

# The Flatworm Spliced Leader 3'-Terminal AUG as a Translation Initiator Methionine<sup>\*[S]</sup>

Received for publication, June 27, 2005, and in revised form, September 27, 2005 Published, JBC Papers in Press, October 17, 2005, DOI 10.1074/jbc.M506963200

Guofeng Cheng<sup>‡</sup>, Leah Cohen<sup>§</sup>, David Ndegwa<sup>§</sup>, and Richard E. Davis<sup>‡§1</sup>

From the <sup>‡</sup>Departments of Pediatrics and Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045 and the <sup>§</sup>Department of Biology, City University of New York Graduate Center, College of Staten Island, Staten Island, New York, 10314

Spliced leader (SL) RNA trans-splicing contributes the 5' termini to mRNAs in a variety of eukaryotes. In contrast with some trans-splicing metazoan groups (e.g. nematodes), flatworm spliced leaders are variable in both sequence and length in different flatworm taxa. However, an absolutely conserved and unique feature of all flatworm spliced leaders is the presence of a 3'-terminal AUG. We previously suggested that the *Schistosoma mansoni* spliced leader AUG might contribute a required translation initiator methionine to recipient mRNAs. Here we identified and examined trans-spliced cDNAs from a large set of newly available schistosome cDNAs. 28% of the trans-spliced cDNAs have the SL AUG in-frame with the major open reading frame of the mRNA. We identified over 40 cDNAs (40% of the SL AUG in-frame clones) that require the SL AUG as an initiator methionine to synthesize phylogenetically conserved N-terminal residues characteristic of orthologous proteins. RNA transfection experiments using several schistosome stages demonstrated that the flatworm SL AUG can serve as a translation initiator methionine *in vivo*. We also present *in vivo* translation studies of the schistosome initiator methionine context and the effect of the spliced leader AUG added upstream and out-of-frame with the main open reading of recipient mRNAs. Overall, our data have provided evidence that another function of flatworm spliced leader trans-splicing is to provide some recipient mRNAs with an initiator methionine for translation initiation.

Spliced leader (SL)<sup>2</sup> RNA trans-splicing is an RNA processing event that forms the mature 5' end of mRNAs (1–3). The two substrates for the reaction are 1) a small RNA, the spliced leader RNA, and 2) a pre-mRNA. The spliced leader RNA consists of a small exon, the spliced leader, ranging in size from 15 to 52 nucleotides. The SL exon is followed by a 5' splice site and an intron without a 3' splice site. The recipient pre-mRNA substrate begins with an outtron (an intron lacking a 5' splice site) followed by an exon or several exons and introns. Trans-splicing adds the spliced leader sequence to the 3' splice site of the outtron, generating the mature 5' end of the mRNA. Spliced leader addition

provides not only the 5'-terminal sequence but also brings a new cap to the mature mRNA. In metazoa, a trimethylguanosine (TMG) cap (m<sup>2,2,7</sup>GpppN) is present on the spliced leader RNA, and thus the mRNAs acquire a TMG cap (4–9).

Trans-splicing is present in some sarcomastigophora protozoa, hydra, nematodes, rotifers, flatworms, and early chordates (8–20). SL addition thus represents a major form of gene expression present in divergent groups of eukaryotes. For some organisms and genes, a function for trans-splicing is known. Trans-splicing in concert with 3' end formation serves to resolve polycistronic transcripts into monocistronic mRNAs for almost all genes in the kinetoplastida, ~20% of *Caenorhabditis elegans* genes, several genes in the early chordate *Oikopleura*, and at least one schistosome gene locus (11, 21–27). In *C. elegans*, two spliced leaders have been identified, SL1 and SL2 (17, 28). SL2 trans-splicing participates in the resolution of polycistronic transcripts (21, 22, 24, 29). However, SL1 addition is responsible for the majority of trans-splicing in nematodes, accounting for 50–90% of trans-splicing in different nematodes. Notably, SL1 trans-splicing is not typically involved in resolution of polycistronic transcripts, and overall, its primary function remains unclear. Recently, we have shown that the addition of the TMG-capped SL1 to test transcripts does not have a significant effect on nematode mRNA stability and translation when compared with non-trans-spliced m<sup>7</sup>G-capped control RNAs (30). However, SL1 trans-splicing may play a role in nematode translation by adding the spliced leader at a relatively conserved distance from the AUG of the recipient RNA endogenous open reading frame that is optimal for translation of TMG-capped RNAs (30). Overall, however, the function of trans-splicing in several organisms remains unknown, and it remains to be determined what other functions trans-splicing may serve.

The spliced leader sequence varies among different trans-splicing organisms. However, within a group of trans-splicing organisms, the SL sequence typically exhibits high sequence conservation. For example, the 22-nucleotide SL1 spliced leader is almost absolutely conserved in all nematodes. The high sequence conservation of the nematode SL may be required as the sequence constitutes part of the SL RNA promoter in *Ascaris* and contributes to the ability to translate mRNAs with the trimethylguanosine cap (30, 31). Trans-splicing has been identified in four major and divergent flatworm groups, triclads, polyclads, trematodes (parasitic flukes), and cestodes (tapeworms) (8, 13–15, 20, 32). Flatworm spliced leaders exhibit a high degree of length and sequence variation, more than observed in some groups of trans-splicing organisms. The SL sequences range from 34 to 52 nucleotides in length (52 nucleotides is the longest SL known) and often have less than 50% nucleotide identity. The length and sequence variation in the SL is also accompanied by some variation in predicted flatworm SL RNA structure. Despite the significant SL length and sequence variation, several phylogenetically conserved nucleotides are present in divergent flatworm spliced

<sup>\*</sup> This work was supported by grants from Ellison Medical Foundation (Grant ID-IA-0037) and the National Institutes of Health (Grant AI49558) and by UCHSC startup funds (to R. E. D.). 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.

[S] The on-line version of this article (available at <http://www.jbc.org>) contains a supplemental table showing schistosome mRNAs using the SL AUG as an initiator methionine.

<sup>1</sup> To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Mail Stop 8101, P. O. Box 6511, 12801 East 17th Ave., Aurora, CO 80045. Tel.: 303-724-3226; Fax: 303-724-3215; E-mail: Richard.Davis@uchsc.edu.

<sup>2</sup> The abbreviations used are: SL, spliced leader; TMG, trimethylguanosine; ORF, open reading frame; UTR, untranslated region; FCS, fetal calf serum; EST, expressed sequence tag.

leaders. Among these are an invariant 3'-terminal AUG. This is a unique characteristic of the flatworm SL not seen in spliced leaders from other organisms. In our original studies on schistosomes, the presence of the 3'-terminal SL AUG led us ask whether a function of trans-splicing in flatworms might be to provide recipient mRNAs with an initiator methionine (33). Our initial sequence characterization of a small set of trans-spliced schistosome RNAs indicated that the majority of trans-spliced mRNAs contained an independent and endogenous 3' initiator methionine for the primary open reading frame of the recipient mRNA. For these mRNAs, the SL AUG was upstream and out-of-frame with the primary open reading frame of the mRNA. Interestingly, we also identified two mRNAs for which the SL AUG was in-frame with the primary open reading frame of the mRNA. However, no additional bioinformatic or functional data were available to support the use of the SL AUG as an initiator methionine.

Recently, large scale EST and cDNA sequencing efforts have been conducted on schistosome mRNAs (34, 35). In addition, during the last decade, genomic, cDNA, and EST sequencing efforts have increased enormously the number of predicted and known protein sequences. These new sequence data have enabled us to further examine the question of whether the flatworm SL AUG serves to provide recipient mRNAs with a required initiator methionine for expression of the endogenous mRNA open reading frame. In the current study, we have provided extensive bioinformatic data demonstrating that the flatworm spliced leader contributes an initiator AUG to some mRNAs necessary for the synthesis of N-terminal residues conserved in the orthologous proteins of other organisms. In addition, we developed and used methods for transient RNA transfection to demonstrate that the schistosome-spliced leader AUG can functionally serve as an initiator methionine *in vivo*. Our results have expanded the known functions of trans-splicing and demonstrate that spliced leader addition in flatworms serves to contribute an initiator methionine for expression of some recipient mRNA open reading frames.

## MATERIALS AND METHODS

**Bioinformatic Analyses**—Schistosome cDNAs with the spliced leader sequence were identified in the NCBI non-redundant and EST databases as well as the *Schistosoma mansoni* GeneDB using BLASTN. Spliced leader cDNAs were screened to eliminate clones containing non-terminal SL sequences. Non-terminal SL sequences can represent as high as 20–25% of the clones.

5'-terminal SL clones were analyzed to determine whether the SL AUG was in-frame with the major open reading frame (ORF) of the mRNA using either the NCBI ORF Finder or MacVector Software (Accelrys, San Diego, CA). cDNAs with two approximately equivalent ORFs were eliminated from the analysis, and only cDNAs with ORFs of at least 150 nucleotides were characterized. Clones with the SL AUG in-frame with the major mRNA ORF were then analyzed by BLASTX against protein databases. Resulting pairwise alignments derived from BLASTX were initially analyzed to determine whether 1) the SL AUG-initiated protein had similarity with proteins in the data base between the SL AUG and the next downstream AUG and 2) if the similarity in these residues was phylogenetically conserved among orthologous proteins. In many cases, the orthologous proteins were extracted from the databases and subjected to multiple sequence alignment with the predicted schistosome proteins to confirm conservation of residues. To be considered a cDNA that uses the SL AUG to initiate open reading frames, a minimum of 5 identical or functionally similar and phylogenetically conserved residues must be present between the SL AUG and the downstream AUG.

The majority of SL AUG-initiated open reading frames exhibit much higher similarity (a higher percentage of identity and similarity between the SL AUG and next downstream AUG) with orthologous proteins. In addition, most assigned SL AUG-initiated proteins are also present in diverse phylogenetic groups representing vertebrates and at least two invertebrate groups.

**RNA Preparation for Transfection**—PCR-generated templates for *in vitro* transcription were synthesized from pRL-null (*Renilla* luciferase) or pGL3 (firefly luciferase) (Promega, Madison, WI) using primers that introduce a T7 promoter and 5'-UTR sequences at the 5' end of the luciferase ORF and a 60- or 85-nucleotide poly-A tail at the 3' end. The *in vitro* transcription reactions (T7 Megascript, Ambion, Austin, TX) typically contained 1  $\mu$ g of template/20- $\mu$ l reaction, 10 mM ATP, 10 mM CTP, 10 mM UTP, 12 mM cap analogue ( $m^7$ GpppG or  $m_3^{2,2,7}$ GpppG), 1.5 mM GTP, 1 $\times$  reaction buffer, and 1 $\times$  T7 polymerase mix and were carried out at 37 °C for 4 h. After DNase I treatment, the mRNA transcripts were extracted with TRIzol (Invitrogen), and the RNAs were precipitated twice, once with isopropyl alcohol and then with 0.5 M ammonium acetate/ethyl alcohol. Precipitated RNAs were further washed with 70% ethyl alcohol, dissolved in water, quantitated spectrophotometrically, and examined by agarose-formaldehyde denaturing gel electrophoresis. Analysis of cap orientation on transcripts was carried out as described (36) and indicated that greater than 80% of the caps were added in the correct orientation.

**Schistosome Materials**—Mice and hamsters infected with *S. mansoni* were provided by an National Institutes of Health contract through the Biomedical Research Institute and Fred Lewis. Additional *S. mansoni*-infected mice and hamsters were generously provided by Edward Pearce and Philip LoVerde, respectively. Adult schistosomes were perfused from the hepatic portal vasculature and maintained in RPMI 1640 with 10% FCS and 200  $\mu$ g/ml penicillin/streptomycin at 37 °C and 5% CO<sub>2</sub>. Eggs were isolated from livers by gentle homogenization of infected livers in 1.5% saline followed by sequential filtering of the eggs through screens as described previously (37, 38). Purified eggs were suspended in spring or MilliQ water to stimulate hatching of the miracidial larval stage. The larval stages were collected by migration toward light and concentrated by gentle centrifugation at 4 °C (38). Miracidia were either used directly for biolistic experiments or transformed into sporocysts by 18–24 h of incubation in MEMSE-J with 1% bovine serum albumin and 200  $\mu$ g/ml gentamycin at 26 °C in 5% CO<sub>2</sub> and O<sub>2</sub> (39, 40). Following 18–24 h of incubation, the transformed sporocysts were maintained at 26 °C in MEMSE-J with 5% FCS and 200  $\mu$ g/ml penicillin/streptomycin in 5% CO<sub>2</sub> and O<sub>2</sub>.

**RNA Biolistics**—Preparation of gold microcarriers and biolistics were performed essentially as described previously (41) with the following modifications: 1) gold particles were spherical  $\sim$ 2.2  $\mu$ m gold from Degussa (10KM)(dmc2 Metals Group, South Plainfield, NJ) and 2) instead of lyophilization of the RNA onto the gold carriers, an alcohol precipitation was used. *In vitro* transcribed RNAs were precipitated onto gold particles using 2.5 M ammonium acetate/ethyl alcohol precipitation using either 0.25–5  $\mu$ g of *Renilla* luciferase reporter RNA or 0.25–5  $\mu$ g of firefly luciferase RNA/1 mg of gold particles. Luciferase activity in worms is RNA dose-dependent. The RNA/gold pellet was washed with 200  $\mu$ l of ice-cold ethyl alcohol, resuspended in ice-cold 100% ethyl alcohol (18  $\mu$ l/mg gold), and spread onto macrocarriers and processed as described previously (41). Binding of RNA onto the gold particles was evaluated by formaldehyde agarose gel analysis to determine the integrity of RNA bound to the beads, and RiboGreen (Molec-

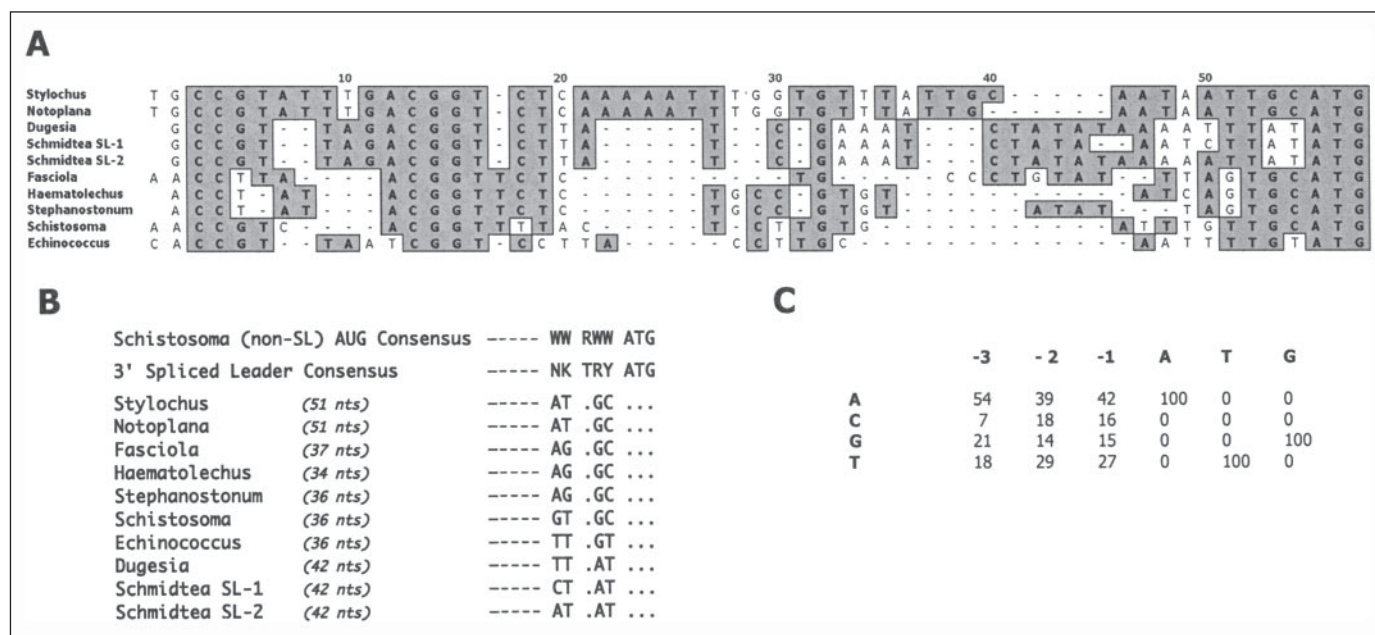


FIGURE 1. **Conservation of flatworm spliced leader 3'-terminal AUG.** A, multiple sequence alignment of flatworm spliced leaders (8, 13–15, 20, 32). B, consensus sequences for schistosome translation initiation (*non-SL* = non-trans-spliced mRNAs) and the 3' ends of flatworm spliced leaders. The translation initiation consensus was determined as described (33) on 300 characterized schistosome mRNAs. The length and 3' ends of the flatworm SLs are illustrated. Dots within the consensus region indicate identical nucleotides (nts). The dashes preceding the consensus represent upstream sequence that is variable in the spliced leaders. C, nucleotide frequencies at positions –1 to –3 for 300 non-trans-spliced schistosome cDNAs.

ular Probes) (42) was used to quantitate RNA loaded onto the particles for data normalization. Analysis of RNA bound to beads indicated that little to no degradation of the RNA occurred prior to particle bombardment and that typically 80–85% of the precipitated RNA was bound onto the gold particles.

Freshly perfused adult schistosomes were incubated in medium in 5% CO<sub>2</sub> at 37 °C for 3 h to overnight prior to bombardment. 30 worm pairs in medium were carefully placed into the center of 35-mm Petri dishes, and just prior to bombardment, the medium was removed. The adults were bombarded with 1 mg of RNA-coated gold particles in a Bio-Rad Biolistic PDS-1000/HE particle delivery system at 15 inches of mercury of chamber vacuum, target distance of 3 cm (stage 1), and 1,550-psi particle acceleration pressure. In many experiments, the adults were immediately bombarded a second time. This effectively doubles the level of expression observed. After bombardment, 2–3 ml of RPMI plus 10% FCS and 200 µg/ml penicillin/streptomycin was added, and the adults were incubated at 37 °C in 5% CO<sub>2</sub> for 3 h before the adults were collected, washed in phosphate-buffered saline, pelleted, and frozen prior to lysis for luciferase assays. Evaluation of gold particle delivery was carried out as described previously (41).

For particle bombardment of miracidia and sporocyst, freshly isolated miracidia or 18–24-h transformed sporocysts were suspended in MEMSE-J with 1% bovine serum albumin and 200 µg/ml gentamycin, and ~1000 miracidia or sporocysts were gently spread onto the center of 35-mm Petri dishes in a minimal volume. Optimal biolistics parameters were determined to be 15 inches of mercury of chamber vacuum, target distance of 6 cm (stage 2), and 450-psi particle acceleration pressure. Following bombardment, 2 ml of Media F or MEMSE-J plus 1% bovine serum albumin or 5% FCS were added, and the miracidia or sporocysts were incubated at 26 °C in 5% CO<sub>2</sub> and O<sub>2</sub> for 3 h before they were collected, washed in media, pelleted, and frozen prior to lysis for luciferase assays. Bombarded miracidia develop into viable sporocysts that can be maintained for several weeks.

**RNA Electroporation of Sporocysts**—600–2000, 18–24-h-old sporocysts were resuspended in 100 µl of MEMSE-J with 1% FCS. The sporocysts were placed into prechilled 0.4-cm electroporation cuvettes, RNA was added to 50 µg/ml, and electroporation was carried out in a BTX ECM 830 (square wave) electroporator (BTX, Holliston, MA) at 280 volts using a 0.5-ms pulse length. Following electroporation, the sporocysts were suspended in 2 ml of MEMSE-J with 5% FCS/200 µg/ml penicillin/streptomycin and incubated at 26 °C in 5% CO<sub>2</sub> and O<sub>2</sub> for 3 h before they were collected, washed in media, pelleted, and frozen prior to lysis for luciferase assays.

**Luciferase and Protein Assays**—Frozen pellets of adult or larval schistosome stages were resuspended in 150 µl of the recommended lysis buffer (Promega, Madison, WI) and homogenized in Kontes disposable PEL-LET PESTLES® with Microtubes (Vineland, NJ) several times using several freeze-thaw cycles and multiple passes with a motor-driven pestle. The lysates were cleared by centrifugation, and aliquots (20 µl) of the supernatant were assayed using the Promega *Renilla*, luciferase, or Dual-Luciferase assay systems with a Sirius luminometer (Zylux Corp., Oak Ridge, IN) (41). Dual-Luciferase assays were used in conjunction with co-transfection of two reporters (firefly and *Renilla* luciferase) as normalization for transfection efficiency and biological variation. Protein assays on lysates were carried out using the Pierce BCA protein assay kit and Compat-Able protein assay preparation reagent set (Pierce). Experimental data were similar with or without normalization to a co-transfected reporter and to protein concentrations in sample lysates. Thus, some experimental data are presented with normalization, whereas other data are not.

## RESULTS

**Conservation of the Flatworm SL AUG and Its Sequence Context**—An alignment of the known flatworm spliced leaders and a subalignment of their 3' ends are presented in Fig. 1, A and B. The alignments illustrate the absolute sequence conservation of the 3'-terminal AUG of flatworm spliced leaders as well as the AUG sequence context. We previously



derived a schistosome translation initiation AUG consensus from a small set of non-trans-spliced schistosome mRNAs. Re-evaluation of this consensus in light of additional sequences (~300) indicates that it is still valid with the larger sampling (33) (Fig. 1C). The non-trans-spliced mRNA translation initiation consensus along with the consensus of the flatworm SL AUG are shown in Fig. 1B. The schistosome translation initiation AUG context for non-trans-spliced mRNAs conforms to that observed in higher eukaryotes and a variety of invertebrates (43–45) with a highly conserved purine in the –3 position (Fig. 1C). Notably, all the flatworm SL AUGs have a U in the –3 position instead of a purine.

**Bioinformatic Analysis of SL cDNAs and ESTs**—We previously noted that the schistosome SL AUG was typically not in-frame with major open reading frames of recipient mRNAs (33). However, we observed two mRNAs (corresponding to ~5% of the trans-spliced mRNAs we originally identified) for which the SL AUG was in-frame with the major ORF of the recipient mRNA. Available bioinformatic data at the time did not support the conservation of the protein sequence between SL AUG and the next 3'-AUG with orthologous protein sequences in databases. The large increase in known and predicted protein sequences from a breadth of organisms, as well new EST and cDNA data for schistosomes (34, 35), led us to further examine whether we could derive additional data to support the use of the conserved flatworm spliced leader AUG as an initiator methionine. We re-examined trans-spliced mRNAs we previously described and identified a large number of new trans-spliced schistosome mRNAs in both EST and nucleotide databases (~250,000 ESTs and cDNAs) (34, 35). These mRNA sequences were derived from two different schistosome species, *S. mansoni* and *Schistosoma japonicum*, which have identical SL sequences. Only a small percentage of the known cDNAs and EST sequences contains the schistosome SL. For example, of the ~150,000 known *S. mansoni* EST sequences, ~245 (0.16%) clones contained the spliced leader, whereas of the ~97,000 *S. japonicum* EST sequences, ~845 (0.9%) have the spliced leader. These values are an under-representation of the frequency of mRNAs with an SL for several reasons. First, unless specialized methods are used, full-length cDNAs are typically under-represented in libraries. Second, many of the *S. mansoni* ESTs represent sampling of internal cDNA fragments and cDNA synthesis, and sampling did not emphasize preparation or analysis of full-length clones (34). Recently, ~9,500 *S. japonicum* cDNAs sequences were deposited into the nucleotide data base. Of these ~9,500 cDNAs, 216, or 2.3% of the clones, have a 5'-terminal spliced leader (270 clones have the SL, but 54 of these clones exhibited the SL in a non-terminal position). For our analyses, we examined all available cDNAs with a 5'-terminal SL in current databases (NCBI nucleotide and dbEST and *S. mansoni* GeneDB) to identify clones that might use the SL AUG as an initiator methionine. However, the 216 *S. japonicum* SL cDNAs derived from the newly deposited 9,500 cDNAs were used to calculate the percentages of clones with the SL AUG in-frame and other frequencies as these are primarily full-length and better characterized mRNAs.

The SL sequence is predominantly out-of-frame with the AUG of the major open reading frame for the majority of the SL mRNAs (72%). 28% of the cDNAs and ESTs contain the SL AUG in-frame with the main open reading frame. Many mRNAs often use the first in-frame AUG as the initiator methionine, particularly when present in an appropriate sequence context. However, there are a growing number of examples in which the first AUG may not be used as the initiation codon (46–48). In addition, as little is known regarding translation initiation in flatworms, it is possible that trans-spliced mRNAs with the potentially poor sequence context of the SL AUG (it lacks a purine at –3 when compared with non-trans-spliced schistosome mRNAs and other invertebrates)

might preferentially initiate translation from a downstream AUG in a better context.

N-terminal sequence for proteins derived from trans-spliced mRNAs containing the SL AUG in-frame with major ORF would provide definitive data regarding the initiator methionine of these proteins. However, these data are exceptionally difficult and expensive to obtain, particularly in parasitic organisms for which limited material is available. In the absence of these data, we chose to examine predicted SL AUG-initiated proteins for conserved amino acid residues between the SL AUG and the next 3'-AUG in orthologous proteins. If a predicted SL AUG-initiated protein contained conserved residues in orthologous proteins between the SL AUG and the next downstream AUG, these data would provide strong support that the SL AUG would be required for translation to produce a protein with N-terminal residues that are phylogenetically conserved for that particular protein. Of those cDNA clones that have the SL AUG in-frame with the major ORF (SL AUG in-frame clones), 35% have an in-frame Met within ~10 amino acids 3' of the SL AUG (10% of all SL clones identified). None of these cDNAs passed the similarity criterion described above for the amino acids between the SL AUG and the next in-frame AUG. These cDNAs may use the SL AUG (see below), or in some cases, the downstream AUG in these mRNAs might be used as the translation initiator methionine via a leaky scanning mechanism (49, 50). Of the remaining SL AUG in-frame clones, ~40% exhibited phylogenetically conserved sequence similarity between the SL AUG and the downstream methionine, supporting the use of the SL AUG as an initiator methionine. Thus, over 40 schistosome mRNAs representing ~10% of the total SL-containing clones were identified in the different databases that meet our criterion (see supplemental Table 1). Alignments illustrating the phylogenetic conservation of SL AUG-initiated N-terminal residues are shown in Fig. 2. These alignments are representative of the types of matches obtained and are presented to illustrate the evidence for conservation of N-terminal residues. Most alignments for predicted proteins initiated by the SL are strongly supportive of the SL AUG to generate a functional protein based on phylogenetic conservation of residues in orthologous proteins (Fig. 2). However, some alignments (~15%) are not as supportive as those shown. The majority of the predicted schistosome proteins are similar to known proteins with a discrete function. Only a small percentage of the proteins is similar to unknown, hypothetical proteins. This suggests that the use of the SL AUG for translation initiation is not generally a property of schistosome-specific proteins. Overall, these data provided strong support that a number of trans-spliced schistosome mRNAs would require the SL AUG as an initiator methionine to produce proteins with the requisite phylogenetically conserved N-terminal residues.

**Luciferase RNA Reporter Transfection**—We developed methods for RNA transfection using biolistics and electroporation into several stages of the schistosome life cycle, including miracidia (a free-swimming stage infecting the freshwater snail intermediate host), sporocysts (derived from miracidia that undergo polyembryony within the intermediate snail host), and adults (the sexual form in the vertebrate host). Following optimization of parameters for both biolistic and electroporation methods (see “Materials and Methods” for conditions used), we prepared luciferase (firefly or *Renilla*) RNAs to evaluate whether the schistosome SL AUG could be used to produce functional luciferase protein *in vivo* as measured by assay for luciferase activity. The initial test RNAs synthesized are illustrated in Fig. 3A. Wild-type firefly or *Renilla* luciferase RNAs were used as controls for transfection and translation. The firefly RNA control (firefly) consisted of an m<sup>7</sup>GpppG-

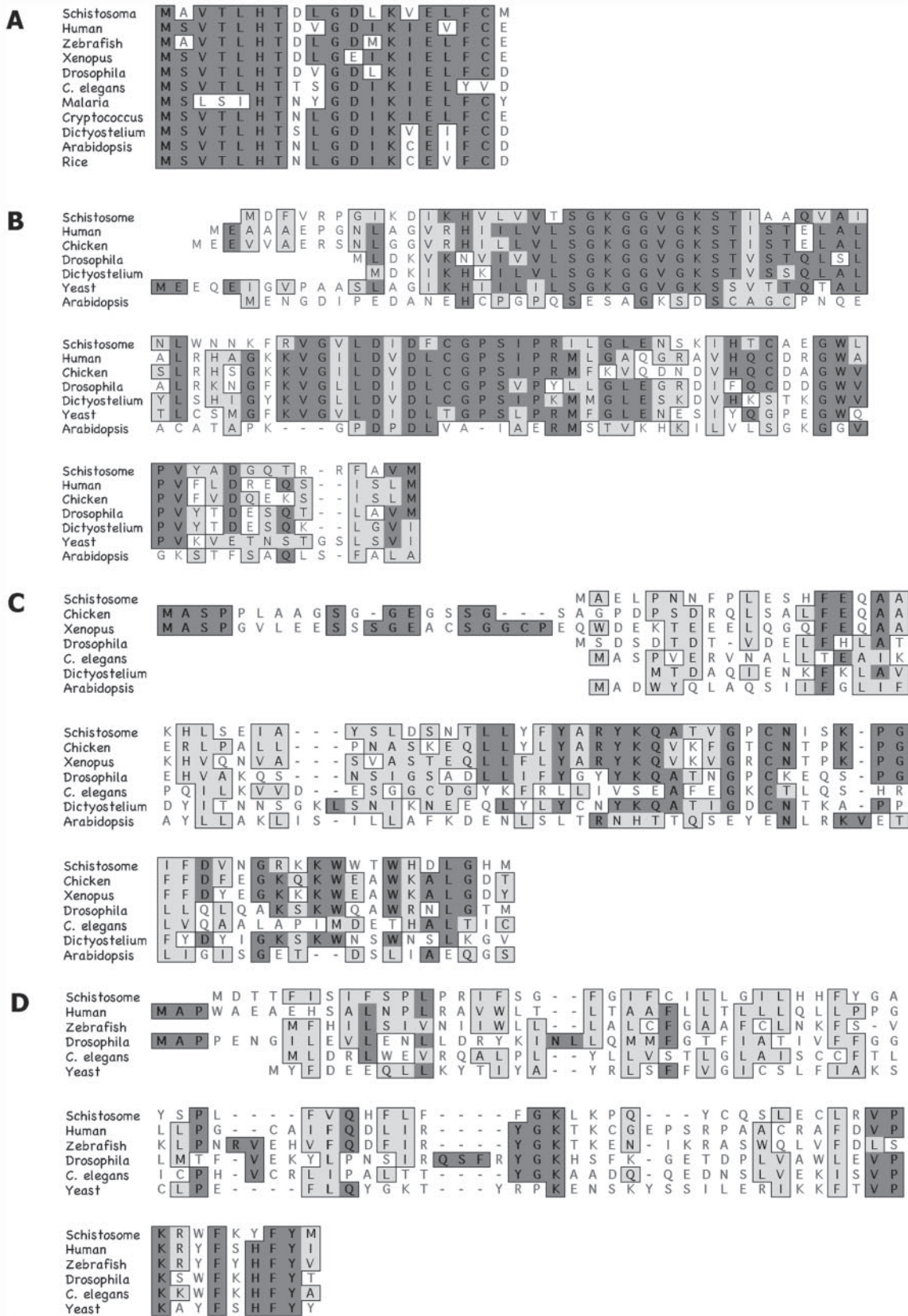
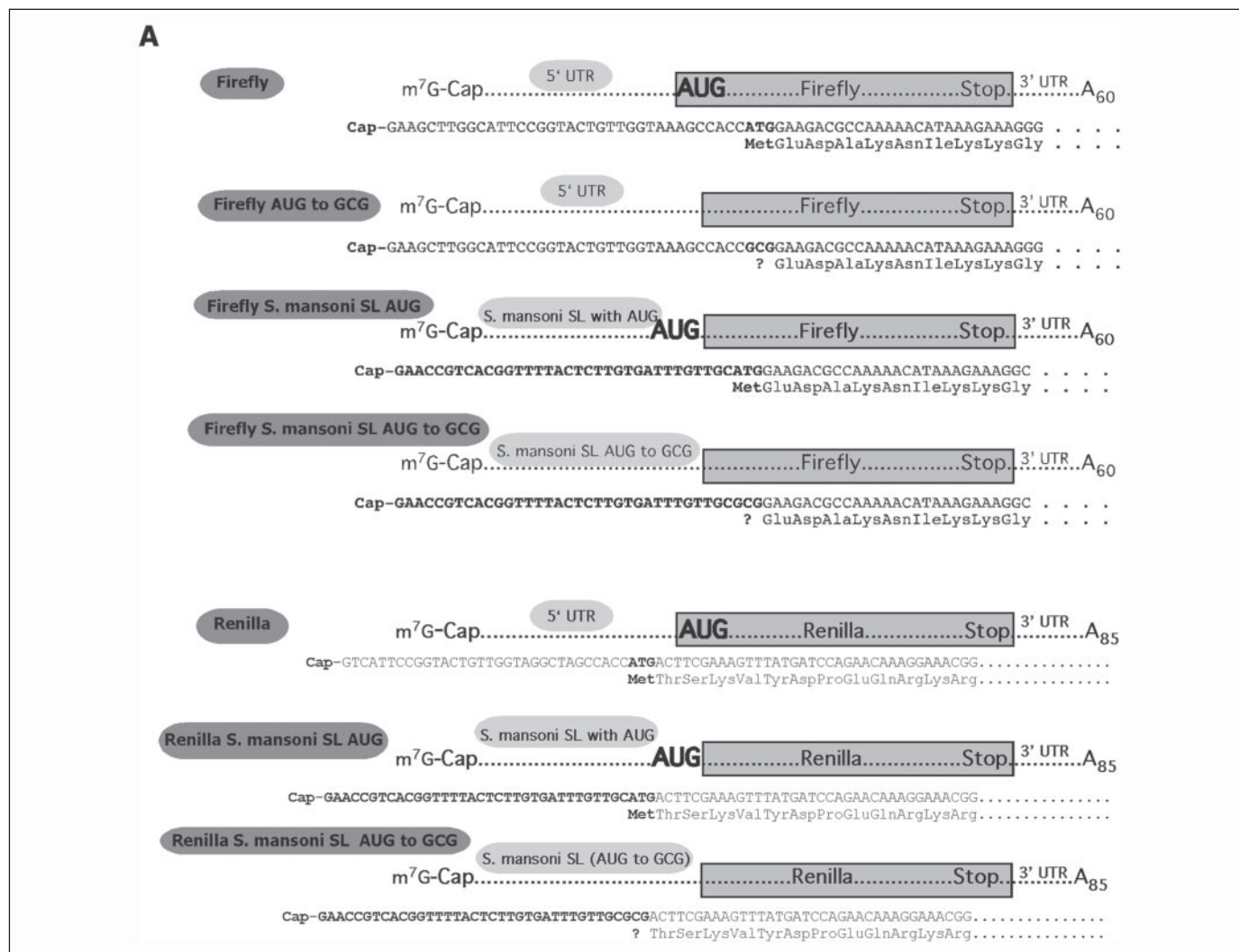


FIGURE 2. **N-terminal residues of SL AUG-initiated proteins are conserved in orthologous proteins.** Predicted proteins encoded by the SL AUG to the next in-frame AUG were aligned with orthologous proteins using the MacVector implementation of Clustal. *Darkly shaded residues* are identical amino acids, and *lightly shaded residues* are functionally similar. These are representative alignments for the proteins identified. *A*, schistosome accession number AY811984, cyclophilin J (peptidylprolyl isomerase protein 3). *B*, schistosome accession number AY223260, nucleotide-binding protein 2 (Mrp, ATPase involved in chromosome partitioning). *C*, schistosome accession number AY815507, acyl CoA binding domain protein. *D*, schistosome accession number AY812760, steroid, 3-oxo-5- $\alpha$ -steroid 4-dehydrogenase (steroid 5- $\alpha$ -reductase).



**FIGURE 3. The schistosome SL AUG can serve as an initiator methionine for *in vivo* luciferase translation.** A, illustration of 5' sequence of RNAs and initiator AUGs for firefly or *Renilla* RNA transfection. B, firefly luciferase activity following biolistic introduction of test RNAs into schistosome adults, miracidia, and sporocysts. Test RNAs were introduced into different schistosome life cycle stages, the schistosomes were incubated for 3 h for translation of the RNA, the schistosomes were collected and lysates were prepared, and the lysates were assayed for luciferase activity. C, normalized firefly luciferase activities in co-transfected, electroporated sporocysts. The illustrated firefly luciferase RNAs were electroporated along with a second *Renilla* RNA as a transfection control into schistosome sporocysts (see *Renilla* in A). Luciferase activity in sporocysts was determined 3 h after transfection using the Dual-Luciferase assay. Firefly luciferase activity is illustrated normalized to both *Renilla* luciferase activity and protein concentration in sporocyst lysates. D, normalized *Renilla* luciferase activities in firefly co-transfected, electroporated sporocysts. Luciferase activity in sporocysts was determined 3 h after transfection using the Dual-Luciferase assay. *Renilla* luciferase activity is illustrated normalized to co-transfected firefly luciferase RNA activity and protein concentration in sporocyst lysates. These are representative experiments illustrating the mean  $\pm$  S.E. with three replicates for each RNA sample evaluated.

capped 36-nucleotide 5'-UTR, firefly luciferase ORF, 50-nucleotide SV40 3'-UTR, and a 3' 60-nucleotide poly(A)-tail. The control *Renilla* RNA (*Renilla*) consisted of an m<sup>7</sup>GpppG-capped 32-nucleotide 5'-UTR, *Renilla* ORF, 50-nucleotide SV-40 3'-UTR, and 3' 85-nucleotide poly(A)-tail. To test whether the spliced leader AUG could initiate translation of luciferase protein, the entire 5'-UTRs of these RNAs were substituted with the 36-nucleotide schistosome SL, placing the SL AUG as the initiator methionine in-frame with the *Renilla* or firefly luciferase ORF (firefly or *Renilla* S. mansoni SL AUG). Test RNAs were introduced into schistosomes either by biolistics (adults, miracidia, and sporocysts) or electroporation (sporocysts). 3 h following RNA introduction, the samples were collected, frozen, lysed, and assayed for luciferase activity. As illustrated in Fig. 3B, introduction of the control firefly luciferase RNAs into different schistosome stages led to significant levels of luciferase activity in parasite lysates 3 h after RNA introduction. Replacement of the positive control RNA 5'-UTR with the schistosome-spliced leader such that the SL AUG replaced the luciferase AUG

also resulted in luciferase activity in the parasite lysates. These data suggested that the schistosome-spliced leader AUG can function as a translation initiator methionine. To further test whether the luciferase activity observed was dependent on the SL AUG as the initiator methionine, we synthesized RNAs with the SL AUG mutated to GCG (firefly S. mansoni SL AUG to GCG and *Renilla* S. mansoni SL AUG to GCG). Mutation of either the wild-type luciferase or schistosome-spliced leader initiator AUG to GCG (AUG to GCG mRNAs) resulted in almost complete loss of luciferase activity in schistosomes (Fig. 3, B–D). These data are consistent with previous experiments demonstrating that the N-terminal 8 amino acids of firefly luciferase are required for luciferase activity (51). Data illustrated in Fig. 3, C and D, were normalized for both transfection efficiency (co-transfection with a second RNA reporter) and protein content in schistosome lysates. Thus, differences in luciferase activity were not a function of variation in transfection frequencies or schistosome material. These data demonstrated that the schistosome SL AUG can serve as an initiator methionine for translation.



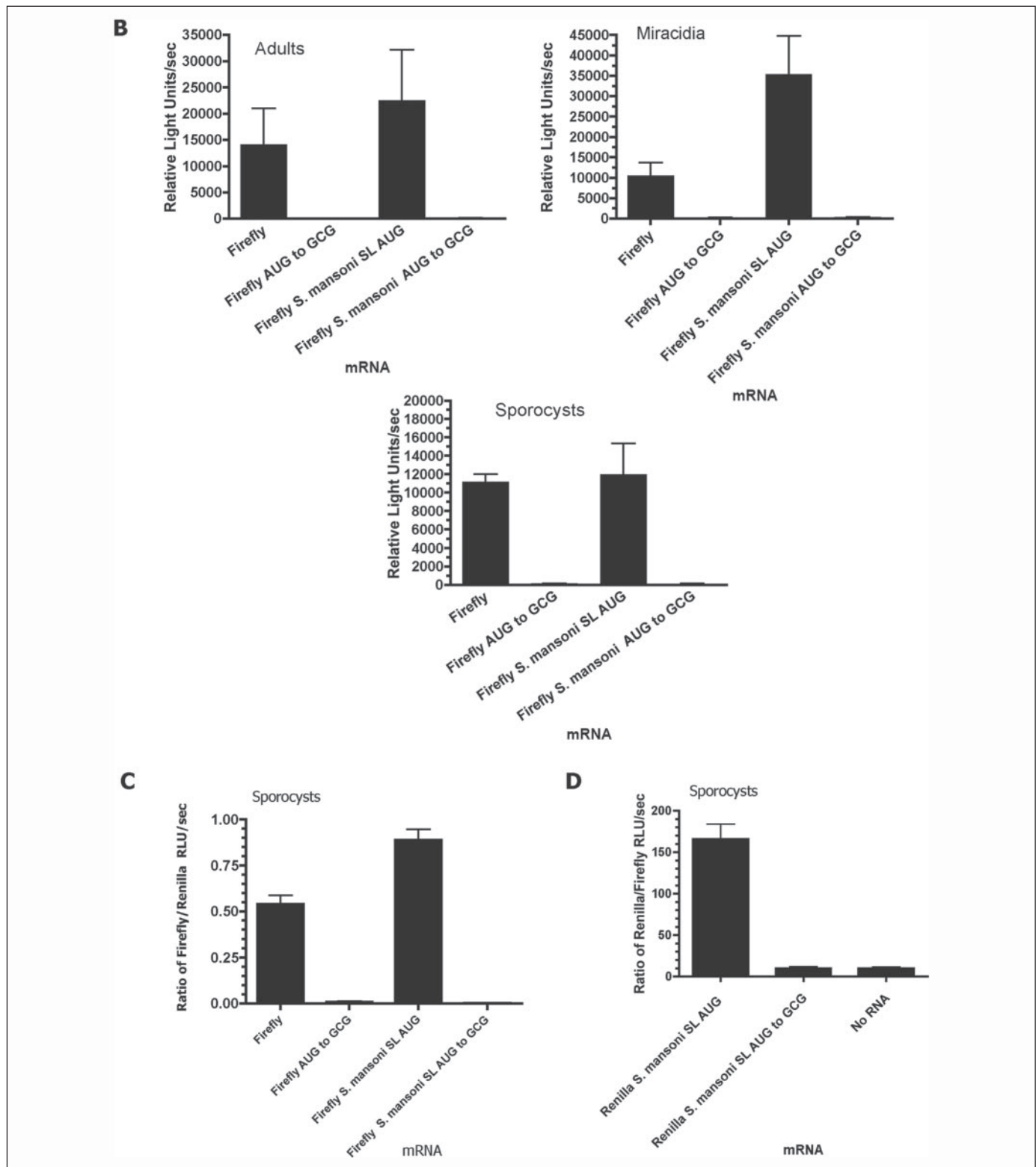


FIGURE 3—continued

**Spliced Leader AUG Context and Translation**—The SL AUG initiation context contains a U at  $-3$ . This is often considered a poor initiation context for some organisms. In a sampling of 300 non-trans-spliced schistosome cDNAs, U was present at  $-3$  18% of the time (Fig. 1C). To further examine the SL AUG initiation context and the contribution of the spliced leader sequence to translation initiation efficiency, we com-

pared the AUG initiation context as a function of the nucleotide at the  $-3$  position and in the context of the spliced leader sequence. As illustrated in Fig. 4, substitution of the SL U at the  $-3$  position to A led to significant increases in translation efficiency (compare SL UGC AUG to SL AGC AUG) for both firefly and *Renilla* reporters with the spliced leader. An A at the  $-3$  position is the preferred nucleotide present 54%

## Flatworm Spliced Leader Initiator Methionine

of the time in non-trans-spliced schistosome cDNAs (Fig. 1C). Thus, the increases in translation efficiency with A at  $-3$  are consistent with the apparent preferred AUG context in non-trans-spliced schistosome cDNAs (Fig. 4).

As the UGC of the SL AGU is apparently a suboptimal context for initiation (when compared with AGC), we examined whether the upstream SL sequence contributed to the translation context by comparing translation of the SL UGC AUG with the first 30 nucleotides of the SL substituted with random nucleotides (N30). Notably, the translation efficiency of the N30 was as efficient as the SL in the 5'-UTR of the reporter RNA. These data indicated that the SL sequence itself does not contribute significantly to the use of the UGC AUG context. This may not be unexpected as 18% of schistosome cDNAs have a U at  $-3$ , and at least for the firefly luciferase RNA, mRNAs with UGC AUG are relatively efficiently translated (Fig. 4A). Overall, these data indicated that although a purine at  $-3$  may be optimal in some contexts, mRNAs with a U at  $-3$  can be relatively efficiently translated.

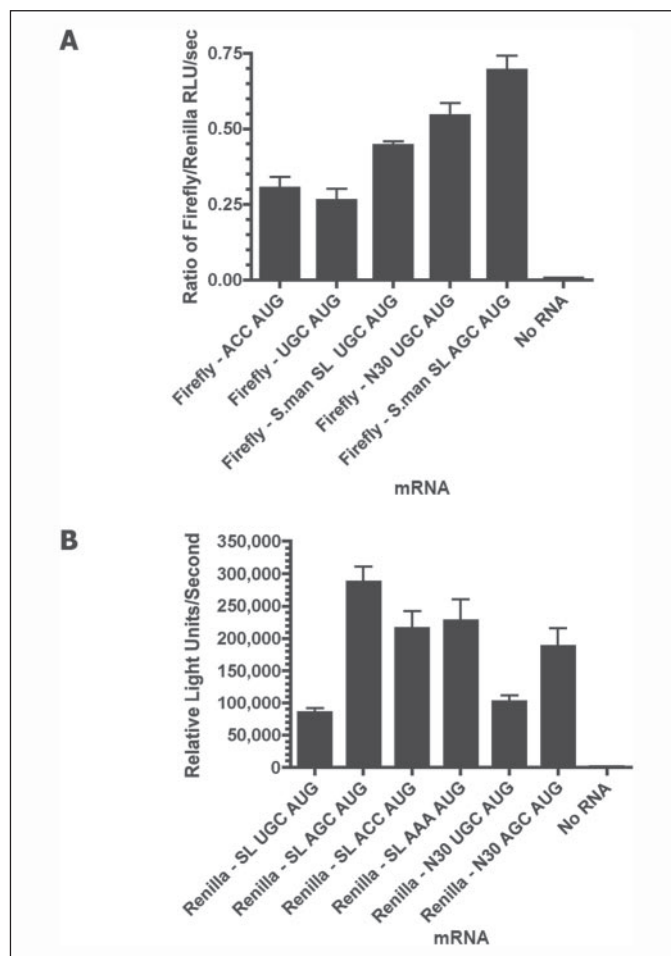
**TMG Cap and in Vivo Translation**—The addition of the flatworm spliced leader to the mRNA also results in the addition of an  $m_3^{2,2,7}$ GpppA cap to the RNA (8). Substitution of an  $m_3^{2,2,7}$ GpppG for the  $m_7$ GpppG cap on the *Renilla* control RNA led to a reduction in translation when compared with the  $m_7$ GpppG capped *Renilla* control RNA (Fig. 5). The addition of the TMG cap in the context of the SL did not improve translation of TMG-capped transcripts. Overall, reporter RNAs with the TMG and spliced leader were translated less efficiently than the  $m_7$ G-capped control RNA.

**SL AUG to Endogenous ORF AUG Distance**—The addition of the spliced leader to recipient mRNAs that do not use the SL AUG as an initiator methionine places an AUG upstream from the major ORF AUG. In other organisms, this is known to be a potentially confounding factor that might require either leaky scanning or, if a short ORF is present, reinitiation mechanisms to facilitate initiation at the major ORF (44, 49, 50). To determine whether the flatworm SL might be added at a conserved distance from the endogenous ORF AUG, we examined 160 *S. japonicum* mRNAs to define the distance between the SL and ORF of endogenous AUG-initiated mRNAs. As shown in Fig. 6A, 50% of the clones acquire the spliced leader within 30 nucleotides of the major open reading frame AUG. The distribution within the 30 nucleotides is further illustrated in Fig. 6B.

**Effect of an Upstream SL AUG on Downstream AUG Translation**—To evaluate the effect of the upstream and out-of-frame SL AUG on translation of a downstream open reading frame, we selected the 5'-UTRs of two native mRNAs for analysis, one encoding a high voltage-activated calcium channel  $\beta$  subunit 2 and another encoding a U2 small nuclear ribonucleoprotein B (Fig. 7A). The distance between the SL AUG and the open reading frame AUG for these mRNAs is 13 and 16 nucleotides, respectively. These distances fall within the typical range of the SL AUG to the ORF AUG as illustrated in Fig. 6. The translation efficiency of reporter RNAs with these 5'-UTRs introduced into sporocysts was compared with the wild-type control firefly 5'-UTR and the SL AUG 5'-UTR. Notably, as illustrated in the Fig. 7B, the upstream and out-of-frame SL AUG did not have a negative effect on firefly reporter translation. In fact, the 5'-UTRs derived from SL AUG out-of-frame mRNAs led to higher levels of luciferase activity in schistosome lysates than either the wild-type or SL AUG translation-initiated reporter.

## DISCUSSION

Our data have demonstrated that the spliced leader AUG is used as an initiator methionine for the open reading frames of some recipient

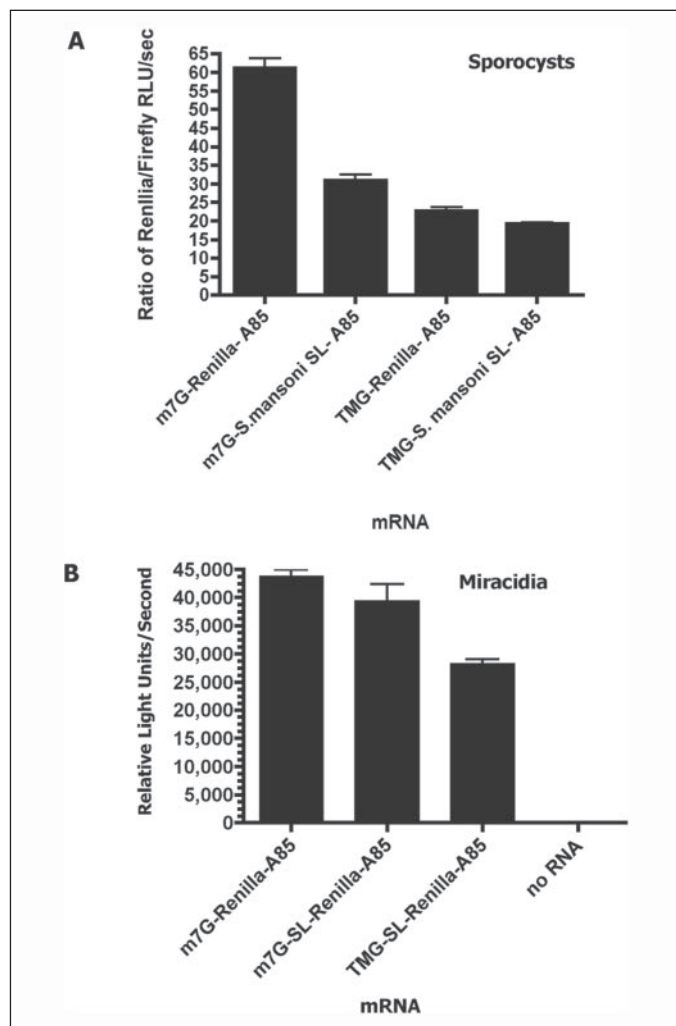


**FIGURE 4. Analysis of the schistosome AUG initiation context.** A, firefly luciferase RNA reporter AUG initiation context. Sporocysts were electroporated with the illustrated firefly RNAs and a co-transfected *Renilla* luciferase RNA control. Luciferase activity was measured in sporocyst lysates 3 h after transfection using the Dual-Luciferase assay. Data were normalized to the co-transfected *Renilla* reporter and protein concentrations in the lysates. The firefly-S. man SL UGC AUG RNA (firefly *S. mansoni* SL AUG) and firefly-ACC AUG RNA (firefly) are illustrated in Fig. 3A. The firefly-S. man SL AGC AUG differs only in the substitution of an A at the  $-3$  position, whereas the firefly-UGC AUG differs by substitution of UG at positions  $-2$  and  $-3$ . Firefly-N30 UGC AUG contains the last 6 nucleotides of the spliced leader (UGC AUG) with the first 30 nucleotides of the spliced leader substituted with random nucleotides. B, *Renilla* luciferase RNA reporter AUG initiation context. Sporocysts were electroporated with the illustrated *Renilla* RNAs, and *Renilla* luciferase activity was measured 3 h after transfection. RNA designations are as in A except that they are *Renilla* RNAs. These are representative experiments illustrating the mean  $\pm$  S.E. with three replicates for each RNA sample evaluated.

schistosome mRNAs. This conclusion has been supported by protein similarity data that require the SL 3'-AUG as an initiator methionine to initiate synthesis of a protein with conserved N-terminal residues characteristic of orthologous proteins. In addition, RNA transfection experiments demonstrated that the schistosome SL AUG can be used as an initiator methionine for translation of test luciferase mRNAs.

**Bioinformatic Analyses of SL AUG-initiated Proteins**—28% of the schistosome SL cDNA clones analyzed have the SL AUG in-frame with the major open reading frame of the recipient mRNA. Approximately ~38% of these clones have a 2nd in-frame Met within ~10 amino acids of the SL AUG. Some or many of these clones could use the SL AUG as the initiator methionine since the RNA transfection data demonstrated that the SL AUG can act as an initiator. In addition, in some contexts, these adjacent AUG codons have been shown to increase translation, and it remains to be determined whether this is true in schistosomes (52). Using relatively strict bioinformatic criteria requiring phylogeneti-

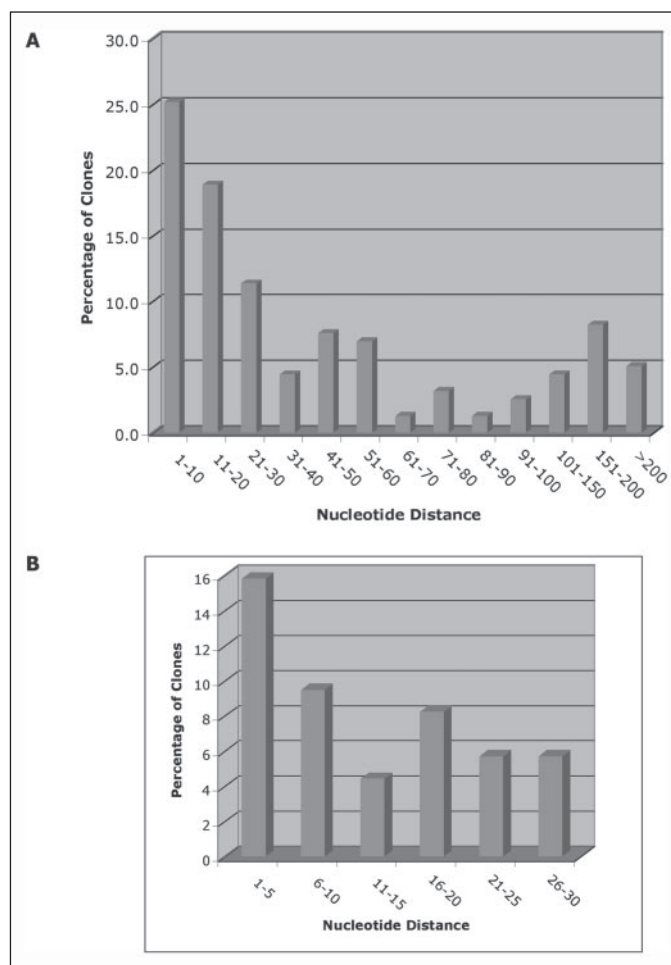




**FIGURE 5. Translation of TMG-SL RNAs in schistosomes.** *A*, sporocysts. The illustrated test *Renilla* RNAs were electroporated with a co-transfected Firefly luciferase RNA into sporocysts and luciferase activity measured 3 h after transfection using the Dual-Luciferase assay. Data were normalized to the co-transfected firefly reporter and protein concentrations in the sporocyst lysates. *B*, miracidia. The illustrated test *Renilla* RNAs were biologically introduced into miracidia, and *Renilla* luciferase activity was measured 3 h after transfection. The RNAs are as illustrated in Fig. 3A except for the substitution of a TMG cap on some RNAs. These are representative experiments illustrating the mean  $\pm$  S.E. with three replicates for each RNA sample evaluated. *m*<sup>7</sup>G, *m*<sup>7</sup>GpppG cap; *TMG*, = *m*<sub>3</sub><sup>2,2,7</sup>GpppG cap.

cally conserved sequence similarity between the residues encoded between the SL AUG and the next in-frame downstream AUG, we have identified a set of ~40 mRNAs that use the SL AUG. This is likely an underestimate, and a number of the mRNAs that do not pass these criteria may actually use the SL AUG as an initiator methionine. Analysis of the types of proteins encoded by clones with the SL AUG in-frame did not reveal any common patterns in types of proteins encoded.

**SL AUG Initiation Context**—Several features are known to influence the efficiency and/or fidelity of translation initiation at an AUG codon in vertebrates including 1) the mRNA cap, 2) the length of the 5'-UTR, 3) the context of the AUG codon, 4) the secondary structure 5' and 3' of the AUG, and 5) the presence of upstream AUGs or ORFs (49, 50). The levels of SL AUG-derived firefly luciferase activity were on average 50% higher than the control 5'-UTR for firefly luciferase. However, in contrast, *Renilla* luciferase activity with the SL AUG 5'-UTR was typically ~30% less than the control *Renilla* 5'-UTR (data not shown). Our interpretation of these data is that the SL AUG provides a reasonably efficient translation initiation context. However, a better understanding of

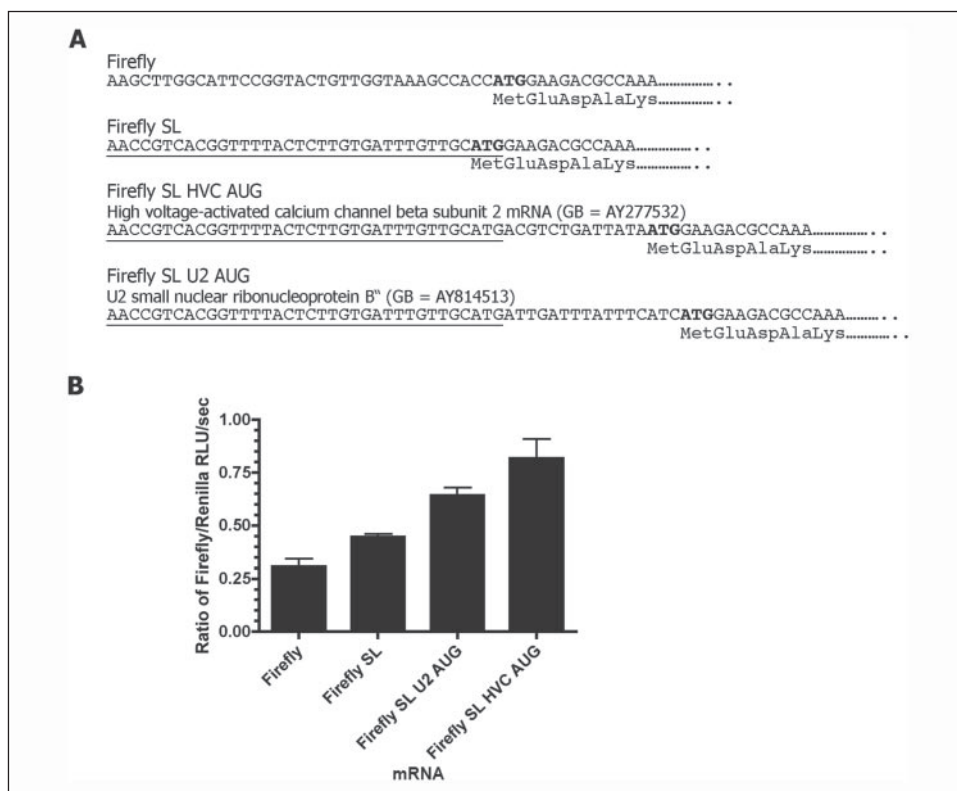


**FIGURE 6. Spliced leader AUG to predicted main open reading frame initiator methionine distance.** A set of 160 trans-spliced *S. japonicum* cDNAs that do not have the spliced leader AUG in-frame with the main open reading frame were analyzed for the nucleotide distance between the SL AUG and the initiator methionine of the ORF. The percentage of clones was plotted against the nucleotide distance. *A*, distance in 10-nucleotide intervals to 100 nucleotides, distance in 50-nucleotide intervals from 101 to 200, and distance greater than 200 nucleotides are illustrated. *B*, distance in 5-nucleotide intervals from 1 to 30 nucleotides is illustrated.

the comparative efficiency of the 5'-UTRs of trans-spliced *versus* non-trans-spliced mRNAs will require additional analyses using several different non-trans-spliced, schistosome mRNA 5'-UTRs.

The schistosome SL AUG initiation context lacks the purine at the -3 position (a U is present) that is typical of initiation contexts for non-trans-spliced schistosome mRNAs, other invertebrates, or the general vertebrate Kozak consensus (GCCA/GCCAUGG) (43, 45, 49, 50). Analysis of non-trans-spliced schistosome AUG contexts indicated that ~18% of the mRNAs have a U at this position. The most common nucleotide at -3 is an A present 54% of the time followed by a G at 21%. The control luciferase RNAs used in this study have a Kozak consensus (ACC) for -1 to -3. Substituting UGC (the SL AUG -1 to -3) for the Kozak context sequence in the control RNA led to only a small drop in translation efficiency (Fig. 4A). Furthermore, efficient use of the SL AUG context (UGC) is not dependent on the upstream spliced leader sequence as a 5'-UTR with 30 random 5'-nucleotides (N30 UGC AUG) substituted for the upstream SL sequence does not have a large effect on translation. However, substitution of an A for U at the -3 position leads to a significant increase in overall luciferase activity (Fig. 4). Thus, the relatively high frequency of A at -3 observed in cDNAs is reflective of a more optimal translation initiation context. Overall, these data indi-

**FIGURE 7. Effect of upstream and out-of-frame SL AUG on luciferase activity.** A, an illustration of the test RNA 5'-UTRs examined. The underlined sequence represents the spliced leader sequence, and the **bold ATG** is the initiator methionine for the main open reading frame. B, luciferase activity of a firefly reporter RNAs as a function of the native, trans-spliced upstream and out-of-frame SL AUG 5'-UTR. The illustrated reporter RNAs were electroporated along with a co-transfected *Renilla* RNA reporter into sporocysts. Luciferase activity 3 h after transfection was measured using a Dual-Luciferase assay, and the firefly activity was normalized to both *Renilla* luciferase and protein concentration. This is a representative experiment illustrating the mean  $\pm$  S.E. with three replicates for each RNA sample evaluated.



cated that aspects of the translation initiation context in schistosomes may be somewhat flexible and are complex.

**The addition of SL AUG Out-of-Frame**—The majority of trans-spliced flatworm mRNAs do not have the spliced leader AUG in-frame with the major open reading frame. As illustrated in Fig. 6, the spliced leader is typically trans-spliced within 30 nucleotides (54% of the mRNAs) of the initiator methionine of the recipient mRNA. No general patterns of encoded proteins were observed for these mRNAs as they appear to represent a diversity of protein types in a variety of cellular pathways. This is also true in the tapeworm *Echinococcus* (53).

The addition of an upstream and out-of-frame SL derived AUG to the 5'-UTR of the recipient mRNAs might require leaky scanning or reinitiation mechanisms for translation of these mRNAs. For some oncogenes and growth factors, an upstream and out-of-frame AUG can lead to low or regulated translation of these mRNAs (44). Addition of the native 5'-UTRs of two trans-spliced mRNAs with the SL AUG upstream and out-of-frame with the reporter AUG did not result in a reduction of overall luciferase activity. In fact, an increase in translation over both the wild-type and the SL AUG 5'-UTR was observed (Fig. 7B). The test mRNAs have 13 or 16 nucleotides between the SL AUG and the reporter AUG. This distance is within the range observed for the majority of trans-spliced mRNAs that have the SL AUG out-of-frame with the ORF. The addition of the SL AUG less than 30 nucleotides from the ORF AUG might be optimal to minimize deleterious effects of an upstream SL AUG. Thus, the flatworm translation machinery may have evolved to accommodate the upstream AUG within a specified distance. However, additional experiments are required to rigorously examine this, and it remains to be determined whether shorter or longer distances between the SL AUG and the ORF initiator methionine have differential effects on translation initiation.

The nematode SL also typically splices close to the recipient mRNA initiator methionine. We recently demonstrated that this spacing is optimal for translation *in vitro* (30). However, the nematode SL does not

contain a 3'-terminal AUG. We previously proposed that heterogeneous transcription initiation from trans-spliced genes may require that trans-splicing trim heterogeneous 5'-UTRs to generate uniform 5'-UTRs with an appropriate translation initiation context lacking additional upstream AUGs (30). The spacing between the SL and initiator methionine in both flatworms and nematodes is consistent with this hypothesis. However, although transcription of *Ascaris* trans-spliced genes appears heterogeneous, which is consistent with this hypothesis, it has been difficult to provide definitive data on the heterogeneous transcription sites (30).

**TMG Cap and Translation**—The addition of the TMG cap leads to a reduction in translation of test mRNAs in schistosomes even when this cap is present on the SL (Fig. 5). This is similar to what we observed for TMG-capped reporters in nematode embryos (30). Thus, under the experimental conditions evaluated, translation of TMG-capped RNAs in both flatworms and nematodes is in general less efficient than m<sup>7</sup>G-capped RNAs.

**Translation Context and RNA Stability**—In the current studies, we have not directly examined the mRNA half-lives for various test mRNAs. The large 50–100-fold differences in luciferase activity observed when the SL AUG is mutated to GCG are not likely explained by differences in RNA stability. Other RNAs examined with 1–2-nucleotide substitutions are also not likely to have significant differences in RNA stability. In addition, in separate studies in nematodes, the SL sequence and random sequence substitutions in the 5'-UTR did not have a significant effect on mRNA stability of the test RNAs (30).<sup>3</sup> Overall, although we think that the current data are reflective of primarily translation initiation efficiency, some differences in mRNA half-life may also contribute to the overall levels of translation we have measured.

<sup>3</sup> L. S. Cohen, M. Mikhli, M. Jankowska, J. Stepinski, E. Darzynkiewicz, and R. E. Davis, submitted

**Conclusions**—SL trans-splicing was also recently described in cestodes, a distinct flatworm group commonly known as tapeworms (13, 32). Analysis of a small set of trans-spliced mRNAs for two genera indicated that several of the trans-spliced mRNAs were likely to use the SL AUG based on the conserved N-terminal residues criterion applied here. A subsequent study also identified a high percentage of SL cDNA clones that were likely to use the SL AUG based on the SL AUG in-frame with an open reading frame of at least 99 nucleotides (53). Our bioinformatic and functional schistosome data provides additional support to suggest that the tapeworm SL AUG is also likely to serve as an initiator methionine for some mRNAs.

In summary, bioinformatic and transfection data demonstrated that the 3'-terminal AUG of the schistosome-spliced leader can be used as an initiator methionine. These findings have added a novel function to the list of functions for trans-splicing.

**Acknowledgments**—We thank Phil LoVerde and Ed Pearce for generously providing protocols and schistosome-infected hamsters and mice and an anonymous reviewer for helpful suggestions to the manuscript. Thanks to Fred Lewis and the Biomedical Research Institute for providing schistosome-infected mice and hamsters and for valuable training, assistance, and support. We thank Anthony Esposito, Melissa Katz, and Nathan Netsanet for contributions to the identification and analysis of trans-spliced schistosome mRNAs. Particular thanks to Phil Newmark for sharing the planarian spliced leader sequences prior to publication and Edward Darzynkiewicz for cap analogs.

## REFERENCES

- Hastings, K. E. (2005) *Trends Genet.* **21**, 240–247
- Nilsen, T. W. (2001) *Trends Genet.* **17**, 678–680
- Davis, R. E. (1996) *Parasitol. Today* **12**, 33–40
- Van Doren, K., and Hirsh, D. (1990) *Mol. Cell. Biol.* **10**, 1769–1772
- Maroney, P. A., Hannon, G. J., and Nilsen, T. W. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 709–713
- Liou, R. F., and Blumenthal, T. (1990) *Mol. Cell. Biol.* **10**, 1764–1768
- Thomas, J. D., Conrad, R. C., and Blumenthal, T. (1988) *Cell* **54**, 533–539
- Rajkovic, A., Davis, R. E., Simonsen, J. N., and Rottman, F. M. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8879–8883
- Stover, N. A., and Steele, R. E. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5693–5698
- Pouchkina-Stantcheva, N. N., and Tunnaclyffe, A. (2005) *Mol. Biol. Evol.* **22**, 1482–1489
- Ganot, P., Kallesoe, T., Reinhardt, R., Chourrout, D., and Thompson, E. M. (2004) *Mol. Cell. Biol.* **24**, 7795–7805
- Vandenberghe, A. E., Meedel, T. H., and Hastings, K. E. (2001) *Genes Dev.* **15**, 294–303
- Brehm, K., Jensen, K., and Frosch, M. (2000) *J. Biol. Chem.* **275**, 38311–38318
- Davis, R. E. (1997) *Mol. Biochem. Parasitol.* **87**, 29–48
- Davis, R. E., Singh, H., Botka, C., Hardwick, C., Ashraf el Meanawy, M., and Villanueva, J. (1994) *J. Biol. Chem.* **269**, 20026–20030
- Tessier, L. H., Keller, M., Chan, R. L., Fournier, R., Weil, J. H., and Imbault, P. (1991) *EMBO J.* **10**, 2621–2625
- Krause, M., and Hirsh, D. (1987) *Cell* **49**, 753–761
- Van der Ploeg, L., Liu, A., Michels, P., De Lange, T., Borst, P., Majumder, H., Weber, H., Veeneman, G., and Van Boom, J. (1982) *Nucleic Acids Res.* **10**, 3591–3604
- Boothroyd, J., and Cross, G. (1982) *Gene (Amst.)* **20**, 279–287
- Zayas, R. M., Bold, T. D., and Newmark, P. A. (2005) *Mol. Biol. Evol.* **22**, 2048–2054
- Blumenthal, T., and Gleason, K. S. (2003) *Nat. Rev. Genet.* **4**, 110–118
- Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M., and Kim, S. K. (2002) *Nature* **417**, 851–854
- Davis, R. E., and Hodgson, S. (1997) *Mol. Biochem. Parasitol.* **89**, 25–39
- Spieth, J., Brooke, G., Kuersten, S., Lea, K., and Blumenthal, T. (1993) *Cell* **73**, 521–532
- Muhich, M. L., and Boothroyd, J. C. (1988) *Mol. Cell. Biol.* **8**, 3837–3846
- Johnson, P. J., Kooter, J. M., and Borst, P. (1987) *Cell* **51**, 273–281
- Liang, X. H., Haritan, A., Uliel, S., and Michaeli, S. (2003) *Eukaryot. Cell* **2**, 830–840
- Huang, X. Y., and Hirsh, D. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8640–8644
- Zorio, D. A., Cheng, N. N., Blumenthal, T., and Spieth, J. (1994) *Nature* **372**, 270–272
- Lall, S., Friedman, C. C., Jankowska-Anyszka, M., Stepinski, J., Darzynkiewicz, E., and Davis, R. E. (2004) *J. Biol. Chem.* **279**, 45573–45585
- Hannon, G. J., Maroney, P. A., Ayers, D. G., Shambaugh, J. D., and Nilsen, T. W. (1990) *EMBO J.* **9**, 1915–1921
- Brehm, K., Hubert, K., Sciotto, E., Garate, T., and Frosch, M. (2002) *Mol. Biochem. Parasitol.* **122**, 105–110
- Davis, R. E., Hardwick, C., Tavernier, P., Hodgson, S., and Singh, H. (1995) *J. Biol. Chem.* **270**, 21813–21819
- Verjovski-Almeida, S., DeMarco, R., Martins, E. A., Guimaraes, P. E., Ojopi, E. P., Paquola, A. C., Piazza, J. P., Nishiyama, M. Y., Jr., Kitajima, J. P., Adamson, R. E., Ashton, P. D., Bonaldo, M. F., Coulson, P. S., Dillon, G. P., Farias, L. P., Gregorio, S. P., Ho, P. L., Leite, R. A., Malaquias, L. C., Marques, R. C., Miyasato, P. A., Nascimento, A. L., Ohlweiler, F. P., Reis, E. M., Ribeiro, M. A., Sa, R. G., Stukart, G. C., Soares, M. B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A. M., Wilson, R. A., Menck, C. F., Setubal, J. C., Leite, L. C., and Dias-Neto, E. (2003) *Nat. Genet.* **35**, 148–157
- Hu, W., Yan, Q., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R., Wang, Z. J., Rong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J., Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J., McManus, D. P., Xue, C. L., Feng, Z., Chen, Z., and Han, Z. G. (2003) *Nat. Genet.* **35**, 139–147
- Cohen, L. S., Mikhli, C., Friedman, C., Jankowska-Anyszka, M., Stepinski, J., Darzynkiewicz, E., and Davis, R. E. (2004) *RNA (N. Y.)* **10**, 1609–1624
- Lewis, F. A., Stirewalt, M. A., Souza, C. P., and Gazzinelli, G. (1986) *J. Parasitol.* **72**, 813–829
- Lewis, F. A. (1998) in *Current Protocols in Immunology* (Coligan, J. E., Kruisbeek, J. A., Margulies, D. E., Shevach, E. M., and Strober, W., eds) pp. 19.1.1–19.1.28, John Wiley & Sons, Inc., New York
- Kawanaka, M., Hayashi, S., and Ohtomo, H. (1983) *J. Parasitol.* **69**, 991–992
- Bixler, L. M., Lerner, J. P., Ivanchenko, M., McCormick, R. S., Barnes, D. W., and Bayne, C. J. (2001) *J. Parasitol.* **87**, 1167–1168
- Davis, R. E., Parra, A., LoVerde, P. T., Ribeiro, E., Glorioso, G., and Hodgson, S. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8687–8692
- Jones, L. J., Yue, S. T., Cheung, C. Y., and Singer, V. L. (1998) *Anal. Biochem.* **265**, 368–374
- Mankad, R. V., Gimelbrant, A. A., and McClintock, T. S. (1998) *Biol. Bull.* **195**, 251–254
- Kozak, M. (1991) *J. Cell Biol.* **115**, 887–903
- Cavener, D. G., and Ray, S. C. (1991) *Nucleic Acids Res.* **19**, 3185–3192
- Peri, S., and Pandey, A. (2001) *Trends Genet.* **17**, 685–687
- Pesole, G., Bernardi, G., and Saccone, C. (1999) *FEBS Lett.* **464**, 60–62
- Wang, X. Q., and Rothnagel, J. A. (2004) *Nucleic Acids Res.* **32**, 1382–1391
- Kozak, M. (1999) *Gene (Amst.)* **234**, 187–208
- Kozak, M. (2002) *Gene (Amst.)* **299**, 1–34
- 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
- Enjolras, N., Rodriguez, M. H., Plantier, J. L., Maurice, M., Attali, O., and Negrier, C. (1999) *Thromb. Haemostasis* **82**, 1264–1269
- Fernandez, C., Gregory, W. F., Loke, P., and Maizels, R. M. (2002) *Mol. Biochem. Parasitol.* **122**, 171–180



**Table 1**

Schistosome mRNAs using the SL AUG as an initiator methionine

Accession number <sup>1</sup> Identification <sup>4</sup>	Amino Acids to	Species <sup>2</sup>	Database <sup>3</sup>	Putative Next AUG <sup>5</sup>
U30265	S. mansoni	NR	Cyclophilin E	51
U30261	S. mansoni	NR	Peptidyl-prolyl cis-trans isomerase E	129
U30264	S. mansoni	NR	G protein-like (WD40 domain)	79
U30260	S. mansoni	NR	Meiotic recombination protein Rec14	76
AF014466	S. mansoni	NR	Glyoxalase I	26S proteasome, non-ATPase regulatory subunit 14 (rpn11)
AF006679	S. mansoni	NR	Hydroxyacyl glutathione hydrolase	64
AY812491	S. japonicum	NR	Beta 1,3-glucuronyltransferase I (GlcAT-I)	76
AY811984	S. japonicum	NR	26S proteasome, non-ATPase regulatory subunit 14 (rpn11)	147
AY809874	S. japonicum	NR	Lysophospholipase I	18
AY223270	S. japonicum	NR	acyl-protein thioesterase-1	68
AY815484	S. japonicum	NR	Transcription factor of the Forkhead/HNF3 family	88
superfamily I			Interleukin enhancer-binding factor ILF-2	68
AY223103	S. japonicum	NR	Cyclophilin J	217
AY810115	S. japonicum	NR	Peptidylprolyl isomerase-like protein 3	95
AY814909	S. japonicum	NR	Phosphatidylglycerophosphate synthase	217
AY811953	S. japonicum	NR	Nucleotide binding protein 2	122
AY815507	S. japonicum	NR	Mrp, ATPases involved in chromosome partitioning	72
AY814485	S. japonicum	NR	Ribonuclease Z (tRNase Z)	45
AY812204	S. japonicum	NR	Metal-dependent hydrolases of the beta-lactamase	116
AY915378	S. japonicum	NR	Protein BAT4 (HLA-B-associated transcript 4)	55
AY813031	S. japonicum	NR	Inorganic polyphosphate/ATP-NAD kinase	84
AY223327	S. japonicum	NR	Uracil DNA glycosylase superfamily	55
AY223068	S. japonicum	NR	FN3, Fibronectin type 3 domain	72
AY812760	S. japonicum	NR	Acyl CoA binding protein (binding domain 6)	135
AY816170	S. japonicum	NR	Complex 1 protein (LYR family)	22
AY808316	S. japonicum	NR	NADH-ubiquinone oxidoreductase complex I proteins	74
AY810672	S. japonicum	NR	Metal-dependent hydrolase of the TIM-barrel fold	44
AAW25300	S. japonicum	NR	ATPase, H <sup>+</sup> transporting, lysosomal accessory protein 2	238
AY223319	S. japonicum	NR	LSM1, snRNP Sm proteins	67
AY812762	S. japonicum	NR	Tudor domain 3 protein	None
AY812717	S. japonicum	NR	LSM10, U7 snRNP-specific Sm-like protein	74
			Steroid, 3-oxo-5-alpha-steroid 4-dehydrogenase	44
			Mitochondrial ribosomal protein L14	238
			Myloid beta precursor protein-binding family A-like	67
			X11 Phosphotyrosine-binding domain	None
			Metallophosphoesterase 1	74
			Angio-associated migratory cell protein	44
			Nucleoside diphosphokinase	238
			xylosylprotein beta 1,4-galactosyltransferase	67
			SNARE associated protein snapin	None

AY809604	S. japonicum	NR	Nuclear mitotic apparatus protein 1	105
AY815282	S. japonicum	NR	Beta-parvin	71
AY809388	S. japonicum	NR	TRAP/Mediator complex component TRAP25	37
AY812338	S. japonicum	NR	Replication protein A3	17
AY808464	S. japonicum	NR	Unknown, Hypothetical Protein	65
Sm12724	S. mansoni	GeneDB	Zinc Finger CCCH-type with G patch domain	None
Sm01547	S. mansoni	GeneDB	FK506-binding protein 4	66
			HSP binding immunophilin	
Sm13309	S. mansoni	GeneDB	Hypothetical proteins	46
Sm13026	S. mansoni	GeneDB	Solute carrier family 25	28
			Mitochondrial deoxynucleotide carrier	
Sm03815	S. mansoni	GeneDB	Acid, beta-galactosidase I	36
Sm29759	S. mansoni	GeneDB	StAR-binding protein,-helix coiled-coil rod homologue	136
AI975462	S. mansoni	Est	mitochondrial methionyl-tRNA synthetase	120
BU800427	S. japonicum	Est	asparaginyl-tRNA synthetase	43
AW191483	S. japonicum	Est	Unknown, Hypothetical Protein	none
BU723836	S. japonicum	Est	Unknown, Hypothetical Protein	72

<sup>1</sup>Accession Number = Database Accession Number

<sup>2</sup>Species = Corresponding Schistosoma species

<sup>3</sup>Database = NR is the non-redundant database at the NCBI; Est is the dbEST at the NCBI; GeneDB is the Schistosoma database at the GeneDB (<http://www.genedb.org/>)

<sup>4</sup>Putative Identification = Identification of encoded protein based on similarity with database proteins

<sup>5</sup>Amino Acids to Next AUG = Number of amino acids to next in frame AUG in the deduced protein