A scheme of eukaryotic phylogeny has been suggested based on the structural and physical linkage of the enzymes that catalyze mRNA cap formation. Here we show that the intracellular parasite *Encephalitozoon cuniculi* encodes a complete mRNA capping apparatus consisting of separate triphosphatase (EcCet1), guanylyltransferase (EcCeg1), and methyltransferase (Ecm1) enzymes, which we characterize biochemically and genetically. The triphosphatase EcCet1 belongs to a metal-dependent phosphohydrolase family that includes the triphosphatase components of the capping apparatus of fungi, DNA viruses, and the malaria parasite *Plasmodium falciparum*. These enzymes are structurally and mechanistically unrelated to the metal-independent cysteine phosphatase-type RNA triphosphatases found in metazoans and plants. Our findings support the proposed evolutionary connection between microsporidia and fungi, and they place fungi and protozoa in a common lineage distinct from that of metazoans and plants. RNA triphosphatase presents an attractive target for antiprotozoal/antifungal drug development.

The 5' cap is a distinctive feature of eukaryotic mRNA. The cap consists of 7-methyl guanosine linked via an inverted 5'-5' triphosphate bridge to the initiating nucleoside of the transcript. Cellular mRNAs are capped via three enzymatic reactions. (i) The 5' triphosphate end of the nascent pre-mRNA is hydrolyzed to a diphosphate by RNA 5' 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).

Although the three capping reactions are universal in eukaryotes, there is a surprising diversity in the genetic organization of the cap-forming enzymes in “lower” versus “higher” eukaryotes, as well as a complete divergence in the structure and catalytic mechanism of the RNA triphosphatase component (1). Metazoans and plants have a two-component capping system consisting of a bifunctional triphosphatase-guanylyltransferase polypeptide and a separate methyltransferase polypeptide, whereas fungi and the protozoan parasite *Plasmodium falciparum* have a three-component system consisting of separate triphosphatase, guanylyltransferase, and methyltransferase gene products (1, 2). 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 (3, 4).

Metazoan and plant RNA triphosphatases belong to the cysteine phosphate enzyme superfamily, which is defined by the conserved phosphate-binding loop motif HCXXXRR(S/T). Mammalian RNA triphosphatase catalyzes a two-step ping-pong 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 (4). 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 new family of metal-dependent phosphohydrolases that embraces the triphosphatase components of the poxvirus, baculovirus, *Chlorella* virus, and *P. falciparum* mRNA capping systems (1, 2, 5–11). 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 (β1 and β11 in Fig. 1) that are required for catalysis by every family member and that 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 (3). The β strands comprising the tunnel walls are displayed over the Cet1 protein sequence in Fig. 1. Each of the eight strands contributes at least one functional constituent of the active site (7, 8, 12). The 15 individual side chains within the tunnel that are important for Cet1 function in *vitro* and *in vivo* are denoted by dots in Fig. 1.

When and how did the RNA triphosphatase component of the capping apparatus diverge during eukaryotic evolution, and can the differences in the organization of the capping apparatus provide a useful guide to phylogenetic relationships among taxa? We are pursuing these questions by systematically identifying the cap-forming enzymes of unicellular eukaryotes and evaluating their structural and mechanistic relatedness to exemplary fungal, viral, and mammalian enzymes. Initial studies revealed that the triphosphatase of the protozoan parasite *P. falciparum* belongs to the yeast-like RNA triphosphatase family (2).

Here we undertook a “post-genomics” inquiry into the capping apparatus of the microsporidian organism *Encephalitozoa*...
zoon cuniculi. Microsporidia are amoeboid obligate intracellular parasites that cause disease in animals and humans (13). *E. cuniculi*, which infects a variety of mammalian hosts, has the smallest genome of any eukaryotic organism characterized to date, consisting of ~2900 kbp of DNA organized in 11 chromosomes (14). Complete sequencing of the *E. cuniculi* genome reveals a highly compact arrangement of 1997 potential protein-coding genes (14). *E. cuniculi* encodes a multisubunit RNA polymerase II with a prototypical carboxy-terminal domain on its largest subunit consisting of 13 tandem repeats of the heptapeptide YSPTSPS. Among its many known or purported functions, the carboxy-terminal domain acts as a landing pad for the targeting of the mRNA capping apparatus to the transcription elongation complex (15).

The pertinent questions about capping in this intracellular parasite are as follows. (i) Does *E. cuniculi* encode a full set of mRNA capping enzymes, or might it encode one or more capping enzymes from the host organism? (ii) If *E. cuniculi* does encode its own capping apparatus, how are the components organized genetically (separately encoded as in fungi or fused in metazoans, plants, and poxviruses)? (iii) Is the triphosphatase component related to the cysteine-phosphatase metazoan-type enzymes or to the metal-dependent yeast-type enzymes? Using genomic sequence information to identify candidate capping enzymes, and biochemical and genetic approaches to establish their activities, we show that *E. cuniculi* has a complete three-component capping system similar to that of fungi. The evolutionary implications are discussed.

**MATERIALS AND METHODS**

**Yeast Expression Plasmids for *E. cuniculi* Capping Enzymes**—The EcoCEG1, EcoCET1, and ECM1 genes encoding the putative RNA guanylyltransferase, RNA triphosphatase, and RNA (guanine-N7) methyltransferase enzymes were amplified by PCR from *E. cuniculi* genomic DNA using oligonucleotide primers designed to introduce an NcoI restriction site at the translation initiation codon and a BamHI site 3' of the translation stop codon of each gene. The PCR products were digested with NcoI and BamHI and then inserted between the NcoI and BamHI sites of the yeast shuttle vector pYX132 (CEN TRP1) to generate pYX142-EcoCEG1, pYX142-EcoCET1, and pYX142-ECM1. The plasmids were sequenced to confirm that the *E. cuniculi* gene inserts were identical to the genomic DNA sequences (14). Expression of the *E. cuniculi* genes in these plasmids is under the control of the constitutive yeast TPI1 promoter. A yeast two μ vector bearing EcoCEG1 was constructed by transferring an AarII-Hael fragment from pYX132-EcoCEG1 into pYX202 (2μ TRP1) to generate pYX202-EcoCEG1. Expression of the EcoCEG1 gene in the 2μ vector driven by the TPI1 promoter.

**Site-directed Mutagenesis**—Alanine substitution mutations were introduced into the EcoCEG1, EcoCET1, and ECM1 genes by the PCR-based two-stage overlap extension method. The mutated genes were inserted into pYX132. The resulting pYX132-EcoCEG1-Ala, pYX132-EcoCET1-Ala, and pYX132-ECM1-Ala plasmid insertions were sequenced completely to confirm the presence of the desired mutation and exclude the acquisition of unwanted changes during amplification and cloning. The EcoCEG1-K63A mutant was transferred into the 2μ TRP1 vector to generate pYX232-EcoCEG1-K63A.

**Recombinant EcoCet1—An NcoI-BamHI fragment containing the EcoCET1 reading frame was excised from pYX132-EcoCET1 and then inserted between the NcoI and BamHI sites of the customized T7 RNA polymerase-based vector pET16bNco (which contains an NcoI site instead of the NdeI in pET16b) so as to fuse the EcoCet1 polypeptide in-frame to the NH2-terminal His6-Smt3 leader. The pET-His6-Smt3-EcoCet1 plasmid was introduced into *E. coli* BL21(DE3). A 500-ml culture of *E. coli* BL21(DE3)/pET-EcoCet1 was grown at 37°C in Luria-Bertani medium containing 0.1 mg/ml ampicillin until the OD600 reached 0.6. The culture was placed on ice for 30 min and then adjusted to 0.2 mM isopropyl-1-thio-β-D-galactopyranoside and 2% (v/v) ethanol. After further incubation for 17 h at 18°C with constant shaking, the cells were harvested by centrifugation, and the pellet was stored at ~80°C. All subsequent procedures were performed at 4°C. Thawed bacteria were resuspended in 30 ml of buffer A (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 10% glycerol). Phenylmethylsulfonyl fluoride and lysozyme were added to final concentrations of 1 mM and 10 mg/ml, respectively. After incubation on ice for 30 min, Triton X-100 was added to a final concentration of 0.1% and the lysate was sonicated to reduce viscosity. Insoluble material was removed by centrifugation for 45 min at 18,000 rpm in a Sorvall SS34 rotor. The soluble extract was mixed for 30 min with 1 ml of Ni2+-nitrilotriacetic acid-agarose (Qiagen) that had been equilibrated with buffer A containing 0.1% Triton X-100 and 50 mM imidazole. The slurry was poured into a column, and the resin was washed with 10 ml of buffer B (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 10% glycerol, 0.1% Triton X-100) and 5 mM imidazole. EcoCet1 was then step-eluted with 1-ml aliquots of 10, 50, 100, and 200 mM imidazole in buffer B. The polypeptide compositions of the column fractions were monitored by SDS-PAGE. The 200 mM imidazole eluate fraction containing EcoCet1 was dialyzed against 50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2 mM DTT, 0.05% Triton X-100, 10% glycerol and then stored at ~80°C. The EcoCet1-Ala mutants were purified using the same procedure. Protein concentrations were determined using the Bio-Rad dye reagent with bovine serum albumin as the standard.

**RESULTS**

**Identification of a Candidate *E. cuniculi* Triphosphatase EcoCet1**—We identified a candidate RNA triphosphatase gene on *E. cuniculi* chromosome 4 encoding a 221-amino acid polypeptide with primary structural similarity to the catalytic domain of fungal, viral, and Plasmidium RNA triphosphatases. We named this *E. cuniculi* gene product EcoCet1 (for *E. cuniculi* capping enzyme triphosphatase 1). EcoCet1 is slightly larger than the "minimal" RNA triphosphatase of *Chlorella* virus PBCV-1 (193 amino acids), but is considerably smaller than the RNA triphosphatases of *S. cerevisiae* (Cet1; 549 amino acids), *C. albicans* (CaCat1; 520 amino acids), and *S. pombe* (Pet1; 303 amino acids). Cet1 and CaCat1 contain large nonessential NH2-terminal extensions (8, 15–17) that are missing from Pet1 (11) and EcoCet1. EcoCet1 includes the two metal-binding motifs characteristic of the fungal, viral, and Plasmidium triphosphatases, plus putative homologs of all of the other β strands that comprise the active site tunnel of Cet1 (Fig. 1). Fourteen of the 15 hydrophilic amino acids that are essential for catalysis by Cet1 are conserved in the *E. cuniculi* polypeptide.

**Metal-dependent Triphosphatase Activity of EcoCet1**—The EcoCet1 gene was cloned into a T7 RNA polymerase-based pET vector so as to place the open reading frame in-frame with an NH2-terminal His tag. The expression plasmid was introduced into *E. coli* BL21(DE3), a strain that contains the T7 RNA polymerase gene under the control of the lac promoter. A new ~30-kDa polypeptide corresponding to His-EcoCet1 was detectable by SDS-PAGE in extracts of IPTG-induced bacteria bearing the pET-EcoCet1 plasmid (data not shown). Initial purification of the His-EcoCet1 protein was achieved by adsorption to nickel-agarose and elution with 200 mM imidazole. The nickel-agarose preparation was highly enriched with respect to the EcoCet1 polypeptide, as judged by SDS-PAGE (Fig. 2A).

---

1 The abbreviations used are: DTT, dithiothreitol; WT, wild-type; 5-FOA, 5-fluoroorotic acid.
We found that recombinant EcCet1 catalyzed the release of $^32\text{P}$ from [$\gamma$$^32\text{P}$]ATP in the presence of manganese and that the extent of ATP hydrolysis was proportional to enzyme concentration; the substrate was hydrolyzed completely at saturating enzyme (Fig. 2B). There was no detectable ATP hydrolysis in the absence of a divalent cation (Fig. 3A). Hydrolysis of 1 mM ATP was optimal at 2 mM MnCl$_2$ and declined slightly at 3–5 mM MnCl$_2$ (Fig. 3A). Cobalt also satisfied the metal requirement and activity was optimal at 2–4 mM CoCl$_2$. No activity was observed when the manganese and cobalt concentrations were below the level of input ATP. Magnesium failed to support ATP hydrolysis at up to 5 mM concentration (Fig. 3A). 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 $^32\text{P}$ from [$\gamma$$^32\text{P}$]ATP was identical to the rate of conversion of [$\alpha$$^32\text{P}$]ATP to [$\alpha$$^32\text{P}$]ADP in a parallel reaction mixture containing the same concentration of EcCet1 (Fig. 3B). We detected no formation of [$\alpha$$^32\text{P}$]AMP during the reaction. Hence, we conclude that EcCet1 catalyzes the hydrolysis of ATP to ADP and $P_1$. EcCet1 also converts [$\gamma$$^32\text{P}$]GTP to [$\gamma$$^32\text{P}$]GDP (data not shown).

Kinetic parameters for EcCet1 were determined by measuring the extent of $^32\text{P}$ formation during a 15-min reaction as a function of input [$\gamma$$^32\text{P}$]ATP concentration in the range of 3–200 $\mu$M (Fig. 3C). A double-reciprocal plot of the data fit well to a linear function (Fig. 3C, inset). We calculated a $K_m$ of 96 $\mu$M ATP and a $V_{max}$ of 3 s$^{-1}$.

EcCet1 Activity Is Abolished by Replacement of Conserved Glutamates with Alanine—EcCet1 residues Glu-32 and Glu-34 in the Cet1 tunnel are denoted above the sequence. Peptide segments with the highest degree of conservation in all seven proteins are highlighted by the shaded boxes. Hydrophilic amino acids that comprise the active site within the Cet1 tunnel are denoted by dots. EcCet1 residues Glu-32 and Glu-34 in beta 11 were targeted for alanine substitution in the present study.
pattern is reprised in EcCet1 (Fig. 1) and suggests that the structural context for its metal-binding site is similar to that of Cet1.

Sedimentation Analysis of EcCet1—The native size of recombinant EcCet1 was analyzed by sedimentation of the protein through a 15–30% glycerol gradient. Marker proteins catalase (248 kDa), bovine serum albumin (66 kDa), and cytochrome c (13 kDa) were included as internal standards in the same gradient. After centrifugation, the polypeptide compositions of the gradient fractions were analyzed by SDS-PAGE (Fig. 4). A plot of the S values of the three standards versus fraction number yielded a straight line (not shown). His-EcCet1 (a 28-kDa polypeptide) sedimented as a single peak coincident with bovine serum albumin. We surmise that EcCet1 is a homodimer. The triphosphatase activity profile paralleled the abundance of the EcCet1 polypeptide (Fig. 4). The same distribution of EcCet1 protein and ATPase activity was observed when the enzyme was sedimented in the absence of internal standards (data not shown). The homodimeric quaternary structure of EcCet1 is a feature that is shared with the RNA triphosphatases of S. cerevisiae and S. pombe (3, 11, 18).

Identification and Characterization of E. cuniculi mRNA Guanylyltransferase EcCeg1—We identified a candidate mRNA guanylyltransferase gene on E. cuniculi chromosome 9 encoding a 364-amino-acid polypeptide with extensive similarity to fungal, protozoal, metazoan, and virus-encoded RNA guanylyltransferases (Fig. 5A). We named this E. cuniculi gene product EcCeg1 (for E. cuniculi capping enzyme guanylyltransferase). EcCeg1 is intermediate in size between the “minimal”...
guanylyltransferase of Chlorella virus PBCV-1 (330 amino acids) and the RNA guanylyltransferases of S. cerevisiae (Ceg1; 459 amino acids), S. pombe (Pce1; 402 amino acids), and C. albicans (Cgt1; 449 amino acids). EcCeg1 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. 5). The polypeptides were visualized by staining with Coomassie Blue dye. The identities of the polypeptides are indicated at the left of the gel. Bottom panel, ATPase activity profile. Reaction mixtures (20 µl) containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 5 mM MgCl2, 0.17 mM NaCl, 1 mM EDTA, 2 mM DTT, 0.05% Triton X-100. The mixture was incubated for 15 min at 37 °C and analyzed by SDS-PAGE along with samples of the input protein mixture (lane L). Polypeptides were visualized by staining with Coomassie Blue dye. The positions and sizes (kDa) of co-electrophoresed marker polypeptides are indicated on the left. The enzyme-[32P]GMP complex is denoted by the arrow on the left. Guanylyltransferase reaction mixtures (20 µl) containing 50 mM Tris-HCl, pH 8.0, 5 mM DTT, 5 mM MgCl2, 0.17 mM EDTA, 2 mM DTT, 0.05% Triton X-100. The reactions were quenched with SDS, and the products were resolved by SDS-PAGE. An autoradiograph of the dried gel is shown (bottom panel). The positions and sizes (kDa) of prestained marker polypeptides are indicated on the left. The enzyme-[32P]GMP complex is denoted by the arrow on the right.

FIG. 5. EcCeg1 guanylyltransferase activity in vitro. A, the amino acid sequence of EcCeg1 is shown with the six nucleotidyltransferase motifs highlighted in boxes. B, recombinant His6-Smt3-EcCeg1 was purified from bacteria by nickel-agarose chromatography. The polypeptide compositions of aliquots (20 µl) of the soluble bacterial lysate (lane L), the nickel-agarose flow-through (lane F), and the wash (lane W) fractions, and the 10, 50, 100, and 200 mM imidazole eluate fractions were analyzed by SDS-PAGE (top panel). The polypeptides were visualized by staining with Coomassie Blue dye. The positions and sizes (kDa) of co-electrophoresed marker polypeptides are indicated on the left. The positions of the His6-Smt3-EcCeg1 polypeptide are indicated by an arrow on the left. Guanylyltransferase reaction mixtures (20 µl) containing 50 mM Tris-HCl, pH 8.0, 5 mM DTT, 5 mM MgCl2, 0.17 mM EDTA, 2 mM DTT, 0.05% Triton X-100. The reactions were quenched with SDS, and the products were resolved by SDS-PAGE. An autoradiograph of the dried gel is shown (bottom panel). The positions and sizes (kDa) of prestained marker polypeptides are indicated on the left. The enzyme-[32P]GMP complex is denoted by the arrow on the right.

The identities of the polypeptides are indicated at the left of the gel. Bottom panel, ATPase activity profile. Reaction mixtures (20 µl) containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 5 mM MnCl2, 1 mM [γ-32P]ATP, and 2 µl of the indicated glycerol gradient fractions were incubated for 15 min at 37 °C. The peaks of the marker proteins are indicated by arrows.

EcCeg1 guanylyltransferase activity in vitro. A, the amino acid sequence of EcCeg1 is shown with the six nucleotidyltransferase motifs highlighted in boxes. B, recombinant His6-Smt3-EcCeg1 was purified from bacteria by nickel-agarose chromatography. The polypeptide compositions of aliquots (20 µl) of the soluble bacterial lysate (lane L), the nickel-agarose flow-through (lane F) and the wash (lane W) fractions, and the 10, 50, 100, and 200 mM imidazole eluate fractions were analyzed by SDS-PAGE (top panel). The polypeptides were visualized by staining with Coomassie Blue dye. The positions and sizes (kDa) of co-electrophoresed marker polypeptides are indicated on the left. The position of the His6-Smt3-EcCeg1 polypeptide is indicated by an arrow on the left. Guanylyltransferase reaction mixtures (20 µl) containing 50 mM Tris-HCl, pH 8.0, 5 mM DTT, 5 mM MgCl2, 0.17 mM EDTA, 2 mM DTT, 0.05% Triton X-100. The reactions were quenched with SDS, and the products were resolved by SDS-PAGE. An autoradiograph of the dried gel is shown (bottom panel). The positions and sizes (kDa) of prestained marker polypeptides are indicated on the left. The enzyme-[32P]GMP complex is denoted by the arrow on the right.

As there are no methods available to manipulate the E. cuniculi genome in vivo, it is not possible to test directly whether the EcCEG1 gene is essential for E. cuniculi viability or to assess its role in mRNA metabolism. To circumvent this performed using the nickel-agarose fraction. Formation of the covalent intermediate required a divalent cation cofactor, either magnesium or manganese (data not shown). The reaction was specific for GTP, insofar as we could detect no formation of a protein-NMP complex during reaction of EcCeg1 with [α-32P]ATP (data not shown).

As there are no methods available to manipulate the E. cuniculi genome in vivo, it is not possible to test directly whether the EcCEG1 gene is essential for E. cuniculi viability or to assess its role in mRNA metabolism. To circumvent this
Microsporidian RNA Capping Enzymes

Problem, we exploited a yeast-based system for genetic analysis of heterologous capping enzymes (22, 23). The system provides the capacity to answer the following questions concerning the EcCeg1. Can the E. cuniculi protein function in the cap-synthetic pathway and sustain the growth of yeast cells that lack the endogenous mRNA guanylyltransferase Ceg1? If so, does complementation of the yeast ceg1 mutation by EcCEG1 depend on its catalytic activity in cap formation?

We cloned the EcCEG1 gene into a yeast 2μ TRP1 plasmid under the transcriptional control of the constitutive yeast TP1I promoter. The function of the EcCEG1 gene was tested by plasmid shuffle in yeast ceg1Δ cells that contain CEG1 on a CEN URA3 plasmid. The ceg1Δ strain is unable to form colonies on medium containing 5-fluoroorotic acid (5-FOA), a drug that selects against the URA3 plasmid, unless it is transformed with a second plasmid bearing CEG1 or a functional homolog from another source. We found that 2μ EcCEG1 supported the growth of ceg1Δ cells on 5-FOA at 30 °C. The 2μ EcCEG1 cells grew as well as wild-type CEG1 cells on rich medium (YPD agar) at 30 °C, but formed tiny colonies at 25 °C, and failed to grow at either 18 °C or 37 °C (Fig. 6). These results show that EcCEG1 encodes a biologically active RNA guanylyltransferase, albeit one that is both cold-sensitive and temperature-sensitive in vivo in yeast. The basis for the conditional growth defects of EcCeg1 yeast cells remains to be clarified.

Motif I of EcCEG1 (KKSQDVGxG) contains the conserved lysine nucleophile to which GMP becomes covalently attached in other well-characterized capping enzymes (19, 20). We mutated the Lys-63 codon to an alanine codon and tested the K63A allele in the plasmid shuffle assay. K63A was unable to support the growth of ceg1Δ on 5-FOA (data not shown). Thus, the catalytic nucleophile is required for in vivo activity of EcCeg1.

Identification and Characterization of E. cuniculi Cap Methyltransferase Ecm1—We identified a candidate cap methyltransferase gene on E. cuniculi chromosome 10 that encodes a 298-amino acid polypeptide with similarity to fungal and human cap methyltransferases (Fig. 7). We named this gene product Ecm1 (Encephalitozoon cap methyltransferase). Ecm1 is considerably smaller than the cap methylating enzymes of humans (Hcm1; 476 amino acids), C. albicans (Ccm1; 474 amino acids), S. cerevisiae (Abd1; 436 amino acids), and S. pombe (Pcm1; 402 amino acids). The fungal and human cap methyltransferases contain nonessential NH2-terminal extensions (23–25) that are missing from Ecm1. Ecm1 contains the conserved AdoMet binding motif VLEXGXGXG, and seven of eight of the amino acids that were found to be essential for the in vivo activity of S. cerevisiae and human cap methyltransferases (23–26) are conserved in the E. cuniculi polypeptide.

We cloned the ECM1 gene into a yeast CEN TRP1 plasmid under the control of the yeast TP1I promoter and tested by plasmid shuffle for its ability to sustain the growth of yeast abd1Δ cells that lack the essential endogenous cap methyltransferase (23). Control cells transformed with a TRP1 ABD1 plasmid grew on 5-FOA, whereas cells transformed with the TRP1 vector were incapable of growth on 5-FOA. The instructive finding was that cells bearing the ECM1 plasmid grew on 5-FOA (data not shown). Thus, the Ecm1 is active as a cap methyltransferase in vivo. Yeast Ecm1 cells grew on rich medium (YPD) at 18, 25, 30, and 37 °C, and colony size was indistinguishable from wild-type ABD1 cells (Table I and Fig. 8).

Alanine mutations were introduced at five residues (Asp-70, Gly-74, Asp-78, Arg-106, and Arg-106) that are conserved in other cap methyltransferase (Fig. 7). The ECM1-Ala genes were cloned into CEN TRP1 vectors and tested by plasmid shuffle for complementation of abd1Δ. One of the mutants, D94A, failed to give rise to FOA-resistant colonies during selection at any temperature tested (Table I). Thus, we conclude that Asp-94 is essential for Ecm1 function. The corresponding Asp-194 residue in S. cerevisiae Abd1 is essential for cap methylation in vivo and in vitro (23–26). Structural modeling of Abd1 based on the known structure of glycine N-methyltransferase suggests that Asp-194 directly coordinates the ribose oxygens of the AdoMet substrate (27).

Three other ECM1 mutants, D70A, D78A and R106A, formed tiny 5-FOA-resistant colonies after 7 days of selection at 25 or 30 °C. The R106A strain formed pinpoint colonies on YPD agar at all temperatures. D78A cells formed pinpoint colonies at 18 and 25 °C, but failed to grow at 37 °C. D70A formed small colonies at 25 and 30 °C, pinpoint colonies at 18 °C, and failed to grow at 37 °C (Table I). Thus, elimination of the Asp-70, Asp-78, and Arg-106 side chains seriously compromised cap methyltransferase activity in vivo. Alanine mutants of equivalent residues in Abd1 (Glu-170, Asp-178, and Arg-206) were lethal or extremely sick in vivo and inactive or partially defective in vitro (25); however, those mutants were tested for function in vivo under the control of the natural ABD1 promoter rather than the strong constitutive TP1I promoter used presently to test the ECM1 mutants. Thus, we suspect that high level expression of the Ecm1 D70A and D78A proteins suffices for slow growth of yeast cells at permissive temperatures, but does not meet a threshold level of cap methyltransferase activity at 37 °C. The R106A protein was defective irrespective of the growth temperature.

The ECM1-G74A mutant grew normally at 25, 30, and 37 °C, but formed smaller colonies than wild-type ECM1 at 18 °C (Fig. 8 and Table I). This result differs from the lethality of the corresponding G74A mutation in ABD1 (24). Note that this glycine is situated within the loop that forms the AdoMet binding site. Whereas the glycine may facilitate adoption of the
Microsporidian RNA Capping Enzymes

---

**Table 1**

| ECM1 allele | Complementation of *abd1Δ* |
|-------------|-----------------------------|
| WT          | + + + + + + + + + + + + + + |
| D70A        | + + + + + + + + + + + + + + |
| G74A        | + + + + + + + + + + + + + + |
| D78A        | + + + + + + + + + + + + + + |
| D94A        | + + + + + + + + + + + + + + |
| R106A       | + + + + + + + + + + + + + + |

**Fig. 7.** Ecm1 is structurally related to fungal cap methyltransferases. The amino acid sequence of Ecm1 from residues 11 to 298 is aligned to the sequences of the COOH-terminal catalytic domains of *S. cerevisiae* Abd1, C. albicans Ccm1, S. pombe Pcm1, and human Hcm1. Gaps in the alignment are indicated by dots. The AdoMet-binding motif is highlighted in the shaded box. Residues Asp-70, Gly-74, Asp-78, and Arg-106 of Ecm1 that were targeted for alanine substitution in the present study are indicated by arrowheads. Asp-94, and Arg-106 of Ecm1 that were targeted for alanine substitution in the present study are indicated by arrowheads.

**Fig. 8.** Ecm1 is active as a cap methyltransferase in vivo in *yeast*. The yeast *abd1Δ* strain YBS40 was transformed with *CEN TRP1* plasmids containing either *ECM1* or *ECM1-Ala* genes. Control transformations were performed with a *CEN TRP1 ABD1* plasmid and the *CEN TRP1* vector without an insert. Single *Trp*+ transformants were patched to agar plates lacking tryptophan and then streaked on agar medium containing 0.75 mg/ml 5-FOA. Individual FOA-resistant colonies bearing the *ABD1*, *ECM1*, or *ECM1-G74A* genes were picked and patched on agar medium lacking tryptophan. Cells were then streaked on YPD agar. The plate were photographed after incubation for 4 days at 30°C.

---

**DISCUSSION**

Obligate intracellular parasites often surrender essential functions to the host cell. Microsporidia, which lack mitochondria but contain several nuclear genes of supposed mitochondrial origin, are believed to have lost the organelle during their evolution from a mitochondrion-bearing ancestor (28). The microsporidian *E. cuniculi*, which has the smallest known eukaryotic genome, lacks the enzymes for several biosynthetic pathways and for the Krebs cycle, and is thus acutely dependent on the host cell for these metabolic functions (13). Here we have shown that *E. cuniculi* retains a full complement of RNA cap-forming enzymes. This was not a foregone conclusion, insofar as another highly derived genome, that of the cryptomonad nucleomorph, apparently encodes only one cap-forming activity (29).

The nucleomorph, a remnant of an "enslaved" algal nucleus in the chloroplast of cryptomonad cells that acquired the organelle via secondary endosymbiosis, has a 550-kb genome with high coding density and only 464 predicted proteins (29). It maintains an intimate symbiosis with the cell nucleus. With respect to RNA synthesis, the nucleomorph encodes three RNA polymerases and several general transcription factors. Interestingly, the largest subunit of the putative RNA polymerase II has no counterpart of the carboxyl-terminal domain. The nucleomorph encodes a 357-amino acid monofunctional RNA guanylyltransferase that contains all six of the diagnostic nucleotidytransferase motifs, but it does not encode a recognizable counterpart of either metazoan or fungal type RNA triphosphatases. Thus, the nucleomorph must rely on nuclear-encoded triphosphatase
and methyltransferase components to cap its mRNAs.

_E. cuniculi_ encodes a three-component mRNA capping apparatus similar to that of fungi and distinct from the two-component capping systems of metazoans and plants. Biochemical characterization of the triphosphatase EcCet1 placed it firmly within the family of metal-dependent phosphohydrolases that includes the RNA triphosphatases of fungi, DNA viruses, and the protozoan _P. falciparum_ (2). The guanylyltransferase EcCeg1 and the methyltransferase Ecm1 are clearly related, in structure and mechanism, to the “standard” guanylyltransferases and cap methyltransferase enzymes found in all eukaryotic species (1). The capping functions of EcCeg1 and Ecm1 were verified genetically by complementation of _S. cerevisiae_ strains deleted for the endogenous capping and methylating enzymes. Initial attempts to complement a yeast triphosphatase-null mutant with the EcCET1 gene were unsuccessful; possible explanations include: (i) failure of the EcCet1 protein to be expressed in yeast in amounts sufficient for cell growth, (ii) failure of EcCet1 to localize to the yeast nucleus or to be targeted to the transcription elongation complex within the nucleus, or (iii) inability of EcCet1 to fulfill an important non-catalytic role of the yeast triphosphatase, i.e. binding to and stabilizing the yeast guanylyltransferase Ceg1 (30). Note that there is no second homolog of a fungal-type triphosphatase in the _E. cuniculi_ proteome.

Capping enzymes are a reasonable focal point for considering eukaryotic evolution because the mRNA cap structure is ubiquitous in eukaryotic organisms, but absent from the bacterial and archaeal kingdoms. Thus, any differences in the capping apparatus between taxa would reflect events that post-date the emergence of ancestral nucleated cells. A heuristic scheme of eukaryotic phylogeny has been proposed 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). EcCet1 is the second biochemically verified example of a metal-dependent RNA triphosphatase family member from a protozoan species, the first being _P. falciparum_ RNA triphosphatase Prt1 (2). Moreover, _E. cuniculi_ and _P. falciparum_ both encode separate triphosphatase and guanylyltransferase enzymes. Therefore, the capping-based phylogeny would group fungi, microsporidia, and _Plasmodia_ (which are classified as Apicomplexa along with other pathogenic parasites _Toxoplasma_ and _Cryptosporidia_) into a common lineage distinguished by a Cet1-like RNA triphosphatase. The evolutionary relationship between fungi and microsporidia has been suggested previously (31, 32), but their connection to Apicomplexa is a novel insight. The capping-based phylogeny places metazoans in a different lineage together with Viridiplantae (exemplified by the metaphya _Arabidopsis_ and the unicellular alga _Chlamydomonas reinhardtii_) because all of these organisms have a cysteine-phosphatase type RNA triphosphatase fused in cis to their guanylyltransferase (1).

Microsporidia cause disease in host species ranging from silk worms to man (15). RNA triphosphatase is an attractive target for anti-microsporidial drug development because: (i) the structure and catalytic mechanism of the triphosphatase family of which EcCet1 is a member is completely different from that of the metazoan capping enzyme and (ii) metazoan genomes encode no identifiable homologs of EcCet1 or other fungal/protozoal RNA triphosphatases. Thus, a mechanism-based inhibitor of EcCet1 should be highly selective for the intracellular parasite and have minimal effect on the human or arthropod host. Additionally, the structural similarity between EcCet1 and other metal-dependent RNA triphosphatases raises the exciting possibility of achieving antiprotozoal and antifungal activity with a single class of mechanism-based inhibitors.

Finally, the coding density of the _E. cuniculi_ genome is remarkably high, and it has been suggested that this organism generally produces smaller versions of protein counterparts encoded by organisms with much larger genomes (14). The _E. cuniculi_ capping enzymes have been pared down to little more than the catalytic domains. Indeed, Ecm1 is the smallest cap methyltransferase that has been identified to date, which makes Ecm1 an attractive target for structure determination by crystallography.

_acknowledgment—_We thank Dr. Emmanuel Cornillot for bioinformatic analysis.

**References**

1. Shuman, S. (2000) _Proc. Natl. Acad. Sci. U. S. A._ 97, 1091–1096
2. Ho, C. K., and Shuman, S. (2001) _Cell_ 105, 51–60
3. Hirt, R. P., Logsdon, J. M., Dorey, M. W., Doolittle, W. F., and Embley, T. M. (2000) _EMBO J._ 19, 1561–1568
4. Hausmann, S., Ho, C. K., Schwer, B., and Shuman, S. (2001) _Proc. Natl. Acad. Sci. U. S. A._ 98, 3050–3055
5. Lima, C. D., Wang, L. K., and Shuman, S. (1999) _Cell_ 99, 533–543
6. Changela, A. H., Ho, C. K., Martins, A., Shuman, S., and Mondragon, A. (2001) _Proc. Natl. Acad. Sci. U. S. A._ 98, 10011–10015
7. Pei, Y., Ho, C. K., Schwer, B., and Shuman, S. (1999) _J. Biol. Chem._ 274, 29865–29874
8. Pei, Y., Lehman, K., Tian, L., and Shuman, S. (2000) _Nucleic Acids Res._ 28, 1885–1892
9. Ho, C. K., Martins, A., and Shuman, S. (2000) _J. Virol._ 74, 5486–5494
10. Ho, C. K., Gong, C., and Shuman, S. (2001) _J. Virol._ 75, 1744–1750
11. Pei, Y., Schwer, B., Hausmann, S., and Shuman, S. (2001) _Nucleic Acids Res._ 29, 387–396
12. Bisailon, M., and Shuman, S. (2001) _J. Biol. Chem._ 276, 17261–17266
13. Vivares, C. P., and Métenier, G. (2001) _BioEssays_ 23, 194–202
14. Katinka, M. D., Duprat, S., Cornillot, E., Métenier, G., Thormarat, F., Prensié, G., Barbe, V., Peyretailade, E., Broitier, P., Winsker, P., Defiac, F., El Alsouli, H., Peyret, F., Saurin, W., Gouy, M., Weissenbach, J., and Vivares, C. P. (2001) _Nature_ 414, 450–453
15. Shuman, S. (1997) _Proc. Natl. Acad. Sci. U. S. A._ 94, 12758–12760
16. Lehman, K., Schwer, B., Ho, C. K., Rouzankina, I., and Shuman, S. (1999) _J. Biol. Chem._ 274, 22668–22678
17. Schwer, B., Lehman, K., Saha, N., and Shuman, S. (2001) _J. Biol. Chem._ 276, 1857–1864
18. Lehman, K., Ho, C. K., and Shuman, S. (2001) _J. Biol. Chem._ 276, 14996–15002
19. Häkkanson, K., Doherty, A. J., Shuman, S., and Wigley, D. B. (1997) _Cell_ 89, 535–553
20. Wang, S. P., Deng, L., Ho, C. K., and Shuman, S. (1997) _Proc. Natl. Acad. Sci. U. S. A._ 94, 9573–9578
21. Mossessova, E., and Lima, C. D. (2000) _Mol. Cell_ 5, 865–876
22. Shuman, S., Liu, Y., and Schwer, B. (1994) _Proc. Natl. Acad. Sci. U. S. A._ 91, 12046–12050
23. Saha, N., Schwer, B., and Shuman, S. (1999) _J. Biol. Chem._ 274, 16553–16562
24. Mao, X., Schwer, B., and Shuman, S. (1998) _Mol. Cell._ 2, 475–480
25. Wang, S. P., and Shuman, S. (1997) _J. Biol. Chem._ 272, 14683–14689
26. Schwer, B., Saha, N., Mao, X., Chen, H. W., and Shuman, S. (2000) _Genetics_ 155, 1561–1576
27. Bugajski, J. M., Feder, M., Radlinska, M., and Rychlewski, L. (2001) _BMC Bioinformatics_ 2, 2
28. Métenier, G., and Vivares, C. P. (2001) _Microbes Infect._ 3, 407–415
29. Douglas, S., Zauer, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T., Wu, X., Behin, M., Cavalieri-Smith, T., and Maier, U. (2001) _Nature_ 410, 1091–1096
30. Hausmann, S., Ho, C. K., Schwer, B., and Shuman, S. (2001) _J. Biol. Chem._ 276, 36116–36124
31. Hirt, R. P., Logsdon, J. D., Dorey, M. W., Doolittle, W. F., and Embley, T. M. (1999) _Proc. Natl. Acad. Sci. U. S. A._ 96, 580–585
32. Van de Peer, Y., Ben Ali, A., and Meyer, A. (2000) _Gene (Amst.)_ 246, 1–8
Characterization of the mRNA Capping Apparatus of the Microsporidian Parasite
Encephalitozoon cuniculi
Stéphane Hausmann, Christian P. Vivarès and Stewart Shuman

J. Biol. Chem. 2002, 277:96-103.
doi: 10.1074/jbc.M109649200 originally published online October 30, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M109649200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 32 references, 20 of which can be accessed free at
http://www.jbc.org/content/277/1/96.full.html#ref-list-1