Mass Spectrometry of mRNA Cap 4 from Trypanosomatids Reveals Two Novel Nucleosides*

(Received for publication, December 26, 1991)

James D. Bangs‡‡, Pamela F. Crain‡, Takeshi Hashizume‡, James A. McCloskey‡‖, and John C. Boothroyd‡‡**

From the ‡Department of Microbiology and Immunology, Stanford School of Medicine, Stanford, California 94305 and the Departments of ††Medicinal Chemistry and ‡‡Biochemistry, University of Utah, Salt Lake City, Utah 84112

Synthesis of mRNA in kinetoplastid protozoa involves the process of trans-splicing, in which an identical 39-41-nucleotide (depending on the species) mini-exon is placed at the 5' end of mature mRNAs. The mini-exon sequence is highly conserved among all members of the Kinetoplastida, nucleotides 1–6 being identical in the four genera so far examined. Prior to trans-splicing, the mini-exon donor RNA is capped by the addition of a (5'-5') triphosphate-linked 7-methylguanosine, followed by modification of the first four transcribed nucleotides. Partial structures have been previously deduced for this cap 4 moiety from *Trypanosoma brucei* and *Leptomonas collosoma*. We have purified enough cap 4 from *T. brucei* and *Crithidia fasciculata* to allow definitive structural analysis by combined liquid chromatography/mass spectrometry and gas chromatography/mass spectrometry. The results, together with the known mini-exon sequence, show that cap 4 in both species has the structure m3G(5')ppp(5')m7AmpAmpCmpmsUmp. The presence of N6,N2'-O-trimethyladenosine and 3,2'-0-dimethyluridine, nucleosides previously unknown in nature, were confirmed by rigorous comparison with synthetic standards. The conservation of cap 4 between these divergent genera suggests that this structure may be common to most if not all Kinetoplastida.

The 5' end of mature mRNA in protozoa of the order Kinetoplastida is formed by the process of trans-splicing (reviewed in Ref. 1). In this process, which is mechanistically similar to cis-splicing, a sequence called the mini-exon (or spliced leader) is transferred from the 5' end of a mini-exon donor RNA (medRNA) to pre-mRNA acceptors. Although the size of the mini-exon varies between species (39–41 nucleotides), as does the medRNA (80–140 nucleotides), the mini-exon sequence is invariant within a species and is highly conserved among trypanosomatids, in general. In particular, the first six nucleotides are identical in the four kinetoplastid genera examined so far.

The extreme 5' end of the mini-exon sequence in *Trypanosoma brucei* bears an unusual mRNA cap structure (2–5). As in other eukaryotes, 7-methylguanosine is linked via a 5'-5' triphosphate bridge to nucleotide 1 of the mini-exon, and the first transcribed nucleotides are modified. What is unusual is the extent of this modification; whereas no more than two modified nucleotides have been described in any other eukaryotic cap structure (6), the trypanosome cap has four consecutive modified nucleotides (and thus by convention is referred to as a cap 4 structure). This is the most highly modified eukaryotic mRNA cap known. Based on the known mini-exon sequence (7) and analyses of cap 4 nucleotides derived from uniformly radiolabeled poly(A') RNA, a partial structure (see Fig. 1) has been deduced (2–5). Nucleosides 2 and 3 are 2'-O-methyladenosine and 2'-O-methylcytidine, respectively. The data also suggested that nucleoside 1, an adenosine derivative (A*), and nucleoside 4, a uridine derivative (U*), are not only 2'-O-modified (presumably by methylation), but also possess uncharacterized base modifications. The medRNA bears an identical structure (2–4, 8) indicating that synthesis occurs on the medRNA and that cap 4 is transferred to mature mRNA as a consequence of the trans-splicing process. A similar, if not identical, cap is found on mRNA from *Leptomonas collosoma* (2) suggesting that this phenomenon may be a general property of all trypanosomatids.

Definitive functions have not been assigned to the trypanosome cap, but it is apparently involved in mRNA processing since inhibition of methylation has been shown to decrease the efficiency of trans-splicing (9). mRNA caps are known to participate in the initiation of translation in other eukaryotes (10), and it is reasonable to suppose that cap 4 plays such a role in trypanosomes. In addition, cap 4 may also enhance mRNA stability by conferring nuclease resistance, as caps do in other systems (11).

If meaningful questions concerning cap 4 function are to be asked, it is essential to have complete knowledge of its structure. With this need in mind, we have specifically placed a radiolabel in the triphosphate bridge of poly(A') RNA from *T. brucei* and a related trypanosomatid, *Crithidia fasciculata*. Labeled caps from these RNAs have been generated by mixed RNase digestion and have been compared at the analytical level. These radiolabeled caps have then been used as tracers for the large scale purification of caps for subsequent analysis by both combined liquid chromatography/mass spectrometry.
FIG. 1. Cap 4 structure of *T. brucei*. The 3' nucleotides of the mini-exon are numbered. The 3' RNase T1 cleavage site that generates free cap 4 (arrow) and the position of the radiolabel introduced by the recapping procedure (triangle) are indicated. m7G, 7-methyl; m, 2'-O-methyl (known for nucleotides 2 and 3, presumed for nucleotides 1 and 4); *, previously unknown base modifications of nucleotides 1 and 4.

[Diagram of cap 4 structure with labeled nucleotides]

FIG. 2. Specific 5' labeling of poly(A)+ mRNA by recapping. Chemically decapped *T. brucei* poly(A)+ RNA was enzymatically recapped with guanylyltransferase and [α-32P]GTP. Without mixing with unlabeled RNA, samples (5 µg) of radiolabeled RNA were subjected to various treatments. Lane 1, mock digestion; Lane 2, mixed RNAs; Lane 3, RNase H and antisense mini-exon oligonucleotide; Lane 4, RNase H alone; Lane 5, tobacco acid pyrophosphatase. Treated RNA samples were fractionated on a 5% polyacrylamide, 7 M urea gel. A photograph of the ethidium bromide-stained gel (Panel A) and an autoradiograph (Panel B) are shown. The mobilities of the four small ribosomal RNAs (Ref. 12) and of 5 S RNA are indicated.

[Image of gel with labeled bands and autoradiograph]

and gas chromatography/mass spectrometry. The results allow a definitive and identical structure to be deduced for fully modified cap 4 from each species and reveal the presence of two nucleosides never before detected in nature.

MATERIALS AND METHODS

RESULTS

Recapping of RNA—We wished to place a specific radiolabel in the cap structure of poly(A)+ RNA from *T. brucei* and *C. fasciculata* to facilitate comparative analyses and for use as a marker for the large scale purification of cap 4 from these species. Poly(A)+ RNA was decapped chemically generating a 5'-triphosphate terminus that is a substrate for subsequent enzymatic recapping with guanylyltransferase and [α-32P]GTP (11). This procedure replaces the 7-methylguanosine cap moiety with guanosine and introduces [32P]phosphate into the triphosphate bridge (see Fig. 1).

Recapped *T. brucei* poly(A)+ RNA was tested by various enzymatic digestions, followed by electrophoresis and autoradiography, to determine if the radiolabel was incorporated as expected (Fig. 2). After recapping, radiolabel was detected predominantly in high molecular weight material as would be expected for labeled poly(A)+ RNA (Panel B, Lane 1). Treatment with mixed RNases resulted in the complete elimination of all ethidium staining (Panel A, compare Lanes 1 and 2) and the release of all radiolabel (Panel B, Lane 2). Treatment of the labeled RNA with RNase H, either with or without antisense mini-exon oligonucleotide, had little effect on bulk RNA as judged by ethidium staining (Panel A, Lanes 3 and 4). However, RNase H treatment released essentially all the incorporated radiolabel in an oligonucleotide-specific manner (Panel B, compare Lanes 3 and 4). The remaining RNase H-resistant bands (Panel B, Lane 3) may be due to incomplete digestion or may represent contaminating RNAs (snRNA, 5 S RNA, etc.) that were not eliminated by the poly(A)+ selection procedure and which would be substrates for the subsequent recapping reaction. Treatment with tobacco acid pyrophosphatase also released all the radiolabel with minimal effect on the integrity of the RNA (Panels A and B, Lane 5). Analysis of the products of tobacco acid pyrophosphatase treatment by PEI cellulose TLC indicated that, as expected, the label was released in the form of GMP. In addition, treatment of decapped RNA with alkaline phosphatase prior to recapping abrogated the subsequent incorporation of radiolabel into high molecular weight RNA. These data unequivocally indicate that the recapping procedure is primarily labeling the 5' end of mini-exon-bearing poly(A)+ RNA, and that the label is incorporated in the triphosphate bridge of the cap structure.

Identification of Free CAP 4—Mixed RNase digestion of poly(A)+ RNA should generate a mixture of products containing nucleoside monophosphates from the bulk of the RNA, nucleosides from the 3' end of each RNA molecule, and a six-nucleotide RNase-resistant species corresponding to the cap 4 structure plus a 3' adenosine residue that is the site of cleavage by RNase T1 (see Fig. 1). In order to identify the latter component, the products of mixed RNase digestion of recapped radiolabeled RNA from both *T. brucei* (Panel A) and *C. fasciculata* (Panel B) were fractionated by anion exchange chromatography (Fig. 3). In each case, a bimodal profile of eluted radioactivity was detected. PEI cellulose TLC analysis revealed that the first DEAE-peak (peak I), eluting at 0.2 M salt, contained [32P]guanylyl nucleotides in the order

[Diagram of DEAE chromatography]

3 Portions of this paper (including "Materials and Methods," part of "Results," and Figs. 7, 8, 10, 11, 12, and 13) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

3 J. D. Bangs, unpublished observations.
GTP > GDP > GMP. This peak represents residual radiolabel that is not removed by desalting (see "Materials and Methods"). The radiolabeled material in the second DEAE-peak (peak II), eluting at 0.4 M salt, was also analyzed by PEI cellulose TLC, before and after treatment with tobacco acid pyrophosphatase (Fig. 4). Untreated peak II, from both C. fasciculata (Lane 3) and T. brucei (Lane 5), migrated as a smear between the origin and the position of GTP, consistent with its behavior on DEAE. Tobacco acid pyrophosphatase treatment converted the radiolabel in each of these samples to a form that co-migrated with GMP (Lanes 4 and 6).

The higher charge and the pyrophosphatase sensitivity of the radiolabeled material in peak II makes it a good candidate for the free cap structure. To confirm that peak II contains cap 4, an identical species was generated from recapped T. brucei poly(A) RNA by an alternative, diagnostic procedure based on anti-mini-exon-directed RNase H cleavage. The radiolabeled RNase H products were purified by gel filtration chromatography2 and analyzed by electrophoresis in parallel with peak II material generated by standard mixed RNase digestion (Fig. 5, Panel A). Peak II migrates as a single species near the bromphenol blue marker (Lane 1), and the RNase H products migrate as a set of discrete higher molecular weight species (Lane 2). This pattern of RNase H products was expected and is consistent with staggered cleavage in the mini-exon at multiple sites downstream of the cap 4 structure. Further treatment with mixed RNases converted the purified RNase H products to a single species that has the same charge, as judged by DEAE-chromatography, and electrophoretic mobility as peak II (Fig. 5, Panel B), indicating that the latter does, in fact, contain cap 4 derived from the 5' end of mini-exon-bearing RNA.

Peak II material from C. fasciculata was further analyzed by high resolution anion exchange chromatography on a Mono Q column (Fig. 6). A single major radioactive peak, eluting between polyadenylate standards of n = 8 and n = 9 (charges of -9 and -10, respectively), was detected. Essentially identical results were obtained when T. brucei peak II was analyzed in a similar manner.3 This is the expected elution position for a RNase-resistant cap 4 structure and indicates that the C. fasciculata cap, which also has the same electrophoretic mobility as the T. brucei cap (see Fig. 8), is RNase-resistant, and therefore 2'-O-modified, to the same extent as that of trypanosomes. These data confirm that mixed RNase digestion is generating a discrete cap 4 species that is apparently of homogeneous size.

---

**Fig. 4.** PEI cellulose chromatography of pyrophosphatase-treated peak II. [α-32P]GTP (Lanes 1 and 2) and samples of Peak II material from C. fasciculata (Lanes 3 and 4) and T. brucei (Lanes 5 and 6) were mock-treated (Lanes 1, 3, and 5) or digested with tobacco acid pyrophosphatase (Lanes 2, 4, and 6) and then analyzed by PEI cellulose chromatography. An autoradiograph is shown, and the mobilities of internal GTP, GDP, and GMP markers are indicated.

---

**Fig. 5.** Generation of T. brucei Peak II by an alternate protocol. Radiolabeled Peak II was generated by the standard procedure of mixed RNase digestion followed by DEAE-chromatography. Radiolabeled recapped poly(A) RNA was also treated with RNase H and antisense mini-exon oligonucleotide, and the released 5' ends were purified by Sephadex G-50 gel chromatography. The purified RNase H products were then subjected to mixed RNase digestion. Samples of digestion products were compared with peak II by fractionation on a 25% polyacrylamide sequencing gel and an autoradiograph is presented. Panels A and B, Lane 1, Peak II; Panel A, Lane 2, RNase H products; Panel B, Lane 2, mixed RNase-treated RNase H products. The mobilities of the RNase H products, cap 4, xylene cyanol (XC), bromphenol (BPB) blue, and GTP are indicated.

**Fig. 6.** Mono Q chromatography of cap 4. Peak II material prepared from radiolabeled recapped poly(A) RNA from C. fasciculata was fractionated by Mono Q chromatography. Fractions were collected and the radioactivity in each was determined by liquid scintillation spectrometry. The elution positions of internal [3H] polyadenylate standards (n = 1 to 11) are indicated by vertical bars.

**Purification of CAP 4—**A scheme was devised for the large scale purification of cap 4 from poly(A) RNA for subsequent analysis by mass spectrometry. Recapped radiolabeled poly(A) RNA was mixed, as a tracer, with unlabeled poly(A) RNA from the same source, either C. fasciculata (10 mg) or T. brucei (2.8 mg) and, following mixed RNase digestion, peak II fractions were prepared by DEAE-chromatography. This procedure was chosen for the first step since it was easily scaled up, it provided a rapid purification of cap 4 from other RNase digestion products, and typically had recoveries of radioactivity, and therefore cap 4, in excess of 90%.

Purified peak II was further fractionated by C4 reversed phase HPLC. This method, which is based on a different
mRNA Cap Structure of Trypanosomatids

molecular parameter than charge (hydrophobicity), gives >90% recovery of loaded peak II radioactivity. The C4 HPLC chromatograms from the large scale preparations of C. fasciculata and T. brucei cap 4 are shown in Fig. 7. In each case, a single peak of radioactivity eluting at about 35 min was detected (Panels C and D, respectively) along with a corresponding peak of UV absorbing material (Panels A and B, respectively), indicating that a significant amount of cap 4 material had been purified. The UV profile for T. brucei cap 4 is, in fact, a tightly spaced doublet; the significance of this finding is not clear. Electrophoretic analysis (Fig. 8) of equivalent samples from each step of the C. fasciculata purification demonstrated that: 1) all radiocontaminants were eliminated from purified cap 4 (Lane 4); 2) mixed RNase digestion generates a free cap species of homogeneous size, consistent with the results of Mono Q analysis; and 3) the cap 4 from T. brucei and the cap 4 from C. fasciculata are the same size (compare cap 4 mobility relative to bromphenol blue mobility in Figs. 5 and 8), again confirming the Mono Q results. By these criteria, therefore, cap 4 prepared in this manner was deemed pure and suitable for mass spectrometric analysis.

Analysis of C. fasciculata mRNA Cap Digests by Directly Combined HPLC/Mass Spectrometry—For further analysis, cap 4 was converted to its component nucleosides by enzymatic hydrolysis. Digests of purified bulk Crithidia mRNA cap 4 produced by treatment with nuclease P1 and alkaline phosphatase, with and without nucleotide pyrophosphatase (total and pyrophosphatase-minus digests, respectively), were analyzed by thermospray LC/MS. Identification of components of the digest was based on HPLC retention times, compared with standardized values determined using the same gradient elution system (13), and mass spectra of each eluant, generally consisting of ions representing the protonated molecule (MH+) and protonated base fragment ions (BH+ in which BH corresponds to the neutral free base). Fig. 9 shows the HPLC chromatograms from analysis of Crithidia digests and synthetic nucleosides, and the corresponding mass values for each component are listed in Table I. Procedures for interpretation of data and comparison with standard mass and retention time data are given in Ref. 13.

As indicated in Fig. 9A and Table I, the total digest was found to contain Cm, G, m7G, A, Am, and two uncharacterized nucleosides, U* and A*. The amount of C. fasciculata RNA analyzed was estimated as 170 ng, equivalent to 20–25 ng/nucleotide component. The retention times and thermospray mass spectra of constituents U* and A* do not correspond to any known RNA nucleoside, based on sequence (14) and composition compilations (15) and on HPLC (13, 16, 17) and mass spectral (13) reference data. When Crithidia RNA cap 4 is digested to nucleosides without nucleotide pyrophosphatase treatment, the predicted products are a "core cap" consisting of m7G and nucleoside 1 joined by the triphosphate bridge, along with nucleosides 2–5 (see Fig. 1). The chromatogram of such a digestion is shown in Fig. 9B. Peaks corresponding to m7G and A* are absent (Fig. 9A, peaks at 16.0 and 33.4 min, respectively), and both G and Am (15.1 and 23.2 min, respectively) are reduced in abundance. Thermospray LC/MS analysis of the pyrophosphatase-minus digest confirmed the presence of Cm, A, Am, and U* in the cap 4 structure and indicates the position of these nucleosides to be downstream from the core cap. The amount of C. fasciculata cap RNA analyzed in this digest was estimated as 85 ng, equivalent to 10–12 ng per component.

A new (broad) peak (the core cap) is observed in the chromatogram of the pyrophosphatase-minus digest, centered at about 29.4 min (Fig. 9B). Nucleotide molecular ion species

![Figure 9](image_url)

**Fig. 9.** C18 reversed phase HPLC of C. fasciculata RNA cap 4 constituents and of synthetic ribonucleoside standards. Detection by UV absorbance at 254 nm. Chromatograms of: products of hydrolysis by nuclease P1, nucleotide pyrophosphatase, and alkaline phosphatase (total digest) (A), products of hydrolysis by nuclease P1 and alkaline phosphatase (pyrophosphatase-minus digest) (B), and nucleosides 3',5'-O-dimethyluridine (m7Um) and N6,N2',2'-O-trimethyladenosine (m7Am) (C).

| Nucleoside               | Retention time | Ion<sup>a</sup> | MH<sup>b</sup> | m/z | BH<sup>i</sup> |
|-------------------------|----------------|-----------------|--------------|-----|--------------|
| 2'-O-Methylcytidine (Cm)| 13.3           | 258             | 112          |     |              |
| Guanosine (G)           | 15.7           |                | 136          |     |              |
| 7-Methylguanosine (m7G) | 16.0           |                | 166          |     |              |
| Adenosine (A)           | 20.0           | 268             | 136          |     |              |
| 2'-O-Methyladenosine (Am)| 23.2          | 282             | 136          |     |              |
| U*                     | 24.2           | 273             | 127          |     |              |
| A*                     | 33.4           | 310             | 164          |     |              |

<sup>a</sup> From the chromatogram shown in Fig. 9A.
<sup>b</sup> From thermospray mass spectra.
<sup>i</sup> MH<sup>+</sup> ions characteristically not observed for G and m7G (13).
cap peak (Fig. 10). The diminution of the Am signal in the pyrophosphatase-minus digest, relative to the total digest, and the presence of Am in the core cap may be related phenomena, as addressed below (see "Discussion"). The presence of G in the total digest and its virtual disappearance in pyrophosphatase-minus digest results, an order of the component nucleosides can be deduced (see Fig. 1). This is the same order as previously deduced for T. brucei and L. collosoma. The unmodified A must be 3'-most as it is the only substrate for ribonuclease cleavage. A* must be at position 1 in the core cap as its release is pyrophosphatase-dependent, as is the release of the other core cap constituent, m7G. The order of the remaining nucleotides, Am, Cm, and U*, is clear from the gene sequence (AACCUG...). Thus, these data demonstrate that the overall structure shown in Fig. 1 applies also for C. fasciculata.

The thermospray mass spectrum of U* (Fig. 11A) indicates a nucleoside of molecular weight 272 (MH+, 273) and free base of mass 126 (BH2, 127). The mass difference of 146 u shows the sugar moiety to be methylribose rather than unsubstituted ribose, which would require a 132-u difference (13, 21). The mass of the base requires the presence of an even number of nitrogen atoms and corresponds to uracil substituted by one methyl group. Based on known forms of post-transcriptional modification in RNA, these data suggest either 3',5'-dimethyluridine (m5Um) or 5',2'-O-dimethyluridine (m3Um). The latter structure is excluded, however, because the observed retention time of U* (24.3 min) differs significantly from that of m5Um, 20.6 min, recorded using the same solvent system.

The thermospray mass spectrum of component A* (Fig. 11B) reveals a molecular weight of 309 (MH+, 310), requiring an odd number of nitrogen atoms. The base mass of 163 (BH+, 164) requires the nucleoside to be an N3-unmethylated nucleoside (from the MH+ - BH2 difference, 146 u) with a base corresponding in mass to dimethyladenine. In addition, the observed 254 nm/280 nm UV absorbance ratio observed for A*, 0.45, is very similar to that of N6,N6-dimethyladenosine, 0.43, determined in the same HPLC solvent system. As a result, a tentative structure assignment of N6,N6,2'-O-trimethyladenosine (m5Am) was made for nucleoside A*. Based on the above conclusions, authentic m5Um and m5Am were prepared by chemical synthesis, for comparison with cap nucleosides U* and A* by LC/MS and GC/MS.

Comparison of Synthetic 3',2'-O-Dimethyluridine and N6,N6,2'-O-Trimethyladenosine with Cap 4 Constituents U* and A* by GC/MS—Rigorous comparison of authentic m5Um and m5Am with cap constituents U* and A* was made by capillary column gas chromatography-electron ionization mass spectrometry of their volatile trimethylsilyl derivatives based on three parameters: 1) precise comparison of gas chromatographic retention times from sequential experiments 2) interpretation of the EI mass spectra of U* and A* in terms of known mass-structure relationships for silylated ribonucleosides (22), and 3) comparison of the EI mass spectra of synthetic and cap nucleosides, recorded under identical experimental conditions, using similar quantities of material (estimated as approximately 0.4 ng of each nucleoside).

Gas chromatographic retention time comparisons were made from mass spectral ion current profiles using the characteristic (M - CH3+) ions (22), which are unique with respect to the proposed Crithidina cap 4 constituents: m/z 401 for m5Um-(TMS)2 and m/z 438 for m5Am-(TMS)2 (structures shown in Fig. 14). The results, shown in Figs. 12 and 13, show the experimentally measured retention times to differ by less than 0.5 s in each case, which is within the 1.5-s instrument cycle time for acquisition of each mass spectrum: 13:37 min for m5Um and U* and 16:10 min for m5Am and A*.

Mass spectra recorded at elution times 13:37 and 16:10 min for trimethylsilylated Crithidina mRNA cap 4 total digest are shown in Fig. 14, A and C, respectively, and are compared with corresponding mass spectra of synthetic nucleoside derivatives, Fig. 14, B and D, respectively. Structure assignments for the principal peaks from each spectrum are listed in Table II and are based on earlier detailed studies of the mass spectra of nucleoside models and their stable isotope-labeled analogs (22). Leading references to additional studies and details of spectral interpretation are given in Ref. 23. In Fig. 14, A and C, the molecular mass values of 416 and 453 are as expected from the addition of two TMS groups (net increase 144 Da) to the molecular masses of U* and A*, respectively, initially measured by LC/MS. In both cases, the fragment ion products reflect ribose methylation and bases of

Fig. 14. Electron ionization mass spectra from GC/MS of C. fasciculata RNA cap 4 constituents and synthetic ribonucleoside standards, following trimethylsilylation. A, Cap nucleoside U*, recorded at 11:37 min in Fig. 12A. B, 3',2'-O-Dimethyluridine (TMS)2, recorded at 11:37 min in Fig. 12B. C, Cap nucleoside A*, recorded at 16:10 min in Fig. 13A. D, N6,N6,2'-O-Trimethyladenosine (TMS)2, recorded at 16:10 min in Fig. 13B.

* P. F. Crain and S. C. Pomerantz, unpublished observations.
by the closely related models. Transfer to the ionized base, is a clear indicator of ribose between an enzymatic digest of Crithidia mRNA cap resulting purine mass spectra (24) and is very similar to that exhibited trimethyladenosine.

N6,N6,2'-O-Trimethyladenosine (TMS)3, the same mass as determined from the initial LC/MS experiment. The characteristic fragment ion m/z 290 (ion S–H), interpreted in analogy to uridine in which H-2' is lost by transfer to the ionized base, is a clear indicator of ribose methylation at O-2' (22) in the mass spectra of both U* and A* derivatives. Ribose methylation is also reflected by ions in which, for example, the sugar fragment m/z 159 is the most abundant ion abundances which would be predicted from spectra of other isomers, leads independently to the conclusion that U* is 3',2'-O-dimethyluridine and A* is N6,N6,2'-O-trimethyladenosine.

Comparison of Synthetic 3',2'-O-Dimethyluridine and N6,N6,2'-O-Trimethyladenosine with Cap 4 Constituents U* and A* by HPLC—Comparison of HPLC retention times between an enzymatic digest of Crithidia mRNA cap resulting from treatment by nuclease P1, alkaline phosphatase, and nucleotide pyrophosphatase, and a mixture of synthetic m3Um and m3Am, is shown in Fig. 9, A and C, respectively. The retention times for each pair are experimentally indistinguishable: 23.4 min for U* and m3Um and 33.4 min for A* and m3Am. In addition, comparison of 254 nm/280 nm absorbance ratios from HPLC profiles show similar correspondence for cap and synthetic nucleosides: U*, 2.78; m3Um, 2.86; A*, 0.445; m3Am, 0.430.

**TABLE II**

| m/z  | Assignment |
|------|------------|
| 416  | M          |
| 401  | M – CH3    |
| 421  | M – CH3OH  |
| 332  | M – (CH3OH – TSMS) |
| 290  | S – H      |
| 271  | 199 – H + TMS |
| 259  | S – CH3OH  |
| 255  | B + TMS + CHCHOCH2-2' |
| 227  | B + TMS + 1'-CHO |
| 220  | B + CH3Si; BH + CHCHOCH2-2' |
| 217  | TMSO(CH3)OTMS |
| 201  | S – TMSOH  |
| 199  | BH + TMS; BH + CH(O)CHOCH2-2' |
| 192  | BH + 1'-CHO |
| 187  | 290 – 5'-CH2OTMS |
| 172  | TMSO(CH3)OCH2-2' |
| 169  | S – (CH3OH + TMSOH) |
| 164  | B + 2H     |
| 163  | B + H      |
| 159  | 2'-CH2O(CH3)OTMS |
| 147  | TMSO(Si(CH3)3) |
| 134  | BH – CH3N  |
| 103  | 5'-CH3OTMS |

* M, molecular ion; B, base fragment; S, sugar fragment; TMS, trimethylsilyl. Positive charge is not indicated.

**TABLE III**

| Nucleoside* | HPLC retention time, min | Mass spectra/m/z/Ion |
|-------------|--------------------------|---------------------|
| 2'-O-Methylcytidine (Cm) | 12.9/112/BH+ | |
| 7-Methylguanosine (mG) | 16.3/166/BH+ | |
| Adenosine (A) | 20.3/136/BH+ | |
| 2'-O-Methyladenosine (Am) | 23.5/136/BH+ | |
| U* | 24.6/127/BH+ | |
| A* | 33.4/310/BH+ | |

* See Fig. 15A.

Low intensity signal.

Analysis of T. brucei mRNA Cap Digest by Directly Combined HPLC/Mass Spectrometry—LC/MS analysis of a T. brucei cap RNA total digest is shown in Fig. 15 and confirms the presence of the same nucleosides found in the C. fasciculata cap RNA, as summarized in Table III. The observed retention times are slightly different from those reported for Crithidia (Table I), but are within the day-to-day variance of the chromatographic system (13) and are internally self-consistent within the same run. Due to the low quantity of the nucleosides analyzed (from approximately 54 ng of cap RNA), mass spectral responses were lower than those obtained for Crithidia, but clearly provide ion profiles with characteristic base or molecular ions for each component as indicated in Table III. A mixture of synthetic m5Um and m3Am was analyzed immediately following the T. brucei measurements, and the observed retention times (Fig. 15B; 24.6 and 33.4 min, respectively) are indistinguishable from those of U* and A*, respectively. Several additional minor UV-absorbing components appear in Fig. 15A, including the 18.5-min eluant, which was absent in the enzyme blank chromatogram; however, neither retention time nor mass spectral data permitted identification as nucleosides.

5 P. F. Crain and J. D. Bangs, unpublished observations.
DISCUSSION

Combined liquid chromatography-mass spectrometry was used for identification of nucleosides released from cap 4 by hydrolysis with nuclease P1, alkaline phosphatase, and nucleotide pyrophosphatase. Structure analysis of enzymatic digests by LC/MS is a powerful extension of either HPLC or mass spectrometry alone, due primarily to the great selectivity afforded by the use of mass as a detection parameter, in conjunction with HPLC which can distinguish isomers and other constituents having the same molecular weight (21). The structurally known constituents shown in Fig. 9A were assigned by comparison of thermospray mass spectral molecular ion (MH*) and base ion (BHt) mass values and chromatographic retention times (listed in Table I) with reference data for RNA nucleosides (13). These data indicated that components U* and A* represent ribose-methylated nucleosides not previously reported as constituents of RNA (13-15, 17). Inference that the base moieties are 3'- or 5-methyluracil and N6,N6-dimethyladenine, respectively, was based on mass and structural precedents for post-transcriptional modifications in RNA (15), but the lack of structural detail inherent in thermospray mass spectra prevented definitive structure assignments from these data alone.

Because of severe restrictions in cap 4 sample quantity, additional experiments which would require isolation of U* and A* were not carried out, and authentic m7Um and m7Am were chemically synthesized for direct comparison with cap nucleosides U* and A*. The primary method of comparison selected was capillary column gas chromatography-electron ionization mass spectrometry, viewed as the single most effective means for unambiguous tests of identity of nucleoside structures (27). This method combines the ability to reproducibly define chromatographic retention time within ±1-2 s, with the extensive and structurally informative fragmentation patterns derived from trimethylsilyl derivatives of nucleosides (22). The applications of this method directly to RNA hydrolysates when constituent nucleosides are present at levels below approximately 10 ng are often unsuccessful, due to the effects of salts and enzymes on the yield of conversion of >>NH to >>N-TMS in the base. However, in the present study, an attempt at trace-level silylation and chromatography was thought feasible because neither putative structure contains a silylation site in the base.

The experimental results conclusively demonstrate that U* is m7Um and A* is m7Am (Fig. 16). Structures not previously known in RNA, based on three elements: 1) close correspondence of gas chromatographic retention times for cap-derived and synthetic nucleosides, within approximately 0.5 s in each case (Figs. 12 and 13); 2) virtual identity of electron ionization mass spectra (Fig. 14, A versus B and C versus D); and 3) structure assignments from electron ionization mass spectra of U* and A* derivatives, listed in Table II. In addition, HPLC retention times of cap nucleosides and synthetic m7Um and m7Am (Fig. 9, A and C) were observed to be indistinguishable (within 2 s) and to exhibit matching 254 nm/280 nm UV absorbance ratios for both sets of nucleosides.

LC/MS analysis of a C. fasciculata pyrophosphatase-minus digest was used to confirm the identity of A* as nucleoside 1 of transcription by two criteria. 1) HPLC peaks for m7G and A* are absent in Fig. 9B, and a new peak representing the core cap constituents is observed at 29.4 min. 2) The latter peak produced characteristic BH2 ions for m7G (m/z 166) and m7Am (m/z 164). In addition, the core cap peak contained a signal for the BH2 ion of Am (m/z 136) which probably accounts for the reduced relative amount of Am in the absence of pyrophosphatase (Fig. 9, compare A and B). At least two possibilities could explain these related findings. In the first, partial nuclease P1 cleavage between positions 1 and 2 would lead to the presence of m7GpppAm. The core cap peak is broad, and co-elution of these structures cannot be ruled out. The relative resistance to hydrolysis by nuclease P1 of nucleosides modified in both the base and ribose has been noted (28) and the presence of contaminating phosphodiesterase 1 (see “Materials and Methods”) in nucleotide pyrophosphatase could account for the full recovery of Am in the total digest. Alternatively, the A residue in position 1 may be undermethylated in some fraction of the cap 4 molecules leading to the presence of m7GpppAm in the core cap peak. These possibilities are not mutually exclusive but scarcity of material precluded resolution of this issue.

The presence of guanosine in the cap analyses was not expected and any explanation of this finding must account for the diminution of the G signal in the analyses of pyrophosphatase-minus digests. This result dictates that the bulk of the guanosine must reside in either the extreme 5'-position (in place of 7-methylguanosine), in position 1 of transcription (in place of trimethyladenosine), or, if partial nuclease P1 products are present in the core cap peak as discussed above, it could also reside in position 2 (in place of 2'-O-methyladenosine). The latter two possibilities, which could result from sequence heterogeneity in the genomic repeats that code for the mini-exon, are unlikely since the presence of guanosine at one of these positions would provide a RNase cleavage site within the cap, and the resulting cap 0 or cap 1 structures would be eliminated by the purification procedure. We, therefore, favor the former possibility, which could result simply from a failure to 7-methylate the core cap guanosine residue following its post-transcriptional attachment. Undermethylation could be a consequence of the general metabolic state of the Crithidia cells, which were harvested in late log phase growth. Alternatively, differential cap methylation may play a role in the regulation of mRNA utilization in these organisms. A small amount of residual guanosine is detected in pyrophosphatase-minus digests. This is most easily explained by previously undetected heterogeneity at position 5 of medRNA transcription.

Based on the sequence of the mini-exon and comparison of the total versus pyrophosphatase-minus digests, a definitive structure can be proposed for the fully methylated Crithidia cap. This structure is identical with that shown in Fig. 1 where A* is N6,N6,2'-O-trimethyladenosine and U* is 3',2'-O-dimethyluridine. Structures for these, heretofore undescribed, naturally occurring nucleosides are shown in Fig. 16.

We have also analyzed the cap 4 of T. brucei and, based on: 1) its nucleoside composition (Table III), 2) the sequence of the T. brucei mini-exon (7), and 3) previous studies of cap from this source (2-5), conclude that its structure is identical with that of the fully methylated cap of C. fasciculata.
implication of these findings, along with the previous comparison of the *T. brucei* and *L. collosoma* caps (2), is that the same cap structure will be found in other, perhaps all, kinetoplastids. No guanosine was found in the *T. brucei* cap, consistent with the observations of Ullu and Tschudi (9) who used a permeabilized trypanosome cell system to radiolabel the cap guanosine residue and demonstrated that all of the label was incorporated into 7-methylguanosine. The presence of guanosine in cap 4 from *C. fasciculata* but not *T. brucei* may have a trivial explanation(s) (different sources, culture, and harvesting conditions, etc.) or may reflect real differences in regulation of methylation between the two species.

Complete information on cap structure will facilitate investigation of both its biosynthesis and its role in the processes of trans-splicing and translational initiation and perhaps in other, as yet unknown, functions. One immediate benefit will be the ability to design synthetic cap analogs as probes of cap function. An obvious extension of this strategy will be the design of cap analogs for use as trypanocidal agents. Inhibition of both its biosynthesis and its role in the processes of regulation of methylation between the two species.

Acknowledgments—We thank J. M. Gregson, J. A. Kowalak, and S. C. Pomerantz for experimental assistance and structural drawings and D. Perelman for purification of *Crithidia* RNA.

REFERENCES

1. Boothroyd, J. C. (1989) in *Nucleic Acids and Molecular Biology* (Eckstein, F., and Lilley, D. M. J., eds) Vol. 3, pp. 216–230, Springer Verlag, Berlin
2. Perry, K. L., Watkins, K. P., and Agabian, N. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 8190–8194
3. Sutton, R. E., and Boothroyd, J. C. (1988) *Mol. Cell. Biol.* 8, 494–496
4. Freistadt, M. S., Cross, G. A. M., Branch, A. D., and Robertson, H. D. (1987) *Nucleic Acids Res.* 15, 9861–9879
5. Freistadt, M. S., Cross, G. A. M., and Robertson, H. D. (1988) *J. Biol. Chem.* 263, 15071–15075
6. Rottman, F. M. (1978) in *International Review of Biochemistry, Biochemistry of Nucleic Acids II* (Clark, B. C., ed) Vol. 17, pp. 45–73, University Park Press, Baltimore
7. Campbell, D. A., Thornton, D. A., and Boothroyd, J. C. (1984) *Nature* 311, 350–355
8. Lenardo, M. J., Dorfman, D. M., and Donelson, J. E. (1985) *Mol. Cell. Biol.* 5, 2487–2490
9. Ullu, E., and Tschudi, C. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 10074–10078
10. Sonenburg, N. (1988) *Prog. Nucleic Acids Res.* 35, 174–207
11. Mizumoto, K., and Kaziro, Y. (1987) *Prog. Nucleic Acids Res.* 34, 1–28
12. Cordingly, J. S., and Turner, M. J. (1980) *Mol. Biochem. Parasitol.* 1, 91–96
13. Pomerantz, S. C., and McCloskey, J. A. (1990) *Methods Enzymol.* 193, 756–764
14. Sproul, M., Hartmann, T., Weber, J., Blank, J., and Zeidler, R. (1989) *Nucleic Acids Res. Sequences Suppl.* 17, r1–r172
15. Bjork, G. R. (1984) in *Processing of RNA* (Apirion, D., ed) pp. 291–330, CRC Press, Boca Raton, FL
16. Buck, M., Connick, M., and Ames, B. N. (1983) *Anal. Biochem.* 129, 1–13
17. Gehrze, C. W., and Kuo, K. C. (1990) in *Chromatography and Modification of Nucleosides* (Gehrke, C. W., and Kuo, K. C. T., eds) Part A, pp. A3–A71, Elsevier Science Publishing Co., Inc., New York
18. Takeda, N., Pomerantz, S. C., and McCloskey, J. A. (1991) *J. Chromatogr.* 562, 225–235
19. Buchich, M. L., Hughes, D. E., Simpson, A., and Simpson, L. (1987) *Nucleic Acids Res.* 15, 3141–3153
20. Gabriel, A., Siiodia, S. S., and Cleveland, D. W. (1987) *J. Biol. Chem.* 262, 16192–16199
21. Edmonds, C. G., Vestal, M. L., and McCloskey, J. A. (1985) *Nucleic Acids Res.* 13, 8197–8206
22. Pang, H., Schram, K. H., Smith, D. L., Gupta, S. P., Townsend, L. B., and McCloskey, J. A. (1982) *J. Org. Chem.* 47, 3932–3933
23. McCloskey, J. A. (1990) *Methods Enzymol.* 193, 825–842
24. McCloskey, J. A. (1974) in *Basic Principles in Nucleic Acid Chemistry* (Ty'o, P. O. P., ed) Vol. I, Chap. 3, Academic Press, New York
25. Pang, H., Schram, K. H., Smith, D. L., Gupta, S. P., Townsend, L. B., and McCloskey, J. A. (1982) *J. Org. Chem.* 47 (Microfilm Suppl.), 53 pp.
26. von Minden, D. L., Liehr, J. G., Wilson, M. H., and McCloskey, J. A. (1974) *J. Org. Chem.* 39, 285–289 and references therein
27. McCloskey, J. A. (1981) *Methods Enzymol.* 193, 771–781
28. Crane, P. F. (1980) *Methods Enzymol.* 193, 782–790
29. Bangs, J. D., Hereid, D., Krakow, J. L., Hart, G. W., and Englund, P. T. (1985) *Proc. Natl. Acad. Sci. U. S. A.* 82, 3207–3211
30. Englund, P. T. (1978) *Cell* 14, 157–168
31. MacDonald, R. J., Swift, G. H., Przybyla, A. E., and Chirgwin, J. M. (1987) *Methods Enzymol.* 152, 219–227
32. Jacobson, A. (1987) *Methods Enzymol.* 152, 254–261
33. Greenburg, J. J., and Burn, V. E. (1988) *Nucleic Acids Res.* 16, 3437–3454
34. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
35. Crain, P. F., and McCloskey, J. A. (1983) *Anal. Biochem.* 132, 124–131
36. Vestal, M. L. (1990) *Methods Enzymol.* 193, 107–130
37. Aoyama, T., and Shioiri, T. (1990) *Tetrahedron Lett.* 31, 5507–5508
38. Dunn, D. B., and Hall, R. H. (1975) in *Handbook of Biochemistry and Molecular Biology*, 3rd Ed., p. 127, CRC Press, Cleveland, OH
39. Hall, R. (1971) *The Modified Nucleosides in Nucleic Acids*, pp. 160–161, Columbia University Press, New York and references therein
40. Holmes, J. L. (1985) *Org. Mass Spectrom.* 20, 169–183
41. Wagner, D., Verheyden, J. P. H., and Moffatt, J. G. (1974) *J. Org. Chem.* 39, 54–59
42. Hall, R. (1971) *The Modified Nucleosides in Nucleic Acids*, pp. 96–97, Columbia University Press, New York and references therein.
null
mRNA Cap Structure of Trypanosomatids

Figure 7. Large Scale Purification of Cap 4. Peak II material from the large scale digestion of poly A⁺ RNA was fractionated by C4 reversed phase HPLC. Fractions were collected and the radioactivity in a sample of each was measured. UV absorbance was monitored at 264 nm. The sensitivity of detection in Panel B was 4x that in Panel A. Panels A and B. UV chromatograms of C. fasciculate and T. brucei peak II, respectively. Panels C and D. radiolabeling of C. fasciculate and T. brucei peak II, respectively.

Figure 8. Electrophoretic Analysis of C. fasciculate Cap 4 Purification. Equivalent samples (1/20th of the total) of each step of the large scale purification of C. fasciculate cap 4 purification were fractionated on a 20% polyacrylamide sequencing gel. The mobilities of xylene cyanol (XC), bromphenol blue (BPB), cap 4, [3²P]GMP and [3²P]GTP are indicated. Lane 1, mixed kinase digest; Lane 2, pooled peak II DEAE fractions; Lane 3, concentrated peak II; Lane 4, HPLC purified cap 4. The presence of large amounts of unlabelled digestion products (nucleotide monophosphates) from the bulk digests distorts the mobility of unincorporated [3²P]GMP and [3²P]GTP at the bottom of Lane 1.

Figure 10. Reconstructed ion chromatograms of core cap components eluting in the 29-30 min region of Figure 9B. Ion assignments: m/z 166, protonated 7-methylguanine; m/z 164, protonated A⁺ base; m/z 106, protonated adenine.

Figure 11. Thermospray mass spectra from directly-combined liquid chromatography/mass spectrometry of C. fasciculate RNA cap 4 total digests. (A) Mass spectrum of the nucleotide eluting at 24.5 min (Figure 9A), designated U⁺. (B) Mass spectrum of the nucleotide eluting at 33.4 min (Figure 9A), designated A⁺.
Figure 12. Comparison of gas chromatographic retention times of U" from C. fasciculata RNA cap 4 total digest and synthetic 5',3' di-methylguanosine, following trimethylation. Ion detection channel: m/z 441, corresponding to (M - CH₃O)⁻ (see figures 1A and 1B). Abscissa (upper) mass spectrometer scan number; (lower) retention time, min, referenced from a 2.0 min post-injection data acquisition delay. (A) Cap nucleoside U", and (B) m'PnU (THOM).