

Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine

Monica Yun Liu^{1,2}, Hedieh Torabifard³, Daniel J Crawford^{1,2}, Jamie E DeNizio^{1,2}, Xing-Jun Cao², Benjamin A Garcia², G Andrés Cisneros^{3,4*} & Rahul M Kohli^{1,2*}

Ten-eleven translocation (TET) enzymes catalyze stepwise oxidation of 5-methylcytosine (mC) to yield 5-hydroxymethylcytosine (hmC) and the rarer bases 5-formylcytosine (fC) and 5-carboxylcytosine (caC). Stepwise oxidation obscures how each individual base forms and functions in epigenetic regulation, and prompts the question of whether TET enzymes primarily serve to generate hmC or are adapted to produce fC and caC as well. By mutating a single, conserved active site residue in human TET2, Thr1372, we uncovered enzyme variants that permit oxidation to hmC but largely eliminate fC and caC. Biochemical analyses, combined with molecular dynamics simulations, elucidated an active site scaffold that is required for wild-type (WT) stepwise oxidation and that, when perturbed, explains the mutants' hmC-stalling phenotype. Our results suggest that the TET2 active site is shaped to enable higher-order oxidation and provide the first TET variants that could be used to probe the biological functions of hmC separately from fC and caC.

The discovery of ten-eleven translocation (TET) enzymes transformed the known repertoire of epigenetic DNA modifications¹. TET enzymes catalyze the oxidation of 5-methylcytosine (mC), the mainstay of the epigenome, into three additional bases: 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxylcytosine (caC)^{1–6}. Mounting evidence suggests that these oxidized mC (ox-mC) bases stably populate mammalian genomes, aid in DNA demethylation, and potentially encode unique epigenetic information^{7–11}. The central questions now facing the field involve the functions of each individual base and the mechanisms governing their formation.

The overall catalytic mechanism of TET enzymes (TET1, TET2, and TET3 in mammals) has been largely inferred from related proteins in the Fe(II)/ α -ketoglutarate (α -KG)-dependent family of dioxygenases, such as AlkB¹². Enzymes in this family couple decarboxylation of α -KG with substrate oxidation via a transient Fe(IV)-oxo intermediate, with succinate and CO₂ as byproducts. TET enzymes apply this general mechanism to not one but three stepwise reactions, raising the question of whether these enzymes are specialized for one particular step of oxidation or for three-step oxidation as a whole. Moreover, stepwise oxidation obscures the function of individual ox-mCs, creating a need to break the linkage between steps in order to study each base in isolation.

The first step of oxidation, conversion of mC to hmC, has so far drawn the most attention, as it best explains the physiological levels of cytosine modifications; in the human genome, mC accounts for approximately 0.6–1% of all bases, hmC is typically 1–5% of mC, and fC and caC are at least 1–2 orders of magnitude rarer than hmC¹⁰. Consistent with these observations, biochemical studies have shown that mC substrate is preferred over hmC and fC, with two- to five-fold differences in K_M and k_{cat} reported for human TET2 (ref. 13). Crystal structures did not reveal substrate-specific interactions that could explain these differences^{13,14}, but computational modeling suggested that hydrogen abstraction is more efficient on mC than on hmC and fC, which adopt unfavorable conformations^{13,15}. Together, these studies portray TET enzymes as predominantly

serving to generate hmC; in this case, decreased capacity for further oxidation would help to maintain stable levels of hmC for epigenetic functions. Indeed, most functional studies on ox-mC bases have focused on hmC in health and disease, with fC and caC considered as fairly negligible.

However, this view does not explain why fC and caC are present at all, and it contrasts with evidence for the importance of higher-order oxidation. Most notably, fC and caC, but not hmC, are substrates for base excision by thymine DNA glycosylase (TDG); the resulting abasic site can be repaired to regenerate unmodified cytosine^{5,16,17}. This is the leading candidate pathway for active DNA demethylation⁷. Apart from being intermediates in demethylation, fC and caC potentially also function as stable epigenetic marks. Genomic sequencing has mapped fC and caC to gene regulatory regions separate from hmC¹⁰, and proteomic analysis has described distinct 'reader' proteins for each ox-mC base^{18,19}. Furthermore, mouse Tet2 is capable of iterative oxidation: it can catalyze multiple rounds of oxidation upon a single encounter with mC-containing DNA without releasing the hmC-containing DNA strand²⁰. Although the prevalence of genomic hmC implies that most encounters are not iterative, this mechanism could allow TET enzymes to generate fC and caC marks without first accumulating hmC. Together, these studies encourage the alternate view that TET enzymes are specialized for making not only hmC but fC and caC as well—and even that conversion of hmC to fC could be the key 'committed' step to DNA demethylation.

To resolve these competing views of TET function, one question comes to the fore: whether TET enzymes are adapted to facilitate higher-order oxidation. The mC-to-hmC step is most favored, but if fC and caC serve important functions, mechanisms should be in place to permit their formation, yet these mechanisms remain largely unknown. They could be extrinsic to TET—e.g., other proteins could recruit TET enzymes or regulate their activity. However, intrinsic features, especially structure–function support for higher-order oxidation, would suggest an enzyme specifically shaped to generate not one but three epigenetic bases.

¹Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ²Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ³Department of Chemistry, Wayne State University, Detroit, Michigan, USA. ⁴Present address: Department of Chemistry, University of North Texas, Denton, Texas, USA. *e-mail: rkohli@upenn.edu or andres@unt.edu

We examined the active site of human TET2 for potential structure–function determinants of stepwise oxidation. In the crystal structures of TET2 bound to DNA, the enzyme is truncated to the minimal regions necessary for catalytic activity (hTET2-CS, residues 1129–1936 Δ 1481–1843) (Fig. 1a)^{13,14}. The target nucleobase is everted out of the DNA duplex and occupies a tunnel-like space in the active site, with the 5-modified group pointing toward the α -KG analog and Fe(II) (Fig. 1b). Although the residues that form this tunnel have no obvious interaction with the 5-modified groups^{13,14}, we hypothesized that they could impact the progress of stepwise oxidation by hydrogen bonding or steric interactions. We therefore targeted two conserved residues located close to the 5-methyl group (Fig. 1a,b). By substituting all 20 amino acids at these positions, notably Thr1372, we uncovered a relationship between the side chain properties and stepwise oxidation activity, including variants that stall oxidation at hmC, with little to no fC or caC formed. Molecular dynamics simulations, coupled with biochemical analyses, revealed that a conserved Thr1372–Tyr1902 active site scaffold is required for efficient fC and caC formation, providing the first evidence that wild-type (WT) TET2 is specifically shaped to enable higher-order oxidation. We further show that mutations along this core scaffold can reconfigure active site interactions to stall oxidation at hmC, which opens opportunities to test the importance of hmC compared to fC and caC in biological and pathological systems.

RESULTS

Saturation mutagenesis at Thr1372

We interrogated the active site of human TET2 by performing saturation mutagenesis, which can comprehensively capture structure–function relationships at a particular residue. Using the hTET2-CS construct, we generated plasmids encoding all 20 natural amino acids at either the Thr1372 or Val1900 positions. The plasmids were transiently transfected into HEK293T cells, and genomic DNA (gDNA) was purified from the cells after 48 h. Using dot blotting to assess the qualitative pattern of genomic cytosine modifications, we found that the Val1900 position is fairly tolerant of mutation, with a variety of mutants showing WT-like stepwise oxidation or reduced overall activity, while bulky and charged residues largely inactivate the enzyme (Supplementary Results, Supplementary Fig. 1a).

We focused our attention on the Thr1372 mutants. TET2 overexpression was confirmed to be uniform by western blotting of cell lysates, with only T1372P having slightly reduced expression (Supplementary Fig. 1b). Dot blotting showed that, more so than at Val1900, mutations at Thr1372 produced distinctive patterns of cytosine oxidation, which cluster based on the biochemical properties of the side chain (Fig. 2a). Replacing Thr1372 with a proline, positively charged (H, K, R), or bulkier hydrophobic residue (I, F, L, M, W, Y) renders TET2 inactive. Only the T1372S mutant, which preserves the side chain hydroxyl group, exhibits WT-like activity. Smaller residues (A, C, G) are proficient at oxidation to fC and caC, but at reduced levels compared to WT. Most remarkably, the acidic or related polar residues (D, E, N, Q) and the nearly isosteric valine permit WT-like formation of hmC but no fC or caC, as detected by dot blot. Given this stalling of oxidation at hmC, Thr1372 appeared to play a unique role in stepwise oxidation.

Nucleoside LC-MS/MS quantifies range of mutant activity

We quantified the cellular activity of all Thr1372 mutants capable of oxidizing at least to hmC. The gDNA was degraded to component nucleosides and analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS) (Supplementary Fig. 2). In 0.1 μ g of HEK293T gDNA, limits of detection in the low femtomole range enabled reliable quantification of 1 in 10^3 – 10^4 of all cytosines. While the total abundance of modified cytosine bases (mC + ox-mCs) was similar across all conditions, the distribution of specific modifications differed significantly. In vector-transfected cells, ox-mC products are

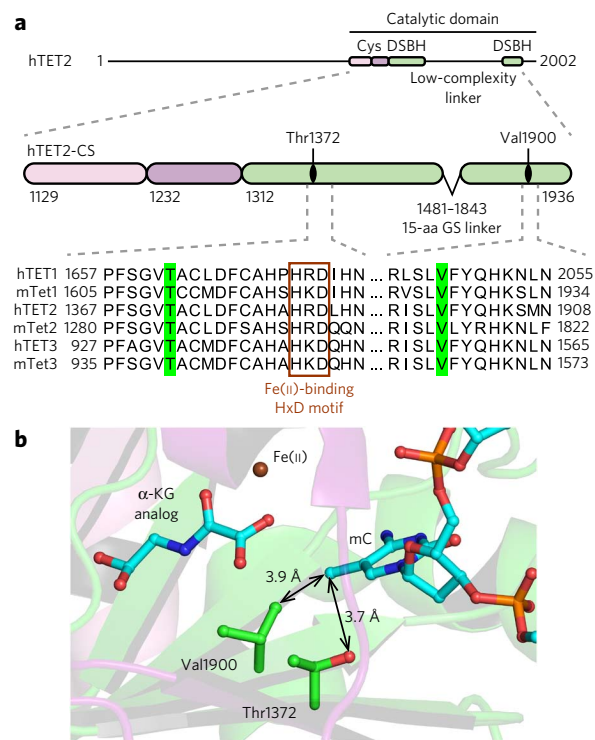


Figure 1 | Thr1372 and Val1900 were targeted for their potential role in TET2-catalyzed cytosine oxidation. (a) Schematic of the hTET2-CS construct (drawn to scale, adapted from ref. 14). The two cysteine-rich domains are shown in pink and purple, and the double-stranded β -helix (DSBH) domain is in green; residues are numbered as in the complete hTET2 protein. Both Thr1372 and Val1900 are conserved across mouse and human TET proteins. (b) Structure of the hTET2-CS active site (PDB 4NM6) highlighting the targets for mutagenesis, Thr1372 and Val1900.

The mC base flips into the active site pocket, pointing toward Fe(II) and the α -KG analog N-oxalylglycine. Shown are the nearest distances between the residues and the 5-methyl carbon.

minimal: $1.6 \pm 1.0\%$ of total cytosine modifications are hmC, with no fC or caC detected (Fig. 2b). Cells overexpressing WT hTET2-CS contain $15.2 \pm 2.8\%$ hmC, $6.0 \pm 1.9\%$ fC, and $5.7 \pm 1.8\%$ caC, demonstrating robust TET-dependent oxidation at a genomic level.

The mutants exhibit a gradient of activity reflected in the fraction of genomic ox-mC bases (Fig. 2b). T1372S is the only mutant with WT-like levels of fC and caC, and hmC levels are slightly higher than WT. T1372A/C/G mutants generate WT-like levels of hmC but only one-third to one-half as much fC and barely detectable caC. Further down the activity gradient, the E/Q/N/D/V mutants produce hmC at levels at least half of that of WT, but fC and caC are near or below detection limits, consistent with the dot blotting results. Among this group, T1372E appears to have the highest activity with WT-like hmC levels and $<1\%$ fC, while T1372V has the lowest, generating half as much hmC but no fC. Finally, the slightly bulkier T1372I mutant resembles the vector control, underscoring the steric constraints at this position. Thus, the LC-MS/MS results more clearly elucidated the patterns seen on dot blot, showing a spectrum of activity among the Thr1372 mutants correlating with the side chain properties, with E/Q/N/D/V mutants stalling oxidation at hmC.

Computational modeling reveals Thr1372–Tyr1902 scaffold

To probe potential mechanisms behind the mutants' effects, we turned to classical molecular dynamics (MD) simulations of all the active Thr1372 variants. We drew from our experience with AlkB^{21,22} to model WT hTET2-CS and the Thr1372 mutants bound to each of the four cytosine derivatives (see Supplementary Figs. 5–9

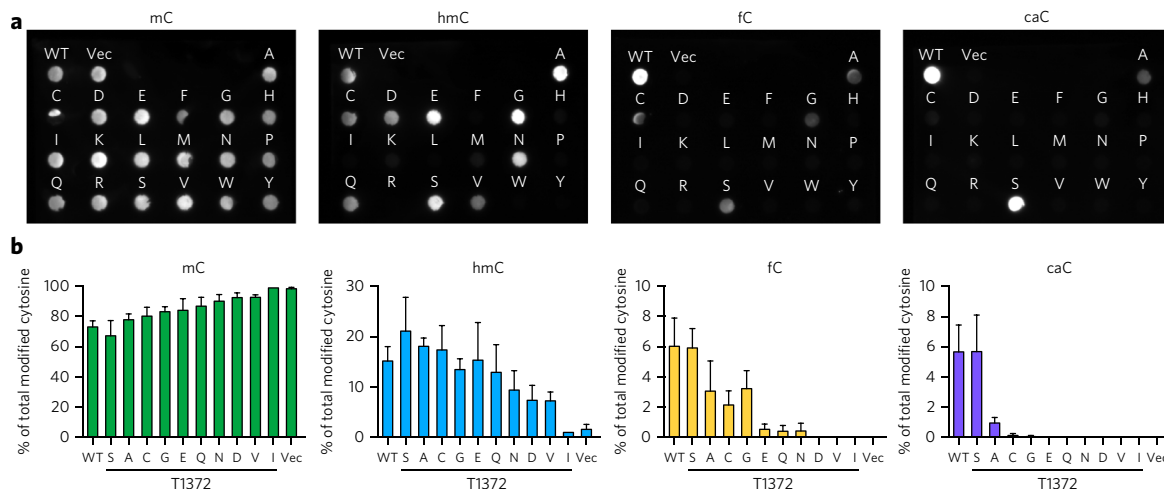


Figure 2 | Screen for mutant activity. (a) Dot blots for mC, hmC, fC, and caC in 400 ng of genomic DNA isolated from transfected HEK293T cells. DNA from cells transfected with WT hTET2-CS or empty vector (vec) was spotted first, followed by the Thr1372 mutants in alphabetical order (uncropped image in **Supplementary Fig. 10a**). Further analysis of mutant phenotypes focused on variants that were capable of oxidation at least to hmC. (b) Genomic levels of mC, hmC, fC, and caC modifications produced by catalytically active Thr1372 mutants, quantified by LC-MS/MS as the percent of total C modifications. Mutants are approximately presented in decreasing order of activity, from WT-like T1372S, to A/C/G that form highly oxidized bases at reduced levels, to E/Q/N/D/V that largely stall at hmC. Shown are the mean and s.d. from independent experiments (WT, $n = 7$; vec, $n = 6$; mutants, $n = 3$; T1372I, $n = 2$).

and **Supplementary Tables 2–8** for details). Our simulations were based on the crystal structure of TET2 in complex with DNA containing mC (PDB 4NM6) (ref. 14), using α -KG and an Fe(II) surrogate (Mg(II)). Our WT models with hmC and fC proved mostly consistent with the more recently published structures of TET2 with these bases¹³; we observe all the key interactions between the enzyme, α -KG, active site metal ion, and DNA substrate for varying durations across our simulations. Furthermore, energy decomposition analysis (EDA) and the root-mean-square (r.m.s.) deviation comparing the simulations to the reference crystal structure show that the cytosine bases stably occupy the active site across time in all our models.

The hmC models in particular revealed distinct patterns of active site interactions in WT, A/C/G, and E/Q/N/D/V mutants, consistent with hmC being the fulcrum of the observed stalling effect. These patterns helped us to define a key structural scaffold in the WT enzyme that is required for efficient stepwise oxidation. This WT active site scaffold consists of a Thr1372–Tyr1902 hydrogen bond that critically supports optimal non-bonded interactions between Tyr1902 and the substrate cytosine base (**Fig. 3a**). The Thr1372–Tyr1902 hydrogen bond is observed in 65% of the simulation time (average over five runs of 50 ns each), and the total non-bonded interaction energy between these residues is -3.37 kcal/mol (**Fig. 3b**). Tyr1902, thus oriented by Thr1372, shows significant non-bonded interaction with the hmC base (-6.10 kcal/mol). This core scaffold is present across all WT models bound to mC, hmC, fC, and caC and remains fully intact in the T1372S mutant, consistent with this mutant's WT-like activity in cells.

All the other mutants eliminate the Thr1372–Tyr1902 hydrogen bond, perturbing the interaction between Y1902 and the substrate base, with a corresponding loss of enzymatic activity. For the A/C/G mutants, loss of the Thr1372–Tyr1902 scaffold appears to weaken interactions between misaligned active site components, as exemplified by T1372A (**Fig. 3a,b**). Combined with the gDNA results, we term the A/C/G phenotype 'low efficiency', since these mutants permit higher-order oxidation but at reduced levels compared to WT.

In our modeling, the E/Q/N/D/V mutants go a step further; they not only eliminate the Thr1372–Tyr1902 scaffold but also elicit new hydrogen bonds specifically with hmC. These new interactions, not present in WT models, position hmC in a different orientation relative to Tyr1902 (**Fig. 3a,b**). For instance, in T1372E, the

Glu1372 hydrogen bonds directly with the 5-hydroxymethyl group for 88% of the simulation time (average over two runs of 50 ns each). Direct hydrogen bonding to hmC is also observed in T1372D and T1372Q, whereas in T1372N and T1372V, the new hydrogen bond is between hmC and other nearby residues (**Supplementary Fig. 5; Supplementary Tables 3b and 4b**). For example, T1372V elicits an hmC–Asp1384 hydrogen bond (38% of simulation time, average over two runs of 50 ns each). We suggest that the loss of the Thr1372–Tyr1902 scaffold, together with new interactions specific to hmC, could contribute to the unique stalling phenotype of T1372E/Q/N/D/V mutants, which we term 'hmC-dominant'.

Biochemical characterization of TET2 variants

With results from cells and MD showing that side chain properties can define WT, low-efficiency, and hmC-dominant phenotypes, we subjected the TET variants to rigorous comparison *in vitro*. We first used driving conditions to compare the maximum extent of the variants' activity and then used limiting conditions to compare the reactivity on mC versus hmC. Representative hTET2-CS variants—WT and T1372S, T1372A, T1372E, and T1372V—were expressed and purified from Sf9 insect cells (**Supplementary Fig. 3a**). To drive oxidation forward, we reacted excess enzyme with limiting substrate: 27-bp oligonucleotides containing a single reactive mC, hmC, or fC duplexed to an unmodified complementary strand. The reaction products were quantified by LC-MS/MS and the results corroborated by three complementary, chemoenzymatic assays (**Supplementary Fig. 3b,c**).

In reactions with 20 nM mC-containing duplexes, 30 μ g/ml (maximally 0.57 μ M) of WT, T1372S, or T1372A convert nearly all substrate to oxidized products in 30 min (**Fig. 4a**). However, while WT and T1372S advance efficiently through stepwise oxidation, turning over $\sim 93\%$ of substrate to fC and caC, T1372A lags behind, forming predominantly hmC (30%) and fC (54%), and only 13% caC. This aligns with the gDNA and modeling results, indicating that low-efficiency mutants are capable of oxidation to caC but at reduced levels compared to WT.

The hmC-dominant T1372E and T1372V mutants show noticeably reduced activity on mC (54% and 76% of mC substrate remaining, respectively), and oxidation products are strongly restricted to hmC, with 4% and 1% conversion to fC, respectively (**Fig. 4a**).

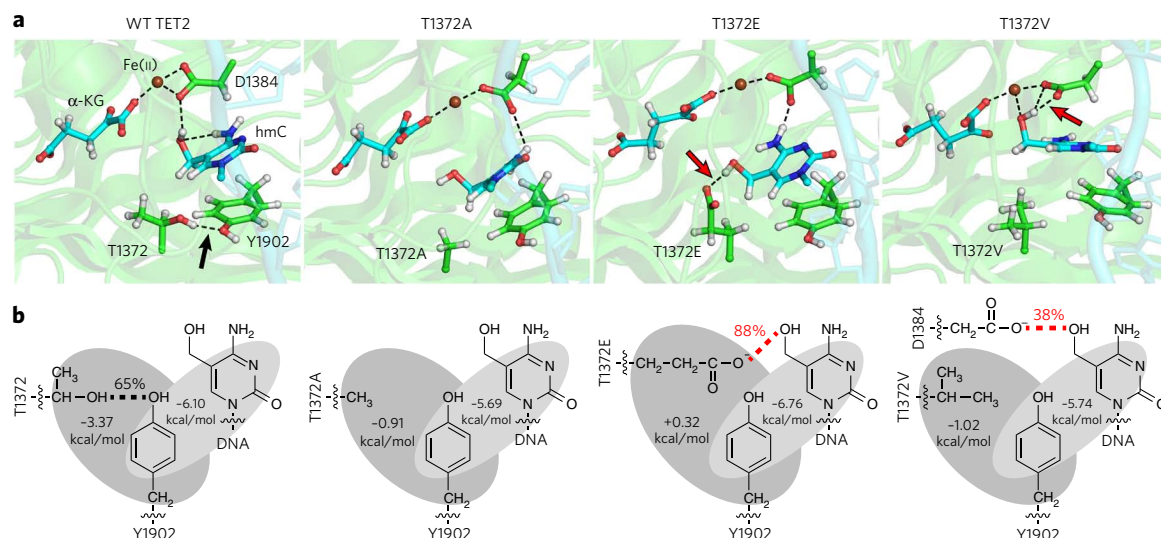


Figure 3 | Molecular dynamics modeling reveals a critical Thr1372-Tyr1902 scaffold that is disrupted in the low-efficiency and hmC-dominant mutants. (a) Selected snapshots from MD simulations highlighting key active site components and hydrogen bonds. In WT enzyme (and T1372S), Thr1372 forms a hydrogen bond (black arrow) with Tyr1902, which orients Tyr1902 for optimal non-bonded interactions with the substrate. Low-efficiency mutants such as T1372A disrupt this scaffold, while hmC-dominant mutants such as T1372E and T1372V not only disrupt the scaffold but also elicit new hydrogen bonds (red arrows) with the 5-hydroxymethyl group of hmC. (b) Simplified scheme of interactions between key residues and hmC, as determined by MD. Hydrogen bonds (dashed lines) are quantified as percentage of simulation time observed. The values are an average over 2–5 simulation runs of 50 ns each (see Online Methods). Non-bonded interactions are indicated in gray, and total energies of interaction are given in kcal/mol. (Additional modeling data in **Supplementary Figs. 5–9** and **Supplementary Tables 2–8**.)

Compared to the gDNA results, where the levels of hmC produced by the E/V mutants are within two-fold of WT (**Fig. 2b**), this indicates that other factors can likely tune the activity of TET2 and/or the levels of hmC in cells. Importantly, the patterns of oxidation and hmC stalling hold true in cells and *in vitro*. T1372E is observed to be slightly more active than T1372V, consistent with the gDNA results, which suggests a trade-off between more hmC production and better stringency of stalling. Time-course analysis further demonstrates that overall reactivity on mC decreases from WT to the low-efficiency T1372A, and the hmC-dominant E/V mutants fail to produce significant fC even after 3 h (**Fig. 4b**; **Supplementary Fig. 3d**). To validate that the hmC-dominant phenotype is not restricted to the truncated CS form of the protein, we also generated the T1372E mutation in the full catalytic domain of TET2 (hTET2-FCD, residues 1129–2002) and noted similar results (**Fig. 4a**).

When all available substrate is hmC, WT and T1372S again convert >93% of substrate to fC and caC. T1372A produces 65% fC and caC, while T1372E and T1372V are able to produce only 8% and 3% fC, respectively. When starting with fC substrate under the same conditions, WT enzymes convert about half of fC to caC, corroborating that the final step of oxidation is the least efficient^{13,23}. T1372A generates 19% caC, ~1/3 of the WT level, while E/V mutants make <3% caC, near or below the detection limits of our assays. These results strongly support our model that the Thr1372-Tyr1902 scaffold is required for WT TET2 activity. Loss of the active site scaffold decreases the activity of low-efficiency mutants and has a more severe effect on hmC-dominant mutants, which do not make significant fC or caC even under driving reaction conditions.

Since TET2 is known to prefer mC over hmC, we next turned to enzyme-limiting conditions to distinguish whether the decrease in overall activity alone was sufficient to explain the restriction of oxidation products to hmC. We compared the reactivity of WT, T1372A, and T1372E mutants on mC compared to hmC by titrating enzyme against 745-bp substrates fully modified with mC or hmC. We chose to simplify our kinetic analysis to measure total oxidation products (i.e., substrate consumed), since iterative oxidation links the kinetics of each oxidation step in ways not easily dissected²⁰.

By this analysis, WT TET2 consumes 2.9 ± 0.2 nmol of mC substrate per milligram enzyme per minute, while activity on hmC decreases 2.6-fold to 1.1 ± 0.1 nmol/mg/min (**Table 1**; **Supplementary Fig. 4**). This mild decrease in activity on hmC is consistent with previously published observations¹³. The T1372A mutant displays similar activity on mC and is only 5.5-fold slower in hmC-to-fC conversion, in line with this mutant's capacity for less efficient higher-order oxidation. By contrast, relative to the most proficient WT reaction, the T1372E mutant is 5.9-fold slower in mC-to-hmC conversion but 48-fold slower in hmC-to-fC conversion. Thus, the hmC-dominant mutant exhibits decreased activity overall, but the usual mild preference for mC substrate is not sufficient to explain the larger loss of activity on hmC, which underlies the stalling effect.

Tyr1902 mutagenesis strongly supports our model

Our MD simulations suggested that active site scaffold mutations could introduce aberrant interactions that contribute to hmC stalling. We were cognizant of the challenges to modeling new interactions with classical MD and therefore subjected this model to an independent test: mutating the other scaffold residue, Tyr1902, to phenylalanine. Our modeling predicts that Y1902F would liberate Thr1372 to form a hydrogen bond directly with hmC (18% of simulation time, average over two runs of 50 ns each), potentially favoring an hmC-dominant phenotype (**Fig. 5a**). Taking the hypothesis one step further, by adding a T1372A mutation to Y1902F, our modeling predicts that the T1372A/Y1902F double mutant could rescue activity by alleviating the aberrant hydrogen bonding interaction.

To test these predictions, we compared the activities of purified T1372A, Y1902F, and T1372A/Y1902F enzymes *in vitro*. The results strikingly confirmed our predictions. Compared to the WT mC-to-hmC reaction, the Y1902F single mutant is 9.9-fold slower at mC-to-hmC conversion and 36-fold slower at hmC-to-fC conversion (**Table 1**; **Supplementary Fig. 4**). Addition of the second T1372A mutation partially restores activity so that the double mutant is only 2.8-fold slower at mC-to-hmC conversion and 14-fold slower at hmC-to-fC conversion. Under driving conditions, the Y1902F mutant leaves 38% of mC substrate unreacted, with products

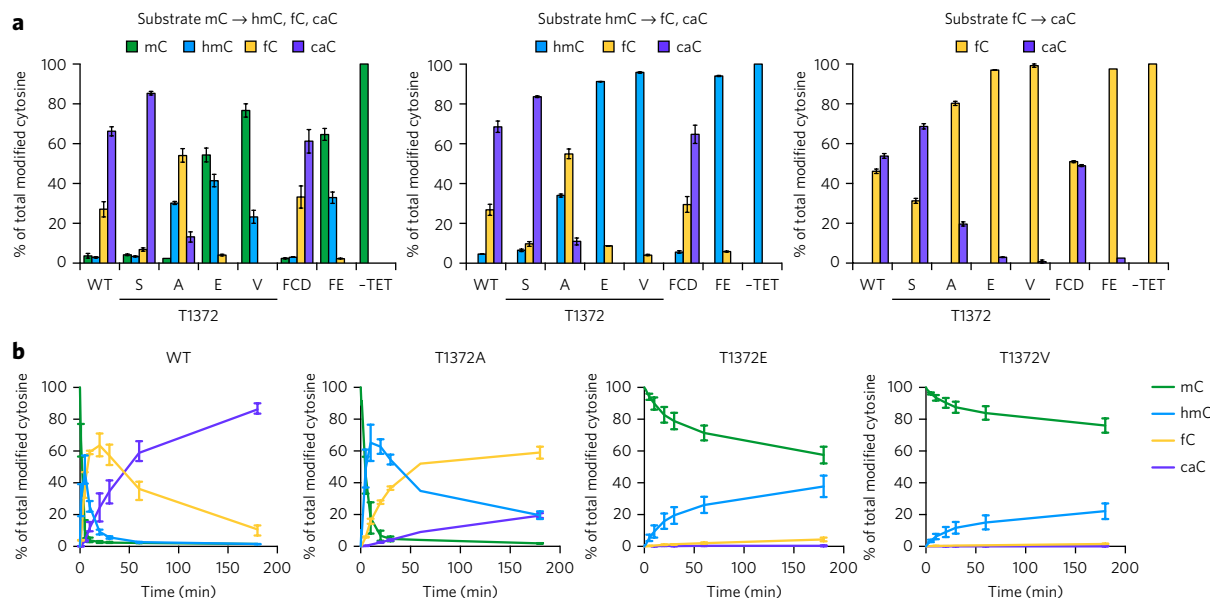


Figure 4 | Biochemical characterization of purified hTET2 mutants. (a) TET2 variants (30 $\mu\text{g}/\text{ml}$) were reacted with 20 nM double-stranded DNA substrates containing mC, hmC, or fC for 30 min. The reaction products were purified, degraded to nucleosides, and quantified by LC-MS/MS. WT and T1372E were also generated in the full catalytic domain of TET2 (FCD and FE, respectively) to confirm that the phenotypes are the same as in the hTET2-CS constructs. Mean values are plotted ($n = 2$), and error bars represent the range. (b) Time course for reactions of 30 $\mu\text{g}/\text{ml}$ purified TET2 on 25 nM mC substrates. Mean values are plotted (WT, $n = 3$; mutants, $n = 2$), and error bars represent the range.

consisting of 49% hmC, 13% fC, and no caC (Fig. 5b)—similar to T1372E/V but with less stringent stalling at hmC. The introduction of a second mutation in the T1372A/Y1902F double mutant rescues activity, such that 97% of mC substrate is consumed, like in the T1372A single mutant.

To complement these LC-MS/MS results, rather than digesting the reaction products to nucleosides, we treated the intact oligonucleotides with purified TDG followed by DNA gel electrophoresis to differentiate strands containing mC or hmC from strands containing fC or caC (Fig. 5c). While Y1902F shows only trace generation of fC and caC, the addition of the second mutation in T1372A/Y1902F restores stepwise oxidation and mirrors the results for T1372A. Thus, our structural modeling correctly predicts the biochemical behavior of the Y1902F and T1372A/Y1902F mutants, strongly supporting both the requirement of the Thr1372–Tyr1902 scaffold for WT stepwise oxidation and the contribution of aberrant active site interactions to the hmC-dominant phenotype.

DISCUSSION

TET-catalyzed stepwise oxidation populates the mammalian epigenome with three ox-mC bases, making it critical to dissect how each individual base forms and functions. Previous studies have elucidated various biases in favor of the first oxidation step, mC-to-hmC conversion, implying that TET enzymes may be primarily adapted for making hmC, with fC and caC as rare oxidative ‘overflow’ products. However, in light of evidence for the importance of fC and caC in active DNA demethylation and as stable epigenetic marks, we asked whether TET enzymes bear structural features that specifically support fC and caC formation. We have now shown that a conserved Thr1372–Tyr1902 active site scaffold is required for

efficient higher-order oxidation by human TET2, suggesting that the enzyme is shaped to enable production of not only hmC but fC and caC as well. We further uncover Thr1372 mutations that effectively abrogate higher-order oxidation by disrupting the active site scaffold; these are the first human TET variants that dissociate the steps of oxidation, providing a new tool to directly test the functions of hmC versus fC and caC.

As a structure–function determinant in TET2, the Thr1372–Tyr1902 scaffold invites comparison to known TET homologs. The Thr–Tyr pair is perfectly conserved across mouse and human TET1, TET2, and TET3 (Supplementary Fig. 1c), raising the possibility that corresponding mutations in TET1 and TET3 could likewise tune TET activity. Notably, while a large number of TET mutations have been identified in various malignancies^{24,25}, we are not aware of any mutations at the Thr1372 or Tyr1902 positions. As an example in more distant homologs, the trypanosomal J-binding protein JBP1 is predicted to have a Thr–Tyr pair while JBP2 has Ser–Tyr (Supplementary Fig. 1c). JBP1 and JBP2 are capable of oxidizing thymine to both 5-hydroxymethyluracil (hmU) and 5-formyluracil, though a glucosyltransferase normally diverts hmU to form base J as part of the trypanosome’s mechanism for immune evasion^{26,27}.

A particularly intriguing exception to the conserved Thr1372–Tyr1902 scaffold is the *Naegleria gruberi* Tet-like protein NgTet1, which is also capable of higher-order oxidation. Using a structure-based algorithm²⁸ recently borne out by crystal structures²³, we found that Thr1372 and Tyr1902 align with Ala212 and Phe295 in NgTet1, respectively (Supplementary Fig. 1c), making NgTet1 analogous to our T1372A/Y1902F double mutant. In NgTet1, it was proposed that these residues form a hydrophobic pocket that accommodates hmC as it rotates from a product to a substrate conformation for

Table 1 | Activity of representative TET2 variants on mC and hmC.

Substrate consumed (nmol/mg/min)	WT	T1372A	T1372E	Y1902F	T1372A/Y1902F
mC	2.9 \pm 0.2	2.9 \pm 0.1	0.48 \pm 0.02	0.29 \pm 0.03	1.0 \pm 0.1
hmC	1.1 \pm 0.1	0.51 \pm 0.03	0.059 \pm 0.006	0.079 \pm 0.025	0.20 \pm 0.02

Values are mean \pm s.e.m. from three independent experiments.

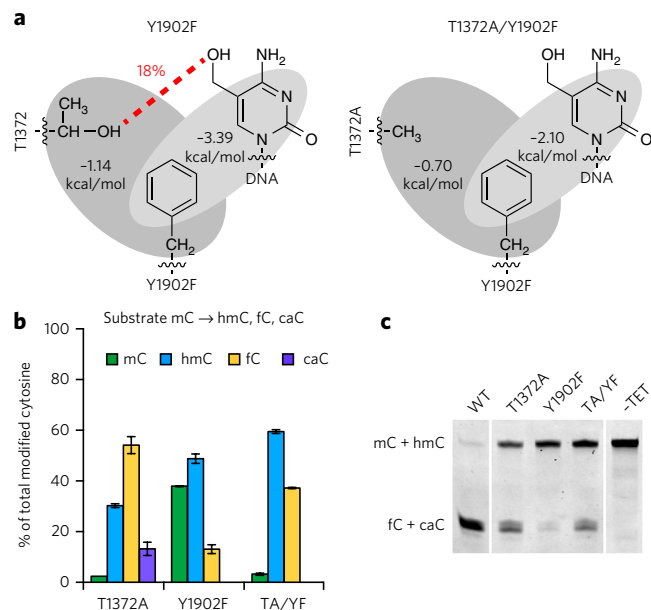


Figure 5 | T1372A/Y1902F double mutant rescues the hmC-dominant phenotype by configuring active site interactions. (a) Our modeling predicts that in the Y1902F single mutant, Thr1372 would hydrogen bond instead with hmC, producing an hmC-dominant phenotype. Addition of a T1372A mutation to Y1902F would remove hydrogen bonding, which is predicted to restore activity. The values shown are an average over 2–3 simulation runs of 50 ns each (see Online Methods). (b) Reaction of 30 μ g/ml purified mutants on 20 nM mC substrate, analyzed by LC-MS/MS. Mean values are plotted ($n = 2$), and error bars represent the range. As predicted by our model, Y1902F mimics hmC-dominant mutants, with relatively low activity on mC and little fC formed. The double mutant (TA/YF) restores activity to resemble the T1372A single mutant. (c) To highlight fC and caC in the reaction products, the purified oligos were treated with recombinant TDG. After alkaline-mediated cleavage at the resulting abasic sites, denaturing PAGE was used to separate intact oligos containing mC and hmC from cleaved oligos that contained fC and caC (uncropped image in **Supplementary Fig. 10b**).

further oxidation²³. The active site mutations A212V/N, which could sterically hinder hmC binding within this pocket, were found to partially stall hmC oxidation. In human TET2, the Ala–Phe double mutant only permits low-efficiency stepwise oxidation, suggesting that the Thr–Tyr dyad may have evolved to fine-tune efficient fC and caC generation. By leveraging this scaffold, our results offer the first variants that produce distinct stepwise oxidation patterns in human TET enzymes.

Our combined computational and biochemical approach shows how T1372E/Q/N/D/V mutants could reconfigure active site interactions to produce the hmC-dominant phenotype, characterized by moderate loss of overall catalytic activity as well as a specific decrement at the hmC-to-fC step. To account for the additional loss of activity on hmC, our modeling most prominently implicates new hydrogen bonding to hmC in these mutants. Our calculations correctly predict the hmC-dominant behavior of the Y1902F single mutant, as well as rescued activity in the T1372A/Y1902F double mutant. Indeed, it is quite unusual that the addition of a second mutation rescues activity of the first, helping to bolster our mechanistic model. We note, however, that other, related mechanisms could also play a role and are not mutually exclusive with this model. Such mechanisms include restriction of substrate–product rotation^{15,23}, changes in protein dynamics (**Supplementary Fig. 9**), and/or altered accessibility of the active site to DNA and α -KG—all of which could occur in concert with aberrant hydrogen bonding

to hmC. These possibilities reflect the complex dynamics of TET–DNA interactions, which remain priorities for future research. Importantly, independent of the mechanism of action, the hmC-dominant Thr1372 mutants fill the need for experimental tools to dissect the individual steps of mC oxidation.

These new TET variants potentially allow for the first direct studies of the epigenetic functions of hmC as distinct from fC and caC. Until now, functional studies have by necessity been all-or-none, showing that loss of one or more TET isozymes can produce diverse phenotypes. Examples range from inability of TET triple-knockout mouse embryonic fibroblasts to undergo reprogramming²⁹ to cancer cell proliferation with loss of TET1 or TET2 (refs. 30,31) to neonatal lethality in TET3-deleted mice³², among others. In many cases, reintroduction of a single active TET isozyme can fully rescue the phenotype. Such systems provide ideal opportunities to introduce low-efficiency and hmC-dominant TET variants to probe whether hmC alone is sufficient to rescue the defect, whether fC and caC are required, or whether interacting enzymes such as TDG are actually the key players.

These *in vivo* applications will bring new challenges as well, such as examining the mutants' activity under more physiological conditions. Our study illustrates one limit to predicting cellular outcomes based on biochemical properties: although all the mutants perturb the enzyme's reactivity *in vitro*, in HEK293T cells the amount of hmC generated can be close to WT. Many explanations are possible, including that HEK293T cell overexpression likely represents a non-steady-state system, in which all cytosine modifications reach unusually high levels with limited means of removing these marks. It will be interesting to see whether normal cells, expressing endogenous TET enzymes, maintain a homeostatic level of ox-mC bases. It will also be important to determine whether the mutant phenotypes in TET2 translate to other TET isoforms, which is needed both for applying the mutants in various biological systems and for helping to address whether TET1, TET2, and TET3 have similar or distinct mechanisms of action. Finally, given recently reported structures of TET2 bound to ox-mC bases¹³, we envision that chemical biology approaches, including additional mutagenesis or unnatural modifications along the Thr1372–Tyr1902–cytosine scaffold, could further hone selectivity for particular bases and potentially uncover TET variants that stringently stall at fC as well or accelerate conversion to caC.

Received 17 December 2015; accepted 29 September 2016; published online 5 December 2016

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

References

- Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
- Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Edn Engl.* **50**, 7008–7012 (2011).
- He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
- Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
- Kohli, R.M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479 (2013).
- Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).
- Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).

10. Wu, H. & Zhang, Y. Charting oxidized methylcytosines at base resolution. *Nat. Struct. Mol. Biol.* **22**, 656–661 (2015).
11. Liu, M.Y., DeNizio, J.E., Schutsky, E.K. & Kohli, R.M. The expanding scope and impact of epigenetic cytosine modifications. *Curr. Opin. Chem. Biol.* **33**, 67–73 (2016).
12. Zheng, G., Fu, Y. & He, C. Nucleic acid oxidation in DNA damage repair and epigenetics. *Chem. Rev.* **114**, 4602–4620 (2014).
13. Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118–122 (2015).
14. Hu, L. *et al.* Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545–1555 (2013).
15. Lu, J. *et al.* A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2). *Phys. Chem. Chem. Phys.* **18**, 4728–4738 (2016).
16. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
17. Weber, A.R. *et al.* Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* **7**, 10806 (2016).
18. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
19. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
20. Crawford, D.J. *et al.* Tet2 catalyzes stepwise 5-methylcytosine oxidation by an iterative and *de novo* mechanism. *J. Am. Chem. Soc.* **138**, 730–733 (2016).
21. Fang, D., Lord, R.L. & Cisneros, G.A. Ab initio QM/MM calculations show an intersystem crossing in the hydrogen abstraction step in dealkylation catalyzed by AlkB. *J. Phys. Chem. B* **117**, 6410–6420 (2013).
22. Fang, D. & Cisneros, G.A. Alternative pathway for the reaction catalyzed by DNA dealkylase AlkB from ab initio QM/MM calculations. *J. Chem. Theory Comput.* **10**, 5136–5148 (2014).
23. Hashimoto, H. *et al.* Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. *Nucleic Acids Res.* **43**, 10713–10721 (2015).
24. Abdel-Wahab, O. *et al.* Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144–147 (2009).
25. Scourzac, L., Mouly, E. & Bernard, O.A. TET proteins and the control of cytosine demethylation in cancer. *Genome Med.* **7**, 9 (2015).
26. Cliffe, L.J. *et al.* JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res.* **37**, 1452–1462 (2009).
27. Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y. & Sabatini, R. Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *J. Biol. Chem.* **289**, 20273–20282 (2014).
28. Pei, J., Kim, B.H. & Grishin, N.V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
29. Hu, X. *et al.* Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14**, 512–522 (2014).
30. Lian, C.G. *et al.* Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* **150**, 1135–1146 (2012).
31. Neri, F. *et al.* TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway. *Oncogene* **34**, 4168–4176 (2015).
32. Gu, T.P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).

Acknowledgments

We thank B. Niedziolka and the Wistar Institute Protein Expression Facility for help with protein expression in Sf9 cells and all members of our labs for insightful discussions. Computing time from Wayne State C&IT and additional mass spectrometry resources from I. Blair's lab are gratefully acknowledged. This work was supported by the Rita Allen Foundation Scholar Award to R.M.K. and by NIH grants (R01 GM110174 to B.A.G., R01 GM108583 to G.A.C., and F30 CA196097 to M.Y.L.).

Author contributions

R.M.K., G.A.C., M.Y.L., and H.T. conceived the experiments. R.M.K., M.Y.L., D.J.C., J.E.D., X.-J.C., and B.A.G. were involved in the design and optimization of biochemical and cellular experiments, which were performed and analyzed by M.Y.L., J.E.D., and R.M.K. For MD simulations, G.A.C., H.T., M.Y.L., and R.M.K. were involved in the design of experiments, which were performed and analyzed by H.T. and G.A.C. The manuscript was written by M.Y.L., R.M.K., H.T., and G.A.C. and edited by all authors.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to R.M.K. or G.A.C.

ONLINE METHODS

Saturation cassette mutagenesis. A codon-optimized hTET2-CS construct (residues 1129–1936 Δ 1481–1843) was designed with an N-terminal FLAG tag and unique restriction sites flanking the Thr1372 and Val1900 codons, purchased as a gene block from Integrated DNA Technologies (IDT), and cloned into a pLEXm vector for mammalian expression. Thirty-eight pairs of complementary oligos encoding all amino acid substitutions at both positions (as well as the Y1902F mutation) were ordered, annealed, and cloned by cassette mutagenesis in place of the WT sequence (**Supplementary Table 1**). Mutations were confirmed by gene sequencing and/or digestion at a unique restriction site within the oligo.

TET2 overexpression in HEK293T cells. HEK293T cells (mycoplasma tested and verified by ATCC) were cultured in DMEM with GlutaMAX (Thermo Fisher Scientific) and 10% FBS (Sigma). Cells were transfected with WT or mutant hTET2-CS or an empty vector control using Lipofectamine 2000 (Thermo) according to the manufacturer's protocol. Media was changed 24 h after transfection, cells were harvested by trypsinization 48 h after transfection and resuspended in PBS, and genomic DNA (gDNA) was purified from four fifths of the collected cells using the DNeasy Blood & Tissue Kit (Qiagen).

Western blot for FLAG-tagged hTET2-CS. One-fifth portion of the transfected cells was lysed using CytoBuster Protein Extraction Reagent (EMD Millipore). The clarified lysates were diluted 50-fold into CytoBuster and run on two 8% SDS–PAGE gels, with WT sample as a standard on each gel. To further standardize the blots, the gels were cut at the 70-kDa marker, so that the upper half contained the Hsp90 control band and the bottom half hTET2-CS. The Hsp90 halves of both gels were transferred together onto a single PVDF membrane, and the two TET halves were transferred onto another membrane using an iBlot Gel Transfer Device (Thermo). Membranes were blocked for 2 h at room temperature with 5% (w/v) milk in Tris-buffered saline with 0.1% (v/v) Tween-20 (TBST), washed 3 \times with TBST, blotted with primary 1:10,000 anti-FLAG M2 (Sigma; cat. no. F1804) or 1:1,000 anti-Hsp90 α/β (Santa Cruz Biotechnology; cat. no. sc-131119) antibodies at 4 °C overnight, washed, blotted with secondary 1:5,000 goat anti-mouse-HRP (Santa Cruz Biotechnology; cat. no. sc-2005) for 2 h, washed, and imaged with Immobilon Western Chemiluminescent HRP Substrate (Millipore) on a Fujifilm LAS-1000 imager with 30-s exposures.

Dot blot for cytosine modifications in gDNA. Purified gDNA from HEK293T cells was diluted to 10 ng/ μ l in Tris–EDTA (TE) buffer, pH 8.0. To this was added $\frac{1}{4}$ volume of 2 M NaOH–50 mM EDTA. The DNA was denatured for 10 min at 95 °C and transferred quickly to ice, followed by addition of 1:1 ice cold 2 M ammonium acetate. Sequi-Blot PVDF membranes (Bio-Rad) were cut to size, wet with MeOH and equilibrated in TE buffer, then assembled into a 96-well Bio-Dot microfiltration apparatus (Bio-Rad). Each well was washed with 400 μ l TE drawn through with gentle vacuum, and 400 ng of gDNA was loaded, followed by another TE wash. Membranes were blocked for 2 h in 5% milk–TBST, washed 3 \times with TBST, and blotted at 4 °C overnight with primary antibodies against each modified cytosine (Active Motif)—1:5,000 mouse anti-mC (cat. no. 39649); 1:10,000 rabbit anti-hmC (cat. no. 39769); 1:5,000 rabbit anti-fC (cat. no. 61223); 1:5,000 rabbit anti-caC (cat. no. 61225). Blots were then washed, incubated with secondary 1:2,000 goat anti-mouse-HRP or 1:5,000 goat anti-rabbit-HRP (Santa Cruz Biotechnology, cat. no. sc-2004) for 2 h, washed, and imaged as described above.

Nano LC–MS/MS analysis of gDNA. Based on published protocols³³, we adapted and optimized LC–MS/MS methods for our systems. To quantify genomic levels of cytosine modifications in HEK293T cells, 20 μ g of purified gDNA was concentrated by ethanol precipitation and degraded to component nucleosides with 20 U DNA Degradase Plus (Zymo) in 20 μ l at 37 °C overnight. A 150 μ m \times 17 cm pre-column and 100 μ m \times 26 cm analytical reverse phase column were made from fused-silica tubing (New Objective) with a Kasil frit: The column was dipped into a 1:3 formamide:Kasil 1624 potassium silicate solution (PQ Corporation), polymerized at 100 °C overnight and trimmed to \sim 3 mm. Using a pressure injection cell, the columns were packed with Supelcosil LC–18-S resin (Sigma). Using this column setup equilibrated

in Buffer A1 (0.1% formic acid in H₂O), the nucleoside mixture was diluted ten-fold into 0.1% formic acid, and 1 μ l was injected onto an EASY-nLC 1000 (Thermo) nano LC. The sample was desalted for 5 min over the pre-column, nucleosides resolved using a gradient of 0–30% of Buffer B1 (0.1% formic acid in acetonitrile) over 30 min at a flow rate of 600 nl/min, and tandem MS/MS performed by positive ion mode electrospray ionization on a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo), with a spray voltage of 2.9 kV, capillary temperature of 275 °C, and normalized collision energy of 30%. Mass transitions were mC 242.11 \rightarrow 126.066 *m/z*; hmC 258.11 \rightarrow 124.051; fC 256.09 \rightarrow 140.046; caC 272.09 \rightarrow 156.041; and T 243.10 \rightarrow 127.050. Standard curves were generated from standard nucleosides (Berry & Associates) ranging from 10 μ M to 5 nM (10 pmol to 5 fmol total) (**Supplementary Fig. 2**). The sample peak areas were fit to the standard curve to determine amounts of each modified cytosine in the gDNA sample and expressed as the percent of total cytosine modifications in each sample.

Molecular dynamics simulations. Forty-four molecular dynamics (MD) simulations were carried out on WT and all experimentally tested mutants (T1372S/C/A/E/Q/N/D/V, Y1902F, T1372A/Y1902F) with all four cytosine derivatives (mC, hmC, fC, caC), α -KG, and Fe(II)/Mg(II) (see **Supplementary Figs. 5–9** and **Supplementary Tables 2–8** for details). All structures were modeled based on WT hTET2-CS bound to mC-containing DNA (PDB 4NM6) (ref. 14). Initially, the PDB structure was evaluated with MOLPROBITY³⁴ to check all possible rotamers, followed by hydrogen atom addition to every system with the Leap program³⁵ using the ff99SB parameter set³⁶ and solvation in a truncated octahedral box of TIP3P water³⁷. In addition, protonation states of titratable residues were tested with PropKa 3.0 (refs. 38–40), which confirmed that the default ionization at pH 7 was correct for all residues. Both coordinated histidines are protonated on ND1. All systems were explicitly neutralized with potassium counterions, which were added to the system using the Leap program. The final system size was \sim 60,000 total atoms with 17–21 counterions. All structures were minimized with 3,000 steps of conjugate gradient, followed by gradual warm-up to 300 K using Langevin dynamics with a collision frequency of 1.0 ps^{–1} in the NVT (particle number, volume, and temperature; canonical) ensemble for 100 ps. All simulations were performed with the GPU version of the pmemd program in AMBER12 (ref. 36). The force-field parameters for all cytosine derivatives (developed in house), α -KG, Fe(II)/Mg(II), and Zn that are not available in the default ff99SB set are provided in **Supplementary Data Set 1**. The iron cation was approximated by using Mg(II) parameters based on the precedent established by our previous studies on AlkB^{21,22}; this approximation was also validated again for our systems (**Supplementary Fig. 7**; **Supplementary Tables 3, 4** and **8**) (refs. 41,42).

Once the systems achieved the target temperature, production MD simulations were performed using Langevin dynamics with a collision frequency of 1.0 ps^{–1} in the NPT (particle number, pressure, and temperature; isothermal–isobaric) ensemble with the Berendsen barostat using a 2-ps relaxation time at 300 K. The production length for each of the simulations was 50 ns, and snapshots were saved every 10 ps; all snapshots were subjected to subsequent analysis (see below). Values reported are generally a time average over calculations from all snapshots. The most relevant simulations were performed 2–5 times for 50 ns each, with the results averaged across all simulations (the number of simulations for each system is denoted in **Supplementary Table 2**). All systems were simulated using the Amberff99SB force field with a 1-fs step size and a 9-Å cutoff for non-bonded interactions. SHAKE was used for all the simulations, and the smooth particle mesh Ewald (PME) method⁴³ was employed to treat long-range Coulomb interactions. Hydrogen bond, root mean square deviation (r.m.s. deviation), and distance analysis on trajectories were carried out using the CPPTRAJ module⁴⁴ available in the AMBER 12 suite, and the trajectories were visualized with the VMD program⁴⁵. Hydrogen bond analysis criteria were (1) angles over 120 degrees and (2) O–H distances less than 3 Å (default cpptraj settings). r.m.s. deviation and distance analysis are presented in **Supplementary Figures 7** and **8**.

Additional analyses to investigate intermolecular interactions in the active site were carried out by non-covalent interaction analysis (NCI) and energy decomposition analysis (EDA). NCI is a visualization tool to identify non-covalent interactions between molecules⁴⁶. The results obtained from the

NCI analysis consist of surfaces between the interacting molecules. These surfaces are assigned specific colors to denote the strength and characteristic of the interactions: green surfaces denote weak interactions (for example, van der Waals (VdW)), blue surfaces strong attractive interactions (for example, hydrogen bonds), and red surfaces strong repulsive interactions. The NCI calculations were performed with the NCI-Plot program⁴⁷. We focused on the hmC systems, and a representative snapshot from every system was subjected to NCI analysis. In all cases, the hmC substrate was considered as a ligand interacting with a spherical region of 10 Å around the binding site. All calculations were obtained with a step size of 0.2 Å for the cube and a cutoff of 5 Å for the calculation of the interactions between the nucleotides and the active site. The NCI analysis for a selected snapshot of WT and all mutants in the presence of hmC are presented in **Supplementary Figure 5**. We further examined the WT and T1372A/E/V mutants in the presence of mC and fC; these NCI analyses are presented in **Supplementary Figure 6**. The snapshots for NCI plots have been selected to highlight the most frequent interactions relevant to the underlying mechanism.

All EDA calculations were carried out with an in-house FORTRAN90 program to determine the non-bonded interactions (Coulomb and VdW interactions) for all the residues^{48–50}. The average non-bonded interaction between a particular cytosine derivative and every other residue, ΔE_{int} , is approximated by $\Delta E_{\text{int}} = \langle \Delta E_i \rangle$, where i represents an individual residue, ΔE_i represents the nonbonded interaction (Coulomb or VdW) between residue i and the particular cytosine derivative, and the broken brackets represent averages over the complete production ensemble obtained from the MD simulations. This analysis has been previously employed for quantum mechanics/molecular mechanics (QM/MM) and MD simulations to study a number of protein systems^{21,22,51–54}. The EDA results for all protein residues with mC, hmC, fC and caC are presented in **Supplementary Table 2**, and specific non-bonded interactions are shown in **Supplementary Table 3**. Hydrogen bond analyses for WT and all mutants with all cytosine bases are shown in **Supplementary Tables 4–7**. As noted, the above-described analyses were performed on each individual snapshot over each individual simulation, and the reported data consist of the averages over all the simulations for each system.

Purification of hTET2 variants from Sf9 insect cells. WT and select hTET2-CS mutants were subcloned into a pFastBac1 vector for expression in Sf9 insect cells as described previously³³. WT and T1372E were also generated in the full catalytic domain (hTET2-FCD, residues 1129–2002). Proteins were expressed for 24 h, and the cell pellet from a 500-ml culture was resuspended in lysis buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 0.2% (v/v) NP-40) with cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche, 1 tablet/10 ml) and 10 U/ml of Benzonase Nuclease (Millipore). Cells were lysed by one freeze-thaw cycle followed by passage through a 20-gauge and then a 25-gauge needle. The lysate was cleared by centrifugation at 20,000g for 30 min, and the supernatant was passed through a 0.2-μm syringe filter. A 250-μl column of anti-FLAG M2 affinity gel (Sigma) was prepared per manufacturer instructions and equilibrated in lysis buffer. The filtered lysate was applied twice to the column under gravity flow, and bound protein was washed with 10 ml then 2 × 5 ml of wash buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 15% (v/v) glycerol). Elutions of 250 μl were collected in wash buffer containing 100 μg/ml 3 × FLAG peptide (Sigma), with each elution incubated on the column for 5 min before collection, until no protein was detected by Bio-Rad Protein Assay and SDS-PAGE. Fractions were pooled, DTT added to 1 mM, and aliquots flash frozen in liquid nitrogen and stored at –80 °C.

TET reactions *in vitro*. For reactions under ‘driving’ conditions, purified TET2 enzymes were reacted with fluorescein (FAM)-labeled, 27-bp oligonucleotides containing a central reactive site (5'-GTA TCT AGT TCA ATC XGG TTC ATA GCA FAM-3', X = mC, hmC, or fC), duplexed with a complementary strand containing an unmodified CpG. Protein concentrations were measured by the Bio-Rad Protein Assay and standardized by diluting in elution buffer. A mixture of 20–25 nM duplexed DNA, 50 mM HEPES, pH 6.5, 100 mM NaCl, 1 mM α-ketoglutarate, 1 mM DTT, and 2 mM sodium ascorbate was pre-warmed to 37 °C. Immediately before the reaction, fresh ammonium iron(II) sulfate (Sigma) was added to 75 μM, and at time $t = 0$, TET2 was added to a

final concentration of 30 μg/ml (maximally 0.57 μM of hTET2-CS and 0.30 μM of hTET2-FCD). Reaction volumes were typically 200–350 μl. After incubation at 37 °C for 30 min (or at designated time points), the reactions were quenched by addition of 8 volumes of 100% ethanol with 2 volumes of Oligo Binding Buffer (Zymo). Reaction products were purified using the Zymo Oligo Clean & Concentrator kit, eluted in LC-MS grade H₂O, and analyzed by LC-MS/MS and/or enzyme-coupled assays³³.

For enzyme titration experiments, substrates were generated by PCR using 5-methyl- or 5-hydroxymethyl-dCTP and standard protocols for Taq polymerase. Each 745-bp amplicon contained a total of 391 modified cytosines (280 in CpG context) and was purified by gel extraction. Reaction conditions were the same as above, except for using 80 ng of PCR substrates and 1.856–72.5 μg/ml of enzyme in a 25-μl reaction. Following randomized analysis by LC-MS/MS, the percentage of total oxidation products (i.e., substrate consumed) was converted to nanomoles based on the known composition of the substrate. Plots were generated of total oxidation products versus enzyme concentration (**Supplementary Fig. 4** and **Supplementary Table 9**), and the slopes from linear regression were compiled in **Table 1**.

Chemoenzymatic assays of TET activity. We designed three chemoenzymatic assays to probe for specific cytosine modifications³³. Concentrated, purified reaction products representing 50 μl of the TET reaction (up to 1.25 pmol) were used for each assay.

To distinguish mC-containing oligos, the restriction enzyme MspI (NEB) was used, which normally cleaves CCGG sites containing C, mC, or hmC, with partial activity on fC and no activity on caC⁶. A combination of aldehyde reactive probe (ARP) (Thermo) and T4 β-glucosyltransferase (βGT) (NEB) were used to protect fC and hmC, respectively, from MspI cleavage, leaving only mC susceptible. The reaction products, along with controls, were treated first with 4.4 μM ARP in 6 mM HEPES, pH 5.0 (10 μl total volume), incubated at 37 °C overnight, then diluted into 20 μl with 1 × CutSmart Buffer (NEB), 2 mM uridine diphosphoglucose (UDP-Glc) and 1:25 volume of βGT for 30 min at 37 °C. To this mixture was added 50 U MspI in 1 × CutSmart Buffer and digestion carried out at 37 °C for >2 h.

To visualize the extent of higher-order oxidation to fC and caC, the reaction products were treated with 25-fold molar excess of thymine DNA glycosylase (TDG) purified as described below, in TDG buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 0.2 mM EDTA, 2.5 mM MgCl₂) for 2–4 h at 37 °C. After the reaction, 1:1 volume of 0.3 M NaOH/0.03 M EDTA was added and the mixture incubated at 85 °C for 15 min to cleave oligos at abasic sites. The TDG mutant N191A, which was previously found to excise fC and not caC⁵⁵, was also purified and used in the same manner to identify fC specifically.

As the final step of all three chemoenzymatic processes, the samples were mixed 1:1 with formamide containing bromophenol blue loading dye, loaded onto a 7 M urea/20% acrylamide/1 × TBE gel pre-warmed to 50 °C, and imaged for FAM fluorescence on a Typhoon 9200 variable mode imager.

LC-MS/MS analysis of reaction products. Concentrated, purified reaction products representing 200 μl of the TET reaction (up to 5 pmol) were degraded to component nucleosides with 1 U DNA Degradase Plus (Zymo) in 10 μl at 37 °C overnight. The nucleoside mixture was diluted ten-fold into 0.1% formic acid, and 20 μl was injected onto an Agilent 1200 Series HPLC with a 5 μm, 2.1 × 250 mm Supelcosil LC-18-S analytical column (Sigma) equilibrated to 50 °C in Buffer A2 (5 mM ammonium formate, pH 4.0). The nucleosides were separated in a gradient of 0–10% Buffer B2 (4 mM ammonium formate, pH 4.0, 20% (v/v) methanol) over 7 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 175 °C, gas flow of 10 l/min, nebulizer at 35 psi, sheath gas temperature of 300 °C, sheath gas flow of 11 l/min, capillary voltage of 2,000 V, fragmentor voltage of 70 V, and delta EMV of +1,000 V. Collision energies were optimized to 10 V for mC, fC, and T; 15 V for caC; and 25 V for hmC. MRM mass transitions and data analysis were as described above.

Purification of hTDG from *E. coli*. We adapted a published protocol⁵⁶ to express and purify WT and N191A TDG from BL21(DE3) cells. 1 L cultures

were grown to OD ~0.6, cooled gradually to 16 °C, induced with 0.25 mM IPTG at OD ~0.8, and grown for another 4 h. Cells were collected by centrifugation, resuspended in 20 ml TDG lysis buffer (50 mM NaPhos, pH 8.0, 300 mM NaCl, 25 mM imidazole) with protease inhibitors, and lysed by four passes on a microfluidizer. The lysate was cleared by centrifugation at 20,000g for 20 min, then passed through a 0.22- μ m syringe filter. A 1-ml column of HisPur cobalt resin (Thermo) was equilibrated in TDG lysis buffer, and the lysate bound by two applications to the column under gravity flow. The column was washed three times with 5 ml of TDG lysis buffer containing 1 M NaCl, then three times with 5 ml of regular TDG lysis buffer. Elutions of 1 ml each were collected in TDG lysis buffer containing increasing concentrations of imidazole: 50, 100, 150, 200, 250, and 500 mM imidazole. Elutions were evaluated by SDS-PAGE and dialyzed overnight at 4 °C into TDG storage buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM DTT, 0.5 mM EDTA, 1% (v/v) glycerol). Final protein concentrations were measured with the Bio-Rad Protein Assay and aliquots stored at -80 °C.

33. Liu, M.Y., DeNizio, J.E. & Kohli, R.M. Quantification of oxidized 5-methylcytosine bases and TET enzyme activity. *Methods Enzymol.* **573**, 365–385 (2016).
34. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
35. Schafmeister, C.E.A.F., Ross, W.S. & Romanovski, V. *The Leap Module of AMBER* (School of Pharmacy, University of California, San Francisco, 1995).
36. Case, D.A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
37. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
38. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. & Baker, N.A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
39. Dolinsky, T.J. *et al.* PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
40. Olsson, M.H., Søndergaard, C.R., Rostkowski, M. & Jensen, J.H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
41. Bradbrook, G.M. *et al.* X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes relating structure to thermodynamics of binding. *J. Chem. Soc., Faraday Trans.* **94**, 1603–1611 (1998).
42. Oda, A., Yamaotsu, N. & Hirono, S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.* **26**, 818–826 (2005).
43. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
44. Roe, D.R. & Cheatham, T.E. III. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
45. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).
46. Johnson, E.R. *et al.* Revealing noncovalent interactions. *J. Am. Chem. Soc.* **132**, 6498–6506 (2010).
47. Contreras-García, J. *et al.* NCIPLOT: a program for plotting non-covalent interaction regions. *J. Chem. Theory Comput.* **7**, 625–632 (2011).
48. Graham, S.E., Syeda, F. & Cisneros, G.A. Computational prediction of residues involved in fidelity checking for DNA synthesis in DNA polymerase I. *Biochemistry* **51**, 2569–2578 (2012).
49. Elias, A.A. & Cisneros, G.A. 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 (2014).
50. Dewage, S.W. & Cisneros, G.A. Computational analysis of ammonia transfer along two intramolecular tunnels in *Staphylococcus aureus* glutamine-dependent amidotransferase (GatCAB). *J. Phys. Chem. B* **119**, 3669–3677 (2015).
51. Cui, Q. & Karplus, M. Catalysis and specificity in enzymes: a study of triosephosphate isomerase and comparison with methyl glyoxal synthase. *Adv. Protein Chem.* **66**, 315–372 (2003).
52. Martí, S. *et al.* Preorganization and reorganization as related factors in enzyme catalysis: the chorismate mutase case. *Chemistry* **9**, 984–991 (2003).
53. Senn, H.M., O'Hagan, D. & Thiel, W. Insight into enzymatic C-F bond formation from QM and QM/MM calculations. *J. Am. Chem. Soc.* **127**, 13643–13655 (2005).
54. Cisneros, G.A. *et al.* Reaction mechanism of the epsilon subunit of *E. coli* DNA polymerase III: insights into active site metal coordination and catalytically significant residues. *J. Am. Chem. Soc.* **131**, 1550–1556 (2009).
55. Maiti, A., Michelson, A.Z., Armwood, C.J., Lee, J.K. & Drohat, A.C. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **135**, 15813–15822 (2013).
56. Morgan, M.T., Bennett, M.T. & Drohat, A.C. Excision of 5-halogenated uracils by human thymine DNA glycosylase. Robust activity for DNA contexts other than CpG. *J. Biol. Chem.* **282**, 27578–27586 (2007).