Structural insights into the interaction of clinically relevant phosphorothioate mRNA cap analogs with translation initiation factor 4E reveal stabilization via electrostatic thio-effect

Marcin Warminski[†], Joanna Kowalska[†,*], Elzbieta Nowak[§], Dorota Kubacka[†], Ryan Tibble[#], Renata Kasprzyk[‡], Pawel J. Sikorski[‡], John D. Gross[#], Marcin Nowotny[§], Jacek Jemielity[‡,*]

[†] Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland; [§] Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, Ksiecia Trojdena 4, 02-109 Warsaw, Poland; [#] Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94158, USA; [‡] Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland;

Supporting Information File

Supplementary tables and figures

Table S1. Data collection and refinement statistics.
Statistics for the highest-resolution shell are shown in parentheses.

	5M7V $m_2^{7,2'O}GppSpG$ D1	**5M7W** $m_2^{7,2'O}GppSpG$ D2	**5M7X*** $m_2^{7,2'O}GppSepG$ D1	**5M7Z*** $m_2^{7,2'O}GppSepG$ D2
Wavelength (Å)	0.9184	0.9184	0.9796	0.9796
Resolution range	36.83 – 1.74 (1.80 – 1.74)	37.00 – 1.97 (2.04 – 1.97)	48.81 – 1.68 (1.74 – 1.68)	36.93 – 1.69 (1.75 – 1.69)
Space group	P 1	P 1	P 1	P 1
Unit cell	38.0 38.1 146.9 88.6 95.4 103.6	38.0 38.0 147.3 96.0 88.3 103.5	38.1 38.2 147.1 88.6 95.3 103.3	38.1 38.1 147.2 88.3 84.1 76.6
Total reflections	137992 (12919)	96609 (8824)	260800 (24427)	337866 (31217)
Unique reflections	77683 (7648)	54200 (5225)	89491 (8718)	88123 (8697)
Multiplicity	1.8 (1.7)	1.8 (1.7)	2.9 (2.8)	3.8 (3.8)
Completeness	0.95 (0.93)	0.96 (0.93)	0.97 (0.94)	0.98 (0.92)
Mean I/sigma(I)	9.7 (1.8)	8.2 (1.8)	14.9 (2.7)	12.0 (2.0)
Wilson B-factor	17.79	24.71	19.65	21.37
R-merge	0.054 (0.394)	0.064 (0.378)	0.049 (0.375)	0.087 (0.611)
R-meas	0.077 (0.557)	0.090 (0.535)	0.060 (0.460)	0.100 (0.709)
CC1/2	0.997 (0.64)	0.996 (0.726)	0.997 (0.804)	0.994 (0.649)
CC*	0.999 (0.884)	0.999 (0.917)	0.999 (0.944)	0.998 (0.887)
Reflections used in refinement	77677 (7648)	54194 (5225)	89485 (8718)	88361 (8695)
Reflections used for R-free	2100 (207)	2100 (202)	2106 (205)	2080 (205)
R-work	0.236 (0.292)	0.213 (0.271)	0.192 (0.270)	0.219 (0.275)
R-free	0.288 (0.351)	0.262 (0.297)	0.219 (0.300)	0.255 (0.282)
CC(work)	0.906 (0.740)	0.930 (0.812)	0.950 (0.867)	0.931 (0.778)
CC(free)	0.833 (0.676)	0.869 (0.804)	0.914 (0.818)	0.910 (0.683)
Number of non-hydrogen atoms	5872	5897	6102	5730
macromolecules	5347	5419	5477	5181
ligands	108	148	148	115
Protein residues	685	695	695	658
RMS(bonds)	0.007	0.007	0.006	0.006
RMS(angles)	0.85	0.90	0.88	0.82
Ramachandran favored (%)	97	97	98	98
Ramachandran allowed (%)	2.5	3.1	2.3	2.5
Ramachandran outliers (%)	0	0	0	0
Rotamer outliers (%)	1.3	2.1	2	2
Clashscore	4.44	5.70	2.99	2.38
Average B-factor	31.5	40.8	34.7	36.3
macromolecules	31.4	40.5	34.0	35.8
ligands	37.4	63.0	51.6	51.6
solvent	32.2	36.8	37.4	38.6
Number of TLS groups	32	28	25	26

* Friedel mates were averaged when calculating reflection statistics.

Table S1 cont. Data collection and refinement statistics.
Statistics for the highest-resolution shell are shown in parentheses.

	5M83 $m_2^{7,2'O}GppSpA$ D1	**5M84** $m_2^{7,2'O}GppSpA$ D2	**5M80** iPr-$m^7GppSpG$ D1	**5M81** iPr-$m^7GppSpG$ D2
Wavelength (Å)	0.9184	0.9184	0.9184	0.9184
Resolution range	36.83 – 1.86 (1.92 – 1.86)	37.16 – 1.85 (1.91 – 1.85)	37.29 – 2.12 (2.19 – 2.12)	48.83 – 1.90 (1.97 – 1.90)
Space group	C 1 2 1	C 1 2 1	P 1	P 1
Unit cell	47.1 59.7 149.2 90 96.3 90	46.9 59.8 149.4 90 95.8 90	38.3 38.3 146.8 87.5 95.4 102.9	38.1 38.1 147.2 88.8 95.8 103.4
Total reflections	223624 (22068)	164916 (16551)	85264 (6939)	181184 (18209)
Unique reflections	34746 (3422)	35356 (3509)	43093 (3914)	62073 (6137)
Multiplicity	6.4 (6.4)	4.7 (4.7)	2.0 (1.8)	2.9 (3.0)
Completeness	1.00 (0.98)	1.00 (0.99)	0.94 (0.87)	0.98 (0.98)
Mean I/sigma(I)	16.6 (2.2)	14.2 (1.4)	10.5 (1.8)	11.9 (2.1)
Wilson B-factor	28.59	32.96	26.42	20.88
R-merge	0.070 (0.848)	0.063 (0.926)	0.068 (0.450)	0.078 (0.557)
R-meas	0.076 (0.922)	0.071 (1.043)	0.093 (0.620)	0.095 (0.677)
CC1/2	0.999 (0.814)	0.999 (0.548)	0.996 (0.662)	0.997 (0.739)
CC*	1 (0.947)	1 (0.841)	0.999 (0.893)	0.999 (0.922)
Reflections used in refinement	34739 (3417)	35350 (3507)	43088 (3914)	62065 (6135)
Reflections used for R-free	817 (79)	1768 (175)	2100 (190)	1462 (145)
R-work	0.206 (0.311)	0.196 (0.363)	0.204 (0.268)	0.178 (0.251)
R-free	0.251 (0.340)	0.228 (0.416)	0.259 (0.308)	0.219 (0.310)
CC(work)	0.948 (0.830)	0.953 (0.723)	0.936 (0.793)	0.960 (0.860)
CC(free)	0.915 (0.883)	0.925 (0.463)	0.898 (0.736)	0.949 (0.764)
Number of non-hydrogen atoms	2968	2984	5690	6534
macromolecules	2743	2771	5192	5796
ligands	74	80	177	238
Protein residues	352	357	683	728
RMS(bonds)	0.007	0.008	0.008	0.007
RMS(angles)	0.82	0.94	0.97	0.83
Ramachandran favored (%)	98	98	96	97
Ramachandran allowed (%)	2.3	1.7	4	2.8
Ramachandran outliers (%)	0	0	0	0
Rotamer outliers (%)	1.5	1.1	1	1
Clashscore	3.02	4.5	5.5	3.67
Average B-factor	51.0	54.6	40.9	36.8
macromolecules	51.0	54.4	41.1	36.0
ligands	67.6	75.4	45.8	62.1
solvent	43.5	47.8	34.9	35.1
Number of TLS groups	15	11	28	37

Table S1 cont. Data collection and refinement statistics.
Statistics for the highest-resolution shell are shown in parentheses.

	6GKJ $m^7GpppApG$	**6GKK** m^7Gpp_SpApG D1	**6GKL** m^7Gpp_SpApG D2
Wavelength (Å)	0.9184	0.9184	0.9184
Resolution range	24.07 – 2.07 (2.14 – 2.07)	22.07 – 1.86 (1.92 – 1.86)	48.79 – 2.20 (2.28 – 2.20)
Space group	P 1	P 1	P 1
Unit cell	38.2 38.2 147.0 84.2 88.6 77.0	38.1 38.2 146.9 88.9 84.6 76.6	38.2 38.2 147.0 88.8 84.7 76.1
Total reflections	200869 (13093)	453130 (31689)	136335 (6010)
Unique reflections	46260 (4430)	66448 (6360)	39200 (3148)
Multiplicity	4.3 (3.0)	6.8 (5.0)	3.5 (1.9)
Completeness	0.94 (0.90)	0.98 (0.93)	0.96 (0.76)
Mean I/sigma(I)	7.1 (1.7)	13.8 (1.4)	6.4 (1.9)
Wilson B-factor	27.78	25.54	33.37
R-merge	0.194 (0.566)	0.100 (1.026)	0.124 (0.398)
R-meas	0.214 (0.692)	0.108 (1.141)	0.145 (0.533)
CC1/2	0.987 (0.638)	0.999 (0.645)	0.991 (0.695)
CC*	0.997 (0.883)	1 (0.885)	0.998 (0.906)
Reflections used in refinement	46228 (4430)	66423 (6356)	39196 (3148)
Reflections used for R-free	1706 (164)	2449 (235)	1445 (116)
R-work	0.200 (0.278)	0.195 (0.297)	0.218 (0.276)
R-free	0.249 (0.367)	0.234 (0.322)	0.257 (0.304)
CC(work)	0.943 (0.773)	0.953 (0.807)	0.930 (0.758)
CC(free)	0.925 (0.427)	0.932 (0.744)	0.907 (0.723)
Number of non-hydrogen atoms	6047	6191	5908
macromolecules	5575	5682	5543
ligands	144	150	175
Protein residues	721	723	727
RMS(bonds)	0.006	0.010	0.006
RMS(angles)	0.84	1.00	0.79
Ramachandran favored (%)	98	98	97
Ramachandran allowed (%)	1.7	2.1	2.9
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	0.92	0.89	0.77
Clashscore	2.68	4.70	4.03
Average B-factor	42.5	43.5	51.3
macromolecules	42.6	43.3	51.1
ligands	52.6	57.0	70.7
solvent	36.1	40.2	40.0
Number of TLS groups	21	25	38

Table S2. Relative affinities of selected di- and trinucleotide mRNA cap analogs and capped RNAs for murine eIF4E determined using various biophysical methods (microscale thermophoresis – MST; fluorescence quenching titration – FQT; isothermal titration calorimetry – ITC).

Compound	P_β config.	$K_{d,app}$ MST /μM [C.I.]	n	K_D FQT /nM	n	K_D ITC /nM [C.I.]	n
m^7GpppG	-	n.d		**106** ± 5[a]		**112** [87, 141]	3
$m_2^{7,2'\text{-}O}GpppG$	-	n.d		**92.6** ± 2.6[a]		n.d	
$m_2^{7,2'\text{-}O}Gpp_SpG$ D1	R_P	**3.2** [2.7, 3.8]	3	**23.2** ± 0.8[a]		**29.2** [22.1, 36.5]	3
$m_2^{7,2'\text{-}O}Gpp_SpG$ D2	S_P	**5.2** [4.5, 6.1]	3	**51.8** ± 5.9[a]		**55.2** [40.2, 74.6]	3
$m_2^{7,2'\text{-}O}Gpp_{Se}pG$ D1	R_P	n.d.		**26.0** ± 0.5[b]		n.d	
$m_2^{7,2'\text{-}O}Gpp_{Se}pG$ D2	S_P	n.d.		**52.6** ± 0.5[b]		n.d	
iPr-m^7GpppG	-	n.d		**345** ± 12[c]		n.d	
iPr-m^7Gpp_SpG D1	R_P	n.d.		**179** ± 3[c]		n.d	
iPr-m^7Gpp_SpG D2	S_P	n.d.		**108** ± 2[c]		n.d	
m^7GpppA	-	**22** [18, 28]	3	**178** ± 3 **213** ± 24[d]	3	n.d	
$m_2^{7,2'\text{-}O}Gpp_SpA$ D1	R_P	**4.9** [4.1, 5.9]	3	**44.0** ± 1.4	3	**49.0** [37.2, 63.3]	1
$m_2^{7,2'\text{-}O}Gpp_SpA$ D2	S_P	**8.1** [6.7, 9.9]	3	**71.9** ± 2.1	3	**69.0** [47.8, 98.0]	1
$m^7GpppApG$	-	**5.4** [4.6, 6.3]	5	**26.6** ± 0.9	3	n.d	
m^7Gpp_SpApG D1	R_P	**1.7** [1.3, 2.0]	4	**7.3** ± 0.4	3	n.d	
m^7Gpp_SpApG D2	S_P	**1.9** [1.7, 2.2]	4	**9.5** ± 0.4	3	n.d	
$m_2^{7,2'\text{-}O}GpppG\text{-}RNA_{35}$	-	**0.14** [0.11, 0.17]	3	n.d.		n.d.	
$m_2^{7,2'\text{-}O}Gpp_SpG$(D1)-$RNA_{35}$	R_P	**0.065** [0.055, 0.085]	3	n.d.		n.d.	
$pppG\text{-}RNA_{35}$	-	no binding		n.d.		n.d.	

n.d. – not determined; C.I. – 68.3 % confidence interval; [a] Kowalska et al. 2008; [b] Kowalska et al. 2009; [c] Warminski et al. 2013; [d] Niedzwiecka et al. 2002;

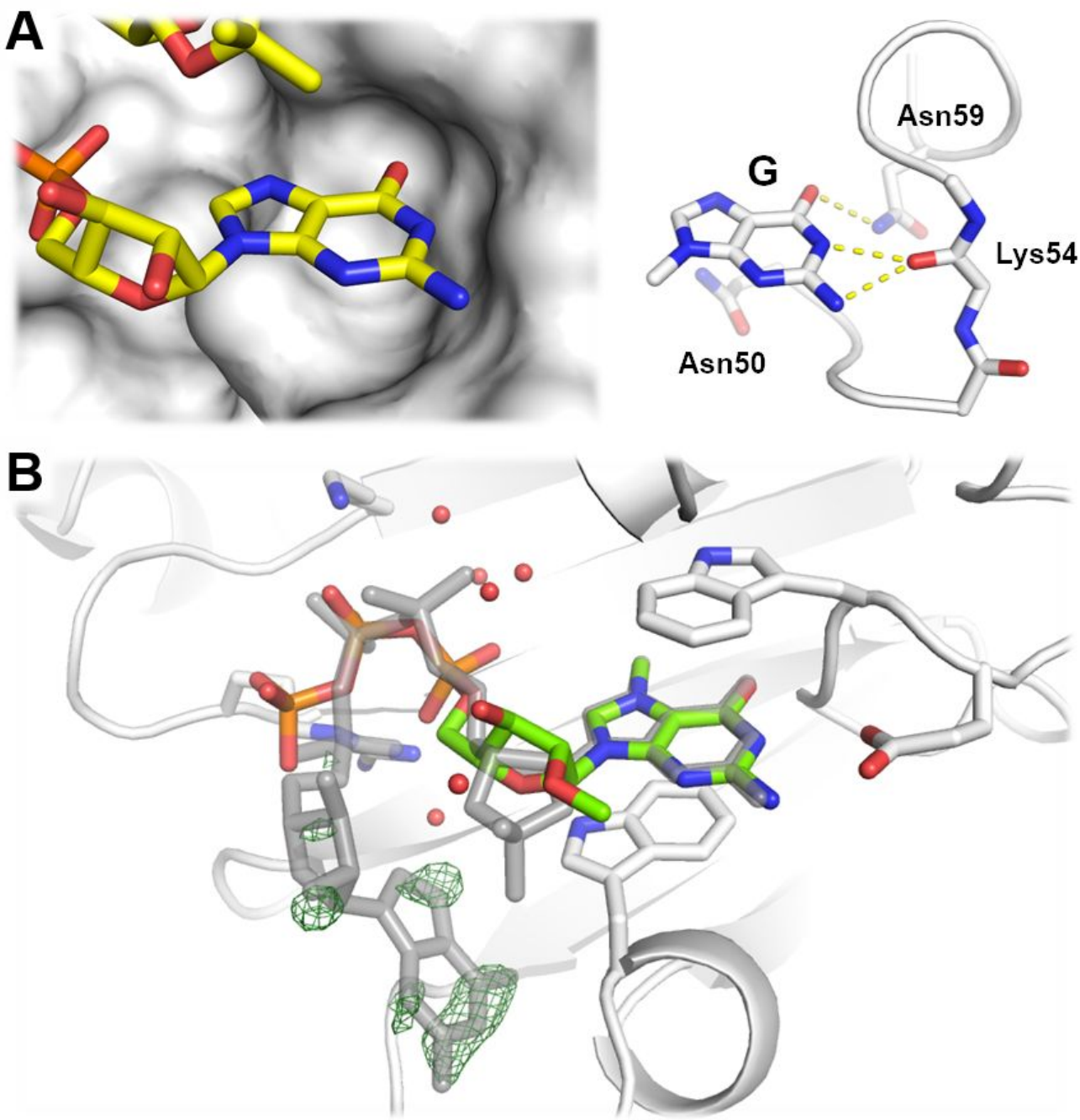


Figure S1. A new binding site of eIF4E revealed by X-Ray structure of eIF4E/iPr-m^{7}Gpp$_S$pG D2-(S_P) complex: A) guanine binding pocket formed by N-terminal loop (residues 50-55) of eIF4E; **B)** superposition of eIF4E/iPr-m^{7}Gpp$_S$pG D2-(S_P) (gray semitransparent sticks) and eIF4E/m$_2^{7,2'\text{-O}}$Gpp$_S$pG D2-(S_P) models with residual electron density map (green mesh contoured at 3σ) calculated from eIF4E/m$_2^{7,2'\text{-O}}$Gpp$_S$pG D2-(S_P) diffraction data.

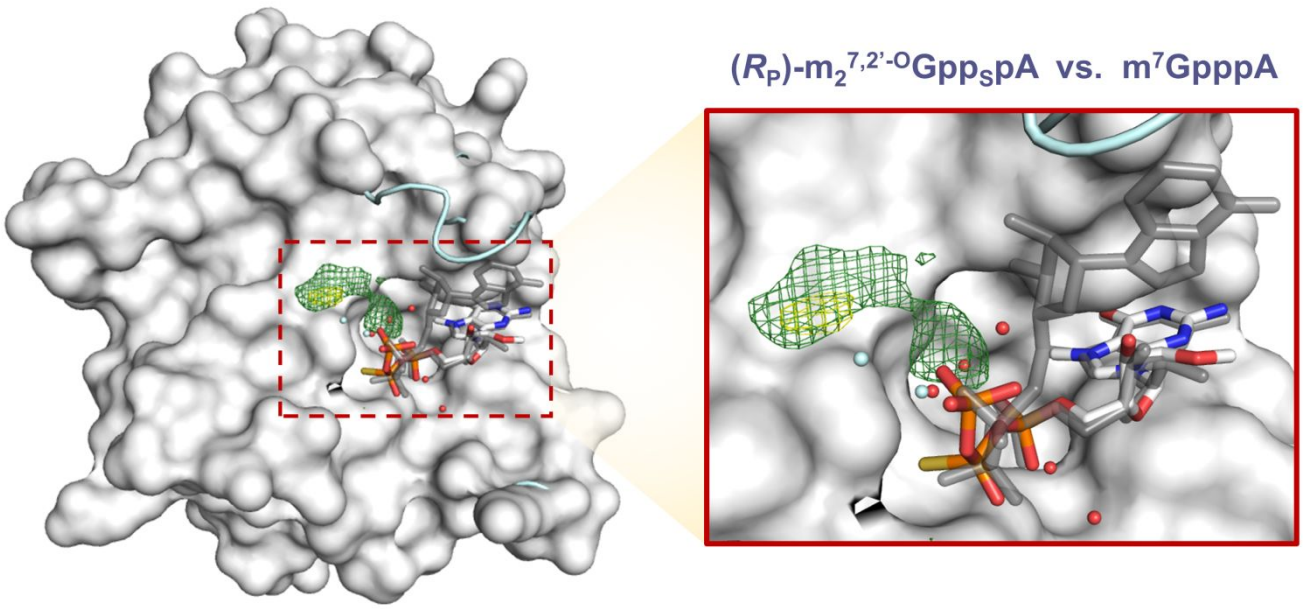


Figure S2. X-Ray structure of eIF4E/$m_2^{7,2'\text{-}O}Gpp_SpA$ D1-(R_P) complex with residual electron density map (green mesh contoured at 3σ) and simulated annealing omit map (F_o-F_c; yellow mesh contoured at 3σ) overlaid with eIF4E/m^7GpppA model (gray sticks; PDB id: 1WKW).

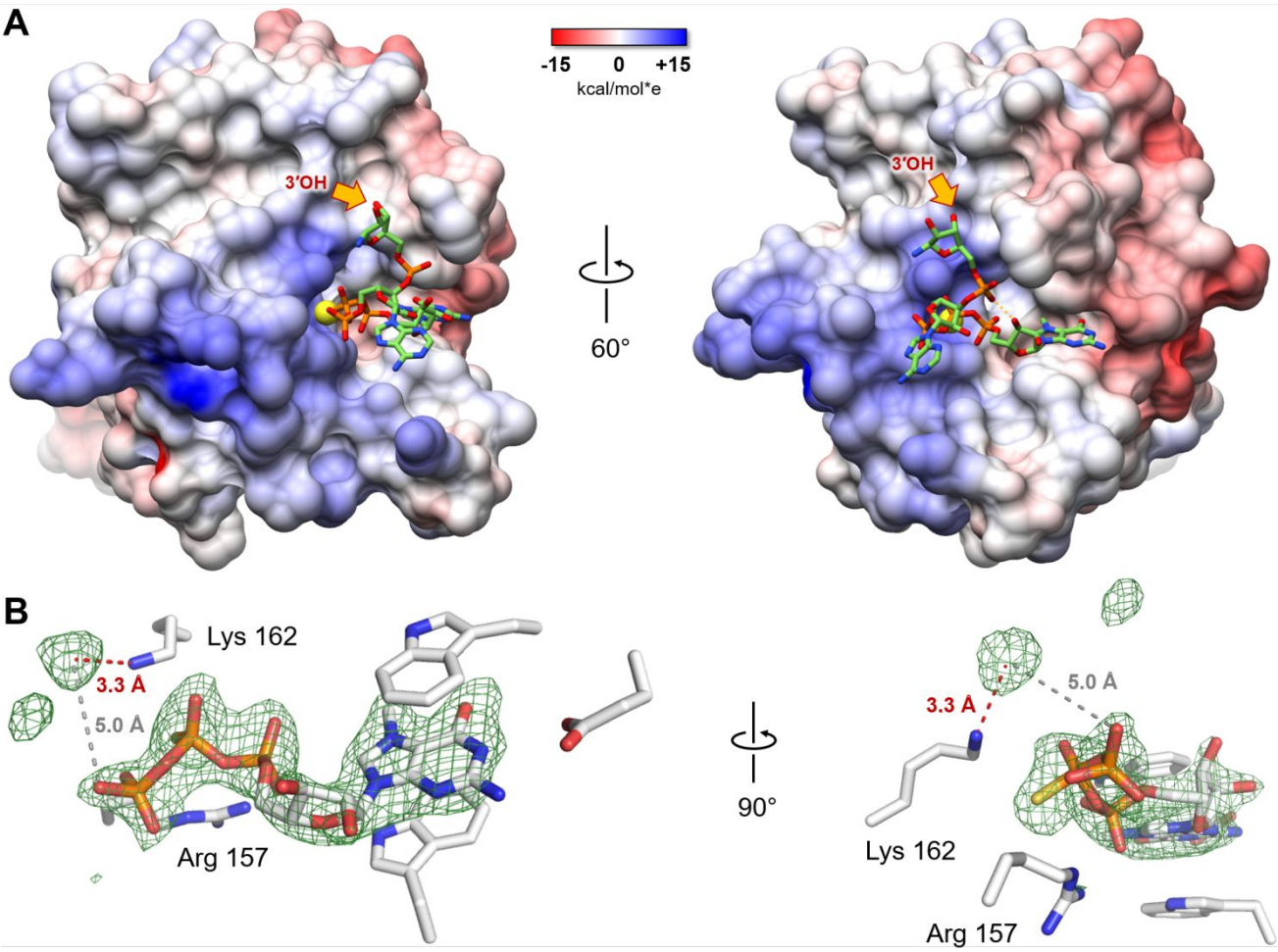


Figure S3. X-Ray structure of eIF4E/m^7Gpp_SpApG D2-(S_P) complex: A) In chain A molecule almost entire portion of the ligand was modeled; protein surface was colored according to coulombic potential calculated using UCSF Chimera; 3'-OH of the third nucleotide was indicated by orange arrow; **B)** In chain B copy only m^7Gppp portion of the ligand was modeled based on the simulated annealing omit map (F_o-F_c; green mesh contoured at 3σ); the distance between Lys162 –NH_3^+ group or 5'-*O* atom of adenosine and center of the density peak that might correspond to phosphodiester linkage of the ligand were denoted in red or grey respectively.

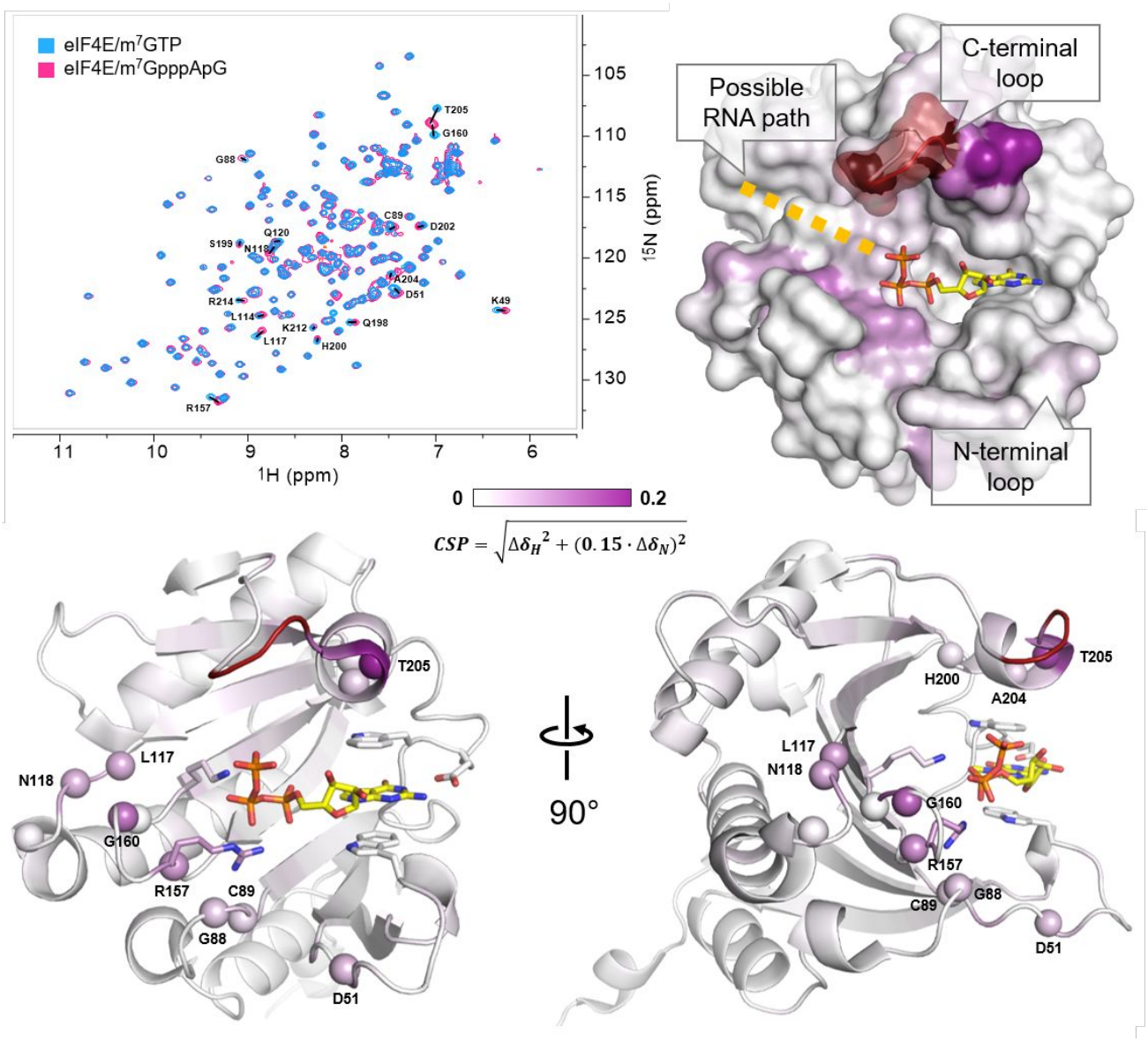


Figure S4. Chemical shift perturbations (CSP) observed upon for eIF4E complexes upon m^{7}GpppApG to m^{7}GTP ligand exchange color coded onto surface of eIF4E (PDB id: 1L8B). CSPs between ^{15}N HSQC spectra of eIF4E/m^{7}GpppApG (pink contours) and eIF4E/m^{7}GTP (blue contours) were calculated as sqrt[$\Delta\delta_H^2$ + (0.2*$\Delta\delta_N$)^2]. Fragment of C-terminal loop that was not assigned in the spectra was colored red. Possible RNA path deduced from CSPs was indicated by a yellow dotted line.

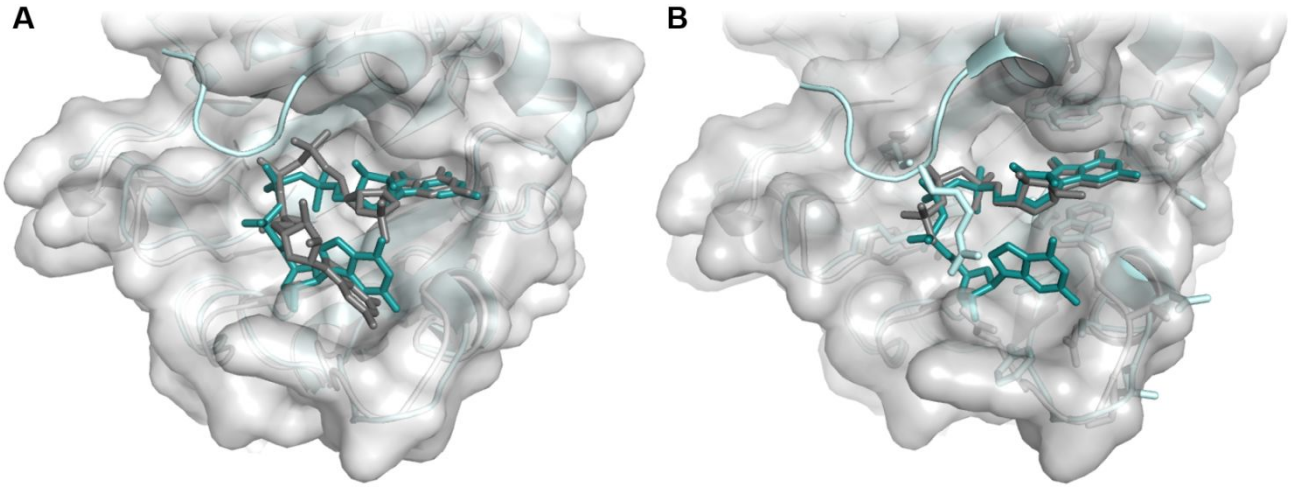


Figure S5. Alignment of murine (*Mus Musculus*) eIF4E in complex with (S_P)-iPr-m^{7}Gpp$_S$pG (panel A) or m$_2^{7,2'\text{-O}}$Gpp$_S$pG D2-(S_P) (panel B) – both colored grey – with eIF4E from *Schistosoma mansoni* in complex with m^{7}GpppG – colored green.

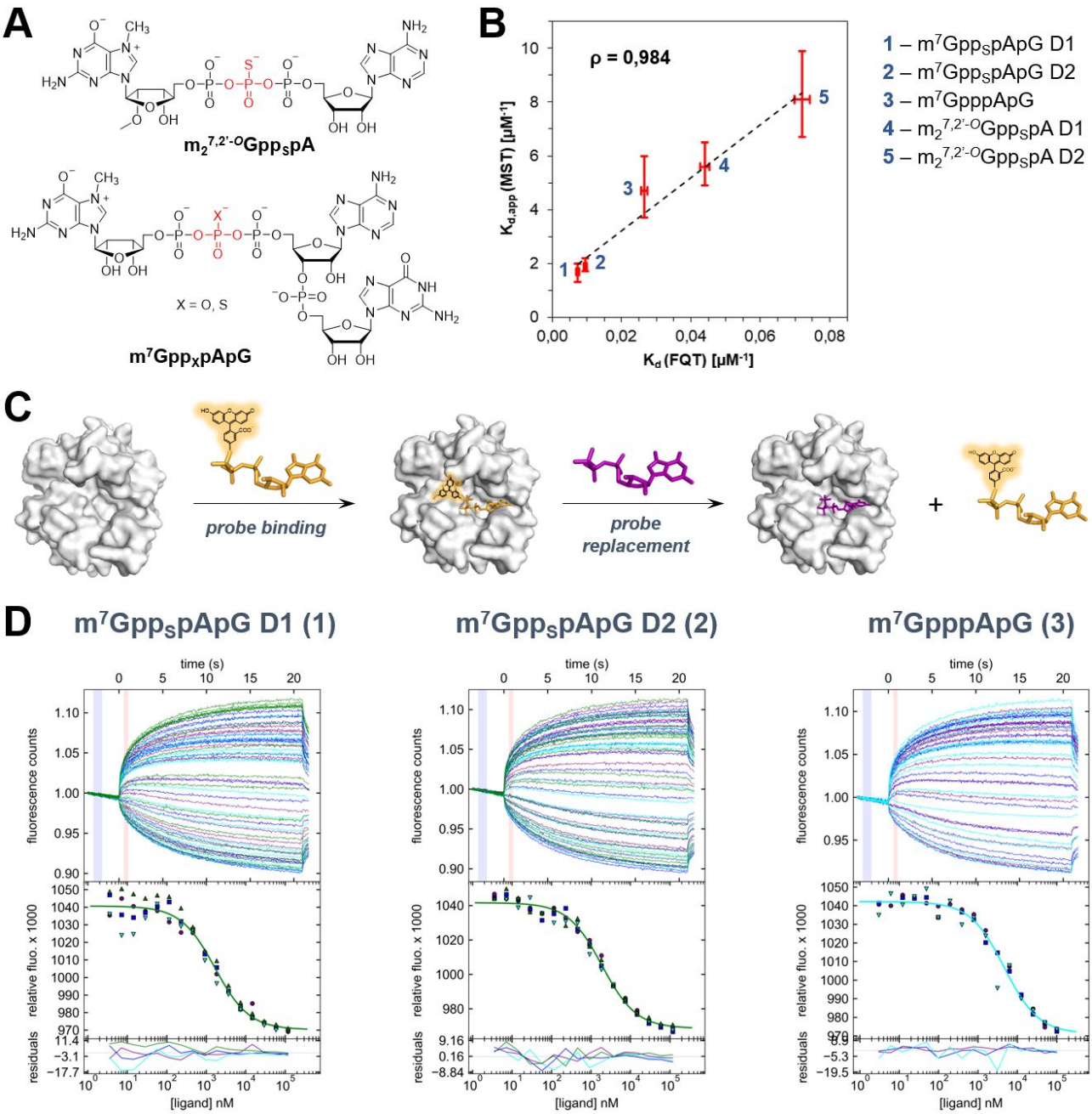


Figure S6. Determination of eIF4E-cap and eIF4E-RNA dissociation constants by microscale thermophoresis (MST) competition assay. A) Chemical structure of di- and trinucleotide cap analogs tested in the assay; **B)** Correlation between apparent K_d values derived from MST competition assay and K_d values obtained from direct binding experiment by FQT (fluorescence quenching titration); **C)** Schematic representation of the competition assay experiment using fluorescent cap analog (orange sticks) as a probe and different cap analogs (purple sticks) as titrant; **D)** Representative MST curves and competitive binding curves obtained in the experiment.

Experimental procedures

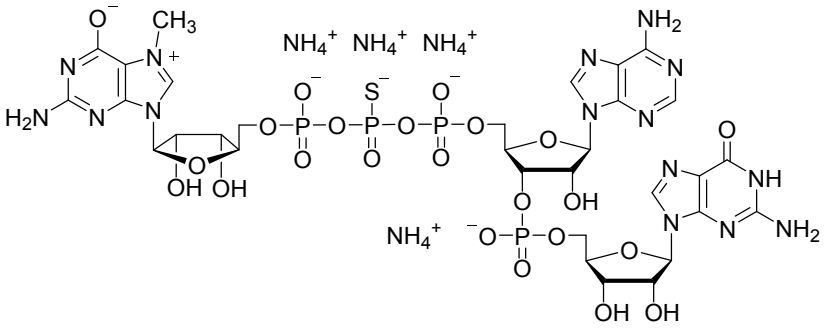
Chemical syntheses

1) Dinucleotide adenosine cap analogs

$m_2^{7,2'\text{-}O}Gpp_SpG$, $m_2^{7,2'\text{-}O}Gpp_{Se}pG$ and iPr-m^7Gpp_SpG were synthesized as reported earlier.[1-3] Diastereomers of $m_2^{7,2'\text{-}O}Gpp_SpA$ were obtained according to the protocol described earlier for $m_2^{7,2'\text{-}O}Gpp_SpG$ analog,[1] using AMP-Im[4] instead of GMP-Im in the last step of the synthesis.

2) Trinucleotide cap analogs

$m^7GpppApG$ was synthesized as described earlier.[5] Synthesis of β-phosphorothioate analog was achieved by activation of pApG into *P*-imidazolide and then coupling with m^7GDP-β-S.[1]

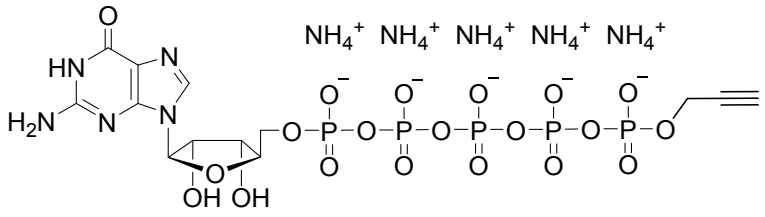


Triethylammonium salt of pApG was synthesized as described earlier.[5] Imidazole (20.6 mg, 303 µmol), 2,2'-dithiodipyridine (24.8 mg, 55.8 µmol), triphenylphosphine (28.3 mg, 108 µmol) and triethylamine (5.2 µL, 37.2 µmol) were added to the solution of pApG (252 OD_{260nm}, 9.30 µmol) in DMF (186 µL) and the solution was stirred at rt for 5 days. The product was precipitated by addition of $NaClO_4$ (11.4 mg) in acetonitrile (1.86 mL), washed with acetonitrile and dried under reduced pressure. Without additional purification, Im-pApG was re-dissolved in DMF (186 µL) followed by addition of m^7GDP-β-S (106 OD_{260nm}, 9.3 µmol) and $ZnCl_2$ (10.1 mg, 74.4 µmol). Reaction was quenched after 1 h by addition of EDTA (27.7 mg, 74.4 µmol) and $NaHCO_3$ (13.8 mg, 165 µmol) in water (1.4 mL). The product was isolated by ion-exchange chromatography on DEAE Sephadex using linear gradient 0-1.2 M TEAB and the diastereomers were separated using semi-preparative RP HPLC on Discovery RP Amide C-16 HPLC column (25 cm x 21.2 mm, 5 µm, flow rate 5.0 mL/min) with linear gradient of acetonitrile in 0.05 M ammonium acetate buffer (pH 5.9). After evaporation and repeated freeze-drying, ammonium salts of m^7Gpp_SpApG D1 (0.99 mg, 26.0 OD_{260nm}, 0.81 µmol) and m^7Gpp_SpApG D2 (1.42 mg, 37.4 OD_{260nm}, 1.17 µmol) were obtained in total yield 43%.

D1: RP-HPLC: R_t = 8.705 min; HRMS ESI (-) calcd. *m/z* for $C_{31}H_{40}N_{15}O_{23}P_4S^-$ $[M\text{-}H]^-$: 1146.10981, found 1146.11096;
D2: RP-HPLC: R_t = 9.045 min; HRMS ESI (-) calcd. *m/z* for $C_{31}H_{40}N_{15}O_{23}P_4S^-$ $[M\text{-}H]^-$: 1146.10981, found 1146.11147;

3) MST probe

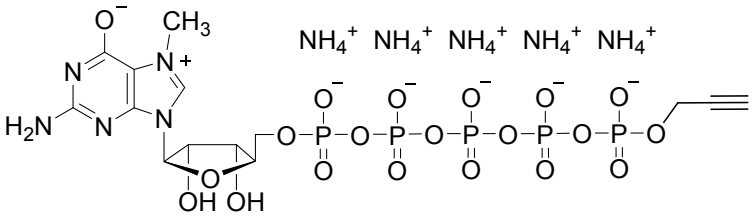
$Gp_5OC_3H_3$



A triethylammonium salt of guanosine tetraphosphate Gp_4 (40 mg, 0.040 mmol), 2-propargyl monophosphate *P*-imidazolide $PO_3C_3H_3$-Im (23 mg, 0.120 mmol) and $ZnCl_2$ (65 mg, 0.477 mmol) were suspended in DMSO (0.75 mL) and the mixture was stirred at a room temperature for two days. The reaction was quenched by addition of Na_2EDTA (180 mg, 0.484 mmol) in H_2O (7 mL), the pH was adjusted to 7 by addition of $NaHCO_3$ and the product was purified by semi-preparative RP HPLC (isocratic elution with 0.05 M ammonium acetate buffer pH 5.9) on Discovery RP Amide C-16 HPLC column (25 cm x 21.2 mm, 5 µm; flow rate 5.0 mL/min). After evaporation and repeated freeze-drying, an ammonium salt of ε-O-(2-propargyl) guanosine pentaphosphate $Gp_5OC_3H_3$ (25.7 mg, 0.032 mmol) was obtained as a white solid in 80% yield.

RP-HPLC: R_t = 2.339 min; HRMS ESI (-) calcd. *m/z* for $C_{13}H_{19}N_5O_{20}P_5^-$ $[M\text{-}H]^-$: 719.93169, found: 719.93288;

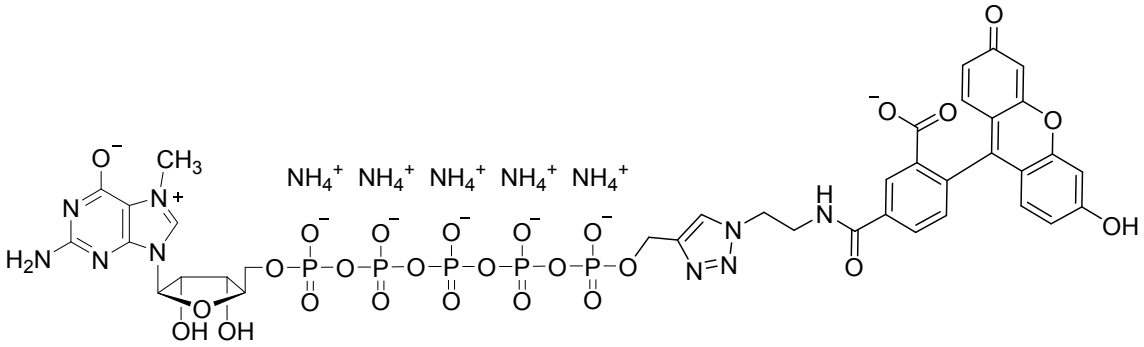
$m^7Gp_5OC_3H_3$



An ammonium salt of ε-O-(2-propargyl) guanosine pentaphosphate $Gp_5OC_3H_3$ (7.3 mg, 9.1 μmol) was dissolved in H_2O (55 μL) and the pH was adjusted to 4 using 30% acetic acid. Then, dimethyl sulfate $(CH_3)_2SO_4$ (8.6 μL, 91 μmol, 10 equivalents) was added in 10 aliquots every 5 min under vigorous stirring. The pH was kept at the level of 4 by addition of 10% solution of KOH. The reaction was quenched after 3 h by dilution with H_2O (1 mL) and the mixture was washed with diethyl ether three times. Then, the pH of the aqueous phase was adjusted to 7.0 using 10% $KOH_{(aq)}$ and the reaction product was purified by semi-preparative RP HPLC (isocratic elution with 0.05 M ammonium acetate buffer pH 5.9) on Discovery RP Amide C-16 column (25 cm x 21.2 mm, 5 μm; flow rate 5.0 mL/min). After evaporation and repeated freeze-drying, an ammonium salt of ε-O-(2-propargyl) 7-methylguanosine pentaphosphate $m^7Gp_5OC_3H_3$ (1.0 mg, 1.25 μmol) was obtained as a white solid in 14% yield.

RP-HPLC: R_t = 3.740 min; HRMS ESI (-) calcd. *m/z* for $C_{14}H_{21}N_5O_{20}P_5^-$ [M-H]$^-$: 733.94734, found: 733.94851;

$m^7Gp_5OC_3(5)FAM$



An aqueous solution (10 μL) of ε-*O*-(2-propargyl) *N*^{7}-methylguanosine pentaphosphate ammonium salt $m^7Gp_5OC_3H_3$ (0.65 mg, 0.81 μmol) was mixed with a solution of *N*-(2-azidoethyl)-5-fluorescein-carboxoamide (5)FAM-N_3 (0.37 mg, 0.84 μmol, 0.06 M) in DMSO (14 μL), followed by the addition of aqueous solutions of 1.0 M $CuSO_4 \cdot 5H_2O$ (0.75 mg, 3.0 μmol, 3.0 μL) and 1.0 M sodium ascorbate (1.2 mg, 6.0 μmol, 6.0 μL) and dilution with 20 μL of DMSO. The reaction was vortexed at room temperature for 6 h and then quenched by addition of an aqueous 0.1 M solution of Na_2EDTA (5.6 mg, 15 μmol). The samples were centrifuged to remove excess of fluorescein dye, which precipitated from the solution. The product was purified by semi-preparative RP HPLC on Grace Vision HT C18 HL column (25 cm x 22 mm, 10 μm, flow rate 5.0 mL/min) with a linear gradient (0–100% in 120 min) of acetonitrile in 0.05 M ammonium acetate buffer pH 5.9. After evaporation and repeated freeze-drying, an ammonium salt of $m^7Gp_5OC_3(5)FAM$ (0.50 mg, 0.4 μmol) was obtained as an orange solid in 50% yield.

RP-HPLC: R_t = 11.290 min; HRMS ESI (-) calcd. m/z for $C_{37}H_{37}N_9O_{26}P_5^-$ [M-H]$^-$: 1178.05433, found: 1178.05609;

Protein expression and purification

Murine (*Mus Musculus*) eukaryotic translation initiation factor 4E (eIF4E, residues 28-217) was expressed and purified as described previously[6] with an additional gel filtration purification step on HiLoad 16/600 Superdex 75 pg (GE Healthcare) using 20 mM HEPES pH 7.2, 100 mM KCl, 0.5 mM EDTA, 2 mM DTT buffer. For crystallization and isothermal titration calorimetry the protein was concentrated to 4.75 mg/ml (Amicon 10,000 molecular weight Ultra Centrifugal Filter Device), aliquoted, flash frozen in liquid nitrogen and stored at -80°C, while for fluorescence quenching titration, 28 μM protein samples supplemented with 10 % glycerol were flash frozen and stored at higher salt concentration reaching 270 mM KCl.

^{15}N labelled murine eukaryotic translation initiation factor 4E (eIF4E, residues 28-217) was expressed in M9 media (pH = 7.4) containing: Na_2HPO_4 (6 g/L), KH_2PO_4 (3 g/L), NaCl (0.5 g/L), ampicillin (100 mg/L), $CaCl_2$

(0.1 mM), $MgSO_4$ (1.5 mM), thiamine (1 mg/L), $^{15}NH_4Cl$ (1 g/L) and D-glucose (2 g/L) according to the protocol analogous to the non-labelled protein. Briefly, 1 L of M9 media was inoculated with 200 µL of E. coli BL21 starter culture in LB media (at OD_{600} = 1.8) and the cells were grown at 37°C overnight to OD_{600} = 0.8. Protein expression was induced with IPTG (1M, 0.5 mL) for 3.5 h at 37°C. Cell were centrifuged at 7000 x *g* and lysed by sonication in lysis buffer (20 mM HEPES/KOH pH 7.5, 100 mM KCl, 1 mM EDTA, 10% glycerol, 2 mM DTT). Inclusion bodies were centrifuged and washed with buffer A1 (1 M guanidinium hydrochloride, 20 mM HEPES/KOH pH 7.2, 10% glycerol, 2 mM DTT) by sonication. The procedure was repeated 3 times and then the inclusion bodies were denatured in buffer A2 (6 M guanidinium hydrochloride, 50 mM HEPES/KOH pH 7.2, 10% glycerol, 2 mM DTT) by sonication, clarified at 20,000 x *g* and diluted with 100 mL of buffer B1 (4 M guanidinium hydrochloride, 50 mM HEPES/KOH pH 7.2, 10% glycerol, 2 mM DTT). The protein was refolded by dialysis into 2 L of buffer B2 (50 mM HEPES/KOH pH 7.2, 100 mM KCl, 0.5 mM EDTA, 1 mM DTT) for 3 h and then into 2 L of fresh buffer B2 overnight. The solution was filtered and loaded onto HiTrap SP XL column (5 mL) pre-equilibrated with IEC buffer A (50 mM HEPES/KOH pH 7.2, 100 mM KCl). The protein was eluted with 0-100% gradient of buffer B (50 mM HEPES/KOH pH 7.2, 1 M KCl), concentrated and the buffer was exchanged to NMR buffer (50 mM NaH_2PO_4 pH 7.4, 100 mM NaCl, 1 mM DTT) using Amicon 10,000 molecular weight Ultra Centrifugal Filter Device. Finally, the protein was concentrated to ca. 250 µL. Stock solution of cap analog and 20 µL of D_2O were added prior to the NMR experiments.

Crystallography

The protein was incubated with 1 mM cap analog at rt for ca. 30 minutes prior to crystallization setup. Thin plate-like crystals were obtained by sitting drop vapor diffusion (200 nL + 200 nL drops, rt) in 96-well plates with 3-lens wells (Swissci) using a pipetting robot Mosquito Crystal (TTP Labtech). Optimization of the crystallization conditions by hanging drop vapor diffusion (1 µL + 1 µL drops, rt) provided diffraction-quality crystals within 1–2 days. Cryoprotectant solution (reservoir solution and glycerol 1:1 v/v) was added to the crystallization drop (1:1 v/v), the crystals were harvested and flash frozen in liquid nitrogen.

The X-Ray diffraction data sets were collected at 100K at Bessy II (Helmholtz-Zentrum Berlin, Germany)[7] Beamline 14.1 using a Dectris PILATUS 6M detector and processed using XDS with XDSAPP GUI.[8, 9] Structures were solved by Molecular Replacement using Phaser[10] and chain B of eIF4E-m^{7}GpppG complex structure (pdb entry 1L8B)[11] as a search model. Data collection and refinement statistics are summarized in Table S2. Ligand dictionaries were generated using ProDRG[12] and JLigand[13]. The model building and ligand fitting was performed in Coot[14] and the structures were refined using phenix.refine[15]. Relatively high R-factor values for some of the structures result from disorder of distal fragments of two protein molecule copies (chain C and D for structures solved in P1 space group or chain B for structures solved in C121 space group). For the same reason, chains C and D (P1 space group) or chain B (C121 space group) contain a large percentage of RSRZ outliers. Some of those residues were deleted from the model during refinement, but further deletion increased the R-factor values considerably. Nonetheless, the electron density map is very well defined for the majority of chain A and chain B (P1 space group) or chain A (C121 space group) residues – including cap binding pocket – and for the core of chain C and chain D (P1 space group) or chain B (C121 space group).

NMR spectroscopy

^{1}H-^{15}N HSQC spectra were recorded at 25°C on Bruker Avance III 500 MHz spectrometer equipped with PA TXI 500S1 H-C/N-D-05 Z probe. The spectra were processed and visualized using MestReNova 12. Resonance assignments were transferred from previously published spectra of apo eIF4E (BMRB id: 7115)[16] and eIF4E/m^{7}GDP complex (BMRB id: 5712)[17]. Chemical shift perturbations were calculated using Mnova Binding as $\mathrm{sqrt}[\Delta\delta_H^2 + (0.15\cdot \Delta\delta_N)^2]$ and mapped onto eIF4E surface (PDB id: 1L8B) using PyMOL.

Fluorescence quenching titration

Fluorescence quenching titration (FQT) assay was performed on Cary Eclipse Agilent Technology spectrofluorimeter. Titration experiments were performed at 20 °C in 50 mM Hepes/KOH buffer pH 7.20 containing 100 mM KCl, 0.5 mM EDTA and 1 mM DTT. Aliquots (1 µl) of the tested ligand were added to 1400 µl solution of 0.1 µM eIF4E. Fluorescence of the measured samples was excited at 280 nm (2.5 nm bandwidth) and detected at 337 nm (4 nm bandwidth). Fluorescence intensities were corrected for sample dilution and inner filter effect.[11] Equilibrium association constants (K_{AS}) were determined by fitting the theoretical dependence of fluorescence intensity on the total concentration of the cap analogue to the experimental data points, according

to the equation described previously.[11] Each experiment was repeated three times and the association constants K_{AS} were calculated as weighted averages, with the weights taken from reciprocal standard deviations squared. The dissociation constants K_D reported in Table 1 and Table S2 were calculated as K_{AS}^{-1}.

Microscale thermophoresis

Direct binding experiment (probe $m^7Gp_5OC_3$(5)FAM–eIF4E)

To evaluate the affinity of $m^7Gp_5OC_3$(5)FAM to eIF4E direct binding MST experiments were performed. Fluorescent probe $m^7Gp_5OC_3$(5)FAM was diluted in MST buffer (50 mM Hepes–KOH pH 7.2, 100 mM KCl, 0.5 mM EDTA, and 0.05% Tween–20) to a concentration 20 nM. A 16–point 1:1 dilution series starting from 1.57 µM of murine eIF4E protein was prepared in MST buffer. Equal volumes (10 µL) of the two components were mixed to obtain the following final concentrations in the assay: 10 nM $m^7Gp_5OC_3$(5)FAM and eIF4E up to 0.785 µM. Samples were loaded into standard Monolith NT.115 Capillaries (cat. no. MO–KO22; NanoTemper Technologies). MST was measured using a Monolith NT.115 instrument (NanoTemper Technologies) at 25°C. Instrument parameters were adjusted to 100% LED Blue power and Medium MST power. Data obtained in 3 independently pipetted measurements were analyzed using PALMIST software (version 1.4.4)[18] in the T-jump mode (cold region range: -3 to -2 s; hot region range: 0.5 to 1.0 s). To determine the K_D values, a standard 1:1 binding model was fitted to the data using PALMIST software (version 1.4.4). Due model limitations the fitting could not yield exact K_D values but revealed that the value is below 2 nM.

Competitive binding assay (probe $m^7Gp_5OC_3$(5)FAM–eIF4E–ligand)

$m^7Gp_5OC_3$(5)FAM was dissolved in MST buffer (50 mM Hepes–KOH pH 7.2, 100 mM KCl, 0.5 mM EDTA, and 0.05% Tween–20) at the final concentration of 20 nM together with murine eIF4E protein at final concentration 100 nM. A 16–point 1:1 dilution series ranging from 1000 µM to 30.4 nM (di- and trinucelotides) or 3.5 µM to 0.1 nM (RNAs) of ligand was prepared in MST buffer. Equal volumes (10 µL) of probe–eIF4E and ligand were mixed to obtain the following final concentrations in the assay: 10 nM probe $m^7Gp_5OC_3$(5)FAM, 50 nM eIF4E and 500 µM to 15.2 nM ligand (di- and trinucleotides) or 1.75 µM to 0.05 nM (RNAs). Samples were loaded into Monolith NT.115 Capillaries (cat. no. MO–KO22; NanoTemper Technologies). MST was measured using a Monolith NT.115 instrument (NanoTemper Technologies) at 25°C. Measurement settings the same as described for the direct binding assay. To determine the $K_{d,\,app}$ values, a standard 1:1 binding model was fitted to the data using PALMIST software (version 1.4.4). Confidence Intervals were determined using error-surface projection (ESP).[18]

Synthesis of capped RNAs by in vitro transcription

Short RNAs were generated on template of annealed oligonucleotides (CAGTAATACGACTCACTATAGGGGAAGCGGGCATGCGGCCAGCCATAGCCGATCA and TGATCGGCTATGGCTGGCCGCATGCCCGCTTCCCCTATAGTGAGTCGTATTACTG) [19], which contains T7 promoter sequence (TAATACGACTCACTATA) and encodes 35-nt long sequence (GGGGAAGCGGGCATGCGGCCAGCCATAGCCGATCA). Typical in vitro transcription reaction (100 µl) was incubated at 37 °C for 2 h and contained: RNA Pol buffer (40 mM Tris-HCl pH 7.9, 6 mM $MgCl_2$, 1 mM DTT, 2 mM spermidine), 10 U/µl T7 RNA polymerase (ThermoFisher Scientific), 1 U/µl RiboLock RNase Inhibitor (ThermoFisher Scientific), 0.5 mM ATP/CTP/UTP, 0.125 mM GTP, 1.25 mM cap analog of interests and 0.1 µM annealed oligonucleotides as a template. Following 2 h incubation, 0.1 U/µl DNase I (ThermoFisher Scientific) was added and incubation was continued for 30 min at 37 °C. The crude RNAs were purified using RNA Clean & Concentrator-25 (Zymo Research). Quality of transcripts was checked on 15% acrylamide/7 M urea gels, whereas concentration was determined spectrophotometrically. To remove in vitro transcription by-products of unintended size RNA samples were gel-purified using PAA elution buffer (0.3 M sodium acetate, 1 mM EDTA, 0.05% Triton X-100), precipitated with isopropanol and dissolved in water.

Isothermal titration calorimetry

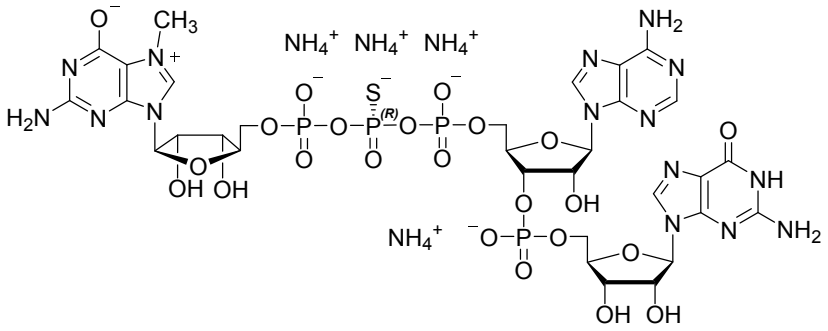
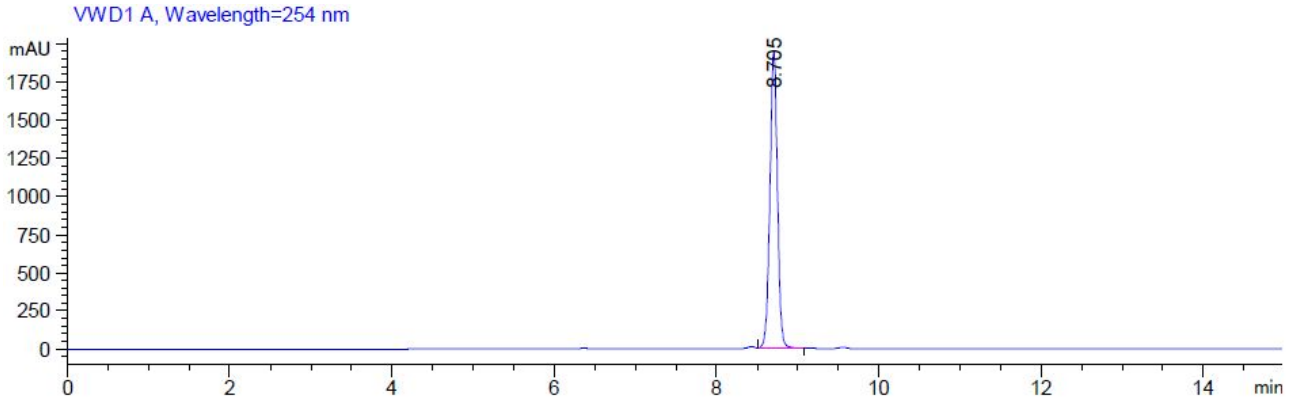
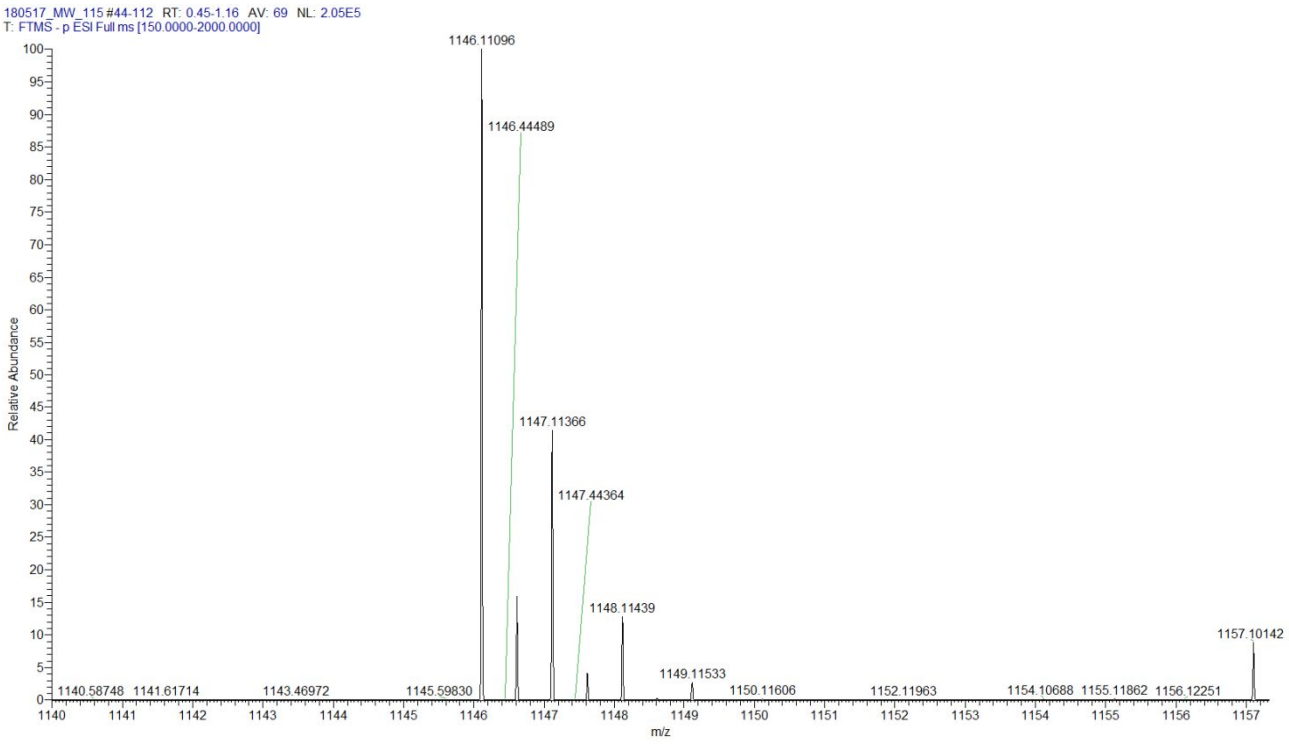
Isothermal titration calorimetry experiments were carried out at 20 °C in 50 mM Hepes/KOH buffer pH 7.20 containing 100 mM KCl on MicroCal™ iTC200 calorimeter with a 200 µl sample cell and 40 µl titration syringe. Ligand was dissolved in the same buffer as protein. 180 µM ligand solution was injected in 150 s time spacing as 35 aliquots of 1ul into stirred (750 rpm) 15 µM protein solution. Automated baseline assignment and peak integration of raw ITC thermograms were accomplished by singular value decomposition and peak-shape analysis using NITPIC.[20] Global ITC analysis was performed from replicate experiments and one site binding

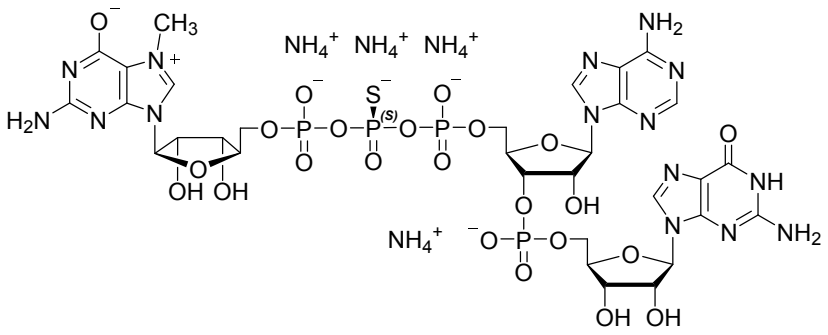
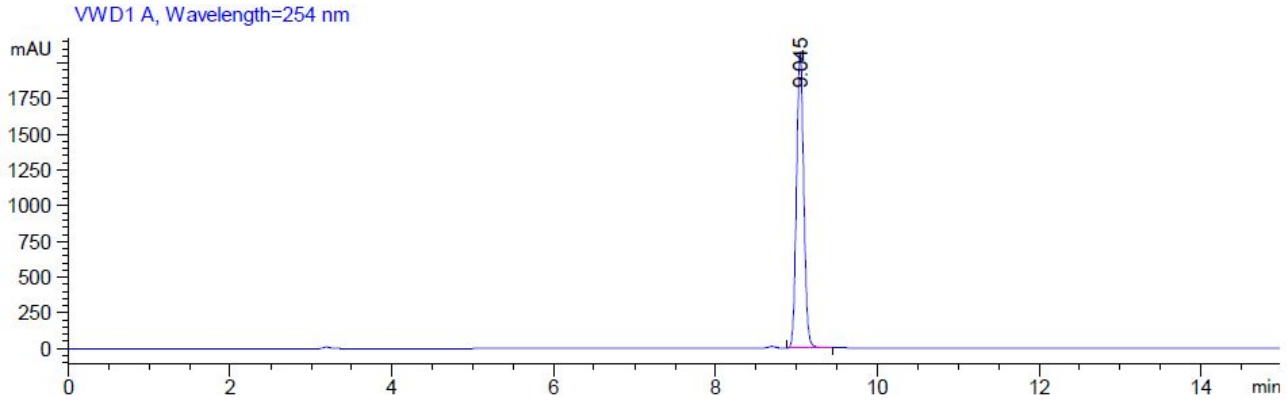
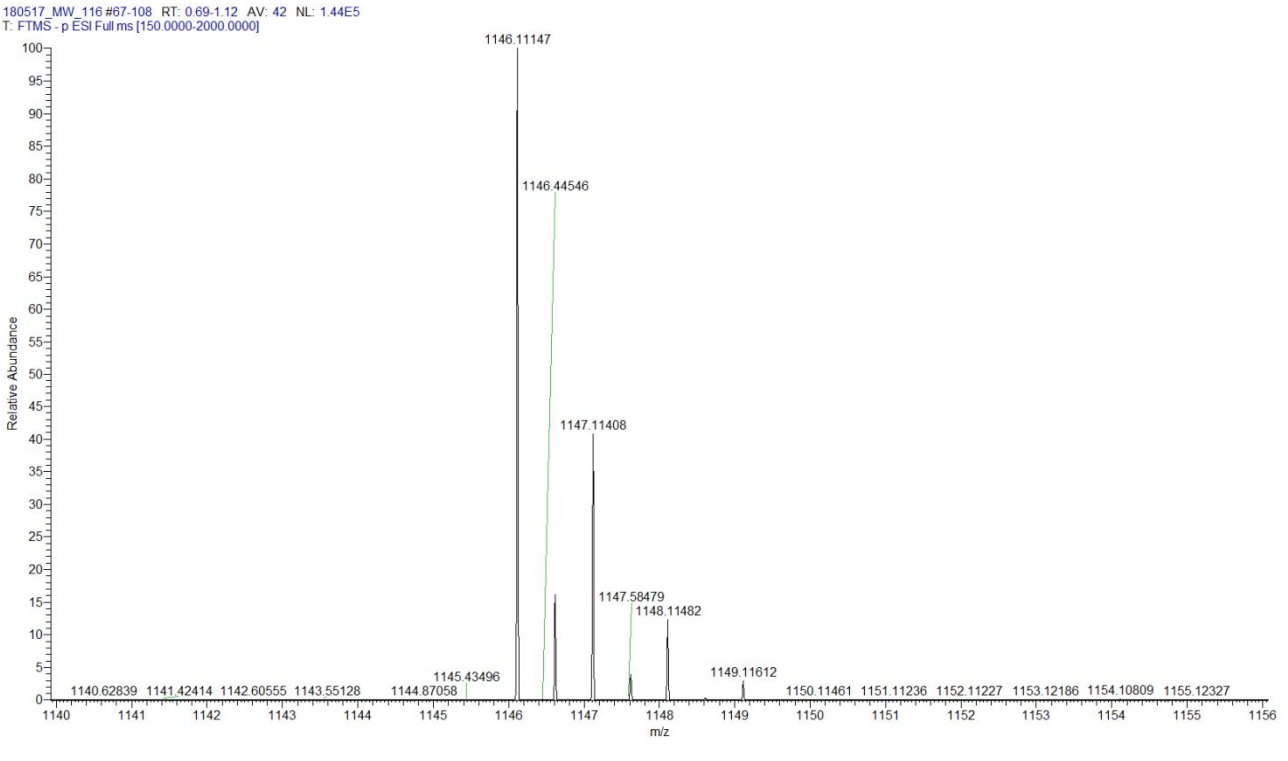
model was fitted to the experimental data with the software SEDPHAT[21, 22] 68.3 % confidence interval was calculated in terms of the dissociation constant.

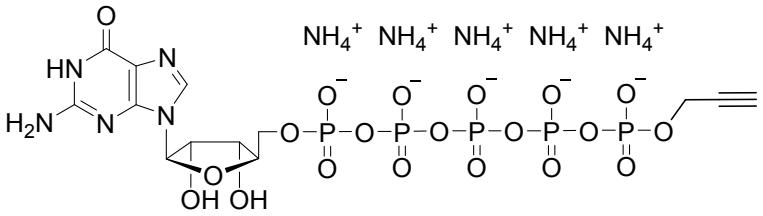
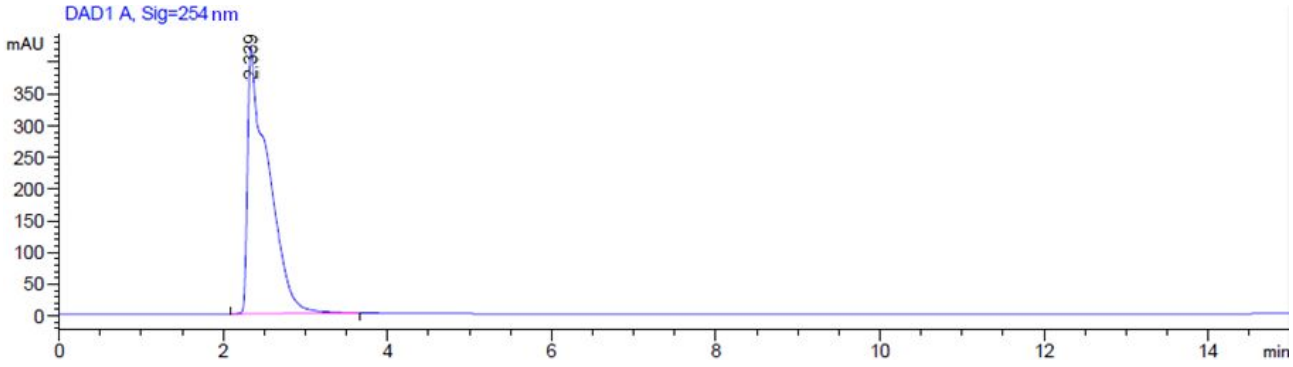
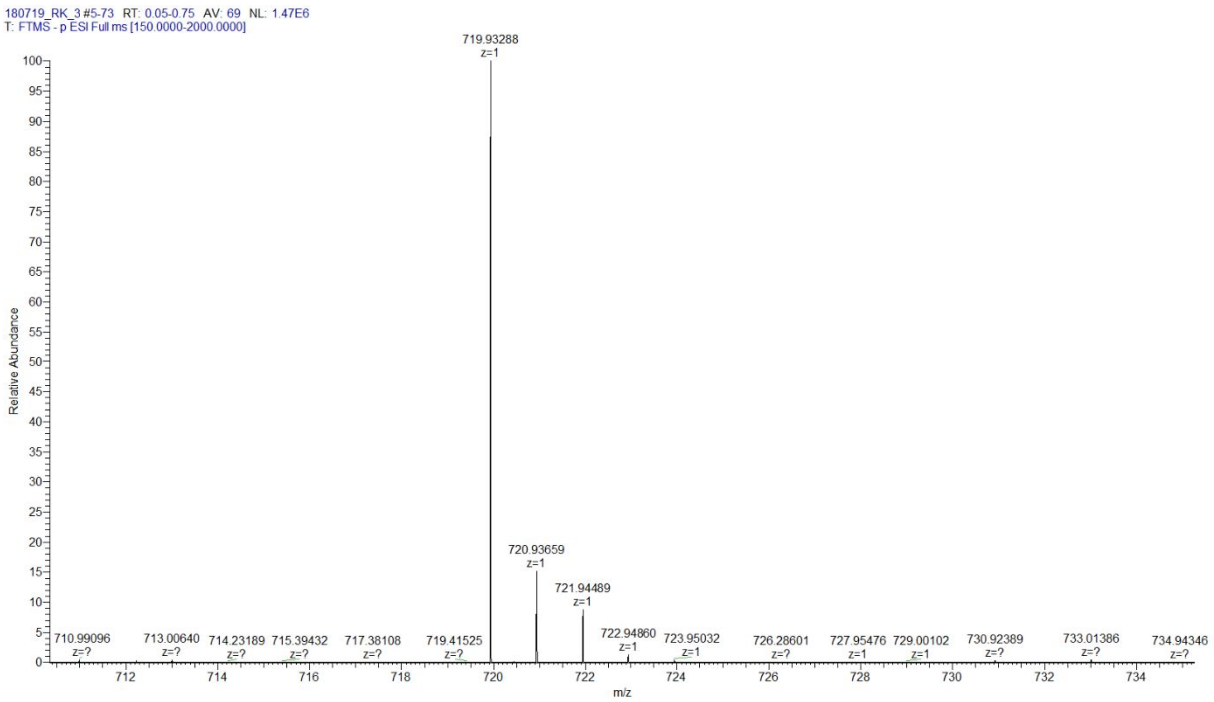
References:

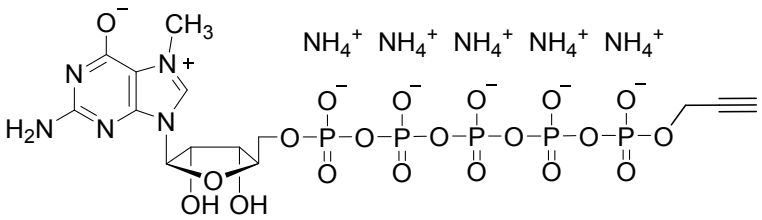
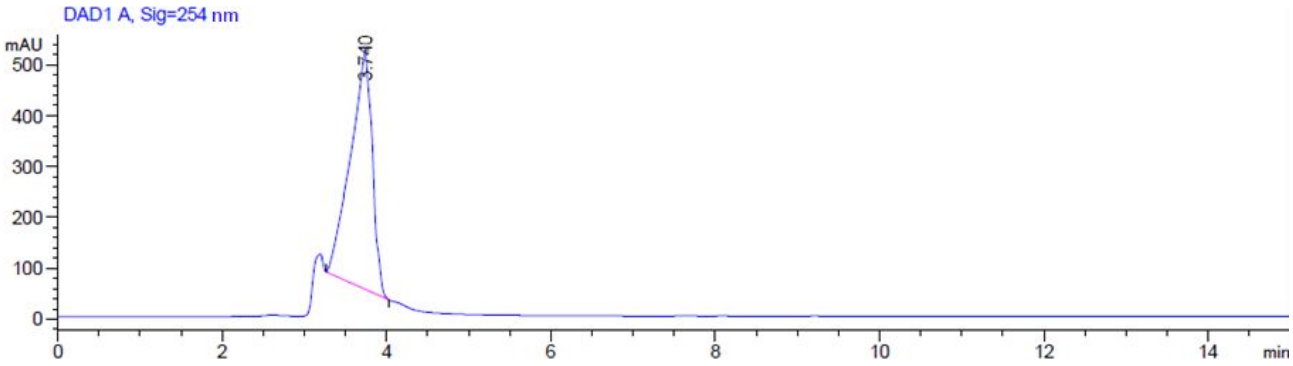
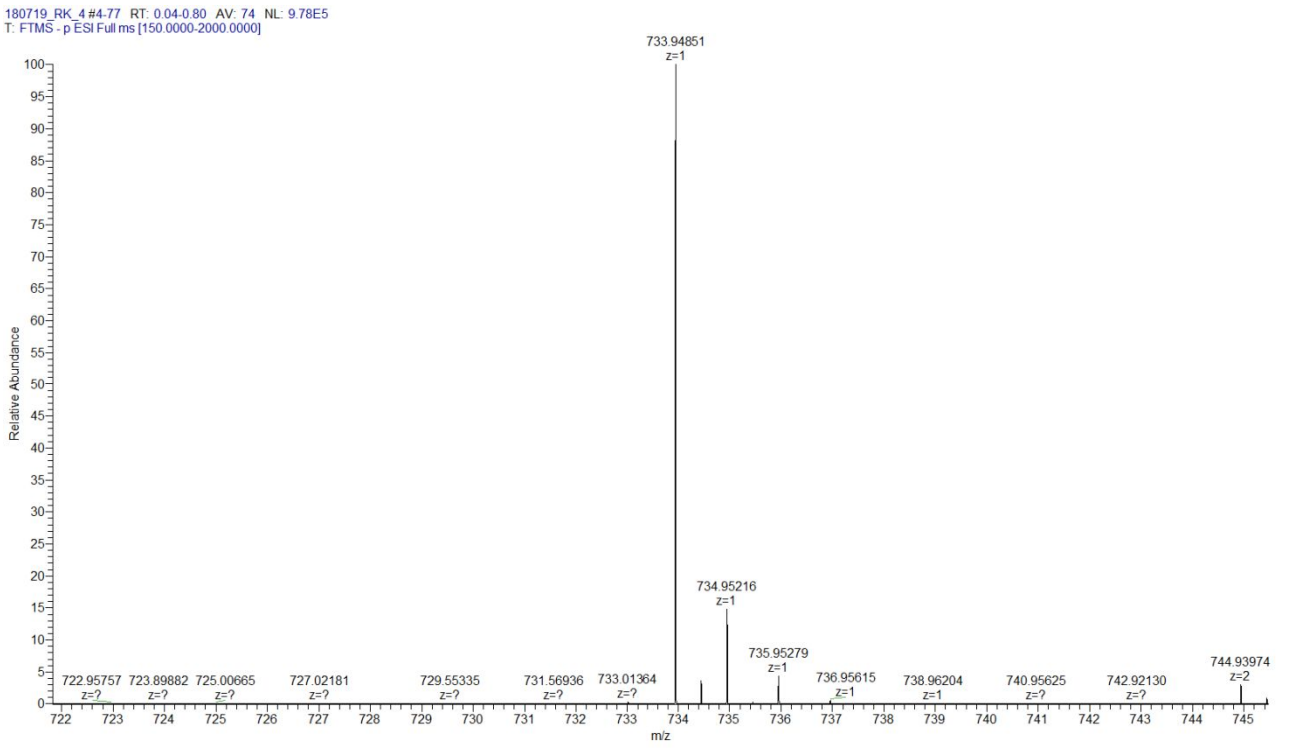
1. Kowalska, J.; Lewdorowicz, M.; Zuberek, J.; Grudzien-Nogalska, E.; Bojarska, E.; Stepinski, J.; Rhoads, R. E.; Darzynkiewicz, E.; Davis, R. E.; Jemielity, J., Synthesis and characterization of mRNA cap analogs containing phosphorothioate substitutions that bind tightly to eIF4E and are resistant to the decapping pyrophosphatase DcpS. *RNA* **2008,** *14* (6), 1119-1131.
2. Kowalska, J.; Lukaszewicz, M.; Zuberek, J.; Darzynkiewicz, E.; Jemielity, J., Phosphoroselenoate Dinucleotides for Modification of mRNA 5′ End. *ChemBioChem* **2009,** *10* (15), 2469-2473.
3. Warminski, M.; Kowalska, J.; Buck, J.; Zuberek, J.; Lukaszewicz, M.; Nicola, C.; Kuhn, A. N.; Sahin, U.; Darzynkiewicz, E.; Jemielity, J., The synthesis of isopropylidene mRNA cap analogs modified with phosphorothioate moiety and their evaluation as promoters of mRNA translation. *Bioorganic & Medicinal Chemistry Letters* **2013,** *23* (13), 3753-3758.
4. Strenkowska, M.; Wanat, P.; Ziemniak, M.; Jemielity, J.; Kowalska, J., Preparation of Synthetically Challenging Nucleotides Using Cyanoethyl P-Imidazolides and Microwaves. *Organic Letters* **2012,** *14* (18), 4782-4785.
5. Sikorski, P. J.; Warminski, M.; Kubacka, D.; Ratajczak, T.; Nowis, D.; Kowalska, J.; Jemielity, J., The identity and methylation status of the first transcribed nucleotide in eukaryotic mRNA 5′ cap modulates protein expression in living cells. *Nucleic Acids Research* **2020**.
6. Zuberek, J.; Wyslouch-Cieszynska, A.; Niedzwiecka, A.; Dadlez, M.; Stepinski, J.; Augustyniak, W.; Gingras, A.-C.; Zhang, Z.; Burley, S. K.; Sonenberg, N.; Stolarski, R.; Darzynkiewicz, E., Phosphorylation of eIF4E attenuates its interaction with mRNA 5′ cap analogs by electrostatic repulsion: Intein-mediated protein ligation strategy to obtain phosphorylated protein. *RNA* **2003,** *9* (1), 52-61.
7. Mueller, U.; Förster, R.; Hellmig, M.; Huschmann, F. U.; Kastner, A.; Malecki, P.; Pühringer, S.; Röwer, M.; Sparta, K.; Steffien, M.; Ühlein, M.; Wilk, P.; Weiss, M. S., The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *The European Physical Journal Plus* **2015,** *130* (7), 141.
8. Krug, M.; Weiss, M. S.; Heinemann, U.; Mueller, U., XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. *Journal of Applied Crystallography* **2012,** *45* (3), 568-572.
9. Kabsch, W., XDS. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (Pt 2), 125-132.
10. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007,** *40* (Pt 4), 658-674.
11. Niedzwiecka, A.; Marcotrigiano, J.; Stepinski, J.; Jankowska-Anyszka, M.; Wyslouch-Cieszynska, A.; Dadlez, M.; Gingras, A.-C.; Mak, P.; Darzynkiewicz, E.; Sonenberg, N.; Burley, S. K.; Stolarski, R., Biophysical Studies of eIF4E Cap-binding Protein: Recognition of mRNA 5′ Cap Structure and Synthetic Fragments of eIF4G and 4E-BP1 Proteins. *Journal of Molecular Biology* **2002,** *319* (3), 615-635.
12. Schuttelkopf, A. W.; van Aalten, D. M. F., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D* **2004,** *60* (8), 1355-1363.
13. Lebedev, A. A.; Young, P.; Isupov, M. N.; Moroz, O. V.; Vagin, A. A.; Murshudov, G. N., JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallographica Section D* **2012,** *68* (4), 431-440.
14. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **2004,** *60* (12 Part 1), 2126-2132.
15. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **2010,** *66* (2), 213-221.
16. Volpon, L.; Osborne, M. J.; Topisirovic, I.; Siddiqui, N.; Borden, K. L. B., Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands. *The EMBO Journal* **2006,** *25* (21), 5138.
17. Miura, T.; Shiratori, Y.; Shimma, N., Letter to the Editor: Backbone resonance assignment of human eukaryotic translation initiation factor 4E (eIF4E) in complex with 7-methylguanosine diphosphate (m7GDP) and a 17-amino acid peptide derived from human eIF4GII. *Journal of Biomolecular NMR* **2003,** *27* (3), 279-280.
18. Scheuermann, T. H.; Padrick, S. B.; Gardner, K. H.; Brautigam, C. A., On the acquisition and analysis of microscale thermophoresis data. *Analytical Biochemistry* **2016,** *496*, 79-93.
19. Coleman, T.; Wang, G.; Huang, F., Superior 5 ' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic Acids Research* **2004,** *32* (1).
20. Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P., High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis. *Analytical Chemistry* **2012,** *84* (11), 5066-5073.
21. Brautigam, C. A.; Zhao, H.; Vargas, C.; Keller, S.; Schuck, P., Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nature Protocols* **2016,** *11* (5), 882-894.
22. Zhao, H.; Piszczek, G.; Schuck, P., SEDPHAT – A platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods* **2015,** *76*, 137-148.

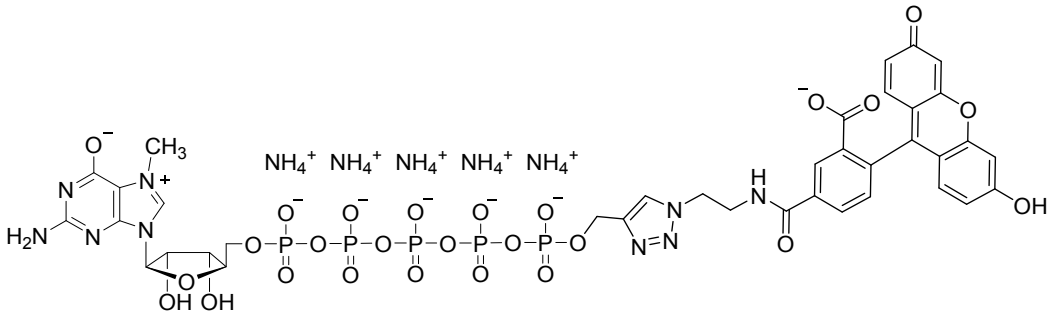
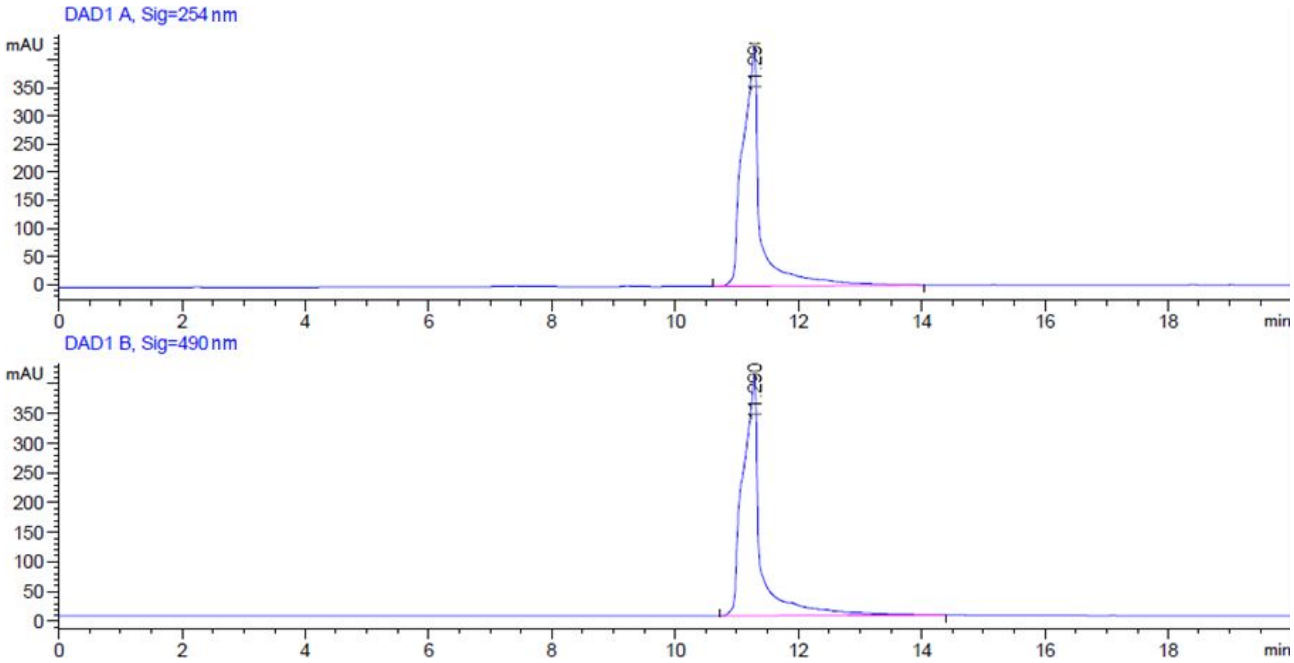
Compounds characterization

	m^7Gpp_spApG D1
Chemical structure	
RP HPLC	
ESI(-) MS: m/z calcd. for $C_{31}H_{40}N_{15}O_{23}P_4S^-$ [M-H]⁻: 1146.10981	

	m^7Gpp_SpApG D2
Chemical structure	
RP HPLC	
***ESI(-) MS: m/z* calcd. for $C_{31}H_{40}N_{15}O_{23}P_4S^-$ $[M-H]^-$: 1146.10981**	

	$Gp_5OC_3H_3$
Chemical structure	
RP HPLC	
***ESI(-) MS: m/z* calcd. *m/z* for $C_{13}H_{19}N_5O_{20}P_5^-$ $[M-H]^-$: 719.9317**	

	$m^7Gp_5OC_3H_3$
Chemical structure	
RP HPLC	
***ESI(-) MS: m/z* calcd. *m/z* for $C_{14}H_{21}N_5O_{20}P_5^-$ $[M-H]^-$: 733.94734**	

	m7Gp$_5$OC$_3$(5)FAM
Chemical structure	
RP HPLC	
***ESI(-) MS: m/z* calcd. for $C_{37}H_{37}N_9O_{26}P_5^-$ [M-H]$^-$: 1178.05433**