DNA Structures

DOI: 10.1002/anie.200806244

Solution Structure of a Locked Nucleic Acid Modified Quadruplex: Introducing the V4 Folding Topology**

*Jakob T. Nielsen, Khalil Arar, and Michael Petersen**

Besides the Watson–Crick double helix, DNA can, if strict sequence requirements are met, adopt structures consisting of three or four strands, triplexes and quadruplexes. G-quadruplexes are four-stranded structures formed by DNA or RNA guanine tracts in which the strands associate through formation and stacking of G-tetrads (Figure 1).[1] G-quadruplexes display a high degree of structural polymorphism with their folding topology depending strongly on nucleotide sequence and environmental factors such as cations.[2] Nucleotide sequences possessing G-tracts with the ability to self-organize in G-quadruplex architectures are found in the context of several important biological processes, for instance, DNA telomere ends, immunoglobulin switch and gene promoter regions, and within the 5′-untranslated regions (5′-UTRs) of mRNA coding for oncoproteins.[3,4] The evidence for in vivo existence of G-quadruplexes is growing with the identification of cellular proteins that recognize and bind quadruplex structures or indeed control the formation of G-quadruplex structures.[5–7]

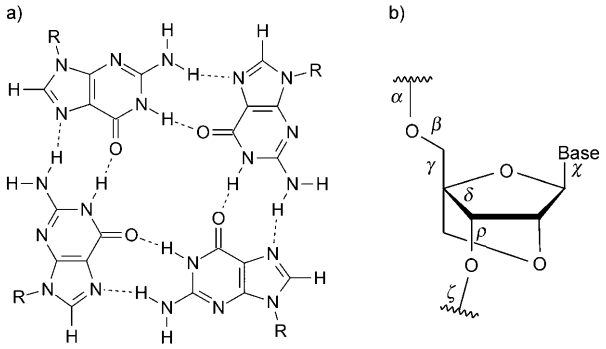

Figure 1. a) Base pairing in a Hoogsteen-paired G-tetrad. b) The chemical structure of LNA with torsion angles labeled.

Chemically modified nucleic acids have attracted much attention for use in oligonucleotide-based therapeutics.[8] A prominent nucleic acid analogue is LNA (locked nucleic acid; Figure 1), in which the modified ribose ring is locked in a C3′-*endo* or *N*-type (RNA-like) sugar pucker by the 2′-*O*,4′-*C*-methylene linkage. LNA is notable owing to its excellent hybridization properties with DNA and RNA and to its potential as an antisense and/or antigene molecule.[9,10]

The structural impact of LNA in parallel quadruplexes has been characterized by us and others,[11,12] but the impact on antiparallel structures is still not well known. Thus we chose as our model system the telomeric sequence from *Oxytricha nova*, d($G_4T_4G_4$) with potassium as a counterion. NMR spectroscopy and X-ray diffraction demonstrated that this DNA sequence forms a dimeric G-quadruplex with antiparallel G-columns and diagonal T_4 loops.[13,14] We systematically investigated the effect of the incorporation of LNA modifications in the G-tracts of the *O. nova* sequence. In general, we obtained ^{1}H NMR spectra with broad imino resonances in the 10.0–12.0 ppm range indicative of quadruplex formation, albeit with conformational heterogeneity present (see the Supporting Information, Figure S1). However, substitution with LNA at positions 2, 4, 10, and 12 led to a species which gave rise to a spectrum with sharp peaks from the imino protons, indicating that a single well-defined species was formed by this oligonucleotide.

The LNA-substituted sequence studied in detail, $GLGLT_4GLGL$, (L = LNA–G), formed a dimeric quadruplex where all guanines partook in G-tetrad formation, as shown by eight sharp guanine imino resonances in the 11.0–12.0 ppm range of the ^{1}H NMR spectrum (Figure 2). The dimeric nature of the quadruplex was verified by a heterogeneous mixing experiment (see the Supporting Information, Figure S2b) and, as indicated by the eight imino resonances, consequent formation of four G-tetrads with twofold symmetry. NMR spectroscopic assignments were carried out by standard methods, including the use of four different inosine-labeled samples (exchanging DNA guanosine units, one by one, with inosine).[15]

The structure of the LNA quadruplex was determined using 784 NOE distance restraints and a torsion-angle dynamics protocol designed to search conformational space thoroughly. The calculated structural ensemble converges well with a root mean square deviation (RMSD) of 1.69 Å (Figure 3) with the G-tetrads being better defined (RMSD 0.74 Å) than the T_4 loops (structure and refinement

[*] Dr. J. T. Nielsen, Dr. M. Petersen
Nucleic Acid Center, Department of Physics and Chemistry
University of Southern Denmark
Campusvej 55, 5230 Odense M (Denmark)
Fax: (+45) 6615-8760
E-mail: mip@ifk.sdu.dk
Dr. K. Arar
Proligo LLC
1 Rue Delaunay, 75011 Paris (France)

[**] We thank The Danish National Research Foundation and the Danish Center for Scientific Computing for financial support and The Instrument Centre for NMR Spectroscopy of Biological Macromolecules at The Carlsberg Laboratory, Copenhagen granted by The Danish Natural Science Research Council for providing spectrometer time at the 800 MHz spectrometer. Atomic coordinates have been deposited in the Protein Data Bank with accession code 2wcn.
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200806244.

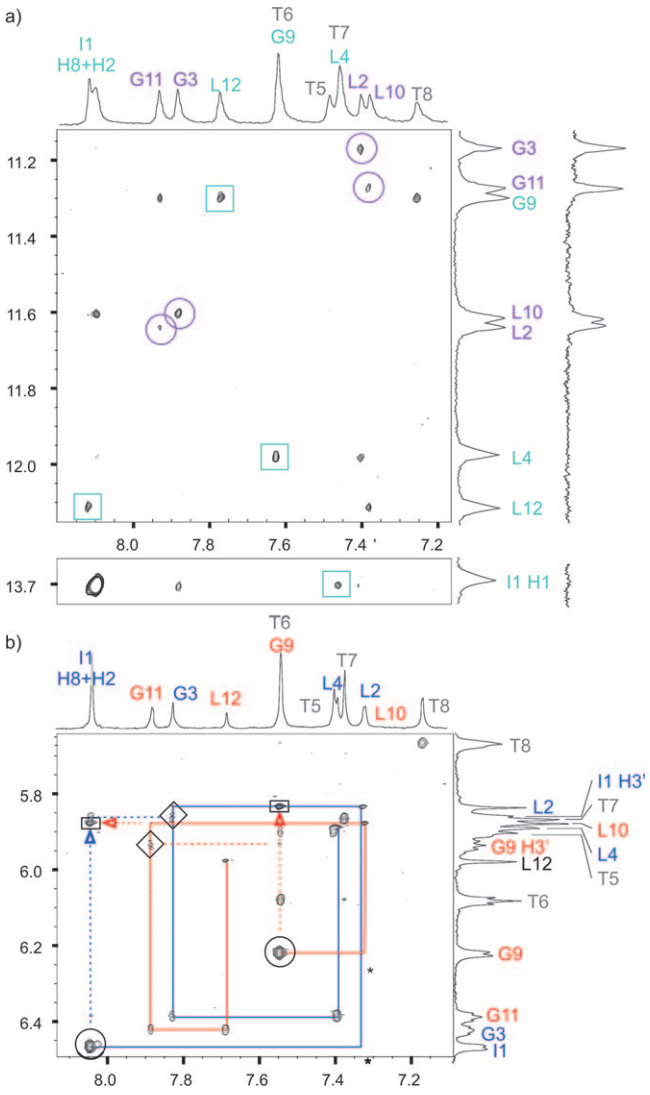

Figure 2. a) Imino to aromatic region of a 300 ms NOESY spectrum of ILGLT4GLGL in H_2O at 5 °C. 1D spectra with resonance assignments are shown at the corresponding axes (aromatic and imino protons horizontal and vertical, respectively). The 1D imino spectrum in D_2O after heating at 80 °C for 24 h is shown to the far right. Cross-peaks between GH1 and GH8 of neighboring bases in the tetrads are labeled with cyan squares (buried tetrads) and purple circles (exposed tetrads). b) Aromatic to H1′/H3′ region of a 300 ms NOESY spectrum of ILGLT4GLGL in D_2O at 25 °C. 1D spectra with resonance assignments are shown at the corresponding axes (aromatic and H1′/H3′ protons horizontal and vertical, respectively). The sequential connectivities between aromatic (H8) and H1′ protons are indicated for I1pL2pG3pL4 (blue lines) and for G9pL10pG11pL12 (red lines). The missing sequential connectivities, I1H1′–L2H8 and G9H1′–L10H8, are marked with asterisks. Important peaks are marked with circles (I1 and G9 H8–H1′), boxes (I1H8–L10*H1′ and G9H8–L2*H1′), and diamonds (I1H3′–G3H8 and G9H3′–G11H8).

statistics are included in Table S1 in the Supporting Information). The structure revealed how each G-stretch folds back in a V-shaped turn[16] and interacts with each of the three other G-stretches through formation of four G-tetrads. We term this new quadruplex folding topology as the V4 fold. The G-tetrads are square planar and stack parallel and coaxially with a distance of 3.3 Å between each other. In the distinct V form of each G_4 stretch, L2/L10 leaps over the tetrad and the chain then makes a sharp turn, reversing its orientation, to connect with G3/G11 (Figure 3 and Figure 4). The four V-shaped loops are interconnected with a fourfold pseudosymmetry and thus create four grooves (of which two are equivalent) in a criss-cross fashion. The grooves are deep and of medium width (4.3–7.1 Å).[17] The V4 fold incorporates features of both parallel and antiparallel quadruplexes in one structure and the remarkable folding topology leads to tetrad steps with both co-aligned (between outer and inner G-tetrads) and anti-aligned (between the two inner G-tetrads) hydrogen bonding, as in parallel and antiparallel quadruplexes, respectively. This alternation in hydrogen-bond orientation is accomplished by the four sharp turns in the backbone and by the mixed *syn*/*anti* and all-*anti* compositions in the outer and inner G-tetrads, respectively (Figures 3 and 4, and Figure S3 in the Supporting Information). In accordance with the V4 fold, the CD spectrum of $GLGLT_4GLGL$ displayed signatures of both parallel (negative and positive bands at ca. 245 and ca. 265 nm, respectively) and antiparallel (broad positive band at ca. 290 nm) stacking (see the Supporting Information, Figure S1).[18,19]

The presence of the V4 fold was backed up by a number of NOE features; sequential NOEs were missing for G1–L2 and G9–L10 and, instead, NOE contacts were detected for G1–L10* and G9–L2* (the asterisk indicates a nucleotide in the symmetry-related strand) and for G1–G3 and G9–G11 (Figure 2). These NOEs define the connection of G1 to G2 leaping over L10* and a chain reversal putting G1 and G3 close in space. The tetrad compositions and hydrogen-bond directions (G1-L12-G9-L4 and L2-G11*-L10-G3*) were deduced from characteristic imino–aromatic NOE contacts (Figure 2). The formation of one exposed and one buried tetrad is consistent with four slowly exchanging imino protons and four rapidly exchanging ones (Figure 2). The chemical shifts of the two G-stretches in each DNA strand (G1pL2pG3pL4 and G9pL10pG11pL12) are similar (see the Supporting Information, Table S2) indicating a pseudosymmetric relationship between the two chain fragments, which is indeed the observed in this structure.

The assembly of four V-shaped G-loops into a quadruplex fold appears to be seamless, preserving G-tetrad formation and stacking between the tetrads by adjusting the malleable DNA sugar-phosphate backbone (see the Supporting Information, Table S3). The abrupt rise between G1/G9 and L2/L10 is accomplished by extending the backbone between G1/G9 C3′ and L2/L10 C5′ maximally. The chain reversal between L2/L10 and G3/G11 is a result of changes in the backbone angles, ε, ζ, α, and β (Figure 1), mirrored in the nonhelical phosphorus chemical shifts for L2pG3 and L10pG11 (see the Supporting Information, Tables S2 and S3). The sugar puckers of G3 and G11 are C2′-*endo*, which ensures the longer intrastrand phosphorus distances required for the chain reversal, whereas the remaining nucleotides partaking in tetrad formation have C3′-*endo* puckers. The sugar puckers in the structure are in agreement with data derived from COSY spectra.

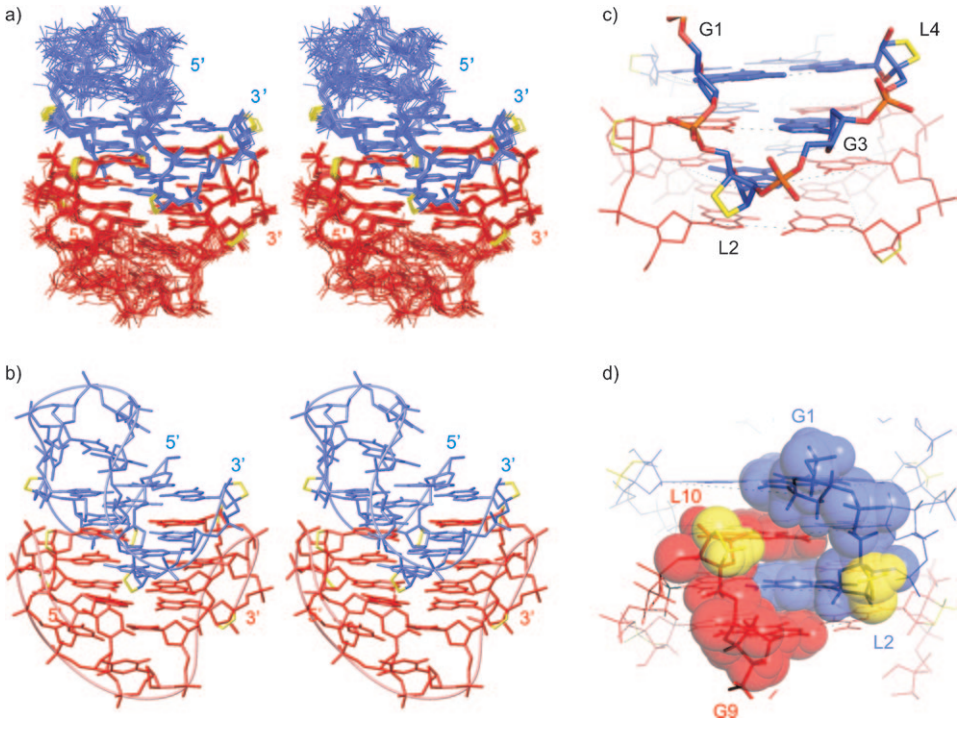

Figure 3. The structure of GLGLT$_4$GLGL. Hydrogen atoms are omitted for clarity. The two equivalent strands are shown in blue and red and the LNA oxymethylene bridges in yellow. a) Stereoview of the 20 structures calculated in the cross-validation procedure. b) Stereoview of a representative structure from the ensemble. c) Close-up view of a V-shaped loop. Phosphate groups are colored in red (oxygen) and orange (phosphorus). d) The intercalated groove; G1pL2 is intercalated with G9*pL10* (shown as transparent spheres).

The T_4 loops are lateral in the LNA quadruplex structure and display a higher degree of flexibility than the G-tetrads. T5 and T8 form a T–T base pair, as confirmed by an imino–imino NOE. Additional NOEs to the imino proton of G4 indicate that the pair stacks on the Watson–Crick edge of G4. The sugar puckers in the loops are primarily *S*-type, as in the diagonal loops of the native d($G_4T_4G_4$) quadruplex.[13]

Because of the C3′-*endo* sugar pucker and the high puckering amplitude (approximately 60°), the *syn* glycosidic conformation is sterically prohibited for LNA nucleotides, owing to a clash between the H3′ and N3 atoms. Positions 2, 4, 10 and 12 correspond to the *anti* positions in the native 12-mer *O. nova* sequence and hence it was possible for the LNA modified quadruplex to attain the same fold-back topology as the native quadruplex. Then it is, of course, a pertinent question why the four LNA sugars change the folding topology of the molecule completely. In a supramolecular complex formed by 16 lipophilic guanosine nucleosides, the tetrads form alternating steps of anti-aligned and co-aligned tetrads.[20] As there are no covalent constraints between tetrads in this system, the tetrads are free to adopt the optimum geometry for stacking and ion coordination interactions. The tetrad geometry of the V4-folded LNA quadruplex resembles the supramolecular complex in terms of tetrad alignment, being co-aligned, anti-aligned, and co-aligned in the three tetrad steps, and the tetrad–tetrad twist (see the Supporting Information). As such, the tetrad arrangement in the V4 fold appears to be intrinsically energetically favorable. It is also noteworthy that the dimeric quadruplex formed by d($G_3T_4G_4$) folds with a topology entirely different from the d($G_4T_4G_4$) quadruplex and contains a single V-shaped loop (Figure 4),[21] suggesting that DNA quadruplexes have some propensity towards forming V-shaped loops themselves and apparently need only small perturbations to change folding topology.

A possible explanation for the change in folding topology could be regulation of the groove hydration by the LNA sugars. In the DNA quadruplex, one groove is exceedingly narrow (3.0 Å between G10pG11 and G10*pG11*) which is disadvantageous for hydration,[6] whereas in the LNA quadruplex all grooves are deep and of medium width. In addition, the intercalation of the G1pL2 and G9*pL10* dinucleotide steps creates an intercalated groove which is also observed in an *i*-motif-like RNA octaplex structure built of r(UpG) steps.[22] The LNA intercalated groove mimics the RNA octaplex groove as sugar puckers are C3′-*endo* and the O2′ oxygen atoms of the LNA sugars could provide solvent anchoring points thus favoring solvation in the intercalated groove.

Based on the existence of intercalated, antiparallel G-tetrads in d(ApG) and U- and G-tetrads in r(UpG) steps,[22,23] it was proposed that a fully intercalated octaplex structure could occur with alternating U- and G-tetrads.[22] The V4-folded LNA-quadruplex lends credence to the possibility of an all-guanine intercalated octaplex with four columns, built on the mould of the G1-L10*-G2-G9* columns. Within the hypothetical octaplex formed by d(GpG) repeats, the 5′-guanosine would adopt a *syn* conformation and the 3′-guanosine a high-*anti* conformation, according to our structure. Glycosidic angles also alternate in the RNA octaplex formed by r(UpG) and in most antiparallel G-quadruplexes. Accordingly, an intercalated octaplex structure could form by integration of two antiparallel G-quadruplexes.

The structure presented herein expands the folding repertoire for quadruplexes and, as it appears to be energetically comparable with the usual fold-back type of quadruplexes, the V4 fold might also exist for DNA quadruplexes in certain environments.

Recently, an RNA G-quadruplex was shown to modulate translation of the N-ras proto-oncogene and it appears very likely that further analysis will identify additional RNA G-quadruplexes implicated in biological processes.[24] Given that

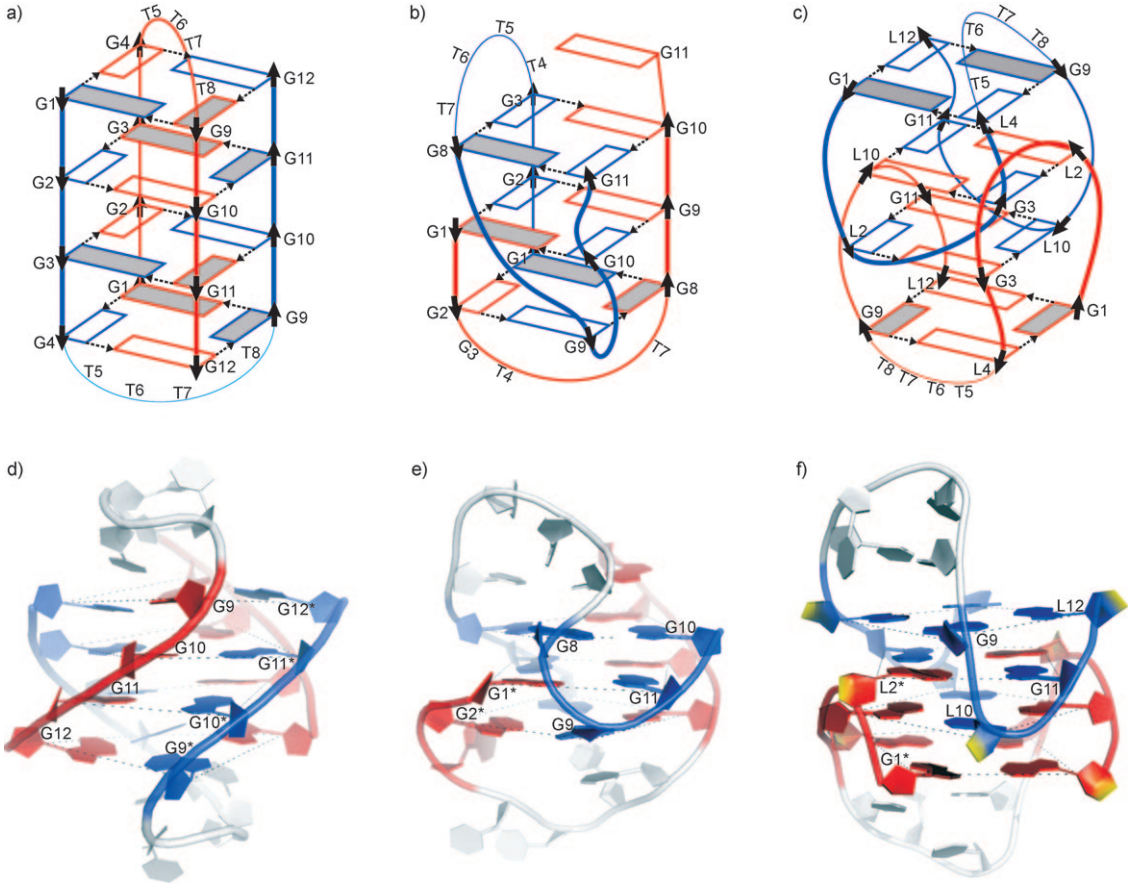

Figure 4. Schematic and molecular representations of dimeric quadruplex structures. a) and d) $G_4T_4G_4$;[13] b) and e) $G_3T_4G_4$;[21] c) and f) $GLGLT_4GLGL$. a), b) and c) Schematic representations: nucleobase tetrads are depicted as rectangles (shaded rectangles indicate nucleotides with *syn* glycosidic conformations), the directions of the strands and hydrogen bonds (from donor to acceptor) within tetrads are indicated with arrows, nucleotides are numbered (LNA nucleobases in bold). d), e) and f) Molecular representations: G-tracts are colored in red and blue, thymine loops in white, and LNA oxymethylene bridges in yellow. Tetrads are emphasized with dashed lines between C1′ atoms. The quadruplexes are viewed into the narrow groove (d), the V-shaped loop (e), and the V-shaped loop and intercalated groove (f).

LNA is an RNA mimic, the structure presented herein might presage the advent of novel folding motifs when atomic-resolution structures of RNA quadruplexes appear.

Experimental Section

NMR spectra were recorded by using Varian Inova spectrometers operating at either 500 (5 mm HCP probe) or 800 MHz (3 mm HCN probe). Sample concentrations were between 1 and 3 mM (strand concentration). Spectra of solutions in D_2O were recorded at 25 °C whereas those in H_2O were recorded at 5 °C. The structure was calculated using X-PLOR. Full experimental details are included in the Supporting Information.

Received: December 21, 2008
Published online: March 23, 2009

Keywords: DNA structures · G-quadruplexes · NMR spectroscopy · nucleic acids · protein folding

[1] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402 – 5415.
[2] *Quadruplex Nucleic Acids* (Eds.: S. Neidle, S. Balasubramanian), The Royal Society of Chemistry, Cambridge, **2006**.
[3] D. J. Patel, A. T. Phan, V. Kuryavyi, *Nucleic Acids Res.* **2007**, *35*, 7429 – 7455.
[4] S. Neidle, G. N. Parkinson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 275 – 283.
[5] R. Erlitzki, M. Fry, *J. Biol. Chem.* **1997**, *272*, 15881 – 15890.
[6] M. P. Horvath, S. C. Schultz, *J. Mol. Biol.* **2001**, *310*, 367 – 377.
[7] K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes, H. J. Lipps, *Nat. Struct. Mol. Biol.* **2005**, *12*, 847 – 854.
[8] J. Kurreck, *Eur. J. Biochem.* **2003**, *270*, 1628 – 1644.
[9] M. Petersen, J. Wengel, *Trends Biotechnol.* **2003**, *21*, 74 – 81.
[10] H. Kaur, B. R. Babu, S. Maiti, *Chem. Rev.* **2007**, *107*, 4672 – 4697.
[11] A. Randazzo, V. Esposito, O. Ohlenschläger, R. Ramachandran, L. Mayol, *Nucleic Acids Res.* **2004**, *32*, 3083 – 3092.
[12] J. T. Nielsen, K. Arar, M. Petersen, *Nucleic Acids Res.* **2006**, *34*, 2006 – 2014.
[13] P. Schultze, F. W. Smith, J. Feigon, *Structure* **1994**, *2*, 221 – 233.
[14] S. Haider, G. N. Parkinson, S. Neidle, *J. Mol. Biol.* **2002**, *320*, 189 – 200.
[15] Full experimental details are available in the Supporting Information.
[16] N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin, D. J. Patel, *J. Mol. Biol.* **2001**, *311*, 1063 – 1079.

[17] Spectra recorded with sodium counterions were largely identical to the spectra recorded with potassium counterions, leading us to believe that the described LNA quadruplex structure is not critically dependent on the identity of the counterion.
[18] V. Dapic, V. Abdomerovic, R. Marrington, J. Peberdy, A. Rodger, J. O. Trent, P. J. Bates, *Nucleic Acids Res.* **2003**, *31*, 2097 – 2107.
[19] A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723 – 2735.
[20] S. L. Forman, J. C. Fettinger, S. Pieraccini, G. Gottareli, J. T. Davis, *J. Am. Chem. Soc.* **2000**, *122*, 4060 – 4067.
[21] M. Crnugelj, P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2003**, *125*, 7866 – 7871.
[22] B. Pan, K. Shi, M. Sundaralingam, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3130 – 3134.
[23] J. Kondo, W. Adachi, S. Umeda, T. Sunami, A. Takenaka, *Nucleic Acids Res.* **2004**, *32*, 2541 – 2549.
[24] S. Kumari, A. Bugaut, J. L. Huppert, S. Balasubramanian, *Nat.Chem.Biol.* **2007**, *3*, 218 – 221.

