Mapping the 5' and 3' ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE)

Xiuwen Liu and Martin A. Gorovsky*

Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received June 23, 1993; Revised and Accepted September 3, 1993

Accession nos +

ABSTRACT

A procedure is described for mapping the ends of RNAs. Using T4 RNA ligase, a DNA (3' end) or RNA (5' end) oligonucleotide is ligated to RNA ends followed by cDNA synthesis, PCR amplification, cloning and sequencing. This method determines 5' ends, 3' polyadenylation sites and the size of poly(A) tails, and should be applicable to non-polyadenylated mRNAs and to non-message RNAs. Analysis of four *Tetrahymena thermophila* histone mRNAs revealed multiple, closely spaced 5' ends consistent with those determined by other methods. Except for a 'CCAAT' box in either orientation 100–200 nucleotides upstream of the transcription start site, no conserved sequence elements were observed in the untranslated 5' region or in sequences immediately flanking the transcription start site. Analysis of the 3' ends of mRNAs encoding four histones, two tubulins and the *Tetrahymena* TATA binding protein confirmed the observations that *Tetrahymena* histone messages are polyadenylated and that poly(A) tails in this organism are short (~50 nt). No canonical poly(A) addition signal was identified. The four histone messages analyzed here contained three sequence elements, TGTGT-TAA-AAGTATT, not found in non-histone messages. Two non-histone messages contained GCATT(N)₁₅ATACC near the poly(A) addition site.

INTRODUCTION

A variety of methods is available for mapping the 5' and 3' ends of RNA molecules, including RNase protection, S1 mapping and primer extension (1). These procedures are subject to various limitations and artifacts. In particular, we have found them difficult to use for mapping the extremely AT-rich (>80%) ends of messages of the ciliated protozoan *Tetrahymena thermophila* (2) and have frequently had to optimize each method for every new mRNA being studied. An alternative procedure based on the rapid amplification of cDNA ends (RACE) has recently been described (3,4). For 3' end analysis, the mRNA is reverse

transcribed using an oligo(dT)-containing adapter-primer. To obtain the 5' ends, mRNA is reverse transcribed with a gene-specific primer and oligo(dA) is synthesized on the 5' end of the cDNA using terminal transferase. For either reaction, the tailed cDNA is then amplified using an oligo(dT) adapter-primer and a gene-specific primer. For 3' ends, this method is necessarily restricted to polyadenylated RNAs and yields no information concerning the actual size of the poly(A) tail itself. Because reverse transcriptases are subject to premature termination, the 5' ends determined with this method may not be the actual transcription start sites.

To facilitate our analyses of transcription units in *Tetrahymena thermophila*, we have developed a new method for mapping 5' and 3' ends. This method involves the ligation of oligonucleotides to the 5' or 3' ends of mRNA before synthesis of cDNA, allowing determination of poly(A) tail length and avoiding the problem of premature termination during reverse transcription. This method has been used to map the ends of a number of *Tetrahymena* mRNAs, some of which had been difficult to map using conventional methods.

MATERIALS AND METHODS

Ligation of synthetic oligonucleotides to 3' ends of total RNA

RNA was isolated from *Tetrahymena thermophila* (Strain CU 428) as described (5). Briefly, 100–200 ml of culture containing exponentially growing cells were collected by centrifugation and washed once with 10 mM Tris–HCl, pH 7.6. Washed cells were resuspended in 8 ml of guanidine isothiocyanate solution (4M guanidine isothiocyanate, 0.5% sodium n-laurylsarcosine, 25 mM sodium citrate, pH 7.0, 0.1% β -mercaptoethanol, 0.1% antifoam A) and passed 10 times through an 18 gauge needle using a 10 ml syringe. RNA was pelleted through a 6.7 M CsCl cushion by centrifugation for 16–18 hr at 35,000 rpm in a Beckman 50 Ti rotor, resuspended in diethylpyrocarbonate (DEPC)-treated water and precipitated in ethanol.

All DNA oligonucleotides were synthesized on an Applied Biosystems Model 392 synthesizer. Oligo 1, which can be any available oligonucleotide (see Results and Fig. 1A), was 5'-C-

* To whom correspondence should be addressed

+GenBank; L18892, L18893, L01415, L01416 and L16957. EMBL; X04755 and X04756.

AAATTTTCATCGATATAAAAATATAAATG-3'. It is derived from the *Saccharomyces cerevisiae* histone HTA1 gene 5' untranslated region plus the initiator ATG (6). Oligo 1 was phosphorylated using T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions and 80 pmols were ligated to 10 μ g (~40 pmol) of total RNA using T4 RNA ligase (GIBCO-BRL or New England Biolabs, see Refs 7,8). Briefly, RNA and Oligo 1 were mixed and heated at 90°C for 2 min. Ligation was done in 100 μ l containing 50 mM HEPES (pH 7.5), 20 mM MgCl₂, 5 mM DTT, 10% DMSO, 10 μ g/ml BSA, 1 unit/ml RNasin (Promega), 240 pmoles of ATP and 100 units of ligase at 16°C for 2 hr. Free oligonucleotides were then eliminated by centrifugal ultrafiltration (Centricon 100 microconcentrators; Amicon) with 5 washes using 2 ml of DEPC-treated H₂O.

Ligation of short RNA transcripts to 5' ends

Tetrahymena RNA was de-capped by β -elimination as described (9). Briefly, 20 μ g of RNA was dissolved in 100 μ l of DEPC treated H₂O at 100°C for 90 sec to dissociate aggregates and then put on ice. An equal volume of freshly prepared sodium periodate solution (2 mM NaIO₄, 1% SDS) was added and the reaction was run for 1 hr in the dark at room temperature. Twenty μ l (1/10 volume) of 0.1 M glycerol was added and incubated for 5 min. The oxidized RNA was ethanol precipitated, dissolved in 100 μ l of DEPC treated H₂O, mixed with an equal volume of cyclohexylamine solution (0.8 M cyclohexylamine, 0.2 M Hepes, 2 mM EDTA) and incubated for 45 min at 45°C. De-capped RNA was treated with calf-intestinal alkaline phosphatase (GIBCO-BRL) to remove pyrophosphate groups and phosphorylated with T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions.

Although oligodeoxyribonucleotides can be efficient donors of 5' phosphate groups during ligation with T4 RNA ligase, they are not good acceptors in this reaction (8). Therefore, we ligated a small RNA fragment of known sequence to the 5' end. Seventy nucleotide long RNA transcripts were synthesized from plasmid Bluescript KS (+) (Stratagene) using T7 RNA polymerase after cleavage of the plasmid with restriction enzyme EcoRI (Fig. 1C). This fragment was then ligated to de-capped RNA under conditions described above for 3' end ligation. (Fig. 1B)

cDNA synthesis

For 3' end analysis, cDNA synthesis was primed with Oligo 2 whose sequence is complementary to that of Oligo 1 (Fig. 1A). For 5' end analysis, cDNA synthesis was primed with random hexamer oligodeoxyribonucleotides (Pharmacia). First strand reverse transcription reactions were carried out according to manufacturer's instructions using MMLV H⁻ reverse transcriptase (Superscript, GIBCO-BRL). Second strand synthesis is not necessary for PCR amplification. Excess primers were removed by centrifugal ultrafiltration as described above.

PCR amplification

One percent (2 of 200 μ l) of the cDNA was used for each PCR reaction with Taq DNA polymerase (Promega) in 50 mM KCl, 1.5 mM MgCl₂, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100 and 0.2 mM each of the four dNTPs. For 3' end amplification, Oligo 2 and an appropriate gene-specific primer corresponding to a sequence near the 3' end of coding sequence was used. For 5' end mapping, primer SK (Stratagene), derived

from the Bluescript plasmid (Fig. 1C) and a gene-specific primer hybridizing close to the 5' end of the coding sequence were used. In most experiments, an initial cycle (1 min at 94°C, 2 min at 45°C, 3 min at 72°C) was followed by thirty-five cycles of the PCR reaction (15 sec at 92°C, 30 sec at 45°C, 1 min at 72°C), followed by 5 min at 72°C. Control reactions omitted cDNA or one of the two primers. Note that PCR reaction conditions may need to be optimized individually for specific primers.

PCR products were separated by electrophoresis on 1–1.5% agarose gels. DNA fragments appearing only in reactions containing both primers and cDNA were eluted and cloned into plasmid vector Bluescript KS (+). In some cases (e.g., closely spaced 5' ends) regions of the gel containing multiple closely spaced fragments were eluted and cloned. Double-stranded sequencing was performed using published protocols (1).

RESULTS

General strategy for RNA ligase mediated amplification of cDNA ends (RLM-RACE)

The overall strategy for synthesizing the ends is illustrated in Fig. 1. For 3' end mapping (Fig. 1A) a DNA oligonucleotide (Oligo 1) is ligated to the ends of total RNA using RNA ligase. A second oligonucleotide (Oligo 2), complementary to the first, is used to synthesize cDNA using reverse transcriptase. A gene-specific oligonucleotide (Oligo 3) and Oligo 2 are then used to amplify the 3' end region of the gene of interest. For 5' end mapping (Fig. 1B), a small RNA transcript (Fig. 1C) is ligated to the 5' ends of RNA molecules from which the caps have been removed by β -elimination (9). Random hexamer oligonucleotides are then used to prime the reverse transcription of all RNA species. A DNA primer (SK, see Fig. 1C), corresponding to part of the small RNA transcript, plus a gene-specific primer (Oligo 4) are then used to amplify the 5' ends of the cDNA. This method yielded specific, defined DNA fragments, sometimes closely spaced (data not shown), for all the *Tetrahymena* genes examined so far (Figs. 2 and 3).

Comparison of 5' and 3' ends determined by RLM-RACE with those mapped by other methods

We have mapped the 5' and 3' ends of mRNAs encoded by a number of single copy genes using conventional methods (RNase protection, S1 protection, primer extension) or RACE and compared them to the ends mapped by the RLM-RACE method described here (Table 1). The RACE and the RLM-RACE methods gave identical results in every case in which both were applied. Ends determined by RLM-RACE (or RACE) were similar but not identical to those identified by conventional methods. We think this reflects inaccuracies in determining the true length of large fragments produced by conventional methods. Since the RLM-RACE and the RACE methods give identical results and are based on direct sequence analyses, it seems likely that these are more precise than conventional methods. Based on these comparisons, we conclude that the RLM-RACE accurately identifies the correct ends of messenger RNAs. Using this method, it is clear that transcripts from *Tetrahymena* genes frequently have multiple 3' and 5' ends. The 3' untranslated regions (UTRs) of *Tetrahymena* messages mapped to date vary in size from 55 to 461 nt and 5' UTRs vary in size from 26 to 60 nt.

Figure 1. Strategies for RLM-RACE amplification of 3'- and 5'-termini of mRNA molecules. Wavy lines represent RNA, straight lines represent DNA, and arrows represent oligonucleotides. AAA...AAA represents poly(A) tails of mRNAs, but this method is also applicable to nonpolyadenylated RNA species. (A), Amplification of 3'-termini. Total cell RNAs are ligated to DNA Oligo 1 using T4 RNA ligase. The complementary oligonucleotide (Oligo 2) is then used to prime cDNA synthesis. Oligo 2 and Oligo 3 (gene-specific) are used in the PCR reaction to amplify the 3'-ends of RNA molecules. (B), Amplification of 5'-termini. RNA caps are removed by β -elimination. The pyrophosphate groups are then removed by calf intestinal alkaline phosphatase (CIP) to leave hydroxyl groups (OH) at the 5'-termini, which are phosphorylated with T4 polynucleotide kinase. These RNAs were then ligated to small RNA transcripts containing the sequence of primer SK (see below). Random primed cDNAs representing various regions of all RNA species are then PCR amplified using primer SK and Oligo 4 (gene-specific). (C), Sequence of the RNA transcript ligated to the 5'-termini of RNAs. This RNA is made by transcribing plasmid Bluescript KS(+) (Stratagene) with T7 RNA polymerase after EcoRI linearization of the plasmid. The transcript initiates from +1 (T7 transcription start site) and terminates at the EcoRI site. SK primer (in bold) is part of the polylinker containing recognition sequences for restriction enzymes XbaI, SpeI, and part of BamHI. The PCR products will contain SmaI and PstI sites as well.

***Tetrahymena* histone and nonhistone messages contain short poly(A) tails**

In most organisms studied to date, cell-cycle regulated histone messages lack poly(A) tails and contain a typical stem-loop structure in the 3' UTR that is important for creation of the 3' end (see 10 for review). Based on the absence of a stem-loop structure in *Tetrahymena* histone mRNAs and their ability to bind to an oligo(dT) column, it has been suggested that *Tetrahymena*

histone messages are polyadenylated (11). However, because the 5' and 3' untranslated regions of *Tetrahymena* histone (and other) messages are extremely ($\geq 80\%$) AT-rich, it also has been suggested (12) that their binding to oligo(dT) may not be due to the presence of a poly(A) tail. In the RLM-RACE method, a primer is attached directly to the message 3' end, enabling direct determination of the existence and length of any poly(A) tail. Sequencing the 3' ends of 18 RLM-RACE clones derived from

Figure 2. Mapping the 5'-termini of 4 histone genes. The genomic sequences of 4 histone genes are aligned at the initiator ATG. Transcription start sites are marked by underlined bold letters, and the numbers under them denote numbers of clones obtained and sequenced for each site. The 'CCAAT' box sequences (TATCCAAT-CAGA) or their inverse sequences (TCTGATTGGATA) are indicated. GenBank accession numbers are: H2A.1 = L18892; H2A.2 = L18893; H4-I = X04756; H4-II = X04755.

A

B

Figure 3. Mapping the 3'-ends and poly(A) tails of *Tetrahymena* mRNAs. The poly(A) tract is indicated by (A)n where n indicates the number of As in each clone sequenced. The multiple poly(A) addition sites for H4-I and TATA binding protein messages are indicated by superscripts. The length of poly(A) tails at each site is indicated by n1, n2, n3, n4. Dots are simply for alignment and do not represent nucleotides. Only 75 bases preceding the poly(A) site are shown, (except for TATA binding protein) and the distances between these 75 bases, and the terminator TGA are indicated for each gene. (A) Alignment of four histone 3'-termini. One conserved sequence motif, TGTGT-TAA-AAGTATT, emerged from this alignment, 26-30 bases upstream of the polyadenylation site. (B), Alignment of 3'-termini for two β -tubulin genes (Btu1 and Btu2) and one TATA binding protein gene (TBP). A bipartite sequence motif GCATT-ATACC was identified immediately upstream of the cleavage/polyadenylation sites for Btu1 and TATA binding protein, and the spacing between the two parts of the element is also identical. This sequence does not exist in Btu2. The 3' ends of the TATA binding protein message were mapped by L.Stargell (16). GenBank accession numbers are: H2A.1 = L18892; H2A.2 = L18893; H4-I = X04756; H4-II = X04755; Btu1 = L01415; Btu2 = L01416; TBP = L16957.

four different histone mRNAs revealed the existence of poly(A) tails in every case, with sizes ranging from 30 (H2A.1) to 76 (H4-II) nt (Fig. 3A). At least two of these histone genes (H4-I, H4-II) have been shown to be cell-cycle regulated (13). These results demonstrate unambiguously that mRNAs encoding major, cell-cycle regulated histones are polyadenylated in *Tetrahymena*.

Previous studies (using RNase digestion) indicated that the majority of *Tetrahymena* messages contained poly(A) tracts averaging about 30 nt in length (14). Again, these estimates are subject to methodological errors and to unknown contributions from A-rich, untranslated regions. Analysis of the 3' ends of 14 RLM-RACE clones derived from three nonhistone mRNAs (Fig. 3B) demonstrates that each contains a poly(A) sequence

ranging in size from 23-95 nt. Thus, the RLM-RACE method confirms the observation that most, if not all, *Tetrahymena* messages contain short poly(A) tails, although they are slightly longer than previously estimated.

Polyadenylated *Tetrahymena* messages lack canonical poly(A) addition signals and may contain other motifs

Polyadenylated messages in most organisms contain a consensus sequence (AAUAA) 10-30 nt from the polyadenylation site. No such sequence is found in the 3' ends of the histone or nonhistone mRNAs analyzed by RLM-RACE (Fig. 3).

The four *Tetrahymena* histone messages analyzed here all contain a conserved, interrupted motif, TGTGT(N)₁₋₈TAA-

Table 1. Comparison of 5' and 3' ends of *Tetrahymena thermophila* messages mapped by different methods

Gene	End	Methods		
		Conventional mapping ^{1,2}	RACE	RLM-RACE
H2A.1	5'	-61, -58 ³	—	-60, -56, -44
H2A.2	5'	-52, -51 ³	—	-49, -48, -43
H4I	5'	-55, -51, -45 ⁴	—	-55, -52, -26
H4II	5'	-59, -54 ⁴	—	-59, -48, -42, -29
H2A.1	3'	473 ³	461 ⁶	461
H2A.2	3'	136, 137 ³	130 ⁶	130
H4I	3'	407 ⁵	—	406, 411
H4II	3'	152 ⁵	151 ⁷	151
Btu1	3'	—	298 ⁸	298
Btu2	3'	—	219 ⁸	219
TBP	3'	—	—	55, 62, 71, 75 ⁹

1. Conventional methods are based on RNase protection, S1 mapping, and/or primer extension.

2. Number of bases upstream from starting ATG (5' end) or downstream from terminator TGA (3' end).

3. Mapped by RNase protection (Liu, X., and Gorovsky, M. unpublished observations).

4. A consensus of RNase protection, S1 mapping and primer extension; see reference (2).

5. By RNase protection; see reference (2).

6. Liu, X., and Gorovsky, M. unpublished observations.

7. Bowen, J., and Gorovsky, M. unpublished observations.

8. Reference (17).

9. Reference (16).

(N)₀₋₁₁AAGTATT, 26–30 nt upstream from the poly(A) addition site (Fig. 3A). The sequence identity extended for 3 additional nucleotides (CTC) in the case of H2A.1 and H4-I. This motif is not present in the 3' UTRs of the nonhistone mRNAs, and it occurred rarely in a search of GenEMBL database suggesting it could be a *Tetrahymena* histone-specific cleavage/polyadenylation signal or could serve some other specific function in the metabolism of histone mRNAs. However the complete motif also was not found in three other *Tetrahymena* histone genes for which extensive 3' untranslated sequence is available (5,12,15). The nonhistone messages encoding the TATA binding protein (16) and one of the two β -tubulins (17) contained identical copies of the sequence GCATT(N)₁₅ATACC near the poly(A) addition site. This sequence is not present in messages from the other β -tubulin gene (Fig. 3B). These observations suggest that, in *Tetrahymena*, the mechanism of 3' end formation probably differs from that in higher eukaryotes where the sequence AAUAAA causes cleavage and polyadenylation 10–30 nt downstream (18,19). *Tetrahymena* may be more similar to *Saccharomyces cerevisiae* where multiple classes of signal appear to determine the position of 3' ends of messenger RNAs (20).

The 5' UTRs of *Tetrahymena* histone mRNAs

The 5' UTRs of all four histone mRNAs examined were highly AT-rich with frequent short oligo(A) stretches. No other conserved sequence elements were observed in the 5' UTRs. Fifteen of 18 cloned H2A and H4 messages started with A; the remaining 3 with T. As previously reported (2,21), analysis of the sequences 5' of the putative transcription start sites failed to reveal consistently positioned TATA boxes. The only conserved element found was a 12 nt 'CCAAT' box (Fig. 2). All four histone genes contained at least one sequence matching this element in at least 11 of the 12 positions, starting 121 to 182 nt upstream of the initiator ATG which is invariably the first ATG in the message. While it seems likely that these 'CCAAT' boxes are promoter elements (21) it seems unlikely that they function to determine the transcription initiation site *per se* since they occur

in both orientations and variable distances from the transcription start sites.

DISCUSSION

We have described a modification of the RACE technique (3,4) for mapping cDNA ends in which a short DNA (3' end) or RNA (5' end) oligonucleotide is ligated directly to the ends of RNA molecules using RNA ligase. This method, referred to as RLM-RACE, was shown to accurately map both the 3' and 5' ends of *Tetrahymena* mRNAs which we have found difficult to map by other techniques, presumably due to the extreme AT-richness of the *Tetrahymena* UTRs.

The RLM-RACE method shares a number of the advantages of the original RACE method for mapping cDNA ends or with subsequent improvements of that method (22). These include precise mapping of ends to single nucleotides based on DNA sequencing and the fact that, for 3' end mapping, once the cDNA is synthesized, it can subsequently be used to map the ends of any mRNAs in the population. In addition, RLM-RACE offers additional advantages over the original RACE technique. Because it is synthesized with non-specific primers, the cDNA for 5' end mapping can also be used to map any mRNA in the population. Because primers are ligated directly to the ends of the RNA being analyzed before reverse transcription and PCR amplification, problems arising from premature termination or snapback addition of bases to the 5' ends during reverse transcription should be avoided. Thus, only cDNA products extending into the small RNA fragment ligated to the 5' ends can be amplified. The RLM-RACE technique also allows determination of the length of the 3' poly(A) tail on polyadenylated mRNAs which is not possible with RACE, where the 3' oligo (dT) adapter primer can hybridize anywhere along the length of the tail. Also, RLM-RACE should be equally suited to 3' end analysis of non-polyadenylated (message or other) RNAs, since the initial ligation of the 3' primer-adaptor requires only a 3' hydroxyl group. The RLM-RACE technique also eliminates the need for oligo(dT)-containing primers which can hybridize to A-rich regions in 5' or 3' UTRs

such as those found in plants, *Tetrahymena*, yeast and *Dictyostelium*.

As described, the RLM-RACE technique, like the RACE technique, can still misidentify 5' ends because of hydrolysis. For mapping the 5' ends of specific messages, this problem is minimized in both methods by selecting specific sized PCR products for subsequent cloning; hydrolytic products should be of diverse sizes. With the RLM-RACE method, it should be possible to minimize this problem by first digesting the RNA with a 5' phosphatase and then using tobacco acid pyrophosphatase (23) instead of β -elimination to remove the cap structure. This enzyme leaves a single, free 5' phosphate group at the end of each de-capped mRNA molecule, obviating the need to phosphorylate the 5' ends prior to ligation. In that case, the 5' hydroxyl ends produced by phosphatase digestion should not be a substrate for RNA ligase. Studies testing this approach are now in progress.

A method has recently been described (23) for sequencing the termini of viral RNAs by circularization of RNA followed by PCR amplification. This method offers some of the same advantages as the RLM-RACE method. However, it requires that the RNA molecule whose ends are being mapped be long enough and sufficiently devoid of secondary structures for the two ends of the RNA molecule to be physically brought together for ligation. Also, for 3' end mapping of mRNA, this method still requires de-capping to enable ligation. Thus, cloning and mapping 3' ends with RLM-RACE is easier. The cyclization method also requires that RNA preparations be ligated at high dilutions to avoid intermolecular reactions competing with intramolecular cyclization. One attempt we made to use the cyclization method to map *Tetrahymena* H2A gene transcripts was unsuccessful, yielding a clone containing the 5' end of the H2A.1 message ligated to an unknown sequence. Another advantage of the RLM-RACE method is that once the cDNA is made it can be stored and used to map any gene for which enough information is available to synthesize gene-specific primers. In the cyclization method, gene-specific primers are required for each cDNA. While it may be possible to store the cyclized RNA as a 'library', the greater susceptibility of RNA to hydrolysis probably would make this a risky approach.

The RLM-RACE method bears some similarity to other anchored PCR methods (22,24) for amplifying cDNAs whose 5' ends are unknown. These methods are similar to RLM-RACE in using (RNA or DNA)ligase to covalently link an 'anchor' oligonucleotide that can be used for subsequent PCR amplification. These methods differ from RLM-RACE in that the anchor is added to the 3' end of first strand cDNA while in RLM-RACE it is added directly to the 5' end of decapped mRNA. Thus, these methods avoid the decapping step but, like the original RACE method, are subject to misidentification of 5' ends due to premature termination or snapback addition of nucleotides to the 5' end during reverse transcription.

We have used the RLM-RACE method to map the ends of a number of abundant messages (histones, tubulins) whose concentrations in *Tetrahymena* are on the order of 10^4 - 10^6 molecules per cell (14,25), as well as the much rarer message for TATA binding protein (16). In all cases we obtained identifiable, specific fragments when DNA was isolated from the PCR amplification step, run on agarose gels and stained. For extremely rare mRNAs, end fragments may be present in too low a concentration to detect or may be indistinguishable from non-specifically amplified fragments after a single round of PCR.

In these cases, additional rounds of PCR amplification using the end primer and a third (sequence-specific) 'nested' primer (26-28) may be necessary as we found to be the case for the TATA binding protein message. We have also observed that the minimum number of PCR cycles required to map the 3' ends was dependent on the distance between the gene-specific primer and the end of the molecule. Thus, it took about ten more PCR cycles to obtain similar amounts of ends from the longer H2A.1 and H4-I messages than from the H2A.2 and H4-II, even though Northern blot analysis indicates all four messages are present in similar amounts (data not shown).

RLM-RACE analysis of the untranslated regions of *Tetrahymena* messages confirmed and extended previous findings. Many messages have multiple 5' and/or 3' ends. Both histone and nonhistone messages contain short poly(A) tails but lack canonical poly(A) addition sequences. Canonical TATA boxes also are lacking. Some histone messages contain a conserved tripartite sequence element in their 3' UTR that is not found in the nonhistone messages, while a different bipartite element was found in the 3' UTR of two (of three) unrelated nonhistone messages examined. This arrangement of conserved sequence elements is similar to that found in *Saccharomyces cerevisiae* where at least three different signal elements (one bipartite) located varying distances from the 3' termini have each been found to interact with preferred sites at or near the polyadenylation site to produce 3' termini (20). As in *Tetrahymena*, no conserved sequence elements have been identified at the polyadenylation site itself. The conserved elements identified in yeast have no homology to those identified here. With the availability of newly developed methods for mass transformation of *Tetrahymena* (29), it should now be possible to analyze the function of the *Tetrahymena* elements *in vivo*.

In summary, we have described a method (RLM-RACE) for mapping the 5' and 3' ends of messenger RNAs. This method was shown to be accurate and to be applicable to abundant and rare *Tetrahymena* mRNAs, some of which have proved to be difficult to map by conventional methods. For 3' end mapping, this method should be applicable to any RNA molecule containing a 3' hydroxyl group. For 5' end mapping, the only requirement is the ability to create a 5' phosphate. Another potential use of this method is to make truly full-length cDNA libraries that can be directionally cloned simply by attaching primer-adapters to both ends followed by PCR amplification. We are currently exploring this possibility.

ACKNOWLEDGEMENTS

This work was supported by Public Health Service Grants GM21793 and GM26973 from the National Institutes of Health and by Grant NP-717A from the American Cancer Society.

REFERENCES

1. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1988) Current Protocols in Molecular Biology. Wiley Interscience, New York.
2. Horowitz, S., Bowen, J.K., Bannon, G.A. and Gorovsky, M.A. (1987) Nucleic Acids Res., **15**, 141-160.
3. Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA, **85**, 8998-9002.
4. Frohman, M.A. (1990) In Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds.), PCR Protocols. Academic Press, San Diego, CA, pp. 28-38.

5. White, E.M., Shapiro, D.L., Allis, C.D. and Gorovsky, M.A. (1988) *Nucleic Acids Res.*, **16**, 179–198.
6. Choe, J., Kolodrubetz, D. and Grunstein, M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1484–1487.
7. Romaniuk, P.J. and Uhlenbeck, O.C. (1983) *Meth. Enzymol.*, **100**, 52–59.
8. Brennan, C.A., Manthey, A.E. and Gumpert, R.I. (1983) *Meth. Enzymol.*, **100**, 38–52.
9. Greenberg, J.R. and Burn, V.E. (1988) *Nucleic Acids Res.*, **16**, 3437–3454.
10. Marzluff, W.F. and Pandey, N.B. (1988) *Trends Biochem. Sci.*, **13**, 49–52.
11. Bannon, G.A., Calzone, F.J., Bowen, J.K., Allis, C.D. and Gorovsky, M.A. (1983) *Nucleic Acids Res.*, **11**, 3903–3917.
12. Nomoto, M., Imai, N., Saiga, H., Matsui, T. and Mita, T. (1987) *Nucleic Acids Res.*, **15**, 5681–5697.
13. Yu, S.-M., Horowitz, S. and Gorovsky, M.A. (1987) *Genes & Dev.*, **1**, 683–692.
14. Calzone, F.J., Angerer, R.C. and Gorovsky, M.A. (1983) *J. Biol. Chem.*, **258**, 6887–6897.
15. Wu, M., Allis, C.D., Richman, R., Cook, R.G. and Gorovsky, M.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8674–8678.
16. Stargell, L.A. and Gorovsky, M.A. (1993) *Mol. Cell. Biol.*, **submitted**.
17. Gaertig, J., Thatcher, T.H., McGrath, K.E., Callahan, R.C. and Gorovsky, M.A. (1993) *Cell Motil. Cytoskel.*, **25**, 243–253.
18. Fitzgerald, M. and Shenk, T. (1981) *Cell*, **24**, 251–260.
19. Wickens, M. and Stephenson, P. (1984) *Science*, **226**, 1045–1051.
20. Russo, P., Li, W.-Z., Hampsey, M., Zaret, K.S. and Sherman, F. (1991) *EMBO J.*, **10**, 563–571.
21. Brunk, C.F. and Sadler, L.A. (1990) *Nucleic Acids Res.*, **18**, 323–329.
22. Dumas, J.B., Edwards, M., Delort, J. and Mallet, J. (1991) *Nucleic Acids Res.*, **19**, 5227–5232.
23. Mandl, C.W., Heinz, F.X., Puchhammer-Stockl, E. and Kunz, C. (1991) *BioTechniques*, **10**, 484–486.
24. Troutt, A.B., McHeyzer-Williams, M.G., Pulendran, B. and Nossal, G.J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 9823–9825.
25. Yu, S.-M. and Gorovsky, M.A. (1986) *Nucleic Acids Res.*, **14**, 7597–7615.
26. Frohman, M.A. and Martin, G.R. (1989) *Technique*, **1**, 165–170.
27. Jain, R., Gomer, R.H. and Murtagh, J.J., Jr. (1992) *BioTechniques*, **12**, 58–59.
28. Frohman, M.A. (1993) *Methods Enzymol.*, **218**, 340–356.
29. Gaertig, J. and Gorovsky, M.A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 9196–9200.