SUPPORTING INFORMATION

A structural rationale for the enhanced catalysis of nonenzymatic RNA primer extension by a downstream oligonucleotide

Wen Zhang[†,#], Chun Pong Tam[†,‡], Lijun Zhou[†,#], Seung Soo Oh[†,#,‖], Jiawei Wang[§] and Jack W. Szostak[†,#,‡,*]

[†]Howard Hughes Medical Institute and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA 02114, USA, [#]Department of Genetics, Harvard Medical School, Boston, MA 02114, USA, [‡] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA and [§]School of Life Sciences, Tsinghua University, Beijing 100084, China

Table of contents

1. LC-MS analysis.

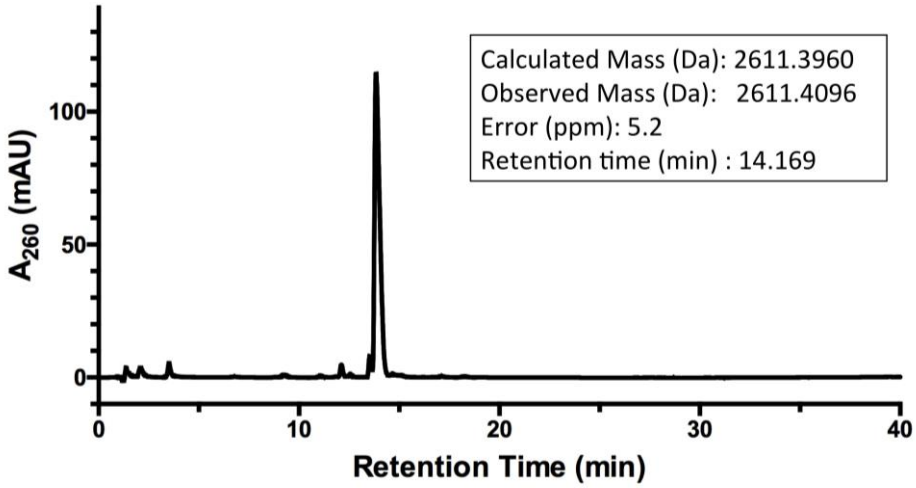

Figure S1. Reversed-phase HPLC analysis of transcript RNA 8mer 5'-ICG-CACCUCA-3' after purified by PAGE-gel. The HPLC analysis was performed on a Xbridge-C18 column (1.0 x 100 mm) with a linear gradient from buffer A (200 mM 1,1,1,3,3,3-hexafluoro-2-propanol with 1.25 mM triethylamine, pH 7.0) to 15% buffer B (Methanol) in 16 min; the RNA retention time was 14.169 min**.**

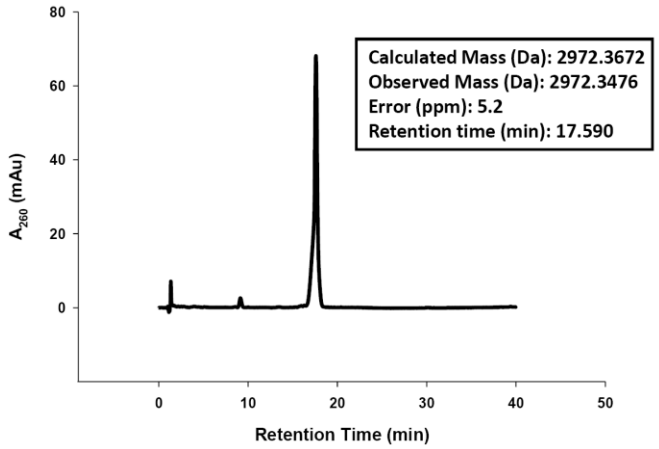

Figure S2. Reversed-phase HPLC analysis of transcript RNA 9mer 5'-GpppG-CACCUCA-3' after purified by PAGE-gel. The HPLC analysis was performed on a Xbridge-C18 column (1.0 x 100 mm) with a linear gradient from buffer A (200 mM 1,1,1,3,3,3-hexafluoro-2-propanol with 1.25 mM triethylamine, pH 7.0) to 20% buffer B (Methanol) in 20 min; the RNA retention time was 17.590 min**.**

2. X-ray crystallography.

2a. Crystallization. Nucleic Acid Mini Screen Kit, Natrix High Throughput Kit and Index High Throughput Kit (Hampton Research, Aliso Viejo, CA), and JBS Classic HTS II (Jena Biosciences, Jena, Germany) were used for screening crystallization conditions by the sitting drop vapor diffusion method. For RNA duplex-monomer crystallization, RNA samples (1 mM) were mixed with ICG (50 mM) at a 1:1 ratio. For RNA hairpin-monomer/RNA hairpin-GpppG crystallization, RNA template (1 mM), RNA downstream helper oligomer (1 mM) and ICG monomer/GpppG ligand (75 mM) were mixed at a 1:1:1 ratio. For crystallization of the RNA hairpin complex, the RNA template-loop-primer strand (0.5 mM) and the downstream GpppG-helper oligomer strand (0.5 mM) were mixed at a 1:1 ratio. All the components were heated to 80^{o}C for 2 min, and then cooled down

slowly to room temperature. All crystals grew at 18°C. For RNA duplex-monomer complexes, 35% MPD was used as cryoprotectant during crystal mounting. For all other complexes, the mother liquor containing 50% glycerol was used as cryoprotectant during crystal mounting. The optimized crystallization conditions are listed below.

Table S1. Optimized conditions for crystallization

	Optimized crystallization conditions
RNA 15mer-ICG (PDB: 5V0H)	50 mM $MgCl_2$, 0.05 M ammonium acetate, 0.05 M tris hydrochloride pH 7.5, 10% v/v (+/-)-2-methyl-2,4-pentanediol
RNA 14mer-ICG (PDB: 5V0J)	50 mM $MgCl_2$, 0.1 M potassium chloride, 0.05 M sodium chloride, 0.1 M HEPES pH 7.5, 15% v/v (+/-)-2-methyl-2,4-pentanediol
RNA H-34-IM (PDB: 5UX3)	50 mM $MgCl_2$, 3 M ammonia sulfate, 100 mM HEPES 7.5
RNA H-34-PO (PDB: 5V9Z)	50 mM $MgCl_2$, 3 M ammonia sulfate, 100 mM HEPES 7.5
RNA H-34-OH (PDB: 5V0O)	50 mM $MgCl_2$, 3 M ammonia sulfate, 100 mM HEPES 7.5
RNA H-35-OH (PDB: 5VCF)	50 mM $MgCl_2$, 0.2 M lithium sulfate monohydrate, 0.1 M BIS-TRIS pH 7.0, 25% w/v Polyethylene glycol 3,350
RNA H-35-PO (PDB: 5VCI)	50 mM $MgCl_2$, 0.2 M lithium sulfate monohydrate, 0.1 M BIS-TRIS pH 7.0, 25% w/v Polyethylene glycol 3,350
RNA H-43-GpppG (PDB: 6AZ4)	50 mM $MgCl_2$, 0.1 M Sodium cacodylate trihydrate pH 6.5, 1.4 M Sodium acetate trihydrate
RNA H-34-GpppG-oligo (PDB: 6BMD)	50 mM $MgCl_2$, 3 M ammonia sulfate, 100 mM HEPES 7.5
RNA H-34-L (PDB: 5UZ6)	50 mM $MgCl_2$, 3 M ammonia sulfate, 100 mM HEPES 7.5
RNA H-35-L (PDB: 5VGW)	50 mM $MgCl_2$, 0.2 M lithium sulfate monohydrate, 0.1 M BIS-TRIS pH 7.0, 25% w/v Polyethylene glycol 3,350

2b. Data collection and structure refinement. All data collection was taken under a stream of nitrogen at 99 K. The data sets were collected at the SIBYLS beamline 821 and 822 at the Advanced Light Source, Lawrence Berkeley National Laboratory. The distances between detector and the crystal were set to 200-300 mm and the collecting wavelength was set to 1 Å. The crystals were exposed for 1 second per image with one degree oscillation, and 180 images were taken for each data set. The data was processed using HKL2000.[1] All the structures were solved by molecular replacement.[2] The RNA models were from our previously reported structures, or based on RNA hairpin structure in the literature. All the structures were refined using Refmac.[3] The refinement protocol included simulated annealing, refinement, restrained B-factor refinement, and bulk solvent correction. During refinement, the topologies and parameters for locked nucleic acids and for the ligand ICG and GpppG (LCC, LCA, LCG, rIMG, GP3) were constructed and applied. After several cycles of refinement, the water and magnesium atoms were added. Data collection, phasing, and refinement statistics of the determined structures are listed in Tables S2 and S3.

Table S2. Data collection statistics.

Structure	RNA 15mer-ICG	RNA 14mer-ICG	H-34-Im complex	H-34-PO complex
PDB code	5V0H	5V0J	5UX3	5V9Z
Space group	P321	R32	$P4_12_12$	$P4_32_12$
Unit cell parameters (Å, º)	42.85, 42.85, 83.84 90, 90, 120	43.39, 43.39, 253.85 90, 90, 120	46.53, 46.53, 156.94 90, 90, 90	45.28, 45.28, 154.57 90, 90, 90
Resolution range, Å (last shell)	50-1.90 (1.97-1.90)	100-1.50 (1.53-1.50)	50-2.50 (2.59-2.50)	50-2.50 (2.59-2.50)
Unique reflections	7272 (702)	15290 (744)	6436 (604)	5769 (418)
Completeness, %	97.4 (95.4)	99.4 (99.9)	98.7 (94.1)	95.4 (73.1)
$\boldsymbol{R_{merge}}$**, %**	13.7 (46.8)	4.1 (57.5)	8.0 (69.5)	10.0 (71.0)
<I/σ(I)>	10 (2.9)	35.4 (1.8)	20.2 (1.0)	16.3 (1.7)
Redundancy	7.3 (4.2)	6.3 (5.6)	5.1 (3.9)	10.6 (7.2)

Structure	H-34-OH complex	H-35-OH complex	H-35-PO complex	H-43-GpppG complex
PDB code	5V0O	5VCF	5VCI	6AZ4
Space group	$P4_32_12$	R3	R3	$P4_12_12$
Unit cell parameters (Å, º)	44.75, 44.75, 154.20 90, 90, 90	71.07, 71.07, 71.27 90, 90, 120	69.61, 69.61, 70.85 90, 90, 120	79.39, 79.39, 73.65 90, 90, 90
Resolution range, Å (last shell)	50-2.70 (2.80-2.70)	50-2.80 (2.90-2.80)	50-2.60 (2.69-2.60)	50-2.98 (3.09-2.98)
Unique reflections	4657 (437)	3296 (316)	3864 (338)	4713 (476)
Completeness, %	97.3 (94.4)	99.7 (100)	97.8 (87.8)	91.6 (95.2)
$\boldsymbol{R_{merge}}$**, %**	10.0 (77.7)	7.2 (49.5)	5.4 (52.6)	13.3 (69.8)
<I/σ(I)>	18.2 (1.6)	17.4 (1.5)	18.4 (1.8)	9.2 (1.6)
Redundancy	9.1 (6.4)	3.3 (3.3)	3.2 (2.4)	4.5 (4.4)

Structure	H-34-GpppG-oligo complex	H-34-L complex	H-35-L complex
PDB code	6BMD	5UZ6	5VGW
Space group	C2	C2	R3
Unit cell parameters (Å, º)	107.39, 63.76, 65.84 90, 122, 90	106.69, 63.07, 65.92 90, 120, 90	71.06, 71.06, 69.78 90, 90, 120
Resolution range, Å (last shell)	50-3.00 (3.11-3.00)	50-2.10 (2.18-2.10)	50-2.40 (2.49-2.40)
Unique reflections	7315 (557)	20855 (1912)	4799 (508)
Completeness, %	95.9 (74.5)	96.9 (89.0)	96.1 (98.8)
$\boldsymbol{R_{merge}}$**, %**	19.0 (37.7)	5.2 (32.3)	7.3 (63.9)
<I/σ(I)>	5.8 (2.0)	15.5 (1.8)	15.0 (2.3)
Redundancy	4.8 (3.2)	2.7 (2.0)	3.7 (3.1)

Table S3. Structure refinement statistics.

Structure	RNA 15mer-ICG	RNA 14mer-ICG	H-34-Im complex	H-34-PO complex
PDB code	5V0H	5V0J	5UX3	5V9Z
Molecules per asymmetric unit	1	1	1	1
Resolution range, Å (last shell)	37.11-1.90	84.62-1.50	44.61-2.50	43.45-2.51
R_{work}**, %**	22.7	23.7	21.8	21.8
R_{free}**, %**	27.2	27.0	26.1	28.1
Number of reflections	6924	14527	6154	5466
Bond length R.M.S., Å	0.026	0.027	0.015	0.013
Bond angle R.M.S.	2.686	2.963	2.564	2.438
Average B-factors, Å^2	43.86	36.47	77.74	73.17

Structure	H-34-OH complex	H-35-OH complex	H-35-PO complex	H-43-GpppG complex
PDB code	5V0O	5VCF	5VCI	6AZ4
Molecules per asymmetric unit	1	1	1	1
Resolution range, Å (last shell)	42.97-2.70	46.58-2.80	45.91-2.60	50.00-2.98
R_{work}**, %**	21.8	16.3	16.2	18.1
R_{free}**, %**	27.7	21.3	21.2	20.7
Number of reflections	4446	3154	3649	4464
Bond length R.M.S., Å	0.012	0.009	0.010	0.009
Bond angle R.M.S.	1.781	1.890	1.787	2.185
Average B-factors, Å^2	86.79	64.45	75.43	73.20

Structure	H-34-GpppG-oligo complex	H-34-L complex	H-35-L complex
PDB code	6BMD	5UZ6	5VGW
Molecules per asymmetric unit	3	3	1
Resolution range, Å (last shell)	50.00-3.00	50.00-2.10	46.15-2.42
R_{work}**, %**	18.3	22.1	20.73
R_{free}**, %**	23.0	26.7	21.29
Number of reflections	6922	19827	4546
Bond length R.M.S., Å	0.012	0.014	0.008
Bond angle R.M.S.	2.060	2.452	1.755
Average B-factors, Å^2	26.41	47.84	68.15

3. References.

(1) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307-326.

(2) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.

(3) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr. D* **1997**, *53*, 240-255.