DNA Structures

DOI: 10.1002/anie.201206522

The G-Triplex DNA**

Vittorio Limongelli, Stefano De Tito, Linda Cerofolini, Marco Fragai, Bruno Pagano, Roberta Trotta, Sandro Cosconati, Luciana Marinelli, Ettore Novellino, Ivano Bertini, Antonio Randazzo, Claudio Luchinat,* and Michele Parrinello**

In memory of Ivano Bertini

Nucleic acids represent the alphabet of the cellular language and through their sequence and topology regulate vital cellular functions. In recent years, it has been found that many variations from the Watson–Crick duplex structure[1] play key roles in many cellular processes. Examples are hairpins,[2] cruciforms,[3] parallel-stranded duplexes,[4] triplexes,[5] G-quadruplexes,[6] and the i-motif.[7] These structures can be formed by nucleotide sequences distributed throughout the whole human genome, their location is not random and often associated with human diseases.[8] These complexes are formed from one to four strands, stabilized by base stacking and hydrogen bond interactions, with a variety of non-standard pairings. For instance, DNA triplexes can present G:G-C, A:A-T, C^+:G-C, and T:A-T pairings, with two strands in the standard Watson–Crick duplex structure (i.e. G-C and A-T) and the third one lying in the major groove of the duplex. In contrast, G-quadruplexes are four-stranded structures stabilized by stacking of two or more guanine tetrads (Figure 1).

These examples highlight the structural polymorphism of DNA and suggest that other structures might exist, perhaps with specific cellular functions that are, to date, unknown. Herein, using metadynamics simulations,[9] we have identified a stable folding intermediate of the thrombin binding aptamer (TBA) quadruplex.[10] This intermediate is characterized by a "G-triplex" structure, having G:G:G triad planes stabilized by an array of Hoogsteen-like hydrogen-bonds (Figure 1). This kind of structure has been already hypothesized in other investigations on different DNA sequences,[11] but never experimentally proven. Herein, for the first time, we have structurally and thermodynamically characterized this DNA structural motif, through a combination of biophysical experiments.

Well-tempered metadynamics[9b] simulations have been used to study the folding of TBA, which is a 15-mer oligonucleotide (5′-dGGTTGGTGTGGTTGG-3′) organized in an anti-parallel monomolecular G-quadruplex with a chair-like structure (Figure 2a). This structure consists of two G-tetrads, able to coordinate a metal ion at the center, connected by two TT loops and a single TGT loop.

Metadynamics accelerates the sampling, adding a bias on a few degrees of freedom of the system, called collective variables (CVs). In such a way, long time scale events, such as ligand/protein docking[12] or protein/DNA folding, can be sampled in an affordable computational time and the free energy surface (FES) of the process can be computed. In the present case, the FES was calculated as a function of two CVs, the radius of gyration CV defined by the oxygen atoms of the guanines forming the G-tetrads and a second CV that counts the number of hydrogen bonds between these guanines (see Supporting Information). Looking at the FES obtained after approximately 80 ns of metadynamics simulation, three main energy minima can be identified (Figure 2b). The deepest one, basin A, corresponds to the experimental G-quadruplex structure of TBA.[13] In the second minimum, basin B, TBA shows a partial opening of the 3′ end with residue G15

[*] Dr. V. Limongelli,[+] S. De Tito,[+] Dr. B. Pagano, Dr. R. Trotta, Dr. S. Cosconati, Dr. L. Marinelli, Prof. E. Novellino, Prof. A. Randazzo
Department of Pharmacy,
University of Naples "Federico II"
Via D. Montesano, 49, 80131 Naples (Italy)
E-mail: antonio.randazzo@unina.it

Dr. V. Limongelli,[+] Prof. M. Parrinello
Department of Chemistry and Applied Biosciences, ETH Zurich
and
Facoltà di Informatica
Istituto di Scienze Computazionali (ICS)
Università della Svizzera Italiana
Via Giuseppe Buffi 13, 6900 Lugano (Switzerland)
E-mail: parrinello@phys.chem.ethz.ch

Dr. L. Cerofolini, Dr. M. Fragai, Prof. I. Bertini, Prof. C. Luchinat
Magnetic Resonance Center (CERM), University of Florence
50019 Sesto Fiorentino, Florence (Italy)
and
Department of Chemistry, University of Florence
50019 Sesto Fiorentino, Florence (Italy)
E-mail: luchinat@cerm.unifi.it

[+] These authors contributed equally to this work.

[**] We acknowledge Anh Tuân Phan for providing an optimized JR HMBC sequence, Roman Reshetnikov and Janez Plavec for useful discussions. Furthermore, we acknowledge that the results of this research have been achieved using the PRACE Research Infrastructure resource JUGENE based in Germany at Forschungszentrum Juelich. This work was supported by the Gabriele Charitable Foundation, Italian Institute of Technology (IIT), Italian M.U.R.S.T., P.R.I.N. 2009, Biologia Strutturale Meccanicistica, Italian Association for Cancer Research (A.I.R.C.), EU FP7, ERC Advanced Grant No. 247075, and a grant from the Swiss National Supercomputing Centre (CSCS) under project ID s386.

Supporting information for this article (experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201206522.

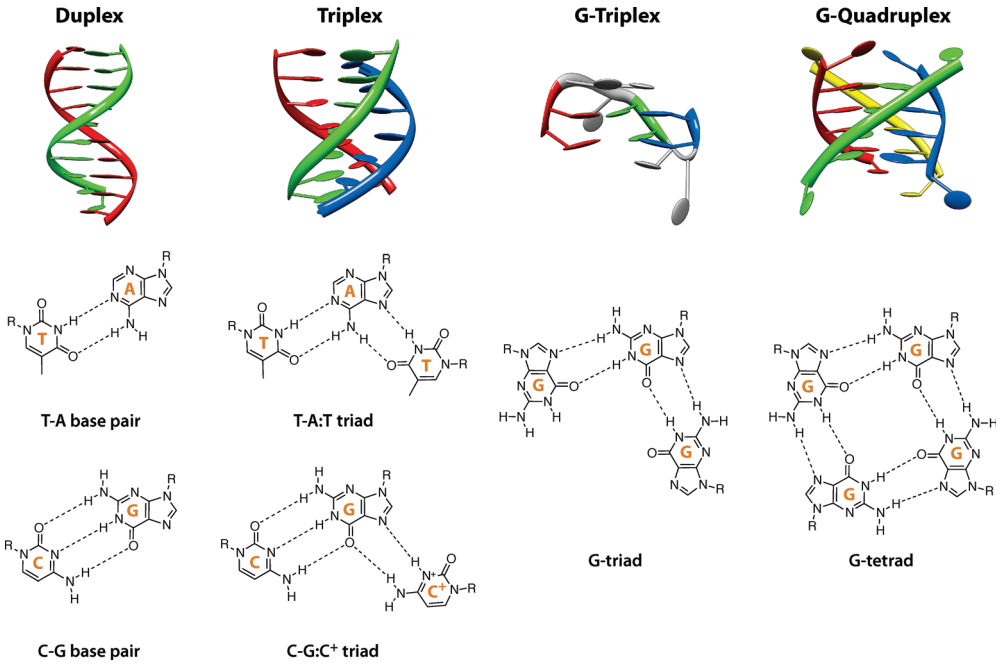

Figure 1. DNA structural motifs. Top: Schematic illustration of the duplex, triplex, G-triplex, and G-quadruplex structures. Bottom: Examples of base pairing: T-A and C-G for the duplex, T-A:T and C-G:C^+ for the triplex (where C-G and T-A form the standard Watson–Crick duplex structure, whereas the colon signs indicate the pairing with the third strand, which lies in the major groove of the duplex), G-triad for the G-triplex, and G-tetrad for the G-quadruplex.

oriented towards the solvent. In this basin G14 moves slightly from its original position, with the oxygen of the base pointing towards the center of the planes formed by G1:G6:G10 and G2:G5:G11 (Figure 2b). In such a way, G14 conserves the hydrogen bond interactions with G2 and G11 and at the same time it partially fills the vacancy in the coordination shell of the metal present in the core. This step can be considered the very first event in the unfolding process of TBA. In basin C, the 3′ end opens completely with G14 and G15 leaving the G-tetrad planes and pointing towards the solvent (Figure 2b). This minimum is approximately 6.5 kcal mol^{-1} higher in energy than basin A and here TBA assumes a number of different conformations owing to the conformational flexibility of the 3′ overhang formed by T13-G14-G15. In fact, these bases rearrange to form a single helix that assumes a number of conformations, all of them stabilized by stacking interactions. While the 3′ overhang is flexible, the rest of TBA is rather stable forming two G:G:G planes, namely G-triads, composed by G1:G6:G10 and G2:G5:G11, that form an array of Hoogsteen-like hydrogen bonds (Figure 1 and 2b). In this conformation, the metal ion is coordinated at the center of the two triads in a way similar to that of the G-quadruplex structure. This structure, named "G-triplex", differs from the known triplex structures not only for the base pairing but also for the structure (Figure 1).

It can be observed that during the metadynamics simulations the 3′ end of TBA opens and closes several times, passing from basin A to basin B and basin C, and then folded again in the G-quadruplex structure, basin A (see Figure 2c and movie in the Supporting Information). Thanks to these recrossing events, the calculated FES is accurate and quantitatively well characterized. The stability of the TBA conformations in the three energy minima A, B, and C, has been further assessed through molecular dynamics simulations (see Supporting Information).

To establish the intrinsic stability of the G-triplex, we removed from the 3′ end of TBA, the last four residues that are highly mobile in basin C. The resulting structure was found stable in an ordinary MD simulation that lasted over 100 ns (see Supporting Information). Its coordinates are provided in the Supporting Information.

This prediction of a stable G-triplex structure needed to be validated by experiments. Thus, we performed a number of experiments on the same truncated form of TBA (5′-GGTTGGTGTGG-3′, **I**). First, **I** was studied using ^{1}H NMR spectroscopy at 25 °C showing the presence of a predominant well-defined hydrogen-bonded structure in solution. This was supported by the presence of four well-defined exchangeable proton signals in the 11.0–12.5 ppm region of the 1D ^{1}H NMR spectrum (Figure 3a). These signals are typical of DNA structures with Hoogsteen hydrogen bonds.[14] On the other hand, the region of aromatic protons (6.5–8.5 ppm) is characterized by the presence of eleven intense signals that can be attributed to the seven guanine H8 and four thymine H6 protons, and by a number of minor signals (Figure 3a). The latter might be due to the presence of unstructured DNA in equilibrium with the structured one. This equilibrium is very sensitive to temperature with the structured form highly favored at low temperature, as shown by the spectrum obtained at 1 °C (Figure 3b). The structure adopted by **I** turned out to be very stable over the course of weeks.

The non-exchangeable base and sugar protons of **I** were assigned through the analysis of 2D NOESY, 2D TOCSY, and 2D COSY NMR spectra (see Table S2 and Supporting Information). Interestingly, the presence of three intense cross-peaks between the H8 proton bases and sugar H1′ resonances for residues G1, G5, and G10 in the NOESY spectrum (900 MHz, $T = 1$ °C, mixing time 100 ms), along with the presence of weak cross-peaks between the same aromatic protons and the H2′ and H2″ protons of their own ribose moiety, indicated that these three residues adopt a *syn* glycosidic angle conformation (Figure 3c). On the contrary, G2, G6, G8, and G11 turned out to have an *anti* glycosidic

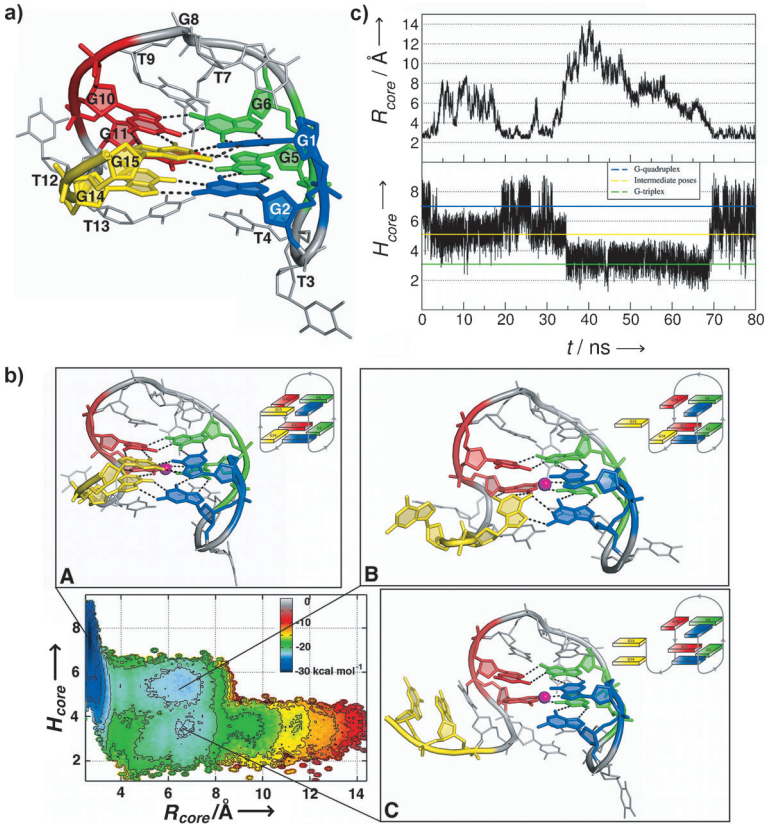

Figure 2. The 3′ end opening of TBA. a) 3D representation of the NMR structure of TBA (PDB ID: 1qdf).[13] b) The free energy surface (FES) of the 3′ end opening of TBA shows three main energy minima: one deep and narrow, basin A, which represents TBA in the G-quadruplex structure; the second one, basin B, which represents an intermediate state; and the last one, basin C, which shows the G-triplex structure formed by the triads G1:G6:G10 and G2:G5:G11. Residue labels are the same as in (a). c) Plots showing the phase space, represented as the H_{core} and R_{core} CV, explored during the metadynamics simulation.

conformation, with the intensity of the cross-peaks mentioned before inverted (Figure 3c). The three H8 peaks of *syn* G residues are downfield shifted with respect to those of the *anti* ones, exactly as reported for TBA.[15] Furthermore, the three *anti*-Gs (G2, G6, and G11) show H8/H2′-H2″ sequential connectivities with the 5′ neighboring *syn*-Gs (G1, G5, and G10). This suggests the presence of the tracks *G1*-G2, *G5*-G6, *G10*-G11 (italic residues have a *syn* conformation) and the formation of a helical structure. The presence of unusual NOE connectivities between a number of Gs and Ts indicates that 5′-TG-3′ and 5′-GT-3′ do not adopt a helical winding, and that the TT and TGT tracts realistically form loops. Finally, the alternation of *syn* and *anti* G residues implies that **I**, as TBA does, folds into an antiparallel structure.

The assignment of the exchangeable imino protons was instead obtained by JR-HMBC[16] correlation experiment (see Figure 4a and Supporting Information). Once the imino protons were assigned, we could determine the folding topology of **I** using the JR-NOESY experiment. This allows us to identify NOE interactions representing the interaction between the imino proton of one base and the H8 proton of the other. Generally, these NOE interactions are diagnostic of the presence of Hoogsteen base pairings. In our case, we observed NOE interactions between G11-NH (12.08 ppm)/G5-H8 (7.44 ppm) and G5-NH (12.23 ppm)/G2-H8 (8.04 ppm). These two correlations indicate that, as in TBA, G5 is involved in the formation of Hoogsteen hydrogen-bonds with both G11 and G2, and this is in agreement with the calculated structure (Figure 4b,c). We have also observed an NOE between G1-NH (12.00 ppm) and G6-H8 (8.26 ppm), indicating that G1 and G6 are also paired. All of this indicates that the structure of **I** could be characterized also by a second G-triad formed by G10, G6, and G1. Because of unfavorable T1 noise, no NOE between G6-NH and G10-H8 could be unambiguously detected. Nevertheless, the fact that the subunit G10-G11 adopts a helical winding strongly supports the idea that G10 also takes part in the second G-triad. The formation of the G-triplex is also supported by the presence of further NOE interactions between NH protons and other exchangeable protons identified by ^{1}H-^{15}N-HSQC experiments (see Supporting Information). In particular, we observed strong NOE interactions between G1-NH/G5-NH and G6-NH/G11-NH, medium intensity NOE interactions between G1-NH/G6-NH and G5-NH/G11-NH, and weak NOE interactions between G5-NH/G6-NH, which also support the formation of the G-triplex, having two G-triads: G1-G6-G10 and G2-G5-G11, characterized by a *syn-anti-syn* and *anti-syn-anti* arrangement of the residues, respectively. Finally, a number of other NOE interactions definitively confirmed the structure of the G-triplex (see Supporting Information).

The thermodynamic stability of **I** was investigated by circular dichroism (CD) and differential scanning calorimetry (DSC) experiments. The CD spectrum of **I** shows two positive bands at 289 nm and 253 nm, and two negative at 235 nm and 265 nm, indicative of a homopolar stacking of the nucleobases,[17] in agreement with the proposed G-triplex structure (see Supporting Information). This spectrum resembles that of TBA (Figure 5a). However, both positive and negative bands are slightly shifted, thus suggesting that the stacking of the bases is not identical to TBA. The CD melting profile of **I**, recorded at the wavelength of maximum absorbance variation upon folding ($\lambda = 289$ nm), shows almost superimposable heating and cooling curves (Figure 5b), therefore the melting profile was reversible, with no significant hysteresis. From these measurements, a melting temperature, T_m, of 33.5(±1.0)°C and a van't Hoff enthalpy change, $\Delta H°_{vH}$, of 145(±15) kJ mol^{-1} are derived (see Supporting Information).

DSC experiments were carried out to characterize the denaturation thermodynamics of **I** with a model-independent analysis method.[18] DSC thermograms for denaturation of **I** (Figure 5c) were obtained at two different heating rates, 0.5 and 1.0 °C min^{-1}. The different heating rate does not alter the

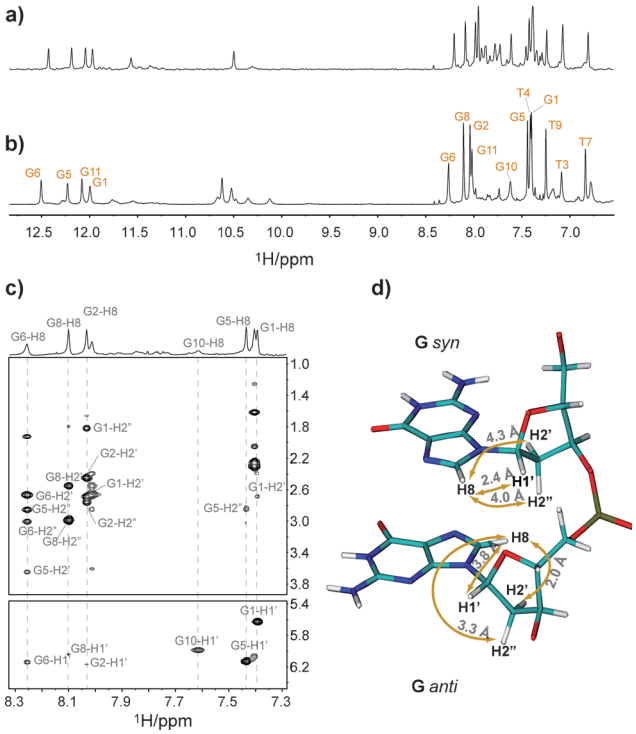

Figure 3. 1D ^{1}H NMR and 2D NOESY spectra of 5′-GGTTGGTGTGG-3′ (**I**; 70 mM KCl, 10 mM KH_2PO_4, 0.2 mM EDTA, pH 7). Imino, amino, and aromatic regions of the 1D ^{1}H NMR spectra of **I** acquired at 25 °C (a) and 1 °C (b). In (b), the signal at 7.63 ppm, attributed to G10-H8, is slightly broad, suggesting that its conformation can vary on the NMR timescale. c) Two expanded regions of the NOESY spectrum (900 MHz, T = 1 °C, mixing time 100 ms). d) Distances of the correlated protons in the NOESY spectrum.

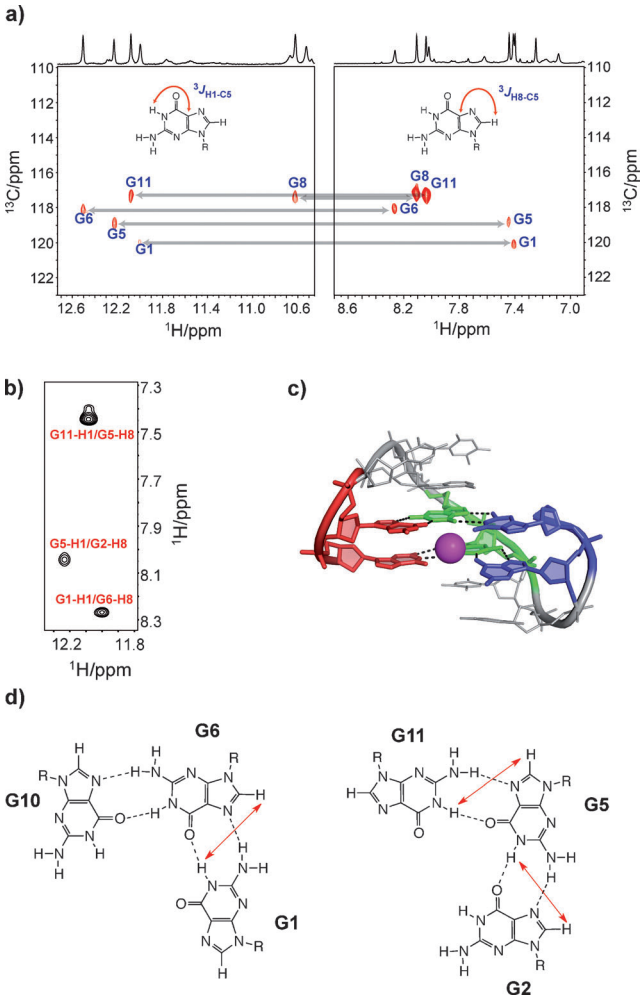

Figure 4. JR-HMBC and JR-NOESY spectra of **I** (70 mM KCl, 10 mM KH_2PO_4, 0.2 mM EDTA, pH 7). a) Expanded regions of the JR-HMBC spectra (600 MHz, T = 1 °C) showing correlations between H8 and H1 guanine protons with C5 carbon atoms. b) Expanded region of the 2D JR-NOESY (900 MHz, T = 1 °C) showing correlations between H8 and H1 protons of paired bases. c) 3D representation of the G-triplex structure adopted by **I** (Supporting Information). d) G-triads involved in the formation of the G-triplex. Red arrows show NOE correlations between H8 and H1 protons of paired bases.

thermodynamic parameters significantly, thereby demonstrating that the investigated process is not kinetically controlled.[19] Furthermore, the unfolding of **I** is a highly reversible process, since the original signal is recovered by rescanning the same sample. The DSC curves show a symmetric shape with a maximum at T_m of 34.0(±0.5)°C, in good agreement with that obtained by CD. The melting temperature was almost concentration-independent, consistent with a structure resulting from unimolecular folding. The integra-

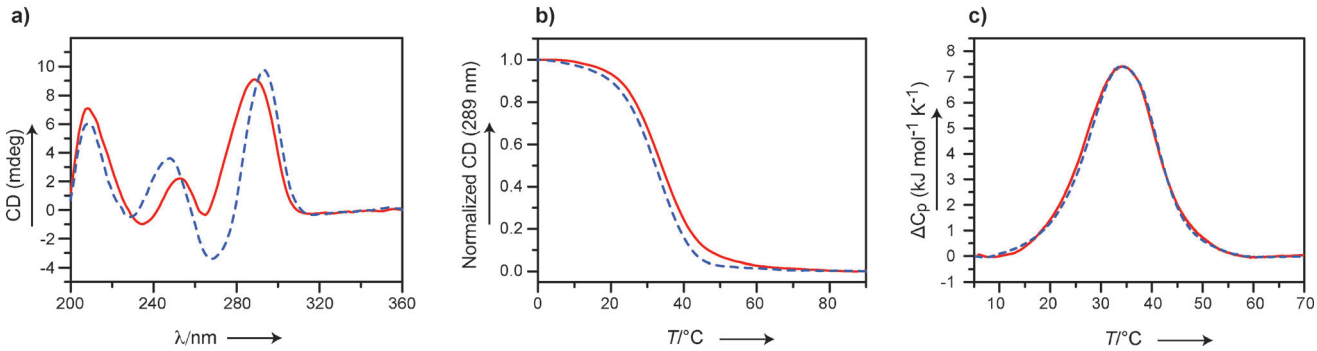

Figure 5. Biophysical characterization of **I**. a) CD spectra of **I** (solid line) and TBA (dashed line) at 1 °C. b) Normalized CD melting curves of **I** recorded at 289 nm at a scan rate of 0.5 °C min^{-1}. Heating and cooling curves are shown with solid and dashed lines, respectively. c) DSC profiles for **I** at 0.5 °C min^{-1} (solid line) and 1.0 °C min^{-1} (dashed line) heating rate. All experiments were performed in a buffer solution containing 10 mM potassium phosphate, 70 mM KCl, and 0.2 mM EDTA (pH 7.0).

tion of the denaturation peak gives a $\Delta H°_{cal}$ of 135(±5) kJ mol^{-1}, which is almost identical to the van't Hoff enthalpy calculated from DSC curves and very close to that calculated from CD. This indicates that the transition of **I** is a two-state process. Finally, the calculated Gibbs energy value, $\Delta G°$, at 298 K is 4(±1) kJ mol^{-1}, and results from the compensation of the favorable enthalpy term with an unfavorable entropy contribution ($\Delta S° = 0.44$-(±0.02) kJ mol^{-1} K^{-1}). The whole set of thermodynamic parameters show that, as expected, the structure of **I** is less stable than that of TBA.[20] In particular, the enthalpy term is lower than the one derived for TBA, probably owing to the greater contribution resulting from the stronger base stacking and larger number of hydrogen bonds involved in the quadruplex structure. This is consistent with the lack of G-tetrads in the structure adopted by **I**.

In summary, during the folding process of the G-quadruplex aptamer TBA, we have observed the formation of the G-triplex structural motif. The existence of this structure has been proven in an 11-mer oligonucleotide, whose structural and thermodynamic properties have been characterized. At variance with the already known triplex structures, the G-triplex presents G:G:G triad planes stabilized by an array of Hoogsteen-like hydrogen bonds. Although this kind of structure was already hypothesized as an intermediate in the folding process of other quadruplex forming sequences,[11] this is the first time that DNA has been unambiguously isolated and structurally characterized in this conformation.

G-rich regions, potentially able to form G-triplex structures, are very abundant in the genome and our study paves the way for further investigations into the presence of these structures in vivo, their biological role, and ways of interacting with them.

Received: August 13, 2012
Published online: January 17, 2013

Keywords: metadynamics · NMR spectroscopy · quadruplex DNA · thermodynamics · triplex DNA

[1] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737 – 738.
[2] G. Varani, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379 – 404.
[3] D. M. J. Lilley, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9513 – 9515.
[4] J. H. van de Sande, N. B. Rmasing, M. W. Germann, W. Elhorst, B. W. Kalish, E. von Kitzing, R. T. Pon, R. C. Clegg, T. M. Jovin, *Science* **1988**, *241*, 551 – 557.
[5] V. Sklená, J. Felgon, *Nature* **1990**, *345*, 836 – 838.
[6] G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876 – 880.
[7] M. Guéron, J. L. Leroy, *Curr. Opin. Struct. Biol.* **2000**, *10*, 326 – 331.
[8] R. D. Wells, *Trends Biochem. Sci.* **2007**, *32*, 271 – 278.
[9] a) A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562 – 12566; b) A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 020603.
[10] O. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas, J. J. Toole, *Nature* **1992**, *355*, 564 – 566.
[11] a) R. D. Gray, R. Buscaglia, J. B. Chaires, *J. Am. Chem. Soc.* **2012**, *134*, 16834 – 16844; b) M. Boncina, J. Lah, I. Prislan, G. Vesnaver, *J. Am. Chem. Soc.* **2012**, *134*, 9657 – 9663; c) T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 14910 – 14918; d) R. Stefl, T. E. Cheatham III, N. Spacková, E. Fadrná, I. Berger, J. Koca, J. Sponer, *Biophys. J.* **2003**, *85*, 1787 – 1804.
[12] a) V. Limongelli, L. Marinelli, S. Cosconati, C. La Motta, S. Sartini, L. Mugnaini, F. Da Settimo, E. Novellino, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1467 – 1472; b) G. Grazioso, V. Limongelli, D. Branduardi, E. Novellino, C. De Micheli, A. Cavalli, M. Parrinello, *J. Am. Chem. Soc.* **2012**, *134*, 453 – 463; c) V. Limongelli, M. Bonomi, L. Marinelli, F. L. Gervasio, A. Cavalli, E. Novellino, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5411 – 5416; d) D. Heckmann, B. Laufer, L. Marinelli, V. Limongelli, E. Novellino, G. Zahn, R. Stragies, H. Kessler, *Angew. Chem.* **2009**, *121*, 4501 – 4506; *Angew. Chem. Int. Ed.* **2009**, *48*, 4436 – 4440.
[13] V. M. Marathias, K. Y. Wang, S. Kumar, T. Q. Pham, S. Swaminathan, P. H. Bolton, *J. Mol. Biol.* **1996**, *260*, 378 – 394.
[14] J. Feigon, F. W. Smith, *Nature* **1992**, *356*, 164 – 168.
[15] P. Schultze, R. F. Macaya, J. Feigon, *J. Mol. Biol.* **1994**, *235*, 1532 – 1547.
[16] A. T. Phan, *J. Biomol. NMR* **2000**, *16*, 175 – 178.
[17] a) S. Masiero, R. Trotta, S. Pieraccini, S. De Tito, R. Perone, A. Randazzo, G. P. Spada, *Org. Biomol. Chem.* **2010**, *8*, 2683 – 2692; b) A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. Webba da Silva, *Angew. Chem.* **2011**, *123*, 10833 – 10836; *Angew. Chem. Int. Ed.* **2011**, *50*, 10645 – 10648; c) A. Randazzo, G. P. Spada, M. Webba da Silva, *Top. Curr. Chem.* **2012**, DOI: 10.1007/128_2012_331.
[18] G. E. Plum, K. J. Breslauer, *Curr. Opin. Struct. Biol.* **1995**, *5*, 682 – 690.
[19] L. Petraccone, B. Pagano, V. Esposito, A. Randazzo, G. Piccialli, G. Barone, C. A. Mattia, C. Giancola, *J. Am. Chem. Soc.* **2005**, *127*, 16215 – 16223.
[20] I. Smirnov, R. H. Shafer, *Biochemistry* **2000**, *39*, 1462 – 1468.

