Published in final edited form as:

Adv Protein Chem Struct Biol. 2019 ; 117: 91–112. doi:10.1016/bs.apcsb.2019.05.002.

Comparison of DNA and RNA substrate effects on TET2 structure

Emmett M. Leddin, G. Andrés Cisneros*

Department of Chemistry, University of North Texas, Denton, TX, United States

Abstract

Ten-eleven translocation (TET) enzymes can perform the stepwise oxidation of 5-methylcytosine (5mC) to 5-carboxylcytosine on both single-stranded (ss) and double-stranded (ds) DNA and RNA. It has been established that TET2 has a preference for ds DNA substrates, but it can catalyze the oxidation reaction on both ssDNA and RNA. The reasons for this substrate preference have been investigated for only a substrate 5mC ribonucleotide in a DNA strand, but not other nucleic acid configurations (*Biochemistry* **58** (2019) 411). We performed molecular dynamics simulations on TET2 with various ss and ds substrates in order to better understand the structural and dynamical reasons for TET2's preference to act on ds DNA. Our simulations show that substrates that have a ribonucleotide experience several disruptions in their overall backbone shape, hydrogen bonding character, and non-bonded interactions. These differences appear to lead to the instability of ribonucleotide in the active site, and provide further rational for TET2's experimental behavior.

1. Introduction

The methylation and demethylation of DNA and RNA is important for many biological functions. Nucleobase methylation is a mainstay epigenetic feature, and it plays a role in tasks such as X-chromosome inactivation and genomic imprinting (Jones, 2012; Kohli & Zhang, 2013; Rasmussen & Helin, 2016; Williams, Christensen, & Helin, 2012).

Most epigenetic methylation marks on DNA occur as 5-methylcytosine (5mC), whereas most RNA epigenetic methylation marks occur as N6-methyladenosine (Chen-Kiang, Nevins, & Darnell, 1979; Jin, Li, & Robertson, 2011). DNA methylation is catalyzed by DNA methyltransferases (DNMTs) acting on deoxycytidine and adenosine, respectively (Bird, 2002).

Several enzymes in the α -ketoglutarate- (α -KG) and Fe(II)-dependent superfamily can perform nucleoside demethylation. The general catalytic cycle for the oxidation of alkyl moieties can be separated into two general stages. The first stage involves the binding of the substrate and co-factors, α -KG and O₂, and formation of a ferryl (Fe(IV)=O) intermediate with the concomitant oxidation of α -KG to succinate and CO₂. Subsequently, the second stage involves the oxidation of the methyl moiety on the substrate. Computational

*Corresponding author: Gerardo.Cisneros@unt.edu.

simulations using hybrid quantum mechanics/molecular mechanics (QM/MM) have identified several features of the catalytic cycle for TET2 and a related enzyme, AlkB (Fang & Cisneros, 2014; Fang, Lord, & Cisneros, 2013; Torabifard & Cisneros, 2018).

Ten eleven translocation (TET) dioxygenases belong to the α -KG- and Fe(II)-dependent superfamily of enzymes. A key structural feature of these enzymes is a conserved double sheet β -helix, also known a jelly-roll fold, in their catalytic cores (Aik, McDonough, Thalhammer, Chowdhury, & Schofield, 2012). The TET name derives from a common translocation between the *TET1* and *MLL* genes (Lorsbach et al., 2003). TET enzymes catalyze the stepwise oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-carboxylcytosine (5caC), and 5-formylcytosine (5fC), sequentially (Fig. 1). These enzymes can perform this oxidation on both methylated DNA and RNA nucleosides, as part of single-stranded (ss) or double-stranded (ds) substrates (Fu et al., 2014; Kizaki & Sugiyama, 2014). A part of the reaction catalyzed by TET enzymes involves a base flipping (eversion) of the nucleic acid, wherein the target residue is flipped out from the substrate strand into the active site (Hong & Cheng, 2016). Base flipping is not limited to TET enzymes, but rather, it is seen across several enzymes that read, write, or otherwise modify DNA bases, including the DNA glycosylases involved in base excision repair (BER) (Brooks, Adhikary, Rubinson, & Eichman, 2013; Drohat & Coey, 2016). The enzymes involved with BER complete the final conversion of the 5caC oxidation product back into dC.

Two of the three TET enzymes, TET1 and TET3, contain a CXXC domain that can bind directly to DNA (Rasmussen & Helin, 2016). While not required for oxidation of 5mC, the CXXC domain has been proposed to increase TET's selectivity (Shen, Song, He, & Zhang, 2014). TET2 does not contain a CXXC binding domain, but is known to partner with IDAX, which does (Ko et al., 2013; Kohli & Zhang, 2013). Recently, a scaffold in the active site of TET2, composed by T1372 and Y1902, has been shown to be required for the sequential oxidation of the substrate (Liu et al., 2016).

It has been shown that TET2 recognizes the cytosine guanine (CpG) dinucleotide sequence of DNA and is unlikely to discriminate between 5mC, 5hmC, and 5fC in the active site (Hu et al., 2015; Tsai & Tainer, 2013). TET2's inability to discern between reactants is proposed to result from the cytosine base adopting nearly identical orientations in the active site. Despite their similarities in orientation, 5mC and 5hmC have been shown to stabilize DNA duplexes (Thalhammer, Hansen, El-Sagheer, Brown, & Schofield, 2011), whereas DNA duplexes containing either 5fC or 5caC destabilize DNA duplexes (Raiber et al., 2015), although this has been challenged (Hardwick et al., 2017). This destabilization leads to what has been deemed F-form DNA, and reduction from 5fC to 5hmC can return the backbone structure to its typical B-form.

While TET2 has been shown to act on both DNA and RNA substrates, the differences between both contexts is poorly understood. Recently, we identified a distinct pattern for selectivity of TET2 complexes, favoring dsDNA and ssDNA, and disfavoring dsRNA and ssRNA (DeNizio, Liu, Leddin, Cisneros, & Kohli, 2019). Through comparing with previous computational work on TET2-DNA complexes (Liu et al., 2016), we uncovered several

residues with large differences in their non-bonded interactions with the target residue, as well as a shift in the DNA backbone structure with the inclusion of an RNA base.

In this work, we extend our computational work to further investigate how a variety of different RNA and DNA substrates impact TET2 complexes and provide a possible explanation for the pattern of selectivity seen experimentally.

2. Methodology

2.1 Structure preparation and parametrization

The crystal structure of a TET2-DNA complex (PDBID: 4NM6) (Hu et al., 2013) was prepared as previously described (DeNizio et al., 2019; Liu et al., 2016). Briefly, the simulation systems involve the following features: a magnesium surrogate for the active site iron, protonation of the His residues coordinating the active site divalent cation on the δ nitrogen, and the insertion of a glycine-serine linker. The linker was introduced using the Modeller program (Šali & Blundell, 1993; Salomon-Ferrer, Götz, Poole, Le Grand, & Walker, 2013).

The 5hmrC, 5scarC, and 5frC ribonucleotide parameters were all generated using PyRED (Bayly, Cieplak, Cornell, & Kollman, 1993; Dupradeau et al., 2010; Vanquelef et al., 2011; Wang, Becker, Cieplak, & Dupradeau, 2013). The published parameters for deoxyribonucleotide cytosine derivatives, α -KG, Mg(II), and Zn(II) were used (Liu et al., 2016), in addition to the 5mrC parameters from our previous work (DeNizio et al., 2019).

2.2 MD simulations

The pmemd.cuda program from the AMBER16 suite (Case et al., 2017; Götz et al., 2012; Salomon-Ferrer et al., 2013) was used to perform molecular dynamics (MD) simulations for all TET2-nucleic acid complexes. Eight systems were created and simulated with the ff99SB force field. The systems considered include: 1) ssDNA with 5mrC (ss 5mrC); 2) ssDNA with 5mdC (ss 5mdC); 3) dsDNA with 5hmrC, 4) 5scarC, or 5) 5frC in the substrate strand with 5mdC in the complementary strand (respectively 5hmrC:5mdC, 5scarC:5mdC, and 5frC:5mdC); 6) dsDNA with 5mrC in both the substrate and complementary strands (5mrC:5mrC); 7) dsDNA with 5mdC in the substrate strand and 5mrC in the complementary strand (5mdC:5mrC); and 8) dsDNA with 5mdC in the substrate and dC in the complementary strand (5mdC:dC).

The LEaP module in AMBER was used to generate each system (Schafmeister, Ross, & Romanovski, 1995). Each system was neutralized to a net charge of zero using potassium ions and solvated with TIP3P explicit water extending at least 12 Å from the system surface in a truncated octahedron (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). The smooth particle mesh Ewald method with a 9 Å direct space cutoff was used to treat long-range electrostatic interactions (Essmann et al., 1995). The same cutoff was used for the Van der Waals interactions.

A 2 fs time step was used for simulations with the Berendsen thermostat in the *NVT* ensemble (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984). A minimum of three

replicates for each system were run for 90 ns of production, resulting in 270 ns of total production time per system, for a total of over 2 s of production simulation time.

2.3 Comparison data

Earlier work studied dsDNA with both 5mdC in substrate and complement (5mdC:5mdC) (Liu et al., 2016). These simulations were minimized and heated in the *NVT* ensemble and run for 40 ns of production using the *NPT* ensemble. Later work looked at dsDNA with 5mrC in the substrate and complement (5mrC:5mdC) using the *NVT* ensemble procedure described here, as well as comparing both procedures to ensure meaningful comparisons could be made (DeNizio et al., 2019).

2.4 Analysis

Different structural analyses, including hydrogen bond analysis, root mean square deviations (RMSDs), root mean square fluctuations (RMSFs), matrix correlations, and nucleic acid backbone torsions and base pair geometries were performed using cpptraj (Babcock, Pednault, & Olson, 1994; Olson et al., 2001; Roe & Cheatham, 2013). Averaging for these analyses was performed using the data.table, abind, and tidyverse packages in R (Dowle & Srinivasan, 2018; Plate & Heiberger, 2016; R Core Team, 2018; Wickham, 2017). VMD and UCSF Chimera were used for data visualization and image creation (Humphrey, Dalke, & Schulten, 1996; Pettersen et al., 2004). An in-house FORTRAN90 program was used to perform an energy decomposition analysis (EDA) (Dewage & Cisneros, 2015; Elias & Cisneros, 2014; Graham, Syeda, & Cisneros, 2012). The non-bonded interactions, either Coulomb or Van der Waals, for each individual residue, i , are computed with every other residue j as per (Eq. 1).

$$E_{int} = \langle E_i - E_j \rangle, \quad (1)$$

where the individual non-bonded components, Coulomb and Van der Waals, for every residue pair i and j are averaged over the generated ensemble, and $\langle \dots \rangle$ denotes an ensemble average.

The sum of the two non-bonded components for each individual E_{int} values provides a qualitative understanding of the non-bonded interactions for every residue pair (Eq. 2).

$$E_{Tot}^{Non-band} = E_{int,Coul} + E_{int,VdW} \quad (2)$$

Systems can be compared relative to each other by taking the difference in the total non-bonded interactions between two systems (Eq. 3).

$$\Delta E_{Tot}^{Non-band} = E_{Tot,ref}^{Non-band} - E_{Tot}^{Non-band} \quad (3)$$

3. Results and discussion

Investigating various nucleosides in a TET2-DNA complex could lead to a better understanding of TET2's preference for (a) ds over ss substrates and (b) DNA over RNA. Because both DNA and RNA undergo methylation to 5mC, it is important to understand if or how the interactions between TET2 and different DNA or RNA substrates change.

The RMSD analysis for each system is shown in Figs. 2 and 3. The ss 5mC and 5mC:5mC systems showed slightly higher overall deviation from the crystal structure than the other systems studied. This may be due to the fact that the lack of the complementary strand reduced the number of contacts between the substrate and the protein resulting in higher flexibility of the ssDNA strand in these systems (*vide infra*).

The RMSF analysis suggests that the overall fluctuation of most protein residues fluctuates similarly between the deoxy and ribo systems. Conversely, the nucleic acids exhibit larger fluctuations in the ribo systems (Fig. 4). This difference suggests that the inclusion of the ribonucleotide creates an instability in the nucleic acid backbone.

3.1 Backbone analysis

In our previous work with the 5mC:5mdC system, we identified a slight shift in the DNA substrate from B-form to A-form, due to the inclusion of the ribonucleotide (DeNizio et al., 2019). To see if a similar pattern emerged for the other systems considered herein, we looked again at the different χ and δ angles throughout the course of the simulations. Many systems showed distinct shifts in the substrate's backbone angle populations in the ribo systems, when compared to their deoxy counterparts (Fig. 5). Remarkably, the 5fdC:5mdC and 5fC:5mdC systems have the same spread of values for χ and δ as 5mC:5mdC. This suggests that the positioning of the target nucleic acid in the active site is more similar between 5fdC:5mdC and 5mC:5mdC than 5mdC:5mdC.

The structure of the DNA helix is impacted in the backbone arrangement and base pairing by the identity of both complement and substrate nucleobases (Packer & Hunter, 1998). Here, we found that changing the composition of, or removing, the complementary strand impacted the substrate strand's backbone angles (Fig. 6). The ssDNA system shared its distinct populations for 5mdC and 5mC with the dsDNA 5mdC:5mdC and 5mC:5mdC systems, respectively. While changing the complementary strand can change the substrate's backbone angles, the complete removal of the complementary strand impacts the backbone stability, and thus the spread of the χ and δ angle populations.

The differences in the backbone and base-pairing arrangements can also be observed by the changes to the shift and slide parameters for the C3:G3' and A4:T4' base pairs (Fig. 7). Each of the ribo systems with a 5mdC complement strand have a wider spread for these values. Alterations to the complementary strand tightened the range for shift and slide, where 5mdC:dC most closely resembled 5mdC:5mdC (Fig. 7E).

As discussed in the introduction, Raiber et al. (2015) uncovered a distinct difference between 5fdC and the other oxidation products, wherein the backbone adopts a structure that

they have termed F-DNA. These findings have been challenged by Hardwick et al. (2017). Our simulations show that the backbone parameters for the 5mrC:5mdC system most closely resemble those from the 5fdC:5mdC and 5frC:5mdC systems. Since this system shifted away from B-form DNA, but does not fit entirely with the definition of A-form DNA, the modified nucleic acid strands exhibit characteristics from A-, B-, and F-form DNA, which may help explain this discrepancy.

3.2 Correlation matrices

Correlation matrices show areas of correlation, anti-correlation, or no correlation for a protein relative to itself. Each of the ds systems simulated here show strong similarities in their self-correlations (Fig. 8).

Building from this, taking the difference of the self-correlations between two systems can provide insight into their dynamical differences. Differences in the correlations between ss 5mdC and ss 5mrC show that the ss 5mrC system has slightly stronger correlation between the DNA-interacting loops (residues 1256-1273 and 1288-1312), but slightly stronger anti-correlation with the rest of the protein (Fig. 9). The ss 5mdC/ss 5mrC difference was nearly identical in magnitude for the comparison between 5fdC:5mdC and 5frC:5mdC. The inherent flexibility of the ssDNA, and the aforementioned conformational difference of ds helices with 5fdC, together may explain the similarities in the correlated and anticorrelated movements of the four systems.

Smaller differences are observed between the 5hmdC:5mdC and 5mrC:5mdC systems (Fig. 10). Compared with the ssDNA systems, the DNA-interacting loops (residues 1256-1273 and 1288-1312) showed slightly stronger correlation between the DNA-interacting loops in the 5hmdC:5mdC system. Comparing each system with an altered complementary strand to the 5mdC:5mdC system showed the most dramatic differences in 5mdC:dC and 5mrC:5mrC, with 5mdC:5mrC exhibiting the same differences more subtly. Each of these systems was less strongly correlated in the DNA-interacting loops than the reference. Taken together, these results suggest that the movements of the DNA-interacting region differ with the different modified cytosine bases, and that the extent of these differences is dependent on both the substrate and complement (or lack thereof) strands.

3.3 Hydrogen bond analysis

The replacement of a deoxy substrate by a ribo substrate induced several changes to the protein's hydrogen bonding interactions. When compared to their respective deoxy systems, both 5hmrC:5mdC and 5carC:5mdC had hydrogen bond between α -KG and R1261 for less time overall, whereas 5frC:5mdC had an increased length of this interaction. A similar pattern, where the hydrogen bond increased in the 5frC:5mdC system and decreased in the 5hmrC:5mdC and 5carC:5mdC systems, was also seen between D1384 and N1387. D1384 is coordinated to the divalent metal center, while N1387 and R1261 both surround the catalytic core (Fig. 11A). Because of this, changes to their orientations induced by the ribo oxygen could impact the demethylation reaction, and thus impede activity.

Both 5hmdC and 5fdC are known to form hydrogen bonds between H1904 and N1387 as part of the stabilization for the flipped in base in the active site (Hu et al., 2015). We

observed decreased prevalence of hydrogen bonds formed for these two residues in the ribo systems (Table 1).

Each of the systems with altered complementary strands (5mdC:dC, 5mdC:5mrC, 5mrC:5mrC, and ss 5mrC) shared several residues with major differences in their hydrogen bonding (Fig. 11B). Each of the altered complementary strand systems can be compared with a reference system most closely matching the system with the unaltered nucleotide to identify the changes in hydrogen bonding patterns. Specifically, 5mdC:dC and 5mdC:5mrC can be compared with 5mdC:5mdC; 5mrC:5mrC can be compared with 5mrC:5mdC; and ss 5mrC with ss 5mdC. The residues that showed hydrogen bonds with longer permanence throughout the production simulation in the reference systems correspond to residues that line the beta sheets enclosing the active site (N1266, E1268, D1376, Y1895, substrate, H1904, G7, T1259). Residues farther from the nucleic acid strand (N1403, T1399, D1412, N1403, E1178, R1176) showed longer lasting hydrogen bonds in the systems with altered complementary strands.

The hydrogen bond between residues E1165 and G1172 was observed for a longer time across simulations in the single stranded ribo system than in the ds altered complementary systems. This interaction takes place in a central area of the protein. The other major change in hydrogen bonding occurred between the substrate and H1904. The presence of that hydrogen bond was diminished by about 30% of the simulation time for 5mdC:5mrC, 5mrC:5mrC, and ss 5mrC. However, for 5mdC:dC, the presence of the substrate - H1904 hydrogen bond increased by about 35%.

TET2 has been shown experimentally to have more oxidation events with 5mdC:dC than any other tested substrate, and that all-deoxy substrates were preferred to DNA-RNA duplexes or chimeras (DeNizio et al., 2019). Here, we identified several hydrogen bonds that change between these contexts to accommodate ribonucleotides in the active site. Our results indicate that residues distal from the nucleic acid strand form longer-lived (or new) hydrogen bonds when a ribonucleotide is present, while breaking hydrogen bonds in the area closest to the strand in order to allow for the additional ribo oxygen, which may provide one explanation for the experimental pattern of activity.

3.4 Energy decomposition analysis

There are several key residues that interact with α -KG (R1261, R1896, and S1898), impact iron chelation (H1382, D1384, and H1881), and recognize DNA (T1372, N1387, Y1902, and H1904) (Table 2). These residues have previously been identified, and some have been observed from mutations in patients with myeloid malignancies (Yin & Xu, 2016). Therefore, the inter-molecular interactions of these key residues are crucial, and changes in these interactions arising from differences in the substrate may have a significant impact in TET2 activity. The most pronounced changes in inter-molecular interactions in these residues include R1261, D1384, and R1896. R1261, for example, has a considerably smaller $E_{Tot}^{Non-bond}$ for the ds 5mC systems, while similarly favoring the ribo system.

An earlier investigation of TET2 mutations identified a loss of activity with the T1372E mutant, stalling at 5hmC, due to the disruption of an active site scaffold that could be

recovered with a secondary Y1902A mutation (Liu et al., 2016). Because a similar stalling pattern arose with the ribo systems experimentally, we were interested in looking at commonalities between the 5hmdC T1372E mutant and 5mrC:5mdC. Consistent with our previous work, we found that the ribo systems did not exhibit any pronounced differences for the intermolecular interactions between the substrate and T1372 or Y1902 regardless of the identity or the position of the ribonucleotide modification (substrate or complement). However, the interaction energies between the substrate residue and the iron-coordinated aspartate (D1384) were much more comparable than any other systems studied here (Table 2). Accordingly, the substrate – D1384 interaction appears to be critical to TET2's function.

Expanding to the entire protein, differences between the deoxy and ribo systems were observed across each of the oxidation products (Fig. 12). Pronounced changes were seen in the ss systems, particularly in the two loops that interact with the DNA (residues 1256-1273, 1288-1312), as well as the three zinc-coordinating regions (Zn1: C1133, C1135, C1221; Zn2: C1193, C1271; Zn3: C1289, C1298, C1358, A1458, H1912) (Hu et al., 2013).

Approximately 39% of the residues between the 5caC systems showed differences ~ 1 kcal mol⁻¹, with over 25% of the protein showing differences ~ 10 kcal mol⁻¹. Thus, the ribo oxygen, in combination with the two oxygens inherent to 5caC, adds additional obstructive interactions to the already crowded active site.

Comparing the systems with changes to the complementary strands, two additional residues (excluding the active site, DNA-interacting loops, and zinc-coordinating regions) had large E values. For $E_{5\text{mdC}:5\text{mdC}-5\text{mrC}:5\text{mrC}}$, these included H1386 (-3.55 ± 0.50 kcal mol⁻¹) and K1462 (8.11 ± 0.58 kcal mol⁻¹). For $E_{5\text{mdC}:5\text{mdC}-5\text{mdC}:dC}$, this again included K1462 (4.05 ± 0.21 kcal mol⁻¹). Both H1386 and K1462 are located in non-structured regions on the same face of the protein as the DNA helix, and are close enough to the DNA strand that their inherent flexibility may impact the magnitude of their interaction (Fig. 13).

Similar to the hydrogen bonding analysis results, the addition of oxygen in the ribonucleotide induced disruptions to the non-bonded interactions, giving rise to instability in the active site. These differences were seen predominantly in the active site and nucleic acid backbone, with farther-reaching effects for the ss systems. Change from 5mdC to dC in the complementary strand better localized these energetic differences to the backbone. Together, both the changes and hydrogen bonding and EDA agree with the experimental preference for ds, deoxy substrates, due to the active site instability of ribonucleotides and, to a lesser extent, the flexibility of ssDNA.

4. Conclusions

We have performed a computational investigation of the effect of the inclusion of various modified ribonucleotides at the substrate and complement positions on the structure and dynamics of TET2. Our results indicate that replacing a modified cytosine deoxynucleotide by a modified cytosine ribonucleotide in the target position of a TET2-DNA complex has a large and fairly consistent impact on the backbone structure, energetics, and dynamics of the system. Our results suggest that ribonucleotide substrates may be disfavored due to their

instability in the active site, brought about by repulsive interactions with the additional oxygen moiety on the ribose ring, and a significant shifting of the substrate structure from a B-form DNA helix into something sharing characteristics with A-form, B-form, and “F-form” DNA. Interestingly, having a dC nucleobase in the complementary strand can help stabilize ribonucleotide substrates, while a ribonucleotide in the complementary strand has a similar overall effect as the inclusion of a ribonucleotide target at the substrate position. Thus, it appears that the composition of both the substrate are important for TET2’s function, and that deviations from typical DNA substrates further complicate the complete oxidation of 5mC.

Acknowledgments

This work was funded by NIH via R01GM108583 and R01GM118501. Computing time from CASCaM with partial support from NSF Grant No. CHE-1531468 is acknowledged. The authors thank Dr. A.R. Walker for assistance with the parametrization of the RNA cytosine-derivatives.

References

- Aik W, McDonough MA, Thalhammer A, Chowdhury R, & Schofield CJ (2012). Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Current Opinion in Structural Biology*, 22(6), 691–700. 10.1016/j.sbi.2012.10.001. [PubMed: 23142576]
- Babcock MS, Pednault EP, & Olson WK (1994). Nucleic acid structure analysis: Mathematics for local cartesian and helical structure parameters that are truly comparable between structures. *Journal of Molecular Biology*, 237(1), 125–156. 10.1006/jmbi.1994.1213. [PubMed: 8133513]
- Bayly CI, Cieplak P, Cornell W, & Kollman PA (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *The Journal of Physical Chemistry*, 97(40), 10269–10280. 10.1021/j100142a004.
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, & Haak JR (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8), 3684–3690. 10.1063/1.448118.
- Bird A (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1), 6–21. 10.1101/gad.947102. [PubMed: 11782440]
- Brooks SC, Adhikary S, Rubinson EH, & Eichman BF (2013). Recent advances in the structural mechanisms of DNA glycosylases. *Biochimica et Biophysica Acta*, 1834(1), 247–271. 10.1016/j.bbapap.2012.10.005 (Special Section: Chemistry and mechanism of phosphatases, diesterases and triesterases). [PubMed: 23076011]
- Case D, Cerutti D, Cheatham T III, Darden T, Duke R, Giese T, et al. (2017). *Amber16* (San Francisco).
- Chen-Kiang S, Nevins JR, & Darnell JE (1979). N-6-methyl-adenosine in adenovirus type 2 nuclear RNA is conserved in the formation of messenger RNA. *Journal of Molecular Biology*, 135(3), 733–752. 10.1016/0022-2836(79)90174-8. [PubMed: 537090]
- DeNizio JE, Liu MY, Leddin EM, Cisneros GA, & Kohli RM (2019). Selectivity and promiscuity in TET-mediated oxidation of 5-methylcytosine in DNA and RNA. *Biochemistry*, 58(5), 411–421. 10.1021/acs.biochem.8b00912 (PMID: 30387995). [PubMed: 30387995]
- Dewage SW, & Cisneros GA (2015). Computational analysis of ammonia transfer along two intramolecular tunnels in *Staphylococcus aureus* glutamine-dependent amidotransferase (GatCAB). *The Journal of Physical Chemistry B*, 119(9), 3669–3677. 10.1021/jp5123568 (PMID: 25654336). [PubMed: 25654336]
- Dowle M, & Srinivasan A (2018). data.table: Extension of ‘data.frame’ [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=data.table> (R package version 1.11.8).
- Drohat AC, & Coey CT (2016). Role of base excision “repair” enzymes in erasing epigenetic marks from DNA. *Chemical Reviews*, 116(20), 12711–12729. 10.1021/acs.chemrev.6b00191. [PubMed: 27501078]

- Dupradeau F-Y, Pigache A, Zaffran T, Savineau C, Lelong R, Grivel N, et al. (2010). The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics*, 12, 7821–7839. 10.1039/C0CP00111B. [PubMed: 20574571]
- Elias AA, & Cisneros GA (2014). Chapter two - computational study of putative residues involved in DNA synthesis fidelity checking in *Thermus aquaticus* DNA polymerase I In Karabencheva-Christova T (Ed.), *Biomolecular modelling and simulations* (Vol. 96, pp. 39–75). Academic Press 10.1016/bs.apcsb.2014.06.003.
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, & Pedersen LG (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19), 8577–8593. 10.1063/1.470117.
- Fang D, & Cisneros GA (2014). Alternative pathway for the reaction catalyzed by DNA dealkylase AlkB from ab initio QM/MM calculations. *Journal of Chemical Theory and Computation*, 10(11), 5136–5148. [PubMed: 25400523]
- Fang D, Lord RL, & Cisneros GA (2013). Ab initio QM/MM calculations show an intersystem crossing in the hydrogen abstraction step in dealkylation catalyzed by AlkB. *The Journal of Physical Chemistry B*, 117(21), 6410–6420. 10.1021/jp403116e (PMID: 23642148). [PubMed: 23642148]
- Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, et al. (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *Journal of the American Chemical Society*, 136(33), 11582–11585. 10.1021/ja505305z (PMID: 25073028). [PubMed: 25073028]
- Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, & Walker RC (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation*, 8(5), 1542–1555. 10.1021/ct200909j (PMID: 22582031). [PubMed: 22582031]
- Graham SE, Syeda F, & Cisneros GA (2012). Computational prediction of residues involved in fidelity checking for DNA synthesis in DNA polymerase I. *Biochemistry*, 51(12), 2569–2578. 10.1021/bi201856m (PMID: 22397306). [PubMed: 22397306]
- Hardwick JS, Ptchelkine D, El-Sagheer AH, Tear I, Singleton D, Phillips SEV, Lane AN, & Brown T (2017). 5-Formylcytosine does not change the global structure of DNA. *Nature Structural & Molecular Biology*, 24(6), 544–552.
- Hong S, & Cheng X (2016). DNA base flipping: A general mechanism for writing, reading, and erasing DNA modifications In Jeltsch A, & Jurkowska RZ (Eds.), *DNA methyltransferases - role and function* (pp. 321–341). Cham: Springer International Publishing 10.1007/978-3-319-43624-1_14.
- Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, et al. (2013). Crystal structure of TET2-DNA complex: Insight into TET-mediated 5mC oxidation. *Cell*, 155(7), 1545–1555. 10.1016/j.cell.2013.11.020. [PubMed: 24315485]
- Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, et al. (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature*, 527, 118–122. 10.1038/nature15713. [PubMed: 26524525]
- Humphrey W, Dalke A, & Schulten K (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33–38. 10.1016/0263-7855(96)00018-5. [PubMed: 8744570]
- Jin B, Li Y, & Robertson KD (2011). DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes & Cancer*, 2(6), 607–617. 10.1177/1947601910393957. [PubMed: 21941617]
- Jones PA (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13, 484–492. 10.1038/nrg3230.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), 926–935. 10.1063/1.445869.
- Kizaki S, & Sugiyama H (2014). CGmCGCG is a versatile substrate with which to evaluate Tet protein activity. *Organic and Biomolecular Chemistry*, 12, 104–107. 10.1039/C3OB41823E. [PubMed: 24162071]
- Ko M, An J, Bandukwala HS, Chavez L, Aijö T, Pastor WA, et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, 497, 122–126. 10.1038/nature12052. [PubMed: 23563267]

- Kohli RM, & Zhang Y (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, 502, 472–479. 10.1038/nature12750. [PubMed: 24153300]
- Liu MY, Torabifard H, Crawford DJ, DeNizio JE, Cao X-J, Garcia BA, et al. (2016). Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. *Nature Chemical Biology*, 13, 181–187. [PubMed: 27918559]
- Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, & Downing JR (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, 17, 637–641. 10.1038/sj.leu.2402834. [PubMed: 12646957]
- Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, et al. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of Molecular Biology*, 313(1), 229–237. 10.1006/jmbi.2001.4987. [PubMed: 11601858]
- Packer MJ, & Hunter CA (1998). Sequence-dependent DNA structure: The role of the sugar-phosphate backbone. *Journal of Molecular Biology*, 280(3), 407–420. 10.1006/jmbi.1998.1865. [PubMed: 9665845]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. 10.1002/jcc.20084. [PubMed: 15264254]
- Plate T, & Heiberger R (2016). abind: Combine multidimensional arrays [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=abind> (R package version 1.4-5).
- R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria Retrieved from <https://www.R-project.org/>.
- Raiber E-A, Murat P, Chirgadze DY, Beraldi D, Luisi BF, & Balasubramanian S (2015). 5-formylcytosine alters the structure of the DNA double helix. *Nature Structural & Molecular Biology*, 22, 44–49. 10.1038/nsmb.2936.
- Rasmussen KD, & Helin K (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes & Development*, 30(7), 733–750. 10.1101/gad.276568.115. [PubMed: 27036965]
- Roe DR, & Cheatham TE (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7), 3084–3095. 10.1021/ct400341p (PMID: 26583988). [PubMed: 26583988]
- Šali A, & Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3), 779–815. 10.1006/jmbi.1993.1626. [PubMed: 8254673]
- Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, & Walker RC (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *Journal of Chemical Theory and Computation*, 9(9), 3878–3888. 10.1021/ct400314y (PMID: 26592383). [PubMed: 26592383]
- Schafmeister CEAF, Ross WS, & Romanovski V (1995). LEAP. San Francisco: University of California.
- Shen L, Song C-X, He C, & Zhang Y (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annual Review of Biochemistry*, 83(1), 585–614. 10.1146/annurev-biochem-060713-035513.
- Thalhammer A, Hansen AS, El-Sagheer AH, Brown T, & Schofield CJ (2011). Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chemical Communications*, 47, 5325–5327. 10.1039/C0CC05671E. [PubMed: 21451870]
- Torabifard H, & Cisneros GA (2018). Insight into wild-type and t1372e TET2-mediated 5hmc oxidation using ab initio QM/MM calculations. *Chemical Science*, 9, 8433–8445. 10.1039/c8sc02961j. [PubMed: 30542593]
- Tsai C-L, & Tainer JA (2013). Probing DNA by 2-OG-dependent dioxygenase. *Cell*, 155(7), 1448–1450. 10.1016/j.cell.2013.12.002. [PubMed: 24360270]
- Vanqualef E, Simon S, Marquant G, Garcia E, Klimerak G, Delepine JC, et al. (2011, 05). 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 Research*, 39(Suppl. 1_2), W511–W517. 10.1093/nar/gkr288. [PubMed: 21609950]

- Wang F, Becker J-P, Cieplack P, & 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.
- Wickham H (2017). tidyverse: Easily install and load the ‘tidyverse’ [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=tidyverse> (R package version 1.2.1).
- Williams K, Christensen J, & Helin K (2012). DNA methylation: TET proteins—guardians of CpG islands? EMBO Reports, 13(1), 28–35. 10.1038/embor.2011.233.
- Yin X, & Xu Y (2016). Structure and function of TET enzymes In Jeltsch A, & Jurkowska RZ (Eds.), DNA methyltransferases - role and function (pp. 275–302). Cham: Springer International Publishing 10.1007/978-3-319-43624-1_12.

Fig. 1. The proposed reaction mechanism for TET enzymes. The conversion from 5mC to 5caC occurs by stepwise oxidation, where the initial reactant “X,” is converted into its immediate product, “Y.”

Fig. 2.
RMSD plots for the ss 5mdC, ss 5mrC, 5mdC:dC, 5mdC:5mrC, and 5mrC:5mrC systems over 90 ns of production. Bézier line smoothing has been applied.

Fig. 3.
RMSD plots for the 5mrC:5mdC, 5hmrC:5mdC, 5carC:5mdC, and 5frC:5mdC systems over 90 ns of production. Bézier line smoothing has been applied.

Fig. 4. RMSF plots for the (A) 5mrC:5mdC, (B) 5hmrC:5mdC, (C) 5scarC:5mdC, and (D) 5frC:5mdC systems. The region with the glycine-serine linker is denoted by “GS linker.”

Fig. 5.
 δ and χ angle populations for the target substrate base in the deoxy systems (A) compared with the ribo systems (B). The 5mdC:5mdC system in (A) is blue and the 5mrC:5mdC system in (B) is orange. In both subfigures, the different 5hmC systems are purple, 5fC systems are gold, and 5caC systems are pink. All replicates are shown, excluding one from 5caC:5mdC exhibiting artifacting.

Fig. 6.

(A) δ and χ angle populations for the target substrate base in the systems with altered complementary strands: 5mdC:dC (pink), 5mdC:5mrC (green), and 5mdC:5mdC (cyan) systems. (B) The original DNA strand, showing the altered residues for substrate (mC6, blue) and complement (mC7', sea green). All replicates are shown.

Fig. 7.

Differences in shift and slide for the C3:G3' and A4:T4' base pairs for (A) 5hmrC:5mdC (purple), (B) 5frC:5mdC (gold), (C) 5carC:5mdC (pink), (D) 5mrC:5mrC (cyan), (E) 5mdC:dC (pink), and (F) 5mdC:5mrC (green). In (A)–(C), the blue corresponds to the deoxy reference system; in (D)–(F), the blue corresponds to 5mdC:5mdC. All replicates, excluding one from 5carC:5mdC with incalculable values, are shown.

Fig. 8. Self-correlation matrices for (A) 5hmrC:5mdC, (B) 5frC:5mdC, (C) 5carC:5mdC, (D) 5mdC:dC, (E) 5mdC:5mrC, and (F) 5mrC:5mrC. Correlated movement is shown in blue, anti-correlated movement is shown in red, and areas with no correlation are yellow.

Fig. 9. Differences in the correlation matrices between ssDNA systems. Areas colored blue are stronger in ss 5mdC, whereas areas colored orange are stronger in the ss 5mrC system.

Fig. 10. Differences in the correlation matrices between 5hmdC:5mdC and 5mrC:5mdC systems. Areas colored blue are stronger in 5hmdC:5mdC system, whereas areas colored orange are stronger in the 5mrC:5mdC system.

Fig. 11. (A) Residues surrounding the active site for the 5hmrC:5mdC, 5frC:5mdC, and 5carC:5mdC that are 20% different between the deoxy and ribo systems. (B) Residues with differences 20% in hydrogen bonding indicated in all systems with altered complementary strands (5mdC:dC, 5mdC:5mrC, 5mrC:5mrC:, and ss 5mrC). Blue residues form hydrogen bonds more prevalent in the reference system, orange residues form hydrogen bonds more prevalent in the systems with altered complementary strand, and pink residues have a system-dependent preference.

Fig. 12. Differences between the total non-bonded interactions (Coulomb and Van der Waals) mapped onto the TET2 structure for (A) ss 5mdC - ss 5mrC, (B) 5hmdC:5mdC - 5hmrC:5mdC, (C) 5fdC:5mdC - 5frC:5mdC, (D) 5cadC:5mdC - 5carC:5mdC, (E) 5mdC:5mdC - 5mdC:5mrC (F) 5mdC:5mdC - 5mrC:5mrC, and (G) 5mdC:5mdC - 5mdC:dC. Residues colored blue are favored in the first system and residues colored orange are favored in the second system. The target nucleotide is shown in black.

Fig. 13. The H1384 and K1462 residues, located in unstructured loops, that exhibit large changes in their total non-bonded interactions for comparisons between the 5mdC:5mdC system and the systems with altered complementary strands.

Table 1

Differences in hydrogen bonding between that exceed 20% of the total production time for the specified system.

Hydrogen acceptor	Hydrogen donor	Absolute difference (%)	System
ASN1387@OD1	5hmC Substrate	33.2950	5hmC and 5hmrC
ASP1384@OD2	ASN1387@ND2	29.3766	5fdC and 5frC
THR1393@OG1	ASN1387@ND2	24.8167	5fdC and 5frC
HIE1904@ND1	TRP1291@N	22.2400	5fdC and 5frC

Table 2

Inter-molecular interaction energy difference, $\Delta E_{T_{at}}^{Non-bond}$, for key residues between dC and rC systems.

	ss		ds			
	$E_{5ndC-5mrC}$	$E_{5hmdC-5mrC}$	$E_{5tdC-5trC}$	$E_{5adC-5earC}$	$\Delta E_{5hmdC-5mrC}^*$	
R1261 [†]	51.78 ± 1.50	4.80 ± 1.10	15.15 ± 0.88	- 51.56 ± 2.15	11.92 ± 1.84	
T1372	1.58 ± 0.15	0.47 ± 0.09	- 0.11 ± 0.07	- 8.11 ± 0.77	21.75 ± 0.06	
H1382 [‡]	1.23 ± 0.07	1.04 ± 0.21	0.78 ± 0.04	5.02 ± 0.13	1.94 ± 0.15	
D1384 [‡]	1.34 ± 1.18	- 13.75 ± 1.24	- 5.52 ± 0.56	45.11 ± 0.93	0.37 ± 2.16	
N1387	- 2.67 ± 0.20	0.64 ± 0.22	- 1.23 ± 0.13	0.61 ± 0.26	- 0.54 ± 0.29	
H1881 [‡]	- 0.52 ± 0.05	- 0.28 ± 0.14	- 0.37 ± 0.08	3.62 ± 0.06	0.06 ± 0.08	
R1896 [†]	6.10 ± 0.07	- 0.32 ± 0.09	- 0.03 ± 0.05	- 21.76 ± 0.27	- 0.01 ± 0.11	
S1898 [†]	- 0.08 ± 0.02	0.20 ± 0.02	0.01 ± 0.01	0.13 ± 0.02	- 0.33 ± 0.02	
Y1902	3.51 ± 0.18	0.87 ± 0.22	- 0.31 ± 0.14	1.57 ± 0.28	- 0.52 ± 0.26	
H1904	1.36 ± 0.29	1.63 ± 0.21	0.29 ± 0.51	- 0.21 ± 0.10	1.75 ± 0.94	

[†], denotes a residue that interacts with α-KG

[‡], signifies a residue that coordinates with Fe(II), and no symbol means that the residue is involved in DNA recognition; $E_{5hmdC-5mrC}$ column denotes that reference 5hmdC system corresponds to the T1372E mutation. All values are the average of each system ± avg. st. dev. in kcal mol⁻¹.