Supplementary Information.

Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I

Swapnil C. Devarkar, Chen Wang, Matthew T. Miller, Anand Ramanathan, Fuguo Jiang, Abdul G. Khan, Smita S. Patel, and Joseph Marcotrigiano.

SI Materials and Methods

RNAs. All RNAs were chemically synthesized and HPLC purified by Bio-Synthesis Inc., TriLink BioTechnologies, or IDT (Supplementary table S1). The synthetic RNAs were analyzed for purity by mass spectrometry (Supplementary figure S8) and HPLC. The lyophilized RNA was resuspended in 20 mM potassium phosphate buffer pH 7.0. Duplex RNA were prepared by mixing complementary ssRNAs in a 1:1.1 ratio, heating to 95°C for 1 min, and slow cooling to 4°C The concentration was determined in 7 M guanidinium HCl using the NanoDrop spectrophotometer at A_{260}.

RNA binding and ATP hydrolysis. The ATP hydrolysis rates were measured at constant RIG-I (5 nM) and increasing RNA concentration (1 nM – 1 μM) in the presence of 1 mM ATP (spiked with [γ-^{32}P] ATP). A time course (0-60 min) of the ATPase reactions was performed in Buffer A at 37°C. The reactions were stopped at desired time points using 4 N HCOOH and analyzed by PEI-Cellulose-F TLC (Merck) developed in 0.4 M potassium phosphate buffer (pH 3.4). The TLC plates were exposed to a phosphorimager plate, imaged on a Typhoon phosphor-imager, and quantified using ImageQuant software. The ATPase rate was determined from the plots of [Pi] produced versus time and dividing the rate of hydrolysis by RIG-I concentration. The ATPase rates were then plotted as a function of RNA concentration and fitted to hyperbolic equation (Equation 1) or quadratic equation (Equation 2) to get the binding affinity ($K_{d,app}$) and the maximal ATPase rate (k_{atpase}), where observed ATPase rate = k_{atpase} x [PR]/[Pt] ; [Pt] is total protein concentration.

$$[\mathrm{PR}] = \frac{[\mathrm{S}]}{K_{\mathrm{d,app}}+[\mathrm{S}]} \qquad \text{(Equation 1)}$$

$$[\mathrm{PR}] = \frac{([\mathrm{P_t}]+[\mathrm{R_t}]+K_{\mathrm{d,app}})-\sqrt{([\mathrm{P}_t]+[\mathrm{R_t}]+K_{\mathrm{d,app}})^2-4[\mathrm{P_t}][\mathrm{R_t}]}}{2} \qquad \text{(Equation 2)}$$

Protein expression and purification. All the protein constructs were expressed in pET28 SUMO vector. Human full-length RIG-I (1–925) and Helicase-RD (232–925) were expressed in *Escherichia coli* strain Rosetta (DE3) (Novagen) as soluble proteins. The soluble fraction of Helicase-RD was purified from the cell lysate using a Ni^{2+}-nitrilotriacetate (Qiagen) column. The recovered protein was then digested with Ulp1 protease to remove the 6xHis–SUMO tag and further purified by hydroxyapatite column (CHT-II, Bio-Rad) and heparin sepharose column (GE Healthcare). Finally, purified Helicase-RD was dialyzed against 50 mM HEPES (pH 7.5), 50 mM NaCl, 5mM $MgCl_2$, 5 mM DTT, 10% glycerol overnight at 4°C, snap frozen in liquid nitrogen, and stored at -80 °C.

Preparation of Helicase-RD-HP RNA complexes. The Cap-0 HP RNA oligonucleotide (TriLink), 5'ppp HP RNA oligonucleotide (BioSynthesis) and 5'OH HP RNA oligonucleotide (IDT) were resuspended with DNase and RNase-free water before use.

The resulting HP RNAs were mixed with purified Helicase-RD in a RNA:protein molar ratio of 1.2:1, incubated on ice for 30 min, and then purified by size-exclusion chromatography (Superdex 200, GE Healthcare) with an elution buffer of 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT.

Crystallization, X-ray diffraction data collection, structure determination and refinement. Crystals of native complex were grown by the hanging-drop vapor diffusion method at 20° C. Aliquots (2 µl) of 8 mg/ml of Helicase-RD–Cap-0 HP RNA complex in 50 mM HEPES (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT were mixed with 2 µl of reservoir solution containing 20% (w/v) PEG 3350, 0.2 M NaSCN, 100 mM MOPS (pH 7.8), 3.5% (v/v) 2,2,2-trifluoroethanol. Crystals appeared after 1-2 days and grew to a maximum size of 0.15 × 0.15 × 0.5 mm over the course of six days. For cryogenic data collection, Helicase-RD-Cap-0 HP RNA crystals were transferred into crystallization solutions containing 15% (v/v) glycerol and 10% (v/v) ethylene glycol as cryoprotectant and then flash-cooled at 100 K.

Crystals of RIG-I Helicase-RD–5'ppp HP RNA complex were grown from 23% (w/v) PEG 3350, 0.25 M KSCN, 100 mM MOPS (pH 7.8). Crystals of RIG-I Helicase-RD–5'OH HP RNA complex were obtained in similar conditions as Helicase-RD-5'ppp HP RNA complex, except 0.3 M NaSCN was used instead of 0.25 M KSCN. Crystals were transferred to crystallization solutions plus 5% (v/v) (2R,3R)-(−)-2,3-butanediol, and then flash-frozen in liquid N_2.

X-ray diffraction data was collected at the National Synchrotron Light Source at Brookhaven National Labs, Cornell High Energy Synchrotron Source at Cornell University, and Advanced Light Source at University of California at Berkeley. Data was processed using MOSFLM and Aimless (1-4). Structure phasing was accomplished by molecular replacement using the coordinates from PDB ID 5E3H and the software package PHENIX (5, 6). The coordinates were refined by iterative rounds of manual refinement using COOT (7) and automated model refinement using PHENIX(8). Figures were generated using PyMol (9) and analysis was performed using PISA (10).

Cell Based IFN-β Reporter Signaling Assays. HEK293T cells were grown in 6-well plates to 60% confluence and cotransfected with firefly luciferase reporter plasmid (pLuc125 / 2.5 µg), Renilla luciferase reporter plasmid (pRL-TK / 500 ng), and a plasmid carrying either the wt RIG-I gene, the H830A RIG-I mutant gene or an empty plasmid under the constitutively active CMV promoter (pcDNA 3.1 / 2 µg). The firefly luciferase gene is under the interferon β promoter and the Renilla luciferase plasmid is under the constitutively active TK promoter. The plasmid transfections were carried out with X-tremeGENE HP DNA Transfection Reagent (Roche). Cells were replated in 96-well plates the next day at 2 x 10^4 cells/well density and transfected with each of the RNA ligands (700 nM final concentration in 110 µl) using Lipofectamine transfection reagent

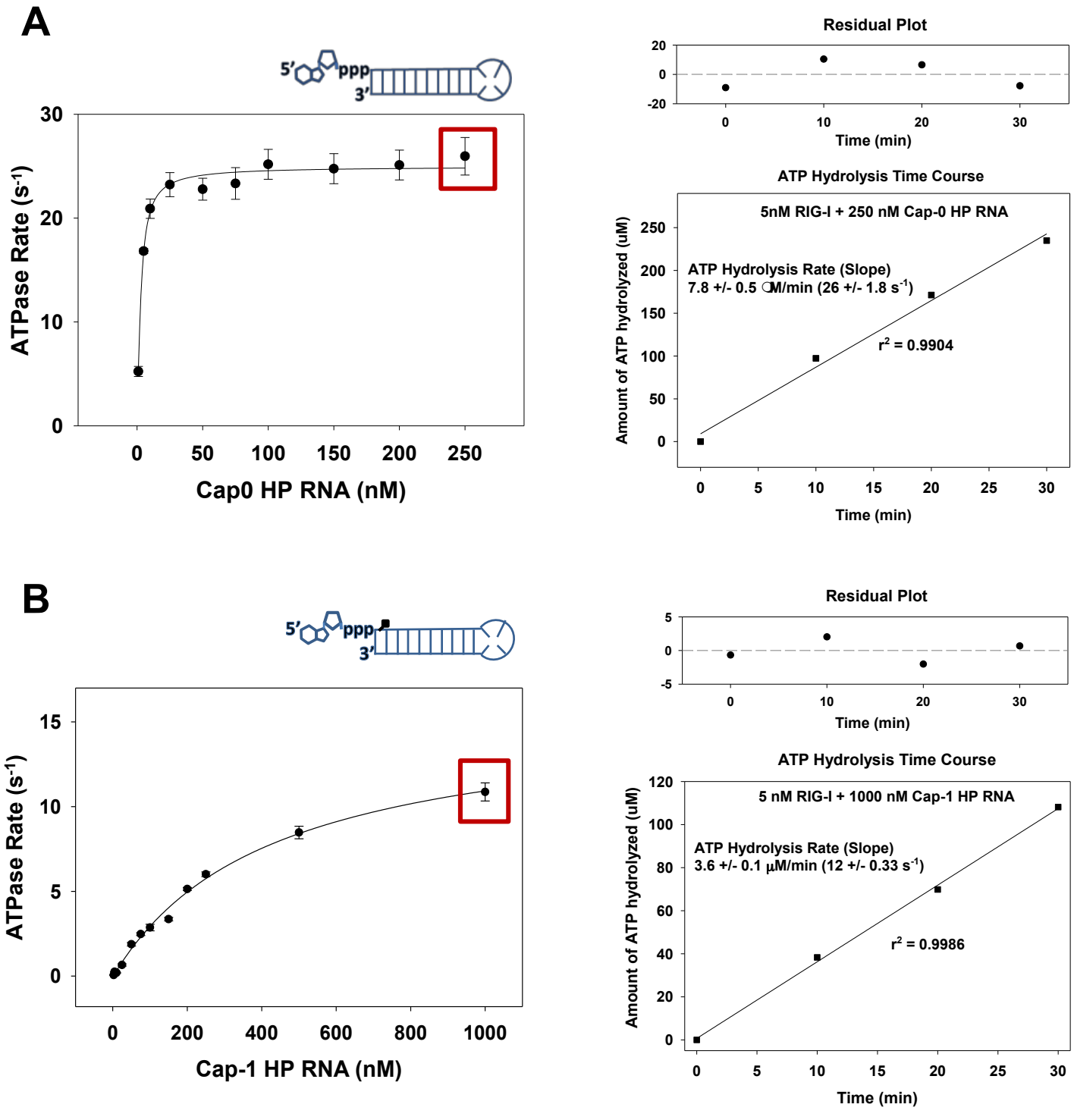
(Life Technologies). After 20 hours the activities of firefly and Renilla luciferases were measured sequentially with the Dual-Luciferase reporter assay (Promega). Data was collected in quadruplicate sets and the relative luciferase activities were calculated. The error bars represents the standard error of the mean (SEM).

Western Blot Analysis. The protein expression levels of wt RIG-I and RIG-I mutants in HEK293T cells was assessed by Western blots. Approximately equal number of cells, transfected with the respective RIG-I constructs, were lysed and an appropriate dilution of the lysates was run on a 4-15% gradient polyacrylamide gels. The samples were transferred to a nitrocellulose membrane (Hybond ECL, GE) and then probed with antibodies against RIG-I and β-actin (loading control). 5% milk solution was used for blocking and then probed with anti-RIG-I antibody (Adipogen) (1:1000 dilution in 1X PBST) or anti-β-actin antibody (Cell Signaling) (1:1000 dilution in 1X PBST) overnight. A secondary antibody against mouse IgGs conjugated with HRP (Cell Signaling) (1:15,000 dilution in 1X PBST) was used for detection of the primary antibodies. The bands were visualized using a chemiluminescent substrate kit (SuperSignal West Pico, Thermo Scientific) and appropriate exposure on an X-ray film (GeneMate).

SI References

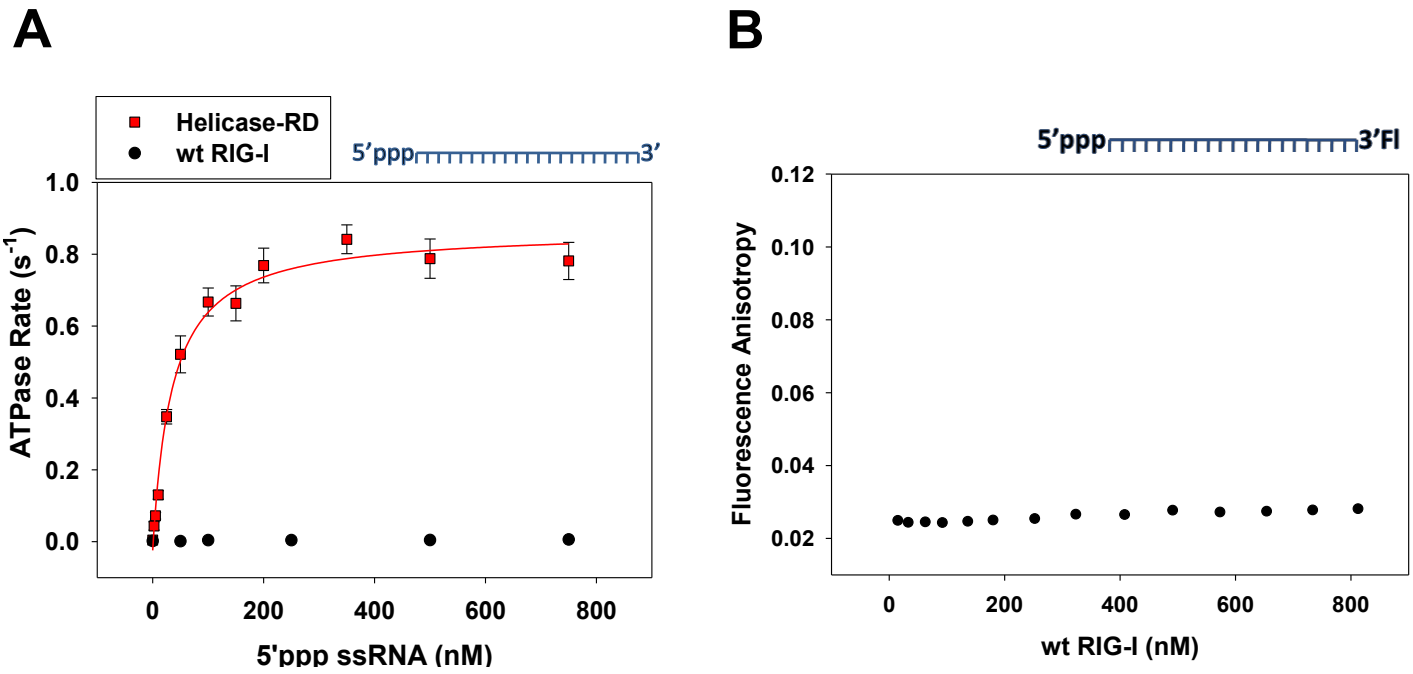
1. Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* 55(Pt 10):1690-1695.
2. Powell HR, Johnson O, & Leslie AG (2013) Autoindexing diffraction images with iMosflm. *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1195-1203.
3. Battye TG, Kontogiannis L, Johnson O, Powell HR, & Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):271-281.
4. Leslie AG (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):48-57.
5. Terwilliger TC*, et al.* (2012) phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J Struct Funct Genomics* 13(2):81-90.
6. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
7. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
8. Afonine PV*, et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):352-367.
9. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
10. Winn MD*, et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235-242.

Supplementary Figure S1



Supplementary Figure S1 (Related to Figure 1 and 3). Binding and ATPase activity of RIG-I for Cap-0 (A, left panel) and Cap-1 HP RNA (B, left panel). The curves are fit to quadratic equation to obtain the k_{atpase} and $K_{d,app}$ values. Error bars are from the time courses of the ATPase rate measurements at each RNA concentration (right panels). The highest RNA concentration used in the titration is highlighted in red and the corresponding time course is shown in the right-hand panels for representative purposes. The time-course experiments were fit to a linear equation and the corresponding residual plot is shown on top.

Supplementary Figure S2

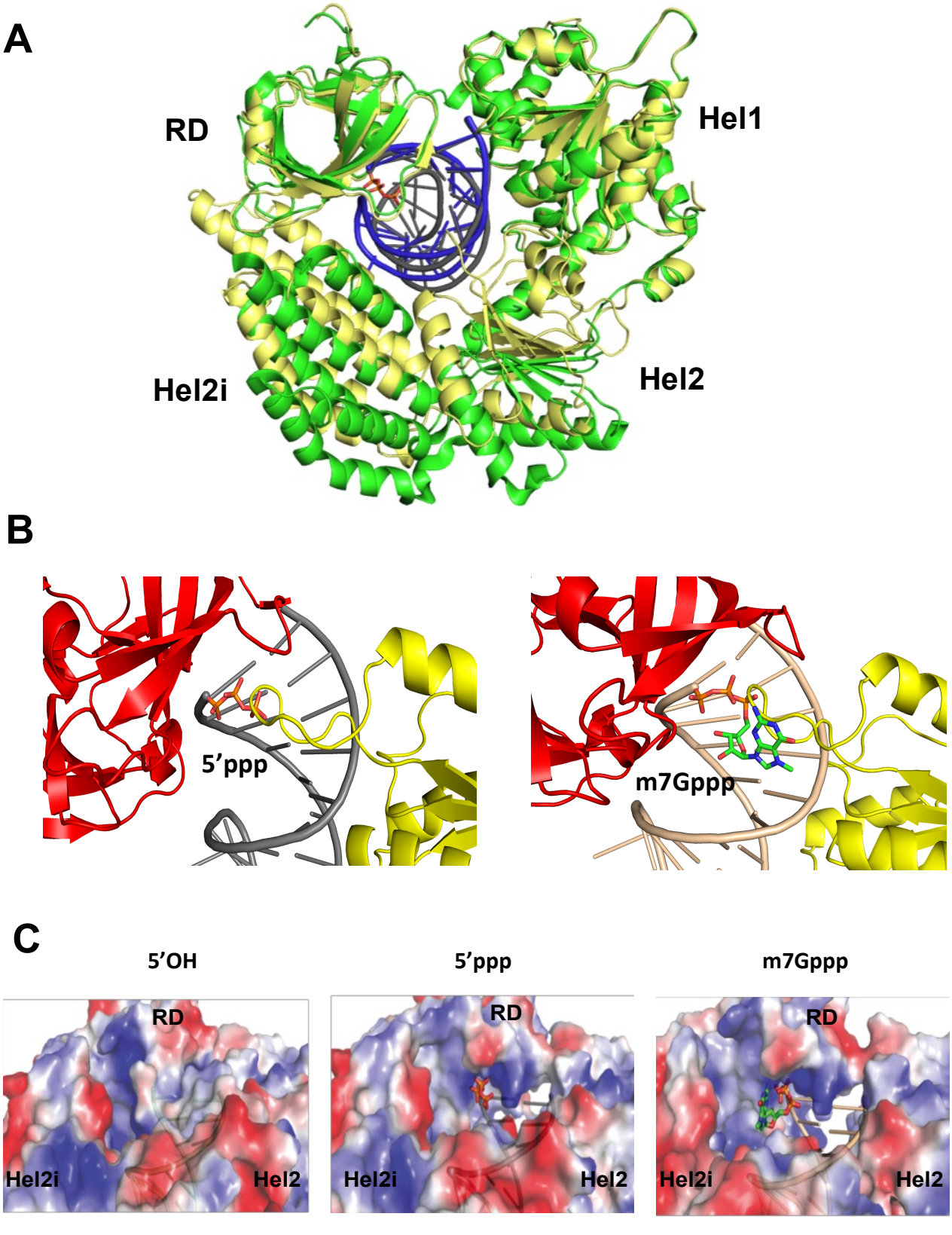


C

	wt RIG-I		Helicase-RD	
RNA Ligand	$K_{d,app}$ (nM)	k_{atpase} (s^{-1})	$K_{d,app}$ (nM)	k_{atpase} (s^{-1})
5'ppp ssRNA	N.B.	N.B.	35 ± 5.5	0.9 ± 0.03
Cap-0 ssRNA	N.B.	N.B.	14 ± 4.4	0.3 ± 0.03

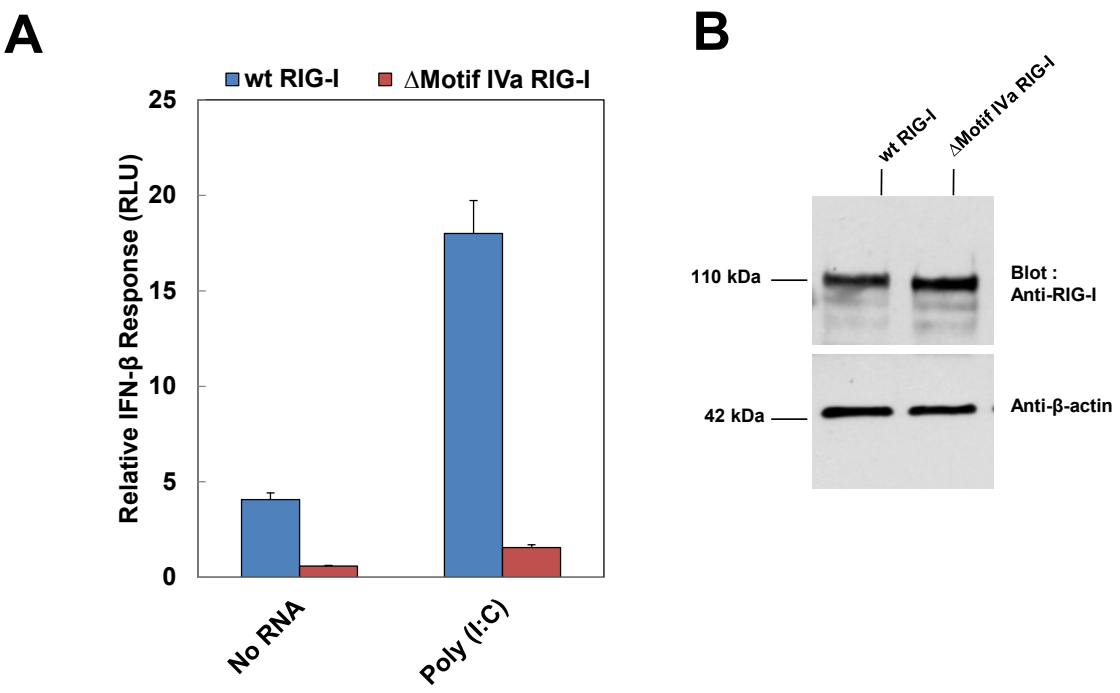
Supplementary Figure S2 (Related to Figure 1). (A) A fixed concentration of wt RIG-I or Helicase-RD protein is titrated with increasing concentration of 27nt single stranded 5'ppp RNA. The binding data are fit to the hyperbolic equation to obtain the k_{atpase} and $K_{d,app}$ values. Error bars are from the time courses of the ATPase rate measurements at each RNA concentration. (B) A fixed concentration of 3' Fluorescein (3'Fl) tagged 10 nt 5'ppp single-stranded RNA (ss10) is titrated with increasing concentration of wt RIG-I. The change in anisotropy of the fluorescein tagged RNA was monitored during the titration for measuring binding. No detectable binding was observed between wt RIG-I and 5'ppp ss10 RNA. (C) The $K_{d,app}$ and k_{atpase} values of Helicase-RD for the respective single-stranded RNAs. wt RIG-I showed no detectable binding for single stranded RNAs (N.B. stands for 'No binding').

Supplementary Figure S3



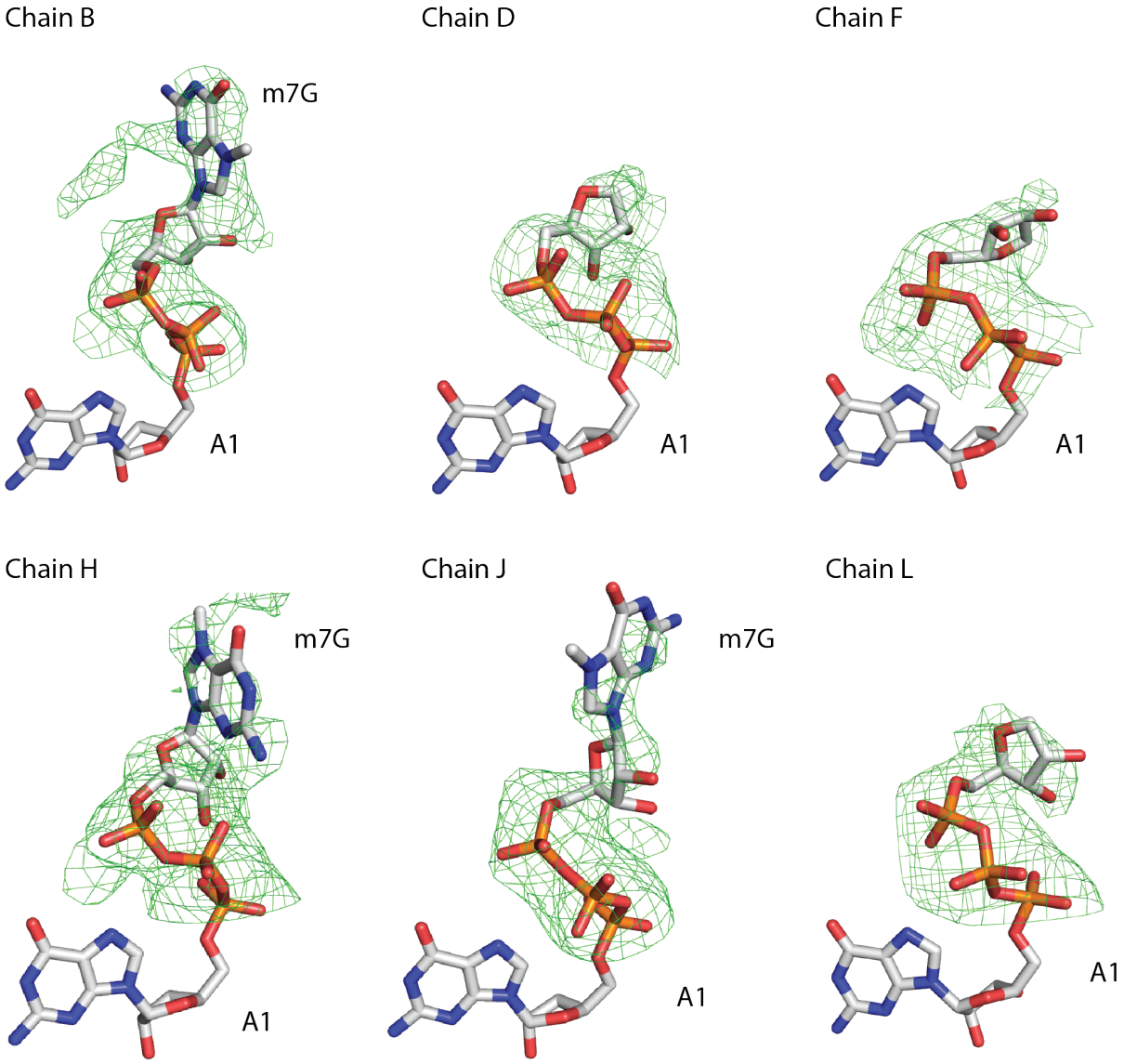
Supplementary Figure S3 (Related to Figure 1 and 2). (A) Overview of the Helicase-RD-5'ppp HP RNA structure (gold) aligned with the previously reported Helicase-RD structure with 5'ppp 8-bp HP RNA (4AY2, green). (B) Crystallographic structure of Helicase-RD with 5'OH HP RNA overlayed on the Helicase-RD-5'ppp HP RNA structure (left) and Helicase-RD-Cap-0 HP RNA structure (right) showing the steric clash of Hel2 extended loop-helix (673-685) formed in presence of 5'OH with the 5'ppp and m7Gppp moiety of 5'ppp and Cap-0 HP RNA respectively. (C) Molecular surface of Helicase-RD in the presence of 5'OH HP RNA (Left), 5'ppp HP RNA (Middle) and Cap-0 HP RNA (Right) colored for electrostatic potential. The view is identical to panel A of figure 2. Disruption of the Hel2 loop and Motif IVa helix (664-685) creates a pocket for the ppp and m7G moieties.

Supplementary Figure S4



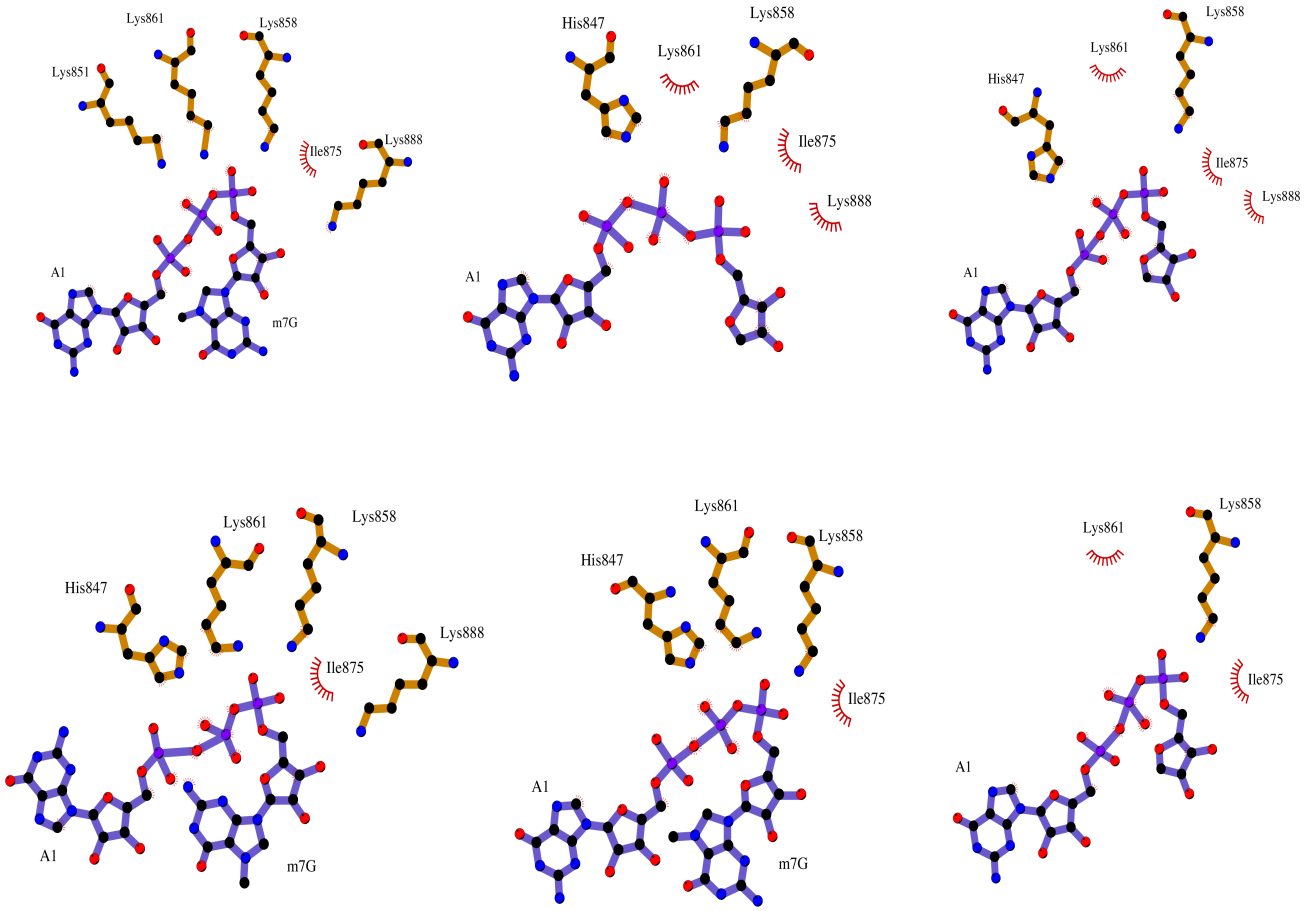
Supplementary Figure S4 (Related to Figure 2). (A) IFN response of wt RIG-I and the Motif IVa deletion mutant (Δ673-685) for Poly(I:C) as well as their respective background signal is shown (B) Western blots for wt RIG-I and the Motif IVa deletion mutant (Δ673-685) as well as their respective loading controls (β-actin) are shown.

Supplementary Figure S5



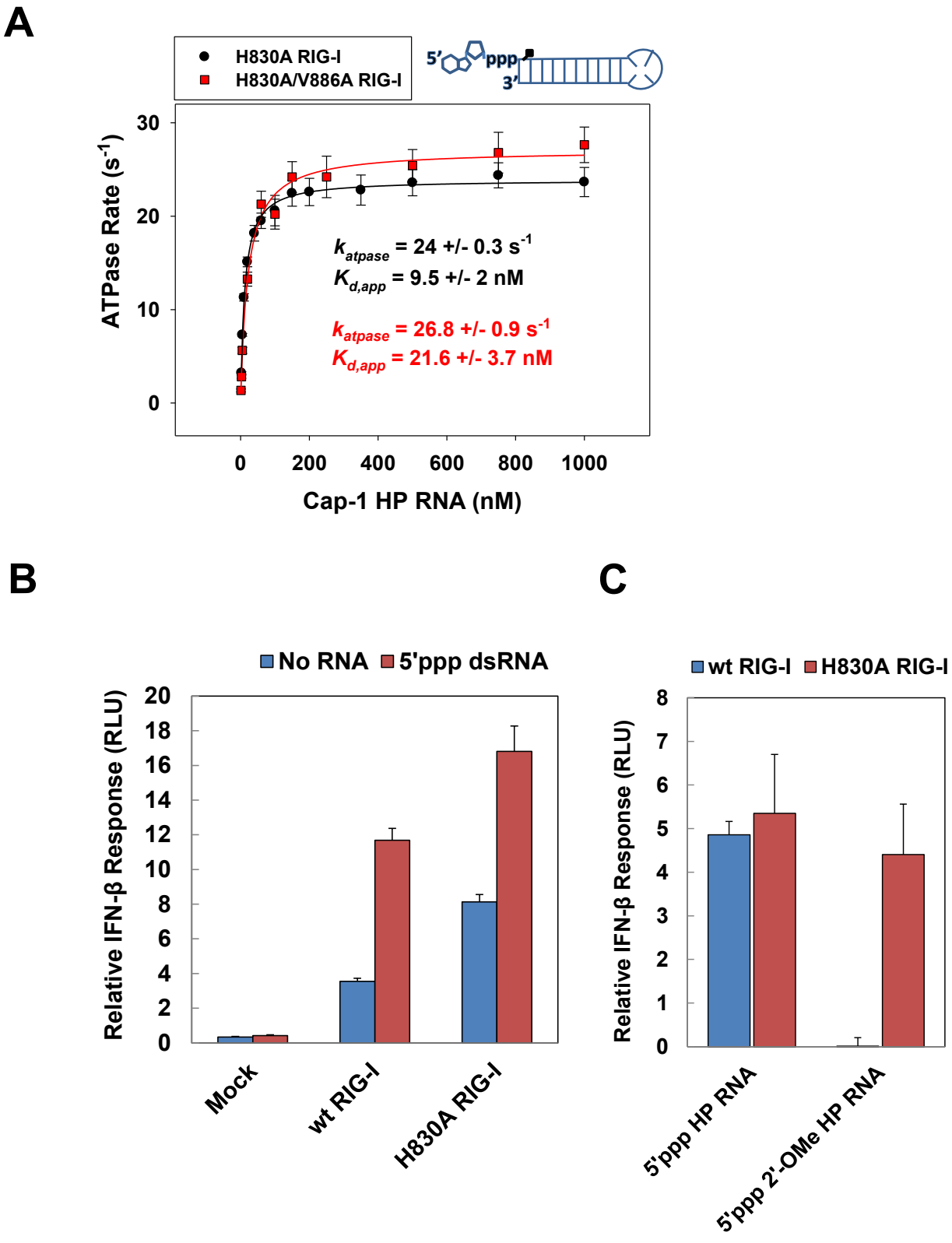
Supplementary Figure S5 (Related to Figure 2). F_o-F_c omit density map calculated without the triphosphate, ribose and m7G moieties; and overlaid with the 5' end of the RNAs from the model. The first adenine of the double-stranded portion (A1), which was included in the model, is shown for orientation purposes. The map is contoured at 2.3σ.

Supplementary Figure S6



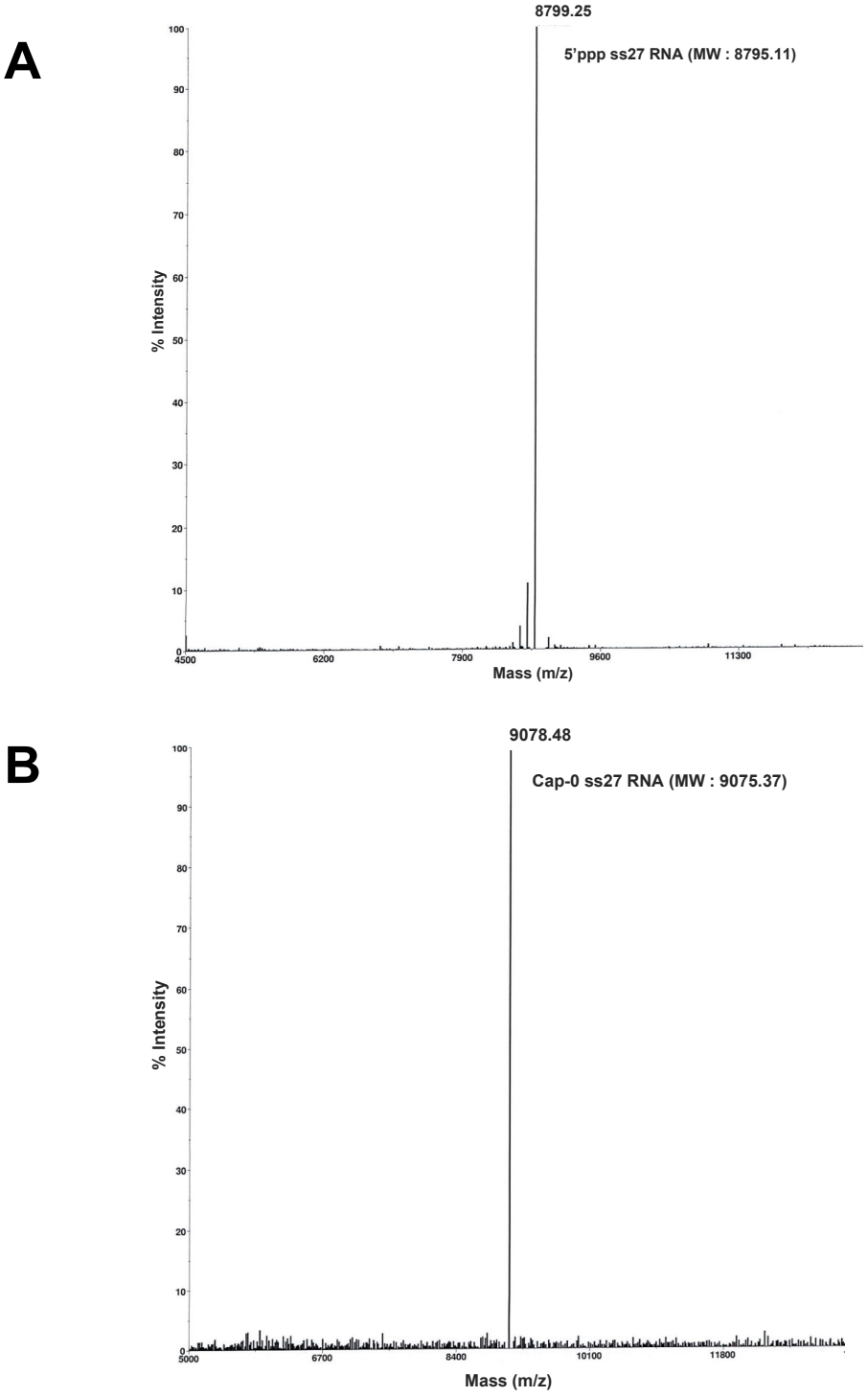
Supplementary Figure S6 (Related to Figure 2). Schematic diagram highlighting RIG-I Helicase-RD residues within 3.6 Å of the 5' end of the RNA in all six complexes in the asymmetric unit. RIG-I Helicase-RD contacts the triphosphate without any specific interactions with m7G base. Given the conformational flexibility of the caps in the six complexes, residues that are 3.6 Å away from the triphosphate are not shown.

Supplementary Figure S7



Supplementary Figure S7 (Related to Figure 3 and 4). (A) Binding and ATPase activity of H830A RIG-I and H830A/V886A RIG-I for Cap-1 HP RNA. The data depicted above is from radiometric ATPase assay wherein a fixed concentration of protein is titrated with increasing concentration of RNA. The curves are fit to the quadratic equation to obtain the k_{atpase} and $K_{d,app}$ values. Error bars are from the time courses of the ATPase rate measurements at each RNA concentration. (B) IFN response of mock transfected (empty plasmid), wt RIG-I, and H830A RIG-I for 5'ppp dsRNA as well as their respective background signal is shown. (C) IFN response of wt RIG-I and H830A RIG-I for 5'ppp HP RNA and 5'ppp 2'-OMe HP RNA is shown.

Supplementary Figure S8



Supplementary Figure S8 (Related to Figure 4)**.** Mass Spectrometric Analysis (MALDI-TOF) of the 27 nucleotide single stranded RNAs carrying 5'ppp (A) and 5'm7Gppp (Cap-0) (B). The theoretical (parentheses) and measured molecular weights (peak) are presented.

Supplementary Table S1.

Sequences of the RNA ligands used in the study.

Sequence	RNA name	RNA end	Source
5’-pppGAAUAUAAUAGUGAUAUUAUAUUC-3’	5’ppp HP RNA	5’ppp	Biosynthesis
5’-(m7G)pppGAAUAUAAUAGUGAUAUUAUAUUC-3’	Cap-0 HP RNA	5'm7Gppp	TriLink
5’-GAAUAUAAUAGUGAUAUUAUAUUC-3’	5’OH HP RNA	5’OH	IDT
5'-pppAUACGUCCUGAUAGUUAGUAUCCAUCG-3'	5'ppp ssRNA	5’ppp	Biosynthesis
5'-AUACGUCCUGAUAGUUAGUAUCCAUCG-3'	5'OH ssRNA	5’OH	Biosynthesis
5'-(m7G)pppAUACGUCCUGAUAGUUAGUAUCCAUCG-3'	Cap-0 ssRNA	5'm7Gppp	Biosynthesis
5'-GCUCGAUGGAUACUAACUAUCAGGACGUAU-3'	complementary ssRNA	5’OH	Biosynthesis
5'-GCUAUACGUCCUGAUAGUUAGUAUCCAUCG-3'	5’ovg ssRNA	5’OH	Biosynthesis
5’-pppmGAAUAUAAUAGUGAUAUUAUAUUC-3’	5’ppp 2'O-Me HP RNA	5’ppp	Biosynthesis
5’-(m7G)pppmGAAUAUAAUAGUGAUAUUAUAUUC-3’	Cap-1 HP RNA	5'm7Gppp	TriLink
5'-(m7G)pppmAUACGUCCUGAUAGUUAGUAUCCAUCG-3'	Cap-1 ssRNA	5'm7Gppp	Biosynthesis
5’-pppCGUGAGACAU-3’Fl	5’ppp ss10 3’Fl RNA	5’ppp / 3’Fluorescein	Biosynthesis

Supplementary Table S2.

Data processing and refinement statistics calculated using PHENIX.

PDB ID	5F9H	5F9F	5F98
5' end	PPP	OH	m7Gppp
Location	BNL (X25)	CHESS (F1)	ALS (8.3.1)
Detector	Pilatus 6M	ADSC Q270	ADSC Q315
Wavelength (Å)	1.100	0.918	1.116
Resolution range (Å)	46.2 - 3.2 (3.4 – 3.2)	30.0 – 2.6 (2.7-2.6)	49.4 – 3.3 (3.4-3.3)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters	111.2, 174.8, 309.3Å	111.5, 174.3, 308.3Å	111.2, 174.2, 308.8Å
Total reflections	1,110,358	2,156,243	1,268,139
Unique reflections	99,666	166,166	90,827
Redundancy	11.1 (10.6)	13.0 (9.5)	14.0 (11.8)
Completeness (%)	99.5 (99.6)	90.6 (80.3)	99.8 (99.7)
Mean I/sigma(I)	9.6 (2.4)	13.0 (2.4)	7.1 (1.9)
R_{merge}	0.28 (1.11)	0.15(1.02)	0.56 (1.95)
R_{meas}	0.29 (1.16)	0.16(1.07)	0.60(2.12)
R_{pim}	0.09 (0.35)	0.04(0.33)	0.21(0.81)
R_{work}	0.22 (0.31)	0.20 (0.29)	0.22 (0.29)
R_{free}	0.27 (0.36)	0.26 (0.38)	0.28 (0.36)
Number of atoms			
All	33,603	34,918	32,719
Protein	30,480	31,642	29,503
RNA	3,108	3,036	3,204
Other	15	114	12
Water	0	126	0
RMS bonds (Å)	0.004	0.002	0.002
RMS angles (°)	0.74	0.58	0.54
Ramachandran favored (%)	97.1	95.3	93.1
Ramachandran allowed (%)	2.8	4.6	6.8
Ramachandran outliers (%)	0.1	0.1	0.1
Clashscore	5.9	2.7	4.6
Average B-factors (Å^2)			
All atoms	66.2	55.1	57.1
Protein atoms	65.5	54.9	56.4
RNA atoms	72.6	57.3	63.3
Other	49.6	78.5	41.8
Water atoms	NA	34.5	NA

(Data for highest resolution range shell are shown in parentheses.)