Supplementary Materials

to

Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases.

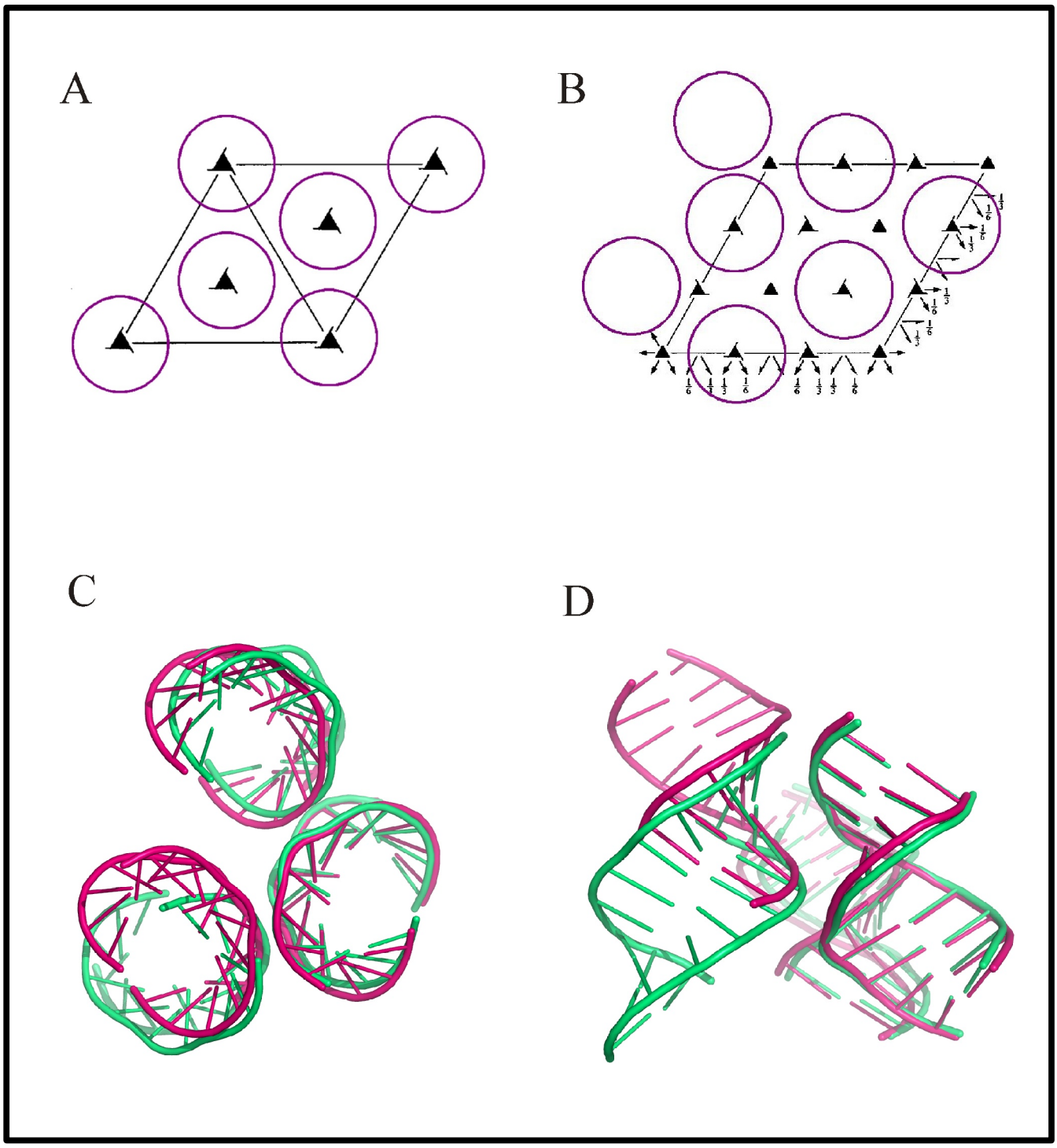
Agnieszka Kiliszek, Ryszard Kierzek, Wlodzimierz J. Krzyzosiak, Wojciech Rypniewski

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

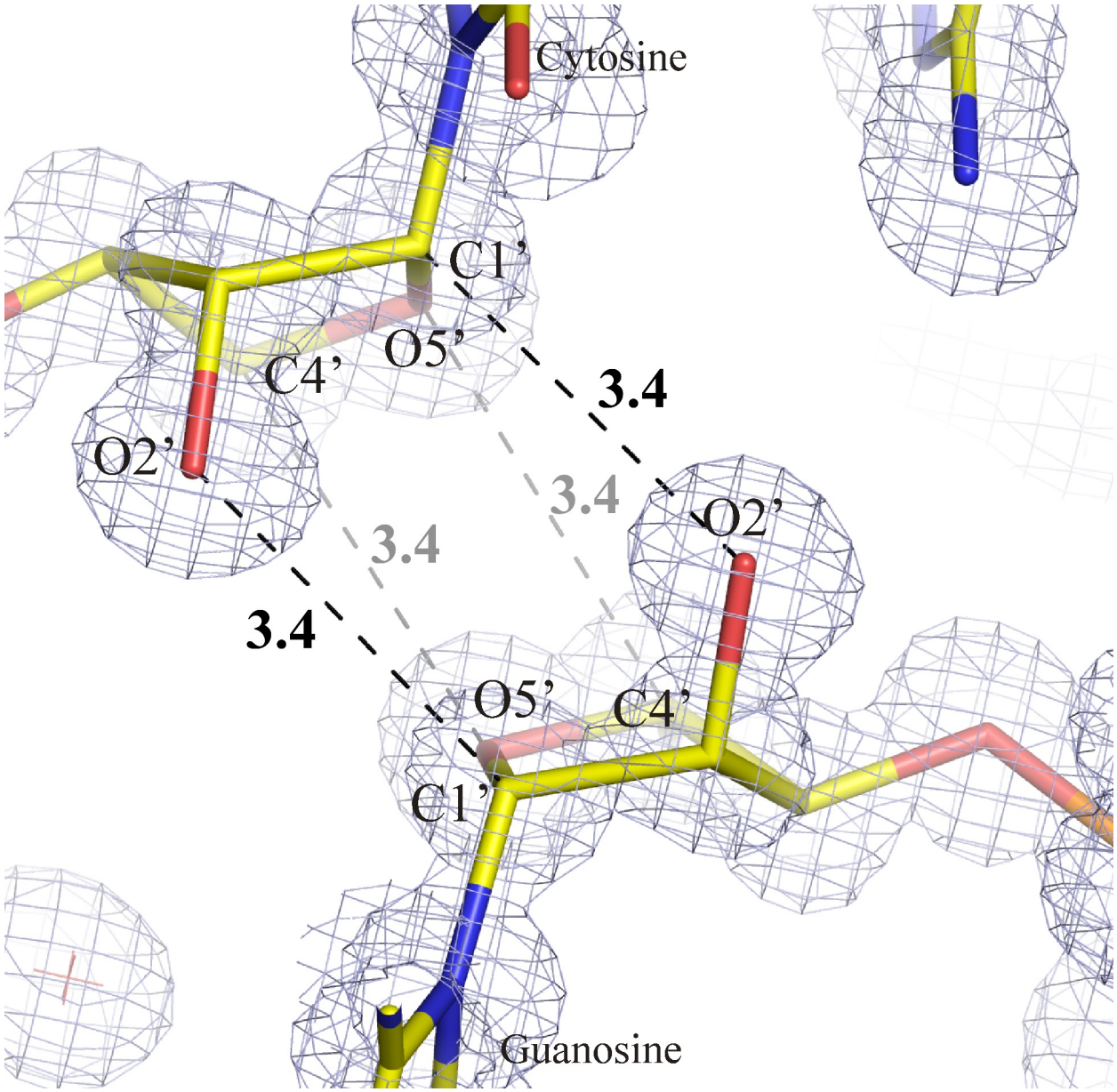
Crystal packing

Both crystal forms were obtained from the same oligomer r(GGCAGCAGCC) under similar conditions. Although the space groups were different, the cell parameters were almost the same and the packing of the molecules also appeared very similar. In the rhombohedral structure, the asymmetric unit contains one RNA strand (chain H). The second strand of the duplex is symmetry-related *via* a crystallographic two-fold axis. The duplexes thus formed stack along the the 3_2 axes generated between three 3-fold axes. The screw axes are coincidental with the helical axes. In the trigonal structure, the asymmetric unit contained tree duplexes: A+B, C+D, E+F. The three helices were also wrapped around the 3_2 crystallographic axes: C+D around the axis along the unit cell edge, A+B and E+F around the two axes in the body of the unit cell (Supplementary Figure 1). In both crystal forms, the duplexes stack end-to-end, forming semi-infinite columns parallel to the *c* cell edge. When the two crystal lattices are compared, the RNA columns appear to superimpose very closely. The difference is that one of them, consisting of duplexes C+D, is shifted and rotated by half a helical turn (5 base pairs) relative to the corresponding duplex in the rhombohedral lattice. The r.m.s. deviations in atomic coordinates between the high resolution structure and the helices in the trigonal structure is 0.9-1.0 Å and the three duplexes in the trigonal structure can be superposed with r.m.s.d. 0.1-0.5 Å.

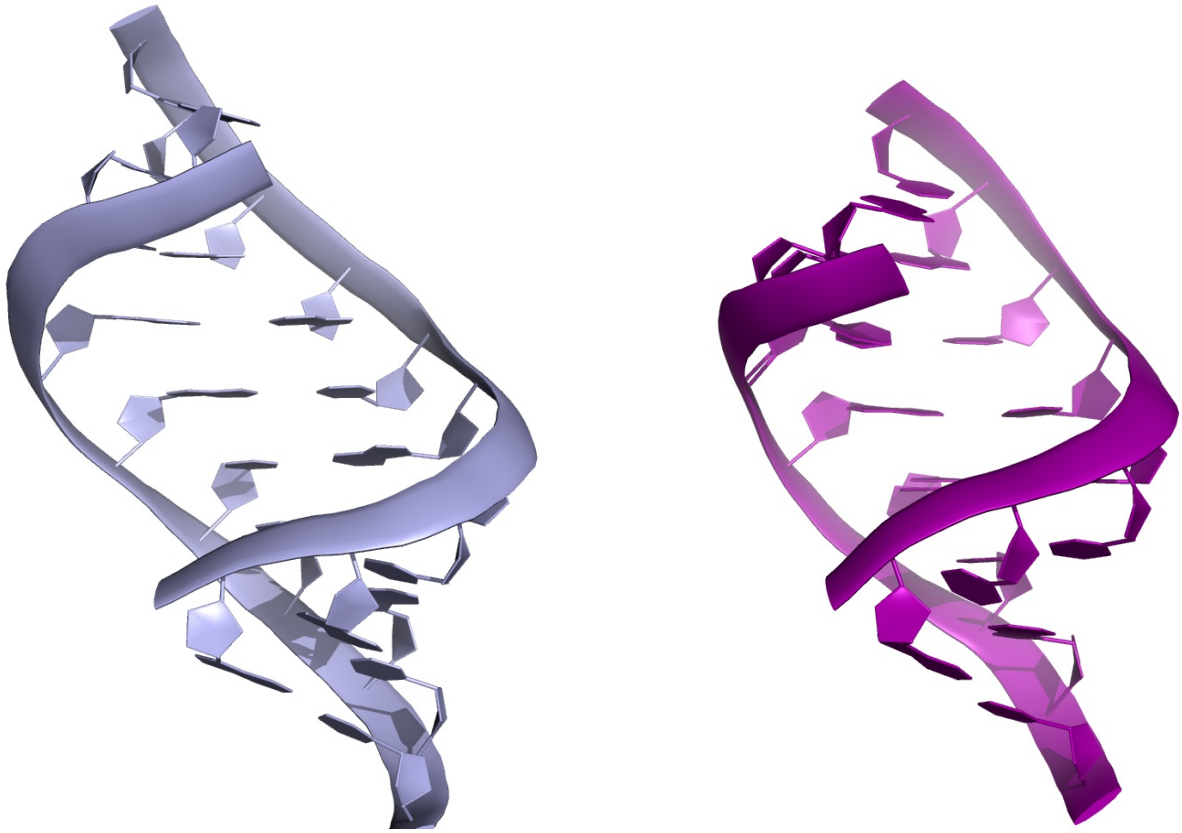
Supplementary Figures and Tables:



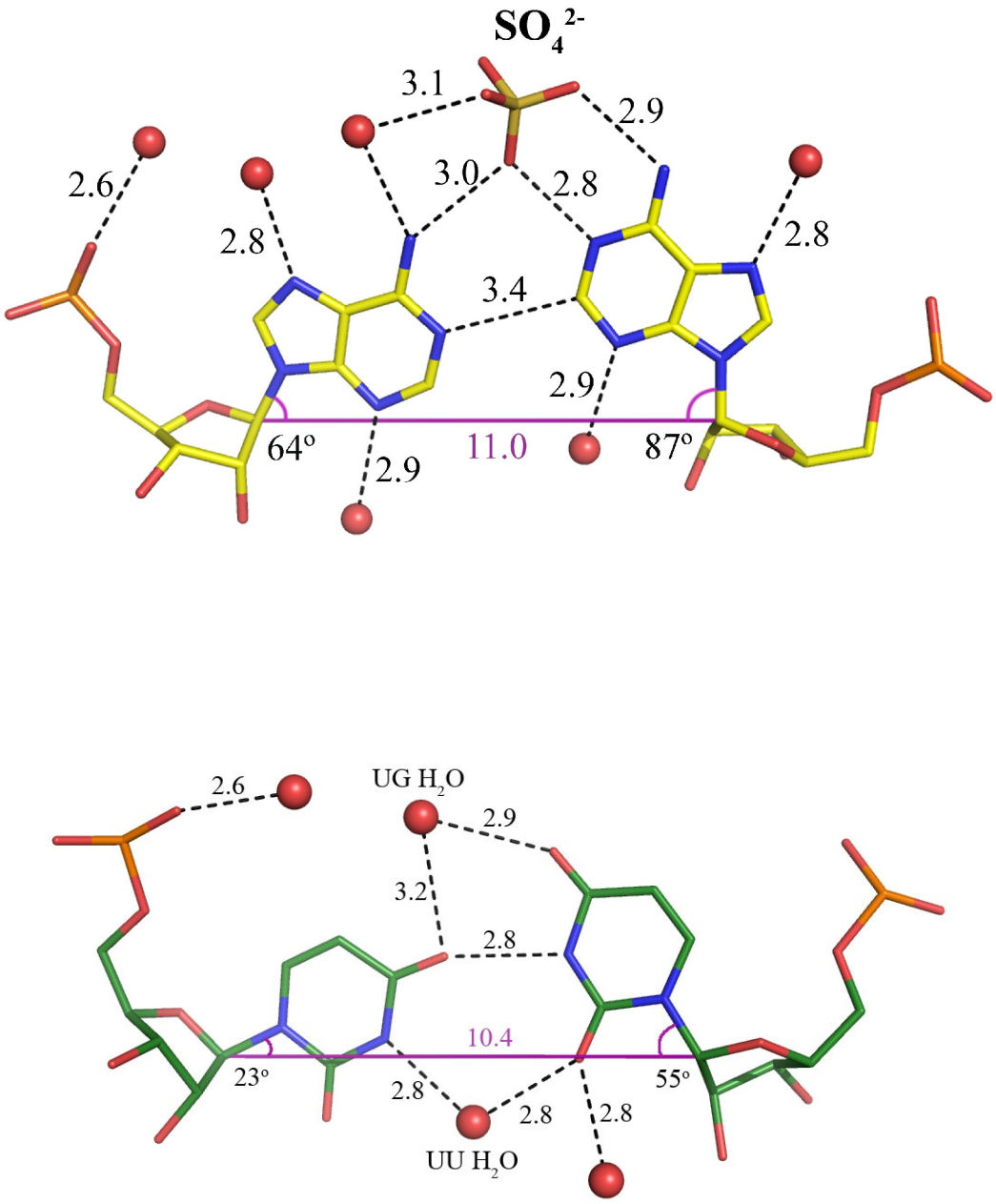
Supplementary Figure 1. Scheme of crystal packing of GGCAGCAGCC duplexes along the helix axes in (A) the trigonal and (B) the rhombohedral lattice. The RNA duplexes form columns represented by violet circles. Although the space groups are different, the packing of molecules shows similarity. (C) and (D) show perpendicular views of the two models superimposed. In the trigonal (*pink*) and rhombohedral (*green*) structure two neighbouring columns can be superimposed while in the third the duplexes are screw-rotated by half a turn.



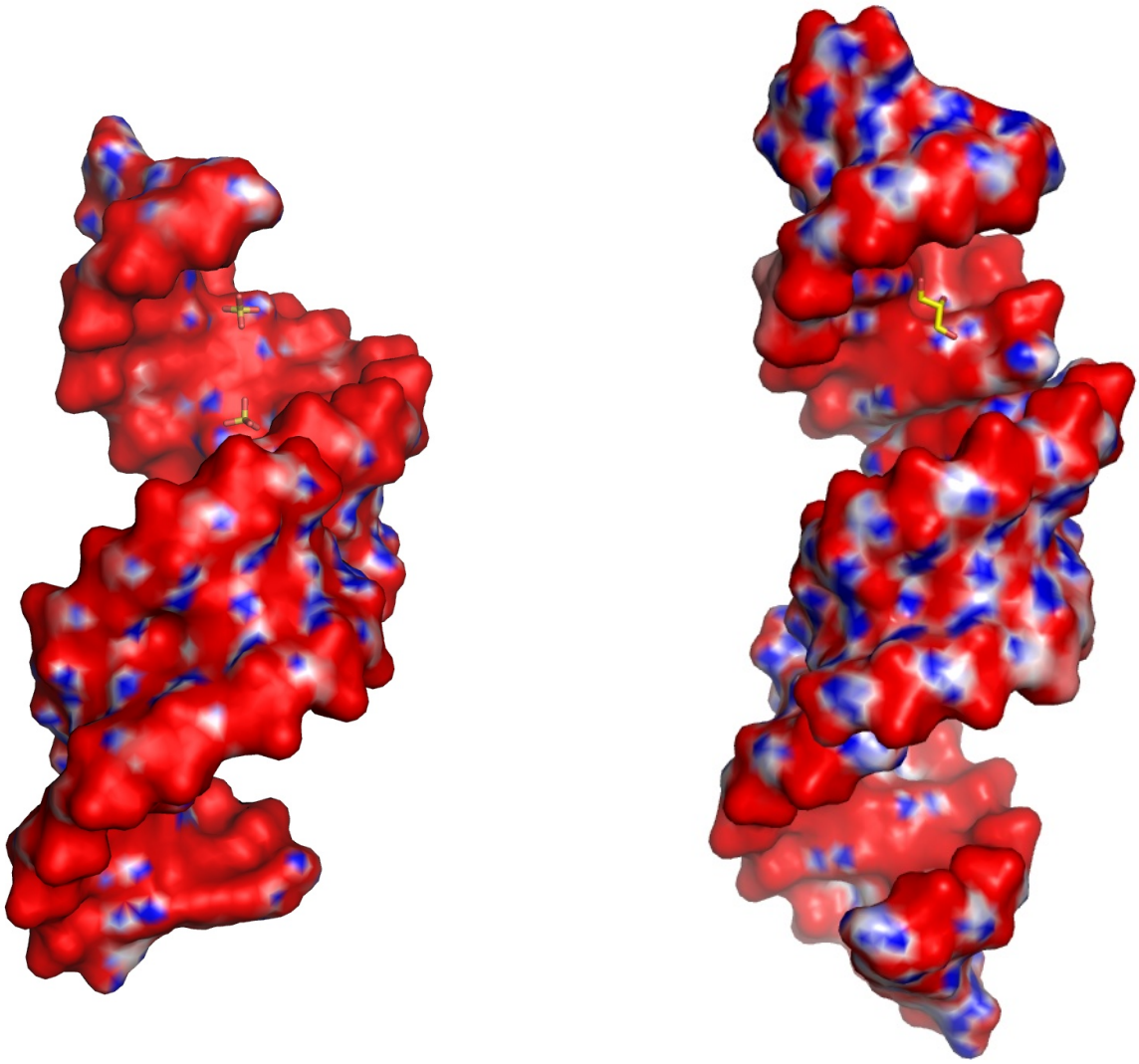
Supplementary Figure 2. Ribose-ribose interactions between two symmetry-related $(GGCAGCAGCC)_2$ molecules ion the rhombohedral structure. The four C-H···O bonds are pairwise symmetric. Donor-acceptor distances are marked in Å. The contours correspond to the $2F_o$-F_c electron density at the 1 σ level.



Supplementary Figure 3. A comparison of the major groove width in the CAG- and CUG-containing duplexes: $(GGCAGCAGCC)_2$ (*left*) and $(GCUGCUGC)_2$ (*right*).



Supplementary Figure 4. A comparison of A-A (*top*) and U-U (*bottom*) base pairs and their hydration patterns.



Supplementary Figure 5. A comparison of the electrostatic potential surfaces of CAG- and CUG-containing duplexes: two stacked $(GGCAGCAGCC)_2$ (*left*) and three $(GCUGCUGC)_2$ (*right*).

Supplementary Table 1. X-ray data statistics.

Crystal	$(GGCAGCAGCC)_2$	$(GGCAGCAGCC)_2$
Beamline	BESSY BL 14.1	EMBL-X13
Temperature (K)	100	100
Wavelength (Å)	0.8000	0.81230
Space group	R32:H	$P3_2$
Unit-cell parameters (Å)	$a=b=46.4$ $c=82.7$	$a=b=47.4$ $c=83.2$
Resolution range (Å)	23.2 – 0.95 (0.97- 0.95)*	18.6-1.9 (1.93 - 1.9)
R_{merge}†	0.036 (0.291)	0.105 (0.518)
No. unique reflections	21756	16346
Completeness (%)	99.9 (98.0)	99.8 (100)
Mosaicity (º)	0.34	0.7
Data redundancy	15.3 (11.0)	6.4 (6.2)
$\langle I/\sigma(I)\rangle$	74.1 (7.7)	18.2 (4.4)
Reflections > 2σ (%)	94.4 (72.9)	81.6 (81.9)
B-factor from Wilson plot ($Å^2$)	6.8	25.7

* 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 2. Helical parameters calculated using 3DNA, based on C1'-C1' distances.

Bp	*Displacement (Å)*				*Angle (º)*				*Twist (º)*				*Rise (Å)*			
	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F
G-C	7.8	8.8	8.8	8.3	13.1	10.8	11.8	13.1	30.0	31.3	32.9	33.2	2.7	2.8	2.7	2.7
G-C	8.7	10.0	10.3	9.8	8.3	7.4	8.4	8.7	30.9	30.6	31.5	32.1	3.0	3.1	3.1	3.1
C-G	9.0	10.2	10.4	10.0	8.0	6.1	7.0	7.4	33.7	30.5	29.6	28.3	2.9	3.0	2.9	3.0
A-A	7.5	8.3	8.7	8.6	10.7	9.0	9.5	9.5	18.8	17.8	17.8	22.4	3.0	2.7	2.8	2.9
G-C	7.4	8.1	8.5	8.4	10.4	10.7	10.2	9.9	33.4	26.0	29.2	25.8	2.4	2.5	2.4	2.4
C-G	7.4	8.4	8.5	8.1	10.4	10.0	10.1	10.7	18.8	22.1	17.9	17.6	3.0	2.9	2.9	2.7
A-A	7.5	8.7	8.7	8.3	10.7	9.4	9.8	9.1	33.7	28.2	28.9	30.4	2.9	3.0	2.9	2.9
G-C	9.0	9.9	10.4	10.1	8.0	7.5	7.2	6.1	30.9	32.6	31.9	30.8	3.0	3.0	3.1	3.1
C-G	8.7	9.8	10.3	10.0	8.3	8.6	8.2	7.2	30.0	33.0	33.0	31.6	2.7	2.7	2.7	2.8
C-G	7.8	8.3	8.8	8.7	13.1	12.9	11.6	10.3	-	-	-	-	-	-	-	-
Average	8.1	9.0	9.3	9.0	10.1	9.2	9.4	9.2	28.9	28.0	28.1	28.0	2.8	2.9	2.8	2.9
s.d.	0.7	0.8	0.9	0.8	2.0	2.0	1.7	2.0	5.9	5.2	6.0	5.2	0.2	0.2	0.2	0.2

Definitions of displacement, angle, twist, rise and opening according to Dickerson *et al.*, J. Mol. Biol. (1989) 205, 787-791.

H and H' are related by crystallographic symmetry

Supplementary Table 3. Sugar, backbone and glycosidic torsion angles* calculated for the $(GGCAGCAGCC)_2$ structures. Successive rows correspond to chains: H of the rhombohedral structure and A, B, C, D, E , F for the trigonal structure. Bold numbers are average values and in brackets are standard deviation values.

	α	β	γ	δ	ε	ζ	χ
1G	---	---	54.8	80.3	-155.1	-68.8	-178.5
	---	---	62.9	78.6	-153.3	-69	-179.2
	---	---	57.4	78	-155	-73.8	177.8
	---	---	58.5	80.6	-153.8	-72.3	-179.7
	---	---	60	83.8	-148.6	-76.5	179.2
	---	---	56.2	79	-156.3	-75.2	178
	---	---	63.9	77.5	-147.8	-74.2	179.3
			59.1 (3.4)	**79.7 (2.1)**	**-152.8 (3.3)**	**-72.8 (3.0)**	**-180.4 (1.4)**
2G	-64.5	177.2	59.2	81.1	-145.3	-67.5	-166.9
	-69.5	179.8	54.1	87	-148.3	-69.6	-163.1
	-70.3	177.7	58.4	84.4	-141.8	-71.7	-167
	-70.3	-179.8	52.9	84.7	-144.1	-71.3	-162.4
	-70.9	178.7	56.8	83.4	-141.3	-72.4	-161.5
	-73.2	175.2	60	85.3	-142.3	-75.2	-166.3
	-66.5	179.5	52.7	87.7	-143.7	-73.7	-163.1
	-69.3 (2.9)	**178.3 (1.8)**	**56.3 (3.1)**	**84.8 (2.21)**	**-143.8 (2.4)**	**-71.6 (2.5)**	**-164.3 (2.3)**
3C	-69	175.8	51.2	76.4	-154.2	-64.7	-163.3
	-67.5	172.3	49.8	80.9	-151.3	-70.6	-166.2
	-66.8	175.7	49.2	79.4	-148.5	-68.5	-164.9
	-69.4	174	50.8	83.4	-146.1	-70	-165.7
	-72	174.4	54.9	82.2	-146.3	-68.5	-164.5
	-64.1	173.2	48.2	79.2	-149	-65.9	-167.3
	-63.8	172.2	46.6	81.9	-153.4	-67	-165.8
	-67.5 (2.9)	**173.9 (1.5)**	**50.1 (2.6)**	**80.5 (2.3)**	**-149.8 (3.2)**	**-67.9 (2.1)**	**-165.4 (1.3)**
4A	-64.6	171.8	57.3	77.1	-161.2	-72.6	-161.9
	-67	169.2	57.6	82.5	-169.7	-80.4	-163.8
	-69.5	169.9	60	80.2	-159.4	-74.4	-165.6
	-67.6	171.8	58	80.3	-165.9	-79.7	-165.9
	-73.1	170.6	62.8	80.5	-165.5	-69.9	-167.2
	-70.5	169.8	60.9	79.8	-161.6	-87.9	-162.2
	-77.3	171.8	65	81.1	-166.9	-67.9	-165.9
	-69.9 (4.2)	**170.7 (1.1)**	**60.2 (2.9)**	**80.2 (1.6)**	**-164.3 (3.7)**	**-76.1 (7.0)**	**-164.6 (2.0)**
5G	153.2	-174.4	179.5	81.2	-151.7	-71.1	-175.5
	135.2	-178.1	-160.7	92.9	-145.8	-78.3	-174.9
	-96.9	175.3	84	68.9	-159.2	-74.8	-174.6
	149.3	178.1	-171.7	90.9	-147.8	-74.5	-177
	-114	-176.4	97.4	67.8	-155.8	-79.1	-178.8
	-221.4	178	-163.9	97	-148.2	-70.2	-171.5
	138.6	-173.2	93.9	66.9	-160.7	-78.9	-179.3
	-169.9 (58.0)	**181.3 (4.3)**	**148.4 (53.5)**	**80.8 (13.0)**	**-152.7 (5.9)**	**-75.3 (3.7)**	**-175.9 (2.7)**
6C	-62.4	176.5	56.5	82.5	-136.7	-69.5	-165.8
	-64.4	172.4	66.9	73.2	-145.1	-66.3	-170
	-57.1	170.4	54.8	80.9	-143.4	-72.8	-164.5
	-60.7	176.8	53.4	76.4	-150	-63.9	-170.8
	-58.6	176.8	50.6	79.8	-146.6	-68.5	-169.5
	-64.9	176.1	60.1	79.4	-145	-71.1	-165.8
	-66.4	175.5	65.5	76.9	-147.9	-65.5	-166.3
	-62.1 (3.4)	**174.9 (2.5)**	**58.3 (6.2)**	**78.4 (3.1)**	**-145.0 (4.2)**	**-68.2 (3.2)**	**-167.5 (2.5)**

	α	β	γ	δ	ε	ζ	χ
7A	-65.1	163.6	52.6	78.9	-175.9	-87	-160.2
	-66	171.6	50.3	80.3	-169.7	-84.2	-161.5
	-57.1	170.2	48.6	75.6	-178.8	-82.5	-166.9
	-68.8	175.4	50.7	81.2	-177.9	-83.1	-163.9
	-63.5	172.2	51.5	79.2	-175.1	-82.2	-166.3
	-61.3	173.8	48.8	80.1	178.7	-82.8	-165.4
	-68.7	170.9	51.7	78.8	-170.9	-79.6	-165.2
	-64.4 (4.2)	**171.1 (3.7)**	**50.6 (1.5)**	**79.2 (1.8)**	**-175.7 (4.2)**	**-83.1 (2.2)**	**-164.2 (2.5)**
8G	-64.9	-175	50.2	76.2	-151.4	-66.7	-157.4
	-60.4	179.6	48.5	78.9	-151.8	-77	-162.3
	-68.6	-176	54.7	76.8	-152	-67.1	-164.3
	-67.1	-175.6	55.7	77.7	-151.6	-74	-162.2
	-67.3	-175.9	55	79.8	-150.6	-75.3	-163.1
	-71.3	-174.1	55.7	75.3	-152	-66.8	-163.8
	-65.7	-179.1	51.6	78.1	-153	-75.1	-159.9
	-66.5 (3.4)	**183.4 (2.3)**	**53.1 (2.9)**	**77.5 (1.6)**	**-151.8 (0.7)**	**-71.7 (4.6)**	**-161.9 (2.4)**
9C	-66.1	171	55.6	76.4	-147.2	-66.1	-162.4
	-63.2	168.5	48.5	80.6	-143.8	-69.9	-161.4
	-70.4	167.6	55.3	77.8	-149.1	-66.4	-161.4
	-65.8	168.9	52.5	82.1	-145.5	-68.8	-159
	-61.9	166.4	49.7	80.1	-146.7	-65.7	-159.3
	-69.6	167.8	53.5	80	-149.8	-68.7	-162
	-62.7	167.1	49.9	76.3	-148.2	-64.3	-160.3
	-65.7 (3.3)	**168.2 (1.5)**	**52.1 (2.8)**	**79.1 (2.2)**	**-147.2 (2.1)**	**-67.1 (2.0)**	**-160.8 (1.3)**
10C	-71.1	175.2	58.4	73.9	---	---	-162.4
	-65.4	179.8	52.5	80.9	---	---	-162.3
	-75.4	178.6	57.4	79.8	---	---	-167.4
	-68.5	178.4	53.4	84.1	---	---	-166.2
	-71.3	-179.9	51.3	83.1	---	---	-164.6
	-69.7	184.5	55.4	77.8	---	---	-167.7
	-64.2	-178.4	46.1	84.7	---	---	-161
	-69.4 (3.8)	**178.5 (2.4)**	**53.5 (4.15)**	**80.6 (3.8)**			**-164.5 (2.7)**

* P $^{\alpha}$ O5′ $^{\beta}$ C5′ $^{\gamma}$ C4′ $^{\delta}$ C3′ $^{\varepsilon}$ O3′ $^{\zeta}$ P

Supplementary Table 4. Minor and major groove widths * according to: direct P-P distances / refined distances which take into account the directions of the sugar-phosphate backbones.

Step	*Minor groove width (Å)*				*Major groove width (Å)*			
	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F
GG/CC	---	---	---	---	---	---	---	---
GC/GC	---	---	---	---	---	---	---	---
CA/AG	17.3/--	17/--	17.3/--	17.3/--	18.5/--	20.7/--	20.2/--	19.5/--
AG/CA	16.9/15.4	16.5/14.9	16.7/15.2	16.8/15.4	22.2/21.4	24.1/24.0	24.4/24.3	24/23.9
GC/GC	16.6/15.1	16.3/14.9	16.4/14.9	16.3/14.9	22.8/22.2	25.2/25.2	26/26.0	25.3/25.3
CA/AG	16.9/15.4	16.8/15.4	16.7/15.1	16.4/14.9	22.2/21.4	24.0/23.9	24.5/24.4	24.2/24.2
AG/CA	17.3/--	17.3/--	17.2/--	17/--	18.5/--	19.5/--	20.2/--	20.7/--
GC/GC	---	---	---	---	---	---	---	---
CC/GG	---	---	---	---	---	---	---	---

* El Hassan and Calladine (1998)