Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

A regular thymine tetrad and a peculiar supramolecular assembly in the first crystal structure of an all-LNA G-quadruplex

Irene Russo Krauss,[a] Gary Nigel Parkinson,[b] Antonello Merlino,[a,c] Carlo Andrea Mattia,[d] Antonio Randazzo,[e] Ettore Novellino,[e] Lelio Mazzarella[a,c] and Filomena Sica[a,c,f]*

[a]Department of Chemical Sciences, University of Naples 'Federico II', Complesso Universitario di Monte Sant'Angelo, Via Cinthia, I-80126 Napoli, Italy, [b]Department of Pharmaceutical and Biological Chemistry, UCL School of Pharmacy, University College London, 29–39 Brunswick Square, London WC1N 1AX, England, [c]Institute of Biostructure and Bioimages, CNR, Via Mezzocannone 16, I-80134 Napoli, Italy, [d]Department of Pharmacy, University of Salerno, Via Ponte Don Melillo, I-84084 Fisciano, Italy, [e]Department of Pharmacy, University of Naples 'Federico II', Via D. Montesano 49, I-80131 Napoli, Italy, and [f]National Institute Biostructures and Biosystems, Inter-University Consortium, Viale Medaglie d'Oro 305, I-00136 Roma, Italy

Correspondence e-mail: filosica@unina.it

Received 22 July 2013
Accepted 13 October 2013

PDB reference: all-LNA quadruplex, 4l0a

Locked nucleic acids (LNAs) are formed by bicyclic ribonucleotides where the O2′ and C4′ atoms are linked through a methylene bridge and the sugar is blocked in a 3′-*endo* conformation. They represent a promising tool for therapeutic and diagnostic applications and are characterized by higher thermal stability and nuclease resistance with respect to their natural counterparts. However, structural descriptions of LNA-containing quadruplexes are rather limited, since few NMR models have been reported in the literature. Here, the first crystallographically derived model of an all-LNA-substituted quadruplex-forming sequence 5′-TGGGT-3′ is presented refined at 1.7 Å resolution. This high-resolution crystallographic analysis reveals a regular parallel G-quadruplex arrangement terminating in a well defined thymine tetrad at the 3′-end. The detailed picture of the hydration pattern reveals LNA-specific features in the solvent distribution. Interestingly, two closely packed quadruplexes are present in the asymmetric unit. They face one another with their 3′-ends giving rise to a compact higher-order structure. This new assembly suggests a possible way in which sequential quadruplexes can be disposed in the crowded cell environment. Furthermore, as the formation of ordered structures by molecular self-assembly is an effective strategy to obtain nanostructures, this study could open the way to the design of a new class of LNA-based building blocks for nanotechnology.

1. Introduction

Locked nucleic acids (LNAs) are some of the most prominent and successful nucleic acid analogues available; they have been widely exploited and are able to introduce specific functions and properties into nucleic acids. LNAs achieve their success by the incorporation of a bicyclic ribonucleotide analogue, where the O2′ and C4′ atoms are linked through a methylene bridge (Supplementary Fig. S1[1]), blocking the sugar to a single accessible C3′-*endo* conformation (Obika *et al.*, 1998; Koshkin *et al.*, 1998; Kumar *et al.*, 1998) and thus promoting an RNA-like conformation. Locked nucleic acids are characterized by a higher thermal stability than their unmodified counterparts; in particular, they are able to confer increasing stability to DNA and RNA duplexes and triplexes, even when only one modified nucleotide is inserted in a hybrid sequence (Braasch & Corey, 2001; Obika *et al.*, 1998; Kurreck *et al.*, 2002; Bondensgaard *et al.*, 2000). Moreover, they are very stable in biological systems, since they are not recognized and digested by nucleases (Doessing & Vester, 2011), are soluble in water (unlike, for instance, peptide nucleic acids),

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: BE5244).

are nontoxic and are easy to synthesize (Kaur *et al.*, 2007; Briones & Moreno, 2012). Additionally, the selective incorporation of the functional group may confer novel physical and chemical properties. Their utility is further extended as dsLNA/DNA and dsLNA/RNA heterodimers have a substantially increased thermal stability compared with traditional dsDNA or dsRNA homodimers. This is made all the more remarkable as the destabilizing effect of base mismatches in LNA-containing heterodimers is much higher than in double-stranded DNA or RNA (Braasch *et al.*, 2002; Kurreck *et al.*, 2002; Kaur *et al.*, 2007). Thus, the incorporation of LNA in different nucleic acid-based therapeutic strategies can take advantage of many of these properties, as extensively reported in antigene (Brunet *et al.*, 2005; Beane *et al.*, 2007, 2008; Zaghloul *et al.*, 2011) and antisense (Jepsen & Wengel, 2004; Kurreck *et al.*, 2002; Straarup *et al.*, 2010; Gupta *et al.*, 2010; Lanford *et al.*, 2010) strategies. LNA modifications have also been incorporated to improve the properties of aptamers (Schmidt *et al.*, 2004; Campbell & Wengel, 2011; Förster *et al.*, 2012; Kanwar *et al.*, 2011), DNAzyme for RNA targeting (Vester *et al.*, 2006; Kaur *et al.*, 2010; Donini *et al.*, 2007; Jakobsen *et al.*, 2007), small interfering RNAs (Sun *et al.*, 2011; Fluiter *et al.*, 2009; Mook *et al.*, 2007) and decoy oligonucleotides (Crinelli *et al.*, 2002; Cogoi *et al.*, 2013), and used effectively for diagnostic purposes (Briones & Moreno, 2012; Kauppinen *et al.*, 2006).

Besides duplex-forming sequences, LNAs can also be effectively incorporated into four-stranded G-rich topologies termed G-quadruplexes. The use of G-quadruplexes as building blocks in nonconventional materials for application in molecular electronics and optoelectronics is a rapidly emerging field. Nucleic acids represent building blocks that are highly functionalized biopolymers that can be assembled nocovalently in various ways, such as through the formation of G-tetrads to build specific complex three-dimensional topologies. Long G-quadruplexes can be used as conducting molecular wires (Miyoshi & Sugimoto, 2011) or as a synthetic transmembrane ion transporter (Kauppinen *et al.*, 2006). Moreover, they can also be used as molecular switches that are sensitive to the addition of ions or small ligands or to pH changes, and as intracellular devices (Vester *et al.*, 2006). Convincing evidence of the existence and the role that putative G-quadruplex-forming sequences can play in modulating biological functions in cells has been rapidly accumulating, extending from ssDNA in telomeres to dsDNA in promoter regions and within mRNA sequences located in both exons and introns (Bugaut & Balasubramanian, 2012; Lam *et al.*, 2013). The stability of distinctive topologies formed from specific DNA and RNA G-rich sequences makes them valuable therapeutic targets, so the effect of LNA incorporation in G-quadruplex-forming sequences to enhance certain desired properties is now being actively evaluated. NMR studies have demonstrated that fully modified LNA oligonucleotides can arrange in G-quadruplex structures as much as DNA strands containing only a few modified residues (Nielsen *et al.*, 2006; Randazzo *et al.*, 2004). Within quadruplexes the rigid bicyclic structure of LNA restrains the base towards an *anti* conformation, allowing single LNA substitutions to have a position-dependent effect on the stability of G-quadruplexes. Structural studies of LNA-substituted thrombin-binding aptamers have clearly demonstrated that a single LNA substitution can either destroy or stabilize an antiparallel chair-like G-quadruplex (Virno *et al.*, 2007; Bonifacio *et al.*, 2008), depending on its position in the sequence. Similarly, LNA incorporation has been rationally used to selectively promote different topologies of the human telomeric sequence (Pradhan *et al.*, 2011). However, it is still apparent that structural studies of LNA containing G-quadruplexes are limited. In particular, a detailed description of the features that are important for the stability of these molecules, such as ion coordination and hydration in the groove region, is still lacking.

Here, we present the crystal structure at 1.7 Å resolution of the G-quadruplex formed by the LNA sequence 5′-TGGGT-3′. Structural details derived from this first high-resolution crystallographic model of an all-LNA quadruplex add important information to our knowledge of the effect of modified nucleotides in quadruplex structure and stability. We observe a regular terminal thymine tetrad, the first to be observed in a crystallographic structure of a G-quadruplex, which is a likely contributor to the high thermal stability of this molecule. Further stabilization of the LNA quadruplex comes from the distinctive and regular hydration pattern that also involves the O2′ atom. Moreover, the unusual stacked arrangement of the two quadruplexes in the asymmetric unit suggests how sequential quadruplexes can arrange in the crowded cell environment, while at the same time the structure provides an opportunity for the design of LNA-based blocks for developing scaffolds with nanotechnology applications.

2. Experimental procedures

2.1. Sample preparation and crystallization

The LNA oligonucleotide 5′-TGGGT-3′ was obtained and purified as described elsewhere (Randazzo *et al.*, 2004). Solutions of the 5-mer at a concentration of about 15 m*M* in 10 m*M* KCl, 10 m*M* potassium phosphate buffer pH 7.0 were annealed by heating at 363 K and slow cooling to 293 K in order to induce formation of the G-quadruplex. The correct folding of the oligonucleotide was verified by circular-dichroism measurements: the presence of a strong positive peak at about 260 nm is characteristic of parallel-stranded G-quadruplexes and was used as evidence of quadruplex formation.

An extensive screening of crystallization conditions was performed using commercial kits such as the Natrix and Nucleic Acid Mini Screen kits from Hampton Research. Crystallization trials were set up at two different temperatures, 277 and 293 K, and using different oligonucleotide concentrations in the range 5–15 m*M* (Russo Krauss *et al.*, 2013). Bushes of thin needles were obtained using 2.0 *M* ammonium sulfate, 10 m*M* magnesium sulfate, 50 m*M* sodium cacodylate pH 6.5 as precipitant solution and a 15 m*M* oligonucleotide

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data-collection statistics	
Space group	$P6_5$
Unit-cell parameters (Å)	
$a = b$	27.670
c	196.830
Resolution limits (Å)	22.00–1.70 (1.73–1.70)
No. of observations	22015
No. of unique reflections	8969
Completeness (%)	96.1 (74.3)
$\langle I/\sigma(I)\rangle$	11.9 (7.7)
Average multiplicity	2.5
$R_{\rm merge}$† (%)	5.3 (6.1)
Mosaicity (°)	0.8
$V_{\rm M}$ (Å^3 Da^{-1})	1.57
Solvent content (%)	47.0
Refinement	
Resolution limits (Å)	22.00–1.70
No. of reflections used in refinement [$F > 0\sigma(F)$]	8945
No. of reflections in working set	8473
No. of reflections in test set	472
R factor/$R_{\rm free}$‡	0.161/0.184
No. of oligonucleotide atoms	904
No. of glycerol atoms	18
No. of water molecules	236
No. of ions	6
R.m.s.d. from ideal values	
Bond lengths (Å)	0.02
Bond angles (°)	3.7
Average B factors (Å^2)	
Oligonucleotide	7.8
Ions	21.9
Solvent	20.5
Glycerol atoms	28.4

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl))$, where $I_i(hkl)$ is the ith intensity measurement of the reflection hkl, including symmetry-related reflections, and $\langle I_i(hkl)\rangle$ is its average. ‡ R factor = $\sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ represent the observed and calculated structure factors, respectively. The R factor was calculated using 95% of the data, which were included in refinement, and $R_{\rm free}$ was calculated using 5% of the data, which were excluded from refinement.

concentration. Optimization of the initial conditions resulted in the growth of well formed diffracting crystals. In particular, the best crystals (dimensions of 0.1 × 0.2 × 0.4 mm) grew after several months from drops consisting of 1 µl LNA solution (9 m*M* in 10 m*M* KCl, 10 m*M* potassium phosphate buffer pH 7.0) and 1 µl reservoir solution (1.7 *M* ammonium sulfate, 10 m*M* magnesium sulfate, 50 m*M* sodium cacodylate pH 6.5) that were equilibrated against 500 µl reservoir solution.

2.2. Data collection and structure determination

Diffraction data were collected in-house on a Saturn 944 CCD detector to 1.7 Å resolution. The X-ray radiation used was Cu $K\alpha$ radiation from a Rigaku MicroMax-007 HF generator. After the addition of 20% glycerol to the harvesting solution, the crystals were flash-cooled at 100 K in supercooled N_2 gas (Oxford Cryosystems) and maintained at 100 K during data collection. Diffraction data were processed using *iMosflm* (Battye *et al.*, 2011) and *SCALA* (Evans, 2006) from the *CCP*4 package (Winn *et al.*, 2011). The initial automated data reduction suggested the hexagonal space group $P6_n22$ (where n = 1 or 5) to be the most likely space group, with unit-cell parameters $a = b = 27.67$, $c = 196.83$ Å, $\alpha = \beta = 90.0$, $\gamma = 120.0°$. However, careful reprocessing using *AIMLESS* (Evans & Murshudov, 2013) suggested an alternative lower symmetry Laue group, $P6_n$. In particular, the $R_{\rm merge}$ calculated on all reflections was significantly lower (9% for $P6_n$ compared with 16% for $P6_n22$). Then, by calculating the agreement among reflections related by the additional symmetry operator of the $P6_n22$ space group, we were finally able to definitively exclude $P6_n22$ as the correct space group [$R_{\rm meas}$ calculated with *POINTLESS* (Evans & Murshudov, 2013) was 35%]. Moreover, when the data were scaled in space group $P6_n$ the program also detected possible twinning, but the twinning fraction did not indicate perfect twinning, which would be expected in the case that the space group was indeed $P6_n22$. $P6_n$ data were then successfully used for automatic molecular replacement with *Phaser* (McCoy *et al.*, 2005). Several search models were used, and eventually a good solution with a log-likelihood gain of 92 was obtained using a model derived from the NMR structure of the LNA quadruplex (PDB entry 1s9l; Randazzo *et al.*, 2004). In particular, the model was formed by the three central guanines of the four strands of the first NMR model (*i.e.* the most representative conformer of the 20-structure ensemble). Since Matthews coefficient calculation (Kantardjieff & Rupp, 2003) suggested the presence of two highly packed quadruplexes in the asymmetric unit, two quadruplexes per asymmetric unit were searched for with 90% sequence identity with respect to the model. Molecular replacement also allowed unambiguous identification of space group $P6_5$ as the correct space group and excluded any twinning. Final statistics of the data collection are provided in Table 1. Post-refinement analysis then revealed that the two molecules in the asymmetric unit are related by an approximate twofold symmetry. In particular, they are related by a 180° rotation around an axis that is rotated approximately 81° with respect to the c axis and is oriented about 15° off the ab diagonal. Therefore, the additional symmetries of the diffraction pattern are owing to the intrinsic symmetry of the structure rather than to twinning or higher crystal symmetry.

2.3. Refinement and structural analysis

The starting model was refined using *CNS* (Brünger *et al.*, 1998) and *REFMAC*5 (Murshudov *et al.*, 2011). Each run was alternated with manual model building using *Coot* (Emsley & Cowtan, 2004). Analysis of electron-density maps calculated with ($F_{\rm o} - F_{\rm c}$) and ($2F_{\rm o} - F_{\rm c}$) coefficients allowed the rebuilding of missing residues, eight thymines for each G-quadruplex, and the positioning of three potassium ions in each quadruplex channel, of several water molecules and three glycerol molecules from the cryoprotecting solution. The final model has R-factor and $R_{\rm free}$ values of 0.161 and 0.184, respectively. Statistics and parameters of the refinement are given in Table 1.

The figures were prepared with *PyMOL* (http://pymol.org).

The coordinates of the structure have been deposited in the Protein Data Bank (entry 4l0a).

w3DNA (Zheng *et al.*, 2009) and *Curves+* (Blanchet *et al.*, 2011) were used to calculate local and overall geometric parameters. The *SUPERPOSE* program from the *CCP*4 package was used to calculate root-mean-square deviations (Krissinel & Henrick, 2004). The geometry of the tetrads was analysed using a home-made program that calculates the least-squares plane of the two tetrads, the root-mean-square deviation of residue atoms from the best planes, the angle that each residue makes with the best plane of the tetrad to which it belongs and the angle between the two tetrad planes and their separation along the chain axis (Russo Krauss *et al.*, 2011).

3. Results

3.1. Overall structure and crystal packing

The structure was solved and refined at 1.7 Å resolution. Detailed statistics of the refinement are reported in Table 1.

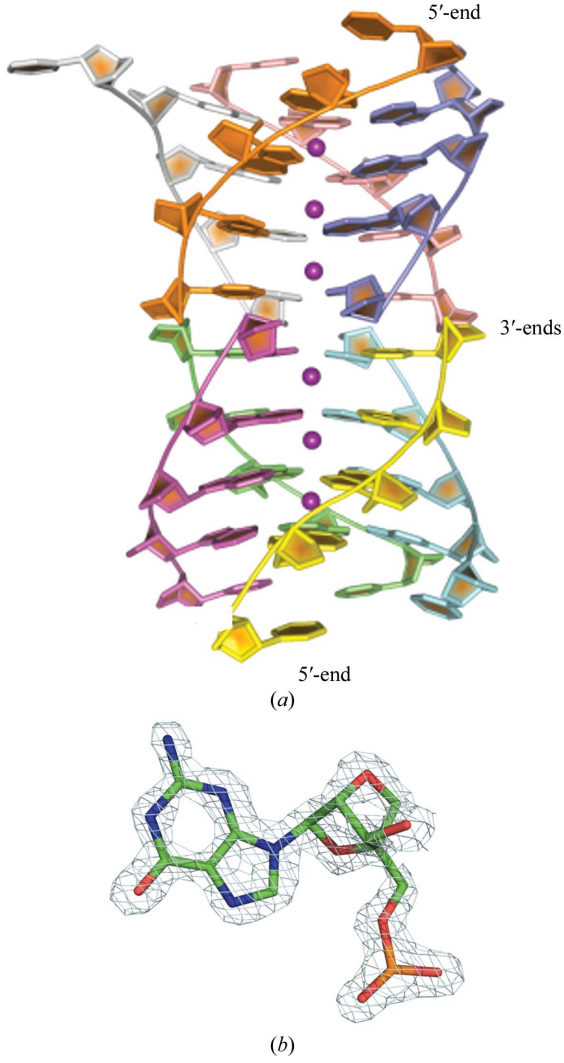

Figure 1
(*a*) Cartoon representation of the two LNA quadruplexes in the asymmetric unit. Different chains are marked in different colours. Potassium ions are also shown as spheres. (*b*) The $2F_o - F_c$ electron-density map of a guanine residue contoured at the 2.0σ level is shown as an example of the high quality of the electron-density maps.

The oligonucleotide is organized in a four-stranded parallel G-quadruplex formed by a three-G-tetrad core and one thymine tetrad (T-tetrad) at the 3′-end. The quadruplex axis is only slightly offset with respect to the crystallographic screw axis. The two quadruplexes in the asymmetric unit (composed of strands *A*, *B*, *C* and *D* and of strands *E*, *F*, *G* and *H*, respectively) are related by a local twofold symmetry approximately orthogonal to the hexagonal *c* axis and stack on top of each other through the T-tetrads at the 3′-end (Fig. 1*a*). With the exclusion of thymine residues at the 5′-terminus, the r.m.s. deviation between the two quadruplexes calculated on all atoms is 0.5 Å; it rises to 4.4 Å when these residues are included in the calculation. All bases adopt an *anti* conformation and are well defined in electron-density maps (Fig. 1*b*), the only exception being Thy D1, which is partially disordered (Supplementary Fig. S2*a*).

Three K^+ ions per quadruplex are sandwiched between the tetrads and are almost equidistant from one another (the distances are in the range 3.5–3.7 Å; Fig. 2). Those placed between G-tetrads show the distorted antiprismatic coordination geometry found in other quadruplex structures (Russo Krauss *et al.*, 2011, 2012; Haider *et al.*, 2002), whereas an approximately cubic coordination geometry is adopted by the potassium ions placed between the T-tetrads and G-tetrads (see below). The six cations in the asymmetric unit are perfectly aligned along the quadruplex axes. Thymines located at the 5′-end adopt very different conformations in the two quadruplexes, with the only exceptions being Thy *A*1 and the corresponding Thy *E*1, whose position is fixed in both cases by a hydrogen bond between O5′ and N2 of Gua *B*2 and Gua *F*2, respectively (Supplementary Fig. S2*b*). These thymine

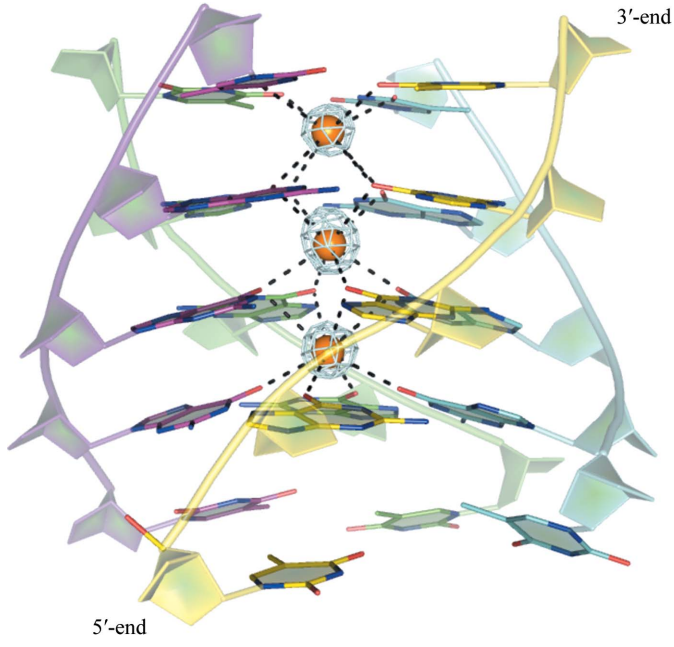

Figure 2
Position and coordination of potassium ions within the LNA G-quadruplex. An $F_o - F_c$ OMIT electron-density map of the ions contoured at the 5.0σ level is also shown. The dotted lines mark the coordination of the ions.

Table 2
Helical parameters for crystallographic and NMR models of similar LNA, DNA and RNA quadruplexes.

	LNA TGGGT X-ray (4l0a)	LNA TGGGT NMR (1s9l)	DNA TGGGGT X-ray (1s45)	DNA (T)TGGGGT NMR (139d)	RNA UGGGGU X-ray (1j8g)	RNA UGGGGU NMR (1rau)
Rise (Å)						
2–3	3.3	3.3	3.0	3.0	2.8	3.2
3–4	3.1	3.4	3.5	3.3	3.7	4.4
4–5	3.4†	3.2†	3.4	2.8	2.8	4.1
Mean	3.2 (3.2)‡	3.4 (3.3)‡	3.3	3.0	3.1	3.9
Twist (°)						
2–3	29.0	28.6	28.0	30.1	27.5	26.5
3–4	29.3	27.8	32.3	36.9	36.6	35.7
4–5	19.5†	23.3†	25.0	23.6	24.7	27.4
Mean	29.2 (25.9)‡	28.2 (25.6)‡	28.4	30.2	29.6	29.9
Groove width§ (Å)						
A–*B*	2.8	3.1	1.5	0.7	3.2	4.2
A–*C*	3.4	3.2	1.6	0.7	3.2	4.3
B–*D*	2.8	3.2	1.6	0.7	3.2	4.1
B–*C*	4.7	3.2	1.1	0.7	3.2	4.3
Mean	3.4	3.2	1.5	0.7	3.2	4.2

† These values refers to the G-tetrad/T-tetrad steps. ‡ These values are obtained by also including the G-tetrad/T-tetrad steps in the averaging. § Groove widths were calculated as the shortest distance between phosphates across the groove minus the sum of their van der Waals radii (5.8 Å).

residues are extensively involved in packing interactions with symmetry-related molecules (Supplementary Table S1). In particular, a stacking interaction between Gua *F*2 and Thy *B*1* connects symmetry-related quadruplexes along the crystallographic screw axis, defining, together with the interaction between Thy *B*5 and Thy *F*5 at the 3′–3′ interface of the two quadruplexes in the asymmetric unit, a virtual continuous strand which winds around the *c* axis and is actually composed of strands *B* and *F* (Supplementary Figs. S3*a* and S3*b*). Another important interaction between quadruplexes along the crystallographic screw axis involves Thy *E*1 and Thy *C*1* (Supplementary Fig. S3*c*), whereas the main contact between symmetry mates within layers of quadruplexes involves Thy *F*1 (Supplementary Fig. S3*d*). Water molecules also extensively contribute to packing contacts within the layers through a complex network of intermolecular interactions.

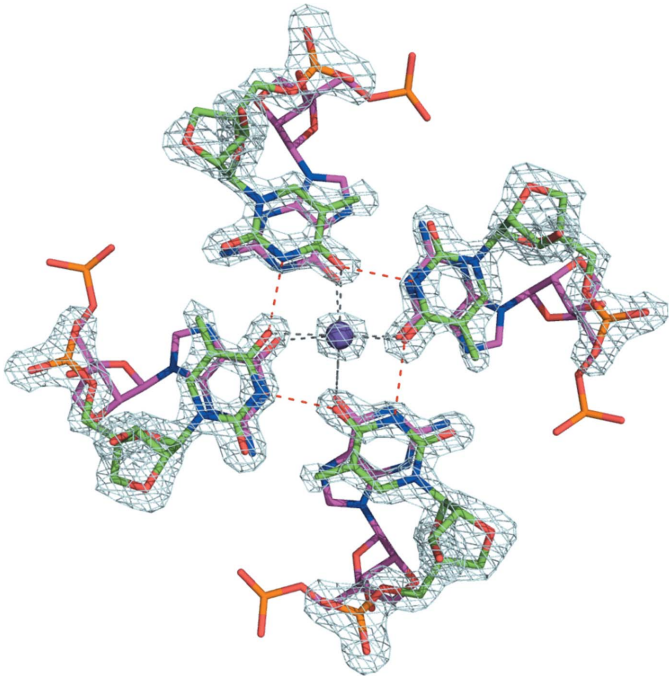

Figure 3
$2F_o - F_c$ electron density, contoured at the 1.5σ level, of the T-tetrad and the potassium ion sandwiched between the T-tetrads and G-tetrads. Interactions between O4 and N3 atoms within the T-tetrad (red dashed lines) are shown in order to mark the regularity of the motif. Interactions between K^+ and the O6 and O4 atoms (grey dashed lines) are also shown to underline the cubic geometry.

3.2. Thymine tetrad

The thymine tetrad observed in the LNA G-quadruplex structure is found at the 3′-end of the sequence. It is stabilized by cyclic O4–N3 hydrogen bonds and by a perfect stacking of the bases on the six-membered rings of the preceding G-quartet (Fig. 3), as indicated by the very low value of the angle between the two tetrad planes (0.2 and 0.8° for the two quadruplexes in the asymmetric unit) and the 3.6 Å separation between the two tetrads. The T-tetrad geometry is very regular: all O4–N3 distances are about 3.0 Å (Supplementary Fig. S4*a*) and the root-mean-square deviations of Thy atoms with respect to the average tetrad plane are very low (0.18 and 0.19 Å for the two quadruplexes in the asymmetric unit, respectively), as are the angles between the thymine planes and the average tetrad plane (which are in the range 4.5–10.1°). The T-tetrad is stabilized by the interaction with a potassium ion that bridges the O4 atoms of four thymines to the O6 atoms of four guanines. The coordination geometry of the cation is different from the distorted antiprismatic coordination observed when it is stacked between two G-tetrads: owing to the relative rotation between the G-tetrad and the T-tetrad (see the twist value reported in Table 2) the coordination geometry becomes approximately cubic, with angles between the coordinating O atoms in the range 83–95° and angles between these and the K^+ ion in the range 64–67° (Fig. 3). Thymine/uracil tetrads at the ends of G-quadruplexes have been reported in previous crystal structures: in particular at the 5′-end of a $(TGGGGT)_4$ DNA quadruplex crystallized in the presence of thallium and sodium ions (Cáceres *et al.*, 2004) and at the 3′-end of the $(UGGGGU)_4$ RNA quadruplex (Deng *et al.*, 2001). However, the tetrad geometry is different in the three cases (Supplementary Fig. S4). In the DNA quadruplex thymines are actually paired and no cyclic hydrogen-bond pattern is observed (Supplementary Fig. S4*b*). In the RNA structure the four uracil residues are connected through a cyclic hydrogen-bond pattern, but the U-tetrad plane is quite distorted (the r.m.s.d. of the uracil atoms with respect to the average tetrad plane is 0.38 Å) and the four bases are tilted (the angle

between the uracil planes and the average tetrad plane is 18.2°) in an almost pyramidal arrangement (Supplementary Fig. S4*c*). A NMR study has also shown and discussed the formation of an U-tetrad at the 3′-end of the RNA telomeric sequence $(UAGGGU)_4$ (Xu *et al.*, 2010*a*). However, detailed structural features of this motif, such as the O4–N3 distance or cation binding, are missing. The only example of a highly regular T-tetrad comparable to the one observed in this crystal structure is the tetrad formed by inner thymines in the NMR model of the DNA telomeric sequence of *Saccharomyces cerevisiae* $(TGGTGGC)_4$ (Patel & Hosur, 1999).

3.3. Quadruplex geometry and LNA conformation

The helical parameters of the LNA G-quadruplex were calculated using *w3DNA* (Zheng *et al.*, 2009). For comparison, the same parameters were also calculated for the NMR model of the same quadruplex (Randazzo *et al.*, 2004) and for two other pairs of NMR–crystallographic structures, one of a DNA G-quadruplex and the other of an RNA quadruplex, whose sequences are similar to that of LNA. As all of the molecules are symmetric and the geometry of the G-tetrads is fixed by the planar hydrogen-bond arrangement, just two parameters are needed to compare the structures: rise and twist. In addition, we calculated the groove widths as the shortest distance between phosphates across the groove minus the sum of their van der Waals radii (5.8 Å). The results are reported in Table 2. The rise values are comparable in the crystallographic and NMR structures of the LNA G-quadruplex, ranging from 3.1 to 3.4 Å, with mean values of 3.2 and 3.4 Å for the crystal and NMR model, respectively. The rise values are also similar in the case of the DNA quadruplex, irrespective of the experimental method used, whereas in the case of the RNA quadruplex a significant difference is observed (3.1 *versus* 3.9 Å for the X-ray and NMR models, respectively). For the twist angles, the mean values are very similar for all of the structures. The only differences are observed for the G-tetrad/T-tetrad step in both models of the LNA quadruplex. In this region the quadruplex is slightly unwound compared with the other structures, as judged by twist values of 19.5 and 23.3° for the X-ray and NMR models, respectively. Finally, significant differences in the mean groove widths of crystallographic and NMR structures of DNA and RNA quadruplexes are found, whereas they are the same in the case of the LNA quadruplex. When the T-tetrad at the 3′-end is included in calculations, the two LNA structures are even more similar to each other.

Backbone conformations for all of the previously described structures were calculated using *Curves+* (Blanchet *et al.*, 2011). Backbone angles classified as *gauche*(+) (g+), *gauche*(−) (g−) and *trans* (t) are reported in Supplementary Table S2. From this analysis it clearly emerges that almost all of the LNA residues in both the crystal and the NMR structures have the conformation α = g−, β = t, γ = g+, δ = g+, ε = t, ζ = g−, which corresponds to an A-form RNA helix. Higher variability is observed in the case of the DNA and RNA crystal structures, although the dominant conformations are B-type (α = g−, β = t, γ = g+, δ = t, ε = t, ζ = t) and A-type, respectively. A small variability between RNA residues is also found in the NMR models, whereas the corresponding DNA structures are highly restrained to the B-type conformation.

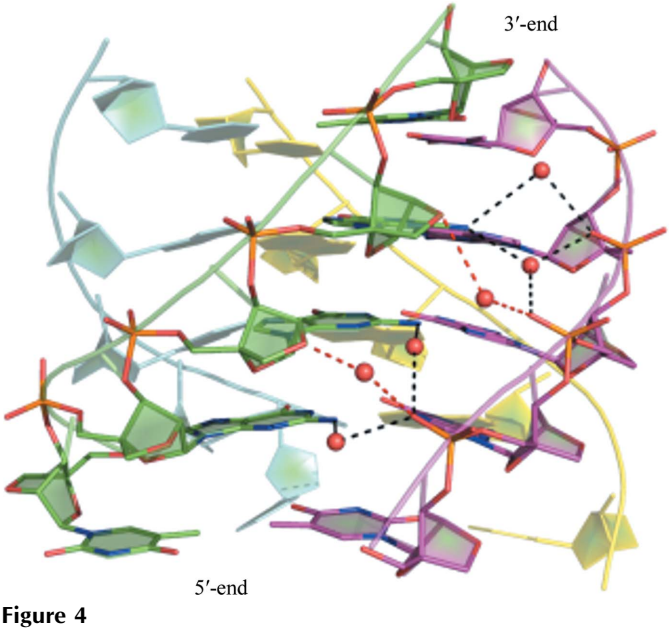

Figure 4
An example of the hydration pattern in one of the LNA G-quadruplex grooves. Water molecules (W) are represented by red spheres. O2′–W–O2P and N2–W–O2P networks are marked in red and black, respectively.

3.4. Hydration

The high resolution of the diffraction data allowed the placement of numerous water molecules. A total of 239 water molecules were located in the asymmetric unit of the crystal, corresponding to an average of about six water molecules per residue. In each groove the O2′–C4′-methylene bridge of one strand faces the phosphate group of the neighbouring strand. Their interaction is mediated by well defined water molecules that create a continuous network of hydrogen bonds (with distances ranging from 2.6 to 3.4 Å) linking atoms of both the backbone and the bases. This hydration pattern is very regular, unlike that found in other DNA and RNA quadruplex structures. The grooves have almost the same width from Gua 2 to Thy 5, and often a single water molecule bridges adjacent strands (Fig. 4). In particular, the most common networks of hydrogen bonds connecting two adjacent strands are N2–water molecule–O2P (O1P in only a few cases) and O2′–water molecule–O2P (O1P in only a few cases) (Fig. 4). At the 5′-end the grooves are wider and more water molecules are involved in connecting different strands. Moreover, in this region water molecules are critically involved in the stabilization of thymine conformations and packing interactions. The preferential hydration of the N2 and N3 atoms of guanines previously observed in other G-quadruplex structures (Phillips *et al.*, 1997; Russo Krauss *et al.*, 2011) is also confirmed in this all-LNA quadruplex. In contrast, in this case the well characterized hydrogen bonds between O4′ and water molecules (Haider *et al.*, 2002; Phillips *et al.*, 1997) play a marginal role in the hydration of grooves: only a few thymine residues and one

guanine residue show this kind of interaction. On the other hand, almost all O2′ atoms are involved in hydrogen bonds to water molecules, underlining the important role of the ribose modification in the hydration of the molecule. It should be stressed that in contrast to what was observed in the RNA quadruplex (Collie, Haider *et al.*, 2010), the highly hydrated O2′ atoms of the LNA moiety do not make intramolecular interactions.

3.5. Interface

The two quadruplexes in the asymmetric unit are closely packed, as indicated by the very low V_M value (1.57 Å^3 Da^{-1}). They face each other with their 3′ T-tetrads. Applying the stacking classification recently used by Lech *et al.* (2013) to classify the stacking between intramolecular or intermolecular G-tetrads, the interface between two neighbouring T-tetrads is a tail-to-tail opposite-polarity stacking. The distance between the two intermolecular T-tetrads is almost the same as that between tetrads belonging to the same quadruplex (3.1–3.4 Å) and the two quadruplexes could also be seen as a whole object, apart from the opposite polarity of the strands. An interesting pattern of interactions connecting the backbones is observed at the interface between the two quadruplexes. In particular, two different interfaces can be identified: one between the strand pairs *A*–*F*, *B*–*E*, *C*–*H* and *D*–*G* and the other between strand pairs *A*–*H*, *B*–*F*, *C*–*G* and *D*–*E*. In the former we find a symmetric O3′–O2P interaction and in the latter an O2′–water molecule–O2′ network (Fig. 5), which once again highlights the role of the modified ribose ring. Altogether, these contacts furnish a strong link between the two quadruplexes.

4. Discussion

Here, we present the first crystallographic structure of a locked nucleic acid G-quadruplex, the all-LNA 5′-TGGGT-3′ sequence, refined at 1.7 Å resolution. The oligonucleotide adopts a solid architecture in which two parallel G-quadruplexes strongly interact, giving rise to a compact higher order structure. The most interesting structural feature that we find in the LNA crystal structure is the presence of a very regular T-tetrad at the 3′-end of the molecule, which can be considered to belong to the core of the quadruplex. In fact, this motif does not perturb the geometry of the quadruplex architecture; in contrast, it is very well integrated into it, as indicated by the rise and twist values and by coordination of the potassium ion sandwiched between the T-tetrad and one of the G-tetrads. Previously, the presence of the methyl group in thymine with respect to uracil was supposed to prevent the formation of a regular plane in which all bases are hydrogen bonded to each other (Cáceres *et al.*, 2004). However, these new data prove that such an arrangement is also allowed with thymines. The formation of the regular T-tetrad plane could be favoured by the peculiar RNA-like conformation of the LNA sugar. Similar to the case of the $(UAGGGU)_4$ RNA quadruplex (Xu *et al.*, 2010*a*), the T-tetrad forms at the 3′-end and not at the 5′-end. However, it should be noted that in the case of RNA the formation of a T-tetrad at the 5′-edge could be prevented by the presence of adenine residues that inhibit the stacking on the adjacent G-tetrad. Thus, since in the LNA sequence both 5′ and 3′ thymines are directly connected to the G-core, our structure suggests an intrinsic preferential formation of these unusual tetrads at the 3′-end for RNA-like strands. Previous studies of the same LNA quadruplex showed that it has a thermal stability that is 20°C higher than that of the corresponding DNA quadruplex but that rather surprisingly is comparable to that of the corresponding RNA quadruplex (Randazzo *et al.*, 2004). This suggests that structural and thermodynamic factors that contribute to the higher stabilization of duplex and/or triplex structures containing LNA residues with respect to the corresponding natural molecules, *i.e.* the pre-organization of the modified strand resulting in a more favourable initiation free energy for strand association, the presence of the O2′,C4′-methylene bridge limiting the ribose flexibility and affecting the entropy of duplex/triplex formation, and the better stacking of aromatic bases in the resulting structures (Kierzek *et al.*, 2009), are not so important in the case of quadruplexes. The stability of the LNA quadruplex was previously ascribed to entropic effects (Petraccone *et al.*, 2006), but other stabilizing factors that are similar for LNA and RNA quadruplexes should exist. Since an U-tetrad has been shown to be crucial for the stabilization of a telomeric RNA G-quadruplex (Xu *et al.*, 2010*b*), the same role is likely to be played by the T-tetrad in the LNA G-quadruplex. Moreover, a well known factor in the stabilization of G-quadruplexes is hydration, which plays a key structural role through the bridging of residues from neighbouring strands. With respect to DNA, both RNA and LNA have the additional O2′, forming networks of hydrogen bonds with water

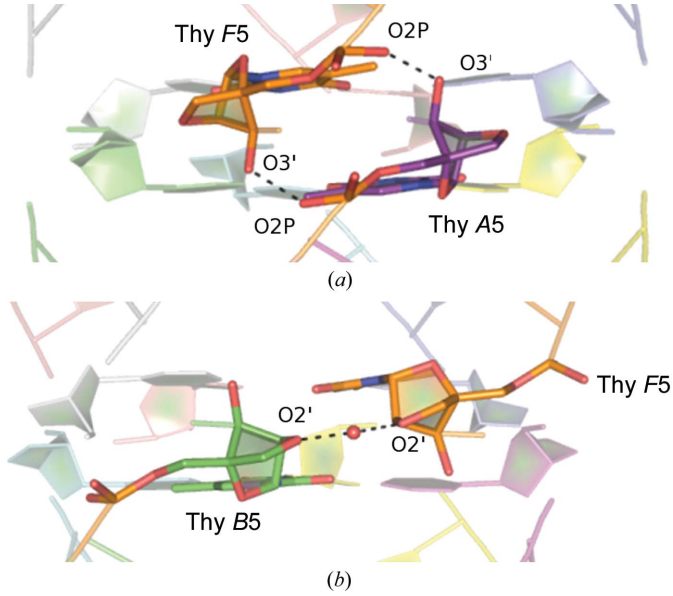

Figure 5
Key backbone interactions between the two quadruplexes in the asymmetric unit: (*a*) the symmetric O3′–O2P interaction between Thy *A*5 and Thy *F*5 and (*b*) the O2′–W–O2′ interaction between Thy *B*5 and Thy *F*5, where W indicates the water molecule represented as a red sphere, are shown as examples.

molecules. We find that the ether O2′ atom of LNA is able to form fewer hydrogen bonds than the hydroxyl group of RNA, but in most cases it forms a regular O2′–water molecule–OP pattern that directly bridges adjacent strands and seems to be peculiar to the LNA moiety.

A comparison of NMR and crystallographic structures of the LNA quadruplex indicates minimal structural differences, thus suggesting that the presence of the locked residue determines a very stable quadruplex architecture that is independent of the experimental conditions (for example, the oligonucleotide concentrations used for crystallography are at least one order of magnitude higher than those used for NMR experiments). This finding is relevant for the future use of LNA-modified sequences as nanodevices or for *in vivo* applications. Since molecular crowding is often observed on the solid surfaces of arrays or sensors (Miyoshi *et al.*, 2002; Miyoshi & Sugimoto, 2008), as well as in cellular compartments, a solid fixed G-quadruplex structure unaffected by crowding conditions is highly desirable.

5. Conclusions

Locked nucleic acids formed by bicyclic ribonucleotides in which the O2′ and C4′ atoms are linked through a methylene bridge have largely been studied as substitutes for natural nucleic acids for therapeutic and diagnostic applications (Doessing & Vester, 2011; Kaur *et al.*, 2007). Despite the great attention towards LNA, few structural studies of LNA-containing quadruplexes have been reported (Randazzo *et al.*, 2004; Virno *et al.*, 2007; Nielsen *et al.*, 2006; Pradhan *et al.*, 2011; Bonifacio *et al.*, 2008). All of these studies were carried out by NMR and have provided only limited information on structural details such as the coordination of ions and, most importantly, the hydration of the molecule. Thus, a clear understanding of the role of the conformational constraint, *i.e.* the presence of the locked moiety, on the physical and structural properties of G-quadruplexes is still lacking. The high-resolution crystal structure and full structural analysis of an all-LNA quadruplex extends our understanding beyond that provided by NMR techniques. First of all, three potassium ions are found bound to the quadruplex, whereas only two would have been predicted on the basis of sequence alone. The additional ion site is principally created by the terminal T-tetrad and is likely to stabilize the quadruplex arrangement, as observed for the RNA quadruplex (Xu *et al.*, 2010*a*). Moreover, the high quality of the diffraction data shown here allows accurate modelling of the hydration structure, which appears to play a determinant role in the stabilization of G-quadruplexes (Miller *et al.*, 2010). Our analysis of the hydration pattern reveals new features characteristic of locked nucleic acids that should provide a significant contribution to the increased stability of LNA G-quadruplexes. In particular, we find that the O2′ atoms of LNA are not involved in any intra-strand interactions, but are mainly involved in hydrogen-bonded interactions to water molecules, which often bridge two adjacent strands.

Finally, it should be recalled that higher order G-quadruplex assemblies have been observed *in vitro* for telomeric DNA and telomeric repeat-containing RNA (TERRA; Collie, Parkinson *et al.*, 2010) and for a number of G-quadruplex-based aptamers. Much recent work on G-quadruplexes is devoted to understanding the way in which these molecular structures can arrange in higher order architectures (Lech *et al.*, 2013; Haider *et al.*, 2008). In this context, the peculiar strong interface formed between the two quadruplexes, which involves the two T-tetrads and polar interactions between backbone atoms, both direct and water-mediated, provides an important contribution. The tail-to-tail interface described here enriches our knowledge of how two quadruplexes can arrange, since only head-to head and head-to-tail interfaces have been observed in crystallographic structures (Cáceres *et al.*, 2004; Collie *et al.*, 2011; Hazel *et al.*, 2006; Lee *et al.*, 2007). Thus, this mode of stacking could represent a possible way in which sequential telomeric quadruplexes arrange. This new interaction mode could also be used to design new molecular entities that are capable of polymerization in supramolecular G-wires. For example, the observed higher order assembly of the two quadruplexes in the asymmetric unit can also be seen as a stable branching point, as usually needed for the design of DNA architectures in nanotechnologies (Endo & Sugiyama, 2009), since the strong 3′–3′ interface between the two quadruplexes leaves the eight 5′-ends free to be functionalized with moieties useful for polymerization. Alternatively, modified oligonucleotides having a 5′–5′ inversion of polarity sites (Galeone *et al.*, 2008) that externally expose only the 3′ edges of the molecules could also favour the polymerization process. All of these will be the subjects of future investigations in our laboratories.

We acknowledge Giosuè Sorrentino and Maurizio Amendola (Institute of Biostructures and Bioimages, CNR, Naples, Italy) for technical assistance. We thank Elda Trimarco for her contribution to the crystallization experiments. The authors acknowledge Associazione Italiana di Cristallografia (AIC) for the Research Fellowship that allowed Dr Irene Russo Krauss to spend two months in the group of Dr Gary N. Parkinson.

References

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Beane, R., Gabillet, S., Montaillier, C., Arar, K. & Corey, D. R. (2008). *Biochemistry*, **47**, 13147–13149.
Beane, R. L., Ram, R., Gabillet, S., Arar, K., Monia, B. P. & Corey, D. R. (2007). *Biochemistry*, **46**, 7572–7580.
Blanchet, C., Pasi, M., Zakrzewska, K. & Lavery, R. (2011). *Nucleic Acids Res.* **39**, W68–W73.
Bondensgaard, K., Petersen, M., Singh, S. K., Rajwanshi, V. K., Kumar, R., Wengel, J. & Jacobsen, J. P. (2000). *Chemistry*, **6**, 2687–2695.
Bonifacio, L., Church, F. C. & Jarstfer, M. B. (2008). *Int. J. Mol. Sci.* **9**, 422–433.
Braasch, D. A. & Corey, D. R. (2001). *Chem. Biol.* **8**, 1–7.
Braasch, D. A., Liu, Y. & Corey, D. R. (2002). *Nucleic Acids Res.* **30**, 5160–5167.

Briones, C. & Moreno, M. (2012). *Anal. Bioanal. Chem.* **402**, 3071–3089.

Brunet, E., Corgnali, M., Perrouault, L., Roig, V., Asseline, U., Sørensen, M. D., Babu, B. R., Wengel, J. & Giovannangeli, C. (2005). *Nucleic Acids Res.* **33**, 4223–4234.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Bugaut, A. & Balasubramanian, S. (2012). *Nucleic Acids Res.* **40**, 4727–4741.

Cáceres, C., Wright, G., Gouyette, C., Parkinson, G. & Subirana, J. A. (2004). *Nucleic Acids Res.* **32**, 1097–1102.

Campbell, M. A. & Wengel, J. (2011). *Chem. Soc. Rev.* **40**, 5680–5689.

Cogoi, S., Zorzet, S., Rapozzi, V., Géci, I., Pedersen, E. B. & Xodo, L. E. (2013). *Nucleic Acids Res.* **41**, 4049–4064.

Collie, G. W., Haider, S. M., Neidle, S. & Parkinson, G. N. (2010). *Nucleic Acids Res.* **38**, 5569–5580.

Collie, G. W., Parkinson, G. N., Neidle, S., Rosu, F., De Pauw, E. & Gabelica, V. (2010). *J. Am. Chem. Soc.* **132**, 9328–9334.

Collie, G. W., Sparapani, S., Parkinson, G. N. & Neidle, S. (2011). *J. Am. Chem. Soc.* **133**, 2721–2728.

Crinelli, R., Bianchi, M., Gentilini, L. & Magnani, M. (2002). *Nucleic Acids Res.* **30**, 2435–2443.

Deng, J., Xiong, Y. & Sundaralingam, M. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 13665–13670.

Doessing, H. & Vester, B. (2011). *Molecules*, **16**, 4511–4526.

Donini, S., Clerici, M., Wengel, J., Vester, B. & Peracchi, A. (2007). *J. Biol. Chem.* **282**, 35510–35518.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Endo, M. & Sugiyama, H. (2009). *Chembiochem*, **10**, 2420–2443.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Fluiter, K., Mook, O. R. & Baas, F. (2009). *Methods Mol. Biol.* **487**, 189–203.

Förster, C., Zydek, M., Rothkegel, M., Wu, Z., Gallin, C., Gessner, R., Lisdat, F. & Fürste, J. P. (2012). *Biochem. Biophys. Res. Commun.* **419**, 60–65.

Galeone, A., Mayol, L., Virgilio, A., Virno, A. & Randazzo, A. (2008). *Mol. Biosys.* **4**, 426–430.

Gupta, N., Fisker, N., Asselin, M. C., Lindholm, M., Rosenbohm, C., Ørum, H., Elmén, J., Seidah, N. G. & Straarup, E. M. (2010). *PLoS One*, **5**, e10682.

Haider, S., Parkinson, G. N. & Neidle, S. (2002). *J. Mol. Biol.* **320**, 189–200.

Haider, S., Parkinson, G. N. & Neidle, S. (2008). *Biophys. J.* **95**, 296–311.

Hazel, P., Parkinson, G. N. & Neidle, S. (2006). *J. Am. Chem. Soc.* **128**, 5480–5487.

Jakobsen, M. R., Haasnoot, J., Wengel, J., Berkhout, B. & Kjems, J. (2007). *Retrovirology*, **4**, 29.

Jepsen, J. S. & Wengel, J. (2004). *Curr. Opin. Drug Discov. Devel.* **7**, 188–194.

Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.

Kanwar, J. R., Roy, K. & Kanwar, R. K. (2011). *Crit. Rev. Biochem. Mol. Biol.* **46**, 459–477.

Kauppinen, S., Vester, B. & Wengel, J. (2006). *Handb. Exp. Pharmacol.* **2006**, 405–422.

Kaur, H., Babu, B. R. & Maiti, S. (2007). *Chem. Rev.* **107**, 4672–4697.

Kaur, H., Scaria, V. & Maiti, S. (2010). *Biochemistry*, **49**, 9449–9456.

Kierzek, E., Pasternak, A., Pasternak, K., Gdaniec, Z., Yildirim, I., Turner, D. H. & Kierzek, R. (2009). *Biochemistry*, **48**, 4377–4387.

Koshkin, A. A., Nielsen, P., Meldgaard, M., Rajwanshi, V. K., Singh, S. K. & Wengel, J. (1998). *J. Am. Chem. Soc.* **120**, 13252–13253.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Kumar, R., Singh, S. K., Koshkin, A. A., Rajwanshi, V. K., Meldgaard, M. & Wengel, J. (1998). *Bioorg. Med. Chem. Lett.* **8**, 2219–2222.

Kurreck, J., Wyszko, E., Gillen, C. & Erdmann, V. A. (2002). *Nucleic Acids Res.* **30**, 1911–1918.

Lam, E. Y., Beraldi, D., Tannahill, D. & Balasubramanian, S. (2013). *Nature Commun.* **4**, 1796.

Lanford, R. E., Hildebrandt-Eriksen, E. S., Petri, A., Persson, R., Lindow, M., Munk, M. E., Kauppinen, S. & Ørum, H. (2010). *Science*, **327**, 198–201.

Lech, C. J., Heddi, B. & Phan, A. T. (2013). *Nucleic Acids Res.* **41**, 2034–2046.

Lee, M. P., Parkinson, G. N., Hazel, P. & Neidle, S. (2007). *J. Am. Chem. Soc.* **129**, 10106–10107.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.

Miller, M. C., Buscaglia, R., Chaires, J. B., Lane, A. N. & Trent, J. O. (2010). *J. Am. Chem. Soc.* **132**, 17105–17107.

Miyoshi, D., Nakao, A. & Sugimoto, N. (2002). *Biochemistry*, **41**, 15017–15024.

Miyoshi, D. & Sugimoto, N. (2008). *Biochimie*, **90**, 1040–1051.

Miyoshi, D. & Sugimoto, N. (2011). *Methods Mol. Biol.* **749**, 93–104.

Mook, O. R., Baas, F., de Wissel, M. B. & Fluiter, K. (2007). *Mol. Cancer Ther.* **6**, 833–843.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nielsen, J. T., Arar, K. & Petersen, M. (2006). *Nucleic Acids Res.* **34**, 2006–2014.

Obika, S., Nanbu, D., Hari, Y., Andoh, J., Morio, K., Doi, T. & Imanishi, T. (1998). *Tetrahedron Lett.* **39**, 5401–5404.

Patel, P. K. & Hosur, R. V. (1999). *Nucleic Acids Res.* **27**, 2457–2464.

Petraccone, L., Erra, E., Randazzo, A. & Giancola, C. (2006). *Biopolymers*, **83**, 584–594.

Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. & Luisi, B. (1997). *J. Mol. Biol.* **273**, 171–182.

Pradhan, D., Hansen, L. H., Vester, B. & Petersen, M. (2011). *Chemistry*, **17**, 2405–2413.

Randazzo, A., Esposito, V., Ohlenschläger, O., Ramachandran, R. & Mayola, L. (2004). *Nucleic Acids Res.* **32**, 3083–3092.

Russo Krauss, I., Merlino, A., Giancola, C., Randazzo, A., Mazzarella, L. & Sica, F. (2011). *Nucleic Acids Res.* **39**, 7858–7867.

Russo Krauss, I., Merlino, A., Randazzo, A., Novellino, E., Mazzarella, L. & Sica, F. (2012). *Nucleic Acids Res.* **40**, 8119–8128.

Russo Krauss, I., Merlino, A., Vergara, A. & Sica, F. (2013). *Int. J. Mol. Sci.* **14**, 11643–11691.

Schmidt, K. S., Borkowski, S., Kurreck, J., Stephens, A. W., Bald, R., Hecht, M., Friebe, M., Dinkelborg, L. & Erdmann, V. A. (2004). *Nucleic Acids Res.* **32**, 5757–5765.

Straarup, E. M., Fisker, N., Hedtjärn, M., Lindholm, M. W., Rosenbohm, C., Aarup, V., Hansen, H. F., Ørum, H., Hansen, J. B. & Koch, T. (2010). *Nucleic Acids Res.* **38**, 7100–7111.

Sun, Z., Xiang, W., Guo, Y., Chen, Z., Liu, W. & Lu, D. (2011). *Biochem. Biophys. Res. Commun.* **409**, 430–435.

Vester, B., Hansen, L. H., Lundberg, L. B., Babu, B. R., Sørensen, M. D., Wengel, J. & Douthwaite, S. (2006). *BMC Mol. Biol.* **7**, 19.

Virno, A., Randazzo, A., Giancola, C., Bucci, M., Cirino, G. & Mayol, L. (2007). *Bioorg. Med. Chem.* **15**, 5710–5718.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Xu, Y., Ishizuka, T., Kimura, T. & Komiyama, M. (2010*a*). *J. Am. Chem. Soc.* **132**, 7231–7233.

Xu, Y., Ishizuka, T., Kimura, T. & Komiyama, M. (2010*b*). *J. Am. Chem. Soc.* **132**, 7231–7233.

Zaghloul, E. M., Madsen, A. S., Moreno, P. M., Oprea, I. I., El-Andaloussi, S., Bestas, B., Gupta, P., Pedersen, E. B., Lundin, K. E., Wengel, J. & Smith, C. I. (2011). *Nucleic Acids Res.* **39**, 1142–1154.

Zheng, G., Lu, X.-J. & Olson, W. K. (2009). *Nucleic Acids Res.* **37**, W240–W246.