RNA (1995), *1*:363-374. Cambridge University Press. Printed in the USA.

Analysis of the structure of *Tetrahymena* nuclear RNAs in vivo: Telomerase RNA, the self-splicing rRNA intron, and U2 snRNA

ARTHUR J. ZAUG and THOMAS R. CECH

Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, USA

ABSTRACT

Dimethyl sulfate modification of RNA in living *Tetrahymena thermophila* allowed assessment of RNA secondary structure and protein association. The self-splicing rRNA intron had the same methylation pattern in vivo as in vitro, indicating that the structures are equivalent and suggesting that this RNA is not stably associated with protein in the nucleolus. Methylation was consistent with the current secondary structure model. Much of telomerase RNA was protected from methylation in vivo, but the A's and C's in the template region were very reactive. Thus, most telomerase is not base paired to telomeres in vivo. Protein-free telomerase RNA adopts a structure different from that in vivo, especially in the template and pseudoknot regions. The U2 snRNA showed methylation protection at the Sm protein-binding sequence and the mRNA branch site recognition sequence. For both telomerase RNA and U2 snRNA, the in vivo methylation pattern corresponded much better to the structure determined by comparative sequence analysis than did the in vitro methylation pattern. Thus, as expected, comparative analysis gives the structure of the RNA in vivo.

Keywords: group I intron; methylation; protein-RNA interactions; ribozyme

INTRODUCTION

In the macronuclei of *Tetrahymena*, the large rRNA is synthesized as a precursor molecule containing a 413-nt group I intron. The intron RNA folds to form an active site for its own splicing reactions (Cech, 1990). Splicing is estimated to occur ~20-times more rapidly in vivo than the optimal rate measured in vitro (Brehm & Cech, 1983; Zhang et al., 1995), leaving open the possibility that proteins further facilitate this RNA-catalyzed reaction in the cell. Indeed, some group I introns form stable RNA-protein complexes that enable splicing under physiological conditions (Gampel et al., 1989; Lambowitz & Perlman, 1990; Weeks & Cech, 1995), whereas in other cases, RNA chaperones may bind transiently to increase the fraction of correctly folded molecules (Coetzee et al., 1994; Herschlag, 1995). Thus, the form of the *Tetrahymena* pre-rRNA intron that is active in vivo remains an open question.

Reprint requests to: Thomas R. Cech, Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, USA.

Another *Tetrahymena* RNA of much current interest is its telomerase RNA, which provides the template for telomeric DNA synthesis (Greider & Blackburn, 1989; Yu et al., 1990). Telomerase is a ribonucleoprotein whose protein and RNA components are both required for activity (Greider & Blackburn, 1987). Among other functions, protein may provide the active site that binds TTP, dGTP, and the template-primer complex and catalyzes primer elongation. The minimal roles of the RNA would then be to act as template and to interact with protein subunits. The secondary structure of the *Tetrahymena* telomerase RNA has been determined by comparative sequence analysis (Romero & Blackburn, 1991; ten Dam et al., 1991), and most of this secondary structure is conserved in hypotrichous ciliates as well (Lingner et al., 1994). There is currently no information regarding what portions of the RNA interact with protein in vivo.

Chemical modification provides a valuable approach to assess RNA structure and RNA-protein interactions (Peattie & Gilbert, 1980; Moazed et al., 1986). Dimethyl sulfate (DMS) enters living cells and reacts with macromolecules including RNA and DNA, thereby pre-

serving information about the accessibility of individual atoms in vivo (Ephrussi et al., 1985; Nick & Gilbert, 1985; Price & Cech, 1987; Climie & Friesen, 1988; Ares & Igel, 1990). Methylation of N-1 of A and N-3 of C, which is not revealed by Maxam-Gilbert chemical cleavage, is readily detected as stops or pauses during primer extension by reverse transcriptase (Inoue & Cech, 1985). Methylation at these positions is readily interpreted because N-1 of A and N-3 of C are directly involved in base pair formation; thus, these positions are modifiable in single-stranded RNA and become protected upon base pairing, tertiary structure formation, or binding by protein (Inoue & Cech, 1985; Moazed et al., 1986).

In the current study, we have treated living *Tetrahymena* with DMS, extracted the RNA, and used specific primers to probe the sites of methylation of the group I intron and telomerase RNAs. As a control, we also probed the U2 snRNA, whose protein-binding interactions in the snRNP are better known. We report evidence for protein interaction with both U2 snRNA and telomerase RNA, but find that the rRNA intron has a modification pattern strikingly similar to that of the deproteinized RNA.

RESULTS

Group I intron

The in vivo methylation pattern of the rRNA intron was determined by primer extension from four different primers. This gave overlapping information, allowing all regions of the RNA to be assessed except for the primer-binding site at the 3′-terminus (350–414) and the positions closest to the 5′ end of the full-length intron. Judging by the low extent to which primer extension proceeded beyond the 5′ end of the intron, the RNA being probed consisted mainly of the excised intron rather than pre-rRNA. A sample of the data is shown in Figure 1.

In the summary in Figure 2, A's and C's methylated in vivo are indicated by arrows or arrowheads superimposed on the phylogenetically established secondary structure model. Approximately half of the A's in single-stranded regions are methylated, as are some of the C's (which are chemically less reactive and therefore often not detected; Inoue & Cech, 1985). Because the secondary structure of this RNA is so well established, most of the unmethylated A's in loop regions are likely to be involved in tertiary interactions. Some of these have been identified: L5b interacts with P6a (Murphy & Cech, 1994), the A-rich bulge that interrupts P5a is involved in a subdomain structure (Murphy & Cech, 1994), L9.1 has been proposed to interact with L2.1 (Been et al., 1987; Banerjee et al., 1993), and L9 has been proposed to interact with base pair 2 (G-C) of P5 in a subgroup of group I introns (Michel & Westhof, 1990).

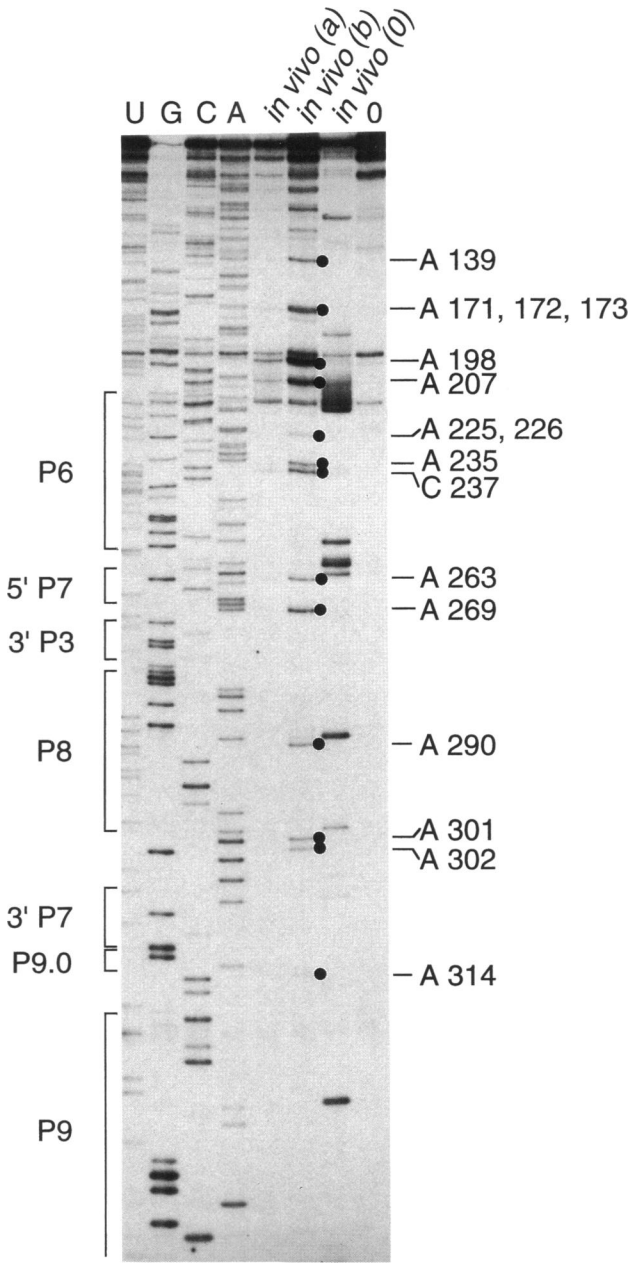


FIGURE 1. Group I intron. Example of in vivo methylation analyzed by primer extension with reverse transcriptase. All lanes show extension of primer IP341-22, which allows the intron RNA to be analyzed in the mixture of total nuclear RNA. Lanes U, G, C, A, dideoxynucleotide sequencing using unmethylated nuclear RNA. Lanes *in vivo (a)* and *in vivo (b)*, RNA methylated in vivo with 40 and 120 mM DMS, respectively. Dots represent primer extension stops that were specific to in vivo methylation (nucleotides identified at the right). Termination at a methylated site occurs one nucleotide prior to the corresponding dideoxy termination product. Lane *in vivo (0)*, a control in which the β-mercaptoethanol stop solution was added to the cells prior to the DMS. Lane 0, primer extension of unmethylated nuclear RNA without dideoxynucleotides.

In marked contrast, the A's and C's in proposed double-stranded regions are almost all protected from

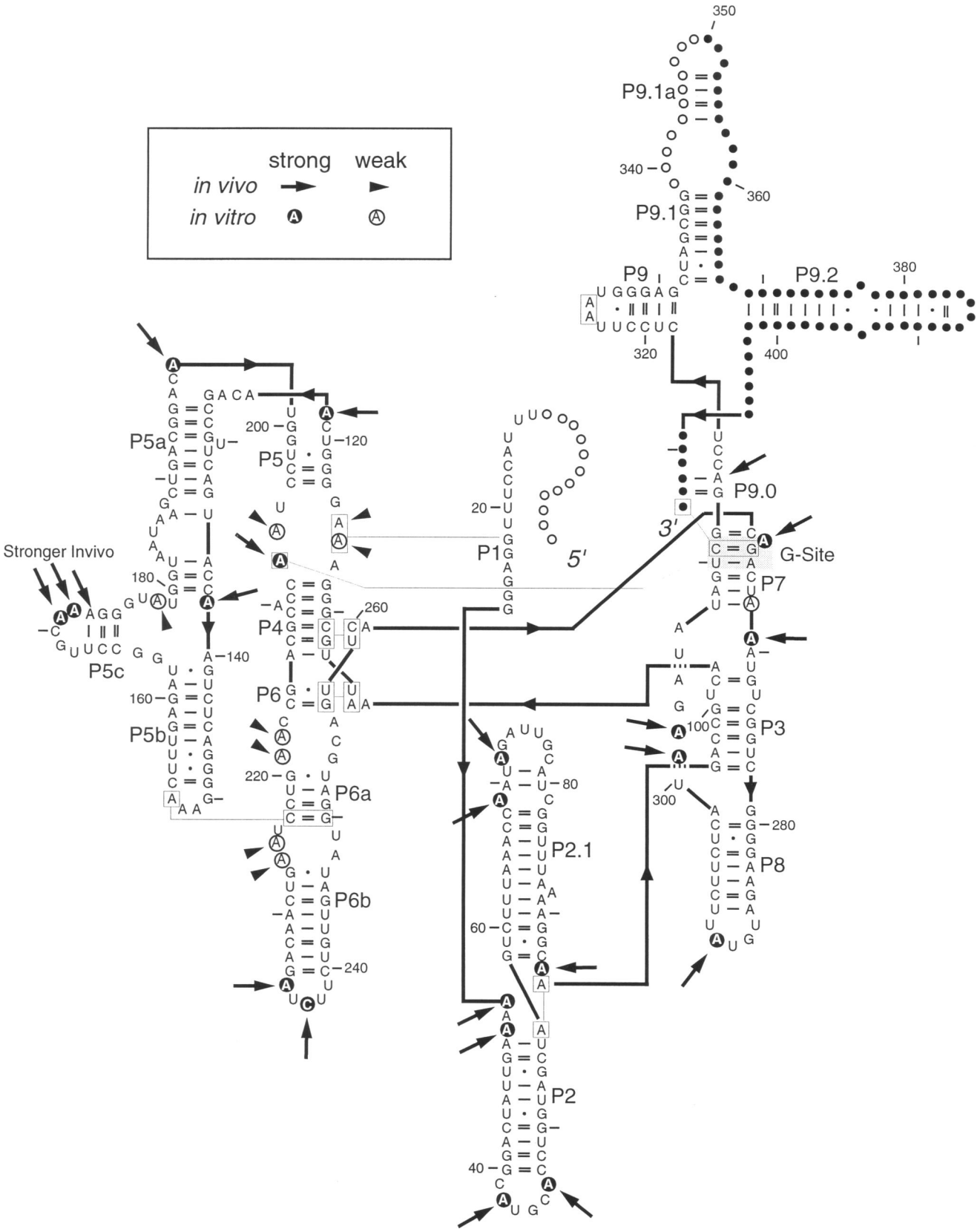


FIGURE 2. Sites of DMS modification of the group I intron superimposed on the secondary structure determined by comparative sequence analysis (Michel & Westhof, 1990). The domain structure representation shown here is from Cech et al. (1994), with light boxes around nucleotides implicated in tertiary interactions. Arrows show sites of modification in vivo; circled nucleotides show sites of modification after deproteinization. Filled dots, nucleotides that could not be evaluated due to primer binding site. Open dots, nucleotides too close to the 3′-terminal primer or too close to the 5′ end to be read with confidence.

methylation. There are only two exceptions, A173 in P5c and A314 in P9.0. The A314 modification suggests that P9.0, which contributes to the second step of RNA splicing (Burke et al., 1990; Michel et al., 1990), may be paired only transiently or only in a specific stage of splicing. A detailed study of this interaction in the case of the *sunY* intron has shown that the P9.0 pairing is present in at least some precursor molecules prior to splicing (Jaeger et al., 1993). However, these authors also show that this interaction confers no additional stability to the intron core, so it is reasonable to think of it as functioning to bring the 3′ splice site "substrate" into the catalytic core for step 2 of splicing. As such, it is reasonable that P9.0 in the *Tetrahymena* intron might be one of the more labile pairing interactions.

In addition, the nuclear RNA was deproteinized prior to DMS treatment and analyzed by primer extension and gel electrophoresis side by side with RNA methylated in vivo (multiple experiments; data not shown). The in vitro and in vivo patterns agreed at 27 of 31 positions (Fig. 2; Table 1). This provides strong evidence that the RNA structure is essentially the same in vivo and in vitro and provides an argument against the RNA being stably associated with protein. A small number of positions were more reactive to DMS in vivo than in vitro: the A's in L5c, A115 in the P4–P5 internal loop, and A314 in P9.0. Such a difference in structure could result if the in vitro conditions overstabilized the RNA structure relative to the state in vivo, or if the RNA folding was slightly different in vitro.

TABLE 1. Correspondence of nucleotides methylated in vivo and in vitro.

RNA	Number of nucleotides methylated[a]				Agreement[b] (%)	Agreement if random[c] (%)
	Only in vitro	Only in vivo	Both	Total		
Group I	1	3	27	31	87	10
Telomerase	21	8	20	49	41	31
U2 snRNA	10	2	13	25	52	20

[a] Both, methylated both in vitro and in vivo; total, methylated in vitro or in vivo or both. Strong and weak extents of methylation were lumped together for this analysis.

[b] Percent correspondence between in vitro and in vivo methylation patterns, calculated as (Both/Total) × 100.

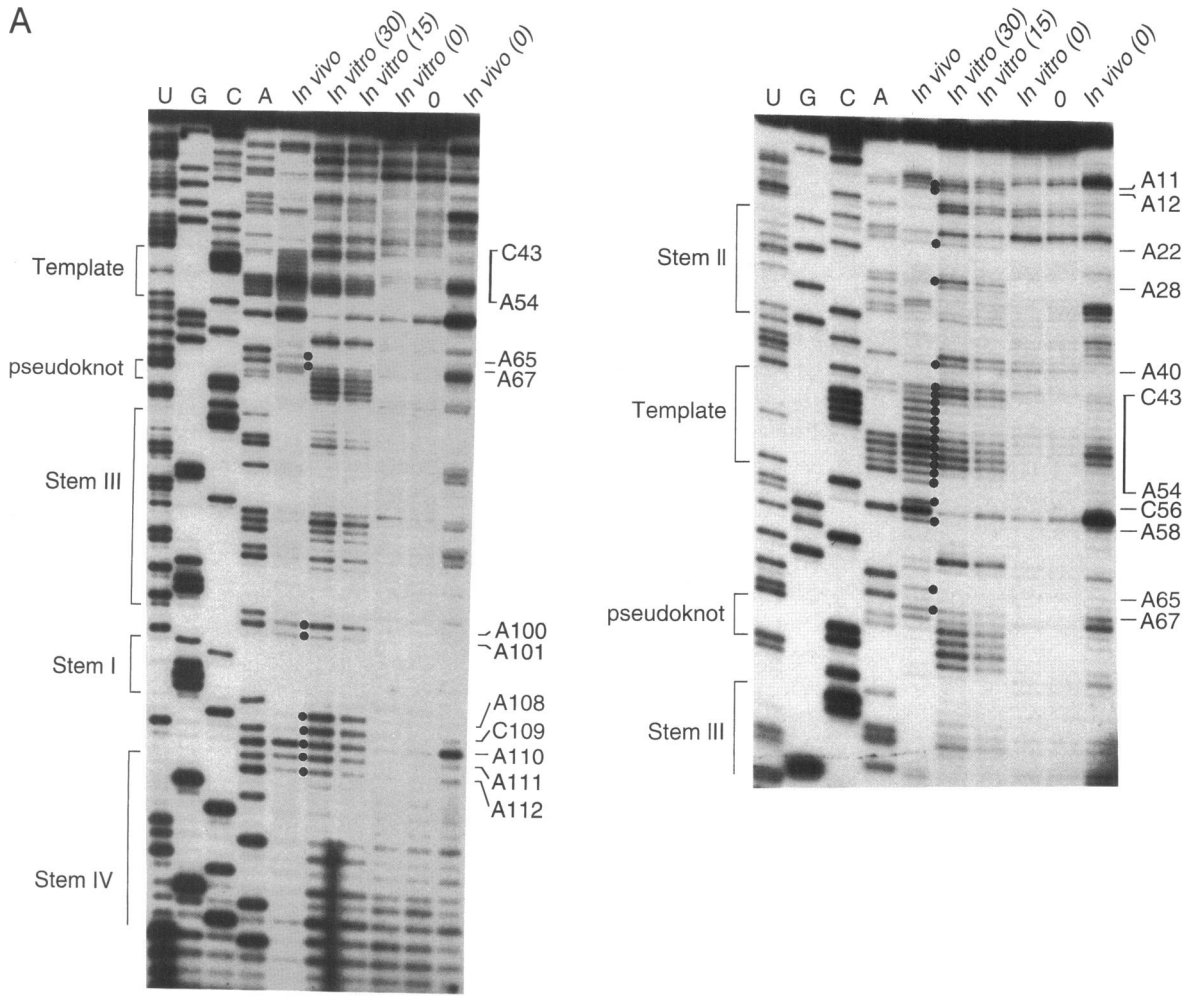
[c] Percent correspondence between the two methylation patterns if both patterns were random, calculated as $P_{\text{in vitro}} \times P_{\text{in vivo}} \times 100/[P_{\text{in vitro}} + P_{\text{in vivo}} - (P_{\text{in vitro}} \times P_{\text{in vivo}})]$. $P_{\text{in vitro}}$ = probability of a given A or C being methylated in vitro if the RNA structure is random = (number of nucleotides methylated in vitro)/n, where n = total number of A's and C's in the molecule. n = 159 for group I RNA, 72 for telomerase RNA, and 55 for U2 RNA. $P_{\text{in vivo}}$ was calculated analogously, substituting "in vivo" for "in vitro" in the equation.

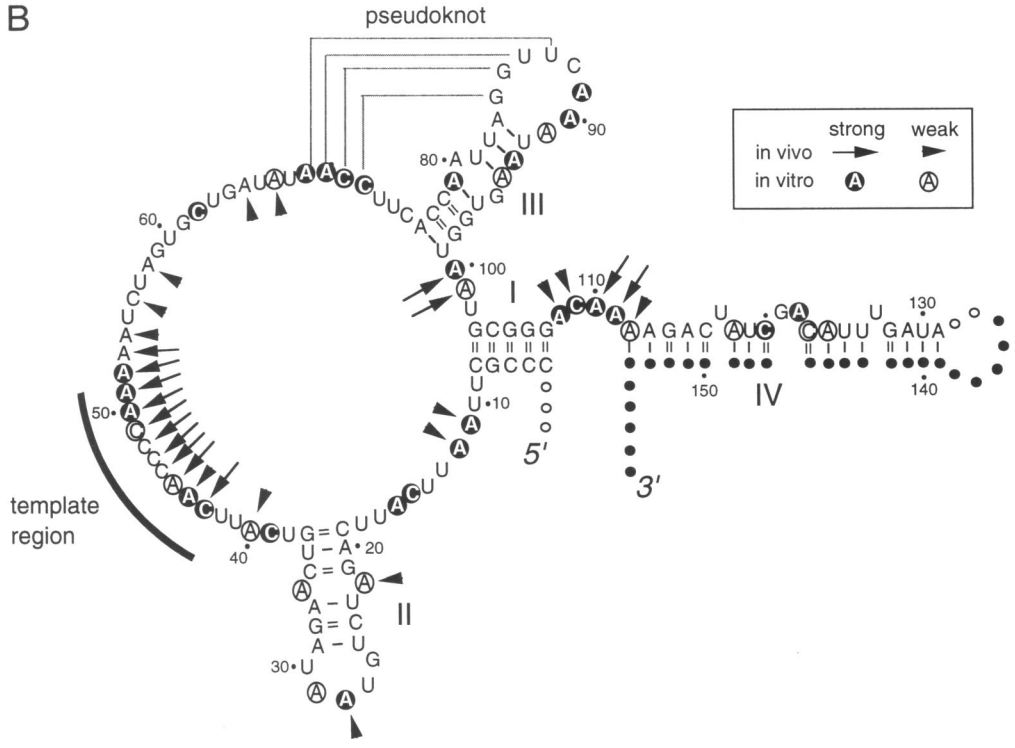
Telomerase RNA

The in vivo methylation pattern of telomerase RNA was analyzed using the same RNA samples used to analyze the group I intron. Primer extension in the presence of ddNTPs clearly revealed the known sequence of telomerase RNA (data not shown). Many of the A's and C's are protected from methylation in vivo, although the template region is strongly modified (Fig. 3A; compare lane *in vivo* with the control lane *in vivo (0)*). The occasional strong bands in the *in vivo (0)* lane (e.g., A110, U68, U57) were not observed when nuclear RNA was first purified and then methylated in vitro (data not shown). Therefore, these do not represent molecules that are normally nicked or truncated in vivo, but they could represent molecules nicked by a nuclease that was activated upon exposure to the β-mercaptoethanol used to quench the DMS prior to deproteinization. The intensities of these reverse transcriptase products were subtracted from those of the "in vivo" distribution prior to analysis, and in a few cases their intensity was great enough to prevent assignment of the state of methylation of a particular position.

In the summary in Figure 3B, A's and C's methylated in vivo are again indicated by arrows superimposed on the phylogenetic secondary structure model (Romero & Blackburn, 1991). In this case, the assignment of strongly and weakly methylated sites was based on quantitation of the amount of radioactivity per band (see figure legend), which agreed closely with visual assessment of band intensities. The structure probing data correspond very closely to the proposed secondary structure: 21 of the 28 proposed single-stranded A's were methylated, whereas only 1 of the 17 proposed

FIGURE 3. (*Facing page.*) Telomerase RNA. **A:** Structure probing in vivo and in vitro. Lanes U, G, C, A, dideoxy chain termination sequencing of an unmethylated T7 RNA polymerase transcript. Lane 0, primer extension in the absence of dideoxynucleotides. Lane *in vivo*, treatment of living *Tetrahymena* with 80 mM DMS. Lane *in vivo (0)*, stop control in which cells were exposed to DMS after the addition of excess β-mercaptoethanol. Lanes *in vitro*, T7 transcripts of telomerase RNA treated with DMS for the time indicated (min). Reverse transcriptase stops one nucleotide prior to a methylated base, so bands are offset one nucleotide from the dideoxy sequencing ladder. Black dots to right of bands indicate stops seen in the *in vivo* lane relative to *in vivo (0)* (the latter being more intensely represented in this gel). **B:** Sites of DMS modification superimposed on the secondary structure determined by comparative sequence analysis. Symbols as in Figure 1, except that strong and weak methylation sites were determined by PhosphorImager quantitation (strong, >50% of maximum intensity; weak, 20–50% of maximum). These criteria were chosen in part because they agreed with visual assessment of band intensities. Template region (C43–A51) includes both the template and the primer alignment region (Greider & Blackburn, 1989; Autexier & Greider, 1994).


A
U G C A In vivo In vitro (30) In vitro (15) In vitro (0) 0 In vivo (0)
Template
pseudoknot
Stem III
Stem I
Stem IV
C43
A54
A65
A67
A100
A101
A108
C109
A110
A111
A112
U G C A In vivo In vitro (30) In vitro (15) In vitro (0) 0 In vivo (0)
Stem II
Template
pseudoknot
Stem III
A11
A12
A22
A28
A40
C43
A54
C56
A58
A65
A67


B
pseudoknot
strong
weak
in vivo
in vitro
I
II
III
IV
5′
3′
template region

base paired A's was methylated (A112, which is at the end of stem IV and thus may undergo breathing). The methylation pattern suggests that the pseudoknot is formed in vivo: the sequence AACC at positions 69–72 and the A's in stem III are not methylated in vivo, whereas they are methylated in vitro. Alternatively or in addition, binding of proteins could directly protect this region of the RNA from methylation in vivo.

Single-stranded A's that are not methylated in vivo are likely to be involved in RNA–protein interactions or as yet unidentified RNA tertiary structure. Candidates include A16 (and the adjacent C), the A's in the hairpin loop of stem III adjacent to the pseudoknot, and bulged A122 in stem IV. Cytosines 39 and 62, methylated in vitro put protected in vivo, are also candidates for such interactions.

For comparison, the positions methylated in vitro are circled in Figure 3B. Here the correspondence with the secondary structure is marginal: 8 of the 17 proposed base-paired A's were methylated, and 4 proposed base-paired C's were methylated. (Methylation of C's is particularly revealing, because they are intrinsically less reactive than A's.) Thus, the central part of stem IV, stem III, and the pseudoknot appear to be unfolded in vitro. Direct comparison of the in vivo and in vitro methylation patterns showed 41% agreement (Table 1). This is slightly higher than the 31% correspondence expected for random superposition of two distributions with this extent of methylation.

We were concerned that the deproteinized telomerase RNA present in the mixture of all nuclear RNAs might have had its structure altered by pairing with other RNA molecules. A DNA template was therefore constructed to allow transcription of homogeneous telomerase RNA, starting at nucleotide A1, using T7 RNA polymerase. The methylation pattern of the purified T7 transcript (Fig. 3A) was essentially identical to that of telomerase RNA in the nuclear RNA mixture (data not shown). This result alleviated the concern that interaction of deproteinized telomerase RNA with other nuclear RNAs might have altered its structure.

To provide a quantitative basis for evaluating the telomerase RNA methylation data, the gels were scanned with a PhosphorImager. Background signal not attributable to methylation was subtracted from each scan, i.e., *in vitro (0)* was subtracted from *in vitro (30)* and *in vivo (0)* from *in vivo*. A portion of these corrected scans is shown in Figure 4. Concerning the in vivo methylation, we conclude: (1) the template region is methylated relatively uniformly, varying within a

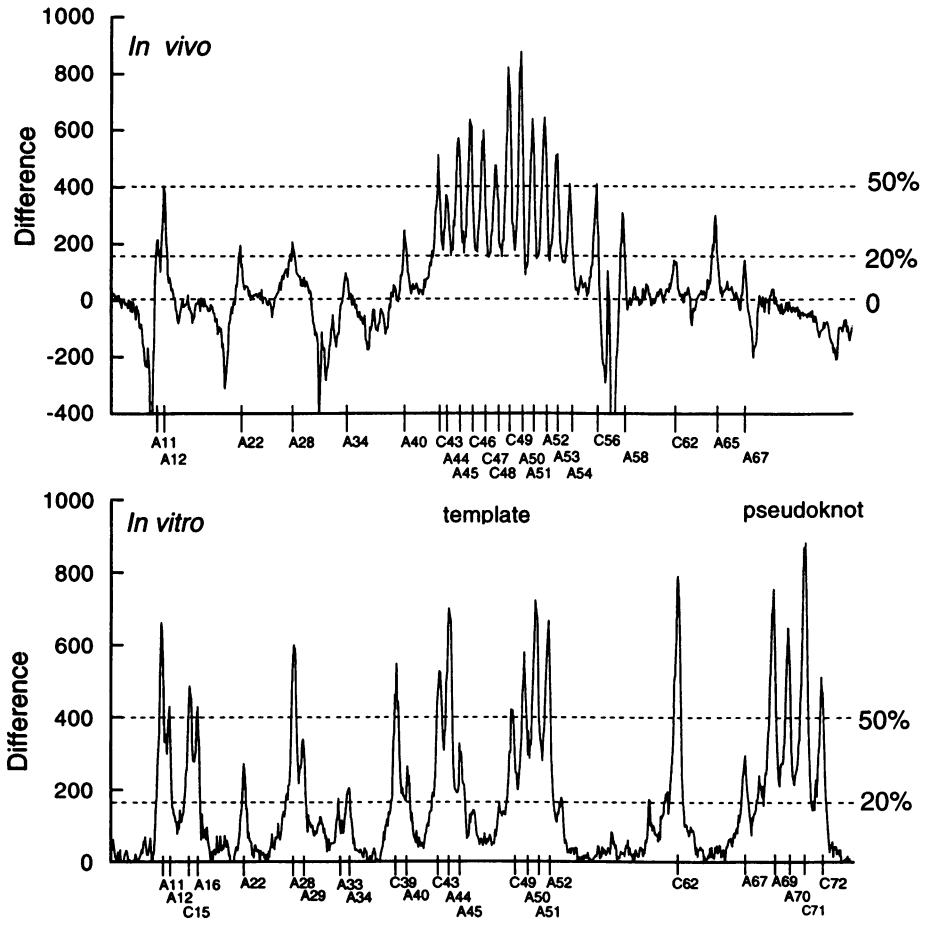


FIGURE 4. Quantitation of telomerase RNA methylation. The gel shown in Figure 3A was scanned using a PhosphorImager. Top scan, *in vivo* methylation and *in vivo (0)* control lanes were normalized to the same total number of counts per lane; Difference = *in vivo* − *in vivo (0)*. Negative peaks represent damage to the nuclear RNA when the β-mercaptoethanol stop solution was added prior to the DMS; this prevents evaluation of the in vivo methylation of these positions. Bottom scan, *in vitro (30)* methylation and *in vitro (0)* control lanes were normalized as above. Difference = *in vitro (30)* − *in vitro (0)*. Top and bottom scans are aligned vertically to facilitate comparison. Dashed lines show levels used to assign strong methylation (>50% of maximum) and weak methylation (20–50%).

twofold range for C43–A54; (2) the template region has a level of methylation as great as any part of the molecule, suggesting full accessibility; (3) the only other residues as intensely methylated as the template are A100 and A101, which lie between stems III and I, and A110 and A111, which join stems I and IV; and (4) the pseudoknot nucleotides A69–C72 have no detectable methylation (<5% that obtained in vitro). Note that the background noise level is such that it is difficult to be confident about methylation levels <20% of that of the maximum (e.g., C62 falls slightly below this standard and is not marked, and A40 falls slightly above this standard and is marked).

U2 snRNA

The in vivo methylation pattern of the U2 snRNA could be read from the 3′ region through the branch point interaction sequence (BPS) from a single primer (Fig. 5A). The methylated positions (arrows) are shown in Figure 5B, superimposed on the secondary structure of the *Tetrahymena* U2 snRNA established by comparative sequence analysis (Ørum et al., 1991; Guthrie & Patterson, 1988). As in the case of the group I intron and telomerase RNAs, the majority of the A's and five C's proposed to be in single-stranded regions were methylated in vivo, whereas none of the 11 A's in proposed duplex regions was methylated. This provides good support for the proposed secondary structure.

The Sm protein-binding site (underlined) contains two A's, A100 and A107, that are methylated in the deproteinized RNA. A107 is protected from methylation in vivo, presumably because of Sm protein binding, whereas A100 is not protected. The BPS that pairs with the branch point sequence in the pre-mRNA contains A35 and A38, both protected from methylation in vivo.

The possibility of a pseudoknot involving loop IIa paired with nucleotides 90–95 (indicated by a question mark in Fig. 5B) has been evaluated in the case of the yeast U2 snRNP by Ares and Igel (1990). They provided both genetic and in vivo chemical probing data supporting the existence of stems IIa and IIb, but contradicting the pseudoknot interaction. Our data provide evidence against the pseudoknot interaction for the majority of *Tetrahymena* U2 snRNPs because A57, A58, and A59 are methylated in vivo as they are with deproteinized RNA. We certainly cannot rule out that the pseudoknot might be formed a small fraction of the time, or that a minor subpopulation of the snRNPs might have it stably formed.

As with the previous RNAs, it is worth comparing the in vivo data with that obtained from DMS treatment of nuclear RNA after deproteinization (Table 1). They agree in 13 of 25 positions (52%), a value significantly higher than the 20% agreement expected for random superposition of two methylation patterns but significantly lower than the 87% agreement obtained with the group I RNA. The difference between in vivo and in vitro patterns has two causes: the accessibility of the Sm site (and perhaps the BPS) upon deproteination is presumably due to disruption of biologically relevant interactions, whereas the accessibility of C50 and A64 in stem IIa and of A122, A138, and A145 in stem III seems likely to result from incorrect folding of these regions after deproteinization.

DISCUSSION

The purpose of this study was to evaluate RNA structure and to identify sites of protein binding for the self-splicing rRNA intron and telomerase RNA in vivo. The approach was to methylate RNA in living *Tetrahymena*, then isolate nuclear RNA and map N-1-methyl A and N-3-methyl C by reverse transcription (Inoue & Cech, 1985; Moazed et al., 1986). For comparison, total deproteinized nuclear RNA and, in the case of telomerase RNA, a homogeneous T7 RNA polymerase transcript were methylated. The U2 snRNA provided a control for an RNA known to exist as an RNP in vivo.

Group I intron

The reason for suspecting protein complexation with the pre-rRNA intron concerns the rate of splicing, which is ~20-fold faster in vivo than under optimal conditions in vitro (compare Brehm & Cech, 1983; Cech & Bass, 1986). Zhang et al. (1995) see a similar rate enhancement when the *Tetrahymena* intron is expressed as part of the 23S rRNA in *Escherichia coli*, suggesting that the factors responsible are not species specific. Considering that a number of group I introns, even self-splicing ones, have protein facilitators in vivo (Gampel et al., 1989; reviewed by Cech, 1990 and by Lambowitz & Perlman, 1990), it seemed quite possible that we would find evidence of protein binding. Instead, we found no sites that were available to methylation in vitro but protected in vivo. This negative result is made more convincing by the positive evidence for protein binding seen with the telomerase and U2 RNAs.

How, then, is the self-splicing rate enhanced in vivo? The protein hypothesis has not been eliminated. It seems possible that a protein could be bound to the intron in vivo but protect no A residues, or could be bound only transiently and therefore difficult to detect in the steady-state population of RNA. Alternatively, the RNA might fold into a similar but more active structure in vivo (note there were three nucleotides more strongly modified in vivo than in vitro, including A115, which has been implicated in positioning the splice site for reaction; Wang et al., 1993). As another alternative, our "optimal" in vitro conditions might have overlooked some small molecule present in the cell that facilitates splicing, such as an unusual polyamine not yet

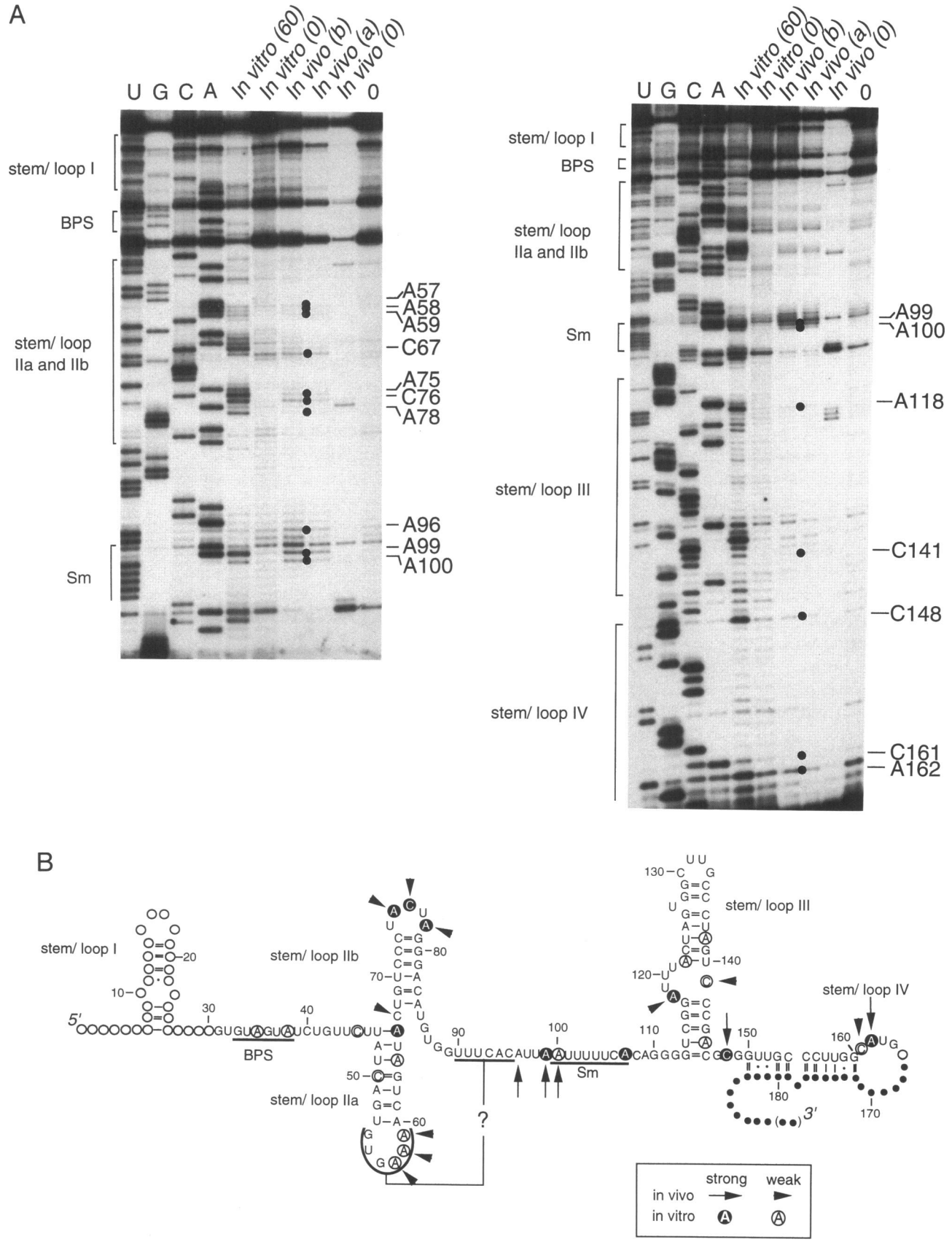


FIGURE 5. U2 snRNA. **A:** Structure probing in vivo and in vitro. Lanes marked as in Figure 1. Dideoxynucleotide sequence and in vitro methylation were performed with deproteinized total nuclear RNA. **B:** Sites of DMS modification superimposed on the secondary structure model of Ørum et al. (1991). Symbols as in Figure 2. Question mark represents a possible pseudoknot interaction contraindicated in yeast U2 snRNA by the data of Ares and Igel (1990).

tested. None of these possibilities seems very plausible, making the mystery of the higher in vivo rate all the more intriguing.

There is striking agreement between the in vitro and in vivo patterns of methylation of the intron. When the in vitro pattern is compared with that determined previously for the circular form of the intron (Inoue & Cech, 1985), there is 91% agreement (62 of 68 A's and C's; the C's were counted only when they were reactive). This agreement seems excellent considering that different forms of the intron were analyzed and that different solution conditions were employed (same $MgCl_2$ concentration, but Inoue and Cech added 58 mM NaCl and methylated at 30 °C rather than room temperature; see Jaeger et al. [1990] for data at other solution conditions).

U2 snRNP

Our in vivo data on the *Tetrahymena* U2 RNA can be compared to the in vivo DMS experiments on yeast U2 reported by Ares and Igel (1990). The methylation patterns are substantially similar, with reaction in the loops of stems IIa and IIb supporting their existence and providing an argument against a pseudoknot involving the IIa loop. The Sm protein-binding sites are difficult to compare; uracil is unreactive with DMS, and the sequences flanking the run of U's differ between yeast and *Tetrahymena*. We interpret A107 as interacting with an Sm protein, because it is methylated in deproteinized RNA and shows no methylation in vivo. Unfortunately, the corresponding nucleotide in yeast U2 is a G, preventing comparison. The sequence GUAGUA that pairs with the pre-mRNA branch site (Parker et al., 1987) is present in the same location in *Tetrahymena* and yeast U2 snRNA. In yeast, both A's are accessible, indicating that the bulk of U2 is not base paired to introns (Ares & Igel, 1990). In contrast, in *Tetrahymena*, the A's are both methylated in deproteinized RNA but protected from methylation in vivo. We do not know the source of this protection and certainly think it would be premature to conclude that most of the U2 snRNP is paired with pre-mRNA in *Tetrahymena*.

Telomerase RNA

The pattern of methylation of A's and C's for telomerase RNA in vivo gives strong support for the secondary structure model of Romero and Blackburn (1991) as extended by ten Dam et al. (1991), who added the pseudoknot. There is only one discrepancy—A112 is weakly methylated and shown as base paired in the model, although it does occur at the terminus of a duplex region. Thus, this structure model (Fig. 3B) appears to be about as good as that of the well-studied group I intron, where there were two A's methylated in vivo that are shown as base paired.

We also conclude that, in rapidly dividing cells, the template region of telomerase RNA is not telomere-bound. Given that the *Tetrahymena* macronucleus contains 18,000–40,000 molecules of telomerase and ~40,000 telomeres (Avilion et al., 1992), it was conceivable that most of the telomerase could have been engaged with telomeres in the steady state. We have recently shown by in situ hybridization that in the *Oxytricha nova* and *Euplotes aediculatus* macronuclei, most of the telomerase is sequestered away from telomeres (Fang & Cech, 1995). If a similar situation pertains in *Tetrahymena*, it would be predictable that the template region of telomerase RNA would be accessible.

Lingner et al. (1994) suggested that the pseudoknot might alternate between two conformations during the primer elongation–translocation cycle. The present data are consistent with the specific model where the pseudoknot is formed when the template is unoccupied. If the suggestion of dynamic pseudoknot is correct, the pseudoknot would then convert to the single stem-loop III upon primer binding or elongation.

Telomerase is an RNP, containing >150 kDa of protein in *Tetrahymena* (Greider & Blackburn, 1987). The location of protein subunits relative to the RNA is of considerable interest. As described in the Results, our methylation analysis implicates protein binding for the hairpin loop of stem III adjacent to the pseudoknot, the bulge around A122 in stem IV, and three sites on the central wheel: C15–A16, C39, and C62. C39 is noteworthy because it resides in the (U)GUCA sequence that precedes the template by two bases and is conserved among phylogenetically distant telomerase RNAs (Lingner et al., 1994; McCormick-Graham & Romero, 1995). These conclusions about protein binding are based on comparison of the methylation data with the structure derived from comparative analysis; by themselves, our data cannot differentiate between direct protein binding and protein-induced RNA structure formation. In addition, we suggest that the template region is likely to be in close proximity to protein in such a manner that its base pairing face is accessible. This conclusion derives from the observation that C46, C47, C48, and C49 are highly methylated in vivo and methylated more than they are in vitro (Fig. 4). This is most easily understood by there being a protein environment that increases the local concentration of DMS or enhances the modification reaction.

Telomerase RNA folds differently in vitro than in vivo

The structure of *Tetrahymena* telomerase RNA in vitro has recently been probed by diethylpyrocarbonate (DEPC) and by nucleases (Bhattacharyya & Blackburn, 1994). Although DEPC probes the N-7 of A, whereas DMS probes its base pairing face, previous studies with the *Tetrahymena* circular IVS (Inoue & Cech, 1985) in-

dicate that they tend to correspond in their identification of single-stranded nucleotides. Our in vitro DMS data are in striking correlation with the DEPC data of Bhattacharyya and Blackburn (1994). All of our major sites (Fig. 3B) are marked on their Figure 1C, and although the relative intensities do not always agree, this may be expected because of the different chemistry of DMS and DEPC reaction.

The correspondence between our in vitro and in vivo methylation patterns is only 41%, which is somewhat higher than the 31% correspondence expected for two random distributions (Table 1). Two types of sites contribute to the low extent of agreement: (1) The 21 nucleotides methylated in vitro but not in vivo can be explained either as the removal of a protein "footprint" or as RNA structural changes occurring upon protein removal. Some of these nucleotides are in the pseudoknot (including stem III), structures that are apparently not stably formed in vitro (see also Bhattacharyya & Blackburn, 1994). (2) The eight nucleotides methylated only in vivo are most simply interpreted as reflecting differences in the RNA structures. More specifically, a portion of the template is methylated only in vivo, presumably because of misfolding in vitro; perhaps it undergoes fortuitous interactions with other nucleotides that are sequestered by bound protein in the RNP. Misfolding of RNA in vitro is not unusual (Uhlenbeck, 1995).

Nevertheless, the incorrect structure of the deproteinized RNA leads to some intriguing questions. Might the low efficiency of reconstitution of active telomerase (Autexier & Greider, 1994) be due to the low proportion of correctly folded RNA? Although incorrectly folded RNA would provide an energetic barrier to protein–RNA complexation, does the RNP nevertheless form spontaneously? Or is it assisted in vivo by RNA chaperones?

One of the tenets of comparative sequence analysis of RNA is that the derived structure should be that which pertains in vivo (e.g., see Fig. 10 of Moazed et al., 1986). The current study provides another strong example. Chemical modification of the RNA in vitro gave little correlation with the proposed secondary structure: 8 of the 17 proposed base paired A's were methylated, approaching the number of 10 that would result from the same number of methyl groups being randomly distributed among the A's. In contrast, chemical modification in vivo modified only 1 of the 17 proposed base-paired A's, and it was at a helix terminus. Thus, the structure determined by Romero and Blackburn (1991) by comparative analysis represents the structure of telomerase RNA in vivo, as it should.

MATERIALS AND METHODS

Growth of cells

Tetrahymena thermophila strain B7 was grown in 1% proteose peptone (Difco) and 0.003% sequestrine (an iron supplement, Ciba-Geigy). One-liter cultures were grown in 2.8-L Fernbach flasks at 30 °C with gentle shaking. Cells were grown to a density of 1.0×10^5 cells/mL, then collected by centrifugation for 10 min at 3,000 rpm ($1,000 \times g$) in a Beckman JA-10 rotor.

Isolation of nuclear RNA

The general method followed that of Higashinakagawa et al. (1975). All steps were performed at 4 °C or on ice. Cell pellets were resuspended in 90 mL TMS buffer (0.01 M Tris-HCl, pH 7.5, 0.01 M $MgCl_2$, 0.003 M $CaCl_2$, 0.25 M sucrose) on ice in a 250-mL siliconized beaker. The cells were stirred by means of a stainless steel stirring rod attached to an overhead motor and lysed by the addition of freshly prepared 1% NP-40 in TMS to give a final NP-40 concentration of 0.16%. After 20 min of stirring, 0.815 g sucrose was added for each 1 mL solution, and stirring continued until the sucrose dissolved (about 20 min). Nuclei were pelleted ($9,000 \times g$, 30 min) in a JS 13 rotor (Beckman), washed in TMS, and resuspended in 0.5 mL H_2O. RNA was extracted by adding 1 mL TRI reagent (Molecular Research Center, Inc.) to resuspended nuclei with vortexing. After the addition of 200 μL of SEVAG (chloroform:isoamyl alcohol, 24:1), the organic and aqueous phases were separated by centrifugation for 10 min at 4 °C. Following the addition of 750 μL isopropanol to the aqueous phase, the RNA was pelleted by centrifugation, washed with 70% ethanol, resuspended in 50 μL H_2O, and quantitated spectrophotometrically.

Treatment of cells with DMS

One-liter cultures of *Tetrahymena thermophila* were grown and harvested as described above and resuspended in fresh room-temperature TMS to a final volume of 4 mL. After the addition of 15–60 μL of DMS (Sigma; 40–160 mM final concentration), the cell suspension was gently rocked for 2 min at room temperature. β-Mercaptoethanol (200 μL of 14.3 M) was added to the cells to quench the DMS (final concentration of 0.7 M). Nuclear RNA was isolated as above using TMS supplemented with 2 mM β-mercaptoethanol. Mock-treated cells had the β-mercaptoethanol quench prior to the addition of DMS, and RNA was isolated as described for DMS-treated cells.

DMS treatment of deproteinized nuclear RNA or T7 RNA transcripts

In a typical DMS treatment, 3–6 μg of nuclear RNA was added to 300 μL of 30 mM Hepes, pH 7.5, and 10 mM $MgCl_2$. The RNA was renatured by incubation at 50 °C for 5–10 min, then allowed to cool to room temperature. After this preincubation, a 100-μL portion was removed and added to 50 μL of "stop mix" consisting of 0.5 M β-mercaptoethanol and 0.75 M NaOAc, pH 5.5. To the remaining 200 μL, 2 μL of diluted DMS (2 μL DMS diluted with 5 μL absolute ethanol) was added and mixed vigorously, giving a final DMS concentration of 42 mM, and incubated at room temperature. At various times (usually 15, 30, and 60 min), 100-μL portions were removed and added to 50 μL of stop mix to quench the DMS. The modified nuclear RNA was precipitated with 3 vol-

umes of ethanol, washed with 70% ethanol, and dissolved in 16 μL H_2O.

Mapping modified adenines and cytosines by primer extension

^{32}P-end-labeled DNA primers for telomerase RNA (TT110-26 and TT134-26), U2 RNA (U2 166-23), and the pre-rRNA intron (IP67-17, IP126-25, IP274-22, and IP341-22) were annealed to DMS-treated RNA in 0.05 M Tris HCl, pH 8.3, 0.06 M NaCl, and 0.01 M DTT by boiling for 1 min and then cooling on ice or at room temperature. (Primer nomenclature: first number indicates nucleotide in the RNA to which the 3′ end of the primer hybridizes, and second number indicates length of primer in nucleotides.) Reverse transcription was initiated by the addition of dNTPs (Pharmacia/LKB) and AMV reverse transcriptase (Life Sciences) at room temperature in 0.05 M Tris-HCl, pH 8.3, 0.006 M MgOAc, 0.06 M NaCl, and 0.01 M DTT, followed by incubation for 15 min at 60 °C for telomerase RNA or 50 °C for U2 and IVS RNA. (The higher temperature was required to inhibit primers from annealing to other RNAs.) In addition, a sample of nuclear RNA (not treated with DMS) or an in vitro T7 RNA polymerase transcript was sequenced using the same primers. Transcription was stopped by the addition of stop mix (95% formamide, 0.1× TBE, 0.025% xylene cyanol, and 0.025% bromophenol blue). Products were run on 8% polyacrylamide (acrylamide: bis-acrylamide, 29:1)/8 M urea gels in 1× TBE and analyzed by autoradiography or PhosphorImager (Molecular Dynamics) scanning.

Construction of template for transcription of telomerase RNA by T7 RNA polymerase

PCR was done in 100-μL reactions containing 10 mM Tris-HCl, pH 8.3, 4 mM $MgCl_2$, 50 mM KCl, 0.2 mM each dNTP, and 2.5 units *Taq* polymerase (Boehringer-Mannheim), using genomic DNA from *T. thermophila* embedded in agarose (Conover & Brunk, 1986) and primers based on the published sequence of telomerase RNA (Greider & Blackburn, 1989). The 5′ primer encoded an *Eco*R I restriction endonuclease site for cloning, a promoter for T7 RNA polymerase, and a hammerhead ribozyme self-cleavage domain to process the 5′ end of the RNA. The 3′ primer encoded an *Ear* I site for termination of T7 RNA transcription and a *Bam*H I site for cloning. DNA was denatured for 4 min at 94 °C and then amplified in 30 cycles, each involving denaturation for 1 min at 94 °C and annealing and polymerization for 2 min at 65 °C. Reactions were finished with a final polymerization step for 10 min at 65 °C and then held at 4 °C. Products were analyzed on a 4% NuSieve (FMC Corporation) agarose gel. A band corresponding to the proper size was purified and cloned into pUC-19. Clones were screened by restriction endonuclease digestion and sequencing (USB Sequenase kits).

ACKNOWLEDGMENTS

We thank J. Lingner, A. Bhattacharyya, and E. Blackburn for comments on the manuscript and A. Sirimarco for manuscript preparation. A.J.Z. is a Senior Associate and T.R.C. is an Investigator of the Howard Hughes Medical Institute. We thank the W.M. Keck Foundation for their generous support of RNA research at the University of Colorado at Boulder.

Received April 5, 1995; returned for revision April 28, 1995; revised manuscript received May 8, 1995

REFERENCES

Ares M Jr, Igel AH. 1990. Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. *Genes & Dev 4*:2132–2145.

Autexier C, Greider CW. 1994. Functional reconstitution of wild-type and mutant *Tetrahymena* telomerase. *Genes & Dev 8*:563–575.

Avilion AA, Harrington LA, Greider CW. 1992. *Tetrahymena* telomerase RNA levels increase during macronuclear development. *Dev Genet 13*:80–86.

Banerjee AR, Jaeger JA, Turner DH. 1993. Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. *Biochemistry 32*:153–163.

Been MD, Barfod ET, Burke JM, Price JV, Tanner NK, Zaug AJ, Cech TR. 1987. Structures involved in *Tetrahymena* rRNA self-splicing and RNA enzyme activity. *Cold Spring Harbor Symp Quant Biol 52*: 147–157.

Bhattacharyya A, Blackburn EH. 1994. Architecture of telomerase RNA. *EMBO J 13*:5721–5731.

Brehm SL, Cech TR. 1983. Fate of an intervening sequence ribonucleic acid: Excision and cyclization of the *Tetrahymena* ribosomal RNA intervening sequence in vivo. *Biochemistry 22*:2390–2397.

Burke JM, Esherick JS, Burfeind WR, King JL. 1990. A 3′ splice site-binding sequence in the catalytic core of a group I intron. *Nature 344*:80–82.

Cech TR. 1990. Self-splicing of group I introns. *Annu Rev Biochem 59*:543–568.

Cech TR, Bass BL. 1986. Biological catalysis by RNA. *Annu Rev Biochem 55*:599–629.

Cech TR, Damberger S, Gutell RR. 1994. Representation of the secondary and tertiary structure of group I introns. *Nature Struct Biol 1*:273–280.

Climie SC, Friesen JD. 1988. In vivo and in vitro structural analysis of the *rplJ* mRNA leader of *Escherichia coli*. *J Biol Chem 263*:15166–15175.

Coetzee T, Herschlag D, Belfort M. 1994. *Escherichia coli* proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes & Dev 8*:1575–1588.

Conover RK, Brunk CF. 1986. Macronuclear DNA molecules of *Tetrahymena thermophila*. *Mol Cell Biol 6*:900–905.

Ephrussi A, Church GM, Tonegawa S, Gilbert W. 1985. B lineage-specific enhancer with cellular factors in vivo. *Science 227*:134–139.

Fang G, Cech TR. 1995. Telomerase RNA localized in the replication band and spherical subnuclear organelles in hypotrichous ciliates. *J Cell Biol*. Forthcoming.

Gampel A, Nishikimi ME, Tzagoloff A. 1989. CBP2 protein promotes in vitro excision of a yeast mitochondrial group I intron. *Mol Cell Biol 9*:5424–5433.

Greider CW, Blackburn EH. 1987. The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell 51*:887–898.

Greider CW, Blackburn EH. 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature 337*:331–337.

Guthrie C, Patterson B. 1988. Spliceosomal snRNAs. *Annu Rev Genet 23*:387–419.

Herschlag D. 1995. RNA chaperones and the RNA folding problem. *J Biol Chem*. Forthcoming.

Higashinakagawa T, Tashiro F, Mita T. 1975. DNA-dependent RNA polymerase from a protozoan, *Tetrahymena pyriformis*. *J Biochem 77*:783–793.

Inoue T, Cech TR. 1985. Secondary structure of the circular form of the *Tetrahymena* ribosomal RNA intervening sequence: A technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci USA 82*:648–652.

Jaeger JA, Zuker M, Turner DH. 1990. Melting and chemical modification of a cyclized self-splicing group I intron: Similarity of structures in 1 M Na^+, in 10 mM Mg^{2+}, and in the presence of substrate. *Biochemistry 29*:10147–10158.

Jaeger L, Westhof E, Michel F. 1993. Monitoring of the cooperative unfolding of the *sunY* group I intron of bacteriophage T4. *J Mol Biol 234*:331–346.

Lambowitz AM, Perlman PS. 1990. Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem Sci 15*:440–444.

Lingner J, Hendrick LL, Cech TR. 1994. Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. *Genes & Dev 8*:1984–1998.

McCormick-Graham M, Romero DP. 1995. Ciliate telomerase RNA structural features. *Nucleic Acids Res 23*:1091–1097.

Michel F, Netter P, Xu MQ, Shub DA. 1990. Mechanism of 3′ splice site selection by the catalytic core of the *sunY* intron of bacteriophage T4: The role of a novel base-pairing interaction in group I introns. *Genes & Dev 4*:777–788.

Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol 216*:585–610.

Moazed D, Stern S, Noller HF. 1986. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J Mol Biol 187*:399–416.

Murphy FL, Cech TR. 1994. GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J Mol Biol 236*:49–63.

Nick H, Gilbert W. 1985. Detection in vivo of protein-DNA interactions within the *lac* operon of *Escherichia coli*. *Nature 313*:795–798.

Ørum H, Nielsen H, Engberg J. 1991. Spliceosomal small nuclear RNAs of *Tetrahymena thermophila* and some possible snRNA-snRNA base-pairing interactions. *J Mol Biol 222*:910–932.

Parker R, Siliciano PG, Guthrie C. 1987. Recognition of the TACT AAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell 49*:229–239.

Peattie DA, Gilbert W. 1980. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci USA 77*:4679–4682.

Price CM, Cech TR. 1987. Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. *Genes & Dev 1*:783–793.

Romero DP, Blackburn EH. 1991. A conserved secondary structure for telomerase RNA. *Cell 67*:343–353.

ten Dam E, van Belkum A, Pleij K. 1991. A conserved pseudoknot in telomerase RNA. *Nucleic Acids Res 19*:6951.

Uhlenbeck OC. 1995. Keeping RNA happy. *RNA 1*:4–6.

Wang JF, Downs WD, Cech TR. 1993. Movement of the guide sequence during RNA catalysis by a group I ribozyme. *Science 260*:504–508.

Weeks KW, Cech TR. 1995. Efficient protein-facilitated splicing of the yeast mitochondrial bI5 intron. *Biochemistry 34*:7728–7738.

Yu GL, Bradley JD, Attardi LD, Blackburn EH. 1990. In vivo alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature 344*:126–132.

Zhang F, Ramsay ES, Woodson SA. 1995. In vivo facilitation of *Tetrahymena* group I intron splicing in *Escherichia coli* pre-ribosomal RNA. *RNA 1*:284–292.