MATERIALS AND METHODS

Purification of RNA. Deprotected and desalted RNA was purchased from Integrated DNA Technologies, Inc. The RNA was dissolved in water and purified by HPLC on a Waters HPLC instrument with an attached UV-Vis detector that monitored absorbance at 220 and 254 nm. A gradient of 10 mM triethylammonium acetate, pH 7.0 (100% to 0%) in acetonitrile over 55 min with a flow rate of 2 ml/min was applied to an X Terra Prep MS C18 column (7.8 ×150 mm, 5 μ m); t_R = 25 min. Fractions containing RNA were lyophilized, dissolved in DEPC-treated water, and desalted by using a Sephadex PD-10 prepacked size exclusion column. Fractions containing RNA were combined and lyophilized. The RNA sample was re-dissolved in DEPC-treated water, and the concentration determined by its absorbance at 260 nm at 95 °C. Molar extinction coefficients were determined by using the Hyther server (Nicolas Peyret and John SantaLucia Jr., Wayne State University, Detroit, MI), which uses parameters based on molar absorptivity of RNA nearest neighbors.(1)

Crystallization of r(CGG) Oligonucleotide. A 1.2 mM solution of RNA duplex (Figure 1a) was prepared in DEPC-treated water. The sample was then annealed at 60 °C for 5 min and left to cool to room temperature by placing the sample on the bench top. A Qiagen Nucleix Suite kit was used to screen conditions that provided high quality crystals using the sitting drop method. Initially, drops contained 0.2 μ L of RNA and 0.2 μ L of the kit's reservoir solution. The conditions that provided the highest quality crystals used a reservoir solution that contained 50 mM magnesium acetate, 50 mM sodium cocadylate, pH 6.5 and 1.3 M lithium sulfate. Crystals appeared after 5 days at a temperature of 18 °C.

Data Collection, Structure Determination and Refinement. Crystals used for data collection were flash frozen by immersion in liquid nitrogen. Diffraction data were collected at beamline 11-1 at the Standford Synchroton Radiation Lightsource (SSRL) under cryoconditions at temperature (100 K) using the MARmosaic 325 CCD detector. Data were processed and scaled using HKL2000(2). An RNA model to fit the diffraction data was manually built with 17 base pairs of double-stranded RNA using

S-2

Coot.(3) The position of the starting model was found by molecular replacement using Molrep (4). The atomic model was refined with PHENIX (5). The structures were refined to 1.4 Å. Statistics for data collection, processing, and refinement are given in Table S-1.

Calculation of structural parameters. Helical parameters, groove widths and torsion angles were calculated using the program 3DNA (*6*). To avoid computational artifacts arising from the non-canonical base pairing, sequence in dependent measures was used based on vector connecting the C1' atoms.

Calculation of Electrostatic Potentials. The Protein Data Bank (PDB) file 3QIQ was taken for the CUG RNA structure (*7*) while AU and CG paired standard duplex RNA were rebuilt using Amber topology parameters. Hydrogen atoms were added to the biomolecules and positioned based on a previously described algorithm.(*8*) After the construction, hydrogen atoms were checked for steric conflicts. Atom partial charges and atomic radii were assigned based on AMBER force-field using the program PDB2PQR.(*9*) Surface electrostatic potential for the RNA models was calculated with Adaptive Poisson-Boltzmann Solver (APBS) - Software.(*10*) For APBS calculations, the RNA molecule was treated as a low dielectric medium within the volume enclosed by its solvent-accessible surface (probe radius = 1.4 Å). A dielectric constant of 2 was used to account for the electronic polarizability effects. The surrounding solvent was treated as a continuum with a dielectric constant of 80. The ion-exclusion radius of 2.0 Å was added to account for ion size on the RNA molecule surface. Ten grid points per square angstrom were used to construct molecular surfaces.

Electrostatic calculations for all the structures were completed at 298 K. In order to calculate the electrostatic surface potential, a sequential focusing multi-grid method was used. This involves solving the equation using a coarse grid, which is then refined to provide a more accurate, finer grid using Dirichlet boundary condition.(*11*)

S-3

Crystal Packing. The 19-mer RNA duplex r (<u>UU</u>GGGC(CGG)₃GUCC)₂ studied herein, has continuous base-paired nucleotides with the central r(CGG) motifs and terminal 5'UU dangling end that forms pairs with an adjacent duplex (Figure S-1). Thus, in the monoclinic structure, the asymmetric unit contains one double stranded RNA that is symmetry-related via a crystallographic two-fold axis with oliomgers stacked end to end, forming infinite antiparallel A-form double helices (Figure S-1). The dangling UU ends form two hydrogen bonded pairs with the dangling UU ends of a neighboring strand, or a 'UU zipper'.(*12*)



Figure S-1: The crystal contact with the neighboring RNA strand showing hydrogen bonding between a 5' UU dangling end with a 5' UU dangling end on another helix.

General features of the structure: Local unwinding around the syn-G's in the GG pairs. Accommodation of purine-purine pairs leads to some local distortions are observed near noncanonically paired syn-G8 and G11 (Figure S-2). This is ascertained by the distortion in the α - and Ytorsional angles of 5'phosphodiester bond of G8 and G11. For example, the α -torsional angles have average values of 148.3°, which is significantly different than standard angle of ca. -60°. The Y-torsional angles are -179° for the *syn*-G's, which is significantly different than the standard angles of ca. 55°. These differences are illustrated in Figure S-3.

Using C1'-C1' vector, sequence independent helical parameters were calculated. Displacement of the middle C1'-C1' from the helix, inclination between this vector and helix, and helical rise by projection of this vector onto the helical axis changes significantly to accommodate the non-canonical GG pairs in this structure as compared to duplex RNA.



Figure S-2: A full view of RNA structure studied herein, also showing the distortion near the GG pair on the backbone.



Figure S-3: Stereo view of the backbone showing α - (148.3°) and Y- (179°) torsional angles of the phosphodiester bond between between C7 and G8 in a 5'C<u>G</u>G/3'G<u>G</u>C motif. The torsional angles have values that are indicative of local unwinding of the helix relative to standard A-form RNA. The arrows indicate positions of local unwinding.

Table S-1: Data collection, phasing, and refinement								
statistics								
Data collection								
Space group	С2							
Cell dimensions								
a, b, c [Å]	a=55.660, b=32.320,							
α, β, γ	α=γ=90.00, p=108.95							
λλ/autologeth [Å]	0.07855							
wavelength [A]	0.97855							
Resolution [Å]	50.00-1.36							
hr.c.	(1.41-1.36)							
R _{sym} or R _{merge} ~[%]	2.6 (25.0) ^a							
Ι/σΙ	32.52 (3.4) ^ª							
Completeness [%]	98.2 (96.1) ^a							
Redundancy	2.9 (2.8) ^a							
Refinement								
Resolution [Å]	27.54-1.36							
No. reflections	24111							
$R_{\text{work}/R_{\text{free}}}^{c}$ [%]	15.4/18.5							
No. atoms								
RNA	1172							
Ligand/ion	14							
Water	278							
B-factors								
RNA	25.38							
Ligand/ion	19.69							
Water	38.80							
R.m.s deviations								
Bond lengths [Å]	0.005							
Bond angles [º]	0.783							
Chiral volume [Å ³]	0.106							

^a Values in parentheses are for the highest resolution shell.

^b $R_{merge} = \sum_{h} \sum_{l} |I(h)| - \langle I(h) \rangle | / \sum_{h} \sum_{l} |I(h)|$, where $I(h)_{l}$ is the l th observation of the reflection h and $\langle I(h) \rangle$ is the weighted average intensity for all observations l of reflection h.

^c $R_{work}=\sum_{h}||F_{obs}(h)|-|F_{cal}(h)|| / \sum_{h}|F_{obs}(h)|$, where $F_{obs}(h)$ and $F_{cal}(h)$ are the observed and calculated structure factors for reflection h respectively.

 $^{d}R_{free}$ was calculated as R_{work} using the 5% of reflections which were selected randomly and omitted from refinement.

Strand I									
Base	α	β	Ŷ	δ	ε	ξ	x		
G-3	-64.4	175.8	49.4	77.2	-149.5	-68.7	-159.3		
G-4	-60.8	169.1	60.4	78.8	-164.5	-60.5	-160.		
G-5	157.5	179.5	-179.7	80.1	-140.1	-75.0	-177.		
C-6	-63.1	168.4	61.7	83.1	-143.9	-68.7	-163.		
C-7	-69.4	-174.6	43.4	75.6	-171.2	-76.9	-154.		
G-8	147.8	-169.8	-178.6	80.8	-141.3	-62.0	9.		
G-9	-71.5	-167.7	53.8	79.2	-148.8	-73.3	-165.		
C-10	-61.2	175.5	53.9	80.3	-163.6	-76.7	-156.		
G-11	146.0	-170.7	-178.2	81.1	-144.5	-62.1	10.		
G-12	-67.7	-172.7	50.2	81.4	-152.8	-71.3	-171.		
C-13	-66.2	-179.0	52.6	79.6	-144.1	-70.4	-162.		
G-14	-68.4	166.0	60.4	78.7	-160.0	-73.4	-171.		
G-15	-71.2	-177.1	55.2	77.3	-151.2	-74.6	-168.		
G-16	-59.5	168.5	58.6	79.7	-150.2	-68.1	-166.		
U-17	-63.6	171.5	53.9	80.9	-145.6	-82.3	-160.		
C-18	-68.8	170.8	55.5	78.2	-150.0	-71.2	-158.		
C-19	-60.0	171.7	55.3	76.4			-157.		
	0010	_, _,,	Strar	nd II			2071		
Base	α	β	Y	δ	ε	ξ	Х		
C-19	-65.0	178.5	49.2	80.4			-148.		
C-18	-67.8	-174.8	53.8	79.8	-153.2	-73.0	-154.		
U-17	-65.9	170.7	61.2	79.9	-160.0	-78.0	-163.		
G-16	-62.2	169.2	60.3	77.7	-147.4	-68.2	-168.		
G-15	-68.9	-179.7	51.7	77.2	-151.6	-71.5	-168.		
G-14	-71.0	176.8	57.1	79.4	-155.6	-74.7	-166.		
C-13	-59.6	168.7	53.4	77.8	-155.2	-64.2	-160.		
G-12	-72.3	-172.8	55.7	79.2	-149.5	-74.2	-162.		
G-11	-66.7	166.4	61.4	78.4	-163.5	-72.7	-170.		
C-10	-66.1	-179.9	51.5	78.1	-146.8	-69.5	-164.		
G-9	-71.5	-172.5	51.2	80.2	-149.3	-68.3	-170.		
G-8	151.5	-170.1	180	82.4	-142.1	-63.9	6.		
C-7	-59.7	178.5	50.4	78.6	-169.0	-75.8	-155.		
C-6	-65.8	175.3	54.2	76.8	-154.2	-73.8	-159.		
G-5	-69.6	-177.5	58.1	77.7	-139.8	-67.6	-169.		
G-4	-70.6	167.5	70.2	77.4	-156.0	-72.3	-173.		



Table S-3: Distances (Å) and angle (º) of atoms for different base pairs of							Table S-4: 0	Blobal helica	al paramete	ers calcula	ted for			
r[<u>UU</u> GGGC(CGG)₃GUCC]								the base pairs of r[<u>UU</u> GGGC(CGG) ₃ GUCC] structure						
Base Pair	λ (II)*(º)	λ (II)*(º)	C1'-C1'(Å)	RN9-YN1(Å)	RC8-YC6(Å)		Base pair	Displac- ement (Å)	Angle (º)	Twist (º)	Rise (Å)			
G3-C19	51.3	55.7	10.8	9.0	9.9		G3-C19	8.82	12.28	28.16	3.19			
G4-C18	56.3	56.0	10.7	9.1	10.0		G4-C18	8.09	11.46	32.47	2.62			
G5-U17	44.2	70.7	10.4	8.8	9.8		G5-U17	6.75	11.40	28.94	2.40			
C6-G16	56.2	56.7	10.6	9.0	10.0		C6-G16	7.06	8.02	31.16	2.87			
C7-G15	56.8	55.9	10.6	9.0	9.9		C7-G15	7.45	8.60	33.16	2.88			
G8+G14	35.3	65.3	11.3	9.5	9.1		G8+G14	7.07	7.95	29.95	2.72			
G9-C13	52.8	54.3	10.8	9.0	9.9		G9-C13	7.20	8.14	30.17	3.24			
C10-G12	54.3	54.1	10.8	9.0	10.0		C10-G12	6.86	9.94	34.81	2.83			
G11+G11	34.6	64.8	11.3	9.4	9.1		G11+G11	6.13	10.17	28.18	2.62			
G12-C10	54.1	55.4	10.7	9.0	9.9		G12-C10	6.56	6.75	33.11	2.91			
C13-G9	54.3	52.8	10.8	9.0	9.9		C13-G9	6.71	6.30	30.30	2.70			
G14+G8	64.0	34.3	11.3	9.5	9.1		G14+G8	6.05	9.99	33.37	2.80			
G15-C7	54.6	54.0	10.8	9.0	10.0		G15-C7	6.48	10.27	29.88	3.13			
G16-C6	54.5	57.6	10.6	9.0	9.9		G16-C6	6.82	7.48	27.46	2.97			
U17-G5	69.2	43.1	10.4	8.8	9.7		U17-G5	7.18	5.34	38.30	2.79			
C18-G4	57.0	54.6	10.6	9.0	9.9		C18-G4	7.16	10.13	26.27	2.90			
C19-G3	57.5	55.6	10.6	9.0	10.0		C19-G3	7.17	12.43					

^{*}Lambda is the virtual angle between C1'-YN1 or C1'-RN9 glycosidic bonds and the base-pair C1'-C1' line.

C1'-C1' is the distance between C1' atoms for each base-pair.

RN9-YN1 is the distance between RN9-YN1 atoms for each base-pair.

RC8-YC6 is the distance between RC8-YC6 atoms for each base-pair.



Local base-pair parameters					Local base-pair step parameters					Local base-pair helical parameters									
bp	Shear (A°)	Stretch (A°)	Stagger (A°)	Buckle (°)	Propeller (°)	Opening (°)	Step	Shift (A°)	Slide (A°)	Rise (A°)	Tilt (°)	Roll (°)	Twist (°)	X-disp (A°)	Y-disp (A°)	h-Rise (A°)	Incl.	Tip (°)	h-Twist (°)
G3-C19	-0.34	-0.09	-0.07	-10.44	-5.94	-1.70	GG/CC	0.64	-1.40	3.07	-0.19	8.24	27.95	-4.37	-1.31	2.55	16.6	0.39	29.12
G4-C18	-0.14	-0.02	0.12	-0.70	-9.8	0.43	GG/UC	1.08	-2.61	3.00	-0.59	8.45	19.57	-9.65	-3.11	1.71	23.49	1.63	21.31
G5-U17	-2.33	-0.47	0.13	2.30	-12.27	1.61	GC/GU	-0.27	-1.52	3.22	0.01	1.61	42.87	-2.24	0.37	3.16	2.20	-0.01	42.90
C6-G16	0.22	-0.06	0.00	3.70	-18.31	2.91	CC/GG	-0.61	-1.97	3.29	-1.87	9.30	30.50	-5.15	0.80	2.63	17.16	3.46	31.91
C7-G15	0.22	-0.10	0.10	0.15	-4.59	-0.12	CG/GG	0.04	-3.25	-1.34	-170.79	31.71	160.42	-1.60	0.09	-1.43	15.86	85.44	178.93
G8+G14	-1.48	-3.61	-0.15	11.94	4.31	87.72	GG/CG	-0.47	-3.69	-3.19	128.85	-110.61	97.02	-2.37	-0.37	-1.04	-56.14	-65.39	173.26
G9-C13	-0.10	-0.07	0.07	-5.76	-5.64	-0.74	GC/GC	-0.14	-1.57	3.19	-0.08	4.46	31.15	-3.68	0.24	2.94	8.25	0.16	31.46
C10-G12	0.16	-0.07	0.12	2.42	-3.1	-1.07	CG/GG	0.44	-3.37	-1.24	-173.11	29.11	140.36	-1.65	0.02	-1.50	14.58	86.68	178.49
G11+G11	-1.38	-3.56	-0.12	13.17	-2.07	89.05	GG/CG	-1.16	-3.03	-3.68	130.7	-108.41	64.11	-2.36	-0.43	-1.07	-55.73	-67.19	171.37
G12-C10	-0.12	-0.10	0.07	-1.15	-6.71	-1.03	GC/GC	-0.17	-1.96	3.23	0.06	2.30	33.86	-3.71	0.30	3.09	3.95	-0.10	33.94
C13-G9	0.06	-0.08	0.02	0.46	-7.58	-1.22	CG/GG	1.27	-3.43	3.14	7.99	5.06	81.71	-2.73	-0.78	3.06	3.86	-6.09	82.16
G14+G8	1.34	3.61	0.06	-12.38	3.07	-90.48	GG/CG	-1.38	0.36	3.26	-1.43	3.54	-20.48	-2.48	-4.42	3.05	-9.84	-3.98	-20.83
G15-C7	-0.17	-0.07	0.10	-3.49	-1.85	-1.08	GG/CC	0.29	-1.83	3.28	1.4	7.23	28.66	-5.03	-0.30	2.75	14.31	-2.77	29.58
G16-C6	-0.32	-0.10	-0.11	-0.11	-12.53	0.12	GU/GC	-0.13	-1.43	3.18	-0.68	6.40	41.94	-2.59	0.12	2.94	8.88	0.94	42.41
U17-G5	2.39	-0.57	-0.03	2.97	-16.49	-0.59	UC/GG	-0.04	-1.99	3.12	2.74	7.83	26.52	-5.79	0.66	2.42	16.55	-5.8	27.77
C18-G4	0.36	-0.10	-0.05	1.84	-7.91	0.41	CC/GG	0.12	-2.06	3.23	0.34	3.18	25.15	-5.58	-0.19	2.96	7.27	-0.77	25.35
C19-G3	0.24	-0.05	-0.16	7.91	1.98	-0.94													
Average	-0.08	-0.32	0.00	0.76	-6.20	4.90	Average	-0.03	-2.17	1.80	-4.79	-5.66	51.96	-3.81	-0.52	1.76	1.95	1.66	67.44
Std. Dev.	1.05	1.53	0.1	6.69	6.42	38.26	Std. Dev.	0.71	1.03	2.54	78.54	41.46	46.78	2.09	1.39	1.84	23.93	39.72	67.36

Table S-5: Helical parameters for different base pairs and steps of r[UUGGGC(CGG)₃GUCC]

Table S-6: Major groove widths according to direct P-										
P distances for the direction of sugar-phosphate										
backbone in the r[<u>UU</u> GGGC(CGG) ₃ GUCC] structure										
and their corresponding AU and CG pair and B-DNA.										
Step	Major Groove (Å)									
	RNA DNA									
	CGG	CG Pair	AU pair	B-DNA						
GG/CC										
GG/UC										
GC/GU	13.2	9.1	9.1	11.4						
CC/GG	13.1	9.1	9.1	11.4						
CG/GG	11.5	9.1	9.1	11.4						
GG/CG	12.2	9.1	9.1	11.4						
GC/GC	11.8 9.1 9.1 11.4									
CG/GG	11.4	9.1	9.1	11.4						
GG/CG	10.9	9.1	9.1	11.4						
GC/GC	10.8	9.1	9.1	11.4						
CG/GG	10.6	9.1	9.1	11.4						
GG/CG	12.1	9.1	9.1	11.4						
GG/CC	11.5	9.1	9.1	12.1						
GU/GC	10.5	9.1	9.1	11.4						
UC/GG										
CC/GG										

nairs along axis in the r[IIIIGGGC(CGG) $_{3}$ GICC]										
structure and their corresponding AU. CG pair and B-										
DNA.										
Step	Inclination (º)									
		RNA DNA								
	CGG	B-DNA								
GG/CC	16.6	11.91	11.91	2.85						
GG/UC	23.49	11.90	11.90	2.85						
GC/GU	2.20	11.54	11.34	2.85						
CC/GG	17.16	11.71	11.90	2.74						
CG/GG	15.86	11.90	12.33	2.85						
GG/CG	-56.14	12.56	12.13	2.97						
GC/GC	8.25	11.90	11.34	2.96						
CG/GG	14.58	11.34	12.33	2.74						
GG/CG	-55.73	12.55	12.13	2.96						
GC/GC	3.95	11.91	11.35	2.96						
CG/GG	3.86	11.34	12.33	2.74						
GG/CG	-9.84	12.56	12.14	2.97						
GG/CC	14.31	11.90	11.90	2.96						
GU/GC	8.88	11.91	11.54	2.85						
UC/GG	16.55	11.55	11.72	2.8						
CC/GG	7.27	11.72	11.91	2.85						
Mean	1.95	11.89	11.89	2.87						
SD	23.93	0.37	0.35	0.09						

 Table S-7: Inclination angle for the direction base

References

- 1. Puglisi, J. D., and Tinoco, I., Jr. (1989) Absorbance melting curves of RNA, *Methods Enzymol. 180*, 304-325.
- 2. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol.* 276, 307-326.
- 3. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Cryst. Section D-Biol.Cryst. 60*, 2126-2132.
- 4. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Cryst. 30*, 1022-1025.
- 5. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr. D Biol Crystallogr. 58*, 1948-1954.
- 6. Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Res.* 31, 5108-5121.
- 7. Kumar, A., Fang, P., Parkesh, R., Guo, M. and Disney, M.D. (2011) Crystal structure of the repeating CUG motif in Myotonic Dystrophy reveals multiple conformations and provides insights into selective drug targeting. (PDB 3QIQ, manuscript in preparation)
- 8. Nielsen, J. E., and Vriend, G. (2001) Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations, *Proteins 43*, 403-412.
- 9. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, *Nucleic Acids Res. 35*, W522-525.
- 10. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. U S A 98*, 10037-10041.
- 11. Cheng, A. H. D., and Cheng, D. T. (2005) Heritage and early history of the boundary element method, *Eng. Anal. Bound Elem. 29*, 268-302.
- 12. Wahl, M. C., Rao, S. T., and Sundaralingam, M. (1996) The structure of r(UUCGCG) has a 5'-UUoverhang exhibiting Hoogsteen-like trans U center dot U base pairs, *Nat. Struct. Biol.* 3, 24-31.