HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2017 July 06.

Published in final edited form as:

Nature. 2017 January 18; 541(7637): 339–346. doi:10.1038/nature21351.

Chemical and structural effects of base modifications in messenger RNA

Emily M. Harcourt[1], **Anna M. Kietrys**[2], and **Eric T. Kool**[2]

[1]Department of Chemistry, Wellesley College, Wellesley, Massachusetts 02481, USA

[2]Department of Chemistry, Stanford University, Stanford, California 94305, USA

Abstract

A growing number of nucleobase modifications in messenger RNA have been revealed through advances in detection and RNA sequencing. Although some of the biochemical pathways that involve modified bases have been identified, research into the world of RNA modification — the epitranscriptome — is still in an early phase. A variety of chemical tools are being used to characterize base modifications, and the structural effects of known base modifications on RNA pairing, thermodynamics and folding are being determined in relation to their putative biological roles.

It is understood that the sequence and structure of messenger RNA have self-regulatory effects, influencing the splicing, translation, cellular localization and longevity of the RNA, and that the sequence and structure of a molecule of mRNA can be changed in its lifetime. Self-regulation has been well studied in riboswitches, which are fragments of RNA with a secondary structure that can be modulated by small molecules to alter splicing, translation and RNA stability[1]. Structural alterations to the purine or pyrimidine rings of nucleobases have the ability to change these same properties, as has long been recognized through modifications in transfer RNA and ribosomal RNA[2,3].

In mRNA, as well as nucleoside modifications that are associated with the 5′-cap (ref. 4), six nucleosides with base modifications have been discovered so far: N^6 -methyladenosine (m^6A) (refs 5 and 6), inosine [7], pseudouridine[8–10], 5-methylcytidine (m^5C) (refs 11 and 12), 5-hydroxymethylcytidine (hm^5C) (refs 13 and 14) and N^1-methyladenosine (m^1A) (refs 15 and 16) (Fig. 1). Cytidine to uridine ('C-to-U') editing is also known to occur in mRNA[17,18]. Although not discussed here, 2′-*O*-methylation of the ribose sugar of various nucleo-sides is another modification that has been recognized in mRNA[6]. Advances in detection have accelerated the pace of identifying new modifications and the discovery of further modifications is on the horizon. Meanwhile, high-throughput sequencing methods have improved our understanding of the distribution and regulation of modifications throughout the transcriptome. The vast amount of data generated by transcriptome-wide sequencing techniques has been complemented in some cases by studies of individual transcripts[19,20].

Correspondence should be addressed to E.T.K. kool@stanford.edu.

The authors declare no competing financial interests.

However, it is widely accepted that continued effort is needed to confirm sites of modification and to quantify the extent of the modification at any particular site.

Determination of the extent of the modification is only one piece of the puzzle; it is also necessary to understand the chemical and structural roles of modified bases. Modifications do not exist as passive marks. Even the addition of small groups such as methyl affects the ability of bases to pair, to stack against neighbouring bases, to adopt one conformation over another, to favour one folded structure over another and to interact with proteins. Such changes affect the biological activities of the RNA molecules that contain them. As yet, these activities largely remain unclear, as do the direct pathways through which mRNA modifications occur, although rapid advances are being made[21,22]. Improvements in detection and quantification together with the consideration of structure, localization and regulation at all levels will help to elucidate the role of each mRNA modification. In this Review, we give an overview of the approaches that are being used to discover and quantify modifications in mRNA. We then discuss these modifications in the context of their structures, asking how changes in shape and chemical functionality might inform the roles of base modifications.

Detection methods

Historically, base modifications have been detected through protocols that used combinations of chemical or enzymatic digestion, radiolabelling and thin-layer chromatography[23]. These intensive techniques were used to detect the presence of m^6A (ref. 6), m^5C (refs 11 and 12) and inosine[24] in mRNAs. However, it is difficult or impossible to use these methods to obtain information about which sites are modified. This is because individual RNA transcripts must be isolated in large enough quantities to enable the use of a combination of RNase enzymes to digest the RNA and localize the modification.

A spate of modifications newly identified in mRNA has been facilitated by improvements in mRNA isolation along with the successful combination of next-generation sequencing techniques with older methods such as immunoprecipitation and base-specific chemistry. Advances in mass spectrometry have also aided the identification and quantification of previously undetected mRNA modifications. However, generalizable methods for site-specific detection and quantification are elusive.

Transcriptome-wide sequencing

In the past five years, a number of studies have provided important transcriptome-wide data on modifications in mRNA and long non-coding RNA[8–10,15,16,25–30]. The innovation in these studies comes from the integration of older, base-specific sequencing techniques with deep sequencing, giving better sensitivity and more comprehensive data than in the past. Three strategies have been combined with deep sequencing for the detection of modifications in mRNA: truncating the products of reverse transcription at the site of modification; altering or detecting altered base-pairing properties at the site of modification; and preferentially enriching modified sequences.

Some modified bases have been detected by conversion to a structure that causes a reverse transcriptase enzyme to stop. For example, both inosine and pseudouridine can undergo base-specific reactions that add bulky groups to the Watson–Crick pairing face, which prevents reverse transcription of the RNA (Fig. 2a). For example, *N*-cyclohexyl-*N′*-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMC) is used to selectively modify pseudouridine at N3 (ref. 31), the nitrogen that typically serves as a hydrogen-bond donor in a Watson–Crick hydrogen bond. When CMC-treated transcripts are compared with untreated transcripts, the location of pseudouridine can be determined by looking for sites of premature truncation[8–10] (Fig. 2b). Inosine sequencing using methods that halt the reverse transcriptase is carried out differently[32,33]. Inosine is a modified form of adenosine but pairs best with cytidine. Because such modification does not occur in every RNA transcript, sites at which inosine is present are therefore sequenced as a mixture of adenosine and guanosine. However, the selective reaction of inosine with acrylonitrile converts inosine into a block for reverse transcription. Only unmodified transcripts (containing adenosine but not inosine) can be read through by reverse transcriptase, whereas the complementary DNA from transcripts that contain inosine is truncated. When cDNA created from acrylonitrile-treated RNA is compared with cDNA generated from untreated RNA, sites of inosine can be identified as those that are read as adenosine in the treated sample but as a mixture of adenosine and guanosine in the control[33].

Other methods of detection rely on altered base pairing, owing to changes in the chemical properties of modified bases. For example, m^5C can be detected by sequencing after treatment with sodium bisulphite: whereas cytosine is converted to uracil, m^5C remains unchanged (Fig. 2c). When the sequences of sodium-bisulfite-treated and untreated RNA transcripts are compared, sites that are read as cytosine can be identified as m^5C (ref. 25) (Fig. 2d).

So far, there is no specific method for detecting the chemical modification m^6A. The methyl group destabilizes pairing with uracil, and a polymerase enzyme has been identified that is slowed approximately tenfold when the group is present[34]. This feature has not yet been developed into a sequencing-based detection method, however. The state of the art for m^6A detection is immunoprecipitation enrichment, in which an anti-m^6A antibody is used to pull down fragments of RNA that contain m^6A. The sequences of the m^6A-enriched RNA pool can be compared to those of an unenriched control pool of transcripts to localize sites of modification on mRNA transcripts to regions of about 100 nucleotides[26,27]. Other efforts have focused on more specific localization. In 2015, m^6A mapping with improved resolution was reported by both Chen *et al.*[35] and Linder *et al.*[36]. Achieved by crosslinking m^6A antibodies to RNA fragments, Linder *et al.* took the method a step further, relying on signature mutations or truncations induced by the crosslinking to identify m^6A sites with single-nucleotide resolution[36].

Immunoprecipitation enrichment can be used in combination with other techniques to provide higher certainty with respect to the location of modified bases. Reported methods for mapping m^1A start with immunoprecipitation enrichment[15,16]. To improve accuracy and sensitivity, each study also takes advantage of the propensity of m^1A to halt polymerases or to induce mismatches[37] owing to the protrusion of the methyl group from the Watson–Crick

hydrogen-bonding face of adenine. The immunoprecipitation-enriched RNA is split into two pools, with one sequenced directly and the other undergoing enzymatic demethylation[16] or Dimroth rearrangement to m^6A (ref. 15) before sequencing. Immunoprecipitation-enriched regions with considerable truncation[16] or a high rate of mismatch[15] in the pool that has been sequenced directly are assigned as containing m^1A. Theoretically, such an approach could be used to determine the precise site of modification; however, the transcriptome-wide studies described here rely on localization-based rather than site-specific mapping.

Genome-wide sequencing has greatly improved our understanding of which transcripts may be modified, as well as the probable location of such modifications. Furthermore, transcriptome-wide mapping has begun to lead to the identification or validation of proteins that 'read' and 'write' base modifications, and overlaying the maps that result can provide insights into the way in which various levels of cellular regulation are interwoven. However, considerable drawbacks and uncertainties are associated with each of the methods developed so far. For example, bisulfite sequencing of m^5C in RNA must balance the unwanted degradation of RNA at high temperatures with the need for denatured (that is, unpaired) RNA and completed reactions to be successful[38]. Acrylonitrile, used in the detection of inosine, and CMC, used in the detection of pseudouridine, both cross-react with other modified bases[39,40].

Detection methods that do not rely on chemical modification also have drawbacks. The success of immunoprecipitation enrichment depends on the specificity of the antibody, and a lack of specificity of anti-m^6A antibodies in purine-rich regions[41] and non-specific affinity for both adenosine and m^1A (ref. 15) have been reported. To account for errors, detection studies set thresholds below which a modification will not be designated, which means that sites of low-frequency modification, sites modified on RNAs of low abundance and sites involved in secondary structures might not be identified. Indeed, it could be that only a fraction of pseudouridine sites have been detected through sequencing[30]. Also, the lack of resolution of immunoprecipitation-based studies and sequence truncation in reverse-transcription-based studies can obscure the presence of several modification sites in close proximity.

Another important factor is the reverse-transcription step that forms part of most sequencing methods. It can be a source of bias because reverse transcriptases may not read through modified nucleotides as well as unmodified ones, meaning transcripts that contain modifications are underrepresented after reverse transcription. In our experience, and that of other researchers in the field, this can lead to underestimation of the modified fraction of RNA, particularly during the use of sequencing methods that involve halting reverse transcriptase. A less investigated corollary is that the presence of several modifications in a single sequence will probably hinder the sequencing of each.

Although transcriptome-wide studies give an idea of the number of transcripts that have been modified, as well as the relative abundance of modification under certain conditions, the ability to accurately quantify modifications at this level is still missing. Exciting breakthroughs in the transcriptome-wide mapping of base modifications therefore need to be

followed up with careful validation, together with the establishment of new methods that enable site-specific quantification.

Mass spectrometry

Generally, and unlike transcriptome-wide sequencing, mass spectrometry does not provide sequence-level information. However, it can be used to identify the existence of modifications and to determine their global abundance. In most cases, RNA is digested into single nucleotides or nucleosides, which are then analysed by high-performance liquid chromatography (HPLC) coupled with tandem mass spectrometry (MS/MS).

The presence of m^1A in mRNA was confirmed through HPLC– MS/MS. Under the alkaline conditions often used during RNA digestion, m^1A can rearrange into m^6A, which may have prevented its earlier identification. Using mild conditions, a peak was found with the retention time and fragmentation pattern of authentic m^1A, with an abundance of 0.015–0.16% of all adenosines[15,16].

Pseudouridine was first identified in mRNA through sequencing, but HPLC–MS/MS was used to quantify pseudouridine in this class of RNA[30]. The results revealed that mRNA contained pseudouridine modification at an abundance of 0.1–0.5% of all uridines, which matched the level of m^6A modification and was considerably higher than the levels suggested by previous studies based on sequencing alone.

Although HPLC–MS/MS has proved its utility in the identification of modified nucleotides, we have yet to see a comparison of the levels of all base modifications using quantitative methods, and the chance to detect new modifications remains.

Quantitative site-specific detection

The global detection of modification levels is an important target, but it relies on selective sequencing (as described previously) and is not always quantitative. One method, termed SCARLET (site-specific cleavage and radioactive labelling followed by ligation-assisted extraction and thin-layer chromatography), enables the detection and quantification of modifications that do not affect Watson–Crick base pairing[42]. The technique was first used to detect m^6A but has also been used to validate sites in the transcriptome-wide sequencing of pseudouridine[30]. SCARLET involves multiple enzymatic steps and site-specific sequences, which must be designed and optimized for each site of interest. In addition to being a time-consuming technique, the reliability of quantification by SCARLET may be affected by the unforeseen effects of nucleoside modification on the activity of any of the several enzymes that are used.

Continued improvements in the site-specific sequencing and quantification of RNA modifications, as well as the further refinement of transcriptome-wide techniques, will aid in our understanding of when and where such modifications occur and expand our ability to correlate modification with function. Even with imperfect detection methods, strides have been made in establishing the function of some modifications. Analysing base structure is the main approach by which we can improve our knowledge of the biological roles of RNA modifications.

Structural effects of modification

It is clear that base modifications cause changes in the reactivity, structure and base-pairing interactions of RNA. However, the degree to which these changes translate into biological effects remains unclear. Of particular interest are dynamic modifications, which have the potential to be added and removed in the lifetime of a single RNA molecule, suggesting that there is exquisite temporal control of biological activity (Box 1). Dynamic modifications could presumably be switched on or off at a single site, whereas non-dynamic modifications only have to occur once to permanently affect the fate of the RNA. However, even those modifications not considered to be truly dynamic (for which no eraser exists) are responsive to changes in cellular conditions, including differentiation and stress. As work continues to elucidate the importance of such modifications, we now focus on the structural effects of each modification in the context of what is known about its biology.

m^6A

The modification m^6A is found mainly near stop codons and in 3′ untranslated regions (UTRs) and has been implicated in numerous mRNA processes and events, including splicing, RNA degradation and the regulation of protein expression levels. There are an estimated 3–5 occurrences of m^6A per molecule of mRNA, mainly in the context of the sequence GGm^6ACU (refs 26 and 27). Several proteins that can recognize m^6A (known as ‘readers’) have been identified, and there is evidence that m^6A can function both as a switch of secondary structure and as a direct recognition element for proteins.

The relatively simple change in structure that results from N^6-methylation has been shown to modulate the accessibility of RNA sequences to RNA-binding proteins[20]. Although it does not change the pattern of hydrogen-bonding donors and acceptors on the base (Table 1), N^6-methylation of adenosine does alter the energetics of the A•U pair. In the preferred unpaired conformation of m^6A, the methyl group is in the *syn* orientation[43] (Table 1). The *anti* conformation, which is required for the Watson–Crick pairing of A with U, elicits an energetic penalty owing to the steric clash between the methyl group and N7, a nitrogen in the purine ring, causing a considerable destabilization of m^6A•U pairs in comparison to A•U pairs[44]. However, unpaired m^6A stacks strongly at the end of a duplex, stabilizing regions that are in transition from double-stranded to single-stranded structures[44]. Consistent with this, cellular RNAs show a decrease in base pairing around sites of m^6A when they undergo mefhylation[45], as well as a structural transition from paired to unpaired in the vicinity of m^6A modifications[44]. The methylated base is therefore proposed to act as a spring-loaded switch, changing from the m^6A•U paired form in the *anti* conformation to the unpaired form in the *syn* conformation, which results in a change in local secondary structure and in biological function[44]. Demethylation can revert the mRNA to its previous form; interestingly, the m^6A demethylases identified so far act only on single-stranded RNA[46,47], which is consistent with the decrease in base pairing in regions that contain m^6A.

The hypothesis that m^6A can serve as a structural switch is also supported by the work of Liu *et al.*, who found that the modification of A to m^6A results in an increase in the accessibility of binding sites for heterogeneous nuclear ribonucleoprotein C (HNRNPC)[20]. Binding of HNRNPC to sites that are modulated by m^6A was, in turn, linked to mRNA

abundance and splicing. For m^{6}A, direct links between a chemical change (modification), a structural change (duplex destabilization) and a biological effect (protein binding) have therefore been established.

Another effect that seems to be attributable directly to the destabilized m^{6}A•U pair is slower pairing of cognate tRNAs with codons that contain m^{6}A. Although structures determined through X-ray crystallography show m^{6}A•U pairs with minimal overall perturbation in the active site of the ribosome, the presence of the modification affects the rate and fidelity of tRNA selection and elongation during translation, probably owing to minor steric effects[48].

The YTH domain-containing family of proteins shows a high affinity of binding to m^{6}A in mRNA[27], especially in the GGm^{6}ACU consensus sequence. Crystal structures of m^{6}A-containing RNA bound to two YTH proteins, YTHDC1 and YTHDF2, reveal that the methyl group of m^{6}A is recognized by an aromatic pocket in the favoured *syn* conformation[49,50]. YTHDC1 was found to regulate mRNA splicing in targeted RNAs[51]. The recognition of m^{6}A-containing RNA by YTHDF2 is linked to the YTHDF2-mediated degradation of such RNAs, a decrease in the lifetime of methylated mRNAs and alterations in ribosome occupancy[52]. YTHDF1, another protein that contains the conserved YTH domain, also binds to m^{6}A and it acts to increase the efficiency of translation[53]. YTHDF1 and YTHDF2 may work in concert to elicit short bursts of the translation of certain transcripts. These examples of binding by YTH domain-containing proteins demonstrate how m^{6}A can trigger a protein-recognition event that affects mRNA fate in the cell.

The modification m^{6}A also has a role in the cap-independent initiation of translation in response to heat shock[54,55]. Clearly, this dynamic modification offers the cell a wide spectrum of tools with which to modulate mRNA maturation, splicing, lifetimes and translational accessibility.

m^{1}A

The unique properties of the modification m^{1}A include a positive charge and a methyl group that fully blocks Watson–Crick pairing (Table 1). These features promote the possibility of strong electrostatic interactions between RNA and protein and the formation of alternative RNA secondary structures. Indeed, m^{1}A is well established as a structural feature in tRNA and rRNA. In tRNA^{Lys}, N^1-alkylation at a specific adeno-sine residue alone is sufficient to trigger the clover-leaf secondary structure, whereas the unmodified sequence forms an extended hairpin[56]. A structural study in RNA generalizes the finding that m^{1}A destabilizes the local duplex: instead of participating in Hoogsteen base pairing, as occurs in DNA[57], m^{1}A remains unpaired, which melts the local duplex[58].

The N^1 methylation of adenine occurs mainly in the 5′ UTR of mRNA and has been correlated with an increase in gene expression and changes in cellular metabolism. The modification m^{1}A, which is around ten times less abundant than m^{6}A, is suggested to play a part in the initiation of translation. The poor pairing ability of m^{1}A, together with its occurrence in GC-rich regions of 5′ UTRs[15,16], strongly suggests that m^{1}A could affect translation by triggering a change in RNA folding that facilitates access to a previously

paired region of RNA. However, such roles and the mechanisms of action of m^1A have yet to be confirmed.

m^5C

So far, the potential roles of m^5C in mRNA remain tenuous. Methylation at position 5 of cytosine has very little effect on pairing but increases the hydrophobicity of the major groove of RNA and probably enhances base stacking[59] (Table 1). The interaction and, presumably, methylation of p16 mRNA by the RNA m^5C methyltransferase NSun2 enhances the stability of the p16 transcript by preventing the binding of proteins that would trigger degradation[19]. However, NSun2 does not have a similar effect on other transcripts[60]. The latest RNA sequencing has revealed the presence of more than 8,000 m^5C sites in both coding and non-coding regions of mRNA. Although methylation sites seem to be distributed randomly in coding regions, considerable enrichment occurs in the 5′ UTR and the 3′ UTR (ref. 25). More information on the transcripts, sequence contexts and secondary structures that contain m^5C is needed before conclusions can be drawn about the chemical or structural effects and the biological consequences of this modification.

The oxidation products of m^5C in RNA — hm^5C and 5-formyl-cytosine — have been observed both *in vitro* and *in vivo*[13,61,62]. These modifications widen the possibility of structural effects through hydrogen bonding or polar interactions in the major groove. Transcriptome mapping and gene analysis of hm^5C indicates the presence of this modification in genes that are involved in basic cellular processes and development[14]. Further work is therefore needed to analyse the roles of hm^5C and 5-formylcytosine in biological processes and to characterize their potentially dynamic nature.

Pseudouridine

Pseudouridine is an isomer of uridine that retains its Watson–Crick base-pairing preference for adenosine. However, there is evidence that pseudouridine both constrains flexibility in single-stranded RNA[63] and provides a small thermodynamic benefit over uridine in pairing with adenosine in double-stranded RNA (Table 1)[64]. Nuclear magnetic resonance structures of a folded RNA molecule show the slow exchange with solvent of the extra hydrogen-bond donor N1 in pseudouridine, which suggests that stabilizing tertiary interactions are formed between N1 and the backbone through a water molecule[63–65]. The idea that this constrained flexibility can modulate function is supported by the work of Chen *et al.*, who found that the incorporation of pseudouridine at important positions in the polypyrimidine tract of adenovirus premRNA reduces splicing. This effect was linked to a decrease in binding of the protein U2 auxiliary factor, which was attributed to the C-3′-*endo* sugar pucker that is favoured by pseudouridine[66]. The fact that pseu-douridine gains an extra hydrogen-bonding group on its non-pairing edge also raises the possibility of selective protein recognition through polar interactions in the major groove.

Sequencing of pseudouridine in human mRNA has led to the identification of several hundred sites[8–10], but analysis by HPLC–MS/MS suggests that pseudouridine is present at a level of 0.2–0.6% of all uridine in mRNA[30]. Although most of the functional pathways involving pseudouridine in mRNA remain unclear, pseudouridylated mRNA sequences are

characterized by enhanced lifetimes in *in vitro* experiments[67] and pseudouridine-containing mRNAs showed about 25% increased expression compared to the same sequences containing uridine[10]. However, whereas a study of an mRNA that contains a pseudouridine modification showed a twofold increase in translation levels compared to those of an unmodified control transcript[67], another study showed a decrease in translation levels of about 30% (ref. 68). Pseudouridylation might therefore participate in the modulation of gene expression and mRNA stability, although its effects could be dependent on the transcript involved, the sequence context and the expression system[10,52].

Other studies have suggested further possible roles for pseudouridine. Various mRNAs containing pseudouridine in place of the uridine of stop codons were read through and decoded as specific amino acids, both *in vitro* and in cells[69]. Interestingly, the anticodon–codon pairs involve previously unknown Watson–Crick/Hoogsteen purine–purine pairs[70] and the basis for the pseudouridine-derived change is unclear, although reduced interactions with release factor proteins owing to a change in the dipole of pseudouridine relative to uridine has been proposed[71]. Evidence of read through both in yeast[69], in which pseudouridine is known to occur in mRNA, and in bacteria[70] suggests that pseudouridine could play a part in changing the coding properties of a transcript. However, such an effect could also be limited, as other studies have shown no change in coding properties on substitution of uridine with pseudouridine[67.68].

Inosine

Inosine modification — commonly referred to as ‘A-to-I’ editing — results in distinctly different base-pairing properties than does adenosine, because inosine pairs most stably with cytidine[72]. However, the I•C pair is expected to be slightly less stable than the A•U pair[73]. Inosine does have the ability to form wobble base pairs with uridine and adenosine, as seen in anticodon pairing, although these are more strongly destabilized compared to the canonical base pairs[74].

It has long been recognized that inosine modifications in the coding region can change the amino acids that are encoded because of the resulting change in base-pairing preference. The best-known example is an A-to-I editing event that changes a glutamine to an arginine in a glutamate receptor in the brain, which leads to a change in calcium permeabil-ity[75]. A-to-I editing can also generate splice-donor and splice-acceptor sites. For example, self-editing has been observed for the intronic pre-mRNA of the enzyme double-stranded RNA-specific editase 1 (also known as ADAR2), resulting in the generation of an alternative 3′ splice-acceptor site and suppression of the enzyme’s expression[76,77].

However, the intrinsic change in the stability of base pairing suggests that inosine modification can also affect the local secondary structure as well as coding and recognition. A-to-I editing in a pairing region would result in a destabilizing I•U wobble pair[78] (Table 1). Indeed, investigations into the ‘unwinding’ of double-stranded RNA found that the cause was inosine modification[79]. ADAR1 and ADAR2, the adenosine deaminase enzymes that are responsible for the conversion of adenosine to inosine in mRNA, act only on double-stranded regions and may continue their activity until the substrate is destabilized[80]. Research focused on RNA editing events suggests that more than 100 million potential

modification sites exist, most of which are found in the repetitive Alu elements of the human transcriptome[81]. Edited double-stranded RNA regions have been detected in a complex with endonuclease V and Tudor staphylococcal nuclease and are linked to the degradation of extensively edited Alu elements, especially in viral infections and stress-response pathways[82–84]. These regions can also be bound by p54^{nrb} and NEAT1 long non-coding RNAs, resulting in their retention in nuclear paraspeckles[85]. Editing plays a part in discriminating between endogenous RNAs: for example, I•U mismatches prevent the oligomerization of the viral double-stranded RNA receptor MDA5 (ref. 86) and the activation of a cytosolic double-stranded RNA cascade response[87], which is typically triggered by viral infection. Nonetheless, a direct connection between changes in secondary structure and the mechanism of action of these biological effects in mRNA has not yet been made.

Uridine

Although uridine is a canonical nucleotide, it can — similar to inosine — arise in mRNA as the result of post-transcriptional deamination (C-to-U editing) rather than through direct incorporation. Unsurprisingly, modification of cytidine to uridine results in a change in pairing preference from guanosine to adenosine, which could drive the unfolding of RNA. Until 2011, *APOB* was the only mRNA known to be an editing target; C-to-U editing changes the coding sequence, resulting in the production of a short isoform of the protein apolipoprotein B[17,18]. However, a transcriptome-wide search uncovered more than 70 new sites of editing, most of which are in the 3′ UTR[88,89]. Similar to inosine modification, uridine editing may change the structure of mRNA by destabilizing double-stranded regions and by promoting new folded structures. So far, only differences at the translation level have been noted[89], and the activity of C-to-U deaminases in double-stranded RNA has not been confirmed.

Outlook

Through the discovery of several base modifications, and with more likely to come, mRNA has joined the ranks of the base-modified RNA families. The burgeoning of transcription-wide modification mapping is the result of important innovations in detection that have greatly increased sensitivity and coverage. Even so, current transcriptome-wide techniques cannot tell us everything we need to know because of biases introduced by RNA structure and abundance. Other than identifying further modifications, the next big improvements in detection may come from the development of more quantitative techniques, including those at the site-specific level. The direct sequencing of RNA — for example, by nanopore[90] — may aid the quest for site-specific detection and quantification.

Because the downstream effects of base modifications arise from chemical alterations, more work is needed to analyse how such changes influence base pairing, helix stability and RNA conformation and folding both quantitatively and at high resolution. Methods that can directly examine how these modifications affect folding in cellular RNAs[91] will also be invaluable in linking the chemistry to the biology.

Although we now know that mRNA is decorated with a considerable level of modification, it remains a substantial task to work out the underlying modification pathways and effects. However, when determining the parts that base modifications play in the splicing, maturation, stability, expression and degradation of mRNA, it is important to keep in mind that all such modifications have a direct impact on RNA structure. Structure, in turn, has the potential to affect function, as has already been shown for both m^6A and inosine. Structural and chemical effects will therefore need to be understood in combination with protein interactions to obtain a complete picture of RNA base modification.

Acknowledgments

We thank the US National Institutes of Health (GM110050 and GM106067) for support. E.M.H. acknowledges the US National Science Foundation for a Predoctoral Fellowship.

References

1. Serganov A, Nudler E. A decade of riboswitches. Cell. 2013; 152:17–24. [PubMed: 23332744]
2. Decatur WA, Fournier MJ. rRNA modifications and ribosome function. Trends Biochem. Sci. 2002; 27:344–351. [PubMed: 12114023]
3. Motorin Y, Helm M. tRNA stabilization by modified nucleotides. Biochemistry. 2010; 49:4934–4944. [PubMed: 20459084]
4. Ramanathan A, Robb GB, Chan S-H. mRNA capping: biological functions and applications. Nucleic Acids Res. 2016; 44:7511–7526. [PubMed: 27317694]
5. 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. [PubMed: 1168101]
6. Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc. Natl Acad. Sci. USA. 1974; 71:3971–3975. [PubMed: 4372599]
7. Morse DP, Bass BL. Detection of inosine in messenger RNA by inosine-specific cleavage. Biochemistry. 1997; 36:8429–8434. [PubMed: 9264612]
8. Carlile TM, et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature. 2014; 515:143–146. [PubMed: 25192136]
9. Lovejoy AF, Riordan DP, Brown PO. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. PLoS ONE. 2014; 9:110799.
10. Schwartz S, et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell. 2014; 159:148–162. [PubMed: 25219674]
11. Dubin DT, Taylor RH. The methylation state of poly A-containing-messenger RNA from cultured hamster cells. Nucleic Acids Res. 1975; 2:1653–1668. [PubMed: 1187339]
12. Sommer S, et al. The methylation of adenovirus-specific nuclear and cytoplasmic RNA. Nucleic Acids Res. 1976; 3:749–765. [PubMed: 1272797]
13. Huber SM, et al. Formation and abundance of 5-hydroxymethylcytosine in RNA. ChemBioChem. 2015; 16:752–755. [PubMed: 25676849]
14. Delatte B, et al. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 2016; 351:282–285. [PubMed: 26816380]
15. Dominissini D, et al. The dynamic N^1-methyladenosine methylome in eukaryotic messenger RNA. Nature. 2016; 530:441–446. Refs 15 and 16 present the first maps of the locations of the modification m^1A in mRNA. [PubMed: 26863196]
16. Li X, et al. Transcriptome-wide mapping reveals reversible and dynamic N^1-methyladenosine methylome. Nature Chem. Biol. 2016; 12:311–316. [PubMed: 26863410]
17. Chen S-H, 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. [PubMed: 3659919]

18. Powell LM, et al. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell. 1987; 50:831–840. [PubMed: 3621347]
19. Zhang X, et al. The tRNA methyltransferase NSun2 stabilizes $p16^{INK4}$ mRNA by methylating the 3′-untranslated region of p16. Nature Commun. 2012; 3:712. [PubMed: 22395603]
20. Liu N, et al. *N*6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature. 2015; 518:560–564. The incorporation of m^6A into a hairpin region is shown to modulate the local structure of RNA, facilitating the binding of HNRNPC. [PubMed: 25719671]
21. Roundtree IA, He C. RNA epigenetics — chemical messages for posttranscriptional gene regulation. Curr. Opin. Chem. Biol. 2016; 30:46–51. [PubMed: 26625014]
22. Gilbert WV, Bell TA, Schaening C. Messenger RNA modifications: form, distribution, and function. Science. 2016; 352:1408–1412. [PubMed: 27313037]
23. Grosjean H, Keith G, Droogmans L. Detection and quantification of modified nucleotides in RNA using thin-layer chromatography. Methods Mol. Biol. 2004; 265:357–391. [PubMed: 15103084]
24. Paul MS, Bass BL. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. EMBO J. 1998; 17:1120–1127. [PubMed: 9463389]
25. Squires JE, et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. 2012; 40:5023–5033. [PubMed: 22344696]
26. Meyer KD, et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell. 2012; 149:1635–1646. This article and ref. 27 were the first to report a deep-sequencing-based transcriptome-wide analysis of an mRNA base modification (m^6A). [PubMed: 22608085]
27. Dominissini D, et al. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. Nature. 2012; 485:201–206. [PubMed: 22575960]
28. Cattenoz PB, Taft RJ, Westhof E, Mattick JS. Transcriptome-wide identification of A > I RNA editing sites by inosine specific cleavage. RNA. 2013; 19:257–270. [PubMed: 23264566]
29. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nature Biotechnol. 2013; 31:458–464. [PubMed: 23604283]
30. Li X, et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nature Chem. Biol. 2015; 11:592–597. A combination of mass spectrometry, enrichment and sequencing by chemical modification are used to improve our understanding of the abundance and localization of pseudouridine in mRNA. [PubMed: 26075521]
31. Ho NW, Gilham PT. Reaction of pseudouridine and inosine with *N*-cyclohexyl-*N*′-β-(4-methylmorpholinium)ethylcarbodiimide. Biochemistry. 1971; 10:3651–3657. [PubMed: 4328867]
32. Sakurai M, et al. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. Genome Res. 2014; 24:522–534. [PubMed: 24407955]
33. Sakurai M, Yano T, Kawabata H, Ueda H, Suzuki T. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nature Chem. Biol. 2010; 6:733–740. [PubMed: 20835228]
34. Harcourt EM, Ehrenschwender T, Batista PJ, Chang HY, Kool ET. Identification of a selective polymerase enables detection of *N*6-methyladenosine in RNA. J. Am. Chem. Soc. 2013; 135:19079–19082. [PubMed: 24328136]
35. Chen K, et al. High-resolution *N*6-methyladenosine (m^6A) map using photo-crosslinking-assisted m^6A sequencing. Angew. Chem. Int. Ed. 2015; 54:1587–1590.
36. Linder B, et al. Single-nucleotide-resolution mapping of m^6A and m^6Am throughout the transcriptome. Nature Methods. 2015; 12:767–772. [PubMed: 26121403]
37. Hauenschild R, et al. The reverse transcription signature of *N*-1-methyladenosine in RNA-Seq is sequence dependent. Nucleic Acids Res. 2015; 43:9950–9964. [PubMed: 26365242]
38. Schaefer M, Pollex T, Hanna K, Lyko F. RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Res. 2009; 37:12. The background of bisulfite sequencing and the changes that are necessary to apply this procedure to RNA are described; both conventional and deep sequencing are then used to detect m^5C sites in tRNA and rRNA.
39. Yoshida M, Ukita T. Modification of nucleosides, nucleotides: VII Selective cyanoethylation of inosine and pseudouridine in yeast transfer ribonucleic acid. Biochim. Biophys. Acta. 1968; 157:455–465. [PubMed: 5665898]

40. Durairaj A, Limbach PA. Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection. Anal. Chim. Acta. 2008; 612:173–181. [PubMed: 18358863]
41. Schwartz S, et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell. 2013; 155:1409–1421. [PubMed: 24269006]
42. Liu N, et al. Probing *N*6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA. 2013; 19:1848–1856. [PubMed: 24141618]
43. Engel JD, von Hippel PH. Effects of methylation on the stability of nucleic acid conformations: studies at the monomer level. Biochemistry. 1974; 13:4143–4158. [PubMed: 4606508]
44. Roost C, et al. Structure and thermodynamics of *N*6-methyladenosine in RNA: a spring-loaded base modification. J. Am. Chem. Soc. 2015; 137:2107–2115. The biophysical effects of m^6A in RNA are analysed, determining that m^6A destabilizes duplexes but stacks strongly on helices, which provides evidence for a structural transition*in vivo*. [PubMed: 25611135]
45. Spitale RC, et al. Structural imprints *in vivo* decode RNA regulatory mechanisms. Nature. 2015; 519:486–490. A method is demonstrated to analyse the structure of RNAs *in vivo*; the authors also show how such a technique can be used to connect RNA modification with structural information. [PubMed: 25799993]
46. Jia G, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nature Chem. Biol. 2011; 7:885–887. [PubMed: 22002720]
47. Zheng G, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell. 2013; 49:18–29. [PubMed: 23177736]
48. Choi J, et al. *N*6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. Nature Struct. Mol. Biol. 2016; 23:110–115. [PubMed: 26751643]
49. Li F, Zhao D, Wu J, Shi Y. Structure of the YTH domain of human YTHDF2 in complex with an m^6A mononucleotide reveals an aromatic cage for m^6A recognition. Cell Res. 2014; 24:1490–1492. [PubMed: 25412658]
50. Xu C, et al. Structural basis for selective binding of m^6A RNA by the YTHDC1 YTH domain. Nature Chem. Biol. 2014; 10:927–929. [PubMed: 25242552]
51. Xiao W, et al. Nuclear m^6A reader YTHDC1 regulates mRNA splicing. Mol. Cell. 2016; 61:507–519. [PubMed: 26876937]
52. Wang X, et al. *N*6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 2014; 505:117–120. [PubMed: 24284625]
53. Wang X, et al. *N*6-methyladenosine modulates messenger RNA translation efficiency. Cell. 2015; 161:1388–1399. [PubMed: 26046440]
54. Zhou J, et al. Dynamic m^6A mRNA methylation directs translational control of heat shock response. Nature. 2015; 526:591–594. [PubMed: 26458103]
55. Meyer KD, et al. 5' UTR m^6A promotes cap-independent translation. Cell. 2016; 163:999–1010.
56. Helm M, Giegé R, Florentz CA. Watson-Crick base-pair-disrupting methyl group (m^1A9) is sufficient for cloverleaf folding of human mitochondrial tRNALys. Biochemistry. 1999; 38:13338–13346. [PubMed: 10529209]
57. Yang H, Lam SL. Effect of 1-methyladenine on thermodynamic stabilities of double-helical DNA structures. FEBS Lett. 2009; 583:1548–1553. [PubMed: 19376116]
58. Zhou H, et al. m^1A and m^1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. Nature Struct. Mol. Biol. 2016; 23:803–810. [PubMed: 27478929]
59. Wang S, Kool ET. Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2'-hydroxyl effects. Biochemistry. 1995; 34:4125–4132. [PubMed: 7535100]
60. Hussain S, et al. NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep. 2013; 4:255–261. [PubMed: 23871666]
61. Fu L, et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J. Am. Chem. Soc. 2014; 136:11582–11585. [PubMed: 25073028]
62. Zhang H-Y, Xiong J, Qi B-L, Feng Y-Q, Yuan B-F. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. Chem. Commun. 2016; 52:737–740.
63. Davis DR. Stabilization of RNA stacking by pseudouridine. Nucleic Acids Res. 1995; 23:5020–5026. [PubMed: 8559660]

64. Kierzek E, et al. The contribution of pseudouridine to stabilities and structure of RNAs. Nucleic Acids Res. 2014; 42:3492–3501. [PubMed: 24369424]

65. Charette M, Gray MW. Pseudouridine in RNA: what, where, how, and why. IUBMB Life. 2000; 49:341–351. [PubMed: 10902565]

66. Chen C, Zhao X, Kierzek R, Yu Y-T. A flexible RNA backbone within the polypyrimidine tract is required for U2AF65 binding and pre-mRNA splicing *in vivo*. Mol. Cell. Biol. 2010; 30:4108–4119. [PubMed: 20606010]

67. Karikó K, et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Mol. Ther. 2008; 16:1833–1840. [PubMed: 18797453]

68. Hoernes TP, et al. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. Nucleic Acids Res. 2016; 44:852–862. [PubMed: 26578598]

69. Karijolich J, Yu Y-TT. Converting nonsense codons into sense codons by targeted pseudouridylation. Nature. 2011; 474:395–398. [PubMed: 21677757]

70. Fernández IS, et al. Unusual base pairing during the decoding of a stop codon by the ribosome. Nature. 2013; 500:107–110. [PubMed: 23812587]

71. Parisien M, Yi C, Pan T. Rationalization and prediction of selective decoding of pseudouridine-modified nonsense and sense codons. RNA. 2012; 18:355–367. [PubMed: 22282339]

72. Nigita G, Veneziano D, Ferro A. A-to-I RNA editing: current knowledge sources and computational approaches with special emphasis on non-coding RNA molecules. Front. Bioeng. Biotechnol. 2015; 3:37. [PubMed: 25859542]

73. Kawase Y, Iwai S, Inoue H, Miura K, Ohtsuka E. Studies on nucleic acid interactions: I Stabilities of mini-duplexes (dG2A4XA4G2-dC2T4YT4C2) and self-complementary d(GGGAAXYTTCCC) containing deoxyinosine and other mismatched bases. Nucleic Acids Res. 1986; 14:7727–7736. [PubMed: 3774544]

74. Lim VI, Curran JF. Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. RNA. 2001; 7:942–957. [PubMed: 11453067]

75. Sommer B, Kohler M, Sprengel R, Seeburg PH. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell. 1991; 67:11–19. [PubMed: 1717158]

76. Rueter SM, Dawson TR, Emeson RB. Regulation of alternative splicing by RNA editing. Nature. 1999; 399:75–80. [PubMed: 10331393]

77. Feng Y, Sansam CL, Singh M, Emeson RB. Altered RNA editing in mice lacking ADAR2 autoregulation. Mol. Cell. Biol. 2006; 26:480–488. [PubMed: 16382140]

78. Serra MJ, Smolter PE, Westhof E. Pronounced instability of tandem IU base pairs in RNA. Nucleic Acids Res. 2004; 32:1824–1828. [PubMed: 15037659]

79. Wagner RW, Smith JE, Cooperman BS, Nishikura K. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells. Proc. Natl Acad. Sci. USA. 1989; 86:2647–2651. [PubMed: 2704740]

80. Lehmann KA, Bass BL. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry. 2000; 39:12875–12884. [PubMed: 11041852]

81. Bazak L, et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. 2014; 24:365–376. [PubMed: 24347612]

82. Weissbach R, Scadden AD. Tudor-SN and ADAR1 are components of cytoplasmic stress granules. RNA. 2012; 18:462–471. [PubMed: 22240577]

83. Ng SK, Weissbach R, Ronson GE, Scadden AD. Proteins that contain a functional Z-DNA-binding domain localize to cytoplasmic stress granules. Nucleic Acids Res. 2013; 41:9786–9799. [PubMed: 23982513]

84. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. Nature Rev. Mol. Cell Biol. 2016; 17:83–96. [PubMed: 26648264]

85. Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding. RNA. Mol. Cell. 2009; 35:467–478. [PubMed: 19716791]

86. Wu B, et al. Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. Cell. 2013; 152:276–289. [PubMed: 23273991]

87. Liddicoat BJ, et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science. 2015; 349:1115–1120. [PubMed: 26275108]

88. Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S, Papavasiliou FN. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. Nature Struct. Mol. Biol. 2011; 18:230–236. [PubMed: 21258325]

89. Blanc V, et al. Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. Genome Biol. 2014; 15:79.

90. Branton D, et al. The potential and challenges of nanopore sequencing. Nature Biotechnol. 2008; 26:1146–1153. [PubMed: 18846088]

91. Spitale RC, et al. RNA SHAPE analysis in living cells. Nature Chem. Biol. 2013; 9:18–20. [PubMed: 23178934]

92. Singer, B., Grunberger, D. Molecular Biology of Mutagens and Carcinogens. Plenum: 1983. p. 45-96.Ch. 4

93. Liu J, et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA *N*6-adenosine methylation. Nature Chem. Biol. 2014; 10:93–95. [PubMed: 24316715]

94. Ping XL. Mammalian WTAP is a regulatory subunit of the RNA *N*6-methyladenosine methyltransferase. Cell Res. 2014; 24:177–189. [PubMed: 24407421]

95. Thiagarajan D, Dev RR, Khosla S. The DNA methyltranferase Dnmt2 participates in RNA processing during cellular stress. Epigenetics. 2011; 6:103–113. [PubMed: 20864816]

96. Huber SM, et al. Formation and abundance of 5-hydroxymethylcytosine in RNA. Chembiochem. 2015; 16:752–755. [PubMed: 25676849]

97. Ito S, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011; 333:1300–1303. [PubMed: 21778364]

98. Henderson JO, Blanc V, Davidson NO. Isolation, characterization and developmental regulation of the human apobec-1 complementation factor (ACF) gene. Biochim. Biophys. Acta. 2001; 1522:22–30. [PubMed: 11718896]

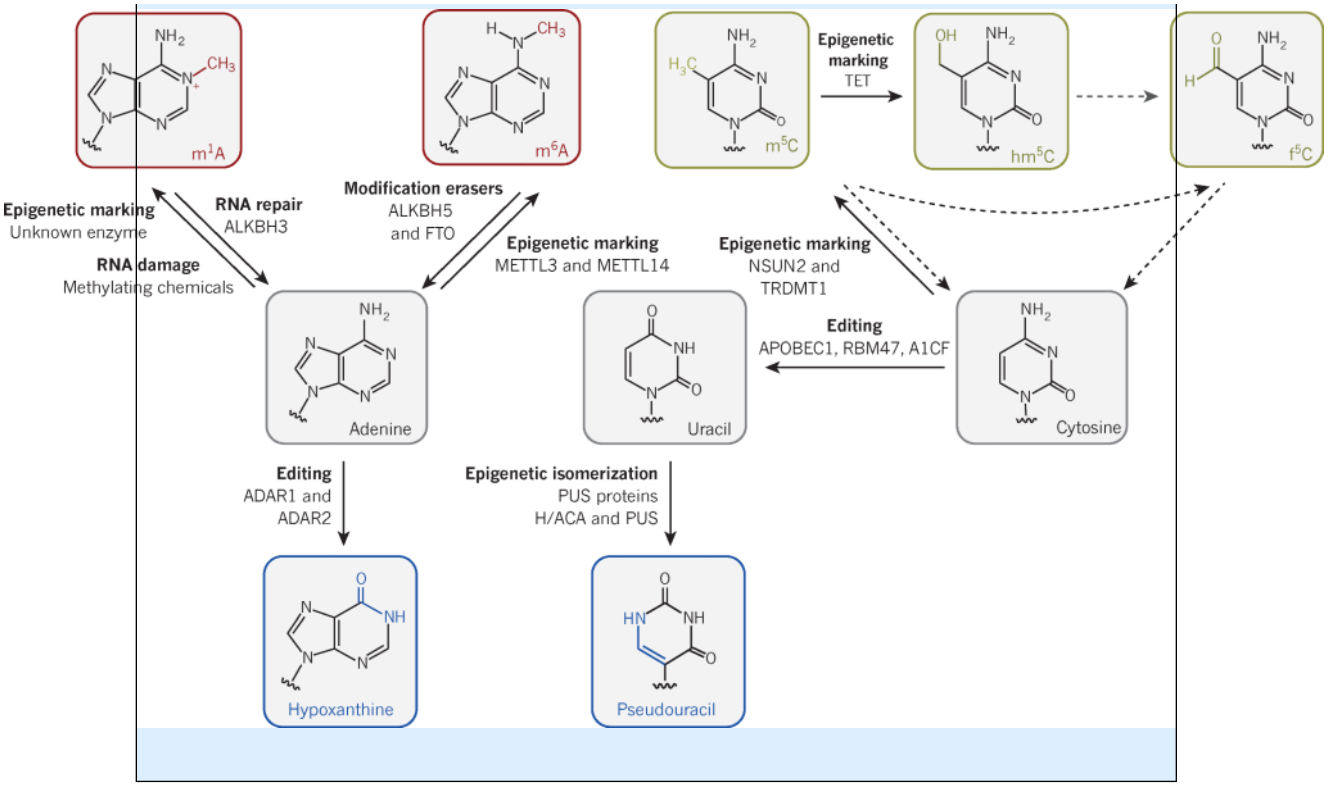
99. Hamma T, Ferré-D'Amaré AR. Pseudouridine synthases. Chem. Biol. 2006; 13:1125–1135. [PubMed: 17113994]

100. Kiss AM, Jady BE, Bertrand E, Kiss T. Human box H/ACA pseudouridylation guide RNA machinery. Mol. Cell. Biol. 2004; 24:5797–5807. [PubMed: 15199136]

BOX 1 Cellular modification pathways

Canonical bases (grey) may undergo a variety of cellular processing events to form modifications that are known to be reversible (red), proposed to be dynamic (green) or not expected to be dynamic (blue). Dynamic modifications with both readers and erasers are of particular interest because a single site might be modified and then unmodified, which enables control within the lifetime of the mRNA — typically a period of hours. Here, we summarize the reported cellular machinery for the forward and reverse modification of mRNA bases (Box Fig.).

The modified base m^1A can arise as a type of RNA damage caused by treatment with a methylating agent[92]. However, the abundance and distribution of this modification suggest that methylation at position 1 of adenine can also be carried out by an unknown enzymatic pathway. The RNA repair enzyme ALKBH3 can remove the N1 methyl group from m^1A (refs 15 and 16). In cells, m^6A is generated by the METTL3– METTL14–WTAP enzymatic complex[93,94] and the N6 methyl group can be removed by the enzymes ALKBH5 (ref. 47) and FTO[46]. Another form of adenine modification is the base hypoxanthine (the nucleoside form of which is called inosine), which arises as the result of mRNA editing provided by the deaminases ADAR1 and ADAR2 (ref. 80). Cytosine undergoes methylation at position C5 to form m^5C and the reaction may be performed by the enzyme NSUN2 (ref. 19) or the methyltransferase TRDMT1 (ref. 95). The modification m^5C may then undergo enzymatic oxidation by TET dioxygenases to form hm^5C (ref. 61). The product of further oxidation to a formyl group (f^5C) has also been detected[96] and may be an intermediate in the conversion of m^5C or hm^5C back to cytosine[97]. Cytosine is edited into uracil by a complex of the proteins APOBEC1, RBM47 and A1CF, which form part of an editosome[98]. Uracil can be isomerized to pseudouracil by various PUS[99] enzymes alone or with the cooperation of H/ACA box ribonucleoproteins[100]



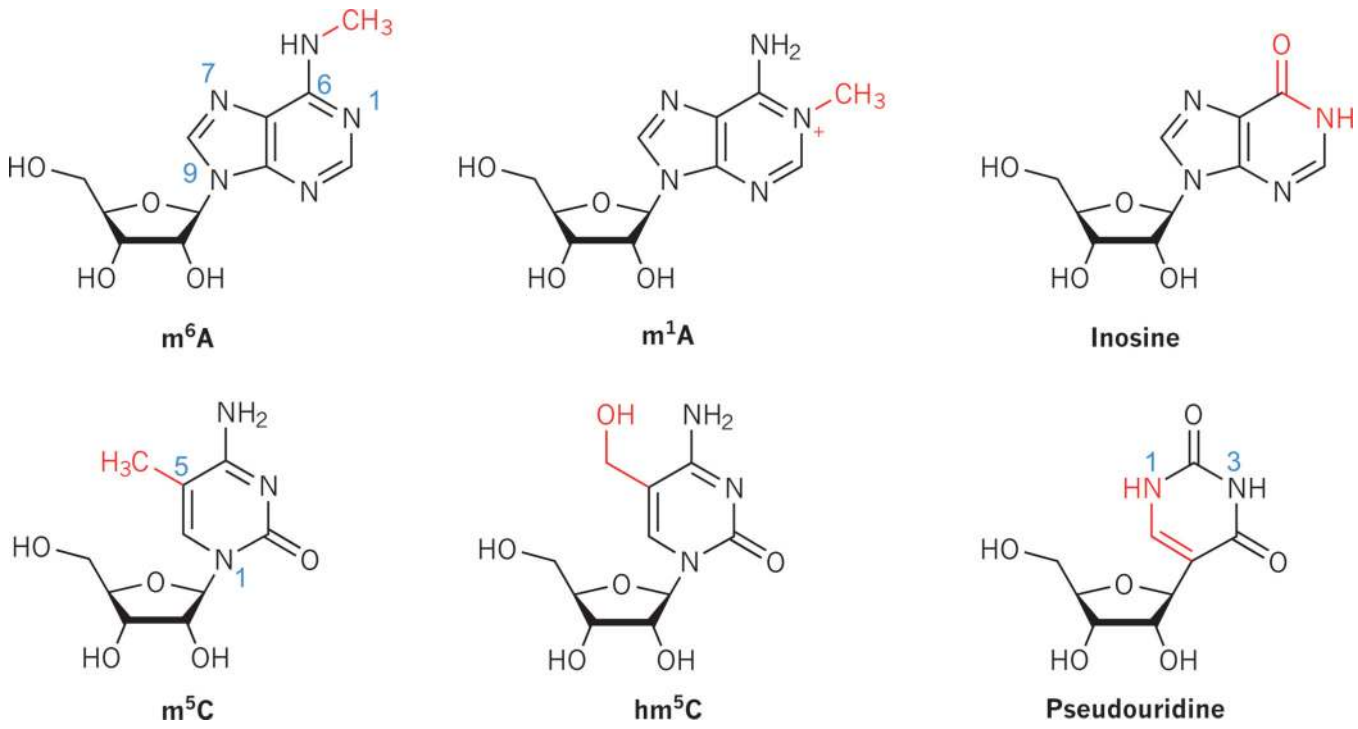


Figure 1. Structures of base-modified nucleosides known to be present at internal positions in mRNA

Six modified bases have been discovered so far: m^6A, m^1A and inosine (top row) and m^5C, hm^5C and pseudouridine (bottom row). Chemical modifications are shown in red. Relevant purine and pyrimidine ring numbering is shown in blue.

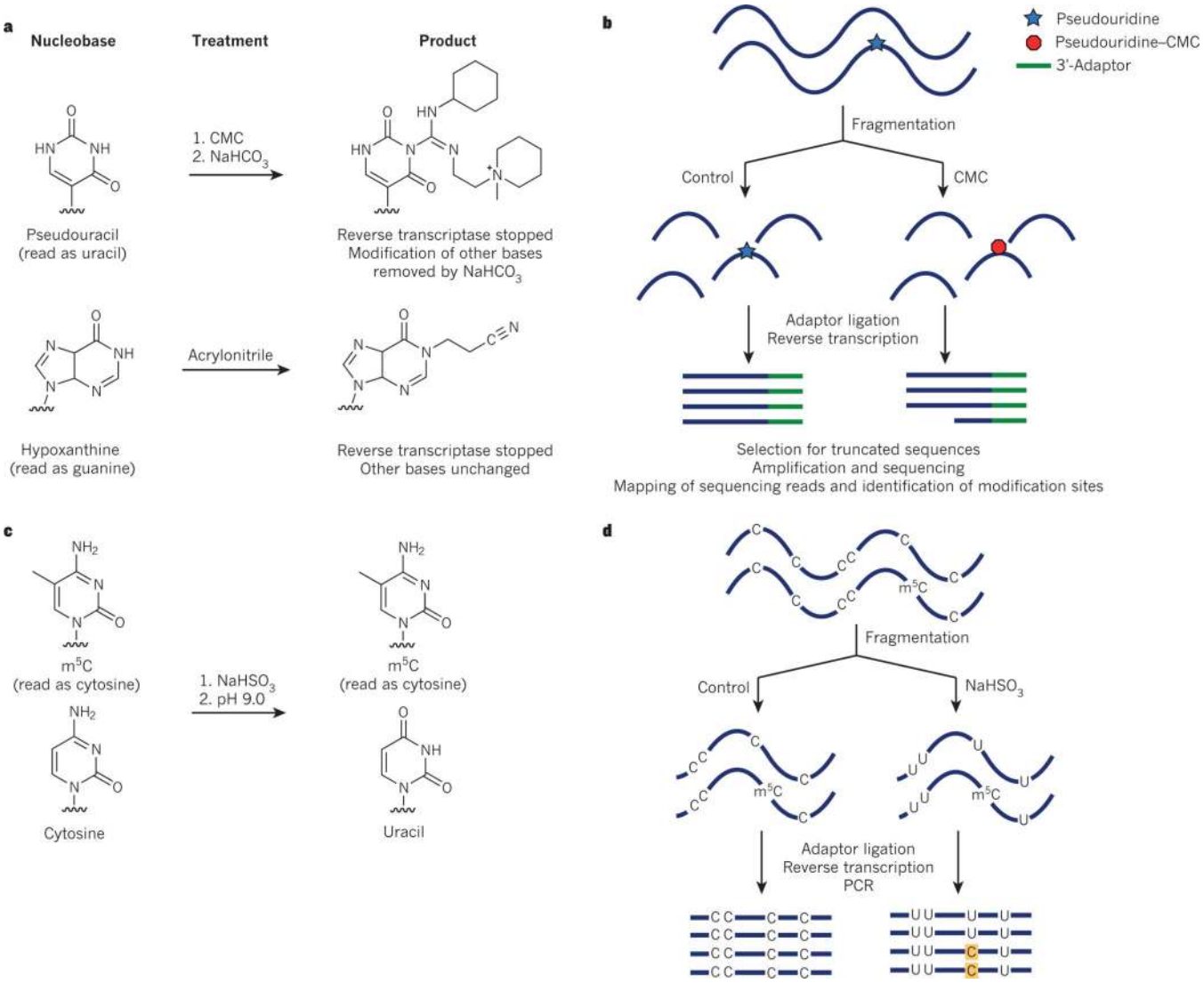


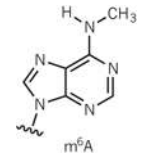
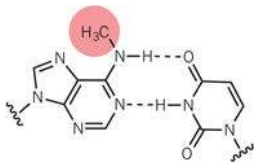
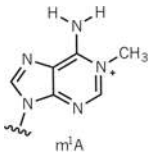
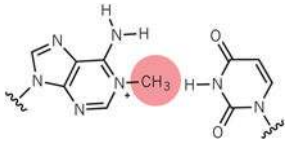
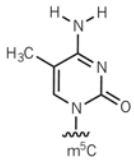
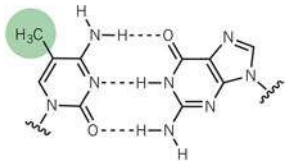
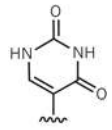
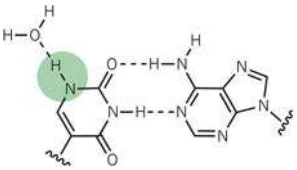
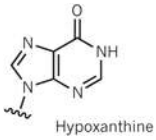
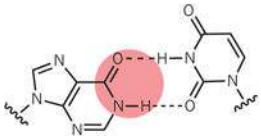
Figure 2. Sequencing by chemical modification for the location of modified bases

a, Chemical treatments are applied to the modified bases pseudouracil and hypoxanthine, and the structures that result prevent the read through of RNA transcripts by reverse transcriptases. **b**, Sequencing by modification-specific termination of reverse transcription. The process for pseudouridine is shown as an example[8–10]. Poly(A)$^+$-enriched mRNA is fragmented and treated with CMC. A 3′-adaptor is ligated to each fragment and then reverse transcription is carried out. Next, truncated complementary DNA is selected by gel electrophoresis, amplified and sequenced. The location of the modification site is determined by comparing the frequency of read termination with and without the CMC treatment. **c**, The effect of sodium bisulfite ($NaHSO_3$) treatment on m^5C and cytosine. The resulting structures are shown on the right. **d**, Bisulfite sequencing[25]. Poly(A)$^+$-enriched mRNA is fragmented and treated with sodium bisulfite. In the treated sample, all cytosines are converted to uracils; however, m^5C is resistant to the treatment. The location of this base

modification can be then identified as the sites that still code as cytosine after sodium bisulfite treatment. C, cytosine; U, uracil.

Table 1

Structural effects of modified nucleotides in RNA

Modified base	Structural effects	Proposed conformation in doubled-stranded RNA	Effects in double-stranded RNA	Protein interactions
 m^{6}A	Watson-Crick base pairing blocked[44] Base stacking enhanced[44]		Base pairing destabilized[44] Helix unwinding[44,45] Adjacent helices stabilized[44]	Binds to hydrophobic pocket (YTH domain proteins)[49,50]
 m^{1}A	Watson-Crick base pairing blocked[58]		Base pairing blocked[58] Helix unwinding[58]	Possible electrostatic nteractions
 m^{5}C	Base stacking enhanced[59]		Base pairing possibly stabilized[49] Increase in major groove hydrophobicity	Possible hydrophobic nteractions
 Pseudouracil	Conformationa flexibility possibly reduced[65]		Base pairing stabilized[63] Increase in major groove polarity Tertiary interactions stabilized[65]	Possible polar interactions in the major groove
 Hypoxanthine	Pairing preference altered[73]		Base pairing destabilized[78] Helix unwinding[75] Non-selective base pairing	Unknown

Modifications that have a destabilizing effect on Watson-Crick base pairing and helix stability are indicated by red circles. Modifications that have a stabilizing effect are indicated by green circles.