Stability and Cations Coordination of DNA and RNA 14-Mer G-Quadruplexes: A Multiscale Computational Approach

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · OCTOBER 2008

Impact Factor: 3.3 · DOI: 10.1021/jp804036j · Source: PubMed

6 AUTHORS, INCLUDING:

Bruno Pagano
University of Naples Federico II
56 PUBLICATIONS **849** CITATIONS
SEE PROFILE

Carlo Andrea Mattia
Università degli Studi di Salerno
104 PUBLICATIONS **1,316** CITATIONS
SEE PROFILE

Concetta Giancola
University of Naples Federico II
126 PUBLICATIONS **2,130** CITATIONS
SEE PROFILE

Franca Fraternali
King's College London
153 PUBLICATIONS **2,657** CITATIONS
SEE PROFILE

Stability and Cations Coordination of DNA and RNA 14-Mer G-Quadruplexes: A Multiscale Computational Approach

Bruno Pagano,[†] Carlo A. Mattia,[†] Luigi Cavallo,[‡] Seiichi Uesugi,[§] Concetta Giancola,[||] and Franca Fraternali*,[⊥]

Dipartimento di Scienze Farmaceutiche, Università di Salerno, via Ponte Don Melillo, Fisciano (SA), I-84084, Italy, Dipartimento di Chimica, Università di Salerno, via Ponte Don Melillo, Fisciano (SA), I-84084, Italy, Department of Environment and Natural Sciences, Yokohama National University, 79-7 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501, Japan, Dipartimento di Chimica, Università di Napoli Federico II, via Cintia, Napoli, I-80126, Italy, and The Randall Division of Cell and Molecular Biophysics, New Hunt's House, SE1 1UL, London, U.K.

Received: May 7, 2008; Revised Manuscript Received: July 20, 2008

Molecular dynamics simulations have been used to study the differences between two DNA and RNA 14-mer quadruplexes of analogous sequences. Their structures present a completely different fold: DNA forms a bimolecular quadruplex containing antiparallel strands and diagonal loops; RNA forms an intrastrand parallel quadruplex containing a G-tetrad and an hexad, which dimerizes by hexad stacking. We used a multiscale computational approach combining classical Molecular dynamics simulations and density functional theory calculations to elucidate the difference in stability of the 2-folds and their ability in coordinating cations. The presence of 2′-OH groups in the RNA promotes the formation of a large number of intramolecular hydrogen bonds that account for the difference in fold and stability of the two 14-mers. We observe that the adenines in the RNA quadruplex play a key role in conserving the geometry of the hexad. We predict the cation coordination mode of the two quadruplexes, not yet observed experimentally, and we offer a rationale for the corresponding binding energies involved.

1. Introduction

Guanine-rich nucleic acid sequences can adopt G-quadruplex structures stabilized by quartet layers of Hoogsteen paired guanine residues (G-tetrads).[1] These structures have applications in areas ranging from supramolecular chemistry to medicinal chemistry.[2] Most importantly, quadruplex-prone sequences are found in biologically significant genomic regions such as telomeres[3] or oncogene promoter regions,[4] and bind to a number of proteins or small molecules.[5,6] DNA G-quadruplex structures have been shown to inhibit telomerase activity[7] and have been supposed to play important roles in functions such as the control of gene expression.[8]

Recent studies have shown that, similarly to DNA, RNA G-quadruplexes play a key biological role in cellular processes.[9] A crucial example is represented by the fragile X mental retardation protein (FMRP) that binds with high affinity to purine-rich mRNA targets forming quadruplexes, suggesting a key role for these mRNAs in the pathogenesis of the fragile X mental retardation syndrome.[10] Short G-rich quadruplex-forming sequence within the 5′ untraslated region of the mRNA for the oncogene *NRAS* seems to act as a repressor for translation of mRNA into the protein p21.[11] These peculiar RNA structures can also modulate gene expression: the gene expression of the Shine−Dalgarno (SD) sequence can be regulated *in vivo*, without altering the GGAGG consensus sequence, by forming RNA quadruplex structures.[12]

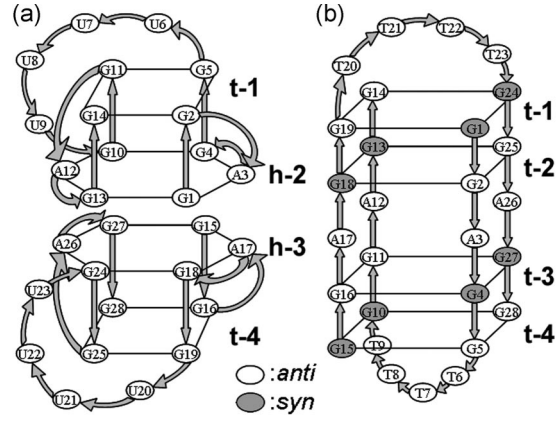

Figure 1. Schematic representation of the quadruplex structures formed by (a) r(GGAGGUUUUGGAGG) and (b) d(GGAGGTTTTGGAGG). Rectangles indicate the G-tetrads (planes t-*n*) and hexagons the hexads (planes h-*n*).

G-quadruplex structures can fold in different topological arrangements depending on the specific sequence, chain length and on the presence of monovalent or divalent cations.[13] In some cases, novel non-G-tetrads have been found like the T-tetrad,[14] A-tetrad,[15] ATAT, and GCGC tetrads.[16]

Recently, it has been found that the RNA sequence r(GGAGGUUUUGGAGG) folds into an unusual intrastrand parallel quadruplex, containing a canonical G:G:G:G tetrad, a U_4 double-chain reversal loop and a G:G(:A):G:G(:A) hexad.[17] Two r(GGAGGUUUUGGAGG) molecules dimerize in solution by stacking through the hexad-hexad interface (Figure 1a). The corresponding DNA 14-mer d(GGAGGTTTTGGAGG) is found to form an intermolecular quadruplex with antiparallel strand

* Corresponding author. E-mail franca.fraternali@kcl.ac.uk.
† Dipartimento di Scienze Farmaceutiche, Università di Salerno.
‡ Dipartimento di Chimica, Università di Salerno.
§ Department of Environment and Natural Sciences.
|| Dipartimento di Chimica, Università di Napoli Federico II.
⊥ The Randall Division of Cell and Molecular Biophysics.

10.1021/jp804036j CCC: $40.75
Published on Web 09/03/2008

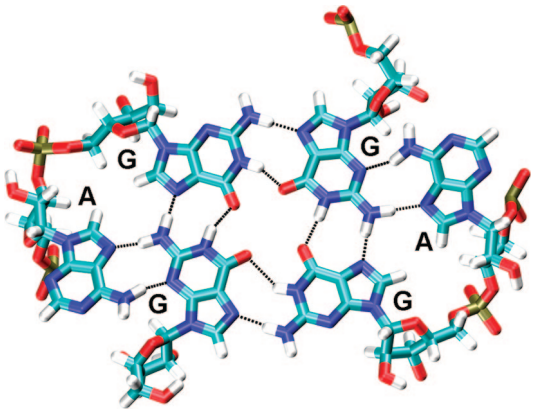

Figure 2. Top view of a hexad plane.

orientations composed by four canonical G-tetrads and two T_4 diagonal loops (Figure 1b).[18]

It is rather surprising that a RNA and its corresponding DNA sequences, which differ only for the presence of thymines instead of uracil bases and for deoxyribose sugars instead of ribose, can form entirely different quadruplex structures. It is known that DNA sequences can form both parallel and antiparallel quadruplex structures since deoxynucleosides can adopt easily both *anti* and *syn* glycosidic bond conformation. On the other hand, it is also known that RNA sequences prefer to form parallel quadruplexes: the observed difficulty of adopting a *syn* conformation does not promote the formation of quadruplexes containing antiparallel strand orientations.

Interestingly, the adenine residues of the DNA quadruplex (D14) are located at the center of the core helix, between two G-tetrads, whereas in the RNA quadruplex (R14) two adenine residues associate with a G-tetrad forming a hexad, creating in such way a larger surface that acts as dimerization surface (Figure 2). The adenine in the GGAGG sequence must therefore play a key role in stabilizing specific topological arrangements resulting from the stacking of hexads.

Considering that GGA triplet repeats have been found in biologically significant parts of DNA[19] and RNA[20] and that the GGAGG sequence has been shown to block the HIV-1 RNA dimerization *in vitro*,[21] it is of particular interest to study in details the folding topology of these sequences. So far, there are relatively little experimental data on RNA quadruplexes and, in particular, the folding of RNA into quadruplex structures is largely unexplored.

In order to improve the understanding of the main factors contributing to the stability of this particular fold of RNA, and to investigate the role played by the adenines in the different topologies of the D14 and R14 quadruplexes, we have used a multiscale computational approach consisting in (a) extensive molecular dynamics (MD) simulations of these molecules in explicit solvent in the presence of Na^+ ions and (b) more accurate DFT calculations in implicit solvent to accurately describe the energy associated to the assembly of the tetrads and hexads and to the cation coordination.

To investigate the ability of the d(GGAGGTTTTGGAGG) sequence to form, as in the case of RNA, a stable dimeric quadruplex architecture comprising two hexads and two tetrads, we constructed in silico a chimeric molecule (hereafter referred to as R-D14) with the DNA sequence of D14 and the structure of R14. The stability of this chimeric structure was investigated by MD simulations.

In recent years, computational studies have proved particularly useful in characterizing a wide range of structures of nucleic acids and their binding modes to cations,[22−24] but so far, no extended MD simulations studies were performed on RNA hexads. Moreover, no evidence at atomic detail is available about the coordination of these structures to cations, and, in particular, about the coordinating role of the hexads. We elucidate here a possible mechanism of coordination to Na^+ ions and we support our findings by density functional theory calculations with accurate description of the π-stacking interactions.[25]

2. Methods

Molecular Dynamics Simulations. All simulations were performed by using the GROMACS package,[26] employing the all-atom force field parm98.[27] The initial structures of the R14 and D14 were generated by using the coordinates of the previously reported structures.[17,18] The chimeric R-D14 molecule was generated in silico from the R14 structure by replacing the uracil bases with thymine bases and the ribose sugars with the deoxyribose sugars. The initial structure of R-D14 was energy minimized *in vacuo* by 1000 steps of the steepest descent method. The molecules were then neutralized with 26 Na+ ions (placed following electrostatic potential values) and solvated in boxes with more than 5000 TIP3P water molecules.[28] Initially, water molecules and ions were relaxed by energy minimization and allowed to equilibrate for 200 ps of MD at 300 K with the solute molecules restrained at their initial geometry. The bonds were constrained by the LINCS[29] algorithm with a force constant of 3000 kJ mol^{-1} nm^{-2}. Finally, the equilibrated systems were subjected to unrestrained MD simulations for 40 ns. Simulations were carried out with periodic boundary conditions at a constant temperature of 300 K. The Berendsen algorithm was applied for temperature and pressure coupling,[30] and the particle mesh Ewald method[31] was used for the calculation of electrostatic contribution to nonbonded interactions. MD trajectories were analyzed using GROMACS analysis tools.

DFT Calculations. The calculations were performed with the TURBOMOLE 5.8 package.[32] We used the GGA functional of Perdew, Burke and Ernzerhof (PBE).[33,34] The double-ζ quality basis set augmented with a polarization shells on each atom, referred as def2-SVP in TURBOMOLE, was used.[35] To speed up the calculations we took advantage of the RI-J approximation,[36] in connection with the auxiliary basis set corresponding to def2-SVP. Solvation in water was modeled using Klamt's COnductor-like Solvation MOdel (COSMO)[37] with a dielectric constant of 78.0 and a solvent radius of 1.4 Å.

Since standard DFT usually fails to correctly reproduce Van der Waals forces,[38,39] which are at the basis of stacking interactions, calculations were performed by including the DFT-D approach developed by Grimme to investigate base−base stacking.[40] Grimme's approach consists in the addition to standard DFT of an empirical correction term describing Van der Waals forces.[40] The specific DFT-D parametrization we used is the one developed by Ducèrè and Cavallo.[25,41] The total interaction energy ΔE was calculated as $\Delta E = E_{XY} - (E^0_X + E^0_Y)$, where E_{XY} is the energy of the optimized X−Y base pair, while E^0_X and E^0_Y are the energies of the isolated and optimized X and Y bases.

We calculated that the DFT + dispersion approach used here[25] results in a Mean Absolute Deviation (MAD) of 1.97 kcal/mol and in a rmsd of 2.39 kcal/mol from the highly accurate interaction energies of the JSCH-2005 database, which comprises 137 systems of biological interest (amino acids and nucleotides).[42] This is a clear improvement with respect to the uncorrected DFT approach that gives a MAD of 2.87 and a rmsd of 3.51 kcal/mol.

3. Results

Three independent MD simulations have been performed in the course of this study. The first two simulations were carried out on the quadruplex structures formed by two molecules of RNA 14-mer r(GGAGGUUUUGGAGG) (R14) and by two molecules of DNA 14-mer d(GGAGGTTTTGGAGG) (D14), whereas a third simulation was performed for a chimeric molecule with the DNA sequence of D14 and the structure of R14 (R-D14). The length of each simulation was 40 ns.

D14 and R14 simulations were started from the high-resolution NMR structures previously published,[17,18] while the starting structure of R-D14 was generated *in silico* from the R14 structure.

Since no evidence at atomic detail is available about the cations coordination of these structures, the molecules were immersed in water solution with the Na^+ ions not directly coordinated to the molecules. The binding mode of the cations to the quadruplex molecules was then revealed by monitoring the spontaneous insertion of the ions in the central channel of the structures under study.

To observe whether extensive MD simulations lead to these structures being disrupted, ions being expelled, or more ions replacing the water molecules in the central channel of the quadruplex, we run the simulations monitoring them for a total of 40 ns for each molecule.

MD Simulation of R14. MD simulation of R14 produces a stable trajectory, the root-mean-square deviation (rmsd) value of the theoretical structure as compared to the starting NMR structure is 2.4 ± 0.1 Å. The final structure of R14 along with the coordinated cations after 40 ns of MD simulation is reported in Figure 3a. The final coordination shows two Na^+ ions coordinated at the carbonyl oxygens of the guanine bases, positioned between the first tetrad (t-1) and hexad (h-2) planes, and the second hexad (h-3) and tetrad (t-4) planes, respectively.

Analyzing the simulation we observe that the central channel of R14, initially without any coordinated ion, becomes hydrated within the first few picoseconds by the entry of water molecules. A schematic illustration of coordination process of Na^+ ions to R14 is shown in Figure 4. After less than 100 ps of simulation we observe one Na^+ ion entering the channel from the bottom and situating itself between planes h-3 and t-4. After 1 ns another cation moves to the top of the channel, coordinating the 4 carbonyl oxygens of the G-tetrad forming t-1, positioned almost coplanarly to the tetrad. This situation remains unperturbed for about 200 ps of simulation (1 to 1.2 ns of the absolute simulated time), then the cation enters the channel displacing one water molecule and positioning between planes t-1 and h-2. No changes in the coordination of cations were observed until end of trajectory, but only small displacements of the cations inside the coordination site. Indeed, during the remaining simulation the two Na^+ ions are temporarily positioned exactly at the midpoints between hexad and tetrad planes, occasionally more close to the hexad plane, never leaving the coordination site.

Interestingly, no ion occupies the quadruplex channel between the planes of the hexads, but a water molecule positions itself in this region after few ps of simulation and remains trapped at the site for the entire simulation.

It is worth noting that the movement of ions from solvent into the central channel has a marked effect on the quadruplex structure and stability. In fact, in the absence of any coordinated cation, the guanine bases initially tend to rotate, slightly deforming the canonical G-tetrad scheme in which the four guanines are associated through a cyclic array of eight hydrogen bonds. Additionally, without any coordinated cation, the two outer G-tetrads at top and bottom of the molecule give rise to a bifurcated hydrogen bond geometry involving N1, N7, and O6 atoms of guanines, as already seen in other MD simulations of quadruplex molecules, but for inner G-tetrads.[43] On the contrary, bifurcated geometry was never observed in the simulation of R14 for the inner guanines forming hexads. Probably, this is due to the adenines hydrogen-bonding to the guanines of the inner G-tetrads. This structural feature confers more rigidity to the guanine bases, and does not allow them to form a bifurcated hydrogen bond geometry in the hexad planes.

All glycosidic bonds of guanine and uracil residues in the starting structure of R14 are in *anti* conformation, while the adenine residues are in a high *anti* conformation.[17] Throughout the trajectory, all glycosidic bond conformations of guanine and adenine residues are conserved, conversely, two uracil residues

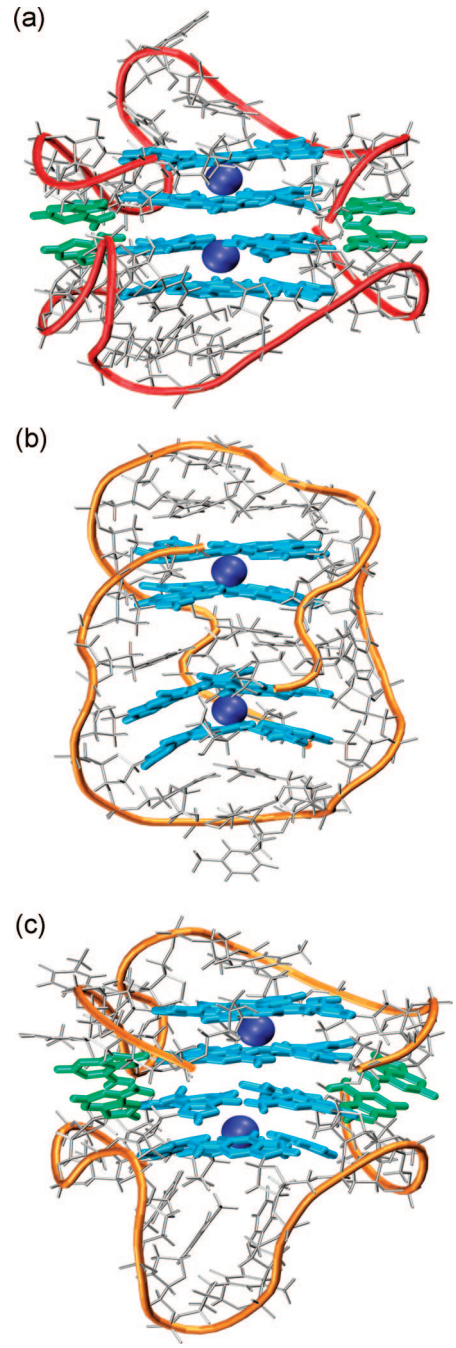

Figure 3. Snapshots of the final structures along with the coordinated cations after 40 ns of simulation for R14 (a), D14 (b) and R-D14 (c).

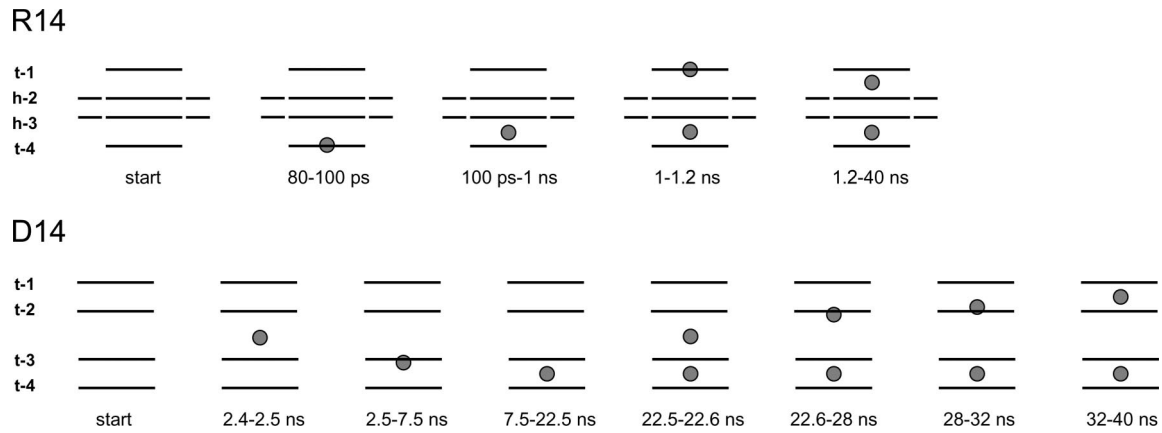

Figure 4. Schematic illustration of coordination process of Na^+ ions (gray spheres) to R14 (top) and D14 (bottom) quadruplex structures during the MD simulation.

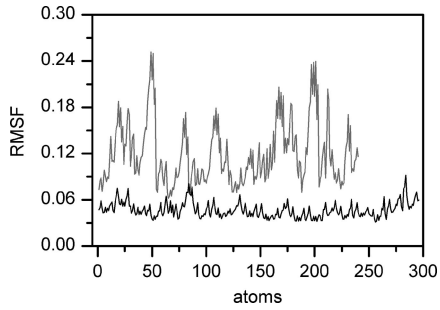

Figure 5. Atomic fluctuations of uracil residues (shaded line) and of guanine and adenine residues (black line) of R14.

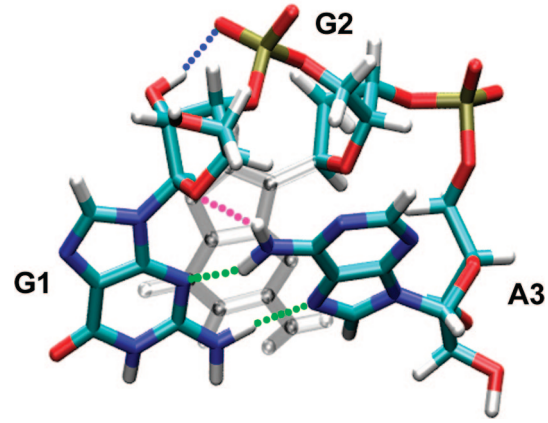

Figure 6. V-turn formed by nucleotides G1−G2−A3 that allows for the hexad formation. The H-bonds between G1 and A3 bases (green dots), the H-bond between NH_2 group of A3 and O4′ atom of ribose (magenta dots), and the H-bond formed by the 2′-OH with the phosphate oxygen (blue dots) are shown.

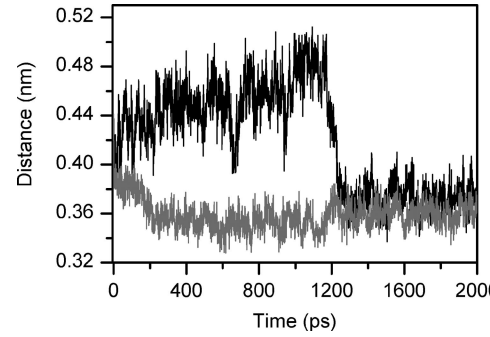

Figure 7. Time evolution, over the first 2 ns of MD simulation, of the relative distances between centers of mass of planes t-1 and h-2 (black line) and h-3 and t-4 (shaded line) of R14.

of the loops (U9 and U21) adopt a *syn* glycosidic bond conformation. This occurs because the two double-chain reversal loops of R14 are very flexible regions. This behavior is described in Figure 5, where the atomic fluctuations (RMSF) of selected R14 residues clearly show the higher fluctuation of uracil residues compared to the rest of the molecule. Visual inspection of the trajectory reveals that the uracil basis are not coplanar and hydrogen bonds are not formed between them. In addition, no uracil bases are stacked on the closest G-tetrads, except for the first residues of each loop (U6 and U20) that are sporadically stacked on the adjacent guanine residues.

During the simulation, the two adenine bases of each hexad are hydrogen-bonded to guanines and stacked to the other two adenines at the dimerization interface. In addition, all the four adenine bases are slightly rotated out of plane of the guanines to form a hydrogen bond between the NH_2 group of adenines and the O4′ atom of the adjacent guanosines (Figure 6).

As previously mentioned, binding to Na^+ leads to a stiffening of the molecule, reorientation of the bases and consequent stabilization of the molecule. Indeed, the presence of cations in the quadruplex channel minimize the electrostatic repulsion between the O6 atoms of the tetrad and hexad planes that, consequently, move closer to each other, as indicated by the distances of the two planes' centers of mass reported in Figure 7, optimizing the stacking interactions. This stabilization is also reflected in the average number of hydrogen bonds between the bases of each plane of R14 that, after ions coordination, stabilizes around an average value of 8 for the tetrads and 12 for the hexads.

The previous observations are summarized by the energy contributions as calculated over the first and last 200 ps of the trajectory. Especially interesting is the comparison between the van der Waals energy (that comprise the stacking contributions) of the molecule that from a starting value of −2150 (±45) kJ/mol decrease to a more favorable −2370 (±40) kJ/mol. Likewise, the RNA−Na^+ electrostatic energy contribution goes from −760 (±80) kJ/mol to a value of −1100 (±100) kJ/mol, contributing to the stabilization of the system.

MD Simulation of D14. MD simulations of D14 result in a stable trajectory, with a rmsd value along the trajectory between starting NMR and theoretical structures of 2.7 ± 0.1 Å. The final structure of D14 along with the coordinated Na^+ ions after 40 ns of MD simulation is shown in Figure 3b. D14 coordinates a total of two Na^+ ions placed between the first (t-1) and second

(t-2) tetrad planes, and between the third (t-3) and fourth (t-4) tetrad planes, respectively.

As for the case of R14, the central channel formed by the tetrads of D14, initially without any coordinated ions, was fully hydrated within the first few picoseconds by the entry of water molecules. Particularly hydrated was the cavity formed between planes t-2 and t-3 where the adenines leave space to water molecules. In fact, the adenine bases of D14 clearly do not participate in the formation of an A-tetrad during the MD simulation and are relatively mobile. The time evolution of the number of hydrogen bonds formed between adenines fluctuates around an average value of 1, indicating the formation of a single interaction that, in particular, is the hydrogen bond formed between the N3 atom of A12 and the NH_2 group of A17.

A schematic illustration of the coordination process of Na^+ ions to D14 is shown in Figure 4. The binding process of the cations to D14 during the simulation occurs on slower time scales than for R14, in fact the first cation takes position at midpoint between planes t-3 and t-4 after 7.5 ns of simulation, while the second cation situates itself between planes t-1 and t-2 after 32 ns. This is probably due to the presence of the diagonal loops that obstruct the channel, not allowing ions to enter from the top and the bottom of the molecule. Indeed, both the first and second ions coordinating the quadruplex pass through the central cavity between planes t-2 and t-3 that, thanks to the flexibility exhibited from the adenines, results quite accessible to the ions. This cavity is not a stable binding site for the cations, which are only temporarily coordinated during the simulation.

After about 2.5 ns of simulated time one Na^+ ion enters the central cavity and situates itself between tetrads t-3 and t-4, but closer to t-3 and coordinating its four carbonyl oxygens. This situation remains unchanged until 7.5 ns, when the cation moves downward and binds exactly at the midpoint between planes. After 22.5 ns of simulation another cation situates itself in the central cavity and coordinates from bottom the carbonyl oxygens of the tetrad forming t-2, in position almost coplanar with the tetrad. Then, at about 28 ns of simulation time the cation moves between tetrads t-1 and t-2, but closer to t-2, and finally at 32 ns situates itself at the midpoint between tetrads. This situation remains unchanged for the rest of the simulation.

Visual inspection of the trajectory before the binding of cations reveals that the guanine bases, as in R14, are subject to an in-plane rotation that slightly distorts the G-tetrads. Although complete ions coordination was fully achieved only after 32 ns, the molecule remains rather stable prior to this. Nevertheless, the coordination of cations has a marked stabilizing effect on the quadruplex, indeed, after Na^+ coordination occurs, the guanines do not rotate anymore. Additionally, the average number of hydrogen bonds between the guanine bases forming the tetrads remains fixed around an average value of 8, while before coordination this value was fluctuating around an average of 5−6 per tetrad.

Throughout the simulation, the *syn*−*syn*−*anti*−*anti* glycosidic bond conformation of the guanine bases in the tetrads is well conserved, so is the *anti* glycosidic bond conformation of thymine residues in the two diagonal loops.

Both loops show a significant low flexibility during the simulations. The first and third thymine residues of each loop are positioned on top of the outer G-tetrads (t-1 and t-4) of the quadruplex and parallel to them, forming stacking interactions with the aromatic rings of the guanine bases. During the simulation, we observed these thymines forming a hydrogen bond between the O2 oxygen and the N3 hydrogen atoms of the first and third thymine residues, with an average donor−acceptor distance of 2.8 Å. The hydrogen bond between the first and third thymine residues of the bottom loop (T6 and T8) remains stable during the entire simulation. Conversely, in the top loop, the first residue of the loop (T20) loses the interaction with the third residue (T22) after 15 ns of simulation and, simultaneously, forms a hydrogen bond with the fourth thymine (T23). This hydrogen bond, formed between the N3 hydrogen atom of T20 and the O4 oxygen of T23, remains stable until end of simulation, with a N3−O4 distance of about 2.8 Å. Moreover, the T22 residue continues to be close to the other two thymines and maintains its coplanar arrangement, stacked on top of the adjacent guanine residues, although it is no longer hydrogen bonded to T20 (Figure 8).

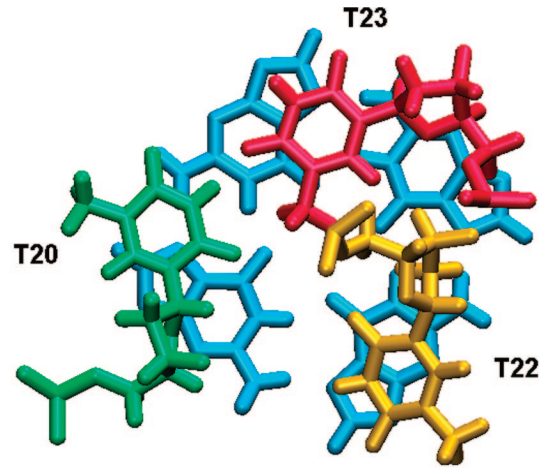

Figure 8. Thymine bases arrangement in the top loop of D14.

As mentioned above, the adenines are quite flexible and two out of four residues (A3 and A26) change the glycosidic bond conformation, while the other two residues (A12 and A17) are more rigid because involved in hydrogen bonding and preserve the glycosidic bond conformation throughout the trajectory. In particular, the starting *anti* glycosidic bond conformations of A12 and A17 are conserved, conversely, the A3 converts from *anti* to *syn* conformation, while the A26 converts from *syn* to *anti* conformation.

The energy contributions, as calculated over the first and last 200 ps of the D14 trajectory, reveal that of particular relevance is, in this case, the DNA−Na^+ electrostatic energy contribution that from an initial −340 (±60) kJ/mol value becomes −1200 (±80) kJ/mol in the final structure.

MD Simulation of R-D14. The R-D14 simulation stabilized to a rmsd value of 2.7 ± 0.1 Å with respect to the starting structure. At the end of MD simulation, the R-D14 molecule shows the same Na^+ coordination pattern as the R14 quadruplex (Figure 3c). Analyzing the trajectory, we observed that after about 200 ps of simulation the first Na^+ ion moves from the solvent to the channel, positioning itself between h-3 and t-4. Next, no other ions coordinate the molecule until 24.5 ns, when a second cation enters the channel placing itself between t-1 and h-2. The absence of any cation coordinated to the bases forming t-1 and h-2 for 24.5 ns is the cause of a large displacement of these residues. In particular, during the simulation the guanine bases of t-1 rotate out of plane, partially disrupting the cyclic array of hydrogen bonds of the tetrad. In addition, the movement of the adenine bases of h-2, involving also the adenines of the stacked h-3, causes the breaking of a number of hydrogen bonds between adenine and guanine bases of both hexads. On the other hand, the guanine bases of the

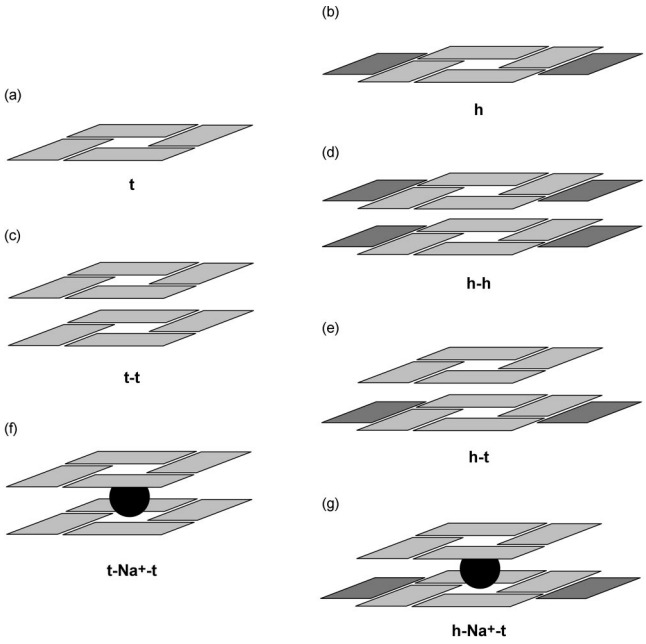

Figure 9. Schematic drawing of the systems considered for DFT calculations. Key: light gray box = guanine; dark gray box = adenine; black ball = Na^+; *t* = tetrad; *h* = hexad.

core of h-2 remain relatively stable, probably thanks to the stacking interactions with the closer guanines of h-3. Despite these marked fluctuation, the R-D14 molecule does not dissociate and the stacking between guanines as well as between adenines at the dimer interface of the two hexads is maintained for all the simulated time. The coordination of the second cation leads the guanine bases of t-1 to return coplanar each other and parallel to the h-2 plane, restoring the hydrogen bonds of the tetrad. In contrast, since the adenines do not return relatively coplanar to the plane of the guanines of the hexad, one of the two hydrogen bonds between adenine and guanine bases of each hexad and the hydrogen bond between the NH_2 group of adenines and the O4′ atom of the close guanosines are not restored.

The *anti* glycosidic bond conformations of all guanine and adenine residues are conserved throughout the trajectory. On the contrary, two thymine residues of the loops (T7 and T22) convert the glycosidic bond conformation from *anti* to *syn*.

Concerning the loops behavior, visual inspection of the simulation reveals that the thymine bases do not form hydrogen bonds between them and that no thymines are stacked on the G-tetrads. At the same time the last two thymines of each loop, although twisted toward the solvent, form stacking interactions.

Energy Contributions to Bases Stacking and Cation Coordination. To analyze in more detail the intra- and intermolecular energy contributions to the two different D14 and R14 folds, we calculated the energy contributions of the structural units typical of each system. We therefore considered the following systems: a tetrad (t) and a hexad (h) alone (Figures 9a and b, respectively), two stacked G-tetrads (t−t, Figure 9c), two stacked hexads (h−h, Figure 9d) and one hexad stacked with a G-tetrad (h−t, Figure 9e), these stacked systems were considered alone and in presence of a sandwiched Na^+ cation (Figure 9, parts f and g). For the sake of simplicity, we start this section discussing first systems in which only the H-bond contribution is relevant, that is adenine interaction with an isolated guanine or with an isolated G-tetrad, then we move to systems where only the stacking contribution is relevant, that is the stacking of two G-tetrads, of two hexads and of one hexad and one G-tetrad. Finally, we will discuss systems with the Na^+ cation.

The H-bond interaction energy calculated for an isolated A−G pair showing the same H-bonds pattern of the hexads, 9.7 kcal/mol, which is in reasonable agreement with the highly accurate value of 10.6 kcal/mol, calculated by Sponer and Hobza at the MP2 level,[44] which cannot be applied to the larger systems described below due to the computational cost. On the other hand, the energy gain associated to the H-bonding of two adenines to a G-tetrad to yield the corresponding hexad, amounts to 17.6 kcal/mol, or 8.8 kcal/mol per adenine molecule. Thus, H-bonding of an adenine to a G-tetrad is slightly unfavored with respect to the H-bonding of an adenine to an isolated guanine. This difference can be associated with slightly higher deformations due to the H-bonding of two adenines to a G-tetrad and also indicates that there is no synergic effect that reinforces the H-bonding interaction.

We now discuss systems in which only stacking interactions are relevant. The stacking energy of the t−t system in the absence of the Na^+ cation (Figure 9c), which is the interaction energy between two G-tetrads, was calculated to be 39.4 kcal/mol, while the stacking energy of the h−h system (Figure 9d), which is the interaction energy between two hexads, was calculated to be 65.7 kcal/mol. The difference of these two stacking energies, 26.3 kcal/mol, equivalent to 13.1 kcal/mol for each pair of stacked adenines, can be assumed to correspond to the stacking contribution of the 4 extra adenines in the h−h system.

At this point we are able to split the total energy gain associated to coordination of 4 adenines to the t−t system to yield the h−h system, which we calculated to be 61.6 or 15.4 kcal/mol per adenine, into the H-bonding and stacking terms discussed above. Indeed, the total interaction energy of 61.6 kcal/mol is in excellent agreement with the value of 61.4 kcal/mol, obtained as the sum of the stacking energy of 2 adenine pairs in the h−h system (2*13.1 kcal/mol = 26.2 kcal/mol), and the H-bonding energy of 4 adenines with the G-tetrads (4*8.8 kcal/mol = 35.2 kcal/mol). This suggests that the stacking and the H-bond terms contribute roughly to 40% and 60% of the total interaction energy of the 4 adenines with two stacked G-tetrads to form the h−h system, respectively. Further, we also calculated the coordination of only two adenines to a G-tetrad of the t−t system, to yield the h−t system in the absence of the Na^+ cation, which results in an energy gain of 18.2 kcal/mol, or 9.1 kcal/mol for each adenine. These values are marginally higher than the interaction energy of two adenines with an isolated G-tetrad, which indicates that stacking interactions between the adenines and the guanines in the h−t system contribute negligibly to the formation of the h−t system.

Moving to structures comprising two stacked G-tetrads sandwiching a Na^+ cation (t−Na^+−t, Figure 9f), we first calculated the energy gain, 23.6 or 12.8 kcal/mol for adenine, due to the coordination of two adenines to one of the G-tetrads to form a structure corresponding to a hexad stacked to a G-tetrad, with a sandwiched Na^+ cation (h−Na^+−t, Figure 9g). In consideration of the calculated interaction energy of 8.8 kcal/mol for each of the adenines in the isolated h−t system, these results suggests that there is an additional energy gain of 4.0 kcal/mol when two adenines coordinate to one of the G-tetrads in the h−Na^+−t system. This stabilizing effect can be associated to the increased H-bonding interaction between the adenines and the G-tetrad in the presence of the cation, probably due to more polarized guanines in the presence of the Na^+ cation. In

conclusions, these results suggest that the formation of the h−t system is favored by the presence of the Na^+ cation by 4 kcal/mol.

Discussion

We have carried out extensive MD simulations of quadruplex structures formed by two oligomers of RNA 14-mer or DNA 14-mer in explicit water and in the presence of Na^+ ions in solution.

The focus of our study was the dynamic behavior of these quadruplex molecules, the binding mode of the ions, and their influence in stabilizing these quadruplex structures, since no experimental evidence is available about the coordination mode to ions for these structures. The study aimed also at improving the understanding of the main factors contributing to the stability of the unusual folding of R14 and, in particular, of the role played by the extra adenine bases in the R14 hexad.

We are aware that a new parametrization (parmbsc0)[45] has been recently developed to correct for the artifactual conformational changes of the phosphodiester backbone that were seen in long simulations, but this paper was published once this work was already in progress. As a check, we have nevertheless calculated the distributions of α/γ torsion angles for the R14 and D14 systems (Supporting Information, Figures S1 and S2) and the occurring α/γ transitions during our performed trajectories (Tables S1 and S2). Indeed a number of transitions to alternative conformations occur for the R14 and D14 backbone during the simulated time, most of them after tenths of nanoseconds.

The main purpose of this study is, however, to study the coordination of cations to the chosen systems and most transitions occur after the ions are coordinated and are not correlated with it, therefore do not affect sensibly the studied processes.

In addition, it should be noted that in comparison with duplex and triplex DNA systems and, in general, to more canonical structures and geometries, DNA and RNA quadruplexes can be formed by a larger variety of α/γ conformations with some heterogeneity and, indeed, some non canonical (i.e., g^-/t, g^+/t and t/t) conformations have been observed.

Both R14 and D14 quadruplexes show a final coordination of two Na^+ ions coordinated in the central channel at the carbonyl oxygens of the guanine bases. Starting without any Na^+ directly coordinated to the quadruplexes, the simulations reveal that in both cases the cations spontaneously move into the central channel of the structures under study, without causing dissociation of the quadruplexes and with a marked stabilizing effect. Interestingly, the coordination process to the two quadruplexes is rather different. Indeed, for the R14 molecule, the Na^+ ions enter the channel from the top and the bottom of the quadruplex, leading to a faster process (occurring within the first 2 ns), whereas, to coordinate the D14, the cations pass through a cavity sited between the inner tetrads, leading to a markedly slower coordination process (coordination was achieved at 32 ns).

Brownian dynamics simulation for the sequence d(GGAGGAG), whose folding is comparable to R14 (but with a hexad motif composed of two different strands), have predicted five potential ion-binding sites, three of them in the channel between the four planes, and other two in the grooves between hexads.[46] In agreement with these predictions, we found two Na^+ ions sandwiched between the planes of R14, conversely, we do not found Na^+ ions between hexads and do not observe in the grooves between hexads of R14 quadruplex any electronegative pocket necessary for the binding of cations.

The adenines do not directly participate to the Na^+ ions coordination in D14, nevertheless, from DFT calculations, we observe an energy gain of 4.0 kcal/mol when two adenines coordinate to one of two stacked G-tetrads in the presence of a sandwiched Na^+ ion. This stabilizing effect is probably due to more polarized guanines in the presence of the cation, that increase the H-bonding interaction between the adenines and the G-tetrad.

The simulation of R14 shows low fluctuations for the residues forming tetrads as well for those forming hexads, especially after cations coordination. In particular, the position of the adenine bases in the hexads and their hydrogen bonding pattern are well conserved. On the other hand, high fluctuations were observed for the residues of the two double-chain reversal loops.

D14 simulation also shows low fluctuations for the guanine residues of the tetrads, whereas high fluctuations were observed for the adenine bases, which are relatively free to move. Conversely, the positions of the thymines in the two diagonal loops are well conserved, revealing a stable nature of these loops.

We have also analyzed the atomic fluctuations of the backbone atoms of the two 14-mers forming the R14 and D14 molecules (Supporting Information, Figures S3 and S4, respectively). The major difference observed in the fluctuations of the two systems involves the loops. Indeed, the atoms of the U4 loops of R14 backbone are much more prone to fluctuate in comparison to the central part of the structure. Conversely, the largest RMSF values of D14 backbone are not associated with the atoms of the T4 loops.

Concerning the R-D14 simulation, some considerations need to be done. We constructed the R-D14 molecule in order to investigate the ability of the DNA sequence d(GGAGGTTTTGGAGG) to form, as in the case of RNA, a stable dimeric quadruplex architecture comprised of two hexads and two tetrads. Patel and co-workers found that the DNA sequence d(GGAGGAG) in Na^+ solution is able to form a G:G(:A):G:G(:A) hexad motif composed of two different strands, also in that case stabilized by the stacking of the two hexads.[46] Moreover, circular dichroism (CD) experiments revealed that at low concentration of cations in solution, the sequence d(GGAGGTTTTGGAGG) shows a similar spectral pattern as r(GGAGGUUUUGGAGG) with a positive band at 262 nm,[47] characteristic of a parallel quadruplex structure, suggesting or at least not excluding the possibility that, under particular conditions, this DNA sequence could folds as the R14 molecule. In silico experiments like this one can help understanding the main factors contributing to the stability of different structural elements in determining a particular fold.

The MD simulation of R-D14 shows that d(GGAGGTTTTGGAGG) DNA is capable of forming cation-stabilized parallel quadruplexes with the same folding of R14. However, the simulation of this hypothetical structure revealed significant differences between the DNA and RNA quadruplexes. Throughout the simulation, the R-D14 molecule appears to be more unstable than the other two quadruplexes when the central channel is empty, although the quadruplex structure is maintained. Moreover, the final cation-stabilized structure shows a partial rupture of the hydrogen bonding pattern of the hexads.

The main difference between the RNA and DNA molecules studied here is to be found in the presence of the hydroxyl groups in the sugar moieties of RNA. In fact, the 2′-OH groups in the RNA promote the formation of a large number of

intramolecular hydrogen bonds that contribute greatly to the stability of the R14 molecule. The total number of intramolecular hydrogen bonds formed by R14 is on average 53, while this number is 38 and 37 for D14 and R-D14, respectively. As expected, we observed long time-residence intramolecular hydrogen bonds between 2′-OH and any of the hydrogen-bond acceptor atoms of the next neighboring nucleotide. From our analysis, the hydrogen bonds with the longest time-residence are the ones formed with phosphate oxygen atoms (Figure 6), and in particular, the hydrogen bonds 2′-OH(G1)···OP(G2), 2′-OH(G10)···OP(G11), 2′-OH(G15)···OP(G16), and 2′-OH-(G24)···OP(G25).

These results suggest that the nucleotides G1−G2−A3 and G10−G11−A12 (and their symmetrical counterparts for the second monomer) forming the "sharp V-shape turn" that allows for the hexad formation,[17] are strongly stabilized by the particular network of hydrogen bonds formed around the adenines, as shown in Figure 6. This network of hydrogen bonds creates a robust scaffold which allows for more stability of the molecular architecture and that infers more rigidity to the stacked bases. In particular, the hydrogen bond formed by the 2′-OH with the phosphate oxygen acts as a "clip", holding in place the geometry of the V-turn.

A more detailed analysis of the contribution of the 2′-OH groups to the rest of the RNA molecule is performed on the evaluation of their contribution to the electrostatic energy when they are involved in hydrogen bonding vs when they are not. Therefore, we calculated the Coulomb short-range contribution of each of the 28 2′-OH groups and the rest of the molecule. The average contribution of the 2′-OH that form hydrogen bond equals to 16.1 kcal/mol, while the average contribution of the 2′-OH that do not form hydrogen bond amounts to 9.7 kcal/mol. The difference between these two values is 6.4 kcal/mol and this is the average contribution of each hydrogen bond formed by the 2′-OH moiety. This result, together with the discussed MD analysis of the hydrogen bonds network formed around the adenines quantifies the role of the 2′-OH groups in the stability of the RNA system.

From our DFT calculations, we were able to quantify the H-bonding contribution of each adenine base to the stability of this scaffold that was calculated to be 8.8 kcal/mol in the absence of Na^+ and 12.8 kcal/mol in the presence of the cation. On the other hand, we found that the stacking interactions between the adenines of a hexad and the guanines of a G-tetrad contribute negligibly to the stability of the system.

To compare MD results with the DFT energetical analysis presented here, we have extracted the energy contributions of the following systems: (a) t−t, (b) h−t, (c) h−h. Furthermore we have isolated the van der Waals and the electrostatic contributions. The first is directly comparable to the contribution of stacking from the DFT while the electrostatic contribution has been calculated to have an estimate of the hydrogen bonding contribution to the total energy. In fact, the implementation of Amber force field in the Gromacs package is such that there is not an explicit term for the hydrogen bond. In consideration of the fact that the hydrogen bond is mostly due to electrostatic interactions we consider this approximation valid.

The t−t system presents a van der Waals energy contribution of 41.8 kcal/mol to be compared with 39.4 kcal/mol from the DFT calculations. The van der Waals energy of the t−h system is 46.0 kcal/mol. This result confirms the fact that "stacking interactions between the adenines and the guanines in the h−t system contribute negligibly to the formation of the h−t system" as deduced from the DFT results. Finally for the h−h system we calculated 58.5 kcal/mol, while the DFT results reports a value of 65.7 kcal/mol.

These results confirm a very good performance of the force field parameters in characterizing stacking interactions with respect to DFT calculations.

For the electrostatic contribution to the energy stability of the tetrad system vs the hexad system we calculate 38.2 and 58.6 kcal/mol, respectively. Therefore we conclude that adding one adenine to the system results in a sensible electrostatic gain, mostly due to a larger number of hydrogen bonds formed, as shown also from our MD analyses.

The loss of hydroxyl groups in the sugar moieties of DNA implies a loss of several hydrogen bonds, however, the results show that d(GGAGGTTTTGGAGG) can folds as the r(GGAG-GUUUUGGAGG) even though the structure is clearly less stable.

It is reasonable to suppose that, during the quadruplex formation, some of these d(GGAGGTTTTGGAGG) fold as the R14 and then convert their structure to a more stable antiparallel structure. In this view, the stable nature of the loops in the antiparallel structure could have a key role in determining the folding topology.

In conclusions, our multiscale computational approach has helped us in clarifying the differences in term of structure and stability between the two diverse folding of the RNA and DNA 14-mer sequences. In particular, we have observed that the adenines in the R14 quadruplex play a key role in conserving the geometry of the hexad and we have also quantified their energetic contribution. In addition to this we were able for the first time to propose a binding mode for each of the 2-folds to Na^+ ions and to quantify the energetics involved in this binding. From our results it is clear that RNA G-quadruplexes deserve more investigations exploring their potential use and role in biological context.

Supporting Information Available: Figures of torsion angles distributions, atomic fluctuations of the backbone atoms and tables of α/γ backbone torsions. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013.
(2) Davis, J. T. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 668.
(3) Neidle, S.; Parkinson, G. N. *Curr. Opin. Struct. Biol.* **2003**, *13*, 275.
(4) Rangan, A.; Fedoroff, O. Y.; Hurley, L. H. *J. Biol. Chem.* **2001**, *276*, 4640.
(5) Fry, M. *Front Biosci.* **2007**, *12*, 4336.
(6) Pagano, B.; Giancola, C. *Curr. Cancer Drug Targets* **2007**, *7*, 520.
(7) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature* **1991**, *350*, 718.
(8) Eddy, J.; Maizels, N. *Nucleic Acids Res.* **2008**, gkm1138.
(9) Shafer, R. H.; Smirnov, I. *Biopolymers* **2001**, *56*, 209.
(10) Schaeffer, C.; Bardoni, B.; Mandel, J. L.; Ehresmann, B.; Ehresmann, C.; Moine, H. *EMBO J.* **2001**, *20*, 4803.
(11) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. *Nat. Chem. Biol.* **2007**, *3*, 218.
(12) Wieland, M.; Hartig, J. S. *Chem. Biol.* **2007**, *14*, 757.
(13) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402.
(14) Patel, P. K.; Hosur, R. V. *Nucleic Acids Res.* **1999**, *27*, 2457.
(15) Patel, P. K.; Koti, A. S.; Hosur, R. V. *Nucleic Acids Res.* **1999**, *27*, 3836.
(16) Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D. J. *J. Mol. Biol.* **2001**, *312*, 1073.
(17) Liu, H.; Matsugami, A.; Katahira, M.; Uesugi, S. *J. Mol. Biol.* **2002**, *322*, 955.
(18) Uesugi, S.; Liu, H.; Kugimiya, A.; Matsugami, A.; Katahira, M. *Nucleic Acids Res. Suppl.* **2003**, 51.

(19) Heller, M.; Flemington, E.; Kieff, E.; Deininger, P. *Mol. Cell. Biol.* **1985**, *5*, 457.
(20) Koch, K. S.; Gleiberman, A. S.; Aoki, T.; Leffert, H. L.; Feren, A.; Jones, A. L.; Fodor, E. J. *Nucleic Acids Res.* **1995**, *23*, 1098.
(21) Skripkin, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4945.
(22) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **2000**, *300*, 1113.
(23) Spackovà, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.* **2001**, *123*, 3295.
(24) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350.
(25) Ducéré, J. M.; Cavallo, L. *J. Phys. Chem. B* **2007**, *111*, 13124.
(26) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.
(27) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
(28) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
(29) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. *J. Comput. Chem.* **1997**, *18*, 1463.
(30) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(31) Cheatham III, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.
(32) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(33) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(34) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.
(35) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
(36) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359.
(37) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799.
(38) Kristyán, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.
(39) Pérez-Jordá, J.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134.
(40) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
(41) Oliva, R.; Tramontano, A.; Cavallo, L. *RNA* **2007**, *13*, 1427.
(42) Jurecka, P.; Sponer, J.; Cerný, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
(43) Spackova, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.* **1999**, *121*, 5519.
(44) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.
(45) Ṕ; erez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.
(46) Kettani, A.; Gorin, A.; Majumdar, A.; Hermann, T.; Skripkin, E.; Zhao, H.; Jones, R.; Patel, D. J. *J. Mol. Biol.* **2000**, *297*, 627.
(47) Liu, H.; Kugimiya, A.; Matsugami, A.; Katahira, M.; Uesugi, S. *Nucleic Acids Res. Suppl.* **2002**, 177.

JP804036J