# **Supporting Information**

### Nakamura et al. 10.1073/pnas.1321312111

### SI Materials and Methods

Preparation of Candida albicans Thg1 (CaThg1) and Mutants. The thg1 gene from Candida albicans was cloned into the Nco1/Xho1 sites of the pET26b vector (Novagen) with an N-terminal His<sub>6</sub> tag. C. albicans encodes a unique seryl-tRNA<sub>CAG</sub> that decodes the leucine codon CTG as serine (1, 2); thus all CTG codons in the cloned C. albicans thg1 gene were altered to the serine codon AGC using a QuikChange mutagenesis kit (Agilent Technologies). A tobacco etch virus (TEV) protease-cleavable site was inserted between the N-terminal His<sub>6</sub>-tag and *thg1* to allow removal of the His-tag for crystallization. The enzyme was overexpressed using Escherichia coli B834 (DE3)-pRARE2 (Novagen). Cells were grown in LB medium containing 25 µg/mL kanamycin and 34 µg/mL chloramphenicol at 37 °C to an  $OD_{600}$  of 0.6. Isopropyl- $\beta$ -Dthiogalactopyranoside was then added to a final concentration of 250 µM for the induction of overnight protein production at 25 °C. Cells were harvested  $(4,000 \times g \text{ for } 15 \text{ min at } 4 \text{ }^\circ\text{C})$  and disrupted by sonication in buffer A [50 mM Tris HCl, pH 7.5, 300 mM NaCl, 1 mM MgCl<sub>2</sub>, 10% (vol/vol) glycerol] with 0.5 mg/mL lysozyme and 0.1 mg/mL DNase I. All of the following purification processes were carried out at 4 °C. Cell debris was removed by centrifugation (40,000  $\times$  g for 1 h), and clarified supernatant was loaded onto a HisTrap HP column (GE Healthcare) preequilibrated with buffer A. The column was washed with buffer A containing 50 mM imidazole, and proteins were eluted with a linear gradient of 50-250 mM imidazole. Pooled fractions were dialyzed against buffer A containing 1 mM DTT. The His<sub>6</sub> tag attached to the N terminus was cleaved with TEV protease overnight and was then removed using a HisTrap HP column. The enzyme fractions were then loaded onto a HiLoad 26/60 Superdex 200 pg column (GE Healthcare) equilibrated with buffer A. Pooled fractions were concentrated by ultrafiltration to a final concentration of 14 mg/mL.

The overexpression vectors of CaThg1 mutants without a TEV protease cleavage site were constructed using a QuikChange mutagenesis kit (Agilent Technologies). All mutations were confirmed by sequencing of the plasmid DNA. CaThg1 variants were purified by Ni Sepharose 6 Fast Flow column chromatography (GE Healthcare) followed by dialysis against buffer B [25 mM Hepes–Na, pH 7.5, 500 mM NaCl, 4 mM MgCl<sub>2</sub>, 1 mM DTT, 50% (vol/vol) glycerol] and stored at –30 °C.

**tRNA Preparation.** *C. albicans*  $\text{tRNA}^{\text{His}}\Delta \text{G}^{-1}$  and its variants and *Saccharomyces cerevisiae*  $\text{tRNA}^{\text{Phe}}$  variant  $\text{tRNA}^{\text{Phe}}_{\text{GUG}}$  were transcribed using T7 RNA polymerase as described previously (3). Transcribed tRNAs were purified by HiTrap DEAE FF column chromatography (GE Healthcare) as described previously (4). The tRNAs eluted at ~250 mM NaCl. The tRNAs were precipitated with ethanol, dissolved in buffer C [25 mM Tris·HCl, pH 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 5% (vol/vol) glycerol] for crystallization or in buffer D [25 mM Hepes–Na, pH 7.5, 125 mM NaCl, 10 mM MgCl<sub>2</sub>, 3 mM DTT, 5% (vol/vol) glycerol] for Thg1 activity assays, and stored at ~80 °C.

**Crystallization and Data Collection.** Crystals of ligand-free CaThg1 were obtained by sitting drop vapor diffusion against 100 mM Bis-Tris propane (pH 8.0), 200 mM NaNO<sub>3</sub>, and 20% (wt/vol) PEG 3350 at 20 °C. Crystals of the CaThg1-ATP complex were obtained by soaking ligand-free CaThg1 crystals in a reservoir solution containing 25 mM ATP and 20 mM MgCl<sub>2</sub> overnight and then moved into cryoprotectant buffer [reservoir solution containing 25 mM ATP, 20 mM MgCl<sub>2</sub>, and 20% (vol/vol) glycerol]. Crystals of the CaThg1-GTP complex were obtained by the same procedure. The diffraction data sets were collected on beamline BL41XU at SPring-8 (Proposal 2011A1299) using a wavelength of 1.0000 Å at -173 °C.

CaThg1 and tRNA<sup>Phe</sup><sub>GUG</sub> were mixed in a molecular ratio of 1:1.2 in buffer C. The resulting mixture was concentrated by ultrafiltration to a final protein concentration of 200  $\mu$ M. Crystals of the CaThg1-tRNA<sup>Phe</sup><sub>GUG</sub> complex were obtained by sitting drop vapor diffusion against 50 mM Bis-Tris propane (pH 6.0), 200 mM CaCl<sub>2</sub>, and 15–20% (wt/vol) PEG 3350 at 20 °C. Crystals of CaThg1-tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup> complex were obtained under the same crystallization conditions as for the CaThg1-tRNA<sup>Phe</sup><sub>GUG</sub> complex. The diffraction data sets of CaThg1-tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup> and CaThg1tRNA<sup>Phe</sup><sub>GUG</sub> were collected on beamline BL1A (Photon Factory) with a wavelength of 1.1000 Å and on beamline BL41XU at SPring-8 (Proposal 2011B1227) with a wavelength of 0.97906 Å, respectively, at –173 °C after crystals were soaked into reservoir solution buffer containing 20% (vol/vol) glycerol. All diffraction data of CaThg1-tRNA, CaThg1-ATP, and CaThg1-GTP complexes were indexed, integrated, scaled, and merged using XDS (5).

**Structure Determination and Refinement.** The structure of the CaThg1-ATP complex was solved by molecular replacement using Molrep in the CCP4 suite (6), with the dimer structure of human Thg1 [PDB ID 3OTE (19)] as a search model. The model of CaThg1 was rebuilt by phenix.autobuild (7) and LAFIRE (8) and modified manually using Coot (9). Structure refinement was performed using phenix.refine (10) and autoBUSTER (11) with noncrystallographic symmetry restraints. Finally,  $R_{work}$  and  $R_{free}$  factors were converged to 17.6% and 20.8%, respectively. The structure of the CaThg1-GTP complex was solved by molecular replacement using Molrep with the tetramer structure of CaThg1 in the CaThg1-ATP complex as a search model. After several cycles of refinement by phenix.refine with noncrystallographic symmetry restraints and manual fitting using Coot,  $R_{work}$  and  $R_{free}$  factors were finally converged to 19.8% and 23.9%, respectively.

The structure of the CaThg1-tRNA<sup>Phe</sup><sub>GUG</sub> complex was solved by molecular replacement using Phaser (12) with the tetramer structure of CaThg1 in the CaThg1-ATP complex as a search model. The model of tetrameric CaThg1 in the CaThg1-tRNAPhe<sub>GUG</sub> complex was refined using phenix.refine and autoBUSTER with reference structure restraints using the CaThg1-ATP complex structure. After refinement of the CaThg1 structure, electron density maps of two tRNA molecules were clearly observed, which allowed us to manually fit the S. cerevisiae tRNA<sup>Phe</sup> structure [PDB ID 1EHZ (28)] into this electron density map using Coot. The models of two tRNA molecules were rebuilt and fitted automatically using NAFIT, a model-fitting program for nucleic acid structures (13). After several cycles of refinement with phenix. refine or autoBUSTER, in which reference structure and noncrystallographic symmetry restraints were applied to each subunit of CaThg1 and tRNA molecules, and manual fitting with Coot,  $R_{\text{work}}$  and  $R_{\text{free}}$  factors finally converged to 25.3% and 28.8%, respectively (Table S1). The structure of the CaThg1-tRNA<sup>His</sup> $\Delta G^{-1}$ complex was obtained by rigid-body refinement using the struc-ture of the CaThg1-tRNA<sup>Phe</sup><sub>GUG</sub> complex as a model. The struc-tural refinement of the CaThg1-tRNA<sup>His</sup> $\Delta G^{-1}$  complex was carried out in the same way as described above, and  $R_{\text{work}}$  and  $R_{\text{free}}$  factors converged to 29.2% and 33.5%, respectively. For all structures, the translation/libration/screw protocol was used in the final cycles of the refinement. Table S1 presents a summary of all data collection and structural refinement statistics. Ramachandran plots of all determined structures were calculated using MolProbity (14) and showed that 97.1–99.8% of the protein residues were in the favored regions, with no residues in the outlier region. All figures were prepared with PyMOL (PyMOL Molecular Graphics System, Version 1.5.0.4).

**Small-Angle X-Ray Scattering.** For preparation of CaThg1-bound tRNA<sup>His</sup> $\Delta G^{-1}$ , CaThg1 without His<sub>6</sub>-tag was mixed with CatRNA<sup>His</sup> $\Delta G^{-1}$  in a molar ratio of 1:1 and then purified by gel filtration using a Superdex 200 16/60 pg column (GE Health-care) in buffer C. The complex fractions were collected and concentrated by ultrafiltration to a final protein concentration of 5 mg/mL with binding buffer C.

Small-angle X-ray scattering (SAXS) measurements were carried out at beamline 10C (Photon Factory). A wavelength of 1.488 Å was used, and the specimen-to-detector distance was 2,034 mm. The scattering vector range was  $0.017 \text{ Å}^{-1} < s < 0.300 \text{ Å}^{-1}$ , where  $s = 4\pi \sin\theta/\lambda$  with 20 being the scattering angle. The conditions of the SAXS measurement were optimized, and final data were measured using 1.0 mg/mL of each sample at 20 °C, and all buffers and samples were exposed for 10 min. The SAXS data sets were normalized to the intensity of the incident beam and processed for background subtraction using the standard procedures with the program package PRIMUS (15).

The SAXS mass of each sample was estimated from I(0)/c [I(0)= scattering intensity at Q = 0; c =concentration] based on the value  $I(0)_{\rm ref}/c_{\rm ref}$  calculated from a solution of the reference protein ovalbumin (molecular weight = 45 kDa) in buffer C. The Rg values of each sample were estimated from scattering profiles in a specific Q range ( $Q \times Rg < 1.3$ ) by Guinier analysis and are summarized in Table S2. The Rg volume and the discrepancies between the calculated and the experimental scattering curves were minimized using the program CRYSOL (16) as described previously (17). The model of CaThg1 bound to two tRNA molecules in antiparallel orientation was built by superposing subunit A with tRNA1 to subunit C. The model of CaThg1 bound to three tRNA molecules was built by superposing subunit A with tRNA1 to subunits B and C. The model of CaThg1 bound to four tRNA molecules was built by superposing subunits A with tRNA1 to subunits B, C, and D.

Gel Filtration Analysis. Gel filtration analysis was performed by fast protein liquid chromatography using a Superdex 200 10/300 GL column (GE Healthcare) at 4 °C equilibrated in buffer B. For preparative assays, samples (150  $\mu$ L) were prepared in the same buffer and preincubated for 30 min at room temperature with 100  $\mu$ M of CaThg1 with or without tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup>. The samples were loaded onto a 24-mL column at 4 °C at a flow rate of 1.0 mL/min.

**Thg1 Assay.** Guanylylation activity of CaThg1 proteins was assayed by monitoring the incorporation of  $[\alpha^{-32}P]$ GTP into tRNA<sup>His</sup> $\Delta G^{-1}$ as described previously (18). Reactions were incubated at 25 °C for 1 h until they reached a plateau. All reactions were carried out in buffer D. Protein concentrations were adjusted to 5  $\mu$ M, and tRNA was added to a final concentration of 10  $\mu$ M, giving a 1:2

- 1. Gomes AC, et al. (2007) A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. *Genome Biol* 8(10):R206.
- Suzuki T, Ueda T, Watanabe K (1997) The 'polysemous' codon: A codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J 16(5):1122–1134.
- Nakamura A, Yao M, Chimnaronk S, Sakai N, Tanaka I (2006) Ammonia channel couples glutaminase with transamidase reactions in GatCAB. Science 312(5782):1954–1958.
- Easton LE, Shibata Y, Lukavsky PJ (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. RNA 16(3):647–653.
- Kabsch W (2010) XDS. Acta Crystallogr D Biol Crystallogr 66(Pt 2):125–132.
  Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. Acta Crystallogr D
- Biol Crystallogr 66(Pt 1):22–25.
  Tervilliger TC, et al. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr D Biol Crystallogr 64(Pt 1):61–69.

molar ratio of enzyme to tRNA. Reactions were analyzed by polyacrylamide gel electrophoresis in gels containing 12% urea and visualized using a Molecular Dynamics Storm 860 phosphoimager (Amersham Biosciences).

**Phylogenetic Tree.** We used the classification scheme provided by the Structural Classification of Proteins database (19). Each obtained structure was than blasted against the Protein Data Bank (PDB) (20) to identify additional palm-domain–containing proteins. Structures were downloaded from PDB. Structures were aligned using the STAMP algorithm as implemented in Visual Molecular Dynamics (VMD) (21, 22).

#### SI Text

**Thg1-tRNA Complex in Solution.** The predicted SAXS curves for parallel and antiparallel tRNA binding were compared with experimental SAXS data (Fig. S3). The calculated SAXS curve from the crystal structure of the CaThg1-tRNA complex in parallel tRNA orietation showed a close fit to the experimental curve of the purified CaThg1-tRNA<sup>His</sup> $\Delta G^{-1}$  complex with a discrepancy  $\chi$  value of 0.184, whereas the model of tRNA binding to CaThg1 in an antiparallel orientation showed a higher discrepancy  $\chi$  value (0.384) than the crystal structure (Fig. S3*B*). These results suggest that the orientation of two tRNA molecules in the present crystal structure is consistent with the solution structure of the CaThg1-tRNA<sup>His</sup> $\Delta G^{-1}$  complex.

**Flexible Region of Thg1.** Thg1 contains a highly flexible region (residues 218–244) with low-sequence conservation. This region is disordered in CaThg1-tRNA structures as is the case in the HsThg1-dGTP structure (PDB ID 30TB) (Fig. 2*B*), whereas this flexible region forms the  $\alpha$ -helix in the CaThg1-GTP structure but a long  $\beta$ -sheet arm ( $\beta 6$  and  $\beta 7$ ) in that of HsThg1 (PDB ID 30TC) (Fig. S5 *A* and *B*).

ATP and GTP Binding in Nucleotide-Binding Site. To understand how eukaryotic Thg1 recognizes and distinguishes ATP and GTP, we solved the structures of CaThg1 in complex with ATP and in complex with GTP. Both structures contained two nucleotide molecules (ATP1 and ATP2 and GTP1 and GTP2, respectively) and three magnesium ions ( $Mg^{2+}A$ ,  $Mg^{2+}B$ , and  $Mg^{2+}C$ ). In the CaThg1-ATP complex, the triphosphate moiety of ATP1 interacts with Mg<sup>2+</sup>A and Mg<sup>2+</sup>B and forms hydrogen bonds with main-chain amides (residues 32-34). The base of ATP1 is coordinated in the nucleotide-binding pocket composed of  $\alpha 2$  and  $\alpha$ 3 and the in-between loop. Superposition of CaThg1-GTP into CaThg1-ATP clearly showed that the base of ATP1 is more deeply embedded in the nucleotide-binding pocket than that of GTP1 and recognized by hydrogen bonds with the main-chain N and O of Glu43 and Asp47 (Fig. S8B). The base of GTP1 makes hydrogen bonds with the main-chain atoms of Glu43 and Asp47, which is also observed in the HsThg1-dGTP structure. Remarkably, unlike the CaThg1-GTP structure, electron density maps in the CaThg1-ATP structure clearly show that the side chain of Lys44 interacts directly with N3 and O2' of ATP1 (Fig. S8 A and B).

- Yao M, Zhou Y, Tanaka I (2006) LAFIRE: Software for automating the refinement process of protein-structure analysis. Acta Crystallogr D Biol Crystallogr 62(Pt 2):189–196.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66(Pt 4):486–501.
- Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix refine. Acta Crystallogr D Biol Crystallogr 68(Pt 4):352–367.
- Blanc E, et al. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1):2210–2221.
- McCoy AJ, et al. (2007) Phaser crystallographic software. J Appl Cryst 40(Pt 4):658–674.
  Yamashita K, Zhou Y, Tanaka I, Yao M (2013) New model-fitting and model-completion programs for automated iterative nucleic acid refinement. Acta Crystallogr D Biol Crystallogr 69(Pt 6):1171–1179.
- Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66(Pt 1):12–21.

- Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. J Appl Cryst 36:1277–1282.
- Svergun D, Barberato C, Koch MHJ (1995) CRYSOL: A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Cryst 28:768–773.
- Nakamura A, et al. (2010) Two distinct regions in Staphylococcus aureus GatCAB guarantee accurate tRNA recognition. Nucleic Acids Res 38(2):672–682.
- Heinemann IU, et al. (2009) The appearance of pyrrolysine in tRNAHis guanylyltransferase by neutral evolution. Proc Natl Acad Sci USA 106(50):21103–21108.
- 19. Lo Conte L, et al. (2000) SCOP: A structural classification of proteins database. *Nucleic Acids Res* 28(1):257–259.
- Berman HM, et al. (2002) The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 58(Pt 6 No 1):899–907.
- O'Donoghue P, Luthey-Schulten Z (2003) On the evolution of structure in aminoacyltRNA synthetases. *Microbiol Mol Biol Rev* 67(4):550–573.
- Russell RB, Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins* 14(2):309–323.
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037–10041.



Fig. S1. Enzymatic reactions catalyzed by Thg1 and DNA polymerase.



Fig. S2. (A) The guanylylation activity using  $[\alpha^{-32}P]$ GTP of CatRNA<sup>His</sup> (C. *albicans* tRNA<sup>His</sup>) variants and SctRNA<sup>Phe</sup> (S. *cerevisiae* tRNA<sup>Phe</sup>) variants. (B) The 2D structure of C. *albicans* tRNA<sup>His</sup> and S. *cerevisiae* tRNA<sup>Phe</sup>. (C) Superposition of tRNA molecules in CaThg1-tRNA<sup>Phe</sup><sub>GUG</sub> (yellow) and CaThg1-tRNA<sup>His</sup> $\Delta G^{-1}$  (gray). (D) Electrostatic surface potential of CaThg1 calculated using APBS software (23). Positively and negatively charged surfaces are colored blue and red, respectively.



**Fig. S3.** The observed SAXS curves (black dots) are compared with theoretical scattering curves calculated from CaThg1-tRNA<sup>His</sup> $\Delta G^{-1}$ . (*A*) Model of the molar ratio of 4:1 Thg1 to tRNA. (*B*) Model of the molar ratio of 4:2 Thg1 to tRNA. The crystal structure of CaThg1-tRNA (red) and the model of CaThg1-tRNA (blue) with two tRNA molecules bound in antiparallel orientation. (*C*) Model of the molar ratio of 4:3 Thg1 to tRNA. (*D*) Model of the molar ratio of 4:4 Thg1 to tRNA.



Fig. 54. Gel-filtration analysis. The gel-filtration analysis of CaThg1 was carried out in the presence (solid line and diamond) or absence (dashed line and squares) of tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup>. The standard curve was calculated using ferritin (440 KDa),  $\gamma$ -globulin (158 kDa), conalbumin (75 kDa), and ovalbumin (44 kDa). The complex of CaThg1 with tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup> eluted with an apparent molecular weight of 173 kDa, consistent with the molecular weight of tetrameric CaThg1 (130 kDa) with two tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup> molecules (48 kDa).



**Fig. S5.** (*A*) The dimer structure of CaThg1-GTP with ordered  $\alpha$ -helix and loop structures (blue). (*B*) The dimer structure of HsThg1 with the long arm composed of two  $\beta$ -strands (blue). (*C*) Superposition of eight monomers of Thg1 in the asymmetric unit of the CaThg1-ATP structure. The palm domains (blue) were well superposed but not the finger domains (cyan), suggesting a flexibility of the finger domain. (*D*) The structural comparison of tRNA molecules in the CaThg1-tRNA structure (yellow) and the crystal structure of *S. cerevisiae* tRNA<sup>Phe</sup> (black) (PDB ID: 1EHZ).



Fig. S6. Stereo views of the simulated annealing composite omit map calculated using Phenix for the anticodon and the acceptor stem recognition are in gray and are contoured at 1.0  $\sigma$ . (A) G34 and U35 recognition. (B) U35 and G36 recognition. (C) The acceptor stem recognition.

## Palm domain

PNAS PNAS

	HsTha1		α1	тт	β1	α2 0000000	α3 00000000000	0000000	β2
	CaTha1		α1 000000		β1	α2 0000000	α3	00000000	β2
	carngr	ŗ	10	20	30 ♦♦	40	5 <u>9</u>	eò	70
euThg1	CaThgl ScThgl HsThgl DdiTLP1	1	YEYVKLFEKEN GGYVRQFETHD FEYVRDFEADD YEYVKSFEQPD	ILLPDTY ILLPQCY ICLAHCW ILLQNVW	IIIRVDGK IVVRIDGKI VVVRLDGRI IVVRIDGR	GEHKESQEYEEEK KEHEESKEYEEAK VEHREAEKHNEAK SEHKETTKHDYAK	. PNDLKALQVMNS# . PNDENALKLMNAC . PNDSRALQLMTKC . PNDDRGLSLMNR#	A A E K LMSKYS CAKNLVLKYK CAQTVMEELE A A LEVCKEFP	. DVMLAYG NDIILAFG . DIVIAYG . DIVIAFG
arThg1	MaThg1 MtThg1 MkThg1		. MKTREIYAEMI ISMREHEVYSKLI VERCDIEVYADI	RCIPP	VVLRADGRI	NFKNTLSGLGFEK GFHRLTESLDFDR	. PYDKTFARAMAD . PYDEAFRDLMIR . PYDRFAFAMAF	TAELFIKKSG CLDLMEEFS	LSPLFAYT P. SLIYT
TLP	DdiTLP2 DdiTLP3	QSNGNFR	I YNKFELFEKAFI I GDRMKSYEDGMI	VFPNIN	FIIRLDGN FIIRLDGH	SFSKLNKQLKLER SFSKFSKVFKKPG	HDERFHESMKET IAWDIRIHQAMVET	TSNNLFSHFI TATALMKTFL	G.CKFIYS PTVVYT
	DdiTLP4	METKDKKKDMELI	<u>IGD</u> RMK <u>S Y E</u> DDMI	KIQIEKNKPE	TIRDGH	SESKFTKNF <mark>NK</mark> P.	. HDIRIHNAMIET	<u>s</u> tv <u>ll</u> kt <b>f</b> m	P TCIYT
	HsTha1	β3 TT		000	α4 000000000	0000000 0000		_	β4
	CaThgl	β3 ππ			α4		T	<b>T</b>	β4
	carnyr	♦ 80		90	100	110	120	130	•
euTha1	CaThg1 ScThg1	DSDEYSFLLRKN ESDEYSFILKSS	· · · · · · · · · · · · · · · · · · ·	CQLYERREMP TLFNRRKDP	KLTTLFSS KLATLFG <mark>S</mark> I	LMSTYYMYFWSQY FFTSNYVALWAKF	FPDKPLHI	. DHLPNFDA . KHLPYFDS	RAVLYPDF RCVAYPNL
	HsThg1 DdiTLP1	QSDEYSFVFKRK ESDEYSFVLKKS		CNWFKRRAS CNLFERRSS	ISSSIVS	QFASSYVFYWRDY XFTSQFVYRWKEY	FEDQPL FGEHEL	LYPPGFDG KYPPTFDS	RVVVYPSN RCVLYPTD
arThg1	MaThg1 MtThg1 MkThg1	FSDEISFLFT FSDEINVLLD FSDELNALIPRG		VPFAGRVE VPFAGRVE VPFSGRVE	LDSVVAS LDSVFSG LTSVSAS	FLGSALTIKL FASSSFTRGALEA CASTYFFRALORH	GFSPV		RLVALQK. RVIPISS. RCVVLTD.
TLP	DdiTLP2 DdiTLP3	FSDEINIVIYNQY FSDEITMCFPSII	PP.	QFLSNRIQ LAYNGKVQ	LISTTSS LISLTAGI	ITSLNFSINLSKK LASTVFYKSITQA	LN <mark>NNNNNN</mark> LY <mark>D</mark> TDKEEKIIDLI	NVFSYFDC LKTATPSFDS	RAFVLPI. RLFVLPSN
	DdiTLP4	FSDEITMCFPSII	DESTLEEGKEIP	LAYSGKVQ	LISLSSG	LASTVFFKSITNA	2Y <mark>D</mark> KDTELNLIKLI	LETCTPHFDA	RIFTLPSN
		Fingers do	main						
	HsThg1	0000000000.00	α5 2000000000000000000000000000000000000	00000	α6 000000	α7 <u>0000000</u>	η1 222 222	α8 β!	5
	CaThg1		α5 000000000000000	00000	0000000			οοοοο	5
	Camberl								
euThg1	ScThg1 HsThg1 DdiTLP1	QTIKDYLSWR.Y QTLKDYLSWR.QZ ENIKDYLSWR.QZ	D T H I NNL YN T H AD C H I NNL YN T H AD C H I NNL YN T Y AD T H I NNL YN T C	WQLIIKCGI WALIQQSGI WALVLKAGP	TPQESEKI TPVQAQGI TPIEAENI	LCGTFSNEKQEI LQGTLAADKNEI LRGTFSDGKNEM	LFS.ECGINYNNEI LFS.EFNINYNNEI LFS.RFNINYNNLI	EMTRKGILL PMTRKGIVL PAETRKGSVI	TRKGE IWQKVDE. FKKPVQE.
arThg1	MaThg1 MtThg1 MkThg1	EEIPEYFHRR.QI DIVPEYFRSR.QI DDLVDYFKWR.OI	LEAWRNFVASWG ESWRNCLNSYA EAWRNHLNSYA	YALRN.EGN WTLRREV WALRE.RGI	GRNEAAK GARRAAEI KPKEAAEI	LKRKKESEIHEM LRGLKSDSLHDL LRGMKAHDVHEL	LFERGINLATLE LFERGVNISRVE LYR.EFGINLGRTE	SWQRRGVII AWQRRGVGV AWORRGILA	SKEARE YRVPHR YRVAVNED
TLP	DdiTLP2 DdiTLP3 DdiTLP4	NEVKDYFIKR.QO DEIRHNLIWRSI QEIVNNLIWRSL	RCYTNFLSSIAS DCKRNSVSQVG DCKRNSVANLG HTNNLVN	SDNGFNTDS] SHFLP TAHFTP	IFNNDS( 	SLDGSKKNIYARE 21HGLSGQEIKKK MLGLNNTEVKKK	YYLEKQGVNLMEI LLL.EKGIDFNDE LLE.EKSIIYENE	PEHFKSGFLV DWYKYGVYL PAWYRFGTYL	F <mark>KNSTN</mark> K <mark>KQ</mark> NYTHK K <u>KE</u> YYTLT
		disordered regio	on (tRNA comp	lex)					
		β6	β7	β8		α9			
	Hsingi CaThql		0000000000	<b>β</b> 8	د بە	α9 η2			
	carngr	230	240	250		260			
• <b>T</b> h = 4	CaThg1 ScThg1	.ETEDEAEL.SKI	QVQRLEKKRKK	AELKIYHVDI	IINDD	SWW <mark>K</mark> SRPWLKD ELFEGY			 
eu i hg1	HsThg1 DdiTLP1	.VMTKEIKLPTEN .TN	AEGKKMAVTRTRT	KPVPLHCDI KRLVIEHVDI		AFWKEHPEILDED: KFWKEYPDILASK	S		· · ·
arThg1	MaThg1 MtThg1	IQGFNI	VSGKEEKSLRRI PIRDEEVSAVRMI	KITQNWEIPH RVKVDLELPI	KFKSEKGII	PFLEKLINRN EFFEGLMR	· · · · · · · · · · · · · · · · · · ·		· · · ·
	DdiTLP2	G TOTAL	ELDYKS.	. PINFKSDI	FIID. YL		ENGNDL ST SNOT		 
TLP	DdiTLP4		EPEKOITAIRSKI	VNCFSFDITE	KLSNSLNF:	ISFKTLPEGFNLE	LIESK		 

**Fig. 57.** Multiple sequence alignment of eukaryotic Thg1 (euThg1), archaeal Thg1 (arThg1), and TLP enzymes. Secondary structures of CaThg1 and HsThg1 are indicated at the top ( $\alpha$ ,  $\alpha$ -helix;  $\beta$ ,  $\beta$ -strand;  $\eta$ ,  $3_{10}$ - turn). The residues interacting with ATP/GTP, magnesium ions, the anticodon loop, and the acceptor stem are marked with circles, diamonds, stars, and triangles. The HINNLYN motif is highlited with an orange box. The species aligned are as follows: CaThg1, *C. albicans*; ScThg1, *S. cerevisiae*; HsThg1, *Homo sapiens*; MaThg1, *Methanosarcina acetivorans*; MtThg1, *Methanothermobacter thermautotrophicus*; MkThg1, *Methanopyrus kandleri*; and DdiTLP1-4, *Dictyostelium discoideum*. Each protein sequence was aligned by CLUSTALW, and the figure was prepared with the program ESPript.



**Fig. S8.** (A) In the CaThg1-ATP structure, two ATP molecules (yellow sticks) were observed in the palm domain. ATP1 binds to the adenylylation site with  $Mg^{2+}$  A and  $Mg^{2+}B$ . ATP2 binds to the incoming nucleotide-binding site with  $Mg^{2+}C$ . The residues interacting with ATP molecules and magnesium ions are shown as stick models. The omit map for two ATP molecules and the side chain of Lys44 is represented as a gray mesh (2.5  $\sigma$ ). (*B*) In the CaThg1-GTP structure, two GTP molecules (magenta sticks) were detected in the palm domain. The omit map for two GTP molecules is represented as a gray mesh (2.5  $\sigma$ ).



**Fig. S9.** (*A*) Stereoview of the proposed guanylylation model, which was generated by superposing the CaThg1-tRNA, CaThg1-ATP, and CaThg1-GTP structures. The arrow indicates possible conformational changes of GTP induced by tRNA binding. (*B*) The proposed base movements of GTP ( $G^{-1}$ ) and A73. The magenta dash line indicates the proposed G-A base pairing between  $G^{-1}$  and A73.

Table S1.	Data collection	and refinement	statistics	(molecular replaced)	placement)
-----------	-----------------	----------------	------------	----------------------	------------

	CaThg1-ATP	CaThg1-GTP	$CaThg1-tRNA^{His}\Delta G^{-1}$	CaThg1-tRNA <sup>Phe</sup> GUG
Data collection				
Space group	<i>P</i> 2 <sub>1</sub>	P2 <sub>1</sub>	<i>P</i> 3 <sub>1</sub> 21	<i>P</i> 3 <sub>1</sub> 21
Cell dimensions				
a, b, c (Å)	74.8, 217.9, 87.2	87.4, 217.6, 140.9	96.5, 96.5, 300.4	96.7, 96.7, 299.3
α, β, γ (°)	90, 113.7, 90	90, 102.3, 90	90, 90, 120	90, 90, 120
Resolution (Å)*	50–2.39 (2.54–2.39)	50–3.03 (3.21–3.03)	50–4.18 (4.43–4.18)	50-3.64 (3.86-3.64)
R <sub>meas</sub> (%)* <sup>,†</sup>	4.3 (30.6)	8.3 (85.6)	12.7 (100.3)	11.3 (118.0)
//σ (/)*	23.93 (5.19)	16.50 (2.03)	12.45 (2.48)	16.68 (2.05)
Completeness (%)*	97.1 (89.8)	98.8 (96.5)	98.8 (94.0)	99.6 (98.1)
Redundancy*	3.8 (3.5)	3.9 (3.9)	10.6 (10.7)	9.5 (9.6)
Refinement				
No. reflections	97,819	98,553	12,539	18,933
R <sub>work</sub> /R <sub>free</sub> (%)	17.62/20.82	19.83/23.93	29.22/33.48	25.29/28.75
No. of atoms				
Protein	17,504	35,254	8,128	8,128
Ligand/ion	520	1072	-	-
Water	365	15	-	-
RNA	-	-	3,102	3,146
B-factors (Å <sup>2</sup> )				
Protein	49.6	84.4	116.7	89.7
Ligand/ion	46.0	88.2	-	-
Water	36.1	42.6	-	-
RNA	-	-	183.1	144.2
Estimated coordinate error (Å) <sup>‡</sup> rmsd from ideal	0.26	0.41	0.53	0.69
Bond lengths (Å)	0.009	0.005	0.003	0.004
Bond angles (°)	1.261	0.999	0.807	0.761

Each data set was collected from a single crystal.

\*Values in parentheses are for the highest-resolution shell.

 ${}^{\dagger}R_{\text{meas}}$  is the redundancy-independent  $R_{\text{sym}}$ .

PNAS PNAS

\*Maximum-likelihood estimates using test set reflections given by phenix.refine.

## Table S2. Estimated experimental molecular weight of CaThg1 and CaThg1 with $tRNA^{His}\Delta G^{-1}$ by SAXS

Sample name	Theoretical mass (kDa)	SAXS mass (kDa)	R <sub>g</sub> (Å)
CaThg1	128	122	41.0 ± 0.2
CaThg1-CatRNA <sup>His</sup> ∆G <sup>-1</sup>	178	178	43.9 ± 0.2

The estimated molecular weight of CaThg1 with tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup> by SAXS was 173 kDa, indicating a 4:2 molar ratio of CaThg1 (130 kDa for tetramer) with two tRNA<sup>His</sup> $\Delta$ G<sup>-1</sup> molecules (48 kDa) in solution.