Supporting Information

Chu et al. 10.1073/pnas.1106610108

Materials and Methods

Guanylyltransferase (GTase) Constructs. His-tagged full-length human capping enzyme (hCE) was expressed and purified as described previously (1). The purified protein (5 mg/mL) was digested with plasmin (0.035 mg/mL) in 50 mM Tri·HCl, pH 7.0, 10 mM NaCl, 0.1% n-octyl-β-D-glucoside, 10 mM 2-mercaptoethanol, 1 mM GTP at 4 °C for 2 h, conditions established to obtain protease-resistant core fragments after many trials at different temperatures, incubation times, protease/hCE ratios, etc. Several other proteases were also tested under various conditions, and hCE was completely digested in most experiments. The GTase domain of hCE corresponding to amino acids 222–574, 224–567, 229–569, or 240–569 was selected based on SDS-PAGE and MS analysis of the protease-resistant core fragments and sequence-based secondary structure predictions. The indicated regions were PCR amplified, sequenced, and cloned into pET20b (Novagen).

GTase Expression. pET20b/GTase transformed BL21(DE3) Star cells (Invitrogen) were inoculated into 50 mL of LB medium containing 100 μg/mL ampicillin (Amp) and cultured overnight at 37 °C. The next morning cultures were transferred to 4 L of LB containing the same concentration of Amp and grown to an OD_{600} of 0.6–0.8. IPTG was added to a final concentration of 1 mM and induction was carried out at 17 °C for 16–18 h. Cells were collected, and stored at −80 °C before use for purification. The truncated proteins expressed well, and those starting with residues 222, 224, or 229 were soluble and had GTase activity similar to the full-length hCE as measured by GTase-GMP formation (15). Trimming to position 240 yielded insoluble protein after nickel (Ni)-nitrilotriacetate (NTA) affinity chromatography, while removing the C terminus to position 551, 475, or 468 resulted in loss of most of the GTase activity.

GTase Purification. Frozen cells were resuspended in 20 mM phosphate buffer (PB), pH 8.0, 10% glycerol, 5 mM $MgCl_2$, 1 mg/mL lysozyme, 1 mM PMSF, 20 mM imidazole, 5 mM 2-mercaptoethanol, and stirred at 4 °C for 30 min. Triton X-100 was added to the lysate to a final concentration of 1%, and the lysate was stirred at 4 °C for another 30 min. The lysate was freeze-thawed once and subjected to sonication to reduce viscosity. Cleared lysate was loaded onto a Ni-NTA column equilibrated with 20 mM PB, pH 7.4, 5 mM $MgCl_2$, 20 mM imidazole, 5 mM 2-mercaptoethanol. Contaminants were eluted with the same buffer mixture except 60 mM imidazole, and GTase was eluted by raising the imidazole to 200 mM. Purified GTase was exchanged into 20 mM Tris·HCl, pH 8.0, 10% glycerol, and loaded onto Heparin Sepharose Fast Flow equilibrated with the same buffer. GTase was eluted using a 0 to 1 M NaCl linear gradient. Purified GTase was pooled, concentrated, and loaded onto a Superdex 75 column equilibrated with 20 mM Tris·HCl, pH 7.5, 0.1 M NaCl, 5 mM 2-mercaptoethanol, 5 mM $MgCl_2$, and was eluted with the same buffer. Fractions containing GTase were collected and concentrated to 10 mg/mL.

Crystallization. The condition 2 M ammonium sulfate, 5% isopropanol produced small crystals of the human GTase (hGTase) constructs 229–567 and 229–569. The crystals were improved using an optimized condition in 1.7 M ammonium sulfate, 10% glucose, 5 mM 2-mercaptoethanol, 5 mM $MgCl_2$, 5% isopropanol and the hanging drop method. The protein concentration ranged from 4 to 10 mg/mL. Crystals were visible 7–14 d after drops were set up and placed in an incubator at 20 °C. The crystals were dipped in the crystallization solution containing an additional 20% glycerol for cryoprotection and cryocooled in liquid N_2. The X-ray diffraction data were collected from the cryocooled crystals at the X25 beamline, Brookhaven National Laboratory (BNL).

X-ray Crystallography. X-ray diffraction data were processed, integrated and scaled using HKL2000 software (2). Molecular replacement (MR) solutions were attempted using the *Chlorella* (PDB ID code 1CNK) and *Candida albicans* (PDB ID code 1P16) partial structures and modeled hGTase based on 1CNK and 1P16 structures. In general, four to seven monomers were found in a typical MR solution; however, no well-defined MR solution could be obtained. Therefore, each MR solution was (*i*) manually checked for presence of twofold noncrystallographic symmetry (NCS), as suggested by self-rotation function analysis of the diffraction data, and (*ii*) subjected for NCS averaging among all molecules per asymmetric unit; the NCS correlation matrices and electron density maps were verified for possible solutions. The correct MR solution consisted of two trimers initially and subsequent search found the seventh monomer using the nucleotide transferase domain of 1CNK-A molecule as the search model. Sevenfold NCS averaging produced a good quality electron density map from almost uninterpretable initial maps. The averaged map was used to build the initial model for the larger domain of hGTase. Gradual extension of the NCS mask in the initial cycles of model building, structure refinement, and NCS averaging located the oligonucleotide/oligosaccharide binding fold domains for individual monomers. The crystallographic software programs Phaser, Molrep, DM, and Refmac were used from the CCP4 program suite (3) for carrying out MR solutions, NCS averaging, and structure refinement. Programs Coot (4) and PyMol (www.pymol.org) were used for molecular modeling and for generating figures, respectively; Adaptive Poisson–Boltzmann Solver (5) was used to calculate the electrostatic potential surfaces. The final cycles of refinement were carried out using the PHENIX suite of programs (6). The phase probability distributions (Hendrickson Lattman coefficients) calculated from NCS averaged phases and figures and merit were used in the structure refinement process. Parts of the loop connecting αD and β9 are disordered in all seven monomers and the β11–β12 hairpins are missing in three monomers. The final structural model of 18,382 atoms was refined against 53,387 reflections to R and R-free of 0.258 and 0.296, respectively (Table S1).

1. Wen Y, Shatkin AJ (1999) Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev* 13:1774–1779.
2. Otwinowski Z, Minor W (2001) DENZO & SCALEPACK. *Crystallography of Biological Macromolecules*, International Tables for Crystallography, eds Rossmann MG, Arnold E (Kluwer, Boston), Vol F, pp 226–235.
3. Laskowski RA, MacArthur MW, Thornton JM (1994) Evaluation of protein coordinate datasets. *From First Map to Final Model*, Proceedings of the CCP4 study weekend, eds Bailey S, Hubbard R, Waller D (Daresbury Laboratory, Warrington, UK), pp 149–159.
4. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60:2126–2132.
5. 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* 98:10037–1004.
6. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66:213–221.

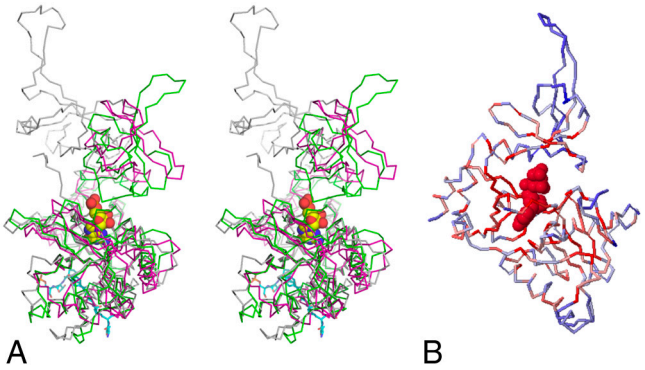

Fig. S1. Comparison of GTase structures from human, *Paramecium bursaria Chlorella* virus 1 (PBCV-1) virus and yeast. (*A*) Stereoview showing the superposition of GTase structures from human (green), PBCV-1 (magenta; PDB ID code 1CKN-A), and *C. albicans* (gray; PDB ID code 1P16-A). The GTP molecule from the PBCV-1 GTase structure is shown as a space-filled model. (*B*) A Cα trace of the hGTase molecule that is color-coded based on amino acid sequence conservation across the species (red, highly conserved to blue, highly diverged). The GTP molecule is shown as a red space-filled model.

Table S1. Diffraction data and refinement statistics

	Human GTase
X-ray source	BNL—NSLS X25
Space group	$P2_1$
Cell constants (a, b, c in Å; β in °)	92.12, 104.66, 149.57; 94.95
Resolution range, Å	50–3.0
No. of unique reflections (no. of observations)	53,410 (189,603)
R_{merge} (in last shell, 3.11–3.0 Å)	0.085 (0.718)
Completeness, % (in last shell)	95.7 (74.8)
Average $I/\sigma(I)$	14.4
Sigma cutoff	$I < -1.0\sigma(I)$
Refinement statistics	
Resolution range, Å	50–3.0
Total no. of atoms	18,382
No. of reflections (R_{free} set)	53,387 (2,724)
R_{work}/R_{free}	0.258/0.296
Root-mean-square deviation	
Bond length, Å	0.011
Bond angle, °	1.43