Structural Basis for Eukaryotic mRNA Modification

Andrew J. Fisher[1,2,*] and Peter A. Beal[1,*]

[1]Department of Chemistry, University of California, One Shields Ave, Davis, CA 95616, USA.

[2]Department of Molecular and Cellular Biology, University of California, One Shields Ave, Davis, CA 95616, USA.

*Correspondence to: ajfisher@ucdavis.edu, pabeal@ucdavis.edu

Highlights:

- All RNA methyltransferases share a common structural core and use *S*-adenosyl methionine as a methyl donor.
- All RNA deaminases exhibit a common structural core and utilize a water-activated zinc ligand and conserved active site residues for catalysis.
- ADAR appears to be unique in that it only modifies nucleotides in double stranded helical structures, through a base-flipping mechanism.

Abstract

All messenger RNAs in eukaryotes are modified co- and post-transcriptionally. They are all capped at the 5′end and polyadenylated at the 3′end. However, many mRNAs are also found to be chemically modified internally for regulation of mRNA processing, translation, stability, and to recode the message. This review will briefly summarize the structural basis for formation of the two most common modifications found at internal sites in mRNAs; methylation and deamination. The structures of the enzymes that catalyze these modifications show structural similarity to other family members within each modifying enzyme class. RNA methyltransferases, including METTL3/METTL14 responsible for N^6-methyladensosine (m^6A) formation, share a common structural core and utilize *S*-adenosyl methionine as a methyl donor. RNA deaminases, including Adenosine Deaminases acting on RNA (ADARs), also share a common structural core and similar signature sequence motif with conserved residues used for binding zinc and catalyzing the deamination reaction. In spite of recent reports of high resolution structures for members of these two RNA-modifying enzyme families, a great deal remains to be uncovered for a complete understanding of the structural basis for mRNA modification. Of particular interest is the definition of factors that control modification site specificity.

Introduction

Post transcriptional modification of RNA is a universal event in molecular biology as it is observed in all three kingdoms of life and present in all types of RNA [1]. The first RNA modified base identified was that of pseudouridine (ψ) in the 1950's [2]. The next two decades saw an exponential growth in the characterization of many different types of RNA modifications [3]**. The most abundant and chemically diverse modified RNA is found in ribosomal RNA (rRNA) and transfer RNA (tRNA) from prokaryotic organisms. Therefore, much of the early effort and analysis in understanding modified RNA nucleotides focused on tRNA and rRNA, where they help stabilize and fine-tune structure and function [1]. Recently, fueled by the advances in mass spectrometry and next-generation sequencing, many more types of nucleoside modifications have been discovered in all classes of RNA from all organisms, including viruses. Currently well over 100 distinct modifications that have been characterized (modomics.genesilico.pl) (mods.rna.albany.edu) [4,5].

Originally thought to be static variants, many RNA modifications can be dynamic and are involved in regulating RNA function and sequence [6]. The modified nucleosides can alter RNA secondary structure by changing hydrogen bonding patterns, affecting base stacking potential, or favoring a specific nucleotide conformation. Some changes can also re-code mRNA by altering the genetic interpretation of the base, which is referred to as "RNA editing." These mRNA "edits" alter the interaction with the tRNA anticodon loop resulting in a translated protein that differs in amino acid sequence predicted from the DNA sequence. Additionally, specific modified RNA bases have been found to recruit "reader" proteins [7,8], which can regulate mRNA location, splicing, and translation efficiency. Therefore, these post-transcriptional modifications referred to as the "epitranscriptome" must be tightly controlled. Indeed, aberrant

regulation of RNA modification has been linked to many different phenotypes and diseases in higher eukaryotes [9,10]*,**.

In eukaryotes, mRNA is transcribed by a large multi-subunit RNA polymerase II complex consisting of 12 subunits in humans [11]. The nascent mRNA is "capped" by addition of guanosine at the 5′-terminal mRNA residue through an unusual 5′,5′-triphosphate linkage [12,13]*. This guanosine gets methylated at the N7 position along with the 2′OH of the next nucleotide. After transcription termination, mRNA is also polyadenylated at the 3′-end with 50-100 adenosines using a large multi-subunit polyadenylate polymerase complex [13,14]. This review will focus on mRNA modification between these ends of mRNA, where the canonical nucleosides are altered by addition of, or exchange of, atoms. The review will specifically center on the two most abundant types of mRNA modification, methylation and deamination catalyzed by methyltransferases (MTase) and deaminases, respectively (Figure 1).

Methylation

The most common type of mRNA modification is methylation, where both the base and 2′-OH group are observed to be methylated in higher eukaryotic mRNA. The methyl donor for all RNA methylation reactions is *S*-adenosyl methionine (SAM), and members of the Methyltransferase class of enzymes (MTase) catalyze the methyl transfer, resulting in methylated nucleotide product and *S*-adenosyl homocysteine (SAH). The structural core of the MTases are all categorized in the SCOP superfamily of S-adenosyl-L-methionine-dependent methyltransferases [15], which contain a three-layer α/β/α core containing a Rossmann fold [16]. All RNA MTases have the same architectural core of a seven-stranded mostly parallel β-sheet with one strand running antiparallel (Figure 2). The topology of all RNA MTases, except human

METTL3/METTL14, are identical to the conserved canonical MTase core, which catalyze methyl transfer to a very broad range of acceptors including; hormones, neurotransmitters, lipids, proteins and nucleic acids [17]. The core of the canonical MTase fold incorporates alternating β-strands and α-helices with a β-strand order of 6-7-5-4-1-2-3, with strand 7 anti-parallel to the other six strands (Figure 2a). β-strands 1-2-3 (with interposed helices) form half of the Rossmann dinucleotide binding domain, with parallel strands 5-4 forming part of the other half (Figure 2a). The SAM molecule for most MTases is situated between the ends of β-strands 4 and 1 (Figure 2).

The human methyl transferase METTL3/METTL14 heterodimer, which produces m^6A in mRNA, has the same architecture as the canonical MTase for each subunit, but the connectivity or topology is different (Figure 2b). It appears the a METTL3/METTL14 ancestral gene may have been circularly permuted so the new N- and C-terminus is now between what was β-strands 2 and 3 of the canonical fold, producing a new topology where the new β-strand 1 structurally corresponds to the old β-strand 3, and the new β-strand 7 is structurally equivalent to the old β-strand 2 (Figure 2b). This ancestral gene duplicated to produce METTL3 and METTL14 both with a β-strand order of 4-5-3-2-6-7-1, with β-strand 5 running anti-parallel. This topology appears to be unique in METTL3/METTL14 because all the other human METTLs (methyltransferase-like proteins) whose structures are known (PDB IDs: 3ckk, 5wjc, 2h00, 4rfq, 4mtl, 4lec, 4lg1) have the canonical MTase fold.

Given the inherent nucleophilic property of nitrogen and oxygen atoms, it has been observed that some nucleic acids can be methylated spontaneously in the presence of SAM alone [18,19], suggesting that the MTases may serve as a catalyst by simply promoting a favorable orientation between the SAM methyl group and the acceptor [20]. However, some MTase

structures have conserved acidic residues in the active site that can serve as a catalytic base to increase the nucleophilicity of the attacking nitrogen or oxygen atoms, which can attack the methyl group of SAM (Figure 2d).

For methyl transfers to less nucleophilic carbon atoms, as in the production of m^5C, the electron-poor carbon-5 in the aromatic ring must be activated by a Michael addition. Like the DNA m^5C MTase reaction [21], an active site cysteine attacks carbon 6 to form a transient intermediate covalently linked to the MTase, creating a C5 nucleophile enamine that can attack the methyl group of SAM (Figure 2e).

m^6A Methyltransferase

Outside the m^7G capping modification of mRNA, the most prevalent type of mRNA modification is methylation of the 6-amino group of adenosine to generate N^6-methyladenosine (m^6A). First identified over 40 years ago [22,23], N6-methylation predominately occurs at the 3´ end of the mRNA and has been detected in mammals, plants, flies, and yeast [24]. Recent evidence reveals that m^6A plays many important roles in development, growth, metabolism, circadian clock regulation, and stem cell differentiation (reviewed in [10,24]*). This is a consequence of m^6A regulating the processing, translation, and stability of cellular mRNA molecules.

Methylation of adenosines in mRNA is believed to be a dynamic process that is governed by three classes of proteins; the "writers" (m^6A MTases), "erasers" (m^6A demethylases), and "readers" (m^6A-binding proteins). In humans, the m^6A MTase is the METTL3/METTL14 heterodimer, while one demethylase is the FTO protein, which has been linked to obesity [25].

The other demethylase is alkB homolog 5 (ALKBH5) [26]. Reader m^6A-binding proteins often contain YTH domains and can be cytoplasmic or nuclear localized [27].

The METTL3/METT14 MTase heterodimer requires the association of the Wilms' tumor 1-associating protein (WTAP) for efficient activity [28]. This METTL3/METT14/WTAP methylates adenosines in a well-defined motif of RRACH (R is A or G, and H can be A, C or U) with a higher abundance of methylation near the stop codons of mRNAs [29,30]. However, not all adenosines in this consensus sequence get methylated suggesting that the primary sequence is likely necessary, but not sufficient for efficient m^6A modification to occur.

Crystal structures were recently determined of the human METTL3/METTL14 heterodimer from two laboratories [31,32], but unfortunately, they only had SAM or SAH bound and no oligonucleotide substrate to clearly define the mRNA binding site. The structures revealed that the two subunits have similar topology and structure to each other but differ from the canonical MTase fold (as discussed above). The METTL3 and METTL4 subunits superimpose together with an rmsd of ~0.8Å over 130 α-carbons. The methyl donor SAM molecule bound in the METTL3 subunit in a similar position to canonical MTases. The structures reveal a patch of positive electrostatic potential across the dimer interface near the active site, which suggests the mRNA would bind in this pocket created by the heterodimer. Near the SAM binding site in METTL3 there is a conserved Asp395, which is part of a conserved DPPW motif, that likely serves as a catalytic base, because altering this residue substantially decreases methyltransferase activity [31].

m^5C Methyltransferase

It has been well established that 5-methylcytosine (m^5C) in DNA plays a role in epigenetic genome activity [33,34]. However relatively recently, m^5C has been found in RNA, both in non-coding and coding RNA [35], however the level m^5C in mRNA is debated [36]. Two MTases were identified to catalyze the methylation of cytosine in mRNA [35]; NSUN2 and DNMT2, both of which also generate m^5C in tRNA [37,38]. The crystal structure is known for human DNMT2, which was originally thought to be a DNA MTase, but in spite having a totally different catalytic mechanism to other RNA MTases, the structure still maintains the canonical MTase fold (Figure 2c) [39]. The catalytic mechanism for the RNA m^5C MTase is similar to the DNA m^5C MTase (Figure 2e) [40], which requires a catalytic cysteine to activate carbon 5 through a Michael addition covalent adduct (Figure 2e).

m^1A Methyltransferase

Very recently a new type of mRNA methylation modification was uncovered, where N1 was found to be methylated in human mRNA to generate the modified nucleotide N^1-Methyladenosine (m^1A) [41,42]. This modification, which is enriched around the start codon, causes the base to be positively charged and adds a bulky methyl group at the Watson-Crick interface affecting base-pairing capabilities. m^1A modification in mRNA was found to be reversible through the demethylase activity of the known DNA/RNA demthylase ALKBH3 [41,42]. While the exact MTase that catalyzes this reaction on mRNA has yet to be identified, several SAM-dependent m^1A MTases are known to catalyze this reaction on tRNA and rRNA [43,44], suggesting one of these or a related MTase may be acting on mRNA.

The crystal structure is known for human TRM61/TRM6 heterodimer, a m^1A MTase that methylates A58 in $tRNA^{Lys}$ [45]. Reminiscent of the METTL3/METTL14 m^6A MTase, both

TRM61 and TRM6 have similar core structures, but with conventional MTase topology. The function of the TRM6 subunit has been firmly established to bind the tRNA molecule. SAM binds to TRM61, which contains a conserved Asp181. This residue is positioned between the bound SAM and tRNA and has been implicated in catalysis.

Deamination

Two bases in mRNA can be hydrolytically deaminated resulting in "editing" of the transcript, due to the fact the modification alters the Watson-Crick hydrogen bonding properties of the nucleotide. Thus, deamination in mRNA can lead to recoding of the message by substituting a different amino acid during translation. Cytidine can be deaminated to generate uridine (C-to-U edit), and adenosine can be deaminated to produce inosine (A-to-I edit) (Figure 1). C-to-U deamination is catalyzed by the enzyme APOBEC (apolipoprotein B mRNA editing catalytic polypeptide-like) and A-to-I edit is catalyzed by ADAR (adenosine deaminase acting on RNA). Although they share limited sequence homology both enzymes have a similar protein structure core consisting of a five-stranded β-sheet and two α-helices that pack against the β-sheet (Figure 4b). Both enzymes also contain an active site zinc ion positioned between the N-terminal end of the two common helices. Both enzymes coordinate the metal in a similar fashion where the active site zinc atom is ligated by two cysteines and one histidine. The fourth ligand is a water molecule that serves as the nucleophile that attacks carbon 4 (cytidine) or 6 (adenosine) of the nucleobase and proceeds through a tetrahedral intermediate (Figure 3). There are two major distinct differences between the two deaminase enzymes, first is the substrate secondary structure specificity, where APOBEC deaminates cytidines in ssRNA, while ADAR deaminates adenosines in regions of dsRNA. The second difference between APOBECs and ADARs, is that

some APOBECS can act on ssRNA and/or ssDNA, while ADAR cannot deaminate adenosines in dsDNA. Finally, one difference between mRNA deamination and methylation modifications, is that unlike many methylation modifications, C-to-U and A-to-I edits are not directly reversible. Reversing the effects of deaminations requires new transcription to generate additional unedited mRNA.

C-to-U Edits

C-to-U editing in mRNA was first identified in apolipoprotein B (apoB) mRNA where the edit occurs only at nucleotides 6666 and 6802 [46,47]. Editing at 6666 changes the glutamine CAA codon to the UAA stop codon resulting in a protein about half the size of full length. The enzyme that catalyzes the apoB mRNA C-to-U edits was identified in 1993 [48], and since then 10 other gene products have been added to the APOBEC family [49]**. However the majority of APOBECs deaminate cytidine in ssDNA, with only two, APOBEC1 and APOBEC3A, known to act on RNA [50]**. Some APOBECs that deaminate DNA may play a role in demethylating m^5C in DNA [51,52]. APOBEC3F, 3G, & 3H have demonstrated anti-viral/HIV properties by editing DNA strands during viral DNA synthesis [53,54].

Crystal structures are known for many APOBECs, but not for APOBEC1 or APOBEC3A, the only two involved in RNA editing. The crystal structures reveal that while the APOBECs only share roughly 15-25% sequence identity, they all have homologous structures to each other, which is similar to the bacterial cytidine deaminase (CDA) that acts on mononucleosides (Figure 3) [55]. The APOBECs have a conserved sequence of HXE-X_{25-30}-PCX_{2-4}C, where the Cys and His residues coordinate the zinc ion and Glu is the catalytic residue. The zinc ion binds between the N-terminal ends of helices α2 and α3 (Figure 3)

where the His and catalytic Glu reside on helix α2 and the two Cys zinc ligands are on helix α3. Variations in length, sequence, and loop dispositions likely dictate the specific APOBECs function and substrate specificity.

The crystal structures have been determined for APOBEC3A and a chimera of APOBEC3A and 3B bound to a short ssDNA oligonucleotide (Figure 3a) [56,57]. To trap the ssDNA bound to the APOBECs, the active site Glu was mutated to Ala. The short oligonucleotides form a sharp bend at the targeted cytidine, which extends deep into the active site with the flanking bases being more exposed to solvent. However, the base on the 5′side of the cytidine does bind in a small cleft between the α1-β1 loop and β4-α4 loop, where an Asp hydrogen bonds to the thymidine base. All the structures reveal that oxygen O2 of the cytidine base hydrogen-bonds to the main chain amide nitrogen of Ala following the conserved zinc-ligating His. It is expected that the catalytic Glu will also hydrogen-bond to the base to confer specificity for cytidine.

A-to-I Edits

Adenosine deamination in RNA was identified around the same time as the C-to-U RNA editing [58]. Over three million A-to-I edits have been identified in the human transcriptome, making it more abundant than C-to-U edits [59]. However, the majority of the A-to-I edits occur in non-coding regions, with roughly 1800 falling in mRNA coding regions. The enzyme that deaminates adenosine is called ADAR (adenosine deaminases that act on RNA), and humans contain three ADAR genes. The ADAR proteins are modular, with all ADARs possessing a homologous catalytic or deamination domain comprising roughly the C-terminal 400 amino acids and N-terminal nucleic acid binding domains. ADAR1 and ADAR2 are expressed in most

tissues, but ADAR3 is enigmatic because it is expressed exclusively in the brain, but appears to have no deaminase activity.

The crystal structure of the deaminase domain of human ADAR2 (ADAR2d) was determined over a decade ago in the absence of RNA [60]*. The structure unexpectedly revealed that ADARs bury an inositol hexakisphosphate (IP_6) in the protein core, detached from the active site. The exact role of IP_6, outside of maintaining a stably folded enzyme, is unknown. The ADAR2d structure contains a seven-stranded mostly parallel β-sheet core with the long α-helix (α1) packed perpendicularly across one side of the β-sheet and eight α-helices on the other side of the β-sheet core, which binds the IP_6 (Figure 4a). At first glance, it appears the ADARs don't resemble APOBEC deaminase structures, but they do share a similar common core comprising a five-stranded β-sheet core together with two α-helices that are involved in zinc binding (Figure 4b). Additionally, the ADARs do have a signature sequence motif similar to APOBECs but with larger insertions. The motif for ADARs is HXE-$X_{\sim 55}$-PC$X_{\sim 65}$C with a large insertion between the catalytic Glu and the first Cys zinc ligand. Additionally, instead of a 2-4 residue spacing between the two Cys zinc ligands in APOBEC, the ADARs contain a much longer insert of roughly 65 residues between the two Cys residues. A fascinating feature of this large insertion is that it contains many residues that interact with the dsRNA substrate, including the base-flipping residue and an RNA-binding loop (Figure 4c) [61]*.

The crystal structure of ADAR2d complexed with a 23 bp dsRNA substrate was recently determined [61]*, which was made possible by substituting the target adenosine with the analogue 8-azanebularine (8-azaN) [62]. The structure revealed the nucleotide analogue is flipped out of the dsRNA duplex by Glu488, where it interacts with both the active site zinc ion and the catalytic residue Glu396. The 8-azaN is also hydrated by the zinc-ligand water molecule

thus mimicking the high-energy tetrahedral intermediate at C6 (Figure 3c). This base-flipping loop approaches the dsRNA from the minor grove where Glu488 intercalates the space vacated by the flipped adenosine and hydrogen bonds to the Watson-Crick face of the complementary-strand orphaned base. This base-flipping loop sequence of Gly-Glu-Gly is conserved in all ADARs including human ADAR3, which has no identified catalytic activity. The action of the base-flipping residue Glu488 and interaction with the orphaned base is reminiscent of a similar mechanism observed for the HhaI DNA MTase, a dsDNA-modifying enzyme that also uses a base-flipping mechanism to access 2′-dC for methylation, but approaches dsDNA from the major groove [63]. The human ADAR2d structure complexed with dsRNA provides a structural basis for substrate specificity dictated by nearest-neighbor preferences. The main chain of glycine residues that flank the base-flipping Glu488 either hydrogen bond to the 2-amino group in the preferred 3′G or occlude a potential 2-amino group of G in the preferred 5'U-A base pair [61]. This contrasts with the APOBECs that have some sequence preferences within each group but vary among different APOBEC families. For example, APOBEC3G prefers the C<u>C</u>C sequence, APOBEC3B favors T<u>C</u> and APOBEC1 is more relaxed in deaminating cytosine after pyrimidines Y<u>C</u> (underlined C is deaminated) [64].

ADARs also contain dsRNA Binding Domains (dsRBDs) of about 70 residues that help direct the ADAR deaminase domain to regions of dsRNA [65]. The NMR structure of ADAR2 dsRBDs bound to dsRNA reveal they can make some sequence-specific contacts to the RNA substrate [66]. The structures suggest that dsRBD bind the minor groove of RNA near the adenosine targeted for deamination. However, additional structural studies are necessary to determine how ADAR deaminase domains and dsRBDs cooperate to bind substrate RNAs.

Conclusions

RNA modifications play many important and diverse functional roles including; directing mRNA to the ribosome to initiate protein translation, stabilizing RNA structure, regulating degradation, increasing the genetic variability in editing, and discriminating self from non-self RNA in the cell [67]. Most modifications in mRNA occur in the untranslated regions (UTR) where they can regulate RNA stability, localization, or protein translation. The majority of mRNA modifications are either methylations or deaminations. While many structures are known for the enzymes that catalyze these mRNA modifications, many still are yet to be determined, and more importantly, a better understanding is demanded to explain how the structure (with potential associated factors) dictates specificity in determining which nucleotides are modified.

References

1. Grosjean H (Ed): *Fine-Tuning of RNA Functions by Modification and Editing*: Springer-Verlag Berlin; 2005.
2. Davis FF, Allen FW: **Ribonucleic acids from yeast which contain a fifth nucleotide**. *J Biol Chem* 1957, **227**:907-915.
3. Grosjean H: **Modification and editing of RNA: historical overview and important facts to remember**. In *Fine-Tuning of RNA Functions by Modification and Editing*. Edited by Grosjean H: Springer Berlin Heidelberg; 2005:1-22. 10.1007/b106848
4. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, et al.: **MODOMICS: a database of RNA modification pathways--2013 update**. *Nucleic Acids Res* 2013, **41**:D262-267.
5. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FA, Fabris D, Agris PF: **The RNA Modification Database, RNAMDB: 2011 update**. *Nucleic Acids Res* 2011, **39**:D195-201.
6. Li S, Mason CE: **The pivotal regulatory landscape of RNA modifications**. *Annu Rev Genomics Hum Genet* 2014, **15**:127-150.
7. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, et al.: **N6-methyladenosine-dependent regulation of messenger RNA stability**. *Nature* 2014, **505**:117-120.
8. Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, Sun HY, Li A, Ping XL, Lai WY, et al.: **Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing**. *Mol Cell* 2016, **61**:507-519.
9. Sarin LP, Leidel SA: **Modify or die?--RNA modification defects in metazoans**. *RNA Biol* 2014, **11**:1555-1567.
10. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM: **The RNA modification landscape in human disease**. *RNA* 2017, **23**:1754-1769.
11. Cramer P: **RNA polymerase II structure: from core to functional complexes**. *Curr Opin Genet Dev* 2004, **14**:218-226.
12. Shatkin AJ: **Capping of eucaryotic mRNAs**. *Cell* 1976, **9**:645-653.
13. Shatkin AJ, Manley JL: **The ends of the affair: capping and polyadenylation**. *Nat Struct Biol* 2000, **7**:838-842.
14. Colgan DF, Manley JL: **Mechanism and regulation of mRNA polyadenylation**. *Genes Dev* 1997, **11**:2755-2766.
15. Murzin AG, Brenner SE, Hubbard T, Chothia C: **SCOP: a structural classification of proteins database for the investigation of sequences and structures**. *J Mol Biol* 1995, **247**:536-540.
16. Rossmann MG, Moras D, Olsen KW: **Chemical and biological evolution of nucleotide-binding protein**. *Nature* 1974, **250**:194-199.
17. Martin JL, McMillan FM: **SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold**. *Curr Opin Struct Biol* 2002, **12**:783-793.
18. Rydberg B, Lindahl T: **Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction**. *EMBO J* 1982, **1**:211-216.

19. Barrows LR, Magee PN: **Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro**. *Carcinogenesis* 1982, **3**:349-351.
20. Motorin Y, Helm M: **RNA nucleotide methylation**. *Wiley Interdiscip Rev RNA* 2011, **2**:611-631.
21. Santi DV, Norment A, Garrett CE: **Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine**. *Proc Natl Acad Sci U S A* 1984, **81**:6993-6997.
22. Desrosiers R, Friderici K, Rottman F: **Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells**. *Proc Natl Acad Sci U S A* 1974, **71**:3971-3975.
23. Perry RP, Kelley DE, Friderici K, Rottman F: **The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus**. *Cell* 1975, **4**:387-394.
24. Jia G, Fu Y, He C: **Reversible RNA adenosine methylation in biological regulation**. *Trends Genet* 2013, **29**:108-115.
25. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, et al.: **N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO**. *Nat Chem Biol* 2011, **7**:885-887.
26. Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH, et al.: **ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility**. *Mol Cell* 2013, **49**:18-29.
27. Patil DP, Pickering BF, Jaffrey SR: **Reading m(6)A in the Transcriptome: m(6)A-Binding Proteins**. *Trends Cell Biol* 2018, **28**:113-127.
28. Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, Adhikari S, Shi Y, Lv Y, Chen YS, et al.: **Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase**. *Cell Res* 2014, **24**:177-189.
29. Csepany T, Lin A, Baldick CJ, Jr., Beemon K: **Sequence specificity of mRNA N6-adenosine methyltransferase**. *J Biol Chem* 1990, **265**:20117-20122.
30. Harper JE, Miceli SM, Roberts RJ, Manley JL: **Sequence specificity of the human mRNA N6-adenosine methylase in vitro**. *Nucleic Acids Res* 1990, **18**:5735-5741.
31. Wang P, Doxtader KA, Nam Y: **Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases**. *Mol Cell* 2016, **63**:306-317.
32. Wang X, Feng J, Xue Y, Guan Z, Zhang D, Liu Z, Gong Z, Wang Q, Huang J, Tang C, et al.: **Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex**. *Nature* 2016, **534**:575-578.
33. Wyatt GR: **Occurrence of 5-methylcytosine in nucleic acids**. *Nature* 1950, **166**:237-238.
34. Suzuki MM, Bird A: **DNA methylation landscapes: provocative insights from epigenomics**. *Nat Rev Genet* 2008, **9**:465-476.
35. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T: **Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA**. *Nucleic Acids Res* 2012, **40**:5023-5033.
36. Legrand C, Tuorto F, Hartmann M, Liebers R, Jacob D, Helm M, Lyko F: **Statistically robust methylation calling for whole-transcriptome bisulfite**

sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res* 2017, **27**:1589-1596.
37. Brzezicha B, Schmidt M, Makalowska I, Jarmolowski A, Pienkowska J, Szweykowska-Kulinska Z: **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**:6034-6043.
38. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH: **Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2**. *Science* 2006, **311**:395-398.
39. Dong A, Yoder JA, Zhang X, Zhou L, Bestor TH, Cheng X: **Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA**. *Nucleic Acids Res* 2001, **29**:439-448.
40. Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, Nellen W, Schaefer M, Helm M: **Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation**. *RNA Biol* 2017, **14**:1108-1123.
41. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, Dai Q, Di Segni A, Salmon-Divon M, Clark WC, et al.: **The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA**. *Nature* 2016, **530**:441-446.
42. Li X, Xiong X, Wang K, Wang L, Shu X, Ma S, Yi C: **Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome**. *Nat Chem Biol* 2016, **12**:311-316.
43. Ozanick S, Krecic A, Andersland J, Anderson JT: **The bipartite structure of the tRNA m1A58 methyltransferase from S. cerevisiae is conserved in humans**. *RNA* 2005, **11**:1281-1290.
44. Peifer C, Sharma S, Watzinger P, Lamberth S, Kotter P, Entian KD: **Yeast Rrp8p, a novel methyltransferase responsible for m1A 645 base modification of 25S rRNA**. *Nucleic Acids Res* 2013, **41**:1151-1163.
45. Finer-Moore J, Czudnochowski N, O'Connell JD, 3rd, Wang AL, Stroud RM: **Crystal Structure of the Human tRNA m(1)A58 Methyltransferase-tRNA(3)(Lys) Complex: Refolding of Substrate tRNA Allows Access to the Methylation Target**. *J Mol Biol* 2015, **427**:3862-3876.
46. Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J: **A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine**. *Cell* 1987, **50**:831-840.
47. Chen SH, Habib G, Yang CY, Gu ZW, Lee BR, Weng SA, Silberman SR, Cai SJ, Deslypere JP, Rosseneu M, et al.: **Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon**. *Science* 1987, **238**:363-366.
48. Teng B, Burant CF, Davidson NO: **Molecular cloning of an apolipoprotein B messenger RNA editing protein**. *Science* 1993, **260**:1816-1819.
49. Wedekind JE, Dance GS, Sowden MP, Smith HC: **Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business**. *Trends Genet* 2003, **19**:207-216.

50. Salter JD, Bennett RP, Smith HC: **The APOBEC Protein Family: United by Structure, Divergent in Function**. *Trends Biochem Sci* 2016, **41**:578-594.
51. Fritz EL, Papavasiliou FN: **Cytidine deaminases: AIDing DNA demethylation?** *Genes Dev* 2010, **24**:2107-2114.
52. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM: **AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation**. *Nat Chem Biol* 2012, **8**:751-758.
53. Olson ME, Harris RS, Harki DA: **APOBEC Enzymes as Targets for Virus and Cancer Therapy**. *Cell Chem Biol* 2018, **25**:36-49.
54. Romani B, Engelbrecht S, Glashoff RH: **Antiviral roles of APOBEC proteins against HIV-1 and suppression by Vif**. *Arch Virol* 2009, **154**:1579-1588.
55. Betts L, Xiang S, Short SA, Wolfenden R, Carter CW, Jr.: **Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex**. *J Mol Biol* 1994, **235**:635-656.
56. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, McCann JL, Starrett GJ, Duffy JV, Demir O, et al.: **Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B**. *Nat Struct Mol Biol* 2017, **24**:131-139.
57. Kouno T, Silvas TV, Hilbert BJ, Shandilya SMD, Bohn MF, Kelch BA, Royer WE, Somasundaran M, Kurt Yilmaz N, Matsuo H, et al.: **Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity**. *Nat Commun* 2017, **8**:15024.
58. Bass BL, Weintraub H: **An unwinding activity that covalently modifies its double-stranded RNA substrate**. *Cell* 1988, **55**:1089-1098.
59. Picardi E, Manzari C, Mastropasqua F, Aiello I, D'Erchia AM, Pesole G: **Profiling RNA editing in human tissues: towards the inosinome Atlas**. *Sci Rep* 2015, **5**:14941.
60. Macbeth MR, Schubert HL, Vandemark AP, Lingam AT, Hill CP, Bass BL: **Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing**. *Science* 2005, **309**:1534-1539.
61. Matthews MM, Thomas JM, Zheng Y, Tran K, Phelps KJ, Scott AI, Havel J, Fisher AJ, Beal PA: **Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity**. *Nat Struct Mol Biol* 2016, **23**:426-433.
62. Haudenschild BL, Maydanovych O, Veliz EA, Macbeth MR, Bass BL, Beal PA: **A transition state analogue for an RNA-editing reaction**. *J Am Chem Soc* 2004, **126**:11213-11219.
63. Klimasauskas S, Kumar S, Roberts RJ, Cheng X: **HhaI methyltransferase flips its target base out of the DNA helix**. *Cell* 1994, **76**:357-369.
64. Saraconi G, Severi F, Sala C, Mattiuz G, Conticello SG: **The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas**. *Genome Biol* 2014, **15**:417.
65. Fierro-Monti I, Mathews MB: **Proteins binding to duplexed RNA: one motif, multiple functions**. *Trends Biochem Sci* 2000, **25**:241-246.

66. Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB, et al.: **The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove**. *Cell* 2010, **143**:225-237.
67. O'Connell MA, Mannion NM, Keegan LP: **The Epitranscriptome and Innate Immunity**. *PLoS Genet* 2015, **11**:e1005687.

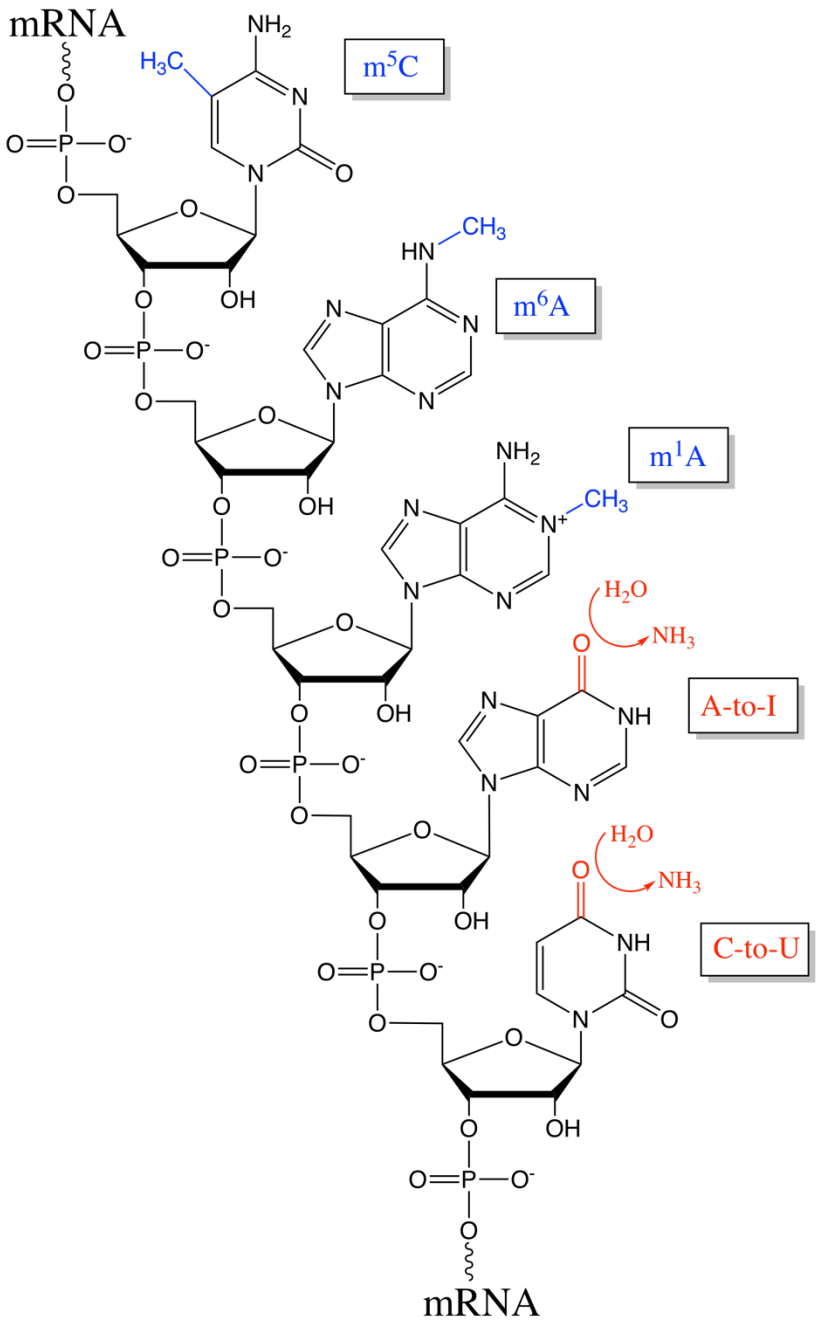
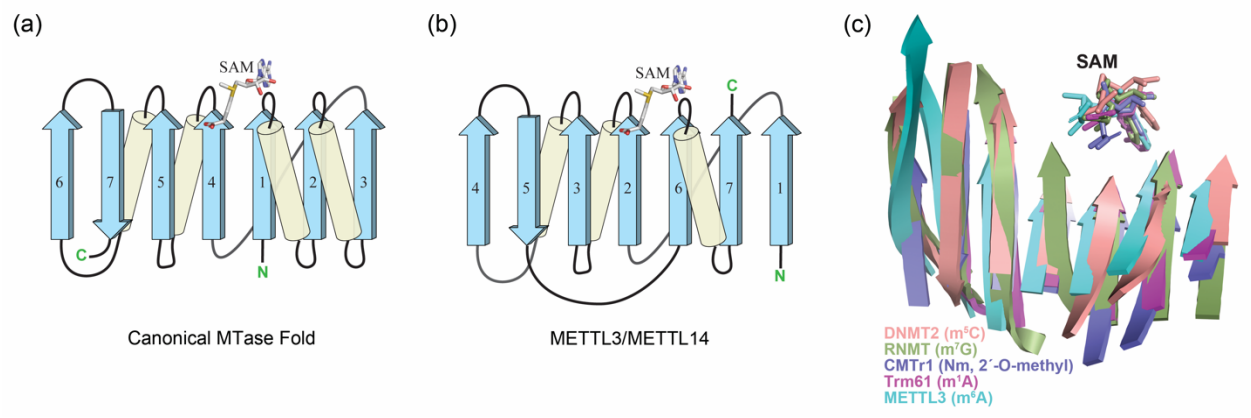
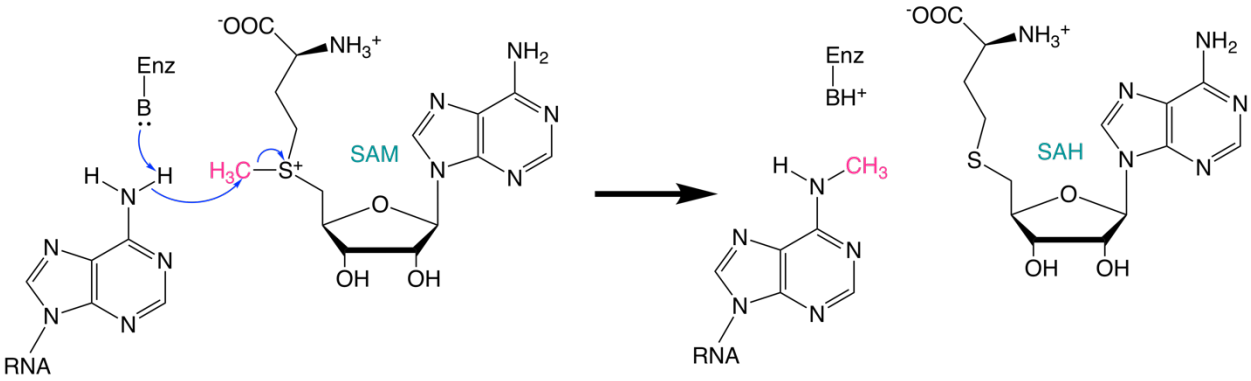


Figure 1. The structures of modified nucleosides identified in mRNA that are discussed in this review. Methylations are colored blue, deaminations in red. Methylated nucleotide abbreviations are indicated in box along with the deamination edits.



(d)



(e)

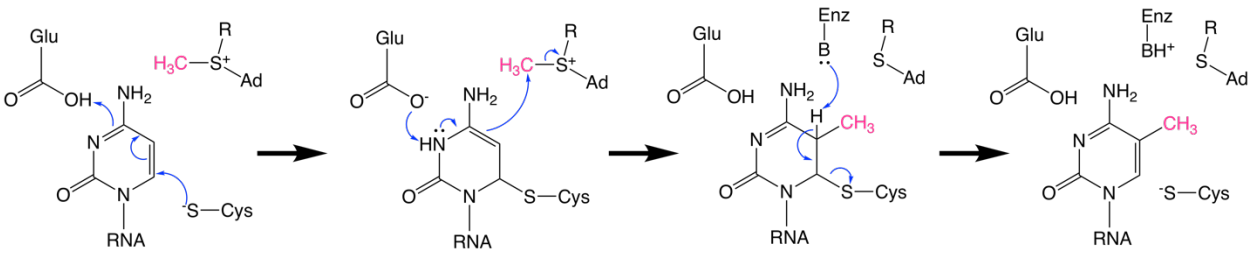


Figure 2. RNA methylation. **(a)** Cartoon illustrating the canonical topology of most RNA MTases. **(b)** The topology of human METTL3/METTL14 m^6A heterodimer MTase showing a similar core, but different connectivity. **(c)** Superposition of the β-sheet core along with the bound SAM/SAH substrate of all known human MTases that modify mRNA, including the Cap MTases RNMP and CMTr1. **(d)** Proposed common mechanism of base-catalyzed methylation of mRNA. **(e)** Proposed mechanism of cytidine C5 methylation. A conserved Cys residue on MTase attacks C6 in a Michael addition, generating an enamine that attacks the methyl group of SAM.

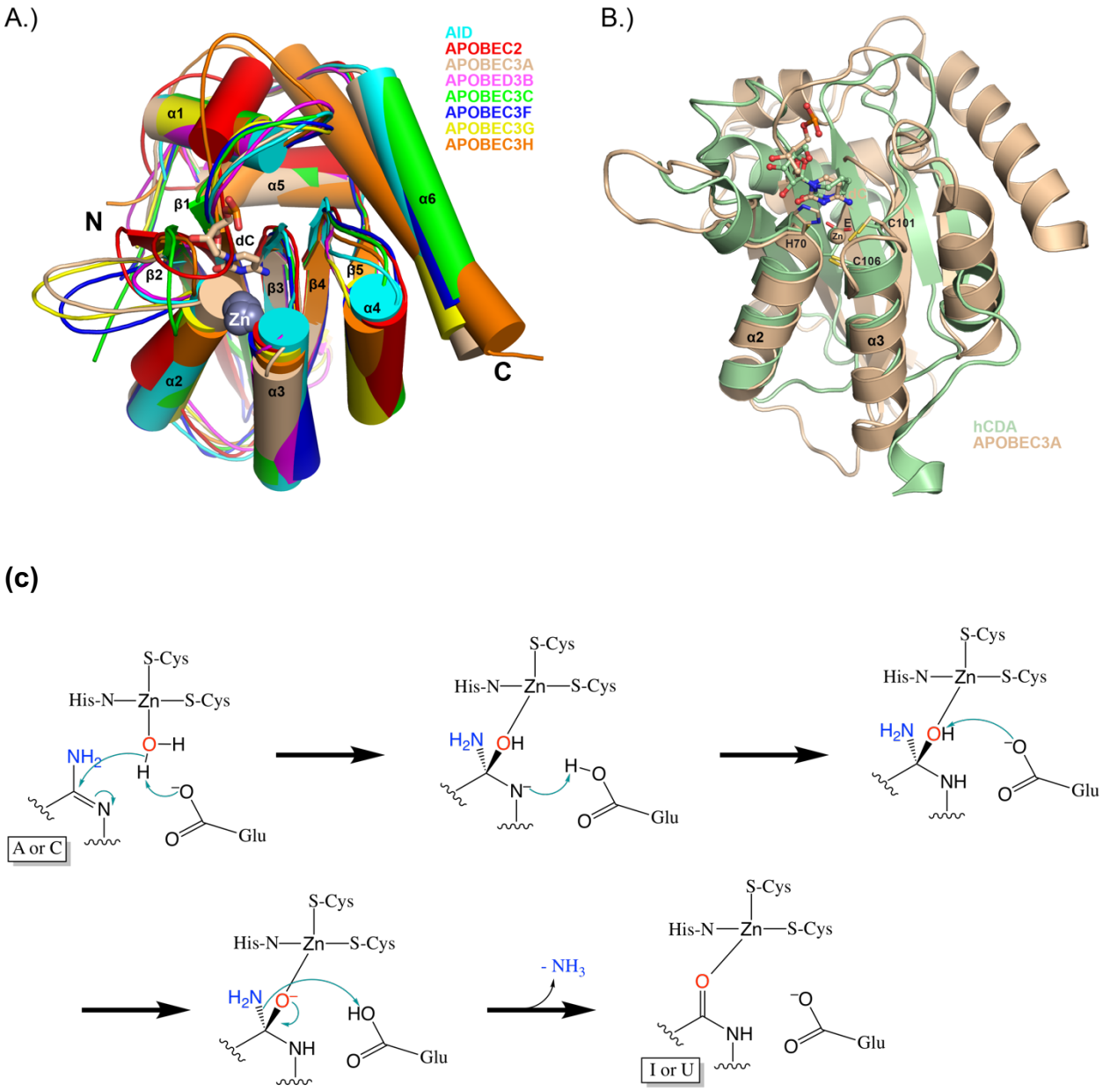


Figure 3. RNA deamination. **(a)** Superposition of known human structures of APOBECs with active site zinc ions shown as gray spheres. Shown in sticks is the cytidine nucleotide of the ssDNA substrate bound to the active site of APOBEC3A. **(b)** Superposition of human APOBEC3A onto the human cytidine deaminase CDA. Nucleotide analogs are shown in sticks with zinc ions as small sphere. **(c)** Common deamination mechanism for both APOBEC and ADAR involving conserved Zn – water ligand that attacks the ring carbon assisted by a Glu residue involve in proton shuttling.

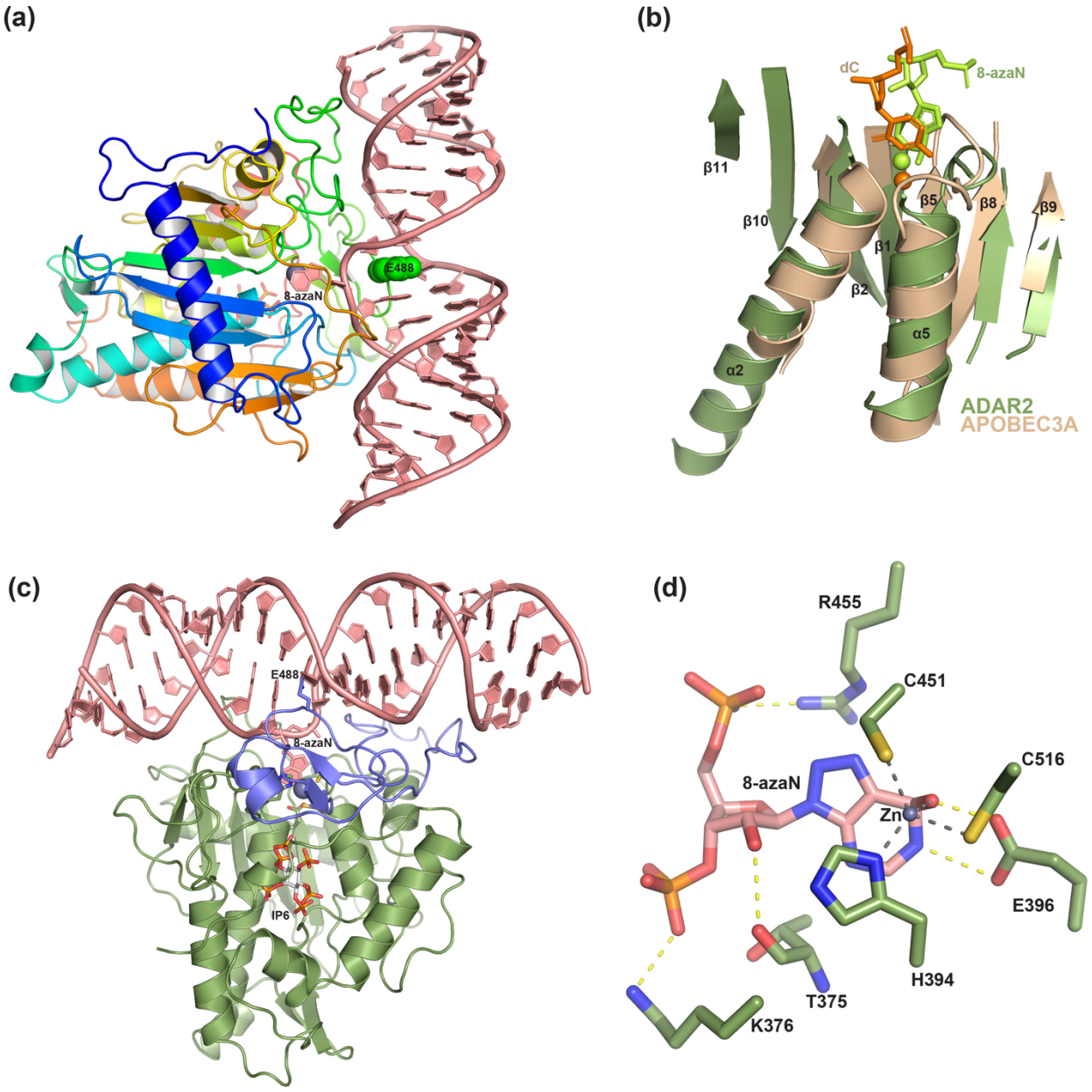


Figure 4. ADAR deamination of RNA. **(a)** Crystal structure of human ADAR2 deaminase domain bound to dsRNA. The base-flipping residue Glu488 is shown in green-colored spheres, which displaces the adenosine analog 8-azaN into the active site. **(b)** Superposition of human ADAR2 deaminase core onto human APOBEC3A. The molecules superimpose with an rmsd of 3.1Å for 64 equivalent α-carbons. Labeled secondary structural elements based on ADAR structure. Note the cytidine nucleotide from the ssDNA-APOBEC structure lies in a similar orientation to the adenosine RNA nucleotide analog 8-azaN. Nucleosides are shown in sticks with active site zinc ions in their respective colored spheres. **(c)** Structure of human ADAR2 showing the location of the IP_6 molecule (sticks with white-colored carbon), and the large insertion between the two Cys zinc ligands. Colored in blue is the sequence between zinc ligands Cys451 and Cys516, which contains the base-flipping residue 488 and the RNA-binding loop. **(d)** ADAR2 active site, showing the hydrated 8-azaN analog bound to the zinc ion and the active

site Glu396. ADAR-RNA hydrogen bonds are represented by yellow dashes, while the zinc ligand bonds by gray dashes.

TOC Graphic