THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 15, Issue of April 9, pp. 14656–14664, 2004
Printed in U.S.A.
© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.

Capped mRNA with a Single Nucleotide Leader Is Optimally Translated in a Primitive Eukaryote, *Giardia lamblia**

Received for publication, September 5, 2003, and in revised form, January 13, 2004
Published, JBC Papers in Press, January 13, 2004, DOI 10.1074/jbc.M309879200

Lei Li and Ching C. Wang‡

From the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-2280

The 5′-untranslated region (5′-UTR) of an mRNA plays an important role in translation initiation in eukaryotes. A minimal length of about 20 nucleotides is required to prevent leaky ribosome scanning. In one of the most primitive eukaryotes, *Giardia lamblia*, however, the mRNAs have 5′-UTRs mostly in the range of 0 to 14 nucleotides without a conserved sequence, which raises the question on how the ribosome could effectively scan such short 5′-UTRs for an accurate initiation of translation. In the present study, we expressed capped transcripts of luciferase gene in *Giardia* trophozoites via transfection and observed that when the 5′-UTR of the transcript was lengthened from 9 to 21 nucleotides, there was a corresponding decrease of translation efficiency. Conversely, shortening of the 5′-UTR from nine nucleotides down to a single nucleotide did not result in any reduced translation or leaky scanning. Translation appeared to initiate exclusively from the first initiation codon located downstream from the cap. Experimental evidence indicated also that a stem-loop structure immediately downstream from the initiation codon exerted significant inhibition on translation initiation when the 5′-UTR consisted of less than seven nucleotides. This inhibitory effect was abolished by increasing the distance between the stem-loop and the cap-G structure either upstream or downstream from the start codon, thus suggesting a spatial requirement for effective ribosome recruitment. Overall, our results suggest an absence of ribosome scanning for AUG in initiating translation in *Giardia*. A capped mRNA with a single nucleotide leader is apparently sufficient for recruiting ribosome and initiating translation.

The 5′-untranslated regions (UTRs)[1] of mRNA molecules in eukaryotes have an average length of about 90 nucleotides in mammals (1) and 52 nucleotides in yeast (2). Shortening the 5′-UTR was found to place the first AUG codon in an increasingly unfavorable context for *in vitro* translation initiation, resulting in leaky scanning (3). The yield of protein initiated from the first AUG codon was progressively decreased, whereas that derived from the second AUG downstream from it was correspondingly increased. In a systematic *in vitro* experiment, about half of the ribosomes bypassed the first AUG codon and initiated instead at a downstream AUG site when the first AUG was within 12 nucleotides from the m^7GpppN cap (3). The leakiness was suppressed when the leader sequence was lengthened to 20 nucleotides. Further lengthening of the 5′-UTR beyond 20 nucleotides indicated a linear increase of the efficiency of translation up to 80 nucleotides (4). The scanning leakiness could be also reduced by introducing a modest amount of secondary structure downstream from the first AUG, presumably via slowing down the rate of ribosome scanning (5).

Giardia lamblia is an anaerobic protozoan that infects the small intestines of humans and many other mammals. It is classified as one of the deeply branched and most primitive eukaryotes based on phylogenetic analysis of its small ribosomal RNA and many other proteins (6–9). Within its small 12-Mb genome, most of the genes reported so far do not contain introns (10). Their transcripts have exceedingly short 5′-UTRs ranging mostly from 0 to 14 nucleotides (11) and similarly short 3′-UTRs of 10 to 30 nucleotides (11). With such unusually short 5′-UTRs that do not share a consensus sequence (12), it will be interesting to learn how an optimal accuracy, as well as efficiency of translation, can be maintained and leaky ribosome scanning prevented in *Giardia*. Similarly short 5′-UTRs have also been identified among the mRNAs from other deeply branched anaerobic protozoa such as *Entamoeba histolytica* (13) and *Trichomonas vaginalis* (14). These short 5′-UTRs could reflect the presence of a unique protein synthetic machinery in these closely related primitive eukaryotes.

In a previous study, we indicated that *a priori* decapping of the poly(A)$^+$ RNA from *G. lamblia* was unnecessary for T4 polynucleotide kinase phosphorylation or ligation to oligonucleotides and suggested that much of it could be uncapped at the 5′-ends (12). But a subsequent report on the natural abundance of antisense transcripts in *Giardia* up to 20% of the total poly(A)$^+$ RNA cast some doubt on the real content of mRNA in the sample (15). A recently completed *Giardia* Genome data base showed that many homologues of yeast and human translation initiation factors are present in *Giardia*. Genes encoding homologues of eIF1 (EAA42222), eIF1A (EAA41151), eIF2 (α, EAA42702; β, EAA38204; γ, AAD04236) (16), eIF2B (α, EAA38784; β, EAA38691; δ, EAA40267; ϵ, EAA37510), eIF3 (δ, EAA39062), eIF4A (EAA36655), eIF5A (EAA37465), eIF4E (EAA39303), the cap-binding protein, and eIF5 (EAA37306), the putative GTPase-activating protein for eIF2, were identified, suggesting that capped mRNA may be present and translated in this organism. Upon a closer scrutiny, however, homologues of other initiation factors such as eIF4B, eIF4G, and eIF4H are apparently missing. eIF4B and eIF4H are the RNA-binding proteins thought to promote eIF4A to unwind the RNA secondary structure near the 5′-end of the mRNA presumably required for ribosome scanning (17). eIF4G is a scaffold protein interacting with several other initiation factors (18). It contains a novel RNA binding region of 40 amino acid residues critical

* This work was supported by Grant AI-30475 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 415-476-1321; Fax: 415-476-3382; E-mail: ccwang@cgl.ucsf.edu.

[1] The abbreviations used are: UTR, untranslated region; RT, reverse transcription; DMS, dimethyl sulfate; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide; nt, nucleotide; SL, stem-loop.

This paper is available on line at http://www.jbc.org

This is an Open Access article under the CC BY license.

for ribosome scanning (19). Assuming a cap-mediated translation initiation in *Giardia*, these missing protein subunits may suggest a simplified protein synthetic machinery that still assembles the mRNA, Met-tRNA^{met}, and the 40 S ribosome subunit into an initiation complex at the site of 5′-cap without scanning down the mRNA in a 5′ to 3′ direction searching for the AUG start codon (20). This model could be used to explain the presence of unusually short and heterogeneous 5′-UTRs in *Giardia* mRNAs.

In an effort to verify this model, we first demonstrated a longer half-life and much higher level of translation of the capped than the uncapped mRNA in *Giardia* and proceeded to indicate an apparent absence of ribosome scanning in this organism. The AUG immediately downstream from the cap is recognized as the only initiation codon despite the short distance down to a single nucleotide in between. This simple 5′-cap-G-AUG structure, which provides an optimal translation efficiency in *Giardia*, suggests a protein synthetic machinery that may bridge the gap between prokaryotes and the advanced eukaryotes.

MATERIALS AND METHODS

Plasmid Construction—The α2 tubulin 5′-UTR (21, 22) located between T7 promoter and firefly luciferase gene in the plasmid *pTub* was replaced with the 5′-UTR sequence from the γ giardin gene, AAGAAA, from *Giardia* (23) by site-directed mutagenesis following the instructions of the manufacturer (Stratagene) to construct the plasmid *pGialuc*. Other mutant plasmids were constructed accordingly with the individual mutations verified by subsequent DNA sequencing. The full-length *Renilla* luciferase gene was amplified by PCR and inserted into a pGEM T easy vector (Promega) with a T7 promoter and an α2 tubulin 5′-UTR located upstream. The recombinant plasmid thus constructed, *pRluc*, was used for further alterations of its 5′-UTR by site-directed mutagenesis as described above.

In Vitro RNA Synthesis—The DNA plasmids were linearized at the 3′-end of firefly luciferase gene with a SacII digestion and purified. Using the mMessage mMachine kit (Ambion), capped transcript was synthesized from 0.5 μg of the linearized plasmid DNA with T7 RNA polymerase in the presence of 12 mM m^{7}G(5′)ppp(5′)G and 3 mM GTP. Poly(A) tails were added to the newly synthesized mRNAs following the instructions from a poly(A) tailing kit (Ambion). The product was purified by the MEGAclear kit (Ambion) and examined for integrity by electrophoresis in a 1.0% agarose gel. The concentration of each mRNA sample was estimated by measuring its absorbance at 260 nm in a Beckman DU7 spectrophotometer.

Transfection of Giardia Trophozoites—*In vitro* culture of *Giardia* strain WB trophozoites was maintained as described previously (24). Serial passages of the *in vitro* culture were performed at an inoculation ratio of 1:13 once every 3 days into fresh culture medium to maintain a continuous logarithmic cell growth. Transfection of the trophozoites with RNA was performed by electroporation as described previously (25, 26) with minor modifications. Logarithmically growing cells were pelleted at 2500 rpm for 10 min and washed twice with the Ca^{2+}-, Mg^{2+}-free phosphate-buffered saline. Cell pellets (5×10^5 cells) were suspended in 400 μl of cytomix buffer, mixed with 20 μg of the *in vitro* transcript and 100 μg of yeast tRNA (Invitrogen) and placed in a 0.2-μm cuvette for 10 min on ice. Electroporation was performed with a Bio-Rad GenePulser set at 0.45-kV voltage and 500-microfared capacitance. Following the electroporation, the cells were transferred into 13 ml of fresh culture medium and incubated at 37 °C for 4 h. Assay of the luciferase activity in the lysate of transfectants were performed by the instructions of the manufacturer (Promega). The luciferase activity was measured with a luminometer (Monolight 2010; Analytical Luminescence Laboratory, San Diego, CA) and registered in relative light units. Protein concentration was determined by the standard Bradford method in a Beckman DU7 spectrophotometer. The specific luciferase activity expressed in each line of *Giardia* transfectants was determined from at least three independent electroporation experiments with calculated standard deviations.

Semi-quantitative Reverse Transcription (RT)-PCR—The total RNA from RNA-transfected *Giardia* trophozoites was extracted, and the concentration of each RNA sample was measured spectrophotometrically. An equal quantity of RNA was included as template in each of the semi-quantitative one-step RT-PCR (Invitrogen) using the firefly luciferase gene primers 5′-ATGGAAGACGCCAAAAACATAAAG-3′ and 5′-TTACACGGCGATCTTTCCGCCC-3′ to amplify the full-length transcript of luciferase gene (GenBank™ accession number U03687). The RT-PCR conditions were set at one cycle of reverse transcription at 50 °C for 30 min and 20 cycles of amplification at 94 °C for 15 s, 50 °C for 30 s, and 72 °C for 1 min. The same RT-PCR on the NADH oxidase gene transcript in the same transfected *Giardia* cells (GenBank™ accession number AF454832) was included as an internal quantitative control using the pair of primers 5′-ATGTCAAACTTACTCTTCAAG-3′ and 5′-TCAGTCCTTCTTGTTTATCGC-3′ for the full-length gene product.

Probing the RNA Secondary Structure by Chemical Modification or Enzymatic Digestion and Followed By Primer Extension—The capped *in vitro* transcript of SacII-restricted GGAUG-Fluc construct was prepared using a T7 RNA polymerase kit (MegaScript; Ambion). The RNA was heated at 70 °C for 15 min and cooled down slowly to room temperature in an ambient water bath for 30 min in a 200-μl non-denaturing probing buffer (80 mM HEPES-KOH, pH 7.8, 10 mM $MgCl_2$, 50 mM KCl). It was quantified spectrophotometrically, and its integrity was verified by electrophoresis on a 1% agarose-formaldehyde gel. Approximately 5 μg of the *in vitro* transcript was used in each 300-μl reaction mixture containing either 0.5% dimethyl sulfate (DMS) or 21 mg/ml 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide (CMCT) (25, 35). The reactions were carried out at 37 °C for 0, 5, 10 and 20 min and terminated by adding 75 μl of the stop buffer (1 M Tris acetate, pH 7.5, 1 M β-mercaptoethanol, 1.5 M sodium acetate, 0.1 mM EDTA). The chemically modified RNA was precipitated by 3 volumes of 100% ethanol.

Enzymatic probing of the RNA structure was carried out essentially as described previously (36). The capped *in vitro* transcript of GGAUG-Fluc (2 μg), together with 4 μg of yeast tRNA, was denatured in the 1× RNA structure buffer (Ambion) at 65 °C for 15 min and then cooled slowly to ambient temperature within 30 min. The renatured RNA sample was digested with RNase V1, which recognizes only double-stranded RNA as substrate (Ambion), at 10-fold serially diluted enzyme concentrations at room temperature for 20 min. The reaction was stopped by adding phenol-chloroform to extract the remaining RNA molecules, which were subsequently pelleted by ethanol in the presence of 0.3 M NaOAc and dissolved in diethyl pyrocarbonate-treated water.

The sites of chemical modification or RNase V1 digestion in the RNA molecule were mapped by primer extension using the reverse transcriptase Superscript III (Invitrogen) with radiolabeled oligonucleotides targeted to the region 20 nt downstream from the 3′-end of the postulated stem-loop in the transcript (see Fig. 4*A*). Briefly, the ^{32}P-5′-end-labeled primers were annealed to ~1 μg of the chemically modified or RNase V1-digested RNA by incubating the mixture at 75 °C for 15 min followed by another 10 min on ice. Primer extension was carried out at 50 °C for 1 h as specified by the manufacturer. Reaction products were analyzed in an 8% denaturing acrylamide gel along with the sequencing ladders prepared by the fmol cycle-sequencing system (Promega). 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. The position of each radiolabeled band in the denaturing gel of primer extension products from RNase V1-digested RNA represented a nucleotide that is cleaved at its 5′-end by the enzyme and had thus the same mobility as that in the corresponding DNA sequencing gel.

RESULTS

The Stability and Translation of Capped Luciferase mRNA in Giardia Trophozoites—*In vitro* transcription of the linearized *pGialuc* followed by polyadenylation of the *in vitro* transcript produced the firefly luciferase mRNA with m^{7}G-(5′)ppp(5′)GGAAGAAACC in the 5′-UTR. This *in vitro* transcript was introduced into *Giardia* trophozoites via electroporation. Semi-quantitative RT-PCR on the total RNA extracted from the trophozoites 0.25, 1, 2, 4, 8, and 20 h after transfection indicated that the capped luciferase transcript was present in *Giardia* at a significantly high level 15 min after electroporation, which was slightly decreased after 1 h and appreciably reduced after 2 h but remained still detectable even after 4 h (Fig. 1*A*). Results from assaying the luciferase activity in the lysate of the transfectants indicated an appreciable and relatively constant level of activity expressed from 1 to 4 h following the transfection (Fig. 1*B*). In a parallel experiment, intro-

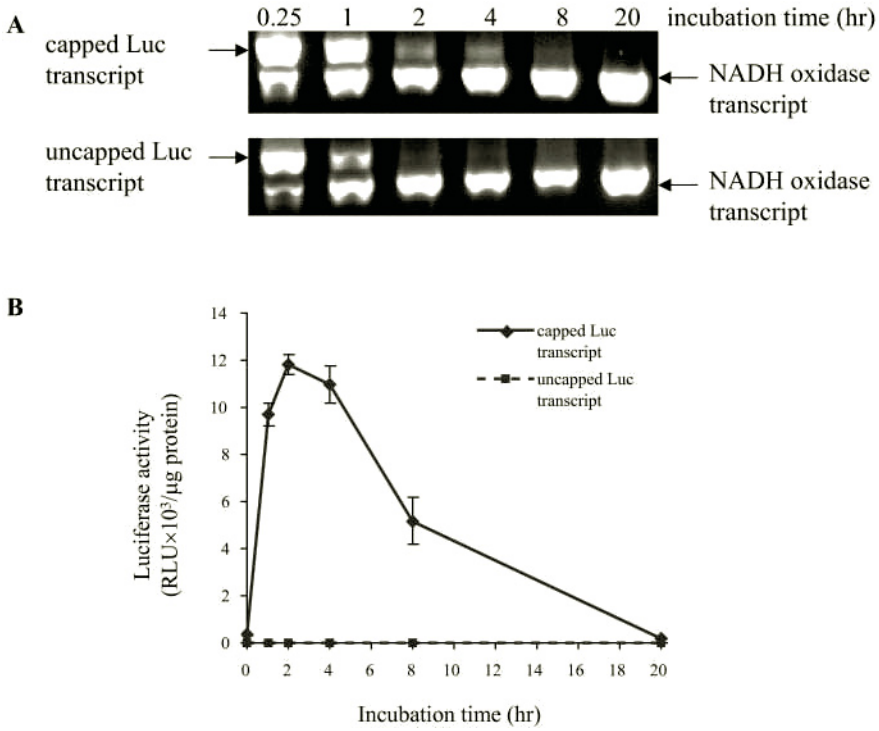

FIG. 1. **Time course analysis of firefly luciferase mRNA and luciferase activity expressed in the transfected *Giardia* trophozoites.** *A*, RNA samples extracted from the transfected *Giardia* trophozoites harvested at different time points after electroporation were used as templates in semi-quantitative RT-PCR for quantitating luciferase mRNA. Quantitation of NADH oxidase mRNA in the same time samples of *Giardia* trophozoite RNA extracts by the same RT-PCR procedure was included as an internal quantitative control. *B*, firefly luciferase activities (*Luc*) were determined from the same time samples of transfected *Giardia* trophozoites in triplicate experiments. For preparation of the capped and uncapped luciferase mRNA used in the electroporation experiments, see "Materials and Methods."

duction of an uncapped, otherwise identical, luciferase mRNA with 5′-GGAAGAAACC in the 5′-UTR into *Giardia* trophozoites resulted in a shorter intracellular half-life of the mRNA. It was rapidly reduced between 15 min and 1 h after transfection and became virtually undetectable by RT-PCR after 2 h (Fig. 1*A*). No luciferase activity could be detected in the cell lysate from 15 min to 1 h after transfection, while the uncapped transcript was still detectable by RT-PCR, suggesting that the uncapped mRNA was not or only poorly translated in contrast to the capped mRNA. Thus, the cap-mediated translation initiation is most likely functioning in *Giardia*. All the subsequent investigations on translation initiation inside *Giardia* trophozoites were thus conducted with capped mRNAs with the luciferase expression routinely monitored 4 h after transfection.

An Increasing Length of 5′-UTR Causes a Corresponding Decrease in Translation Efficiency in Giardia—The 5′-UTR of capped luciferase mRNA studied in Fig. 1 consists of an $m^7G(5')ppp(5')G$ structure as the cap separated from the initiation codon by a stretch of GAAGAAACC, which is the 5′-UTR of γ-giardin mRNA AAGAAA from *Giardia* flanked by a 5′-G and a 3′-CC (23). This capped 9-nt UTR enabled translation of its downstream open reading frame to an appreciable level of luciferase activity in *Giardia* (Fig. 1*B*). This 5′-UTR was extended to 15, 21, and 27 nucleotides by duplicating, triplicating, and quadrupling the γ-giardin 5′-UTR sequence AAGAAA in *pGialuc* through site-directed mutagenesis. *In vitro* transcripts of these three mutants were capped and polyadenylated and individually introduced into *Giardia* trophozoites. Levels of the transcripts inside *Giardia* trophozoites 4 h after transfection were monitored by semi-quantitative RT-PCR and shown to be essentially identical to one another (Fig. 2*A*), indicating a similar *in vivo* stability. The results on their translation efficiencies, presented in Fig. 2*B*, indicate that, contrary to what was observed in the *in vitro* mammalian system (3), an increase in the length of 5′-UTR leads to a decrease in translation. A 15-nucleotide 5′-UTR led to 58.7 ± 4.5% (Fig. 2*B*, *2*) of the expression level of that from a 9-nucleotide 5′-UTR (Fig. 2*B*, *1*), whereas 21- and 27-nucleotide 5′-UTRs reduced the levels of expression to 15.8 ± 3.5% (Fig. 2*B*, *3*) and 23.2 ± 7.4% (Fig. 2B, *4*), respectively. This is the first instance, as far as we are aware, that an increase of 5′-UTR from 9 to 27 nucleotides exerts an inhibitory effect, instead of an enhancing effect, on translation. The observation, however, goes along well with the fact that most 5′-UTRs in the mRNAs of *Giardia* are between 0 to 14 nucleotides in length (11).

Time courses of the gradual disappearance of these transcripts in *Giardia* within the ensuing 4 h post-transfection were also monitored by the same procedure described in Fig. 1*A*. The results presented in Fig. 2*C* indicate that the rate of disappearance of the transcript with the highest level of expression (Fig. 2*C*, *1*) is essentially identical to that with the lowest level of expression (Fig. 2*C*, *3*). They thus rule out the possibility that the different levels of expression could be because of different time courses of degradation among the transcripts.

The Initiation Site Can Be a Single Nucleotide Downstream from the Cap without Causing Reduced Translation or Leaky Ribosome Scanning—The confined length of 5′-UTR for an optimal translation initiation in *Giardia* raised the question on how much it can be further shortened without compromising the translation efficiency. The original AUG in luciferase mRNA was converted to UUG whereas a new AUG was placed 6 nucleotides upstream from it and separated from the cap by only a single G residue (Fig. 3*A*, *3*). This chimeric mRNA was introduced into *Giardia* via transfection and reached 87.7 ± 9.7% of the translation level of that with the original 9-nucleotide 5′-UTR (Fig. 3*A*, *2*). There is thus little space required between the cap and AUG for an optimal translation initiation in *Giardia*.

When the downstream UUG in transcript 3 was converted back to AUG in transcript 4, *i.e.* when a second AUG was introduced 6 nucleotides downstream from the first one that was only one nucleotide away from the cap, the chimeric mRNA thus constructed was expressed at 94.5 ± 22% (Fig. 3*A*, *4*) of the control level in *Giardia*. However, when the distance between the two AUGs was shortened to 5 nucleotides, the level of expression was reduced to 1.6 ± 0.5% (Fig. 3*A*, *5*) of the control, suggesting that most, if not all, of the translation was

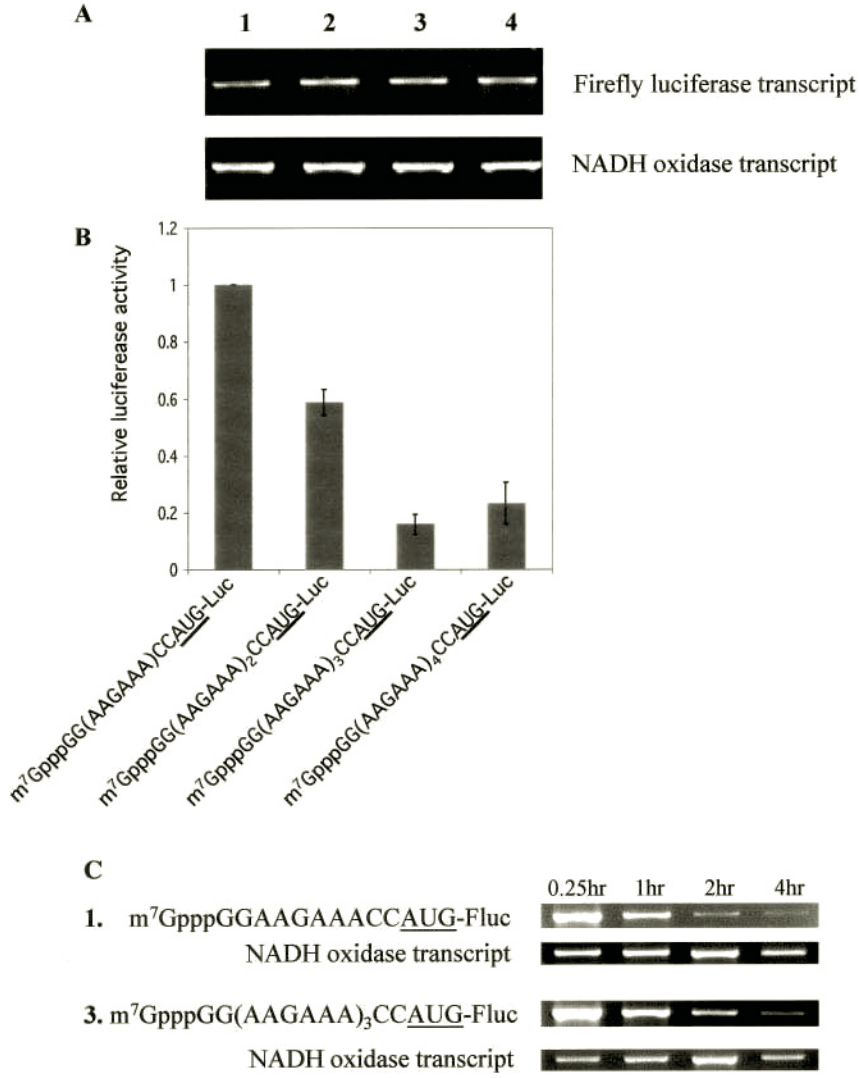

FIG. 2. **Effects from extending the length of 5′-UTR in firefly luciferase mRNA on its expression in transfected *Giardia*.** *A*, RNA samples extracted from the transfected *Giardia* trophozoites 4 h after electroporation were used as templates in semi-quantitative RT-PCR for monitoring levels of the luciferase mRNA. Quantitation of NADH oxidase mRNA was included as a sampling control. *B*, capped *in vitro* luciferase transcripts *1*, *2*, *3*, and *4* with 5′-UTRs of 9, 15, 21, and 27 nts, respectively, were synthesized and used to transfect *Giardia* trophozoites. The transfected cells were harvested 4 h after electroporation and assayed for firefly luciferase activity. Each data value was derived from three independent electroporation experiments. *Luc* designates the open reading frame of firefly luciferase mRNA without its initiation codon. *C*, time samples of RNA extracted from transcripts *1* and *3* transfected *Giardia* cells (0.25, 1, 2, and 4 h after transfection) were used as the templates in semi-quantitative RT-PCR for monitoring the changed levels of luciferase mRNA with time. Results from RT-PCR on the NADH oxidase transcript among the same time samples were included as the sampling control.

initiated from the first AUG, which was now out of phase from the downstream luciferase open reading frame. This aborted translation was not rescued by doubling or tripling the second initiation codon 5 nucleotides downstream from the first AUG (Fig. 3*A*, *6* and *7*), thus indicating a lack of or very little ribosome scanning leakiness that could initiate translation from the downstream AUGs. The absence of such leakiness and the short spacing of only a single nucleotide separating the cap and AUG for optimal translation suggest absence of ribosome scanning during translation initiation in *Giardia*.

The time courses of degradation and eventual levels of individual transcripts in *Giardia* 4 h after the transfection in the above experiments were also examined with semi-quantitative RT-PCR as described in Figs. 1 and 2. Results in Fig. 3*B* indicate that transcripts with varying levels of expression (*2*, *4*, and *5*) have essentially the same diminishing time course and the same level after 4 h. The distinct levels of expression of these transcripts are thus most likely attributed to the different translation efficiencies.

A Postulated Secondary Structure Immediately Downstream from and Including the Start Codon Is Inhibitory to Translation Initiation on a 5′-Capped mRNA with a Single Nucleotide Leader—The absence of ribosome scanning suggests that 5′-cap-G-AUG alone may constitute an adequate basic structure in the mRNA for both 40 S ribosome subunit recruitment and translation initiation in *Giardia*. Accomplishment of such a task should, however, still require some downstream space from the cap for binding of the various initiation factors and recruitment of the ribosome. Thus, instead of preventing scanning leakiness in mammals (5), a secondary structure immediately downstream from the initiation codon is expected to have an inhibitory effect on initiating translation of a mRNA with a very short 5′-UTR in *Giardia*. The native firefly luciferase mRNA has an MFOLD predicted stem-loop structure (SL) immediately downstream from and including the start codon (Fig. 4*A*). When the distance between the 5′-cap and the start codon of luciferase mRNA is shortened from 9 to 7 nucleotides, a 98.8% expression (Fig. 4*A*, *2*) of the mRNA is still achieved, which is, however, reduced to 46.6 ± 2.9% (Fig. 4*A*, *3*) with 5 nucleotides and 22.7 ± 1.8% (Fig. 4*A*, *4*) and 28.1 ± 3.6% (Fig. 4*A*, *6*) with 3 nucleotides and 1 nucleotide in the 5′-UTR, respectively. To contend with the possibility that the reduced expression of transcript 4 in Fig. 4*A* could be attributed to the enriched GC content in the shortened sequence, its CC residues were replaced with AA in transcript 5. Expression of the latter in transfected *Giardia* was 45.4 ± 9.0% of the control (Fig. 4*A*, *5*), which was higher than that of transcript 4, suggesting that the base composition may have some effect on translation initiation. An enriched GC content in the sequence (transcript 4) has an equivalent inhibitory effect as if the stem-loop is pulled 2 nucleotides closer to the cap (transcript 6). The shortened transcripts have, however, consistently lower levels of expression than that of the longer transcripts 1 and 2.

To verify whether the inclusion of initiation codon in the

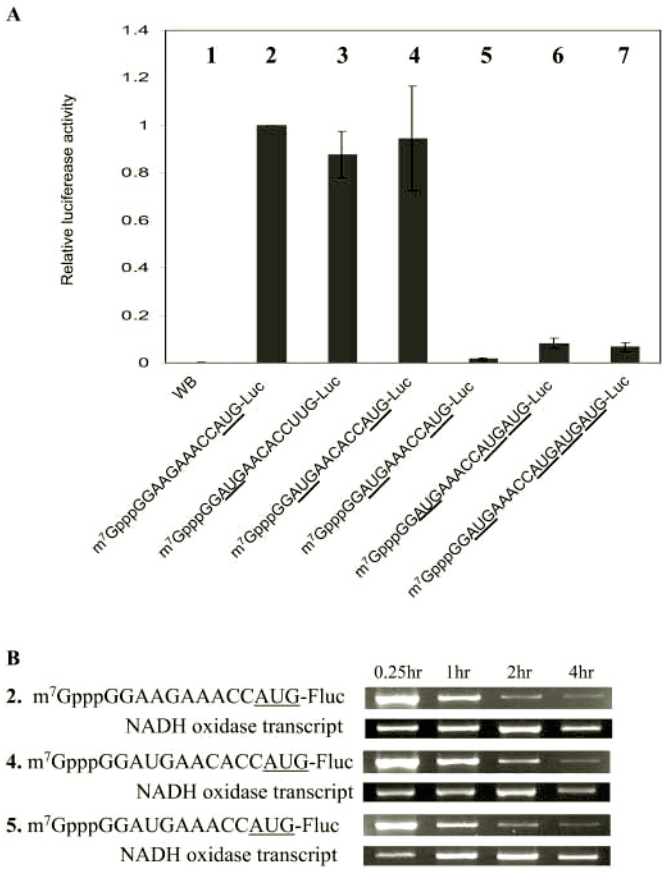

FIG. 3. **Effects from shortening the length of 5′-UTR and introducing additional downstream initiation codons in firefly luciferase mRNA on its expression in transfected *Giardia*.** *A*, various capped *in vitro* firefly luciferase transcripts with initiation codons located at different positions were prepared and used individually to transfect *Giardia* trophozoites. The firefly luciferase activity from each transfectant was assayed 4 h after electroporation. Each data value was derived from at least three independent electroporation experiments. *Luc* designates the open reading frame of firefly luciferase mRNA without its initiation codon. *WB*, *Giardia* WB strain. *B*, semi-quantitative RT-PCR analysis of the time samples of RNA extracted from transcripts *2*, *4*, and *5* transfected *Giardia* trophozoites at 0.25, 1, 2, and to 4 h after electroporation. RT-PCR quantitation of NADH oxidase mRNA was included as a sampling control at each time point.

postulated stem structure (see Fig. 4*A*) may have contributed to the reduced translation, an additional AUG initiation codon was placed immediately upstream from the original firefly luciferase AUG (transcript 7) with the anticipation that the newly added AUG, which is apparently not included in the stem structure, will be the real initiation codon (see Fig. 3*A*). Transcript 7 was expressed at 68.6 ± 5.8% of the control level (Fig. 4*A*), which is 2.4 times that of the transcript without the extra AUG (Fig. 4*A*, *6*), thus indicating that inclusion of AUG in a secondary structure does reduce translation initiation. This conclusion is further supported by the 66.7 ± 18.9% expression level of transcript 8 (Fig. 4*A*), in which an AAGAAA sequence was inserted immediately downstream from the original luciferase AUG in transcript 6 to free the initiation codon from the postulated stem structure. Translation initiation was not, however, further enhanced when an additional AAGAAA was inserted immediately upstream from AUG in transcript 8 (Fig. 4*A*, *9*), which resulted in only 51.0 ± 15.2% of the control level, confirming the previous observation that lengthening the distance between the cap and initiation codon does not enhance translation initiation but reduces it instead.

The N-terminal portion of firefly luciferase unfortunately plays an essential role in its catalytic function (27). It cannot be altered extensively enough to destroy the predicted stem-loop structure in the encoding mRNA without eliminating the enzymatic activity as well. The *Renilla* luciferase mRNA (*Rluc*), which does not have a predicted secondary structure within 10 nts downstream from the start codon, was used to replace the firefly luciferase mRNA in the study (Fig. 4*B*). The result indicates that when there is only 5′-cap-G placed in front of the initiation codon, the *Renilla* luciferase activity was expressed at 97.5 ± 21% of the 9-nucleotide 5′-UTR control (Fig. 4*B*, compare *2* and *1*). Furthermore, when the predicted stem-loop secondary structure in the firefly luciferase mRNA (see Fig. 4*A*, *SL*) was introduced between the 5′-UTR and AUG in transcripts *1* and *2* in Fig. 4*B* (transcripts *3* and *4*), the former retained the same level of expression as those of transcripts *1* and *2* whereas expression of transcript *4* was reduced to 44.9 ± 9.8% of the control (Fig. 4*B*, *3* and *4*).

The time courses of degradation and final levels of the transcripts presented in Fig. 4 were also examined by semi-quantitative RT-PCR. The results (not shown) indicated no detectable difference among any of them and thus confirm that the variations in luciferase expression observed in Fig. 4 should represent the changing efficiencies in translation initiation.

Probing the Secondary Structure at the 5′-End of Firefly Luciferase mRNA by Chemical Modification and Enzymatic Digestion Followed By Primer Extension—Because there is always the uncertainty whether an MFOLD-predicted RNA secondary structure really exists in nature, we determined the actual secondary structure at the 5′-end of firefly luciferase mRNA under native conditions using a combination of chemical modification and RNase V1 enzymatic digestion of the RNA molecule followed by a primer extension reaction. The capped GGAUG-Fluc *in vitro* transcript was treated with DMS or CMCT under the native condition. DMS methylates unpaired A residues at N1 and acts weakly on unpaired C residues at the N3 positions, whereas CMCT modifies the N1 group of unpaired G residue and the N3 group of unpaired U residues (25, 35). The bases thus modified can no longer be recognized by the reverse transcriptase in a subsequent primer extension reaction, and each modified base can be identified as a reverse transcription stop in subsequent gel electrophoresis. The oligonucleotide terminating at the modified base moves one nucleotide ahead of the corresponding DNA ladder, because primer extension stops immediately before the modified base. The results from a chemical modification and primer extension experiment on the capped GGAUG-Fluc transcript are shown in Fig. 5*A*. When the region involving the predicted loop structure was examined by primer extension after the chemical modification (Fig. 5*A*), essentially all the bases in it were shown as the transcriptional stops except for U21 and C32, suggesting that the region is largely in an unpaired single-stranded form. The two regions postulated to form the stem with a bulge in the middle were found free of chemical modification except for C13 that is at the end of the stem structure, which could be attributed to a partial denaturation of the stem structure at the border region with the loop. The AUG initiation codon at the top of the gel in Fig. 5*A* was free of chemical modification and is thus most likely involved in the stem structure. The bases in the postulated bulge region demonstrated little accessibility by the chemicals and are likely involved in as yet unidentified Watson-Crick base pairings. Taken together, the results, presented in Fig. 5*A* and illustrated in Fig. 5*B*, lend a strong experimental support to the presence of the predicted stem-loop structure at the 5′-end of firefly luciferase mRNA with the AUG located at the base of the stem structure.

To further verify the conclusion from chemical modification experiments, the GGAUG-Fluc transcript structure was also

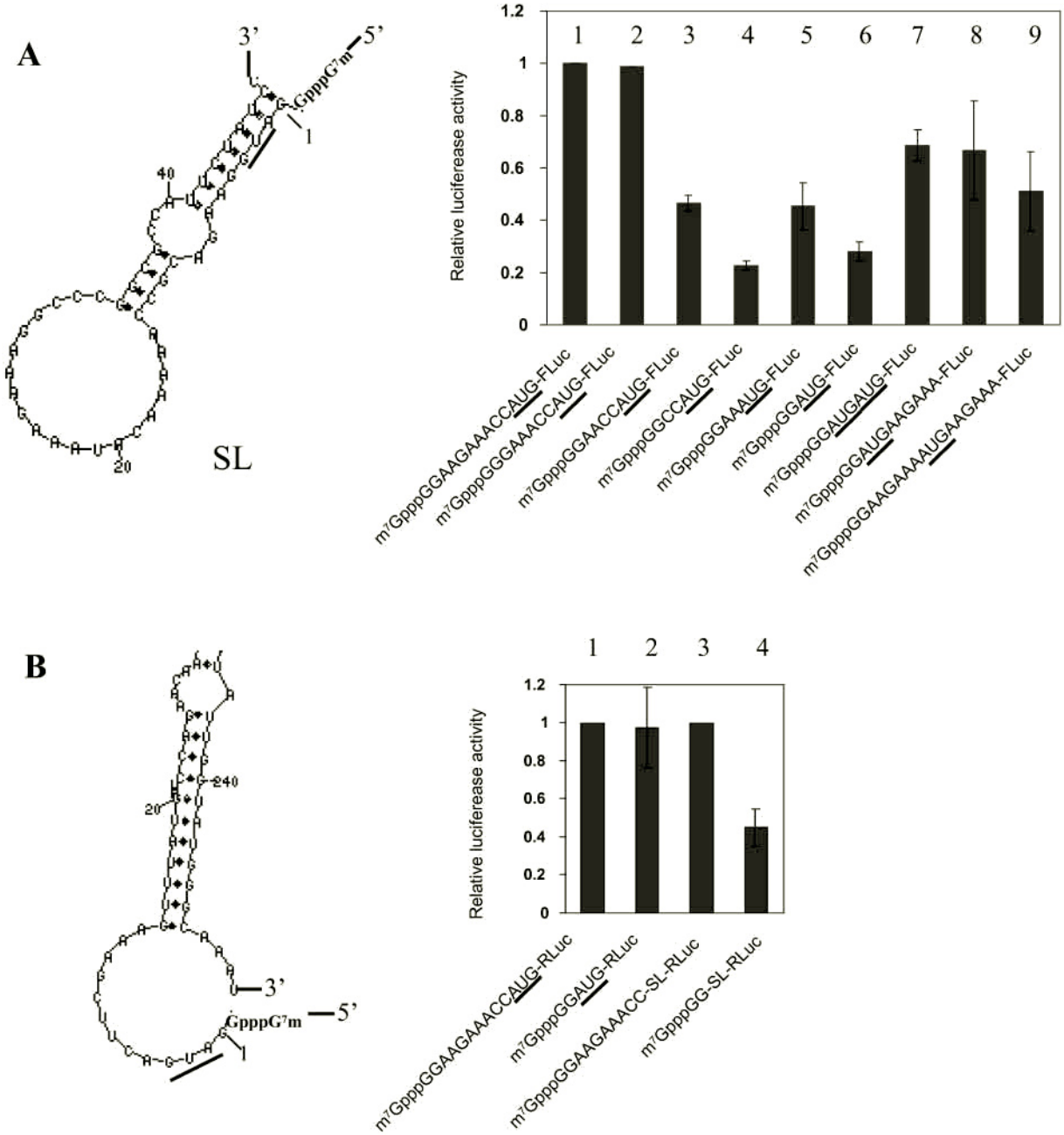

FIG. 4. **The effect of a secondary structure immediately downstream from and involving the start codon of the transcript on the translation initiation in *Giardia*.** The SL structure downstream from and involving the initiation codon of firefly luciferase mRNA (*A*) and the secondary structure downstream from the AUG in *Renilla* luciferase mRNA (*B*) were predicted by MFOLD software (39). *A*, the 5′-UTR of capped firefly luciferase mRNA is shortened from 9, 7, 5, and 3 nt to 1 nt (transcripts *1*, *2*, *3*, *4*, and *6*). Transcript *5* has the CC residues in transcript *4* changed to AA. One additional AUG codon was inserted immediately upstream from the original luciferase AUG in transcript *6* to construct transcript *7*, whereas AAGAAA was inserted immediately downstream from the start codon in transcript *6* to make transcript *8*. Transcript *9* was derived from transcript *8* by introducing another AAGAAA between the cap and AUG. Each transcript was introduced into *Giardia* trophozoites by electroporation, and the luciferase activity thus expressed was assayed after a 4-h incubation. *B*, the 5′-UTR in capped *Renilla* luciferase mRNA was shortened from 9 nts (transcript *1*) to 1 nt (transcript *2*). The SL structure from the 5′-end of firefly luciferase mRNA (*A*) involving the AUG codon was used to replace that in transcripts *1* and *2* to produce transcripts *3* and *4*. Expressions of these transcripts in transfected *Giardia* were compared 4 h after transfection. The data points were each derived from at least three independent transfection experiments. *Fluc* and *Rluc* designate the open reading frames of firefly luciferase and *Renilla* luciferase mRNAs without initiation codon, respectively.

probed with RNase V1 (Fig. 5*C*), known to hydrolyze only RNA structures in stacked helical conformations (36). The results indicate that the enzyme-digested RNA molecule provided reverse transcriptase stops at positions G1, A2, G4, U42, C43, U44, U46, and C47 in the postulated stem 1 structure and at positions C10, G11, C12, C13, G34, and G37 in the postulated stem II (Fig. 5*D*). The few nucleotides in the predicted stems that were found resistant to RNase V1 digestion (Fig. 5*D*) were, however, also found resistant to chemical modification (Fig. 5*B*), suggesting that they could be shielded from the enzyme by certain steric hindrance. The nucleotides in the postulated loop and bulge structures, with but three exceptions, all turned out to resist RNase V1 digestion (Fig. 5*D*), thus suggesting their single-stranded nature. For the RNase V1 digestible C32 in the loop and C38 and C39 in the bulge, they also turned out resistant to chemical modification (Fig. 5*B*). These three bases could be thus involved in a stacked conformation generated between stems I and II. The good agreements between data

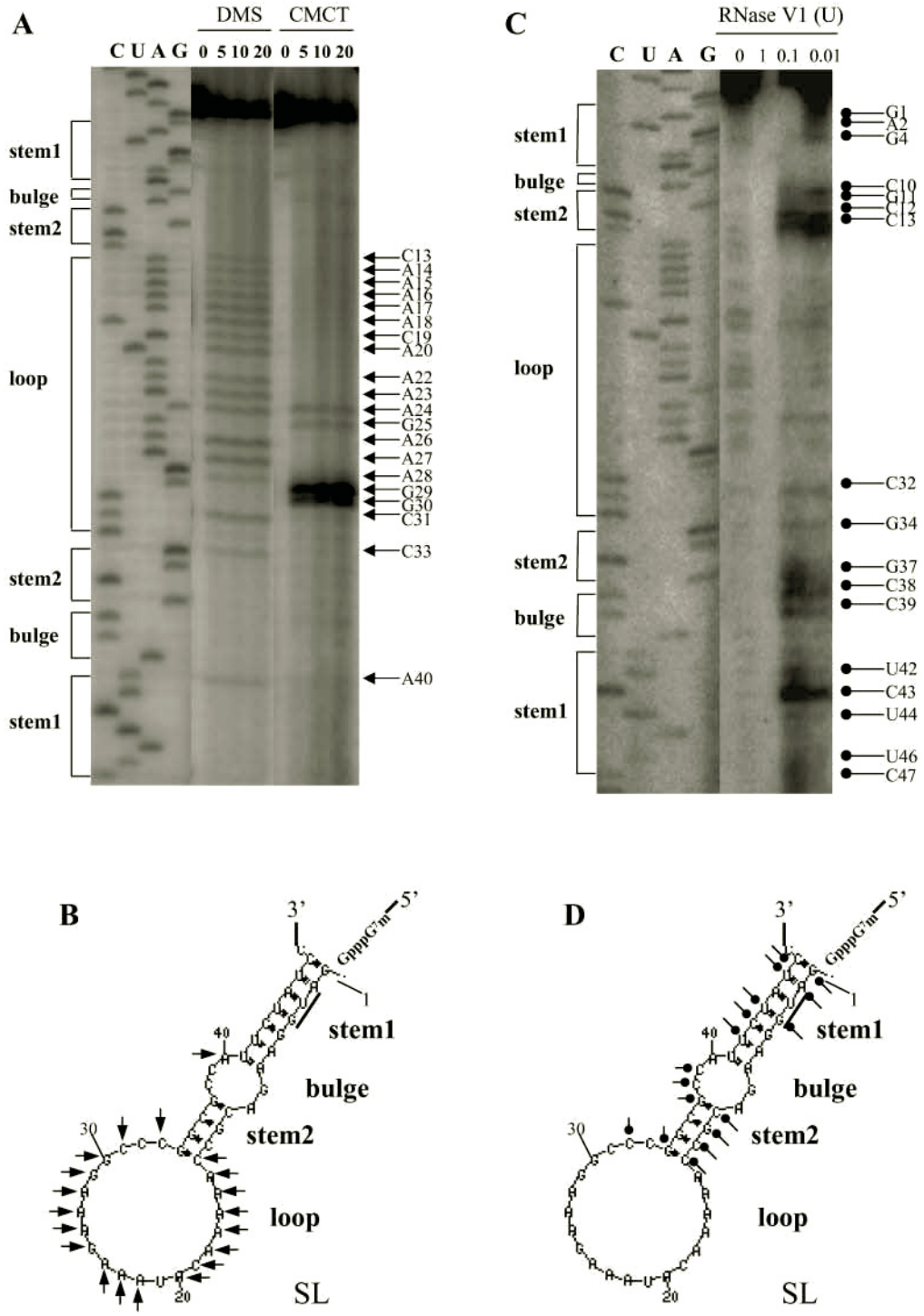

FIG. 5. **Chemical and enzymatic probing of the m^{7}GpppGG-AUG-Fluc transcript and structural analysis of the 5′-end region of firefly luciferase mRNA by primer extension.** *A*, the chemically modified RNA molecule (*A* and *C* by DMS, *U* and *G* by CMCT) was probed by reverse transcription using an end-labeled primer complementary to the sequence 20 nucleotides downstream from the last nucleotide in the postulated stem-loop structure presented in Fig. 4*A*. Durations of incubation in each chemical modification reaction are indicated in minutes *above* each *lane* of gel electrophoresis. The *left side* of the gel labeled *CUAG* at the *top* represents the corresponding sequencing ladder of the DNA plasmid. *Arrows* on the *right* indicate transcriptional stops identified from the primer extension representing the chemically modified bases in the treated RNA molecule, which migrate at a distance 1 nucleotide short of that in the corresponding DNA ladder. *B*, summary of the results from chemical modification confirming the postulated secondary structure in the 5′-end of m^{7}GpppGG-AUG-Fluc transcript. The chemically modified bases are indicated by *arrows*. *C*, structural probing of secondary structures in the 5′-end of the m^{7}GpppGG-AUG-Fluc transcript by RNase V1 digestion. Products from RNase V1 digestion at different concentrations of the enzyme were analyzed by primer extension using a radiolabeled primer hybridizing to nt 67–89 in the RNA template. Labels on the *right* indicate transcriptional stops generated by the enzymatic digestion of the RNA template and identified with the DNA sequencing ladder on the *left*. The positions of the predicted stem and loop regions are in *brackets* on the *left side* of the ladder. *D*, summary of the results from RNase V1 digestion confirming the postulated secondary structure of the 5′-end of m^{7}GpppGG-AUG-Fluc transcript. Points of enzymatic hydrolysis are demonstrated by *solid circles* attached to *bars*. *Numbers* in *B* and *D* designate the nucleotide positions relative to the 5′-end of the open reading frame in firefly luciferase mRNA.

from chemical modification and enzyme digestion indicate the actual presence of the predicted stem-loop structure at the 5′-end of firefly luciferase mRNA including the AUG start codon at the base of the stem structure.

DISCUSSION

In the current investigation, we employed the experimental technique of introducing mRNAs into *Giardia* trophozoites via electroporation and monitoring its *in vivo* expression. The results thus obtained on translation initiation differ significantly from those observed in the *in vitro* systems derived from yeast and mammals (28, 29). The mRNA translation efficiency in *Giardia* decreases with increasing lengths of 5′-UTR beyond nine nucleotides, whereas the mammalian *in vitro* system demonstrated reduced translation efficiency accompanied with increasing leakiness of ribosome scanning with a 5′-UTR shorter than 20 nts (3). The 5′-UTR in *Giardia* mRNA can be reduced to a single nucleotide without affecting its translation efficiency or causing any leakiness, suggesting that ribosome scanning involved in translation initiation among the more advanced eukaryotes may not exist in *Giardia*. The absence from *Giardia* of eIF4B, eIF4H, and eIF4G homologues that are involved in ribosome scanning (19, 30) further supports this conclusion. The minimal inclusion of a single G residue in the 5′-UTR in our current study was attributed to a technical necessity in the T7 RNA polymerase-catalyzed *in vitro* transcription, because a change of the first two G residues downstream from the T7 transcription start site would significantly reduce the yield of the *in vitro* transcript (31). Because many *Giardia* mRNA species are without any 5′-UTR at all (11), it is not unlikely that a simple conjugation between the 5′-cap and the initiation codon may constitute the essential and sufficient structural basis in the mRNA to initiate translation in *Giardia*. Presumably, following the binding of eIF4E to the cap, the 40 S ribosome subunit, associated with various other initiation factors and Met-tRNA^{met}, is recruited to the 5′-cap (30). The AUG codon immediately downstream from the 5′-cap will be recognized by the bound complex without going through a ribosome scanning from the 5′- to the 3′-direction along the mRNA. Our current data, indicating a certain spatial requirement for optimal translation initiation when a secondary structure is located immediately downstream from the initiation codon (Fig. 4), support this conclusion.

These events illustrated above distinguish the protein synthesis in *Giardia* from those in the other eukaryotes and prokaryotes. In the prokaryotic kingdom, selection of the correct start site is often guided by base pairing of the 5′ Shine-Dalgarno sequence in the mRNA with the 3′-terminal tail of 16 S ribosomal RNA in the 30 S subunit (32). Leaderless mRNAs have been, however, also identified in bacteria. They are recognized and translated by a ribosome-IF2-fMet-tRNA complex similar to, but distinctive from, that identified in typical eukaryotes (33). The *Archaea*, which are the closest relatives to the nuclear components of eukaryotes, also have very short or no 5′-UTR in their uncapped mRNAs (34). Though they frequently utilize Shine-Dalgarno sequences to identify translation initiation codons, they also possess a regiment of initiation factors that are clearly homologous to those of the eukaryotes that participate in identifying the initiation codon. Remarkably, the translation initiation factors identified in *Methanobacterium jannaschii* have been eIF1, eIF1A, eIF2, and eIF4A without eIF2B, eIF4B, eIF4H, eIF4G, or eIF5 (34), which are a somewhat simplified replica of the initiation factor profile identified in *Giardia* (see the Introduction). These initiation factors may provide an adequate complex for initiating translation of the uncapped mRNAs in *M. jannashii* apparently without either a 5′-UTR or ribosome scanning.

Giardia mRNAs do not have a Shine-Dalgarno-like sequence or any other consensus sequence at the 5′-ends (11). Differing from the *Archaea*, a cap structure at the 5′-end of mRNA is probably involved in recruiting the small ribosome subunit and subsequent recognition of the downstream initiation codon. It is not known whether an additional RNA base pairing between the mRNA and the 3′-end of the 16 S-like ribosomal RNA in *Giardia* is involved in guiding the identification of AUG. In a previous study (12), we identified in a giardiavirus a 13-nucleotide downstream box in the coding region of uncapped transcript that can complement a 15-nucleotide sequence at the 3′-end of the 16 S-like ribosomal RNA in *Giardia*. This downstream box is not involved in any secondary structure of the viral transcript (35) but is essential for initiating translation of the viral transcript (12), suggesting that a Shine-Dalgarno type interaction could be involved in translation initiation, albeit by a sequence downstream from the initiation codon. Further analysis indicated that translation of the uncapped giardiaviral transcript in *Giardia* is initiated at an internal ribosomal entry site in the mRNA covering parts of both the 5′-UTR and the downstream region including the downstream box (36). Within the complex secondary structures of this internal ribosomal entry site, there is a stem-loop structure located 21 nucleotides upstream from the AUG and another stem-loop 7 nucleotides downstream from it, both playing essential roles in translation initiation. The very location of the AUG between the two stem-loops cannot be altered; a mere shift upstream or downstream by three nucleotides completely abrogated translation initiation (36). Thus, the 31-nucleotide stretch between the two stem-loops, which could accommodate the size of a bound small ribosome subunit well (37, 38), may be the very site for recruited ribosome small subunit to bind where the initiation codon in the mRNA may fit into the precise site of the bound ribosome for initiating the translation. This observation from the internal ribosomal entry site in a viral transcript in *Giardia* has rendered further support to our tentative conclusion that translation initiation from cellular mRNAs may occur efficiently without ribosomal scanning in *Giardia*. A recent survey of about 20 randomly chosen *Giardia* cDNAs indicated that there is no second in-frame AUG within 100 nucleotides downstream from the first initiation codon.[2] It could be another indirect indication of the absence of ribosome scanning in the protein synthesis in *Giardia*.

In summary, we have identified in one of the most primitive eukaryotes *Giardia* a unique mechanism of translation initiation that probably falls somewhere between that of a prokaryote and a eukaryote with a potentially close link with that in the *Archaea*. The ease in conducting these experiments *in vivo* and the likelihood in turning them into a prototype of the protein synthetic machinery representing a transition from the prokaryote into the eukaryote would carry a great potential for more interesting findings in future studies.

Acknowledgments—We thank Dr. Srinivas Garlapati of UCSF for the plasmid *pTub* and constructive suggestions for this manuscript.

REFERENCES

1. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148
2. Cigan, A. M., and Donahue, T. F. (1987) *Gene* **59,** 1–18
3. Kozak, M. (1991b) *Gene Expr.* **1,** 111–115
4. Kozak, M. (1991a) *Gene Expr.* **1,** 117–125
5. Kozak, M. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8301–8305
6. Adam, R. D. (2001) *Clin. Microbiol. Rev.* **14,** 447–475
7. Inagaki, Y., and Ford Doolittle, W. (2000) *Mol. Biol. Evol.* **17,** 882–889
8. Roger, A. J., Svard, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D., and Sogin, M. L. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 229–234
9. Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A., and Peattie, D. A. (1989) *Science* **243,** 75–77
10. Nixon, J. E., Wang, A., Morrison, H. G., McArthur, A. G., Sogin, M. L., Loftus,

[2] S. Garlapati and C. C. Wang, unpublished data.

B. J., and Samuelson, J. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 3701–3705
11. Adam, R. D. (2000) *Int. J. Parasitol.* **30,** 475–484
12. Yu, D. C., Wang, A. L., Botka, C. W., and Wang, C. C. (1998) *Mol. Biochem. Parasitol.* **96,** 151–165
13. Bruchhaus, I., Leippe, M., Lioutas, C., and Tannich, E. (1993) *DNA Cell Biol.* **12,** 925–933
14. Davis-Hayman, S. R., Shah, P. H., Finley, R. W., Lushbaugh, W. B., and Meade, J. C. (2000) *Parasitol. Res.* **86,** 608–612
15. Elmendorf, H. G., Singer, S. M., and Nash, T. E. (2001) *Nucleic Acids Res.* **29,** 4674–4683
16. Keeling, P. J., Fast, N. M., and McFadden, G. I. (1998) *J. Mol. Evol.* **47,** 649–655
17. Pestova, T. V., and Kolupaeva, V. G. (2002) *Genes Dev.* **16,** 2906–2922
18. Dever, T. E. (1999) *Trends Biochem. Sci.* **24,** 398–403
19. Prevot, D., Decimo, D., Herbreteau, C. H., Roux, F., Garin, J., Darlix, J. L., and Ohlmann, T. (2003) *EMBO J.* **22,** 1909–1921
20. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241
21. Elmendorf, H. G., Singer, S. M., Pierce, J., Cowan, J., and Nash, T. E. (2001b) *Mol. Biochem. Parasitol.* **113,** 157–169
22. Kirk-Mason, K. E., Turner, M. J., and Chakraborty, P. R. (1989) *Mol. Biochem. Parasitol.* **36,** 87–99
23. Nohria, A., Alonso, R. A., and Peattie, D. A. (1992) *Mol. Biochem. Parasitol.* **56,** 27–37
24. Wang, A. L., and Wang, C. C. (1986) *Mol. Biochem. Parasitol.* **21,** 269–276
25. Li, L., Wang, A. L., and Wang, C. C. (2001) *J. Virol.* **75,** 10612–10622
26. Yu, D. C., Wang, A. L., Wu, C. H., and Wang, C. C. (1995) *Mol. Cell Biol.* **15,** 4867–4872
27. Wang, X. C., Yang, J., Huang, W., He, L., Yu, J. T., Lin, Q. S., Li, W., and Zhou, H. M. (2002) *Int. J. Biochem. Cell Biol.* **34,** 983–991
28. Kozak, M. (2002a) *Mamm. Genome* **13,** 401–410
29. Kozak, M. (2002b) *Gene* **299,** 1–34
30. Dever, T. E. (2002) *Cell* **108,** 545–556
31. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15,** 8783–8798
32. Shine, J., and Dalgarno, L. (1975) *Eur. J. Biochem.* **57,** 221–230
33. Moll, I., Grill, S., Gualerzi, C. O., and Blasi, U. (2002) *Mol. Microbiol.* **43,** 239–246
34. Dennis, P. P. (1997) *Cell* **89,** 1007–1010
35. Garlapati, S., Chou, J., and Wang, C. C. (2001) *J. Mol. Biol.* **308,** 623–638
36. Garlapati, S., and Wang, C. C. (2002) *RNA* **8,** 601–611
37. Sachs, M. S., Wang, Z., Gaba, A., Fang, P., Belk, J., Ganesan, R., Amrani, N., and Jacobson, A. (2002) *Methods* **26,** 105–114
38. Kozak, M. (1997) *EMBO J.* **16,** 2482–2492
39. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol.* **288,** 911–940