Supplementary Data

to

Crystallographic characterisation of CCG repeats

Agnieszka Kiliszek, Ryszard Kierzek, Wlodzimierz J. Krzyzosiak, Wojciech Rypniewski

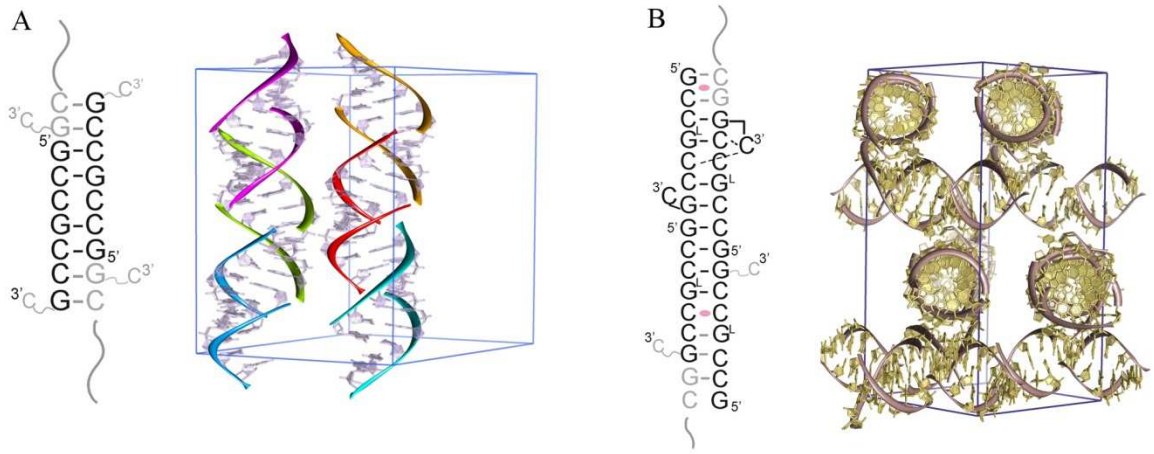
Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

Conformation of the 8C residues in the C+D duplex

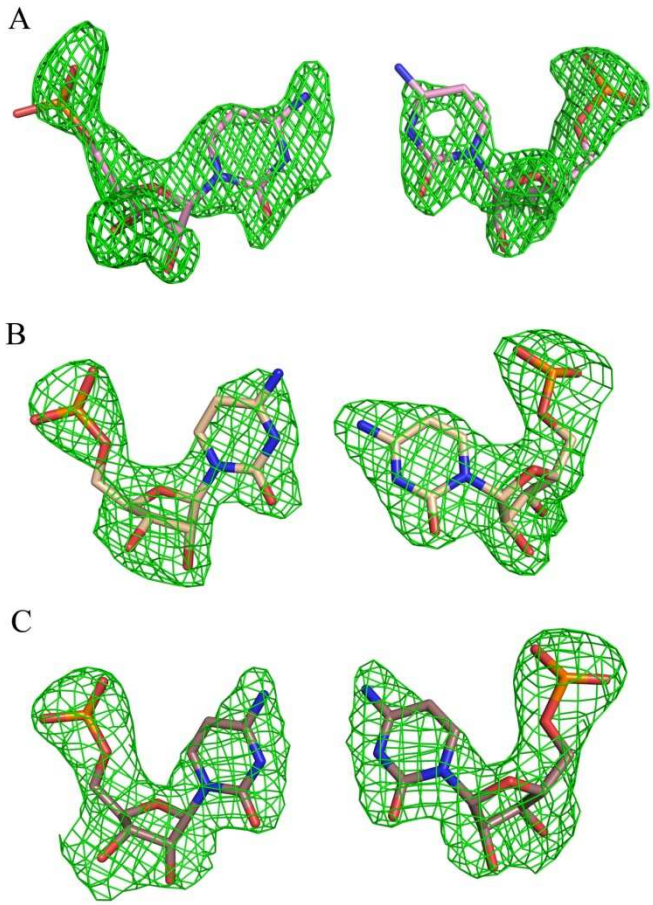
In the C+D duplex of the LNA-containing structure the 8C nucleotides at the 3'-end are flipped out of the double helix and fold back into the minor groove. The backbone between 7G and 8C bends sharply and reverses direction (Supplementary Figure S3A). Moreover, the conformation of adjacent guanosine residues is changed from the typical C3'-*endo* to C2'-*endo*, characteristic of the B-form. The ribose ring of 8C in chain C protrudes into the solvent space. The pyrimidine ring points into the minor groove and forms a bifurcated H-bond between its *exo*-amino group and two O2' atoms, one from 5C of chain D and the other from 6C of chain C (Supplementary Figure S3B). The other 8C is arranged differently. The '5' side' of the nucleotide fits on the surface of the minor groove. The cytidine interacts with the adjacent 7G residue *via* three H-bonds: between the O2 of cytidine and *exo*-amino group of G and between their sugar rings: O4'•••C1' and C4'•••O4' (Supplementary Figure S3C). Both the flipped cytidine residues interact and form intermolecular lattice contacts. They stack, but the pyrimidine rings are not parallel. The angle between the cytosine planes is approximately 60° which indicates stacking *face-to-edge*. In addition, a hydrogen bond is observed between O2' (C from chain D) and N4 (C from chain C). Two 8C-8C interactions, together with an H-bond formed between the O2' atoms of 5C residues, are involved in crystal contacts between symmetry-related duplexes (Supplementary Figure S3D).

Solvent interactions

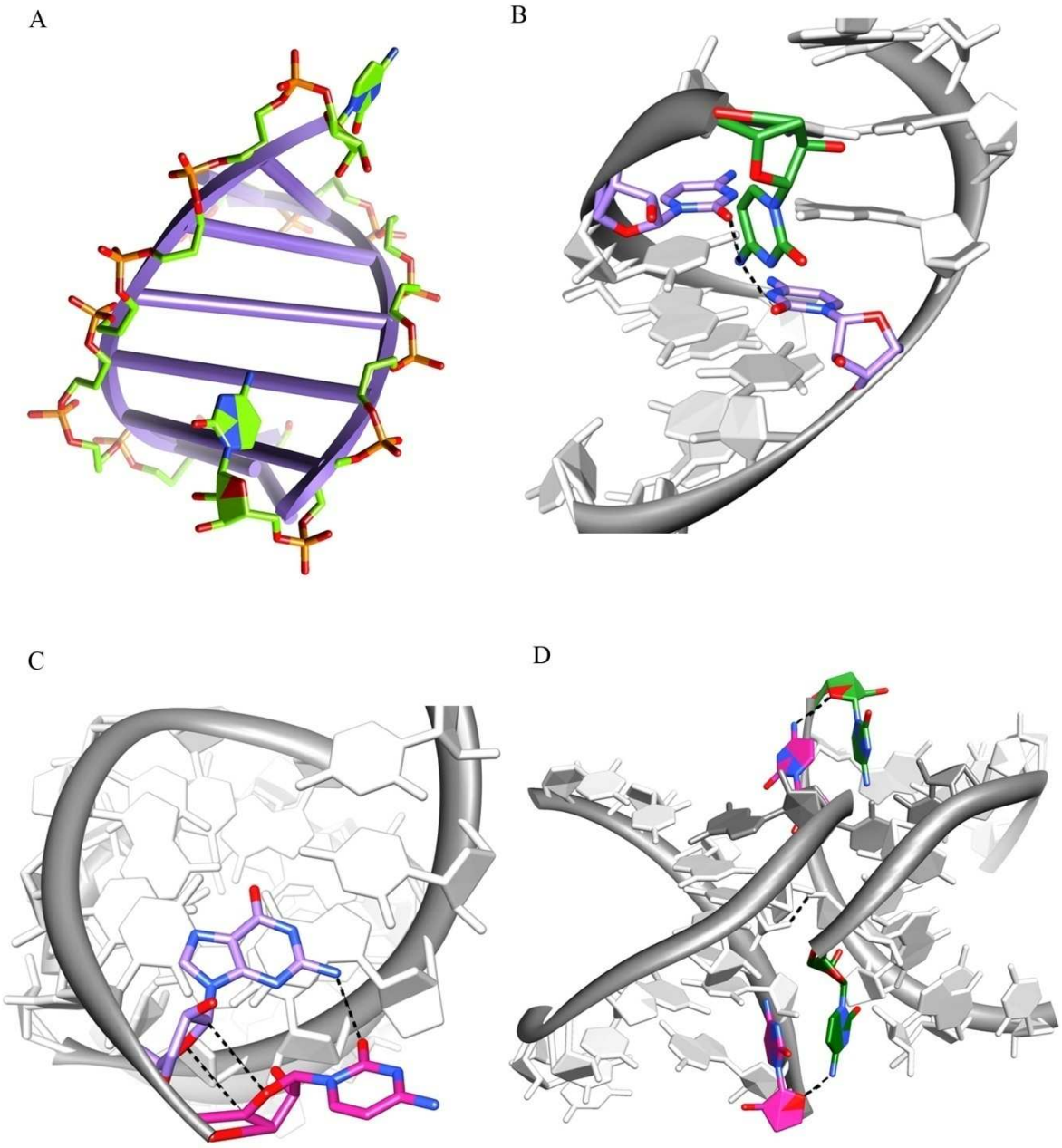
Each of the two crystallographic models contains 16 and 14 ordered water molecules, and the LNA-containing structure has in addition two sulphate ions. One interacts with the major groove of the C+D duplex and it forms hydrogen bonds with the *exo*-amino groups of 2C and 3C (chain C) and with one water molecule (Supplementary Figure S5A). The second sulphate ion is in the minor groove, at the backbone distortion associated with G^L. One H-bond is with the *exo*-amino group of the guanosine. Another oxygen atom of SO_4^{2-} interacts with N3 and with O2' of the methylene bridge. Some weak H-bonds are observed with O4' of adjacent 5G and with flipped 8C from chain C, and with a water molecule (Supplementary Figure S5B). The occupancy factor of the sulphate ions are 0.6 for each.



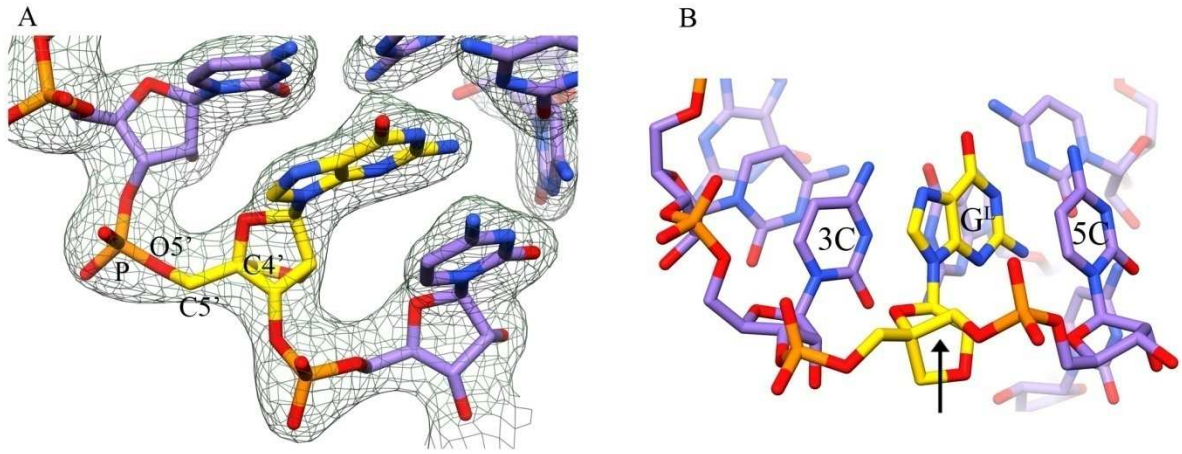
Supplementary Figure S1. Crystal packing of native duplexes in the rhombohedral unit cell (A) and LNA-containing duplexes in the tetragonal cell (B).



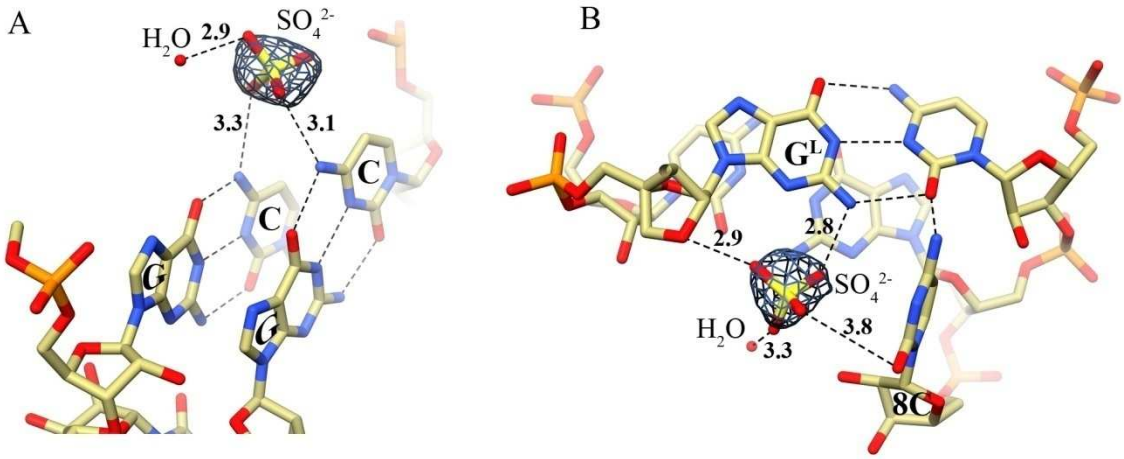
Supplementary Figure S2. The $F_o - F_c$ omit map of C-C pairs, contoured at the 2.9 σ level, for the duplexes: A+B (A), C+D (B) and E+E' (C).



Supplementary Figure S3. The terminal cytidine residues are flipped out of the helix, into the minor groove (A). The flipped 8C residue from strand C interacts with two citidine residues in the duplex (B), whereas the flipped 8C residues from strand D forms three H-bonds with the preceding guanosine residue (C). Contacts between symmetry-related duplexes in the LNA-containing structure involves stacking interactions between 5C and 8C residues and hydrogen bonds (D).



Supplementary Figure S4. The LNA residue ($4G^L$) of strand D in the $2F_o$-F_c electron density contoured at the 1σ level (A). The residue's torsion angles α, β and γ take unusual values and atoms C4', C5' O5' and the phosphorus atom are not co-planar and the phosphate group is shifted toward the neighbouring residue. As a result, the ribose ring of the LNA residue is shifted (arrow) toward the major groove (B).



Supplementary Figure S5. Sulphate ions in the LNA-containing structure, in the major groove (A) and in the minor groove (B). The $2F_o$-F_c electron density map is contoured at the 1σ level. The hydrogen bonding distance are given in Å.

Supplementary Table S1. X-ray data statistics.

Crystal	$(GCCGCCGC)_2$	$(GCCG^{L}CCGC)_2$
Beamline	BESSY BL 14.2	BESSY BL 14.1
Temperature (K)	100	100
Wavelength (Å)	0.9184	0.9184
Space group	R3:H	$P3_322$
Unit-cell parameters (Å)	a=b=40.6 c=56.1	a=b=42.0 c=81.7
Resolution range (Å)	20.0 – 1.54 (1.57- 1.54)*	20.0 – 1.95 (1.98 - 1.95)
R_{merge}†	0.04 (0.711)	0.067 (0.673)
No. unique reflections	5101	5762
Completeness (%)	99.9 (100)	99.8 (97.5)
Mosaicity (º)	0.27	0.8
Data redundancy	4.3 (3.8)	28.3 (15.6)
$\langle I/\sigma(I)\rangle$	26.14 (2.1)	23.7 (2.2)
Reflections > 2σ (%)	89.8 (54.6)	88.3 (54.0)
B-factor from Wilson plot ($Å^2$)	29.7	50.7

* Values in parenthesis are for the last resolution shell.

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the observed individual and mean intensities of a reflection with the indices *hkl*, respectively, $\sum_i$ is the sum over *i* measurements of a reflection with the indices *hkl*, and $\sum_{hkl}$ is the sum over all reflections

Supplementary Table S2. Helical parameters calculated using 3DNA, based on C1'- C1' vectors. The core of the helix (see main text) is marked in bold.

Pairs in native duplex	Pairs in modified duplex	Displacement (Å)			Angle (º)			Twist (º)			Rise (Å)		
		A+B	C+D	E +E'	A+B	C+D	E +E'	A+B	C+D	E +E'	A+B	C+D	E+E'
G-C	C-G	7.0	7.1	6.1	10.9	8.6	10.4	33.8	38.0	33.3	2.5	2.3	2.6
G-C	**C-G**	6.6	7.1	6.3	12.8	11.4	11.7	35.9	30.3	34.4	2.4	3.8	2.9
C-G	**G^L-C**	6.3	6.7	5.8	15.2	6.6	11.5	35.1	29.6	33.5	2.6	2.7	2.7
C-C	**C-C**	6.6	5.9	5.9	14.9	7.4	10.4	32.3	38.7	33.5	2.6	2.6	2.7
G-C	**C-G^L**	6.8	5.1	5.8	14.6	14.0	11.5	32.1	15.0	34.4	2.6	3.9	2.9
C-G	**G-C**	6.4	5.4	6.3	15.1	18.5	11.7	39.7	49.3	33.3	2.5	2.7	2.6
C-G	G-C	5.6	6.5	6.1	16.2	13.3	10.4						
average	**average**	**6.5**	**6.3**	**6.0**	**14.2**	**11.4**	**11.1**	**34.8**	**33.5**	**33.7**	**2.5**	**3.0**	**2.7**
s.d.	s.d.	0.5	0.8	0.2	1.8	4.2	0.6	2.8	11.5	0.5	0.1	0.7	0.1
avg. for core	**avg. for core**	**6.5**	**6.0**	**6.0**	**14.5**	**11.6**	**11.3**	**33.9**	**28.4**	**33.9**	**2.6**	**3.3**	**2.8**
s.d.	s.d.	0.2	0.8	0.2	1.0	4.9	0.5	1.9	9.8	0.6	0.1	0.7	0.1

Supplementary Table S3. Local base pair parameters calculated using 3DNA.

Pairs in native duplex	Pairs in modified duplexes	Propeller (º)			Buckle (º)			Opening (º)		
		A+B	C+D	E+E'	A+B	C+D	E +E'	A+B	C+D	E+E'
G-C	C-G	-14.6	-7.5	-13.9	-3.1	-4.2	0.8	8.3	-1.2	-1.2
G-C	C-G	-14.1	-1.2	-8.3	-3.5	-16.8	-5.3	-2.9	-5.9	-2.3
C-G	G^L-C	-11.6	-14.0	-18.7	-2.5	0.0	-5.3	-3.4	-1.5	-2.0
C-C	C-C	-12.7	-20.5	-20.4	0.5	-13.0	0.0	-11.5	-22.3	-7.9
G-C	C-G^L	-9.7	-13.8	-18.7	2.2	-11.2	5.3	3.3	-0.6	-2.0
C-G	G-C	-14.0	2.8	-8.3	3.3	23.2	5.3	-0.7	0.0	-2.3
C-G	G-C	-12.9	-9.9	-13.9	0.9	13.1	-0.8	0.5	-0.6	-1.2
average	**average**	**-12.8**	**-9.0**	**-14.7**	**-0.5**	**-3.7**	**0.1**	**-1.2**	**-5.3**	**-2.9**
s.d	s.d.	1.9	8.0	5.0	2.9	14.7	4.3	6.7	8.0	2.5

Supplementary Table S4. Roll calculated using 3DNA.

Steps in native duplex	Steps in modified duplexes	Roll (º)		
		A+B	C+D	E+E’
GG/CC	CC/GG	6.2	3.0	5.5
GC/GC	CG^{L}/CG	3.7	15.1	15.2
CC/CG	$G^{L}C$/CC	11.7	6.3	0.3
CG/CC	CC/$G^{L}C$	10.2	5.3	0.3
GC/GC	CG/CG^{L}	7.6	10.3	15.2
CC/GC	GG/CC	10.2	6.3	5.5
average	**average**	**8.3**	**7.7**	**7.0**
s.d.	s.d.	3.0	4.3	6.7

Supplementary Table S5. C-C pair present in the 30 models of the ribosomal subunit of *Deinococcus radiodurans* found by FRABASE.

PDB code	residue number	α (°)	β (°)	γ (°)	λ (°)	residue number	α (°)	β (°)	γ (°)	λ (°)	C1'-C1' (Å)
1J5A	1219	-67,6	170,8	53	56,9	1253	-292,5	200	279,8	11,9	11,1
1JZX	1219	-67,4	170,9	52,9	57,0	1253	-292,8	199,6	280,3	12,5	11,2
1JZY	1219	-67,4	171,3	52,5	57,2	1253	-293,2	199,1	280,5	13,1	11,2
1JZZ	1219	-67,5	170,6	52,8	56,8	1253	-292,5	198,9	279,9	12,1	11,1
1K01	1219	-67	170,9	52,2	57,1	1253	-292,9	198,2	280,4	13,1	11,2
1NJM	1219	-63,3	168,3	55,1	63,0	1253	-298,2	216,7	285,4	4,4	10,9
1NJN	1219	-63,8	168,3	55,2	63,3	1253	-298,8	216,9	285,5	4,7	10,9
1NJO	1219	-63,4	168,2	55	63,1	1253	-298,3	216,8	285,3	4,4	10,9
1NJP	1219	-63,4	168,3	55,1	63,1	1253	-298,3	216,8	285,4	4,5	10,9
1NKW	1219	-42,3	151,3	50,7	71,1	1253	-304,6	201,2	287,5	17,5	11,4
1NWX	1219	-46,6	152,5	53	70,4	1253	-302,6	200,5	285,5	15,8	11,5
1NWY	1219	-46,5	152,4	53	70,4	1253	-302,7	200,5	285,5	15,8	11,5
1OND	1219	-42,2	148,9	51,3	70,3	1253	-306,6	204	289,1	18,4	11,4
1P9X	1219	-69,4	177	51,7	59,0	1253	-303	201,8	282,6	12,3	11,8
1PNU	1219	-63,7	166,7	55,7	63,1	1253	-299	216,4	284,5	2,8	11,1
1PNY	1219	-63,7	166,7	55,7	63,1	1253	-299	216,4	284,5	2,8	11,1
1Y69	1219	-72,9	201,6	56,9	65,7	1253	-294,9	193	286,7	31,3	10,3
2AAR	1219	-68,2	182,4	51,7	65,8	1253	-288,4	183,6	282,1	26,0	10,8
2O43	1219	-51,4	178,4	45,2	57,2	1253	-163,2	153,6	162,8	25,5	11,5
2O45	1219	-55,5	191,8	49,8	50,6	1253	-186	229,7	163,7	32,3	11,2
2OGM	1219	-63,1	163	56,2	71,6	1253	-290,2	173,9	289,9	28,8	11,3
2OGN	1219	-50,7	173,4	49,1	55,3	1253	-356,9	133	39,3	38,0	11,0
2OGO	1219	-55,2	165,1	56,3	70,5	1253	-289,9	187,6	285,1	28,8	11,4
2ZJP	1219	-59,7	180,7	49,6	57,5	1253	-311	229,1	271,9	22,5	11,0
2ZJQ	1219	-66,1	185,8	49,9	62,1	1253	-304,6	187,1	290,4	27,6	11,0
2ZJR	1219	-65,8	183,3	49,6	60,3	1253	-321,4	208,5	286,8	29,4	11,0
3CF5	1219	-67,9	193	47,3	62,5	1253	-297	180,1	286,8	31,4	10,6
3DLL	1219	-88	193,5	57,8	64,3	1253	-270,5	186,4	254,1	19,2	11,7
3FWO	1219	-59,9	183,2	51,5	53,1	1253	-296,2	182,1	292	36,3	11,1
3JQ4	1219	-75,1	198,1	46,6	66,0	1253	-64,4	184,9	61,4	42,5	10,4
average		**-62,2**	**173,9**	**52,4**	**62,2**		**-283,7**	**197,2**	**259,8**	**19,5**	**11,1**
s.d.		10,0	13,7	3,2	5,7		54,0	20,6	65,0	11,5	0,3
standard values for A-RNA		-68,0	178,0	54,0			-68,0	178,0	54,0		

Supplementary Table S6. Comparison of characteristic features of double helical CCG , CGG (1), CAG (2) and CUG repeats (3).

Feature	CCG	CGG	CAG	CUG
Helix form	A	A	A	A
Helical twist* (°)	33.5-35 (core 28-34)	30-32	28.5 ± 5.7	33.6 ± 4.1
Major groove width† (Å)	15.0 (native), 15.8 (LNA-containing)	17.9 ± 0.9,17.8 ± 2.5, 14.3 (three models)	23.8 ± 0.2	12.7 ± 2.3
Minor groove width† (Å)	15.7 and 14.3	15.8 ± 0.5, 15.4 ± 0.5, 16.1	15.3 ± 0.1	15.7 ± 0.4
Average C1'-C1' distance for N-N§ (and for the other pairs) (Å)	10.7-10.9 (10.6)	11.3 (10.7)	11.0 (10.7)	10.4 (10.5)
Local effect of N-N on helicity	Not observed or local unwinding compensated elsewhere along the duplex	Local unwinding, compensated elsewhere along the duplex	Unwinding	Not observed
N-N pairing interaction	None or one hydrogen bond *exo* -N4···N3	N1H···O6 and *exo* -N2H···N7 and *intra* -molecular: *exo* -N2H···O2	C2-H2···N1 hydrogen bond	N3-H3···O4hydrogen bond
Manner of accommodating N-N (according to direction of the glycosidic bond, λ)	Not observed or elevated propeller value of C-C pair	One G is in *syn* conformation (inclined towards the major groove)	One A turned towards major groove (“thumbs-up”)	One U inclined towards minor groove
Effect of N-N conformation on neighbouring N-N	Not determined	Symmetric arrangements are clearly favoured	Co-operativity: A-A pairs in consecutive repeats have alternative conformations	No effect: each U-U takes one of two possible conformation independently
Electrostatic profile	Negtive potential around the C-C pairs; alternating stripes of positive and negative potential due to C-G pairs	Alternating stripes of positive and negative potential due to C-G pairs	Alternating stripes of positive and negative potential due to C-G pairs	Alternating stripes of positive and negative potential due to C-G pairs
Observed ligand affinity of N-N	Not observed	Sulphate or Ca^{2+} binding in major groove	Sulphate binding in major groove	Sulphate or glycerol through-water binding in major groove
Exposed functional groups of N-N:				
- major groove	None or N4 *exo* -amino of one C	G(anti) O6 carbonyl G(*syn*) O6 carbonyl, N1 amino, N2 *exo* -amino	First A N6 amino Second A N1 imino, N6 amino	First U O4 carbonyl
- minor groove	None or O2 carbonyl of one C and O2 carbonyl, N3 amino of second C	G(*anti*) N3 imino, N2 *exo* -amino	First A N3 imino Second A N3 imino	First U O2 carbonyl Second U O2 carbonyl, N3 amino

* Average for A-RNA is 33.1 (4).

† The values given are the “refined” widths, according to the program 3DNA (5).

§ G-G or A-A or U-U base pair.

References:

1. Kiliszek, A., Kierzek, R., Krzyzosiak, W.J. and Rypniewski, W. (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res.*, 39: 7308–7315.

2. Kiliszek, A., Kierzek, R., Krzyzosiak, W.J. and Rypniewski, W. (2010) Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res.*, 38, 8370-8376.

3. Kiliszek, A., Kierzek, R., Krzyzosiak, W.J. and Rypniewski, W. (2009) Structural insights into CUG repeats containing the 'stretched U-U wobble': implications for myotonic dystrophy. *Nucleic Acids Res.*, 37, 4149-4156.

4. Bloomfield,V.A., Crothers,D.M. and Tinoco,I. (2000) Nucleic Acids: Structures, Properties, and Functions. University Science Books, Sausalito.

5. Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z. *et al*. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, 313, 229-237.