that of the flanking target strand. (A) Workflow showing the substrates and procedures used in the experiments. (B) Activity of TET2-CS with the Series 1 substrates. Reaction products were quantified by LC-MS/MS and expressed as the percentage of each base relative to the amount of total modified cytosine (DNA or RNA). (C) The reaction in (B) was repeated at three enzyme concentrations, and the reaction products were expressed as total oxidation events: the generation of 5hmC counts once, 5fC counts twice, and 5caC counts three times. (D-E) The same experiments using Series 2 substrates. Data for individual oxidation products used to generate (C and E) can be found in Figure S3A,B. All error bars represent standard deviation for 3–5 independent replicates.

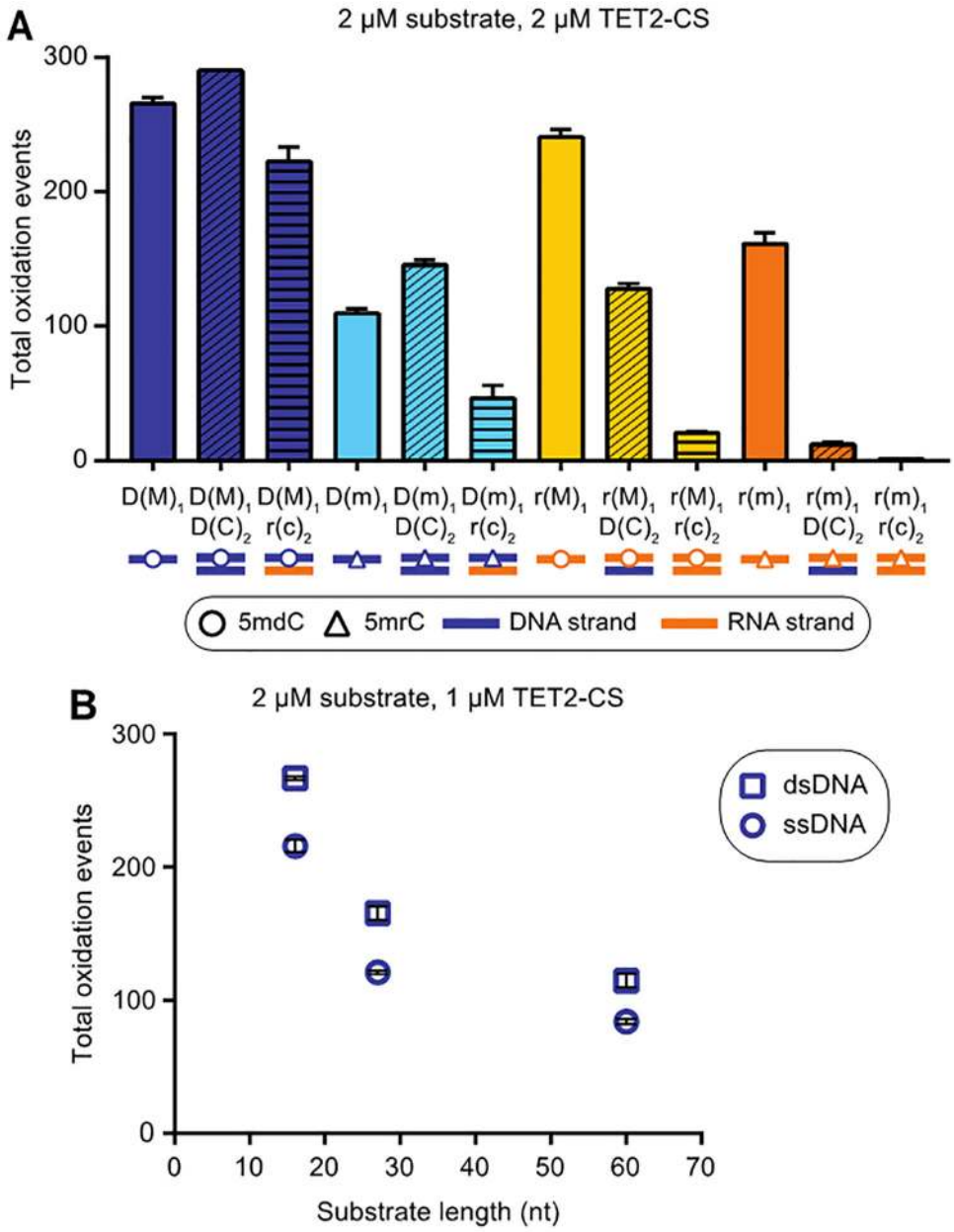


Figure 2.

Double-stranded DNA is the preferred TET substrate, while dsRNA is strongly disfavored. (A) Activity of TET2-CS on Series 1 substrates, either single-stranded or duplexed to a non-reactive DNA or RNA complementary strand. (B) Comparison of TET2-CS activity on sequence-matched ssDNA versus dsDNA substrates of different lengths. Error bars represent standard deviation for 2–3 independent replicates.

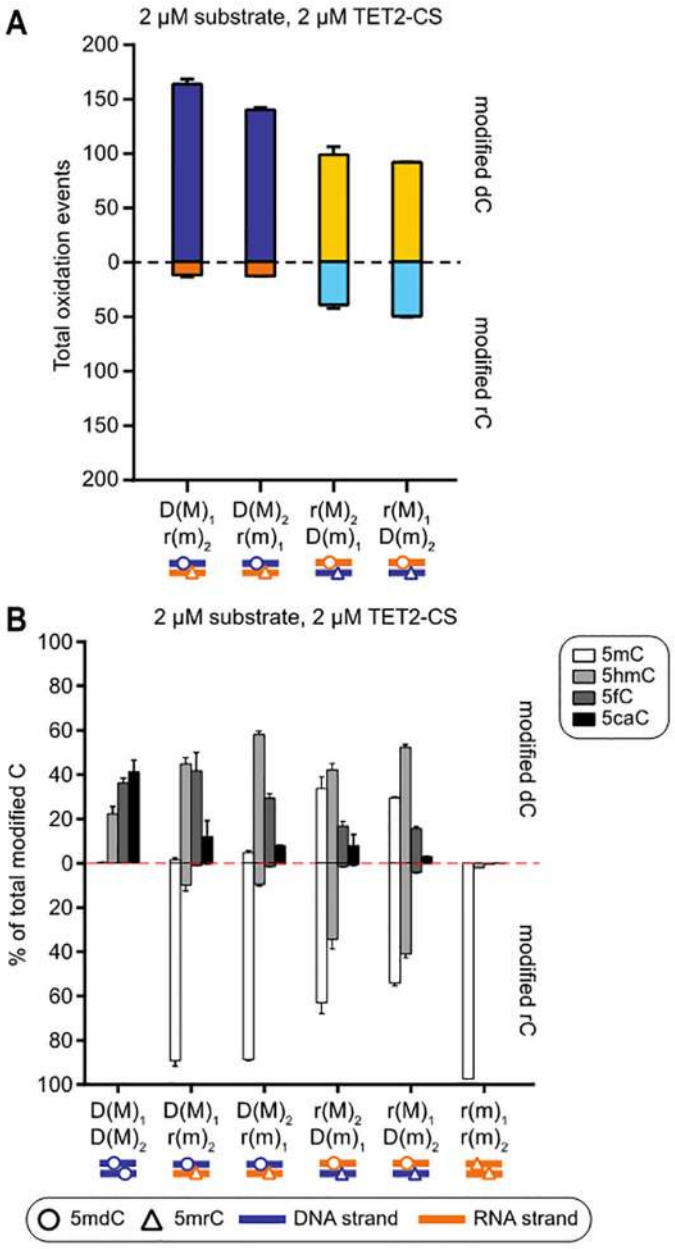


Figure 3.

For duplexes with two reactive strands, TET2 activity on each strand is largely independent of the other, complementary strand. (A) Series 1 and 2 substrates, each containing 5mC, were duplexed to one another. Reaction products are shown as the total oxidation events on each strand. (B) Reaction products from the same reactions are also shown as percentages of each modification for DNA (top half of graph) or RNA (bottom half of graph) targets. Also included for comparison are results for symmetrically methylated dsDNA and dsRNA, showing the reaction products for both strands combined.

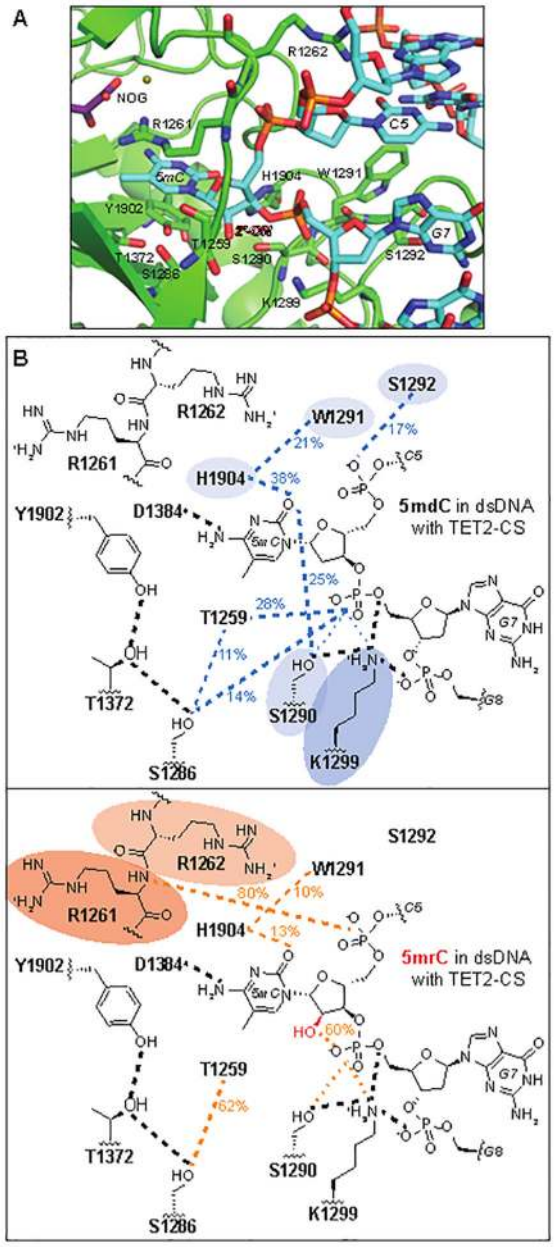


Figure 4.

Modeling 5mrC as the target base in dsDNA bound to hTET2 results in dynamic and structural changes. (A) The active site of TET2-CS (green) bound to dsDNA (teal) containing 5mrC. (B) H-bond network and energetic differences in models of the active site with either 5mdC (top) or 5mrC (bottom). H-bonds (present >10% simulation time in at least one system) are designated by dashed lines, including those that occur in both models (<10% difference, black). When there is more than one bond that contributes to an interaction between two molecules, it is shown as a dotted line and the simulation time present is not shown (see Figure S9 for all contributing data). For H-bonds that occur in both systems but at different levels (≥10% difference), the bond line is thinner in the system in which it is less prevalent. Significantly different non-bonded interactions (sum of Coulombic

and van der Waals energies $\geq|1|$ kcal/mol) between a TET residue and the target 5mC in the two systems are highlighted according to the relative magnitude of the difference (see Table S1 for values).

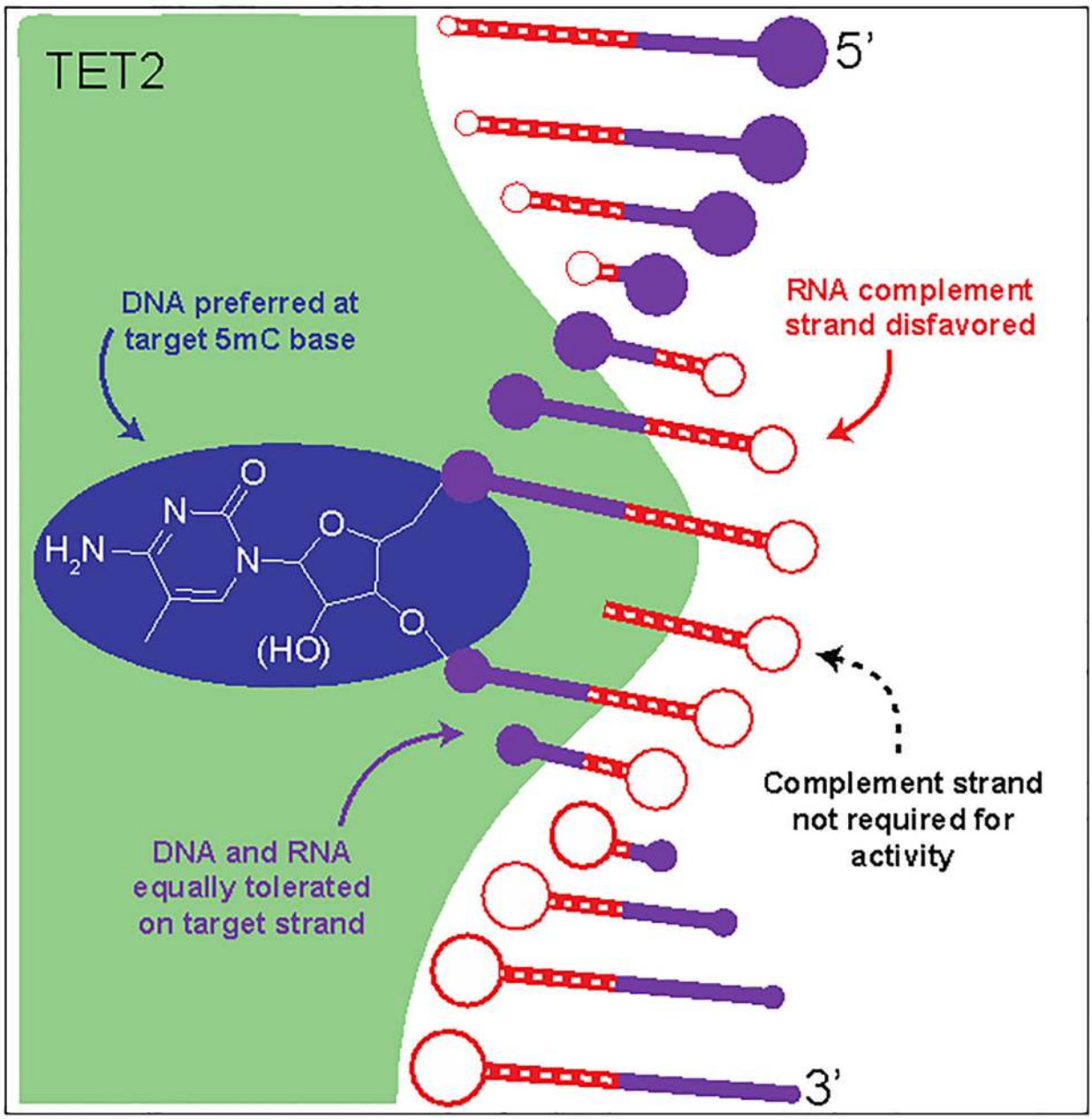


Figure 5.

Summary of substrate determinants of TET reactivity from biochemical findings. The 2'-OH present in 5mrC, but absent in 5mdC, is shown in parentheses. The complement strand is shown as a dashed line to indicate that it is not required for TET activity.