DETERMINATION OF NOVEL GENE TARGETS OF TETS VIA NON-CANONICAL PROTEIN AND DNA MODIFICATIONS AND SELECTIVE INHIBITION

by

Kirsten E. Shriver

B.A. Chemistry and Spanish, Grinnell College, 2012

M.Ed. Southern Methodist University, 2014

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Kirsten Shriver

It was defended on

November 22, 2016

and approved by

W. Seth Horne, Associate Professor, Chemistry

W. Seth Childers, Assistant Professor, Chemistry

Thesis Director: Kabirul Islam, Assistant Professor, Chemistry

DETERMINATION OF NOVEL GENE TARGETS OF TETS VIA NON-CANONICAL PROTEIN AND DNA MODIFICATIONS AND SELECTIVE INHIBITION

Kirsten Shriver, M.S.

University of Pittsburgh, 2017

Abstract:

The Ten-Eleven translocation (TET) family of enzymes mediates the oxidation of DNA methylation, but the specific gene targets and the role of these enzymes in human development and disease are not well understood. The overall goals of this work are to profile TET-specific gene targets and to develop potent and specific small molecule inhibitors for TET family members. To accomplish these goals, a photo-crosslinking system and non-canonically modified DNA substrates will be developed as chemical tools to elucidate TET member-specific gene targets. Additionally, the development of specific, small molecule inhibitors will be explored. Photo-crosslinkable TET enzymes were developed, via incorporation of a photo-activable unnatural amino acid into the catalytic domain; three mutant enzymes were found to crosslink to methyl cytosine containing DNA substrates. This crosslinking system will be generalized across family members and applied in cells to identify novel gene targets of TET enzymes. DNA substrates with non-canonical modifications were tested in the development of a DNMT-TET tandem system to functionalize oxidation products. Ethyl, allyl, and propyl substituted DNA substrates were successfully oxidized by both wild type and mutant TET proteins. Potential inhibitor compounds were screened against the TET2 wild type enzyme, and a few compounds indicated slight inhibitory activity. A fragment based approach was employed to design potential linked inhibitor compounds, which will be synthesized and tested against TET enzymes.

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

1.1 BACKGROUND AND SIGNIFICANCE

The human genome is regulated by a team of enzymes which install, remove, and interpret chemical modifications on histones, DNA, and RNA; the study of these modifications and their effects on gene expression is known as epigenetics.[1] Control of gene expression is fundamental to all major biological processes, including the regulation of embryonic development and oncogenesis.[2] Epigenetic modifications can operate on vastly different scales, from silencing entire chromosomes to specific chemical changes on single bases.[1] One highly regulated set of chemical modifications is the acetylation and methylation of histone tails, DNA bases, and RNA bases.[3] DNA methylation has important regulatory roles in gene expression and development across plants, animals, and other organisms.[4]

DNA is typically methylated at areas of symmetrical cytosine and guanine pairs, known as CpG islands, on genomic DNA.[5] These methylation marks are often clustered together near transcription start sites and along promoter regions of genes, and generally result in repressed genomic loci.[1,6] Once the methyl marker is removed, the gene expression at that locus generally increases.[7] Enzymes known as DNA methyl transferases (DNMTs), with cofactor S-adenosyl methionine (SAM), are responsible for converting cytosine residues to five methyl cytosine (5mC) residues through the installation of the methyl group at the five position on cytosine

residues.[8] There are multiple DNMTs with different roles in installing and maintaining genomic methylation patterns. DNMT3a and DNMT3b are known as *de novo* methyltransferases which establish patterns of DNA methylation during development; DNMT1 is the methyl transferase responsible for maintenance methylation, ensuring methylation patterns are maintained through cycles of cell division.[8,9] While the process of cytosine methylation and the roles of DNMTs in generating and maintaining a methylated state of genomic loci are well accepted, the processes, roles, and players involved in demethylation are not well characterized.[3,10]

The Ten-Eleven translocation (TET) family of enzymes is known to be significantly involved in the epigenetic regulation of DNA methylation.[10] This family is made up of enzymes that catalyze the demethylation of 5mC, through successive oxidation of the methyl group to a hydroxyl (5hmC), formyl (5fC), and carboxyl (5caC) group (Figure 1).[7,11–13,10]

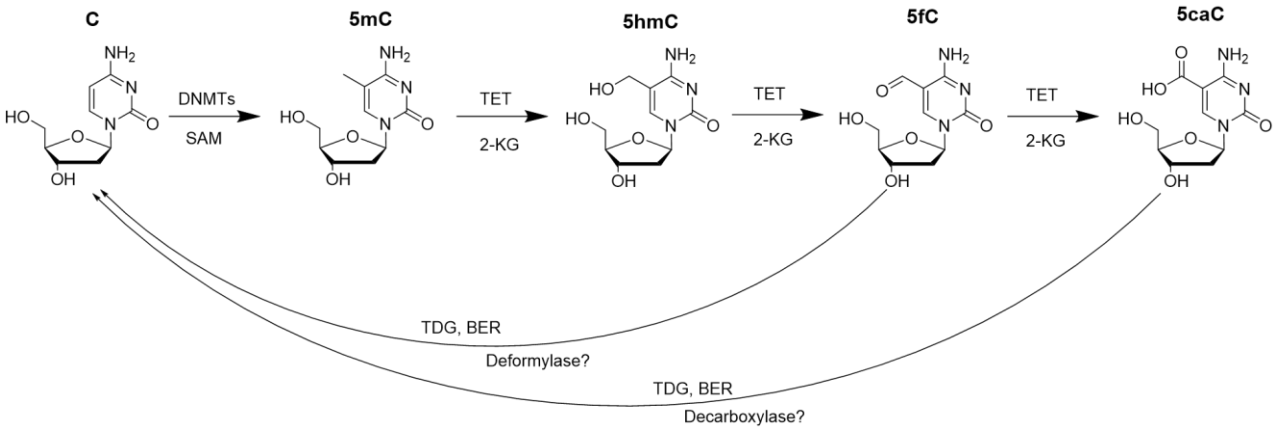


Figure 1. Oxidation of 5mC to 5caC by TETs

The process of oxidation is well characterized, but the removal of hydroxyl, formyl, and carboxyl groups and the restoration of unmodified cytosine is not well understood.[10] Thymine-DNA glycosylase (TDG) has been shown to remove formyl and carboxyl cytosines, followed by restoration of cytosine by base excision repair (BER).[13] 5hmC, 5fC, and 5caC have been

considered to be intermediates in the process of active DNA demethylation, but further work has indicated that these intermediates could be additional stable epigenetic markers.[6,14–16]

TET1, the original TET family member, was first identified in 2002 by scientists studying chromosomal abnormalities in acute myeloid leukemia.[17] In rare cases of the disease, a translocation event between chromosome 10 and chromosome 11 formed a fusion product between TET1 and the mixed lineage leukemia (MLL) gene.[15,17] While this gene was determined to be a member of a novel well-conserved protein family,[17] its enzymatic activity remained undiscovered until 2009, when Tahiliani and coworkers showed that human TET1 oxidized 5mC to 5hmC.[7,15] Later work by Ito, *et al*. and He, *et al*. showed that mouse TET1, TET2, and TET3 performed the same oxidation of 5mC to 5hmC , and that all three TETs produced additional oxidation products, 5-formyl cytosine (5fC) and 5-carboxy cytosine (5caC).[11–13]

Epigenetic methylation patterns are established in early development and differentiation in order to ensure that differentiated cells retain their unique identities and that these patterns are maintained through cycles of cell division.[18] However, in order to reproduce, cells must regain their pluripotent identity.[18] After fertilization, a wave of demethylation removes parental methylation patterns from oocyte and sperm DNA, which grants pluripotency to the zygote and allows it to form all of the cell types of the developing organism[18,19]. Studies in mouse models have tracked the changes in methylation patterns throughout early embryogenesis.[7,18] In mouse zygotes, between fertilization and the 8-16 cell stage, there is a decrease in 5mC and concurrent increase in 5hmC, 5fC, and 5caC; over time the levels of 5hmC and other intermediates are seen to decrease as demethylation is completed.[15] TET3 expression is high during these developmental stages.[15,20] Gu *et al*., have shown that TET3 is responsible for the sharp decrease

in 5mC in zygotes, as maternal TET3 is known to oxidize the methylated male pro-nucleus, resulting in demethylation.[19] The deletion of maternal TET3 disrupts prenatal development; although embryos implant normally, these embryos are less likely to reach full term development. TET3 conditional knockout mice were also shown to have increased developmental deformities, with some even resulting in non-viability.[19] This suggests that TET3-mediated demethylation is required for successful development during its earliest stages. By the blastula stage, TET3 expression falls, and TET1 and TET2 expression are high.[20] At this point, there is an increase in 5mC and 5hmC levels,[15,19] triggering the cells of the developing embryo to differentiate into different cell types and gain their own methylation patterns.[18]

Due to their roles in pluripotency and differentiation, TET enzymes are expected to play a role in embryonic stem cells (ESCs). Both TET1 and TET2 are expressed in ESCs, while TET3 is expressed as cells differentiate and is present in various cell types.[21] TET1 and TET2 are influential in modifying methylation and hydroxymethylation patterns in ESCs, and have been shown to cause changes in quantity of 5mC and 5hmC.[15,21] The developmental stage-specific expression of TET proteins and their non-redundant functions in developing mice suggests that they act on different genes for cellular differentiation.[20,21] In triple TET knockout mouse ESCs, Dawlaty, *et al*. have shown that while ESC morphology is maintained, upon differentiation, triple knock out cells are more poorly differentiated than WT controls and have significant hypermethylation in promotor regions of genes with significant roles in development, indicating that TETs have a critical contribution in differentiation of ESCs.[21]

In addition to early development, differentiation, and stem-cell pluripotency, TET enzymes are known to be involved in the regulation of methylation patterns in cancer. Abnormal expression and somatic mutation of TET proteins have been linked to various types of cancer,

such as melanoma[2] and leukemia.[21,22] Relatedly, a loss of 5hmC is regarded as one of the "hallmarks" of oncogenesis.[2,23] It is hypothesized that changes in methylation patterns downregulate tumor suppressor genes, and the concurrently observed decreased 5hmC levels are related to both these changes in methylation patterns and TET expression or activity.[7] In melanoma, a genome-wide loss of hmC was observed in tandem with downregulation of IDH2 and all three TETs, with TET2 downregulated most significantly.[2] The loss of hmC was reversible with the restoration of either isocitrate dehydrogenase 2 (IDH2) or TET2, indicating the involvement of multiple biological processes in the loss of 5hmC and hypermethylation in melanomas.[2] Thienpont *et al.,* showed that in solid tumors, hypoxia and the subsequently established tumor microenvironment down regulates TET activity due to the severely low oxygen levels, resulting in a loss of 5hmC, thus allowing for hypermethylation at promotor regions, likely suppressing tumor suppressor genes, and ultimately contributing to oncogenesis.[23] Mechanistic consideration indicates that lack of oxygen does not allow for proper coordination of the enzyme, small molecule cofactor, and DNA substrate, preventing the oxidation reaction and subsequent demethylation from occurring.[2,23]

TET enzymes belong to the Fe (II) and 2-ketoglutarate (2-KG)-dependent dioxygenase superfamily, meaning both Fe (II) and 2-KG are critical for these enzymes to carry out their oxidative functions.[3,4,15] The small molecule cofactor, 2-KG, is synthesized *in vivo* by isocitrate dehydrogenase (IDH).[2] All members of the TET family contain a conserved C-terminal catalytic domain, responsible for demethylation,[20,5] which consists of a cysteine-rich region, a double stranded beta helix (DSBH) and a spacer with unknown function (Figure 2).[20] Together the cysteine rich region and DSBH are responsible for the oxidative catalytic activity of the TET enzymes.[20,5] TET1 and TET3 contain an N-terminal CXXC zinc finger domain which

recognizes DNA substrates, including CpG islands. TET2 lacks this CXXC domain, which was evolutionarily separated from TET2 and is now coded as a separate gene known as IDAX. IDAX acts as a cofactor, binding to CpG islands and recruiting TET2.[20,24] While the combination of the cysteine-rich and DSBH regions are responsible for the oxidative catalytic activity of TET enzymes, it is the role of the CXXC domains to recognize DNA substrates.[20,24,25]

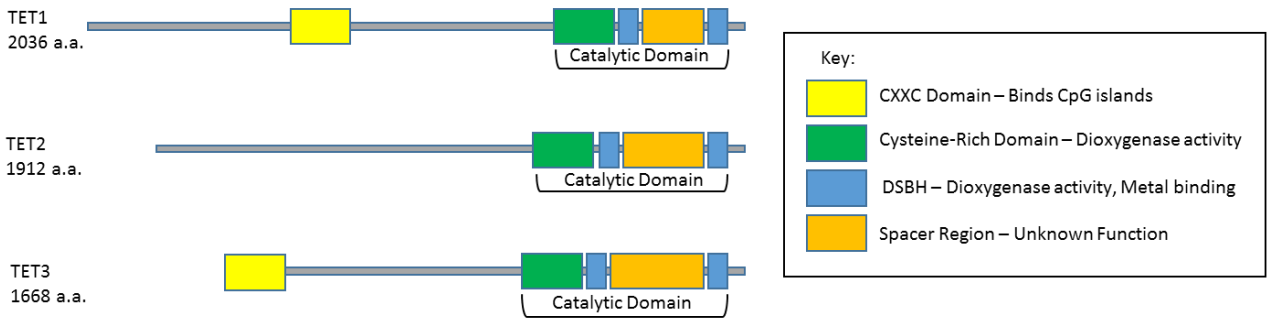


Figure 2. TET Domain Organization and Functions[20]

The mechanism of oxidation proposed for TET and similar enzymes is an Fe (II) dependent catalytic cycle, involving the coordination of 2-KG, water, two histidine residues, and an aspartic acid residue (Figure 3).[10] The oxidative addition of molecular oxygen forms a peroxo intermediate, which releases a molecule of CO_2 to result in the ferroxo complex. A radical mechanism allows for the addition of oxygen to the 5mC and the reduction of the iron to Fe (III). The reductive elimination of the oxidized cytosine returns the complex to Fe (II), which is coordinated to the histidine and aspartic acid residues and succinate. The replacement of succinate with 2-KG and water allows for further substrate oxidation.[10] While the chemical mechanism of how TET oxidizes 5mC to demethylate DNA is understood, how each TET member acts within their endogenous biological context is unknown.[10]

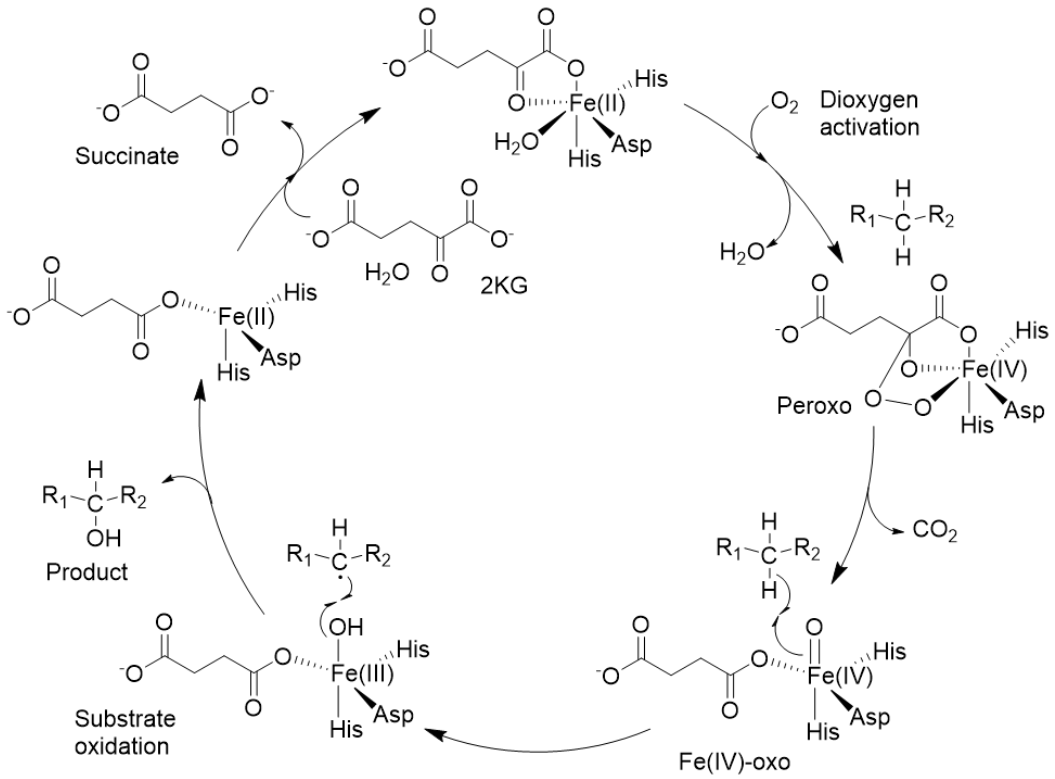


Figure 3. Mechanism of Oxidation[10]

1.2 GOALS AND SPECIFIC AIMS

TET enzymes play critical roles in gene expression, development, and cancer. Researchers have broadly studied the overall functions of TETs using expression patterns and knockout models,[19,21] and we understand the fine mechanistic details of how TET enzymes oxidize methylated cytosines.[10] However, between these broad and fine approaches, there is a gap in our knowledge; we do not know the specific genes regulated by individual TET family members in specific contexts and times in development, differentiation, and cancer. Our methods to identify TET-specific gene targets allow for resolution of gene targets that are specifically regulated by the enzymatic activity of TET enzymes; current methods used to

identify DNA-protein interacting partners, such as chromatin immunoprecipitation (ChIP), cannot distinguish between interactions resulting in enzymatic activity and those that may interact though non-enzymatic localization or binding. This work bridges gaps in current knowledge about the roles, functions, and behaviors of TET proteins. The overall goals of my projects are to profile TET-specific gene targets and to develop potent and specific small molecule inhibitors for TET family members. These goals will be achieved through the following specific aims:

Specific Aim 1: Identify TET member-specific gene targets by developing photo-crosslinkable TET enzymes

The member-specific roles of TET enzymes in development and oncogenesis need to be better understood. By identifying the specific genes that are demethylated by TET proteins, precise roles of TET members in development and pathogenesis can be determined. A photo-crosslinking approach, via incorporation of a photo-activable unnatural amino acid into the catalytic domain of TET proteins is being employed. A photo-crosslinkable amino acid will allow for crosslinking to specific gene targets, identifying member specific roles of TET enzymes.

Specific Aim 2: Develop a DNMT-TET tandem system to functionalize DNA substrates through non-canonical DNA modifications

Because the profile of TET-specific activity along genomic loci is not well characterized, it is rational to approach this challenge through multiple strategies. The introduction of non-canonical modifications on DNA via DNMTs that can be further oxidized by TETs allows for

orthogonal functionalization of DNA modifications to selectively profile gene targets of TETs. Functionalized DNA modifications can be reacted with different probes for visualization and pull-down applications. By identifying specific TET gene targets, the roles of TET enzymes in biological processes will be elucidated.

Specific Aim 3: Study the role of TETs in development by developing TET specific inhibitors

The role of TET enzymes in developmental processes is not well understood. To gain a more complete understanding of how TETs regulate aspects of development, potent and specific small molecule inhibitors need to be developed. A fragment-based design principle, combining analogs of nucleotides and 2-KG cofactor-like compounds as fragments to increase potency and specificity is being used to develop small molecule inhibitors. Rationally designed small molecule inhibitors will allow the activity of TET proteins to be perturbed on a rapid timescale suitable to the dynamic embryonic developmental process to further elucidate the role of TETs in human biology.

By working to identify both the genes regulated by TET proteins, as well as develop a small molecule inhibitor for these proteins, we will gain new insights and tools for the study of TET proteins and their roles in human development and diseases such as cancer.

2.0 SPECIFIC AIM 1: TET2 PHOTO-CROSSLINKING

2.1 OVERVIEW OF PHOTO-CROSSLINKING

A photo-crosslinking approach is used to determine the specific genes regulated by TET family members. Utilizing technology pioneered by Schultz and coworkers, an unnatural amino acid (UAA), *para*-azidophenylalanine (pAzF), is incorporated into the catalytic pocket of TET enzymes with the use of an evolved, orthogonal, *Methanococcus jannaschii* TyrRS-tRNACUATy pair (Figure 4). [26–28]

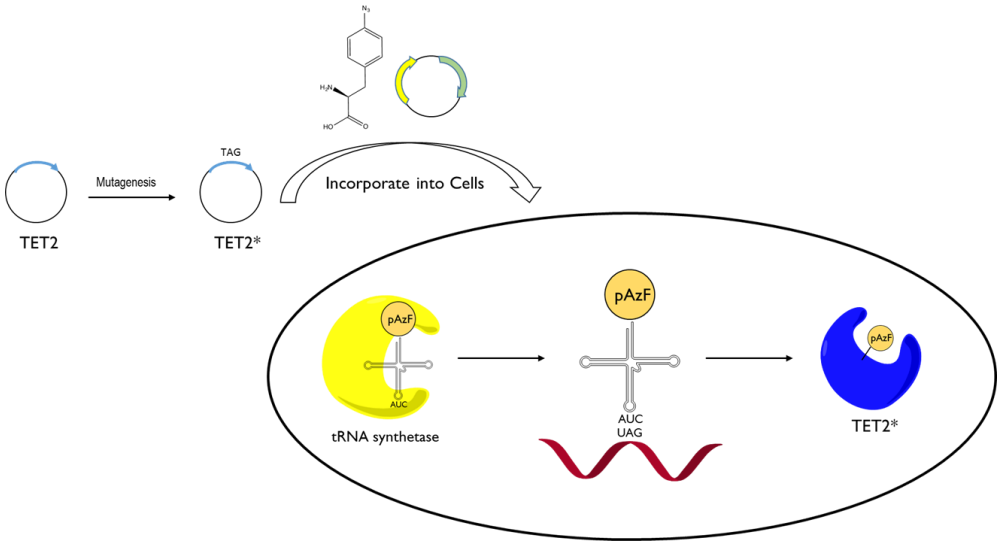


Figure 4. Incorporation of pAzF into TET2

When exposed to ultraviolet (UV) light, the azide in pAzF is activated, inducing crosslinking of the enzyme to the DNA substrate via a reactive nitrene intermediate (Figure 5). [26–28] Once the DNA substrate has been captured, the TET enzyme and its substrate can be pulled

down using an anti-TET antibody. Bound DNA is then sequenced to determine which genes are direct TET targets.[27]

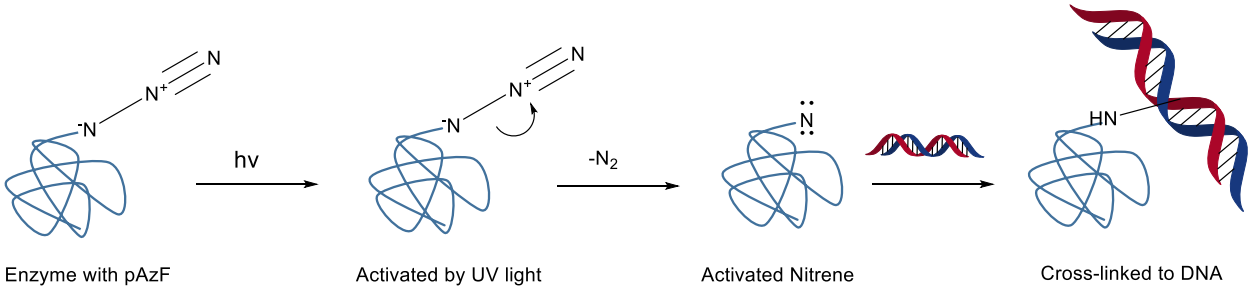


Figure 5. Cross-Linking Mechanism

2.2 DESIGN, EXPRESSION, AND VALIDATION OF pAzF MUTANTS

The photo-crosslinkable UAA pAzF was strategically introduced into TET2 by replacing a native residue within close proximity to the bound 5mC-containing substrate. The crystal structure of TET2 with the inhibitor N-oxalylglycine (NOG) and methylated cytosine DNA substrate was analyzed to gain insight into which residues could be replaced without disrupting the enzyme–substrate interactions (PDB: 4NM6, Figure 6A).[5] Because the DNA binds across the active site with the 5mC flipped into the enzymatic pocket, amino acids shown to be critical for catalytic activity were avoided to maintain native-like enzymatic efficiency.[5] Additionally, for the covalent linkage to be formed, the positioning of the pAzF needed to be in close proximity to the substrate DNA in order to facilitate covalent bond formation. In selecting specific residues for mutagenesis, each residue was considered for its proximity to the DNA binding sites, possible role in DNA binding, and role in catalytic activity. From this analysis, fourteen amino acids were selected for site directed mutagenesis, replacing the natural amino acid with an amber codon, TAG, resulting in a nonsense mutation (Figure 6A).

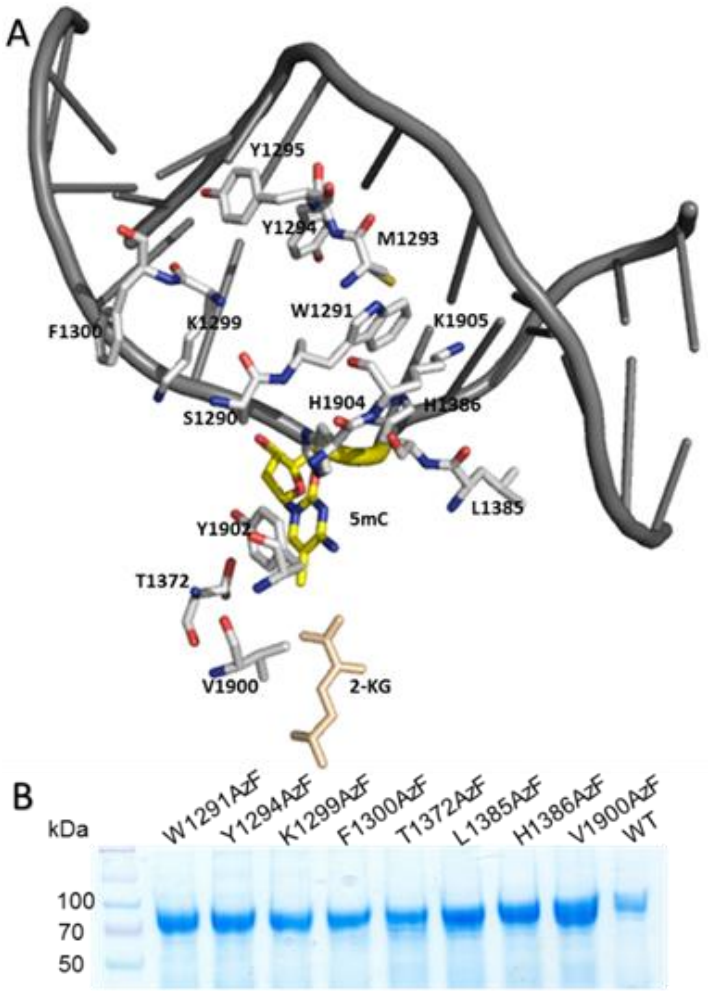


Figure 6. Crosslinking Mutants and Expressed pAzF-Containing Mutants[19]

The TAG mutations allow for the incorporation of the UAA, pAzF, during protein expression in the presence of an orthogonal aminoacyl tRNA synthetase and tRNA pair.[26] All fourteen TET2 mutant proteins were expressed in the presence of pAzF in *Escherichia. Coli* and purified. The purity of proteins was determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 6B). The inclusion of the pAzF in all mutant proteins was confirmed through in-gel fluorescence (Figure7).

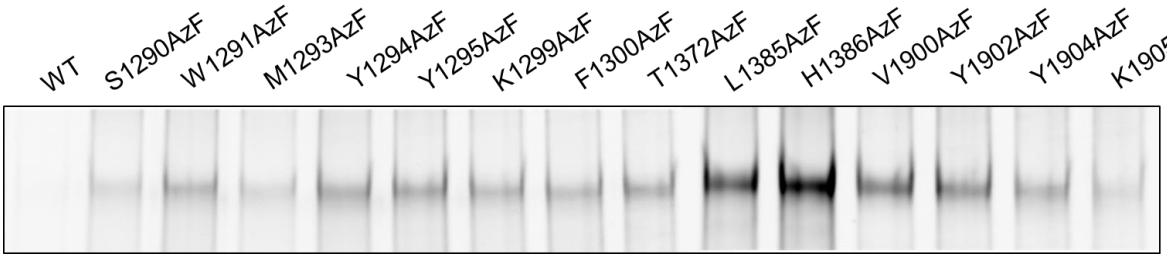


Figure 7. In-Gel Fluorescence of pAzF TET2 Enzymes

Copper-catalyzed click chemistry was used to couple the azide-containing protein to a fluorophore probe. The clicked protein was separated by SDS-PAGE and visualized. The alkyne fluorophore used for click chemistry was synthesized from a piperazine rhodamine derivative to form an alkynl amide (Figure 8).

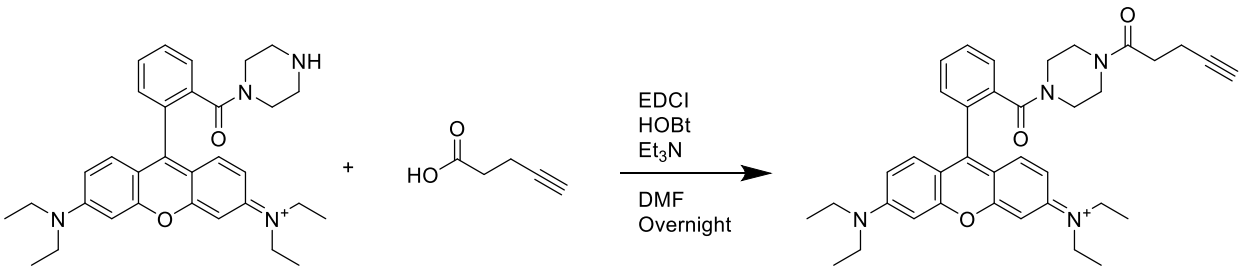


Figure 8. Synthesis of Alkynated Fluorophore

2.3 ACTIVITY AND CROSSLINKING ASSAYS

To test our ability to pull down TET gene targets, we tested our photo crosslinking assay *in vitro* with synthesized DNA substrates. Initial activity screenings were carried out with an eight nucleotide DNA oligo with the following sequence: 5'–CAC–mCGG–TG–3', which was previously optimized in the Islam lab by Dr. B. Sudhamalla and S. Salazar. For crosslinking experiments, the previous eight nucleotide sequence was extended by four nucleotides to accommodate a 5' biotin tag and maintain self-complementarity. The unmethylated biotin DNA

substrate was synthesized with the following sequence: 5'–Biotin–CAC–ACC–GGT–GTG–3'. The methylated cytosine sequence was: 5'–Biotin–CAC–ACmC–GGT–GTG–3'. DNA was purified and verified by Matrix Assisted Laser Desorption/Ionization – Time of Flight (MALDI-TOF) mass spectrometry. Single-stranded DNA was then annealed to give the double-stranded DNA substrates used in subsequent assays.

Once the mutants were generated and the recombinant protein was purified, it was necessary to test their activity, as active mutants should have a longer residence time in the catalytic pocket, and thus would be more likely to successfully crosslink. The activity assay was developed in the Islam lab (unpublished results) and was used to test mutant proteins for activity. In order to quantify the oxidation efficiency via MALDI mass spectrometry, a standard curve was created from the analysis of a series of mixtures of the synthesized 5mC and 5hmC DNAs of known concentrations (Figure 9). Regression analysis indicates that the ionization potentials of the two molecules are equivalent by MALDI and thus relative quantities can be read from the mass spectrometry data.

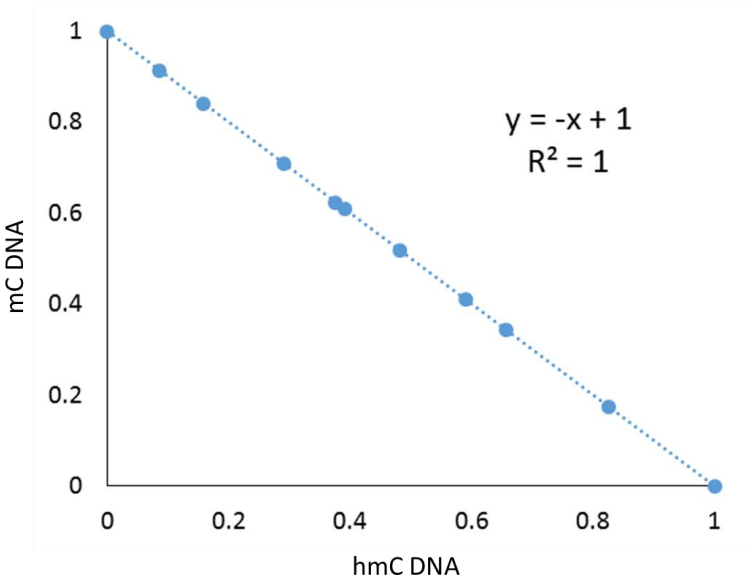


Figure 9. Standard Curve of 5mC vs. 5hmC DNAs

All fourteen TET2 pAzF containing mutants were tested for activity against the native eight nucleotide doubled-stranded DNA with symmetrical 5mC-containing CpG nucleotides. Nine mutants showed activity comparable to wild type activity (Figure 10). These nine mutants were further tested with the biotin-tagged DNA substrate, where all showed comparable levels of activity.

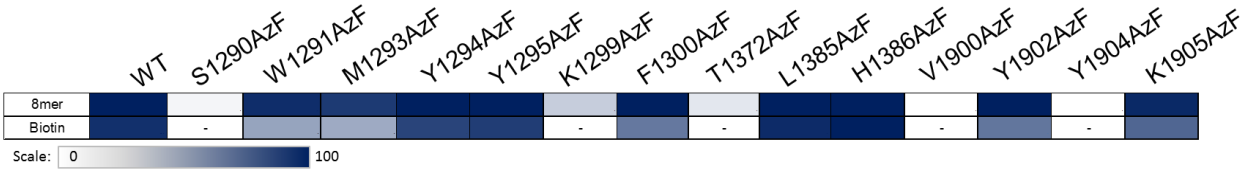


Figure 10. Activity Assay Data for AzF Mutant Proteins

All pAzF-containing enzymes were tested for crosslinking with symmetrically methylated DNA. Successfully crosslinked mutants for TET2 included Y1294AzF, H1386AzF, and Y1902AzF (Figure 11); no inactive mutants successfully crosslinked with methylated DNA substrates. H1386AzF, the strongest crosslinking mutant, was also tested with non-methylated substrates. Relative quantification indicated that H1386AzF was shown to crosslink 11.5 times more strongly to methylated cytosine substrates than to unmodified cytosine substrates (Figure 11). These data indicate that crosslinking is selective for 5mC substrates over unmodified cytosine.

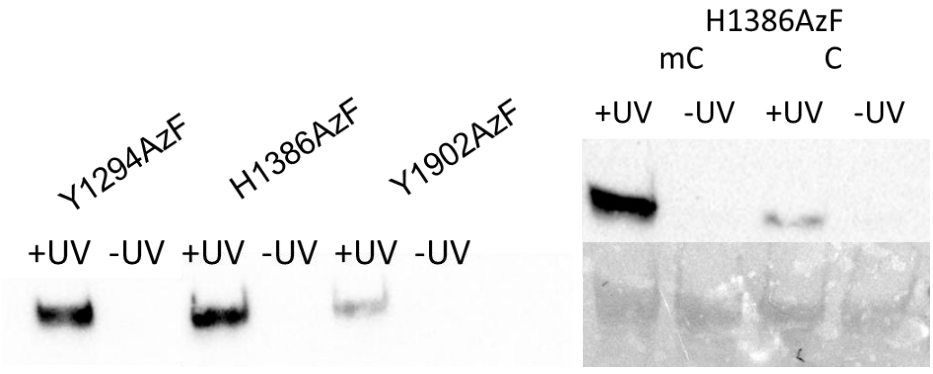


Figure 11. Successfully Crosslinked Mutants

2.4 FUTURE DIRECTIONS

To expand this proof of concept to the other TET family members, the equivalent catalytic domain crosslinking mutants for TET1 and TET3 will be generated and tested *in-vitro*. With a comprehensive set of *in-vitro* data, this crosslinking system will then be applied in cells. Given the scope of TET influence in development, pluripotency, and somatic cells, different cell types, including mouse ESCs or induced pluripotent stem cells can be utilized for in-cell experiments. To carry out crosslinking experiments in cells, first full-length constructs will be transfected into cells. Cells will be lysed and soluble proteins will be separated by SDS-PAGE and analyzed by western blotting with enzyme or tag-specific antibodies to validate that the expression system is working *in vivo*. Next, the equivalent amber codon containing crosslinking mutants will be generated in the full-length constructs, transfected, and expressed *in vivo* in the presence of pAzF. Once the in-cell protein expression systems are optimized to ensure significant over-expression of the target enzyme in cells, crosslinking will be induced to activate the incorporated azide to crosslink to genomic DNA. Cells will then be collected and lysed; crosslinked proteins and DNA will be pulled down and subjected to genomics analysis. After sequencing pulled-down DNA and identifying gene targets, selected targets will be validated. Additional data will be collected by carrying out a parallel analysis via ChIP-seq. Taken together, genomics data and subsequent pathway analysis can be used to shed light on the role of TET gene regulation on broader biological processes. From these analyses, specific gene targets can be identified and probed via additional experiments to determine specific gene roles in biological contexts.

3.0 SPECIFIC AIM 2: DEVELOPMENT OF DNMT-TET TANDEM SYSTEM TO FUNCTIONALIZE DNA SUBSTRATES

3.1 INITIAL SCREENING OF MODIFIED DNA SUBSTRATES

Another approach to determining the specific localization of TET activity on chromosomal DNA is through the incorporation of non-canonical DNA modifications to create a "bumped" DNA substrates which can be further functionalized. Protein methyltransferases, such as G9a, have been previously engineered with SAM analogs to install non-canonical modifications at methylation sites on proteins to identify novel substrates of specific enzymes.[29] Our system will establish this ability with DNMT and DNA substrates. We expect that DNMTs are capable of installation of unnatural DNA modifications along chromosomal DNA, which can be subsequently oxidized by TET enzymes. This oxidized DNA can then be further selectively oxidized or functionalized, followed by subsequent imaging or pull-downs (Figure 12).

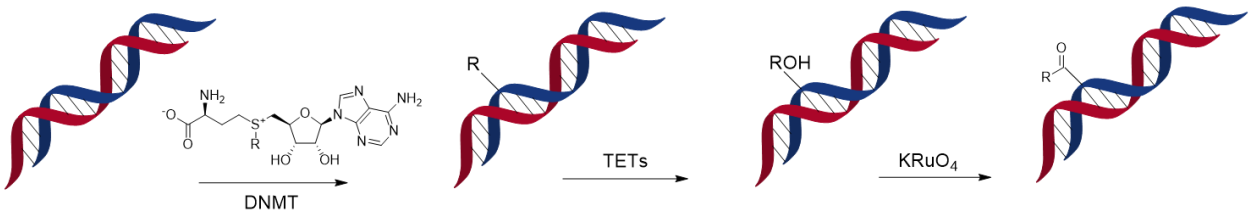


Figure 12. Overview of Proposed DNMT-TET Tandem System

Because the activity of TET enzymes towards unnatural cytosine modifications was unknown, synthetic DNA containing allyl, ethyl, or propyl-substituted cytosines, synthesized by

Dr. D. Dey of the Islam Lab, in place of methyl-substituted cytosines, were screened *in vitro* for oxidation activity against the TET2 wild type enzyme (Figure 13). The wild type enzyme fully converted methylated and ethylated substrates to oxidized products, but not the allylated or propylated substrates (Figure 14). To find an enzyme that would efficiently oxidize these larger allylated and propylated substrates, mutant enzymes with extended active sites (EASs) were employed (Figure 15).

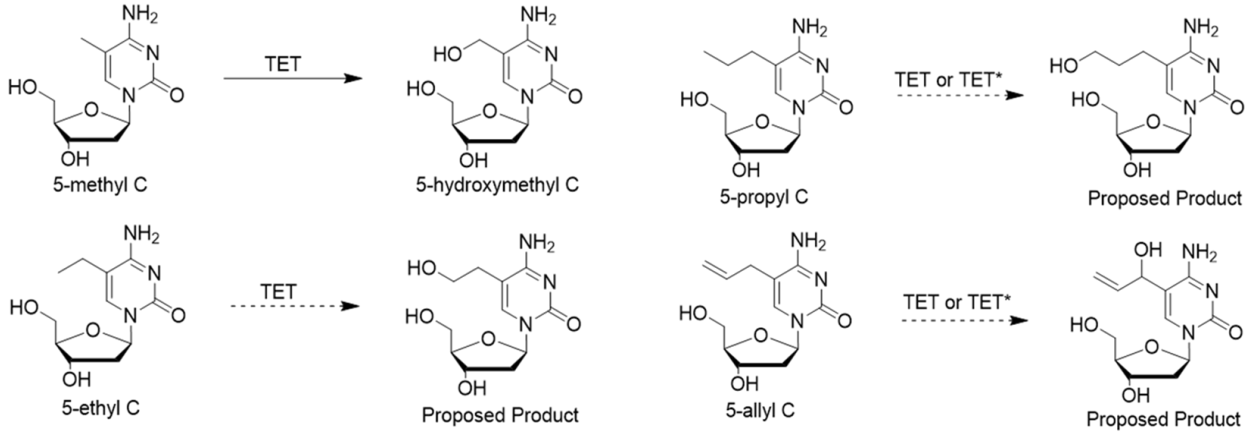


Figure 13. Methyl, Allyl, Ethyl, and Propyl DNA Substitutions

A total of forty-two mutants, the majority of which were previously expressed and purified by Dr. B. Sudhamalla, of the Islam lab, were screened with methyl, allyl, ethyl, and propyl-containing DNA to determine oxidation efficacy via activity assay. Samples were analyzed by MALDI-TOF mass spectrometry, and seven mutants were found to oxidize these substrates at around 20-40% efficacy (Figure 14). Although some mutant enzymes showed oxidative activity towards the allyl and propyl substrates, MALDI data was unclear as there were significant salt patterns for both starting material and the oxidized product in the spectra.

Mutant	Methyl	Allyl	Ethyl	Propyl
WT				
Neg				
T1372A				
T1372G				
T1372S				
C1374A				
C1374G				
A1379G				
N1387A				
T1393A				
T1393G				
T1393S				
V1395A				
V1395G				
T1397A				
T1397G				
H1416A				
H1416G				
L1872A				
L1872G				
T1883A				
T1883G				

Mutant	Methyl	Allyl	Ethyl	Propyl
T1883S				
S1898A				
S1898G				
V1900A				
V1900G				
Y1902A				
Y1902F				
Y1902G				
Y1902S				
Y1902T				
Y1902V				
A1379G T1883S				
T1393A V1395A				
T1393G V1395A				
T1393S V1395A				
T1393S V1395G				
V1395G T1883S				
V1900A V1395A				
V1395G T1883S A1379G				
V1395G T1393S A1379G				
V1395G T1393S T1883S				
V1395G T1883S T1393S A1379G				

Scale: 0 100

Figure 14. Screening of TET2 EAS Mutants with Methyl, Allyl, Ethyl, and Propyl DNA

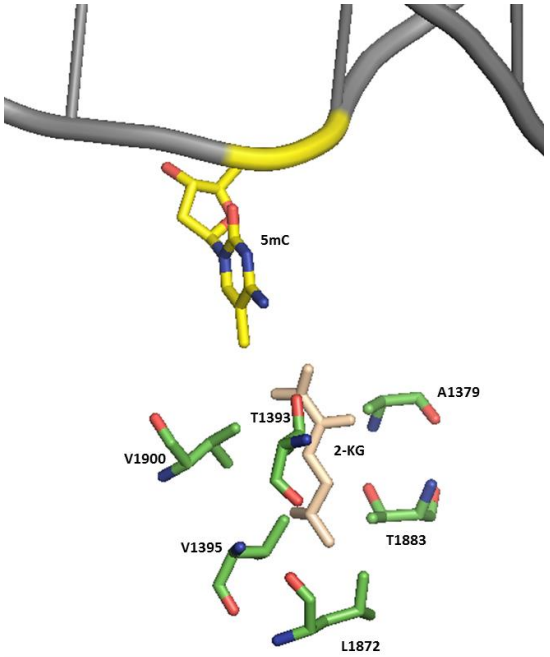


Figure 15. Positions of Most Active TET2 Mutants (PDB: 4NM6)

3.2 ASSAY OPTIMIZATION

To increase the efficacy of these *in vitro* oxidations, a buffer screening was performed with the most active mutants (Figure 15), using Tris buffer in place of the previous HEPES pH 8.0 buffer. The use of Tris buffer in place of HEPES buffer showed a drastic drop in oxidation activity. Additional screening involved testing three pH values within the HEPES buffering capacity: 7.0, 7.5, and 8.0. A pH of 7.5 showed increased activity for both the wild type enzyme and mutants tested (Figure 16), and this condition was then used in future assays.

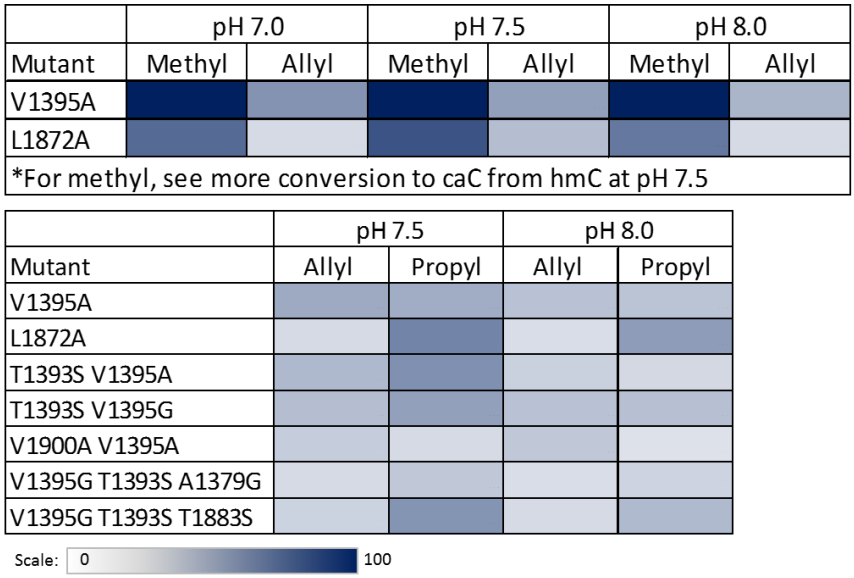


Figure 16. pH Optimization Assay with Most Active Mutants

In MALDI spectra, significant salt patterns were observed, which potentially obscured oxidation products and affected quantification. To improve spectral quality and overall clarity of the data, the sodium chloride concentration was decreased by 100mM in the assay buffer for the wild type TET2 with methylated DNA. With these assay conditions showing no loss of activity, the sodium chloride component of the assay buffer was removed to screen all four DNA substrates with wild type and mutant proteins. These new conditions showed increases in oxidation,

particularly for the propyl substrate, with some mutants able to fully convert this DNA to the oxidized form (Figure 16). Interestingly, the wild type enzyme also shows a clear ability to oxidize allyl and propyl DNA under these reduced ionic conditions

3.3 FUTURE DIRECTIONS

All DNA substrates are being re-screened with all mutants and the wild type enzymes to generate a complete data set for these lesser salt conditions. After completed screening to indicate oxidation capacity of TET enzymes with regard to these modifications, the structure of the enzymatically oxidized substrates will be characterized. DNA substrates will be digested and purified to determine the specific location of oxidation by NMR. Oxidized DNA substrates will be further functionalized using potassium perrutenate, which has been shown to selectively oxidize 5hmC in DNA in the process of oxidative bisulfite sequencing developed by the Balasubramanian lab.[30] The aldehyde, ketone, or Michael-acceptor oxidized cytosine-containing DNA will then be reacted with a probe or pull-down substrate such as a hydroxylamine fluorophore or a biotinylated compound to allow for direct visualization of the effects of TET-specific enzymatic activity of along the genome or for pull-down applications. To develop a full proof of concept model for a DNMT-TET tandem system, the installation of unnatural aliphatic groups by DNMT3 will be tested and confirmed by mass spectrometry. Recombinant DNMT3 will be expressed, purified, and tested with SAM cofactor analogs[29] previously synthesized in the Islam lab, to determine the ability of DNMT3 to install ethyl, propyl, and allyl modifications on DNA substrates.

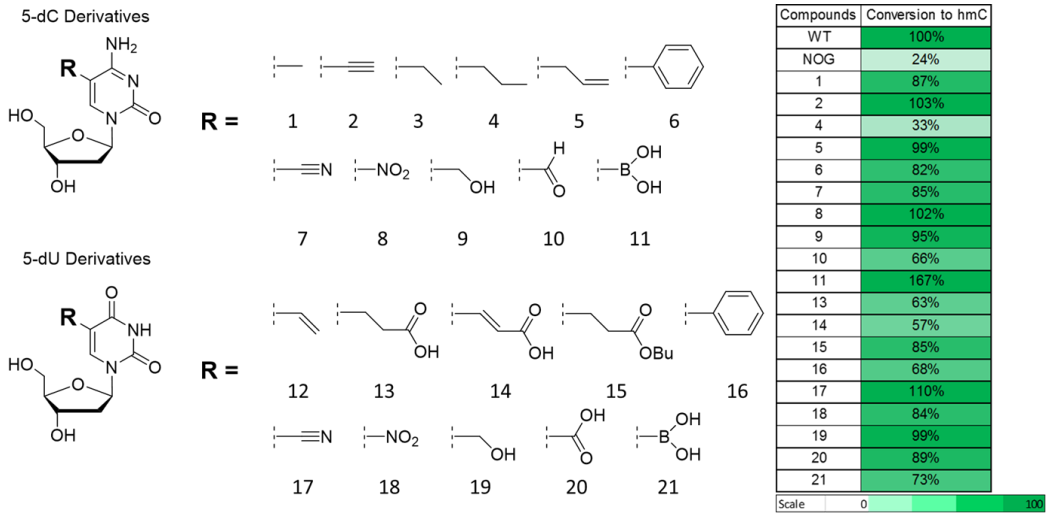
4.0 SPECIFIC AIM 3: TET-SPECIFIC INHIBITOR DEVELOPMENT

4.1 PRELIMINARY INHIBITION SCREENINGS

A small molecule inhibitor specific to TET family enzymes has yet to be discovered. The use of small molecules allows us to perturb biological systems by targeting individual enzyme families with highly resolved temporal control. While inhibitors such as NOG and 2-hydroxyglutarate (2-HG) are shown to be potent inhibitors of TETs, they are competitive inhibitors in the 2-KG binding pocket of other 2-KG dependent enzymes, such as such as histone demethylases, KDM4s, and RNA demethylase, FTO.[3,5,25,31] This redundancy poses a significant challenge to studying such enzymes and their critical roles in development and disease.[32] Traditional knockout studies cannot elucidate the depths of TET activity and significance due to the increase in fetal mortality and developmental defects and delays. This makes the temporal control and reversibility offered by a small molecule inhibitor critical to fully elucidate the functions of these enigmatic enzymes.

To address the goal of developing a potent and selective inhibitor for TET proteins, a library of nucleotide derived compounds available in the Islam laboratory were screened for inhibition potential against the wild type TET2 protein using a MALDI-based assay (Figure 17). As the native substrate of TET2 is methylated cytosine, it is expected that these nucleotide derivatives would competitively inhibit the binding of 5mC containing DNA by interacting with

the active site of the enzyme at the position typically occupied by the methylated cytosine.[5,31,33] This would lead to a lack of oxidation of the methylated substrate, and thus a lack of DNA demethylation activity.[34] Promising compounds from this preliminary screening will be further developed using a fragment-based approach to develop a specific and potent inhibitor for TET proteins. The compounds that are identified as weak inhibitors will be linked with another logical fragment, such as NOG, 2-KG, or succinate, to make a more potent and selective inhibitor.



Compounds	Conversion to hmC
WT	100%
NOG	24%
1	87%
2	103%
4	33%
5	99%
6	82%
7	85%
8	102%
9	95%
10	66%
11	167%
13	63%
14	57%
15	85%
16	68%
17	110%
18	84%
19	99%
20	89%
21	73%

Figure 17. Screening Library and Heat Map of Inhibition Data

The initial screening was carried out with a library of compounds previously synthesized in the Islam lab by Dr. D. Dey. The compounds are cytosine and uracil derivatives, which are expected to mimic the methylated cytosine substrate and occupy the DNA binding site to perturb the binding of the native 5mC substrate, thus inhibiting the enzymatic oxidation of 5mC to 5hmC. The screening was carried out using the eight nucleotide symmetrical methylated cytosine containing DNA strand described above. The enzyme and inhibitor compound were pre-incubated together before the addition of the 2-KG cofactor and DNA substrate. These screenings show single fragment inhibition in the 1-5 mM range, with the propylated compound,

Compound 4, indicating the strongest inhibition via the least conversion of 5mC to 5hmC. Preliminary screening results indicate multiple compounds showing some inhibition activity at 5mM concentrations (Figure 17), via less conversion to oxidized 5hmC relative to the uninhibited wild type enzyme.

4.2 DESIGN OF LINKED INHIBITORS

In order to be a viable small molecule inhibitor, the proposed compounds must have significant potency, ideally in the nano to pico molar range. To achieve this, the initial approach was altered to employ the principles of fragment based inhibitor design and consideration of transition state analogs. We thus designed prototype linked inhibitor compounds that would incorporate aspects of both the DNA substrate and the 2-KG cofactor (Figure 18). Combination of the active DNA substrate and 2-KG cofactor-like structure as one inhibitor will afford an additive enthalpic effect without inducing new entropic barriers. This will lower the overall Gibbs free energy for the system, leading to a more favorable inhibition.

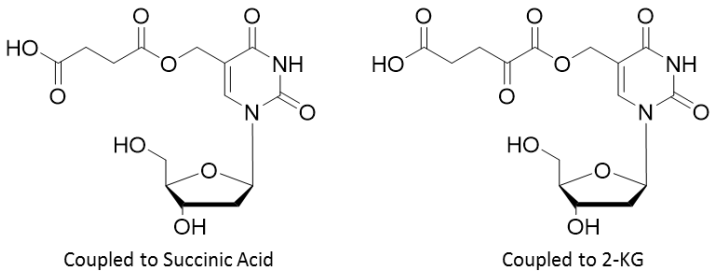


Figure 18. Prototype Linked Inhibitors

4.3 FUTURE DIRECTIONS: PROPOSED INHIBITOR SYNTHESIS

The first attempts at syntheses were with unprotected hydroxymethylated deoxyuridine that had been synthesized in the Islam lab by Dr. S. Wang using known methods.[35] Due to the presence of three free hydroxyl moieties in this compound, the high polarity made purification difficult. Using conditions previously used by Dr. D. Dey in the Islam Lab to selectively acetylate the hydroxymethyl, coupling with succinic acid, and 2-ketoglutarate was attempted. These reactions produced significantly complex reaction mixtures that were more difficult to purify than the preceding step, limiting their utility as a broader approach to build a library of inhibitory compounds. From these reactions, it was determined that protection of the 3' and 5' hydroxyl groups was necessary for an efficient and more broadly applicable synthesis.

To synthesize generalized linked inhibitor compounds, two routes are proposed, one beginning with deoxyuridine (Figure 19), and the other with deoxythymidine (Figure 20). Both routes ultimately arrive at tethered compounds that mimic the nucleotide and the cofactor binding in the catalytic site of TET enzymes and can be generalized to yield a library of analogs.

The first step of the deoxyuridine route (Figure 19) employs a cyclic protection, tethering both the 3' and 5' hydroxyl groups together with a single silyl ether.[36] This is followed by a Baylis-Hillman reaction with formaldehyde to install the hydroxymethyl at the five position on the deoxyuridine.[35] Esterification with succinic acid, NOG, or 2-KG, among other possible compounds, followed by deprotection will yield the final linked compounds.

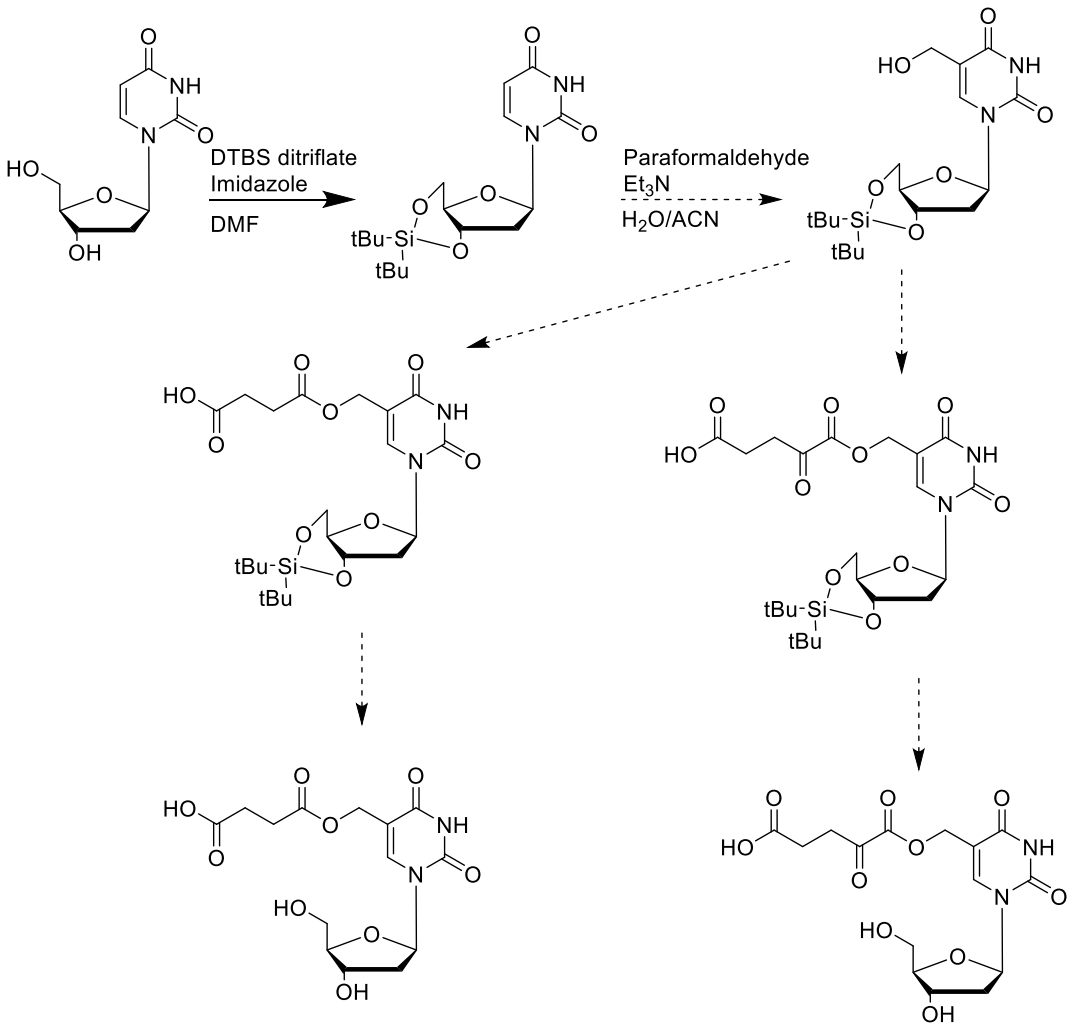


Figure 19. Synthesis via Deoxyuridine

The first step of the deoxythymidine route (Figure 20) is TBS protection of both free hydroxyl groups,[37] followed by bromination of the methyl position.[37] Substitution with a carboxylate and subsequent deprotection will yield the final linked compounds in this route. Once synthesized, the inhibitor compounds will be tested for inhibition activity *in vitro*, and IC_{50} values will be determined for strong inhibitors. Promising compounds will be further developed into cell-permeable compounds and applied in cell experiments.

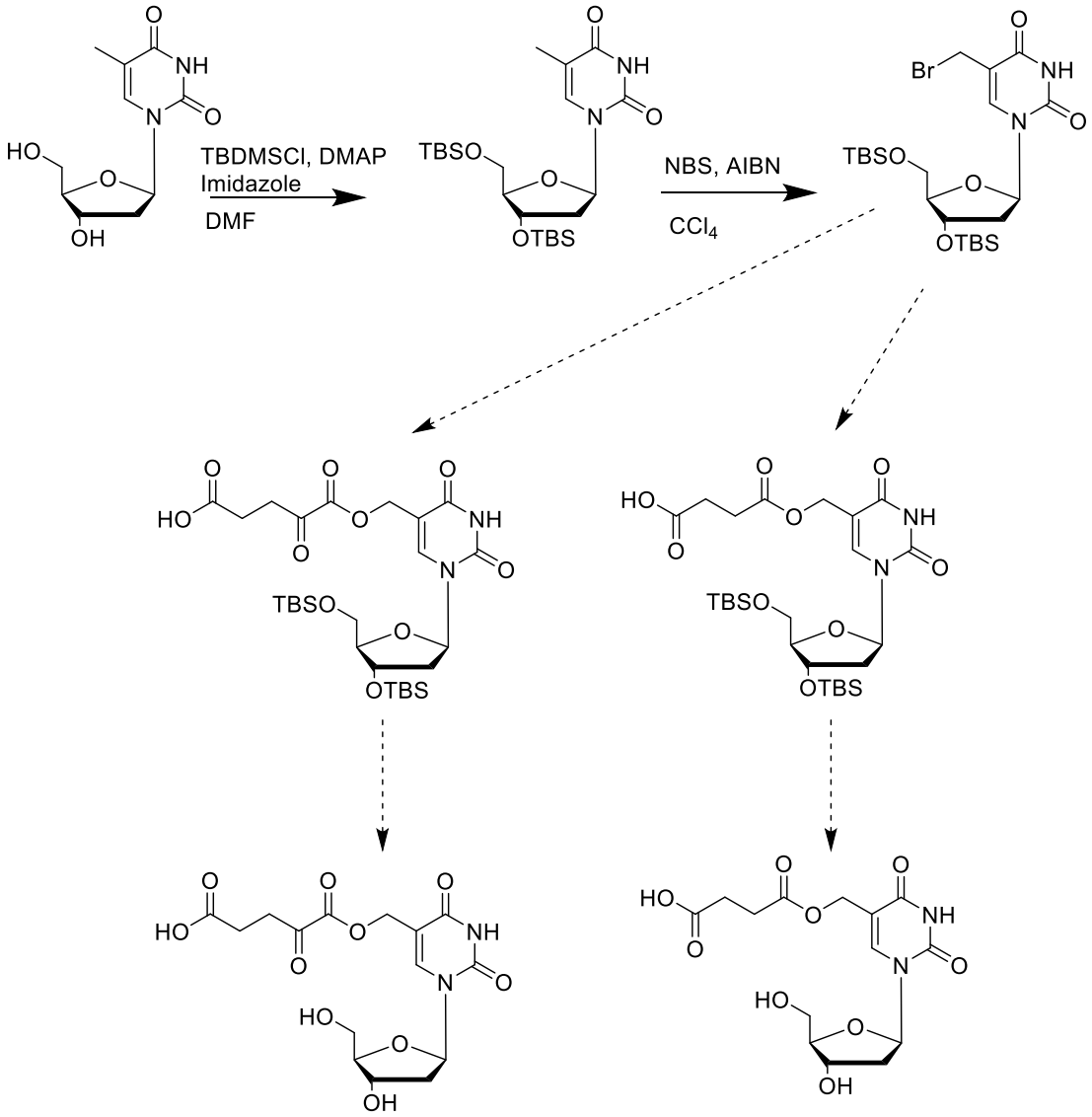


Figure 20. Synthesis via Deoxythymidine

5.0 SUMMARY

In summary, as a proof of concept, recombinant TET2 pAzF containing mutants have been designed, expressed, and purified. TET2 Y1294AzF, H1386AzF, and Y1902AzF were shown to crosslink to methyl cytosine containing DNA substrates. TET crosslinking system will be generalized across family members and applied in cells. Crosslinking to chromosomal DNA will identify novel gene targets of TET enzymes. Additionally, ethyl, allyl and propyl substituted DNA substrates were screened for optimal activity with TET2 WT and mutant enzymes. Under optimized conditions, propylated substrates were oxidized most readily. These substrates will be selectively further oxidized and tagged to allow for visualization of specific genomic targets of TET enzymes. Lastly, a library of potential inhibitor compounds was screened against the TET2 WT enzyme. Few compounds indicated slight inhibitory activity at 5mM concentrations. Based on this result, a fragment based approach was employed to design linked inhibitor compounds, which will be synthesized and tested for inhibition activity against TET enzymes.

6.0 EXPERIMENTAL

Mutagenesis and Expression of TET2 and its variants. The C-terminal 6xHis-tagged TET2 catalytic domain bacterial expression construct pPEI-TET2 was obtained from Yanjui Xu and evolved Methanococcus Jannaschii p-Azido-L-phenylalanine RS (2 copies+tRNA) expression vector pEVOL-pAzF (Addgene ID: 31186) were obtained from Addgene.[27] The wild type TET2 plasmid was transformed into E. Coli BL21 (DE3) competent cells (Invitrogen) using pPEI kanamycin-resistant vector. A single colony was picked up and grown overnight at 37^{o}C in 10 mL of Luria-Bertani (LB) broth in the presence of 50 µg/mL kanamycin (Kan). The culture was diluted 100-fold and allowed to grow at 37^{o}C to an optical density (OD600) of ~0.8, and protein expression was induced overnight at 17^{o}C with 0.1 mM IPTG in an Innova 44® Incubator shaker (New Brunswick Scientific). Proteins were purified as follows: harvested cells were resuspended in 15 mL lysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, 25 mM imidazole, Lysozyme, DNase, and Roche protease inhibitor cocktail). The cells were lysed by pulsed sonication (Qsonica-Q700), and centrifuged at 13000 rpm for 40 min at 4^{o}C. The soluble extracts were subject to Ni-NTA agarose resin (Thermo) per manufacturer's instructions. After rinsing with washing buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 25 mM imidazole), proteins were eluted with a buffer containing 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 400 mM imidazole. Proteins were further purified by size

exclusion chromatography (Superdex-200) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 10% glycerol. Purified proteins were concentrated using Amicon Ultra-10k centrifugal filter device (Merck Millipore Ltd.). The protein concentration was determined using Bradford assay kit (BioRad Laboratories) with BSA as a standard. The concentrated proteins were stored at -80ºC before use. TET2 EAS variants were generated using the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies). The resulting mutant plasmids were confirmed by DNA sequencing. TET2 EAS variants were expressed and purified as stated for TET2-WT.

The wild-type TET2 gene was mutagenized using the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) to generate a series of clones with an Amber suppressor codon (TAG) at positions S1290, W1291, M1293, Y1294, Y1295, K1299, F1300, T1372, L1385, H1386, V1900, Y1902, and K1905. The resulting mutant plasmids were confirmed by DNA sequencing. To express TET2 variants carrying p-Azido-L-phenylalanine (pAzF) at specific positions, we co-transformed E. Coli BL21 star (DE3) cells (Invitrogen) with pEVOL-pAzF and TET2 Amber Variants S1290, W1291, M1293, Y1294, Y1295, K1299, F1300, T1372, L1385, H1386, V1900, Y1902, and K1905. Cells were recovered in 200µL of SOC medium for 1 h at 37ºC before being plated on an LB agar plate containing Kan (50 µg/mL) and chloramphenicol (Chl) (35 µg/mL). A single colony was selected and grown overnight at 37ºC in 10 mL LB broth in presence of 50 µg/mL Kan and 35 µg/mL Chl. The overnight culture was centrifuged for 10 min at 1000g, and 8 mL of LB broth was subsequently removed. The cell pellet was resuspended in the remaining 2 mL LB broth and was used to inoculate 1L of GMML medium supplemented with 50 µg/mL Kan and 35 µg/mL Chl.[38] Cells were allowed to grow at 37ºC in an incubator shaker (225 rpm) until OD600 reached ~0.8. pAzF was added to the

culture to a final concentration of 1 mM and the culture was cooled to 17ºC for 30 min while shaking at 225 rpm. The culture was then induced with 0.05% arabinose and continued to shake at 17ºC for an additional 30 min. Finally, the protein expression was induced by the addition of 0.25 mM IPTG and allowed to shake at 225 rpm for 20 h at 17ºC. All the mutant proteins were purified as described above for the wild type TET2 protein.

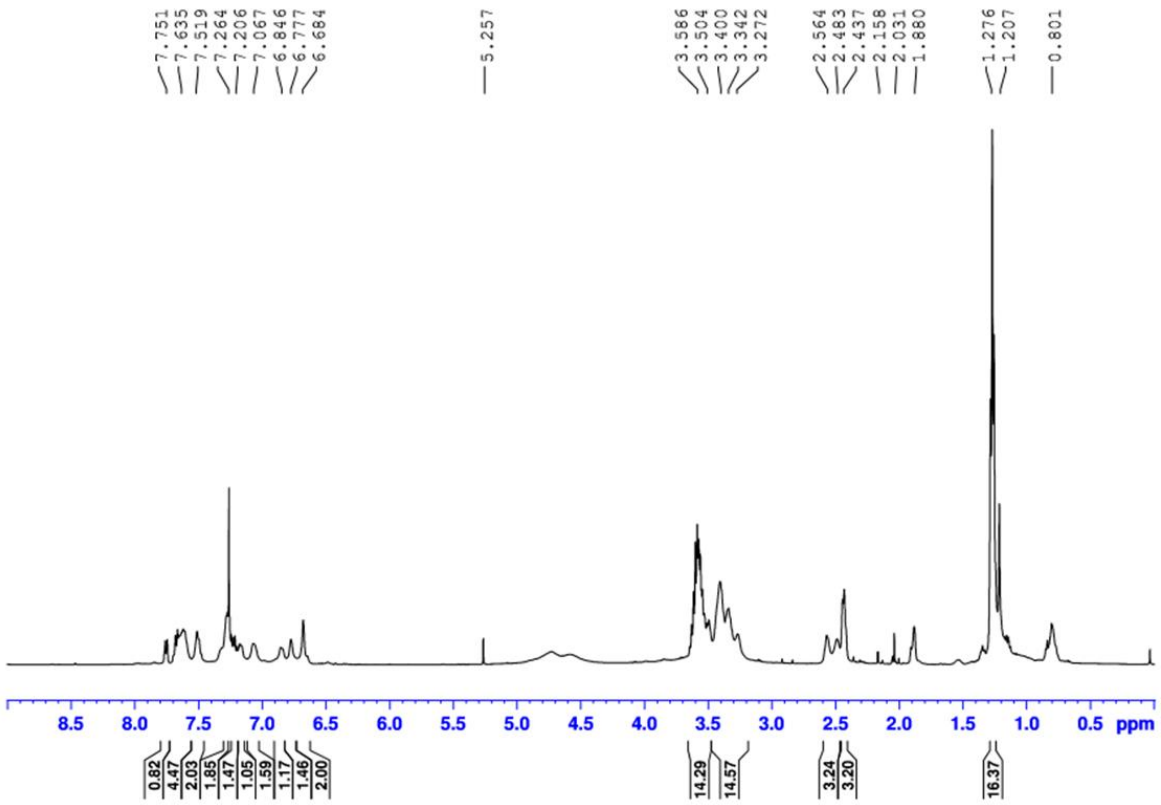
TET2 Activity Assay. To monitor enzymatic activity, a demethylation activity assay was utilized and activity was observed through MALDI-TOF. Each DNA demethylation assay sample included 10 µM enzyme, 10 µM DNA, 1 mM 2-KG, 100mM NaCl, 1mM ATP, 1mM DTT, 50 mM HEPES pH 8, 100 µM $(NH_4)_2Fe(SO_4)_2$, and 2 mM L-ascorbic acid with a total assay volume of 20 µl. Fe (II) and L-ascorbic acid were prepared freshly. The enzyme was added last to the assay sample and briefly centrifuged. The samples were incubated at 37°C for one hour. Samples were denatured at 100°C for 10 minutes and incubated for 10 mins with BT AG® 50W-X8 Resin (BioRad, cat# 143-5441). To observe activity, 1 µL of assay sample was applied to the MALDI plate followed by 1 µL of 3HPA matrix. The sample was analyzed using Voyager on the Reflector Positive mode. The negative control included all components of the assay except for the enzyme. TET2 extended active site assays were performed the same as listed above, except HEPES was pH 7.5, 100mM NaCl was omitted, and samples were incubated for 3 hours.

Copper Catalyzed Click Chemistry. Each click reaction included 1 mM CuSO4, 100 µM TBTA, 2 mM Na-Ascorbate, 100 µM Alkyne, and 5µg enzyme. Ascorbate solutions were prepared freshly. Samples were centrifuged and incubated for 3 hours at 37°C at 400 rpm in a ThermoTop ThermoMixer C (Eppendorf). Samples were separated by SDS-PAGE and visualized on a ChemiDoc MP Imaging system using TAMRA fluorophore excitation

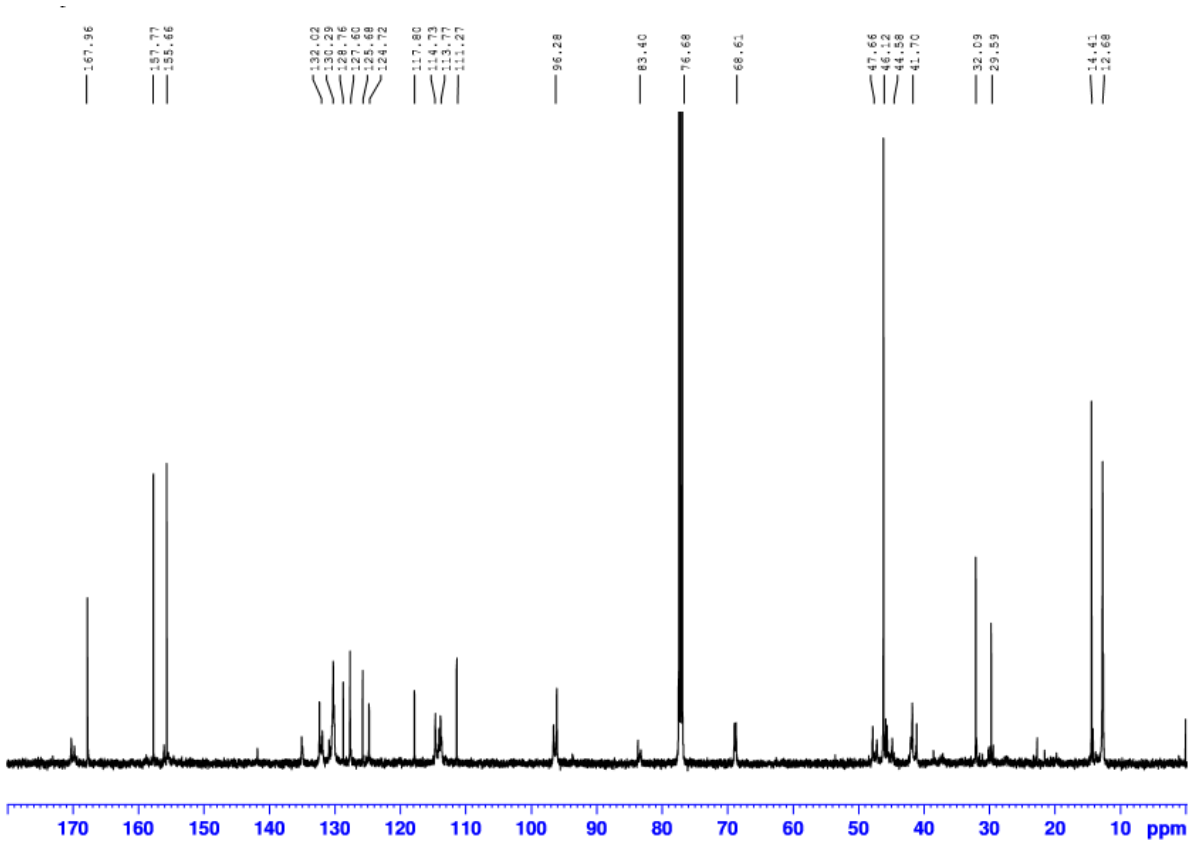
wavelength (Emission filter 605/50, Light: green Epi illumination). The gel was subsequently stained with coomassie brilliant blue staining solution to confirm the presence of protein in all the samples.

Synthesis of N-(6-(diethylamino)-9-(2-(4-(pent-4-ynoyl)piperazine-1-carbonyl)phenyl)-3H-xanthen-3-ylidene)-N-ethylethanaminium. A mixture of Rhodamine B piperazine amide (50.3mg, 98.3umol), 4-Pentynoic acid (14.38 mg, 146.57 µmol, 1.5eq), and EDCI (33.6 mg, 216.4 µmol, 2eq) was dissolved in anhydrous DMF (1 mL). Triethylamine (40.80 µL, 293.15 µmol, 3eq) was added and the solution stirred at rt overnight. The reaction mixture was diluted with DCM (75 mL), washed with HCl (1M, 3x25 mL) and brine (25 mL), dried over Na_2SO_4, filtered, and evaporated *in vacuo*. The residue was purified by Silica gel chromatography (20:1 DCM:MeOH) to yield the final product (38.5 mg, 66.6%) as a deep purple solid.

^{1}H NMR (CHLOROFORM-*d*, 500MHz):



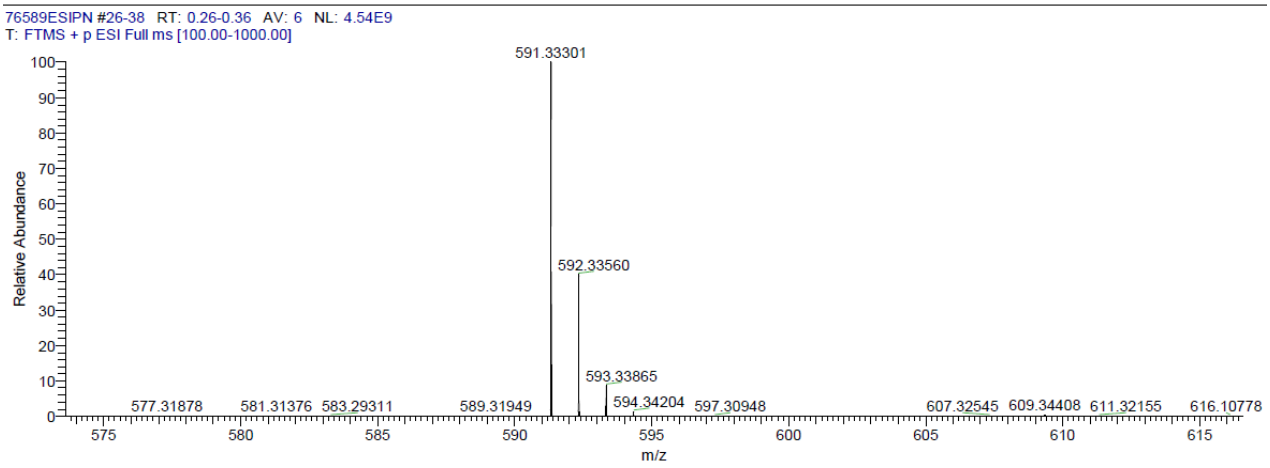
^{13}C NMR (CHLOROFORM-*d*, 125MHz):



HRMS: m/z as $[M+H]^+$ calculated for $C_{37}H_{43}O_3N_4$, found 591.33301

76589ESIPN#27-37 RT: 0.26-0.36 AV: 6
T: FTMS + p ESI Full ms [100.00-1000.00]
m/z= 590.00000-593.00000

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
591.33301	4569678336.0	100.00	591.33297	0.07	$C_{37}H_{43}O_3N_4$



Crosslinking Assay. Each assay included 10 µM enzyme, 10 µM DNA, 100 µM 2-KG, 100mM NaCl, 1mM ATP, 1mM DTT, 50 mM HEPES pH 8, 100 µM $(NH_4)_2Fe(SO_4)_2$, and 2 mM L-ascorbic acid with a total assay volume of 20 µL. Fe (II) and L-ascorbic acid were prepared freshly. Enzyme and DNA substrate were aliquoted, and buffer solutions with 2-KG were added last. Samples were centrifuged briefly and incubated on ice in the presence or absence of UV light for 20 mins (365 nm, VWR Transiluminator). For Western Blotting, assay samples were separated by SDS-PAGE and transferred onto a 0.2 µm nitrocellulose membrane (Bio-Rad Laboratories) at a constant voltage of 68 mA for 16 hours at 4°C. Membranes were blocked with 15 mL of TBST buffer (50 mM Tris HCl pH 7.4, 200 mM NaCl, 0.01% Tween-20) with 5% nonfat dry milk for 1 hr. at room temperature with gentle shaking. The blocking buffer was then removed and membranes were washed with 20 mL of TBST buffer. Immunoblotting was performed with 1:500 dilution of primary antibody (Anti-biotin, Pierce) overnight at 4°C. The antibody solutions were removed and membranes were washed three times with TBST buffer. The blots were then incubated with HRP-conjugated secondary antibody Donkey anti-Goat IgG with 5% nonfat dry milk (1:10000 dilution) in TBST for 1.5 h at room temperature. After washing, protein bands were visualized by chemiluminescence using VISIGLO HRP Chemiluminescent substrates A and B (cat# N252-120ML and N253- 120ML, aMReSCO) following manufacturer's protocol.

Inhibition Assay. Each DNA demethylation assay sample included 5 µM enzyme, 10 µM DNA, 5mM inhibitor, 100 mM 2-KG, 100 mM NaCl, 1 mM ATP, 1 mM DTT, 50 mM HEPES pH 8, 100 µM $(NH_4)_2Fe(SO_4)_2$, and 2 mM L-ascorbic acid with a total assay volume of 20 µl. Fe (II) supplement and L-ascorbic acid were prepared freshly. The enzyme and all buffer components were incubated for 30 minutes on ice before addition of DNA and 2-KG. Samples

were briefly centrifuged and incubated for 3 hours at 37°C. To observe inhibition/activity, 1 µL of assay sample was applied to the MALDI plate followed by 1 µL of 3HPA matrix. The sample was analyzed using Voyager on the Reflector Positive mode.

DNA Synthesis: DNA oligonucleotides were synthesized using standard DNA phosphoramidite monomers (Glen Research) using 1H-tetrazole as the activator reagent in an Expedite Nucleic Acid Synthesis System (PerSeptive Biosystems) with DMT-ON protocol. 5mC phosphoramidite was purchased from Carbosynth and a biotin phosphoramidite was synthesized by Dr. S. Wang according to known procedures.[39] To ensure good coupling, extended five minute coupling was used for 5mC bases and fifteen minute coupling was used to incorporate biotin. The crude oligonucleotide was cleaved from the solid support resin and deprotected by incubating with 30% ammonium hydroxide at room temp for 36 hours. Preliminary purification and DMT deprotection were carried out using Poly PakII purification cartridge (Glen Research) and further purified by HPLC using a C-18 column. Fractions were collected and concentrated by SpeedVac concentrator followed by lyophilization, and re-dissolved in RNAase free water. The quality and purity of synthesized DNA was monitored by MALDI-TOF.

BIBLIOGRAPHY

(1) Allis, C. D.; Jenuwein, T. *Nat. Rev. Genet.* **2016**, *17* (8), 487–500.

(2) Lian, C. G.; Xu, Y.; Ceol, C.; Wu, F.; Larson, A.; Dresser, K.; Xu, W.; Tan, L.; Hu, Y.; Zhan, Q.; Lee, C.; Hu, D.; Lian, B. Q.; Kleffel, S.; Yang, Y.; Neiswender, J.; Khorasani, A. J.; Fang, R.; Lezcano, C.; Duncan, L. M.; Scolyer, R. A.; Thompson, J. F.; Kakavand, H.; Houvras, Y.; Zon, L. I.; Mihm, M. C.; Kaiser, U. B.; Schatton, T.; Woda, B. A.; Murphy, G. F.; Shi, Y. G. *Cell* **2012**, *150* (6), 1135–1146.

(3) Loenarz, C.; Schofield, C. J. *Nat. Chem. Biol.* **2008**, *4* (3), 152–156.

(4) Wu, H.; Zhang, Y. *Cell* **2014**, *156* (1), 45–68.

(5) Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y. *Cell* **2013**, *155* (7), 1545–1555.

(6) Branco, M. R.; Ficz, G.; Reik, W. *Nat. Rev. Genet.* **2012**, *13* (1), 7–13.

(7) 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. *Science* **2009**, *324* (5929), 930–935.

(8) Bagci, H.; Fisher, A. G. *Cell Stem Cell* **2013**, *13* (3), 265–269.

(9) Scourzic, L.; Mouly, E.; Bernard, O. A. *Genome Med.* **2015**, *7* (1).

(10) Shen, L.; Song, C.-X.; He, C.; Zhang, Y. *Annu. Rev. Biochem.* **2014**, *83* (1), 585–614.

(11) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333* (6047), 1300–1303.

(12) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. *Nature* **2010**, *466* (7310), 1129–1133.

(13) 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. *Science* **2011**, *333* (6047), 1303–1307.

(14) Iurlaro, M.; McInroy, G. R.; Burgess, H. E.; Dean, W.; Raiber, E.-A.; Bachman, M.; Beraldi, D.; Balasubramanian, S.; Reik, W. *Genome Biol.* **2016**, *17*, 141.

(15) Pastor, W. A.; Aravind, L.; Rao, A. *Nat. Rev. Mol. Cell Biol.* **2013**, *14* (6), 341–356.

(16) Bachman, M.; Uribe-Lewis, S.; Yang, X.; Williams, M.; Murrell, A.; Balasubramanian, S. *Nat. Chem.* **2014**, *6* (12), 1049–1055.

(17) Lorsbach, R. B.; Moore, J.; Mathew, S.; Raimondi, S. C.; Mukatira, S. T.; Downing, J. R. *Leukemia* **2003**, *17* (3), 637–641.

(18) Messerschmidt, D. M.; Knowles, B. B.; Solter, D. *Genes Dev.* **2014**, *28* (8), 812–828.

(19) Gu, T.-P.; Guo, F.; Yang, H.; Wu, H.-P.; Xu, G.-F.; Liu, W.; Xie, Z.-G.; Shi, L.; He, X.; Jin, S.; Iqbal, K.; Shi, Y. G.; Deng, Z.; Szabó, P. E.; Pfeifer, G. P.; Li, J.; Xu, G.-L. *Nature* **2011**, *477* (7366), 606–610.

(20) Tan, L.; Shi, Y. G. *Dev. Camb. Engl.* **2012**, *139* (11), 1895–1902.

(21) Dawlaty, M. M.; Breiling, A.; Le, T.; Barrasa, M. I.; Raddatz, G.; Gao, Q.; Powell, B. E.; Cheng, A. W.; Faull, K. F.; Lyko, F.; Jaenisch, R. *Dev. Cell* **2014**, *29* (1), 102–111.

(22) Lu, X.; Zhao, B. S.; He, C. *Chem. Rev.* **2015**, *115* (6), 2225–2239.

(23) Thienpont, B.; Steinbacher, J.; Zhao, H.; D'Anna, F.; Kuchnio, A.; Ploumakis, A.; Ghesquière, B.; Van Dyck, L.; Boeckx, B.; Schoonjans, L.; Hermans, E.; Amant, F.; Kristensen, V. N.; Koh, K. P.; Mazzone, M.; Coleman, M. L.; Carell, T.; Carmeliet, P.; Lambrechts, D. *Nature* **2016**, *537* (7618), 63–68.

(24) Ko, M.; An, J.; Bandukwala, H. S.; Chavez, L.; Äijö, T.; Pastor, W. A.; Segal, M. F.; Li, H.; Koh, K. P.; Lähdesmäki, H.; Hogan, P. G.; Aravind, L.; Rao, A. *Nature* **2013**, *497* (7447), 122–126.

(25) Xu, Y.; Xu, C.; Kato, A.; Tempel, W.; Abreu, J. G.; Bian, C.; Hu, Y.; Hu, D.; Zhao, B.; Cerovina, T.; Diao, J.; Wu, F.; He, H. H.; Cui, Q.; Clark, E.; Ma, C.; Barbara, A.; Veenstra, G. J. C.; Xu, G.; Kaiser, U. B.; Liu, X. S.; Sugrue, S. P.; He, X.; Min, J.; Kato, Y.; Shi, Y. G. *Cell* **2012**, *151* (6), 1200–1213.

(26) Zhang, Y.; Wang, L.; Schultz, P. G.; Wilson, I. A. *Protein Sci. Publ. Protein Soc.* **2005**, *14* (5), 1340–1349.

(27) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124* (31), 9026–9027.

(28) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35* (1), 225–249.

(29) Islam, K.; Chen, Y.; Wu, H.; Bothwell, I. R.; Blum, G. J.; Zeng, H.; Dong, A.; Zheng, W.; Min, J.; Deng, H.; Luo, M. *Proc. Natl. Acad. Sci.* **2013**, *110* (42), 16778–16783.

(30) Booth, M. J.; Ost, T. W. B.; Beraldi, D.; Bell, N. M.; Branco, M. R.; Reik, W.; Balasubramanian, S. *Nat. Protoc.* **2013**, *8* (10), 1841–1851.

(31) Hu, L.; Lu, J.; Cheng, J.; Rao, Q.; Li, Z.; Hou, H.; Lou, Z.; Zhang, L.; Li, W.; Gong, W.; Liu, M.; Sun, C.; Yin, X.; Li, J.; Tan, X.; Wang, P.; Wang, Y.; Fang, D.; Cui, Q.; Yang, P.; He, C.; Jiang, H.; Luo, C.; Xu, Y. *Nature* **2015**, *527* (7576), 118–122.

(32) Yao, Y.; Chen, P.; Diao, J.; Cheng, G.; Deng, L.; Anglin, J. L.; Prasad, B. V. V.; Song, Y. *J. Am. Chem. Soc.* **2011**, *133* (42), 16746–16749.

(33) Copeland, R. A.; Solomon, M. E.; Richon, V. M. *Nat. Rev. Drug Discov.* **2009**, *8* (9), 724–732.

(34) Xu, W.; Yang, H.; Liu, Y.; Yang, Y.; Wang, P.; Kim, S.-H.; Ito, S.; Yang, C.; Wang, P.; Xiao, M.-T.; Liu, L.; Jiang, W.; Liu, J.; Zhang, J.; Wang, B.; Frye, S.; Zhang, Y.; Xu, Y.; Lei, Q.; Guan, K.-L.; Zhao, S.; Xiong, Y. *Cancer Cell* **2011**, *19* (1), 17–30.

(35) Xu, X.; Yan, S.; Hu, J.; Guo, P.; Wei, L.; Weng, X.; Zhou, X. *Tetrahedron* **2013**, *69* (46), 9870–9874.

(36) Dai, Q.; Song, C.-X.; Pan, T.; He, C. *J. Org. Chem.* **2011**, *76* (10), 4182–4188.

(37) Bornemann, B.; Marx, A. *Bioorg. Med. Chem.* **2006**, *14* (18), 6235–6238.

(38) Farrell, I. S.; Toroney, R.; Hazen, J. L.; Mehl, R. A.; Chin, J. W. *Nat. Methods* **2005**, *2* (5), 377–384.

(39) Pon, R. T. *Tetrahedron Lett.* **1991**, *32* (14), 1715–1718.