Yeast-like mRNA Capping Apparatus in *Giardia lamblia*

Received for publication, October 25, 2004, and in revised form, November 18, 2004
Published, JBC Papers in Press, November 19, 2004, DOI 10.1074/jbc.M412063200

Stéphane Hausmann‡, Melissa A. Altura§, Matthew Witmer§, Steven M. Singer§, Heidi G. Elendorf§, and Stewart Shuman¶

From the ¶Molecular Biology Program, Sloan-Kettering Institute, New York, New York 10021 and the §Department of Biology, Georgetown University, Washington, D. C. 20057

A scheme of eukaryotic phylogeny has been suggested based on the structure and physical linkage of the RNA triphosphatase and RNA guanylyltransferase enzymes that catalyze mRNA cap formation. Here we show that the unicellular pathogen *Giardia lamblia* encodes an mRNA capping apparatus consisting of separate triphosphatase and guanylyltransferase components, which we characterize biochemically. We also show that native *Giardia* mRNA have blocked 5'-ends and that 7-methylguanosine caps promote translation of transcripts of fungi, microsporidia, and protozoa such as *Plasmodium* and *Trypanosoma*. The tunnel enzymes adopt a unique active-site fold and are structurally and mechanistically unrelated to the cysteine-phosphatase-type RNA triphosphatases found in metazoans and plants, which comprise part of a bifunctional triphosphatase-guanylyltransferase fusion protein. All available evidence now points to the separate tunnel-type triphosphatase and guanylyltransferase as the aboriginal state of the capping apparatus. We identify a putative tunnel-type triphosphatase and a separate guanylyltransferase encoded by the red alga *Cyanidioschyzon merolae*. These findings place fungi, protozoa, and red algae in a common lineage distinct from that of metazoans and plants.

The m^7^GpppN cap is a defining feature of eukaryotic mRNA that is formed via three enzymatic reactions: (i) the 5'-triphosphate end of the pre-mRNA is hydrolyzed to a diphosphate by RNA triphosphatase; (ii) the diphosphate RNA end is capped with GMP by RNA guanylyltransferase; and (iii) the GpppN cap is methylated by RNA (guanine-N7) methyltransferase (1). Capping enzymes are a good focal point for considering eukaryotic evolution, because the cap structure is thought to be ubiquitous in eukaryotic organisms but absent from the bacterial and archaeal domains of life. Thus, any differences in the capping apparatus between eukaryal taxa would reflect events that post-date the emergence of ancestral nucleated cells. Previously, we proposed a scheme of eukaryotic phylogeny based on two features of the mRNA capping apparatus: the structure and mechanism of the triphosphatase component (metal-dependent fungal-type versus metal-independent cysteine-phosphatase type) and whether the triphosphatase is covalently fused to the guanylyltransferase component (2, 3).

Metazoans and plants have a two-component capping system consisting of a bifunctional triphosphatase-guanylyltransferase polypeptide and a separate methyltransferase polypeptide, whereas fungi, microsporidia, and protozoa (e.g. *Plasmodium falciparum* and *Trypanosoma brucei*) have a three-component system consisting of separate triphosphatase, guanylyltransferase, and methyltransferase polypeptides (2, 4, 5). The primary structures and biochemical mechanisms of the fungal and mammalian guanylyltransferases and cap methyltransferases are conserved. However, the atomic structures and catalytic mechanisms of the fungal and mammalian RNA triphosphatases are completely different (6–8).

Metazoan and plant RNA triphosphatases belong to the cysteine-phosphatase enzyme superfamily, which is defined by the conserved phosphate-binding loop motif H(S/T)XXH(R/K) (7). Mammalian RNA triphosphatase catalyzes a two-step phosphoryl transfer reaction in which the conserved cysteine of the signature motif attacks the γ-phosphorus of triphosphate-terminated RNA to form a covalent protein-cysteinyl-S-phosphate intermediate and expel the diphosphate RNA product (7, 8). The covalent phosphoenzyme intermediate is hydrolyzed to liberate inorganic phosphate. The reaction does not require a divalent cation cofactor.

In contrast, the RNA triphosphatases of fungal species such as *Saccharomyces cerevisiae*, *Candida albicans*, and *Schizosaccharomyces pombe* are strictly dependent on a divalent cation. The fungal enzymes belong to a family of metal-dependent phosphohydrolases that embraces the triphosphatase components of the *P. falciparum*, *T. brucei*, *Encephalitozoon cuniculi*, *Chlorella* virus, poxvirus, and baculovirus mRNA capping systems (2, 4, 5, 9–13). The signature biochemical property of this enzyme family is the ability to hydrolyze nucleoside triphosphates to nucleoside diphosphates and inorganic phosphate in the presence of manganese or cobalt. The defining structural features of the metal-dependent RNA triphosphatases are two glutamate-containing motifs that are required for catalysis by every family member and which comprise the metal-binding site. The crystal structure of the *S. cerevisiae* RNA triphosphatase Cet1 revealed a novel tertiary structure in which the active site is situated within a topologically closed hydrophilic tunnel composed of eight antiparallel β strands (6).

The radically different flavors of RNA triphosphatase exemplify an unusual situation in nature, whereby two different structural solutions to an enzyme chemistry problem (γ-phosphate hydrolysis, a fairly pedestrian reaction) have evolved within a single biological pathway. When and how did the RNA triphosphatase component of the capping apparatus diverge during eukaryotic evolution and will the differences in the organization of the capping apparatus continue to provide a useful guide to phylogenetic relationships among taxa as more genomic information becomes available? We are pursuing these questions by identifying the cap-forming enzymes of uni-
cellular eukaryotes and evaluating their structural and mechanistic relatedness to exemplary fungal and mammalian enzymes.

Here we focus on the amitochondriate protozoan parasite *Giardia lamblia*. In addition to being a significant enteric pathogen of humans, *Giardia* is posited to occupy a deeply branching position in eukaryal phylogeny. Analysis of the sequence of the 12-Mb *Giardia* genome is providing important insights to the early origins of RNA processing mechanisms, such as spliceosomal pre-mRNA splicing, which are regarded as uniquely eukaryal (14). Although it has been a matter of some debate whether *Giardia* mRNAs even have a 5′ cap structure (15, 16), we show here that *Giardia* does possess the enzymatic machinery for cap synthesis, modifies the 5′-ends of its mRNAs, and exploits the m7G cap for enhanced translation of a reporter mRNA in vivo. Biochemical characterization of the separately encoded triphosphatase and guanylyltransferase of *Giardia* shows that they resemble the capping enzymes of fungi and protozoa.

Mining of other newly available genome databases permitted the identification of candidate cap-forming enzymes in the unicellular red alga *C. merolae* (17) and the water mold *Phytophthora ramorum*. The red alga separately encodes a tunnel family RNA triphosphatase and a guanylyltransferase, whereas *P. ramorum* appears to encode a bifunctional fusion enzyme consisting of an N-terminal tunnel-type-triphasphatase domain and a C-terminal guanylyltransferase domain. We integrate these findings into an expanded cap-centric scheme of eukaryotic phylogeny.

**Experimental Procedures**

*Giardia Cultivation*—Trophozoites of the *G. lamblia* isolate WB, clone WB/1267 (ATCC number 50582), were cultured essentially as described previously (33) with a slight modification. Cells were maintained anaerobically in glass culture tubes at 37 °C in Keister’s modified TYI-S-33 medium supplemented with antibiotics in which the sodium phosphate buffer was replaced by a 25 mM sodium bicarbonate buffer. This change had no detectable effect on parasite growth.

**Recombinant G. lamblia RNA Triphosphatase**—The open reading frame encoding GICet1 was PCR-amplified from genomic DNA and cloned into a T7 RNA polymerase-based bacterial expression vector pET16b-Nco (which contains an NcoI site instead of a NdeI site) so as to fuse the GICet1 protein to an N-terminal 14-amino acid leader peptide (MGHHHHHHHHHHHHHSA). pET-His10-GICet1 was transformed into *E. coli* BL21-CodonPlus(DE3). A 500-ml culture amplified from a single transformant was grown at 37 °C in LB medium containing 50 μg/ml kanamycin and 50 μg/ml chloramphenicol until the A600 reached 0.6. The cultures were adjusted to 2% ethanol and 0.2 mM isopropyl β-D-thiogalactopyranoside and then incubated at 17 °C for 20 h with constant shaking. Cells were harvested by centrifugation, and the pellets were stored at −80 °C. All subsequent procedures were performed at 4 °C. Thawed bacteria were resuspended in 25 ml of buffer A (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10% glycerol). Cell lysis was achieved by the addition of lysozyme and Triton X-100 to final concentrations of 100 μg/ml and 0.1%, respectively. The lysates were sonicated to reduce viscosity, and insoluble material was removed by centrifugation. The soluble extracts were applied to 1-ml columns of nickel-nitrotriacetic acid-agarose resin (Qiagen) that had been equilibrated with buffer A containing 0.1% Triton X-100. The columns were washed with 10 ml of the same buffer and then eluted stepwise with 2.5-ml aliquots of buffer A containing 50, 100, 250, and 500 mM imidazole. The polypeptide compositions of the column fractions were monitored by SDS-PAGE. The 54-kDa recombinant His6-Smt3-GICet1 polypeptide was recovered predominantly in the 250 and 500 mM imidazole fractions. The 250 mM imidazole eluates were dialyzed against buffer A containing 2 mM DTT and 0.01% Triton X-100 and then stored at −80 °C. The protein concentrations were determined by SDS-PAGE analysis of serial dilutions of the GICet1 preparation and a BSA standard. The gels were stained with Coomassie Blue, and the staining intensities of the His6-Smt3-GICet1 and BSA polypeptides were quantified with a Fuji film FLA-5000 digital imaging and analysis system. GICet1 concentrations were calculated by interpolation to the BSA standard curve.

**Triphosphatase Assay**—Reaction mixtures (20 μl) containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 2 mM MnCl2, 1 mM [γ-32P]ATP, and GICet1 as specified were incubated for 20 min at 37 °C. An aliquot (2 μl) of the mixture was applied to a polyethyleneimine-cellulose TLC plate, which was developed using 0.5 M LiCl, 1 M formic acid. The radiolabeled material was visualized by autoradiography, and 32P incorporation was quantified by scanning the TLC plate with a Fujix BAS2500 imager. Alternatively, reaction mixtures (20 μl) containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 2 mM MnCl2, 1 mM unlabeled NTP, and GICet1 as specified were incubated for 20 min at 37 °C. The reactions were quenched by adding 1 ml of malachite green reagent (Biomol Research Laboratories, Plymouth Meeting, PA). Phosphate release was determined by measuring A600 and interpolating the value to a phosphate standard curve.

**Recombinant G. lamblia Guanylyltransferase**—The open reading frame encoding GICet1 was PCR-amplified from genomic DNA and cloned into a T7 RNA polymerase-based bacterial expression vector pET16b-Nco (which contains an NcoI site instead of a NdeI site) so as to fuse the GICet1 protein to an NH2-terminal 14-amino acid leader peptide (MGHHHHHHHHHHHHHSA). pET-His10-GICet1 was transformed into *E. coli* BL21-CodonPlus(DE3). A 500-ml culture amplified from a single transformant was grown at 37 °C in LB medium containing 50 μg/ml kanamycin and 50 μg/ml chloramphenicol until the A600 reached 0.6. The culture was adjusted to 2% ethanol and 0.2 mM isopropyl β-D-thiogalactopyranoside and then incubated at 17 °C for 20 h. The recombinant His6-GICet1 polypeptide was recovered predominantly in the 250 and 500 mM imidazole fractions. The 250 mM imidazole eluates were dialyzed against buffer A containing 2 mM DTT and 0.01% Triton X-100 and then stored at −80 °C. Protein concentrations were determined using the Bio-Rad dye reagent with bovine serum albumin as the standard.

---

**Table I**

| Gene       | Primer name          | Sequence | PCR     | Size expected bp |
|------------|----------------------|----------|---------|-----------------|
| α2 β-Tubulin | TUB800 | GATGGGCGCGTTACGACCAGGCGGG | Primary | 800 |
| Glutamate   | GDI800 | CTCCTCGTTGACCCCGTCCTTGTG | Primary | 800 |
| dehydrogenase | GDI600 | CGGCGCTGATGAAAGACCGCCCGAC | Nested | 600 |
| NOR-sense   | 5H3NOR | GCGGGATCCCTACTCGGAAGACTCTAGGA | Primary | 400 |
| NOR-antisense | 32FS | CCAAAAGGCCCATCGTGTCAGCA | Nested | 300 |
| NOR-antisense | 32R4 | CCGGAGAGCCTCGCCACAGTGCGTT | Nested | 800 |
| NOR-antisense | 32R6 | ATAGGGATACATTACAAACACACTAG | Nested | 250 |

---

T. Abbreviations used are: DTT, dithiothreitol; BSA, bovine serum albumin; RACE, rapid amplification of cDNA ends; CIP, calf intestinal phosphatase; TAP, tobacco acid pyrophosphatase; aa, amino acid; NOR, NADH oxidoreductase; TMG, 2,2,7-trimethylguanosine.
**Guanylyltransferase Assay**—Guanylyltransferase activity was assayed by the formation of the covalent enzyme-GMP intermediate. Reaction mixtures (20 μl) containing 50 mM Tris-HCl, pH 8.0, 5 mM DTT, 5 mM MgCl₂, and [³²P]GTP and enzyme as specified were incubated at 37 °C. The reaction was halted by addition of SDS to 1% final concentration. The samples were analyzed by SDS-PAGE. Transfer of labeled GMP to the GlCeg1 protein was visualized by autoradiography of the dried gel and quantified by scanning the gel with a Fujix BAS2500 imager.

**Glycerol Gradient Sedimentation**—An aliquot (40 μl) of the nickel-agarose preparation of GlCeg1 was mixed with catalase (40 μg), BSA (40 μg), and cytochrome c (40 μg). The mixture was applied to a 4.8-ml 15–30% glycerol gradient containing 50 mM Tris-HCl, pH 8.0, 0.2 M NaCl, 1 mM EDTA, 2 mM DTT, 0.05% Triton X-100. The gradient was centrifuged for 14 h at 4 °C in a Beckman SW50 rotor at 50,000 rpm. Fractions (0.19 ml) were collected from the bottom of the tube. The polypeptide compositions of the gradient fractions were analyzed by SDS-PAGE.

**Cap-dependent 5'-RACE Assay**—The state of the 5'-end of specific *Giardia* transcripts was determined by using the GeneRacer™ kit from Invitrogen, Inc. (Carlsbad, CA). Total RNA was isolated from *Giardia* using RNA-STAT 60 according to the manufacturer's protocol (Tel-Test, Inc., Friendswood, TX) and then treated with DNase I (Promega, Inc., Madison, WI) to remove all remaining DNA. Four separate reactions were established in parallel for each transcript to be assayed: 1) untreated; 2) treated with calf intestinal phosphatase (CIP); 3) treated with tobacco acid pyrophosphatase (TAP); and 4) treated with CIP and TAP. To remove all unmodified phosphates from the 5'-end of transcripts, RNA in the CIP and CIP + TAP tubes was initially dephosphorylated (10 units of CIP per 1 μg of total RNA) at 50 °C for 60 min. Untreated and TAP-only tubes were processed in parallel without the addition of CIP. To remove modified phosphate structures from the 5'-end of RNA transcripts and leave a single 5'-phosphate behind, CIP was then removed by phenol:chloroform extraction. RNA in the TAP and CIP + TAP samples was subsequently treated with TAP (0.5 unit/1 μg of total RNA) at 37 °C for 60 min. Untreated and CIP-treated samples were processed in parallel without the addition of enzyme.

To provide an anchor for subsequent PCR reactions, 0.25 μg of the GeneRacer™ 5'-Nested Primer (5'-GGACACTGACATGGACTGAAGTGTA-3') and another gene-specific primer. The PCR products were analyzed by agarose gel electrophoresis and visualized by staining the gel with ethidium bromide.

**Luciferase Expression Assays**—Plasmids pT7Luc and pT7LucA50 (34) were kindly provided by Dr. D. Gallie. Transcription from either plasmid yields a luciferase transcript with an 8-nucleotide 5'-leader. Transcription from pT7LucA50 results in a 50-nucleotide 3'-poly(A) tail, whereas transcripts produced from pT7Luc have no poly(A) tail. The plasmids were linearized by digestion with BamHI and DraI, respectively, and mRNA was synthesized by using AmpliScribe and AmpliCap T7 in vitro transcription kits (Ambion Inc., Austin, TX). RNA was

**Fig. 1. Capping enzymes from *G. lamblia* and *C. merolae*.** The complete predicted amino acid sequences of the putative triphosphatase and guanylyltransferase enzymes of *G. lamblia* and *C. merolae* are shown. The two signature metal-binding motifs of the tunnel family of RNA triphosphatases and the six signature nucleotidyl transferase motifs of the RNA guanylyltransferases are highlighted in shaded boxes.
recovered after DNase I digestion, phenol extraction, and ammonium acetate precipitation. RNA was quantified by UV absorbance and RNA integrity was judged by denaturing agarose gel electrophoresis. For transfection experiments, 10 million trophozoites were resuspended in 800 μl of cytomix with 50 units of RNasin (Promega) and 10 μg of yeast tRNA and then kept on ice (25). 20 μg of luciferase transcripts were added and cells were electroporated at 800 V, 1260 μF, and 500 V in a 0.4-cm cuvette using a BTX ECM600 electroporator (Harvard Apparatus, Holliston, MA). Cells were kept on ice an additional 10 min and then placed in 15 ml of fresh medium for the indicated times. Cells were harvested, and luciferase activity was determined as described (35).

RESULTS

Identification of a Candidate G. lamblia Triphosphatase GlCet1—We identified a candidate Giardia RNA triphosphatase gene (Genbank accession number AACB01000007) on the third iteration of a psi-blast search of the NCBI database with the T. brucei RNA triphosphatase TbCet1. The amino acid sequences of the G. lamblia triphosphatase (GlCet1) and the C. merolae triphosphatase (CmCet1) are aligned to the sequences of the catalytic domains of RNA triphosphatases of S. cerevisiae (ScCet1 and ScCth1), C. albicans (CaCet1), S. pombe (SpCet1), Giberella zeae (GzCet1), Neoropsora crassa (NCet1), F. palficarum (PFPrt1), C. viridis (cvRtp1), E. cuniculi (EcCet1), L. major (LmCet1), and T. brucei (TbCet1). The β strands that form the triphosphate tunnel are denoted above the sequence. Gaps in the alignment are indicated by dashes. Inter-strand inserts in PfPrt1, LmCet1, and CmCet1 are omitted from the alignment and denoted by dots. Peptide segments with the highest degree of conservation in all seven proteins are highlighted by the shaded boxes. Essential hydrophilic amino that comprise the active site within the Cet1 tunnel are denoted by dots.
Metal-dependent Triphosphatase Activity of GlCet1—The putative triphosphatase was produced in bacteria as a His<sub>10</sub>-Smt3 fusion protein and partially purified from a soluble bacterial extract by adsorption to nickel-agarose and elution with 250 mM imidazole. The nickel-agarose preparation was enriched with the enzyme. Recombinant wild-type GlCet1 catalyzed the release of ATP from [γ<sup>32</sup>P]ATP in the presence of manganese; the extent of ATP hydrolysis was proportional to enzyme concentration (Fig. 3B). The ATPase activity of the E58A mutant was <1% of the activity of wild-type enzyme. We conclude that the observed phosphohydrolase activity is intrinsic to GlCet1.

GlCet1 catalyzed no detectable ATP hydrolysis in the absence of a divalent cation (Fig. 4A). ATP activity was optimal at 2.5 mM MnCl<sub>2</sub> and declined slightly at 5 mM MnCl<sub>2</sub>. Cobalt also satisfied the metal requirement; activity was optimal at 2.5–5 mM CoCl<sub>2</sub>. No activity was observed when the manganese and cobalt concentrations were below the level of input ATP. Magnesium supported only feeble ATP hydrolysis at up to 5 mM concentration (Fig. 4A). Specificity for NTP hydrolysis in the presence of manganese and cobalt is characteristic (indeed, virtually diagnostic) of the fungal-type RNA triphosphatase family.

The rate of release of <sup>32</sup>P from [γ<sup>32</sup>P]ATP was identical to the rate of conversion of [α<sup>32</sup>P]ATP to [α<sup>32</sup>P]ADP in a parallel reaction mixture containing the same concentration of GlCet1 (Fig. 4C), indicating that GlCet1 catalyzes the hydrolysis of ATP to ADP and P<sub>i</sub>. From the initial rate, we estimate a turnover number of 3.4 s<sup>−1</sup>. We detected only low levels of [α<sup>32</sup>P]AMP formation during the reaction. NTP specificity was examined by colorimetric assay of the release of P<sub>i</sub> from unlabeled ATP, GTP, CTP, or UTP. GlCet1 displayed nearly equivalent specific activity with each of the four ribonucleoside triphosphates (Fig. 4B).

Identification and Characterization of Giardia Guanylyltransferase GlCeg1—We identified a candidate mRNA guanylyltransferase gene (GenBank<sup>TM</sup> accession number AACB01000002) encoding a 511-aa polypeptide (GenBank<sup>TM</sup> accession number EAA42847) with extensive similarity to fungal, protozoal, metazoan, and virus-encoded RNA guanylyltransferases (Fig. 1). We named this gene product GlCeg1 (gapping enzyme guanylyltransferase). GlCeg1 contains the six conserved peptide motifs (I, III, IIIa, IV, V, and VI) that comprise the active site of GTP binding and nucleotidyl transfer (Fig. 1). Recombinant His<sub>10</sub>-GlCeg1 was produced in bacteria and isolated from a soluble bacterial extract by nickel-agarose chromatography. The 59-kDa Giardia polypeptide adsorbed to the affinity resin and was eluted with 50–250 mM imidazole (Fig. 5A).

All known cellular guanylyltransferases employ a ping-pong reaction mechanism of nucleotidyl transfer through a covalent enzyme-(lysyl-N)-GMP intermediate (1). Hence, guanylyltransferase activity can be detected with high sensitivity and specificity by label transfer from [α<sup>32</sup>P]GTP to the enzyme. To assay guanylyltransferase activity of the expressed GlCeg1 protein, we incubated the initial soluble bacterial extract and the nickel-agarose column fractions with [α<sup>32</sup>P]GTP and a divalent cation. This resulted in the formation of an SDS-stable ~60-kDa nucleotidyl-protein adduct (Fig. 5A). The guanylyltransferase activity eluted in parallel with the GlCeg1 protein.

Formation of the covalent intermediate was optimal at pH 8.0 to 9.0 in Tris-HCl buffer and was sharply reduced at pH ~6.5 in Tris acetate buffer (Fig. 6A). Enzyme-GMP formation
shown in the lower panel stained gel is shown in the eluate fractions were analyzed by SDS-PAGE. The Coomassie Blue-H9262 quots (20 nickel-agarose chromatography. The polypeptide compositions of ali-
Ceg1.

flow-through (optima, which were 0.3–5 mM MnCl2 and 5 mM MgCl2 (Fig. 6 Calcium was unable to support guanylylation activity at up to 5 mM concentration. The extent of enzyme guanylylation achieved at 0.09 μM GTP. The nucleotidyltransferase reaction was specific for GTP, insofar as we could detect no formation of a protein-NMP complex during reaction of GlCeg1 with [α-32P]GTP (data not shown).

The forward guanylylation reaction of capping enzymes is in equilibrium with the reverse pyrophosphorylase reaction, in which PPi attacks the enzyme-GMP complex to regenerate GTP. We observed that inclusion of low concentrations of PPi in the guanylyltransferase reaction mixture exerted a stimulatory effect on the extent of label transfer from [α-32P]GTP to the GlCeg1 protein. A 4-fold stimulation was evident at the optimal concentration of 10 μM PPi (Fig. 6D). We ascribe this stimulation to pyrophosphorylase of the population of enzyme-GMP present in the recombinant protein preparation, which thereby regenerates additional apoenzyme to react with the radiolabeled GTP substrate. From the data in Fig. 6D, we estimated that 48% of the enzyme molecules were labeled with [32P]GTP in the presence of 10 μM PPi.

The native size of GlCeg1 was gauged by sedimentation through a 15–30% glycerol gradient. Marker proteins catalase (248 kDa), BSA (66 kDa), and cytochrome c (13 kDa) were included as internal standards. Guanylyltransferase activity sedimented as a single discrete peak between BSA and cytochrome c (Fig. 5B); the activity profile coincided with the abundance of the GlCeg1 polypeptide (data not shown). An S value of 4.1 was determined for GlCeg1 by interpolation to the internal standard curve. We surmise that GlCeg1 is a monomer in solution.

5’ Modification of Endogenous Giardia Transcripts—The existence of capping machinery in Giardia prompted us to query by 5’-RACE whether native Giardia transcripts have blocked 5’-ends. We examined sense transcripts that code for tubulin, glutamate dehydrogenase, and NADH oxidoreductase (NOR), and an antisense transcript for NOR that is one of many such sterile transcripts present in Giardia (36). 5’-RACE experiments were designed to distinguish transcripts that contain a blocked 5’-end (presumed to be a cap) from transcripts that have unblocked 5’-monophosphate, di- or triphosphate termini (Fig. 7). Only RNAs having 5’-monophosphate termini can be ligated to the synthetic oligonucleotide that provides an anchor for reverse transcription-PCR amplification. The experimental logic was as follows: 1) if the transcript has a 5’-monophosphate end, then a PCR product should appear in the RACE reactions programmed by both the untreated and TAP-treated RNA samples but not in the reactions programmed by CIP or CIP + TAP-treated RNA samples; 2) if the transcript has an unblocked 5’-triphosphate or diphosphate end, then a PCR product should appear in the RACE reactions programmed by TAP-treated RNA but not in reactions containing the CIP-treated, CIP + TAP-treated, or untreated RNAs; 3) if the transcript has a blocking 5’ cap structure, then a PCR product will appear in RACE reactions programmed by both the TAP-treated and CIP + TAP-treated RNAs but not in reactions containing the untreated or CIP-treated RNAs. As shown in Fig. 7, we consistently observed PCR products for all of the sense transcripts and the antisense-NOR transcript in the TAP and CIP + TAP treated samples, but not in the CIP-treated or untreated samples. These results indicate that endogenous Giardia RNAs have blocked 5’-ends, consistent with the presence of a cap.

Cap-dependent Translation of mRNA Transfected into Giardia in Vivo—To evaluate whether m7G caps were utilized by Giardia to promote translation, we synthesized luciferase transcripts in vitro with and without 5’-m7G caps and 3’-poly(A) tails. We found, in agreement with the recent report from Li and Wang (25), that luciferase activity was detected...
after transfection of capped mRNA, but not after transfection of uncapped mRNA (Fig. 8). Luciferase production was evident within 1 h post-transfection and activity persisted for at least 20 h. The presence of a 50-nucleotide poly(A) tail was also required for expression of the reporter mRNA. We conclude that caps and poly(A) tails can be utilized by a putative 397-aa metal-dependent RNA triphosphatase 
PcmCet1 (Fig. 1). CmCet1 contains the two signature glutamate-containing metal-binding motifs and putative homologs of the β strands that comprise the triphosphate tunnel of yeast Cap1, including the conserved catalytic side chains (Fig. 2). A psi-blast search using the CmCet1 sequence itself readily identify all fungal metal-dependent RNA triphosphatases as well as viral and protozoal enzyme, which we named CmCet1 (Fig. 1). CmCet1 contains the six nucleotidyltransferase motifs that comprise the RNA 5’-ends, and gene-specific transcripts were amplified by reverse transcription-PCR. For less abundant transcripts (glutamate dehydrogenase (Gdh), NOR sense, and NOR antisense), one round of nested PCR was performed. The expected outcomes of treating a 5’-triphosphate-terminated RNA (GpppRNA), a 5’-triphosphate-terminated RNA (pRNA), and a 5’-monophosphate-terminated RNA (pRNA) with CIP alone, TAP alone, or CIP then TAP are depicted at right. The ligatable pRNA species that can give rise to a 5’-RACE product are highlighted in shaded boxes.

Fig. 6. Characterization of the Giardia guanylyltransferase. A, pH dependence. Reaction mixtures (20 μl) containing either 50 mM Tris-acetate, pH 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, or 7.0, or 50 mM Tris-HCl, pH 7.0, 7.5, 8.0, 8.5, 9.0 or 9.5, 5 mM DTT, 5 mM MgCl2, 5 μM [α-32P]GTP, and 1 μg of GlCeg1 were incubated for 15 s at 37 °C. E-GMP formation is plotted as a function of pH. B, divalent cation requirement. Reaction mixtures (20 μl) containing 50 mM Tris-HCl, pH 8.0, 5 mM DTT, 5 mM MgCl2, 1 μg of GlCeg1, and [α-32P]GTP, 1 μg GlCeg1, and CaCl2, MgCl2, or MnCl2 as specified were incubated for 15 s at 37 °C. E-GMP formation is plotted as a function of divalent cation concentration. C, GTP dependence. Reaction mixtures (20 μl) containing 50 mM Tris-HCl, pH 8.0, 5 mM DTT, 5 mM MgCl2, 1 μg of GlCeg1, and [α-32P]GTP as specified were incubated for 10 min at 37 °C. E-GMP formation is plotted as a function of GTP concentration. D, effect of pyrophosphate. Reaction mixtures (20 μl) containing 50 mM Tris-HCl, pH 8.0, 5 mM MgCl2, 5 mM DTT, 5 μM [α-32P]GTP, 1 μg GlCeg1, and PPi, as specified were incubated for 10 min at 37 °C. E-GMP formation is plotted as a function of PPi concentration.

Fig. 7. 5’-RACE analysis of Giardia transcripts. Trophozoite RNA treated with CIP + TAP (lanes 1), TAP alone (lanes 2), CIP alone (lanes 3) or an untreated control (lanes 4) was used for 5’-RACE as described under “Experimental Procedures.” An anchor primer was ligated to the RNA 5’-ends, and gene-specific transcripts were amplified by reverse transcription-PCR. For less abundant transcripts (glutamate dehydrogenase (Gdh), NOR sense, and NOR antisense), one round of nested PCR was performed. The expected outcomes of treating a 5’-capped RNA (GpppRNA), a 5’-triphosphate-terminated RNA (pRNA), and a 5’-monophosphate-terminated RNA (pRNA) with CIP alone, TAP alone, or CIP then TAP are depicted at right. The ligatable pRNA species that can give rise to a 5’-RACE product are highlighted in shaded boxes.

Fig. 8. Cap-dependent translation in Giardia. Giardia trophozoites were electroporated with luciferase RNAs synthesized with and without 5’-caps and 3’-poly(A) tails as indicated. Parasites were transfected with pTUB-luc plasmid DNA (36) as a positive control for luciferase reporter expression (pDNA) or mock transfected with no RNA as a negative control. Luciferase activity was determined at 1, 2, 4, 8, and 20 h post-transfection as indicated. Light units were normalized to 20 μg of RNA per transfection. Values are averages of two luciferase readings from single transfections and data are representative of three independent experiments.

whether plants, green algae, and red algae are monophyletic with respect to the organization of their mRNA capping systems.

Fig. 8. Cap-dependent translation in Giardia. Giardia trophozoites were electroporated with luciferase RNAs synthesized with and without 5’-caps and 3’-poly(A) tails as indicated. Parasites were transfected with pTUB-luc plasmid DNA (36) as a positive control for luciferase reporter expression (pDNA) or mock transfected with no RNA as a negative control. Luciferase activity was determined at 1, 2, 4, 8, and 20 h post-transfection as indicated. Light units were normalized to 20 μg of RNA per transfection. Values are averages of two luciferase readings from single transfections and data are representative of three independent experiments.
that comprise the active site of prototypical capping enzymes (Fig. 1). We also identified a putative 373-aa cap guanine-N7 methyltransferase encoded by C. merolae locus CMP091C (data not shown). We surmise that the red alga has a yeast-like three-component mRNA capping apparatus that includes a tunnel family RNA triphosphatase. The capping apparatus of this unicellular red alga is substantially diverged from that of plants.

**DISCUSSION**

Capping in Giardia—An initial report that bulk Giardia poly(A)-containing RNAs do not have a blocking cap structure at their 5'-ends (15) raised the prospect that either: (i) the mRNA cap did not exist in the last common ancestor of all present eukarya, but arose subsequent to the branching of Giardia, or (ii) the cap did exist in the ancestral eukaryote but was lost in the lineage that gave rise to Giardia. It was hypothesized that the mechanism of translation initiation in Giardia was fundamentally different from that of other eukarya in being cap-independent (15). However, other analyses of a specific Giardia transcript highlighted two distinct 5' starts, one of which had either a 5'-triphosphate or blocked 5'-end, whereas the other had a 5'-monophosphate. (16). Our analysis of four additional transcripts, including one antisense transcript, indicates that all four have blocked 5'-ends consistent with cap-like modifications of the primary transcripts. In the absence of direct analysis of the cap structure of specific Giardia mRNAs, we regard the demonstration that Giardia transcripts have blocked 5'-ends and that the parasite encodes an mRNA capping apparatus as reasonable evidence that this primitive eukaryote does indeed cap its mRNAs. Our biochemical characterization of the recombinant Giardia triphosphatase and guanyltransferase proteins highlights their similarities to the well characterized fungal mRNA capping enzymes. Although the present study was limited to the enzymes that are predicted to catalyze the first two steps in formation of the m7GpppN cap in Giardia, we, and others (22), also identified a 422-aa putative Giardia cap methyltransferase (GenBank™ AACB01000009) that resembles the biochemically and structurally characterized cap methyltransferases of fungi, metazoa, and microsporidia (data not shown) (23, 24). Our initial attempts to produce catalytically active Giardia cap methyltransferase in bacteria were not successful. Nonetheless, we surmise that Giardia can synthesize m7GpppN caps.

This conclusion is consistent with a recent study (25) and our own data showing that a m7GpppN capped reporter mRNA, but not an uncapped reporter RNA, was translated in Giardia trophozoites in vivo when introduced into cells by electroporation. This result, and the fact that Giardia encodes two homologs of the cap-binding translation initiation factor eIF4E (data not shown), would argue that Giardia is not an exception to the general reliance on the cap structure for optimal gene expression in eukarya. The hypothesis that Giardia caps its RNAs is also consistent with an early study identifying several small RNAs in Giardia that reacted with antibody to the 2,2,7-trimethylguanosine (TMG) cap; the TMG-capped species were tentatively identified as Giardia snRNAs with possible roles in RNA processing (26). Available evidence suggests that TMG caps are formed by post-transcriptional N2-methylation of the standard m7GpppN cap (27). In this light, it is noteworthy that Giardia encodes a putative homolog of the yeast TMG cap synthase, Tgs1 (data not shown).

As current methods only allow two alleles in the tetraploid Giardia genome to be targeted in vivo, it is not possible to test directly whether the GICEG1 and GICEC1 genes are essential for Giardia viability or to assess their role in mRNA metabo-

lism. We attempted to exploit a surrogate genetic system by querying whether expression of the Giardia proteins could sustain the growth of S. cerevisiae cells that lack the endogenous guanylyltransferase Ceg1 or the triphosphatase Cet1. We found that neither GICEg1 nor GICEc1 could complement the respective yeast null mutations when they were expressed under the control of a constitutive yeast promoter on a centromeric plasmid (data not shown). GICEg1 was also unable to complement the yeast ceg1Δ mutation when provided on a multicopy plasmid (data not shown). A possible explanation for the failure of the Giardia enzymes to function in yeast is that they are unable to interact with yeast RNA polymerase II. It is noteworthy that the largest subunit of Giardia RNA polymerase II (28) lacks a carboxy-terminal domain, a component of tandem repeats of the heptapeptide YSPTSPS, which is characteristic of fungal, plant, and metazoan RNA polymerase II (29). Among its many known or imputed functions, the carboxy-terminal domain acts as a landing pad for the targeting of the guanylyltransferase component of the fungal and metazoan capping apparatus to the transcription elongation complex (30, 31).

G. lamblia is a significant cause of water-borne enteritis in humans. Although not yet validated as essential for Giardia growth, the triphosphatase GICEt1 is a potential target for anti-giardial drug discovery because: (i) the structure and catalytic mechanism of the triphosphatase family of which GICEt1 is a member is completely different from that of the human host capping enzyme and (ii) humans encode no plausible homolog of GICEt1 or other tunnel family RNA triphosphatases. The similarity between GICEt1 and other metal-dependent RNA triphosphatases raises the possibility of achieving broad-spectrum antiprotozoal and antifungal activity with a single class of mechanism-based inhibitors.

An Expanded Cap-centric Phylogeny of Eukarya—A scheme of eukaryotic phylogeny was suggested (2, 3) based on two features of the mRNA capping apparatus: the structure and mechanism of the triphosphatase component (metal-dependent “tunnel” type versus metal-independent cysteine-phospha-
tase type) and whether the triphosphatase is physically linked in cis to the guanylyltransferase component (Fig. 9). By these simple criteria, relying on stark differences in the same metabolic pathway, we arrived at different relationships among taxa than those suggested by comparisons of sequence variations among proteins that are themselves highly conserved in all eukaryotes (32). For example, the capping-based phylogeny placed metazoans (specifically arthropods, nematodes, and vertebrates) in a common lineage with plants (exemplified by Arabidopsis thaliana) because these organisms have a cysteine-phosphatase-type RNA triphosphatase fused in cis to their guanylyltransferase (Fig. 9). Fungi, microsporidia, and many protozoa (including Plasmodium, Trypanosoma, and, as shown here, Giardia) fall into a different lineage distinguished by a tunnel family RNA triphosphatase that is physically separate from RNA guanylyltransferase. In contrast to the cap-centric view, the protein sequence variation-based scheme (32) places fungi in the same supergroup as metazoa and puts Plasmodium and Trypanosoma nearer to plants.

Recent progress in plant genome sequencing has consolidated the conservation of bifunctional capping enzymes in plants and metazoans. Searching the plant databases at TIGR (tigrblast.tigr.org/tgi) with an Arabidopsis capping enzyme revealed clear evidence of a cysteine-phosphatase/guanylyltransferase fusion in many species of higher plants, including rice, tomato, potato, maize, barley, and wheat (data not shown). It is therefore highly instructive that the plastid-containing red algae, which shares an evolutionary history with plants (20), has a different capping apparatus akin to that of yeast and unicellular protozoa. The implication is that the plant/metazoan-type capping apparatus appeared subsequent to branching of the red alga from the ancestral algal/plant lineage. A key question is whether the differences in capping enzymes between fungi/red algae/protozoa and metazoa/plants reflect a transition from unicellularity to multicellularity. We suspect this is not the case, insofar as the unicellular green alga Chlamydomonas reinhardtii encodes a cysteine-phosphatase-type RNA triphosphatase along with a guanylyltransferase. However, because we do not find in the available genome data base a continuous genomic clone or cDNA that includes both the triphosphatase and the guanylyltransferase active sites, we can only tentatively place the green alga in the plant/metazoan group with respect to capping. It is conceivable that a capping enzyme composed of a cysteine-phosphatase-type RNA triphosphatase arose first in green algae and was retained in the ensuing plant lineage and transferred horizontally to a separate ancestor of the metazoan lineage. Alternatively, green algae, plants, and metazoas might have descended from a common ancestral lineage.

Given that Giardia is regarded as a basal eukaryote, the available evidence strongly favors the model that a three-component capping system with a metal-dependent tunnel family triphosphatase is the ancestral state from which other eukarya evolved by gene fusion and gene replacement with a novel analog (Fig. 9). It was noted previously (3) that certain protozoa (e.g., Entamoeba histolytica) acquired a bifunctional capping enzyme by fusion of the ancestral triphosphatase and guanylyltransferase genes (Fig. 9). The candidate E. histolytica capping enzyme is composed of an N-terminal tunnel family triphosphatase domain fused to a guanylyltransferase domain. A similar bifunctional tunnel family triphosphatase/guanylyltransferase fusion appears to exist in Dictostelium discoideum. By mining the genome data base for the water mold P. ramorum (www.jgi.doe.gov/genomes), we find that this unicellular organism also encodes a bifunctional tunnel-type triphosphatase/guanylyltransferase fusion (data not shown). Phytophthora is classified as a Stramenipole, a kingdom that includes diatoms and brown algae.

Metazoas and plants have experienced a different and more radical gene rearrangement that transferred a cysteine-phosphatase domain into the same transcription unit as the guanylyltransferase. A pathway of evolution was suggested, entailing the appearance of a new cysteine phosphatase enzyme (e.g., via duplication and mutation of one of the protein phosphatase genes present in lower eukarya) that gained the capacity to hydrolyze an RNA 5′-phosphate instead of, or in addition to, a phosphoprotein (3). The subsequent fusion of the cysteine phosphatase to the guanylyltransferase then allowed for the loss of the tunnel family enzyme from the genome of a putative common metazoan/plant ancestor or else the divergence of the protein to a point that it is no longer discernable as Cet1-like. An alternative explanation is that the cysteine-phosphatase/guanylyltransferase fusion arose first in either the green algal/plant (or metazoan) lineage and was subsequently transferred horizontally to the metazoan (or green algal/plant) ancestor. Another alternative explanation is that green algae/plants and metazoas independently experienced this gene fusion in distant branches of the phyllogenetic tree.

The cap-centric scheme of eukaryotic evolution is useful for its simplicity and because answers to questions about missing evolutionary links and the order of events in the progression from fungal and protozoal-type to metazoan and plant-type capping systems are emerging as more genomes are sequenced. A key outstanding issue is the nature of the mRNA capping apparatus in the earliest branching metazoan organisms.

REFERENCES

1. Shuman, S. (2000) Prog. Nucleic Acids Res. Mol. Biol. 66, 1–40
2. Ho, C. K., and Shuman, S. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 3050–3055
3. Shuman, S. (2002) Nat. Rev. Mol. Cell. Biol. 3, 619–625
4. Ho, C. K., and Shuman, S. (2001) J. Biol. Chem. 276, 46182–46186
5. Hausmann, S., Vivares, C. P., and Shuman, S. (2002) J. Biol. Chem. 277, 96–103
6. Lima, C. D., Wang, L. K., and Shuman, S. (1999) Cell 99, 533–543
7. Takagi, T., Monre, C. R., Diehn, F., and Buratowski, S. (1997) Cell 88, 867–873
8. Changela, A., Ho, C. K., Martins, A., Shuman, S., and Mondragon, A. (2001) J. Biol. Chem. 276, 2575–2586
9. Ho, C. K., Pei, Y., and Shuman, S. (1998) J. Biol. Chem. 273, 34151–34156
10. Bisaillon, M., and Shuman, S. (2001) J. Biol. Chem. 276, 17261–17266
11. Gong, C., and Shuman, S. (2002) J. Biol. Chem. 277, 15317–15324
12. Gong, C., and Shuman, S. (2003) Virology 309, 125–134
13. Martins, A., and Shuman, S. (2003) Nucleic Acids Res. 31, 1455–1463
14. Nixon, J. F., Wang, A., Morrison, H. G., McGrath, A. G., Sogin, M. L., Luntz, B. J., and Samuelson, J. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 3701–3705
15. Yu, D. C., Wang, A. L., Botka, C. W., and Wang, C. C. (1998) Mol. Biochem. Parasitol. 96, 151–165
16. Knodler, L. A., Svard, S. G., Silberman, J. D., Davids, B. J., and Gillin, F. D. (1999) Mol. Microbiol. 34, 327–340
17. Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, P., Nishida, K., Yoshida, Y., Nishimura, Y., Makao, S., Kobayashi, T., Monomura, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka, M., Hasegawa, T., Kubota, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takana, H., Ohta, N., Kuriyama, H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogawa, H., and Kohara, Y., and Kuriwai, T. (2004) Nature 428, 653–657
18. Mosevicsa, E., and Lima, C. D. (2000) Mol. Cell. Biol. 20, 1253–1264
19. Gong, C., Martins, A., and Shuman, S. (2003) J. Biol. Chem. 278, 14655–14662
20. McFadden, G. I., and van Dooren, G. G. (2004) Mol. Biochem. Parasitol. 66, 49–57
21. Bujnicki, J., Feder, M., Radlinska, M., and Rychlewski, L. (2001) BMC Bioinformatics 2, 2
22. Fabrega, C., Hausmann, S., Shen, V., Shuman, S., and Lima, C. D. (2004) Mol. Cell 13, 77–89
23. Saha, N., Schwer, B., and Shuman, S. (1999) J. Biol. Chem. 274, 16553–16562
24. Li, L., and Wang, C. C. (2004) J. Biol. Chem. 279, 14656–14664
25. Niu, X. X., Hartshorne, T., He, X. Y., and Agabian, N. (1994) Mol. Biochem. Parasitol. 66, 49–57
26. Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., and Bordonné, R. (2002) Mol. Cell 9, 891–901
28. Seshadri, V., McArthur, A. G., Sogin, M. L., and Adam, R. D. (2003) *J. Biol. Chem.* **278**, 27804–27820
29. Stiller, J. W., and Cook, M. S. (2004) *Eukaryotic Cell* **3**, 735–740
30. Ho, C. K., and Shuman, S. (1999) *Mol. Cell* **3**, 405–411
31. Fabrega, C., Shen, V., Shuman, S., and Lima, C. D. (2003) *Mol. Cell* **11**, 1549–1561
32. Baldauf, S. L., Roger, A. J., Wenk-Siefart, I., and Doolittle, W. F. (2000) *Science* **290**, 972–977
33. Keister, D. B. (1983) *Trans. R. Soc. Trp. Med. Hyg.* **77**, 487–488
34. Gallie, D. R. (1991) *Genes Dev.* **5**, 2108–2116
35. Singer, S. M., Yee, J., and Nash, T. E. (1998) *Mol. Biochem. Parasitol.* **92**, 59–69
36. Elmendorf, H. G., Singer, S. M., and Nash, T. E. (2001) *Nucleic Acids Res.* **29**, 4674–4683