EUKARYOTIC CELL, Apr. 2005, p. 742–754 Vol. 4, No. 4
1535-9778/05/$08.00+0 doi:10.1128/EC.4.4.742–754.2005

Structural Elements in the 5′-Untranslated Region of Giardiavirus Transcript Essential for Internal Ribosome Entry Site-Mediated Translation Initiation†

Srinivas Garlapati and Ching C. Wang*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94107-2280

Received 3 February 2005/Accepted 15 February 2005

Translation of uncapped giardiavirus (GLV) mRNA in *Giardia lamblia* requires the presence of a 5′-untranslated region (5′-UTR) and a viral capsid coding region. We used dicistronic viral constructs to show that the downstream 253 nucleotides (nt) of the 5′-UTR plus the initial 264-nt capsid coding region constitute an internal ribosome entry site (IRES). Predicted secondary structures in the 253-nt 5′-UTR include stem-loops U3, U4a, U4b, U4c, and U5. Chemical and enzymatic probing analysis confirmed the presence of all predicted stem-loops except U4a. Disruption of stem-loop structures U3 and U5 by site-directed mutagenesis resulted in a drastic reduction in translation of a monocistronic viral transcript, which could be restored by compensatory sequence changes. Mutations disrupting stem-loops U4b and U4c do not exert an appreciable effect on translation, but certain sequences in the U4a region and in U4b do appear to play important roles in the IRES. Structural analysis also suggests that an 8-nt U3 loop sequence (nt 147 to 154) pairs with an 8-nt downstream sequence (nt 168 to 175) to form a pseudoknot. Disruption of this pseudoknot by mutagenesis resulted in a drastic reduction in translation, which could be restored by compensatory sequence changes. This study has defined the secondary structure in the 5′-UTR of the IRES. Together with the previous results, we have now completed analysis of the entire structure of GLV IRES and fully defined the functionally essential structural elements in it.

Giardiavirus (GLV), which specifically infects the trophozoites of the protozoan parasite *Giardia lamblia*, harbors a double-stranded RNA genome of 6,277 bp (23). The plus-strand viral transcript is flanked by a 367-nucleotide (nt) 5′-untranslated region (5′-UTR) and a 301-nt 3′-UTR and directs the translation of a major 100-kDa capsid protein (Gag) and a minor 190-kDa fusion protein (Gag-Pol) via a −1 ribosomal frameshift (10, 24). The ability of purified GLV to infect and the capability of its plus-strand RNA to transfect *G. lamblia* trophozoites, resulting in intracellular proliferation of infectious GLV particles, are the major distinguishing features of this virus among *Totiviridae* (23).

Translation of GLV transcript in *Giardia* is initiated on a unique internal ribosome entry site (IRES) element that contains sequences from a part of the 5′-UTR and a portion of the capsid coding region (6). By expressing dicistronic viral transcripts in transfected *Giardia*, we showed previously that both the 5′-UTR and the downstream coding region are required for mediating internal ribosome entry (6). Functional and structural analysis of the 264-nt coding region of the IRES were conducted, and several structural elements essential for translation initiation were identified (4). They include a 13-nt downstream box (DB) at positions 66 to 78 that complements (with two gaps) a 15-nt sequence near the 3′ end of *Giardia* 16S-like rRNA (29), stem-loops I (nt 11 to 35), II (nt 144 to 164), III (nt 166 to 182), and IVA (nt 193 to 215), and a novel pseudoknot structure between loop II and a region downstream from stem-loop IVA (4, 5). Mutations that destroyed the stems in stem-loops I and II or disrupted the pairing of DB with rRNA resulted in a drastic reduction in IRES-mediated translation.

The MFOLD-predicted secondary structures of the 5′-UTR in GLV mRNA include eight stem-loops: U1, U1a, U2, U3, U4a, U4b, U4c, and U5 (Fig. 1). Previous deletion analysis suggested that only the 253-nt downstream portion of the 5′-UTR (nt 114 to 367), involving stem-loops U3 to U5, were required for the IRES function (6). In the present study, we employed chemical and enzymatic probing techniques to verify the MFOLD-predicted secondary structures from U3 to U5 in GLV IRES. Using a combination of site-directed mutagenesis and deletion analysis, we tested also the roles of individual secondary structures in the 5′-UTR portion of GLV IRES.

MATERIALS AND METHODS

RNA secondary structure prediction. Secondary structure predictions for the 367-nt 5′-UTR from GLV mRNA were made using web implementation of the MFOLD algorithm (30), incorporating version 3.0 of the Tuner rules (14).

Chemical probing of RNA structure. Chemical probing was carried out essentially as previously described (15, 20) except that the reactions were carried out at 37°C (8). A 631-nt in vitro transcript from pC631, containing the entire 367-nt 5′-UTR and the downstream 264-nt capsid coding sequence (4), was analyzed in this study. For dimethyl sulfate (DMS) and KE (kethoxal) probing, 10 μg of the RNA sample was 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 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide (CMCT) probing, the same amount of RNA was suspended in 150 μl of BMK buffer (70 mM potassium borate [pH 8.0], 50 mM KCl, 10 mM $MgCl_2$). RNA samples were first denatured at 65°C for 15 min and slowly renatured at ambient temperature for an hour prior to the addition of the chemical probes.

* Corresponding author. Mailing address: Dept. of Pharmaceutical Chemistry, University of California San Francisco, Mission Bay Genentech Hall, 600 16th St., San Francisco, CA 94107-2280. Phone: (415) 476-1321. Fax: (415) 476-3382. E-mail: ccwang@cgl.ucsf.edu.

† Supplemental material for this article may be found at http://ec.asm.org/.

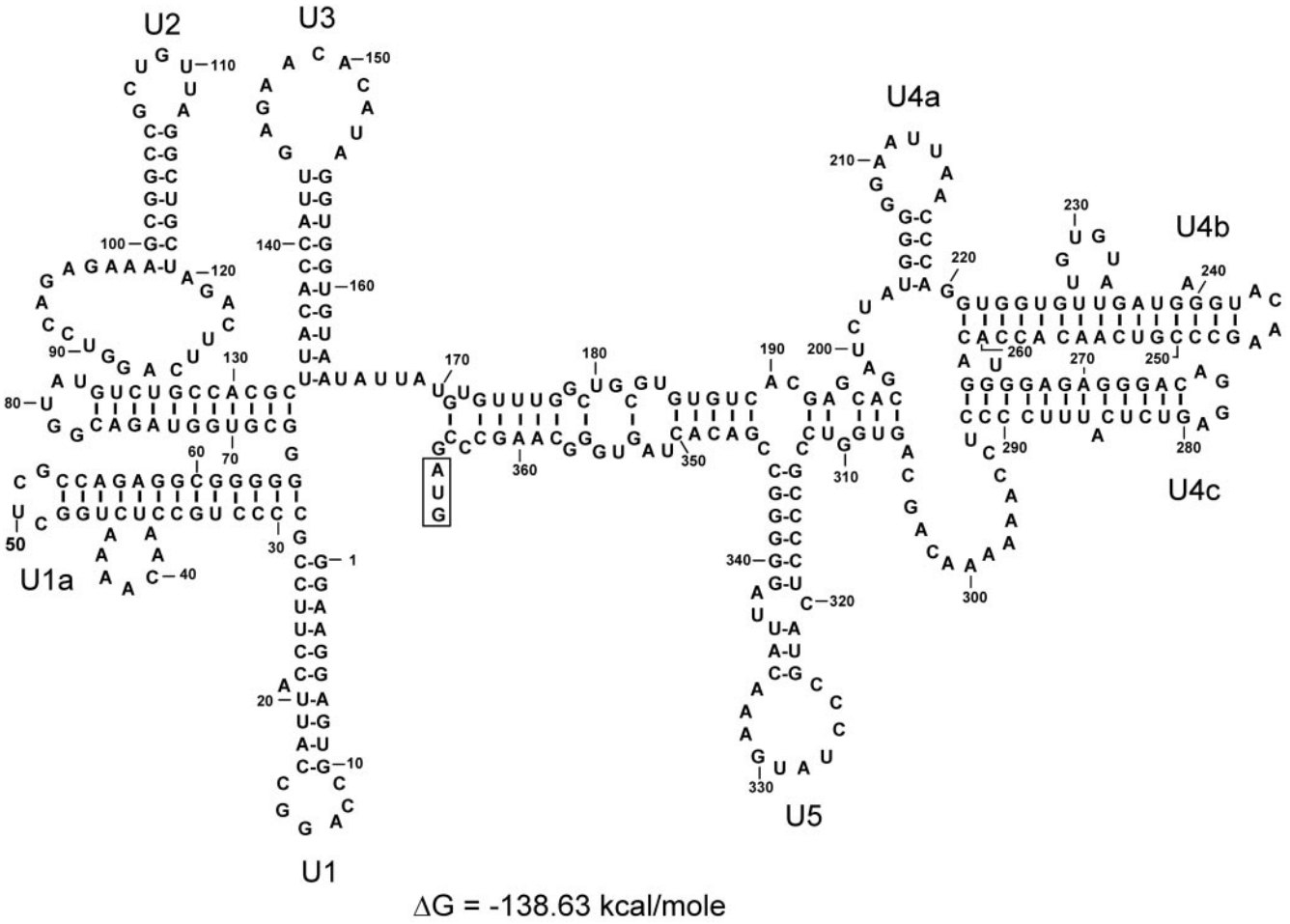

FIG. 1. Optimal secondary structure of the 5′-UTR of GLV mRNA predicted by the minimum free-energy minimization program MFOLD (14, 30). The stem-loop structures were designated U1 to U5. The boxed region indicates the initiation codon at the downstream end of 5′-UTR.

DMS (12 μl, diluted 1:12 in ethanol) was added to the renatured RNA in 300 μl of HMK buffer, incubated at 37°C for 0, 5, and 10 min, and stopped with 75 μl of 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% (vol/vol) ethanol was added to the renatured RNA in 270 μl of HMK buffer, incubated as described above, and stopped by adding 25 mM potassium borate (pH 7.2). CMCT modification was carried out by mixing the RNA sample with an equal volume (150 μl) of freshly prepared CMCT (42 mg/ml) in BMK buffer, incubating, and stopping as described for DMS modification. The treated RNA samples were precipitated in 2.5 volumes of ethanol with 0.3 M sodium acetate, redissolved in nuclease-free water, and extracted once with phenol-chloroform and twice with chloroform-isoamyl alcohol. The extracted aqueous phase was precipitated with 2.5 volumes of ethanol and 0.3 M sodium acetate. The chemically modified RNA was then used as template in a primer extension reaction to identify the chemically modified bases.

Enzymatic probing of RNA structure. Enzymatic probing was carried out essentially as described previously (3, 5, 15). Approximately 5 μg of the RNA sample was initially denatured at 65°C for 15 min in a probing buffer (80 mM HEPES [pH 7.5], 50 mM KCl, 10 mM $MgCl_2$) followed by a slow cooling to ambient temperature for an hour. RNase V1 (Ambion) and RNase T1 (Ambion) were each serially diluted in the probing buffer, titrated (in units) to identify the optimal concentrations for RNA probing, and finalized at 0.1 to 0.2 units of RNase V1 and 1 to 2 units of RNase T1 for the subsequent structure probing experiments. The enzymatic digestions, performed in a final volume of 100 μl at 37°C for a period of 20 min, were stopped by adding phenol-chloroform, and the digested RNA was extracted and recovered by ethanol precipitation in the presence of 0.3 M NaOAc and 10 μg of yeast tRNA. The pellets were dissolved in diethyl pyrocarbonate-treated water and subjected to primer extension analysis to determine the enzyme-cleaved sites in the RNA molecule.

Primer extension. Primer extension was carried out as previously described (15). Four ^{32}P-end-labeled primers, complementing nucleotides 194 to 211, 242 to 259, 304 to 322, and 369 to 386 in the RNA, were each annealed to 5 μg of an RNA sample 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 units of M-MLV reverse transcriptase (Invitrogen). The radiolabeled products were analyzed by 8% denaturing polyacrylamide gel electrophoresis. Sequencing ladders generated by the fmol-cycle sequencing system (Promega) were included as a reference. The chemically modified bases were each identified as a reverse transcription stop with a higher mobility 1 nucleotide short of that in the corresponding DNA sequencing gel, because primer extension would stop in front of the modified base.

Site-directed mutagenesis. Plasmid construct pC631-luc has the full-length GLV cDNA cloned into a pGEM-T vector and a full-length luciferase gene inserted between nt 631 and 4256 of the cDNA (28). The 631-nt sequence upstream from the luciferase gene thus consists of the entire 367-nt GLV 5′-UTR and the downstream 264-nt capsid encoding region. The luciferase gene is fused in-frame with the upstream 631-nt fragment (28). Site-directed mutagenesis of the 5′-UTR in the fragment was carried out essentially as described previously (4) using a QuikChange site-directed mutagenesis kit (Stratagene). Individual mutations were verified by DNA sequencing.

In vitro transcription. pC631-luc and its mutants were each linearized with NruI at the 3′ end of GLV cDNA and used as a template for in vitro synthesis of transcripts using a MegaScript T7 transcription kit (Ambion).

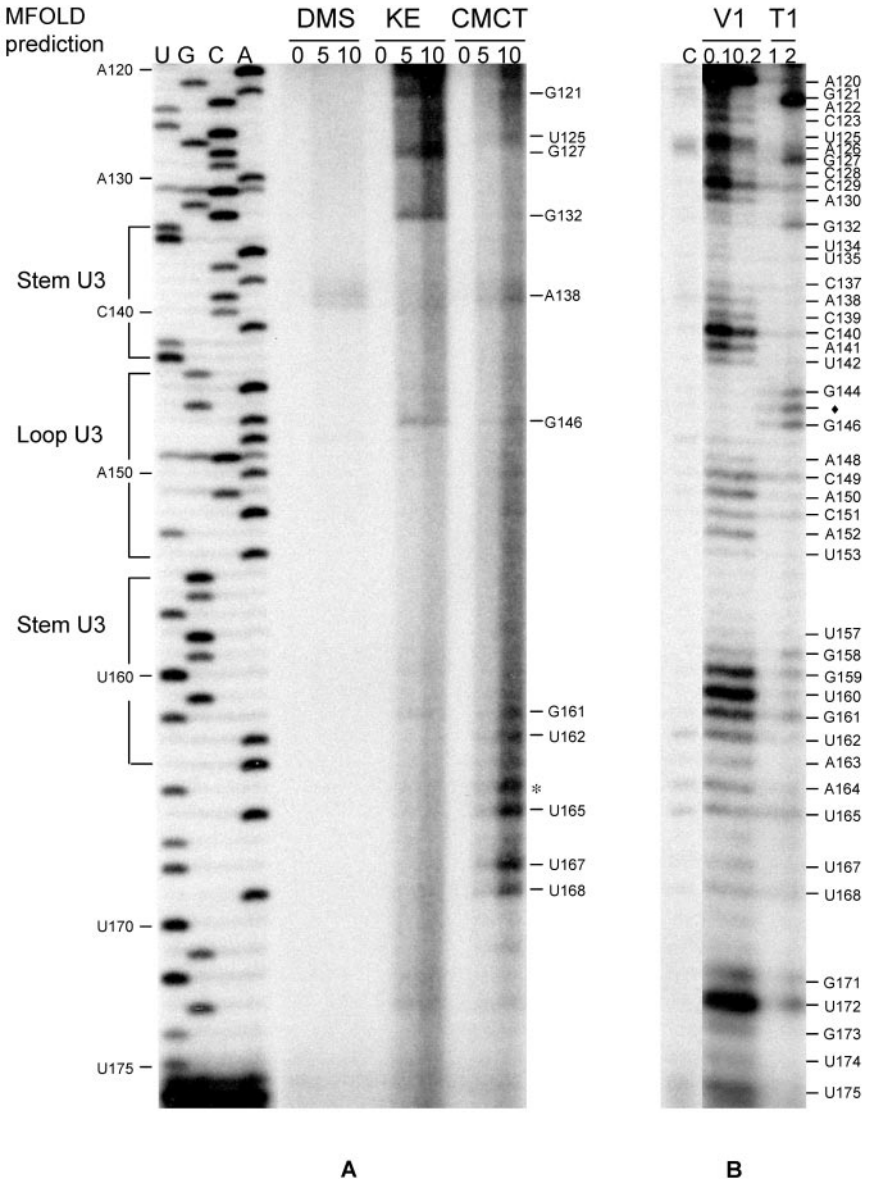

FIG. 2. Structural probing of stem-loop U3 and the pseudoknot structure with chemical modification and enzymatic digestion. (A) Chemical modification of A and C (by DMS), G (by KE), and U and G (by CMCT) was monitored by reverse transcription with a radiolabeled primer hybridizing to positions 194 to 211 in the 5′-UTR. Durations for chemical modification (in minutes) are indicated above each lane. Bases indicated on the right side of the gel are chemically modified. (B) Products from RNaseV1 (V1) and RNaseT1 (T1) digestion were analyzed in primer extension as described above. The units of RNase used are indicated by the numbers above each lane. Bases identified on the right side of the figure indicate points of digestion by RNaseV1 and/or RNase T1. ♦, an unusual digestion of A residue by RNase T1; *, CMCT-modified A. DNA ladders on the left are for base identification.

Transfection of *Giardia* trophozoites. The in vitro transcripts were each introduced into a GLV-infected WB strain of *G. lamblia* trophozoites (WBI) by electroporation as described previously (4, 27). Approximately 4×10^6 trophozoites were transfected with 100 μg of the in vitro transcript. Each transcript was used in triplicate in every duplicated transfection experiment.

Luciferase assay. The transfected *G. lamblia* trophozoites were lysed and assayed for luciferase activity 16 h posteletroporation as described previously (27). Transfectants, in triplicate from two independent transfection experiments, were examined with the pC631-luc transfectant as the positive control. Luciferase activity was calculated in relative light units (RLU) per μg of crude lysate protein as determined by the Bradford method (2).

Northern blot analysis. Total RNA was extracted from transfected *G. lamblia* WBI trophozoites 16 h posttransfection, as described previously (28) and used for Northern blotting following the standard procedures (18). A HindIII/XhoI fragment from pC631-luc containing the *luc* gene sequence (28) was labeled with [α-^{32}P]dCTP using Rediprime II random primer labeling system (Amersham) and used as the probe. Hybridization was carried out at 42°C for 12 h, and the blots were washed under high stringency followed by autoradiography with an exposure time of from 12 to 72 h (18). The intensity of hybridization was determined using a Storm phosphorimager scanner (Amersham Biosciences) and normalized with the 16S rRNA loads (determined using an AlphaImager densitometer) with the ethidium bromide-stained gels.

RESULTS

Probing the secondary structures in the 5′-UTR of GLV IRES. To verify the presence of MFOLD-predicted secondary

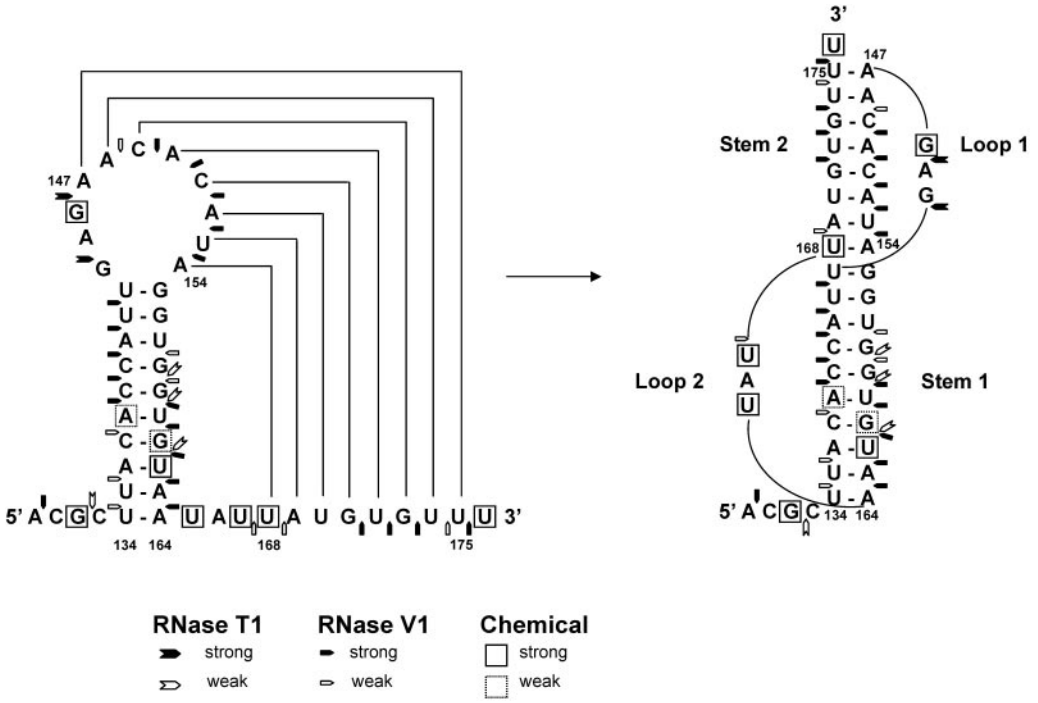

FIG. 3. The proposed pseudoknot structure. Chemically modified bases are boxed, the sites of strong RNaseV1 digestion are indicated by pentagons, and strong RNaseT1 digestion is indicated by arrows.

structure U3, U4a, U4b, U4c, and U5 in GLV IRES, a 631-nt RNA molecule containing the IRES (i.e., the 367-nt 5′-UTR and the downstream 264-nt capsid coding region) was synthesized in vitro for chemical and enzymatic probing. The RNA was incubated with DMS, which methylates unpaired A residues at N-1 and unpaired C residues at N-3; KE, which modifies unpaired G residues at N-1; and CMCT, which modifies the N-3 of unpaired U and the N-1 of unpaired G residues (15). Similarly, the RNA was incubated with ribonucleases RNase V1, which cleaves the nucleotides in base-paired or stacked conformation, and RNase T1, which cleaves unpaired G residue on the 3′ side (3). The site of chemical modification or enzymatic cleavage was identified as a reverse transcription stop in subsequent primer extension analysis.

The predicted stem-loop U3 (nt 134 to 164) has a 10-bp stem and an 11-nt loop. Nucleotides 134 to 143 and 155 to 164 in the stem were resistant to chemical modification (Fig. 2A), except for A138, G161, and U162 located near the bottom of the stem. The predicted stem structure was also sensitive to hydrolysis by RNase V1, but the mid-portion of it appeared more susceptible (Fig. 2B), suggesting that stem U3 is present in the RNA molecule as predicted, but base pairings seem somewhat weaker toward the two ends of the stem.

In the predicted loop U3 (nt 144 to 154), however, all of the residues except G146 were found to be resistant to chemical modification (Fig. 2A), and the nucleotides 148 to 153 in the loop were also hydrolyzed by RNase V1 (Fig. 2B). This suggests that the loop is apparently involved in Watson-Crick base pairing with another unidentified complementary sequence. An inspection of the region surrounding stem-loop U3 revealed a downstream 8-nt region (168 to 175) that could form Watson-Crick base pairings with the 8-nt sequence 147 to 154 in loop U3, resulting in a pseudoknot structure (Fig. 3). Residues 169 to 175 were found to be resistant to chemical modification (Fig. 2A), whereas strong RNaseV1 hydrolysis of G171 to U175 and weak hydrolysis of U168 was observed (Fig. 2B). Thus, this postulated pseudoknot structure may indeed exist in the RNA molecule. According to conventional pseudoknot nomenclature (7, 21, 22), this structure should be a classical H-type pseudoknot (Fig. 3) with stem U3 as stem 1 and the stem formed between nt 147 to 154 and 168 to 175 as stem 2 (Fig. 3). The two stems are joined by loop 1 (G144 to G146) and loop 2 (U165 to U167). RNase T1 hydrolysis of G144 and G146 in loop 1 (Fig. 2B) and CMCT modification of U165 and U167 in loop 2 (Fig. 2A) both supported the single-stranded nature of these two loop structures.

Downstream of the pseudoknot structure, a predicted stem-loop U4a (nt U204 to A219) with a short (4-bp) stem and a large (8-nt) loop (Fig. 1), was subjected to chemical probing and enzyme digestion. The results showed that all the bases in the stem-loop, other than C216 to C218, were significantly modified by chemicals (Fig. 4A) and weakly hydrolyzed (residues U204 and G206 to G208) by RNase V1, except for C216 to C217, which were strongly hydrolyzed by the enzyme (Fig. 4B). The strong chemical modification and weak hydrolysis by RNase V1 could only suggest that U4a is probably in a stacked (as RNase V1 also digests bases in stacked conformation) but unpaired conformation (see Fig. 8). Lack of chemical modification of C216 to C218 and significant RNase V1 hydrolysis of C216 to C217 suggest that they are base pairing with other bases not yet identified in the RNA molecule (Fig. 4A and B).

The predicted stem U4b (nt 221 to 242 and 247 to 261) exhibited significant resistance to chemical modification except for the predicted bulge (nt 228 to 233), in which U230, G231,

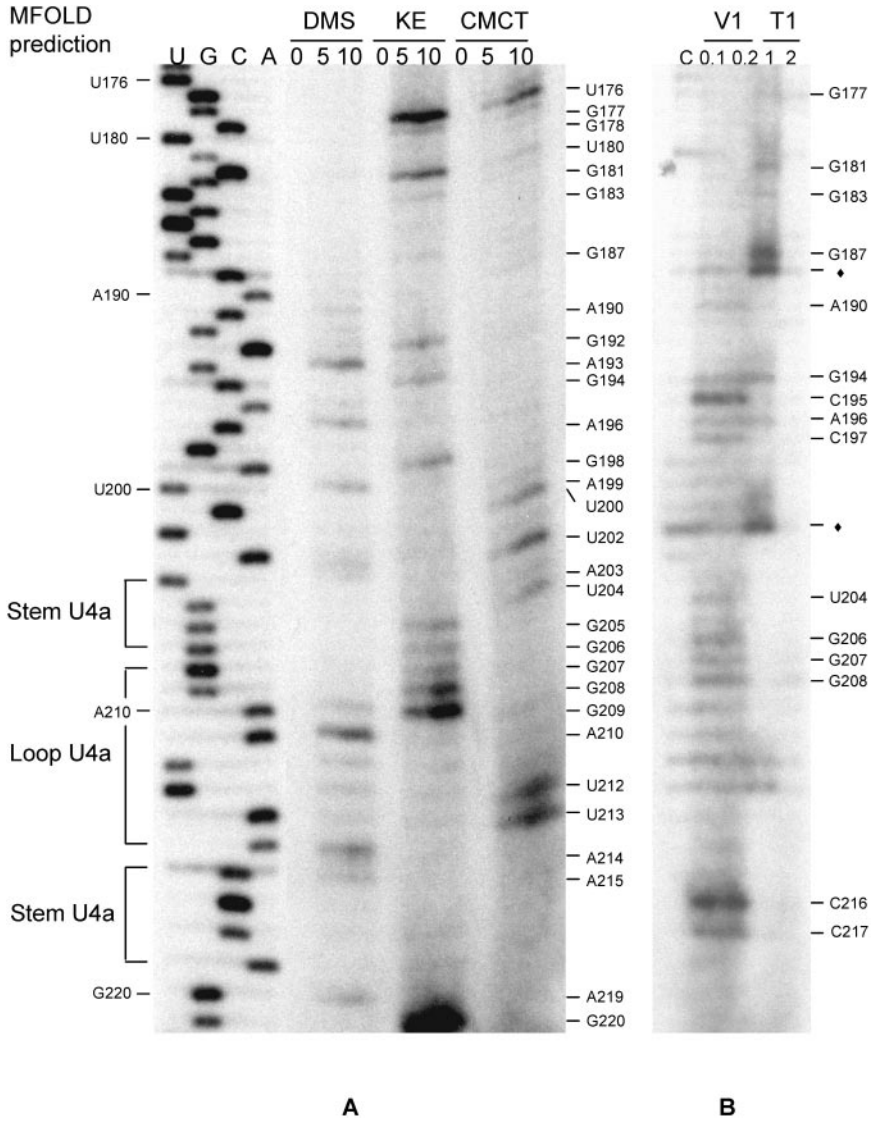

FIG. 4. Chemical modification and enzymatic digestion of the predicted stem-loop U4a. Chemical modification (A) and enzyme digestion (B) were monitored by reverse transcription with a radiolabeled primer hybridizing to positions 242 to 259 in the 5′-UTR. Bases indicated on the right sides of the figures are modified by chemicals (A) or digested by ribonucleases (B). The symbol ♦ indicates an unusual digestion of U residue by RNase T1. DNA ladders on the left are for base identification.

and U232 were chemically modified (Fig. 5A). The predicted single mismatched base A239 in the stem was also modified by DMS. The stem was sensitive to RNase V1 digestion. Residues G224 to U242 (including the bulge and single mismatch) and C249 to A260 were hydrolyzed by RNase V1, whereas only G231 in the bulge was hydrolyzed by the single-strand specific RNase T1 (Fig. 5B). Residues (A243 to A246) in the predicted loop U4b were modified by chemicals but not digested by RNase V1 (Fig. 5A and B). Overall, the experimental data support the presence of the predicted stem-loop U4b. However, some experimental ambiguities, such as the chemical modification and the RNase V1 digestion of G251 to U252 in the predicted stem (Fig. 5A and B), were used as constraints in a new MFOLD prediction, which resulted in a modified stem-loop U4b with a slightly enlarged loop (U242 to G247) and a shifted bulge from U228–A233 to U230–U236 to accommodate the experimental data (Fig. 8).

The single nucleotide A262 that separates stem-loop U4b from U4c (Fig. 1) was modified by DMS as anticipated (Fig. 6A). In U4c, bases in the loop (nt 276 to 279) were chemically modified, and G277 to G278 were also digested by RNase T1 (Fig. 6A and B). Residues G263 to C275 and G280 to C292 in the stem were hydrolyzed by RNase V1 (Fig. 6B) and were relatively resistant to chemical modification, except for the bulged U265 and A285, which were both chemically modified (Fig. 6A). The presence of stem-loop U4c is thus verified (Fig. 8).

The segment U293 to C313, separating stem-loops U4c and U5, was chemically modified and not hydrolyzed by RNase V1 (Fig. 7A and B), thus indicating its single-stranded nature as

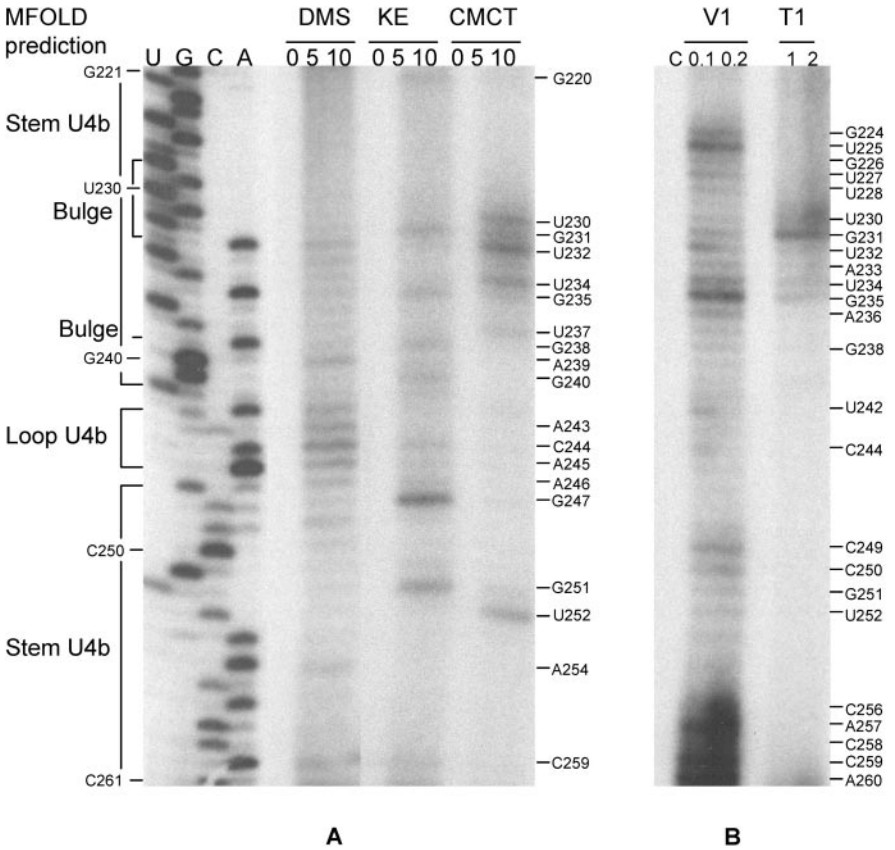

FIG. 5. Chemical modification and enzymatic digestion of the predicted stem-loop U4b. Chemical modification (A) and enzyme digestion (B) were monitored by reverse transcription with a radiolabeled primer hybridizing to positions 304 to 322 in the 5′-UTR. Bases indicated on the right sides of the figures are modified by chemicals (A) or digested by enzymes (B). DNA ladders on the left are for base identification.

predicted (Fig. 1 and 8). For the predicted stem-loop U5, the stem (G314 to G323 and C334 to C344) was unambiguously modified by the chemicals at G314, U319, C320, and U322 on one arm and U336, U337, G342, and G343 on the other (Fig. 7A) and digested by RNase V1 only at C316 to C320 on one side and U337 and C344 to C345 on the other (Fig. 7B). These data suggest considerable wobbling among the base pairings around the predicted bulge in the stem structure. In the assumed U5 loop, residues C325 to A333 were chemically modified, and the residue G330 was digested by RNase T1. However, C325 to U327 were also digested by RNase V1, suggesting a stacked conformation for these residues (Fig. 7A and B). The essence of a stem-loop U5 structure was thus confirmed by the experimental data.

Overall, the experimental data (Fig. 2 to 7) have largely confirmed the MFOLD-predicted secondary structures in the 5′-UTR portion of GLV IRES (Fig. 1) except for stem-loop U4a and the need for a slight modification of U4b (Fig. 8). The data also revealed a potential pseudoknot structure which MFOLD was unable to identify.

Structure-function analysis of stem-loop U3 and the pseudoknot. To analyze the role of stem-loop U3 in IRES function, we deleted the entire U3 stem-loop, which resulted in loss of luciferase translation to 1% of the wild-type control, thus suggesting an essential role of U3 in the IRES (Table 1). Introduction of a bulge into the stem by substituting C139/C140 with G139/G140 or by changing G158/G159 to C158/C159 (Fig. 9) reduced the downstream luciferase expression to 2.1% and 9.2% of the control, respectively (Table 1). The loss was recovered to 139% of the control (Table 1), when the two mutations were combined to restore the stem structure (Fig. 9). However, Northern analysis showed that the C139G/C140G and G158C/G159C mutant transcripts diminished to an undetectable level 16 h posttransfection into *Giardia* (Fig. 10A), whereas this rapid disappearance did not occur to the double mutant C139G/C140G–G158C/G159C transcript (Fig. 10A). The essential role of stem-loop U3 in the IRES is thus probably in maintaining the stability of transcript.

The structure-function of the putative pseudoknot was also analyzed. When the entire pseudoknot region was deleted, luciferase translation was completely abolished and the transcript level was reduced to 20% of the control (Table 1; Fig. 10A). Starting at the 3′ end of the pseudoknot stem, we progressively disrupted the base pairings toward the junction of the two stems by mutagenesis (Fig. 3). An A-to-C mutation (1A) at position 147 and a U-to-G mutation (1B) at position 175 reduced the luciferase expression to 23.9% and 23.5% of the control, respectively (Table 2). The loss was recovered to 86% by restoring the lost base pairing with alternate base-pair G175 · C147 (1R) at the disrupted location (Table 2). The single mutants 1A and 1B showed also significantly diminished transcript stabilities in transfected *Giardia* (Fig. 10B), which

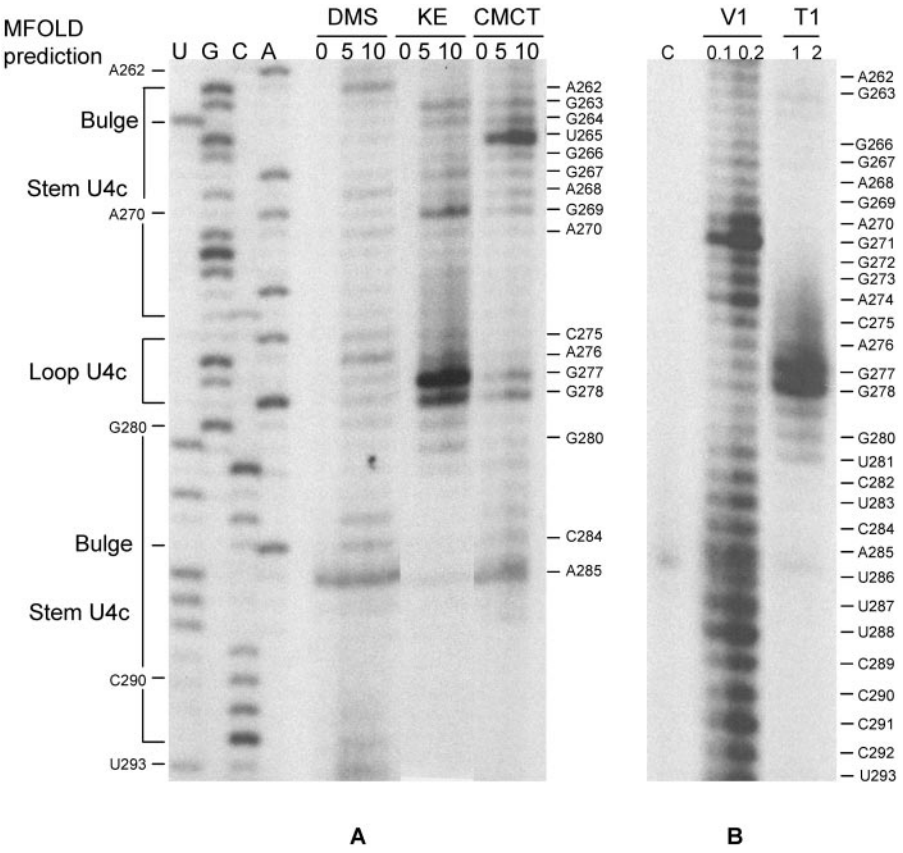

FIG. 6. Chemical modification and enzymatic digestion of the predicted stem-loop U4c. Chemical modification (A) and enzyme digestion (B) were monitored by reverse transcription with a radiolabeled primer hybridizing to positions 304 to 322 in the 5′-UTR. Bases on the right sides of the figures are chemically modified (A) or enzymatically digested (B). DNA ladders on the left are for base identification.

were restored to 50% of the wild-type level by the combined mutations (1R) that presumably reestablished the pseudoknot stem (Fig. 10B). Mutant transcripts A148C/C149A (2A) or G173U/U174G (2B) were translated to 2% and 12% of the wild-type level, while their transcripts were reduced to 25% and 30%, respectively (Table 2 and Fig. 10B). The activity was restored to 125% and transcript level to 130% when the two mutations were combined (2R) (Table 2). Similarly, two triple mutants C149G/A150U/C151G (3A) and G171C/U172A/G173C (3B) had luciferase expression at 2.5% (both) and transcript levels at 25% and 30%, respectively, but were restored to 54% expression and 44% transcript level when combined (3R) (Table 2 and Fig. 10B).

In another set of mutations (Table 2, mutations 4, 5, and 6), however, the outcomes were somewhat different. C151U/A152G (4A) and U170C/G171A (4B) demonstrated decreased translation (2% and 3%, respectively) and transcript levels (12% and 14%, respectively), but the recovered translation (50%) in combined mutations (4R) was not accompanied by a similarly improved transcript stability (16%) (Fig. 10B). Thus, the pseudoknot may have a role in translation initiation in addition to its stabilizing effect on the transcript. In a previous study, a pseudoknot-deleted mutant (nt 126 to 176) of a dicistronic viral transcript was found to be fully stable in the transfected *Giardia*, even though translation of the downstream cistron was significantly compromised (6). Thus, the role of this pseudoknot in maintaining IRES-mediated translation initiation goes beyond stabilizing the mRNA inside *Giardia*.

U153C (5A) and A169G (5B) mutations reduced luciferase expression to 5.5% and 9.7% and transcript level to 23% and 20%, respectively. A restoration of base-pairing at the 153 · 169 position, however, resulted in only 3.1% translation and 20% of the transcript level (Table 2; Fig. 10B). Another compensatory mutation, A154C/U168G (6R), also failed to restore the severely compromised luciferase translation and transcript stability exhibited by the single mutants 6A and 6B (Table 2; Fig. 10B). The discrepancy between these two instances and the previous results suggests that the end of the pseudoknot stem near the junction with stem U3 requires the presence of both the stem structure and the original sequence. These two particular base pairs at the end of the pseudoknot are thus not only base specific but also polarity specific.

Structure-function analysis of U4a. To test the potential role of U4a in translation initiation, the entire U4a region was removed (Fig. 9), which reduced translation to 7.7% of the wild-type level without affecting transcript stability (Table 1; Fig. S1 in the supplemental material). Mutation G205C/G206C (Fig. 9) reduced translation of the transcript to 17.3% without compromising stability of the transcript. The presumed compensatory mutation G205C/G206C–C217G/C218G further reduced luciferase expression to 4.8% while the transcript remained stable (Table 1; Fig. S1 in the supplemental material).

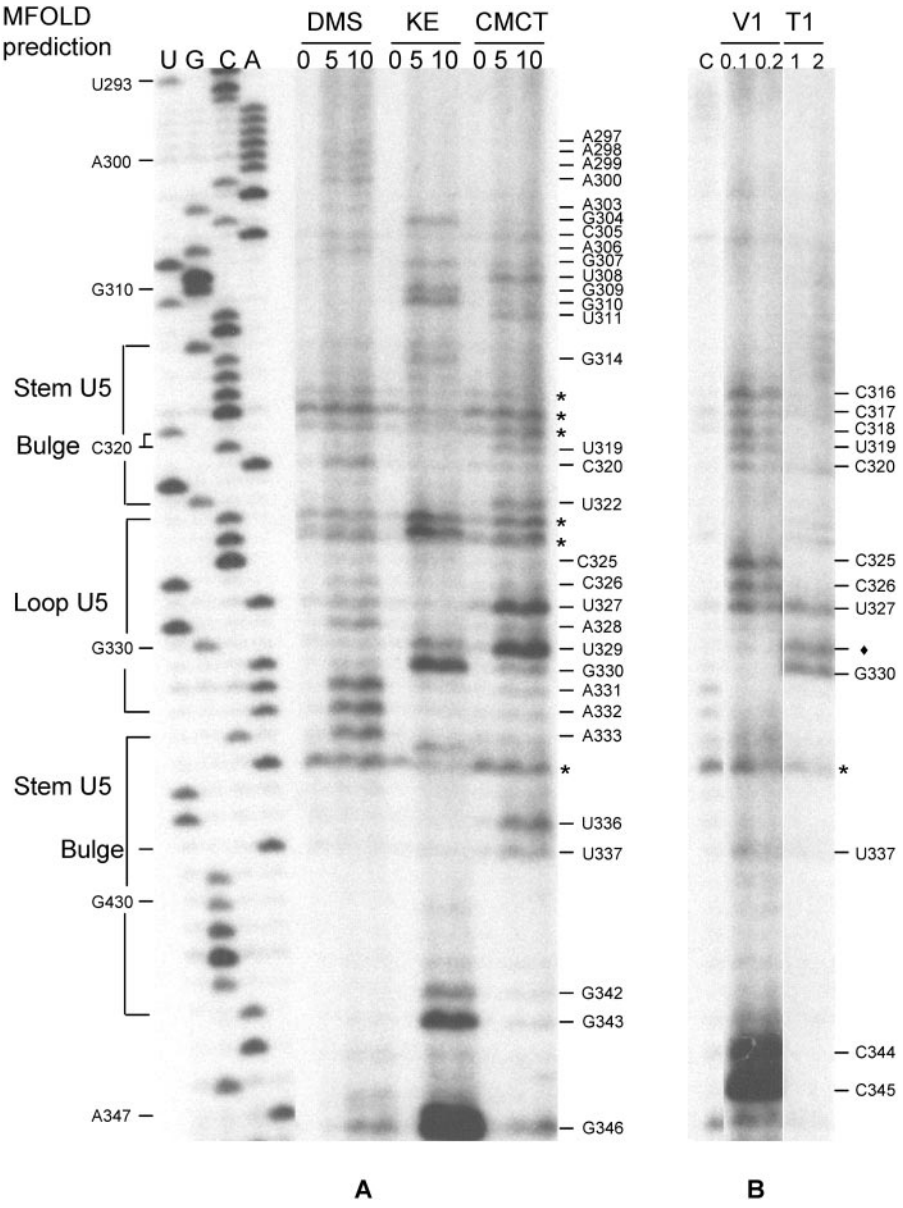

FIG. 7. Chemical and enzymatic probing of the predicted stem-loop U5. Chemical modification (A) and enzyme digestion (B) were monitored by reverse transcription with a radiolabeled primer hybridizing to positions 369 to 386 in the capsid coding region. Bases on the right sides of the figures are chemically modified (A) or enzymatically digested (B). The symbol (♦) indicates an unusual digestion of U residue by RNase T1. The ladders on the left are for base identification. The sensitivity of some of the residues (C315 to C318, G323 to C324, and C332) to chemical modification cannot be resolved due the presence of background in the control lanes (*). They have been attributed to strong structural features that cause the premature termination of reverse transcription (20).

It confirms the outcome from previous structure analysis that stem-loop U4a does not exist (Fig. 4 and 8), whereas the sequence in a certain part of the region may play a role in the IRES function. To further test this possibility, a C217A mutation was introduced to disrupt the hypothetical stem (Fig. 9), but it resulted in no apparent loss of luciferase translation (Table 1). To identify the boundaries of this sequence-specific region, a series of base substitutions were made from G206 to U213. A significant drop in translation was observed upon mutations in the G206-to-A211 region (Table 1), indicating that the G205-to-A211 sequence is essential for translation initiation.

Structure-function analysis of stem-loops U4b and U4c. When the entire stem-loop U4b was deleted, luciferase translation was decreased to 12.6% of the control but the transcript remained stable (Table 1; Fig. S1 in the supplemental material). The introduction of a bulge into the lower U4b stem (U225C) resulted in only a partial loss of activity, to 58.3% of that of the control (Fig. 9; Table 1). A similar minimal loss of translation activity was observed with U237A/G238C and

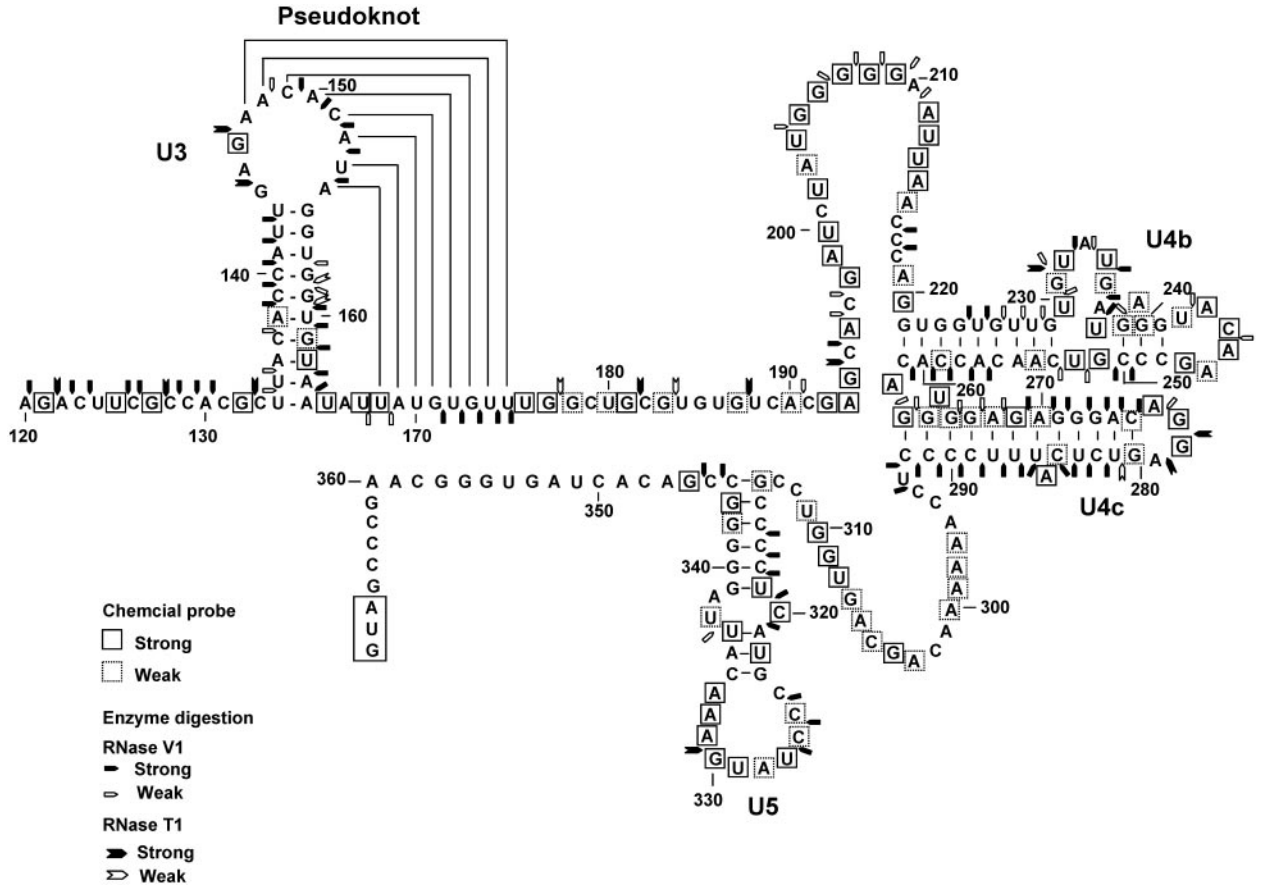

FIG. 8. The secondary structure of GLV mRNA 5′-UTR derived from the experimental results shown in Fig. 2 to 7. Numbers indicate nucleotide positions in the 5′-UTR. Chemically modified bases are boxed, sites of RNaseV1 digestion are indicated by pentagons, and sites of strong RNaseT1 digestion are indicated by arrows.

G240C/G241C in the upper stem region above the single stranded bulge (Table 1), suggesting that the stem structure is not essential for activity. However, mutations in the loop C244U/A245G and in the upper stem regions C248G/C249G and C250U/G251A resulted in significant reductions in translation, to 11.7%, 7%, and 10% of the control, respectively, without affecting transcript stability (Table 1; Fig. S1 in the supplemental material), suggesting that a particular sequence in C244 to G251 may play an important role in translation initiation.

A similar observation was made on stem-loop U4c. Upon its deletion (Fig. 9), the transcript was translated at 30.5% of the control while remaining stable (Table 1; Fig. S1 in the supplemental material), whereas creation of a large bulge in the stem in the G271C/G272C/G273C mutant resulted in 88.6% translation efficiency (Fig. 9; Table 1).

Structure-function analysis of stem-loop U5. A complete removal of stem-loop U5 from the transcript resulted in 0.1% of the wild-type translation efficiency, whereas the transcript remained stable (Table 1; Fig. S1 in the supplemental material). C316G/C317G or G341C/G342C (Fig. 9), aimed at disrupting the stem structure, led to a drastic decrease of translation, down to 0.04% and 0.3% of the control, without affecting transcript stability. These losses were recovered to 92.6% by combining the two mutations to restore the stem structure (Fig. 9; Table 1). Since U5 is the stem-loop immediately upstream from the initiation codon, it may play a more pivotal role in initiating translation of the downstream transcript than the other secondary structures.

DISCUSSION

Structural analysis and site-directed mutagenic dissections have given us insights into the structure and function of the 5′-UTR portion of the GLV IRES. The structure starts near A120, excluding the upstream stem-loops U1, U1a, and U2 (6). Stem-loop U3 lies 13 nucleotides downstream from A120, with a portion of its loop associating with a downstream sequence to form a pseudoknot. This structural complex is apparently important in stabilizing the transcript in vivo as well as in initiating translation. U4a does not exist as a stem-loop but it contains an essential sequence for IRES function. The structures of stem-loops U4b and U4c may not have an important function, but a sequence in part of the loop and the upper stem in U4b may play an important role in the IRES-mediated translation. Stem-loop U5 appears to be a crucial structure in the IRES, with possible direct involvement in initiation factor binding and ribosome recruitment.

TABLE 1. Relative luciferase activities of mutant transcript-transfected *Giardia* and the stability of the mutant transcripts

Stem-loop and mutation	RLU (%)[a]	Remark	Stability[c]
Stem-loop U3	100		+
C139G/C140G	2.1 ± 0.2	Wobbled stem	−
G158C/G159C	9.2 ± 4.2	Wobbled stem	−
C139G/C140G–G158C/G159C[b]	139.0 ± 17.5	Restored stem	+
Δ 134–164	1.0 ± 0.2	Entire U3 deleted	−
Δ 134–175	1.2 ± 0.14	Entire pseudoknot deleted	−
Stem-loop U4a	100		+
G205C/G206C	17.3 ± 0.9	Wobbled stem	+
G205C/G206C–C217G/C218G[b]	4.8 ± 0.2	Restored stem	+
C217A	94.6 ± 5.4	A bulge in stem	ND
G206C/G207C	2.7 ± 0.6	Disrupted stem	ND
G208C/G209C	1.0 ± 0.05	Loop mutation	+
A210C/A211C	17.2 ± 0.6	Loop mutation	ND
U212C/U213C	54.7 ± 1.5	Loop mutation	ND
Δ 204–219	7.7 ± 0.8	Entire U4a deleted	+
Stem-loop U4b	100		+
U225C	58.3 ± 6.0	Shortened stem	ND
U237A/G238C	63.5 ± 4.0	Loop mutation	ND
G240C/G241C	54.4 ± 5.7	Loop mutation	ND
C248G/C249G	7.4 ± 0.2	Loop mutation	+
C250U/G251A	10.6 ± 0.5	Loop mutation	ND
C244U/A245G	11.7 ± 2.0	Loop mutation	+
Δ 221–261	12.6 ± 0.8	Entire U4b deleted	+
Stem-loop U4c	100		+
G271C/G272C/G273C	88.6 ± 6.6	Wobbled stem	ND
Δ 263–292	30.5 ± 2.1	Entire U4c deleted	+
Stem-loop U5	100		+
C316G/C317G	0.04 ± 0.01	Wobbled stem	+
G341C/G342C	0.3 ± 0.1	Wobbled stem	+
C316G/C317G–G341C/G342C[b]	92.6 ± 3.4	Restored stem	+
Δ 314–344	0.1 ± 0.02	Entire U5 deleted	+

[a] Values are averages of relative luciferase activities obtained from at least two independent transfection experiments with the wild type (pC631luc) as the positive control (100%). The wild type and the mutants were all transfected in triplicate in each experiment.
[b] Restorative mutations.
[c] Stability of the transcripts as determined by Northern blot analysis. −, loss of the transcript; +, presence of the transcript. See Fig. 10 for stem-loop U3 and Fig. 1 in the supplemental material for stem-loops U4a to U5. ND, not determined.

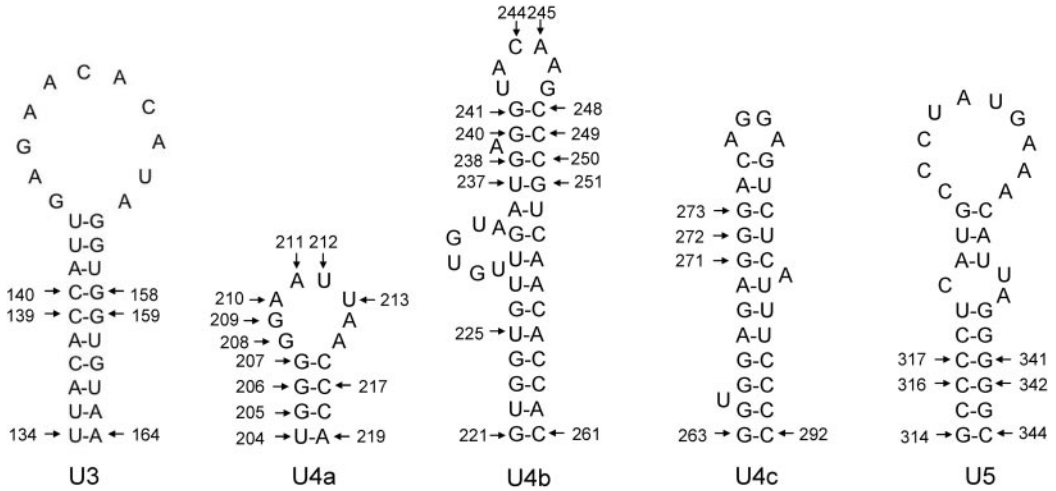

FIG. 9. Structures of individual MFOLD-predicted stem-loops U3, U4a, U4b, U4c, and U5 as indicated in Fig. 1, with arrows indicating the positions of individual site-directed mutations. Results from analyzing these mutants are presented in Table 1.

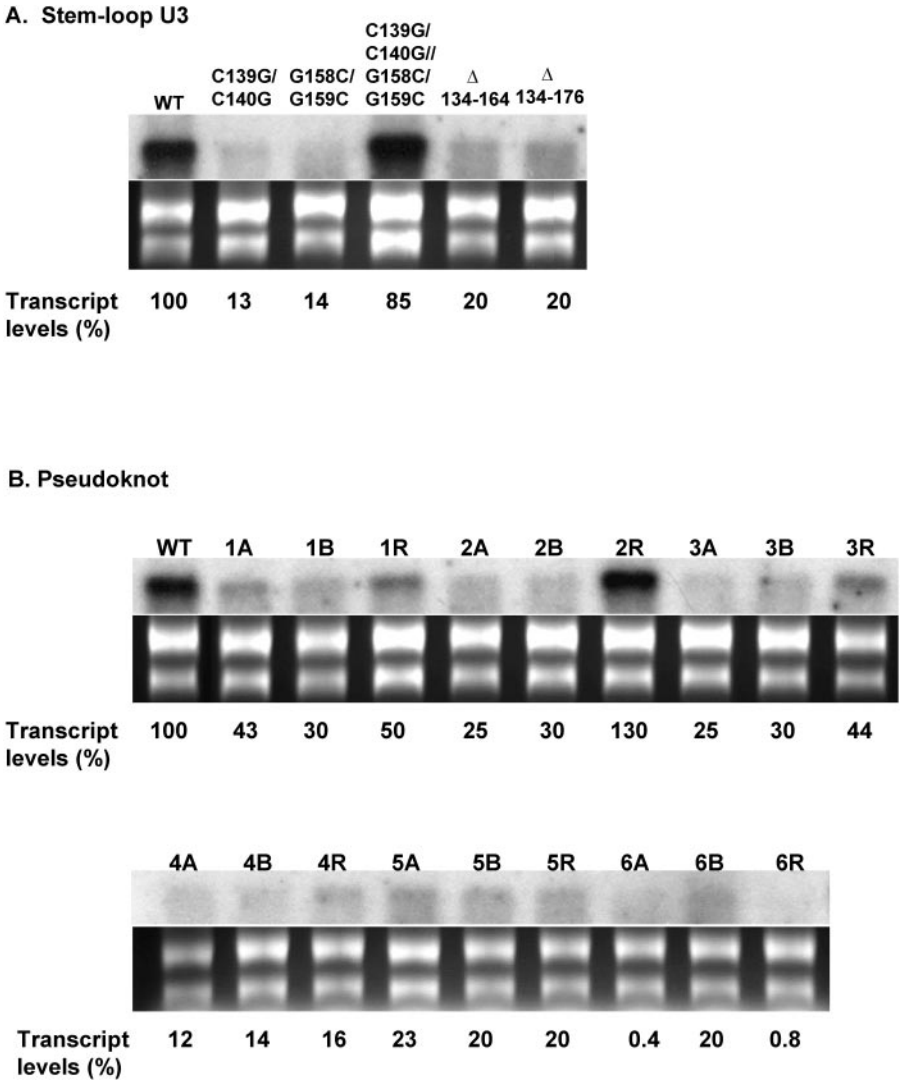

FIG. 10. Northern blot analysis showing varied stabilities among the stem-loop U3 (A) and pseudoknot mutant (B) transcripts in *Giardia* cells. Total RNA was extracted from transfected *Giardia* trophozoites 16 h after electroporation and analyzed by Northern blotting using [α-^{32}P]-*luc* DNA as a probe. The same samples, stained with ethidium bromide in gel, were used as sampling controls. Mutations are indicated on the top of each lane for stem-loop U3 mutants (A) and pseudoknot mutants (B) (see Table 2). The transcript level determined for each mutant (see Materials and Methods) is indicated on the bottom of each lane.

Pseudoknot structures in RNA molecules are known to play important functions in mRNA translation including ribosomal frame-shift (9, 12), read-through of the *gag*-termination codon (1), and IRES-mediated initiation of translation (17, 19, 25, 26). A unique feature of the pseudoknot identified in this study is its role in IRES function as well as in stabilization of the viral transcript. Discrepancies between these two effects of the pseudoknot are apparent. For instance, the recovery mutants 1R and 4R in Table 2 show a significant increase in translation efficiency but did not show a corresponding increase in the message levels that was close to being as great as that for the single mutants 1A, 1B, 4A, and 4B. Apparently, the recovery mutants with altered base pairings are sufficient for initiating translation but not for stabilizing the transcript. Another interesting feature of the pseudoknot lies in the base-specific and orientation-specific pairs U168 · A154 and A169 · U153 at the U3 stem-pseudoknot stem junction. A precedent of it was observed in the pseudoknot involved in ribosomal frame-shift in beet western yellow virus mRNA, in which the U13 · A25 base pair at the helical junction could not be replaced even with the pseudoknot structure maintained (9).

It has been proposed that viral IRESs have a modular organization. Each module, while inactive on its own, performs a precise function in concert with the others (16). For instance, domains in the 3′ end of FMDV IRES bind to translation initiation factors, while the 5′ and central domains maintain the overall architecture of the IRES (16). Disruption of the structures in the latter by point mutations, deletions, or insertions inactivates the IRES by destroying essential RNA-protein interactions (13) and/or long-range RNA-RNA interactions of the IRES (16). Thus, the overall RNA secondary structure of IRES needs to be maintained in order to retain its

TABLE 2. Relative luciferase activities expressed by pseudoknot mutant transcript-transfected *Giardia*[a]

Mutation	nt 147–154	nt 168–175	RLU (%)[b]	Stability[c]
None (wild type)	AACACAUA	UAUGUGUU	100	+
1A	**c**ACACAUA	UAUGUGUU	23.9 ± 2.8	−
1B	AACACAUA	UAUGUG**g**	23.5 ± 1.7	−
1R	**c**ACACAUA	UAUGUG**g**	86.5 ± 8.2	+
2A	A**ca**ACAUA	UAUGUGUU	2.1 ± 0.3	−
2B	AACACAUA	UAUGU**ug**U	12.4 ± 0.8	−
2R	A**ca**ACAUA	UAUGU**ug**U	125.1 ± 1.6	+
3A	AA**gug**AUA	UAUGUGUU	2.5 ± 1.6	−
3B	AACACAUA	UAU**cac**UU	2.5 ± 0.1	−
3R	AA**gug**AUA	UAU**cac**UU	54.3 ± 2.1	+
4A	AACA**ug**UA	UAUGUGUU	1.9 ± 0.2	−
4B	AACACAUA	UA**ca**UGUU	2.7 ± 0.1	−
4R	AACA**ug**UA	UA**ca**UGUU	49.7 ± 3.7	−
5A	AACACA**c**A	UAUGUGUU	5.5 ± 0.7	−
5B	AACACAUA	U**g**UGUGUU	9.7 ± 0.5	−
5R	AACACA**c**A	U**g**UGUGUU	3.1 ± 0.3	−
6A	AACACAU**c**	UAUGUGUU	1.3 ± 0.1	−
6B	AACACAUA	**g**AUGUGUU	0.2 ± 0.1	−
6R	AACACAU**c**	**g**AUGUGUU	0.6 ± 0.1	−

[a] The letters in boldface lowercase are the substituted bases.
[b] Each value is an average from at least two independent transfection experiments with the wild type (pC631luc) as positive control (100%). The wild type and the mutants were all used to transfect cells in triplicate in each experiment.
[c] Stability of the transcripts as determined by Northern blot analysis. −, loss of the transcript; +, presence of the transcript. See Fig. 10 for details.

function. Our current structural study suggests that GLV IRES could have a similar modular organization. In the 5′-UTR, stem-loop U3 and the pseudoknot could function by maintaining the architecture of the 5′-UTR in IRES. It could also bind to protein factors essential for translation initiation. The essential sequence elements in the U4a region and stem-loop U4b could also function by binding to translation initiation factors. Stem-loop U5 and the downstream stem-loop I in the coding region are separated by 31 nucleotides with the initiation codon located at the very center (6). Previous studies indicated that the location of this initiation codon cannot be shifted at all, and a move upstream or downstream by a mere three nucleotides completely abrogated translation initiation (5, 6). Since the length of 31 nucleotides is roughly equivalent to the diameter of a 40S ribosomal subunit, it is possible that the latter is recruited by the IRES to be located precisely between stem-loops U5 and I without any further ribosomal scanning. The lack of ribosomal scanning in translation initiation was recently demonstrated in *G. lamblia* (11). A simple structure of ^{m7}cap-AUG at the 5′-end of a mRNA is sufficient for initiating translation. A short 5′-UTR is tolerated up to a length of 9 nucleotides. Further extension of the 5′-UTR reduces the efficiency of translation initiation drastically.

Further downstream from stem-loop I are a DB forming base pairs with the 3′-end of the 16S-like rRNA in *Giardia*, stem-loops II, III, and IVA, a pseudoknot between loop II, and a downstream sequence, each performing an essential role in GLV IRES (4, 5). The entire length of GLV IRES extends from −253 to 264, totaling 517 nucleotides (Fig. 11). It is a unique IRES, because it expands across both the 5′-UTR and the coding region. We postulate that such an unusual IRES structure is required because of the absence of ribosomal scan-

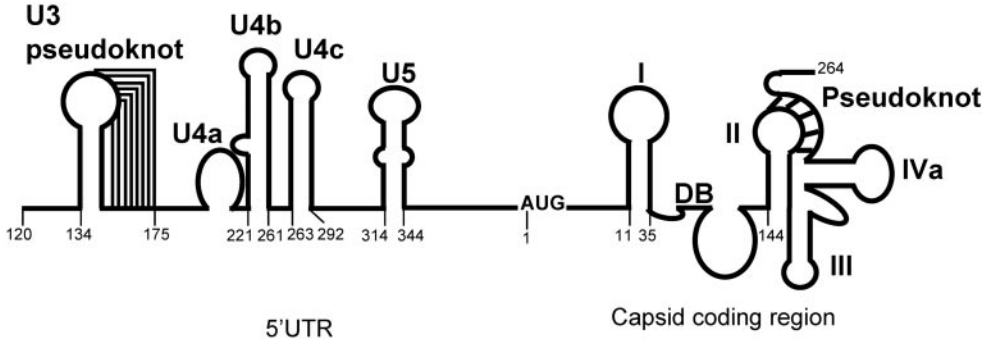

FIG. 11. Secondary structure of GLV IRES verified by experimental data.

ning in the translation initiation in *Giardia*. GLV IRES will be a useful tool for dissecting the detailed mechanism of translation initiation in this primitive eukaryote.

ACKNOWLEDGMENTS

We thank Potter Wickware for critical reading of the manuscript.

This work was supported by grant AI-30475 from the National Institutes of Health.

REFERENCES

1. **Alam, S., N. M. Wills, J. A. Ingram, J. F. Atkins, and R. F. Gesteland.** 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.
2. **Bradford, M. M.** 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.
3. **Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J. P. Ebel, and B. Ehresmann.** 1987. Probing the structure of RNAs in solution. Nucleic Acids Res. **15:**9109–9128.
4. **Garlapati, S., J. Chou, and C. C. Wang.** 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.
5. **Garlapati, S., and C. C. Wang.** 2002. Identification of an essential pseudoknot in the putative downstream internal ribosome entry site in giardiavirus transcript. RNA **8:**601–611.
6. **Garlapati, S., and C. C. Wang.** 2004. Identification of a novel internal ribosome entry site in giardiavirus that extends to both sides of the initiation codon. J. Biol. Chem. **279:**3389–3397.
7. **Hilbers, C. W., P. J. Michiels, and H. A. Heus.** 1998. New developments in structure determination of pseudoknots. Biopolymers **48:**137–153.
8. **Joseph, S., and H. F. Noller.** 1998. EF-G-catalyzed translocation of anticodon stem-loop analogs of transfer RNA in the ribosome. EMBO J. **17:**3478–3483.
9. **Kim, Y., L. Su, S. Maas, A. O'Neill, and A. Rich.** 1999. Specific mutations in a viral pseudoknot drastically change ribosomal frameshifting efficiency. Proc. Natl. Acad. Sci. USA **96:**14234–14239.
10. **Li, L., A. L. Wang, and C. C. Wang.** 2001. Structural analysis of the −1 ribosomal frameshift elements in giardiavirus mRNA. J. Virol. **75:**10612–10622.
11. **Li, L., and C. C. Wang.** 2004. Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. J. Biol. Chem. **279:**14656–14664.
12. **Liphardt, J., S. Napthine, H. Kontos, and I. Brierley.** 1999. Evidence for an RNA pseudoknot loop-helix interaction essential for efficient −1 ribosomal frameshifting. J. Mol. Biol. **288:**321–335.
13. **Lopez de Quinto, S., and E. Martinez-Salas.** 2000. Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation *in vivo*. RNA **6:**1380–1392.
14. **Mathews, D., J. Sabina, M. Zuker, and D. Turner.** 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. **288:**911–940.
15. **Moazed, D., S. Stern, and H. F. Noller.** 1986. Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. J. Mol. Biol. **187:**39916.
16. **Ramos, R., and E. Martinez-Salas.** 1999. Long-range RNA interactions between structural domains of the aphthovirus internal ribosome entry site (IRES). RNA **5:**1374–1383.
17. **Rijnbrand, R., T. van der Straatten, P. A. van Rijn, W. J. Spaan, and P. J. Bredenbeek.** 1997. Internal entry of ribosomes is directed by the 5′ noncoding region of classical swine fever virus and is independent of the presence of an RNA pseudoknot upstream of the initiation codon. J. Virol. **71:**451–457.
18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Sasaki, J., and N. Nakashima.** 2000. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. Proc. Natl. Acad. Sci. USA **97:**1512–1515.
20. **Stern, S., D. Moazed, and H. F. Noller.** 1998. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. Methods Enzymol. **164:**481–489.
21. **ten Dam, E. B., K. Pleij, and L. Bosch.** 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. Virus Genes **4:**121–136.
22. **ten Dam, E. B., K. Pleij, and D. Draper.** 1992. Structural and functional aspects of RNA pseudoknots. Biochemistry **31:**11665–11676.
23. **Wang, A. L., and C. C. Wang.** 1986. Discovery of a specific double-stranded RNA virus in *Giardia lamblia*. Mol. Biochem. Parasitol. **21:**269–276.
24. **Wang, A. L., H. M. Yang, K. A. Shen, and C. C. Wang.** 1993. Giardiavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-pol-like fusion protein by translation frameshift. Proc. Natl. Acad. Sci. USA **90:**8595–8599.
25. **Wang, C., S. Y. Le, N. Ali, and A. Siddiqui.** 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.
26. **Wilson, J. E., T. V. Pestova, C. U. T. Hellen, and P. Sarnow.** 2000. Initiation of protein synthesis from the A site of the ribosome. Cell **102:**511–520.
27. **Yu, D. C., A. L. Wang, C. H. Wu, and C. C. Wang.** 1995. Virus-mediated expression of firefly luciferase in the parasitic protozoan *Giardia lamblia*. Mol. Cell. Biol. **15:**4867–4872.
28. **Yu, D. C., and C. C. Wang.** 1996. Identification of *cis*-acting signals in the giardiavirus (GLV) genome required for expression of firefly luciferase in the parasitic protozoan *Giardia lamblia*. RNA **2:**824–834.
29. **Yu, D. C., A. L. Wang, C. W. Botka, and C. C. Wang.** 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.
30. **Zuker, M.** 1989. On finding all suboptimal folding of an RNA molecule. Science **244:**48–52.