MOLECULAR AND CELLULAR BIOLOGY, Sept. 1985, p. 2298–2306
0270-7306/85/092298-09$02.00/0

Vol. 5, No. 9

# Precise Localization of m$^6$A in Rous Sarcoma Virus RNA Reveals Clustering of Methylation Sites: Implications for RNA Processing

SUSAN E. KANE AND KAREN BEEMON*

*Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218*

Received 29 April 1985/Accepted 12 June 1985

**N$^6$-methyladenosine (m$^6$A) residues are present as internal base modifications in most higher eucaryotic mRNAs; however, the biological function of this modification is not known. We describe a method for localizing and quantitating m$^6$A within a large RNA molecule, the genomic RNA of Rous sarcoma virus. Specific fragments of $^{32}$P-labeled Rous sarcoma virus RNA were isolated by hybridization with complementary DNA restriction fragments spanning nucleotides 6185 to 8050. RNA was digested with RNase and fingerprinted, and individual oligonucleotides were analyzed for the presence of m$^6$A by paper electrophoresis and thin-layer chromatography. With this technique, seven sites of methylation in this region of the Rous sarcoma virus genome were localized at nucleotides 6394, 6447, 6507, 6718, 7414, 7424, and 8014. Further, m$^6$A was observed at two additional sites whose nucleotide assignments remain ambiguous. A clustering of two or more m$^6$A residues was seen at three positions within the RNA analyzed. Modification at certain sites was found to be heterogeneous, in that different molecules of RNA appeared to be methylated differently. Previous studies have determined that methylation occurs only in the sequences Gm$^6$AC and Am$^6$AC. We observed a high frequency of methylation at PuGm$^6$ACU sequences. The possible involvement of m$^6$A in RNA splicing events is discussed.**

In addition to having 5′-methylated cap structures, many eucaryotic mRNAs contain internal methylated bases. N$^6$-methyladenosine (m$^6$A), the major form of modified nucleoside, has been detected in most higher eucaryotic cell mRNAs, in viral mRNAs, and in the virion RNA of retroviruses (2). Methylation of internal adenosines occurs in the nucleus very rapidly after synthesis of the heterogeneous nuclear RNA precursor to mRNA and before the RNA is spliced (9, 28). While the function of this internal methylation is not known, several observations suggest a significant role for m$^6$A in RNA metabolism.

Some implications for the function of internal base modifications come from methylation inhibition studies, using either cycloleucine or *S*-tubercidinylhomocysteine to block internal methylation of RNA. These experiments suggest an involvement of m$^6$A in RNA processing events or in the transport of mRNA from the nucleus to the cytoplasm (6, 13, 35). It is not clear from these studies, however, that the effect of the inhibitors is directly related to the methylation state of the mRNA being examined. Secondary effects of the drugs have not been ruled out. Furthermore, the requirement for internal methylation in mRNA is not absolute, as mature messages without any detectable m$^6$A residues do exist, including globin and histone mRNAs (24, 27).

In all cases characterized to date, m$^6$A occurs only in the sequences Gm$^6$AC and Am$^6$AC (7, 12, 31, 38). Preliminary localization studies with specific RNAs have determined that the distribution of m$^6$A within a molecule is nonrandom. Rous sarcoma virus (RSV), for example, has an average of 10 to 15 m$^6$A residues per virion RNA molecule (4, 14, 36). These methylated bases have been detected only in the 3′ half of the genomic RNA (4), even though sequence analysis shows that putative methylation sites (GAC and AAC) are widely dispersed throughout the molecule (32). Likewise, internal methylation of bovine prolactin mRNA is confined to the 3′ two-thirds of the message, with a high concentration of m$^6$A in the 3′ noncoding region (17). Simian virus 40 (SV40) late mRNAs, on the other hand, each contain an average of three m$^6$As, all near the 5′ ends of those messages in translated sequences (7). Adenovirus mRNAs are methylated in the 5′ two-thirds of the molecules, and m$^6$As seem, in this system, to be conserved during mRNA processing (9, 33).

We have developed a method for directly identifying m$^6$A residues within the genomic RNA of RSV. We have precisely mapped seven such sites in the genome, all within protein coding regions. In addition, we have analyzed the extent of modification at individual methylation sites. Earlier studies with SV40 (7) and with RSV (12) suggest that methylation at some specific sites is heterogeneous, with base modification occurring less than 100% of the time. We now present direct evidence of such heterogeneity. Six of seven m$^6$A sites identified were at PuGACU sequences, suggesting a preferred sequence for methylation in RSV RNA. Further, we observed clustering of two or more m$^6$A residues at several specific regions in the RNA.

## MATERIALS AND METHODS

**Cells and viruses.** Prague strain, subgroup B, (PR-B) RSV virus stocks were originally provided by P. K. Vogt. PR-C RSV was recloned from virus stocks provided by R. Guntaka. Virus was grown on gs$^-$ chf$^-$ chicken embryo fibroblasts (CEF) that were prepared from eggs supplied by SPAFAS, Inc., Norwich, Conn. A full-length clone of Schmidt-Ruppin strain, subgroup A (SR-A) RSV provirus in pBR322 (SRA-2) was obtained from J. M. Bishop (11). Full-length PR-C RSV DNA in pBR322 (pATV8) was obtained from R. Guntaka (18). M13 mp11 (23) and mp18 (25) and *Escherichia coli* JM103 (22) and JM109 were provided by R. C. Huang.

**Radioactive labeling of RSV RNA.** Plates (diameter, 15 cm) of CEF infected with PR-B were preincubated for 1 to 2 h in 15 ml of medium 199 without methionine plus 2% dialyzed calf serum–20 μM guanosine–30 μM adenosine–20 mM sodium formate (for [*methyl*-$^3$H]methionine labeling) or in medium 199 plus 2% calf serum (for [$^3$H]uridine labeling). After preincubation, medium was removed and replaced

* Corresponding author.

with 15 ml of the same medium containing 3 mCi of [*methyl*-$^3$H]methionine (83 Ci/mmol Amersham) or [$^3$H]uridine (44 Ci/mmol, ICN). Virus samples were collected after 12 h, cells were incubated with additional labeling medium, and virus samples were collected after another 12 h. Medium 199 plus 2% tryptose phosphate broth–1% calf serum–1% chicken serum (2:1:1) was added, and virus samples were collected after 12 more h. For $^{32}$P labeling, 9-cm plates of CEF infected with PR-C RSV were preincubated for 1 h in 7 ml of medium 199 without phosphate plus 2% dialyzed calf serum–1% dimethyl sulfoxide. Medium was removed and replaced with 7 ml of the same medium containing 3 mCi of $^{32}P_i$ (carrier free, ICN). Virus samples were collected after 20 to 24 h, medium 199 (2:1:1) plus 1% dimethyl sulfoxide was added for another 20 to 24 h, and virus samples were again collected.

**Preparation of RNA.** Virus collections were pooled, and cells were removed by centrifugation at 5,000 rpm for 20 min. Virus was pelleted in a Beckman SW27 rotor at 26,000 rpm for 90 min or in a Beckman SW41 rotor at 39,000 rpm for 40 min. Pellets were suspended in 1 ml of standard buffer (0.1 M NaCl, 10 mM Tris-hydrochloride [pH 7.4], 1 mM EDTA) and extracted three times with phenol-chloroform (1:1) in the presence of 1% sodium dodecyl sulfate, 0.1 ml of β-mercaptoethanol, and 50 μg of carrier RNA. Virion RNA was precipitated with ethanol, and 70S genomic RNA was isolated by centrifugation in a linear 10 to 25% sucrose gradient (4). Peak fractions were pooled and precipitated with ethanol.

**Gel purification of DNA.** pATV8 DNA was digested with *Ava*I restriction endonuclease, and the resulting fragments were separated on a 1.5% low-melting-temperature agarose preparative gel. Individual fragments were extracted from the gel by standard procedures (21). Some of the *Ava*I fragments were cut again with additional restriction enzymes and put through another round of gel purification to isolate *Ava*I subfragments. The purity of each DNA preparation was verified by hybridization in solution with $^{32}$P-labeled RNA and resolution of resulting DNA-RNA hybrids by agarose gel electrophoresis.

**Subcloning into M13.** pATV8 DNA was digested with *Pst*I, and the resulting fragments were ligated into the *Pst*I site of the M13 vector, mp11. DNA was used to transform *Escherichia coli* JM103. White plaques were picked, and single-stranded phage DNAs were screened for the desired insert orientation by hybridization with $^{32}$P-labeled viral RNA. Individual positive clones were characterized by restriction analysis to identify their *Pst*I inserts. In a separate round of cloning, pATV8 was digested with *Xba*I, and a DNA fragment spanning nucleotides (nt) 3378 to 6861 was isolated from low-melting-temperature agarose. This fragment was ligated into M13 mp18, and DNA was used to transform *E. coli* JM109.

**Preparation of DNA filters.** Subcloned fragments of pATV8 were isolated as single- or double-stranded M13 DNA by standard procedures (30). Single-stranded DNA (20 μg) was diluted into 2 ml of 5× SSC (1× SSC = 0.15 M NaCl plus 0.015 M sodium citrate [pH 7.0]) and filtered onto 13-mm nitrocellulose circles, and the filters were baked. Double-stranded replicative-form DNA was used for purifying cloned insert from vector DNA. After digesting replicative-form DNA with *Pst*I or *Xba*I, insert DNA was isolated either by low-melting-temperature agarose gel purification, as before, or by directly transferring the insert DNAs from agarose gels to nitrocellulose strips by the method of Southern (34). In some cases, the insert DNA was further digested into subfragments before running on a gel for purification or transfer. Gel-purified DNA was denatured in 0.1 N NaOH for 30 min at room temperature before being filtered onto nitrocellulose circles as for single-stranded DNA. For reuse in multiple hybridizations, filters were washed in 2× SSC–0.1 N NaOH for 30 min at room temperature and then rinsed five times with 2× SSC and once with water (21).

**Solution hybridization.** Solution hybridizations were performed essentially by the method of Casey and Davidson (8). Radioactively labeled RNA was precipitated in the same tube with the appropriate DNA and then suspended in 25 μl of hybridization buffer (80% formamide, 400 mM NaCl, 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid), pH 6.4], 1 mM EDTA, 0.1% sodium dodecyl sulfate). For $^3$H experiments, PR-B RSV 70S RNA labeled with [$^3$H]methionine or [$^3$H]uridine (250,000 cpm) was hybridized with total DNA from restriction digests of SR-A RSV DNA. RNA and DNA were denatured at 80°C for 10 min and immediately transferred to 57°C for 3 h. Hybridization mix was diluted to 400 μl with 0.3 M NaCl–10 mM Tris (pH 7.4)–1 mM EDTA (buffer A), precipitated with ethanol, and suspended in 15 μl of buffer A. Unhybridized RNA was digested for 1 to 2 h at 37°C with 0.5 μg of RNase A. Hybrid fragments were separated on 1.5% agarose gels. Gels were subjected to fluorography, as described by Laskey (20), and were exposed to presensitized Kodak X-Omat AR (XAR-5) film. For $^{32}$P experiments, PR-C RSV 70S RNA ($10^6$ cpm) was hybridized with individual gel-purified restriction fragments of PR-C RSV DNA. RNA and DNA were denatured at 80°C for 10 min and transferred to 52°C for 12 to 16 h. Hybridization mix was diluted and precipitated, as for $^3$H experiments, and unhybridized RNA was digested with 5 U of RNase $T_1$. RNA-DNA hybrids were separated from digested nucleotides by Bio-Rad P-30 or P-100 column chromatography (21). Fractions containing hybrids were pooled, precipitated with ethanol, and resuspended in 3 μl of TE (10 mM Tris [pH 7.4], 1 mM EDTA). Hybrids were denatured by heating at 90°C for 1.5 min and quenched, releasing RNA for fingerprinting.

**Filter hybridization.** Filter hybridizations were performed essentially by the method of Gillespie and Spiegelman (15). All hybridizations were carried out in siliconized glass vials (13 by 57 mm). Multiple hybridizations were performed in a single vial. DNA filters were prehybridized for 3 h at 42°C in 50% formamide, 5× Denhardt solution (1× Denhardt = 0.02% bovine serum albumin, 0.02% polyvinyl-pyrolidine, 0.02% Ficoll)–1% sodium dodecyl sulfate–2× SSC–single-stranded salmon sperm DNA (100 μg/ml)–polyadenylic acid (100 μg/ml). Hybridizations were in 50% formamide–0.7 M NaCl–10 mM PIPES (pH 6.6)–4 mM EDTA–0.1% sodium dodecyl sulfate containing $1 \times 10^6$ to $2 \times 10^6$ cpm of $^{32}$P-labeled RNA. Vials were placed at 65°C for 10 min to denature RNA, then transferred immediately to 42°C, and hybridized with shaking for 16 to 20 h. Filters were removed from hybridization buffer (buffer was saved for subsequent rounds of hybridization) and washed with 3 volumes of 0.3 M NaCl–10 mM Tris-hydrochloride (pH 7.4) (buffer B). Filters were incubated for 3 h at room temperature in 0.3 ml of buffer B containing 10 U of RNase $T_1$ and then washed with buffer B until $^{32}$P in the wash was at background levels. Hybridized RNA was eluted by boiling in water for 1 min and quenching. RNA was precipitated with ethanol and suspended in 20 μl of TE for fingerprinting.

**RNA fingerprinting and oligonucleotide analysis.** RNA from hybridizations, resuspended in TE, was digested with 5 U of RNase $T_1$ for 1 h at 37°C. Oligonucleotides were

separated in two dimensions as described by Barrell (3). The first dimension was on cellulose acetate strips (Schleicher & Schuell) in pH 3.5 buffer (5% acetic acid, 0.5% pyridine). Strips were presoaked in buffer containing 7 M urea. The second dimension was either on DEAE-cellulose paper (Whatman DE81) in 7% formic acid or on DEAE-cellulose plates by homochromatography with homomix c (3). The fingerprint was exposed to XAR-5 film and candidate oligonucleotides were identified and eluted with 30% triethylammonium carbonate, pH 10.0 (3). Eluted oligonucleotides were dried, washed with water, and then resuspended in 10 μl of 50 mM ammonium acetate (pH 4.5) plus 5 U of RNase $T_1$, 0.04 U of RNase $T_2$, and 0.5 μg of RNase A. After digestion at 37°C for 1 h, the 3'-phosphate mononucleotides were dried in a desiccator, washed, and resuspended in water for spotting onto 3MM paper (46- by 57-mm sheets). A marker lane of $^{32}$P-labeled RNA digested with RNase A-RNase $T_1$-RNase $T_2$ was also spotted. Mononucleotides were separated by electrophoresis in pH 3.5 buffer. The marker lane was visualized by autoradiography and used as a guide for excising the Ap/m$^6$Ap spot from each sample lane. Nucleotides were eluted from 3MM with water, dried, and assayed for m$^6$Ap by thin-layer chromatography on cellulose plates. Plates were developed in isopropanol–5% ammonium acetate (pH 3.5) (65:57.5) to separate m$^6$Ap and Ap. The identity of oligonucleotides on fingerprints was determined by partial sequence analysis. $T_1$ oligonucleotides were eluted and digested with RNase A, and the products were electrophoresed on DEAE-cellulose paper at pH 3.5 (3).

## RESULTS

**Localization of *methyl*-$^3$H-labeled RSV RNA to restriction fragments.** Methylation of RSV RNA consists of a 5'-methylated cap sequence and internal methylated m$^6$A residues (14). The genomic RNA of PR-B RSV contains an average of 12 m$^6$A residues per 9,500 nt, whereas 15 and 10 m$^6$A residues have been observed in B77 and SR-A RSV RNAs, respectively (4, 14, 36). The majority of these modified bases are localized in the 3' half of the molecule, with about five m$^6$A's, on the average, in the *src* region of the genome (4).

To confirm and extend these findings by a more direct method, PR-B RSV genomic RNA was labeled in vivo with either [*methyl*-$^3$H]methionine, so that $^3$H would be incorporated only into methylated nucleotides, or with [$^3$H]uridine. The labeled RNA was hybridized in solution to cloned SR-A RSV DNA that had been digested with various restriction enzymes. Unhybridized RNA was digested with RNase, and the remaining DNA-RNA hybrid fragments were separated on agarose gels.

Figure 1 illustrates the results from four such hybridizations. As expected, *methyl*-$^3$H-labeled RNA predominantly hybridized to restriction fragments corresponding to the 3' half of the genome and [$^3$H]uridine-labeled RNA hybridized to all fragments. The intensity differences between hybrid bands in the [$^3$H]uridine lanes denoted differences in hybridization efficiencies for different restriction fragments, probably due to variable GC contents or to variable numbers of mismatches between PR-B and SR-A virus strains. Nonetheless, there were some weakly hybridizing [$^3$H]uridine bands which showed up strongly in the [$^3$H]methionine lanes, and, conversely, some strongly hybridizing [$^3$H]uridine bands which were much less prominent in the [*methyl*-$^3$H]methionine lanes. Bands having a high [*methyl*-$^3$H]/[$^3$H]uridine ratio appeared to be heavily methylated. For example, when DNA was digested with *Bam*HI, fragments spanning nt 3710 to 9300 and 1 to 525 (Fig. 1, bands 6640 and 4040) were highly methylated. (These numbers are approximate, based on the SRA-2 DNA restriction map determined by DeLorbe et al. [11].) Fragments extending from nt 525 to 3710, on the other hand, were not labeled with [*methyl*-$^3$H]-methionine (Fig. 1, *Bam*HI, bands 1350 and 1800). Similar results were obtained from the other hybridizations. From these, we localized *methyl*-$^3$H-labeled residues between nt 4500 and the 3' end of the RNA, in the *env* and *src* regions of the genome. *methyl*-$^3$H label also was detected in the 5' end of the virion, probably due to methylations in the cap sequence.

This approach to localization proved to be inadequate for a number of reasons. The differential hybridization efficiencies between fragments made the quantitation of m$^6$A levels difficult. Furthermore, the level of resolution of the technique, limited by the sensitivity of *methyl*-$^3$H detection, was not sufficient to localize m$^6$A onto small restriction fragments. This low sensitivity also made it necessary to use very long exposure times—up to 18 months—to see an adequate signal in the *methyl*-$^3$H lanes. It should be noted that on these long exposures, we saw faint labeling of nearly all of the restriction fragments in the *methyl*-$^3$H lanes. We did not determine whether this was due to nonspecific hybridization or to low levels of methylation in the corresponding regions of the RNA.

**Method for precise localization of m$^6$A residues.** The following procedure was devised as a more exact means of localizing m$^6$A residues. For all subsequent studies, RNA from the PR-C strain of RSV and DNA from a full-length clone of PR-C (pATV8) were used. The entire nucleotide sequence of this clone has been determined (32). A map of relevant restriction endonuclease sites is presented in Fig. 2. DNA from subclones or from gel-purified restriction fragments of pATV8 was immobilized on individual nitrocellulose filters and hybridized with genomic RNA labeled in vivo with $^{32}P_i$. In some experiments, hybridization was performed in solution. Unhybridized RNA was removed, and hybrids were denatured to release [$^{32}$P]RNA corresponding to the region of the genome being analyzed (see above). This RNA was digested with RNase $T_1$, and the resulting oligonucleotides were separated in two dimensions. RNA from each genomic region was fingerprinted both by paper electrophoresis and by homochromatography in the second dimension. Representative $T_1$ fingerprints are depicted in Fig. 3. Since the sequence of each isolated RNA fragment is known from the sequence of pATV8 and since the target site for methylation is either GAC or AAC (12), it was possible to predict which $T_1$ oligonucleotides might contain one or more m$^6$A residues.

These candidate oligonucleotides were eluted from the RNA fingerprint and further digested with a mixture of RNases A, $T_1$, and $T_2$. The resulting 3'-phosphate mononucleotides were analyzed in two dimensions to separate m$^6$Ap from unmodified nucleotides (see above; Fig. 4). Methylated residues were thereby located within specific oligonucleotides and within the genomic sequence. The sequence of each positive oligonucleotide was confirmed by eluting that same spot from a parallel $T_1$ fingerprint, digesting with RNase A, and separating the resulting products by electrophoresis on DEAE-cellulose paper in pH 3.5 buffer (3) (data not shown). For each fragment of RNA, the two fingerprinting methods led to the localization of m$^6$A within the same oligonucleotides.

Using this technique, we analyzed a heavily methylated

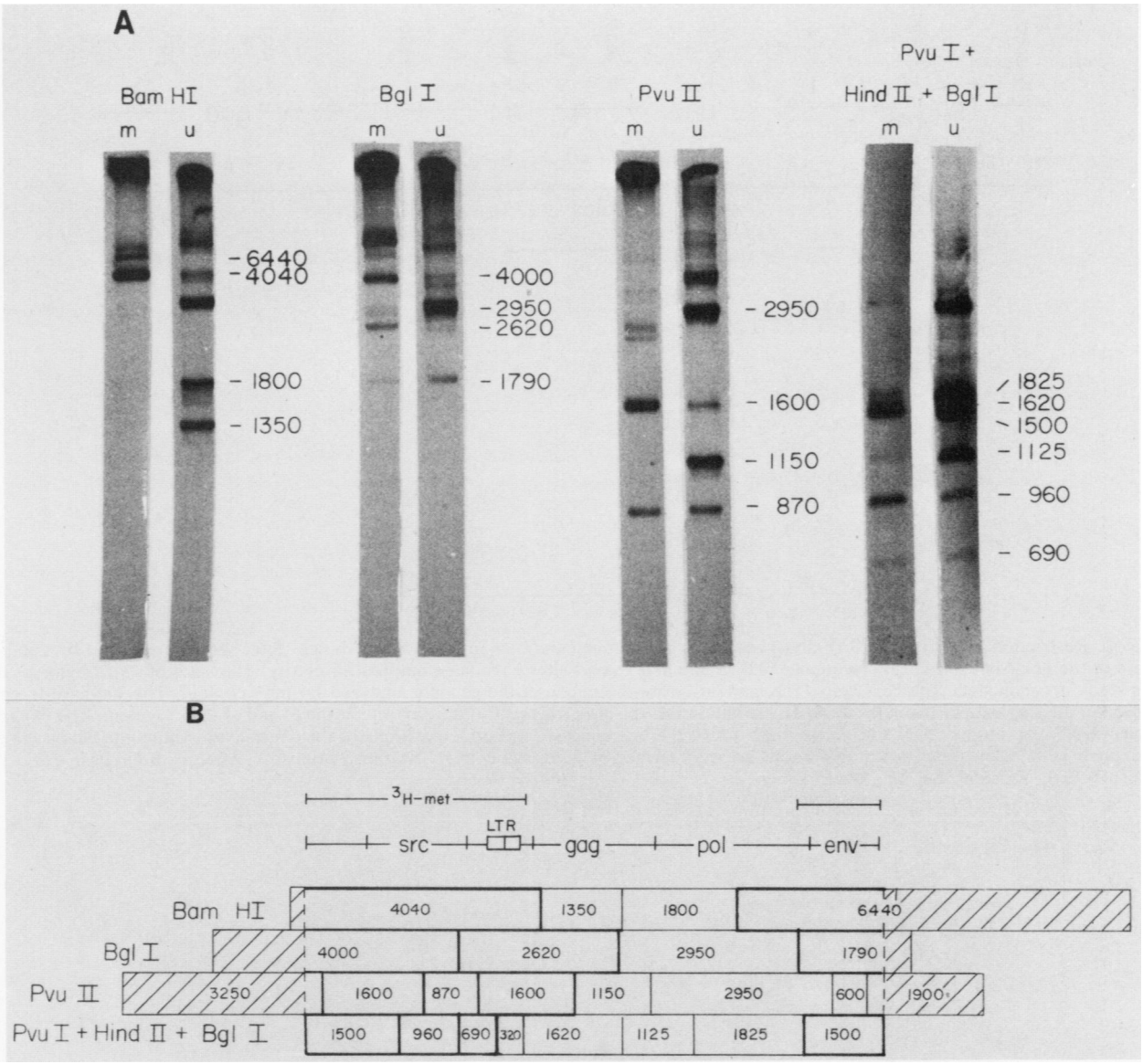


FIG. 1. Localization of methylated regions of the RSV genome. PR-B RSV RNA labeled with [*methyl*-$^3$H]methionine or [$^3$H]uridine was hybridized in solution with SR-A DNA that had been digested with the designated restriction endonucleases. (A) RNA-DNA hybrid fragments were separated on a 1.5% agarose gel, and radioactivity was visualized by fluorography. m, [*methyl*-$^3$H]methionine; u, [$^3$H]uridine. Lanes representing *methyl*-$^3$H labeling were exposed for 1 year (*Bam*HI, *Bgl*I, and *Pvu*II) or 1.5 years (*Pvu*I + *Hin*dII + *Bgl*I). [$^3$H]uridine lanes were exposed for 1 month. (B) A restriction map of SRA-2 showing expected fragment sizes resulting from restriction endonuclease digestions. Crosshatched regions indicate pBR322 sequences present in SRA-pBR322 junction fragments. Positions of the corresponding fragments on the agarose gels, detected by ethidium bromide staining, are indicated in (A). Extra, unidentified hybrid bands were seen reproducibly in this type of experiment. These bands were not further characterized. Fragments hybridizing with *methyl*-$^3$H-labeled RNA are outlined in bold ink on the restriction map in (B). These fragments include the 5′-methylated cap sequence as well as sequences mapping between nt 4500 and the 3′ terminus of the viral genome, as indicated above the map.

region of PR-C genomic RNA, extending from nt 6185 to 8050, and located seven sites of internal methylation. These lie at positions 6394, 6447, 6507, 6718, 7414, 7424, and 8014 (nt numbering is as described by Schwartz et al. [32]). Initially, hybridizations were performed in solution with gel-purified restriction fragments from the full-length clone, pATV8. This led to the identification of sites 7414 and 7424 within RNA complementary to a *Hin*dII-*Ava*I fragment spanning nt 6983 to 7459. When this RNA was digested with RNase $T_1$, nt 7414 migrated as (G)ACUG in a two-dimensional fingerprint, and 7424 was found as (G)ACUUG (Fig. 3). Each of these oligonucleotides is present only once in this 476-base region of RNA, so the identification of the m⁶A sites was unambiguous here.

Solution hybridization of a 503-base *Ava*I fragment (nt 7916 to 8419) led to the detection of an m⁶A signal in an oligonucleotide migrating as $AC_2U_2G$. However, both ACUCUG (A = 8014) and ACUUCG (A = 8339) are present in this fragment, and both are candidates for methylation. Thus, this identification of an m⁶A site was ambiguous.

**Filter hybridizations with *Pst*I subclones.** To clearly identify this m⁶A site and to confirm the results for positions 7414 and 7424, we generated a set of *Pst*I restriction fragments to use in hybridizations (Fig. 2). In addition, since gel-purified

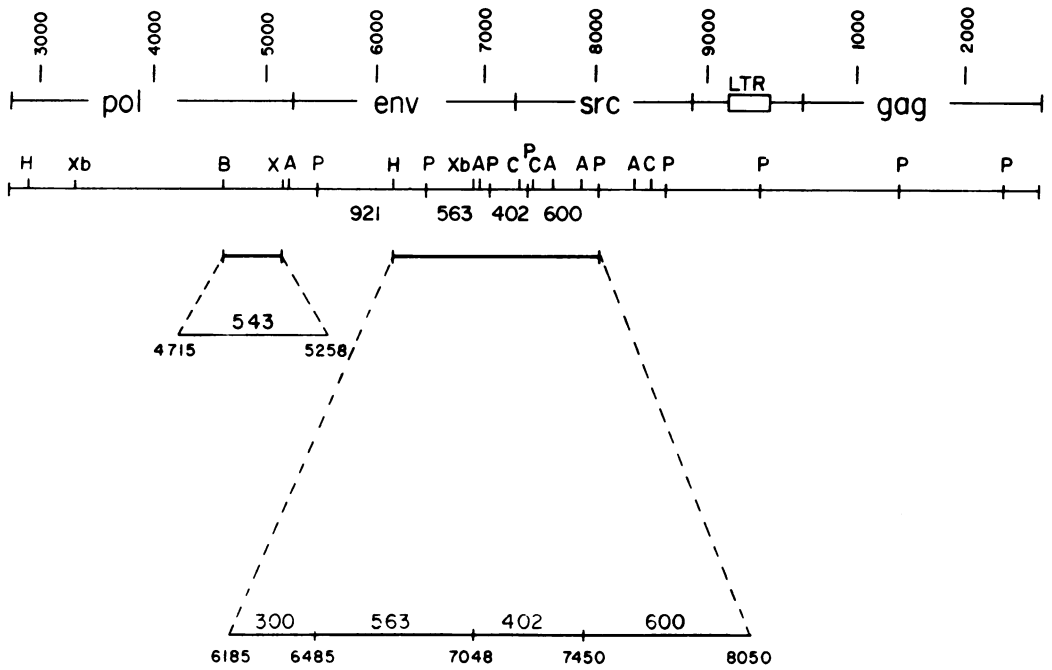


FIG. 2. Restriction map of pATV8. Relevant restriction endonuclease sites in pATV8 are shown. Sites were determined by computer analysis of the published PR-C RSV sequence (32). Some of the *Ava*I sites have been omitted for clarity. The sizes of *Pst*I fragments used in this study are indicated. The bars below the map designate the region of the genome analyzed for $m^6A$ content. The 300-base fragment discussed in the text extends from the *Hin*dIII site within the 921-base fragment to the *Pst*I site at the 3′ end of that fragment. The expanded portions of the map indicate DNA fragment sizes above the line and the nucleotide positions of their respective endpoints below the line. Approximate gene delineations and nt numbering are as described by Schwartz et al. (32). Abbreviations: A, *Ava*I; B, *Bam*HI; C, *Hin*dII; H, *Hin*dIII; P, *Pst*I; X, *Xho*I; and Xb, *Xba*I.

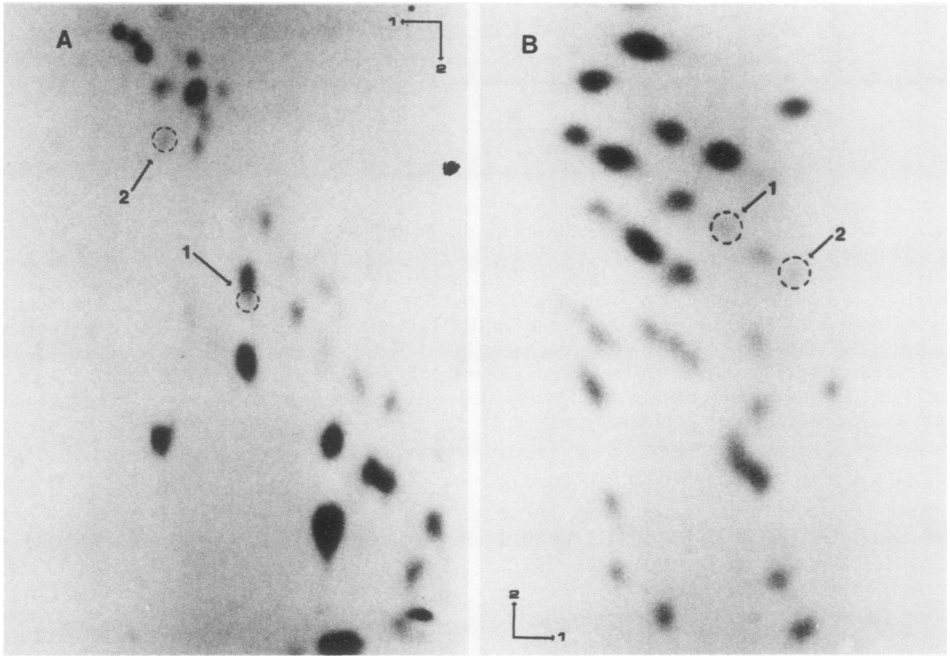


FIG. 3. $T_1$ oligonucleotide fingerprints. Representative $T_1$ fingerprints of $^{32}P$-labeled RNA isolated by hybridization with a DNA restriction fragment are shown. The first dimension in both cases was on cellulose acetate in 5% acetic acid–0.5% pyridine (pH 3.5). The second dimension was on DEAE-cellulose paper in 7% formic acid (A) or on DEAE-cellulose plates by homochromatography (B). The region represented here is the 402-base *Pst*I fragment, nt 7048 to 7450. The circled oligonucleotides were found to be positive for $m^6A$. Circled oligonucleotides: 1, ACUG (nt 7414); 2, ACUUG (nt 7424).

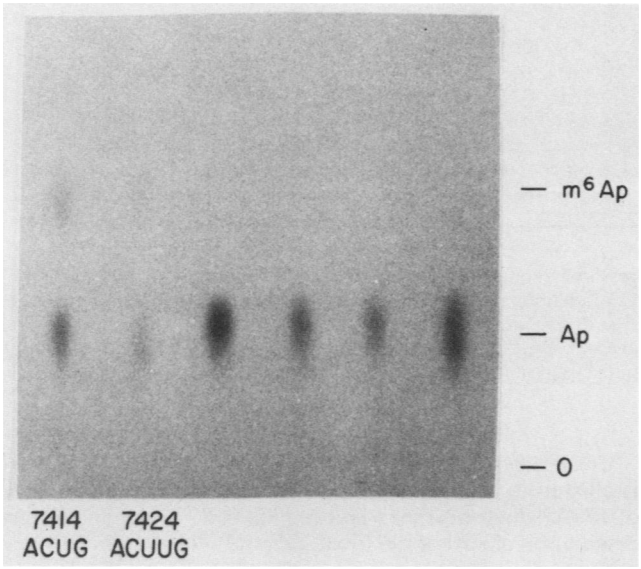


FIG. 4. Separation of m⁶Ap and Ap by thin-layer chromatography. Oligonucleotides eluted from a fingerprint were digested to mononucleotides, and m⁶Ap was separated from unmodified nt as detailed in the text. Presented here is an example of the final thin-layer chromatographic dimension resolving m⁶Ap and Ap. Shown are the m⁶Ap/Ap spots from six oligonucleotides eluted from a $T_1$ fingerprint of the 402-base *Pst*I fragment. Oligonucleotide sequences and nt positions of the two positive methylation sites are indicated. Bars designate positions of cold markers. O, Origin.

fragments are often contaminated with trace amounts of other DNA fragments, we subcloned the *Pst*I fragments into M13 (see above). From these clones, either single-stranded phage DNAs or purified inserts from replicative-form DNA were used in filter hybridizations with viral RNA. There were additional advantages to this filter hybridization procedure. Many different DNA filters could be used in a single hybridization, and both DNA filters and unhybridized RNA could be reused in subsequent rounds of hybridization. This allowed very efficient use of relatively small amounts of starting materials.

Filter hybridizations with the 402-base *Pst*I fragment subclone (nt 7048 to 7450) again identified positions 7414 and 7424 as sites of methylation. The 600-base *Pst*I fragment (nt 7450 to 8050) conveniently includes nt 8014 but does not contain nt 8339. In this fragment, a $T_1$ oligonucleotide migrating as $AC_2U_2G$ is uniquely identified as ACUCUG, with the A being in position 8014. In fact, hybridizations with the 600-base fragment detected m⁶A in the ACUCUG oligonucleotide at base 8014. The complementary hybridization with the 612-base fragment, which harbors nt 8339, has not been performed. A second oligonucleotide in the 600-base region was also identified as having m⁶A. However, this oligonucleotide migrated as $AC_2UG$ and could be either ACUCG (A = 7890) or ACCUG (A = 7981). We have not been able to resolve these two sites, so assignment of this m⁶A remains ambiguous.

Hybridizations with the 563-base *Pst*I fragment (nt 6485 to 7048) and with a subfragment spanning nt 6485 to 6861 have unambiguously identified nt 6718 as an m⁶A residue, present in the $T_1$ oligonucleotide (G)ACUG. In addition, m⁶A was detected in a large $T_1$ oligonucleotide from this region, identified as (G)ACUUCUUG (A = 6507).

The 921-base *Pst*I fragment was further divided into two subfragments by digestion with *Hin*dIII. Filter hybridizations with the 300-base subfragment (nt 6185 to 6485) revealed two more m⁶A positions. One resided in a unique $T_1$ oligonucleotide, (G)ACUAG, with m⁶A at residue 6394. The other site was in oligonucleotide (G)ACUUAUUG (A = 6447).

Another M13 subclone of pATV8 containing PR-C sequences 3378 to 6861 (*Xba*I fragment) was used to isolate a 543-base *Bam*HI-*Xho*I fragment spanning nt 4715 to 5258. In this region, we detected m⁶A in an oligonucleotide migrating as ACACUG (nt 4976). However, we cannot rule out contamination of this spot by oligonucleotide ACUAAG (nt 5044), which runs next to ACACUG in both types of fingerprinting techniques utilized here. The location of this m⁶A residue therefore remains ambiguous.

A summary of the m⁶A sites located by these fragment hybridizations is presented in Table 1. Additional analysis of these *Pst*I fragments led to the characterization of many putative methylation sites which clearly were not modified. The corresponding oligonucleotides repeatedly were negative in our assay for m⁶A. Several other candidate oligonucleotides either were not tested or yielded ambiguous results (data not shown).

**Heterogeneity of methylation sites.** During the course of the localization experiments, we realized that methylation at some of the sites appeared to be heterogeneous. These sites were not necessarily methylated in every molecule of RNA. For position 7424, for example, the corresponding oligonucleotide is ACUUG, which contains only one A. In the assay for m⁶A, however, we detected a signal both at Ap and at the m⁶Ap position on the final thin-layer chromatograph (Fig. 4). This could only occur if some of the RNA molecules were methylated at nt 7424 while others were not. Similar results were obtained for the other methylated nucleotides; however, the degree of heterogeneity varied from site to site. There seemed to be two classes of modified bases, those which were highly methylated (positions 6718, 7414, 7424, and 8014) and those which showed low levels of methylation (nt 4976 or 5044, 6394, 6447, 6507, and 7890 or 7981). We were not able to precisely quantitate the extent of heterogeneity at any site because of the low levels of radioactivity used in these experiments.

TABLE 1. Sites of m⁶A localization[a]

| Nt | Sequence |
|---|---|
| Positive m⁶A sites | |
| 6394 | (GAG)ACUAG |
| 6447 | (GGG)ACUUAUUG |
| 6507 | (UUG)ACUUCUUG |
| 6718 | (AGG)ACUG |
| 7414 | (UGG)ACUG |
| 7424 | (CGG)ACUUG |
| 8014 | (AAG)ACUCUG |
| Ambiguous m⁶A sites | |
| 4976 | (GGG)ACACUG |
| 5044 | (GUG)ACUAAG |
| 7890 | (GGG)ACUCG |
| 7981 | (GGG)ACCUG |

[a] Sequences are the $T_1$ oligonucleotides in which the modified bases reside, plus three nt to the 5′ side of the m⁶A site (in parentheses). Some oligonucleotides were shown to contain m⁶A, but their identities on $T_1$ fingerprints could not be determined unambiguously. The derivation of all oligonucleotides is described in the text.

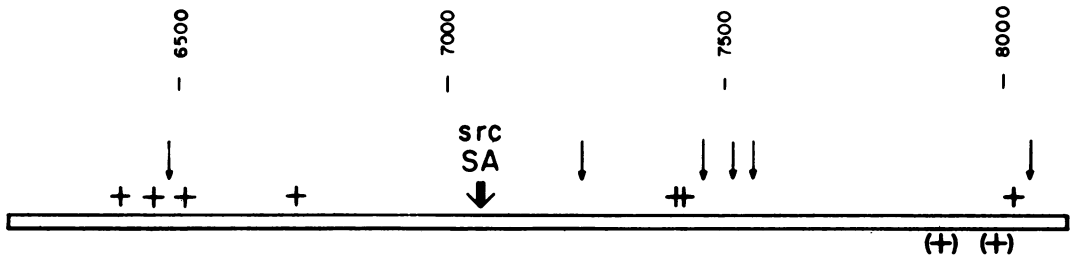


FIG. 5. Schematic representation of methylation sites. The region of the RSV genome spanning nt 6185 to 8150 is depicted. The seven m$^6$A residues detected in this region are designated with + above the line, and site identifications which remain ambiguous are indicated with (+) below the line. The ambiguous site at position 4976 or 5044 is not included in this figure. The exact nt positions of all sites are given in Table 1. Thin arrows refer to points on the genome which are potential splice acceptor sites, based on the consensus sequence PyPyNCAG/Pu (5). The true *src* gene splice acceptor site is indicated by a bold arrow. The sequences at nt 6522 and 6738 (see text) are not strict splice consensus sequences and are not depicted here.

## DISCUSSION

We have developed a method for precisely locating m$^6$A residues within an RNA molecule and have applied it to the analysis of RSV genomic RNA. We observed clustering of m$^6$A sites at nt 6394, 6447, and 6507, at nt 7414 and 7424, and at nt 8014 with either 7890 or 7981. Since our method is quantitative, we were able to determine that methylation is heterogeneous; a given site is not modified in all molecules.

**Methylation sequence.** By identifying m$^6$A residues within RNAs, we hope to come closer to defining a possible role for internal mRNA methylation. The m$^6$A residues which we have located in genomic RNA are all in *src* and *env* gene coding regions. The remaining, unidentified sites are likewise clustered in the 3′ half of the genome (Fig. 1 and reference 4). In other systems there are differing results on the overall distribution of m$^6$A within RNA (1, 7, 17). In addition to the minimal recognition sequences, GAC and AAC, some higher order of specificity is needed to explain this varied positioning of m$^6$A in different RNAs. In mouse L cells, $^{G}_{A}$m$^6$AC is most often preceded by a purine and followed by A, C, or U (31). Likewise, in HeLa cells, the nucleotide on the 3′ side of the methylation sequence (Pum$^6$AC) is A, C, or U (38). In the B77 strain of RSV, a purine comes before $^{G}_{A}$m$^6$AC ca. 85% of the time, and GAC is used most frequently (12).

Our localization results allowed further sequence characterization of methylation sites in PR-C RSV RNA, derived from virus grown in CEF. The oligonucleotides which contain m$^6$A residues are (GAG)ACUAG, (GGG)ACUUAUUG, (UUG)ACUUCUUG, (AGG)ACUG, (UGG)ACUG, (CGG)ACUUG, and (AAG)ACUCUG. Therefore, we can identify a preferred methylation sequence of PuGACU for these GAC-specific sites. In a random sequence, PuGACU would occur once every 512 nt. In fact, in the entire PR-C genome, there are 19 such sequences, a frequency of 1 every 490 nt. In the 1865 bases that we analyzed (nt 6185 to 8050), this sequence was present nine times, more than twice the expected frequency. Of these nine, six have been identified as unambiguous sites of methylation, one is among the possibilities for the ambiguous sites, and two remain uncharacterized for m$^6$A (Table 1, Fig. 5). All of the highly methylated m$^6$A residues (see above) reside in a PuGACU sequence. Neither position 6507 nor the candidates in the ambiguous 4976 or 5044 oligonucleotide are within such a sequence (Table 1), but these are both only minor sites of methylation. The 5,000-nt *gag* and *pol* region of the genome, which is deficient in methylation, has only seven PuGACU sequences (1 every 715 nt).

**Clustering of m$^6$A residues.** It is possible, then, that there is clustering of preferred methylation sequences in a region of RNA which becomes highly modified. A schematic representation of part of the PR-C genome (Fig. 5) indicates that m$^6$A residues are also grouped together. This phenomenon could be important functionally. In addition, it might reveal something about the conformation of the RNA in the cell. Presumably, a methylase will only act upon those sites which are accessible to it. Sequences that are not methylated may remain unmodified because they are buried within the folded RNA molecule. Methylation could be a useful probe, then, for RNA secondary and tertiary structure.

**Heterogeneity of internal methylation.** We have shown that internal base modification is a heterogeneous process; not all RNA molecules have the same methylation pattern. This suggests that the total number of different m$^6$A sites will be larger than the average number of 10 to 15 m$^6$A residues per RSV RNA molecule (4, 14, 36). If methylation does play a role in RNA metabolism, it is possible that there are multiple functional methylation sites. The observed clustering of m$^6$A residues might be a manifestation of just such a phenomenon. That is, methylation at any of the sites within a cluster, or domain, could be sufficient for the proper functioning of that domain.

Heterogeneity might be expected if the methylation pattern in the RSV system were involved in targeting full-length viral RNA for packaging into virions, for translation, or for splicing. Different patterns could signal different functions. These processes themselves might be inefficient, however, thus allowing RNA destined to be translated or spliced to instead be packaged into virions. Similarly, heterogeneity might be a result of incomplete or incorrect methylation of certain molecules, which are still packaged. If methylation is a reflection of secondary and tertiary structure, then the heterogeneity could simply be an indication of multiple RNA conformations within the cell. Finally, the observed heterogeneity could result from the random loss of base modifications.

**Models for m$^6$A function.** Studies with undermethylated RNA have led other investigators to propose a positive role for m$^6$A in regulating mRNA processing events (6, 13, 35). Stoltzfus and Dane (35) saw a buildup of unspliced viral RNA and a decreased level of spliced subgenomic message when they treated RSV-infected CEF with cycloleucine to inhibit internal methylation. They suggested a role for m$^6$A in mRNA splicing. In SV40-infected cells, Finkel and Groner (13) found that cycloleucine treatment led to a reduction in the level of nuclear 19S SV40 mRNA, while the amount of small, fragmented polyadenylated RNA in the nucleus in-

creased. At the same time, overall cytoplasmic levels of SV40-specific and cellular mRNA decreased. From this, they concluded that methylation is somehow involved in preparing mRNA for transport from the nucleus to the cytoplasm. HeLa cells treated with *S*-tubercidinylhomocysteine showed an increase in the time it takes for polyadenylated RNA to appear in the cytoplasm (6). Again, this suggests a role for m$^6$A in processing or transport of mRNA.

Another interpretation of the data, however, is that methylation functions as a negative regulator of mRNA processing (9). It is intriguing to note, in this light, recent data concerning intron branch formation as part of the splicing mechanism. Both mammalian globin RNA and adenovirus RNA have been used as substrates in in vitro splicing experiments, with comparable results (26, 29). Upon cleavage at a 5′ splice site, the 5′ end of the intron forms a 5′→2′ bond with an internal nucleotide in the intron, thus creating a branch at that nucleotide. Similar branched structures have been characterized in the nuclear polyadenylated RNA of HeLa cells (37). In more than 80% of these structures, the base at which branching occurs is adenosine, frequently within either GA*C or AA*C sequences, where * designates the branched residue. In general, the branch point is situated 25 to 40 bases upstream from the 3′ splice site (29). It has been proposed that m$^6$A may be a positive signal for branch formation (39). However, the preferred sequence that we have derived for RSV RNA methylation (PuGACU) is similar, though not identical, to the consensus sequence derived for branch sites (PyNPyPyPuAPy) (19, 29, 39), suggesting that not all branches contain m$^6$A. This and the detection of m$^6$A residues in exons argue against a positive role for m$^6$A in branch site selection. It is possible, on the other hand, that methylation at branch points can occur under certain circumstances or that sequences which closely resemble branch sequences are methylated as a means of preventing their utilization. This could represent one means of regulating splice site selection.

Interestingly, most of the methylation sites which we have identified are upstream of sequences which fit into a consensus splice acceptor sequence, PyPyNCAG/Pu (5). Figure 5 is a schematic representation of the relationship between the m$^6$A residues (including ambiguous sites) and such putative splice sites in the region of the genome that we analyzed. Specifically, downstream from position 7414 or 7424 is the sequence CCUGCAG/A ("splice site" at nt 7457), and position 8014 precedes the sequence CCUGCAG/G (nt 8057). The m$^6$As at positions 6394 and 6447 are followed by the sequence UCCUGCAG/A at nt 6491. The closest fits to the consensus sequence after m$^6$As 6507 and 6718 are CUUCUAG/C at position 6522 and CUUGUAG/U at nt 6738, respectively. For comparison, the sequences surrounding the two functional splice acceptor sites in full-length RSV RNA are UUUGCAG/G (nt 5078), preceding the *env* gene, and GCUGCAG/G (nt 7054), in front of the *src* gene (32).

Even though genomic RNA is an unspliced molecule, no methylation has been detected in the RNA regions preceding the true *src* acceptor site. Preceding the *env* acceptor site at nt 5078 is the oligonucleotide ACUAAG (A = 5044), which remains at least a formal candidate for containing m$^6$A (see above; Table 1). However, even if this residue is methylated, it is one of the minor sites of modification described above. Therefore, perhaps methylation acts as a means of protecting against splicing at sites that are never to be utilized, while some other mechanism is used to regulate splicing at functional splice sites. It will be interesting to determine whether intracellular RSV RNAs, some of which are precursors to the spliced *env* and *src* mRNAs, are methylated near their splice acceptors.

A splice protection model of this sort is consistent with the overwhelming presence of m$^6$A residues in coding regions and with the existence of unmethylated, yet spliced, messages. It is important to note, however, that we are not suggesting just a single mechanism for splicing or that methylation serves as the only means of regulating that activity. Indeed, methylation might not be involved in splicing at all. It may be associated with RNA stability, secondary structure, transport, or packaging in the case of RNA viruses. Possibly, there may be no function for m$^6$A.

In the RSV system, we are able to analyze a single RNA species which serves multiple functions. Full-length RSV RNA can be packaged into virions as genomic RNA, it can remain unspliced and serve as mRNA for the *gag* gene, or it can be spliced to subgenomic *env* or *src* mRNAs (16). No major differences have yet been detected between the full-length molecules which serve each of these functions, but the possibility of more subtle differences, such as methylation, remains to be investigated.

## ACKNOWLEDGMENTS

We thank R. Guntaka for pATV8 and J. M. Bishop for plasmid SRA-2, R. C. Huang, R. Christy, and S. Arrigo for assistance with M13 cloning, and B. Mattingly and A. Doering for technical assistance.

This work was supported by National Institutes of Health (NIH) grant CA33199 from the National Cancer Institute, Basil O'Connor Starter Research Grant 5-399 from the March of Dimes, and by an NIH Biomedical Research Support Grant. S.E.K. was supported by an NIH Predoctoral Training Grant. K.B. is the recipient of a Faculty Research Award from the American Cancer Society.

## LITERATURE CITED

1. **Aloni, Y., R. Dhar, and G. Khoury.** 1979. Methylation of nuclear simian virus 40 RNAs. J. Virol. **32**:52–60.
2. **Banerjee, A. K.** 1980. 5′-Terminal cap structure in eucaryotic messenger ribonucleic acids. Microbiol. Rev. **44**:175–205.
3. **Barrell, B. G.** 1971. Fractionation and sequence analysis of radioactive nucleotides, p. 751–795. *In* G. L. Cantoni and D. R. Davies (ed.), Procedures in nucleic acid research, vol. 2. Harper & Row, Publishers, New York.
4. **Beemon, K., and J. Keith.** 1977. Localization of N$^6$-methyladenosine in the Rous sarcoma virus genome. J. Mol. Biol. **113**:165–179.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
6. **Camper, S. A., R. J. Albers, J. K. Coward, and F. M. Rottman.** 1984. Effect of undermethylation on mRNA cytoplasmic appearance and half-life. Mol. Cell. Biol. **4**:538–543.
7. **Canaani, D., C. Kahana, S. Lavi, and Y. Groner.** 1979. Identification and mapping of N$^6$-methyladenosine containing sequences in simian virus 40 RNA. Nucleic Acids Res. **6**:2879–2899.
8. **Casey, J., and N. Davidson.** 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res. **4**:1539–1552.
9. **Chen-Kiang, S., J. R. Nevins, and J. E. Darnell.** 1979. N$^6$-methyladenosine in adenovirus type 2 nuclear RNA is conserved in the formation of messenger RNA. J. Mol. Biol. **135**:733–752.
10. **Darnell, J. E.** 1979. Transcription units for mRNA production in eucaryotic cells and their DNA viruses. Prog. Nucleic Acid Res. Mol. Biol. **22**:327–353.
11. **DeLorbe, W. J., P. A. Luciw, H. M. Goodman, H. E. Varmus,**

**and J. M. Bishop.** 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. J. Virol. **36**:50–61.
12. **Dimock, K., and C. M. Stoltzfus.** 1977. Sequence specificity of internal methylation in B77 avian sarcoma virus RNA subunits. Biochemistry **16**:471–478.
13. **Finkel, D., and Y. Groner.** 1983. Methylations of adenosine residues ($m^6A$) in pre-mRNA are important for formation of late simian virus 40 mRNAs. Virology **131**:409–425.
14. **Furuichi, Y., A. J. Shatkin, E. Stavnezer, and J. M. Bishop.** 1975. Blocked, methylated 5'-terminal sequence in avian sarcoma virus RNA. Nature (London) **257**:618–620.
15. **Gillespie, D., and S. Spiegelman.** 1965. A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. J. Mol. Biol. **12**:829–842.
16. **Hayward, W. S., and B. G. Neel.** 1981. Retroviral gene expression. Curr. Top. Microbiol. Immunol. **91**:217–276.
17. **Horowitz, S., A. Horowitz, T. W. Nilsen, T. W. Munns, and F. R. Rottman.** 1984. Mapping of $N^6$-methyladenosine residues in bovine prolactin mRNA. Proc. Natl. Acad. Sci. U.S.A. **81**:5667–5671.
18. **Katz, R. A., C. A. Omer, J. H. Weis, S. A. Mitsialis, A. J. Faras, and R. V. Guntaka.** 1982. Restriction endonuclease and nucleotide sequence analyses of molecularly cloned unintegrated avian tumor virus DNA: structure of large terminal repeats in circle junctions. J. Virol. **42**:346–351.
19. **Keller, E. A., and W. A. Noon.** 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc. Natl. Acad. Sci. U.S.A. **81**:7417–7420.
20. **Laskey, R. A.** 1980. The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis. Methods Enzymol. **65**:363–371.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Messing, J., B. Gronenborn, B. Muller-Hill, and P.-H. Hofschneider.** 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a Hind II fragment of the *lac* regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **74**:3642–3646.
23. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
24. **Moss, B., A. Gershowitz, L. A. Weber, and C. Baglioni.** 1977. Histone mRNAs contain blocked and methylated 5' terminal sequences but lack methylated nucleosides at internal positions. Cell **10**:113–120.
25. **Norrander, J., T. Kemp, and J. Messing.** 1983. Construction of improved M13 vectors using dideoxynucleotide-directed mutagenesis. Gene **26**:101–106.
26. **Padgett, R. A., M. A. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp.** 1984. Lariat RNAs as intermediates and products in the splicing of mRNA precursors. Science **225**:898–903.
27. **Perry, R. P., and K. Scherrer.** 1975. Methylated constituents of globin mRNA. FEBS Lett. **57**:73–78.
28. **Revel, M., and Y. Groner.** 1978. Post-transcriptional and translational controls of gene expression in eukaryotes. Annu. Rev. Biochem. **47**:1079–1126.
29. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell **38**:317–331.
30. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in a single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
31. **Schibler, U., D. E. Kelley, and R. P. Perry.** 1977. Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. J. Mol. Biol. **115**:695–714.
32. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
33. **Sommer, S., M. Salditt-Georgieff, S. Bachenheimer, J. E. Darnell, Y. Furuichi, M. Morgan, and A. J. Shatkin.** 1976. The methylation of adenovirus-specific nuclear and cytoplasmic RNA. Nucleic Acids Res. **3**:749–765.
34. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
35. **Stoltzfus, C. M., and R. W. Dane.** 1982. Accumulation of spliced avian retrovirus mRNA is inhibited in S-adenosylmethionine-depleted chicken embryo fibroblasts. J. Virol. **42**:918–931.
36. **Stoltzfus, C. M., and K. Dimock.** 1976. Evidence for methylation of B77 avian sarcoma virus genome RNA subunits. J. Virol. **18**:586–595.
37. **Wallace, J. C., and M. Edmonds.** 1983. Polyadenylated nuclear RNA contains branches. Proc. Natl. Acad. Sci. U.S.A. **80**:950–954.
38. **Wei, C.-M., and B. Moss.** 1977. Nucleotide sequence at the $N^6$-methyladenosine sites of HeLa cell messenger ribonucleic acid. Biochemistry **16**:1672–1676.
39. **Zeitlin, S., and A. Efstratiadis.** 1984. In vivo splicing products of the rabbit β-globin pre-mRNA. Cell **39**:589–602.