Supporting Information

Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP

Nikita Vasilyev, Anna Polonskaia, Jennifer C. Darnell, Robert B. Darnell, Dinshaw J. Patel and Alexander Serganov

Table S1. Crystallographic statistics for RGG-*sc1* complex

Dataset	$[Ir(NH_3)_6]^{3+}$	Cs^+	Native
Data collection*			
Spacegroup	*P* 2_12_12	*P* 2_12_12	*P* 2_12_12
Cell dimensions			
a, b, c (Å)	57.65, 130.60, 36.70	56.57, 129.85, 36.79	57.06, 130.02, 36.75
α = β = γ (°)	90	90	90
Wavelength (Å)	1.10528	1.4938	1.1053
Resolution (Å) [a]	20.00-3.10 (3.27-3.10)	20.00-2.80 (2.96-2.80)	20.00-3.00 (3.18-3.00)
R_{merge} (%) [a]	4.9 (14.8)	7.4 (44.9)	11.1 (48.1)
<I>/σ<I> [a]	26.7 (8.6)	15.3 (4.0)	13.3 (3.9)
Completeness (%) [a]	99.0 (98.4)	99.2 (97.7)	99.5 (99.1)
Unique reflections [a]	9,688 (1,519)	12,880 (2,004)	10,516 (1,645)
Redundancy [a]	7.6 (7.7)	5.2 (5.1)	7.4 (7.5)
Data after anisotropy correction* [b]			
Resolution (Å) [a]	20.00-3.10 (3.18-3.10)	20.00-2.80 (2.87-2.80)	20.00-3.00 (3.08-3.00)
R_{merge} (%) [a]	4.9 (24.2)	7.4 (50.9)	11.1 (66.5)
<I>/σ<I> [a]	26.8 (7.45)	15.3 (4.4)	12.9 (3.3)
Completeness (%) [a]	98.9 (98.8)	97.3 (69.5)	99.4 (98.7)
Unique reflections [a]	9,588 (677)	12,553 (639)	10,504 (783)
Redundancy [a]	7.6 (7.7)	5.3 (5.4)	7.4 (7.3)
Refinement			
Resolution (Å)	20.0-3.1	20.0-2.8	20.0-3.0
Number of reflections [c]			
Working set	5,091	6,616	5,576
Test set	243	331	265
R_{work} / R_{free} (%) [c]	20.1 / 22.7	20.2 / 22.6	19.5 / 22.1
Number of atoms			
RNA	1,518	1,518	1,518
Protein	170	170	164
Cations	7	7	7
Water	-	8	4
Average B-factors (Å^2)			
RNA	47.3	37.6	50.3
Protein	44.0	32.8	41.1
Cations	46.8 for K^+ 97.2 for $[Ir(NH_3)_6]^{3+}$	16.4 for K^+ 88.4 for Cs^+	40.5 for K^+
Water	-	19.2	18.2
R.M.S.D. from ideal geometry			
Bond length (Å) [c]	0.006	0.004	0.003
Bond angles (°) [c]	1.217	1.287	0.846
Coordinates error [d]	0.37	0.25	0.35

* Values in data collection and anisotropy correction are reported by XSCALE;
[a] Values in parentheses are for highest-resolution shell;
[b] Anisotropy correction was performed using Diffraction Anisotropy Server (http://services.mbi.ucla.edu/anisoscale);
[c] Values reported by PHENIX.REFINE;
[d] Coordinate errors based on maximum likelihood are reported by PHENIX.REFINE.

Table S2. Comparison of G-quadruplexes in crystal structures of the RGG-*sc1* complex using pair_fit function in PyMol

Model 1	Model 2	Residues	Atoms	r.m.s.d.
Native chain A	Native chain A	6-30	P+O5+C5+C4+C3+O3	0.00
	Native chain C	6-30	P+O5+C5+C4+C3+O3	3.14
	Cs^+ chain A	6-30	P+O5+C5+C4+C3+O3	0.86
	Cs^+ chain C	6-30	P+O5+C5+C4+C3+O3	3.10
	$[Ir(NH_3)_6]^{3+}$ chain A	6-30	P+O5+C5+C4+C3+O3	0.90
	$[Ir(NH_3)_6]^{3+}$ chain C	6-30	P+O5+C5+C4+C3+O3	3.10
Native chain C	Native chain A	6-30	P+O5+C5+C4+C3+O3	3.14
	Native chain C	6-30	P+O5+C5+C4+C3+O3	0.00
	Cs^+ chain A	6-30	P+O5+C5+C4+C3+O3	3.18
	Cs^+ chain C	6-30	P+O5+C5+C4+C3+O3	1.44
	$[Ir(NH_3)_6]^{3+}$ chain A	6-30	P+O5+C5+C4+C3+O3	2.94
	$[Ir(NH_3)_6]^{3+}$ chain C	6-30	P+O5+C5+C4+C3+O3	0.87
Cs^+ chain A	Native chain A	6-30	P+O5+C5+C4+C3+O3	0.86
	Native chain C	6-30	P+O5+C5+C4+C3+O3	3.18
	Cs^+ chain A	6-30	P+O5+C5+C4+C3+O3	0.00
	Cs^+ chain C	6-30	P+O5+C5+C4+C3+O3	3.16
	$[Ir(NH_3)_6]^{3+}$ chain A	6-30	P+O5+C5+C4+C3+O3	1.28
	$[Ir(NH_3)_6]^{3+}$ chain C	6-30	P+O5+C5+C4+C3+O3	3.16
Cs^+ chain C	Native chain A	6-30	P+O5+C5+C4+C3+O3	3.10
	Native chain C	6-30	P+O5+C5+C4+C3+O3	1.44
	Cs^+ chain A	6-30	P+O5+C5+C4+C3+O3	3.16
	Cs^+ chain C	6-30	P+O5+C5+C4+C3+O3	0.00
	$[Ir(NH_3)_6]^{3+}$ chain A	6-30	P+O5+C5+C4+C3+O3	2.87
	$[Ir(NH_3)_6]^{3+}$ chain C	6-30	P+O5+C5+C4+C3+O3	1.45
$[Ir(NH_3)_6]^{3+}$ chain A	Native chain A	6-30	P+O5+C5+C4+C3+O3	0.90
	Native chain C	6-30	P+O5+C5+C4+C3+O3	2.94
	Cs^+ chain A	6-30	P+O5+C5+C4+C3+O3	1.28
	Cs^+ chain C	6-30	P+O5+C5+C4+C3+O3	2.87
	$[Ir(NH_3)_6]^{3+}$ chain A	6-30	P+O5+C5+C4+C3+O3	0.00
	$[Ir(NH_3)_6]^{3+}$ chain C	6-30	P+O5+C5+C4+C3+O3	2.94
$[Ir(NH_3)_6]^{3+}$ chain C	Native chain A	6-30	P+O5+C5+C4+C3+O3	3.10
	Native chain C	6-30	P+O5+C5+C4+C3+O3	0.87
	Cs^+ chain A	6-30	P+O5+C5+C4+C3+O3	3.16
	Cs^+ chain C	6-30	P+O5+C5+C4+C3+O3	1.45
	$[Ir(NH_3)_6]^{3+}$ chain A	6-30	P+O5+C5+C4+C3+O3	2.94
	$[Ir(NH_3)_6]^{3+}$ chain C	6-30	P+O5+C5+C4+C3+O3	0.00

Table S3. Comparison of crystal and NMR structures of the RGG-*sc1* complex using pair_fit function in PyMol

Model 1	Model 2	R.M.S.D. RNA P	R.M.S.D. RNA backbone (P+O5+C5+C4+C3+O3)	R.M.S.D. Peptide Cα	R.M.S.D. Peptide backbone (N+CA+C+O)
K^+ #1	NMR #1	2.52	3.33	2.30	2.02
	NMR #2	2.39	3.29	2.01	2.11
	NMR #3	2.13	2.98	2.46	2.46
	NMR #4	1.97	2.81	2.52	2.55
	NMR #5	2.37	3.19	2.27	2.30
	NMR #6	2.25	2.72	2.32	2.18
	NMR #7	2.65	3.39	2.12	2.05
	NMR #8	2.08	2.82	2.29	2.43
	NMR #9	2.17	2.77	2.61	2.57
	NMR #10	2.12	2.95	2.11	1.98
K^+ #2	NMR #1	2.71	3.07	2.14	2.22
	NMR #2	2.44	2.75	2.04	2.10
	NMR #3	2.21	3.10	2.48	2.46
	NMR #4	1.92	2.86	2.59	2.51
	NMR #5	2.51	2.56	2.32	2.28
	NMR #6	2.27	3.06	2.16	2.30
	NMR #7	2.83	3.07	2.12	2.09
	NMR #8	2.09	2.90	2.44	2.33
	NMR #9	2.31	2.98	2.58	2.58
	NMR #10	2.20	3.10	2.07	2.07
Cs^+ #1	NMR #1	2.71	3.29	2.13	2.01
	NMR #2	2.42	3.26	2.03	2.09
	NMR #3	2.18	2.94	2.48	2.45
	NMR #4	1.91	2.77	2.58	2.53
	NMR #5	2.49	3.15	2.30	2.29
	NMR #6	2.27	2.66	2.14	2.17
	NMR #7	2.83	3.37	2.11	2.04
	NMR #8	2.07	2.79	2.44	2.42
	NMR #9	2.30	2.76	2.58	2.56
	NMR #10	2.17	2.91	2.06	1.96
Cs^+ #2	NMR #1	2.50	3.09	2.31	2.22
	NMR #2	2.38	2.77	2.01	2.09
	NMR #3	2.10	3.10	2.46	2.45
	NMR #4	1.93	2.87	2.52	2.48
	NMR #5	2.34	2.57	2.26	2.25
	NMR #6	2.22	3.08	2.33	2.29
	NMR #7	2.63	3.08	2.13	2.07
	NMR #8	2.06	2.92	2.27	2.30
	NMR #9	2.14	2.98	2.59	2.55
	NMR #10	2.10	3.12	2.12	2.06
Ir #1	NMR #1	2.80	3.37	2.16	2.03
	NMR #2	2.51	3.33	2.01	2.03
	NMR #3	2.27	2.95	2.41	2.36
	NMR #4	2.01	2.80	2.50	2.44
	NMR #5	2.59	3.21	2.27	2.22
	NMR #6	2.35	2.70	2.17	2.16
	NMR #7	2.92	3.46	2.09	1.99

	NMR #8	2.13	2.76	2.37	2.31
	NMR #9	2.39	2.85	2.50	2.46
	NMR #10	2.23	2.94	2.04	1.92
Ir #2	NMR #1	2.53	2.88	2.31	2.23
	NMR #2	2.44	2.52	2.04	2.08
	NMR #3	2.15	2.88	2.47	2.39
	NMR #4	1.98	2.62	2.54	2.43
	NMR #5	2.36	2.32	2.30	2.24
	NMR #6	2.30	2.88	2.35	2.29
	NMR #7	2.65	2.89	2.15	2.07
	NMR #8	2.09	2.67	2.31	2.29
	NMR #9	2.16	2.87	2.61	2.49
	NMR #10	2.10	2.88	2.13	2.06

Table S4. Comparison of G-quadruplexes in crystal and NMR structures of RGG-*sc1* complex using pair_fit function in PyMol

Model 1	Model 2	Residues	Atoms	r.m.s.d.
Native chain A	NMR1	6-30	P+O5+C5+C4+C3+O3	3.76
	NMR2	6-30	P+O5+C5+C4+C3+O3	3.64
	NMR3	6-30	P+O5+C5+C4+C3+O3	3.27
	NMR4	6-30	P+O5+C5+C4+C3+O3	3.11
	NMR5	6-30	P+O5+C5+C4+C3+O3	3.56
	NMR6	6-30	P+O5+C5+C4+C3+O3	2.93
	NMR7	6-30	P+O5+C5+C4+C3+O3	3.82
	NMR8	6-30	P+O5+C5+C4+C3+O3	3.17
	NMR9	6-30	P+O5+C5+C4+C3+O3	3.10
	NMR10	6-30	P+O5+C5+C4+C3+O3	3.29
Native chain C	NMR1	6-30	P+O5+C5+C4+C3+O3	3.31
	NMR2	6-30	P+O5+C5+C4+C3+O3	2.99
	NMR3	6-30	P+O5+C5+C4+C3+O3	3.36
	NMR4	6-30	P+O5+C5+C4+C3+O3	3.07
	NMR5	6-30	P+O5+C5+C4+C3+O3	2.68
	NMR6	6-30	P+O5+C5+C4+C3+O3	3.29
	NMR7	6-30	P+O5+C5+C4+C3+O3	3.33
	NMR8	6-30	P+O5+C5+C4+C3+O3	3.31
	NMR9	6-30	P+O5+C5+C4+C3+O3	3.36
	NMR10	6-30	P+O5+C5+C4+C3+O3	3.53
Cs^+ chain A	NMR1	6-30	P+O5+C5+C4+C3+O3	3.79
	NMR2	6-30	P+O5+C5+C4+C3+O3	3.70
	NMR3	6-30	P+O5+C5+C4+C3+O3	3.30
	NMR4	6-30	P+O5+C5+C4+C3+O3	3.19
	NMR5	6-30	P+O5+C5+C4+C3+O3	3.67
	NMR6	6-30	P+O5+C5+C4+C3+O3	2.90
	NMR7	6-30	P+O5+C5+C4+C3+O3	3.81
	NMR8	6-30	P+O5+C5+C4+C3+O3	3.16
	NMR9	6-30	P+O5+C5+C4+C3+O3	2.76
	NMR10	6-30	P+O5+C5+C4+C3+O3	3.30
Cs^+ chain C	NMR1	6-30	P+O5+C5+C4+C3+O3	3.27
	NMR2	6-30	P+O5+C5+C4+C3+O3	3.56
	NMR3	6-30	P+O5+C5+C4+C3+O3	3.64
	NMR4	6-30	P+O5+C5+C4+C3+O3	3.46
	NMR5	6-30	P+O5+C5+C4+C3+O3	2.81
	NMR6	6-30	P+O5+C5+C4+C3+O3	3.33
	NMR7	6-30	P+O5+C5+C4+C3+O3	3.29
	NMR8	6-30	P+O5+C5+C4+C3+O3	3.65
	NMR9	6-30	P+O5+C5+C4+C3+O3	3.88
	NMR10	6-30	P+O5+C5+C4+C3+O3	3.71
$[Ir(NH_3)_6]^{3+}$ chain A	NMR1	6-30	P+O5+C5+C4+C3+O3	3.70
	NMR2	6-30	P+O5+C5+C4+C3+O3	3.59
	NMR3	6-30	P+O5+C5+C4+C3+O3	3.14
	NMR4	6-30	P+O5+C5+C4+C3+O3	3.02
	NMR5	6-30	P+O5+C5+C4+C3+O3	3.38
	NMR6	6-30	P+O5+C5+C4+C3+O3	2.79
	NMR7	6-30	P+O5+C5+C4+C3+O3	3.81
	NMR8	6-30	P+O5+C5+C4+C3+O3	3.03
	NMR9	6-30	P+O5+C5+C4+C3+O3	3.26

	NMR10	6-30	P+O5+C5+C4+C3+O3	3.18
$[Ir(NH_3)_6]^{3+}$ chain C	NMR1	6-30	P+O5+C5+C4+C3+O3	3.17
	NMR2	6-30	P+O5+C5+C4+C3+O3	2.77
	NMR3	6-30	P+O5+C5+C4+C3+O3	3.15
	NMR4	6-30	P+O5+C5+C4+C3+O3	2.85
	NMR5	6-30	P+O5+C5+C4+C3+O3	2.51
	NMR6	6-30	P+O5+C5+C4+C3+O3	3.16
	NMR7	6-30	P+O5+C5+C4+C3+O3	3.14
	NMR8	6-30	P+O5+C5+C4+C3+O3	3.03
	NMR9	6-30	P+O5+C5+C4+C3+O3	3.31
	NMR10	6-30	P+O5+C5+C4+C3+O3	3.29
min				2.51
max				3.88

Table S5. Intermolecular hydrogen bond distances in the crystal structure of the RGG-*sc1* complex

Peptide residue	Atom	Distance, Å	Atom	RNA nucleotide
Arg8	N	3.0-3.1	OP2	C2
	O	3.2-3.3	N4	
	NE*	3.2	OP2	U3
	NH1*	2.9		
Arg10	N	3.1-3.2	O4	
	NE	3.3-3.4	O6	G31
		3.1-3.2	N7	
	NH2	3.1-3.2	N7	
		3.3	OP2	
	NH1	2.7-2.9	OP2	C30
		3.0-3.2	OP2	G29
Gly11	N	3.1-3.2	O6	G4
	O	2.8	N4	C5
Gly12	N	2.9-3.0	OP2	G4
Gly14	N	3.1-3.3	O	W2 (bridging A17 and U28)
	O	2.8-2.9	O2'	U28
Arg15	NE	2.7-2.8	O6	G7
	NH1	3.0-3.1	N7	
		3.3	OP2	G6
	NH2**	2.9	O	W3 (bridging phosphates of C5, G6 and O2' of A17)
Gly16	N	2.9	OP2	G29
Gln17	N	3.0-3.1	O2'	U28

The range of hydrogen bond distances was determined based on distances in both monomers of the asymmetric unit.

* Alternative conformation of guanidinium group of Arg8 in two monomers places either Nε or Nη1 at the distance of a hydrogen bond with a non-bridging oxygen atom of nucleotide U3.

** The water molecule W3 was placed in one monomer.

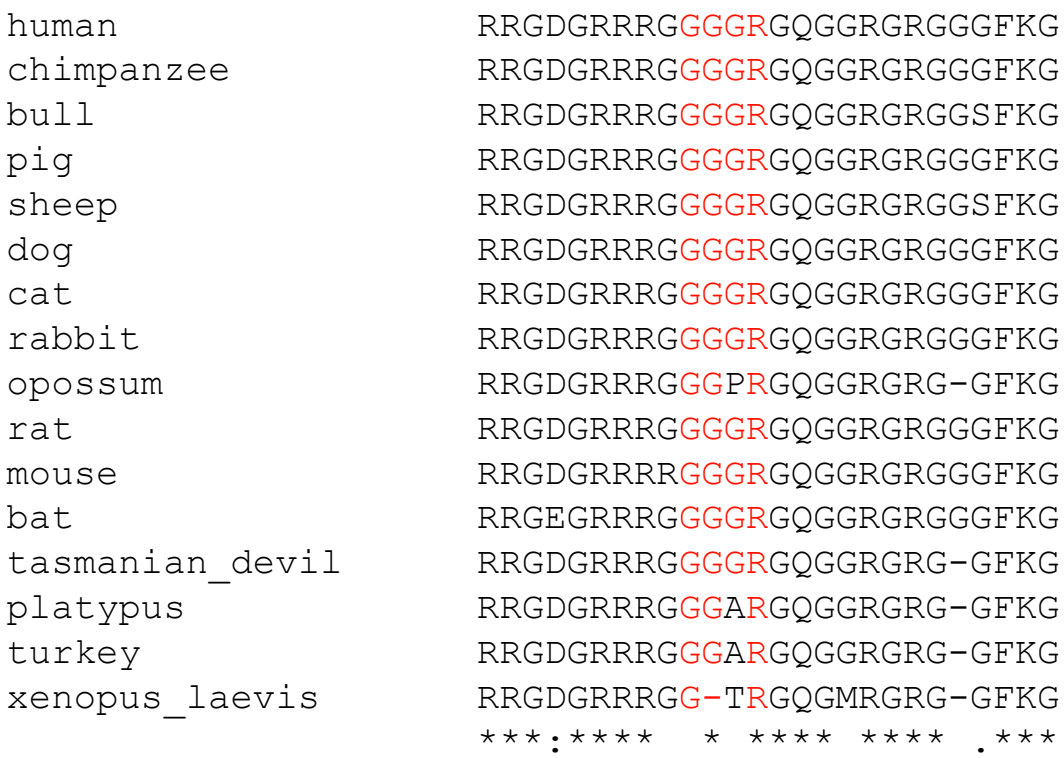

Figure S1. ClustalO alignment of RGG boxes from FMRPs. The RGG box is virtually identical in mammals and highly conserved across vertebrates. Residues in red correspond to arginines and glycines involved in the formation of a β-turn in the *sc1*-RGG complex structure.

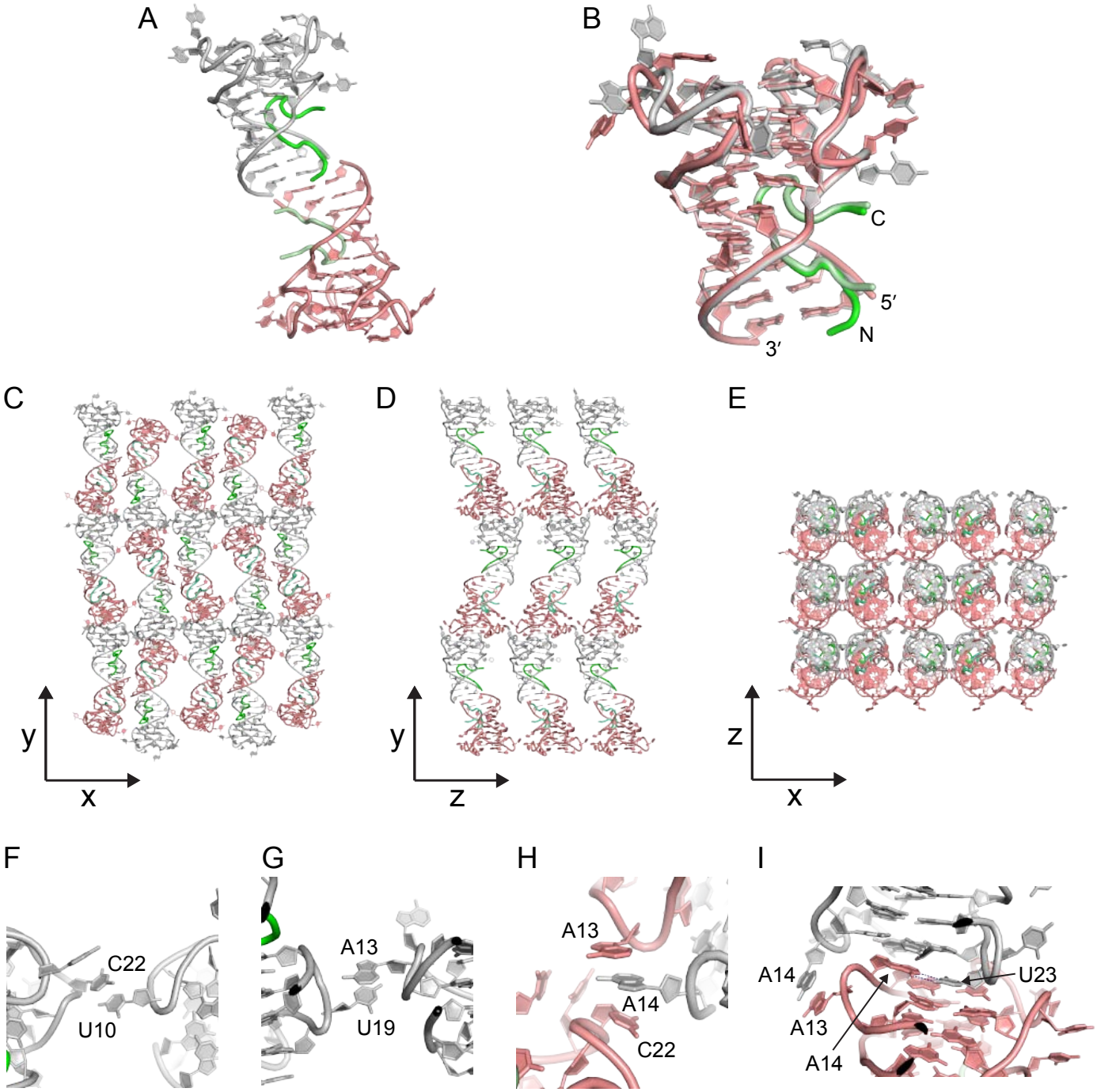

Figure S2. Asymmetric unit and crystal packing in the crystal of the RGG-*sc1* complex. Monomer 1 is in grey and green, monomer 2 is in pink and light green. (A) Asymmetric unit showing two molecules of the complex. The ends of RNA duplexes from two complexes participate in packing interactions. These close contacts position N-terminal regions of peptides in proximity to each other and orient them towards the groove of the neighboring RNA molecule. Removal of N-terminal amino acids of the RGG motif and the R3A mutation prevent steric clashes and unfavorable electrostatic interactions and significantly improves the quality of the crystals. (B) All-atom superposition of the complexes from the asymmetric unit. (C-E) Packing of the complexes in the crystal shown in three orientations. (F-I) Zoomed-in views of crystal packing interactions highlighting different conformations of the RNA loops.

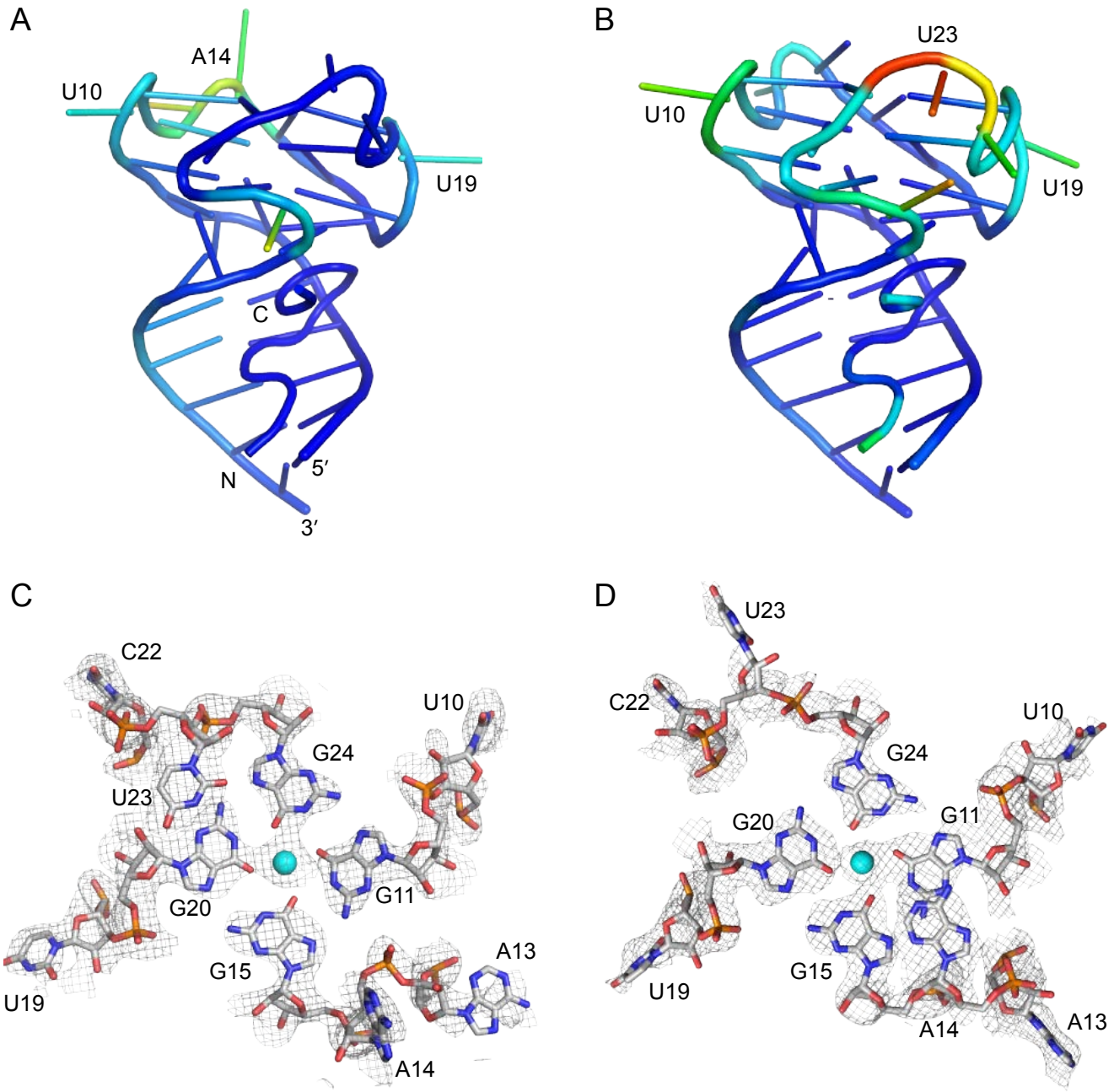

Figure S3. Conformations of loops in the crystal of the RGG-*sc1* complex. (A, B) Structure of the complex colored according with crystallographic B-factors from low (blue) to high (orange) values. Monomer 1 and 2 are shown in panels A and B, respectively. (C, D) Simulated annealing omit $2F_o$-F_c electron density map contoured at 1 σ level shown with the refined model of the complex. Monomers 1 and 2 are shown in panels C and D, respectively.

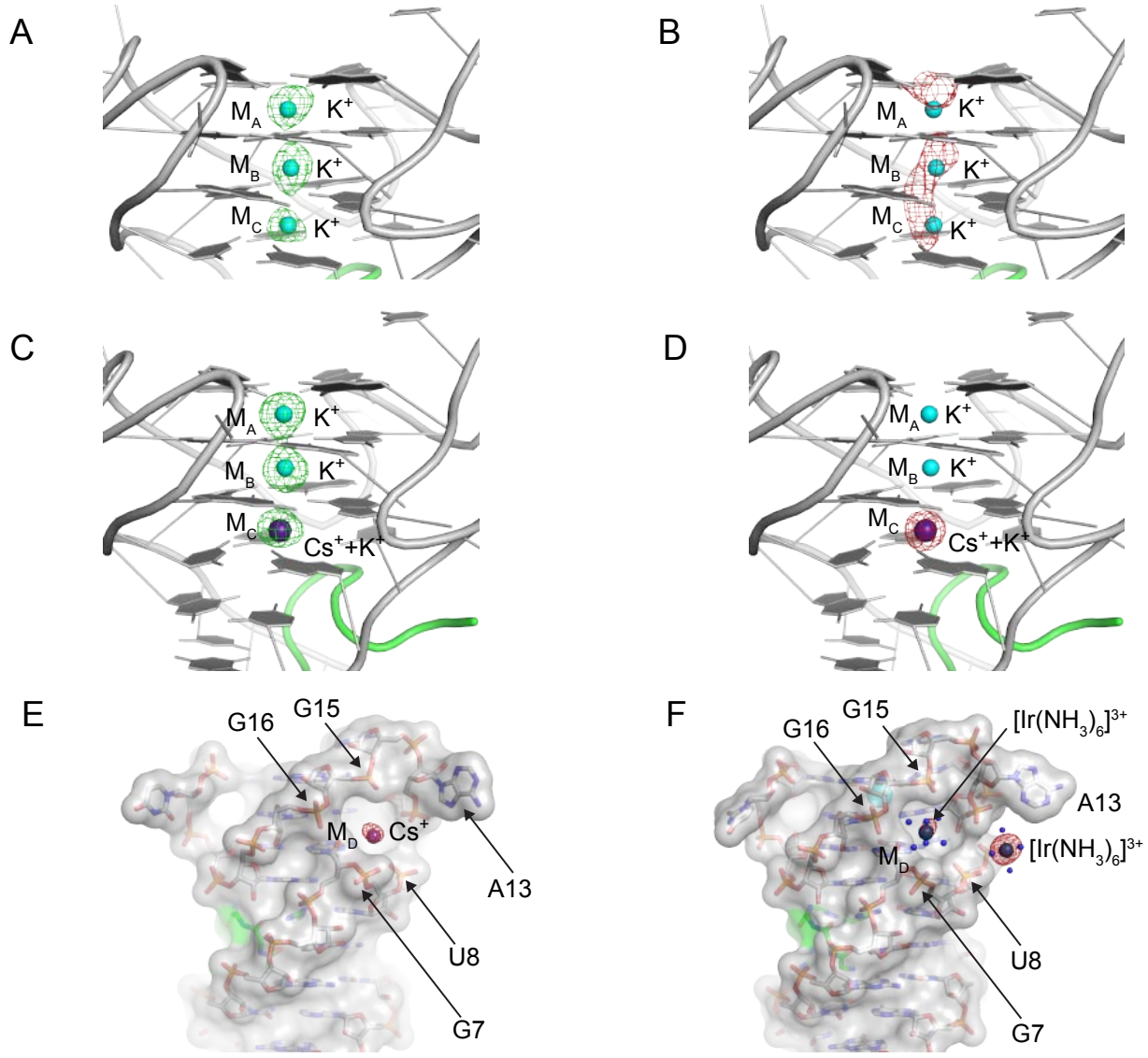

Figure S4. Cation binding sites. (A) Native structure shown with simulated annealing omit F_o-F_c map (green mesh) contoured at 5 σ level around cations. (B) Native structure shown with anomalous difference map contoured at 2.5 σ level (red mesh) around K^+ cations. This image shows similar anomalous signal in crystals with K^+ in M_A, M_B, and M_C sites. (C) Cs^+ structure shown with a map prepared as in panel A. (D) Cs^+ structure shown with anomalous map contoured at 5 σ level (red mesh) around cations. This image shows that in the crystal grown in the presence of both K^+ and Cs^+, the M_C site demonstrates an increase in the anomalous signal. To quantify this incerase, we performed following calculations. Because of difficulties in obtaing diffracting crystals, we did not collect native and Cs^+ data at the same wave length and had to compare data sets collected at 1.11 and 1.49 Å wave lengths, respectively. The anomalous signal of K^+ is 1.73 fold stronger at the second wave length. In agreement with the prediction, we observed ~1.8 fold incerase in anomalous signal in sites M_A and M_B in the Cs^+ structure, indicative of K^+ cations in these sites with almost the same occupancy as in the native structure. The anomalous signal was however 3.3 fold stronger in the M_c site. This value is 1.83 fold higher than expected for a K^+ cation, suggestive of a partial substition of K^+ by Cs^+ in the M_c site. Similar calculations and corrections revealed ~1.6 and ~1.4 fold gains in the amomalous signal in the M_c site of the second monomer and in the M_D site, suggestive of a partial substition of K^+ by Cs^+ in these two sites as well. (E) Overall structure of RNA and peptide in semi-transparent surface representation shown with a Cs^+ cation (purple) in the M_D site and anomalous difference electron density map (red) contoured at 4 σ level. Crystals were grown in the presence of both K+ and Cs^+ cations. (F) Similar view of the structure from the crystals soaked in $[Ir(NH_3)_6]^{3+}$ cations. Cations are in blue spheres. Anomalous map (red) is contoured at 8 σ level.

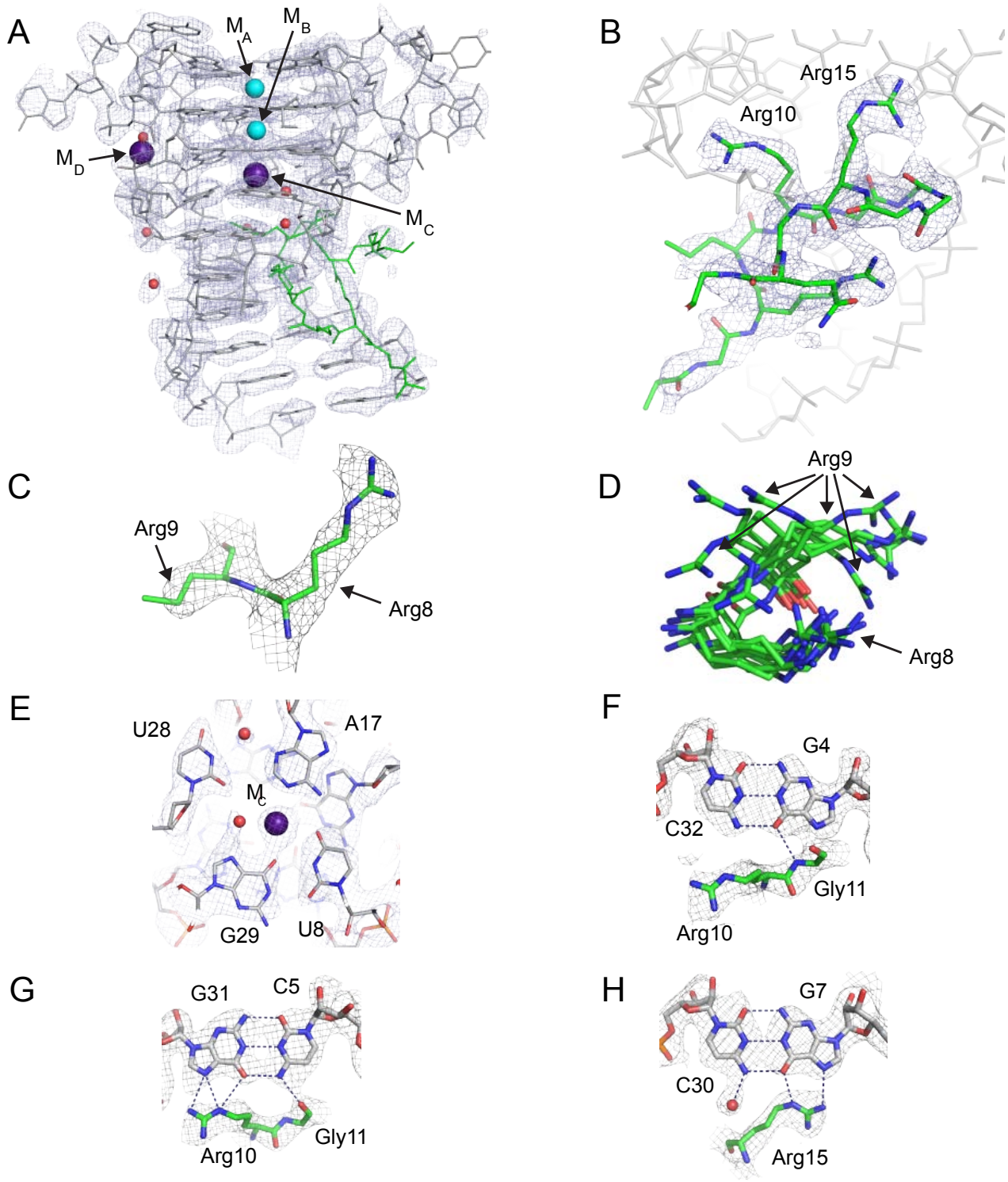

Figure S5. Simulated annealing composite omit $2F_o$-F_c electron density maps contoured at 1 σ level and shown with the refined model of the RGG-*sc1* complex structure derived from the "CsCl crystal" data. (A) Overall structure with RNA in gray, peptide in green, K^+ cations in cyan and Cs^+ cations in purple. (B) Zoomed-in view of the peptide. (C) Zoomed-in view of Arg8 and Arg9 of the RGG peptide. The side chain of Arg9 is oriented towards solution and is partially disordered as evidenced by the missing density for the guanidinium group. (D) Zoomed-in view showing the conformational enemble of Arg8 and Arg9 in the NMR structure of the *sc1*-RGG complex [37]. Note that different position of Arg8 in the NMR structure was defined by several intermolecular NOEs between the side chain of Arg8 and the sugar of G1 and the base of C2. Arg9 did not have intermolecular NOEs and its conformation was not unamigiously determined. (E) Zoomed-in view of the mixed base tetrad. (F-H) Views of the RNA-peptide interactions.

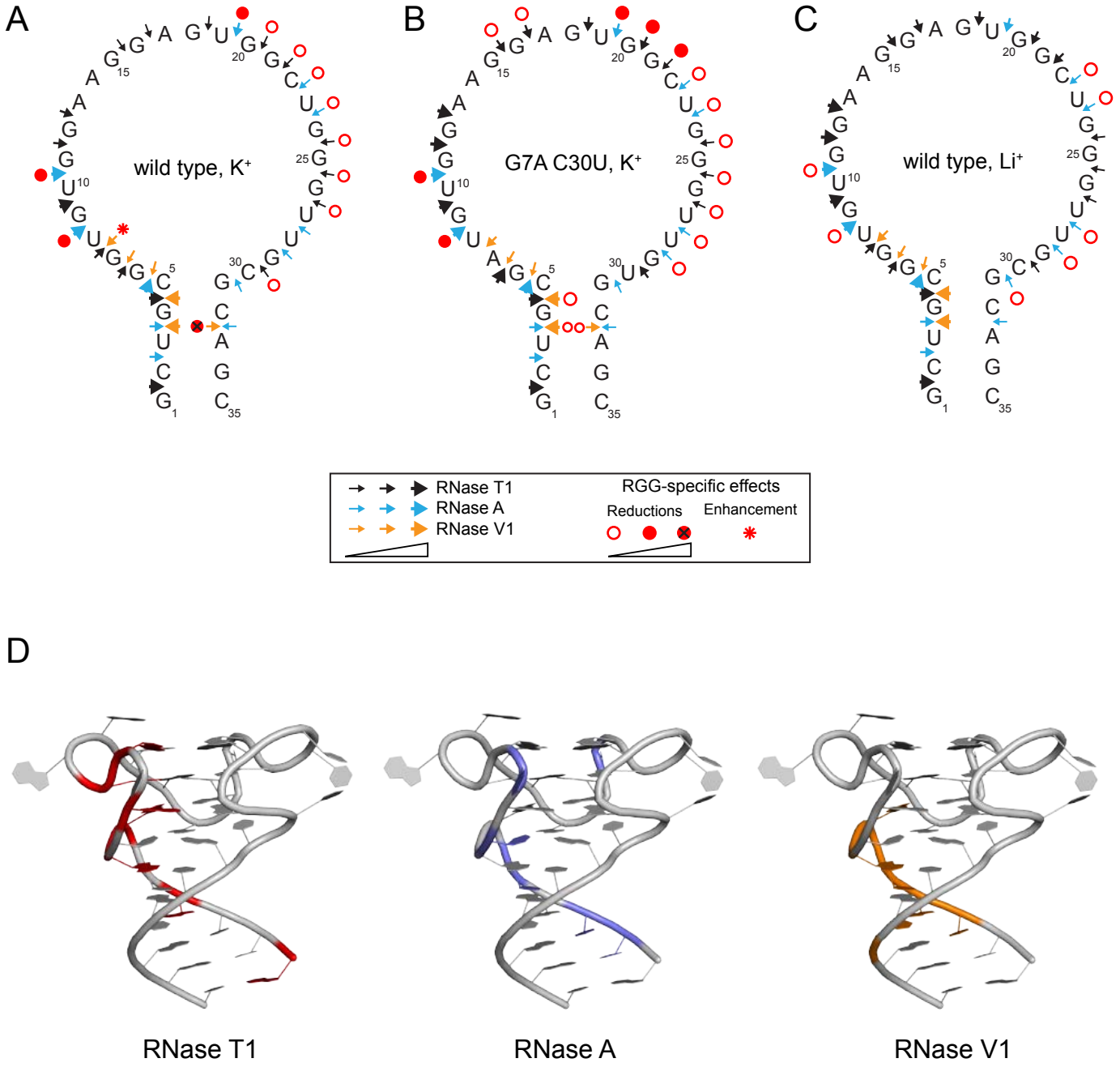

Figure S6. Summary of structure probing experiments. A-C, Nuclease cleavages and protections by the RGG peptide are shown on schematics of the wild type *sc1* RNA in the presence of K^+ (A), the G7A-C30U mutant in the presence of K^+ (B), and the wild type *sc1* RNA in the presence of Li^+ (C). (D) Projections of major and moderate nuclease cleavages on the X-ray structure of *sc1* RNA from the RGG-*sc1* complex. Cleavages by RNases T1, A, and V1 are shown in red, blue and orange colors in the left, middle and right panels, respectively. Note that *sc1* RNA likely adopts dynamic conformations in the free state and does not stably form the conformation observed in the crystal structure of the complex. Therefore, observed nuclease cleavage patterns should be compared with the three-dimensional structure of the RNA with caution.

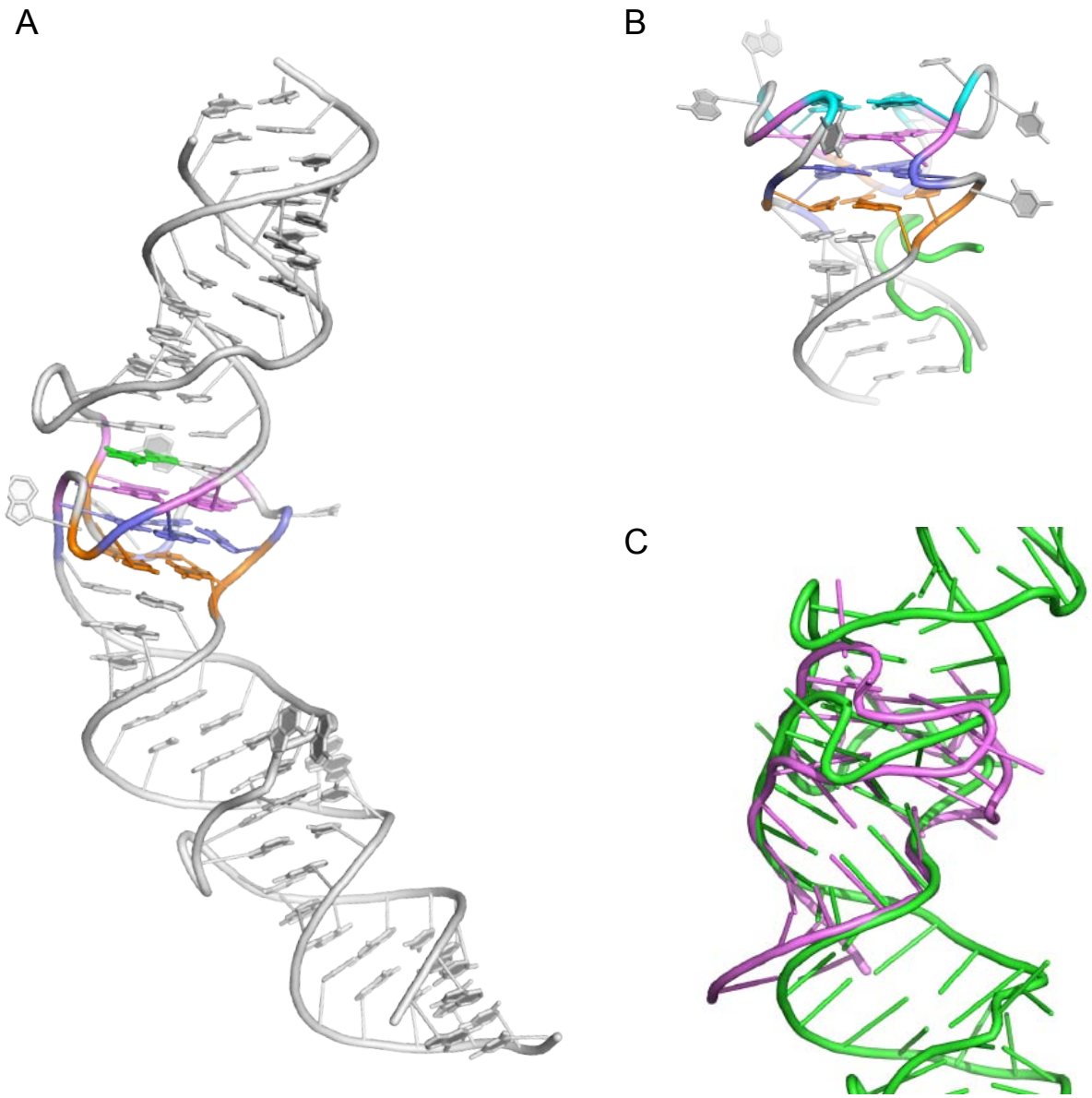

Figure S7. Comparison of G-quadruplexes from the structures of *sc1*-RGG complex (current study) and Spinach RNA (PDB ID 4TS2). (A) Spinach RNA in cartoon representation. Mixed tetrad is in orange, G-quartets are in blue and light magenta. Bound ligand is in green. (B) Structure of the *sc1*-RGG complex. Colored regions correspond to the regions shown in panel A. (C) Manual superposition of Spinach RNA (green) and *sc1* RNA (light magenta) on the blue G-quartet. Superposition highlights different conformations of G-quadruplexes adopted in the two structures.