RNA (2002), *8*:601–611. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S135583820202071X

Identification of an essential pseudoknot in the putative downstream internal ribosome entry site in giardiavirus transcript

SRINIVAS GARLAPATI and CHING C. WANG

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, USA

ABSTRACT

Enhanced translation of giardiavirus-luciferase chimeric mRNA in *Giardia lamblia* requires the initial 264-nt viral capsid coding region as a putative internal ribosomal entry site (IRES). Essential structural elements in this site include (1) a downstream box (DB) complementary to the anti-DB at the 3′ end of 16S-like rRNA, (2) stem-loops I, II, III, and IVA, and (3) a pentanucleotide 5′-UCUCC-3′ immediately downstream from stem loop IVA. A search for the structural role of the pentanucleotide suggested that it may form a pseudoknot with another pentanucleotide 5′-GGAGA-3′ in loop II. Alteration of the two pentanucleotides by site-directed mutagenesis resulted in a drastic reduction in translation of the transcript. But the loss was recovered by compensatory changes in the two sequences, suggesting Watson–Crick base pairings between them. Results from in vitro enzymatic and chemical structural probing supported the presence of such a pseudoknot 143 nt downstream from the initiation codon. Minor repositioning of this codon led invariably to a complete loss of translation, suggesting that the initiation site is confined within a rigid position defined by all the structural elements in the IRES including the pseudoknot. This is the first pseudoknot of its kind shown to play an important role in a downstream IRES of a viral transcript. The finding is particularly interesting because it could reflect a unique feature of translation initiation in *Giardia*, which is known to have exceedingly short (1–6 nt) 5′ untranslated regions in its mRNAs.

Keywords: giardiavirus (GLV); internal ribosome entry site (IRES); pseudoknot; RNA structure probing

INTRODUCTION

Giardiavirus (GLV) is a double-stranded (ds)RNA virus of the *Totiviridiae* family, which specifically infects trophozoites of the most primitive protozoan parasite *Giardia lambila* (Wang & Wang, 1986). Its dsRNA genome of 6277 bp encodes two polypeptides; a major 100-kDa capsid protein (Gag) and a minor 190-kDa fusion protein (Gag-Pol) via a −1 ribosomal frameshift (Wang et al., 1993; Li et al., 2001). The coding region in GLV (+)-strand RNA is flanked by a 367-nt 5′ untranslated region (UTR) and a 301-nt 3′ UTR . The ability of purified GLV to infect, and the capability of its (+)-stranded RNA to transfect *G. lamblia* trophozoites, resulting in intracellular proliferation of infectious GLV particles, are the major distinguishing features of this virus among the totiviruses (Wang & Wang, 1986). It has turned the GLV (+)-strand RNA into an effective transfection vector for high-level expression of foreign mRNAs in GLV-infected *G. lamblia* (Yu et al., 1995; Yu & Wang, 1996).

Reprint requests to: Prof. Ching C. Wang, Department of Pharmaceutical Chemistry, University of California–San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0446, USA; e-mail: ccwang@cgl.ucsf.edu.

In vitro-transcribed chimeric mRNA containing a full-length firefly luciferase transcript flanked by the 367-nt GLV 5′ UTR and a 2022-nt 3′ terminus of GLV (+)-strand RNA can be introduced into GLV-infected *G. lamblia* trophozoites via electroporation (Yu et al., 1995). The chimeric mRNA thus introduced undergoes vigorous replication and transcription but only a basal level of translation, resulting in an expressed luciferase activity barely above the background level (Yu et al., 1995; Yu & Wang, 1996). This relatively poor translation efficiency was, however, enhanced by 5,000-fold when the initial 264 nt of the capsid-coding region in GLV mRNA were fused in frame with, and upstream from, the luciferase mRNA (Yu & Wang, 1996).

In our earlier studies, we have identified a 13-nt downstream box (DB) sequence within the initial 264-nt capsid coding region of GLV mRNA at position 66–78 that complements a 15-nt sequence (with two gaps) between nt 1382 and 1396 in the V9 region near the 3′ end of the 16S-like ribosomal RNA (rRNA) of *G. lamblia* (Yu et al., 1998). Deletion or scrambling of this DB sequence led to a significant loss of translation efficiency. However, although DB is located 66–78 nt downstream of the start codon, inclusion of the first 98 nt

of the downstream region encompassing the entire DB exerted little enhancement of translation of the chimeric transcript (Yu & Wang, 1996). It was the increment of coding region from nt 111 to 264 that resulted in an exponential increase of translation efficiency up to 5,000-fold (Yu & Wang, 1996).

The entire 264-nt segment of GLV mRNA has since been thoroughly analyzed for additional structural and sequence-specific elements that may contribute to the enhanced translation. Results from chemical probing and mutational analysis of the MFOLD-predicted stem-loops I (nt 11–35), II (nt 144–164), and III (166–182) verified their presence in the RNA molecule and indicated their essential involvement in enhanced translation (Garlapati et al., 2001). Similar experiments also demonstrated the presence and involvement of another stem-loop, IVA, in translation enhancement that was not predicted by MFOLD (Garlapati et al., 2001). A pentanucleotide sequence, 5′-UCUCC-3′, immediately downstream from stem-loop IVA was also found essential for translation enhancement; alteration of its sequence invariably leads to a significantly reduced translation (Garlapati et al., 2001). Data from chemical probing of semidenatured RNA indicated that the three C residues in the pentanucleotide are apparently involved in Watson–Crick base pairings. A search for a complementary pentanucleotide sequence within the 264-nt RNA fragment revealed another sequence, 5′-GGAGA-3′, in the loop region of stem-loop II that could form a canonical Watson–Crick base pairing with it. This complementary sequence is 58 nt upstream from the first pentanucleotide and their interaction will result in a tertiary structure known as a pseudoknot. Here, we present data from an extensive mutational analysis coupled with evidence from enzymatic and chemical probing of the RNA structure under more native conditions to demonstrate the existence of such an essential structure for enhanced translation. The highly complex secondary and tertiary structures of this RNA molecule resulting from this additional "pseudoknot" suggested a highly rigid downstream IRES-like structure for recruiting the small ribosomal subunit and precise positioning of the start codon for translation initiation. Mutational analysis indicated that the initiation codon could not be moved at all without precipitously decreasing translation efficiency, and thus supports such a postulation.

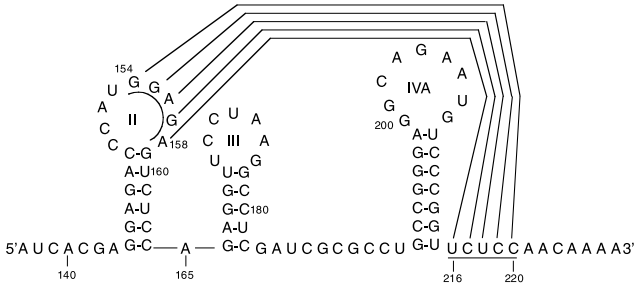

FIGURE 1. The region (nt 137–227) within GLV IRES encompassing the structural (stem-loops II, III, and IVA) and sequences (pentanucleotides nt 154–158 and 216–220, underlined) elements participating in the formation of a putative pseudoknot complex. The five Watson–Crick base pairings in the predicted pseudoknot stem are connected with lines.

RESULTS

Identification by site-directed mutations of a putative pseudoknot structure in GLV mRNA that is essential for translation enhancement

In our previous investigation, we observed that any alteration of the sequence 5′-UCUCC-3′ (nt 216–220, Fig. 1) in the 264-nt capsid-encoding region of GLV mRNA resulted in a drastic reduction in luciferase expression in the transfected *Giardia* (Garlapati et al., 2001). The loss of activity in two of the previously analyzed mutants, U216A/C217A and C219U/C220G, was not due to disruption of any identifiable stem-loop structure. It was due to alteration of the specific sequence, suggesting that the latter is probably involved in base pairing with another sequence in the transcript (Garlapati et al., 2001). A search in the 264-nt region by visual inspection revealed a sequence, 5′-GGAGA-3′, in the loop of stem-loop II (nt 154–158) that could form a Watson–Crick base pairing with the sequence 5′-UCUCC-3′ (Fig. 1). To test this possibility, a compensatory mutation was introduced at G157U/A158U in the mutant U216A/C217A. The double mutation resulted in a significant recovery of the luciferase activity from 4.9% to 41.0% of the wild-type level (Table 1). Compensatory mutations were also introduced at G154C/G155A to determine if they could complement the mutations C219U/C220G and C219U/C220G/C229A/A230C (Table 1). The luciferase activity in mutant C219U/C220G was restored from 2.5% to 55.0%, whereas a 90% recovery of luciferase activity was achieved in mutant C219/C220G/C229A/A230C (Table 1). The mere partial recovery in the former case could be due to a competition between the unaltered C229/A230 and the altered C154/A155 in compensatory mutation to base pair with U219/G220, whereas this competition was absent in the latter case and thus resulted in a much enhanced recovery. Two double substitutions, U216A/C217G and G157C/A158U, were shown to each result in a drastic loss of luciferase activity to 3.8% and 4.2% of the wild type (Table 1). However, when the two double substitutions were combined, luciferase expression was restored to 76% of the wild type (Table 1). Two other double substitutions, C219A/C220G and G155C/G156U, each exhibiting 5.0% and 4.1% luciferase expression, demonstrated 61% of the wild type when combined (Table 1). Similarly, two triple mutants, U216A/C217G/U218A and A156U/G157C/A158U, each expressing luciferase activity at 3.7% and

TABLE 1. Relative luciferase activities expressed by various pseudoknot mutant transcript-transfected *Giardia.*[a]

	Nt 154–158	Nt 216-220	RLU(%)[b]
Wild type	GGAGA	UCUCC	100
	GGAGA	**a a**UCC	4.9 ± 0.7
	GGA**u u**	UCUCC	10.6 ± 1.9
	GGA**u u**	**a a**UCC	41.2 ± 5.3
	GGAGA	UCU**u g**	2.5 ± 0.2
	c a AGA	UCUCC	8.3 ± 0.2
	c a AGA	UCU**u g**	54.6 ± 0.8
	c a AGA	UCU**u g**[c]	90.0 ± 18.4
	GGAGA	**a g**UCC	3.8 ± 0.9
	GGA**c u**	UCUCC	4.2 ± 0.1
	GGA**c u**	**a g**UCC	76.3 ± 5.3
	GGAGA	**a g a**CC	3.7 ± 1.9
	GG**u c u**	UCUCC	4.2 ± 0.1
	GG**u c u**	**a g a**CC	63.5 ± 7.6
	GGAGA	UCU**a g**	5.0 ± 0.9
	c u AGA	UCUCC	4.1 ± 0.4
	c u AGA	UCU**a g**	61.3 ± 0.5
	GGAGA	UCUC**g**	10.2 ± 7.0
	c GAGA	UCUCC	9.1 ± 6.5
	c GAGA	UCUC**g**	78.3 ± 8.4
	UGGAGA	UCUC**g u**	7.7 ± 0.9
	a c GAGA	UCUCCA	11.4 ± 4.3
	a c GAGA	UCUC**g u**	62.0 ± 10.3
	GGAGA	**g g a g a**	2.8 ± 0.3
	u c u c c	UCUCC	4.2 ± 1.0
	u c u c c	**g g a g a**	96.5 ± 16.1
	GGAGA	**g g a c a**	5.8 ± 0.3
	u g u c c	UCUCC	5.0 ± 0.6
	u g u c c	**g g a c a**	63.7 ± 2.4

[a]The letters in lower case and bold are the substituted bases.
[b]Each value represents an average from at least two independent transfection experiments with the wild type (pC631luc) as positive control (100%). Wild type and mutants were each used to transfect cells in triplicates in each experiment.
[c]With an additional mutation C229A/A230C to eliminate the potential competition between C229/A230 and C154/A155 to base pair with U219/G220.

4.3% of the wild type, had the expression restored to 63% in combination (Table 1). A conversion of 5′-UCUCC-3′ to 5′-GGAGA-3′ and vice versa each resulted in a drastically reduced luciferase expression to 2.8% and 4.2% (Table 1). But a combination of the two changes, resulting in a swapping of the two pentanucleotides, restored the luciferase activity to 96.5% of the wild type (Table 1). Finally, substitutions of 3′-UCUCC-3′ by 5′-GGACA-3′ and 5′-GGAGA-3′ by 5′-UGUCC-3′, each resulting in 4.7% and 6.0% luciferase expression, were restored to 63% expression in combination (Table 1). There is thus little doubt that a mutation that leads to a loss of any of the five Watson–Crick base pairings between the two pentanucleotides invariably causes a precipitous decrease in translation efficiency, whereas a restoration of the five base pairings regardless of sequence changes always brings back the wild-type efficiency. The results supported presence of the postulated pseudoknot structure and also illustrated its functional importance within the putative IRES.

The tertiary hydrogen bonding between two pentanucleotides induces formation of a pseudoknot structure. According to the original pseudoknot nomenclature (ten Dam et al., 1990, 1992; Hilbers et al., 1998), this putative pseudoknot in our hands is similar to the classical H-type pseudoknot with only minor variations. By the classical H-type pseudoknot structure, the stem in stem-loop II could be considered as the pseudoknot stem 1, whereas the stem structure generated by pairing the two pentanucleotides could be regarded as pseudoknot stem 2 (Fig. 6). The part of loop II sequence immediately upstream from the pentanucleotide 5′-GGAGA-3′ could be taken as pseudoknot loop 1, whereas pseudoknot loop 2 spans 58 nt, including stem-loops III and IVA separated by an 10-nt single-stranded region (Fig. 6). For the ease of discussion, the stem formed by the two pentanucleotides will be referred to as the pseudoknot stem and the rest of the stem structures will remain as previously designated (Garlapati et al., 2001).

Another possible explanation, that the two pentanucleotides may have formed a “kissing complex,” is less likely, because the latter has been defined as a base-pair formation between two hairpin loops (Chang & Tinoco, 1994). Whereas 5′-GGAGA-3′ constitutes a part of the loop in stem-loop II, 5′-UCUCC-3′ is located in a single-stranded region immediately downstream from stem-loop IVA (Garlapati et al., 2001). A pseudoknot will be a more appropriate description of such a base-pairing complex.

To ascertain that the decreased luciferase expression was not a consequence of the corresponding amino acid changes in the luciferase fusion protein, a double mutant, U216A/C219G, was prepared to disrupt two base pairings in the pseudoknot stem without affecting the encoded peptide sequence. The mutant transcript led to a 1.6% luciferase expression. A triple mutant, U216A/C217G/C219U, with three mispairings generated in the pseudoknot stem that change the encoded leucine to a closely related isoleucine resulted in a decrease in luciferase expression to 2.7%. Thus, the disrupted luciferase expression is caused by destabilization of the pseudoknot structure in the transcript rather than a change of peptide sequences in the encoded fusion protein. It is also highly unlikely that any loss of luciferase expression among the mutants could be attributed to reduced transcription or enhanced degradation of mutant transcripts, as we have observed in northerns no appreciable differences between the levels of mutant transcript and that of the wild type (data not shown).

Enzymatic probing of the putative pseudoknot structure

To evaluate the presence of a postulated pseudoknot structure in the 264-nt GLV mRNA fragment, the latter was subjected to partial hydrolysis by four different ribonucleases. RNaseV, known to hydrolyze only RNA in helical conformations either base paired or single stranded but stacked, was used to detect double-stranded and helical RNA regions and generate RNA fragments carrying 5′-phosphates (Ehresmann et al., 1987). To probe the single-stranded regions, RNaseT1, which cleaves the 3′-phosphates from unpaired G residue (Ehresmann et al., 1987), RNaseT2, a nonspecific endoribonuclease hydrolyzing single-stranded RNA with a preference for phosphodiester bonds on the 3′-side of an A residue, and RNaseA, which cleaves 3′-phosphates from unpaired C and U residues (Blackburn, 1979), were used in the present study.

In the case of the well-established stem-loop II structure from our previous study, RNaseV digestion was observed in the predicted base-paired region of stem II at positions A146, G147, A148, U160, C161, C163, and C164 (Figs. 2 and 6). None of the G residues in the stem region were susceptible to RNaseT1, whereas most of the C and U residues in stem II were resistant to RNaseA (Figs. 2 and 6). In the loop II region, however, little RNaseV cleavage was evident, including the pentanucleotide from G154 to A158 in the postulated pseudoknot stem structure (Figs. 2 and 6). The G and A residues in this pentanucleotide were, however, also relatively resistant to RNaseT1 and RNaseT2 digestion. Resistance of this pentanucleotide to all three ribonuclease digestions suggests that it could be shielded from the enzymes by certain steric hindrance imposed by the pseudoknot loop, which could be tested by chemical probing (see below). C151, A152, and U153 in the same loop II immediately upstream from 5′-GGAGA-3′ were susceptible to RNaseA and RNaseT2 (Figs. 2 and 6), suggesting that they are in a single-stranded region accessible to the enzymes. The other arm in the putative pseudoknot stem, 5′-UCUCC-3′, was readily digested by RNaseV and was resistant to RNaseA (Fig. 3), thus suggesting its presence on the side of a stem structure that is accessible to RNaseV.

Data in Figure 2 indicate also that the stem region in stem-loop III is highly susceptible to RNaseV but resistant to RNaseT1, RNaseT2, and RNaseA. The RNaseV cleavage sites at G166, A167, G168, G169, C179, C180, U181, and C182 represent the precise stem III structure (Figs. 2 and 6). The loop III sequence is not susceptible to RNaseV hydrolysis as anticipated. It shows, however, hydrolysis at positions U171, C172, and C173 by RNaseA and U171 and C172 by RNaseT2, suggesting the single-stranded nature in this region (Figs. 2 and 6). The single residue between stem II and stem III, A165, was susceptible to RNaseV cleavage

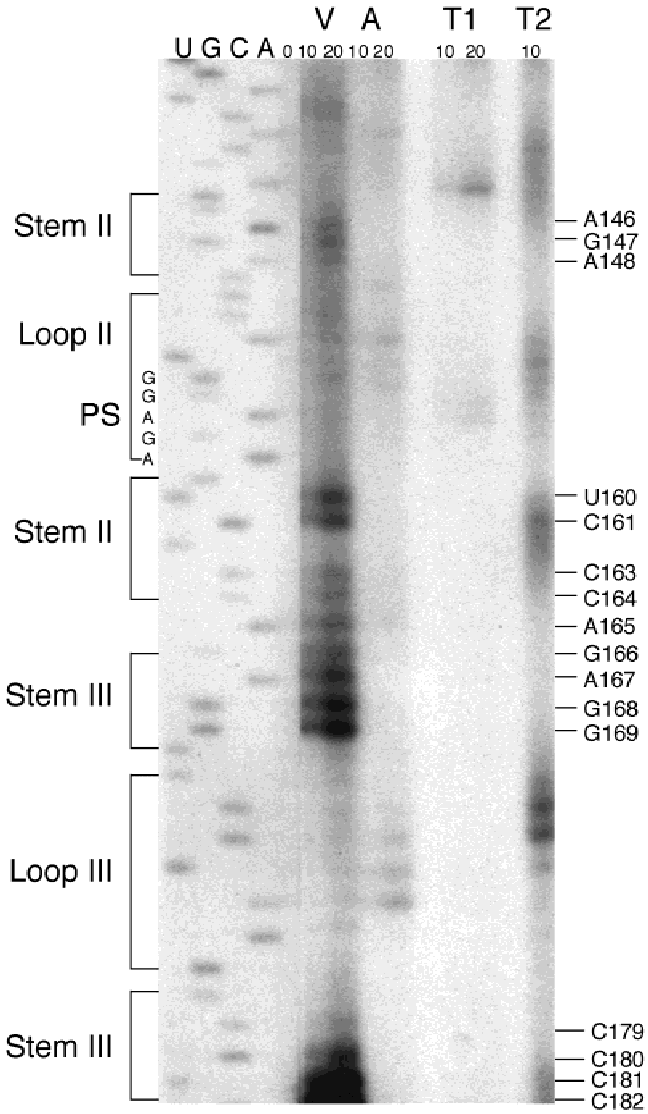

FIGURE 2. Structural probing of the 264-nt GLV mRNA fragment by ribonuclease digestion. Products from digestion of the 264-nt RNA fragment with RNaseV (V), RNaseA (A), RNaseT1 (T1), and RNaseT2 (T2) were analyzed in primer extension using a radiolabeled primer hybridizing to nt 193–213 in the RNA template. Durations of ribonuclease digestions are indicated in number of minutes above each lane. Bars, letters, and numbers on the right indicate points of digestion by RNaseV identified with the help of DNA sequence ladders presented on the left. The positions of predicted stem and loop regions are bracketed on the left side of the sequencing ladder.

(Fig. 2), which could be attributed to a stacked conformation generated by two adjacent stems on either side of the single A residue (Fig. 6).

Similarly, nucleotides in stem IVA immediately upstream from the pentanucleotide 5′-UCUCC-3′, C212, G213, G214, and U215, and those in the other arm, C194, C195, G196, G197, G198, and GA199, were all digested by RNaseV while remaining totally resistant to RNaseT1, RNaseT2, and RNaseA digestion (Figs. 3 and 6). In loop IVA, G204 was partially cleaved by RNaseT1; A203, A205, A206, and U207 were readily

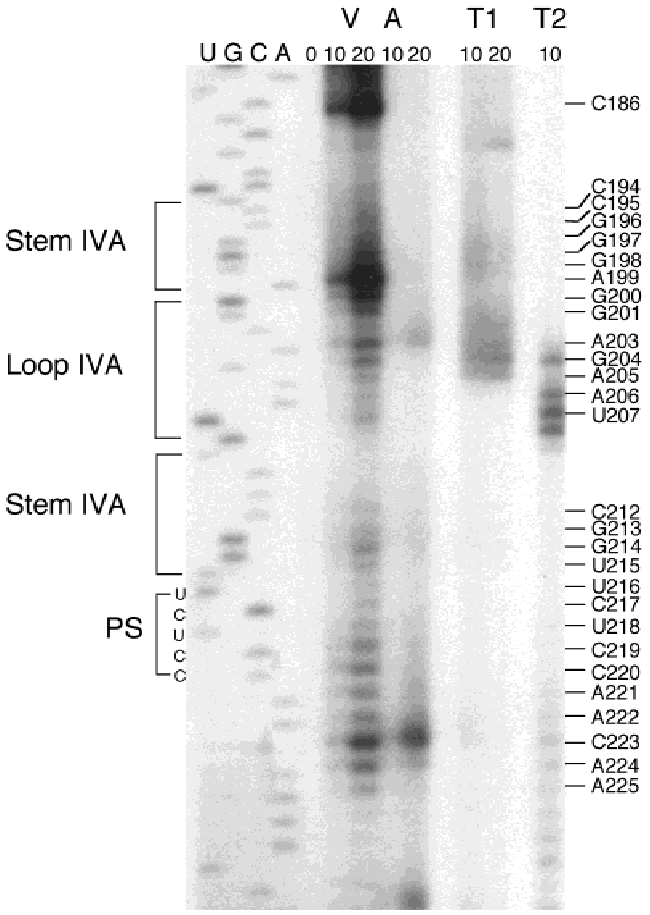

FIGURE 3. Enzymatic probing of secondary structures in the 264-nt GLV mRNA fragment. Products from RNaseV (V), RNaseA (A), RNaseT1 (T1), and RNaseT2 (T2) digestion of the RNA molecule were analyzed by reverse transcription reactions using a radiolabeled primer hybridizing to nt 252–274 in the RNA template. Durations of ribonuclease digestions are listed in minutes on top of each lane. Bars, letters, and numbers on the right side designate points in the RNA substrate hydrolyzed by RNaseV and identified with the help of DNA sequencing ladders on the left. The positions of predicted stem and loop regions are bracketed on the left side of the sequencing ladder.

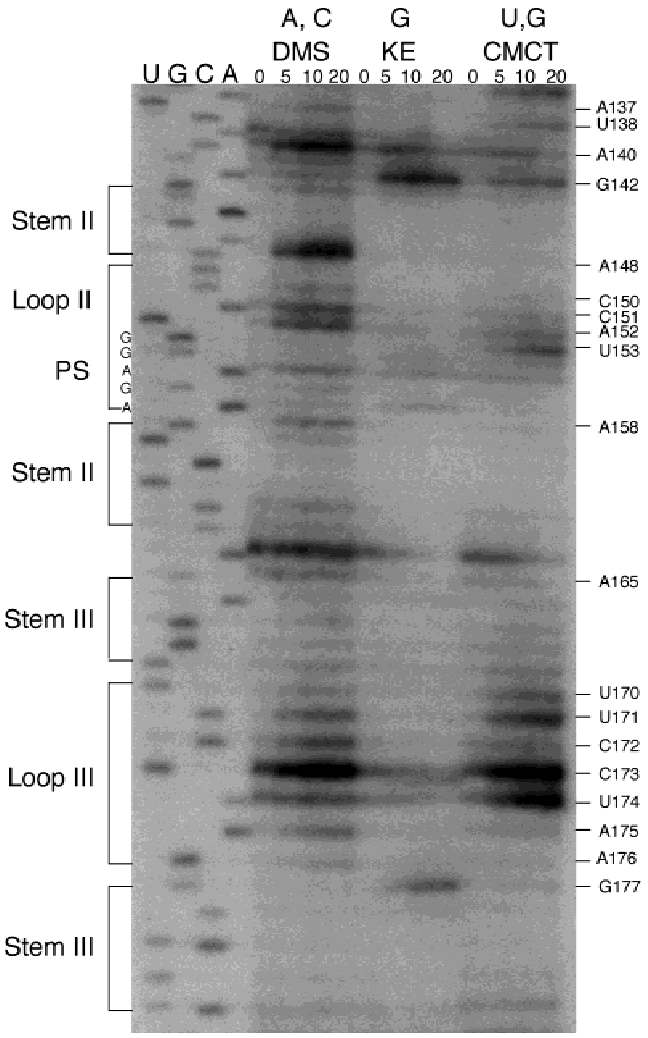

FIGURE 4. Chemical probing of secondary structures in the 264-nt GLV mRNA fragment. Chemical modifications of A and C (by DMS), G (by KE), and U and G (by CMCT) were monitored by reverse transcriptase using radiolabeled primer hybridizing to position 193–213. Time-points for modification in minutes are indicated above each lane. Positions indicated are the modified bases that correspond to the bases shown alongside the DNA sequencing ladder. The positions of stems and loops are shown on the left side of the figure.

digested by RNaseT2; and C202 was hydrolyzed by RNaseA as would have been predicted (Figs. 3 and 4). G200, G201, A203, G204, A205, A206, and U207 were, however, also slowly hydrolyzed by RNaseV beginning from the 5′ end and progressing gradually toward the 3′ end over a 20-min incubation time (Fig. 3), as if the loop was also involved in forming a stem with another yet unidentified single-stranded structure during the time course of RNaseV digestion. Because this potential artifact would not occur in chemical probing, the structure of this particular region in loop IVA was reexamined by the latter method as well (see below).

The predicted single-stranded 10-nt stretch (183–192) between stems III and IVA showed little digestion by any of the ribonucleases tested. Only G187 showed weak digestion with RNaseT1 and C186 showed limited digestion with RNaseV. Because this region is flanked by two stem-loops, its accessibility to enzyme molecules is also questionable and could be examined by chemical probing under similar native conditions in subsequent experiments.

Chemical probing of the putative pseudoknot structure

The putative pseudoknot structure was also evaluated by probing the RNA with chemical agents. DMS (dimethyl sulphate), which methylates unpaired A residue at N-1 and unpaired C residue at N-3, KE (kethoxal), which modifies only unpaired G residue at

N-1, and CMCT (1-cyclohexyl-3-(-2-morpholinoethyl)-carbodiimide), which modifies the N-3 group of unpaired U and the N-1 group of unpaired G residue, were used in the present study. In our previous probing experiments, we probed the RNA molecule under semidenatured conditions, and observed that only the three C residues in the 5′-UCUCC-3′ pentanucleotide were inaccessible to chemical probing, whereas the two Us and the 5 nt in 5′-GGAGA-3′ were all chemically modified (Garlapati et al., 2001). In our present study, we used the same slow renaturing protocol under native conditions as for the enzymatic probing (i.e., in the presence of 10 mM $MgCl_2$ instead of 1 mM EDTA; Garlapati et al., 2001), to try to stabilize the putative pseudoknot structure. A major difference between the outcomes from under these two conditions is that bases in the pentanucleotide 5′-GGAGA-3′ in loop II were no longer labeled by the chemicals (Fig. 4). Only a weak DMS labeling of A158 at the 3′ end of the pentanucleotide was observed, suggesting that the bases are mostly in paired conformation. The weak modification of the last residue in the pentanucleotide could still be due to partial denaturation of the pseudoknot stem under the current in vitro conditions. Similarly, all the five residues in the other pentanucleotide, 5′-UCUCC-3′, were not labeled at all under the present conditions (Fig. 5). Thus, under more native conditions, both of the pentanucleotides postulated to be involved in a pseudoknot formation are in a base-pairing conformation.

Loop IVA, containing residues 200–208, had residues 202–207 chemically modified under semidenatured conditions (Garlapati et al., 2001) and had 200–208 modified by the same chemicals under more native conditions in the present study (Fig. 5). Its susceptibility to RNaseT1, RNaseT2, and RNaseA digestions, as seen in Figure 3, further confirms the single-stranded nature of loop IVA. The slow and time-dependent RNaseV digestion of this loop sequence observed in Figure 3 could indeed be attributed to a likely conformational change of the RNA brought about by the initial RNaseV hydrolysis of the molecule.

The bases G183, A184, U185, C186, G187, G189, and U192 in the previously identified single-stranded 10-nt region separating stem III and stem IVA also became chemically modified under the native condition (Fig. 5), verifying its presence in a single-stranded form. Its relative insensitivity toward digestion by RNaseT1, RNaseT2, and RNaseA shown in Figure 2 could thus be taken as an indication that the sequence, flanked by two stem loops, is inaccessible to these enzymes.

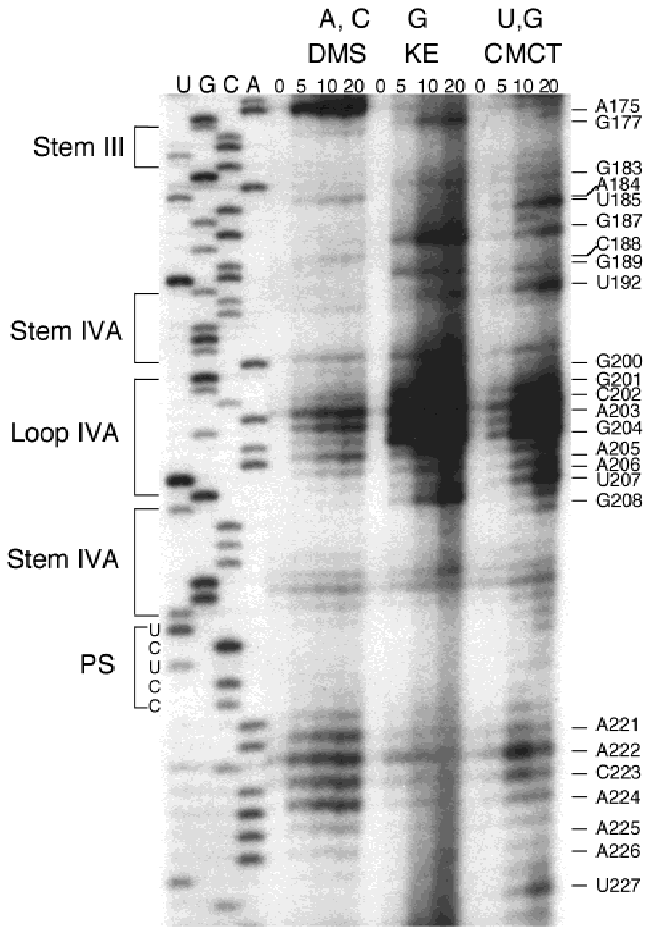

FIGURE 5. Chemical probing of the secondary structures in the 264-nt GLV mRNA fragment. Chemical modifications of A and C (by DMS), G (by KE), and U and G (by CMCT) were monitored by reverse transcriptase using radiolabeled primer hybridizing to position 252–274. Time-points for modification in minutes are indicated above each lane. Positions indicated are the modified bases that correspond to the bases shown alongside the DNA sequencing ladder. The positions of stems and loops are shown on the left side of the figure.

Further structure-function analysis of the pseudoknot by site-directed mutations

The enzymatic probing experiments indicated that A221 and A222 immediately downstream from the pentanucleotide 5′-UCUCC-3′ could be slowly digested by RNaseV (Figs. 3 and 6) suggesting that A221 could base pair with U153 in loop II, immediately upstream to the other pentanucleotide sequence, 5′-GGAGA-3′, to extend the length of base pairings by one more (Fig. 6). An A152C/U153C substitution in the previous study was found without significant effect on translation, suggesting a lack of functional importance of the postulated U153/A221 base pair (Garlapati et al., 2001). In the current investigation, two double substitutions, U153A/G154C and C220G/A221U, were each found to cause decrease of luciferase expression to 11% and 8%, which was restored to 78% of the wild-type level by combining the two (Table 1). Single substitutions G154C and C220G each expressed luciferase at 9.1% and 10.2% of the wild type but restored the activity to 78.3% of the wild type in combination (Table 1). These nearly identical results between double and single substitutions suggest that the extra sixth base pairing between

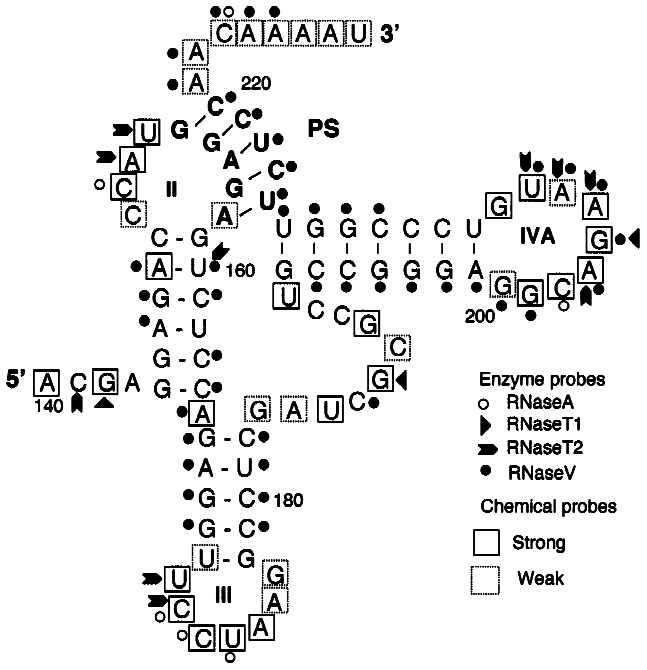

FIGURE 6. Summary of the results from enzymatic and chemical probing of the pseudoknot structure in the nt 140–228 region of GLV mRNA. Enzymatic digestion data from RNaseV (after 20 min), RNaseA (after 20 min), RNaseT1 (after 20 min), and RNaseT2 (after 10 min) indicating points of hydrolysis are represented by closed circles, open circles, closed triangles, and arrowheads, respectively. Bases that are strongly modified by various chemicals are boxed in continuous line. Bases boxed in broken line demonstrated weaker chemical modification. Numbers indicate nucleotide positions. PS: pseudoknot stem.

		RLU (%)
WT	UGG GCA AGC CCG **AUG** GUG UGG GGG ACU GGA.....	100
A1C	UGG GCA AGC CCG CUG GUG UGG GGG ACU GGA.....	0.5±0.7
C-3A	UGG GCA AGC **AUG** CUG GUG UGG GGG ACU GGA.....	0.07±0.01
G-5T	UGG GCA **AUG** CCG CUG GUG UGG GGG ACU GGA.....	1.47±0.3
G4A	UGG GCA AGC CCG CUG **AUG** UGG GGG ACU GGA....	0.7±0.3
T7A	UGG GCA AGC CCG CUG GUG **AUG** GGG ACU GGA....	0.04±0.02

FIGURE 7. Effects of changing position of the initiation codon in the GLV-luciferase chimeric mRNA on luciferease expression in transfected *Giardia*. Arrows indicate the original position of the initiation codon in wild-type GLV transcript. Open arrows indicate the shifted positions of initiation codon in the mutant transcripts. RLU(%) indicates the average of relative luciferase activity obtained from at least two independent transfection experiments with the wild-type (pC631-luc) transcript as the positive control with RLU(%) = 100.

153U and 221A does not make a significant contribution to the stability of the presumed pseudoknot structure.

Translation is initiated at a precise position in the viral transcript

The pseudoknot is located 143 nt downstream from the initiation codon and yet its disruption leads to a drastic loss of translation efficiency (Yu et al., 1998; Garlapati et al., 2001). Together with the other essential structural elements in this loosely defined 264-nt downstream IRES [stem-loop I, the downstream box (DB) and stem-loops II, III, and IVA located 8, 63, 141, 162, and 190 nt downstream from the initiation codon, respectively; Yu et al., 1998; Garlapati et al., 2001], the pseudoknot may form a rather rigid complex structure for recruiting the small ribosomal subunit of *Giardia*. It predicts an inflexible location of the initiation codon in the transcript for an efficient initiation of translation. There is only one AUG codon located at the 5′ end of the IRES. We mutated it to CUG and observed a virtual abolishment of translation down to 0.5% of the wild-type level as anticipated (Fig. 7). This loss of luciferase expression was not recovered when an AUG codon was placed either one or two codons upstream or one or two codons downstream from its original position now occupied by CUG (Fig. 7). This requirement of a highly precise location for an AUG to initiate translation suggests that the viral IRES dictate a highly defined geometric arrangement for recruiting the *Giardia* protein synthetic machinery to initiate translation at a precise position.

DISCUSSION

We have identified, by site-directed mutagenesis, the essential role of a putative downstream pseudoknot structure in the GLV transcript for its efficient translation. Results from enzymatic probing provided some support for the existence of such a pseudoknot structure. The asymmetrical RNaseV digestion on the pseudoknot stem could be due to the documented C residue preference by RNaseV (van Dijk et al., 2000), or alternatively, a yet unidentified steric hindrance mak-

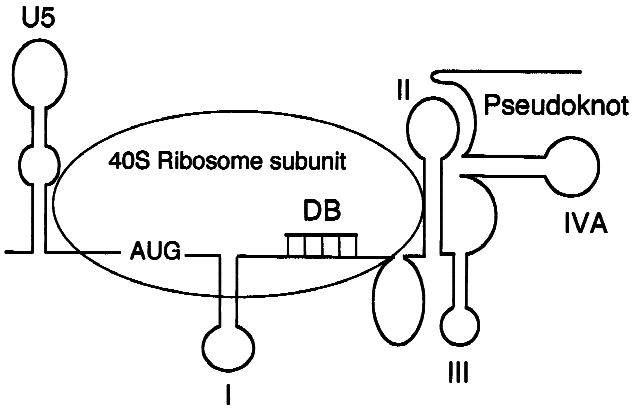

FIGURE 8. A structural model of the IRES in GLV mRNA.

ing the other GGAGA arm relatively inaccessible for RNaseV, as observed with other pseudoknots (Alam et al., 1999; Liphardt et al., 1999). The GGAGA arm is unlikely in a single-stranded region, because it was also resistant to RNaseT1 and RNaseT2 digestions. Furthermore, results from chemical probing indicate that both arms are in base paired structures, thus confirming the presence of the postulated pseudoknot stem. This stem and stem IVA are continuous on one strand of their double helices and branched out on the other. The common strand, rich in C residues, was readily digested by RNaseV. The other arm in the pseudoknot stem is continuous with one of the arms in stem II. In most of the H-type pseudoknots, the stems are usually coaxially stacked with a single-stranded loop in between (Hilbers et al., 1998). In the present case, however, it is not clear if the pseudoknot stem, stem IVA, and stem II could be placed coaxially or at an angle to one another, as there is an adjacent stem-loop III involved in the complex as well (see Fig. 6).

Our experimental results also indicated that loop IVA is in a single-stranded form. Base substitutions in the loop IVA sequence showed that G200C/G201C/C202A and A203U/G204C resulted in 2.2% and 82% of luciferase expression (data not shown), suggesting the functional importance of G200/G201/C202. Because it happens to be complementary to an upstream sequence G189/C190/C191, which was not digested by any of the ribonucleases (Fig. 3), we tried another mutant, G189U/C190G/C191G, which resulted also in a drastic reduction in luciferase expression to 3.6% (data not shown). A combination of the two mutations, however, maintained a luciferase expression at a similarly low level (data not shown), thus suggesting that the important roles these two trinucleotides may play in translation initiation is probably not by hybridizing with each other to generate another pseudoknot stem. Data from chemical probing indicate that the nt 183–192 segment between stem-loops III and IVA is in a single-stranded conformation. Its relative inaccessibility to enzymes could be caused by the two flanking structural complexes with pseudoknot stem and stem IVA on one side and stems II and III on the other.

The single A165 at the junction between stems II and III was strongly modified by DMS (Fig. 4) but was also the site of attack by RNaseV, indicating that the residue is in a single strand with a stacked conformation (Ehresmann et al., 1987). It also suggests that the two stems are probably in the same axis but pointing toward opposite directions (see Fig. 6).

A pseudoknot is known to play a variety of regulatory functions in controlling protein synthesis. In translation initiation, a pseudoknot is usually positioned in the nonprotein coding 5′ UTR overlapping with the ribosome binding site and/or the initiation codon in the mRNA, thereby facilitating recognition (Rijnbrand et al., 1997) of the initiation codon by the ribosomal complex (Tang & Draper, 1989; Philippe et al., 1993; Ehresmann et al., 1995). In the IRES of hepatitis C virus (HCV) transcript, a pseudoknot upstream from the AUG codon is required for internal initiation of translation (Wang et al., 1995); mutations that disrupted the pseudoknot structure dramatically reduced internal initiation of translation. A spacing requirement of 11 to 12 nt between the initiator AUG and an upstream pseudoknot motif in the 5′ UTR appears to be essential for translation initiation (Wang & Siddiqui, 1995; Wang et al., 1995). Similarly, a pseudoknot in the 5′ UTR of classical swine fever virus (CSFV) transcript is an essential element for its IRES function (Rijnbrand et al., 1997). Recently, it was reported that in cricket paralysis virus (PSIV) a pseudoknot present just upstream of the capsid-coding region is required for an unusual translation initiation at a non-AUG codon (Sasaki & Nakashima, 2000; Wilson et al., 2000). The ribosome 40S subunits were unable to recognize an AUG codon placed 7 or 19 nt downstream of an inactivated authentic translation initiation codon (Rijnbrand et al., 1997). These examples indicate that, among these viruses, the ribosome is usually directed by a pseudoknot-containing IRES in 5′ UTR to bind directly to or near the initiator AUG codon (Reynolds et al., 1996). In our present study, the pseudoknot-containing IRES is located primarily downstream from the initiation codon and the location of the latter has virtually no flexibility for translation initiation, thus suggesting that the ribosome is directly recruited onto the initiation codon without a prior ribosomal scanning (Fig. 8). The structures surrounding the initiation codon thus must play a major role in precisely positioning the ribosome onto the AUG codon. In our previous study, we showed that stem-loop I, located 8 nt downstream of the initiation codon, plays an essential role in translation. Our recent preliminary results indicated also that a stem-loop structure (stem-loop U5) in the immediate 5′ UTR, 21 nt upstream from the AUG triplet, also plays an essential function for translation initiation (data not shown). The spacing (including the AUG triplet itself) between the two stem-loop structures U5 and I is estimated to be 33 nt, which could provide the immediately necessary secondary structural environment for a direct recruitment of ribosome to the initiator AUG (Fig. 8). However, a 13-nt DB sequence located 64 nt downstream of AUG codon, 30 nt downstream from stem-loop I, and complementary to a 15-nt sequence at the 3′ end of the 16S-like rRNA (anti-DB) was found to be essential for translation of the viral transcript in *Giardia* (Yu et al., 1998). Similar sequences have not yet been identified in the IRES of HCV and CSFV, but evidence for base pairings between the IRES and the 18S rRNA was obtained from reconstitution studies (Pestova et al., 1998). It is possible that the interaction between DB and anti-DB helps positioning the recruited 40S ribosome subunit to fit the initiation codon in the transcript into the translation initiation site. How-

ever, because DB is located downstream from stem-loop I, the latter may be involved with binding to the ribosome complex, but may not be a part of the specific geometric arrangement deterring ribosomal movement. The presence of a pseudoknot, stem-loops II, III, and IVA complex structure further downstream from the DB could serve the function of retarding the downstream movement of ribosome prior to a sound and error-proof recognition of the initiation codon (Kozak, 1990).

The presence of a potential IRES downstream from the initiation codon has been also observed among some of the other viral transcripts. For hepatitis A virus, inclusion of more than 114 nt of its 5′ capsid-coding transcript enhances the translation by three- to fourfold (Graff & Ehrenfeld, 1998). Inclusion of the first 266 nt of the 5′ capsid-coding sequence of the Sinbdis viral transcript enhances translation by 10-fold (Frolov & Schlesinger, 1996). Our current identification of putative downstream IRES in GLV mRNA may yet carry an additional biological significance beyond an interesting phenomenon among the viruses. The dependence of GLV on downstream IRES for translation could also reflect one of the uniquely primitive features of the translation machinery and mechanisms operating in its host, *Giardia*. The cellular mRNAs in this primitive protozoan are mostly uncapped (Yu et al., 1998) and have very short 5′ UTR in the range of 1 to 6 nt (Adam, 1991). In the absence of a cap and a substantial 5′ UTR, how mRNA molecules can recruit the 40S ribosome subunit to the precise decoding region to initiate translation has not been clearly understood. The GLV transcript is translated by the same machinery as the cellular mRNA of *Giardia*, and translation of the viral mRNA is known to exert no inhibitory effect on translation of cellular mRNA (Wang & Wang, 1986). Insights obtained from localizing the IRES in the downstream region and analyzing the structure–function relationship of the IRES in GLV mRNA may shed some light on whether the cellular mRNA in *Giardia* could also depend on similar elements present downstream of the initiation codon for translation initiation. Thus, translation of GLV transcript may serve as a model system for understanding the translational initiation mechanisms in this most primitive and one of the earliest diverging and living eukaryotes known to man.

MATERIALS AND METHODS

Site-directed mutagenesis

Plasmid construct pC631-luc has the GLV genomic cDNA cloned into a pGEM-T vector and a full-length luciferase gene inserted between nt 631 and 4256 of the viral genome cDNA (Yu & Wang, 1996). The sequence upstream from the luciferase gene thus consists of cDNA encoding the 367-nt GLV 5′ UTR and the first 264-nt capsid-encoding region of GLV mRNA. The luciferase gene is fused in frame with the upstream cDNA encoding the 264-nt viral mRNA (Yu & Wang, 1996). Site-directed mutagenesis of the 264-nt sequence in the present study was carried out essentially as described previously (Garlapati et al., 2001) using a QuickChange site-directed mutagenesis kit following the manufacturer's instructions (Stratagene). Individual mutations were each verified by DNA sequencing. The location of each mutation and the specific nucleotides involved are indicated throughout the text.

In vitro transcription

The mutant plasmids of pC631-luc were each linearized with *Nru*I at the 3′ end of GLV genomic cDNA and used as template for in vitro synthesis of transcripts using a MegaScript T7 transcription kit (Ambion).

Transfection of *Giardia* trophozoites

The in vitro transcripts of various GLV-luciferase chimeric cDNA mutants were each introduced into GLV-infected WB strain *G. lamblia* trophozoites (WBI) by electroporation as described (Yu et al., 1995; Garlapati et al., 2001). Approximately 4×10^6 trophozoites were transfected with 100 μg of the in vitro-synthesized transcript. Mutant and wild-type transcripts were each used in triplicate for transfection study in every duplicated experiment.

Luciferase assay

The transfected *G. lamblia* trophozoites were lysed and assayed for luciferase activity 20 h posteletroporation as described (Yu et al., 1995). Each mutant transfectant in triplicate from two independent transfection experiments was examined, with the wild-type pC631-luc transfectant as a positive control. Luciferase activity was calculated in relative light units (RLU) per microgram of crude lysate protein determined by the Bradford method (Bradford, 1976).

Enzymatic probing of RNA structure

Enzymatic probing was carried out essentially as described (Moazed et al., 1986; Stern et al., 1988). cDNA encoding the 264-nt RNA fragment was amplified by polymerase chain reactions (PCR) using pC631-luc as template, and the amplified product was cloned into a pGEM-T-easy vector (Promega). The recombinant plasmid was linearized by *Hin*dIII and transcribed in vitro to produce the 264-nt transcript using the T7 Megascript transcription kit (Ambion). The RNA was initially denatured at 65 °C for 15 min in a probing buffer (160 mM HEPES, pH 7.5, 50 mM KCl, 10 mM $MgCl_2$) followed by a slow cooling to ambient temperature for an hour. Approximately 10 μg of the renatured RNA was used for each enzymatic digestion. RNaseV (Pharmacia), RNaseT1 (Life Technologies, Inc.), RNaseT2 (Life Technologies), and RNaseA (Ambion) were each diluted in the probing buffer and the enzymatic activities were titrated (in units or nanograms) to identify the optimal probing condition for each enzyme. The following optimal amount of each enzyme was determined and used in subsequent experiments: 0.35 U of RNaseV, 0.1 U of RNaseT1, 2.5 U of RNaseT2, and 12 ng of

RNaseA. All the enzymatic digestions were performed in a final volume of 100 μL at 37 °C for a period of 10 to 20 min. The reactions were stopped by adding phenol-chloroform to the digestion mixture, and the RNA substrates were extracted and recovered by ethanol precipitation in the presence of 0.3 M NaOAc and 10 μg yeast tRNA. RNA pellets were dissolved in DEPC-treated water and subjected to primer extension analysis to determine the enzyme-cleaved sites in the RNA molecule.

Chemical probing of RNA structure

Chemical probing was carried out essentially as previously described (Moazed et al., 1986; Stern et al., 1988; Garlapati et al., 2001) except that the RNA substrate was renatured to assume its native structure prior to chemical probing as in the enzymatic probing. Approximately 10 μg of the 264-nt transcript were used for each type of modification. For DMS and KE probing, 10 μg of the RNA sample were suspended in 300 μL and 270 μL of HMK buffer (160 mM HEPES, pH 7.2, 50 mM KCl, 10 mM $MgCl_2$), respectively. For CMCT probing, the same amount of RNA sample was suspended in 150 μL of BMK buffer (70 mM potassium-borate, pH 8.0, 50 mM KCl, 10 mM $MgCl_2$). The RNA samples were denatured in their respective buffers at 65 °C for 15 min and slowly renatured to ambient temperature for an hour.

DMS modification of RNA was carried out by adding 12 μL of DMS (diluted 1:12 in ethanol) to the renatured RNA in 300 μL of HMK buffer, and incubated at 37 °C for varying lengths of time (0, 5, 10, and 20 min). The reaction was stopped by adding 75 μL of the DMS stop buffer (1 M Tris-acetate, pH 7.5, 1 M β-mercaptoethanol, 1.5 M sodium acetate, 0.1 mM EDTA). For KE treatment, 30 μL of KE at 37 mg/mL in 20% (v/v) ethanol was added to the renatured RNA in 270 μL of HMK buffer. The reaction was carried out as described and stopped by stabilizing the modified RNA with 40 mM potassium borate, pH 7.0. CMCT modification was carried out by adding an equal volume (150 μL) of freshly prepared CMCT (42 mg/mL) in BMK buffer to the RNA sample, incubating, and stopping as described for DMS modification. The treated RNA samples were precipitated by adding 2.5 vol of ethanol with 0.3 M sodium acetate, redissolved in nuclease free water, and extracted once with phenol-chloroform mixture and twice with chloroform-isoamylalcohol mixture. The extracted aqueous phase was reprecipitated with 2.5 vol of ethanol and 0.3 M sodium acetate.

Primer extension

Primer extension was carried out as previously described (Garlapati et al., 2001). Two ^{32}P-end-labeled primers, complementing nucleotides 193–213 and 252–274 in the chimeric RNA, were each annealed to 5 μg of RNA by incubating at 65 °C for 15 min followed by an additional 10 min on ice. Primer extension was carried out at 42 °C for 1 h using 200 U of M-MLV reverse transcriptase (Life Technologies). The radiolabeled products were analyzed in 8% denaturing polyacrylamide gel electrophoresis. Sequencing ladders generated by the femtomole-cycle sequencing system (Promega) were included as reference. The position of each band in the lanes of enzyme-digested RNA corresponded to a base in the sequencing ladder that is cleaved at either the 3′ end (digested by RNaseT1, RNaseT2, or RNaseA) or the 5′ end (digested by RNaseV) thus terminating the primer extension. The chemically modified bases were each identified as a reverse transcriptase stop with a higher mobility 1 nt short of that in the corresponding DNA sequencing gel, because primer extension would stop in front of the modified base.

ACKNOWLEDGMENTS

We thank Jim Chou for his technical assistance and Alice Wang for her critical reading of the manuscript. This work was supported by a National Institutes of Health Grant AI-30475.

Received December 19, 2001; returned for revision February 5, 2002; revised manuscript received February 20, 2002

REFERENCES

Adam RD. 1991. The biology of *Giardia* spp. *Microbio Rev 55*:706–732.

Alam SL, Wills NM, Ingram JA, Atkins JF, Gesteland RF. 1999. Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukemia virus. *J Mol Biol 288*:837–852.

Blackburn P. 1979. Ribonuclease inhibitor from human placenta: Rapid purification and assay. *J Biol Chem 254*:12484–12487.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem 72*:248–254.

Chang K-Y, Tinoco I Jr. 1994. Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome. *Proc Natl Acad Sci USA 91*:8705–8709.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res 15*:9109–9128.

Ehresmann C, Philippe C, Westhof E, Benard L, Portier C, Ehresmann B. 1995. A pseudoknot is required for efficient translational initiation and regulation of the *Escherichia coli* rpsO gene coding for ribosomal protein S15. *Biochem Cell Biol 73*:1131–1140.

Frolov I, Schlesinger S. 1996. Translation of Sindbis virus mRNA: Analysis of sequences downstream of the initiating AUG codon that enhance translation. *J Virol 70*:1182–1190.

Garlapati S, Chou J, Wang CC. 2001. Specific secondary structures in the capsid-coding region of giardiavirus transcript are required for its translation in *Giardia lamblia*. *J Mol Biol 308*:623–638.

Graff J, Ehrenfeld E. 1998. Coding sequences enhance internal initiation of translation by hepatitis A virus RNA in vitro. *J Virol 72*:3571–3577.

Hilbers CW, Michiels PJ, Heus HA. 1998. New developments in structure determination of pseudoknots. *Biopolymers 48*:137–153.

Kozak M. 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci USA 87*:8301–8305.

Li L, Wang AL, Wang CC. 2001. Structural analysis of the −1 ribosomal frameshift elements in giardiavirus mRNA. *J Virol 75*:10612–10622.

Liphardt J, Napthine S, Kontos H, Brierley I. 1999. Evidence for an RNA pseudoknot loop-helix interaction essential for efficient −1 ribosomal frameshifting. *J Mol Biol 288*:321–335.

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.

Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CU. 1998. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes & Dev 12*:67–83.

Philippe C, Eyermann F, Benard L, Portier C, Ehresmann B, Ehresmann C. 1993. Ribosomal protein S15 from *Escherichia coli* mod-

ulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc Natl Acad Sci USA 90*:4394–4398.
Reynolds JE, Kaminski A, Carroll AR, Clarke BE, Rowlands DJ, Jackson RJ. 1996. Internal initiation of translation of hepatitis C virus RNA: The ribosome entry site is at the authentic initiation codon. *RNA 2*:867–878.
Rijnbrand R, van der Straaten T, van Rijn PA, Spaan WJ, Bredenbeek PJ. 1997. Internal entry of ribosomes is directed by the 5′ noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. *J Virol 71*:451–457.
Sasaki J, Nakashima N. 2000. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc Natl Acad Sci USA 97*:1512–1515.
Stern S, Moazed D, Noller HF. 1988. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol 164*:481–489.
Tang CK, Draper DE. 1989. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell 57*:531–536.
ten Dam E, Pleij K, Draper D. 1992. Structural and functional aspects of RNA pseudoknots. *Biochemistry 31*:11665–11676.
ten Dam EB, Pleij CW, Bosch L. 1990. RNA pseudoknots: Translational frameshifting and readthrough on viral RNAs. *Virus Genes 4*:121–136.
van Dijk EL, Sussenbach JS, Holthuizen PE. 2000. Distinct RNA structural domains cooperate to maintain a specific cleavage site in the 3′-UTR of IGF-II mRNAs. *J Mol Biol 300*:449–467.
Wang AL, Wang CC. 1986. Discovery of a specific double-stranded RNA virus in *Giardia lamblia*. *Mol Biochem Parasitol 21*:269–276.
Wang AL, Yang HM, Shen KA, Wang CC. 1993. Giardiavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-pol-like fusion protein by a translation frameshift. *Proc Natl Acad Sci USA 90*:8595–8599.
Wang C, Le SY, Ali N, Siddiqui A. 1995. An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5′ noncoding region. *RNA 1*:526–537.
Wang C, Siddiqui A. 1995. Structure and function of the hepatitis C virus internal ribosome entry site. *Curr Top Microbiol Immunol 203*:99–115.
Wilson JE, Pestova TV, Hellen CUT, Sarnow P. 2000. Initiation of protein synthesis from the A site of the ribosome. *Cell 102*:511–520.
Yu DC, Wang AL, Botka CW, Wang CC. 1998. Protein synthesis in *Giardia lamblia* may involve interaction between a downstream box (DB) in mRNA and an anti-DB in the 16S-like ribosomal RNA. *Mol Biochem Parasitol 96*:151–165.
Yu DC, Wang AL, Wu CH, Wang CC. 1995. Virus-mediated expression of firefly luciferase in the parasitic protozoan *Giardia lamblia*. *Mol Cell Biol 15*:4867–4872.
Yu DC, Wang CC. 1996. Identification of *cis*-acting signals in the giardiavirus (GLV) genome required for expression of firefly luciferase in *Giardia lamblia*. *RNA 2*:824–834.