Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 6624–6628, July 1994
Biochemistry

Active site of the mRNA-capping enzyme guanylyltransferase from *Saccharomyces cerevisiae*: Similarity to the nucleotidyl attachment motif of DNA and RNA ligases

(nucleotidyl transfer/covalent catalysis)

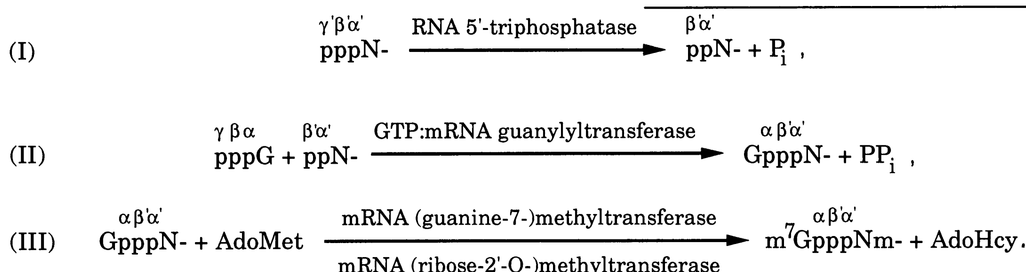
LUCILLE D. FRESCO* AND STEPHEN BURATOWSKI*†

The Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142

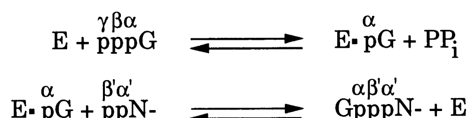
Communicated by Phillip A. Sharp, March 8, 1994

ABSTRACT **Nascent mRNA chains are capped at the 5′ end by the addition of a guanylyl residue to form a G(5′)ppp(5′)N. . . structure. During the capping reaction, the guanylyltransferase (GTP:mRNA guanylyltransferase, EC 2.7.7.50) is reversibly and covalently guanylylated. In this enzyme–GMP (E-GMP) intermediate, GMP is linked to the ε-amino group of a lysine residue via a phosphoamide bond. Lys-70 was identified as the GMP attachment site of the *Saccharomyces cerevisiae* guanylyltransferase (encoded by the *CEG1* gene) by guanylylpeptide sequencing. *CEG1* genes with substitutions at Lys-70 were unable to support viability in yeast and produced proteins that were not guanylylated *in vitro*. The CEG1 active site exhibits sequence similarity to the active sites of viral guanylyltransferases and polynucleotide ligases, suggesting similarity in the mechanisms of nucleotidyl transfer catalyzed by these enzymes.**

Capping of eukaryotic pre-mRNA is the earliest detectable RNA processing event (1, 2). The cap consists of an inverted 7-methylguanosine residue linked via a 5′–5′ triphosphate bridge to the first transcribed residue. Capping proceeds by the sequential action of several enzymatic activities (reviewed in refs. 3 and 4). Nascent RNA molecules synthesized by RNA polymerase II carry 5′-triphosphate termini. The capping enzyme possesses a RNA 5′-triphosphatase activity that removes the terminal phosphate and a guanylyltransferase (GTP:mRNA guanylyltransferase, EC 2.7.7.50) activity that catalyzes the transfer of a GMP moiety from GTP to the 5′-diphosphoryl group of the RNA. Subsequent methylation of the cap (5–7) may serve to prevent reversal of the guanylyl transfer reaction (8). These reactions may be summarized as follows:



The transfer of GMP from GTP to the diphosphate termini of mRNA (transguanylylation, reaction II) proceeds through two reversible steps that involve an enzyme guanylylate (E-GMP) intermediate (9–11):



The covalent E-GMP intermediate (E•pG) can be conveniently identified by radioactive labeling of the protein with [α-^{32}P]GTP.

Interest in the mechanism of capping has focused attention on the E-GMP complex. Cellular mRNA guanylyltransferases from human, rat liver, calf thymus, *Artemia salina*, wheat germ, and *Saccharomyces cerevisiae* all form E-GMP covalent complexes (reviewed in ref. 3). In addition, cytoplasmic viruses encode their own capping enzymes. The guanylyltransferases of vaccinia virus (12, 13), reovirus (11, 14, 15), African swine fever virus (16), rotavirus (17), bluetongue virus (18), and Shope fibroma virus (19) have also been shown to form E-GMP complexes. The E-GMP complex has been shown to act as a *bona fide* intermediate in the transguanylation reaction (10, 11). The nucleotidyl attachment site is a lysine in the rat liver (13), reovirus (20), and vaccinia virus guanylyltransferases (13, 21). The nucleotidyl–enzyme linkage is a phosphoamide bond between the catalytic lysine and GMP (10, 12–14, 21, 22), suggesting that the nucleotidyl transfer mechanism proceeds via nucleophilic attack of the ε-amino group of the lysine on the α phosphate of GTP (23). The extensive conservation of the capping reaction suggests that the GMP–lysine bond is a feature of all guanylyltransferases.

The *S. cerevisiae* capping enzyme is a heterodimer of 80-kDa and 52-kDa subunits possessing 5′-triphosphatase and guanylyltransferase activities, respectively (24, 25). The essential gene *CEG1* encodes the small subunit (26) and is the only cellular capping enzyme gene described to date. The gene for the large subunit has not yet been identified. The *S. cerevisiae* E-GMP linkage exhibits the characteristics of a phosphoamide bond (13)—i.e., alkali-stability, acid-lability, and acidic hydroxylamine cleavage susceptibility (23). These properties are consistent with a GMP–lysine bond, although they could also be due to phosphoamide linkage to a histidine or arginine.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ORF, open reading frame; WT, wild type; 5-FOA, 5-fluoroorotic acid; E-GMP, enzyme–GMP complex.

*Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115.
†To whom reprint requests should be addressed.

In the course of polynucleotide ligation by DNA and RNA ligases, a covalent adenylyl–enzyme intermediate is formed. AMP is attached to the enzyme by a phosphoamide bond with a lysine that lies within a conserved motif (reviewed in ref. 27). The sequence surrounding the site of GMP attachment in the vaccinia virus mRNA guanylyltransferase is similar to the motif containing the active-site lysine residues of DNA and RNA ligases (16, 28). In contrast, the guanylyltransferase active site from reovirus (20) does not obviously conform to this consensus. While no extensive sequence similarities to the viral capping enzymes are apparent, the guanylyltransferase of *S. cerevisiae* contains a short sequence that is similar to the proposed consensus motif for nucleotidyl attachment (16, 28).

To determine the guanylylated residue of the yeast capping enzyme, we performed phosphopeptide sequencing and examined the effects of replacement of the suspected active-site lysine. We report here that the site of guanylyl attachment in yeast CEG1 is Lys-70. The active-site sequence conforms to a motif for nucleotidyl attachment found in enzymes that catalyze nucleotidyl transfer to nucleic acid substrates.

MATERIALS AND METHODS

Plasmids. The *CEG1* gene was amplified from *S. cerevisiae* genomic DNA by the polymerase chain reaction (PCR) using oligonucleotide primers derived from the published genomic sequence (26). The entire *CEG1* gene was amplified by using oligonucleotide primers CEG1-A5′ and CEG1-C3′. The CEG1 open reading frame (ORF) was amplified by using primers CEG1-$B_2$5′ and CEG1-C3′, creating an *Nco* I site at the initiation codon in the process. The sequences of PCR primers are as follows: CEG1-A5′, 5′-GGGAGATCTCACTGAAAGGCAGATGATC-3′; CEG1-$B_2$5′, 5′-GGGAGATCTGCCATGGTTTTAGCAATGGAAAGTAGAG-3′; and CEG1-C3′, 5′-GGGAGATCTATAACACTTTCACCCTCGCTTCTC-3′.

The PCR product containing the entire *CEG1* gene was digested with *Bgl* II and ligated into the *Bam*HI site of pRS315, pRS316 (29), or pBSKS(+), generating pRS315-CEG1, pRS316-CEG1, and pBSKS(+)-CEG1, respectively. The PCR product containing the *CEG1* ORF was digested with *Bgl* II and inserted into the *Bam*HI site of pBSKS(+) to create plasmid pBSKS(+)-CEG1orf.

For bacterial production of CEG1 protein and to facilitate purification by Ni^{2+} chelate chromatography, the ORF was cloned downstream of a phage T7 promoter and engineered to carry seven histidine residues at the N terminus. The *Nco* I–*Hin*dIII *CEG1* ORF fragment from pBSKS(+)-CEG1orf was cloned into the *Nco* I and *Hin*dIII sites of pET11d-His_7 to create pET11-HisCEG1orf. For expression of *CEG1* mutants encoding a replacement for Lys-70, the 1.1-kb *Eco*RI fragments from pRS315-CEG1 mutant plasmids carrying the Lys-70 mutation (see below) were inserted into *Eco*RI-cleaved pET11-HisCEG1orf. Plasmid sequences were verified by DNA sequencing with dideoxynucleotides (Sequenase 2.0; United States Biochemical).

A *CEG1* disruption allele was constructed as follows: a 1.6-kb *Pst* I–*Spe* I *CEG1* fragment from pBSKS(+)-CEG1 was inserted into the *Pst* I and *Spe* I sites of pBSKS(+) from which the polylinker sequences between the *Eco*RI and *Xho* I sites were deleted. The resulting plasmid [pBSKS(+)(ΔRI-X)-CEG1] was digested with *Eco*RI and *Xho* I, removing the region encoding amino acids 36 through 397. A 2.4-kb *Eco*RI–*Sal* I fragment carrying the *HIS3* gene was inserted into the plasmid, generating pBSKS(+)(ΔRI-X)-ceg1Δ1::HIS3.

Mutant alleles of *CEG1* were created by oligonucleotide-directed mutagenesis as described (30). pRS315-CEG1 was used to transform the *dut⁻ ung⁻* *Escherichia coli* strain CJ236, and uracil-containing single-stranded phagemid DNA was isolated following infection with M13K07 helper phage. Mutagenic oligonucleotides were phosphorylated by using T4 polynucleotide kinase and annealed to the phagemid DNA, and extension and ligation reactions were performed as described (30). *E. coli* strain S28 (dut^+ ung^+) was transformed with the reaction products to select against the uracil-containing parental strand. Plasmid DNA was isolated by alkaline lysis miniprep procedure, and mutants were identified by DNA sequencing.

The sequences of the mutagenic oligonucleotides, designed to change the amino acids at Lys-59 and Lys-70 are as follows: CEG1-K59mut5′, 5′-CTGATGTGGAAGAGABBCTGCTTGCGCATG-3′, where B = T, C, or G; and CEG1-K70mut5′, 5′-GATTACTACGTTTGTGAGABNACAGATGGTCTGCGGG-3′, where N = T, C, G, or A.

Yeast Methods. Plasmids were introduced into yeast by using a modified lithium acetate transformation protocol (31). Media preparation, tetrad dissection, genomic DNA isolation, plasmid shuffling, and all other yeast manipulations were performed by standard methods (30, 32). 5-Fluoroorotic acid (5-FOA) was purchased from PCR Incorporated.

Yeast Strains. Yeast strains were CTY143 (*MAT***a**/*MATα ura3-52/ura3-52 leu2-3,112/leu2-3,112 his3Δ200/his3Δ200*), YSB225 (*MAT***a**/*MATα ura3-52/ura3-52 leu2-3,112/leu2-3,112 his3Δ200/his3Δ200 CEG1/ceg1Δ1::HIS3*), YSB238 (*MAT***a**/*MATα ura3-52/ura3-52 leu2-3,112/leu2-3,112 his3Δ200/his3Δ200 CEG1/ceg1Δ1::HIS3* {pRS316-CEG1}), and YSB244, (*MAT***a** *ura3-52 leu2-3,112 his3Δ200 ceg1Δ1::HIS3* {pRS316-CEG1}).

CTY143 was transformed with *Pst* I/*Sma* I-digested pBSKS(+)(ΔRI-X)-ceg1Δ1::HIS3. His^+ transformants were selected after 3–6 days to generate YSB225. Disruption of *CEG1* was verified by Southern analysis (data not shown). Sporulation resulted in 2:2 segregation of viability, and all viable spores were His^-. YSB225 was transformed with plasmid pRS316-CEG1 to create YSB238. A Ura^+ His^+ haploid strain (YSB244) derived from YSB238 was used for plasmid shuffling experiments.

Purification of Recombinant Guanylyltransferase. Polyhistidine-tagged CEG1 (His_7-CEG1) proteins were expressed by using a T7 promoter/polymerase system in *E. coli* strain pGP1-2/K38 (33). Strains carrying bacterial expression plasmids were grown at 30°C to an OD_{595} ≈0.5 and induced by warming the cultures to 42°C for 15 min. Cells were incubated at 37°C and harvested 3 hr after induction. Lysates were prepared, and soluble recombinant proteins were purified by chromatography over a nitrilotriacetic acid-agarose column containing immobilized Ni^{2+} ions (Ni^{2+}-NTA-agarose) as recommended by the manufacturer (Qiagen, Chatsworth, CA). Extracts were incubated in batch with resin for 2–12 hr on a rotator at 4°C. The resin was washed successively with BC100 [20 mM Tris chloride, pH 7.9/20% (vol/vol) glycerol/100 mM KCl/0.5 mM phenylmethylsulfonyl fluoride] containing 5 and 10 mM imidazole and then was poured into a column. Bound protein was eluted with BC100 containing 100 mM imidazole. Fractions were collected and assayed for formation of the E-GMP complex as described below. At this step, CEG1 protein was ≈25% pure.

For further purification, peak fractions from the Ni^{2+}-NTA-agarose column were pooled and further purified by Mono S fast protein liquid chromatography (FPLC; Pharmacia) using a gradient of 10–750 mM KCl in buffer A (20 mM Hepes·KOH, pH 7.9/10% glycerol/2 mM EDTA/1 mM dithiothreitol/0.5 mM phenylmethylsulfonyl fluoride). Fractions were assayed for E-GMP complex formation and His_7-CEG1 activity peaked near 100 mM KCl. Proteins were quantitated by gel electrophoresis, Coomassie blue staining, and comparison with standards. After these two columns, the His_7-CEG1 protein was at least 90% pure. Yields were typically in the range of 1 mg of protein per liter of bacteria.

Assay of Covalent E-GMP Complex Formation. One microliter of each fraction was assayed for E-GMP complex formation in 10-μl reaction mixtures containing 10% glycerol, 5 mM dithiothreitol, 5 mM $MnCl_2$, 20 mM Tris chloride (pH 7.5), and 3 μCi of [α-^{32}P]GTP (3000 Ci/mmol; 1 Ci = 37 GBq; NEN/DuPont). After incubation at 30°C for 15 min, 10 μl of 2× sample buffer (2× = 0.125 M Tris chloride, pH 6.8/4% SDS/20% glycerol/10% 2-mercaptoethanol) was added, and samples were boiled and loaded onto an SDS/10% polyacrylamide gel. Radioactively labeled E-GMP complexes were visualized by autoradiography of the wet gel.

Sequencing of the Guanylyl Peptide. His_7-CEG1 (6.6 μg) was incubated in a typical reaction mixture as described above containing [α-^{32}P]GTP. In parallel, 100 μg of His_7-CEG1 was incubated with unlabeled GTP in a final volume of 100 μl. E-GMP complexes were cleaved in separate reactions with modified sequencing grade trypsin (Boehringer Mannheim) at an enzyme/protein ratio of 1:15. Reactions were carried out for 18 hr at 37°C in 100 mM Tris chloride, pH 8/2 mM $CaCl_2$, conditions that ensured complete digestion. The resulting tryptic peptides were mixed and electrophoresed on an alkaline 40% acrylamide gel designed to resolve phosphopeptides (34, 35). A control reaction lacking His_7-CEG1 was electrophoresed in an adjacent lane for comparison. The guanylylated peptide was localized by autoradiography, cut out of the gel, eluted in 500 μl of distilled water, and spotted onto Immobilon-N (Millipore) membrane as described (34). After washing with 50% methanol, the radioactive spots were cut out, and the attached peptide was sequenced with an Applied Biosystems model 475 microsequencer.

RESULTS

Identification of the CEG1 Active Site by Phosphopeptide Sequencing. To identify the active-site lysine in CEG1, His_7-CEG1 protein was produced and purified from *E. coli* crude extract by Ni^{2+} chelate and Mono S FPLC (Pharmacia) chromatography. His_7-CEG1 protein was radioactively labeled by guanylylation with [α-^{32}P]GTP. The E-GMP complex was then digested to completion with the sequence-specific protease trypsin. The labeled GMP-peptide was purified by alkaline 40% acrylamide gel electrophoresis (Fig. 1*A*), spotted onto a positively charged membrane (Immobilon N), and microsequenced directly (34, 35).

The partial sequence obtained from the peptide was Xaa-(Leu or Phe)-Ala-His-(Asp or Leu)-Tyr-(Tyr or Val)-Val-, where Xaa signifies an unidentified residue (Fig. 1*B*). Sequence further into the peptide was not interpretable. The obtained sequence of the guanylylated peptide matches residues 60–67 of CEG1. Since trypsin cleaves after the basic amino acid residues lysine and arginine, the guanylylated peptide must include amino acids 60–70 (Fig. 1*B*). In addition, the peptide would extend to amino acid 75 if the attached nucleotidyl group blocked cleavage by trypsin, as previously observed for the ligase-AMP intermediates (36, 37). The GMP residue is linked to CEG1 by a phosphoamide bond (13). Therefore, the guanylylated residue is predicted to be either His-63 or Lys-70, since these are the only amino acids in the peptide capable of forming a phosphoamide bond. To date, all of the characterized nucleotidyl attachment sites of guanylyltransferases and polynucleotide ligases are lysine residues. Furthermore, the *S. cerevisiae* Lys-70 is conserved in the *Schizosaccharomyces pombe* guanylyltransferase, while His-63 is not (see below). For these reasons, we believe that Lys-70 must be the guanylyl attachment site in CEG1.

Replacements of Lys-70 Abolish CEG1 Activity *in Vivo*. To confirm that Lys-70 is the site of guanylylation, mutant alleles substituted at this position were assayed for the ability to support viability *in vivo* by plasmid shuffling (Fig. 2). As a control, proteins containing substitutions for the nearby Lys-59 were also tested. *CEG1* alleles carried on the *LEU2* plasmid pRS315, or pRS315 alone (vector), were used to transform YSB244, a haploid yeast that harbors WT *CEG1* on a *URA3* plasmid (pRS316-CEG1), which complements the chromosomal *CEG1* null allele. His^+ Leu^+ transformants were purified and replica-plated to medium containing 5-FOA to select against cells that retained the *URA3* plasmid. Strains

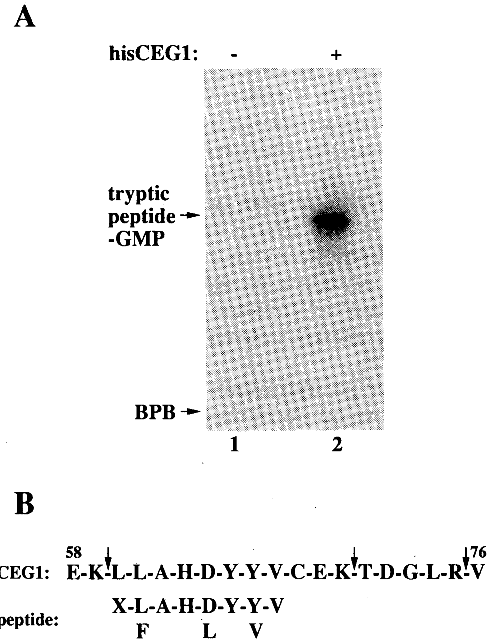


FIG. 1. Proteolytic mapping of the guanylylated peptide. (*A*) Tryptic digestions containing no His_7-CEG1 (hisCEG1) protein (−, lane 1) or purified His_7-CEG1 protein (+, lane 2) were electrophoresed on a 40% acrylamide alkaline gel. The mobilities of the tryptic ^{32}P-labeled peptide–GMP complex (tryptic peptide–GMP) and of the bromphenol blue dye (BPB) are indicated. (*B*) Peptide sequence in single-letter amino acid code obtained from the guanylylated peptide is aligned beneath the corresponding amino acid sequence of CEG1. X represents a residue of undetermined identity. Numbers specify the amino acid position within CEG1. Predicted sites of trypsin cleavage are indicated by arrows (see text).

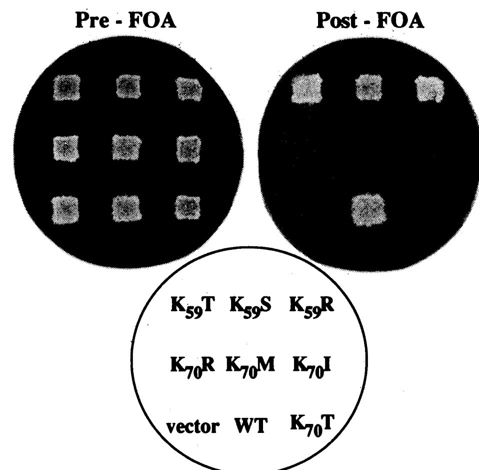


FIG. 2. Substitutions for Lys-70 are not tolerated *in vivo*. (*Upper*) Viability of *CEG1* mutant yeast strains was tested before (Pre-FOA) and after (Post-FOA) removal of the wild-type (WT) *CEG1* gene by plasmid shuffling. (*Lower*) A key indicating the allele of *CEG1* is presented, where subscript numbers indicate the relevent amino acid residue of CEG1, the single-letter amino acid code indicates the WT (first letter) and the substituted (second letter) residues, and "vector" represents the plasmid lacking any *CEG1* insert.

containing either the WT gene or mutant gene carrying a replacement for Lys-59 (Lys-59 → Thr, Ser, or Arg) fully complemented the *CEG1* null allele. In contrast, strains that harbor the pRS315 vector or replacements for Lys-70 (Lys-70 → Met, Ile, or Thr) were inviable. Even the conservative replacement of lysine by arginine abolished *CEG1* function *in vivo*.

Substitutions at Lys-70 Abrogate Guanylylation *in Vitro*. Substitutions of the active-site lysine abolish nucleotidyl–enzyme complex formation in vaccinia virus guanylyltransferase and DNA ligase (28), T4 RNA ligase (38), and human DNA ligase I (39). We tested the ability of recombinant His$_7$-CEG1 proteins substituted at Lys-70 to be guanylylated *in vitro*. *E. coli* extracts containing WT His$_7$-CEG1 or proteins substituted at position 70 with isoleucine or threonine, as well as a control extract, were passed through Ni^{2+}-NTA-agarose. The partially purified proteins were resolved by gel electrophoresis and visualized by Coomassie blue staining (Fig. 3*A*, lanes 3–5). Purified His$_7$-CEG1 was used as a size standard (lane 2). The mutant proteins were produced at levels equal to the WT protein, suggesting that the amino acid substitutions did not cause gross disturbances in folding or stability *in vivo*. The proteins were assayed for E-GMP complex formation by using [α-^{32}P]GTP (Fig. 3*B*). In contrast to the WT protein (lanes 4 and 8), guanylylation of the mutant CEG1 proteins was undetectable (lanes 9 and 10) even upon longer exposure of the gel (data not shown). As expected, no activity was evident in the control extract (lane 7). Therefore, Lys-70 is essential for guanylylation of the *S. cerevisiae* capping enzyme *in vitro*.

DISCUSSION

We present several lines of evidence demonstrating that Lys-70 is the guanylylated residue in the *S. cerevisiae* capping enzyme. A tryptic digest of the E-GMP complex generated one guanylylated peptide containing a single lysine residue at position 70. CEG1 mutant proteins substituted at this lysine were unable to function *in vivo* and were not guanylylated *in vitro*. Several evolutionary arguments support the hypothesis that Lys-70 is the active-site residue. A lysine is the guanylylated residue in the rat, vaccinia virus, and reovirus guanylyltransferases (13, 20, 21). Also, in the guanylyltransferase of *Sch. pombe*, the position corresponding to Lys-70 is conserved (see Fig. 4). Furthermore, the sequence surrounding Lys-70 conforms to a motif that appears to be conserved in a family of nucleotidyltransferases (see below). Therefore, biochemical, genetic, and evolutionary data all indicate that Lys-70 is the active-site residue of the *S. cerevisiae* mRNA-capping enzyme guanylyltransferase.

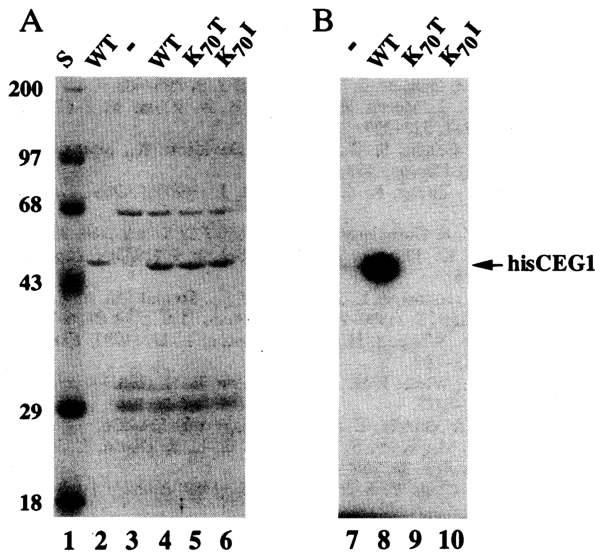


FIG. 3. CEG1 mutant proteins with replacements for Lys-70 (shown as in Fig. 2) are not guanylylated *in vitro*. (*A*) Coomassie blue-stained gel containing protein molecular weight standards (S, lane 1), purified His$_7$-CEG1 (hisCEG1) (WT, lane 2), Ni^{2+}-NTA-column eluates of bacterial extracts producing no His$_7$-CEG1 protein (−, lane 3), WT His$_7$-CEG1 protein (lane 4), His$_7$-CEG1 with Thr-70 (K$_{70}$T, lane 5), or His$_7$-CEG1 with Ile-70 (K$_{70}$I, lane 6). (*B*) E-GMP complex formation in Ni^{2+}-NTA-column eluates containing no His$_7$-CEG1 protein (−, lane 7), WT His$_7$-CEG1 protein (lane 8), His$_7$-CEG1 with Thr-70 (K$_{70}$T, lane 9), or His$_7$-CEG1 with Ile-70 (K$_{70}$I, lane 10). For both panels, the sizes of the molecular weight standards in kDa and the mobility of His$_7$-CEG1 are indicated.

Enzyme	Start	Sequence	End
Cellular Guanylyltransferases (GTP)			
S. cerevisiae	63	H D Y Y V C E **K** T **D G** L R V L M F I	80
S. pombe	60	K N Y F V C E **K** S **D G** I R C L L Y M	77
Viral Guanylyltransferases (GTP)			
***Vaccinia**	253	E N L Y A V T **K** T **D G** I P I T I R V	270
SFV	249	T N L Y V T T **K** T **D G** V G V L I T V	266
ASFV	275	V G Y Y V T D **K** A **D G** I R G I A V I	292
***Reovirus**	219	G V L V H Y D **K** P T **N G** H H Y L L G T	237
RNA Ligases (ATP)			
***T4 RNA**	93	V D Y I L T **K** E **D G** S L V S T Y L	109
S. cerevisiae* **tRNA	107	G P Y D V T I **K** A **N G** C I I F I S G	124
DNA Ligases (ATP)			
***human**	561	A A F T C E Y **K** Y **D G** Q R A Q I H A	578
***bovine**		A A F T C Q Y **K** T **D G** Q R	
S. pombe	409	A A F T C E Y **K** Y **D G** E R A Q V H F	426
S. cerevisiae	412	E T F T S E Y **K** Y **D G** E R A Q V H L	429
ASFV	144	T D P I V Q R **K** R **N G** V R A V A C Q	161
ASFV	144	T D P I V H G **K** R **N G** V R A V A C Q	161
***Vaccinia**	224	S G M F A E V **K** Y **D G** E R V Q V H K	241
D. ambivalens	254	N I A L V D Y **K** Y **D G** E R G Q I H K	271
T7	27	G Y L I A E I **K** Y **D G** V R G N I C V	44
T3	27	G Y L I A D C **K** Y **D G** V R G N I V V	44
T6	152	F P A F A Q L **K** A **D G** A R C F A E V	169
T4	152	F P A F A Q L **K** A **D G** A R C F A E V	169
DNA Ligases (NAD+)			
E. coli	108	V T W C C E L **K** L **D G** L A V S I L Y	125
Z. mobilis	137	V I C T V E P **K** I **D G** L S C S L R Y	154
T. thermophilus	111	F A Y T V E H **K** V **D G** L S V N L Y Y	128

FIG. 4. Sequence alignment of known and putative nucleotidyl attachment active sites of mRNA guanylyltransferases and polynucleotide ligases. Letters represent the single-letter amino acid code, and numbers represent the positions of the amino acid residues. The most highly conserved amino acids of the nucleotidyltransferase motif Lys-Xaa-Asx-Gly are highlighted, and the active-site lysines are indicated in boldface type. Asterisks denote experimentally determined nucleotidyl attachment sites of *S. cerevisiae* (this paper), vaccinia virus guanylyltransferase (28, 40), reovirus guanylyltransferase (20), T4 RNA ligase (38, 41, 42), *S. cerevisiae* tRNA ligase (37), human DNA ligase I (39), bovine DNA ligase (36), and vaccinia virus DNA ligase (28). Enzymes are grouped as cellular guanylyltransferases [*S. cerevisiae* (26), *Sch. pombe* (L.D.F., Sue-Ann Woo, and S.B., unpublished data)], viral guanylyltransferases [vaccinia virus (43), Shope fibroma virus (SFV) (19), African swine fever virus (ASFV) (16), reovirus λ2 (44)], RNA ligases [bacteriophage T4 (41, 45), *S. cerevisiae* tRNA ligase (46)], or DNA ligases [human DNA ligase I (47), bovine (36), *Sch. pombe* (48), *S. cerevisiae* (49), ASFV BA71V isolate (50), ASFV Malawi Lil isolate (51), vaccinia virus (52), *Desulfurolobus ambivalens* (53), bacteriophage T7 (54), bacteriophage T3 (55), bacteriophage T6 (56), bacteriophage T4 (57), *E. coli* (58), *Zymomonas mobilis* (59), and *Thermus thermophilis* (60)]. Donor molecules for nucleotidyl transfer are indicated in parentheses after each class of enzyme.

As pointed out by Cong and Shuman (28) and by Pena *et al* (16), sequences around the known and putative active sites of mRNA guanylyltransferases exhibit a short stretch of

similarity to the active-site motif of ATP/NAD$^+$-dependent ligases (reviewed in ref. 27; see Fig. 4). The sequence Lys-Xaa-Asx-Gly is conserved without exception in both families of enzymes. In the guanylyltransferases, several other positions are at least partially conserved. The residue following the active-site lysine is usually a threonine or a serine, while tyrosine or phenylalanine appears four residues upstream. Several other positions near this motif appear to favor hydrophobic residues. Only the reovirus guanylyltransferase diverges significantly, although its active site may be aligned with the consensus sequence by allowing the insertion of an extra amino acid position into the core motif (see Fig. 4).

Sequence similarity between the active sites of guanylyltransferases and polynucleotide ligases suggests that a common catalytic mechanism is employed in the capping and ligation reactions. Formation of phosphodiester bonds by ligases proceeds by nucleophilic attack of the active-site lysine on ATP (or NAD$^+$) with formation of a covalent enzyme–adenylyl (E-AMP) intermediate and release of pyrophosphate (or NMN). The activated adenylyl moiety is transferred to the phosphate residue at the 5′ end of the first substrate nucleic acid strand, activating it for attack by the ribose 3′-OH from the second substrate polynucleotide. Thus, the AMP is released as the polynucleotides are joined (reviewed in ref. 27).

The nucleotidyl transfer mechanisms of guanylyltransferases and ATP- or NAD$^+$-dependent DNA and RNA ligases are similar in the following respects: (*i*) covalent phosphoamide intermediates are formed between the ε-amino group of lysine and the 5′-phosphate of a nucleotide, (*ii*) the transferred mononucleotides are purines. (*iii*) the nucleotidyl acceptor molecules are polynucleotides with 5′ phosphates, (*iv*) transfer of the mononucleotides to the substrates generates 5′,5′-phosphoric anhydride bridge structures (GpppN. . . or AppN. . .), and (*v*) divalent cations Mg^{2+} or Mn^{2+} are required for activity.

The parallels between the capping and ligation reactions suggest that the similarities in active-site sequences represent true homology. The motif proposed for nucleotidyl attachment in the DNA and RNA ligases can be extended to include guanylyltransferases. One can only speculate whether these sequence similarities represent divergent evolution from an ancestral nucleotidyl transferase or convergent evolution due to constraints on active-site structure imposed by the chemistry of nucleotidyl transfer.

We thank Sue-Ann Woo for assistance with dideoxy sequencing reactions and Jun Sakai and Paul Matsudaira of the Whitehead Institute Microsequencing Facility for peptide sequencing and analysis. We thank David Bartel, Robin Buratowski, Paul Matsui, Charles Query, and Sue-Ann Woo for critical reading of the manuscript. L.D.F. was supported by a postdoctoral fellowship from the American Cancer Society (PF-3505). This work was supported by a grant (NP-856) from the American Cancer Society to S.B. The Whitehead Institute Fellows program (S.B.) is supported by grants from the Lucille P. Markey Charitable Trust and the W. M. Keck Foundation.

1. Hagler, J. & Shuman, S. (1992) *Science* **255**, 983–986.
2. Rasmussen, E. B. & Lis, J. T. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7923–7927.
3. Mizumoto, K. & Kaziro, Y. (1987) *Prog. Nucleic Acid Res. Mol. Biol.* **34**, 1–28.
4. Furuichi, Y. & Shatkin, A. J. (1989) *Methods Enzymol.* **180**, 164–176.
5. Ensinger, M. J. & Moss, B. (1976) *J. Biol. Chem.* **251**, 5283–5291.
6. Mizumoto, K. & Lipmann, F. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4961–4965.
7. Langberg, S. R. & Moss, B. (1981) *J. Biol. Chem.* **256**, 10054–10060.
8. Shuman, S. & Moss, B. (1990) *Methods Enzymol.* **181**, 170–180.
9. Shuman, S. (1982) *J. Biol. Chem.* **257**, 7237–7245.
10. Mizumoto, K., Kaziro, Y. & Lipman, F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1693–1697.
11. Mao, Z. & Joklik, W. K. (1991) *Virology* **185**, 377–386.
12. Shuman, S. & Hurwitz, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 187–191.
13. Toyama, R., Mizumoto, K., Nakahara, Y., Tatsuno, T. & Kaziro, Y. (1983) *EMBO J.* **2**, 2195–2201.
14. Shatkin, A. J., Furuichi, Y., LaFiandra, A. J. & Yamakawa, M. (1983) in *Double-Stranded RNA Viruses*, ed. Bishop, J. M. (Elsevier, New York), pp. 43–54.
15. Cleveland, D. R., Zarbl, H. & Millward, S. (1986) *J. Virol.* **60**, 307–311.
16. Pena, L., Yañez, R. J., Revilla, Y., Viñuela, E. & Salas, M. L. (1993) *Virology* **193**, 319–328.
17. Pizarro, J. L., Sandino, A. M., Pizarro, J. M., Fernandez, J. & Spencer, E. (1991) *J. Gen. Virol.* **72**, 325–332.
18. le Blois, H., French, T., Mertens, P. P. C., Burroughs, J. N. & Roy, P. (1992) *Virology* **189**, 757–761.
19. Upton, C., Stuart, D. & McFadden, G. (1991) *Virology* **183**, 773–777.
20. Fausnaugh, J. & Shatkin, A. J. (1990) *J. Biol. Chem.* **265**, 7669–7672.
21. Roth, M. J. & Hurwitz, J. (1984) *J. Biol. Chem.* **259**, 13488–13494.
22. Venkatesan, S. & Moss, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 340–344.
23. Shabarova, Z. A. (1970) *Prog. Nucleic Acid Res. Mol. Biol.* **10**, 145–182.
24. Wang, D. & Shatkin, A. J. (1984) *Nucleic Acids Res.* **12**, 2303–2315.
25. Itoh, N., Yamada, H., Kaziro, Y. & Mizumoto, K. (1987) *J. Biol. Chem.* **262**, 1989–1995.
26. Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S. & Mizumoto, K. (1992) *J. Biol. Chem.* **267**, 9521–9528.
27. Lindahl, T. & Barnes, D. E. (1992) *Annu. Rev. Biochem.* **61**, 251–281.
28. Cong, P. & Shuman, S. (1993) *J. Biol. Chem.* **268**, 7256–7260.
29. Sikorski, R. S. & Hieter, P. (1989) *Genetics* **122**, 19–27.
30. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1991) *Current Protocols in Molecular Biology* (Wiley, New York).
31. Gietz, D., St. Jean, A., Woods, R. A. & Schiestl, R. H. (1992) *Nucleic Acids Res.* **20**, 1425.
32. Guthrie, C. & Fink, G. R. (1991) *Methods Enzymol.* **194**, 1–933.
33. Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1074–1078.
34. Dadd, C. A., Cook, R. G. & Allis, C. D. (1993) *BioTechniques* **14**, 266–273.
35. West, M. H. P., Wu, R. S. & Bonner, W. M. (1984) *Electrophoresis* **5**, 133–138.
36. Tomkinson, A. E., Totty, N. F., Ginsburg, M. & Lindahl, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 400–404.
37. Xu, Q., Teplow, D., Lee, T. D. & Abelson, J. (1990) *Biochemistry* **29**, 6132–6138.
38. Heaphy, S., Singh, M. & Gait, M. J. (1987) *Biochemistry* **26**, 1688–1696.
39. Kodama, K., Barnes, D. E. & Lindahl, T. (1991) *Nucleic Acids Res.* **19**, 6093–6099.
40. Niles, E. G. & Christen, L. (1993) *J. Biol. Chem.* **268**, 24986–24989.
41. Rand, K. N. & Gait, M. J. (1984) *EMBO J.* **3**, 397–402.
42. Thøgerson, H. C., Morris, H. R., Rand, K. N. & Gait, M. J. (1985) *Eur. J. Biochem.* **147**, 325–329.
43. Niles, E. G., Condit, R. C., Caro, P., Davidson, K., Matusick, L. & Seto, J. (1986) *Virology* **153**, 96–112.
44. Seliger, L. S., Zheng, K. & Shatkin, A. J. (1987) *J. Biol. Chem.* **262**, 16289–16293.
45. Koonin, E. V. & Gorbalenya, A. E. (1990) *FEBS Lett.* **268**, 231–234.
46. Westaway, S. K., Phizicky, E. M. & Abelson, J. (1988) *J. Biol. Chem.* **263**, 3171–3176.
47. Barnes, D. E., Johnston, L. H., Kodama, K., Tomkinson, A. E., Lasko, D. D. & Lindahl, T. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6679–6683.
48. Barker, D. G., White, J. H. M. & Johnston, L. H. (1987) *Eur. J. Biochem.* **162**, 659–667.
49. Barker, D. G., White, J. H. M. & Johnston, L. H. (1985) *Nucleic Acids Res.* **13**, 8323–8337.
50. Yañez, R. J. & Vinuela, E. (1993) *Virology* **193**, 531–536.
51. Hammond, J. M., Kerr, S. M., Smith, G. L. & Dixon, L. K. (1992) *Nucleic Acids Res.* **20**, 2667–2671.
52. Smith, G. L., Chan, Y. S. & Kerr, S. M. (1989) *Nucleic Acids Res.* **17**, 9051–9062.
53. Kletzin, A. (1992) *Nucleic Acids Res.* **20**, 5389–5396.
54. Dunn, J. J. & Studier, F. W. (1981) *J. Mol. Biol.* **148**, 303–330.
55. Schmitt, M. P., Beck, P. J., Kearney, C. A., Spence, J. L., DiGiovanni, D., Condreay, J. P. & Molineux, I. J. (1987) *J. Mol. Biol.* **193**, 479–495.
56. Kaliman, A. V., Zimin, A. A., Nazipova, N. N., Kryukov, V. M., Tanyashin, V. I., Kraev, A. S., Mironova, M. V., Skyrabin, K. G. & Baev, A. A. (1988) *Dokl. Akad. Nauk SSSR* **299**, 737–742.
57. Armstrong, J., Brown, R. S. & Tsugita, A. (1983) *Nucleic Acids Res.* **11**, 7145–7156.
58. Ishino, Y., Shinagawa, H., Makino, K., Tsunasawa, S., Sakiyama, F. & Nakata, K. (1986) *Mol. Gen. Genet.* **204**, 1–7.
59. Shark, K. B. & Conway, T. (1992) *FEMS Microbiol. Lett.* **75**, 19–26.
60. Lauer, G., Rudd, E. A., McKay, D. L., Ally, A., Ally, D. & Backman, K. C. (1991) *J. Bacteriol.* **173**, 5047–5053.