Nucleic Acids Research, 1999, Vol. 27, No. 24 **4671–4678**

An essential surface motif (WAQKW) of yeast RNA triphosphatase mediates formation of the mRNA capping enzyme complex with RNA guanylyltransferase

C. Kiong Ho, Kevin Lehman and Stewart Shuman*

Molecular Biology Program, Sloan-Kettering Institute, New York, NY 10021, USA

Received September 21, 1999; Revised and Accepted October 28, 1999

ABSTRACT

Saccharomyces cerevisiae **RNA triphosphatase (Cet1p) and RNA guanylyltransferase (Ceg1p) interact *in vivo* and *in vitro* to form a bifunctional mRNA capping enzyme complex. Cet1p binding to Ceg1p stimulates the guanylyltransferase activity of Ceg1p. Here we localize the guanylyltransferase-binding and guanylyltransferase-stimulation functions of Cet1p to a 21-amino acid segment from residues 239 to 259. The guanylyltransferase-binding domain is located on the protein surface, as gauged by protease sensitivity, and is conserved in the *Candida albicans* RNA triphosphatase CaCet1p. Alanine-cluster mutations of a WAQKW motif within this segment abolish guanylyltransferase-binding *in vitro* and Cet1p function *in vivo*, but do not affect the triphosphatase activity of Cet1p. Proteolytic footprinting experiments provide physical evidence that Cet1p interacts with the C-terminal domain of Ceg1p. Trypsin-sensitive sites of Ceg1p that are shielded from proteolysis when Ceg1p is bound to Cet1p are located between nucleotidyl transferase motifs V and VI.**

INTRODUCTION

Capping of mRNA entails three enzymatic reactions in which the 5′ triphosphate end of pre-mRNA is hydrolyzed to a 5′ diphosphate by RNA triphosphatase, then capped with GMP by RNA guanylyltransferase, and methylated by RNA (guanine-7) methyltransferase (1). The budding yeast *Saccharomyces cerevisiae* encodes a three-component capping apparatus consisting of separate triphosphatase (Cet1p), guanylyltransferase (Ceg1p) and methyltransferase (Abd1p) gene products. Each of the cap-forming activities is essential for yeast cell growth (2–7).

The yeast RNA triphosphatase Cet1p forms a heteromeric complex with the yeast RNA guanylyltransferase Ceg1p (8–10). The binding of Cet1p to Ceg1p serves two purposes. First, Cet1p–Ceg1p interaction stimulates guanylyltransferase activity by enhancing the affinity of Ceg1p for GTP and increasing the extent of formation of the Ceg1p–GMP reaction intermediate (9). Second, the physical tethering of Cet1p to Ceg1p facilitates recruitment of the triphosphatase to the RNA polymerase II elongation complex. The yeast guanylyltransferase Ceg1p binds to the phosphorylated C-terminal domain (CTD) of the largest subunit of RNA polymerase II, whereas Cet1p by itself does not bind to the phosphorylated CTD (11–13).

The 549-amino acid Cet1p protein consists of three domains: (i) a 230-amino acid N-terminal segment that is dispensable for catalysis *in vitro* and for Cet1p function *in vivo*; (ii) a protease-sensitive segment from residues 230 to 275 that is dispensable for catalysis, but essential for Cet1p function *in vivo*; and (iii) a catalytic domain from residues 275 to 539 (10). A homodimeric quaternary structure for the biologically active fragment Cet1(231–549)p was suggested based on analysis of the purified recombinant enzyme by glycerol gradient sedimentation (10). Cet1(231–549)p binds *in vitro* to Ceg1p to form a 7.1S triphosphatase–guanylyltransferase complex that is surmised to be a trimer consisting of two molecules of Cet1(231–549)p and one molecule of Ceg1p. The more extensively truncated protein Cet1(276–549)p, which cannot support cell growth, sediments as a monomer and does not interact with Ceg1p (10). These results implicate the segment of Cet1p from residues 230 to 275 in both Cet1p homodimerization and binding to the guanylyltransferase. The interaction of Cet1p with Ceg1p does not require a functional triphosphatase active site in Cet1p, insofar as Cet1p–Ceg1p complex formation is unaffected by mutations in the catalytic domain that abrogate RNA triphosphatase activity (14).

Two lines of genetic evidence indicate that the Cet1p–Ceg1p interaction is important *in vivo*. First, several temperature-sensitive *ceg1* mutations are suppressed in an allele-specific manner by overexpression of *CET1* (9,13). In turn, the temperature-sensitive *cet1*-(*K250A-W251A*) mutation can be suppressed by overexpression of *CEG1* (10). This *cet1-ts* mutation is located with the segment of Cet1p that is suspected to mediate guanylyltransferase-binding. Fifteen other *cet1-ts* alleles with missense changes mapping elsewhere in the protein were not suppressed by *CEG1* overexpression (10). Second, the *in vivo* function of Cet1(275–549)p, which does not bind to Ceg1p *in vitro*, is completely restored by fusion of Cet1(275–549)p to the guanylyltransferase domain of the mouse capping enzyme (10). This result shows that the need for Ceg1p-binding by yeast RNA triphosphatase can by bypassed when the triphosphatase catalytic domain is delivered to the RNA polymerase II elongation complex by linkage in *cis* to the mammalian guanylyltransferase.

*To whom correspondence should be addressed. Tel: +1 212 639 7145; Fax: +1 212 717 3623; Email: s-shuman@ski.mskcc.org

Here we use synthetic peptide ligands to localize the guanylyltransferase-binding and guanylyltransferase-stimulation functions of Cet1p. Alanine cluster mutations implicate the Cet1p sequence 247–WAQKW–251 in guanylyltransferase-binding *in vitro* and in Cet1p function *in vivo*. The Ceg1p-binding domain of Cet1p is conserved in *Candida albicans* RNA triphosphatase. Indeed, the Cet1p peptide binds avidly *in vitro* to the *C.albicans* RNA guanylyltransferase and binding is abrogated by mutation of the WAQKW motif.

MATERIALS AND METHODS

Guanylyltransferase expression and purification

The *S.cerevisiae CEG1* gene was inserted into a customized T7-based expression plasmid (a derivative of pET16b) in such a way as to fuse the 459-amino acid Ceg1p polypeptide in frame to an N-terminal 29-amino acid leader peptide (MGSH-HHHHHHHHHSSGHIEGRHSRRASVH) containing 10 consecutive histidine codons (His-tag) and a serine-phosphorylation site (RRASV) for protein kinase A. Recombinant Ceg1p was expressed in *Escherichia coli* BL21(DE3) and purified from a soluble bacterial lysate by nickel–agarose chromatography as described (10). The *C.albicans CGT1* gene was PCR-amplified from a genomic library clone using a sense primer designed to introduce an *Nde*I site at the translation start codon and an antisense primer that introduced a *Bam*HI site immediately 3′ of the stop codon. The PCR product was digested with *Nde*I and *Bam*HI and then inserted into yeast expression plasmid pYN132 to yield plasmid pYN-Cgt1. The *CGT1* gene was excised from pYN-Cgt1 with *Nde*I and *Bam*HI and then inserted into the T7-based expression vector pET16b so as to fuse the 449-amino acid Cgt1p polypeptide to a N-terminal 21-amino acid leader peptide containing the His-tag. Recombinant Cgt1p was expressed in *E.coli* BL21(DE3) and purified from a soluble bacterial lysate by nickel–agarose chromatography. The nickel–agarose preparations of Cet1p and Cgt1p were dialyzed against buffer containing 50 mM Tris–HCl (pH 8.0), 50 mM NaCl, 2 mM DTT, 10% glycerol, 0.05% Triton X-100 and then stored at –80°C. Recombinant mouse guanylyltransferase Mce1(211–597)p was purified as described (15). Protein concentrations were determined using the Bio-Rad dye binding reagent with bovine serum albumin as the standard.

Cet1 peptides

Peptides composed of Cet1p residues 232–265, 239–265 or 232–259 were synthesized in the Sloan-Kettering Microchemistry Core Laboratory on a Perkin-Elmer Biosystems 431A automated peptide synthesizer using standard Fmoc chemistry. Addition of biotin to the N-terminal amino group of the peptide was performed using reagents purchased from AnaSpec (San Jose, CA) according to the vendor's instructions. The peptides were purified on a preparative scale by reverse phase HPLC and the purity of the material was confirmed by analytic scale reverse phase HPLC as described (16). The molecular weight of each peptide was analyzed by MALDI-TOF mass spectrometry. The measured masses were in agreement with the calculated theoretical masses within the limits of calibration of the instrument. The lyophilized peptides were dissolved in TE (10 mM Tris–HCl, pH 8.0, 1 mM EDTA) and stored at 4°C. The UV absorbance profiles of the Cet1 peptides revealed a large peak at <220 nM and a minor peak at 274–280 nM (the latter being attributable to tyrosine and tryptophan in the peptides). The molar concentrations of the peptides were calculated from the absorbance at 274 nM using the extinction coefficients of 1.4×10^3 M^{-1} for free tyrosine and 5.6×10^3 M^{-1} for free tryptophan.

Peptide-affinity chromatography

The biotinylated Cet1 peptides (1 nmol) were adsorbed to 0.6 mg of streptavidin coated magnetic beads (Dynabeads M280 streptavidin; Dynal) in 50 µl of binding buffer (25 mM Tris–HCl, pH 8.0, 50 mM NaCl, 1 mM DTT, 5% glycerol, 0.03% Triton X-100). The Dynabeads have a binding capacity of 650–900 pmol of free biotin per mg according to the manufacturers specifications; hence the amount of biotinylated Cet1 peptide was sufficient to saturate the bead-bound streptavidin. After incubation for 30 min on ice, the beads were concentrated using a horseshoe magnet and then washed three times with 0.5 ml of binding buffer to remove any unbound peptide.

Affinity chromatography was performed by mixing 4 µg of guanylyltransferase with 0.6 mg of Cet1 peptide beads (estimated to contain 390–540 pmol of peptide) in 50 µl of binding buffer. After incubation for 20–30 min on ice, the beads were concentrated by microcentrifugation for 15 s and then held in place with a magnet as the supernatant was withdrawn. The beads were resuspended in 0.5 ml of binding buffer and subjected to two cycles of concentration and washing. After the third wash, the beads were resuspended in 50 µl of binding buffer. Aliquots (20 µl) of the input protein sample, the unbound supernatant fraction and the bead bound fraction were mixed with 3 µl of SDS sample buffer (200 mM Tris–HCl, pH 6.8, 8% SDS, 140 nM β-mercaptoethanol, 40% glycerol), heated at 90°C for 3 min and then analyzed by electrophoresis through a 15% polyacrylamide gel containing 0.1% SDS. Polypeptides were visualized by staining the gel with Coomassie blue dye.

Guanylyltransferase assay

Guanylyltransferase activity was assayed by the formation of a covalent Ceg1-GMP intermediate (9). Reaction mixtures (20 µl) containing 50 mM Tris–HCl, pH 8.0, 5 mM $MgCl_2$, 0.17 µM [α-^{32}P]GTP, Ceg1p and Cet1 peptides as specified were incubated for 10 min at 37°C. The reaction was halted by addition of SDS to 1% final concentration. The samples were analyzed by SDS–PAGE. Transfer of labeled GMP to Ceg1p was visualized by autoradiography of the dried gel and quantitated by scanning the gel with a FUJIX BAS2500 phosphorimager.

Mutagenesis of yeast RNA triphosphatase

Alanine-cluster mutations were introduced into the *CET1(201–549)* gene by PCR as described (7,10). The mutated genes were inserted into the yeast *CEN TRP1* plasmid pCET1-5′3′, where expression of the inserted gene is under the control of the natural *CET1* promoter. The inserts were sequenced completely to exclude the acquisition of unwanted mutations during amplification and cloning. The *W247A-Q249A* and *K250A-W251A* genes were excised from their respective pCET1-5′3′ plasmids with *Nde*I and *Bam*HI and inserted into the yeast expression vector pYN132 (*CEN TRP1*).

In this vector, expression of *CET1(201–549)* is driven by the strong constitutive yeast *TPI1* promoter. The *in vivo* activity of the mutated *CET1* alleles was tested by plasmid shuffle as described (9). Yeast strain YBS20 (*trp1 ura3 leu2 cet1::LEU2* p360-Cet1[*CEN URA3 CET1*]) was transformed with *CEN TRP1* plasmids containing the wild-type and mutant alleles of *CET1(201–549)*. A control transformation was performed using the *TRP1* vector. Trp+ isolates were selected and then streaked on agar plates containing 0.75 mg/ml of 5-fluoroorotic acid (5-FOA). Growth was scored after 7 days of incubation at 25, 30 and 37°C. Lethal mutants were those that failed to form colonies on 5-FOA at any temperature.

*Nde*I–*Bam*HI fragments encoding mutated versions of Cet1(201–549)p were excised from the respective pCET1-5′3′ plasmids and inserted into pET16b. Wild-type Cet1(201–549)p and the W247A-Q249A mutant were expressed in *E.coli* BL21(DE3) at 37°C by IPTG-induction for 3 h. The K250-W251A mutant was expressed at 17°C by IPTG-induction for 20 h in the presence of 2% ethanol (14). The proteins were purified from soluble bacterial lysates by nickel–agarose chromatography as described (7,14). The 0.2 M imidazole eluate fractions containing Cet1(201–549)p were dialyzed against 50 mM Tris–HCl (pH 8.0), 50 mM NaCl, 2 mM DTT, 10% glycerol, 0.05% Triton X-100 and then stored at –80°C.

Triphosphatase reaction mixtures (10 μl) containing 50 mM Tris–HCl (pH 7.5), 5 mM DTT, 2 mM $MnCl_2$, 1 mM [γ-^{32}P]ATP and Cet1(201–549)p as specified were incubated for 15 min at 30°C. The reactions were quenched by adding 2.5 μl of 5 M formic acid. An aliquot (2.5 μl) of the mixture was applied to a polyethyleneimine-cellulose TLC plate, which was developed with 0.5 M LiCl and 1 M formic acid. The release of ^{32}Pi from [γ-^{32}P]ATP was quantitated by scanning the TLC plate with a phosphorimager.

Proteolytic footprinting of kinase-tagged Ceg1p

Ceg1p containing an N-terminal kinase tag was ^{32}P end-labeled using protein kinase A, which specifically phosphorylates the serine in the RRASV sequence of the N-terminal tag. Kinase reaction mixtures (100 μl) containing 50 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 0.2 mM [γ-^{32}P]ATP, 50 μg of Ceg1p and 5 U of protein kinase A (38 kDa catalytic subunit; New England Biolabs) were incubated for 5 min at 30°C. The mixture was adjusted to 50 mM EDTA and placed on ice. The labeled Ceg1p was resolved from free ATP by gel filtration of the sample through a 1-ml column of Sephadex G-50 at 4°C. The radiochemical purity of the ^{32}P-Ceg1p eluting in the void volume was confirmed by SDS–PAGE. The concentration of ^{32}P-Ceg1p was calculated from the signal intensity of the ^{32}P-Ceg1p protein and the specific activity of the ATP substrate used in the kinase reaction.

Footprinting reactions were performed as follows. Reaction mixtures (15 μl) containing 50 mM Tris–HCl (pH 8.0), 0.5 μg of ^{32}P-Ceg1p, 4 μg of either Cet1(201–549)p, Ceg1p or bovine serum albumin, and 0, 1, 5 or 10 ng of trypsin were incubated for 15 min at 22°C. The reaction was quenched by the addition of SDS sample buffer and the digests were analyzed by electrophoresis through a 15% polyacrylamide gel containing 0.1% SDS. ^{32}P-labeled tryptic fragments were visualized by autoradiography.

RESULTS

A synthetic peptide, Cet1(232–265), binds to Ceg1p and stimulates guanylyltransferase activity

Overlapping segments of Cet1p from residues 205 to 265 and 230 to 275 have been implicated in Ceg1p-binding using a variety of biochemical and genetic assays (10,13). To better understand the structural basis for Cet1p–Ceg1p interaction and the attendant stimulation of guanylyltransferase activity, we prepared a synthetic peptide ligand containing the 34-amino acid segment from Cet1p residues 232–265 (Fig. 1). An N-terminal biotin was added during chemical synthesis so that the peptide could be linked to streptavidin beads for affinity chromatography purposes. Peptide-containing beads and control streptavidin beads lacking peptide were then incubated with purified recombinant yeast guanylyltransferase Ceg1p (guanylyltransferase concentration 1.4 μM) in buffer containing 50 mM NaCl. The beads were recovered by centrifugation and held in place with a magnet while the supernatant containing free guanylyltransferase was withdrawn. The beads were washed with buffer containing 50 mM NaCl. The bead-bound material was eluted from the beads with 1% SDS. The input guanylyltransferase protein (L) and the free (F) and bead-bound (B) fractions were then analyzed by SDS–PAGE. The salient finding was that the Ceg1p bound nearly quantitatively to beads containing the Cet1(232–265) peptide, but not at all to streptavidin-alone control beads (Fig. 1A) or to beads containing a biotinylated 28mer peptide composed of four tandem copies of the CTD heptamer YSPTSPS (not shown). Note that the streptavidin polypeptide was stripped off the beads by 1% SDS and recovered in the bound eluate fraction (Fig. 1A). This experiment shows that a defined synthetic peptide can be used to study the Cet1p–Ceg1p interaction and that the Cet1p segment from 232 to 265 suffices for binding to the yeast guanylyltransferase.

Previously, we showed that the binding of full-length Cet1p or Cet1(201–549)p to Ceg1p stimulates the guanylyltransferase activity of Ceg1p (9). Here we tested the effects of the Cet1(232–265) peptide on guanylyltransferase activity. Ceg1p was incubated for 15 min on ice with various concentrations of the peptide and aliquots of the mixtures were assayed for enzyme–GMP complex formation. Ceg1p–GMP formation was stimulated ~11-fold by 50 nM peptide (Fig. 1B). The fold-stimulation was proportional to peptide concentration in the range 12–50 nM and plateaued at 50–100 nM. Maximal effect was attained at an ~2:1 molar ratio of peptide to guanylyltransferase. We conclude that binding of Cet1p to Ceg1p and stimulation of guanylyltransferase activity are mediated by the same small peptide domain.

Effects of mutations within the Ceg1p-binding domain on Cet1p function *in vivo*

The sequence of the guanylyltransferase-binding peptide of *S.cerevisiae* RNA triphosphatase (amino acids 232–265) is conserved in the RNA triphosphatase of *C.albicans* (Fig. 2A). Previously, we reported an analysis of the effects of several alanine cluster mutations within this conserved region on Cet1p function *in vivo* (10). Yeast cells expressing *CET1(201–549)* alleles *K237A-P238A*, *K240A-Y241A* and *P245A-I246A*

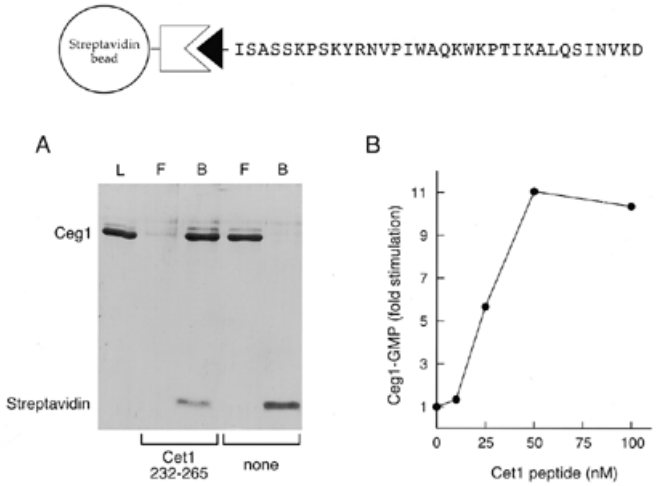

Figure 1. Synthetic peptide Cet1(232–265) binds to Ceg1p and stimulates guanylyltransferase activity. The sequence of the Cet1(232–265) peptide is shown. An N-terminal biotin anchors the peptide to a streptavidin-coated magnetic bead. (**A**) Affinity chromatography. Ceg1p was incubated with either streptavidin beads alone (none) or beads with attached Cet1(232–265) peptide. Chromatography was performed as described in Materials and Methods. Aliquots of the input guanylyltransferase protein fraction (L) (equivalent to 40% of total material loaded), free-unbound fraction (F) (40% of supernatant) and the bead-bound fraction (B) (40% of SDS eluate) were analyzed by SDS–PAGE. A Coomassie blue stained gel is shown. The positions of Ceg1p and streptavidin are indicated on the left. (**B**) Guanylyltransferase activity. Ceg1p (300 ng) was preincubated for 15 min on ice with either 0, 2, 5, 10 or 20 pmol of Cet1(232–265) peptide in 10 µl of binding buffer. An aliquot (1 µl) of each sample was then assayed for guanylyltransferase activity. The signal intensities of the Ceg1p–GMP complexes were normalized to the signal intensity of the 'no peptide' control reaction (defined as 1.0). The fold stimulation is plotted as a function of the final concentration of Cet1(232–265) peptide in the guanylyltransferase reaction mixture, which contained 25 nM Ceg1p.

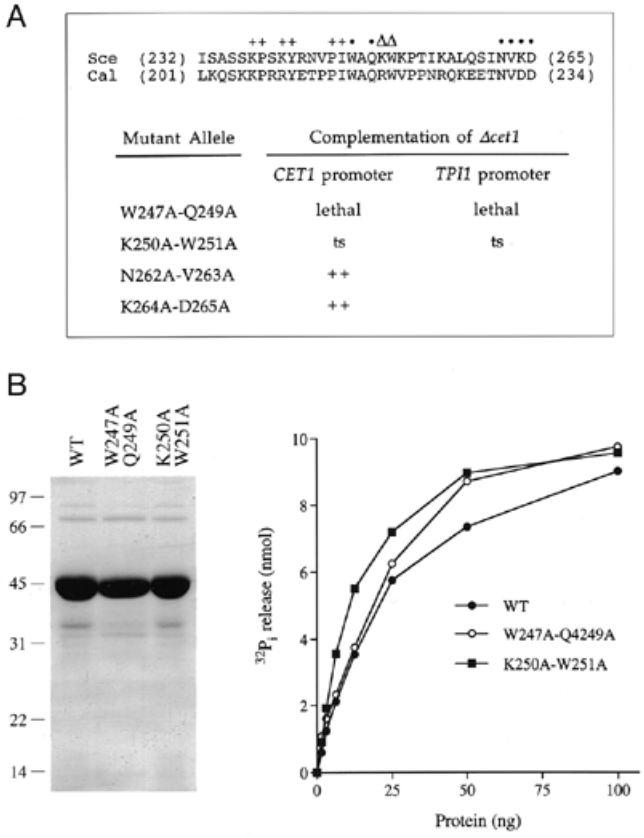

Figure 2. Effects of Ala-cluster mutations within the Ceg1p-binding domain on Cet1p function. (**A**) *In vivo* mutational effects. The amino acid sequence of *S.cerevisiae* (Sce) Cet1p from residues 232 to 265 is aligned with the homologous segment of *C.albicans* (Cal) CaCet1p. Amino acid pairs in Cet1p that were mutated to alanine in a prior study (10) and found to have no effect on *CET1* function *in vivo* are denoted by + above the sequence. The K250A-W251A mutation eliciting a temperature-sensitive phenotype is denoted by Δ. Paired residues targeted for double-alanine replacement in the present study are indicated by dots. The indicated Ala-cluster alleles of *CET1(201–549)* were tested by plasmid shuffle for growth complementation of *cet1*Δ cells on medium containing 5-FOA. Expression of the mutant genes was driven either by the natural *CET1* promoter or, where indicated, by the strong constitutive *TPI1* promoter. (**B**) Triphosphatase activity. (Left) Aliquots (4 µg) of the dialyzed nickel–agarose preparations of wild-type Cet1(201–549)p and mutants W247A-Q249A and K250A-W251A were electrophoresed through a 12.5% polyacrylamide gel containing 0.1% SDS. Polypeptides were visualized by staining with Coomassie blue dye. The positions and sizes (in kDa) of marker proteins are indicated on the left. (Right) ATPase activity was assayed as described in Materials and Methods. Release of ^{32}Pi from [γ-^{32}P]ATP is plotted as a function of input protein for each enzyme preparation.

grew as well as wild-type *CET1(201–549)* cells at all temperatures tested (the residues mutated are denoted by + above the Cet1p sequence in Fig. 2A), whereas *K250A-W251A* cells displayed a temperature-sensitive (*ts*) growth phenotype (Δ in Fig. 2A) that was suppressed by overexpression of *CEG1* (10). We have now extended the alanine-cluster mutagenesis of the guanylyltransferase-binding domain to three other pairs of amino acids that are conserved in Cet1p and CaCet1p. The mutated residues are denoted by dots in Figure 2A. The mutant genes *W247A-Q249A*, *N262A-V263A* and *K264A-D265A* were cloned into a *CEN TRP1* vector under the control of the *CET1* promoter and then tested by plasmid shuffle for complementation of a *cet1*Δ strain. The *N262A-V263A* and *K264A-D265A* mutants were viable after selection on 5-FOA at 25, 30 or 37°C. *N262A-V263A* and *K264A-D265A* cells grew as well as wild-type cells on rich medium (YPD) at 25, 30 and 37°C. In contrast, the *W247A-Q249A* cells failed to give rise to FOA-resistant colonies at all temperatures tested (25, 30 and 37°C); thus the *W247A-Q249A* change was lethal.

Some lethal mutations of *CET1* can be suppressed by expression of the mutant allele under the control of the strong constitutive *TPI1* promoter instead of the natural *CET1* promoter (10). We found that *CET1(W247A-Q249A)* was still lethal at all temperatures tested (16, 25, 30 and 37°C) when its expression was driven by the *TPI1* promoter. The *TPI1-CET1(K250A-W251A)* mutant was still temperature-sensitive, i.e. *TPI1-CET1(K250A-W251A)* cells grew at 25 and 30°C, but not at 37°C. The *in vivo* mutational effects point toward the conserved peptide WAQKW as a key constituent of the guanylyltransferase-binding site on yeast RNA triphosphatase.

In order to address whether mutations within the WAQKW motif might affect the catalytic activity of Cet1(201–549)p, the wild-type, W247A-Q249A and K250A-W251A proteins were expressed in bacteria as His-tagged fusions and purified from soluble lysates by nickel–agarose column chromatography (Fig. 2B). Triphosphatase activity was determined by assaying manganese-dependent ATP hydrolysis as a function of input enzyme. The specific activities of W247A-Q249A and

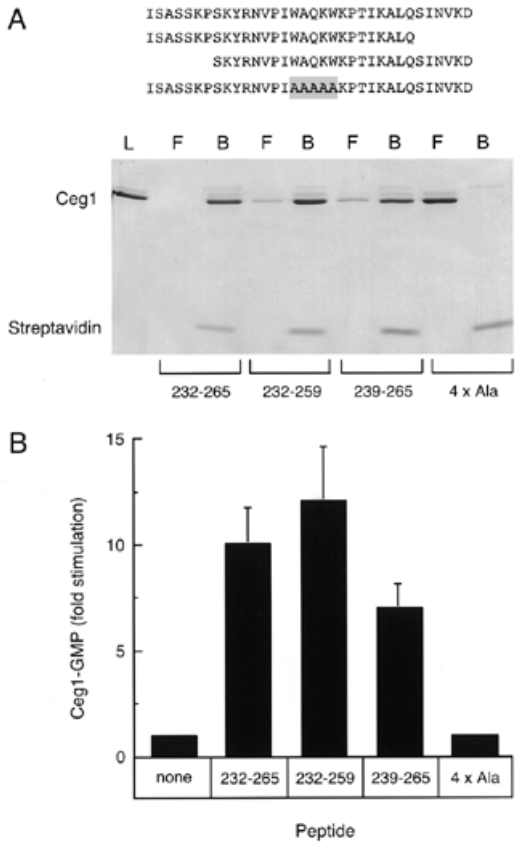

Figure 3. Effect of Cet1 peptide deletions and mutations on guanylyltransferase-binding. (**A**) Affinity chromatography. The amino acid sequences of the synthetic peptides Cet1(232–265), Cet1(232–259), Cet1(239–265) and Cet1(232–265)-4× Ala are shown. Purified Ceg1p was incubated with streptavidin beads containing the indicated biotinylated Cet1 peptides. Chromatography was performed as described in Materials and Methods. Aliquots of the input guanylyltransferase protein fraction (L) (equivalent to 40% of total material loaded), free-unbound fraction (F) (40% of supernatant) and the bead-bound fraction (B) (40% of SDS eluate) were analyzed by SDS–PAGE. A Coomassie blue stained gel is shown. The positions of Ceg1p and streptavidin are indicated on the left. (**B**) Guanylyltransferase activity. Ceg1p (300 ng) was preincubated for 15 min on ice with 50 pmol of Cet1 peptide in 10 µl of binding buffer. An aliquot (1 µl) of each sample was then assayed for guanylyltransferase activity. The signal intensities of the Ceg1p–GMP complexes were normalized to the signal intensity of a 'no peptide' control reaction (defined as 1.0). The data are the averages of the extents of stimulation from four separate experiments. Standard error bars are shown.

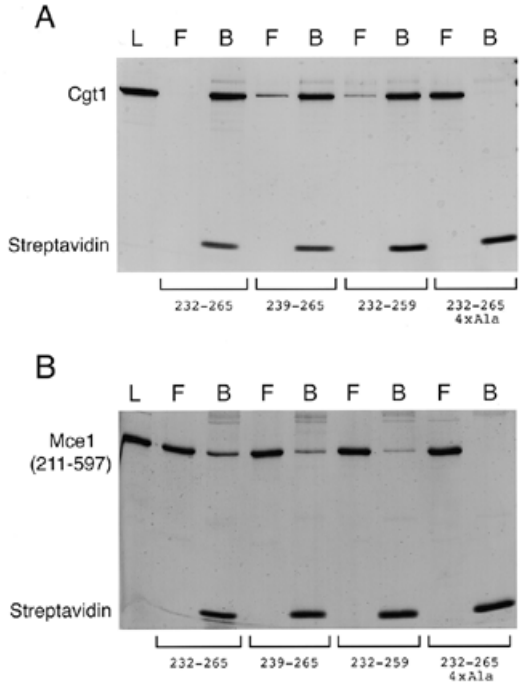

Figure 4. The conserved triphosphatase peptide binds to *C.albicans* and mouse RNA guanylyltransferases. *Candida albicans* guanylyltransferase Cgt1p (**A**) and mouse guanylyltransferase Mce1(211–597)p (**B**) were incubated with streptavidin beads containing the indicated biotinylated Cet1 peptides. Chromatography was performed as described in Materials and Methods. Aliquots of the input guanylyltransferase protein fraction (L) (equivalent to 40% of total material loaded), free-unbound fraction (F) (40% of supernatant) and the bead-bound fraction (B) (40% of SDS eluate) were analyzed by SDS–PAGE. Coomassie blue stained gels are shown.

K250A-W251A were comparable to that of wild-type Cet1(201–549)p (Fig. 2B). Thus, the *in vivo* phenotypes are not simply attributable to a catalytic defect. Moreover, an analysis of the native size of the W247A-Q249A and K250A-W251A proteins by glycerol gradient sedimentation indicated that the mutant proteins sedimented as homodimers, as did the wild-type Cet1(201–549)p (data not shown).

Effect of peptide deletions and mutations on guanylyltransferase-binding *in vitro*

Having validated the concept of using a synthetic peptide to study Cet1p–Ceg1p interactions, we sought to better define the guanylyltransferase-binding site by testing truncated and mutated Cet1 peptides as affinity ligands for Ceg1p. Biotinylated peptides composed of Cet1p amino acids 239–265 or 232–259 bound Ceg1p (Fig. 3A) and stimulated the guanylyltransferase activity of Ceg1p (Fig. 3B), implying that the Ceg1p-binding epitope is situated within the overlapping 21-amino acid segment from 239 to 259.

To address whether the WAQKW motif is involved in Ceg1p-binding, we introduced four alanines in lieu of Trp247, Gln249, Lys250 and Trp251 in the context of the 34mer Cet1(232–265) peptide. The instructive finding was that the 4× Ala cluster mutation abrogated both Ceg1p-binding and stimulation of guanylyltransferase activity (Fig. 3).

The conserved triphosphatase peptide binds to *C.albicans* RNA guanylyltransferase

The *C.albicans* RNA triphosphatase (CaCet1p) and RNA guanylyltransferase (Cgt1p) are structurally similar to the *S.cerevisiae* Cet1p and Ceg1p enzymes. Moreover, *S.cerevisiae cet1*Δ and *ceg1*Δ null mutations can be complemented by *CaCET1* and *CGT1*, respectively (17,18). To determine if the domain mediating the triphosphatase–guanylyltransferase interaction in *S.cerevisiae* is conserved functionally in other fungal capping systems, we evaluated the yeast triphosphatase peptides as ligands for the *C.albicans* guanylyltransferase. Cgt1p was expressed in bacteria as a His-tagged fusion protein and purified from soluble extracts by nickel–agarose chromatography. The enzyme preparation was nearly homogeneous with respect to the 54-kDa His-Cgt1p polypeptide (Fig. 4A, lane L) and the enzyme was catalytically active in

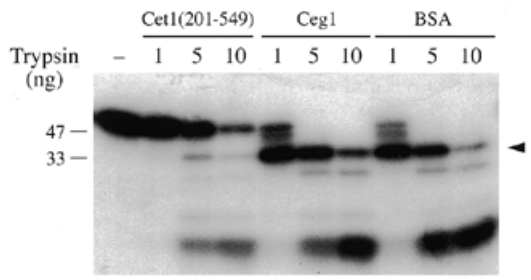

Figure 5. Proteolytic footprinting of a Cet1p-binding site on Ceg1p. Tryptic footprinting of ^{32}P-Ceg1p mixed with unlabeled Cet1(201–549)p, Ceg1p or BSA was performed as described in Materials and Methods. The ^{32}P-labeled digestion products were resolved by SDS–PAGE. An autoradiograph of the gel is shown. The amount of trypsin included in each reaction is indicated above the lanes. A control reaction mixture without trypsin is shown in lane (–). The positions and sizes (in kDa) of prestained marker polypeptides are indicated on the left.

transguanylylation (not shown). We found that Cgt1p was bound nearly quantitatively to beads containing the 34mer Cet1(232–265) peptide, but not at all to beads containing to the 34mer peptide with a 4× Ala mutation in the WAQKW motif (Fig. 4A). Cgt1p, like Ceg1p, also bound to the truncated peptides Cet1(239–265) and Cet1(232–259). We conclude that the guanylyltransferase-binding domain of *S.cerevisiae* RNA triphosphatase (amino acids 239–259) is conserved in other fungi and we infer that the target site on the guanylyltransferase for Cet1p peptide-binding is also likely to be conserved among fungal capping enzymes.

Peptide binding studies were performed with the guanylyltransferase domain of the mouse capping enzyme, Mce1p(211–597). Mce1p is a bifunctional 597-amino acid polypeptide composed of autonomous RNA triphosphatase (amino acids 1–210) and guanylyltransferase (amino acids 211–597) domains (15). Mce1(211–597)p bound weakly to beads containing the guanylyltransferase-binding peptide of yeast RNA triphosphatase, and not at all to the beads alone or to beads containing the 4× Ala mutant peptide (Fig. 4B). Most of the input mouse guanylyltransferase was recovered in the unbound fraction and very little was retained by the Cet1(232–265) peptide; this stands in contrast to the quantitative binding of the yeast guanylyltransferase to the Cet1(232–265) peptide when both capping enzymes were tested in parallel. The low-affinity interaction of mouse guanylyltransferase with the yeast RNA triphosphatase peptide *in vitro* may explain the genetic observation that yeast cells containing a capping apparatus composed of Cet1p and Mce1(211–597)p are viable, but very slow growing (15). We have since found that the slow growth phenotype of *MCE1(211–597)* cells can be suppressed by overexpression of *CET1* (C.K. Ho, B. Schwer and S. Shuman, unpublished results).

Binding studies conducted using purified recombinant *Chlorella* virus guanylyltransferase, a monofunctional monomeric enzyme (19), revealed a low extent of binding of the viral capping enzyme to beads containing the Cet1(232–265) peptide (data not shown), similar to what was observed in Figure 4B for the mouse guanylyltransferase.

Proteolytic footprinting of the Cet1p-binding site on Ceg1p

As a first step in defining the triphosphatase binding site on the yeast guanylyltransferase, we performed proteolytic footprinting of Ceg1p in the presence and absence of Cet1p. Recombinant guanylyltransferase containing a 5-amino acid N-terminal 'protein kinase tag' (RRASV) was ^{32}P end-labeled by *in vitro* reaction with protein kinase A (PKA) and [γ-^{32}P]ATP (20). PKA specifically phosphorylates the serine in the tag. Control experiments verified that PKA did not phosphorylate recombinant Ceg1p that lacked the kinase tag (not shown). Aliquots (0.5 μg) of end-labeled Ceg1p were pre-incubated with 4 μg of unlabeled Cet1(201–549)p, 4 μg of unlabeled Ceg1p or 4 μg of bovine serum albumin (BSA) and the mixtures were digested with 1, 5 or 10 ng of trypsin. The reaction products were analyzed by SDS–PAGE and the labeled N-terminal fragments were detected by autoradiography (Fig. 5). Digestion of the two 'control' mixtures of labeled guanylyltransferase with either cold Ceg1p or BSA by 1 ng of trypsin resulted in conversion of the input 54 kDa ^{32}P-Ceg1p into a prominent ^{32}P-labeled cleavage product of ~35 kDa (denoted by the arrowhead in Fig. 5). Digestion with 5 and 10 ng of trypsin resulted in cleavage at secondary sites closer to the N-terminus to yield a predominant low molecular weight fragment and two minor species of intermediate mobility (Fig. 5). The instructive finding was that pre-incubation of ^{32}P-Ceg1p with Cet1(201–549)p protected the primary tryptic site from proteolysis. This was evinced by (i) the complete resistance of ^{32}P-Ceg1p to digestion by 1 ng of trypsin, (ii) the persistence of a significant fraction of intact ^{32}P-Ceg1p at 5 ng of trypsin (this level of trypsin being sufficient to cleave all of the Ceg1p in the control mixtures to smaller fragments) and (iii) failure to accumulate significant amounts of the 35 kDa cleavage product even at 10 ng of trypsin (Fig. 5). The protection by Cet1(201–549)p was selective for the primary site, insofar as treatment with 5 and 10 ng of trypsin still resulted in the production of the prominent low molecular tryptic fragment and the minor fragment migrating just above it (Fig. 5). It was reported previously that initial scission of Ceg1p by trypsin occurs at two closely-spaced sites, Arg304-Leu305 and Lys306-His307, to yield an N-terminal fragment of 35 kDa and a carboxyl fragment of 20 kDa (1). The simplest interpretation of the present results is that the Cet1p-binding site on Ceg1p is located at or near Arg304 and Lys306, the primary tryptic sites protected from digestion in the presence of RNA triphosphatase, whereas secondary tryptic sites located close to the N-terminus (as yet unmapped) are not part of the Cet1p–Ceg1p interface. Consistent with this interpretation, we found that the 34mer Cet1(232–265) peptide was able to protect Ceg1p from proteolysis by trypsin, whereas the 4× Ala mutant peptide was not (data not shown).

DISCUSSION

We have co-localized the guanylyltransferase-binding and guanylyltransferase-stimulation functions of yeast RNA triphosphatase to a 21-amino acid peptide from residues 239 to 259. Within this peptide is a WAQKW motif that is implicated by alanine-cluster mutagenesis in Cet1p–Ceg1p interaction *in vivo* and *in vitro*. The correlation between the *in vivo* lethality of the W247A-Q249A mutation and the abrogation of

guanylyltransferase-binding and stimulation by a WAQKW-to-alanine cluster mutation of the Cet1(232–265) peptide provides the strongest evidence to date that the triphosphatase–guanylyltransferase interaction is essential in yeast, i.e. because the alanine-cluster mutations in the WAQKW motif did not affect either the catalytic activity of yeast triphosphatase or its self-interaction to form a homodimer.

Available data concerning yeast triphosphatase structure are consistent with the guanylyltransferase-binding epitope being located on the surface of the protein. Limited proteolysis of Cet1(201–549)p showed that the Tyr241–Arg242 peptide bond is highly sensitive to chymotryptic cleavage and that additional sites of accessibility to chymotrypsin and trypsin are clustered nearby at Lys256–Ala257, Leu258–Gln259 and Lys264–Asp265 (10). In contrast, the distal C-terminal portion of the protein is relatively protease-insensitive. We have recently crystallized a biologically active form of yeast RNA triphosphatase and determined its structure by X-ray diffraction (C. Lima, L. Wang and S. Shuman, submitted for publication). The structure reveals that the Cet1p segment from residues 241 to 259 is located on the protein surface and that the Trp247 side chain in the WAQKW motif is solvent-exposed (and thus in a position to interact directly with Ceg1p).

Our finding that the Cet1(232–265) peptide binds tightly to fungal guanylyltransferases and only weakly to the mammalian and *Chlorella* virus guanylyltransferases has interesting implications for the evolution of the capping apparatus. Fungal, metazoan and DNA virus guanylyltransferases share a catalytic mechanism of covalent nucleotidyl transfer and they have a common active site composed of six conserved motifs (21,22). The *in vivo* effects of mutations within the nucleotidyl transferase motifs of yeast Ceg1p are in accord with the molecular contacts between GTP and guanylyltransferase revealed in the crystal structure of the *Chlorella* virus capping enzyme (22,23). Thus, it is very likely that all cellular and DNA virus-encoded guanylyltransferases share the same overall structural fold. Nonetheless, it is clear from sequence alignments that the conserved components of the guanylyltransferase catalytic site are embellished by intervening peptide segments that diverge in viruses, fungi and mammals.

It is surprising that a metazoan guanylyltransferase has any affinity for the guanylyltransferase-binding domain of yeast RNA triphosphatase, because the metazoan guanylyltransferase and triphosphatase activities are linked in *cis* within a single polypeptide and the metazoan RNA triphosphatase is structurally and mechanistically unrelated to the fungal RNA triphosphatase. Metazoan RNA triphosphatase is instead related to the superfamily of protein phosphatases that catalyzes phosphoryl transfer through a cysteinyl-phosphate intermediate (22,24–26). The vestigial low-affinity interaction of mouse guanylyltransferase with the Cet1 peptide suggests that guanylyltransferases evolved from an ancestral enzyme that did interact *in trans* with a triphosphatase, but that selection for such interaction was relaxed during the emergence of metazoa, because of a gene fusion event that linked the triphosphatase and guanylyltransferase domains *in cis*. We cannot at present discern any homolog of fungal RNA triphosphatase in genomic databases for metazoans, including the completed genome of the nematode *Caenorhabditis elegans.* The fusion of metazoan guanylyltransferase to a PPase-like RNA triphosphatase

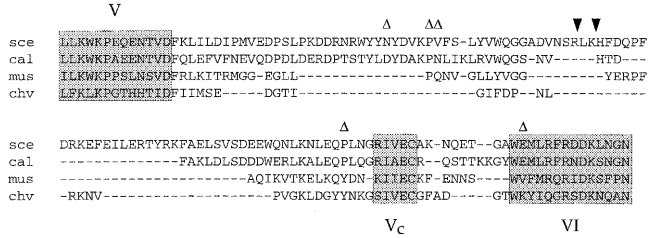

Figure 6. Alignment of a candidate triphosphatase-binding region of *S.cerevisiae* guanylyltransferase with the corresponding segments of *C.albicans*, mouse and *Chlorella* virus guanylyltransferases. Nucleotidyl transferase motifs V and VI are shown in shaded boxes, as is motif V_c, which is conserved in the cellular and *Chlorella* virus guanylyltransferases (22). The primary sites of tryptic cleavage of Ceg1p are denoted by arrowheads above the sequence. Residues mutated in *ceg1-ts* alleles that are suppressed by overexpression of Cet1p are indicated by Δ above the sequence.

may have allowed for the loss of a Cet1p-like enzyme from the metazoan genome, or else have permitted the divergence of such a protein to a point that it serves a new function in metazoan species and is no longer discernible as Cet1p-like by available search programs.

Our proteolytic footprinting experiments provide the first physical evidence that yeast triphosphatase interacts with the C-terminal domain of yeast guanylyltransferase. The trypsin-sensitive sites of Ceg1p that are shielded from proteolysis when Ceg1p is bound to Cet1p (denoted by arrowheads in Fig. 6) are located between nucleotidyl transferase motifs V and VI. An alignment of the sequences of the *S.cerevisiae*, *C.albicans*, mouse and *Chlorella* virus guanylyltransferases underscores the point that this intervening region is variable in length and poorly conserved, except for motif V_c (Fig. 6). Segments present in the two fungal guanylyltransferases, which are absent from the mouse and *Chlorella* virus proteins, may well comprise the high-affinity Cet1p-binding surface of Ceg1p. Several *ceg1-ts* alleles containing missense mutations located between motifs V and VI in the C-terminal domain can be suppressed by overexpression of Cet1p (9,13). Several of the amino acids that are mutated in these *ceg1-ts* alleles are situated just proximal to the tryptic cleavage sites (denoted by Δ in Fig. 6). Thus, biochemical and genetic experiments implicate the same segment of Ceg1p in triphosphatase-binding. The results presented here set the stage to define in molecular detail the triphosphatase–guanylyltransferase interface by crystallizing yeast guanylyltransferase bound to the Cet1(232–265) peptide.

ACKNOWLEDGEMENTS

We thank San San Yi and Scott Geromanos for expert peptide synthesis and Iaroslava Rouzankina for contributions to the proteolytic footprint analysis. This research was supported by NIH grant GM52470.

REFERENCES

1. Shuman,S. (1995) *Prog. Nucleic Acid Res. Mol. Biol.*, **50**, 101–129.
2. Shibagaki,Y., Itoh,N., Yamada,H., Nagata,S. and Mizumoto,K. (1992) *J. Biol. Chem.*, **267**, 9521–9528.
3. Schwer,B. and Shuman,S. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 4328–4332.

4. Mao,X., Schwer,B. and Shuman,S. (1996) *Mol. Cell. Biol.*, **16**, 475–480.
5. Wang,S.P. and Shuman,S. (1997) *J. Biol. Chem.*, **272**, 14683–14689.
6. 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.
7. Ho,C.K., Pei,Y. and Shuman,S. (1998) *J. Biol. Chem.*, **273**, 34151–34156.
8. Itoh,N., Yamada,H., Kaziro,Y. and Mizumoto,K. (1987) *J. Biol. Chem.*, **262**, 2989–2995.
9. Ho,C.K., Schwer,B. and Shuman,S. (1998) *Mol. Cell. Biol.*, **18**, 5189–5198.
10. Lehman,K., Schwer,B., Ho,C.K., Rouzankina,I. and Shuman,S. (1999) *J. Biol. Chem.*, **274**, 22668–22678.
11. McCracken,S., Fong,N., Rosonina,E., Yankulov,K., Brothers,G., Siderovski,D., Hessel,A., Foster,S., Shuman,S. and Bentley,D.L. (1997) *Genes Dev.*, **11**, 3306–3318.
12. Cho,E., Takagi,T., Moore,C.R. and Buratowski,S. (1997) *Genes Dev.*, **11**, 3319–3326.
13. Cho,E., Rodriguez,C.R., Takagi,T. and Buratowski,S. (1998) *Genes Dev.*, **12**, 3482–3487.
14. Pei,Y., Ho,C.K., Schwer,B. and Shuman,S. (1999) *J. Biol. Chem.*, **274**, 28865–28874.
15. Ho,C.K., Sriskanda,V., McCracken,S., Bentley,D., Schwer,B. and Shuman,S. (1998) *J. Biol. Chem.*, **273**, 9577–9585.
16. Ho,C.K. and Shuman,S. (1999) *Mol. Cell*, **3**, 405–411.
17. Yamada-Okabe,T., Shimmi,O., Doi,R., Mizumoto,K., Arisawa,M. and Yamada-Okabe,H. (1996) *Microbiology*, **142**, 2515–2523.
18. Yamada-Okabe,T., Mio,T., Matsui,M., Kashima,Y., Arisawa,M. and Yamada-Okabe,H. (1998) *FEBS Lett.*, **435**, 49–54.
19. Ho,C.K., Van Etten,J.L. and Shuman,S. (1996) *J. Virol.*, **70**, 6658–6664.
20. Kelman,Z., Nattinis,V. and O'Donnell,M. (1995) *Methods Enzymol.*, **262**, 430–442.
21. Shuman,S. and Schwer,B. (1995) *Mol. Microbiol.*, **17**, 405–410.
22. Wang,S.P., Deng,L., Ho,C.K. and Shuman,S. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 9573–9578.
23. Håkansson,K., Doherty,A.J., Shuman,S. and Wigley,D.B. (1997) *Cell*, **89**, 545–553.
24. Takagi,T., Moore,C.R., Diehn,F. and Buratowski,S. (1997) *Cell*, **89**, 867–873.
25. Yue,Z., Maldonado,E., Pillutla,R., Cho,H., Reinberg,D. and Shatkin,A.J. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 12898–12903.
26. Wen,Y., Yue,Z. and Shatkin,A.J. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 12226–12231.