Solution Structure of a Unique G-Quadruplex Scaffold Adopted by a Guanosine-Rich Human Intronic Sequence

Vitaly Kuryavyi and Dinshaw J. Patel

1Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
*Correspondence: pateld@mskcc.org
DOI 10.1016/j.str.2009.10.015

SUMMARY

We report on the solution structure of an unprecedented intramolecular G-quadruplex formed by the guanosine-rich human chl1 intronic d(G3-N-G4-N2-G5-N-G3-N) 19-mer sequence in K+-containing solution. This G-quadruplex, composed of three stacked G-tetrads containing four syn guanines, represents a new folding topology with two unique conformational features. The first guanosine is positioned within the central G-tetrad, in contrast to all previous structures of unimolecular G-quadruplexes, where the first guanosine is part of an outermost G-tetrad. In addition, a V-shaped loop, spanning three G-tetrad planes, contains no bridging nucleotides. The G-quadruplex scaffold is stabilized by a T≡G≡A triple stacked over the G-tetrad at one end and an unpaired guanosine stacked over the G-tetrad at the other end. Finally, the chl1 intronic DNA G-quadruplex scaffold contains a guanosine base intercalated between an extended G-G step, a feature observed in common with the catalytic site of group I introns. This unique structural scaffold provides a highly specific platform for the future design of ligands specifically targeted to intronic G-quadruplex platforms.

INTRODUCTION

The diversity of possible G-quadruplex folds adopted by single-stranded guanosine-rich sequences containing four tracts of consecutive guanosine residues interspersed by loop sequences is currently under intensive investigation (reviewed in Davis, 2004; Burge et al., 2006; Phan et al., 2006; Patel et al., 2007; Neidle, 2009; Balasubramanian and Neidle, 2009; Lipps and Rhodes, 2009). We were interested in addressing the solution structure of a consensus motif for DNA promoter sequences found in protooncogenes (reviewed in Qin and Hurley, 2008), composed of four successive G3 guanine tracts (in bold), G-G-G-N-G-G-G-N-N-N-N-G-G-G-N-N-N-N-G-G-G-G, with the aim of probing for new G-quadruplex folds. During sequence design within the framework of this concept, we obtained a well-resolved imino proton nuclear magnetic resonance (NMR) spectrum for the 19-mer sequence with a central A-A segment, which contains successive G3, G4, G6, and G3 guanine tracts as outlined: G1-G2-G3-T4-G5-G6-G7-G8-A9-A10-G11-G12-G13-G14-T15-G16-G17-G18-T19. This sequence (Figure 1A) exhibited imino proton NMR characteristics of a single conformation in K+-containing solution, amenable for structural characterization (Figure 1B). A search for the corresponding sequence pattern in the human genome revealed its presence in the 5′-intron of the human chl1 gene with the exact sequence spanning positions 214.844 to 214.863 on chromosome 3 (Kent et al., 2002): C-G-G-G-C-G-G-G-G-A-A-G-G-G-T-G-G-G-A.

This observation is relevant considering the recent insight from bioinformatics searches of conserved guanosine-rich tracts with the potential to form polymorphic G-quadruplex folds in the first intron of human genes (Eddy and Maizels, 2008, 2009).

In addition, the CHL1 gene product belongs to the family of FANCJ helicases, a group of DNA-dependent ATPases capable of catalytically resolving G-quadruplex-forming blocks generated at guanosine-rich tracts, that can impede DNA replication (Wu et al., 2008, 2009).

Here we report on the solution structure of the chl1 intronic G-quadruplex in K+-containing aqueous solution and demonstrate that it adopts an unprecedented G-quadruplex fold exhibiting two unanticipated conformational features, one of which has not been observed previously in the unimolecular G-quadruplex topology literature (reviewed in Burge et al., 2006; Patel et al., 2007).

Surprisingly, the G-quadruplex solution structure of this guanosine-rich consensus sequence identified for a particular intron in the human genome exhibits certain features in common with the catalytic site of group I introns.

RESULTS

NMR Spectra and Spectral Assignments

The imino proton NMR spectrum of the chl1 intronic sequence in 50 mM K+-containing H2O buffer solution at 25 °C is characterized by 12 well-resolved resonances in the 11 to 12 ppm range (Figure 1B), a spectral region consistent with guanosine inosine N-H protons hydrogen bonded to nitrogen acceptors (N-H•••O), as observed for G-tetrads involved in G-quadruplex formation.

We pursued a strategy of initially assigning the guanosine imino protons by incorporating 2% 15N-labeled guanosines one at a time into the oligonucleotide sequence (Figure 1C) and then correlating these unambiguous assignments with guanosine H8 protons (Figure 1D), through long-range through-bond J-coupling experiments (Figure 1E) (Phan, 2000; Phan and Patel, 2002). The resulting guanosine imino proton assignments

Structure 18, 73–82, January 13, 2010 ©2010 Elsevier Ltd All rights reserved 73
of the \textit{chl1} intronic sequence are listed above the control NMR spectrum in Figure 1B and are analyzed in the context of the four guanosine tracks, G1-G2-G3, G5-G6-G7-G8, G11-G12-G13-G14, and G16-G17-G18, within the 19-mer sequence (Figure 1A). We anticipated that three guanosines within each track would be involved in G-quadruplex formation. Nevertheless, the imino proton of G3 at 10.55 ppm, which originates in the first G1-G2-G3 tract, is both broad and resonates upfield of the 11 to 12 ppm region (Figure 1B), suggestive of a guanosine imino proton not involved in N-H...O hydrogen bond formation with another guanosine. In addition, the imino proton of G8, though narrow, resonates upfield at 9.95 ppm, indicative of a guanosine imino proton hydrogen bonded to thymine oxygen acceptors. These results are consistent with involvement of G1-G2, G5-G6-G7, G11-G12-G13-G14, and G16-G17-G18 guanosine tracts in G-quadruplex formation. Thus, rather than an expected G5-G2-G3-G6 alignment of guanosine tracts required to form a standard G-quadruplex, the \textit{chl1} intron sequence adopts an unanticipated G2-G5-G3-G6 alignment of guanosine tracts.

The remaining non-exchangeable base and sugar protons of the \textit{chl1} intronic sequence were assigned using standard protocols by through-space (NOESY) and through-bond (COSY, TOCSY, \textsuperscript{13}C-GHSQC) experiments in 50 mM K\textsuperscript+ containing \textsuperscript{2}H\textsubscript{2}O buffer solution. In case of spectral overlap, resonance assignments were facilitated following 8-bromoguanosine (at positions 1, 5, 11, and 16) for guanosine substitutions, as well as deoxyuridine (at position 15) for thymine substitutions. The expanded NOESY contour plot (200 ms mixing time) in 50 mM K\textsuperscript+ containing \textsuperscript{2}H\textsubscript{2}O buffer at 25°C correlating base and sugar H\textsubscript{1} protons, together with assignment connectivities (base proton to its own and 5'-flanking sugar H\textsubscript{1} protons), is shown in Figure 2A. The complete tabulation of base and sugar proton chemical shifts are listed in Table S1 (available online).

The expanded NOESY stacked plot (50 ms mixing time) correlating base with sugar H\textsubscript{1} NOEs of the \textit{chl1} intronic sequence in 50 mM K\textsuperscript+ containing \textsuperscript{2}H\textsubscript{2}O buffer at 25°C exhibits four strong cross peaks for G1, G5, G11, and G16 residues (Figure 2B), indicative of syn torsion angles for these four guanosines (short base to sugar H\textsubscript{1} distances of 2.5 Å) (Patel et al., 1982) in the folded structure of the G-quadruplex in solution. In addition, four guanosine imino protons, assigned to G1, G6, G17, and G13, exchange slowly, as recorded 45 min after transferring...
(following lyophilization) the chl1 intronic sequence from 50 mM K+-containing H2O solution to its 2H2O counterpart at 25°C (Figure 2C). These results establish that four guanosines, G1, G6, G17, and G13, must originate from the central G-tetrad (slowest exchanging guanosine imino protons due to shielding from exchange with solvent) of the G-quadruplex.

**G-Tetrad Alignments**

We have assigned guanosines to the three G-tetrad planes by monitoring NOEs between guanosine imino and H8 protons between adjacent guanosines around individual G-tetrad planes (Figure 3A) in the NOESY spectrum (300 ms mixing time) of the chl1 intronic G-quadruplex in 50 mM K+-containing H2O solution at 25°C (Figure 3B). Thus, the imino proton of G1 shows an NOE to the H8 proton of G13 (Figure 3B, labeled G1/G13 in orange), the imino proton of G13 shows an NOE to the H8 proton of G17, the imino proton of G17 shows an NOE to the H8 proton of G6, and the imino proton of G6 shows an NOE to the H8 proton of G1, thereby assigning the guanosines and their order around the G1-G13-G17-G6 G-tetrad plane (Figures 3B–3D, assignments in orange). Related NOE tracings identify the G7-G11-G14-G18 (Figures 3B–3D, assignments in red) and G2-G5-G16-G12 (assignments in blue) G-tetrads that form the G-quadruplex. Since the imino protons of the G1-G13-G17-G6 G-tetrad exhibit the slowest exchange rate (Figure 2C), this central G-tetrad must be bracketed on either side by stacked G7-G11-G14-G18 and G2-G5-G16-G12 G-tetrads, resulting in the backbone tracing from G1 to G18 as shown in Figure 3D.

**G-Quadruplex Folding Topology**

The chl1 intronic G-quadruplex folding topology shown in Figure 3D contains syn guanosines at G1, G5, G11, and G16, and anti-guanosines at the remaining positions, consistent with the short mixing time NOESY data (Figure 2B). Unexpectedly, the first guanosine in the sequence, G1, is positioned within the central G-tetrad of the G-quadruplex. The G1-G2 segment forming the first column spans the lower two G-tetrad planes and is followed by the G3-T4 segment, which forms an edgewise loop, leading into the G5-G6-G7 segment of the second column that spans all three G-tetrad planes. Next, the G8-A9-A10 segment forms a second edgewise loop, leading into the G11-G12 step with its extended sugar-phosphate backbone, whereby G11 of the top G-tetrad is directly connected in a V-shaped extended alignment to G12 of the bottom G-tetrad, located on adjacent columns. The G12-G13-G14 segment that forms the third column is connected to the G16-G17-G18 segment that forms the fourth column of the G-quadruplex by a single-residue (T15) double chain reversal loop.

**Structure Calculations**

Initial distance-restrained and subsequent intensity-restrained molecular dynamics calculations (see Experimental Procedures) of the solution structure of the chl1 intronic G-quadruplex were guided by exchangeable and non-exchangeable proton restraints, with the restraints listed by category in Table 1. The ensemble of 17 refined superpositioned structures is shown in...
stereo in Figure 4A, with a representative refined structure shown in a ribbon representation in Figure 4B and a surface representation in Figure 4C. The ensemble of refined structures is well converged, exhibiting pair-wise root mean square deviation (rmsd) values in the 0.47 range for all nucleotides except for the T15 residue (Table 1), which is confined to a cluster of energetically preferred conformations pointing outwards from the groove (Figure 3A).

Both the top (G8-A9-A10) and bottom (G3-T4) lateral loops are structured within the chl1 intronic G-quadruplex fold. The top loop is stabilized by a T19*G8*A10 triple (Figure 5A), which stacks on the top G7*G11*G14*G18 tetrad (Figure 5B), while G3 of the bottom loop stacks on the G2*G5*G16*G12 tetrad (Figure 5C). Finally, A9 is unpaired and stacks on top of the T19*G8*A10 triple (Figure 5D).

**Analysis of Modified Sequences**
We have systematically probed specific structural elements within the proposed chl1 intronic G-quadruplex fold by base substitutions so as to gain insights into the range of tolerated mutations by the adopted scaffold and to derive a consensus sequence for bioinformatics analysis.

**Figure 3. Assignment of Individual G-Tetrads within the chl1 G-Quadruplex**
(A) Characteristic guanosine imino-H8 NOE connectivity patterns around a Gα*Gβ*Gγ*Gδ tetrad as indicated by arrows.
(B) Expanded contour plot showing imino to H8 connectivities of chl1 intronic 19-mer sequence in 50 mM K+ and 5 mM phosphate-H2O buffer (pH 6.8) at 25°C. Cross-peaks identifying connectivities within individual G-tetrads are color coded, with each cross-peak framed and labeled with the residue number of guanosine imino proton in the first position and that of the H8 proton in the second position.
(C) Guanosine imino-H8 connectivities observed for G1*G13*G17*G6 (orange), G7*G11*G14*G18 (red), and G2*G5*G16*G12 (blue) G-tetrads.
(D) Schematic representation of the chl1 intronic G-quadruplex fold. Anti-bases are indicated in cyan, while syn bases are indicated in magenta.
(F) Schematic of a V-shaped loop spanning three G-tetrad planes.
of the 3’-terminal residue T19, which is involved in T19•G8•A10 triple formation, by C, A, and G results in progressive deterioration of imino proton spectral quality, despite retaining the overall spectral pattern (Figure S4), consistent with T19 contributing to the stabilization of the G8-A9-A10 edgewise loop of the G-quadruplex fold, as a consequence of hydrogen bonding to G8 of the sheared GwA non-canonical pair (Figure 5A). As a result of the mutation studies, the consensus sequence compatible with the fold reported here is \( G_{3^\prime}NG_{4}NAG_{3}(A/C/T)G_{4}(A/C/T) \) and might provide useful constraints in future bioinformatics searches for intronic G-quadruplex folds.

**DISCUSSION**

The recent demonstration of guanosine-rich repeats with the potential for formation of polymorphic G-quadruplex scaffolds within the first intron of human genes (Eddy and Maizels, 2008, 2009) has highlighted the need to elucidate the diversity of G-quadruplex folding topologies adopted by human intronic guanosine-rich repeats.

To this end, we have solved the solution structure of a unique intramolecular G-quadruplex formed by the chl1 human intronic sequence in K\(^+\) solution. Inspection of the sequence G1-G2-G3-T4-G5-G6-G7-G8-A9-G10-G11-G12-G13-G14-T15-G16-G17-G18-T19 shows that four guanosine-rich tracts are separated by single (T4), double (A9-A10), and single (T15) connecting segments, suggesting the potential for formation of an all-parallel-stranded G-quadruplex involving three stacked G-tetrads connected by double-chain-reversal loops, as reported for the four guanosine-rich c-myc (Phan et al., 2004; Ambrus et al., 2005) and c-kit (Phan et al., 2007) sequences. However, this is not the case since we observe a new G-quadruplex folding topology containing two unanticipated conformational features, one of which has not been reported previously for unimolecular G-quadruplexes.

### First Guanosine is Positioned within the Central G-Tetrad

The first striking feature is that the first guanosine G1 is part of the central G-tetrad (Figure 3D), in contrast to the first guanosine having always been reported to be part of a terminal G-tetrad in all unimolecular G-quadruplexes reported to date (reviewed in Patel et al., 2007; Neidle, 2009) Thus, G11 and the G1-G2 step that are part of the same column are not connected to each other, in contrast to the unbroken linkage for the three other G5-G6-G7, G12-G13-G14, and G16-G17-G18 columns of the chl1 intronic G-quadruplex (Figure 3D). It should be noted that an earlier report did identify a bimolecular G-quadruplex where the first guanosine was also part of a central G-tetrad for one of the strands (Crmugelj et al., 2003).

### A V-Shaped Loop Spanning Three G-Tetrad Planes

Our group first identified the V-shaped loop topology for G-quadruplex folds (Zhang et al., 2001), where guanosines at adjacent corners of stacked G-tetrads are directly connected, resulting in a missing support column. The first example of a V-shaped loop within a G-quadruplex scaffold linked two adjacent G-tetrads as shown in Figure 3E (Zhang et al., 2001), with later examples identified for dimeric G-quadruplexes formed as per the given text.
by d(G₃T₄G₄) (Cmugelj et al., 2003) and d(GLGLT₄GLGL), where L is a locked nucleic acid (Nielsen et al., 2009), containing V-shaped loops that spanned three G-tetrad planes as shown in Figure 3F. The V-shaped loop motif represented in Figure 3F is observed for the unimolecular chl1 intronic G-quadruplex (Figure 3D), where G11 belongs to the top G-tetrad of one column while G12 belongs to the bottom G-tetrad of an adjacent column, accompanied by an extension of the backbone at the G11-G12 step. In essence, a second striking feature is that the G1 base is partially intercalated between two guanosine bases, G11 and G12, within the folding topology of the chl1 intronic G-quadruplex (Figure 4B).

**Stacked T19•G8•A10 Triple Stabilizes Edgewise G8-A9-A10 Loop**

The T19•G8•A10 triple involves a sheared G8•A10 pair alignment involving pairing of the minor groove edge of G8 with the major groove edge of A10 (Figure 5A), as reflected in the NOEs within the sheared pair and with the stacked adjacent G7•G11•G14•G18 tetrad. The O4 carbonyl group of T19 is hydrogen bonded to the imino proton of G8 in the T19•G8•A10 triple, readily explaining the upfield shifted 9.95 ppm imino proton chemical shift of G8, characteristic of a guanosine imino proton hydrogen bonded to a thymine oxygen acceptor (N-H•••O). Replacement of T19 by A and G, but not C, results in deterioration in imino proton spectral quality.
(Figure S4). The O4 carbonyl is common to both T and C, but not to the purines, emphasizing its importance as an acceptor for the hydrogen bond from the imino proton of G8. In addition, the larger size of the purine ring at position 19 would be more difficult to accommodate as part of a triple on top of the adjacent G7-G11-G14-G18 tetrad.

Consistent with the structure, unpaired A9, which stacks on the T19-G8-A10 triple, can be replaced by G, T, and C without impacting on the imino proton NMR spectrum (Figure S1). Since A9 is well defined in the structure, the conformation of the connecting sugar-phosphate is more important for the G8-A9-A10 edgewise loop than the base at position 9.

Stacked G3 Stabilizes Edgewise G3-T4 Loop

The two-residue bottom edgewise loop is composed of residues G3 and T4. They form a configuration where T4 is stacked on top of the G3, which in turn is stacked on the G2•G5•G16•G12 G-tetrad (Figure 5C). We observe a single average resonance for the two amino protons of G3, in contrast to separate hydrogen-bonded and exposed amino protons for guanines involved in G-tetrad formation. The average amino protons of G3 exhibit cross-peaks to the imino protons of all four guanosines in the G2•G5•G16•G12 G-tetrad (Figure 3B, peaks j, k, l, and m), consistent with the overlap geometry shown in Figure 5C, where the G3 amino protons are stacked over the center of the adjacent G-tetrad.

Replacement of G3 by I3, T3, or A3 results in doubling of the imino proton NMR spectra, including the upfield shifted imino proton of G8 at 9.95 ppm (Figure S3), indicative of an equilibrium between two G-quadruplex folds for the substituted analogs. Guanosine has an amino proton at the 2-position, which is missing in inosine and adenosine. It is conceivable that a K+ ion may be sandwiched between G3 and the G2•G5•G16•G12 G-tetrad and that monovalent cation coordination can serve to stabilize the presence of an amino group at the 2-position of the purine ring to stabilize a single G-quadruplex conformation. The proposed coordinated K+ cation could also slow down the exchange of the imino proton of G3, thereby accounting for the observation of a broadened resonance at 10.55 ppm (Figure 1B).

A Double-Chain-Reversal Loop

The parallel alignments of the G12-G13-G14 and G16-G17-G18 segments are bridged by a single residue (T15) double-chain-reversal loop in the structure of the chl1 intronic G-quadruplex (Figures 4A and 4B). Such double-chain-reversal loops were first reported a decade and a half ago by our group for the four-guanosine repeat Tetrhyymena G-quadruplex (Wang and Patel, 1994), but came into greater prominence with their observation in the crystal structures of two- and four-guanosine repeat segments in the G-binding motif in the catalytic core of group I intron ribozymes (Adams et al., 2004; Guo et al., 2004; Golden et al., 2004; reviewed in Stahley and Strobel, 2006). Indeed, in the 3.1 Å crystal structure of Azoarcus group I intron in complex with both 5’- and 3’-splice sites, corresponding to the splicing intermediate before the exon ligation step, the terminal nucleotide of the intron, dG206, partially intercalates between G128 and A127, as part of an anchored G-binding motif (Adams et al., 2004; Figure 6A). Partial intercalation of dG206 between purine bases (Figures 6B and 6C) and its pairing with the Hoogsteen edge of G130 (Figure 6D) in the G-tetrad (Figure 6G) in the Azoarcus group I intron (Adams et al., 2004) has striking parallels with partial intercalation of G1 between G11 and G12 (Figures 6E and 6F), with G1 paired with the Hoogsteen edge of G13 as part of the G1•G13•G17•G6 G-tetrad (Figure 6G) in the chl1 intronic G-quadruplex. One difference between these two alignments is that the sugar ring oxygens of dG206 in the intron (Figure 6B) and G1 in the G-quadruplex (Figure 6E) are oriented in opposite directions.

A Proposed Model for Phosphoryl Transfer Mediated by a G-Quadruplex Scaffold

A unique feature of the group I catalytic pocket is the in-line position of the activated 5’-exon 2’-OH and the conformationally constrained scissile phosphate at the intron-3’-exon junction, which is bracketed on either side by Mg2+ cations (Figure 6A; Stahley and Strobel, 2005). Such alignments, involving an constrained guanosine, define splice site selection during catalyzed phosphoryl transfer between guanosine and a substrate RNA.
strand (reviewed in Stahley and Strobel, 2006). We propose that the guanosine flanking the scissible phosphate can be equally well locked into position either through partial intercalation (Figures 6B and 6C) and base triple formation (Figure 6D), as observed for UdG206 in the group I intron (Adams et al., 2004), or through partial intercalation (Figures 6E and 6F) and G-tetrad formation (Figures 6G), as observed for G1 in the chl1 intronic G-quadruplex.

We therefore propose a model of a higher order RNA structure where a constrained guanosine as part of a G-quadruplex could participate in the splicing reaction at a quadruplex-duplex junction as shown schematically in Figure S5. The RNA duplex is on the left and the RNA G-quadruplex is on the right, with the 5′ and 3′ ends of the RNA labeled in Figure S5. A distinction between the experimentally determined group I intron splice site (Figure 6A) and the proposed model of the quadruplex-duplex junctional splice site (Figure S5B; expanded view of Figure SSA) is that the attacking nucleophile is a 3′-OH group in the former structure and a 5′-OH group in the latter model. The model in Figure SSA has been proposed in the spirit of stimulating further experiments at comprehensive molecular engineering of higher order G-quadruplex folds and clearly needs experimental validation.

Our structure of the chl1 intronic G-quadruplex raises the issue of whether a DNA segment can be accommodated 5′ to the first guanosine that is part of the central G-tetrad of the chl1 intronic DNA G-quadruplex. Our modeling studies shown in Figure SSA indicate that there is ample room to accommodate such a DNA segment without perturbing this unique G-quadruplex fold.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation**

The unlabeled and the site-specific low-enrichment (2% uniformly N15-labeled) oligonucleotides were synthesized and purified as described previously (Phan, 2000; Phan and Patel, 2002). The strand concentration of the sample was typically in the range 0.5 to 5.0 mM and was dissolved in 50 mM KCl and 5 mM potassium-phosphate buffer (pH 6.8).

**NMR Spectroscopy**

Experiments were performed on 600 MHz Varian NMR spectrometers with data recorded at 25°C. Guanosine base resonances were assigned unambiguously by using site-specific low-enrichment labeling and through-bond correlation at natural abundance (Phan, 2000; Phan and Patel, 2002). Assignments for some residues were verified and confirmed in independently synthesized...
samples with specific substitutions. Spectral assignments were also assisted and supported by COSY, TOCSY, $^{13}$C-HSQC, and NOESY spectra. Interproton distances involving exchangeable protons were categorized as strong (2.8 to 5.2 Å), medium (3.15 to 5.85 Å), or weak (3.5 to 6.5 Å) based on cross-peak intensities recorded in a NOESY spectrum (50 and 300 ms mixing time) in H$_2$O solution. Interproton distances involving non-exchangeable protons were measured from NOE build-ups using NOESY experiments recorded at five mixing times (50, 100, 150, 200, and 250 ms) in H$_2$O solution.

**Structure Calculations**

The structures of the ch1 intronic G-quadruplex were calculated using the X-PLOR program (Brünger, 1992) as described previously (Phan et al., 2007). The initial folds guided by NMR restraints listed in Table 1 were obtained using torsion dynamics. The structures were further refined by Cartesian dynamics and, finally, using relaxation matrix refinement.

The initial structure consisted of an extended DNA strand with randomized chain torsion angles of constituent nucleotides whose angles and bonds were set in accordance with most updated measurements (Geilbin et al., 1996; Clore et al., 1996).

**Torsion Dynamics**

In the heating stage, the regularized extended DNA chain was subjected to 60 ps of torsion-angle molecular dynamics at 40,000 K using a hybrid energy function composed of geometric and NOE terms. The van der Waals (vdW) component of the geometric term was set to 0.1, thus facilitating torsional bond rotations, while the NOE term included NOE-derived distances with a scaling factor of 150. The structures were then slowly cooled from 40,000 K to 1 K during 0.6 ps, while the force constants for the distance restraints were kept at 1 Kcal mol$^{-1}$ ā$^{-2}$. The resulting average structure was subjected to minimization.

**Distance Restrained Molecular Dynamics**

Cartesian molecular dynamics was initiated at 300 K and the temperature was gradually increased to 1000 K during 7 ps. The system was equilibrated for 0.5 ps, while the force constants for the distance restraints were kept at 1 Kcal mol$^{-1}$ ā$^{-2}$. Subsequently, the force constants were linearly scaled up to 150 during 17.5 ps. The system was then slowly cooled to 300 K in 14 ps and equilibrated at 300 K for 12 ps. The coordinates saved every 0.5 ps during the last 4.0 ps were averaged. The resulting average structure was subjected to minimization until the gradient of energy was less than 0.1 Kcal mol$^{-1}$. The flat planarity restraints imposed on the G-tetrads with the weight 10 kcal mol$^{-1}$ ā$^{-2}$ before the heating process were removed at the beginning of the equilibration stage. The electrostatic term was excluded from the energy function to increase the weight of covalent geometry terms during minimization process. The dihedral and hydrogen-bonding restraints for G-tetrad formation were maintained throughout the computations. The 17 best structures were selected at this stage.

**Relaxation Matrix Intensity Refinement**

To account for spin diffusion effects, all 17 distance refined structures were next subjected to the energy minimization with back-calculation of the NOESY spectra with X-PLOR (Nilges et al., 1991). The relaxation matrix was set up for the non-exchangeable protons, with the exchangeable imino and amino protons replaced by deuterons. NOE intensity volumes from 178 non-exchangeable cross-peaks for each of five mixing times (50, 100, 150, 200, and 250 ms) were used as restraints, with uniform upper and lower bounds of ±30%. Dynamics was started at 5 K, and the system was heated up to 300 K in 0.6 ps. During the subsequent relaxation, the force constant for NOE intensities was gradually increased from 0 to 300 Kcal mol$^{-1}$ ā$^{-2}$ with simultaneous decrease of the distance force constant of non-exchangeable protons from 50 to 30 Kcal mol$^{-1}$ ā$^{-2}$. The force constant for exchangeable protons and hydrogen bonds was kept at 100 Kcal mol$^{-1}$ ā$^{-2}$. After equilibration at 300 K for 3.0 ps, the resulting structure was subjected to minimization until the gradient of energy was less than 0.1 Kcal mol$^{-1}$. The NMR R-factor ($R_\text{ref}$) improved from initial value of 6.0% to 2.0% with simultaneous improved structure converyency.

**Proposed Model for Phosphoryl Transfer Mediated by a G-Quadruplex Scaffold**

The ch1 intronic RNA G-quadruplex was modeled by rmdf fitting of individual RNA nucleotides onto the DNA G-quadruplex structure (presented here) using common base atoms as superposition targets. All torsion angles of DNA residues except the sugar pucker pseudorotation angle (P) were copied into RNA nucleotides. The junction between the G-quadruplex and duplex was modeled using the Azoarcus group I intron junction as a prototype, with inversion of duplex orientation in order to form the connection with the 5′-end G1 of the ch1 G-quadruplex. The initial model was built manually and subsequently refined by energy minimization using the Discover module of Insight II software.

**ACCESSION NUMBERS**

The coordinates for the ch1 intronic G-quadruplex have been deposited in the Protein Data Bank under accession number 2KPR.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes five figures, one table, and Supplemental Text and can be found with this article online at doi:10.1016/j.str.2009.01.015.

**ACKNOWLEDGMENTS**

The research was supported by the New York Structural Biology Center supported by NIH grant GM66354. V.K. identified the potential of the ch1 intronic DNA sequence for G-quadruplex formation, as well as recorded and interpreted the NMR spectra and undertook the computations leading to structure determination under the guidance of D.J.P. V.K. noted the similarities with group I intronic sequences and formulated the proposed model for phosphoryl transfer at a quadruplex-duplex junction. D.J.P. and V.K. jointly wrote the paper.

Received: August 26, 2009
Revised: October 19, 2009
Accepted: October 22, 2009
Published: January 12, 2010

**REFERENCES**


