

Supporting Information

Stable and Conserved G-Quadruplexes in the Long Terminal Repeat Promoter of

Retroviruses

Emanuela Ruggiero[†], Martina Tassinari[†], Rosalba Perrone[‡], Matteo Nadai[†] and Sara N. Richter^{†,*}

[†]*Department of Molecular Medicine, University of Padua, via Aristide Gabelli 63, 35121 Padua, Italy*

[‡]*Buck Institute for Research on Aging, 8001 Redwood Boulevard, Novato, California 94945, United States*

Content:

Table S1	page S2
Table S2	page S3
Table S3	page S6
Table S4	page S8
Figure S1	page S9
Figure S2	page S10
Figure S3	page S11
Figure S4	page S12
Figure S5	page S13

Table S1. List of all RVs analyzed in this study, with the corresponding acronyms

	Virus	Acronym	Virus	Acronym
<i>alpha-RVs</i>	Avian leukosis virus	ALV-A	Rous sarcoma virus	RSV
	Avian leukemia virus	ALV-J	Avian myelocytomatosis virus	AMV
	Y73 sarcoma virus	YAV	Fujinami sarcoma virus	FSV
	UR2 sarcoma virus	UR2SV	Avian carcinoma virus	ACV
<i>beta-RVs</i>	Bovine retrovirus CH15	BoRV-CH15	Mason-Pfizer monkey virus	MPMV
	Desmodus rotundus endogenous retrovirus	DrERV	Jaagsiekte sheep retrovirus	JSRV
	Ovine enzootic nasal tumor virus	ENTV-1	Mouse mammary tumor virus	MMTV
	Enzootic nasal tumour virus of goats	ENTV-2	Simian retrovirus 4	SRV-4
	Squirrel monkey retrovirus	SMRV		
<i>delta-RVs</i>	Simian T-lymphotropic virus 1	STLV-1	Human T-lymphotropic virus 1	HTLV-1
	Simian T-lymphotropic virus 2	STLV-2	Human T-lymphotropic virus 2	HTLV-2
	Simian T-lymphotropic virus 3	STLV-3	Bovine leukemia virus	BLV
<i>gamma-RVs</i>	Friend murine leukemia virus	F-MuLV	Moloney murine sarcoma virus	MuSV
	Moloney murine leukemia virus	Mo-MLV	Feline leukemia virus	FeLV
	Baboon endogenous virus	BaEV	Gibbon ape leukemia virus	GALV
	Mus musculus mobilized endogenous polytropic provirus	Mus	Murine osteosarcoma virus	MSV
	Koala retrovirus	KoRV	Abelson murine leukemia virus	A-MuLV
	RD114 retrovirus	RD-114	Spleen focus-forming virus	SFFV
	Woolly monkey sarcoma virus	WMSV	Porcine endogenous retrovirus E	PERV-E
	Reticuloendotheliosis virus	REV		
<i>epsilon-RVs</i>	Walleye dermal sarcoma virus	WDSV	Walleye epidermal hyperplasia virus 2	WEHV-2
	Walleye epidermal hyperplasia virus 1	WEHV-1		
<i>spuma-RVs</i>	Human spumaretrovirus (or human foamy virus)	HFV	Feline foamy virus	FFV
	Simian foamy virus	SFV	Equine foamy virus	EFV
	Macaque simian foamy virus	M-SFV	Bovine foamy virus	BFV
	African green monkey simian foamy virus	SFV-3		

Table S2. PQS analysis performed with QuadBase2 within the LTR region of RVs. G₃ tracts are reported in red and bold, non-overlapping bulged G₃ tracts (e.g. GGxG) are reported in blue and bold, overlapping bulged G₃ tracts (eg.GXGGG) are underlined.

	Virus	Sequence	Strand	
<i>alfa-RVs</i>	ALV-A	<i>NF</i> ^a		
	ALV-J	GGTGCACACCAATGTGGTGGGAGGTAAAATGGCG	-	
	YAV	<i>NF</i>		
	UR2SV	GGTGCACACCAATGTGGTGAGTGGTAAAATGGCG	-	
	RSV	<i>NF</i>		
	AMV	<i>NF</i>		
	FSV	<i>NF</i>		
	ACV	GGTGGAAGTAAGGTGGTATGATCGTGGTATGATCGTGG	+	
<i>beta-RVs</i>	BoRV-CH15	<i>NF</i>		
	DrERV	GGTGAGGCGACTGCATGGGGGAGCTTGAAAGGGCAGCGCTGCACTGCGG GGGCAGCGCTGCACTGCGGAGGAGGGGTGAGGAGGG	- -	
	ENTV-1	<u>GCGGGGACAACCTGCGGAGGGTTAAGTCCTGGGAG</u>	+	
	SMRV	GGCGTGGTGGCGGCCACCAATGGAGGACCTGATCACGGG GGGTTTCCTTATATAGGGAGGGGAGAGGGTAGAGAGGGG GTGGATGAGAAAAGGATGGGACAAGAGGCG	+ - -	
	ENTV-2	<i>NF</i>		
	MPMV	GGAGGAGGGAGTGGGAATTGAAAGG	-	
	JSRV	<i>NF</i>		
	MMTV	GGGGCTATTGGGGGAAGTTGCGGTTTCGTGCTCGCAGGG GCGGCTTGTAAGAGGAAGTTGGCTGTGGTCCTTGCCTCAGGAGG GGAGACAGGTGGTGGCAACCAGGGACTTATAGGGG <u>GAGGGTCACCGGGGTCTGCGGGGGG</u>	+ - + -	
	SRV-4	<u>GGCGGGAAGGAAAAGGAAACGTCA</u> GCGGACGCTGGG	-	
	<i>delta-RVs</i>	STLV-1	GGGAGGAAATGGGTGGTTCCCGGTGGCTTCCCGGAGG GGAGTCTATAAAAGCGTGGAGACAGTTCAGGAGGGG CGGCGTGGATGGCGGCCTCAGGTAGGGCGGC <u>GGGCGCGTAGGAG</u> GGCGGACGCAGTTCAGGAGGCACCACAGGC <u>GGGAGGCGG</u>	- + - -
		STLV-2	GGGGAAGTGGGTAAGGGCGAGGTTATTTTTCTGGAG GGGCCAGTGGTGCAGGGAGGGG GGGTGTTTTGGGCCTCTCCGGGAGGGG	- - +
		STLV-3	<i>NF</i>	

gamma-RVs	HTLV-1	GGGAGGAAATGGG <u>TGGT</u> CCCGGTGGCTTCCCGAGG GGAG TCTATAAAAGCGT GGGG ACAGTTCAGGAGGGG GGCG TGGAT GGCG GCCTCAGGTAAG GGCG <u>GC</u> GGG CGCGTGAAGGAG GGCG GACGTAGTTCAGGAGGCACCACAGGC GG AG GGCG	- + - -
	HTLV-2	GGGGG AGGGACGTCA GGG CCGTGG GGG GAAGT GGG TAA GGG TGAGG GGCG TCTCAACCGAGGCCGCTT GGAG AGGATGGATCCCCAAGGTG GAGG AGACTTACCTT GGG AGGAGCAGTTCAGGAGGTGCCGAGGCCGGG GGCG TATGTTGGCCCC GGG ACGCCTGCCAAGA GGG CGGCCGTTGAGG GGCG TCCC GGGG CCAACATACGCC GTGG AGCGCAGCAA GGG CTAGGG	- - - - - +
	BLV	GGGTGTGG ATTTTT GGG AAAGGGGAAGTT GGGG AGGTAC GGGG GTGG AGACGGTCA GAGG ACCG GAG ATAGAGCTCGCG GGCG	- -
	F-MuLV	GGCG ACTCAGTCTATCG GAG ACT GGCG CGCCGAGTGA GGGG	-
	Mo-MLV	GGAG ACT GGCG CCCCGAGTGA GGGG TGT GGG GGCG ACTCAGTCAATCG GAG ACT GGCG CTGCAACTGCAAG AGG GGGGG TCTTTCATTT GGGGG CTCGTCC GGG ATCGGG GGTGGTGGG TC GGT GTCCT GGG CAGGGG GGG ACGTCTCCCA GGG TT GGCG CC GGG TG GGG AGACGTCCCA GGG ACTTCC GGGGG CCGTTTTT GTGG	- - + - - +
	BaEV	GGGTCTGGG TTGCAG GGG TCGGG GGGGTGGG ATAG GTG CTAGCCCC GGGG AGGTCT GGGG	- -
	Mus	GGG ACAG GGGG CCAAATATCG GTG GTCAAGCACCT GGG GGGTATGGG <u>AGGGG</u> TAC GAGG AAA GGG GGG CT GGGG CT GGGG AGCAAAAAG GGCGG	+ - -
	KoRV	NF	
	RD-114	GCGG TTA GGG T GGG CGCATAAG GGG GCGG CT GGGG ACTTTCGGGCTA GGGTGGGG CGCATAAG GGG GGG TTGCGAAG GGG CTGATGCAACTC GGGG CCCGGG GGTGGTTGGGG TTGT GGG TAATTTTCGTCCC GGGG AAGCTT GGG	- - - -
	WMSV	GGCG ACTCAGTCTCTCG GAG ACT GGCG CGCCGAGT GGAG	-
	REV	<u>GTGGG</u> AGGG AGCTCC GGGGGGGGG	-
	MuSV	GAGG CTTTATT GGG AATAC GGG TACCC GGGCG GGGG GTCTTTCATTT GGGG GCTCGTCC GGG ATT GGAG GGTGGTGGG TC GGT GTCCT GGG CAGGGG	- + -

		GGG ACGTCTCCCA GGG TT GCGG CC GGG TG	-
	FeLV	GTGG TCTC GGT GTTCTGT GGG CG GGGG TCTCATCGCC GAGG	+
	GALV	<i>NF</i>	
	MSV	GGAG GTAAGAG GAGG CTGGAAAGTACC GGG ACTA GGG	+
		GGCG ACTCAGTCTATC GGAG GACT GGCG CGCCGAGTGT GGGG	-
	A-MuLV	GGAG GACT GGCG CGCCGAGTGA GGGG TTGT GGG	-
	SFFV	GGCG ACTCAGTCTGT GGAG GACT GGCG CGCCGAGTGA GGGG	-
	PERV-E	<i>NF</i>	
<i>epsilon-RVs</i>	WDSV	<i>NF</i>	
	WEHV-1	<i>NF</i>	
	WEHV-2	<i>NF</i>	
<i>spuma-RVs</i>	HSRV	GAGG AGCAG GAG TATTT GGG TCAC GAGG	-
	SFV	GAGG AGCT GGAG TATTT GGG TAGT GAGG	-
	M-SFV	<i>NF</i>	
	SFV-3	<i>NF</i>	
	FFV	<i>NF</i>	
	EFV	GAGG ATCGTCATCAG GAGG AATCTTCGCCA GGG T GAGG	-
	BFV	GTGG TAAACATAA GGG TCCTCACGACAGT GGAGAGGTG	-
	FFV	<i>NF</i>	

^a*NF*: Not Found.

Table S3. Sequences used for base conservation analysis of RVs PQS. For each analyzed virus, accession numbers of the different strains are provided, with the reference strains reported in bold.

	Virus	Accession number	Virus	Accession number
<i>beta-RVs</i>	DrERV	NC_027117.1 KP175582.1 KP175581.1	MPMV	NC_001550.1 M12349.1 AF033815.1 M11841.1
	ENTV-1	NC_007015.1 GU292318.1 GU292317.1 GU292316.1 GU292315.1 GU292314.1 FJ744149.1 FJ744148.1 FJ744147.1 FJ744146.1 KC189895.1 FJ744150.1	MMTV	NC_001503.1 AF228552.1 AF228551.1 AF228550.1 M15122.1 AF033807.1 L37517.1
	SMRV	NC_001514 M23385.1	SRV-4	NC_014474.1 AB920340.1 FJ979639.1 FJ979638.1 FJ971077.1 AB920339.1
<i>delta-RVs</i>	STLV-2	NC_001815.1 U90557.1 Y14570.1	BLV	LC080652.1 LC080651.1 LC005616.1 LC005615.1 AB934283.1 AB934282.1 HE967303.1 HE967301.1 HE967302.1 AB987702.1 AF257515.1 FJ914764.1 KT122858.1 AF033818.1 AP018032.1 AP018031.1 AP018030.1 AP018029.1 AP018028.1 AP018027.1 AP018026.1 AP018025.1 AP018024.1 AP018023.1 AP018022.1 AP018021.1 AP018020.1 AP018019.1
	HTLV-2	NC_001488.1 AF412314.1 M10060.1 GU212854.1 AF074965.1 AF139382.1 AF326584.1 L11456.1 AF326583.1 Y14365.1 Y13051.1		
	BLV	NC_001414.1 EF600696.1 LC154849.1 LC154848.1 K02120.1 LC080675.1 LC080674.1 LC080673.1 LC080672.1 LC080671.1 LC080670.1 LC080669.1 LC080668.1 LC080667.1		

		LC080666.1 LC080665.1 LC080664.1 LC080663.1 LC080662.1 LC080661.1 LC080660.1 LC080659.1 LC080658.1 LC080657.1 LC080656.1 LC080655.1 LC080654.1 LC080653.1		AP018018.1 AP018017.1 AP018016.1 AP018015.1 AP018014.1 AP018013.1 AP018012.1 AP018011.1 AP018010.1 AP018009.1 AP018008.1 AP018007.1 AP018006.1
<i>gamma-RVs</i>	BaEV	NC_001815.1 D10032.1 AB979448.1 X05470.1	RD-114	NC_009889.1 AB705393.1 AB559882.1 EU030001.1 AB705392.1
	Mus	NC_029853.1 FJ544577.2 FJ544578.2 FJ544576.1	MuSV	NC_001488.1 J02266.1 AF033813.1
	REV	NC_006934.1 DQ387450.1 DQ003591.1 AY842951.1 KJ909531.1 KJ909530.1 KJ756349.1 KF709431.1 JX912710.1 GQ415646.2 FJ439120.1 FJ439119.1 KU204703.1 KU204702.1 GQ375848.1 JQ804915.1 KX278301.1 KU641115.1 KF305089.1 FJ496333.1	MoMLV	NC_001501.1 AF462057.1 AF033811.1 M63653.1

Table S4. Oligonucleotides used in the Taq polymerase stop assay

	Sequence
RV1	GGGCAGCGCTGCACTGCGGAGGAGGGGTGAGGAGGGCTGCATATAAGCAGCTGCTTTTTGCC
RV5	GGAGGAGGGAGTGGGAATTGAAGGGCTGCATATAAGCAGCTGCTTTTTGCC
RV7	GAGGGTCACCGGGGTCTGCGGGGGGCTGCATATAAGCAGCTGCTTTTTGCC
RV9	GGGCCAGTGGTGCAGGGAGGGGCTGCATATAAGCAGCTGCTTTTTGCC
RV12	GGGGAAGTGGGTAAGGGTGAGGCTGCATATAAGCAGCTGCTTTTTGCC
RV18	GGGTCTGGGTTGCAGCGGTCGGGCTGCATATAAGCAGCTGCTTTTTGCC
RV22	GGGCTGGGGCTGGGGAGCAAAAAGCGGGCTGCATATAAGCAGCTGCTTTTTGCC
RV26	GTGGGAGGGAGCTCCGGGGGGGGCTGCATATAAGCAGCTGCTTTTTGCC
non-G4 cnt	TTGTCGTAAAGTCTGACTGCGAGCTCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCC
Primer	GGCAAAAAGCAGCTGCTTATATGCAG

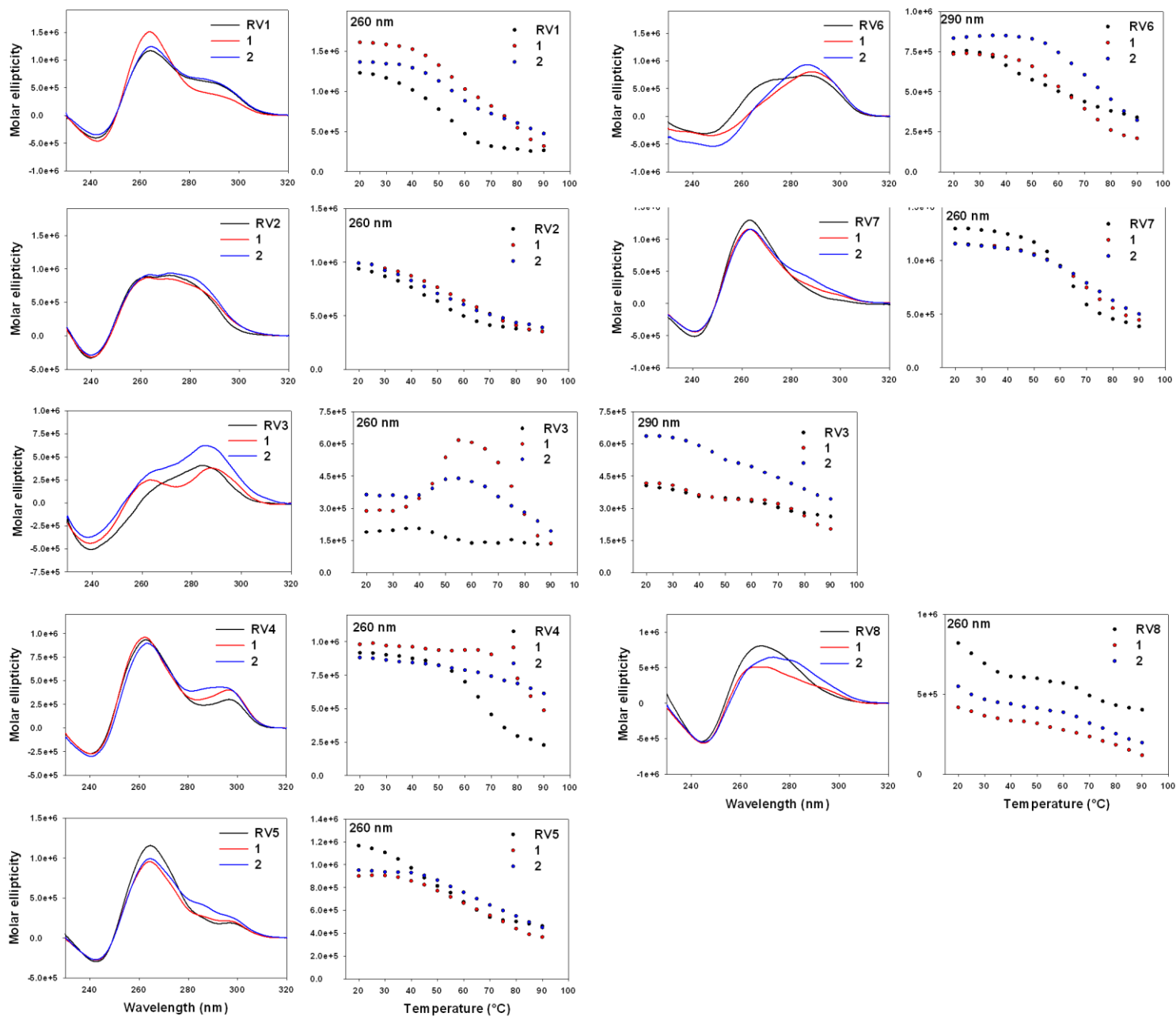


Figure S1. CD spectra of beta-RV G4 sequences in the absence or presence of G4 ligands **1** and **2**, and plots of molar ellipticity vs temperature, measured at the indicated wavelengths.

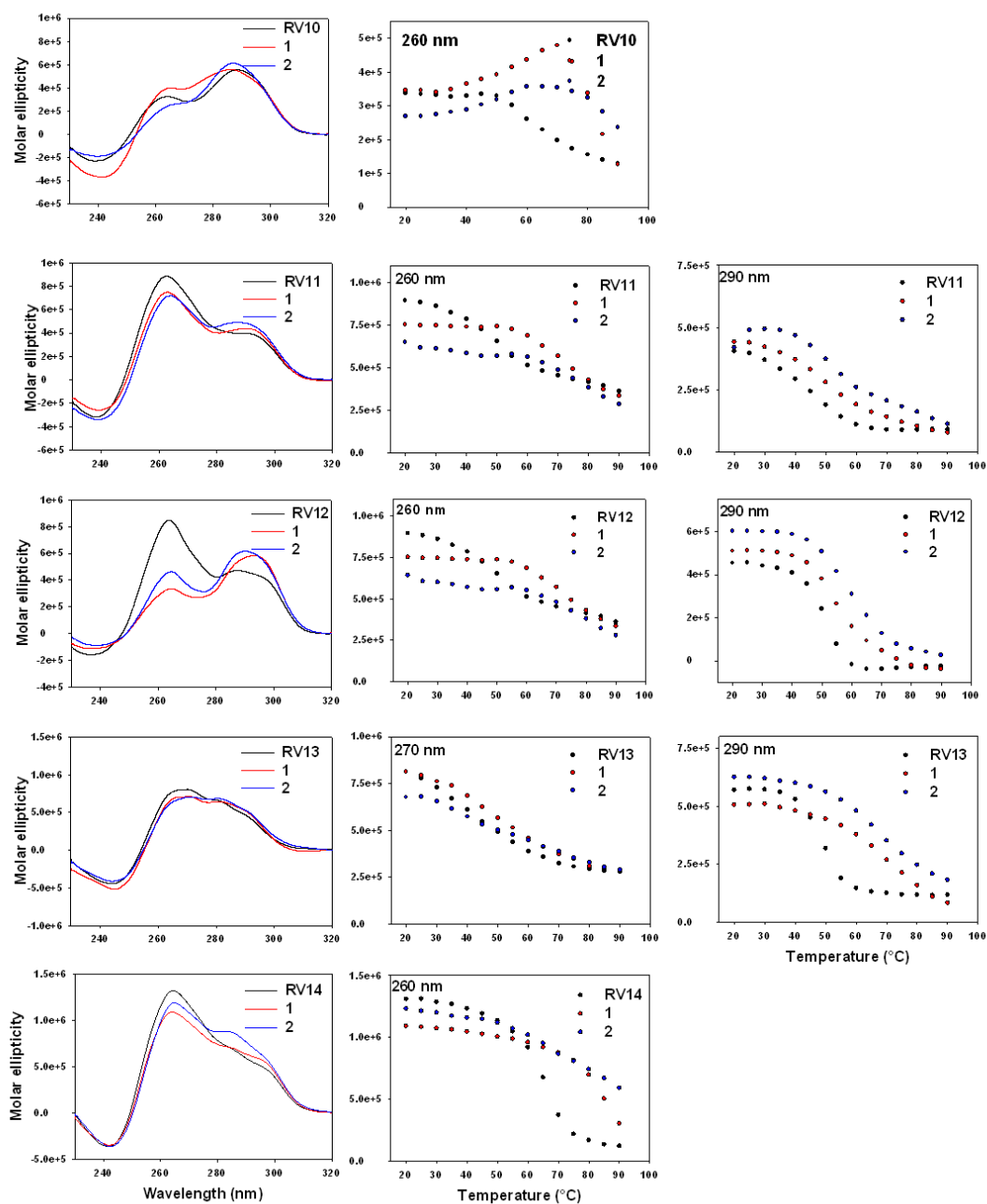


Figure S2. CD spectra of delta-RV G4 sequences in the absence or presence of G4 ligands **1** and **2**, and plots of molar ellipticity vs temperature, measured at the indicated wavelength.

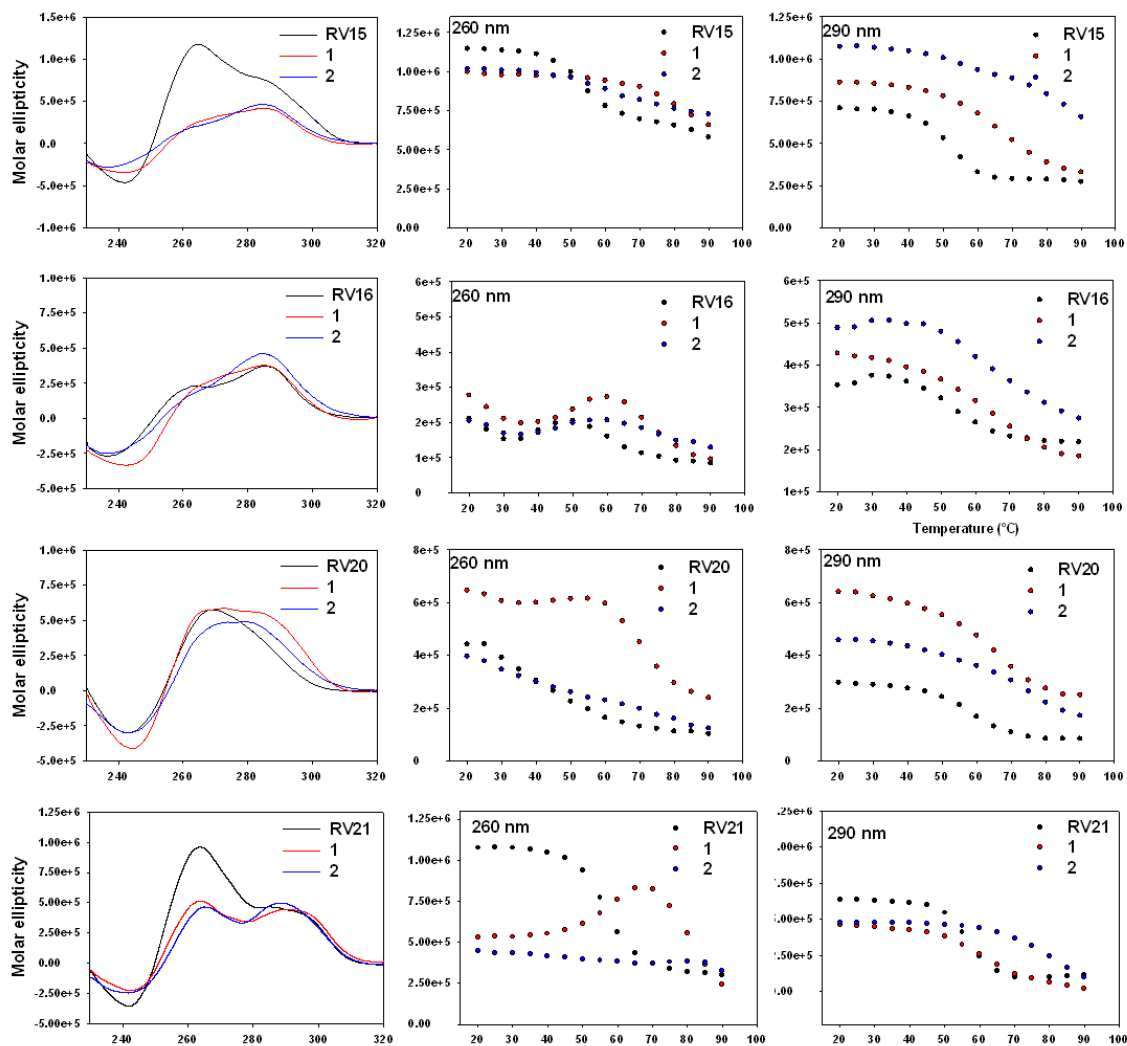


Figure S3A. CD spectra of gamma-RV G4 sequences with two transitions, performed in the absence or presence of G4 ligands **1** and **2**, and plots of molar ellipticity vs temperature, measured at the indicated wavelength.

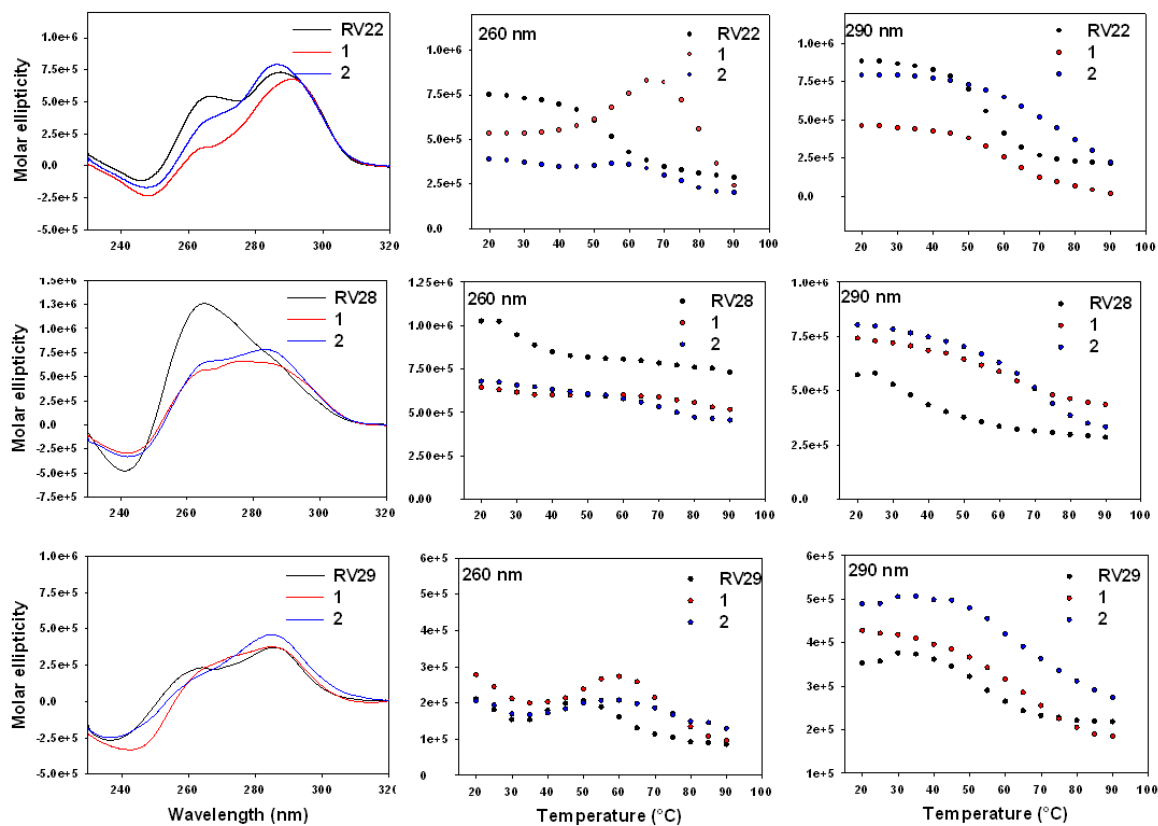


Figure S3B. CD spectra of gamma-RV G4 sequences with two transitions, performed in the absence or presence of G4 ligands **1** and **2**, and plots of molar ellipticity vs temperature, measured at the indicated wavelength.

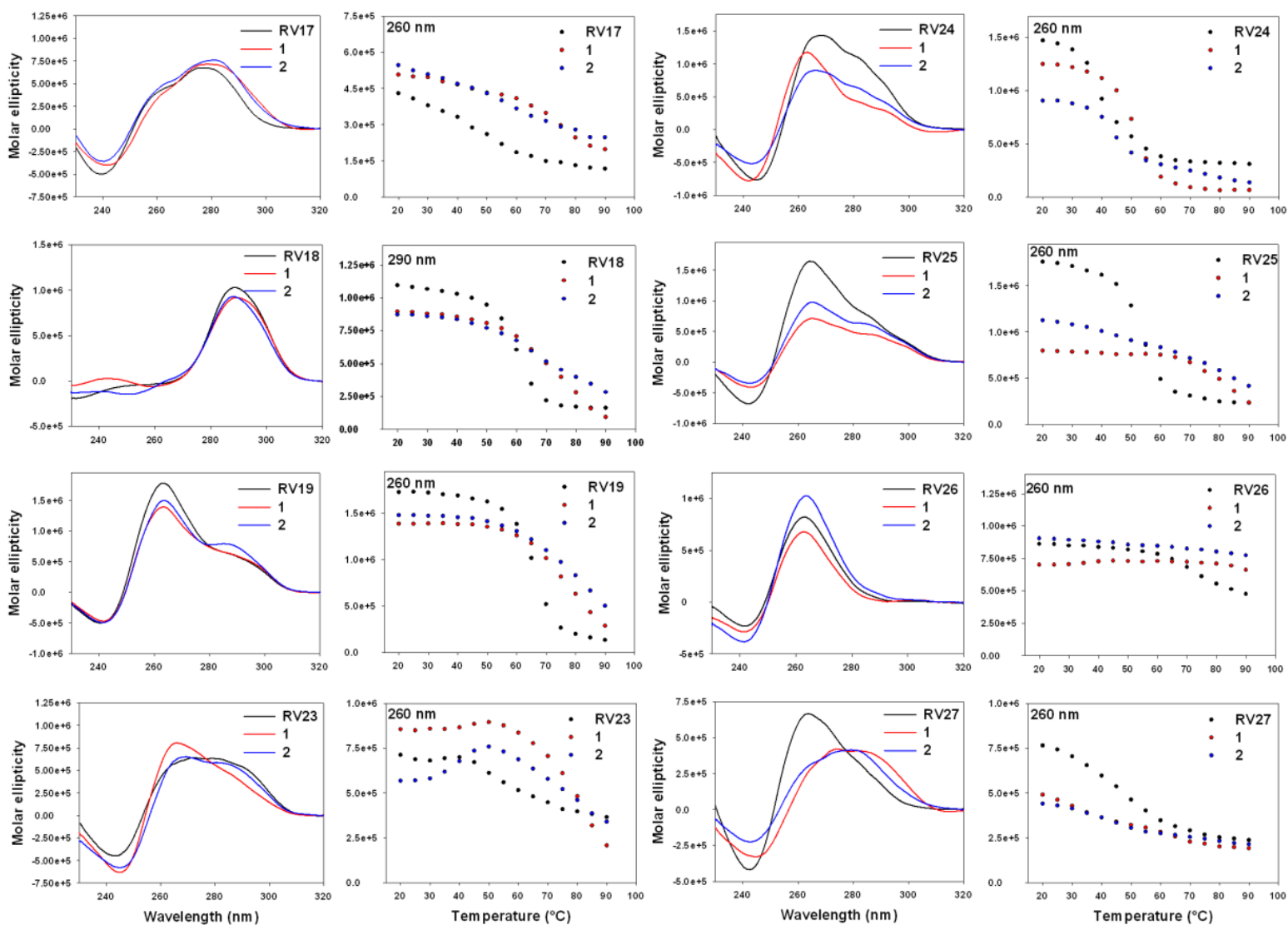


Figure S4. CD spectra of gamma-RV G4 sequences with single transition, performed in the absence or presence of G4 ligands 1 and 2, and plots of molar ellipticity vs temperature, measured at the indicated wavelength.

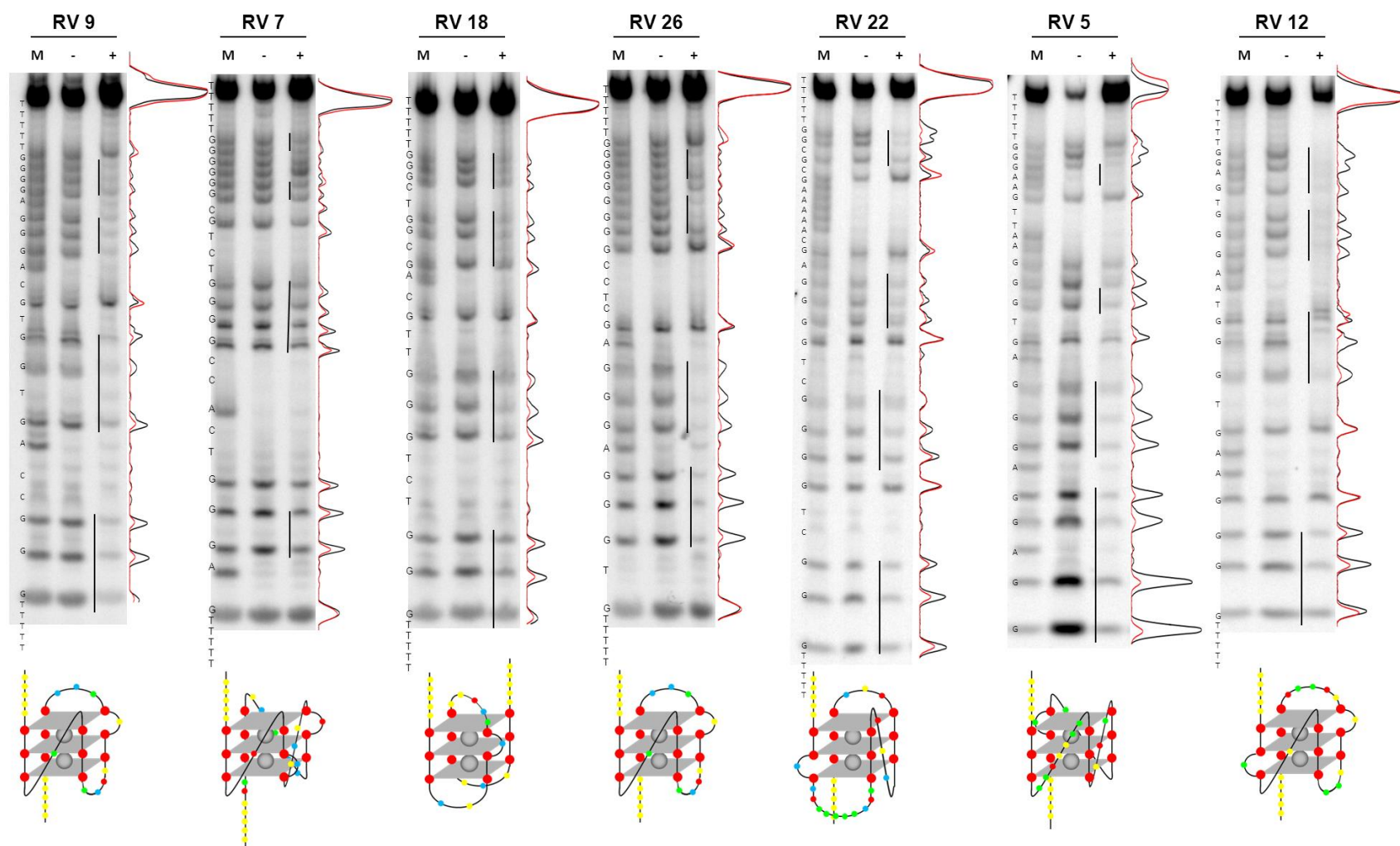


Figure S5. DMS protection analysis of RV G4s folded in the absence (-) or presence (+) of KCl. M is a marker lane obtained with the Maxam and Gilbert sequencing protocol. Proposed G4 structure is shown below the gel image. G, T, C, and A bases are shown in red, yellow, blue, and green, respectively. The densitograms show quantification of cleaved bands intensity in the absence (black lines) and presence (red lines) of KCl.