CHARACTERIZATION OF 5-METHYLCYTOSINE DIOXYGENASE TET2 AND RESCUE OF MUTANT TET2 ACTIVITY BY USING TURBO CO-SUBSTRATE

A DISSERTATION IN
Pharmaceutical Sciences
and
Molecular Biology and Biochemistry

Presented to the Faculty of the University of Missouri-Kansas City
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

by

CHAYAN BHATTACHARYA

MS, Cell and Molecular Biology
University of Missouri-Kansas City, 2021

M.Sc., Molecular Biology and Genetics
Presidency University, Calcutta, India, 2013

Kansas City, Missouri
2023

CHARACTERIZATION OF 5-METHYLCYTOSINE DIOXYGENASE TET2 AND RESCUE OF MUTANT TET2 ACTIVITY BY USING TURBO CO-SUBSTRATE

Chayan Bhattacharya, Candidate for the Doctor of Philosophy Degree

University of Missouri-Kansas City, 2023

ABSTRACT

Regulation of epigenetic transcription mediated by 5-methylcytosine (5mC) plays a critical role in eukaryotic development. Demethylation of these epigenetic marks is accomplished by sequential oxidation by ten-eleven translocation dioxygenases (TET1-3), followed by thymine-DNA glycosylase-dependent base excision repair. Inactivation of the TET2 gene due to genetic mutations or other epigenetic mechanisms is associated with poor prognosis in patients with diverse cancers, especially hematopoietic malignancies. Herein, we describe an efficient single-step purification of enzymatically active untagged human TET2 dioxygenase using cation-exchange chromatography. We further provide a liquid chromatography-tandem mass spectrometry (LC-MS/MS) approach that can separate and quantify the four normal DNA bases (A, T, G, and C), as well as the four modified cytosine bases (5-methyl, 5-hydroxymethyl, 5-formyl, and 5-carboxyl). This method can be used to evaluate the activities of wild-type and mutant TET2 dioxygenases.

In the mammalian genome, cytosine methylation predominantly occurs at CpG sites. In addition, several recent studies have uncovered extensive C5 cytosine methylation (5mC) at non-CpG (5mCpH, where H = A/C/T) sites. Little is known about the enzyme responsible for the active demethylation of 5mCpH sites. Using a highly sensitive and quantitative LC–MS/MS method, we demonstrated that human TET2, an iron (II)- and

2OG-dependent dioxygenase, which is a frequently mutated gene in several myeloid malignancies, as well as in a number of other types of cancers, can oxidize 5mCpH sites in double-stranded DNA in vitro. Similar to the oxidation of 5mCpG, the oxidation of 5mC at CpH sites produces 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) bases in DNA. After 5mCpG, which is the preferred substrate, TET2 prefers 5mCpC as a substrate, followed by 5mCpA and 5mCpT. Because the TDG/BER pathway can convert 5fCpH and 5caCpH to an unmodified cytosine base in DNA[1], our results suggest a novel demethylation pathway of 5mCpH sites initiated by TET2 dioxygenase.

TET isoforms (TET1-3) play a critical role in epigenetic transcriptional regulation. In addition, mutations in TET2 are frequently detected in patients with glioma and myeloid malignancies. TET isoforms can oxidize 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine via iterative oxidation. However, little is known about the preference for DNA length and configuration as the optimum substrates for TET isoforms. We used a highly sensitive LC-MS/MS-based method to compare the substrate preferences of the TET isoforms. To this end, four DNA substrate sets (S1, S2, S3, and S4) with different sequences were chosen. In addition, four different lengths of DNA substrates comprising 7-, 13-, 19-, and 25-mer nucleotides were synthesized in each set. Each DNA substrate was further used in three different configurations, that is, double-stranded symmetrically methylated, double-stranded hemi-methylated, and single-stranded single-methylated, to evaluate their effects on TET-mediated 5mC oxidation. We demonstrated that mouse TET1 (mTET1) and human TET2 (hTET2) had the highest preference for 13-mer dsDNA substrates. Increasing or decreasing the length of the dsDNA substrate reduced

the product formation. In contrast to their dsDNA counterparts, the length of the ssDNA substrates did not exhibit a consistent pattern of 5mC oxidation. Finally, we showed that the substrate specificity of TET isoforms correlates with their DNA-binding efficiency. Our results demonstrated that mTET1 and hTET2 prefer 13-mer dsDNA as a substrate over ssDNA.

The pathogenesis of malignant evolution has been linked to the epigenetic repression of tumor suppressor genes (TSG). One such epigenetic mechanism observed in myelodysplastic syndrome (MDS) is the acquired progressive methylation of CpG islands in gene promoters, leading to transcriptional repression. The TET family of hydroxylases/dioxygenases, which includes TET1-3, has recently been identified as iron (II)- and 2OG-dependent dioxygenases (2-OGDDs). *TET2* gene mutations occur frequently in myeloid malignancies, such as myelodysplastic syndrome (MDS), chronic myelomonocytic leukemia (CMML), myeloproliferative neoplasms, and secondary acute myeloid leukemia derived from these conditions. In cancer cells, the normal regulation of 2-OGDDs activity is disrupted, leading to changes in gene expression and epigenetic modifications. The 2OG analogs mimic the chemical structure of 2OG and modulate the activity of 2-OGDDs. 2OG analogs have been shown to restore the activity of clinically relevant 2-OGDDs and modulate the epigenetic landscape of cancer cells. The objective of this study was to develop effective strategies using 2OG analogs to enhance the activity of TET2 harboring point mutations at the R1896 residue. A library of 11 compounds was designed to mimic the chemical structure of the 2OG. By screening these compounds, eight 2OG analogs that could specifically rescue the mutant TET2 activity were identified. The catalytic activity of the TET2 R1896S mutant was enhanced by up to 90% compared with

that of the wild-type enzyme. Furthermore, we demonstrated that TET2 clinical mutations in the R1896A and R1896F residues could be rescued using this alternate co-substrate approach.

The crystal structure of the catalytic domain of TET2 revealed that the two zinc fingers bring together the cysteine-rich domain and the double-stranded β-helix fold domain, creating a compact catalytic domain. The cysteine-rich domain stabilizes DNA above the double-stranded β-helix fold domain, and the catalytic cavity allows for insertion of 5mC, with the methyl group positioned towards the catalytic Fe (II) for oxidation. This cysteine-rich domain is conserved among all three TET family members and catalyzes substrate oxidation. Numerous TET2 mutations are documented, primarily within the dioxygenase domain. However, there is limited information regarding the structural and functional effects of these TET2 missense variants. The TET2-DNA crystal structure identified 12 amino acid residues in the active site that interact with the DNA substrate. In this study, these residues were mutated to alanine to disrupt their interaction with DNA and to investigate the resulting oxidation patterns of CpG and non-CpG substrates. Mutated residues affected the activity of hTET2, particularly on CpG substrates. Notably, three active-site mutations (Y1295A, R1302A, and H1904A) showed significantly higher oxidation efficiencies on CpG substrates than the wild-type enzyme. By performing alanine scanning on the DNA-interacting residues, it was found that the catalytic cavity of TET2 can accommodate different DNA sequences, and stability of the TET2-DNA complex relies on multiple amino acid residues. This suggests that a single-point mutation can be compensated for without rendering the enzyme catalytically inactive. However, further in vivo studies and molecular dynamics simulations are required to confirm these findings.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "Characterization of 5-methylcytosine dioxygenase Tet2 and rescue of mutant Tet2 activity by using turbo co-substrate", presented by Chayan Bhattacharya, candidate for the Doctor of Philosophy degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Mridul Mukherji, D. Phil,
Primary discipline Advisor
Division of Pharmacology and Pharmaceutical Sciences

Gerald J. Wyckoff, Ph.D.
Co-Discipline Advisor
Division of Pharmacology and Pharmaceutical Sciences

Kun Cheng, Ph.D.
Division of Pharmacology and Pharmaceutical Sciences

Ryan Mohan, Ph.D.
School of Biological and Chemical Sciences

Xiaolan Yao, Ph.D.
School of Biological and Chemical Sciences

CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGEMENTS

During my doctoral studies at UMKC, I have been supported by several people, and it is my pleasure to thank them. First, my supervisor, Dr. Mridul Mukherji, gave this opportunity and accepted me into his research group. I would like to thank him for his guidance, support, expertise, understanding, and assistance. I appreciate his vision and invaluable suggestions while writing the research papers, and I would also like to thank him for his constant push to work harder and perform better science.

I express my sincerest thanks to Drs. Gerald J. Wyckoff, Kun Cheng, Ryan Mohan and Xiaolan Yao for serving on my supervisory committee and their constant support and encouragement throughout this journey and especially through the difficult times. I would also like to thank Dr. William Gutheil for helping us with LC-MS/MS analysis and providing valuable suggestions to overcome scientific challenges.

I am especially thankful to staff members of School of Pharmacy (SOP) Joyce Johnson, Sharon Self, Norma Aguierre, Dr. Sarah Beach, Jane Poe, for their continued assistance.

I sincerely appreciate the expertise, insights, and tireless support of Dr. Navid J. Ayon, Dr. Amardeep Sharma, Dr. Shivani Gargvanshi, and Vidit Minda. They have been a constant source of support whenever we need help with mass spectrometry. I would also extend my gratutute to my DPPS friends Hari Mani Krishna Veerubhotla, Mayank Sharma, Shivani Gargvanshi, Amardeep Sharma and my seniors Navid J. Ayon and Abhirup Mandal. Their friendships and shared experiences shaped me into the person I am today. I am grateful for the memories we have created and the valuable life lessons we have learned. Their unwavering belief in me has been instrumental to my personal and academic growth.

I would like to express my heartfelt appreciation to Dr. Aninda S. Dey, my invaluable friend, and my lab mate. His unwavering support, insightful discussions, and technical assistance throughout this research endeavor have been instrumental in its success. His dedication and collaborative spirit enriched my scientific journey. I am grateful for his friendship and invaluable contributions to my Ph.D. journey.

I want to extend my sincere gratitude to my parents, who were the sole reason for my achievements, and this would not have been possible without your support, blessings, and sacrifices. I thank you for everything and I am deeply indebted to you.

CHAPTER 1

INTRODUCTON

1.1 2-oxoglutarate dependent dioxygenases and their role in epigenetics**:**

2-Oxoglutarate dependent dioxygenases (2-OGDD) are a superfamily of enzymes that play diverse roles in many biological processes. These enzymes are involved in the regulation of hypoxia-inducible factor-mediated adaptation to hypoxia, extracellular matrix formation, epigenetic regulation of gene transcription, and regulation of DNA and histone demethylation[2]. They are also involved in biosynthesis of secondary metabolites, such as meroterpenoids in fungi[3]. The 2-OGDD superfamily is the second largest enzyme family in the plant genome, and its members are involved in various plant-specific processes, such as hormone biosynthesis, cell wall modification, and stress response[4].

The 2-OGDD enzymes are characterized by their dependence on 2-oxoglutarate and molecular oxygen for their activity. These enzymes catalyze hydroxylation of a wide range of substrates, including proteins, nucleic acids, and small molecules. The hydroxylation reaction is coupled to the decarboxylation of 2-oxoglutarate to succinate and incorporation of one oxygen atom into the substrate. The reaction is dependent on the presence of iron and ascorbate, which act as cofactors for the enzyme[5].

The 2-OGDD enzymes are involved in many biological processes, including cancer. Dysregulation of 2-OGDDs in cancer, by genetic or epigenetic alterations, can lead to changes in expression of genes involved in cell proliferation, differentiation, and survival. For example, 2-OGDD enzyme, prolyl hydroxylase domain-containing protein 2 (PHD2), is involved in regulation of hypoxia-inducible factor (HIF) stability. Under normoxic conditions, PHD2 hydroxylates HIF-α, leading to its degradation by the

proteasome[3]. Under hypoxic conditions, PHD2 activity is inhibited, leading to the stabilization of HIF-α and the activation of genes involved in angiogenesis, glucose metabolism, and cell survival[5]. Dysregulation of 2-OGDDs in cancer can lead to changes in expression of genes involved in cell proliferation, differentiation, and survival. For example, the 2-OGDD enzyme lysine demethylase 5A (KDM5A) is overexpressed in various types of cancers, including breast, lung, and prostate cancers. KDM5A demethylates histone H3 lysine 4 (H3K4), leading to activation of genes involved in cell proliferation and survival[6]. The 2-OGDD enzyme lysine demethylase 6A (KDM6A) is a tumor suppressor that is frequently mutated or deleted in various types of cancer, including ovarian, bladder, and pancreatic cancers. KDM6A demethylates histone H3 lysine 27 (H3K27), leading to repression of genes involved in cell proliferation and survival[7].

The 2-OGDD enzymes are also involved in regulation of DNA and histone demethylation. Enzymes, such as the ten-eleven translocation (TET) family of DNA demethylases and the Jumonji C (JmjC) family of histone demethylases, catalyze hydroxylation of methylated cytosine and lysine residues, respectively. The hydroxylation reaction leads to removal of the methyl group and generation of formaldehyde and succinate as byproducts. Demethylation of DNA and histones is important for regulation of gene expression and cellular differentiation[8].

In conclusion, 2-oxoglutarate dependent dioxygenases are a superfamily of enzymes that play diverse roles in many biological processes, including regulation of hypoxia-inducible factor-mediated adaptation to hypoxia, extracellular matrix formation, epigenetic regulation of gene transcription, and regulation of DNA and histone demethylation. Dysregulation of 2-OGDDs in cancer cells can lead to changes in expression of genes

involved in cell proliferation, differentiation, and survival. These enzymes are also involved in biosynthesis of secondary metabolites in fungi and in various plant-specific processes. The hydroxylation reaction catalyzed by 2-OGDD enzymes is dependent on the presence of iron and ascorbate and is coupled to decarboxylation of 2-oxoglutarate to succinate and incorporation of one oxygen atom into the substrate[9].

1.2 Structure of 2-oxoglutarate dependent dioxygenases**:**

The structure of 2OGDDs is important for their functions. These enzymes have a conserved double-stranded beta-helix (DSBH) fold that supports residues coordinating iron. The DSBH fold comprises eight anti-parallel β-strands, forming a β-sandwich structure composed of two four-stranded anti-parallel β-sheets (Figure 1). The DSBH fold is common to all 2-OGDD enzymes and is responsible for the ability of these enzymes to activate oxygen in an iron-dependent manner. The DSBH fold is also found in enzymes other than 2-OGDDs, such as ribonucleotide reductases and DNA polymerases. The DSBH fold is thought to provide a scaffold for binding of metal ions and substrates, as well as for protein-protein interactions[6,7,10,11].

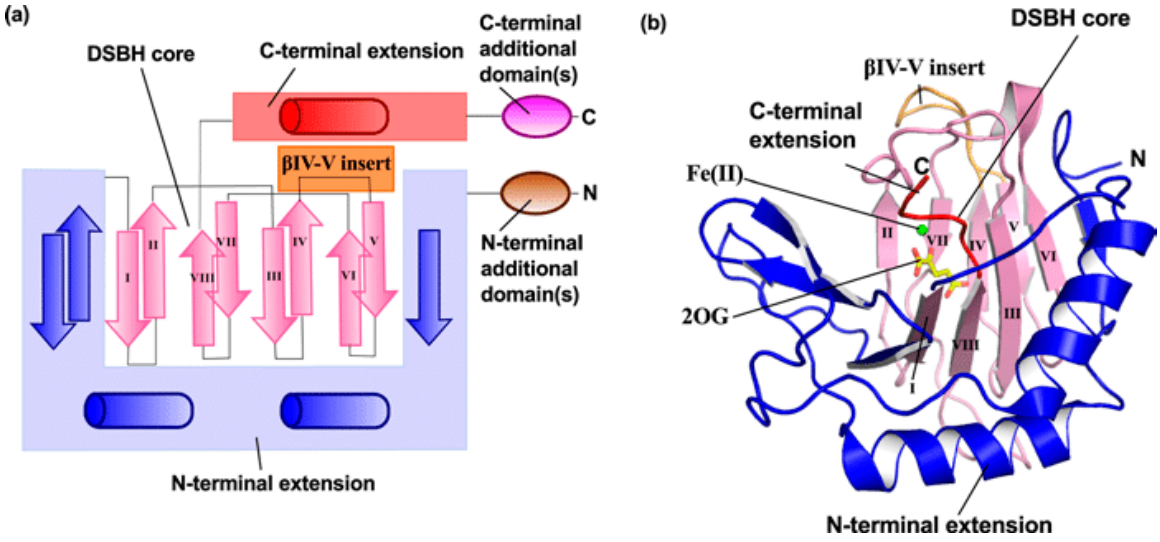


Figure 1: A general active site structure of 2-OG dependent dioxygenases showing the DSBH core (e.g; AlkB, PDB ID: 3BIE)[12]. Panel (a) represents the orientation of β-sheets in the DSBH core (in pink) along with the N-terminal and C-terminal extensions (in blue). Panel (b) depicts three-dimensional structure of the DSBH active site core bound to Fe (II) and 2OG.

The highly conserved HXD/E...H motif consists of two histidine residues and one aspartic or glutamic acid residue separated by a variable number of amino acids, followed by another histidine residue. The HXD/E motif is involved in coordinating the iron ion in the active site of the enzyme, which is essential for enzyme activity. The iron ion is coordinated by two histidine residues and an aspartic or glutamic acid residue, which act as ligands, and the second histidine residue, which stabilizes the iron ion. The HXD/E motif is conserved in all 2-OGDD enzymes, including prolyl hydroxylases (PHDs), lysine demethylases (KDMs), and the ten-eleven translocation (TET) family of DNA demethylases. This motif is important for the ability of enzymes to catalyze hydroxylation of a wide range of substrates, including proteins, nucleic acids, and small molecules. The hydroxylation reaction is coupled with the decarboxylation of 2-oxoglutarate to succinate and the incorporation of one oxygen atom into the substrate. The reaction is dependent on the presence of iron and ascorbate, which act as enzyme cofactors[10,13,14].

In conclusion, the structure of 2OGDDs is characterized by a conserved double-stranded β-helix fold that supports residues coordinating iron. The DSBH fold is common to all 2-OGDD enzymes and is responsible for the ability of these enzymes to activate oxygen in an iron-dependent manner. The DSBH fold is also found in other enzymes that are not 2-OGDDs and is thought to provide a scaffold for the binding of metal ions and substrates, as well as for the formation of protein-protein interactions.

1.3. Discovery of TET enzymes:

The discovery of the ten-eleven translocation (TET) family of enzymes is a major breakthrough in epigenetics. TET enzymes are known for their common translocation in

cancers[15]. The TET family comprises three members: TET1, TET2, and TET3. These enzymes indirectly remove epigenetic marks on cytosine by converting 5mC to 5-hydroxylcytosine, which can then be further oxidized to 5-formylcytosine and 5-carboxylcytosine. TET enzymes are α-ketoglutarate (αKG)-dependent dioxygenases that oxidize 5-methyldeoxycytosine, thereby participating in passive and active demethylation[16,17]. The catalytic domains of TET enzymes contain specific regions and binding sites that are crucial for their function. It consists of multiple functional domains, including the CXXC, cysteine-rich (Cys-rich), and double-stranded β-helix (DSBH) domains (Figure 2). The CXXC domain, found at the N-terminus of TET enzymes, is responsible for DNA recognition and binding. This domain plays a critical role in targeting TET enzymes to specific genomic regions, allowing them to modify their DNA methylation patterns. The Cys-rich domain, located within the catalytic domain, contains conserved cysteine residues that coordinate binding of iron (Fe2+) ions. These iron ions are essential for hydroxylation reactions performed by TET enzymes. The DSBH domain, also known as the J-binding domain, is involved in binding of α-ketoglutarate (α-KG), a co-substrate required for TET enzyme activity. Binding of α-KG is necessary for catalytic conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and subsequent steps in the active DNA demethylation process. Furthermore, all three TET enzymes share a similar overall three-dimensional structure, characterized by a central catalytic core surrounded by various functional domains. This structural conservation is consistent with the enzymatic activities and biological functions of 2OGDDs.

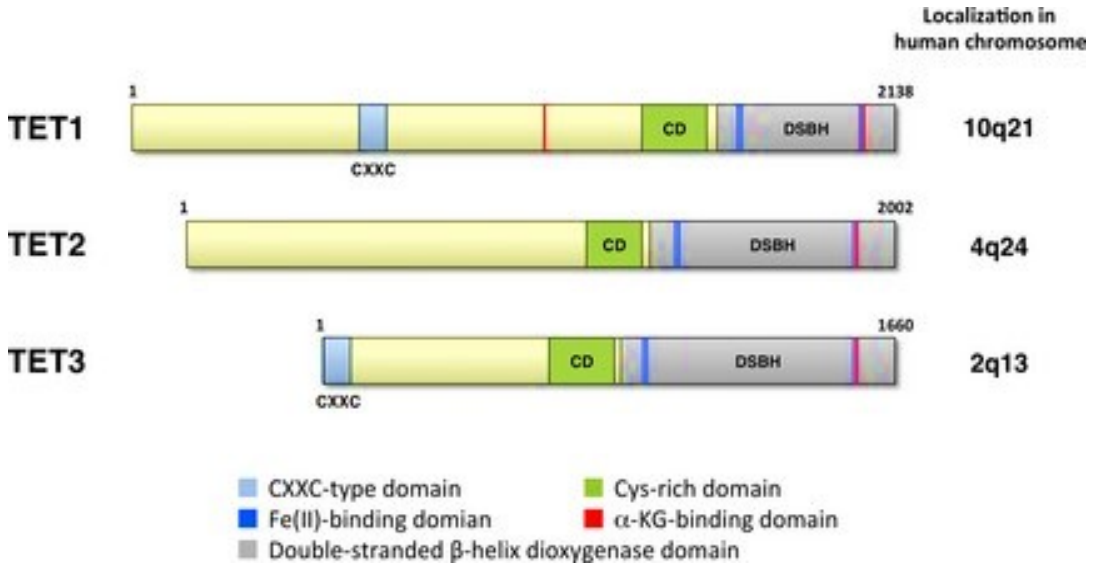


Figure 2: Domain architecture of human TET1, TET2 and TET3 along with their chromosomal location in the human genome[18]. All three members of the TET family contains a Cys-rich domain (shown in green) and a DSBH domain (shown in grey) to form the catalytic core. Only TET1 and TET3 but not TET2 contains the CXXC domain at the N-terminal. Within the DSBH core, Fe (II) binding site (HXD) has been represented in blue whereas, the 2-OG binding site (RXS) is represented in red.

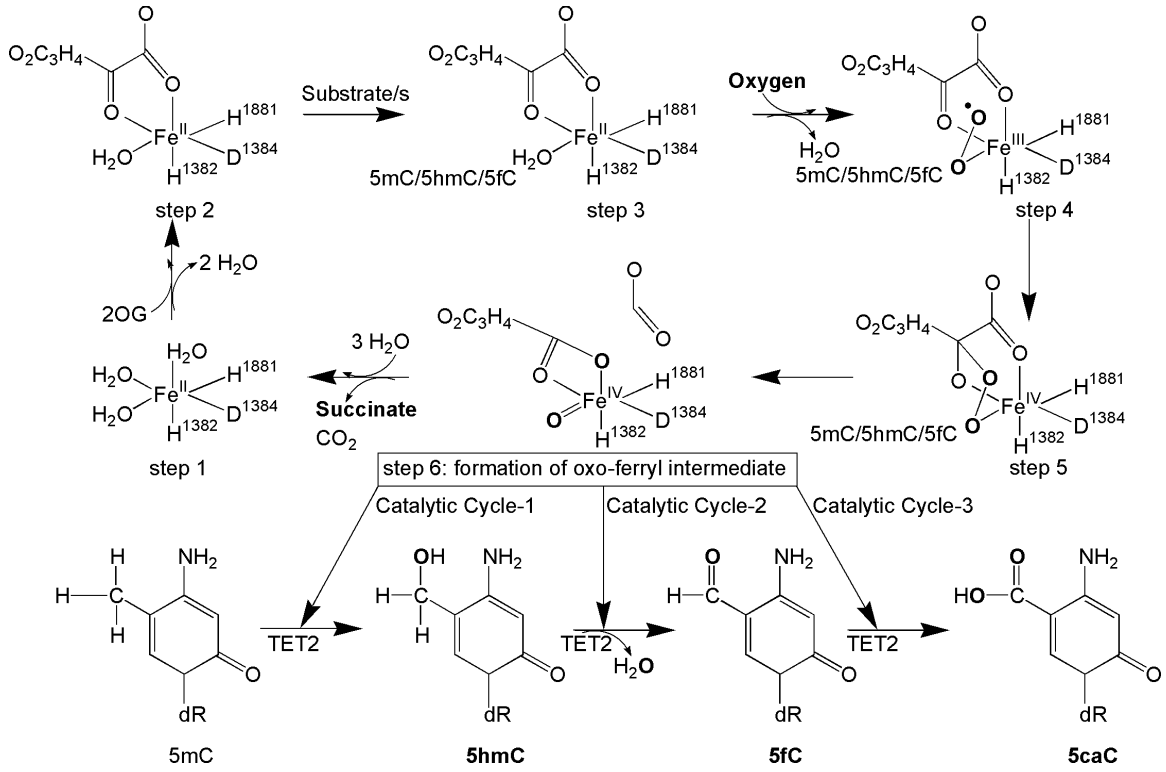


Figure 3: A catalytic cycle of Fe (II), 2-OG-dependent dioxygenases showing sequential binding of the cofactors and substrates leading to the formation of reactive oxo-ferryl [Fe (IV)=O] intermediate. The two histidines and one aspartate residues, which form the active site core, is called the 2-His-1-carboxylate facial triad. The conversion of 5mC to 5caC is mediated by three iterative catalytic/oxidation cycles by TET family of dioxygenases generating 5hmC and 5fC as intermediates.

TET enzymes are involved in immune cell development, affecting stem cell self-renewal and lineage commitment to terminal differentiation. Loss-of-function mutations in genes encoding TET enzymes occur frequently in hematopoietic malignancies. In the oxidation reaction, TET proteins coordinate one oxygen atom from the molecular oxygen (O2) to the hydroxyl group of the substrate (hydroxylation) and the other oxygen atom to 2-OG, leading to the formation of succinate and carbon dioxide[19–21].

1.4. Role of 5-methylcytosine in epigenetics:

Cytosine methylation is a process in which a methyl group is added to the 5th carbon of the cytosine ring of DNA. This process is catalyzed by DNA methyltransferase enzymes (DNMTs) and occurs at the cytosine bases of eukaryotic DNA, which are converted to 5-methylcytosine. The methylation of cytosine residues can affect gene expression and cellular differentiation. When DNA is methylated, nearby histones are deacetylated, resulting in compounded inhibition of transcription (Figure 4). Conversely, DNA demethylation can lead to increased gene expression[22]. Cytosine methylation can also affect the binding of transcription factors (TFs) to DNA[23,24]. Methylation of cytosine in a gene regulatory element can change DNA conformation and affect the binding of TFs. Methylation can also provide a mechanism for regulating gene expression by directly affecting the binding. Methylation changes DNA structure and dynamics, and high levels of methylation can induce cooperative effects[25,26].

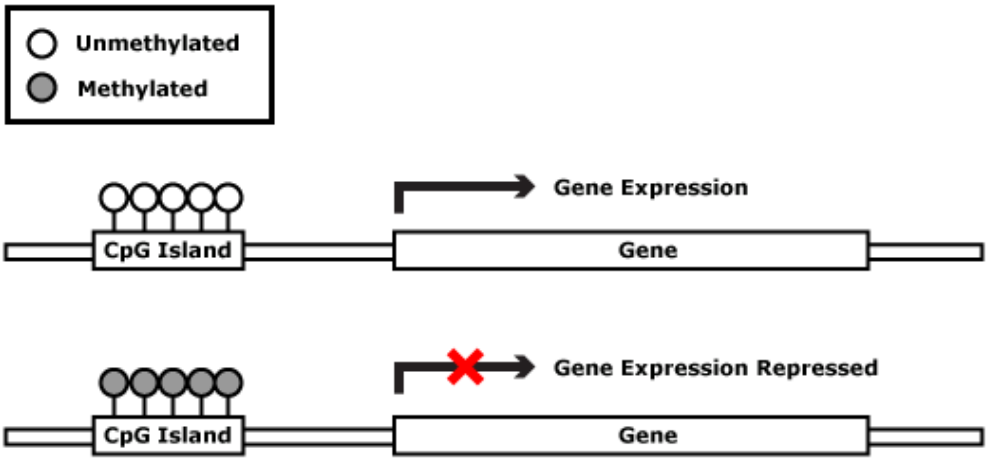


Figure 4: Effect of cytosine methylation on DNA expression. A CpG island located upstream of a gene promoter can regulate the expression of the downstream coding region. Hypermethylated CpG islands can repress gene expression, whereas the removal of methylation marks can reinitiate the transcription of downstream genes.

DNA methylation decreases the propensity of DNA to form nucleosomes and can affect the arrangement of nucleosomes near the beginning of transcription. Methylation has the potential to modify the configuration of nucleosomes, resulting in more compact wrapping of DNA around histones. Additionally, methylation can influence various aspects of nucleosome behavior, including the assembly, movement, stability, positioning, and three-dimensional structure of chromatin[27–29]. Methylation can also alter the structure of nucleosomes, leading to tighter wound configuration of DNA around histones[30,31].

1.5 Role of TET2 in developmental diseases and cancer:

TET2-mediated DNA demethylation influences the accessibility of transcription factors to target gene promoters either by directly affecting their binding sites or by remodeling the chromatin structure[32–34]. TET2 loss impairs normal gene expression programs, leading to altered cellular function and disrupted hematopoiesis. Secondly, TET2 interacts with other epigenetic modifiers and transcription factors to maintain chromatin integrity and regulate gene expression. For instance, TET2 collaborates with Polycomb Repressive Complex 2 (PRC2) to ensure appropriate gene silencing and differentiation of hematopoietic cells[35,36]. TET2 mutations can disrupt this interaction, resulting in abnormal gene silencing and impaired cellular differentiation.

The tumor suppressor role of TET2 is evident in various cancer types, particularly hematopoietic malignancies. TET2 mutations and deletions have been identified in myeloid malignancies such as myelodysplastic syndrome (MDS), myeloproliferative neoplasm (MPN), and acute myeloid leukemia (AML)[37,38]. These mutations often lead to loss of TET2 function, resulting in abnormal DNA methylation patterns and altered gene

expression profiles (Figure 5). Dysregulation of key genes involved in cell proliferation, differentiation, and apoptosis contributes to the initiation and progression of these cancers. Furthermore, TET2 mutations are associated with poor prognosis and adverse clinical outcomes in myeloid malignancies[39,40]. TET2-deficient cells exhibit increased self-renewal capacity and skewed differentiation potential, leading to expansion of abnormal hematopoietic stem/progenitor cells[34,41]. Dysregulated hematopoiesis promotes clonal evolution and development of malignant clones. Although TET2 mutations are most observed in hematopoietic malignancies, recent studies identified TET2 alterations in solid tumors. TET2 mutations are observed in various cancers, including breast, lung, colon, and bladder[42,43]. These findings suggest a broader role for TET2 in tumorigenesis of hematopoietic tissues.

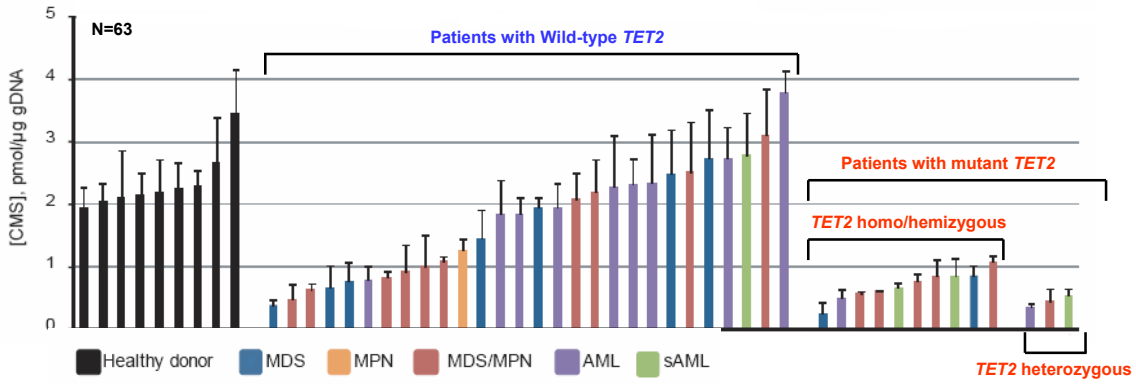


Figure 5: Effect of TET2 mutation on the genomic level of 5hmC [44]. Patients with mutant TET2, regardless of the constellation, had reduced levels of 5hmC as compared to controls, as did a proportion of patients with WT TET2. The 'X' axis represents patients with and without TET2 mutations whereas the 'Y' axis represents presence of 5hmC marks in the genomic DNA.

Furthermore, TET2 has been implicated in other diseases including cardiovascular disorders, neurodegenerative diseases, and immune dysregulation[45,46]. Altered TET2 expression or function can influence the pathogenesis of these conditions by affecting DNA methylation patterns and gene expression profiles. During embryonic development, TET2 plays a critical role in maintaining cellular pluripotency and in orchestrating lineage specification. TET2 helps establish and maintain DNA methylation patterns that are essential for proper cell fate determination and differentiation. Dysregulated TET2 expression or function can disrupt these processes and lead to developmental abnormalities. For instance, TET2 mutations are observed congenital heart defects, neural tube defects, and other structural anomalies[47,48].

In the context of neurodevelopmental disorders, TET2 has gained attention because of its involvement in regulation of neuronal development and synaptic plasticity. TET2-mediated DNA demethylation is crucial for proper expression of genes involved in neuronal maturation, synaptic formation, and synaptic plasticity. Disruptions in TET2 function, caused by mutations or altered expression levels are associated with neurodevelopmental conditions, including autism spectrum disorders (ASDs), intellectual disability, and schizophrenia[49,50].

Furthermore, TET2 is implicated in regulation of immune and inflammatory responses during development. Immune dysregulation is observed in various developmental disorders, including autoimmune diseases and immunodeficiency. TET2 influences differentiation and function of immune cells through its effect on DNA methylation and gene expression. Alterations in TET2 activity may contribute to immune dysregulation and development of immune-related disorders[51,52].

The precise mechanisms by which TET2 dysfunction contributes to developmental diseases remain unclear. However, the role of TET2 in regulating DNA methylation dynamics and gene expression is critical for normal development. Disruptions in TET2-mediated epigenetic regulation can result in altered cellular differentiation, impaired neural development, and aberrant immune response. Understanding the role of TET2 in developmental diseases holds great promise for the advancement of diagnostic and therapeutic strategies. Further research is needed to uncover the specific molecular pathways influenced by TET2 and to identify potential therapeutic targets to mitigate the impact of TET2-related developmental disorders.

In conclusion, TET2 plays a critical role in maintaining normal cellular functions through its involvement in DNA demethylation and gene regulation. TET2 mutations and deletions are tumor suppressor genes that contribute to development and progression of various cancers, particularly those of the hematopoietic system. Additionally, TET2 alterations have been identified in non-hematopoietic tumors and other diseases, indicating their broad significance in disease pathogenesis. Further research on the molecular mechanisms underlying TET2 function will enhance our understanding of its role in cancer and disease, potentially leading to the development of targeted therapies.

1.6 Aim of the thesis

The aim of the dissertation is functional characterization of TET2 mutations and rescue of mutant TET2 activity by 2OG analogs. In chapter 2, we developed a one-step cation exchange chromatography method for hTET2 and optimized its *in vitro* assay conditions. In chapter 3, we determined substrate preference of human TET2 (hTET2) and investigated the effect of substrate DNA length and sequence on hTET2 mediated 5mC oxidation. By systematically varying length and configuration of DNA substrates, we

demonstrated that mTET1 and hTET2 have an optimum activity with 13-mer double stranded DNA (dsDNA) substrates. Although hTET2 is versatile enough to recognize single stranded DNA (ssDNA) and dsDNA both as its substrates, dsDNA clearly is a preferred substrate when it comes to higher oxidation products such as 5fC and, 5caC. Using a sensitive LC–MS/MS-based method, we also demonstrated multiple lines of evidence that the human TET2 can oxidize 5mCpH sites in DNA. In chapter 4, we rescued TET2 clinical mutations in R1896S, R1896A, and R1896F residues using an alternate co-substrate approach. We identified three 2OG analogs in which the 5-carboxylate group was replaced with an aliphatic chain, and these compounds enhanced the activity of hTET2 R1896S by up to 70-80% compared to the wild-type enzyme. Finally, in chapter 5, we performed alanine scanning on hTET2 active site residues and characterized these mutant TET2 activity on CpG as well as non-CpG substrates. TET isoforms are particularly important in embryonic development and stem cell maintenance, where they regulate the dynamic changes in DNA methylation patterns that are critical for proper cell fate specification and tissue differentiation. Therefore, understanding the significance of TET enzymes has important implications for our understanding of both normal development and disease. Overall, this dissertation contributes to our understanding of TET mediated 5mC oxidation in various length and sequence of DNA substrates and also exhibits a strategy to enhance mutant TET2 activity by structurally modified 2OG analogs.

CHAPTER 2

CLONING, EXPRESSION, PURIFICATION OF TET METHYLCYTOSINE DIOXYGENASE 2 AND DEVELOPMENT OF AN *IN VITRO* ASSAY SYSTEM

2.1 Introduction

The predominant site for methylation (5mCpG) in mammalian genomes is the C5 position of cytosine bases within CpG dinucleotides[53]. Recent studies have also discovered extensive cytosine methylation (5mC) at non-CpG sites (5mCpH, where H = A, T, or C)[54,55]. The modification of 5mC acts as a transcriptional silencer at endogenous retrotransposons and gene promoters, and plays crucial roles in various processes, such as X chromosome inactivation, gene imprinting, nuclear reprogramming, and tissue-specific gene expression[56–60]. DNA methyltransferases carry out cytosine methylation at the C5 position, whereas TET1-3 5mC oxidases initiate the removal of 5mC marks by converting them into sequential oxidation steps:5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC)[61–65]. Thymine-DNA glycosylase then replaces 5fC or 5caC with unmodified cytosine via the base excision repair pathway[63,66].

In humans, the TET2 gene is frequently mutated in various hematopoietic malignancies, including myelodysplastic syndromes (MDS), myeloproliferative neoplasms (MDS-MPN), and acute myeloid leukemia (AML) originating from MDS and MDS-MPN[41,67,68]. Patients with TET2 mutations exhibit lower levels of 5hmC modification in their bone marrow DNA than those with the wild-type *TET2* gene. *TET2*-knockout mouse models have been developed to understand the role of TET2 in normal hematopoiesis and myeloid transformation[34,45,47,69]. These mice initially show normal and viable characteristics, but develop diverse hematopoietic malignancies as they age, leading

to early death. Studies in these mouse models have revealed the significance of wt-TET2 in normal hematopoietic differentiation, demonstrating that both heterozygous hematopoietic stem cells (TET2+/- HSCs) and homozygous TET2-/- HSCs result in various hematopoietic malignancies. Therefore, haploinsufficiency of TET2 dioxygenase alters HSC development, leading to hematopoietic malignancies[18,36,45–47,70].

Similar to TET2 mutant mice, patients with leukemia exhibit haploinsufficiency of TET2 dioxygenase activity. These patients have mostly heterozygous somatic mutations, including frameshift and nonsense mutations scattered throughout the TET2 gene body, while missense mutations tend to cluster in the dioxygenase domain[71,72]. However, there is limited characterization of both wild-type and mutant forms of TET2 in the literature, mainly because of challenges in producing TET2 dioxygenase and conducting its assay. In this study, we present a straightforward purification method for native TET2 dioxygenases using ion exchange chromatography. Additionally, we optimized and utilized a quantitative LC-MS/MS assay to measure the enzymatic activity of the native TET2 dioxygenase.

2.2 Materials and methods

Cloning of GST tagged TET2: A human TET2 (1129-1936) clone with residues 1481-1843 replaced by a 15-residue GS linker (designated as TET2 1129-1936 Δ1481-1843 DNA) was obtained as we have previously described[73]. Briefly, after PCR and purification on the agarose gel, the TET2 1129-1936 Δ1481-1843 DNA fragment was cloned into the pJET1.2 vector and verified by DNA sequencing.

Forward primer sequence:

5’GAATTCCATATGTCTGTTCTCAATAATTTTATAG3’

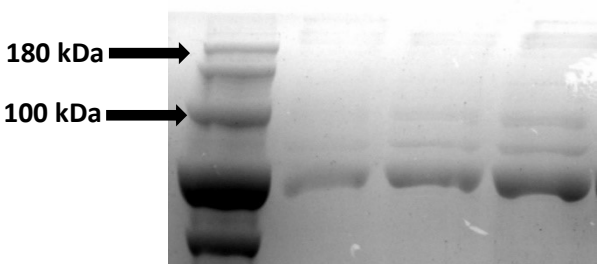
Reverse primer sequence:

5’CTCGAGGCGGCCGCGTCGACTCAGCCATACTTTTCACAC3’

To produce the GST-tagged human TET2 protein, the minimal catalytic domain of human TET2 (TET2 1129-1936 Δ1481-1843 DNA) was inserted into pGEX4T-1 vector using EcoR1 and Xho1 restriction enzymes. The final recombinant clone was sequence verified. The recombinant vectors were transformed into *E. coli* BL21(DE3) cells for protein expression. Cells were grown in Luria-Bertani broth containing 100 μg/ml ampicillin at 37°C, until the $O.D._{600}$ of the culture reached 0.8. The optimized condition for recombinant protein expression was found to be 0.25 mM IPTG induction of bacterial culture ($O.D._{600}$ = 0.8) for 16 hrs at 17°C.

Purification of GST-TET2 protein: The recombinant TET2 protein was purified from 18 liter of culture. Bacterial cells were pelleted and either stored at –80°C or lysed immediately for protein purification. This step onwards, all the steps of protein purification were performed on ice. The cells were resuspended in 30 ml of lysis buffer (PBS-10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, pH 7.5) and sonicated at an intensity of 20 for 6 mins with 6 pulses every 1 minute (Sonic Dismembrator 550, Fisher Scientific). The lysate was spun down at 4700 rpm for 60 min (Beckman J2-HS centrifuge) and the soluble fraction was collected and filtered through 0.45μ filter. Glutathione Agarose resin was used for purification of GST tagged proteins. The resin (100 ml) was packed into a XK26/20 FPLC column (Pharmacia now GE Healthcare, Piscataway, New Jersey) and equilibrated in 10 bed volumes of wash buffer (PBS-10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, pH 7.5) by a Akta FPLC system (Pharmacia/GE Healthcare). Soluble fraction was loaded on the equilibrated column at 0.2ml/min of flow rate. Following this, the column was washed with 5-10 bed volumes of wash buffer (flow rate 1ml/min) till the flow through was clear. Elution of the bound protein was carried out using a gradient (flow rate 1 ml/min)

from the wash buffer to the elution buffer (50 mM Tris-HCl, 33 mM Reduced Glutathione, pH 7.5). 0 to 100 % of elution buffer was reached in 10 min, followed by holding at 100% elution buffer for 90 min and then at 0% elution buffer for 30 min. Samples of cell lysate before and after column loading, along with all the wash and elution fractions (30 ml each) were collected and analyzed by SDS-PAGE (Figure 6). Selected fractions containing the desired protein were concentrated to 20 ml using Labconco Freeze dry system, Freezone 4.5 (Kansas City, Missouri).



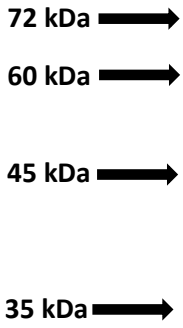


Figure 6: Purified fractions of GST-TET2 on a 10% SDS-PAGE gel. Lane 1 shows the molecular marker whereas, lanes 2, 3, and 4 are different fractions of the purified GST-TET2 protein.

Purification of untagged TET2 dioxygenase: hTET2 catalytic domain ((TET2 1129-1936, 1481-1843) was already cloned already into pDEST14 vector by site specific recombination technique. This study was performed previously by other laboratory members. The pDEST14 expression vector containing untagged hTET2 dioxygenase (TET2 1129-1936, Δ1481-1843) was transformed into chemically competent E. coli BL21 (DE3) cells. Following transformation, cells were plated on LB agar containing ampicillin. Isolated colonies were chosen from LB agar plates and grown in LB media containing ampicillin. The bacterial cells were harvested for plasmid DNA isolation. Transformation of recombinant DNA was confirmed by DNA sequencing.

Purification of the untagged hTET2 was performed using cation-exchange chromatography. The isoelectric point of the hTET2 catalytic domain is 7.49. Thus, we used MOPS buffer pH 6.5 as the mobile phase, where hTET2 was weakly positively charged. SP Sepharose, a strong cation exchanger, was used as the stationary phase. Binding was performed on pre-equilibrated SP Sepharose for 2 h, followed by 10 column volumes of washing (MOPS pH 6.5). Finally, the protein of interest was eluted using 300mM NaCl. Purified fractions of the protein (hTET2 M. W= 54.6 kDa) were subjected to SDS-PAGE analysis (Figure 7).

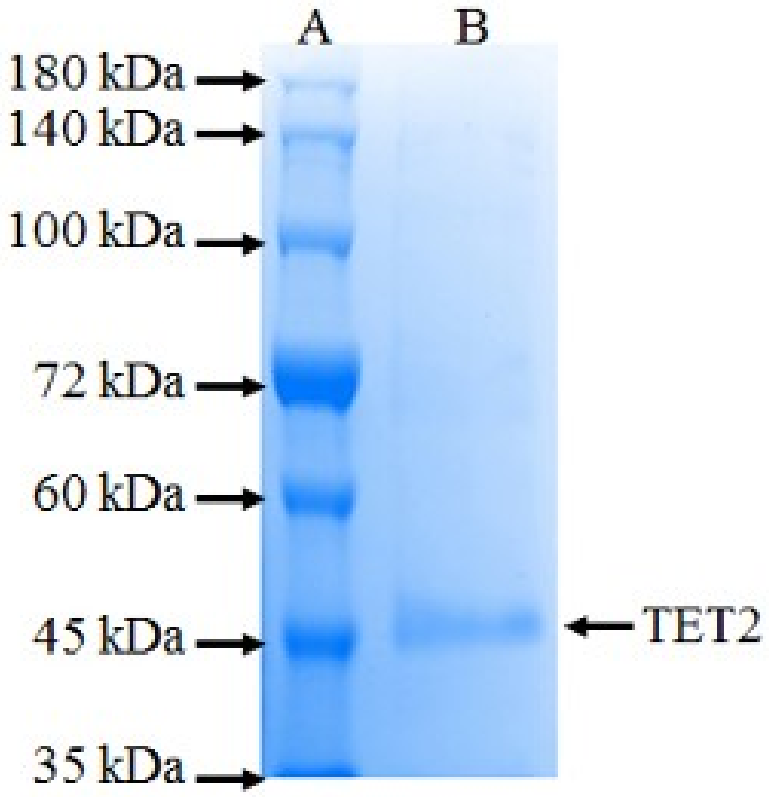


Figure 7: Purified untagged hTET2 fraction (M. W= 54.6 kDA) on a 10% SDS-PAGE gel. Lane A is the marker and lane B is the purified fraction of hTET2 (54 Kda).

5mC oxidation reactions: The TET2-mediated *in vitro* oxidation reactions were performed using a 25-mer double stranded symmetrically methylated DNA. dsDNA substrates were generated by hybridizing single stranded oligomers (sense strand: 5'-AGCCCGCGCCG/iMe-dC/GCCGGTCGAGCGG-3', antisense strand: 5'-CCGCTCGACCGGCG/iMe-dC/GGCGCGGGCT-3') on a thermal cycler using a stepdown PCR method (95 °C to 37 °C with 5 °C stepwise decrease and a hold for 5 minutes at every step). The *in vitro* TET enzymatic reactions were performed in a reaction mixture containing 50 mM HEPES (pH 8.0), 75 μM FeSO4, 1 mM 2OG, 2 mM ascorbate, 1 mM ATP and 1 mM DTT in 100 μl of total volume. An equal amount of DNA substrate (0.2 pM) and purified TET2 enzyme (20 μM) were added to each reaction mixture. After 1 hour of incubation at 37 °C, TET2 catalyzed oxidation reactions were quenched with 5 μl of 500 mM EDTA. Total DNA from the reaction mixture was separated using the superior oligo purification kit (Superior Scientifics, Lenexa, KS) following manufacturer's protocol. The isolated DNA (20 μl) was digested with 2 units of DNase I and 60 units of S1 nuclease in 40 μl at 37 °C for 12 hours to produce individual nucleotide monophosphates. Following the digestion, 2 units of calf intestinal alkaline phosphatase (CIAP) were added in the samples followed by incubation for an additional 4 hours at 37 °C to remove the terminal phosphate groups to obtain nucleosides. The nucleosides mixture weas quantified by LC-MS/MS. LC-MS/MS assay was developed and performed by Navid J Ayon and Aninda S. Dey.

2.3 Results and discussion

Dynamic modulation of 5mC in DNA by TET-family dioxygenases plays a crucial role in epigenetic transcriptional regulation[47,49,52,71]. TET2 dioxygenase is commonly

mutated in hematopoietic malignancies. To investigate the function of TET2 in normal development and disease, we cloned its minimal catalytically active domain into pGEX4T1 vector and expressed it in E. coli. GST tagged hTET2 was purified by affinity chromatography. Untagged TET2 dioxygenase was produced in E. coli BL21 (DE3) cells, constituting approximately 5% of the total soluble protein, as determined by SDS-PAGE analysis. Given the relatively high isoelectric point (~7.49) of the catalytic domain of TET2 compared to that of most native E. coli proteins, we devised an efficient purification process employing cation exchange chromatography, resulting in >90% purity of the TET2 enzyme in a single step[74–76].

Mutations in TET2 are among the most frequently observed genetic alterations in patients with diverse hematopoietic malignancies. Numerous TET2 mutations, including nonsense, frameshift, and missense mutations, have been identified in patients[67,77–80]. Patients harboring TET2 mutations exhibit reduced levels of genomic 5hmC in the bone marrow compared to those harboring wild-type TET2. Experimental studies using mutant TET2 knock-in replicated the impact of these mutations on 5hmC levels in transfected cells. Insights from TET2-knockout mouse models have demonstrated an inverse correlation between TET2 enzyme levels and the progression of hematopoietic malignancies. Consistently, recent research by Zhang et al. revealed downregulation of TET2 expression serves as a potential prognostic and predictive biomarker in cytogenetically normal acute myeloid leukemia. Despite mounting evidence highlighting the fundamental role of TET2 in normal hematopoiesis and myeloid transformation, the biochemical characterization of wild-type and mutant TET2 remains limited owing to challenges associated with the production and assay of active TET2.

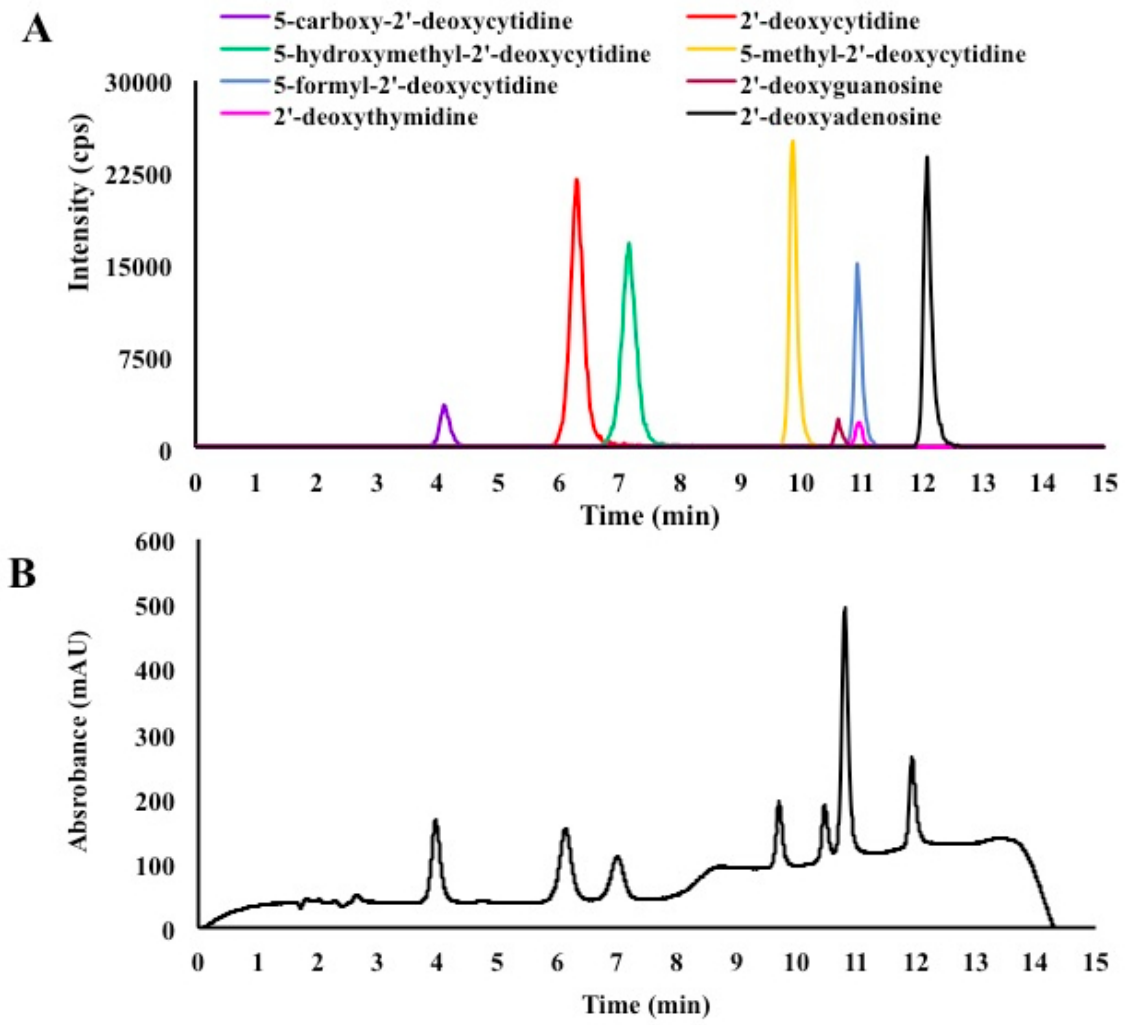


Figure 8: Mass spec chromatogram (panel A) and HPLC chromatogram (panel B) showing separated peaks of all four natural nucleosides as well as the modified cytosine bases.

Additionally, challenges persist in quantifying the activities of wild-type and mutant TET2 dioxygenases. Most studies have relied on antibody-based assays such as dot blots and enzyme-linked immunosorbent assays (ELISA), which employ a single antibody to detect specific cytosine modifications. However, these assays do not offer a comprehensive understanding of the catalytic reactions facilitated by the TET isoforms. Consequently, LC-MS/MS-based assays have emerged as the sole means for quantifying different cytosine modifications. In this study, we developed a novel liquid chromatography method capable of separating four normal DNA bases (A, T, G, and C) as well as four modified cytosine bases (5mC, 5hmC, 5fC, and 5caC). To quantify the eight nucleosides resulting from TET2-catalyzed reactions, we integrated our improved liquid chromatography method with tandem mass spectrometry (Figure 8). This highly sensitive LC-MS/MS assay was employed to determine the activity of the recombinant untagged human TET2 enzyme and its mutants.

CHAPTER 3

CHARACTERIZATION OF TET METHYLCYTOSINE DIOXYGENASE 2 MEDIATED 5MC OXIDATION IN NON-CpG CONTEXT

3.1 Introduction

The methylation of cytosine at carbon-5 (5mC) in the CpG dinucleotide (5mCpG) of DNA plays crucial roles in various processes, such as X-chromosome inactivation, gene imprinting, nuclear reprogramming, and tissue-specific gene expression in mammalian cells[57–59]. Dynamic regulation of 5mCpG methylation also has significant implications for pluripotency, differentiation, and development[22,31,53,81]. The human genome contains more than 20 million CpG dinucleotides, with 70-80% of CpG cytosines methylated[82]. The delicate balance between cytosine methylation and demethylation within CpG dinucleotides is vital for establishing the epigenetic pattern of a cell, and any disruption in methylation patterns can lead to pathological conditions including cancer[40,41,83,84].

Recent genome-wide bisulfite sequencing studies at single-base resolution have revealed a considerable amount of 5mC at CpH sites in almost all human cells[85,86]. Notably, in human embryonic stem cell lines, approximately 67.85%, 6.68%, 1.48%, and 0.63% of all CpG, CpA, CpT, and CpC sites are methylated, respectively[87]. CpH methylation marks, including those on CpA, CpT, and CpC, account for a significant portion of the methylation marks in brain cells, neurons, and different embryonic stem cells (ESCs). Similar to 5mCpG, the presence of 5mCpH marks is generally associated with reduced transcription[47,49–52,82–84]. These markers are also linked to genomic imprinting, regulation of inter-chromosomal interactions, and prediction of genes that escape X-inactivation[88–90]. Moreover, non-CpG methylation has been correlated with gene expression in various

cancer types, and the presence of 5hmC in the context of CpH has been reported in human cells[91–94]. Taken together, these initial studies indicate the important roles of 5mCpH marks in the mammalian genome.

The methylation of 5mC within CpG dinucleotides is carried out by DNA methyltransferases (DNMTs), while its removal is initiated by the ten-eleven translocation (TET) family of dioxygenases, specifically TET1-3[62,95,96]. TET proteins are a class of dioxygenases that depend on iron (II) and 2-oxoglutarate (2OG) to catalyze the stepwise oxidation of 5mC to generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC)[51,63]. The 5fC and 5caC residues are ultimately replaced by unmodified cytosine bases through the thymine DNA glycosylase (TDG)/base excision repair (BER) pathway. While the role of TET2 (and its homologs TET1 and TET3) in oxidizing 5mC at CpG sites has been extensively studied, its involvement in 5mC oxidation at non-5mCpG (or 5mCpH) sites remains unestablished.

Notably, in vitro studies using oligonucleotides have shown that DNMT3A, an enzyme responsible for methylation, can methylate CpH markers[97,98]. In vivo studies have also demonstrated a strong correlation between the presence of 5mCpH markers and the expression of DNMT3A, DNMT3B, and DNMT3L[99–101]. Additionally, research on ESCs has provided compelling evidence that DNMT3A and DNMT3B methylate CpH sites (51, 123)[102,103]. Although passive demethylation of 5mCpH sites may occur during cell division, dynamic active demethylation has been observed in some instances. However, it has been shown the human TET2 enzyme efficiently oxidizes 5mCpG sites but shows negligible oxidation of 5mCpA and 5mCpC sites[72]. In contrast, a TET homolog called NgTET1 from N. gruberi, which is structurally similar to human TET2, can oxidize DNA

containing both 5mCpG and 5mCpA sites with comparable efficiency[104]. These conflicting results raise questions regarding the bases tolerated at the +1 position in TET-mediated 5mC oxidation of DNA substrates.

This study presents findings from a highly sensitive LC–MS/MS-based assay, demonstrating that human TET2 can oxidize 5mCpH sites in double-stranded DNA. Similar to the oxidation of 5mCpG, the oxidation of 5mC in CpH leads to the formation of 5hmC, 5fC, and 5caC marks. As the human TDG enzyme can remove 5fC and 5caC from CpH sites, allowing their replacement with unmodified cytosine along with the recently discovered deformylation and decarboxylation pathways, our results indicate a novel demethylation pathway of 5mCpH sites initiated by TET2 dioxygenase[105–109]. These findings contribute to our understanding of the emerging role of non-CpG methylation in gene expression and cancer.

3.2 Materials and methods

Purification of GST-tagged Human TET2 catalytic domain: A human TET2 (1129-1936) clone with residues 1481-1843 replaced by a 15-residue GS linker (designated as TET2 1129-1936 Δ1481-1843 DNA) was obtained as we have previously described (33). Briefly, after PCR and purification on the agarose gel, the TET2 1129-1936 Δ1481-1843 DNA fragment was cloned into the pJET1.2 vector and verified by DNA sequencing. To produce the GST-tagged human TET2 protein, the minimal catalytic domain of human TET2 (TET2 1129-1936 Δ1481-1843 DNA) was inserted into pGEX4T-1 vector using EcoR1 and Xho1 restriction enzymes. The final recombinant clone was sequence verified. The recombinant vectors were transformed into *E. coli* BL21(DE3) cells for protein expression. Cells were grown in Luria-Bertani broth containing 100 μg/ml ampicillin at 37°C, until the $O.D._{600}$ of the culture reached 0.8. The optimized condition for recombinant

protein expression was found to be 0.25 mM IPTG induction of bacterial culture ($O.D._{600}$ = 0.8) for 16 hrs at 17°C.

The recombinant TET2 protein was purified from 18 liter of culture. Bacterial cells were pelleted and either stored at –80°C or lysed immediately for protein purification. This step onwards, all the steps of protein purification were performed on ice. The cells were resuspended in 30 ml of lysis buffer (PBS-10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, pH 7.5) and sonicated at an intensity of 20 for 6 mins with 6 pulses every 1 minute (Sonic Dismembrator 550, Fisher Scientific). The lysate was spun down at 4700 rpm for 60 min (Beckman J2-HS centrifuge) and the soluble fraction was collected and filtered through 0.45µ filter. Glutathione Agarose resin was used for purification of GST tagged proteins. The resin (100 ml) was packed into a XK26/20 FPLC column (Pharmacia now GE Healthcare, Piscataway, New Jersey) and equilibrated in 10 bed volumes of wash buffer (PBS-10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, pH 7.5) by a Akta FPLC system (Pharmacia/GE Healthcare). Soluble fraction was loaded on the equilibrated column at 0.2ml/min of flow rate. Following this, the column was washed with 5-10 bed volumes of wash buffer (flow rate 1ml/min) till the flow through was clear. Elution of the bound protein was carried out using a gradient (flow rate 1 ml/min) from the wash buffer to the elution buffer (50 mM Tris-HCl, 33 mM Reduced Glutathione, pH 7.5). 0 to 100 % of elution buffer was reached in 10 min, followed by holding at 100% elution buffer for 90 min and then at 0% elution buffer for 30 min. Samples of cell lysate before and after column loading, along with all the wash and elution fractions (30 ml each) were collected and analyzed by SDS-PAGE. Selected fractions containing the desired protein were

concentrated to 20 ml using Labconco Freeze dry system, Freezone 4.5(Kansas City, Missouri).

***In vitro* TET2 assay:** The TET2-mediated *in vitro* oxidation reactions were performed using 25-mer double stranded hemimethylated substrates. dsDNA substrates were generated by hybridizing single stranded oligomers on a thermal cycler using a stepdown PCR method (95 °C to 37 °C with 5 °C stepwise decrease and a hold for 5 minutes at every step). Substrate DNA sequences are provided in Table 1. The *in vitro* TET2 enzymatic reactions were performed in a reaction mixture containing 50 mM HEPES (pH 8.0), 75 μM FeSO4, 1 mM 2OG, 2 mM ascorbate, 1 mM ATP and 1 mM DTT in 100 μl of total volume. An equal amount of DNA substrate (0.2 pM) and purified TET2 enzyme (20 μM) were added to each reaction mixture. After 1 hour of incubation at 37 °C, TET catalyzed oxidation reactions were quenched with 5 μl of 500 mM EDTA.

Table 1: Sequence of DNA substrates used in this study.

DNA substrates	Sequence
5mCpG substrate	5'GCGCCGGTCGTA/iMe-dC/GGCCGCTCCCGC3' 5'GCGGGAGCGGCCGTACGACCGGCGC3'
5mCpA substrate	5'GCGCCGGTCGTA/ iMe-dC/AGCCGCTCCCGC3' 5'GCGGGAGCGGCTGTACGACCGGCGC3'
5mCpC substrate	5'GCGCCGGTCCTG/ iMe-dC/CCCCGCTCCCGC3' 5'GCGGGAGCGGGGGCAGGACCGGCGC3'
5mCpT substrate	5'GCGCCGGTCCTG/ iMe-dC/TCCCGCTCCCGC3' 5'GCGGGAGCGGGAGCAGGACCGGCGC3'

Quantitative LC-MS/MS-based analysis of TET2 activity: An LC-MS/MS method was developed for the separation and quantification for different cytosine nucleotides using a Phenomenex Gemini 5µ C18 column (Phenomenex, Torrance, CA). A gradient elution method was employed with solvent A (10 mM ammonium acetate, pH 3.5) and solvent C (100% acetonitrile with 0.1% acetic acid) as mobile phase solvents. The chromatographic elution was performed with a gradient of 0% solvent C in 4.5 min followed by 0-90% solvent C in 4.5-7.5 minutes and a post equilibration with solvent A for 7 minutes using a flow rate of 0.3 mL/min. The UV detector was set at 280 nm. MS detection was done in the positive ESI mode on an AB Sciex API 3200 Q-Trap tandem mass spectrometer (Concord, Ontario, Canada) equipped with an ESI source. The mass-spectrometry parameters were optimized by infusing different cytosine nucleosides and using the automated quantitative optimization routine in the Analyst software. The optimized source parameters were as follows: heater temperature 550°C, ion source voltage 5500 V, curtain gas 50.00, nebulizer gas (GS1) 50.00, and sheath gas (GS 2) 50.00. Tandem mass-spectrometric analyses were performed using nitrogen as the collision gas (CAD--Medium). For each parent ion nucleoside (Q1), the most intense product ion (Q3) was selected and used in an LC-MS/MS analysis.

The TET2 reaction was quenched and prepared for LC-MS/MS analysis by separating the DNA from the reaction mixture using Zymo Oligo purification columns (Zymo research, Irvine, CA) according to the manufacturer's protocol. The separated DNA was denatured by heating at 55C for 2 min and digested with 2 units of DNAse I and 60 units of S1 nuclease at 37°C for 12 h to produce individual nucleotide-monophosphates. Following the digestion, 2 units of calf intestinal alkaline phosphatase (CIAP) was added

at 37°C for 12 h to remove the terminal phosphate groups from nucleoside-monophosphates to obtain nucleosides. All modified cytosine nucleosides were quantified using the LC-MS/MS method described above. 5hmC formation in our reactions is expressed as 5hmC peak area/ peak area of cytosine/ no. of cytosine residues in the substrate (ΔhmC/ΔC/#C). We used this formula for 5hmC quantification to nullify any inconsistency in our mass spec data.

3.3 Results and discussion

Non-CpG methylation includes methylation at cytosines followed by adenine, thymine, or another cytosine. Non-CpG methylation was initially described in the plant genome. It is suggested to be prevalent in human embryonic stem cells (ES cells) and brain tissue and comprises 0.02% of total methyl-cytosine in differentiated somatic cells[105,110]. The genome of H1 (male) human ESCs is heavily methylated at non-CG sites: 25% of methylated sites are in a non-CG context. The DNA methylation profile of human ESCs revealed a strong correlation between non-CG methylation and pluripotency[111]. However, non-CpG methylation has been far less extensively studied than CpG methylation. The prevalence of non-CpG methylation in the human genome is still not entirely clear, but it is suggested to be far more prevalent within the methylome than previously thought[112]. The Ten-eleven-translocation (TET) family of enzymes can oxidize the fifth base of DNA, 5-methylcytosine (mC) sequentially, to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxylcytosine (caC). The discovery that TET proteins are biochemically able to oxidize 5mC made it possible to test if this enzyme protein may be responsible for the specific loss of 5mC in non-CpG islands.

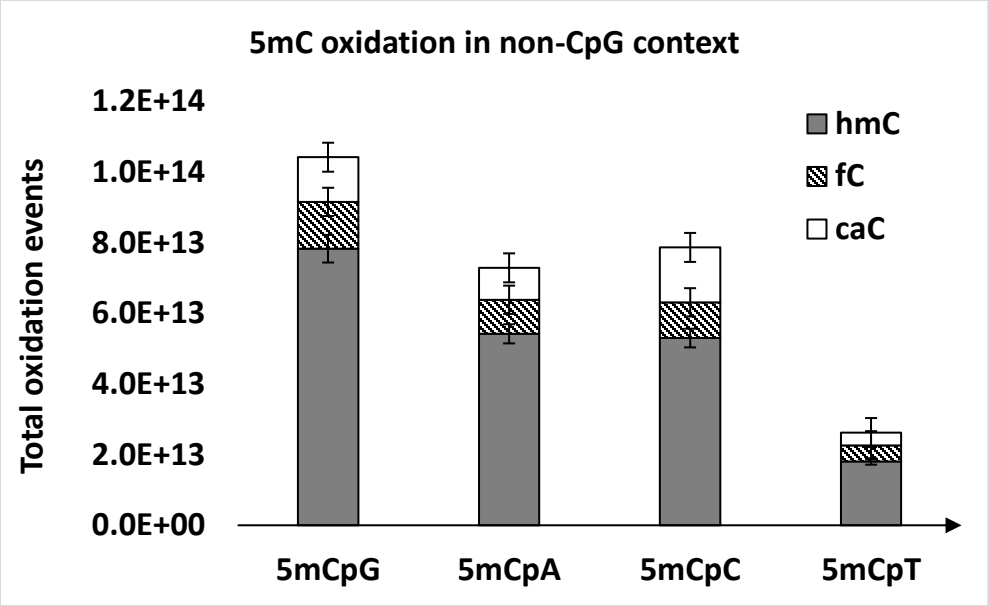


Figure 9: Oxidation of 5mCpG and 5mCpH (here H refers to A/C/T) substrates by hTET2. 5mC oxidation at the non-CpG substrates by hTET2 exhibits that the enyme is versatile enough to oxidize metylation marks in CpG as well as non-CpG substrates. 'Y' axis corresponds to total oxidation events (TOE). The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 10). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

To test this hypothesis, we designed 25 mer double stranded hemimethylated substrates, in which the 5mC site was followed by an adenine, thymine, or another cytosine. The DNA substrates were incubated for an hour with TET2, and the reaction was quenched. The substrate DNA was purified from the reaction mixture, and the DNA was digested into nucleosides. The formation of the 5mC products was quantified using LC-MS/MS. Surprisingly, similar to the oxidation of 5mC at CpG sites, three product peaks with identical mass characteristics and elution profiles to the 5hmC, 5fC, and 5caC bases were detected by LC-MS/MS. Oxidation of 5mC in non-CpG sites was significant breakthrough, because a previous study reported that the human TET2 predominantly oxidized 5mCpG sites (>85% 5mC oxidized), while the oxidation of 5mCpA was negligible (<2% 5mC oxidized). The previous observation was reasoned by a specific hydrogen bond between the phosphate group of guanine with TET2 active site residue S1290[72]. We observed that TET2 can significantly oxidize 5mC in CpA (~70%) compared to 5mCpG. Specifically, little difference was observed in the formation of 5hmC and 5fC, while the amount of 5caC, in the case of 5mCpA, was significantly less than in the case of 5mCpG. Intrigued by these results, we further tested the activity of the TET2 catalytic domain with dsDNA sequences containing one 5mC residue in CpC and CpT sites. Our experiments demonstrated that the TET2 catalytic domain can oxidize 5mC in CpH sites. Taken together, our results suggest that TET2 can initiate the oxidation cascade at 5mCpH sites which may further lead to complete demethylation by the TDG/BER pathway.

CHAPTER 4

SUBSTRATE DNA LENGTH REGULATES THE ACTIVITY OF TET 5-METHYLCYTOSINE DIOXYGENASES

4.1 Introduction

Methylation of cytosine residues (5mC) in DNA plays critical roles during metazoan development by regulating gene expression. The 5mC marks constitute approximately 4% of all cytosine residues present in a human genome and are predominantly present in the CpG islands[113,114]. This reversible epigenetic mark when present upstream of a promoter region generally causes transcriptional repression of the downstream genes, whereas its removal restores transcription[22,81,114]. Cytosine methylation is catalyzed by S-adenosylmethionine dependent DNA methyl transferases (DNMT), while the demethylation cascade is initiated by the TET enzymes. TET enzymes belong to a class of iron (II) and 2-oxoglutarate (2OG) dependent dioxygenases that catalyze oxidation of 5mC into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) sequentially in a three-step oxidation reaction. In the active demethylation pathway, thymine-DNA glycosylase hydrolyses the N-glycosidic bond to convert 5fC and 5caC into basic sites that are further processed through the base-excision repair pathway to reinstate unmodified cytosine [4–6]. Recent studies have shown that 5fC and 5caC can also undergo direct deformylation and decarboxylation to reinstate unmodified cytosine[107,115].

In eukaryotes, DNA is wrapped around histone isoforms to form nucleosomes. Tightly wrapped DNA around the nucleosome is called the core DNA whereas the DNA connecting two nucleosomes is referred as the linker DNA. Normally the nucleosomal

DNA is less accessible to enzymes and transcription factors compared to the linker DNA[116]. However, nucleosomal architecture is highly dynamic in nature as sliding of nucleosomes regulate accessibility of the linker DNA. A methylation study demonstrated that the maintenance DNA methyltransferase, DNMT1 possesses DNA-length sensing ability as it preferentially recognizes nucleosomes with symmetric linker DNA with a minimum length of 20 nucleotides[117]. Another study compared the activity of *de novo* methyltransferases DNMT3A and DNMT3B on reconstituted nucleosomes[118]. This study demonstrated that DNMT3A preferentially methylates the linker DNA region of nucleosomes whereas DNMT3B targets the core nucleosomal DNA for cytosine methylation. Although the linker DNA contains significantly higher methylation marks compared to the core DNA region.

Similarly, methylase assisted bisulfite sequencing was used to detect strand specific distribution of TET mediated 5mC oxidation products in the mouse genome[93,119]. Interestingly, 5fC and 5caC marks, which are higher order oxidation products of 5mC, were enriched at the enhancers of accessible chromatin regions suggesting that local chromatin structure can regulate TET mediated 5mC oxidation. Consistent with these results, a study with mTET1 has found the enzyme to be more efficient in oxidizing 5mC marks located in the linker DNA region due to their higher accessibility[120]. Together, these findings suggest that the linker region of nucleosomes harboring 5mC sites is the preferred substrate for TET enzymes because of their higher accessibility. This also suggests that the length of the linker DNA may also regulate TET mediated 5mC oxidation. A previous study by Kizaki *et al* using mTET1 has revealed that the enzyme prefers a 6-mer dsDNA as a substrate compared to a longer 20-mer counterpart[121]. In addition, they claimed that ssDNA was a

preferred substrate over dsDNA. However, this evaluation was exclusively based on the formation of 5hmC using reverse phase HPLC. To this end, by systematically varying length and configuration of DNA substrates, we demonstrate that mTET1 and hTET2 have an optimum activity with 13-mer dsDNA substrates. Increasing or decreasing the length of dsDNA substrates reduce TET-mediated product formation. Furthermore, contrary to the results of Kizaki *et al*, our results demonstrate that mTET1 and hTET2 prefer dsDNA as a substrate over ssDNA.

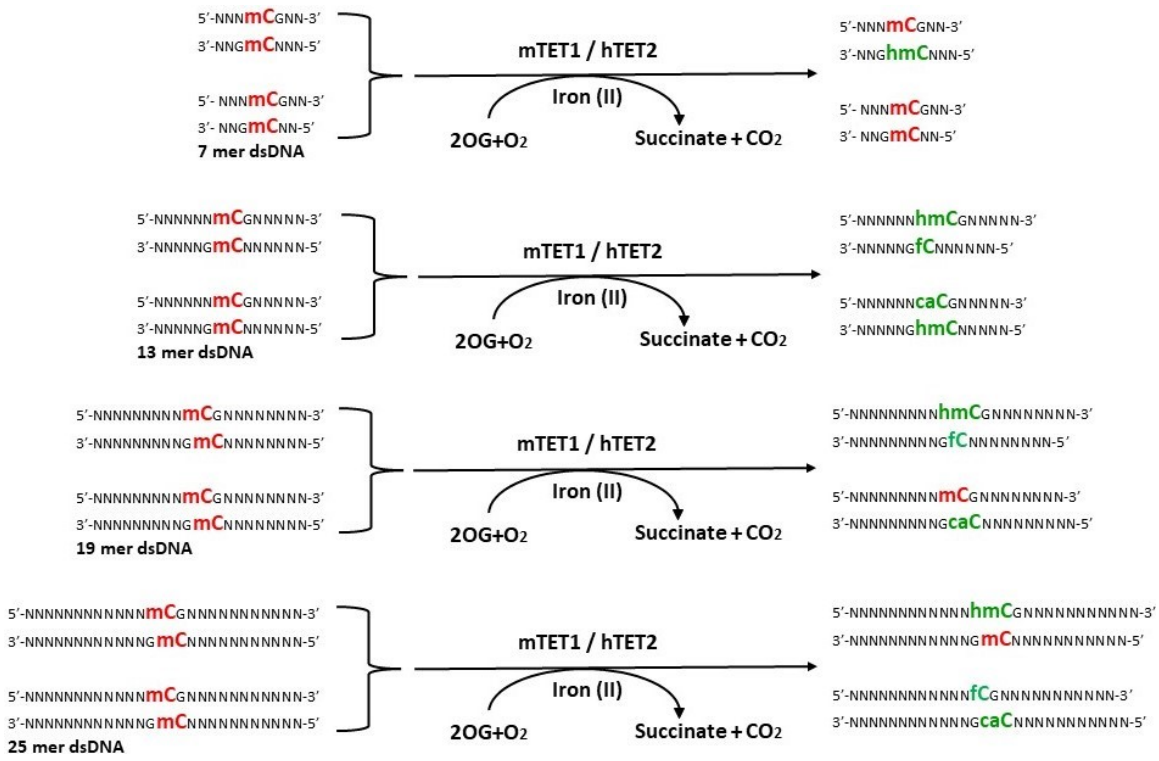


Figure 10: Graphical abstract depicting how substrate DNA length regulates TET-mediated 5mC oxidation. Our results demonstrated that 13 mer DNA substrates undergo more oxidation events than their smaller and longer counterparts (7-, 19, and 25 nucleotide long DNA substrates). 13 mer dsDNA also produces higher levels of oxidation products, such as 5fC and 5caC, which can be further replaced with unmodified cytosines.

4.2 Materials and methods

Chemicals and reagents: All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Escherichia coli (E. coli) expression strain BL21 (DE3) was purchased from Novagen (now EMD Millipore, Billerica, MA). Chromatography solvents such as water and methanol were purchased from Fisher scientific (Waltham, MA). All growth media were from Difco Laboratories (Detroit, MI). SP Sepharose fast flow was purchased from GE Health Care (Chicago, IL). 3-(*N*-morpholino) propanesulfonic acid was purchased from Carbosynth US LLC (San Diego, CA). DNA oligomers were purchased from Integrated DNA Technologies (San Diego, CA). Ultrafiltration membranes were purchased from Millipore (Bedford, MA).

Cloning, expression, and purification of hTET2: The C-terminal dioxygenase domain of hTET2 (hTET2 1129-1936, Δ1481-1843, the minimal catalytically active domain), was cloned into the pDEST14 vector using the site-specific recombination technique as described previously[73]. Untagged hTET2 was purified by cation exchange chromatography on SP Sepharose, as described in our previous article[122]. Fractions containing purified untagged hTET2 were pooled together and concentrated using ultrafiltration membranes (NMWL 30 KDa).

Cloning, expression, and purification of his-tagged mouse TET1 catalytic domain: A plasmid encoding the His-tagged mTET1 catalytic domain (mTET1 1367- 2057) was purchased from Addgene (plasmid #81053). For overexpression, the mTET1 construct was transformed into BL21 (DE3) cells. The cells were grown in LB medium with streptomycin antibiotic until an OD 0.6 was reached, and protein expression was induced for 16 hours at 20 °C by the addition of 0.5mM isopropyl-1-thio-D-galactopyranoside. For the

purification, the cell pellet was resuspended in a sonication/wash buffer (50mM HEPES pH 6.8, 500mM NaCl, 1mM DTT, 35mM imidazole, 10mM alpha ketoglutarate). The cell lysis was done by sonication (8 cycles, 30 seconds with 20% power, 3 minutes off). The lysate was centrifuged for 1 hour at 4,700 rpm at 4 °C. After that, binding to Ni-NTA beads (GE) was performed with constant rotation for 1.5 hour. After washing with the sonication/wash buffer, the bound mTET1 was eluted an elution buffer (50mM HEPES pH 6.8, 500mM NaCl, 1 mM DTT, 300mM imidazole, 10mM alpha ketoglutarate) and dialyzed against a dialysis buffer (50mM HEPES pH 6.8, 300mM NaCl, 1 mM DTT). Aliquots of the dialyzed mTET1 protein were stored at −80 °C. The quality of the mTET1 protein was verified using SDS polyacrylamide gels. Expression and purification of mTET1 was carried away by Aninda S. Dey.

***In vitro* TET reactions:** All dsDNA substrates were generated by hybridizing single stranded oligomers on a thermal cycler using a stepdown PCR method (95 °C to 37 °C with 5 °C stepwise decrease and a hold for 5 minutes at every step). The *in vitro* TET enzymatic reactions were performed in a reaction mixture containing 50 mM HEPES (pH 8.0), 75 μM FeSO4, 1 mM 2OG, 2 mM ascorbate, 1 mM ATP and 1 mM DTT in 100 μl of total volume. An equal amount of DNA substrate (0.2 pM) and purified TET enzyme (2.5 μg) were added to each reaction mixture. After 1 hour of incubation at 37 °C, TET catalyzed oxidation reactions were quenched with 5 μl of 500 mM EDTA. Total DNA from the reaction mixture was separated using the superior oligo purification kit (Superior Scientifics, Lenexa, KS) following manufacturer's protocol. The isolated DNA (20 μl) was digested with 2 units of DNase I and 60 units of S1 nuclease in 40 μl at 37 °C for 12 hours to produce individual nucleotide monophosphates. Following the digestion, 2 units of calf intestinal alkaline

phosphatase (CIAP) were added in the samples followed by incubation for an additional 4 hours at 37 °C to remove the terminal phosphate groups to obtain nucleosides.

Liquid chromatography of nucleosides under different MS/MS modes: The LC-MS/MS analysis was performed on a Sciex 3200 QTrap mass spectrometer (Foster City, CA) coupled to a Shimadzu UFLC LC-20 system (Columbia, MD) using electrospray ionization source (ESI) and run with Analyst v 1.6.2 software. For chromatographic separation of nucleosides in ion-switching mode, a water/methanol based solvent system was used where solvent A was water (adjusted to pH 3.5 using formic Acid) and solvent B was methanol (adjusted to pH 3.0 using formic Acid). The gradient used was 0% B (0-1 minute), 0-2% B (1-12 minute), 2-30% B (12-17 minute), 30% B (17-18 minute), 30-0% B (18-18.5 minute), followed by an equilibration at 0% B for 4.5 minutes at a flow rate of 0.3 mL/minute. Nucleosides were separated on a C18 column (Dimension 150×2 mm, particle size 5µM, Pore Size 100 Å). Standard curves, the limit of detection (LOD), lower limit of quantification (LLOQ) and matrix effect for all eight nucleosides in the positive and negative mode has been calculated by our developed method. Optimization of mass spectrometric parameters and standard curves for modified cytosines have been published in our previous work[123].

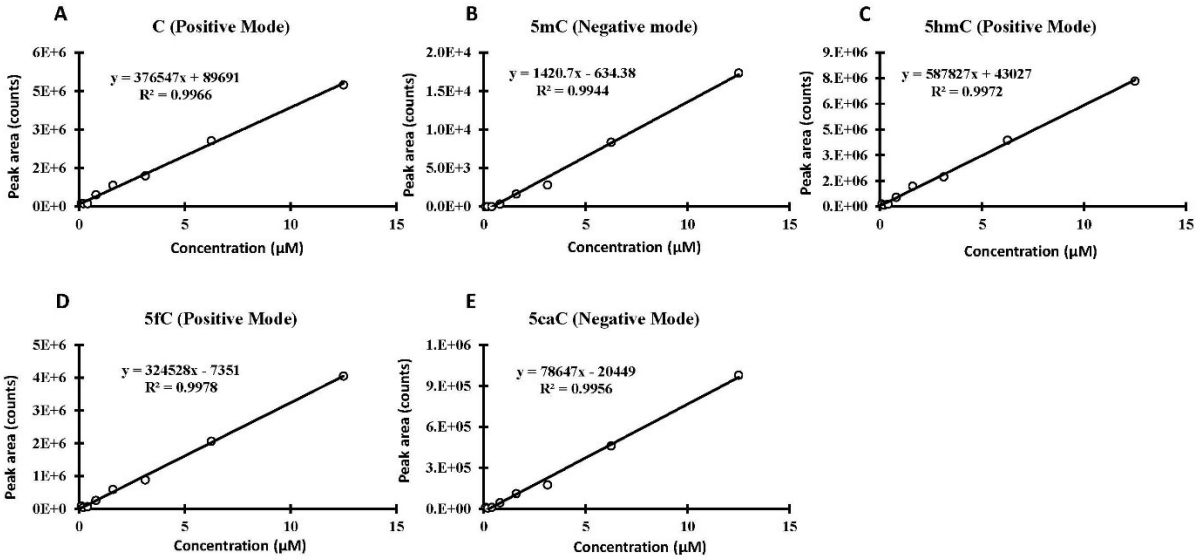


Figure 11: Standard curve for cytosine (dC) and the modified cytosine bases (5mC, 5hmC, 5fC, 5caC). The 'X' axes represent nucleoside concentration, and the 'Y' axes represent peak area. These standard curves were used to quantify 5mC oxidation products from *in vitro* TET reactions.

The hTET2 substrate preference was compared using "total-oxidation-events (TOE)". The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (∆5hmC/∆C/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 10). Furthermore, the catalytic cycle of TET2 dioxygenase generates 5hmC after one oxidation reaction, 5fC after two oxidation reactions, and 5caC after three oxidation reactions. As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard deviations and standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

4.3 Results and discussion

In addition to the fundamental roles played by TET isoforms in normal developmental processes (described earlier), mutations in the hTET2 enzyme are frequently observed in myeloid malignancies such as MDS-myeloproliferative neoplasms (MDS-MPN) and acute myeloid leukemia derived from MDS and MDS-MPN (sAML)[124]. Dysregulation of hTET2 activity is also detected in early stages of other types of leukemia and glioma[41]. Further, *TET2* homozygous null mice demonstrates a significant decrease in the levels of 5hmC and a corresponding increase in the levels of 5mC in the genomic DNA of bone marrow cells[69]. Despite these critical emerging roles of TET dioxygenase in health and diseases, little is known about the effect of substrate DNA length on TET-mediated 5mC oxidation. Studying the optimum substrate of TET enzymes is important for understanding their enzymatic efficiency, functional properties, epigenetic regulation, dsease implications, and for enabling targeted modifications of DNA methylation. These

insights contribute to our overall understanding of epigenetic processes and may have broad implications for both basic research and potential therapeutic applications.

Therefore, it is important to identify factors that regulate hTET2-mediated 5mC oxidation. To determine if hTET2 has any preference for DNA length as a substrate, four DNA sets (S1, S2, S3, S4) of different sequences were chosen. The DNA sequence for each set was chosen from previously reported TET substrates in the literature. In addition, in each set, four different lengths of DNA substrates comprising 7-, 13-, 19-, and 25-mer nucleotides were synthesized. Thus, in each substrate set (S1, S2, S3, and S4), there were five different substrate DNA molecules (1 molecule of DSDM, 2 molecules of DSHM, and 2 molecules of SSSM) of each length (7,13,19 and 25 nucleotides). Each DNA substrate was further used in three different configurations, i.e. double stranded symmetrically-methylated (DSSM substrates i.e., S1-7^{dssm}, S1-13^{dssm}, S1-19^{dssm}, S1-25^{dssm}, and so on), double stranded hemi-methylated (DSHM substrates i.e., S1-$7F^{dshm}$, S1-$7R^{dshm}$, S1-$13F^{dshm}$, S1-$13R^{dshm}$, S1-$19F^{dshm}$, S1-$19R^{dshm}$, S1-$25F^{dshm}$, S1-$25R^{dshm}$, and so on), and single stranded single-methylated (SSSM substrates i.e., S1-$7F^{sssm}$, S1-$7R^{sssm}$, S1-$13F^{sssm}$, S1-$13R^{sssm}$, S1-$19F^{sssm}$, S1-$19R^{sssm}$, S1-$25F^{sssm}$, S1-$25R^{sssm}$, and so on) to evaluate their effect on 5mC oxidation (Table 2). All DNA substrates were designed such that the 5mC residue was present at the center of every DNA substrate.

Table 2: List of substrate DNA sequences, used in this study. Substrate sets: S1, S2, S3, S4.

DNA substrate nomenclature	DNA sequence
S1-7F^{sssm}	5’-CCG/iMe-dC/GCC-3’
S1-7R^{sssm}	5’-GG/iMe-dC/GCGG-3’
S1-13F^{sssm}	5’-GCGCCG/iMe-dC/GCCGGT-3’
S1-13R^{sssm}	5’-ACCGG/iMe-dC/GCGGCGC-3’
S1-19F^{sssm}	5’-CCCGCGCCG/iMe-dC/GCCGGTCGA-3’
S1-19R^{sssm}	5’-TCGACCGG/iMe-dC/GCGGCGCGGG-3’
S1-25F^{sssm}	5'-AGCCCGCGCCG/iMe-dC/GCCGGTCGAGCGG-3'
S1-25R^{sssm}	5'-CCGCTCGACCGGCG/iMe-dC/GGCGCGGGCT-3'
S1-7F^{dshm}	5’-CCG/iMe-dC/GCC-3’ 5’-GGCGCGG-3’
S1-7R^{dshm}	5’-CCGCGCC-3’ 5’-GG/iMe-dC/GCGG-3’
S1-13F^{dshm}	5’-GCGCCG/iMe-dC/GCCGGT-3’ 5’-ACCGGCGCGGCGC-3’
S1-13R^{dshm}	5’-GCGCCGCGCCGGT-3’ 5’-ACCGG/iMe-dC/GCGGCGC-3’
S1-19F^{dshm}	5’-CCCGCGCCG/iMe-dC/GCCGGTCGA-3’ 5’-TCGACCGGCGCGGCGCGGG-3’
S1-19R^{dshm}	5’-CCCGCGCCGCGCCGGTCGA-3’ 5’-TCGACCGG/iMe-dC/GCGGCGCGGG-3’
S1-25F^{dshm}	5'-AGCCCGCGCCG/iMe-dC/GCCGGTCGAGCGG-3' 5'-CCGCTCGACCGGCGCGGCGCGGGCT-3'
S1-25R^{dshm}	5'-AGCCCGCGCCGCGCCGGTCGAGCGG-3' 5'-CCGCTCGACCGGCG/iMe-dC/GGCGCGGGCT-3'
S1-7^{dssm}	5’-CCG/iMe-dC/GCC-3’ 5’-GG/iMe-dC/GCGG-3’
S1-13^{dssm}	5’-GCGCCG/iMe-dC/GCCGGT-3’ 5’-ACCGG/iMe-dC/GCGGCGC-3’
S1-19^{dssm}	5’-CCCGCGCCG/iMe-dC/GCCGGTCGA-3’ 5’-TCGACCGG/iMe-dC/GCGGCGCGGG-3’
S1-25^{dssm}	5'-AGCCCGCGCCG/iMe-dC/GCCGGTCGAGCGG-3' 5'-CCGCTCGACCGGCG/iMe-dC/GGCGCGGGCT-3'
S2-7F^{sssm}	5’-GAG/iMe-dC/GGC-3’
S2-7R^{sssm}	5’-GC/iMe-dC/GCTC-3’
S2-13F^{sssm}	5’-GTCGAG/iMe-dC/GGCCGC-3’
S2-13R^{sssm}	5’-GCGGC/iMe-dC/GCTCGAC-3’
S2-19F^{sssm}	5’-CCGGTCGAG/iMe-dC/GGCCGCTCC-3’
S2-19R^{sssm}	5’-GGAGCGGC/iMe-DC/GCTCGACCGG-3’
S2-25F^{sssm}	5'-GCGCCGGTCGAG/iMe-dC/GGCCGCTCCCGC-3'
S2-25R^{sssm}	5'-GCGGGAGCGGCCG/iMe-dC/TCGACCGGCGC-3'
S2-7F^{dshm}	5’-GAG/iMe-dC/GGC-3’ 5’-GCCGCTC-3’

S2-7R^{dshm}	5’-GAGCGGC-3’ 5’-GC/iMe-dC/GCTC-3’
S2-13F^{dshm}	5’-GTCGAG/iMe-dC/GGCCGC-3’ 5’-GCGGCCGCTCGAC-3’
S2-13R^{dshm}	5’-GTCGAGCGGCCGC-3’ 5’-GCGGC/iMe-dC/GCTCGAC-3’
S2-19F^{dshm}	5’-CCGGTCGAG/iMe-dC/GGCCGCTCC-3’ 5’-GGAGCGGCCGCTCGACCGG-3’
S2-19R^{dshm}	5’-CCGGTCGAGCGGCCGCTCC-3’ 5’-GGAGCGGC/iMe-DC/GCTCGACCGG-3’
S2-25F^{dshm}	5'-GCGCCGGTCGAG/iMe-dC/GGCCGCTCCCGC-3' 5'-GCGGGAGCGGCCGCTCGACCGGCGC-3'
S2-25R^{dshm}	5'-GCGCCGGTCGAGCGGCCGCTCCCGC-3' 5'-GCGGGAGCGGCCG/iMe-dC/TCGACCGGCGC-3'
S2-7^{dssm}	5’-GAG/iMe-dC/GGC-3’ 5’-GC/iMe-dC/GCTC-3’
S2-13^{dssm}	5’-GTCGAG/iMe-dC/GGCCGC-3’ 5’-GCGGC/iMe-dC/GCTCGAC-3’
S2-19^{dssm}	5’-CCGGTCGAG/iMe-dC/GGCCGCTCC-3’ 5’-GGAGCGGC/iMe-DC/GCTCGACCGG-3’
S2-25^{dssm}	5'-GCGCCGGTCGAG/iMe-dC/GGCCGCTCCCGC-3' 5'-GCGGGAGCGGCCG/iMe-dC/TCGACCGGCGC-3'
S3-7F^{sssm}	5’-CTC/iMe-dC/GGT-3’
S3-7R^{sssm}	5’-AC/iMe-dC/GGAG-3’
S3-13F^{sssm}	5’-CAGCTC/iMe-dC/GGTCAC-3’
S3-13R^{sssm}	5’-GTGAC/iMe-dC/GGAGCTG-3’
S3-19F^{sssm}	5’-TTTCAGCTC/iMe-dC/GGTCACGCT-3’
S3-19R^{sssm}	5’-AGCGTGAC/iMe-dC/GGAGCTGAAA-3’
S3-25F^{sssm}	5’-AGCTTTCAGCTC/iMe-dC/GGTCACGCTCGC-3’
S3-25R^{sssm}	5’-GCGAGCGTGAC/iMe-dC/GGAGCTGAAAGCT-3’
S3-7F^{dshm}	5’-CTC/iMe-dC/GGT-3’ 5’-ACCGGAG-3’
S3-7R^{dshm}	5’-CTCCGGT-3’ 5’-AC/iMe-dC/GGAG-3’
S3-13F^{dshm}	5’-CAGCTC/iMe-dC/GGTCAC-3’ 5’-GTGACCGGAGCTG-3’
S3-13R^{dshm}	5’-CAGCTCCGGTCAC-3’ 5’-GTGAC/iMe-dC/GGAGCTG-3’
S3-19F^{dshm}	5’-TTTCAGCTC/iMe-dC/GGTCACGCT-3’ 5’-AGCGTGACCGGAGCTGAAA-3’
S3-19R^{dshm}	5’-TTTCAGCTCCGGTCACGCT-3’ 5’-AGCGTGAC/iMe-dC/GGAGCTGAAA-3’
S3-25F^{dshm}	5’-AGCTTTCAGCTC/iMe-dC/GGTCACGCTCGC-3’ 5’-GCGAGCGTGACCGGAGCTGAAAGCT-3’
S3-25R^{dshm}	5’-AGCTTTCAGCTCCGGTCACGCTCGC-3’ 5’-GCGAGCGTGAC/iMe-dC/GGAGCTGAAAGCT-3’

S3-7^{dssm}	5’-CTC/iMe-dC/GGT-3’ 5’-AC/iMe-dC/GGAG-3’
S3-13^{dssm}	5’-CAGCTC/iMe-dC/GGTCAC-3’ 5’-GTGAC/iMe-dC/GGAGCTG-3’
S3-19^{dssm}	5’-TTTCAGCTC/iMe-dC/GGTCACGCT-3’ 5’-AGCGTGAC/iMe-dC/GGAGCTGAAA-3’
S3-25^{dssm}	5’-AGCTTTCAGCTC/iMe-dC/GGTCACGCTCGC-3’ 5’-GCGAGCGTGAC/iMe-dC/GGAGCTGAAAGCT-3’
S4-7F^{sssm}	5’-GAC/iMe-dC/GGA-3’
S4-7R^{sssm}	5’-TC/iMe-dC/GGTC-3’
S4-13F^{sssm}	5’-CGTGAC/iMe-dC/GGAGCT-3’
S4-13R^{sssm}	5’-AGCTC/iMe-dC/GGTCACG-3’
S4-19F^{sssm}	5’-CAGCGTGAC/iMe-dC/GGAGCTGAT-3’
S4-19R^{sssm}	5’-ATCAGCTC/iMe-dC/GGTCACGCTG-3’
S4-25F^{sssm}	5’-AGCCAGCGTGAC/iMe-dC/GGAGCTGATCGC-3’
S4-25R^{sssm}	5’-GCGATCAGCTC/iMe-dC/GGTCACGCTGGCT-3’
S4-7F^{dshm}	5’-GAC/iMe-dC/GGA-3’ 5’-TCCGGTC-3’
S4-7R^{dshm}	5’-GACCGGA-3’ 5’-TC/iMe-dC/GGTC-3’
S4-13F^{dshm}	5’-CGTGAC/iMe-dC/GGAGCT-3’ 5’-AGCTCCGGTCACG-3’
S4-13R^{dshm}	5’-CGTGACCGGAGCT-3’ 5’-AGCTC/iMe-dC/GGTCACG-3’
S4-19F^{dshm}	5’-CAGCGTGAC/iMe-dC/GGAGCTGAT-3’ 5’-ATCAGCTCCGGTCACGCTG-3’
S4-19R^{dshm}	5’-CAGCGTGAC/iMe-dC/GGAGCTGAT-3’ 5’-ATCAGCTC/iMe-dC/GGTCACGCTG-3’
S4-25F^{dshm}	5’-AGCCAGCGTGAC/iMe-dC/GGAGCTGATCGC-3’ 5’-GCGATCAGCTCCGGTCACGCTGGCT-3’
S4-25R^{dshm}	5’-AGCCAGCGTGACCGGAGCTGATCGC-3’ 5’-GCGATCAGCTC/iMe-dC/GGTCACGCTGGCT-3’
S4-7^{dssm}	5’-GAC/iMe-dC/GGA-3’ 5’-TC/iMe-dC/GGTC-3’
S4-13^{dssm}	5’-CGTGAC/iMe-dC/GGAGCT-3’ 5’-AGCTC/iMe-dC/GGTCACG-3’
S4-19^{dssm}	5’-CAGCGTGAC/iMe-dC/GGAGCTGAT-3’ 5’-ATCAGCTC/iMe-dC/GGTCACGCTG-3’
S4-25^{dssm}	5’-AGCCAGCGTGAC/iMe-dC/GGAGCTGATCGC-3’ 5’-GCGATCAGCTC/iMe-dC/GGTCACGCTGGCT-3’

DNA substrates were subjected to *in vitro* oxidation by hTET2 and mTET1 catalytic domains as described in the experimental section. After enzymatic reactions, DNA was purified on silica columns, and its concentrations were measured on nanodrop. These analyses demonstrated lower recovery for shorter DNAs compared to their longer counterparts. The recovery loss for every DNA sequence was adjusted later. Purified DNA was enzymatically converted to produce nucleosides, which were subjected to mass spectrometry analysis using our previously described ion switching method[124]. In this method, the positive ion mode was used for the detection of 5hmC and 5fC while 5caC was detected in the negative ion mode. The amount of 5mC oxidation products (5hmC, 5fC, and 5caC) were quantified by calculating the peak area of each product. At this stage, the percentage of DNA loss for every reaction during the purification process was normalized in our mass spectrometry data. The normalized peak areas were used to calculate TOE, as described in the experimental section. The substrate preference of TET isoforms was compared using TOE values.

Oxidation of double stranded symmetrically methylated DNA substrates by hTET2: Due to base complementarity CpG islands are palindromic in nature, and when both DNA strands contain 5mCpG sites it's called symmetrically methylated DNA. Previous studies have established that symmetric methylation of palindromic CpG dyads is most prevalent in the mammalian genome[102]. To evaluate the substrate preference of hTET2, dsDNA containing 5mCpG sites on both strands were synthesized and used as DSSM substrates. Results from these experiments demonstrated that TET2-mediated 5mC oxidation varies according to the substrate DNA length (Figure 12). The 7-mer dsDNA substrate produced the lowest amount of products across all four substrate sets (TOE: S1-7^{dssm} = 7.17E+12,

S2-7^{dssm} = 6.63E+13, S3-7^{dssm} = 5.12E+12, S4-7^{dssm} = 1.33E+13). On the other hand, 13-mer DSSM substrates produced the highest amount of oxidation products (TOE: S1-13^{dssm} = 4.18E+13, S2-13^{dssm} = 3.71E+14, S3-13^{dssm} = 6.57E+13, S4-13^{dssm} = 8.48E+13). As substrate length increased to 19- and 25-mer the amount of oxidation products went down (TOE: S1-19^{dssm} = 2.4E+13, S2-19^{dssm} = 1.65E+14, S3-19^{dssm} = 3.61E+13, S4-19^{dssm} = 2.81E+13) (S1-25^{dssm} = 1.89E+13, S2-25^{dssm} = 1.45E+14, S3-25^{dssm} = 2.13E+13, S4-25^{dssm} = 2.78E+13), respectively. Based on these results, we conclude that 13 nucleotides DSSM substrate is the optimum length for hTET2-mediated 5mC oxidation compared to its shorter and longer counterparts.

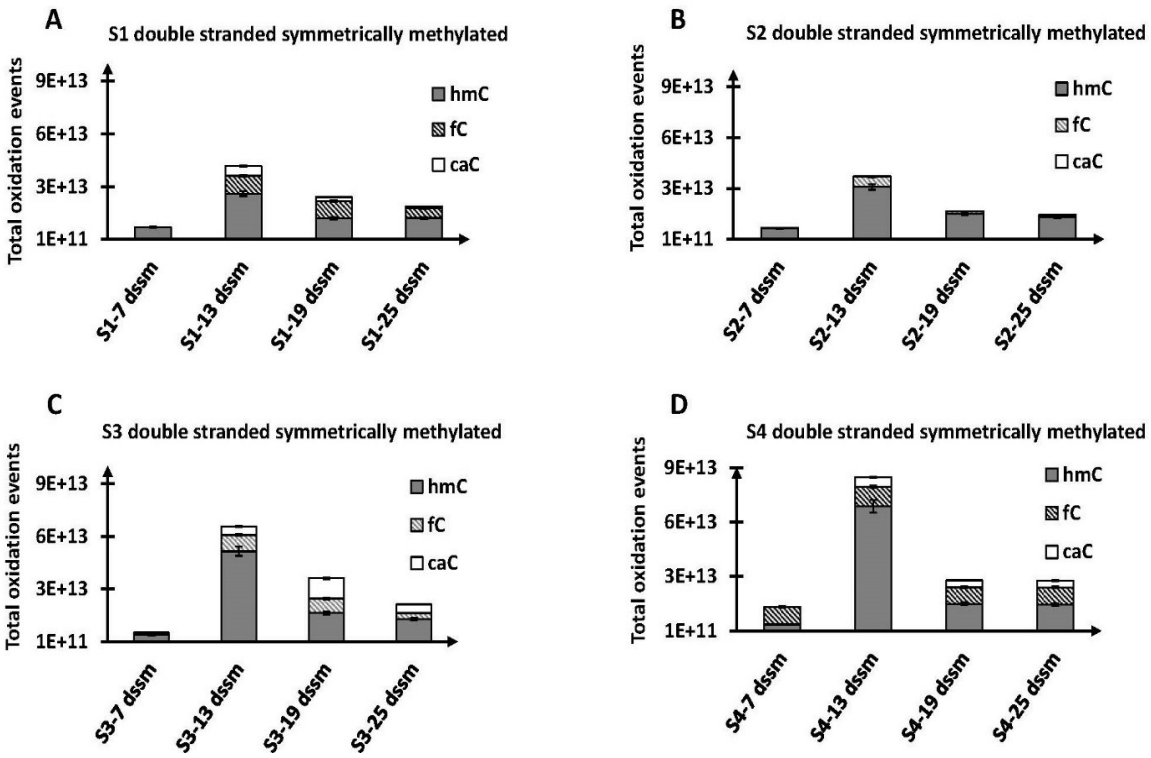


Figure 12: Effect of DNA length on hTET2-mediated 5mC oxidation of double stranded symmetrically methylated (DSSM) substrates sets where, 13 mer DNA substrates undergo the highest number of oxidation events. (A) S1 DSSM substrates, (B) S2 DSSM substrates, (C) S3 DSSM substrates, and (D) S4 DSSM substrates. The 'X' axes represent varying length of DNA substrates while the 'Y' axes represent total oxidation events. The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

Oxidation of double stranded hemimethylated DNA substrates by hTET2: More than 80% of 5hmC located in the palindromic 5mCpG islands are asymmetrically oxidized[102,125]. This suggests that selective oxidation of 5hmC marks can occur in one of the strands of the dsDNA. Furthermore, double stranded hemimethylated DNA can also result from passive demethylation by replicative dilution within the cell. To evaluate the substrate preference of hTET2, dsDNA containing 5mCpG sites on one of the strands keeping the complementary strand unmethylated was synthesized and used as DSHM substrate. Results of these studies also demonstrated that 13-mer dsDNA undergoes highest number of oxidation events (TOE: S1-13F^{dshm} = 3.74E+13, S1-13R^{dshm} = 3.52E+13, S2-13F^{dshm} = 4.64E+13, S2-13R^{dshm} = 5.18E+13, S3-13F^{dshm} = 5.98E+13, S3-13R^{dshm} = 7.25E+13, S4-13F^{dshm} = 3.19E+13, S4-13R^{dshm} = 2.6E+13) among all four substrate sets compared to 7-, 19-, and 25-mer DSHM substrates. However, no significant difference was observed in 5mC oxidation between the forward and complimentary DSHM substrates (Figure 13). These results suggest that 5mC oxidation by hTET2 is not strand specific, which aligns with the previous literature[126].

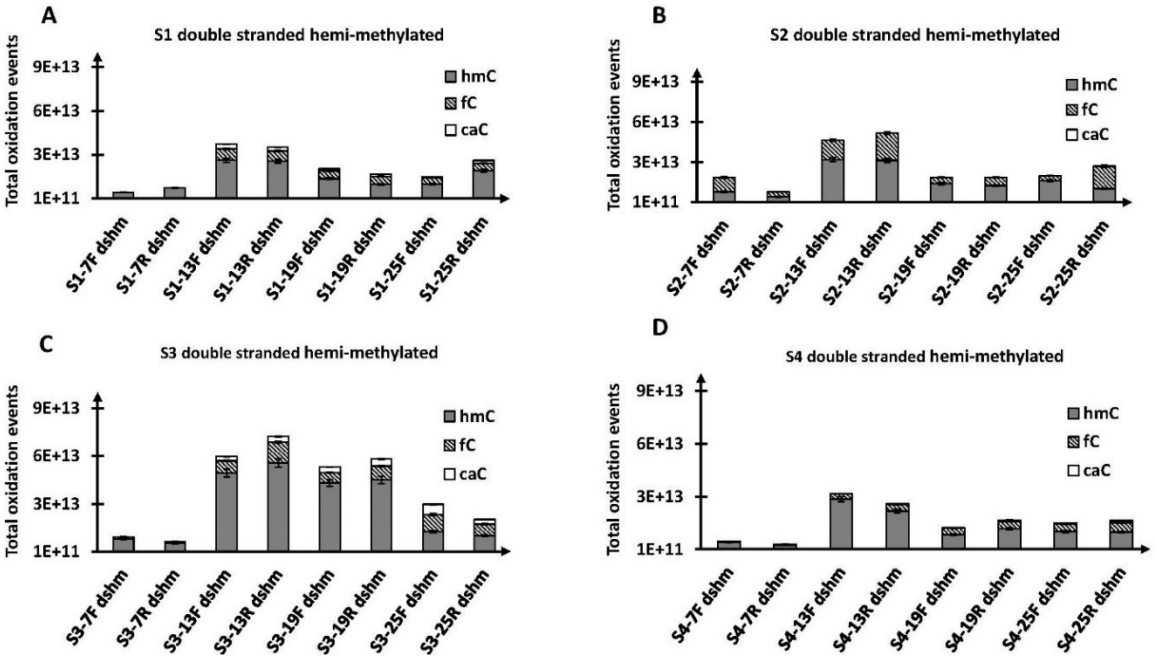


Figure 13: Effect of DNA length on hTET2-mediated 5mC oxidation of double stranded hemi-methylated (DSHM) substrates sets where, 13 mer DNA substrates undergo the highest number of oxidation events. (A) S1 DSHM substrates, (B) S2 DSHM substrates, (C) S3 DSHM substrates, and (D) S4 DSHM substrates. The 'X' axes represent varying length of DNA substrates while the 'Y' axes represent total oxidation events. The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (∆5hmC/∆C/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

Oxidation of single stranded single methylated DNA substrates by hTET2: Single stranded DNA is prevalent during cellular events such as replication and transcription. Interestingly, the presence of 5mC marks in ssDNA can stimulate the *de novo* methylation of downstream CpG marks[127,128]. To evaluate the substrate preference of hTET2, ssDNA sequences containing 5mCpG sites were synthesized and used as single stranded single methylated DNA substrates. Results from these experiments demonstrated that the length of SSSM substrates did not have a consistent pattern of 5mC oxidation by hTET2 in contrast to their dsDNA counterparts (Figure 14). This could be due to secondary structures formed by Watson-Crick hydrogen bonding as well as non-canonical base pairings in ssDNA, which may interfere between the 5mC target site and the hTET2 dioxygenase. Further, in the case of SSSM substrates, hTET2 predominantly produced 5hmC as a product while the levels of higher order oxidation products (i.e., 5fC and 5caC) were negligible. Thus, although hTET2 was able to catalyze 5mC oxidation in both ssDNA and dsDNA substrates, it preferred dsDNA substrates to its single stranded counterparts.

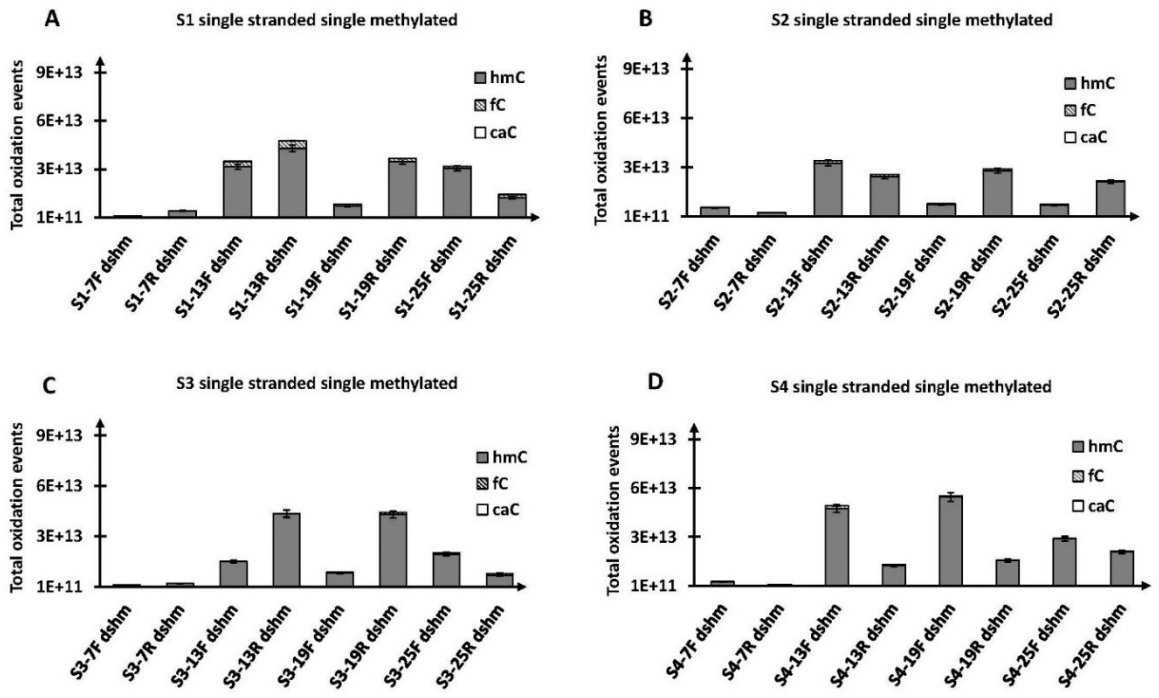


Figure 14: Effect of DNA length on hTET2-mediated 5mC oxidation of single stranded single methylated (SSSM) substrates sets. (A) S1 SSSM substrates, (B) S2 SSSM substrates, (C) S3 SSSM substrates, and (D) S4 SSSM substrates. The 'X' axes represent varying length of DNA substrates while the 'Y' axes represent total oxidation events. The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

Effect of substrate DNA length and configuration on mTET1-mediated 5mC oxidation: A previous study by Kizaki *et al* on mTET1 reported that 6-mer DNA is a better substrate compared to 20-mer counterpart, and it preferred ssDNA over dsDNA based on the formation of only 5hmC[121]. Therefore, experiments were performed with mTET1 and 7-,13-, 19-, and 25-mer DSSM, DSHM, and SSSM substrates. Results of these experiments demonstrated, in agreement with our hTET2 results, that mTET1 preferred 13-mer DSSM and DSHM substrates compared to their shorter and longer counterparts (Figures 15-17). Although mTET1 recognized both ssDNA and dsDNA as substrates, the length of SSSM substrates didn't have a predictable effect on the oxidation of 5mC by mTET1, like hTET2. These discrepancies in results by Kizaki *et al* could be due to the detection of 5hmC only in their study[121]. Since 5hmC itself is a substrate in TET-mediated stepwise oxidation, depletion of 5hmC could correlate with production of higher order oxidized products, i.e., 5fC and/or 5caC.

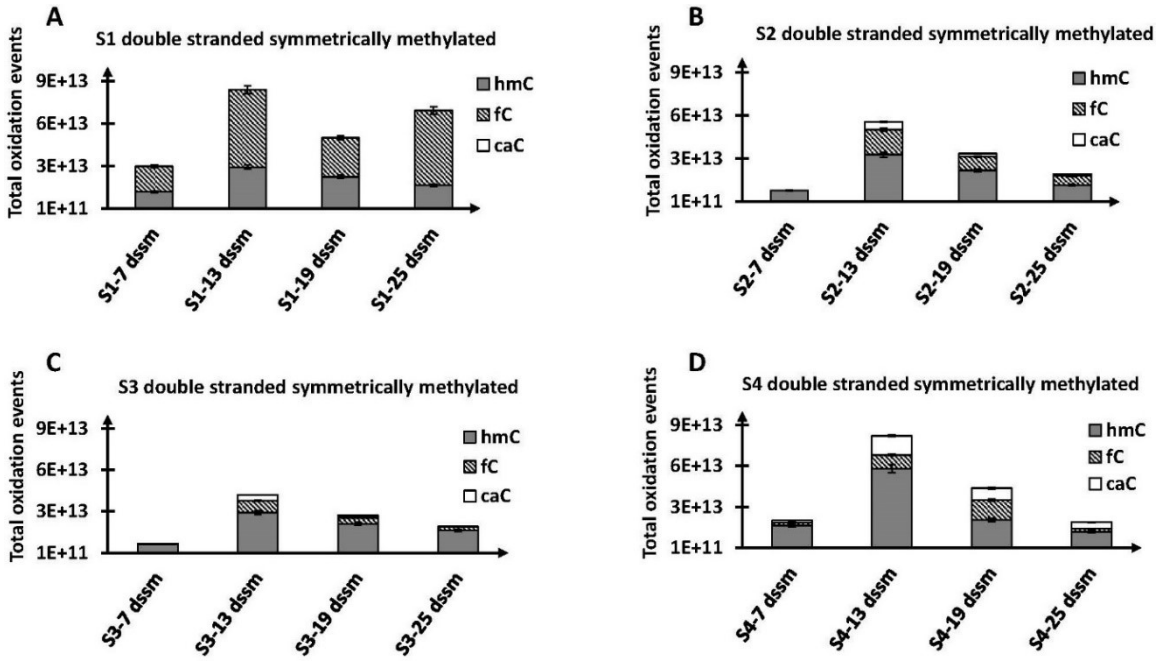


Figure 15: Effect of substrate DNA length on mTET1 mediated 5mC oxidation across double stranded symmetrically methylated DNA configurations where, 13 mer DNA substrates undergo the highest number of oxidation events. (A) S1 DSSM substrates, (B) S2 DSSM substrates, (C) S3 DSSM substrates, and (D) S4 DSSM substrates. The 'X' axes represent varying length of DNA substrates while the 'Y' axes represent total oxidation events. The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

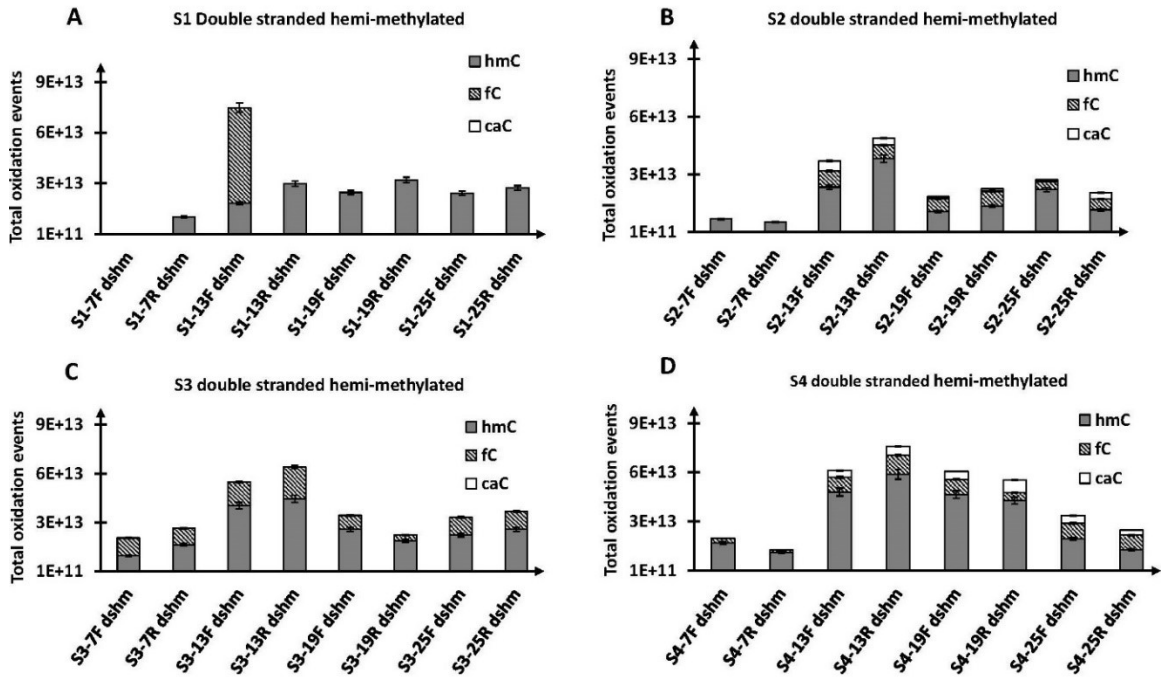


Figure 16: Effect of substrate DNA length on mTET1 mediated 5mC oxidation across double stranded hemi methylated DNA configurations where, 13 mer DNA substrates undergo the highest number of oxidation events. (A) S1 DSHM substrates, (B) S2 DSHM substrates, (C) S3 DSHM substrates, and (D) S4 DSHM substrates. The 'X' axes represent varying length of DNA substrates while the 'Y' axes represent total oxidation events. The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

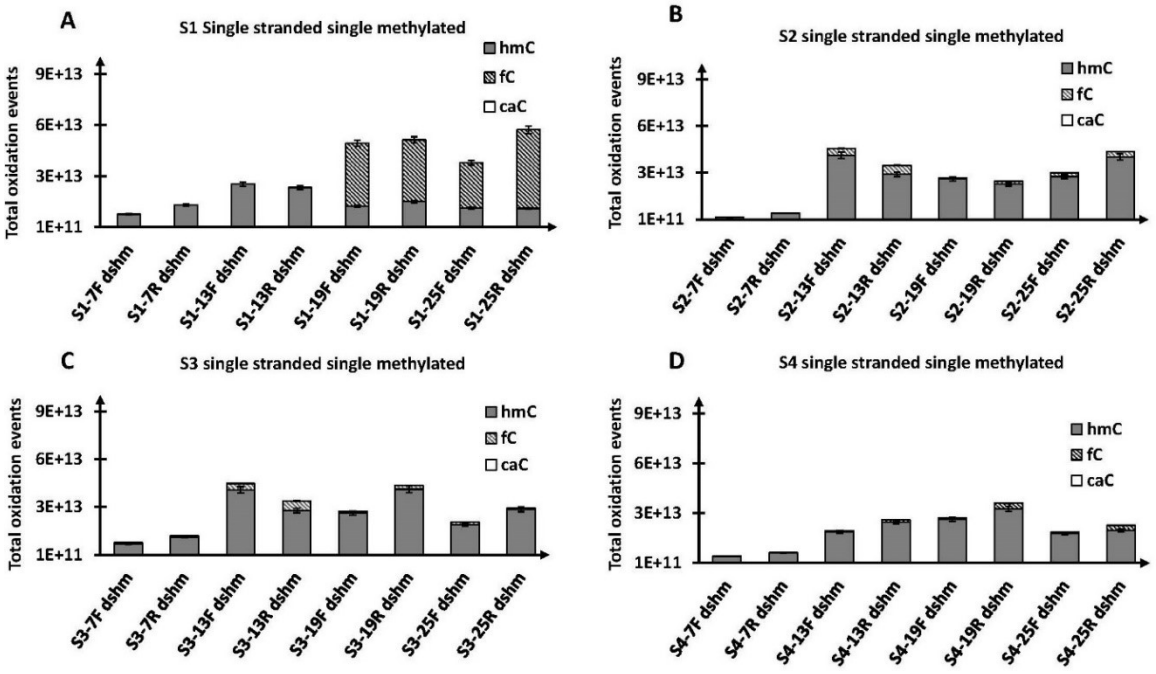


Figure 17: Effect of substrate DNA length on mTET1 mediated 5mC oxidation across single stranded methylated DNA configurations. (A) S1 SSSM substrates, (B) S2 SSSM substrates, (C) S3 SSSM substrates, and (D) S4 SSSM substrates. The 'X' axes represent varying length of DNA substrates while the 'Y' axes represent total oxidation events. The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

A report by Kizaki *et al* claimed that mTET1 prefers 6-mer DNA over a 20-mer substrate. In addition, this study showed that mTET1 prefers ssDNA over dsDNA as a substrate and has a specific requirement of ATP in TET-mediated 5mC oxidation. Our results demonstrate that the length and configuration of substrate DNA can regulate TET-mediated 5mC oxidation. The 13-mer DSSM and DSHM DNA substrates produced the highest amount of oxidative products compared to its shorter and longer counterparts suggesting that the 13-mer nucleotide is the optimum length of DNA for TET interaction and catalysis. However, the effect of DNA length in case of SSSM substrates on TET-mediated 5mC oxidation was less predictable. This could be due to the secondary structures formed by SSSM DNA substrates. Furthermore, our results demonstrate that mTET1 and hTET2 prefer dsDNA as a substrate over ssDNA. We also demonstrate that substrate specificity of TET isoforms correlates with their binding efficiency to dsDNA substrates. Finally, several other studies have reported that ATP stimulates TET-mediated 5mC oxidation *in vitro*[72,121,129]. However, in the proposed TET reaction mechanism there is no specific role of ATP in the oxidation reaction [5]. Therefore, to evaluate any specific role of ATP in the TET-mediated 5mC oxidation, ATP was replaced with dATP and other dNTP analogues (i.e., dGTP, dCTP, and dTTP) in the *in vitro* assay. Results from these experiments demonstrated that along with ATP the activity of hTET2 was also enhanced in the presence of all four dNTP analogues (Figure 18). Therefore, our hypothesis is that ATP and dNTP analogues enhance TET activity by maintaining the iron-(II) homeostasis using the negative charge present on the phosphate group.

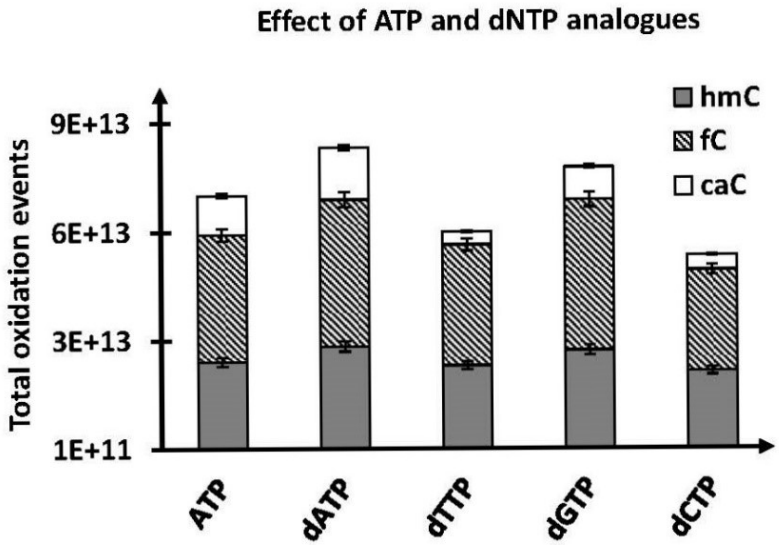


Figure 18: The effect of ATP and dNTP analogues on TET2 mediated 5mC oxidation. The 'X' axis represents the different dNTP analogues and 'Y' axis represents total oxidation events. The amount of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

Although, it has been reported that TET isoforms can oxidize 5mC present in dsDNA along with ssDNA and RNA, little is known about the minimum substrate length required for the optimum activity of TET isoforms. Here, by systematically varying length and configuration of DNA substrates combined with a highly sensitive LC-MS/MS method, we demonstrate that mTET1 and hTET2 have highest preference for 13-mer dsDNA substrates. Increasing or decreasing the length of dsDNA substrate reduces product formation. In contrast to their dsDNA counterparts, the length of ssDNA substrates did not have a predictable effect on 5mC oxidation. Furthermore, our results demonstrate that mTET1 and hTET2 prefer dsDNA as a substrate over ssDNA. Finally, we show that substrate specificity of TET isoforms correlates with their DNA binding efficiency.

It is important to note that the length of linker DNA can vary from 5-60 nucleotides depending on the context and specific protein-DNA interactions involved[130]. Therefore, identifying that hTET2 and mTET1 prefer double-stranded 13-mer DNA as substrates is significant because it provides insight into the minimal substrate DNA length, probably within the linker region, with which TET proteins interact for optimum oxidative product formation. It is worth mentioning that TETs work in conjunction with various regulatory proteins and factors that influence their binding and activity. These factors encompass DNA-binding proteins, chromatin structure, and specific DNA sequence motifs. The presence of these additional factors can impact the specific DNA length and context required for optimal TET activity. TETs exhibit their effects on individual DNA molecules as well as larger genomic regions like CpG islands, which consist of frequently methylated CpG sites. Additionally, the DNA sequence context, including neighboring nucleotides and specific DNA motifs, can impact TET activity. Certain DNA sequences can facilitate the

binding and recognition of TET enzymes and other associated DNA-binding proteins, thereby altering the minimum DNA length required for their activity. Thus, the minimum length of DNA required for DNA demethylase activity *in vivo* may vary depending on the specific TET enzyme and its associated factors. Importantly, our results demonstrate that the size of the target region can influence the 5mC oxidation states (*i.e.,* 5hmC, 5fC, and 5caC) which can dictate fundamental epigenetic regulatory processes in health and diseases.

CHAPTER 5

USE OF 2-OXOGLUTARATE ANALOGS FOR RESCUING MUTATIONS IN TET METHYLCYTOSINE DIOXYGENASE 2

5.1 Introduction

Cytosine methylation is a chemical modification that occurs in eukaryotic DNA, in which a methyl group (-CH3) is added to the carbon 5 position of the cytosine base. This modification can occur in specific regions in the DNA sequence called CpG sites, where a cytosine base is followed by a guanine base in the 5' to 3' direction. Cytosine methylation (5mC) plays an important role in gene regulation by affecting the accessibility of DNA to transcription factors and other regulatory proteins. Methylation at CpG (5mCpG) sites is associated with gene silencing as it can inhibit the binding of transcription factors and recruit proteins that repress transcription[131,132]. In contrast, unmethylated CpG sites are typically associated with active gene expression. In addition to its role in gene regulation, cytosine methylation plays a critical role in genomic imprinting, X-chromosome inactivation, and DNA repair. Aberrant methylation patterns, such as global hypomethylation or hypermethylation at specific CpG sites, are associated with various diseases, including cancer and developmental disorders. Overall, cytosine methylation is a crucial epigenetic modification that contributes to the regulation of gene expression and maintenance of genomic stability[2,19,21,41]. Although 5mC sites have primarily been found at CpG sites, recent studies have shown that cytosine methylation marks can occur at non-CpG sites (5mCpH; H refers to A, C, or T) where the 5mC base is followed by adenine, cytosine, or thymine. The 5mCpH marks are cell type-specific and are enriched in the enhancers, promoters, and gene bodies of the eukaryotic genome. However, the role of

5mCpH methylation in development and disease conditions remains to be elucidated[91,92,102,110,133–136].

TET proteins (TET1, TET2, and TET3) belong to the family of non-heme iron (II), 2OG-dependent dioxygenases (2-OGDDs). TET enzymes catalyze the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), as well as to other oxidized derivatives of 5mC, including 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5fC and 5caC undergoes further excision by the DNA repair enzyme thymine DNA glycosylase (TDG) leading to the restoration of the unmodified cytosine base. This process is known as active DNA demethylation, as it can reverse the DNA methylation mark and facilitate gene expression[61]. Interestingly, recent studies by us and other research groups have demonstrated that TET2 is sufficiently versatile to oxidize 5mC marks in both CpG and non-CpG contexts[105,133]. Epigenetic repression of tumor suppressor genes (TSG) has been implicated in pathogenesis of malignant evolution. One such epigenetic mechanism, observed in MDS, is acquired progressive methylation of CpG islands of gene promoters, leading to transcriptional repression. In contrast to studies on genomic lesions, the mechanisms regulating aberrant epigenetic silencing and the causes of putative epigenetic instability have not yet been conceptualized[77,137,138]. The TET family (TET1-3) of hydroxylases/ dioxygenases was recently identified as an iron (II), 2OG-dependent enzymes. TET2 gene mutations occur frequently in myeloid malignancies such as MDS (≈10% of cases), chronic myelomonocytic leukemia (CMML, ≈50%), myeloproliferative neoplasms (≈20%), and secondary acute myeloid leukemias derived from these conditions (sAML, ≈20%). Human TET2 converts 5mC to 5hmC; consequently, patients with TET2

mutations show low levels of genomic 5hmC in the bone marrow compared to those with WT-TET2[41].

Non-heme iron (II), 2OG-dependent dioxygenases (2-OGDDs), constitute the largest subgroup of dioxygenases and oxidases in humans that regulate numerous processes, including gene expression and chromatin structure[5]. These enzymes were first identified in prokaryotes, where they catalyze diverse oxidation reactions[10–12]. However, recent studies in human cells have shown that these proteins primarily utilize hydroxylation reactions to control a number of epigenetic processes and transcription. Oxidation of the prime substrate by these enzymes is coupled to the conversion of 2OG into succinate and CO_2. However, in cancer cells, normal regulation of 2-OGDDs activity is disrupted, leading to changes in gene expression and epigenetic modifications. 2-oxoglutarate (2OG) analogs are molecules that mimic the chemical structure of 2OG and can modulate the activity of 2-OGDDs. 2OG analogs have been shown to restore the activity of clinically relevant 2-OGDDs and modulate the epigenetic landscape of cancer cells[139,140]. For instance, cancer-associated 2-oxoglutarate analogs (e.g., succinate, fumarate, 2-hydroxyglutarate) have been shown to inhibit histone lysine demethylases, suggesting that these enzymes are susceptible to inhibition by naturally occurring 2OG derivatives[5,141]. The objective of this study was to develop effective strategies using 2OG analogs to enhance the activity of TET2 harboring point mutations at the R1896 residue. In this study, a library of 11 compounds was designed which mimic the chemical structure of 2OG. By screening these compounds, eight 2OG analogs were identified that could specifically rescue mutant TET2 activity. The catalytic activity of the TET2 R1896S mutant was enhanced by up to 90% compared with that of the wild-type enzyme. Furthermore, we

were able to rescue TET2 clinical mutations in the R1896A and R1896F residues using this alternate co-substrate approach.

5.2 Materials and methods:

Chemicals and Reagents: All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Escherichia coli (E. coli) expression strain BL21 (DE3) was purchased from Novagen (now EMD Millipore, Billerica, MA). Chromatography solvents such as water and methanol were purchased from Fisher scientific (Waltham, MA). All growth media were from Difco Laboratories (Detroit, MI). SP Sepharose fast flow was purchased from GE Health Care (Chicago, IL). 3-(*N*-morpholino) propanesulfonic acid was purchased from Carbosynth US LLC (San Diego, CA). DNA oligomers were purchased from Integrated DNA Technologies (San Diego, CA). Ultrafiltration membranes were purchased from Millipore (Bedford, MA).

Cloning, Expression, and Purification of WT TET2 and mutants: The C-terminal TET2 dioxygenase domain (TET2 1129-1936, Δ1481-1843, the minimal catalytically active domain), was cloned into the pDEST14 vector using the site-specific recombination technique as described previously[73]. In protein engineering, site-directed mutagenesis is used to generate DNA sequences with mutated codons, insertions, or deletions. In this widely used method, mutations are generated by PCR using a pair of oligonucleotide primers designed with mismatching nucleotides at the center of the primers as shown in Table 3.

Table 3: Primer DNA sequences used in site directed mutagenesis.

List of mutations	Primer DNA sequence
R1896A forward	5'CACCCCACCGCGATCTCCCTCGTCTTTTACCAGCAT3'
R1896A reverse	5' GAGGGAGATCGCGGTGGGGTGATTCCTATTGGGATTC 3'
R1896C forward	5' CACCCCACCTGCATCTCCCTCGTCTTTTACCAGCAT 3'
R1896C reverse	5' GAGGGAGATGCAGGTGGGGTGATTCCTATTGGGATTC 3'
R1896D forward	5' CACCCCACCGACATCTCCCTCGTCTTTTACCAGCAT 3'
R1896D reverse	5' GAGGGAGATGTCGGTGGGGTGATTCCTATTGGGATTC 3'
R1896E forward	5' CACCCCACCGAGATCTCCCTCGTCTTTTACCAGCAT 3'
R1896E reverse	5' GAGGGAGATCTCGGTGGGGTGATTCCTATTGGGATTC 3'
R1896F forward	5' CACCCCACCTTCATCTCCCTCGTCTTTTACCAGCAT 3'
R1896F reverse	5' GAGGGAGATGAAGGTGGGGTGATTCCTATTGGGATTC 3'
R1896G forward	5' CACCCCACCGGGATCTCCCTCGTCTTTTACCAGCAT 3'
R1896G reverse	5' GAGGGAGATCCCGGTGGGGTGATTCCTATTGGGATTC 3'
R1896H reverse	5' CACCCCACCCATATCTCCCTCGTCTTTTACCAGCAT 3'
R1896H forward	5' GAGGGAGATATGGGTGGGGTGATTCCTATTGGGATTC 3'

R1896I reverse	5’ CACCCCACCATCATCTCCCTCGTCTTTTACCAGCAT 3’
R1896I forward	5’ GAGGGAGATGATGGTGGGGTGATTCCTATTGGGATTC 3’
R1896K forward	5’ CACCCCACCAAGATCTCCCTCGTCTTTTACCAGCAT 3’
R1896K reverse	5’ GAGGGAGATCTTGGTGGGGTGATTCCTATTGGGATTC 3’
R1896L forward	5’ CACCCCACCTTGATCTCCCTCGTCTTTTACCAGCAT 3’
R1896L reverse	5’ GAGGGAGATCAAGGTGGGGTGATTCCTATTGGGATTC 3’
R1896M forward	5’ CACCCCACCATGATCTCCCTCGTCTTTTACCAGCAT 3’
R1896M reverse	5’ GAGGGAGATCATGGTGGGGTGATTCCTATTGGGATTC 3’
R1896N forward	5’ CACCCCACCAACATCTCCCTCGTCTTTTACCAGCAT 3’
R1896N reverse	5’GAGGGAGATGTTGGTGGGGTGATTCCTATTGGGATTC 3’
R1896P forward	5’ CACCCCACCCCGATCTCCCTCGTCTTTTACCAGCAT 3’
R1896P reverse	5’GAGGGAGATCGGGGTGGGGTGATTCCTATTGGGATTC 3’
R1896Q forward	5’ CACCCCACCCAGATCTCCCTCGTCTTTTACCAGCAT 3’
R1896Q reverse	5’ GAGGGAGATCTGGGTGGGGTGATTCCTATTGGGATTC 3’
R1896S forward	5’ CACCCCACCACGATCTCCCTCGTCTTTTACCAGCAT 3’

R1896S reverse	5’ GAGGGAGATCGTGGTGGGGTGATTCCTATTGGGATTC 3’
R1896T forward	5’ CACCCCACCACGATCTCCCTCGTCTTTTACCAGCAT 3’
R1896T reverse	5’GAGGGAGATCGTGGTGGGGTGATTCCTATTGGGATTC 3’
R1896V forward	5’ CACCCCACCGTG ATCTCCCTCGTCTTTTACCAGCAT 3’
R1896V reverse	5’ GAGGGAGATCACGGTGGGGTGATTCCTATTGGGATTC 3’
R1896W reverse	5’ CACCCCACCTGGATCTCCCTCGTCTTTTACCAGCAT 3’
R1896W reverse	5’ GAGGGAGATCCAGGTGGGGTGATTCCTATTGGGATTC 3’
R1896Y forward	5’ CACCCCACCTACATCTCCCTCGTCTTTTACCAGCAT 3’
R1896Y reverse	5’ GAGGGAGATGTAGGTGGGGTGATTCCTATTGGGATTC 3’

The R1896 residue of WT TET2 was mutated into all other 19 amino acids and these TET2 mutants were cloned into the same pDEST14 vector. WT TET2 enzyme and its mutants were purified by cation exchange chromatography on SP Sepharose, as described in our previous article[122]. Fractions containing purified untagged TET2 were pooled together and concentrated using ultrafiltration membranes (NMWL 30 KDa).

***In vitro* TET2 Reaction:** A 25-mer double stranded DNA used as TET2 substrate was generated by hybridizing single stranded oligomers on a thermal cycler. The *in vitro* TET2 enzymatic reactions were performed using a reaction matrix containing 50 mM HEPES (pH 8.0), 75 μM FeSO4, 1 mM 2-oxoglutarate, 2 mM ascorbate, 1 mM ATP and 1 mM DTT in 100 μl of total reaction volume. Equal amount of DNA substrate was added according to designated concentration and the reaction was started by addition of 2.5 μg purified TET2. After 1 hour of incubation at 37 °C, TET2 catalyzed oxidation reactions were quenched with 5 μl of 500 mM EDTA. Total dsDNA from the reaction mixture was separated using the superior oligo purification kit following manufacturer's protocol (Superior Scientifics, Lenexa, KS). The isolated DNA (20 μl) was digested with 2 units of DNase I and 60 units of S1 nuclease in 40 μl at 37 °C for 12 hours to produce individual nucleotide monophosphates. Following the digestion, 2 units of calf intestinal alkaline phosphatase (CIAP) was added in the samples followed by incubation for an additional 4 hours at 37 °C to remove the terminal phosphate groups to obtain nucleosides.

The amount of oxidation products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (∆5hmC/∆C/#C). The amount of each oxidative product (picomoles) was

calculated using the standard curve (Figure S1). The activity of mutant TET2 enzymes were quantified the same way and the amount of total oxidation product (picomoles) was presented in comparison with the WT TET2 activity. Standard deviations and standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

Liquid Chromatography of Nucleosides under different MS/MS modes: The LC-MS/MS analysis was performed on Sciex 3200 QTrap mass spectrometer (Foster City, CA) coupled to a Shimadzu UFLC LC-20 system (Columbia, MD) using electrospray ionization source (ESI) and run with Analyst v 1.6.2 software. For chromatographic separation of nucleosides in ion-switching mode, water/methanol based solvent system was used where solvent A was water (adjusted to pH 3.5 using formic Acid) and solvent B was methanol (adjusted to pH 3.0 using formic Acid). The gradient used was 0% B (0-1 minute), 0-2% B (1-12 minute), 2-30% B (12-17 minute), 30% B (17-18 minute), 30-0% B (18-18.5 minute), followed by a 4.5-minute equilibration at 0% B at a flow rate of 0.3 mL/minute. Nucleosides were separated on C18 column (Dimension 150×2 mm, particle size 5µM, Pore Size 100 Å). Standard curves, the limit of detection (LOD), lower limit of quantification (LLOQ) and matrix effect for all eight nucleosides in the positive and negative mode has been calculated by our developed method. Optimization of mass spectrometric parameters and standard curves for modified cytosines have been published in our previous work[123].

5.3 Results and discussion

Tet2 mutations are commonly found in myeloid cancers, including myelodysplastic syndromes, chronic myelomonocytic leukemia, and myeloproliferative neoplasm[28,29]. These mutations can be caused by various factors such as splice site mutations, out-of-

frame insertions or deletions and base substitutions[142–145]. The reported crystal structures of 2OG-dependent dioxygenases allowed prediction of hTET2 active site showing a β-strand core containing eight sheets folded into a distorted "jellyroll" motif with an active site containing two histidine and a carboxylate ligand, occupying one face of the iron (II) coordination sphere[10,146]. Of the remaining three sites of the iron (II) coordinate, the carboxyl, and keto oxygens of 2OG bind to the metal center in bidentate manner. The last sites were occupied by O_2, trans to any one of three ligands of the 2-His-1-carboxylate facial triad. The 5-carboxylate group of 2OG binds to guanidino group of arginine-R1896 and hydroxyl group of serine-1898 (present on the eighth β-sheet of the "Jellyroll") from the hTET2 enzyme (Figure 19). Several clinical mutations have been identified in the hTET2 R1896 and S1898 residues (*e.g.,* R1896M and S1898F) which leads to higher 5mC levels in patients[41]. However, limited characterizations of these mutations have been reported in the literature partly due to lack of a quantitative assay.

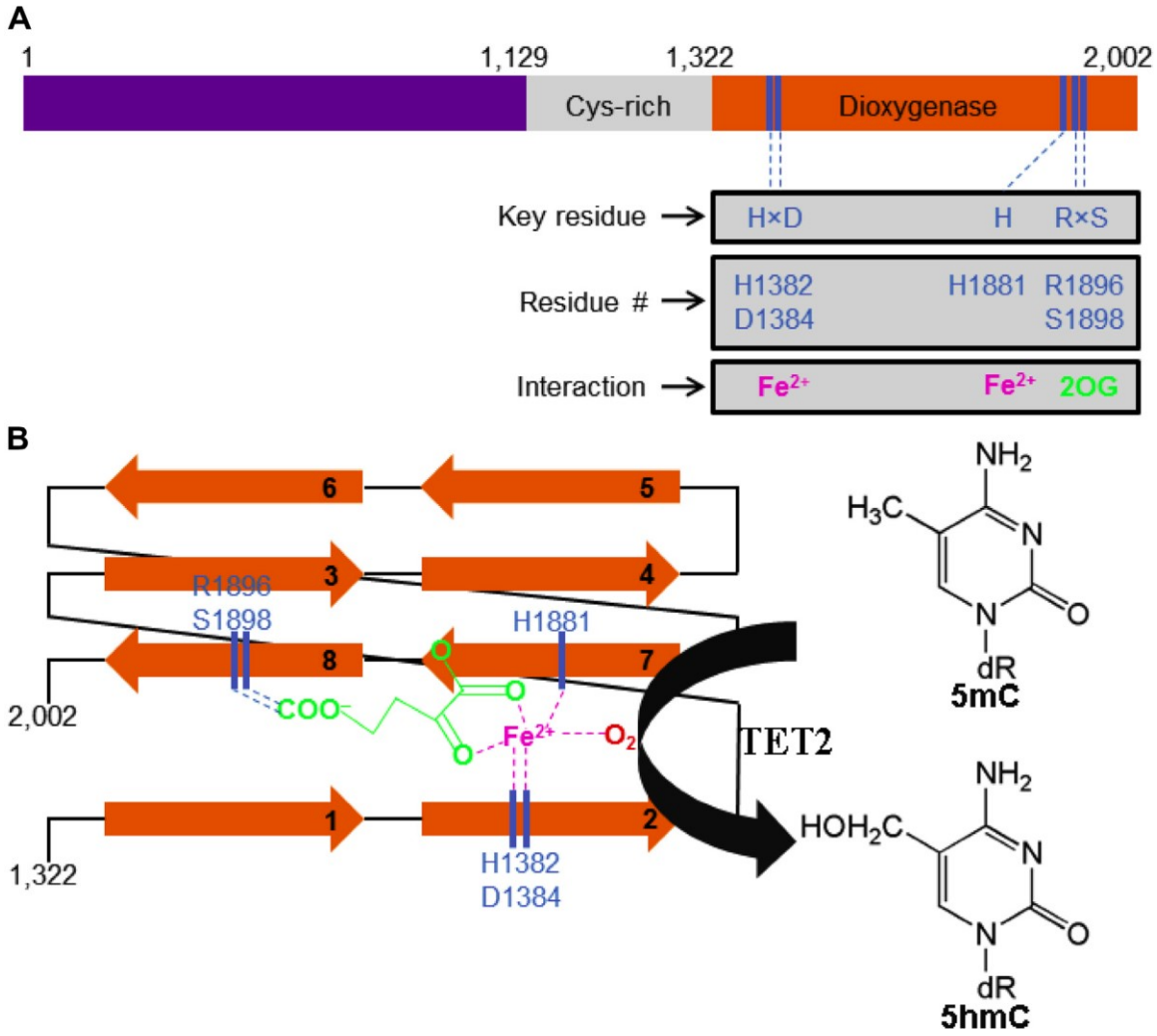


Figure 19: Panel A provides a graphical depiction of the complete structure of TET2, highlighting both the cysteine-rich domain and catalytic dioxygenase domain whereas, panel B presents an active site representation of TET2, emphasizing the presence of eight antiparallel β-sheets that form the DSBH domain also referred as the jelly roll motif [71]. This motif serves as the binding site for the co-substrate, 2OG, as well as the cofactors, Fe (II) and O2.

Recent studies have reported that modified 2OG analogs can rescue or inhibit the catalytic activity of various 2OGDD family mutants[139,140]. Successful attempts have been made to modify 2OG (natural cofactor for 2OGDDs) at carbons 3, 4, and 5, resulting in natural and nonnatural 2OG analogs. For instance, Phytanoyl-CoA 2- hydroxylase (PAHX) is frequently mutated in Refsum's disease where point mutations occur in the R275 residue. This arginine residue is involved in interaction with the 5-carboxylate of 2OG. 2OG analogs have been used to rescue mutations in R275 residue of PAHX by replacing the 5-carboxylate group of 2OG with aliphatic chains[75]. Similarly, another recent study has reported that mutant enzymes of the KDM4 family of histone lysine demethylases (KDM4A-E) can utilize 2OG analogs and the 2OG derivatives do not alter activity of the wild type (WT) enzymes. By performing alanine scanning, this study identified point mutations in the active site of KDM4A-E which can be rescued by 2OG derivatives, carrying aliphatic modification at C4[147].

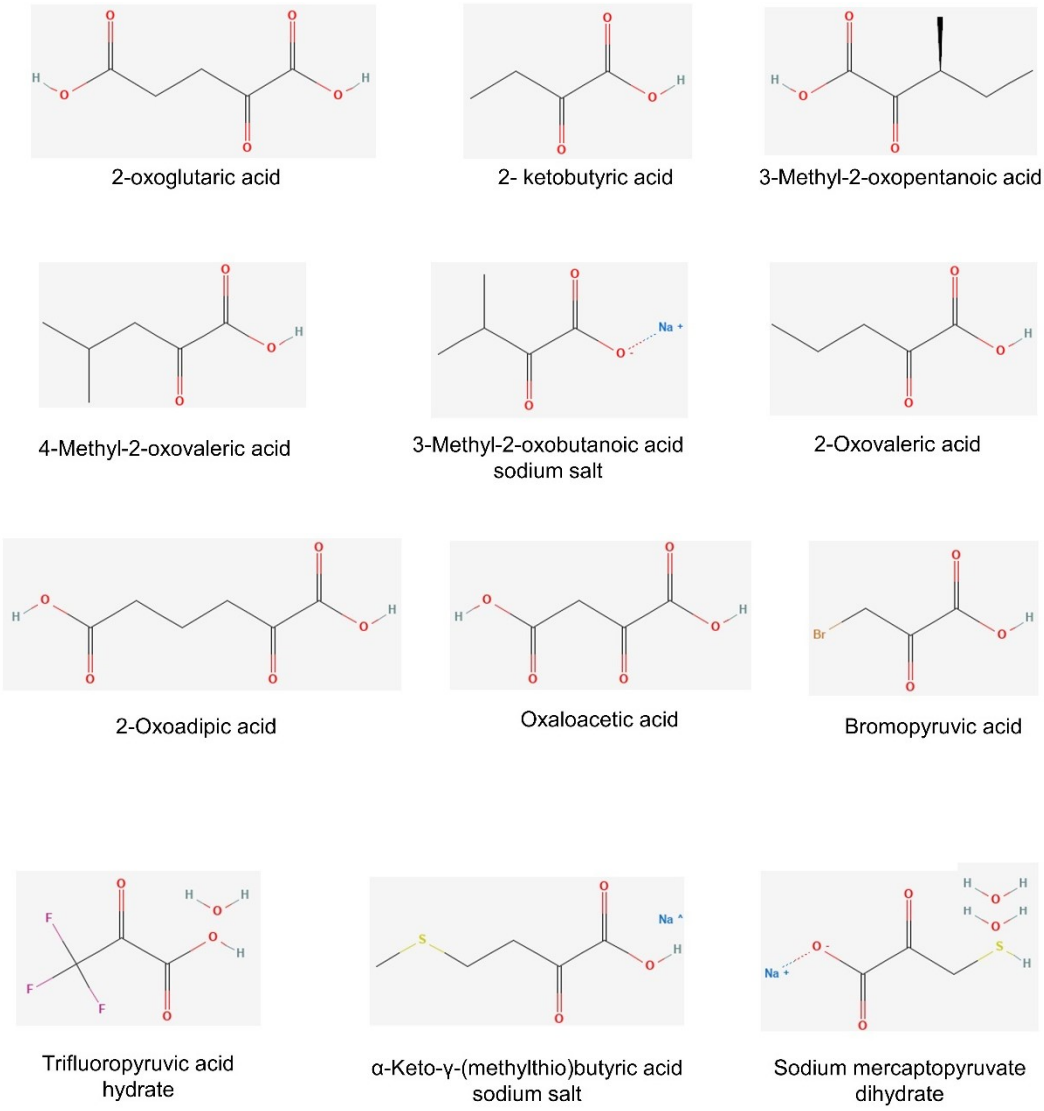


Figure 20: Chemical structure of 2OG analogs used for rescuing mutant TET2 activity.

The crystal structure of the TET2-DNA complex showed that the 5-carboxylate moiety of 2OG is stabilized by residues H1416, R1896, and S1898 of TET2[72]. Mutations in the residue R1896 have been repeatedly identified in patients with myeloid cancers[41]. Interestingly, some of the most frequently mutated residues, i.e., TET2- R1896M, -R1896G, and -R1896S, (-R1896M/S/G) bind the 5-carboxylate group of 2OG in the TET2 dioxygenase active site. The clinical mutations of TET2 in R1896 residue result in an unfavorable interaction between sidechains of M1896/ G1896/S1896 and 5-carboxylate of 2OG. It is tempting to speculate that the impaired binding of 2OG in these mutants possibly account for the loss of TET2 activity in patients with myeloid malignancies. Previous attempts to rescue mutations in the 2OGDD family of enzymes inspired us to investigate the effect of 2OG analogs in TET2 R1896 mutations. For the initial screening, three clinical mutations at the R1896 residue were selected: R1896G, R1896M, and R1896S. The mutant enzymes were cloned by site-directed mutagenesis on the minimum catalytically active domain of human TET2 (1129-1936, Δ1481-1843). Next, we designed a set of 2OG analogs (Figure 20) and used them as replacements for 2OG in our in vitro assay. Catalytic activity of the mutants in the presence of 2OG analogs was measured based on the total oxidation products formed in the *in vitro* reactions. Preliminary screening of 2OG analogs revealed that W.T. TET2 activity is most efficient in presence of 2OG. Replacing 2OG with its derivatives significantly diminished the oxidation efficiency of W.T. TET2 to 25% or lower. W.T. TET2 loss of activity was evident across all 11 2OG analogs (Figure 21).

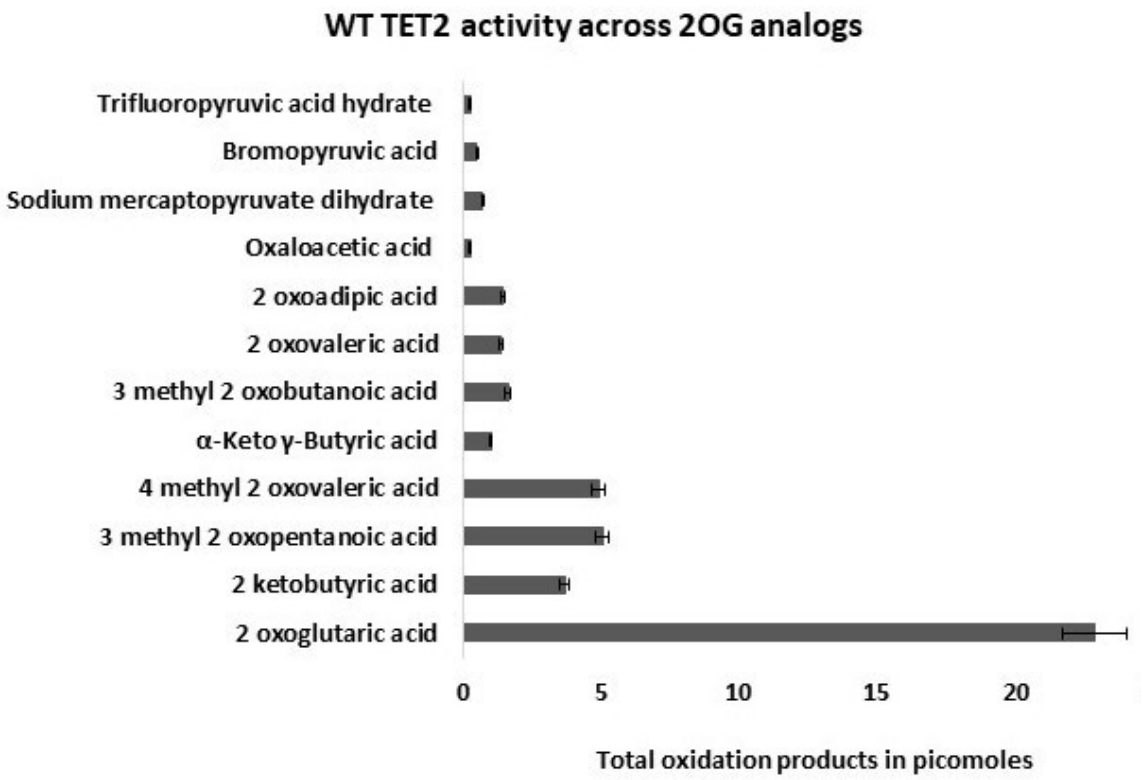


Figure 21: W.T. TET2 activity across the 2OG analogs. The enzyme demonstrated highest catalytic activity in presence of 2OG and replacing 2OG with the modified analogs lowered its activity by ≥ 4 folds. The 'X' axis represents total oxidation products in picomoles. The amount of products formed during hTET2-mediated oxidation were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figures.

However, replacing the C5 carboxyl group with 1-2 carbon long aliphatic chain moderately enhanced the TET2 mutants R1896G and R1895M. For instance, TET2 R1896G in the presence of 2OG achieved 30% oxidation efficiency compared with that of the WT enzyme. However, replacing 2OG with 3-methyl 2-oxopentanoic acid rescued its activity by more than 8% (Figure 22). A few other 2OG analogs, such as 2-ketobutyric acid, 4-methyl 2-oxovaleric acid, and α-keto γ-butyric acid, demonstrated similar enhancement of the TET2 mutant R1896G. In the case of TET2 R1896M, we observed an almost 13% enhancement in activity when 2OG was replaced with 2-ketobutyric acid (Figure 23).

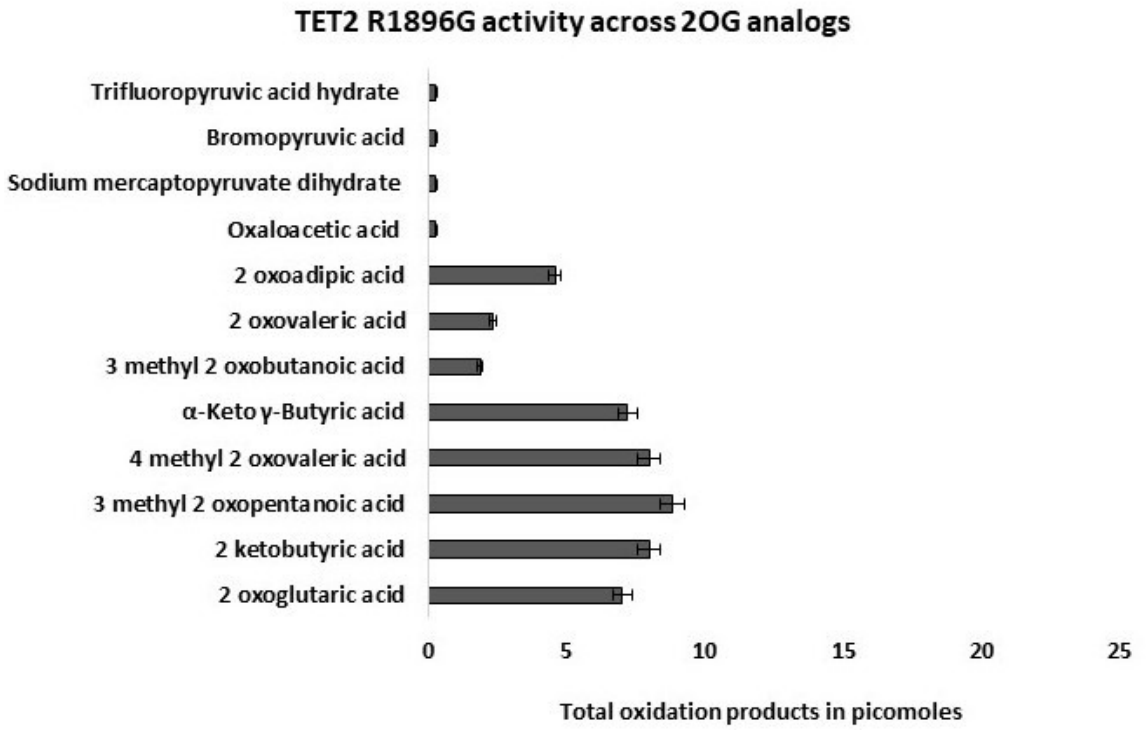


Figure 22: TET2 R1896G activity across 2OG analogs. The mutant enzyme demonstrated a lower catalytic activity than the wild-type enzyme in the presence of 2OG. However, replacing 2OG with modified analogs such as 2-ketobutyric acid, 3 methyl 2 oxopentanoic acid, and 4-methyl 2-oxovaleric acid slightly enhanced TET2 R1896G activity. The 'X' axis represents total oxidation products in picomoles. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (∆5hmC/∆C/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figures.

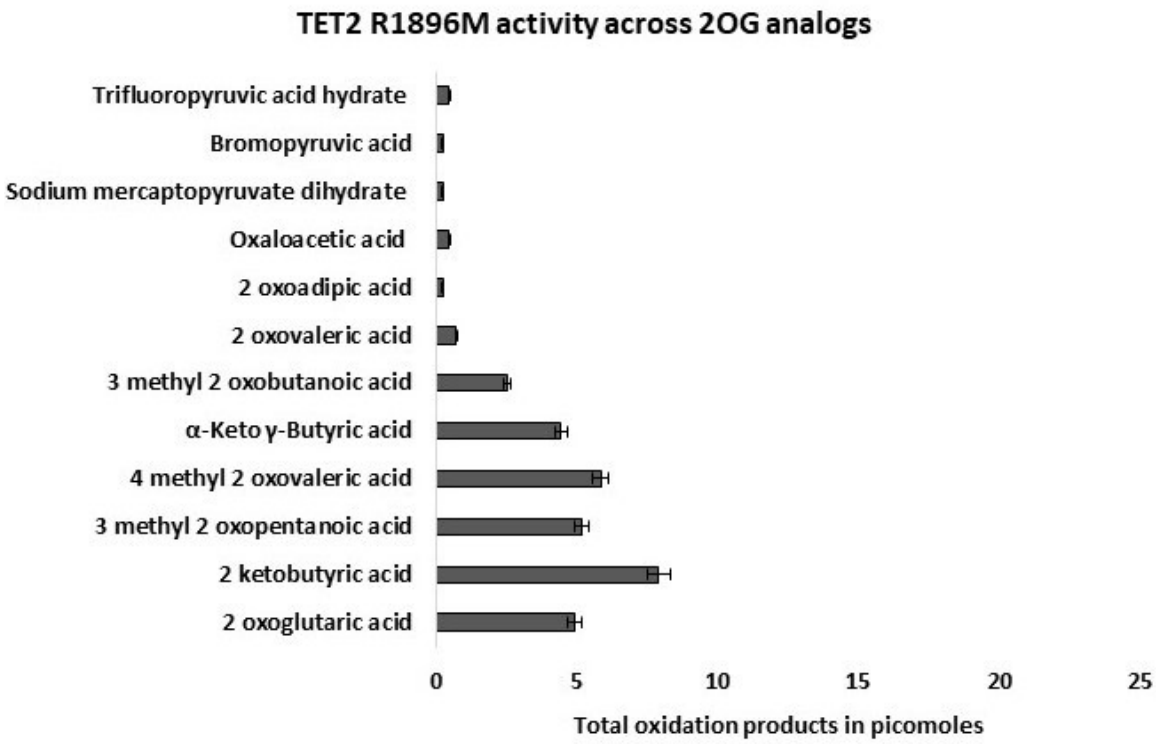


Figure 23: TET2 R1896M activity across 2OG analogs. The mutant enzyme demonstrated a lower catalytic activity than the wild-type enzyme in the presence of 2OG. However, replacing 2OG with modified analogs such as 2-ketobutyric acid, 4-methyl 2-oxovaleric acid slightly enhanced TET2 R1896G activity. The 'X' axis represents total oxidation products in picomoles. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figures.

The most remarkable enhancement of catalytic activity was observed for TET2 R1896S when the natural co-substrate 2OG was replaced by its derivatives. TET2 catalytic activity declined by almost 75% when arginine-1896 was mutated to serine because of the unfavorable interaction between the 5-carboxylate of 2OG and serine. Removal of 5-carboxylate group in 2OG or its replacement by 1-2 carbons long aliphatic side chains significantly rescued catalytic activity of TET2 R1896S (Figure 24). 2OG analogs such as 2-ketobutyric acid and 4-methyl 2-oxovaleric acid rescued activity of this mutant enzyme up to 93% and 84% respectively compared to the activity of the WT enzyme with 2OG (Figure 25).

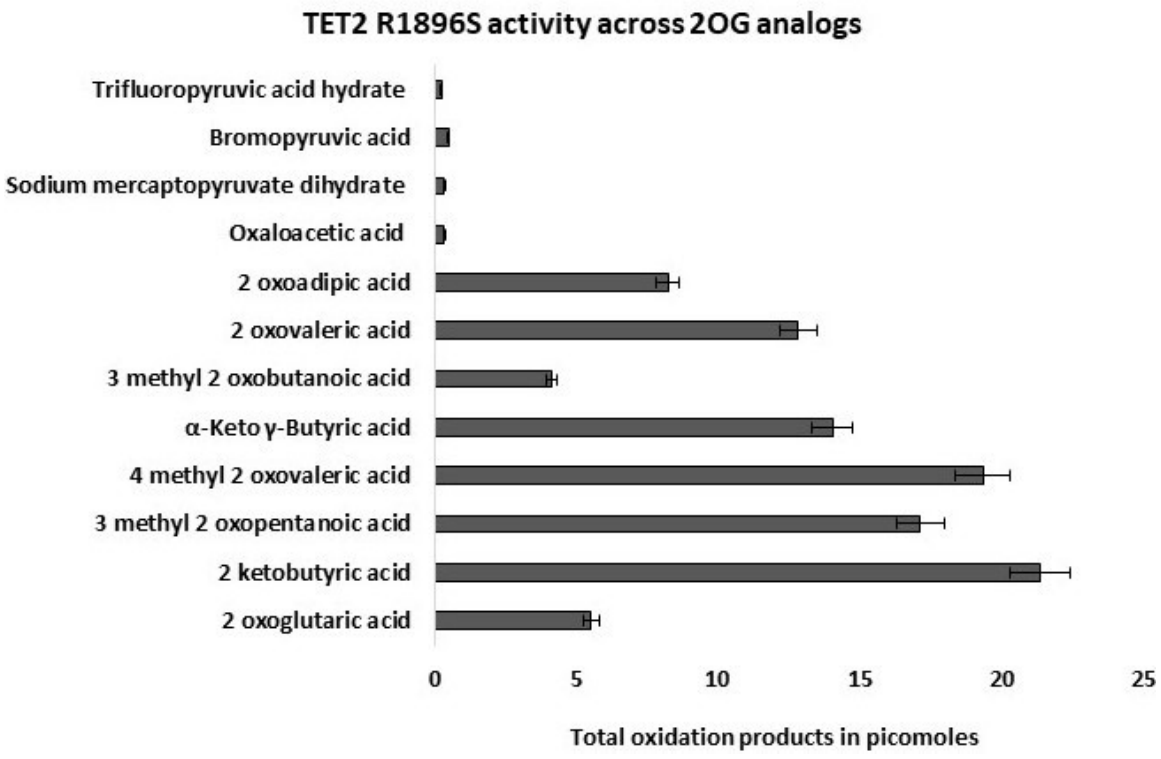


Figure 24: TET2 R1896S activity across the 2OG analogs. The mutant enzyme demonstrated a lower catalytic activity than the wild-type enzyme in the presence of 2OG. However, replacing 2OG with modified analogs such as 2-ketobutyric acid, 3 methyl 2 oxopentanoic acid, 4-methyl 2-oxovaleric acid, α-keto γ-butyric acid, 2 oxovaleric acid, and 2 oxoadipic acid significantly enhanced TET2 R1896G activity. The 'X' axis represents total oxidation products in picomoles. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figures.

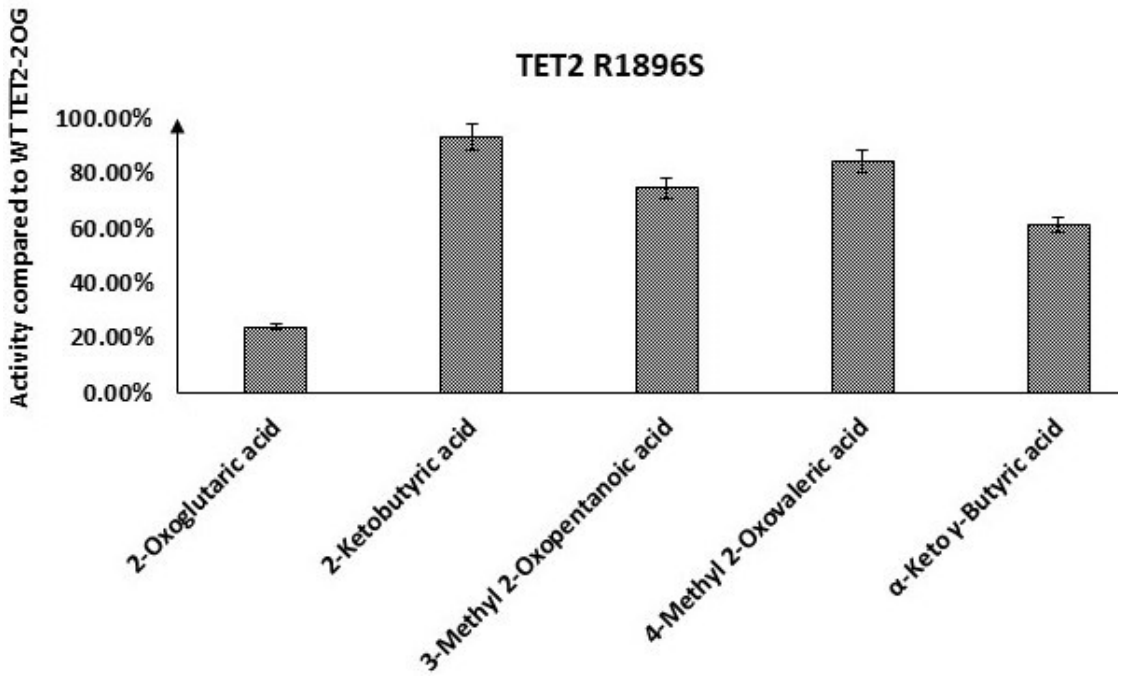


Figure 25: TET2 R1896S activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. The mutant enzyme demonstrated a lower catalytic activity than the wild-type enzyme in the presence of 2OG. However, replacing 2OG with modified analogs such as 2-ketobutyric acid, 3 methyl 2 oxopentanoic acid, 4-methyl 2-oxovaleric acid, α-keto γ-butyric acid, 2 oxovaleric acid, and 2 oxoadipic acid enhanced its activity by more than 4 folds. The 'X' axis represents total oxidation products in picomoles. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figures.

Of the 11 2OG derivatives, we identified six compounds that significantly enhanced and rescued the catalytic activity of TET2 R1896S. Based on this finding, we identified four 2OG analogs (2-ketobutyric acid, 3-methyl 2-oxopentanoic acid, 4-methyl 2-oxovaleric acid, and α-keto γ-butyric acid) that most efficiently rescued TET2 R1896S activity and tested these compounds against all possible mutations in the R1896 residue. Point mutations were created by site-directed mutagenesis on the same DNA backbone as *TET2* 1129-1936, Δ 1481-1843. The catalytic activity of the TET2 R1896 mutants was characterized in the presence of the natural co-substrate 2OG, and its analogs. All mutants exhibited reduced 5mC oxidation efficiency in the presence of 2OG compared with that of the WT enzyme. Additionally, two other mutants, TET2 R1896A and TET2 R1896F, exhibited enhanced catalytic activity when the natural co-substrate 2OG was replaced with 2OG derivatives. TET2 R1896A (Figure 26) produced almost three-fold more 5-mC oxidation products when conjugated with 2OG derivatives whereas R1896F produced two-fold more oxidation products in the presence of 2OG derivatives (Figure 27).

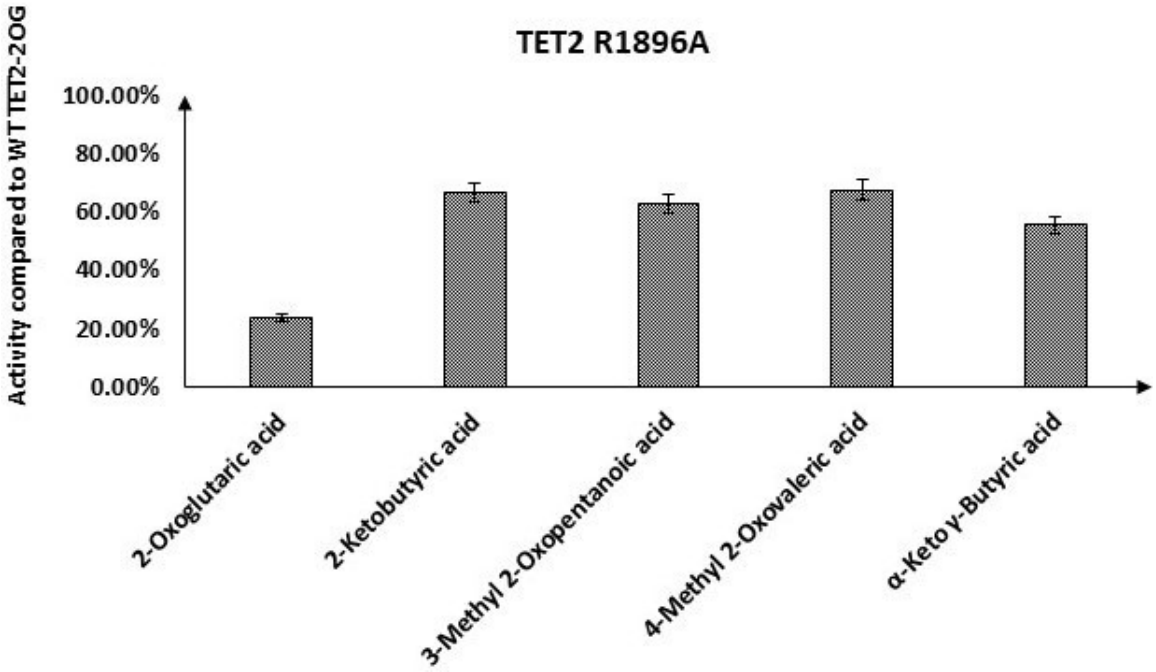


Figure 26: TET2 R1896A activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. The mutant enzyme demonstrated a lower catalytic activity than the wild-type enzyme in the presence of 2OG. However, replacing 2OG with modified analogs such as 2-ketobutyric acid, 3 methyl 2 oxopentanoic acid, 4-methyl 2-oxovaleric acid, α-keto γ-butyric acid, 2 oxovaleric acid, and 2 oxoadipic acid enhanced its activity by ≥ 3 folds. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figures.

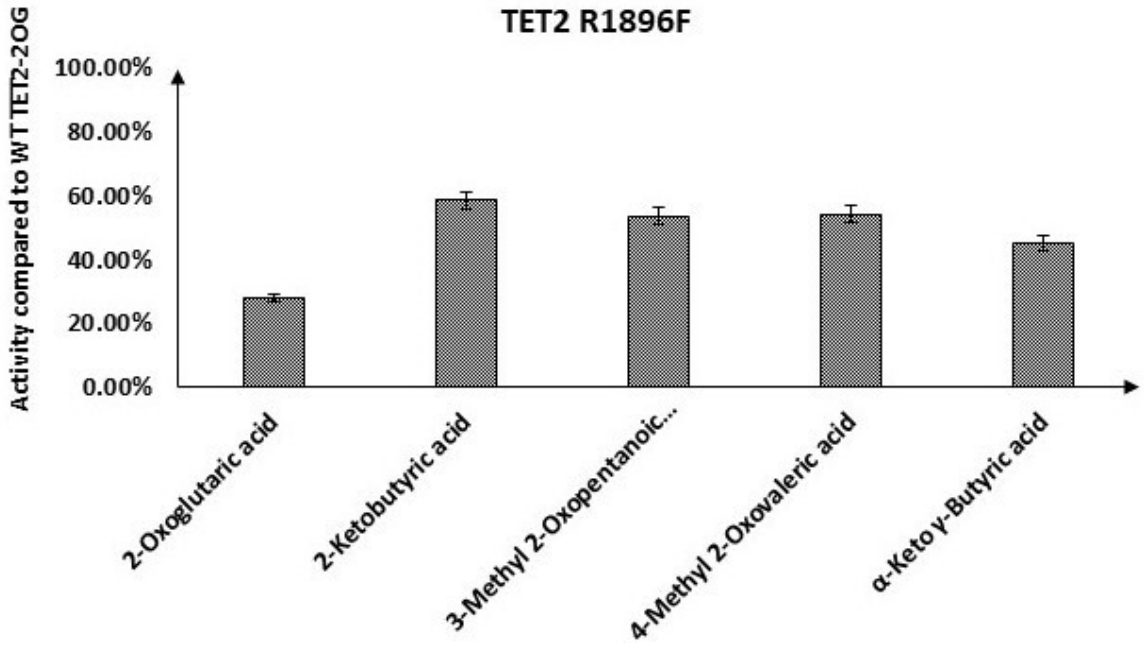


Figure 27: TET2 R1896F activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. The mutant enzyme demonstrated a lower catalytic activity than the wild-type enzyme in the presence of 2OG. However, replacing 2OG with modified analogs such as 2-ketobutyric acid, 3 methyl 2 oxopentanoic acid, 4-methyl 2-oxovaleric acid, α-keto γ-butyric acid, 2 oxovaleric acid, and 2 oxoadipic acid enhanced its activity by ≤ 3 folds. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (∆5hmC/∆C/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figures.

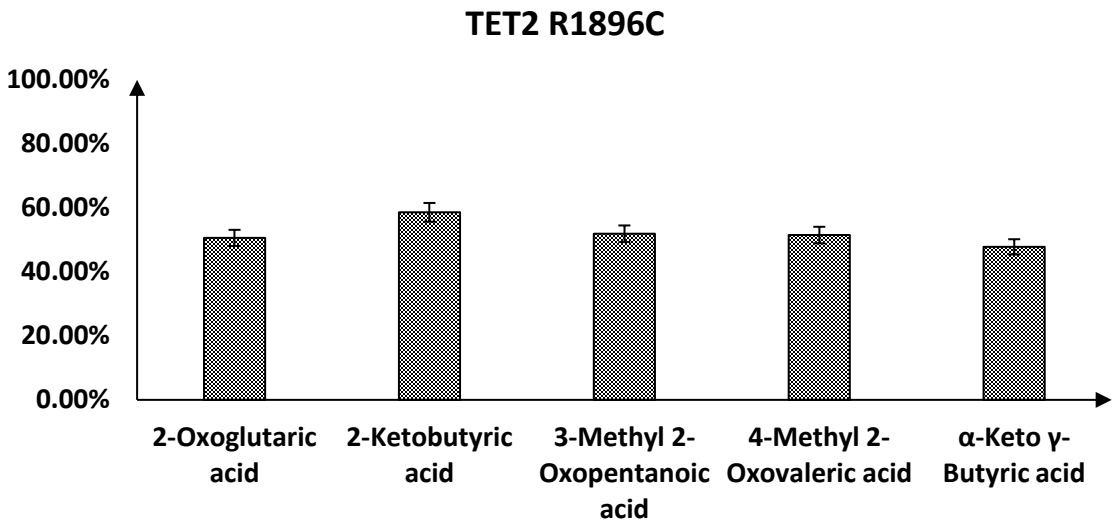


Figure 28: TET2 R1896C activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with the other 2OG derivatives did not have a significant change in catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

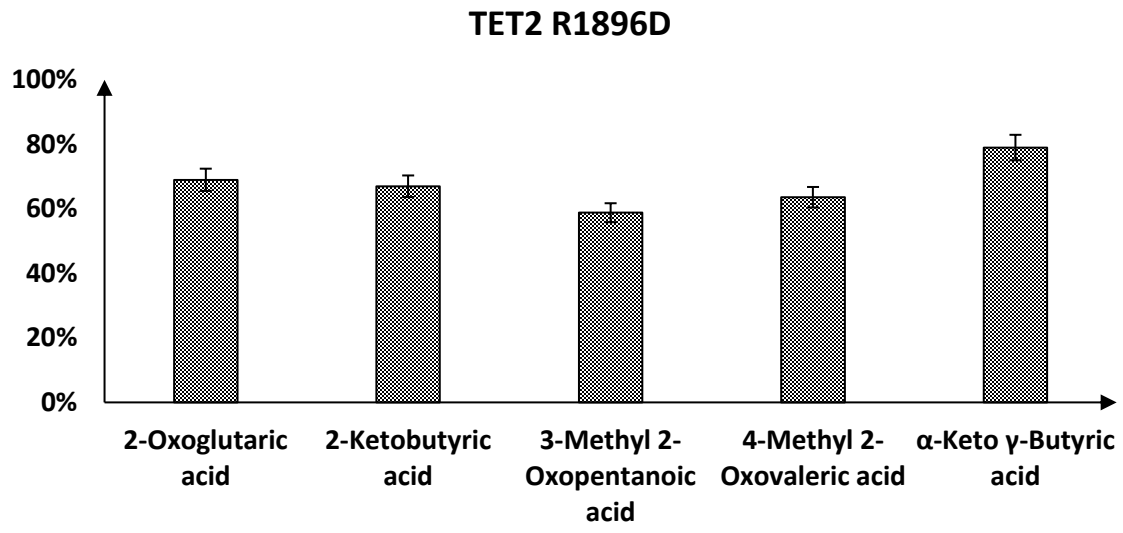


Figure 29: TET2 R1896D activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with α-keto γ-butyric acid slightly enhanced catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

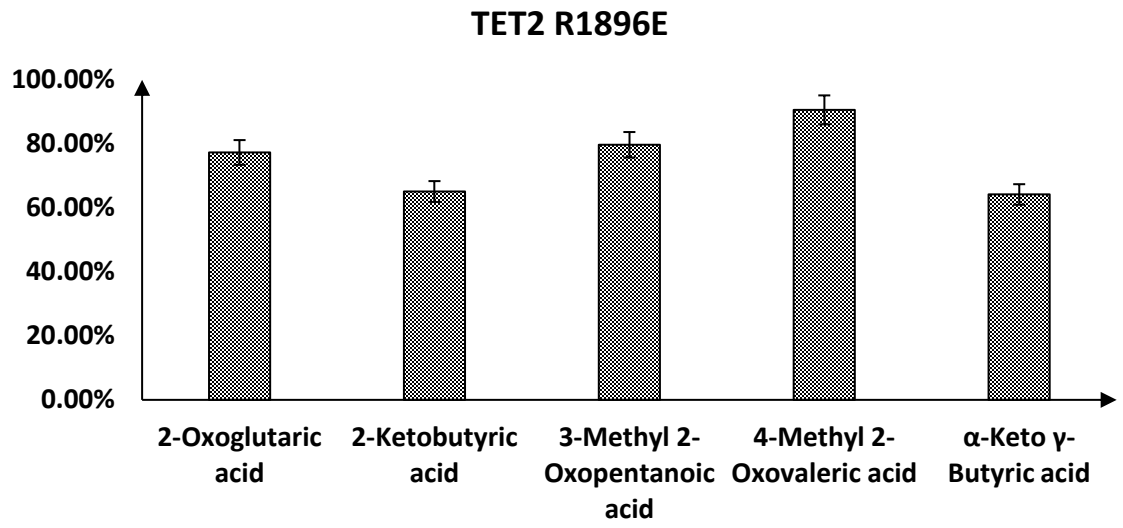


Figure 30: TET2 R1896E activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 4-methyl 2- oxovaleric acid slightly enhanced catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

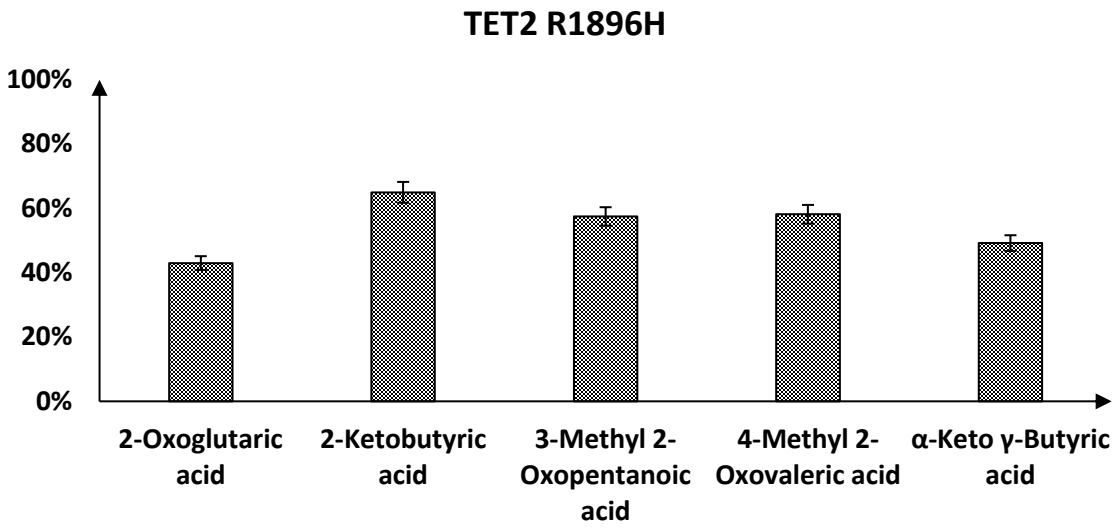


Figure 31: TET2 R1896H activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with the other 2OG derivatives slightly enhanced catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

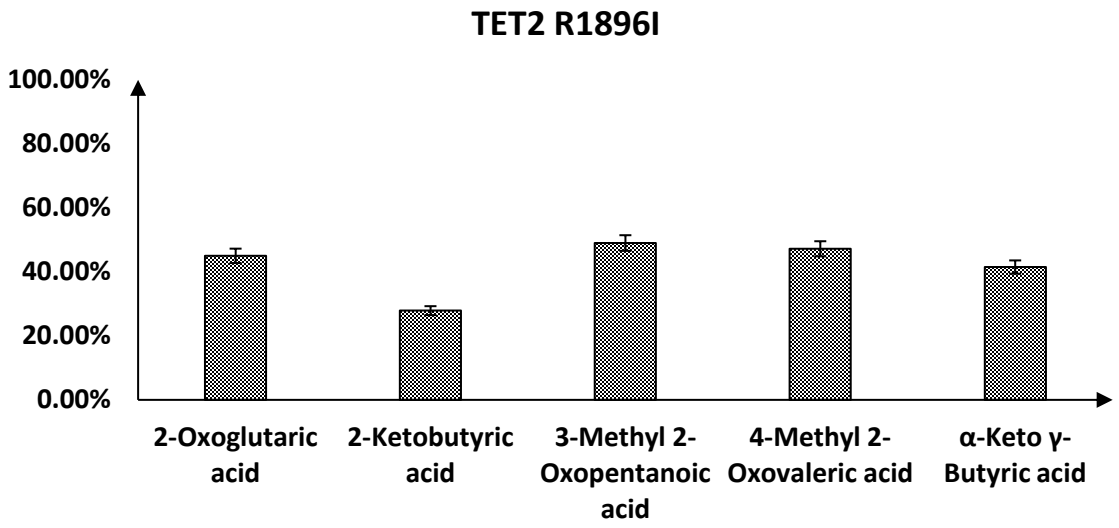


Figure 32: TET2 R1896I activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 2-ketobutyric acid lowered the mutant enzyme activity. However, the other 2OG derivatives did not have any significant effect on catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

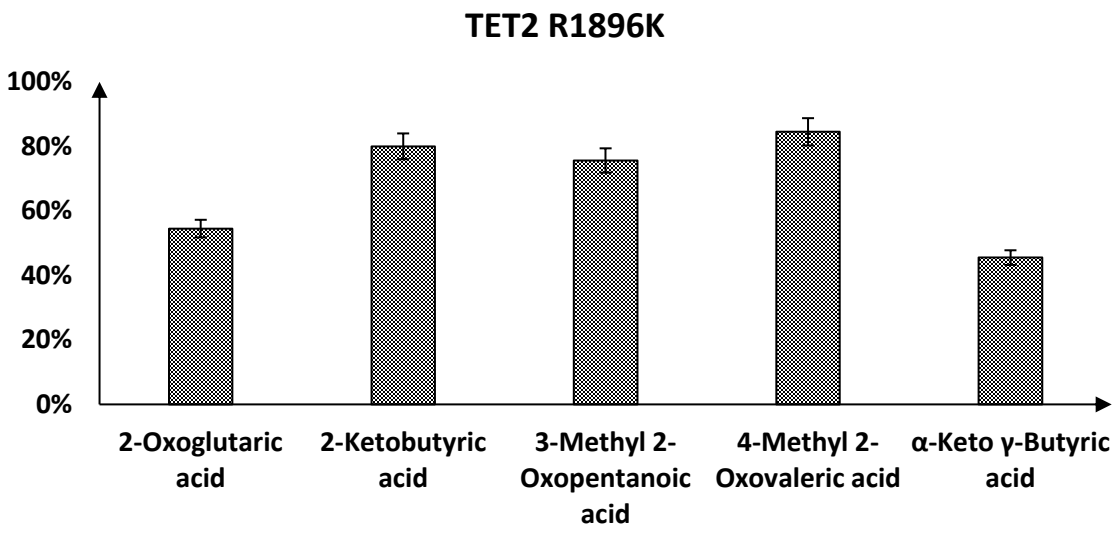


Figure 33: TET2 R1896K activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 2-ketobutyric acid, 3-methyl 2-oxopentanoic acid, 4-methyl 2-oxovaleric acid slightly enhanced catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

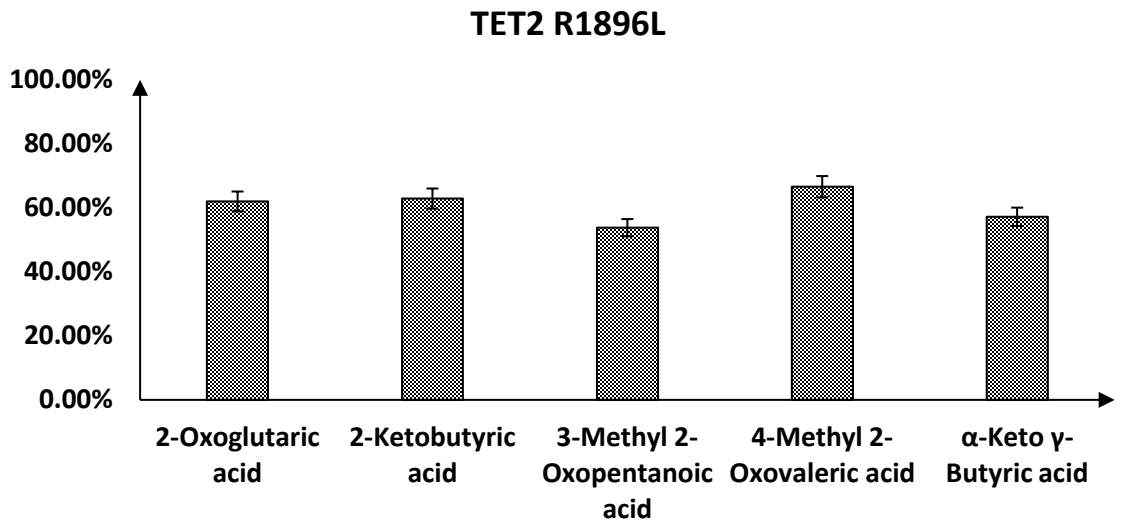


Figure 34: TET2 R1896L activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with the 2OG derivatives did not have any significant effect on catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

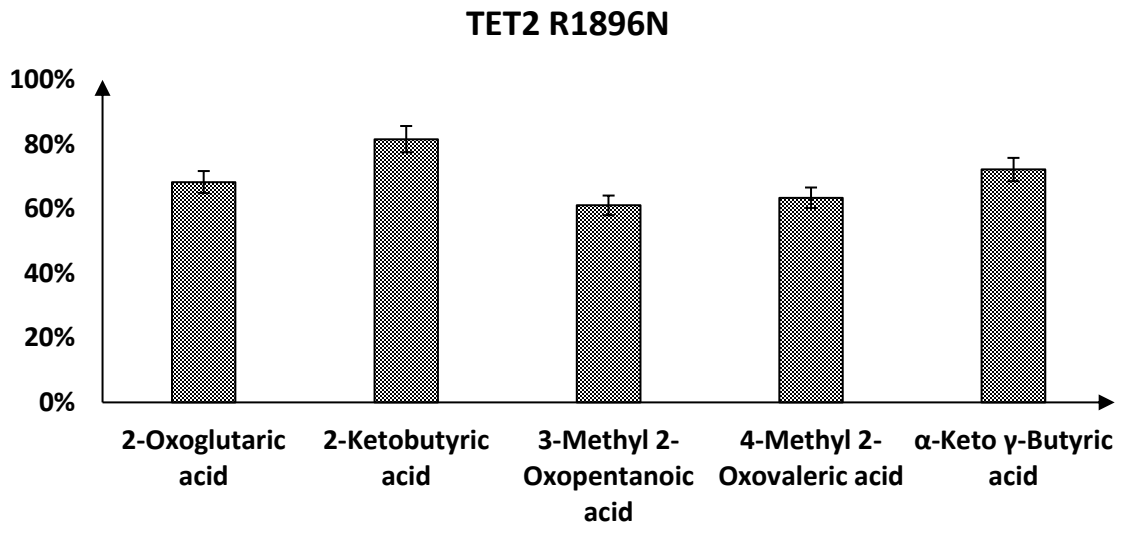


Figure 35: TET2 R1896N activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 2-ketobutyric acid slightly enhanced catalytic activity of the mutant enzyme. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

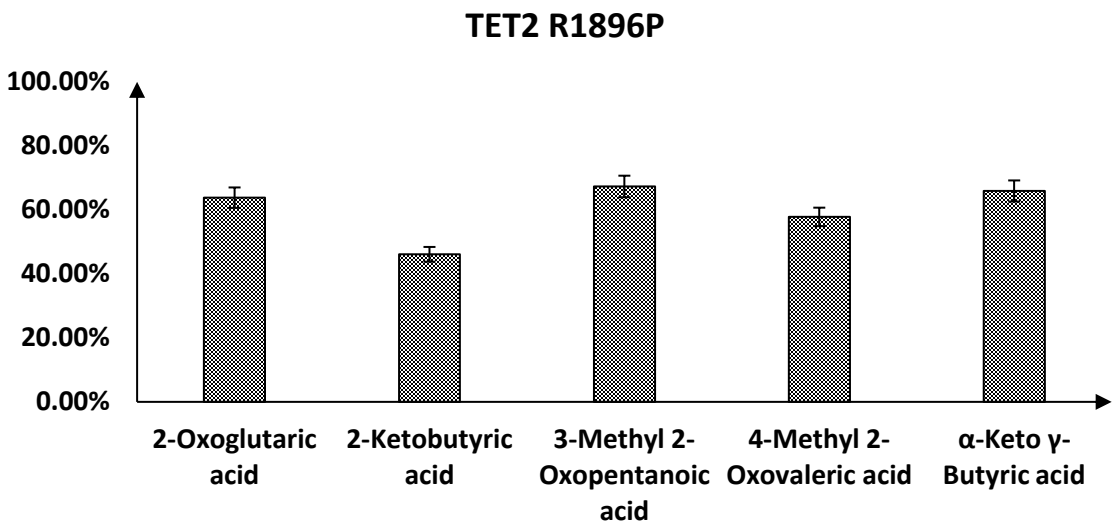


Figure 36: TET2 R1896P activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 2-ketobutyric acid slightly lowered catalytic activity of the mutant enzyme. However, 3-methyl 2-oxopentanoic acid, 4-methyl 2-oxovaleric acid and α-keto γ-butyric acid did not have any significant effect on TET2 R1896P activity. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

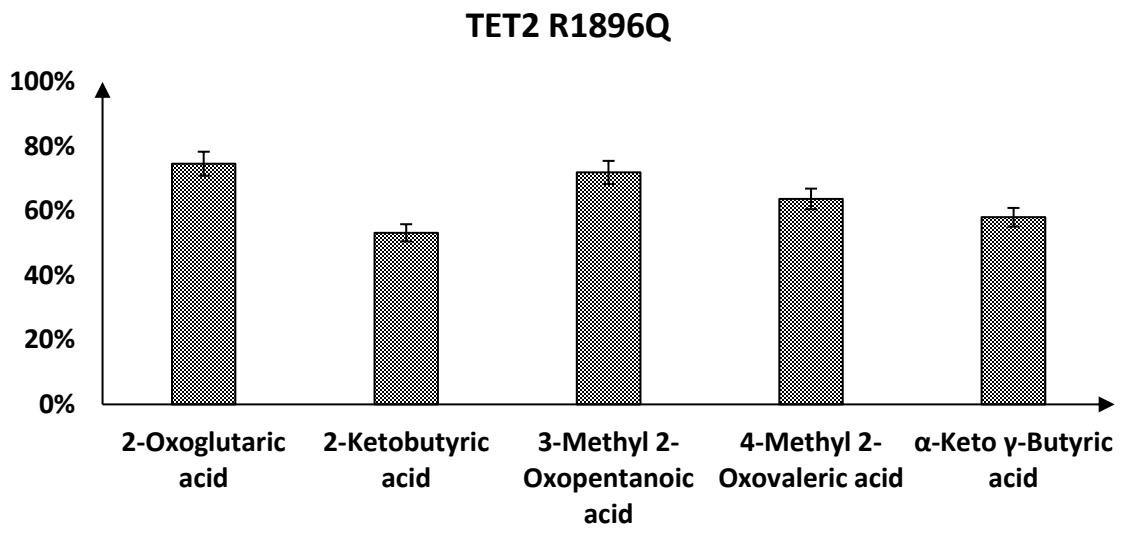


Figure 37: TET2 R1896Q activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 2-ketobutyric acid slightly lowered catalytic activity of the mutant enzyme. However, 3-methyl 2-oxopentanoic acid, 4-methyl 2-oxovaleric acid did not have any significant effect on TET2 R1896P activity. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

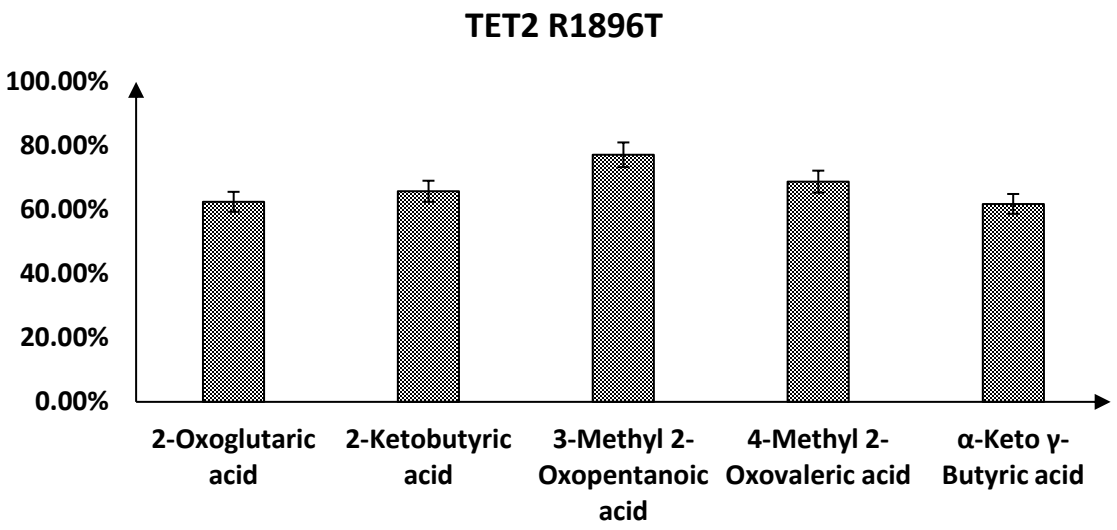


Figure 38: TET2 R1896T activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 3-methyl 2-oxopentanoic acid slightly enhanced TET2 R1896T activity. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

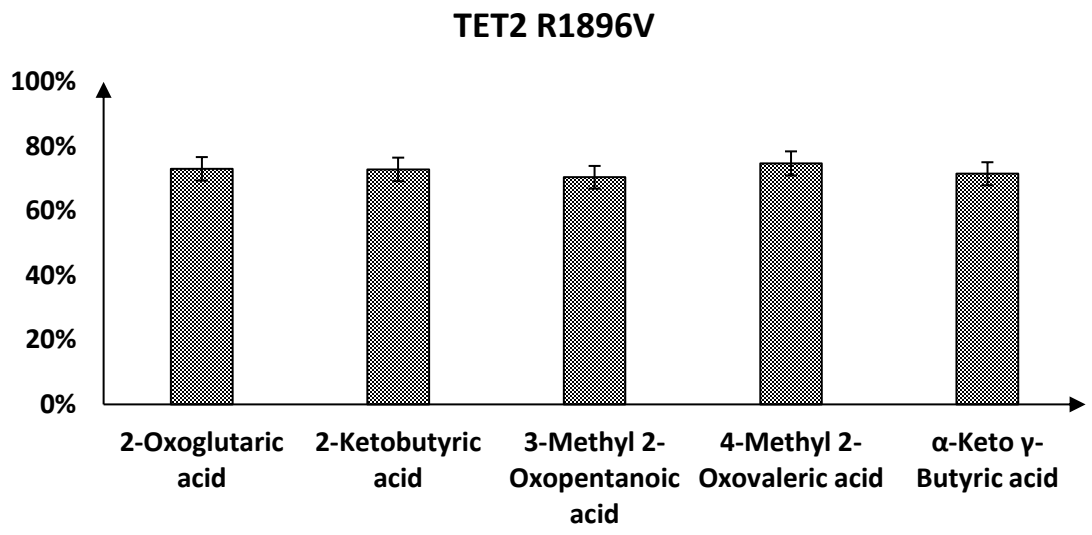


Figure 39: TET2 R1896V activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with other 2OG derivatives did not have any significant effect on TET2 R1896V activity. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

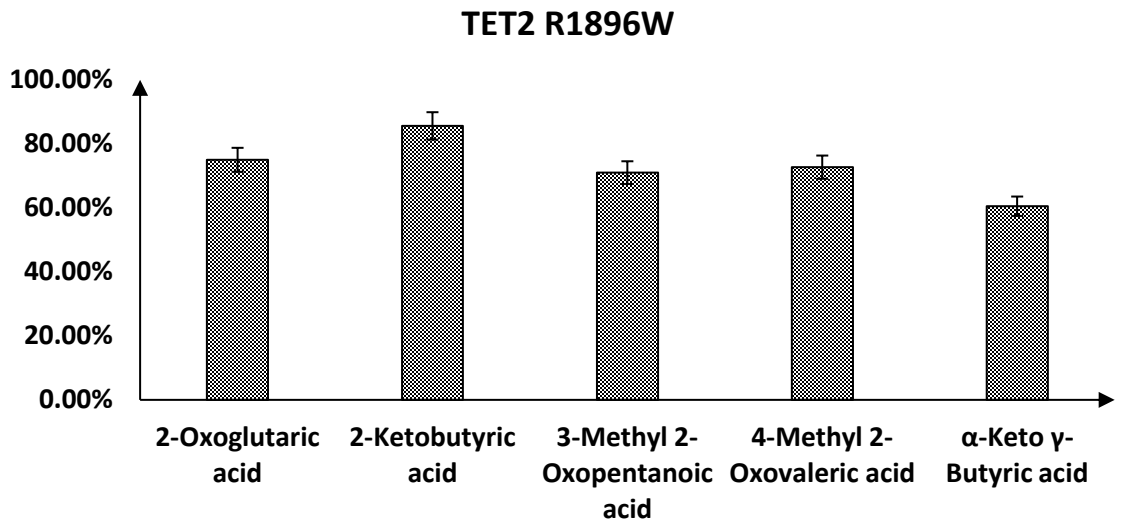


Figure 40: TET2 R1896W activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with other 2OG derivatives did not have any significant effect on TET2 R1896W activity. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

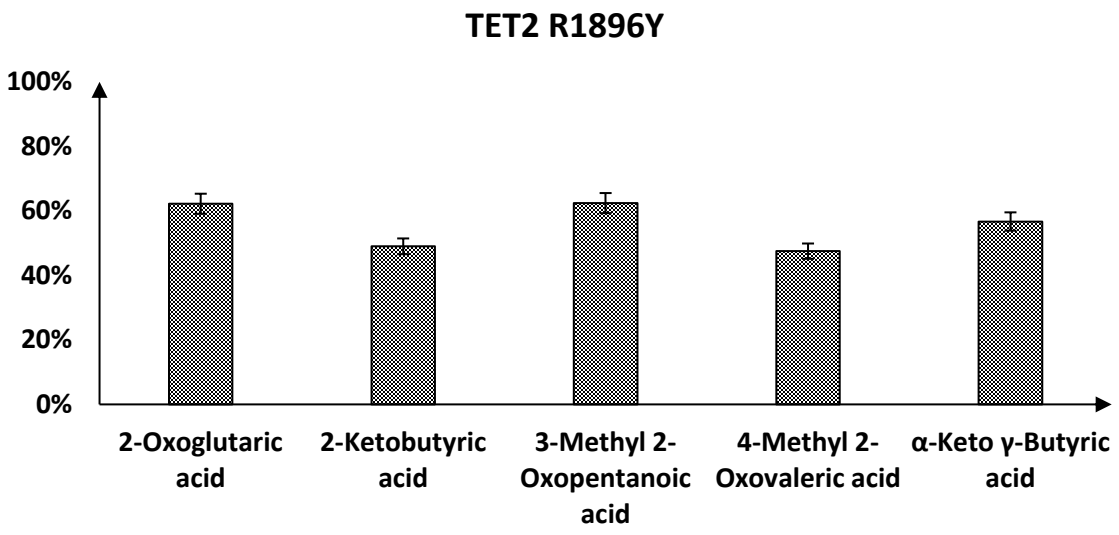


Figure 41: TET2 R1896Y activity across the 2OG analogs. Total oxidation has been expressed in comparison with the activity of W.T. TET2 in presence of 2OG. Replacing 2OG with 2-ketobutyric acid and 4-methyl 2-oxovaleric acid slightly lowered catalytic activity of the mutant enzyme. However, 3-methyl 2-oxopentanoic acid and α-keto γ-butyric acid did not have any significant effect on TET2 R1896Y activity. The amount of product formed during hTET2-mediated oxidation was normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using a standard curve (Figure 11). Standard errors were calculated from triplicate experiments and are represented in the figure.

However, replacing the C5 carboxylate moiety with an aliphatic chain, marginally enhanced the catalytic activity of the mutants R1896C/D/E/H/I/K/L/N/P/Q/T/V/W/Y (figures 28-41). Functional characterization of these TET2 mutants carrying point mutations in the R1896 residue helped us identify natural 2OG analogs that can specifically rescue mutant TET2 activity. Furthermore, the identified 2OG derivatives are products of amino acid metabolism and are naturally abundant in eukaryotic cells, suggesting that chemically modified 2OG analogs may act as useful tools to modulate mutant TET2 activity in disease conditions. The importance of such chemical tool is highly significant in cancer therapeutics as well as in assay development for the TET family of dioxygenases.

CHAPTER 6

FUNCTIONAL CHARACTERIZATION OF ACTIVE SITE MUTATIONS IN TET METHYLCYTOSINE DIOXYGENASE 2

6.1 Introduction

The C5 position of cytosine bases within CpG dinucleotides (mCG) is the predominant methylation site in mammalian genomes[53]. DNA methylation plays critical roles in tissue-specific gene expression, X chromosome inactivation, gene imprinting, and nuclear reprogramming[59]. 5mCs, especially when clustered, are important transcriptional silencers at gene promoters and endogenous retrotransposons in the genome[54,56,88]. Cytosine methylation is catalyzed by a family of DNA methyltransferases, and deficiencies in these enzymes result in profound developmental defects[60]. The removal of 5mC marks is initiated by TET family of dioxygenases (TET1-3)[61,62]. TET enzymes oxidize 5mC into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) through iterative oxidation steps. Finally, thymine-DNA glycosylase (TDG) replaces 5caC with unmodified cytosine via the base excision repair (BER) pathway[63,65,71]. The crystal structure of TET2 bound to dsDNA demonstrates that the dioxygenase active site consists of three residues (H1382xD1384…H1881) binding to one side of the iron(II) coordination sphere. In addition, iron (II) binds to two exogenous ligands (i.e., 2OG and O2). Two iron (II) coordinates are occupied by the carboxyl and keto oxygen atoms of 2OG, whereas oxygen binds to the last available iron (II) coordinate opposite to the TET2 residues constituting the 2-His-1-carboxylate triad. Thus, TET2 uses iron (II) and atmospheric oxygen as essential cofactors and 2OG as a cosubstrate[72].

A few recent studies have also uncovered extensive C5 cytosine methylation (5mC) in non-CG sites (mCH, where H = A, T, or C), reaching levels similar to that of mCG in the neuronal genome[54,55]. TET2 initiates the 5mC oxidation cascade in both CpG and non-CpG island demethylation[133]. TET2 mutations affect non-CpG island DNA methylation of enhancers and other regulatory elements, indicating that TET2 also plays a role in non-CpG island DNA methylation[148]. In a study on chronic myelomonocytic leukemia (CMML), TET2 mutations were found to affect 5mC methylation marks in non-CpG islands. The distribution of TET2-specific differentially methylated CpGs and non-CpG islands was determined in a study on chronic myelomonocytic leukemia (CMML). This study analyzed the methylomes of TET2 wild-type (TET2-WT) and mutant (TET2-MT) cases of CMML and identified 409 TET2-specific differentially methylated 5mC sites located within non-CpG islands, 86% of which were hypermethylated in TET2-MT cases[149]. TET2, a dioxygenase tumor suppressor, is one of the most frequently mutated genes in myelodysplastic syndrome (MDS). A high degree of TET2 mutations has been observed in several myeloid malignancies, such as myeloproliferative neoplasms (MDS-MPNs) and acute myeloid leukemia derived from MDS and MDS-MPN (sAML)[41,68,79]. Patients with TET2 mutations show low levels of genomic 5hmC in the marrow compared with those with wild-type TET2. Clinically described TET2 mutations include frameshift, nonsense, and missense mutations. Studies using a TET2-knockout mouse model have demonstrated that the progression of myeloid malignancies depends on TET2 activity. Restoration of TET2 activity reverses aberrant hematopoietic stem and progenitor cell proliferation in vitro and in vivo and in leukemia[47,69,70,150].

TET2 is located on chromosome 4q24 and consists of 11 exons. The open reading frame consists of 6009 nucleotides and the protein encoded by this gene comprises 2002 amino acid residues[151]. Crystal structure of TET2 catalytic domain shows that two zinc fingers bring the cysteine-rich and double-stranded β-helix fold domains together to form a compact catalytic domain. The cysteine-rich domain stabilizes the DNA above the double-stranded β-helix fold domain and 5mC is inserted into the catalytic cavity with the methyl group orientated towards the catalytic Fe (II) for oxidation. The methyl group is not involved in TET2-DNA contacts so that the catalytic cavity allows TET2 to accommodate 5mC derivatives for further oxidation[72]. The cysteine-rich domain is highly conserved among all three TET family members and is involved in substrate oxidation. In the past ten years, a few hundred TET2 mutations have been documented in scientific literature[151]. Most of the frameshift or nonsense mutations are concentrated within the dioxygenase domain, rendering the protein inactive. However, limited information is available regarding the structural and functional consequences of these TET2 missense variants. TET2-DNA crystal structure was solved in 2013 and this study identified twelve amino acid residues in the TET2 active site which interact with the substrate DNA (Figure 42)[72]. The active site of 2OG-dependent dioxygenase TET2 is flexible enough to accommodate modified nucleotide bases. Here, we mutated these amino acid residues into alanine, which may in turn perturb their interaction with DNA to produce a differential oxidation pattern for CpG and non-CpG substrates. We characterized these mutants based on their efficiency in 5mC oxidation in CpG and non-CpG contexts. As expected, mutating the active site residues to alanine altered hTET2 activity. However, this effect was more prominent for CpG substrates. Notably, three active-site mutations, Y1295A, R1302A,

and H1904A, demonstrated significantly higher oxidation efficiencies on CpG substrates than the wild-type enzyme.

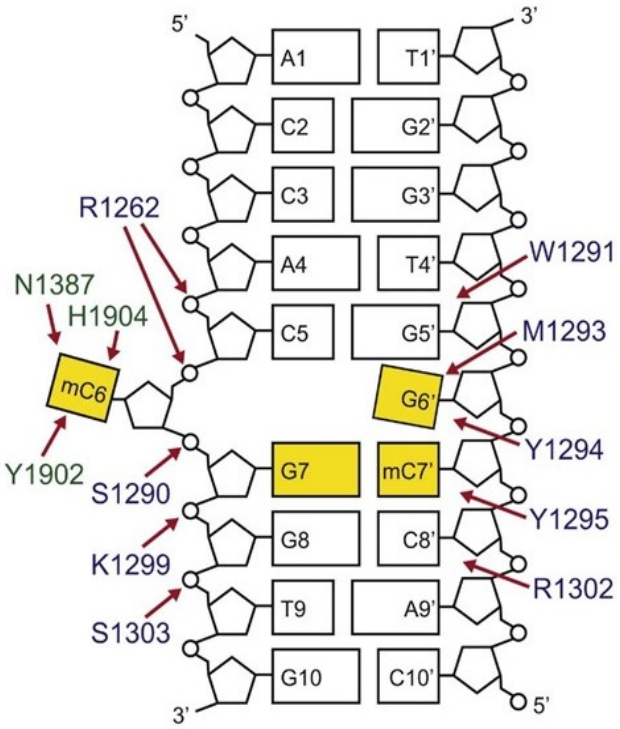


Figure 42: Interaction between TET2 active site residues and the substrate DNA within the catalytic cavity of the enzyme[72].

6.2 Materials and methods

Chemicals and Reagents: All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Escherichia coli (E. coli) expression strain BL21 (DE3) was purchased from Novagen (now EMD Millipore, Billerica, MA). Chromatography solvents such as water and methanol were purchased from Fisher scientific (Waltham, MA). All growth media were from Difco Laboratories (Detroit, MI). SP Sepharose fast flow was purchased from GE Health Care (Chicago, IL). 3-(*N*-morpholino) propanesulfonic acid was purchased from Carbosynth US LLC (San Diego, CA). DNA oligomers were purchased from Integrated DNA Technologies (San Diego, CA).

Cloning, Expression, and Purification of WT TET2 and mutants: The C-terminal TET2 dioxygenase domain (TET2 1129-1936, Δ1481-1843, the minimal catalytically active domain), was cloned into the pDEST14 vector using the site-specific recombination technique as described previously[73]. In protein engineering, site-directed mutagenesis is used to generate DNA sequences with mutated codons, insertions, or deletions. In this widely used method, mutations are generated by PCR using a pair of oligonucleotide primers designed with mismatching nucleotides at the center of the primers. TET2 active site residues R1262, N1387, H1904, Y1902, S1290, K1299, S1303, W1291, M1293, Y1294, Y1295, R 1302 were chosen for site directed mutagenesis. WT TET2 enzyme and its mutants were purified by cation exchange chromatography on SP Sepharose, as described in our previous article (Figure 42)[122].

Table 4: Primer DNA sequences used in site directed mutagenesis.

List of mutants	Primer DNA sequence
TET2-R1262A-fwd	5’CTCACCAATCGCGCGTGTGCCTTGAATGAAGAGAGAACT3’
TET2-R1262A-rev	5’CAAGGCACACGCGCGATTGGTGAGCGTGCCGTATTTCCT3’
TET2-N1387A-fwd	5’AGAGACTTGCACGCCATGCAGAATGGCAGCACATTGGTA3’
TET2-N1387A-rev	5’ATTCTGCATGGCGTGCAAGTCTCTGTGGGCATGAGCACA3’
TET2-H1904A-fwd	5’GTCTTTTACCAGGCTAAGAGCATGAATGAGCCAAAACAT3’
TET2-H1904A-rev	5’CATGCTCTTAGCCTGGTAAAAGACGAGGGAGATCCTGGT3’
TET2-Y1902A-fwd	5’TCCCTCGTCTTTGCCCAGCATAAGAGCATGAATGAGCCA3'
TET2-Y1902A-rev	5’CTTATGCTGGGCAAAGACGAGGGAGATCCTGGTGGGGTG3’
TET2-S1290A-fwd	5’TCTTTTGGTTGTGCATGGAGCATGTACTACAATGGATGT3’
TET2-S1290A-rev	5’CATGCTCCATGCACAACCAAAAGAGAAGGAGGCACCACA3’
TET2-K1299A-fwd	5’TACAATGGATGTGCGTTTGCCAGAAGCAAGATCCCAAGG3’
TET2-K1299A-rev	5’TCTGGCAAACGCACATCCATTGTAGTACATGCTCCATGA3’
TET2-S1303A-fwd	5’AAGTTTGCCAGAGCCAAGATCCCAAGGAAGTTTAAGCTG3’
TET2-S1303A-rev	5’TGGGATCTTGGCTCTGGCAAACTTACATCCATTGTAGTA3’
TET2-W1291A-fwd	5’TTTGGTTGTTCAGCGAGCATGTACTACAATGGATGTAAG3’
TET2-W1291A-rev	5’GTACATGCTCGCTGAACAACCAAAAGAGAAGGAGGCACC3’
TET2-M1293A-fwd	5’TGTTCATGGAGCGCGTACTACAATGGATGTAAGTTTGCC3’
TET2-M1293A-rev	5’CCATTGTAGTACGCGCTCCATGAACAACCAAAAGAGAAG3’
TET2-Y1294A-fwd	5’TCATGGAGCATGGCCTACAATGGATGTAAGTTTGCCAGA3’
TET2-Y1294A-rev	5’CATCCATTGTAGGCCATGCTCCATGAACAACCAAAAGAG3’
TET2-Y1295A-fwd	5’TCATGGAGCATGTACTACAATGGATGTAAGTTTGCCAGA3’

TET2-Y1295A-rev	5’CATCCATTGCCGTACATGCTCCATGAACAACCAAAAGAG3’
TET2-R1302A-fwd	5’TGTAAGTTTGCCGCAAGCAAGATCCCAAGGAAGTTTAAG3’
TET2-R1302A-rev	5’GGGATCTTGCTTGCGGCAAACTTACATCCATTGTAGTAC3’

***In vitro* TET2 Reaction:** A 25-mer double stranded DNA used as TET2 substrate was generated by hybridizing single stranded oligomers on a thermal cycler. The *in vitro* TET2 enzymatic reactions were performed using a reaction matrix containing 50 mM HEPES (pH 8.0), 75 μM FeSO4, 1 mM 2-oxoglutarate, 2 mM ascorbate, 1 mM ATP and 1 mM DTT in 100 μl of total reaction volume. Equal amount of DNA substrate was added according to designated concentration and the reaction was started by addition of 2.5 μg purified TET2. After 2 hours of incubation at 37 °C, TET2 catalyzed oxidation reactions were quenched with 5 μl of 500 mM EDTA. Total dsDNA from the reaction mixture was separated using the superior oligo purification kit following manufacturer's protocol (Superior Scientifics, Lenexa, KS). The isolated DNA (20 μl) was digested with 2 units of DNase I and 60 units of S1 nuclease in 40 μl at 37 °C for 12 hours to produce individual nucleotide monophosphates. Following the digestion, 2 units of calf intestinal alkaline phosphatase (CIAP) was added in the samples followed by incubation for an additional 4 hours at 37 °C to remove the terminal phosphate groups to obtain nucleosides.

Table 5: Sequence of DNA substrates.

DNA Substrates	Sequence
5mCpA Substrate	Forward strand:5'-GCGCCGGTCCTGmCACCCGCTCCCGC-3' Reverse strand:5'-GCGGGAGCGGGTGCAGGACCGGCGC-3'
5mCpG Substrate	Forward strand:5'-GCGCCGGTCCTGmCGCCCGCTCCCGC-3' Reverse strand:5'-GCGGGAGCGGGCGCAGGACCGGCGC-3'

The amount of oxidation products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g., 5hmC) and dividing it by the area represented by one deoxycytidine

residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using nucleoside standards. The activity of mutant TET2 enzymes were quantified the same way and the amount of total oxidation product (picomoles) was presented in comparison with the WT TET2 activity. Standard deviations and standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

Liquid Chromatography of Nucleosides under different MS/MS modes: The LC-MS/MS analysis was performed on Sciex 3200 QTrap mass spectrometer (Foster City, CA) coupled to a Shimadzu UFLC LC-20 system (Columbia, MD) using electrospray ionization source (ESI) and run with Analyst v 1.6.2 software. For chromatographic separation of nucleosides in ion-switching mode, water/methanol based solvent system was used where solvent A was water (adjusted to pH 3.5 using formic Acid) and solvent B was methanol (adjusted to pH 3.0 using formic Acid). The gradient used was 0% B (0-1 minute), 0-2% B (1-12 minute), 2-30% B (12-17 minute), 30% B (17-18 minute), 30-0% B (18-18.5 minute), followed by a 4.5-minute equilibration at 0% B at a flow rate of 0.3 mL/minute. Nucleosides were separated on C18 column (Dimension 150×2 mm, particle size 5μM, Pore Size 100 Å).

6.3 Results and discussion

TET2 is a tumor suppressor gene that encodes a 2-oxoglutarate/Fe (II)-dependent dioxygenase that catalyzes iterative oxidation of 5mC marks in eukaryotic DNA. Mutations in TET2 are associated with various myeloid malignancies, hematological disorders, and myeloproliferative disorders[144]. Full-length TET2 comprises an N-terminal domain and a C-terminal catalytic domain consisting of a Cys-rich region and a double-stranded beta-helix (DSBH) fold. The function of the N-terminal domain of TET2 is not well understood,

but it has been shown to be heavily post-translationally modified, pointing to the important regulatory functions of this region in TET2 regulation and function in DNA. The crystal structure of TET2 bound to methylated DNA showed that the N-terminal domain of TET2 is not involved in DNA binding or catalytic activity[72]. The Cys-rich domain of TET2 is a part of the C-terminal catalytic domain, which consists of a double-stranded beta-helix (DSBH) fold and is responsible for substrate oxidation. The Cys-rich domain stabilizes the DNA above the DSBH core, and two zinc fingers bring the Cys-rich and DSBH domains together to form a compact catalytic domain. The TET2-DNA complex is further stabilized by two loops L1 and L2 which are parts of the cys-rich domain. The two loops are highly conserved among members of the TET subfamily[72].

In this study, we chose DNA-interacting residues in the TET2 active site and performed alanine scanning on these residues to characterize differential 5mC oxidation by these mutants on CpG and non-CpG substrates. Residue R1262 was from loop L1, and all other DNA-interacting residues (N1387, H1904, Y1902, S1290, K1299, S1303, W1291, M1293, Y1294, Y1295, R 1302) were located in loop L2. All these active site residues were mutated into alanine by site directed mutagenesis and mutant proteins were purified by cation exchange chromatography as described in the 'Materials and methods' section. Using a sensitive LC-MS/MS technique, we characterized the 5mC oxidation pattern of wild-type and mutant TET2 enzymes on 25-base pair double-stranded hemimethylated DNA substrates. The substrate DNA was designed such that the 5mC site resides in the middle (Table 5). Following enzymatic reactions with TET2, DNA was isolated and transformed into nucleosides, which were then subjected to mass spectrometry.

Point mutations W1291A and S1303A significantly lowered the 5mC oxidation of CpG substrates. However, mutations S1290A, M1293A, K1299A, R1262A, Y1294A, N1387A, and Y1902A had little or no effect on the catalytic activity. Because TET2 binds to DNA through extensive hydrophobic interactions and hydrogen bonds, single- or double-point mutations may not be sufficient to abrogate TET2-DNA interactions. Interestingly, three active-site mutations, Y1295A, R1302A, and H1904A, demonstrated significantly higher oxidation efficiencies on CpG substrates than the wild-type enzyme. This enhancement in the catalytic activity of these mutant enzymes is directly correlated with the formation of 5fC and 5caC, which are higher-order oxidation products that may lead to complete demethylation of the substrate DNA (Figure 43).

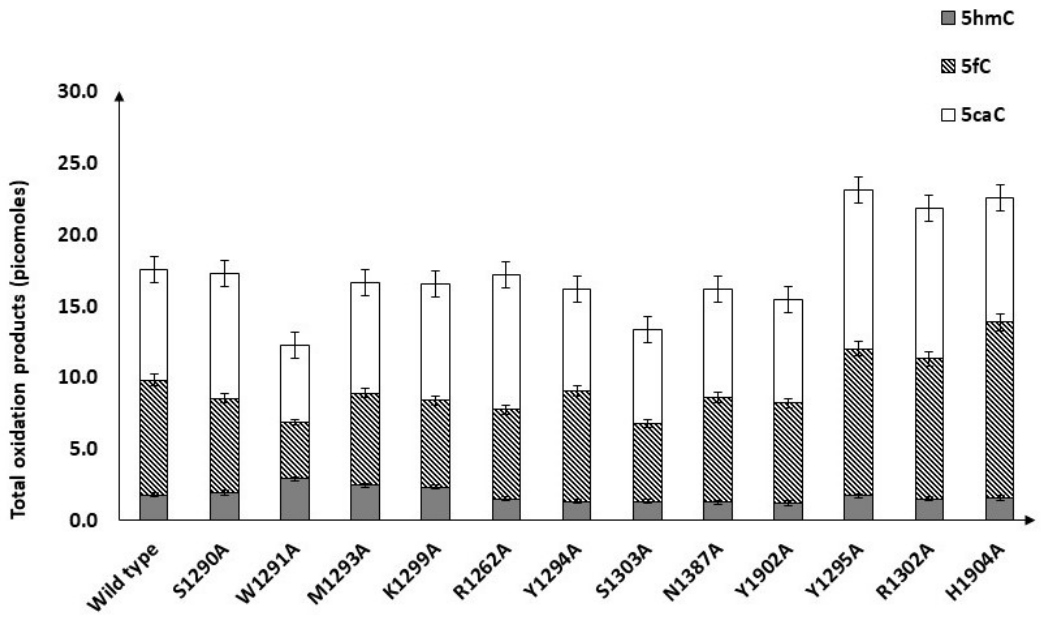


Figure 43: Effect of TET2 active site mutations on CpG substrate. "Y" axis represents total oxidation products in picomoles. The 'Y' axes represent total oxidation events. The amounts of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 10). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

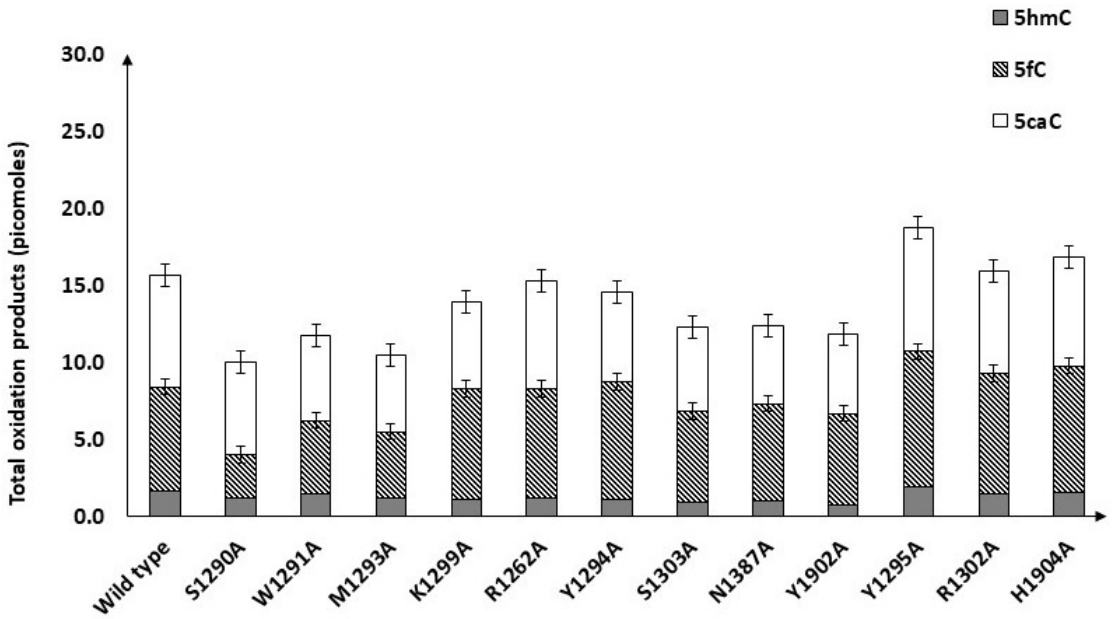


Figure 44: Effect of TET2 active site mutations on CpA substrate. "Y" axis represents total oxidation products in picomoles. The 'Y' axes represent total oxidation events. The amounts of products formed during hTET2-mediated oxidation reactions across different DNA substrates were normalized by calculating the peak area of each product (e.g. 5hmC) and dividing it by the area represented by one deoxycytidine residue (Δ5hmC/ΔC/#C). The amount of each oxidative product (picomoles) was calculated using the standard curve (Figure 10). As a result, TOE was calculated by [(1 × number of 5hmC molecules) + (2 × number of 5fC molecules) + (3 × number of 5caC molecules)]. Standard errors were calculated for each oxidative product from triplicate experiments and are represented in the figures.

W. T. TET2 demonstrated a similar preference towards CpG and CpA substrates and produced slightly lower amounts of oxidation products for the CpA substrate. The TET2 mutants demonstrated an overall reduction in 5mC oxidation for CpA substrates. The catalytic activity of the mutants S1290A and M1293A was significantly lower (almost by 2 folds) in the presence of CpA substrates. Mutations in the residues K1299, R1262, and Y1294 had no significant effect on 5mC oxidation, irrespective of CpG or non-CpG substrates. TET2-Y1295A demonstrated a slight enhancement in catalytic activity compared to that of the wild-type enzyme, which was consistent with both CpG and CpA substrates (Figure 44). In conclusion, alanine scanning of DNA-interacting residues demonstrated that the catalytic cavity of TET2 is flexible enough to accommodate varying sequences of DNA substrates, and the stability of the TET2-DNA complex is collectively supported by all these amino acid residues. Thus, the effect of a single point mutation is compensated for without rendering the enzyme catalytically inactive. However, these findings need to be supported by in vivo studies, as well as molecular dynamics simulations.

CHAPTER 7

CONCLUSIONS

Here, we discuss the role of TET family dioxygenases, particularly TET2, in epigenetic transcriptional regulation and their involvement in hematopoietic malignancies. To overcome the challenges of quantifying and comparing mutant TET2 activity, we developed a novel liquid chromatography method integrated with tandem mass spectrometry to quantify modified cytosine bases by TET2 mediated 5mC oxidation. Our study also demonstrated that TET2 can initiate the oxidation of 5mC not only at CpG sites, but also at non-CpG sites, which is a significant breakthrough compared to previous observations. TET2 can significantly oxidize 5mC at CpA sites, although the amount of 5caC formed was less than that in the case of 5mCpG. Additionally, we showed that the TET2 catalytic domain can oxidize 5mC at the CpC and CpT sites, suggesting that TET2 can initiate the oxidation cascade at the 5mCpH sites, leading to demethylation by the TDG/BER pathway. By investigating the substrate preferences of the TET isoforms, we found that mTET1 and hTET2 had a higher preference for 13-mer dsDNA substrates. We observed that altering the length of the dsDNA substrates reduced product formation, whereas altering the length of the ssDNA substrates did not result in a consistent pattern of 5mC oxidation.

The crystal structure of the TET2-DNA complex revealed the importance of the residue R1896 in stabilizing the 5-carboxylate moiety of 2OG. Mutations in this residue are frequently observed in myeloid cancers, leading to impaired 2OG binding and a loss of TET2 activity. We investigated the effect of 2OG analogs on rescuing TET2 R1896 mutations and found that certain analogs enhanced the catalytic activity of TET2 mutants.

We identified specific 2OG analogs that efficiently rescued TET2 R1896S activity, suggesting their potential use in modulating mutant TET2 activity under diseased conditions. To characterize the differential 5mC oxidation by TET2 mutants on CpG and non CpG substrates, alanine scanning was performed on the DNA-interacting residues in the TET2 active site. We found that certain mutations significantly lowered 5mC oxidation of CpG substrates, whereas others had little or no effect. Notably, the mutations Y1295A, R1302A, and H1904A demonstrated higher oxidation efficiencies on CpG substrates, leading to the formation of higher-order oxidation products.

Overall, our study provides insight into the function and activity of TET2 and its mutants in DNA demethylation, substrate preference, and the potential use of 2OG analogs to modulate mutant TET2 activity. These findings contribute to our understanding of epigenetic regulatory processes and have implications for cancer therapeutics and assay development involving the TET family of dioxygenases. However, further in vivo studies and molecular dynamics simulations are required to support these findings.

REFERENCES

(1) Denizio, J. E.; Dow, B. J.; Serrano, J. C.; Ghanty, U.; Drohat, A. C.; Kohli, R. M. TET-TDG Active DNA Demethylation at CpG and Non-CpG Sites HHS Public Access. *J Mol Biol* **2021**, *433* (8), 166877. https://doi.org/10.1016/j.jmb.2021.166877.

(2) Losman, J. A.; Koivunen, P.; Kaelin, W. G. 2-Oxoglutarate-Dependent Dioxygenases in Cancer. *Nat Rev Cancer* **2020**, *20* (12), 710–726. https://doi.org/10.1038/S41568-020-00303-3.

(3) Hewitson, K. S.; Granatino, N.; Welford, R. W. D.; McDonough, M. A.; Schofield, C. J. Oxidation by 2-Oxoglutarate Oxygenases: Non-Haem Iron Systems in Catalysis and Signalling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2005**, *363* (1829), 807–828. https://doi.org/10.1098/RSTA.2004.1540.

(4) Kawai, Y.; Ono, E.; Mizutani, M. Evolution and Diversity of the 2–Oxoglutarate-Dependent Dioxygenase Superfamily in Plants. *The Plant Journal* **2014**, *78* (2), 328–343. https://doi.org/10.1111/TPJ.12479.

(5) Crake, R. L. I.; Burgess, E. R.; Royds, J. A.; Phillips, E.; Vissers, M. C. M.; Dachs, G. U. The Role of 2-Oxoglutarate Dependent Dioxygenases in Gliomas and Glioblastomas: A Review of Epigenetic Reprogramming and Hypoxic Response. *Front Oncol* **2021**, *11*, 313. https://doi.org/10.3389/FONC.2021.619300/BIBTEX.

(6) Vissers, M. C. M.; Kuiper, C.; Dachs, G. U. Regulation of the 2-Oxoglutarate-Dependent Dioxygenases and Implications for Cancer. *Biochem Soc Trans* **2014**, *42* (4), 945–951. https://doi.org/10.1042/BST20140118.

(7) Kaelin, W. G. Cancer and Altered Metabolism: Potential Importance of Hypoxia-Inducible Factor and 2-Oxoglutarate-Dependent Dioxygenases. *Cold Spring Harb Symp Quant Biol* **2011**, *76*, 335–345. https://doi.org/10.1101/SQB.2011.76.010975.

(8) Dong, C.; Zhang, H.; Xu, C.; Arrowsmith, C. H.; Min, J. Structure and Function of Dioxygenases in Histone Demethylation and DNA/RNA Demethylation. *urn:issn:2052-2525* **2014**, *1* (6), 540–549. https://doi.org/10.1107/S2052252514020922.

(9) Salminen, A.; Kauppinen, A.; Kaarniranta, K. 2-Oxoglutarate-Dependent Dioxygenases Are Sensors of Energy Metabolism, Oxygen Availability, and Iron Homeostasis: Potential Role in the Regulation of Aging Process. *Cellular and Molecular Life Sciences* **2015**, *72* (20), 3897–3914. https://doi.org/10.1007/s00018-015-1978-z.

(10) McDonough, M. A.; Loenarz, C.; Chowdhury, R.; Clifton, I. J.; Schofield, C. J. Structural Studies on Human 2-Oxoglutarate Dependent Oxygenases. *Curr Opin Struct Biol* **2010**, *20* (6), 659–672. https://doi.org/10.1016/j.sbi.2010.08.006.

(11) Schofield, C. J.; Zhang, Z. Structural and Mechanistic Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes. *Curr Opin Struct Biol* **1999**, *9* (6), 722–731. https://doi.org/10.1016/S0959-440X(99)00036-6.

(12) Aik, W. S.; Chowdhury, R.; Clifton, I. J.; Hopkinson, R. J.; Leissing, T.; McDonough, M. A.; Nowak, R.; Schofield, C. J.; Walport, L. J. Introduction to Structural Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes. *RSC Metallobiology* **2015**, *2015-January* (3), 59–94. https://doi.org/10.1039/9781782621959-00059.

(13) Hewitson, K. S.; Granatino, N.; Welford, R. W. D.; McDonough, M. A.; Schofield, C. J. Oxidation by 2-Oxoglutarate Oxygenases: Non-Haem Iron Systems in Catalysis and Signalling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2005**, *363* (1829), 807–828. https://doi.org/10.1098/rsta.2004.1540.

(14) Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J. Structural Studies on 2-Oxoglutarate Oxygenases and Related Double-Stranded Beta-Helix Fold Proteins. *J Inorg Biochem* **2006**, *100* (4), 644–669. https://doi.org/10.1016/J.JINORGBIO.2006.01.024.

(15) Melamed, P.; Yosefzon, Y.; David, C.; Tsukerman, A.; Pnueli, L. Tet Enzymes, Variants, and Differential Effects on Function. *Front Cell Dev Biol* **2018**, *6* (MAR), 22. https://doi.org/10.3389/FCELL.2018.00022/BIBTEX.

(16) Parker, M. J.; Weigele, P. R.; Saleh, L. Insights into the Biochemistry, Evolution, and Biotechnological Applications of the Ten-Eleven Translocation (TET) Enzymes. *Biochemistry* **2019**, *58* (6), 450–467. https://doi.org/10.1021/ACS.BIOCHEM.8B01185/ASSET/IMAGES/LARGE/BI-2018-01185U_0008.JPEG.

(17) Gerecke, C.; Egea Rodrigues, C.; Homann, T.; Kleuser, B. The Role of Ten-Eleven Translocation Proteins in Inflammation. *Front Immunol* **2022**, *13*, 1143. https://doi.org/10.3389/FIMMU.2022.861351/BIBTEX.

(18) Nakajima, H.; Kunimoto, H. TET2 as an Epigenetic Master Regulator for Normal and Malignant Hematopoiesis. *Cancer Sci* **2014**, *105* (9), 1093–1099. https://doi.org/10.1111/CAS.12484.

(19) Tan, L.; Shi, Y. G. Tet Family Proteins and 5-Hydroxymethylcytosine in Development and Disease. *Development* **2012**, *139* (11), 1895–1902. https://doi.org/10.1242/DEV.070771.

(20) Williams, K.; Christensen, J.; Pedersen, M. T.; Johansen, J. v; Cloos, P. A. C.; Rappsilber, J.; Helin, K. TET1 and Hydroxymethylcytosine in Transcription and DNA Methylation Fidelity. *Nature* **2011**, *473* (7347), 343–348. https://doi.org/10.1038/nature10066.

(21) An, J.; Rao, A.; Ko, M. TET Family Dioxygenases and DNA Demethylation in Stem Cells and Cancers. *Exp Mol Med* **2017**, *49* (4), e323. https://doi.org/10.1038/emm.2017.5.

(22) Robertson, K. D. DNA Methylation and Human Disease. *Nat Rev Genet* **2005**, *6* (8), 597–610. https://doi.org/10.1038/NRG1655.

(23) Yin, Y.; Morgunova, E.; Jolma, A.; Kaasinen, E.; Sahu, B.; Khund-Sayeed, S.; Das, P. K.; Kivioja, T.; Dave, K.; Zhong, F.; Nitta, K. R.; Taipale, M.; Popov, A.; Ginno, P. A.; Domcke, S.; Yan, J.; Schübeler, D.; Vinson, C.; Taipale, J. Impact of Cytosine Methylation on DNA Binding Specificities of Human Transcription Factors. *Science (1979)* **2017**, *356* (6337). https://doi.org/10.1126/SCIENCE.AAJ2239/SUPPL_FILE/AAJ2239_YIN_SM_TABLES_S1-S6.XLSX.

(24) Jin, J.; Lian, T.; Gu, C.; Yu, K.; Gao, Y. Q.; Su, X. D. The Effects of Cytosine Methylation on General Transcription Factors. *Sci Rep* **2016**, *6*. https://doi.org/10.1038/SREP29119.

(25) Medvedeva, Y. A.; Khamis, A. M.; Kulakovskiy, I. V.; Ba-Alawi, W.; Bhuyan, M. S. I.; Kawaji, H.; Lassmann, T.; Harbers, M.; Forrest, A. R. R.; Bajic, V. B. Effects of Cytosine Methylation on Transcription Factor Binding Sites. *BMC Genomics* **2014**, *15* (1), 1–12. https://doi.org/10.1186/1471-2164-15-119/FIGURES/3.

(26) Hognon, C.; Besancenot, V.; Gruez, A.; Grandemange, S.; Monari, A. Cooperative Effects of Cytosine Methylation on DNA Structure and Dynamics. *Journal of Physical Chemistry B* **2019**, *123* (34), 7365–7371. https://doi.org/10.1021/ACS.JPCB.9B05835/ASSET/IMAGES/LARGE/JP9B05835_0007.JPEG.

(27) Li, S.; Peng, Y.; Landsman, D.; Panchenko, A. R. DNA Methylation Cues in Nucleosome Geometry, Stability and Unwrapping. *Nucleic Acids Res* **2022**, *50* (4), 1864–1874. https://doi.org/10.1093/NAR/GKAC097.

(28) Portella, G.; Battistini, F.; Orozco, M. Understanding the Connection between Epigenetic DNA Methylation and Nucleosome Positioning from Computer Simulations. *PLoS Comput Biol* **2013**, *9* (11), e1003354. https://doi.org/10.1371/JOURNAL.PCBI.1003354.

(29) Machado, A. C. D.; Zhou, T.; Rao, S.; Goel, P.; Rastogi, C.; Lazarovici, A.; Bussemaker, H. J.; Rohs, R. Evolving Insights on How Cytosine Methylation Affects Protein–DNA Binding. *Brief Funct Genomics* **2015**, *14* (1), 61. https://doi.org/10.1093/BFGP/ELU040.

(30) Costello, J. F.; Plass, C. Methylation Matters. *J Med Genet* **2001**, *38* (5), 285–303. https://doi.org/10.1136/JMG.38.5.285.

(31) Chen, C.; Wang, Z.; Ding, Y.; Wang, L.; Wang, S.; Wang, H.; Qin, Y. DNA Methylation: From Cancer Biology to Clinical Perspectives. *Frontiers in Bioscience - Landmark* **2022**, *27* (12), 326. https://doi.org/10.31083/J.FBL2712326/2768-6698-27-12-326/FIG5.JPG.

(32) Wu, H.; D'Alessio, A. C.; Ito, S.; Xia, K.; Wang, Z.; Cui, K.; Zhao, K.; Eve Sun, Y.; Zhang, Y. Dual Functions of Tet1 in Transcriptional Regulation in Mouse Embryonic Stem Cells. *Nature* **2011**, *473* (7347), 389–394. https://doi.org/10.1038/NATURE09934.

(33) Hon, G. C.; Song, C. X.; Du, T.; Jin, F.; Selvaraj, S.; Lee, A. Y.; Yen, C. A.; Ye, Z.; Mao, S. Q.; Wang, B. A.; Kuan, S.; Edsall, L. E.; Zhao, B. S.; Xu, G. L.; He, C.; Ren, B. 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. *Mol Cell* **2014**, *56* (2), 286–297. https://doi.org/10.1016/J.MOLCEL.2014.08.026.

(34) Quivoron, C.; Couronné, L.; Della Valle, V.; Lopez, C. K.; Plo, I.; Wagner-Ballon, O.; Do Cruzeiro, M.; Delhommeau, F.; Arnulf, B.; Stern, M. H.; Godley, L.; Opolon, P.; Tilly, H.; Solary, E.; Duffourd, Y.; Dessen, P.; Merle-Beral, H.; Nguyen-Khac, F.; Fontenay, M.; Vainchenker, W.; Bastard, C.; Mercher, T.; Bernard, O. A. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* **2011**, *20* (1), 25–38. https://doi.org/10.1016/J.CCR.2011.06.003.

(35) Thomson, J. P.; Ottaviano, R.; Unterberger, E. B.; Lempiainen, H.; Muller, A.; Terranova, R.; Illingworth, R. S.; Webb, S.; Kerr, A. R. W.; Lyall, M. J.; Drake, A. J.; Rolandwolf, C.; Moggs, J. G.; Schwarz, M.; Meehan, R. R. Loss of Tet1-

Associated 5-Hydroxymethylcytosine Is Concomitant with Aberrant Promoter Hypermethylation in Liver Cancer. *Cancer Res* **2016**, *76* (10), 3097–3108. https://doi.org/10.1158/0008-5472.CAN-15-1910.

(36) Cimmino, L.; Dawlaty, M. M.; Ndiaye-Lobry, D.; Yap, Y. S.; Bakogianni, S.; Yu, Y.; Bhattacharyya, S.; Shaknovich, R.; Geng, H.; Lobry, C.; Mullenders, J.; King, B.; Trimarchi, T.; Aranda-Orgilles, B.; Liu, C.; Shen, S.; Verma, A. K.; Jaenisch, R.; Aifantis, I. TET1 Is a Tumor Suppressor of Hematopoietic Malignancy. *Nat Immunol* **2015**, *16* (6), 653–662. https://doi.org/10.1038/NI.3148.

(37) Delhommeau, F.; Dupont, S.; Valle, V. Della; James, C.; Trannoy, S.; Massé, A.; Kosmider, O.; Le Couedic, J.-P.; Robert, F.; Alberdi, A.; Lécluse, Y.; Plo, I.; Dreyfus, F. J.; Marzac, C.; Casadevall, N.; Lacombe, C.; Romana, S. P.; Dessen, P.; Soulier, J.; Viguié, F.; Fontenay, M.; Vainchenker, W.; Bernard, O. A. Mutation in TET2 in Myeloid Cancers . *New England Journal of Medicine* **2009**, *360* (22), 2289–2301. https://doi.org/10.1056/nejmoa0810069.

(38) *Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies - PubMed*. https://pubmed.ncbi.nlm.nih.gov/19420352/ (accessed 2023-06-05).

(39) Bejar, R.; Lord, A.; Stevenson, K.; Bar-Natan, M.; Pérez-Ladaga, A.; Zaneveld, J.; Wang, H.; Caughey, B.; Stojanov, P.; Getz, G.; Garcia-Manero, G.; Kantarjian, H.; Chen, R.; Stone, R. M.; Neuberg, D.; Steensma, D. P.; Ebert, B. L. TET2 Mutations Predict Response to Hypomethylating Agents in Myelodysplastic Syndrome Patients. *Blood* **2014**, *124* (17), 2705–2712. https://doi.org/10.1182/BLOOD-2014-06-582809.

(40) *TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs) - PubMed*. https://pubmed.ncbi.nlm.nih.gov/19666869/ (accessed 2023-06-05).

(41) Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; MacIejewski, J. P.; Rao, A. Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2. *Nature* **2010**, *468* (7325), 839–843. https://doi.org/10.1038/NATURE09586.

(42) Beerman, I.; Rossi, D. J. Epigenetic Control of Stem Cell Potential During Homeostasis, Aging, and Disease. *Cell Stem Cell* **2015**, *16* (6), 613. https://doi.org/10.1016/J.STEM.2015.05.009.

(43) Huang, H.; Jiang, X.; Li, Z.; Li, Y.; Song, C. X.; He, C.; Sun, M.; Chen, P.; Gurbuxani, S.; Wang, J.; Hong, G. M.; Elkahloun, A. G.; Arnovitz, S.; Wang, J.; Szulwach, K.; Lin, L.; Street, C.; Wunderlich, M.; Dawlaty, M.; Neilly, M. B.; Jaenisch, R.; Yang, F. C.; Mulloy, J. C.; Jin, P.; Liu, P. P.; Rowley, J. D.; Xu, M.; He, C.; Chen, J. TET1 Plays an Essential Oncogenic Role in MLL-Rearranged Leukemia. *Proc Natl Acad Sci U S A* **2013**, *110* (29), 11994–11999. https://doi.org/10.1073/PNAS.1310656110/-/DCSUPPLEMENTAL/PNAS.201310656SI.PDF.

(44) Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; MacIejewski, J. P.; Rao, A. Impaired Hydroxylation of 5-

Methylcytosine in Myeloid Cancers with Mutant TET2. *Nature* **2010**, *468* (7325), 839. https://doi.org/10.1038/NATURE09586.

(45) Moran-Crusio, K.; Reavie, L.; Shih, A.; Abdel-Wahab, O.; Ndiaye-Lobry, D.; Lobry, C.; Figueroa, M. E.; Vasanthakumar, A.; Patel, J.; Zhao, X.; Perna, F.; Pandey, S.; Madzo, J.; Song, C.; Dai, Q.; He, C.; Ibrahim, S.; Beran, M.; Zavadil, J.; Nimer, S. D.; Melnick, A.; Godley, L. A.; Aifantis, I.; Levine, R. L. Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell* **2011**, *20* (1), 11–24. https://doi.org/10.1016/J.CCR.2011.06.001.

(46) Ko, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Thompson, E. C.; Hastie, R.; Tsangaratou, A.; Rajewsky, K.; Koralov, S. B.; Rao, A. Ten-Eleven-Translocation 2 (TET2) Negatively Regulates Homeostasis and Differentiation of Hematopoietic Stem Cells in Mice. *Proc Natl Acad Sci U S A* **2011**, *108* (35), 14566–14571. https://doi.org/10.1073/PNAS.1112317108/-/DCSUPPLEMENTAL/PNAS.201112317SI.PDF.

(47) Ko, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Thompson, E. C.; Hastie, R.; Tsangaratou, A.; Rajewsky, K.; Koralov, S. B.; Rao, A. Ten-Eleven-Translocation 2 (TET2) Negatively Regulates Homeostasis and Differentiation of Hematopoietic Stem Cells in Mice. *Proc Natl Acad Sci U S A* **2011**, *108* (35), 14566–14571. https://doi.org/10.1073/PNAS.1112317108/-/DCSUPPLEMENTAL/PNAS.201112317SI.PDF.

(48) Huang, H.; Jiang, X.; Li, Z.; Li, Y.; Song, C. X.; He, C.; Sun, M.; Chen, P.; Gurbuxani, S.; Wang, J.; Hong, G. M.; Elkahloun, A. G.; Arnovitz, S.; Wang, J.; Szulwach, K.; Lin, L.; Street, C.; Wunderlich, M.; Dawlaty, M.; Neilly, M. B.;

Jaenisch, R.; Yang, F. C.; Mulloy, J. C.; Jin, P.; Liu, P. P.; Rowley, J. D.; Xu, M.; He, C.; Chen, J. TET1 Plays an Essential Oncogenic Role in MLL-Rearranged Leukemia. *Proc Natl Acad Sci U S A* **2013**, *110* (29), 11994–11999. https://doi.org/10.1073/PNAS.1310656110/-/DCSUPPLEMENTAL/PNAS.201310656SI.PDF.

(49) Lio, C. W.; Zhang, J.; González-Avalos, E.; Hogan, P. G.; Chang, X.; Rao, A. Tet2 and Tet3 Cooperate with B-Lineage Transcription Factors to Regulate DNA Modification and Chromatin Accessibility. *Elife* **2016**, *5* (NOVEMBER2016). https://doi.org/10.7554/ELIFE.18290.

(50) Zhang, Q.; Zhao, K.; Shen, Q.; Han, Y.; Gu, Y.; Li, X.; Zhao, D.; Liu, Y.; Wang, C.; Zhang, X.; Su, X.; Liu, J.; Ge, W.; Levine, R. L.; Li, N.; Cao, X. Tet2 Is Required to Resolve Inflammation by Recruiting Hdac2 to Specifically Repress IL-6. *Nature* **2015**, *525* (7569), 389–393. https://doi.org/10.1038/NATURE15252.

(51) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **2011**, *333* (6047), 1300–1303. https://doi.org/10.1126/SCIENCE.1210597.

(52) Ficz, G.; Branco, M. R.; Seisenberger, S.; Santos, F.; Krueger, F.; Hore, T. A.; Marques, C. J.; Andrews, S.; Reik, W. Dynamic Regulation of 5-Hydroxymethylcytosine in Mouse ES Cells and during Differentiation. *Nature 2011 473:7347* **2011**, *473* (7347), 398–402. https://doi.org/10.1038/nature10008.

(53) Suzuki, M. M.; Bird, A. DNA Methylation Landscapes: Provocative Insights from Epigenomics. *Nature Reviews Genetics 2008 9:6* **2008**, *9* (6), 465–476. https://doi.org/10.1038/nrg2341.

(54) Lister, R.; Mukamel, E. A.; Nery, J. R.; Urich, M.; Puddifoot, C. A.; Johnson, N. D.; Lucero, J.; Huang, Y.; Dwork, A. J.; Schultz, M. D.; Yu, M.; Tonti-Filippini, J.; Heyn, H.; Hu, S.; Wu, J. C.; Rao, A.; Esteller, M.; He, C.; Haghighi, F. G.; Sejnowski, T. J.; Behrens, M. M.; Ecker, J. R. Global Epigenomic Reconfiguration during Mammalian Brain Development. *Science (1979)* **2013**, *341* (6146). https://doi.org/10.1126/SCIENCE.1237905/SUPPL_FILE/PAPV2.PDF.

(55) Guo, J. U.; Su, Y.; Shin, J. H.; Shin, J.; Li, H.; Xie, B.; Zhong, C.; Hu, S.; Le, T.; Fan, G.; Zhu, H.; Chang, Q.; Gao, Y.; Ming, G. L.; Song, H. Distribution, Recognition and Regulation of Non-CpG Methylation in the Adult Mammalian Brain. *Nature Neuroscience 2013 17:2* **2013**, *17* (2), 215–222. https://doi.org/10.1038/nn.3607.

(56) Jaenisch, R.; Bird, A. Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals. *Nature Genetics 2003 33:3* **2003**, *33* (3), 245–254. https://doi.org/10.1038/ng1089.

(57) Schultz, M. D.; He, Y.; Whitaker, J. W.; Hariharan, M.; Mukamel, E. A.; Leung, D.; Rajagopal, N.; Nery, J. R.; Urich, M. A.; Chen, H.; Lin, S.; Lin, Y.; Jung, I.; Schmitt, A. D.; Selvaraj, S.; Ren, B.; Sejnowski, T. J.; Wang, W.; Ecker, J. R. Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation. *Nature 2015 523:7559* **2015**, *523* (7559), 212–216. https://doi.org/10.1038/nature14465.

(58) Bonasio, R.; Tu, S.; Reinberg, D. Molecular Signals of Epigenetic States. *Science* **2010**, *330* (6004), 612–616. https://doi.org/10.1126/SCIENCE.1191078.

(59) Feng, S.; Jacobsen, S. E.; Reik, W. Epigenetic Reprogramming in Plant and Animal Development. *Science* **2010**, *330* (6004), 622. https://doi.org/10.1126/SCIENCE.1190614.

(60) Reik, W. Stability and Flexibility of Epigenetic Gene Regulation in Mammalian Development. *Nature 2007 447:7143* **2007**, *447* (7143), 425–432. https://doi.org/10.1038/nature05918.

(61) Iyer, L. M.; Tahiliani, M.; Rao, A.; Aravind, L. Prediction of Novel Families of Enzymes Involved in Oxidative and Other Complex Modifications of Bases in Nucleic Acids. *Cell Cycle* **2009**, *8* (11), 1698–1710. https://doi.org/10.4161/CC.8.11.8580.

(62) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **2009**, *324* (5929), 930–935. https://doi.org/10.1126/SCIENCE.1170116.

(63) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, N.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **2011**, *333* (6047), 1303–1307. https://doi.org/10.1126/SCIENCE.1210944.

(64) Ponnaluri, V. K. C.; Maciejewski, J. P.; Mukherji, M. A Mechanistic Overview of TET-Mediated 5-Methylcytosine Oxidation. *Biochem Biophys Res Commun* **2013**, *436* (2), 115–120. https://doi.org/10.1016/j.bbrc.2013.05.077.

(65) Tamanaha, E.; Guan, S.; Marks, K.; Saleh, L. Distributive Processing by the Iron(II)/α-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases. *J Am Chem Soc* **2016**, *138* (30), 9345–9348. https://doi.org/10.1021/JACS.6B03243/ASSET/IMAGES/LARGE/JA-2016-03243C_0003.JPEG.

(66) Wu, X.; Zhang, Y. TET-Mediated Active DNA Demethylation: Mechanism, Function and Beyond. *Nat. Rev. Genet.* **2017**, *18* (9), 517–534. https://doi.org/10.1038/nrg.2017.33.

(67) Langemeijer, S. M. C.; Kuiper, R. P.; Berends, M.; Knops, R.; Aslanyan, M. G.; Massop, M.; Stevens-Linders, E.; Van Hoogen, P.; Van Kessel, A. G.; Raymakers, R. A. P.; Kamping, E. J.; Verhoef, G. E.; Verburgh, E.; Hagemeijer, A.; Vandenberghe, P.; De Witte, T.; Van Der Reijden, B. A.; Jansen, J. H. Acquired Mutations in TET2 Are Common in Myelodysplastic Syndromes. *Nat Genet* **2009**, *41* (7), 838–842. https://doi.org/10.1038/NG.391.

(68) Smith, A. E.; Mohamedali, A. M.; Kulasekararaj, A.; Lim, Z. Y.; Gäken, J.; Lea, N. C.; Przychodzen, B.; Mian, S. A.; Nasser, E. E.; Shooter, C.; Westwood, N. B.; Strupp, C.; Gattermann, N.; Maciejewski, J. P.; Germing, U.; Mufti, G. J. Next-Generation Sequencing of the TET2 Gene in 355 MDS and CMMLpatients Reveals Low-Abundance Mutant Clones with Early Origins, but Indicates No Definite

Prognostic Value. *Blood* **2010**, *116* (19), 3923–3932. https://doi.org/10.1182/blood-2010-03-274704.

(69) Li, Z.; Cai, X.; Cai, C. L.; Wang, J.; Zhang, W.; Petersen, B. E.; Yang, F. C.; Xu, M. Deletion of Tet2 in Mice Leads to Dysregulated Hematopoietic Stem Cells and Subsequent Development of Myeloid Malignancies. *Blood* **2011**, *118* (17), 4509–4518. https://doi.org/10.1182/BLOOD-2010-12-325241.

(70) Moran-Crusio, K.; Reavie, L.; Shih, A.; Abdel-Wahab, O.; Ndiaye-Lobry, D.; Lobry, C.; Figueroa, M. E.; Vasanthakumar, A.; Patel, J.; Zhao, X.; Perna, F.; Pandey, S.; Madzo, J.; Song, C.; Dai, Q.; He, C.; Ibrahim, S.; Beran, M.; Zavadil, J.; Nimer, S. D.; Melnick, A.; Godley, L. A.; Aifantis, I.; Levine, R. L. Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell* **2011**, *20* (1), 11–24. https://doi.org/10.1016/J.CCR.2011.06.001.

(71) Ponnaluri, V. K. C.; Maciejewski, J. P.; Mukherji, M. A Mechanistic Overview of TET-Mediated 5-Methylcytosine Oxidation. *Biochem Biophys Res Commun* **2013**, *436* (2), 115–120. https://doi.org/10.1016/J.BBRC.2013.05.077.

(72) Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **2013**, *155* (7), 1545–1555. https://doi.org/10.1016/J.CELL.2013.11.020.

(73) Jaiswal, M.; Bhar, S.; Vemula, H.; Prakash, S.; Ponnaluri, V. K. C.; Gutheil, W. G.; Mukherji, M. Convenient Expression, Purification and Quantitative Liquid Chromatography-Tandem Mass Spectrometry-Based Analysis of TET2 5-

Methylcytosine Demethylase. *Protein Expr Purif* **2017**, *132*, 143–151. https://doi.org/10.1016/j.pep.2017.02.003.

(74) Mukherji, M.; Kershaw, N. J.; Schofield, C. J.; Wierzbicki, A. S.; Lloyd, M. D. Utilization of Sterol Carrier Protein-2 by Phytanoyl-CoA 2-Hydroxylase in the Peroxisomal α Oxidation of Phytanic Acid. *Chem Biol* **2002**, *9* (5), 597–605. https://doi.org/10.1016/S1074-5521(02)00139-4.

(75) Mukherji, M.; Kershaw, N. J.; MacKinnon, C. H.; Clifton, I. J.; Wierzbicki, A. S.; Schofield, C. J.; Lloyd, M. D. 'Chemical Co-Substrate Rescue' of Phytanoyl-CoA 2-Hydroxylase Mutants Causing Refsum's Disease. *Chemical Communications* **2001**, No. 11, 972–973. https://doi.org/10.1039/B101039P.

(76) Mukherji, M.; Chien, W.; Kershaw, N. J.; Clifton, I. J.; Schofield, C. J.; Wierzbicki, A. S.; Lloyd, M. D. Structure-Function Analysis of Phytanoyl-CoA 2-Hydroxylase Mutations Causing Refsum's Disease. *Hum Mol Genet* **2001**, *10* (18), 1971–1982. https://doi.org/10.1093/HMG/10.18.1971.

(77) Ferrone, C. K.; Blydt-Hansen, M.; Rauh, M. J. Age-Associated TET2 Mutations: Common Drivers of Myeloid Dysfunction, Cancer and Cardiovascular Disease. *Int J Mol Sci* **2020**, *21* (2). https://doi.org/10.3390/IJMS21020626.

(78) Ferrone, C. K.; Blydt-Hansen, M.; Rauh, M. J. Age-Associated TET2 Mutations: Common Drivers of Myeloid Dysfunction, Cancer and Cardiovascular Disease. *Int J Mol Sci* **2020**, *21* (2). https://doi.org/10.3390/IJMS21020626.

(79) Langemeijer, S. M. C.; Kuiper, R. P.; Berends, M.; Knops, R.; Aslanyan, M. G.; Massop, M.; Stevens-Linders, E.; van Hoogen, P.; van Kessel, A. G.; Raymakers, R. A. P.; Kamping, E. J.; Verhoef, G. E.; Verburgh, E.; Hagemeijer, A.; Vandenberghe,

P.; de Witte, T.; van der Reijden, B. A.; Jansen, J. H. Acquired Mutations in TET2 Are Common in Myelodysplastic Syndromes. *Nat Genet* **2009**, *41* (7), 838–842. https://doi.org/10.1038/ng.391.

(80) Guo, Z.; Zhang, S. kai; Zou, Z.; Fan, R. hua; Lyu, X. dong. Prognostic Significance of TET2 Mutations in Myelodysplastic Syndromes: A Meta-Analysis. *Leuk Res* **2017**, *58*, 102–107. https://doi.org/10.1016/j.leukres.2017.03.013.

(81) Moore, L. D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38* (1), 23–38. https://doi.org/10.1038/npp.2012.112.

(82) Jabbari, K.; Bernardi, G. Cytosine Methylation and CpG, TpG (CpA) and TpA Frequencies. *Gene* **2004**, *333* (SUPPL.), 143–149. https://doi.org/10.1016/j.gene.2004.02.043.

(83) Song, C.-X.; He, C. Balance of DNA Methylation and Demethylation in Cancer Development. *Genome Biology 2012 13:10* **2012**, *13* (10), 1–3. https://doi.org/10.1186/GB-2012-13-10-173.

(84) Portela, A.; Esteller, M. Epigenetic Modifications and Human Disease. *Nat Biotechnol* **2010**, *28* (10), 1057–1068. https://doi.org/10.1038/NBT.1685.

(85) Laurent, L.; Wong, E.; Li, G.; Huynh, T.; Tsirigos, A.; Ong, C. T.; Low, H. M.; Sung, K. W. K.; Rigoutsos, I.; Loring, J.; Wei, C. L. Dynamic Changes in the Human Methylome during Differentiation. *Genome Res* **2010**, *20* (3), 320. https://doi.org/10.1101/GR.101907.109.

(86) Ravichandran, M.; Jurkowska, R. Z.; Jurkowski, T. P. Target Specificity of Mammalian DNA Methylation and Demethylation Machinery. *Org Biomol Chem* **2018**, *16* (9), 1419–1435. https://doi.org/10.1039/C7OB02574B.

(87) Ziller, M. J.; Müller, F.; Liao, J.; Zhang, Y.; Gu, H.; Bock, C.; Boyle, P.; Epstein, C. B.; Bernstein, B. E.; Lengauer, T.; Gnirke, A.; Meissner, A. Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genet* **2011**, *7* (12), e1002389. https://doi.org/10.1371/JOURNAL.PGEN.1002389.

(88) Schultz, M. D.; He, Y.; Whitaker, J. W.; Hariharan, M.; Mukamel, E. A.; Leung, D.; Rajagopal, N.; Nery, J. R.; Urich, M. A.; Chen, H.; Lin, S.; Lin, Y.; Jung, I.; Schmitt, A. D.; Selvaraj, S.; Ren, B.; Sejnowski, T. J.; Wang, W.; Ecker, J. R. Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation. *Nature* **2015**, *523* (7559), 212–216. https://doi.org/10.1038/nature14465.

(89) Schultz, M. D.; He, Y.; Whitaker, J. W.; Hariharan, M.; Mukamel, E. A.; Leung, D.; Rajagopal, N.; Nery, J. R.; Urich, M. A.; Chen, H.; Lin, S.; Lin, Y.; Jung, I.; Schmitt, A. D.; Selvaraj, S.; Ren, B.; Sejnowski, T. J.; Wang, W.; Ecker, J. R. Corrigendum: Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation. *Nature* **2016**, *530* (7589), 242. https://doi.org/10.1038/NATURE16179.

(90) Patil, V.; Ward, R. L.; Hesson, L. B. The Evidence for Functional Non-CpG Methylation in Mammalian Cells. *Epigenetics* **2014**, *9* (6), 823–828. https://doi.org/10.4161/EPI.28741.

(91) Xiao, W.; Liu, X.; Niu, X.; Li, C.; Guo, Y.; Tan, J.; Xiong, W.; Fan, L.; Li, Y. The Frequency of CpG and Non-CpG Methylation of Notch3 Gene Promoter Determines

Its Expression Levels in Breast Cancer Cells. *Exp Cell Res* **2020**, *386* (2), 111743. https://doi.org/10.1016/j.yexcr.2019.111743.

(92) Li, C.; Xiong, W.; Liu, X.; Xiao, W.; Guo, Y.; Tan, J.; Li, Y. Hypomethylation at Non-CpG/CpG Sites in the Promoter of HIF-1α Gene Combined with Enhanced H3K9Ac Modification Contribute to Maintain Higher HIF-1α Expression in Breast Cancer. *Oncogenesis* **2019**, *8* (4), 26. https://doi.org/10.1038/s41389-019-0135-1.

(93) Sun, Z.; Dai, N.; Borgaro, J. G.; Quimby, A.; Sun, D.; Corrêa, I. R.; Zheng, Y.; Zhu, Z.; Guan, S. A Sensitive Approach to Map Genome-Wide 5-Hydroxymethylcytosine and 5-Formylcytosine at Single-Base Resolution. *Mol Cell* **2015**, *57* (4), 750–761. https://doi.org/10.1016/J.MOLCEL.2014.12.035.

(94) Kinde, B.; Gabel, H. W.; Gilbert, C. S.; Griffith, E. C.; Greenberg, M. E. Reading the Unique DNA Methylation Landscape of the Brain: Non-CpG Methylation, Hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A* **2015**, *112* (22), 6800–6806. https://doi.org/10.1073/PNAS.1411269112.

(95) Bestor, T. H. The DNA Methyltransferases of Mammals. *Hum Mol Genet* **2000**, *9* (16), 2395–2402. https://doi.org/10.1093/HMG/9.16.2395.

(96) Schübeler, D. Function and Information Content of DNA Methylation. *Nature 2015 517:7534* **2015**, *517* (7534), 321–326. https://doi.org/10.1038/nature14192.

(97) Ramsahoye, B. H.; Biniszkiewicz, D.; Lyko, F.; Clark, V.; Bird, A. P.; Jaenisch, R. Non-CpG Methylation Is Prevalent in Embryonic Stem Cells and May Be Mediated by DNA Methyltransferase 3a. *Proc Natl Acad Sci U S A* **2000**, *97* (10), 5237–5242. https://doi.org/10.1073/PNAS.97.10.5237.

(98) Gowher, H.; Jeltsch, A. Enzymatic Properties of Recombinant Dnmt3a DNA Methyltransferase from Mouse: The Enzyme Modifies DNA in a Non-Processive Manner and Also Methylates Non-CpA Sites. *J Mol Biol* **2001**, *309* (5), 1201–1208. https://doi.org/10.1006/jmbi.2001.4710.

(99) Barrès, R.; Osler, M. E.; Yan, J.; Rune, A.; Fritz, T.; Caidahl, K.; Krook, A.; Zierath, J. R. Non-CpG Methylation of the PGC-1alpha Promoter through DNMT3B Controls Mitochondrial Density. *Cell Metab* **2009**, *10* (3), 189–198. https://doi.org/10.1016/J.CMET.2009.07.011.

(100) Ichiyanagi, T.; Ichiyanagi, K.; Miyake, M.; Sasaki, H. Accumulation and Loss of Asymmetric Non-CpG Methylation during Male Germ-Cell Development. *Nucleic Acids Res* **2013**, *41* (2), 738–745. https://doi.org/10.1093/NAR/GKS1117.

(101) Shirane, K.; Toh, H.; Kobayashi, H.; Miura, F.; Chiba, H.; Ito, T.; Kono, T.; Sasaki, H. Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. *PLoS Genet* **2013**, *9* (4), e1003439. https://doi.org/10.1371/JOURNAL.PGEN.1003439.

(102) Arand, J.; Spieler, D.; Karius, T.; Branco, M. R.; Meilinger, D.; Meissner, A.; Jenuwein, T.; Xu, G.; Leonhardt, H.; Wolf, V.; Walter, J. In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genet* **2012**, *8* (6), e1002750. https://doi.org/10.1371/journal.pgen.1002750.

(103) Lee, J. H.; Park, S. J.; Nakai, K. Differential Landscape of Non-CpG Methylation in Embryonic Stem Cells and Neurons Caused by DNMT3s. *Sci Rep* **2017**, *7* (1). https://doi.org/10.1038/S41598-017-11800-1.

(104) Pais, J. E.; Dai, N.; Tamanaha, E.; Vaisvila, R.; Fomenkov, A. I.; Bitinaite, J.; Sun, Z.; Guan, S.; Corrêa, I. R.; Noren, C. J.; Cheng, X.; Roberts, R. J.; Zheng, Y.; Saleh, L.; Rao, A. Biochemical Characterization of a Naegleria TET-like Oxygenase and Its Application in Single Molecule Sequencing of 5-Methylcytosine. *Proc Natl Acad Sci U S A* **2015**, *112* (14), 4316–4321. https://doi.org/10.1073/PNAS.1417939112/SUPPL_FILE/PNAS.201417939SI.PDF.

(105) DeNizio, J. E.; Dow, B. J.; Serrano, J. C.; Ghanty, U.; Drohat, A. C.; Kohli, R. M. TET-TDG Active DNA Demethylation at CpG and Non-CpG Sites. *J Mol Biol* **2021**, *433* (8). https://doi.org/10.1016/J.JMB.2021.166877.

(106) 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* **2011**, *286* (41), 35334–35338. https://doi.org/10.1074/JBC.C111.284620.

(107) Iwan, K.; Rahimoff, R.; Kirchner, A.; Spada, F.; Schröder, A. S.; Kosmatchev, O.; Ferizaj, S.; Steinbacher, J.; Parsa, E.; Müller, M.; Carell, T. 5-Formylcytosine to Cytosine Conversion by C-C Bond Cleavage in Vivo. *Nat Chem Biol* **2018**, *14* (1), 72–78. https://doi.org/10.1038/NCHEMBIO.2531.

(108) Feng, Y.; Xie, N. Bin; Tao, W. B.; Ding, J. H.; You, X. J.; Ma, C. J.; Zhang, X.; Yi, C.; Zhou, X.; Yuan, B. F.; Feng, Y. Q. Transformation of 5-Carboxylcytosine to Cytosine Through C–C Bond Cleavage in Human Cells Constitutes a Novel Pathway for DNA Demethylation. *CCS Chemistry* **2021**, *3* (4), 994–1008. https://doi.org/10.31635/CCSCHEM.020.202000286.

(109) Papin, C.; Ibrahim, A.; Le Gras, S.; Velt, A.; Stoll, I.; Jost, B.; Menoni, H.; Bronner, C.; Dimitrov, S.; Hamiche, A. Combinatorial DNA Methylation Codes at Repetitive Elements. *Genome Res* **2017**, *27* (6), 934–946. https://doi.org/10.1101/GR.213983.116.

(110) Jang, H. S.; Shin, W. J.; Lee, J. E.; Do, J. T. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes (Basel)* **2017**, *8* (6), 2–20. https://doi.org/10.3390/GENES8060148.

(111) He, Y.; Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu Rev Genomics Hum Genet* **2015**, *16*, 55. https://doi.org/10.1146/ANNUREV-GENOM-090413-025437.

(112) Titcombe, P.; Murray, R.; Hewitt, M.; Antoun, E.; Cooper, C.; Inskip, H. M.; Holbrook, J. D.; Godfrey, K. M.; Lillycrop, K.; Hanson, M.; Barton, S. J. Human Non-CpG Methylation Patterns Display Both Tissue-Specific and Inter-Individual Differences Suggestive of Underlying Function. *Epigenetics* **2022**, *17* (6), 653–664. https://doi.org/10.1080/15592294.2021.1950990/SUPPL_FILE/KEPI_A_1950990_SM0731.ZIP.

(113) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front Genet* **2018**, *9*. https://doi.org/10.3389/fgene.2018.00640.

(114) Shi, D. Q.; Ali, I.; Tang, J.; Yang, W. C. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front Genet* **2017**, *8* (JUL), 270761. https://doi.org/10.3389/FGENE.2017.00100/BIBTEX.

(115) Feng, Y.; Chen, J.-J.; Xie, N.-B.; Ding, J.-H.; You, X.-J.; Tao, W.-B.; Zhang, X.; Yi, C.; Zhou, X.; Yuan, B.-F.; Feng, Y.-Q. Direct Decarboxylation of Ten-Eleven Translocation-Produced 5-Carboxylcytosine in Mammalian Genomes Forms a New Mechanism for Active DNA Demethylation. *Chem Sci* **2021**, *12* (34), 11322–11329. https://doi.org/10.1039/d1sc02161c.

(116) Peng, Y.; Li, S.; Onufriev, A.; Landsman, D.; Panchenko, A. R. Binding of Regulatory Proteins to Nucleosomes Is Modulated by Dynamic Histone Tails. *Nat Commun* **2021**, *12* (1), 5280. https://doi.org/10.1038/s41467-021-25568-6.

(117) Schrader, A.; Gross, T.; Thalhammer, V.; Längst, G. Characterization of Dnmt1 Binding and DNA Methylation on Nucleosomes and Nucleosomal Arrays. *PLoS One* **2015**, *10* (10), e0140076. https://doi.org/10.1371/journal.pone.0140076.

(118) Takeshima, H.; Suetake, I.; Shimahara, H.; Ura, K.; Tate, S.; Tajima, S. Distinct DNA Methylation Activity of Dnmt3a and Dnmt3b towards Naked and Nucleosomal DNA. *J Biochem* **2006**, *139* (3), 503–515. https://doi.org/10.1093/jb/mvj044.

(119) Wu, H.; Wu, X.; Shen, L.; Zhang, Y. Single-Base Resolution Analysis of Active DNA Demethylation Using Methylase-Assisted Bisulfite Sequencing. *Nat Biotechnol* **2014**, *32* (12), 1231–1240. https://doi.org/10.1038/nbt.3073.

(120) Kizaki, S.; Zou, T.; Li, Y.; Han, Y.-W.; Suzuki, Y.; Harada, Y.; Sugiyama, H. Preferential 5-Methylcytosine Oxidation in the Linker Region of Reconstituted Positioned Nucleosomes by Tet1 Protein. *Chemistry* **2016**, *22* (46), 16598–16601. https://doi.org/10.1002/chem.201602435.

(121) Kizaki, S.; Sugiyama, H. CGmCGCG Is a Versatile Substrate with Which to Evaluate Tet Protein Activity. *Org Biomol Chem* **2014**, *12* (1), 104–107. https://doi.org/10.1039/c3ob41823e.

(122) Bhattacharya, C.; Dey, A. S.; Ayon, N. J.; Gutheil, W. G.; Mukherji, M. Efficient Purification and LC-MS/MS-Based Assay Development for Ten-Eleven Translocation-2 5-Methylcytosine Dioxygenase. *J Vis Exp* **2018**, No. 140. https://doi.org/10.3791/57798.

(123) Dey, A. S.; Ayon, N. J.; Bhattacharya, C.; Gutheil, W. G.; Mukherji, M. Positive/Negative Ion-Switching Based LC-MS/MS Method for Quantification of Cytosine Derivatives Produced by the TET-Family 5-Methylcytosine Dioxygenases. *Biol Methods Protoc* **2020**. https://doi.org/10.1093/biomethods/bpaa019.

(124) Tefferi, A.; Lim, K.-H.; Abdel-Wahab, O.; Lasho, T. L.; Patel, J.; Patnaik, M. M.; Hanson, C. A.; Pardanani, A.; Gilliland, D. G.; Levine, R. L. Detection of Mutant TET2 in Myeloid Malignancies Other than Myeloproliferative Neoplasms: CMML, MDS, MDS/MPN and AML. *Leukemia*. July 2009, pp 1343–1345. https://doi.org/10.1038/leu.2009.59.

(125) Huang, Y.; Pastor, W. A.; Shen, Y.; Tahiliani, M.; Liu, D. R.; Rao, A. The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. *PLoS One* **2010**, *5* (1), e8888. https://doi.org/10.1371/journal.pone.0008888.

(126) Crawford, D. J.; Liu, M. Y.; Nabel, C. S.; Cao, X.-J.; Garcia, B. A.; Kohli, R. M. Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de Novo Mechanism. *J Am Chem Soc* **2016**, *138* (3), 730–733. https://doi.org/10.1021/jacs.5b10554.

(127) Ashton, N. W.; Bolderson, E.; Cubeddu, L.; O'Byrne, K. J.; Richard, D. J. Human Single-Stranded DNA Binding Proteins Are Essential for Maintaining Genomic Stability. *BMC Mol Biol* **2013**, *14*, 9. https://doi.org/10.1186/1471-2199-14-9.

(128) Christman, J. K.; Sheikhnejad, G.; Marasco, C. J.; Sufrin, J. R. 5-Methyl-2'-Deoxycytidine in Single-Stranded DNA Can Act in Cis to Signal de Novo DNA Methylation. *Proc Natl Acad Sci U S A* **1995**, *92* (16), 7347–7351. https://doi.org/10.1073/pnas.92.16.7347.

(129) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **2011**, *333* (6047), 1303–1307. https://doi.org/10.1126/science.1210944.

(130) Lynch, M.; Marinov, G. K. The Bioenergetic Costs of a Gene. *Proc Natl Acad Sci U S A* **2015**, *112* (51), 15690–15695. https://doi.org/10.1073/PNAS.1514974112/SUPPL_FILE/PNAS.1514974112.SAPP.PDF.

(131) Dammann, R.; Strunnikova, M.; Schagdarsurengin, U.; Rastetter, M.; Papritz, M.; Hattenhorst, U. E.; Hofmann, H.-S.; Silber, R.-E.; Burdach, S.; Hansen, G. CpG Island Methylation and Expression of Tumour-Associated Genes in Lung Carcinoma. *Eur J Cancer* **2005**, *41* (8), 1223–1236. https://doi.org/10.1016/j.ejca.2005.02.020.

(132) Shi, D.-Q.; Ali, I.; Tang, J.; Yang, W.-C. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front Genet* **2017**, *8*, 100. https://doi.org/10.3389/fgene.2017.00100.

(133) Dey, A. S.; Bhattacharya, C.; Guan, Y.; Jha, B. K.; Mukherji, M. Demethylation of Non-CpG Sites in DNA Is Initiated by TET2 5-Methylcytosine Dioxygenase. *DNA* . 2021. https://doi.org/10.3390/dna1010004.

(134) Guo, J. U.; Su, Y.; Shin, J. H.; Shin, J.; Li, H.; Xie, B.; Zhong, C.; Hu, S.; Le, T.; Fan, G.; Zhu, H.; Chang, Q.; Gao, Y.; Ming, G.; Song, H. Distribution, Recognition and Regulation of Non-CpG Methylation in the Adult Mammalian Brain. *Nat Neurosci* **2014**, *17* (2), 215–222. https://doi.org/10.1038/nn.3607.

(135) Ziller, M. J.; Müller, F.; Liao, J.; Zhang, Y.; Gu, H.; Bock, C.; Boyle, P.; Epstein, C. B.; Bernstein, B. E.; Lengauer, T.; Gnirke, A.; Meissner, A. Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genet* **2011**, *7* (12), e1002389. https://doi.org/10.1371/journal.pgen.1002389.

(136) Lee, J. H.; Saito, Y.; Park, S. J.; Nakai, K. Existence and Possible Roles of Independent Non-CpG Methylation in the Mammalian Brain. *DNA Research* **2020**, *27* (4), 1–11. https://doi.org/10.1093/DNARES/DSAA020.

(137) Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; MacIejewski, J. P.; Rao, A. Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2. *Nature* **2010**, *468* (7325), 839–843. https://doi.org/10.1038/NATURE09586.

(138) Figueroa, M. E.; Abdel-Wahab, O.; Lu, C.; Ward, P. S.; Patel, J.; Shih, A.; Li, Y.; Bhagwat, N.; Vasanthakumar, A.; Fernandez, H. F.; Tallman, M. S.; Sun, Z.; Wolniak, K.; Peeters, J. K.; Liu, W.; Choe, S. E.; Fantin, V. R.; Paietta, E.; Löwenberg, B.; Licht, J. D.; Godley, L. A.; Delwel, R.; Valk, P. J. M.; Thompson, C. B.; Levine, R. L.; Melnick, A. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **2010**, *18* (6), 553. https://doi.org/10.1016/J.CCR.2010.11.015.

(139) Brewitz, L.; Tumber, A.; Nakashima, Y.; Schofield, C. J. Novel 2-Oxoglutarate Analogues Modulate the Epigenetic Activity of the Cancer-Related Human Enzyme Aspartate/Asparagine-β-Hydroxylase. *The FASEB Journal* **2020**, *34* (S1), 1–1. https://doi.org/10.1096/FASEBJ.2020.34.S1.00468.

(140) Nakashima, Y.; Brewitz, L.; Tumber, A.; Salah, E.; Schofield, C. J. 2-Oxoglutarate Derivatives Can Selectively Enhance or Inhibit the Activity of Human Oxygenases. *Nature Communications 2021 12:1* **2021**, *12* (1), 1–17. https://doi.org/10.1038/s41467-021-26673-2.

(141) Laukka, T.; Myllykoski, M.; Looper, R. E.; Koivunen, P. Cancer-Associated 2-Oxoglutarate Analogues Modify Histone Methylation by Inhibiting Histone Lysine Demethylases. *J Mol Biol* **2018**, *430* (18 Pt B), 3081–3092. https://doi.org/10.1016/J.JMB.2018.06.048.

(142) Jiang, S. Tet2 at the Interface between Cancer and Immunity. *Communications Biology 2020 3:1* **2020**, *3* (1), 1–6. https://doi.org/10.1038/s42003-020-01391-5.

(143) Delhommeau, F.; Dupont, S.; Valle, V. Della; James, C.; Trannoy, S.; Massé, A.; Kosmider, O.; Le Couedic, J.-P.; Robert, F.; Alberdi, A.; Lécluse, Y.; Plo, I.; Dreyfus, F. J.; Marzac, C.; Casadevall, N.; Lacombe, C.; Romana, S. P.; Dessen, P.; Soulier, J.; Viguié, F.; Fontenay, M.; Vainchenker, W.; Bernard, O. A. Mutation in TET2 in Myeloid Cancers . *New England Journal of Medicine* **2009**, *360* (22), 2289–2301. https://doi.org/10.1056/NEJMOA0810069/SUPPL_FILE/NEJM_DELHOMMEAU_2289SA1.PDF.

(144) Kosmider, O.; Gelsi-Boyer, V.; Cheok, M.; Grabar, S.; Della-Valle, V.; Picard, F.; Viguié, F.; Quesnel, B.; Beyne-Rauzy, O.; Solary, E.; Vey, N.; Hunault-Berger, M.; Fenaux, P.; Mansat-De Mas, V.; Delabesse, E.; Guardiola, P.; Lacombe, C.; Vainchenker, W.; Preudhomme, C.; Dreyfus, F.; Bernard, O. A.; Birnbaum, D.; Fontenay, M. TET2 Mutation Is an Independent Favorable Prognostic Factor in Myelodysplastic Syndromes (MDSs). *Blood* **2009**, *114* (15), 3285–3291. https://doi.org/10.1182/BLOOD-2009-04-215814.

(145) Chou, W. C.; Chou, S. C.; Liu, C. Y.; Chen, C. Y.; Hou, H. A.; Kuo, Y. Y.; Lee, M. C.; Ko, B. S.; Tang, J. L.; Yao, M.; Tsay, W.; Wu, S. J.; Huang, S. Y.; Hsu, S. C.; Chen, Y. C.; Chang, Y. C.; Kuo, Y. Y.; Kuo, K. T.; Lee, F. Y.; Liu, M. C.; Liu, C. W.; Tseng, M. H.; Huang, C. F.; Tien, H. F. TET2 Mutation Is an Unfavorable Prognostic Factor in Acute Myeloid Leukemia Patients with Intermediate-Risk Cytogenetics. *Blood* **2011**, *118* (14), 3803–3810. https://doi.org/10.1182/BLOOD-2011-02-339747.

(146) Iyer, L. M.; Tahiliani, M.; Rao, A.; Aravind, L. Prediction of Novel Families of Enzymes Involved in Oxidative and Other Complex Modifications of Bases in Nucleic Acids. *Cell Cycle* **2009**, *8* (11), 1698. https://doi.org/10.4161/CC.8.11.8580.

(147) Breski, M.; Dey, D.; Obringer, S.; Sudhamalla, B.; Islam, K. Engineering Biological C-H Functionalization Leads to Allele-Specific Regulation of Histone Demethylases. *J Am Chem Soc* **2016**, *138* (41), 13505–13508. https://doi.org/10.1021/JACS.6B08653/SUPPL_FILE/JA6B08653_SI_001.PDF.

(148) Pérez, C.; Martínez-Calle, N.; Martín-Subero, J. I.; Segura, V.; Delabesse, E.; Fernandez-Mercado, M.; Garate, L.; Alvarez, S.; Rifon, J.; Varea, S.; Boultwood, J.; Wainscoat, J. S.; Cigudosa, J. C.; Calasanz, M. J.; Cross, N. C. P.; Prósper, F.; Agirre, X. TET2 Mutations Are Associated with Specific 5-Methylcytosine and 5-Hydroxymethylcytosine Profiles in Patients with Chronic Myelomonocytic Leukemia. *PLoS One* **2012**, *7* (2), e31605. https://doi.org/10.1371/JOURNAL.PONE.0031605.

(149) Yamazaki, J.; Jelinek, J.; Lu, Y.; Cesaroni, M.; Madzo, J.; Neumann, F.; He, R.; Taby, R.; Vasanthakumar, A.; Macrae, T.; Ostler, K. R.; Kantarjian, H. M.; Liang, S.; Estecio, M. R.; Godley, L. A.; Issa, J. P. J. TET2 Mutations Affect Non-CpG Island DNA Methylation at Enhancers and Transcription Factor-Binding Sites in Chronic Myelomonocytic Leukemia. *Cancer Res* **2015**, *75* (14), 2833–2843. https://doi.org/10.1158/0008-5472.CAN-14-0739/651401/AM/TET2-MUTATIONS-AFFECT-NON-CPG-ISLAND-DNA.

(150) Quivoron, C.; Couronné, L.; Della Valle, V.; Lopez, C. K.; Plo, I.; Wagner-Ballon, O.; Do Cruzeiro, M.; Delhommeau, F.; Arnulf, B.; Stern, M. H.; Godley, L.; Opolon,

P.; Tilly, H.; Solary, E.; Duffourd, Y.; Dessen, P.; Merle-Beral, H.; Nguyen-Khac, F.; Fontenay, M.; Vainchenker, W.; Bastard, C.; Mercher, T.; Bernard, O. A. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* **2011**, *20* (1), 25–38. https://doi.org/10.1016/J.CCR.2011.06.003.

(151) Bussaglia, E.; Antón, R.; Nomdedéu, J. F.; Fuentes-Prior, P. TET2 Missense Variants in Human Neoplasia. A Proposal of Structural and Functional Classification. *Mol Genet Genomic Med* **2019**, *7* (7). https://doi.org/10.1002/MGG3.772.

VITA

Chayan Bhattacharya is a PhD candidate in Pharmaceutical Sciences at the School of Pharmacy (SOP), University of Missouri-Kansas City (UMKC). Prior to this, he received his bachelor's degree in microbiology (Honors) from West Bengal State University, India. During the undergraduate program, he was exposed to various allied modules, such as biochemistry, general microbiology, and immunology. Subsequently, he discovered his interest in molecular biology and completed his master's degree in molecular biology and Genetics with a GPA of 4.0 on a 4-point scale from Presidency University, one of the prestigious institutes in India.

During his master's course, he took the opportunity to explore his compatibility with biological research by participating in two independent research projects on toxicology and cancer biology. His first introduction in research was during his master's degree in 2012 in the field of toxicology. This study assessed the genotoxic and cytotoxic effects of Multiwalled Carbon Nanotubes (MWCNT) on Allium cepa, which he performed at the University of Calcutta under the supervision of Prof. Anita Mukherjee. In the same year, after completing his work, he joined the Chittaranjan National Cancer Institute for another project under the supervision of Dr. Chinmay Kumar Panda. Here his research was on mutation analysis of the Epidermal Growth Factor Receptor (EGFR) in triple negative breast cancer (TNBC) cases. In 2013, after his master's degree, he got the opportunity to join St. John's Research Institute, Bangalore, as a research assistant, where his work was based on the seroepidemiology of dengue virus and the implications of high mobility group box protein 1 as a biomarker of disease severity under the supervision of Dr. Ujjwal Neogi.

Chayan joined the interdisciplinary PhD program at the UMKC School of Pharmacy in Fall 2016. His doctoral research focuses on the substrate preference of 5-methylcytosine dioxygenase tet2 and rescue of mutant tet2 activity using a turbo co-substrate. His research provided multiple lines of evidence that human TET2 can oxidize 5mCpH sites (CpA, CpC, and CpT) in DNA. They also identified 2OG derivatives in which the 5-carboxylate group was replaced with an aliphatic chain, and these compounds enhanced the activity of hTET2 clinical mutations R1896S, R1896A, and R1896F by 3 to 4 folds. His work has been published in various peer-reviewed journals, including the Journal of Visualized Experiments, DNA, Cell Biochemistry & Function.

LIST OF FIRST AUTHOR PUBLICATIONS

- Bhattacharya, C., Dey, A. S., Ayon, N. J., Gutheil, W. G., Mukherji, M. Efficient Purification and LC-MS/MS-based Assay Development for Ten-Eleven Translocation-2 5-Methylcytosine Dioxygenase. J. Vis. Exp. (140), e57798, doi:10.3791/57798 (2018).
- Dey, A.S.; Bhattacharya, C.; Guan, Y.; Jha, B.K.; Mukherji, M. Demethylation of Non-CpG Sites in DNA Is Initiated by TET2 5-Methylcytosine Dioxygenase. *DNA* **2021**, *1*, 26-36. https://doi.org/10.3390/dna1010004.
- Bhattacharya C, Dey AS, Mukherji M. Substrate DNA length regulates the activity of TET 5-methylcytosine dioxygenases. Cell Biochem Funct. 2023 Jun 22. doi: 10.1002/cbf.3825. Epub ahead of print. PMID: 37349892.