The Caenorhabditis elegans mRNA 5'-capping enzyme: In vitro and in vivo characterization

Toshimitsu Takagi**, Amy K. Walker***, Chika Sawa**, Felix Diehn**, Yasutaka Takase**, T. Keith Blackwell***, and Stephen Buratowski**

From the **Department of Biological Chemistry and Molecular Pharmacology and ***Center of Blood Research and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

Running title: C. elegans mRNA capping enzyme

*This work was supported by National Institutes of Health grants GM56663 to S.B. and GM62891 to T.K.B.

**Senior Postdoctoral Fellow of the American Cancer Society, Massachusetts Division. Present address: Department of Automated Biotechnology, Merck Research Laboratories, North Wales, PA 19454, USA

***Postdoctoral Fellow For Research Abroad of the Japan Society for the Promotion of Science

****Present address: Mayo Clinic, Rochester, MN 55901, USA

*****Present address: Laboratory of Seeds Finding Technology, Eisai Co., Ltd., Ibaraki 300-2635, Japan

******A Scholar of the Leukemia And Lymphoma Society. To whom correspondence should be addressed: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Ave., Boston, MA 02115, USA; Tel.: 617-432-0696; Fax: 617-738-0516; E-mail: steveb@hms.harvard.edu
The abbreviations used are: RTPase, RNA 5'-triphosphatase; GTase, GTP::mRNA guanylyltransferase; PTP, protein tyrosine phosphatase; RNAi, RNA interference; PCR, polymerase chain reaction; NTPase, nucleotide phosphohydrolase; 5-FOA, 5-fluoroorotic acid; pol II, RNA polymerase II; CTD, carboxy-terminal domain of the largest subunit of RNA polymerase II; CTD-P, phosphorylated CTD; HA, Influenza virus hemagglutinin; TLC, thin-layer chromatography; PEI, polyethyleneimine; PMSF, phenylmethylsulfonyl fluoride; CIP, calf intestine alkaline phosphatase; *C. elegans, Caenorhabditis elegans*; EST, Expressed Sequence Tag; *S. cerevisiae, Saccharomyces cerevisiae; S. pombe, Schizosaccharomyces pombe; C. albicans, Candida albicans*; ORF, open reading frame; P-TEFb, positive transcription elongation factor b; CDK, cyclin dependent kinase; DAPI, 4', 6'-diamidino-2-phenylindole hydrochloride; DIC, differential interference microscopy; SR protein, serine/arginine rich protein; CBP, C/EBP Binding Protein.

\(^1\) T. Takagi, unpublished observation

\(^2\) T. Takagi, unpublished observation

\(^3\) T. Takagi, unpublished observation
Abstract

Eukaryotic mRNA capping enzymes are bifunctional, carrying both RNA triphosphatase (RTPase) and guanylyltransferase (GTase) activities. The Caenorhabditis elegans CEL-1 capping enzyme consists of an amino (N)-terminal region with RTPase activity and a carboxy (C)-terminal region that resembles known GTases. However, CEL-1 has not previously been shown to have GTase activity. Cloning of the CEL-1 cDNA shows that the full-length protein has 623 amino acids, including an additional 38 residues at the C-termini and 12 residues at the N-termini not originally predicted from the genomic sequence. Full-length CEL-1 has RTPase and GTase activities, and the cDNA can functionally replace the capping enzyme genes in Saccharomyces cerevisiae. The CEL-1 RTPase domain is related by sequence to protein tyrosine phosphatases (PTP), so mutagenesis of residues predicted to be important for RTPase activity were carried out. CEL-1 uses a mechanism similar to PTPs, except that there was not an absolute requirement for a conserved acidic residue that acts as a proton donor. CEL-1 shows a strong preference for RNA substrates of at least three nucleotides in length. RNA-mediated interference in C. elegans embryos shows that lack of CEL-1 causes development to arrest with a phenotype similar to that seen when RNA polymerase II elongation activity is disrupted. Therefore, capping is essential for gene expression in metazoans.
Introduction

Most eukaryotic and viral mRNAs are modified at their 5’ end by a “cap” structure which consists of a 7-methylguanosine moiety attached to the 5’ terminus via a 5’-5’ linkage (1). Three sequential enzymatic activities are required to form the "cap 0" structure, m7GpppN-. First, an RNA 5’-triphosphatase (RTPase) removes the gamma-phosphate from the 5’ end of the RNA substrate to leave a diphosphate end. Next, a GTP::mRNA guanylyltransferase (GTase) catalyzes transfer of GMP from GTP, resulting in a 5’-5’ linkage: GpppNp-. These two activities are typically associated and copurify as mRNA capping enzyme. A third protein, RNA (guanine-7-)-methyltransferase, adds a methyl group to the N-7 position of the guanine cap (1,2).

Previously we characterized a putative capping enzyme gene, which we named CEL-1, that emerged from the Caenorhabditis elegans genome sequencing project (3). The open reading frame (ORF) of this gene originally predicted by the Nematode Sequencing Project was 573 amino acids. The C-terminal 340 amino acids exhibit very strong similarity to yeast and viral GTases. The N-terminal region has significant sequence similarity to the protein tyrosine phosphatase (PTP) family, including the active site consensus motif (I/V)HCxxGxxR(S/T)G (4-7). We proved that the isolated N-terminal region (residues 1-236) of CEL-1 has RTPase activity (3, 8). However, we were unable to demonstrate that the C-terminal region had GTase activity.

The ORF used in the earlier study was based on predictions of exons within genomic sequence. Since then, the C. elegans Expressed Sequence Tag (EST) database has produced several cDNA sequences predicted to encode a longer form of CEL-1 that has additional residues at both the N- and C-termini. Protein produced from the longer ORF fully substitutes for the Saccharomyces cerevisiae GTase and RTPase in vivo. The longer CEL-1 C-terminal domain has GTase activity in vitro. We further characterized the isolated RTPase domain both in vivo and in vitro. We analyzed its catalytic properties, including the effect of RNA chain length on the activity. Mutagenesis confirms a mechanistically conserved role for key residues found in both the RTPase and PTPs. Surprisingly, the CEL-1 RTPase did not require linkage to the GTase for targeting to pre-mRNA in S. cerevisiae. Finally, we used RNA-mediated interference (RNAi) to demonstrate that CEL-1 is essential in vivo for development of the C. elegans.
EXPERIMENTAL PROCEDURES

DNA cloning methods – Supplementary tables listing oligonucleotides and plasmids used in this study are available at our lab website (http://tfiib.med.harvard.edu/publications.htm). Polymerase chain reactions (PCR) for construction of plasmids and site-directed mutagenesis were carried out with Vent DNA polymerase (New England BioLabs).

Cloning of CEL-1 cDNA – To obtain a full-length cDNA for CEL-1, a two-step Mega-Primer PCR was performed with two EST clones, yk786d02 and yk798b08 (supplied by Dr. Y. Kohara, National Institute of Genetics, Mishima, Japan). In the first reaction, a 1.2-kb fragment was amplified from the 5’ region of CEL-1 using CEL-1upstreamA and Cel1-4 as primers and yk786d02 as template. The product was purified and used in the second PCR reaction as a megaprimemer. The second reaction used yk798b08 as template and used 3’CEL-1longer and CEL-1upstreamA as primers for secondary amplification. A 1.8-kb product was subcloned into pCR-Script SK(+) (Stratagene) to produce pBS-CEL-1. To create CEL-1(1-585), 3’CEL-1longer was replaced with Celeg. CE-C.

Genetic manipulation of Saccharomyces cerevisiae - S. cerevisiae strains used in this study were YSB244 (MATa ura3-52 leu2-3,112 his3Δ200 cet1Δ1::His3 {pRS316-CEG1}) (9), YSB533 (MATa ura3-52 leu2Δ1 trp1Δ63 his3Δ200 lys2Δ202 cet1Δ1::TRP1 {pRS316-CET1}) (10) and YSB719 (MATα ura3-52 leu2Δ1 trp1Δ63 his3Δ200 lys2Δ202 cet1Δ1::TRP1 ceg1Δ3::LYS2 {pRS316-CEG1-CET1}) (11). We introduced plasmids into these strains using a modified lithium acetate transformation protocol (12). Media preparation, plasmid shuffling with 5-fluoroorotic acid (5-FOA), and other yeast manipulations were carried out by standard methods (13, 14).

Yeast whole-cell extract preparation and protein analysis – Whole-cell extracts were prepared from S. cerevisiae. Immunoprecipitation and subsequent enzyme-GMP formation assay were as previously described (11).

Site-directed mutagenesis - Site-directed mutagenesis were carried out using either single-stranded phagemid (15, 16) or PCR (17). In both strategies, the plasmid pBS-CEL-1(13-
(3, listed as pBS-CEL-1 therein) was used as template. Mutations were verified by dideoxy-DNA sequencing.

To change Arg142 to Lys (R142K) or Ala (R142A), Asp76 to Asn (D76N), and Glu111 to Gln (E111Q), phagemid mutagenesis was performed with mutagenic oligonucleotides CEL1-R142K, CEL1-R142A, CEL1-D76N or CEL1-E111Q. The resulting plasmids, pBS-CEL-1 R142K, pBS-CEL-1 R142A, and pBS-CEL-1 D76N and pBS-CEL-1 E111Q, served as template for subsequent PCR reactions. Except for D76N, we amplified 0.7-kb fragments corresponding to RTPase domain containing those mutations using Celeg. CE-B and CEL-1T222stop as primers. For the wild-type and the active site cysteine mutants, we used pBS-CEL-1(13-585), pET-his7CEL-1(13-248) C136S, and pET- his7CEL-1(13-248)C136A as templates for the same PCR reaction. Each product was subcloned into pCR-Script SK(+) to generate pBS-CEL-1(13-221) version 2, pBS-CEL-1(13-221)C136S version 2, pBS-CEL-1(13-221)C136A version 2, pBS-CEL-1(13-221)R142K version 2, pBS-CEL-1(13-221)R142A version 2, and pBS-CEL-1(13-221)E111Q version 2, respectively. We further subcloned inserts from these plasmids into yeast expression vector pAD5 (11) and bacterial expression vector pSBET-his7 (18). For D76N, a 0.7-kb fragment was amplified with pBS-CEL-D76N as template and Celeg. CE-B and CEL1(3’Sac) as primers. The product was digested with Nco I and Sac I and subcloned into pAD5.

PCR-mediated site-directed mutagenesis was used to change Asp112 to Asn (D112N). In the first reaction, a 0.3-kb fragment was amplified with CEL-1D112N and CEL-1T222stop. In the second reaction, primers were Celeg. CE-B and the product from the first reaction as a megaprimer. The product was subcloned onto pCR-Script SK(+) to produce pBS-CEL-1(13-221)D112N.

Recombinant protein production and purification – Mce full-length protein and the RTPase domains of Mce and CEL-1 were expressed using a T7 promoter/polymerase system (19). E. coli strain BL21(DE3) was transformed with the appropriate expression plasmids and cultured in 500-ml media at 37 °C to an OD600= 0.5. The proteins were induced as described (18). All further operations were at 0-4 °C. Lysate was prepared by sonication in buffer B [50 mM Tris-HCl, pH 7.9, 10 % (v/v) glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF)] with 300 mM KCl and 0.5 %
NP-40. After incubating soluble extracts (100,000 x g supernatant fraction) with 2 ml of Ni$^{2+}$-NTA-agarose resin (Qiagen) for 2 hours on a rotator, the resin was poured into a column (1.5 x 2.5 cm) and extensively washed with buffer B with 20 mM KCl and 25 mM imidazole. Bound proteins were eluted with buffer B containing 20 mM KCl and 600 mM imidazole, immediately supplemented with 1 mM EDTA and 1 mM DTT. Proteins were further purified by chromatography with a column (1.2 x 9.0 cm) of Heparin-Sepharose CL-6B (Pharmacia). After washing the resin with 20 mM KCl in buffer C [20 mM Tris-HCl, pH 7.9, 1 mM EDTA, 1 mM DTT, 10 % (v/v) glycerol, 1 mM PMSF], proteins were eluted with a 100 ml linear gradient of 20 to 300 mM KCl in buffer B. Purified proteins were visualized by SDS-PAGE and Coomassie Brilliant Blue staining and by immunoblot analysis with monoclonal anti-polyhistidine antibody (anti 6xHis, Clontech).

**Preparation of substrate for RTPase assay** – [$\gamma$-$^{32}$P] and [$\alpha$-$^{32}$P]ATP-terminated oligoribonucleotides for the RTPase assay were synthesized with the DNA primase protein of bacteriophage T7 (8, 20). The standard reaction (100 µl) contained 40 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 10 mM DTT, 50 µg/ml BSA, 50 mM potassium glutamate, 2 mM dTTP, 2 mM CTP, 0.3 mM [$\gamma$-$^{32}$P] or [$\alpha$-$^{32}$P]ATP (500-1,000 cpm/pmol) (NEN/DuPont), 1 nM synthetic oligonucleotide template, and 1 µM (hexamer) T7 primase (21). The sequences of the oligonucleotides used are: for the diribonucleotide pppApC: 5'- (C)$_3$GTC(T)$_{25}$-3'; for the trinucleotide pppApCpC: 5'- (C)$_4$GGTC(T)$_{25}$-3'; for the tetraribonucleotide pppApCpCpC: 5'- (C)$_5$GGGGTC(T)$_{25}$-3'; for the pentaribonucleotide pppApCpCpCpC: 5'- (C)$_6$GGGGGTCT(T)$_{25}$-3'. The reaction was incubated at 37 °C overnight. After extraction with phenol-chloroform (1:1), RNAs were precipitated with ethanol and dissolved in 500 µl of buffer A [20 mM Tris-HCl, pH 7.9, 7 M urea] containing 100 mM NaCl. RNAs were further purified by chromatography with a column (1.0 x 27.0 cm) of DEAE-Sephadex A-25 (Pharmacia) pre-equilibrated with buffer A with 100 mM NaCl. The column was washed with 300 ml of buffer A with 120 mM NaCl, and RNAs were eluted with a 700 ml linear gradient of 150 to 350 mM NaCl in buffer A. RNAs were analyzed with electrophoresis on a 36 % polyacrylamide-3M urea gel and autoradiography. Samples were then pooled and dialyzed against 10 mM Tris-HCl (pH 8.0)/ 1 mM EDTA with Spectra/Por 7.
The RNAs were then lyophilized and dissolved in water. Previously, we used triethylamine-bicarbonate (TEA-HCO$_3^-$) buffer for the DEAE-Sephadex A-25 column chromatography (8). However, we observed that residual amounts of TEA-HCO$_3^-$ contaminating the purified substrate inhibit RTPase activity, so the protocol was modified as indicated.

[$\gamma^3P$]GTP-terminated RNA was synthesized by in vitro transcription of linearized plasmid template (3) using recombinant polyhistidine-tagged T7 RNA polymerase (22).

Enzymatic assays—GTase, RTPase, and nucleotide phosphohydrolase (NTPase) activities were assayed as previously described (3, 8, 18).

RNAi analysis—In vitro synthesized double stranded (ds)RNA (Megascript, Ambion) was produced using pBS-CEL-1(13-221) and pBS-CEL1(1-551) and injected at 1.0 µg/µl into young adults (2-8 fertilized embryos). Equivalent results were obtained with cel-1 RNAs covering the RTPase or GTase domains. ama-1(RNAi) and cel-1(RNAi) embryos were analyzed at 24 and 36 hours post injection, respectively, when uniform populations of arrested embryos appeared. For immunostaining, embryos were collected from dissected hermaphrodites 36 hours after injection. Because most analyses were performed before terminal embryonic developmental arrest, RNAi effectiveness was confirmed by monitoring sibling embryos that were allowed to develop.

Immunostaining—for α-CDK-9 (23, 24) and α-Pol II (25) staining, embryos were subjected to 2% paraformaldehyde fixation and freeze cracked before treating with methanol. Washes and antibody incubations were performed in PBT (1X phosphate buffered saline, 1% Triton X-100, 1% BSA) prior to staining. For α-Pol II staining (25), embryos were subjected to 2% paraformaldehyde fixation and freeze cracked before treating with methanol. Washes and antibody incubations were performed in PBT (1X phosphate buffered saline, 1% Triton X-100, 1% BSA) prior to staining. Staining with P-CTD (α-PSer5) (26) and H5 (α-PSer2) (Babco) was performed as in Walker, et al., 2001 (27). Images were captured with a Zeiss AxioSKOP2 microscope and AxioCam digital camera, and antibody staining intensities were compared over a range of exposure times. Pixel intensities were standardized using Adobe Photoshop 6.0.
RESULTS

Detection of GTase activity in C. elegans nuclear extract – To determine the gel mobility of the native capping enzyme, we incubated C. elegans nuclear extract with [α-32P]GTP to form the covalent GTase-[32P]GMP intermediate (Fig. 1). The major labeled band was about 70 kDa (lane 2). Complex formation was dependent on the presence of divalent cation (lanes 3 and 4), and specific to guanine nucleotide (lanes 5 and 6). The reaction rapidly reached completion within a short period of time at 0°C (lane 7). Therefore, the C. elegans capping enzyme appears to be a single protein of approximately 70 kDa.

Identification of an extended C. elegans capping enzyme open reading frame – Previously, we (3) and others (28) found that the predicted gene C03D6 (GenBank accession no. Z75525) from the C. elegans genome sequencing project had significant similarity to the yeast GTase. Similar capping enzyme genes from other metazoans were described (29-33). All of these proteins contain motifs found in the GTase proteins/domains of yeast and virus (2, 28, 34-36; Fig. 2A) as well as an N-terminal domain related to the PTPs.

The protein encoded by ORF for CEL-1 is somewhat shorter at the C-terminus than its homologues from other species (Fig. 2A). Using the predictions of exon structure, we amplified a CEL-1 ORF from C. elegans cDNA. Although we could demonstrate RTPase activity of the N-terminal domain, we were not able to show GTase activity of the C-terminal domain. Therefore, we searched the EST database (http://www.ncbi.nlm.nih.gov/BLAST/) for naturally occurring CEL-1 cDNAs and found SP9F10 (accession no. BE228078). This cDNA encodes an ORF containing the previously predicted amino acids 534-573 of CEL-1, but which has an extra 38 residues at the C-terminus due to a splicing event that was not predicted from the genomic sequence. The C. elegans EST database server at the DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_BLAST.html) contains clone yk786d02 containing sequences for the 5’ region of CEL-1. This encodes the previously identified RTPase domain (3, 8; designated residues 1-236 in those papers). However, there is an in-frame initiation codon 36-base upstream of the one previously believed to be the translation start site (3; Fig. 2B).
Combining this new cDNA information, full-length CEL-1 is predicted to have an additional 12 residues at the N-termini and 38 residues at the C-termini, for a total of 623 amino acids. The predicted molecular weight is 72 kDa, in good agreement with the size of the GTase detected in worm extract (Fig. 1). We therefore have renumbered the CEL-1 amino acids, and the previously analyzed shorter protein (3, 8) will herein be referred to as CEL-1(13-585).

**CEL-1 is a bifunctional capping enzyme with both RTPase and GTase activities** - Unlike the metazoan enzyme, capping enzyme in the yeast *S. cerevisiae* is a complex of RTPase and GTase subunits (37). These polypeptides are encoded by the *CET1* and *CEG1* genes, respectively, both of which are essential for cell viability (38, 39). Ceg1 is related by sequence to the viral and metazoan GTases (2, 36). In contrast, Cet1 is not related to PTPs or metazoan RTPase domains (40).

We tested whether CEL-1 can function in place of *CEG1* and *CET1* in *S. cerevisiae* (Fig. 3A). The *ceg1Δcet1Δ* strain YSB719 (11) was transformed with a high copy plasmid expressing HA epitope-tagged CEL-1 from the constitutive *ADH1* promoter. An expression construct for the mouse capping enzyme (Mce) was used as a positive control. After shuffling out the *CEG1/CET1* plasmid with 5-FOA, growth was observed with cells expressing either Mce or CEL-1. In contrast, expression of CEL-1(1-585) did not rescue cells. Therefore, CEL-1 with the extended C-terminus, but not the shorter form previously analyzed, has both capping enzyme RTPase and GTase activities.

A truncation mutant of Mce containing only the GTase domain (residues 211-597) can replace *CEG1* in *S. cerevisiae* (41). The corresponding domain of CEL-1 (residues 221-623, see Fig. 2A) supports viability (Fig. 3B) of the *ceg1Δ* strain YSB244 (9). Whole-cell extracts from yeast expressing CEL-1 were assayed for capping enzyme-GMP complex (Fig. 3C). A protein of about 70 kDa was detected (left panel, lane 2), the same size as that of the complex detected in *C. elegans* nuclear extract (Fig. 1). We also assayed extracts from yeast expressing either epitope-tagged CEL-1(222-623) or CEL-1(222-585) for the presence of protein by immunoblot using anti-HA monoclonal antibody 12CA5 (42) and for enzyme-GMP complex formation. Both proteins were detected in the immunoblot (data not shown). In contrast, GTase activity was
detectable with CEL-1(222-623) but not CEL-1(222-585) (Fig. 3C, right panel, lane 3 and data not shown). These results again show that the residues 586-623 are important for CEL-1 GTase activity.

_The CEL-1 RTPase domain can function independently of the GTase domain in vivo_

We previously showed that CEL-1(13-248) has RTPase activity _in vitro_ (3, 8). Although full-length CEL-1 can simultaneously replace yeast Ceg1 and Cet1, we tested whether over-expression of CEL-1 derivatives could rescue _cet1Δ_ strain YSB533 (10). Neither CEL-1 derivatives 1-221, 13-221, nor 1-585 could replace _CET1_ (Fig. 4B and data not shown). However, this was not surprising because GTase activity was being supplied by Ceg1. The GTase associates with the phosphorylated C-terminal domain of the largest subunit of pol II (CTD-P) (29, 30, 41, 43). Ceg1 by itself is inactive on CTD-P unless it is interacting with the central region (aa. 235-265) of Cet1 (10, 11). In contrast, the GTase domain of Mce (Mce(211-597)) is activated by binding to CTD-P (44). The GTases from other fungi, _Schizosaccharomyces pombe_ (pce1) and _Candida albicans_ (Cgt1), do not require RTPase activation (45, 46).

To avoid the complications of Ceg1 dependence upon Cet1 interaction, the CEL-1 derivatives were assayed in cells expressing GTases from other organisms. In _cег1_ _Δcет1Δ_ cells expressing pce1, Cgt1, or Mce(211-597), the CEL-1 fragments 1-221, 13-221, or 1-585 could support viability (Fig. 4 and data not shown). When over-expressed in _Δcег1_ _Δcет1_ strain in combination with Mce(211-597), cells containing Cet1 formed colonies after one day, while cells containing either CEL-1(13-221) or Mce(1-210) formed colonies after 3 days (Fig. 4A). Both _S.pombe_ pce1 and _C.albicans_ Cgt1 GTases could also combine with the CEL-1 or Mce RTPase domains to support viability, but the
Chlorella virus GTase A103R (34) could not (Fig. 4B). A103R could not complement a ceg1Δ strain either, even though enzyme-GMP complex was detectable in lysates.

To determine whether the mouse and C. elegans RTPase domains were interacting with the GTases, or instead were independently acting at mRNA 5' ends, immunoprecipitation experiments were carried out. Epitope-tagged Cet1, CEL-1(13-221) or Mce(1-210) were co-expressed with either pce1 or Cgt1. The RTPases were immunoprecipitated with 12CA5 (Fig. 4C). The top panel shows that each of the RTPases was expressed and immunoprecipitated efficiently. The precipitate was assayed for guanylylation with [α-32P]GTP to detect any GTase (Fig. 4C, middle panel). Lane 2 shows HA-Cet1 coprecipitates the C. albicans GTase Cgt1. In contrast, lane 3 shows that HA-Cet1 does not interact with the S. pombe GTase pce1. These findings confirm our earlier observation that pce1 functions without any interaction with RTPase protein (46). Under the same conditions, neither pce1 (lanes 4 and 6) nor Cgt1 (lanes 5 and 7) are coprecipitated with the RTPases CEL-1(13-221) or Mce(1-210). Therefore, we conclude that the overexpressed metazoan RTPase domains can function in vivo without linkage to the GTase domain.

Presumably, CEL-1(13-221) and Mce(1-210) do not support growth of a cet1Δ strain because they do not bind and activate Ceg1. To confirm this, regions of Cet1 were fused to the RTPase domain of CEL-1(13-221). The chimeras were co-expressed along with Ceg1 over-expressed from a 2µm plasmid in a cet1Δceg1Δ strain (Fig. 5). Cet1(1-225)-CEL-1(13-221) cannot functionally replace Cet1. In contrast, Cet1(1-265)-CEL-1(13-221), which contains the Ceg1-interaction region (aa. 235-265), can support viability. When amino acids 1-265 are derived from the mutant cet1-446 (P245A W247A), which is disrupted for the ability to interact with Ceg1 (11), the ability to replace Cet1 was disrupted. All of the chimeric RTPases support viability of the same strain if Mce(211-597) replaces Ceg1 as the GTase (Fig. 5, right), showing that these proteins are functional. These data support our earlier conclusion that residues 235-265 of Cet1 are primarily required for proper function of Ceg1.

Mutation analysis of CEL-1 RTPase domain -
Members of the PTP superfamily, including the metazoan capping enzyme RTPases, have an active site consensus sequence of \((I/V)HCxxGxxR(S/T)G\). The nucleophilic cysteine attacks the phosphate. The arginine residue contributes to transition-state stabilization via the formation of hydrogen bonds with two oxygens of the phosphate (47). Outside the active site motif, a conserved aspartic acid residue serves to stabilize the leaving group (48-50). In the PTPs, this acidic residue is believed to act as a general acid, donating a proton to the leaving group oxygen of the substrate’s tyrosine residue.

In order to examine the degree of mechanistic conservation between the PTPs and RTPases, we analyzed CEL-1 derivatives mutated at conserved residues important for the PTP mechanism. Arginine142 in the consensus motif was mutated. Also, Aspartate 76, Glutamate 111, and Aspartate 112 were also mutated because they were candidates for the proton-donating acidic residue. Arg142 and Asp76 are conserved in all of the PTP-like RTPases, while Glu111 and Asp112 are not (Fig. 2B). The histidine-tagged mutants C136S, R142K, D76N, E111Q, and D112N were purified from \textit{E. coli} (Fig. 6A). Their RTPase activities were tested with \(\gamma^{32}\text{P}\)GTP-terminated RNA (Fig. 6B). C136S and R142K, mutated in key active site residues, were inactive. The activities of E111Q, D112N and D76N were about 20\%, 50\% and 10\%, respectively, of that of the wild-type protein.

We also tested if these mutants can support viability of a \textit{ceg1 cef1} strain that was also expressing Mce(211-597) (Fig. 6C). CEL-1 mutants C136S, C136A, R142K, or R142A could not support viability, whereas E111Q or D112N grew as well as the wild-type strain. The \textit{in vivo} phenotypes correlated well with \textit{in vitro} results. D76N, which had only 10\% of wild-type activity \textit{in vitro}, did not support viability. Immunoblotting of whole-cell extracts confirmed that all the mutants were expressed, although some variability in levels was observed (Fig. 6D). The differences in ability to support cell growth did not correlate with protein expression, since the non-functional C136S, R142K, and R142A mutants were expressed at greater levels than the functional E111Q and D112N proteins.

Residues Glu111 and Asp112 are not highly conserved and do not seem to be vital \textit{in vivo} (Fig. 6C). In contrast, Asp76 is conserved and a D76N mutant can not support viability in
yeast. However, the D76N mutant still has partial activity in vitro, indicating that the carboxylate side chain is not absolutely required. We speculate that the reduced activity of D76N is not sufficient to rescue cells in the heterologous yeast system. Mutation of the equivalent residue in Mce or BVP (Asp66 of Mce and Asp60 of BVP) only slightly diminished the activity (51-53). Therefore, in contrast to the PTPs, general acid catalysis may not be essential for the mechanism of the RNA phosphatases.

**Effect of chain length of RNA on RTPase domain of metazoan capping enzyme**

The DNA primase of bacteriophage T7 uses a DNA template to make short RNA primers (2-10 nucleotides) that begin with the sequence pppApC (21). Using this system, substrates of various sizes were prepared for RTPase assays. We previously reported that CEL-1(13-248) efficiently uses a trinucleotide substrate (pppApCpC), but not a mononucleotide (pppA) (8). Here, we prepared dimer (pppApC), trimer (pppApCpC), tetramer (pppApCpCpC), and pentamer (pppApCpCpCpC) RNAs (Fig. 7A). First, the activity of CEL-1(13-221) was tested with dinucleotide labeled at either the gamma (p**pp**ApC; bold denotes the position of the radioactive phosphate) or alpha (pp**p**ApC) positions (Fig. 7B). CEL-1(13-221) releases the terminal phosphate to leave a diphosphate end. This activity was not seen with CEL-1(13-221)C136S, mutated at the active site cysteine. CEL-1(13-221) did not have detectable NTPase activity under these conditions (Fig. 7C).

The pH optimum of the RTPase reaction was about pH 8.0, and the reaction was severely inhibited below pH 7.0 (data not shown). Sodium vanadate is an inhibitor of PTPs that acts as a transition-state mimic (54). Vanadate also inhibited CEL-1(13-221), with 60% inhibition observed at 1 µM (data not shown). Like PTPs, CEL-1 RTPase is independent of, and in fact inhibited by, divalent cations (data not shown). Similar inhibition was reported for Mce, PIR1, and BVP (51, 55).

Next, we tested the effect of RNA chain length (Fig. 8). CEL-1(13-221) hydrolyzes the beta-gamma phosphodiester bond of trinucleotide more efficiently than that of dinucleotide. Little difference was seen between tri-, tetra-, and pentanucleotides. With a double-reciprocal plot, the $k_{cat}/K_m$ values with ATP, dinucleotide, trinucleotide, and tetranucleotide were calculated to be 5.5
x 10, 0.5 x 10^5, 4.6 x 10^5, 5.1 x 10^5 M^-1 s^-1, respectively. The same length dependence was observed with full-length Mce and Mce(1-210) (data not shown), indicating that two phosphodiester bonds are necessary for optimal fit of the RNA substrate into the active site of metazoan capping enzyme RTPases.

**CEL-1 is essential for embryonic development and CTD Ser 2 phosphorylation** - To investigate the requirement for **CEL-1 in vivo**, we used RNA interference (RNAi) to inhibit its expression during embryogenesis (56). The early embryo provides an advantageous system for analyzing functions of essential transcription or mRNA processing machinery components *in vivo*. The initial stages of *C. elegans* development are orchestrated by maternally derived proteins and mRNAs, making it possible for embryos to survive until approximately the 100-cell stage when new mRNA transcription is prevented (27, 57).

**cel-1(RNAi)** embryos arrested development after forming approximately 100 cells that lacked any signs of differentiation (Fig. 9A). This terminal arrest phenotype is very similar to that observed when the pol II large subunit or various other broadly essential mRNA transcription factors are inhibited by RNAi (23, 24, 27, 57). However, early cell division timing and cleavage planes were normal in **cel-1(RNAi)** embryos, suggesting that these embryos contained appropriate maternal mRNA stores (not shown). One abnormality was the cell cycle period of the endodermal precursor cells Ea and Ep, which was shortened compared to wild type. This particular cell cycle abnormality characteristically occurs in response to broad defects in early embryonic transcription, including mutation or RNAi knockdown of the *C. elegans* orthologs of the transcription elongation factor genes *spt5* and *spt6* (23, 24, 27, 57). Together, the data suggest that lack of **cel-1** activity may significantly impair new embryonic mRNA production.

To further characterize how the process of mRNA production was affected in **cel-1(RNAi)** embryos, phosphorylation of RNA polymerase II was analyzed. The C-terminal domain (CTD) of the pol II large subunit consists of repeats based in the consensus sequence YSPTSPS (42 copies in *C. elegans*). The CTD interacts with mRNA processing factors, linking them to the transcribing polymerase (58, 59). Near promoters, the CTD repeat is primarily phosphorylated on Serine 5 by the TFIIH kinase, recruiting mRNA capping enzyme (26, 60, 61). As Pol II moves
away from the promoter, the CTD phosphorylation shifts primarily to Serine 2 (60). During metazoan transcription, CTD Serine 2 is phosphorylated primarily by the kinase P-TEFB (CDK-9/Cyclin T) (23, 62). CTD Ser 5 and Ser 2 phosphorylation can be specifically detected in embryonic nuclei by staining with the P-CTD and H5 antibodies, respectively (26, 27, 63), which we refer to as α-PSer5 and α-PSer2 for clarity (Fig. 9C and D).

In the early *C. elegans* embryo, the appearance of both α-PSer5 and α-PSer2 staining depends upon transcription. Staining with α-PSer5 and α-PSer2 is not detected in embryonic nuclei until the three-to-four cell stage, when new mRNA transcription begins (63). At later stages, the patterns and intensity of this staining closely parallel transcription activity in embryonic cells. For example, both types of staining are eliminated or reduced in tandem by RNAi depletion of transcription initiation factors such as TFIIB (*ttb-1*) (24, 27). In contrast, when the elongation factor CDK-9 is depleted by RNAi, Ser 5 phosphorylation levels appear normal but Ser 2 phosphorylation is undetectable (23).

In *cel-1(RNAi)* embryos, total levels of the Pol II large subunit AMA-1 are unaffected (Fig. 9B), but CTD phosphorylation was highly abnormal. As in wild type embryos, in *cel-1(RNAi)* embryos Ser 5 phosphorylation was detectable as bright punctate staining pattern in somatic nuclei (Fig. 9C). In contrast, in *cel-1(RNAi)* embryos levels of specific α-PSer2 staining were dramatically reduced, to a level only slightly higher than the background observed in *ama-1(RNAi)* embryos (Fig. 9D). Levels of the CTD Ser 2 kinase CDK-9 appeared to be normal in *cel-1(RNAi)* embryonic nuclei, arguing that the drop in CTD phosphorylation was not an indirect effect (Fig. 9B). The specific and substantial defect in CTD Ser 2 phosphorylation suggests that when CEL-1 levels are depleted, the normal progression of CTD phosphorylation during transcription is disrupted at most or possibly all genes.
Discussion

Here we characterize the full-length C. elegans capping enzyme, CEL-1. Based on cDNAs in the EST databases, CEL-1 is a 623 amino acid protein with both RTPase and GTase activities. This matches the size of the enzyme-GMP intermediate detected in C. elegans nuclear extract (Fig. 1). Interestingly, multiple cDNAs for human and Xenopus capping enzymes have been described, possibly produced by alternative splicing of mRNA (31-33). These cDNA variants encode an intact N-terminal RTPase domain, but have either internal deletions or truncations in the C-terminal GTase domain. As a result, the proteins from these variants would only have RTPase activity. PCR analyses showed that these short forms are expressed, but their physiological function is unknown. To date, no cDNAs corresponding to a shortened capping enzyme have been found in the C. elegans EST database. CEL-1 was originally predicted to have 573 amino acids. CEL-1(1-585) has RTPase activity but does not complement a S. cerevisiae GTase mutant ceg1 (Fig. 3). Therefore, CEL-1 residues 586-623 are not required for protein stability or proper localization, but are essential for GTase activity.

The CEL-1 RTPase domain (3) was the founding member of a subfamily of PTP-like RNA phosphatases. This subfamily includes the capping enzyme RTPases and RNA tri- and diphosphatases whose functions are unknown (Fig. 2B). All members contain a nucleophilic cysteine necessary for activity (3, 8, 41, 64, 51, 55, 65). A phosphocysteine intermediate was detected with Mce and BVP (52, 53). Other PTP-like enzymes with substrates other than phosphotyrosine have been reported; these include the phosphoinositide phosphatase PTEN/MMAC1 and myotubularin (66), and S. cerevisiae arsenite reductase Acr2 (67).

In PTPs, the formation and hydrolysis of the phosphocysteine intermediate of PTP requires transition-state stabilization by the arginine residue within the consensus motif (4-7). Mutagenesis of Mce (51, 52), BVP (53) and CEL-1 (Fig. 6) show that this residue is essential for RNA phosphatase activity, providing further evidence that the PTP and RNA phosphatases use the same enzymatic mechanism. On the other hand, mutation of a conserved aspartic acid residue in the RTPases (Asp76 of CEL-1, Asp66 of Mce and Asp60 of BVP; see Fig. 2B) only
slightly diminishes activity (Fig. 6; 51-52). The equivalent mutations in PTPs lower activity by $10^2$ to $10^5$ fold (4-7). X-ray crystallography of Mce(1-210) shows that Asp66 is positioned differently from the essential general acid aspartate loop described for PTPs (52). Apparently, the RTPase mechanism does not conserve the function of this residue.

Both CEL-1(13-221) and Mce(1-210) can remove the gamma-phosphate from the 5'-end of a dinucleotide (Fig. 7 and data not shown). However, maximal activity is observed on substrates that are three nucleotides or longer (Fig. 8 and data not shown). The \textit{S. cerevisiae} RTPase Cet1 is unrelated to PTPs and its reaction mechanism is different from that of metazoan RTPases (40). However, Cet1 also acts on dinucleotide and trinucleotide RNAs efficiently (18). Diphosphate-ended oligonucleotides such as ppApG, ppGpC, and ppGpCpC are active as guanylyl acceptors for mammalian and yeast GTases (68-72). Structural studies on RNA polymerase II suggest that RNA exits polymerase in the vicinity of the CTD (73), where capping enzymes will be bound. Capping occurs around the time mRNAs are about 30 nucleotides in length (74, 75). Therefore, capping enzyme probably recognizes the first few phosphodiester bonds of nascent RNA that emerge from the body of pol II and immediately caps the mRNA.

RTPases and GTases are typically linked with each other, either on the same protein (metazoans) or in a complex (yeast). In \textit{S. cerevisiae}, the interaction between the GTase (Ceg1) and RTPase (Cet1) subunits is essential for cell viability. Cet1 cannot be replaced by the RTPase domains from MCE or CEL-1, presumably because these RTPases cannot interact with Ceg1. It was originally proposed that the primary role of the linkage between GTase and RTPase on a single polypeptide was speculated to guide RTPase to pol II transcription complex (43, 64). However, both CEL-1(13-221) and Mce(1-210) can support viability when Ceg1 is replaced with Mce(211-597), \textit{S. pombe} pce1, or \textit{C. albicans} Cgt1 (Fig. 4A and B). As we did not detect any tight interaction between these RTPases and GTases (Fig. 4C), we conclude that the metazoan RTPase domain can be targeted to pre-mRNA and function without any linkage to GTase. The primary function of the Cet1 interaction with Ceg1 is instead required for the activity of Ceg1 (10, 11). Other fungal and metazoan GTases do not require an interaction with RTPase for activity (45, 46).
Although we found that the link between RTPase and GTase domains is not absolutely required for the capping enzymes of metazoans or fungi other than S. cerevisiae, this does not mean that the interaction is unimportant. To substitute for Cet1 in vivo, it was necessary to overexpress the isolated metazoan RTPase domain with a strong promoter and a high copy plasmid (76 and this study). In contrast, a low copy plasmid of the full-length enzyme was sufficient for rescuing a \textit{cet}1\Delta strain (52). Transfection experiments showed that Mce(1-210) is mostly cytoplasmic in mammalian cells (51). This may also be true in \textit{S. cerevisiae}.

Overexpression may be necessary to drive sufficient amounts of RTPase into the nucleus and into proximity with the mRNA 5’ end. Alternatively, RTPases may independently bind pol II or a pol II-associated protein. HIV-1 Tat protein binds to both full-length MCE as well as the isolated GTase and RTPase domains (77). There could be a corresponding cellular protein(s) that mediates the association of RTPase domain with the pol II complex or RNA chain. Whatever mechanism is used, isolated RTPase domains function more efficiently in vivo when it is linked to a GTase domain.

Finally, we examined the requirement for CEL-1 in vivo using RNA-mediated inactivation of the gene. \textit{Cel-1(RNAi)} embryos arrest development with a phenotype that is characteristic of a broad transcription defect. A similar phenotype is seen upon RNAi knockdown of ama-1 (Pol II), \textit{ttb-1} (TFIIB), or multiple TAFs (23, 24, 27). One \textit{CEL-1(RNAi)} phenotype is strikingly different from effects seen upon depletion of basal initiation factors. In those cases, levels of CTD phosphorylation at both serine 5 and serine 2 were lowered in parallel, often reduced to undetectable levels. For example, in \textit{ttb-1(RNAi)} embryos, in which basal factor TFIIB is knocked down, both serine 5 and serine 2 phosphorylation are reduced to background (23, 27). In \textit{cel-1(RNAi)} embryos, CTD serine 5 phosphorylation appears to be relatively unaffected while serine 2 phosphorylation is dramatically reduced (Fig. 9B). The only other example of this “uncoupling” of CTD serine 5 and 2 phosphorylation occurred when we depleted either of the P-TEFb components, CDK-9, or Cyclin T (23). Levels of the CDK-9 kinase appear normal in \textit{cel-1(RNAi)} embryos, however. Because serine 5 phosphorylation occurs primarily near the promoter, the generally normal levels in \textit{cel-1(RNAi)} embryos suggest that transcription initiation
may be close to normal. The markedly decreased levels of serine 2 phosphorylation, a modification linked to elongation phase, suggests that the absence of capping enzyme interrupts the progression of transcription. It will be interesting to determine whether lack of capping enzyme decreases the efficiency with which P-TEFb or other elongation factors are recruited to transcribed genes. This would be the latest of many connections have recently emerged between transcription elongation and mRNA processing.

**Acknowledgements** - We would like to thank Drs. Aaron J. Shatkin and Fabio Piano for communicating their results prior to publication, Gary Ruvkun for supplying us *C. elegans* nuclear extract, Dale Wigley for pET-A103R, Gerhard Wagner and Kylie Walter for pT7-911Q, Takahiro Kusakabe for designing the sequences of synthetic oligonucleotides for preparation of short RNA with T7 DNA primase, and Robin Buratowski for help with the plasmid and oligo tables.
References
1. Furuichi, Y., and Shatkin, A. J. (2000) Adv. Virus Res. 55, 135-184
2. Shuman, S. (2000) Prog. Nucl. Acids Res. Mol. Biol. 66, 1-40
3. Takagi, T., Moore, C. R., Diehn, F., and Buratowski, S. (1997) Cell 89, 867-873
4. Denu, J. M., Stuckey, J. A., Saper, M. A., and Dixon, J. E. (1996) Cell 87, 361-364
5. Fauman, E. B., and Saper, M. A. (1996) Trends. Biochem. Sci. 21, 413-417
6. Barford, D., Das, A. K., Egloff, M.-P. (1998) Annu. Rev. Biophys. Biomol. Struct. 27, 133-164
7. Denu, J. M., and Dixon, J. E. (1998) Curr. Opinion Chem. Biol. 2, 633-641
8. Takagi, T, Taylor, G. S., Kusakabe, T., Charbonneau, H., and Buratowski, S. (1998) Proc. Natl. Acad. Sci. U.S.A. 95, 9808-9812
9. Fresco, L. D., and Buratowski, S. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 6624-6628
10. Cho, E.-J., Rodriguez, C.R., Takagi, T., and Buratowski, S. (1998) Genes Dev. 12, 3482-3487
11. Takase, Y., Takagi, T., Komarnitsky, P. B., and Buratowski, S. (2000) Mol. Cell. Biol. 24, 9307-9316
12. Gietz, D., St. Jean, A., Woods, R. A., and Schiestl, R. (1992) Nucleic Acids Res. 20, 1425
13. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smit, J. A., Struhl, K. eds. (1991) Current protocols in molecular biology. John Wiley and Sons, New York.
14. Guthrie, C., and Fink, G. (1991) Meth. Enzymology 194
15. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Meth. Enzymol. 154, 367-382
16. Vieira, J., and Messing, J. (1987) Meth. Enzymol. 153, 3-11
17. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Gene 77, 51-59
18. Rodriguez, C. R., Takagi, T., Cho, E.-J., and Buratowski, S. (1999) Nucleic Acid Res. 27, 2181-2188
19. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Meth. Enzymol. 185, 60-89
20. Matsuo, H., Moriguchi, T., Takagi, T., Kusakabe, T., Buratowski, S., Sekine, M., Kyogoku, Y., and Wagner, G. (2000) J. Am. Chem. Soc. 122, 2417-2421
21. Kusakabe, T., and Richardson, C. C. (1997) *J. Biol. Chem.* **272**, 5943-5951

22. Ichetovkin, I. E., Abramochkin, G., and Shrader, T. E. (1997) *J. Biol. Chem.* **272**, 33009-33014

23. Shim, E.Y., Walker, A.K., Shi, Y., and Blackwell, T.K. (2002) *Genes Dev.* **16**, 2135-2146

24. Shim, E.Y., Walker, A.K., and Blackwell, T.K. (2002) *J. Biol. Chem.* **277**, 30413-30416

25. Bellier, S., Dubois, M.F., Nishida, E., Almouzni, G., Bensaude, O. (1997) *EMBO J.* **16**, 6250-6252

26. Schroeder, S.C., Schwer, B., Shuman, S. and Bentley, D. (2000) *Genes Dev.* **14**, 2435-2440

27. Walker, A.K., Rothman, J.H., Shi, Y., and Blackwell, T.K. (2001) *EMBO J.* **20**, 5269-5279

28. Wang, S. P., Deng, L., Ho, C. K., and Shuman, S. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9573-9578

29. Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., and Shatkin, A. J. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12898-12903

30. McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Amgen EST Program, Shuman, S., and Bentley, D. L. (1997) *Genes Dev.* **11**, 3306-3318

31. Tsukamoto, T., Shibagaki, Y., Murakoshi, T., Suzuki, M., Nakamura, A., Gotoh, H., and Mizumoto, K. (1998) *Biochem. Biophys. Res. Commun.* **243**, 101-108

32. Yamada-Okabe, T., Doi, R., Shimmi, O., Arisawa, M., and Yamada-Okabe, H. (1998) *Nucleic Acid Res.* **26**, 1700-1706

33. Yokoska, J., Tsukamoto, T., Miura, K., Shiokawa, K., and Mizumoto, K. (2000) *Biochem. Biophys. Res. Commun.* **268**, 617-624

34. Hakansson, K., Doherty, A. J., Shuman, S., and Wigley, D. B. (1997) *Cell* **89**, 545-554

35. Hakansson, K., and Wigley, D. B. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1505-1510

36. Shuman, S., and Schwer, B. (1995) *Mol. Microbiol.* **17**, 405-410

37. Itoh, N., Yamada, H., Kaziro, Y., and Mizumoto, K. (1987) *J. Biol. Chem.* **262**, 1989-1995

38. Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S., and Mizumoto, K. (1992) *J. Biol. Chem.* **267**, 9521-9528
39. Tsukamoto, T., Shibagaki, Y., Imajoh-Ohmi, S., Murakoshi, T., Suzuki, M., Nakamura, A.,
    Gotoh, H., and Mizumoto, K. (1997) Biochem. Biophys. Res. Commun. 239, 116-122
40. Lima, C. D., Wang, L. K., and Shuman, S. (1999) Cell 99, 533-543
41. Ho, C. K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998) J.
    Biol. Chem. 273, 9577-9585
42. Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., Lerner, R. A., and
    Wigler, M. (1988) Mol. Cell. Biol. 8, 2159-2165
43. Cho, E.-J., Takagi, T., Moore, C. R., and Buratowski, S. (1997) Genes Dev. 11, 3319-3326
44. Ho, C. K., and Shuman, S., (1999) Mol. Cell 3, 405-411
45. Pei, Y., Hausmann, S., Ho, C. K., Schwer, B., and Shuman, S. (2001) J. Biol. Chem. 276,
    28075-28082
46. Takagi, T., Cho, E.-J., Janoo, R. T. K., Polodny, V., Takase, Y., Keogh, M.-C., Woo, S.-A.,
    Fresco-Cohen, L. D., Hoffman, C. S., and Buratowski, S. (2002) Eukaryot. Cell 1, 448-
    457
47. Zhang, Z.-Y., Wang, Y., Wu, L., Fauman, E. B., Stuckey, J. A., Schubert, H. L., Saper, M. A.,
    and Dixon, J. E. (1994) Biochemistry 33, 15266-15270
48. Zhang, Z.-Y., Wang, Y., and Dixon, J. E. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 1624-1627
49. Hengge, A. C., Sowa, G. A., Wu, L., and Zhang, Z.-Y. (1995) Biochemistry 34, 13982-13987
50. Denu, J. M., Zhou, G., Guo, Y., and Dixon, J. E. (1995) Biochemistry 34, 3396-3403
51. Wen, Y., Yue, Z., and Shatkin, A. J. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 12226-12231.
52. Changela, A., Ho, C. K., Martins, A., Shuman, S., and Mondragon, A. (2001) EMBO J. 20,
    2575-2586
53. Martins, A., and Shuman, S. (2000) J. Biol. Chem. 275, 35070-35076
54. Denu, J. M., Lohse, D. L., Vijayalakshmi, J., Saper, M. A., and Dixon, J. E. (1996) Proc. Natl.
    Acad. Sci. U. S. A. 93, 2493-2498
55. Deshpande, T., Takagi, T., Hao, L., Buratowski, S., and Charbonneau, H. (1999) J. Biol.
    Chem. 274, 16590-16594
79. Yuan, Y., Li, D.-M., and Sun, H. (1998) *J. Biol. Chem.* **273**, 20347-20353

80. Sheng, Z., and Charbonneau, H. (1993) *J. Biol. Chem.* **268**, 4728-4733

81. Sikorski, R. S. and Hieter, P. (1989) *Genetics* **122**, 19-27

82. Victor, M., Bei, Y., Gay, F., Calvo, D., Mello, C., and Shi, Y. (2002) *EMBO Report* **3**, 50-55

83. Bregman, D. B. X., Du, L., van der Zee, S., and Warren, S. L. (1995) *J. Cell. Biol.* **129**, 287-298

84. Patturajan, M., Schulte, R. J., Sefton, B. M., Berezney, R., Vincent M., Bensaude, O., Warren, S. L., and Corden, J. L. (1998) *J. Biol. Chem.* **273**, 4689-4694
Figure Legends

FIG. 1. Detection of the capping enzyme-GMP covalent intermediate in C. elegans nuclear extract. 3 µM [α-32P]GTP was incubated at 30 °C for 10 min with 20 µg of C. elegans nuclear extract protein in 20 mM Tris-Cl (pH 7.5)/ 2 mM MnCl2/ 5 mM DTT (lanes 2-7). Lane 1 is a control without protein. MnCl2 was omitted in lanes 3 and 4, and an additional 2 mM EDTA was added in lane 4. Lanes 5 is the same as lane 2 except 30 µM of unlabeled ATP was added. In lane 6, 30 µM of unlabeled GTP was added. In lane 7, the reaction mixture was incubated on ice instead of 30 °C for 10 min. After the reaction, proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and proteins covalently bound to GMP were visualized by autoradiography.

FIG. 2. Protein sequence alignment of CEL-1 with capping enzymes from various species. Similarity analysis was carried out on the National Center for Biotechnology Information Web server using the BLAST algorithm (78). Sequence alignments were made using SEQVU. Letters represent the single-letter amino acid code, and numbers represent the positions of the amino acid residues. Boxed residues denote identities, and shaded residues signify similarity to the CEL-1 amino acid sequence. (A) Alignment of C-terminal GTase regions. The following amino acid sequences are shown: CEL-1 (this study); xCAP1 (AF218793, residues 207-598 of Xenopus laevis capping enzyme (33)); MCE (AF025653, residues 211-597 of mouse capping enzyme (29, 30); HCE (AB009022, residues 211-597 of human capping enzyme (29); Drosophila (AE003495, residues 223-649 of an Drosophila melanogaster open reading frame); Arabidopsis (AC009326, residues 294-653 of an Arabidopsis thaliana open reading frame). The asterisk shows the active site lysine residue (36). Motifs that are highly conserved in GTases (36) are designated by bars above and below the sequences. The arrow shows the position of the previously predicted stop codon of CEL-1 (3). (B) Alignment of N-terminal RTPase regions. The following deduced amino acid sequences are shown: CEL-1 (this study); xCAP1 (AF218793, residues 1-206); MCE (AF025653, residues 1-210); HCE (AB009022, residues 1-210); Drosophila (AE003495, residues...
The amino acid sequences of two PTP-like RNA phosphatases are also shown: human PIR1 (AF023917, residues 31-169 (ref. 79)); and baculovirus BVP (L22858 (80)). The PTP active site consensus motif is marked by a line. Arrows and asterisks show the residues of CEL-1 mutated in this study. The diagonal arrow shows the position of the initiation methionine previously reported (3).

**FIG. 3.** The CEL-1 GTase domain can replace Ceg1 in *S. cerevisiae*. (A) Complementation of a double deletion mutant (ceg1Δcet1Δ) lacking both the GTase and RTPase. The CEG1/CET1 shuffling strain YSB719 (ref. 11) was transformed with the following LEU2/2μm plasmids: vector (pAD5); MCE (pGL-MCE, expressing mouse capping enzyme Mce1 from the GPD promoter); CEL-1 (pAD5-CEL1, expressing CEL-1 from the ADH1 promoter); and CEL-1(1-585) (pAD5-CEL1(1-585), expressing residues 1-585 of CEL-1 from the ADH1 promoter). Leu" transformants were tested for growth after shuffling out pRS316-CEG1-CET1 (ref. 11). (B) Complementation of *ceg1Δ* by the C-terminal region of CEL-1. The CEG1 shuffling strain YSB244 (ref. 9) was transformed with following LEU2/2μm plasmids: vector (pAD5); CEG1 (pAD5-CEG1); MCE(211-594) (pAD5-MCE(211-594) (ref. 11)); and CEL-1(221-623) (pAD5-CEL1(222-623), expressing CEL-1(222-623) from the ADH1 promoter). Leu" transformants were tested for growth after plasmid shuffling (9). (C) Formation of a CEL-1-GMP intermediate. 10 μg of yeast whole cell extract protein was incubated for 10 min at 30 °C with 3 μM [α-32P]GTP. After the reaction, the proteins were separated by SDS-PAGE and analyzed by autoradiography. Extracts were prepared from: left panel, YSB719 transformed with pAD5 (lane 1) or pAD5-CEL1 (lane 2); right panel, YSB244 transformed with pAD5 (lane 1), pAD5-CEG1 (lane 2) or pAD5-CEL1(222-628) (lane 3). Note that at this exposure, the endogenous Ceg1 is not visible.

**FIG. 4.** The isolated metazoan RTPase domain can replace Cet1 in *S. cerevisiae*. (A) Complementation by plasmid shuffling. YSB719 was transformed with pDB20H-MCE(211-597). A His" isolate was subsequently transformed with the following LEU2/2μm plasmids: vector (pRS425 (ref. 81)); CET1 (pAD5-CET1); CEL-1(13-221) (pAD5-CEL1(13-221), expressing CEL-
1(13-221) tagged with an HA epitope from the ADH1 promoter); MCE(1-210) (pAD5-MCE(1-210), expressing Mce(1-210) tagged with the HA epitope from the ADH1 promoter). Leu+ His+ transformants were tested for growth in the presence of 5-FOA to shuffle out pRS316-CEG1-CET1. **(B)** YSB719 was transformed with (Left plate) pAD5-CEL-1(13-221) or (Right plate) pAD5-MCE(1-210). Leu+ isolates were subsequently transformed with the following HIS3/2µm plasmids: vector (pRS423) (ref. 81)), CEG1 (pRSH-CEG1), Mce(211-597) (pDB20H-MCE(211-597)), CGT1 (pRSH-CGT1), pce1* (pDB20H-pce1*), or A103R (pDB20H-A103R, expressing the Chlorella virus GTase A103R from the ADH1 promoter). Growth was tested after plasmid shuffling. **(C)** Immunoprecipitation and GTase-GMP intermediate formation assay. YSB719 transformants selected by 5-FOA were grown in selective media and whole-cell extracts were prepared. Lane 1 has extract from YSB719 transformed with pAD5 and pDB20H-pce1*. This extract was prepared from cells without being selected by 5-FOA. Lanes 2 and 3 have extracts from the same strain transformed with pAD5-CET1 and either pRSH-CGT1 (lane 2) or pDB20H-pce1* (lane 3), respectively. Lane 4 is from cells containing pAD5-CEL-1(13-221) and pDB20H-pce1*. Lane 5 is from cells containing pAD5-CEL-1(13-221) and pRSH-CGT1. Lane 6 is from cells containing pAD5-MCE(1-210) and pDB20H-pce1*. Lane 7 is from cells containing pAD5-MCE(1-210) and pRSH-CGT1. Immunoprecipitation was carried out with 10 µg of extract protein, Protein A Sepharose beads, and monoclonal antibody 12CA5. Precipitates were incubated for 15 min at 30°C with 3 µM [α-32P]GTP, and processed to SDS-PAGE. Proteins were transferred to nitrocellulose membrane and analyzed by immunoblotting with 12CA5 (upper panel) and autoradiography to detect radiolabeled GTase-GMP covalent intermediate (middle panel). The lower panel shows the result of GTase-GMP formation with 20 µg of whole-cell extract protein. The asterisks denote position of radioactive phosphate. The only combination that shows an interaction is S. cerevisiae Cet1 and the C. albicans Cgt1 (lane 2).

**FIG. 5.** Addition of the Ceg1-interacting region from Cet1 onto CEL-1 RTPase domain allows it to function in the presence of Ceg1. YSB719 was transformed with (left) pRSH-CEG1 or (right) pDB20H-MCE(211-597). His+ isolates were subsequently transformed with the following
LEU2/2µm plasmids: vector (pAD5); CET1 (pAD5-CET1); CET1(1-225)-CEL-1(13-221) (pAD5-CET1(1-225)-CEL-1(13-221)), expressing a fusion protein containing residues 1-225 of Cet1 and CEL-1(13-221) tagged with the HA epitope; CET1(1-265)-CEL-1(13-221) (pAD5-CET1(1-265)-CEL-1(13-221)); cet1-446(1-265)-CEL-1(13-221) (pAD5-cet1-446(1-265)-CEL-1(13-221)), expressing a fusion protein consisting of residues 1-265 from the cet1-446 mutant [11(Takase et al., 2000)] and CEL-1(13-221) tagged with the HA epitope; and CEL-1(13-221) (pAD5-CEL-1(13-221)). Leu⁺ His⁺ transformants were grown in the presence of 5-FOA, and the plates were incubated for 3 days at 30 °C.

**FIG. 6.** RTPase activities of CEL-1(13-221) mutants. (A) Purification of recombinant polyhistidine tagged CEL-1(13-221) protein. The peak fractions of the Heparin-Sepharose CL-6B column were analyzed by SDS-PAGE and visualized by Coomassie Brilliant Blue staining (top panel) or immunoblotting with monoclonal anti-polyhistidine antibody (bottom panel). 400 ng and 50 ng of protein were loaded for the top and bottom panels, respectively. Lane 1, wild-type CEL-1(13-221); lane 2, C136S; lane 3, R142K; lane 4, D76N; lane 5, E111Q; lane 6, D112N. For reasons that are not clear, the D76N mutant shows slightly altered mobility. (B) RTPase assay. The wild-type and mutated CEL-1(13-221) proteins were incubated at 30 °C for 10 min with 1 µM of termini of a \[^{32}\text{P}\]GTP-labeled 65 nucleotide RNA. Reaction mixtures were analyzed by thin layer chromatography (TLC) on PEI-cellulose plates. Released phosphate was detected by autoradiography and radioactive spots on were cut out and quantitated by liquid scintillation counting. Relative amounts of released phosphate were plotted versus protein amount. (C) In vivo analysis of CEL-1(13-221) mutants by plasmid shuffling. YSB719 carrying pDB20H-MCE(211-597) was transformed with the vector pAD5 or derivatives expressing the indicated CEL-1 alleles. Leu⁺ His⁺ transformants were tested for growth in the presence of 5-FOA. Plates were incubated for 4 days at 30 °C. (D) Immunoblot analysis of *S. cerevisiae* whole-cell extracts. YSB719 cells carrying pDB20H-MCE(211-597) and CEL-1(13-221) derivatives were grown in selective media, but without shuffling out the CEG1/CET1 plasmid. Extracts were prepared and 10-µg protein was analyzed by SDS-PAGE and immunoblotting with the anti-HA antibody12CA5.
Lane 1, pAD5; lane 2, pAD5-CEL-1(13-221); lane 3, pAD5-CEL-1(13-221)C136S; lane 4, pAD5-CEL-1(13-221)C136A; lane 5, pAD5-CEL-1(13-221)R142K; lane 6, pAD5-CEL-1(13-221)R142A; lane 7, pAD5-CEL-1(13-221)D76N; lane 8, pAD5-CEL-1(13-221)E111Q; lane 9, pAD5-CEL-1(13-221)D112N.

FIG. 7. Substrate specificity and catalytic properties of the CEL-1 RTPase domain. (A)
Substrates for the RTPase assay, [γ-32P]ATP-terminated oligoribonucleotides (pppA(pC)pC, n=0-3) were synthesized with T7 DNA primase and purified with DEAE-Sephadex A-25 column chromatography. Each RNA was resolved by electrophoresis on a 36 % polyacrylamide gel containing 3 M urea, and 32P label (asterisk) was detected by PhosphorImager. Lanes 2-5, 80 pmol of termini of dimeric (pppApC), trimeric (pppApCpC), tetrameric (pppApCpCpC), and pentameric (pppApCpCpCpC) RNAs, respectively. In lane 1, [γ-32P]ATP was loaded as a marker.

(B) RTPase activity. (Left panel) CEL-1(13-221) releases the gamma-phosphate. The indicated amounts of the wild-type protein (CEL-1(13-221); lanes 3-5) and the active site cysteine mutant (CEL-1(13-221)C136S; lanes 6-8) were incubated for 10 min at 30 °C with 0.1 µM (of termini) [γ-32P]ATP-terminated dinucleotide RNA (pppApC). The reaction mixtures were analyzed by TLC on PEI-cellulose plates, and 32P label (asterisk) was detected by Phosphor Imager. In lane 1 of each panel, the substrate was incubated with 1 unit of calf intestinal phosphatase (CIP) to determine the position of free phosphate. (Right panel) CEL-1(13-221) leaves a 5’ diphosphate end.
Reactions in lanes 1-8 were identical to those in lanes 1-8 of the left panel, except [γ-32P]ATP-terminated dinucleotide RNAs were replaced with [α-32P]ATP-terminated dinucleotide RNA (pppApC). The position of ppApC was determined using ADP-terminated dimer as standards (8).

(C) NTPase activity. Reactions in lanes 1-5 were identical to those in lanes 1-5 in the left panel of (B), except [γ-32P]ATP-terminated dinucleotide RNAs were replaced with 5 µM [γ-32P]ATP (pppA).
Incubation was for 30 min at 30 °C.

FIG. 8. Effect of RNA chain length on the CEL-1 RTPase. The indicated amounts of CEL-1(13-221) was incubated for 10 min at 30 °C with 1 µM of termini of [α-32P]ATP-terminated di-, tri-,
tetra-, and pentanucleotides. The reaction mixtures were analyzed by TLC on PEI-cellulose plates. $^{32}$P label was detected by autoradiography, and radioactive spots on PEI-cellulose plate were cut out and counted by liquid scintillation.

**FIG. 9.** Cel-1 is required for embryonic development and CTD Serine 2 phosphorylation. (A) Terminal developmental arrest of *cel-1(RNAi)* embryos. Wild type (WT) and RNAi embryos were examined by differential interference (DIC) microscopy. The WT embryo in the left panel is about to hatch, but *ama-1(RNAi)* (Pol II) and *cel-1(RNAi)* embryos each arrested with approximately 100 cells. Embryos measure ~50 µm. (B) Expression of RNA Pol II and the CDK-9 kinase in WT and *cel-1(RNAi)* embryos. Embryos were stained simultaneously with DAPI to visualize DNA, a CDK-9 antibody (23), and an antibody to the unphosphorylated Pol II large subunit (8WG16). (C) CTD serine 5 phosphorylation in *cel-1(RNAi)* embryos. Wild type or RNAi embryos (presented in rows) were stained with DAPI, an SR protein antibody (16M3) as a control, or an antibody to CTD phosphoserine 5 (pCTD) (26), as indicated. An expanded $\alpha$-PSer5 stained somatic nucleus (white arrow) is shown in the right column. Representative embryos at comparable stages are presented. (D) *Cel-1* is required for CTD serine 2 phosphorylation. Wild type or RNAi embryos were stained with DAPI, a CBP-1 antibody for a staining control (82), and an antibody to CTD phosphoserine 2 (H5) (83, 84). An expanded $\alpha$-PSer2 stained somatic nucleus is shown as in (C). In the transcriptionally silent germline precursor (red arrow), only weak cross-reactivity with perinuclear germline P granules is detected. Mitotic nuclei, which also cross-react with $\alpha$-PSer2 in the absence of the Pol II epitope (23, 63), are marked with asterisks. Representative embryos of comparable stages are shown.
vector

\[ \text{CET1} \]
\[ \text{CEL-1(13-221)} \]
\[ \text{CET1(1-225)-CEL-1(13-221)} \]
\[ \text{cet1-446(1-265)-CEL-1(13-221)} \]
\[ \text{CET1(1-265)-CEL-1(13-221)} \]

\[ \text{CEG1} \]
\[ \text{MCE(211-597)} \]
The Caenorhabditis elegans mRNA 5′-capping enzyme: In vitro and in vivo characterization
Toshimitsu Takagi, Amy K. Walker, Chika Sawa, Felix Diehn, Yasutaka Takase, T. Keith Blackwell and Stephen Buratowski

J. Biol. Chem. published online February 7, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M212101200

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