The Caenorhabditis elegans mRNA 5′-Capping Enzyme

IN VITRO AND IN VIVO CHARACTERIZATION*

Received for publication, November 27, 2002, and in revised form, January 31, 2003
Published, JBC Papers in Press, February 7, 2003, DOI 10.1074/jbc.M212101200

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 N-terminal region with RTPase activity and a 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 inSaccharomyces cerevisiae. The CEL-1 RTPase domain is related to sequence to protein-tyrosine phosphatases; therefore, mutagenesis of residues predicted to be important for RTPase activity was carried out. CEL-1 uses a mechanism similar to protein-tyrosine phosphatases, 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.

Most eukaryotic and viral mRNAs are modified at their 5′ end by a “cap” structure that 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)3 removes the γ-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 (LV/HCXXGXR(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) data base 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 theSaccharomyces 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 C. elegans.

EXPERIMENTAL PROCEDURES

DNA Cloning Methods—Supplementary tables listing oligonucleotides and plasmids used in this study are available at the Buratowski laboratory web site.3 PCR for construction of plasmids and site-directed

GTase, GTP::mRNA guanylyltransferase; PTP, protein-tyrosine phosphatase; RNAi, RNA interference; 5-FOA, 5-fluoroorotic acid; pol II, RNA polymerase II; CTD, C-terminal domain of the largest subunit of RNA polymerase II; CTD-P, phosphorylated CTD; HA, influenza virus hemagglutinin; EST, Expressed Sequence Tag; ORF, open reading frame; P-TEFb, positive transcription elongation factor b; CDK, cyclin-dependent kinase; kb, kilobase; DTT, dithiothreitol; ds, double-stranded; MCE, mouse capping enzyme; BVP, baculoviral PTP.


* This work was supported by National Institutes of Health Grants GM56663 (to S. B.) and GM62891 (to T. K. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Senior Postdoctoral Fellow of the American Cancer Society, Massachusetts Division. Present address: Dept. of Automated Biotechnology, Harvard Medical School, Boston, Massachusetts 02115. Tel.: 617-432-0696; Fax: 617-738-0516; E-mail: steveb@hms.harvard.edu.

§§ A scholar of the Leukemia And Lymphoma Society. To whom correspondence should be addressed: Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Ave., Boston, MA 02115. Tel.: 617-432-0696; Fax: 617-738-0516; E-mail: steveb@hms.harvard.edu.

1 The abbreviations used are: RTPase, RNA 5′-triphosphatase;
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 yk789b08 (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-1 upstream and CEL-1 downstream primers, and the product was purified and used in the second PCR as a megaprimer. The second reaction used yk789b08 as template and used 3′ CEL-1 longer and CEL-1 upstream 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-1 longer was replaced with Cegl, CE-C. A degenerate codon of C. elegans capping enzyme was used in this study were YSBS44 (MATa ura3-52 leu2-3,113 his3Δ200 yk789b08 ceg1Δ::HIS3 [pRS316-CEG1]) (9), YSBS53 (MATa ura3-52 leu2Δ1 trp1Δ63 his3Δ200 lys2Δ202 cel1Δ::TRP1 [pRS316-CEG1]) (10), and YSB719 (MATa ura3::52 leu2Δ1 trp1Δ63 his3Δ200 lys2Δ202 cel1Δ::TRP1 ceg1Δ::LYS2 [pRS316-CEG1-CEL1]) (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 from C. elegans, immunoprecipitation, and subsequent enzymatic-GMP assays were performed as described previously (11). Site-directed Mutagenesis—Site-directed mutagenesis was carried out using either single-stranded phagemid (15, 16) or PCR (17). In both cases, the plasmids pBS-CEL-1-(1-585) (Ref. 3, listed as pBS-CEL-1 therein) was used as the template. Mutations were verified by dideoxy DNA sequencing.

To change Arg142 to Lys (R142K) or Ala (R142A), Asp76 to Asn (D76N), and Glu-111 to Gln (E111Q), phagamid 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, pBS-CEL-1-D76N, and pBS-CEL-1-E111Q, served as template for subsequent PCR. Except for D76N, we amplified 0.7-kb fragments corresponding to the CTD domain, containing those mutations using Cegl, CE-B, and CEL-1T222stop as primers. For the wild type and the active site cysteine mutants, we used pBS-CEL-1-(1-585), pET-his7CEG1-(13-248)C136S, and pET-his7CEG1-(13-248)C136A as templates for the same PCR. Each product was subcloned into pCR-Script SK(+) to generate pBS-CEL-1-(1-153)-221 version 2, pBS-CEL-1-(1-221)C136S version 2, pBS-CEL-1-(1-221)C136A version 2, pBS-CEL-1-(1-221)R142K version 2, pBS-CEL-1-(1-221)R142A version 2, and pBS-CEL-1-(1-221)E111Q version 2, respectively. We further subcloned inserts from these plasmids into yeast expression vector pAD5 (11) and bacterial expression vector pSET-His (18). For D76N, a 0.7-kb fragment was amplified with pBS-CEL-D76N as template and Cegl, CE-B, and CEL1-(3′)Sac as primers. The product was digested with NcoI and SacI and subcloned into pAD5.

PCR-mediated site-directed mutagenesis was used to change Asp142 to Lys (D142K) and Ala (D142A), Arg76 to Asn (R76N), and Asp111 to Asn (E111N), 5′ PCR product was amplified with CEL1-D142K and CEL1-D142A, or CEL1-R76N, and CEL1-R76A, or CEL1-E111K, and CEL1-E111N, respectively. The standard reaction (100 ml) contained 40 mM Tris-Cl, pH 7.5, 10 mM MgCl2, 10 mM DTT, 50 lM bovine serum albumin, 50 mM potassium glutamate, 2 mM dTT, 2 mM CTP, 0.3 mM [γ-32P]ATP or [α-32P]ATP (500–1000 cpm/pmol) (PerkinElmer Life Sciences), 1 mM synthetic oligonucleotide template, and 1 lM (hexamer) T7 primase (21). The sequences of the oligonucleotides used were: ddirribonucleotide ppApCpC, 5′-G(CGGCGT-CGTTCG)-3′; trinucleotide ppApCpCpC, 5′-G(CGGCGT-CGTTCG)-3′; tetraarribonucleotide pppApCpCpCpC, 5′-G(CGGCGT-CGTTCG)-3′; and pentarribonucleotide pppApCpCpCpCpC, 5′-G(CGGCGT-CGTTCG)-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 l of buffer A (20 mM Tris-HCl, pH 7.9, 7 mM urea) containing 100 mM NaCl. RNAs were further purified by chromatography with a column (1.0 × 7.0 cm) of DEAE-Sepharose A-25 (Amersham Biosciences) 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 700-ml linear gradient of 150–300 mM NaCl in buffer A. RNAs were analyzed with electrophoresis on a 38% polyacrylamide, 3 mM urea gel and autoradiography. Samples were then pooled and dialyzed against 10 mM Tris-HCl, pH 8.0, 1 mM EDTA with Spectra/Per 7 ( Spectrum). The RNAs were then lyophilized and dissolved in water. Previously, we used triethylamine bicarbonate (TEA-HCO3) for buffer for the DEAE-Sepharose A-25 column chromatography (8). However, we observed that residual amounts of TEA·HCO3− contaminating the purified substrate, inhibit RTPase activity, and therefore, the protocol was modified as indicated. [γ-32P]GTP-terminated RNA was synthesized by in vitro transcription of linearized plasmid template (3) using recombinant polynucleotide-activated T7 RNA polymerase (22).

Enzymatic Assays—GTPase, RTPase, and nucleotide phosphohydrolase assays 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-CEL-1-(1-585) and injected at 1.0 μg/ml into young adults (2–8 fertilized embryos). Equivalent results were obtained with cel-1 RNAs covering the RTPase or GTPase domains. ama-1(RNAi) and cel-1(RNAi) embryos were analyzed at 24 and 36 hr post-injection, respectively, when uniform populations of arrested embryos appeared. For immunostaining, embryos were collected from dissected hermaphrodites 36 hr 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% parafomaldehyde fixation and freeze-cracked before treating with methanol. Washes and antibody incubations were performed in PBT (1× phosphate-buffered saline, 1% Triton X-100, 1% bovine serum albumin) before staining. Staining with the phosphorylated C-terminal domain (CTD-P) (α-Ser5-P-2) (26) and If5 (α-Ser5-P-2) (27) antibodies was performed as described by Roger et al. (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 \([a-\text{32P}]\)GTP to form the covalent GTase-\([32\text{P}]\)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 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 ~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 number Z75225) 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 viruses (2, 28, 34–36; Fig. 2A) as well as an N-terminal domain related to the PTPs.

The protein encoded by the 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 data base (www.ncbi.nlm.nih.gov/BLAST/) for naturally occurring CEL-1 cDNAs and found SP9F10 (accession number 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 data base server at the DNA Data Bank of Japan (www.ddbj.nig.ac.jp/c-elegans.html/CE_BLAST.html) contains clone yk786802 containing sequences for the 5’ region of CEL-1. This encodes the previously identified RTPase domain (Refs. 3 and 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 (Ref. 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 mass 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-(1–585).

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 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 terminal, 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 211–585) 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 overexpression 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 (amino acids 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 ceg1Δcet1Δ 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 overexpressed in ceg1Δcet1Δ strain in combination with MCE-(211–597), cells containing Cet1 formed colonies after 1 day, whereas cells containing either CEL-1-(1–221) 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 cet1Δ strain even though enzyme-GMP complex was detectable in lysates.3

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

3 T. Takagi, unpublished observation.
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 [γ-32P]GTP to detect any GTase activity. The lower panel shows the result of GTase-GMP formation with 20–30 μg of whole-cell protein.

Previously, CEL-1-(13–221) and MCE-(1–210) do not support growth of a cet1 Δ 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 chimera were co-expressed along with Ceg1 overexpressed from a 2-μm plasmid in a ceg1Δ 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, can support viability. While amino acids 1–265 are derived from the mutant ceg1-146 (P245A/W247A), which is disrupted for the ability to interact with Ceg1 (11), the ability to replace Cet1 was disrupted. All

**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 LEU2/2-μm plasmids: vector (pRS425 (81)); CET1 (pAD5-CET1); CEL-1-(13–221) (pAD5-CEL-1-(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 pRS425-HIS3 (pAD5-CEL-1-(13–221) (left plate) or pAD5-MCE-(1–210) (right plate)). Leu+ isolates were subsequently transformed with the following HIS2/2-μm plasmids: vector (pRS425 (81)), CEG1 (pRS425-HIS3), MCE-(211–597), and pRSH-CGT1. B, YSB719 was transformed with pAD5-CET1-(13–221) (left plate) or pAD5-MCE-(1–210) (right plate). Leu+ isolates were subsequently transformed with the following HIS2/2-μm plasmids: vector (pRS425 (81)), CET1 (pRS425-HIS3), MCE-(211–597), CGT1 (pRS425-HIS3), and pRSH-CGT1, pcl1-1 (pDB20H-pcl1-1), or A103R (pDB20H-A103R, expressing the Chlorella virus GTase A103R from the ADH1 promoter). Expression 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-pcl1-1. 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-pcl1-1 (lane 3), respectively. Lane 4 is from cells containing pAD5-CET1-(13–221) and pDB20H-pcl1-1. Lane 5 is from cells containing pAD5-CET1-(13–221) and pDB20H-pcl1-1. Lane 6 is from cells containing pAD5-MCE-(1–210) and pDB20H-pcl1-1. 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 μCi [γ-32P]GTP and SDS-PAGE. Proteins were transferred to nitrocellulose membrane and analyzed by immunoblotting with 12CA5 (upper panel) and autoradiography to detect radio-labeled GTase-GMP complex (middle panel). The lower panel shows the result of GTase-GMP formation with 20 μg of whole-cell extract protein. The asterisks denote the position of radioactive phosphate. The only combination that shows an interaction is S. cerevisiae Cet1 and the C. albicans Cgt1 (lane 2).

**Fig. 5.** The 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) pRS425-CEG1 or (right) pDB20H-MCE-(211–597). His+ isolates were subsequently transformed with the following LEU2/2-μm 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)); celt1-1 (pAD5-CET1-(1–265)-CEL-1-(13–221)); CET1-(1–446)-CEL-1-(13–221) (pAD5-CET1-(1–446)-CEL-1-(13–221)), expressing a fusion protein consisting of residues 1–446 from the cet1-146 mutant (11) and CEL-1-(13–221) tagged with the HA epitope); and CEL-1-(13–221) (pAD5-CET1-(1–225)-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.
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)H-CXXGXXR(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 tyrosine residue.

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. Arginine 142 in the consensus motif was mutated. Also, aspartate 76, glutamate 111, and aspartate 112 were mutated because they are candidates for the proton-donating acidic res-
idue. Arg-142 and Asp-76 are conserved in all of the PTP-like RTPases, whereas Glu-111 and Asp-112 are not (Fig. 2B). The histidine-tagged mutants C136S, R142K, D76N, E111Q, and D112N were purified from *E. coli* (Fig. 6A). Their RTPase activities were tested with *γ*-32P[ATP]-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, that of the wild-type protein.

We also tested if these mutants can support viability of a *ceg1* -1 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 *in vivo* phenotypes correlated well with *in vitro* results. D76N, which had only 10% of wild-type activity *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 Glu-111 and Asp-112 are not highly conserved and do not seem to be vital *in vivo* (Fig. 6C). In contrast, Asp-76 is conserved, and a D76N mutant cannot 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 (Asp-66 of MCE and Asp-60 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 ppApC (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 γ-pppApC; bold denotes the position of the radioactive phosphate) or α-(pppApC) 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 nucleotide phosphohydrolase 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, PRI1, and BVP (51, 55).

Next we tested the effect of RNA chain length (Fig. 8). CEL-1-(13–221) hydrolyzes the β-γ phosphodiester bond of trinucleotide more efficiently than that of dinucleotide. Little difference was seen between trimer-, tetramer-, and pentanucleotides. With a double-reciprocal plot, the *Km*/*Vmax* values with ATP, dinucleotide, trinucleotide, and tetrancleotide were calculated to be 5.5 × 10−5, 0.5 × 10−5, 4.6 × 10−5, 5.1 × 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 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 ~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 with 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 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 transcription factor IIH 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-Ser(P)-5 and α-Ser(P)-2 for clarity (Fig. 9, C and D).

In the early *C. elegans* embryo, the appearance of both α-Ser(P)-5 and α-Ser(P)-2 staining depends upon transcription. Staining with α-Ser(P)-5 and α-Ser(P)-2 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 (tb-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-Ser(P)-2 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 activi-
C. elegans mRNA Capping Enzyme

Fig. 9. Cell-1 is required for embryonic development and CTD serine 2 phosphorylation. A, terminal developmental arrest of cell-1(RNAi) embryos. Wild-type (WT) and RNAi embryos were examined by differential interference (DIC) microscopy. The wild-type embryo in the left panel is about to hatch, but ama-1(RNAi) (pol II) and cell-1(RNAi) embryos each arrested with ~100 cells. Embryos measure ~50 μm. B, expression of RNA pol II and the CDK-9 kinase in wild-type and cell-1(RNAi) embryos. Embryos were stained simultaneously with 4′,6-diamidino-2-phenylindole hydrochloride (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 cell-1(RNAi) embryos. Wild-type or RNAi embryos (presented in rows) were stained with 4′,6-diamidino-2-phenylindole hydrochloride, an serine/arginine-rich protein antibody (16M3) as a control, or an antibody to CTD phosphoserine 5 (26), as indicated. An expanded α-Ser(P)-5 stained somatic nucleus (white arrow) is shown in the right column. Representative embryos at comparable stages are presented. D, cell-1 is required for CTD serine 2 phosphorylation. Wild type or RNAi embryos were stained with 4′,6-diamidino-2-phenylindole hydrochloride, a CEBP-binding protein (CBP)-1 antibody for a staining control (82), and an antibody to CTD phosphoserine 2 (H5) (83, 84). An expanded α-Ser(P)-2-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 α-Ser(P)-2 in the absence of the pol II epitope (23, 63), are marked with asterisks. Representative embryos of comparable stages are shown.

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 data base. 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 ceg1Δ (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, 51, 55, 64, 65). A phosphocysteine intermediate was detected with MCE and BVP (52, 53). Other PTP-like enzymes with substrates other than phosphoarginine 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) shows 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 (Asp-76 of CEL-1, Asp-66 of MCE, and Asp-60 of BVP; see Fig. 2B) only slightly diminishes activity (Fig. 6; Refs. 51–52). The equivalent mutations in PTPs lower activity by 103–105-fold (4–7). X-ray crystallography of MCE (1–210) shows that Asp-66 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 γ-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 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 diphosphate 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 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 to guide GTase 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), S. pombe pce1, or C. albicans Cgt1 (Fig. 4, A and B). Because 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 catalytic enzymes of metazoa 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 (Ref. 76 and this study). In contrast, a low copy plasmid of the full-length enzyme was sufficient for rescuing a ced-1Δ strain (52). Transfection experiments showed that MCE–1–210 is mostly cytoplasmic in mammalian cells (51). This may also be true in 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. cel-1(1RNAi) 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), ttk-1 (TFIIB), or multiple TAFs (23, 24, 27). One cel-1(1RNAi) 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 ttk-1(1RNAi) embryos, in which basal factor TFIIB is knocked down, both serine 5 and serine 2 phosphorylation are reduced to background (23, 27). In cel-1(1RNAi) embryos, CTD serine 5 phosphorylation appears to be relatively unaffected, whereas 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 phosphorylation occurred when we depleted either of the P-TEFb components, CDK-9, or cyclin T (23). Levels of the CDK-9 phosphorylation occurred when we depleted either of the P-TEFb components, CDK-9, or cyclin T (23).

1. Acknowledgments—We thank Drs. Aaron J. Shatkin and Fabio Piano for communicating results before publication, Gary Ruvkun for supply of C. elegans mRNA Capping Enzyme.
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

doi: 10.1074/jbc.M212101200 originally 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

This article cites 83 references, 50 of which can be accessed free at http://www.jbc.org/content/278/16/14174.full.html#ref-list-1