View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Digital.CSIC

"Specific loop modifications of the thrombin binding aptamer trigger the formation of parallel structures" Aviñó, A., Portella, G., Ferreira, R., Gargallo, R., Mazzini, S., Gabélica, V., Orozco, M., Eritja, R. The FEBS J., 281(4), 1085-1099 (2014). PMID: 24304855, doi: 10.1111/febs.12670

SPECIFIC LOOP MODIFICATIONS OF THE THROMBIN BINDING APTAMER TRIGGER THE FORMATION OF PARALLEL STRUCTURES

Anna Aviñó[1,#],Guillem Portella[2,3,#], Ruben Ferreira[1], Raimundo Gargallo[5], Stefania Mazzini[6], Valerie Gabélica[7], Modesto Orozco,[2,3,4*]and Ramon Eritja[1*]

[1]Institute for Advanced Chemistry of Catalonia (IQAC), CSIC, Networking Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Jordi Girona 18-26 08034 Barcelona, Spain.

[2]Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10, 08028 Barcelona, Spain

[3]Joint IRB-BSC Program in Computational Biology, Barcelona, Spain

[4]Departament de Bioquímica i Biologia Molecular, Facultat de Biologia, University of Barcelona, 08028 Barcelona, Spain

[5]Department of Analytical Chemistry, University of Barcelona, 08028 Barcelona, Spain

[6]Department of Food, Environmental and Nutritional Sciences (DEFENS), Section of Chemical and Biomolecular Sciences, University of Milan, Via Celoria 2, 20133 Milan, Italy

[7]Physical Chemistry and Mass Spectrometry Laboratory, Department of Chemistry, University of Liège, Belgium. (Present addresses: (1) Univ. Bordeaux, IECB - ARNA laboratory, F-33600 Pessac, France, (2) INSERM, U869 - ARNA laboratory, F-33000 Bordeaux, France.

* To whom correspondence should be addressed. Tel: 34-93-4006145; Fax: 34-93-2045904; Email: recgma@cid.csic.es. Correspondence may also be addressed to Modesto Orozco; Tel 34-93-4039942; Email: modesto.orozco@irbbarcelona.org

[#]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors

Running Title: loop modified TBAs

Keywords: G-quadruplex, thrombin binding aptamer, DNA polymorphism, nucleic acid structure, thermal stability

Abbreviations list: Ac: acetate; CCS: collision cross section; CD: circular dichroism; DC: self-diffusion coefficients; DSS: 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt; DOSY: diffusion ordered spectroscopy; EHSS: exact hard-spheres scattering; ESI-MS: electrospray ionization mass spectrometry; FID: free induction decay; IMS: ion mobility spectroscopy; HPLC: high performance liquid chromatography; LNA: locked nucleic acids ; MD: molecular dynamics; MeOH: methanol; NMR: nuclear magnetic resonance; NOE: nuclear Overhauser effect; NOESY: nuclear Overhauser effect

spectroscopy; PAGE: polyacrylamide gel electrophoresis; Pur: purine; Pyr: pyrimidine; TBA: thrombin binding aptamer; Tm: meting temperature; TOCSY: total correlation spectroscopy; RMSD: root-mean-square deviation; UNA: unlocked nucleic acids; UV: ultraviolet.

ABSTRACT

Guanine-rich sequences display a large structural variability with folds ranging from duplex to triplex and quadruplex helices. Quadruplexes are polymorphic and can display multiple stoichiometries, parallel and antiparallel strand alignments, with different topological arrangements. We analyze here the equilibrium between intramolecular antiparallel and intermolecular parallel G-quadruplexes in the thrombin binding aptamer (TBA) sequence. Our theoretical and experimental studies demonstrate that an apparently simple modification at the loops of TBA induces a large change in the monomeric antiparallel structure of TBA to yield a parallel G-quadruplex exhibiting a novel T-tetrad. Present results illustrate the extreme polymorphism of G-quadruplexes and the ease to manipulate their conformation in solution by nucleotide modification.

INTRODUCTION

DNA is an extremely polymorphic molecule able to adopt a multitude of helical structures [1-3]. Guanine-rich oligonucleotides lead to especially polymorphic nucleic acids, which can adopt both A and B type duplexes, different types of triplexes and quadruplexes [1-3]. G-quadruplexes are very stable structures generated by the association of steps of four guanines (G-quartets) held together by eight hydrogen bonds. These structures are stabilized by intra and inter-molecular stacking and strong electrostatic interactions with cations allocated within the structure [4,5]. One of the most intriguing characteristics of G-quadruplexes is their structural variability [6-9] depending on the length of the polyG track, the loops that connect the quartets, the environment and the *syn*/*anti* preference of the purines in the quartets, the G-quadruplex can appear in both “antiparallel” and “parallel” arrangements. Within the first family of structures, the quadruplex is characterized by opposite neighboring strand orientations and different conformations around the glycosidic bond of the guanosines in the G-quartets. In contrast, in the parallel quadruplex all the strands have the same direction and all the guanines are in *anti* conformation. In addition, hybrid structures formed combining both arrangements have been also described as well as multimeric G-quadruplex formed by association of several strands [4,5]. It is important to shed light on the factors that govern the specific folding and stability adopted by quadruplexes of a given sequence. Although considerable effort is being made in this regard, no general rules have been established to date.

G-quadruplexes are structures of large biological relevance, which contribute to the stabilization of telomers [10,11], and accordingly to the preservation of chromosomal integrity. Furthermore, aromatic molecules, able to bind to G-quadruplexes are excellent inhibitors of telomerase, a major target in cancer [3,4,12]. G-quadruplexes acting as aptamers [13] can be powerful inhibitors of different proteins, such as HIV integrase [14] or thrombin [15,16], with various applications in diagnostics and

therapeutics. Particularly, large research efforts have been focussed in the 15-base long thrombin binding aptamer (TBA) sequence, which was derived in experiments as an inhibitor of thrombin, a key enzyme in the blood clotting cascade [15]. TBA forms in aqueous solution an intramolecular, antiparallel G-quadruplex with two stacked G-quartets connected by a central TGT, and two lateral TT loops, defining a chair-like conformation [17,18]. The adopted structure is identical in potassium and in ammonium ions [19]. Guanines in the two tetrads display an alternating *syn/anti* conformation (*syn*-dG at positions dG1, dG5, dG10 and dG14, *anti*-dG at positions dG2, dG6, dG11 and dG15) and are stabilized by the presence of small monovalent cations in the central channel [19]. The stoichiometry and interaction motif between thrombin and TBA have been controversial for several years [18,20,21]. It is currently assumed that the two TT loops interact with the thrombin anion exosite I in a 1:1 stoichiometry [23]. The loops connecting G-quartets have a strong influence on the folding and stability of intramolecular quadruplexes. Different loop lengths and sequences can either stabilize or destabilize G-quadruplex [23-28]. For example, reduction of the two lateral TT loops of TBA to a single T caused complete disruption of the quadruplex structure, while the extension to TTT does not alter stability [22]. Significant changes in the stability have also been found when thymidines in the lateral loops have been changed to other coding nucleotides [29,30], or to derivatives with modified backbones [31-35]. For example, the replacement of several thymidines at the loop positions by 2'-F-arabino derivatives leads always to more stable G-quadruplex [31], while the substitution by LNA (Locked Nucleic Acids), UNA (Unlocked Nucleic Acids), and acyclic thymidines showed a position dependent effect on TBA stability [32-35].

In this paper we systematically analysed the effect of (coding) nucleotide substitution in the two lateral loops of TBA. The most remarkable result is the double T→G substitution that leads to a dramatic change in the topology of the quadruplex that moves from an intramolecular antiparallel arrangement to an intermolecular parallel one. A wide repertoire of experiments and theoretical calculations have allowed us to characterize and rationalize the nature of this conformational change induced by loop changes, which up to date was believed to occur only upon changes in the solvent or in the conformational preferences of nucleotides in the central tetrads [36-40]. The unexpected results reported here illustrate the extreme conformational promiscuity of G-DNA quadruplexes and the possibility to manipulate its conformational landscape by simple mutations.

RESULTS AND DISCUSSION

Thermal UV and Circular Dichroism studies of TBA derivatives.

We have synthesized up to 7 variants of TBA to exhaustively explore the effect of nucleotide modification in the lateral loops (Table 1).

We used circular dichroism (CD) spectroscopy to characterize the nature of the structures of the different TBA derivatives. TBA and most of the derivatives showed clear antiparallel quadruplex spectra (see CD spectra in KCl/ cacodylate buffer in Figure 1A), with positive bands around 250 and

295 nm and a negative band around 270 nm, while TBA derivatives containing two or more guanines at the lateral loops (TBA2G[4,13], TBA2G[3,12] and TBA4G[3,4,12,13]) displayed very different CD spectra, which match with those expected for a parallel G-quadruplex (note the positive band at 265 nm and the negative one at 245 nm).

As described elsewhere [11, 29-35, 38], TBA forms a very stable structure with a melting temperature around 50 degrees in potassium buffer (monitored by UV absorbance, Table 2). Melting profiles of TBA are reversible without hysteresis (Figure S1), suggesting that melting is tracing a single and fast intramolecular process. Changes of thymine by the other pyrimidine, cytosine (TBA2C[4,13]) at positions 4 and 13 have little impact on the stability of the structure (Table 2 and Figure S1). On the contrary, the substitution of thymine by adenine (TBA2A[4,13]) induces a large destabilization of the TBA structure, which is again fully reversible, suggesting that we are still analyzing the same simple intramolecular folding/unfolding process found for TBA.

Results above strongly suggest that in the lateral loop there are no clear preferences between pyrimidines, but the presence of a purine is not very well tolerated, probably due to its larger size. However, experiments in TBA2G[4,13] in which two thymines are substituted by two guanines show a completely different situation. On the one hand, this substitution increases the thermal stability (ΔTm= +10 degrees; see Table 2). On the other hand, melting profiles show a clear hysteresis (Figure S1), suggesting that the TBA→TBA2G[4,13] (or the TBA→TBA2G[3,12]) substitutions induce a severe structural change in the quadruplex, resulting in a structure with a slower folding kinetics. To further analyze these results we synthesized and performed melting experiments with a derivative where only one thymine was substituted by one guanine (TBA1G[4]), and another where all thymines in lateral loops were substituted by guanines (TBA4G [3,4,12,13]). The first derivative showed the behavior of a Pyr→Pur transversion, i.e. a destabilization of the structure, but maintenance of the melting profile. On the contrary, the second derivative leads to a dramatic increase in the melting temperature, the resulting structure being so stable that structural transition cannot be observed (see Table 2 and Figure S1).

It is clear that melting experiments suggest that TBA structure is quite permissive to Pyr transitions in the lateral loops, while Pyr→Pur transversions destabilize the TBA structure, except in the case of two or four guanine substitutions. In that case, a more stable and complex structure, with a slower folding kinetics was formed. Melting experiments at different oligonucleotide concentrations strongly suggest that two guanine substitutions altered the overall molecularity of the quadruplex, as the Tm increases with the concentration only for the TBA2G derivatives (Figure S2). In addition, thermal stability of TBA2G[4,13] derivative also increased with the K^+ concentration (Table S4). This could be explained by the improvement of the stacking interactions between G-tetrads.

Thermal denaturation and circular dichroism were also analysed in ammonium conditions. Stabilities of the modified TBAs are shown in Table S5. Tm values were lower than in potassium buffer but the Tm differences (ΔTm) relative to unmodified TBA were the same in both buffer conditions. The CD profile of the TBAs in ammonium buffer differed from those in the potassium buffer (Figure 1B).

Specifically, TBA2G[4,13] showed CD spectra with two positive bands, a higher band at ~ 260 nm corresponding to a parallel quadruplex structure and a lower band at ~ 295 nm corresponding to an antiparallel structure. This fact agrees with a partial conversion to the antiparallel state. Note that such a change is not detected in the ultra-stable TBA4G[3,4,12,13] parallel quadruplex, where the antiparallel/parallel stability difference seems to be very large, therefore precluding any ion-dependent polymorphism [17,19]. In addition, TBA2A[4,13], which was similar to TBA when dissolved in potassium buffer, had CD spectra showing a mixture of parallel and antiparallel structure when in ammonium buffer. The rest of the TBA derivatives, TBA, TBA2G[3,12], TBA1G[4], TBA2C[4,13] and TBA4G[3,4,12,13] had similar CD spectra in ammonium buffer to those in potassium buffer.

TBA and TBA2G[4,13] were also examined using thermal denaturation and circular dichroism under dehydrating agent (MeOH), which is known to favor compact structures (such as the parallel quadruplex) [41,42]. Thermal stability of TBA2G[4,13] was higher than TBA and increased with the percentage of MeOH in both compounds (Table S6). The addition of MeOH to the TBA in ammonium buffer increased the amplitude of the CD bands corresponding to antiparallel quadruplex structures (Figure 1C). This finding has been also described using PEG 200 [41]. More importantly, the addition of MeOH to TBA2G[4,13], that in the absence of MeOH showed a mixture of parallel and antiparallel conformations, caused the complete formation of parallel structure (Figure 1D).

Melting experiments and CD spectroscopy clearly suggest that addition of two or more guanines to the lateral loops of TBA induces a transition from monomeric antiparallel to multimeric parallel G-quadruplex. In addition, stabilities derived from thermodynamic parameters using a simple folded and unfolded state were obtained in order to compare all the derivatives (Table 2). Results from TBA2G derivatives were approximate and should be interpreted cautiously, as minor multimeric species have been also observed (see below).

NMR studies of TBA derivatives

NMR measurements of derivatives TBA, TBA2C[4,13] and TBA2A[4,13] give rise to similar spectra (Figure 2), which fit well into the known characteristics of antiparallel quadruplexes [25, 38]. A temperature increase leads to the loss of quadruplex signals at temperatures (TBA and TBA2C[4,13] around 50ºC and TBA2A[4,13] around 40°C (Figure 2A, 2B), which match well the Tm derived from UV melting experiments.

We have measured the NMR spectra of TBA2G[4,13] at different K^+ concentrations ranging from 5-100 mM KCl, and we have obtained essentially the same spectra. The spectra of TBA2G[4,13] showed a severe and generalized line broadening suggesting that several species are in equilibrium (Figure 2C). Visible broad NH signals were found in the 10.5-11.5 ppm. These signals remain visible even at a high temperature. A quantitative analysis of the TBA2G spectra is difficult due to the existence of different conformers.

Spectra of TBA, TBA2A and TBA2C can be analyzed to obtain general structural information (see proton assignments in Tables S7-S8). These derivatives showed NH signals between 11.5-12.5 ppm that correspond to antiparallel conformation. In addition, the C4 and C13 NH imino signals in TBA2C present at 5°C disappear when we increased the temperature to 15°C (data not shown). The same effect occurs to the imino protons of the thymines in the loops, indicating that they are more exposed to the solvent. We did not find chemical shift signals for the imino protons of A4 and A13. The 1D NMR spectrum of TBA2A showed a main antiparallel monomeric species in solution, together with another minor monomeric species, as proved by low intensity signals present both in the NH imino and in the aromatic protons region (Fig. 2B). The presence of A4 and A13 affected the NH imino protons of thymines in the loops. No NH signals were observed for T12, T7 and T3 which implies exposure to a major solvent. Similarly, the path of NOE connectivity for TBA2C between aromatic protons and H1' of the ribose of G5, G10 and G14 was characterized by strong intensities as expected for *syn* conformation, whereas weak intraresidue intensities found for G2 and G6 were associated with an *anti* conformation. No sequential connections were found between residues G2 and T3, T3 and C4, C4 and G5, T7 and G8, T9 and G10 or C13 and G14. We observed, characteristic long range interactions i.e. H8 of G11 with NH imino proton of C13, H8 of G2 with H2',H2'' and NH of C4. These contacts were confirmed by the NOEs between H6 of C4 with H1', H2' and H2'' of G2. We also detected interactions between H8 of G8 and H1', H2'' and NH of G6 and between H1'T9 and H8 of G15. Moreover the methyl group of T9 gave a NOE signal with NH imino protons of G6, G10, G1 and G15 (Figure S4). All these findings pointed to similarity in the backbone conformation with the unmodified TBA. We also detected other interactions between aromatic protons of C4 and G2, C13 and G11.

The spectral assignment for TBA2A[4,13] was difficult due to the large overlap of NH and aromatic protons signals. Nevertheless, some NOE contacts were found between H8 of G2 and ribose protons of G6, H8 of G15 and H1' of T9, MeT9 and NH imino protons of G15, G10, G1 and G6. Finally, NMR DOSY experiments allowed us to determine self-diffusion coefficients (DC): for TBA, TBA2C and TBA2A we obtained DC values of 1.86 x 10^{-6} cm^2s^{-1} (expected value for monomeric form 1.77 x 10^{-6} cm^2s^{-1}), while for TBA2G we obtained a smaller value (1.26 x 10^{-6} cm^2s^{-1}, expected value for dimeric form 1.26 x 10^{-6} cm^2s^{-1}), suggesting the existence of a dimer structure for TBA2G.

ESI-MS and IMS of TBA derivatives

Figure 3 shows the ESI-IMS-MS results for unmodified TBA and TBA2G[4,13], in aqueous solution (top) and 60% methanol solution (bottom). The data are represented in 2D, with m/z separation horizontally, and ion mobility drift time (dt) separation vertically.

ESI-IMS-MS of modified and unmodified TBA spectra from purely aqueous ammonium acetate in the negative mode (Figure 3A) gave ion signals corresponding to the intramolecular G-quadruplex at charge states 3-, 4- (major), and 5-. For TBA2G[4,13] (Figure 3B), an additional charge state distribution is detected, corresponding to dimeric G-quadruplexes (noted D), at charge states 5-, 6- and 7-. ESI-IMS-MS spectra recorded for TBA in dehydrating conditions (Figure 3C) were similar to

those in aqueous conditions. There is a very slight relative increase of a dimer signal, indicated by shadows on the 2D plot, but the relative signal is too weak to extract the corresponding mass data. In contrast, for the TBA2G[4,13] (Figure 3D), the addition of methanol leads to the disappearance of the monomer, and the enlargement of the peaks corresponding to the dimer, which is the major species. Trimer (noted T) and tetramer (noted Q) ion series are also detectable.

In the M^{3-} monomeric forms of TBA and TBA2G, we found 0, 1 of 2 ammonium ions bound to TBAs, in agreement with previous results [42]. For the charge state 3-, the collision cross section does not change with the ammonium ion count [43]. As the lowest charge state is also the least prone to gas-phase conformational changes due to Coulomb repulsion, we used M^{3-} to compare gas-phase experimental and theoretical collision cross sections. With the same reasoning, we used the collision cross section of the dimer D^{5-} peak to find out the plausible gas-phase structure of the dimer. At least four ammonium ions are preserved in the dimer in both conditions with or without methanol. (see blue insets). Interestingly, more ammonium ions are preserved in the presence of MeOH (see insert in Fig. 3D, the distribution maximum is with 5 ammonium ions). This may be due to an increasing cation activity in 60%/40% methanol/water compared to 100% water.

We obtained the collision cross sections (CCS) for different gas phase ions from ion mobility experiments. Based on our previous results [44-48] such observables provide valuable information on the shape of monomers and dimers, both in the gas phase and in solution. Results shown in Table 3 match well with those derived from extended MD simulations in the gas phase (see section next section), confirming that detected signal correspond to the expected monomeric antiparallel (low CCSs), and mostly dimeric parallel G-quadruplexes (large CCS found for TBA2G[4,13]).

All-atoms resolution theoretical studies

The spectroscopic evidences just described suggest that the solution structure of TBA derivatives is the result of a subtle balance between at least three species: a single stranded oligonucleotide, a monomeric antiparallel quadruplex and a dimeric parallel quadruplex (Figure 4). Clearly, as the concentration of oligo increases, dimeric structures become more populated, but the entropic advantages of an intramolecular process favors the antiparallel quadruplex even in the high NMR concentration regime. However, the complete equilibrium can be modulated by the introduction of mutations at the loops. The physical reasons are however unclear, which forced us to perform theoretical calculations.

Extended MD simulations in aqueous solution of TBA, TBA2A[4,13], TBA2C[4,13] and TBA2G[4,13] result in stable trajectories that do not significantly deviate from the starting structures generated from the known NMR conformation of TBA in water (Figure S6). Our simulations support the experimental observation that the change of nucleotides (T→A, C or G) at the lateral loop might destabilize the structure, but does not lead to disruption of the antiparallel arrangement in physiologic-like conditions. We also performed MD simulations for models of dimeric parallel quadruplex for TBA, TBA2A[4,13], TBA2C[4,13], TBA2G[4,13] and TBA4G[3,4,12,13] sequences. The intermolecular TBA2G[4,13] and

TBA4G[3,4,12,13] quadruplexes remain fully stable for 0.4 microseconds, in a conformation that resembles very closely a canonical parallel quadruplex (Figure 5) with the central channel occupied by 3-4 cations. On the contrary, the parallel quadruplex generated from TBA, TBA2A[4,13] and TBA2C[4,13] sequences leads to unstable trajectories, characterized by a large increase in RMSd (above 0.5 nm from the original model structure), loss of ions in the central channel and complete corruption of structure at the end of the trajectory, where dissociation of the two strands is more evident for TBA (see embedded representations in Figure 5). Furthermore, the CCS of the main charge states for TBA (M^{3-}) and TBA2G dimers (D^{5-}) computed from MD simulations in gas-phase agrees well with the experimental measurements (Table 3).

MD simulations strongly support that a dimer of TBA4G[3,4,12,13] can form a very stable parallel quadruplex thanks to an un-interrupted tract of six G-tetrads that maintains a strong hydrogen bond network and coordinates around five ions in the central cavity. According to our MD simulations, the presence of one thymine tetrad in the central part of the intermolecular parallel G-quadruplex is well tolerated (the case of TBA2G[4,13] sequence), despite the reduction of hydrogen bonds as compared to a guanine tetrad. Ion coordination between T- and G-tetrad is tolerated and is consistent with ESI-MS data. On the contrary, our simulations strongly suggest that TBA, TBA2C[4,13] and TBA2A[4,13] do not tolerate well the inclusion of two consecutive non-guanine tetrads in the central steps, and thus they do not form intermolecular parallel quadruplexes, and the intramolecular antiparallel quadruplex dominates the conformational landscape of these oligonucleotides.

To gain a more quantitative explanation of the role of different nucleotides in the stabilization of the parallel quadruplex, we computed the changes that nucleotide mutation in the loop produces in the relative stability of unfolded single strand oligonucleotide, antiparallel G-quadruplex and a model parallel quadruplex. We used standard thermodynamic cycle to extract the differential binding free energy between the sequences from the reversible works associated with alchemical transformations between nucleotides in different environments (unfolded, antiparallel and parallel; see method section, SI material and figure S3). The results (summarized in Table 4) clearly suggest that the mutations for T to purines (G or A) do not dramatically affect the intrinsic stability of the antiparallel quadruplex (i.e. do not change the relative stability of the antiparallel and unfolded conformations). As expected from unbiased MD, the same T→G mutation that was innocuous for the stability of the antiparallel conformation, dramatically affects the stability of the dimeric parallel quadruplex. In fact, the mutation TBA→TBA2G[4,13] stabilizes the intermolecular parallel quadruplex by more than 47 kJ/mol (more than 60 kJ/mol if we take the completely unfolded state as reference). Assuming additivity the multiple mutation TBA→TBA4G[3,4,12,13] would imply an extreme stabilization of more than 90 kJ/mol for the formation of the hypothetical parallel quadruplex (120 kJ/mol if the unfolded oligo is taken as reference). Note that these estimates are done in the absence of external loops (already considered in unbiased MD simulations), suggesting that the origin of the extreme stabilization of parallel quadruplexes stems from the formation of alternative tetrad (T4 in the case of TBA2G[4,13] and G4 in the case of TBA4G[3,4,12,13], Figure 5), and not from spurious loop effects. Theoretical simulations

provide then a quantitative and clear explanation on the surprising effect of loop mutations in the stability and in the overall topology of G-DNA quadruplexes.

.

Proposed structures for TBA derivatives

Substitutions of T4 and T13 by A, C and G generated distinct G-quadruplex structures depending on the nucleobase. Experimental results showed that substitution by A and C generates antiparallel quadruplex structures. TBA2C[4,13] resulted in a stable quadruplex derivative similar to TBA, an observation that implies that the two cytosines are well accommodated in the overall antiparallel structure. By contrast, TBA2A[4,13] showed a destabilized structure as reflected by our spectroscopic data. Figure 4 summarizes the hypothetical structures for TBA derivatives observed in this manuscript.

Double guanine substitution derivatives (TBA2G[4,13], TBA2G[3,12]) resulted in a more stable parallel quadruplex structures. DOSY experiments and the agreement between simulated and experimental CCS for TBA2G[4,13] suggested that the parallel G-quadruplex is originated by the dimerization of two TBA2G[4,13] oligonucleotides. Note the CCS of TBA2G[4,13] is less than the sum of the constituent TBA monomers both in theory and experiment. In addition, PAGE analysis of TBA derivatives showed a main band with lower migration for TBA2G derivatives indicating a larger structure than TBA (Figure S5).

Consequently, we proposed a parallel structure with five G-tetrads and a T-tetrad and two TGT propeller loops (Figure 4). Spectroscopic data showed that the preferred structure is a dimer but other parallel multimers are not discarded. NMR imino protons are broad and several ammonium ions are observed in the ESI-MS spectra. Multimer quadruplex may correspond to the stacking of several parallel dimers. Experimentally 5 ammonium ions were observed in the ESI-MS spectra, suggesting that additional ammonium ions could be accommodated between T- and G-tetrads. To our knowledge, this coordination has not been described to date.

The complete rearrangement of two unfolded structures is required for the formation of the parallel dimer. This dimer was previously observed in TBA derivatives carrying riboguanosines or acyclic thymidines derivatives [35,49]. In addition, some authors have described the formation a bimolecular antiparallel quadruplex for unmodified TBA [50] in contrast with earlier findings and NMR results [29,18-19].

In the proposed dimer, the four thymidines are intercalated by stacking between two G-tetrads. This T-tetrad also contributes to the overall stability of the structure. Nevertheless, the broad signal of the dimer at 10-11 ppm does not allow assignment of the NH imino protons of the T-tetrad. A similar tetrad has been described by NMR in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisae* telomeric repeats d(TGGTGGC). The T-tetrad is formed by four thymidines held together by O4-H3 H bonds in a plane and is well accommodated in the center of the G-quadruplex and flanked by two G-tetrads on each side. The T-tetrad is less stable than the G-tetrad (only four hydrogen bonds

and a smaller surface for stacking interactions) [51]. This thymidine arrangement was also studied by theoretical studies [52]. TBA4G[3,4,12,13] derivative forms a very stable parallel quadruplex structure for which we propose a preferentially very compact dimer structure with 6 G-tetrads.

In addition to the proposed dimer, multimers are also observed. The presence of multimeric species may be due to stacking of dimers on the G-tetrads at the end of the quadruplex. These multimeric forms are not observed in the chair-like TBA structure due to the presence of the TT and TGT loops near the G-tetrads that prevents stacking. In this way the presence of multimeric species is in agreement with the parallel dimer structure suggested for the TBA2G and TBA4G derivatives.

In the case of TBA1G[4] derivative, where T4 was substituted by G4, it is not possible to form an extra G-tetrad and the parallel dimer is expected not to form. CD confirmed the antiparallel structure and the Tm, and the thermodynamic parameters showed that this derivative was destabilized. However, the Tm was increased at high concentration. This observation and the PAGE experiment (Figure S5) implied that equilibrium of complex multimeric species was present in solution.

CONCLUSIONS

G-quadruplexes are very polymorphic structures whose major conformation in solution can change due to multiple reasons. By combining experimental and theoretical tools we demonstrate here that the conformation of the thrombin aptamer sequence can change completely from a chair-like intramolecular antiparallel quadruplex to a parallel quadruplex by mutations in the lateral loop region. The present results might force us to revise our assumptions on the nature of chromosomal quadruplexes, since apparently small sequence alterations can make inter-molecular parallel quadruplexes competitive with the intra-molecular antiparallel ones, opening new potential roles of G-quadruplexes in genomic recombination and telomere fusion processes.

EXPERIMENTAL PROCEDURES

Oligonucleotides prepared for this study (see Table 1), correspond to the TBA modified in positions 3, 4, 12 and 13 that were originally occupied by thymidines. The modified TBAs were synthesized on an automated RNA/DNA synthesizer using β-cyanoethylphosphoramidite chemistry and following standard protocols. The oligonucleotides were purified by HPLC and characterized by mass spectrometry.

Ultraviolet absorbance and Circular Dicroism measurements.

Oligonucleotides at 3 μM were re-suspended in either potassium buffer (10 mM sodium cacodylate 100 mM KCl pH 7.0) or ammonium buffer (100 mM ammonium acetate pH 7.0) and were then annealed at 85 ºC and slowly cooled to room temperature. TBA and TBA2G[4,13] were also studied under dehydrating conditions. For this purpose the oligonucleotides were dissolved in the ammonium buffer with an increasing percentage of MeOH prior to analysis. The thermal curves were carried out

following the absorption change at 295 nm from 15ºC to 80ºC with a linear temperature ramp of 0.5 ºC/min on a JASCO V-650 spectrophotometer equipped with a Peltier temperature control. Thermal denaturation experiments were analysed at a range concentrations of (1.5, 3, 15 and 30 μM). The thermal denaturation-renaturation profile (hysteresis) was studied at 3 μM at 0.5 ºC/min.

Thermodynamic parameters were calculated from UV-melting experiments as described elsewhere [53] using Matlab program (R2009b version, Math-Works, Natick, MA, USA). This analysis assumes simple two-state equilibrium between the folded and unfolded states. Thermodynamic analysis of TBA2G[4,13] and TBA2G[3,12] assumed a self-complementary bimolecular process and for the rest of the derivatives an intramolecular process.

The CD spectra were recorded on a JASCO spectropolarimeter J-810. Spectra were registered at 25ºC over a range of 220-320 nm with a scanning speed of 50 nm/min, a response time of 4 s, 0.5 nm data pitch and 1nm bandwidth. The samples (4 μM) were dissolved in potassium or ammonium buffer and annealed before recording spectra. To analyse the effect of crowding agents TBA and TBA2G[4,13] were dissolved in ammonium buffer with an increasing percentage of MeOH (20, 40, 60 %) prior to analysis.

NMR spectroscopy.

NMR spectra were recorded on a Bruker AV600 spectrometer operating at a frequency of 600.10 MHz for ^{1}H. ^{1}H NMR spectra were recorded at variable temperature ranging from 5°C to 65°C. Chemical shifts (δ) were measured in ppm and referenced to external DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) set at 0.00 ppm. Estimated accuracy is within 0.02 ppm.

For NMR experiments, TBAs were dissolved in H_2O/D_2O (90:10 v/v) at a concentration range of 0.7-1.0 mM in the presence of 10 mM potassium phosphate buffer and 5 mM KCl (pH 6.7). The sequential ^{1}H assignments for all oligonucleotides were performed by applying well established procedures for the analysis of thrombin binding aptamer [29]. All the protons were attributed for TBA2C[4,13] and partially attributed to TBA2A[4,13]. The melting experiments were performed using ^{1}H NMR spectroscopy at variable temperature. Phase-sensitive NOESY spectra (t_{mix}=300 ms) were acquired at 5°C in TPPI mode, with 4K x 512 complex FIDs, a spectral width of 15.000 Hz for H_2O solutions, recycling delay of 1.5 s, and 120 scans. TOCSY spectra were acquired with the use of a MLEV-17 [54] spin-lock pulse (field strength 10000 Hz, 60 ms total duration). All spectra were transformed and weighted with a 90° shifted sine-bell squared function to 4K x 1K real data points. For the H_2O suppression, the excitation sculpting sequences from standard Bruker pulse program libraries were used. Pseudo two-dimensional DOSY [55] experiments were performed using the pulse-program “stebpgp1s”, diffusion delay: 0.12-0.45s; gradient pulse: 1.5ms; number of increments: 64. Temperature 25ºC. Raw data were processed using the standard DOSY software present in the Bruker library (TOPSPIN v. 1.3). A calibration curve was obtained using samples with a molecular weight (MW) ranging from 180 to 23.500 as standards, as previously reported [56]. The values of the diffusion coefficients obtained for TBA, TBA2C and TBA2A were 1.86 x 10^{-6} cm^2s^{-1} ($10^{-9.73}$ m^2s^{-1}). The expected value for the monomeric form was 1.77 x 10^{-6} cm^2s^{-1} ($10^{-9.75}$ m^2s^{-1}). The values obtained for

TBA2G were 1.26 x 10^{-6} cm^2s^{-1} ($10^{-9.90}$ m^2s^{-1}), (expected value for the dimeric form 1.26 x 10^{-6} cm^2s^{-1}, $10^{-9.90}$ m^2s^{-1}).

Electrospray mass spectrometry (ESI-MS) and Ion mobility spectrometry (IMS). The ESI-MS experiments were carried using a SYNAPT HDMS (Waters, Manchester, UK) recorded in ion mobility mode. The capillary voltage was set to -2.2 kV; cone voltage = 30 V; extraction cone = 4 V; source pressure= 3.15 mbar; source and desolvation temperatures = 40ºC and 60°C, respectively; trap and transfer voltages = 6 V and 4 V, respectively. The ion mobility cell was supplied with N_2 to reach a pressure of 0.532 mbar in the IMS cell (instrument pirani reading). The wave height was 8 V and the wave speed 300 m/s. The bias voltage for ion introduction into the IMS cell was 10 V. SYNAPT HDMS was calibrated in collision cross section with oligonucleotides, as described previously [57]. Oligonucleotides TBA, TBA2G[4,13], TBA2C[4,13] and TBA2A[4,13] were folded at 200 μM in 100 mM ammonium acetate buffer and injected at a final strand concentration of 5 μM and at a rate of 140 μL h/1 at room temperature. Higher mobility (higher charge or lower collision cross section) translates in shorter drift time. Here ion mobility is used to separate the charge state series corresponding to different stoichiometries. For example, if a given m/z can correspond to a monomer M^{z-} and a dimer D^{2z-} (double the mass and double the charge), given that collision cross sections of dimers are almost always less than twice than those of monomeric forms, and that the charge is double for the dimer, the mobility is higher for the dimer than for the monomer. The higher the molecularity, the lower the drift time. As a consequence, multimers separate on different diagonals on the 2D plots.

Molecular dynamics simulations.

Model preparation. We obtained the initial structure of TBA in its native antiparallel fold from the PDB structure 148D. The same PDB structure was used to model the TBA2G, TBA2A and TBA2C mutations in monomeric antiparallel form by exchanging residues in positions 4 and 13 from thymine to guanine, adenine or cytosine, respectively. We used the PDB structure 2JT7 –sequence $(TG_4T)_4$– as a template to model a parallel G-quadruplex fold built of four 8-mer chains of the sequence TG_6T. We changed the third and/or fourth G-tetrads of the original oligonucleotide to T-tetrads to obtain then $(TGGTTGGT)_4$ and $(TGGTGGGT)_4$, which were used as model systems for the dimeric form of TBA and TBA2G. We used these structures for free energy calculations (see below), as the lack of loops allows isolating the effect of the mutations. We also build parallel G-quadruplexes of sequence $(GGTTGG)_4$, $(GGTAGG)_4$, $(GGTCGG)_4$, $(GGTGGG)_4$ and $(GGGGGG)_4$ based on 2JT7 PDB. In these parallel quadruplexes we connected the 3' termini of the strand 1 with the 5' termini of strand 2 with a manually placed TGT fragment, thus completing the TBA, TBA2A[4,13],TBA2C[4,13], TBA2G[4,13], and TBA4G sequences. Using the same operation between strands 3 and 4 we finally obtained the dimeric G-quadruplex models. More details can be found in the SI material.

We immersed each DNA structure previously described in an equilibrated water box with an amount of potassium ions that both fill the cavities between each guanine tetrad and also ensures overall charge neutrality. The structures were optimized, thermalized and pre-equilibrated using a slightly

modified version of our standard protocol [58] (see SI) prior to an extended equilibration of 10 ns. Production runs are in the range of 0.2-0.7 μsec (see Suppl. Table S1) for detailed enumeration of all MD simulations. All simulations were performed using the Gromacs-4.5 software [59], the force fields describing the interactions for DNA were generated based on the pam99-parmBSC0 [60,61] parameters, SPC/E [62] model was used to describe the water molecules.

Gas phase simulations. We used the last structures of our aqueous MD simulations of TBA and the parallel dimer of TBA2G as initials models for gas phase calculations. For TBA we used one starting conformation, while for the dimer of TBA2G we used two extracted from our two different MD simulations. As done in previous works [44-47], we selected the most probable ionic states based on our Mass Spectrometry results (see below), which suggested a charge of 5- on TBA and of 10- for TBA2G parallel dimer. Since there are two cations inside TBA, the five cations in TBA2G parallel dimer, the total charge of the systems is, respectively, 3- and 5-. We selected two (TBA) and three (dimer TBA2G) charge distributions that optimized the placement of the protonated phosphates by reducing Coulomb repulsion between anionic phosphates (see Supp. Table S2). For TBA, we performed five different MD simulations (using different starting velocities) of 400 ns for each one of the two different charge distributions. For the TBA2G dimer, we performed three replicates for each one of the three charge distributions, using two different starting configurations, all of them 400 ns long. From each trajectory, we extracted 400 equally spaced structures and computed the predicted collision cross section (CCS) using the exact hard-spheres scattering (EHSS) approximation [63], which were then averaged to obtain the ensemble CCS.

Differential binding free energies. We determined the effect of the T→G substitution in the loops to the differential stability of monomer antiparallel and dimeric parallel quadruplexes. For this purpose, we used standard thermodynamic cycles (Figures S3, S6 and S7 and full details in SI material) to compute the reversible work associated to the T→G conversion in three different environments: i) a single stranded d(GGTGG) oligonucleotide, ii) an antiparallel TBA quadruplex (T_4 and T_{13}) iii) a parallel quadruplex $(TGGTTGGT)_4$ (T_4). From such reversible works the impact of T→G mutation in the relative stability of unfolded, antiparallel and parallel quadruplexes was determined taking advantage of the state function properties of the free energy (see Figure S3). We have also determined the difference in free energy of formation between TBA2A and TBA by transforming A→G in a single stranded d(GGAGG) oligonucleotide and in the antiparallel TBA2A. The reversible works associated to the alchemical transformations T→G, or A→G, in the different structures were computed using a thermodynamic integration method in its discrete formalism (DTI) [64].

FUNDING

This work was supported by the Spanish Ministry of Science and Innovation, MICINN (CTQ2010-20541, CTQ2012-38616, BIO2009-10964 and Consolider E-Science), the Generalitat de Catalunya, (2009/SGR/208), the University of Milano (PUR 2009 Funds), the funds de la Recherche Scientifique-

FNRS (VG research associate position and FRFC grant 2.4528.11) and PRIN09 (2009Prot-2009J54YAP_005). RF is a recipient of a FPI predoctoral contract (MICINN) and a STSM from COST (G4net, MP0802). GP is a recipient of a Sara Borrell postdoctoral fellowship. Collaborative research was funded by a Cost action (G4net, MP0802) and an Italian-Spanish collaborative action (IT2009-0067). CIBER-BBN is an initiative funded by the VI National R&D&i Plan 2008-2011, Iniciativa Ingenio 2010, Consolider Program, CIBER Actions and financed by the Instituto de Salud Carlos III with assistance from the European Regional Development Fund.

REFERENCES

1. Saenger W (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New-York, NY.

2. Bloomfield VA, Crothers DM & Tinoco I (2000) *Nucleic Acids: Structures Properties and Functions*. University Science Books. Sausalito, CA.

3. Neidle S (2002) *Nucleic Acids Structure and Recognition*. Oxford University Press. NY.

4. Neidle S & Balasubramanian S (2006) *Quadruplex Nucleic Acids*, Royal Society of Chemistry: London.

5. Shafer RH & Smirnov I (2001) Biological aspects of DNA/RNA quadruplexes. *Biopolymers* **56**, 209-227.

6. Petraccone L, Pagano B & Giancola C (2012) Studying the effect of crowding and dehydration on DNA G-quadruplexes. *Methods* **57**, 76-83.

7. Miller MC, Buscaglia R, Chaires JB, Lane AN & Trent JO (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3′ sequence of d($AG_3(TTAG_3)_3$). *J Am Chem Soc* **132**, 17105–17107.

8. Krishnan-Ghosh Y, Liu D & Balasubramanian S (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J Am Chem Soc* **126**, 11009–11016.

9. Neidle S (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr Opin Struc Biol* **19**, 239-250.

10. Wang Y & Patel DJ (1993) Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex *Structure* **1**, 263-282.

11. Parkinson GN, Lee MPH & Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876-880.

12. Monchaud D & Teulade-Fichou MP (2008) A hitchhiker's guide to G-quadruplex ligands. *Org Biomol Chem* **6**, 627-636.

13. Gatto B, Palumbo M & Sissi C (2009) Nucleic acid aptamers based on the G-quadruplex structure: therapeutic and diagnostic potential. *Curr Med Chem* **16**, 1248-1265.

14. Phan AT, Kuryavyi V, Ma JB, Faure A, Andéola ML & Patel DJ (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase*. Proc Natl Acad Sci USA* **102**, 634-639.

15. Bock LC, Griffin LC, Latham JA, Vermaas EH & Toole JJ (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564-566.

16. Padmanabhan K, Padmanabhan KP, Ferrara JD, Sadler JE & Tulinsky A (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J Biol Chem* **268**, 17651-17654.

17. Macaya RF, Schultze P, Smith FW, Roe JA & Feigon J (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex in solution. *Proc Natl Acad Sci USA* **90** 3745-3749.

18. Kelly JA , Feigon J & Yeates TO (1996) Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d(GTTGGTGTGGTTGG). *J Mol Biol* **256**, 417–422.

19. Trajkovski M, Šket P & Plavec J (2009) Cation localization and movement within DNA thrombin binding aptamer in solution. *Org Biomol Chem* **7**, 4677-4684.

20. Padmanabhan K & Tulinsky A (1996) An ambiguous structure of a DNA 15-mer thrombin complex. Acta Crystallogr Sect D Biol Crystallogr **D52**, 272–282.

21. Krauss IR, Merlino A, Giancola C, Randazzo A, Mazzarella L & Sica F (2011) Thrombin-aptamer recognition: a revealed ambiguity. *Nucleic Acids Res* **39**, 7858-7867.

22. Krauss IR, Merlino A, Randazzo A, Novellino E, Mazzarella L & Sica F (2012) High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Res* **40**, 8119-8128.

23. Smirnov I & Shafer RH (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry **39***, 1462–1468.

24. Hazel P, Huppert J, Balasubramanian S & Neidle S (2004) Loop-length-dependent folding of G-quadruplexes. *J Am Chem Soc* **126**, 16405–16415.

25. Smargiasso N , Rosu F, Hsia W, Colson P, Baker ES, Bowers MT, De Pauw E & Gabelica, V (2008) G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J Am Chem Soc* **130**, 10208–10216.

26. Guédin A, Gros J, Alberti P & Mergny JL(2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res* **38**, 7858–7868.

27. Risitano A & Fox KR (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res* **32**, 2598–2606.

28. Cevec M & Plavec J (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry* **44**, 15238–15246.

29. Schultze P, Macaya RF & Feigon J (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *J Mol. Biol* **235**, 1532-1547.

30. Nagatoishi S, Isono N, Tsumoto K & Sugimoto N (2011) Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie* **93**, 1231-1238.

31. Peng CG & Damha MJ (2007) G-quadruplex induced stabilization by 2'-deoxy-2'-fluoro-D-arabinonucleic acids (2'F-ANA). *Nucleic Acids Res* **35**, 4977-4988.

32. Virno A, Randazzo A, Giancola C, Bucci M, Cirino G & Mayol L (2007) A novel thrombin binding aptamer containing a G-LNA residue. *Bioorg Med Chem* **15**, 5710-5718.

33. Bonifacio L, Church F C & Jarstfer MB (2008) Effect of locked-nucleic acid on a biologically active G-quadruplex. A structure-activity relationship of the thrombin aptamer. *Int J Mol Sc* **9**, 422-433.

34. Pastenak A, Hernandez FJ, Rasmussen LM, Vester B & Wengel J (2011) Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer. *Nucleic Acids Res* **39**, 1155-1164.

35. Coppola T, Varra M, Oliveiro G. Galeone A, D'Isa G, Mayol L, Morelli E, Bucci RM, Vellecco V, Cirino G & Borbone N (2008) Synthesis and biological properties of new TNA analogues containing an acyclic nucleotide. *Biorg Med Chem* **16**, 8244-8253.

36. Spacková N, Cubero E, Sponer J & Orozco M (2004) Theoretical study of the guanine-> 6-thioguanine substitution in duplexes, triplexes, and tetraplexes. *J Am Chem Soc* **126**, 14642-14650.

37. Martín-Pintado N, Yahyaee M, Deleavey GF, Portella G, Orozco M, Damha MJ & Gonzalez C (2013) Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J Am Chem Soc* **135**, 5344-5347.

38. Saneyoshi H, Mazzini S, Aviñó A, Portella G, González C, Orozco M, Marquez V, & Eritja R (2009) Conformationally rigid nucleoside probes help inderstand the role of sugar pucker and nucleobase orientation in the thrombin binding aptamer. *Nucleic Acids Res* **37**, 5589-5601.

39. Zhang AY & Balasubramanian S (2012) The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. *J Am Chem Soc* **134**, 19297-19308.

40. Mukundan VT & Phan AT (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J Am Chem Soc* **135**, 5017-5028. .

41. Nagatoishi S, Tanaka Y & Tsumoto K (2007) Circular dichroism spectra demonstrate formation of the thrombin-binding DNA aptamer G-quadruplex under stabilizing-cation-deficient conditions. *Biochem Biophys Res Comm* **352**, 812-817.

42. Ferreira R, Marchand A & Gabelica V (2012) Mass spectrometry and ion mobility spectrometry of G-quadruplexes. A study of solvent effects on dimer formation and structural transitions in the telomeric DNA sequence d(TAGGGTTAGGGT). *Methods* **57**, 56-63.

43. Balthasart F, Plavec J & Gabelica V (2013) Ammonium ion binding to DNA G-quadruplexes: Do electrospray mass spectra faithfully reflect the solution-phase species?. *J Am Soc Mass Spectrom* **24**, 1-8.

44. Rueda M, Kalko S, Luque FJ & Orozco M (2003) The structure and dynamics of DNA in gas phase. *J Am Chem Soc* **125**, 8007-8014.

45. Rueda M, Luque FJ & Orozco M (2005) Nature of minor-groove binders-DNA complexes in the gas phase. *J Am Chem Soc* **127**, 11690-11698.

46. Rueda M, Luque FJ & Orozco M (2006) G-quadruplexes can maintain their structure in the gas phase. *J Am Chem Soc* **128**, 3608-3619.

47. Arcella A, Portella G, Ruiz M, Eritja R, Vilaseca M, Gabelica V & Orozco M (2012) The structure of triplex DNA in the gas phase. *J Am Chem Soc* **134**, 6596-6606.

48. Meyer T, de la Cruz X & Orozco M (2009) An atomistic view to the gas phase proteome. *Structure* **17**, 88-95.

49. Tang CF & Shafer RH (2006) Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J Am Chem Soc* **128**, 5966-5973.

50. Fialova M, Kypr J & Vorlickova M (2006) The thrombin binding aptamer GGTTGGTGTGGTTGG forms a bimolecular guanine tetraplex. *Biochem Biophys Res Comm* **344**, 50–54.

51. Patel PK & Hosur RV (1999) NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by Saccharomyces cerevisiae telomere repeats. *Nucleic Acids Res* **27**, 2457-2464.

52. Gu J & Leszczynski J (2001) A theoretical study of thymidine and uracil tetrads: structures, properties, and interactions with the monovalent K^+ cation. *J Phys Chem* **105**, 10366-10371.

53. Puglisi JD & Tinoco I (1989) Absorbance melting curves of RNA. *Methods Enzymol* **180**, 304-325.

54. Bax A. & Davis DG (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J Magn Reson* **65**, 355-360.

55. Morris KF & Johnson Jr CS (1992) Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *J Am Chem Soc* **114**, 3139-3141.

56. Mazzini S, Scaglioni L, Animati F & Mondelli R (2010) Interaction between double helix DNA fragments and the new antitumor agent sabarubicin, Men10755. *Bioorg Med Chem* **18**, 1497-1506.

57. Rosu F, Gabelica V, Joly VL, Gregoire G & De Pauw E (2010) Zwitterionic i-motif structures are preserved in DNA negatively charged ions produced by electrospray mass spectrometry. *Phys Chem Chem Phys* **12**, 13448-13454.

58. Shields GC, Laughton CA & Orozco M (1997) Molecular dynamics simulations of the d(T.A.T) triple helix. *J Am Chem Soc* **119**, 7463-7469.

59. Hess B, Kutzer C, van der Spoel D & Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**, 435-447.

60. Cheatham TE III, Cieplak P & Kollman PA (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn* **16**, 845-862.

61. Pérez A, Marchan I, Svozil D, Sponer J, Cheatham TE III, Laughton CA & Orozco M (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* **92**, 3817-3829.

62. Berendsen HJC, Grigera JR & Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* **91**, 6269-6271.

63. Shvartsburg AA, Mashkevich SV, Baker ES & Smith RD (2007) Optimization of algorithms for ion mobility calculations. *J Phys Chem A* **111**, 2002-2010.

64. Kirkwood JG (1935) Statistical mechanics of fluid mixtures. *J Chem Phys* **3**, 300-313

TABLE AND FIGURES LEGENDS

Table 1. Sequences of the TBA derivatives used in this study

Table 2. Melting temperatures and thermodynamic parameters obtained from denaturing curves for TBA derivatives in potassium buffer.

Table 3. Collision cross sections for the TBA derivatives ions at 3- charge state in ammonium acetate buffer determined experimentally, and computed from MD-simulations.

Table 4. Changes in free energy of different processes associated with mutations at loops. For example: ΔΔG antiparallel → parallel dimer (T→G) means the change in relative stability of antiparallel and parallel quadruplexes occurring as a consequence of the mutation of a T in the lateral loop of TBA into a G. The negative value (-48 kJ/mol) means that the parallel dimer is highly stabilized by the T→G change.

Fig. 1. CD spectra of modified TBAs in A): 10 mM cacodylate, 100 mM KCl pH 7 buffer and B) 100 mM ammonium acetate buffer. C) CD spectra of TBA in 100 mM ammonium acetate buffer with an increasing percentage of MeOH. D) CD spectra of TBA2G[4,13] in 100 mM ammonium acetate buffer with an increasing percentage of MeOH.

Fig. 2. Imino protons region. NMR melting experiment on (A) TBA2C[4,13], (B) TBA2A[4,13] and TBA2G[4,13] in H_2O, 10 mM potassium phosphate buffer and 5 mM KCl (pH 6.7).

Fig. 3. ESI-IMS-MS 2D spectra for (A) natural TBA in 100% aqueous 100 mM NH_4OAc, (B) TBA2G[4,13] in 100% aqueous 100 mM NH_4OAc, (C) natural TBA in 40% aqueous 100 mM NH_4OAc and 60% MeOH, (C), and (D) TBA2G[4,13] in 40% aqueous 100 mM NH_4OAc and 60% MeOH. Peak annotation showing theoretical peak location: M = monomer, D = dimer, T = trimer, Q = tetramer. The insets of Figure 3B and 3D are the mass spectra extracted for the peak D^{5-}, and show the number distribution of ammonium ions preserved in the dimers in both conditions.

Fig. 4. The figure illustrates the impact of nucleotide substitutions on the structure of TBA derivatives. The transition between the antiparallel monomer and the parallel dimer structures occurs via an unfolded intermediate, as direct conversion is kinetically very unfavorable. The left and right halves of the sketch show that the main structure corresponds to the parallel dimers if the positions 3,12 and 4,13 are occupied by two or four guanines (in the case of a complete substitution with guanines, the equilibrium is prominently displaced towards the parallel dimer). On the upper and lower halves we illustrate the equilibrium that correspond to TBA, TBA2C[4,13], TBA2A[4,13] and TBA1G[4] where the parallel dimer is not stable and the main structure is the monomeric antiparallel fold. The lower thermal stability of TBA2A[4,13] and TBA1G[4] compared to TBA and TBA2C[4,13] is indicated by a slight shift towards the unfolded state.

Fig. 5. (A) Time evolution of the root mean square deviation (RMSD) with respect to the core tetrads of each initial structure for the parallel dimers of TBA4G[3,4,12,13], TBA2G[4,13], TBA2A[4,13], TBA2C[4,13], and TBA (i.e., the RMSD excluding the atoms that belong to the loops). Whereas TBA4G[3,4,12,13] and TBA2G[4,13] adopt a stable conformation in a folded state after partially loosing one layer of G-tetrad, TB2A[4,13], TBA2C[4,13] and TBA rapidly and completely unfold, loosing the virtually all the G-tetrads (TBA2A[4,13] retains two of them due to the time scale of the simulations). The embedded molecular models show sketches of the end stated of each MD simulation. The potassium ions are shown as purple spheres. (B) Side (above) and top (below) close-up views of the central tetrads for TBA4G[3,4,12,13], TBA2G[4,13] and TBA after relaxation of the loops, prior to the unrestrained MD simulations. The rightmost figures illustrate a top view of the T-

tetrad (above) and a G-tetrad (below). For clarity, thymine carbons are colored in green, while guanine carbon atoms are colored in gray.

Fig. 1

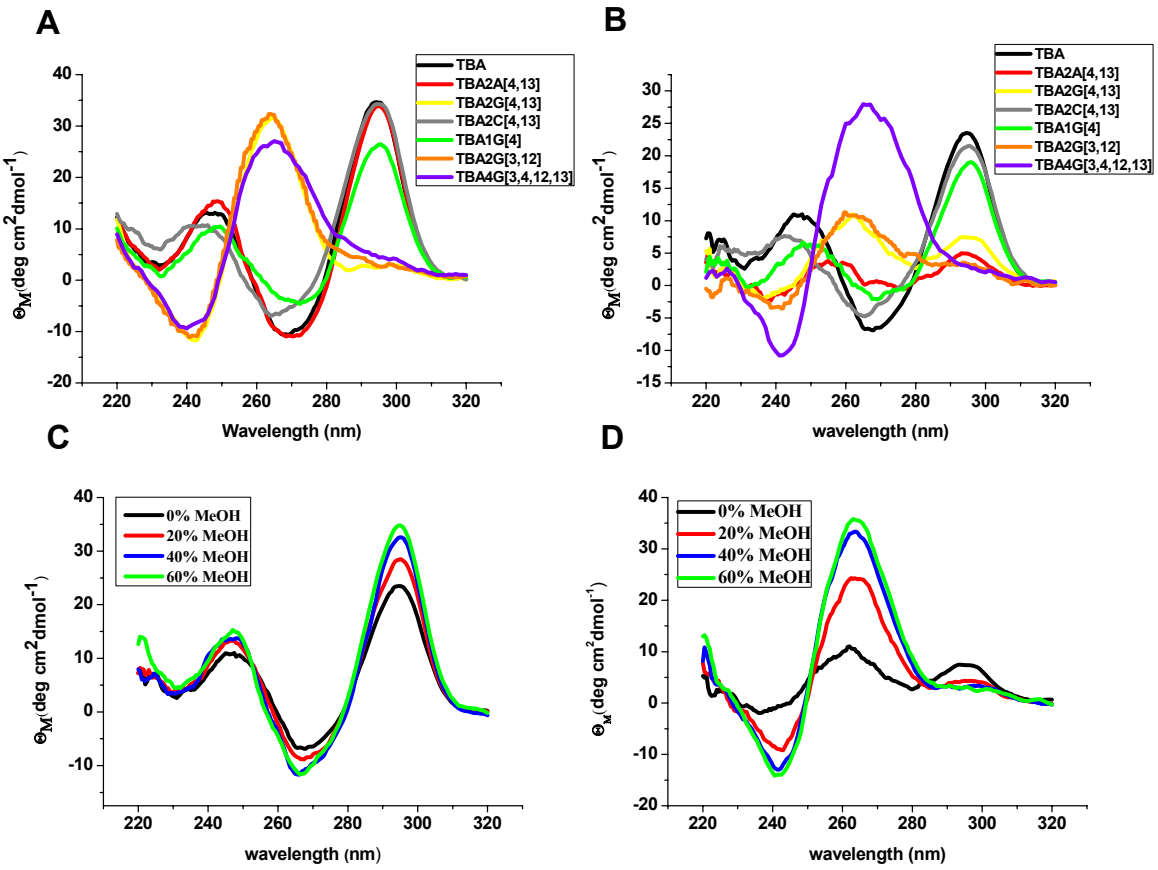
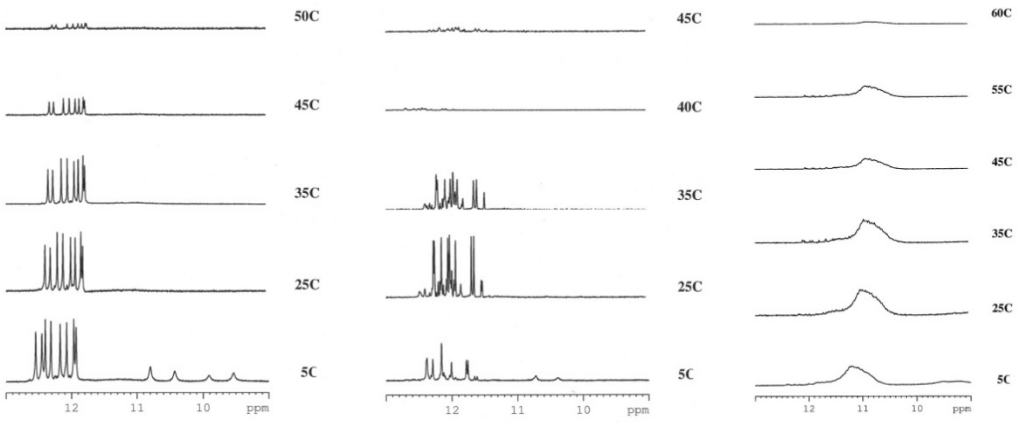


Fig. 2


50C
45C
35C
25C
5C
12
11
10
ppm
45C
40C
35C
25C
5C
12
11
10
ppm
60C
55C
45C
35C
25C
5C
12
11
10
ppm

(A) (B) (C)

Fig. 3

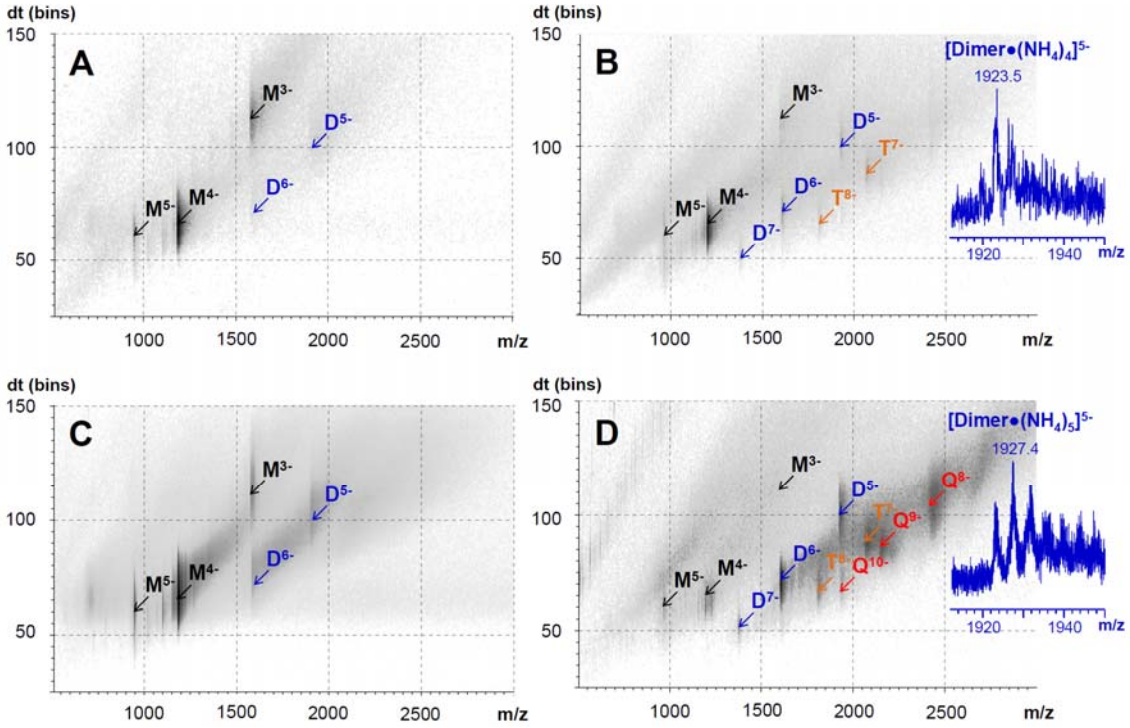

dt (bins)
150
100
50
A
M3-
M5-
M4-
D5-
D6-
1000
1500
2000
2500
m/z
dt (bins)
B
[Dimer•(NH4)4]5-
1923.5
M3-
M5-
M4-
D5-
D6-
D7-
T7-
T8-
1920
1940
m/z
dt (bins)
C
M3-
M5-
M4-
D5-
D6-
dt (bins)
D
[Dimer•(NH4)5]5-
1927.4
M3-
M5-
M4-
D5-
D6-
D7-
T7-
T8-
Q8-
Q9-
Q10-

Fig. 4

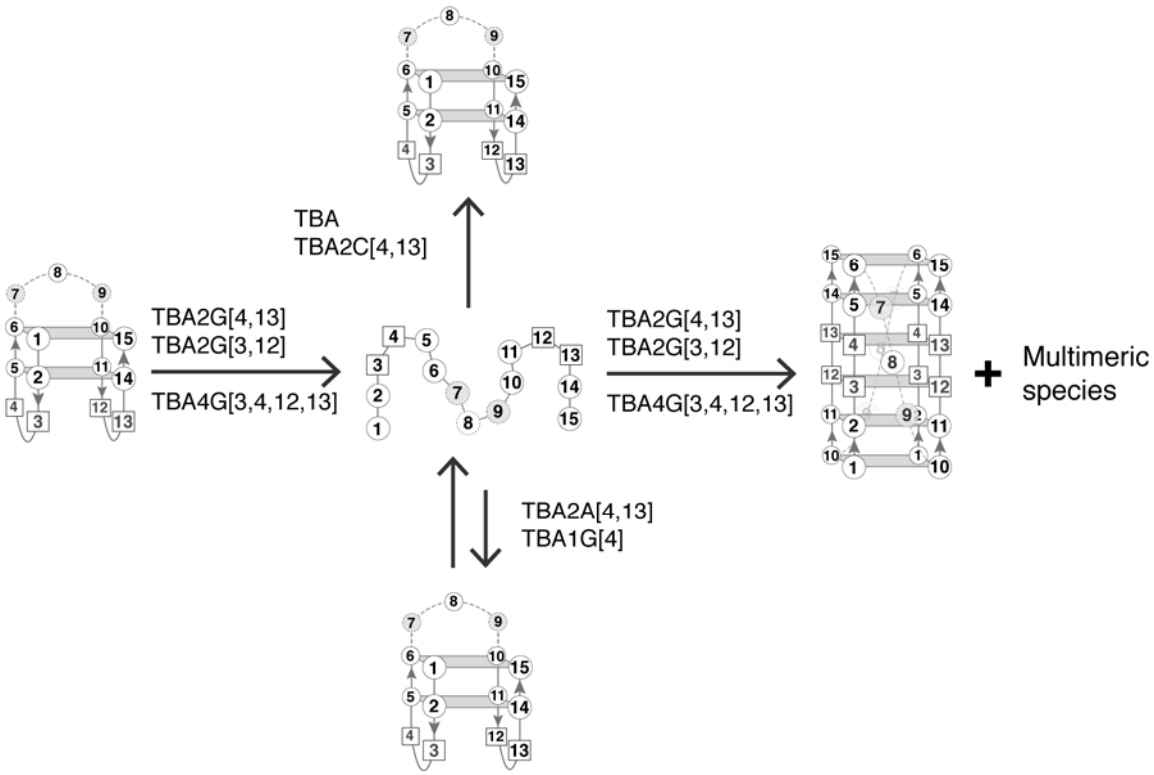

TBA
TBA2C[4,13]
TBA2G[4,13]
TBA2G[3,12]
TBA4G[3,4,12,13]
TBA2G[4,13]
TBA2G[3,12]
TBA4G[3,4,12,13]
Multimeric species
TBA2A[4,13]
TBA1G[4]

Fig. 5

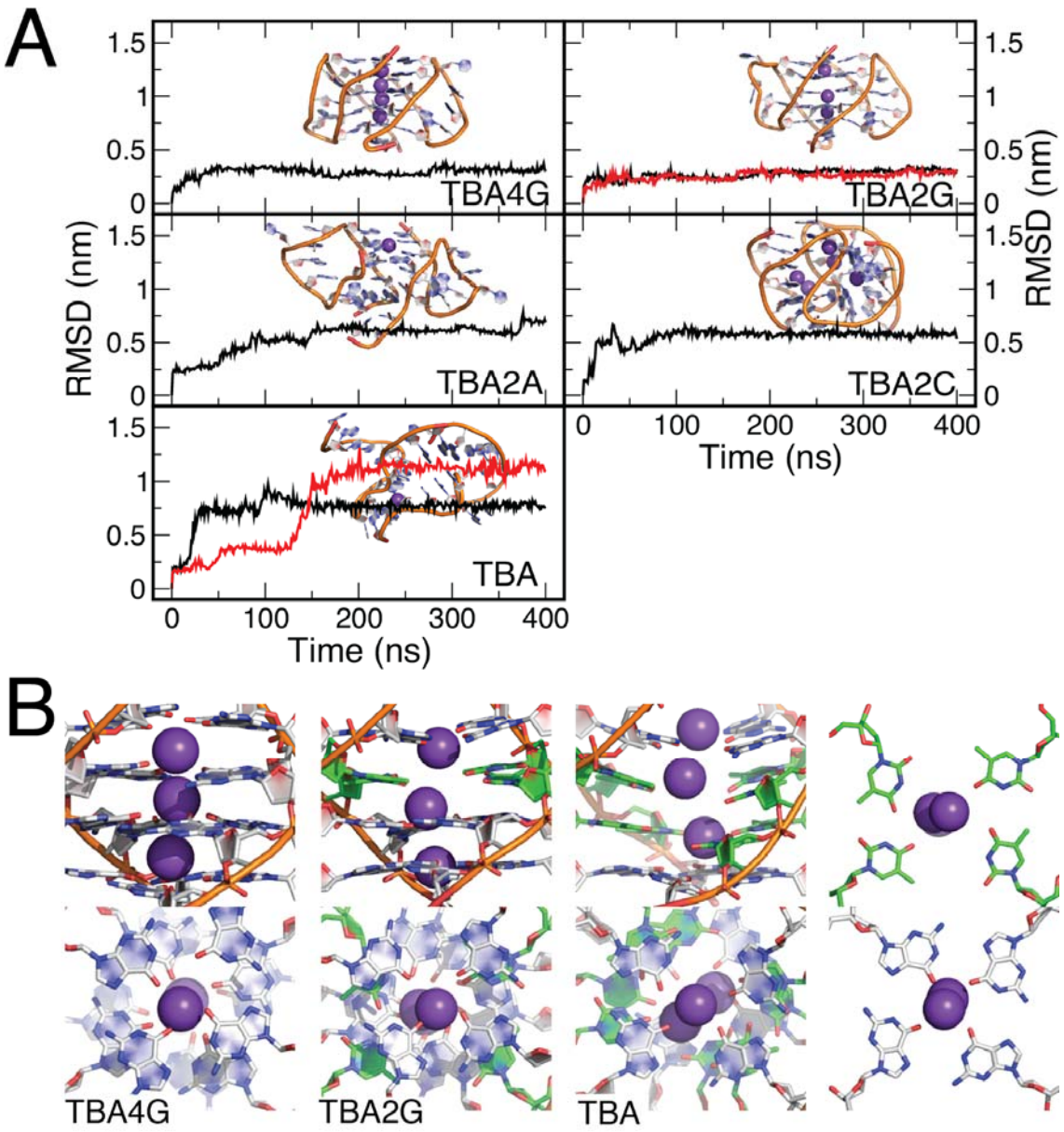

A
RMSD (nm)
Time (ns)
TBA4G
TBA2G
TBA2A
TBA2C
TBA
B
TBA4G
TBA2G
TBA

Table 1. Sequences of the TBA derivatives used in this study

Name	**Sequence**
TBA	GGTTGGTGTGGTTGG
TBA2A[4,13]	GGTAGGTGTGGTAGG
TBA2C[4,13]	GGTCGGTGTGGTCGG
TBA1G[4]	GGTGGGTGTGGTTGG
TBA2G[4,13]	GGTGGGTGTGGTGGG
TBA2G[3,12]	GGGTGGTGTGGGTGG
TBA4G[3,4,12,13]	GGGGGGTGTGGGGGG

Table 2. Melting temperatures and thermodynamic parameters obtained from denaturing curves for TBA derivatives in potassium buffer.

Name	**Tm (ΔTm)[a] (ºC)**	**ΔHº (Kcal/mol)**	**ΔSº (cal/Kmol)**	**ΔGº(25ºC) (Kcal/mol)**	**Hysteresis[b]**
TBA	50.1	-45.1±0.3	-139.6±1.0	-3.5	No
TBA2A [4,13]	40.1 (-10)	-33.2±0.5	-106.1±1.6	-1.6	No
TBA2C [4,13]	51.5 (+1.4)	-42.5±0.7	-131.0±2.4	-3.4	No
TBA1G [4]	47.9(-2.2)	-39.8±0.9	-124.0±2.9	-2.8	No
TBA2G [4,13]	59.2 (+9.1)	-63.4±0.9	-165.3±2.6[c]	-14.1	Yes
TBA2G [3,12]	58.5 (+8.4)	-58.9±2.5	-152.3±7.6[c]	-13.5	Yes
TBA4G [3,4,12,13]	>80	-	-	-	Yes

[a] ΔTm is the Tm change in the melting process between the modified TBAs and unmodified TBA.

[b] Hysteresis refers to the difference between the heating and cooling process with a temperature gradient of 0.5 ºC/min. Uncertainty ± 0.5°C

[c]. Apparent values using the approximation of a two-state equilibrium

Table 3. Collision cross sections for the TBA derivatives ions at 3- charge state in ammonium acetate buffer determined experimentally, and computed from MD-simulations.

	Ω (Å^2) Experimental[a]	**Ω (Å^2) MD-simulation**[b]
TBA	564 ± 1	569± 4
TBA 2A [4, 13]	564 ± 1	
TBA 2C [4, 13]	548 ± 1	
TBA 2G [4,13]	563 ± 1	
TBA 2G [4,13][c]	899 ± 2	928± 4

[a] Average values and standard deviations obtained from measurements on the Synapt G1 HDMS at bias voltage of 15V. The prediction error (95% confidence) is 1.6% in all cases. [b] MD values obtained from ensemble of more than 4 msec obtained for replicates with different charge distributions and initial structures (see Methods) defined for the ionic state determined from ESI-MS measurements. [c] Collision cross sections calculated for the dimer ion at 5- charge state

Table 4. Changes in free energy of different process associated with mutations at loops. For example: $\Delta\Delta G_{\text{antiparallel} \rightarrow \text{parallel dimer}}$ (T→G) means the change in relative stability of antiparallel and parallel quadruplexes occurring as a consequence of the mutation of a T in the lateral loop of TBA into a G. The negative value (-48 kJ/mol) means that the parallel dimer is highly stabilized by the T→G change.

Mutation	**Free Energy difference (kJ/mol)**
$\Delta\Delta G_{\text{antiparallel} \rightarrow \text{parallel dimer}}$ (T→G)	-48 +/- 2
$\Delta\Delta G_{\text{unfolded} \rightarrow \text{parallel dimer}}$ (T→G)	-62 +/- 1
$\Delta\Delta G_{\text{unfolded.} \rightarrow \text{antiparallel}}$ (T→G)	-7 +/- 1
$\Delta\Delta G_{\text{unfolded.} \rightarrow \text{antiparallel}}$ (A→G)	3 +/- 2

Supporting Information

SPECIFIC LOOP MODIFICATIONS OF THE THROMBIN BINDING APTAMER TRIGGER THE FORMATION OF PARALLEL STRUCTURES

Anna Aviñó, Guillem Portella, Ruben Ferreira, Raimundo Gargallo, Stefania Mazzini, Valerie Gabélica, Modesto Orozco and Ramon Eritja

Contents

Fig. S6. Structures of TBA (A) and the parallel dimer of TBA2G (B) as found at the end of our gas-phase MD simulations of 400ns**.**

Fig. S7. Time evolution of the root mean square deviation (RMSD) with respect to the initial structure for monomeric form of TBA, TBA2G[4,13], TBA2A[4,13], and TBA2C[4,13], in water solution**.**

Fig. S8. Schematic illustration of the steps followed to build the simulation boxes containing intermolecular G quadruplexes**.**

Table S1. List of MD simulations performed in this work**.**

Table S2. List of residue numbers of un-protonated nucleotides along the TBA and TBA2G[4,13] sequences in our gas-phase MD simulations**.**

Table S3. CCS ($Å^2$) obtained using the exact hard-spheres scattering (EHSS) approximation from the ensemble of structures generated in our gas phase MD simulations for TBA and TBA2G[4,13] dimer in their most populated charge states (5- and 10-, respectively, on the nucleic acids)**.**

Table S4. Melting temperatures of TBA2G[4,13] in phosphate buffer (10 mM) with increasing K^+ concentration**.**

Table S5. Melting temperatures and thermodynamic parameters obtained from denaturing curves in ammonium buffer for TBA derivatives.

Table S6. Melting temperatures and thermodynamic parameters obtained from denaturing curves in ammonium buffer for TBA and TBA2G[4,13] with an increasing percentage of MeOH**.**

Table S7. Proton chemical shift of TBA2C[4,13].

Table S8. Proton chemical shift of TBA2A[4,13].

Appendix S1.

Model building for the computational studies

The initial structure of TBA in its native antiparallel fold was obtained from the PDB structure 148D. The same PDB structure was used to model the TBA2A[4,13], TBA2C[4,13], TBA2G[4,13] mutations in monomeric antiparallel form by exchanging residues in positions 4 and 14 from thymine to guanine, adenine and cytosine, respectively.

To model a parallel G-quadruplex fold formed by combination of two TBA/TBA2G monomers we first built a model for a parallel G-quadruplex containing 6 G-tetrads along the central core using the PDB structure 2JT7 $(TG_4T)_4$ as template. This structure does not contain any loop connecting the different strands involved in G-quadruplex formation. We elongated the number of tetrads to eight by translational and rotational fitting of the ligand-free 2JT7 structure to itself by using two sets of shifted-but-overlapping groups of atoms along the DNA sequence. By merging the two overlapping chains we obtained an 8mer G-quadruplex of the sequence $(TG_6T)_4$. We morphed the third and/or fourth G-tetrads to T-tetrads to obtain $(TGGTTGGT)_4$ and $(TGGTGGGT)_4$, which were used as model systems for the dimeric form of TBA and TBA2G. We employed these structures for free energy calculations and for assessing the time dependence of their structural stability.

We obtained the initial models for the full-sequence dimeric G-quadruplex TBA and TBA2G (see scheme in figure S6) also based on 2JT7 PDB, by following the procedure described above. We first prepared parallel G-quadruplexes of sequence $(GGTTGG)_4$, $(GGTAGG)_4$, $(GGTCGG)_4$, $(GGTGGG)_4$ and $(GGGGGG)_4$ which constitute the core of the G-quadruplexes. Then, we connected the 3' termini of the strand 1 with the 5' termini of strand 2 with a manually placed TGT fragment, thus completing the TBA, TBA2A[4,13], TBA2C[4,13], TBA2G[4,13], and TBA4G[3,4,12,13] sequence. Using the same operation between strands 3 and 4 we finally obtained the dimeric G-quadruplex models.

Each DNA structure described (TBA, TBA2G, TBA2C, TBA2A, $(TGGTTGGT)_4$, $(TGGTGGGT)_4$, dimeric TBA and dimeric TBA2G) was immersed in a previously equilibrated water box using an octahedral cage such that the box boundaries are positioned at least 2 nm away from any DNA atom. The amount of potassium ions added ensured both overall charge neutrality and filled the cavities between each guanine tetrad (these internal ions were manually placed to avoid long equilibration times required to spontaneously capture them from the solvent). All modeled structures were first relaxed using and extended version of our standard protocol (46) first by energy minimization (with DNA backbone fixed by soft (500 kJ/mol/nm^2), followed by thermalization and pre-equilibration (5 ns) using the same soft restraints on the DNA backbone for all the antiparallel G-quadruplexes. For parallel dimeric G-quadruplexes we applied restraints only to the stem, and we extended the pre-equilibration runs to 10 ns. The resulting structures were then subjected to a final 10 ns equilibration prior to production runs (table S1).

Gas phase simulations

As stated in the methods section, we used the last structures of our molecular dynamics simulations of TBA and the parallel dimer of TBA2G as initials models for our gas phase calculations. For TBA we used one starting conformation, whereas we used two different conformations, extracted from our two different MD simulations for the dimer of TBA2G. In order to best model the behavior of DNA in the gas phase we must take into account that partial neutralization of the phosphate backbone takes place under mild desolvation conditions. As done in previous works (54), we selected the most probable charge states based on the Ion Mobility Mass Spectrometry measurements performed for this work. Taking into account that TBA contains 2 potassium cations in the central cavity and 5 of them are present in the TBA2G dimer, the charge state (i.e. the total charge of the nucleic acid not including the cations) for TBA was set to 5-, and we used 10- for the TBA2G parallel dimer.

We selected two charge distributions for TBA (i.e. two different arrangements of the phosphate protonation) and three for the each of the two structures of the TBA2G parallel dimer. The location of the non-protonated phosphate groups was assigned based on charge symmetry considerations, placing the charge groups as far as possible, assuming that they are representative of the ensemble of quasi-degenerated charge states detected experimentally for a given total charge. The exact distribution of charge phosphates can be found in table S2. For TBA, for each charge distribution we performed five different MD simulations of 400 ns, using different initial velocities for each one. In the case of TBA2G parallel dimer, we performed three MD replicates for each charge distribution and initial configuration, each one with different starting velocities (same trajectory length as before). From each trajectory, we extracted 400 equally-spaced structures and computed the predicted collision cross section (CCS) using the exact hard-spheres scattering (EHSS) approximation (55). A complete list of simulations can be found in table S1, and the results of our CCS calculations are summarized in table S3, decomposed by charge distribution.

All simulations were carried out using the Gromacs-4.5 softwar3 (47). For the simulations in solution we have used periodic boundary conditions and the particle mesh Ewald method[5] for the long-range electrostatics, together with a cut-off of 1.0 nm for the short-range repulsive and attractive dispersion interactions, which were modeled via a Lennard-Jones potential. In the gas phase simulations we also used periodic boundary conditions, but with a large cubic box (each side has 200 nm), large enough to mimic vacuum conditions, and a very large cut-off distance (20 nm) to effectively capture all the electrostatic and Lennard-Jones interactions. The Settle algorithm (66) was used to constrain bond lengths and angles of water molecules, and P-Lincs (67) was used for all other bond lengths (both for the aqueous and gas phase simulations), allowing a time step of 2 fs. The temperature was kept constant at 300 K in the simulations carried out in water, and at 340 K for the simulations in gas phase. For both type of simulations we have used the thermostat method of Bussi et al. (68). The pressure was controlled by coupling the simulation box to a pressure bath at P= 1 atm.

The force fields describing the interactions for DNA were generated based on the amber99SB+parmBSC0 (48, 49) parameters, and we used the SPC/E (50) model to describe the water molecules.

Free energy differences based on Thermodynamic Integration calculations

We obtained the binding free energy difference between two parallel G-quadruplex containing the sequence $(TGGTGGGT)_4$ and $(TGGTTGGT)_4$. These two idealized G-quadruplexes mimic the number of guanine tetrads of TBA and TBA2G dimers if folded as a parallel G-quadruplex, while avoiding including the effect of the loops connecting the strands.

We have computed the reversible work associated to the alchemical transformation between two DNA sequences, e.g., converting thymine to guanine, both in the bound and un-bound state. As a reference state, we have used the work associated with the T→G transformation (TBA→TBA2G transformation) for the monomeric antiparallel G-quadruplex state, and also the work associated to the T→G transformation in a single stranded d(GGXGG) oligo, where X is mutated from T to G. We have also determined the difference in free energy of formation between TBA2A and TBA by transforming A→G in a single stranded d(GGTGG) oligonucleotide and in the antiparallel TBA2A.

The reversible works associated to the alchemical transformations described above was computed using the thermodynamic integration method in its discrete formalism (56) (DTI), where the free energy between two states is λ=0 and λ=1 is computed by

$$\Delta G_{0\to 1} = \int_0^1 \left\langle \frac{\partial \mathrm{H}}{\partial \lambda} \right\rangle_\lambda d\lambda$$

, i.e. by integration of the derivative of the energy of the system as function of the state parameter λ, also known as coupling parameter.

To describe the interactions that are involved in the alchemical transformation, we used the mixed-topology libraries of Seeliger *et al.* (69) adapted for the parmbsc0 forcefield. In this approach each pair of nucleic acid mutations is described in terms of a hybrid residue, which contain a super set of all atoms necessary to represent a linear combination of both initial and final state depending on the coupling parameter. For the mutation, the variation of λ is discretized in 21 windows, i.e. dλ=0.05, and the final ΔG is computed via numerical integration. We have used soft-core potentials implemented in Gromacs to avoid singularities in the Lennard-Jones and Coulomb potentials, with α=0.3, σ=0.25 and a soft-core power of 1. The initial structures for each window were obtained from

a 10 ns simulation in which λ was continuously varied from 0 to 1. From this simulation we extracted frames corresponding to a given λ value, and we energy minimized the system in those configurations. We simulated each window for 10 ns, and we discarded the first nanosecond of simulation. For each window we collected 9 estimates for $\left\langle \frac{\partial \mathrm{H}}{\partial \lambda} \right\rangle_{\lambda}$ by using 9 blocks of 1ns, which were then integrated through the entire mutation pathway to obtain mutation free energies (with associated statistical errors). Integration of the derivative of the system energy and its associated statistical errors were computed using the implementation of the Bennett's acceptance ratio (70) implemented in Gromacs.

References

65. Darden T, York D & Pedersen L (1993) Particle mesh Ewald- an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys* **98**, 10089–10092.

66. Miyamoto S & Kollman PA (1992) Settle- an analytical version of the shake and rattle algorithm for rigid water models. *J. Comp. Chem* **13**, 952–962

67. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput* **4**, 116–122

68. Bussi G, Donadio D & Parrinello M (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys* **126**, 014101

69. Seeliger D, Buelens F P, Goette M, De Groot B L & Grubmüller H (2011) Towards computational specificity screening of DNA-binding proteins. *Nucleic Acids Res* **39**, 8281-8290.

70. Bennett C (1976) Efficient estimation of free-energy differences from Monte-Carlo data. *J. Comput. Phys* **22**, 245–268.

Fig. S1. Melting (in black) and annealing (in red) curves of TBA derivatives. A TBA, B TBA2A[4,13], C TBA2G[4,13], D TBA2C[4,13], E TBA2G[3,12], F TBA1G[4] and G TBA4G[3,4,12,13]. Buffer: 10mM sodium cacodylate 100 mM KCl pH 7

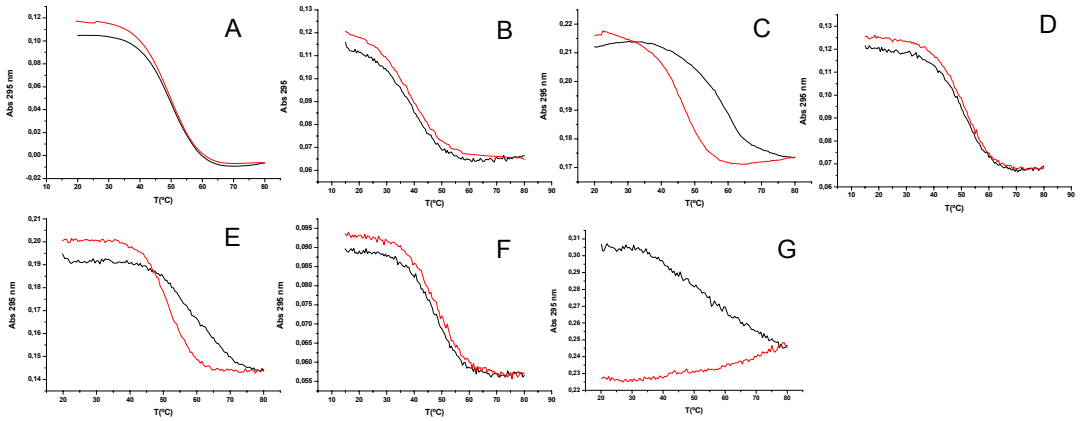


Fig. S2. UV denaturation data (*T*m) of modified TBAs at different concentrations. Uncertainty ± 0.5°C. Buffer: 10mM sodium cacodylate 100 mM KCl pH 7

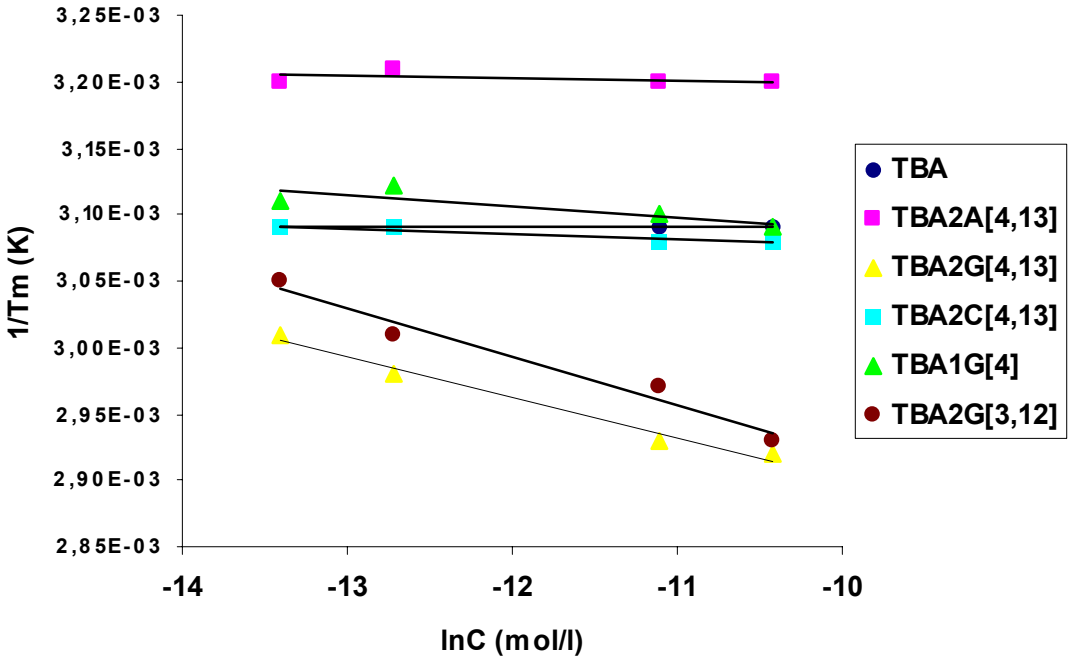


Fig. S3. Sketch of the free energy cycle used to compute the differential binding free energy of TBA and TBA2G dimer.

The cycle on the left hand side characterizes the differential binding free energy for the formation of the intramolecular quadruplexes (associated equations in the surrounded by continuous line), whereas the cycle on the right describes the formation of the intermolecular quadruplexes (dashed lines). For the calculations describing the intermolecular equilibrium we did not use the complete dimeric TBA sequence, but we kept the nearest neighbors in the parallel G-quadruplex core.

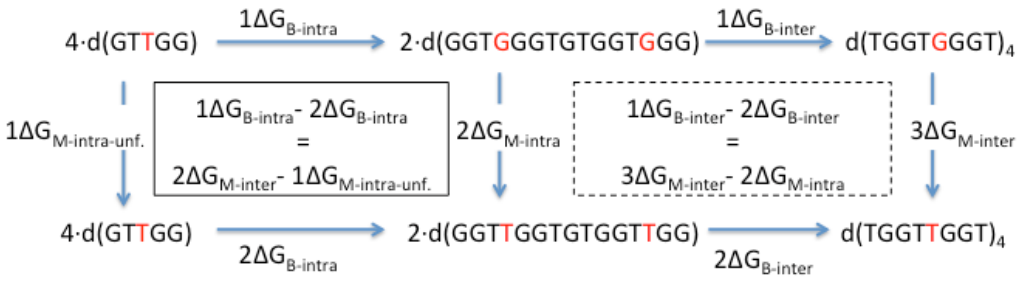


Fig. S4. Portion of NOESY spectrum of TBA2C[4,13] in 10 mM potassium phosphate buffer and 5 mM KCl (pH 6.7), *T* = 5 °C. Region displayed contains cross-peaks between hydrogen-bonded G imino protons and T9 methyl and H2' and H2'' resonances.

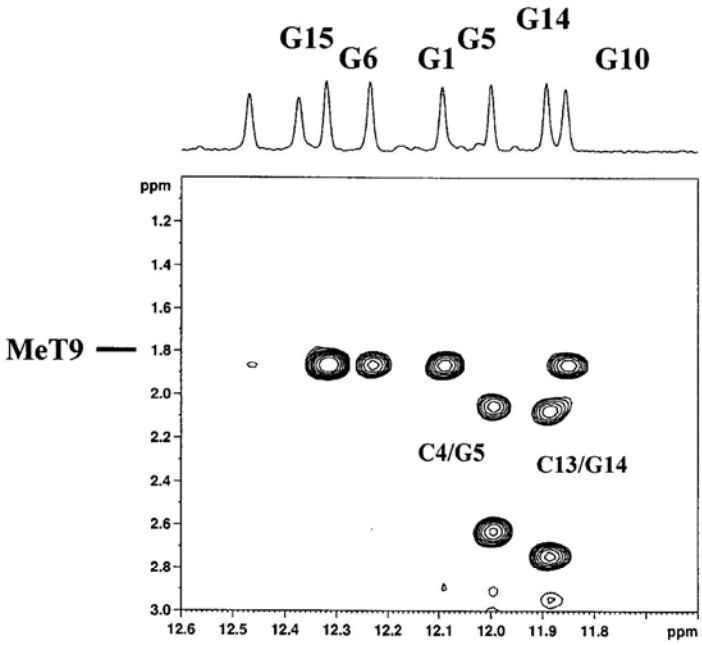
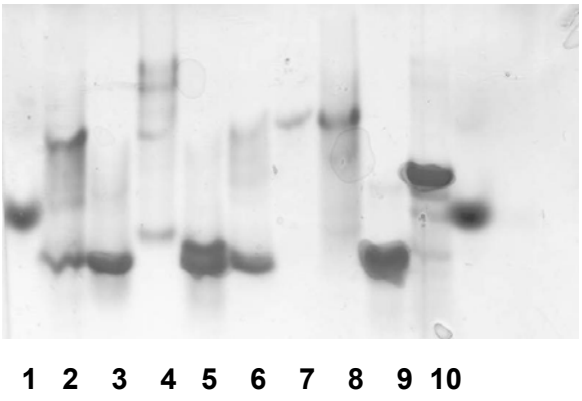


Fig. S5. 20 % PAGE of TBA derivatives



1: Bromophenolblue (BPB) dye; 2: TBA1G[4]; 3: TBA2C[4,13]; 4: TBA4G[3,4,12,13]; 5: TBA2A[4,13]; 6: TBA; 7: TBA2G[3,12]; 8: TBA2G[4,13]; 9: ODN 15 mer; 10: ODN 30 mer; 11: BPB dye.

TBA derivatives (0.3µM) were dissolved in 10 mM potassium phosphate buffer 100 mM KCl pH 7 and loaded on the native 20% polyacrylamide gel (acrylamide/bisacrylamide 19:1). The running buffer was the same. The gel was kept at 15ºC at 150 mV in a SE-600 Hoefer Scientific apparatus. After the electrophoresis, oligodeoxynucleotides were stained by STAINS-ALL (Sigma) using a solution of 20 mg of the dye in formamide/water (55:45) (100 ml). The remaining blue stains were visualized and photographed.

Fig. S6. Structures of TBA (A) and the parallel dimer of TBA2G (B) as found at the end of our gas-phase MD simulations of 400ns. Each structure corresponds to a different combination of charge distribution and initial structure. The structural rearrangement takes place mainly at the beginning of the simulations, the first 20 ns, although we can not discard further deviations on a much longer time scale. As explained in the text, the net charge of the TBA is -5, and TBA2G has a net charge of -10. A given charge state is achieved by protonation of the phosphate groups, and the exact distribution of the protonated states is listed in table S7.

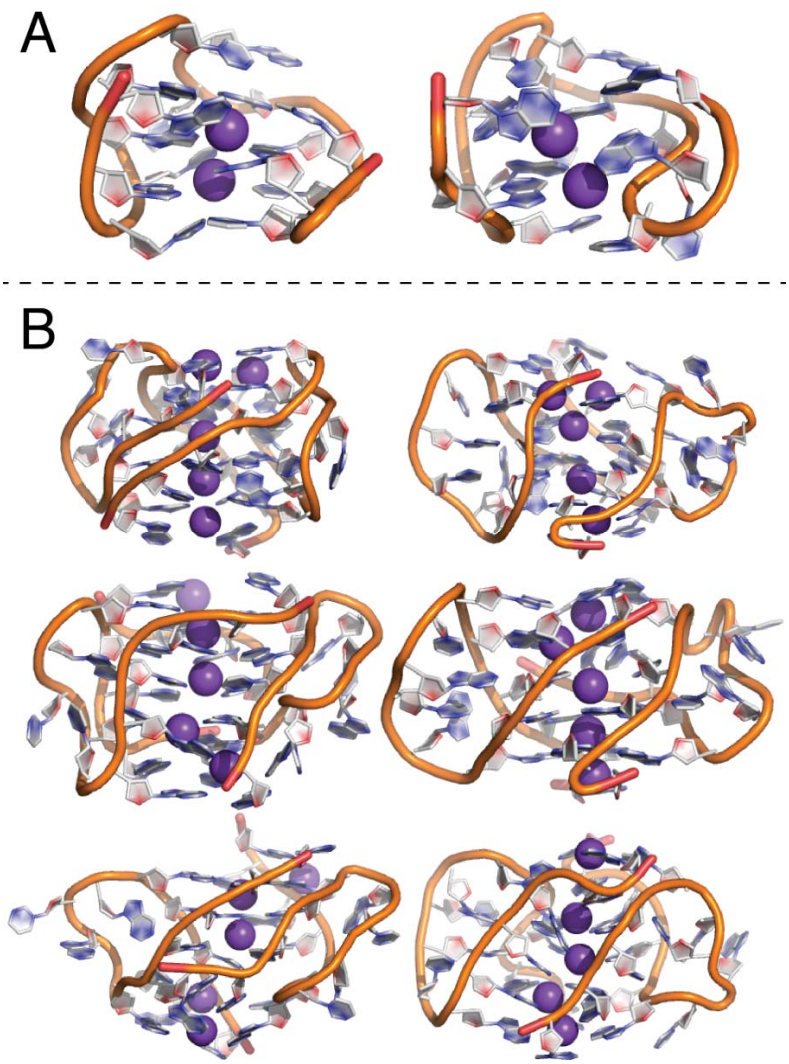


Fig. S7. Time evolution of the root mean square deviation (RMSD) with respect to the initial structure for monomeric form of TBA, TBA2G, TBA2A and TBA2C in water solution. In the time scale of the simulation all these structures are stable. The figures on the right hand side show the structures as found at the end of the simulations. The potassium ion that is always present at the center of the two G-tetrads as a purple sphere.

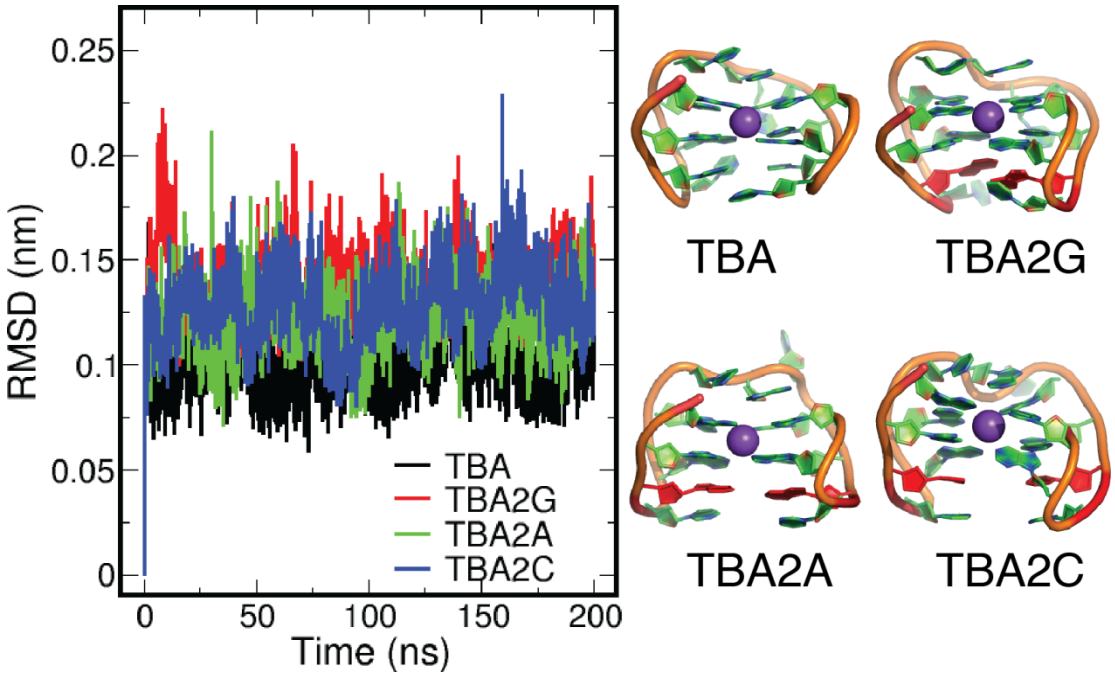


Fig. S8. From left to right in descending order, schematic illustration of the steps followed to build the simulation boxes containing intermolecular G quadruplexes, like $(TGGTGGT)_4$ or the dimer of TBA2G[4,13], as described in the text and in the SI material. Other setups involving intramolecular G-quadruplexes were constructed following a similar procedure.

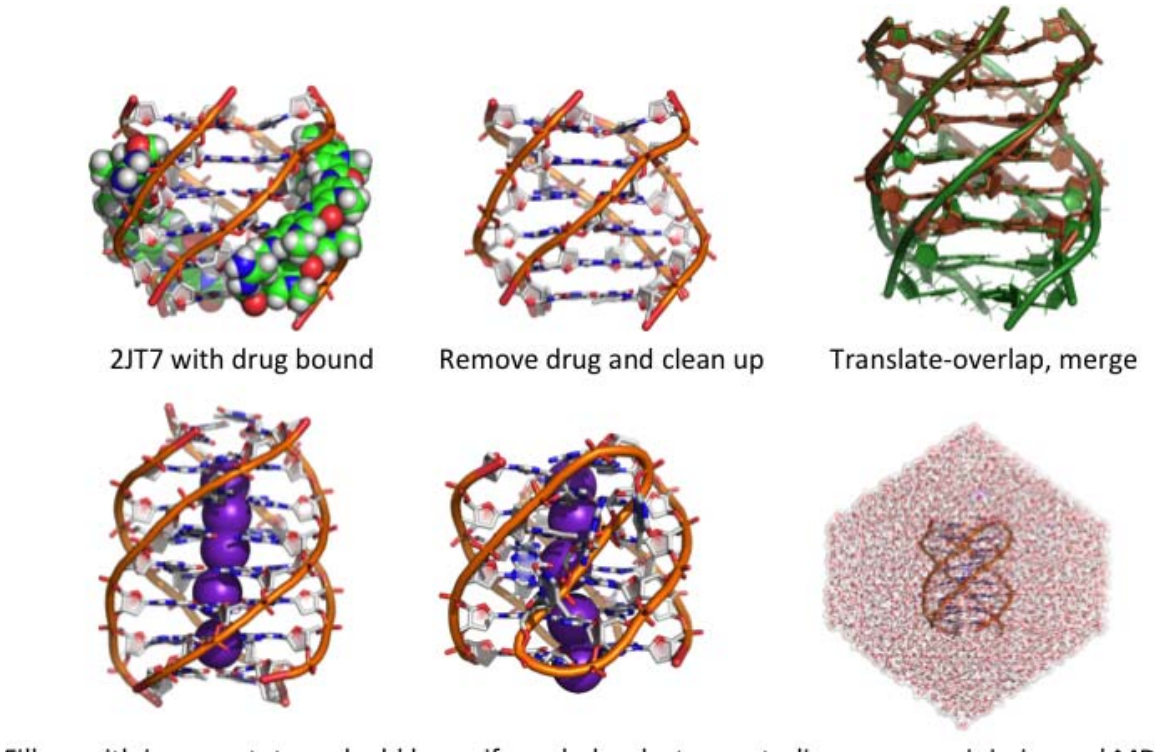


Table S1. **List of MD simulations performed in this work.**

For each system, we describe the number of replicates (slightly different initial structures and velocities) and the simulation length of each replicate. G-quadruplex dimers refer to the dimers folded in a parallel arrangement.

System	Replicates	Simulation time (ns)
Monomeric TBA	1	200
Mon. TBA gas phase	10	400
Monomeric TBA2G[4,13]	1	200
Monomeric TBA2A[4,13]	1	200
Monomeric TBA2C[4,13]	1	200
$(TGGTTGGT)_4$	1	700
$(TGGTGGGT)_4$	1	700
Dimer TBA	2	400
Dimer TBA2G[4,13]	2	400
Dim. TBA2G[4,13] gas phase	18	400

Dimer TBA2C[4,13]	1	400
Dimer TBA2A[4,13]	1	400
Dimer TBA4G[3,4,12,13]	1	400

Table S2. **List of residue numbers of un-protonated nucleotides along the TBA and TBA2G sequences in our gas-phase MD simulations.** In the TBA2G dimer, the residue numbers of the second strand are numbered as if they would form a continuous chain in the 5' to 3' direction.

Species	Charge distribution 1 (CD1)	Charge distribution 2 (CD2)	Charge distribution 3 (CD3)
TBA	3,5,8,10,13	2,4,8,12,14	-
TBA2G	3,5,8,10,13,18,20,23,25, 28	2,5,8,11,14,17,20,23,26, 29	3,6,9,12,15,18,21,24,27, 30

Table S3 Collision cross section (CCS, in Å^2) obtained using the exact hard-spheres scattering (EHSS) approximation[3] from the ensemble of structures generated in our gas phase MD simulations for TBA and TBA2G[4,13] dimer in their most populated charge states (5- and 10-, respectively, on the nucleic acids).

As explained in the text, we generated different charge distributions for a given charge state, and we performed independent simulations fore ach charge distribution and initial structure. The third column reports the time and replicate averaged for each charge distribution. The fourth column lists the averages over different charge distributions, and the fifth column displays the overall average, taking into account different initial structures. The last column gives the value of the CCS extracted from our ESI-MS experiments for the most populated charge states.

Structure	Charge distribution (CD)	CCS (Replicate time-averaged)	CCS averaged over CD	CCS averaged over full ensemble	Exp. value
TBA	CD-1 (5 replicates)	580 +/- 2	569 +/- 4	569 +/- 4	564 +/- 1

	CD-2 (5 replicates)	559 +/- 2			
TBA2G dimer st-1	CD-1 (3 replicates) CD-2 (3 replicates) CD-3 (3 replicates)	902 +/- 1 939 +/- 1 928 +/- 1	922 +/- 3	928 +/- 4	899 +/- 2
TBA2G dimer st-2	CD-1 (3 replicates) CD-2 (3 replicates) CD-3 (3 replicates)	927 +/- 1 926 +/- 1 949 +/- 1	933 +/- 3		

Table S4. Melting temperatures of TBA2G[4,13] in phosphate buffer 10mM with increasing potassium concentration

	5 mM KCl	50 mM KCl	100 mM KCl
Tm (ºC)[a]	44	50	54

[a] Concentration of TBA2G[4,13]= 3µM

Table S5. Melting temperatures and thermodynamic parameters obtained from denaturing curves in ammonium buffer for TBA derivatives.

Name	Tm (ºC) (ΔTm)[a]	ΔHº (Kcal/mol)	ΔSº (cal/Kmol)	ΔGº (25ºC) (Kcal/mol)
TBA	33.6	-39.1±0.5	-127±1.9	-1.1
TBA2A[4,13]	nd	-	-	-

TBA2C[4,13]	35.7(+2.1)	-29.7±0.8	-96.6±2.7	-0.9
TBA2G[4,13]	42.7(+9.1)	-52.4±1.7	-140.5±3.5	-10.5
TBA1G[4]	30.9(-2.7)	-31.6±0.4	-103.9±1.5	-0.6
TBA2G[3,12]	40.7(+7.1)	-44.8±1.0	-117.5±3.2	-9.8

[a] ΔTm is the Tm change in the melting process between the modified TBAs and the unmodified TBA.

Table S6. Melting temperatures and thermodynamic parameters obtained from denaturing curves in ammonium buffer for TBA and TBA2G[4,13] with an increasing percentage of MeOH. Uncertainty ± 0.5°C

TBA:

Name	Tm (ºC)	ΔHº (Kcal/mol)	ΔSº (cal/Kmol)	ΔGº (25ºC) (Kcal/mol)
TBA	33.6	-39.1±0.5	-127±1.9	-1.1
TBA 20%MeOH	42.7	-47.1± 0.6	-149.3±2.0	-2.6
TBA 40%MeOH	48.7	-54.9±0.5	-170.7±1.7	-4.0
TBA 60%MeOH	53.4	-54.5±0.6	-167.0±2.1	-4.7

TBA2G[4,13]:

Name	Tm (ºC)	ΔHº (Kcal/mol)	ΔSº (cal/Kmol)	ΔGº (25ºC) (Kcal/mol)
TBA2G[4,13]	42.7	-52.4±1.7	-140.5±3.5	-10.5
TBA2G[4,13] 20%MeOH	50.8	-54.1± 0.8	-141.5±2.5	-11.9
TBA2G[4,13] 40%MeOH	52.7	-70.1±1.7	-186.5±5.2	-14.5
TBA2G[4,13] 60%MeOH	67.8	-82.6±2.6	216.5±7.7	-18.0

Table S7. **Proton chemical shift of TBA2C[4,13] in 10 mM potassium phosphate buffer and 5 mM KCl (pH 6.7), T=5°C.**

The chemical shifts are unambiguously assigned: T3 is coincident with G8 and G8 was assigned by the NOE interaction with H1' T9. The same H1'T9 allow to assign the NH T9 which is found coincident with T12

	H8/H6	H1'	H2'	H2''	H3'	Me6/H5	NH
G1	7.45	6.15	2.90	2.90	4.96	--	12.09
G2	8.36	6.11	3.04	2.35	5.15	--	12.47[a]
T3	8.00	6.11	2.17	2.55	4.87	2.06	10.36

C4	7.06	6.07	2.06	2.62	4.88	6.84	9.45
G5	7.52	6.12	3.54	2.84	4.89	--	11.99
G6	7.75	5.98	2.90	2.63	5.13	--	12.23
T7	7.96	6.51	2.53	2.64	4.89	2.00	10.36
G8	7.53	5.80	2.03	2.36	4.83	--	10.36
T9	7.32	5.90	2.02	2.46	4.67	1.86	10.72
G10	7.46	6.11	3.80	2.90	4.95	--	11.85
G11	8.37	6.01	3.79	2.91	5.15	--	12.37[a]
T12	8.00	6.11	2.34	2.55	5.15	2.06	10.72
C13	7.09	6.11	2.08	2.75	5.15	6.95	9.85
G14	7.52	6.12	3.67	2.99	n.d.	--	11.89
G15	8.11	6.21	2.73	2.46	4.82	--	12.35

[a] NH G2 and NH G11 could be exchanged.

Table S8. Proton chemical shift of TBA2A[4,13] in 10 mM potassium phosphate buffer and 5 mM KCl (pH 6.7), T=5°C

	H8/H6	H1'	Me6 /H5	NH
G1	7.24	5.94	---	11.94
G2	8.12	5.97	--	12.21
T3	7.73	6.04	2.08	
A4			--	

G5	7.31	6.25	--	11.59
G6	7.55	5.95	--	11.94
T7	7.75	6.37	1.88	
G8	7.30	5.62	--	
T9	7.05	5.68	1.60	
G10				11.79
G11	8.12			11.56
T12				
A13	7.68			
G14	7.32	6.01		
G15	7.92	6.01		12.08