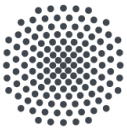


Universität Stuttgart

Biochemical Characterisation of TET DNA Hydroxylases

Von der Fakultät 4: Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Mirunalini Ravichandran

aus Mayavaram, Indien

Hauptberichter: Jun.-Prof. Dr. Tomasz Jurkowski

Mitberichter: Prof. Dr. Matthias Bochtler

Mitberichter: Prof. Dr. Markus Morrison

Tag der mündlichen Prüfung: 18.05.2017

Institut für Biochemie der Universität Stuttgart

2017

Erklärung

Hiermit versichere ich, das ich diese Arbeit selbst verfasst und dabei keine anderen als angegebenen Quellen und Hilfsmittel verwendet habe.

Stuttgart, Januar 2017

Mirunalini Ravichandran

Table of Contents

I. ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my supervisor Dr. Tomasz Jurkowski for his tremendous support and guidance throughout my PhD studies. I would like to thank him for the assistance and encouragement he provided both during my master and doctoral studies.

Next, I would like to thank Prof. Dr. Matthias Bochtler, Prof. Dr. Markus Morrison and Prof. Dr. Stefan Nussberger for kindly accepting my request to be the co-referees for my thesis and my special thanks goes to Prof. Dr. Matthias Bochtler, who agreed to travel from Poland to be a part of my thesis committee.

I am grateful to Prof. Dr. Albert Jeltsch for his financial support and guidance during the initial part of my thesis. I would also like to thank Prof. Dr. Stefan Nussberger, Prof.Dr. Dieter Wolf, Prof. Dr. Monilola Olayioyole and Prof. Dr. Christina Wege for accepting to review my thesis.

I am thankful to our collaborators Dr. Wolf Reik, Dr. Timothy Hore, Dr. Ferdinand von Meyenn for their contribution in the ascorbic acid project and I would like to thank Dr. Mark Helm and Jacob Dominik for the LC-MS measurements. I am thankful to Dr Yi Zhang for providing the TET plasmid constructs.

I would like to express my gratitude to Dr Renata Jurkowska for her stimulating scientific discussion, motivation and suggestions. I am thankful to her for helping me with the mouse embryonic stem cells experiment.

I would like to thank the lab colleagues (both current and old) for their help and contribution. I am grateful to my lab friends: Sara, Rebekka, Raluca, Radu, Peter, Phil and others, who have been helpful all the time and made me feel very welcome in Germany. I am thankful to them for many wonderful evenings we spent together both inside and outside the lab. There are no words to express my gratitude to Elisabeth, Max and Rebekka for all the help they provided me with the official stuffs.

I would like to thank Dragica and Regina for all the technical support, especially Dragica for being very friendly and helpful all the time. I am very thankful to the funding agency DFG for the financial support.

I would like to thank my friends in India, especially Radha, AK, Sathiya and others who provided moral support during my PhD thesis. I would like to thank my host parents Heidi and Dieter Piechowiak in Bremen for their continuous support and motivation.

I am really grateful to Rolf and Birgit for being such wonderful landlords and friends, who make me feel Oberaichen my second home. Last but not least; I would like to thank my parents, who provided all the support, strength, love and motivation. I dedicate this thesis to my parents without whom this would not have been possible.

II. LIST OF PUBLICATIONS

Hore, T. A*., von Meyenn, F*., Ravichandran, M*., Bachman, M., Ficz, G., Oxley, D., Santos, F., Balasubramanian, S., Jurkowski, T. P., and Reik, W. (2016). Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. **Proceedings of the National Academy of Sciences**. PMID: 27729528 (*equal contribution)

Jurkowski TP., **Ravichandran, M.**, Stepper P. (2015). Synthetic epigenetics-towards intelligent control of epigenetic states and cell identity. **Clinical Epigenetics.** PMID: 25741388

Delatte B., Jeschke J., Defrance M., Bachman M., Creppe C., Calonne E., Bizet M., Deplus R., Marroquí L., Libin M., **Ravichandran M.**, Mascart F., Eizirik DL., Murrell A., Jurkowski TP., Fuks F. (2015). Genome-wide hydroxymethylcytosine pattern changes in response to oxidative stress. **Scientific reports.** PMID: 26239807

III. ZUSAMMENFASSUNG

Die DNA Methylierung von CpG Dinukleotiden spielt eine wichtige Rolle in der Entwicklung von Säugetieren. Die kürzliche Entdeckung der TET-Enzyme zeigte, dass DNA Demethylierung durch die schrittweise Oxidierung von 5-Methylcytosin (5mC) zu 5-Hyrdoxymethylcytosin (5hmC), 5-Formylcytosin (5fC) und schließlich zu 5-Carboxylcytosin (5-caC) stattfinden kann, gefolgt von der Abtrennung der höher oxidierten Basen durch die Thymin DNA Glycosylase (TDG) und den Mechanismus der Basenexzisionsreparatur. Genetische Studien zeigten, dass die TET-Enzyme in zahlreichen biologischen Prozessen wie der Transkriptionskontrolle, der Differenzierung hämatopoetischer Stammzellen und der Entwicklung von embryonalen- und Urkeimzellen involviert sind, und dass sie häufig in Krebs dereguliert sind. Während die biologischen Funktionen der TET-Enzyme ausgiebig erforscht wurden, ist nur wenig über deren biochemischen Eigenschaften bekannt. In dieser Arbeit wurde die Biochemie der TET-Enzyme detailliert untersucht, wobei der Fokus auf dem katalytischen und kinetischen Verhalten lag, was zum Verständnis der molekularen Mechanismen der TET-Enzyme beitragen könnte. Zuerst wurde ein *in-vitro* System etabliert, welches einen neuartigen Mikrotiterplattentest zur Quantifizierung der durch die TET-Enzyme katalysierten Oxidationsprodukte einschließt. Um einen Machbarkeitsnachweis zu erbringen, wurde dieser mit mehreren Analoga von α-Ketoglutarat, Intermediaten des Zitrat-Zyklusses und Onkometabolite getestet. Darüber hinaus wurde der Einfluss von zweiwertigen Metallionen *in vitro* und *in vivo* getestet und es wurde nachgewiesen, dass die Zugabe von Nickel-Ionen zu Säugerzellen den 5hmC-Spiegel, aufgrund der Inhibierung der TET-Enzyme durch die Verdrängung des Fe^{2+} aus dem katalytischen Zentrum, verringert. Mit detaillierten biochemischen Studien wurde außerdem gezeigt, dass Ascorbinsäure (AscA) die Aktivität der TET-Enzyme durch effizientes Recyceln von Fe^{2+} moduliert, was die bisherige Annahme, dass AscA ein gebunder Cofaktor der TET-Enzyme sei, in Frage stellt. Schließlich wurde durch einen biochemischen Ansatz, gefolgt von n*ext generation sequencing* und bioinformatischen Analysen, das katalytische Verhalten der TET-Enzyme auf der Ebene einzelner Moleküle aufgeklärt. Mithilfe von linearer, doppelsträngiger DNA, welche mehrere 5mC-Substrate in verschiedenen Sequenzkontexten enthält, konnte gezeigt werden, dass die TET-Enzyme von Säugetieren die 5mC-Substrate sowohl im CG- und nicht-CG-Kontext oxidieren. Dabei ist wichtig, dass sowohl die TET-Enzyme von Säugetieren und *Naegleria gruberi* Tet1-like

dioxygenase (nTet) starke und unterschiedliche Präferenzen der flankierenden Sequenzen aufweisen. Zusätzlich wurde gezeigt, dass die TET-Enzyme (von Säugetieren und nTet) die Substrate auf der DNA distributiv katalysieren könnten.

IV. ABSTRACT

Methylation of DNA in CpG dinucleotide plays an important role in mammalian development. The recent discovery of TET enzymes showed that DNA demethylation can occur through stepwise oxidation of 5-methylcytosine (5mC) to 5-hyrdoxymethylcytosine (5hmC), 5-formylcytosine (5fC) and finally to 5-carboxylcytosine (5-caC) followed by the removal of the higher oxidised bases by Thymine DNA glycosylase (TDG) and base excision repair mechanism. Genetic studies revealed that the TET enzymes are involved in numerous biological processes such as transcriptional regulation, hematopoietic stem cell differentiation, embryonic, primodial germ cells (PGCs) development and are commonly misregulated in cancer. While the biological functions of TET enzymes have been studied extensively, very little is known about their biochemical properties. In this body of work, the biochemistry of TET enzymes was investigated in detail with the focus on their catalytic and kinetic behaviour, which would allow us to understand the molecular mechanisms of TET enzymes. First of all, an *in vitro* system including a novel plate assay to quantify the oxidation products catalysed by TET enzymes, was established. As a proof of principle, several analogues of α-Ketoglutarate, the intermediates of citric acid cycle (oncometabolites) were tested. Moreover, the effect of divalent metal ions was tested both *in vitro* and *in vivo* and it was demonstrated that the addition of nickel ions to mammalian cells decreased the level of 5hmC through inhibition of TET enzymes by displacing the Fe^{2+} from the catalytic centre. Furthermore, using detailed biochemical studies, it was demonstrated that ascorbic acid (AscA) modulates the activity of TET enzymes through efficient recycling of Fe^{2+}, which challenges the existing view of AscA as a bound cofactor of TET enzymes. Finally, using biochemical approach followed by next generation sequencing and bioinformatics analysis, the catalytic behaviour of TET enzymes on single molecule level was elucidated. Using linear double stranded DNA containing multiple 5hmC-substrates in different flanking sequence context, it was shown that mammalian TET enzymes oxidize 5hmC substrates in both CG and non-CG context. Importantly, both mammalian TET enzymes and *Naegleria gruberi* Tet1 like dioxygenase (nTet) showed a strong and distinct flanking sequence preference. In addition, it was shown that TET enzymes (both mammalian and nTet) might catalyse the substrates on DNA in distributive manner.

V. LIST OF ABBREVIATIONS

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
aa	amino acids
AdoMet	S-adenosine-L-methionine
AML	acute myeloid leukemia
AscA	ascorbic acid
BER	base excision repair
B-ME	ß-mercaptoethanol
C-4PH	collagen-4-prolyl hydroxylase
CcrM	cell cycle-regulated DNA adenine methyl transferase
CTCF	CCCTC binding factor
DHG	D-α-hydroxyglutarate
DMOG	dimethyl oxalylglycine
DNMTs	DNA methyltransferases
DSBH	double stranded beta helix
DTT	dithiothreitol
EBF1	early B cell factor 1
ECL	enhanced chemiluminescence
EDC	1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride
EMT	epithelial-mesenchymal transition
fCAB	5fC chemically assisted bisulfite sequencing
FH	fumarate hydratase
FL	fetal liver
GIST	gastrointestinal stromal tumour
GlnNAc	N-acetyl glucosamine
H2B	histone 2B
HCC	hepatocellular carcinoma
HQ	hydroquinone

HRP	horse radish peroxidase
HT	high throughput
ICM	inner cell mass
IDAX	inhibition of the dvl and axin complex
IDH	isocitrate dehydrogenase
iPS	induced pluripotent stem cells
IPTG	isopropyl ß-D-1-thiogalactpyranoside
LB	Luria Bertani (LB)
LC-MS	liquid chromatography and mass spectrometry
L-glu	L-glutathione
LHG	L-α-hydroxyglutarate
LINE	long interspersed nuclear element
LTR	long terminal repeat
MAB	methyl-assisted bisulfite sequencing
MALDI-TOF	matrix assisted laser desorption and ionisation- time of flight
MBD4	methyl-CpG binding protein 4
mES	mouse embryonic stem (mES)
MET	mesenchymal to epithelial transformation
MLL	mixed lineage leukemia
MTases	methyltransferases
NAC	N-acetyl L-cysteine
NC	nitro cellulose
NGS	next generation sequencing
NOG	N-oxalylglycine (NOG)
nTet	Tet-like dioxygenase 1 from *Naegleria gruberi*
OGT	O-linked N-acetylglucosamine (GlcNAc) transferase
OxBS-seq	oxidative bisulfite sequencing
PAR	poly-ADP ribose
PARP	poly-ADP ribose polymerase
PCR	polymerase chain reaction
PGC	primodial germ cells

PN	pronuclear stage
RAR	retinoic acid receptor
RARE	retinoic acid receptor element
REST	RE1-silencing transcription factor
SAH	S-adenosine-L-homocysteine
SDH	succinate dehydrogenase
SINE	small interspersed nuclear element
SMRT	small molecule, real-time sequencing
T4-BGT	T4 phage ß-glucosyltransferase
TAB-seq	TET assisted bisulfite sequencing
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TF	transcription factor
TFEB	transcription factor EB
TKO	triple knockout
TLC	thin layer chromatography
Treg	T regulatory cells
TSS	transcription start site
WT	wilm's tumour
αKG	α-Ketoglutarate

1. INTRODUCTION

1.1. DNA methylation

DNA methylation is one of the best studied epigenetic modifications present in both prokaryotes and eukaryotes. In prokaryotes, DNA methylation can be found at both adenine (N6 position) and cytosine (N4 and C5 position) bases and is involved in a host defence mechanism (reviewed in Wilson and Murray 1991), whereas in mammals it is mostly limited to the C5 position of cytosine, though methylation at the N6 position of adenine has recently been reported in mammals (Ratel et al. 2006; Wu et al. 2016b). 5-methylcytosine (5mC) is also called the fifth base and in the mammalian genome, it is found primarily within a CG dinucleotide context. The human genome is comprised of approximately 56 million CG sites, out of which 60-80% are methylated. Around 10% of the CG sites are present in CG-rich regions called CpG island, and are associated with 70% of the promoters including housekeeping and developmentally regulated genes, which are often unmethylated in a tissue specific manner (Saxonov et al. 2006). The presence of 5mC on CG-rich promoters is generally associated with gene repression, and is important for various biological processes such as parent specific gene imprinting, X-chromosomal inactivation in females, silencing of repeat elements and stem cell differentiation (reviewed in Jurkowska et al. 2011).

Methylation of DNA is carried out by a group of enzymes called methyltransferases (MTases), which utilize S-adenosine-L-methionine (AdoMet) as a donor and transfer the methyl group from AdoMet to cytosine, leading to the formation of 5mC and S-adenosine-L-homocysteine (SAH). The transferred methyl group on cytosine is located in the major groove of the DNA and it does not interfere with the Watson-Crick base pairing. Moreover, the methyl group in the major groove can interact with DNA binding proteins and provide an extra layer of information without altering the nucleotide sequence (reviewed in Goll and Bestor 2005).

DNA methylation in mammals is achieved by the enzymes called DNA methyltransferases (DNMTs), which include so called *de novo* methyltransferases such as DNMT3A, DNMT3B, catalytically inactive member DNMT3L and the maintenance methyltransferase DNMT1. The methylation mark on unmethylated DNA is initially set by DNMT3s on CG dinucleotide during development and in primordial germ cells (PGCs). DNMT3A and DNMT3B have both overlapping and distinct functions. DNMT3A together with DNMT3L is essential for the

methylation of imprinting genes (Delaval and Feil 2004; Hore et al. 2007). Both DNMT3A and DNMT3B are required for the methylation of repeat elements. DNMT3B specifically methylates minor satellite repeats in pericenteromeric region of the chromatin and DNMT3A is necessary for the methylation of transposable elements such as long interspersed nuclear element (LINEs), short interspersed nuclear elements (SINEs) and long terminal repeats (LTR). Moreover, both DNMT3A and DNMT3B are necessary for the methylation of gene promoters during cellular differentiation (reviewed in Smith and Meissner 2013).

After each cell division, the methylation level is halved in the DNA of both daughter cells, as the nascent DNA strand lacks the methyl marks, and the methylation pattern is re-established on the newly synthesized DNA strand by the maintenance methyltransferase DNMT1, which is upregulated during the S-phase (Detich et al. 2001). DNMT1 has a (up to 40-fold) preference for hemi-methylated substrate over unmethylated CG *in vitro* (Jeltsch 2006), and is recruited to hemi-methylated substrate by its interaction partner UHRF1 (Sharif et al. 2007). However, this classical mode of methylation proposed by Holiday & Pugh and Riggs (Holliday and Pugh 1975; Riggs 1975) has been revised based on the studies in the recent past, which demonstrated that the functional segregation of these two groups of enzymes is not very strict, as both DNMT3s and DNMT1 are required for both *de novo* and maintenance methylation (reviewed in Jeltsch and Jurkowska 2014). Deletion of any of these three DNMTs is embryonically lethal (Li et al. 1992; Okano et al. 1999), and a disturbance in the DNA methylation pattern is often implicated in various diseases such as cancer, immune disorders, etc. (Baylin and Jones 2011), which emphasizes its vital role in normal cellular functions.

1.2. DNA demethylation

5mC was considered relatively stable due to the chemical nature of the C-C bond, and demethylation of DNA was believed to occur through replication-dependent dilution due to the absence or improper functioning of the maintenance methylation machinery. This notion changed in the year 2000, when global loss of the methylation mark was detected in mouse zygotes, and this was observed independent of DNA replication. For instance, during mammalian development DNA demethylation occurs at two stages. The first occurrence is during early embryogenesis in the paternal genome, following fertilization and preceding DNA replication, conferring totipotency to the developing embryo, and the methylation pattern is

re-established in the preimplantation stages (Mayer et al. 2000a; Oswald et al. 2000). The second occurrence is during the germ cell specification that includes demethylation of imprinting genes (Hajkova et al. 2002; Yamazaki et al. 2003). Furthermore, active DNA demethylation has also been observed at specific loci in T cells, neurons and other cells (Bruniquel and Schwartz 2003; Martinowich et al. 2003). Although active DNA demethylation has been observed in different biological contexts in mammals, the molecular mechanism remained largely unclear. Previous reports on the involvement of BER in excising 5mC in chicken embryos (Jost 1993) and the demonstration of DNA demethylation through direct excision of the methyl group containing base by the DEMETER/ROS1 family of DNA glycosylases in plants had raised the possibility of the presence of similar pathways in mammals (reviewed in Zhu 2009). However, the search for an orthologous glycosylase, which could excise the methyl group in mammals, was not fruitful, as the enzymes suspected to have the glycosylase activity on 5mC, such as TDG and MBD4, had a stronger activity on T/G mismatch repair and around a 30-40 fold weaker activity on 5mC *in vitro* than on T:G mismatches (Zhu et al. 2000b, 2000a). Thus the enzyme responsible for active DNA demethylation in mammals remained enigmatic for a long time.

1.3. Discovery of TET enzymes and their reaction products

A major breakthrough came in 2009, when a group of enzymes called Ten-Eleven Translocation (TET) was shown to oxidize 5mC to 5-hydroxymethylcytosine (5hmC) both *in vitro* and in mES cells, in which 5hmC constitutes about 0.03% of total nucleotides (Tahiliani et al. 2009). In parallel, another independent group identified the presence of higher levels of 5hmC (about 0.6% and 0.2% of total nucleotides) in mouse purkinje neurons and granule cells respectively (Kriaucionis and Heintz 2009). These two seminal discoveries suggested a possible pathway for active DNA demethylation.

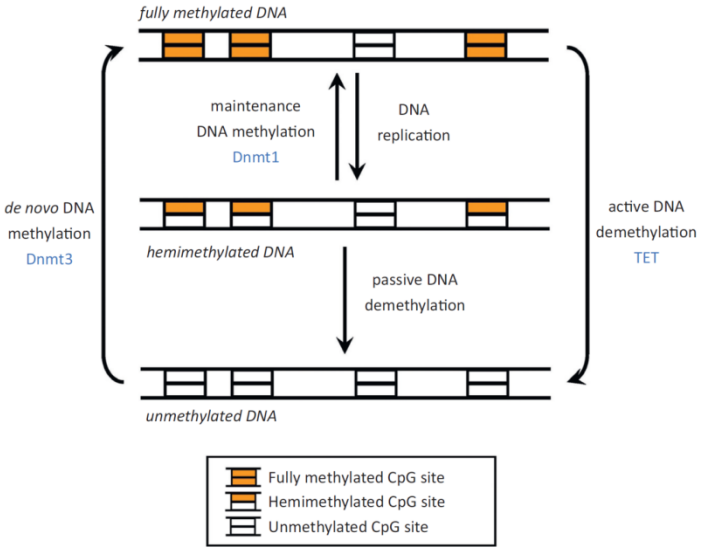


Figure 1: Model of DNA methylation and demethylation pathway, taken from (Jeltsch and Jurkowska 2014). The unmethylated DNA is methylated by the DNMT3 family of enzymes during early development. After each cell division, the methylation mark is re-established on the newly synthesized daughter strand by DNMT1. During global reprogramming, the DNA methylation mark is lost by both replication-dependent dilution and active demethylation initiated by the TET family of enzymes.

Although the gene coding for TET1 (also known as CXXC6) was known to be a fusion partner of MLL involved in acute myeloid leukaemia (AML), the function of the gene was not characterized until 2009 (Ono et al. 2002). TET enzymes were identified based on the computational search for DNA modifying enzymes using the predicted dioxygenase domain of JBP1 and JBP2 that oxidize the methyl group of thymine to 5-hydroxymethyluracil in trypanosomes (Cliffe et al. 2009; Iyer et al. 2009).They belong to the iron and α-ketoglutarate dependent dioxygenase family (Fe^{2+}/αKG-DO). Bioinformatics searches further revealed that TET enzymes are distributed across the metazoans that have DNA methylation marks and are also present in fungi and algae (Iyer et al. 2009).

The oxidized base 5hmC was first described in 1952, when it was identified in the genomes of T-even bacteriophages (T2 and T4) as a modified base, which gets further glucosylated and provides protection against cleavage by bacterial restriction enzymes (Wyatt and Cohen 1952). Later, Penn and colleagues demonstrated that 5hmC was also found in adult rats, mice and frogs and that it accounted for ~ 15% of total cytosine (Penn et al. 1972). Yet this finding was disregarded, as it could not be reproduced by another group (Kothari and Shankar 1976). Thereafter, the formation of 5hmC in mammalian cells was thought to be the result of oxidative damage until its rediscovery in 2009. Later studies showed that 5hmC is present in

different mouse tissue types such as heart, kidney, lung, muscle and the highest level is found in the brain and ES cells (Globisch et al. 2010).

Two years later, it was demonstrated that similar to the thymidine salvage pathway (Smiley et al. 2005), TET enzymes can oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al. 2011; He et al. 2011). Furthermore these higher oxidized bases (5fC and 5caC) are recognized and excised by the thymine DNA glycosylase (TDG), which trigger the base excision repair pathway (BER) to replace the abasic site to an unmodified cytosine (He et al. 2011), thereby completing the DNA demethylation pathway.

1.4. Structural organization of TET enzymes

The mammalian TET family comprises of three members, namely: TET1, TET2 and TET3 that share similar domain architecture (Figure 2). These are large proteins harbouring the catalytic activity at their C-terminus, which is composed of a double-stranded β helix domain (DSBH) and a cysteine-rich region preceding the DSBH. In metazoan TETs, the DSBH is interrupted by a large unstructured region, which is less conserved than the DSBH domain and is believed to involve in protein-protein interaction. The amino terminus of both TET1 and TET3 contain a CXXC domain. But the TET2 protein lacks the CXXC domain, which was lost during evolution after gene duplication and inversion, and is now coded as another protein named IDAX (inhibition of the dvl and axin complex) (Iyer et al. 2009; Pastor et al. 2013).

The core catalytic domain (DSBH) forms the characteristics dioxygenase domain and contains the binding sites for Fe^{2+} and αKG, which are essential for the catalytic activity. The amino acids which are crucial for Fe^{2+} and αKG binding in the DSBH and Zn^{2+} binding in the Cys-rich domain are conserved among the family members and all the three enzymes are shown to be catalytically active *in vitro* (Iyer et al. 2009; Tahiliani et al. 2009).

Tet-like dioxygenases are also present in the unicellular amoebaflagellate *Naegleria gruberi*. In total there are eight Tet-like enzymes in this organism. All of them are single domain proteins with various lengths, yet they all have a conserved region of around 210 residues. The Tet-like dioxygenase 1 (nTet) shares a sequence similarity of about 39% with the mammalian TET1 and contains the conserved Fe^{2+} and αKG binding motif (Hashimoto et al. 2014).

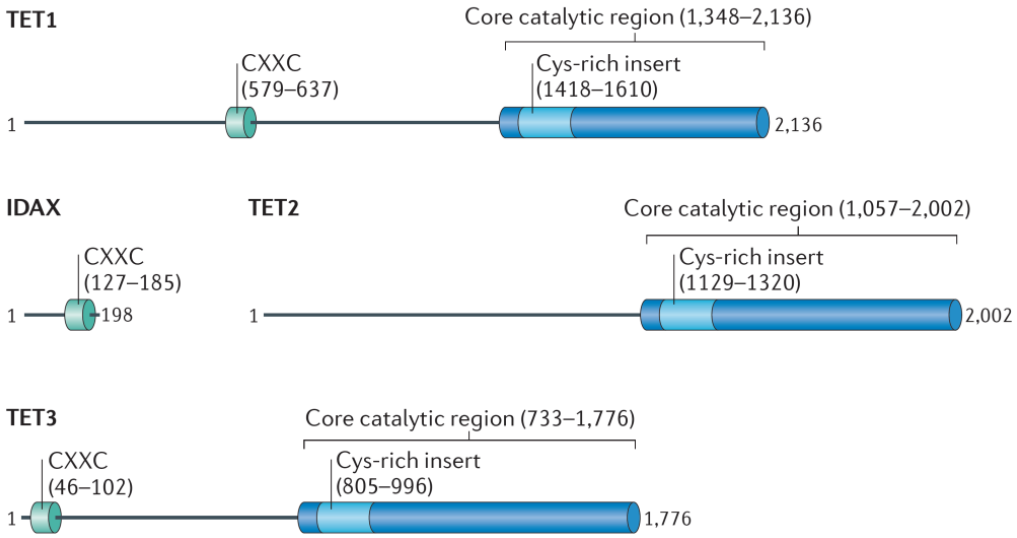


Figure 2: The domain architecture of Homo sapiens TET family members, taken from (Pastor et al. 2013). There are three members in the family, namely TET1, TET2 and TET3. The catalytic domain resides in the C-terminus and is interfered by the Cys-rich region. TET1 and TET3 have an N-terminus CXXC domain, whereas the CXXC domain of TET2 was lost during evolution.

The crystal structure of the human TET2 catalytic domain and nTet was solved recently (Hu et al. 2013; Hashimoto et al. 2014). Crystal structure of the truncated, active version of the human TET2 catalytic domain (PDB ID: 4NM6) in complex with 5mC containing DNA (Figure 3) reveals that the catalytic domain containing a DSBH core forms a globular structure, which is stabilized by the flanking region of DSBH and the Cys-rich region. Unlike the Cys-rich region in other proteins, the TET Cys-rich region does not form an independent domain, but wraps around the DSBH core. Interestingly, residues from both the DSBH and the Cys-rich region are involved in coordinating zinc ions, stabilizing the overall structure of the enzyme and proving crucial for catalysis (Hu et al. 2013). The DNA is bound above the DSBH core, which is enriched in basic and hydrophobic amino acids. Similar to DNA MTases and DNA repair enzymes (Klimasauskas et al. 1994; Yang et al. 2008), TET enzymes utilize a base flipping mechanism to position the target base in the catalytic pocket for the oxidation reaction. Once the methyl group is located in the catalytic pocket, it is oriented towards the catalytic iron and αKG which facilitate the catalytic turnover (Hu et al. 2013; Hashimoto et al. 2014).

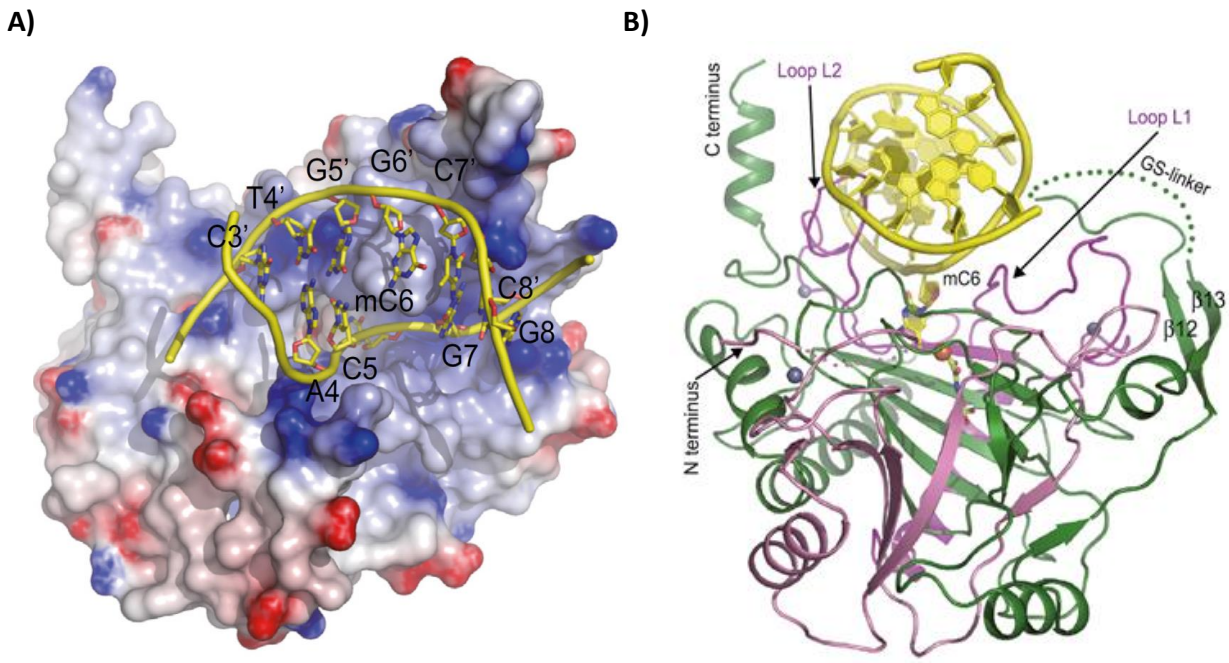


Figure 3: Structure of the hTET2 catalytic domain in complex with 5mC containing DNA, taken from (Hu et al. 2013). A) The TET2 catalytic domain is represented as electrostatic surface. DNA is shown in yellow and the DNA binding surface of the protein is rich in hydrophobic (white) and basic (blue) residues. B) The structure of TET2-DNA is represented in ribbon form. DNA is shown in yellow. The bases of DNA and NOG (αKG analogue) are represented as sticks. The dotted line in green indicates the removed unstructured region. The red and three grey balls indicate the position of catalytic iron and zinc cations respectively.

Analysis of the interaction of human TET2 with DNA indicates that besides the target 5mC within a CpG dinucleotide context, the enzyme does not interact with other neighbouring bases (Hu et al. 2013). Intriguingly, the enzyme does not recognize the methyl group of the target cytosine suggesting that this would allow TET2 to generate higher oxidation of 5hmC to 5fC and 5caC (Hu et al. 2013). Additionally, TET enzymes have been shown to oxidize the methyl group of thymine (T) to 5-hydroxymethyl uracil (5hmU) (Pfaffeneder et al. 2014), however the physiological relevance still needs to be validated.

Crystal structure of human TET2 in complex with 5hmC and 5fC substrates indicates that, although the active site of the enzyme can accommodate 5hmC and 5fC, TET enzymes prefer 5mC over 5hmC and 5fC. This may be due to the intermolecular bonding between the hydroxyl group and the co-substrate or the intramolecular bonding of the carbonyl group with the N4 nitrogen of cytosine that may prevent the hydrogen abstraction from the carbon (Hu et al. 2015). Supporting this, the investigation of the catalytic efficiency of hTET2-CD on substrates containing 5mC, 5hmC and 5fC showed that TET2 was more active on 5mC than

5hmC and 5fC (Hu et al. 2015). Interestingly, the recent biochemical analysis on hTET2 suggested that T1372 and Y1902 form an active site scaffold through hydrogen bonding and mutation of T at position 1372 to the amino acid E/Q/N/D/V stalls the oxidation of 5hmC to 5fC/5caC due to hydrogen bonding between T1372 mutant and –OH moiety of 5hmC. The mutant T1372S behaves similar to the WT while the other mutants lead to complete loss of activity (Liu et al. 2017).

Moreover, the solved crystal structure of Tet-like dioxygenase 1 from *Naegleria gruberi* (nTet PDB accession number- 5CG8) with a 5hmC substrate demonstrated that the size of the hydrophobic pocket is important for further oxidation. For 5hmC (the first oxidation product) to become a substrate, the hydroxyl group (OH-CH_2-) has to rotate away from the iron to the hydrophobic pocket to expose the carbon for hydrogen abstraction from a target CH_2. Decreasing the hydrophobic pocket size by a point mutation decreases the higher oxidation capacity of nTet, implying that spatial constraint can affect the activity. Moreover, nTet also exhibited lower activity on 5hmC and 5fC substrates compared to 5mC (Hashimoto et al. 2015).

1.5. Catalytic mechanism of TET enzymes

All three TET enzymes follow the same catalytic mechanism, which is typical for Fe^{2+}/αKG dependent dioxygenases. As the target atom has a C-C bond and not a C-N bond, as in jumonji containing histone demethylases and adenine demethylases (reviewed in Ozer and Bruick 2007), the oxidation product of TET enzymes is stable and cannot be released spontaneously as formaldehyde, thus resulting in the hydroxylation and not demethylation. The reactions catalysed by TET enzymes are shown in figure 4.

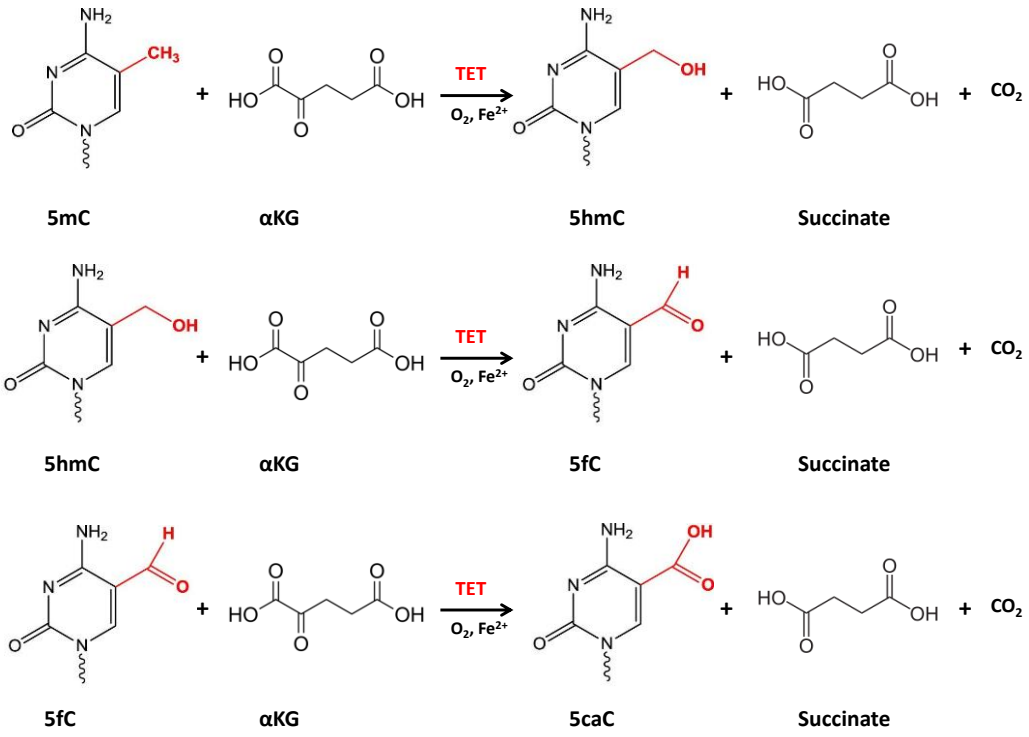


Figure 4: The reactions catalysed by TET enzymes. TET enzymes catalyse the successive oxidation of 5mC to 5hmC, 5fC and 5caC respectively together with the formation of succinate and CO_2.

TET enzymes perform the catalysis through a radical mechanism, as it was first explained for the collagen 4-prolyl hydroxylases (C-4PH) (de Jong et al. 1982). Fe^{2+} binds to the active site coordinated by the HXD motif and the other three coordination sites are occupied by the water molecules. The catalysis is initiated by the binding of αKG to the active pocket through the conserved arginine residue followed by the binding of substrate DNA. Binding of both αKG and the DNA substrate displaces the water molecules from Fe^{2+} site. Oxygen molecule binds to the already formed enzyme-substrate-co-substrate complex and initiates the reaction through the nucleophilic attack on the uncoordinated oxygen atom on the second carbon of αKG. This leads to the decarboxylation of αKG and the cleavage of oxygen molecule concomitant with the formation of a highly reactive ferryl-oxo intermediate together with the release of CO_2. As a next step, the highly reactive ferryl-oxo species abstracts the hydrogen from the neighbouring DNA substrate resulting in Fe^{3+}-hydroxide formation and a substrate radical that eventually resulted in the generation of hydroxylated DNA and Fe^{2+}. This is followed by the dissociation of the product, succinate (Figure 5) The next oxidation steps also follow similar process, but the exchange of succinate with αKG is necessary to start the new catalytic cycle. The detailed mechanism of how the exchange of succinate and the higher oxidation occur is currently unknown.

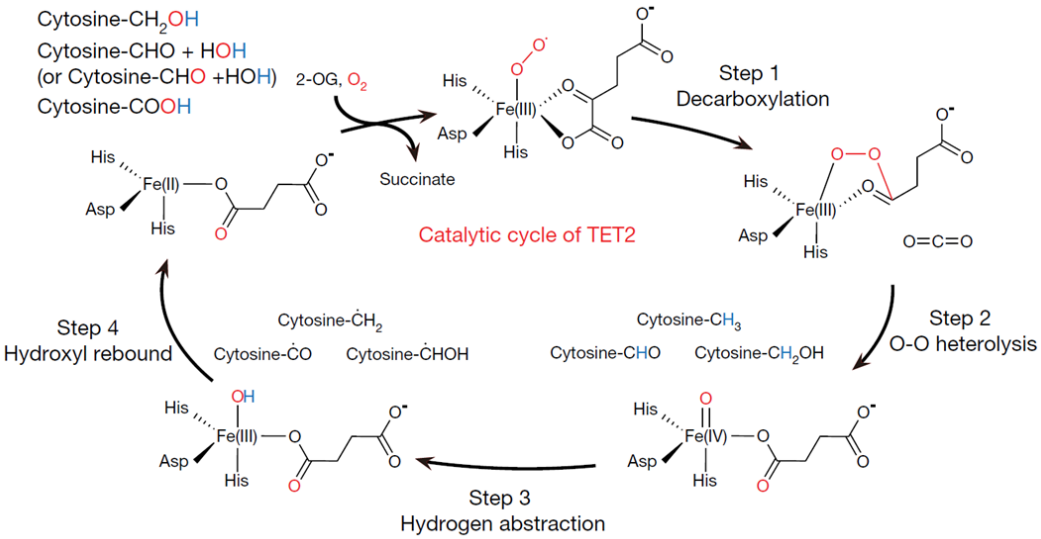


Figure 5: The catalytic mechanism of TET enzymes, taken from (Hu et al. 2015). Different steps involved in the oxidation of 5mC, 5hmC or 5fC DNA is shown here. In the presence of oxygen, Fe^{2+} and αKG, TET enzymes split the oxygen molecule, one oxygen atom is incorporated into the substrate DNA and the other oxygen goes to αKG leading to the formation of 5hmC, 5fC or 5caC together with succinate and CO_2.

1.6. Function of ascorbic acid on TET activity

Besides Fe^{2+}, O_2 and αKG, some enzymes in the family of Fe^{2+}/αKG dependent dioxygenases additionally require ascorbic acid (AscA) for their full activity (Ozer and Bruick 2007). The best example of this category is collagen-4 prolyl hydroxylase (C-4PH) (Myllyla et al. 1984). Biochemical investigations revealed that AscA was a necessary cofactor for the enzymatic reaction and in the absence of AscA the oxidation reaction of C-4PH stopped abruptly or proceeded at a very slow rate and eventually led to its inactivation (Nietfeld and Kemp 1981; de Jong et al. 1982). However, the molecular mechanism was unclear and it was believed to happen through AscA mediated reduction of the catalytic iron, which underwent oxidation through the uncoupled reaction (oxidation in the absence of substrate) (Hallberg et al. 1989). In line with this, AscA was found to be consumed in stoichiometry during the *in vitro* enzymatic reaction of C-4PH (Myllyla et al. 1984). However, another study observed a significant increase of Fe^{3+} generation even after the addition of prime substrate, indicating that the coupled oxidation might also lead to iron oxidation (McNeill et al. 2005). Consequently, AscA mediated rescue of the dioxygenase activity has been highly debated.

Interestingly, similar to C-4PH, addition of AscA has been shown to increase the TET mediated formation of 5hmC in mammalian cell culture studies (Blaschke et al. 2013; Yin et al. 2013;

Minor et al. 2013; Dickson et al. 2013). Using biochemical experiments, Yin and co-workers showed that TET enzymes require AscA as a co-factor that can bind to the enzyme and enhances their activity (Yin et al. 2013). Moreover, no increase in the activity of TET enzymes was observed with other reducing agents (Blaschke et al. 2013; Yin et al. 2013). This specific requirement of AscA over other reducing agents by TET enzymes, together with the studies suggesting the binding of AscA to the active centre of C-4PH (de Jong et al. 1982) and anthocyanidin synthase (Wilmouth et al. 2002), supported the role of AscA as a cofactor. However, the detailed mechanism of how, when and where AscA binds to the TET catalytic centre during the catalytic cycle is not clear.

1.7. TET mediated DNA demethylation pathways

Identification of TET enzymes and their reaction products have paved the way for DNA demethylation through a direct enzymatic action on methyl group. Since their discovery numerous plausible DNA demethylation pathways involving TETs have been investigated both *in vivo* and *in vitro*. Intriguingly, the formation of 5hmC via TET enzymes has been shown to facilitate the passive dilution of modified bases, as DNMT1 is less active on hemi-5hmC containing DNA (> 60-fold) *in vitro* (Valinluck et al. 2004; Hashimoto et al. 2012). This contributes to the replication-dependent loss of methylation induced by 5hmC formation. However, this observation has been contradicted by recent reports. Firstly, *in vitro* studies showed that the DNMT1 interaction partner UHRF1 binds 5hmC (Frauer et al. 2011; Iurlaro et al. 2013; Spruijt et al. 2013) thereby targeting DNMT1 to hemi-5hmC containing DNA. Secondly, unlike DNMT1, DNMT3A and DNMT3B are not sensitive to hemi-5hmC DNA and can re-methylate hemi-5hmC containing DNA (Hashimoto et al. 2012; Ji et al. 2014). These observations argue against 5hmC mediated passive dilution and require further investigation.

The other suggested pathway involves TET-TDG-BER mediated DNA demethylation, where the higher oxidation products 5fC and 5caC are excised by TDG, followed by the restoration of the abasic site with an unmodified cytosine by the BER machinery (Ito et al. 2011; He et al. 2011). It is suggested to be the main active DNA demethylation pathway triggered by TET enzymes. This is supported by the observations that the knockdown of TDG in mES cells results in an up to 10-fold increase of 5fC and 5caC (Cortellino et al. 2011; Raiber et al. 2012) and overexpression of TDG leads to the opposite effect with no significant changes in the level

of 5mC or 5hmC (Nabel et al. 2012). Moreover, *in vitro* investigation of TDG demonstrated that TDG removes 5fC and 5caC in CpG dinucleotide context more efficiently than the mismatch repair T:G (Maiti and Drohat 2011), suggesting that the main function of TDG is the excision of the oxidized bases over T:G removal. Nevertheless, TDG-BER mediated demethylation cannot account for genome-wide demethylation, as TDG is not highly expressed in the zygote and loss of TDG does not affect the demethylation in zygote (Guo et al. 2014). This indicates that other DNA glycosylases may be involved. Furthermore, TDG and the BER mechanism may compromise the genome stability by introducing multiple nicks and single or double strand breaks on the DNA, while processing multiple substrates. Besides BER, nucleotide excision repair (NER) protein GADD45A has also been implicated in DNA demethylation and it has been shown to interact with TET (Arab et al. 2014; Li et al. 2015; Kienhöfer et al. 2015). These observations suggest that both NER and BER may coordinate with TET enzymes in processing the oxidized base, but further experimental evidence is needed.

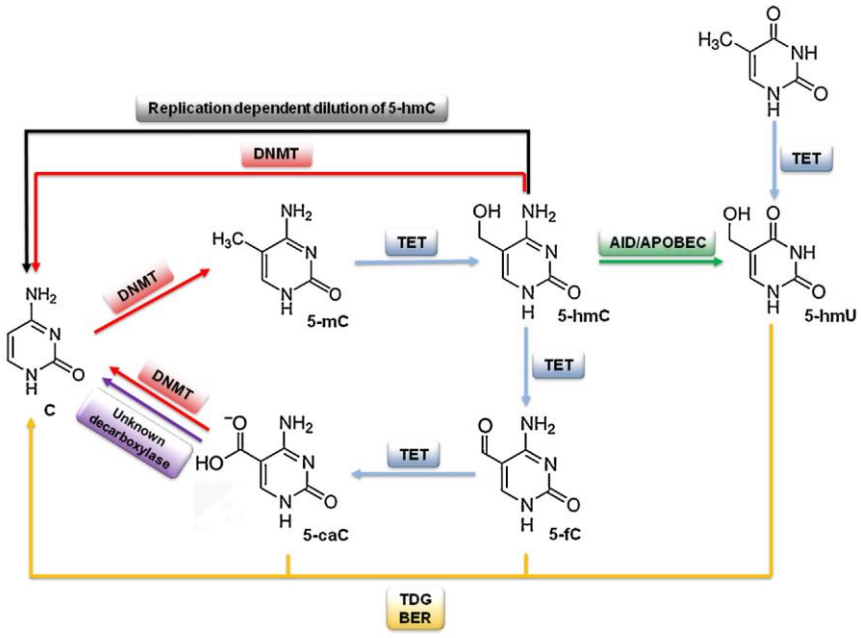


Figure 6: Possible DNA demethylation pathways triggered by TET enzymes, taken from (Rawłuszko-Wieczorek et al. 2015).

Another possible pathway involves the AID/APOBEC mediated deamination of 5hmC to 5hmU, which is removed by DNA glycosylases such as SMUG, MBD4, NEIL or TDG (Cortellino et al. 2011; Banerjee et al. 2011; Olinski et al. 2016). Supporting this view, a study conducted in the mouse brain (which has the highest level of 5hmC) reported that the TET mediated

formation of 5hmC can be processed further by AID/APOBEC. Moreover, both work together at the locus specific DNA demethylation of neuronal activity induced genes in mouse dentate gyrus (Guo et al. 2011). Alternatively, the capacity of TET enzymes to oxidize thymine to 5hmU has also been shown to trigger DNA demethylation through DNA glycosylases and BER (Pfaffeneder et al. 2014). However, this pathway of DNA demethylation still remains controversial due to the inconsistencies in the reported results and needs to be clarified by further studies. On the other hand, another putative pathway involving the removal of the carboxyl group from 5caC similar to the thymidine salvage pathway (Smiley et al. 2005) has also been proposed. However, no such decarboxylase has been identified so far in mammals.

Interestingly, DNA demethylation involving both DNMTs and TETs has been described. In the absence of AdoMet, the methyltransferase DNMT1 was shown to remove 5hmC as formaldehyde (Liutkeviciute et al. 2009) and a similar result was shown for DNMT3A and DNMT3B, however it required non-physiological concentration of H_2O_2 (Chen et al. 2012). Moreover, DNMT3A was reported to convert 5caC to C in the absence of SAM (Liutkevičiutė et al. 2014). It should be noted that these conclusions were drawn based on *in vitro* experiments and need to be validated *in vivo*.

In summary, it has been increasingly clear that the TET mediated formation of oxidized bases are processed by glycosylase-BER machinery, but the contribution of different glycosylases and the factor which triggers the downstream processes in different cell types is not yet clearly understood.

1.8. Methods to map the oxidized Bases

Despite having a substantial interest in mapping the oxidized bases (5hmC, 5fC and 5caC) genome-wide, initial studies were hindered by the lack of an efficient assay to discriminate all these oxidized bases at single base resolution. Methods such as thin layer chromatography (TLC) (Tahiliani et al. 2009; Kriaucionis and Heintz 2009), liquid chromatography-coupled mass spectrometry (LC-MS) (Münzel et al. 2011), the methylation or hydroxymethylation sensitivity based cleavage analysis (Szwagierczak et al. 2010; Kinney et al. 2011), dot blot assay using antibodies specific for the modified bases (Xu et al. 2011a) and single molecule real time sequencing (SMRT) (Flusberg et al. 2010) that relies on the modification dependent kinetics

of the polymerase were used. These methods are not very quantitative (Dot blot or TLC), not high throughput (HT) or require expensive and sophisticated machinery (LC-MS).

Moreover, the genome-wide studies have been conducted based on the enrichment coupled DNA precipitation methods using antibodies against 5hmC, 5fC and 5caC (Shen et al. 2013; Shen and Zhang 2013) or further modification of 5hmC using chemical or enzymatic treatment such as sodium bisulfite, glucosylation, peroxidate oxidation, biotinylation or a combination of chemical modification and biotinylation and reduced representation based method (for a detailed review refer Song et al. 2013; Shen and Zhang 2013; Wu and Zhang 2015). Even though the results obtained from these studies have greatly enhanced our knowledge about TET induced oxidation of 5mC, there are numerous drawbacks associated with each method, the main one being the inability to map the bases at single base resolution. This motivated the researchers to develop efficient assays, which could discriminate 5mC from the oxidized bases at single base resolution. To achieve this, the modification coupled bisulfite sequencing method has been developed, where the oxidized base is modified using enzymatic or chemical modification before the bisulfite treatment. Moreover, advancement in the sequencing technology has allowed gaining deeper insight on the genome wide scale. To date, there are several methods available, such as oxidative bisulfite sequencing (OxBS-seq - for 5hmC) (Booth et al. 2012), TET assisted bisulfite sequencing (TAB-seq - for 5hmC) (Yu et al. 2012), 5fC chemically assisted bisulfite sequencing (fCAB - for 5fC) (Xia et al. 2015), 5caC chemically assisted bisulfite sequencing (caCAB - for caC) (Wu et al. 2014), methyl-assisted bisulfite sequencing (MAB - for both 5fC and 5caC) (Guo et al. 2014; Wu et al. 2014) and caMAB - for 5caC (Wu et al. 2016a). A brief summary of each method is listed in table 1. However it should be noted that there is no method available to date which can discriminate all 5 modified bases at the same time.

Table 1: Brief procedure of the methods used to map the oxidized bases at single base resolution

Method	References	Mapped base	Procedure
OxBS-seq	Booth et al. 2012	5hmC	Oxidation of 5hmC to 5fC with potassium peruthenate ($KRuO_4$) followed by bisulfite sequencing
TAB-seq	Yu et al. 2012	5hmC	Glucosylate 5hmC using T4-BGT and oxidize 5mC to 5fC or 5caC using TET enzymes followed by bisulfite sequencing
caCAB	Wu et al. 2014	5caC	Chemical modification using EDC followed by bisulfite sequencing
fCAB	Xia et al. 2015	5fC	Treatment with $EtONH_2$ or $NaBH_4$ followed by bisulfite sequencing
MAB	Guo et al. 2014; Wu et al. 2014	5fC and 5caC	Treatment of the DNA with M.SssI (converts C to 5mC) followed by bisulfite sequencing
caMAB	Wu et al. 2016	5caC	Treatment of the DNA with M.SssI followed by $NaBH_4$ and bisulfite sequencing

1.9. Genome-wide distribution of TET and their reaction products

Initial quantification of the level of 5hmC has revealed that it is detectable in most of the tissue tested. Unlike 5mC, which is relatively constant and makes up 4% of total cytosine in somatic cells (Ehrlich et al. 1982), the level of 5hmC is lower than of 5mC and varies between tissue types. It is abundant in ES cells and in the brain and constitutes between 0.4 and 0.7% of total cytosine. It is present in moderate levels in the liver, kidney, colon and the lung (0.40-0.65%) and in negligible levels in the heart, breast and placenta (0.05-0.06%) (Kriaucionis and Heintz 2009; Münzel et al. 2011; Kinney et al. 2011; Li et al. 2011; Nestor et al. 2012). Studies have shown that cancer cells often contain low levels of 5hmC, which has become a hallmark of certain type of cancers, for example, in melanoma (Lian et al. 2012).

Contrary to 5hmC, the initial attempts to quantify the levels of 5fC and 5caC were not successful due to their low abundance and they were believed to be recycled rapidly. However, the development of more sensitive techniques has enabled the researchers to quantify the levels of the oxidized bases and they are 10-100 fold lower than the level of 5hmC (0.02-0.002%) (Pfaffeneder et al. 2011; Bachman et al. 2015). Interestingly, despite their low abundance (especially of 5fC), both 5hmC and 5fC could be detected after several cell divisions (Bachman et al. 2014, 2015), suggesting that they might be a stable mark.

To better understand the roles of the oxidized bases, genome wide profiling was performed and studies were conducted in the brain and ES cells because of their high abundance in these cell types. 5hmC is generally enriched in euchromatin, especially at the transcriptional start sites (TSSs), promoters (with moderate and low CG content), and gene bodies in both ES cells and in brain tissue. There is no overlap of 5mC and 5hmC detected on promoters and transcription factors (TFs) binding sites (Szulwach et al. 2011; Jin et al. 2011; Pastor et al. 2011; Xu et al. 2011b; Ficz et al. 2011; Booth et al. 2012; Neri et al. 2013). 5hmC on gene bodies is positively correlated with gene expression in both brain and ES cells (Jin et al. 2011). However, 5hmC on the promoters can have different outcomes on transcription depending on the cellular context (Song et al. 2011b), suggesting that the 5hmC mark can recruit different factors and can regulate the transcription in both ways.

In ES cells, 5hmC is mostly prevalent in developmentally regulated genes marked with bivalency (PRC2 target gene), TF binding regions, active enhancers and CTCF binding sites, but not on housekeeping genes (Robertson et al. 2011; Becker et al. 2011; Yu et al. 2012; Neri et al. 2013), whereas in the brain, 5hmC is present in higher levels on poised enhancers primed for activation than on active enhancers (Wen et al. 2014). Interestingly, the overall level of 5hmC observed around TSSs and gene promoters in neuronal cells are lower compared to ES cells (Szulwach et al. 2011; Song et al. 2011a; Jin et al. 2011), indicating that 5hmC in somatic cells has cell type specific roles. For example, in nerve cells 5hmC is enriched on the promoters of genes responsible for hypoxia, angiogenesis, nerve function and ion transport but not on testis specific genes (Jin et al. 2011; Song et al. 2011a). Interestingly, 5hmC is enriched on the intron-exon boundary of 5’ splicing sites in the brain, suggesting a potential involvement in splicing (Wen et al. 2014).

Similar to 5hmC, the profiling of 5fC in ES cells showed that it is enriched at CpG island (CGI) containing promoters, TSSs (marked with H3K4me3), exons and gene regulatory elements especially at the poised enhancers (Raiber et al. 2012; Song et al. 2013; Neri et al. 2015b). However, the analysis of 5fC in different tissues have suggested that 5fC is enriched at exons and enhancers of actively expressed genes in a tissue type specific fashion (Iurlaro et al. 2016; Wu et al. 2014). The distribution of 5fC is dependent on TDG to some extent, as the depletion of TDG leads to increased 5fC levels (Raiber et al. 2012; Shen et al. 2013; Neri et al. 2015b). Interestingly, 5fC enriched sites show increased binding of p300 (Song et al. 2013) and 5fC

containing promoters are positively correlated to gene expression (Raiber et al. 2012), indicating that the 5fC mark can recruit chromatin factors and may exhibit a distinct regulatory function.

Genome wide mapping of TET1 occupancy showed that there is a significant overlap between the co-occurrence of both 5hmC and TET1 in the genome. In ES cells, TET1 is localised in 60% of bivalent genes marked with both H3K4me3 and H3K27me3 and on the promoter of PRC2 occupied genes carrying H3K27me3, but not on the promoters harbouring H3K4me3 alone (Pastor et al. 2011; Williams et al. 2011). TET1 is present on CG rich promoters with intermediate and low CG content, but not with high CG density (Xu et al. 2011b); for example CG poor promoters of pluripotency inducing TFs like NANOG, ESRRB, TCL1, KLF4 (Wu et al. 2011). The depletion of TET1 in ES cells leads to decreased occupancy of EZH2 - the PRC2 component and hypermethylation of the promoters of pluripotency inducing TFs (Ficz et al. 2011). This suggests that TET1 regulates both the developmentally regulated genes and the TFs inducing pluripotency (Wu et al. 2011). TET enzymes also co-localise with Sin3A independent of 5hmC (Neri et al. 2013). Furthermore, TET and TDG co-localize with 5fC and 5caC in ES cells (Neri et al. 2015b), which suggests that TET-TDG mediated DNA demethylation is necessary for proper differentiation and development.

1.10. Biological functions of TET enzymes

Since the discovery of TET enzymes, numerous studies have been conducted to study the biological role of TET enzymes. Genetic studies strongly suggest that despite having common catalytic activity, the different TET enzymes perform both overlapping and distinct cellular functions in a cell type dependent manner. This is probably partially due to the different expression patterns of the family members and their localization on the genome. TET3 is present in higher levels in oocyte and zygote (Wossidlo et al. 2011) and it starts declining in the inner cell mass (ICM), whereas TET1 and TET2 are highly expressed in both the ICM and ES cells (Ito et al. 2010). After differentiation, the overall level of all three TET enzymes decreases, but they are nevertheless present in a wide variety of somatic cells. For example, TET2 is abundant in hematopoietic cells (Ko et al. 2010) and TET2 and TET3 are upregulated in neuronal cells (Szwagierczak et al. 2010; Santiago et al. 2014).

1.10.1. Genetic studies in ES cells

Genetic studies conducted to elucidate the physiological relevance of TET enzymes have shown that knockout of either TET1 or TET2 in mice has no apparent abnormalities in growth (Dawlaty et al. 2011; Koh et al. 2011). However, mice lacking both TET1 and TET2 have developmental abnormalities with only 40% of mice embryos surviving and the mice contained low 5hmC levels with defects in imprinting (Dawlaty et al. 2013). This suggests that TET1 and TET2 are compensating for one another to some extent. Moreover, the examination of the developmental potential on triple knock out (TKO) ES cells showed that the loss of all three TET enzymes was not lethal, but severely affected the embryonic development and compromised the differentiation capacity of ES cells (Dawlaty et al. 2014). However, the deletion of TET3 in mice led to neonatal lethality and mice with conditional TET3 knockout in oocytes were less fertile and the offspring suffered from developmental failures (Gu et al. 2011), strongly suggesting the importance of TET3 during early development and that its loss cannot be fully compensated by other family members.

1.10.2. Roles of TET enzymes in haematopoiesis

The loss of TET2 in adult hematopoietic cells and in fetal liver (FL) HSC has resulted in a low level of 5hmC, increased self-renewal capacity and expansion of the hematopoietic stem cells (HSC) pool. Moreover, it has been shown to skew the cell differentiation towards the monocyte or macrophage lineage (Ko et al. 2011; Quivoron et al. 2011; Moran-crusio et al. 2012; Kunimoto et al. 2012). Furthermore, Klug et al demonstrated that the loss of TET2 affected the DNA demethylation of enhancers regulating tumour suppressor genes in HSC and a similar effect was also observed in primary monocytes (Klug et al. 2013). These results suggest that TET2 is necessary for the proper development and differentiation of HSC. Interestingly, Cimmino et al showed that the deletion of TET1 in mice decreased the level of 5hmC, increased the level of 5mC and changed the expression of genes involved in B cell lineage specification, chromosome maintenance and DNA repair and resulted in the development of B cell lymphoma with increased self-renewal capacity. (Cimmino et al. 2015). This difference between loss of TET1 and TET2 suggests that these two proteins affect the expression level of different lineage specific genes and may not compensate for loss of one

another, further suggesting that any perturbation in TET expression may lead to a diseased state, especially cancer (Jeschke et al. 2016).

1.10.3. TET in other cell types

The relative abundance of TET and 5hmC in the brain has prompted researchers to investigate the roles of TET enzymes in the brain tissues. The loss of TET1 in mice has affected the gene expression of a group of neuronal activity related genes and shows defects in memory extinction, spatial memory and neuronal plasticity (Rudenko et al. 2013; Zhang et al. 2013). TET1 has been implicated in memory formation and cocaine addiction (Kaas et al. 2013; Feng et al. 2015). TET enzymes have also been implicated in the immune system, where the TET mediated oxidation regulates the expression of cytokines in Th cells (Th1 and Th17) (Ichiyama et al. 2015) and are involved in maintaining the immune homeostasis (Yang et al. 2015b). TET2 has been shown to regulate the muscle cell differentiation through the oxidation of 5mC at *MYOD* and *SRF* genes (Liu et al. 2013b). TET enzymes can act as co-activators or co-repressors independent of their catalytic activity (Tsai et al. 2014; Zhang et al. 2015; Xue et al. 2016), which suggests that apart from the gene activity modulation through the oxidation of 5mC, TET enzymes regulate the gene regulation by interacting with other epigenetic players. Altogether, these studies show that TET enzymes play a diverse roles in a cell type dependent manner and any disturbances can affect the normal function of the cells.

1.11. Involvement of TET in reprogramming

1.11.1. TET enzymes in zygote and PGCs

The presence of active DNA demethylation has been known since 2000, when extensive loss of methylation independent of DNA replication has been observed at several loci in the paternal pronucleus (Mayer et al. 2000b; Oswald et al. 2000). After the identification of TET mediated oxidation of 5mC, it was shown that DNA demethylation could occur through the oxidation of 5mC (Wossidlo et al. 2011). In line with the previous observation, the paternal genome has high levels of 5hmC and low levels of 5mC, whereas the maternal pronucleus shows almost no 5hmC and high levels of 5mC (Wossidlo et al. 2011; Iqbal et al. 2011). Using the conditional knockout mice, it has been shown that the TET3 enzyme present in the oocyte (Gu et al. 2011; Ficz et al. 2011) was responsible for the appearance of 5hmC. Further examination have revealed that one of the polycomb group proteins called PGC7/STELLA is

bound to H3K9me2 marks and protect the methylated loci in the maternal genome and some of the imprinting loci in the paternal genome from oxidation by interacting with the catalytic domain of TET3 enzyme. (Nakamura et al. 2007, 2012; Bian and Yu 2014). However, recent studies conducted in mouse zygotes at different developmental stages have reported that both paternal and maternal genome undergo both active and passive demethylation (Guo et al. 2014; Shen et al. 2014). Detailed kinetics revealed that DNA demethylation in zygote occurs in 2 phases. The first phase is TET independent; while the second phase of DNA demethylation involves both TET and cytidine deaminase (AID/APOBEC) mediated demethylation. Interestingly, the second phase largely coincides with the replication-dependent demethylation (Santos et al. 2013). More recent report have suggested that during the pronuclear (PN) stages PN2-PN3, accumulation of 5hmC does not coincide with the disappearance of 5mC and it occurs only after a major drop in the methylation mark, indicating the presence of TET-independent active DNA demethylation pathway (Amouroux et al. 2016).

In mammals, the second wave of demethylation occurs in primordial germ cells (PGCs), which are derived from the pluripotent epiblast and are precursors of the gametes (Lawson and Hage 1994). Studies conducted so far have suggested that the demethylation in PGCs occurs in 2 phases, which involves both active and passive demethylation (Popp et al. 2010; Seisenberger et al. 2013; Hackett et al. 2013). The first phase of demethylation is believed to be mostly TET-independent and it occurs due to the downregulation of histone H3K9me2 and the DNA methylation machineries (Hajkova et al. 2008; Kurimoto et al. 2008; Yamaguchi et al. 2013). The second phase of reprogramming occurs once the PGCs reache the gonadal ridges and a drastic DNA demethylation occurs with only 2% of DNA methylation mark left which includes imprinted genes, heterochromatin marks and reactivation of the inactive X chromosomes (Hajkova et al. 2010, 2002; Guibert et al. 2012; Seisenberger et al. 2012). TET1, TET2, TDG and BER proteins are highly expressed at this stage (Hajkova et al. 2010; Hackett et al. 2013) and oxidize 5mC to 5hmC which can be lost either through DNA replication (Hackett et al. 2013) or further oxidized to 5fC and 5caC by TET enzymes followed by quick recycling by TDG, indicating the involvement of active DNA demethylation. Genetic studies have shown that TET1 and TET2 have a distinct locus specific functions.in the PGCs, as loss of TET1, both TET1/2 and TDG leads to imprinting defects (Yamaguchi et al. 2012, 2013; Piccolo

et al. 2013), suggesting that TET-TDG mediated demethylation is necessary to establish proper imprinting (Cortellino et al. 2011; Cortázar et al. 2011).

1.11.2. Reprogramming *in vitro*

Reprogramming can be mimicked *in vitro* through different mechanisms, which includes the addition of transcription factors (Takahashi and Yamanaka 2006), somatic cell nuclear transfer (Gurdon et al. 1958) and cell fusion (Reik and Surani 2015). Out of these three, the induced pluripotent stem (iPS) cells generation through the addition of the transcription factors has been quite successful and has been used as a tool to elucidate the molecular mechanism of epigenetic reprogramming.

TET1 and TET2 are implicated in the generation of iPS cells and these cells are reported to have high levels of 5hmC. The depletion of TET1 and TET2 have resulted in the reduced efficiency in reprogramming (Koh et al. 2011; Doege et al. 2012; Gao et al. 2013; Reik and Surani 2015) and thus, it is believed that the TET mediated DNA demethylation is required to remove the barrier set by DNA methylation against pluripotency. However, studies conducted to elucidate the roles of TET enzymes in reprogramming have shown that TET together with TDG is involved in a locus specific DNA demethylation induced by TET enzymes. (Gao et al. 2013; Hu et al. 2014). This is further supported by the observation that the loss of TET1/2 in ES cells grown in 2i media has no significant increase in 5mC levels (Ficz et al. 2013). Moreover, recent investigation on mechanism underlying reprogramming (from serum grown ES cells shifted to 2i media) has revealed an intriguing result that the global DNA demethylation is driven due to the impairment of DNA methylation maintenance machinery and the contribution of active DNA demethylation through TET and AID is surprisingly very limited (von Meyenn et al. 2016). Taken together, all these emerging evidences suggest that TET enzymes trigger the locus specific DNA demethylation that plays a pivotal role in reprogramming (both *in vitro* and *in vivo*), but are not responsible for global DNA demethylation.

1.12. Recruitment of TET enzymes

Despite a tremendous progress in understanding the physiological relevance of TET enzymes and their reaction products, there is limited understanding on the recruitment and the regulation of TET enzymes. The presence of the CXXC domain on the N-terminus of TET1 and

TET3 is believed to be partly responsible for targeting these enzymes to the CpG region, as the CXXC domain containing proteins such as DNMT1, MLL1, CFP1 have been shown to be recruited to unmethylated CpG via the CXXC domain (Thomson et al. 2010; Blackledge et al. 2013). In agreement with this, DNA binding studies have shown that TET1 bound to CpG-rich DNA irrespective of the modification (C, 5mC or 5hmC) (Zhang et al. 2010; Xu et al. 2011b), whereas the TET3-CXXC domain has been reported to bind unmodified C in both CG and non-CG context, with a slightly higher preference for CG (Xu et al. 2012; Jin et al. 2016). Interestingly, a recent study has demonstrated that the TET3-CXXC domain can bind 5caC and moreover, the TET3-FL preferentially binds to the TSS of the genes involved in base excision repair mechanism (Jin et al. 2016). This suggests that TET3 may be specifically targeted to these loci through the CXXC domain or by other interacting proteins (Jin et al. 2016) and oxidize 5mC on the transcriptional start sites (TSSs).

Unlike TET1 and TET3, TET2 lacks the CXXC domain and may depend on other proteins or TFs for locus specific recruitment. Supporting this idea, TET2 has been shown to interact with the transcription factor Wilm's tumor (WT) and Early B cell factor 1 (EBF1) and modulates the target gene expression through its oxidation (Rampal et al. 2014; Guilhamon et al. 2013; Wang et al. 2015).

Likewise, NANOG-dependent recruitment of TET1 and TET2 has also been suggested to promote the expression of genes involved in reprogramming and lineage commitment (Costa et al. 2013). Furthermore, a recent study by Perera et al in mouse retinal cells has demonstrated that RE1-silencing transcription factor (REST) recruits a TET3 isoform lacking the CXXC domain, which then interact with the histone methyltransferase NSD3 and activates its target genes (Perera et al. 2015). TET enzymes have been shown to interact with proteins involved in base excision repair pathway such as TDG, PARP1, MBD4, NEIL etc (Müller et al. 2014). Furthermore, all three TET enzymes are reported to associate with O-linked ß-D-N-acetylglucosamine (O-GlcNAc) transferse (OGT). It has been suggested that TETs recruit OGT to the chromatin and TET-OGT interaction promotes the OGT activity (Chen et al. 2013; Vella et al. 2013; Fujiki et al. 2011). Taken all these results together, it is increasingly clear that TET enzymes do not function alone and they interact with multiple other proteins in a contextual manner and thereby modulate the gene expression.

1.13. Regulation of TET enzymes

TET enzymes can be regulated at both transcriptional and protein level. In ES cells, they are mostly regulated at the transcriptional level by the action of ES cell-specific transcription factors OCT4, NANOG and Myc (Neri et al. 2015a). Also, they are regulated by the binding of miRNA to the 3'UTR of their transcripts in differentiated cells (Fu et al. 2013; Cheng et al. 2013).

In addition, the activity of TET enzymes can be regulated at the protein level. TET2 was shown to be downregulated through IDAX-dependent recruitment of TET2 to the DNA and subsequent activation of caspases (Ko et al. 2013). In addition to caspases, another proteolytic system called calpains was also shown to regulate all TET enzymes (Wang and Zhang 2014). Furthermore, the post translational modifications such as glucosylation (GlnNAc) and phosphorylation could also modulate the levels of TET enzymes, as TET enzymes were shown to be heavily modified in the N-terminus by OGT and not yet identified kinases (Bauer et al. 2015). All TET enzymes can also undergo monoubiquitination on a conserved lysine residues (K1299 in TET2) by the VPRBP-CLR4 complex, which increases the DNA binding to the chromatin, thus acts as a potential regulator of TET enzymes (Nakagawa et al. 2015).

On the other hand, the TET enzymatic activity can be regulated based on the availability of cellular metabolic factors. For example, the availability of αKG (the co-substrate of TETs) can influence the TET-induced oxidation and alters the gene expression (Oermann et al. 2012). Other factors like polyADP-ribose (PAR) and GlcNAc can also regulate TET enzymes (Kim et al. 2005; Le May et al. 2012; Zhang et al. 2014; Ciccarone et al. 2014). Alternatively, extrinsic factors such as vitamin C (AscA) and vitamin A can also influence the activity of TET enzymes, as it was evidenced by several groups (Blaschke et al. 2013; Minor et al. 2013; Yin et al. 2013; Gudas 2013).

1.14. Roles of TET enzymes in cancer

After the functional characterization of TETs in 2009 (Tahiliani et al. 2009), it has become evident that TET enzymes (in particular TET2) are misregulated in a wide variety of cancer (Weissmann et al. 2012). As TET2 is highly expressed in hematopoietic stem cells (HSC), it is often implicated in haematological cancers and undergoes chromosomal alterations and mutations (missense, non-sense, insertion, microdeletion and frameshift) in many myeloid

and lymphoid malignancies (Viguié et al. 2005; Huang and Rao 2014). Majority of the mutations are non-sense mutations (according to COSMIC database), which are accumulated in the catalytic domain, thereby abrogating the catalytic activity (for a detailed review refer Scourzic et al. 2015). In line with this, 5hmC has been reported to be low in numerous haematological cancer (Rampal et al. 2014), as well as in various solid tumours such as in the liver, breast, prostate and the pancreas (Haffner et al. 2011; Lian et al. 2012; Liu et al. 2013a; Yang et al. 2013), especially on the promoters, gene bodies and the regulatory elements (Rasmussen et al. 2015; Jeschke et al. 2016). Genetic studies have demonstrated that the loss of TET2 leads to increased self-renewal capacity of HSC and skew the cells towards monocytic lineage, which eventually result in malignant transformation (Langemeijer et al. 2009; Ko et al. 2010; Kunimoto et al. 2012).

Besides TET2, TET1 and TET3 have also been mutated in cancer, though the frequency is less compared to TET2 (Abdel-Wahab et al. 2009). The decrease or loss of TET1 is implicated in solid tumours such as the liver, colorectal and the prostate cancer (Kudo et al. 2012; Liu et al. 2013a), indicating the tumour suppressor role of TET enzymes. In contrast, TET1 is also found to be upregulated in MLL rearranged leukaemia (Huang and Rao 2014), suggesting that TET1 can act as a tumour suppressor or tumour promotor based on the cellular contexts.

Apart from mutations found in *TET* genes, mutations on the proteins interacting with TET enzymes have also been found in cancer. For example, the TF-WT1 (interacts with TET2 and TET3) is often mutated in mutual exclusivity to TET enzymes (Rampal et al. 2014; Wang et al. 2015). Moreover, mutation on the lysine residue which undergoes monoubiquitination or the residue important for the CLR4-VPRBP complex binding has also been observed in leukaemia (Nakagawa et al. 2015). Alternatively, the misregulation of miRNA that modulate the level of TET enzymes are also found in hepatocellular carcinoma (HCC) and haematopoietic cancer (Cheng et al. 2013; Lin et al. 2015; Chuang et al. 2015). Furthermore, overexpression of IDAX is reported in breast and colorectal cancer that can negatively regulate TET2 protein level (Knappskog et al. 2011).

Finally, mutations on the metabolic genes such as isocitrate dehydrogenase (IDH1/2), succinate dehydrogenase (SDH) and fumarate hydratase (FH) have also been observed in cancer such as gliomas (>85%), melanoma, renal carcinoma and gastrointestinal stromal

tumour (GIST) (Oermann et al. 2012; Guilhamon et al. 2013; Abdel-Wahab et al. 2013). IDH1/2 mutations have been shown to affect the synthesis of αKG by inactivating the enzyme or by acquisition of a new gain of function that leads to the conversion of αKG to 2-HG which acts as a competitive inhibitor to the co-substrate αKG (Figueroa et al. 2010; Xu et al. 2011a; Ye et al. 2013). Similarly, mutations on *SDH* and *FH* gene can lead to the accumulation of succinate and fumarate respectively, which are structurally similar to αKG and inhibit the catalytic activity of αKG dependent dioxygenases including TET enzymes, especially succinate, as it can induce product inhibition (Xiao et al. 2012). Interestingly, a recent study has demonstrated that the accumulation of fumarate due to mutation on *FH* gene reduces the TET-mediated oxidation and thereby induces the epithelial-mesenchymal transition (EMT) in cells, which is one of the phenotypes associated with metastasis and tumour invasion in renal cancer (Sciacovelli et al. 2016) thus, elucidating the potential mechanism of *FH* and SDH mutation-induced carcinogenesis.

Altogether, these results suggest that the negative regulation of TET enzymes affect the 5mC oxidation and in combination with other mutants, they alter the epigenetic landscape and accelerate the disease progression. However, how the redued TET enzymatic activity contributes to the disease progression still remains to be elucidated.

1.15. Biochemical properties of TET enzymes

Although the biological relevance of TET enzymes has been studied extensively in both diseased state and in normal cells, little is known about the biochemical characteristics of TET enzymes. There are only handful of studies conducted on the catalytic properties of TET enzymes. For example, *in vitro* studies on the substrate preference of TET enzymes have suggested that both the mammalian TETs and the nTet prefer CG substrates over non-CG substrates (Hu et al. 2013; Hashimoto et al. 2014). This is agreeable as TET enzymes catalyse 5mC that occurs predominantly in CpG dinucleotide, although 5mC occurs in non-CG context in ES cells and in the brain (Ku et al. 2011). In line with this, 5hmC has also been found in non-CG context (Ficz et al. 2011; Lister et al. 2013) indicating that TET enzymes may also able to oxidize the substrate bases in non-CG context. However, there is no comprehensive analysis of TETs substrate specificity and flanking sequence available. Based on the crystal structure studies, it was conceived that TET enzymes showed almost no specificity beyond the CpG

dinucleotide (Hu et al. 2013). Using *in vitro* studies, it was demonstrated that hTET2-CD exhibited no difference in activity between CG and AT rich substrates in CpG dinucleotide context (Hu et al. 2015). However, it should be mentioned that the experiments performed on two substrates with saturating conditions with regard to TET enzyme over the substrate do not allow probing of difference sequences flanking the target substrates. Hence, proper biochemical experiments encompassing different flanks should be done to have a detailed insight into the sequence preference.

Another important characteristic of an enzyme is the inter and intra site processivity. Studies performed to analyse the kinetic behaviour of TET enzymes using MALDI-TOF suggested that TET oxidized 5mC to 5caC in an iterative manner i.e., without the enzyme dissociating from the substrate between the oxidation steps (Crawford et al. 2016). But this result was contradicted by another group, where they observed the opposite trend in which the oxidation of 5mC to 5caC occurred in a distributive manner i.e. the enzyme dissociates from the target 5mC between different oxidation steps and additionally they showed that TET enzymes, including the evolutionarily distant homologue nTet catalysed the multiple substrates on single DNA strand in a distributive manner (Tamanaha et al. 2016). Although these experimental evidences provide useful insights into the biochemical properties of TET enzymes, there are many unexplainable observations that need to be addressed still. For example, which factors determine the processivity on the substrate? Are TET enzymes insensitive to the sequence flanking CpG dinucleotide? Are there any inherent differences among the family members which determine the processivity of the substrate? Therefore, detailed biochemical investigations are needed to examine the catalytic characteristics of TET enzymes.

2. AIMS OF THE STUDY

2.1. Setting up an *in vitro* system to study the biochemical properties of TET enzymes

The first aim of this thesis was to establish a proper *in vitro* system that could be used to investigate the properties of TET enzymes that govern their function, and would help us to understand the molecular mechanism of this group of enzymes.

2.2. The effect of divalent metal ions on the activity of TET enzymes

As TET enzymes require Fe^{2+} for the catalysis, the aim of this project was to investigate the effect of divalent metal ions on the activity of TET enzymes both *in vitro* and *in vivo*.

2.3. Investigation of the role of ascorbic acid and retinol on the activity of TET enzymes

The main aim of this study was to elucidate the mechanism by which AscA (vitamin C) and retinol (vitamin A) influence the catalytic activity of TET enzymes *in vitro*.

2.4. Investigation of the biochemical and kinetic behaviour of TET enzymes

The aim of this work was to study the catalytic and kinetic behaviour of TET enzymes on a linear double stranded DNA containing multiple 5hmC substrates embedded in different sequence contexts in competition to one another, and to analyse the substrate specificity and the flanking sequence preference of both the mammalian TET enzymes and the Tet-like dioxygenase from *Naegleria gruberi*.

3. MATERIALS AND METHODS

3.1. Cloning of the catalytic domain of TET enzymes into bacterial expression vector pET28a (+)

The catalytic domain of mouse TET1 (mTET1-CD), TET2 (mTET2-CD) and TET3 (mTET3-CD) were cloned into the bacterial expression vector pET28a (+) using the conventional restriction enzyme cloning. The inserts were amplified from pFastBac vectors containing the catalytic domains of the respective TET enzymes, kindly provided by Dr. Yi Zhang (Harvard Hughes Medical Institute, New York, USA). The details of the primers used for the amplification are listed in table 2. After amplification, all the three inserts and the pET28a (+) vector were digested with the respective enzymes. The Tet1-like dioxygenase from *Naegleria gruberi* (nTet) was cloned by digesting the synthetic construct containing nTet (generated based on the accession number XP_002667965.1, MWG operon, Germany) that was delivered in a pEX-A2 vector (Table 2). The digested inserts and the vector were ligated using T4 DNA ligase (Fermentas) following the manufacturer's protocol. The ligated products were cleaned using a PCR clean up kit (Macherey Nagel), transformed into an electro-competent *E.coli* strain XL-1 BLUE MRF' and were grown on LB-agar plates supplemented with appropriate antibiotics overnight, at 37 °C. The colonies were screened for the presence of insert and the positive clones were verified further by Sanger sequencing (MWG-Operon, Germany).

Table 2: Details of the primers and restriction sites that are used for cloning the catalytic domains of TET enzymes into the bacterial expression vector, pET28a (+)

Construct name	Forward primer 5'--> 3'	Reverse primer 5'--> 3'	Restriction site
mTET1-CD	GCTTCGCATATGGAAGCTGCACCCT GTGACTGTG	AGTCTTGCGGCCGCTTAGACCCAACG ATTGTAGGGTCC	NdeI/NotI
mTET2-CD	TACCAGTCAGCCAAAGTCAGAATGG CAAATGTGAAG	TCGCCTGAATTCTCATACAAATGTGTT GTAAGGCCCTG	EcoRI/NheI
mTET3-CD	TACCACATATGCGCCCTTTAGAGTTC CCTACCTG	CGCCTGAATTCCTAGATCCAGCGGCT GTAGGG	NdeI/EcoRI
mTET1-CD-KRAK	GGTGGGTCGGGTGGCGGTGGCAGC GAGGACGATAACTTGCCTCAACTTG	GCCGCCACTCCCGCCACCTCCTTTGGC CCTCTTCCCACATCG	NA
mTET2-CD- Del5	CAGAGCGATAACGCGATCGCCATGT CAGACCACGTGTCTCAGAAAAATC	GCCGCCACTCCCGCCACCTCCGGAAT GCTTTCGAGCTGCTTTCTTGGTCTTGG C	NA
mTET3-CD-Del1	GGTGGGTCGGGTGGCGGTGGCAGC GAACTGTGGTCGGACAGTGAACAC	GCCGCCACTCCCGCCACCTCCGGACTT GGCAGGCTCAGGCA	NA
nTet	-	-	NdeI/XhoI

For generation of the shorter version of the catalytic domains lacking the unstructured regions, the borders were selected based on the multiple sequence alignment (Figure 7). The constructs were generated following the ExSite method using pET28a (+) vector containing the respective insert as the template and the unstructured region was replaced with the linker rich in glycine and serine (GGGGSGGGGSGGGGS) of length 15 amino acids (aa). For details regarding the border of the constructs, refer table 8.

3.2. Expression and purification of the catalytic domain of TET family of enzymes

pET28a (+) vector containing the catalytic domain (the complete catalytic domain or the shorter version) was transformed into the *E.coli* expression strain BL21 (DE3) CodonPlus RIL (Novagen). After successful transformation, a single colony was selected and inoculated into 20 ml of Luria Bertani (LB) media containing the antibiotic kanamycin (25 µg/ml). After the cells were grown at 37 °C for ~7-8 hours, the pre-culture was transferred into a fresh flask containing one litre of LB media (1:1000 dilution) with kanamycin (25 µg/ml) and 1X trace metals solution (the trace metals solution was prepared as mentioned in Studier 2005). The cultures were grown at 37 °C until it reached the OD_{600} of ~ 0.6-0.8, subsequently the cultures were transferred to 20 °C and the protein expression was induced by adding isopropyl ß-D-1-thiogalactpyranoside (IPTG) at a final concentration of 0.5 mM. The cells were further grown at 20 °C for ~14-15 hours (overnight). After overnight induction, the cells were harvested by centrifugation (Sorvall Lynx 6000, Thermo Scientific) and washed once with the STE buffer containing 10 mM Tris-HCl pH 8.0, 1 mM EDTA and 100 mM NaCl to remove the left over LB media. The harvested cells were stored at -20 °C until further use.

For the purification procedure, the cells containing His_{6x}-tagged TET protein were thawed on ice and suspended in the sonication buffer supplemented with the protease inhibitor cocktail (1 mM of AEBSF, 0.8 µM of aprotinin, 50 µM of bestatin, 15 µM of E-64, 5 mM of EDTA, 20 µM of leupeptin, 10 µM of pepstatin A, prepared in-house). Next, the cells were lysed by sonication (25-30 cycles, 40% power, 15 s ON and 45 s OFF, Bandelin Sonoplus sonicator) and the lysate was clarified by centrifugation at ~38,800 g for 75 minutes using Sorvall Lynx 6000 centrifuge (Thermo Scientific). The cleared lysate was loaded onto pre-equilibrated column containing the Ni-NTA agarose beads (Genaxxon, Germany). Afterwards, the beads containing the protein were washed with ~150-200 ml of wash buffer and eluted using a high

concentration of imidazole (Table 3-4). The highly concentrated fractions of the eluted protein were pooled together and dialysed against the dialysis buffer I for 3 hours. The dialysed protein was aliquoted, flash frozen in liquid nitrogen and stored at -80 °C until further use. The optimized buffer condition used for the purification of different TET proteins is listed in table 3-4.

Table 3: The buffer conditions used for the purification of mammalian TET proteins as His_{6x} fusion

Components	Sonication/wash buffer	Elution buffer	Dialysis I
HEPES, pH 6.8	50 mM	50 mM	50 mM
Imidazole	35 mM	300 mM	0 mM
DTT	1 mM	1 mM	1 mM
NaCl	500 mM	500 mM	300 mM
Glycerol	10 %	10 %	10 %

Table 4: The buffer condition used for the purification of nTet as His_{6x} fusion

Components	Sonication/wash buffer	Elution buffer	Dialysis I	Dialysis II
Kpi pH 7.4	20 mM	20 mM-	-	-
Bis Tris, pH 7.4	-	-	20 mM	20 mM
Imidazole	20 mM	300 mM	-	-
KCl	500 mM	500 mM	250 mM	250 mM
DTT	1 mM	1 mM	1 mM	1 mM
Glycerol	10%	10%	10%	50%

3.3. Western blot analysis of the purified TET proteins as His_{6x} tag fusion

The quality of the purified proteins was verified using western blot analysis. The purified proteins were resolved on SDS-PAGE (12%) and the separated proteins were transferred onto nitrocellulose (NC) membrane (GE Healthcare) using the GoRun Trans Blot Turbo transfer system (BioRad). The successful transfer was verified using Ponceau staining. Afterwards, the membrane was washed with PBST (137 mM NaCl, 2.7 mM KCl, 10 mM NaH2PO4, 2 mM K2HPO4, pH, 7.4) and was blocked overnight with 10% skimmed milk (Roth). The membrane was washed once with PBST and the primary antibody (anti-His, 1:2000, Roche) was added and incubated at room temperature for 2 hours on a rocking platform. Subsequently, the membrane was washed 3 times for 5 minutes with PBST and incubated with the secondary antibody (anti-mouse-HRP from goat, 1:10,000, GE Health care) conjugated to horse radish peroxidase (HRP) for one hour at room temperature. Following the incubation with secondary

antibody, the washing steps were repeated and the membrane was developed using enhanced chemiluminescence (ECL) solution (Thermo Scientific).

3.4. Generation of the substrates used in this study

3.4.1. Preparation of the substrates carrying biotin for the ELISA based plate assay

For the ELISA based plate assay, the substrate of length 86 bp was amplified from pUC19 vector using the primers DpnII_pUC19_f: 5'-GAGTAAACTTGGTCTGACAGTTACCA-3' and DpnII_pUC19_r: 5'-CAACTATGGATGAACGAAATAGACAGAT-3' by PCR using Taq DNA polymerase (cycling conditions: step 1: 95 °C-3 minutes, step 2: 95 °C-30 seconds, step 3: 58 °C-40 seconds, step 4: 72 °C-15 seconds, go to step 4-30 times, step 5: 72 °C-5 minutes, step 8: 4 °C- hold). The forward primer carried biotin molecule at its 5' end. For the substrates containing 5mC, 5hmC, 5fC and 5caC, dNTPs mixture containing the respective modified nucleotides (5mdCTP, 5hmdCTP (NEB), 5fdCTP or 5cadCTP (TriLink technologies)) was included instead of the dCTP for PCR amplification. The successful amplification was verified using 10% native acrylamide gel electrophoresis and purified using a PCR purification kit (Macherey Nagel) according to the manufacturer's protocol.

3.4.2. Preparation of the CG-rich substrate

The CG-rich substrate of size 176 bp containing 22 CG sites was prepared from the DNA template that was originally amplified from CpG island upstream of the human gene *SUHW1* (the DNA template was kindly provided by Dr. Renata Jurkowska, currently in BioMedX, Heidelberg, Germany), using the primers CGRich_noCG_F: 5'-CATCATCCCCAAGGCCTTCC-3' and CGRich_NoC_R: 5'-CCCTCCTCCTTCTCAATTTAACCC-3'. The substrates carrying different modifications were amplified by PCR using the dNTPs mixture containing the respective modified dCTPs (5mdCTP, 5hmdCTP (NEB), 5fdCTP or 5cadCTP (TriLink technologies)) instead of dCTP using Taq DNA polymerase, betaine (1 M) and DMSO (5 %). The cycling conditions are as follows, step 1: 95 °C-5 minutes, step 2: 95 °C-30 seconds, step 3: 55 °C-35 seconds, step 4: 72 °C-45 seconds for 5hmC and 1 minute for 5mC, 5fC and 5caC, go to step 4-30 times, step 5: 72 °C-5 minutes, step 8: 4 °C- hold The amplified PCR products were purified using a PCR clean up kit (Macherey Nagel) following the manufacturer's protocol and used in the subsequent experiments.

3.4.3. Preparation of the substrate for the processivity analysis of TET enzymes using the restriction cleavage method

The processivity of TET enzymes was tested on the DNA substrate of length 120 bp containing two CCGG sites, amplified from pET28a (+) vector by PCR with the primers Proc_TET_F: 5'-CCCTGACGGGCTTGTCTGC-3', Proc_TET_R: 5'-ATGAGCTTTACCGCAGCTGCC-3' using Taq DNA polymerase (cycling conditions: step 1: 95 °C-3 minutes, step 2: 95 °C-30 seconds, step 3: 68 °C-35 seconds, step 4: 72 °C-20 seconds, go to step 4-30 times, step 5: 72 °C-5 minutes, step 8: 4 °C- hold). After the successful amplification, the internal C of two C<u>C</u>GG sites were methylated using HpaII methyltransferase (NEB) according to the manufacturer's protocol. The methylated substrate was purified using a PCR clean up kit (Macherey Nagel) as per the manufacturer's protocol and the methylation of the DNA was validated by HpaII (NEB) cleavage to ensure complete methylation.

3.4.4. Preparation of the substrate used for Liquid Chromatography-Mass Spectrometry (LC-MS)

To quantify the amount of the oxidized bases generated after TET treatment using LC-MS, the substrate carrying a hemi-modified CpG site was prepared by annealing the complementary synthetic oligonucleotides (Integrated DNA Technologies, IDT). Upper strand: 5'-GGCGCGGCTC**<u>X</u>**<u>G</u>AGCTAGGAG-3', Lower strand: 5'-CTCCTAGCTC**<u>G</u>**GAGCCGCGCC-3', X= 5mC or 5hmC. For this, 20 µM of each oligo was added to the solution containing 10 mM HEPES, pH 7.4, 50 mM NaCl in a total volume of 100 µl, heated to 85 °C for 5 minutes and cooled down slowly to room temperature. The annealed DNA was then stored at -20 °C until further use.

3.5. The enzymatic activity of TET proteins

3.5.1. The reaction condition for the mammalian TET proteins

To analyse the activity of mammalian TET enzymes, 0.15-0.5 µM of the substrate DNA was incubated with 0.5-3 µM of the respective TET enzymes in a reaction mixture containing HEPES buffer at pH 6.8 (unless otherwise specified), varying concentration of Fe^{2+} or Fe^{3+}, 1 mM αKG, 1 mM ascorbic acid, 150 mM NaCl and the reaction was performed at 37 °C. Aliquots from the reaction mixture were taken at different time points and the 5hmC/5fC signal was assessed, as described below.

3.5.2. The reaction condition for nTET

To analyse the activity of nTet enzyme, the respective substrate DNA (0.4-0.5 µM) was incubated with 0.15-3 µM of nTet in a reaction mixture containing Bis-Tris, pH 6.0, 75 µM Fe^{2+}, 1 mM αKG, 1 mM ascorbic acid, 100 mM NaCl and the reaction was performed at 34 °C.

3.6. Detection of the oxidized DNA bases using the ELISA based plate assay

To quantify the amount of 5mC oxidation catalysed by TET enzymes, an ELISA based plate assay, which relies on biotin-avidin interaction, was developed in this study. The 96-well microtiter plate (Lumitrac 600, high binding, Greiner Bio One) was coated with avidin (1.3 µg of avidin from egg white per well, Sigma Aldrich) suspended in 0.1 M $NaHCO_3$, pH 9.6 for at least 12 hours. Before loading the DNA containing biotin molecule, the wells were washed three times with PBST (137 mM NaCl, 2.7 mM KCl, 10 mM NaH2PO4, 2 mM K2HPO4, pH, 7.4) supplemented with 500 mM NaCl (PBST-500). After washing, the wells were filled with stop solution containing 100 µl of 0.05 M NaOH. 2 µl aliquots containing 0.36 pmole (20 ng) of the modified DNA or the DNA treated with TET enzymes was loaded into each well. The biotinylated DNA was incubated with avidin coated plate for 1.5 hours on a shaking platform and washed three times with PBST-500 (PBST supplemented with 500 mM NaCl). After washing, the wells were blocked with 300 µl of 2% BSA (Roth) dissolved in PBST for an hour. Following blocking, the wells were incubated with 100 µl of the primary antibody against specific mark (primary antibody, dissolved in PBST). After 1.5 hours of incubation with the primary antibody, the wells were washed again, and the secondary antibody conjugated to HRP (for antibodies detail- refer table 5) was added and incubated at room temperature for an hour. Following the secondary antibody incubation, the washing steps were repeated and the signal was developed using ECL solution using the 2300 EnSpire Multimode plate reader (Perkin Elmer).

Table 5: Details of antibodies used for the ELISA based plate assay

Target	Antibody	Dilution	Company
5hmC	Primary-rabbit Secondary-anti rabbit developed in goat	1:10,000 1:5000	Active motif GE Healthcare
5fC	Primary-rabbit Secondary- anti rabbit developed in goat	1:2500 1:5000	Active Motif GE Health care
5caC	Primary-rabbit Secondary- anti rabbit developed in goat	1:2000 1:5000	Active motif GE Healthcare

3.7. Generation of the standard curve

The standard curve to calculate the amount of DNA needed for the plate assay was done by adding different amounts of the respective DNA (86 bp 5hmC or 5fC modified DNA, amplified from pUC substrate-refer 3.4.1 for details). DNA ranging from 0.001 ng to 10 ng was prepared in a solution containing PBST with 2% BSA and added to separate wells. After adding the DNA, the procedure, as described in section 3.6, was followed and an antibody against 5hmC or 5fC was used. The obtained chemiluminescence was plotted as a function of the amount of DNA used and fitted using the linear regression in MS-Excel.

3.8. The reaction condition used to study the inhibitory potential of αKG analogues and metal ions on the activity of TET enzymes

To check the effect of inhibitors on the TET enzymatic activity, the reaction was performed with 0.4 µM of fully methylated DNA (86 bp), 2 µM mTET1-CD and other components as mentioned in section 3.5.1 with little changes. For the αKG analogues and the oncometabolites, 0.6 mM of αKG and inhibitors such as N-oxalylglycine (NOG), dimethyl oxalylglycine (DMOG), L-α-hydroxyglutaric acid disodium salt (LHG), D-α-hydroxyglutaric acid disodium salt (DHG), sodium succinate dibasic and disodium fumarate (all purchased from Sigma-Aldrich) were used in the reaction mixture at a concentration of 0, 1, 2, 4, 8, 16, 32, 64 mM. For the metal ions, 10µM of Fe^{2+} and divalent metal ions such as cadmium chloride (Roth), calcium chloride (Roth), cobalt (II) chloride hexahydrate (Sigma-Aldrich), manganese chloride tetrahydrate (Roth), zinc chloride (Sigma-Aldrich) and nickel sulfate hexahydrate (Roth) were used in the reaction at concentrations such as 0, 0.1, 1, 10, 100, 1000 µM. The oxidation of 5mC to 5hmC in the presence of various concentrations of the inhibitors was measured using the ELISA based plate assay (as described in 3.6). The initial velocity of each reaction was calculated using the linear regression and the rate of the reaction was plotted as a function of the inhibitor concentration. The apparent IC50 value was calculated by least-square fitting of the obtained data using MS-Excel following the equation (Ci) = BL +100 *IC50/ (IC50+Ci), which follows the simple inhibition of enzyme by an inhibitor. BL represents baseline, Ci is the concentration of the inhibitor, IC50 is the concentration of the inhibitor to achieve 50% activity of the enzyme compared to the control where no inhibitor was added.

3.9. Treatment of the mammalian cells with nickel ions

The effect of nickel ions was tested on mammalian cells such as HEK293 and E14 mouse embryonic stem (E14-mES) cells. HEK293 cells were grown in DMEM media (Sigma-Aldrich) containing 10% fetal bovine serum (Sigma), 100 units/ml penicillin, 100 μg/ml streptomycin (Sigma), 4 mM glutamine (Sigma) at 37 °C, 5% CO_2 for maximum of 5 days. The mouse ES cell line E14 (129/Ola) (a kind gift from Wolf Reik's laboratory, Babraham Institute, Cambridge, UK) were grown in a feeder-free conditions at 37 °C and 5% CO_2 in complete ES medium consisting of DMEM, 4.5 g/litre glucose, 4 mM L-glutamine and 110 mg/litre sodium pyruvate, 15% fetal bovine serum, 100 units/ml penicillin, 100 μg/ml streptomycin, 0.1 mM non-essential amino acids, 50 μM β-mercaptoethanol, 103 units of LIF (Cambridge Stem Cell Institute).

Equal amount of the cells were seeded in T25 flasks and the cells were grown until it reached 70% confluency. The cells were then treated with different concentrations of $NiCl_2$ (0-800 μM) and were allowed to grow at 37 °C, 5% CO_2 for maximum 5 days. The medium was exchanged on alternate days supplemented with the appropriate concentration of $NiCl_2$. The cells were harvested after day 1 and day 5 and used for further analysis.

3.10. Dot blot to detect the levels of 5hmC in $NiCl_2$ treated cells

Dot blot assay was used to quantify the changes of cellular 5hmC levels upon metal ion treatment. The genomic DNA was extracted from the cells (HEK293 cells or mES cells) treated with different concentrations of nickel ($NiCl_2$) using the QIAamp DNA mini kit (Qiagen) following the manufacturer's instructions. The extracted DNA (with serial dilution) was spotted on the Hybond N+ membrane (Amersham) using the dot blot apparatus (Bio-Rad). The membrane was dried at 80 °C for 30 minutes followed by blocking with 10% skimmed milk in PBST for an hour. Afterwards, the primary antibody (anti-5hmC, Active Motif, 1:10000-diluted in 1% skimmed milk in PBST) was added and the membrane was incubated overnight at 4 °C on a rocking platform. Subsequently, the membrane was washed 3 times for 5 minutes with PBST and incubated with the secondary antibody (anti-rabbit-HRP, 1:10000 diluted in 1% skimmed milk in PBST) at room temperature for an hour on a rocking platform. The signal was detected using ECL (Thermo Scientific). Later, the membrane was stained with methylene

blue solution (0.3 M sodium acetate, pH 5.5, 0.02% methylene blue-from Riedel-De Haen) and used as loading control.

3.11. Isolation of RNA, preparation of cDNA and qPCR from HEK293 cells

HEK293 cells treated with different concentrations of $NiCl_2$ (0, 10, 30, 100, 300 µM) was harvested after 5 days and RNA was isolated from the cell pellets using RNeasy Plus Mini kit (Qiagen). 500-1000 ng of RNA was converted into cDNA using M-MuLV reverse transcriptase (NEB) and RT-qPCR was performed using the oligos and the cycling conditions as described in table 6 using Bio-Rad CFX connect Real–Time System. The cycle number differences among the $NiCl_2$ treated samples were normalized to the reference gene REEP5 using ΔΔCt analysis.

Table 6: The oligonucleotides used to quantify the level of mRNA expression of TET1, TET2 and TET3 in HEK293 cells

Target	Forward Primer (5'-->3')	Reverse primer (5'-->3')
hTET1	TACCGACAGAAGATGCACCC	GGGCTTGGGCTTCTACCAAA
hTET2	GGCTGACAAACTCTACTCGGA	AAAGAGAAGGAGGCACCACAG
hTET3	CTGGAAAACTGTGGCGCTTG	GGCTCCTTCACCAGCCTTTAT
REEP5	TTTGGCTACCCAGCCTACATC	CAACAGGAAGCCACACTTCAGC

Cycling condition: Step 1: initial denaturation 95 °C-3 min, Step 2: 95 °C-10 sec, Step 3: 61 °C-30 sec, repeat step 2-3 for 40 times, 95 °C-10 min, melting curve- 65 to 95 °C, with 0.05 increments for 5 sec. The reaction was performed using SsoFast EvaGreen supermix (BioRad) according to manufacturer's protocol.

3.12. *In vitro* kinetics to study the effect of retinol and different reducing agents on the activity of TET enzymes

The effect of different reducing agents, retinol and retinoic acid on the activity of TET enzymes was analysed by adding the appropriate amount of the supplements (Table 7) to the reaction mixture containing 0.4 µM DNA substrate (86 bp), 0.5-2 µM TET enzymes (mTET1-CD and mTET2-CD-Del5- 2µM and mTET3-CD-Del1- 0.5 µM) and the reaction was performed as described in the section 3.5.1. For the pre-incubation studies, the reaction was initiated by adding the enzyme and other components after incubating the iron (Fe^{2+} or Fe^{3+}) with the appropriate reducing agent for 30 minutes. The oxidation of 5mC to 5hmC was measured using the ELISA based plate assays as mentioned in the section 3.6.

The initial rate of the reaction (retinol, retinoic acid and other reducing agents) was calculated using the linear regression and the overall activity was normalized to the control reaction containing no supplements under Fe^{2+}or Fe^{3+} conditions.

Table 7: Details of the concentration of different reducing agents and vitamin A derivatives used

Reducing agent and vitamin A derivatives	Final concentration	Company
ß-Mercaptoethanol (B-ME)	1 mM	Roth
Dithiothreitol (DTT)	1 mM	Roth
Hydroquinone (HQ)	50 µM	Sigma-Aldrich
TCEP (tris-(2-carboxyethyl)phosphine)	1 mM	Sigma
Potassium Iodide (KI)	10 µM	Roth
L-glutathione (L-glu)	1 mM	AppliChem
L-ascorbic acid sodium salt (AscA)	1 mM	Sigma-Aldrich
D-Isoascorbic acid (D-AscA)	1 mM	Sigma
N-acetyl L-cysteine (NAC)	1 mM	Sigma-Aldrich
Retinol	0.2 or 100 µM	Sigma-Aldrich
Retinoic Acid	0.2 or 100 µM	Sigma-Aldrich

3.13. Kinetics to probe the uncoupled oxidation reaction

To check the uncoupled oxidation, the reaction was performed with 2 µM mTET1-CD, as described in the section 3.5.1, with little changes. The enzyme and other reaction mixture except the substrate DNA was incubated at room temperature for different time points such as 0, 1, 5, 10 and 30 minutes. The reaction was initiated by adding 0.4 µM substrate (86 bp fully methylated DNA) and the kinetics was performed for 20 minutes. Samples were taken at defined time points and the oxidation of 5mC to 5hmC was measured using the ELISA based plate assay (refer section 3.6).

3.14. The competition assay to check the binding affinity of Fe^{3+}

The binding affinity of Fe^{3+} to the catalytic pocket of TET enzymes was analysed using the competition assay. The reaction was performed, as described in the section 3.5.1, with 0.4 µM substrate, 10 µM Fe^{2+} and increasing concentration of Fe^{3+} (0, 1, 5, 10, 50, 100, 1000, 2000, 5000 µM). The activity of mTET1-CD in the presence of different concentrations of Fe^{3+} was analysed using the ELISA based plate assay (refer section 3.6). The initial rate of the reaction was calculated using the linear regression in MS-Excel and the overall activity was normalized to the reaction containing no Fe^{3+}.

3.15. Measurement of the amount of Fe^{2+} using ferrozine method

To quantify the amount of Fe^{2+} present in the reaction mixture, the ferrozine method (Verschoor and Molot 2013) was used. 10 µM Fe^{2+} was added to the buffer solution containing

50 mM HEPES either at pH 6.8 or pH 8.0. The reaction mixture was placed at 37 °C and 500 µl of the reaction mix was withdrawn from the reaction mix at different time points and added to the stop solution containing 100 µl of 0.5 M ferrozine (Sigma-Aldrich) and 100 µl of 2 M ammonium acetate, pH 6.0, which prevents the further oxidation of Fe^{2+} during the course of the reaction by decreasing the pH of the mixture. The reaction mixture was incubated at room temperature for 20 minutes to ensure full colour development. The absorbance of the reaction mixture at different time points was measured at 562 nm using 10-mm glass cuvettes with uv-vis spectrophotometer (Hitachi U-2810 spectrophotometer). The solution containing 50 mM HEPES, pH 6.8 or pH 8.0 without Fe^{2+} was used as baseline. The standard curve was generated by measuring the absorbance at several defined concentration of Fe^{2+}. Based on the standard curve, the concentration of Fe^{2+} was derived and plotted as a function of time. The data was fitted using single exponential function in MS-Excel.

3.16. Measurement of the intrinsic fluorescence of mTET1-CD using fluorescence spectrophotometer

The binding of ascorbic acid to mTET1-CD was tested using the fluorescence quenching, as described in Yin et al. 2013. Increasing concentration of ascorbic acid was added to the reaction mixture containing all components, as mentioned in the section 3.5.1 and 2 µM mTET1-CD except the substrate DNA. The reaction mixture was incubated at room temperature for 20 minutes and the fluorescence intensity was detected using F-2500 Fluorescence spectrophotometer. As a negative control, unrelated proteins such as EcoDam, CcrM and hDNMT2 were used at a concentration of 2 µM, and their intrinsic fluorescence in the presence of ascorbic acid was measured.

The samples were excited at 280 nm and the fluorescence spectrum was measured between 290-400 nm with excitation and emission bandwidth set at 2.5 nm, scan speed at 200 nm/min and integration time set to 0.2 seconds. The obtained fluorescence at different ascorbic acid concentrations was plotted as a function of wavelength using MS-Excel.

3.17. The binding assay to measure the apparent dissociation constant of Fe^{2+}

To determine the binding affinity of Fe^{2+} to TET enzymes, enzymatic reactions were performed using 2 µM mTET1-CD with increasing concentration of Fe^{2+} (0, 0.25, 0.5, 1, 2, 4, 6, 8, 10, 50, 100 µM), as described in the section 3.5.1. The initial rate of each reaction was

determined from the kinetics using linear regression. The apparent K_d value was calculated using the least square fit to bimolecular binding equilibrium model. As the purified mTET1-CD from bacteria contained co-purified Fe^{2+}, an additional variable was included in the calculation to taken into account for the bound Fe^{2+}.

3.18. The inhibitory effect of ascorbic acid on the activity of TET enzymes

To analyse the inhibitory effect of ascorbic acid, different concentrations of ascorbic acid such as 0, 0.01, 0.05, 0.2, 0.5, 1, 2, 5, 10 mM was added to the reaction mixture and the reaction was carried out, as described in the section 3.5.1, using 2 µM of mTET1-CD. The initial rate of the reaction was extracted from the kinetic data for each ascorbic acid concentration using the linear regression. The rate of the reaction was plotted as a function of ascorbic acid concentration, and the apparent IC50 value was calculated by the least-square fitting of the data in MS-Excel using the equation (Ci) = BL +100 *IC50/(IC50+Ci), which follows the simple inhibition of the enzyme by an inhibitor. BL represents baseline, Ci is the concentration of the inhibitor, IC50 is the concentration of the inhibitor to achieve 50% activity of the enzyme compared to the control, where no inhibitor was added.

3.19. *In vitro* reaction condition of TET enzymatic activity on the CG-rich substrate

The activity of TET enzymes on the CG-rich substrate containing 5mC or5hmC was performed using 0.15 µM of 5mC or 5hmC-modified substrate DNA (refer section 3.4.2) incubated with 2 µM of mTET1-CD, mTET2-CD-Del5, mTET3-CD-Del1 and the other reaction components, as described in the section 3.5, in a total volume of 40 µl for different time points such as 1, 5 and 40 minutes. After TET reaction, the DNA was incubated with 0.3 µl of proteinase K (NEB, 800 units /ml)) for an hour at 50 °C and the proteinase K was inactivated at 70°C for 20 minutes. The DNA was purified using a DNA purification kit (Macherey Nagel). Following the DNA purification, the methylated TruSeq LT Illumina adapters were attached using T4-DNA ligase (Fermentas) following the manufacturer's protocol. The DNA with the adapters was subjected to bisulfite conversion using the EZ-DNA Methylation-Lightning kit (Zymo Research). The bisulfite treated samples were amplified using the primers Kapa-f: 5'-AATGATACGGCGACCACCGA-3' and Kapa-r: 5'-CAAGCAGAAGACGGCATACGA-3'. The successful amplification was verified using electrophoresis and quantified by NEBNext kit

using the primers mentioned above. The samples were pooled together and sent for sequencing on an Illumina MiSeq 2x300 bp.

3.20. Kinetics performed to quantify the amount of oxidized bases using LC-MS

To quantify the amount of 5hmC, 5fC and 5caC catalysed by the TET enzymes using LC-MS, 0.5 µM of the hemi-methylated or hemi-hydroxymethylated DNA (refer section 3.4.4) was treated with 3 µM TET enzymes (mTET1-CD, mTET1-CD-KRAK, mTET2-CD-Del5, mTET3CD-Del1, nTet) and the reaction was performed, as described in the section 3.5, in a total volume of 280 µl. At defined time points (1, 5, 10, 20 and 40 minutes), 40 µl of the sample was taken out and the reaction was stopped by adding 5µl of 100 mM EDTA solution and flash frozen in liquid nitrogen. The enzyme was then inactivated at 95 °C for 5 minutes and subsequently treated with 1 µl of proteinase K (NEB, 800 units/ml) at 50°C for an hour. The DNA was purified using ethanol precipitation and analysed using LC-MS.

3.21. Detection of TET activity using the restriction digestion method

To analyse the conversion of 5mC to 5hmC, 5fC and 5caC by TET enzymes using the restriction cleavage method, 0.6 µM of M.HpaII methylated DNA containing two CCGG sites was treated with 1-3 µM of TET enzymes using the *in vitro* reaction condition, as described in the section 3.5. After TET treatment, the DNA was glucosylated using T4 ß-glucosyltransferase (T4 BGT) (2 µM of in house purified T4-BGT) with 40 µM of uridine phosphate glucose in the CutSmart buffer (from NEB) at 37 °C for 2 hours. Following the glucosylation reaction, the DNA was digestion with MspI (NEB) and analysed using 10% native acrylamide gel.

3.22. Bioinformatics analysis of the sequencing results

The quality control check on the next generation sequences (NGS) in fastq format was done using the quality control tool FastQC (Babraham Bioinformatics). The adapters from the sequences were trimmed and quality controlled using Trim Galore (Babraham Bioinformatics). Later the files were demultiplexed using the custom script and processed further using BiQ Analyzer HT (Max-Plank-Institut Informatik), where all Cs in CpN contexts were analysed and exported to an MS-Excel readable file. The obtained data was later analysed in MS-Excel, where sequences were sorted based on number of undefined bases and sequences containing more than three undefined bases were removed. Moreover,

sequences containing more than 3 unconverted C in primer binding regions (as they carry unmodified Cs) were removed, which would allow us to remove the reads that were matching the lower strand. The average oxidation at each C was calculated and the overall oxidation for each time point is obtained by subtracting the oxidation at respective time point from the control. Based on total oxidation, the flanking sequence preference, sequence specificity, and the processivity/distributivity were extracted.

4. RESULTS

4.1. Establishment of an *in vitro* system to investigate the biochemical properties of TET enzymes

4.1.1. Expression and purification of TET enzymes

The discovery of TET enzymes in 2009 was a major breakthrough in DNA demethylation process, as TET enzymes were shown to initiate the active DNA demethylation by oxidizing 5mC to 5hmC, 5fC and 5caC respectively (Tahiliani et al. 2009; Ito et al. 2011). Since their identification, the significance of different members of the TET family and their reaction products were studied in various biological contexts. In this thesis, I was interested in understanding the biochemical behaviour underpinning the molecular function of TET enzymes.

To achieve this, establishment of an *in vitro* system is necessary. So, I started setting up the *in vitro* system by cloning, expression and purification of active recombinant proteins from *E.coli*. The catalytic domain of mouse TET1 was purified successfully from *E.coli*, after optimisation at both expression and purification procedures. Like TET1-CD, the expression and purification of the catalytic domains of TET2 and TET3 were quite challenging. However, after several rounds of optimisation, the catalytic domains of mouse TET2 and TET3, which are larger than TET1, were purified from *E.coli,* yet these proteins were not active and showed a lot of degradation, suggesting a poor quality of the recombinant proteins. I hypothesized that this could be due to the presence of the large unstructured part within the catalytic domain that could affect the protein folding in *E.coli*. Sequence alignment among the TET family members showed that the unstructured region was not conserved (Iyer et al., 2009) and moreover, the human TET2 catalytic domain lacking the low complexity region was shown to be active (Hu et al., 2013). Taking these studies into consideration, the unstructured region was removed from all three mouse TET enzymes. The borders for deletion of the unstructured region were chosen based on the sequence alignment (Figure 7). These newly designed constructs were cloned into *E.coli* expression vector pET28a (+), expressed in BL21 (DE3) CodonPlus RIL strain and purified as His_{6x} fusion protein using Ni-NTA agarose beads. Consequently, the deletion of the unstructured region greatly improved the expression and the quality of all three TET proteins after purification.

In addition to mammalian TET enzymes, the Tet-like dioxygenase-1 from a free-living amoebaflagellate *Naegleria gruberi* (nTet) was also cloned in pET28a (+) vector from a synthetic gene (nTet in pEX-A2 vector, from MWG Operon), over expressed and purified from *E.coli* like other TET proteins as mentioned above. Unlike the mammalian TETs, nTet is a small, single-domain protein with 39% sequence similarity to mouse and human TET1. It lacks the low complexity region within the DSBH domain and the Cys-rich region, yet the functional residues that are important for the DNA, αKG binding and Fe^{2+} coordination are conserved and the enzyme was shown to be highly active (Hashimoto et al. 2014). SDS-PAGE analysis of all TET proteins used in this study, which were purified from *E.coli* is shown in figure 8. The purified proteins ran at the expected size according to the theoretical mass calculations. The size and the border detail of each construct is tabulated (Table 8).

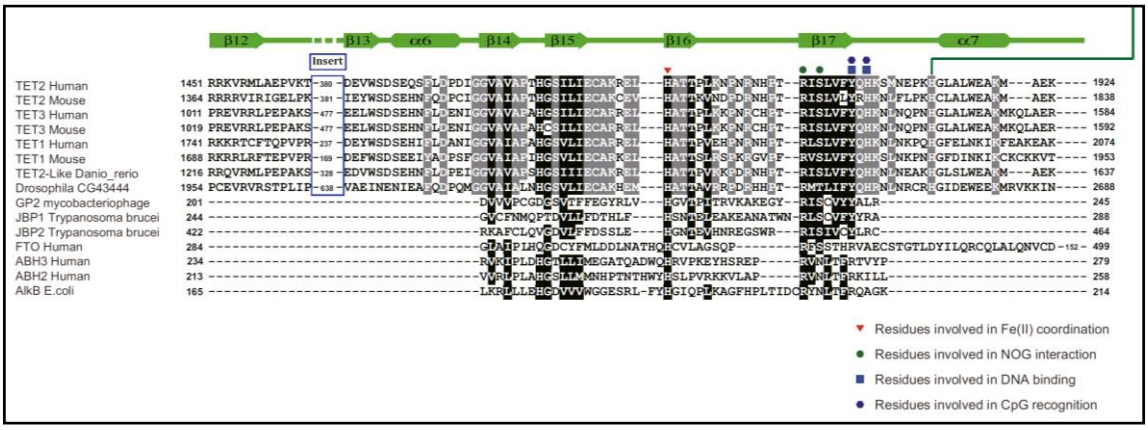


Figure 7: Snapshot of the structure based sequence alignment around the unstructured region, taken from Hu et al., 2013. The sequence alignment of TET enzymes and other αKG dependent dioxygenases containing the DSBH is shown. The conserved regions are highlighted. The unstructured low complexity region is marked with the blue box. The residues involved in the DNA, αKG binding, Fe^{2+} coordination and CpG recognition are indicated by specific symbols.

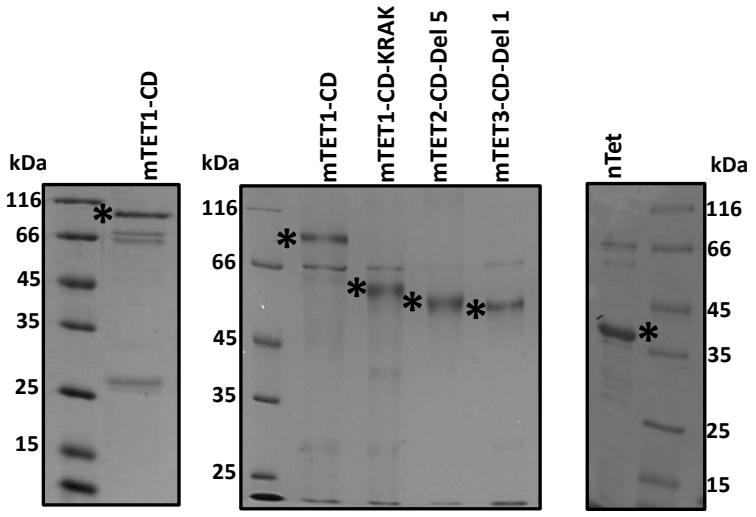


Figure 8: SDS-PAGE containing all TET proteins purified from *E.coli* used in this study. The constructs cloned in pET28a (+) vector were overexpressed and purified from *E.coli* BL21 (DE3) CodonPlus RIL strain. The quality of the purified proteins was verified using 12% SDS-PAGE electrophoresis and stained using Coomassie brilliant blue G250 solution. The size and the construct details are tabulated (Table 8). Asterisk represents the protein of interest. KRAK and Del are the constructs having unstructured region removed.

The identity of the purified proteins was further verified by western blotting using the antibody against His_{6x}-tag (Figure 9 A). The western blot analysis showed a single band for the purified proteins at the expected size, which indicates that there is no significant degradation of the proteins from the C-terminus. This suggests that the other unspecific bands seen in the SDS-PAGE (Figure 8) are either the co-purified *E.coli* proteins or the truncated versions of the protein from the N-terminus. Furthermore, the activity of the purified proteins was tested *in vitro* using the 5mC-containing DNA as substrate. The purified TET enzymes were catalytically active (Figure 9 B), though the level of activity differed, with mTET1-CD, mTET1-CD-KRAK, mTET3-CD-Del1 and nTet showed almost equal activity on 5mC containing DNA, whereas mTET2-CD-Del5 showed ~ 3-3.5 times lower activity compared to other enzymes.

Table 8: Details of TET proteins from mouse and *Naegleria gruberi*, cloned in pET28a (+) vector and purified from *E.coli* as His_{6x} tag fusion

Enzyme	Borders (amino acids)	Expected size (kDa)
mTET1-CD	1367-2038	76.6
mTET1-CD-KRAK	1367-2038-Δ (1679-1869)	61.3
mTET2CD-Del5	1044-1920-Δ (1407-1766)	61.5
mTET3CD-Del1	693-1668- Δ (1302-1509)	59.1
nTet	1-321	37

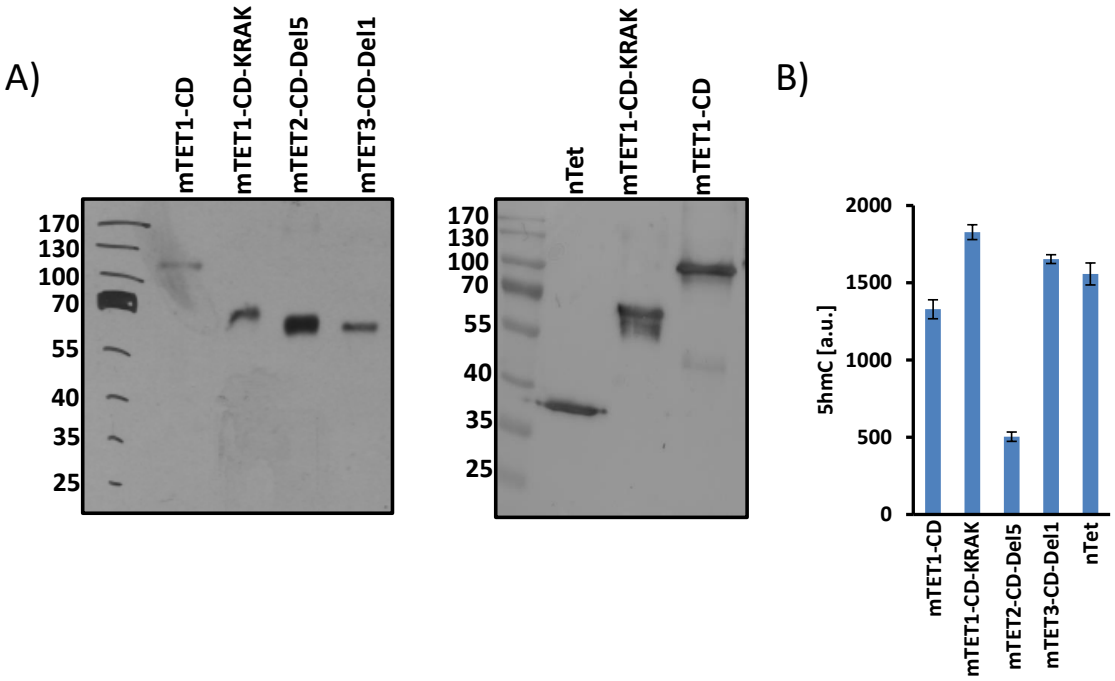


Figure 9: Analysis of different TET proteins purified from *E.coli*. A) Western blot analysis of all TET proteins purified from *E.coli*. The purified proteins were analysed using antibody against His_{6x}-tag. The protein purified is intact as no degradation is observed B) The activity of the purified TET enzymes tested *in vitro* on the methylated substrate.

4.1.1. Establishment of an ELISA-based plate assay to quantify the level of 5mC-oxidation products

Since the discovery of TET enzymes and the oxidized bases (Tahiliani et al. 2009; Kriaucionis and Heintz 2009), numerous studies have been conducted to map their distribution in the genome. The initial studies were focused on 5hmC, as it is relatively abundant in the genome and is detectable in almost all the tissues tested (Globisch et al. 2010; Nestor et al. 2012). However, the methods to detect 5hmC were limited at that time. They were not suitable for high throughput (HT) kinetic assays (2D TLC), provided semi-quantitative data (dot-blot) or required expensive machines like LC-MS.

In addition, the classical bisulfite analysis, which is used to discriminate between C and 5mC, cannot be employed to identify 5hmC, as both 5mC and 5hmC behave in the same way after the bisulfite treatment. Therefore, most of the initial studies were based on pull down using antibodies or chemical modification. The other methods based on the restriction endonucleases are quite successful in detecting 5hmC. However, they are not quantitative, intrinsically sequence dependent and cannot represent the level of oxidation products in all sequence contexts. Moreover, some enzymes require additional modification to discriminate

5mC from the oxidized products (for example- glucosylation of 5hmC using T4-BGT) that increases the amount of the initial material used, as there is sample loss at each step (for details on all the methods, refer Pastor et al. 2013). For global quantifications, mass spectrometry and dot-blot methods were used. Although these methods provide useful information, there are a few disadvantages associated with them. Mass spectrometry is quite expensive and cannot be used easily for the high throughput analysis and requires isotope labelled internal standards for reliable quantification. On the other hand, dot-blot based method is semi-quantitative and show variability based on the concentration of the sample used. For example, antibody used in the dot blot prefers sample with high signal intensity, which leads to signal bias (Ko et al. 2010). There is no practical assay available to study the kinetics of TET catalysed reaction in a simple yet, reliable manner that can be used in a HT mode.

Although the recent advancements in technology helped the researchers to develop few sensitive sequencing based methods to map 5hmC, 5fC and 5caC (Booth et al. 2012; Xia et al. 2015; Wu et al. 2016a), they cannot be used frequently, as they are rather expensive, laborious (takes few days to finish the entire protocol) and require special scientific equipment. So in this project, I set out to establish an easy, efficient and reliable method, which would allow one to study the biochemical properties of TET enzymes. I have developed an ELISA-like microtiter plate (96 or 384) assay that relies on the DNA modification-specific antibodies to quantify the level of the oxidized bases. In this method, the biotin-avidin interaction has been exploited that allowed to tether the product DNA to the surface of the microtiter plate using biotin molecule attached to the DNA. The TET treated DNA was added to the denaturing solution containing 0.05 M NaOH to separate the DNA strands. After binding and denaturation of the DNA strands, a modification-specific primary antibody (anti-5hmC, anti-5fC and anti-5caC) was used followed by the secondary antibody conjugated to HRP and the signal was detected using chemiluminescence. Each step in the workflow was carefully optimized and the schematic representation of the entire procedure is shown in figure 10. This method offers a quick and an easy workflow that can be used to detect different DNA modifications on the same plate. Moreover, it is cost effective, high throughput, does not require any special reagents and is advantageous over the dot blot by offering robust quantitative results.

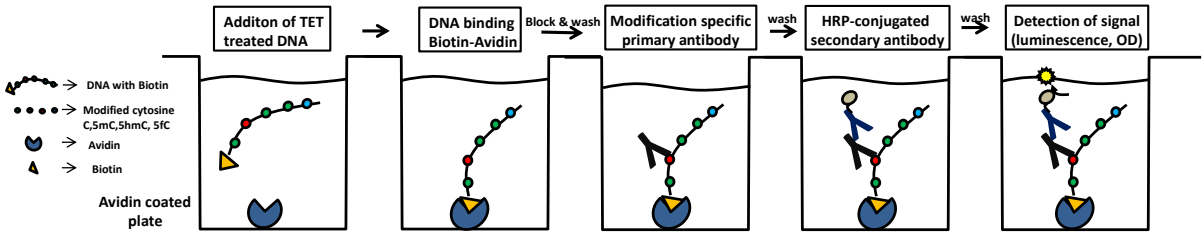


Figure 10: Schematic representation of the ELISA-based plate assay. The biotinylated DNA carrying different modifications is bound to avidin. The oxidized base is detected using the modification specific antibody followed by the secondary antibody and the signal is detected using chemiluminescence.

Since the assay requires antibodies to detect the modification, as a first step the specificity of the commercial antibody (anti-5hmC and anti-5fC) was tested. For this, differently modified DNA (86 bp) was prepared in which every cytosine (except the primer binding region) was substituted by the modified nucleotide (d5caCTP, d5fCTP, d5hmCTP, d5mCTP) during the amplification of the substrate DNA by PCR. As can be seen in figure 11 A, the modified DNA was successfully amplified and they all migrated at the expected size. Next, the DNA carrying different modifications were spotted on the membrane and the specificity of the antibody (anti-5hmC, anti-5fC) was analysed. Dot blot analysis (Figure 11 B) confirmed that both antibodies are specific towards their mark, as there was no signal for the other modified DNA (except the target modification). Moreover, the presence of 5hmC was detected for the 5mC DNA treated with mTET1-CD purified from *E.coli* (Figure 11 B-upper panel right).

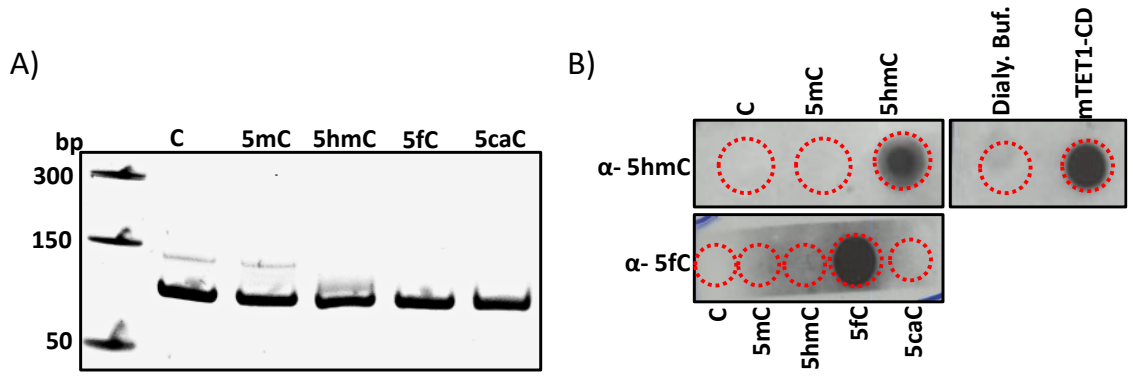


Figure 11: Amplification of the substrate DNA and analysis of the specificity of the antibodies. A) Gel showing the double stranded DNA of length 86 bp carrying different modifications at every cytosine in the sequence except the primer binding regions amplified using PCR. B) Dot blot with antibody against 5hmC (top panel) and 5fC (bottom panel). The DNA with different modifications was used to check the specificity of the antibodies. The results show that the antibodies are specific towards their modifications, as both antibodies (anti-5hmC and anti-5fC) do not recognize other modified DNA except the desired marks. Dot blot on top right panel shows that mTET1-CD catalysed the 5mC containing DNA to 5hmC.

Next, the amount of DNA required per well was optimized. The goal was to determine the amount of DNA that can be added to a well to achieve a reliable and robust signal insensitive to fluctuations in the amount of DNA added. To achieve this, a standard curve was generated by adding different concentrations of the modified DNA and the detected chemiluminescence was plotted as a function of the amount of DNA. It is evident from figure 12, the signal intensity increased linearly with increased amount of the DNA added (for both 5hmC and 5fC containing DNA) and slowly started saturating after 5 ng of the DNA. Based on this experiment, 20 ng of DNA per well was chosen for further assays in order to minimize the signal fluctuations, potentially caused by the variability of DNA added. It should be noted here that the fully modified DNA substrate was used and the signal saturation represents the saturation of the binding surface with the DNA in the well.

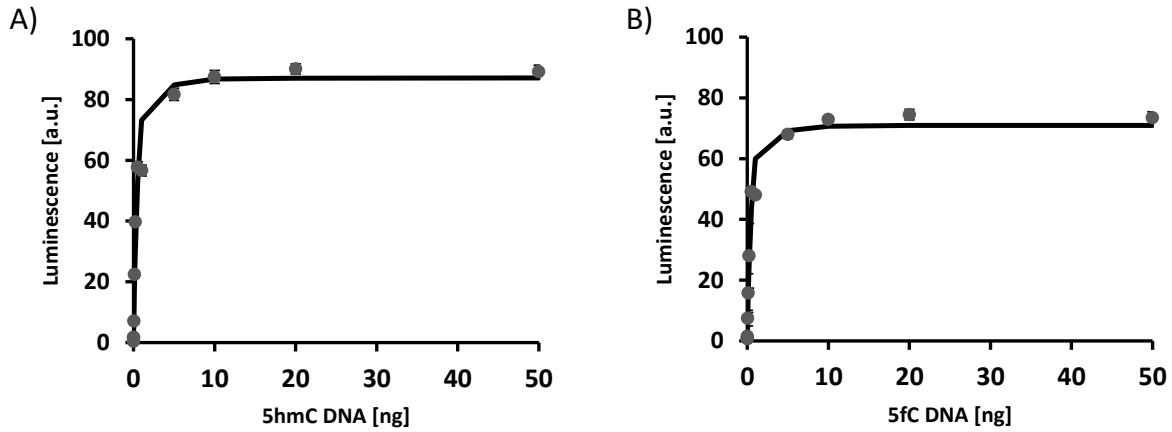


Figure 12: Generation of the standard curve. The detected chemiluminescence which is proportional to the level of oxidized base (A) 5hmC and (B) 5fC is plotted as a function of the amount of DNA. The 5hmC containing DNA is used as a substrate for panel A and the 5fC containing DNA is used as a substrate for panel B. The reaction was performed by adding increasing concentration of the biotinylated DNA. The signal was detected using the ELISA-based plate assay and fitted exponentially using the least square fit. Error bar represents ± SD, n=3 for 5hmC and n=2 for 5fC.

The optimized protocol was employed to follow the time resolved kinetics of the reaction catalysed by TET enzymes. The recombinant mTET1-CD purified from *E.coli* was incubated with 5mC containing DNA and other components necessary for the reaction (for reaction details, refer the section 3.5). 2 µl of the reaction mix containing 20 ng or 0.36 pmole of DNA was taken at different time points and loaded into separate wells of a 96-well microtiter plate. The signal was detected using anti-5hmC, anti-5fC and anti-caC antibodies and was plotted as a function of time (Figure 13). The levels of TET reaction products 5hmC, 5fC and 5caC increased over time, which shows that the plate assay can be used to study the enzymatic

reaction catalysed by TET enzymes. However the total amount of 5hmC produced is more compared to the higher oxidation products 5fC and 5caC.

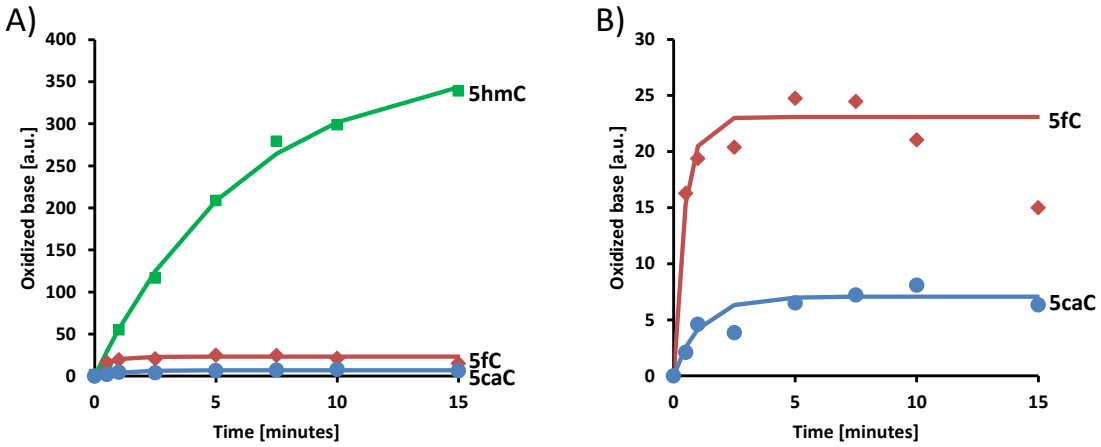


Figure 13: Time resolved kinetics catalysed by TET enzymes A) The level of oxidized bases catalysed by mTET1-CD, fitted exponentially using the least square fit is plotted as a function of time. B) The levels of 5fC and 5caC are shown separately. The reaction was performed using 0.4 µM of the methylated DNA containing multiple 5mC and 2 µM of mTET1-CD for 15 minutes at 37 °C.

4.1.2. Effect of inhibitors on the activity of mTET1-CD

Since the kinetics assay with mTET1-CD worked successfully, as a proof of principle, this protocol was used to study the effect of inhibitors on the activity of mTET1-CD. As there are no TET-specific inhibitors available to date, the inhibitors of αKG-dependent dioxygenase were screened, which include analogues of the co-substrate αKG such as N-oxalylglycine (NOG), Dimethyl oxalylglycine (DMOG). In addition to these, the oncometabolites, 2-hydroxyglutarate (L and D forms) and the citric acid cycle intermediates, succinate and fumarate, which were shown to accumulate in the cancerous cells such as glioma and AML (Xu et al. 2011a; Figueroa et al. 2010) due to the mutation of the respective enzyme, were also tested. These analogues are structurally similar to the co-substrate αKG and are believed to inhibit TET enzymes by displacing αKG from the catalytic centre. The structure of all the inhibitors used in this study together with the co-substrate αKG and enzymes that catalyse the reaction are shown in figure 14.

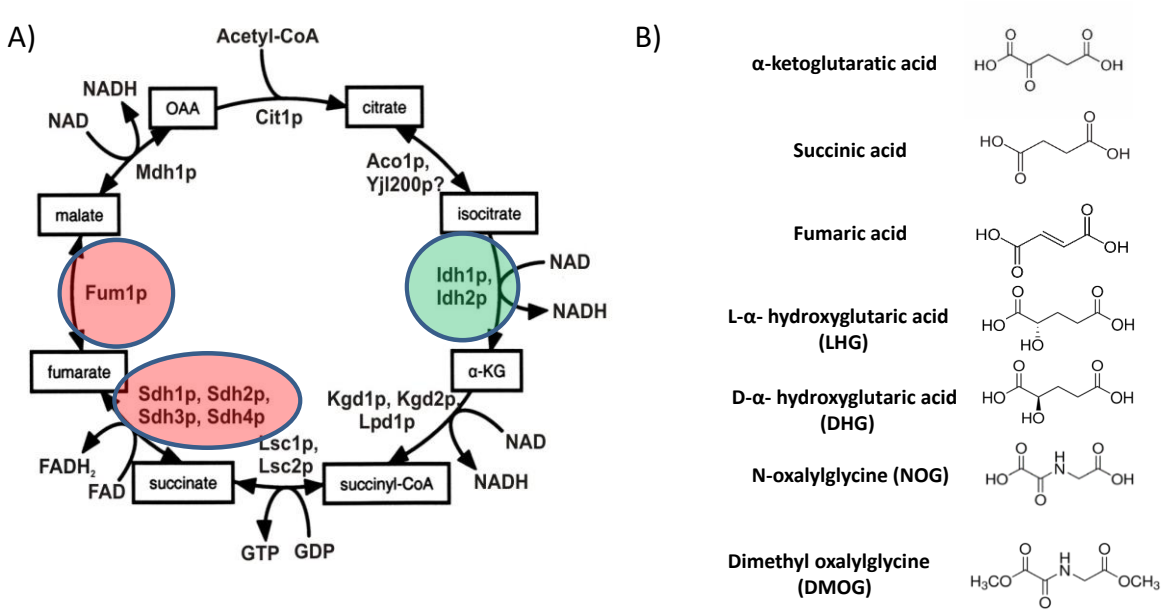


Figure 14: The schematic representation of the citric acid cycle and the chemical structure of the analogues of αKG used in this study. A) Enzymes in the citric acid cycle which catalyse the formation of 2-HG, succinate and fumarate are highlighted. The green circle represents the enzymes isocitrate dehydrogenase (IDH1/2), which catalyse the conversion of isocitrate to α-ketoglutarate (αKG). Specific mutation on these enzymes leads to a gain of function and converts αKG to 2HG (Koivunen et al. 2012). The red circle represents succinate dehydrogenase (SDH) and fumarate hydratase (FH), which converts succinate to fumarate and fumarate to malate respectively and these intermediates often accumulate in cancer cells due to the loss of function mutation (Xiao et al. 2012). B) The structure of α-ketoglutaric acid (αKG) and the derivatives of αKG, the oncometabolite 2HG, succinate and fumarate used in this study.

For the inhibitor assay, the kinetics was performed with different concentrations of the inhibitor with 2 µM mTET1-CD and 0.6 mM of the co-substrate αKG. The apparent IC50 value was calculated by plotting the initial velocity of the reaction as a function of the inhibitor concentration.

From the figure 15 A, it can be seen that the addition of NOG inhibited the activity of mTET1-CD leading to a decrease in the reaction velocity in a concentration dependent manner. The calculated apparent IC50 (at 0.6 mM aKG) for NOG was around 81 ± 30 µM, which indicates that NOG competes with αKG and inhibits the activity of TET enzymes efficiently. Interestingly, DMOG, which is an esterified version of NOG, did not inhibit the activity of mTET1-CD even at a concentration as high as 64 mM (Figure 15 B). This suggests that DMOG does not compete for αKG binding site *in vitro*.

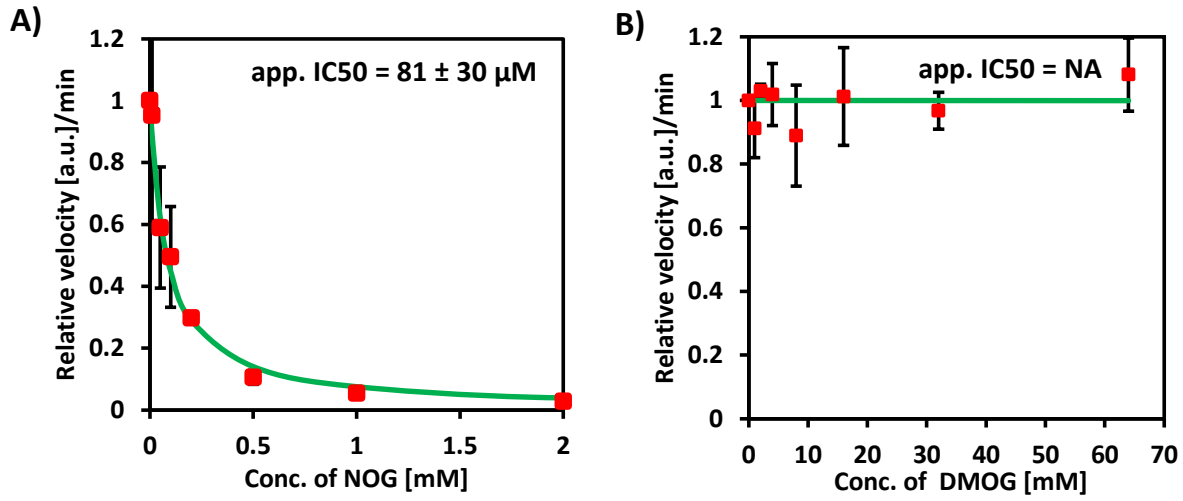


Figure 15: The derivatives of αKG inhibits the activity of mTET1-CD. The activity of mTET1-CD at different concentrations of (A) N-oxalylglycine (NOG) and (B) Dimethyl oxalylglycine (DMOG) is fitted using the least square fit. Please note the different concentration scale for NOG. The reactions were performed in the presence of 0.6 mM aKG. Error bar represents ± SD, n=3.

Likewise, the kinetics performed with the oncometabolites showed that both L-HG and D-HG inhibited the activity of mTET1-CD (Figure 16 A-B). The apparent IC50 value determined for L-HG and D-HG were 6.4 ± 0.9 mM and 3.6 ± 0.9 mM respectively, which are quite high (in mM) compared to the inhibitor NOG. These results suggest that these metabolites are weak inhibitors of TET enzymes and should be present at high concentration to effectively block the intracellular activity of TET enzymes. Indeed, 2-HG is reported to accumulate in mM range in gliomas carrying the IDH mutants (Ye et al. 2013). Moreover, the app. IC50s calculated for L-HG and D-HG are in the mM range, as it was reported in an independent study (Koivunen et al. 2012), showing that the method is reliable, since it matches the results of an independent study.

Furthermore, the inhibitory effect of succinate and fumarate was tested on mTET1-CD. Succinate is the reaction product of αKG-dependent dioxygenases family including TET enzymes and an excess of succinate can lead to product inhibition of the enzymatic activity by occupying the αKG binding pocket.

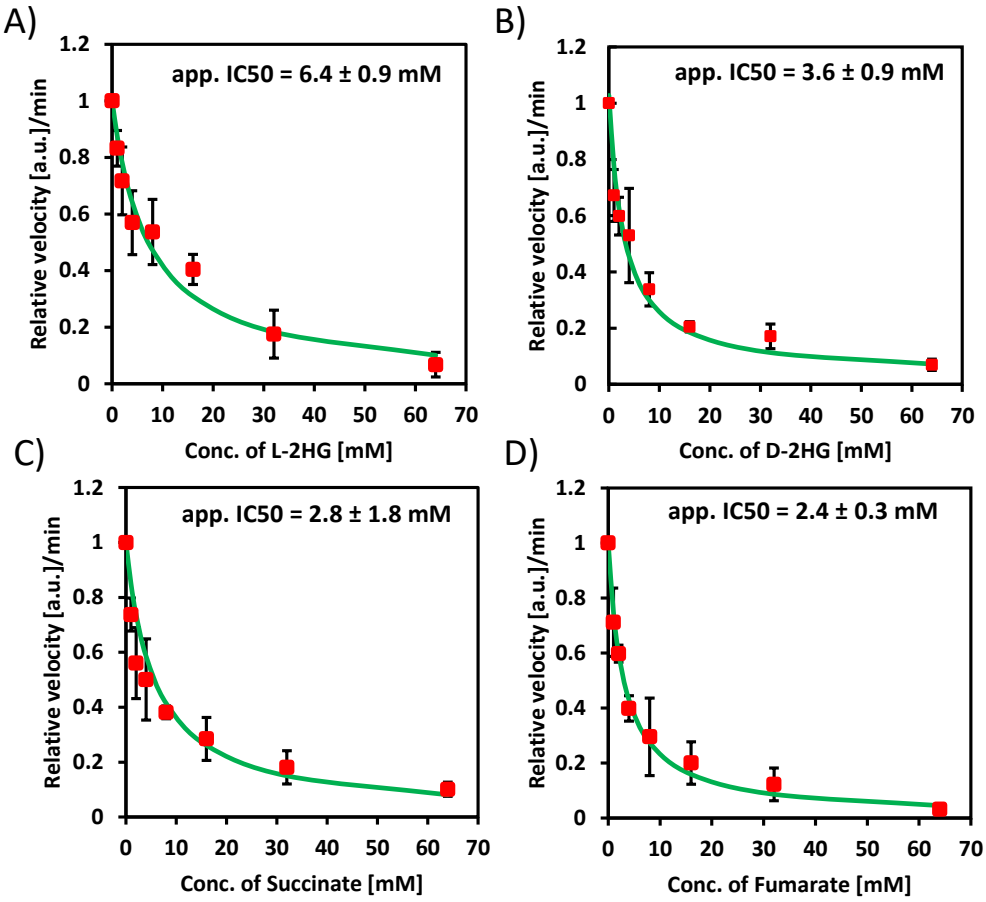


Figure 16: The derivatives of αKG inhibits the activity of mTET1-CD. The figure shows that L-hydroxyglutarate (L-HG), D-hydroxyglutarate (D-HG), succinate and fumarate inhibit the activity of mTET1-CD in a concentration dependent manner. The reactions were performed in the presence of 0.6 mM αKG. Error bar represents ± SD, n=3.

Moreover, the studies showed that the accumulation of succinate and fumarate caused by the acquired mutations on the respective enzymes (SDH and FH) can inhibit several αKG dependent dioxygenases including TET enzymes and the inhibition was demonstrated *in vitro* as well (Laukka et al. 2016). In agreement with this, both succinate and fumarate inhibited the activity of mTET1-CD in a concentration dependent manner. The calculated apparent IC50 value for succinate and fumarate were 2.8 ± 1.8 mM and 2.4 ± 0.3 mM respectively (Figure 16 C-D).

Table 9: The apparent IC50 values of the inhibitors used in this study

Inhibitors	IC50 of mTET1-CD
L-hydroxyglutarate (LHG)	6.4 ± 0.9 mM
D-hydroxyglutarate (DHG)	3.6 ± 0.9 mM
N-oxylglycine (NOG)	80 ± 30 μM
Dimethyloxalylglycine (DMOG)	No inhibition detected (ND)
Succinate	2.8± 1.8 mM
Fumarate	2.4 ± 0.3 mM

Overall, the inhibitors kinetics showed that except DMOG, other inhibitors used here could block the activity of TET enzymes *in vitro,* but their inhibitory potential differed, with NOG being the most potent inhibitor compared to the other inhibitors tested here. Interestingly, the inhibition caused by succinate and fumarate was more efficient than LHG and DHG. Altogether, these results showed that the ELISA-based plate assay established in this study can be successfully employed to study the TET enzyme kinetics or potentially any other DNA modification enzymes depending on the available modification specific antibodies. As a proof of concept, the tested inhibitors recapitulated the published results (Koivunen et al. 2012; Laukka et al. 2016). Thus, this assay is perfectly suited to screen any specific inhibitors of TET enzymes in a high-throughput manner or to follow the reaction kinetics of TET mutants.

4.2. Effect of divalent metal ions on the activity of TET enzymes

TET enzymes require Fe^{2+} in their catalytic centre for the catalysis. Due to their dependency on Fe^{2+}, other metal ions could compete for the iron binding site and inhibit their activity. Using both cell culture and *in vivo* studies, it was demonstrated that the exposure of divalent nickel ions increase the chromatin compaction and DNA methylation. It was suggested that the cellular exposure of nickel ions target the heterochromatin and increase the DNA methylation level, which affect the expression of tumour suppressor gene (Chen et al. 2006; Ke et al. 2006, 2008a; Sun et al. 2013). However the mechanism of nickel-induced increase in DNA methylation is not completely understood. Studies showed that the treatment of cells with nickel ions (both soluble-$NiCl_2$ and insoluble NiS) induce other epigenetic changes such as decrease in acetylation, increase in histone H3K9me2 methylation, ubiquitination and phosphorylation of histone H3 at serine 10 (H3S10) (Chen et al. 2006; Ke et al. 2008a, 2008b, 2006). Moreover, treatment of mammalian cells with nickel was reported to decrease the cellular iron level, which eventually affects the Fe-S cluster dependent enzymes therefore, resulting in an altered cellular metabolic processes/pathways such as oxidative phosphorylation (Chen et al. 2005; Chen and Costa 2006). Interestingly, another study reported that the increase in nickel concentration inside the cells inhibits histone H3K9 demethylase JMJD1A by displacing the Fe^{2+} from its catalytic centre (Chen et al. 2010a). By taking all these observations into consideration, together with the dependency of TET enzymes on Fe^{2+} and their function as tumour suppressor, it was hypothesized that the divalent metal ions could inhibit the activity of TET enzymes by replacing the ferrous iron (Fe^{2+}) from its binding pocket and hence contribute to the increased methylation level. To verify this, six divalent metal ions: cadmium, cobalt, zinc, nickel, manganese and calcium were selected and their influence on the activity of TET enzymes was investigated.

4.2.1. Nickel ions inhibit the activity of mTET1-CD *in vitro*

To test the inhibitory effect of divalent metal ions on mTET1-CD activity, increasing concentrations of these selected metal ions were used (0, 0.1, 1, 10, 100, 1000 µM). The reactions were conducted in the presence of 10 µM Fe^{2+} supplemented with respective metal ions, 2 µM mTET1-CD and other necessary components for the reaction, as described in the section 3.8. The time resolved kinetics were performed using the ELISA-based plate assay and the level of 5mC oxidation catalysed by mTET1-CD was followed using specific antibody. The

results showed (Figure 17) that for all metal ions except calcium, with their increased concentration in the reaction, decrease in the activity of mTET1-CD was observed. However, the strength of inhibition differed depending on the metal ions used. The apparent inhibitory concentration (IC50) was calculated for each metal ion tested and is shown in table 10. Comparison of the app. IC50 values calculated for the tested metal ions showed that cadmium was the most potent inhibitor with an app. IC50 value of 0.87 ± 0.14 µM followed by zinc with an app. IC50 value of 2.26 ± 0.3 µM, suggesting that both these metal ions have high affinity for the TET catalytic centre. The app.IC50 values for cobalt (15.6 ± 0.27 µM), nickel (19 ± 1 µM) and manganese (57 ± 15 µM) indicate that cobalt ions show stronger inhibition than nickel and manganese. The app. IC50 values obtained in this study for nickel ion is in line with the reported IC50 values for JMJD1A (25 µM) (Chen et al. 2010a). Interestingly, calcium ions did not inhibit TET enzymes even at a concentration as high as 10 mM suggesting that it does not bind in the enzymes catalytic centre.

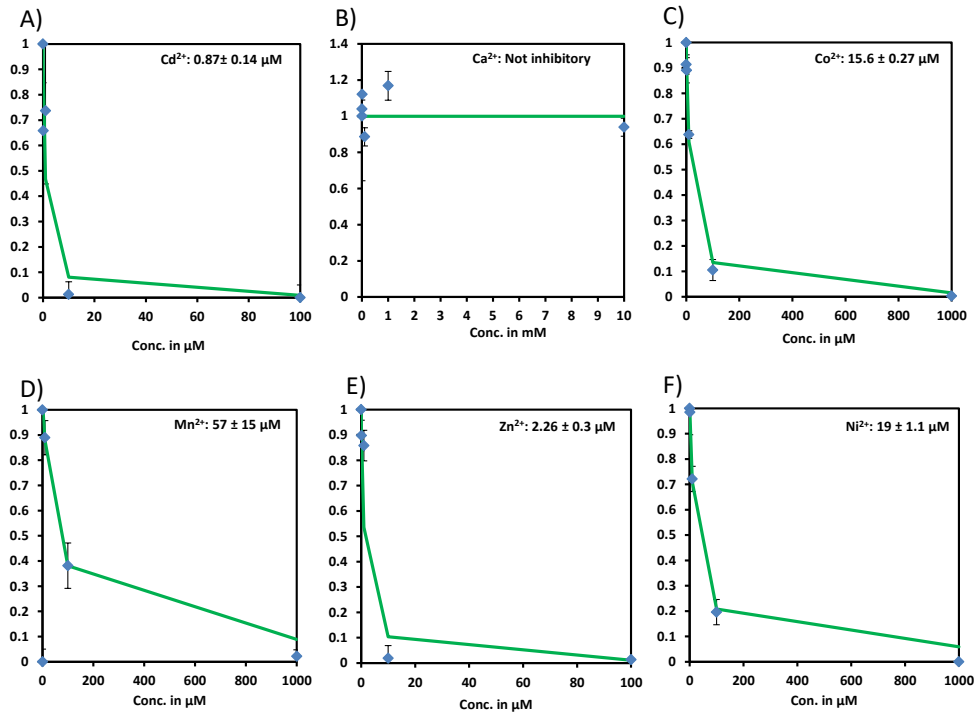


Figure 17: The inhibitory effect of different divalent metal ions on the activity of mTET1-CD. The activity of mTET1-CD (with 10 µM Fe^{2+} in the presence of various metal ions (A) cadmium (B) calcium (C) cobalt (D) manganese (E) zinc (F) nickel was analysed using the ELISA-based plate assay. The rate of oxidation of 5mC to 5hmC at different concentrations of the metal ions was normalized to the control and the IC50 value was calculated. Error bar represents standard deviation from at least 3 individual experiments. Note that the scale for calcium concentration is in mM, whereas for other metal ions it is in µM.

Table 10: The apparent IC50 values of divalent metal ions for mTET1-CD

Divalent metal ion	App. IC50 [μM]
Cadmium	0.87 ± 0.14
Zinc	2.26 ± 0.3
Cobalt	15.6 ± 0.27
Nickel	19 ± 1
Manganese	57 ± 15
Calcium	Not inhibitory

4.2.2. Nickel ions decrease the level of 5hmC in the mammalian cells

Since the metal ions inhibited mTET1-CD *in vitro*, their effect on TET mediated 5hmC production was tested in mammalian cell culture system. For this purpose, HEK293 and mES cells were chosen. To check the toxicity of these metal ions, different concentrations (0, 200, 400, 600, 800 μM) of each metal ion was added to HEK293 cells and their growth monitored. Cadmium, manganese and zinc at these concentrations were toxic to the cells and died within 24 hours of incubation. On the other hand, cobalt and nickel (up to 800 μM) did not show acute toxicity to the cells during the first 24 hours of cultivation. Based on the toxicity test, further experiments were continued with nickel ions. However, the cells treated with nickel at concentrations above 400 μM grew slower, were detaching from the culture plates and started dying after 24 hours (Figure 18).

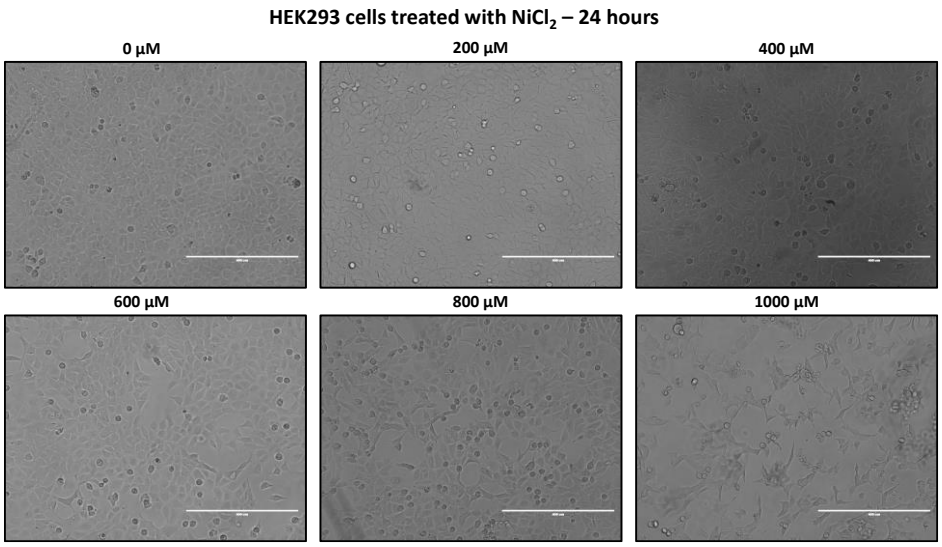


Figure 18: Representative images of HEK293 cells after 24 hours treatment with different concentration of $NiCl_2$. Treatment of HEK293 cells with $NiCl_2$ until 400 μM did not affect their growth, whereas at 600 μM and 800 μM the cell growth was retarded and at 1000 μM, the cells started dying.

Even though the cells treated with high concentration of nickel were dying, I still wanted to check the effect of nickel on 5hmC (yet one should keep in mind that cell death due to other effects cannot be excluded). The cells treated up to 800 µM for 24 hours were harvested and the genomic DNA was extracted from these cells. The level of 5hmC was detected using dot blot with the antibody against 5hmC. As a loading control, the same blot was later stained with methylene blue (Figure 19 A-B). Comparing the loading control and the level of 5hmC, it is evident that for similar amount of DNA loaded on the blot, level of 5hmC decreases in a concentration dependent manner. Quantification of the 5hmC signal (Figure 19 C) showed that there was almost four fold decrease in the level of 5hmC in the cells treated with 800 µM nickel ions compared to the control (Table 11). This indicates that nickel ions indeed affect the activity of TET enzymes and the effect is more pronounced for high concentrations of nickel (600 and 800 µM), yet these observations should be considered with caution, as other effects related to nickel toxicity could be overlapping.

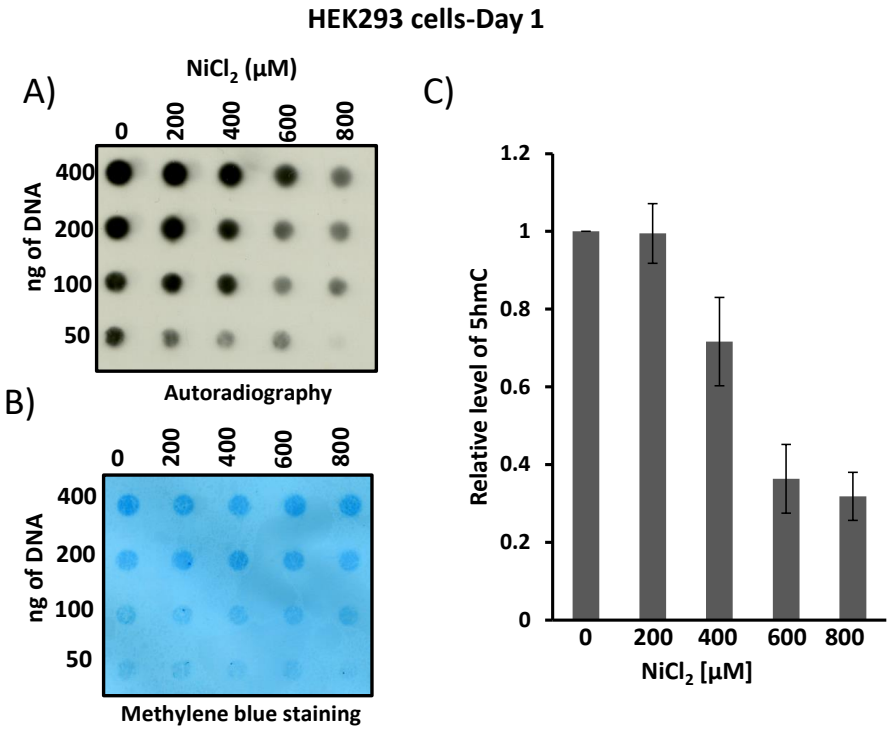


Figure 19: Representative dot blot and quantification of the signal showing the inhibitory effect of nickel on TET enzymes in HEK293 cells. A) The figure shows the dot blot where the level of 5hmC is detected using anti-5hmC antibody. The genomic DNA was extracted from HEK293 cells treated with nickel chloride ($NiCl_2$) of concentration up to 800 µM for 24 hours. B) The figure shows the blot stained with methylene blue, used as a loading control. C) Quantification of 5hmC signal from the DNA extracted from HEK293 cells treated with different concentrations of nickel ions for 24 hours. Error bar represents ± SD, n=2.

As HEK293 cells treated with nickel concentrations > 400 µM did not survive for more than 24 hours, further experiment were conducted with low concentrations of $NiCl_2$ (0, 10, 30, 100 and 300 µM), at which the cells grew normally. Moreover, the cells were treated with nickel ions for 5 days instead of 1 day, to simulate an extended exposure to nickel. Nickel exposure (up to 300 µM) for 5 days was not deleterious to the cells and the cells grew similar to the control cells cultivated without nickel ions added (Figure 20). After 5 days of $NiCl_2$ treatment, the level of 5hmC was measured using dot blot analysis. The result (Figure 21 A-B) showed that the prolonged exposure of nickel in HEK293 cells (compared with day 1 and day 5 for HEK293 cells) decreased the level of 5hmC further in a concentration dependent manner. This is also evident from the quantification (Figure 21 C), where there are ~2 and 5-fold differences in the level of 5hmC in the DNA extracted from the cells treated with 100 and 300 µM $NiCl_2$ respectively for 5 days compared to the control cells (Table 11). This indicates that similar to JMJD1A (Chen et al. 2010a), the exposure of cells to nickel ion could affect the activity of TET enzymes in HEK293 cells. Moreover, the time dependent decrease in the level of 5hmC and normal cellular growth up to five days of nickel treatment (up to 300 µM) suggest that the loss of 5hmC is also cell division dependent. This suggests that the addition of nickel ions to the cells affects the activity of TET enzymes and decreases the level of 5hmC, which then diluted passively with each cell division.

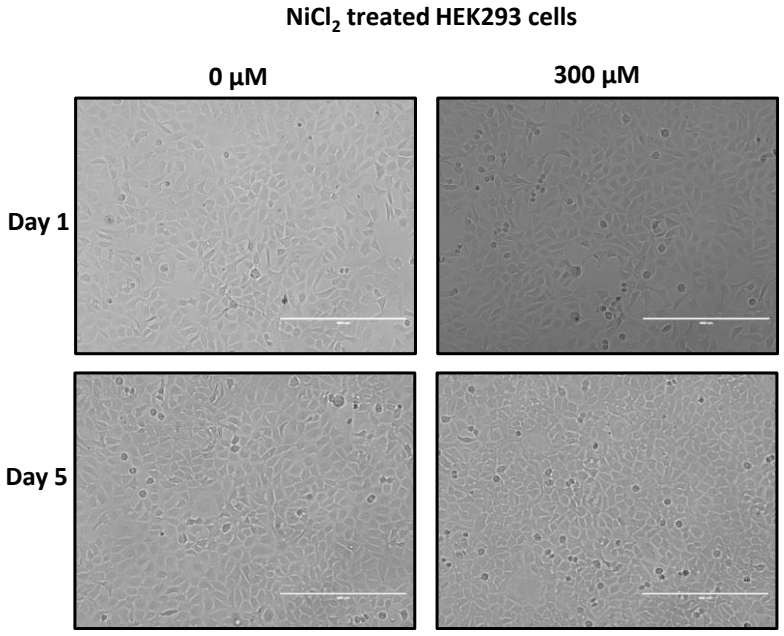


Figure 20: Representative image of HEK293 cells treated with 300 µM $NiCl_2$. HEK293 cells treated with or without nickel ions (up to 300 µM) for 5 days. After 5 days of treatment, the cells treated with nickel ions does not show any growth defect and look similar to the control cells.

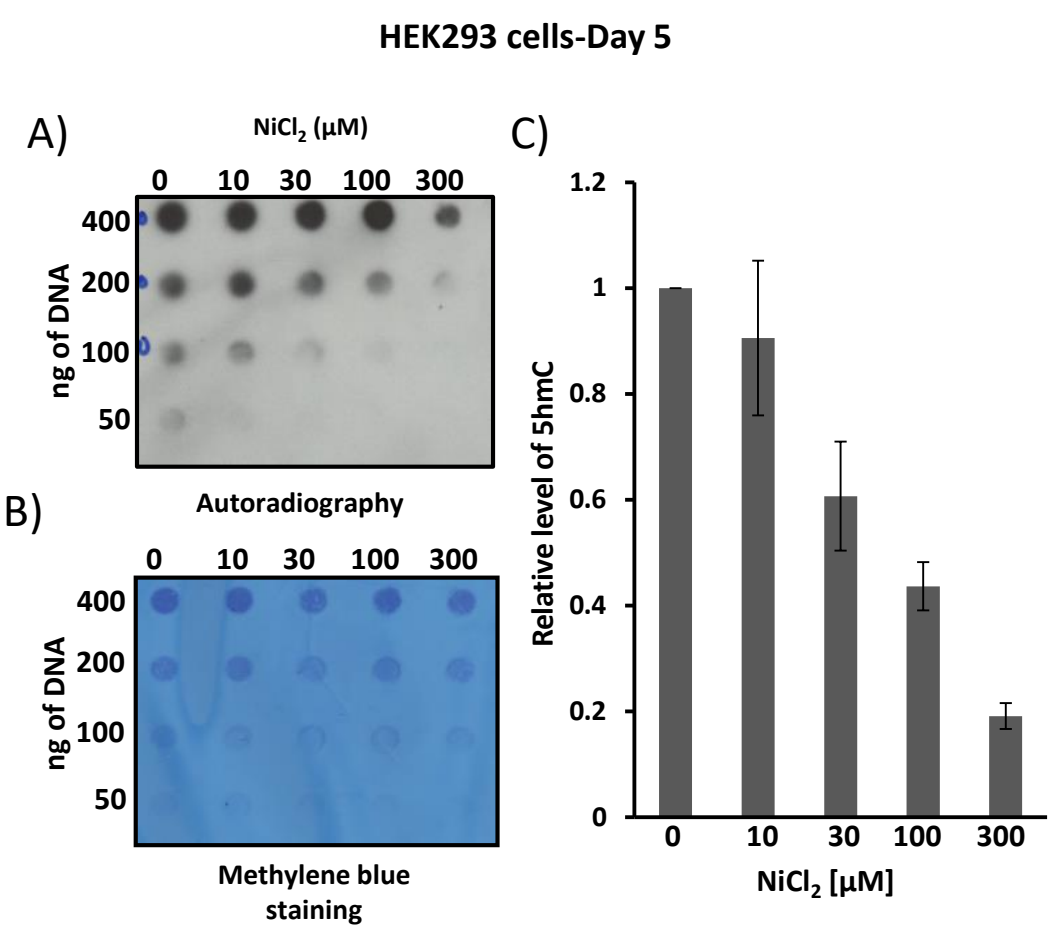


Figure 21: Representative dot blot and quantification of the inhibitory effect of nickel on TET enzymes in HEK293 cells. A) The figure shows the dot blot where the level of 5hmC is detected using the antibody against 5hmC. The genomic DNA was extracted from HEK293 cells treated with nickel chloride of concentration up to 300 µM for five days. B) The figure shows the same blot used for figure A but later stained with methylene blue, used as a loading control. C) Quantification of 5hmC signal from the DNA treated with different concentrations of nickel ions from HEK293 cells after 24 hours. Error bar represents ± SD, n=2.

Table 11: Quantification of the level of 5hmC on the genomic DNA extracted from HEK293 cells treated with $NiCl_2$ for 1 and 5 days

$NiCl_2$ (µM)	Relative level of 5hmC-Day 1	$NiCl_2$ (µM)	Relative level of 5hmC-Day 5
	1	0	1
200	0.99 ± 0.07	10	0.91 ± 0.15
400	0.71 ± 0.11	30	0.61 ± 0.1
600	0.36 ± 0.09	100	0.43± 0.05
800	0.31 ± 0.06	300	0.19 ± 0.02

Since HEK293 cells express TET enzymes at a relatively low level, the effect of nickel ion was further tested on mouse embryonic stem cells (mES cells), which express high levels of TET1 and TET2, and contain high level of 5hmC in their genome (Ito et al. 2010; Koh et al. 2011).

Similar to HEK293 cells, the toxicity of nickel ions on mES cells was tested. mES cells were more sensitive to nickel ions than HEK293 cells and died at 250 µM of $NiCl_2$ after 24 hours. Therefore, mES cells were treated with concentrations < 250 µM such as 10, 50 and 100 µM for 5 days that did not show any significant effect on the ESC growth. Similar to the previous experiment, the level of 5hmC was detected using dot blot on the genomic DNA extracted from mES cells treated with $NiCl_2$ for 5 days. As can be seen in figure 22, the level of 5hmC decreased in a concentration dependent manner. Altogether, these experiments in cultured cells reconfirmed the *in vitro* results that the exposure of mammalian cells to nickel ions decreased the level of 5hmC and it might be due to the inhibition of TET enzymes, as TET enzymes are mainly responsible for the formation of 5hmC.

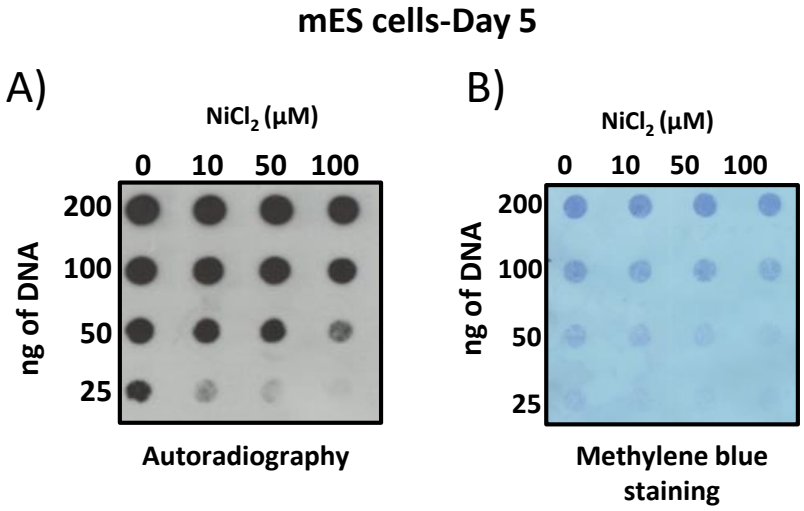


Figure 22: Representative dot blot showing the inhibitory effect of nickel on TET enzymes in mouse embryonic stem cells (mES cells). A) The figure shows the dot blot where the level of 5hmC is detected using anti-5hmC antibody. The genomic DNA was extracted from mouse embryonic stem cells treated with nickel chloride of concentration up to 100 µM for five days. B) The figure shows the same blot used for figure A but later stained with methylene blue, used as a loading control.

4.2.3. Nickel ions do not alter the mRNA expression of TET 1-3 significantly

The *in vitro* and cell culture experiments showed an overall decrease in the level of 5hmC on nickel treatment, which might be due to the impairment of TET enzymes. However, this decrease in 5hmC could also be the result of reduced TET mRNA expression or their protein level being affected by nickel treatment. To verify this, qPCR was performed with the cDNA prepared from HEK293 cells treated with $NiCl_2$ (up to 300 µM) for 5 days. Using specific primers, the expression level of all three TET enzymes (TET1, TET2 and TET3) was analysed and compared to the expression of the control gene REEP5. Figure 23 shows the mRNA

expression of all three TET enzymes from the cells treated with different concentrations of nickel ions (0, 10, 30, 100, 300 µM). Nickel concentration up to 100 µM did not significantly affect the mRNA expression of TET enzymes, whereas cells treated with 300 µM of $NiCl_2$ showed decreased level of TETs mRNA and the effect was more pronounced (two fold decrease) for TET2 compared to TET1 and TET3. This suggests that at concentrations > 300 µM, nickel ions affect the mRNA expression to some extent.

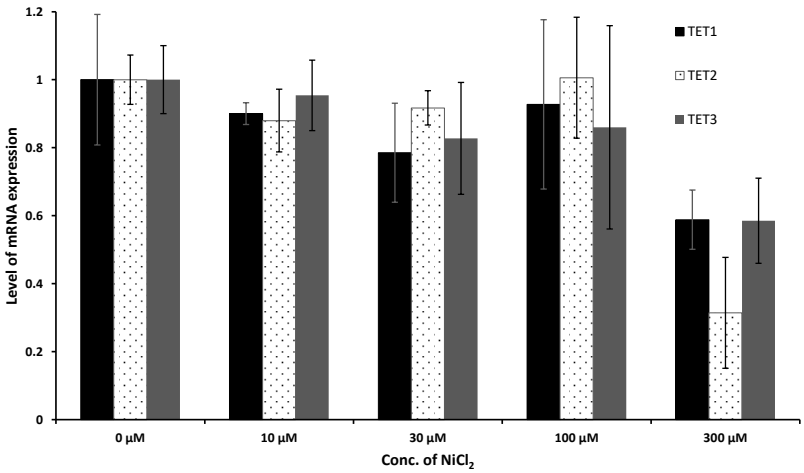


Figure 23: Quantitative PCR to measure the level of mRNA expression of all three TET enzymes. The figure shows the level of mRNA expression of TET1, TET2 and TET3 for the cells treated with different concentrations of nickel for five days. The expression level of TETs is normalized to the expression level of the gene REEP5. Error bar represents standard deviation of three experiments (technical repeat).

Taken together, these results showed that the nickel ions inhibited the catalytic activity of TET enzymes *in vitro* and also in cell culture. Interestingly, nickel ions did not affect the mRNA level of all three TET enzymes until 100 µM, whereas at 300 µM, the expression of TET enzymes was decreased around 1.7- 2 fold.

4.3. Investigation of the role of ascorbic acid and retinol on the activity of Tet enzymes

Ascorbic acid (AscA), commonly known as vitamin C is a reducing agent, which converts Fe^{3+} to Fe^{2+} and is a specific cofactor of the enzyme collagen prolyl 4-hydroxylase (C-P4H). The initial findings showed that the deficiency of AscA inactivated C-4PH, which with time leads to scurvy (Svirbely and Szent-Györgyi 1933). Using *in vitro* studies, it was demonstrated that AscA was absolutely necessary for the enzyme catalysis and was suspected to rescue the enzyme from uncoupled oxidation (oxidation without prime substrate) (Myllyla et al. 1984). Following this discovery, many other members of αKG and Fe^{2+}-dependent dioxygenase family of enzymes were shown to depend on AscA for their full catalytic activity (Ozer and Bruick 2007). It was suggested that AscA was required to maintain the iron in Fe^{2+} state in the enzyme's catalytic centre. However, the exact molecular mechanism was not known.

In the recent years, several groups showed that the supplementation of AscA to the cell culture media enhances the TET activity and increases the genome-wide level of 5hmC. These reports suggested that AscA functioned as a specific cofactor of TET enzymes (Blaschke et al. 2013; Chung et al. 2010; Minor et al. 2013; Yin et al. 2013). Moreover, using biochemical studies, it was demonstrated that similar to C-P4H, AscA enhanced the activity of TET enzymes as a bound co-factor (Yin et al. 2013).

4.3.1. Ascorbic acid does not stimulate the activity of TET enzymes *in vitro*

However, when I performed the reactions with TET enzymes in the presence of AscA, I did not observe any stimulation in their activity under the optimised reaction condition. Addition of 1 mM AscA to the reaction mixture did not increase the activity of mTET1-CD compared to the control reaction, where no AscA was supplemented (Figure 24). To make sure that the observed effect was not an artefact, the experiment was repeated several times and same effect was observed i.e., there was no significant increase in the activity of TET enzymes after adding 1 mM AscA. This prompted me to further investigate the role of AscA acid on the TET enzymatic activity.

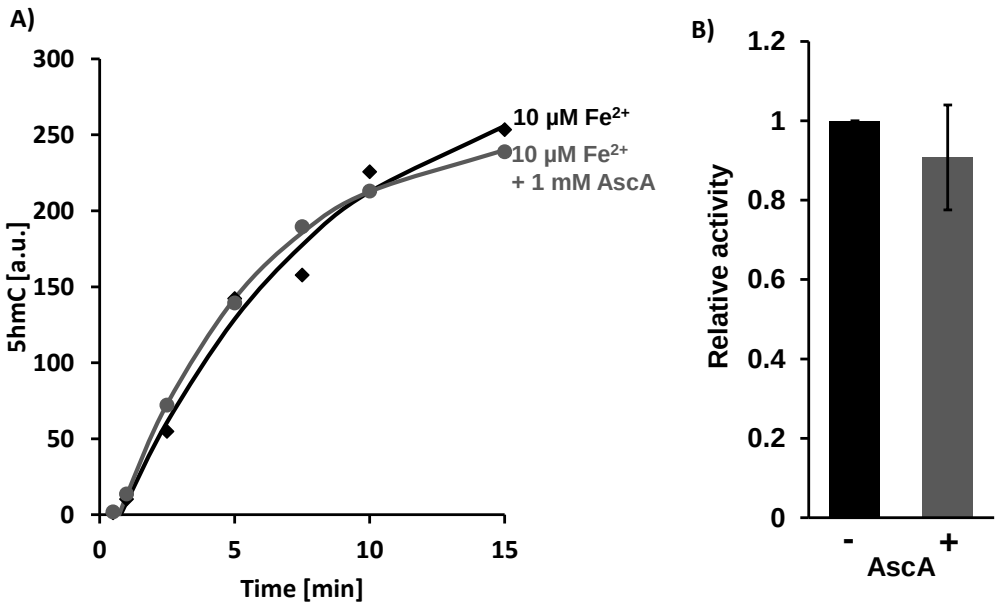


Figure 24: No stimulation in the activity of mTET1-CD in the presence of AscA. A) The primary data showing the oxidation of 5mC to 5hmC catalysed by mTET1-CD over time in the presence of 1 mM AscA (grey curve) and absence of AscA (black curve) with 100 µM Fe^{2+}. a.u. represents arbitrary units B) The relative activity of mTET1-CD with 100 µM Fe^{2+} supplemented with or without AscA (1 mM). Error bar represents standard deviation, n=4.

To verify whether the simulation of TET enzymes by AscA as observed by others (Blaschke et al. 2013; Yin et al. 2013) could be due to the reduction of Fe^{3+} to Fe^{2+}, as AscA is a well-known reducing agent of iron, the kinetics was repeated with mTET1-CD in the presence or absence of 1 mM AscA under Fe^{2+} or Fe^{3+} (100 µM) condition. The result obtained under Fe^{2+} condition was consistent with the previous observation (Figure 25) that the oxidation rate of 5mC to 5hmC was unaffected by the presence or absence of AscA. Whereas, under Fe^{3+} condition (100 µM) in the absence of AscA, no catalytic activity was observed with mTET1-CD (Figure 25 A- red bar). This was expected, as Fe^{3+} does not support the catalysis of TET enzymes and reconfirms the dependence of TET on Fe^{2+} for the catalysis. On the other hand, increase in the oxidation of 5mC to 5hmC was observed over time for the reaction containing Fe^{3+} with 1 mM AscA and importantly, the rate of the reaction was almost equal to the rate of the reaction observed under Fe^{2+} condition (Figure 25). This suggests that the stimulation of TET enzymatic activity by AscA might be due to the reduction of Fe^{3+} to Fe^{2+}.

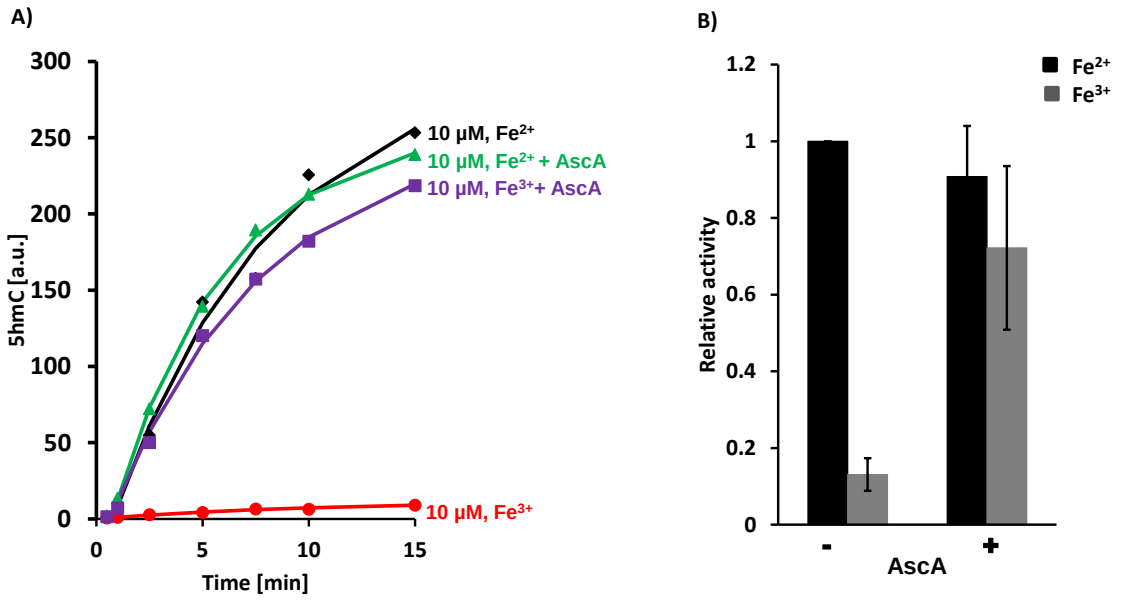


Figure 25: The kinetics performed with mTET1-CD in the presence or absence of AscA (1mM) under Fe^{2+} or Fe^{3+} condition (100 µM) showing AscA reduces Fe^{3+} to Fe^{2+} efficiently. A) The kinetics shows the oxidation of 5mC to 5hmC catalysed by mTET1-CD under different conditions. B) The relative activity of mTET1-CD under Fe^{2+} or Fe^{3+} condition in the presence or absence of AscA. No stimulation of mTET1-CD was observed under Fe^{2+} conditions with AscA, whereas the addition of AscA restored the activity of mTET1-CD under Fe^{3+}. Error bar represents standard deviation, n=4. a.u. represents arbitrary unit.

4.3.2. Ascorbic acid is not a specific cofactor of TET enzymes

Encouraged by the results that AscA could rescue the TET activity by reducing the oxidised iron, I wanted to prove that the reducing capacity of AscA was responsible for the stimulatory effect rather than directing binding of AscA to TET enzymes. Moreover, the proposal of AscA being a cofactor was based on the observation that except AscA, other reducing agents such as spermidine, DTT, vitamin B1, vitamin E, glutathione, NADPH, and L-cysteine did not enhance the activity of TET enzymes in cell culture experiments (Yin et al. 2013; Blaschke et al. 2013; Minor et al. 2013). To verify whether AscA could be substituted by other reducing agents under the optimised reaction condition used in this study, B-ME, DTT, hydroquinone, L-glutamine, N-acetyl L-cysteine, TCEP and KI were used instead of AscA (for details refer the section 3.12). The initial kinetics performed with all reducing agents under Fe^{2+} or Fe^{3+} conditions showed (Figure 26 A) that except AscA (both L- and D-form), the other reducing agents had little effect on the activity of mTET1-CD, as there was no full restoration of the mTET1-CD activity. As other reducing agents show different propensity to reduce ferric iron (Petrat et al. 2003), most likely owing to their chemical nature and the substrate specificity, I

reasoned that additional time would be required for them to reduce Fe^{3+} ions in the reaction mixture. To test this further, Fe^{3+} was incubated with different reducing agents for 30 minutes without mTET1-CD and the kinetics were initiated by adding the enzyme. The resulting kinetics (Figure 26 B) confirmed that this extended pre-incubation of reducing agents with Fe^{3+} restored the activity of mTET1-CD for almost all except TCEP and KI (Figure 26 A). This suggests that the other reducing agents can substitute AscA in the reaction and thus, AscA is not a specific cofactor of TET enzymes.

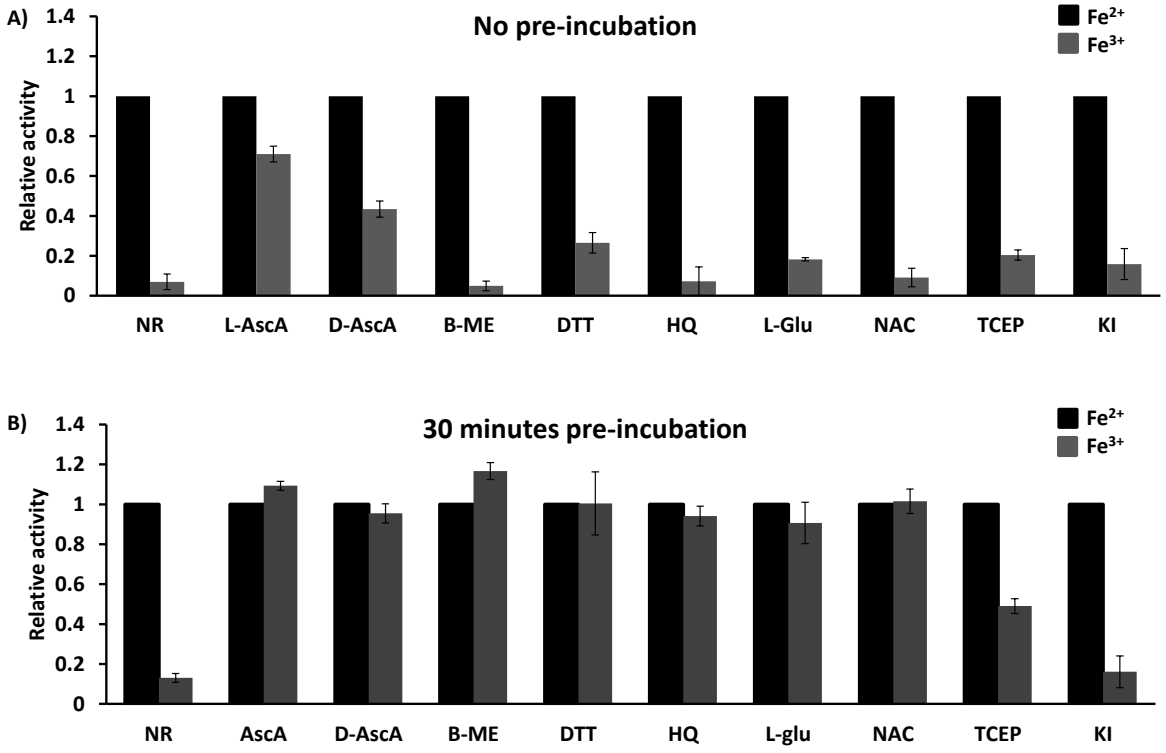


Figure 26: The effect of different reducing agent on the activity of mTET1-CD. A) The relative activity of mTET1-CD without pre-incubation under Fe^{2+} (black bars) or Fe^{3+} (grey bars) condition (100µM), in the presence of different reducing agents. B) The relative activity of mTET1-CD under Fe^{2+} (black bars) or Fe^{3+} (grey bars) condition (100µM) after incubating for 30 minutes with different reducing agents. NR represents no reducing agent added to the reaction, B-ME-ß-mercaptoethanol, DTT-Dithiothreitol, HQ-hydroquinone, TCEP-tris (2-carboxyethyl) phosphine, NAC-N-acetyl L-cysteine, L-glu-L-glutathione, KI-potassium iodide. Error bar represents standard deviation, n=4.

Studies on C-P4H suggested that it can catalyse the uncoupled decarboxylation of the co-substrate αKG i.e., the reaction in the absence of prime substrate (also known as partial reaction). It requires AscA at an almost stoichiometric amounts, as the partial reaction generate Fe^{3+} which will inactivate the enzyme (Myllyla et al. 1984). To check whether TET enzymes undergo a partial reaction, the purified mTET1-CD was incubated in the absence of

substrate and AscA at room temperature up to 30 minutes. After incubation, the reaction was initiated by addition of the substrate DNA. Contrary to C-P4H, incubation of mTET1-CD without substrate and AscA did not affect the activity of mTET1-CD (Figure 27) indicating that there was no substantial partial catalysis occurring and that there was no absolute requirement of AscA for TET enzymes to perform the catalysis.

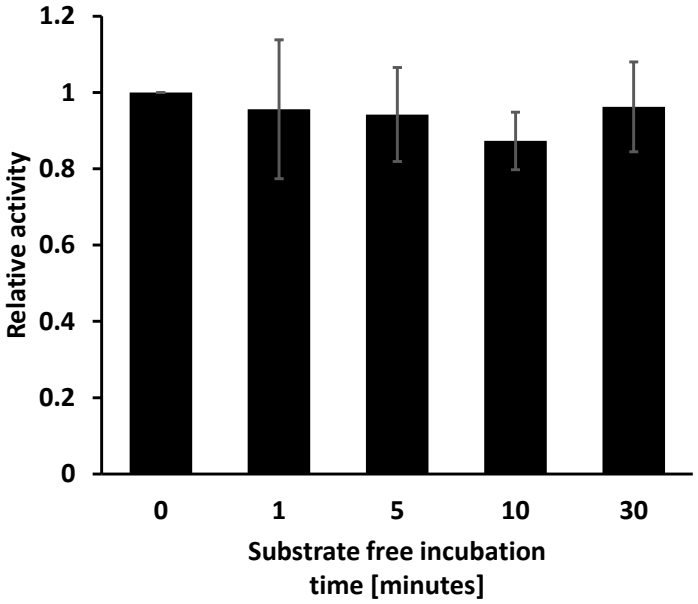


Figure 27: Probing for uncoupled oxidation reaction. The enzyme mTET1-CD is incubated for different time period without the substrate DNA in the absence of AscA. The reaction was started by adding the substrate DNA and the oxidation of 5mC to 5hmC catalysed by mTET1-CD was measured. Error bar represent standard deviation, n=3.

4.3.3. pH plays an important role in the *in vitro* activity of TET enzymes

The previous experiments showed that AscA was required only when Fe^{2+} was limited in the reaction mixture. This prompted to compare the reaction conditions used in this study with the condition reported by Yin et al (Yin et al. 2013). Comparison of two reaction conditions showed a few differences (Table 12). For example, HEPES pH 6.8 without DTT and ATP were used in this study, whereas in the published report HEPES pH 8.0 with 1 mM DTT and 1 mM ATP were used. As the rate of Fe^{2+} oxidation is strongly dependent on pH and at pH 8, the oxidation of Fe^{2+} to Fe^{3+} is ~100x times faster than at pH 6.8 (Morgan and Lahav 2007), it was hypothesized that the stimulatory effect upon addition of AscA observed by others was due to the reduction of Fe^{3+}, which is rapidly generated at pH 8.0, by AscA.

To verify this, kinetics were performed at both conditions (Figure 28). Interestingly, when the kinetics performed with or without AscA under Yin et al conditions were compared, increase in the TET enzymatic activity was observed in the presence of 1 mM AscA compared to the control reaction (red bars in figure 28). This in agreement with the published report, where ~ five-fold increase in activity was obtained with AscA compared to the control (Yin et al. 2013). However, when the reactions performed with Yin et al condition were compared with the reactions performed under the optimised condition used in this study, the overall activity of mTET1-CD was ~15 fold higher than the activity obtained with Yin et al (Yin et al. 2013). Importantly, no significant difference in mTET1-CD activity was observed between AscA treated and untreated sample under the optimised reaction condition used in this study. This indicated that the reaction condition, especially the pH, played a major role in the difference observed with the effect of AscA between two studies.

Table 12: Comparison of the conditions used by Yin et al (Yin et al. 2013) and conditions used in this study. Different components used in the reaction (except DNA and Enzyme) are listed below.

Components	Conditions used in this study	Condition used by Yin et al(Yin et al. 2013)
HEPES	0.5 M, pH 6.8 (optimal pH) or pH 8.0	0.5 M, pH 8.0
αKG	1 mM	1 mM
Fe^{2+}	10 µM or 100 µM	10 µM or 100 µM
Ascorbic acid	1 mM	1 mM
DTT	Not used	1 mM
ATP	Not used	1mM
Sodium chloride (NaCl)	150 mM	50 mM

To make sure that the difference was due to the pH and not due to other components in the reaction, the kinetics was repeated with the optimised reaction condition used in this study, but at pH 8.0, instead of pH 6.8. Interestingly, the results obtained with pH 8.0 under the reaction conditions used in this study (Figure 28 B- grey bar) were quite similar to the results obtained using Yin et al condition (Yin et al. 2013), which pinpointed that pH was the key factor and the difference in pH of the buffer resulted in the different effect observed with TET activity.

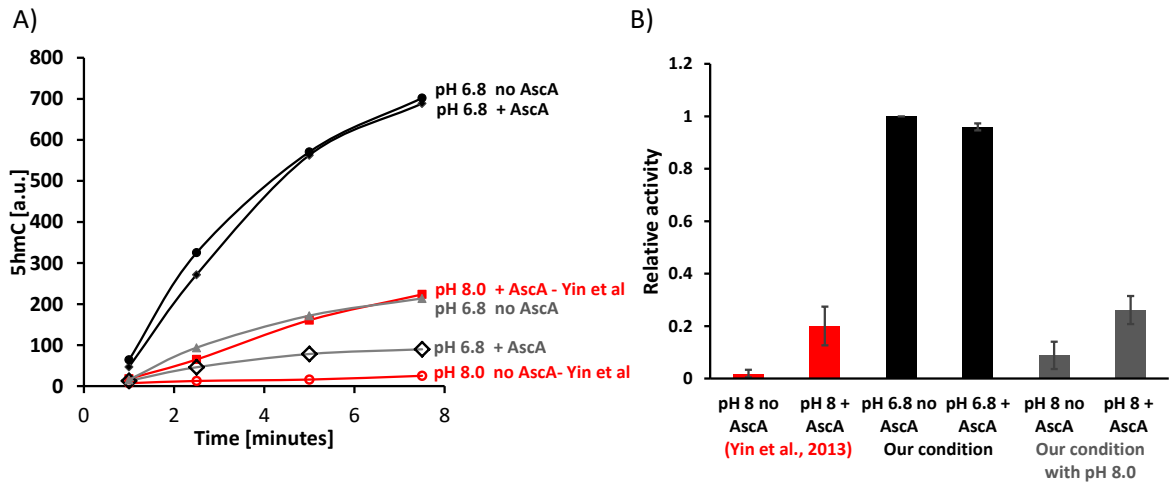


Figure 28: The kinetics performed with mTET1-CD under different conditions. The overall activity of the enzyme is higher at pH 6.8 compared to pH 8.0. A) The primary kinetics showing the oxidation of 5mC to 5hmC catalysed by mTET1-CD in the presence or absence of AscA under different conditions. B) The relative activity of mTET1-CD under different conditions with 10 µM Fe^{2+}, supplemented with or without AscA. The black bar represents the conditions used in this study, the red bar represents the conditions used by Yin et al (Yin et al. 2013) (red bars), the grey bar represents the conditions used in this study, but with pH 8.0. Error bar represents standard deviation, n=3. a.u. represents arbitrary unit.

To verify the effect of pH on Fe^{2+} oxidation, the kinetics of iron oxidation under TET reaction conditions was followed at pH 6.8 and pH 8.0 using ferrozine method (Verschoor and Molot 2013). Quantification of Fe^{2+} at two different pHs showed that at pH 8.0, the concentration of Fe^{2+} in the reaction decreased rapidly and the level decreases to almost zero within five minutes. While at pH 6.8, the concentration of Fe^{2+} stayed constant for at least 30 minutes, which can also be seen with the colour developed after ferrozine binds to Fe^{2+} (Figure 29 A-C). This result provides the explanation for the differences observed between two pH i.e. at pH 8.0, Fe^{2+} undergoes rapid oxidation and the addition of AscA rescues the activity of the enzyme by partially regenerating Fe^{2+}; whereas at pH 6.8, the level of Fe^{2+} is stable, hence the overall activity is higher and no stimulation is observed, which reinforces that AscA is merely a reducing agent.

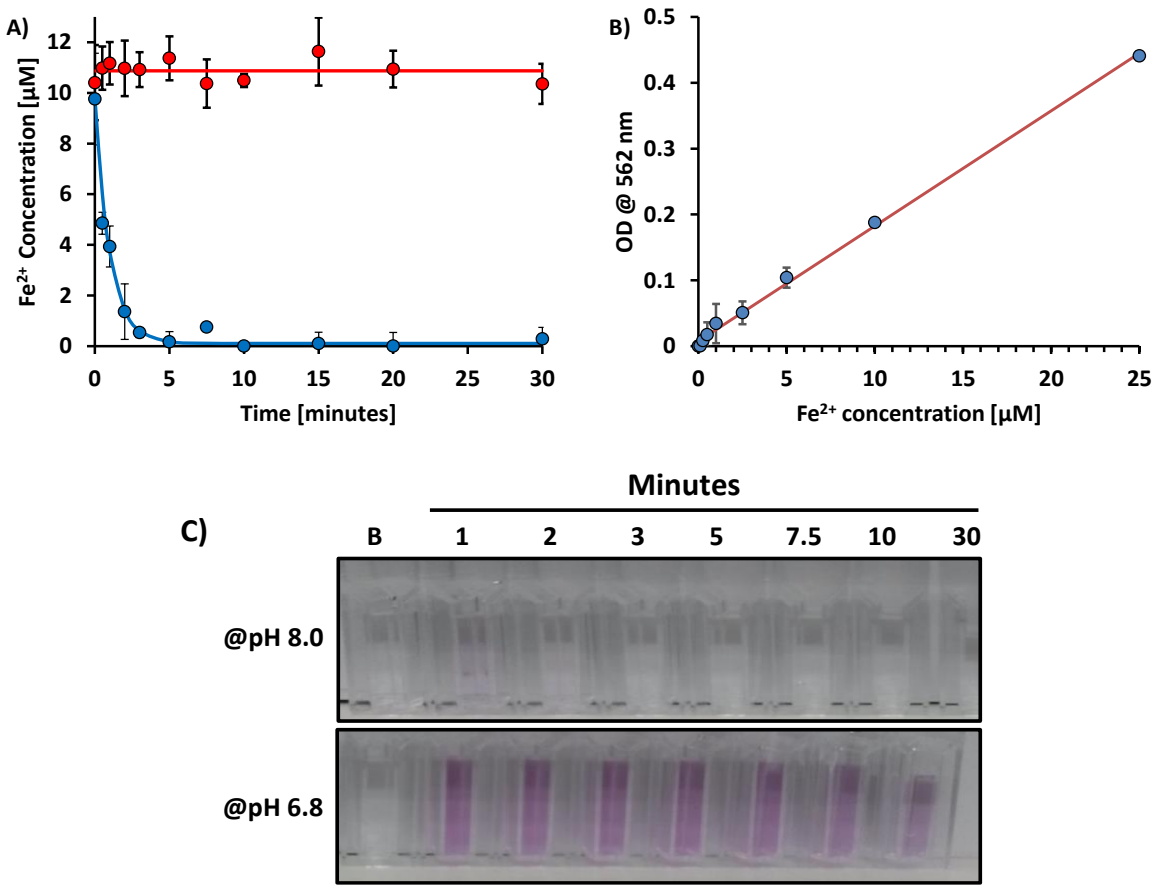


Figure 29: Oxidation of Fe^{2+} in a pH dependent manner. A) The concentration of Fe^{2+} as a function of time at pH 6.8 (red line), pH 8.0 (blue line) measured using ferrozine method (Verschoor and Molot 2013). B) The standard curve represents the OD measured at 562 nm using ferrozine method (Verschoor and Molot 2013), as a function of Fe^{2+} concentration. Error bars represents standard deviation, n=2. C) The picture shows the concentration of Fe^{2+} present at pH 8.0 and pH 6.8 measured using ferrozine method. The colour development represents the concentration of Fe^{2+} present in the solution. At pH 8.0, the concentration of Fe^{2+} decreases rapidly (almost zero-within 5 minutes), while at pH 6.8, Fe^{2+} stays constant (stayed pink) until 30 minutes.

In order to show that the effect obtained with AscA was not only specific for mTET1-CD, the kinetic assays were performed with other two family members in the presence or absence of AscA under Fe^{2+} or Fe^{3+} condition. Similar to mTET1-CD, the addition of 1 mM AscA showed no significant stimulation in the activity of mTET2-CD and mTET3-CD (Figure 30). This indicates that the effect of AscA on the enzyme activity is common to all three members of TET family.

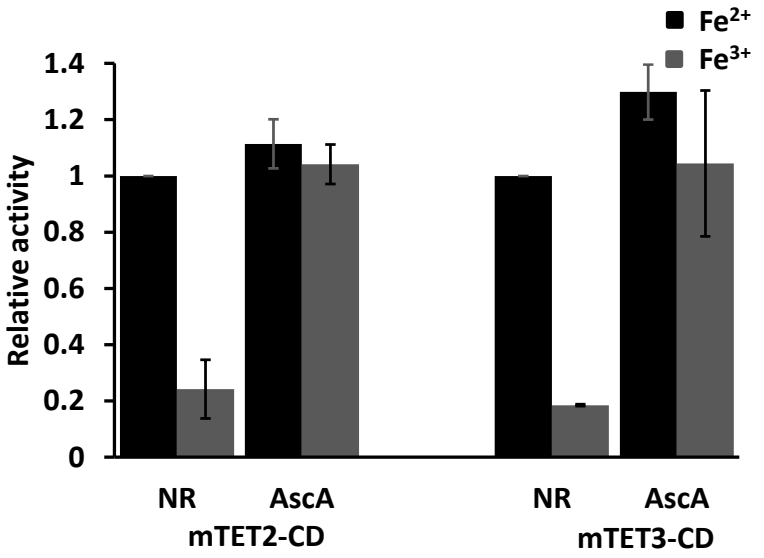


Figure 30: No stimulation in the activity of mTET2-CD and mTET3-CD is observed after addition of AscA under Fe^{2+} (100µM). AscA restores the activity under Fe^{3+} condition. The figure shows the relative activity of mTET2-CD and mTET3-CD under Fe^{2+} (black bars) or Fe^{3+} (grey bars) condition (100 µM) in the presence or absence of AscA (1 mM). NR represents no reducing agent included in the reaction. Error bar represents standard deviation, n=3.

4.3.4. Ascorbic acid does not stimulate the TET activity as a bound cofactor

The experimental evidences obtained so far suggested the role of AscA as a reducing agent However, I wanted to check whether Fe^{3+} is reduced by AscA in its free form in the solution or as bound to the enzyme's catalytic centre. For this, the binding affinity of Fe^{3+} to the catalytic pocket was analysed, as this indirectly provides information about where the reduction of Fe^{3+} might takes place (in the solution or as enzyme-bound manner). To this end, the competition assay was performed with fixed concentration of Fe^{2+} (10µM) and increasing concentrations of Fe^{3+} (0-5000 µM). As it is evident from figure 31, more than 100 fold excess of Fe^{3+} (around 1000 µM) was needed to displace around half of the bound Fe^{2+} in the catalytic pocket (as 1000 µM Fe^{3+} was required to reduce the activity of mTET1-CD by half). This indicates that Fe^{3+} does not bind strongly to the active centre.

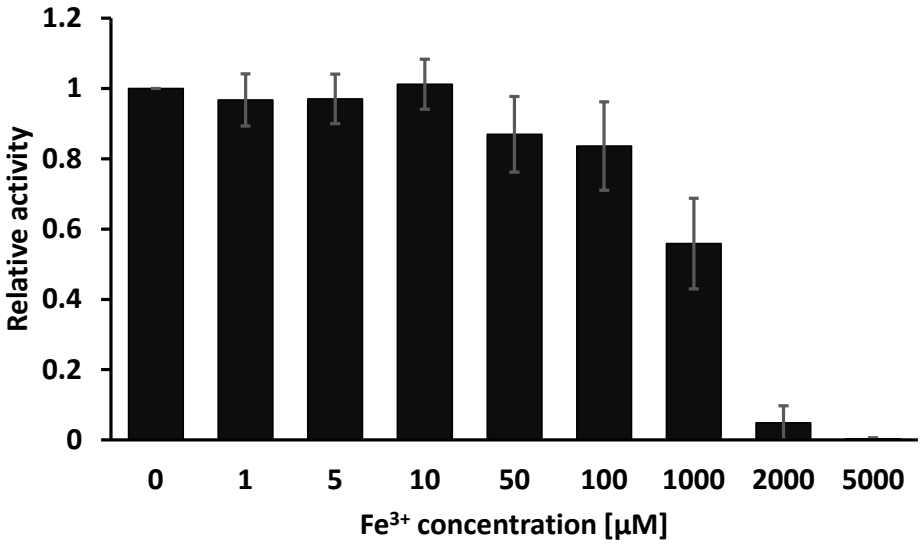


Figure 31: The competition assay to check the binding affinity of Fe^{3+}. The relative activity of mTET1-CD was measured with the reaction containing 10 µM Fe^{2+} and increasing concentrations of Fe^{3+}. Error bar represent standard deviation, n=4.

Moreover, inspection of the human TET2 catalytic domain crystal structure (Hu et al. 2013) revealed that after the binding of Fe^{2+}, the substrate base and the co-substrate αKG in the active centre, there is no sufficient space for the AscA to bind in the catalytic centre (Figure 32 A). Therefore, it was hypothesised that in order for AscA to access the catalytic iron, it would have to compete with either the substrate or the co-substrate for its binding, suggesting that at working concentration, due to the lack of sufficient space in the enzyme active pocket, AscA would rather behave as a competitive inhibitor than an enhancer of the catalytic activity. To validate this experimentally, the activity measurement of mTET1-CD was performed with different concentrations of AscA (Figure 32 B). From the result, no significant decrease in the activity was observed up to 2 mM AscA in the reaction and the calculated app.IC50 was 8.3 mM, which is quite high compared to the working concentration suggesting a non-specific inhibition of mTET1-CD, potentially by sequestering Fe^{2+} by AscA at high concentrations of AscA. Thus, this experiment indicates that AscA may not bind within the active centre of TET enzymes.

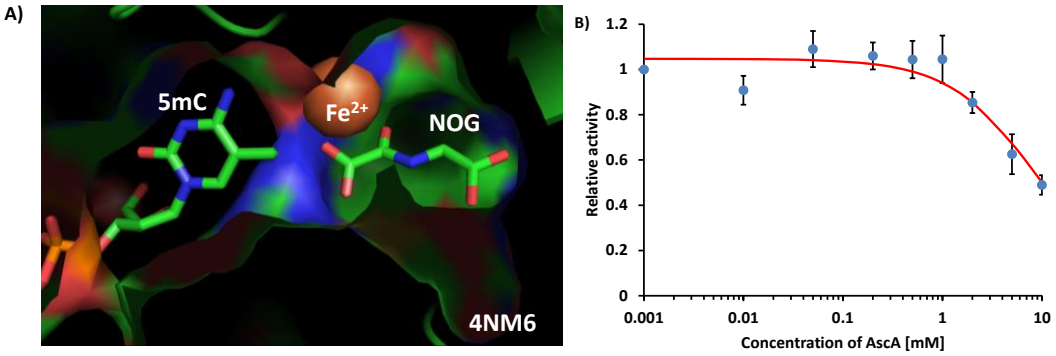


Figure 32: AscA may not bind to TET enzymes A) Crystal structure of the human TET2 catalytic domain (PDB ID: 4NM6) centre around the catalytic iron. The substrate 5mC and DMOG- an analogue of the co-substrate αKG are shown. B) The relative activity of mTET1-CD with increasing concentrations of AscA. Error bar represents standard deviation, n=3.

In the report published by Yin and co-workers (Yin et al. 2013), fluorescence quenching experiment was shown as a proof of AscA binding to TET enzymes. It was demonstrated that the binding of AscA to mTET2-CD decreased the intrinsic fluorescence of the protein in a concentration dependent manner. This effect was highly pronounced for the L and D-form, although the D form had lower fluorescence intensity than the L-form, but the same effect was not observed for the oxidation resistant derivatives of AscA (Yin et al. 2013). Similar experiment was performed in this study to check the fluorescence intensity upon addition of AscA. Different concentrations of AscA (0, 100, 200, 300, 500, 750 and 1000 µM) was tested with 2 µM of mTET1-CD enzyme. As reported, the fluorescence quenching was observed with increased concentration of AscA (Figure 33- panel 1) in a concentration dependent manner. However, when other unrelated proteins, which do not bind AscA and whose function is independent of AscA, such as EcoDam, CcrM and DNMT2 were used, similar fluorescence quenching was observed (Figure 33). This shows that the quenching observed is not due to the binding of AscA and is most likely an artefact, since this effect was consistent with all proteins tested, regardless of their AscA requirement.

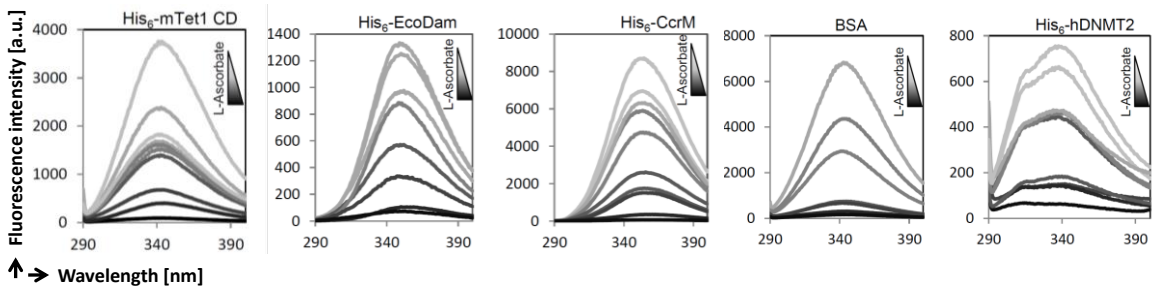


Figure 33: Representative image of the fluorescence quenching experiment. Decrease in the fluorescence intensity was observed with increase in AscA concentration, irrespective of the proteins used. a.u. represents arbitrary units. Note that the scale of y axis is different.

4.3.5. Ascorbic acid stimulates the activity of TET enzymes by providing more Fe^{2+} inside the cells

Although no stimulation was observed with AscA under optimized *in vitro* condition, increase in the level of 5hmC was observed in the cell culture experiments upon addition of AscA (Blaschke et al. 2013; Minor et al. 2013; Yin et al. 2013). As AscA can reduce Fe^{3+} to Fe^{2+}, an intracellular increase of Fe^{2+} would enhance the activity of TET enzymes. To better understand this, Michaelis-Menten kinetics was performed and the activity of mTET1-CD was measured with increasing amount of Fe^{2+}. The $K_{d\ (app)}$ of Fe^{2+} calculated from the experiment was 0.41 ± 0.05 µM (Figure 34), which falls in the range of labile iron reported in the resting erythroid cells (Epsztejn et al. 1999). This observation suggests that the alteration of Fe^{2+} in this range inside the cells can increase the activity of TET induced production of 5hmC.

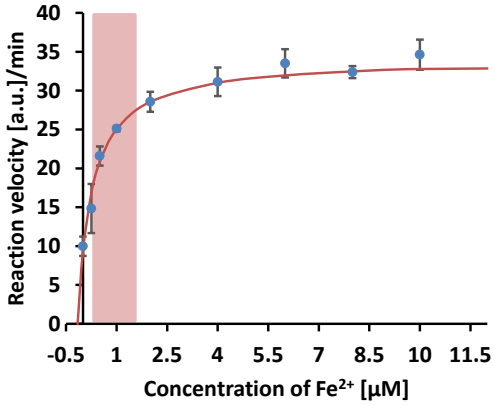


Figure 34: Determination of the apparent dissociation constant of Fe^{2+} using Michaelis-Menten reaction. The reaction was performed with increasing concentrations of Fe^{2+} and the respective initial velocity was plotted as a function of the concentration of Fe^{2+}. The calculated $K_{d\ (app)}$ is 0.41 ± 0.05 µM. The shaded part represents the range of Fe^{2+} (0.2-1.5 µM) present in the labile iron pool in resting erythroid cells (Epsztejn et al. 1999).

Overall, these results demonstrated in this chapter using an extensive biochemical assays show that AscA potentiates the activity of TET enzymes through recycling of Fe^{2+} and importantly it does not act as a bound co-factor. Moreover, it was demonstrated that under optimal reaction condition (when there is enough Fe^{2+}), AscA is not strictly necessary for the activity of TET enzymes, but when Fe^{2+} is below the level required by the enzyme, AscA helps the enzyme by providing more Fe^{2+} through efficient recycling of oxidized iron.

4.3.6. Retinol and retinoic acid increase the level of TET2 and TET3, but do not alter the catalytic activity of TET enzymes

Besides AscA, our collaborators (Dr. Wolf Reik and co-workers, Babraham Institute, Cambridge, UK) observed that retinol - the most common form of vitamin A, increased the efficiency of iPS cells generation through the TET induced formation of 5hmC. Moreover, they observed that the addition of both AscA and retinol synergistically increased the TET mediated formation of 5hmC, thereby, improving the derivation of induced pluripotent cells. Detailed investigation revealed that retinol increases the mRNA expression of TET2 and TET3 by increasing the binding of RAR enzyme to its RAR element (RARE) present on the intron.

To verify whether retinol or retinoic acid affects the enzymatic activity of TET enzymes, in this study, I performed the *in vitro* kinetics using the purified mTET1-CD. Addition of retinol or retinoic acid did not result in increased production of 5hmC under Fe^{2+} or Fe^{3+} condition, even after co-incubating the vitamin A for 30 minutes with mTET1-CD and ferrous iron (Figure 35). This suggests that both retinol and retinoic acid do not affect the catalytic activity of the enzyme. This is in line with the cellular studies observed by Dr. Wolf Reik and co-workers that the forced expression of either TET1 or TET2 (through transfecting the gene encoded in plasmids) in TET TKO cells in the presence of retinol did not stimulate the level of 5hmC.

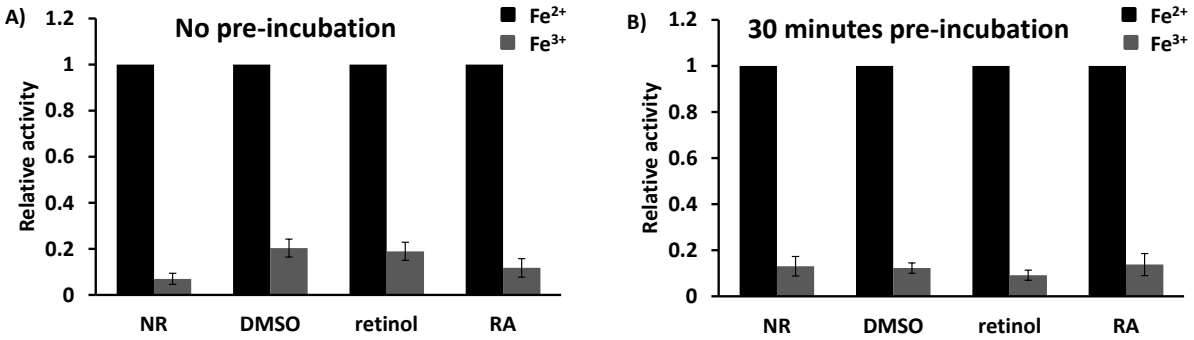


Figure 35: The catalytic activity of mTET1-CD is unaffected by retinol and retinoic acid. A) The relative activity of mTET1-CD performed under Fe^{2+} or Fe^{3+} conditions in the presence of either retinol or retinoic acid. B) The relative activity of mTET1-CD performed under Fe^{2+} or Fe^{3+} conditions after incubating the enzyme and iron with retinol or retinoic acid for 30 minutes. DMSO was used as a control, as the solution was dissolved in the solvent DMSO. NR represents the control reaction, where no form of vitamin A is added. Error bar represents standard deviation, n=4.

Collectively, these results show an interesting phenomenon that both vitamin A and vitamin C modulate the TET mediated formation of 5hmC. AscA enhances the catalytic activity of TET enzymes by providing Fe^{2+}, while vitamin A increases the expression level of *TET2* and *TET3* and both vitamins together enhances the iPS cell generation in a concentration dependent manner through the erasure of epigenetic memory.

4.4. Investigation of the biochemical and kinetic behaviour of TET enzymes

Most of the studies on TET enzymes were focused on elucidating their biological role and their reaction products; however, the intrinsic biochemical properties of TET enzymes that govern their function are not completely understood. For example, little is known about how TET enzymes choose the target sequence. Do TET enzymes exhibit any flanking sequence preference (in the context of CpG sites)? How specific are they towards CpG sites? How do they prefer different modified base (5mC, 5hmC and 5fC)? How do they catalyse the stepwise oxidation on one site (5mC to 5caC) and how do they oxidize multiple 5mC (also 5hmC or 5fC) substrates on a single DNA strand (from one 5mC/5hmC/5fC substrate to another), in a distributive or in a processive manner? Although the solved crystal structure of both hTET2-CD and nTet in complex with the DNA (both 5mC and 5hmC) have provided some insight into the behaviour of TET enzymes, (Hu et al. 2013; Hashimoto et al. 2014, 2015; Hu et al. 2015) the detailed biochemical evidence is still lacking.

In this work, the catalytic and kinetic behaviour of TET enzymes was investigated using single molecule approach under the optimised reaction condition, which would provide further insight into the function of TET enzymes in a genomic context. Existing biochemical evidences showed that TET enzymes especially hTET2, preferred the substrates in CG context over non-CG context and had weak or almost no preference towards different flanking sequences (Hu et al. 2013). These conclusions were drawn based on the structural study, which showed that the TET enzymes did not interact with the neighbouring bases other than the target 5mC. This was further supported by an *in vitro* experiment with the target 5mC in different dinucleotide context CpN (N=A or T or G or C) (Hu et al. 2013). However, these experiments were performed on few substrates with a defined flanking sequences that did not probe all potential sequence contexts and therefore provides a very limited view. Moreover, the CpG flanking sequence strongly influences other DNA modifying enzymes, like DNMT3A or DNMT3B (Handa and Jeltsch 2005; Jurkowska et al. 2011b) and hence intrinsically provides a bias for faster and more efficient modification of the preferred sites.

In order to address this issue comprehensively, the activity of TET enzymes was tested on the DNA containing multiple substrate bases (5mC/5hmC) with the target C in both CG and non-CG contexts and analysed at single molecule and base resolution. The main idea of this

experiment was to offer TET enzymes, a collection of substrates containing 5mCs/5hmCs embedded in different sequence flanks and monitor the 5mC/5hmC oxidation catalysed by TET enzymes on each target cytosine using bisulfite conversion and next generation sequencing (NGS). Using this approach, 5mC/5hmC oxidation pattern would be generated that would provide a blueprint of the TETs activity and their sequence preference. Comparing the level of oxidation at each cytosine would provide information on the enzymes' preference towards different substrate.

To achieve this, a synthetic "CG-rich" substrate DNA was generated, where each cytosine (C) in the sequence was substituted with 5hmC or 5mC during PCR amplification step (refer the section 3.4.2). The modified substrate DNA was treated with 2 µM of different TET enzymes: mTET1-CD, mTET2-CD-Del5, mTET3-CD-Del1 and nTet and the reactions were stopped at increasing time points. Subsequently, the purified DNA was ligated with Illumina adapters. As a first attempt, the 5hmC containing DNA was used as a substrate and its oxidation to 5fC or 5caC was analysed. Bisulfite sequencing cannot distinguish 5mC and 5hmC bases, as both these bases are read as C, however 5fC and 5caC can be deaminated to uracil and can be read as thymine (T). The readout of 5mC and its oxidation products after the bisulfite treatment is shown in table 13. Following the bisulfite treatment and PCR amplification, the samples were sequenced on the Illumina MiSeq platform.

Table 13: The readout of modified cytosine bases after bisulfite sequencing

Base	Direct sequencing	After bisulfite treatment
C	C	T
5mC	C	C
5hmC	C	C
5fC	C	T
5caC	C	T

The activity of different TET enzymes on 5hmC containing CG-rich substrate (both forward and reverse) is summarized in figure 36. Each row represents a different sequencing sample (collected at different time point or treated with a different TET enzymes) and each column represents an individual cytosine within the substrate DNA sequence. The presented results for each experiment were obtained from at least 10,000 sequence reads obtained during the Illumina MiSeq 2x300 PE sequencing. The CG-rich double stranded substrate DNA had a total of 122 Cs within both forward and reverse strand. Out of 122 Cs, only 100 were modified

during PCR amplification step. Remaining 22 unmodified cytosine were present within the PCR primers labelled as the primer binding region - first 10 cytosine in the forward strand and 12 cytosine in the reverse strand, served as an internal control for bisulfite conversion. The results were colour coded, 5hmC is represented as red and the level of oxidation of 5hmC to 5fC/5caC is represented by the colour gradient changing from red to blue with blue being 90-100% oxidized, red being not oxidized (0-10% oxidized) and the intermediate colour (from red to blue) represents 10-90% of oxidation. The blue on the primer binding region represents the unmodified cytosine.

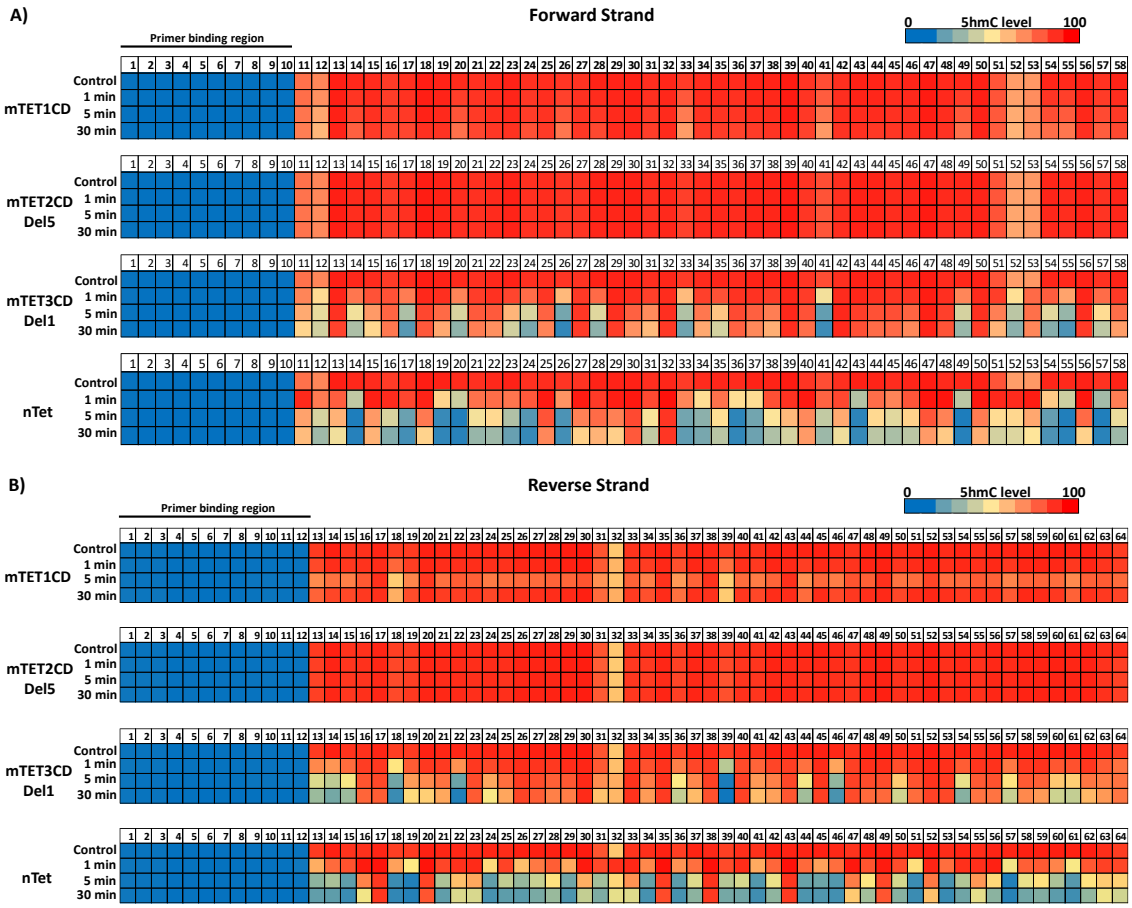


Figure 36: The achieved 5hmC oxidation patterns obtained for different TET enzymes represented as heat maps. The CG-rich substrate containing multiple 5hmC modification incorporated using PCR is treated with different TET enzymes (from both mouse and *Naegleria gruberi*) and the activity of each enzyme at all cytosine at different time points are analysed and represented here. Each column represents an individual cytosine which is analysed and numbered. The blue represents either unmodified cytosine (Cs present in the oligos) or 5fC or 5CaC and the red represent 5hmC. The primer binding region contains unmodified cytosine and serves as a control for bisulfite conversion. A) represents all cytosine present in the forward strand and B) represents all cytosine present in the reverse strand of the CG-rich substrate from 5' to 3'. Each square is the result of at least 10,000 sequence reads.

Table 14: The average bisulfite conversion of all unmodified C in the sequence (22 Cs)

Time (minutes)	mTET1 CD	mTET2 CD-Del5	mTET3 CD-Del1	nTET
0	98.8	98.8	98.8	98.6
1	98.8	98.6	98.7	98.7
5	98.8	98.8	98.5	98.7
30	98.7	98.7	98.7	98.8

The conversion of unmodified cytosine (C) to thymine (T) after bisulfite treatment was calculated for all 22 Cs on the primer binding region (Table 14). This shows that the bisulfite conversion was successful and the overall conversion rate was around 99%. Moreover, the average level of 5hmC was also calculated. The control experiment or time point '0', where no TET enzyme was added (Figure 36) showed the total 5hmC content of about 90% (Table 15), suggesting that not all cytosine (C) were substituted with 5hmC during PCR amplification.

Table 15: The average oxidation of all 5hmC (100 Cs in total except the Cs on the oligo) on the CG-rich substrates by different TET enzymes.

Time (minutes)	mTET1-CD		mTET2-CD-Del5		mTET3-CD-Del1		nTET	
	5hmC	5fC/5cac	5hmC	5fC/5cac	5hmC	5fC/5cac	5hmC	5fC/5cac
0	89.92	0.00	89.92	0.00	89.92	0.00	89.92	0.00
1	88.49	1.43	89.30	0.65	81.92	8.00	77.90	12.02
5	85.97	3.95	89.40	0.55	68.67	21.25	45.20	44.72
30	80.35	9.57	88.92	1.03	60.07	29.85	34.56	55.36

Interestingly, the incorporation of unmodified Cs was not completely random and it accumulated in the sequence containing stretches of Cs, suggesting that 5hmC was less efficiently incorporated in these DNA sequence and the 5hmdCTP used for amplification in dNTPs mixture contained either some of the unmodified dCTP or 5fdCTP/5cadCTP. Nevertheless, after treating the 5hmC modified DNA with TET enzymes, conversion of 5hmC to 5fC/5caC was observed for all the enzymes used. Importantly, the accumulation of the reaction products increased over time, as can be seen from the change in the colour gradient (red to blue, where red is 5hmC and blue is oxidation of 5hmC to 5fC or 5caC). Interestingly, the results showed that the pattern of oxidation varied for different TET enzymes (Figure 36). First of all, not all TET enzymes oxidised the CG-rich DNA substrate equally efficiently (Table 15). Importantly, not all 5hmCs on the CG-rich substrate were oxidised equally and the rate of oxidation differed among 5hmC sites on the DNA strand. This emerging pattern of oxidation

for an individual TET enzyme indicates that there are differences in substrate specificity among TET enzymes.

Comparison of the overall oxidation on the CG-rich DNA after 30 minutes reaction showed that nTet oxidized around 60% of the total 5hmC substrate, whereas mTET3-CD-Del1 showed ~ 30% oxidation followed by mTET1-CD (~10%) and mTET2-CD-Del5 being the lowest with an average oxidation of around 1%.

However, a closer look at the 5hmC- site specific oxidation by different members of the mammalian TET revealed that despite the enzymes having different total level of 5hmC oxidation, the modification pattern was similar. For example, to mention a few sites, all three mammalian TET enzymes (TET1, TET2 and TET3) oxidized 5hmC at C14 (GCA**CG**CCT), C20 (ACA**CG**CGC), C26 (AGA**CC**TGC), C33 (TGC**CG**GCG) on the upper strand, C18 (AGC**CG**TCG), C31 (GCT**CC**TAG), C36 (AGC**CG**CGC), C39 (CGC**CG**ATC) on the lower strand. (Note: here all C in the sequence are 5hmC and the target 5hmCpN dinucleotide is highlighted). This suggests the different catalytic efficiencies of different members, where mTET3-CD-Del1 oxidized 5hmC substrates more efficiently than mTET1-CD and mTET2-CD-Del5.

On the other hand, the pattern of oxidation exhibited by nTet was different compared to the mammalian TETs. nTet oxidised almost all 5hmCs on the DNA equally, whereas mTET3-CD-Del1 showed a more distinct pattern with few sites being oxidized very efficiently (equal to or greater than nTet) and many others not. This suggests that the mammalian TETs display strict substrate selectivity, while nTet shows relaxed substrate selectivity. Further analysis of the data revealed an interesting phenomenon on the substrate preference between mTET3-CD-Del1 and nTet. Although nTet oxidized many of the 5hmCs (with different level of oxidation) and the overall oxidation on the CG-rich substrate was almost twice more than mTET3-CD-Del1, there were many noticeable differences observed between the two enzymes. For example, to name a few, C28 (CTG**CG**ATT), C41 (CTC**CG**AGC) on the forward strand and C39 (CGC**CG**ATC) on the reverse strand were strongly oxidized by mTET3-CD-Del1, while the same sites were weakly oxidized by nTet. This suggests that the mammalian TETs behave differently than the Tet-like dioxygenase from the amoebaflagellate, *Naegleria gruberi*. Since the overall activity of mTET3-CD-Del1 and nTet was higher than the other two enzymes, further analysis was performed on mTET3-CD-Del1 and nTet.

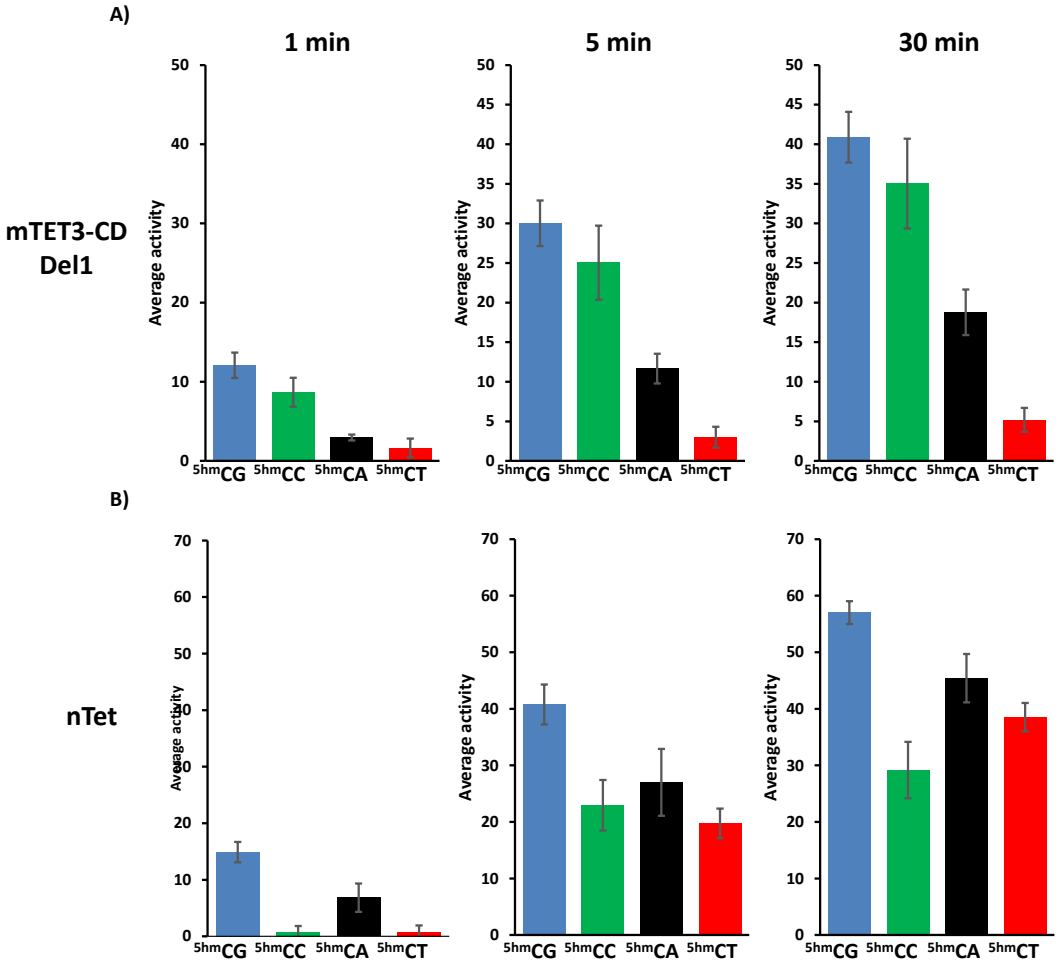


Figure 37: CpN sequence preference of A) mTET3-CD-Del1 and B) nTet. The average activity of TET enzymes on 5hmC containing DNA at 1, 5 and 30 minutes in different sequence contexts (^{5hm}CpN, where N= G, C, T, A). Error bars represents standard deviation from at least 10,000 sequences analysed.

As a next step, the substrate specificity of mTET3-CD-Del1 and nTet in 5hm**C**pN (N= A, T, G or C) context was analysed. The sequence specificity for mTET3-CD-Del1 and nTet was derived by calculating the average activity of the respective enzymes on the target sites (5hmC) having the base G, C, A or T following the target 5hmC (Figure 37). Figure 37 A shows that mTET3-CD-Del1 has higher preference for the substrate in CG context compared to non-CG context, which is in line with the published report (Hu et al. 2013). However, besides CG, mTET3-CD-Del1 oxidized 5hmCs in non-CG context as well. Interestingly, the 5hmC substrates in CC contexts were almost equally modified as 5hmC substrates in CG context, while CA was moderately oxidized and the 5hmC substrates in CT context were least oxidized. This observation contradicts the published report, where it was shown that hTET2-CD had weak activity (8 fold less activity) on the sequence in CC context and almost no activity in CA context

(Hu et al. 2013). Similar analysis on nTet showed that it highly oxidized substrates in CG context followed by CA>CT>CC (Figure 37 B).

Interestingly, comparison of the substrate specificity in CpN context between mTET3-CD-Del1 and nTet showed that both enzymes highly preferred the substrate in CG context, but they exhibited different substrate specificity for the targets in non-CG context. mTET3-CD-Del1 oxidized 5hmC substrate in non-CG sequence in the following order CC>CA>CT, whereas nTet oxidized the substrates in CA context more than CT and CC, and contrast to mTET3-CD-Del1, CC is the least oxidized by nTet.

To understand the substrate specificity better, the oxidation of 5hmC at each CpN (N= A or T or G or C) site on the DNA was sorted according to the on-site activity, divided into three major groups (0-40%, 41-70%, 71-100% of oxidation) and the total number of 5hmC sites within each group was calculated. From figure 38, it can be seen that for mTET3-CD-Del1, there is a binominal distribution of highly oxidized and less oxidized sites. The number of 5hmC sites which were less oxidized (where the oxidation is less than 40%) was more compared to the sites which were highly oxidized (oxidation-71-100%). Whereas for nTet, no such difference was observed and many 5hmC sites (in CpN context) underwent > 40% oxidation, again indicating that nTet has low substrate selectivity and can oxidize many substrates, while mTET3-CD-Del1 has more stringent substrate preference. This also explains the overall higher oxidation exhibited by nTet (~55%) compared to mTET3-CD-Del1 (~30%).

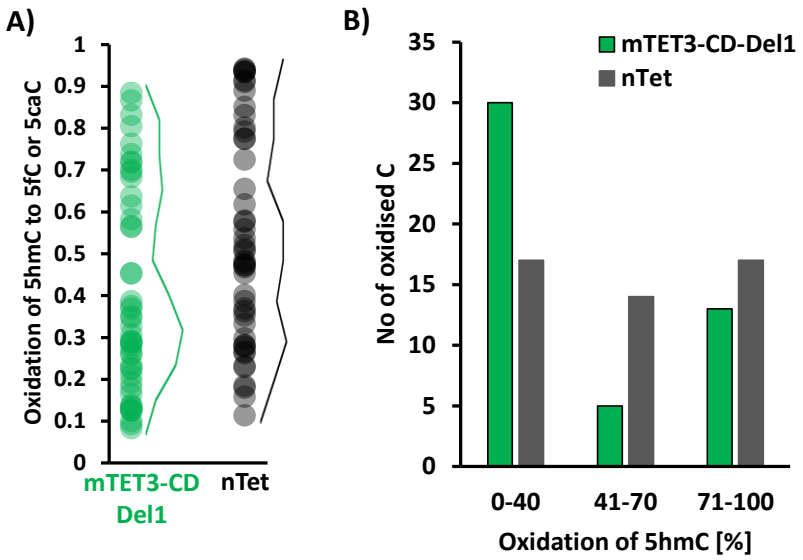


Figure 38: The average oxidation of all 5hmC on the CG-rich substrate. A) The average 5hmC oxidation of an individual CpN sites by mTET3-CD-Del1 and nTet is sorted based on the level of oxidation. Each circle represents one 5hmC site (in CpN context), which is oxidized to the level shown in y axis and the line graph next to the circle represents the number of sites with the corresponding level of oxidation. B) The oxidation level is divided into three groups, 0-40%, 41-70%, and 71-100% and the total number of 5hmC sites which are oxidized is represented as bar chart.

In order to closely investigate these differences observed among 5hmCs of mTET3-CD-Del1 and nTet, the total level of oxidation at each 5hmC embedded in a different flanking sequence context on both the upper and lower strand was sorted from the lowest to the highest for each enzyme independently. Figure 39 shows the percentage of oxidation for each analysed 5hmC of both mTET3-CD-Del1 and nTet (the reverse strand is shown here as an example).

One striking aspect (in figure 39 A) was that although the 5hmC substrates in CG context were highly preferred over non-CG by mTET3-CD-Del1, not all 5hmC sites in CpG context were equally oxidized. There were some 5hmC substrates in CG context oxidized to high level and some which were not oxidized at all by mTET3-CD-Del1 and the same pattern was observed for 5hmC substrates in both CC and CA context. This indicates that the enzyme mTET3-CD-Del1 is sensitive to the flanking sequences beyond the CpN dinucleotide.

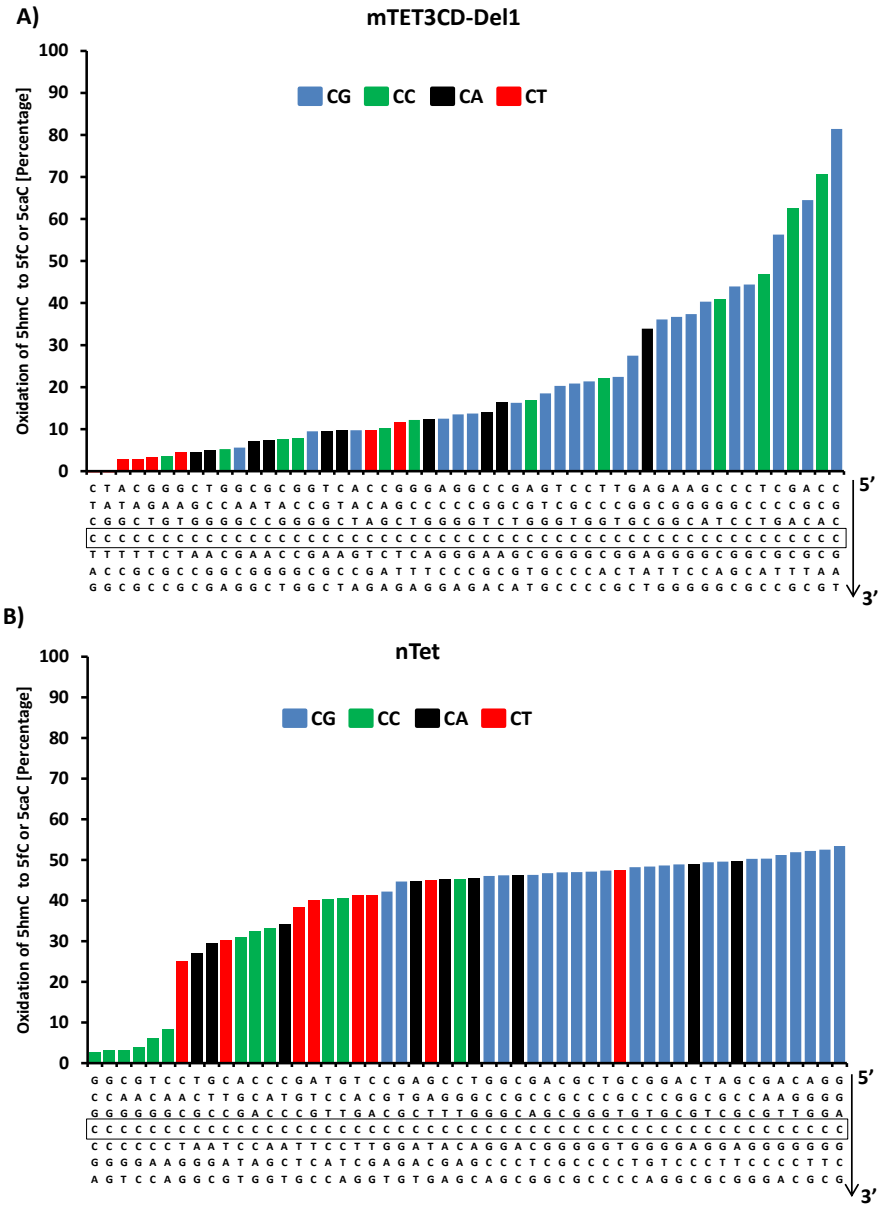


Figure 39: The level of oxidation of each 5hmCpN site on the CG-rich substrate. The oxidation of each 5hmC by (A) mTET3CD-Del1 and (B) nTet is calculated. The oxidation of 5hmC to 5fC or 5caC after 30 minutes treatment with TET enzymes compared to the control is calculated and the activity is sorted from the lowest to the highest. As an example, the reverse strand of the CG-rich substrate is show here. The sequence below represents the analysed 5hmC at position 4 (highlighted with the box) and the sequence flanking C on both sides.

The observed pattern for nTet was different, where most of the 5hmC substrates (in both CG and non-CG contexts) were nicely oxidised and very few 5hmC substrates showed low level of oxidation (Figure 39 B). Moreover, comparison of two enzymes showed that the difference in the level of oxidation between the highly preferred substrate and the least preferred substrate was more for mTET3-CD-Del1 than nTet, reinforcing that nTet can oxidize wide

variety of 5hmC substrates embedded in different sequence contexts i.e., less substrate specificity, while mTET3-CD-Del1 cannot.

To check the sequence preference beyond CpN dinucleotide globally, the flanking sequence preference for mTET3-CD-Del1 and nTet on the CG-rich substrate was extracted by taking different composition of the flanks into account. This is calculated based on the average oxidation of the target 5hmC having each base at a particular position of the flanks, for example, the activity on the sequences having guanine (G) at position +1 regardless of the other bases at other positions was averaged and is shown here .This indicates the preference of a particular base at that defined position to the overall preference (Figure 40).

A)

mTET3-CD-Del1

	-3	-2	-1	+1	+2	+3	+4
G	15.63	29.40	14.68	30.02	12.77	20.61	19.54
A	26.07	4.82	45.87	11.65	38.12	11.54	21.05
T	23.87	15.98	20.81	2.49	32.91	23.37	22.56
C	24.16	21.63	23.58	25.04	16.29	24.88	22.07
	[ACTG]	[GC]	[A]	[GC]	[AT]	[CTG]	[TCAG]

B)

nTet

	-3	-2	-1	+1	+2	+3	+4
G	29.75	40.93	25.69	40.77	22.39	33.22	29.15
A	35.12	25.82	62.14	26.99	31.03	20.03	32.71
T	30.88	17.95	35.73	19.73	35.10	17.96	28.95
C	30.29	28.56	24.43	22.94	35.41	35.25	32.71
	[ATCG]	[G]	[A]	[GA]	[CTA]	[CG]	[ACGT]

Figure 40: The flanking sequence preference of mTET3-CD-Del1 and nTet. The flanking sequence preference for TET enzymes is extracted by considering the different flanks of all 100 5hmC sites on the CG-rich substrate. The numbers are obtained by calculating the oxidation of each 5hmC (position 0, not shown here) having a particular base at a particular position regardless of the bases at other positions and this is done for 3 bases upstream of the analysed 5hmC (-3 to -1) and 4 bases downstream of the analysed 5hmC (+1 to +4), which indicates the preference of that base over others. Based on the extracted numbers, the preferred base was selected and shown in square brackets with the highly preferred followed by the less preferred base.

The position 0, not shown in the figure 40 is the target 5hmC that was oxidised by TET enzymes and +1 to +4 represents the bases upstream of the target C (5' to 3') and -1 to -3 represents the bases that precede the target C. Analysis of the extracted data for mTET3-CD-Del1 showed that the sequence containing the base adenine (A) at -1 position showed higher activity compared to others followed by the base T, while at +2 position either base A or T was

preferred. At position -2, the enzyme accepted G or C, whereas at +3, G or C or T was permitted. Beyond -2 and +3 positions, the enzyme did not show any preference for a particular base. This suggests that the first two bases flanking the CpG dinucleotide on both sides influence the enzyme activity and especially having A or T at these positions (-1 and +2) favours the activity of mTET3-CD-Del1, while having G is least favoured.

The flanking sequence preference calculated for nTet showed that similar to mTET3-CD-Del1, nTet preferred the base A at -1 position. It could tolerate base C, T or A at +2 position with slightly higher preference for C than others. Interestingly, at -2 position, the base G was highly preferred over others and +3 position C or G was favoured than A or T. Similar to mTET3-CD-Del1, no significant preference was observed beyond -2 and +3 positions for nTet. Comparison of the flanking sequence preferences between mTET3-CD-Del1 and nTet showed that both enzymes preferred the base A at -1 position, whereas they had slightly different preference at -2, +1 and +2 positions suggesting that these differences observed between two enzymes could account for the difference in the substrate selectivity.

The flanking sequence preference observed in this study is quite informative; however the experiment was performed with single DNA substrate. Therefore, it would be important to probe additional substrates, which would allow probing the sequence preference in many different sequence contexts. Nevertheless, the results obtained here indicate that both mTET3-CD-Del1 and nTet enzymes have different sequence selectivity, with nTet having wide preference compared to mTET3-CD-Del1.

As single DNA substrate provides many 5hmC sites in competition to each other, this system allows direct comparison of the kinetic rates of distinct substrates simultaneously. The total oxidation at an individual 5hmC sites was compared and 5hmC substrates which underwent high, medium and low oxidation in both CG and non-CG context by mTET3-CD-Del1 and nTet were chosen and the initial rate of oxidation was calculated (Figure 41). Comparison of the initial rate of oxidation of 5hmC substrates in CpN context by mTET3-CD-Del1 revealed that the least oxidized 5hmC substrate (GCG**CT**CTG) was ~240-fold slower than the highly oxidized substrate (CGG^{5hm}**CG**ATC) and around 60-fold slower than the moderately oxidized substrate (AGA^{5hm}**CC**TGC). Comparison of the initial rate of oxidation in the CG context showed that the

least active CG substrate (GAG**CG**CGC) showed 50-fold slower oxidation compared to the highly oxidized substrate (CGG**CG**ATC).

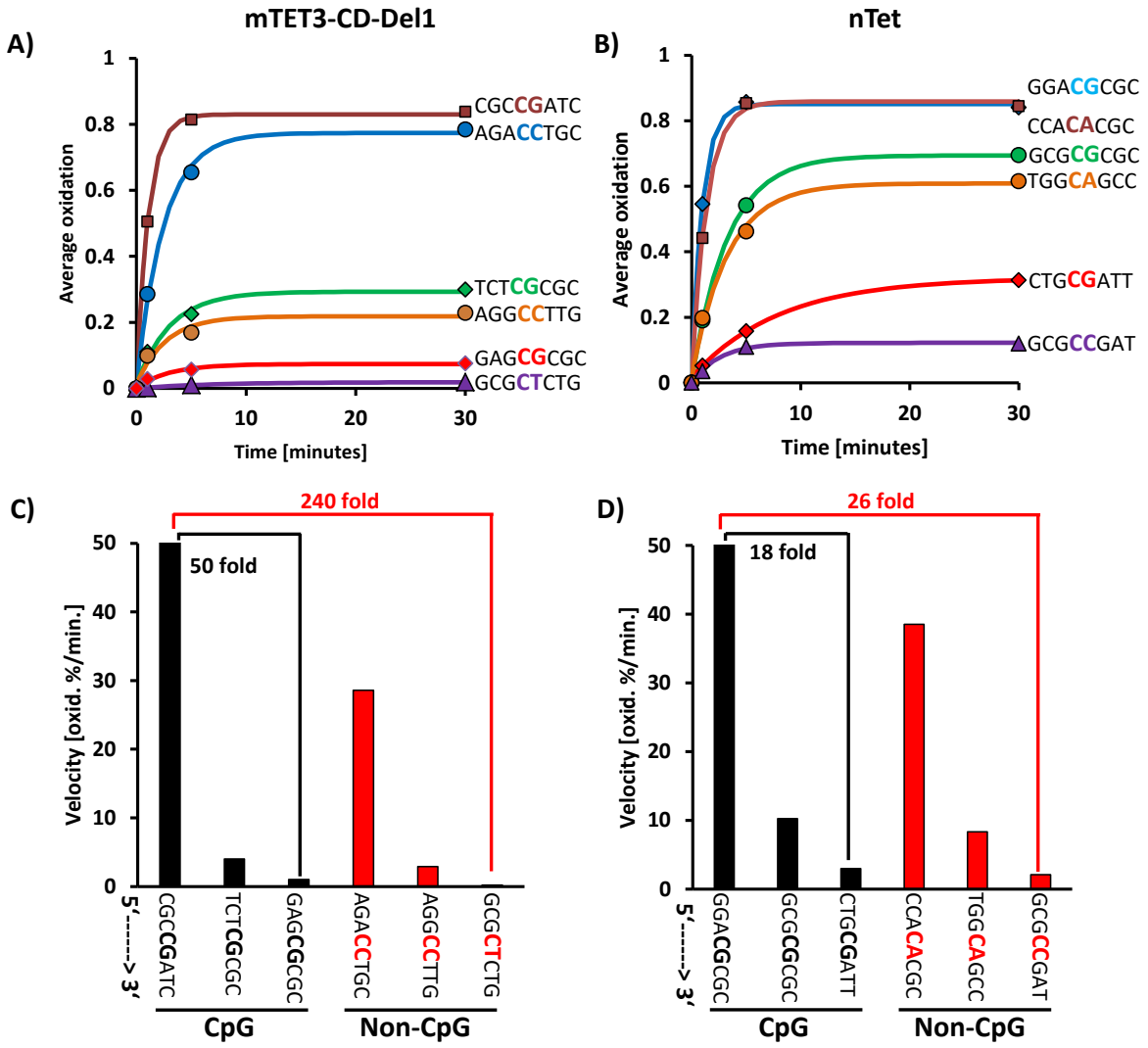


Figure 41: The rate of oxidation of different substrates by mTET3-CD-Del1 and nTet. 5hmC substrates which underwent high, medium and low oxidation in bot CG and non-CG context by both enzymes are selected and the initial velocity of the reaction was calculated. Figure A and B show the exponential fit of the reaction catalysed by mTET3-CD-Del1 and nTet respectively. The sequence of the substrate is shown with the target 5hmC in CpN dinucleotide highlighted. The initial rate of each reaction was calculated using the linear fit and is shown in figure (C) for mTET3-CD-Del1 and (D) for nTet.

In contrast, nTet did not show a drastic difference in the rate of oxidation of substrates compared to mTET3-CD-Del1. There was only 26-fold difference in the rate between the most (GGA**CG**CGC) and the least oxidized substrate (GCG**CC**GAT) and around 4-fold difference in the rate between the least and the moderately oxidized substrate (CCA**CA**CGC) by nTet. Similarly, in the CG context, the least oxidized CG (CTG**CG**ATT) showed 18-fold difference when compared to the highly oxidized substrate sequence (GGA**CG**CGC) (Figure 41). As

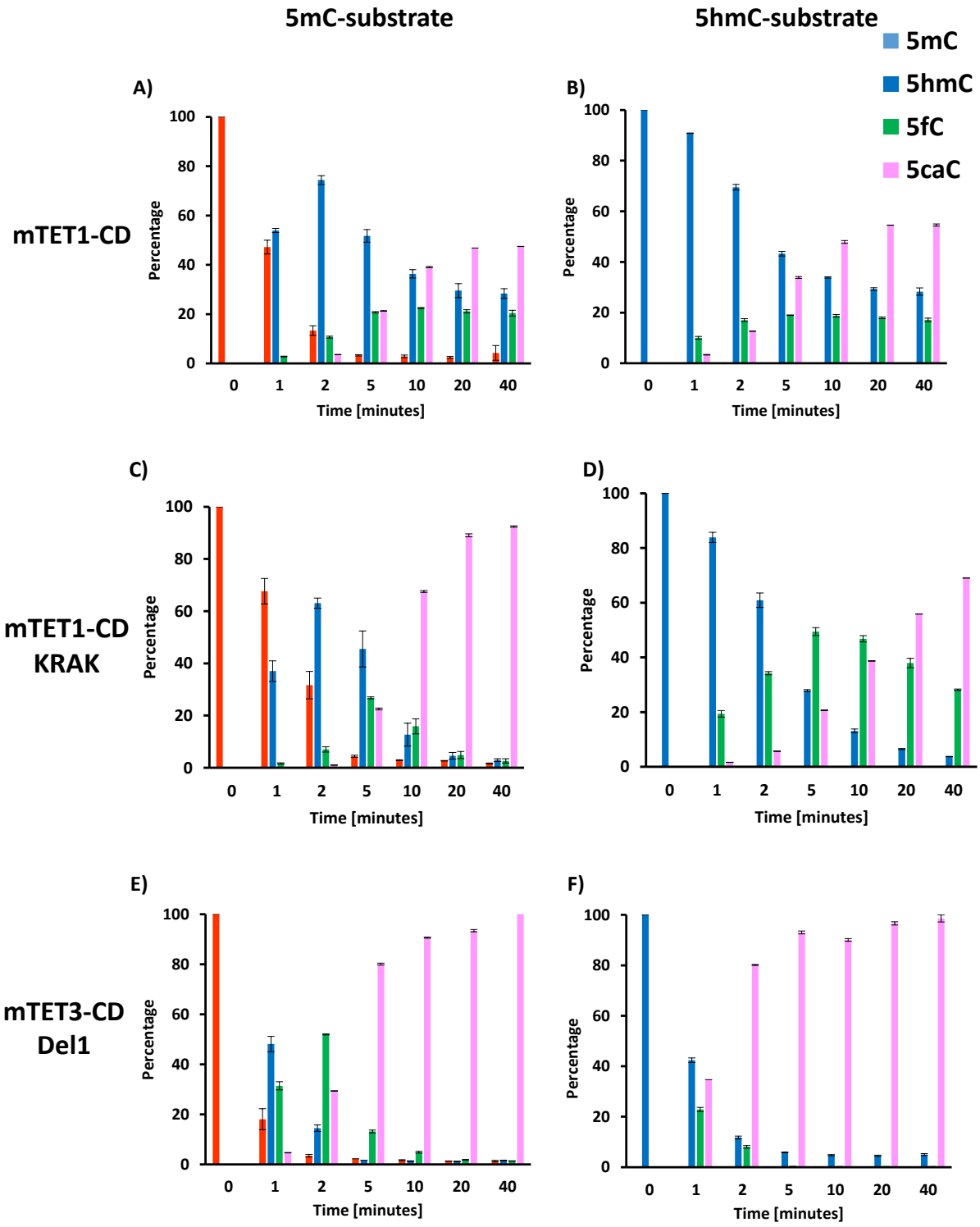
mentioned above, the difference in the oxidation rate is huge (~240-fold) for mTET3-CD-Del1, while it is not that drastic (~ 26-fold) for nTet. This strongly suggests that mTET3-CD-Del1 can oxidize at high level on selective 5hmC substrates, while nTet can perform oxidation on wide varieties of 5hmC substrate. Closer analysis showed that the highly oxidized sequences had base A or T flanking the CpG dinucleotide. On the other hand, substrate having base G at -1 position underwent the least oxidation by both enzymes and this effect was more pronounced for mTET3-CD-Del1.

To analyse the observed preference further and to look at other oxidized 5mC products catalysed by TET enzymess, the kinetics of TET enzymatic activity was analysed on the highly preferred substrate and the accumulation of different oxidized base was quantified by liquid chromatography and mass spectrometry (LC-MS) (LC-MS analysis was performed by Dr. Mark Helm and Dominik Jacob, Institute for Pharmacy and Biochemistry, Johannes Gutenberg University, Mainz, Germany). The substrate with the flanking sequence CTC**C**GAGC (the target C is highlighted) was chosen based on the substrate preference analysis and the hemi-modified substrate of length 20 bp containing a single CpG site carrying 5mC or 5hmC was designed. For the reaction, 0.5 µM of the substrate was treated with 3 µM of TET enzymes (mTET1-CD, mTET1-CD-KRAK, mTET2-CD-Del5, mTET3-CD-Del1, nTet) for different time periods and the substrate was processed further as mentioned in the section 3.20.

The results obtained from the LC-MS (Figure 42) quantification showed that all TET enzymes were active except mTET2-CD-Del5, which showed weak activity and can be seen from the accumulation of 5hmC without any further oxidation of 5hmC to 5fC and 5caC (Figure 42 I-J). Interestingly, mTET1-CD, which was less active on the previous experiment (with the CG-rich substrate) showed better activity on both 5mC and 5hmC DNA, where it oxidized around 95% of 5mC to 5hmC, 5fC and 5caC and around 70% of 5hmC to 5fC and 5caC. Moreover, the shorter version of mTET1-CD, called as mTET1-CD-KRAK was also tested on both 5mC and 5hmC substrates. It oxidized almost all of 5mC and 5hmC substrates and interestingly, the accumulation of all three oxidation species with both 5mC substrate and 5hmC substrate was more for the shorter version of mTET1-CD (mTET1-CD-KRAK) than mTET1-CD, which could be most likely attributed to the better quality of the protein. Likewise, mTET3-CD-Del1 and nTet oxidized almost all 5mC and 5hmC substrates.

An intriguing observation was that mTET3-CD-Del1 oxidized both 5mC and 5hmC substrates rapidly to 5caC within 10 minutes. The step wise oxidation of 5mC to 5caC was rapid for mTET3-CD-Del1 and the accumulation of 5fC seemed to be transient and it disappeared very quickly. Whereas, for enzymes such as mTET1-CD, mTET1-CD-KRAK and nTet, the accumulation of 5caC was slow, concomitant with the disappearance of 5fC, and moreover, 5caC did not reach 100% within 40 minutes. Another interesting observation was that the formation of 5hmC (the first oxidation reaction) was faster for mTET1-CD, mTET1-CD-KRAK and nTet compared to the formation of 5fC or 5caC (second and third oxidation reaction).

Moreover, comparison of the oxidation of 5mC and 5hmC substrates showed that almost all 5mC were oxidized quickly by all enzymes except mTET2-CD-Del5 and the level was reduced to almost 0 within 5 minutes (Figure 42). But the oxidation of 5hmC substrate was slower compared to 5mC. Interestingly, mTET1-CD and nTet oxidized only around 80% of 5hmC substrate, whereas mTET3-CD-Del1 and mTET1-CD-KRAK oxidized almost all 5hmC substrates within 40 minutes. This phenomenon of different oxidation efficiency exhibited on 5mC and 5hmC substrates is in line with the published data for hTET2-CD (Hu et al. 2015). However, in contrast to the published report, where they showed only 25% of 5hmC oxidation by mTET1-CD even at concentration as high as 12.5 µM in 40 minutes (Hu et al. 2015), our data shows that the difference is not that drastic as it was reported.


5mC-substrate
5hmC-substrate
5mC
5hmC
5fC
5caC
A)
B)
C)
D)
E)
F)
mTET1-CD
mTET1-CD KRAK
mTET3-CD Del1
Percentage
Time [minutes]
0
1
2
5
10
20
40
100
80
60
40
20

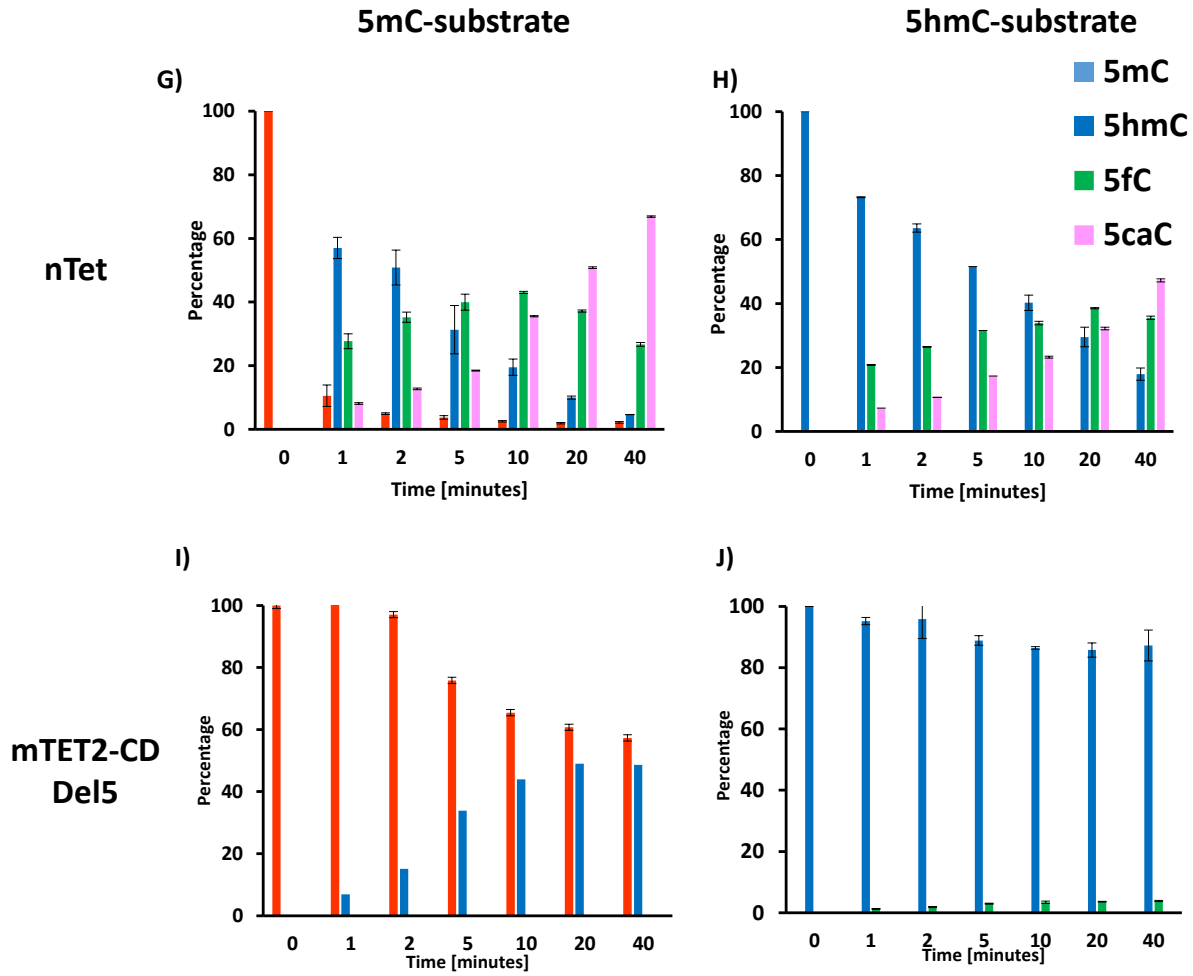


Figure 42: Quantification of the reaction products oxidized by different TET enzymes using LC-MS. 20 bp long DNA substrate having hemi-modification (either 5mC or 5hmC) was used as a substrate (0.5 µM) and treated with different TET enzymes (3 µM) for different time points. The accumulation of the oxidized bases treated with (A+B) mTET1-CD, (C+D) mTET1-CD-KRAK, (E+F) mTET3-CD-Del1, (G+H) nTet, (I+J) mTet2-CD-Del5 was detected using LC-MS. Error bar represents standard deviation, from 2 individual experiments.

Next, the rate of the reaction was calculated for all enzymes except mTET2-CD-Del5 with both 5mC and 5hmC substrates. The rate of the first oxidation step i.e., 5mC to 5hmC and 5hmC to 5fC was calculated using the linear fit (Figure 43) and the values are tabulated (Table 16). Comparison of the initial reaction velocities with 5mC substrates showed that both nTet and mTET3-CD-Del1 oxidized 5mC efficiently with the initial rate of 89%/min and 83%/min respectively and they were almost two times faster than mTET1-CD and mTET1-CD-KRAK. However, the rate of oxidation of 5hmC substrate was generally slow for all enzymes compared to the 5mC substrate (Table 16). The difference was more pronounced for nTet and mTET1-CD, where the initial velocity for 5hmC was 5-fold and 4-fold less than the 5mC substrate respectively, while the difference in the oxidation of 5hmC was only 1.4 fold for

mTET3-CD-Del1 and 2.3 fold for mTET1-CD-KRAK. Out of all TET enzymes, mTET3-CD-Del1 catalysed both 5mC and 5hmC substrates to 5caC rapidly than other enzymes and this behaviour of rapid oxidation by mTET3-CD-Del1 suggests that it may oxidize the substrate CG site in a processive manner i.e., once the enzyme encounters the substrate, it does not dissociate from the substrate until it oxidizes the target 5mC to 5caC. Whereas other enzymes may dissociate from the DNA before oxidizing the next reaction. But further experiments are required to validate this.

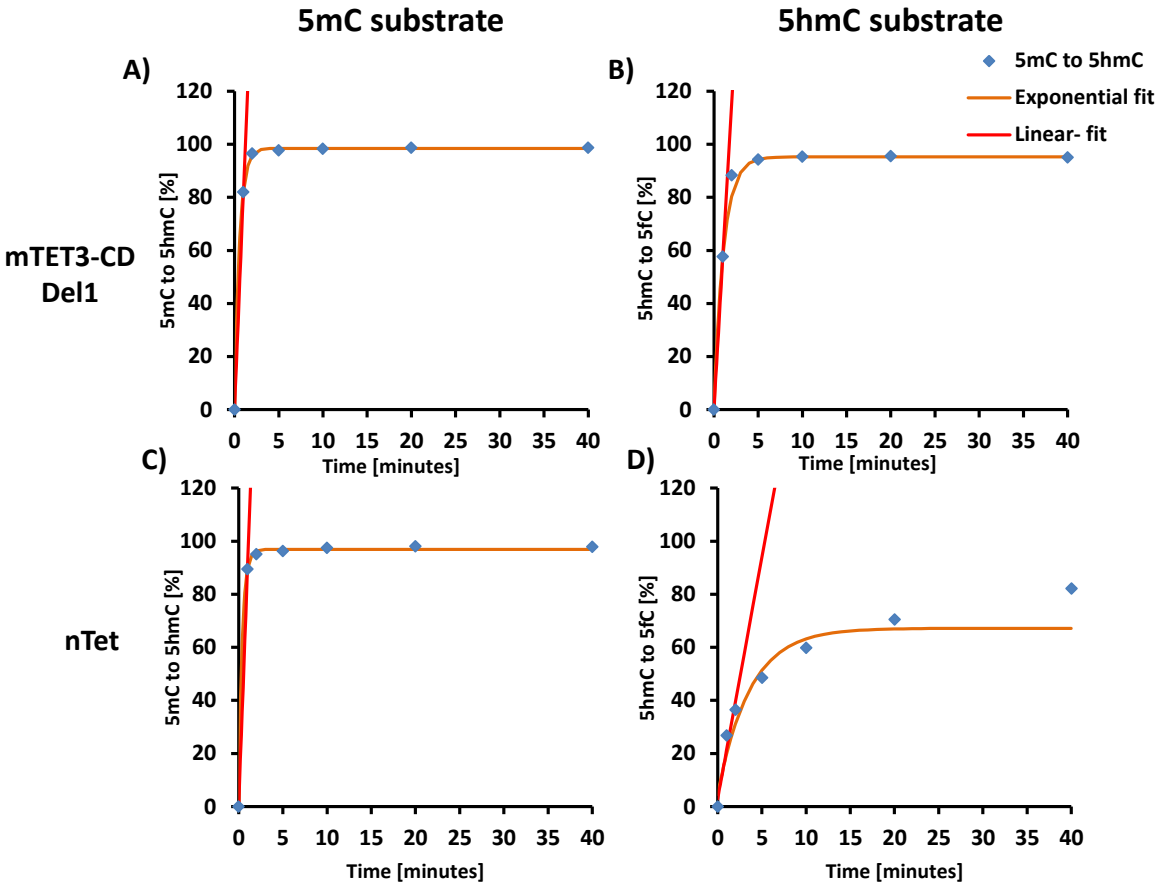


Figure 43: The rate of oxidation of the substrates, 5mC and 5hmC to 5hmC and 5fC respectively over time calculated based on LC-MS data. A) The amount of 5hmC accumulated over time by treating the 5mC substrate with mTET3-CD-Del1. B) The amount of 5fC accumulated over time by treating the 5hmC substrate with mTET3-CD-Del1. C) The amount of 5hmC accumulated over time by treating the 5mC substrate with nTet. D) The amount of 5fC accumulated over time by treating the 5hmC substrate with nTet. The data was fitted exponentially (orange line) using the least square fit. The initial rate of the reaction was calculated using the linear fit (red line).

Table 16: The initial rate of the reaction catalysed by TET enzymes calculated using the linear fit for both 5mC and 5hmC substrates

Enzyme	Rate of the reaction with 5mC substrate (%/min)	Rate of the reaction with 5hmC substrate (%/min)
mTET1-CD	43.3	11.4
mTET1-CD-KRAK	34.2	14.3
mTET3-CD-Del1	81.9	57.6
nTet	89.4	18.1

Next, the mode of oxidation of multiple 5mC substrates on a single DNA strand was analysed. In general, DNA modifying enzymes can catalyse the substrate base on single DNA molecule either in a processive or distributive manner. The processive mode of catalysing the substrate is defined as the ability of the enzyme to catalyse the multiple 5mC/5hmC substrates on a single DNA strand without dissociating from it, before proceeding to the next DNA molecule. The distributivity involves the dissociation of the enzyme from the DNA molecule after each catalytic cycle and reattaches in a stochastic manner to potentially another DNA molecule, thereby catalysing the substrate base (5mC/5hmC) on different strands.

To analyse the processivity between two target sites on a single DNA molecule by TET enzymes, the restriction enzyme cleavage method was used. The rationale for this experiment is that when the enzyme modifies DNA processively, it results in a cooperative oxidation of both substrates (in case of 2 target sites on one DNA) on the same DNA molecule before moving to the next DNA molecule. This result in the DNA molecules showing a binominal distribution of oxidation, thus having either both or none of the sites oxidized. The intermediate state of having one site oxidized and the other not oxidized is very transient. Whereas in the case of distributivity, the enzymes starts with oxidizing in a stochastic manner, therefore the partially oxidized substrates will accumulate before observing complete oxidation of both sites. Following the reaction in a time resolved manner and later treating the oxidized DNA with a modification sensitive restriction enzyme will yield different patterns of digested fragments for processive and distributive, before all the target sites are oxidized. In short, processive oxidation gives rise to a fully digested DNA to undigested, whereas distributive oxidation gives intermediate bands besides other bands.

For this, the DNA substrate of length 127 bp containing two CCGG sites was methylated with HpaII methyltransferase resulting in the 5mC modification of the internal C of C**C**GG sequence.

Importantly, the two CCGG sites are separated by a distance that yields different fragments size, which can be differentiated from one another by electrophoresis, after digesting the DNA using the restriction enzyme MspI. The enzyme MspI can cleave both 5mC and 5hmC within its recognition site (C**C**GG- the modified C is highlighted), but it is blocked by 5fC and 5caC or the glucosylated 5hmC (5ghmC). To differentiate 5hmC from 5mC, 5hmC base was modified further using T4-BGT, where a glucose moiety is added to 5hmC (5ghmC) and cannot be cleaved by MspI. The schematic representation of the restriction sites, length of the DNA fragments after digestion and the exemplary gels showing the processive or distributive mode of oxidizing the substrate is shown in figure 44.

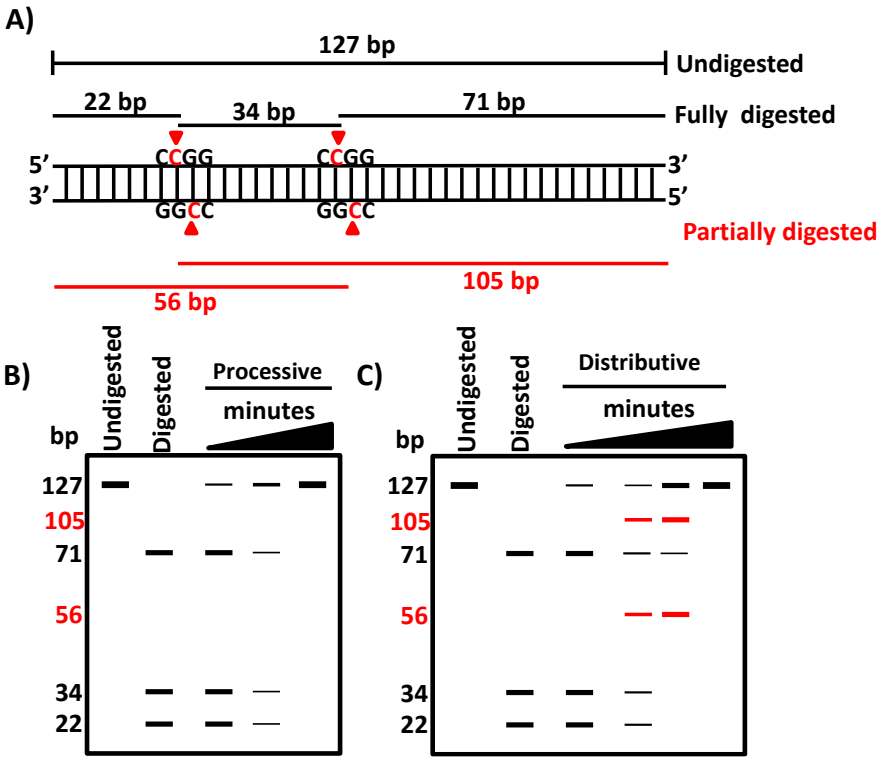


Figure 44: Scheme representing the substrate used and the processive or distributive mode of catalysing the substrate. A) Representation of the DNA containing two MspI restriction site ($C^{5m}CGG$) with the methylated C highlighted in red. The total length of DNA is 127 bp and digestion with MspI will yield three bands at 22 bp, 34 bp and 71 bp. When only one site is oxidized, this will lead to intermediate bands at size 56 bp and 105 bp and when both sites are oxidised, it will lead to complete protection of the CCGG sites against MspI cleavage. B) The theoretical gel showing the scheme of the processive mode of catalysing. When the enzyme catalyse in a processive manner, cleavage with MspI of the DNA treated with the enzyme over time will give rise to the pattern with fully digested to partially digested and undigested to completely protected from digestion C) The theoretical gel showing the distributive mode of catalysing. When the enzyme catalyse in a distributive manner, the DNA treated with TET enzyme will yield intermediate bands (shown in red) along with the digested and undigested, until all the sites are oxidised that give rise to the fully protected DNA of length 127 bp.

The results obtained after treating the substrate DNA catalysed by TETs with or without T4-BGT followed by digestion with MspI is shown in figure 45. The undigested DNA and the methylated DNA treated with MspI were used as controls and they migrated at the expected size. The lane containing the undigested methylated DNA showed one band corresponding to the size of 127 bp, whereas the methylated DNA, digested with MspI yielded 3 bands of around 22, 34 and 71 bp.

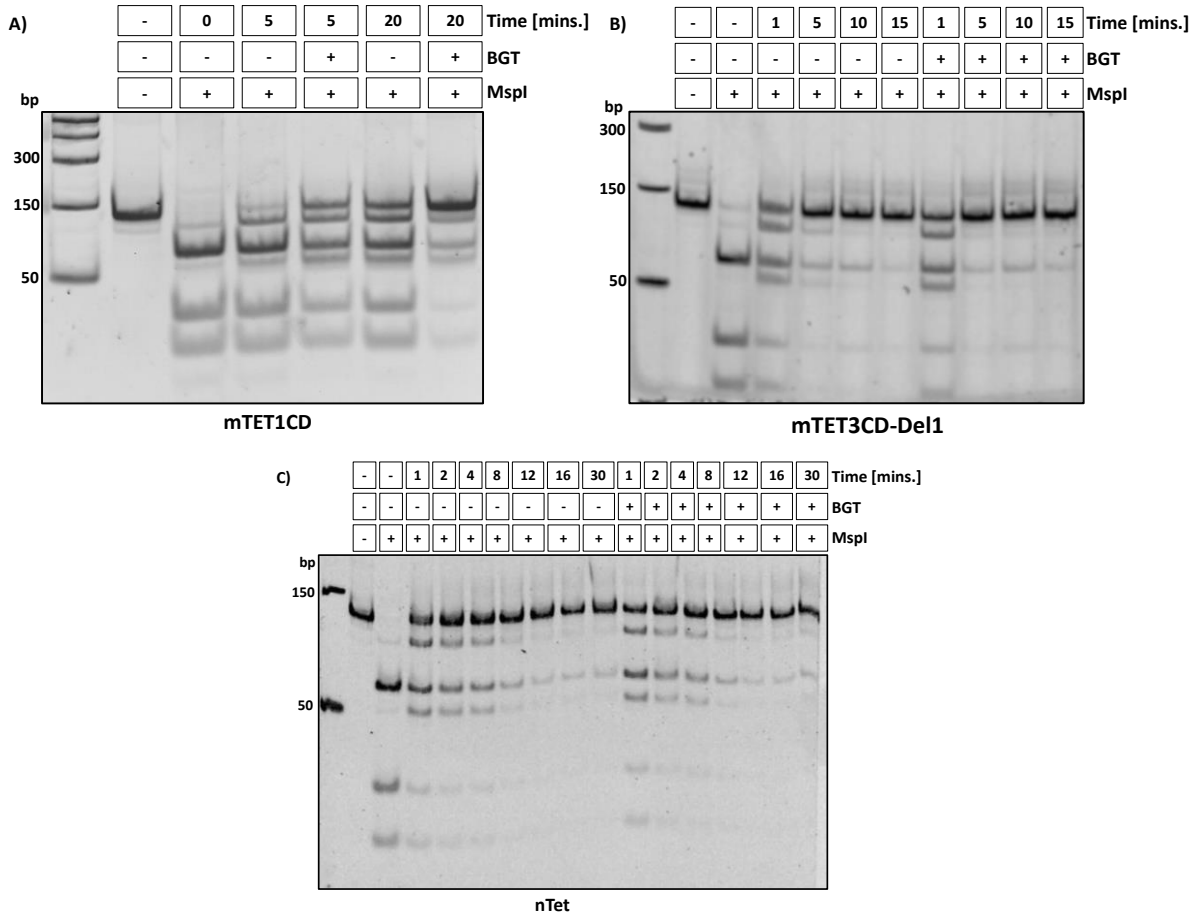


Figure 45: The processivity of TET enzymes analysed using restriction enzymes. The substrate DNA containing two $C^{5m}CGG$ sites was treated with A) mTET1-CD B) mTET3-CD-Del1 and C) nTet and glucosylated by T4-BGT. After glucosylation, the DNA was subjected to digestion using the modification sensitive enzyme MspI and analysed using 10% native acrylamide gel.

Comparison of the TET treated DNA with the control DNA (both digested and undigested) showed that all the three TET (mTET1-CD, mTET3-CD-Del1 and nTet) modified DNA generated intermediate bands that migrated around, 56 bp, and 105 bp in addition to the bands corresponding to the size of the digested and undigested fragments. Importantly, the intensity of the intermediate bands changed over time before progressing to full protection and it looked similar to the figure 44 C-distributive. This suggests that the enzyme processes

one site at a time and does not directly proceed to the second site. Although similar pattern was observed for all three TET enzymes, the level of oxidation differed among three enzymes. mTET1-CD treated DNA did not reach complete protection, while the DNA treated with mTET3-CD-Del1 and nTet reached almost complete protection after 5 minutes and 8 minutes respectively. Moreover, comparison of the DNA treated only with TET (but not T4-BGT) between nTet and mTET3-CD-Del1 showed that mTET3-CD-Del1 oxidized almost all 5mC to 5fC or 5caC faster (within 5 minutes) compared to nTet (8-12 minutes). This can be evidenced from the appearance of band at size corresponding to the full length and intermediate bands at 105 bp and 56 bp, as the intermediate bands are the results of the partial protection against MspI due to the oxidation of 5mC to 5fC or 5caC. These evidences suggest that mTET1-CD, mTET3-CD-Del1 and nTet process multiple substrates on a single DNA molecule in a distributive manner and moreover, in line with LC-MS results (Figure 42), mTET3-CD-Del1 oxidizes 5mC to 5fC/5caC faster than nTet and mTET1-CD.

Alternatively, the pattern of 5hmC oxidation by TET enzymes was analysed from the bisulfite sequencing data. As bisulfite sequencing is a single molecule analysis method, this allowed to study how each of the sequenced strands was processed by TETs. Figure 46 represents the oxidation of the upper strand of CG-rich substrate containing 48 5hmC sites in both CG and non-CG context by mTET3-CD-Del1 at different time points (0, 1, 5 and 30 minutes) of 1000 randomly chosen sequences. Each row represents an individual 5hmC site and each column represents different sequence reads. The oxidation of 5hmC to 5fC or 5caC was represented by yellow and 5hmC by blue. As can be seen from figure 46, the pattern of oxidation were random and importantly not all 5hmC sites in a sequence were oxidized.

Similar analysis was performed on the DNA treated with mTET1-CD and nTet. As an example, the DNA treated with mTET1-CD for 5 minutes and the DNA treated with nTet for 1 minute are shown in figure 47. Similar to mTET3-CD-Del1, the oxidation on 5hmC target sites of the DNA strand for these two enzymes did not have any stretches of oxidized site, as it would be expected for the enzymes catalysing in a processive manner. However, there is a possibility that the enzymes skip unfavourable sites and proceed to the next site. By taking this into consideration, the oxidation pattern of the 5hmC sites, which underwent high level of oxidation by TET enzyme, was compared.

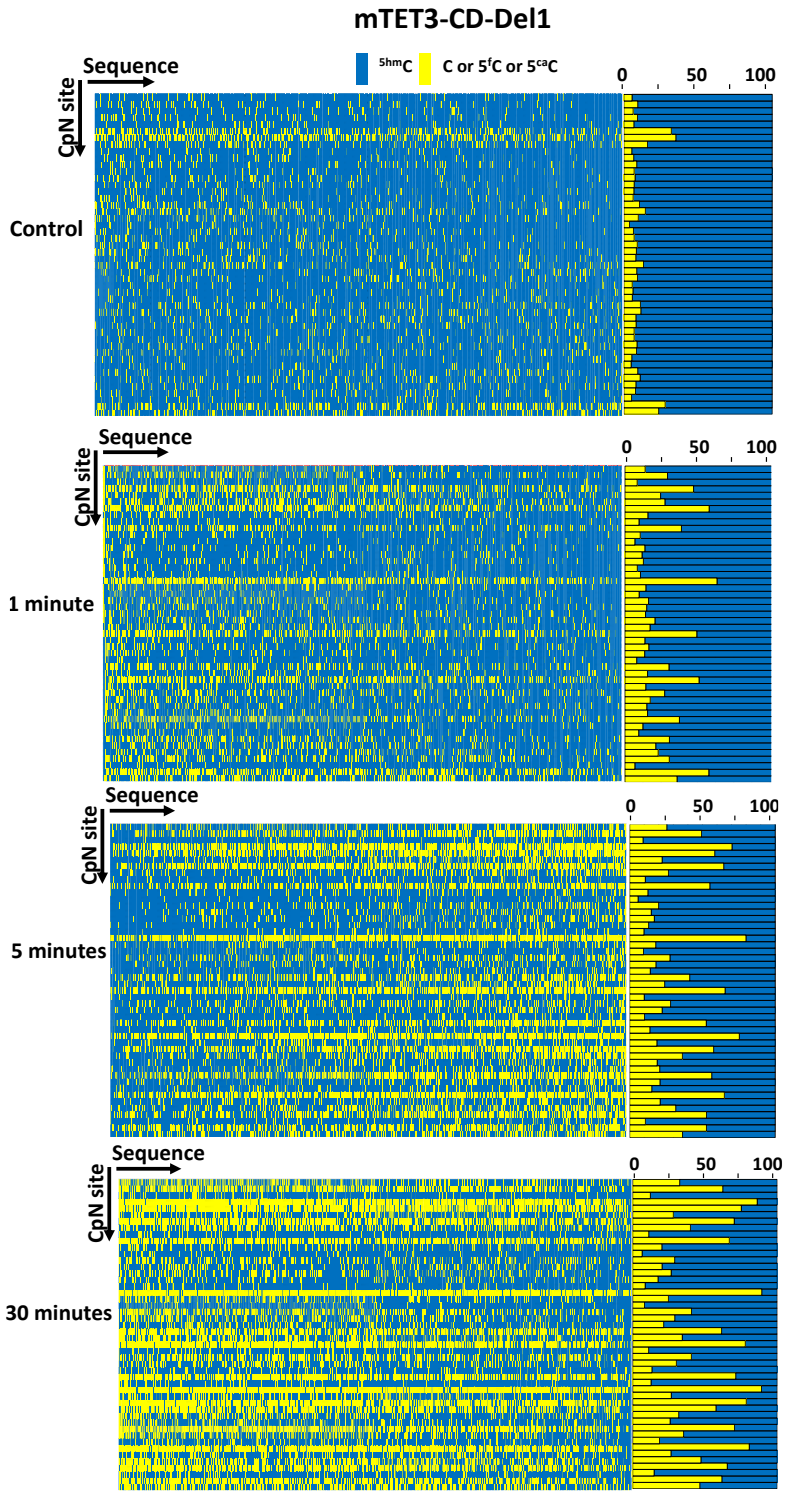


Figure 46: The oxidation pattern of mTET3-CD-Del1 on the upper strand of the CG-rich substrate. The oxidation of 5hmC to 5fC or 5caC by mTET3-CD-Del1 at different time points on all 48 5hmC sites on the upper strand of the CG-rich DNA in both CG and non-CG context is represented here. The site oxidized by the enzyme is represented in yellow and non-oxidized site is represented in blue. Each figure represents the oxidation pattern on 5hmC sites of thousand randomly chosen sequences.

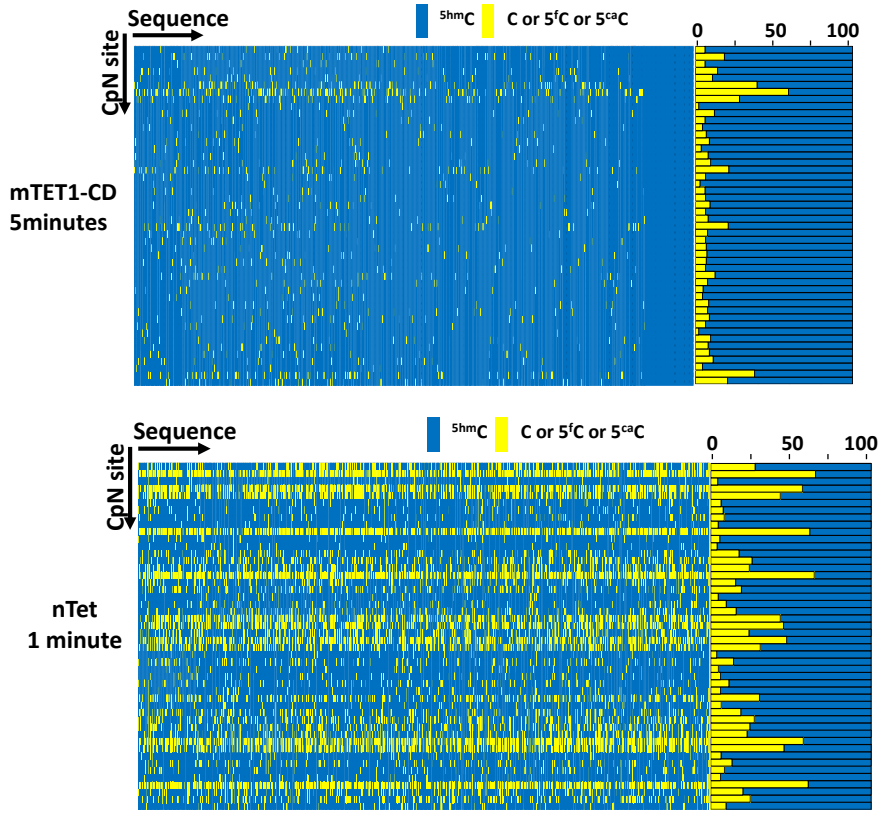


Figure 47: The oxidation pattern of mTET1-CD and nTet on the upper strand of the CG-rich substrate. The oxidation of 5hmC to 5fC or 5caC by mTet1-CD and nTet (after 5 minutes and 1 minute treatment respectively) on all 48 5hmC sites on the upper strand of the CG-rich DNA in both CG and non-CG context is represented here. The site oxidized by the enzyme is represented in yellow and non-oxidized site is represented in blue. Each figure represents the oxidation pattern on 5hmC sites of thousand randomly chosen sequences.

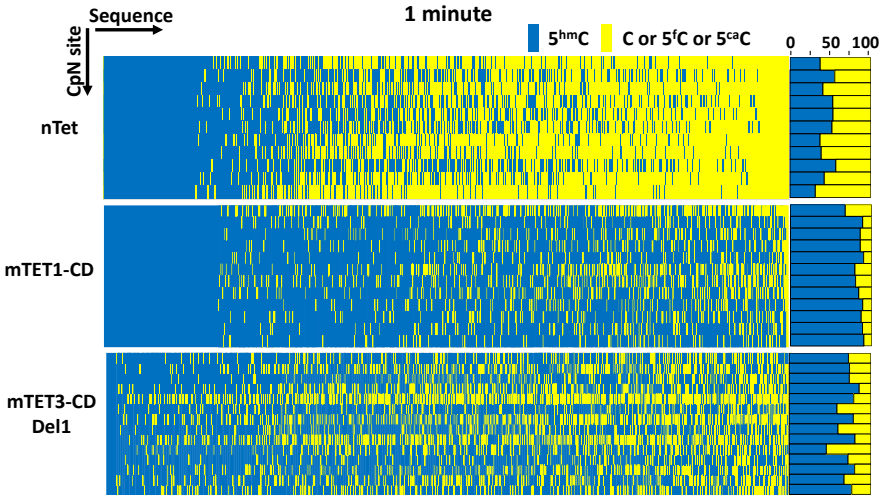


Figure 48: The processivity of TET enzymes analysed from the bisulfite sequencing. Each row represents the analysed C and each column represents different sequence reads. 5hmC sites which were oxidized at high level by TET enzymes were chosen and the pattern of the oxidation was analysed. Yellow represents the oxidation of 5hmC to 5fC and 5caC and blue represents 5hmC.

Comparison of an individual CpN sites, which underwent high level of oxidation, suggests that even on highly oxidized sites, the enzymes seemed to oxidize the 5hmC sites on DNA molecule stochastically and moreover no stretches of oxidation was observed. This observation together with the restriction digestion analysis suggests that TET enzymes may oxidize multiple 5hmC substrates on a DNA molecule in a distributive manner. However, further experiments conducted in the presence of a competitive DNA are required to validate this.

Overall in this study, using detailed biochemical assays followed by NGS and bioinformatics studies that would allow one to follow the oxidation kinetics of TET enzymes at base resolution on a single DNA molecule was elucidated. The experimental results provided a blueprint of the catalytic behaviour of both mammalian TET enzymes and nTet .These results challenge the existing reports, which suggested that TET enzymes were insensitive to the sequences flanking the target CpG site, by demonstrating that TET enzymes showed preference to the flanking sequences and exhibited different level of activity on different substrates. For example, mTET3-CD-Del1 oxidized the sequence in the following order CG**CG**TCC>CGT**CC**AGC>CAG**CC**GT>GGG**CG**TCA. Moreover, it was shown that besides the target substrate in CpG dinucleotide context, mTET3 oxidized the substrates in CC (CGT**CC**AGC) and CA (GC**CA**ATC) context. In addition to the mammalian TET enzymes, the catalytic behaviour of Tet1-like dioxygenase from amoebaflagellate N*aegleria gruberi* was also studied. These results unveiled the previously unknown properties of mammalian TET enzymes, which would help to better understand where and how TET enzymes function in a genomic context. However, further investigation of other members of both mouse and human TET enzymes using single molecule analysis with substrates in other possible sequence contexts (and different modifications) will further reveal other interesting aspects regarding the activity of TET family of enzymes.

5. DISCUSSION

5.1. Establishment of *in vitro* system to investigate the biochemical properties of TET enzymes

In this study, an efficient *in vitro* system was established and employed to investigate the biochemical properties of TET enzymes. For the large scale protein preparation, the *E.coli* expression system was used. The expression of TET enzymes in *E.coli* system was optimized and the recombinant proteins were purified successfully with good quality, even after single purification step, and are catalytically active (Figure 8B). Although some *E.coli* contaminants were also co-purified, these were identified as *E.coli* chaperons and might not affect the activity of the enzyme.

Furthermore, a new method based on ELISA has been established in this project, which is reliable and efficient in detecting the DNA modification (in this case, 5mC oxidized bases). The developed method is quite efficient over other existing methods, as it is quick, follows easy work flow, does not require complex machines (like LC-MS) and is economical. Moreover, this method allows the processing of 96 or 384 samples simultaneously in a high throughput fashion. Unlike the dot blot, this method overcomes the disadvantage of the sample bias, as each sample is treated independent of one another. Furthermore, it does not involve radioactive isotope labelling or further modifications (like glucosylation) and requires low amount of DNA. However, the specificity of each antibody needs to be tested before its application and the conditions need to be optimized. Nevertheless, it can be used to quantify different DNA modifications efficiently and quantitatively.

As a proof of principle, this method was used to study the effect of inhibitors on the activity of mTET1-CD. For this purpose, the known inhibitors of αKG and Fe^{2+} dependent dioxygenases were assayed. The TET kinetics at different concentrations of the inhibitors, in the presence of 0.6 mM αKG was followed and the apparent IC50 was calculated (Table 9). According to the apparent IC50 values, NOG is the potent inhibitor of all the inhibitors tested in this study. When compared to the concentration of αKG (0.6 mM), the app. IC50 value of NOG calculated is in µM range (80 ± 30 µM), indicating that it is able to compete efficiently for the αKG binding site within the TET enzyme. Moreover, the apparent IC50 of NOG obtained for TET enzymes

in this study is in a similar range to the IC50 of NOG reported for other αKG and Fe^{2+}dependent dioxygenases (Xu et al. 2011a), suggesting that a similar mode of binding is adopted.

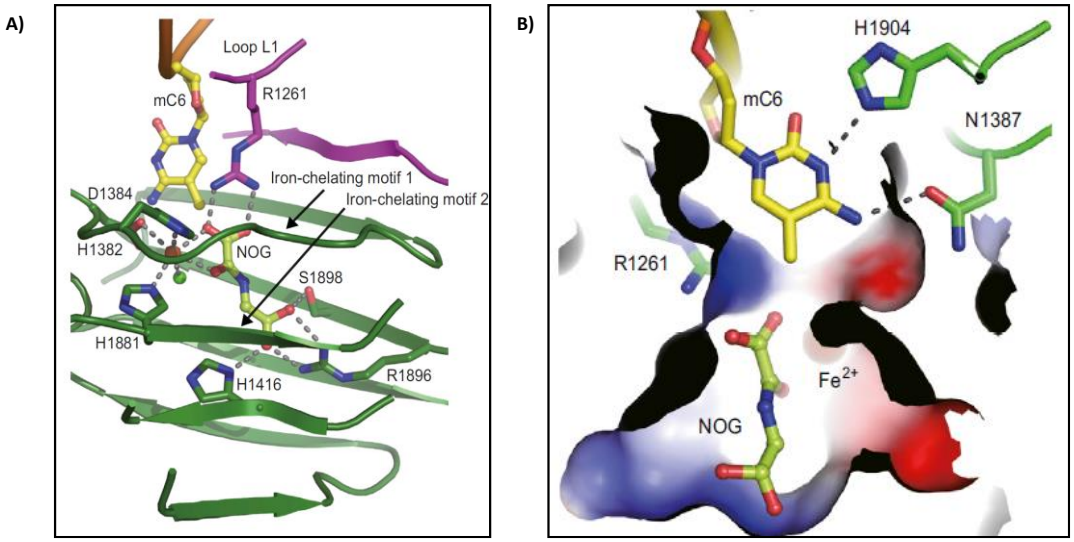


Figure 49: Structure of the DBSH and the catalytic cavity of human TET2 catalytic pocket taken from Hu et al (Hu et al. 2013). A) The DSBH core of human TET2 is represented as ribbons. NOG (an analogue of αKG), the substrate DNA and other important residues are represented as sticks, while the dashed line represents the hydrogen bonds. The Green ball represents the crystallographic water, and the nitrogen and the oxygen atoms are coloured in blue and red respectively. B) The surface representation of the catalytic pocket of TET2. The methyl group is oriented towards the catalytic iron (shaded red ball) and NOG.

Interestingly, at *in vitro* conditions, DMOG did not inhibit the activity of TET enzymes even at a concentration as high as 64 mM. DMOG is the esterified form of NOG, that allows the cell permeability (Schiller et al. 2014). However, the lack of inhibition *in vitro* suggests that the additional methyl group disfavours its interaction within the TET catalytic pocket (as NOG is shown to interact with the TET enzymes through the hydrogen bonding in figure 49A). The close up view of the catalytic pocket (Figure 49B) indicates that the additional methyl group may result in a steric hindrance and without removing the additional methyl groups, it may not fit in the catalytic pocket, thereby resulting in no inhibitory effect. However, the inhibition of TET activity by DMOG in cell culture studies (Amouroux et al. 2016) indicates that DMOG is deesterified inside the cells.

The apparent IC50 of the oncometabolite 2HG, both L-2HG and D-2HG in the presence of 0.6 mM αKG are in the range of mM. This indicates that these are rather weak inhibitors of TET enzymes, when compared to NOG. Moreover, the reported intracellular concentration of αKG is in the range of µM (265 µM in renal cortical cells, as reported by Pritchard 1995). Therefore,

high concentrations of the metabolites has to accumulate in the cell for an efficient inhibition to occur. In agreement with this, studies have shown that 2-hydroxyglutarate indeed could be accumulated up to 10 mM in various gliomas. These gliomas harbour gain of function mutations on *IDH* gene that can synthesise 2HG (Chowdhury et al. 2011; Inoue et al. 2016). The app. IC50 of L-2HG and D-2HG reported in this study are higher than the published value for TET enzymes, which is in the range of 1 mM and 4mM respectively (Koivennen et al., 2012).This can be due to the different reaction conditions under which the activity was tested. Similar to 2HG, succinate and fumarate also showed an inhibitory effect with the app. IC50 value of 2.8 ± 1.8mM and 2.4 ± 0.3mM respectively. Comparison of the app. IC50 of succinate and fumarate shows that these two inhibitors have similar efficiency in inhibiting the TET enzymes, which is in line with the published data, although the IC50 of succinate and fumarate for the TET enzymes is higher than the reported value. For example, the IC50 for succinate was 540 µM and for fumarate, it was 390 µM (Laukka et al. 2016).

This observed difference in IC50 can be explained by different concentration of αKG and different reaction conditions used to test the TET enzymatic activity. For example, 120 µM of αKG was used to test the inhibitory effect of both 2-LHG and 2-DHG (Koivunen et al. 2012) and 240 µM of αKG was used to test the inhibitory effect of succinate and fumarate (Laukka et al. 2016), whereas in the present study, 600 µM of αKG was used . This difference in the concentration of αKG could be the reason, as higher amount of αKG leads to increase in the IC50 value of the inhibitor in a competitive inhibition. Nevertheless, comparison of all the inhibitors shows that NOG is the most potent inhibitor among the tested ones, whereas 2-HG (both L and D form), succinate and fumarate can be regarded as weak inhibitors. Interestingly, comparison of 2-HG with succinate and fumarate shows that both succinate and fumarate show stronger inhibition than 2-HG.

Altogether, the established *in vitro* system (the purified protein and the ELISA-based plate assay) are useful tools that can be efficiently used to investigate the biochemistry of TET mediated hydroxylases. The effect of inhibitors is shown as an example in this study, which can be extended further to screen other small molecules in a high throughput mode, inhibitors of TET and can also be used to study the reaction catalysed by the TETs or their mutants under different conditions. Thus, this established *in vitro* system has opened up the possibility to address many biochemical questions in the field of TET enzymes.

5.2. The effect of divalent metal ions on the activity of TET enzymes

Studies of carcinogenic metals like nickel, cadmium, arsenic and chromium have shown that they can alter the cellular epigenome patterns that can play an important role in transforming the cells to malignancy (Arita and Costa 2009). This has been supported by the experimental evidence, where nickel ions have been shown to alter the epigenetic signals, in particular the histone and the DNA modifications. Moreover, exposure of the cultured cells to nickel leads to an increase in the genome-wide DNA methylation, histone H3K9 dimethylation, ubiquitination (H2A and H2B) and phosphorylation H3S10 (Ke et al. 2006; Sun et al. 2013). Furthermore, nickel exposure decreases the iron level in the cells by 40%. Interestingly, nickel decreases the activity of histone demethylase JMJD1A, which is also a member of αKG and Fe^{2+} dependent dioxygenase, by competing with Fe^{2+} for its binding pocket (Chen et al. 2010a, 2010b).

These observations prompted us to investigate whether the addition of divalent metal ions affects the activity of TET enzymes, owing to their dependence on Fe^{2+} for their catalysis. Obtained *in vitro* results showed that cadmium, zinc, cobalt, manganese and nickel decreased the activity of TET enzymes in a concentration dependent manner, while calcium did not show any inhibitory effect. The calculated IC50 probed at 10μM Fe^{2+} indicates that cadmium and zinc are the most potent inhibitors of the TET enzymes with an app. IC50 value of 0.87 ± 0.14 μM and 2.26 ± 0.3 μM respectively (at 10 μM Fe^{2+}). This low IC50 (lower than the conc. of Fe^{2+} used) suggests that these two metal ions may have higher binding affinity to the catalytic centre than Fe^{2+} and can readily displace the iron at physiological concentrations in the cell or binding of these metal ions can induce conformation changes on the protein and affects the catalysis. Interestingly, cadmium was reported to bind to the cysteine rich region (Vestergaard et al. 2008).This suggests that cadmium can also bind to the cysteine rich region of TET enzymes and influences the overall activity, as the Cys-rich region is involved in stabilizing the overall structure of the TET catalytic domain and is important for the catalysis (Hu et al. 2013).

The calculated app.IC50 of cobalt (15.6 ± 0.27 μM), nickel (19 ± 0.1 μM) and manganese (57 ± 15 μM) suggests that they are less potent TET inhibitors than cadmium and zinc. Interestingly, no inhibition was observed with calcium even at 10 mM, (1000 times higher

than the Fe^{2+} used in the reaction). This difference in the app.IC50 value among the metal ions can be due to the different coordination chemistry of the metal ions on the enzyme and the 'no inhibitory effect' of calcium suggests that it does not interfere with iron binding site.

The decrease in mTET1-CD activity observed with increase in the concentration of metal ions (Cd^{2+}, Zn^{2+}, Co^{2+},Ni^{2+}, Mn^{2+}) suggests a competitive inhibition for metal binding site on TET enzymes. In agreement with the *in vitro* results, exposure of mammalian cells (HEK293 and mES cells) to nickel ions decreased the overall level of 5hmC in both concentration and incubation time dependent manner (Figure 19-22). The gradual loss of 5hmC on the DNA extracted from nickel treated cells, in both HEK293 and mES cells compared to the control cells (~5-fold difference) indicates that the TET enzymes are affected since they are mainly responsible for the production of 5hmC in the cells. Moreover, the initial 5hmC in the genome could be diluted further by the cell division, as the cells treated with nickel ions for 5 days lost more 5hmC than the cells treated for 1 day and the cells grew normally. Furthermore, qPCR results (Figure 23) showed no significant decrease in the level of TETs expression until 100 µM after 5 days of nickel treatment, suggesting that the loss of 5hmC observed at these concentrations might be mainly due to the impairment in the activity of TET enzymes. However, the cells treated with 300 µM of nickel showed around two-fold decrease in all three TET mRNA levels suggesting that besides the catalytic activity, the mRNA level was also affected at 300 µM.

Reduced mRNA expression at 300 µM could be due to up or down regulation of some unknown miRNAs that regulates the TET mRNA expression, as miRNAs (miR-222 and miR-152) have been reported to alter the expression of tumour suppressor genes and DNMT1 in nickel induced carcinogenesis (Sun et al. 2013; Ji et al. 2013). On the other hand, it could be due to pleiotropic effect enforced on other epigenetic pathways, for example on the histone demethylases, which is also a member of the αKG and Fe^{2+} dependent dioxygenases family. Other possibility would be the induction of hypoxia by nickel ions (Costa et al. 2005) that would affect the TET expression/function. Another interesting, yet rather unlikely possibility would be that the TET enzymes binding nickel ions are preferentially cleared from the cells through the degradation machinery, however this still needs to be investigated.

In addition to nickel inhibiting TET enzymes directly, decrease in the cellular Fe^{2+} concentration can also contribute to the reduction in TET activity, as nickel exposure is reported to decrease the level of iron by 40% in lung carcinoma by altering the expression of the iron regulating proteins (Chen and Costa 2006). It would be interesting to investigate how nickel exposure alters the local DNA methylation in the genome and which genes are dysregulated. It would also be interesting to investigate whether the nickel mediated reduction of 5hmC and concomitant increase in the methylation of DNA in mES cells could have an effect on ES cell differentiation.

To summarize, in this study, it was shown that the cells exposed to nickel ions showed decreased level of 5hmC through the inhibition of TET enzymes. As TET enzymes are DNA hydroxylases and require 5mC as the substrate, decrease in the activity of TET enzymes will lead to accumulation of the DNA methylation thus, providing a potential explanation for the nickel mediated DNA methylation gain observed by others (Arita and Costa 2009). Moreover, it suggests that an increase in methylation due to nickel ions, potentially on the tumour suppressor genes can contribute to the cancer development, as TET mediated DNA demethylation is involved in maintaining the steady state level of DNA methylation. Additionally, decrease in 5hmC due to the impairment of TET enzymes has been observed in numerous cancers (Lian et al. 2012; Jeschke et al. 2016).

5.3. Investigation of the role of ascorbic acid (AscA) and retinol on the activity of TET enzymes

Ascorbic acid (AscA), commonly known as vitamin C, is a reducing agent. It has been shown to increase the level of 5hmC by stimulating the activity of TET enzymes in cultured cells and *in vitro* conditions. These reports suggested that AscA was a specific cofactor of TET enzymes, as other reducing agents did not exhibit the same effect (Blaschke et al. 2013; Yin et al. 2013; Minor et al. 2013). In this project, I investigated the mechanism of AscA mediated enhancement of the activity of TET enzymes and found that instead of being a cofactor, AscA increases the activity of TET enzymes through efficient recycling of Fe^{3+} to Fe^{2+}, thereby supporting the TET activity both *in vitro* and in the cells.

The results presented in this study showed that when sufficient Fe^{2+} was present in the reaction, the addition of AscA did not increase the level of 5hmC production (Figure 24). However, when $Fe^{2+}$was limited, for example, when the reaction was performed with Fe^{3+}, the addition of AscA rescued the activity of TET enzymes (Figure 25) suggesting that the enhancement of the TET activity by AscA is simply through the reduction of iron. Moreover, the results showed that the effect was not specific to AscA alone, and other reducing agents could rescue the activity of TET enzymes, however they required more time than AscA (Figure 26). Out of the eight reducing agents tested ß-mercaptoethanol, DTT, N-acetyl L-cysteine, hydroquinone and L-glutamine could rescue the activity of TET enzymes after incubating with Fe^{3+} for 30 minutes, whereas the reactions containing TCEP or KI showed very little effect on TET activity. These differences in the efficiencies may reflect the capacity of different reducing agents to reduce the iron (Petrat et al. 2003). Extended incubation provided an additional time necessary to reduce sufficient amount of $Fe^{3+}$to Fe^{2+} to support the catalysis. This observation supports the fact that the reduction of iron is responsible for the effect rather than the direct binding of AscA to TET enzymes and also provides an explanation for the lack of effect observed by others with the other reducing agents (Blaschke et al. 2013; Yin et al. 2013; Minor et al. 2013). Moreover, unlike C-P4H (Myllyla et al. 1984), no loss in mTET1-CD activity, due to uncoupled oxidation was observed until 30 minutes (Figure 27).This reinforces that AscA is not strictly required for TET enzymes' activity.

In this study, it has been shown that the pH is an important factor in the *in vitro* reaction condition (Figure 28). At pH 8.0 (pH used in Yin et al., 2013, reactions), almost all Fe^{2+} were

oxidised within 5 minutes, while at pH 6.8 (with or without AscA) Fe^{2+} remained constant for at least 30 minutes (Figure 29). This finding provides an explanation for the effect observed by Yin et al (Yin et al. 2013). The authors performed the reaction at a non-optimal condition (at pH 8.0) at which iron underwent rapid oxidation which caused reduction in the activity of TET enzymes. Supplementation of AscA under these conditions partially restored the activity of TET enzymes by recycling oxidized Fe^{2+}, which was interpreted as a stimulation in the TET activity by AscA. Whereas, under the sufficient Fe^{2+} condition (at pH 6.8), the overall activity was high (~15-fold) and AscA was not a necessary component. This indicates that AscA enhances the activity through reduction of iron and helps to reinstate the activity of TET enzymes. Although this is contradictory to the existing report on AscA mediated enhancement of TET activity, these results are in complete agreement with another study, where the authors reported that the addition of redox active quinones increased the level of TET mediated formation of 5hmC by providing more Fe^{2+} through reduction of enzyme free Fe^{3+}(Coulter et al. 2013).

Furthermore, to prove that there was no enzyme bound reduction of iron by AscA, a competition assay was performed to measure the binding affinity of Fe^{3+} (Figure 31). The results showed that $Fe^{3+}$was very weakly bound within the TET catalytic pocket. This indicates that once Fe^{2+} gets oxidized, it can rapidly be released from the enzyme, thus suggesting that the enzyme bound reduction of Fe^{3+} to Fe^{2+} by AscA is less likely. Moreover, closer view of the crystal structure of hTET2-CD (Hu et al. 2013) showed that the active centre was tightly packed with the target base, the co-substrate αKG and the catalytic iron, and there was not enough space for AscA to fit in. This suggests that if AscA is bound inside the enzymes' catalytic pocket in close proximity to the iron ion, it will result in the displacement of either the substrate base or the co-substrate αKG and therefore can cause rather a competitive inhibition than a stimulation in the activity of TET enzymes. Based on this, the reaction conducted with different concentrations of AscA showed that, at working concentration (1-2 mM of AscA), no significant stimulation nor inhibition in the TET activity was observed (Figure 32B). However, at very high AscA concentrations, reduced catalytic activity of TET1 was observed. The high app. IC50 calculated for AscA (8.3 mM) suggests that the inhibition in TET activity can be due to the sequestering of Fe^{2+} by AscA.

In addition to the kinetic studies, Yin and colleagues (Yin et al. 2013) investigated the AscA binding to TET by measuring the AscA induced quenching of intrinsic tryptophan/tyrosine fluorescence. Interestingly, the fluorescence quenching experiment reproduced in this work showed that AscA supplementation to the reactions proportionally decreased the intrinsic fluorescence of mTET1-CD indeed, potentially suggesting a direct interaction (as reported by Yin et al., 2013). However, a similar fluorescence quenching was observed with proteins that were unrelated to AscA function such as EcoDam, CcrM, DNMT2 . This suggests that the observed effects are rather connected to an unspecific interaction of polar AscA with the proteins (Figure 33). Altogether, these results suggest that AscA is not a bound co-factor, and only serves as a reducing agent that helps recycling oxidized iron and at an optimal reaction conditions, Fe^{2+} is not oxidized quickly, and hence, it is dispensable for the catalytic activity.

TET enzymes are sensitive to the concentration of Fe^{2+} in the reaction, which can explain the effects related to AscA supplementation in the cell culture experiments. Additionally the apparent dissociation constant $K_{d\ (app)}$ calculated using Michaelis-Menten reaction showed that the value (0.41 ± 0.05 µM) fell within the range of (0.2-1.5 µM) cellular Fe^{2+} labile iron pool (Figure 34), implying that increase in the concentration of Fe^{2+} within this range can modulate the activity of TET enzymes. Moreover, reductants like thiols and AscA can release ferritin bound Fe^{3+} in the cell (Sirivech et al. 1974), which further strengthens the results, as AscA mediated extraction and the reduction of Fe^{3+} can provide more Fe^{2+} in the cells that are tightly regulated otherwise. This increase in Fe^{2+} can modulate the activity of many Fe^{2+} dependent dioxygenase including TET enzymes.

Besides vitamin C, Vitamin A was also reported to enhance the reprogramming efficiency of iPS cells (Yang et al. 2015a). Interestingly, our collaborators (Dr. Wolf Reik, Dr. Timothy Hore, Dr. Ferdinand von Meyenn, Babraham Institute, Cambridge, UK) observed an additive effect on TET mediated formation of 5hmC in a dose dependent manner after the addition of both AscA and vitamin A (both retinol and retinoic acid), which further increased the reprogramming efficiency of iPS cells. They identified that the increased 5hmC level observed in ES cells treated with vitamin A was due to an increased mRNA expression of TET2/3 proteins through binding of retinoic acid receptor (RAR) to their retinoic acid receptor element (RARE) located on the introns of *TET2* and *TET3* gene. Vitamin A supplementation caused a marked increase in the TET2 and TET3 mRNA levels in the treated cells. Moreover, after reintroducing

the TET enzymes, whose expression was controlled by an independent promoter in TET TKO cells, no TET stimulatory effect was observed. This suggests that indeed vitamin A influences the expression of TET2/3 from their native promoters. Using *in vitro* experiments, it was shown that vitamin A (retinol and retinoic acid) did not directly affect the activity of TET enzymes *in vitro* (Figure 35) even after 30 minutes incubation.

Combining the results obtained in this study with the data obtained from the cellular studies (performed by Dr.T.Hore, University of Otago, New Zealand, and Dr. F. von Meyenn, Babraham Institute, Cambridge, UK) show an interesting observation that both vitamin C (AscA) and vitamin A enhance the activity of TET enzymes, but they regulate the TET enzymes through different pathways. Presence of AscA modulates the catalytic activity of the TET proteins, not as a bound cofactor but as a provider of the reduced iron (Fe^{2+}), and retinol does not influence the protein activity directly, but increases the mRNA expression of TET2 and TET3 by enhancing the binding of RAR to its element (RARE) present on *TET2* and *TET3* gene (Figure 50). Most importantly, the addition of both AscA and retinol show a synergetic effect in a concentration dependent manner and enhance the reprogramming efficiency of pluripotent stem cells.

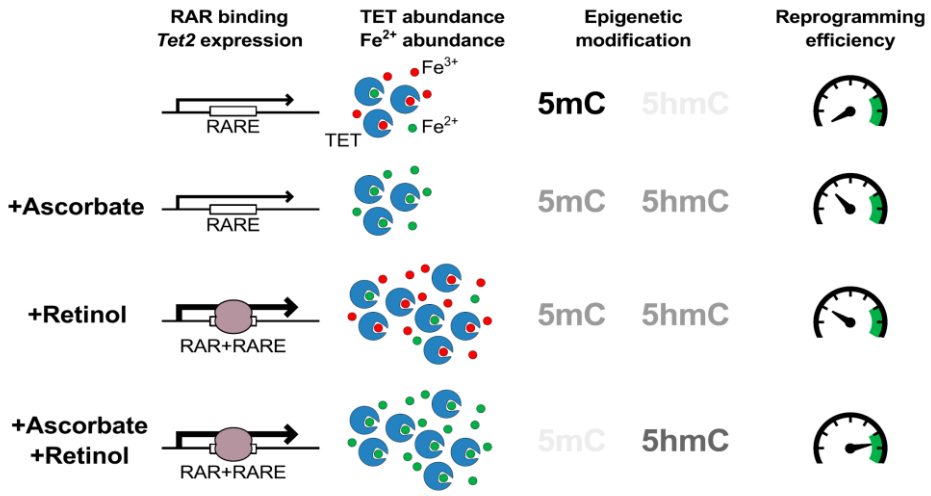


Figure 50: Model showing the regulation of TET enzymes by both vitamin A and vitamin C. Ascorbic acid potentiates the 5hmC production by reducing Fe^{3+} (red circle) to Fe^{2+} (green circle) and provides more cofactor for TET enzymes, while vitamin A increases the mRNA expression of TET2 and TET3 by increasing the binding of RAR (brown shaded part) to the RAR element on their intron (Black big arrow). This model is part of the publication (Hore et al. 2016).

Overall, the study conducted here provides a clear explanation on the mechanism of how AscA potentiates the activity of TET enzymes. It would be worth to investigate whether other iron dependent dioxygenases also behave the same away, as most of the conclusions derived for other dioxygenases on AscA requirement was based on *in vitro* studies. The work presented in this chapter is part of the published report (Hore et al. 2016).

5.4. Investigation of the biochemical and kinetic behaviour of TET enzymes

In this study, the catalytic behaviour of TET enzymes was investigated in detail using single molecule sequencing approach and further confirmed by biochemical experiments on the DNA containing multiple 5hmC (100 in total including forward and reverse strand) present within CG and non-CG contexts. Although PCR was used to incorporate the 5hmC, approx. 90% of the cytosine outside the primer binding region was modified with 5hmC, while 10% still remained unmodified. This suggests that the commercial 5hmdCTPs might not be pure and contain either unmodified dCTP or 5fdCTP/5cadCTP. Interestingly, the incorporation of the unmodified cytosine was not random and they mostly occurred in the region containing stretches of C indicating that the DNA polymerase might show lower efficiency in introducing multiple 5hmCs in a row than unmodified cytosine. The kinetics performed on 5hmC containing DNA using different TET enzymes: mTET1-CD, mTET2-CD-Del5, mTET3-CD-Del1 and nTet showed that all the enzymes could oxidize 5hmC to 5fC and 5caC. The level of oxidation observed with each enzyme was different (Table 15), with mTET2-CD-Del5 showing the weakest activity (5hmC oxidation) of about 1% (of total modified Cs) followed by mTET1-CD, mTET3-CD-Del1 and nTet with the overall oxidation of 5hmC in CpN context of about 10%, 30% and 55% respectively on the CG-rich substrate. The low activity of mTET2-CD-Del5 observed in both bisulfite sequencing (Figure 36) and the kinetics measured using LC-MS (Figure 42) suggests that the quality of the protein from *E.coli* can be improved further. As this enzyme is the shorter version of mTET2-CD, generated by deleting the low complexity region based on sequence alignment, optimizing the borders may help in improving the activity of the enzyme. Interestingly, mTET1-CD, which oxidized the CG-rich substrate only ~ 10% (measured using bisulfite sequencing) showed higher activity on the substrate with single CpG site (measured using LC-MS kinetics), with almost all the 5mC substrates and 70% of the 5hmC substrate getting oxidized. This can be explained by the fact that the substrate with a single CpG may have a preferred sequence context, which results in the increased oxidation efficiency observed. Whereas the CG-rich DNA used for bisulfite sequencing contains multiple 5hmC substrates in CpN context (where N=A, T, G or C), where not all the substrates have been equally preferred, potentially leading to the overall lower global oxidation level. This indicates that it might not be due to the poor catalytic efficiency of the enzyme per se, but rather due to the sequence specificity.

Out of all three mammalian TET enzymes, mTET3-CD-Del1 showed the highest activity on CG-rich substrate and showed a strong sequence selectivity. Investigation of the sequence specificity revealed that TET enzymes were able to oxidize 5hmC/5mC in almost all CpN dinucleotide contexts. Among all, CG dinucleotide was highly preferred and interestingly some CC sites were oxidized with almost equal efficiency as CG sites. Weaker oxidation efficiency was observed for substrates in CA context. Substrates in CT context were the least preferred, however still approx. 3.5% of all CTs were converted. Furthermore, this indicates that TET enzymes may not be very stringent in oxidizing the CG dinucleotide alone and may have the capacity to oxidize the sequence in non-CG context, thus providing the possibility of oxidizing the non-CG methylation, which is reported to present in oocytes, zygote and the brain (Ficz et al. 2011; Guo et al. 2013; Shirane et al. 2013). Moreover, the detailed analysis of the sequence preference revealed a striking difference in the conversion rate of CG embedded in a distinct flanking sequence context. Some CG sites were highly oxidized (for example CG**CG**TCC), whereas some CGs were least oxidized (GAG**CG**CGC) (Figure 39). This suggests that besides recognizing the central dinucleotide, the TET enzymes might be sensitive to the sequence flanks in which the substrate 5mC is embedded. This observation contradicts the existing reports, which suggest that hTET2 has no or weak flanking sequence preference (Hu et al. 2013, 2015). Their conclusion was drawn mainly based on the structural data, where the hTET2 catalytic domain does not show any direct contact with the bases except the target CpG dinucleotide and most of the contacts between the DNA and the enzyme is mediated through the hydrogen bonds on the phosphate backbone or through the hydrophobic interaction (Hu et al. 2013). This observation, together with the biochemical evidence showing an equal activity of hTET2 on both CG-rich and AT-rich substrate in a CG context leads to the conclusion that the TET enzymes are not sensitive to the flanking sequence (Hu et al. 2015). It should be noted that these were tested only on few substrates and hence there is a possibility that the other sequence composition can influence the outcome. Moreover the sequence of both the substrates tested (CCGG and ACGT) in a CpG context are the sequences well preferred by TET enzymes (as it was observed from bisulfite sequencing). Another possible reason for the observed discrepancy between two studies might be due to the fact that, in this study, the reactions were performed in competition to each other, meaning that the enzyme is free to choose between 5hmC sites within different sequence contexts. In addition, the enzyme concentration was ~100-fold lower than the

available 5hmC in the DNA, whereas Hu et al (Hu et al. 2013, 2015) reactions contained only one 5hmC in the substrate and approx. 1:1 molar ratio of substrate to enzyme was used and in this condition, the relative differences between the substrates specificity might be minimized by the experimental system used.

Detailed analysis of the sequencing results revealed an interesting flanking sequence preference for mTET3-CD, where the enzyme preferred A or T flanking the CpG dinucleotide on both sides, especially having the base A at -1 position showed higher activity (Figure 40A). This is intriguing, as in the crystal structure of both human TET2 and nTet, DNA distortion of about 40% and 65% respectively was observed after the enzyme binding (Hu et al. 2013; Hashimoto et al. 2014). Based on this observation, it is conceivable that DNA having A or T flanks can be easily bent and acts as an indirect readout (Jurkowski et al. 2007) thus, adopting the active conformation, which would then be efficiently catalysed by TET enzymes. This also suggests that in the genomic context, TET enzymes could more efficiently oxidize 5mC on the sequence which are already bent, for example near the CTCF binding sites.

Moreover, the flanking sequence of mTET3-CD-Del1 obtained from this study matches quite well with the *in vivo* occurrence of 5hmC (ACGT) (Sérandour et al. 2016). In addition, the sequence preferred by mTET3-CD was reported to be the preferred sequence of DNMT3A (ACGT) (Handa and Jeltsch 2005). This agrees with the finding that TET2 and DNMT3A act as the direct competitor (Zhang et al. 2016) and TET enzymes can oxidize the methyl marks deposited by DNMT3a on TSSs, as TET enzymes are localized near TSSs (Xu et al. 2011b). Furthermore, the observed sequence motif matches with the binding motif reported for TET3-FL. This suggests that besides the N-terminal CXXC domain, which is believed to target TET1 and TET3 to specific sequences in the genome (Xu et al. 2012; Jin et al. 2016), the catalytic domain by itself shows a strong substrate selectivity that contributes to the TETs function in the cell. Interestingly, the sequence preference of mTET3-CD-Del1 also matches with the binding motif of the transcriptional factors such as Earyl B-cell factor (EB) and transcription factors binding E-box sequences (Jin et al. 2016; Sérandour et al. 2016). This suggests that TET mediated demethylation may help TFs to bind to the particular sequence, as many TFs are sensitive to the methylation marks in their recognition sequence (for example c-myc (CA**CG**TC) and TFEB (TCA**CG**TG)). In line with this, the recent publication demonstrated that MAX- the partner of TF Myc, recognized 5caC in CG of the E-box sequence (Wang et al. 2016),

thus suggesting that the TET mediated demethylation could modulate the activity of transcriptional network.

The flanking sequence preference obtained for nTet (Figure 40B) showed that like mTET3, nTet was also sensitive to the flanks, although the difference between the most oxidized 5hmC substrate and the least oxidized substrate was not high. The motif obtained showed a preference for 2 nucleotides downstream and 3 nucleotides upstream of the target site (Figure 40). This contradicts the report by Pais et al, where the authors claimed that only the base upstream of target C (-1 position) had pronounced effect and nucleotide at +1 position had no drastic effect (Pais et al. 2015). This discrepancy might be due to the different experimental set up and the authors checked the effect on -1 and +1 positions ((one nucleotide up and downstream of target C, total 8 substrates) and the other positions were ignored, while in this study, the substrate DNA containing 100 5hmCs in a different flanking sequence contexts was analysed.

Moreover, comparison of the flanking sequence preference of mTET3-CD-Del1 with nTet showed a number of interesting differences. Although both enzymes exhibited similar preference at -1 position, there was a difference at +2 position. At this position, nTet accepted base C or T or A almost equally, while mTET3-CD preferred T or A more. This additional acceptance of C at +2 position might confer flexibility in substrate selection for nTet, which could potentially explain the overall higher activity displayed by nTet on the CG-rich substrate in which many of the 5hmCs were oxidized efficiently compared to mTET3-CD-Del1. Moreover, nTet showed different substrate specificity towards non-CG sites compared to mTET3-CD-Del1. For example, the 5hmC substrate in CC context (GCG**CC**GAT) was the least preferred by nTet, whereas the 5hmC substrate in CT context (GCG**CT**CTG) was least preferred by mTET3-CD-Del1 (Figure 37). Moreover, the calculation of the initial velocity of oxidation showed that nTet showed only 26-fold difference between the least and the most preferred substrate (CG and non-CG contexts), whereas mTET3-CD-Del1 showed around 240 fold difference. These differences observed between mTET3-CD-Del1 and nTet suggest that the intrinsic property of TET enzymes might determine the sequence specificity and mouse TET3-CD-Del1 is more stringent in substrate selectivity and behave differently than nTet.

Moreover, the kinetics experiments performed on the highly preferred substrate containing a single CpG site (reaction products measured by LC-MS) showed that all TET enzymes except mTET2-CD-Del5 were able to oxidize both 5mC and 5hmC-containing substrates (Figure 42). Interestingly, the accumulation of higher oxidation product catalysed by mTET1-CD-KRAK (the shorter version of mTET1-CD) was more than mTET1-CD, suggesting that the quality of the full catalytic domain might decline over time, while the shorter version having the unstructured region removed could efficiently oxidize the substrate further. Moreover, mTET3-CD-Del1 performed all three oxidation reactions quickly, which is evident from the accumulation of the final oxidized base 5caC to almost 100% after 20 minutes, without any accumulation of 5fC visible during the experimental time (Figure 42). Whereas nTet, mTET1-CD and mTET1CD-KRAK oxidized the first reaction faster compared to the other higher oxidation reactions, and both 5fC and 5caC accumulated at a slower rate. Moreover, 5fC was not completely oxidized to 5caC. These results are in contrast to the recent report (Hu et al. 2015) and show a fundamental functional difference between TET1/2 and TET3 enzymes. This intriguing observation suggests that mTET3-CD-Del1 might processively oxidize the single target site. This would provide a molecular explanation for the fast occurrence of 5caC in biological context, as TET3 is involved in the demethylation of early embryogenesis (Gu et al. 2011). So the conversion of 5mC to 5caC in a single encounter would speed up the demethylation process. On the other hand, the stepwise oxidation of the single target site in a distributive manner might be more practical from the enzyme point of view. The exchange of succinate to αKG after each oxidation would be easier, when the target base dissociates from the DNA, as αKG is buried deeply in the catalytic pocket of the enzymes (Hu et al. 2013). However, the accumulation of 5hmC and 5fC as stable marks (Bachman et al. 2014, 2015) in the genome of mES, where TET1 and TET2 are highly expressed suggests that, different members of the family may have different processivity (on single site) and thereby perform different biological functions. For example TET3 might oxidize 5hmC to 5caC, whereas TET1 and TET2 might oxidize 5hmC to 5fC and 5caC in a distributive manner. Although the kinetics observed with the two mammalian TETs partially supports this hypothesis, further experiments are needed to validate this.

Interestingly, comparison of the kinetics performed with both 5mC and 5hmC substrates showed that 5mC is a better substrate for TET enzymes than 5hmC. Although the overall

accumulation of oxidized base was lower with the 5hmC substrate than the 5mC, the difference was not as drastic as it was reported (Hu et al. 2015). In the published report, only 25% of the 5hmC substrate was converted, while in our experiments around 70-80% of the 5hmC substrates were oxidized (whereas almost all 5mC substrates were oxidized). Moreover, the initial velocity of the reaction calculated with 5mC and 5hmC substrates for different TET enzymes showed that the maximum difference observed was around 4-fold and 5-fold between 5mC and 5hmC for mTET1-CD and nTet respectively, while mTET1-CD-KRAK and mTET3-CD-Del1 showed around 2.3 and 1.4 fold differences respectively. This difference might be due to the non-optimal reaction conditions (pH 8.0) used in the reported experiment. Since, the reaction performed with 5hmC as the substrate is slower than 5mC (for all enzymes tested except mTET3-CD-Del1), the reaction condition may have a greater influence, as pH 8.0 accelerates the iron oxidation and drastically reduces the activity of TET enzymes, which may have resulted in a drastic decrease in the TET activity on 5hmC compared to 5mC.

Furthermore, the analysis of processing multiple substrate bases on a single DNA molecule by TET enzymes using the restriction enzyme cleavage experiments suggest that TET enzymes (mTET1-CD, mTET3-CD-Del1 and nTet) might oxidize the distinct sites in a distributive manner. This observation was further corroborated by the bisulfite sequencing results, where the multiple C analysed showed stochastic distribution of oxidation on single molecules. Such oxidation pattern can be explained by the dissociation of the enzyme from the DNA before moving to the next modification site. The distributive behaviour of TET enzymes between substrates could support its cellular function, as sliding of the enzyme on DNA in order to modify numerous sites cooperatively would be hampered by numerous other proteins densely occupying the genomic DNA. However, further experiments performed with a competitive substrate and different molar concentration of enzyme and substrate could confirm the distributivity of TET enzymes between substrates. While these experiments were carried out, another group reported that TET enzymes performed the catalysis on different substrate in a distributive manner (Tamanaha et al. 2016), which further strengthens our observation on distributive mode of oxidizing the substrates.

This is an ongoing project and the data presented here represents the initial observation of the behaviour of TET enzymes on CG-rich substrate. As follow up experiments, the kinetics of

multiple DNA substrates carrying both 5mC and 5hmC modifications with many possible flanks, treated with all the members of mouse TET enzymes and nTet is under progress. This will provide further insight into the behaviour of different members in the family and how different steps of oxidation occur. Nevertheless, the data presented in this study provided many useful insights about the catalytic properties of TET enzymes in both mammalian TETs and *Naegleria*Tet. To the best of our knowledge, this is the first *in vitro* experiment, where the kinetics of TET enzymes were analysed at single base resolution on the synthetic substrates and the DNA oxidation is performed in competition. The detailed analysis revealed many interesting features of TET enzymes and how it could catalyse the substrate. The shorter, but still the catalytically active version of mTET3-CD was studied intensively together with nTet. Similar studies on other two members in the family would shed light on how different family members of the TET enzymes regulate the epigenome. It would also be interesting to perform similar studies on RNA, as TET enzymes have been shown to oxidize 5mC on RNA (Fu et al. 2014; Basanta-Sanchez et al. 2017).

6. CONCLUSION

The results presented in this thesis elucidated various biochemical properties of TET enzymes. First of all, an efficient *in vitro* assay was established and employed to investigate the biochemistry of TET enzymes. Examining the inhibitory effects of divalent metal ions suggested that the exposure of mammalian cells to nickel ions resulted in decreased 5hmC, indicating that the nickel mediated 5hmC reduction might alter the epigenome of the cells and eventually could lead to a diseased state. Furthermore, this study provided the mechanistic detail of AscA mediated enhancement of TET activity using detailed biochemical assays. Finally, using single molecule approach, it was demonstrated that mammalian TET enzymes have sequence preferences beyond CpN dinucleotide. On the whole, using detailed biochemical experiments, this study uncovered previously unknown properties of how the DNA hydroxylases oxidize the DNA, and how they are regulated by extrinsic factor (AscA). The knowledge gained from these studies have shed light on the enzymatic behaviour of TET enzymes and have provided a new perspective on how the TET enzymes function *in vivo*. These new findings would be helpful in understanding the molecular mechanism of TET mediated demethylation in different cellular contexts, which would provide a deeper understanding of how the DNA modifications modulate the epigenome in both normal and diseased state. Thus, this study has provided a new direction in the TET mediated DNA demethylation and further investigation would unveil more interesting aspects, which would lead to a more complete understanding of the methylation dynamics and how it can be tuned to overcome the diseased state, as observed for example, in cancer.

7. REFERENCES

Abdel-Wahab O, Gao J, Adli M, Dey a, Trimarchi T, Chung YR, Kuscu C, Hricik T, Ndiaye-Lobry D, Lafave LM, et al. 2013. Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. *J Exp Med* **210**: 2641–2659.

Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, Malinge S, Yao J, Kilpivaara O, Bhat R, et al. 2009. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**: 144–7.

Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PWS, D'Souza Z, Nakayama M, Matsuda M, Turp A, Ndjetehe E, et al. 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol* **18**: 225–233.

Arab K, Park YJ, Lindroth AM, Schäfer A, Oakes C, Weichenhan D, Lukanova A, Lundin E, Risch A, Meister M, et al. 2014. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol Cell* **55**: 604–14.

Arita A, Costa M. 2009. Epigenetics in metal carcinogenesis: nickel, arsenic, chromium and cadmium. *Metallomics* **1**: 222–8.

Bachman M, Uribe-Lewis S, Yang X, Burgess HE, Iurlaro M, Reik W, Murrell A, Balasubramanian S. 2015. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat Chem Biol* **11**: 3–6.

Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 2014. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* **6**: 1049–1055.

Banerjee D, Mandal SM, Das A, Hegde ML, Das S, Bhakat KK, Boldogh I, Sarkar PS, Mitra S, Hazra TK. 2011. Preferential repair of oxidized base damage in the transcribed genes of mammalian cells. *J Biol Chem* **286**: 6006–6016.

Basanta-Sanchez M, Wang R, Liu Z, Ye X, Li M, Shi X, Agris PF, Zhou Y, Huang Y, Sheng J. 2017. TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. *ChemBioChem* **18**: 72–76.

Bauer C, Göbel K, Nagaraj N, Colantuoni C, Wang M, Müller U, Kremmer E, Rottach A, Leonhardt H. 2015. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J Biol Chem* **290**: 4801–12.

Baylin SB, Jones PA. 2011. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**: 726–34.

Becker H, Maharry K, Radmacher MD, Mrózek K, Metzeler KH, Whitman SP, Schwind S, Kohlschmidt J, Wu YZ, Powell BL, et al. 2011. Clinical outcome and gene- and microRNA-expression profiling according to the wilms tumor 1 (WT1) single nucleotide polymorphism rs16754 in adult de novo cytogenetically normal acute myeloid leukemia: A cancer and leukemia group B study. *Haematologica* **96**: 1488–1495.

Bian C, Yu X. 2014. PGC7 suppresses TET3 for protecting DNA methylation. *Nucleic Acids Res* **42**: 2893–2905.

Blackledge NP, Thomson JP, Skene PJ. 2013. CpG island chromatin is shaped by recruitment of ZF-CxxC proteins. *Cold Spring Harb Perspect Biol* **5**: a018648.

Blaschke K, Ebata KT, Karimi MM, Zepeda-Martínez J a, Goyal P, Mahapatra S, Tam A, Laird DJ, Hirst M, Rao A, et al. 2013. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**: 222–6.

Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, Balasubramanian S. 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**: 934–7.

Bruniquel D, Schwartz RH. 2003. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat Immunol* **4**: 235–240.

Chen CC, Wang KY, Shen CKJ. 2012. The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem* **287**: 33116–33121.

Chen H, Costa M. 2006. Effect of soluble nickel on cellular energy metabolism in A549 cells. *Exp Biol Med (Maywood)* **231**: 1474–1480.

Chen H, Davidson T, Singleton S, Garrick MD, Costa M. 2005. Nickel decreases cellular iron level and converts cytosolic aconitase to iron-regulatory protein 1 in A549 cells. *Toxicol Appl Pharmacol* **206**: 275–87.

Chen H, Giri NC, Zhang R, Yamane K, Zhang Y, Maroney M, Costa M. 2010a. Nickel ions inhibit histone demethylase JMJD1A and DNA repair enzyme ABH2 by replacing the ferrous iron in the catalytic centers. *J Biol Chem* **285**: 7374–83.

Chen H, Ke Q, Kluz T, Yan Y, Costa M. 2006. Nickel ions increase histone H3 lysine 9 dimethylation and induce transgene silencing. *Mol Cell Biol* **26**: 3728–37.

Chen H, Kluz T, Zhang R, Costa M. 2010b. Hypoxia and nickel inhibit histone demethylase JMJD1A and repress Spry2 expression in human bronchial epithelial BEAS-2B cells. *Carcinogenesis* **31**: 2136–2144.

Chen Q, Chen Y, Bian C, Fujiki R, Yu X. 2013. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**: 561–4.

Cheng J, Guo S, Chen S, Mastriano SJ, Liu C, D'Alessio AC, Hysolli E, Guo Y, Yao H, Megyola CM, et al. 2013. An extensive network of TET2-targeting MicroRNAs regulates malignant hematopoiesis. *Cell Rep* **5**: 471–81.

Chowdhury R, Yeoh KK, Tian Y-M, Hillringhaus L, Bagg EA, Rose NNR, Leung IKH, Li XS, Woon ECY, Yang M, et al. 2011. The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep* **12**: 463–9.

Chuang KH, Whitney-Miller CL, Chu CY, Zhou Z, Dokus MK, Schmit S, Barry CT. 2015. MicroRNA-494 is a master epigenetic regulator of multiple invasion-suppressor microRNAs by targeting ten eleven translocation 1 in invasive human hepatocellular carcinoma tumors. *Hepatology* **62**: 466–480.

Chung T-L, Brena RM, Kolle G, Grimmond SM, Berman BP, Laird PW, Pera MF, Wolvetang EJ. 2010. Vitamin C promotes widespread yet specific DNA demethylation of the epigenome in human embryonic stem cells. *Stem Cells* **28**: 1848–1855.

Ciccarone F, Valentini E, Bacalini MG, Zampieri M, Calabrese R, Guastafierro T, Mariano G, Reale A. 2014. Poly (ADP-ribosyl) ation is involved in the epigenetic control of TET1 gene transcription. *Oncotarget* **5**.

Cimmino L, Dawlaty MM, Ndiaye-Lobry D, Yap YS, Bakogianni S, Yu Y, Bhattacharyya S, Shaknovich R, Geng H, Lobry C, et al. 2015. TET1 is a tumor suppressor of hematopoietic malignancy. *Nat Immunol* **16**: 653–62.

Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, Sweeney K, Sabatini R. 2009. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res* **37**: 1452–1462.

Cortázar D, Kunz C, Selfridge J, Lettieri T, Saito Y, MacDougall E, Wirz A, Schuermann D, Jacobs AL, Siegrist F, et al. 2011. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**: 419–423.

Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, et al. 2011. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**: 67–79.

Costa M, Davidson TL, Chen H, Ke Q, Zhang P, Yan Y, Huang C, Kluz T. 2005. Nickel carcinogenesis: Epigenetics and hypoxia signaling. *Mutat Res - Fundam Mol Mech Mutagen* **592**: 79–88.

Costa Y, Ding J, Theunissen TW, Faiola F, Hore T a, Shliaha P V, Fidalgo M, Saunders A, Lawrence M, Dietmann S, et al. 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**: 370–4.

Coulter JB, O'Driscoll CM, Bressler JP. 2013. Hydroquinone increases 5-hydroxymethylcytosine formation through ten eleven translocation 1 (TET1) 5-methylcytosine dioxygenase. *J Biol Chem* **288**: 28792–28800.

Crawford DJ, Liu MY, Nabel CS, Cao X-JJ, Garcia BA, Kohli RM. 2016. Tet2 catalyzes stepwise 5-methylcytosine oxidation by an iterative and de novo mechanism. *J Am Chem Soc* **138**: 730–733.

Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, Gao Q, Powell BE, Cheng AW, Faull KF, Lyko F, et al. 2014. Loss of tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell* **29**: 102–111.

Dawlaty MM, Breiling A, Le T, Raddatz G, Barrasa MI, Cheng AW, Gao Q, Powell BE, Li Z, Xu M, et al. 2013. Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Dev Cell* **24**: 310–323.

Dawlaty MM, Ganz K, Powell BE, Hu Y-CC, Markoulaki S, Cheng AW, Gao Q, Kim J, Choi S-WW, Page DC, et al. 2011. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**: 166–175.

de Jong L, Albracht SP, Kemp a. 1982. Prolyl 4-hydroxylase activity in relation to the oxidation state of enzyme-bound iron. The role of ascorbate in peptidyl proline hydroxylation. *Biochim Biophys Acta* **704**: 326–332.

Delaval K, Feil R. 2004. Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev* **14**: 188–195.

Detich N, Ramchandani S, Szyf M. 2001. A conserved 3'-untranslated element mediates growth regulation of DNA methyltransferase 1 and inhibits its transforming activity. *J Biol Chem* **276**: 24881–90.

Dickson KM, Gustafson CB, Young JI, Züchner S, Wang G. 2013. Ascorbate-induced generation of 5-hydroxymethylcytosine is unaffected by varying levels of iron and 2-oxoglutarate. *Biochem*

Biophys Res Commun **439**: 522–527.

Doege CA, Inoue K, Yamashita T, Rhee DB, Travis S, Fujita R, Guarnieri P, Bhagat G, Vanti WB, Shih A, et al. 2012. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**: 652–5.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* **10**: 2709–21.

Epsztejn S, Glickstein H, Picard V, Slotki IN, Breuer W, Beaumont C, Cabantchik ZI. 1999. H-ferritin subunit overexpression in erythroid cells reduces the oxidative stress response and induces multidrug resistance properties. *Blood* **94**: 3593–3603.

Feng J, Shao N, Szulwach KE, Vialou V, Huynh J, Zhong C, Le T, Ferguson D, Cahill ME, Li Y, et al. 2015. Role of Tet1 and 5-hydroxymethylcytosine in cocaine action. *Nat Neurosci* **18**: 536–544.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore T a, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**: 398–402.

Ficz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul Y-L, Walter J, et al. 2013. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**: 351–9.

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, et al. 2010. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**: 553–67.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* **7**: 461–465.

Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, Antes I, Leonhardt H. 2011. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* **6**: 1–8.

Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, Cai Q, Ji D, Jin S-G, Niedernhofer LJ, et al. 2014. Tet-Mediated Formation of 5-Hydroxymethylcytosine in RNA. *J Am Chem Soc* **136**: 11582–11585.

Fu X, Jin L, Wang X, Luo A, Hu J, Zheng X, Tsark WM, Riggs AD, Ku HT, Huang W. 2013. MicroRNA-26a targets ten eleven translocation enzymes and is regulated during pancreatic cell differentiation. *Proc Natl Acad Sci U S A* **110**: 17892–7.

Fujiki R, Hashiba W, Sekine H, Yokoyama A, Chikanishi T, Ito S, Imai Y, Kim J, He HH, Igarashi K, et al. 2011. GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* **480**: 557–560.

Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, Kou X, Zhang Y, Huang H, Jiang Y, et al. 2013. Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming. *Cell Stem Cell* **12**: 453–469.

Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, Brückl T, Biel M, Carell T. 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**: 1–9.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**: 481–514.

Gu T-P, Guo F, Yang H, Wu H-P, Xu G-F, Liu W, Xie Z-G, Shi L, He X, Jin S, et al. 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**: 606–610.

Gudas LJ. 2013. Retinoids induce stem cell differentiation via epigenetic changes. *Semin Cell Dev Biol* **24**: 701–705.

Guibert S, Forné T, Weber M. 2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res* **22**: 633–41.

Guilhamon P, Eskandarpour M, Halai D, Wilson G a, Feber A, Teschendorff AE, Gomez V, Hergovich A, Tirabosco R, Fernanda Amary M, et al. 2013. Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nat Commun* **4**: 2166.

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu T-P, Hu B, et al. 2014. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell* **15**: 447–458.

Guo JU, Ma DK, Mo H, Ball MP, Jang M-H, Bonaguidi MA, Balazer JA, Eaves HL, Xie B, Ford E, et al. 2011. Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci* **14**: 1345–1351.

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, et al. 2013. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* **17**: 215–222.

Gurdon JB, Elsdale TR, Fischberg M. 1958. Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei. *Nature* **182**: 64–65.

Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA. 2013. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**: 448–52.

Haffner MC, Chaux A, Meeker AK, Esopi DM, Gerber J, Pellakuru LG, Toubaji A, Argani P, Iacobuzio-Donahue C, Nelson WG, et al. 2011. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* **2**: 627–37.

Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F, Lee C, Almouzni G, Schneider R, Surani MA. 2008. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**: 877–81.

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA. 2002. Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**: 15–23.

Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA. 2010. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**: 78–82.

Hallberg L, Brune M, Rossander L. 1989. The role of vitamin C in iron absorption. *Int J Vitam Nutr Res Suppl* **30**: 103–8.

Handa V, Jeltsch A. 2005. Profound Flanking Sequence Preference of Dnmt3a and Dnmt3b Mammalian DNA Methyltransferases Shape the Human Epigenome. *J Mol Biol* **348**: 1103–1112.

Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, Zhang X, Cheng X. 2012.

Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* **40**: 4841–4849.

Hashimoto H, Pais JE, Dai N, Corrêa IR, Zhang X, Zheng Y, Cheng X. 2015. Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. *Nucleic Acids Res* **43**: 10713–21.

Hashimoto H, Pais JE, Zhang X, Saleh L, Fu Z-Q, Dai N, Corrêa IR, Zheng Y, Cheng X. 2014. Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**: 391–5.

He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, et al. 2011. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science (80-)* **333**: 1303–1307.

Holliday R, Pugh J. 1975. DNA modification mechanisms and gene activity during development. *Science (80-)* **187**: 226–32.

Hore TA, Rapkins RW, Graves JAM. 2007. Construction and evolution of imprinted loci in mammals. *Trends Genet* **23**: 440–448.

Hore TA, von Meyenn F, Ravichandran M, Bachman M, Ficz G, Oxley D, Santos F, Balasubramanian S, Jurkowski TP, Reik W. 2016. Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. *Proc Natl Acad Sci U S A* **113**: 12202–12207.

Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, Shi YG, Zhu J, Wang P, Xu Y. 2013. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155**: 1545–1555.

Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, Lou Z, Zhang L, Li W, Gong W, et al. 2015. Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**: 118–22.

Hu X, Zhang L, Mao S-Q, Li Z, Chen J, Zhang R-R, Wu H-P, Gao J, Guo F, Liu W, et al. 2014. Tet and TDG Mediate DNA Demethylation Essential for Mesenchymal-to-Epithelial Transition in Somatic Cell Reprogramming. *Cell Stem Cell* **14**: 512–522.

Huang Y, Rao A. 2014. Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet* **30**: 464–474.

Ichiyama K, Chen T, Wang X, Yan X, Kim B-S, Tanaka S, Ndiaye-Lobry D, Deng Y, Zou Y, Zheng P, et al. 2015. The Methylcytosine Dioxygenase Tet2 Promotes DNA Demethylation and Activation of Cytokine Gene Expression in T Cells. *Immunity* **42**: 613–626.

Inoue S, Lemonnier F, Mak TW. 2016. Roles of IDH1/2 and TET2 mutations in myeloid disorders. *Int J Hematol* **103**: 627–33.

Iqbal K, Jin S-G, Pfeifer GP, Szabó PE. 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* **108**: 3642–3647.

Ito S, D'Alessio AC, Taranova O V, Hong K, Sowers LC, Zhang Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**: 1129–1133.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. 2011. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (80-)* **333**: 1300–1303.

Iurlaro M, Ficz G, Oxley D, Raiber E-A, Bachman M, Booth MJ, Andrews S, Balasubramanian S, Reik W. 2013. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol* **14**: R119.

Iurlaro M, Mcinroy GR, Burgess HE, Dean W, Raiber E, Bachman M, Beraldi D, Balasubramanian S, Reik W. 2016. In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biol* **17**: 141.

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

Jeltsch A. 2006. On the Enzymatic Properties of Dnmt1: Specificity, Processivity, Mechanism of Linear Diffusion and Allosteric Regulation of the Enzyme. *Epigenetics* **1**: 63–66.

Jeltsch A, Jurkowska RZ. 2014. New concepts in DNA methylation. *Trends Biochem Sci* **39**: 310–8.

Jeschke J, Collignon E, Fuks F. 2016. Portraits of TET-mediated DNA hydroxymethylation in cancer. *Curr Opin Genet Dev* **36**: 16–26.

Ji D, Lin K, Song J, Wang Y, Jaenisch R, Bird A, Robertson KD, Okano M, Bell DW, Haber DA, et al. 2014. Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol Biosyst* **10**: 1749.

Ji W, Yang L, Yuan J, Yang L, Zhang M, Qi D, Duan X, Xuan A, Zhang W, Lu J, et al. 2013. MicroRNA-152 targets DNA methyltransferase 1 in NiS-transformed cells via a feedback mechanism. *Carcinogenesis* **34**: 446–53.

Jin S-G, Zhang Z-M, Dunwell TL, Harter MR, Wu X, Johnson J, Li Z, Liu J, Szabó PE, Lu Q, et al. 2016. Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep* **14**: 493–505.

Jin SG, Wu X, Li AX, Pfeifer GP. 2011. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res* **39**: 5015–5024.

Jost JP. 1993. Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. *Proc Natl Acad Sci U S A* **90**: 4684–8.

Jurkowska RZ, Jurkowski TP, Jeltsch A. 2011a. Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem* **12**: 206–222.

Jurkowska RZ, Siddique AN, Jurkowski TP, Jeltsch A. 2011b. Approaches to Enzyme and Substrate Design of the Murine Dnmt3a DNA Methyltransferase. *ChemBioChem* **12**: 1589–1594.

Jurkowski TP, Anspach N, Kulishova L, Nellen W, Jeltsch A. 2007. The M.EcoRV DNA-(adenine N6)-methyltransferase uses DNA bending for recognition of an expanded EcoDam recognition site. *J Biol Chem* **282**: 36942–52.

Kaas GA, Zhong C, Eason DE, Ross DL, Vachhani R V, Ming G, King JR, Song H, Sweatt JD. 2013. Hydroxylation , Active DNA Demethylation , Gene Transcription , and Memory Formation. *Neuron* **79**: 1086–1093.

Ke Q, Davidson T, Chen H, Kluz T, Costa M. 2006. Alterations of histone modifications and transgene silencing by nickel chloride. *Carcinogenesis* **27**: 1481–8.

Ke Q, Ellen TP, Costa M. 2008a. Nickel compounds induce histone ubiquitination by inhibiting histone deubiquitinating enzyme activity. *Toxicol Appl Pharmacol* **228**: 190–199.

Ke Q, Li Q, Ellen TP, Sun H, Costa M. 2008b. Nickel compounds induce phosphorylation of histone H3 at serine 10 by activating JNK-MAPK pathway. *Carcinogenesis* **29**: 1276–81.

Kienhöfer S, Musheev MU, Stapf U, Helm M, Schomacher L, Niehrs C, Schäfer A. 2015. GADD45a physically and functionally interacts with TET1. *Differentiation* **90**: 59–68.

Kim MY, Zhang T, Kraus WL. 2005. Poly (ADP-ribosyl) ation by PARP-1 : " PAR-laying " NAD + into a nuclear signal. **14**: 493–505.

Kinney SM, Chin HG, Vaisvila R, Bitinaite J, Zheng Y, Estve PO, Feng S, Stroud H, Jacobsen SE, Pradhan S. 2011. Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *J Biol Chem* **286**: 24685–24693.

Klimasauskas S, Kumar S, Roberts RJ, Cheng X. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**: 357–69.

Klug M, Schmidhofer S, Gebhard C, Andreesen R, Rehli M. 2013. 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. *Genome Biol* **14**: R46.

Knappskog S, Myklebust LM, Busch C, Aloysius T, Varhaug JE, Lonning PE, Lillehaug JR, Pendino F. 2011. RINF (CXXC5) is overexpressed in solid tumors and is an unfavorable prognostic factor in breast cancer. *Ann Oncol* **22**: 2208–2215.

Ko M, An J, Bandukwala HS, Chavez L, Aijö T, Pastor W a, Segal MF, Li H, Koh KP, Lähdesmäki H, et al. 2013. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**: 122–6.

Ko M, Bandukwala HS, An J, Lamperti ED, Thompson EC, Hastie R. 2011. Homeostasis and Differentiation of Hematopoietic Stem Cells in Mice. *Proc Natl Acad Sci* **108**: 14566–71.

Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, et al. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**: 839–843.

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, et al. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**: 200–13.

Koivunen P, Lee S, Duncan CG, Lopez G, Lu G, Ramkissoon S, Losman JA, Joensuu P, Bergmann U, Gross S, et al. 2012. Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**: 484–488.

Kothari RM, Shankar V. 1976. 5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity. *J Mol Evol* **7**: 325–9.

Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (80-)* **324**: 929–930.

Ku CS, Naidoo N, Wu M, Soong R. 2011. Studying the epigenome using next generation sequencing. *J Med Genet* **48**: 721–30.

Kudo Y, Tateishi K, Yamamoto K, Yamamoto S, Asaoka Y, Ijichi H, Nagae G, Yoshida H, Aburatani H, Koike K. 2012. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci* **103**: 670–676.

Kunimoto H, Fukuchi Y, Sakurai M, Sadahira K, Ikeda Y, Okamoto S, Nakajima H. 2012. Tet2 disruption leads to enhanced self-renewal and altered differentiation of fetal liver hematopoietic stem cells.

Sci Rep **2**: 1–10.

Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M. 2008. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* **22**: 1617–35.

Langemeijer S, Kuiper R, Berends M, Knops R, Aslanyan M, Massop M, van Hoogen P, van Kessel AG, Raymakers R, Verburgh E, et al. 2009. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Leuk Res* **33**: 838–842.

Laukka T, Mariani CJ, Ihantola T, Cao JZ, Hokkanen J, Kaelin WG, Godley LA, Koivunen P. 2016. Fumarate and Succinate Regulate Expression of Hypoxia-inducible Genes via TET Enzymes. *J Biol Chem* **291**: 4256–65.

Lawson KA, Hage WJ. 1994. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* **182**: 68-84-91.

Le May N, Iltis I, Amé J-C, Zhovmer A, Biard D, Egly J-M, Schreiber V, Coin F. 2012. Poly (ADP-ribose) glycohydrolase regulates retinoic acid receptor-mediated gene expression. *Mol Cell* **48**: 785–98.

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915–926.

Li W, Liu M, Li W, Liu M. 2011. Distribution of 5-Hydroxymethylcytosine in Different Human Tissues. *J Nucleic Acids* **2011**: 1–5.

Li Z, Gu TP, Weber AR, Shen JZ, Li BZ, Xie ZG, Yin R, Guo F, Liu X, Tang F, et al. 2015. Gadd45a promotes DNA demethylation through TDG. *Nucleic Acids Res* **43**: 3986–3997.

Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu WW, Tan L, Hu Y, Zhan Q, et al. 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of Melanoma. *Cell* **150**: 1135–1146.

Lin LL, Wang W, Hu Z, Wang LW, Chang J, Qian H. 2015. Erratum to: Negative feedback of miR-29 family TET1 involves in hepatocellular cancer. *Med Oncol* **32**: 39.

Lister R, Mukamel E a, Nery JR, Urich M, Puddifoot C a, Nicholas D, Lucero J, Huang Y, Dwork AJ, Schultz MD, et al. 2013. Global Epigenomic Reconfi guration During Mammalian Brain Development. *Science* **341**: 629–643.

Liu C, Liu L, Chen X, Shen J, Shan J, Xu Y, Yang Z, Wu L, Xia F, Bie P, et al. 2013a. Decrease of 5-Hydroxymethylcytosine Is Associated with Progression of Hepatocellular Carcinoma through Downregulation of TET1. *PLoS One* **8**: 1–9.

Liu MY, Torabifard H, Crawford DJ, DeNizio JE, Cao X-J, Garcia BA, Cisneros GA, Kohli RM. 2017. Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. *Nat Chem Biol* **13**: 181–187.

Liu R, Jin Y, Tang WH, Qin L, Zhang X, Tellides G, Hwa J, Yu J, Martin KA. 2013b. Ten-eleven translocation-2 (TET2) is a master regulator of smooth muscle cell plasticity. *Circulation* **128**: 2047–57.

Liutkevičiutė Z, Kriukienė E, Ličytė J, Rudytė M, Urbanavičiutė G, Klimašauskas S. 2014. Direct decarboxylation of 5-Carboxylcytosine by DNA C5- Methyltransferases. *J Am Chem Soc* **136**: 5884–5887.

Liutkeviciute Z, Lukinavicius G, Masevicius V, Daujotyte D, Klimasauskas S. 2009. Cytosine-5-

methyltransferases add aldehydes to DNA. *Nat Chem Biol* **5**: 400–402.

Maiti A, Drohat AC. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J Biol Chem* **286**: 35334–35338.

Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y, Fan G, Sun YE. 2003. DNA Methylation-Related Chromatin Remodeling in Activity-Dependent Bdnf Gene Regulation. *Science (80-)* **302**.

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. 2000a. Demethylation of the zygotic paternal genome. *Nature* **403**: 501–2.

Mayer W, Smith A, Fundele R, Haaf T. 2000b. Spatial separation of parental genomes in preimplantation mouse embryos. *J Cell Biol* **148**: 629–34.

McNeill LA, Flashman E, Buck MRG, Hewitson KS, Clifton IJ, Jeschke G, Claridge TDW, Ehrismann D, Oldham NJ, Schofield CJ. 2005. Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol Biosyst* **1**: 321–4.

Minor E a., Court BL, Young JI, Wang G. 2013. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J Biol Chem* **288**: 13669–13674.

Moran-crusio K, Reavie L, Shih A, Abdel-wahab O, Ndiaye-lobry D, Lobry C, Figueroa ME, Vasanthakumar A, Patel J, Zhao X, et al. 2012. Tet2 loss leads to increased hematopoietic stem cell self- renewal and myeloid transformation. *Cancer* **20**: 11–24.

Morgan B, Lahav O. 2007. The effect of pH on the kinetics of spontaneous Fe(II) oxidation by O2 in aqueous solution - basic principles and a simple heuristic description. *Chemosphere* **68**: 2080–2084.

Müller U, Bauer C, Siegl M, Rottach A, Leonhardt H. 2014. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res* **42**: 8592–604.

Münzel M, Globisch D, Carell T. 2011. 5-Hydroxymethylcytosine, the Sixth Base of the Genome. *Angew Chemie - Int Ed* **50**: 6460–6468.

Myllyla R, Majamaa K, Gunzler V. 1984. Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. *J Biol Chem* **259**: 5403–5405.

Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM. 2012. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* **8**: 751–8.

Nakagawa T, Lv L, Nakagawa M, Yu Y, Yu C, D'Alessio AC, Nakayama K, Fan H-Y, Chen X, Xiong Y. 2015. CRL4VprBP E3 Ligase Promotes Monoubiquitylation and Chromatin Binding of TET Dioxygenases. *Mol Cell* **57**: 247–260.

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M, et al. 2007. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat Cell Biol* **9**: 64–71.

Nakamura T, Liu Y-J, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T. 2012. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**: 415–419.

Neri F, Incarnato D, Krepelova A, Dettori D, Rapelli S, Maldotti M, Parlato C, Anselmi F, Galvagni F, Oliviero S. 2015a. TET1 is controlled by pluripotency-associated factors in ESCs and downmodulated by PRC2 in differentiated cells and tissues. *Nucleic Acids Res* **43**: 6814–26.

Neri F, Incarnato D, Krepelova A, Rapelli S, Anselmi F, Parlato C, Medana C, Dal Bello F, Oliviero S. 2015b. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep* **10**: 674–683.

Neri F, Incarnato D, Krepelova A, Rapelli S, Pagnani A, Zecchina R, Parlato C, Oliviero S. 2013. Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol* **14**: R91.

Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR. 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res* **22**: 467–477.

Nietfeld JJ, Kemp A. 1981. The function of ascorbate with respect to prolyl 4-hydroxylase activity. *Biochim Biophys Acta* **657**: 159–67.

Oermann EK, Wu J, Guan K-L, Xiong Y. 2012. Alterations of metabolic genes and metabolites in cancer. *Semin Cell Dev Biol* **23**: 370–380.

Okano M, Bell DW, Haber DA, Li E. 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**: 247–257.

Olinski R, Starczak M, Gackowski D. 2016. Enigmatic 5-hydroxymethyluracil: Oxidatively modified base, epigenetic mark or both? *Mutat Res Mutat Res* **767**: 59–66.

Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. 2002. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res* **62**: 4075–4080.

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J. 2000. Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**: 475–8.

Ozer A, Bruick RK. 2007. Non-heme dioxygenases: cellular sensors and regulators jelly rolled into one? *Nat Chem Biol* **3**: 144–153.

Pais JE, Dai N, Tamanaha E, Vaisvila R, Fomenkov AI, Bitinaite J, Sun Z, Guan S, Corrêa IR, Noren CJ, et al. 2015. Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. *Proc Natl Acad Sci* **112**: 4316–4321.

Pastor WA, Aravind L, Rao A. 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* **14**: 341–56.

Pastor WWA, Pape UJU, Huang Y, Henderson HRHHR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, et al. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**: 394–397.

Penn NW, Suwalski R, O'Riley C, Bojanowski K, Yura R. 1972. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J* **126**: 781–90.

Perera A, Eisen D, Wagner M, Laube SK, Künzel AF, Koch S, Steinbacher J, Schulze E, Splith V, Mittermeier N, et al. 2015. TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression. *Cell Rep* **11**: 283–294.

Petrat F, Paluch S, Dogruöz E, Dörfler P, Kirsch M, Korth H-G, Sustmann R, de Groot H. 2003. Reduction of Fe(III) ions complexed to physiological ligands by lipoyl dehydrogenase and other flavoenzymes in vitro: implications for an enzymatic reduction of Fe(III) ions of the labile iron pool. *J Biol Chem* **278**: 46403–13.

Pfaffeneder T, Hackner B, Truss M, Münzel M, Müller M, Deiml CA, Hagemeier C, Carell T. 2011. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* **50**: 7008–12.

Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, Truss M, Steinbacher J, Hackner B, Kotljarova O, et al. 2014. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* **10**: 574–81.

Piccolo FM, Bagci H, Brown KE, Landeira D, Soza-Ried J, Feytout A, Mooijman D, Hajkova P, Leitch HG, Tada T, et al. 2013. Different Roles for Tet1 and Tet2 Proteins in Reprogramming-Mediated Erasure of Imprints Induced by EGC Fusion. *Mol Cell* **49**: 1023–1033.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**: 1101–5.

Pritchard JB. 1995. Intracellular alpha-ketoglutarate controls the efficacy of renal organic anion transport. *J Pharmacol Exp Ther* **274**: 1278–84.

Quivoron C, Couronne L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, Do Cruzeiro M, Delhommeau F, Arnulf B, Stern MH, et al. 2011. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* **20**: 25–38.

Raiber E-A, Beraldi D, Ficz G, Burgess HE, Branco MR, Murat P, Oxley D, Booth MJ, Reik W, Balasubramanian S. 2012. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol* **13**: R69.

Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, Li Y, Ahn J, Abdel-Wahab O, Shih A, et al. 2014. DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia. *Cell Rep* **9**: 1841–1855.

Rasmussen KD, Jia G, Johansen J V., Pedersen MT, Rapin N, Bagger FO, Porse BT, Bernard OA, Christensen J, Helin K. 2015. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev* **29**: 910–922.

Ratel D, Ravanat J-L, Berger F, Wion D. 2006. N6-methyladenine: the other methylated base of DNA. *BioEssays* **28**: 309–315.

Rawłuszko-Wieczorek AA, Siera A, Jagodziński PP. 2015. TET proteins in cancer: current “state of the art.” *Crit Rev Oncol Hematol* **96**: 425–36.

Reik W, Surani MA. 2015. Germline and Pluripotent Stem Cells. *Cold Spring Harb Perspect Biol* **7**: a019422.

Riggs AD. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenet Genome Res* **14**: 9–25.

Robertson AB, Dahl JA, Vagbo CB, Tripathi P, Krokan HE, Klungland A. 2011. A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids*

Res **39**: e55.

Rudenko A, Dawlaty MM, Seo J, Cheng AW, Meng J, Le T, Faull KF, Jaenisch R, Tsai L-H. 2013. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**: 1109–22.

Santiago M, Antunes C, Guedes M, Sousa N, Marques CJ. 2014. TET enzymes and DNA hydroxymethylation in neural development and function — How critical are they? *Genomics* **104**: 334–340.

Santos F, Peat J, Burgess H, Rada C, Reik W, Dean W. 2013. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* **6**: 39.

Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**: 1412–7.

Schiller R, Scozzafava G, Tumber A, Wickens JR, Bush JT, Rai G, Lejeune C, Choi H, Yeh T-L, Chan MC, et al. 2014. A cell-permeable ester derivative of the JmjC histone demethylase inhibitor IOX1. *ChemMedChem* **9**: 566–71.

Sciacovelli M, Gonçalves E, Johnson TI, Zecchini VR, da Costa ASH, Gaude E, Drubbel AV, Theobald SJ, Abbo SR, Tran MGB, et al. 2016. Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition. *Nature* **537**: 544–47.

Scourzic L, Mouly E, Bernard OA. 2015. TET proteins and the control of cytosine demethylation in cancer. *Genome Med* **7**: 1–16.

Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W. 2012. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol Cell* **48**: 849–862.

Seisenberger S, Peat JR, Hore T a, Santos F, Dean W, Reik W. 2013. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci* **368**: 20110330.

Sérandour AA, Avner S, Mahé EA, Madigou T, Guibert S, Weber M, Salbert G, Tahiliani M, Koh K, Shen Y, et al. 2016. Single-CpG resolution mapping of 5-hydroxymethylcytosine by chemical labeling and exonuclease digestion identifies evolutionarily unconserved CpGs as TET targets. *Genome Biol* **17**: 56.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, et al. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**: 908–12.

Shen L, Inoue A, He J, Liu Y, Lu F, Zhang Y. 2014. Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell* **15**: 459–470.

Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, Zhang K, Zhang Y. 2013. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**: 692–706.

Shen L, Zhang Y. 2013. 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr Opin Cell Biol* **25**: 289–96.

Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H, Sasaki H, Matsui Y, et al. 2013.

Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases ed. M.S. Bartolomei. *PLoS Genet* **9**: e1003439.

Sirivech S, Frieden E, Osaki S. 1974. The release of iron from horse spleen ferritin by reduced flavins. *Biochem J* **143**: 311–5.

Smiley JA, Kundracik M, Landfried DA, Barnes VR, Axhemi AA. 2005. Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a Rhodotorula glutinis cDNA library and iso-orotate decarboxylase from Neurospora crassa. *Biochim Biophys Acta* **1723**: 256–64.

Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**: 204–20.

Song C-X, Szulwach KE, Dai Q, Fu Y, Mao S-QQ, Lin L, Street C, Li Y, Poidevin M, Wu HPH, et al. 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**: 678–691.

Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C-H, Zhang W, Jian X, et al. 2011a. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* **29**: 68–72.

Song CX, Sun Y, Dai Q, Lu XY, Yu M, Yang CG, He C. 2011b. Detection of 5-Hydroxymethylcytosine in DNA by Transferring a Keto-Glucose by Using T4 Phage ß-Glucosyltransferase. *ChemBioChem* **12**: 1682–1685.

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Münzel M, Wagner M, Müller M, Khan F, et al. 2013. Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**: 1146–1159.

Studier FW. 2005. Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expr Purif* **41**: 207–234.

Sun H, Shamy M, Costa M. 2013. Nickel and epigenetic gene silencing. *Genes (Basel)* **4**: 583–95.

Svirbely JL, Szent-Györgyi A. 1933. The chemical nature of vitamin C. *Biochem J* **27**: 279–85.

Szulwach KE, Li X, Li Y, Song CX, Han JW, Kim S, Namburi S, Hermetz K, Kim JJ, Rudd MK, et al. 2011. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* **7**: e1002154.

Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38**: e181.

Tahiliani M, Koh KP, Shen Y, Pastor W a, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930–935.

Takahashi K, Yamanaka S. 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**: 663–676.

Tamanaha E, Guan S, Marks K, Saleh L. 2016. 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* **138**: 9345–9348.

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, et al. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1.

Nature **464**: 1082–6.

Tsai Y-P, Chen H-F, Chen S-Y, Cheng W-C, Wang H-W, Shen Z-J, Song C, Teng S-C, He C, Wu K-J, et al. 2014. TET1 regulates hypoxia-induced epithelial-mesenchymal transition by acting as a co-activator. *Genome Biol* **15**: 513.

Valinluck V, Tsai H-H, Rogstad DK, Burdzy A, Bird A, Sowers LC. 2004. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res* **32**: 4100–8.

Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, Roberto A, Christensen J, Bonaldi T, Helin K, et al. 2013. Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Mol Cell* **49**: 645–656.

Verschoor MJ, Molot LA. 2013. A comparison of three colorimetric methods of ferrous and total reactive iron measurement in freshwaters. *Limnol Oceanogr Methods* **11**: 113–125.

Vestergaard M, Matsumoto S, Nishikori S, Shiraki K, Hirata K, Takagi M. 2008. Chelation of Cadmium Ions by Phytochelatin Synthase: Role of the Cystein-rich C-Terminal. *Anal Sci* **24**: 277–281.

Viguié F, Aboura A, Bouscary D, Ramond S, Delmer A, Tachdjian G, Marie JP, Casadevall N. 2005. Common 4q24 deletion in four cases of hematopoietic malignancy: early stem cell involvement? *Leukemia* **19**: 1411–5.

von Meyenn F, Iurlaro M, Habibi E, He C, Reik W, Stunnenberg HG. 2016. Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells Article. *Mol Cell* **62**: 1–14.

Wang D, Hashimoto H, Zhang X, Barwick BG, Lonial S, Boise LH, Vertino PM, Cheng X. 2016. MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res*.

Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L, Wang P, Yang H, Ma S, Lin H, et al. 2015. WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation. *Mol Cell* **57**: 1–12.

Wang Y, Zhang Y. 2014. Regulation of TET protein stability by calpains. *Cell Rep* **6**: 278–284.

Weissmann S, Alpermann T, Grossmann V, Kowarsch A, Nadarajah N, Eder C, Dicker F, Fasan A, Haferlach C, Haferlach T, et al. 2012. Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia* **26**: 934–942.

Wen L, Li X, Yan L, Tan Y, Li R, Zhao Y, Wang Y, Xie J, Zhang Y, Song C, et al. 2014. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol* **15**: R49.

Williams K, Christensen J, Pedersen MT, Johansen J V., Cloos PAC, Rappsilber J, Helin K, Manuscript A. 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**: 343–348.

Wilmouth RC, Turnbull JJ, Welford RW., Clifton IJ, Prescott AG, Schofield CJ. 2002. Structure and Mechanism of Anthocyanidin Synthase from Arabidopsis thaliana. *Structure* **10**: 93–103.

Wilson GG, Murray NE. 1991. Restriction and Modification Systems. *Annu Rev Genet* **25**: 585–627.

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J. 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic

reprogramming. *Nat Commun* **2**: 241.

Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**: 389–393.

Wu H, Wu X, Shen L, Zhang Y. 2014. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat Biotechnol* **32**: 1231–1240.

Wu H, Wu X, Zhang Y. 2016a. Base-resolution profiling of active DNA demethylation using MAB-seq and caMAB-seq. *Nat Protoc* **11**: 1081–1100.

Wu H, Zhang Y. 2015. Charting oxidized methylcytosines at base resolution. *Nat Struct Mol Biol* **22**: 656–61.

Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, Liu Y, Byrum SD, Mackintosh SG, Zhong M, et al. 2016b. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**: 329–33.

Wyatt GR, Cohen SS. 1952. A new pyrimidine base from bacteriophage nucleic acids. *Nature* **170**: 1072–3.

Xia B, Han D, Lu X, Sun Z, Zhou A, Yin Q, Zeng H, Liu M, Jiang X, Xie W, et al. 2015. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat Methods* **12**: 1047–1050.

Xiao M, Yang H, Xu W, Ma S, Lin H, Zhu H, Liu L, Liu Y, Yang C, Xu Y, et al. 2012. Inhibition of a-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev* **26**: 1326–1338.

Xu W, Yang H, Liu Y, Yang Y, Wang PP, Kim S-H, Ito S, Yang C, Wang PP, Xiao M-T, et al. 2011a. Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* **19**: 17–30.

Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S, et al. 2011b. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Mol Cell* **42**: 451–464.

Xu Y, Xu C, Kato A, Tempel W, Abreu JG, Bian C, Hu Y, Hu D, Zhao B, Cerovina T, et al. 2012. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for xenopus eye and neural development. *Cell* **151**: 1200–1213.

Xue S, Liu C, Sun X, Li W, Zhang C, Zhou X, Lu Y, Xiao J, Li C, Xu X, et al. 2016. TET3 Inhibits Type I IFN Production Independent of DNA Demethylation. *Cell Rep* **16**: 1096–105.

Yamaguchi S, Hong K, Liu R, Shen L, Inoue A, Diep D, Zhang K, Zhang Y. 2012. Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**: 443–447.

Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y. 2013. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**: 460–4.

Yamazaki Y, Mann MRW, Lee SS, Marh J, McCarrey JR, Yanagimachi R, Bartolomei MS. 2003. Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. *Proc Natl Acad Sci U S A* **100**: 12207–12.

Yang C-G, Yi C, Duguid EM, Sullivan CT, Jian X, Rice PA, He C. 2008. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**: 961–965.

Yang H, Liu Y, Bai F, Zhang J-Y, Ma S-H, Liu J, Xu Z-D, Zhu H-G, Ling Z-Q, Ye D, et al. 2013. Tumor

development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**: 663–9.

Yang J, Wang W, Ooi J, Campos LS, Lu L, Liu P. 2015a. Signalling Through Retinoic Acid Receptors is Required for Reprogramming of Both Mouse Embryonic Fibroblast Cells and Epiblast Stem Cells to Induced Pluripotent Stem Cells. *Stem Cells* **33**: 1390–1404.

Yang R, Qu C, Zhou Y, Konkel JE, Shi S, Liu Y, Chen C, Liu S, Liu D, Chen Y, et al. 2015b. Hydrogen Sulfide Promotes Tet1- and Tet2-Mediated Foxp3 Demethylation to Drive Regulatory T Cell Differentiation and Maintain Immune Homeostasis. *Immunity* **43**: 251–263.

Ye D, Ma S, Xiong Y, Guan K-L. 2013. R-2-hydroxyglutarate as the key effector of IDH mutations promoting oncogenesis. *Cancer Cell* **23**: 274–6.

Yin R, Mao SQ, Zhao B, Chong Z, Yang Y, Zhao C, Zhang D, Huang H, Gao J, Li Z, et al. 2013. Ascorbic acid enhances tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J Am Chem Soc* **135**: 10396–10403.

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, et al. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**: 1368–1380.

Zhang D, Iyer LM, Burroughs a. M, Aravind L. 2014. Resilience of biochemical activity in protein domains in the face of structural divergence. *Curr Opin Struct Biol* **26**: 92–103.

Zhang H, Zhang X, Clark E, Mulcahey M, Huang S, Shi YG. 2010. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res* **20**: 1390–1393.

Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, Zhao D, Liu Y, Wang C, Zhang X, et al. 2015. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* **525**: 389–93.

Zhang R-R, Cui Q-Y, Murai K, Lim YC, Smith ZD, Jin S, Ye P, Rosa L, Lee YK, Wu H-P, et al. 2013. Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* **13**: 237–45.

Zhang X, Su J, Jeong M, Ko M, Huang Y-H, Park HJ, Guzman A, Lei Y, Huang Y-H, Rao A, et al. 2016. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet* **48**: 1014–23.

Zhu B, Zheng Y, Angliker H, Schwarz S, Thiry S, Siegmann M, Jost JP. 2000a. 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res* **28**: 4157–65.

Zhu B, Zheng Y, Hess D, Angliker H, Schwarz S, Siegmann M, Thiry S, Jost JP. 2000b. 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc Natl Acad Sci U S A* **97**: 5135–9.

Zhu J-K. 2009. Active DNA Demethylation Mediated by DNA Glycosylases. *Annu Rev Genet* **43**: 143–166.