Viral Epitranscriptomics

Edward M. Kennedy, David G. Courtney, Kevin Tsai, Bryan R. Cullen
Department of Molecular Genetics and Microbiology and Center for Virology, Duke University Medical Center, Durham, North Carolina, USA

ABSTRACT

Although it has been known for over 40 years that eukaryotic mRNAs bear internal base modifications, it is only in the last 5 years that the importance of these modifications has begun to come into focus. The most common mRNA modification, the addition of a methyl group to the N6 position of adenosine (m6A), has been shown to affect splicing, translation, and stability, and m6A is also essential for embryonic development in organisms ranging from plants to mice. While all viral transcripts examined so far have been found to be extensively m6A modified, the role, if any, of m6A in regulating viral gene expression and replication was previously unknown. However, recent data generated using HIV-1 as a model system strongly suggest that sites of m6A addition not only are evolutionarily conserved but also enhance virus replication. It is therefore likely that the field of viral epitranscriptomics, which can be defined as the study of functionally relevant posttranscriptional modifications of viral RNA transcripts that do not change the nucleotide sequence of that RNA, is poised for a major expansion in scientific interest and may well fundamentally change our understanding of how viral replication is regulated.

KEYWORDS

Posttranscriptional gene regulation, RNA modification, N6-methyladenosine, mRNA function, mRNA stability, HIV-1

While over 100 different modified bases have been identified on RNA transcripts in mammalian cells, the majority of these are restricted to noncoding RNAs, especially tRNAs. However, at least 10 distinct modified bases have now been reported to occur in mammalian mRNAs (1). In addition to the 7-methylguanosine cap that is added at the 5′ end of all cellular mRNAs, these include N6-methyladenosine (m6A), 2′-O-methyladenosine (Am), N6-2′-O-methyladenosine (m6Am), pseudouridine, and 5-methylcytosine. Of these, by far the most prevalent internal modified base found on mRNAs is m6A, and recent work has now begun to reveal how m6A affects mRNA function and how to precisely map the m6A residues present on mRNAs (1–3). m6A is also highly prevalent on a wide range of different viral RNA species (4–13), and recently, the first reports demonstrating a significant phenotypic effect of these m6A modifications have been published (10–14). Therefore, we will focus this review entirely on m6A and how this particular modification might affect different aspects of the viral life cycle.

m6A was first reported to be present on cellular mRNAs in 1975 with ~3 internal m6A residues found on the average ~2.2-kb transcript (15, 16). However, we now know that many cellular mRNAs, including mRNAs encoding housekeeping genes, lack any m6A residues, while highly regulated mRNAs may contain 10 or more (2, 3). The first demonstration of m6A residues on viral mRNAs soon followed and, using the biochemical approaches available at that time, a range of mRNAs encoded by several nuclear DNA and RNA viruses were then shown to bear fairly high levels of m6A, with the eight influenza A virus (IAV) mRNAs bearing an average of three m6A residues each (4–9). Subsequent work looking at each individual IAV mRNA revealed that IAV mRNAs actually contain from 1 to 8 m6A residues each (Table 1) (5). Investigators looking at transcripts encoded by the Rous sarcoma virus (RSV) also demonstrated that the RSV genomic RNA contained at least 8 m6A residues and were able to map two of these (7,
However, mutagenesis of these two m6A sites did not produce any phenotypic effect (17). In the absence of a more facile method to map the precise location of m6A on transcripts, and in the absence of information about which cellular factors produce and detect m6A residues, the field of viral epitranscriptomics, which can be defined as the study of functionally relevant posttranscriptional modifications of viral RNA transcripts that do not change the nucleotide sequence of that RNA, then became largely quiescent for almost 2 decades. During this time, researchers looking at aspects of gene regulation and development in a number of organisms were able to gradually identify several factors relevant to m6A addition and function and, perhaps most importantly, to develop techniques that map m6A sites with near single-nucleotide resolution.

The addition to m6A occurs predominantly in the nucleus and is mediated by the enzyme methyl transferase-like 3 (METTL3) together with several cofactors that have been reported to include METTL14, WTAP, KIAA1429, and RBM15/RBM15B (Fig. 1) (18–22). The human nucleus contains at least two proteins able to detect m6A residues, called YTHDC1 and YTHDC2 (23–26). YTHDC1, known as YT521-B in Drosophila, has been proposed to regulate mRNA splicing and is required for transcriptional repression by the long noncoding RNA XIST, which is heavily m6A modified (21, 24–26). Once exported from the nucleus, m6A residues on mRNAs are bound by three related cytoplasmic proteins, called YTHDF1, YTHDF2 and YTHDF3, which are believed to mediate the phenotypic effects of m6A on mRNA stability and translation (Fig. 1)(2, 3, 27).

In addition to METTL3 and its associated cofactors, referred to as m6A “writers,” and the various m6A-binding proteins, referred to as m6A “readers,” at least two proteins, ALKBH5 and FTO, have been proposed to function as m6A demethylases or “erasers” (28, 29). However, recent data suggest that FTO actually selectively demethylates the m6Am residues located at position 2 in many mRNAs and has a very limited ability to demethylate internal m6A residues (30). Nevertheless, the existence of at least one m6A eraser, the largely nuclear ALKBH5, means that m6A has the ability to function as a dynamic mRNA modification that can be added or removed in response to stress or other signals (2, 3).

A major reason why m6A has become a focus of research interest relates to the profound cellular phenotypes observed when m6A addition is perturbed. Loss of m6A addition is embryonic lethal in plants (31) and strongly perturbs development and sex determination in Drosophila (25, 26, 32). Moreover, loss of m6A addition blocks the differentiation of mammalian embryonic stem cells (33, 34). Importantly, the m6A addition machinery is evolutionarily conserved in all multicellular organisms examined thus far and is also present in fungi, including the yeast Saccharomyces cerevisiae (32), thus highlighting its potential importance.

### TABLE 1 Viruses encoding RNAs with reported m6A residues

<table>
<thead>
<tr>
<th>Virus</th>
<th>No. of m6A residues</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA viruses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influenza A virus</td>
<td>~24</td>
<td>4, 5</td>
</tr>
<tr>
<td>Avian sarcoma virus</td>
<td>13–15</td>
<td>6</td>
</tr>
<tr>
<td>Rous sarcoma virus</td>
<td>10–12</td>
<td>7</td>
</tr>
<tr>
<td>Feline leukemia virus</td>
<td>NA*</td>
<td>46</td>
</tr>
<tr>
<td>HIV-1</td>
<td>10–14</td>
<td>10, 11</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td>~16</td>
<td>12</td>
</tr>
<tr>
<td>Flavivirusesb</td>
<td>5–12</td>
<td>12, 13</td>
</tr>
<tr>
<td>DNA viruses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenovirus</td>
<td>NA</td>
<td>8, 45</td>
</tr>
<tr>
<td>SV40</td>
<td>NA</td>
<td>9, 43</td>
</tr>
<tr>
<td>Herpes simplex virus 1</td>
<td>NA</td>
<td>44</td>
</tr>
</tbody>
</table>

*NA, not available or not applicable.

bIncluding Zika virus, yellow fever virus, Dengue virus, and West Nile virus, all of which were reported to contain multiple internal m6A residues.
The sequence specificity of the writer proteins that add m6A to mRNAs is not entirely clear, though it has been known for some time that the minimal sequence context is 5'-Rm6AC-3' (where R is a purine) (6). A larger consensus sequence, 5'-RRm6ACH-3' (where H is A, C, or U), has also been suggested (2, 3, 35), and evidence indicates that 5'-Gm6AC-3' is generally preferred over 5'-Am6AC-3' (6, 10). Yet, at most 10% of the consensus m6A sites found on mRNAs are actually modified and, despite the random distribution of consensus target sites, m6A residues are also, for currently unclear reasons, concentrated in the 3' untranslated region (UTR) of cellular mRNAs (36, 37).

A major step forward in the study of m6A was the development of techniques to map the adenosine residues that are actually modified. The first reported technique, called Me-RIP-seq (37, 38), uses a commercially available antiserum that specifically recognizes m6A. With this protocol, mRNAs are first purified by poly(A) selection and are then fragmented to ~100 to 200 nucleotide (nt) pieces. The fragmented RNA is then
incubated with the m6A-specific antiserum, which enables the selective immunopre-
cipitation (IP) of m6A-containing RNA fragments. These are collected, subjected to deep
sequencing, and then mapped onto the relevant genome or mRNA transcript using
bioinformatics. The problems with Me-RIP-seq are 2-fold. First, because this technique
is completely reliant on the relatively weak interaction between the antibody and m6A,
the purification steps that can be performed are not that rigorous, resulting in signif-
ificant nonspecific RNA background. Second, the precision of m6A site mapping that can
be achieved is only 100 to 200 nt. As 5′-RAC-3′ sequences are expected to occur by
chance every 32 nt, this technique cannot map m6A sites precisely and cannot
distinguish between single m6A sites and m6A clusters.

The second technique used for m6A mapping, PA-m6A-seq, also relies on the same
m6A-specific antiserum but uses poly(A)-containing mRNA derived from cells that have
been pulsed with the highly photoactivatable uridine analog 4-thiouridine (4SU) (39).
Once the antibody has been bound to the purified 4SU-labeled mRNA population, the
antibody is cross-linked to the RNA by a pulse of UV light. The resultant RNA:protein
complexes can then be rigorously purified prior to digestion with T1 RNase to remove
RNA sequences that are not protected by the bound antibody. The antibody is then
removed by proteinase K treatment, and the resultant ~30-nt RNA fragments are deep
sequenced. An additional advantage of this variation on the photoactivatable
ribonucleoside-enhanced cross-linking and IP (PAR-CLIP) (40) protocol is that the
cross-linked 4SU residue is misread by reverse transcriptase as a C, so that any residual
contaminating RNA fragments can be discarded during bioinformatic analysis by
including only reads bearing single U-to-C mutations. Other major advantages of
PA-m6A-seq are the resultant extremely low background and the increased resolution of
~30 nt. Despite claims that PA-m6A-seq can identify m6A residues at single-
nucleotide resolution, the prevalence of the 5′-RAC-3′ motif means that there are quite
often 2 or even 3 candidate A residues within the mapped m6A peak. This remains a
problem, although targeted mutagenesis of individual 5′-RAC-3′ consensus sequences,
followed by a repeat of the PA-m6A-seq analysis, represents one effective way to
resolve this issue. Recently, another method for mapping m6A residues by cross-linking
m6A-specific antibodies to RNA molecules, referred to as m6A individual-nucleotide-
resolution cross-linking and IP (miCLIP), has been reported (41) that, as made clear by
its name, claims single-nucleotide mapping of m6A sites. The key to this level of
resolution is the authors’ finding that UV cross-linking followed by reverse transcription
specifically and uniquely results in the introduction of a C-to-T mutation at the cytosine
present in the m6A consensus sequence 5′-Rm6AC-3′ in the ~40-nt-long reads ob-
tained, thus enabling the unequivocal bioinformatic identification of m6A residues on
transcripts of interest.

A final method used to map m6A sites relies on the fact that the cytoplasmic
YTHDF1, YTHDF2, and YTHDF3 reader proteins (Fig. 2) are all known to specifically bind
to m6A (2, 3). A form of PAR-CLIP in which cells are pulsed with 4SU, cross-linked using
UV light, and then subjected to immunoprecipitation of a YTHDF protein, followed by
RNase T1 and proteinase K treatment and cDNA synthesis, can therefore identify
precisely the YTHDF protein-binding sites on mRNAs (10). These sites should then
define functionally relevant m6A residues. As in the case of PA-m6A-seq, YTHDF
PAR-CLIP again maps m6A sites with ~30-nt resolution and gives rise to almost no
background. It remains theoretically possible that YTHDF1, YTHDF2, or YTHDF3 might
also bind to RNA sites that lack m6A, though we have not so far observed this
phenomenon.

While the techniques described above can accurately map m6A residues on all
expressed RNA transcripts in a cell, they are at best semiquantitative and the actual
level of m6A modification at any given site is therefore uncertain. One published
 technique called “site-specific cleavage and radioactive-labeling followed by ligation-
assisted extraction and thin-layer chromatography” (SCARLET) has been reported to
enable the quantification of the level of m6A at specific sites on RNAs (42). However, as
implied by its name, this procedure is technically complex, is expensive to perform, and
assesses the level of m^6^A at one adenosine at a time. It is therefore clear that a simpler high-throughput approach that measures the level of m^6^A modification at multiple sites on the transcriptome simultaneously would represent an important technical advance.

Armed with the ability to inhibit m^6^A addition or function using RNA interference (RNAi) or gene editing and to map and mutate specific m^6^A residues on RNAs, it is now possible to begin to ask precisely how individual m^6^A residues, and the process of m^6^A addition in general, affect viral replication and gene expression. While only a few articles have appeared so far using this kind of approach, it appears likely that the emerging field of viral epitranscriptomics is not only poised for a major expansion but also has the potential to greatly influence our understanding of how viruses regulate their life cycle.

NUCLEAR RNA AND DNA VIRUSES

As the cellular proteins that add m^6^A to transcripts reside in the nucleus at steady state (2, 3) (Fig. 2), one might anticipate that, if viral RNAs are indeed m^6^A modified, this would primarily or exclusively occur for RNAs generated by nuclear DNA or RNA viruses. In fact, analyses of three DNA viruses (adenovirus, herpes simplex virus type 1, and SV40), four retroviruses (the closely related avian sarcoma virus and Rous sarcoma virus as well as HIV-1 and feline leukemia virus), and the orthomyxovirus influenza A virus (IAV) have revealed m^6^A residues present at levels that are at least as high as the number of m^6^A residues detected on cellular mRNAs (4–11, 43–46). Moreover, in HIV-1, where m^6^A residues have been mapped at near single-nucleotide resolution, the consensus m^6^A addition sites that are utilized are highly conserved across HIV-1 isolates (10). Given the plasticity of the HIV-1 genome, this conservation clearly implies that m^6^A facilitates some aspect of the replication cycle of HIV-1 and, by extension, of other nuclear viruses that express m^6^A-modified transcripts (Table 1). We note that m^6^A addition has been proposed to affect mRNA splicing (24–26), stability (27, 47), and translation (47–50), to modify RNA structure (51), and to inhibit the recognition of viral RNAs by Toll-like receptors and RIG-I (52, 53), and so m^6^A could positively regulate several aspects of the viral life cycle. Indeed, knockdown of the METTL3 and/or METTL14 m^6^A writers using RNA interference (RNAi) has been reported to inhibit HIV-1 replication up to 5-fold, while knockdown of the ALKBH5 m^6^A demethylase enhanced HIV-1 replication up to 8-fold (11, 14). Similarly, in CD4-positive T cells, overexpression
of the predominant cytoplasmic reader protein YTHDF2 enhanced HIV-1 replication, while knockout of YTHDF2 by gene editing inhibited HIV-1 replication by 2-fold or more (10). We note that one group has reported, in contrast, that all of the YTHDF proteins can inhibit HIV-1 replication (14). However, this group exclusively analyzed replication of an HIV-1 variant bearing the firefly luciferase (FLuc) indicator gene in place of nef, and we have observed that FLuc actually contains prominent m6A modification sites that may well affect how YTHDF proteins affect the replication of this HIV-1-derived lentiviral vector (E. M. Kennedy and B. R. Cullen, unpublished results). We therefore believe it is essential that experiments addressing how m6A affects virus replication use wild-type viruses rather than viral mutants that have been modified to express an exogenous indicator gene. In conclusion, the prevalence and conservation of m6A residues on nuclear DNA and RNA viruses, combined with the limited number of reports looking at how m6A affects the replication of HIV-1, clearly suggest that m6A addition enhances viral gene expression and, hence, replication. However, the mechanistic basis for this positive effect currently remains unclear. We anticipate that ongoing efforts to precisely map m6A sites on viral transcripts, combined with the targeted mutagenesis of these m6A addition sites, will shed additional light on this question in the near future.

**CYTOPLASMIC RNA VIRUSES**

As noted above and demonstrated in Fig. 2, the cellular m6A writers METTL3, METTL14, and WTAP are all localized to the nucleus at steady state (2, 3). However, it has also been reported that METTL3 and METTL14 can be detected in the cytoplasm (12, 13, 54), suggesting that these proteins have the ability to shuttle between the nucleus and the cytoplasm and/or to enter the cytoplasm in response to stress. If the m6A writers are indeed able to access the cytoplasm, then this raises the possibility that cytoplasmic viruses might also encode mRNAs bearing m6A residues. In fact, analyses of hepatitis C virus (HCV) and several different flaviviruses, including Zika virus, Dengue virus, yellow fever virus, and West Nile virus, have revealed at least 5 and up to 16 m6A modification sites on the RNA genomes of these viruses (12, 13). In the case of HCV, the effect of m6A modifications has been analyzed in detail, and surprisingly and in marked contrast to HIV-1, knockdown of METTL3 and METTL14 mRNA using RNAi enhanced the production of infectious HCV virions, and knockdown of the mRNAs encoding the YTHDF proteins had a similar positive effect (12). Interestingly, HCV mRNA translation and RNA replication were both unaffected, thus suggesting that m6A on HCV RNAs might directly regulate the production of infectious HCV virions. Indeed, immunofluorescence analysis of HCV-infected cells showed that YTHDF proteins and the HCV structural proteins colocalize to the lipid droplets that function as sites of HCV virion morphogenesis, consistent with a direct role for m6A in regulating HCV virion production (12). Similarly, in the case of Zika virus, knockdown of METTL3 or METTL14 mRNA was also reported to enhance the production of Zika virions, while knockdown of ALKBH5 mRNA exerted an opposite inhibitory effect (13).

In general, viruses, especially RNA viruses that rely on virally encoded, error-prone RNA-dependent RNA polymerases, can rapidly evolve to inactivate sequences present on the viral RNA genome, such as targets for small interfering RNAs, that inhibit their replication in cis (55). Similarly, m6A addition to viral RNAs, which requires the consensus sequence 5'-RRm6ACH-3', would also be easy for a virus to avoid if m6A indeed exerted an inhibitory effect in cis. It could be argued that for a virus that establishes long-term persistent infections, such as HCV, it might be advantageous to downregulate the rate of viral replication so as to mitigate host immune responses. However, this argument makes little sense in the case of Zika virus or the other flaviviruses listed in Table 1, which cause acute infections marked by high viremia, which are generally rapidly cleared by the host adaptive immune response. Thus, the fact that multiple m6A residues have been detected on all the flaviviruses analyzed so far (Table 1) argues that m6A addition has been selected for, rather than against, during flavivirus evolution. The observation that m6A can inhibit the release of infectious virions by Zika virus-infected cells is therefore difficult to understand. It is possible that m6A, as noted above, enables
Zika virus to avoid the viral RNA-induced activation of innate antiviral immune responses, which might balance or enhance viral replication in vivo (52, 53). However, one would then have to argue that these antiviral responses have been lost in the cells used to analyze Zika virus growth in culture. Indeed, the Vero cells that were exclusively used by Linchinchi et al. (13) are known to be unable to mount an interferon response (56). Additional experiments using other cells that are fully competent to mount antiviral innate immune responses and cells in which m6A addition has been knocked out by gene editing, rather than knocked down using RNAi, are needed to resolve this conundrum.

m6A AS A TARGET FOR ANTIVIRAL THERAPY

If m6A indeed normally functions to enhance viral replication, as implied by the conservation of m6A on transcripts produced by diverse virus families (Table 1) and also supported by data generated using HIV-1 (10, 11), then m6A addition presents itself as a possible target for antiviral agents. The advantage of drugs that inhibit cellular proteins required for virus replication is that they make it very difficult for the virus to evolve resistance, while the disadvantage is that they can inhibit the normal physiological function of that protein and, hence, cause toxicity. So, is m6A addition a potential target for antiviral drug development? In fact, several lines of data suggest that this might be the case. Specifically, the S-adenosylhomocysteine (SAC) hydrolase inhibitor 3-deazaadenosine (DAA) has been shown to inhibit m6A addition and to act as a broad antiviral inhibitor (57–61).

The inhibition of SAC activity by DAA results in the accumulation of SAC in cells, which in turn results in depletion of S-adenosylmethionine (SAM), the methyl donor used by METTL3 to generate m6A (57) (Fig. 1). As SAM is used as a methyl donor by a wide range of cellular methylases, DAA is clearly not a specific inhibitor of m6A formation, though mRNA capping has been shown to be unaffected by DAA treatment (62). So, is DAA too toxic to use as an antiviral? In fact, several papers have reported using DAA to inhibit the replication of diverse viruses, including Rous sarcoma virus, HIV-1, respiratory syncytial virus, parainfluenza virus, vesicular stomatitis virus, measles virus, and reovirus (57–61), in cultured cells at concentrations that did not show any detectable cytopathic effects. Even more impressively, DAA was found to effectively block respiratory syncytial virus replication in cotton rats (58) and Ebola virus-induced fatality in mice (63, 64) at doses that did not give rise to any evident toxicity. Importantly, DAA is not incorporated into cellular nucleic acids (57) and does not have the structure expected for a nucleoside that can function as a chain terminator. Thus, it appears probable that it is indeed the inhibition of SAC hydrolase activity that underlies this inhibitory effect, although whether m6A addition is indeed the key target for DAA remains uncertain (57). Nevertheless, these observations are consistent with the hypothesis that m6A addition plays an important positive role in the life cycle of a wide range of viruses and suggest that an inhibitor that can specifically target METTL3 activity, rather than SAM-dependent methylation in general, might be well tolerated and could prove to be an effective broad-spectrum antiviral, especially for viruses that cause acute infections and disease. Given recent data suggesting that excessive m6A modification of cellular mRNAs might also contribute to the progression of some forms of cancer, such as acute myeloid leukemia (65), efforts to identify specific inhibitors of m6A addition would seem to be very timely.

CONCLUSIONS AND FUTURE DIRECTIONS

While the emerging field of viral epitranscriptomics is clearly in its infancy, we nevertheless feel that the limited data reported thus far are consistent with the hypothesis that m6A will emerge as a ubiquitous modification of viral RNA transcripts that profoundly influences several different aspects of the viral life cycle. Exactly how m6A exerts its phenotypic effects at a mechanistic level is still largely unclear in not only the viral but also cellular context, but there is no question that this area has now become the subject of an intense research effort that has begun to clarify aspects of
this problem. Clearly, the next step will be to precisely map and then mutate m6A residues found on different viral genomes and then study the phenotypic consequences. Obviously, it will be critical to ensure that any observed inhibition of viral replication is indeed due to loss of m6A rather than to the inactivation of some other cis-acting RNA sequence. To control for this potential problem, and assuming that the observed phenotype is not too severe, one could test m6A-deficient viral mutants not only in wild-type cells but also in METTL3 knockout cells, where the mutant and parental viruses should replicate at equivalent levels. Such specific viral mutants should then enable a precise definition of how the addition of m6A to viral mRNAs regulates viral gene expression and replication.

ACKNOWLEDGMENTS

This work was supported in part by a National Institutes of Health grant (R21-AI130574) to B. R. Cullen. K. Tsai was supported by National Cancer Institute grant T32-CA009111.

We thank Joy Marshall for assistance with immunofluorescence analysis.

REFERENCES


