Isotropic B factor distribution in alternate conformations of UGGGGU octamers.

Structural variations and solvent structure of r(UGGGGU) quadruplexes stabilized by Sr^{2+} ions.

Alastair C. Fyfe¹, Pete W. Dunten², Monika M. Martick³, and William G.
Scott¹

¹Department of Chemistry and Biochemistry, University of California,
Santa Cruz

²Stanford Synchrotron Radiation Lightsource

³Department of Molecular and Cell Biology, University of California,
Berkeley

Abstract

Guanine-rich sequences can, under appropriate conditions, adopt a distinctive, four-stranded, helical fold known as a G-quadruplex. Interest in quadruplex folds has grown in recent years as evidence of their biological relevance has accumulated from both sequence analysis and function-specific assays. The folds are unusually stable and their formation appears to require close management to maintain cell health; regulatory failure correlates with genomic instability and a number of cancer phenotypes.

Biologically relevant quadruplex folds are anticipated to form transiently in mRNA and in single-stranded, unwound DNA. To elucidate factors, including bound solvent,

that contribute to the stability of RNA quadruplexes, we examine, by x-ray crystallography and SAXS, the structure of a previously-reported tetramolecular quadruplex, UGGGGU stabilized by Sr^{2+} ions. Crystal forms of the octameric assembly formed by this sequence exhibit unusually strong diffraction and anomalous signal enabling the construction of reliable models to a resolution of 0.88 Å. The solvent structure confirms hydration patterns reported for other nucleic acid helical conformations and provides support for the greater stability of RNA quadruplexes relative to DNA. Novel features detected in the octameric RNA assembly include a new crystal form, evidence of multiple conformations and structural variations in the 3' U-tetrad, including one that leads to the formation of a hydrated internal cavity.

1 Introduction

Nucleic acid sequences in which guanines are both abundant and distributed in accordance with a distinct but permissive pattern have a propensity to fold into a G-quadruplex motif [1] comprised of stacks of planar G-quartets, as outlined in Fig. 1. Once formed, such structures are stabilized by a complement of electrostatic, stacking and hydrogen-bonding forces that render them unusually stable relative to other nucleic acid conformations. Though many examples of *in vitro* folded structures obtained from G-rich sequences are available [2, 3], the extent to which such structures occur *in vivo* and their physiological roles remain open to question.

Although our current understanding is limited, the following generalizations appear broadly applicable: a growing inventory of structures confirms the stability and conformational diversity of G-quadruplex folds; sequences with the potential to adopt a quadruplex fold are abundant, though globally counterselected, and exhibit a markedly non-uniform genomic distribution suggestive of selective pressure for and against their formation in particular genomic regions; function-specific assays associate quadruplex folds with a range

of cellular functions, notably regulation of genes associated with growth and development; lastly, the formation and unwinding of quadruplex folds appear to require close management to maintain cell health — regulatory failure correlates with a number of cancer phenotypes [4, 5, 6].

Here we examine, by x-ray crystallography and SAXS, an RNA hexamer, UGGGGU, whose conformation adopts a guanine quadruplex fold. The crystallographic data analyzed in this work, three data sets whose resolution ranges from 0.88 to 1.01 Å, display unusually high resolution relative to current PDB depositions, particularly for RNAs. Thus, independently of insights into quadruplex structure, these data provide an opportunity for reliable quantification of RNA structure and hydration parameters, and this analysis is therefore of more general interest.

A TGGGGT repeat is synthesized by *Tetrahymena* telomerase and was the focus of the initial elucidation of telomerase activity in chromosome maintenance [7]. Its four-stranded, parallel quadruplex conformation has been the subject of numerous structural studies [8, 9]. The RNA counterpart examined here has been the subject of previous NMR [10] and X-ray studies [11]. The DNA and RNA forms share a similar arrangement over the central stack of guanine quartets, but differ in the conformation of capping nucleotides.

Biologically relevant quadruplex folds are anticipated to form transiently in mRNA and in single-stranded, unwound DNA [4]. In both cases formation of stable structures in the substrate may interfere with recognition or processivity by polymerases. This effect has been documented in translational elongation and ribosomal frame shifting [12] and stalled transcription [13]. Understanding effects that contribute to the stability of quadruplex structures can help assess the roles they play. Two features of the present structure bear on the general issue of quadruplex stability: variability in the U tetrad formed by the capping nucleotides and evidence for helix stabilization by ribose O2' atoms.

Capping the regularly-spaced column of stacked G4 quartets poses an interesting struc-

tural challenge for quadruplex nucleic acid conformation [14]. A thymine tetrad at the 5' terminal of a parallel-stranded TGGGGT quadruplex was reported by Cáceres and co-workers [15]. Recently a thymine tetrad was also found in the locked-nucleic acid (LNA) form of the TGGGGT repeat [16]. However, as discussed by Cáceres *et al*, tetrad formation by terminal thymines is only one among several competing arrangements as evident in PDB entries 352D, 244D, 1S45 and 1S47. In contrast, the RNA form of this sequence appears to invariably favor U tetrads as a capping conformation. Formation of U tetrads and the unusual stability of the resulting structure was first reported by Cheong and Moore [10]. More recently, formation of U tetrads was shown to contribute to stabilization in human telomeric RNA repeats, UAGGGU [17, 18]. Overall, U tetrads appear to provide a more stable solution for capping the quadruplex complex than the splayed-out thymines evident in the DNA structures. However, as reported in Section 2.4, there is considerable variation in how U tetrads stack over guanines quartets.

Thermodynamic studies indicate the RNA form of quadruplex structures is often more stable than DNA [19, 20]. In Section 2.8 we show that in guanine residues the ribose O2' hydroxyl is heavily involved in networks of water-mediated hydrogen bonds that link quadruplex chains. The hydrogen bonds form both inter and intra-helix contacts that stabilize the right-handed helical fold formed from the core guanine residues. Taken together, these two effects likely to contribute to the unusual stability exhibited by this structure and provide insight into related RNA sequences.

2 Results and Discussion

2.1 Oligo Assemblies

Crystallographic data reduction and refinement statistics for the three data sets examined are shown in Table 1. Two of the data sets share the tetragonal, $P 4_2 2_1 2$, space group of the

Space group:	P 4 21 2		C 2 2 21
Data set	P1A	P1B	P2C
Cell: a, b c ($\text{\AA}$)	36.65 - 37.08	36.62 - 37.38	37.66 54.20 95.68
Resolution range ($\text{\AA}$)	36.68-0.88 (0.93-0.88)	18.31-0.92 (0.97-0.92)	95.68-1.01 (1.06-1.01)
R_{meas}	0.082(0.249)	0.153(0.631)	0.103(0.946)
R_{pim}	0.024(0.157)	0.033(0.205)	0.030(0.364)
Number of observations	139429(2889)	391879(15631)	624238(46394)
Number of unique observations	19031(1684)	17321(1759)	51294(7088)
Mean(I)/sd(I)	18.500(2.400)	21.900(5.300)	18.000(2.600)
CC(1/2)	0.998(0.945)	0.989(0.824)	0.997(0.748)
Completeness (%)	93.50(58.80)	95.00(69.30)	99.20(94.90)
Multiplicity	7.30(1.70)	22.60(8.90)	12.20(6.50)
Refinement			
R work	0.093	0.091	0.102
R free	0.098	0.109	0.119
Bond length RMSD ($\text{\AA}$)	0.013	0.011	0.012
Bond angle RMSD ($^{\circ}$)	1.744	1.568	1.670
Chiral volume RMSD ($\text{\AA}^3$)	0.106	0.096	0.100

Table 1: *Crystallographic data reduction and refinement statistics [21]. Entries in parentheses apply to the outermost shell.*

previously reported structure (PDB 1J8G) [11], whereas the third exhibits a novel crystal form in an orthorhombic space group, $C 2 2 2_1$.

In all unit cell arrangements observed, the overall crystal packing is arranged in infinite columns comprised of coaxially-stacked, oblong, octamers. The columns are parallel to the cell **c** and **a** axes in tetragonal and orthorhombic space groups respectively. Each octamer consists of two stacked and intercalated tetramers. Four UGGGU oligos with parallel polarity wrapped about a central axis with a right-handed helical twist make up a tetramer, as shown in Fig. 2 (a) and (b). The two tetramers that form the octameric unit are arranged with opposing polarity. Thus the center of the octamer is formed by the intercalated 5' termini of two tetramers whereas the 3' termini form the ends of the octamer.

Within each oligo, the four guanine residues and the 3' uridine point towards the central axis and the orbit resulting from rotation about the four-fold yields the stacked G-quartets characteristic of a G4 fold as shown in Figure 2(b). The 5' and 3' terminal uridines adopt very different conformations. The vector corresponding to the C1'-N1 bond of the 3' uridine points towards the central axis but diagonally away from it in the 5' uridine, though in both cases the χ angle is in the *anti* range. The G-quartets formed by four hydrogen-bonded guanine bases are approximately planar and all nucleotides adopt an *anti* conformation about the glycosidic bond with χ values around -145° , near the center of the *anti* range.

Cations that stabilize the stacked G-quartets are positioned along the four-fold axis. In these structures, Sr^{2+} cations are invariably positioned between two G-quartets, as shown in magenta in Figure 2(b) and (d), and Na^+ cations coordinate the 3' U tetrads.

2.2 Column Arrangement and Crystal Packing

The columns formed by stacked octamers are arranged differently in the two space groups observed as shown in Figure 3. In the tetragonal space group, each column is in contact with four neighbors. The cations at the center of the neighboring columns form a square

1
2
3
4 and a relatively large volume of bulk solvent fills the space near the center of this square.
5
6 In the orthorhombic space group, each column is in contact with six neighbors. The column
7
8 centers form a hexagon and nearly all the solvent volume is taken up by interstitial waters
9
10 that form networks of hydrogen bonds around each column.
11
12
13

14 In both cases, the overall symmetry of the crystal results from application of multiple
15
16 symmetry operations to the hexamer chains. However the division between crystallographic
17
18 and non-crystallographic operations differs between space groups. In the tetragonal case,
19
20 two strands, related by an NCS two-fold, occur in the asymmetric unit (ASU). Octamers
21
22 result from rotation of these strands about the space group's four-fold axis.
23
24

25 Orthorhombic space groups include no crystallographic four-fold operator. Here, eight
26
27 RNA chains occur in the ASU. Of these, four form a complete tetramer whose chains are
28
29 related by four-fold NCS. The partner tetramer is formed by a crystallographic two-fold.
30
31 The remaining four chains in the ASU, which are related by two NCS two-fold operations,
32
33 form an adjacent octamer by rotation around the space group's two-fold axis.
34

35 Though data sets P1A and P1B share the $P 4_1 2_1 2$ space group of structure 1J8G, the
36
37 unit cell is only approximately half as large. Whereas the 1J8G unit cell spans two octamers
38
39 along the **c** axis, here it only spans a single octamer.
40
41
42

43 2.3 SAXS

44
45
46 Solution small angle scattering data were collected for UGGGGU samples over a range of
47
48 concentrations and exposure times as described in the Methods section. Scattering curves
49
50 indicated good contrast and a mono-disperse solution with little aggregation or radiation
51
52 damage yielding data amenable to more detailed analysis. The Kratky curves indicate a
53
54 folded conformation. Intensity and Kratky plots for a representative sample are shown in
55
56 Fig 4(a)-4(b) and derived parameters are given in Table 2.
57
58
59
60
61
62
63
64
65

1
2
3
4 The estimates of R_g and D_{max} are in good agreement with the dimensions of the octamer
5 augmented by a hydration layer. The molecular weight estimate of 14.3kDa, obtained via
6 the volume-of-correlation method [22], is in reasonable agreement with the value of 17.2kDa
7 expected for an octamer and, along with the R_g and D_{max} estimates, excludes larger mul-
8 timeric assemblies in solution.
9

10
11 The fit of the molecular envelope obtained from an *ab initio* model to the crystallographic
12 structure is shown in Fig. 4(c). A comparison of the experimental scattering curve, merged
13 over all concentrations, with predicted scattering from octameric and tetrameric models
14 derived from structure P1B is shown in Fig. 4(d).
15

16
17 Overall, these scattering data suggest the preferred conformation in solution is an oc-
18 tameric assembly with dimensions in good agreement with the crystal form. This result
19 differs from the tetramer conformation observed via NMR analysis by Cheong and Moore
20 [10], a difference likely due to the identity of the central coordinating cation, K^+ versus Sr^{2+} .
21 The backbone geometry of the NMR structure, as reported in their Table IV, shows only
22 rough agreement with the torsion angles of structures P1A, P1B and P2C (Supplemental
23 Table 6). It is noteworthy that angles for which they report the widest variation among the
24 three NMR models for nucleotides G5 and U6, G5- ζ and U6- α , are the same angles that show
25 pronounced variation in conformations 3'U-A, 3'U-B and 3'U-C of the crystal structures as
26 discussed in Section 2.4.
27

28
29 At higher scattering angles, discrepancy in intensity between that predicted for the oc-
30 tamer derived from the crystallographic model and the experimental solution data is appar-
31 ent. Conformations of the octameric assemblies in solution could differ from the crystallo-
32 graphic model, for example, by populating multiple U-tetrad conformations.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Guinier-based $I(0)(\text{cm}^{-1})$, R_g (Å)	4707.3 ± 6.05	14.77 ± 0.002
P(R)-based $I(0)(\text{cm}^{-1})$, R_g (Å)	4654 ± 1.3	14.63 ± 0.003
D_{max}	45.82	
Porod volume	16711	

Table 2: *Derived parameters from SAXS data collection.*

2.4 Structural Variation and Disorder

Guanine tetraplex structures are distinguished among nucleic acid conformations by remarkable stability. As outlined in Fig. 1, three distinct bonding effects contribute to this stability: eight hydrogen bonds that link the Hoogsteen face of each guanine to the Watson-Crick face of its neighbor occur in each quartet; a universally present, dehydrated, central cation forms electrostatic links to the O6 carbonyl oxygen atoms of four or eight surrounding guanines; thirdly, the planes of adjacent quartets are offset at angles that favor π - π stacking interactions [3, 14].

The stability of stacked G quartets is evident in the distribution of B factors shown in Figure 5. The base moiety of the four pairs of G quartets coordinated by a central Sr^{2+} cation form the most stable elements in the structure. Conversely, the more disordered parts of the octamer assembly are the outward-directed 5' uridines and the 3' U-tetrads at either end of the octamer. Among atoms with the highest B factors, there is clear evidence of structural variation; three instances are prominent.

The first involves alternate arrangements of 3' U tetrad. Superposition over the first 5 residues in 5'-3' order of sixteen oligo chains from four structures, P1A, P1B, P2C and 1J8G with THESEUS [23] is shown in Fig. 6(a). There is very good structural alignment over three guanine residues and minor variation in the 3' guanine and 5' uridine. The average, pairwise least-square RMSD over these residues is 1.687Å. However the 3' uridine can adopt three distinct conformations, labeled 3'U-A, 3'U-B, 3'U-C and colored in red, green and magenta. The conformations are not equally populated, with 3'U-A appearing most frequently. 3'U-A

1
2
3
4 and 3'U-B are related by mirror symmetry and result from a nearly 180 difference in ζ and α
5 angles flanking the phosphodiester linkage between the last two residues. Average values for
6 3'U-A residue 5 ζ and residue 6 α are 58.4 and 159.0 relative to 284.6 and 298.2 for 3'U-B.
7
8
9

10 Conformation 3'U-C results from a significantly smaller α angle coupled with a different
11 sugar pucker. This oligo chain is unique in adopting a c3'endo pucker at residue 6 with phase
12 angle of 21.6.
13
14
15

16 Octamers can form from different combinations of these conformations, utilizing either
17 two inward pointing U-tetrads, as occurs in structures P1A and 1J8G, or with one inward
18 U-tetrad and one outward, as occurs in structure P1B.
19
20
21

22 Crystal packing requires the stacking of octamers into columns. At the junction between
23 two stacked octamers, a void forms in one of two ways. If both 3' U tetrads point toward
24 the center of the respective octamer in a concave arrangement, the void forms between two
25 octamers. In contrast, a convex arrangement at one end of the octamer results when the
26 octamer includes chains with both inward pointing and outward pointing 3' U tetrads. In
27 this case an internal void forms between the outward pointing U tetrad and its neighboring G
28 tetrad, within the same octamer. In Fig. 5 structures P1A and P1B exhibit, respectively, the
29 concave and convex conformations. We did not observe a crystal form exhibiting a convex
30 arrangement at both ends of the octamer, though the *ab initio* model derived from SAXS
31 data does not seem to preclude this possibility.
32
33
34
35
36
37
38
39
40
41
42
43

44 The RNA structure surrounding the void between octamers is the least well-ordered
45 region of the octamer as evident in Fig. 5. The occurrence of different conformations for
46 the 3' uridine shown in Fig. 6(a) suggests an ensemble of distinct arrangements. However,
47 within a given crystal, a single conformation occurs predominantly and there is insufficient
48 density to model alternate conformations of the 3' U tetrad. Selection of "concave" or
49 "convex" octamer conformations appears fixed for a given crystal packing. Though some
50 water molecules are apparent in the void, they are not clearly resolved and significant residual
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

density remains unmodeled in this region.

A second instance of structural variation occurs in the the phosphate group that connects uridine 1006 to guanine 1005 in chain B of structure P1A as shown in Figure 6(b). A similar alternate conformation was detected in the TGGGGT structure deposited as PDB entry 352D [8], but not in a previously published UGGGGU structure, 1J8G [11]. The 5' uracils are the least ordered bases in the octamer. However there is little evidence to support extension of the alternate conformation beyond the phosphate group. Density for the corresponding *O1P* and *O2P* oxygens only becomes visible at about the 1.5σ level and displacement of the *O3'* and *O5'* oxygens appears minimal.

A third instance of structural variation occurs in structure P2C in chains 'A' and 'F'. Once again, the 3' uridines are involved, however here the alternate conformations extend over the entire terminal residue rather than being limited to the phosphodiester linkage. In chain 'A', the alternate conformation begins at *O3'* of the n-2 guanine, residue 104. A key distinguishing characteristic of the two conformations are different values of the ζ torsion angle at the linkage between residues 104 and 105: -82.2 for conformation A and -64.4 for conformation B. In chain 'F' the alternate conformation begins at *C1'* of the n-1 guanine, residue 605, and a distinguishing feature of the two conformations is the difference in ϵ values, -148.5 versus -160.7. In both cases, the net effect of the two conformations is a lateral shift of the 3' residue that maintains the angle of the base relative to the central cation axis.

It is noteworthy that the alternate conformations in structure P2C appear to be near the limit of what can be observed by diffraction in this relatively disordered part of the structure. Crystal packing in the orthorhombic space group enables observation of multiple chains within the same tetramer component: chains 'A'-'D' belong to the same tetramer assembly, chains 'E' and 'F' belong to another and chains 'G' and 'H' belong to a third; furthermore, the two tetramers in chains 'E'-'H' belong to the same octamer assembly. Though the abundance of independent chains within the ASU allows observation of multiple constituents

of the capping 3' U-tetrads, alternate conformations could only be modeled for one chain within each tetramer.

2.5 Cation Coordination and Quartet Stacking

The close packing of bases characteristic of quartet structures places the carbonyl oxygens, i.e., the O6 atoms in G-quartets, and the O4 atoms in U-quartets, within unusually close proximity. The repulsion between the partial negative charges residing on neighboring oxygen atoms is screened by inner-sphere coordination of dehydrated cations, such as Sr^{2+} , in the central channel formed by the stacked quartets. Coordination observed in the UGGGGU structure exhibits three distinct patterns, one defined by the four central guanine quartets and two others defined by quartets formed by the 3' uridines that cap the ends of the octameric unit. As described previously, the 5' uridines rotate away from the central channel and thus do not participate in quartets or cation coordination.

The coordination geometry of the guanine quartets is invariant across all structures examined in this work and is summarized in Figure 7(a). The coordinating cation is always Sr^{2+} , centrally positioned between the eight O6 carbonyl oxygens of two adjacent G-quartets in a bipyramidal-antiprism geometry. The Sr-O6 distance, at 2.60 Å, is shorter than O6-O6 distances to atoms in either the same or adjacent quartets, 3.16 Å and 3.54 Å respectively. The guanines in each quartet are tilted away from a common plane and towards the Sr^{2+} cation; a cation occurs only between every other quartet. This alternating distribution, which allows a separation of 6.43 Å between adjacent cations appears to be unique to Sr^{2+} , and accommodates their electrostatic repulsion [24]. Though Sr^{2+} has no direct biological relevance, this cation is known to confer high stability to the resulting quadruplex structure [25] thus enabling more detailed examination of the fold's geometry and of the capping U-tetrads.

Within each octamer, 32 O6 carbonyl oxygens are wrapped about the four central Sr^{2+}

1
2
3
4 cations in a helical arrangement as shown in Figures 7(b)-7(c). As evident from the distri-
5 bution of B factors summarized in Figure 5, cation coordination by G quartets yields the
6 most stable, well-ordered, part of the overall structure.
7
8
9

10
11
12 Coordination of uridine O4 atoms follows a similar but less well-defined pattern. Here an
13 Na^+ cation with split-occupancy centered on the central axis serves to screen accumulated
14 charge. The U-tetrad's four O4 carbonyl atoms define a square of approximately 3.27 Å.
15
16 Two adjacent sites near the center of the square are alternately occupied by a Na^+ cation.
17
18 The most frequently occupied site is coplanar with the oxygens whereas the site with minor
19 occupancy lies 2.2-2.3 Å out of the plane, away from the octamer. As discussed in Section
20 2.4 the two 3' U tetrads which occur at each octamer-octamer boundary can adopt either a
21 concave or convex disposition. In the concave arrangement, the two U tetrads tilt away from
22 one another towards the center of their respective octamer. In this case, the coordination
23 pattern by the central Na^+ cation is similar for the two tetrads. However in the convex
24 arrangement, both U-tetrads tilt towards the same octamer. In this case, the two U-tetrads
25 are brought close to one another and the two alternate position of a single Na^+ cation appear
26 to coordinate both U-tetrads.
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 2.6 Geometric Outliers

42
43
44 Maintaining the regularity of stacked guanine and uridine tetrads in the octameric unit im-
45 poses constraints which tax the flexibility of the RNA backbone. Accordingly, the structures
46 exhibit some features which, though well-supported by the data, are unusual with respect to
47 nucleotide geometry: not all nucleotide conformations match the clusters of torsion angles
48 used in the definition of rotamers for the RNA backbone [26]; there is a marked deviation
49 from planarity between the guanine and sugar moieties; some bonds and angles fall outside
50 the center point of distributions derived from reference structures.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

The RNA backbone has been shown to be rotameric [26]. Nevertheless for certain UGGGGU residues, rotamer assignment as implemented in *molprobability* version 4.1 and *suitename* version 0.3.070628 [26] does not identify a known cluster. For example, for structure P1B, no rotamer assignment is reported for residues 1004 in chain 'A' and 2004 in chain B. For both, the $\delta_{-1}\delta\gamma$ classification is "23p". Inspection indicates that the observed ϵ_{-1} torsion angles of the two residues, with values of -170 and -172 degrees respectively, do not match any peak among currently defined "23p" rotamers.

The strong electrostatic interaction between O6 and Sr^{2+} atoms distorts both the planar arrangement of atoms within individual guanine residues and the planar arrangement of residues in a G4 quartet. The deviation from a plane is most pronounced at the ribose C1'. For example, in structure P1A, the connecting ribose C1' of residues 1002 and 1003 are markedly below and above the plane defined by the guanine atoms: seven and eight times the plane's RMS deviation respectively.

An instance where the orderly arrangement of stacked quartets strains the RNA backbone occurs in structure P1A. The "concave" U-tetrad conformation of chain "A" requires the terminal uridine to stack above the adjacent 3' guanine in a direction opposed to the strand's helical twist. This geometry places a tight constraint on the connecting phosphodiester bond. Two conformations of the C3'-O3'-P-O5'-C5' linkage are evident. Though both are well-supported by 2Fo-Fc density as shown in Fig. 6(c), the C2'-C3-O3 angle and C5'-O5' bond length for one of the conformations falls over five standard deviations outside the centroid of the reference distribution. A similar situation occurs at the 3' guanine-uridine linkage of the tetramer formed by the structure's "B" chain.

2.7 Solvent Ions

Ordered Sr^{2+} and Ca^{2+} ions are evident in the solvent channels between adjacent octamers. They are distributed at regular intervals and appear to serve two distinct functions: stabi-

lizing the geometry of 3' and 5' uridines or stabilizing octamer assembly and crystal packing via contacts that link strands within an octamer and between adjacent octamers.

Examples of the first instance are residues 9004 and 2106 in structure P1B. Residue 9004 forms bifurcated contacts with the ribose O2' and O3' oxygens of the chain's 3' uridine, residue 1006. A contact to water 9111 places the cation near the center of a triangle defined by the three oxygens. Five water molecules, 9110, 9108, 9102, 9112 and 9101 are evenly distributed in a plane perpendicular to this triangle. Each forms contacts with both the cation and either the ribose O2' or O3' oxygens. The location of the metal and the regularly arranged cluster of water molecules suggest a role in stabilizing the unusual backbone geometry that enables formation of the terminal U-tetrad. Residue 2106 appears to serve a similar function near the 5' terminal uridine. A short, 2.48 Å, contact links the Ca²⁺ cation with the uridine O4 and both atoms share similar, low, B-factors of 7.88 and 8.11 respectively.

The locations of other metals suggest roles in stabilizing both the helical twist of neighboring chains which forms the octameric assembly and the packing of adjacent columns that enables crystal formation. Representative examples are Ca²⁺ residues 2104 and 2105 in structure P1B. Residue 2104 forms near-equidistant contacts, of 2.40 and 2.47 Å with the OP1 oxygens of guanine residues from octamers in adjacent columns. The two oxygen atoms are separated by a distance of only 2.80 Å. Contacts to four adjoining waters, 2204, 2209, 2214 and 2225 form a network that further stabilizes cross-column contact. Similarly, residue 2105 forms contacts, of distances 2.39, 2.34 and 2.29 Å with the phosphate oxygens of residues in three distinct chains; two are OP2 oxygens from residues 1002 and 2002 in different chains of the same octamer, the third is the OP1 oxygen from residue 2005 in a neighboring column.

Interestingly, in a number of cases, solvent positions that support inter and intra-octamer linkage are alternately occupied by a metal cation or by a water molecule separated by an Ångström or less. In most instances, the X-ray anomalous signal unambiguously distin-

guishes the identity of the two alternately-occupied scattering peaks and corroborates metal assignment for one of the two peaks.

2.8 Hydration: Distribution of Water Contacts

Hydration of the octameric assembly largely agrees with expectations based on similar nucleic acid structures. No crystallographically-detectable waters are observed in the central channel within each octamer, whereas water contacts occur throughout the external surface and at the cavity formed between stacked octamers. Closer examination of the distribution of water contacts indicates that they are not uniformly distributed over the external surface.

To examine this point, a list of potential contacts was constructed by considering the neighborhood of solvent atoms in each of the UGGGGU structures. As described in Section 4, each contact record includes the source solvent atom, a target atom that lies within 3.6 Å and attributes of the two atoms. The list of 1852 possible water contacts was used to define a simple hydration index for each atom type: the ratio of the number of contacts to the number of occurrences. The results are summarized in Supplemental Table 4.

In Fig. 8(a), the RNA atoms from structure P1A have been colored by value of the atom type's hydration index with thin, blue, lines indicating a value less than 0.25. In Fig. 8(b) the full symmetry-expanded octamer and its surface are colored by hydration value and modeled waters are included.

Distinctly high values of the hydration index occur at uridine C5' atoms. Inspection suggests that the network of water molecules near C5' stabilize the terminal uridines. At the 3' end, hydration of C5' serves to maintain orientation of the capping U-tetrads whereas at the 5' end, C5' hydration helps maintain the strand intercalation that stabilizes octamer assembly.

As expected, phosphate oxygens, which protrude into solvent away from the octamer, exhibit high hydration scores. Significantly, the hydration index of ribose O2' guanine atoms

falls close to this peak, underscoring a stabilizing structural feature accessible to RNA over DNA G-quadruplexes.

2.9 Hydration: Role of the Ribose O2' Hydroxyl Group

Residency times for bound waters are fleeting relative to those of macromolecular atoms [27]. Thus repeated exchange of waters at a given site that occurs frequently enough to permit observation of the site by diffraction implies a structural role for a water molecule at that location. Inspection of water linkages in the UGGGGU data sets confirms three well-established roles for occupied sites: helix stabilization, crystal packing and charge-shielding of phosphates along the RNA backbone.

Structural features mediated through water-based hydrogen bonding can resist definitive classification because of promiscuity and plasticity. Nevertheless, even an approximate and coarse-grained classification is sufficient to allow some general patterns to emerge. To this end we analyzed the distribution of waters across the three data structures examined, focusing on hydration of ribose O2' oxygens. As documented above, the ribose O2' hydroxyl is a well-hydrated group in the UGGGGU structures: on average, each instance is the target of contacts from two water molecules. This finding is consistent with previous analyses of RNA hydration [28, 29]. Hydrogen bonding associated with the 2'-OH group has been suggested to account for greater rigidity of RNA A-duplexes [28]. Experimental data indicates RNA G-quadruplexes are more stable than corresponding DNA structures [30]. We hypothesized that hydration of O2' contributes to this effect and analyzed the list of potential water contacts to determine whether the data support this hypothesis.

For analysis, we aggregated the 1852 possible contacts from 238 waters in two ways. The first considered whether a given water molecule makes contacts exclusively with atoms of a single octamer or whether its contacts span two or more adjacent octamers, typically symmetry-related copies. For example, for structure P1A, the octamer formed by the two

strands in the ASU is linked by water contacts to five adjacent octamers. The distinction between the two groups of water molecules is illustrated in Fig. 9(a). The octamer formed by the ASU strands is shown in thick yellow lines and adjacent octamers linked by water contacts in thin blue. Water molecules limited to intra-octamer contacts are shown in magenta and waters with possible inter-octamer contacts are shown in teal. Most, but not all, intra-octamer waters make contact with the ASU-defined octamer. This classification assigns a given water to one of three disjoint subsets depending on whether its contacts are inter-octamer, intra-octamer or part of a secondary hydration shell which includes no contacts to RNA.

Within each subset, we considered whether the target of a potential water-based contact is a phosphate oxygen or a ribose O2' oxygen. These atom types are among the most frequent targets found as hydration contacts, though the overall distribution of target atom types is quite broad (Supplemental Table 5). The secondary classification does not partition waters into mutually exclusively subsets: among the 238 waters combined from three data sets, 143 include contacts to phosphate oxygens, 96 to O2' oxygens, 44 to both types of targets and 43 to neither.

The properties of potential water contacts classified according to these two levels of aggregation are summarized in Table 3.

As expected, contacts among waters in the secondary hydration layer exhibit higher B factors at both source and target though the mean contact distance remains unaffected. More interestingly, the analysis reveals differences between waters bound to O2' oxygens relative to those bound to the phosphate group. Two differences arise: contacts to O2' are more frequent among waters involved in intra-octamer binding and the locations of those waters are more tightly bound as evidenced by lower B factors. Whereas only 15 of 79(18%) of inter-octamer waters include O2' contacts, 81 of 153(53%) intra-octamer waters do. Though the B factors of contact targets among phosphate oxygens and O2' hydroxyl groups

	Count	Contacts	Distance	Solvent B Factor	Target B Factor
All	238	1852	3.07(0.33)	17.15(8.91)	12.83(7.58)
Phosphate	143	200	2.94(0.29)	17.51(8.40)	10.84(2.70)
Ribose O2'	96	126	2.98(0.30)	14.95(7.80)	11.23(7.43)
Secondary shell	6	16	2.79(0.33)	33.86(16.42)	20.96(10.09)
Inter-octamer	79	648	3.08(0.33)	18.73(10.27)	14.14(8.58)
Phosphate	54	92	2.98(0.31)	17.25(9.28)	10.67(2.78)
Ribose O2'	15	16	2.98(0.37)	18.49(10.14)	15.23(14.77)
Intra-octamer	153	1188	3.07(0.33)	16.06(7.52)	12.00(6.77)
Phosphate	89	108	2.90(0.26)	17.73(7.62)	10.99(2.64)
Ribose O2'	81	110	2.98(0.29)	14.44(7.31)	10.65(5.53)

Table 3: *Contact Properties. Properties of overall water contacts in comparison to three disjoint subsets: waters in contact with phosphate group oxygens, with the ribose O2' hydroxyl group and with waters in a secondary hydration shell. For each group, the number of contacts found among all three structures with the mean and standard deviation of the contact distance, target B factor and solvent B factor are listed.*

are very similar, the B factors of the waters are significantly lower for contacts to the O2' group. This difference, though minor, is significant at the 0.005 level among all waters and at 0.001 among intra-octamer waters.

Overall, these data indicate the displacement of solvent positions near the O2' group, most of which form intra-octamer contacts, is more tightly constrained than displacement of positions involved in charge-shielding along the RNA backbone. An example of the strand-linking stabilization provided by water contacts to the O2' hydroxyl groups in structure P1B is illustrated in Fig. 9(b). The O2' groups of residues 1001 and 1002 from chain 'A' and 2001 and 2002 from chain 'B' form a network of hydrogen bonds with waters 9505 and 9521.

3 Conclusions

The findings reported here contribute to the understanding of quadruplex folds and RNA structure. Evidence that mammalian telomeric repeats, TTAGGG, are selectively transcribed [31] has increased interest in RNA quadruplexes. The r(UAGGGU) repeat was

1
2
3
4 shown to adopt a parallel-stranded G-quadruplex conformation capped and stabilized by
5
6 a U-tetrad [17, 18]. Here we show that the terminal U-tetrads in the previously-reported
7
8 UGGGGU crystal structure [11] exhibit considerable conformational variability.
9

10 Detailed analysis of hydration patterns in the structures reported confirms that the O2'
11
12 hydroxyl group plays a major role in stabilization of inter and intra-strand contacts that
13
14 support the quadruplex helical fold. Comparisons of the stability of similar DNA and RNA
15
16 quadruplex-forming sequences have shown that the RNA form often exhibits greater stability
17
18 [19, 20]. Our analysis supports the hypothesis that networks of water-mediated contacts
19
20 within the helical grooves of RNA quadruplexes contribute to enhanced stability. However
21
22 our investigation is limited to the structures presented. A more extensive comparison of
23
24 hydration across diverse DNA and RNA quadruplexes is needed to assess the importance of
25
26 this effect.
27
28

29 The high resolution is unusual for RNA, and contributes to the understanding of RNA
30
31 geometry. In particular, some features of these structures, including bond lengths, angles,
32
33 and tuples of torsion angle values used for rotamer assignment that are flagged as outliers
34
35 by current structure validation software are shown to be legitimate expressions of RNA
36
37 conformation.
38
39

40 The SAXS-derived solution structure for UGGGGU shows good agreement with the
41
42 octameric assembly observed in the crystal structure and suggests mono-disperse collections
43
44 of the octamer readily form in solution.
45
46

47 48 49 **4 Materials and Methods** 50 51

52 RNA oligos were purchased from Dharmacon and deprotected per the manufacturer's in-
53
54 structions. Crystals were grown using previously reported conditions [11]. Usable crystals
55
56 grew within one to two weeks. Variation in the neighborhood of these conditions did not
57
58
59
60
61
62
63
64
65

1
2
3
4 appear to affect crystal quality consistently. However, high mosaicity and irregular growth
5
6 yielding multiple lattices were a common problem and required screening multiple crystals
7
8 to obtain high-quality data sets. Growth of irregular crystals was significantly improved by
9
10 a simple microseeding protocol [32].
11

12
13 The thirty three data sets used for analysis were collected on beamlines 7-2, 9-1,9-2
14
15 and 11-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) from seventeen single
16
17 crystals selected from a larger pool of candidates. For several crystals, multiple image batches
18
19 were collected to optimize anomalous differences or improve data statistics via high and low
20
21 resolution passes. Each of the final P1A, P1B and P2C data sets results from images collected
22
23 from individual single crystals.
24

25
26 For several crystals, multiple data sets were collected, either to optimize anomalous
27
28 differences or to improve data statistics via high and low resolution passes.
29

30
31 Data reduction used *mosflm*, *scala*, *pointless* and *aimless* from the CCP4 [33, 21] suite
32
33 as well as programs from the PHENIX and *labelit* [34, 35] suites.
34

35
36 Preliminary model fitting was based on iterative rigid body refinement with *phenix.refine*
37
38 [35] in the tetragonal space group and by molecular replacement with *phaser* [36] in the
39
40 orthorhombic space group.
41

42
43 Structure refinement relied on the *refmac* [37], and *shelxl* [38] programs. Occupancy
44
45 refinement with *shelxl* was applied to all solvent, to alternate conformations, and, in certain
46
47 chains, to the terminal, less-ordered, uridine residues. B factors were refined individually
48
49 and anisotropically other than in regions where weak density did not support the additional
50
51 parameterization. Non-hydroxyl hydrogens were modeled and refined with *shelxl* except for
52
53 residues modeled in weak density. Refinement solutions obtained from *shelxl* were input to
54
55 *refmac* for bulk-solvent modeling.
56

57
58 Anomalous difference maps were computed with the CCP4 *fft* program and with the
59
60 *phaser* log-likelihood gradient module and used for identification of solvent ions.
61
62
63
64
65

1
2
3
4 Waters were placed by means of a novel approach based on partitioning density difference
5 maps into Morse-Smale complexes [39]. Briefly, a preliminary model was built and prelimi-
6 nary waters and ions placed via the *coot* [40] “findwaters” command. All waters were then
7 removed and the resulting difference map was partitioned into basins bounded by zero flux
8 surfaces [41]. Basins enclosing an acceptable total density with a peak within a plausible
9 distance of non-hydrogen atoms of the existing working model were selected as candidate
10 waters. Those yielding improvements in model quality were retained for further rounds of
11 refinement. The procedure was iterated outward from the working model to uncover waters
12 in outer hydration shells.
13
14

15 A number of adjacent solvent sites were detected from their elongated profile in difference
16 density maps and modeled as alternate split-occupancy locations. Partial occupancy of
17 waters and solvent ions was modeled, however occupancies that refined over 0.9 were assigned
18 full occupancy and waters whose occupancies refined below 0.20 and which did not fall on
19 special positions were discarded.
20
21

22 Solvent contact statistics were calculated in two steps. First solvent positions were
23 scanned to determine all atoms, including symmetry equivalents, lying within 3.6 Å [42].
24 For solvent positions with alternate occupancy, both positions were considered but atoms
25 with alternate occupancy within the radial distance sphere were excluded. Thus, for each
26 potential contact, either the source or target but not both may occupy an alternate position.
27 The resulting list of potential contacts were annotated with properties of the source and
28 target atoms, including occupancy, B factor, and peak height in the associated 2Fo-Fc map
29 and analyzed with R [43]. Aggregating potential contacts over structures P1A, P1B and
30 P2C yielded a data set of 2364 contact records and 267 solvent entries. For most analyses,
31 we focused on the subset of 1852 contacts from 238 water molecules.
32
33

34 SAXS data were collected on SSRL beamline 4-2 for an RNA concentration series in the
35 range of 0.4 to 4 mg/ml in a solution of 10mM sodium cacodylate, 10mM MgCl₂, 40mM
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 SrCl₂ and 20mM CacCl₂. Samples were allowed to equilibrate against buffer overnight in
5
6 50μL dialysis buttons (Hampton Research). For each sample, scattering from ten one second
7
8 exposures of sample and two exposures of buffer was recorded. Data analysis relied on the
9
10 ATSAS suite of programs [44]. *Ab initio* molecular models were built from fifteen iterations of
11
12 DAMMIF, aligned with SUPCOMB, averaged with DAMAVER and filtered with DAMFILT.
13
14 The symmetry expanded octamer derived from structure P1B was fit to the resulting bead
15
16 model with the CHIMERA “Fit-toMap” routine [45]. Predicted scattering profiles for models
17
18 derived from crystallographic coordinates were computed with FoXS [46].
19
20
21
22

23 **5 Acknowledgments**

24
25
26 Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Lab-
27
28 oratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic
29
30 Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular
31
32 Biology Program is supported by the DOE Office of Biological and Environmental Research,
33
34 and by the National Institutes of Health, National Institute of General Medical Sciences
35
36 (including P41GM103393) . The contents of this publication are solely the responsibility of
37
38 the authors and do not necessarily represent the official views of NIGMS or NIH.
39
40
41
42
43

44 **6 Accession numbers**

45
46
47 Data sets P1A, P1B and P2C have been deposited in the PDB with codes 4RKV, 4RJ1 and
48
49 4RNE respectively.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

- [1] D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, *Nature* 334 (6180) (1988) 364–366.
- [2] A. T. Phan, V. Kuryavyi, D. J. Patel, DNA architecture: from G to Z, *Curr Opin Struct Biol* 16 (3) (2006) 288–298.
- [3] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, *Nucleic Acids Res* 34 (19) (2006) 5402–5415.
- [4] N. Maizels, Quadruplexes and the Biology of G-Rich Genomic Regions, in: S. Neidle, S. Balasubramanian (Eds.), *Quadruplex Nucleic Acids*, 2006.
- [5] J. L. Huppert, Hunting G-quadruplexes, *Biochimie* 90 (8) (2008) 1140–1148.
- [6] S. Neidle, *Therapeutic Applications of Quadruplex Nucleic Acids*, Academic Press, 2012.
- [7] C. W. Greider, E. H. Blackburn, A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis, *Nature* 337 (6205) (1989) 331–337.
- [8] K. Phillips, Z. Dauter, A. I. Murchie, D. M. Lilley, B. Luisi, The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution, *J Mol Biol* 273 (1) (1997) 171–182.
- [9] G. Laughlan, A. I. Murchie, D. G. Norman, M. H. Moore, P. C. Moody, D. M. Lilley, B. Luisi, The high-resolution crystal structure of a parallel-stranded guanine tetraplex, *Science* 265 (5171) (1994) 520–524.

- [10] C. Cheong, P. B. Moore, Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures, *Biochemistry* 31 (36) (1992) 8406–8414.
- [11] J. Deng, Y. Xiong, M. Sundaralingam, X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent $\text{Sr}(2+)$ ions at subatomic resolution (0.61 Å), *Proc Natl Acad Sci U S A* 98 (24) (2001) 13665–13670.
- [12] T. Endoh, N. Sugimoto, Unusual -1 ribosomal frameshift caused by stable RNA G-quadruplex in open reading frame, *Anal Chem* 85 (23) (2013) 11435–11439.
- [13] B. P. Belotserkovskii, R. Liu, S. Tornaletti, M. M. Krasilnikova, S. M. Mirkin, P. C. Hanawalt, Mechanisms and implications of transcription blockage by guanine-rich DNA sequences, *Proc Natl Acad Sci U S A* 107 (29) (2010) 12816–12821.
- [14] C. J. Lech, B. Heddi, A. T. Phan, Guanine base stacking in G-quadruplex nucleic acids, *Nucleic Acids Res* 41 (3) (2013) 2034–2046.
- [15] C. Cáceres, G. Wright, C. Gouyette, G. Parkinson, J. A. Subirana, A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl^+/Na^+ ions, *Nucleic Acids Research* 32 (3) (2004) 1097–1102.
- [16] I. Russo Krauss, G. N. Parkinson, A. Merlino, C. A. Mattia, A. Randazzo, E. Novellino, L. Mazzearella, F. Sica, A regular thymine tetrad and a peculiar supramolecular assembly in the first crystal structure of an all-LNA G-quadruplex, *Acta Crystallogr D Biol Crystallogr* 70 (Pt 2) (2014) 362–370.
- [17] T. Kimura, Y. Xu, M. Komiyama, Human telomeric RNA r(UAGGGU) sequence forms parallel tetraplex structure with U-quartet, *Nucleic Acids Symp Ser (Oxf)* (53) (2009) 239–240.

- [18] Y. Xu, T. Ishizuka, T. Kimura, M. Komiyama, A U-tetrad stabilizes human telomeric RNA G-quadruplex structure, *J Am Chem Soc* 132 (21) (2010) 7231–7233.
- [19] J. L. Mergny, L. Lacroix, UV Melting of G-Quadruplexes, *Curr Protoc Nucleic Acid Chem* Chapter 17.
- [20] A. Joachimi, A. Benz, J. S. Hartig, A comparison of DNA and RNA quadruplex structures and stabilities., *Bioorganic & medicinal chemistry* 17 (19) (2009) 6811–6815.
- [21] P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality., *Science* (New York, N.Y.) 336 (6084) (2012) 1030–1033.
- [22] R. P. Rambo, J. A. Tainer, Accurate assessment of mass, models and resolution by small-angle scattering, *Nature* 496 (7446) (2013) 477–481.
- [23] D. L. Theobald, D. S. Wuttke, THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures, *Bioinformatics* 22 (17) (2006) 2171–2172.
- [24] J. Plavec, Metal Ion Coordination in G-Quadruplexes, in: N. Hadjiliadis, E. Sletten (Eds.), *Metal complexes: DNA interactions*, Blackwell Publishing Ltd, 2009.
- [25] P. Tran, A. De Cian, J. Gros, R. Moriyama, J.-L. Mergny, Tetramolecular Quadruplex Stability and Assembly, in: J. B. Chaires, D. Graves (Eds.), *Quadruplex Nucleic Acids*, Vol. 330 of *Topics in Current Chemistry*, Springer Berlin Heidelberg, 2013, pp. 243–273.
- [26] J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. HersHKovits, L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook, H. M. Berman, RNA Ontology Consortium, RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution), *RNA* 14 (3) (2008) 465–481.

- [27] M. Levitt, B. H. Park, Water: now you see it, now you don't, *Structure* 1 (4) (1993) 223–226.
- [28] H. M. Berman, B. Schneider, Nucleic Acid Hydration, in: S. Neidle (Ed.), *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press, 1999.
- [29] M. Sundaralingam, B. Pan, Hydrogen and hydration of DNA and RNA oligonucleotides, *Biophys Chem* 95 (3) (2002) 273–282.
- [30] J. Mergny, J. Gros, A. De Cian, A. Bourdoncle, F. Rosu, B. Sacca, L. Guittat, S. Amrane, M. Mills, P. Alberti, M. Takasugi, L. Lacroix, Energetics, Kinetics and Dynamics of Quadruplex Folding, in: S. Neidle, S. Balasubramanian (Eds.), *Quadruplex Nucleic Acids*, 2006.
- [31] S. Schoeftner, M. Blasco, Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II, *Nature Cell Biology* 10 (2) (2008) 228–236.
- [32] T. Bergfors, Succeeding with seeding: some practical advice, in: R. J. Read, J. L. Sussman (Eds.), *Evolving Methods for Macromolecular Crystallography*, Vol. 245 of NATO Science Series, Springer Netherlands, 2007, pp. 1–10.
- [33] Collaborative Computational Project, Number 4. 1994, *The CCP4 Suite: Programs for Protein Crystallography* (1994).
- [34] N. K. Sauter, R. W. Grosse-Kunstleve, P. D. Adams, Robust indexing for automatic data collection, *Journal of Applied Crystallography* 37 (3) (2004) 399–409.
- [35] P. D. Adams, R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, T. C. Terwilliger, PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr D Biol Crystallogr* 58 (Pt 11) (2002) 1948–1954.

- [36] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software, *J Appl Crystallogr* 40 (Pt 4) (2007) 658–674.
- [37] G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr D Biol Crystallogr* 53 (Pt 3) (1997) 240–255.
- [38] G. M. Sheldrick, A short history of SHELX, *Acta Crystallogr A* 64 (Pt 1) (2008) 112–122.
- [39] H. King, K. Knudson, N. Mramor, Generating Discrete Morse Functions from Point Data, *Experimental Mathematics* 14 (4) (2005) 435–444.
- [40] P. Emsley, K. Cowtan, Coot: Model-Building Tools for Molecular Graphics, *Acta Crystallographica Section D - Biological Crystallography* 60 (2004) 2126–2132.
- [41] Fyfe AC, Scott WG, Volumetric segmentation of crystallographic electron density maps, In preparation.
- [42] K. D. Cowtan, The Clipper Project, in: *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, Vol. 40, 2002.
- [43] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2013).
- [44] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, D. I. Svergun, New developments in the *ATSAS* program package for small-angle scattering data analysis, *Journal of Applied Crystallography* 45 (2) (2012) 342–350.
- [45] T. Goddard, C. Huang, T. Ferrin, Visualizing density maps with UCSF Chimera, *Journal of Structural Biology* 157 (1) (2007) 281–287.

- 1
2
3
4 [46] D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, Accurate SAXS profile
5 computation and its assessment by contrast variation experiments., Biophysical journal
6 105 (4) (2013) 962–974.
7
8
9
10
11 [47] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, Conforma-
12 tional analysis of nucleic acids revisited: Curves+, Nucleic Acids Res 37 (17) (2009)
13 5917–5929.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Atom Type	Occurrences	Potential Contacts	Ratio
C5'_U	22	79	3.591
C2'_G	46	106	2.304
C5_U	22	50	2.273
OP2_G	47	96	2.043
O2'_G	46	90	1.957
OP1_G	47	92	1.957
C4'_G	47	91	1.936
C5'_G	47	87	1.851
C1'_G	47	86	1.830
C3'_U	22	39	1.773
O2'_U	22	36	1.636
C5_G	48	78	1.625
O4_U	22	33	1.500
C4'_U	22	28	1.273
C6_U	22	28	1.273
N2_G	48	60	1.250
O3'_U	22	24	1.091
O5'_U	21	20	0.952
O3'_G	45	40	0.889
O4'_G	47	39	0.830
OP1_U	9	7	0.778
C2'_U	22	16	0.727
N3_G	48	31	0.646
O2_U	22	14	0.636
C3'_G	46	29	0.630
C1'_U	22	13	0.591
OP2_U	9	5	0.556
C4_U	22	11	0.500
N3_U	22	11	0.500
O5'_G	47	16	0.340
N1_U	22	7	0.318
C8_G	48	15	0.312
P_G	47	10	0.213
O4'_U	22	3	0.136
C2_G	48	5	0.104
C2_U	22	2	0.091
C6_G	48	2	0.042
C4_G	48	2	0.042
N7_G	48	1	0.021
N1_G	48	1	0.021

Table 4: Hydration index by atom type: for each atom type in structures P1A, P1B and P2C the number of occurrences of atoms of that type, the number of potential water contacts that have an atom of that type as target and the ratio of contacts to occurrences is shown. Potential contacts to modeled hydrogen atoms are credited to the atom to which the hydrogen is attached. Atoms with alternate conformations are excluded from the list of potential targets.

Group/Atom	Occurrence:Contacts	O	Ca	Sr
backbone	372:412(1.11)	375 3.07(0.34)	32 2.76(0.54)	5 2.78(0.43)
C5'	69 :43(0.62)	43 3.44(0.10)	- -	- -
O3'	67 :79(1.18)	67 3.28(0.24)	9 2.66(0.48)	3 2.85(0.59)
O5'	68 :36(0.53)	36 2.95(0.33)	- -	- -
OP1	56 :120(2.14)	109 2.88(0.26)	9 2.46(0.16)	2 2.66(0.04)
OP2	56 :111(1.98)	105 2.97(0.33)	6 2.32(0.07)	- -
P	56 :23(0.41)	15 3.48(0.09)	8 3.54(0.07)	- -
base	608:168(0.28)	163 3.21(0.30)	3 2.80(0.63)	2 2.50(0.00)
C2_G	48 :5(0.10)	5 3.55(0.03)	- -	- -
C2_U	22 :2(0.09)	2 3.35(0.13)	- -	- -
C4_G	48 :2(0.04)	2 3.52(0.09)	- -	- -
C4_U	22 :13(0.59)	12 3.44(0.21)	1 3.52(-)	- -
C5_G	48 :1(0.02)	1 3.39(-)	- -	- -
C5_U	22 :15(0.68)	15 3.45(0.12)	- -	- -
C6_G	48 :2(0.04)	2 3.46(0.05)	- -	- -
C6_U	22 :11(0.50)	11 3.35(0.15)	- -	- -
C8_G	48 :2(0.04)	2 3.51(0.00)	- -	- -
N1_G	48 :1(0.02)	1 3.46(-)	- -	- -
N1_U	22 :7(0.32)	7 3.51(0.07)	- -	- -
N2_G	48 :20(0.42)	20 3.29(0.20)	- -	- -
N3_G	48 :31(0.65)	31 3.08(0.26)	- -	- -
N3_U	22 :2(0.09)	2 3.40(0.25)	- -	- -
N7_G	48 :1(0.02)	1 3.51(-)	- -	- -
O2_U	22 :14(0.64)	14 3.01(0.24)	- -	- -
O4_U	22 :39(1.77)	35 2.95(0.26)	2 2.44(0.06)	2 2.50(0.00)
furanose	411:304(0.74)	287 3.17(0.30)	14 3.21(0.38)	3 3.03(0.47)
C1'	69 :38(0.55)	38 3.36(0.09)	- -	- -
C2'	68 :57(0.84)	52 3.31(0.13)	5 3.59(0.01)	- -
C3'	68 :15(0.22)	11 3.43(0.15)	3 3.33(0.23)	1 3.58(-)
C4'	69 :18(0.26)	18 3.49(0.08)	- -	- -
O2'	68 :134(1.97)	126 2.98(0.30)	6 2.85(0.24)	2 2.76(0.04)
O4'	69 :42(0.61)	42 3.23(0.24)	- -	- -
solvent	221:529(2.39)	473 2.94(0.35)	35 2.52(0.25)	21 2.59(0.10)
CA	3 :7(2.33)	7 2.50(0.19)	- -	- -
O	201:502(2.50)	446 2.97(0.35)	35 2.52(0.25)	21 2.59(0.10)
SR	17 :20(1.18)	20 2.62(0.22)	- -	- -

Table 5: *Solvent contact distances: the distribution of solvent contacts across all structures examined. Each non-bonding contact links a solvent ion/water, with full or partial occupancy, and a target atom. Target atoms are grouped into four classes, only those with full occupancy are considered. The number of occurrences of atoms of that type, the number of contacts to such atoms, and the average number of contacts per atom across all structures are shown in the first column. The three columns on the right show the number of contacts and the average and standard deviation of the contact distance for HOH, Ca, and Sr respectively.*

	α	β	γ	δ	ϵ	ζ	χ
U1			52.8(1.3)	79.2(1.2)	-155.7(3.5)	-89.5(1.7)	-145.2(2.6)
G2	-97.0(1.5)	63.8(3.0)	175.7(1.8)	82.9(1.7)	-153.1(2.4)	-71.7(2.3)	-167.2(1.8)
G3	-69.0(2.2)	-171.3(3.0)	59.5(1.2)	149.7(1.4)	-170.0(1.8)	-117.6(1.4)	-120.8(1.8)
G4	-77.0(1.6)	176.3(1.3)	45.6(1.2)	76.0(0.7)	-151.8(3.5)	-74.2(3.7)	-154.6(1.4)
G5-A	-67.1(3.5)	138.2(105.6)	50.5(1.4)	80.4(3.7)	-158.3(2.9)	59.5(5.5)	-148.7(1.9)
G5-B	-71.4	175.8	50.4	75.9	-152.5	-78.7	-148.8
G5-C	-69.2	174.4	49.0	75.8	-163.5	43.0	-148.9
U6-A	154.8(16.6)	151.8(7.3)	54.7(14.2)	141.8(2.5)			-135.2(4.2)
U6-B	-62.1	171.6	64.1	122.5			-130.8
U6-C	73.0	176.0	52.2	90.6			-159.4

Table 6: *Torsion angles. Torsion angles averaged over twelve chains from structures P1A, P1B and P2C [47]. For residues G5 and U6, angles are computed separately for conformations 3'U-A (ten chains), 3'U-B and 3'U-C (one chain each).*

List of Figures

1	Guanine quartet secondary structure and quartet stacking.	34
2	Octamer assembly for UGGGGU quadruplexes.	35
3	Arrangement of quadruplex columns in tetragonal ($P4_212$) and orthorhombic ($C222_1$) space groups.	36
4	SAXS	37
5	B factor distribution.	38
6	Alternate Conformations	39
7	Cation coordination by Guanine	40
8	xx	41
9	Bound waters in contact with the ribose $O2'$ hydroxyl group in UGGGGU data set P1A.	42

Figure 1: *Guanine quartet secondary structure and quartet stacking, coordinates are from data set P1B.*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 2: *Components of UGGGGU octamer assembly. (a) An isolated RNA hexamer, the 3' uridine that participates in a U-tetrad is shown in blue and the 5' uridine involved in tetramer intercalation is shown in cyan. Two central Sr^{2+} cations involved in G-quartet stabilization are shown in magenta. (b) The tetramer component produced by rotation of an individual strand about a four-fold axis. (c) Two adjacent, stacked, strands related by two-fold symmetry and (d) the resulting octamer. Coordinates correspond to data set P1B.*

Figure 3: *Arrangement of quadruplex columns in (a) tetragonal ($P4_212$) and (b) orthorhombic ($C222_1$) space groups. Stacked octamers are shown with the central cation axis directed towards the reader. The asymmetric unit is shown in lighter colors.*

Figure 4: SAXS analysis. (a) Intensity of buffer-adjusted scattering for UGGGU samples at concentrations of 0.4 (red), 2 (green) and 4 (blue) mg/ml. (b) The corresponding Kratky plots (c) The ab initio model derived by DAMMIF from merged scattering data displayed as a black mesh and fit to the symmetry-expanded octamer from structure P1B. (d) Comparison of the experimental scattering curve, merged over all concentrations and shown in blue, relative to predicted scattering from octameric and tetrameric models derived from structure P1B, shown in red and green.

Figure 5: *Isotropic B factor distribution of octamer assemblies for data set (a) P1A and (b) P1B.*

(a)

(b)

(c)

Figure 6: *Structural variation in UGGGGU conformation. (a) Superposition of 16 RNA strands from four structures over residues 1-5 in 5'-3' order [23] identifies three principal conformations for the 3' uridine: 3'U-A (red), 3'U-B (green) and 3'U-C (magenta). (b) Zeta and alpha torsion angles for the three conformations fall into distinct clusters. (c) Alternate conformation of a phosphate group in UGGGGU structure P1A. Isocontours of a σ_A -weighted 2Fo-Fc density map contoured at 2.0σ are shown in blue, conformation A is shown in magenta and B in cyan.*

Figure 7: *Cation coordination by guanine quartets. (a) Coordination of a Sr^{2+} cation, shown in green, between the eight guanine O6 atoms of two G-quartets, shown in magenta. (b,c) Views of a UGGGGU octamer emphasizing the regular arrangement of O6 atoms about the central cation induced by the structure's helical twist. Coordinates are from the refinement of data set P1B.*

Figure 8: *Hydration by atom type.* (a) The two strands from structure P1A are shown with atom types colored by hydration index value (Supplemental Table 4). Atoms with a hydration score below 0.25 are shown by thinner blue lines. (b) The hydration-based coloring scheme extended to the symmetry-expanded P1A octamer with waters included; the molecular surface is also colored by hydration value.

Figure 9: Waters involved in intra and inter-octamer contacts. (a) The octamer formed from the ASU strands of structure P1A is shown in thick yellow lines and adjacent octamers linked by water contacts are shown in thin blue lines. Waters whose potential contacts are limited to a single octamer are colored magenta, waters which form possible cross-octamer contacts are colored teal. (b) A network of hydrogen bonds involving two waters and four O2' residues stabilizes cross-strand linkage in structure P1B.