**Supplementary information, Table S1** Data collection and refinement statistics

| | | Gemin5-WD | Gemin5-WD/$m^7$GpppG | Gemin5-WD/Sm RNA |
|---|---|---|---|---|
| **Data collection** | | | | |
| Space group | | $P2_1$ | $P2_1$ | $P2_1$ |
| Cell Parameters (Å) | a | 76.5 | 113.2 | 76.8 |
| | b | 91.4 | 91.0 | 92.1 |
| | c | 112.6 | 153.7 | 110.8 |
| (°) | β | 100.5 | 100.6 | 99.9 |
| Resolution (Å) | | 30.0-3.00 (3.16-3.00) | 50.0- 2.57 (2.61 -2.57) | 48.8-2.49 (2.57- 2.49) |
| $^aR_{merge}$ | | 0.11(0.88) | 0.07 (0.73) | 0.12 (0.94) |
| $CC_{1/2}$ | | 0.99 (0.77) | 0.99 (0.72) | 0.99 (0.60) |
| I/sigI | | 11.4 (2.2) | 16.4 (2.0) | 9.6 (1.6) |
| Completeness | | 99.0 (99.6) | 97.9 (71.1) | 96.9 (78.2) |
| Redundancy | | 5.8 (6.0) | 7.0 (5.5) | 5.7 (5.1) |
| **Refinement** | | | | |
| Resolution (Å) | | 30.0-3.0 (3.09-3.00) | 50.0-2.57 (2.59-2.57) | 48.8-2.49 (2.54-2.49) |
| No. of reflections | | 30,397 (2625) | 96,106 (2075) | 51,648 (2351) |
| $^bR_{work}/^cR_{free}$ (%) | | 20.0 / 26.0 | 17.1 / 22.8 | 20.1 / 25.6 |
| R.m.s deviation | | | | |
| Bond length (Å) | | 0.004 | 0.009 | 0.009 |
| Bond angle (°) | | 0.7 | 1.1 | 1.0 |
| Ramachandran Plot | | | | |
| % favored | | 94.4 | 95.2 | 97.1 |
| % allowed | | 5.0 | 4.0 | 2.4 |
| % outliers | | 0.6 | 0.7 | 0.5 |
| No. of atoms | | | | |
| Protein | | 10146 | 20376 | 10302 |
| Sm RNA | | - | - | 282 |
| $m^7$Gppp | | - | 132 | - |
| Other ligands/ions | | 6 | 18 | 6 |
| Water | | - | 120 | 153 |

Values in parentheses correspond to those of highest resolution shell.
$^aR_{merge} = \sum|I_j - \langle I \rangle| / \sum I_j$, where $I_j$ is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.
$^bR_{work} = \sum||F_o| - |F_c|| / \sum|F_c|$, where $F_o$ denotes the observed structure factor amplitude, and $F_c$ denotes the structure factor amplitude calculated from the model.
$^cR_{free}$ is as for $R_{work}$ but calculated with 5.0% of randomly chosen reflections omitted from the refinement.

## Material and Methods

***Protein expression and purification:*** Nucleotide sequence corresponding to human Gemin5-WD (aa 1-740) was cloned into pFastBac HTA (Thermo Fisher Scientific) and expressed as His-tagged proteins in Sf9 insect cells using Sf-900™ III SFM medium (Gibco). The mutants were created using QuikChange II XL Site-Directed

Mutagenesis Kit (Agilent Technologies). Wild type and mutant proteins were purified by HisPur™ Cobalt Resin (Thermo Fisher Scientific), ion exchange and gel filtration chromatography (GE healthcare). The purified proteins were concentrated to 15 mg $mL^{-1}$ and stored in a buffer containing 20 mM Tris pH 8.0, 150 mM NaCl and 10% glycerol.

***Crystallization and Structure determination:*** The Gemin5-WD apo crystals were obtained in the crystallization condition containing 0.1 M Na/K phosphate pH 6.2, 20% PEG 3350 by hanging-drop vapor diffusion method at 15 °C. The crystals were cryo-protected using mother liquor supplemented with 30% (v/v) glycerol before flash freezing in liquid nitrogen. The Gemin5-WD/$m^7$GpppG complex crystals were grown by mixing 1 uL protein/Cap mixture (1:2 molar ratio) with 1 uL reservoir solution containing 0.1 M Bis-Tris propane pH 6.5, 0.2 M Na/K phosphate and 20% PEG 3350 and 30% (v/v) glycerol supplemented reservoir was used as cryo-protectant. Gemin5-WD/Sm RNA complex is mixed in molar ratio of 1:1.5 and purified by S200 10/300GL chromatography (GE healthcare). The crystallization and cryo conditions are same as for Gemin5-WD/$m^7$GpppG. The X-ray diffraction data of the crystals were collected on beamlines PX-I (SLS, PSI, Switzerland), ID-23-1 (ESRF, Grenoble, France) and BL13B1 (NSRRC, Taiwan). The data were processed by XDS[1]. The Gemin5-WD apo structure was solved by molecular replacement with PHASER[2] using Protein Data Bank entry 2ymu as a model. The polarity of the RNA and identity of the nucleotides are defined unambiguously by checking the simulated-annealing omit map and the possible interactions of the nucleotides with the neighboring protein residues. The structures were further manually built using COOT[3], refined using Phenix[4] and REFMAC5[5]. Structure validation was performed using Molprobity[6] and PROCHECK[7]. The statistics for the diffraction data and structure refinement are listed in Supplementary information, Table S1.

***Cap-affinity chromatography:*** Cap-binding reactions were carried out with 150 μg of purified protein (wild type or mutant Gemin5-WD) with 20 μL agarose immobilized with gamma-Aminophenyl-$m^7$GTP (C10-spacer) (Jena Biosciences) in the buffer with 10 mM Tris pH 8.0 and 150 mM NaCl. The reaction was incubated at 4 °C for 1 h. After washing the beads for 5 times, 25 μL of buffer supplementary with 1 mM $m^7$GpppG analog was used to elute the protein bound. The input and eluted proteins were subjected to SDS-PAGE.

***Fluorescence anisotropy assay:*** The measurements were performed using Sm RNA with 6-carboxy-fluorescein (6-FAM) labeled at the 3′ ends (Integrated DNA Technologies). Reactions of 100 µL were carried out in a 96-well flat bottom black plate by the incubation of 100 nM of the labeled RNA with increasing protein (from 100 nM to 6.4 µM) at ambient temperature for 30 minutes prior to measurements, in a buffer containing 10 mM Tris pH 8.0 and 150 mM NaCl. Then the samples were measured using Safire II fluorescent plate reader (Tecan) by using the excitation and emission wavelengths were 470 and 535 nm, respectively. Dissociation constants (Kd) values were determined by fitting the experimental data to a binding equation describing a single-site binding using Prism software 4.0 (GraphPad).

***RNA binding assay:*** Native acrylamide electrophoretic mobility shift assay (EMSA) was used to detect the RNA binding. The experiments were performed by the addition of Sm RNA (nt 118-127 of pre-U4) with 3′ end 6-FAM (Integrated DNA Technologies) or longer U4-RNA (nt 85-145 of pre-U4, prepared by in-vitro transcription) to the equivalent amounts of wild type Gemin5-WD in the absence or presence of 2.5-fold excess of $m^7$GpppG and incubated on ice for 1h. Then the samples were electrophoresed at 80 V for 120 min in 8% TBE native gel and quantified with Chemidoc (Bio-rad).

References


1 Kabsch W. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:125-132.
2 Storoni LC, McCoy AJ, Read RJ. *Acta Crystallogr D Biol Crystallogr* 2004; **60**:432-438.
3 Emsley P, Cowtan K. *Acta Crystallogr D Biol Crystallogr* 2004; **60**:2126-2132.
4 Adams PD, Afonine PV, Bunkoczi G *et al. Acta Crystallogr D Biol Crystallogr* 2010; **66**:213-221.
5 Murshudov GN, Skubak P, Lebedev AA *et al. Acta Crystallogr D Biol Crystallogr* 2011; **67**:355-367.
6 Chen VB, Arendall WB, 3rd, Headd JJ *et al. Acta crystallographica Section D, Biological crystallography* 2010; **66**:12-21.
7 Laskowski RA, Macarthur MW, Moss DS, JM T. *J Appl Cryst* 1993; **26**:283-291.