Supporting Information

A bacterial DNA quadruplex with exceptional K^+ selectivity and unique structural polymorphism

Charlotte Rehm, Isabelle T. Holder, Andreas Groß, Filip Wojciechowski, Maximilian Urban, Malte Sinn, Malte Drescher, Jörg S. Hartig*

Department of Chemistry and Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Universitätsstr. 10, 78457 Konstanz, Germany

Content:

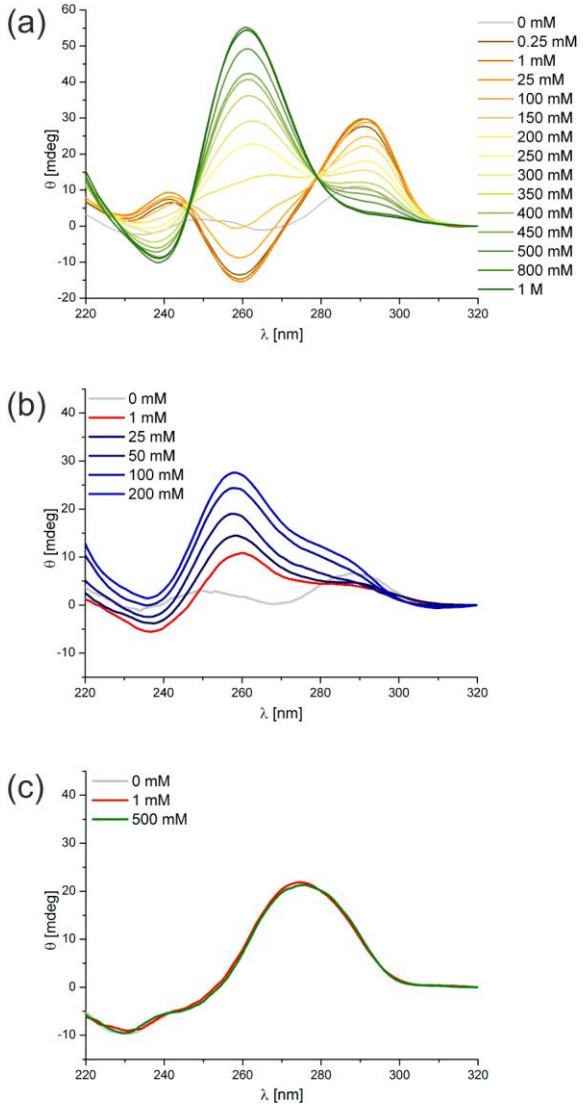


Fig. S1: Additional CD spectra of d[$(G_4CT)_3G_4$] and d[$(C_4AG)C_4$]

(a): CD spectra of 5 µM d[$(G_4CT)_3G_4$] in the presence of KCl, concentrations ranging from 0 to 1 M KCl. **(b)**: CD spectra of 5 µM d[$(G_4CT)_3G_4$] in the presence of $MgCl_2$, concentrations ranging from 0 mM (gray), 1 mM (red), 25 to 200 mM (blue) $MgCl_2$. With increasing concentrations $MgCl_2$ stabilizes a secondary structure, presumably a quadruplex with parallel topology. **(c)**: CD Spectra of 5 µM d[$(C_4AG)C_4$] in the presence of 0 mM (gray), 1 mM (red) to 500 mM (green) KCl.

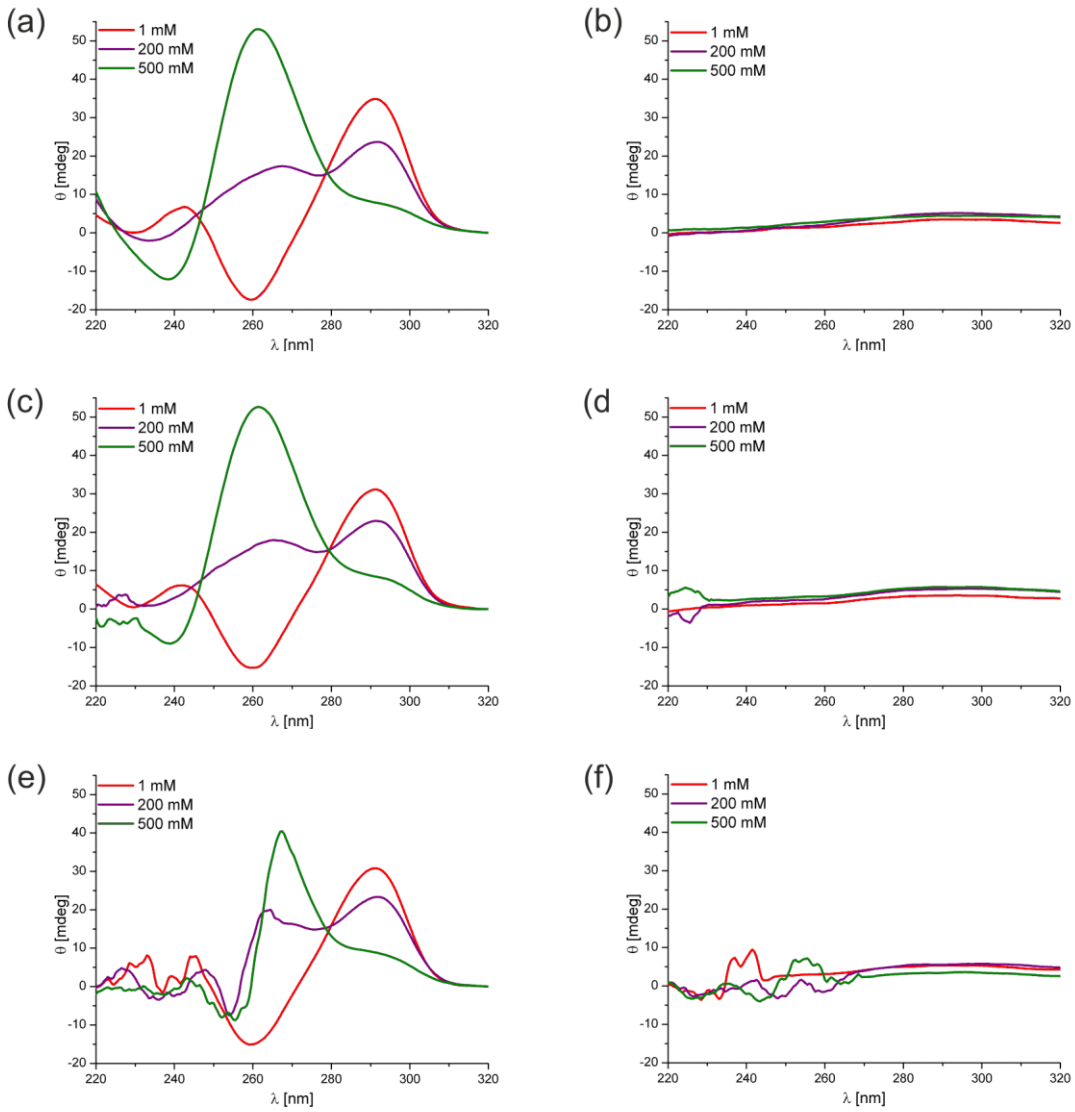


Fig. S2: Additional CD spectra of d[(G$_4$CT)$_3$G$_4$] in presence of different anions

(a): CD spectra of 5 µM d[(G$_4$CT)$_3$G$_4$] in the presence 1 mM (red), 200 mM (violet) and 500 mM (green) KF. Structural transition from the antiparallel to the parallel conformation occurs with increasing concentration of KF. **(b)**: CD spectra of buffer solution contraining 10 mM tris-HCl, pH 7.2 with 1 mM (red), 200 mM (violet) and 500 mM (green) KF. **(c)**: CD spectra of 5 µM d[(G$_4$CT)$_3$G$_4$] in the presence 1 mM (red), 200 mM (violet) and 500 mM (green) KBr. Structural transition from the antiparallel to the parallel conformation occurs with increasing concentration of KBr. **(d)**: CD spectra of buffer solution contraining 10 mM tris-HCl, pH 7.2 with 1 mM (red), 200 mM (violet) and 500 mM (green) KBr. KBr interferes with ellipticity signal below 230 nm. **(e)**: CD spectra of 5 µM d[(G$_4$CT)$_3$G$_4$] in the presence 1 mM (red), 200 mM (violet) and 500 mM (green) KI. Although KI disturbs with the recorded ellipticity, the signal at 290 nm can still be used to monitor the change in topology from the antiparallel to the parallel conformation that occurs with increasing concentration of KI. **(f)**: CD spectra of buffer solution contraining 10 mM tris-HCl, pH 7.2 with 1 mM (red), 200 mM (violet) and 500 mM (green) KI. KI interferes with ellipticity signal below 250 nm for 1 mM KI and 260 nm for 200 and 500 mM KI.

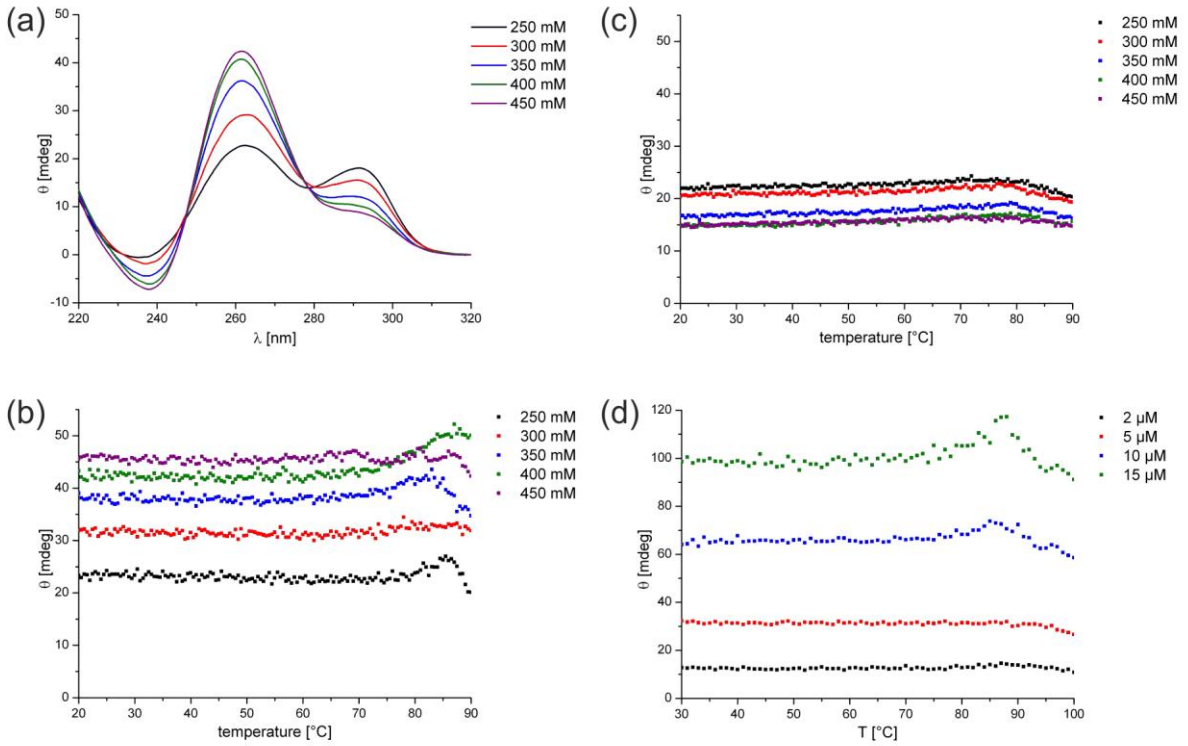
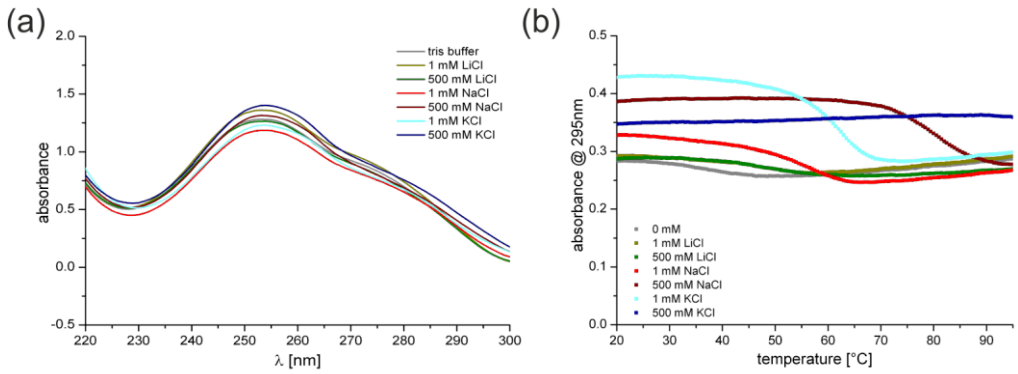


Fig. S3: **Additional CD spectra and melting profiles of d[(G$_4$CT)$_3$G$_4$] in presence of high concentrations of K$^+$**

(a): CD spectra of 5 µM d(G$_4$CT)$_3$G$_4$ in the presence of KCl, concentrations ranging from 250 mM (black), 300 mM (red), 350 mM (blue), 400 mM (green) to 450 mM (violet). A mixture of antiparallel and parallel conformations is detected at 250 mM KCl. With further increasing concentrations of KCl the remaining signal for the antiparallel conformer at 290 nm decreases, signal for the parallel conformer at 260 nm increases. **(b)**: Ellipticity at 260 nm during thermal denaturation of d(G$_4$CT)$_3$G$_4$. The signals only start to decrease above 85 °C and higher. **(c)**: Ellipticity at 290 nm during thermal denaturation of d(G$_4$CT)$_3$G$_4$. D: Ellipticity at 260 nm during thermal denaturation of the parallel conformer of d(G$_4$CT)$_3$G$_4$ in the presence of 500 mM KCl and varying oligo concentrations 2 µM (black), 5 µM (red), 10 µM (blue) and 15 µM (green). The signals only start to decrease above 90 °C and the oligo is not fully denatured.



	0 mM	1 mM LiCl	500 mM LiCl	1 mM NaCl	500 mM NaCl	1 mM KCl
$T_{1/2}$ [°C]	32.7	38.6	46.3	50.8	79.8	59.3

Fig. S4: **UV melting profiles of d[$(G_4CT)_3G_4$] in presence of Li^+, Na^+ and K^+**

(a) UV spectra of 6 µM d$(G_4CT)_3G_4$ in the presence of tris buffer only (gray), 1 mM LiCl (green), 500 mM LiCl (dark green), 1 mM NaCl (red), 500 mM NaCl (dark red), 1 mM KCl (blue) and 500 mM KCl (dark blue). (b) UV melting curves of 6 µM d$(G_4CT)_3G_4$ in the presence of tris buffer only (gray), 1 mM LiCl (green), 500 mM LiCl (dark green), 1 mM NaCl (red), 500 mM NaCl (dark red), 1 mM KCl (blue) and 500 mM KCl (dark blue). Temperatures corresponding to the half-maximal decay of absorbance $T_{1/2}$ were determined as shown in table.

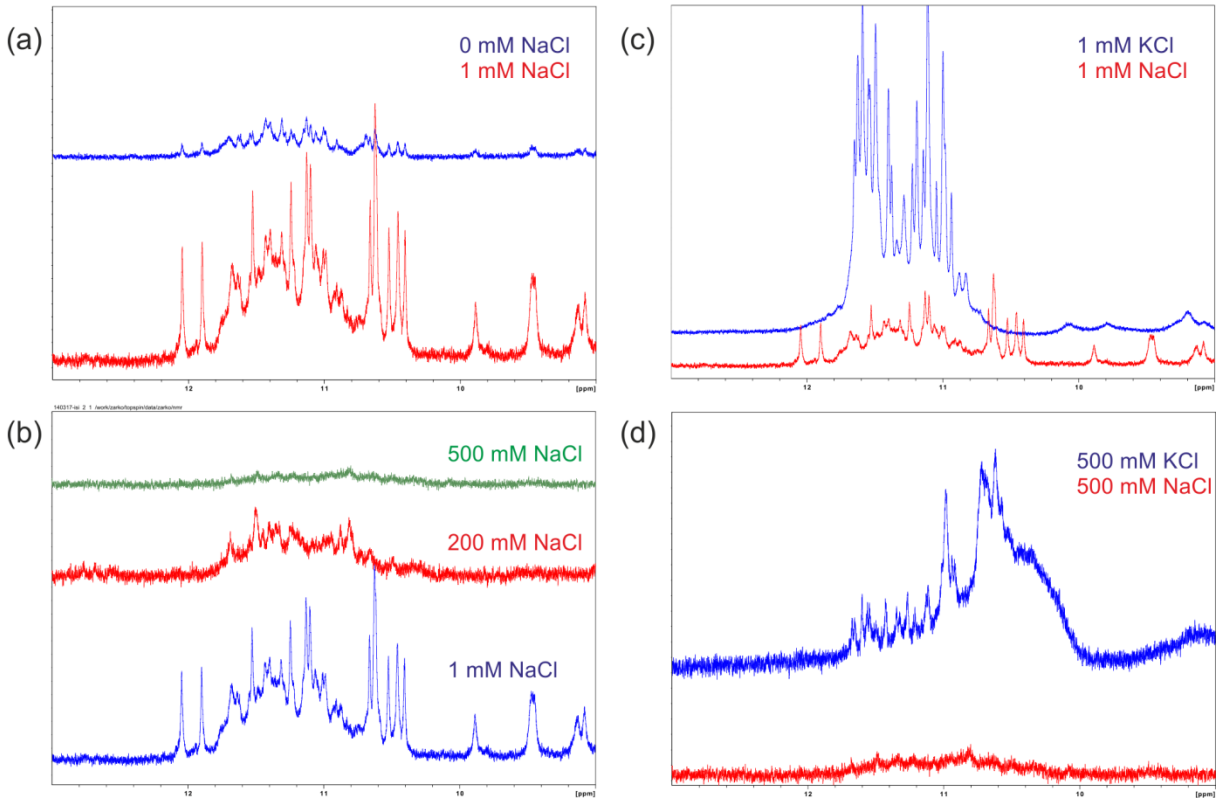


Fig. S5: **NMR spectra of d[$(G_4CT)_3G_4$] in the presence of Na^+**

NMR was measured on a Bruker advanced III 600 spectrometer. (a) ^{1}H-NMR spectra of 65 µM d$(G_4CT)_3G_4$ folded without (blue) and in the presence of 1 mM (red) NaCl. (b) ^{1}H-NMR spectra of 65 µM d$(G_4CT)_3G_4$ folded in the presence of 1 mM (blue) NaCl., 200 mM (red), 500 mM (green) NaCl. (c) Comparison of ^{1}H-NMR spectra of 65 µM d$(G_4CT)_3G_4$ folded in the presence of 1 mM (red) NaCl or 1 mM KCl (blue). (d) Comparison of ^{1}H-NMR spectra of 65 µM d$(G_4CT)_3G_4$ folded in the presence of 500 mM (red) NaCl or 500 mM KCl (blue).

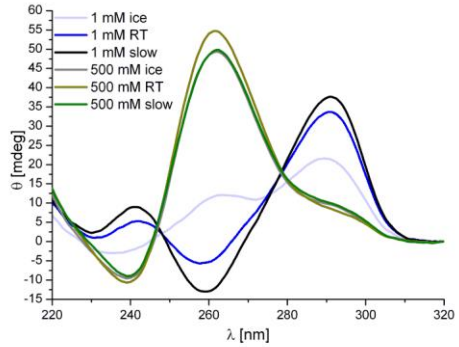
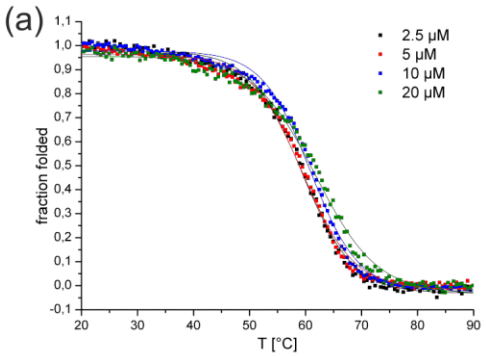
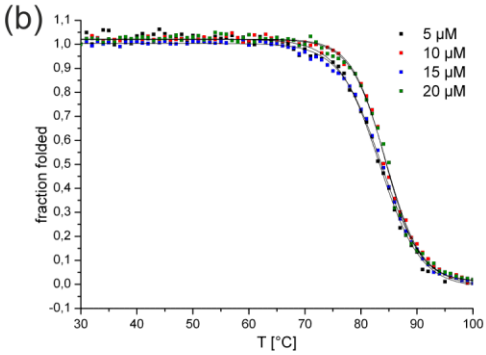


Fig. S6: **CD spectra of d[(G$_4$CT)$_3$G$_4$] folded with different cooling rates**

d[(G$_4$CT)$_3$G$_4$] was denatured with 1 mM (blues) or 500 mM (greens) KCl present, cooled to room temperature over night or immediately transferred to room temperature or ice. At 500 mM KCl the parallel conformations forms under all conditions, the antiparallel conformation at 1 mM KCl requires slow renaturation.



c (oligo) [µM]	$T_{1/2}$ [°C]
2.5	59.3
5	59.4
10	60.8
15	61.7



c (oligo) [µM]	$T_{1/2}$ [°C]
5	83.3
10	84.4
15	83.6
20	84.3

Fig. S7: **Melting profiles of d[(G$_4$CT)$_3$G$_4$] in presence of low concentrations of K$^+$**

(a): Normalized CD melting curves of the antiparallel conformer of d(G$_4$CT)$_3$G$_4$ using 2.5 µM (black), 5 µM (red), 10 µM (blue) and 20 µM (green) oligo in the presence of 1 mM KCl. Temperatures corresponding to the half-maximal decay of ellipticity $T_{1/2}$ were determined as shown in table. ´**(b)**: Normalized CD melting curves of the antiparallel conformer of d(G$_4$CT)$_3$G$_4$ using 5 µM (black), 10 µM (red), 15 µM (blue) and 20 µM (green) oligo in the presence of 25 mM KCl. Temperatures corresponding to the half-maximal decay of ellipticity $T_{1/2}$ were determined as shown in table.

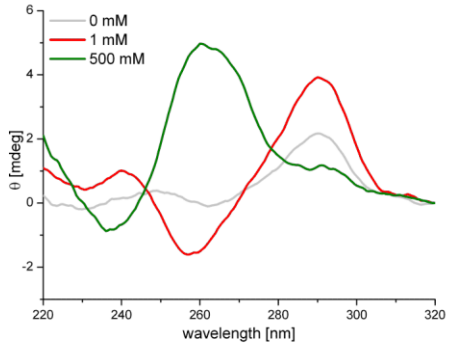


Fig. S8: CD spectra of spin-labeled d[(G$_4$CT)$_3$G$_4$] in presence of K$^+$

CD spectra of 5 µM spin-labeled d(G$_4$CT)$_3$G$_4$ in 10 mM tris-HCl, pH 7.2 with 0 mM (gray), 1 mM (red) and 500 mM (green) KCl, total volume 200 µL measured in a 1 mm pathlength cell.

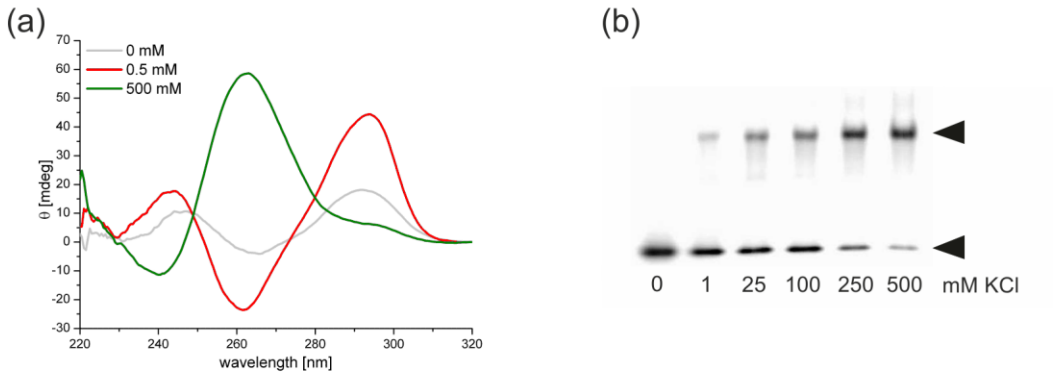


Fig. S9: Analysis of d[(G$_4$CT)$_3$G$_4$] in presence of K$^+$ by EMSA

(a): CD spectra of 5 µM d[(G$_4$CT)$_3$G$_4$] in 1xTBE pH 8.5 with 0 mM (gray), 1 mM (red) and 500 mM (green) KCl. **(b)**: EMSA of 5'-radiolabeled d[(G$_4$CT)$_3$G$_4$] in the presence of increasing KCl concentrations. Samples were subjected to electrophoresis directly after folding. The band for the parallel GQP increases with increasing K$^+$ concentrations, simultaneously the band for the antiparallel conformer decreases.

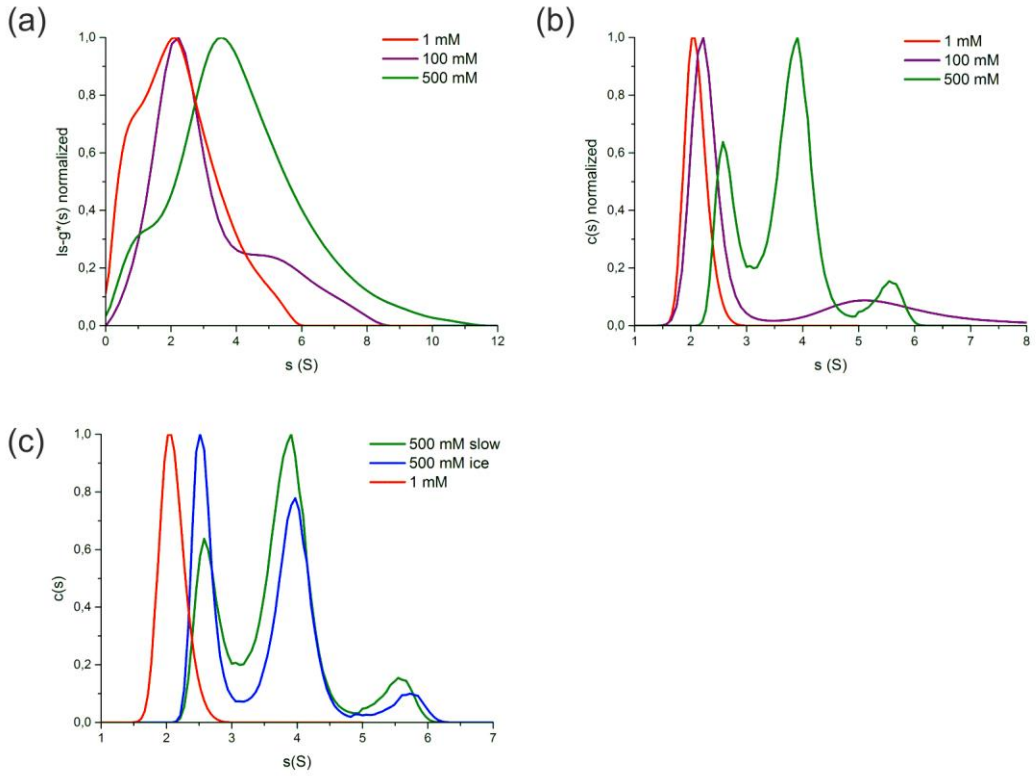


Fig. S10: **Analysis of d[(G$_4$CT)$_3$G$_4$] by AUC**

(a): Sedimentation coefficient distribution of d[(G$_4$CT)$_3$G$_4$] folded in the presence of 1 mM (red), 100 mM (violet), 500 mM (green) KCl as determined by analytical ultracentrifugation. 500 mM KCl sample was diluted to 100 mM KCl before centrifugation to adjust density of the solution to 100 mM sample. **(b)**: Sedimentation coefficient distribution of d[(G$_4$CT)$_3$G$_4$] folded in the presence of 1 mM (red), 100 mM (blue), 500 mM (green) KCl using a partial-specific volume of 0.55 mL/g for all GQP species. **(c)**: Sedimentation coefficient distribution of d[(G$_4$CT)$_3$G$_4$] folded in the presence of 1 mM KCl (red), 500 mM KCl and immediate transfer to ice after denaturation (blue) and 500 mM KCl and slow renaturation (green). A partial-specific volume of 0.55 mL/g was used for determination of c(s) for all GQP species.

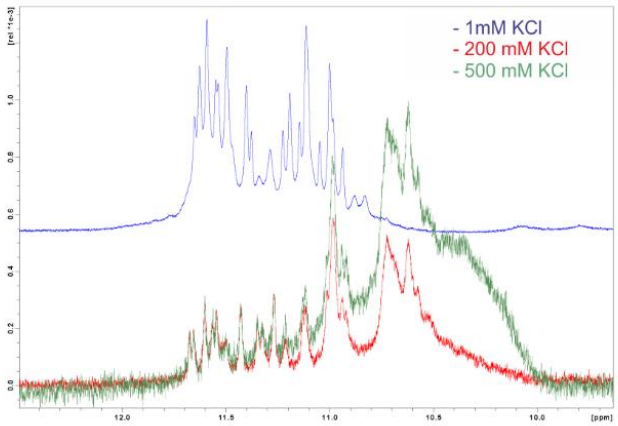


Fig. S11: **NMR spectra of d[(G$_4$CT)$_3$G$_4$] in the presence of K$^+$**

^{1}H-NMR spectra of 65 µM d(G$_4$CT)$_3$G$_4$ folded in the presence of 1 mM (blue), 200 mM (red), 500 mM (green) KCl. NMR was measured on a Bruker advanced III 600 spectrometer.

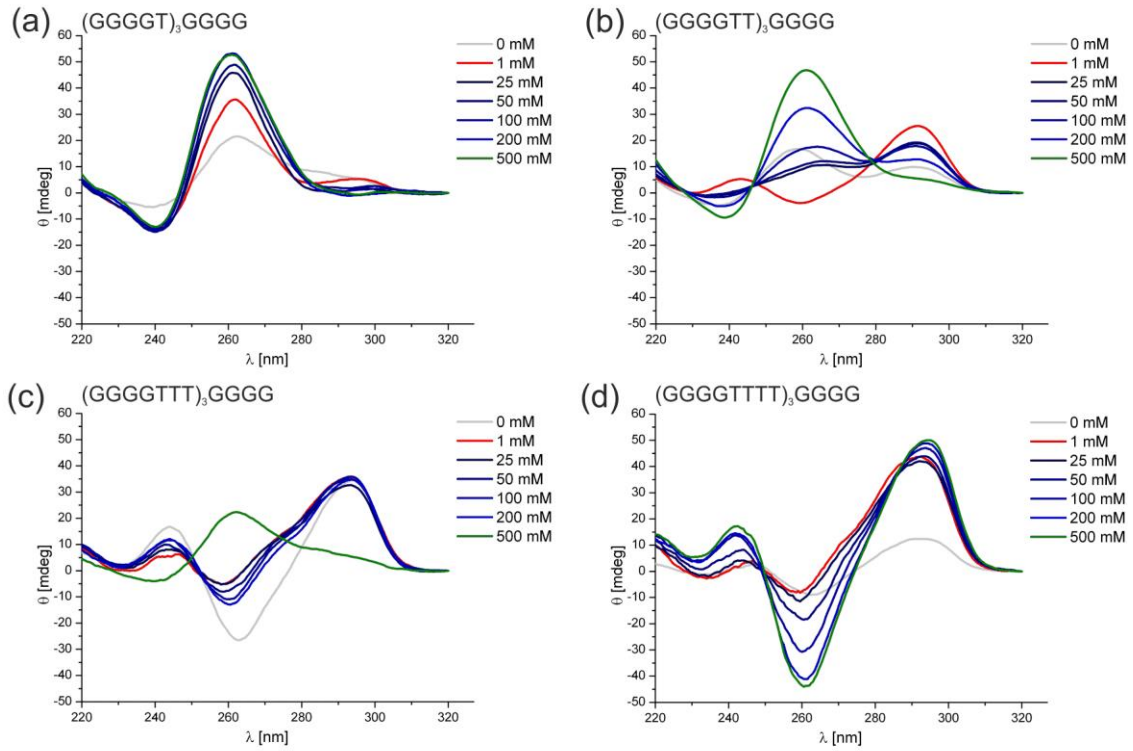


Fig. S12: CD-spectra of quadruplex sequences with varying T-loop length

(a): d[$(G_4T)_3G_4$], **(b)**: d[$(G_4T_2)_3G_4$], **(c)**: d[$(G_4T_3)_3G_4$], **(d)**: d[$(G_4T_4)_3G_4$]. Spectra were acquired in the presence of increasing concentrations of KCl from 0 mM (gray), 1 mM (red), 25 to 200 mM (blue) to 500 mM (green).

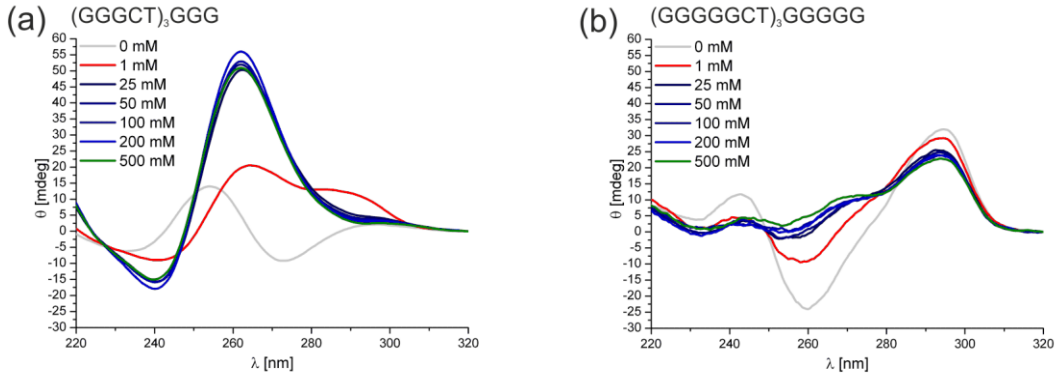


Fig. S13: **CD-spectra of quadruplex sequences with varying G-tract length**

(a): d[$(G_3CT)_3G_3$], **(b)**: d[$(G_5CT)_3G_5$]. Spectra were acquired in the presence of in increasing concentrations of KCl from 0 mM (gray), 1 mM (red), 25 to 200 mM (blues) to 500 mM (green).

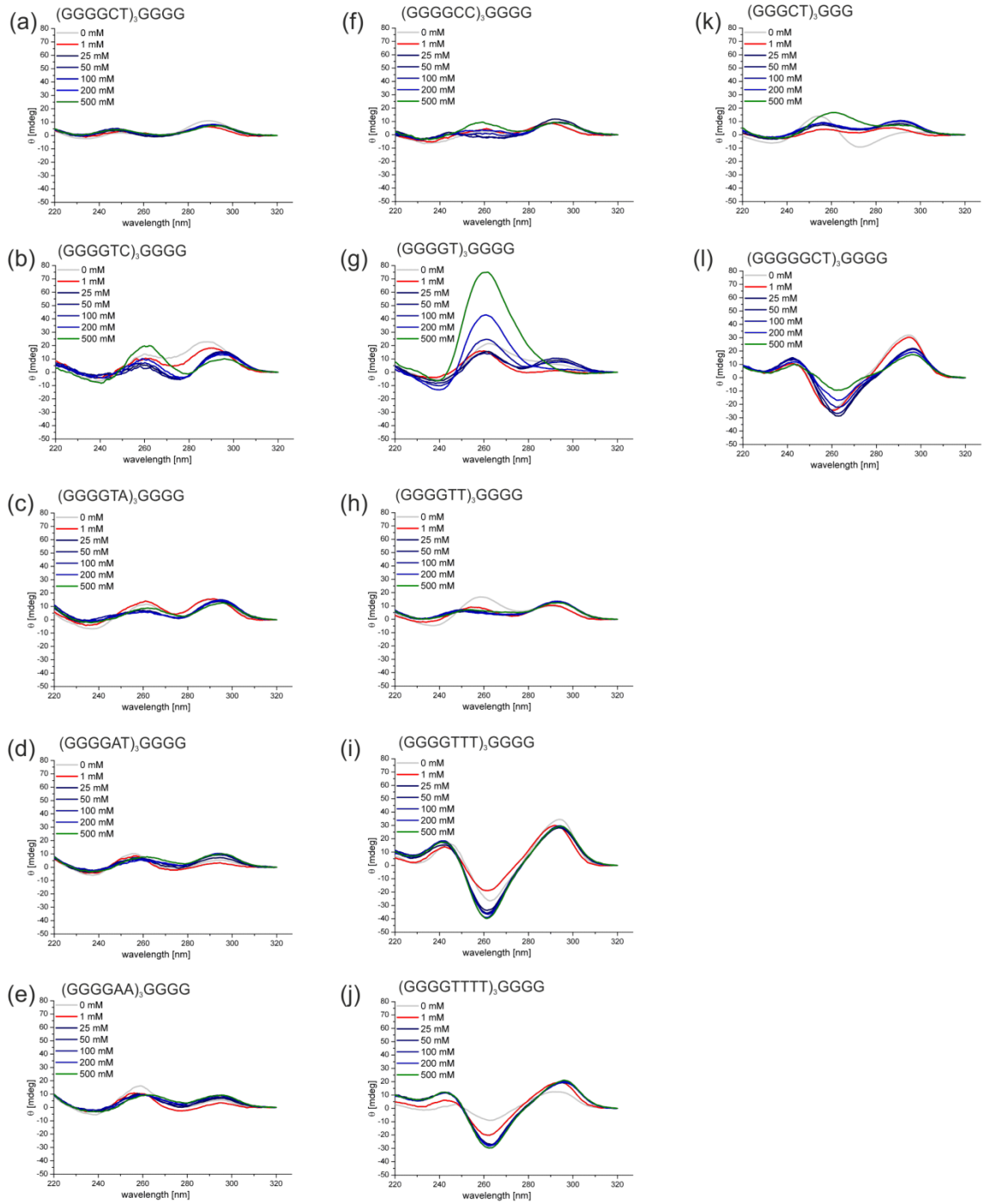


Fig. S14: **CD-spectra of all oligonucleotides used in this study in the presence of NaCl**

(a): d[$(G_4CT)_3G_4$], **(b)**: d[$(G_4TC)_3G_4$], **(c)**: d[$(G_4TA)_3G_4$], **(d)**: d[$(G_4AT)_3G_4$], **(e)**: d[$(G_4A_2)G_4$], **(f)**: d[$(G_4C_2)_3G_4$], **(g)**: d[$(G_4T)_3G_4$], **(h)**: d[$(G_4T_2)G_4$], **(i)**: d[$(G_4T_3)_3G_4$], **(j)**: d[$(G_4T_4)_3G_4$], **(k)**: d[$(G_3CT)_3G_3$], **(l)**: d[$(G_5CT)_3G_5$]. All spectra were acquired in the presence of increasing concentrations of NaCl ranging from 0 mM (gray), 1 mM (red), 25 to 200 mM (blues) and 500 mM (green).

Strain	protein	protein description	position start	end	sequence	strand	amino acid sequence
Bacillus infantis NRRL B-14911	AGX04358.1	hypothetical protein	2289827	2289806	$(GGGGCT)_3GGGG$	neg	GWGWG
Burkholderia ambifaria AMMD chromosome 2	ABI90376.1	heavy metal translocating P-type ATPase	1822405	1822426	$(GGGGCT)_3$ GGGG	pos	GAGAGAG
Burkholderia cenocepacia MC0-3 chromosome 2	ACA93829.1	hypothetical protein	1705109	1705130	$(GGGGCT)_3$ GGGG	pos	GLGLGLG
Frankia alni str. ACN14A chromosome	CAJ60943.1	hypothetical protein putative HNH endonuclease domain	2501669	2501690	$(GGGGCT)_3$ GGGG	pos	WGWGWGWG
Frankia symbiont of Datisca glomerata	AEH10209.1	FAD dependent oxidoreductase	3421555	3421516	$(GGGGCT)_6GGGG$	neg	GAGAGAGAGAGAG
Gloeobacter kilaueensis JS1	AGY58131.1	glycosyl transferase family 39	1987505	1987484	$(GGGGCT)_3$ GGGG	neg	GLGLGLG
Gluconacetobacter xylinus NBRC 3288 DNA	BAK83422.1	hypothetical protein	1165429	1165450	$(GGGGCT)_3$ GGGG	pos	WGWGWGWG
Phycisphaera mikurensis NBRC 102666 DNA	BAM04236.1	hypothetical protein	2452916	2452895	$(GGGGCT)_3$ GGGG	neg	GLGLGLG
Ralstonia solanacearum CFBP2957 plasmid RCFBPv3_mp	CBJ53914.1	protein of unknown function	1336019	1335998	$(GGGGCT)_3$ GGGG	neg	WGWGWGWG
Ralstonia solanacearum str. PSI07 chromosome	CBM10292.1	hypothetical protein	3425125	3425146	$(GGGGCT)_3$ GGGG	pos	GLGLGLG
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91	CAR35995.1	glutathione-regulated potassium-efflux system protein KefC	100073	100094	$(GGGGCT)_3$ GGGG	pos	GLGLGLG
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91	CAR36596.1	sensor protein KdpD	757600	757573	$(GGGGCT)_5$	neg	GLGLGLGLG
Salmonella enterica subsp. enterica serovar Gallinarum/pullorum str. CDC1983-67	AGU63044.1	glutathione-regulated potassium-efflux system protein KefC	100108	100129	$(GGGGCT)_3$ GGGG	pos	GLGLGLG
Salmonella enterica subsp. enterica serovar Gallinarum/pullorum str. RKS5078	AET52547.1	glutathione-regulated potassium-efflux system protein KefC	100109	100130	$(GGGGCT)_3$ GGGG	pos	GLGLGLG
Shigella flexneri 2002017	ADA75072.1	Membrane-spanning protein of hydrogenase 3	2855294	2855253	$(GGGGCT)_7$	neg	GLGLGLGLGLGLG
Shigella flexneri 2a str. 301	AAN44231.2	membrane-spanning protein of hydrogenase 3	2816692	2816653	$(GGGGCT)_6GGGG$	neg	GLGLGLGLGLGLG
Stackebrandtia nassauensis DSM 44728	ADD41590.1	YD repeat protein	2008682	2008723	$(GGGGCT)_7$	pos	GAGAGAGAGAGAGA
Stigmatella aurantiaca DW4/3-1	ADO75292.1	uncharacterized protein	9222581	9222602	$(GGGGCT)_3$ GGGG	pos	GLGLGLG
Xanthomonas oryzae pv. oryzae PXO99A	ACD60711.1	acetylxylan esterase	4273877	4273856	$(GGGGCT)_3$ GGGG	neg	RGWGWGWG

Table S1. GGGGCT repeat sequences within protein encoding sequences

Table shows the occurrence of potential quadruplex forming sequences in bacterial genomes within protein encoding regions. Only complete sequences of bacterial genomes were taken into account. Start and end point of the sequence as well as location on plus (pos) or minus (neg) strand is given. Amino acid sequence, which a repeat sequence encodes, and Genbank accession number of the respective protein are given.

Strain	protein	protein description	position start	end	sequence	strand
Acidothermus cellulolyticus 11B	ABK52391.1	cellulose-binding, family II	668279	668258	$(GGGGCT)_3GGGG$	neg
Actinosynnema mirum DSM 43827	ACU36234.1	hypothetical protein	2552254	2552281	$(GGGGCT)_4GGGG$	pos
Adlercreutzia equolifaciens DSM 19450 DNA	BAN76597.1	polysaccharide deacetylase	842549	842528	$(GGGGCT)_3GGGG$	neg
Azospirillum lipoferum 4B plasmid AZO_p1	CBS88380.1	putative hybrid sensor histidine kinase	63134	63167	$(GGGGCT)_6GGGG$	pos
Azotobacter vinelandii CA	AGK13378.1	D-Alanyl-D-Alanine carboxypeptidase	800884	800863	$(GGGGCT)_3GGGG$	neg
Azotobacter vinelandii CA6	AGK17736.1	D-Alanyl-D-Alanine carboxypeptidase	800876	800855	$(GGGGCT)_3GGGG$	neg
Azotobacter vinelandii DJ	ACO77088.1	D-Alanyl-D-Alanine carboxypeptidase	800864	800843	$(GGGGCT)_3GGGG$	neg
Bdellovibrio bacteriovorus complete genome, strain HD100	CAE77747.1	conserved hypothetical protein	61375	61402	$(GGGGCT)_5GGGG$	pos
Burkholderia phenoliruptrix BR3459a chromosome 1	AFT85453.1	exodeoxyribonuclease V alpha subunit	1572212	1572191	$(GGGGCT)_3GGGG$	neg
Burkholderia sp. CCGE1001 chromosome 1	ADX54801.1	exodeoxyribonuclease V, alpha subunit	1558469	1558404	$(GGGGCT)_{11}$	neg
Burkholderia vietnamiensis G4 chromosome 1	ABO54384.1	protein of unknown function DUF1228	1473363	1473342	$(GGGGCT)_3GGGG$	neg
Deinococcus gobiensis I-0 plasmid P2	AFD27512.1	hypothetical protein	197513	197492	$(GGGGCT)_3GGGG$	neg
Delftia sp. Cs1-4	AEF89628.1	cytochrome c class I	2987497	2987476	$(GGGGCT)_3GGGG$	neg
Desulfitobacterium hafniense DCB-2	ACL20815.1	hypothetical protein	2996050	2996071	$(GGGGCT)_3GGGG$	pos
Frankia sp. CcI3	ABD11955.1	hypothetical protein	3060256	3060235	$(GGGGCT)_3GGGG$	neg
Pelotomaculum thermopropionicum SI DNA	BAF58268.1	hypothetical membrane protein	82967	82938	$(GGGGCT)_5$	neg
Ralstonia solanacearum CFBP2957 plasmid RCFBPv3_mp	CBJ53915.1	conserved exported protein of unknown function	1336019	1335998	$(GGGGCT)_3GGGG$	neg
Salinispora arenicola CNS-205	ABV97463.1	hypothetical protein	1802368	1802389	$(GGGGCT)_3GGGG$	pos
Sanguibacter keddieii DSM 10542	ACZ22171.1	ATP-dependent DNA helicase RecQ	2508401	2508342	$(GGGGCT)_{10}$	neg
Shigella dysenteriae Sd197	hycC Gene	hycC	2702838	2702793	$(GGGGCT)_8$	neg
Streptomyces collinus Tu 365	AGS72154.1	hypothetical protein	6159803	6159824	$(GGGGCT)_3GGGG$	pos
Streptomyces davawensis strain JCM 4913	CCK29538.1	hypothetical protein	5737369	5737390	$(GGGGCT)_3GGGG$	pos
Streptomyces rochei plasmid pSLA2-L DNA	BAC76555.1	probable plasmid partitioning protein, ParB	161471	161498	$(GGGGCT)_5GGGG$	pos

Table S2. GGGGCT repeat sequences complementary to coding regions

Table shows the occurrence of potential quadruplex forming sequences in bacterial genomes that are located complementary to protein coding regions. Only complete sequences of bacterial genomes were taken into account. Start and end point of the sequence as well as location on plus (pos) or minus (neg) strand is given, as well as Genbank accession number of the respective protein.

strain	position start	end	sequence	strand	5' gene	protein description	orient	dis. 5' [bp]	3' gene	protein description	orient	dis. 3' [bp]
Amycolatopsis mediterranei RB	6108902	6108923	$(GGGGCT)_3GGGG$	pos	AGT86094.1	two-component system response regulator	pos	41	AGT86095.1	cupin domain-containing protein	neg	195
Amycolatopsis mediterranei S699	6108967	6108988	$(GGGGCT)_3GGGG$	pos	AFO78966.1	two-component system response regulator	pos	38	AFO78967.1	cupin domain-containing protein	neg	195
Amycolatopsis mediterranei S699	6099040	6099061	$(GGGGCT)_3GGGG$	pos	AEK44080.1	two-component system response regulator	pos	41	AEK44081.1	cupin domain-containing protein	neg	195
Delftia acidovorans SPH-1,	2403406	2403439	$(GGGGCT)_5GGGG$	pos	ABX34829.1	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	pos	123	ABX34830.1	glycolsyl transerase family 39	pos	26
Frankia sp. CcI3	1296252	1296231	$(GGGGCT)_3GGGG$	neg	ABD10466.1	conserved hypothetical protein	pos	805	ABD10467.1	FAD linked-oxidase like	neg	10
Geobacter sp. M18	4943641	4943608	$(GGGGCT)_6GGGG$	neg	ADW15616.1	conserved hypothetical protein	neg	200	ADW15617.1	helicase c2	pos	93
Geobacter sp. M21	1968533	1968554	$(GGGGCT)_3GGGG$	pos	ACT17742.1	hypothetical protein	neg	270	ACT17743.1	thioesterase superfamily protein (275bp)	neg	275
Geobacter sp. M21	4437666	4437761	$(GGGGCT)_{10}GGGACT-(GGGGCT)_4GGGG$	pos	ACT19869.1	FRG domain protein	neg	116	ACT19869.1	conserved hypothetical protein	neg	33
Pseudomonas denitrificans ATCC 13867	3371781	3371760	$(GGGGCT)_3GGGG$	neg	AGI24854.1	hypothetical protein	neg	123	AGI24855.1	hypothetical protein	neg	238
Streptomyces avermitilis MA-4680 DNA	2499489	2499547	$(GGGGCT)_9GGGG$	pos	BAC69758.1	hypothetical protein	pos	31	BAC69759.1	hypothetical protein	neg	16
Treponema pallidum subsp. pallidum DAL-1	150542	150563	$(GGGGCT)_3GGGG$	pos	AEZ60452.1	hypothetical protein	pos	48	AEZ60453.1	hypothetical protein	pos	120
Treponema pallidum subsp. pallidum str. Chicago	150548	150569	$(GGGGCT)_3GGGG$	pos	TPChic_0127a	TPChic_0127a gene	pos	48	ADD72282.1	conserved hypothetical protein	pos	189
Treponema pallidum subsp. pallidum str. Nichols	149332	149353	$(GGGGCT)_3GGGG$	pos	AAC65119.1	predicted coding region TP0127	pos	48	AAC65120.1	predicted coding region TP0128	pos	120
Treponema pallidum subsp. pallidum str. Nichols	150545	150566	$(GGGGCT)_3GGGG$	pos	AGN75349.1	hypothetical protein	pos	68	AGN75350.1	hypothetical protein	pos	120
Verrucosispora maris AB-18-032	6353913	6353824	$(GGGGCT)_{15}$	neg	AEB47716.1	diguanylate cyclase with gaf sensor	pos	9	AEQ94537.1	aspartate-semialdehyde dehydrogenase	neg	109
Xanthomonas oryzae pv. oryzicola BLS256,	295882	295903	$(GGGGCT)_3GGGG$	pos	AEQ94536.1	two-component system response regulator protein ntrC	pos	187	AEQ94537.1	superoxide dismutase	pos	299

Table S3. GGGGCT repeat sequences within untranslated regions

Table shows the occurrence of potential quadruplex forming sequences in bacterial genomes that are located in untranslated regions between two genes. Only complete sequences of bacterial genomes were taken into account. Start and end point of the repetitive sequence as well as location on plus (pos) or minus (neg) strand is given. The proteins encoded by the next genes located 5' and 3' of the repeat sequence, their location on the geneome as well as the distance between the genes and the repeat sequence, and Genbank accession numbers of the proteins are listed.

Human genome, GRCh38 Primary Assembly	Accession Number	Gene/RNA	strand	position start	end	sequence	strand
Homo sapiens chromosome 1	NC_000001.11	PEX14	positive	10630209	10630188	$(G_4CT)_3G_4$	negative
		CASZ1	negative	10656809	10656786	$(G_4CT)_4$	negative
		--		90547895	90547916	$(G_4CT)_3G_4$	positive
		SLC44A3	positive	94820362	94820341	$(G_4CT)_3G_4$	negative
		SLC44A3	positive	94854900	94854877	$(G_4CT)_4$	negative
		KIAA0907	negative	155917585	155917612	$(G_4CT)_5G_4$	positive
		--		224865257	224865228	$(G_4CT)_5$	negative
Homo sapiens chromosome 2	NC_000002.12	NCOA1	positive	24597500	24597521		positive
		--		42810131	42810108	$(G_4CT)_4$	negative
		GPR1 antisense RNA	positive	206229916	206229937		positive
		Tensin 1	positive	217858985	217858962	$(G_4CT)_4$	negative
		--		234260112	234260087	$TC(G_4CT)_4$	negative
Homo sapiens chromosome 3	NC_000003.12	--		40791360	40791401	$(G_4CT)_7$	positive
		CACNA2D2	negative	50489619	50489654	$(G_4CT)_6$	positive
		PLXNA1	positive	127003141	127003174	$(G_4CT)_5G_4$	positive
		--		127392119	127392142	$(G_4CT)_4$	positive
Homo sapiens chromosome 5	NC_000005.10	uncharact. LOC101929505	negative	16428163	16428138	$CT(G_4CT)_4$	negative
		SSBP2	negative	81750273	81750252	$(G_4CT)_3G_4$	negative
		CXXC5	positive	139682352	139682325	$(G_4CT)_4G_4$	negative
		KCNIP1	positive	170485812	170485789	$CT(G_4CT)_3G_4$	negative
	NC_000006.12	--		71886633	71886610	$(G_4CT)_4$	negative
		--		168851680	168851703	$(G_4CT)_4$	positive
Homo sapiens chromosome 7	NC_000007.14	--		128739101	128739124	$(G_4CT)_4$	positive
		--		129122728	129122749	$(G_4CT)_3G_4$	positive
		KRBA1	positive	149730262	149730241	$(G_4CT)_3G_4$	negative
		--		151880833	151880858	$CT(G_4CT)_4$	positive
		DPP6	positive	154305082	154305019	$(G_4CT)_{10}G_4$	negative
		--		155467865	155467836	$(G_4CT)_5$	negative
Homo sapiens chromosome 8	NC_000008.11	SDCBP	positive	58553559	58553590	$CT(G_4CT)_5$	positive
		--		66379273	66379300	$(G_4CT)_4G_4$	positive
Homo sapiens chromosome 9	NC_000009.12	--		41647291	41647266	$CT(G_4CT)_4$	negative
		--		61668692	61668696	$(G_4CT)_4$	negative
		--		61674981	61674952	$(G_4CT)_5$	negative
		--		61861242	61861217	$CT(G_4CT)_4$	negative
		--		65390307	65390276	$CT(G_4CT)_4$	negative
		--		67719561	67719586	$CT(G_4CT)_4$	positive
		DAB2IP	positive	121731667	121731696	$(G_4CT)_4$	positive
		TOR1A	Negative	129823518	129823491	$(G_4CT)_4G_4$	negative
		TOR1A	negative	129823720	129823699	$(G_4CT)_3G_4$	negative
		ABCA2	negative	137017955	137017932	$(G_4CT)_4$	negative
Homo sapiens chromosome 10	NC_000010.11	CCNY	positive	35357216	35357189	$(G_4CT)_4G_4$	negative
		ANKRD30BP3	positive	45180318	45180351	$(G_4CT)_5G_4$	positive
		C10orf11	positive	75972870	75972891	$(G_4CT)_3G_4$	positive
		--		129191495	129191468	$(G_4CT)_5$	negative
		MTG1	negative	133402253	133402276	$(G_4CT)_4$	positive
Homo sapiens chromosome 11	NC_000011.10	DUSP8	negative	1563802	1563781	$(G_4CT)_3G_4$	negative
		CTSD	negative	1753983	1753960	$(G_4CT)_4$	negative
		NAV2	positive	19555868	19555843	$CT(G_4CT)_4$	negative
		CD82	positive	44594416	44594391	$CT(G_4CT)_4$	negative
		--		69471960	69471937	$(G_4CT)_4$	negative
		--		116526305	116526282	$(G_4CT)_4$	negative
Homo sapiens chromosome 12	NC_000012.1	CD27 antisense RNA 1	negative	6448967	6448946	$(G_4CT)_3G_4$	negative

		CD27	positive				
		--		45611265	45611308	$CT(G_4CT)_7$	positive
		ASIC1	positive	50074323	50074346	$(G_4CT)_4$	positive
		LOC102724030	positive	54159175	54159150	$CT(G_4CT)_4$	negative
		--		82686971	82686996	$CT(G_4CT)_4$	positive
Homo sapiens chromosome 13	NC_000013.11	--		18249046	18249021	$CT(G_4CT)_4$	negative
		--		28819787	28819832	$(G_4CT)_7G_4$	positive
		long intergenic non-protein coding RNA 1070	positive	112197836	112197813	$(G_4CT)_4$	negative
Homo sapiens chromosome 14	NC_000014.9	TTLL5	positive	75812859	75812884	$CT(G_4CT)_4$	positive
		FOXN3	negative	89177541	89177568	$CT(G_4CT)_5$	positive
		BCL11B	negative	99188538	99188585	$CT(G_4CT)_8$	positive
Homo sapiens chromosome 15	NC_000015.10	--		69470248	69470279	$CT(G_4CT)_5$	positive
		ZNF710	positive	90001660	90001639	$(G_4CT)_3G_4$	negative
		ADAMTS17		100005977	100005952	$CT(G_4CT)_4$	negative
Homo sapiens chromosome 16	NC_000016.10	--		769266	769237	$(CTG_4)_5$	negative
		--		1004382	1004361	$(G_4CT)_3G_4$	negative
		SBK1	positive	28307068	28307037	$CT(G_4CT)_4G_4$	negative
		STX1B	negative	31001489	31001510	$(G_4CT)_3G_4$	positive
		--		83951197	83951174	$(G_4CT)_4$	negative
		GSE1	positive	85236028	85236049	$CT(G_4CT)_4$	positive
		ZFPM1	positive	88497286	88497259	$(G_4CT)_4G_4$	negative
Homo sapiens chromosome 17	NC_000017.11	NUP88	negative	5394652	5394675	$(CTG_4)_4$	positive
		ASGR1	negative	7176964	7176937	$(G_4CT)_4G_4$	negative
		mRNA-phosphoinositide-3-kinase, regulatory subunit 6	negative	8860458	8860437	$(G_4CT)_3G_4$	negative
		--		17504224	17504245	$(G_4CT)_3G_4$	positive
		LOC100420851	positive	36178069	3617811	$(G_4CT)_4$	positive
		--		36441000	36441023	$(G_4CT)_4$	positive
		--		37922975	37922950	$(G_4CT)_5$	negative
		LOC440434	negative	38256739	38256700	$(G_4CT)_6G_4$	negative
		PPY	negative	43941365	43941386	$(G_4CT)_3G_4$	positive
		DBF4B	positive	44749287	44749266	$(G_4CT)_3G_4$	negative
		NPEPPS	positive	47531714	47531741	$(G_4CT)_3G_4$	positive
		mRNA-transmembrane protein 105	negative	81313145	81313168	$(G_4CT)_4$	positive
		HGS	positive	81700665	81700636	$(G_4CT)_4$	negative
		FOXK2	positive	82556560	82556591	$CT(G_4CT)_4$	positive
Homo sapiens chromosome 18	NC_000018.10	--		37779372	37779351	$(G_4CT)_3G_4$	negative
		CTIF	positive	48699394	48699435	$(G_4CT)_7$	positive
		--		63610584	63610605	$(G_4CT)_3G_4$	positive
Homo sapiens chromosome 19	NC_000019.10	MED16	negative	877797	877768	$(G_4CT)_5$	negative
		MED16	negative	877977	877954	$(G_4CT)_4$	negative
		MED16	negative	878257	878236	$(G_4CT)_3G_4$	negative
		MED16	negative	878381	878358	$(G_4CT)_4$	negative
		MED16	negative	878871	878848	$(G_4CT)_4$	negative
		MED16	negative	879149	879122	$(G_4CT)_4G_4$	negative
		DOT1L	positive	2181797	2181766	$CT(G_4CT)_5$	negative
		mRNA-cytochrome c oxidase subunit VIIa polypeptide 1	negative	36152531	36152552	$(G_4CT)_3G_4$	positive
		SIPA1L3	positive	37929180	37929205	$CT(G_4CT)_4$	positive
		--		55781144	55781171	$(G_4CT)_3G_4$	positive
Homo sapiens chromosome 20	NC_000020.11	--		4489603	4489626	$(G_4CT)_4$	positive
		MMP24 antisense RNA 1	negative	35277785	35277818	$(G_4CT)_5G_4$	positive
		--		50217617	50217596	$(G_4CT)_3G_4$	negative
		PHACTR3	positive	59587754	59587775	$(G_4CT)_3G_4$	positive
		CDH4	positive	61611206	61611185	$(G_4CT)_3G_4$	negative
		OGFR OGFR antisense RNA 1	positive negative	62805395	62805428	$(G_4CT)_5G_4$	positive

Homo sapiens chromosome 22	NC_000022.11	LOC388849	negative	20150427	20150448	$(G_4CT)_3G_4$	positive
		ZDHHC8P1	negative	23393126	23393105	$(G_4CT)_3G_4$	negative
		ATXN10	positive	45752567	45752596	$(G_4CT)_4$	positive
		--		49748986	49748953	$(G_4CT)_5G_4$	negetive
		--		49929463	49929490	$(G_4CT)_4G_4$	positive
Homo sapiens chromosome X	NC_000023.11	AVPR2	positive	153905331	153905352	$(G_4CT)_3G_4$	positive

Table S4. GGGGCT repeat sequences the human genome (Description at the end of table)

Table S4. GGGGCT repeat sequences in the human genome

Table shows the occurrence of potential quadruplex forming sequences in the human genome. Only sequences of Human genome, GRCh38 Primary Assembly were taken into account. Start and end point of the repetitive sequence as well as location on positive or complement (negative) strand is given (last column). In case a repeat occurred within or complementary to a protein coding region or non-coding RNA, the gene or RNA as well as its location on the on positive or complement (negative) strand is listed (fourth column).

Experimental Section

Materials. Oligonucleotides (Table 1) for CD measurements and melting assays were synthesized by Sigma Aldrich (Steinheim, Germany) at the 1 µmol scale with HPLC purification. Prior to use in native Polyacrylamide gel electrophoresis (PAGE) oligonucleotides were purified by denaturating polyacrylamide-urea gel electrophoresis.

Table 1. List of oligonucleotides used in this study

Name	Sequence (5'-...-3')
$d[(G_4CT)_3G_4]$	d(GGGGCTGGGGCTGGGGCTGGGG)
$d[(G_4TC)_3G_4]$	d(GGGGTCGGGGTCGGGGTCGGGG)
$d[(G_4T)_3G_4]$	d(GGGGTGGGGTGGGGTGGGG)
$d[(G_4T_2)_3G_4]$	d(GGGGTTGGGGTTGGGGTTGGGG)
$d[(G_4T_3)_3G_4]$	d(GGGGTTTGGGGTTTGGGGTTTGGGG)
$d[(G_4T_4)_3G_4]$	d(GGGGTTTTGGGGTTTTGGGGTTTTGGGG)
$d[(G_3CT)_3G_3]$	d(GGGCTGGGCTGGGCTGGG)
$d[(G_5CT)_3G_5]$	d(GGGGGCTGGGGGCTGGGGGCTGGGGG)
$d[(G_4C_2)_3G_4]$	d(GGGGCCGGGGCCGGGGCCGGGG)
$d[(G_4AT)_3G_4]$	d(GGGGATGGGGATGGGGATGGGG)
$d[(G_4TA)_3G_4]$	d(GGGGTAGGGGTAGGGGTAGGGG)
$d[(G_4A_2)_3G_4]$	d(GGGGAAGGGGAAGGGGAAGGGG)
$d[(C_4AG)_3C_4]$	d(CCCCAGCCCCAGCCCCAGCCCC)

Circular dichroism (CD) measurements. CD spectra were recorded on a JASCO-J815 spectropolarimeter equipped with a MPTC-490S/15 multicell temperature unit using a 1 cm optical path and a reaction volume of 600 µL. Oligonucleotides were prepared as a 5 µM solution in 10 mM Tris-HCl pH 7.5 buffer unless otherwise noted, supplemented with either KCl, NaCl, LiCl or $MgCl_2$ as noted and denatured by heating to 98 °C for 5 min, followed by slow cooling to 20 °C over night. Scans were performed at 20 °C over a wavelength range of 220–320 nm (5 accumulations) with a scanning speed of 500 nm/min, 0.5 s response time, 0.5 nm data pitch and 1 nm bandwidth. The buffer spectrum was subtracted and all spectra zero-corrected at 320 nm.

Thermal denaturation. For thermal denaturation oligonucleotides were prepared as previously described. Samples were heated from 20 °C to 100 °C with a heating rate of 0.5 °Cmin^{-1}. The CD signal at 260 nm or 290 nm was recorded every 0.5 °C. The temperature of the half-maximal decay of ellipticity $T_{1/2}$ was obtained from the normalized ellipticity decrease.

UV thermal denaturation. 6 µM oligonucleotides were prepared as previously described. Thermal denaturation was performed with a Cary 100 Bio UV-Visible Spectrophotometer. UV spectra from 220 nm to 300 nm were recorded prior to melting. Samples were heated from 20 °C to 95 °C with a heating rate of 0.5 °Cmin^{-1}. UV absorbance at 295 nm was recorded every 1.0 °C. The temperature of the half-maximal decrease in absorbance $T_{1/2}$ was obtained from the normalized absorbance decrease.

Radioactive labeling. Prior to use in native PAGE oligonucleotides were 5'-end-labeled with γ-^{32}P-ATP by phosphate exchange reaction through T4-polynucleotide kinase (NEB), as described in the manufacturer's protocol. The products were purified by gel filtration (Sephadex G-25 column). After purification the DNA concentration was adjusted to 5 µM with unlabeled oligonucleotide.

Electrophoretic Mobility Shift Assay (EMSA). To investigate the influence of the potassium concentration on secondary structure formation of $d(G_4CT)_3G_4$ analytical native polyacrylamide gels were run. 5 µM radioactively labeled quadruplex DNA in 10 mM Tris-HCl was folded by heating up to 95 °C for 5 min and cooling down within 12 hours in the presence of different KCl concentrations as indicated. Polyacrylamide gels containing 16% acrylamide 40 (19:1) with a thickness of 0.4 mm were poured in 1xTBE supplemented with 100 mM KCl. The folded oligonucleotide was mixed 1:1 with native PAGE loading buffer (40% (v/v) glycerol, 0.25% (w/v) bromphenolblue, 0.25% (w/v) xylencyanol), 4 µL of the samples were loaded on the gel. Separation was carried out at a voltage of 300-400 V in 1xTBE buffer supplemented with 100 mM KCl. The running time was 6 h at 4 °C. Gels were analyzed by autoradiography.

Synthesis of spin-labeled d[$(G_4CT)_3G_4$]. The syntheses of DNA oligomers were performed on an ABI 394 DNA/RNA synthesizer with commercially available reagents. The spin-labeled DNA was synthesized using an extended coupling time for the spin-labeled phosphoramidite (10 min total in several pushes). 5-(ethylthio)-1H-tetrazole (0.25 M in CH_3CN) was used as the activator, 0.02 M iodine in THF/Pyridine/Water as the oxidizer, and 3% (v/v) dichloroacetic acid in CH_2Cl_2 as the deblock solution. The DNA oligomers were cleaved from the solid support and with concentrated aqueous ammonia at 55 ºC for 12 h. The crude oligomers were lyophilized and then purified twice by HPLC. First by ion-exchange chromatography (DNAPac® PA100) under denaturing conditions; eluent A: 25 mM NaOH pH 12.0, eluent B: 25 mM NaOH, 1 M NaCl, pH 12.0 with a gradient of 0 to 40% B in 35 minutes. The oligomers were lyophilized, and desalted using a Sep-Pak C18 Classic Cartridge using standard conditions and purified by RP-HPLC using A = 0.05 M triethylammonium acetate in H_2O, pH 7.0 and B = methanol. ESI-MS calculated for 7146.3, found 7146.3.

ERP measurements. G-quadruplex samples containing 80 µM DNA oligomers were folded as described earlier without KCl or in the presence of 1 mM and 500 mM KCl, by heating up to 98 °C and slowly cooling down to room temperature. 20% (v/v) glycerol was added and samples were filled into Bruker Q-band capillaries and shock frozen in liquid nitrogen. Q-band DEER measurements were performed at T = 50 K on a Bruker ELEXSYS E580 spectrometer equipped with an EN 5107D2 Q-band probehead (Bruker Biospin) and a helium gas flow system (CF935, Oxford Instruments) using the following pulse sequence:

$$\left(\frac{\pi}{2}\right)_{obs} \cdots \tau_1 \cdots (\pi)_{obs} \cdots t \cdots (\pi)_{pump} \cdots (\tau_1 + \tau_2 - t) \cdots (\pi)_{obs} \cdots \tau_2 \cdots echo.$$

The pump pulse (length 28 ns) was set to the maximum of the nitroxide spectrum and the observer pulses (length 24/48/48 ns) were shifted 49 MHz higher. All samples were measured at $\tau_1 = 256$ ns and $\tau_2 = 2.6$ µs. Shot repetition time was set to 4 µs. Accumulation time per sample was 10 hours.

EPR data analysis. The DEER curves were analysed using the DeerAnalysis2013.2 software package (G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham,C. R. Timmel, D. Hilger, H. Jung, DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data, Appl. Magn. Reson. 2006, 30, 473–498). As a reference for the calculation of the number of coupled spins a 100% doubly labeled oligo(para-phenylenethinylen) was used.

NMR measurements. NMR spectra were acquired at 278 K on a Bruker Avance III 600 MHz spectrometer equipped with a TCI-H/C/N triple resonance cryoprobe. 65 µM $d(G_4CT)_3G_4$ was dissolved in Tris buffer supplemented with different concentrations of KCl or NaCl and containing 5 % Vol. D_2O as field lock. G-quadruplexes were folded as described earlier, by heating up to 98 °C and slowly cooling down to room temperature. Because high salt concentration disturbed the NMR measurements, 325 µM $d(G_4CT)_3G_4$ folded with 500 mM KCl or NaCl was diluted to 200 mM KCl or NaCl prior to the measurement with the final oligo concentration being 65 µM, because of the high stability of the parallel conformer formed in the presence of KCl dilution should not interfere with the structure formed.1D-proton spectra were acquired with 32'000 data points using 10k accumulated scans due to low sample concentration, and processed with an exponential line broadening window function. Solvent suppression was achieved by excitation sculpting (Hwang TL, Shaka AJ, Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. J. Magn. Reson. Series A, 1995, 112, 275-279.). Acquired data was processed and analyzed using Bruker Topspin and MestReNova software.

Analytical Ultracentrifugation (AUC). Sedimentation velocity profiles were recorded using a Beckman Optima XLA ultracentrifuge equipped with absorbance optical system and An-60-Ti rotor. All samples were PAGE-purified before folding. Samples were prepared with 5 µM oligonucleotide and folded as previously described. Except samples folded in 500 mM KCl, here samples were prepared using 25 µM oligonucleotide as the sample was diluted to 100 mM KCl before measurement resulting in the same oligonucleotide concentration and comparable solution viscosity and density to samples folded directly in 100 mM KCl. Dilution should not interfere with the structure formed because of the high stability of the parallel conformer. Sedimentation velocity profiles were recorded at 25 °C and 50,000 rpm. Primary data were processed using Sedfit fit to appropriate physical models (Schuck, P. (2000). "Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling." Biophysical Journal 78(3): 1606-1619.). A value of 0.55 g/ml was used for *v*. Buffer viscosity and density was calculated with Ultrascan 3 (Demeler B, UltraScan version 3. A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. The University of Texas Health Science Center at San Antonio, Department of Biochemistry. http:/www.ultrascan.uthscsa.edu). The partial specific volume used for determination of c(s) for all GQP species is 0.55 mL/g (Hellman LM, Rodgers DW, Fried MG, Phenomenological partial-specific volumes for G-quadruplexes, Eur Biophys J, 2010, 39(3):389-96).

nBLAST search. Bacterial genomes were searched for the quadruplex motif using the nBLAST web server (http://blast.ncbi.nlm.nih.gov) (Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.) and the following parameters: query

sequence “GGGGCTGGGGCTGGGGCTGGGG”, database “nucleotide collection (nr/nt)”, organism “bacteria (taxid:2)”. The search included 20,755,639 sequences and an effective sequence space used of 52,298,456,252 nt. Only hits from fully sequenced bacterial genomes with 100% identity were taken in account, all showing a bit score = 44.1 (bit score = 40.06 required) and an E-value = 0.003. Hits were then divided into groups according to their occurrence within protein coding sequences, being complementary to protein coding sequences or occurrence in untranslated regions. Similarly, the human genome were searched for the quadruplex motif using the following parameters: query sequence “GGGGCTGGGGCTGGGGCTGGGG”, database “Human genomic plus transcript (Human G + T)”. The search included 102,358 sequences and an effective sequence space used of 37,516,328,584 nt. Only hits from GRCh38 Primary Assembly with 100% identity were taken in account, all showing a bit score = 44.1 and an E-value = 0.002.