Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome

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

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

ABSTRACT

The CGG repeats are present in the 5′-untranslated region (5′-UTR) of the fragile X mental retardation gene *FMR1* and are associated with two diseases: fragile X-associated tremor ataxia syndrome (FXTAS) and fragile X syndrome (FXS). FXTAS occurs when the number of repeats is 55–200 and FXS develops when the number exceeds 200. FXTAS is an RNA-mediated disease in which the expanded CGG tracts form stable structures and sequester important RNA binding proteins. We obtained and analysed three crystal structures of double-helical CGG repeats involving unmodified and 8-Br modified guanosine residues. Despite the presence of the non-canonical base pairs, the helices retain an A-form. In the G–G pairs one guanosine is always in the *syn* conformation, the other is *anti*. There are two hydrogen bonds between the Watson–Crick edge of G(*anti*) and the Hoogsteen edge of G(*syn*): O6·N1H and N7·N2H. The G(*syn*)-G(*anti*) pair shows affinity for binding ions in the major groove. G(*syn*) causes local unwinding of the helix, compensated elsewhere along the duplex. CGG helical structures appear relatively stable compared with CAG and CUG tracts. This could be an important factor in the RNA's ligand binding affinity and specificity.

INTRODUCTION

Tandem repeats of the CGG trinucleotide motif are abundant in human genome and occur in numerous genes and transcripts (1). The repeat tracts are often polymorphic in length in human population and may play a regulatory role in gene expression. The CGG repeats present in the 5′-untranslated region (5′-UTR) of the fragile X mental retardation gene *FMR1* are associated

with several distinct phenotypes (2). In normal population the number of CGG repeats varies in the range 5–54 (2,3); 45–54 repeats fall into a subclass named 'the grey zone' associated with an increased likelihood of inter-generational pathogenic expansions (2,4). Tracts of 55–200 CGGs are premutations that can cause progressive neurodegenerative disorder fragile X-associated tremor ataxia syndrome (FXTAS) in elderly males (5,6). Female premutation carriers are at risk of developing premature ovarian failure (7). More than 200 CGG repeats are full mutations resulting in fragile X syndrome (FXS), the most common inherited mental retardation syndrome in man (8).

FXTAS is an RNA-mediated disease in which the level of FMR1 mRNA is significantly elevated (9,10). The expanded tracts form stable structures and sequester important RNA binding proteins which are normally required for splicing and other cellular processes (11,12). The protein sequestration results in the formation of intranuclear inclusions in neurons and astrocytes (12,13) and triggers a dynamic formation of aggregates and deregulation of alternative splicing of a number of genes in model cellular systems (11). In addition, the presence of stable CGG structures inhibits *FMR1* translation at the initiation step, resulting in a deficiency of the encoded FMRP protein (9,14,15). FMRP is an important RNA binding protein involved in mRNA trafficking between the cell nucleus and the cytoplasm and in regulating translation at the synapse (16).

The structure of the CGG repeats in FMR1 transcripts is a hairpin whose stem is formed by alternating C–G, G–C and the non-canonical G–G base pairs (17). Isolated $(CGG)_{20}$ repeats are thermodynamically the most stable hairpins of all the $(CNG)_{20}$ (N stands for C or G or A or U) repeats (18,19) which means that G–G pairs are the strongest of all homobasic interactions. According to an NMR study (20), the opposing G–G bases are highly dynamic in CGG repeat hairpin, having one G residue in *anti* and the other in *syn* conformation.

*To whom correspondence should be addressed. Tel: +48-61-8528503; Fax: +48-61-8520532; Email: wojtekr@ibch.poznan.pl

Short CGG repeats were shown to form a duplex structure (20,21) and in the presence of potassium ions—G-tetraplexes (22–24). Thus, an RNA structure formed by CGG repeats is less clearly defined than in the case of CUG repeats (18,19,25) and CAG repeats (19,26–28) for which crystal structures have been determined (29–31).

In this study, we report three crystal structures of double-helical CGG repeats containing native and 8-Br modified guanosine residues.

MATERIALS AND METHODS

Synthesis, purification and crystallization of oligoribonucleotides

GCGGCGGC, GC(8-BrG)GCGGC and GC(8-BrG)GC GGCGGC oligomers were synthesized on an Applied Biosystems DNA/RNA synthesizer, using cyanoethyl phosphoramidite chemistry. Commercially available C and G phosphoramidites with 2′-*O*-tetrbutyldimethylsilyl were used for the synthesis of RNA (Glen Research, Azco, Proligo). The phosphoramidite 8-Br guanosine was synthesized according to Proctor *et al.* (32). The details of deprotection and purification of oligoribonucleotides were described previously (33).

All crystals were grown by the hanging drop/vapour diffusion method at 19°C. A single crystal of GC(8-BrG)GCGGCGGC grew in 10 months. The reservoir initially contained 10 mM MgCl₂ 50 mM Na cacodylate, pH 6.0 and 1.0 M Li₂SO₄. The crystallization drop initially contained 3 µl of RNA at 10 mg/ml and 1 mM MgCl₂, and 3 µl of the reservoir solution. Crystals of GCGGCGGC grew in several days from the same solution as above, but involved crystal seeding and the starting RNA concentration of 2.4 mg/ml. The crystals grew as clusters of small needles which were then used for seeding. The seeded crystals grew from a similar solution but with the RNA concentration of 1.2 mg/ml. They appeared as single blocks but were in fact clusters of crystals that had to be separated. Crystals of GC(8-BrG)GCGGC grew in 2 months. The reservoir initially contained 10 mM CaCl₂, 0.2 M NH₄Cl, 50 mM Tris–HCl at pH 8.5 and 30% w/v PEG 4000. The crystallization drop initially contained 2 µl of RNA at 10 mg/ml and 1 mM MgCl₂, and 2 µl of the reservoir solution.

X-ray data collection, structure solution and refinement

X-ray diffraction data were collected at 100 K: from GCG GCGGC on BL 14.2 beam line at the BESSY synchrotron (Berlin) to the resolution of 2.05 Å; from GC(8-BrG)GCG GCGGC and GC(8-BrG)GCGGC on EMBL X11, DESY, Hamburg, to the resolution of 1.45 Å and 0.97 Å, respectively. The crystals were cryoprotected by 20% glycerol (v/v) in the mother liquor. The data were integrated and scaled using the program suite DENZO/SCALEPACK (34).

The structure of GC(8-BrG)GCGGC was solved first. SHELXD was used to identify the positions of the Br atoms by analysing the Patterson function based on the anomalous signal (35). The program identified two sites at least five times higher than any other peaks. SHELXE was

used to identify the correct enantiomorph and to calculate initial phases (35), the calculated electron density map was uninterpretable in terms of an atomic model, but showed parallel columns of density indicating stacked RNA duplexes. DM was used for density modification but the resulting maps were still uninterpretable (36). The phases from DM were used in a free-atom phase refinement with 'shaking' of the model, using ARP/wARP (37). In a total of 100 cycles of adding/removing atoms with nine round of shaking interspersed, the R-factor/R-free statistics initially remained apparently random but after 25 cycles started dropping and in the final 50 cycles collapsed to 0.188/0.256. The free-atom electron density map showed the RNA atoms clearly resolved and a well-defined solvent structure; the map looked very similar to the final map for the refined model. Both GC(8-BrG)GCGGCGGC and GCGGCGGC structures were solved by molecular replacement using PHASER (38). The initial models were poor but sufficient for the purpose of refinement and model extension. The manual rebuilding and map inspection were done using Coot (39).

All three structures were refined using Refmac5 (40) and Phenix (41). The final model of [GC(8-BrG)GCGGC]₂ was refined without restraints and with anisotropic temperature factors. The last few cycles of the [GC(8-BrG)GC GGC]₂ refinement were performed using all data, including the R_{free} set. The other two models, GCGGC GGC and GC(8-BrG)GCGGCGGC, were refined using isotropic B-factors.

Helical parameters were calculated using 3DNA (42). Sequence-independent measures were used based on vectors connecting the C1′ atoms of the paired residues, to avoid computational artefacts arising from non-canonical base pairing. Program PBEQ-Solver (43) was used to calculate electrostatic potential map. All pictures were drawn using PyMOL v0.99rc6 (44). The coordinates of the crystallographic models have been deposited with the Protein Data Bank (PDB) with accession codes 3R1C, 3R1D and 3R1E.

RESULTS AND DISCUSSION

The overall structures

The RNA in all three crystal structures has the form of duplexes stacking end-to-end and forming bundles of parallel columns. The crystals of the native RNA contain 18 (GCGGCGGC)₂ duplexes in the P1 unit cell. Each column consists of all the 18 independent duplexes stacked consecutively (Supplementary Figure 1). In the atomic resolution C2 structure, there is one [GC(8-BrG)GCGGC]₂ duplex in the asymmetric unit. The other monoclinic crystal contains five crystallographically independent RNA strands. They form three (GC(8-BrG)GCGGCGGC]₂ duplexes, the third consisting of two symmetry-related strands. The native RNA and [GC(8-BrG)GCGGCGGC]₂ structures contain sulphate ions from the crystallization medium, while [GC(8-BrG)GCGGC]₂ crystals have Ca²⁺. All the ions interact in the major groove with the G–G pairs (details below). The final models are summarized in Table 1.

Table 1. Summary of the X-ray data and model refinement for (GCGGCGGC)$_2$, [GC(8-BrG)GCGGCGGC]$_2$ and [GC(8-BrG)GCGGC]$_2$

Crystal	GCGGCGGC	GC(8-BrG)GCGGC	GC(8-BrG)GCGGCGGC
Beam line	BESSY BL 14.2	EMBL-X11	EMBL-X11
Wavelength (Å)	0.9200	0.8126	0.8126
Space group	P1	C2	C2
Cell parameters	$a = 39.7$, $b = 76.9$, $c = 85.4$ Å, $\alpha = 90.0$, $\beta = 88.6$, $\gamma = 77.3°$	$a = 50.7$, $b = 22.5$, $c = 44.2$ Å, $\beta = 117.8°$	$a = 118.6$, $b = 28.6$, $c = 61.8$ Å, $\beta = 118.0°$
Resolution range (Å)	20.0–2.05 (2.09–2.05)[a]	20.0–0.97 (0.99–0.97)	20.0–1.45 (1.47–1.45)
Mosaicity (°)	0.3	1	0.3
Exposure time per image (s)	5–10	40	50
R_{merge}[b]	0.093 (0.488)	0.132 (0.934)	0.112 (0.620)
$<I/\sigma(I)>$	17 (2.8)	9 (2.5)	10 (3.3)
Completeness (%)	98.4 (97.5)	95.7 (93.4)	99.8 (100.0)
No. unique reflections	60 328	24 974	32 941
Overall multiplicity	4.4 (3.5)	5.4 (4.7)	7.5 (6.3)
Reflections > 3σ (%)	71 (40)	77 (29)	77 (50)
B-factor from Wilson plot (Å^2)	33	7.8	21.2
R_{work}[c]	21.56	13.66	23.21
R_{free}[c]	25.71	–	27.02
No. RNA atoms	6304	370	1617
Ions	7 SO$_4^{2-}$	2 Ca^{2+}	8 SO$_4^{2-}$
No. water molecules	524	103	207
Other solvent	–	1 glycerol	–
r.m.s. deviation from ideal values			
Bond lengths (Å)	0.005	0.011	0.006
Bond angles (°)	1.145	1.746	1.282
PDB code	3R1C	3R1E	3R1D

[a]Values in brackets are for the highest resolution shell.
[b]$R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - <I(hkl)>|/\Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ and $<I(hkl)>$ are the observed individual and mean intensities of a reflection with indices *hkl*, respectively, Σ_i is the sum over the individual measurements of a reflection with indices *hkl* and Σ_{hkl} is the sum over all reflections.
[c]R_{free} was calculated using 5% of the total reflections chosen randomly and omitted from the refinement.

Non-canonical G–G pairing and its consequences on the duplex

Despite the presence of the non-canonical base pairs, the helices have a typical A-form, with values of helical twist in the range 30–32 for all the duplexes, the sugar pucker C3′-*endo* or, in some cases, C2′-*exo* and Zp values of 2.63 ± 0.22 Å (for the A-form it should be more than 1.5 Å). The helices contain C–G and G–C base pairs with typical Watson–Crick interactions. Between them are the non-canonical G–G pairs in which one guanosine is always in the *syn* conformation while the other is *anti* (Figure 1), experimentally shown to be the preferred arrangement in double helical context (45–47). The *syn-anti* geometry of this base pair can be described as G/G *cis* Watson–Crick/Hoogsteen according to the nomenclature proposed by Leontis and Westhof (48). There are two hydrogen bonds between the guanosine residues: carbonyl oxygen is bonded to N1H, and N7 to the *exo*-amino group. All the H-bond distances are in the range 2.6–3.3 Å. The conformation of the G(*syn*) residue is additionally stabilized by a hydrogen bond between the *exo*-amino function and its phosphate oxygen atom (3.0 Å).

The *syn-anti* arrangement avoids the steric clash between the two bulky guanines within the helical structure. This is evident from the C1′–C1′ distances between the paired residues: 11.3 ± 0.1 Å for G–G, compared with 10.7 ± 0.2 Å for the canonical C–G pairs. The angle λ of
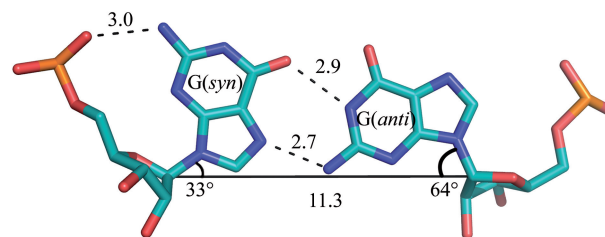

Figure 1. Non-canonical G(*syn*)–G(*anti*) pair. Hydrogen bonds are drawn with dashed lines. Solid line connecting C1′ atoms gives a measure of strand separation. Angles λ are marked (see text).

the glycosidic bond with the line connecting the C1′ atoms of each pair is $33 \pm 4°$ for the G(*syn*) residues and $64 \pm 3°$ for G(*anti*), compared with $54 \pm 3°$ for the other residues. This means that the G(*syn*) is shifted towards the minor groove and the G(*anti*) towards to major groove, which optimizes the H-bonding interactions between the Hoogsteen and Watson–Crick edges, while avoiding the clash between the carbonyl oxygen atoms (Figure 1).

The G(*syn*) residues also show unusual α and γ backbone torsion angles. The α angle, representing a rotation about the P-O5′ bond, is +*ac*, +*ap*, or −*ap* (in one case) instead of the typical −*sc*. The angles range 107–182° with the average value of 142°, almost half a turn from the mean −*sc* value of −60°. The γ angle, about the C5′-C4′ bond, is +*ap* or −*ap* in the G(*syn*)

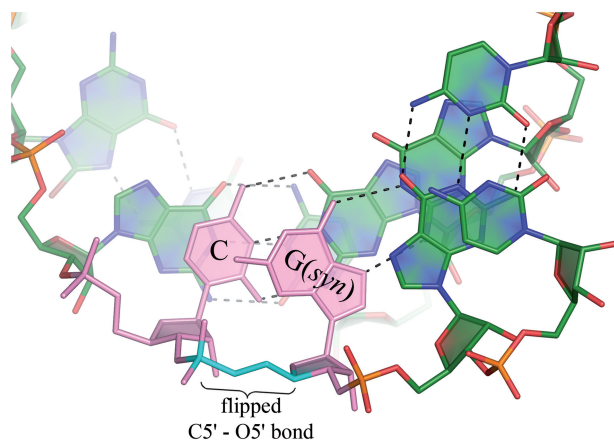**Figure 2.** The G(*syn*) residue in a helical context is nearly co-planar with the adjacent cytosine (pink). The torsion angles α, γ (cyan) are flipped with respect to typical values for A-form and the helix is locally unwound.

flipped
C5' - O5' bond

residues, ranging 152° to −152°. This is ~120° from the usual +*sc*. This amounts to flipping of the O5′-C5′ bond and a local 'straightening' of the sugar-phosphate backbone (Figure 2). The unusual torsion of the backbone is similar to the 'extended' conformation observed by Haran *et al.* (49) in a CG step of a DNA helix. In the present study this conformation seems to be necessary for the simultaneous (i) Hoogsteen–Watson–Crick pairing of the guanosines; (ii) the H-bond between *exo*-amino group of G(*syn*) with the phosphate O; and (iii) stacking against the neighbouring cytosine. These effects explain the accommodation of the G–G pair within the helix.

The local straightening of the sugar-phosphate backbone, amounting to a local unwinding of the helix, is compensated elsewhere along the duplex, and the overall statistics do not deviate from typical (Supplementary Tables 1–4). This is different from the case of CAG repeats (31) where a similar inversion of the α and γ angles is associated with the overall unwinding of the helix and broadening of the major groove to >20 Å (measured as the distance between lines connecting P atoms). For the native CGG-containing duplexes, the average width of the major groove was 17.9 ± 0.9 Å, for the longer Br-modified duplex it was 17.8 ± 2.5 Å and for the shorter modified structure 14.3 Å. The values for the minor groove were 15.8 ± 0.5 Å, 15.4 ± 0.5 Å and 16.1 Å, respectively. The values are not out of the ordinary for the A form.

Another effect of the G(*syn*) conformation is that the guanine has no stacking interaction with the downstream G–C pair and it stacks against the preceding pair of residues (Figure 3).

The effect of bromination and the distribution of G–G conformers

All the 8-Br modified guanosine residues are in the *syn* conformation, therefore the pairs are 8-BrG(*syn*)-G(*anti*). As observed before, duplexes containing the modified residues are more stable than the

corresponding native duplexes. The melting temperature of [GC(8-BrG)GCGGC]$_2$ is 13°C higher than for (GCGG CGGC)$_2$ (21). In structural terms this is probably due to a restricting effect of Br on the conformational freedom and excluding the unfavourable G(*anti*)-G(*anti*) interactions. The crystallographic structures bear this out in the sense that the modified duplexes contain only well ordered 8-BrG(*syn*)-G(*anti*) pairs, while in the native duplexes each G–G pair is observed in one of three possible arrangements: G(*syn*)-G(*anti*), G(*anti*)-G(*syn*) or a statically disordered mixture of the two (in 2 out of 36 pairs). The three base-pairing arrangements occur with different frequencies and in various combinations of pairs along the native duplex. Symmetric arrangements are clearly favoured: *anti-syn* followed by *syn-anti* or *vice versa* (in 14 out of 18 cases). There is a slight preference, which can be fortuitous, for the former arrangement (8 cases as opposed to 6). Of the remaining four duplexes, two are clearly asymmetric and two contains statically disordered G–G pairs (Supplementary Table 5). In the longer modified duplexes, containing unmodified G–G pair in the middle, two of the native G–G pairs are disordered, showing both possible conformations; the third is ordered.

Apart from restricting the conformational freedom, and thus defining the conformation of the G–G pair, there is little crystallographic evidence that bromination alters the structure compared with the native RNA. The native and brominated duplexes can be superposed with an r.m.s. deviation of ~1 Å. The similarity is also reflected in helical parameters. In terms of interactions with the solvent, the Br atom seems to displace a water molecule, which in native G(*syn*) is located ~3.2 Å from C8, in the minor groove, but its main effect appears to be steric. In terms of H-bonding capacity, bromination alters the pKa of guanosine from 9.3 to 8.4 (50), but this is likely to be insignificant.

Solvent interactions and hydration

The exposed Watson–Crick edges of G(*syn*) residues interact with sulphate anions. In some instances in the native RNA structure and in the longer modified duplex, the sulphate appears well ordered, but in most cases its orientation is poorly defined. The anions can be distinguished from water molecules by the size and shape of the electron density and the interaction distance from the RNA (Figure 4b). In the high resolution structure, where there was no sulphate in the crystallization medium, an inner complex is observed involving the carbonyl oxygen atom with a hydrated calcium cation (Figure 4a). The conditions in the crystallization medium are far from physiological, nevertheless the observed complexes indicate a potential of the solvent-exposed Hoogsteen–Watson–Crick edge for attracting charged species, especially as no interactions with ions are observed elsewhere in the structures. In the absence of sulphate, the G(*syn*) Watson–Crick edge is hydrated by three water molecules that form a crest co-planar with the guanine. In the absence of Ca^{2+}, the G(*anti*) H-edge is hydrated by two- or three-ordered water molecules.
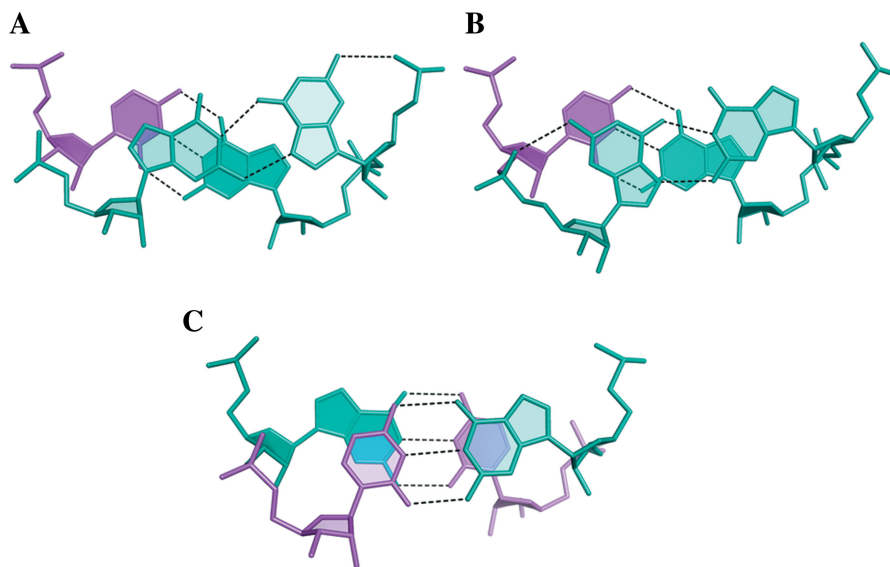
Figure 3. Stacking interactions in the CGG duplex structure. (**A** and **B**)The non-canonical G–G pairs (aquamarine); (**C**) the canonical CG/GC step.
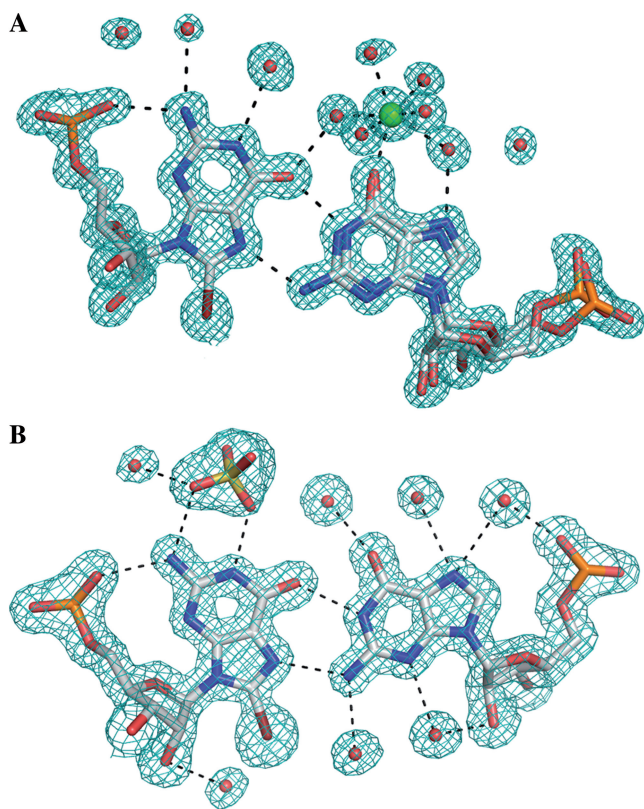

Figure 4. The hydration of G(*syn*)–G(*anti*) pairs in the GC(8-BrG)GCGGC (**A**) and in the GC(8-BrG)GCGGCGGC (**B**) structures. Ca^{2+} (green) is bound directly to the carbonyl oxygen atom of G(*anti*); a sulphate anion interacts with the WC edge of G(*syn*). The 2*Fo-Fc* electron density map is contoured at 1σ level.

Interestingly, in all the examined instances of the structure which have sulphate anions in the major groove, the width of the grooves are similar and close to 18 Å (see above), as opposed to the Ca^{2+}-containing structure,

in which the groove is narrower by nearly 4 Å. It is possible that the presence of the anions stabilizes the width of the major groove.

Electrostatic surface potential

The electrostatic surface potential shows the already familiar pattern of alternating stripes of positive and negative potential in the minor groove, similar to the previously observed distribution in CUG and CAG repeats (Figure 5). The pattern is due primarily to the C–G and G–C pairs rather than to the interposed N–N pairs. The major groove in the CGG structures is mostly electronegative with positive areas generated each by the Watson–Crick edge of G(*syn*) and the *exo*-amino group of the preceding cytidine residue. The binding of the sulphate ions corresponds very closely with the electropositive features associated with G(*syn*) residues and their calculated surface potential indeed appears higher than for the adjacent cytosines, which have not attracted any ions. The difference in binding potential for ligands can be explained by the exposed Watson–Crick edge of the guanine in this position. In addition, the guanines have stacking interactions only on one side, while the cytosines are engaged on both their surfaces.

It is harder to explain, just by examining the electrostatic potential, what distinguishes CGG from CUG or CAG, and why CGG is not recognized by MBNL1 protein (51–53). An answer could lie in the stability of the CGG duplex. The non-canonical G–G pair contains two H-bonds, as opposed to a single bond in U–U and a weak C–H···N bond in A–A. Therefore the CGG tracts could be more stable in the duplex form and less accessible to the protein which appears to bind single-stranded RNA (54). This is consistent with thermodynamic parameters: ΔG of duplexes containing CGG is markedly higher than for the other three repeats, which are similar (18,19,21).
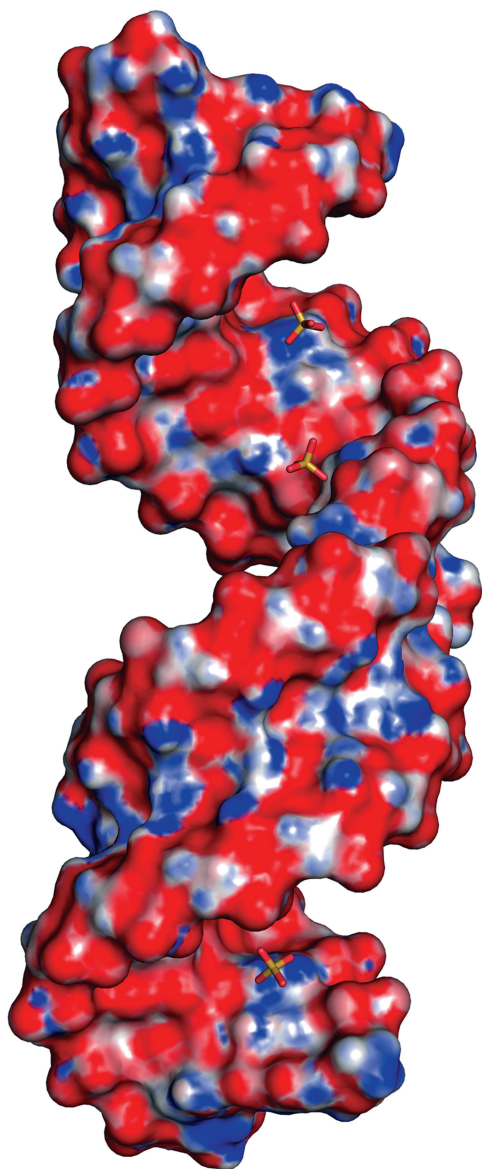
Figure 5. The electrostatic surface potential of two consecutive duplexes of the [GC(8-BrG)GCGGCGGC]$_2$ structure. Red is negative, blue is positive. Sulphate anions (sticks) are shown interacting in the major groove.

CONCLUSIONS

This is a third in a series of studies aimed at profiling CNG structures at crystallographic resolution to facilitate drug design and provide 3D templates for rationalizing biochemical and cytological observations. The need for detailed RNA structures is clearly signalled in the literature on rational design of CNG-binding ligands (55–57).

The crystal structures presented here of unmodified and modified CGG repeats are consistent and allow a general description of double-stranded CGG tracts and a comparison with CUG and CAG structures reported previously (29,31). The foremost common feature is that all the known CNG structures form A-helices stabilized by C–G and G–C pairs acting as sturdy struts. The variety is provided in between by the non-canonical pairs.

The 'accommodation problem' is solved differently for each N–N pair, but in every case the disruption of the helix is surprisingly small. The bulky guanines fit within the helical constraints by a 180° flip about the glycosidic bond of one of the bases. The equally bulky adenines retain the *anti*-conformation, but are shifted out of the helical axis, towards the major groove. In the one known example of paired pyrimidines, the relatively small uracil rings remain some distance apart, making only one direct hydrogen bond instead of two bonds which they can make in different environments.

Despite the general resilience of the A-form, characteristic differences can be observed between the CNG duplex structures. The parameters that seem to be specially sensitive are helical twist and major groove width. In CGG helices, we observe local unwinding of the helix around the G–G pairs, as opposed to the more general unwinding in the case of CAG structures. The major groove width may depend on the nature of the ligand which occupies it, bound to the G–G pair. This work provides examples of bound calcium and sulphate ions; another example of sulphate interacting with G–G pairs and possible affecting the groove width is provided by Adamiak and colleagues (58).

Thermodynamic stability is an important property that is difficult to investigate in crystallography, but some measure is provided by the nature and the count of hydrogen bonds and the extent of stacking interactions. In this respect, CGG helical structures appear relatively stable compared with CAG and CUG tracts—in agreement with calorimetric studies (19,21). This could be an important factor in the RNA's ligand binding affinity and specificity.

So far, the three known CNG repeats have been observed to bind small ligands in the major groove. Glycerol, sulphate anion and Ca^{2+} cations were found to be associated with N–N pairs of the CNG duplexes. The interactions depend on hydrogen bonds formed with the functional groups exposed in the major groove, characteristic of each N–N pair and to some extent on the electrostatic charge distribution. These can be taken as the main indicators in designing specific ligands.

A detailed comparison of CGG, CAG and CUG duplex structures is provided in Supplementary Table 6.

ACCESSION NUMBERS

3R1C, 3R1D, 3R1E.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. Kozlowski,P., de Mezer,M. and Krzyzosiak,W.J. (2010) Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.*, **38**, 4027–4039.
2. Fu,Y.H., Kuhl,D.P., Pizzuti,A., Pieretti,M., Sutcliffe,J.S., Richards,S., Verkerk,A.J., Holden,J.J., Fenwick,R.G. Jr, Warren,S.T. *et al.* (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.
3. Dombrowski,C., Levesque,S., Morel,M.L., Rouillard,P., Morgan,K. and Rousseau,F. (2002) Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.*, **11**, 371–378.
4. Zhong,N., Ju,W., Pietrofesa,J., Wang,D., Dobkin,C. and Brown,W.T. (1996) Fragile X "gray zone" alleles: AGG patterns, expansion risks, and associated haplotypes. *Am. J. Med. Genet.*, **64**, 261–265.
5. Hagerman,R.J., Leehey,M., Heinrichs,W., Tassone,F., Wilson,R., Hills,J., Grigsby,J., Gage,B. and Hagerman,P.J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*, **57**, 127–130.
6. Jacquemont,S., Hagerman,R.J., Leehey,M., Grigsby,J., Zhang,L., Brunberg,J.A., Greco,C., Des Portes,V., Jardini,T., Levine,R. *et al.* (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.*, **72**, 869–878.
7. Sherman,S.L. (2000) Premature ovarian failure among fragile X premutation carriers: parent-of-origin effect? *Am. J. Hum. Genet.*, **67**, 11–13.
8. Glass,I.A. (1991) X linked mental retardation. *J. Med. Genet.*, **28**, 361–371.
9. Tassone,F., Hagerman,R.J., Taylor,A.K., Gane,L.W., Godfrey,T.E. and Hagerman,P.J. (2000) Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.*, **66**, 6–15.
10. Kenneson,A., Zhang,F., Hagedorn,C.H. and Warren,S.T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.*, **10**, 1449–1454.
11. Sellier,C., Rau,F., Liu,Y., Tassone,F., Hukema,R.K., Gattoni,R., Schneider,A., Richard,S., Willemsen,R., Elliott,D.J. *et al.* (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J*, **29**, 1248–1261.
12. Iwahashi,C.K., Yasui,D.H., An,H.J., Greco,C.M., Tassone,F., Nannen,K., Babineau,B., Lebrilla,C.B., Hagerman,R.J. and Hagerman,P.J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**, 256–271.
13. Greco,C.M., Hagerman,R.J., Tassone,F., Chudley,A.E., Del Bigio,M.R., Jacquemont,S., Leehey,M. and Hagerman,P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, **125**, 1760–1771.
14. Primerano,B., Tassone,F., Hagerman,R.J., Hagerman,P., Amaldi,F. and Bagni,C. (2002) Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. *RNA*, **8**, 1482–1488.
15. Chen,L.S., Tassone,F., Sahota,P. and Hagerman,P.J. (2003) The (CGG)n repeat element within the 5′ untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of a downstream reporter. *Hum. Mol. Genet.*, **12**, 3067–3074.
16. De Rubeis,S. and Bagni,C. (2010) Fragile X mental retardation protein control of neuronal mRNA metabolism: Insights into mRNA stability. *Mol. Cell. Neurosci.*, **43**, 43–50.
17. Napierala,M., Michalowski,D., de Mezer,M. and Krzyzosiak,W.J. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.*, **33**, 451–463.
18. Sobczak,K., de Mezer,M., Michlewski,G., Krol,J. and Krzyzosiak,W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.*, **31**, 5469–5482.
19. Sobczak,K., Michlewski,G., de Mezer,M., Kierzek,E., Krol,J., Olejniczak,M., Kierzek,R. and Krzyzosiak,W.J. (2010) Structural diversity of triplet repeat RNAs. *J. Biol. Chem.*, **285**, 12755–12764.
20. Zumwalt,M., Ludwig,A., Hagerman,P.J. and Dieckmann,T. (2007) Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. *RNA Biol.*, **4**, 93–100.
21. Broda,M., Kierzek,E., Gdaniec,Z., Kulinski,T. and Kierzek,R. (2005) Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry*, **44**, 10873–10882.
22. Khateb,S., Weisman-Shomer,P., Hershco,I., Loeb,L.A. and Fry,M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
23. Khateb,S., Weisman-Shomer,P., Hershco-Shani,I., Ludwig,A.L. and Fry,M. (2007) The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.*, **35**, 5775–5788.
24. Ofer,N., Weisman-Shomer,P., Shklover,J. and Fry,M. (2009) The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.*, **37**, 2712–2722.
25. Napierala,M. and Krzyzosiak,W.J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem.*, **272**, 31079–31085.
26. Michlewski,G. and Krzyzosiak,W.J. (2004) Molecular architecture of CAG repeats in human disease related transcripts. *J. Mol. Biol.*, **340**, 665–679.
27. Sobczak,K. and Krzyzosiak,W.J. (2004) Imperfect CAG repeats form diverse structures in SCA1 transcripts. *J. Biol. Chem.*, **279**, 41563–41572.
28. Sobczak,K. and Krzyzosiak,W.J. (2005) CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J. Biol. Chem.*, **280**, 3898–3910.
29. 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.
30. Mooers,B.H., Logue,J.S. and Berglund,J.A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl Acad. Sci. USA*, **102**, 16626–16631.
31. 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.
32. Proctor,D.J., Kierzek,E., Kierzek,R. and Bevilacqua,P.C. (2003) Restricting the conformational heterogeneity of RNA by specific incorporation of 8-bromoguanosine. *J. Am. Chem. Soc.*, **125**, 2390–2391.
33. Xia,T., SantaLucia,J. Jr, Burkard,M.E., Kierzek,R., Schroeder,S.J., Jiao,X., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719–14735.

34. Otwinowski,Z.M., W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–325.

35. Sheldrick,G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 479–485.

36. Cowtan,K. (1994) An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.

37. Lamzin,V.S. and Wilson,K.S. (1993) Automated refinement of protein models. *Acta Crystallogr. D Biol. Crystallogr.*, **49**, 129–147.

38. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

39. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

40. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

41. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

42. 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.

43. Jo,S., Vargyas,M., Vasko-Szedlar,J., Roux,B. and Im,W. (2008) PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.*, **36**, W270–W275.

44. DeLano,W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, Palo Alto, CA, USA.

45. Burkard,M.E. and Turner,D.H. (2000) NMR structures of r(GCAGGCGUGC)2 and determinants of stability for single guanosine-guanosine base pairs. *Biochemistry*, **39**, 11748–11762.

46. Burkard,M.E., Xia,T. and Turner,D.H. (2001) Thermodynamics of RNA internal loops with a guanosine-guanosine pair adjacent to another noncanonical pair. *Biochemistry*, **40**, 2478–2483.

47. SantaLucia,J. Jr, Kierzek,R. and Turner,D.H. (1991) Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U.U and C.C.+ pairs. *Biochemistry*, **30**, 8242–8251.

48. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.

49. Haran,T.E., Wang,A.H.-J. and Rich,A. (1987) The crystal structure of d(CCCCGGGG): a new A-form variant with an extended backbone conformation. *J. Biomol. Struct. Dyn.*, **2**, 397–412.

50. Ioele,M., Bazzanini,R., Chatgilialoglu,C. and Mulazzani,Q.G. (2000) Chemical Radiation Studies of 8-Bromoguanosine in Aqueous Solutions. *J. Am. Chem. Soc.*, **122**, 1900–1907.

51. Ho,T.H., Savkur,R.S., Poulos,M.G., Mancini,M.A., Swanson,M.S. and Cooper,T.A. (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell Sci.*, **118**, 2923–2933.

52. Kino,Y., Mori,D., Oma,Y., Takeshita,Y., Sasagawa,N. and Ishiura,S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.*, **13**, 495–507.

53. Yuan,Y., Compton,S.A., Sobczak,K., Stenberg,M.G., Thornton,C.A., Griffith,J.D. and Swanson,M.S. (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.*, **35**, 5474–5486.

54. Teplova,M. and Patel,D.J. (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.*, **15**, 1343–1351.

55. Arambula,J.F., Ramisetty,S.R., Baranger,A.M. and Zimmerman,S.C. (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc. Natl Acad. Sci. USA*, **106**, 16068–16073.

56. Lee,M.M., Childs-Disney,J.L., Pushechnikov,A., French,J.M., Sobczak,K., Thornton,C.A. and Disney,M.D. (2009) Controlling the specificity of modularly assembled small molecules for RNA via ligand module spacing: targeting the RNAs that cause myotonic muscular dystrophy. *J. Am. Chem. Soc.*, **131**, 17464–17472.

57. Pushechnikov,A., Lee,M.M., Childs-Disney,J.L., Sobczak,K., French,J.M., Thornton,C.A. and Disney,M.D. (2009) Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J. Am. Chem. Soc.*, **131**, 9767–9779.

58. Rypniewski,W., Adamiak,D.A., Milecki,J. and Adamiak,R.W. (2008) Noncanonical G(syn)-G(anti) base pairs stabilized by sulphate anions in two X-ray structures of the (GUGGUCUGAU GAGGCC) RNA duplex. *RNA*, **14**, 1845–1851.