UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mass Spectrometric Study of RNA Epigenetic Modifications

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Environmental Toxicology

by

Lijuan Fu

June 2015

Dissertation Committee:

Dr. Yinsheng Wang, Chairperson
Dr. Connie Nugent
Dr. Wenwan Zhong

The Dissertation of Lijuan Fu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

First of all, I would like to show my appreciation to my supervisor, Professor Yinsheng Wang. His guidance, help and support enabled me to investigate a completely new bioanalytical field and complete this thesis work. His enthusiasm and dedication to science are contagious and motivational for me. I appreciate all his contributions of time, ideas and help on both my professional and personal growth.

I would also like to show my gratitude to my guidance and dissertation committee members, Professors Connie Nugent, Li Fan and Wenwan Zhong. Their suggestions and discussion with my research in the past years are really appreciated and helpful. I am also grateful for the technical supports from staff members at UC Riverside. I thank Dr. Dan Borchardt in the Analytical Chemistry Instrumentation Facility at UCR for his help with NMR measurements and data processing. Thank Mr. Ron New and Dr. Richard Kondrat's help for using Thermo TSQ Vantage. Additionally, I would like to thank my collaborators, including Professor Stephen Baylin at John Hopkins University for kindly providing the Tet2 expression plasmids, Professor Guoliang Xu at the Chinese Academy of Sciences of China for providing the Tet1 and Tet3 expression plasmids, Professor Laura J. Niedernhofer at Scripps Florida for providing mouse tissue samples.

Next, I would like to show my sincere thanks to the current and former members of the Wang group. Especially, I would like to thank Dr. Yongsheng Xiao for his help to set up

the Nano-LC system. I would also like to thank Dr. Hongxia Wang for teaching me a lot of basic techniques in this lab and Dr. Bifeng Yuan for his discussion and suggestions when I started my research here. I also feel lucky to have support and discussion from Shuo Liu, Ming Huang and Dr. Qian Cai, Dr. Candace Guerrero, and Dr. Nicholas Amato. In addition, I would like to appreciate Pengcheng Wang for the help with the synthesis of standards used in this thesis and Yang Yu for giving me a lot of priority to use the instrument in the last quarters. Moreover, I want to show my gratitude to all other former and current group members: Dr. Jin Wang, Dr. John Prins, Dr. Changjun You, Dr. Jianshuang Wang, Dr. Ashley Swanson, Dr. Lei Guo, Dr. Xiaoxia Dai, Dr. Debin Ji, Dr. Qianqian Zhai, Dr. Jun Wu, Dr. Lin Li, Dr. Nathan Price, Dr. Nissana Aderson, Dr. Renee William, Dr. Tao Bing, Dr. Xianggang Jiang, Preston Williams, Zi Wang, Yuxiang Cui, Jiabin Wu, Weili Miao.

Lastly, I am so appreciated my parents, my grandparents, my sister and my other family members for their endless love, continuous encouragement and support. I also feel so fortunate to have my husband, Yunhua, and my son, Zhixuan, who were accompanied with me during my most frustrated time during my PhD study. Also, I would like to appreciate my parents-in-law who help me in my last 2 years in Riverside. Without their help to take care of my son, I could not be able to finish my PhD thesis. Furthermore, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

COPYRIGHT ACKNOWLEDGEMENTS

The text and figures in Chapter 3, in part or in full, are a reprint of the material as it appears in *J Am Chem Soc.,* 2014, 136 (33), pp. 11582-5. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research which forms the basis of this chapter.

DEDICATION

To my husband Yunhua Liu, my son Zhixuan Liu, my mother Cuilin Li, my grandma Xiuying Li and my sister Liping Fu. I appreciate so much to have all of you in this world.

ABSTRACT OF THE DISSERTATION

Mass Spectrometric Study of RNA Epigenetic Modifications

by

Lijuan Fu

Doctor of Philosophy, Graduate Program in Environmental Toxicology
University of California, Riverside, June 2015
Dr. Yinsheng Wang, Chairperson

The rising interest in understanding the functions, regulation and maintenance of the epitranscriptome calls for robust and accurate analytical methods for the identification and quantification of post-transcriptionally modified nucleosides in RNA. Mass spectrometry has become a very powerful tool for bioanalysis which can elucidate the structure of substances and provide quantitative measurements. The LC-MS-based analytical method, in combination with genetic manipulation, may facilitate the studies in the area of epitranscriptome. In this thesis, I focus on the development of novel MS-based strategies to identify and quantify post-transcriptional modifications present in total RNA and mRNA isolated from mammalian tissues and cultured human cells. Additionally, by using these analytical methods, I was able to discover new enzymes involved in demethylation of mono-methylated cytosine in RNA both *in vitro* and *in vivo*.

In Chapter 2, an LC-MS/MS/MS coupled with the stable isotope-dilution method was developed for the sensitive and accurate quantifications of 5-methylcytidine (m^5C), 2'-*O*-methylcytidine (Cm), N^6-methyladenosine (m^6A) and 2′-*O*-methyladenosine (Am)

in RNA isolated from mammalian cells and tissues. Our results showed that the distributions of these four methylated nucleosides are tissue-specific. We also found that the levels of m^5C, Cm and Am are significantly lower (by 6.5-43 fold) in mRNA than in total RNA isolated from HEK293T cells, whereas the level of m^6A was slightly higher (by 1.6 fold) in mRNA than in total RNA.

In Chapter 3, I first demonstrated that Tet enzymes can catalyze the formation of 5-hydroxymethylcytidine (5-hmrC), 5-formycytidine (5-forC) and 5-carboxycytidine (5-carC) from m^5C *in vitro*. Subsequently, I established a sensitive and accurate LC-MS/MS/MS with the isotope-dilution method to measure the level of 5-hmrC *in vivo* and further demonstrated that the catalytic domains of all three Tet enzymes as well as full-length Tet3 could induce the formation of 5-hmrC in human cells.

In Chapter 4, I selected four Fe(Π)- and 2-oxoglutarate-dependent dioxygenases, including FTO, ALKBH5, ALKBH2 and ALKBH3, to test their demethylase activity towards m^5C in RNA in human cells by using the analytical methods established in Chapters 2 & 3. Our results showed that, the level of 5-hmrC is significantly decreased whereas the level of m^5C is significantly increased in *Alkbh3*$^{-/-}$ cells. Our results suggested that ALKBH3 was involved in the demethylation of m^5C in RNA.

Table of Contents

Lists of Figures

Chapter 1

General Overview

1.1 Epitranscriptome

In the central dogma of molecular biology, RNA plays important roles in biological system through transmitting genetic information and regulating biological processes. Various types of RNAs function in the biological processes including ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), messenger RNAs (mRNAs), small nuclear RNAs (snRNAs), etc. These RNA species carry more than 100 types of structurally distinct post-transcriptional modifications [1]. According to their functions, the RNA modifications can be categorized into three main types: "(1) modifications that enforce certain RNA structures and tune RNA biogenesis which includes modifications on rRNA and snRNA; (2) modifications that expand the RNA vocabulary and refine molecular recognition, like the one at the decoding region in tRNA; (3) modifications that code dynamic regulatory information on top of the primary sequence, such as modifications on mRNA" [2].

Most investigations related to the function of RNA modifications are confined to tRNA and rRNA due to their high abundance (rRNA and tRNA comprise 80% and 15% of total RNA, respectively [3]) and as they are easy to obtain. The functions of RNA modifications were commonly regarded as fine-tuning the structure while being considered static and unalterable after their covalent attachment [4]. However, in 2010, it was first proposed that reversible RNA epigenetic modifications (Figure 1), as an

analogue of the reversible modifications occurring on DNA [5] and histone proteins [6], may be involved in gene regulation [7]. Additionally, the formation of N^6-methyladenosine in mRNA was found to be reversible and has been intensively studied [8-11]. However, our knowledge about this additional regulatory layer of biology between DNA and protein is still very limited and these initial studies have led to the birth of the field of RNA epigenetics.

Among the nearly 150 types of post-transcriptional modifications of RNA, one common group is the mono-methylated cytidine and adenosine, which include 5-methylcytidine (m^5C), 2′-*O*-methylcytidine (Cm), N^6-methyladenosine (m^6A) and 2′-*O*-methyladenosine (Am) [1]. In this dissertation, I will focus on these four mono-methylated ribonucleosides because of their high abundance in RNA species and because little is known about their functions. In the rest of this chapter, I will first summarize the present studies about the functions of these four RNA modifications in rRNA, tRNA and mRNA. Next I will briefly summarize the enzymes involved in the deposition of m^5C, Cm and Am, while primarily focusing on the enzymes involved in the deposition, recognition and demethylation of m^6A.

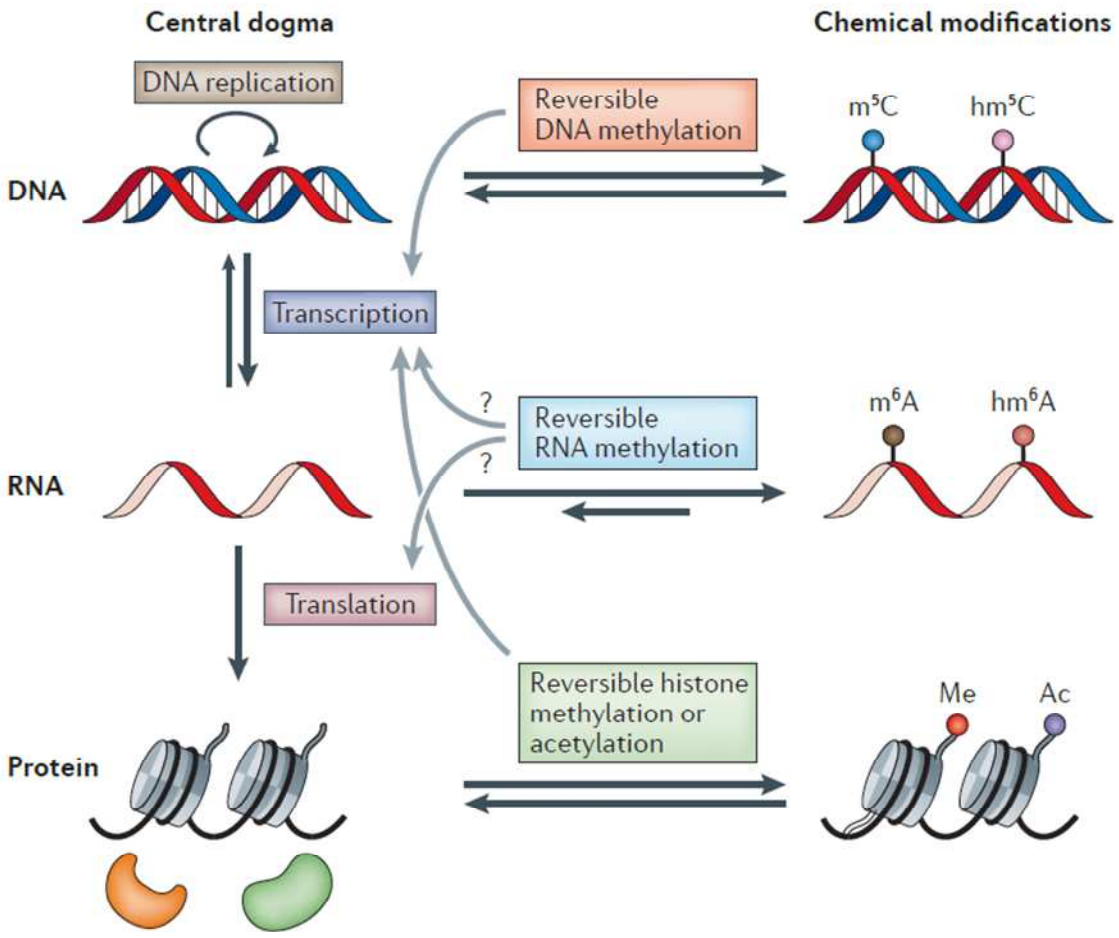

Figure 1.1 Reversible chemical modifications that regulate the flow of genetic information [4].

1.2 The functions of m^5C, Cm, m^6A and Am in RNA

1.2.1The functions of m^5C in RNA

5-methylcytosine (m^5C) is the most well-characterized epigenetic modification in genomic DNA and has served as a well-known epigenetic biomarker due to its important roles in various biological processes [12]. However, the regulatory functions of m^5C in RNA remain unclear, especially in mRNA.

A general role of nucleotide modification in tRNA is structural modulation and metabolic stabilization [13]. The m^5C site at position 40 of yeast $tRNA^{Phe}$ was reported to be involved in internal-site Mg^{2+} binding and tRNA secondary structure stabilization [14, 15]. It was also reported that the rapid degradation of pre-existing $tRNA^{Val(AAC)}$ accompanied by its de-aminoacylation was associated with the trm8 (N^7-methylguanine methyltransferase)-and trm4 (yeast tRNA methyltransferase 4)-mediated methylation on tRNA [16]. In addition, tRNA stability and protein synthesis were promoted by tRNA cytosine methylation mediated by two methyltransferases: TRDMT1 (tRNA aspartic acid methyltransferase 1, also known as DNMT2) and NSUN2 (tRNA and mRNA cytosine-C5-methyltransferase) [17]. Furthermore, emerging findings have shown that dynamic tRNA modifications can affect codon selection and translation outcome. Recent studies in yeast revealed that the level of m^5C on tRNA can dynamically respond to exposure from different toxic chemicals [18]. In response to the oxidative stress induced by hydrogen peroxide, there was an increase in m^5C level at position 34 of $tRNA^{Leu(\underline{C}AA)}$ (the first residue in the anticodon region) which enhanced the

translation of UUG-rich transcripts [19]. Several studies have shown that RNA methylation by DNMT2 protected tRNA against stress-induced cleavage, thereby facilitating the regulation of siRNA homeostasis during stress response [19, 20].

Most rRNA modifications are concentrated in the functional regions such as peptidyl transferase center. The m^5C in rRNA was generally thought to participate in tRNA recognition, fine-tuning the structure and peptidyl transfer [21]. The widespread presence and abundance of m^5C in mRNA and non-coding RNA (ncRNA) was confirmed recently and the sequencing data showed that m^5C sites are enriched in untranslated regions and near Argonaute binding sites [22]. The region of m^5C sites located in mRNA is the central component of miRNA/RISC complex, implicating that m^5C may be involved in miRNA degradation pathway [22]. Furthermore, m^5C sites were found to be present in two well-known long ncRNAs, HOTAIR and XIST, and the m^5C in these ncRNAs was shown to disturb RNA-protein interaction *in vitro* [23].

1.2.2 The functions of Cm and Am in RNA

Cm and Am belong to the 2′-*O*-methylation group which comprise a large number of post-transcriptional RNA modifications. The prevailing hypothesis of the biological functions of these modifications on rRNA is that they may play roles during ribosome biogenesis, as well as fine-tuning the structure and function of ribosome [2]. These modifications on tRNA also help to prevent the hydrolysis of the phosphodiester backbone and stabilize the 3′ endo conformation [24]. 2′-*O*-methylation also occurs on plant microRNA (miRNA) which has been suggested to serve as a protective

mechanism against 3′→5′ degradation and 3′-uridylation [25-27]. On the cap-structure of mRNA, the 2′-*O*-methylation was reported to be involved in discrimination of self and non-self mRNA dependent on the RNA sensor Mda 5 and type 1 interferon [28, 29].

1.2.3 The functions of m^6A in RNA

In 1974, m^6A was first discovered and reported to be the most prevalent internal modification on eukaryotic mRNA [30, 31]. However, investigations into the functions of m^6A have lagged for decades because of the low abundance of mRNA, leading to difficulty in detecting m^6A. In 2010, a breakthrough in the functional role of m^6A was made with the discovery and characterizations of m^6A writers, erasers and readers together with the development of high-throughput assays that profile this methylation on a transcriptome-wide scale [8, 10, 11, 32, 33]. Based on the current studies, the functions of m^6A are proposed to be related to mRNA processing, such as pre-mRNA splicing, mRNA stability, translation, turnover and nuclear export [34]. Two m^6A demethylases, FTO (Fat mass and obesity-associated proteins) and ALKBH5 (Probable α-ketoglutarate-dependent dioxygenase ABH5), were first reported in 2011 and 2013, respectively. FTO was found to partially co-localize with m^6A methyltransferase, MT-A70, in nuclear speckles and the *Alkbh5* knockout mice showed impaired fertility resulting from apoptosis that affects meiotic metaphase-stage spermatocytes, implicating that m^6A play roles in energy homeostasis and spermatogenesis [10, 33]. In addition, m^6A methyltransferases heterodimer-METTL3-METTL14 (human methyltransferase-like 3 and 14) core-complex was found to interact with a splicing

regulator-WTAP (Wilms' tumor 1-associating protein), suggesting the involvement of m^6A in pre-mRNA splicing [11]. Through the high-throughput assay, it was also found that silencing the m^6A methyltransferase significantly affects gene expression and alternative splicing patterns, resulting in modulation of the p53 signaling pathway and apoptosis [8].

Emerging studies showed that a main function of m^6A is likely to affect transcript stability. First, m^6A can be selectively recognized by the human YTH domain family 2 (YTHDF2) and the binding of YTHDF2 leads to the localization of the methylated mRNA from the translatable pool to mRNA decay site, such as processing bodies [32]. Second, statistical analysis showed that the methylation at the N^6-position of adenine generally associates with mRNA species with shorter half-lives [4]. In addition, two other recent studies on yeast also indicated that the YTH domain homologs in fission yeast and budding yeast, Mmi and YDR37C, bind with m^6A RNA and regulate the transcript stability [35, 36].

m^6A may also regulate other cellular process. It was found that knockdown of Mettl3 and Mettl14 (m^6A methyltransferase) in mouse embryonic stem cells (mESCs) led to loss of their self-renewal capability [37]. It was also reported that Mettl3 knockout naive embryonic stem cells failed to adequately terminate their naive state, and subsequently undergo aberrant and restricted lineage priming at the post-implantation stage, leading to early embryonic lethality [9]. These studies suggest that the m^6A on mRNA affects embryonic cell differentiation. Additionally, the

electrophoretic mobility shift assay (EMSA) results showed that YTHDF2 exhibited weaker binding toward RNA containing FTO-mediated oxidized derivatives of m^6A, N^6-hydroxymethyladenosine (hm^6A) and N^6-formyladenosine (f^6A), implicating the functions of m^6A in gene expression regulation [38].

1.3 RNA methyltransferase of m^5C, Cm, m^6A and Am

1.3.1 m^5C -methyltransferases (m^5C-MTases)

RNA m^5C-MTases use SAM (*S*-adenosyl-L-methionine) as a methyl group donor to form m^5C in RNA. The m^5C-MTases consist of a large number of proteins which can be divided into four major sub-families based on their structure-function relationship: Nop2/Nol1, YebU/Trm4, RsmB/Ynl022c and PH1991/NSUN6 [39]. Most known m^5C-MTases are functional in bacterial RNA and almost all of them have conserved homologues in most eukaryotic organisms. However, only two m^5C-MTases have been found to catalyze the formation of m^5C in higher eukaryotic RNA: NOP2/Sun domain protein 2 (NSUN2) and tRNA aspartic acid MTase 1 (TRDMT1). Since this dissertation focuses on mammalian RNA, I will primarily discuss these two methyltransferases.

NOP2/Sun domain protein 2 (NSUN2)

NSUN2 is the human homologue of yeast Trm4 (Multi-site-specific tRNA:(cytosine-C5-methyltransferase). It has been demonstrated that yeast Trm4 is responsible for complete m^5C methylation of yeast tRNA [40]. However, NSUN2 has narrower specificity on tRNA and selectively methylates the wobble position of

$tRNA^{Leu(CAA)}$ prior to intron splicing [41]. In 2012, the first m^5C profile in the entire human transcriptome showed that NSUN2 is also responsible for methylation in $tRNA^{Asp(GUC)}$ and two mRNA, *CINP* (cyclin-dependent kinase 2 interacting protein mRNA) and *NAPRT1* (nicotinate phosphoribosyl-transferase domain containing 1) [22]. Many studies have shown that NSUN2 had very important roles in cell cycle and cell differentiation. The expression level of NSUN2 is up-regulated upon Myc activation and present at the highest level in the S phase of the cell cycle [42]. The depletion of NSUN2 specifically blocked meiotic progression of germ cells into the pachytene stage [43] and RNA–NSUN2 protein complex is required for correct spindle assembly [44]. NSUN2 was also shown to stimulate a sub-population of stem cells to leave the hair bulge and become committed progenitor cells in the hair germ [45].

tRNA aspartic acid MTase 1 (TRDMT1)

TRDMT1 is also named as DNMT2 (DNA methyltransferase-2) due to its sequence homology to DNA m^5C-MTases [46]. However, TRDMT1 has been demonstrated to possess enzymatic activity in methylating cytosine 38 of $tRNA^{Asp-GTC}$, $tRNA^{Val-AAC}$ and $tRNA^{Gly-GCC}$ [47, 48]. Recently, two mRNAs, the KRT18 mRNA and the KRT18 pseudogene mRNA, displayed moderate enrichment with TRDMT1 immuno-precipitates in the 5-azacytidine–mediated RNA immunoprecipitation (Aza-IP) assay, implicating the possible methyltransferase activity of TRDMT1 on mRNA [49]. Furthermore, TRDMT1 has been implicated in stress-response. Drosophila DNMT2 mutants had reduced viability under stress conditions, and DNMT2 relocalized to stress

granules following heat shock [48]. In addition, there is some RNA-mediated non-mendelian inheritance of phenotypes occurring in mice which required DNMT2 expression [50, 51].

1.3.2 2′-*O*-methyltransferases (2′-*O*-MTase)

The 2′-*O*-methyltation in the cap-structure was catalyzed by 2′-*O*-methyltransferases, where SAM was used as the methyl donor. The 2′-*O*-MTase in viruses has been widely studied because of its importance of antiviral agents. The structure of 2′-*O*-MTase is conserved in all flavivirus, consisting of three motifs: (1) an N-terminal domain; (2) a core domain with a typical structure of SAM-dependent MTase; and (3) a C-terminal domain [52]. Another 2′-*O*-MTase which has been intensively investigated is HUA ENHANCER 1 (HEN1). HEN 1 was first identified to play a role in the specification of stamen and carpel identities during the flower development in *Arabidopsis thaliana* [53]. Subsequently, HEN1 was found to methylate the 2′-*O*-position at the 3′ termini of miRNAs and all types of siRNAs in plants [27, 54]. The human cap1 and cap2 2′-*O*-MTases, hMTr1 (also known as FTSJD2 and ISG95) and hMTr2 (also known as FTSJD1), were recently confirmed [55, 56]. The crystal structure of the active hMTr1 catalytic domain was then obtained and the mechanism of specific recognition of capped RNA was revealed to be significantly different from viral enzymes [57].

1.3.3 m^{6}A-methyltransferases

Initially, the proteins functioning to methylate the N^6-position of adenosine in mammalian RNA were characterized as a multi-component methyltransferase complex (~200 kDa) where METTL3 (a 70 kDa protein, also known as MT-A70) was the only known component [58]. Decades later, two other components, METTL14 and WTAP, were identified [11]. All three components are crucial to the methyltransferase activity. Individual knockdown of each component led to an over 30% decrease of m^{6}A levels in mRNA. METTL14 is a close homologue of METTL3 and the combination of these two proteins leads to substantially increased methylation activity *in vitro* [11]. WTAP was initially identified as a splicing factor and it is critical for cell cycle progression and early mammalian embryonic development [59]. Until now, the functional details about this multi-component methyltransferase complex are yet to be fully understood. Why does the deposition m^{6}A require two methyltransferases unlike one for other methylation sites? How does WTAP enhance the methylation activity of the complex? There are still many aspects awaiting further investigation.

1.4 m^{6}A readers and erasers in mammals

m^{6}A is the first known dynamic RNA modification. The enzymes involved in recognition and removal of this methylation have been intensely studied in recent years. Therefore, in this section, I will describe the readers and erasers for m^{6}A.

1.4.1 m^{6}A readers

Three m^{6}A-selective binding proteins were identified through an RNA affinity chromatography approach, including two YTH-family proteins, YTHDF2 and YTHDF3, and ELAVL1 (also known as HUR) [8]. Subsequently, YTHDF1 was identified as another m^{6}A-selective binding protein. It was revealed that YTHDF2 is the primary protein which most mRNA bind to and the binding sites were shown to localize around stop codons and at 3´UTRs with a conserved GAC[U>A] motif. The binding of YTHDF2 leads to the localization of the methylated mRNA from the translatable pool to mRNA decay site, such as processing bodies, suggesting a relationship between YTHDF2 and RNA degradation. This YTHDF2-mediated RNA degradation pathway is unique from other mRNA degradation pathways since it is dependent on the methylation of the target mRNA [32].

1.4.2 m^{6}A erasers

Two m^{6}A erasers, FTO and ALKBH5, have been discovered since 2011. Before that, the FTO gene was found to be associated with body mass index and the risk of obesity in multiple populations [60-62]. FTO also plays a role in development and it was reported that *Fto*-knockout mice have elevated postnatal lethality and retarded growth [63]. FTO was first reported to demethylate N^3-methylthymine in single-stranded DNA [64] and single-stranded RNA *in vitro* [65]. Subsequently, Jia et al. [10] discovered that FTO can demethylate m^{6}A in RNA and DNA *in vitro*. They also showed that the siRNA-mediated knockdown of FTO led to an elevated level of m^{6}A in

mRNA, while overexpression of FTO resulted in a decreased level of m^6A in mRNA. Since m^6A in genomic DNA is present at a very low level (a few parts per million), the authors concluded that m^6A in RNA is a major physiological substrate of FTO [10].

FTO belongs to the Fe(II) and α-ketoglutarate-dependent AlkB family of proteins [64]. The crystal structure of FTO revealed that it contains an amino-terminal AlkB-like domain which is required for the catalytic activity of the protein [66]. These findings prompted Zheng et al. [33] to biochemically test the catalytic activity of other human homologs of the AlkB family towards m^6A-carrying single-stranded RNA. They found that the ALKBH5 can completely demethylate m^6A. Furthermore, they identified the enzymatic activity of ALKBH5 *in vivo*. ALKBH5 knockdown in human cells led to an increased level of m^6A in mRNA and accelerated export of these mRNA from the nucleus to the cytoplasm. In addition, they found that the expressed level of ALKBH5 was the highest in mouse testes and *Alkbh*-5 knockout male mice displayed impaired fertility [33].

Subsequently, the He group reported that two intermediates, N^6-hydroxymethyladenosine (hm^6A) and N^6-formyladenosine (f^6A), were generated through the FTO-catalyzed oxidation pathway of m^6A but not through the ALKBH5-catalyzed oxidation pathway [38]. These two FTO-mediated derivatives were shown to impact protein-RNA interaction, implicating the role of m^6A in dynamically tuning the function or status of mRNA [38].

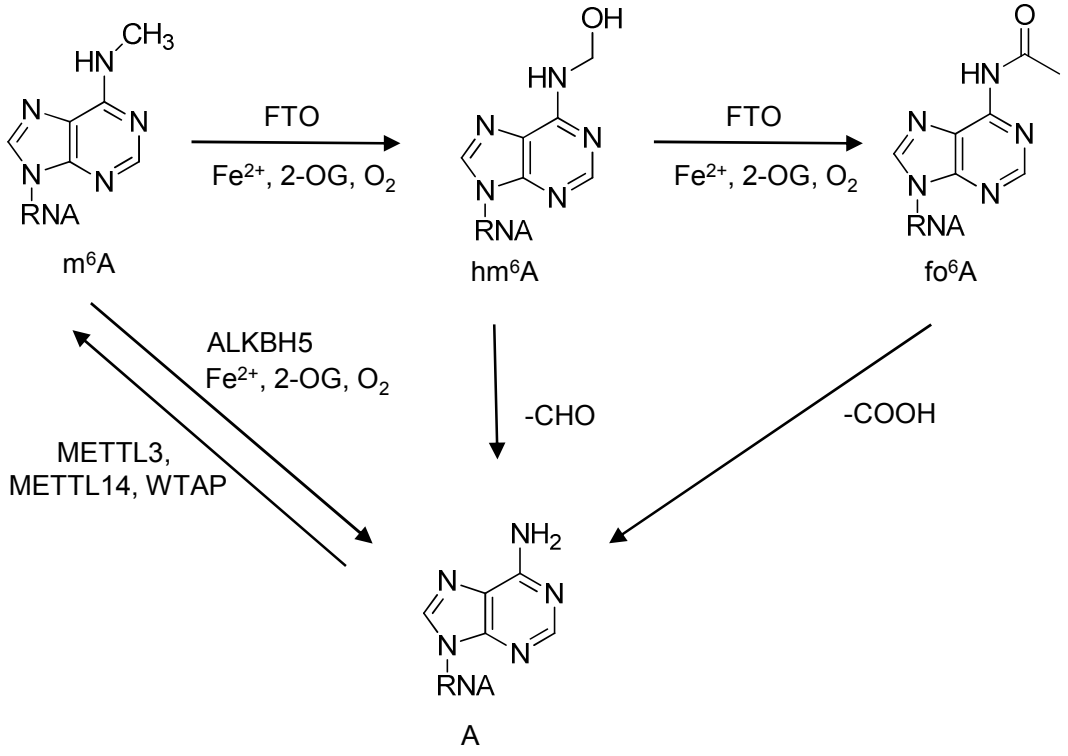

Figure 1.2 FTO and ALKBH5 remove the methyl group of m^6A by oxidation. FTO oxidizes m^6A to form metastable products of N^6-hydroxymethyladenosine (hm^6A) and N^6-formyladenosine (f^6A), which decompose back to adenosine.

1.5 Fe(II)- and 2-oxoglutarate-dependent dioxygenases

The Fe(II)- and 2-oxoglutarate-dependent dioxygenases (2-ODDs) are non-heme proteins belonging to a large super-family which can perform hydroxylation or demethylation of their substrates through oxidation. Recently, two types of 2-ODDs have attracted a lot of attentions, i.e. ALKBH-family proteins (ALKBH1-8 and FTO) and ten-eleven translocation family of Fe(II)- and 2-oxoglutarate-dependent dioxygenases (Tet 1-3).

1.5.1 ALKBH-family proteins

The *E. coli* AlkB protein belongs to the superfamily of α-KG- and Fe(II)-dependent dioxygenases [67, 68]. Until now, nine human homologues, ABH1-8 and FTO, have been identified [64, 69-71]. Many studies have shown that the *E. coli* AlkB protein has an impressive repertoire of substrates, including both DNA and RNA adducts. The DNA substrates of AlkB protein include alkyl adducts on the N^2 position of guanine, N^4 position of cytosine [72], N^6 position of adenine [73], as well as 1-methyladenine (1-meA) [68], 3-methylcytosine (3-meC) [74], 1-methylguanine (1-meG) and 3-methylthymine (3-meT) [75]. The RNA targets of AlkB protein include 1-meA and 3-meC [76]. However, only five human homologues are confirmed to have similar repair functions to AlkB on DNA and RNA. In addition to FTO and ALKBH5, ALKBH2, ALKBH3 and ALKBH8 are another three proteins [10, 33, 76-79]. Particularly, ALKBH2 prefers double-stranded DNA (dsDNA) substrates over single-stranded DNA (ssDNA) ones and was found to be primarily responsible for repairing 1-

meA base lesions in genomic DNA [80], whereas both hALKBH3 and AlkB are more active with ssDNA and single-stranded RNA (ssRNA) substrates [76, 81]. Recently, the X-ray crystal structure of ALKBH2-dsDNA complex was solved, and the structure revealed that the ALKBH2 protein adopts a commonly observed base-flipping mechanism with a finger residue which intercalates inside the DNA duplex to fill the gap left by the flipped base. Meanwhile, the X-ray crystal structure of AlkB-dsDNA complex showed that the AlkB protein squeezes the DNA duplex to eliminate the gap left by base flipping. This distortion imposed by AlkB on DNA explains its preference to flexible ssDNA over relatively rigid duplex DNA. These studies elucidated the mechanisms underlying the substrate preferences of these proteins [82].

ALKBH8 protein is the last known human homolog of *E. coli* AlkB which contains dioxygenase function on nucleotides. This protein was found to catalyze the hydroxylation of 5-methoxycarbonylmethyluridine (mcm^5U) at the wobble position of tRNA both *in vitro* and *in vivo* [78, 79]. Interestingly, ALKBH8 specifically hydroxylates mcm^5U into (*S*)-$mchm^5U$ in $tRNA^{Gly(UCC)}$ *in vivo* [79]. Apart from its dioxygenase activity, ALKBH8 was also reported to possess methyltransferase activity. The endogenous levels of mcm^5U in RNA were reduced in ALKBH8-depleted human cells [83]. In addition, a functional tRNA methyltransferase of ALKBH8 required the interaction of ALKBH8 with a small accessory protein, TRM112 [84]. Both methyltransferase and oxygenase activities are absent in the *Alkbh8*-/- mice [79].

For the other members of human homologues of AlkB family, ALKBH1 is a histone dioxygenase that acts specifically on histone H2A [85] and ALKBH4 remove a mono-methylation in actin (K84me1) [86].

1.5.2 TET (ten-eleven translocation) family proteins (Tet 1-3)

TET family proteins were discovered through iterative sequence profile searches using the predicted oxygenase domains of JBP1 and JBP2, enzymes of the 2OG- and Fe(II)-dependent dioxygenase superfamily which catalyze the formation of base J (β-D-glucosylhydroxymethyluracil) in trypanosomes [87]. Meanwhile, TET proteins were characterized to catalyze the oxidation of the methyl group on m^5C in DNA to form 5-hmC (5-hydroxymethylcytosine) [87]. Subsequently, it was shown that TET could mediate the reversal of DNA m^5C methylation through iterative oxidation of m^5C to 5hmC, 5-formylcytosine (5fC), and 5-carboxylcytosine (5-caC), as shown in Figure 3 [88, 89]. The oxidative derivatives, 5fC and 5caC, can be recognized and cleaved by thymine DNA glycosylase (TDG) and the resulting abasic site can be restored to normal cytosine through the base-excision repair (BER) pathway, which results in the active DNA demethylation of m^5C [89-91]. Alternatively, it was proposed that 5-hmC can be deaminated to generate 5-hmU by AID (activation-induced deaminase)/APOBEC (apolipoprotein B mRNA-editing enzyme complex) families of cytidine deaminase [92, 93]. The 5-hmU but not 5-hmC can be readily recognized and cleaved by TDG then replaced with unmethylated cytosine through BER pathway [91, 93, 94].

The crystal structures of a catalytically active truncated human TET2 and a Naegleria Tet-like dioxygenase have been reported [95, 96]. The results showed that the catalytic domain of TET proteins harbors three components, i.e., the double-stranded β-helix (DSBH) fold which is similar to all characterized Fe(II)/α-KG-dependent dioxygenases, a conserved iron-binding motif and a unique cysteine-rich domain [5, 95, 96]. Based on the similar structure and dioxygenase function between TET proteins and the ALKBH family proteins, it is highly possible that the demethylase activity of TET protein may also function on RNA methylation, especially for m^5C in RNA.

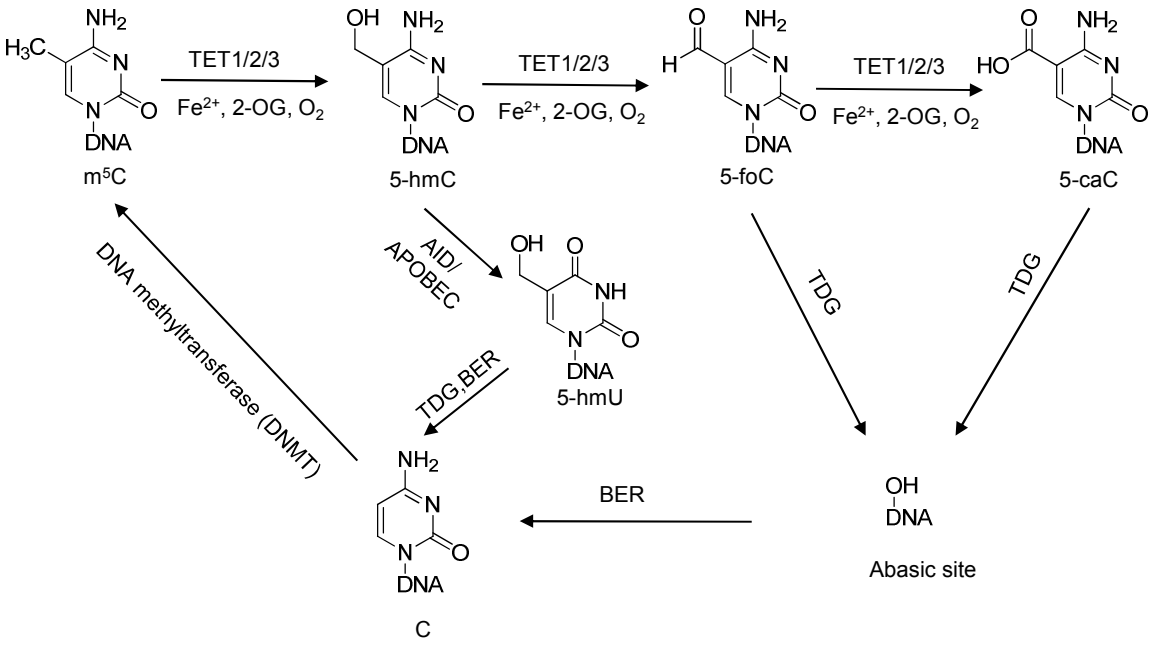

Figure 1.3 TET1, TET2, and TET3 can oxidize DNA 5-methylcytosine (m^5C) to the chemically stable intermediates 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC) in a stepwise manner. 5-foC and 5-caC can be recognized and cleaved by TDG to form an abasic site which can be restored by BER pathway to unmethylated cytosine. Alternatively, 5-hmC can be deaminated by AID/APOBEC deaminase to produce 5-hmU, which can be repaired to unmethylated cytosine by TDG and BER pathway.

1.6 Mass spectrometry-based detection of RNA modifications

Mass spectrometry has been widely used for quantifying DNA modifications, yet it has been much less frequently used for the quantitation of RNA modifications. Therefore, I will mainly use the examples on DNA modifications to introduce the application of mass spectrometry.

LC-MS/MS coupled with the use of stable isotope-labeled internal standards (isotope dilution LC-MS/MS) is considered as a golden standard for DNA adduct quantitation [97]. The method has several superior advantages. First, the measured levels of the analytes are not affected by alterations in sample matrices or LC-MS/MS conditions because of the addition of stable isotope-labeled standards to the nucleoside mixture. Moreover, the analytes and corresponding internal standards are analyzed simultaneously by LC-MS/MS under identical conditions. Any variations in experimental conditions after enzymatic digestion and during LC-MS/MS analysis do not affect the analytical accuracy. Second, this method allows for the unambiguous identification of each analyte. Both the analytes and their isotope-labeled standards co-elute and yield the same fragmentation patterns, thereby offering unequivocal chemical specificity for analyte identification. Among the various types of mass analyzers, triple quadrupole is the most commonly used for DNA adduct analysis because of its relatively low cost, compact size, efficient ion transmission, excellent sensitivity, and the ease of operation. In addition, 3-dimensional ion trap and linear ion trap are also commonly used. Unlike triple quadrupole, these two types of ion-trap mass analyzers

can perform multi-stage tandem mass spectrometry experiments (MS^n) which can provide additional structural information [97]. The LC-MS/MS coupled with the isotope-dilution technique has been applied to limited studies for the quantification of RNA modifications. Brucke et al. [98] employed an isotope-based mass-spectrometric method to quantitatively measure the levels of 11 tRNA modifications in tumor cells and non-tumorigenic tissues. Another example is that the D_3-m^6A was added in the digestion mixture to quantify the level of m^6A in mRNA isolated from human cells [9, 11].

Novel approaches have been developed to improve the sensitivity of quantifying DNA adducts *in vivo*, i.e. capillary HPLC coupled with nanoelectrospray ionization MS (capLC-nanoESI) and nanoflow HPLC-nanospray ionization mass spectrometry (NanoLC-nanoESI). The capLC-nanoESI involves using a capillary HPLC column (0.3-0.8 mm I.D.) and flow rates of 5-10 μL/min with a flow splitter and a nanoelectrospray ion source. The LOD can reach a few amol or a few tens of adducts per 10^9 nucleosides in matrix [97, 99]. The NanoLC-nanoESI uses nanobore HPLC columns (0.025-0.1 mm I.D.) with a nanoelectrospray and the flow rate was used at 100-500 nL/min. The sensitivity with the use of this method was increased by more than 10 times compared to the one using capLC-nanoESI. The LOD can reach to lower than 1 amol for some adducts and lower than 5 adducts per 10^9 nucleosides in matrix can be detected [97, 100]. Until now, there is no reported application of capLC-nanoESI or NanoLC-nanoESI for analyzing RNA modifications. The superior sensitivity of these methods should have great advantages on the detection and analyses of RNA modifications

which are present in the RNA species of low abundance, such as mRNA and small RNA.

Mass spectrometry can also be applied to sequence oligodeoxyribonucleotides (ODNs) and to locate the modification sites in ODNs. In MS/MS, the major cleavages of ODNs occur at the *N*-glycoside bond and 3′ C-O bond to afford the formation of [a_n-base] and w_n ions (Figure 1.4) [101]. The sequence of ODNs can be determined by comparing the mass difference between neighboring w_n or [a_n-base] ions, which allows for the identification of the sandwiched nucleotide. Compared to the ODN sequencing, the sequencing of oligoribonucleotides by MS/MS is more challenging because, aside from the w_n and [a_n-base] ions, y_n, [d_n-H_2O], and c_n ions are also generated [102-104]. Nevertheless, the MS-based sequencing method can still provide adequate information about the sequence of RNA.

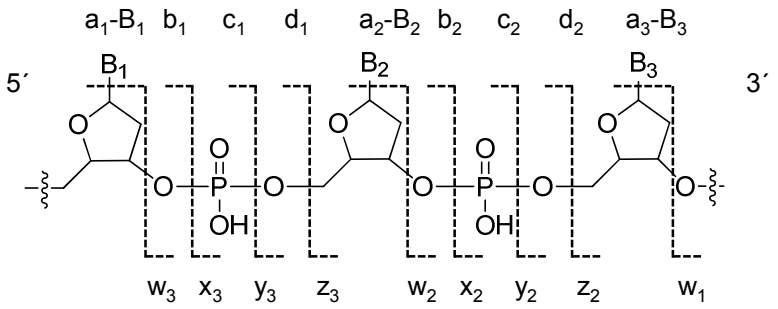

Figure 1.4 The nomenclature for fragment ions observed for oligodeoxynucleotides.

1.6 Scope of this dissertation

Mass spectrometry has become a very powerful tool in bioanalysis, which can allow for structure elucidation of compounds and provide quantitative measurements. The LC-MS-based analytical method, in combination with genetic manipulation, may facilitate the future discovery of proteins involved in the maintenance and regulation of these RNA modifications. In this thesis, I focus on the development of novel MS-based strategies to identify and quantify post-transcriptional modifications present in total RNA and mRNA isolated from mammalian tissues and cultured human cells. Additionally, by using these analytical methods, I was able to discover new enzymes involved in metabolism of mono-methylated m^5C both *in vitro* and *in vivo*.

In Chapter 2, an LC-MS/MS/MS coupled with the stable isotope-dilution method was developed for the sensitive and accurate quantifications of 5-methylcytidine (m^5C), 2′-*O*-methylcytidine (Cm), N^6-methyladenosine (m^6A) and 2′-*O*-methyladenosine (Am) in RNA isolated from mammalian cells and tissues. In Chapter 3, I first demonstrated that Tet enzymes can catalyze the formation of 5-hydroxymethylcytidine (5-hmrC) from m^5C *in vitro* by using a biochemical assay. Subsequently, I established a sensitive and accurate LC-MS/MS/MS with the isotope-dilution method for measuring the levels of 5-hmrC *in vivo* and further demonstrated that the catalytic domains of all three Tet enzymes as well as full-length Tet3 could induce the formation of 5-hmrC in human cells. In Chapter 4, I selected three Fe(Π)- and 2-oxoglutarate-dependent dioxygenases, FTO, ALKBH5 and ALKBH3, and tested

their demethylase activity towards m^5C in RNA. The results showed that, ALKBH3 may constitute another enzyme for the demethylation of m^5C in RNA.

References

1. Machnicka, M.A., et al., MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res, 2013. 41(Database issue): p. D262-7.

2. Wang, X. and C. He, Dynamic RNA modifications in posttranscriptional regulation. Mol Cell, 2014. 56(1): p. 5-12.

3. Lodish H, B.A., Zipursky SL, et al., Molecular Cell Biology. 4th edition. Section 11.6 Processing of rRNA and tRNA. W. H. Freeman and Company, 2000: p. section 11.6.

4. Fu, Y., et al., Gene expression regulation mediated through reversible m(6)A RNA methylation. Nat Rev Genet, 2014. 15(5): p. 293-306.

5. Shen, L., et al., Mechanism and function of oxidative reversal of DNA and RNA methylation. Annu Rev Biochem, 2014. 83: p. 585-614.

6. Shi, Y., Histone lysine demethylases: emerging roles in development, physiology and disease. Nat Rev Genet, 2007. 8(11): p. 829-33.

7. He, C., Grand challenge commentary: RNA epigenetics? Nat Chem Biol, 2010. 6(12): p. 863-5.

8. Dominissini, D., et al., Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature, 2012. 485(7397): p. 201-6.

9. Geula, S., et al., m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science, 2015.

10. Jia, G., et al., N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol, 2011. 7(12): p. 885-7.

11. Liu, J., et al., A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol, 2014. 10(2): p. 93-5.

12. Suzuki, M.M. and A. Bird, DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet, 2008. 9(6): p. 465-76.

13. Helm, M., Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res, 2006. 34(2): p. 721-33.

14. Chen, Y., et al., 5-Methylcytidine is required for cooperative binding of Mg2+ and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA. Biochemistry, 1993. 32(38): p. 10249-53.

15. Dao, V., R.H. Guenther, and P.F. Agris, The role of 5-methylcytidine in the anticodon arm of yeast tRNA(Phe): site-specific Mg2+ binding and coupled conformational transition in DNA analogs. Biochemistry, 1992. 31(45): p. 11012-9.

16. Alexandrov, A., et al., Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell, 2006. 21(1): p. 87-96.

17. Tuorto, F., et al., RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol, 2012. 19(9): p. 900-5.

18. Chan, C.T., et al., A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet, 2010. 6(12): p. e1001247.

19. Chan, C.T., et al., Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun, 2012. 3: p. 937.

20. Durdevic, Z., et al., The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila. Cell Rep, 2013. 4(5): p. 931-7.

21. Chow, C.S., T.N. Lamichhane, and S.K. Mahto, Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. ACS Chem Biol, 2007. 2(9): p. 610-9.

22. Squires, J.E., et al., Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res, 2012. 40(11): p. 5023-33.

23. Amort, T., et al., Long non-coding RNAs as targets for cytosine methylation. RNA Biol, 2013. 10(6): p. 1003-8.

24. Kawai, G., et al., Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. Biochemistry, 1992. 31(4): p. 1040-6.

25. Kamminga, L.M., et al., Hen1 is required for oocyte development and piRNA stability in zebrafish. EMBO J, 2010. 29(21): p. 3688-700.

26. Li, J., et al., Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol, 2005. 15(16): p. 1501-7.

27. Yu, B., et al., Methylation as a crucial step in plant microRNA biogenesis. Science, 2005. 307(5711): p. 932-5.

28. Daffis, S., et al., 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature, 2010. 468(7322): p. 452-6.

29. Zust, R., et al., Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat Immunol, 2011. 12(2): p. 137-43.

30. Desrosiers, R., K. Friderici, and F. Rottman, Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci U S A, 1974. 71(10): p. 3971-5.

31. Desrosiers, R.C., K.H. Friderici, and F.M. Rottman, Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5' terminus. Biochemistry, 1975. 14(20): p. 4367-74.

32. Wang, X., et al., N6-methyladenosine-dependent regulation of messenger RNA stability. Nature, 2014. 505(7481): p. 117-20.

33. Zheng, G., et al., ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell, 2013. 49(1): p. 18-29.

34. Jia, G., Y. Fu, and C. He, Reversible RNA adenosine methylation in biological regulation. Trends Genet, 2013. 29(2): p. 108-15.

35. Harigaya, Y., et al., Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature, 2006. 442(7098): p. 45-50.

36. Kang, H.J., et al., A novel protein, Pho92, has a conserved YTH domain and regulates phosphate metabolism by decreasing the mRNA stability of PHO4 in Saccharomyces cerevisiae. Biochem J, 2014. 457(3): p. 391-400.

37. Wang, Y., et al., N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat Cell Biol, 2014. 16(2): p. 191-8.

38. Fu, Y., et al., FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. Nat Commun, 2013. 4: p. 1798.

39. Motorin, Y., F. Lyko, and M. Helm, 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. Nucleic Acids Res, 2010. 38(5): p. 1415-30.

40. Motorin, Y. and H. Grosjean, Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: identification of the gene and substrate specificity of the enzyme. RNA, 1999. 5(8): p. 1105-18.

41. Brzezicha, B., et al., Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). Nucleic Acids Res, 2006. 34(20): p. 6034-43.

42. Frye, M. and F.M. Watt, The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. Curr Biol, 2006. 16(10): p. 971-81.

43. Hussain, S., et al., The mouse cytosine-5 RNA methyltransferase NSun2 is a component of the chromatoid body and required for testis differentiation. Mol Cell Biol, 2013. 33(8): p. 1561-70.

44. Hussain, S., et al., The nucleolar RNA methyltransferase Misu (NSun2) is required for mitotic spindle stability. J Cell Biol, 2009. 186(1): p. 27-40.

45. Blanco, S., et al., The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate. PLoS Genet, 2011. 7(12): p. e1002403.

46. Wilkinson, C.R., et al., The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. Nucleic Acids Res, 1995. 23(2): p. 203-10.

47. Goll, M.G., et al., Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science, 2006. 311(5759): p. 395-8.

48. Schaefer, M., et al., RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev, 2010. 24(15): p. 1590-5.

49. Khoddami, V. and B.R. Cairns, Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat Biotechnol, 2013. 31(5): p. 458-64.

50. Kiani, J., et al., RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2. PLoS Genet, 2013. 9(5): p. e1003498.

51. Rassoulzadegan, M., et al., RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature, 2006. 441(7092): p. 469-74.

52. Dong, H., et al., Flavivirus RNA methylation. J Gen Virol, 2014. 95(Pt 4): p. 763-78.

53. Chen, X., et al., HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. Development, 2002. 129(5): p. 1085-94.

54. Yang, Z., et al., HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Res, 2006. 34(2): p. 667-75.

55. Werner, M., et al., 2'-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res, 2011. 39(11): p. 4756-68.

56. Belanger, F., et al., Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. J Biol Chem, 2010. 285(43): p. 33037-44.

57. Smietanski, M., et al., Structural analysis of human 2'-O-ribose methyltransferases involved in mRNA cap structure formation. Nat Commun, 2014. 5: p. 3004.

58. Bokar, J.A., et al., Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA, 1997. 3(11): p. 1233-47.

59. Little, N.A., N.D. Hastie, and R.C. Davies, Identification of WTAP, a novel Wilms' tumour 1-associating protein. Hum Mol Genet, 2000. 9(15): p. 2231-9.

60. Scuteri, A., et al., Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet, 2007. 3(7): p. e115.

61. Dina, C., et al., Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet, 2007. 39(6): p. 724-6.

62. Frayling, T.M., et al., A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science, 2007. 316(5826): p. 889-94.

63. Fischer, J., et al., Inactivation of the Fto gene protects from obesity. Nature, 2009. 458(7240): p. 894-8.

64. Gerken, T., et al., The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science, 2007. 318(5855): p. 1469-72.

65. Jia, G., et al., Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. FEBS Lett, 2008. 582(23-24): p. 3313-9.

66. Han, Z., et al., Crystal structure of the FTO protein reveals basis for its substrate specificity. Nature, 2010. 464(7292): p. 1205-9.

67. Mishina, Y., E.M. Duguid, and C. He, Direct reversal of DNA alkylation damage. Chem Rev, 2006. 106(2): p. 215-32.

68. Falnes, P.O., R.F. Johansen, and E. Seeberg, AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature, 2002. 419(6903): p. 178-82.

69. Tsujikawa, K., et al., Expression and sub-cellular localization of human ABH family molecules. J Cell Mol Med, 2007. 11(5): p. 1105-16.

70. Kurowski, M.A., et al., Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. BMC Genomics, 2003. 4(1): p. 48.

71. Sanchez-Pulido, L. and M.A. Andrade-Navarro, The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. BMC Biochem, 2007. 8: p. 23.

72. Li, D., et al., Removal of N-alkyl modifications from N(2)-alkylguanine and N(4)-alkylcytosine in DNA by the adaptive response protein AlkB. Chem Res Toxicol, 2013. 26(8): p. 1182-7.

73. Li, D., et al., Exocyclic carbons adjacent to the N6 of adenine are targets for oxidation by the Escherichia coli adaptive response protein AlkB. J Am Chem Soc, 2012. 134(21): p. 8896-901.

74. Trewick, S.C., et al., Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature, 2002. 419(6903): p. 174-8.

75. Falnes, P.O., Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. Nucleic Acids Res, 2004. 32(21): p. 6260-7.

76. Aas, P.A., et al., Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature, 2003. 421(6925): p. 859-63.

77. Duncan, T., et al., Reversal of DNA alkylation damage by two human dioxygenases. Proc Natl Acad Sci U S A, 2002. 99(26): p. 16660-5.

78. Fu, Y., et al., The AlkB domain of mammalian ABH8 catalyzes hydroxylation of 5-methoxycarbonylmethyluridine at the wobble position of tRNA. Angew Chem Int Ed Engl, 2010. 49(47): p. 8885-8.

79. van den Born, E., et al., ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. Nat Commun, 2011. 2: p. 172.

80. Ringvoll, J., et al., Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. EMBO J, 2006. 25(10): p. 2189-98.

81. Falnes, P.O., et al., Substrate specificities of bacterial and human AlkB proteins. Nucleic Acids Res, 2004. 32(11): p. 3456-61.

82. Yang, C.G., et al., Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. Nature, 2008. 452(7190): p. 961-5.

83. Fu, D., et al., Human AlkB homolog ABH8 Is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. Mol Cell Biol, 2010. 30(10): p. 2449-59.

84. Songe-Moller, L., et al., Mammalian ALKBH8 possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding. Mol Cell Biol, 2010. 30(7): p. 1814-27.

85. Ougland, R., et al., ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. Stem Cells, 2012. 30(12): p. 2672-82.

86. Li, M.M., et al., ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. Nat Commun, 2013. 4: p. 1832.

87. Tahiliani, M., et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, 2009. 324(5929): p. 930-5.

88. Ito, S., et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science, 2011. 333(6047): p. 1300-3.

89. He, Y.F., et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science, 2011. 333(6047): p. 1303-7.

90. Zhang, L., et al., Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol, 2012. 8(4): p. 328-30.

91. Maiti, A. and A.C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem, 2011. 286(41): p. 35334-8.

92. Cortellino, S., et al., Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell, 2011. 146(1): p. 67-79.

93. Guo, J.U., et al., Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell, 2011. 145(3): p. 423-34.

94. Bennett, M.T., et al., Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. J Am Chem Soc, 2006. 128(38): p. 12510-9.

95. Hu, L., et al., Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell, 2013. 155(7): p. 1545-55.

96. Hashimoto, H., et al., Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature, 2014. 506(7488): p. 391-5.

97. Tretyakova, N., et al., Quantitation of DNA adducts by stable isotope dilution mass spectrometry. Chem Res Toxicol, 2012. 25(10): p. 2007-35.

98. Bruckl, T., et al., Parallel isotope-based quantification of modified tRNA nucleosides. Angew Chem Int Ed Engl, 2009. 48(42): p. 7932-4.

99. Liu, X., M.A. Lovell, and B.C. Lynn, Detection and quantification of endogenous cyclic DNA adducts derived from trans-4-hydroxy-2-nonenal in human brain tissue by isotope dilution capillary liquid chromatography nanoelectrospray tandem mass spectrometry. Chem Res Toxicol, 2006. 19(5): p. 710-8.

100. Goggin, M., et al., Molecular dosimetry of 1,2,3,4-diepoxybutane-induced DNA-DNA cross-links in B6C3F1 mice and F344 rats exposed to 1,3-butadiene by inhalation. Cancer Res, 2009. 69(6): p. 2479-86.

101. McLuckey, S.A., G.J. Van Berkel, and G.L. Glish, Tandem mass spectrometry of small, multiply charged oligonucleotides. J Am Soc Mass Spectrom, 1992. 3(1): p. 60-70.

102. Ni, J., et al., Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry. Anal Chem, 1996. 68(13): p. 1989-99.

103. Pomerantz, S.C. and J.A. McCloskey, Detection of the common RNA nucleoside pseudouridine in mixtures of oligonucleotides by mass spectrometry. Anal Chem, 2005. 77(15): p. 4687-97.

104. Huang, T.Y., J. Liu, and S.A. McLuckey, Top-down tandem mass spectrometry of tRNA via ion trap collision-induced dissociation. J Am Soc Mass Spectrom, 2010. 21(6): p. 890-8.

Chapter 2

Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography-Tandem Mass Spectrometry Coupled with the Stable Isotope-dilution Method

Introduction

More than 100 types of post-transcriptional modifications are known to be present in RNA and they play very important roles in the metabolic and regulatory processes of RNA. The biological functions of individual types of RNA modifications and their contributions to gene regulation remain largely unknown.[1] Among these RNA modifications, mono-methylated cytidine and adenosine, including 5-methylcytidine (m^5C), 2′-*O*-methylcytidine (Cm), N^6-methyladenosine (m^6A) and 2′-*O*-methyladenosine (Am) commonly occur in all RNA species.[2]

Previous investigations about the mono-methylated ribonucleosides have been mostly confined to transfer RNA (tRNA) and ribosomal RNA (rRNA), especially for cytidine modifications. It has been reported that m^5C contributes to the stabilization of secondary structures, codon recognition, and aminoacylation of tRNA.[3-5] In rRNA, m^5C sites have been thought to regulate translational fidelity and tRNA recognition.[6] Cm in tRNA was found to prevent the hydrolysis of the phosphodiester backbone,[7] whereas the Cm located in the cap structure of mRNA inhibits its 5′→3′ degradation[8] and distinguishes self from non-self RNA.[9]

Aside from their function in tRNA and rRNA, recent studies suggested that nucleobase methylations in mRNA may also play a very important role in gene regulation. In this vein, transcriptome-wide mapping studies have revealed a widespread occurrance of m^5C and m^6A in messenger RNA (mRNA) and non-coding RNA.[10-12] Sequencing data indicate that m^6A is localized around stop codons and present in both 3′-untranslated regions (3′-UTRs) and long internal exons, [10] whereas the m^5C sites are enriched in untranslated regions and near Argonaute protein binding sites.[12] Those studies suggested that m^6A may modulate pre-mRNA splicing, mRNA stability, translation, turnover and nuclear export, [13-15] whereas m^5C may play role in microRNA (miRNA)-mediated mRNA degradation and regulate the interactions of long non-coding RNA with chromatin-associated protein complexes.[12, 16]

The recent identification of enzymes fostering RNA methylation and demethylation highlights the importance of furthering our current understanding of the role of RNA methylation in gene regulation. Two mRNA demethylases, FTO (Fat mass and obesity-associated protein) and ALKBH5 (Alkylated DNA repair protein alkB homolog 5) were found to demethylate m^6A.[17, 18] Subsequently, human YTH domain family proteins (YTHDF1-3) were shown to bind to m^6A and affect the stability of m^6A-harboring mRNA.[19] Furthermore, the heterodimeric METTL3-METTL14 (human methyltransferase-like 3 and 14) core-complex was observed to deposit m^6A on mammalian nuclear RNAs.[20] In addition to these regulatory proteins of m^6A, the ten-eleven translocation family of Fe(II)- and 2-oxoglutarate-dependent dioxygenases 3 (Tet3)

can induce the formation of 5-hydroxymethylcytidine from m^5C in cellular RNA.[21] Additionally, TRDMT1 (tRNA aspartic acid methyltransferase 1) and NSUN2 (NOP2/Sun domain family, member 2) have been identified to be the cytosine-5-methyltransferase in tRNA and mRNA.[12, 22] Taken together, the identification and characterizations of proteins involved in the deposition, recognition and demethylation of m^6A and m^5C provide strong support for a reversible post-transcriptional modification pathway of RNA, which may constitute an important, yet under-appreciated mechanism of gene regulation. [23, 24]

To better exploit the mechanisms of RNA epigenetics, a robust analytical method is needed to assess the occurrence of these modifications in cellular RNA. Traditional methods for analyzing RNA modifications include ^{32}P-labelling and two-dimensional thin-layer chromatography,[25] dot-blot,[18] and capillary electrophoresis coupled with laser-induced fluorescence (CE-LIF) detection.[26] Apart from being tedious, semi-quantitative, and low-throughput, these methods require a large amount of RNA and are not compatible for the analysis of RNA species of low abundance. Recently, high-performance liquid chromatography coupled with a triple-quadrupole mass spectrometer, along with the use of external standards, was employed to measure m^6A in mRNA [17, 18] and other RNA modifications in tRNA and small RNA, and femtomole level of sensitivity was achieved.[27, 28] We reason that the application of stable isotope-labeled internal standards will offer unambiguous and accurate measurements of these post-transcriptional modifications in cellular RNA species. Herein, we developed an LC-MS/MS/MS coupled with the stable isotope-dilution method to achieve sensitive,

accurate and simultaneous quantifications of the global levels of the mono-methylated cytidine and adenosine in RNA. By using this method, we quantified the levels of m^5C, Cm, m^6A, and Am in total RNA isolated from cultured human cells and mammalian tissues, and in mRNA isolated from HEK293T human embryonic kidney cells.

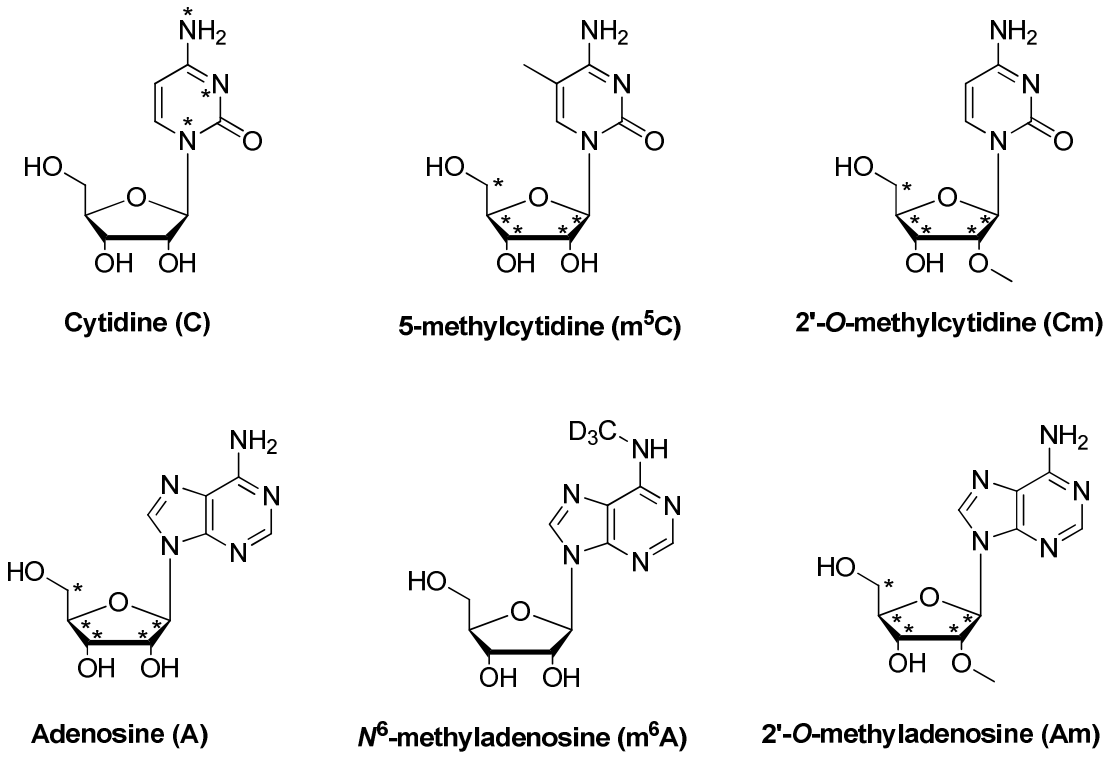

Figure 2.1 The chemical structures of the stable isotopic-labeled nucleosides. Asterisks (*) indicate the sites of 15N and 13C labelings; D = deuterium.

Experiment Section

Materials

All chemicals and enzymes, unless otherwise specified, were purchased from Sigma-Aldrich (St. Louis, MO) and New England Biolabs (Ipswich, WA). Erythro-9-(2-hydroxy-3-nonyl) adenine (EHNA) hydrochloride was purchased from Tocris Bioscience (Ellisville, MO). $^{15}N_3$-cytidine-5′-triphosphate was obtained from Sigma-Aldrich (St. Louis, MO) and all other stable isotope-labeled nucleoside starting materials were from Cambridge Isotope Laboratories (Tewksbury, MA). Mouse tissues were obtained from 19-21 week old mice. The HEK293T embryonic kidney cells, MCF7 human breast cancer cells, HCT116 human colon cancer cells, HeLa human cervical cancer cells, WM-266-4 human melanoma cells, and cell culture reagents were purchased from ATCC (Manassas, VA).

Syntheses of Stable Isotope-labeled Ribonucleosides

The stable isotope-labeled nucleosides employed in this study are shown in figure 2.1.

$^{15}N_3$-cytidine

$^{15}N_3$-cytidine-5′-triphosphate was treated with alkaline phosphatase in 50 mM Tris-HCl buffer (pH 8.9) at 37°C for 2 hrs. The enzyme was removed by chloroform extraction and the aqueous layer was dried in a Speed-vac. The resulting $^{15}N_3$-cytidine was purified by HPLC.

5-methyl-$^{13}C_5$-cytidine and $^{13}C_5$-adenosine

Ribose-$^{13}C_5$-cytidine (5.0 mg, 0.020 mmol) was fully acetylated by treating with acetic anhydride (40 μL, 0.409 mmol) at 60°C for 3 hrs in 1-mL anhydrous pyridine. The resulting crude tetra-acetylated ribose-$^{13}C_5$-cytidine was dissolved in anhydrous acetonitrile (1 mL) in the presence of 5-methyl-N^4-benzoylcytosine (7.4 mg, 0.032 mmol) or N^6-benzoyladenine (9.7 mg, 0.040 mmol) and stirred at room temperature for 10 min. Bis-trimethylsilylacetamide (20 μL, 0.070 mmol) was added and the reaction mixture was heated to 70°C. After stirring at 70°C for 15 min, TMS-triflate (4 μL, 0.020 mmol) was added to the reaction flask and the reaction was refluxed at 70°C for 4 hrs. The solvent was removed, and the resulting crude mixture was dissolved in 2 M ammonia in methanol (4 mL) and stirred at 40°C for 24 hrs. Subsequently, 30% ammonium hydroxide (4 mL) was added to the reaction mixture and stirred at 40°C for 48 hrs. The resulting crude 5-methyl-$^{13}C_5$-cytidine and $^{13}C_5$-adenosine were evaporated of solvent and purified by HPLC using a reverse-phase Alltima C18 column (5 μm in particle size, Grace Davison, Deerfield, IL). The purified 5-methyl-$^{13}C_5$-cytidine and $^{13}C_5$-adenosine were confirmed by LC-MS and MS/MS analysis (Figure 2.2). A portion of the purified $^{13}C_5$-adenosine was then used to synthesize 2′-*O*-methyl-$^{13}C_5$-adenosine.

2′-O-methyl-$^{13}C_5$-cytidine, 2′-O-methyl-$^{13}C_5$-adenosine and D_3-N^6-methyladenosine

2′-*O*-methyl-$^{13}C_5$-cytidine, 2′-*O*-methyl-$^{13}C_5$-adenosine were synthesized according to established procedures.[29] 6-Chloro-9-(β-D-ribofuranosyl)purine was

synthesized following published procedures [30] and then reacted with D_3-methylamine to yield D_3-N^6-methyladenosine.[31]

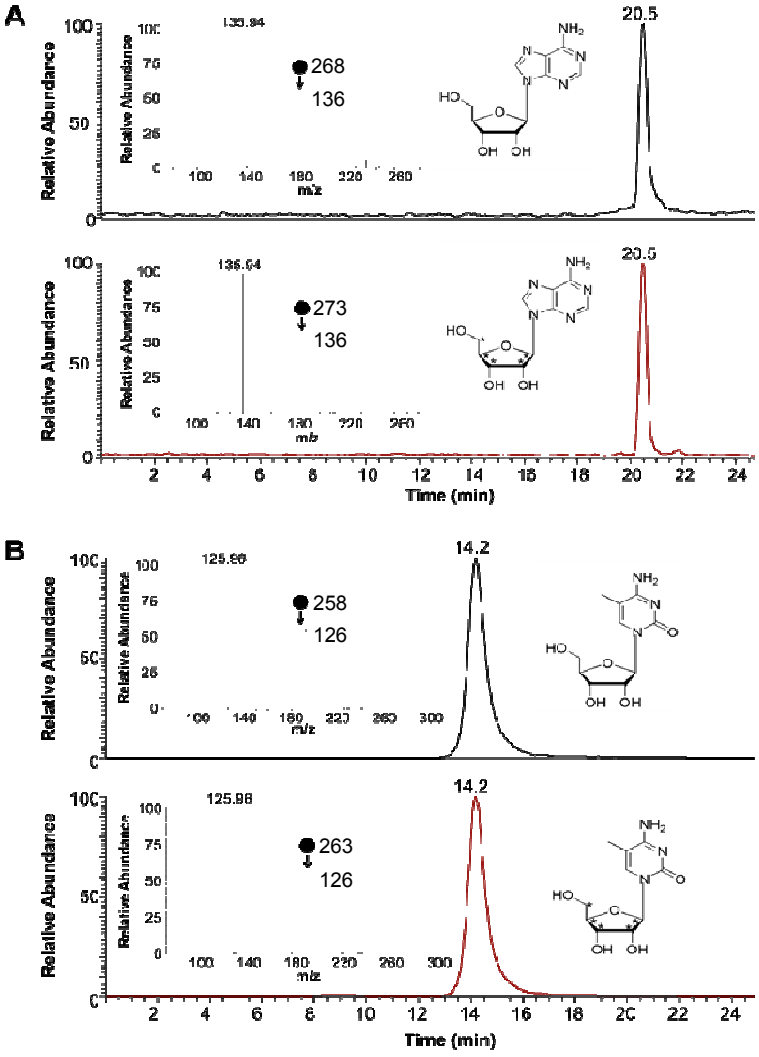

Figure 2.2. LC-MS/MS results for the analyses of unlabeled and purified stable isotope-labeled adenosine (A) and cytidine (B). Shown are the selective-ion chromatograms for monitoring the indicated transitions for the labeled and unlabeled nucleosides. The MS/MS are shown in the insert.

Isolation of total RNA and mRNA

Total RNA was isolated from mammalian cells and tissues using TRI Reagent® following the manufacturer's recommended procedures. The poly(A) messenger RNA (mRNA) was extracted using PolyATtract® mRNA Isolation Systems (Promega), immediately followed with the removal of rRNA contaminations by using RiboMinus Transcriptome Isolation Kit (Invitrogen). The mRNA concentrations were measured using UV spectrophotometry. The quality of mRNA was analyzed using an Agilent 2100 Bioanalyzer equipped with an RNA PicoChip.

Digestion of RNA

To 500 ng of RNA were added 0.05 unit of nuclease P1, 0.125 nmol of EHNA, 0.0000625 unit of phosphodiesterase 2 and 1 μL solution containing 300 mM sodium acetate (pH 5.6) and 10 mM zinc chloride. The EHNA was added to minimize the potential deamination of adenosine. Doubly distilled water (ddH_2O) was added to the reaction mixture to reach a final volume of 10 μL. The reaction mixture was incubated at 37°C for 4 hrs. To the resulting mixture were subsequently added 0.05 unit of alkaline phosphatase, 0.005 unit of phosphodiesterase 1, 1.5 μL of 0.5 M Tris-HCl buffer (pH 8.9) and ddH_2O to reach a final volume of 15 μL. After digestion at 37°C for 2 hrs, the resulting digestion mixture was dried using a Speed-vac and the dried residue was reconstituted in 500 μL of ddH_2O.

For the quantification of m^5C and Cm in total RNA, to a 5-μL aliquot of the digestion mixture of total RNA (5 ng) were added 25.5 fmol of 5-methyl-$^{13}C_5$-cytidine,

19.6 fmol of 2′-*O*-methyl-$^{13}C_5$-cytidine and 3.3 pmol of $^{15}N_3$-labeled cytidine. For the quantification of m^6A and Am in total RNA, to a 1-μL aliquot of total RNA (1 ng) were added 8.5 fmol D_3-N^6-methyladenosine, 6.9 fmol of 2′-*O*-methyl-$^{13}C_5$-adenosine and 1.555 pmol $^{13}C_5$-labeled adenosine. For the quantification of m^5C and Cm in mRNA, to a 10-μL aliquot of the digestion mixture of mRNA (10 ng) were added 5.1 fmol of 5-methyl-$^{13}C_5$-cytidine, 9.7 fmol of 2′-*O*-methyl-$^{13}C_5$-cytidine and 3.3 pmol of $^{15}N_3$-labeled cytidine. For the quantification of m^6A and Am in mRNA, 8.5 fmol D_3-N^6-methyladenosine, 3.45 fmol of 2′-*O*-methyl-$^{13}C_5$-adenosine and 1.555 pmol $^{13}C_5$-labeled adenosine were added to a 2-μL aliquot of the digestion mixture of mRNA. All enzymes used for RNA digestions were subsequently removed by chloroform extraction. The resulting aqueous layer was dried and reconstituted in ddH_2O. For the total RNA samples, ¼ of the total sample was used for nLC-MS^3 analysis, while for the mRNA, ½ of the total sample was subjected to nLC-MS^3 analysis.

LC-MS^3 Analyses of m^5C, Cm, m^6A and Am

LC-MS^3 measurements were conducted on an LTQ XL linear ion trap mass spectrometer equipped with a nanoelectrospray ionization source and coupled to an EASY-nLC II (Thermo Fisher Scientific, San Jose, CA, USA). The temperature for the ion transport tube of the mass spectrometer was maintained at 275°C. The instrument was operated in the positive-ion mode, the spray, capillary, and tube lens voltages were 2.0 kV, 12 V, and 100 V, respectively. The sensitivities for detecting the four mono-

methylated ribonucleosides were optimized by varying the normalized collision energy and activation Q of the LTQ mass spectrometer (Table 2.1).

For the measurements of m^5C and Cm, the samples were loaded onto a pre-column (150 μm × 70 mm) packed with porous graphitic carbon (PGC, 5 μm in particle size, Thermo Fisher Scientific). The samples were then eluted, at a flow rate of 2.5 μL/min, onto an in-house packed Zorbax SB-C18 column (75 μm × 250 mm, 5 μm beads, 100 Å in pore size, Agilent). A solution of 0.1% (v/v) formic acid in water (solution A) and a solution of 0.1% (v/v) formic acid in acetonitrile (solution B) were used as the mobile phases. The modified nucleosides were separated using a gradient of 0-5% B in 5 min, 5-15% B in 37 min, 15% B in 18 min, 15-90% B in 1 min, and finally at 90% B for 10 min. The flow rate was 300 nL/min.

The measurements of m^6A and Am were conducted in a similar way except that a 150 μm × 50 mm pre-column and a 75 μm × 150 mm analytical column, which were both packed with Magic AQ reversed-phase C18 resin (5 μm beads, 100 Å in pore size; Michrom BioResources, Auburn, CA, USA), were used. The modified nucleosides were separated using a gradient of 0-10% B in 50 min follow by 10% B for another10 min, and the flow rate was 300 nL/min.

compounds	SIM (MS^3 or MS^2) of compounds		Scan Width	Optimized LTQ parameters		LOD, amol	LOQ, amol
	unlabeled	Isotope-labeled		Normalized Collision Energy	Activation Q		
C	244→112	247→115	3	37	0.25		
	112→ 95	115→ 97	2	40	0.50		
m^5C	258→126	263→126	3	37	0.25	2.8±0.4	9.4±1.5
	126→108	126→108	2	40	0.50		
Cm	258→112	263→112	3	37	0.25	2.2±0.7	7.2±2.3
	112→ 95	112→ 95	2	40	0.50		
A	268→136	273→136	3	35	0.25		
m^6A	282→150	285→153	3	39	0.31	0.68±0.2	1.9±0.6
	150→ 94	153→ 94	2	35	0.37		
Am	282→136	287→136	3	39	0.31	1.0±0.4	3.4±1.2
	136→ 94	136→ 94	2	35	0.37		

Table 2.1 Optimized instrument conditions for LC-ESI-MS3 analysis of m^5C, Cm, m^6A and Am and detection limits of these modified nucleosides. The LOQ (limits of quantitation, which is defined as the amount of analyte that gives rise to a signal-to-noise ratio of 10) represents the means and standard deviations of the results from three measurements of the unlabeled standards. The LOD (limits of detection, which is defined as the amount of analyte that gives rise to a signal-to-noise ratio of 3) represents the means and standard deviations of the results from three measurements of the unlabeled standards. A constant activation time of 30 ms was employed for all the measurement.

Results

Nano-LC-MS/MS/MS analyses of m^5C, Cm, m^6A and Am

We set out to develop an nLC-MS/MS/MS in combination with the stable isotope-dilution method for the accurate assessment of the levels of m^5C, Cm, m^6A and Am in total RNA isolated from cultured cells and tissues. First, we examined the efficiencies of the pre-columns packed with various stationary phase materials including Zorbax SB-C18, Magic C18, Magic C18-AQ and porous graphitic carbon (PGC) in trapping the modified nucleosides. We found that m^5C and Cm could be retained very well on the PGC trapping column, but not on the other three types of packing materials. On the other hand, Magic C18-AQ displayed the most efficient trapping of m^6A and Am.

We next tested the performance of the analytical columns packed with the four types of stationary phase materials. Our results showed that, even though m^5C and Cm could be efficiently trapped on the PGC column, the use of PGC as the packing material for the analytical column yielded poor reproducibility and displayed severe issues with matrix interferences for the modified cytidines. On the other hand, the analytical columns packed with Magic C18 AQ or Zorbax SB-C18 exhibited excellent reproducibility and low matrix interferences when used with a slow gradient, with the Zorbax SB-C18 analytical column displaying slightly better performance overall. Therefore, we chose the combination of PGC trapping column with Zorbax SB-C18 analytical column for the analyses of m^5C and Cm. On the other hand, we found that the use of Magic C18 AQ as

the stationary phase material for both the trapping and analytical columns was the most suitable for measuring the two modified adenosine derivatives, i.e., m^6A and Am.

Upon collisional activation, the $[M+H]^+$ ions of the four unlabeled methylated ribonucleosides readily eliminate the ribose moiety to yield the protonated ions of the nucleobase portions (i.e., m/z 126, 112, 150 and 136 for m^5C, Cm, m^6A and Am, respectively). Further collisional activation of the ions of m/z 126 and 112 leads to the facile losses of NH_3 and H_2O, yielding the fragment ions of m/z 109 and 108 in the MS^3 for m^5C, and the fragment ions of m/z 95 and 94 in the MS^3 for Cm (Figure 2.3, b). On the other hand, collisional activation of the ion of m/z 150 results in the loss of $C_2H_4N_2$ and HCN to yield product ions of m/z 94 and 123 in the MS^3 of m^6A. Additionally, further collisional activation of the ions of m/z 136 of Am gave rise to the elimination of $C_2H_4N_2$ and NH_3, yielding ions of m/z of 94 and 119, respectively, in the MS^3 (Figure 2.3, d). The fragment ions of m/z 108, 95, 94 and 94 observed in the MS^3 of these modified nucleosides were chosen for the quantification of the levels of m^5C, Cm, m^6A and Am, respectively (see representative SICs in Figure 2.3, a & c). The nearly identical elution time and similar MS^3 spectra for the analytes and their stable isotope-labeled counterparts, permit for the unambiguous identification and reliable quantification of the four ribonucleosides in the digestion mixture of total RNA. Calibration curves for the quantifications of rC, m^5C, Cm, rA, m^6A, and Am are shown in Figure 2.6 & 2.7.

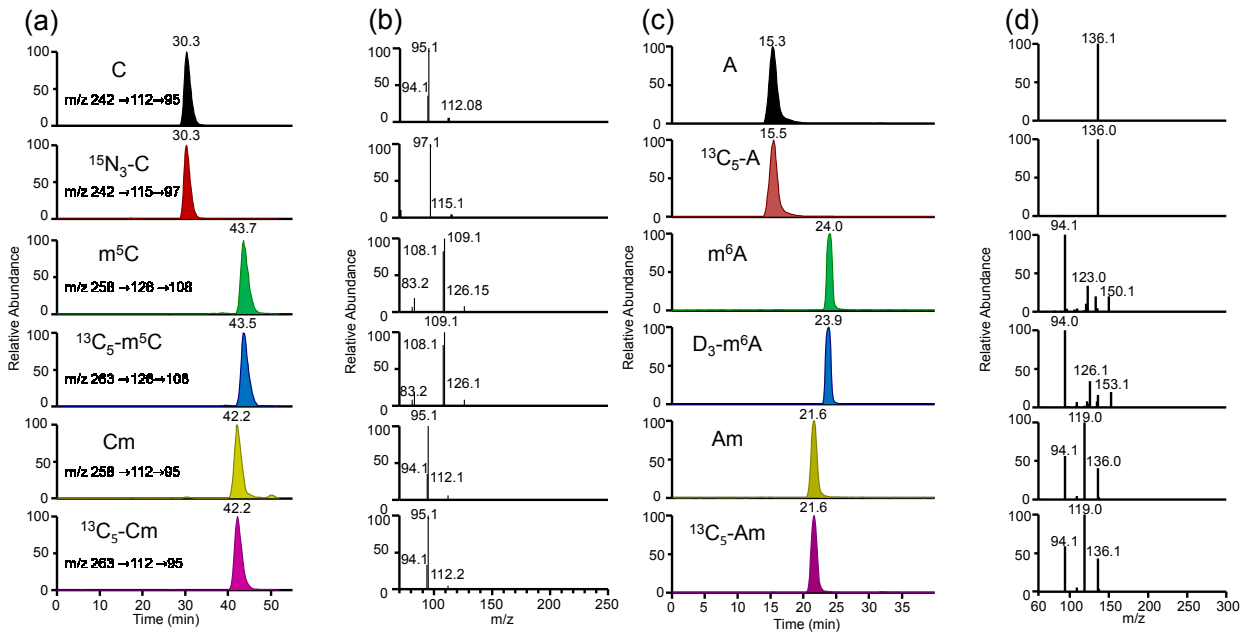

Figure 2.3 Representative LC-MS/MS/MS results for the quantifications of m^{5}C, Cm, m^{6}A and Am in mouse brain. Shown are the selective-ion chromatograms for monitoring the indicated transitions for the analytes and the stable isotope-labeled standards (a & c), and the corresponding MS/MS/MS for the analytes and internal standards (b & d).

Quantification of m^5C, Cm, m^6A and Am in total RNA isolated from mammalian tissues

We first assessed the levels of the four methylated ribonucleosides in total RNA isolated from different mouse tissues. Our results showed that the levels of m5C were 0.29, 0.41, 0.93, and 0.51 modifications per 100 cytidines in the total RNA isolated from mouse pancreas, spleen, heart, and brain tissues, respectively, while the corresponding levels of Cm were 0.66, 0.68, 0.62, and 0.66 modifications per 100 cytidines, respectively (Figure 2.4, a). In addition, the levels of Am (at levels of 1.98, 1.93, 0.97 and 1.24 modifications per 100 adenosines, respectively) were found to be significantly higher than those of m6A (at levels of 0.065, 0.061, 0.064 and 0.070 modifications per 100 adenosines, respectively. Figure 2.4, c). Interestingly, the levels of m5C are significantly higher in the heart than in other three types of mouse tissues. However, the levels of Cm and Am from the mouse heart are lower than those measured from the other three types of mouse tissues. Together, these results suggest that the distributions of these methylation ribonucleosides are tissue-specific.

Quantification of m^5C, Cm, m^6A and Am in total RNA of human cancer cells

To evaluate if the levels of m^5C, Cm and m^6A and Am vary among different cancer cells, we isolated total RNA from four different human cancer cell lines, digested them with enzymes and subjected the resulting digestion mixture to LC-MS^3 analyses. Our results showed that the levels of m^5C were 0.22, 0.34, 0.32, and 0.25 modifications per 100 cytidines in total RNA isolated from HeLa, WM-266-4, MCF7, and HCT116

cells, respectively, whereas the levels of Cm were consistently higher, at 0.57, 0.57, 0.56, and 0.55 modifications per 100 cytidines, respectively (Figure 2.4, b). Additionally, the levels of m^6A in total RNA isolated from these four cell lines were 0.059, 0.064, 0.070 and 0.065 modifications per 100 adenosines, respectively, whereas the levels of Am were 1.32, 1.38, 1.63 and 1.33 modifications per 100 adenosines, respectively (Figure 2.4, d). Thus, these results indicate that the levels of 2′-O-methylated rA and rC are significantly higher in all human cancer cell lines compared to their respective mono-methylated nucleobase modifications (m^6A and m^5C). Additionally, these results indicate that the levels of m^6A, Am, and Cm are similar among the human cancer cell lines, while the levels of m^5C appear to be cell-specific. Finally, to investigate the levels of these RNA modifications in cancer cells relative to healthy non-cancerous cells, we measured the levels of these four mono-methylated nucleosides in the total RNA isolated from HEK293T cells (Figure 2.5, a). Our results showed that the levels of m^6A, Am, and Cm are not significantly different between healthy and cancerous cell lines except the level of Am between HEK293T and MCF7 cells ($p=0.01$). Intriguingly, the level of m^5C in total RNA of HEK293T cells (0.28 modifications per 100 cytidines) is slightly higher than HeLa and HCT116 cells, but slightly lower than those observed for the WM-266-4 and MCF-7 cells. Taken together, these results suggest that the m^5C levels in total RNA may be cell-specific and altered in cancerous cells, whereas the levels of m^6A, Am, and Cm are similar in cultured cancer cells and HEK293T cells.

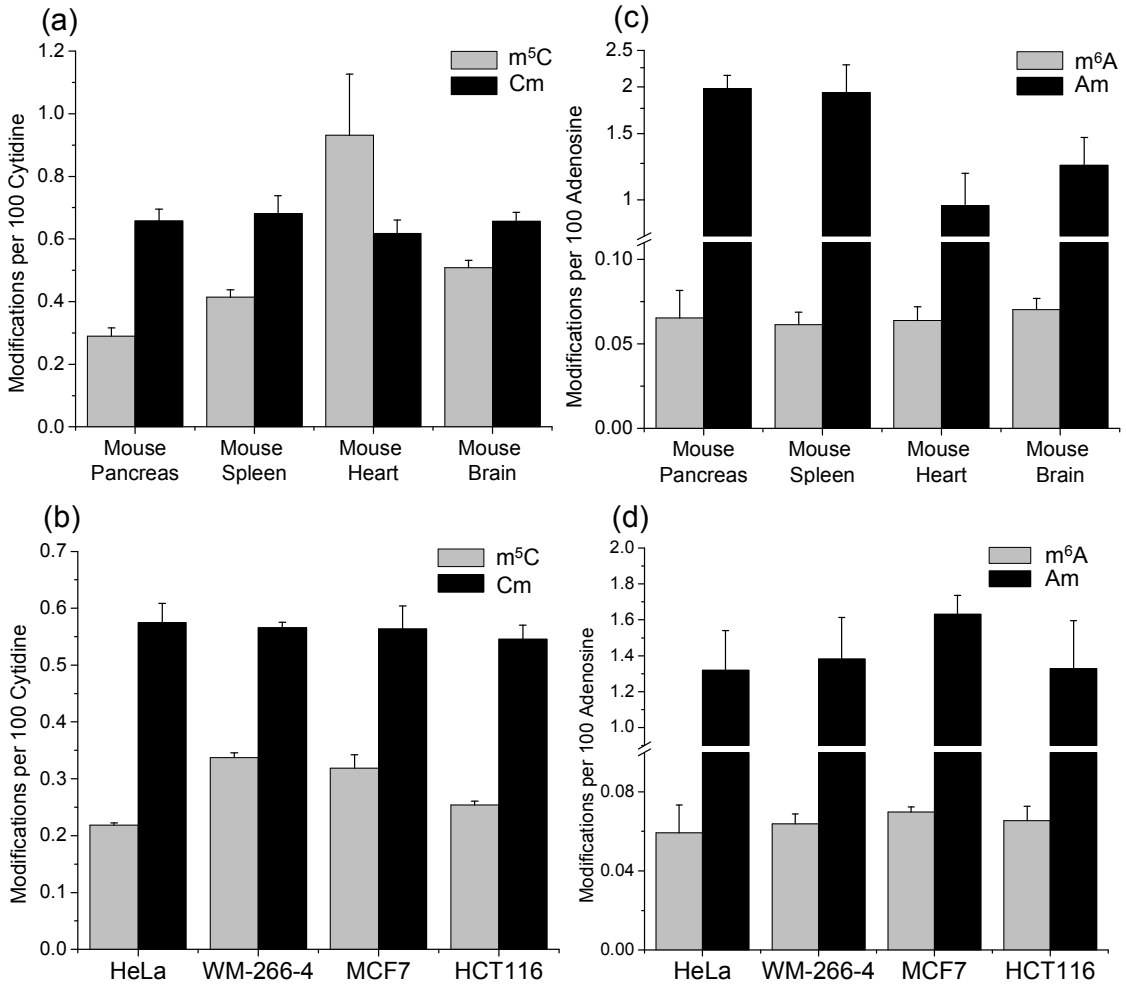

Figure 2.4 Quantification results for the levels of m^5C and Cm (a), m^6A and Am (c) in total RNA isolated from mouse tissues (n≥3). The tissue types include mouse pancreas, spleen, heart, brain. Quantification results for the levels of m^5C and Cm (b), m^6A and Am (d) in cultured cancer cells (n=3). The data represent the means and standard deviations of results from at least three separate mouse tissues or 3 individual RNA samples extracted from cultured human cells.

Quantification of m^5C, Cm, m^6A and Am in mRNA from HEK293T cells

Lastly, we compared the global levels of these four mono-methylated ribonucleosides in mRNA isolated from HEK293T cells. Our results showed that the levels of m^5C, Cm and Am in mRNA are lower than those in total RNA by16, 6.5, and 43 folds, respectively. However, the level of m^6A in mRNA is 1.6-fold higher than that in total RNA. This observation is consistent with previous reports showing that m^6A is the most abundant methylation product in mRNA.[32-34] However, our measured level of m^6A in mRNA from HEK293T cells (0.11 per 100 adenosines) is significantly lower than the previously reported level of m^6A (~ 0.4 per 100 adenosine).[18] This difference might be attributed to the methods through which the levels of m^6A were quantified. Here, we employed isotope-labeled internal standards for the quantification, which offers unambiguous identification and accurate measurement of the levels of the analyte. On the other hand, external standards were utilized in the previous report,[18] where the measured levels of the modified nucleosides could be potentially influenced by matrix effects, such as ion suppression during LC-MS analysis. Our quantification results also showed that the levels of m^5C, Cm, and Am in mRNA were 0.017, 0.086, and 0.024 modifications per 100 cytidines, respectively. These results demonstrate a higher level of Cm than m^5C in both total RNA and mRNA of HEK293T cells. However, the relative level of m^6A and Am displayed an opposite trend in total RNA and mRNA, with the level of m^6A being higher than Am in mRNA, and lower than Am in total RNA.

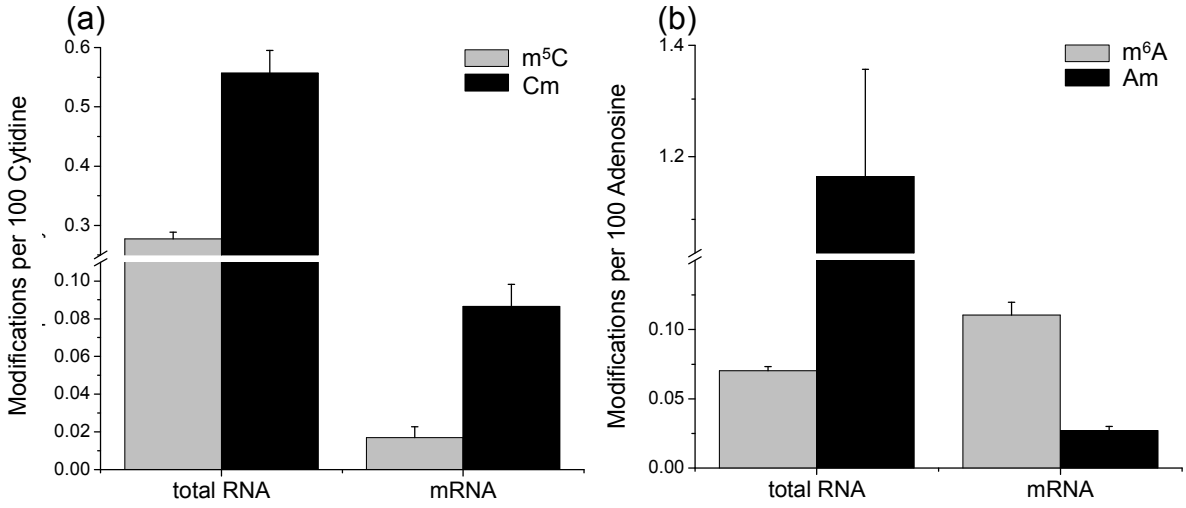

Figure 2.5 Quantification results for the levels of m^5C and Cm (a), m^6A and Am (b) in total RNA and mRNA isolated from HEK 293T cells. The data represent the mean and standard deviation of measurement results for at least three separate total RNA and mRNA samples.

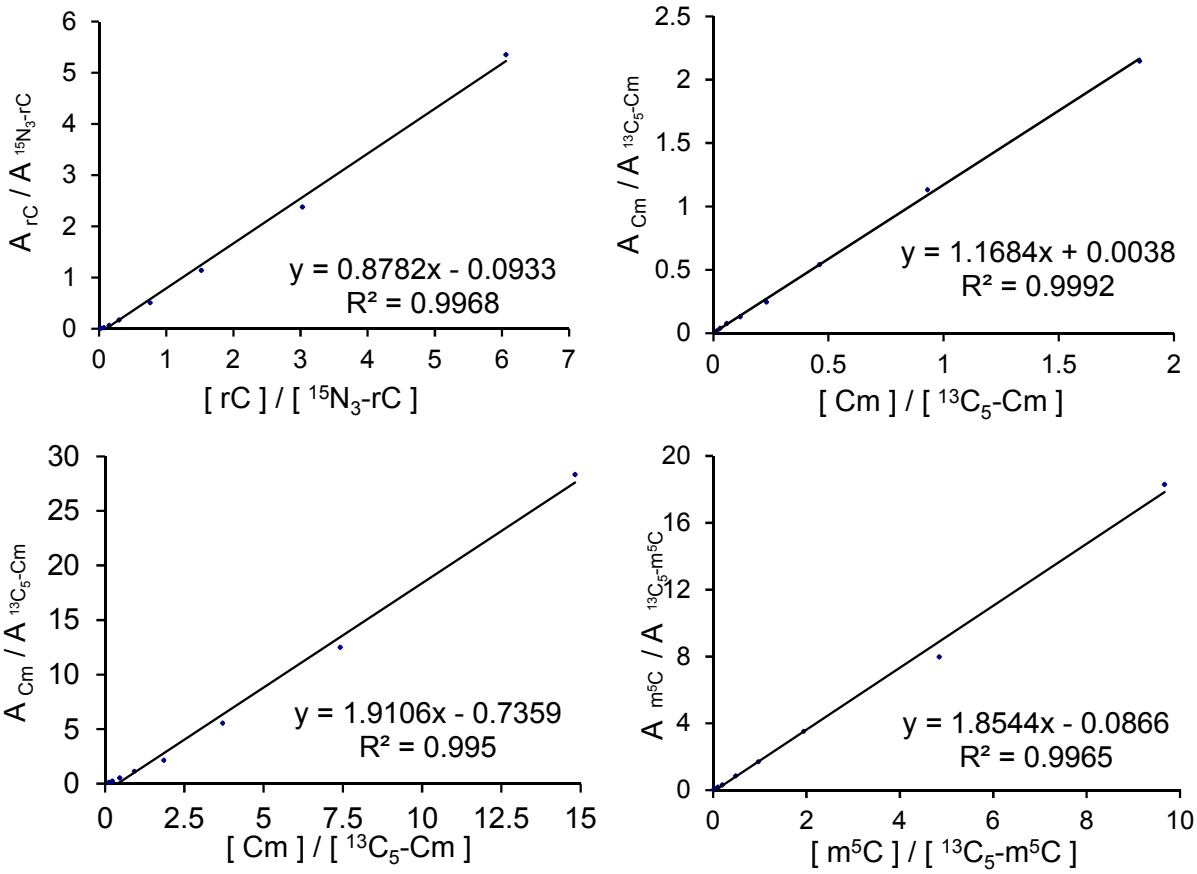

Figure 2.6 Calibration curves for the quantification of rC, m^5C and Cm in RNA. The amounts of internal standards were 3300, 25.5 and 19.4 fmol, respectively and the amounts of unlabeled rC,m^5C and Cm ranged from 49.5 fmol - 20.0 pmol, 0.3 - 144.0 fmol and 0.2 - 246.6 fmol, respectively.

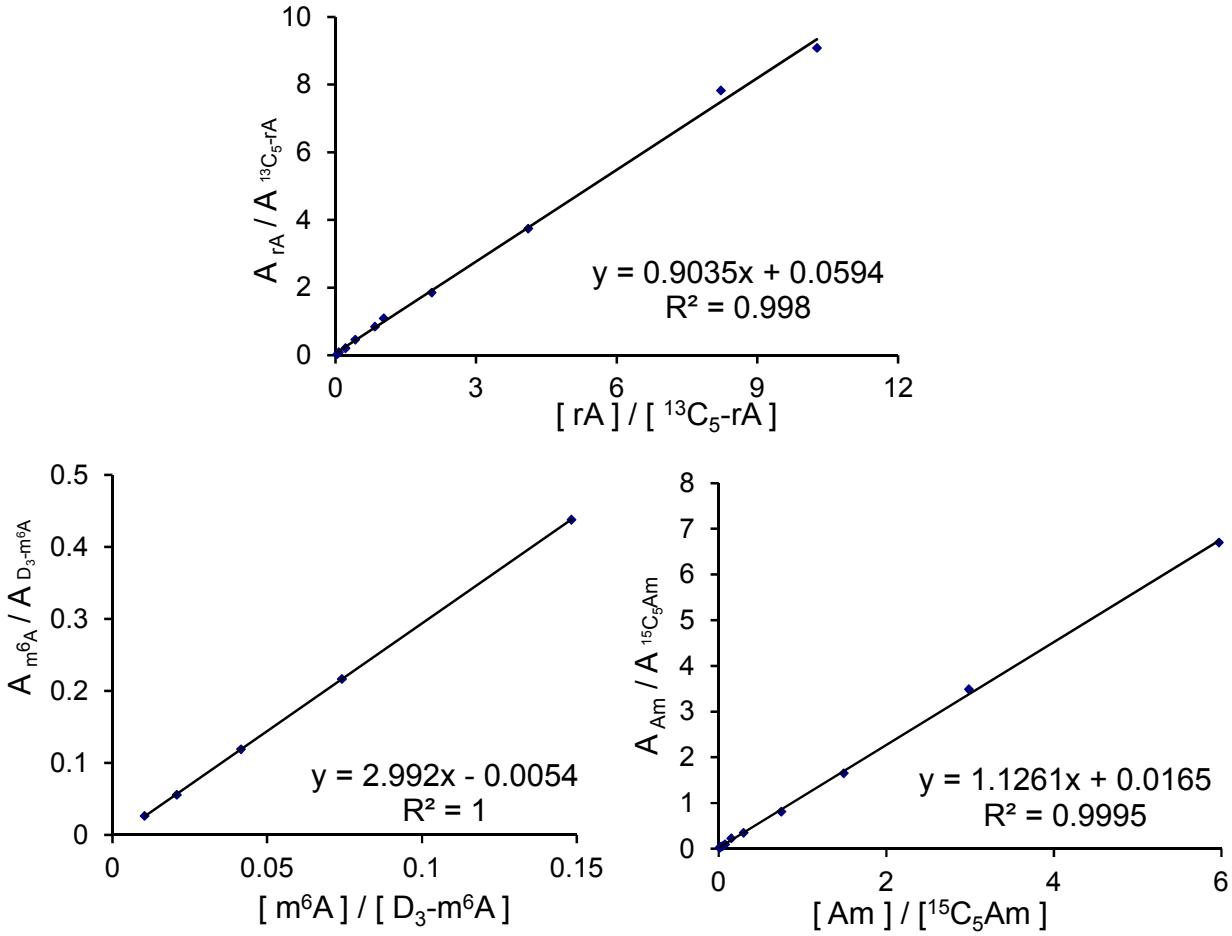

Figure 2.7 The calibration curves for the quantifications of rA, m^6A and Am in RNA. The amounts of internal standards were 1555 , 8.5 and 6.9 fmol, respectively and the amounts of unlabeled rA, m^6A and Am ranged from 52.9 fmol - 16.0 pmol, 0.09 - 1.2 fmol and 0.07 - 41.2 fmol, respectively.

Discussions

In this study, we developed an LC-MS/MS/MS coupled with stable isotope-dilution method to detect the levels of m^5C, Cm, m^6A and Am in total RNA isolated from cultured mammalian cells and tissues, as well as in mRNA isolated from HEK293T cells. This method has several advantages compared to previously reported methods. First, the measured levels of the analytes are not affected by alterations in sample matrices or LC-MS/MS/MS conditions because of the addition of stable isotope-labeled standards to the nucleoside mixture. Moreover, the analytes and corresponding internal standards are analyzed simultaneously by LC-MS/MS/MS under identical conditions. Any variations in experimental conditions after enzymatic digestion and during LC-MS/MS/MS analysis will not affect the analytical accuracy. Second, our method allows for the unambiguous identification of each analyte. Both the analytes and their isotope-labeled standards co-elute and yield the same fragmentation patterns, offering unequivocal chemical specificity for analyte identification. Lastly, this method displays superior sensitivity. The limits of quantification for m^5C, Cm, m^6A and Am with our method was found to be 9.4±1.5 amol, 7.2±2.3 amol, 1.9±0.6 amol and 3.4±1.2 amol, respectively (Table 2.1). The RNA amount used in previous methods was reported to be a few μg, wheras nucleoside mixture from digestion of only 0.5 ng of total RNA and less than 10 ng of mRNA were used for the analyses with this method.

Together, we developed a robust LC-MS^3 coupled with the stable isotope-dilution method for the quantifications of m^5C, Cm, m^6A, and Am in total RNA isolated from

mammalian tissues and cultured human cells. We were also able to accurately measure the levels of four mono-methylated ribonucleosides in mRNA isolated from HEK293T cells. To our knowledge, this is the first report about the global levels of m^5C, Cm and Am in mRNA. It can be envisaged that this method can be generally applicable for examining the role of proteins involved in the deposition, recognition and demethylation of those RNA methylations.

References

1. Dominissini, D., Genomics and Proteomics. Roadmap to the epitranscriptome. Science, 2014. 346(6214): p. 1192.

2. Machnicka, M.A., et al., MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res, 2013. 41(Database issue): p. D262-7.

3. Helm, M., Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res, 2006. 34(2): p. 721-33.

4. Motorin, Y. and M. Helm, tRNA stabilization by modified nucleotides. Biochemistry, 2010. 49(24): p. 4934-44.

5. Squires, J.E. and T. Preiss, Function and detection of 5-methylcytosine in eukaryotic RNA. Epigenomics, 2010. 2(5): p. 709-15.

6. Chow, C.S., T.N. Lamichhane, and S.K. Mahto, Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. ACS Chem Biol, 2007. 2(9): p. 610-9.

7. Kawai, G., et al., Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. Biochemistry, 1992. 31(4): p. 1040-6.

8. Cougot, N., et al., 'Cap-tabolism'. Trends Biochem Sci, 2004. 29(8): p. 436-44.

9. Daffis, S., et al., 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature, 2010. 468(7322): p. 452-6.

10. Dominissini, D., et al., Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature, 2012. 485(7397): p. 201-6.

11. Edelheit, S., et al., Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. PLoS Genet, 2013. 9(6): p. e1003602.

12. Squires, J.E., et al., Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res, 2012. 40(11): p. 5023-33.

13. Jia, G., Y. Fu, and C. He, Reversible RNA adenosine methylation in biological regulation. Trends Genet, 2013. 29(2): p. 108-15.

14. Niu, Y., et al., N6-methyl-adenosine (m6A) in RNA: an old modification with a

novel epigenetic function. Genomics Proteomics Bioinformatics, 2013. 11(1): p. 8-17.

15. Pan, T., N6-methyl-adenosine modification in messenger and long non-coding RNA. Trends Biochem Sci, 2013. 38(4): p. 204-9.

16. Amort, T., et al., Long non-coding RNAs as targets for cytosine methylation. RNA Biol, 2013. 10(6): p. 1003-8.

17. Zheng, G., et al., ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell, 2013. 49(1): p. 18-29.

18. Jia, G., et al., N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol, 2011. 7(12): p. 885-7.

19. Wang, Y., et al., N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat Cell Biol, 2014. 16(2): p. 191-8.

20. Liu, J., et al., A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol, 2014. 10(2): p. 93-5.

21. Fu, L., et al., Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc, 2014. 136(33): p. 11582-5.

22. Goll, M.G., et al., Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science, 2006. 311(5759): p. 395-8.

23. He, C., Grand challenge commentary: RNA epigenetics? Nat Chem Biol, 2010. 6(12): p. 863-5.

24. Wang, X. and C. He, Dynamic RNA modifications in posttranscriptional regulation. Mol Cell, 2014. 56(1): p. 5-12.

25. Keith, G., Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. Biochimie, 1995. 77(1-2): p. 142-4.

26. Cornelius, M.G. and H.H. Schmeiser, RNA analysis by MEKC with LIF detection. Electrophoresis, 2007. 28(21): p. 3901-7.

27. Su, D., et al., Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. Nat Protoc, 2014. 9(4): p. 828-41.

28. Yan, M., et al., A high-throughput quantitative approach reveals more small RNA modifications in mouse liver and their correlation with diabetes. Anal Chem, 2013. 85(24): p. 12173-81.

29. Kiyoshi Yamauchi , T.N., Masayoshi Kinoshita, Methylation of nucleosides with trimethylsulfonium hydroxide. Effects of transition metal ions. J. Org. Chem., 1980. 45(19): p. 3865-3868.

30. Hiroki Kumamoto, H.H., Hiromichi Tanaka, Satoru Shindoh, Keisuke Kato, Tadashi Miyasaka, Kazuki Endo, Haruhiko Machida & Akira Matsuda, Synthesis of 2-Alkynylcordycepins and Evaluation of Their Vasodilating Activity. Nucleosides and Nucleotides, 1998. 17(1-3): p. 15-27.

31. Wang, J. and Y. Wang, Synthesis and characterization of oligodeoxyribonucleotides containing a site-specifically incorporated N6-carboxymethyl-2'-deoxyadenosine or N4-carboxymethyl-2'-deoxycytidine. Nucleic Acids Res, 2010. 38(19): p. 6774-84.

32. Carroll, S.M., P. Narayan, and F.M. Rottman, N6-methyladenosine residues in an intron-specific region of prolactin pre-mRNA. Mol Cell Biol, 1990. 10(9): p. 4456-65.

33. Desrosiers, R., K. Friderici, and F. Rottman, Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci U S A, 1974. 71(10): p. 3971-5.

34. Narayan, P. and F.M. Rottman, An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. Science, 1988. 242(4882): p. 1159-62.

Chapter 3

Tet3-mediated oxidation of 5-methylcytosine on RNA

Introduction

In this section, I will use 5-mrC to indicate the m^5C in RNA and 5-mdC to show the m^5C in DNA in order to distinguish them. As we mentioned in Chapter 1, RNA is known to carry more than 100 distinct types of modifications, and these modifications modulate the structure and functions of RNA [1]. In this vein, it was found that methylation at the N^6 of adenine and oxidative demethylation of the resulting N^6-methyladenine by two members of the ALKBH family dioxygenases, i.e. FTO and ALKBH5, may be relevant in the epigenetic control of gene regulation [2-5]. Aside from N^6-methyladenosine, 5-methylcytidine (5-mrC) has long been known to be present in RNA [6]. Recent sequencing studies revealed the widespread presence of 5-mrC in both coding and non-coding RNA [7, 8], with more than 8,000 candidate m^5C sites being identified in mRNA, implicating this RNA methylation in gene regulation [7-9].

Recent studies showed that the ten-eleven translocation (Tet) family of Fe (II) - and 2-oxoglutarate-dependent dioxygenases in mammals could induce the sequential oxidation of 5-methyl-2'-deoxycytidine (5-mdC) to yield 5-hydroxymethyl-2'-deoxycytidine (5-hmdC), 5-formyl-2'-deoxycytidine (5-fodC) and 5-carboxyl-2'-deoxycytidine (5-cadC) [10-15]. In this context, it is worth noting that 5-hmdC, instead of dC, is incorporated into genomic DNA of T-eleven bacteriophage from the 5-hmdC

triphosphate, and the 5-hmdC in DNA is further glucosylated, which serve as an important mechanism for the bacteriophages to protect their DNA from degradations by host and phage factors [16]. In addition, a recent study revealed that cytosine 5-methyltransferases were capable of adding formaldehyde to the C5 position of cytosine to yield 5-hydroxymethylcytosine in DNA [17]. In mammalian cells, the oxidized derivatives of 5-mdC may constitute alternative epigenetic marks as they could be recognized by unique cellular proteins [13, 18-20]. In addition, 5-formylcytosine and 5-carboxylcytosine are readily recognized by thymine DNA glycosylase, and the subsequent action by the base excision repair machinery converts an initially methylated cytosine to its unmethylated counterpart [13, 21], which may contribute to active cytosine demethylation in mammals. Aberrant Tet-mediated oxidation of 5-mdC in DNA is known to be associated with human diseases including cancer [22-24], In addition, the genome of *D. melanogaster* lacks a homologue of the mammalian DNA methyltransferases Dnmt1, Dnmt3a or Dnmt3b, but it encodes the RNA methyltransferase Dnmt2 and a conserved Tet homologue [25]. Interestingly, ALKBH family enzymes, which are another family of Fe (II) - and 2-oxoglutarate-dependent enzymes, can oxidize the *N*-alkylated nucleobases in both DNA and RNA [26-31]. These findings, along with the structural similarity between human Tet2 and ALKBH-family enzymes [32], prompted us to hypothesize that the Tet family enzymes may also be capable of oxidizing the methyl group of 5-mrC in RNA (Figure 3.1).Here, we set up experiments to investigate the formation of 5-HmrC, 5-ForC and 5-CarC mediated with Tet proteins both *in vitro* and *in vivo*.

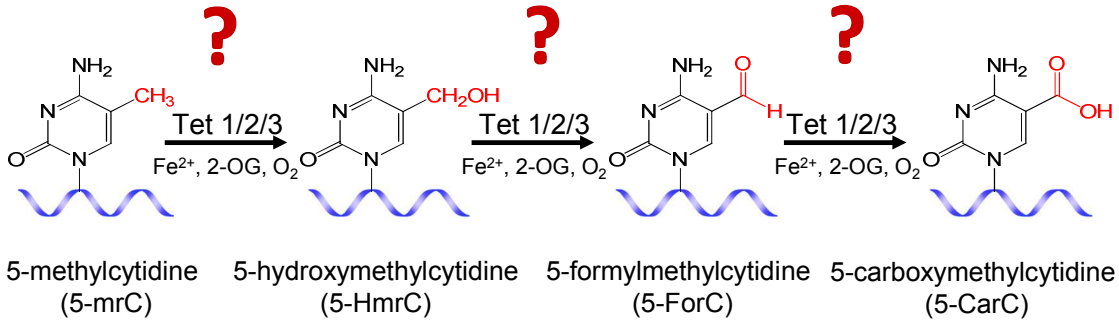

Figure 3.1 Proposed oxidative demethylation of 5-mrC to 5-hmrC, 5-forC and 5-carC in RNA by Tet1 in the presence of Fe^{2+} and 2-OG.

Experimental Procedure

Materials

All chemicals and enzymes unless otherwise specified were from Sigma-Aldrich (St. Luis, MO). Erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA) hydrochloride and [1,3-$^{15}N_2$]-cytidine were obtained from Tocris Bioscience (Ellisville, MO) and Cambridge Isotope Laboratories (Andover, MA), respectively. 5-methylcytidine, 5-formycytidine, and 5-caboxycytidine were obtained from Berry & Associates (Dexter, MI). The HEK293T human embryonic kidney cells, HeLa human cervical cancer cells, and WM-266-4 human melanoma cells, and cell culture reagents were purchased from ATCC (Manassas, VA, USA). All oligodeoxyribonucleotides and oligoribonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Expression vectors for the catalytic domain of human Tet1 (amino acids 1418-2136) and its corresponding mutant (H1672Y/D1674A) were previously described [33]. Expression vectors for the full-length mouse Tet1 and its corresponding mutant (H1620Y/D1622A) contained amino acids 1-2007. Expression vectors for the catalytic domain of human Tet2 and its catalytically inactive mutant (H1302Y/D1304A) contained amino acids 1129-2002, and those of the full-length human Tet2 and its catalytically inactive mutant (H1382Y/D1384A) contained amino acids 1-2002 [22]. Expression vectors for the catalytic domain of mouse Tet3 contained amino acids 697-1669, and those of the full-length mouse Tet3 and its corresponding mutant (H950Y/D952A) contained amino acids 1-1669.

Mouse embryonic stem (ES) cells and mouse ES cells with the depletion of all three *Tet* genes or *Tdg* gene were described elsewhere [34]. Mouse tissues used in the present study were from 19-21 week old mice. The human brain tissues used here were described previously [35].

Synthesis of [1, 3-$^{15}N_2$]-5-hydroxymethylcytidine

The titled compound was prepared at a microscale from [1,3-$^{15}N_2$]-cytidine following previously described procedures for the synthesis of isotope-labeled 5-hydroxymethyl-2'-deoxycytidine [36].

Synthesis of $^{15}C_5$-5-formylcytidine and $^{15}C_5$-5-carboxycytidine

The titled compounds were prepared at a microscale from $^{13}C_5$-5-methylcytidine ($^{13}C_5$-5-mrC) following previously described procedures with minor modifications [37]. The systhesis of $^{13}C_5$-5-methylcytidine will be reported in elsewhere. $^{13}C_5$-5-methylcytidine (75 nmol) was dissolved in 3 ml aqueous solution that was saturated with 2-methyl-1,4-naphthoquinone (MQ). The solution was then transferred to a 3.5-cm I.D. petri dish, irradiated on ice for 30 min. The solution was exposed to air during irradiation and two 15-WS Spectroline light tubes emitting at 365 nm were used (Spectronics Corporation, Westbury, NY). The distance between the light tubes and the sample solution was ~3 cm. The resulting solution was dried by using Speed-vac and the dried residue was dissolved in water and subjected to HPLC analysis.

HPLC separation was performed on an Agilent 1100 system with a Kinetex XB-C18 column (4.6× 250 mm, 5 μm in particle size and 100 Å in pore size; Phenomenex In.,

Torrane, CA, USA). Water (solution A) and methanol (solution B) were used as mobile phases, and the flow rate was 0.4 mL/min. A gradient of 20 min 0% B, 1 min 0-1% B, 16 min 1%B and 3min 1-5%B, then 5 min 5-7%B was employed. A typical HPLC trace is depicted in Figure 3.2. The $^{13}C_5$-5-formylcytidine ($^{13}C_5$-5-ForC) and $^{13}C_5$-5-carboxycytidine ($^{13}C_5$-5-CarC) were eluted out at at 3.9 min and 56.5 min. The corresponding fractions were collected and the identities and concentrations of the modified nucleosides were confirmed by mass spectrometric analyses (Figure 3.3 & 3.4).

Figure 3.2 HPLC trace of isolation of labeled 5-forC and 5-carC from the mixture of MQ-sensitized photoreaction of 13C5-5-methylcytidine. The fractions at retention times of 3.9 min and 56.5 are labeled 5-forC and 5-carC, respectively.

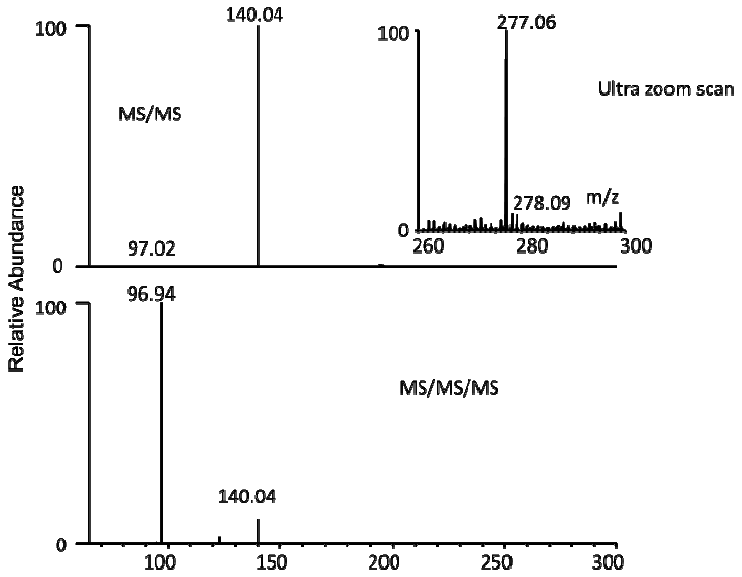

Figure 3.3 MS/MS and MS/MS/MS characterizations of 5-forC, which monitor the fragmentation of the [M+H]+ ion of the 5-forC (top) and the further fragmentation of the protonated nucleobase (bottom), respectively. Displayed in the inset of is the positive-ion electrospray ionization mass spectrum (ESI-MS) for 5-forC.

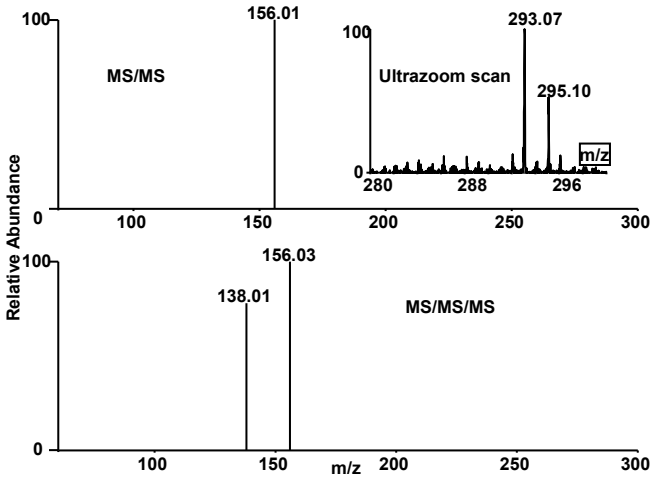

Figure 3.4 MS/MS and MS/MS/MS characterizations of 5-carC, which monitor the fragmentation of the [M+H]+ ion of the 5-carC (top) and the further fragmentation of the protonated nucleobase (bottom), respectively. Displayed in the inset of is the positive-ion ESI-MS for 5-carC.

Biochemical assay of Tet1-mediated oxidation of 5-mrC in RNA and 5-mdC in DNA

The Tet1-mediated RNA oxidation assay was conducted with the use of the 5mC Tet1 Oxidation Kit (Wisegene, IL, USA), where a 12.5-μL reaction mixture contained 200 pmol of m^5C-bearing single-stranded RNA, 5'-UUUCAGCUC(5-mrC)GGUCACGCUC-3', the catalytic domain of mouse Tet1 (with amino acids 1367 to 2039) and reagents 1 and 2 included with the kit. The reaction was incubated at 37°C for 30 min and the enzyme was removed immediately afterwards by chloroform extraction. The reaction mixture was subsequently digested with nuclease P1 at 37°C for 4 hrs and then with alkaline phosphatase at 37°C for 2 hrs. The resulting ribonucleoside mixture was then subjected to HPLC analysis to determine the extent of conversion 5-mrC to 5-hmrC, and the HPLC conditions were described below in the HPLC enrichment section.

Tet1-mediated reactions were also performed using a 11mer 5-mrC-containing RNA, AGCUC(5-mC)GGUCA, and the reaction mixtures were analyzed directly by LC-MS and MS/MS to identify the reaction products and to quantify the levels of conversions of 5-mrC to 5-hmrC. To this end, 10 pmol of RNA was incubated with 0.125 μL of mTet1 protein, along with reagents 1 and 2 included with the kit in a total volume of 10 μL, at 37°C for the indicated periods of time. In order to measure the further conversions of m^5C to 5-forC and 5-carC, the volume of mTet1 protein was increased to 0.5 μL to increase the conversion efficiency. The reaction mixtures were immediately frozen on dry ice, and the enzyme in the mixture was subsequently removed using chloroform extraction. The residual chloroform was removed by using a speed-vac for 5

min, and the remaining aqueous solution was subjected directly to LC-MS analysis on an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific), where MS/MS and the higher-resolution "ultra-zoom-scan" MS were acquired for the $[M - 3H]^{3-}$ ions of the starting 11mer RNA as well as the corresponding RNA with the 5-mrC being oxidized to 5-hmrC, 5-forC, or 5-carC (Figure 3.8, Figure 3.10 and Figures 3.12-3.14). The intensities of the monoisotopic peak and the +1 isotopic peak were employed for calculating the percentages 5-mrC, 5-hmrC, and 5-forC in the reaction mixtures (5-carC-containing RNA was not detectable under the reaction conditions used). In this regard, the molecular weight of 5-hmrC and 5-forC differ by only 2 Da; thus, the +2 and +3 isotopic peaks of the 5-forC-bearing RNA overlap respectively with the monoisotopic and +1 isotopic peaks of the 5-hmrC-containing RNA. Therefore, we subtracted the intensities of +2 and +3 isotopic peaks of the 5-forC-harboring RNA from the intensities of the monoisotopic and +1 isotopic peaks of the 5-hmrC-containing RNA, considering that the relative abundances of the monoisotopic, +1, +2, and +3 isotopic peaks of the $[M - 3H]^{3-}$ ion of the 5-forC-carrying RNA (with elemental composition of $C_{107}H_{135}N_{41}O_{65}P_{10}$) follow the ratios of 73.9:100:76.8:42.9.

Similar reactions were performed for a single 5-mC-containing duplex DNA with the same sequence context, i.e., d(AGCTC(5-mC)GGTCA)/d(GTGACCGGAGCTG) under identical reaction conditions, and the levels of the conversions of 5-mdC to 5-hmdC, 5-fodC and 5-cadC were determined by LC-MS using similar procedures as described above for the single-stranded RNA.

Cell culture, transfection and RNA extraction

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (ATCC). HeLa and WM-266-4 cells were cultured in Eagle's Minimum Essential Medium (ATCC). All cells were incubated at 37°C in 5% CO_2 atmosphere. The culture medium was supplemented with 10% fetal bovine serum and 100 IU/mL penicillin. The HEK293T cells were seeded in 6-well plates at 50-60% confluence level and, 24 hrs later, the cells were transfected with 1.5 μg or 7.5 μg plasmid for overexpressing the full-length (FL) or catalytic domain (CD), or the corresponding catalytically inactive mutants (FL-m and CD-m), of the three Tet family enzymes, except that 3.0 μg plasmids were employed for full-length Tet2 and its catalytically inactive mutant, using Lipofectamine 2000 (Invitrogen). Control experiments were also conducted by transfecting cells with 1.5 μg pGEM-T Easy plasmid (Pormega). The cells were harvested for RNA extraction 48 hrs after plasmid transfection. Total RNA was isolated from mammalian cells and tissues using TRI Reagent® according to the manufacturer's recommended procedures. The RNA pellet was dissolved in RNase-free water and its concentration measured by UV absorbance at 260 nm.

DNA was also isolated from the same cells and the levels of 5-hmdC in the DNA samples were measured using LC-MS/MS/MS, as described previously [38].

Enzymatic digestion of total RNA

One unit of nuclease P1, 0.00125 unit of phosphodiesterase 2, 2.5 nmol of EHNA and a 20-μL solution containing 300 mM sodium acetate (pH 5.6) and 10 mM zinc

chloride were added to 10 μg of RNA, where EHNA was added to minimize the deamination of adenosine. The above digestion was continued at 37°C for 4 hrs. To the digestion mixture were then added 1 unit of alkaline phosphatase, 0.0025 unit of phosphodiesterase 1, and 40 μL of 0.5 M Tris-HCl buffer (pH 8.9). The digestion mixture was incubated at 37°C for 2 hrs. To the mixture was then added 200 fmol of [1,3-$^{15}N_2$]5-hmrC and 44 fmol $^{13}C_5$-5-forC. The enzymes in the digestion mixture were subsequently removed by chloroform extraction. The resulting aqueous layer was dried, reconstituted in doubly distilled water, and subjected to off-line HPLC separation for the enrichment of the 5-hmrC, 5-forC and 5-carC.

HPLC enrichment

HPLC analysis was performed on a Beckman HPLC system with pump module 125 and a UV detector (module 126). A 4.6×250 mm Alltima HP C18 column (5 μm in particle size, Grace Davison, Deerfield, IL) was used. A solution of 10 mM ammonium formate (solution A) and methanol (solution B) were used as mobile phases, and the flow rate was 0.7 mL/min. A gradient of 42 min 0-0.5% B and 27 min 0.5-8% B was employed. A typical HPLC trace is depicted in Figure 3.5. A minor DNA contamination (less than 5% for most RNA samples) was often observed for the total RNA isolated from mammalian cells and tissues based on the chromatograms obtained during the HPLC enrichment. Correction for DNA contamination was made for each RNA sample according to the peak areas of the 2'-deoxynucleosides relative to those of ribonucleosides. The HPLC fractions eluting at 12.5-16 min were pooled for 5-hmrC, 48-

52 min were pooled for 5-forC and 5-7 min were pooled for 5-carC. The collected fractions were dried in the Speed-vac, redissolved in H_2O, and injected for LC-MS/MS/MS or LC-MS/MS analysis.

LC-MS/MS/MS Analysis of 5-hmrC

A 0.5 mm × 250 mm Zorbax SB-C18 column (particle size, 5 μm, Agilent) was used for the separation of the above-enriched 5-hmrC fraction, and the flow rate was 8.0 μL/min, which was delivered by using an Agilent 1100 capillary HPLC pump (Agilent Technologies). A solution of 0.1% (v/v) formic acid in water (solution A) and a solution of 0.1% (v/v) formic acid in methanol (solution B) were used as mobile phases, and a 30-min linear gradient of 0-70% B was employed. The effluent from the LC column was directed to an LTQ linear ion-trap mass spectrometer, which was set up for monitoring the fragmentation of the labeled and unlabeled 5-hmrC in the positive-ion mode. The temperature for the ion transport tube was maintained at 275 °C, the sheath gas flow rate was 15 arbitrary units, and no auxiliary gas was used. The electrospray, capillary, and tube lens voltages were 5 kV, 16 V, and 60 V, respectively; each MS^3 scan was composed of three microscans, and the maximum time for each microscan was 250 ms. The normalized collision energies were 42% and 27% in MS/MS and MS/MS/MS, respectively, and activation Q was 0.25 in both MS/MS and MS/MS/MS.

LC-MS/MS Analysis of 5-forC

LC-MS/MS measurement of 5-forC was conducted on a triple quadrupole mass spectrometer equipped with a nanoelectrospray ionization source coupled to an EASY-

NLC II system (Thermo Fisher Scientific, San Jose, CA, USA). Samples were automatically loaded onto a home-made trapping column (150 μm × 70 mm) packed with PGC packing material (5 μm beads, Thermo Electron Corporation, USA) at a flow rate of 2.5 μL/min and eluted to an in-house packed PGC column (75 μm × 250 mm, 5 μm beads, 100 Å pore size, Agilent). A solution of 0.1% (v/v) formic acid in water (solution A) and a solution of 0.1% (v/v) formic acid in acetonitrile (solution B) were used as mobile phases. The modified nucleosides were separated with a gradient of 5 min 0– 20% B, 40 min 20-90% B, 10 min 90% B. The flow rate was 300 nL/min. The temperature for the ion transport tube was maintained at 270 °C, The electrospray voltage was 1.8 KV and the collision energy was 18 V. The Q1 and Q3 resolution was 0.7 Da. A scan width of 0.5 Da was used for the product ion detection.

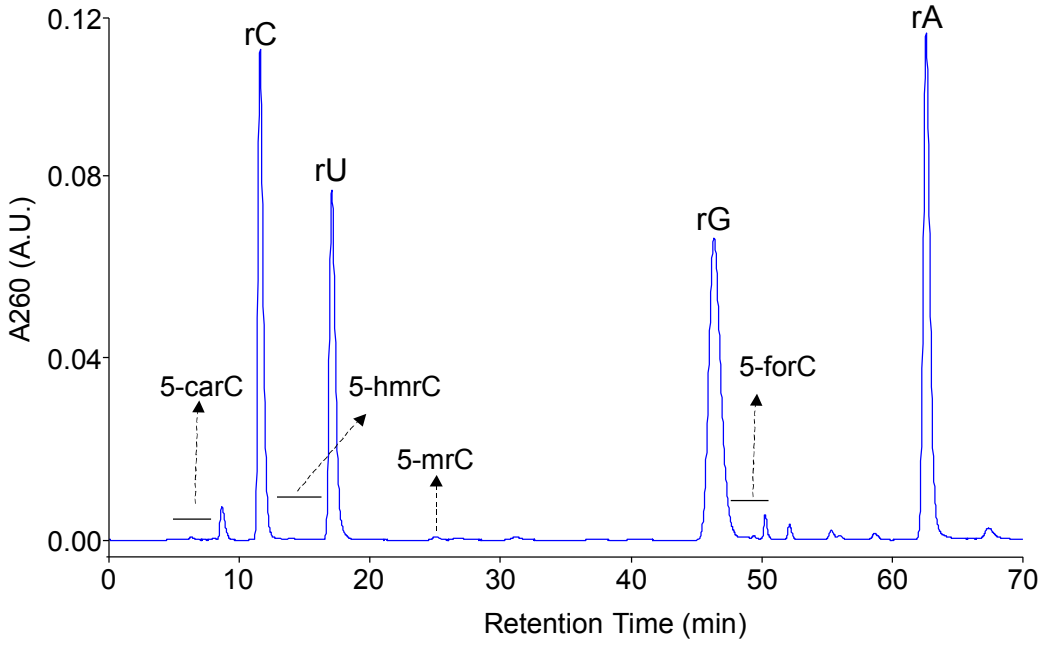

Figure 3.5 A representative HPLC trace for the enrichment of 5-hmrC and for the quantification of 5-mrC from the enzymatic digestion mixture of total RNA isolated from cells or tissues. Shown is the trace for the nucleoside mixture of a RNA sample isolated from mouse brain.

Results

Tet-mediated formation of 5-HmrC, 5-ForC and 5-CarC in single-strand RNA *in vitro*

To explore this possibility, we first assessed the capability of recombinant catalytic domain of mouse Tet1 protein in inducing the oxidation of 5-mrC in RNA by conducting an *in vitro* reaction with the use of a single-stranded RNA carrying a single 5-mrC. HPLC analysis of the nucleoside mixture from the enzymatic digestion of the RNA isolated from the reaction mixture revealed the formation of 5-hydroxymethylcytidine (5-hmrC), which is accompanied with a decrease in the level of 5-mrC (Figure 3.6a). However, no significant peaks corresponding to 5-forC and 5-CarC were observed in the UV-spectrum from HPLC analysis which may due to the poor sensitivity of HPLC analysis or low conversion from 5-hmrC to 5-forC and 5-carC. The identities of the two nucleosides (i.e., 5-hmrC and 5-mrC) were confirmed by mass spectrometric analyses (Figure 3.6 , Figure 3.7). Thus, this biochemical assay demonstrated that Tet1 is able to oxidize 5-mC in single-stranded RNA *in vitro*.

We next investigated the relative efficiencies of the catalytic domain of Tet1 in oxidizing 5-mrC in RNA and 5-mdC in DNA. To this end, we conducted another *in-vitro* reaction by using a 11-mer RNA sequence, AGCUC(5-mrC)GGUCA, or a duplex DNA with a single 5-mdC situated in the same sequence context. We then subjected the reaction mixtures directly to LC-MS and MS/MS analyses (Figure 3.8-3.14). Quantification results based on LC-MS data revealed 5-hmrC as the major product

formed when single-stranded RNA was employed as the substrate, though we were able to detect very low level of 5-forC at 40 min (Figure 3.8 and Figure 3.9). It is of note that omitting Fe^{2+} in the reaction buffer led to a decrease in the formation of 5-hmrC by ~4 fold, whereas exclusion of 2-oxoglutarate in the reaction buffer nearly abolished the Tet1-catalyzed formation of 5-hmrC (Figure 3.15), supporting that 5-hmrC arises from the Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase activity of Tet1. For the duplex DNA substrate, we, however, observed a rapid formation of 5-hmdC and then 5-fodC, which is accompanied by the nearly complete loss of 5-mdC. Furthermore, 5-hmdC and 5-fodC were almost completely converted to 5-cadC at later time points (Figures 3.9 b). This finding is consistent with Tet enzyme's capability in the sequential oxidation of 5-mdC to 5-hmdC, 5-fodC, and 5-cadC [39, 40]. These results, therefore, supported that the Tet-mediated oxidation of 5-mrC in RNA is much less efficient than the corresponding oxidation of 5-mdC in duplex DNA. We also found that Tet1 displayed a higher activity toward single-stranded DNA than single-stranded RNA in the same sequence context (Figure 3.16). Comparison of the extents of oxidation of 5-mdC in single vs. double-stranded DNA showed that the oxidation of 5-mdC is more facile in the latter substrate, which could be attributed to the preferential binding of Tet1 to duplex DNA. Thus, the less efficient oxidation of 5-mrC to 5-hmrC in RNA than the corresponding oxidation of 5-mdC in duplex DNA could be partly due to the less favorable binding of Tet1 to single-stranded RNA. Future structural determination of the Tet protein-RNA complex, along with the known structure of Tet2-DNA complex [32], may provide some mechanistic insights about this difference.

Based on the previous observations, we reason that 5-forC and 5-carC might also be produced when an elevated level of Tet1 is used for the reaction. Therefore, we incubated the 11 mer m^5C-containing RNA with 1 μL Tet1 protein, which is 10 times more than what we used for the above reaction. Our LC-MS results indicated that, under these conditions, significant amount of 5-forC- and 5-carC-containing 11mer RNA were formed (Figure 3.10 & 3.11). The yields of these two products reached to a plateau after 60 min which may be attributed to the loss of enzymatic activity. After 60 min, the 5-hmrC product was been completely consumed. Comparison of the extents of oxidation of 5-mdC in single-stranded RNA vs. double-stranded DNA showed that the Tet1-mediated oxidation of 5-mrC in RNA experienced iterative oxidation pathway, similar as what was observed in double-stranded DNA: from methyl group to hydroxymethyl group, then to fomyl group and further to carboxyl group.

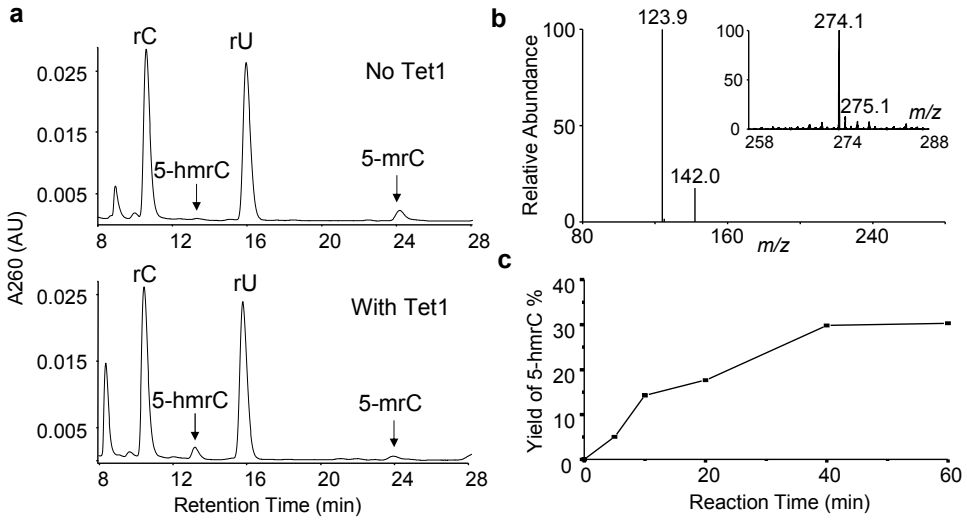

Figure 3.6 (a) As shown by HPLC profiles of digested substrates, a new peak which is corresponded to 5-hmrC showed and the intensity of 5-mrC peak was significantly diminished with treatment of Tet1 protein. (b) MS/MS/MS profile of 5-hmrC fraction and the insert is the ultrazoom scan profile. (c)The plot of time dependent Tet1-mediated yield of 5-hmrC.

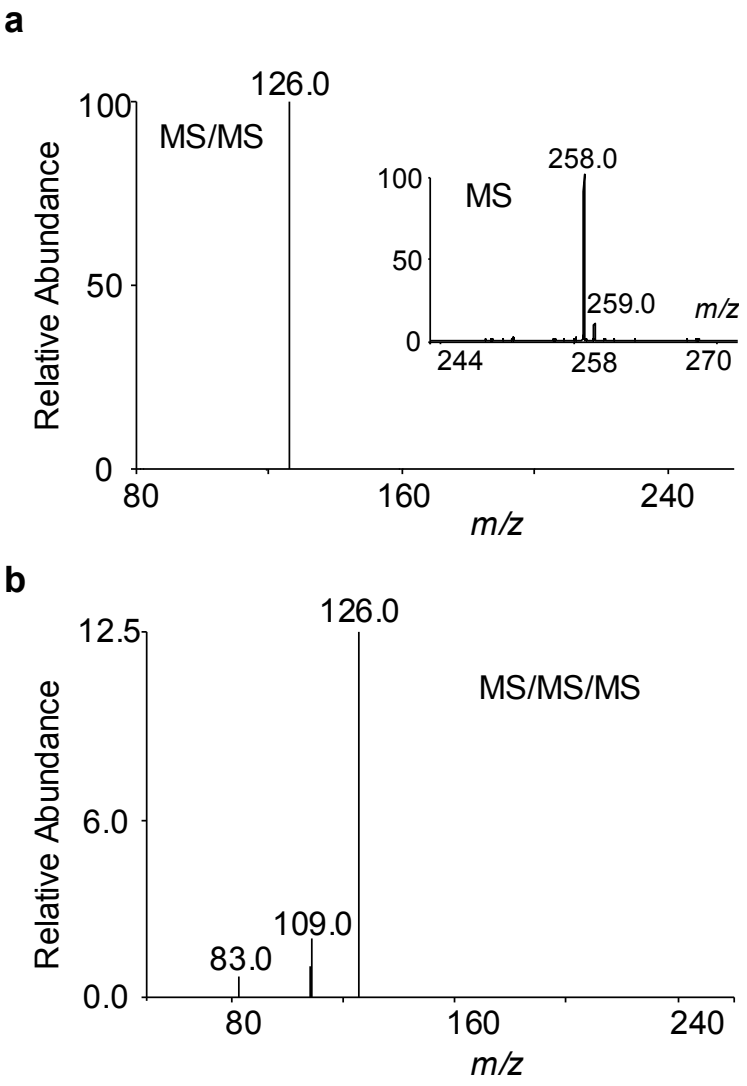

Figure 3.7 Positive-ion ESI-MS/MS (a) and MS/MS/MS (b) of the 5-mrC fraction from the HPLC separation of the nucleoside mixture of RNA isolated from the in-vitro Tet1-oxidation assay. The inset in (a) gives the high-resolution "ultra-zoom scan" MS for 5-mrC.

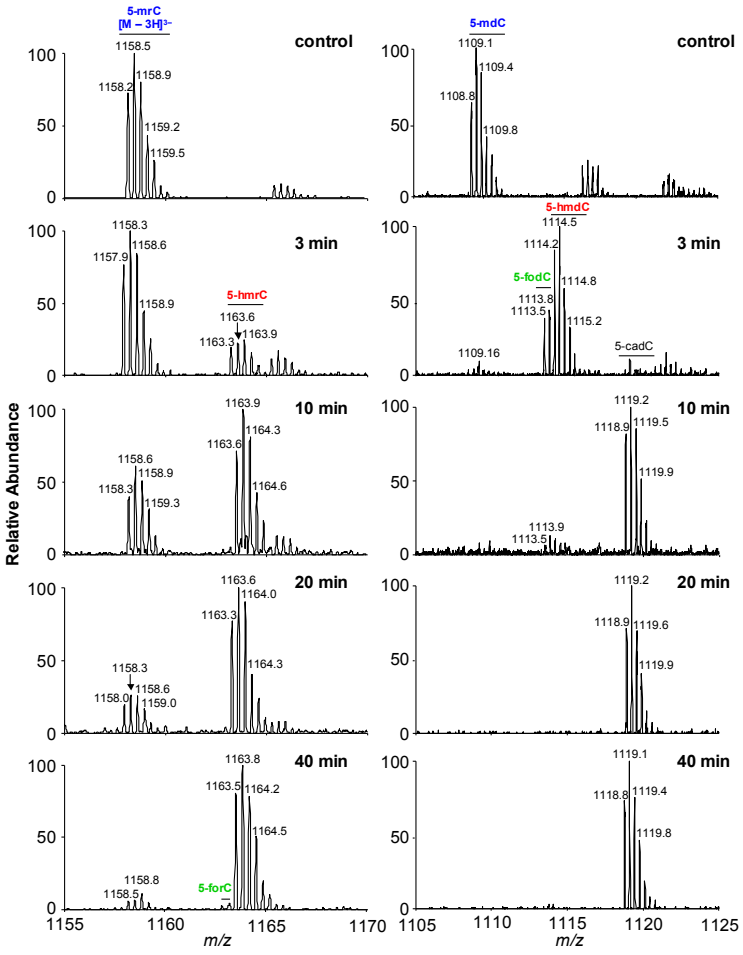

Figure 3.8 LC-MS for monitoring the Tet1-mediated oxidation of 5-mrC in a single-stranded RNA, AGCUC(5-mrC)GGUCA (left) and a duplex DNA, d(AGCTC(5-mdC)GGTCA) /d(TGACCGGAGCT) (right). Shown are the higher-resolution "ultra-zoom-scan" MS results for monitoring the $[M\text{-}3H]^{3-}$ ions of the initial 5-mC-bearing 11mer RNA (left) or DNA (right), together with their oxidation products, where the 5-mC is oxidized to 5-hmC, 5-foC, and 5-caC. The peaks at around m/z 1166 and m/z 1117 for the control samples in the left and right panels are attributed to the Na+ ion adduct, i.e., the $[M+Na+\text{-}4H]^{3-}$ ions, of the 5-mrC-containing RNA and 5-mdC-bearing DNA strand, respectively.

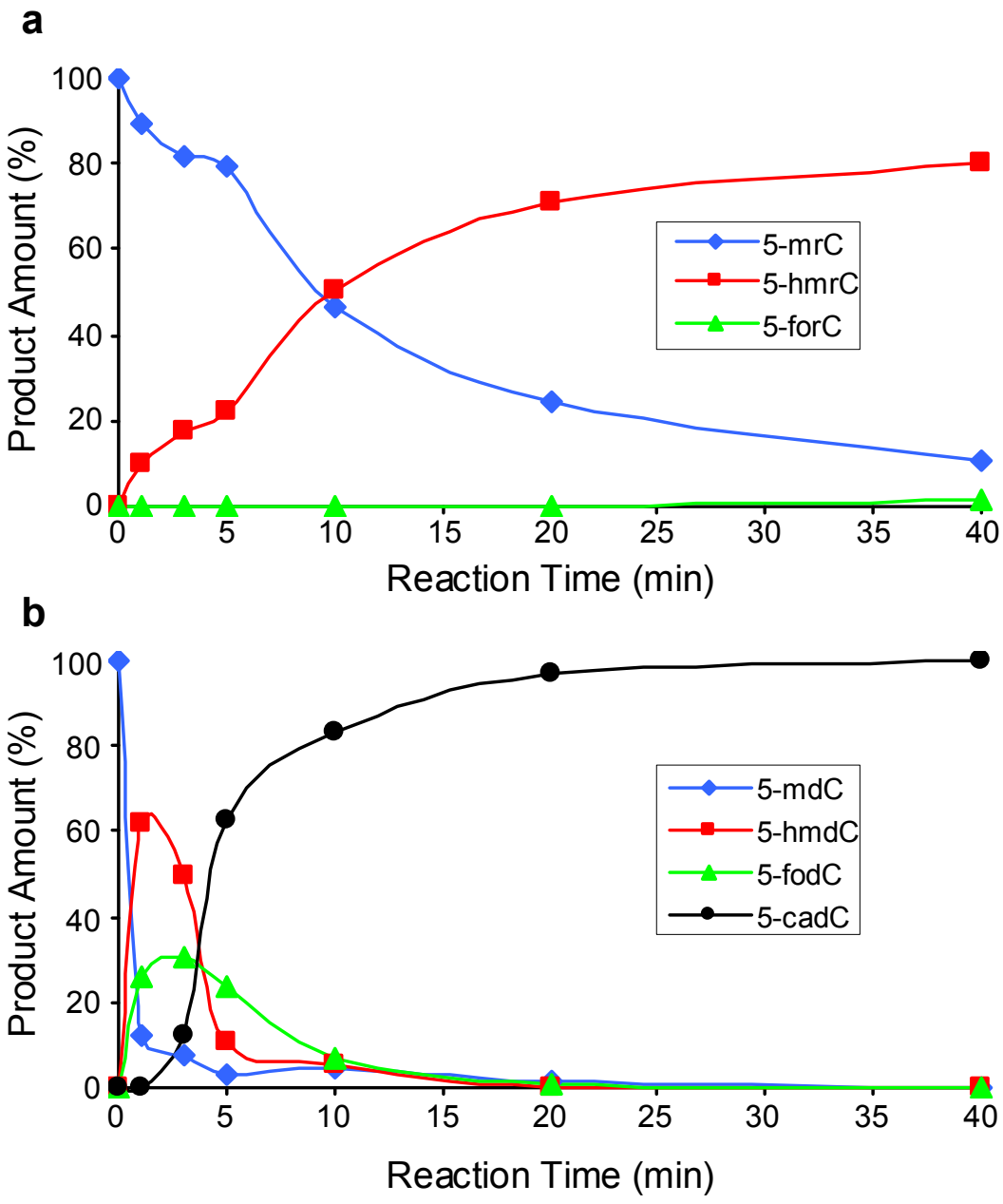

Figure 3.9 Time-dependent formation of oxidation products of 5-mrC in single-stranded RNA, AGCUC(5-mrC)GGUCA (a), and of 5-mdC in duplex DNA, d(AGCTC(5-mdC)GGTCA) /d(GTGACCGGAGCTG) (b). The products were quantified from LC-MS analyses. 0.125 μL Tet1 protein was used.

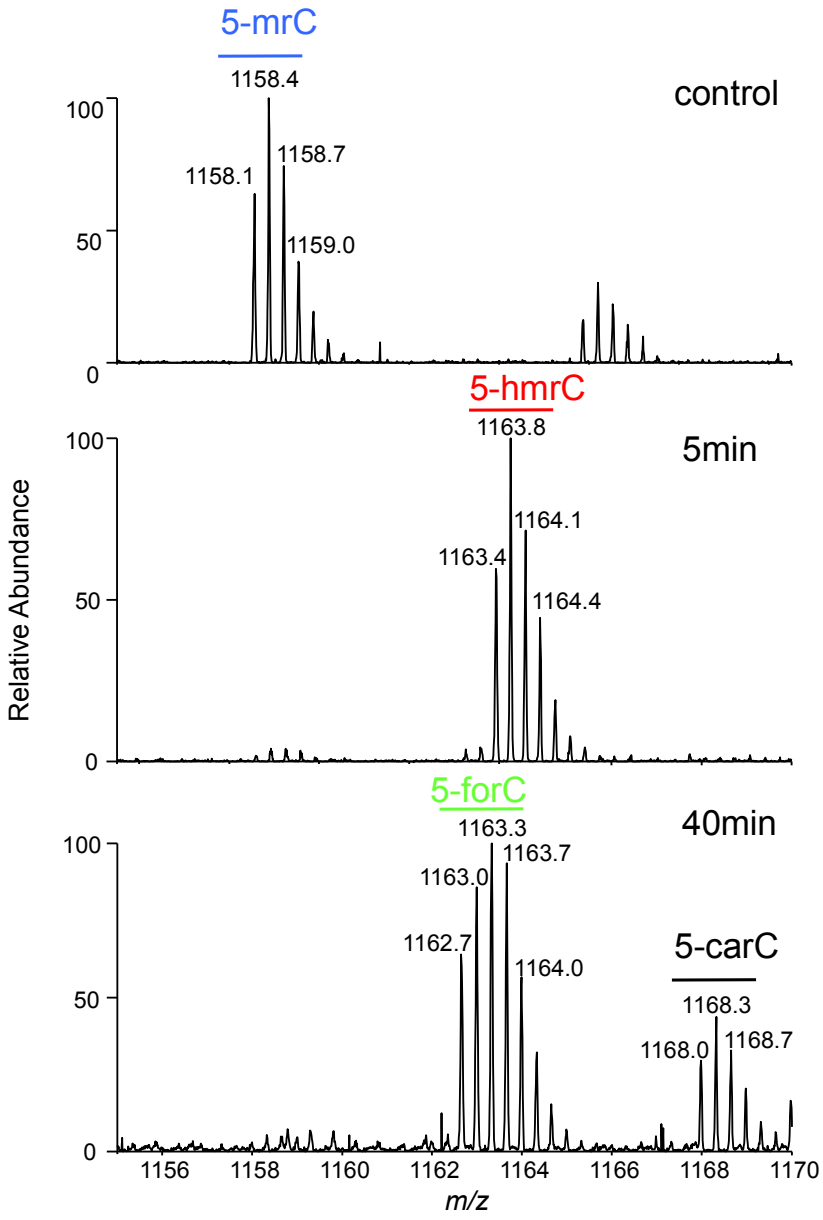

Figure 3.10 LC-MS for monitoring the Tet1-mediated oxidation of 5-mrC in a single-stranded RNA with more Tet1 enzyme. Shown are the higher-resolution "ultra-zoom-scan" MS results for monitoring the [M-3H]3- ions of the initial 5-mC-bearing 11mer RNA together with their oxidation products, where the 5-mC is oxidized to 5-hmrC, 5-forC, and 5-carC.

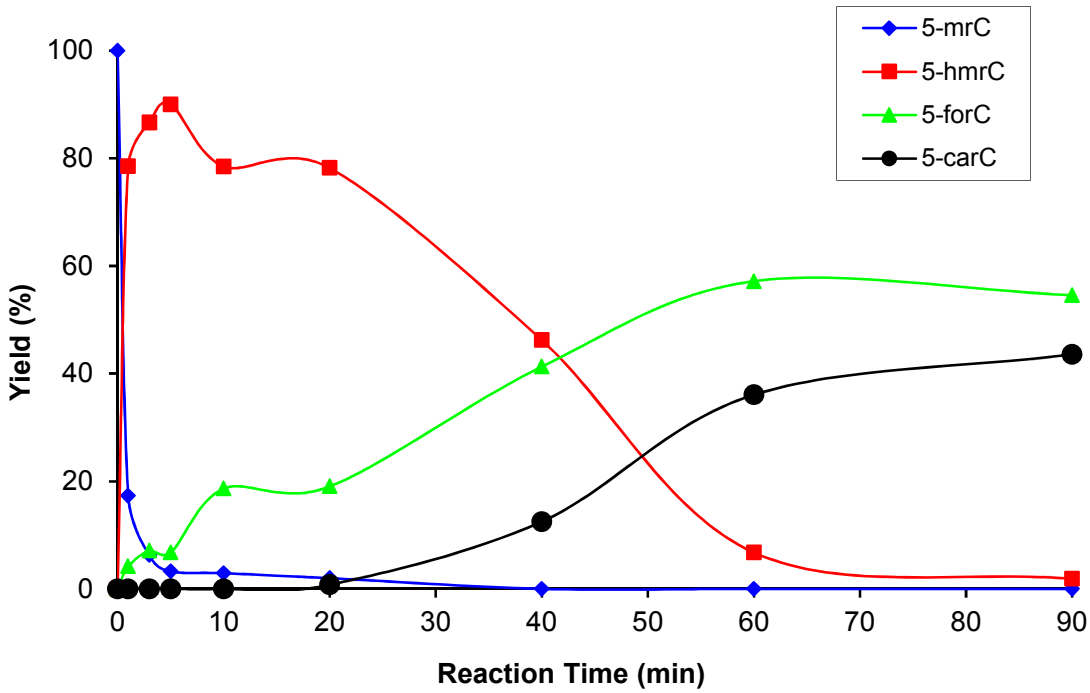

Figure 3.11 Time-dependent formation of oxidation products of 5-mrC in single-stranded RNA, AGCUC (5-mrC) GGUCA. The products were quantified from LC-MS analyses. 1 μL Tet1 protein was used.

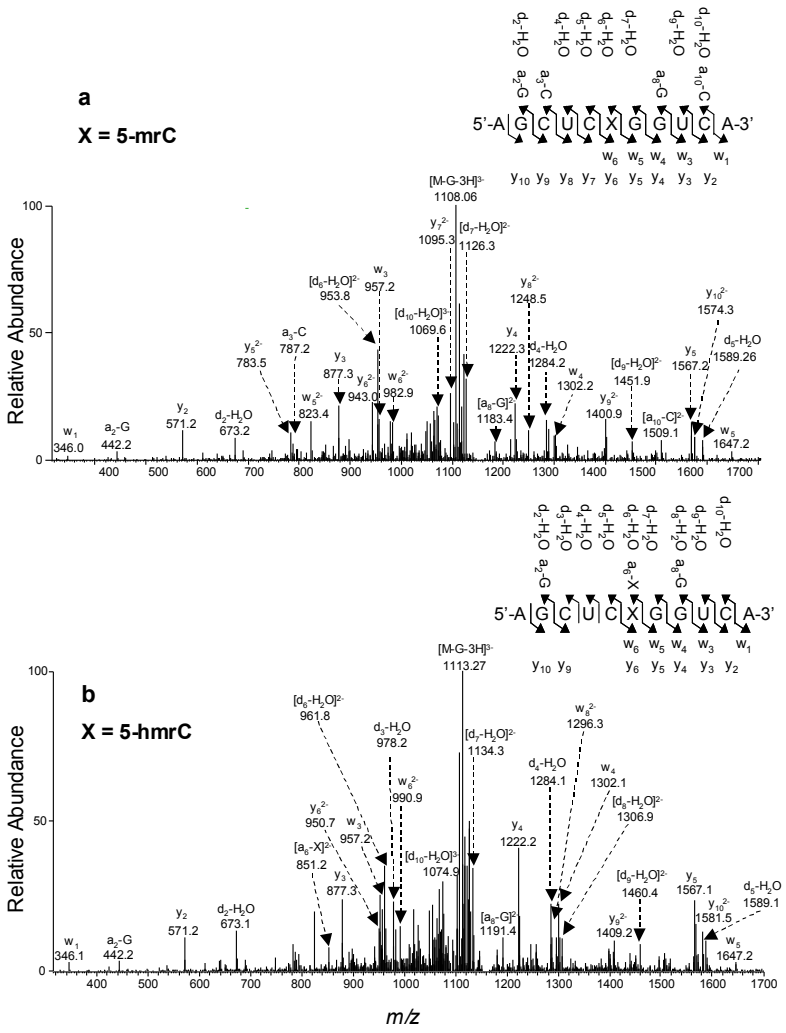

Figure 3.12 ESI-MS/MS for the $[M - 3H]^{3-}$ ions of AGCUCXGGUCA, where 'X' is a 5-mrC (a) or 5-hmrC (b) found in the Tet1-catalyzed reaction mixture of the 11-mer single-stranded RNA. The *m/z* values of fragment ions for RNA were calculated using the Mongo Oligo Mass Calculator v2.06 (http://mods.rna.albany.edu/masspec/Mongo-Oligo). The mass difference between the neighboring $[d_n - H_2O]$, w_n, or y_n ions defines the identity of the nucleotide flanked by the two neighboring ions. In particular, the mass difference between the w_5 and w_6 ions, between the y_5 and y_6 ions, or between $[d_5 - H_2O]$ and $[d_6 - H_2O]$ ions, corresponds to the residue mass of 5-mrC-5'-monophosphate (a) or 5-hmrC-5'-monophosphate (b).

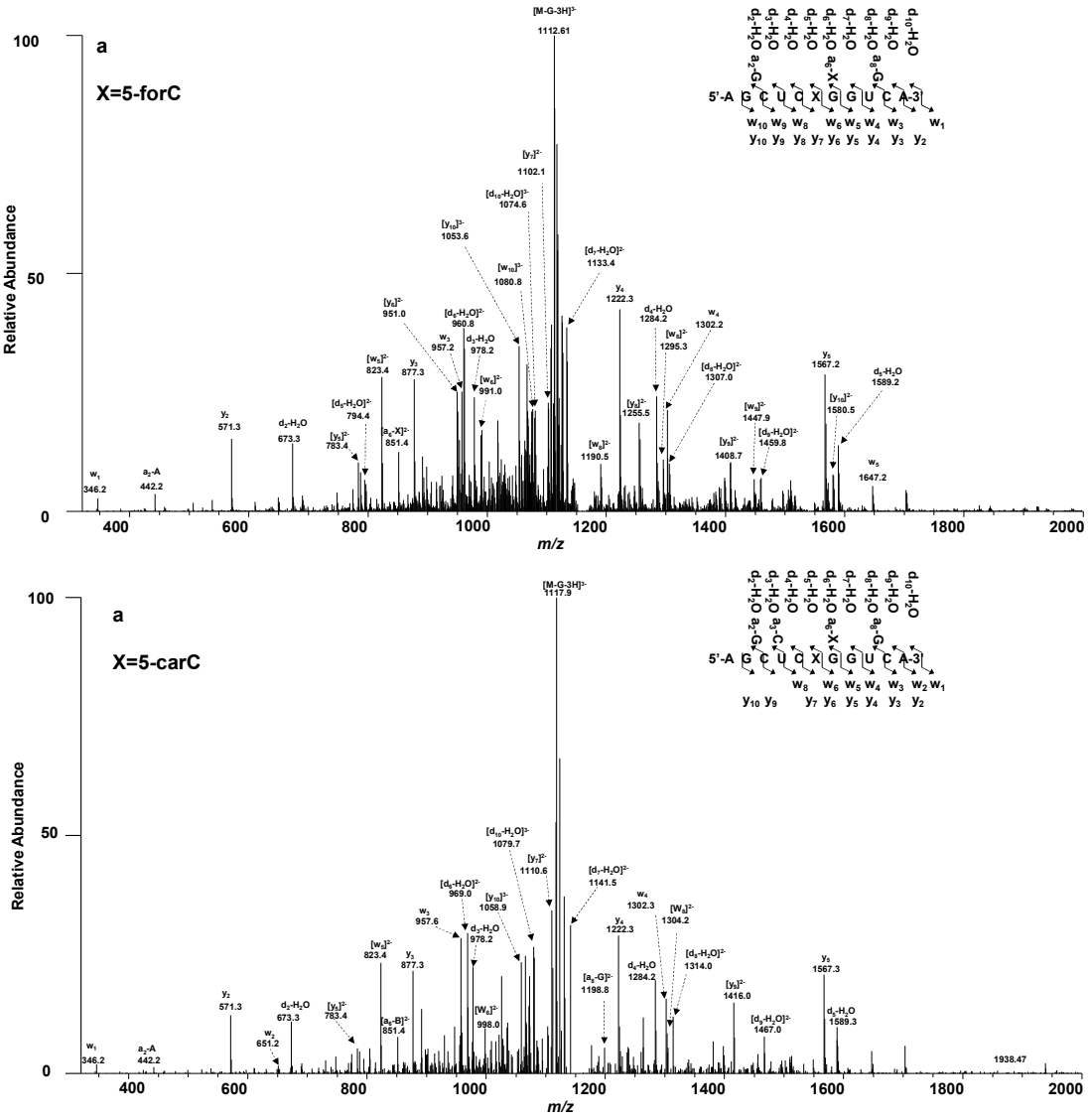

Figure 3.13 ESI-MS/MS for the [M – 3H]3- ions of AGCUCXGGUCA, where 'X' is a 5-forC (a) or 5-carC (b) found in the Tet1-catalyzed reaction mixture of the 11-mer single-stranded RNA. The m/z values of fragment ions for RNA were calculated using the Mongo Oligo Mass Calculator v2.06 (http://mods.rna.albany.edu/masspec/Mongo-Oligo). The mass difference between the neighboring [dn – H2O], wn, or yn ions defines the identity of the nucleotide flanked by the two neighboring ions.

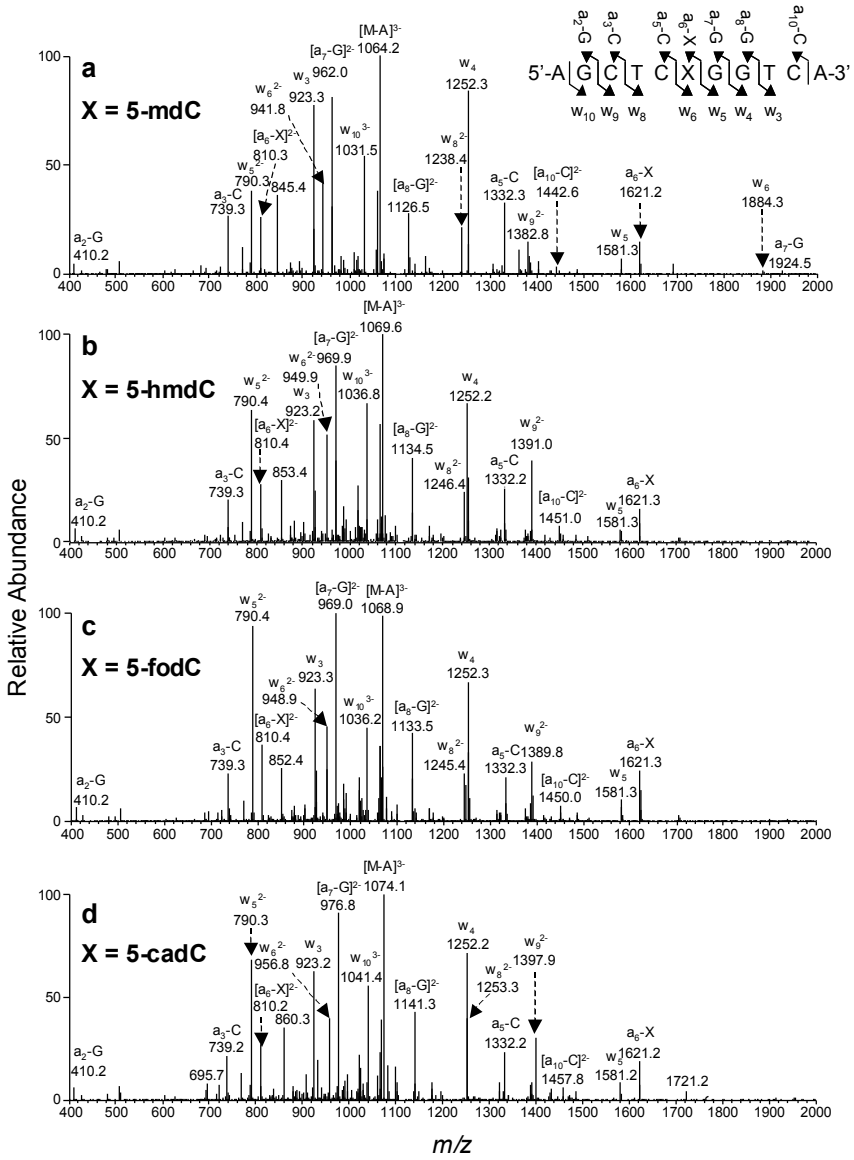

Figure 3.14 ESI-MS/MS for the [M – 3H]3- ions of d(AGCTCXGGTCA) found in Tet1-catalyzed reaction mixture of the 11-mer duplex DNA, where 'X' is a 5-mdC (a), 5-hmdC (b), 5-fodC (c), or 5-cadC (d) (on next page). Collisional activation of deprotonated ions of ODNs led to the loss of nucleobases (A, C, or G) and subsequent cleavages of the 3' C-O bond of the same nucleoside to give [an – Base] and its complementary wn ions 7; the mass difference between the neighboring [an – Base] or wn ions defines the identity of the nucleotide flanked by the two neighboring ions. For instance, the mass difference between the w5 and w6 ions, or between a6 – X and a7 – G ions, corresponds to the residue mass of 5-mdC-5'-monophosphate (a) or its corresponding oxidized derivatives (b-d).

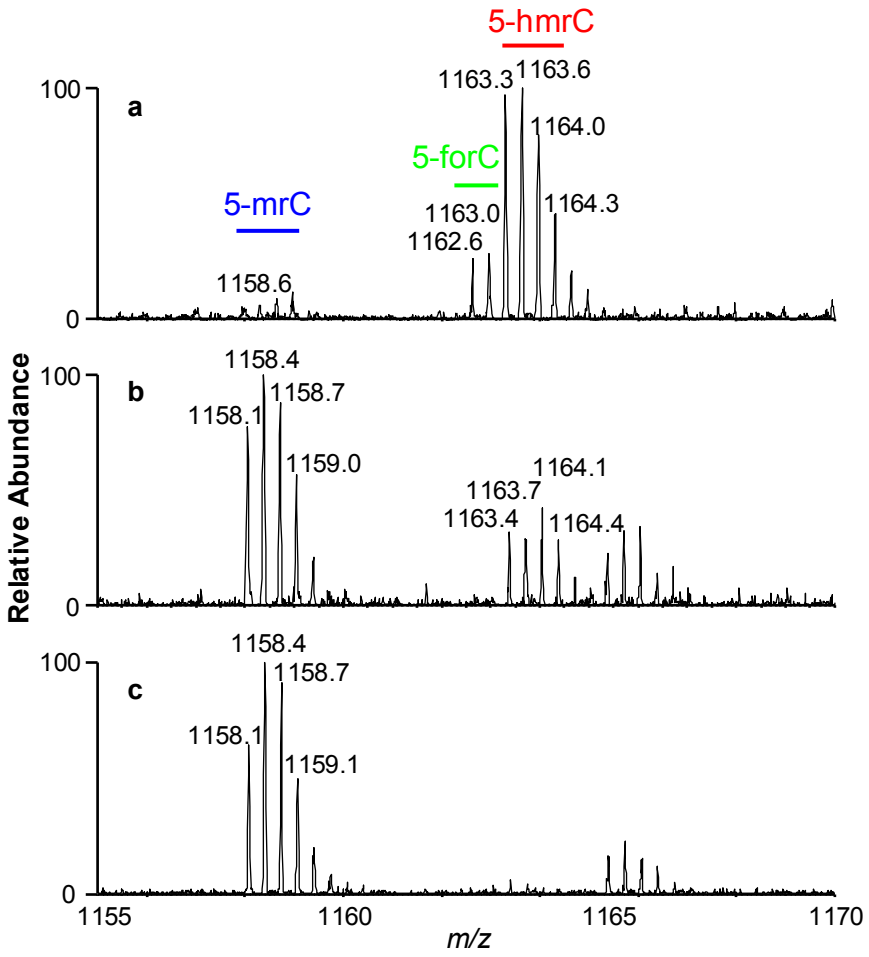

Figure 3.15 LC-MS for monitoring the Tet1-mediated oxidation of 5-mrC in a single-stranded RNA, AGCUC(5-mrC)GGUCA in complete Tet1 reaction buffer (a), or in the same buffer without the addition of Fe^{2+} (b) or 2-oxoglutarate (c). Shown are the higher-resolution "ultra-zoom-scan" MS results for monitoring the $[M-3H]^{3-}$ ions of the initial 5-mC-bearing 11mer RNA, together with their oxidation products, where the 5-mrC is oxidized to 5-HmrC or 5-ForC.

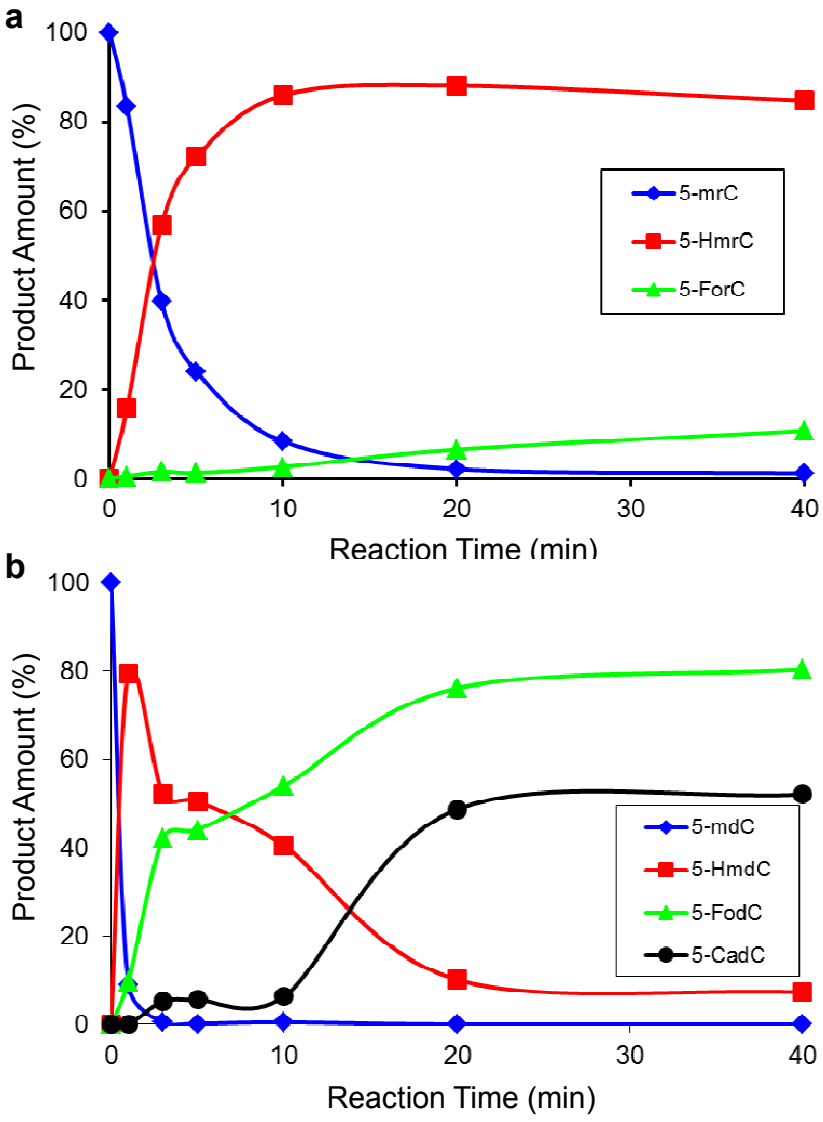

Figure 3.16. Time-dependent formation of oxidation products of 5-mrC in single-stranded RNA, AGCUC(5-mrC)GGUCA (a), and of 5-mdC in single-stranded DNA, d(AGCTC(5-mdC)GGTCA) (b).

LC-MS/MS/MS analysis of 5-hmrC and 5-forC in total RNA

To further examine if the oxidation also happened in vivo, we first set out to develop LC-MS/MS/MS combined with stable isotope-dilution methods for the accurate assessment of levels of 5-HmrC in total RNA. The co-elution of the analyte with the stable isotope-labeled standard at 10.1-10.2 min, together with the similar fragment ions for the analyte and internal standard, allowed for the unambiguous identification of 5-hmrC (Figure 3.17). Similar as what we described previously for the quantification of 5-hmdC in DNA [41], we monitored the fragmentation of the protonated ion of modified nucleobase (i.e., the ion of *m/z* 142, which is the major fragment ion found in the MS/MS of the protonated ion of 5-hmrC) in MS/MS/MS, which displayed the facile loss of a H_2O molecule (i.e., the ion of *m/z* 124, Figure 3.17 top and the inset). The corresponding fragment ion was found for the isotope-labeled standard, with the exception of a 2-Da mass shift introduced by ^{15}N-labeling to the nucleobase portion (Figure 3.17 bottom and the inset).

Subsequently, we developed an LC-MS/MS combined with stable isotope-dilution methods for the accurate assessment of levels of 5-forC on a triple quadrupole mass spectrometer. In this vein, we monitored the fragmentation of the protonated ion of 5-forC in MS/MS, which displayed the loss of the ribose moiety with a transition *m/z* 272.1 → 140.0 (Figure 3.18). The corresponding fragmentation from the protonated isotope-labeled standard was found which showed a loss of the $^{13}C_5$-labeled ribose

moiety (a 5-Da mass shift compared to the unlabeled ribose moiety) and the transition is *m/z* 277.1 → 140.0 (Figure 3.18, inset).

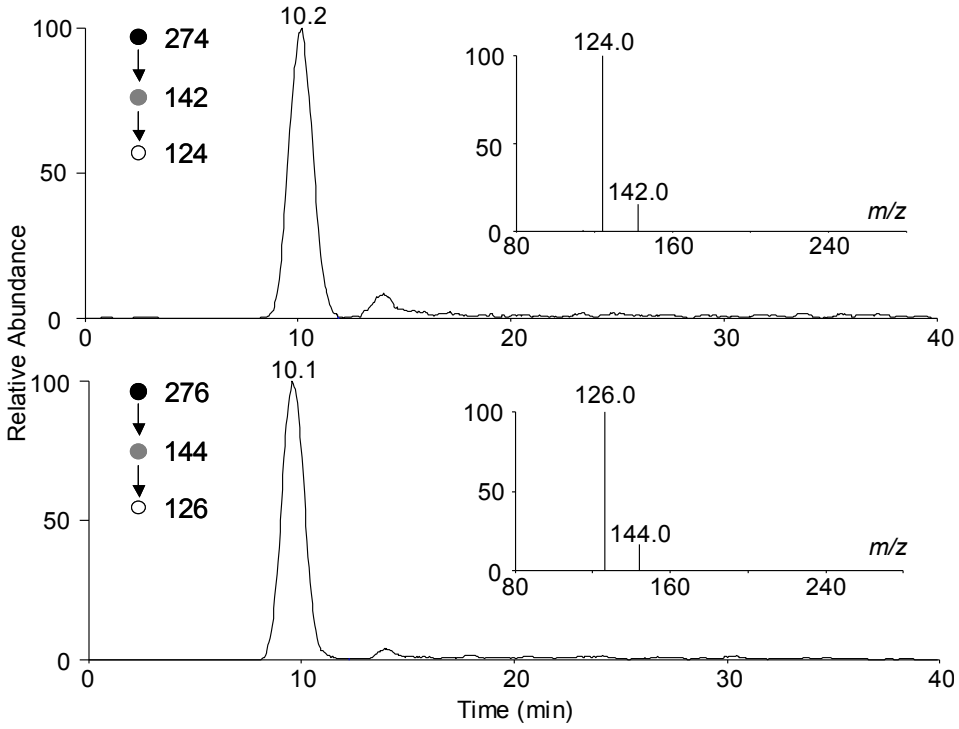

Figure 3.17 Representative LC-MS/MS/MS results for the quantification of 5-hmrC in cellular and tissue RNA. Shown are the selected-ion chromatograms for monitoring the indicated transitions for the analyte (a) and the isotope-labeled standard (b), and the insets give the corresponding MS/MS/MS for the analyte and internal standard. The RNA sample used for this analysis was from mouse brain.

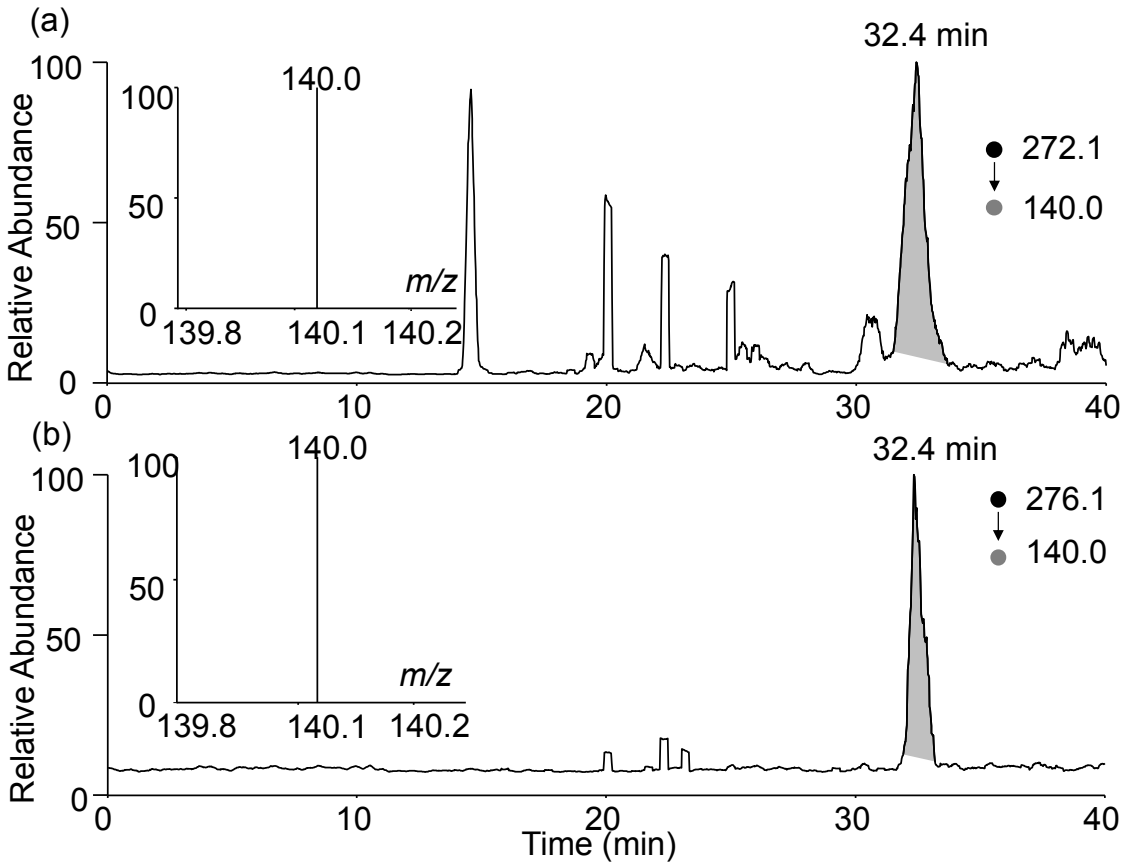

Figure 3.18 Representative LC-MS/MS results for the quantification of 5-forC in total RNA. Shown are the selected-ion chromatograms for monitoring the indicated transitions for the analyte (a) and the isotope-labeled standard (b), and the insets give the corresponding MS/MS for the analyte and internal standard. The RNA sample used for this analysis was from HEK293T overexpressed with Tet3-CD.

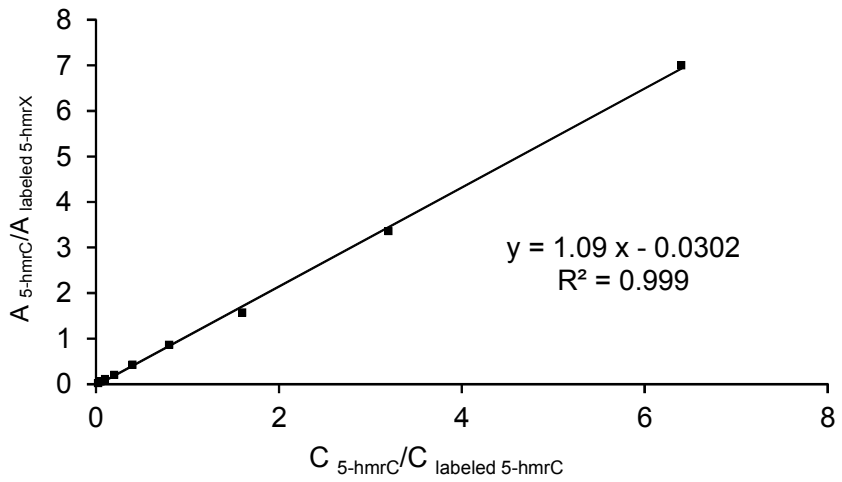

Figure 3.19 Calibration curve for the quantification of 5-hmrC in RNA. The amount of internal standard was 200 fmol, and the amount of unlabeled 5-hmrC ranged from 5 to 1280 fmol.

Tet-mediated formation of 5-HmrC in single-strand RNA *in vivo*

To further assess the function of Tet1 in this oxidation, we overexpressed the catalytic domain of Tet1 (Tet1-CD) or its inactive mutant (Tet1-m) in HEK293T cells [41], isolated total RNA from the cells, digested it to mononucleosides, and quantified the levels of 5-hmrC in the resulting nucleoside mixture by using LC-MS/MS/MS with the isotope dilution method (Figures 3.17 & 3.19). Our LC-MS/MS/MS quantification results revealed that the catalytic activity of Tet1 conferred a marked elevation in the level of 5-hmrC, as the RNA samples isolated from HEK293T cells transfected with wild-type Tet1 carried significantly higher levels of 5-hmrC (11.9 modifications per 10^6 ribonucleosides) than those isolated from cells transfected with the mutant form of Tet1 or a control pGEM-T vector (at 2.0 and 1.9 modifications per 10^6 ribonucleosides, respectively. (Figure 3.20a) Likewise, overexpression of the catalytic domains of Tet2 and Tet3 also led to significant elevations in the levels of 5-hmrC in HEK293T cells (Figure 3.20a).

Considering that other domains of Tet proteins may also be involved in regulating their substrate accessibility, we next assessed the levels of 5-hmrC in cells overexpressing individually the three full-length Tet proteins. Indeed our results demonstrated that the overexpression of full-length Tet3, but not Tet1 or Tet2, could result in substantially elevated level of 5-hmrC in RNA, where the levels of 5-hmrC were 4.1 and 1.8 modifications per 10^6 nucleosides in HEK293T cells overexpressing the full-length Tet3 and its catalytically inactive mutant, respectively (Figure 3.21a). In this regard, it is

important to note that all three full-length Tet proteins are functional, as manifested by marked increases in the levels of 5-hmdC in genomic DNA isolated from cells overexpressing any of the three full-length Tet proteins (Figure 3.21b). Along this line, it is worth noting that Tet1 and Tet2 are localized in the nucleus, whereas Tet3 is localized in both the cytosol and the nucleus [42].

To further exploit the roles of Tet enzymes in inducing 5-hmrC *in vivo*, we measured the levels of 5-hmrC in total RNA isolated from wild-type mouse embryonic stem (ES) cells and Tet-null ES cells where Tet1, Tet2, and Tet3 were genetically deleted ($Tet^{-/-}$). Our results demonstrated that removal of all three Tet activities led to a significant decline in the level of 5-hmrC in total RNA (from 1.4 to 0.82 modifications per 10^6 ribonucleosides, Figure 3.22), whereas knockout of the thymine DNA glycosylase gene ($Tdg^{-/-}$) did not lead to apparent change in 5-hmrC level (Figure 3.22). The relatively small difference in the levels of 5-hmrC in the wild-type and $Tet^{-/-}$ ES cells is in line with the relatively low level of expression of Tet3 in ES cells [12]. In addition, the presence of appreciable levels of 5-hmrC in $Tet^{-/-}$ ES cells suggests that other enzyme(s) might also be involved in oxidizing 5-mC to 5-hmrC in mammalian cells, though we cannot formally exclude the possibility that some of the 5-hmrC may also be induced by cellular reactive oxygen species. Together, the above results support that Tet enzymes contribute to the oxidation of 5-mrC in RNA to 5-hmrC *in vivo*.

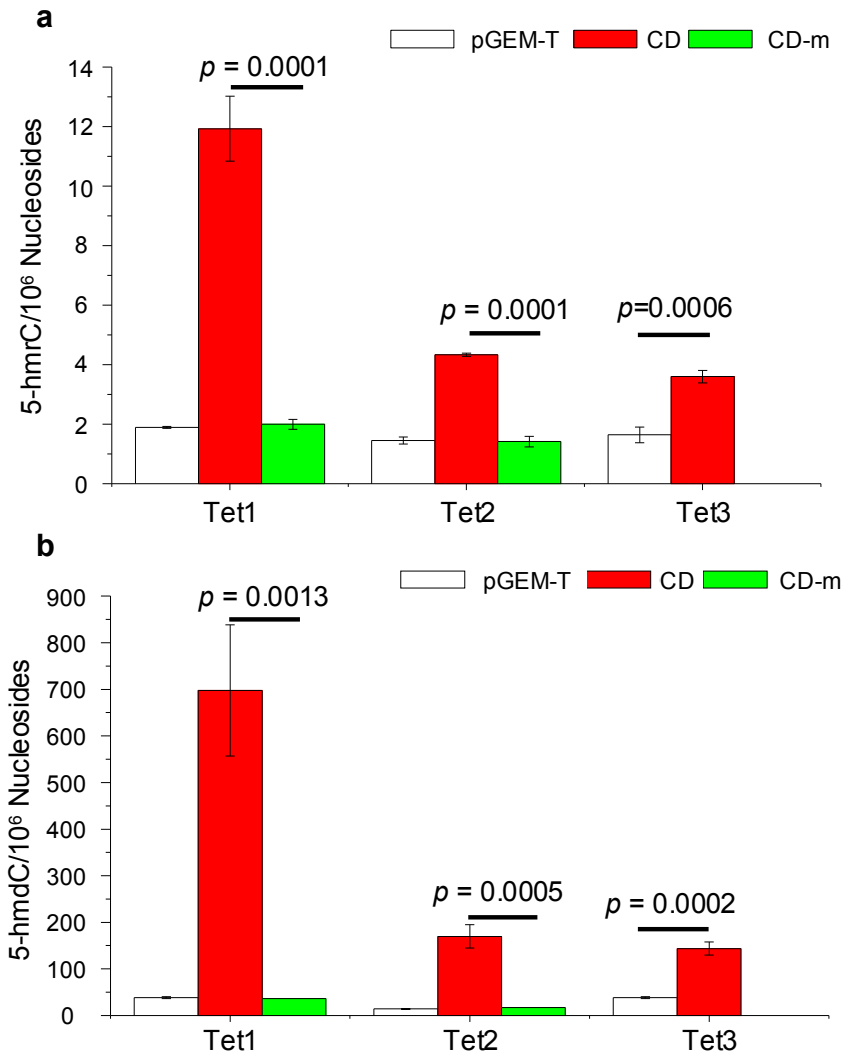

Figure 3.20 The levels of 5-hmrC and 5-hmdC in HEK293T cells overexpressing individually the catalytic domain (CD) of Tet proteins, or their catalytically inactive mutants (CD-m). 'pGEM-T' refers to DNA samples from HEK293T cells transfected with the control pGEM-T Easy plasmid. The data represent the means and standard deviations of three independent transfection and measurement results. The *p* values were calculated using unpaired two-tailed Student's *t*-test.

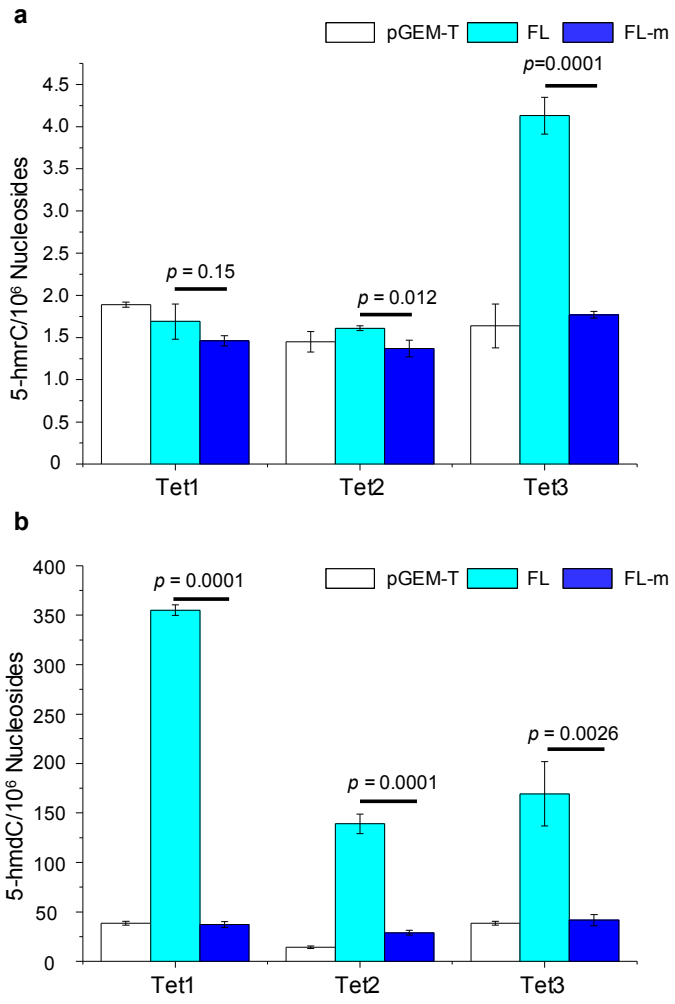

Figure 3.21 The levels of 5-hmrC and 5-hmdC in HEK293T cells overexpressing individually the full-length (FL) Tet proteins, or their catalytically inactive mutants (FL-m). 'pGEM-T' refers to DNA samples from HEK293T cells transfected with the control pGEM-T Easy plasmid. The data represent the means and standard deviations of three independent transfection and measurement results. The *p* values were calculated using unpaired two-tailed Student's *t*-test.

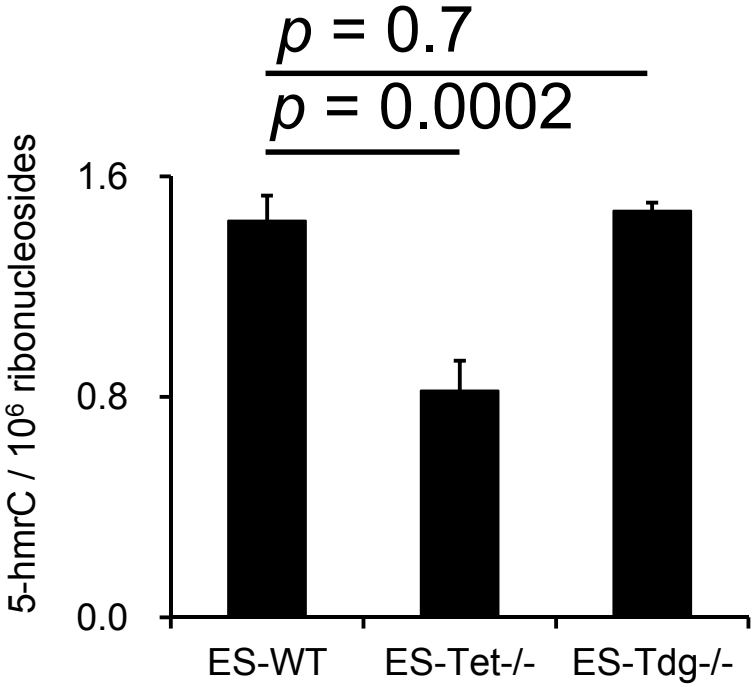

Figure 3.22 The quantification results for the levels of 5-hmrC in wild-type, *Tet*-null and *Tdg*$^{-/-}$ mouse ES cells. The data represent the mean and standard deviation of three measurement results. The *P* values were calculated using unpaired two-tailed *t*-test.

Detection of 5-forC in single-strand RNA *in vivo*

According to the previous *in vitro* and *in vivo* results for 5-hmrC, we conclude that the formation of 5-forC requires more Tet enzymes and Tet3 could be the enzymes mediating the demethylation pathway of 5-mrC *in vivo*. Our *in vitro* results showed that 5-forC and 5-carC were significantly generated with more Tet1 enzymes (Figure 3.10 & 3.11. To further investigate the Tet-mediated oxidation, we overexpressed the catalytic domain of Tet3 (Tet3-CD), full-length Tet3 (Tet3-FL) and its inactive mutant (Tet3-FL-m) in HEK293T cells with more plasmids (7.5 μg). Subsequently, we isolated total RNA from the cells, digested it to mononucleosides, and quantified the levels of 5-forC in the resulting nucleoside mixture by using LC-MS/MS with the isotope dilution method. Our LC-MS/MS quantification results showed that no significant increase in 5-forC was observed in the total RNA isolated from HEK293T cells overexpressing Tet3-FL compared to the cells overexpressing with Tet3-FL-m (Figure 3.23). It is worth noting that only a 4-folds increase in the level of 5-hmrC was observed in cells transfected with 7.5 μg Tet3-FL plasmids and a 2-fold increase in the level of 5-hmrC was observed in cells transfected with 1.5 μg Tet3-FL plasmids, suggesting that Tet3 may not be expressed in cells at a higher level, or 5-hmrC and 5-forC are intermediates which may be rapidly converted to unmethylated rC after their formation. Currently we are not able to detect 5-carC in cells because of the poor sensitivity of the detection method and the extremely low trapping efficiency for this modified nucleoside on HPLC column.

It is worth noting that the levels of 5-hmrC in total RNA isolated from HEK293T cells overexpressing Tet3-FL or Tet3-FL-m are significantly higher than those in total RNA from HEK293T cells transfected with a control plasmid (pGEM-T). These results suggest that the enzymatic activity of Tet3 in conversion from 5-hmrC to 5-forC may require other domains in addition to the catalytic domain. Nevertheless, the experiment needs to be repeated and further confirmed. In the future, we will measure the level of 5-forC in total RNA isolated from wild-type mouse embryonic stem (ES) cells and Tet-null ES cells, which may provide additional insight into the role of Tet proteins in the oxidation of 5-mrC to 5-forC in cells.

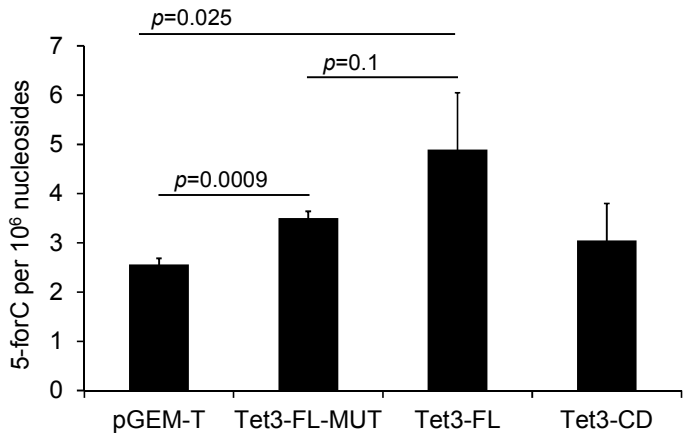

Figure 3.23 The levels of 5-forC in HEK293T cells overexpressing the catalytic domain (CD) of Tet3 proteins (Tet3-CD), the full-length Tet 3 proteins (Tet3-FL), or the catalytically inactive mutants of full-length Tet 3 proteins (Tet3-FL-m). 'pGEM-T' refers to DNA samples from HEK293T cells transfected with the control pGEM-T Easy plasmid. The data represent the means and standard deviations of results from three independent transfection and measurements. The p values were calculated using unpaired two-tailed Student's t-test.

Quantification of 5-HmrC in mammalian tissues and cultured cancer cells

Having demonstrated the enzymatic activity of Tet1 toward 5-mC in RNA, we next assessed the occurrence of 5-hmrC in RNA isolated from various mouse and human tissues by using LC-MS/MS/MS (Figure 3.24). In this vein, it is of note that 5-hmrC was previously detected in rRNA isolated from wheat seedlings [43, 44]. Our results showed that 5-hmrC could be readily detected in RNA samples isolated from all the tissue types we tested, including brain, heart, pancreas, and spleen, with the level being the highest in the heart (3.9 modifications per 10^6 ribonucleosides, Figure 3.24a). In addition, 5-hmrC could be detected in human brain RNA at a frequency of 1.4 per 10^6 ribonucleosides (Figure 3.24a). 5-hmrC could also be found in cultured human cancer cells, including the HeLa cervical cancer cells (at 0.68 modifications per 10^6 ribonucleosides) and WM-266-4 melanoma cells (1.6 modifications per 10^6 ribonucleosides, Figure 3.24b).

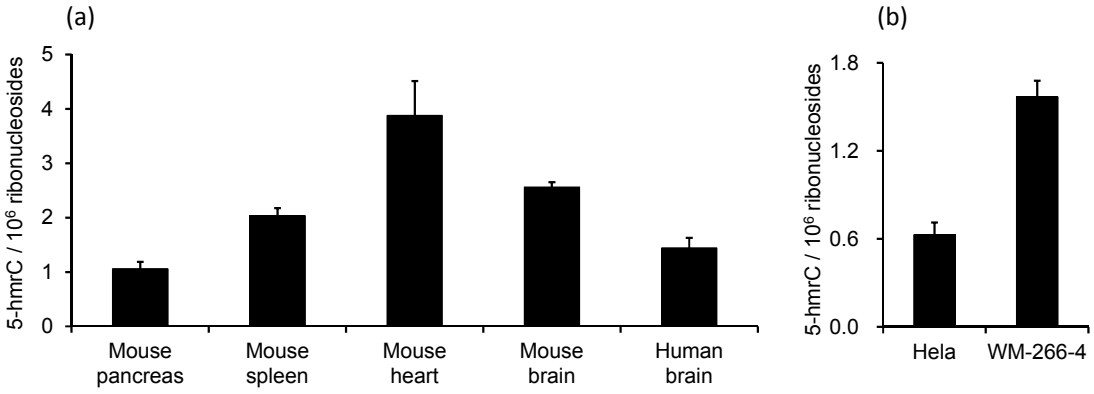

Figure 3.24 (a) Quantification results for the level of 5-hmrC in wild-type human brain tissue and different wild-type mouse tissue (n=3). (b) Quantification results for the level of 5-hmrC in Hela and WM-266-4 cells (n=3). The data represent the mean and standard deviation of the measurement results. The *P* values were calculated using unpaired two-tailed *t*-test.

Discussion

Taken together, we demonstrate that Tet enzymes can catalyze the formation of 5-hmrC from 5-mrC both *in vitro* and *in vivo*. We also determined, for the first time, the levels of 5-hmrC in tissue and cellular RNA by using a sensitive and accurate LC-MS/MS/MS with the isotope-dilution method. Our results revealed that the level of this modification occurs at a frequency of approximately one 5-hmrC per 5,000 5-mrC. Recent bisulfite sequencing data showed the widespread presence of 5-mrC in both coding and non-coding RNA [45]. The presence of appreciable level of 5-hmrC in cellular RNA and the involvement of Tet-family enzymes in inducing this modification suggest that the function of Tet enzymes is not restricted to the epigenetic regulation at the DNA level, but perhaps can also be extended to RNA. In addition, 5-hmrC may also participate in the epigenetic regulation of gene expression. The present work sets the stage for future studies in defining the distribution and site-specific localization of 5-hmrC in different RNA species (i.e., rRNA, mRNA and tRNA), and the function of this 5-mrC oxidation in RNA biology. The relative levels of 5-hmrC in RNA are lower than those of 5-hmdC in DNA. It will be of particular importance to determine whether the 5-hmrC is a stable oxidation product or occurs transiently, possibly as an intermediate step in a pathway leading towards 5-mC decay in RNA, or is perhaps a signal that mediates RNA degradation. Both scenarios could explain the relatively low level of this modification at steady state *in vivo*.

We also successfully detected the 5-forC in total RNA from HEK293T cells and the basal level of 5-forC is around 2.4 per 10^6 nucleosides, which is comparable to the level of 5-hmrC. Considering that the level of 5-fodC is 50 times lower than the level of 5-hmdC in genomic DNA [35], the relatively high level of 5-forC in total RNA may not only serve as an intermediate derivative of Tet-mediated 5-mrC demethylation pathway, but also be involved in other epitranscriptomic processes.

References

1. Cantara, W.A., et al., The RNA Modification Database, RNAMDB: 2011 update. Nucleic Acids Res., 2011. 39(Database issue): p. D195-201.

2. Jia, G., et al., N^6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol., 2011. 7(12): p. 885-7.

3. Zheng, G., et al., ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell, 2013. 49(1): p. 18-29.

4. Jia, G., Y. Fu, and C. He, Reversible RNA adenosine methylation in biological regulation. Trends Genet., 2013. 29(2): p. 108-15.

5. Schwartz, S., et al., High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell, 2013. 155(6): p. 1409-21.

6. Motorin, Y., F. Lyko, and M. Helm, 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. Nucleic Acids Res., 2010. 38(5): p. 1415-30.

7. Squires, J.E., et al., Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res., 2012. 40(11): p. 5023-33.

8. Amort, T., et al., Long non-coding RNAs as targets for cytosine methylation. RNA Biol., 2013. 10(6): p. 1003-8.

9. Hussain, S., et al., NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep., 2013. 4(2): p. 255-61.

10. Pfaffeneder, T., et al., The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed. Engl., 2011. 50: p. 7008-7012.

11. Tahiliani, M., et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, 2009. 324(5929): p. 930-5.

12. Ito, S., et al., Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature, 2010. 466(7310): p. 1129-33.

13. He, Y.F., et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science, 2011. 333(6047): p. 1303-7.

14. Ito, S., et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science, 2011. 333(6047): p. 1300-3.

15. Kriaucionis, S. and N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science, 2009. 324(5929): p. 929-30.

16. Warren, R.A., Modified bases in bacteriophage DNAs. Annu. Rev. Microbiol., 1980. 34: p. 137-58.

17. Liutkeviciute, Z., et al., Cytosine-5-methyltransferases add aldehydes to DNA. Nat. Chem. Biol., 2009. 5(6): p. 400-2.

18. Yildirim, O., et al., Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell, 2011. 147(7): p. 1498-510.

19. Frauer, C., et al., Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS One, 2011. 6(6): p. e21306.

20. Spruijt, C.G., et al., Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell, 2013. 152(5): p. 1146-59.

21. Maiti, A. and A.C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chem., 2011. 286(41): p. 35334-8.

22. Ko, M., et al., Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature, 2010. 468(7325): p. 839-43.

23. Jin, S.G., et al., 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer Res., 2011. 71(24): p. 7360-5.

24. Chen, M.L., et al., Quantification of 5-methylcytosine and 5-hydroxymethylcytosine in genomic DNA from hepatocellular carcinoma tissues by capillary hydrophilic-interaction liquid chromatography/quadrupole TOF mass spectrometry. Clin. Chem., 2013. 59(5): p. 824-32.

25. Dunwell, T.L., et al., The mysterious presence of a 5-methylcytosine oxidase in the Drosophila genome: Possible explanations. Cell Cycle, 2013. 12(21): p. 3357-65.

26. Aas, P.A., et al., Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature, 2003. 421(6925): p. 859-63.

27. Falnes, P.O., et al., Substrate specificities of bacterial and human AlkB proteins. Nucleic Acids Res, 2004. 32(11): p. 3456-61.

28. Lee, D.H., et al., Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. J. Biol. Chem., 2005. 280(47): p. 39448-59.

29. Westbye, M.P., et al., Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. J. Biol. Chem., 2008. 283(36): p. 25046-56.

30. Jia, G., et al., Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. FEBS Lett., 2008. 582(23-24): p. 3313-9.

31. Yi, C., C.G. Yang, and C. He, A non-heme iron-mediated chemical demethylation in DNA and RNA. Acc. Chem. Res., 2009. 42(4): p. 519-29.

32. Hu, L., et al., Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell, 2014. 155(7): p. 1545-55.

33. Guo, J.U., et al., Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell, 2011. 145(3): p. 423-34.

34. Hu, X., et al., Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell, 2014. 14(4): p. 512-22.

35. Liu, S., et al., Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. Nucleic Acids Res, 2013. 41(13): p. 6421-9.

36. LaFrancois, C.J., J. Fujimoto, and L.C. Sowers, Synthesis and characterization of isotopically enriched pyrimidine deoxynucleoside oxidation damage products. Chem Res Toxicol, 1998. 11(1): p. 75-83.

37. Liu, Z., Y. Gao, and Y. Wang, Identification and characterization of a novel cross-link lesion in d(CpC) upon 365-nm irradiation in the presence of 2-methyl-1,4-naphthoquinone. Nucleic Acids Res, 2003. 31(18): p. 5413-24.

38. Liu, S., et al., Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. Nucleic Acids Res., 2013. 41: p. 6421-9.

39. Hashimoto, H., et al., Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature, 2013. 506(7488): p. 391-5.

40. Kizaki, S. and H. Sugiyama, CGmCGCG is a versatile substrate with which to evaluate Tet protein activity. Org. Biomol. Chem., 2014. 12(1): p. 104-7.

41. Liu, S., et al., Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. Nucleic Acids Res., 2013. 41: p. 6421-6429.

42. Arioka, Y., et al., Activation-induced cytidine deaminase alters the subcellular localization of Tet family proteins. PLoS One, 2012. 7(9): p. e45031.

43. Racz, I., I. Kiraly, and D. Lasztily, Effect of light on the nucleotide composition of rRNA of wheat seedlings. Planta, 1978. 142(3): p. 263-7.

44. Limbach, P.A., P.F. Crain, and J.A. McCloskey, Summary: the modified nucleosides of RNA. Nucleic Acids Res., 1994. 22(12): p. 2183-96.

45. Squires, J.E., et al., Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res, 2012. 40(11): p. 5023-33.

Chapter 4

ALKBH3 catalyzes the oxidation of 5-methylcytidine to 5-hydroxymethylcytidine in RNA

Introduction

Ten-eleven translocation family proteins (TET) are Fe(II)- and 2-oxoglutarate-dependent dioxygenases. DNA cytosine methylation can be reversed through TET-mediated oxidation pathway in which the m^5C was first oxidized to 5hmC, then to 5-formylcytosine (5fC), and 5-carboxylcytosine (5-caC) [1-3]. In Chapter 3, we have demonstrated that m^5C in RNA can be oxidized to form 5-hmrC by TET proteins both *in vitro* and *in vivo*. Furthermore, TET proteins can catalyze the stepwise oxidation from 5-hmrC to 5-forC and 5-carC *in vitro*. However, the presence of appreciable levels of 5-hmrC in *Tet*$^{-/-}$ ES cells and in species that do not possess detectable 5-hmdC in their DNA and lack TET homologues in their genomes, such as *C. elegans* and *A. thaliana.*[4], suggests that other enzyme(s) might also be involved in oxidizing m^5C to 5-hmrC in RNA from mammalian cells.

Human ALKBH-family proteins, a subfamily of Fe(II)- and α-KG-dependent dioxygenases, contain nine homologues: ALKBH1-8 and FTO [5-8]. Five of them have been confirmed to contain dioxygenase activity on RNA methylations. ALKBH2 prefers double-stranded DNA (dsDNA) substrates over single-stranded DNA (ssDNA) ones and was found to be primarily responsible for repairing 1-meA base lesions in genomic DNA [9]. However, a moderate demethylase activity of ALKBH2 toward

single-stranded RNA was also reported (10-50 fold lower than toward dsDNA) [10]. ALKBH3 was shown to repair 1-methyladenine and 3-methylcytosine in RNA [11]. FTO and ALKBH5 have been reported to remove m^6A in mRNA [12, 13]. In addition, ALKBH8 was found to oxidize 5-methoxycarbonylmethyluridine (mcm^5U) to 5-methoxycarbonylhydroxymethyluridine ($mchm^5U$) in tRNA [14, 15]. Because the substrates of ALKBH2, ALKBH3, FTO and ALKBH5 are methyl group on the nucleobase moiety, we select these four proteins to test their demethylase activity towards m^5C in RNA.

Experiment Section

Materials

All chemicals and enzymes, unless otherwise specified, were purchased from Sigma-Aldrich (St. Louis) or New England Biolabs (Ipswich). The HEK293T human embryonic kidney cells and cell culture reagents were purchased from ATCC (Manassas). Expression vectors for the catalytic domain of human FTO (amino acids 223-1740) and its corresponding catalytically inactive mutant (H231A/D233A) were previously described [12]. Expression vectors for the catalytic domain of human ALKBH5 (amino acids 692-1876) and its corresponding catalytically inactive mutant (with the deletion of the amino-terminal 66 amino acids) were previously described [13]. Mouse embryonic fibroblast (MEF) cells and MEF cells with the depletion of *Alkbh2* or *Alkbh3* gene were described elsewhere [16].

Cell culture, transfection and RNA extraction

MEF and HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (ATCC). All cells were incubated at 37°C in 5% CO_2 atmosphere. The culture medium was supplemented with 10% fetal bovine serum and 100 IU/mL penicillin. The HEK293T cells were seeded in 6-well plates at 50-60% confluence level and, at 24 hrs later, the cells were transfected with 1.5 μg

or 6 μg plasmid for overexpressing the catalytic domain, or the corresponding catalytically inactive mutants, of the FTO and ALKBH5 enzymes, using Lipofectamine 2000 (Invitrogen). Control experiments were also conducted by transfecting cells with pGEM-T Easy plasmid (Pormega). The cells were harvested for RNA extraction at 48 hrs after plasmid transfection.

Isolation and Digestion of total RNA and mRNA

The isolation and enzymatic digestion of total RNA and mRNA samples for mrC and mrA analysis followed the same description as described in Chapter 2. The digestion of RNA samples for 5-hmrC analysis followed the previous description in Chapter 3. The quality of mRNA was analyzed using an Agilent 2100 Bioanalyzer equipped with an RNA Pico-Chip.

HPLC enrichment of 5-hmrC

HPLC analysis was performed on a Beckman HPLC system with pump module 125 and a UV detector (module 126), and a 4.6×250 mm Alltima HP C18 column (5 μm

in particle size, Grace Davison, Deerfield, IL) was used, following the conditions described in detail in Chapter 3.

LC-MS^3 Analyses of m^{5}C, Cm, m^{6}A and Am

LC-MS^3 measurements were conducted on an LTQ XL linear ion trap mass spectrometer equipped with a nanoelectrospray ionization source and coupled to an EASY-nLC II (Thermo Fisher Scientific, San Jose, CA, USA), as described in Chapter 2.

LC-MS^3 Analysis of 5-hmrC

LC-MS^3 measurements of 5-hmrC were conducted on an LTQ linear ion-trap mass spectrometer equipped with a regular electrospray ionization source and coupled to a micro-flow rate Agilent 1100 capillary HPLC pump (Agilent Technologies), following the description in Chapter 3.

Results

FTO and ALKBH5 showed no activity on the formation of 5-hmrC

We first tested the hydryoxylase activity of FTO and ALKBH5 towards m^5C in RNA *in vivo*. We overexpressed the catalytic domain of FTO and ALKBH5 or their inactive mutants (FTO-m and ALKBH5-m) in HEK293T cells, isolated total RNA from the cells, digested them to mononucleosides, and quantified the levels of 5-hmrC in the resulting nucleoside mixture by using LC-MS/MS/MS with the isotope-dilution method. Our LC-MS/MS/MS quantification results revealed that overexpression of FTO or ALKBH5 in HEK293T cells did not induce an elevation in the level of 5-hmrC. As shown in the figure 4.1, the RNA samples isolated from HEK293T cells transfected with wild-type FTO carried similar levels of 5-hmrC as those isolated from cells transfected with the mutant form of FTO (both at 1.6 modifications per 10^6 ribonucleosides), which is comparable to the basal level of 5-hmrC in HEK293T cells. Likewise, overexpression of the catalytic domains of ALKBH5 did not lead to an increase in the level of 5-hmrC in HEK293T cells.

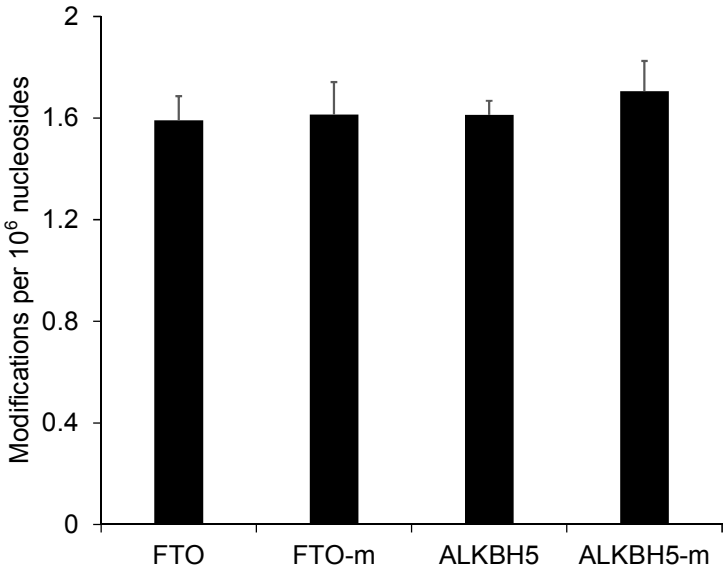

Figure 4.1 The levels of 5-hmrC in total RNA isolated from HEK293T cells overexpressing individually the catalytic domain of FTO and ALKBH5, or their catalytically inactive mutants (FTO-m and ALKBH5-m). The data represent the means and standard deviations of three independent transfection and measurement results.

Demethylase activity of ALKBH3 on m^5C in total RNA

Next, we assessed the capability of ALKBH2 and ALKBH3 in hydroxylating m^5C in RNA. To this end, we measured the levels of 5-hmrC in total RNA isolated from wild-type MEF cells (WT), ALKBH2 knockout MEF cells (*Alkbh2*$^{-/-}$), and ALKBH3 knockout MEF cells (*Alkbh3*$^{-/-}$) [16]. Our results demonstrated that removal of ALKBH3 led to a significant decline in the level of 5-hmrC in total RNA, whereas knockout of the ALKBH2 did not lead to an apparent change in the level of 5-hmrC (Figure 4.2). To further confirm the role of ALKBH3 in the removal of m^5C in RNA, we measured the levels of m^5C in total RNA isolated from WT, *Alkbh2*$^{-/-}$, and *Alkbh3*$^{-/-}$ cells. Our results indicated that the level of m^5C in total RNA isolated from *Alkbh3-/-* cells was significantly increased, but not in the total RNA from *Alkbh2*$^{-/-}$ cells, compared to the level in the total RNA from WT cells (Figure 4.3, b). Meanwhile, we monitored the levels of Cm, m^6A and Am in total RNA isolated from those three cell lines. Surprisingly, we found that the levels of Cm in the total RNA from *Alkbh3*$^{-/-}$ cells was significantly higher than that isolated from WT cells (Figure 4.3, b), whereas the levels of m^6A and Am are similar in the total RNA isolated from the WT and *Alkbh3*$^{-/-}$ cells (Figure 4.3, a). Our data support that ALKBH3 may contribute to the oxidation of m^5C in total RNA to 5-hmrC *in vivo*. In addition, our results suggest that ALKBH3 may also be involved in the removal of Cm in total RNA *in vivo*.

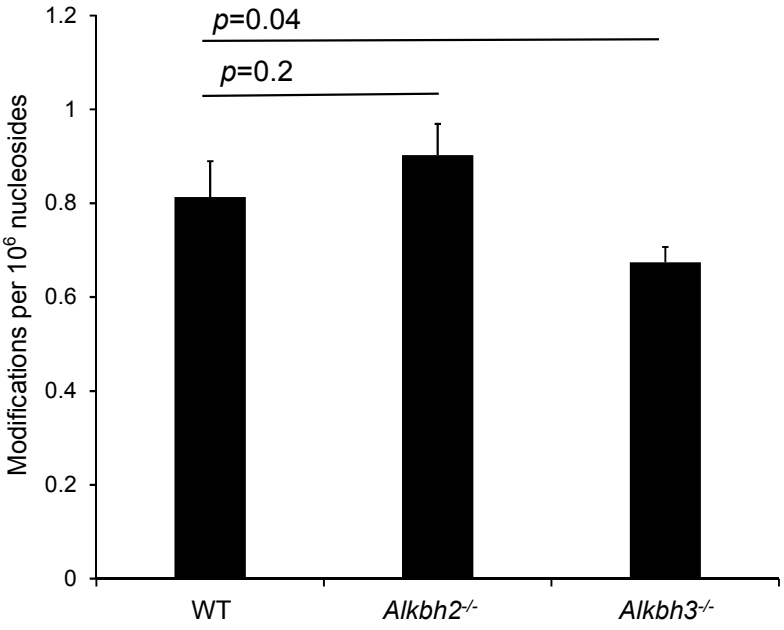

Figure 4.2 The quantification results for the levels of 5-hmrC in total RNA isolated from wild-type, *Alkbh2*$^{-/-}$ and *Alkbh3*$^{-/-}$ MEF cells. The data represent the mean and standard deviation of the measurement results from three biological replicates. The P values were calculated using unpaired two-tailed t-test.

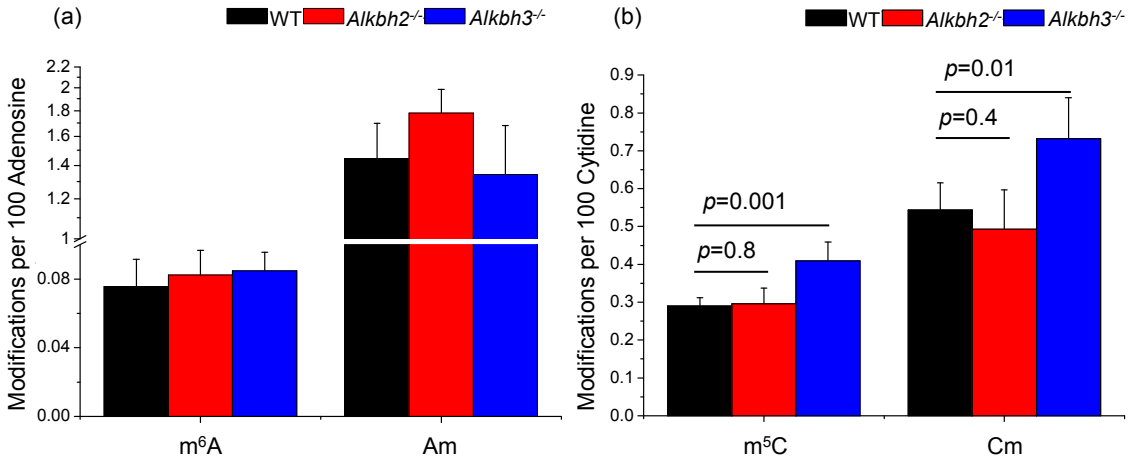

Figure 4.3 The quantification results for the levels of m^5C, Cm, m^6A and Am in total RNA samples isolated from wild-type, *Alkbh2*$^{-/-}$ and *Alkbh3*$^{-/-}$ MEF cells. The data represent the mean and standard deviation of the measurement results from five biological replicates. The P values were calculated using unpaired two-tailed t-test.

Demethylase activity of ALKBH3 on m^5C in mRNA

We also measured the levels of m^5C in mRNA isolated from WT, *Alkbh2*$^{-/-}$, and *Alkbh3*$^{-/-}$ cells. Our results indicated that the levels of m^5C in mRNA isolated from *Alkbh3*$^{-/-}$ and *Alkbh2*$^{-/-}$ cells were increased (Figure 4.4, b). Meanwhile, we found that the levels of Cm in the mRNA isolated from both *Alkbh2*$^{-/-}$ and *Alkbh3*$^{-/-}$ cells were higher than that in the mRNA from WT cells (Figure 4.4, b). In addition, no increase was observed for the levels of m^6A and Am in the mRNA from *Alkbh2*$^{-/-}$ *and Alkbh3*$^{-/-}$ cells compared to the WT cells (Figure 4.4, a). However, the variation of the m^6A in mRNA isolated from *Alkbh3-/-* cells is large. Further experiments are needed to confirm the findings made for mRNA.

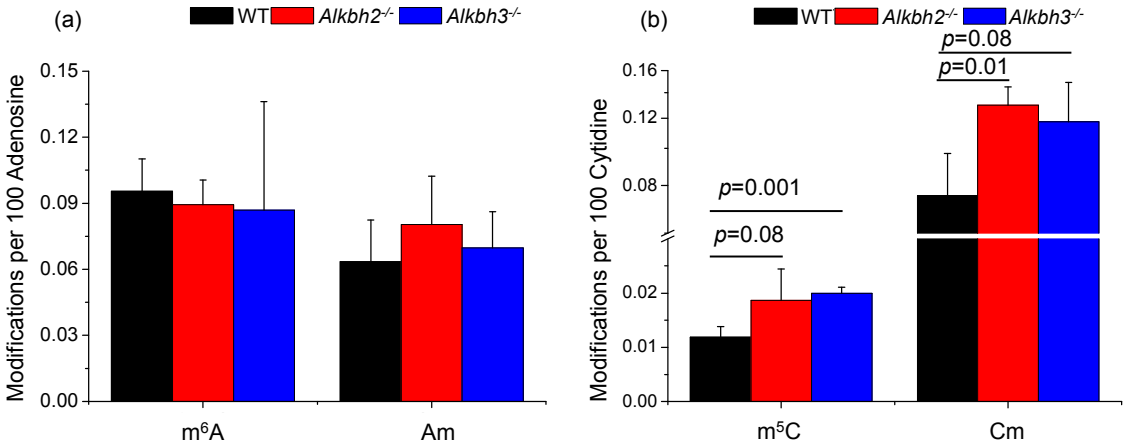

Figure 4.4 The quantification results for the levels of m^5C, Cm, m^6A and Am in mRNA isolated from wild-type, *Alkbh2*$^{-/-}$ and *Alkbh3*$^{-/-}$ MEF cells. The data represent the mean and standard deviation of the measurement results from 3 or 4 biological replicates. The P values were calculated using unpaired two-tailed t-test.

Discussion

In this section, I tested the roles of FTO, ALKBH5, ALKBH2, and ALKBH3 in the induction of 5-hmrC *in vivo*. Our results demonstrated that ALKBH3 was involved in the formation of 5-hmrC in cells. However, further studies should be done to validate this potential important finding. First, biochemical experiments are required to examine whether purified ALKBH3 enzyme is capable of oxidizing m^5C in vitro and if this oxidation is α-KG-and Fe(II)-dependent. Second, it is important to examine whether the level of 5-hmrC in RNA could be increased with the overexpression of the catalytic domain of ALKBH3, but not with the overexpression of inactive mutant form. Third, in order to further confirm the role of ALKBH3 in mRNA, the levels of 5-hmrC in mRNA isolated from wild-type, *Alkbh2*$^{-/-}$ and *Alkbh3*$^{-/-}$ MEF cells should be compared. Meanwhile, the levels of 5-hmrC in mRNA isolated from cells overexpressed with wild-type ALKBH3 and mutant ALKBH3 should also be compared. Last, 5-hmrC is present an appreciated level in total RNA isolated from *Alkbh3*$^{-/-}$ MEF cells. It is possible that the hydroxylase activity of ALKBH3 towards m^5C is restricted to specific RNA species, e.g. mRNA. Therefore, the distribution of 5-hmrC in different types of RNA species isolated from wild-type and *Alkbh3*$^{-/-}$ MEF cells should be measured and compared.

References

1. Zhang, L., et al., Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol, 2012. 8(4): p. 328-30.

2. Maiti, A. and A.C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem, 2011. 286(41): p. 35334-8.

3. He, Y.F., et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science, 2011. 333(6047): p. 1303-7.

4. Liu, S., et al., Detection of oxidation products of 5-methyl-2'-deoxycytidine in Arabidopsis DNA. PLoS One, 2013. 8(12): p. e84620.

5. Tsujikawa, K., et al., Expression and sub-cellular localization of human ABH family molecules. J Cell Mol Med, 2007. 11(5): p. 1105-16.

6. Kurowski, M.A., et al., Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. BMC Genomics, 2003. 4(1): p. 48.

7. Gerken, T., et al., The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science, 2007. 318(5855): p. 1469-72.

8. Sanchez-Pulido, L. and M.A. Andrade-Navarro, The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. BMC Biochem, 2007. 8: p. 23.

9. Ringvoll, J., et al., Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. EMBO J, 2006. 25(10): p. 2189-98.

10. Lee, D.H., et al., Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. J Biol Chem, 2005. 280(47): p. 39448-59.

11. Aas, P.A., et al., Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature, 2003. 421(6925): p. 859-63.

12. Jia, G., et al., N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol, 2011. 7(12): p. 885-7.

13. Zheng, G., et al., ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell, 2013. 49(1): p. 18-29.

14. van den Born, E., et al., ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. Nat Commun, 2011. 2: p. 172.

15. Fu, Y., et al., The AlkB domain of mammalian ABH8 catalyzes hydroxylation of 5-methoxycarbonylmethyluridine at the wobble position of tRNA. Angew Chem Int Ed Engl, 2010. 49(47): p. 8885-8.

16. Nay, S.L., et al., Alkbh2 protects against lethality and mutation in primary mouse embryonic fibroblasts. DNA Repair (Amst), 2012. 11(5): p. 502-10.

Chapter 5

Summary and Future Directions

The rising interest in understanding the functions, regulation and maintenance of the epitranscriptome calls for robust and accurate analytical methods for the identification and quantification of post-transcriptionally modified nucleosides in RNA. Mass spectrometry has become a very powerful tool for bioanalysis which can elucidate the structure of substances and provide quantitative measurements. The LC-MS-based analytical method, in combination with genetic manipulation, may facilitate the studies in the area of epitranscriptome. In this thesis, I focus on the development of novel MS-based strategies to identify and quantify post-transcriptional modifications present in total RNA and mRNA isolated from mammalian tissues and cultured human cells. Additionally, by using these analytical methods, I was able to discover new enzymes involved in demethylation of mono-methylated cytosine in RNA both *in vitro* and *in vivo*.

In Chapter 2, an LC-MS/MS/MS coupled with the stable isotope-dilution method was developed for the sensitive and accurate quantifications of 5-methylcytidine (m^5C), 2'-*O*-methylcytidine (Cm), N^6-methyladenosine (m^6A) and 2′-*O*-methyladenosine (Am) in RNA isolated from mammalian cells and tissues. Our results showed that the distributions of these four methylated nucleosides are tissue-specific. We also found that the levels of m^5C, Cm and Am are significantly lower (by 6.5-43 fold) in mRNA than in total RNA isolated from HEK293T cells, whereas the level of m^6A was slightly higher (by 1.6 fold) in mRNA than in total RNA.

In Chapter 3, I first demonstrated that Tet enzymes can catalyze the formation of 5-hydroxymethylcytidine (5-hmrC), 5-formycytidine (5-forC) and 5-carboxycytidine (5-carC) from m^5C *in vitro*. Subsequently, I established a sensitive and accurate LC-MS/MS/MS with the isotope-dilution method to measure the level of 5-hmrC *in vivo* and further demonstrated that the catalytic domains of all three Tet enzymes as well as full-length Tet3 could induce the formation of 5-hmrC in human cells.

In Chapter 4, I selected four Fe(Π)- and 2-oxoglutarate-dependent dioxygenases, including FTO, ALKBH5, ALKBH2 and ALKBH3, to test their demethylase activity towards m^5C in RNA in human cells by using the analytical methods established in Chapters 2 & 3. Our results showed that, the level of 5-hmrC is significantly decreased whereas the level of m^5C is significantly increased in *Alkbh3*$^{-/-}$ cells. Our results suggested that ALKBH3 was involved in the demethylation of m^5C in RNA.